

POLITECNICO DI MILANO  
Facoltà di Ingegneria Industriale

Corso di laurea in Ingegneria Aeronautica



**RESIDUAL BASED DISCRETIZATION OF  
MADSEN AND SØRENSEN SYSTEM OF  
BOUSSINESQ EQUATIONS FOR  
NON-HYDROSTATIC WAVE  
PROPAGATION**

Relatore:

Prof. Alberto Matteo Attilio GUARDONE

Correlatore:

Prof. Mario RICCHIUTO

Tesi di Laurea di

Andrea Gilberto FILIPPINI

Matr. 762975

Anno Accademico 2011-2012



# Contents

<b>Abstract</b>	<b>ix</b>
<b>Sommario</b>	<b>xi</b>
<b>Introduction</b>	<b>1</b>
<b>1 Governing Equations</b>	<b>7</b>
1.1 Nonlinear Shallow Water Equations (NSWE) . . . . .	7
1.2 Madsen and Sørensen Boussinesq-type Equations . . . . .	10
1.3 Model Dispersion Analysis . . . . .	15
<b>2 Weighted Residuals Methods</b>	<b>23</b>
<b>3 Linear Scalar Advection Equation</b>	<b>29</b>
3.1 Finite Difference Methods . . . . .	30
3.2 Galerkin Finite Element Method . . . . .	34
3.3 Residual Distribution Method . . . . .	39
3.4 Stabilized Upwind Methods . . . . .	43
3.5 Conclusions . . . . .	51

<b>4</b>	<b>Linear Scalar Advection with Dispersion Equation</b>	<b>53</b>
4.1	Finite Difference Methods . . . . .	54
4.2	Galerkin finite element method . . . . .	57
4.3	Residual distribution method . . . . .	59
4.4	Stabilized Upwind methods . . . . .	64
4.5	Conclusions . . . . .	70
<b>5</b>	<b>Madsen and Sørensen system discretisation</b>	<b>73</b>
5.1	Finite difference methods . . . . .	74
5.2	Galerkin finite element method . . . . .	79
5.3	Residual Distribution . . . . .	85
5.4	Stabilized Upwind methods . . . . .	90
5.5	Conclusions . . . . .	98
<b>6</b>	<b>Time discretization</b>	<b>101</b>
6.1	Crank-Nicolson time integration method . . . . .	103
6.2	<i>BDF3</i> time integration method . . . . .	110
<b>7</b>	<b>Numerical implementation</b>	<b>113</b>
7.1	Soliton Wave Generator . . . . .	114
7.2	Boundary conditions . . . . .	119
7.3	Internal Wave Generation . . . . .	121
<b>8</b>	<b>Numerical Tests and Results</b>	<b>125</b>
8.1	Convergence Analysis . . . . .	125

8.2	Head-on Collision of Two Solitary Waves . . . . .	130
8.3	Propagation over a Shelf . . . . .	134
8.4	Periodic Wave Propagation over a Submerged Bar . . . . .	139
	<b>Conclusions</b>	<b>143</b>
A	<b>Demonstration of the Equality 2.5</b>	<b>149</b>
B	<b>Auxiliary Variables Choice for the Linear Advection-Dispersion Problem</b>	<b>151</b>
C	<b>Space-Discretization of the Linearized MS System</b>	<b>155</b>
D	<b>Auxiliary Variables Choice in the MS Dispersion Term Discretization</b>	<b>159</b>
E	<b>Galerkin Space-Discretization of the MS System</b>	<b>161</b>
F	<b>Space-Time Discretization of the Linearized MS System</b>	<b>163</b>
	<b>Bibliography</b>	<b>165</b>



# List of Figures

1.1	Schematic representation of the advection and convergence contribution in equation (1.2). In the figure $H$ is assume to locally increase.) . . . . .	8
1.2	Sketch of the free surface flow problem, main parameters description. . . . .	11
1.3	Free surface time history illustrating a soliton wave propagation over a bed variation ( $g$ ); comparison between the solutions computed by means of NSWE (left) and MS (right) models. .	13
1.4	Free surface time history illustrating a soliton wave propagation over a bed variation ( $g$ ); comparison between the solutions computed by means of NSWE (left) and MS (right) models. .	14
1.5	Comparison of the dispersion characteristics of the Peregrine's Boussinesq, extended Boussinesq and shallow water models with respect to the linear wave theory, based on the representation of the percentage of error in phase velocity computation.	20
2.1	Representation of the linear finite basis function at the node $i$ .	25
3.1	Dispersion $w$ against the wavenumber $k$ of the finite difference schemes $\mathcal{FD}2$ and $\mathcal{FD}4$ , compared to the scalar linear advection model one, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	34
3.2	Representation of the linear finite basis function over the element $e$ . . . . .	35

3.3	Dispersion $w$ against the wavenumber $k$ of the Galerkin finite element scheme, compared to the scalar linear advection model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	38
3.4	Dispersion $w$ against the wavenumber $k$ of the central $\mathcal{RD}$ scheme, compared to the scalar linear advection model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	43
3.5	Representation of the upwind linear finite basis function over the element $e$ . . . . .	46
3.6	Dispersion $w$ against the wavenumber $k$ of the stabilized upwind schemes, compared to the scalar linear advection model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	50
3.7	Diffusion $\xi$ against the wavenumber $k$ of the stabilized upwind $SU/\mathcal{PG}$ and $U/\mathcal{RD}$ schemes, compared to the scalar linear advection model, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	50
4.1	Dispersion $w$ against the wavenumber $k$ of the finite difference $\mathcal{FD}2$ and $\mathcal{FD}4$ schemes, compared to the scalar linear advection with dispersion model, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	56
4.2	Dispersion $w$ against the wavenumber $k$ of the Galerkin finite element scheme, compared to the scalar linear advection with dispersion model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	59
4.3	Dispersion $w$ against the wavenumber $k$ of the central $\mathcal{RD}$ scheme, compared to the scalar linear advection with dispersion model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	63

4.4	Dispersion $w$ against the wavenumber $k$ of the stabilized upwind $SU/PG$ and $U/RD$ schemes, compared to the scalar linear advection with dispersion model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	69
4.5	Diffusion $\xi$ against the wavenumber $k$ of the stabilized upwind $SU/PG$ and $U/RD$ schemes, compared to the scalar linear advection with dispersion model, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	70
5.1	Dispersion $w$ against the wavenumber $k$ of the finite difference $FD2$ and $FD4$ schemes, compared to Airy theory value (1.25), when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	78
5.2	Dispersion $w$ against the wavenumber $k$ of the Galerkin finite element scheme, compared to Airy theory value and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	84
5.3	Dispersion $w$ against the wavenumber $k$ of the central $RD$ scheme, compared to Airy theory value and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	89
5.4	Dispersion $w$ against the wavenumber $k$ of the stabilized upwind $SU/PG$ and $U/RD$ schemes, compared to Airy theory value and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	97
5.5	Diffusion $\xi$ against the wavenumber $k$ of the stabilized upwind $SU/PG$ and $U/RD$ schemes, compared to Airy theory value and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes $N = 5$ . . . . .	97
7.1	Numerical solution of the integration of the second order differential equation of $q$ using the Matlab <sup>®</sup> solver <i>ode113</i> . . . . .	116

7.2	(a): attempts of reconstruction of the tail of the soliton shape by means of the numerical integration of the first order equation (7.7); (b): close-up of the interest region. . . . .	117
7.3	Optimal solution found for the soliton tail reconstruction and comparison between the final and the initial configuration (a); close-up of the interest region (b). . . . .	118
7.4	(a): evolution of the exact $a = 0.2$ m solitary wave; (b): comparison between the exact and the numerically computed solitary wave plotted against the spatial domain at the final time of the simulation. . . . .	119
7.5	sketch of the internal wave generation process. . . . .	122
8.1	Spatial distribution of the error function $e_\eta(x)$ over the domain, parametrised on the growing number of nodes of used for the discretisation (a) and close-up of the region of major contribution to the global error value (b). . . . .	126
8.2	Convergence curves computed for the schemes integrated in time with the Crank-Nicolson method. . . . .	128
8.3	Convergence curves computed for the schemes integrated in time with the $BDF3$ method. . . . .	129
8.4	Free surface profiles of solitary waves with $a = 0.2$ , propagating in opposite directions in a channel of constant depth, plotted at different times of the simulations and finally comparing with the theoretical result. . . . .	131
8.5	Overlapping of the several results obtained with the different developed methods and close-up of the peak regions to highlight the difference between the solutions computed. . . . .	133
8.6	Sketch of the submerged shelf test. . . . .	134
8.7	Time evolution of solution computed for the submerged shelf test. The time step between two consecutive figures is not constant, in order to highlight the most relevant phenomena. . . . .	135

8.8	Splitting of a solitary wave propagating over a submerged shelf: (a) numerical computation using $SU/\mathcal{PG}$ scheme with $\mathcal{CN}$ time integration; (b) test results presented in [21]. . . . .	137
8.9	Overlapping of the several results of the submerged shelf test, obtained integrating the different schemes with the $\mathcal{CN}$ method (a), and close-up of the peaks regions to highlight the difference between the solutions computed (b). . . . .	138
8.10	Overlapping of the several results of the submerged shelf test, obtained integrating the different schemes with the $BDF3$ method (a), and close-up of the peaks regions to highlight the difference between the solutions computed (b). . . . .	139
8.11	Sketch of the computational configuration of the numerical test of the propagation over a submerged bar. . . . .	140
8.12	Time evolution at the gauges coordinates of the surface elevation of sinusoidal wave encountering a submerged bar, with initial amplitude $a = 0.01$ m and period $T = 2.02$ s. In figure, $\circ$ represents the experimental data, $\text{---}$ is the central Galerkin, $\text{---}$ is the $SU/\mathcal{PG}$ and $\text{---}$ is the $U/\mathcal{RD}$ scheme. . . . .	141
8.13	Time evolution at the gauges coordinates of the surface elevation of sinusoidal wave encountering a submerged bar, with initial amplitude $a = 0.01$ m and period $T = 2.02$ s. In figure, $\circ$ represents the experimental data, $\text{---}$ is the central Galerkin, $\text{---}$ is the $SU/\mathcal{PG}$ and $\text{---}$ is the $U/\mathcal{RD}$ scheme. . . . .	142



# Abstract

The accurate numerical simulation of wave propagation in near-shore zones requires to consider both highly nonlinear and dispersive wave interactions in order to accurately predict wave diffraction, reflection and harmonic generation. Boussinesq equations are the simplest model able to offer a mathematical description of all these effects. Among the great variety of Boussinesq-type mathematical models, the set of extended Boussinesq equations of Madsen and Sørensen (MS) is here considered, which is valid over a large range of applicability, thank to a better approximation of the physical dispersion parameters of the problem.

A technique to separate the space and the time discretizations of the problem, formulating them one after the other, is used. Numerical weighted residual based methods are proposed in order to spatially discretize the MS system and are, for semplicity, firstly developed and applied to the simple scalar linear advection and scalar linear advection-dispersion models. In particular, a linear finite element approximation of the solution is used in order to implement a central Galerkin and a Residual Distribution ( $\mathcal{RD}$ ) schemes, which are then stabilized by means of a residual upwind term, producing a kind of Streamline Upwind Petrov-Galerkin ( $SU/PG$ ) scheme on one hand, and an Upwind Residual Distribution ( $U/\mathcal{RD}$ ) scheme on the other.

The Madsen and Sørensen system is rewritten in a form suitable for the application of the previously mentioned numerical methods. These preserve the accuracy and dispersion properties investigated for the previous simpler models. The integration in time is accomplished by implementing the Crank-Nicolson and the third order Backward Differentiation Formula ( $BDF3$ ) techniques, using Newton's iterations to solve the implicit character of these methods.

The performance of the numerical models are demonstrated by means of a comparison with theoretical and experimental results, showing an excellent agreement, which confirms also that, within the shallow water framework, a linear finite element approximation method is sufficiently accurate for the representation of the wave propagation problem.



# Sommario

Lo studio della propagazione ondosa vicino alla costa, dove la forma delle onde viene fortemente influenzata dall'andamento della batimetria, non può essere compiuto utilizzando il celebre *Nonlinear Shallow Water System*. Infatti, processi di natura dispersiva come la diffrazione, la riflessione e l'interazione armonica fra onde non vengono descritti e rappresentati da questo tipo di equazioni. Le equazioni di Boussinesq offrono la più semplice modellizzazione matematica di tali effetti ma sono comunque limitate a campi debolmente nonlineari e dispersivi. Per estenderne il campo di applicabilità, svariati modelli sono stati elaborati nel tempo sulla base delle equazioni di Boussinesq costruite da Peregrine. I modelli di Nwogu e Madsen-Sørensen sono quelli che hanno ricevuto maggiore attenzione dalla comunità scientifica.

In questa tesi verrà mostrato uno studio delle proprietà dispersive dei modelli menzionati, motivando in questo modo la scelta di utilizzare il modello proposto da Madsen e Sørensen per creare uno schema numerico in grado di risolvere, ad alti gradi di accuratezza, il problema della descrizione dei maggiori effetti della propagazione ondosa vicino alle coste.

In seguito, ci si porrà il problema di integrare e risolvere in via numerica il sistema monodimensionale di Madsen-Sørensen. A questo scopo verrà utilizzata una tecnica per discretizzare separatamente, prima nello spazio e poi nel tempo, il sistema di equazioni scelto. Il metodo numerico da realizzare dovrà risultare potenzialmente estendibile al caso bidimensionale, con la possibilità di eseguire il calcolo su griglie non strutturate e adattive. L'utilizzo di griglie adattive per seguire il movimento del fronte d'onda e della linea d'acqua sulle coste servirebbe ad offrire una risoluzione migliore di quanto non si potrebbe ottenere tramite una griglia fissa, a pari numero di elementi. Inoltre nella selezione del metodo numerico si è tenuto conto dei precedenti lavori svolti sul sistema *Shallow Water* poichè, una volta introdotto un criterio per l'individuazione del cosiddetto *wave breaking*, sarà necessario risolvere tale sistema, e non più quello di Boussinesq, per gestire la propagazione degli urti.

In seguito a queste considerazioni, e a una generica introduzione ai metodi ai residui pesati, verranno presentati nella tesi i metodi centrati agli elementi finiti lineari di Galerkin e ai residui ( $\mathcal{RD}$ ) e subito applicati ai più semplici problemi di advezione e advezione-dispersione scalare lineare, confrontando le loro proprietà di accuratezza e dispersione con quelle dei comuni metodi alle differenze finite del secondo e quarto ordine. In questo modo i vantaggi dovuti alla presenza della matrice di massa di Galerkin saranno evidenziati. A valle di considerazioni sulla stabilità numerica nella discretizzazione di problemi dominati dal flusso advettivo, tali schemi saranno stabilizzati mediante l'introduzione di un termine *upwind* ai residui pesati in grado di produrre, attraverso una particolare scelta dei pesi, un metodo *Streamline Upwind Petrov-Galerkin* ( $SU/PG$ ), attraverso la stabilizzazione del Galerkin centrato, ed un metodo *upwind* completamente ai residui ( $U/RD$ ), se applicato al precedente schema  $\mathcal{RD}$  centrato. Verrà inoltre mostrato come l'introduzione dell'*upwinding* non alteri l'accuratezza del corrispondente metodo centrato, avendo anche un debole effetto sulle proprietà dispersive. Nel caso  $SU/PG$ , esso addirittura migliora l'approssimazione della dispersione del modello continuo operata dallo schema, ma si registra tuttavia l'aggiunta di una diffusione artificiale che invece non è riscontrata nel caso  $U/RD$ , in cui le caratteristiche dispersive e diffusive rimangono pressochè identiche a quelle del corrispettivo metodo centrato.

Verrà mostrato come riscrivere il più complesso sistema di equazioni di Madsen-Sørensen in una formulazione contenente più bassi ordini di derivazione spaziale, definendo e introducendo una o più variabili ausiliare. Verrà dimostrato come tale scelta non incida sul grado di accuratezza della soluzione e anche che venga fatta nell'ottica di ridurre al minimo l'aumento del costo computazionale dovuto all'introduzione di nuove incognite nel sistema.

In questo modo i metodi sviluppati potranno essere direttamente applicati al modello di Madsen-Sørensen, preservando gli ordini di accuratezza e le proprietà di stabilità precedentemente osservate. L'integrazione temporale delle equazioni verrà effettuata utilizzando il metodo di Crank-Nicolson ( $\mathcal{CN}$ ) e Backward Differential Formula del terzo ordine ( $\mathcal{BDF3}$ ). La caratteristica implicita di questi schemi verrà risolta attraverso il processo iterativo di Newton, la cui soluzione ad ogni passo temporale rappresenta la parte con il più alto costo computazionale, motivo per cui si adotterà una tecnica di congelamento della matrice Jacobiana del sistema per alleggerire la procedura.

Uno studio di convergenza sarà realizzato per indagare il vantaggio dell'impiego del metodo  $\mathcal{BDF3}$ , di più alto ordine, rispetto all'A-stabile  $\mathcal{CN}$ . L'insorgenza di instabilità si rileva, infatti, negli schemi centrati rispetto ai più robusti metodi che utilizzano l'*upwind*, giustificando così l'introduzione del termine di stabilizzazione e confermando la bontà del suo impiego.

Attenzione verrà posta sull'implementazione delle condizioni iniziali e al contorno. Ci si soffermerà sulle modalità di calcolo della forma esatta del solitone da utilizzare per i test seguenti e si spiegherà come condizioni al contorno periodiche e di uscita, tramite l'assorbimento senza riflessione delle informazioni che raggiungano il bordo, sono state applicate ai metodi proposti. Viene inoltre mostrata una tecnica per la generazione di onde periodiche impiegate per il confronto con i dati sperimentali reperiti in letteratura. Infine, le proprietà di nonlinearità e dispersione dei metodi saranno testate e confrontate con risultati teorici e sperimentali, mostrando un eccellente accordo con essi.



# Introduction

The interest in accurate simulations of nonlinear and dispersive water waves in realistic environments is rapidly increasing in coastal and ocean engineering for the study of the impact of waves on the coasts or the design of coastal structures. The mathematical and numerical modelling of these phenomena has advanced in the last decades, improving the capacity of accurately predicting near-shore wave processes, including shoaling and runup, diffraction, reflection and harmonic interaction. Due to the simplicity whereby different layouts can be constructed and tested, compared to the longer times and higher cost for rebuilding physical models' operations, numerical modelling have now largely replaced laboratory experiments, even if the latter are needed for numerical models validation.

As it is well known, the numerical solution of a realistic problem would require the solution of the incompressible Navier-Stokes system of equations over a large three-dimensional spatial domain with a free-surface boundary. To avoid the complexity and computational resources needed to solve the full Navier-Stokes system of equations, the shallow water nature of this kind of problem has been used to simplify the governing equations through a depth averaging approximation and a two-dimensional restriction. Therefore, numerical models for wave propagation and interaction are usually based on the so-called depth-averaged shallow water equations (NSWE). This model reproduces fairly well the important aspects of wave propagation and the general characteristics of the runup processes, although it is not appropriate in water environments where dispersion effects become more important than nonlinearity. As a consequence, near-shore wave propagation and transformation cannot be modelled within the shallow water system of equations and higher order models have to be considered. Indeed, waves generated in open ocean undergo drastic changes close to shore, where the shape of the free surface is affected by the shape of the sea bed and an accurate prediction of wave activity requires to take into account both nonlinear and dispersive effects.

Using the free surface displacement and the depth averaged velocity as dependent variables, Peregrine derived the standard Boussinesq equations for variable depth. This model includes dispersion terms and it is more suitable to represent waters where dispersion affects the free surface. Numerical models built upon the Boussinesq-type equations include nonlinearity and frequency dispersion and have a potential to handle near-shore processes. However they are limited to weakly nonlinear and weakly dispersive waves assuming  $\mathcal{O}(\epsilon) = \mathcal{O}(\mu^2)$ , where nonlinearity  $\epsilon$  represent the ratio of wave amplitude to depth, and dispersion  $\mu$  is the ratio of water depth to wavelength. Standard Boussinesq approximations break down when the depth is larger than one fifth of the equivalent deep water wavelength and as such are limited to relative shallow water. In addition, the weakly nonlinear assumption limits the largest wave height that can be accurately modelled. Considerable efforts have been made in the last years to extend the validity and applicability of this class of equations to higher water depths and to shorter waves. Several alternative formulations have been presented, however, the ones developed by Madsen-Sørensen and Nwogu received the most attention of the scientific community. The set of equations considered by Madsen and Sørensen introduces a Padé approximation of the linear dispersion relation into the momentum equation. This originates extra third-order terms in the momentum equation with a free parameter, and results in equations suitable for water as deep as  $\mu = 0.5$ . The other set of equations, derived by Nwogu from the three dimensional Euler equations, is formulated in terms of the surface elevation and the horizontal velocity evaluated at a reference depth, chosen to minimize wave propagation errors from linear theory. The two approaches have identical linear dispersion characteristics that show good agreement with the linear wave theory and, as all Boussinesq-type equations, satisfy approximate conservation laws, unlike the nonlinear shallow-water equations which satisfy exact conservation laws for non-dispersive waves. Although these extended systems have improved dispersion characteristics, they are formally of the same accuracy as the original system and hence restricted to the shallow water environment. Moreover, significant efforts have been made in recent years towards the advancement of the nonlinear and dispersive properties of Boussinesq-type models by including higher order nonlinear and dispersive terms [25], [23] which, in turns, are more difficult to integrate and requires a longer computational effort.

In conclusion, the NSWE system is usually suitable for solving the wave propagation problem far from the coast and near the interface between the water and the coast, since it is able to reproduce accurately the discontinuities that originate at the shoreline. On the contrary, a system of equations

of Boussinesq-type is required to handle nonlinearities and dispersion effects. The goal of the present work is to develop a numerical weighted residual based method for the one-dimensional Madsen and Sørensen model, accurate and robust respect to the choice of the time-step. Several methods are going to be presented and compared with respect to their truncation and dispersion error analysis and stability properties. This work is part of an ongoing effort at INRIA (Institut National de Recherche en Informatique et en Automatique) for the development of a future numerical integration scheme for the Madsen-Sørensen system of equation over unstructured two-dimensional meshes for the description of the wave propagation and transformation in the near-shore region. The final code will have to be coupled with a NSWE solver, to be used outside the properly near-shore range and for the description of the wave breaking phenomena. In order to accomplish this, an investigation on the existing numerical schemes which well solved the nonlinear shallow water problem has been done, with the proposal to then extend and apply them to the Madsen and Sørensen problem.

Several numerical techniques have been developed in the last decades for solving NSWE system. The most widely used are based on a finite volume discretization with high order reconstruction and the use of limiters in order to guarantee the absence of oscillations in the numerical solution [6], [27], [4], [31], [2].

Interesting alternatives more recently emerged are based on a discontinuous or continuous finite element representation of the variables. The discontinuous Galerkin finite element method, or  $\mathcal{DG}$  method, is based on a variational formulation of the equations. After integration by parts, the fluxes in the integrals over cell boundaries are replaced by numerical fluxes in order to take into account the discontinuity of the interpolation [44], [13]. Limiters can be used to handle discontinuities [4], [43]. Methods based on a continuous finite element interpolation are instead obtained via a global variational formulation of the problem with the addition of proper stabilizing terms proportional to the element residual [8], [20], [17], [18], [16], [37], [38]. Compared to  $\mathcal{DG}$ , these have the advantage of requiring a small number of unknowns for a given formal order of accuracy due to the continuous interpolation of the degrees of freedom. In particular we have that, taking into account a square domain partitioned on small square elements, such that in each edge of the domain the number of the elements is  $N$ , the  $Q^1$  bilinear continuous finite element approximation have to deal with  $(N+1)^2$  unknowns, which reflects the number of degrees of freedom of the mesh. The  $\mathcal{DG}$  method, instead, since it has to approximate the nodal value of the solution over each of the cells adjacent to the node itself, results in roughly  $4N^2$  unknowns, meaning about four

times more than the respective continuous finite element approach. Among these methods we can mention the streamline upwind Galerkin, or  $SU/P\mathcal{G}$ , and the Galerkin least square, or  $\mathcal{GLS}$  [5]. In the following we will also refer to these methods as continuous residual based, or continuous weighted residual methods.

In this framework, the aim of this work is to investigate the application of the weighted residual methods, applying them on the solution of the extended Boussinesq systems, in particular of the Madsen and Sørensen model. The requirements that our method must fulfil are: the capability of dealing with an unstructured mesh, to allow local mesh adaptation especially in two-dimensional problems; high accuracy and especially low dispersion error; efficiency and robustness with respect to the choice of the timestep in the shallow water limit. The developed methods must also have some potential for including discontinuity capture capabilities and robust treatment of the moving shorelines.

Most of the numerical techniques already developed for the Boussinesq equations are based on the finite difference ( $\mathcal{FD}$ ) methods, e.g. [40], [11], [35], [34], [32], [15]. The popularity of these schemes derives from the ease whereby high order derivatives can be approximated and to the structure of the associated linear systems which can be efficiently solved. However, the main drawback of these schemes is that structured grids have to be used in two-dimensional simulations, even in modelling irregular geometries, fact which can lead to a loss of accuracy. In two-dimensional complex domains, the use of unstructured spatial grids, which can be locally adapted to the geometrical features, depth profiles or complex boundaries gives many advantages and has been put forward as a strategy to obtain more cost-effective models. It was estimated in [33] that the potential reduction factor of the cost, compared to structured meshes, is of the order of 10-20. The computation over unstructured meshes is not possible with these schemes, which thus do not comply with the first requirement stated above.

The most natural candidates for unstructured methods are finite element (FE) or finite volume (FV) schemes. The finite volume discretization methods lead usually to a complex and expensive reconstruction procedure especially on unstructured meshes, even in the absence of high order derivatives, which, however, characterise the extended Boussinesq models [28], [41]. In fact, in order to compute the value on a single element of the mesh, it could become necessary to take into account the contributions of not only the cells adjacent to the node itself, but also of those adjoining to the first group of cells. This large stencil will become even wider if more accurate solutions are requested. Theoretically a computational method would have, instead, a

stencil as compact as possible, capable of guaranteeing the properties already expressed, i.e. low dispersion error and ability both to capture discontinuities and to provide an accurate and robust description of the shorelines.

The use, instead, of FE methods in the solution of the Peregrin's Boussinesq equations ([10]) and of the extended Boussinesq-type models has increased in the last ten years with promising results in terms of accuracy and efficiency; see for example the  $\mathcal{DG}$  works [9], [3], or [45], [33], [42].

A good candidate would be, thus, a method based on a continuous finite element approximation, having a stencil as compact as possible and including stabilization terms providing robust approximation in the shallow water limit. [38], [36] respect all these criteria. In the present work we are going to extend their methods, developed on the nonlinear shallow water model, and to apply them on the hydrostatic case, in the Madsen and Sørensen context.

**Outline of the Thesis :** The numerical models used to solve the propagation of sea waves toward the coast are introduced in the next Chapter 1. There, the Nonlinear system of Shallow Water Equations is first developed and the dispersion properties of this model are investigated and compared with respect to the Madsen and Sørensen ones. In this way, the extended Boussinesq model is shown to be more near to the trend of dispersion assumed by the Airy wave theory, extending the range in which the model is applicable to compute the solution of the propagation problem in the near-shore zone. Many different spatial discretizations are possible although only those that can be extended to 2D unstructured triangular grids will be considered here in detail. A general introduction on the weighted residual methods will be presented in Chapter 2. In the following, numerical methods, elaborated to finally discretize the one-dimensional Madsen and Sørensen model, will be first introduced to spatially discretize the simpler linear scalar advection and linear scalar advection-dispersion problems, respectively in Chapter 3 and 4. The accuracy of these numerical schemes and the analysis of their dispersion properties are also going to be accomplished, and the analogies which emerge from the methods development are going to be largely discussed and investigated.

The numerical modelling of the one-dimensional form of the MS equations will be considered in Chapter 5. Numerical models developed in the previous chapters can be directly applied to this more complex problem and their performance will be shown.

The problem to integrate in time the semi-discretized schemes obtained in Chapter 5 is going to be discussed in the following Chapter 6. There, two

implicit linear multistep methods will be implement, paying the cost of solving an iterative cycle with a Newton kind method, in order to take advantage of their stronger stability properties with respect to the explicit methods. Through the use of the Crank-Nicolson ( $CN$ ) method, which is an A-stable second order method in time, more robust schemes want to be obtained even if the upwind stabilization is not present. The third order Backward Differentiation Formula ( $BDF3$ ), instead, is going to be implement in order to reach better convergence orders, affecting only weakly the stability of the scheme.

Finally, particular attention is then posed in Chapter 7 to the numerical setting of the initial and boundary conditions for the techniques developed. The computational instruments, useful to realize the numerical tests are illustrated and the schemes validation is going to be accomplish, in Chapter 8, by comparison with theoretical and experimental data, testing the goodness of the nonlinear and dispersive properties shown all along the work.

# Chapter 1

## Governing Equations

In this chapter the Nonlinear Shallow Water system of Equations (NSWE) and the extended Boussinesq Madsen and Sørensen model (MS) are going to be presented. We are going to define the variables of the problem and to give an explanation of the various terms and parameters which compose the equations of the two systems. Models analogies and differences, mentioned in the previous introduction, are going to be investigated more in detail in this chapter. We will test the capabilities of describing a soliton wave propagation over a non-constant bathymetry in the near-shore zone by the two methods, showing and commenting the results. At the end of the chapter, a dispersion analysis of the models is presented and a comparison with respect to the linear wave theory is carried out. The results will justify the application of MS system in the near-shore wave propagation problems and will give the exact expression of the dissipation and dispersion parameters of the models, which will be the references for the following discretization schemes development.

### 1.1 Nonlinear Shallow Water Equations (NSWE)

The nonlinear shallow water equations are a non-dispersive depth-averaged approximation of the three-dimensional water wave problem, in the hypothesis of a horizontal length scale much larger than the vertical one. They are therefore applicable to problems where small-amplitude waves propagate in a fluid which is shallow with respect to the wave length. Under this condition, the vertical velocity of the fluid is negligible and the horizontal velocity  $u$  is roughly constant throughout the depth of water layer. By assuming incom-

pressible flow, and uniform density  $\rho$ , the continuity equation takes the form of the following nonlinear scalar advection equation for the variable  $H$ :

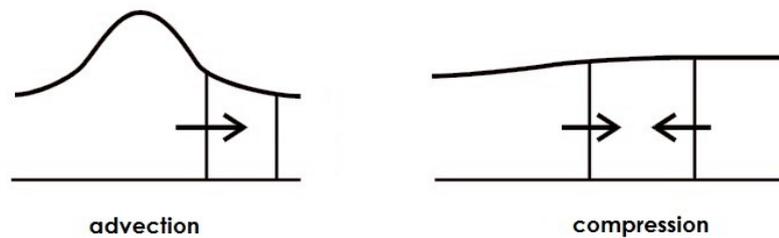
$$\partial_t H + \partial_x(uH) = 0 \quad (1.1)$$

where  $H$  is the total water depth and the operators  $\partial_t$  and  $\partial_x$  denote partial derivatives with respect to time and space.

By differentiating by parts the second term and rearranging, we obtain:

$$\partial_t H + u\partial_x H + H\partial_x u = 0 \quad (1.2)$$

This equation expresses the local rate of change of the surface height in terms of two contributions: the first term,  $u\partial_x H$ , represents the advection of height  $H$ , the second one,  $H\partial_x u$ , expresses the volumetric variation in terms of compression and expansion. The two contributions are depicted schematically in Fig. 1.1.



**Figure 1.1:** Schematic representation of the advection and convergence contribution in equation (1.2). (In the figure  $H$  is assumed to locally increase.)

It can be shown from the momentum equation that vertical pressure gradients do not differ from hydrostatic ones; the hydrostatic pressure arises simply from the weight of the fluid above, due to the action of the gravitational acceleration  $g$ , and this is the force that re-establishes the equilibrium of the system (*gravity waves*). Indeed, the horizontal pressure gradients are due to the displacement of the pressure surface, implying that the horizontal velocity field is constant throughout the depth of the fluid. In such a context,

integrating over the fluid depth allows the vertical velocity to be removed from the equations. Excluding the effect of friction or other source terms beside the ground bathymetry, the momentum equation takes the form:

$$\frac{du}{dt} + g\partial_x H = 0 \quad (1.3)$$

Here  $g$  is simply the magnitude of gravitational acceleration, while the derivative  $d/dt$  is the material derivative, so that equation shows how the velocity  $u$  of the marked volume changes as the volume itself moves around. By converting (1.3) into a form that describes how  $u$  changes in a fixed control volume (Eulrian framework), since  $u = u(x, t)$ , it is possible to apply the chain rule and write:

$$\frac{du}{dt} = \partial_t u + \frac{dx}{dt} \partial_x u = \partial_t u + u \partial_x u \quad (1.4)$$

and thus to write the final form of equation (1.3), which is:

$$\partial_t u + u \partial_x u + g \partial_x H = 0 \quad (1.5)$$

Similarly of what observed in (1.2), two terms affect the local rate of change of velocity: the pressure gradient term  $g\partial_x H$  and the advection of momentum  $u\partial_x u$ . The two equations (1.2) and (1.5) can be solved together for the two unknowns  $u(x, t)$  and  $H(x, t)$  in order to determine the system solution in time, for given initial and boundary conditions. The equations are clearly nonlinear (through the advective terms) and, in their standard NSW form, are usually written in the *depth-discharged* ( $H, q$ ) form for one horizontal dimension ( $x$ ), as follows:

$$\begin{cases} \partial_t H + \partial_x q = 0 \\ \partial_t q + \partial_x \left( \frac{q^2}{H} \right) + g \partial_x \left( \frac{H^2}{2} \right) = 0 \end{cases} \quad (1.6)$$

where  $q = Hu$  is also called *discharge*, since it measures the flow rate of water past a point.

If  $h(x)$  denotes a now variable depth for still water (equal to minus the bathymetry level up to a constant defining the zero level in the  $z$  direction,

1.2), a new term is to be considered in the momentum equation of the system (1.6), which becomes:

$$\partial_t q + \partial_x \left( \frac{q^2}{H} \right) + g \partial_x \left( \frac{H^2}{2} \right) - gH \partial_x h(x) = 0 \quad (1.7)$$

In this case, since  $\eta(x, t)$  is the variation in water depth level with respect to the reference level of calm water, we have that  $H(x, t) = \eta(x, t) + h(x)$ , and the entire system can be rewritten in terms of  $q(x, t)$  and the new variable  $\eta(x, t)$  as:

$$\begin{cases} \partial_t \eta + \partial_x q = 0 \\ \partial_t q + \partial_x \left( \frac{q^2}{H} \right) + gH \partial_x \eta = 0 \end{cases} \quad (1.8)$$

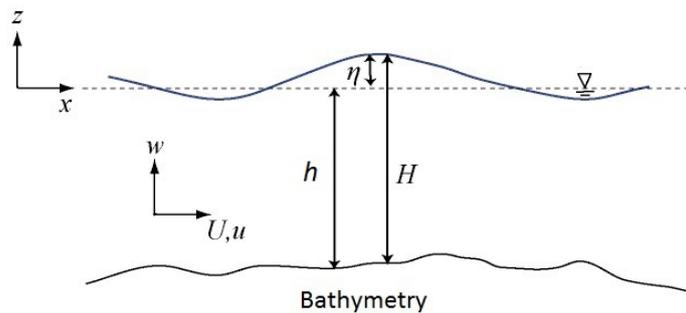
The nonlinear shallow water system of equation (1.8) well reproduces the propagation of very long waves near the coast, where the depth is small enough that frequency dispersion effects are negligible if compared to the ones due to nonlinearity, under the hydrostatic pressure assumption.

## 1.2 Madsen and Sørensen Boussinesq-type Equations

The enhanced Boussinesq equations, proposed by Madsen and Sørensen, are  $\mathcal{O}(\epsilon, \mu^2)$  accurate, recalling that nonlinearity  $\epsilon$  represent the ratio of wave amplitude to depth and dispersion  $\mu$  is the ratio of water depth to wavelength, and have improved dispersion properties thanks to a mathematical manipulation of the dispersive terms [35]. The system of one-dimensional nonlinear equations of Madsen and Sørensen, in the general case with variable bathymetry  $h = h(x)$ , comes in the following form:

$$\begin{cases} \partial_t \eta + \partial_x q = 0 \\ \partial_t q - Bh^2 \partial_{x^2 t} q - \frac{1}{3} h \partial_x h \partial_{xt} q + \partial_x (uq) + gH \partial_x \eta + \\ \quad - \beta g h^3 \partial_{x^3} \eta - 2\beta g h^2 \partial_x h \partial_{x^2} \eta = 0 \end{cases} \quad (1.9)$$

where the symbols  $\eta(x, t)$  and  $h(x)$  indicate the surface elevation and the distance between the bathymetry and the still water level, as shown in Fig. 1.2, while  $H(x, t) = \eta(x, t) + h(x)$  and  $q(x, t)$  are the same variables present in the NSW system; for the sake of clarity, the first term stands for the total water depth, while the latter represents the product of the depth averaged velocity of the water  $u(x, t)$  with  $H(x, t)$ . In addition, the brief notation  $\partial_{x^n}$  will be used within this work in order to indicate the recursive application of the partial derivative with respect to  $x$  for  $n$  times.



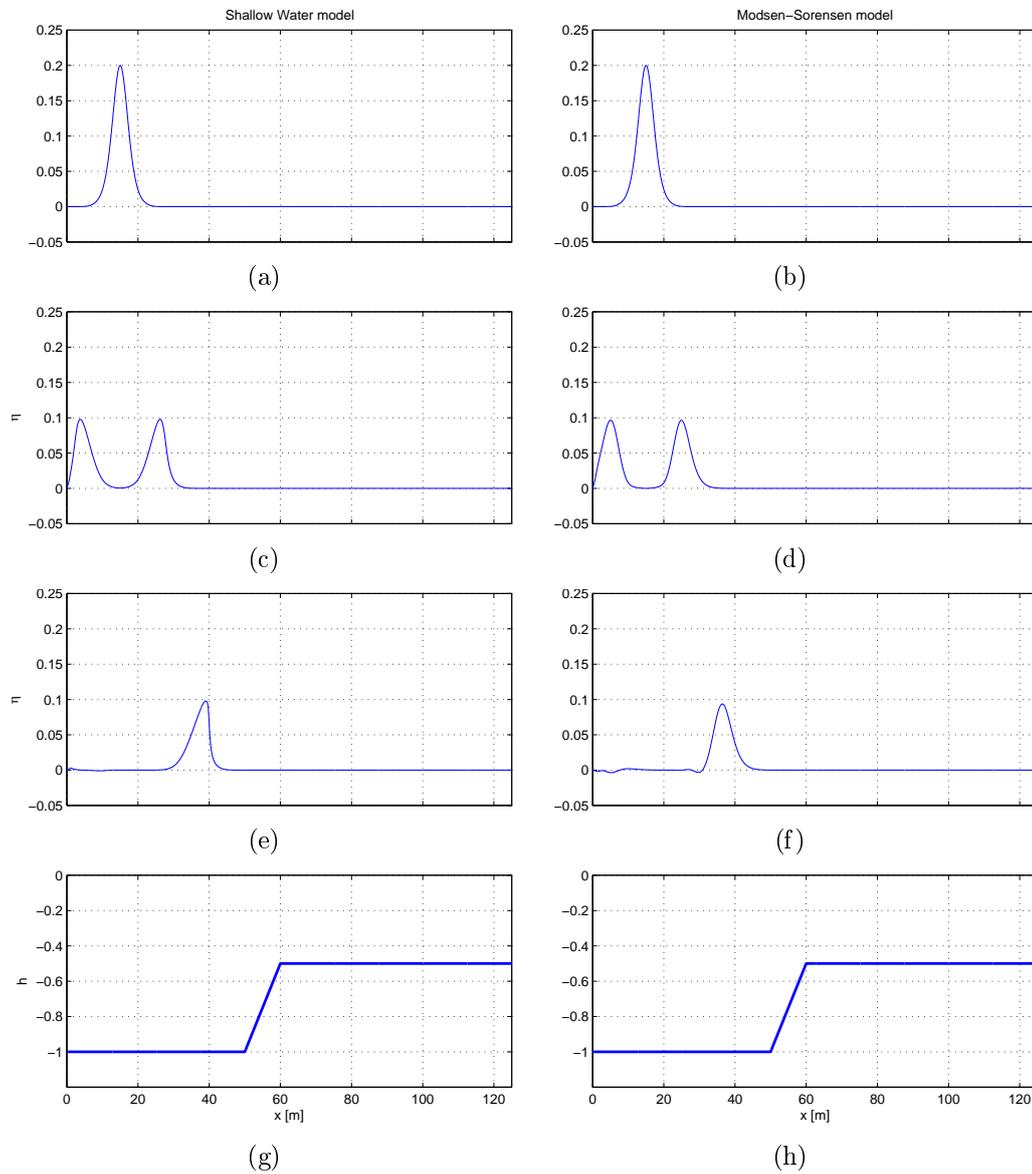
**Figure 1.2:** Sketch of the free surface flow problem, main parameters description.

This model is weakly nonlinear, preserving the same shallow water terms  $\partial_x(uq)$  and  $gH\partial_x\eta$  which are the only nonlinear in the system, being, thus, different from the fully nonlinear models, like the one proposed in [7]. It also contains additional dispersive terms in the momentum equation, which are linear with respect to the unknowns  $\eta$  and  $q$  of the system. These are pre-multiplied by two numerical parameters  $B$  and  $\beta$ , whose values are obtained by optimizing the dispersion properties of the linearized model with respect to the Airy wave theory. In such a way, the two parameters assume the values  $\beta \approx 0.066667$  and  $B = \frac{1}{3} + \beta$ . Assuming different values for both  $B$  and  $\beta$  from the one proposed, the dispersive and diffusive properties of the models will sensibly change; in particular, in the case of constant bathymetry  $h_0$ , with the specific choice of  $\beta = 0$ , and hence  $B = \frac{1}{3}$ , system (1.9) reduces to the original Boussinesq system of equations for constant bathymetries derived by Peregrine. Thus, the Madsen and Sørensen model can be seen as an optimization of the dispersive properties of the previous Peregrine's model, through the introduction of the parameters  $B$  and  $\beta$ , and it extends its range

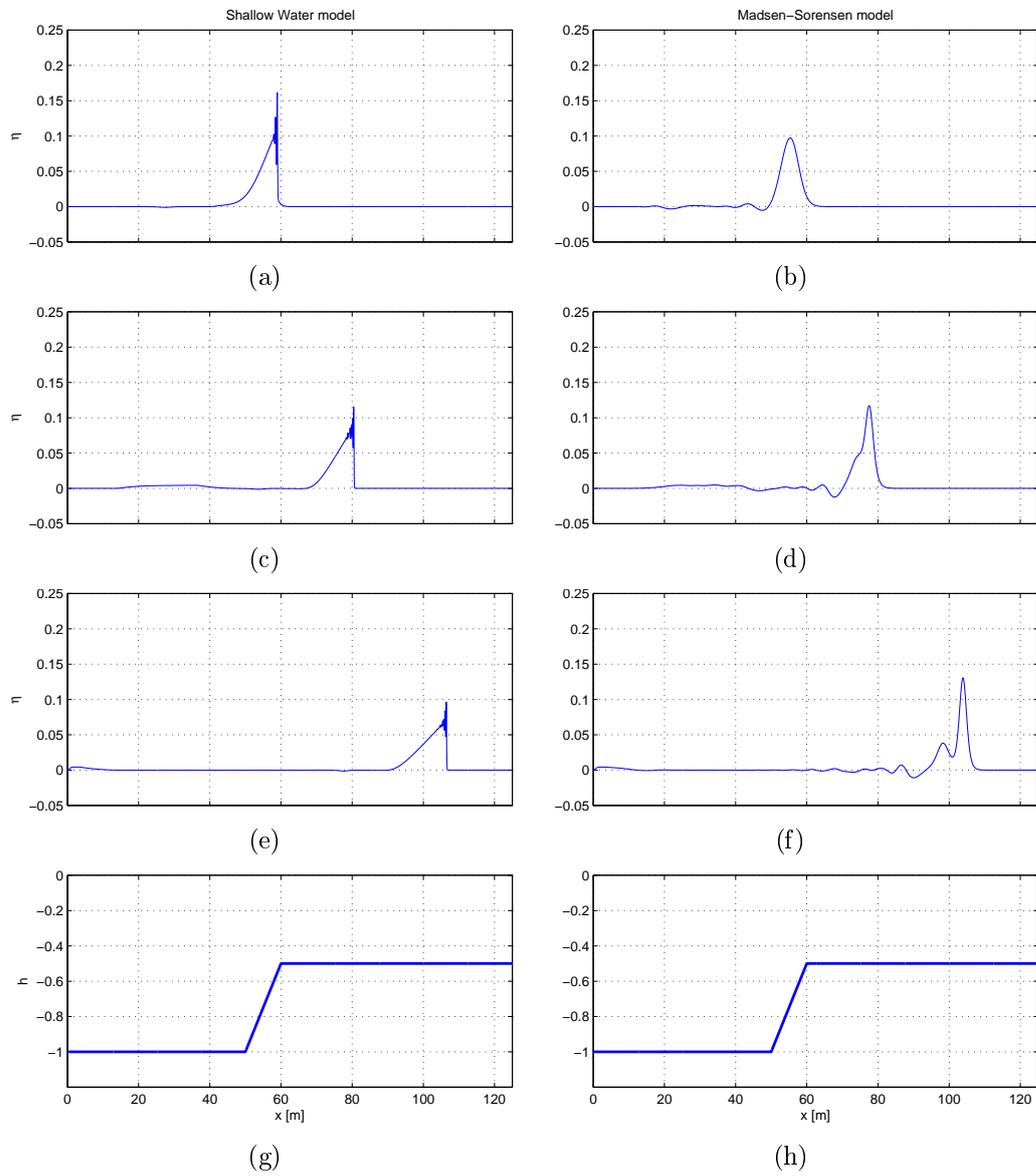
of applicability.

In order to motivate what states above about the inaccurate description given by the NSWE to the near-shore wave propagation problem, we show here an example comparing the two solutions computed by using both the NSWE and the MS systems. The test is very near to what will be reposed at the end of the work as a test case for the development methods validation and it consists on a long wave propagation over a shelf. The test simulates the propagation of a soliton wave with amplitude  $a = 0.2$  m in otherwise still water of depth  $h_0 = 1$  m. Figures 1.3 and 1.4 show the time evolution of the solitary wave from its initial configuration. The initial velocity of the soliton is zero, thus two gravity waves with opposite directions and half of the original wave amplitude move on from the initial point. In pictures of Figure 1.3 the simple propagation over constant bathymetry is shown. The soliton shape derived from MS model maintains a symmetric and almost unchanged shape, excluding a small amplitude reduction due to the dispersion effect of the model, while the shallow water simulation shows a clear nonlinear distortion of the wave shape which leads to a shock formation.

The pictures of Figure 1.4 show, instead, the very different behaviours of the two models in the description of the propagation over a variation in the bathymetry. This is a typical example which clearly shows that the NSWE model is inadequate to describe such a phenomenon, where the dispersive character is predominant with respect to the nonlinear one. The better dispersion properties of the Madsen-Sørensen model, which will be also investigated in the following, allow to obtain a more accurate description of the flow physics. The shape variation of the wave consists in an increment of the height and steepness which, in the limit, under certain conditions, lead to a wave breaking which then needs a shallow-water-type model in order to be computed. The dispersion, instead, leads to the fragmentation of the wave into more waves, each travelling, at its own speed. This effect is completely absent in the NSWE solution. In fact, the regular wave is transformed into a wave which dissipates energy (which is a good solution only in the case of a wave breaking phenomenon) and no sensibility to the presence of the shelf in the bathymetry is shown. In particular, the wave fragmentation due to the dispersion character of the interaction between the wave and the bathymetry is completely absent.



**Figure 1.3:** Free surface time history illustrating a soliton wave propagation over a bed variation (g); comparison between the solutions computed by means of NSWE (left) and MS (right) models.



**Figure 1.4:** Free surface time history illustrating a soliton wave propagation over a bed variation (g); comparison between the solutions computed by means of NSW (left) and MS (right) models.

### 1.3 Model Dispersion Analysis

The scope of this section is to present a comparison between the dispersion characteristics proper of the NSW and of the Madsen and Sørensen models with those of the linear wave theory, or Airy wave theory. As mentioned before, this property is of paramount importance in the oceanography research and it influences the range of applicability of the proposed methods.

A detailed study would require to take into account three specific quantities: the phase velocity  $C$ , which is the velocity of a point on the free surface; the group velocity  $C_g$ , which governs the propagation of energy in a train wave; the shoaling gradient  $s$ , which relates the change in wave amplitude to the change in water depth. However, admitting that the study of the other parameters does not lead to different or more restricted results [42], in this work we focused our analysis on the only phase velocity defined as:

$$C = \frac{\omega}{k} \quad (1.10)$$

with  $k = 2\pi/\lambda$  the wave number, and  $\omega$  the wave frequency.

Assuming a solution  $W = [H, q]^t$  expressed as a Fourier mode:

$$W = W_0 e^{\nu t + j k x} \quad (1.11)$$

for which, being  $j$  the imaginary unit and  $\nu$  a complex number defined as  $\nu = \xi + j\omega$ ,  $\omega$  represent the phase of the mode itself, while the just introduced  $\xi$  represent the rate of damping/amplification. Using the above definition of  $\nu$ , and being  $\omega = \pm C/k$ , for (1.10), it is possible to come to the following statement:

$$W = W_0 e^{\xi t} e^{j k (x + C(k)t)} \quad (1.12)$$

where the real number  $W_0 e^{\xi t}$  reflects the amplitude variation in the Fourier mode propagation, while the phase  $x + C(k)t$  shows explicitly the dependence of the phase velocity from the wave number:  $C = C(k)$ . This dependence results in a completely different behaviour of the solution in time. In fact, under the hypothesis of  $C = \text{const}$ , hence independent from  $k$ , all the characteristic frequencies of the initial solution are advected at the same velocity, without any differentiation in their phases which could lead to a dispersion

effect. Instead, in the case in which  $C = C(k)$ , each frequency will advance at its own velocity producing the differentiation in phase and the dispersion of the initial solution.

### 1.3.1 Linear Shallow Water System

Considering now the system of nonlinear shallow water equations (1.6), by introducing small perturbations to the initial state  $[h_0, u_0]$ , namely by assuming a solution in the form  $H = h_0 + h'$  and  $u = u_0 + u'$  (such that  $|h'|/|u'| \ll 1$  and  $|h'u'| = \mathcal{O}(\epsilon^2)$ ), with  $u_0 = 0$ , by neglecting high order terms, it is possible to obtain the linearized form of the NSW system as follows:

$$\begin{cases} \partial_t h' + h_0 \partial_x u' = 0 \\ \partial_t u' + g \partial_x h' = 0 \end{cases} \quad (1.13)$$

This system of linear hyperbolic partial differential equations, rewritten in the vector form:

$$\partial_t W(x, t) + \partial_x F(W(x, t)) = 0 \quad (1.14)$$

where  $W = [h' \ u']^t$  is the vector of the unknowns and  $F(W) = [h_0 u' \ gh']^T$  is the vector of the fluxes. The chain rule can be applied in order to handle the second term of (1.14) and making appear the Jacobian matrix  $A = dF(W)/dW$  of the system. In the present linearised shallow water case, this matrix has the simple form:

$$A = \begin{bmatrix} 0 & h_0 \\ g & 0 \end{bmatrix} \quad (1.15)$$

Its eigenvalues  $\lambda_{1,2} = \pm\sqrt{gh_0}$  represent the celerity  $c$  of the waves travelling in the two opposite verses of propagation of the problem, since it is one-dimensional. The solution of the system (1.13) can be performed using the classical characteristic method [26] and reads:

$$\begin{cases} h(x, t) = h_0(x - ct) \\ u(x, t) = \sqrt{\frac{g}{h_0}} h(x, t) \end{cases} \quad (1.16)$$

which is the simple linear advection of a perturbation, without any dissipation, which travels at the constant speed  $c$ . The above confirms what observed in the previous section about the non-dispersive character of the NSW. If we insert a Fourier mode solution in this model, equation (1.14) is translated into:

$$\nu W + jkAW = 0 \quad (1.17)$$

Handling equation (1.17), it can be formulated as  $(\nu I + jkA)W = 0$ , whose non-trivial solution leads to investigate the eigenvalue problem  $\nu I + jkA = 0$ , associated to the characteristic polynomial:

$$\nu^2 + K^2 c^2 = 0 \quad (1.18)$$

Using the definition of  $\nu$  and solving separately the real and the imaginary part of the equation, these operations allow to discover the expressions for the parameter  $\xi$  and  $\omega$ :

$$\begin{aligned} \xi_{SW} &= 0 \\ \omega_{SW}^2 &= k^2 c^2 \end{aligned} \quad (1.19)$$

which means that, for the system of nonlinear shallow water equations, the dissipation effect is null, since  $\xi$  is always zero, and there is no dispersion effect of the initial solution, since, using the definition (1.10) and the second of (1.19), we can obtain  $C_{SW} = \frac{\omega}{k} = c = \sqrt{gh_0} = \text{const}$  for all frequencies which thus travel at the same advection velocity with no phase lagging.

### 1.3.2 Linearized Madsen and Sørensen System

We mentioned above that the Madsen-Sørensen system of Boussinesq equations, with respect to the NSWE, conserves dispersive properties that allow it to well reproduce water behaviour also where dispersion effects become more important than nonlinearity. In order to reproduce the same linear dispersion analysis done for the nonlinear shallow water system, we now report below the linearized form of the MS system:

$$\begin{cases} \partial_t \eta + h_0 \partial_x u = 0 \\ \partial_t u - B h_0^2 \partial_{xxt} u + g \partial_x \eta - \beta g h_0^2 \partial_{xxx} \eta = 0 \end{cases} \quad (1.20)$$

As before, we set the vector of variables  $W = [\eta, u]^t$  which allows us to rewrite the previous system (1.20) in the same vectorial form of (1.14) and, after the application of the chain rule, gain the definition of the Jacobian matrix of the system  $A$  as:

$$A = \begin{bmatrix} 0 & h_0 \\ g - \beta g h_0^2 \partial_{xx} & -B h_0^2 \partial_{xt} \end{bmatrix} \quad (1.21)$$

Using equation (1.12) and passing through the reformulation of the problem in the analogous form of (1.17), the new eigenvalue problem can than be set:

$$\nu I + \begin{bmatrix} 0 & j k h_0 \\ j k g - (j k)^3 \beta g h_0^2 & -(j k)^2 \nu B h_0^2 \end{bmatrix} = 0 \quad (1.22)$$

The solution of the characteristic polynomial associated to (1.22) in its real and imaginary parts, setting  $\mu = k h_0$  and  $c = \sqrt{g h_0}$  like in the shallow water case, leads to the final expressions:

$$\begin{aligned} \xi_{MS} &= 0 \\ \omega_{MS}^2 &= k^2 c^2 \frac{1 + \beta \mu^2}{1 + B \mu^2} \end{aligned} \quad (1.23)$$

Comparing the two different expression of  $\omega$  in equations (1.19) and (1.23), it becomes clear that the two models, while having the same null dissipation property since  $\xi_{SW} = \xi_{MS} = 0$ , show a very different dispersive character. In fact, the Madsen-Sørensen system results  $k$ -dependent thank to the term  $(1 + \beta\mu^2)/(1 + B\mu^2)$ , which is a function of  $k$  through  $\mu$ , the phase velocity, which can be computed from  $\omega$  through the definition given in eq(1.10):  $C = C(k)$ . Therefore, each frequency contained in the initial solution will propagate with a different velocity  $C$ . The dispersive character of this model can be tuned through the modification of the values of the parameters  $B$  and  $\beta$ , which appear to influence the value of  $C(k)$  through the (1.23). Usually the standard values for these parameters are chosen in order to match the dispersive behaviour of the linear steady wave theory (or Airy wave theory). Airy wave theory, in fluid dynamics, describes surface gravity waves of small amplitude in a liquid medium of arbitrary depth. The limits of validity of this theory are related to two dimensionless parameters: the wave steepness (given by the ratio of the wave amplitude  $A$  with its wavelength  $\lambda$  and which must be small, i.e.  $A/\lambda \ll 1$ ), and the relative depth  $A/H$  (which must be small too, and represents the ratio between the wave height and the water depth  $H$ ). The derivation of Airy theory gives the expression for the fluid surface elevation as:

$$\eta(x, t) = \frac{A}{2} \cos(kx - \omega t) \quad (1.24)$$

Setting also  $\mu = kh_0$ , for the linear wave theory we obtain:

$$\omega_{\text{Airy}}^2 = gk \tanh(\mu) \quad (1.25)$$

which implies:

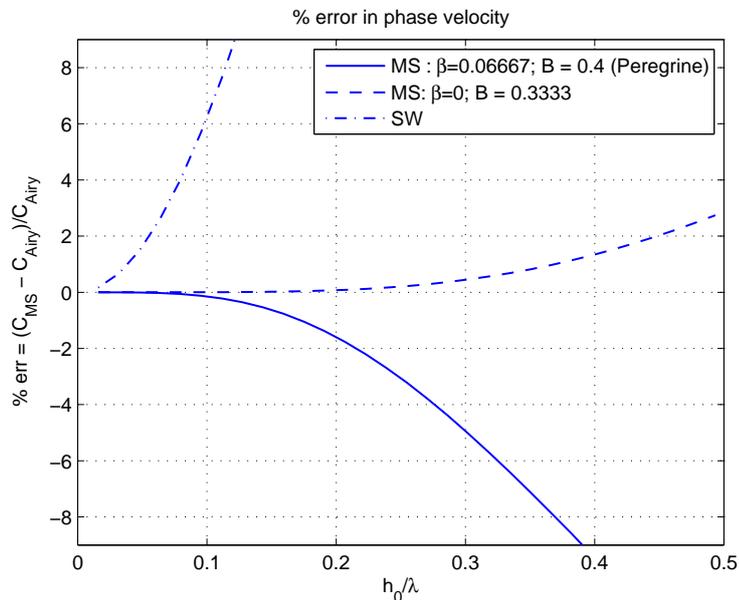
$$C_{\text{Airy}}^2 = \frac{\omega_{\text{Airy}}^2}{k^2} = gh_0 \frac{\tanh(\mu)}{\mu} \quad (1.26)$$

Thus we use the last two relations in order to compare the dispersion properties of the models presented. Fig.1.5 shows the behaviour of the phase velocity  $C$  over the variation of the adimensional quantity  $h_0/\lambda$ . The value of this parameter is strictly connected, through the definition of the wavelength  $k = 2\pi/\lambda$ , to the value of the previously defined  $\mu$ , which is used to derive the Boussinesq-type models and to define their range of applicability

(the limit depth for a model validity).

The phase velocity computed by the Madsen and Sørensen model is very close to the linear wave theory one; on the contrary, the nonlinear shallow water system appears to be in extremely poor agreement with it. Since  $C_{SW} = gh_0 = \text{const}$ , the NSW model cannot reproduce the dispersion of frequency in the near-shore region. The MS model shows, instead, a better agreement to Airy theory, also with respect to the Peregrine Boussinesq-type equations. Figure 1.5 shows the error in phase velocity determined by the relation:

$$\text{err} = 100 \left( \frac{C - C_{\text{Airy}}}{C_{\text{Airy}}} \right) \quad (1.27)$$



**Figure 1.5:** Comparison of the dispersion characteristics of the Peregrine’s Boussinesq, extended Boussinesq and shallow water models with respect to the linear wave theory, based on the representation of the percentage of error in phase velocity computation.

Other informations can be derived from Figure 1.5. The lack of accuracy given by the NSW model, is there shown, whose phase velocity is constant over the whole range and, hence, its error with respect to the linear theory rapidly diverges and becomes greater than 5% after  $h/\lambda = 0.1$ . Fig.1.5 above shows also how the accuracy of the phase velocity computed by the Madsen and Sørensen model is strongly dependent on the value of the parameter  $\beta$ ,

changing also the range of application of the model itself over different water depths. When  $\beta = 0$  and it is set  $B = \beta + 1/3$ , the Peregrine's model is obtained. Phase velocity computed by this model shows a very low accuracy, and the value of the error exceeds the 5% for depths greater than  $3/10$  the wavelength. The optimum correspondence in the frequency  $\omega$  is found with  $\beta = 0.066667$ , since in this case the error with respect to the Airy theory's phase velocity is less than 2% over the whole range of the adimensional parameter  $h_0/\lambda$  considered.



## Chapter 2

# Weighted Residuals Methods

We consider the numerical solution of the enhanced Boussinesq equations of Madsen and Sorensen, introduced in Section 1.2, in order to give an accurate description to the dispersive wave propagation problem in near-shore zones. Solving this system of equations requires to operate a space-time discretization process. In order to accomplish this, weighted residuals methods are here developed.

Let us consider the generic system of hyperbolic nonlinear Partial Differential Equations (PDEs) for the set of conservative variables  $\mathbf{u}$  in its *strong form*:

$$\partial_t \mathbf{u} + \nabla \cdot \mathcal{F}(\mathbf{u}) = 0 \tag{2.1}$$

In order to be solved, system (2.1) must be coupled with initial and boundary conditions on a bounded spatial domain that lead to a well posed problem. In general, the classical solution of the PDEs' system shown must belong to the space of functions which are continuous with partial derivatives continuous too, up to the order of the maximum derivative present in the system, one in the case of system (2.1). In the case of the Madsen-Sørensen equations, we have already focused the attention on the fact that a third order derivative is also present, which means that an even higher level of regularity is requested to the solution. In other words,  $\mathbf{u}$  must be a very smooth function. Thus, physically discontinuous solutions, which satisfy the original integral form of the conservation laws, are not admissible solutions of the system of PDEs in the classical sense in all points, since the derivatives are not defined at the discontinuities. In fact, differential equations are derived from the original integral ones by imposing additional smoothness assumptions on the solution,

while integral form continue to be valid even for discontinuous solutions. In order to broaden the class of admissible functions, it is worthwhile to consider an integral or *weak form* of the conservation law. This again involves integrals and allows discontinuous solutions but it is easier to be treated than the original integral form of the conservation laws.

Weighted residuals methods are a general approach to the derivation of weak forms for a given PDEs' system. The *residual* of equation (2.1) is defined as:

$$\mathcal{R}(\mathbf{u}_h) = \partial_t \mathbf{u}_h + \nabla \cdot \mathcal{F}(\mathbf{u}_h) \quad (2.2)$$

such that  $\mathcal{R}(\mathbf{u}) = 0$  if  $\mathbf{u}$  is the exact solution of the PDE equation. In this way the magnitude of the residual  $\mathcal{R}(\mathbf{u}_h)$  represent also a measure of the accuracy of an approximate solution computed by means of a numerical discretization.

The basic idea in developing the weak form of (2.1) in the continuum is to rewrite the differential system in such a way that less regularity  $\mathcal{F}$  is required to the solution. In order to accomplish this, the system of PDEs is firstly pre-multiplied by a set of smooth *weights*, or *test functions*  $\mathbf{v}$ , and then integrated one or more times over the spatial domain  $\Omega$ ; the third step consists of using integration by parts to move the derivatives from the conservative variables  $\mathbf{u}$  onto the smooth test functions. Being  $V$  an infinite-dimensional function space with some properties linked to the boundary conditions which are to be added to (2.1), the solution procedure for the PDEs system results in looking for the set of functions  $\mathbf{u} \in V$  such that  $\forall \mathbf{v} \in V$  the following expression is satisfied:

$$\int_{\Omega} \mathbf{v} \partial_t \mathbf{u} dx - \int_{\Omega} \mathcal{F}(\mathbf{u}) \cdot \nabla \mathbf{v} dx = \mathcal{BC}s \quad (2.3)$$

where  $\mathcal{BC}s$  includes all the boundary terms, which normally arise from the integration by parts.

Discretizing the above problem means working with the finite-dimensional subspace  $V_h$  of  $V$  making an approximation to the solution. Denoting with the subscript  $_h$  mathematic elements which belong to the discretized problem, the formulation of the previous weak form for the system (2.1) of PDEs becomes:

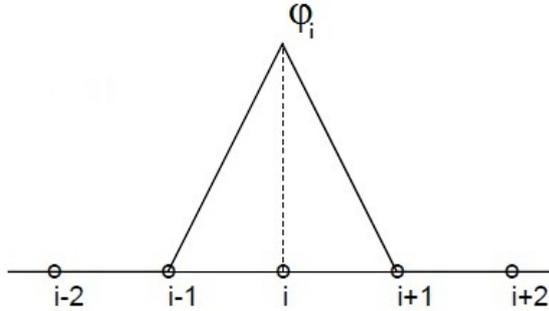
"look for the function  $\mathbf{u}_h \in V_h$  such that  $\forall \mathbf{v}_h \in V_h$  the following expression is satisfied":

$$\int_{\Omega_h} \mathbf{v}_h \partial_t \mathbf{u}_h d\mathbf{x} - \int_{\Omega} \mathcal{F}(\mathbf{u}_h) \cdot \nabla \mathbf{v}_h d\mathbf{x} = \mathcal{BC}_s \quad (2.4)$$

Considering the one-dimensional problem, we now partition the domain into a set of  $N$  finite elements and we define a base of linear weighting functions ( $\varphi_i(x)$ ,  $i = 0, \dots, N$ ) like the one shown in Figure 2.1, such that the test function  $\varphi_i(x)$  is non-zero only over the two elements adjacent node  $i$  and it is identically null outside and on nodes  $i + 1$  and  $i - 1$  themselves.

In this environment, being  $f(u_h) = f_h$  the one-dimensional flux of (2.1), if  $f_h \in C^0$  on each element's closure and  $f_h \in C^\infty$  on the element itself, the following statement becomes true:

$$- \int_{\Omega} f_h \partial_x \varphi_i = \int_{\Omega} \varphi_i \partial_x f_h \quad (2.5)$$



**Figure 2.1:** Representation of the linear finite basis function at the node  $i$ .

A brief explanation of passages which demonstrate (2.5) can be found in Appendix A.

Just making the hypothesis of working with continuous fluxes and using the equality 2.5, a new formulation can be written. Thus, renouncing to transfer first order derivatives to the weighting functions  $\mathbf{v}_h$ , we have:

look for  $\mathbf{u}_h \in V_h$  such that  $\forall \mathbf{v}_h \in V_h$  is:

$$\int_{\Omega_h} \mathbf{v}_h (\partial_t \mathbf{u}_h + \nabla \cdot \mathcal{F}(\mathbf{u}_h)) dx = \widetilde{\mathcal{BC}}s \quad (2.6)$$

where the expression of the boundary terms  $\widetilde{\mathcal{BC}}s$  is in general different from the previous  $\mathcal{BC}s$ .

Recognising in the above expression the residual definition (2.2), (2.6) becomes:

$$\int_{\Omega_h} \mathbf{v}_h \mathcal{R}(\mathbf{u}_h) dx = \widetilde{\mathcal{BC}}s \quad (2.7)$$

which is the *residual form* associated to the strong form (2.1).

Under sufficient smoothness of the involved function spaces, the equivalence between the strong and the weighted-residual forms can be demonstrated. This equivalence plays a fundamental role in the construction of approximate solutions (including finite element solutions) to the underlying problem. Residual form (2.7) can thus be built without considering the variational form (2.4). In this respect, the residual method becomes more similar to a finite volume method in which total exchanged fluxes between cells are computed.

Various approximation methods, such as Galerkin or Residual Distribution methods, can be derived by appropriately restricting the admissible form of the weighted functions and the actual solution.

In the following sections we will restrict our analysis to the one-dimensional Madsen-Sørensen's system of Boussinesq-type equations and we will use a technique for decoupling the spatial and temporal discretization process. We will assume *periodic boundary conditions*, which is equivalent to a one-dimensional domain of infinite length, thus setting to zero the  $\widetilde{\mathcal{BC}}s$  term on the right hand side of equation (2.7). In Chapter 7 we will discuss better how the boundary conditions of the numerical schemes were set for the several test cases proposed.

Given a suitably designed computational mesh, the system of partial differential equations is firstly semi-discretized in space, reducing the problem to a system of ordinary differential equation in time. The continuous functions  $\mathbf{u}(\mathbf{x}, t)$  is, thus, approximated by a finite number of nodal values  $\mathbf{u}_i$  which may be associated with vertices, edges, faces, cells or control volumes. Due to the fact that governing equations model an unsteady process, these degree of freedom are time-dependent  $\mathbf{u}_i(t)$  and should be updated step-by-step solving this final system with an appropriate time integration method. Due to the

complexity of the system of equations we have to deal with, we will show in the following the development of the numerical schemes for simplified problems, illustrating in particular their properties and connections and making comparisons between the results, starting with the simple one-dimensional linear scalar advection equation and passing, then, to the one-dimensional scalar advection-dispersion problem.



## Chapter 3

# Linear Scalar Advection Equation

We first consider the one-dimensional linear scalar advection equation for generic variable  $u$ , which can be written as:

$$\partial_t u + a \partial_x u = 0 \quad (3.1)$$

with  $a$  the constant velocity of propagation of the information. It is the simplest hyperbolic partial differential equation, whose solution can be analytically found using the characteristic method in all the considered space-time domain. We will compute a weighted residual numerical discretization of (3.1), in order to show some important properties and similarity relations for the numerical schemes we are going to develop for the more complex Madsen-Sørensen system.

**Dispersion Analysis :** It is important that the schemes developed below respects the dispersion and dissipation properties of the original model. For this reason these properties are investigated in this paragraph in the same way as it was done in the previous chapter for the NSWE and MS models. Thus, assuming a solution of (3.1), expressed as a Fourier mode, like in (1.11), with  $\xi$  the rate of damping and  $\omega$  the phase of the complex  $\nu = \xi + j\omega$ , replacing this mode in the advection equation we can obtain the relation:

$$\nu + jak = 0 \quad (3.2)$$

which shows that for that advection equation:

$$\xi = 0 \tag{3.3}$$

$$\omega = -ak \tag{3.4}$$

### 3.1 Finite Difference Methods

Finite difference schemes will not be developed here for the Madsen-Sørensen system, which is the subject of this study. Despite this, they are here mentioned, for the sake of clarity, in order to support the following Galerkin finite element, whose accuracy properties will be, in this way, demonstrated to be better.

The finite difference method is developed by first partitioning a given one-dimensional domain  $\Omega$  into a set of  $N + 1$  non-coincident nodes, including the extremes of the domain  $X_1$  and  $X_2$ . Assuming the nodes to be equispaced, a constant grid size  $\Delta x$  can be defined as  $(X_2 - X_1) / N$ . The unknown function  $u$  is approximated over this set of points in terms of its nodal values, and its spatial derivatives are approximated directly at each point by using Taylor series expansions in space: this allows to relate the derivative at each point to the solution evaluated at the two adjacent nodes, and therefore to deal with a global equations system with banded matrix structure, which can efficiently be solved. The finite difference approach, for nodes  $i+1$  and  $i-1$ , leads respectively to:

$$f(x_{i+1}) = f(x_i) + \frac{df}{dx}(x_i)(\Delta x) + \mathcal{O}(\Delta x^2) \tag{3.5}$$

$$f(x_{i-1}) = f(x_i) - \frac{df}{dx}(x_i)(\Delta x) + \mathcal{O}(\Delta x^2) \tag{3.6}$$

if Taylor expansion is limited to the first order terms. By subtracting (3.6) from (3.5) and rearranging, we come to the finite difference approximation of the second truncation error accuracy order for the first derivative for node  $i$ :

$$\frac{df}{dx}(x_i) = \frac{1}{2\Delta x} (f_{i+1} - f_{i-1}) + \mathcal{O}(\Delta x^2) \tag{3.7}$$

The order of accuracy may be increased by amplifying on one hand the polynomial expansion, i.e, up to the third order of derivation, and on the other hand the number of involved nodes:

$$\begin{aligned}
 f(x_{i+1}) &= f(x_i) + \frac{df}{dx}(x_i)(\Delta x) + \frac{d^2f}{dx^2}(x_i)\frac{(\Delta x)^2}{2} + \frac{d^3f}{dx^3}(x_i)\frac{(\Delta x)^3}{6} + \mathcal{O}(\Delta x^4) \\
 f(x_{i-1}) &= f(x_i) - \frac{df}{dx}(x_i)(\Delta x) + \frac{d^2f}{dx^2}(x_i)\frac{(\Delta x)^2}{2} - \frac{d^3f}{dx^3}(x_i)\frac{(\Delta x)^3}{6} + \mathcal{O}(\Delta x^4) \\
 f(x_{i+2}) &= f(x_i) + \frac{df}{dx}(x_i)(2\Delta x) + \frac{d^2f}{dx^2}(x_i)\frac{(4\Delta x)^2}{2} + \frac{d^3f}{dx^3}(x_i)\frac{(8\Delta x)^3}{6} + \mathcal{O}(\Delta x^4) \\
 f(x_{i-2}) &= f(x_i) - \frac{df}{dx}(x_i)(2\Delta x) + \frac{d^2f}{dx^2}(x_i)\frac{(4\Delta x)^2}{2} - \frac{d^3f}{dx^3}(x_i)\frac{(8\Delta x)^3}{6} + \mathcal{O}(\Delta x^4)
 \end{aligned} \tag{3.8}$$

The linear combination of the previous third order expansions leads to the following expression for the first derivative for node  $i$ :

$$\frac{df}{dx}(x_i) = \frac{1}{12\Delta x} (f_{i-2} - 8f_{i-1} + 8f_{i+1} - f_{i+2}) + \mathcal{O}(\Delta x^4) \tag{3.9}$$

Equation (3.9) shows that in order to increase the order of accuracy of the derivative approximation, a stencil enlargement is required. This increases the computational cost of the calculus and results in much more complicated boundary conditions that must be posed.

Relations (3.7) and (3.9) can be used to obtain a space-discrete version of the linear scalar advection equation proposed, where only a first order spatial derivative is present. They represent central finite difference formulae, since their stencil are centred on the node  $i$ . The schemes which results from their use are:

$$\frac{d(u_i)}{dt} + \frac{a}{2\Delta x} (u_{i+1} - u_{i-1}) = 0 \tag{3.10}$$

$$\frac{d(u_i)}{dt} + \frac{a}{12\Delta x} (u_{i-2} - 8u_{i-1} + 8u_{i+1} - u_{i+2}) = 0 \tag{3.11}$$

Despite their great simplicity due to their explicit form, which does not require any matrix inversion in order to be solved, when discontinuous or

near-discontinuous feature raise in a physical system, it is possible for such centred schemes to perform badly as they do not take into account the direction of propagation of the physical quantities. Upwind method must thus be developed in such cases, but they will be investigated in the following.

**Truncation error :** Let  $u$  be a smooth classical solution, Taylor approximation of the nodal values:

$$\begin{aligned}
 u_{i+2} &= u_i + 2\Delta x \partial_x u_i + 2\Delta x^2 \partial_{x^2} u_i + \frac{4\Delta x^3}{3} \partial_{x^3} u_i + \frac{2\Delta x^4}{3} \partial_{x^4} u_i + \frac{4\Delta x^5}{15} \partial_{x^5} u_i + \dots \\
 u_{i+1} &= u_i + \Delta x \partial_x u_i + \frac{\Delta x^2}{2} \partial_{x^2} u_i + \frac{\Delta x^3}{6} \partial_{x^3} u_i + \frac{\Delta x^4}{24} \partial_{x^4} u_i + \frac{\Delta x^5}{120} \partial_{x^5} u_i + \dots \\
 u_{i-1} &= u_i - \Delta x \partial_x u_i + \frac{\Delta x^2}{2} \partial_{x^2} u_i - \frac{\Delta x^3}{6} \partial_{x^3} u_i + \frac{\Delta x^4}{24} \partial_{x^4} u_i - \frac{\Delta x^5}{120} \partial_{x^5} u_i + \dots \\
 u_{i-2} &= u_i - 2\Delta x \partial_x u_i + 2\Delta x^2 \partial_{x^2} u_i - \frac{4\Delta x^3}{3} \partial_{x^3} u_i + \frac{2\Delta x^4}{3} \partial_{x^4} u_i - \frac{4\Delta x^5}{15} \partial_{x^5} u_i + \dots
 \end{aligned}
 \tag{3.12}$$

which can be used in order to compute a truncation error analysis of the central finite difference schemes (3.10) and (3.11). In particular, substituting the previous expression directly in the equations, one can obtain:

$$\partial_t u_i + a \partial_x u_i = -\frac{a\Delta x^2}{6} \partial_x^3 u_i + \mathcal{O}(\Delta x^4) \tag{3.13}$$

$$\partial_t u_i + a \partial_x u_i = -\frac{a\Delta x^4}{30} \partial_x^5 u_i + \mathcal{O}(\Delta x^6) \tag{3.14}$$

Imaging to substitute in this last couple of equations, the exact solution of the advection problem  $\partial_t u_i + a \partial_x u_i$ , we would thus obtain in the first case a second order truncation terms, which is the degree of accuracy of the spatial discretization method adopted, while in the second one the fourth order is reached. Finite difference approximation requires the cost of a scheme stencil enlargement to be payed, in order to reach higher order of accuracy.

**Dispersion Analysis:** The dispersion analysis of the two schemes (3.10) and (3.11) results substituting a discretized formulation of the Fourier solution  $u(x, t) = \bar{u}e^{\nu t + jkx}$  into them. Defined  $m$  as the index spanning all the nodes of the stencil centred in node  $i$  equations (3.10) and (3.11) refer to. Consequently, for the node  $i+m$  it is possible to write:

$$u_{i+m} = \bar{u}e^{\nu t}e^{jkm\Delta x} = e^{\nu t}e^{jm\mu\Delta x} \quad (3.15)$$

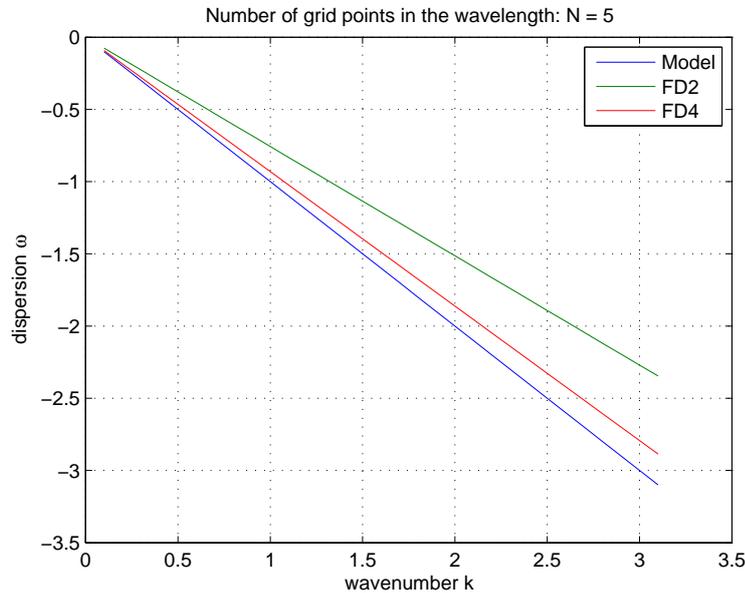
After the definition of  $\mu_{\Delta x} = k\Delta x$ , the final expression of the respective values of the dispersion and damping coefficients, after a series of algebraic passages and manipulation, is:

$$\xi_{\mathcal{FD}2} = 0 \qquad \xi_{\mathcal{FD}4} = 0 \quad (3.16)$$

$$\omega_{\mathcal{FD}2} = -ak \frac{\sin \mu_{\Delta x}}{\mu_{\Delta x}} \qquad \omega_{\mathcal{FD}4} = -ak \frac{\sin \mu_{\Delta x}}{3\mu_{\Delta x}} (4 - \cos \mu_{\Delta x}) \quad (3.17)$$

where subscript  $\mathcal{FD}2$  stands for value related to the finite difference scheme of the second order of accuracy and  $\mathcal{FD}4$  means the one related to the fourth order of accuracy method. Figure 3.1 plots the expression of the dispersion parameter  $\omega$  as function of the mode wavenumber  $k$ , and shows the best agreement reached with the fourth order difference approximation of the first derivative, with respect to the second one. A better performance in the dispersion properties description of the model is also possible, but a deep mesh refinement is needed. In this case also the second order scheme shows a good agreement with the exact model.

Instead, regarding the dissipation property, null values of  $\xi$  correctly reproduce the model absence of dissipation in both the schemes.



**Figure 3.1:** Dispersion  $w$  against the wavenumber  $k$  of the finite difference schemes  $\mathcal{FD}2$  and  $\mathcal{FD}4$ , compared to the scalar linear advection model one, when the wavelength is discretized with a number of grid nodes  $N = 5$ .

## 3.2 Galerkin Finite Element Method

Finite element method makes use of a spatial discretization and a weighted residual formulation to compute a system of matrix equations, whose solution yields an approximation of the original boundary value problem. In Galerkin finite element method, the solution shows the best approximation property. Given a finite one-dimensional domain  $\Omega$  between extremes  $X_1$  and  $X_2$ , the finite element method is first developed partitioning this domain into a set of  $N + 1$  non-coincident nodes, including the extremes of the domain  $X_1$  and  $X_2$ . Each pair of adjacent nodes defines a spatial interval called *element* such that the set of the  $N$  elements built, or *mesh*, completely covers the spatial domain. In the following, the mesh will be assumed to be always *uniform* with a  $\Delta x$  element length constant in all the domain.

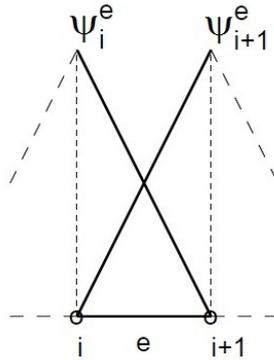
Galerkin finite element discretization can be implemented starting from the weak formulation of (3.1); consequently, recalling what just said in the previous section, where the weak form has been derived for the generic system of hyperbolic nonlinear Partial Differential Equations (PDEs) (2.1), we obtain:

$$\int_{\Omega} \varphi_i \partial_t u dx - \int_{\Omega} au \partial_x \varphi_i dx = \mathcal{BCs} \quad (3.18)$$

where the script  $\mathcal{BCs}$  replace the expression of some terms which depend on the value of the function at the boundaries of the problem domain and which can be easily set to zero by i.e. imposing periodic boundary conditions to the domain.

The spatial discretization of variable  $u$  is introduced through a set of linear basis functions  $(\psi_i(x); i = 1, \dots, N + 1)$  which are defined on each node so that they interpolate nodal values of  $u$ :

$$\psi_i(x_j) = \delta_{i,j} \quad (3.19)$$



**Figure 3.2:** Representation of the linear finite basis function over the element  $e$ .

being  $\delta_{i,j}$  the Kronecker delta. This basis functions are so built in order to assume the exact value of the variables at the node  $i$  and are non-zero on the only two adjacent elements they belong to, as it was shown in Figure 2.1. By the opposite point of view, for each element  $e$  of the mesh, only the basis functions centred on the two adjacent nodes are non-zero on the element itself. This two functions, see Figure 3.2, define a linear interpolation  $u_e(x)$  over the element in terms of the nodal values  $\psi_i$  e  $\psi_{i+1}$  such that for the set of conservative variables  $u$ :

$$u_e(x) = \psi_i^e u_i + \psi_{i+1}^e u_{i+1} \quad (3.20)$$

Such a linear approximation is piecewise continuous over the mesh, although its first derivative will be discontinuous in correspondence of the nodes but can be uniquely defined on each element. The spatial discretized approximation of the unknown function in all the domain is then produced by an interpolation over the set of basis functions:

$$u(x, t) \approx \sum_{j=1}^{N+1} \psi_j(x) u_j(t) \quad (3.21)$$

Galerkin finite element method is a weighted residual method in which weighting functions  $\varphi_i$ , in the weak formulation, are chosen from the same function space as the basis functions  $\psi_j$  used for variables approximation. In this case they are, thus, still  $\mathcal{P}^1$  functions of the kind shown in Figure 3.2 and this lead to the writing:

$$\left( \int_{X_1}^{X_2} \varphi_i \varphi_j dx \right) \frac{du_j}{dt} - \left( \int_{X_1}^{X_2} \varphi_j \partial_x \varphi_i dx \right) a u_j = 0 \quad (3.22)$$

Within the finite element framework this global equation can be assembled efficiently by considering each element in turn. The integral over the spatial domain  $[X_1, X_2]$  can thus be split in the sum of integrals over the single elements, resulting in a element-by-element final assembly procedures. Using (3.20), the solution over each element is interpolated using only the two local basis functions. All the other basis functions are, in fact, zero on the element and their contributions to the integral are null.

To solve the definite integrals present in equation (3.22) in an approximate way a quadrature formula is needed. The Galerkin mass matrix is obtained integrating the terms in the exact way by using the Simpson's rule. These operations, due to the form of the selected basis function, are not particularly complicated and, after assembling the element integrals into a global equation system, the result is a system of ordinary differential equations in time, which for each node  $i$  of the mesh results:

$$\frac{\Delta x}{6} \frac{d(u_{i-1})}{dt} + \frac{2\Delta x}{3} \frac{d(u_i)}{dt} + \frac{\Delta x}{6} \frac{d(u_{i+1})}{dt} + \frac{a}{2} (u_{i+1} - u_{i-1}) = 0 \quad (3.23)$$

The procedure described up to now produces a center nodal equation analogous to the center finite difference schemes with the additional advantage of having a mass matrix term which multiplies the time derivatives and which augments the accuracy of the scheme in a way that will be detailed in the truncation error analysis. In the finite difference method, in fact, these terms are directly replaced by nodal values such that the finite element discretization would be identical to the fully second order finite difference method when a mass lumping operation is made on the Galerkin mass matrix.

**Truncation Error :** Recalling Taylor approximations of the nodal values (3.8) and substituting them into the Galerkin spatially discretized scheme for the linear scalar advection problem, the following equation is found:

$$\begin{aligned} \partial_t u_i + a \partial_x u_i = & - \frac{\Delta x^2}{6} \partial_{x^2} (\partial_t u_i + a \partial_x u_i) + \\ & - \frac{\Delta x^4}{24} \partial_{x^4} \left( \frac{1}{3} \partial_t u_i + \frac{a}{5} \partial_x u_i \right) + \mathcal{O}(\Delta x^6) \end{aligned} \quad (3.24)$$

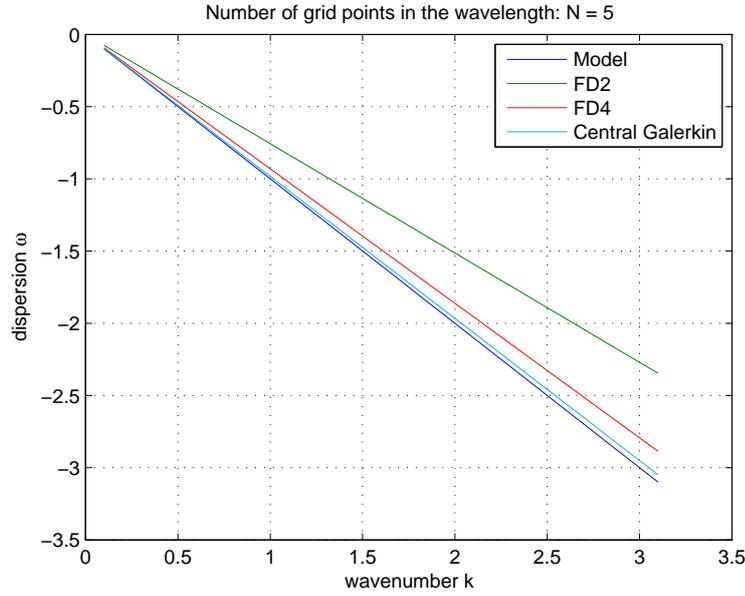
Imaging to substitute in (3.24) an exact solution of equation (3.1), a fourth order truncation error is found, which confirm the increase of the method accuracy given by the Galerkin mass matrix, since the approximation of the advective flux  $a \partial_x u$  results in the same expression derived for the second order finite difference scheme. Fourth order accuracy is here reached paying the cost of a the mass matrix inversion. However, Galerkin mass matrix is constant and the computational cost of this inversion results very low, since it can be done, in an efficiently way, only one time at the beginning of the integration procedure and then used all along the rest of the computational calculus. Nevertheless, as we will show in the following, this additional cost will be share by the FD schemes when the discretizations of terms characterized by mixed space-time derivatives must be accomplished.

**Dispersion Analysis :** Galerkin finite element method shows also better dispersion properties compared to the previous finite difference schemes. Using the Fourier mode of the generic variable  $u(x, t) = \bar{u} e^{\nu t + j k x}$  and (3.15) to express nodal value in the  $i+m$  degree of freedom of the mesh, time derivative becomes:

$$\partial_t u_{i+m} = \nu e^{\nu t + jm\mu\Delta x} = \nu u e^{jm\mu\Delta x} \quad (3.25)$$

Thus using (3.15) and (3.25) in (3.23), the following statement can be derived:

$$\frac{1}{6}\nu e^{j\mu\Delta x} + \frac{2}{3}\nu + \frac{1}{6}\nu e^{-j\mu\Delta x} + \frac{a}{2\Delta x} (e^{j\mu\Delta x} - e^{-j\mu\Delta x}) = 0 \quad (3.26)$$



**Figure 3.3:** Dispersion  $w$  against the wavenumber  $k$  of the Galerkin finite element scheme, compared to the scalar linear advection model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes  $N = 5$ .

which, through some manipulation, results in the expressions for  $\xi$  and  $\omega$ :

$$\xi_{CG} = 0 \quad (3.27)$$

$$\omega_{CG} = -ak \frac{\sin \mu\Delta x}{\mu\Delta x} \frac{3}{2 + \cos \mu\Delta x} \quad (3.28)$$

Obviously, the expression for  $\omega$  reflects the same second order as finite difference, but a second factor appears now, which depends directly from the mass

matrix and that improves sensibly the dispersion properties of the scheme. The following Figure 3.3 provides a direct comparison between (3.28) and the analogous expressions obtained for the finite difference schemes in the previous section. It can be seen that Galerkin finite element method performs even better than a fourth order finite difference scheme on the same spatial grid.

The dissipation parameter  $\xi$  is, instead, still correctly null.

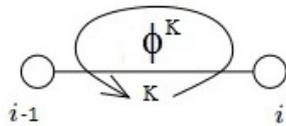
### 3.3 Residual Distribution Method

In the previous section the development of the Galerkin finite element scheme for the linear scalar advection problem was presented. We introduced the weak formulation of the original system (3.18) using  $\mathcal{P}^1$  function space for both basis and weighting functions and we then solved the integrals used Simpson's rule to approximate the definite integrals of the equations system (3.22). In this section a Residual Distribution scheme ( $\mathcal{RD}$ ) is developed [36], starting from the residual form of the (3.1):

$$\int_{\Omega} \omega_i (\partial_t u_h + a \partial_x u_h) dx = 0 \quad (3.29)$$

As already explained, this form is valid outside the variational formulation context making the hypothesis to work with continuous flux  $a \partial_x u$ . The residual distribution approach, which computes a numerical approximation of the solution, can be developed from (3.29), making the hypothesis of  $\omega_i = \text{const}$  in three main passages.

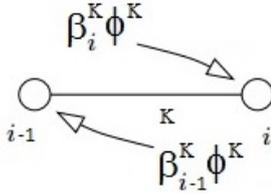
Given the initial values of the solution in the nodes of the mesh:



1. residual  $\Phi^K$  is computed on the initial nodal values  $\forall$  element  $K$  of the mesh:

$$\Phi^K = \int_K (\partial_t u_{h|_K} + a \partial_x u_{h|_K}) dx \quad (3.30)$$

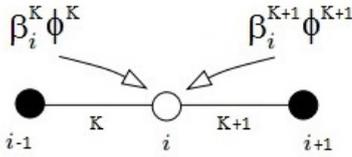
- 2 the amount of residual computed in each cell  $K$  is distributed between the nodes which belong to  $K$  in such a way that, denoting with  $\Phi_i^K$  the fraction of residual assigned to the node  $i \in K$ , where the fraction is expressed by the weighting coefficient  $\beta_i^K$ , for conservation and consistency requirements:



$$\sum_i \Phi_i^K = \Phi^K \quad (3.31)$$

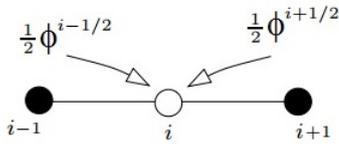
$$\Phi_i^K = \beta_i^K \Phi^K \quad (3.32)$$

- 3  $\forall$  nodes  $i \in \Omega_h$  nodal values are computed assembling residual fluxes from the adjacent elements. Being  $K_i$  the spatial domain which includes all the elements containing node  $i$  as a degree of freedom:



$$\sum_{K \in K_i} \Phi_i^K = 0 \quad (3.33)$$

Recalling (3.32), since we are working in the framework of a one-dimensional problem and each element  $K$  contains only two degrees of freedom, taking for  $\forall i \in K$  and  $\forall K \in \Omega$  a constant value for the distribution coefficient  $\beta = \frac{1}{2}$ , a central residual distribution scheme results. In fact, the residual computed over each cell will be divided and distributed in equal part on the two nodes of the cell. Each node  $i$  of the cell, will thus receive half of the amount of the residual from the right adjacent element and half from the left one, such that the final expression (3.33) of the scheme results in:



$$\frac{1}{2} \Phi^{i-1/2} + \frac{1}{2} \Phi^{i+1/2} = 0 \quad (3.34)$$

The same central  $\mathcal{RD}$  scheme (3.34) can be derived in analogy with what has already been done for the central Galerkin finite element method, given a more explicit expression to compute the fluxes  $\Phi_i^K$ . In fact, recalling (3.29),

using the Galerkin linear basis functions  $\varphi_i$  presented in the previous section instead of the generical constant weighting functions  $\omega_i$ , the expression for the node  $i$  of the grid will be:

$$\int_{i-1}^{i+1} \varphi_i (\partial_t u_h + a \partial_x u_h) dx = 0 \quad (3.35)$$

If now the middle point quadrature rule is used, instead of the Simpson one, to approximate definite integrals of the (3.35), since the new quadrature rule represents the exact evaluation of definite integral of linear functions, the resulting expression will be coincident with that of equation (3.23), except for the coefficient of the mass matrix. We have:

$$\frac{\Delta x}{4} \frac{d(u_{i-1})}{dt} + \frac{\Delta x}{2} \frac{d(u_i)}{dt} + \frac{\Delta x}{4} \frac{d(u_{i+1})}{dt} + \frac{a}{2} (u_{i+1} - u_{i-1}) = 0 \quad (3.36)$$

Finally, being  $\Phi^{i-\frac{1}{2}}$  the residual computed over the left adjacent cell to node  $i$ , having it the extremes in  $i$  and  $i-1$ , it can be assigned the following values:

$$\Phi^{i-\frac{1}{2}} = \Delta x \frac{d(u^{i-\frac{1}{2}})}{dt} + a (u_i - u_{i-1}) \quad (3.37)$$

such that (3.36) can be rewritten as:

$$\frac{1}{2} \Phi^{i-\frac{1}{2}} + \frac{1}{2} \Phi^{i+\frac{1}{2}} = 0 \quad (3.38)$$

and the connection with (3.34) is shown.

**Truncation Error :** A further confirmation of the great accuracy property guaranteed by the Galerkin mass matrix comes from the accuracy order fall (from fourth to second) caused by the changing of the matrix coefficients, which is due to the different quadrature rule used in the integration of (3.35). Using Fourier expansion to express the degrees of freedom  $i-1$  and  $i+1$ , as done for the previous methods, the following expression is obtained:

$$\begin{aligned} \partial_t u_i + a \partial_x u_i = & -\frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{2} \partial_t u_i + \frac{a}{3} \partial_x u_i \right) + \\ & -\frac{\Delta x^4}{24} \partial_{x^4} \left( \frac{1}{2} \partial_t u_i + \frac{a}{5} \partial_x u_i \right) + \mathcal{O}(\Delta x^6) \end{aligned} \quad (3.39)$$

which shows a  $\mathcal{O}(\Delta x^2)$  truncation error of the scheme, if the exact solution for the linear scalar advection problem is assume to be substituted to the first term of the expression.

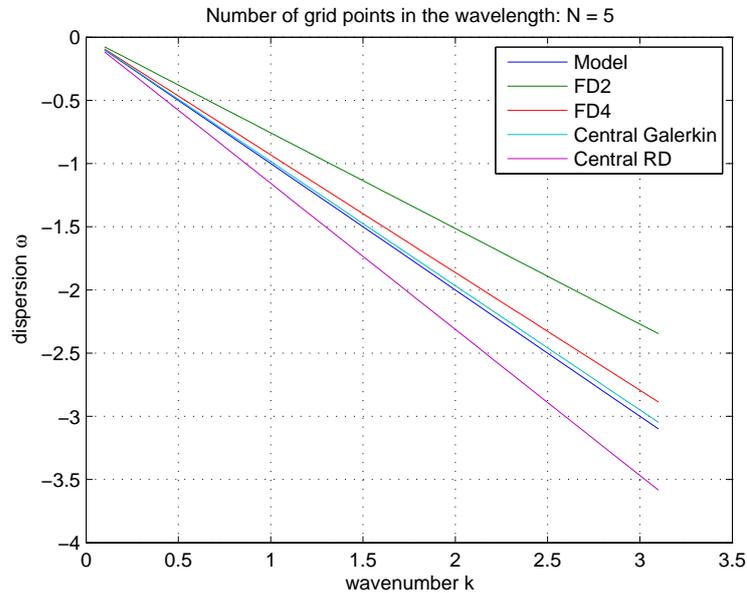
**Dispersion Analysis :** The dispersion analysis of the central residual distribution scheme (3.38) requires the same passages used for the Galerkin method, since the two schemes are almost the same. Consequently, by applying the same procedure, a very similar expression for parameters  $\xi$  and  $\omega$  can be found:

$$\xi_{\mathcal{RD}} = 0 \quad (3.40)$$

$$\omega_{\mathcal{RD}} = -ak \frac{\sin \mu_{\Delta x}}{\mu_{\Delta x}} \frac{2}{1 + \cos \mu_{\Delta x}} \quad (3.41)$$

Comparing (3.41) with (3.28), it becomes evident that the main difference between the two expressions consists in the factor which results from the mass matrix formulation. We have already shown how the residual distribution mass matrix worsens the high properties guaranteed by Galerkin ones; this is true also for the dispersion accuracy (see Figure 3.4).

Central residual distribution scheme (3.38) performed worst with respect to the Galerkin fourth order accuracy method, but shows a better reconstruction of the properties of the model with respect to the other second order finite difference scheme. The damping parameter  $\xi$  is always correctly null.



**Figure 3.4:** Dispersion  $w$  against the wavenumber  $k$  of the central  $\mathcal{RD}$  scheme, compared to the scalar linear advection model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes  $N = 5$ .

### 3.4 Stabilized Upwind Methods

In the previous development of central schemes to spatially discretize the scalar linear advection hyperbolic problem, we have shown how the Galerkin method leads to the best accuracy property and how this property is linked to the form of its mass matrix. However, in advective transport problems like the one we are studying, the matrix associated with the advective terms will not be symmetric. In particular in some regimes, especially in advection dominated fluxes where high Peclet or Reynolds local numbers are reached, it could happen that the best approximation property could be lost and solution could be corrupted by spurious node-to-node oscillations. The local Peclet number  $Pe_L$  represents the ratio between transport and diffusion terms. It can be defined and computed as:

$$Pe_L = \frac{|a|L}{\nu} \tag{3.42}$$

where  $\nu$  represent the diffusivity coefficient,  $a$  is the advection velocity and  $L$  a local reference length, such as the size of the grid element  $\Delta x$ . In accordance to its definition, instabilities in central methods computations originate when  $Pe_L > 1$  [12]. Equation (3.1) is a pure advection equation, which means that Peclet local number will be theoretically infinite and the solution computed could result inaccurate.

Unless the time integration is carefully chosen (e.g. an implicit method) the Galerkin method shows spurious modes for time-dependent advection. These modes cannot be eliminated for steady state computations, and are related to the uniqueness of the solution of the algebraic system associated to the scheme. In particular, the matrix associated to the bilinear Galerkin form  $b(u, v) = - \int_{\Omega} au_h \partial_x v_h dx$  results singular; this implies an infinite number of solutions which originate spurious modes.

A way to eliminate this kind of instabilities and oscillations in the solution is to make a severe refinement of the mesh, such that advection no longer dominates on the element level. Of course mesh refinement is not always possible and deeply influences the computational cost of the solution calculation, for these reasons an alternative to the Galerkin formulation is needed which does not exhibit these spurious oscillations. Additional stabilizing terms can also be used to control oscillations in presence of large gradients and local discontinuous profiles [19], [38]. Upwind differentiation of the advective term allows to obtain a solution free from spurious contaminations and instabilities: upwind differencing amounts to approximating the advective derivatives with solution values at the upstream and central nodes of a three points stencil. In dependence of the value of the advection velocity  $a$ , equation (3.1) can be discretized as:

$$\frac{d(u_i)}{dt} + \frac{a^+}{\Delta x} (u_i - u_{i-1}) = 0 \quad (3.43)$$

$$\frac{d(u_i)}{dt} + \frac{a^-}{\Delta x} (u_{i+1} - u_i) = 0 \quad (3.44)$$

although this approximation of the first spatial derivative is only first order accurate and, thus, it is necessary to renounce to the second order accuracy of the central finite difference methods (3.10). This loss of accuracy is manifested in over-diffuse solutions.

Upwinded advective terms can be built by simply adding artificial diffusion to a central scheme. In fact a combination of a central scheme with an upwind diffusive term gives better results with respect to a central scheme or

an upwind one alone. However upwind stabilisation will be useless or even can produce worse results in flows where Peclet number is  $Pe_L < 1$ , due to the fact that central Galerkin is already a stable and more accurate method for these fluxes.

In the finite element framework the diffusive upwind term can be computed in several different ways. The first idea is to employ modified weighting functions  $\omega_i$  to achieve the upwind effect. In general the element upstream a node is weighted more heavily than one downstream. In the one-dimensional case, like the one here treated, artificial diffusion, when optimally selected, balances the negative diffusion inherent in the Galerkin method, thus resulting in the exact nodal solutions for the model problem. However multidimensional generalisation can be unsuccessful due to crosswind diffusion problems. In the Streamline Upwind Petrov-Galerkin weighted residual method (*SU/PG*) the basic idea is to add diffusion (or artificial viscosity) in the flow direction only. Thus, standard Galerkin weighted functions must be modified by adding a streamline upwind perturbation, which again acts only in the flow direction. The modified weighting functions can then be applied to all the terms of the governing equation, resulting in this way in a consistent weighted residual formulation [36], [38].

In Galerkin formulation weighting functions are considered to be continuous across inter-elements boundaries. The streamline upwind Petrov-Galerkin method requires, instead, discontinuous weighting functions  $\omega_i$  which can be obtained summing to the continuous central functions  $\varphi_i$  the discontinuous streamline upwind contributions  $\delta\varphi_i$ :

$$\omega_i = \varphi_i + \delta\varphi_i \tag{3.45}$$

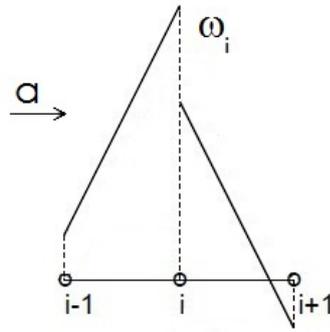
For what discussed above, the general *SU/PG* formulation for the one-dimensional linear scalar advection problem results in the following form:

$$\int_{\Omega} \varphi_i \partial_t u_h dx + \int_{\Omega} \varphi_i a \partial_x u_h dx + \sum_{K \in \Omega} \int_K \delta\varphi_i (\partial_t u_h + a \partial_x u_h) dx = 0 \tag{3.46}$$

being the expression for the streamline upwind weighting operator:

$$\delta\varphi_i = a \partial_x \varphi_i \tau \tag{3.47}$$

$\tau$  is a scale parameter which controls the quantity of artificial viscosity inserted into the model and is set such as  $\tau = \Delta x / (2|a|)$ . The upwind term added to the Galerkin scheme in equation (3.46), increases the weight of the left flux, and diminishes the right one, under the hypothesis of a positive advection velocity  $a > 0$ , since the gradient  $\partial_x \varphi_i$  is positive over the left adjacent cell of node  $i$ , and is negative over the right one. Therefore the one-dimensional Galerkin linear weighted functions of Figure 2.1 are modified obtaining the discontinuous  $\omega_i$  which assumes a shape like the one shown in the following Figure 3.5.



**Figure 3.5:** Representation of the upwind linear finite basis function over the element  $e$ .

Introducing the residual  $\Phi_K$  defined as in (3.30), and being  $CS$  the central scheme which has to be stabilized, formulation (3.46) can be rewritten, referred to a generic internal node of the mesh, as:

$$CS + \frac{\text{sign}(a)}{2} \Phi^{i-\frac{1}{2}} - \frac{\text{sign}(a)}{2} \Phi^{i+\frac{1}{2}} = 0 \quad (3.48)$$

Let now  $CS$  be the central Galerkin finite element discretization of the linear scalar advection (3.1), then the scheme resulting can be recast as:

$$\frac{5\Delta x}{12} \frac{d(u_{i-1})}{dt} + \frac{2\Delta x}{3} \frac{d(u_i)}{dt} - \frac{\Delta x}{12} \frac{d(u_{i+1})}{dt} + a(u_i - u_{i-1}) = 0 \quad (3.49)$$

In (3.49), as a confirmation of the fact that an upwind weight has been applied to the Galerkin central finite element scheme, it could be noted that

the advection flux results approximated with the first order upwind finite difference (3.43) and that the coefficients of the mass matrix are no more symmetric in a way such that, since  $a$  has been assumed positive, more weight is given to the upwind node  $i-1$  and less to the downwind one  $i+1$ . The form of equation (3.48) suggests to create a  $SU/PG$ -type stabilized scheme using the just developed central residual distribution scheme, instead of the Galerkin one. In this case, in fact, the method which could be obtained, would be an upwind residual distribution scheme, respecting the formulation:

$$\sum_{K \in \Omega} \beta_i^K \Phi^K = 0 \quad (3.50)$$

with the distribution coefficients over the cell  $\beta_i^K$  now given by the sum of the central weighting functions and the upwind ones in perfect analogy with the  $SU/PG$  just developed:

$$\beta_i^K = \frac{1}{2} + \frac{\text{sign}(a)}{2} \quad (3.51)$$

In this case, the upwind scheme will simply result:

$$\Phi^{i-\frac{1}{2}} = \frac{\Delta x}{2} \frac{d(u_{i-1})}{dt} + \frac{\Delta x}{2} \frac{d(u_i)}{dt} + a(u_i - u_{i-1}) = 0 \quad (3.52)$$

which even reduces the stencil of only two nodes, although diminishing the high order accuracy guaranteed by the  $SU/PG$  method, as it will be shown in the following.

**Troncation Error :** Streamline upwind Petrov-Galerkin method is a consistently stabilized method, due to the fact that, in contrast with what seen for finite difference upwind method (3.43) and (3.44), the order of the central Galerkin approximation will not be affected. Truncation error analysis made on the scheme related to equation (3.49) reveals that a third order accuracy is preserved:

$$\begin{aligned}
 \partial_t u_i + a \partial_x u_i &= \frac{\Delta x}{2} \partial_x (\partial_t u_i + a \partial_x u_i) + \\
 &\quad - \frac{\Delta x^2}{6} \partial_{x^2} (\partial_t u_i + a \partial_x u_i) + \\
 &\quad + \frac{\Delta x^3}{12} \partial_{x^3} \left( \partial_t u_i + \frac{a}{2} \partial_x u_i \right) + \mathcal{O}(\Delta x^4)
 \end{aligned} \tag{3.53}$$

The analogous upwind method made by central residual distribution scheme shows instead a second order accuracy, which, even being lower with respect to the streamline upwind Petrov-Galerkin one, at least preserves the truncation accuracy of its central scheme.

$$\begin{aligned}
 \partial_t u_i + a \partial_x u_i &= \frac{\Delta x}{2} \partial_x (\partial_t u_i + a \partial_x u_i) + \\
 &\quad - \frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{2} \partial_t u_i + \frac{a}{3} \partial_x u_i \right) + \mathcal{O}(\Delta x^3)
 \end{aligned} \tag{3.54}$$

**Dispersion Analysis :** Dispersion analysis of the stabilized methods results in more complicated expression for variables  $\xi$  and  $\omega$  of these schemes. Taking into account the  $\mathcal{SU}/\mathcal{PG}$  scheme (3.52), written separating the central scheme contribution from the stabilizing term in the form:

$$\begin{aligned}
 \frac{\Delta x}{6} (\partial_t u_{i-1} + 4 \partial_t u_i + \partial_t u_{i+1}) + \frac{a}{2} (u_{i+1} - u_{i-1}) + \\
 + \frac{\Delta x}{4} (\partial_t u_{i-1} - \partial_t u_{i+1}) - \frac{a}{2} (u_{i+1} - 2u_i + u_{i-1}) = 0
 \end{aligned} \tag{3.55}$$

the generic Fourier mode can now be inserted into (3.55) obtaining the complex equation:

$$\frac{2 + \cos(\mu_{\Delta x})}{3} \nu + i a k \frac{\sin \mu_{\Delta x}}{\mu_{\Delta x}} - \frac{i \nu}{2} \sin \mu_{\Delta x} + \frac{a k}{\mu_{\Delta x}} (1 - \cos(\mu_{\Delta x})) = 0 \tag{3.56}$$

with  $\nu = \xi + j\omega$  and  $\mu_{\Delta x} = k\Delta x$ . Cause to the upwind term, the previous equation results in a coupled system of the two equations for the real and

imaginary part. This lead to an expression for the damping  $\xi$ , which is no more zero as in the central Galerkin scheme, while the expression of  $\omega$  very clearly shows the two contributions given by the central and stabilization terms. In particular we have:

$$\xi_{SU/\mathcal{P}\mathcal{G}} = \frac{ak \frac{\sin^2 \mu_{\Delta x}}{2\mu_{\Delta x}} - ak(1 - \cos \mu_{\Delta x}) \frac{2 + \cos(\mu_{\Delta x})}{3\mu_{\Delta x}}}{\frac{(2 + \cos \mu_{\Delta x})^2}{9} + \frac{\sin^2 \mu_{\Delta x}}{4}} \quad (3.57)$$

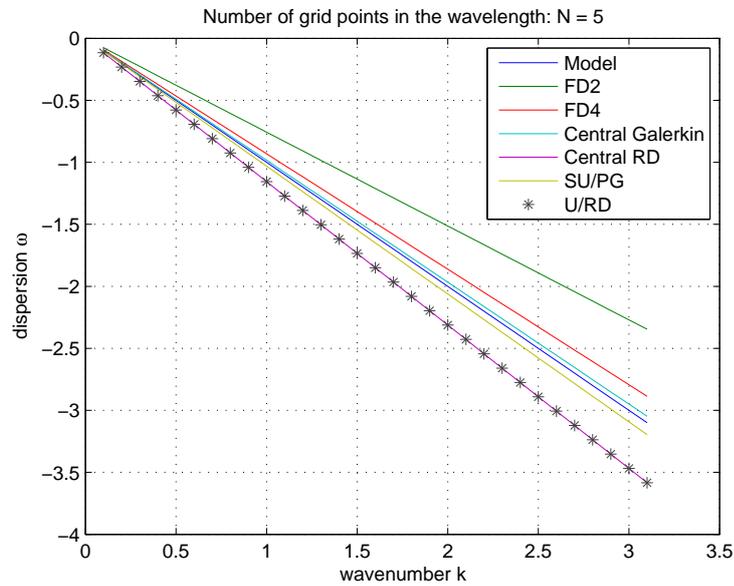
$$\omega_{SU/\mathcal{P}\mathcal{G}} = \frac{-ak \frac{\sin \mu_{\Delta x}}{\mu_{\Delta x}} + \frac{\sin \mu_{\Delta x}}{2} \xi_{SU/\mathcal{P}\mathcal{G}}}{\frac{2 + \cos \mu_{\Delta x}}{3}} \quad (3.58)$$

The same analysis, done for the upwind residual distribution scheme (3.52), gives similar results:

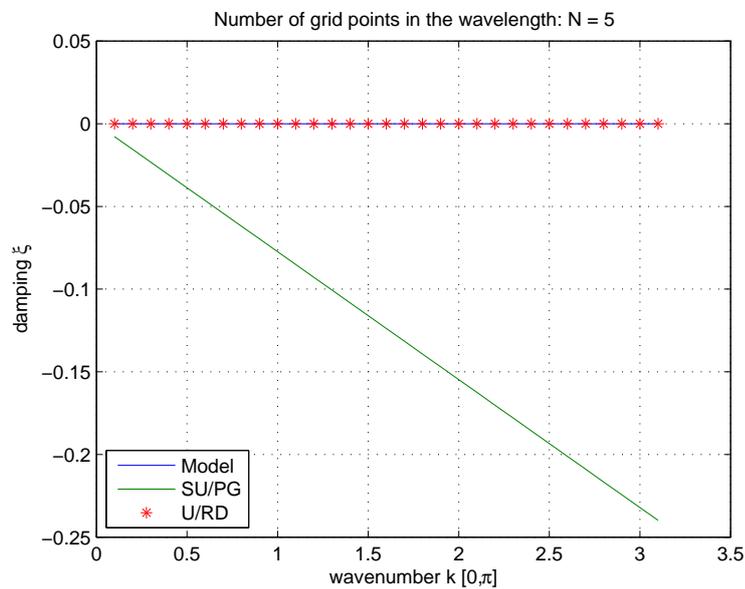
$$\xi_{U/\mathcal{R}\mathcal{D}} = \frac{ak \frac{\sin^2 \mu_{\Delta x}}{2\mu_{\Delta x}} - ak(1 - \cos \mu_{\Delta x}) \frac{1 + \cos(\mu_{\Delta x})}{2\mu_{\Delta x}}}{\frac{(1 + \cos \mu_{\Delta x})^2}{4} + \frac{\sin^2 \mu_{\Delta x}}{4}} \quad (3.59)$$

$$\omega_{U/\mathcal{R}\mathcal{D}} = \frac{-ak \frac{\sin \mu_{\Delta x}}{\mu_{\Delta x}} + \frac{\sin \mu_{\Delta x}}{2} \xi_{U/\mathcal{R}\mathcal{D}}}{\frac{1 + \cos \mu_{\Delta x}}{2}} \quad (3.60)$$

Figure 3.6) compares the expression just found with those of the central schemes above showing that, as for the truncation order accuracy, the streamline upwind Petrov-Galerkin method slightly worsens the accurate Galerkin dispersion description with respect to the exact model, however showing a much better behaviour than a first order completely upwind scheme thank to its Galerkin-type mass matrix. The upwind residual distribution scheme, instead, surprisingly seems to not modify dispersion and dissipation properties of the central scheme, since even the complicated expression (3.58) results, at the end, to be equal to zero from Figure 3.7.



**Figure 3.6:** Dispersion  $w$  against the wavenumber  $k$  of the stabilized upwind schemes, compared to the scalar linear advection model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes  $N = 5$ .



**Figure 3.7:** Diffusion  $\xi$  against the wavenumber  $k$  of the stabilized upwind  $SU/PG$  and  $U/RD$  schemes, compared to the scalar linear advection model, when the wavelength is discretized with a number of grid nodes  $N = 5$ .

## 3.5 Conclusions

In this section, we have spatially discretized the one-dimensional linear scalar advection equation using different numerical methods. We have shown that Galerkin  $\mathcal{P}^1$  finite element method results in a central scheme with the major accuracy order compared to central difference and residual distribution schemes which use the same approximation of the first order derivative present in the equation. The cost, which has to be paid for this higher accuracy property consists in the inversion of the mass matrix which, however, can be done efficiently due to the fact that it is constant and, once computed, the inverse matrix can be used all along the calculus. It has been illustrated how Galerkin fourth order accuracy and better dispersion properties are directly linked to the form of its mass matrix and that a finite difference scheme can result by directly lumping its mass matrix, as a central residual distribution scheme can be obtained by directly integrating the weak form using a middle point quadrature rule; but these methods are only second order accurate.

We have spoken about the raise of instabilities and spurious oscillations in advection dominated fluxes — like the ones described by equation (3.1), when they are discretized using centred schemes — and, thus, of the necessity of adding upwinding terms of streamline artificial diffusion deriving, in such a way, a stabilized  $SU/\mathcal{P}\mathcal{G}$  method. In particular the upwind term has been obtained changing the weighting functions of the variational formulation. This fact produces a stabilization term in a residual form which suggests the idea of exchanging the centred scheme previously used in the  $SU/\mathcal{P}\mathcal{G}$  derivation (Galerkin) with the  $\mathcal{RD}$  one. An upwind residual distribution scheme was, hence, derived in perfect analogy with the streamline upwind artificial diffusion theory for advection dominated problems stabilisation. Both are consistently stabilized methods, since they conserved the accuracy order of their centred part and are, for this reason, preferable with respect to complete upwind schemes.

Dispersion analysis of all this techniques shows that central Galerkin finite element methods reaches the best performance in the description of the dispersive behaviour of the model. This high performance are only weakly worsened by adding a stabilisation term in the  $SU/\mathcal{P}\mathcal{G}$  scheme, which, however, introduces some non-physical dissipation in the scheme, which results in a no more null value of the parameter  $\xi$  for this upwind method, while it remains surprisingly zero for the upwind residual distribution scheme.



## Chapter 4

# Linear Scalar Advection with Dispersion Equation

One-dimensional advection-dispersion equation is here considered:

$$\partial_t u + a \partial_x u - \alpha \partial_{x^2} u = 0 \quad (4.1)$$

we recall that the brief notation  $\partial_{x^2}$  is used to indicate the recursive application of the partial derivative with respect to  $x$  for two times, while  $\partial_{xt}$  indicates the mixed derivative with respect to space and time.

Making the hypothesis of flow velocity  $a$  and dispersivity  $\alpha$  constant and positive, the local Peclet number of the flow defined in the same way as in (3.42) still has an infinite value, and therefore the necessity of stabilizing the central schemes with upwind terms is still present and it is desirable to avoid an excessive refinement of the mesh.

In the following, the central schemes, already proposed for the scalar advection case, are going to be developed in the same way of the previous chapter, showing the major difficulties which arise from the presence, in (4.1), of a second order spatial derivative term. After this, repeating what already done in section 3.4, diffusive upwind terms will be added to the centred schemes in order to give them a major stabilization against spurious oscillations.

**Dispersion Analysis :** In the scalar linear advection-dispersion equation, replacing the Fourier mode  $u = \bar{u} e^{\nu t + j k x}$  in (4.1), a very simple expression

can be obtained:

$$(1 + \alpha k^2)\nu + jak = 0 \quad (4.2)$$

which shows that for the advection with dispersion case:

$$\xi = 0 \quad (4.3)$$

$$\omega = -\frac{ak}{1 + \alpha k^2} \quad (4.4)$$

This means that the damping  $\xi$  still remains the same as the simple advection problem and only the dispersion parameter  $\omega$  is affected by the additional term  $\alpha\partial_{x^2}u$ , which thus gives to the model a higher dispersion property.

## 4.1 Finite Difference Methods

The development of a finite difference scheme for spatially discretized equation (4.1) follows directly what has been done for the scalar advection equation. In fact first spatial derivative of the advective term  $a\partial_x u$  is approximated using the already known formulae (3.7) or (3.9). For the dispersive term, instead, a formula for the approximation of the second order spatial derivative is needed. Finite difference formulae which allow to obtain the second and fourth truncation error accuracy order are listed below:

$$\frac{d^2 f}{dx^2}(x_i) = \frac{1}{\Delta x^2} (f_{i-1} - 2f_i + f_{i+1}) + \mathcal{O}(\Delta x^2) \quad (4.5)$$

$$\frac{d^2 f}{dx^2}(x_i) = \frac{1}{12\Delta x^2} (-f_{i-2} + 16f_{i-1} - 30f_i + 16f_{i+1} - f_{i+2}) + \mathcal{O}(\Delta x^4) \quad (4.6)$$

Using (4.5) and (4.6) to discretize the second order spatial derivative, together with the respective (3.7) and (3.9) to discretize the first order one, the following second and fourth order schemes can be obtained:

$$\frac{d(u_i)}{dt} + \frac{a}{2\Delta x} (u_{i+1} - u_{i-1}) - \frac{\alpha}{\Delta x^2} \left( \frac{d(u_{i-1})}{dt} - 2\frac{d(u_i)}{dt} + \frac{d(u_{i+1})}{dt} \right) = 0 \quad (4.7)$$

$$\begin{aligned} \frac{d(u_i)}{dt} + \frac{a}{12\Delta x} (u_{i-2} - 8u_{i-1} + 8u_{i+1} - u_{i+2}) + \\ - \frac{\alpha}{12\Delta x^2} \left( -\frac{d(u_{i-2})}{dt} + 16\frac{d(u_{i-1})}{dt} - 30\frac{d(u_i)}{dt} + 16\frac{d(u_{i+1})}{dt} + \right. \\ \left. - \frac{d(u_{i+2})}{dt} \right) = 0 \end{aligned} \quad (4.8)$$

Since the formulae used for the approximation of derivatives are centred, central finite difference schemes result. They could, thus, perform bad when near-discontinuous behaviour raises in the flux or when advection dominated flux are analyzed. As anticipated in the previous chapter and observed here, the dispersive terms generates a stiffness matrix which has to be inverted in order to obtain the explicit expression for  $d(u_i)/dt$ , which allows to compute the solution. The previous advantage of this class of methods, that was due to the fact that any matrix inversion was needed in order to solve the system, is now lost.

**Truncation Error :** As for the scalar advection case, a truncation error analysis of the schemes (4.7) and (4.8) confirms that the enlargement of the stencil allows to raise up to fourth order. A more accurate numerical solution could be computed, but an increase of the computational cost must be paid.

$$\begin{aligned} \partial_t u_i + a\partial_x u_i - \alpha\partial_{xxt} u_i = -\frac{\Delta x^2}{6} \partial_{x^2} \left( a\partial_x u_i - \frac{\alpha}{2} \partial_{xxt} u_i \right) + \\ - \frac{\Delta x^4}{120} \partial_{x^4} \left( a\partial_x u_i - \frac{\alpha}{3} \partial_{xxt} u_i \right) + \mathcal{O}(\Delta x^6) \end{aligned} \quad (4.9)$$

$$\partial_t u_i + a\partial_x u_i - \alpha\partial_{xxt} u_i = -\frac{\Delta x^4}{30} \partial_{x^4} (a\partial_x u_i - \alpha\partial_{xxt} u_i) + \mathcal{O}(\Delta x^6) \quad (4.10)$$

**Dispersion Analysis :** Since schemes (4.7) and (4.8) differ from the analogous of the linear scalar advection equation only because of the dispersive term, and due to the fact that this discretized term only results in a term

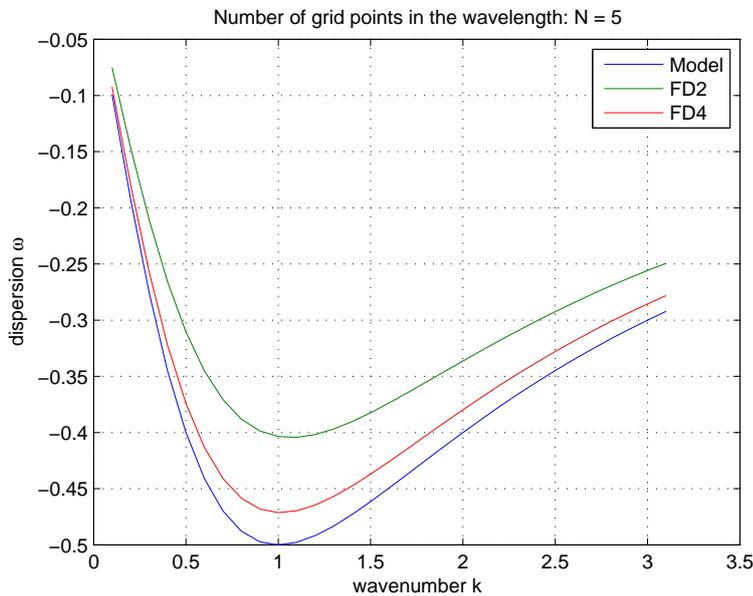
multiplying the complex number  $\nu$ , since it contains a time derivative, this affects only the dispersive property of the schemes, leaving the dissipation one unchanged with respect to the advection case and thus still null. The expression of  $\xi$  and  $\omega$  of the second and fourth order finite difference schemes are thus listed below. A comparison between the dispersion of the two schemes with the exact model one is then shown in Figure 4.1.

$$\xi_{\mathcal{FD}2} = 0 \quad (4.11)$$

$$\omega_{\mathcal{FD}2} = -\frac{ak \frac{\sin \mu_{\Delta x}}{\mu_{\Delta x}}}{1 + \frac{2\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x})} \quad (4.12)$$

$$\xi_{\mathcal{FD}4} = 0 \quad (4.13)$$

$$\omega_{\mathcal{FD}4} = -\frac{ak \frac{\sin \mu_{\Delta x}}{3\mu_{\Delta x}} (4 - \cos \mu_{\Delta x})}{1 + \frac{\alpha k^2}{6\mu_{\Delta x}^2} (15 - 16 \cos \mu_{\Delta x} + \cos 2\mu_{\Delta x})} \quad (4.14)$$



**Figure 4.1:** Dispersion  $w$  against the wavenumber  $k$  of the finite difference  $\mathcal{FD}2$  and  $\mathcal{FD}4$  schemes, compared to the scalar linear advection with dispersion model, when the wavelength is discretized with a number of grid nodes  $N = 5$ .

## 4.2 Galerkin finite element method

All the discussions and considerations made in Section 3.2 about the weighted residual Galerkin method are still valid for the present advection with dispersion problem. Thus, Galerkin finite element discretization starts first from the weak formulation of the discretized advection-dispersive equation, which is:

$$\int_{\Omega} \psi_i \partial_t u_h dx - \int_{\Omega} a u_h \partial_x \psi_i dx + \int_{\Omega} \alpha \partial_x \psi_i \partial_{xt} u_h dx = \mathcal{BC}s \quad (4.15)$$

In (4.15),  $\mathcal{BC}s$  stands for boundary evaluated terms which derive from the integration by part and which can result null if appropriate boundary conditions are imposed to the problem.

Now, recalling the use of  $\mathcal{P}^1$  centred basis function  $\varphi_j$  to approximate the solution in the  $N + 1$  degree of freedom of the domain, since in the Galerkin method both weighting functions  $\psi_i$  and basis functions  $\varphi_j$  must be chosen from the same functional space, (4.15) becomes:

$$\begin{aligned} \left( \int_{X_1}^{X_2} \varphi_i \varphi_j dx \right) \frac{du_j}{dt} - \left( \int_{X_1}^{X_2} \varphi_j \partial_x \varphi_i dx \right) a u_j + \\ + \left( \int_{X_1}^{X_2} \partial_x \varphi_i \partial_x \varphi_j dx \right) \alpha \frac{du_j}{dt} = 0 \end{aligned} \quad (4.16)$$

In this expression a stiffness matrix arises from the dispersion term. Its value is easy to be computed due to the particular form of the chosen functions and, as in the case of the mass matrix, the integral over the entire spatial domain can be split into a sum of integrals over the elements which then result in a element-by-element assembly procedure. In practice, using the Simpson quadrature formula to solve definite integrals of (4.16) in the exact way, the following expression centred on node  $i$  of the mesh is found:

$$\begin{aligned} \frac{\Delta x}{6} \frac{d(u_{i-1})}{dt} + \frac{2\Delta x}{3} \frac{d(u_i)}{dt} + \frac{\Delta x}{6} \frac{d(u_{i+1})}{dt} + \frac{a}{2} (u_{i+1} - u_{i-1}) + \\ - \frac{\alpha}{\Delta x} \left( \frac{d(u_{i+1})}{dt} - 2 \frac{d(u_i)}{dt} + \frac{d(u_{i-1})}{dt} \right) = 0 \end{aligned} \quad (4.17)$$

Simpson integration results into the central second order finite difference approximation for both first and second order spatial derivatives. The Galerkin scheme obtained can thus degenerate to the second finite difference discretization method (4.7) when a mass lumping procedure is applied.

**Truncation Error :** Truncation error analysis of this scheme shows that the fourth approximation order of the solution, which characterises the method in the scalar advection case, is here lost. The second order approximation of the spatial derivative of the dispersive term does not allow the Galerkin mass matrix to recover a higher accuracy order. The expression of truncation error becomes:

$$\begin{aligned} \text{TE}_{CG} = & -\frac{\Delta x^2}{6} \partial_{x^2} \left( \partial_t u_i + a \partial_x u_i - \frac{\alpha}{2} \partial_{x^2 t} u_i \right) + \\ & -\frac{\Delta x^4}{12} \partial_{x^4} \left( \frac{1}{6} \partial_t u_i + \frac{a}{10} \partial_x u_i - \frac{\alpha}{30} \partial_{x^2 t} u_i \right) + \mathcal{O}(\Delta x^6) \end{aligned} \quad (4.18)$$

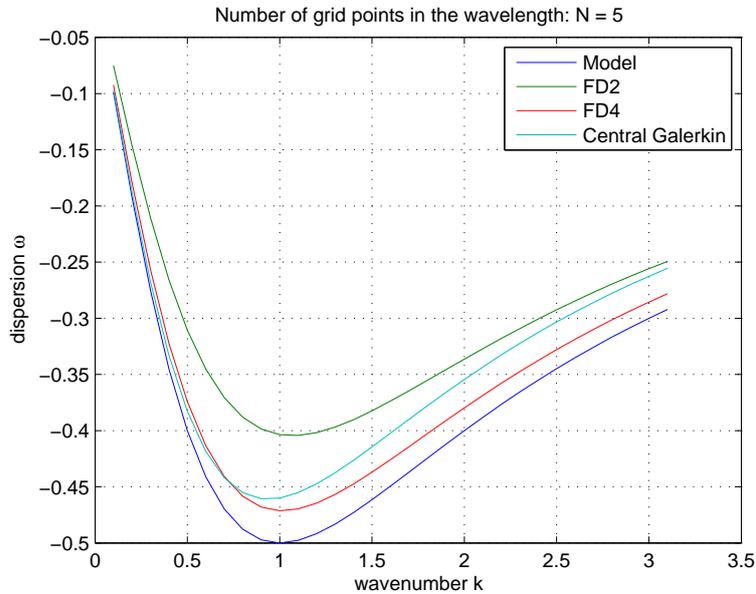
**Dispersion Analysis :** As far as the dispersion properties of the scheme is concerned, the Fourier analysis of the method returns the values:

$$\xi_{CG} = 0 \quad (4.19)$$

$$\omega_{CG} = -\frac{ak \frac{\sin \mu_{\Delta x}}{\mu_{\Delta x}}}{\frac{(2 + \cos \mu_{\Delta x})}{3} + 2 \frac{\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x})} \quad (4.20)$$

For this method, like for the previous finite difference scheme and for the exact model, the addition of the dispersive term to the linear scalar advection equation does not introduce any dissipation property which, therefore, remains exactly null. The dispersion character of the scheme produces a little worse simulation of the diffusivity of the model with respect to the previous case because of the lower accuracy order. However Figure 4.2 shows that the schemes perform better, or at least at the same level, of a high accuracy fourth order scheme ( $\mathcal{FD4}$ ) until wavenumbers values of about  $k \approx 0.8$ . For

higher values of  $k$ , however, even if the behaviour is such that it reaches the second order scheme approximation, a better performance with respect to the second order finite difference scheme ( $\mathcal{FD}2$ ) results.



**Figure 4.2:** Dispersion  $w$  against the wavenumber  $k$  of the Galerkin finite element scheme, compared to the scalar linear advection with dispersion model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes  $N = 5$ .

### 4.3 Residual distribution method

The derivation of a central residual distribution method of the kind of equation (3.34) for the scalar linear advection with dispersion equation is more complicated with respect to what has been done for the other central schemes, simply operating the discretization of the dispersive term  $\alpha \partial_{x^2 t} u$  and adding this to the original scheme computed for the scalar advection equation case. In the same previous way, what has to be done is to evaluate the residual  $\Phi^K$  on each element on the mesh. This residual is now:

$$\Phi^K = \int_K (\partial_t u_{h|_K} + a \partial_x u_{h|_K} - \alpha \partial_{x^2 t} u_{h|_K}) dx \quad (4.21)$$

and must then be distributed in equal parts between the nodes which belong to the same elements, such that, from the point of view of each node  $i$  of the mesh, it is possible to sum all the residual fluxes which come from the adjacent nodes in order to obtain its nodal value. We have already illustrated that in the one-dimensional case it means that the residual computed on a cell must be split between the two degrees of freedom of the element such that the distribution coefficients values assume the value  $\beta_i^K = \frac{1}{2}$  and the spatial discretization scheme simply results in:

$$\frac{1}{2}\Phi^{i-\frac{1}{2}} + \frac{1}{2}\Phi^{i+\frac{1}{2}} = 0 \quad (4.22)$$

In order to compute the exact expression of the spatially discretized residuals  $\Phi^{i-\frac{1}{2}}$  and  $\Phi^{i+\frac{1}{2}}$  of equation (4.22), the following formulation of the original problem has to be taken into account:

$$\int_{\Omega} \omega_i (\partial_t u_h + a \partial_x u_h - \alpha \partial_{x^2 t} u_h) dx = 0 \quad (4.23)$$

Since the possibility to use integration by parts to transfer one spatial derivative on the weighting functions  $\omega_i$  is not allowed and  $\mathcal{P}^1$  basis functions are used to approximate the nodal values of the problem variable, a possibility is to rewrite the strong formulation of the advection-dispersion equation in a lower order form, introducing a new variable  $w$  to replace the first or the second spatial derivative of  $u$ . In Appendix B it is shown that the choice of the type of reconstruction results completely arbitrary, thus both maintaining the mixed derivative and decreasing the order of the equation to one does not influence the approximation of the dispersion term which, independently on the choice, results in the expression:

$$-\frac{\alpha}{4\Delta x} \left( \frac{d(u_{i+2})}{dt} - 2 \frac{d(u_i)}{dt} + \frac{d(u_{i-2})}{dt} \right) \quad (4.24)$$

Anticipating the choice that will be done in the complete Madsen-Sørensen context, we now decide to set:

$$w = \partial_{x^2} u \quad (4.25)$$

The auxiliary variable, defined in such a way, increases the numbers of unknowns of the original equation (4.1). The new definition (4.25) must now be discretized, integrated and solved together with the scalar linear advection-dispersion equation; the system we are going to deal with, thus, results:

$$\begin{cases} \partial_t u_i + a \partial_x u_i - \alpha \partial_t w_i = 0 \\ w_i = \partial_{x^2} u_i \end{cases} \quad (4.26)$$

Imaging to compute their residual formulation using a  $\mathcal{P}^1$  weighting function of the kind of Galerkin finite element scheme, and solving the definite integrals on the spatial domain using the middle point quadrature rule, since a residual distribution scheme is searched, the following spatially discretized schemes can be found:

$$\begin{cases} \frac{\Delta x}{4} \frac{d(u_{i-1})}{dt} + \frac{\Delta x}{2} \frac{d(u_i)}{dt} + \frac{\Delta x}{4} \frac{d(u_{i+1})}{dt} + \frac{a}{2} (u_{i+1} - u_{i-1}) \\ \quad - \frac{\alpha \Delta x}{4} \left( \frac{d(w_{i+1})}{dt} - 2 \frac{d(w_i)}{dt} + \frac{d(w_{i-1})}{dt} \right) = 0 \\ \frac{\Delta x}{4} w_{i-1} + \frac{\Delta x}{2} w_i + \frac{\Delta x}{4} w_{i+1} = \frac{1}{\Delta x} (u_{i+1} - 2u_i + u_{i-1}) \end{cases} \quad (4.27)$$

The mass lumping technique can be applied on the auxiliary equation of the system in order to reduce the computational work added to solve it. Computing directly the value of  $w_i$  through the second order finite difference approximation (4.5) decreases the accuracy of the reconstruction, however the following analysis shows that the spatial truncation error of the scheme is not significantly affected by this approximation. The first equation of the system (4.27), instead, can now be directly reorganized and written in the desired residual distribution form:

$$\frac{1}{2} \Phi^{i-\frac{1}{2}} + \frac{1}{2} \Phi^{i+\frac{1}{2}} = 0 \quad (4.28)$$

$\Phi^{i-\frac{1}{2}}$  is defined as the central residual flux computed on the one-dimensional cell, enclosed between nodes  $i$  and  $i-1$ :

$$\Phi^{i-\frac{1}{2}} = \Delta x \frac{d(u_{i-\frac{1}{2}})}{dt} + a(u_i - u_{i-1}) + \Delta x \frac{d(w_{i-\frac{1}{2}})}{dt} \quad (4.29)$$

**Truncation Error :** Truncation error of the residual distribution scheme must be computed first of all by writing the spatially discretized method in the following form, using (B.11), which is demonstrated in Appendix A to be equivalent to finite different approximation of the dispersive term made after the introduction of the auxiliary variable  $w$ :

$$\begin{aligned} \frac{\Delta x}{4} \frac{du_{i-1}}{dt} + \frac{\Delta x}{2} \frac{du_i}{dt} + \frac{\Delta x}{4} \frac{du_{i+1}}{dt} + \frac{a}{2}(u_{i+1} - u_{i-1}) + \\ - \frac{\alpha}{4\Delta x} \left( \frac{du_{i+2}}{dt} - 2\frac{du_i}{dt} + \frac{du_{i-2}}{dt} \right) = 0 \end{aligned} \quad (4.30)$$

Now, introducing in (4.30) the Taylor expansions shown in (3.8), the following truncation error expansion can be found:

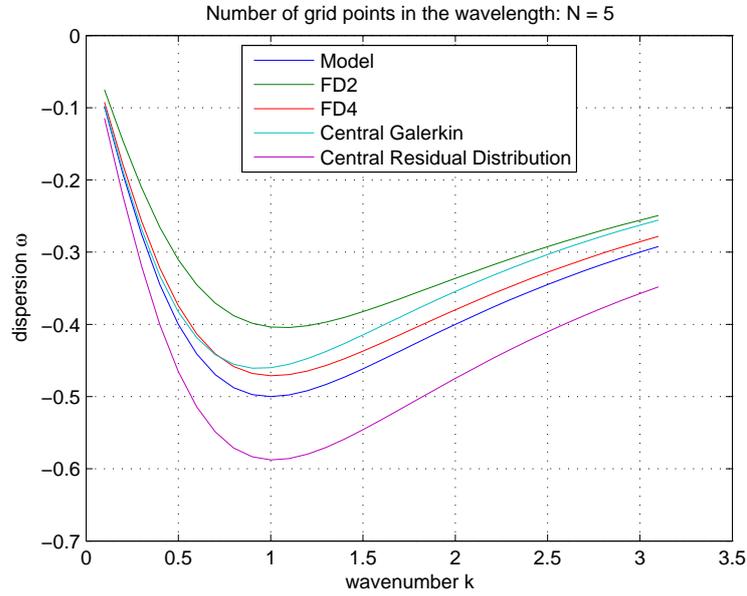
$$\begin{aligned} \text{TE}_{\mathcal{RD}} = & -\Delta x^2 \partial_{x^2} \left( \frac{1}{4} \partial_t u_i + \frac{a}{6} \partial_x u_i - \frac{\alpha}{3} \partial_{x^2 t} u_i \right) + \\ & - \frac{\Delta x^4}{24} \partial_{x^4} \left( \frac{1}{2} \partial_t u_i + \frac{a}{5} \partial_x u_i - \frac{32}{3} \alpha \partial_{x^2 t} u_i \right) + \\ & + \mathcal{O}(\Delta x^6) \end{aligned} \quad (4.31)$$

which shows that the resulting central scheme is still a second order accuracy method.

**Dispersion Analysis :** The study of the dispersive and dissipative properties of the central residual distribution method moves on from the scheme written in the form (4.30) too. The Fourier analysis, in this case leads to the expression:

$$\xi_{\mathcal{RD}} = 0 \tag{4.32}$$

$$\omega_{\mathcal{RD}} = -\frac{ak \frac{\sin \mu_{\Delta x}}{\mu_{\Delta x}}}{\frac{(1 + \cos \mu_{\Delta x})}{2} + \frac{\alpha k^2}{\mu_{\Delta x}^2} \sin \mu_{\Delta x}} \tag{4.33}$$



**Figure 4.3:** Dispersion  $w$  against the wavenumber  $k$  of the central  $\mathcal{RD}$  scheme, compared to the scalar linear advection with dispersion model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes  $N = 5$ .

Making a comparison with (3.40) and (3.41), it can be seen that, exactly in the same way of other central spatially discretized schemes, the value of  $\xi$  is not perturbed by the presence of the dispersive term and that the main difference is the expression now assumed by  $\omega$ . Figure 4.3 provides a comparison between the dissipative property of the several developed central schemes and those of the exact model, and it is shown that the central  $\mathcal{RD}$  scheme performs exactly as a second order scheme.

## 4.4 Stabilized Upwind methods

In this work, a type of streamline upwind Petrov-Galerkin method is used to stabilize the central schemes, since a robust discretization scheme is searched, whose stability could be independent from the time integration scheme used, and spurious oscillations which originate from discontinuous or near - discontinuous solutions are wanted to be avoided. The advection-dispersion model is still a non-dissipative model, so the local Peclet number, which represent a ratio between the advection and diffusion velocities on the elements space scale, is going to be infinite and the flux will be always advection dominated, thus creating the already explained stability problems in the central spatially discretized schemes.

In the previous scalar linear advection case, the stabilization technique was realized by means of the residual of the equation. Here the aim is to apply the same procedure to the advection with dispersion problem, thus, adding an artificial streamline residual based term to the central semi-discrete Galerkin and  $\mathcal{RD}$  schemes of Sections (4.2) and (4.3).

Recalling the general  $SU/PG$  formulation for the one-dimensional linear scalar advection problem (3.48), in the advection with dispersion discrete problem it becomes:

$$\int_{\Omega} \varphi_i \partial_t u_h dx + \int_{\Omega} \varphi_i a \partial_x u_h dx + \int_{\Omega} \partial_x \varphi_i \alpha \partial_{xt} u_h dx + \sum_K \left( \int_K \delta \varphi_i \Phi^K dx \right) = 0 \quad (4.34)$$

being the streamline upwind weighting operator defined as  $\delta \varphi_i = a \partial_x \varphi_i \tau$ , with the scale parameter set to  $\tau = \Delta x / (2|a|)$ , and being  $\Phi^K$  the residual of the advection-dispersion equation computed on the cell  $K$  and defined by (4.21).

The dispersion term  $\alpha \partial_{x^2 t} u_h$  can be computed by means of its weak formulation in the central Galerkin part of the scheme, but needs to be managed in a different way inside the upwind term, since an integration by part, which could transfer at least one derivative to the weighting function, is not possible due to the  $\mathcal{P}^1$ -type functions chosen. The problem is analogous to the one we dealt with in the previous section for the development of the central  $\mathcal{RD}$  scheme.

A new auxiliary variable was there introduced in order to decrease the order of the dispersive term. It was defined such that it assumed the reconstructed value of the second order spatial derivative of the problem variable computed using the second order finite difference formula. Here, we proceed in exactly the same way, augmenting the number of the variables of the problem which, thus, required to be coupled with another equation in order to be solved. Using the same definition of the auxiliary variable  $w$ , the system which results, written in its weak formulation, is:

$$\left\{ \begin{array}{l} \int_{\Omega} \varphi_i \partial_t u_h dx + \int_{\Omega} \varphi_i a \partial_x u_h dx + \int_{\Omega} \partial_x \varphi_i \alpha \partial_{xt} u_h dx + \\ \quad + \frac{a \Delta x}{2|a|} \int_{\Omega} \partial_x \varphi_i (\partial_t u_{h|K} + a \partial_x u_{h|K} - \alpha \partial_t w_{h|K}) dx = 0 \\ \int_{\Omega} \varphi_i w_h dx = - \int_{\Omega} \partial_x \varphi_i \partial_x u_h dx \end{array} \right. \quad (4.35)$$

Solving the integrals of this system, assuming a positive value of the advection speed ( $a > 0$ ), the following semi-discrete system can be found:

$$\left\{ \begin{array}{l} \frac{\Delta x}{6} \frac{du_{i-1}}{dt} + \frac{2\Delta x}{3} \frac{du_i}{dt} + \frac{\Delta x}{6} \frac{du_{i+1}}{dt} + \frac{a}{2} (u_{i+1} - u_{i-1}) + \\ \quad + \frac{\alpha}{\Delta x} \left( \frac{du_{i+1}}{dt} - 2 \frac{du_i}{dt} + \frac{du_{i-1}}{dt} \right) + \\ \quad + \frac{\Delta x}{4} \frac{du_{i-1}}{dt} - \frac{\Delta x}{4} \frac{du_{i+1}}{dt} - \frac{a}{2} (u_{i+1} - 2u_i + u_{i-1}) + \\ \quad + \frac{\Delta x}{4} \frac{dw_{i+1}}{dt} - \frac{\Delta x}{4} \frac{dw_{i-1}}{dt} = 0 \\ \frac{\Delta x}{4} w_{i-1} + \frac{\Delta x}{2} w_i + \frac{\Delta x}{4} w_{i+1} = \frac{1}{\Delta x} (u_{i+1} - 2u_i + u_{i-1}) \end{array} \right. \quad (4.36)$$

which, using the technique of lumping the mass matrix of the second equation of the system to compute the values of  $w_{i+1}$  and  $w_{i-1}$  and substituting them into the first equation, results in the more simple expression:

$$\begin{aligned} \frac{5\Delta x}{12} \frac{du_{i-1}}{dt} + \frac{2\Delta x}{3} \frac{du_i}{dt} - \frac{\Delta x}{12} \frac{du_{i+1}}{dt} + a(u_i - u_{i-1}) + \\ + \frac{\alpha}{\Delta x} \left( \frac{1}{4} \frac{du_{i+2}}{dt} - \frac{3}{2} \frac{du_{i+1}}{dt} + 2 \frac{du_i}{dt} - \frac{1}{2} \frac{du_{i-1}}{dt} - \frac{1}{4} \frac{du_{i-2}}{dt} \right) = 0 \end{aligned} \quad (4.37)$$

Using the central residual distribution method derived above to compute the centred part of the  $\mathcal{SU}/\mathcal{PG}$  scheme instead of Galerkin, an upwind residual distribution method will results, as it has already been shown for the scalar advection case. This scheme can be written in the simple form:

$$\sum_{K \in \Omega} \beta_i^K \Phi^K = 0 \quad (4.38)$$

assuming for the distribution coefficient of the scheme the expression:

$$\beta_i^K = \frac{1}{2} + \frac{\text{sign}(a)}{2} \quad (4.39)$$

which results by adding the upwind weight to the centred one.

To compute the exact spatial-discrete expression of this method, the same steps made in the development of the centred scheme must be accomplished. This implies that the use of the variable  $w$ , in order to reconstruct the second order derivative of the problem variable  $u$ , is still needed. The expression of the scheme is found to be:

$$\begin{aligned} \Phi^{i-\frac{1}{2}} = \frac{\Delta x}{2} \frac{du_i}{dt} + \frac{\Delta x}{2} \frac{du_{i-1}}{dt} + a(u_i - u_{i-1}) + \\ - \frac{\alpha \Delta x}{2} \left( \frac{dw_i}{dt} + \frac{dw_{i-1}}{dt} \right) = 0 \end{aligned} \quad (4.40)$$

Through a mass lumped reconstruction of the value of  $w_i$ , the previous (4.40) becomes finally:

$$\begin{aligned} \frac{\Delta x}{2} \frac{du_i}{dt} + \frac{\Delta x}{2} \frac{du_{i-1}}{dt} + a(u_i - u_{i-1}) + \\ - \frac{\alpha}{2\Delta x} \left( \frac{du_{i+1}}{dt} - \frac{du_i}{dt} - \frac{du_{i-1}}{dt} + \frac{du_{i-2}}{dt} \right) = 0 \end{aligned} \quad (4.41)$$

which denote a translation of the stencil of the scheme in the upwind direction with respect to the central residual distribution case.

**Truncation Error :** The Truncation error of these streamline upwind schemes, calculated using Taylor approximation of nodal values of the different degrees of freedom which appear in formulae (4.37) and (4.41), results for the  $SU/\mathcal{PG}$  scheme in the second order accuracy expression:

$$\begin{aligned} \partial_t u_i + a\partial_x u_i - \alpha\partial_{x^2 t} u_i = & \frac{\Delta x}{2} \partial_x (\partial_t u_i + a\partial_x u_i - \alpha\partial_{x^2 t} u_i) + \\ & - \frac{\Delta x^2}{6} \partial_{x^2} \left( \partial_t u_i + a\partial_x u_i - \frac{\alpha}{2} \partial_{x^2 t} u_i \right) + \mathcal{O}(\Delta x^3) \end{aligned} \quad (4.42)$$

For the upwind residual distribution scheme the truncation error is still of the second order, since the resulting modify expression of the schemes results:

$$\begin{aligned} \partial_t u_i + a\partial_x u_i - \alpha\partial_{x^2 t} u_i = & \frac{\Delta x}{2} \partial_x (\partial_t u_i + a\partial_x u_i - \alpha\partial_{x^2 t} u_i) + \\ & - \frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{2} \partial_t u_i + \frac{a}{3} \partial_x u_i - \frac{2\alpha}{3} \partial_{x^2 t} u_i \right) + \mathcal{O}(\Delta x^3) \end{aligned} \quad (4.43)$$

The previous (4.42) and (4.43) confirm the fact that this upwind methods, built up on the central schemes by adding a stabilization term to the equation, do not affect their initial accuracy.

**Dispersion Analysis :** The analysis of the dispersion parameters gives opposite results for the two methods. Figure 4.4 gives a representation of

the  $\omega_{SU/\mathcal{P}\mathcal{G}}$  and  $\omega_{U/\mathcal{R}\mathcal{D}}$  (whose formulae are listed below) together with the trends of all the other centred schemes presented to discretize the linear scalar advection-dispersion problem. In particular we have that  $SU/\mathcal{P}\mathcal{G}$  performs much better even than the high accurate central  $\mathcal{F}\mathcal{D}4$  scheme, while the  $U/\mathcal{R}\mathcal{D}$  gives a worse description of the dispersive properties of the models also compared to the second accuracy order central  $\mathcal{R}\mathcal{D}$  scheme. This last scheme appears, thus, not to be particularly indicated for the solution of the advection-dispersion problem.

The analysis of the dissipation parameter (Figure 4.5) shows instead that a non-physical dissipation has been added to the  $SU/\mathcal{P}\mathcal{G}$  model. Effect which seems to be totally absent in the upwind residual distribution scheme which, indeed, is able to give stability to its central scheme maintaining an optimal description of the dissipation parameter of the physical model.

The expressions of the dissipation and dispersive parameters, computed for the two upwind models developed in this section, are reported below:

$$\xi_{SU/\mathcal{P}\mathcal{G}} = \frac{N_{SU/\mathcal{P}\mathcal{G}}^\xi}{D_{SU/\mathcal{P}\mathcal{G}}^\xi} \quad (4.44)$$

$$N_{SU/\mathcal{P}\mathcal{G}}^\xi = ak \frac{\sin^2 \mu_{\Delta x}}{2\mu_{\Delta x}} \left( 1 + 2 \frac{\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x}) \right) + \quad (4.45)$$

$$-\frac{ak}{\mu_{\Delta x}} (1 - \cos \mu_{\Delta x}) \left( \frac{(2 + \cos \mu_{\Delta x})}{3} + 2 \frac{\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x}) \right) \quad (4.46)$$

$$D_{SU/\mathcal{P}\mathcal{G}}^\xi = \left( \frac{2 + \cos \mu_{\Delta x}}{3} + 2 \frac{\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x}) \right)^2 + \quad (4.47)$$

$$+ \frac{\sin^2 \mu_{\Delta x}}{4} \left( 1 + 2 \frac{\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x}) \right)^2 \quad (4.48)$$

$$\omega_{SU/\mathcal{P}\mathcal{G}} = \frac{-ak \frac{\sin \mu_{\Delta x}}{\mu_{\Delta x}} + \frac{\sin \mu_{\Delta x}}{2} \left( 1 + 2 \frac{\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x}) \right) \xi_{SU/\mathcal{P}\mathcal{G}}}{\frac{(2 + \cos \mu_{\Delta x})}{3} + 2 \frac{\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x})} \quad (4.49)$$

The same analysis, carried out for the upwind residual distribution scheme (4.41), gives similar results:

$$\xi_{U/RD} = \frac{N_{U/RD}^{\xi}}{D_{U/RD}^{\xi}} \quad (4.50)$$

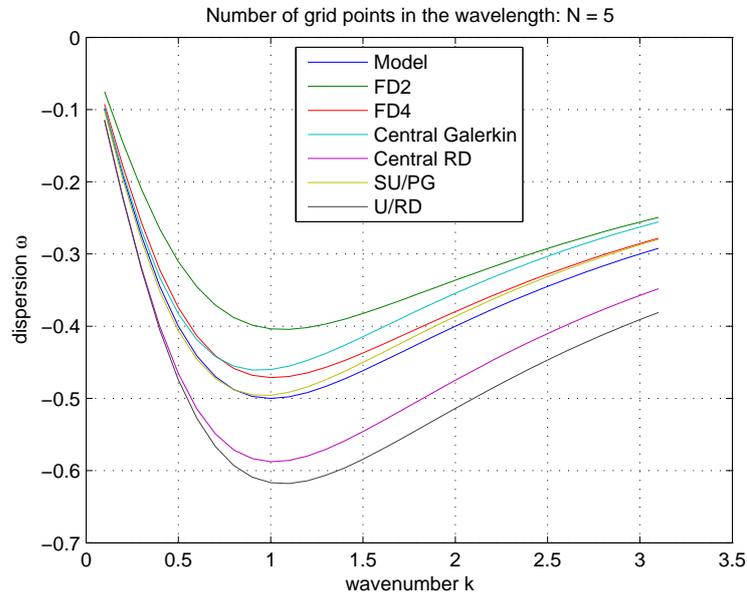
$$N_{U/RD}^{\xi} = ak \frac{\sin^2 \mu_{\Delta x}}{\mu_{\Delta x}} \left( \frac{1}{2} + \frac{\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x}) \right) + \quad (4.51)$$

$$-\frac{ak}{\mu_{\Delta x}} (1 - \cos \mu_{\Delta x}) \left( \frac{(1 + \cos \mu_{\Delta x})}{2} + \frac{\alpha k^2}{\mu_{\Delta x}^2} \sin^2 \mu_{\Delta x} \right) \quad (4.52)$$

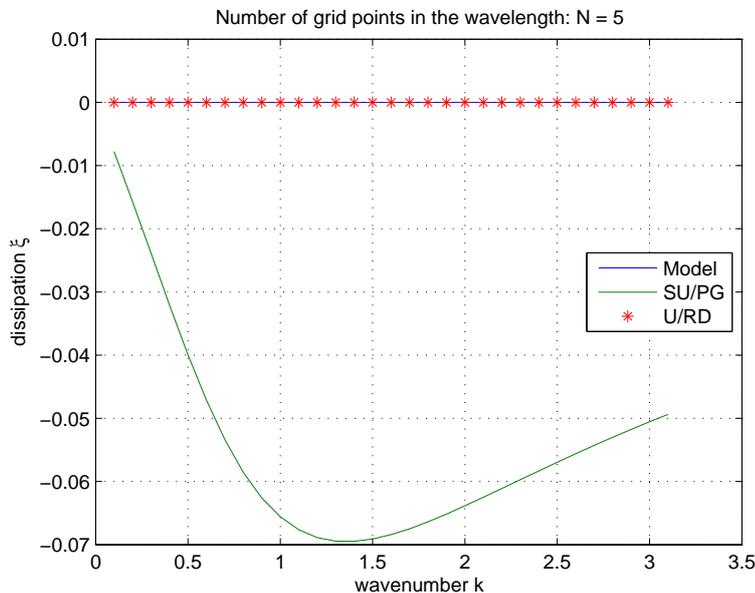
$$D_{U/RD}^{\xi} = \left( \frac{1 + \cos \mu_{\Delta x}}{2} + \frac{\alpha k^2}{\mu_{\Delta x}^2} \sin^2 \mu_{\Delta x} \right)^2 + \quad (4.53)$$

$$+ \sin^2 \mu_{\Delta x} \left( \frac{1}{2} + \frac{\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x}) \right)^2 \quad (4.54)$$

$$\omega_{U/RD} = \frac{-ak \frac{\sin \mu_{\Delta x}}{\mu_{\Delta x}} + \frac{\sin \mu_{\Delta x}}{2} \left( 1 + 2 \frac{\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x}) \right) \xi_{U/RD}}{\frac{(2 + \cos \mu_{\Delta x})}{3} + 2 \frac{\alpha k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x})} \quad (4.55)$$



**Figure 4.4:** Dispersion  $w$  against the wavenumber  $k$  of the stabilized upwind  $SU/PG$  and  $U/RD$  schemes, compared to the scalar linear advection with dispersion model and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes  $N = 5$ .



**Figure 4.5:** Diffusion  $\xi$  against the wavenumber  $k$  of the stabilized upwind  $SU/PG$  and  $U/RD$  schemes, compared to the scalar linear advection with dispersion model, when the wavelength is discretized with a number of grid nodes  $N = 5$ .

## 4.5 Conclusions

Linear scalar advection with dispersion equation was here discretized in space applying all the schemes previously developed in the simple scalar advection case. In this procedure, some complications arised from the second order spatial derivative of the dispersive term in the development of the central  $RD$  method and of the stabilized schemes. In fact, due to the  $\mathcal{P}^1$  basis function chosen for the approximation of the solution over the grid cells, the introduction of a new auxiliary variable was needed to reformulate the equation in a lower order form before the integration. In order to accomplish this, the auxiliary variable introduced had to assume the reconstructed value of the first or second spatial derivative of the main variable and transformed the problem into a system of partial differential equations augmenting the number of the unknowns. It has been demonstrated in Appendix A that the choice related to the derivative to be reconstructed has no influence on the final schemes expressions.

All the schemes developed were found to suffer from a second order trun-

cation error and had to deal with the inversion of, at least, the stiffness matrix of the semi-discrete system which arises from the dispersion term. This fact increases, in particular, the computational cost of the finite difference schemes which, in the previous case, should not compute any matrix inversion. Hence, in this context the cost of the *SU/PG* scheme appears to be not so higher with respect to the finite difference one, having also the advantage to be more robust.

The analysis of the dispersive parameters, instead, showed a very good agreement with the original model for the Galerkin and streamline upwind Petrov-Galerkin schemes. In fact, the former performed as a fourth order accurate scheme in the wavenumbers range till  $k \approx 0.8$  beyond which this accuracy decreased even though remaining better with respect to the *FD2* scheme. The latter is able to attain the best results at all (see Figure 4.4) even revealing a non-physical non-null trend of the dissipative parameter  $\omega$ , which, instead, did not affect the upwind residual distribution scheme. This last scheme, however, appears to be inadequate for this kind of problem solution since its dispersion error results quite big with respect to those of the other schemes proposed.



# Chapter 5

## Madsen and Sørensen system discretisation

The extended one-dimensional Boussinesq Madsen and Sørensen model has already been presented in Section 1.2, but for convenience it is here rewritten:

$$\begin{cases} \partial_t \eta + \partial_x q = 0 \\ \partial_t q - Bh^2 \partial_{x^2 t} q - \frac{1}{3} h \partial_x h \partial_{xt} q + \partial_x(uq) + gH \partial_x \eta + \\ \qquad \qquad \qquad - \beta gh^3 \partial_{x^3} \eta - 2\beta gh^2 \partial_x h \partial_{x^2} \eta = 0 \end{cases} \quad (5.1)$$

A description of the variables and parameters, which compare in system (5.1), has already been given and a dispersion analysis of this model has also been accomplished. In particular, the dissipation and dispersion parameters of the model appeared to be:

$$\begin{aligned} \xi_{MS} &= 0 \\ \omega_{MS}^2 &= k^2 c^2 \frac{1 + \beta \mu^2}{1 + B \mu^2} \end{aligned} \quad (5.2)$$

In the following sections we are going to extend the space-discrete schemes developed for the two analysed problems — the semi-discretized in space one-dimensional scalar linear advection and the advection with dispersion — in

order to apply them to the more complex MS one. After the development of all the different methods and the discussion of their properties, the resulting system of ordinary differential equations in time is going to be integrated by means of a time integration scheme.

## 5.1 Finite difference methods

Finite difference methods can be easily developed even to discretize the complex Madsen-Sørensen system of equations simply by directly replacing the spatial derivatives present in (5.1) with the discrete expressions illustrated in Sections 3.1 and 4.1. However, a third order spatial derivative term is included in the MS model. Depending on the accuracy order desired for the discretization, the second and fourth order accuracy formulae for its approximation are:

$$\frac{d^3 f}{dx^3}(x_i) = \frac{1}{2\Delta x^3} (-f_{i-2} + 2f_{i-1} - 2f_{i+1} + f_{i+2}) + \mathcal{O}(\Delta x^2) \quad (5.3)$$

$$\begin{aligned} \frac{d^3 f}{dx^3}(x_i) = \frac{1}{8\Delta x^3} (f_{i-3} - 8f_{i-2} + 13f_{i-1} - 13f_{i+1} + 8f_{i+2} - f_{i+3}) + \\ + \mathcal{O}(\Delta x^4) \end{aligned} \quad (5.4)$$

These formulae lead to some new centred schemes and this is clear by looking at the symmetry of the coefficients which characterize the expressions (5.3) and (5.4). Their application in the discretization of the term  $\partial_{x^3}\eta$  is going also to enlarge the stencil of the schemes which result. In particular, applying the second order approximations (3.7), (4.5) and (5.3), the following semi-discrete scheme can be found:

$$\begin{aligned}
\frac{d\eta_i}{dt} + \frac{1}{2\Delta x} (q_{i+1} - q_{i-1}) &= 0 \tag{5.5} \\
\frac{dq_i}{dt} - \frac{Bh_i^2}{\Delta x^2} \left( \frac{dq_{i-1}}{dt} - 2\frac{dq_i}{dt} + \frac{dq_{i+1}}{dt} \right) + \\
&- \frac{h_i}{12\Delta x^2} (h_{i+1} - h_{i-1}) \left( \frac{dq_{i+1}}{dt} - \frac{dq_{i-1}}{dt} \right) + \\
&+ \frac{1}{2\Delta x} (u_{i+1}q_{i+1} - u_{i-1}q_{i-1}) + \frac{gH_i}{2\Delta x} (\eta_{i+1} - \eta_{i-1}) + \\
&- \beta \frac{gh_i^3}{2\Delta x^3} (-\eta_{i-2} + 2\eta_{i-1} - 2\eta_{i+1} + \eta_{i+2}) + \\
&- \beta \frac{gh_i^2}{\Delta x^3} (h_{i+1} - h_{i-1}) (\eta_{i+1} - 2\eta_i + \eta_{i-1}) = 0
\end{aligned}$$

The use of (3.9), (4.6) and (5.4) will, instead, lead to the higher accurate scheme:

$$\begin{aligned}
\frac{d\eta_i}{dt} + \frac{1}{12\Delta x} (q_{i-2} - 8q_{i-1} + 8q_{i+1} - q_{i+2}) &= 0 \tag{5.6} \\
\frac{dq_i}{dt} - \frac{Bh_i^2}{12\Delta x^2} \left( -\frac{dq_{i-2}}{dt} + 16\frac{dq_{i-1}}{dt} - 30\frac{dq_i}{dt} + 16\frac{dq_{i+1}}{dt} - \frac{dq_{i+2}}{dt} \right) + \\
&- \frac{h_i}{432\Delta x^2} (h_{i-2} - 8h_{i-1} + 8h_{i+1} - h_{i+2}) \\
&\quad \left( \frac{dq_{i-2}}{dt} - 8\frac{dq_{i-1}}{dt} + 8\frac{dq_{i+1}}{dt} - \frac{dq_{i+2}}{dt} \right) + \\
&+ \frac{1}{12\Delta x} (u_{i-2}q_{i-2} - 8u_{i-1}q_{i-1} + 8u_{i+1}q_{i+1} - u_{i+2}q_{i+2}) + \\
&+ \frac{gH_i}{12\Delta x} (\eta_{i-2} - 8\eta_{i-1} + 8\eta_{i+1} - \eta_{i+2}) + \\
&- \beta \frac{gh_i^3}{8\Delta x^3} (\eta_{i-3} - 8\eta_{i-2} + 13\eta_{i-1} - 13\eta_{i+1} + 8\eta_{i+2} - \eta_{i+3}) + \\
&- \beta \frac{gh_i^2}{72\Delta x^3} (h_{i-2} - 8h_{i-1} + 8h_{i+1} - h_{i+2}) \\
&\quad (-\eta_{i-2} + 16\eta_{i-1} - 30\eta_i + 16\eta_{i+1} - \eta_{i+2}) = 0
\end{aligned}$$

Albeit very simple to be derived, this class of schemes is not here applied to any test case since they present difficulties in modelling irregular geometries

in two space dimensions, due to the fact that representation with structural grids reveals a lack of accuracy and local mesh adaptation techniques cannot be developed for such schemes. Moreover they do not present any computational advantage since, as in the advection-dispersion case, a matrix inversion has to be computed in any case.

In the following paragraphs the properties of these two schemes will be investigated on the simpler linearised Madsen and Sørensen system which differs from the complete model for the absence of the nonlinear terms and for the assumption of a constant bathymetry value  $h = h_0 = \text{const.}$  The schemes, which results from the application to the linear problem of the second and fourth order finite different approximations of the spatial derivatives, are listed in Appendix C.

**Truncation Error :** In order to compute the truncation error analysis of the finite difference schemes (C.3) and (C.4), listed in Appendix C, which are applied to the linearised Madsen and Sørensen problem (1.20), the standard Taylor series expansions in space (3.12) have to be introduced into the schemes. Rearranging the terms, the following modified expression of the (C.3) system appears:

$$\partial_t \eta_i + h_0 \partial_x u_i = -\frac{h_0 \Delta x^2}{6} \partial_{x^3} u_i - \frac{h_0 \Delta x^4}{120} \partial_{x^5} u_i + \mathcal{O}(\Delta x^6) \quad (5.7)$$

$$\begin{aligned} \partial_t u_i - B h_0^2 \partial_{x^2 t} u_i + g \partial_x \eta_i - \beta g h_0^2 \partial_{x^3} \eta_i = \\ -\frac{\Delta x^2}{6} \partial_{x^2} \left( -\frac{B h_0^2}{2} \partial_{x^2 t} u_i + g \partial_x \eta_i - \frac{7}{5} \beta_i g h_0^2 \partial_{x^3} \eta_i \right) + \\ -\frac{\Delta x^4}{40} \partial_{x^4} \left( -\frac{B h_0^2}{9} \partial_{x^2 t} u_i + \frac{g}{3} \partial_x \eta_i - \beta_i g h_0^2 \partial_{x^3} \eta_i \right) + \\ + \mathcal{O}(\Delta x^6) \end{aligned} \quad (5.8)$$

while for the system (C.4), the expression becomes:

$$\partial_t \eta_i + h_0 \partial_x u_i = \frac{h_0 \Delta x^4}{30} \partial_{x^5} u_i + \mathcal{O}(\Delta x^6) \quad (5.9)$$

$$\begin{aligned} \partial_t u_i - B h_0^2 \partial_{x^2 t} u_i + g \partial_x \eta_i - \beta g h_0^2 \partial_{x^3} \eta_i = \\ = \Delta x^4 \partial_{x^4} \left( -\frac{2}{135} B h_0^2 \partial_{x^2 t} u_i + \frac{g}{30} \partial_x \eta_i - \frac{653}{10080} \beta_i g h_0^2 \partial_{x^3} \eta_i \right) + \\ + \mathcal{O}(\Delta x^6) \end{aligned} \quad (5.10)$$

The substitution of the solution of the original equations of the system  $\partial_t \eta + h_0 \partial_x u = 0$  and  $\partial_t u - B h_0^2 \partial_{x^2} u + g \partial_x \eta - \beta g h_0^2 \partial_{x^3} \eta = 0$  results in the truncation error expression for both the schemes. Due to the fact that all the spatial derivatives are discretized by means of second order approximation formulae in one case and by fourth order accuracy expressions in the other, not being there any mass matrix which multiplies time derivative terms, we are not surprised to discover that the truncation error of the two methods are coherently respectively  $\mathcal{O}(\Delta x^2)$  and  $\mathcal{O}(\Delta x^4)$ . Compared to the scalar linear advection with dispersion equation, the stencil of both the schemes results augmented because of the exigency to approximate a higher order spatial derivative. From the truncation error analysis it becomes still clear, in the same way as for the two other simplified cases, that under the payment of the cost of an enlargement of the stencil, this  $\mathcal{FD}$  schemes could achieve higher order of accuracy.

**Dispersion Analysis :** In the Section 1.3.2 the dispersion analysis of the linearised Madsen and Sørensen system has been investigated and it was showed that this extended Boussinesq model gives zero damping and a dispersion with the expression recalled by (5.2). Taking into account finite difference spatial discretisation (C.3) and (C.4), the Fourier analysis on this methods gives the results:

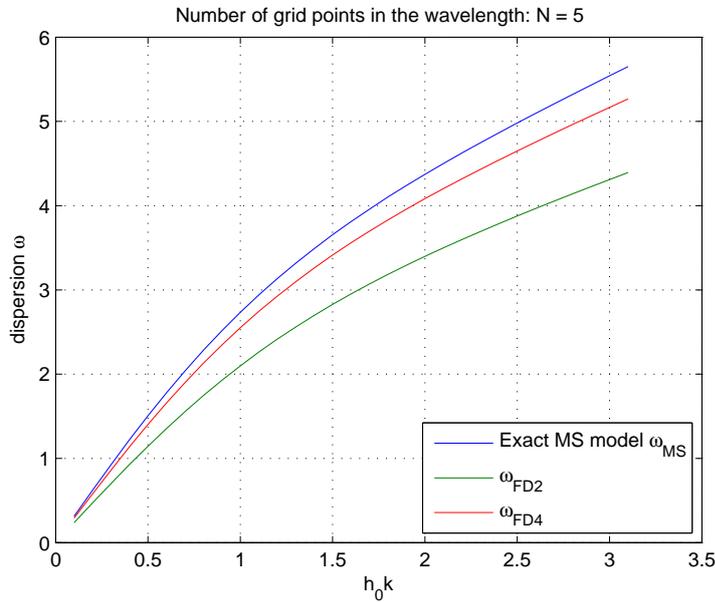
$$\begin{aligned} \xi_{\mathcal{FD}2} &= 0 & \xi_{\mathcal{FD}4} &= 0 \\ \omega_{\mathcal{FD}2} &= \frac{N_{\mathcal{FD}2}}{D_{\mathcal{FD}2}} & \omega_{\mathcal{FD}4} &= \frac{N_{\mathcal{FD}4}}{D_{\mathcal{FD}4}} \end{aligned} \quad (5.11)$$

where the expression of the quantities  $N_{\mathcal{FD}2}$ ,  $D_{\mathcal{FD}2}$ ,  $N_{\mathcal{FD}4}$  and  $D_{\mathcal{FD}4}$  are listed below:

$$\begin{aligned} N_{\mathcal{FD}2} &= c^2 k^2 \frac{\sin^2 \mu_{\Delta x}}{\mu_{\Delta x}^2} + 2\beta g h_0^3 k^4 \frac{\sin^2 \mu_{\Delta x}}{\mu_{\Delta x}^4} (1 - \cos \mu_{\Delta x}) \\ [5pt] D_{\mathcal{FD}2} &= 1 + 2B \frac{h_0^2 k^2}{\mu_{\Delta x}^2} (1 - \cos \mu_{\Delta x}) \end{aligned} \quad (5.12)$$

$$\begin{aligned}
 N_{\mathcal{FD4}} &= \frac{c^2 k^2}{36 \mu_{\Delta x}} (8 \sin \mu_{\Delta x} - \sin 2\mu_{\Delta x})^2 + \\
 &\quad + \beta \frac{h_0^2 k^2}{\mu_{\Delta x}^2} (\sin 3\mu_{\Delta x} - 8 \sin 2\mu_{\Delta x} + 13 \sin \mu_{\Delta x}) \quad (5.13) \\
 D_{\mathcal{FD4}} &= 1 + B \frac{h_0^2 k^2}{6 \mu_{\Delta x}^2} (15 - 16 \cos \mu_{\Delta x} + \cos 2\mu_{\Delta x})
 \end{aligned}$$

Therefore, expressions contained in (5.11) ensure about the well coherent representation of the non dissipative character of the model by the two finite difference schemes. The description of the dispersive properties is given in Figure 5.1, where the variation of the quantities  $\omega_{\mathcal{FD2}}$ ,  $\omega_{\mathcal{FD4}}$  and  $\omega_{MS}$  (the exact model dispersion) is shown as function of the product between the wavenumber  $k$  and the constant bathymetry depth  $h_0$ . The Figure shows clearly that the fourth order accurate scheme gives also the more accurate dispersion description, even if it requires a higher computational cost due to its less compact stencil.



**Figure 5.1:** Dispersion  $w$  against the wavenumber  $k$  of the finite difference  $\mathcal{FD2}$  and  $\mathcal{FD4}$  schemes, compared to Airy theory value (1.25), when the wavelength is discretized with a number of grid nodes  $N = 5$ .

## 5.2 Galerkin finite element method

In the present section, the Galerkin finite element spatial discretization method is going to be developed in the full extended Boussinesq Madsen and Sørensen context. We are going to base our discussion on what already presented and explained for the scalar linear advection and linear advection with dispersion cases. Thus, recalling first the one-dimensional MS system in its strong form, given in (5.1), the weak formulation of the problem must be take into account, which states:

$$\int_{\Omega} \varphi_i \partial_t \eta dx - \int_{\Omega} q \partial_x \varphi_i dx = 0 \quad (5.14)$$

$$\begin{aligned} \int_{\Omega} \varphi_i \partial_t q dx + B \int_{\Omega} \partial_x (\varphi_i h^2) \partial_{xt} q dx + \frac{1}{3} \int_{\Omega} \partial_x h \partial_x (\varphi_i h) \partial_t q dx + \\ - \int_{\Omega} u q \partial_x \varphi_i dx - g \int_{\Omega} \eta \partial_x (\varphi_i H) dx + \\ + \beta g \int_{\Omega} \partial_x (\varphi_i h^3) \partial_{x^2} \eta dx + \\ + 2\beta g \int_{\Omega} \partial_x h \partial_x (\varphi_i h^2) \partial_x \eta dx = 0 \end{aligned} \quad (5.15)$$

which does not contain any boundary term, which would be originated by the integration by parts, if, for example, a set of periodic boundary conditions had been imposed to the spatial domain of the problem. This means the suppression of one degree of freedom since:

$$\begin{aligned} \eta(X_1, t) = \eta(X_2, t) &\Rightarrow \partial_x \eta(X_1, t) = \partial_x \eta(X_2, t) \\ q(X_1, t) = q(X_2, t) &\Rightarrow \partial_x q(X_1, t) = \partial_x q(X_2, t) \end{aligned} \quad (5.16)$$

A first look given to the weak form (5.15) immediately reveals that problems will emerge during the discretization procedure, if the set of linear basis functions  $\varphi_j$ , chosen to interpolate the nodal values of the variables functions, is taken from the  $\mathcal{P}^1$  functional space, like it has been done up to now. In fact, in this case, there would be no way to solve the third order spatial derivative of the variable  $\eta$  and it would be required to the regularity of the basis functions  $\varphi_j$  to be high enough for them to belong to a  $\mathcal{H}^2$  Sobolev space over the spatial domain  $\Omega$ , where  $v \in \mathcal{H}^k(\Omega)$  if  $v \in L^2(\Omega)$  and

$(d^i v/dx^i, i = 1, \dots, k) \in L^2(\Omega)$ . Therefore, methods involving  $\mathcal{H}^2$  quadratic and  $\mathcal{H}^3$  cubic spline functions can be used in this case to solve the problem. The most compact  $\mathcal{H}^2$  function is the Hermite cubic [14], where the function and its first derivative are continuously interpolated over the mesh. However the aim here is to design a scheme that can be in the future generalized to a two-dimensional finite element method based on unstructured triangular grids. Although some work has been done on extending Hermite-type interpolation onto unstructured grids [22], the additional complexity makes them not feasible for calculation with large meshes.

An alternative to use a higher degree of continuity in the finite element basis functions is to rewrite the equation in a lower order form, e.g. by introducing a new variable. The strategy has the same nature to the one that was presented in Section 4.3 when a way to discretize the dispersion term in a residual distribution context was searched. The same solution adopted there is here reproposed, hence, to introduce an auxiliary variable to the problem whereby to decrease the derivation order of this term and allow the use of linear  $\mathcal{P}^1$  weighted and basis functions. Proceeding in this way, the new variable we are going to introduce could assume, in a free way, the approximated value of the first or the second spatial derivative of  $\eta$ . In Appendix B we have demonstrate that the two different choices lead, at the end, exactly to the same discrete expression of the dispersion term. The same equality can be demonstrate also in this case (see Appendix D), under the hypothesis of a constant bathymetry. Otherwise, it can be demonstrated that a small difference arises due to the quadrature formula used to solve the definite integrals, but it appears not to change the solution in a sensible way. Following [42], we thus introduce the definition of the auxiliary variable  $w_\eta$  as:

$$w_\eta = \partial_{xx}\eta \tag{5.17}$$

This variable extends the number of the unknowns of the system, but the continuity restrictions on the test space have been dropped. The situation requires the addition of a new algebraic equation to the system in order to find the solution. Definition (5.4) can be used for this scope, and the increased system in the weak formulation appears as:

$$\int_{\Omega} \varphi_i \partial_t \eta dx - \int_{\Omega} q \partial_x \varphi_i dx = 0 \tag{5.18}$$

$$\begin{aligned}
& \int_{\Omega} \varphi_i \partial_t q dx + B \int_{\Omega} \partial_x (\varphi_i h^2) \partial_{xt} q dx + \frac{1}{3} \int_{\Omega} \partial_x h \partial_x (\varphi_i h) \partial_t q dx + \\
& - \int_{\Omega} u q \partial_x \varphi_i dx - g \int_{\Omega} \eta \partial_x (\varphi_i H) dx + \\
& + \beta g \int_{\Omega} w \partial_x (\varphi_i h^3) dx + \\
& + 2\beta g \int_{\Omega} \partial_x h \partial_x (\varphi_i h^2) \partial_x \eta dx = 0
\end{aligned} \tag{5.19}$$

$$\int_{\Omega} \varphi_i w dx + \int_{\Omega} \partial_x \varphi_i \partial_x \eta dx = 0 \tag{5.20}$$

Now, since the unknowns  $\eta(x, t)$  and  $u(x, t)$  are spatially approximated by means of the set of  $\mathcal{P}^1$  basis functions  $\varphi_j$ , we need to use the weak formulation to transfer one spatial derivative to the test functions only for terms characterized by a second spatial derivative, like the dispersive ones  $-Bh^2 \partial_{x^2 t} q$  and  $-2\beta g h^2 \partial_x h \partial_{x^2} \eta$ .

The spatial discretized approximation of the unknowns in all the domain is produced by an interpolation over the set of basis functions as in (3.21):

$$\eta(x, t) \approx \sum_{j=1}^{N+1} \psi_j(x) \eta_j(t) \quad q(x, t) \approx \sum_{j=1}^{N+1} \psi_j(x) q_j(t) \tag{5.21}$$

The Galerkin finite element discretization can now be computed by splitting the integrals over the entire spatial domain into a sum of integrals over the single cells and accomplishing a final assembly procedure, computing the solution over each element, like in (3.20) using the only two basis functions local to it. The same discretization procedure must be applied to equation (5.20) too. This generates an auxiliary algebraic system whose mass matrix is approximated in diagonal form through a lumping technique trying to reduce the extra-work required to solve it, but with the price of decreasing the accuracy of the approximation. However it is going to be shown, in the following, that this fact does not affect the spatial truncation error of the global scheme. The central semi-discrete scheme resulted is shown in its extended form in Appendix E, but it can be simplified in one of the kind:

$$A_{ij} \dot{y}_i - f_i = 0 \tag{5.22}$$

This is a system of ordinary differential equations in time, where  $y(t)$  contains all the system unknowns, matrix  $A$  results in a constant coefficient matrix, containing the general Galerkin mass matrix shown for the scalar linear advection and linear advection with dispersion cases, and  $f(y, t)$  is a nonlinear function of  $y$ . The localised nature of the one-dimensional finite element spatial approximation proposed will generate a matrix  $A$  which is also banded and which can be solved much more efficiently than a full matrix.

**Truncation Error :** Like in Section 5.1, the truncation error of the scheme is computed on the linearized MS system (1.20). Assuming a smooth classical solution which allows to use Taylor series expansions in space for nodal values of the variables in the several points of the stencil, and then differentiating equation (5.17) and using it to replace the  $w_{\eta,i}$  terms in equation (C.4), we obtain the following modified form of the scheme:

$$\begin{aligned} \partial_t \eta_i + h_0 \partial_x u_i = & -\frac{\Delta x^2}{6} \partial_{x^2} (\partial_t u_i + h_0 \partial_x u_i) + \\ & -\frac{\Delta x^4}{24} \partial_{x^4} \left( \frac{1}{3} \partial_t u_i + \frac{h_0}{5} \partial_x u_i \right) + \mathcal{O}(\Delta x^6) \end{aligned} \quad (5.23)$$

$$\begin{aligned} \partial_t u_i - Bh_0^2 \partial_{x^2 t} u_i + g \partial_x \eta_i - \beta g h_0^2 \partial_{x^3} \eta_i = & \\ & -\frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{3} \partial_t u_i - \frac{Bh_0^2}{6} \partial_{x^2 t} u_i + \frac{1}{3} g \partial_x \eta_i - \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\ & -\frac{\Delta x^4}{8} \partial_{x^4} \left( \frac{1}{9} \partial_t u_i - \frac{Bh_0^2}{45} \partial_{x^2 t} u_i + \frac{g}{15} \partial_x \eta_i - \frac{1}{5} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\ & + \mathcal{O}(\Delta x^6) \end{aligned} \quad (5.24)$$

Equation (5.23) and (5.24) results very close to the previous (5.7) and (5.8). In fact the only difference stands in the presence of the mass matrix and, as already underlined in the previous sections, the two schemes results the same if a mass lumping technique is applied to the first one. Equations (5.23) and (5.24) show a truncation error which is in the order of  $\mathcal{O}(\Delta x^4)$  regarding the first equation of the system, while in the order of  $\mathcal{O}(\Delta x^2)$  for the former one, thus, confirming the statement that the Galerkin mass matrix increases the accuracy of a second order finite difference scheme.

**Dispersion Analysis :** We consider now the estimation of the Fourier properties of the semi-discrete system (C.4). The objective is to recall a correspondence between the dumping and the phase speeds obtained by directly replacing a Fourier mode into the differential equation with those obtained by analyzing the modified equation of the scheme. Considering a Fourier mode of the generic variable  $u(x, t) = u_0 e^{\nu t + j k x}$ , we can write:

$$u_{i+m} = e^{\nu t} e^{j k m \Delta x} = e^{\nu t} e^{j m \mu_{\Delta x}} \quad (5.25)$$

where it is setted:  $\mu_{\Delta x} = k \Delta x$ . The time derivative can be written as:

$$\partial_t u_{i+m} = \nu e^{\nu t + j m \mu_{\Delta x}} = \nu u e^{j m \mu_{\Delta x}} \quad (5.26)$$

which allows us to obtain:

$$\left( \frac{e^{-j \mu_{\Delta x}}}{6} + \frac{2}{3} + \frac{e^{j \mu_{\Delta x}}}{6} \right) \nu \eta + \frac{h_0}{2 \Delta x} (e^{j \mu_{\Delta x}} - e^{-j \mu_{\Delta x}}) u = 0 \quad (5.27)$$

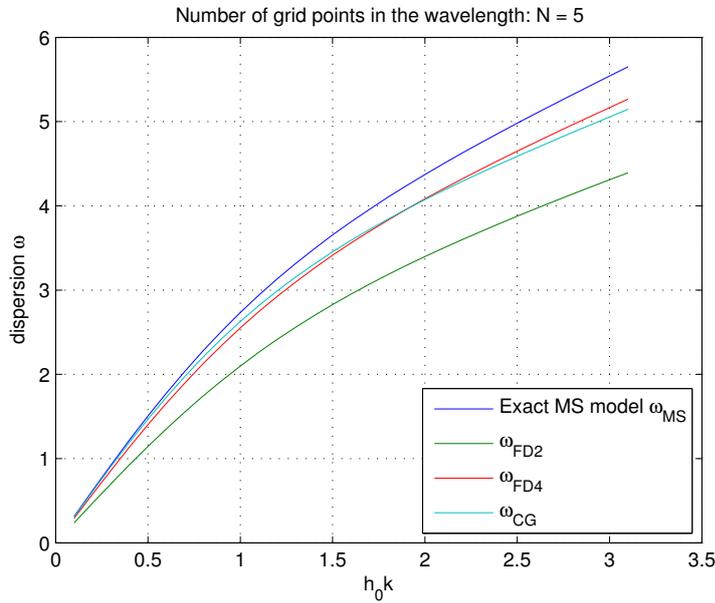
$$\begin{aligned} & \left( \frac{e^{-j \mu_{\Delta x}}}{6} + \frac{2}{3} + \frac{e^{j \mu_{\Delta x}}}{6} \right) \nu u - \frac{B h_0^2}{\Delta x^2} (e^{j \mu_{\Delta x}} - 2 + e^{-j \mu_{\Delta x}}) \nu u + \\ & + \frac{g}{2 \Delta x} (e^{j \mu_{\Delta x}} - e^{-j \mu_{\Delta x}}) \eta + \\ & - \frac{\beta g h_0^2}{2 \Delta x^3} (e^{j 2 \mu_{\Delta x}} - 2 e^{j \mu_{\Delta x}} + 2 e^{-j \mu_{\Delta x}} - e^{-j 2 \mu_{\Delta x}}) \eta = 0 \end{aligned} \quad (5.28)$$

This system can be written in matrix form introducing the vector of the variables  $[\eta, u]$ . After this, recalling the definition  $\nu = \xi + j \omega$ , the analytical determination of the roots corresponding to this eigenvalue problem, after many and long manipulations, leads to this final expression of the damping  $\xi$  and phase  $\omega$  which appear to be:

$$\xi_{CG} = 0 \quad (5.29)$$

$$\omega_{CG} = c^2 k^2 \frac{\sin^2(\mu_{\Delta x})}{\mu^2} \frac{1 + 2\beta\mu^2 \frac{1 - \cos(\mu_{\Delta x})}{\mu_{\Delta x}^2}}{\frac{2 + \cos(\mu_{\Delta x})}{3} \left( \frac{2 + \cos(\mu_{\Delta x})}{3} + 2B\mu^2 \frac{1 - \cos(\mu_{\Delta x})}{\mu_{\Delta x}^2} \right)} \quad (5.30)$$

The null value of the damping variable  $\xi$  means that the spatial discretized model is non-dissipative, in perfect analogy with the Madsen-Sørensen system's Fourier properties studied in the Section 1.3. The comparison between the  $\omega_{CG}$  function of the numerical scheme, just obtained, with the one of the original model, presented in Figure 5.2, shows a good agreement between the dispersion characteristic of the two systems. The trend of  $\omega_{CG}$  appears to give a description of the dispersive properties which is almost the same of a fourth order accuracy scheme, being, in contrast, much better than the  $\mathcal{FD}2$  schemes despite the two methods adopt the same approximation of the spatial derivatives.



**Figure 5.2:** Dispersion  $w$  against the wavenumber  $k$  of the Galerkin finite element scheme, compared to Airy theory value and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes  $N = 5$ .

### 5.3 Residual Distribution

Recalling what already said about the residual discretization technique in the previous chapters, this method is based on the evaluation of the residual over each cell  $K$  of the grid. The residual values are then partitioned between the nodes which belong to the same cell, in a way which is dependent on the distribution coefficients  $\beta_i^K$  and such that, for consistency:

$$\sum_{K \in K_i} \Phi_i^K = \sum_{K \in K_i} \beta_i^K \Phi^K = 0 \quad (5.31)$$

Since in this section a centred residual scheme is searched, then the distribution coefficients assume a value  $\beta_i^K = 1/2$ , because two is the number of the degrees of freedom which belong to the cell  $K$ , in order to attribute to each of them the same fraction of the entire value of the residual computed. It is very clear that in a one-dimensional problem, excluding the edges of the space domain, each degree of freedom of the mesh  $i$  belongs to the two cells adjacent to it and, thus, receives a residual flux from each of them. The semi-discrete form of the scheme, written for the node  $i$  is, hence, always the same:

$$\frac{1}{2}\Phi^{i-\frac{1}{2}} + \frac{1}{2}\Phi^{i+\frac{1}{2}} = 0 \quad (5.32)$$

The aspect of the first equation, which composes the MS system, is very similar to the scalar advection equation discretised in Chapter 3. There are no difficulties in computing the weighted residual fluxes which upgrade the solution in each node  $i$ . The residual flux computed over the cell  $K$  takes the form:

$$\Phi_\eta^K = \int_K (\partial_t \eta_h + \partial_x q_h) dx \quad (5.33)$$

Taking into account the element having nodes  $i$  and  $i-1$  as extremes, the integration with the middle point rule, following the procedure illustrated in Section 3.3 leads to the semi-discrete residual flux expression:

$$\Phi_\eta^{i-\frac{1}{2}} = \Delta x \frac{d\eta_{i-\frac{1}{2}}}{dt} + (q_i - q_{i-1}) \quad (5.34)$$

Difficulties arise in the treatment of the second equation belonging to the MS system, due to the dispersive and high space derivatives terms here present. In order to give a discrete expression to the residuals  $\Phi_q^K$ , the same procedure developed in Section 4.3 must be taken into account to deal with these terms. In particular, in Section 4.3 we have already dealt with dispersive terms of the kind  $\alpha \partial_{x^2 t} u$ . The Madsen and Sørensen system of PDEs (5.1) contains an analogous term  $:Bh^2 \partial_{x^2 t} q$ , which, therefore, can be discretized using the same technique previously illustrated. Thus, an auxiliary variable  $w_q$  needs to be introduced to decrease the order of derivation of this term. Following the steps already presented in Section 4.3, we define  $w_q = \partial_{x^2} q$  and rewrite the dispersive term in a lower order form as:  $Bh^2 \partial_t w_q$ , which can be now integrated using  $\mathcal{P}^1$  basis and weighted functions, without passing through the weak formulation.

The same technique can be applied in order to decrease the order of spatial derivative which characterize the terms:  $\beta gh^3 \partial_{x^3} \eta$  and  $2\beta gh^2 \partial_x h \partial_{x^2} \eta$ , which in the actual form cannot be solved without the use of the weak formulation or a set of basis interpolation functions with more restrictive continuity properties. In the previous section, the reconstruction of the second space derivative of the variable  $\eta$  has already been introduced, but it was applied on the only third space derivative term, since the weak form was used to integrate the other one. In the residual context, instead, the same reconstruction is going to be compute and applied to solve both these terms.

In conclusion we have to double the number of unknowns, defining two auxiliary variables  $w_\eta$  and  $w_q$  for the reconstruction of the second space derivatives of the two original variables  $\eta$  and  $q$ :

$$w_\eta = \partial_{xx} \eta \quad (5.35)$$

$$w_q = \partial_{xx} q \quad (5.36)$$

The residual related to the  $q$ -equation, computed on the cell  $K$  of the grid, becomes:

$$\begin{aligned}
\Phi_q^K = \int_K \left( \partial_t q_h - Bh^2 \partial_t w_{q,h} - \frac{1}{3} h \partial_x h \partial_{xt} q_h + \right. \\
\left. + \partial_x (uq)_h + gH \partial_x \eta_h - \beta g h^3 \partial_x w_{\eta,h} + \right. \\
\left. - 2\beta g h^2 w_{\eta,h} \partial_x h \right) dx \tag{5.37}
\end{aligned}$$

The integration of this expression is now possible in the  $\mathcal{P}^1$  interpolation function space and, accomplishing it using the middle point quadrature rule, leads to the semi-discrete form:

$$\begin{aligned}
\Phi_q^{i-\frac{1}{2}} = \Delta x \frac{dq_{i-\frac{1}{2}}}{dt} - \Delta x B h_{i-\frac{1}{2}}^2 \frac{dw_{i-\frac{1}{2}}}{dt} + \\
- \frac{1}{3\Delta x} h_{i-\frac{1}{2}} (h_i - h_{i-1}) \left( \frac{dq_i}{dt} - \frac{dq_{i-1}}{dt} \right) + \\
+ ((uq)_i - (uq)_{i-1}) + gH_{i-\frac{1}{2}} (\eta_i - \eta_{i-1}) + \\
- \beta g h_{i-\frac{1}{2}}^3 (w_{\eta,i} - w_{\eta,i-1}) + \\
- 2\beta g h_{i-\frac{1}{2}}^2 (h_i - h_{i-1}) w_{\eta,i-\frac{1}{2}} \tag{5.38}
\end{aligned}$$

The two definitions (5.35) and (5.36) are integrated over the domain using the middle point rule too. A more explicit approximation is then introduced, such that the matrices multiplying  $w_{\eta,j}$  and  $w_{q,j}$ , namely the mass matrices in the finite element terminology, are approximated in diagonal form or lumped. In this way, like in Section 5.2, the  $\mathcal{FD}2$  reconstruction of the second space derivatives of the quantities  $\eta$  and  $q$  results:

$$w_{\eta,i} = \frac{1}{\Delta x^2} (\eta_{i+1} - 2\eta_i + \eta_{i-1}) \tag{5.39}$$

$$w_{q,i} = \frac{1}{\Delta x^2} (q_{i+1} - 2q_i + q_{i-1}) \tag{5.40}$$

Finally, using the expressions of the flux residuals (5.34) and (5.38) together with the reconstruction of the high order space derivatives (5.39) and (5.40), the ultimate form of the central residual distribution scheme, referred to an

internal  $i$  node of the grid, for the one-dimensional Madsen and Sørensen system of equations can be written. It stands:

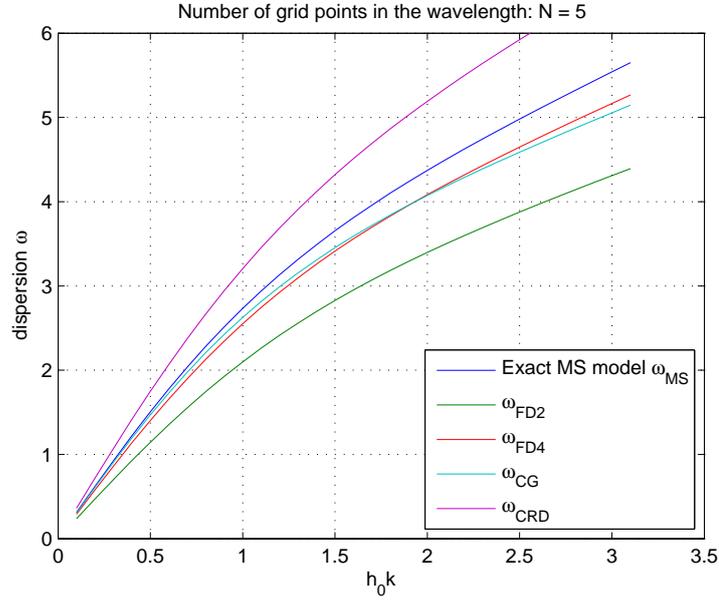
$$\begin{aligned}
 \frac{1}{2}\Phi_\eta^{i-\frac{1}{2}} + \frac{1}{2}\Phi_\eta^{i+\frac{1}{2}} &= 0 \\
 \frac{1}{2}\Phi_q^{i-\frac{1}{2}} + \frac{1}{2}\Phi_q^{i+\frac{1}{2}} &= 0 \\
 w_{\eta,i} &= \frac{1}{\Delta x^2}(\eta_{i+1} - 2\eta_i + \eta_{i-1}) \\
 w_{q,i} &= \frac{1}{\Delta x^2}(q_{i+1} - 2q_i + q_{i-1})
 \end{aligned} \tag{5.41}$$

**Truncation Error :** The scheme (5.41), when applied to the linearised problem (C.1), requires the residual fluxes  $\frac{1}{2}\Phi_\eta^K$  and  $\frac{1}{2}\Phi_u^K$  to be computed using expressions in (C.6). Substituting the Taylor expansions in space into the resulting system for the generic internal node, and rearranging the terms:

$$\begin{aligned}
 \partial_t \eta_i + h_0 \partial_x u_i &= -\frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{2} \partial_t u_i + \frac{h_0}{3} \partial_x u_i \right) + \\
 &\quad -\frac{\Delta x^4}{24} \partial_{x^4} \left( \frac{1}{2} \partial_t u_i + \frac{h_0}{5} \partial_x u_i \right) + \\
 &\quad + \mathcal{O}(\Delta x^6)
 \end{aligned} \tag{5.42}$$

$$\begin{aligned}
 \partial_t u_i - Bh_0^2 \partial_{x^2 t} u_i + g \partial_x \eta_i - \beta g h_0^2 \partial_{x^3} \eta_i &= \\
 -\Delta x^2 \partial_{x^2} \left( \frac{1}{4} \partial_t u_i - \frac{1}{3} \frac{Bh_0^2}{2} \partial_{x^2 t} u_i + \frac{1}{6} g \partial_x \eta_i - \frac{1}{4} \beta g h_0^2 \partial_{x^3} \eta_i \right) &+ \\
 -\frac{\Delta x^4}{4} \partial_{x^4} \left( \frac{1}{3} \partial_t u_i - \frac{8}{45} Bh_0^2 \partial_{x^2 t} u_i + \frac{g}{30} \partial_x \eta_i - \frac{1}{10} \beta g h_0^2 \partial_{x^3} \eta_i \right) &+ \\
 + \mathcal{O}(\Delta x^6) &
 \end{aligned} \tag{5.43}$$

(5.42) and (5.43) state that the truncation error of the scheme is  $\mathcal{O}(\Delta x^2)$  for both the equations of the original MS system and its expressions are:



**Figure 5.3:** Dispersion  $w$  against the wavenumber  $k$  of the central  $\mathcal{RD}$  scheme, compared to Airy theory value and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes  $N = 5$ .

$$\text{TE}_{\mathcal{RD}}^{\eta} = \frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{2} \partial_t u_i + \frac{h_0}{3} \partial_x u_i \right) + \mathcal{O}(\Delta x^4) \quad (5.44)$$

$$\begin{aligned} \text{TE}_{\mathcal{RD}}^q &= \Delta x^2 \partial_{x^2} \left( \frac{1}{4} \partial_t u_i - \frac{1}{3} \frac{B h_0^2}{2} \partial_{x^2 t} u_i + \frac{1}{6} g \partial_x \eta_i - \frac{1}{4} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\ &+ \mathcal{O}(\Delta x^4) \end{aligned} \quad (5.45)$$

**Dispersion Analysis :** Using the definition of  $\mu$  and  $\mu_{\Delta x}$  given in the previous sections, the analysis of the dispersion and dissipation of the scheme leads to the following results:

$$\xi_{\mathcal{RD}} = 0 \quad (5.46)$$

$$\omega_{RD} = c^2 k^2 \frac{\sin^2(\mu_{\Delta x})}{\mu^2} \frac{1 + 2\beta\mu^2 \frac{1 - \cos(\mu_{\Delta x})}{\mu_{\Delta x}^2}}{\frac{1 + \cos(\mu_{\Delta x})}{2} \left( \frac{1 + \cos(\mu_{\Delta x})}{2} + \frac{B\mu^2}{2} \frac{1 - \cos(2\mu_{\Delta x})}{\mu_{\Delta x}^2} \right)} \quad (5.47)$$

Due to the centred characteristic of the scheme and to the non-dissipative character of the equations under analysis, the dissipation parameter  $\xi_{RD}$  is correctly null, while the trend of the dispersion  $\omega_{RD}$  is shown in Figure 5.3. The trend, compared to the other centred schemes here presented, shows a worse representation of the dispersion properties of the model with respect to the higher accurate Galerkin and  $\mathcal{FD}4$  schemes, being however better in comparison to the  $\mathcal{FD}2$  scheme, which has the same truncation error order of accuracy.

## 5.4 Stabilized Upwind methods

In the previous chapter, some stabilized upwind methods have been presented for the advection with dispersion problem. The system of extended Boussinesq equations we are dealing with now presents very similar characteristic to the equation studied there. Therefore, the exigency of stabilizing the centred schemes just developed in particular flow conditions, in order to create semi-discrete schemes which are robust independently from the method used for the time integration, is still present. Hence, streamline upwind diffusion technique is applied with the same observations made for the previous advection-dispersion case, thus, using the residual of the equation in order to influence the stability of a given central scheme.

Due to the fact that the upwind stabilized term is computed taking into account the verse of propagation of the information, in equation (4.34) the advection velocity  $a$  was taken into account and included in the upwind weighting functions definition (3.47). Here, we have to deal with a system of PDEs, which, from the theory, may be characterised by two different directions of propagation of the information. Due to the intrinsic hyperbolic nature of the problem, we will use the Jacobian matrix of the shallow water system in order to predict these behaviour of the solution and compute the upwind term.

Being  $\mathcal{R}(\bar{\eta}, \bar{q}) = [\mathcal{R}_\eta(\bar{\eta}, \bar{q}) \ \mathcal{R}_q(\bar{\eta}, \bar{q})]^T$  the residual of the MS system in the sense of the definition (2.2), and being  $\mathcal{R}_\eta(\bar{\eta}, \bar{q})$  and  $\mathcal{R}_q(\bar{\eta}, \bar{q})$  the vectors of the residuals of the two equations of the system, the spatially discrete scheme

that we are going to developed can be written in the general form:

$$\int_{\Omega} \varphi_i \mathcal{R} dx + \frac{\Delta x}{2} \text{sign}(A) \int_{\Omega} \partial_x \varphi_i \mathcal{R} dx = 0 \quad (5.48)$$

being  $A$  the Jacobian matrix of the nonlinear shallow water system. The first integral of equation (5.48) results in the centred Galerkin or residual distribution schemes just developed in the previous sections, while the second, which represents the real upwind term, is independent on the kind of the central scheme used, exactly as in Sections 3.4 and 4.4. In order to be accomplished, its solution requires, in fact, that the residual  $\mathcal{R}$  would not contain any term characterised by a spatial derivative order higher than one. As already explained before, this is due to the particular function space chosen for the weighting functions  $\varphi_i$  and to the fact that a spatial derivative is already applied to them. Thus, auxiliary variables have to be defined to decrease the high order derivative terms present in the  $q$ -equation of the MS system, similarly to what done in the previous section for the residual distribution scheme development. The result reached in this way is very near to what shown in the Section 5.3; the dimension of the system is doubled by means of the introduction of  $w_{\eta}$  and  $w_q$  for the reconstruction of the second order spatial derivative of the original  $\eta$  and  $q$  variables and the same expression of the residual fluxes  $\Phi_{\eta}^K$  and  $\Phi_q^K$  can be found solving the integrals.

In particular, in each system related to the general internal node  $i$  of the mesh, the upwind term is made by the two contributions given by the left and right cells adjacent to the node itself, respectively  $i-1/2$  and  $i+1/2$ . In this context, the 2x2 square matrix  $A$  is computed locally on both cells such that the resulted scheme, related to the  $i$  node, can be written in the form:

$$CS + \frac{\text{sign}(A_{i-\frac{1}{2}})}{2} \begin{bmatrix} \Phi_{\eta}^{i-\frac{1}{2}} \\ \Phi_q^{i-\frac{1}{2}} \end{bmatrix} - \frac{\text{sign}(A_{i+\frac{1}{2}})}{2} \begin{bmatrix} \Phi_{\eta}^{i+\frac{1}{2}} \\ \Phi_q^{i+\frac{1}{2}} \end{bmatrix} \quad (5.49)$$

being  $CS$  the centred scheme the stabilization is applied to and using (5.34) and (5.38) for the residuals computation.  $CS$  represent the MS system spatially discretized with a central scheme; thus, referred to a generic internal node, it becomes a vector including the two semi-discrete equations  $CS_{\eta}$  and  $CS_q$  of the system, written for the same reference node. If  $CS_{\eta}$  and  $CS_q$  assume the expressions of the Galerkin finite element discretization (E.1), the streamline upwind Petrov-Galerkin method is obtained. Otherwise, if

the central residual distribution approximation (5.41) is there substituted, an upwind residual distribution method is reached.

Finally the  $\text{sign}(A_K)$  matrix is computed using a spectral decomposition of the local  $A_K$ , resulting in:

$$\begin{aligned} \text{sign}(A_K) &= R\text{sign}(\Lambda)L \\ &= \frac{1}{2c_K} \begin{bmatrix} 1 & 1 \\ u_K - c_K & u_K + c_K \end{bmatrix} \\ &\quad \text{sign} \left( \begin{bmatrix} u_K - c_K & 0 \\ 0 & u_K + c_K \end{bmatrix} \right) \\ &\quad \begin{bmatrix} u_K + c_K & -1 \\ c_K - u_K & 1 \end{bmatrix} \end{aligned} \quad (5.50)$$

where  $R$  and  $L$  are respectively the matrices of the right and left eigenvectors of  $A_K$ ,  $\Lambda$  is a diagonal matrix, which contains the eigenvalues of  $A_K$ ,  $u_K$  and  $c_K$  are the values of the local velocity and celerity of the flow.

The scheme is applied to the linearised Madsen and Sørensen problem in Appendix C. The resulting form (C.8) will be used in the following paragraphs in order to study the properties of these methods and to compare them with those of the central schemes illustrated in the previous sections of this chapter.

**Truncation Error :** Substituting the standard Taylor expansions in space into the system (C.8) and rearranging the terms, in the  $SU/PG$  case we obtain:

$$\begin{aligned} \partial_t \eta_i + h_0 \partial_x u_i &= \frac{c \Delta x}{g} \partial_x (\partial_t u_i - B h_0^2 \partial_{x^2 t} u_i + g \partial_x \eta_i - \beta g h_0^2 \partial_{x^3} \eta_i) + \\ &\quad - \frac{\Delta x^2}{6} \partial_{x^2} (\partial_t u_i + h_0 \partial_x u_i) + \\ &\quad + \frac{c \Delta x^3}{2g} \partial_{x^3} \left( \frac{1}{3} \partial_t u_i - \frac{1}{2} B h_0^2 \partial_{x^2 t} u_i + \frac{1}{6} g \partial_x \eta_i - \frac{1}{3} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\ &\quad - \frac{\Delta x^4}{24} \partial_{x^4} \left( \frac{1}{3} \partial_t u_i + \frac{h_0}{5} \partial_x u_i \right) + \mathcal{O}(\Delta x^5) \end{aligned} \quad (5.51)$$

$$\begin{aligned}
\partial_t u_i - Bh_0^2 \partial_{x^2 t} u_i + g \partial_x \eta_i - \beta g h_0^2 \partial_{x^3} \eta_i &= \frac{g \Delta x}{c} \partial_x (\partial_t \eta_i + h_0 \partial_x u_i) + \\
&- \frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{3} \partial_t u_i - \frac{Bh_0^2}{6} \partial_{x^2 t} u_i + \frac{1}{3} g \partial_x \eta_i - \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\
&+ \frac{g \Delta x^3}{6c} \partial_x \left( \partial_t \eta_i + \frac{h_0}{2} \partial_x u_i \right) + \\
&- \frac{\Delta x^4}{8} \partial_{x^4} \left( \frac{1}{9} \partial_t u_i - \frac{Bh_0^2}{45} \partial_{x^2 t} u_i + \frac{g}{15} \partial_x \eta_i - \frac{1}{5} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\
&+ \mathcal{O}(\Delta x^5)
\end{aligned} \tag{5.52}$$

which means that the expression of the truncation error:

$$\begin{aligned}
\text{TE}_{SU/\mathcal{PG}}^\eta &= \frac{c \Delta x^3}{2g} \partial_{x^3} \left( \frac{1}{3} \partial_t u_i - \frac{1}{2} Bh_0^2 \partial_{x^2 t} u_i + \frac{1}{6} g \partial_x \eta_i - \frac{1}{3} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\
&- \frac{\Delta x^4}{24} \partial_{x^4} \left( \frac{1}{3} \partial_t u_i + \frac{h_0}{5} \partial_x u_i \right) + \mathcal{O}(\Delta x^5)
\end{aligned} \tag{5.53}$$

$$\begin{aligned}
\text{TE}_{SU/\mathcal{PG}}^u &= - \frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{3} \partial_t u_i - \frac{Bh_0^2}{6} \partial_{x^2 t} u_i + \frac{1}{3} g \partial_x \eta_i - \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\
&+ \frac{g \Delta x^3}{6c} \partial_x \left( \partial_t \eta_i + \frac{h_0}{2} \partial_x u_i \right) + \\
&- \frac{\Delta x^4}{8} \partial_{x^4} \left( \frac{1}{9} \partial_t u_i - \frac{Bh_0^2}{45} \partial_{x^2 t} u_i + \frac{g}{15} \partial_x \eta_i - \frac{1}{5} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\
&+ \mathcal{O}(\Delta x^5)
\end{aligned} \tag{5.54}$$

is  $\mathcal{O}(\Delta x^3)$ , relating to the  $\eta$ -equation, and  $\mathcal{O}(\Delta x^2)$ , relating to the  $q$ -equation of the MS system. These results confirm that the kind of stabilisation added to the central Galerkin finite element discretization does not decrease, or decrease only weakly, its order of accuracy. The results, especially the one related to the  $\eta$ -equation, appear to be in strong analogy with respect to what obtained in the linear scalar advection case.

The same observations arise from the analysis of the  $\mathcal{U}/\mathcal{RD}$  scheme. In this case we have:

$$\begin{aligned}
 \partial_t \eta_i + h_0 \partial_x u_i &= \frac{c \Delta x}{g} \partial_x (\partial_t u_i - B h_0^2 \partial_{x^2 t} u_i + g \partial_x \eta_i - \beta g h_0^2 \partial_{x^3} \eta_i) + \\
 &\quad - \frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{2} \partial_t u_i + \frac{h_0}{3} \partial_x u_i \right) + \\
 &\quad + \frac{c \Delta x^3}{2g} \partial_{x^3} \left( \frac{1}{3} \partial_t u_i - \frac{1}{2} B h_0^2 \partial_{x^2 t} u_i + \frac{1}{6} g \partial_x \eta_i - \frac{1}{3} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\
 &\quad - \frac{\Delta x^4}{24} \partial_{x^4} \left( \frac{1}{2} \partial_t u_i + \frac{h_0}{5} \partial_x u_i \right) + \mathcal{O}(\Delta x^6)
 \end{aligned} \tag{5.55}$$

$$\begin{aligned}
 \partial_t u_i - B h_0^2 \partial_{x^2 t} u_i + g \partial_x \eta_i - \beta g h_0^2 \partial_{x^3} \eta_i &= \frac{g \Delta x}{c} \partial_x (\partial_t \eta_i + h_0 \partial_x u_i) + \\
 &\quad - \Delta x^2 \partial_{x^2} \left( \frac{1}{4} \partial_t u_i - \frac{1}{3} \frac{B h_0^2}{2} \partial_{x^2 t} u_i + \frac{1}{6} g \partial_x \eta_i - \frac{1}{4} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\
 &\quad + \frac{g \Delta x^3}{6c} \partial_x \left( \partial_t \eta_i + \frac{h_0}{2} \partial_x u_i \right) + \\
 &\quad - \frac{\Delta x^4}{4} \partial_{x^4} \left( \frac{1}{3} \partial_t u_i - \frac{8}{45} B h_0^2 \partial_{x^2 t} u_i + \frac{g}{30} \partial_x \eta_i - \frac{1}{10} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\
 &\quad + \mathcal{O}(\Delta x^6)
 \end{aligned} \tag{5.56}$$

which, imaging to substituting the exact solution of the MS system, leads the expression of the truncation error:

$$\begin{aligned}
 \text{TE}_{\mathcal{U}/\mathcal{RD}}^\eta &= - \frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{2} \partial_t u_i + \frac{h_0}{3} \partial_x u_i \right) + \\
 &\quad + \frac{c \Delta x^3}{2g} \partial_{x^3} \left( \frac{1}{3} \partial_t u_i - \frac{1}{2} B h_0^2 \partial_{x^2 t} u_i + \frac{1}{6} g \partial_x \eta_i - \frac{1}{3} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\
 &\quad - \frac{\Delta x^4}{24} \partial_{x^4} \left( \frac{1}{2} \partial_t u_i + \frac{h_0}{5} \partial_x u_i \right) + \mathcal{O}(\Delta x^5)
 \end{aligned} \tag{5.57}$$

$$\begin{aligned}
 \text{TE}_{u/\mathcal{RD}}^q &= -\Delta x^2 \partial_{x^2} \left( \frac{1}{4} \partial_t u_i - \frac{1}{3} \frac{Bh_0^2}{2} \partial_{x^2 t} u_i + \frac{1}{6} g \partial_x \eta_i - \frac{1}{4} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\
 &+ \frac{g \Delta x^3}{6c} \partial_x \left( \partial_t \eta_i + \frac{h_0}{2} \partial_x u_i \right) + \\
 &- \frac{\Delta x^4}{4} \partial_{x^4} \left( \frac{1}{3} \partial_t u_i - \frac{8}{45} B h_0^2 \partial_{x^2 t} u_i + \frac{g}{30} \partial_x \eta_i - \frac{1}{10} \beta g h_0^2 \partial_{x^3} \eta_i \right) + \\
 &+ \mathcal{O}(\Delta x^5)
 \end{aligned} \tag{5.58}$$

**Dispersion Analysis :** The analysis of the dispersion properties of both these upwind schemes leads to the same form of the dispersion and dissipation parameters  $\xi$  and  $\omega$ :

$$\xi = -\frac{A_\nu}{2A_{\nu^2}} \tag{5.59}$$

$$\omega = \sqrt{\frac{A_s}{A_{\nu^2}} - \xi^2} \tag{5.60}$$

In the  $SU/\mathcal{PG}$  scheme, the expressions of the coefficients  $A_{\nu^2}$ ,  $A_\nu$  and  $A_s$  are found to be:

$$\begin{aligned}
 A_{\nu^2} &= \frac{(2 + \cos \mu_{\Delta x})}{3} \left( \frac{(2 + \cos \mu_{\Delta x})}{3} + 2B\mu^2 \frac{(1 - \cos(\mu_{\Delta x}))}{\mu_{\Delta x}^2} \right) + \\
 &+ \frac{\sin^2 \mu_{\Delta x}}{4} \left( 1 + 2B\mu^2 \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}^2} \right)
 \end{aligned} \tag{5.61}$$

$$\begin{aligned}
 A_\nu &= ck \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}} \left( \frac{(2 + \cos \mu_{\Delta x})}{3} + 2B\mu^2 \frac{(1 - \cos(\mu_{\Delta x}))}{\mu_{\Delta x}^2} \right) + \\
 &+ 2\beta c \mu^2 \frac{(1 - \cos \mu_{\Delta x})^2}{\mu_{\Delta x}^3} \left( \frac{(2 + \cos \mu_{\Delta x})}{3} + 2B\mu^2 \frac{(1 - \cos(\mu_{\Delta x}))}{\mu_{\Delta x}^2} \right) + \\
 &- ck \frac{\sin^2 \mu_{\Delta x}}{2\mu_{\Delta x}} \left( 1 + 2B\mu^2 \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}^2} \right) \left( 1 + 2\beta \mu^2 \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}^2} \right) + \\
 &+ ck \frac{(1 - \cos \mu_{\Delta x})^2}{6\mu_{\Delta x}}
 \end{aligned} \tag{5.62}$$

$$\begin{aligned}
 A_s = c^2 k^2 \frac{(1 - \cos \mu_{\Delta x})^2}{\mu_{\Delta x}^2} \left( 1 + 2\beta \mu^2 \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}^2} \right) + \\
 + c^2 k^2 \frac{\sin^2 \mu_{\Delta x}}{\mu_{\Delta x}^2} \left( 1 + 2B \mu^2 \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}^2} \right)
 \end{aligned} \tag{5.63}$$

while for the  $\mathcal{U}/\mathcal{RD}$  scheme, these other expressions result:

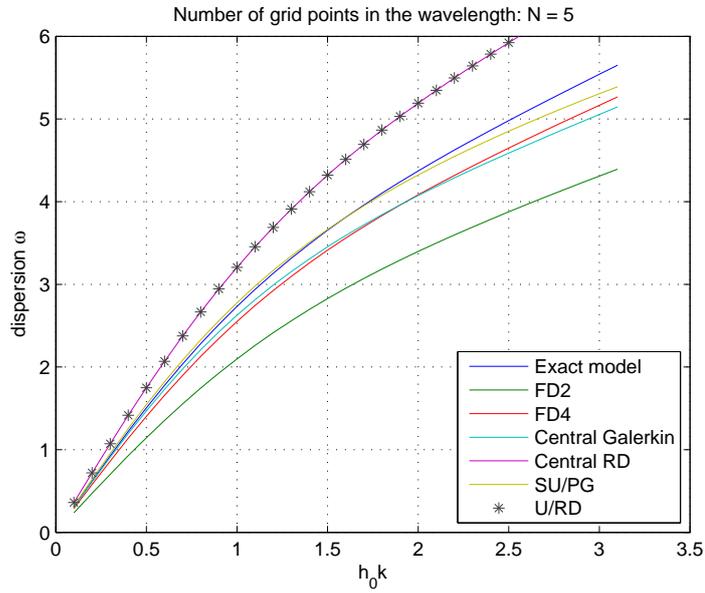
$$\begin{aligned}
 A_{\nu^2} &= \frac{(1 + \cos \mu_{\Delta x})}{2} \left( \frac{(1 + \cos \mu_{\Delta x})}{2} + B \mu^2 \frac{(1 - \cos (2\mu_{\Delta x}))}{2\mu_{\Delta x}^2} \right) + \\
 &+ \frac{\sin^2 \mu_{\Delta x}}{4} \left( 1 + 2B \mu^2 \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}^2} \right) \\
 A_{\nu} &= ck \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}} \left( \frac{(1 + \cos \mu_{\Delta x})}{2} + B \mu^2 \frac{(1 - \cos (2\mu_{\Delta x}))}{2\mu_{\Delta x}^2} \right) + \\
 &+ 2\beta c \mu^2 \frac{(1 - \cos \mu_{\Delta x})^2}{\mu_{\Delta x}^3} \left( \frac{(1 + \cos \mu_{\Delta x})}{2} + B \mu^2 \frac{(1 - \cos (2\mu_{\Delta x}))}{2\mu_{\Delta x}^2} \right) + \\
 &- ck \frac{\sin^2 \mu_{\Delta x}}{2\mu_{\Delta x}} \left( 1 + 2B \mu^2 \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}^2} \right) \left( 1 + 2\beta \mu^2 \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}^2} \right)^2
 \end{aligned} \tag{5.64}$$

$$\begin{aligned}
 A_s &= c^2 k^2 \frac{(1 - \cos \mu_{\Delta x})^2}{\mu_{\Delta x}^2} \left( 1 + 2\beta \mu^2 \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}^2} \right) + \\
 &+ c^2 k^2 \frac{\sin^2 \mu_{\Delta x}}{\mu_{\Delta x}^2} \left( 1 + 2B \mu^2 \frac{(1 - \cos \mu_{\Delta x})}{\mu_{\Delta x}^2} \right)
 \end{aligned}$$

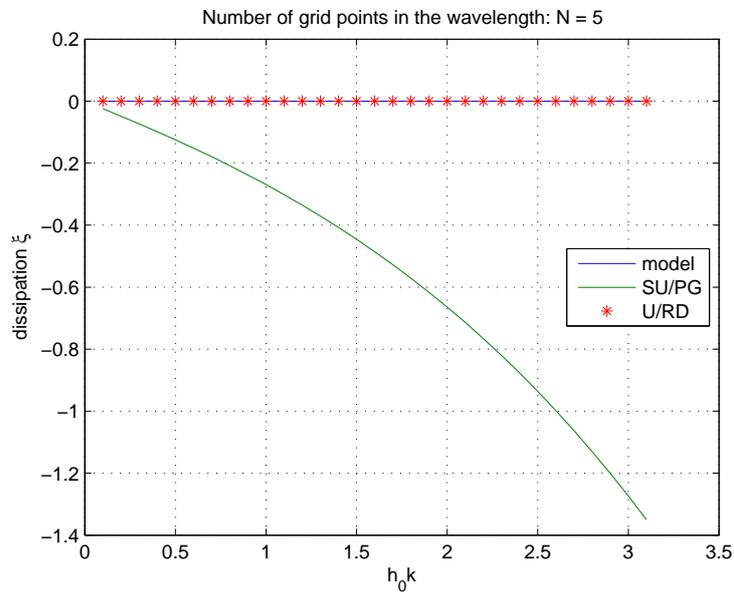
Figure 5.4 and 5.5 shows a comparison between the trend of the previous expressions of  $\xi$  and  $\omega$ , in dependence of a variation of the wavenumber  $k$ . Directly related the two pictures with Fig.4.4 and 4.5 respectively, a strong analogy can be seen and similar observations can be done.

In fact, in the context of the Boussinesq extended equations of Madsen and Sørensen as well as in the simple linear scalar advection with dispersion case, the  $\mathcal{SU}/\mathcal{PG}$  scheme is able to give the best approximation of the dispersive properties of the model, even with respect to the more accurate  $\mathcal{FD4}$ . Its dissipation, however, shows a non-physical trend which reflects the artificial addition of dissipation in order to accomplish the stabilisation.

On the other hand, the  $\mathcal{U}/\mathcal{RD}$  appears to not change the characteristics of the central  $\mathcal{RD}$  scheme in both the dispersive and dissipation fields.



**Figure 5.4:** Dispersion  $w$  against the wavenumber  $k$  of the stabilized upwind  $SU/PG$  and  $U/RD$  schemes, compared to Airy theory value and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes  $N = 5$ .



**Figure 5.5:** Diffusion  $\xi$  against the wavenumber  $k$  of the stabilized upwind  $SU/PG$  and  $U/RD$  schemes, compared to Airy theory value and to the previous mentioned schemes, when the wavelength is discretized with a number of grid nodes  $N = 5$ .

## 5.5 Conclusions

In this chapter the system of extended Boussinesq-type equations of Madsen and Sørensen has been discretized in space applying the schemes already introduced in the previous chapters for the simpler linear scalar advection and advection with dispersion cases. The diverse semi-discrete systems of ordinary differential equations in time, obtained in the various sections of this chapter, are thus ready to be integrated in time with an appropriate integration method, such as those that will be described in the next chapter, with the approximate solution which will be computed at any time by substituting the  $\eta_j(t)$  and  $q_j(t)$  in the (5.21) interpolations over the set of basis functions. In this context, difficulties arise from the exigency of approximating high order space derivatives of both variables of the problem  $\eta$  and  $q$ , which generally leads to schemes characterised by more extended stencils. The exigency of decreasing the order of these terms through the reconstruction of second order derivatives, introducing auxiliary variables into the problem, increases the dimension of the system doubling the number of variables and equations. However, the approximation of the high order spatial derivative seems to not affect the accuracy of the resulted schemes, exactly as in the scalar linear advection-dispersion case.

The central Galerkin finite element approximation results in a high accuracy fourth order scheme regarding the  $\eta$ -equation, being only second order accurate respect to the second equation of the system. Its dispersive properties, however, show a high accordance with respect to the exact dispersion trend of the model, much better than the  $\mathcal{FD}2$  which one hand adopts the same reconstruction of the space derivatives, but which on the other hand is not characterised by the presence of the Galerkin mass matrix. The stabilization of the Galerkin scheme, using the residual of the system and the Jacobian matrix of the shallow water system, in order to compute the direction of propagation of the informations, leads to a weakly less accurate scheme, whose dispersion characteristics are, instead, very close to the exact solution ones even showing an artificial trend of the diffusive parameter.

On the other hand, the central  $\mathcal{RD}$  scheme derived in Section 5.3 and his upwind version (Section 5.4) give second order accuracy schemes in both the equations of the system. Their dispersion properties appears to be exactly the same, which means that the stabilization term does not worsen neither the accuracy nor the dispersion of the central scheme. Even the dissipation parameter remains null in this case and does not show any non-physical trend.

However their properties are worse with respect to Galerkin ones, confirming

that Galerkin mass matrix gives the better approximation of the solution and of the dispersive properties of the model with respect to other schemes which use the same order of approximation of the spatial derivatives.



# Chapter 6

## Time discretization

We are trying to compute an accurate description of the phenomena which govern the propagation of gravity waves in the near-shore zone, where the system of shallow water equations becomes inaccurate due to the fact that it is not able to describe its dispersion characteristics, which are no more negligible with respect to the nonlinear ones. For this reason, we have chosen the Madsen-Sørensen system of extended Boussinesq-type equations, having accomplished a study of the dispersion properties of this model through a comparison with Airy wave theory.

(1.9) is a nonlinear system of partial differential equations, which, in order to be solved, requires a space-time discretization. In the previous chapter, several methods were applied to spatially discretised MS system. Each of the semi-discrete schemes derived there represents a system of ordinary differential equations in time which must be now integrated with a proper time integration scheme. The general system we have now to deal with, referred to the generic internal node  $i$  of the mesh, is one of the kind:

$$A_{ij}\dot{y}_j - f_i = 0 \tag{6.1}$$

The integration in time of (6.1) leads to a complete discrete model which can finally be implemented and numerically solved. The approximation to the time derivative  $\dot{y}$  is determined by the time integration method used and to the order of that method, but has the general form:

$$\dot{y}_j = \frac{y_j - z_j}{\alpha\Delta t} \tag{6.2}$$

where the exact form of  $z_j$  depends on the time integration algorithm that it is being used and depends on informations inherited from the previous time steps. In particular  $z$  does not depend on the current solution  $y$ . Also the constant  $\alpha$  depends on the time integration scheme used, while  $\Delta t$  represents the current time step value.

Since in general implicit time integration methods are chosen for the better stability properties which can be reached, the cost of solving a nonlinear problem with an iterative method must be paid. The equation system is supplied to the time integration algorithm in the form of a residual vector  $r$ :

$$r_i = A_{ij}\dot{y}_j - f_i \quad (6.3)$$

The solution is advanced in time by using in (6.3) the approximation (6.2) of the time derivative and than by requiring the residual at next time level to be sufficiently small. The resulting system of nonlinear equations is solved using the classical iterative Newton's method which requires the solution, at each iteration step, of the linear system:

$$J_{ij}(y)\Delta y_j = -\alpha\Delta t r_i(y) \quad (6.4)$$

where  $\Delta y$  is the solution increment at the current iteration step and  $J_{ij}(y)$  is the nonlinear Jacobian matrix of the system given by:

$$J_{ij} = \alpha\Delta t \frac{dr_i}{dy_i} = \alpha\Delta t \frac{df_i}{dy_i} - A_{ij} \quad (6.5)$$

An exact Newton method would require  $J_{ij}$  to be recomputed after each Newton iteration but this would require a matrix factorisation step at each iteration which would be very expensive in general. This is the most computationally part of the time integration process thus convergence properties of a scheme are an important field to be investigated. In order to diminish the computational effort of the iteration process, it is possible to maintain the Jacobian constant as long as possible and only to recompute it when the error given by its approximations exceeds a certain tolerance. In particular the computation of the Jacobian can be needed in the initial stages when the order of time integration has not been reached yet and is still growing.

Many time-integration methods can be coupled with the description just given, but this choice strongly influences the stability properties of the schemes.

In fact, the semi-discrete methods proposed are weakly stable since the real part of the amplification coefficient ( $\xi$ ) is null, or weakly negative, for all of them. This means that they must be accurately integrated in time if an uncontrolled divergence of the solution is wanted to be avoid. In particular, an A-stable method maintains the modes of the solution in the negative part of the complex plan, while, on the contrary, this is not guarantee by a not A-stable scheme which can, thus, produce oscillations due to spurious modes if the time step is too coarse. In this work a Crank-Nicolson ( $\mathcal{CN}$ ) implicit multistep and A-stable method and a third order Backward Differentiation Formula ( $\mathcal{BDF3}$ ) are implemented. The properties of the resulting numerical schemes are investigated in the following sections.

## 6.1 Crank-Nicolson time integration method

We present, in this section, the integration of the MS semi-discrete systems of ordinary differential equation in time, developed in the previous chapter, using the Crank-Nicolson method. It is an implicit linear one-step method which, for the generic Cauchy problem  $A\dot{u} = f(u, t)$ , can compute the value of the variable at the time  $n+1$  as:

$$A\Delta^n u = \frac{h}{2} f\left(u(t^{n+\frac{1}{2}}), t^{n+\frac{1}{2}}\right) \quad (6.6)$$

where  $\Delta^n u = u(t^{n+1}) - u(t^n)$  is the increment and  $h = \Delta t/\Delta x$ . Since the function  $f$  depends on  $u^{n+1}$ , this is a nonlinear problem, which means that its solution must be computed as the zero of the function  $r(w)$ , being:

$$r(w) = A(w - u^n) - \frac{h}{2} f(w, t^{n+1}) - \frac{h}{2} f(u^n, t^n) \quad (6.7)$$

such that  $r(w) = 0$  if  $w = u^{n+1}$ . Finding the zero of (6.7) means necessary to use an iterative method as the one described above. Here the classical iterative Newton's method is applied and written in the form of the system:

$$w^{k+1} = w^k - J^{-1}r(w^k) \quad (6.8)$$

where  $J$  is the Jacobian matrix of the problem and  $w^k$  is the solution at the

iteration step  $k$ . The loop starts with  $w^1 = w^n$  and ends setting  $w^{k_{end}} = u^{n+1}$ , when the convergence conditions are satisfied, which means when  $\log(\|r(w^{k+1})\|/\|r(w^1)\|)$  or  $\|r(w^{k+1})\|_\infty$  become less than a fixed tolerance value. Thus the convergence properties of the methods are an important field of study and the properties of the numerical schemes here developed will be investigated in the following sections.

Paying the cost of solving a nonlinear problem, better stability properties can be reached. Von Neumann stability analysis, in fact shows that Crank-Nicolson scheme has the advantage, with respect to explicit time-integration schemes, of being neutrally stable, which means that errors remain constant as the computations are carried forward. The method is also A-stable which means that the resulting schemes would be stable and a low CFL number is not needed for stability but is however required for numerical accuracy.

The expression of the matrix  $A$  and vector  $f$  in (6.3) are strictly dependent on the method chosen to spatially discretize the MS system, between the ones derived in Chapter 5. However, in general, matrix  $A$  gathers the mass matrix terms together with the ones resulting from the discretization of the operators  $Bh^2\partial_{x^2}$  and  $\frac{1}{3}h\partial_x h\partial_x$ . The vector  $f$ , instead, contained all the other terms discretized in space following the schemes of Chapter 5:  $\partial_x(uq)$ ,  $gH\partial_x\eta$ ,  $\beta gh^3\partial_{x^3}\eta$  and  $2\beta gh^2\partial_x h\partial_{x^2}\eta$ .

### 6.1.1 Truncation Error

The form of the the matrix  $A$  and vector  $f$  of equation (6.6) depend on the space discretization scheme adopted. Here, the only study of the truncation error of these methods are presented. In order to accomplish this, the Taylor expansion series in time:

$$u^{n+1} = u^n + \Delta t \partial_t u^n + \frac{\Delta t^2}{2} \partial_{t^2} u^n + \frac{\Delta t^3}{6} \partial_{t^3} u^n + \frac{\Delta t^4}{24} \partial_{t^4} u^n + \frac{\Delta t^5}{120} \partial_{t^5} u^n + \dots \quad (6.9)$$

has to be taken into account together with the analogous spatial ones in (3.12). The study of the accuracy properties is going to be realized in the linearised field of the MS problem, similarly to what has been done in the previous chapters for all the space-discrete schemes. The space-time discrete schemes, which give a numerical approximation to the linearised MS system and are here used to accomplish the analysis, can be found in Appendix F.

**Galerkin FE scheme with  $\mathcal{CN}$  time integration :** Taking into account system (F.1), introducing the Taylor expansions and rearranging the resulted terms, the following expression can be found:

$$\begin{aligned}
 \partial_t \eta + h_0 \partial_x u = & -\frac{\Delta t}{2} \partial_t (\partial_t \eta + h_0 \partial_x u) - \frac{\Delta x^2}{6} \partial_{x^2} (\partial_t \eta + h_0 \partial_x u) + \\
 & -\frac{\Delta t^2}{2} \partial_{t^2} \left( \frac{1}{3} \partial_t \eta + \frac{h_0}{2} \partial_x u \right) - \frac{\Delta t^3}{12} \partial_{t^3} \left( \frac{1}{2} \partial_t \eta + h_0 \partial_x u \right) + \\
 & -\frac{\Delta t \Delta x^2}{12} \partial_{x^2 t} (\partial_t \eta + h_0 \partial_x u) + \\
 & -\frac{\Delta x^4}{24} \partial_x^4 \left( \frac{1}{3} \partial_t \eta + \frac{h_0}{5} \partial_x u \right) + \mathcal{O}(\Delta x^5, \Delta t^4)
 \end{aligned} \tag{6.10}$$

$$\begin{aligned}
 \partial_t u - B h_0^2 \partial_{x^2 t} u + g \partial_x \eta - \beta g h_0^2 \partial_{x^3} \eta = & \\
 & -\frac{\Delta t}{2} \partial_t (\partial_t u - B h_0^2 \partial_{x^2 t} u + g \partial_x \eta - \beta g h_0^2 \partial_{x^3} \eta) \\
 & -\frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{3} \partial_t u - \frac{1}{6} B h_0^2 \partial_{x^2 t} u + \frac{1}{3} g \partial_x \eta - \frac{1}{2} \beta g h_0^2 \partial_{x^3} \eta \right) + \\
 & -\frac{\Delta t^2}{2} \partial_{t^2} \left( \frac{1}{3} \partial_t u - \frac{1}{3} B h_0^2 \partial_{x^2 t} u + \frac{1}{2} g \partial_x \eta - \frac{1}{2} \beta g h_0^2 \partial_{x^3} \eta \right) + \\
 & + \mathcal{O}(\Delta x^3, \Delta t^3)
 \end{aligned} \tag{6.11}$$

Imaging to substitute the exact solution  $[\eta, u]$  of the original system (1.9), the expression of the truncation error of the scheme is obtained:

$$\begin{aligned}
 \text{TE}_{CG}^\eta = & -\frac{\Delta t^2}{2} \partial_t^2 \left( \frac{1}{3} \partial_t \eta + \frac{h_0}{2} \partial_x u \right) - \frac{\Delta x^4}{24} \partial_x^4 \left( \frac{1}{3} \partial_t \eta + \frac{h_0}{5} \partial_x u \right) + \\
 & + \mathcal{O}(\Delta x^6, \Delta t^4)
 \end{aligned} \tag{6.12}$$

$$\begin{aligned}
 \text{TE}_{CG}^u = & -\frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{3} \partial_t u - \frac{1}{6} B h_0^2 \partial_{x^2 t} u + \frac{1}{3} g \partial_x \eta - \frac{1}{2} \beta g h_0^2 \partial_{x^3} \eta \right) \\
 & -\frac{\Delta t^2}{2} \partial_{t^2} \left( \frac{1}{3} \partial_t u - \frac{1}{3} B h_0^2 \partial_{x^2 t} u + \frac{1}{2} g \partial_x \eta - \frac{1}{2} \beta g h_0^2 \partial_{x^3} \eta \right) + \\
 & + \mathcal{O}(\Delta x^4, \Delta t^3)
 \end{aligned} \tag{6.13}$$

(6.12) and (6.13) confirm that Crank-Nicolson time-integration method is second order accurate in time and also show that the Galerkin accuracy in space is preserved, since the truncation error is still fourth order for the  $\eta$ -equation and second order for the  $q$ -one.

**$\mathcal{RD}$  scheme with  $\mathcal{CN}$  time integration :** The truncation error of the semi-discrete residual distribution scheme of Section 5.3 was second order accurate in space. The integration with the  $\mathcal{CN}$  method in time and the study of the accuracy properties of this fully discrete scheme leads to analogous results to the ones obtained before for the central Galerkin finite element method. In particular the Taylor expansion of the discrete system, written for the node  $i$  of the mesh, leads after some manipulations to the expressions:

$$\begin{aligned} \partial_t \eta + h_0 \partial_x u = & -\frac{\Delta t}{2} \partial_t (\partial_t \eta + h_0 \partial_x u) - \frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{2} \partial_t \eta + \frac{h_0}{3} \partial_x u \right) + \\ & -\frac{\Delta t^2}{2} \partial_{t^2} \left( \frac{1}{3} \partial_t \eta + \frac{h_0}{2} \partial_x u \right) + \mathcal{O}(\Delta x^4, \Delta t^3) \end{aligned} \quad (6.14)$$

$$\begin{aligned} \partial_t u - Bh_0^2 \partial_{x^2 t} u + g \partial_x \eta - \beta g h_0^2 \partial_{x^3} \eta = & \\ & -\frac{\Delta t}{2} \partial_t (\partial_t u - Bh_0^2 \partial_{x^2 t} u + g \partial_x \eta - \beta g h_0^2 \partial_{x^3} \eta) \\ & -\Delta x^2 \partial_{x^2} \left( \frac{1}{4} \partial_t u - \frac{1}{3} Bh_0^2 \partial_{x^2 t} u + \frac{1}{6} g \partial_x \eta - \frac{1}{4} \beta g h_0^2 \partial_{x^3} \eta \right) + \\ & -\frac{\Delta t^2}{2} \partial_{t^2} \left( \frac{1}{3} \partial_t u - \frac{1}{3} Bh_0^2 \partial_{x^2 t} u + \frac{1}{2} g \partial_x \eta - \frac{1}{2} \beta g h_0^2 \partial_{x^3} \eta \right) + \\ & + \mathcal{O}(\Delta x^4, \Delta t^3) \end{aligned} \quad (6.15)$$

The use of the exact solution of the linearised system in (6.15), gives the expressions of the truncation errors for the two equations of the MS system:

$$\begin{aligned} \text{TE}_{RD}^\eta = & -\frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{2} \partial_t \eta + \frac{h_0}{3} \partial_x u \right) - \frac{\Delta t^2}{2} \partial_{t^2} \left( \frac{1}{3} \partial_t \eta + \frac{h_0}{2} \partial_x u \right) + \\ & + \mathcal{O}(\Delta x^4, \Delta t^3) \end{aligned} \quad (6.16)$$

$$\begin{aligned}
 \text{TE}_{RD}^u &= -\Delta x^2 \partial_{x^2} \left( \frac{1}{4} \partial_t u - \frac{1}{3} B h_0^2 \partial_{x^2 t} u + \frac{1}{6} g \partial_x \eta - \frac{1}{4} \beta g h_0^2 \partial_{x^3} \eta \right) \\
 &\quad - \frac{\Delta t^2}{2} \partial_{t^2} \left( \frac{1}{3} \partial_t u - \frac{1}{3} B h_0^2 \partial_{x^2 t} u + \frac{1}{2} g \partial_x \eta - \frac{1}{2} \beta g h_0^2 \partial_{x^3} \eta \right) + \quad (6.17) \\
 &\quad + \mathcal{O}(\Delta x^4, \Delta t^3)
 \end{aligned}$$

Exactly as in the previous Galerkin case, the spatial approximation accuracy is preserved since the second order of the semi-discrete scheme of Section 5.3 is maintained. Moreover, the  $\mathcal{CN}$  integration method is confirmed to be second order accurate in time also in this case.

**Stabilized upwind schemes with  $\mathcal{CN}$  time integration :** The analysis of the truncation error of the stabilized  $\mathcal{SU}/\mathcal{PG}$  and  $\mathcal{U}/\mathcal{RD}$  schemes leads to analogous results with respect to the previous two schemes. After the integration in time, the two methods maintain the general common form (F.3). The details about the expressions which have to be given to the central scheme  $\mathcal{CS}$  or to the residuals  $\Phi_\eta^K$  and  $\Phi_u^K$ , there present, in order to obtain the two different schemes, can be found in Appendix F. Here, recalling the complete expression of the two stabilized upwind schemes for a generic internal node of the grid, the Taylor expansions in space and time are computed in the case of the  $\mathcal{SU}/\mathcal{PG}$  scheme first:

$$\begin{aligned}
 \partial_t \eta + h_0 \partial_x u &= \Delta x \frac{c}{g} \partial_x \left( \partial_t u - B h_0^2 \partial_{x^2 t} u + g \partial_x \eta - \beta g h_0^2 \partial_{x^3} \eta \right) + \\
 &\quad - \frac{\Delta t}{2} \partial_t (\partial_t \eta + h_0 \partial_x u) - \frac{\Delta x^2}{6} \partial_{x^2} (\partial_t \eta + h_0 \partial_x u) + \quad (6.18) \\
 &\quad - \frac{\Delta t^2}{2} \partial_{t^2} \left( \frac{1}{3} \partial_t \eta + \frac{h_0}{2} \partial_x u \right) + \\
 &\quad + \mathcal{O}(\Delta x^3, \Delta t^3)
 \end{aligned}$$

$$\begin{aligned}
 \partial_t u - Bh_0^2 \partial_{x^2 t} u + g \partial_x \eta - \beta g h_0^2 \partial_{x^3} \eta = & \\
 + \Delta x \frac{g}{c} \partial_x (\partial_t \eta + h_0 \partial_x u) + & \\
 - \frac{\Delta t}{2} \partial_t (\partial_t u - Bh_0^2 \partial_{x^2 t} u + g \partial_x \eta - \beta g h_0^2 \partial_{x^3} \eta) & \\
 - \frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{3} \partial_t u - \frac{1}{6} Bh_0^2 \partial_{x^2 t} u + \frac{1}{3} g \partial_x \eta - \frac{1}{2} \beta g h_0^2 \partial_{x^3} \eta \right) + & \quad (6.19) \\
 - \frac{\Delta t^2}{2} \partial_{t^2} \left( \frac{1}{3} \partial_t u - \frac{1}{3} Bh_0^2 \partial_{x^2 t} u + \frac{1}{2} g \partial_x \eta - \frac{1}{2} \beta g h_0^2 \partial_{x^3} \eta \right) + & \\
 + \mathcal{O}(\Delta x^3, \Delta t^3) &
 \end{aligned}$$

and in the case of the  $\mathcal{U}/\mathcal{RD}$  in the following:

$$\begin{aligned}
 \partial_t \eta + h_0 \partial_x u = \Delta x \frac{c}{g} \partial_x (\partial_t u - Bh_0^2 \partial_{x^2 t} u + g \partial_x \eta - \beta g h_0^2 \partial_{x^3} \eta) + & \\
 - \frac{\Delta t}{2} \partial_t (\partial_t \eta + h_0 \partial_x u) - \frac{\Delta x^2}{2} \partial_{x^2} \left( \frac{1}{2} \partial_t \eta + \frac{h_0}{3} \partial_x u \right) + & \quad (6.20) \\
 - \frac{\Delta t^2}{2} \partial_{t^2} \left( \frac{1}{3} \partial_t \eta + \frac{h_0}{2} \partial_x u \right) + & \\
 + \mathcal{O}(\Delta x^3, \Delta t^3) &
 \end{aligned}$$

$$\begin{aligned}
 \partial_t u - Bh_0^2 \partial_{x^2 t} u + g \partial_x \eta - \beta g h_0^2 \partial_{x^3} \eta = & \\
 + \Delta x \frac{g}{c} \partial_x (\partial_t \eta + h_0 \partial_x u) + & \\
 - \frac{\Delta t}{2} \partial_t (\partial_t u - Bh_0^2 \partial_{x^2 t} u + g \partial_x \eta - \beta g h_0^2 \partial_{x^3} \eta) & \\
 - \Delta x^2 \partial_{x^2} \left( \frac{1}{4} \partial_t u - \frac{1}{3} Bh_0^2 \partial_{x^2 t} u + \frac{1}{6} g \partial_x \eta - \frac{1}{4} \beta g h_0^2 \partial_{x^3} \eta \right) + & \quad (6.21) \\
 - \frac{\Delta t^2}{2} \partial_{t^2} \left( \frac{1}{3} \partial_t u - \frac{1}{3} Bh_0^2 \partial_{x^2 t} u + \frac{1}{2} g \partial_x \eta - \frac{1}{2} \beta g h_0^2 \partial_{x^3} \eta \right) + & \\
 + \mathcal{O}(\Delta x^3, \Delta t^3) &
 \end{aligned}$$

The following expression of the truncation errors of the two schemes for the two equations of the linearized MS system can now be found. For the  $\mathcal{SU}/\mathcal{PG}$  we have:

$$\begin{aligned} \text{TE}_{SU/\mathcal{PG}}^\eta &= -\frac{\Delta t^2}{2}\partial_{t^2}\left(\frac{1}{3}\partial_t\eta + \frac{h_0}{2}\partial_x u\right) + \\ &+ \frac{\Delta x^3}{2}\frac{c}{g}\partial_{x^3}\left(\frac{1}{3}\partial_t u - \frac{1}{2}Bh_0^2\partial_{x^2t}u + \frac{1}{6}g\partial_x\eta - \frac{1}{3}\beta gh_0^2\partial_{x^3}\eta\right) + \\ &+ \mathcal{O}(\Delta x^4, \Delta t^3) \end{aligned} \quad (6.22)$$

$$\begin{aligned} \text{TE}_{SU/\mathcal{PG}}^u &= -\frac{\Delta x^2}{2}\partial_{x^2}\left(\frac{1}{3}\partial_t u - \frac{1}{6}Bh_0^2\partial_{x^2t}u + \frac{1}{3}g\partial_x\eta - \frac{1}{2}\beta gh_0^2\partial_{x^3}\eta\right) + \\ &- \frac{\Delta t^2}{2}\partial_{t^2}\left(\frac{1}{3}\partial_t u - \frac{1}{3}Bh_0^2\partial_{x^2t}u + \frac{1}{2}g\partial_x\eta - \frac{1}{2}\beta gh_0^2\partial_{x^3}\eta\right) + \\ &+ \mathcal{O}(\Delta x^3, \Delta t^3) \end{aligned} \quad (6.23)$$

In the  $\mathcal{U}/\mathcal{RD}$  case, indeed:

$$\begin{aligned} \text{TE}_{\mathcal{U}/\mathcal{RD}}^\eta &= -\frac{\Delta x^2}{2}\partial_{x^2}\left(\frac{1}{2}\partial_t\eta + \frac{h_0}{3}\partial_x u\right) - \frac{\Delta t^2}{2}\partial_{t^2}\left(\frac{1}{3}\partial_t\eta + \frac{h_0}{2}\partial_x u\right) + \\ &+ \mathcal{O}(\Delta x^3, \Delta t^3) \end{aligned} \quad (6.24)$$

$$\begin{aligned} \text{TE}_{\mathcal{U}/\mathcal{RD}}^u &= -\Delta x^2\partial_{x^2}\left(\frac{1}{4}\partial_t u - \frac{1}{3}Bh_0^2\partial_{x^2t}u + \frac{1}{6}g\partial_x\eta - \frac{1}{4}\beta gh_0^2\partial_{x^3}\eta\right) + \\ &- \frac{\Delta t^2}{2}\partial_{t^2}\left(\frac{1}{3}\partial_t u - \frac{1}{3}Bh_0^2\partial_{x^2t}u + \frac{1}{2}g\partial_x\eta - \frac{1}{2}\beta gh_0^2\partial_{x^3}\eta\right) + \\ &+ \mathcal{O}(\Delta x^3, \Delta t^3) \end{aligned} \quad (6.25)$$

which express the preservation of the spatial order of accuracy of the previous semi-discrete schemes developed in Section 5.4 and state one more time that the Crank-Nicolson integration in time leads to a second order accurate scheme in time. Comparing the expressions of the truncation error formulae just found with those of the central schemes listed before, we can observe that the upwind term, computed in the general way (5.48) taking into account the residual of the equation, does not affect the expression of any  $\text{TE}^u$  or  $\text{TE}^\eta$ , at least at low order terms, which thus remain exactly the same of the  $CS$  chosen scheme. The only exception to this is given by the  $\text{TE}_{SU/\mathcal{PG}}^\eta$  where the fourth order of accuracy in space is decreased to a third order one, which has, in any case, still a small influence.

## 6.2 $\mathcal{BDF3}$ time integration method

The backward differentiation formula ( $\mathcal{BDF}$ ) is a family of implicit methods for the numerical integration of ordinary differential equations  $A\dot{u}(t^n) = f(u(t^n), t^n)$ . They are linear multistep methods, as well as the Crank-Nicolson method, but they approximate the derivative function  $\dot{u}$  of the problem's variable at a certain time  $t^n$  with the values of the variable function itself  $u(t)$  at the same time  $t^n$  and the earlier times. The general form of this class of schemes can be written as:

$$\sum_{j=1}^p \frac{1}{j} A \Delta^j u^{n+1} = h f(u^{n+1}, t^{n+1}) \quad (6.26)$$

where  $\Delta$  denotes the increment operator  $\Delta u^n = u^n - u^{n-1}$ ;  $\Delta^2 u^n = \Delta u^n - \Delta u^{n-1}$ ;  $p$  represent the number of steps of the scheme and  $u^{n+1} = u(t^{n+1})$ . More explicitly, three steps backward differentiation formula states:

$$\frac{11}{6} A \Delta u^{n+1} - \frac{7}{6} A \Delta u^n + \frac{2}{6} A \Delta u^{n-1} = h f(u^{n+1}, t^{n+1}) \quad (6.27)$$

The order of this class of methods coincides with their number of steps, which for our purposes is three, and the A-stability property characterizes the first two methods of this family. A close inspection of the stability domain for the higher order  $\mathcal{BDF}$  methods, included the third order one, shows that some parts very close to the imaginary axis do not produce stable solutions. These areas increase their extensions with the order of the scheme and are thus very thin in the case of  $p = 3$ . However the implementation of this method to integrate the schemes proposed, will produce instabilities especially in the centred ones, which thus will need low CFL values or high refined meshes in order to produce stable and accurate results.

Applying the  $\mathcal{BDF3}$  method (6.27) to the several MS systems of ordinary differential equation in time, developed in Chapter 5, means that matrix  $A$  and vector  $f$  of (6.27) assume a different expression in dependence of the method chosen for the spatial discretization of the original continuum system (1.9), among the ones derived in this work. However we can say that in general  $A$  and  $f$  are structured in the same way as was explained for the  $\mathcal{CN}$  integration method.

In addition, this class of methods are implicit multistep schemes, which means that a Newton iteration process of the kind described before need

to be accomplished on each step in order to compute the solution. Again this is the most computationally expensive part of the solution process. A variable  $w^k$  which change his value at each iteration  $k$  has to be introduced. The cycle starts with  $w^1 = u^n$  and finishes at  $w^{end} = u^{n+1}$  when the same requirements sets on the residual increment or value are satisfied. At each step we have, thus, to solve for each generic internal node  $i$  the problem:

$$\frac{11}{6}A_{ij}(w_j - u_j^n) - \frac{7}{6}A_{ij}\Delta u_j^n + \frac{2}{6}A_{ij}\Delta u_j^{n-1} = hf_i(w_j) \quad (6.28)$$

with a proper Newton form of the kind of (6.4), where  $r_i$  is now:

$$r_i = \frac{11}{6}A_{ij}(w_j - u_j^n) - \frac{7}{6}A_{ij}\Delta u_j^n + \frac{2}{6}A_{ij}\Delta u_j^{n-1} - hf_i(w_j) \quad (6.29)$$

In particular the  $\mathcal{BDF3}$  method is here used to integrate in time the semi-discretized schemes developed in Chapter 5, as an alternative to the previous Crank-Nicolson method, in order to investigate the possibility to obtain a higher order of the numerical convergence to the exact solution. We have shown, in fact, in the previous section, that  $\mathcal{CN}$  is a second order method in time; using the  $\mathcal{BDF3}$ , a third order method is searching. A comparison between the convergence properties of these two integration methods is illustrated in the following, after a brief discussion about numerical implementation of the proposed test.



# Chapter 7

## Numerical implementation

In Chapters 5 and 6 respectively the spatial and temporal discretization of the Madsen and Sørensen system has been developed in order to obtain a fully discrete system which could be implemented and solved in a proper way. In Chapter 5, in particular, we have shown that the solution of the Madsen and Sørensen system using the finite element approach necessary needs the rewriting of the system in a low-order form introducing auxiliary variables, resulting in additional equations to be solved. Mass-lumping technique is used on this adding equations in order to reduce the rise in the computational cost of the solution.

The extended system so found can be characterized by three PDEs with three variables (two physical and one auxiliary), as in the case of the Galerkin method, or by four PDEs with four (two physical and two auxiliary), as in the central  $\mathcal{RD}$  or in all the stabilized upwind methods proposed. The solution of the extended system can be accomplished directly including the auxiliary variable  $w$  into the vector of the variables of the system  $U = [\eta \ q \ w]^T$ . In such a way the coupled system can be solved altogether, at each iteration step of the Newton iteration process, computing the value of  $W^k = [\eta^k \ q^k \ w^k]^T$  and upgrading the value of the solution at the end of the cycle as  $W^{end} = U^{n+1}$ . This choice is not the cheapest one.

The other possibility, which is here implement, is to separately compute the value of the auxiliary variables of the problems, at each iteration step  $k$ , using the values of the physical variables computed at the previous step  $\eta^{k-1}$ ,  $q^{k-1}$ . The original system of two equations can thus be solved in a coupled way using the just computed value of  $w^k$  in order to obtain  $\eta^k$  and  $q^k$ . The procedure ends when the convergence requirements are satisfied, so the value of the solution can be finally upgraded and the time step incremented.

In the following sections we are going to illustrate the environment of the elements implemented in order to realize the numerical tests for the validation of the schemes proposed. A method for the computation of the exact form of the soliton wave for the Madsen and Sørensen model, which is going to be used for the test cases, will be shown.

Then a brief treatment of the boundaries conditions setted on the test cases proposed in this work will follow, before concluding the chapter with an illustration of a method of generating waves internally to the domain, useful to compute the solutions in tests where the input wave is constant.

## 7.1 Soliton Wave Generator

A unique standard definition of soliton is difficult to be found in literature, but we refer to it as a self-reinforcing solitary wave that maintains its shape while it travels at its constant speed, caused by the cancellation of nonlinear and dispersive effect in the medium where it propagates. Three are the main properties of a soliton wave: it has a permanent form, is localized within a certain region (such that it decades or approximates a constant value very far in the domain) and can strongly interact with other solitons emerging from the collisions unchanged, except for a small phase shift.

In the context of the NSWE system, the analytical form of a soliton wave is well known [29] and described by the formula:

$$\eta(\zeta) = a \operatorname{sech} \left( \sqrt{\frac{3a}{4h_0^3}} (\zeta - X_0)^2 \right) \quad (7.1)$$

$$u(\zeta) = \sqrt{\frac{g}{h_0}} \eta(\zeta) \quad (7.2)$$

where  $a$  is the soliton wave amplitude,  $h_0$  the level of the bathymetry,  $X_0$  is the coordinate of the maximum and  $\zeta = x - Ct$  is a moving coordinate with the celerity  $C$  of the wave.

The soliton shape described by (7.1) cannot be applied in the MS context. The study of the propagation of this kind of solution performed using the Madsen and Sørensen model shows a not complete agreement of the obtained

wave with the three general features of a soliton wave. In particular a sensible amplitude decreasing accompanied by a dispersive wake can be easily observed. These effects can be reduced through a mesh refinement, but this strongly increases the computational coast of the simulations without deleting completely the problem.

The exact solution of a soliton wave shape for MS system of equations needs to be found in order to achieve better test results. Following what is performed in (as in [21], we first take into account the MS system making the hypothesis of constant bathymetry  $h_0$ :

$$\partial_t \eta + \partial_x q = 0 \tag{7.3}$$

$$\partial_t q - Bh_0^2 \partial_{x^2 t} q + \partial_x(uq) + g(h_0 + \eta) \partial_x \eta - \beta gh_0^3 \partial_{x^3} \eta = 0$$

$h_0$  is a data of the problem and contributes, together with the will value of the amplitude  $a$  to uniquely define the shape of the soliton which will result from this procedure. Assuming that, for a solitary wave solution, it is possible to write:

$$\eta = \eta(x - Ct) = \eta(\zeta) \qquad q = q(x - Ct) = q(\zeta) \tag{7.4}$$

being  $\zeta = x - Ct$  a moving coordinate with  $C$  the celerity of the wave. Using (7.4), the system (7.3) reduces to:

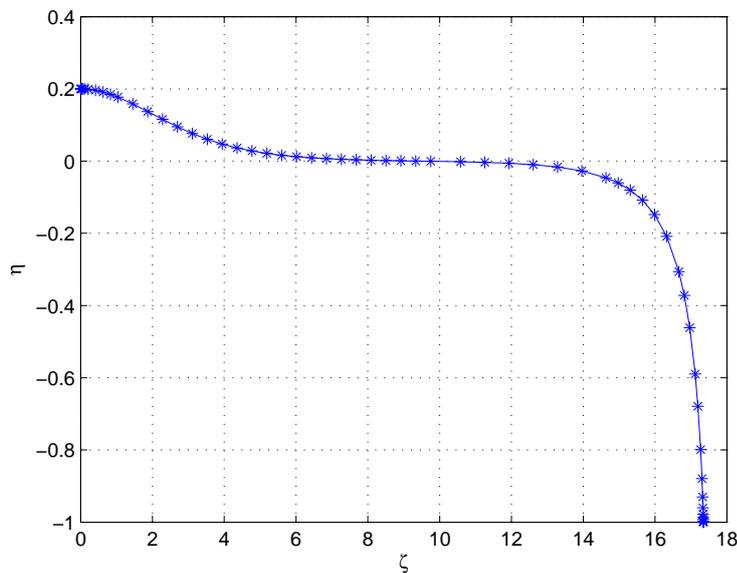
$$\begin{aligned} -C\eta' + q' &= 0 \\ -Cq' + CBh_0^2 q''' + \left( \frac{q^2}{h_0 + \eta} \right)' + gh_0 \eta' + \frac{g}{2} (\eta^2)' - \beta gh_0^3 \eta''' &= 0 \end{aligned} \tag{7.5}$$

indicating with the primes the derivatives with respect to the ordinate  $\zeta$ . (7.5) is now a system of ODEs which can now be integrated.

$$\begin{aligned} -C\eta + q &= 0 \\ -Cq + CBh_0^2 q'' + \frac{q^2}{h_0 + \eta} + gh_0 \eta + \frac{g}{2} \eta^2 - \beta gh_0^3 \eta'' &= 0 \end{aligned} \tag{7.6}$$

Substituting (7.6) into (7.6), a second order ordinary differential equation for the flux  $q$  can be found. Setting the initial conditions  $q(0) = q_{max}$  and  $q'(0) = 0$ , being  $q_{max} = C \cdot a$  through the (7.6), with  $a$  the desired wave amplitude, the second order differential equation is solved numerically in Matlab<sup>®</sup> by means of the function *ode113*. Once obtained the solution  $q$  of the (7.6), the wave profile  $\eta$  of the soliton is then recollected via (7.6).

Since  $q$  and  $\eta$  derives from a numerical integration of the ODEs, we do not have their analytical expressions but only vectors of values which have to be interpolated over the nodes of the mesh in order to attribute to any degree of freedom of the grid a value of the solution. A particular attention has to be paid on the tails of the solution returned by the Matlab<sup>®</sup> integrator. In order to avoid a leakage error, it is required to the solution and its derivative to be null and smooth at the tails of the soliton. This kind of error, in fact, affects not only the quality of the solution but also the possibility to reconstruct the correct convergence order of the schemes, due to the fact that it cannot be decreased by using standard techniques, e.g. mesh refinement, time step reduction...

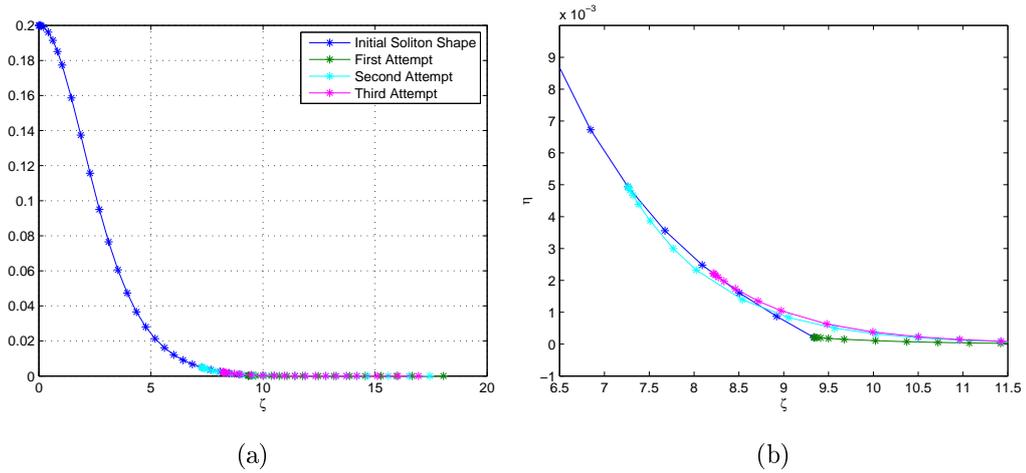


**Figure 7.1:** Numerical solution of the integration of the second order differential equation of  $q$  using the Matlab<sup>®</sup> solver *ode113*.

The solution returned by the Matlab<sup>®</sup> solver in general reproduces only one half of the soliton shape and is composed by a physical and a non-physical

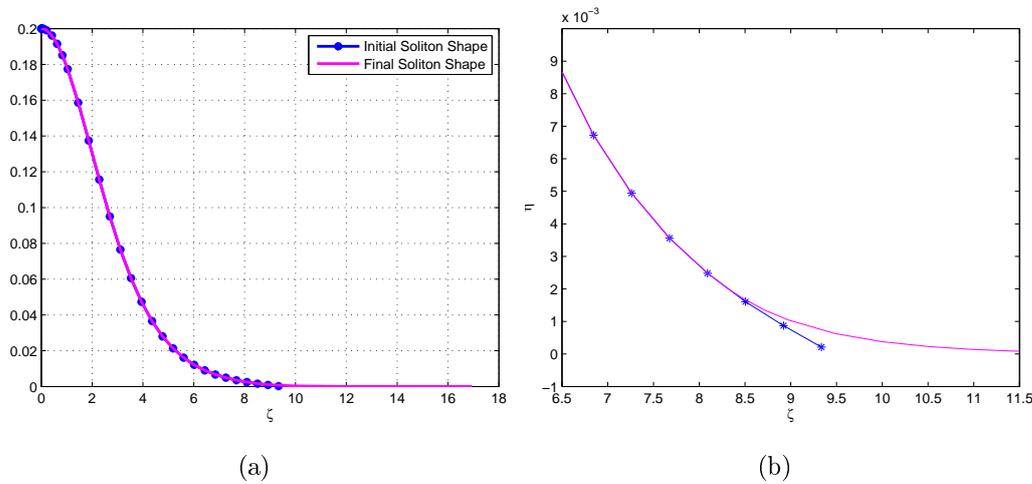
part (see figure 7.1). Necessarily only the first interesting fraction, where the value of the solution is positive, has to be taken into account, cutting the rest of the points computed and thus being forced to treat the tail edge in order to give to the final shape a continuous trend to zero. In this work, the idea is to reconstruct a different shape for the tail by solving the integration of a new first order ordinary differential equation, setting as initial condition the value of a point of the tail previously computed. The first order ODE we are going to integrate takes origin directly from the second order ones previously described and has the form [21]:

$$\frac{1}{2}(q')^2 \left( -CB + \frac{\beta gh_0}{C} \right) = - \left( \frac{C^2}{h_0} \right) q + \left( \frac{g}{2Ch_0} \right) q^2 + \left( \frac{g}{6C^2h_0^2} \right) q^3 + C^3 \ln \left( \frac{Ch_0 + q}{Ch_0} \right) \quad (7.7)$$



**Figure 7.2:** (a): attempts of reconstruction of the tail of the soliton shape by means of the numerical integration of the first order equation (7.7); (b): close-up of the interest region.

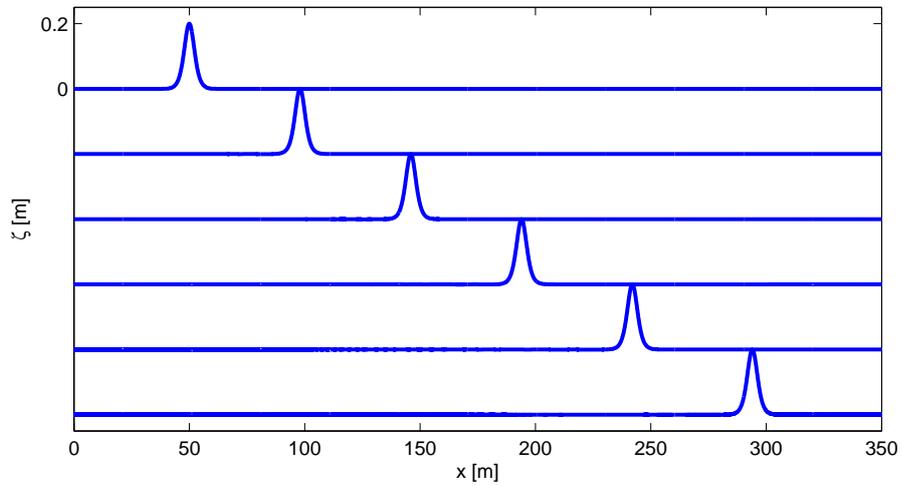
The choice of the point of the tail the integration of (7.7) has to start from is not trivial. Figure 7.3 shows how it exist an optimum point which produce an extension to the tail with good enough continuity properties. Since an analytical method to find this point does not exist, a good compromise was reached after several attempts. In figure 7.2 the final results of the half of the soliton shape, computed with this process, and merging the blue initial solution with the magenta tail of figure 7.3, is shown.



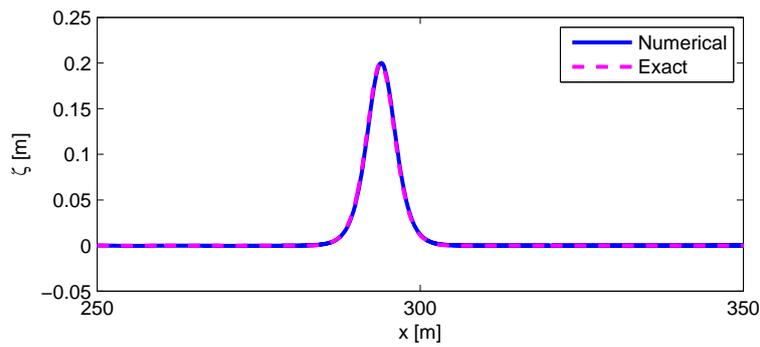
**Figure 7.3:** Optimal solution found for the soliton tail reconstruction and comparison between the final and the initial configuration (a); close-up of the interest region (b).

To test the goodness of the proposed solution, the propagation of a solitary wave of amplitude  $a = 0.2$  m in still water of constant depth  $h_0 = 1$  m has been calculated using the above method to generate the soliton shape and the MS equations to solve the whole domain. The results of the test (see figure 7.4) show a good amplitude and shape conservation without any sensible dispersion effect during the propagation. In figure 7.4(a) the profiles, computed after constant time steps, are deliberately stacked with the vertical spacing of 0.2 m to show clearly the non-changing amplitude of the propagating wave. Figure 7.4(b) shows a comparison between the numerically predicted and the analytical wave profiles plotted against the  $x$  axis at the final time of the simulation  $t = 69.6$  s. The two profiles coincide, confirming the ability of the numerical model to propagate information at the physically correct speed. Problems which arise during the propagation of the soliton often involves trailing edge dispersion and consequently reduction in wave height and celerity. Those phenomena mostly originate from approximation errors in the governing equations with respect to nonlinearity and dispersion or a combination of numerical errors from poorly balanced schemes and truncation of numerical approximation.

All the following test cases, performed in the other sections of this work, will be computed using the soliton shape generator process here developed, which simply solves the shape of the soliton wave of amplitude  $a$ , fixed by the user, able to travel in a stable way over a constant bathymetry configuration  $h_0$ , data of the test case.



(a)



(b)

**Figure 7.4:** (a): evolution of the exact  $a = 0.2$  m solitary wave; (b): comparison between the exact and the numerically computed solitary wave plotted against the spatial domain at the final time of the simulation.

## 7.2 Boundary conditions

In this work, two kinds of boundary conditions are implemented in order to compute the test cases proposed: the periodic and the outflow boundaries.

**Periodic Boundary Conditions :** When these kinds of boundary conditions are implemented, the informations which arrive to an extreme of the

domain are reintroduced on the other side, as if the solution we are looking at were periodic with a period coincident with the domain of computation, assumed to have non-reflecting boundaries.

The way to implement this is quite simple because it is enough to set a series of equality conditions on the values of the solution and its derivatives between left and right boundaries of the domain.

More in detail, imagining to have subdivided the domain into a set of  $N+1$  nodes, node 1 and node  $N+1$ , which are the extremes of the domain, have to be considered as the same point, surrounded by the previous node  $N$  on the left and by the following node 2 on the right. In according with this assumption, the schemes expressions referred to the extreme nodes of the domain result identically formulated.

In addition, depending on the stencil extension of the scheme used, there will be one or more nodes set near the boundaries such that, if we write the equation of the scheme referred to these nodes themselves, there will be at least one index in the formula which does not correspond to any real degree of freedom of the domain. The solution to this problem provided by the implementation of the periodic boundary conditions consists of ideally merging the two boundaries so that the missing nodes of the formula appear to be on the other side of the domain.

The drawback of this procedure is that the non-zero elements of the matrices of the schemes are no more concentrated along the main diagonals of each matrix or of its sub-blocks, but appear to be much sparser.

**Outflow Boundary Conditions :** The aim of an outflow boundary is to ideally let exit from the domain the informations which trough their propagations arrive to the boundary. In practice they are realized by a non-physical restriction of the spatial domain to a suitably smaller one where practical computations can be performed. It is fundamental for these boundaries to minimize non-physical reflection of the information back into the computational domain.

In this work, a viscous damping layer, termed *sponge layer* [42], is introduced near the boundary in the form of a viscous term which increases in magnitude closer to the outflow boundary. This viscous term is added to the free surface equation in the form illustrated below:

$$\begin{aligned}
 \partial_t \eta + \partial_x q &= \nu \partial_{x^2} \eta \\
 \partial_t q - Bh^2 \partial_{x^2 t} q - \frac{1}{3} h \partial_x h \partial_{xt} q + \partial_x(uq) + gH \partial_x \eta + & \quad (7.8) \\
 - \beta gh^3 \partial_{x^3} \eta - 2\beta gh^2 \partial_x h \partial_{x^2} \eta &= \nu \partial_{x^2} q
 \end{aligned}$$

being  $\nu = 0$  everywhere but in  $x \in [X_s, X_2]$ , being  $X_2$  the extreme of the domain, where it is defined as:

$$\nu = n_1 \frac{e^{\left(\frac{x - X_s}{X_2 - X_s}\right)^{n_2}} - 1}{e^1 - 1} \quad (7.9)$$

Here  $n_1$  and  $n_2$  are constants used to fine tune the amount of viscosity, which is problem dependent and has to be determined by attempts on each test case so as to effectively damp the outgoing waves without allowing significant reflections. The width of the sponge layer is another parameter which has to be varied in order to provide sufficient damping for the cases considered. Since the presence of a sponge layer adds a new term to the each of the equations of the MS system. It is important to underline that, although formally modifying the expression of the residual, these terms are not take into account when the residual of the equation is used in order to compute the stabilization upwind term.

### 7.3 Internal Wave Generation

For the test cases in which the solution of the interaction between a train of periodic input waves with a variable bathymetry wants to be compute, the implementation of an oscillating wave-making internal source function is required in order to produce such train of waves. To obtain the desired oscillation signal in the wave generating area, following [42], an analogy with a time varying sea bed is drawn, and a new term is added to the  $\eta$ -equation that simulates the lowest order effects of a time varying depth profile. The free surface equation of the MS system, thus, becomes:

$$\partial_t \eta + \partial_t h_I + \partial_x q = 0 \quad (7.10)$$

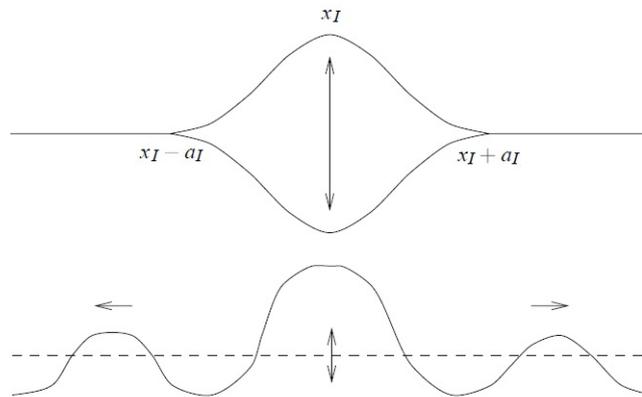
where the form of the sea bed variation  $h_I$  is taken to be:

$$h_I(x, t) = 2C f_I(x) a \cos(\omega t) \quad (7.11)$$

where  $a$ ,  $C$  and  $\omega$  are respectively the amplitude, the celerity and the frequency of the required waves and the spatial function  $f_I(x)$  is given by, still following [42]:

$$f_I(x) = \frac{b_I}{\sqrt{\pi}} e^{-b_I(x-x_I)^2} \quad (7.12)$$

These expressions simulate an undulating Gaussian hill centered at the position  $x = x_I$ . The width of the generation region results strongly dependent on the value attributed to the parameter  $b_I$  which, in addition, influence also the amplitude of the oscillation. In particular  $b_I$  has to be chosen such that the function  $f_I(x)$  is effectively zero at the interfaces between the generation region and the physical test domain:  $x = x_I \pm a_I$  (see the schematic representation in figure 7.5). Function  $f_I(x)$  is also normalized, such that its integral is one over this region.



**Figure 7.5:** sketch of the internal wave generation process.

The wave generator just described does not reproduce the correct expected train of periodic waves, but a certain transient is present from the beginning of the oscillation till the generation of a real periodic train of waves. The

generator produces both a left and a right travelling wave (as represented in figure 7.5) well reproducing the period and wavelength of the waves even if an unclear and inexplicated non-correct generation of the amplitudes is registered. The several parameters which belong to the expressions written before have thus to be set accurately in a proper way for each test application.

Exactly in the same way as for the sponge layer added term, the term  $\partial_t h_I(x, t)$  changes the expression of the residual of the continuity equation of the system. Even if this term will be treated in the discretization procedure as the similar  $\partial_t \eta$ , it will not be taken into account for the stabilization term computation.



# Chapter 8

## Numerical Tests and Results

In this final chapter of the work, the properties of the schemes developed and analysed all along this report are going to be finally tested and verified. A convergence study of the methods will be accomplished in order to verify the schemes precision in space-time description for a sequence of refined grids. The ability of these methods to well reproduce nonlinear and dispersion property of the MS model is going to be tested too, computing the solutions of the collision of two solitary waves and of the propagation of a soliton over a shelf in the bathymetry.

Finally, at the end of the chapter, the methods of integration proposed are validated using the experimental data of a very common test originally proposed by [39]: the propagation of a periodic wave over a submerged bar.

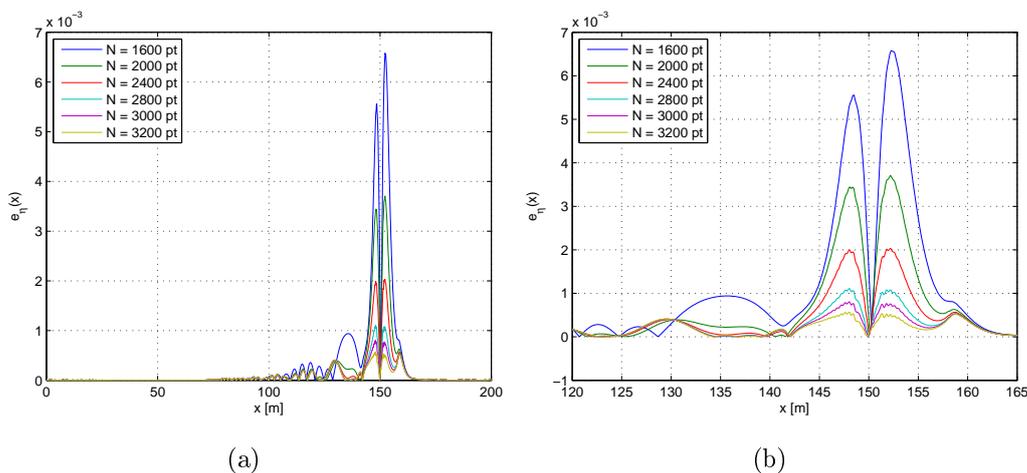
### 8.1 Convergence Analysis

As an indication of the accuracy and efficiency of the proposed numerical approaches, we consider the simple propagation of a solitary wave of amplitude  $a = 0.2$  m over an undisturbed depth  $h = 1$  m. The wave is initially centred at  $x = 50$  m and the spatial domain considered is  $x \in [0, 200]$  m. The soliton shape, realized via the procedure described in Section 7.1, is used as the initial condition and the error  $e(\eta)$  is computed by comparing the resulted solution after  $t = 30$  s with the same initial shape translated into the spatial domain of the quantity  $C \cdot t$ .

The following figure 8.1 shows the form of the error function  $e_\eta(x)$ , measured at the final time of the simulation, when the theoretical solution states that

the soliton should result centred at the spatial coordinate  $x = 150\text{m}$ . In the figure, the effect of the refinement of the grid on this error functions is shown and represented by the different curves, obtained increasing the number of the points used to discretize the spatial domain.

The error is concentrated in the representation of the exact shape of the propagated soliton wave, but other non-trivial contributions come from the existence of a long dispersive wake. The mesh refinement produces a more accurate description of the soliton, demonstrated by the strong decrease of the error in the region characterized by its presence. However the error appears to be less influenced by the grid step width in the wake zone and in the region of the tails of the soliton, especially the advanced one. In such regions of the domain,  $e_\eta(x)$  does not decrease any more under a certain limit even if the mesh is strongly refined. For the wake, this is due to the fact that the dispersive wake is produced directly by the model applied to solve the problem and is only partially due to the numerical integration. Regarding the tails of the soliton, instead, the problem which arises is the one already discussed in the previous Section 7.1. The error, in fact is due to the particular form assumed by the soliton and to the not analytical solution adopted to give to his shape a continuous trend to zero. This error is, thus, linked to the particular form of the input wave given to the problem and cannot be reduced.



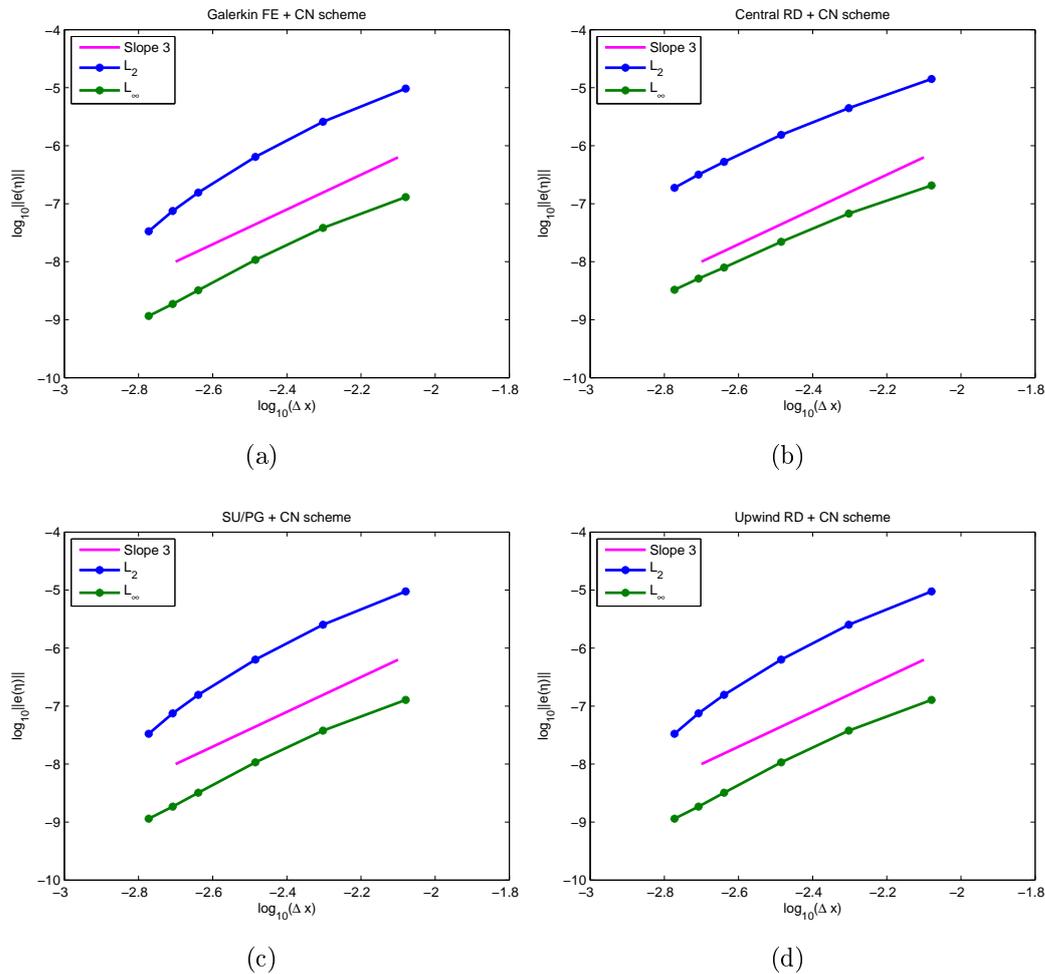
**Figure 8.1:** Spatial distribution of the error function  $e_\eta(x)$  over the domain, parametrised on the growing number of nodes of used for the discretisation (a) and close-up of the region of major contribution to the global error value (b).

Since a study of the convergence order of the schemes is wanted to be computed, it is important to understand this behaviours because, over a certain value of the number of the points  $N$  used for the discretization, the contributes to the global error, which comes from this non-numerical effects, become of the same order of magnitude or even greater with respect to the others. From the moment in which this happens the order of convergence of the scheme computed becomes not representative of the real order of the scheme.

In order to perform a convergence study of the schemes proposed and verify the schemes precision in the space-time description for a sequence of refined grids, the major requirements is to satisfy a consistency refinement property [1]. This property requires the maximum distance across the grid cells to decrease consistently with an increase of the total number of grid data points. Here, the points of the mesh are equispaced and the mesh refinement is accomplished by increasing the number of degrees of freedom from  $N = 1600$  to  $N = 3200$ .

In figure 8.2,  $L_2$  and  $L_\infty$  norms of the error and the asymptotic orders of convergence are presented for the four central Galerkin, central  $\mathcal{RD}$ ,  $\mathcal{SU}/\mathcal{PG}$  and  $\mathcal{U}/\mathcal{RD}$  schemes integrated in time with the Crank-Nicolson method. It can be seen that the schemes do not originate instabilities, confirming the good property of the A-stable  $\mathcal{CN}$  method, and the asymptotic order obtained for all the schemes is closed to the third and even better.

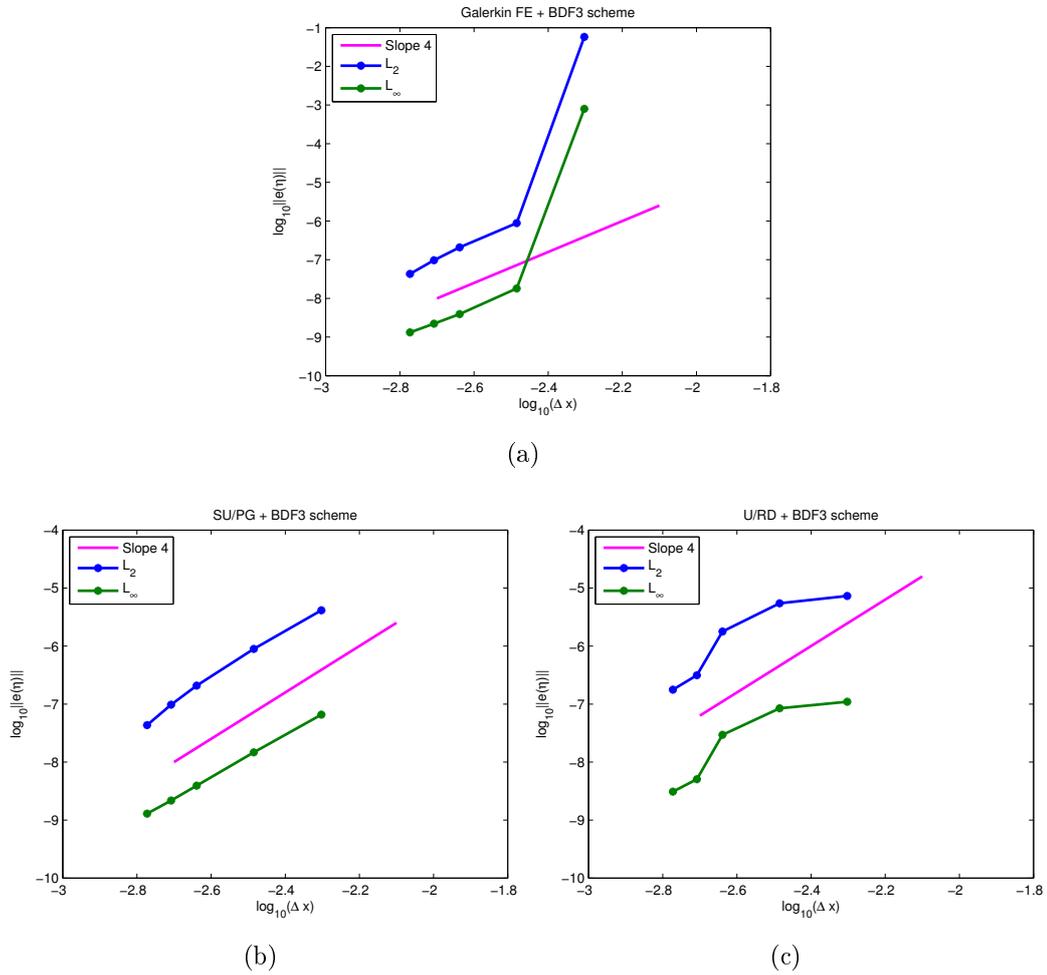
When the  $\mathcal{BDF3}$  method is used to integrate in time the same schemes, instead of the  $\mathcal{CN}$  method, the central Galerkin and residual distribution schemes presents strong oscillations in the solutions for meshes not sufficiently refined and the solution can be computed only with the stabilized  $\mathcal{SU}/\mathcal{PG}$  and  $\mathcal{U}/\mathcal{RD}$  methods. In particular, this behaviour can be observed in figure 8.3(a) were the trend of convergence of the central Galerkin finite element scheme is shown. The figure clearly shows the high error levels obtained on coarse grids, due to the presence of large numerical oscillations. These instabilities disappear on fined meshes where a fourth order slope is recovered. It is quite likely that these instabilities are related to the lack of A-stability of the  $\mathcal{BDF3}$  method which shows for larger timesteps when using the Galerkin scheme in space. In the picture, the trend assumes a very high slope at the beginning before reaching and maintaining a kind of fourth order convergence slope. This is due to the fact that the refinement of the mesh planes the numerical oscillations which characterized the solution in coarse grids; thus, the real convergence order of the scheme appears only



**Figure 8.2:** Convergence curves computed for the schemes integrated in time with the Crank-Nicolson method.

after the effects of these instabilities are cancelled. In addition, it was found impossible for the centred  $\mathcal{RD}$  scheme of Section 5.3 to reach convergence when  $\mathcal{BDF3}$  time integration method is applied; in fact, the graphic of its trend of convergence is not present between the others of figure 8.3.

The  $\mathcal{SU}/\mathcal{PG}$ , instead, shows a very nice and monotone convergence curve of slope very close to four, while the  $\mathcal{U}/\mathcal{RD}$  scheme yields a poorer, yet monotone, grid convergence with averaged slope close to three. In this sense, the  $\mathcal{SU}/\mathcal{PG}$  appears to be the most robust and, hence, the most promising method with respect to the other proposed in this work.



**Figure 8.3:** Convergence curves computed for the schemes integrated in time with the  $BDF3$  method.

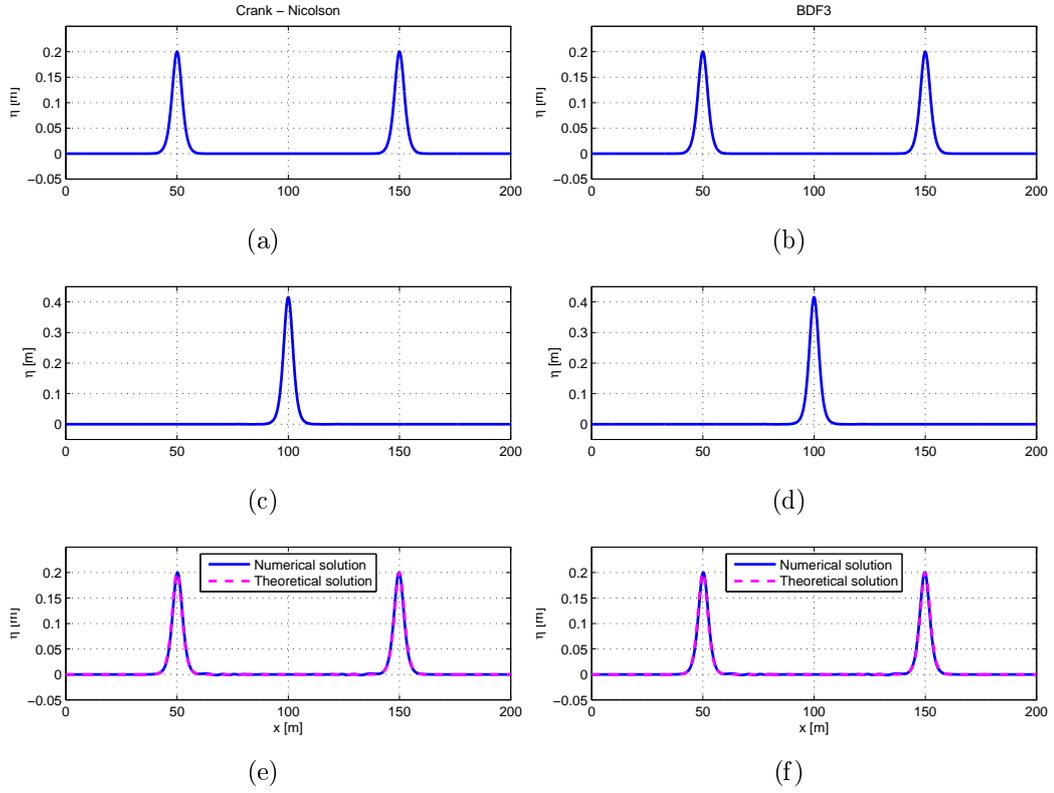
The analysis of the convergence properties of the schemes showed that effectively an higher order of convergence can be reached by integrating in time the space-discretized schemes of Chapter 5 using a  $BDF3$  method, instead of a  $CN$  one. In this procedure, the real advantage of having stabilized the centred schemes becomes more evident, since less refined mesh can here be used to solve the propagation problem without the arise of oscillations which make the solution inaccurate, and an high order scheme can be implement.

## 8.2 Head-on Collision of Two Solitary Waves

A common test for the Boussinesq-type and non-hydrostatic models is the simulation of the interaction of two identical solitary waves propagating in opposite directions. By this numerical test the ability of the scheme to describe the soliton property, already expressed in Section 7.1, to remain unchanged after an interaction with another soliton, wants to be investigated. The collision of the two waves presents additional challenges to the model by a sudden change of the nonlinear and frequency dispersion characteristics. The computation uses a grid spacing of  $\Delta x = 0.08$  m and a time step of integration of  $\Delta t = 0.015$  s. The domain is assumed to be  $x \in [0, 200]$  m with the initial position of the soliton as represented in figures 8.4(a) and 8.4(b). Both the  $\mathcal{CN}$  and the  $\mathcal{BDF3}$  time integration are implemented for each of the four space-discrete schemes proposed.

Figure 8.4 represents the simulated surface profiles, computed by the integration in time of the  $\mathcal{SU}/\mathcal{PG}$  scheme with both the  $\mathcal{CN}$  and the  $\mathcal{BDF3}$  methods, at the beginning of the test, in the instant in which the two waves experiment the maximum amplitude of the constructive interference, and at the end when the two waves have exchanged their initial positions. Immediately one can observe that no differences can be appreciated between the results given by the two integration methods, which, in addition, show both an optimal agreement with the theoretical solution at the final time of the simulation.

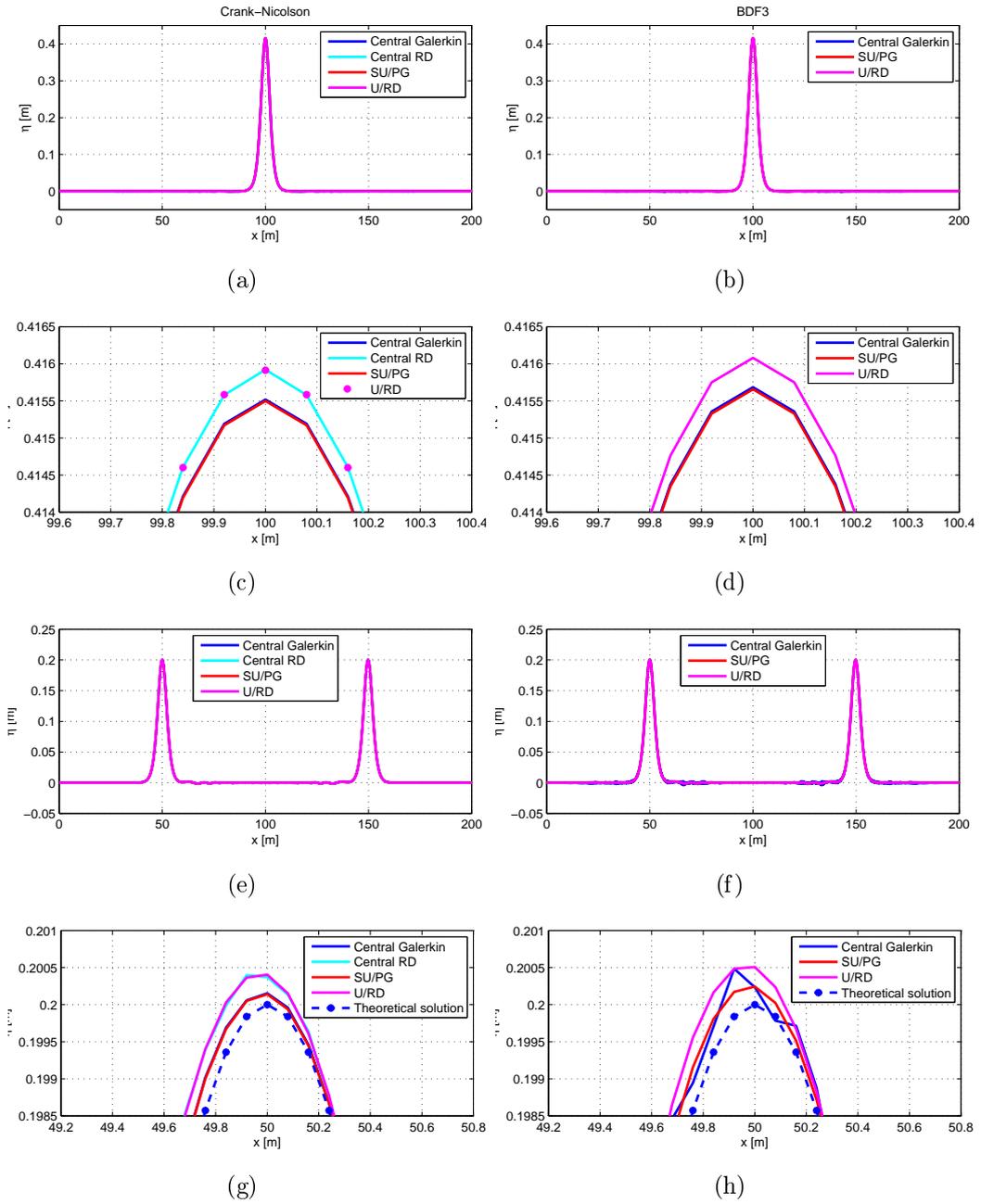
Figures 8.5 illustrate, instead, a comparison between the results computed by the several schemes proposed which gives an almost identical description of the phenomenon. The final solution is characterised by the absence of spurious oscillations and of non-physical dispersion and the final shape of the two waves does not register any relevant variation of their amplitude. Only the residuals schemes show a weak variation of the phase and overestimate of the amplitude with respect to the Galerkin and stabilized Galerkin methods and to the exact analytical solution.



**Figure 8.4:** Free surface profiles of solitary waves with  $a = 0.2$ , propagating in opposite directions in a channel of constant depth, plotted at different times of the simulations and finally comparing with the theoretical result.

Figure 8.5 reports a comparison between the free surface profile computed by all the schemes at the half (figures 8.5(a) and 8.5(b)) and at the end of the simulation (figures 8.5(e) and 8.5(f)). A zoom of the region of the wave crest in both the cases is shown in order to take a better vision on the differences between the several methods. All the schemes appear to solve the problem accurately and with a very close behaviour, but instabilities arise in the centred schemes when they are integrated in time using the  $BDF3$  method. In fact, the central  $\mathcal{RD}$  scheme does not appear completely in the graphics and Galerkin shows some big oscillations which, indeed, are completely absent in the stabilized upwind schemes (see figures 8.5(f) and 8.5(h)). It can also be observed that the residual distribution schemes give exactly the same results, integrating with the  $\mathcal{CN}$  method. This is not a surprise, because is due to the dispersion properties of the two schemes that were shown in Section 5.4 to be identical.

All the schemes computes a very thin increase of the amplitude value of the two solitons after the mutual interaction and show a weak phase error respect to the theoretical solution. A dispersive wake is also present but almost null.



**Figure 8.5:** Overlapping of the several results obtained with the different developed methods and close-up of the peak regions to highlight the difference between the solutions computed.

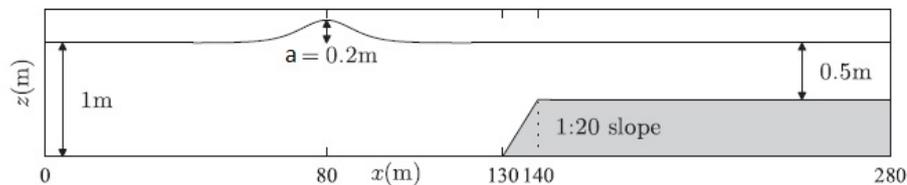
The uniform channel allows examination of the balance between nonlinearity and dispersion without interferences from additional nonlinear effects due to

an irregular bathymetry. The numerical model must handle the equilibrium between amplitude and frequency dispersion to propagate the wave profile at constant shape and speed.

These observations demonstrates the capability of the proposed numerical procedures to handle the strong nonlinearity and dispersion of solitary waves and maintain stable and accurate results along the time of the computation are valid independently from the scheme of time integration chosen.

### 8.3 Propagation over a Shelf

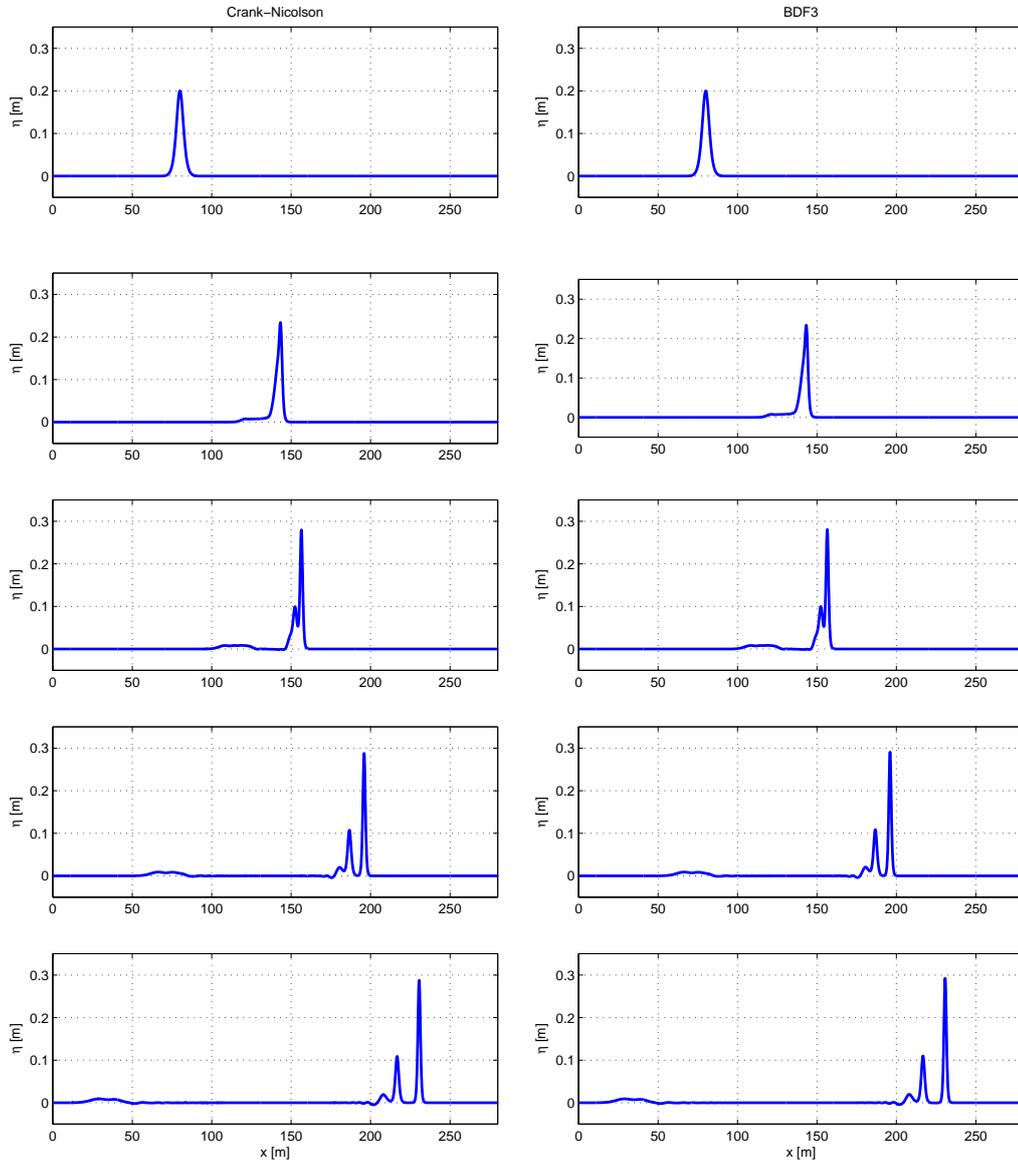
Here we repropose the same test illustrated in Section 1.2 where the differences between the solutions computed by the NSWE and MS systems, due to the different dispersion properties of the two models, wanted to be shown. Hence, a solitary wave of amplitude  $a = 0.2$  m is input in otherwise still water level of depth  $h = 1$  m. The solitary wave starts at the coordinate  $x = 80$  m and then propagates onto a shelf where the still water depth is reduced to  $h = 0.5$  m after a ramp characterised by a slope of 1 : 20 (figure 8.6).



**Figure 8.6:** Sketch of the submerged shelf test.

The soliton propagates unperturbed in the deep water, but when it reaches the shelf, it experiences the reduction of the water depth adjusting to the new level by splitting into several waves. Also a reflected wave is generated, which travels back from the shelf characterised by a very long wavelength and small amplitude. For computing this test case a domain of  $x \in [0, 280]$  m was used, together with a space grid of  $\Delta x = 0.1$  m and a time step of  $\Delta t = 0.029$  s.

The solution computed shows a very good agreement with the same test



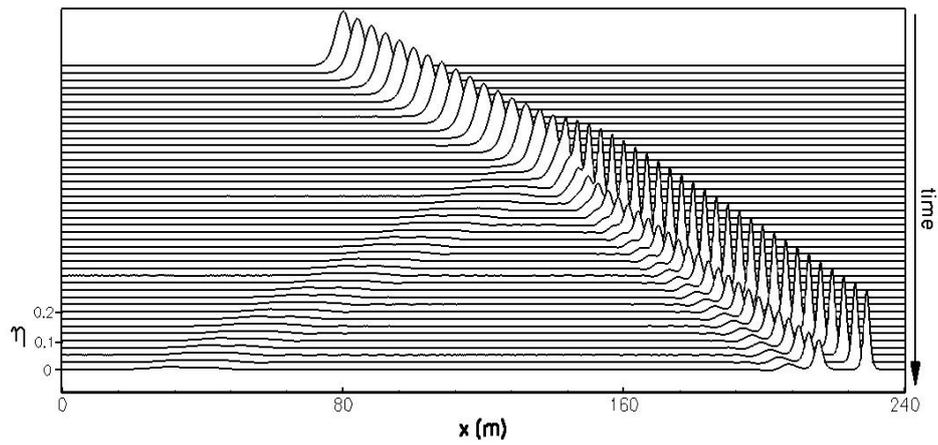
**Figure 8.7:** Time evolution of solution computed for the submerged shelf test. The time step between two consecutive figures is not constant, in order to highlight the most relevant phenomena.

results reported in [21]. In figure 8.7 Crank-Nicolson and  $BDF3$  time integration of the  $SU/PG$  scheme can be compared. Both the solutions well reproduces the wave propagation and transformation overall the domain. The effect of the bathymetry transforms the shape and slope of the soliton, increasing its amplitude and steepness and generating a train of secondary waves of minor amplitude due to the dispersion. The input wave, in fact, which is able to travel unperturbed along the region of constant bathymetry  $h = 1$  m, represents a stable solution for the Madsen-Sørensen system for that configuration of the sea bottom, because it is obtained by directly integrating the system with the procedure explained in Section 7.1. When the bathymetry raise and the new bottom level is set to  $h = 0.5$  m, the system automatically adjust the solution, under the conservation laws constraint, searching for a new stable solution over the new bathymetry configuration. The new soliton wave, which originates from this adjustment, represents a new stable solution, potentially recoverable using the procedure of Section 7.1. In the transient of its formation, a dispersion of the slower frequencies takes place. In Section 1.3.2 we have shown that the dispersion description given by the Madsen and Sørensen system results in a  $k$ -dependent value of the phase velocity  $C(k)$ . Hence, a train of waves takes form from the original solitary wave. Each of them originates from the wake that the one immediate forward left behind during its propagation at its own velocity.

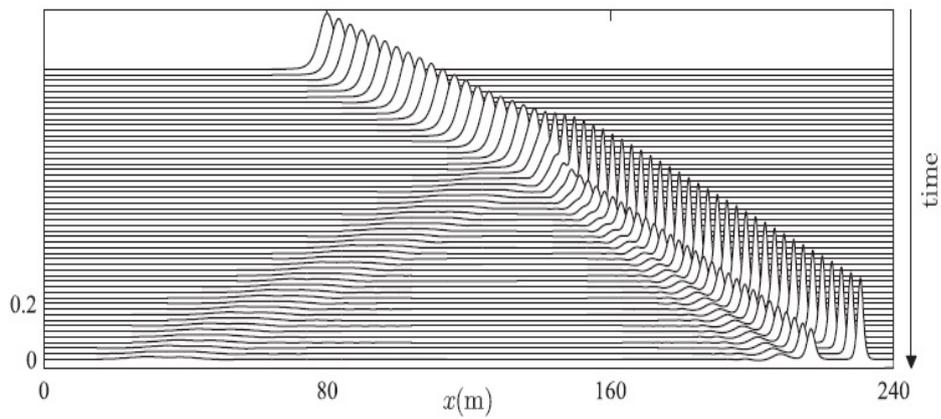
The phenomenon is well represented in the sequential representation of the test in figure 8.7. The solitary wave, in fact, changing its shape becomes a train of different waves. Each element of this dispersive wake travels with an own different speed; in the figure in fact we can see that the distant between the several waves increase, such that a train of stable solitons results in the end.

Figure 8.8 compare the propagation and fragmentation of the solitary wave computed by the  $SU/PG$  scheme with  $CN$  time integration (a) with respect to the test results presented in [21] (b). The agreement between the two solutions is very strong and is a symptom of the goodness of the scheme proposed in the representation of the high dispersive character of this test.

In the above figure 8.9, the results computed by the several schemes proposed in Chapter 3 and integrated in time with the  $CN$  method are compared. All of them give an accurate description of the physical aspect of the phenomenon. However, the dispersion analysis of the schemes described in Chapter 3, and riassumed by the figures 5.4 and 5.5, it is here confirmed by the zoom in figure 8.9(b). In the pictures, in fact, the difference between the curves are principally due to the fact that the several schemes solved the dispersion

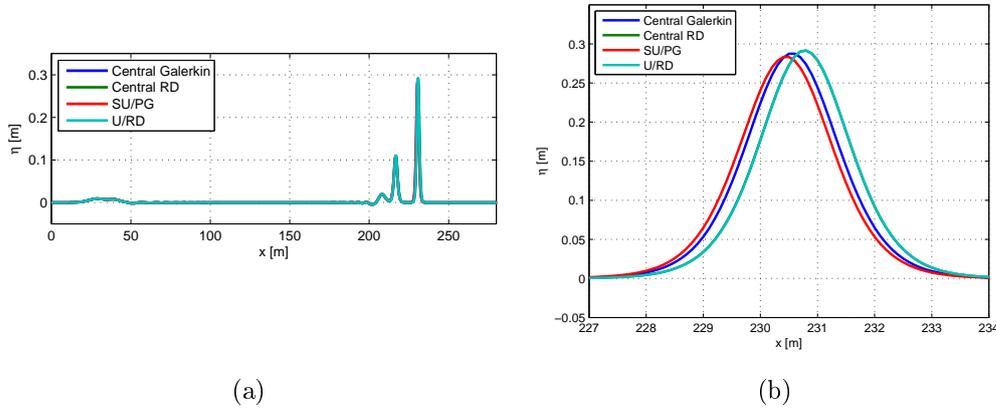


(a)



(b)

**Figure 8.8:** Splitting of a solitary wave propagating over a submerged shelf: (a) numerical computation using  $SU/\mathcal{PG}$  scheme with  $CN$  time integration; (b) test results presented in [21].

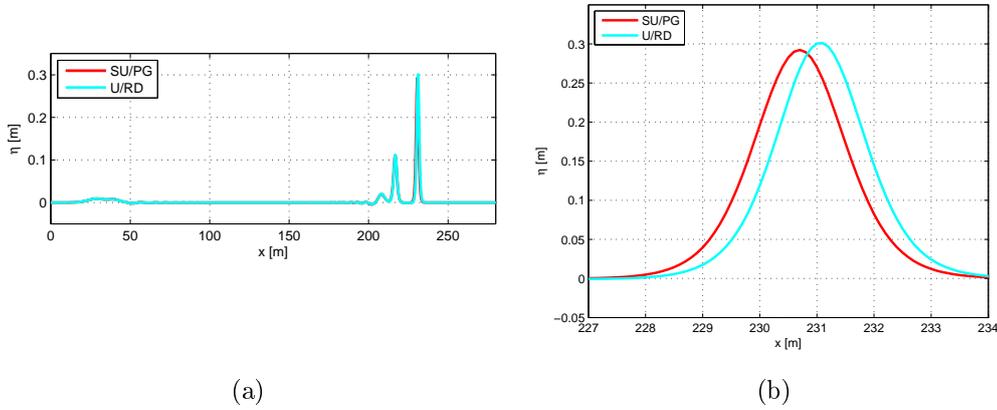


**Figure 8.9:** Overlapping of the several results of the submerged shelf test, obtained integrating the different schemes with the  $\mathcal{CN}$  method (a), and close-up of the peaks regions to highlight the difference between the solutions computed (b).

properties of the MS model with different levels of accuracy. This fact generates the differences in computing the amplitude and celerity of the solitons in the transitory between the two bathymetry levels which create also the error in phase. In particular, it can be seen that the residual distribution schemes give exactly the same result, confirming what was shown in figure 5.4 about the equality of their dispersive properties. Between the Galerkin and the  $SU/PG$  curves, instead, not only the effect of a better dispersive description explains the differences, but also the existence of an artificial dissipation in the  $SU/PG$  method which weakly decrease the soliton amplitude computed.

Figure 8.10, instead, shows the same comparison between the schemes which are here integrated in time using the  $BDF3$  method. The central Galerkin and  $RD$  schemes appear to be not stable for this choice of the grid space and time step of the integration, and that is why in the figure only the curves related to the upwind schemes are shown. The same consideration as before between the dispersion properties of the two schemes can be done. Globally, the results only weakly differ from the ones computed using the other time integration method.

Looking to the strong deformation in amplitude and steepness experienced by the waves in this test case, a more realistic description of the phenomenon should contain a model able to individuate and manage the eventual generation of shocks. This model, in fact, as largely explained in the introduction,



**Figure 8.10:** Overlapping of the several results of the submerged shelf test, obtained integrating the different schemes with the *BDF3* method (a), and close-up of the peaks regions to highlight the difference between the solutions computed (b).

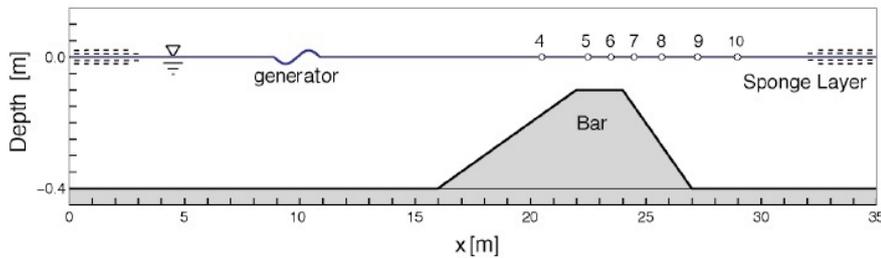
is not able to describe the wave breaking and a switch on the use of the shallow water model should be done for this purpose. The methods developed in this work are robust even in the description of the NSWE model, hence, this could be one of the possible future advancements.

## 8.4 Periodic Wave Propagation over a Submerged Bar

[39] described a series of laboratory experiments to investigate wave propagation and dispersion over a submerged bar. The experimental data collected can be used here to validate the schemes proposed. In particular the experiment having the configuration shown in figure 8.11 is here presented. The test consists in a periodic wave of period  $T = 2.02$  s and amplitude  $a = 0.01$  m which propagates into an initially undisturbed region of depth before reaching a bar of the shape and proportions given in the figure.

The numerical set for this test case needs the use of a periodic internal wave generator, like the one described in Section 7.3 and centred at the coordinate  $x = 10$  m of the domain  $x \in [0, 35]$  m. It is found that the wave generated by this method has a different amplitude compared to the one expected, while

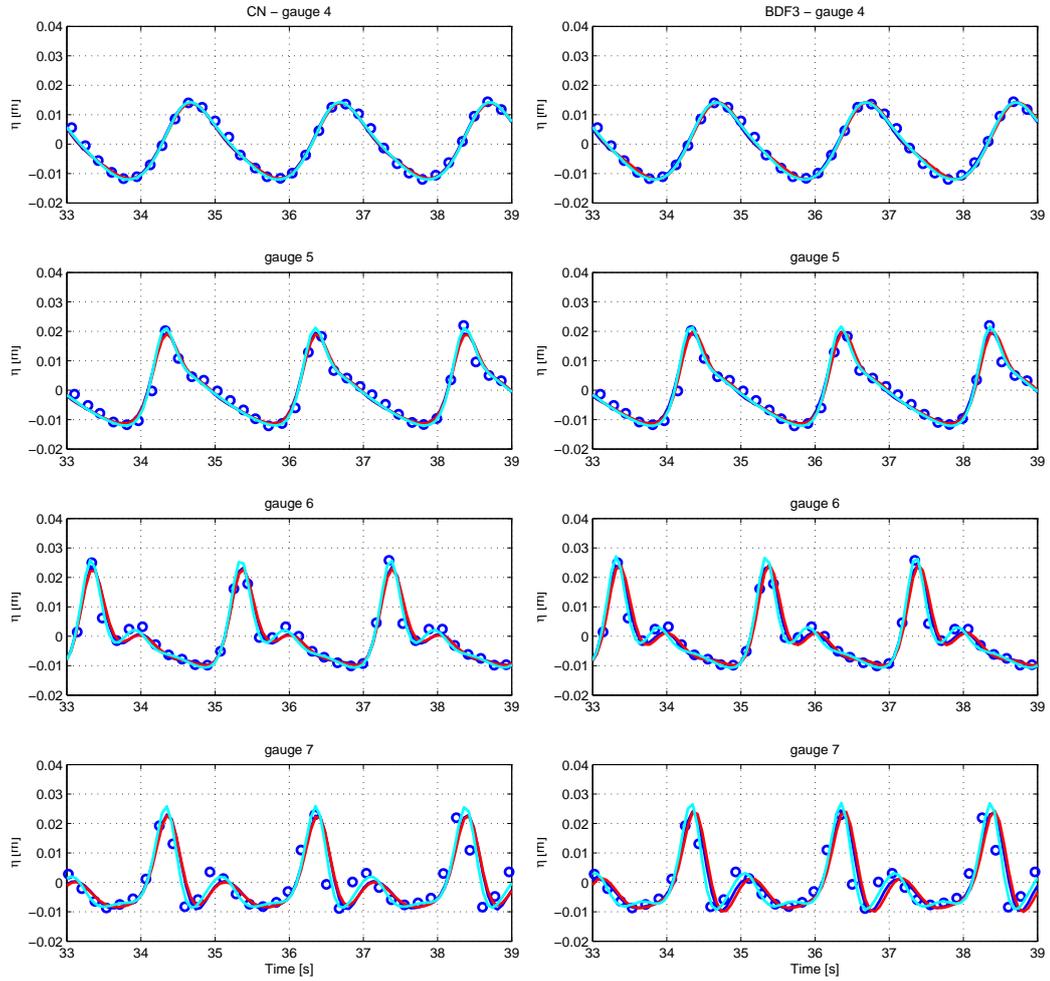
instead the period and wavelength were described correctly. An accurate setting of the parameters of the source function were done, which at the end allowed a more accurate representation of the periodic input wave will. Two sponge layers of the kind described in Section 7.2 are then used at the two boundaries of the domain, extended on the regions  $x \in [0, 3]$  m on the left and  $x \in [32, 35]$  m on the right, to absorb any wave reaching the boundaries.



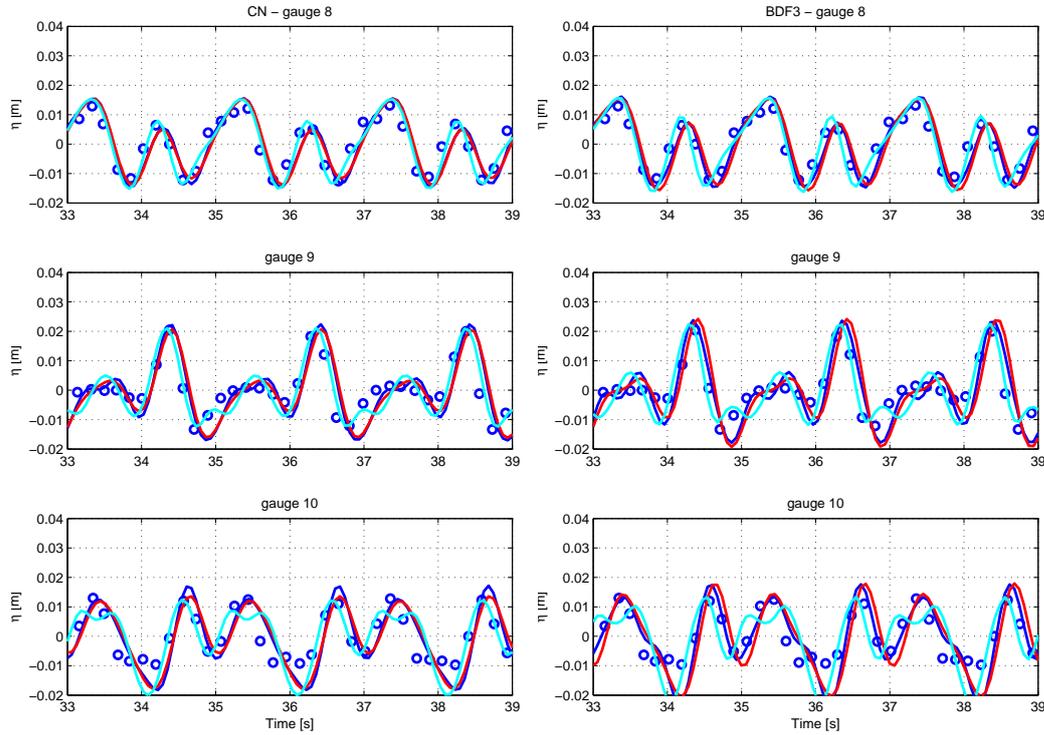
**Figure 8.11:** Sketch of the computational configuration of the numerical test of the propagation over a submerged bar.

The comparison between numerical solution and the experimental data is made up on the time histories of the free surface profile measured with gauges in the positions denoted in figure 8.11 by the numbers 4, 5, 6, 7, 8, and 9 and correspondent to the physical coordinates  $x_4 = 20.5$  m,  $x_5 = 22.5$  m,  $x_6 = 23.5$  m,  $x_7 = 24.5$  m,  $x_8 = 25.7$  m, and  $x_9 = 27.3$  m. The equations are integrated using a grid size  $\Delta x = 0.04$  m and a time step  $\Delta t = 0.0323$ . The computational domain was made long enough in order to avoid any interference with reflected waves that might be generated at the boundaries during the calculus.

Figures 8.12 and 8.13 show the surface elevation of the sinusoidal wave transformation over the submerged bar. Each graphic of the two figures illustrates the trend in time of the surface elevation on the experimental gauge position. In such a way numerical solutions and experimental data are directly comparable on the various sections of the test.



**Figure 8.12:** Time evolution at the gauges coordinates of the surface elevation of sinusoidal wave encountering a submerged bar, with initial amplitude  $a = 0.01$  m and period  $T = 2.02$  s. In figure,  $\circ$  represents the experimental data,  $\text{—}$  is the central Galerkin,  $\text{—}$  is the  $SU/PG$  and  $\text{—}$  is the  $U/RD$  scheme.



**Figure 8.13:** Time evolution at the gauges coordinates of the surface elevation of sinusoidal wave encountering a submerged bar, with initial amplitude  $a = 0.01$  m and period  $T = 2.02$  s. In figure,  $\circ$  represents the experimental data,  $—$  is the central Galerkin,  $—$  is the  $SU/PG$  and  $—$  is the  $U/RD$  scheme.

The schemes performed very well for all the gauges up to the crest of the bar (gauge 8), maintaining a good agreement with the experimental results and showing only a weak error in the phase shift despite the strong nonlinearity; it should be noted that part of the discrepancies between the laboratory data and the models results might be due to the experimental setup. In addition, very thin differences appears between the solutions computed via the use of the  $BDF3$  method, instead of the  $CN$  one and only the central residual distribution scheme results surprisingly unstable for this particular choice of integration parameter for both the cases.

However, as the waves pass the back slope of the bar, some discrepancies with the data appear, especially at the section of gauge 10. This is due to the activation of a numerical phenomenon which decoupled the higher harmonics from the primary longer wave; these are released as free waves propagating generating inaccuracy. The central Galerkin and the upwind residual distribution schemes appears to be more sensible to this kind of problems, while

the  $SU/\mathcal{PG}$  one show a more robust trend with respect to them.

This test case has provided a severe test of the models as nonlinearity has initially steepened the waves on the up-slope and then the increasing depth behind the bar has decomposed the waves into short wave components. This resulted in a rapidly varying profile behind the bar with the exact form depending crucially on the dispersive characteristics of the numerical models analysed. As expected from the theoretical limitations of the present MS model, some discrepancies show up as higher harmonics are released behind the bar. Nevertheless the agreement is still quite reasonable and the discretization schemes proposed for its integration showed the conservation and well representation of the nonlinear and dispersive properties of the original continuous model.



# Conclusions

In this thesis the numerical approximation of nonlinear and dispersive extended Boussinesq wave equations was considered. We have introduced in Chapter 1 the problem of the numerical modelling of the wave propagation in the near-shore zone with the NSW system and it was shown that there is a variety of equation systems, termed Boussinesq-type, which give more accurate solutions to this problem and which can be compared by considering their linear dispersion characteristics. A compromise has to be reached between the theoretical accuracy of the mathematical model and the practicalities of the numerical solution of the systems. In this work a system of extended one-dimensional Boussinesq equations due to Madsen and Sørensen was taken into account: its mathematical properties were compared to the original Boussinesq system and it was shown that the MS equations were significantly more accurate.

Weighted residuals numerical methods to discretize in space nonlinear and dispersive wave equations were introduced in Chapter 2 and then developed and implemented for the simpler cases of linear scalar advection and linear scalar advection with dispersion problems in the following Chapters 3 and 4. In particular, finite difference second and fourth order schemes were developed in order to compare their accuracy and dispersion properties, with respect to the ones of more complex finite element and residual distributions schemes, implemented after. The finite difference schemes was not considered further due to the fact that only those discretization methods which could be later applied to unstructured two-dimensional meshes were of interest here. However, we have shown how those methods which use the same second order reconstruction of the spatial derivatives can actually reach higher truncation error orders, thank to particular coefficients of the mass matrices and can give more accurate approximations of the discretized model dispersion. In particular we have shown that the Galerkin finite element mass matrix gives to its central discretized scheme the best accuracy properties, up to the fourth order in the scalar advection case, using the half of the stencil to accomplish

this, compare to the  $\mathcal{FD4}$  scheme.

We have then introduced a way to stabilize the central Galerkin and  $\mathcal{RD}$  schemes by means of a residual upwind term. We have shown that a particular choice of the weighting functions which pre-multiply this term allows to obtain, in one case, a kind of streamline upwind Petrov-Galerkin scheme and, in the other case, an upwind residual distribution scheme. It has been shown that the use of this technique in order to stabilize a central scheme does not affect the accuracy order of the centred part and that only in the  $\mathcal{SU/PG}$  method some non-physical diffusion is added to the model. The accuracy and the dispersive properties of all these spatially discretized schemes were investigated and shown.

The numerical solution of the full Madsen and Sørensen one-dimensional system was described in Chapter 5. It was shown that the methods presented in the previous chapter could be directly applied to this case, using a technique to decrease the high derivative order of the system, by means of the introduction of one — or in some cases two — auxiliary variables: their role was to assume the reconstructed values of the spatial derivatives of the original variable of the system, increasing the number of variables and the dimension of the system. Even with this manipulation, the schemes maintained the same properties underlined in the previous chapters.

The techniques to spatially discretise the system of equations we dealt with gave as results a series of system of ODEs in time, which were integrated using implicit linear multistep methods of the second ( $\mathcal{CN}$ ) and third ( $\mathcal{BDF3}$ ) order. In Chapter 6 the way to efficiently implement these schemes is shown: a solution adopted to speed-up the calculation consisted of freezing the expression of the Jacobian matrix of the Newton iteration cycle as long as possible, and recomputing it only when the error produced by its approximation exceeds a given tolerance. The convergence study confirmed the expected better convergence trend for schemes using the  $\mathcal{BDF3}$  time integration, however this appeared to be less stable, since in certain configurations of the grid size  $\Delta x$  and time step  $\Delta t$  only the stabilized upwind methods could compute accurately the solution, confirming the goodness of the operated stabilisation. This behaviour is due to the lack of A-stability when the  $\mathcal{BDF3}$  integration method is applied, in fact, all the schemes result stable if integrated by the  $\mathcal{CN}$  method. In this case low CFL numbers are not needed for stability but only for numerical accuracy.

The numerical methods have been finally tested against theoretical and experimental results in Chapter 8. Very good agreements were obtained, confirming that the schemes proposed are able to solve highly nonlinear and dispersive problems with an high order of accuracy. Weak differences emerge among the results of the several used techniques, depending on their degree

of accuracy and especially on their goodness to well reproduce the model dispersion properties. In particular the residual distribution central and up-wind schemes appeared to provide a weakly worse description with respect to the more accurate one given by Galerkin based methods. In addition, the  $SU/\mathcal{P}\mathcal{G}$  scheme better represents the theoretical and experimental shape of the free surface, since it gives the best approximation of the original model dispersion, being also more robust against instabilities which can arise when a higher order time integration method is used, instead of the second order  $\mathcal{CN}$ . For these reasons the  $SU/\mathcal{P}\mathcal{G}$  scheme appears to be the most promising scheme with respect to the ones proposed and, since it is also suitable for the shallow water system solution [38], it is the best candidate to be developed further.

**Future works :** The integration methods proposed for the one-dimensional MS system of extended Boussinesq equations are able to well solve the highly nonlinear and dispersive problem of the wave propagation over a variable bathymetry, giving an accurate description of the free surface water level shape all over the computational one-dimensional domain, but the present preliminary results could originate a wide variety of developments and improvements.

In order to build a comprehensive method, capable of solving all the propagation phenomenon in the near-shore zones, both the management of the moving shoreline and a criteria useful to detect and well solve the breaking wave points have to be included. A possibility is the addition of artificial viscosity to the original model (like in []). However the particular class of developed methods within this work is able to well solve also the NSWE too. In this way, a switch into the shallow water system, once the wave breaking is detected ([30]) would be easy to be implemented and is expected to give better results.

High order models can be treated with the proposed methods, extending the range of applicability of the model here obtained by adding to the MS equations genuinely nonlinear dispersion terms ([7]) and dealing with the Green-Naghdi system of equations.

Other future works could regard the numerical implementation, with the use of the same techniques explained here, of the two-dimensional Madsen and Sørensen problem. This was, in fact, one of the main purposes whereby the choice of the discretization methods was made. In this huge field, many upgrades could be investigated and accomplished. In particular, adaptive dynamic meshes could be implemented in order to follow the wave fronts

and moving shorelines, giving them better reconstructions. By the numerical point of view, higher order finite elements ( $\mathcal{P}^2$  basis functions or hermite hierarchical finite elements or al.) could be used for the numerical approximation of the solution in space; higher order time handling ( $\mathcal{BDF}$  type integration or space-time finite elements) could be tested. Finally, the parallelization of the code could allow the computation over huge domains and large scale applications (e.g. several kilometres off shore).

# Appendix A

## Demonstration of the Equality 2.5

In Chapter 2 we use the integration by parts to transfer the spatial derivative from the discrete value of the flux  $f_h$  to the weighting function of the weak formulation  $\varphi_i$ . This possibility is originated if  $f_h \in C^0$  on each element's closure and  $f_h \in C^1$  on the element itself, using the relation (2.5):

$$\int_{\Omega} \varphi_i \partial_x f_h dx = - \int_{\Omega} f_h \partial_x \varphi_i dx \quad (\text{A.1})$$

In fact, given the particular linear form of the weighting functions on the compact stencil  $[i-1, i+1]$ , represented in figure 2.1, the integral term extended over the whole domain  $\Omega$  becomes an integral over the compact support of the weighting function:

$$- \int_{\Omega} f_h \partial_x \varphi_i dx = - \int_{i-1}^{i+1} f_h \partial_x \varphi_i dx \quad (\text{A.2})$$

However,  $\partial_x \varphi_i$  is a discontinuous function in the interval  $[i-1, i+1]$ , thus, in order to be computed, the integral in equation (A.2) must be partitioned in two over the discontinuity:

$$- \int_{\Omega} f_h \partial_x \varphi_i dx = - \int_{i-1}^i f_h \partial_x \varphi_i dx - \int_i^{i+1} f_h \partial_x \varphi_i dx \quad (\text{A.3})$$

In this way the function  $\partial_x \varphi_i$  is now continuous over each the two intervals

$[i-1, i]$  and  $[i, i+1]$ . Now using the property of the derivation of the product of functions, we have:

$$-\int_{i-1}^i f_h \partial_x \varphi_i dx = -\int_{i-1}^i \partial_x (\varphi_i f_h) dx + \int_{i-1}^i \varphi_i \partial_x f_h dx \quad (\text{A.4})$$

$$-\int_i^{i+1} f_h \partial_x \varphi_i dx = -\int_i^{i+1} \partial_x (\varphi_i f_h) dx + \int_i^{i+1} \varphi_i \partial_x f_h dx \quad (\text{A.5})$$

Recollecting the two expressions and solving some terms, we thus obtain:

$$\begin{aligned} -\int_{i-1}^{i+1} f_h \partial_x \varphi_i dx &= (\varphi_i f_h) |_{i-1} - (\varphi_i f_h) |_i + \int_{i-1}^i \varphi_i \partial_x f_h dx + \\ &+ (\varphi_i f_h) |_i - (\varphi_i f_h) |_{i+1} + \int_i^{i+1} \varphi_i \partial_x f_h dx \end{aligned} \quad (\text{A.6})$$

Simplifying some terms and remembering that the weighting function  $\varphi_i$  is identically null when evaluated in nodes  $i+1$  and  $i-1$  (see figure 2.1), the final expression states:

$$\begin{aligned} -\int_{\Omega} \partial_x \varphi_i f_h dx &= \int_{i-1}^i \varphi_i \partial_x f_h dx + \int_i^{i+1} \varphi_i \partial_x f_h dx = \\ &= \int_{\Omega} \varphi_i \partial_x f_h dx \end{aligned} \quad (\text{A.7})$$

which demonstrate that it is possible, under the initial hypotheses of continuity of the function  $f_h$  and about form of the basis functions  $\varphi_i$ , to exchange the spatial derivative between the two functions through the only change of the sign of the integral.

## Appendix B

# Auxiliary Variables Choice for the Linear Advection-Dispersion Problem

We want to introduce an auxiliary variable  $w$  in order to reduce the order of the dispersive term of the linear advection with dispersion partial differential equation. The possibilities are to reconstruct separately the first or second order spatial derivative contained in this term and then coupled this reconstruction with the modified equation of the problem. In this appendix it will be shown that the two choices lead to the same finite difference approximation of the term  $-\alpha\partial_{xxt}u$  of the equation (4.1) and are, thus, completely equivalent in this case.

We can thus define two auxiliary variables  $w_1$  and  $w_2$  which will assume the expressions:

$$w_{1,i} = \partial_x u_i \tag{B.1}$$

$$w_{2,i} = \partial_{xx} u_i \tag{B.2}$$

The different choices lead to the two different residual formulations of the dispersive term  $-\alpha\partial_{xxt}$  listed below:

$$- \int_{\Omega} \omega_i \alpha \partial_{xt} w_{1,i} dx \quad (\text{B.3})$$

$$- \int_{\Omega} \omega_i \alpha \partial_t w_{2,i} dx \quad (\text{B.4})$$

If  $\mathcal{P}^1$  weighting functions of the kind of the Galerkin finite element scheme are chosen, solving the definite integrals on the spatial domain using the middle point quadrature rule, since a residual distribution form for this terms is searched, the results will be:

$$- \frac{\alpha}{2} \left( \frac{dw_{1,i+1}}{dt} - \frac{dw_{1,i-1}}{dt} \right) \quad (\text{B.5})$$

$$- \frac{\alpha \Delta x}{4} \left( \frac{dw_{2,i+1}}{dt} + 2 \frac{dw_{2,i}}{dt} + \frac{dw_{2,i-1}}{dt} \right) \quad (\text{B.6})$$

Equation (B.1) and (B.2) must be discretized too and, being part of a coupled system with the advection-dispersion equation, they are going to be integrated in the domain used the same middle point rule. This will lead to the expression:

$$\frac{\Delta x}{4} w_{1,i-1} + \frac{\Delta x}{2} w_{1,i} + \frac{\Delta x}{4} w_{1,i+1} = \frac{1}{2\Delta x} (u_{i+1} - u_{i-1}) \quad (\text{B.7})$$

$$\frac{\Delta x}{4} w_{2,i-1} + \frac{\Delta x}{2} w_{2,i} + \frac{\Delta x}{4} w_{2,i+1} = \frac{1}{\Delta x^2} (u_{i+1} - 2u_i + u_{i-1}) \quad (\text{B.8})$$

The mass lumping technique can be use on equations (B.7) and (B.8), obtaining for the auxiliary variables  $w_1$  and  $w_2$  the second order finite difference approximation respectively of the first and second order derivative:

$$w_{1,i} = \frac{1}{2\Delta x} (u_{i+1} - u_{i-1}) \quad (\text{B.9})$$

$$w_{2,i} = \frac{1}{\Delta x^2} (u_{i+1} - 2u_i + u_{i-1}) \quad (\text{B.10})$$

Using (B.9) and (B.10) to compute the values of the auxiliary variables in the  $i+1$ ,  $i$  and  $i-1$  nodes of the mesh, substituting their expression into (B.5) and

(B.6), the equality between the two approaches is demonstrated since both the two expressions approximate the dispersive term of the equation (4.1) through the form:

$$-\frac{\alpha}{4\Delta x} \left( \frac{du_{i+2}}{dt} - 2\frac{du_i}{dt} + \frac{du_{i-2}}{dt} \right) \quad (\text{B.11})$$

which, with respect to the second order finite difference approximation (4.5), enlarges the stencil of the scheme.



# Appendix C

## Space-Discretization of the Linearized MS System

In this Appendix, the several semi-discrete schemes, derived in Chapter 5 for the MS systems of equations and shown in Appendix, are applied on the simpler linearised case under the assumption of constant bathymetry  $h = h_0$ . We have already shown the form of this model in the Section 1.3.2, but for convenience it is rewritten in the following:

$$\begin{cases} \partial_t \eta + h_0 \partial_x u = 0 \\ \partial_t u - B h_0^2 \partial_{xxt} u + g \partial_x \eta - \beta g h_0^2 \partial_{xxx} \eta = 0 \end{cases} \quad (\text{C.1})$$

***FD2*** : The *FD2* scheme is obtained simply applying the second order accuracy finite different approximation of the spatial derivatives present in the two equations of the system. The final expression of the scheme obtained is thus:

$$\begin{aligned}
\frac{d\eta_i}{dt} + \frac{h_0}{2\Delta x} (u_{i+1} - u_{i-1}) &= 0 \\
\frac{du_i}{dt} - \frac{Bh_0^2}{\Delta x^2} \left( \frac{du_{i-1}}{dt} - 2\frac{du_i}{dt} + \frac{du_{i+1}}{dt} \right) &+ \\
+ \frac{g}{2\Delta x} (\eta_{i+1} - \eta_{i-1}) &+ \\
- \beta \frac{gh_0^3}{2\Delta x^2} (-\eta_{i-2} + 2\eta_{i-1} - 2\eta_{i+1} + \eta_{i+2}) &= 0
\end{aligned} \tag{C.2}$$

**$\mathcal{FD4}$**  : Similarly, approximating the spatial derivatives using the fourth order finite difference formulae, instead of the second ones, the  $\mathcal{FD4}$  scheme will result:

$$\begin{aligned}
\frac{d\eta_i}{dt} + \frac{h_0}{12\Delta x} (u_{i-2} - 8u_{i-1} + 8u_{i+1} - u_{i+2}) &= 0 \\
\frac{du_i}{dt} - \frac{Bh_0^2}{12\Delta x^2} \left( -\frac{du_{i-2}}{dt} + 16\frac{du_{i-1}}{dt} - 30\frac{du_i}{dt} + 16\frac{du_{i+1}}{dt} - \frac{du_{i+2}}{dt} \right) &+ \\
+ \frac{g}{12\Delta x} (\eta_{i-2} - 8\eta_{i-1} + 8\eta_{i+1} - \eta_{i+2}) &+ \\
- \beta \frac{gh_0^3}{8\Delta x^2} (\eta_{i-3} - 8\eta_{i-2} + 13\eta_{i-1} - 13\eta_{i+1} + 8\eta_{i+2} - \eta_{i+3}) &= 0
\end{aligned} \tag{C.3}$$

**Galerkin finite element** : In the following the system (C.1) is spatially discretized with the central Galerkin finite element method discussed in Section 5.2. The auxiliary variable  $w$  is the same defined by (5.17) and the scheme results in:

$$\frac{\Delta x}{6} \frac{d\eta_{i-1}}{dt} + \frac{2\Delta x}{3} \frac{d\eta_i}{dt} + \frac{\Delta x}{6} \frac{d\eta_{i+1}}{dt} + \frac{h_0}{2} (u_{i+1} - u_{i-1}) = 0 \tag{C.4}$$

$$\begin{aligned}
\frac{\Delta x}{6} \frac{du_{i-1}}{dt} + \frac{2\Delta x}{3} \frac{du_i}{dt} + \frac{\Delta x}{6} \frac{du_{i+1}}{dt} + \frac{g}{2} (\eta_{i+1} - \eta_{i-1}) &+ \\
- \frac{Bh_0^2}{\Delta x} \left( \frac{du_{i-1}}{dt} - 2\frac{du_i}{dt} + \frac{du_{i+1}}{dt} \right) &+ \\
- \beta \frac{gh_0^2}{2\Delta x^2} (-\eta_{i-2} + 2\eta_{i-1} - 2\eta_{i+1} + \eta_{i+2}) &= 0
\end{aligned} \tag{C.5}$$

**Central Residual Distribution :** In Section 5.3 the development of a central residual distribution scheme for the system of MS non-hydrostatic equations is proposed and discussed. It resulted in the system of four PDEs (5.41). The same form of the system is still valid in the linearised case, the expressions of the residual fluxes  $\Phi_\eta^K$  and  $\Phi_q^K$ , however, must be rewritten. First of all the residual flux  $\Phi_q^K$  becomes  $\Phi_u^K$  in the linearised case since the variable  $q$  is no more present in system (C.1), but the only  $u$  remains. As a consequence, also the auxiliary variable  $w_q$ , is here introduced with the same expression of (5.36) for the reconstruction of the second space derivative of the velocity  $u$ .

Then, taking into account the cell formed by the nodes  $i$  and  $i-1$ , we can derive the following form of the residuals:

$$\begin{aligned}\Phi_\eta^{i-\frac{1}{2}} &= \Delta x \frac{d\eta_{i-\frac{1}{2}}}{dt} + h_0 (u_i - u_{i-1}) \\ \Phi_u^{i-\frac{1}{2}} &= \Delta x \frac{du_{i-\frac{1}{2}}}{dt} - \Delta x B h_0^2 \frac{dw_{u,i-\frac{1}{2}}}{dt} + \\ &\quad + g (\eta_i - \eta_{i-1}) - \beta g h_0^2 (w_{\eta,i} - w_{\eta,i-1})\end{aligned}\tag{C.6}$$

with which the central residual distribution scheme for the linearised problem can than be written in the same form of (5.41).

**Stabilized Upwind schemes :** The general form of the streamline upwind schemes, derived computing the sign of the local Jacobian matrix of the relative shallow water system, was given in (5.49). It was characterised by the adding of a stabilisation term, based on the value of the residual fluxes  $\Phi_\eta^K$  and  $\Phi_q^K$ , to a chosen *CS* (centred scheme). The expression of these fluxes was there computed using definitions (5.34) and (5.38).

In the linearised context, the Jacobian matrix of the linear shallow water system assumes the simpler form (1.3.1), being, thus, constant in all the domain since also the bathymetry  $h_0$  is assumed to be constant. Hence, the computation of  $\text{sign}(A)$  through the (5.50) results in:

$$\text{sign}(A) = \begin{bmatrix} 0 & c/g \\ g/c & 0 \end{bmatrix}\tag{C.7}$$

and (C.7) gives to (5.49) the simpler form:

$$\begin{bmatrix} CS_\eta \\ CS_u \end{bmatrix} + \frac{c}{2g} \begin{bmatrix} \Phi_u^{i-\frac{1}{2}} \\ \Phi_\eta^{i-\frac{1}{2}} \end{bmatrix} - \frac{g}{2c} \begin{bmatrix} \Phi_u^{i+\frac{1}{2}} \\ \Phi_\eta^{i+\frac{1}{2}} \end{bmatrix} \quad (\text{C.8})$$

where the formulae (C.6) are used to compute the residual fluxes  $\Phi_\eta^K$  and  $\Phi_u^K$  of the upwind terms, which are then add to a central Galerkin or  $\mathcal{RD}$  spatial discretization of the linear system (C.1).

# Appendix D

## Auxiliary Variables Choice in the MS Dispersion Term Discretization

The discretization of the Madsen and Sørensen system of equations leads to deal with a third order spatial derivative term. When the use of Galerkin finite element discretization is requested and the use of  $\mathcal{P}^1$  basis and weighting functions is preferred, the need of an auxiliary variable definition arises in order to decrease the derivation order of this term. Exactly as in Appendix C two possible reconstructions can be adopted:

$$w_{\eta 1,i} = \partial_x \eta_i \quad (\text{D.1})$$

$$w_{\eta 2,i} = \partial_{xx} \eta_i \quad (\text{D.2})$$

Applying these different reconstructions to the weak formulation of the third derivative term, in the two cases it assumes the form:

$$\beta g \int_{\Omega} \partial_x (\varphi_i h^3) \partial_x w_{\eta 1,i} dx \quad (\text{D.3})$$

$$\beta g \int_{\Omega} \partial_x (\varphi_i h^3) w_{\eta 2,i} dx \quad (\text{D.4})$$

When the hypothesis of a constant bathymetry is done, hence  $h(x) = h_0 =$

const, the solution of these integrals leads to the following expressions:

$$\beta \frac{gh_0^3}{\Delta x} (w_{\eta 1, i+1} - 2w_{\eta 1, i} + w_{\eta 1, i-1}) \quad (\text{D.5})$$

$$\beta \frac{gh_0^3}{2} (w_{\eta 2, i+1} - w_{\eta 2, i-1}) \quad (\text{D.6})$$

In order to compute the values of the variables  $w_{\eta, i}$ , which must be substitute in (D.5) and (D.6), the definitions of  $w_{\eta 1}$  and  $w_{\eta 2}$  need to be written in the weak formulation too and then integrated over the domain. This procedure, followed by the use of the mass lumping technique, allows to obtain for the variables  $w_{\eta 1, i}$  and  $w_{\eta 2, i}$  analogous expressions to that in equations (B.9) and (B.10).

The values of these auxiliary variable can, now, be computed in the  $i+1, i$  and  $i-1$  nodes of the grid and substitute in (D.5) and (D.6). Analytical development leads both the expression to be identically equal to:

$$\beta \frac{gh_0^3}{2\Delta x^2} (\eta_{i+2} - 2\eta_{i+1} + 2\eta_{i-1} - \eta_{i-2}) \quad (\text{D.7})$$

which simply reflects the second order finite difference discretization of the third order spatial derivative (see (5.3)) integrated over the domain.

In conclusion, under the hypothesis of  $h = \text{const}$  the two different reconstructions leads to the same final discretised expression of the third spatial derivative term  $\partial_{x^3}\eta$ . In cases in which this hypothesis results no more verified, a small difference in the two approaches originates in dependence on the quadrature formula used to solve the definite integrals (D.3) and (D.4), but it seems not to affect the goodness of the solution.

## Appendix E

# Galerkin Space-Discretization of the MS System

In Section 5.2 the Madsen and Sørensen system has been semi-discretized in space developing a central Galerkin finite element method, using linear basis function and through the introduction of an auxiliary variable  $w_\eta$  in order decrease the spatial derivation order of the system. The resulting scheme, in the case of the complete nonlinear MS system and with variable bathymetry  $h(x)$ , is listed below:

$$\begin{aligned}
& \frac{\Delta x}{6} \frac{d\eta_{i-1}}{dt} + \frac{2\Delta x}{3} \frac{d\eta_i}{dt} + \frac{\Delta x}{6} \frac{d\eta_{i+1}}{dt} + \frac{1}{2} (q_{i+1} - q_{i-1}) = 0 \\
& \frac{\Delta x}{6} \frac{dq_{i-1}}{dt} + \frac{2\Delta x}{3} \frac{dq_i}{dt} + \frac{\Delta x}{6} \frac{dq_{i+1}}{dt} + \frac{1}{2} (u_{i+1}q_{i+1} - u_{i-1}q_{i-1}) + \\
& \quad + \frac{B}{3\Delta x} (h_i^2 + h_i h_{i-1} + h_{i-1}^2) \left( \frac{dq_i}{dt} - \frac{dq_{i-1}}{dt} \right) + \\
& \quad + \frac{B}{3\Delta x} (h_i - h_{i-1}) (2h_i + h_{i-1}) \left( \frac{dq_i}{dt} - \frac{dq_{i-1}}{dt} \right) + \\
& \quad - \frac{B}{3\Delta x} (h_i^2 + h_i h_{i+1} + h_{i+1}^2) \left( \frac{dq_{i+1}}{dt} - \frac{dq_i}{dt} \right) + \\
& \quad + \frac{B}{3\Delta x} (h_{i+1} - h_i) (2h_i + h_{i+1}) \left( \frac{dq_{i+1}}{dt} - \frac{dq_i}{dt} \right) + \\
& \quad - \frac{1}{18\Delta x} (h_i - h_{i-1}) (2h_i + h_{i-1}) \left( \frac{dq_i}{dt} - \frac{dq_{i-1}}{dt} \right) + \\
& \quad - \frac{1}{18\Delta x} (h_{i+1} - h_i) (2h_i + h_{i+1}) \left( \frac{dq_{i+1}}{dt} - \frac{dq_i}{dt} \right) \tag{E.1} \\
& \quad + \frac{g}{6} (2H_i + H_{i-1}) (\eta_i - \eta_{i-1}) + \frac{g}{6} (2H_i + H_{i+1}) (\eta_{i+1} - \eta_i) + \\
& \quad - \frac{\beta g}{6} \left( \frac{1}{4} (h_i + h_{i-1})^3 + h_i^3 \right) (w_{\eta,i} - w_{\eta,i-1}) + \\
& \quad - \frac{\beta g}{6} \left( \frac{1}{4} (h_i + h_{i+1})^3 + h_i^3 \right) (w_{\eta,i+1} - w_{\eta,i}) + \\
& \quad + \frac{2\beta g}{3\Delta x^2} (h_i - h_{i-1}) (h_i^2 + h_i h_{i-1} + h_{i-1}^2) (\eta_i - \eta_{i-1}) + \\
& \quad + \frac{2\beta g}{3\Delta x^2} (h_i - h_{i-1})^2 (2h_i + h_{i-1}) (\eta_i - \eta_{i-1}) + \\
& \quad - \frac{2\beta g}{3\Delta x^2} (h_{i+1} - h_i) (h_i^2 + h_i h_{i+1} + h_{i+1}^2) (\eta_{i+1} - \eta_i) + \\
& \quad + \frac{2\beta g}{3\Delta x^2} (h_{i+1} - h_i)^2 (2h_i + h_{i+1}) (\eta_{i+1} - \eta_i) = 0 \\
& w_{\eta,i} - \frac{1}{\Delta x^2} (\eta_{i+1} - 2\eta_i + \eta_{i-1}) = 0
\end{aligned}$$

# Appendix F

## Space-Time Discretization of the Linearized MS System

This Appendix was created to show the fully expressions obtained integrating with the Crank-Nicolson method the semi-discrete schemes illustrated in Appendix C. The resulted expressions written in the following are, thus, a numerical approximation of the linearised MS system (C.1) and they were used in Section 6.1.1 in order to accomplish the study of the properties of the more complex schemes applied to the complete nonlinear Madsen and Sørensen system (1.9).

**Galerkin FE scheme with  $\mathcal{CN}$  time integration :** The system of space-time discretized equation, written below, is obtained first spatially discretizing the linearized version of the MS system of equation with the Galerkin finite element scheme of Section 5.2, then integrating the results with the Crank-Nicolson method explained in Section 6.1. The final expression is then related to a generic inner node  $i$  of the grid, to obtain:

$$\begin{aligned}
& \frac{1}{6\Delta t} (\eta_{i-1}^{n+1} - \eta_{i-1}^n) + \frac{2}{3\Delta t} (\eta_i^{n+1} - \eta_i^n) + \frac{1}{6\Delta t} (\eta_{i+1}^{n+1} - \eta_{i+1}^n) + \\
& \qquad \qquad \qquad + \frac{h_0}{4\Delta x} (u_{i+1}^{n+\frac{1}{2}} - u_{i-1}^{n+\frac{1}{2}}) = 0 \\
& \frac{1}{6\Delta t} (u_{i-1}^{n+1} - u_{i-1}^n) + \frac{2}{3\Delta t} (u_i^{n+1} - u_i^n) + \frac{1}{6\Delta t} (u_{i+1}^{n+1} - u_{i+1}^n) + \\
& \qquad - \frac{Bh_0^2}{\Delta t \Delta x^2} ((u_{i+1}^{n+1} - u_{i+1}^n) - 2(u_i^{n+1} - u_i^n) + (u_{i-1}^{n+1} - u_{i-1}^n)) + \\
& \qquad + \frac{g}{4\Delta x} (\eta_{i+1}^{n+\frac{1}{2}} - \eta_{i-1}^{n+\frac{1}{2}}) - \frac{\beta gh_0^2}{4\Delta x^3} (w_{\eta,i+1}^{n+\frac{1}{2}} - w_{\eta,i-1}^{n+\frac{1}{2}}) = 0
\end{aligned} \tag{F.1}$$

The definition of the variable  $w_\eta$  is coherent to what set all along this work.

**$\mathcal{RD}$  scheme with  $\mathcal{CN}$  time integration :** In the residual distribution environment, the general form of the scheme remains the system of the four PDEs (5.41), but the definitions of the residual fluxes  $\Phi^K$  are, for instance:

$$\begin{aligned}
\Phi_\eta^{i-\frac{1}{2}} &= \frac{1}{2\Delta t} (\eta_{i-\frac{1}{2}}^{n+1} - \eta_{i-\frac{1}{2}}^n) + \frac{h_0}{2\Delta x} (u_i - u_{i-1})^{n+1} + \frac{h_0}{2\Delta x} (u_i - u_{i-1})^n \\
\Phi_u^{i-\frac{1}{2}} &= \frac{1}{2\Delta t} (u_{i-\frac{1}{2}}^{n+1} - u_{i-\frac{1}{2}}^n) - \frac{Bh_0^2}{\Delta t} (w_{u,i-\frac{1}{2}}^{n+1} - w_{u,i-\frac{1}{2}}^n) + \\
& \qquad + \frac{g}{2\Delta x} (\eta_i - \eta_{i-1})^{n+1} + \frac{g}{2\Delta x} (\eta_i - \eta_{i-1})^n + \\
& \qquad - \frac{\beta gh_0^2}{2\Delta x} (w_{\eta,i} - w_{\eta,i-1})^{n+1} - \frac{\beta gh_0^2}{2\Delta x} (w_{\eta,i} - w_{\eta,i-1})^n
\end{aligned} \tag{F.2}$$

**Stabilization Upwind schemes with  $\mathcal{CN}$  time integration :** In Appendix C a general form of the two stabilized upwind schemes developed in this work is given through (C.8). In the previous we have already explained, in fact, that the difference between the, what are here called,  $\mathcal{SU}/\mathcal{PG}$  method and  $\mathcal{U}/\mathcal{RD}$  method regards the only choice of the centred scheme which is chosen to be stabilized. We have shown, in fact, that the discretization of the upwind term leads in each case to the same result (if, of course, the same choice of the auxiliary variables introduced in the system would be always coherent with the what is here chosen).

Applying the Crank-Nicolson method to integrate in time the system (C.8), leads to the final expression of the stabilized space-time discrete schemes. In order to generalize these expressions, they can be still rewritten in the same form as (C.8):

$$\begin{bmatrix} CS_\eta \\ CS_u \end{bmatrix} + \frac{c}{2g} \begin{bmatrix} \Phi_u^{i-\frac{1}{2}} \\ \Phi_\eta^{i-\frac{1}{2}} \end{bmatrix} - \frac{g}{2c} \begin{bmatrix} \Phi_u^{i+\frac{1}{2}} \\ \Phi_\eta^{i+\frac{1}{2}} \end{bmatrix} \quad (\text{F.3})$$

However, even if (F.3) is formally equal to (C.8), the two contexts to which they belong make them deeply different. Here, in fact, the upwind term is computed with the expressions of the residuals  $\Phi_\eta^K$  and  $\Phi_u^K$  given in (F.2). The development of a  $SU/\mathcal{PG}$  space-time discrete scheme for the linearised MS system can then be reached by replacing the term  $CS$  with the central Galerkin space-time discrete scheme (F.1). On the other hand, the choice of a centred residual distribution scheme using the same form of the residual fluxes (F.2) leads to a  $U/\mathcal{RD}$  space-discrete scheme.

[24]



# Bibliography

- [1] M. Kazolea A.I. Delis. Finite volume simulations of waves formed by sliding masses. Communications in numerical methods in engineering, 27, 732-735, 2011.
- [2] D. Ambrosi. Approximation of shallow water equations by roe's riemann solver. International Journal for numerical methods in fluids, 20, 157-168, 1995.
- [3] H.B. Bingham T. Warburton A.P. Engsig-Karupp, J.S. Hesthaven. Dg-fem solution for nonlinear wave-structure interaction using boussinesq-type equations. Coastal Engineering, 55, 197, 2008.
- [4] E. Audusse and M.-O. Bristeau. A well-balanced positivity preserving second-order scheme for shallow water flows on unstructured meshes. Journal of Computational Physics, 206(1):311 – 333, 2005.
- [5] M. Behr. Stability analysis of scalar advection-diffusion equation. <http://www.cats.rwth-aachen.de/library/research/technotes>.
- [6] A. Bermudez and M.E. Vazquez. Upwind methods for hyperbolic conservation laws with source terms. Computers & Fluids, 23(8):1049 – 1071, 1994.
- [7] P. Bonneton, E. Barthelemy, F. Chazel, R. Cienfuegos, D. Lannes, F. Marche, and M. Tissier. Recent advances in serre-green naghdi modelling for wave transformation, breaking and runup processes. European Journal of Mechanics - B/Fluids, In Press, Corrected Proof:–, 2011.
- [8] P. Brufau and P. Garcia-Navarro. Unsteady free surface flow simulation over complex topography with a multidimensional upwind technique. Journal of Computational Physics, 186(2):503 – 526, 2003.

- [9] S.J. Sherwin C. Eskilsson. Spectral/hp discontinuous galerkin methods for modelling 2d boussinsq equations. Journal of Computational Physics 210, 566, 2006.
- [10] L. Quartapelle D. Ambrosi. A taylor-galerkin method for simulating nonlinear dispersive water waves. Journal of computational physics, 146, 546-569, 1998.
- [11] P.A. Madsen D.R. Fuhrman. Simulation of nonlinear wave run-up with a high order boussinesq model. Coastal Engineering 55, 139, 2008.
- [12] A. Ern and J.-C. Guermond. Theory and practice of finite elements, volume 159 of Applied Mathematical Sciences. Springer, 2004.
- [13] A. Ern, S. Piperno, and K. Djadel. A well-balanced runge–kutta discontinuous galerkin method for the shallow-water equations with flooding and drying. Int. J. for Numerical Methods in Fluids, 58(1):1–25, 2008.
- [14] G.j. Fix G. Strang. an analysis of the finite element method. Prantice-Hall, 1973.
- [15] J.T. Kirby G. Wei. A time-dependent numerical code for extended boussinesq equations. Journal of Waterway, Port, Coastal, and Ocean Engineering 120, 251-261, 1995.
- [16] P. Garcia-Navarro, M. Hubbard, and A. Priestley. Genuinely multi-dimensional upwinding for the 2D shallow water equations. J. Comp. Phys., 121(1):79–93, 1995.
- [17] G. Hauke. A symmetric formulation for computing transient shallow water flows. Computer Methods in Applied Mechanics and Engineering, 163(1-4):111–122, 1998.
- [18] M. Hubbard and M.J. Baines. Conservative multidimensional upwinding for the steady two-dimensional shallow-water equations. J. Comput. Phys., 138:419–448, 1997.
- [19] T.J.R. Hughes and M. Mallet. A new finite element formulation for CFD III: the generalized streamline operator for multidimensional advective-diffusive systems. Comp. Meth. Appl. Mech. Engrg., 58:305–328, 1986.
- [20] T.J.R. Hughes and T.E. Tezduyar. Development of time-accurate finite element techniques for first order hyperbolic systems with emphasis on the compressible euler equations. Comp. Meth. Appl. Mech. Engrg., 45(1-3):217–284, 1984.

- [21] P.H. Taylor J. Orszaghova, A.G.L. Borthwick. From the paddle to the beach - a bousinesq shallow water numerical wave tank based on mad-sen and sørensen's equations. Journal of Computational Physics, 231, 328-344, 2012.
- [22] J.F.T. Pittman J. Petera. Isoparametric hermite elements. International journal for numerical methods in engineering, 37, 3489-3519, 1994.
- [23] J.T. Kirby. Boussinesq models and applications to near-shore wave propagation, surf zone processes and wave-induced currents. Advances in Coastal Modelling, 67, 1-41, 2003.
- [24] A. Kurganov and E. Tadmor. Solution of two-dimensional Riemann problems without Riemann solvers. Numerical Methods for Partial Differential Equations, 18:548–608, 2002.
- [25] D. Lannes and P. Bonneton. Derivation of asymptotic two-dimensional time-dependent equations for surface water wave propagation. Physics of Fluids, 21, 2009. 016601 doi:10.1063/1.3053183.
- [26] R. LeVeque. Finite volume methods for hyperbolic problems. Cambridge, 2002.
- [27] R.J. LeVeque. Balancing source terms and flux gradients in high-resolution godunov methods: the quasi-steady wave-propagation algorithm. J. Comput. Phys., 146(1):346–365, 1998.
- [28] I.K. Nikolos C.E. Synolakis M. Kazolea, A.I. Delis. An unstructured finite volume numerical scheme for extended 2d boussinesq-type equations. Coastal Engineering 69, 42-66, 2012.
- [29] M. Petti M. Tonelli. Hybrid finite volume - finite difference scheme for 2dh improved boussinesq equations. Coastal Engineering, 56, 609-620, 2009.
- [30] F. Marche, P. Bonneton, P. Fabrie, and N. Seguin. Evaluation of well-balanced bore-capturing schemes for 2D wetting and drying processes. Int. J. Numer. Meth. Fluids, 53:867–894, 2007.
- [31] I.K. Nikolos and A.I. Delis. An unstructured node-centered finite volume scheme for shallow water flows with wet/dry fronts over complex topography. Computer Methods in Applied Mechanics and Engineering, 198(47-48):3723 – 3750, 2009.

- [32] O. Nwogu. An alternative form of the boussinesq equations for near-shore wave propagation. Journal of Waterway, Port, Coastal, and Ocean Engineering 119, 618-638, 1994.
- [33] L. Sørensen O.R. Sørensen, H. Shaffer. Boussinesq-type modelling using an unstructured finite element technique. Coastal Engineering, 50, 182, 2004.
- [34] H. Liu P.A. Madsen, H. Bingham. A new boussinesq method for fully nonlinear waves from shallow to deep water. Journal of Fluid Mechanics 462, 1, 2002.
- [35] O.R. Sørensen P.A. Madsen. A new form of the boussinesq equations with improved linear dispersion characteristics. part 2: a slowing varying bathymetry. Coastal Engineering 18, 183-204, 1992.
- [36] M. Ricchiuto. Contributions to the development of residual discretization for hyperbolic conservation laws with applications to shallow water flows. HDR Thesis, University of Bordeaux, 2011.
- [37] M. Ricchiuto, R. Abgrall, and H. Deconinck. Application of conservative residual distribution schemes to the solution of the shallow water equations on unstructured meshes,. J. Comput. Phys., 222:287–331, 2007.
- [38] M. Ricchiuto and A. Bollermann. Stabilized residual distribution for shallow water simulations. J. Comput. Phys, 228(4):1071–1115, 2009.
- [39] J.A. Battjes S. Beji. Numerical simulations of nonlinear-wave propagation over a bar. Coastal Engineering, 23, 1-16, 1994.
- [40] K. Nadoka S. Beji. A formal derivation and numerical modelling of the improved boussinesq equations for varying depth. Ocean Engineering 23, 691, 1996.
- [41] R. Volker. Boussinesq-type model for near-shore wave processes in fringing reef environment. PhD Thesis, University of Hawaii and Manoa, 2010.
- [42] M.A. Walkley. A numerical method for extended boussinesq shallow-water wave equations. PhD Thesis, University of Leeds, 1999.
- [43] Y. Xing and C.-W. Shu. High-order finite volume weno schemes for the shallow water equations with dry states. Advances in Water Resources, 34(8):1026 – 1038, 2011.

- [44] Y. Xing, X. Zhang, and C.-W. Shu. Positivity-preserving high order well-balanced discontinuous galerkin methods for the shallow water equations. Advances in Water Resources, 33(12):1476 – 1493, 2010.
- [45] Y.X. Yu G. Lai... Y.S. Li, S.X. Liu. Numerical modelling of boussinesq equations by finite element method. Coastal Engineering 37, 97, 1999.