Numerical modelling of gravel-bed river morphodynamics

Guglielmo Stecca

UNIVERSITÀ DEGLI STUDI DI TRENTO
Dipartimento di Ingegneria Civile
e Ambientale

2012

Doctoral thesis in **Environmental Engineering** (XXIV cycle)
Faculty of Engineering, **University of Trento**
Year: **2012**
Supervisor: **Dr. Guido Zolezzi, Dr. Annunziato Siviglia, Prof. Eleuterio F. Toro**

Front cover: Rakaia River, New Zealand, July 2011

Università degli Studi di Trento
Trento, Italy
2012

Ma se io avessi previsto tutto questo, dati causa e pretesto, forse farei lo stesso

Francesco Guccini, L'Avvelenata

Abstract

This thesis is about the development and testing of a novel two-dimensional numerical model (the GIAMT2D model) able to address the hydro-morphodynamic evolution of gravel-bed rivers. The model solves the two-dimensional hyperbolic system of partial differential equations (PDEs) arising from the shallow water-Exner model, describing free surface shallow flows over erodible bed, with suitable closure relations for bedload transport. A coupled formulation of the mathematical problem, which is needed in order to correctly handle sediment transport in Froude trans-critical flow conditions, is implemented, resulting in a non-conservative hyperbolic problem, which requires the adoption of a path-conservative scheme.

A drawback of the fully-coupled shallow water-Exner model is that in general the solution of the Riemann problem is not easily available, at least if complex empirical sediment transport formulae are applied, which makes the upwind approach inadequate for designing numerical approximations to the solutions. Adoption of the more general, Riemann solver-free centred approach is thus required, the drawback being that centred schemes are significantly less accurate than upwind schemes in some specific cases, namely for intermediate waves and computations at low CFL number. In GIAMT2D an original centred upwind-biased scheme (UPRICE2-Cδ) is applied, recovering accuracy typical of upwind methods, still being able to include any bedload transport formula. The proposed scheme results from original studies in applied mathematics, presented in the first part of the thesis, concerning the development of upwind-biased variations of the centred FORCE scheme for the solution of hyperbolic systems of PDEs, in conservative and non-conservative form. The performance of these schemes is thoroughly assessed in a suite of tests for the shallow water equations.

The GIAMT2D model embeds the UPRICE2-Cδ scheme extended to second-order accuracy in the ADER framework, inserted in a robust second-order preserving splitting technique for the treatment of frictional source terms, and includes an original wetting-and-drying procedure. The model performance is checked in well-established classical test cases with fixed and movable bed. These applications highlight the capability of the model in correctly and accurately solving the equations in

various cases, e.g. in computations at low local CFL number, in the solution of wet-dry fronts with fixed and movable bed and in the prediction of sediment transport in Froude trans-critical conditions.

The concept of "morphodynamic benchmark" is introduced for the purpose of assessing the model performance in reproducing basic river morphodynamic processes for which established theoretical and experimental knowledge is available. Unit processes with utmost importance for gravel-bed river morphodynamics, like free and forced bar instability and the stability of channel bifurcations, are chosen for this aim. In this novel approach for assessing the model capabilities, the numerical solutions satisfactorily compare with approximate analytical morphodynamic solutions and laboratory data.

Having proved that the model is able to reproduce the salient features of these classical morphodynamic solutions, an original morphodynamic study is finally carried out, concerning the non-linear interaction of free and forced bars in straight channels, for which a mature analytical theory is not available at present. The numerical runs of GIAMT2D are used to validate the research hypotheses developed on the basis of existing analytical theories and satisfactorily compare with field observations.

Contents

List of Figures

List of Tables

1 Introduction

Rivers, flowing through alluvial fans and valleys under the action of gravity, represent the most effective terrestrial transporting agent [105]. Alluvial gravel-bed rivers are typically found in mountain and piedmont areas, characterized by values of the longitudinal slope that are intermediate between those of steep mountain streams and of lowland floodplain reaches, where typically fine sediment prevails. The action of water on an erodible bed, characterised mainly by cohesionless sediments, induces sediment transport through the hydrodynamic shear exerted by flow at the fluid-bottom interface. The combination of slope and bed material commonly found in alluvial gravel-bed rivers determines two key properties of gravel-bed rivers: the first one is the possible occurrence of near Froude-critical conditions under bar- and channel-forming events that reshape the bed and planform morphology; the second is that sediment transport mainly occurs as bedload (even though a consistent fraction of the total sediment load is carried in suspension, but apparently with limited relevance for morphological changes). The instability of flow-driven transport generates a variety of complex patterns in the river bed and in the planform morphology. Bedforms scaling with the channel width, called *bars*, represent the fundamental morphological unit of river bedform and transport dynamics. Bars can be either migrating (associated to the inherent instability of the system), or steady, i.e. triggered by local or distributed perturbations of the planform configuration. The dynamic evolution of the riverbed topography is a consequence of different processes mutually interacting at different scales: the actual bar dynamics of gravel-bed rivers is likely to be determined by the non-linear interactions of free and forced bars. Natural rivers assume single-thread (meandering) or multi-thread (braiding) planimetric configurations, or configurations which are intermediate between the two (wandering, transitional). The evolution of braided rivers is determined by a variety of morphodynamic phenomena, which can be decomposed into fundamental unit processes. Among those, channel bifurcations and bar instability are of utmost relevance. An example of the resulting pattern in the case of a gravel-bed braided river is given in Fig. 1.1.

The complexity exhibited by river patterns often discloses surprisingly regular bed and planform morphologies, which has suggested their study through mathematical modelling. The mathematical

Figure 1.1: Bar pattern in a gravel-bed braided river (the Tagliamento River, Italy).

modelling of hydro-morphodynamic evolution problems at local and reach scale in gravel-bed rivers is commonly achieved by coupling a model for free-surface flow with a volumetric sediment balance for sediments, or Exner equation, including empirical relations for sediment transport, even though more complicated models, based on finer simulation of turbulence, have been applied [88]. Flow is often modelled by the Reynolds-averaged Navier-Stokes equations or by their further approximations, such as the shallow water equations, based on the assumption of hydrostatic pressure distribution. We denote the coupled model based on the latter equations as *the shallow water-Exner model*. A distinctive feature of commonly employed flow-transport models is non-linearity, which can be related to the complexity and interactive character of hydro-morphodynamic processes in real rivers.

Two approaches are available for the solution of the resulting systems of partial differential equations (PDEs), namely the numerical approach, providing an approximation to the *solutions*, and the analytical approach, providing approximations to the *equations*. Analytical models often employ solution techniques based on perturbative methods, in which the unknowns are expanded in Taylor series up to the desired order and high-order terms are neglected, thus obtaining linear or weakly non-linear solutions. These models provide deep insight of the fundamental behaviour of the physical system, being able to identify the controlling parameters and to detect the fundamental physics. Moreover, they are very useful in the calibration and validation of numerical models.

However, in order to keep the problem amenable to analytical treatment, they usually focus on simplified planimetric configurations, often neglecting fully non-linear interactions, which makes their solutions unfit to reproduce complex processes such as the bar dynamics of braided rivers [87, 39].
A complementary modelling approach is physical modelling at laboratory scale, providing direct observations of morphodynamic processes in controlled conditions. Experimental observations give direct evidence on the morphodynamic processes in river systems, thus being preliminary to mathematical modelling. The outcomes of analytical theories, identifying the controlling parameters in morphodynamic evolution processes, are fruitfully applied in the interpretation of experimental results, while experimental data provide fundamental benchmarks to test analytical theories and numerical experiments.

In the last decades, the adoption of numerical models, able to integrate the system of governing equations fully preserving their non-linearity, has widely taken place in hydro-morphodynamic modelling. The general increase in the available computational power and continuous improvement in numerical techniques has allowed the development of more refined numerical models, characterised by higher accuracy, incorporation of an increasing number of physical processes and increasing dimensionality [88]. With respect to the analytical approach, numerical techniques allow much more flexibility in the choice of planimetric configurations and in the incorporation of more morphodynamic processes, whose interactions are directly taken into account. However, the main drawback of numerical approaches is their being to some extent "blind", not themselves providing immediate insight, detection of general behaviours and controlling parameters. Moreover, unwise application of numerical models may be misleading if the numerical features of the solution scheme are mistaken for physical features of the considered process [88]. Thus the design of an effective numerical model for morphodynamic applications implies the adoption of a multi-disciplinary approach. In detail, it requires knowledge of the mathematical properties of the system of PDEs, of accurate, stable, efficient and most suitable solution strategies, of the physical processes to be interpreted and of the outcomes of physical and analytical modelling, whose results shall be used as benchmarks to test the numerical model.

The general aim of the present Ph.D. research is to develop a novel numerical tool able to address hydro-morphodynamic problems in gravel-bed rivers. The numerical solution sought shall satisfy three broad categories of requirements. In detail, the model shall be

- generally applicable for the simulation of hydro-morphodynamic processes in gravel-bed rivers;

- able to correctly and accurately solve the system of governing equations;
- able to reproduce the morphodynamic behaviour of gravel-bed rivers.

The first category of requirements concerns the above mentioned peculiarities of flow and transport conditions in gravel-bed rivers, which must be correctly handled by the model. The numerical model shall account for two-dimensional free-surface flow in erodible channels characterised by bedload transport by integrating the two-dimensional shallow-water Exner model. For bedload transport, plenty of different empirically-derived formulae exist [115], whose application and test may be required by the engineering practice, while new relations may have to be calibrated in some cases in order to fit the available data. Thus the solver shall be able to incorporate these formulae in computations and allow easy switch between them.
The planimetric and topographic complexity of gravel-bed rivers, together with unsteadiness in the flow regime and occurrence of trans-critical flows, poses some requirements over the domain discretisation and solution strategies. Polygonal computational domains of irregular shape, reproducing complex planform configurations, shall be properly discretised, using for instance unstructured meshes (see the domain discretisation by unstructured triangular meshes in a bifurcation-junction configuration in Fig. 1.2). High variability in the local flow stage due to topographic complexity may result in the presence of dry regions in the computational domain, whose extent and location may change in time due to morphological evolution and unsteadiness in the flow regime (see Fig. 1.3). Algorithms designed for the treatment of standard fully-wet domains may fail in computations carried out in regions characterised by vanishing depth. Thus the model should incorporate a wetting-and-drying procedure, able to deal with these features in adaptive and robust manner.
The possible occurrence of trans-critical flows requires particular care in the modelling of sediment transport. Whereas in general, outside the trans-critical Froude region, the bed interacts only weakly with the free surface, this is not true in trans-critical conditions, when the characteristic time scale for the propagation of bed waves is comparable to that of free-surface waves [109, 84]. In this case the mathematical problem is fully coupled and a coupled solution strategy, in which the hydrodynamic and morphodynamic problem are solved synchronously, shall be implemented.

The second category concerns the properties of the numerical solution of the governing equations. A basic requirement for numerical schemes is conservativity, which guarantees through the Lax-Wendroff theorem [78] that the numerical solution will converge to the weak solution of the problem as the grid is progressively refined. This implies that the scheme will be able to correctly predict the speed of waves, either smooth or discontinuous, arising in computations. In the framework of finite volumes, the construction of conservative schemes is achieved in straightforward manner by discretising the conservation-law form of systems of PDEs, provided such

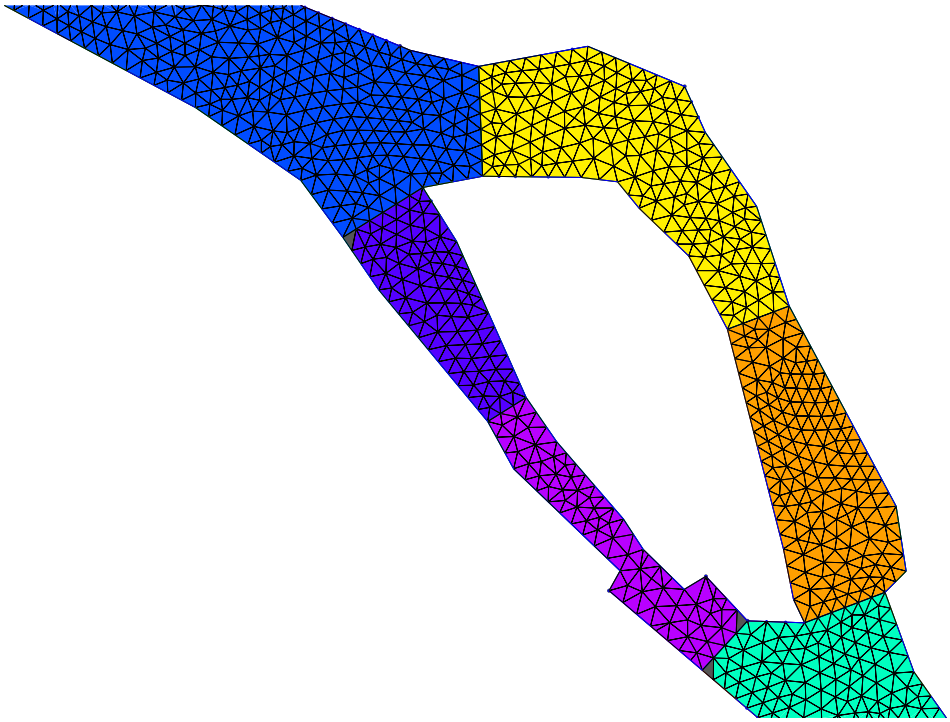

Figure 1.2: Computational domain of complex polygonal shape used for numerical simulations in a bifurcation-junction configuration on the Armea River (Italy), discretised with an unstructured triangular mesh. From Tubino *et al.* [128].

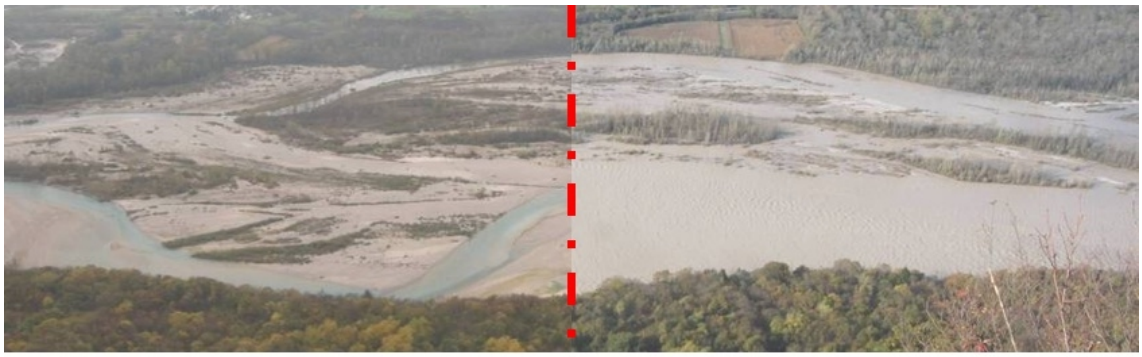

Figure 1.3: Temporal variation in the wet and dry patterns due to unsteady flow regime during a flood event in the Tagliamento River (Italy).

formulation is available. Unfortunately, the shallow water-Exner model, which has recently been proved to belong to the class of hyperbolic systems of PDEs for the range of flow data typically encountered in rivers [35], does not admit a conservation law-form if a coupled formulation, which is mandatory when dealing with trans-critical flows, is adopted [110]. In this case, adoption of the path-conservative framework [42, 95], although still questionable from the point of view of mathematical proof of convergence [1], allows to recover an acceptable definition of weak solutions and to develop shock-capturing methods.
Beyond convergence towards weak solutions, natural requirements for the numerical scheme are accuracy (low numerical dissipation), computational efficiency and generality, i.e. in the present case the possibility of including any bedload transport formula in computations. The behaviour of the numerical solver with respect to these three (somehow contradictory) requirements is deeply influenced by the nature of the solution scheme, which must provide the best trade-off between them. In the framework of finite volume methods, two approaches are available for designing numerical schemes for hyperbolic systems of PDEs: the *upwind* approach, based on the application of Riemann solvers, and the *centred* approach, which does not require any information on the system eigenstructure [112, 113]. Upwind methods are generally more accurate than centred methods, the drawbacks being higher computational cost, higher complexity resulting in more difficult implementation and less generality. The use of upwind methods is restricted to the problems for which the solution of the Riemann problem is viable: since this is not the case for the shallow water-Exner model, at least if complex sediment transport formulae are used, application of the centred approach for the solution of the hydro-morphodynamic problem is preferable [22, 21].
The numerical model presented in this thesis is based on an refined *upwind-biased centred* scheme, resulting from original research in applied mathematics. The scheme partially uses upwind information without the application of any Riemann solver, thus being characterised by accuracy comparable to that of upwind methods, still retaining the generality of centred methods, i.e. the possibility of applying any bedload transport formula. The development of upwind-biased schemes is first conducted with reference to the case of homogeneous hyperbolic systems of conservation laws, with applications to a suite of classical tests for the inviscid shallow water equations with fixed bed. A path-conservative version is then presented, suitable for applications to the hydro-morphodynamic problem. Extension of schemes to second-order accuracy provides sharp solution profiles at a still reasonable computational cost.

The development of a computational tool able to correctly and accurately solve the system of governing equations is only an intermediate objective: the final goal in this thesis is to build a numerical model able to reproduce the real morphodynamic behaviour of gravel-bed rivers. Therefore one of the main distinctive features of the present research is the production of novel outcomes both

in the field of applied mathematics and river morphodynamics.
The capability of the model is carefully assessed in tests aiming at reproducing the morphodynamics observed in controlled conditions, by checking its performance against morphodynamic benchmarks given by well-established analytical solutions and laboratory experiments. Since few examples of numerical applications of this kind are available in the literature, this represents an innovative approach for testing the model. The morphodynamic benchmarks have been chosen in relation to the dynamics of free and forced bars and channel bifurcations. As previously mentioned, a wealth of theories exist which give simplified analytical solutions for these unit morphodynamic processes. More specifically, in the present work we check the performance of the numerical model in the development of free bars in straight channels against the linear free bar theory [33] and we prove that the model is able to correctly address the morphodynamic influence of a localised variation in planform in agreement with the linear theory of spatial bars [141]. Moreover, we focus on numerical assessment of channel bifurcation stability, comparing the numerical results with experimental [11] and analytical [14] results.
These applications aim at reproducing individual processes. However, the morphodynamic evolution of real gravel-bed rivers is also determined by their non-linear interaction. In particular, focusing on single-thread rivers, experimental and theoretical evidence indicates that the actual bar dynamics is the result of the non-linear interaction of free and forced bars [68, 127, 97]. However, a morphodynamic theory able to fully account for these interaction processes in the basic straight-channel configuration is not available at present [142]. Thus, after assessing the capabilities of the numerical model, we conduct an original study which aims at addressing such interactions by integrating the outcomes of existing analytical theories, remotely-sensed field data and numerical modelling.

The present thesis outlines as follows. In chapter 2 we introduce the shallow water-Exner model, governing the hydro-morphodynamic evolution of rivers, and review its properties. We show that the system of governing equations is hyperbolic under some assumptions which are reasonable in gravel-bed rivers and therefore devote some attention to recalling the mathematical properties of hyperbolic systems of PDEs.
The first part of the thesis is devoted to original research in applied mathematics. In chapter 3 we develop novel upwind-biased centred schemes for systems of conservation laws and assess their performance in well-established test cases for the two-dimensional shallow water equations. Then in chapter 4 we extend the schemes developed in the previous chapter for the solution of hyperbolic systems in non-conservative form.
Chapter 5 bridges the mathematical part to the morphodynamic applications which are considered in the second part of the thesis. Here, based on the scheme developed in the previous chapter,

we build our hydro-morphodynamic model, regarded as GIAMT2D, and assess its performance employing classical test cases for fixed and movable bed. Chapter 6 is devoted to reviewing the main achievements in gravel-bed river morphodynamics obtained by analytical and physical modelling. Here we present the theories of free and forced bars and analytical and laboratory results concerning the stability of channel bifurcations. Then in chapter 7 we validate the model against these morphodynamic benchmarks. Finally, in chapter 8 we conduct an original morphodynamic study concerning the interaction of free and forced bars in straight channels.

2 The mathematical model

The mathematical model considered in this thesis, accounting for two-dimensional shallow flows in open erodible channels characterised by bedload sediment transport, results from the coupling of the two-dimensional depth-averaged shallow water equations with the Exner equation, for short referred to as the *two-dimensional shallow water-Exner model*. It is composed of a system of four partial differential equations (PDEs), together with algebraic closure relations for friction and bedload transport.

This chapter is about the presentation of the model and explanation of its mathematical behaviour. In detail, we aim at proving that the considered system of PDEs, together with the proposed closure relations, belongs to the class of hyperbolic systems, at least for the range of flow data which are of interest in practical applications in rivers. This point is of crucial relevance since it deals with assessing well-posedness of the mathematical problem, i.e. assessing the existence and uniqueness of the solution for the initial value problem. Before proving hyperbolicity of the shallow water-Exner model, we define the concept of hyperbolic systems and review their properties. To this aim, we preliminarily introduce the two-dimensional inviscid shallow water equations, governing free-surface flow over fixed horizontal bed, which represent an example of a hyperbolic system of PDEs.
Hyperbolic systems may be either conservative, i.e. admit a conservation-law form, or non-conservative. The shallow water-Exner model cannot be expressed in conservation-law form if a coupled formulation of the hydrodynamic and morphodynamic problem is adopted. Therefore in our review we devote some attention to the definition of weak solutions of non-conservative systems and to the related debate in the literature concerning convergence of numerical methods to these solutions. These issues will be of practical relevance in the development of the numerical model.

The present chapter outlines as follows. In section 2.1 we introduce the two-dimensional shallow water-Exner model. Then in section 2.2 we present the inviscid shallow water equations, which are used in 2.3 for defining the class of hyperbolic systems of PDEs and reviewing their basic

properties. In section 2.4 we prove that the shallow water-Exner model is hyperbolic within the range of flow data which is typical of gravel-bed rivers. Finally, in section 2.5 we show that the coupled shallow water-Exner system does not admit a conservation-law form, thus being a two-dimensional non-conservative hyperbolic system of PDEs.

2.1 The shallow water-Exner model

In this section we present the mathematical model, describing two-dimensional free-surface water flow over erodible bed. The riverbed is assumed to be composed of well-graded sediments characterised by a unique value of particle size D_s. The grain size of sediments transported by water flow is the same as that of sediments composing the riverbed. Solid transport is characterized by bedload only, thus ignoring the role of suspension.
The model is composed of a system of four partial differential equations (the governing equations) and closure relations for friction and bedload transport. In this thesis we consider algebraic closure relations only, relating friction and bedload transport to the local values of the hydrodynamic variables.
In section 2.1.1 we present the governing equations, while the closure relations are presented in section 2.1.2.

2.1.1 Governing equations

The governing equations are obtained imposing mass conservation for the fluid and solid phases and the momentum principle to a dilute suspension of sediment particles flowing in open channels. The water mass conservation equation (continuity equation) reads

$$\partial_t (H-\eta) + \partial_x q_x + \partial_y q_y = 0 \tag{2.1}$$

and the momentum equations in the x and y direction respectively read

$$\partial_t q_x + \partial_x \left(\frac{q_x^2}{H-\eta} + \frac{1}{2} g H^2 - g H \eta \right) + \partial_y \left(\frac{q_x q_y}{H-\eta} \right) + g H \, \partial_x \eta + g (H-\eta) S_{fx} = 0 , \tag{2.2}$$

$$\partial_t q_y + \partial_x \left(\frac{q_y q_x}{H-\eta} \right) + \partial_y \left(\frac{q_y^2}{H-\eta} + \frac{1}{2} g H^2 - g H \eta \right) + g H \, \partial_y \eta + g (H-\eta) S_{fy} = 0 . \tag{2.3}$$

Equations (2.1)-(2.3) constitute the two-dimensional shallow water model. Finally, the sediment mass conservation equation (two-dimensional Exner equation) is

$$(1-\lambda_p) \, \partial_t \eta + \partial_x q_{sx}^* + \partial_y q_{sy}^* = 0 . \tag{2.4}$$

In equations (2.1)-(2.4), x and y are the two Cartesian planimetric coordinates of space and t is time, $H(x,y,t)$ $[m]$ denotes water free-surface elevation, $\eta(x,y,t)$ $[m]$ is bed elevation, $q_x(x,y,t)$ and $q_y(x,y,t)$ $\left[m^2 s^{-1}\right]$ represent water discharge per unit width in the x and y direction respectively, $S_{fx}(x,y,t)$ and $S_{fy}(x,y,t)$ $[-]$ are the dimensionless friction terms in the x and y direction, $q^*_{sx}(x,y,t)$ and $q^*_{sy}(x,y,t)$ $\left[m^2 s^{-1}\right]$ represent sediment discharge per unit width in the x and y direction, $g = 9.81\, ms^{-2}$ is the acceleration due to gravity and λ_p $[-]$ is bed porosity. Moreover, we denote with $D(x,y,t) = H - \eta$ $[m]$ the water depth and with $u(x,y,t) = \frac{q_x}{H-\eta}$ and $v(x,y,t) = \frac{q_y}{H-\eta}$ $\left[ms^{-1}\right]$ the x and y components of velocity. The notation is illustrated in Fig. 2.1.

The final form of the system is obtained after performing slight modifications to equations (2.1) and (2.4). First in (2.4) we incorporate the porosity term $\frac{1}{1-\lambda_p}$ inside bedload discharge, thus redefining the sediment discharge terms as

$$q_{sx} = \frac{q^*_{sx}}{1-\lambda_p} \qquad , \qquad q_{sy} = \frac{q^*_{sy}}{1-\lambda_p} \, . \tag{2.5}$$

Then we replace equation (2.1) with one equation resulting from the linear combination of (2.1) and (2.4) representing joint conservation of water and sediment mass. Finally, we obtain the shallow water-Exner model in the form

$$\begin{cases} \partial_t H + \partial_x (q_x + q_{sx}) + \partial_y (q_y + q_{sy}) = 0 \\ \partial_t q_x + \partial_x \left(\dfrac{q_x^2}{H-\eta} + \dfrac{1}{2} g H^2 - g H \eta \right) + \partial_y \left(\dfrac{q_x q_y}{H-\eta} \right) + g H \, \partial_x \eta + g (H-\eta) S_{fx} = 0 \\ \partial_t q_y + \partial_x \left(\dfrac{q_y q_x}{H-\eta} \right) + \partial_y \left(\dfrac{q_y^2}{H-\eta} + \dfrac{1}{2} g H^2 - g H \eta \right) + g H \, \partial_y \eta + g (H-\eta) S_{fy} = 0 \\ \partial_t \eta + \partial_x q_{sx} + \partial_y q_{sy} = 0 \end{cases} . \tag{2.6}$$

System (2.6) presents eight unknowns (H, q_x, q_y, η, S_{fx}, S_{fy}, q_{sx}, q_{sy}) and only four equations, thus being four times undetermined. Next, we shall introduce closure relations for the evaluation of the friction and sediment discharge terms.

2.1.2 Closure relations

For sake of simplicity, here we introduce a vectorial formulation for discharge $\vec{q}$, bedload discharge $\vec{q}_s$ and dimensionless friction $\vec{S}_f$

$$\vec{q} = (q_x, q_y) \qquad , \qquad \vec{q}_s = (q_{sx}, q_{sy}) \qquad , \qquad \vec{S}_f = (S_{fx}, S_{fy}) \tag{2.7}$$

and define the magnitude of the above vectors as

$$q = \sqrt{q_x^2 + q_y^2} \qquad , \qquad q_s = \sqrt{q_{sx}^2 + q_{sy}^2} \qquad , \qquad S_f = \sqrt{S_{fx}^2 + S_{fy}^2} \, . \tag{2.8}$$

We remark that velocity as well is a vector parallel to discharge:

$$\vec{u} = (u, v) = \frac{1}{H - \eta} (q_x, q_y) \, . \tag{2.9}$$

Such notation allows two-dimensional projection of classical one-dimensional closure relations, as we shall see next.

2.1.2.1 Friction models

For the evaluation of the magnitude of friction S_f we adopt one-dimensional algebraic relations such as the Manning law, which reads

$$S_f = \frac{n^2 q^2}{(H - \eta)^{10/3}} \, , \tag{2.10}$$

where $n \left[m^{-1/3} s \right]$ is the Manning roughness coefficient. The Strickler formula relates the n coefficient to median grain size D_s in the form

$$\frac{1}{n} = Ks = \frac{21.1}{D_s^{1/6}} \, . \tag{2.11}$$

Alternatively, we consider the Chézy friction law, reading

$$S_f = \frac{q^2}{g C_h^2 (H - \eta)^3} \, . \tag{2.12}$$

In (2.12) the Chézy coefficient $C_h \, [-]$ can be taken as constant in the range $C_h \sim 10 \div 20$. Otherwise, we can assume logarithmic dependence of C_h on the water depth $D = H - \eta$ in the form

$$C_h = 6 + 2.5 \ln \left(\frac{D}{k_{C_h} D_s} \right) \, , \tag{2.13}$$

where k_{C_h} is usually $k_{C_h} = 2.5$.

For the projection of the one-dimensional friction laws (2.10) and (2.12) in the x and y direction, consistently with the depth-averaged approach in which we neglect the role of three-dimensional

circulations, we assume that $\vec{S}_f$ is parallel to $\vec{q}$, thus obtaining

$$S_{fx} = \frac{n^2 q_x q}{(H-\eta)^{10/3}} \quad , \quad S_{fy} = \frac{n^2 q_y q}{(H-\eta)^{10/3}} \tag{2.14}$$

if the Manning law (2.10) is used and

$$S_{fx} = \frac{q_x q}{g C_h^2 (H-\eta)^3} \quad , \quad S_{fy} = \frac{q_y q}{g C_h^2 (H-\eta)^3} \tag{2.15}$$

if the Chézy law (2.12) is used.

2.1.2.2 Bedload transport formulae

In order to define the magnitude of the sediment discharge vector q_s (2.8), we assume that

- sediment discharge is in equilibrium with local hydrodynamic conditions (instantaneous adaptation);
- sediment discharge is locally equivalent to sediment transport capacity.

The first hypothesis, which is widely accepted when dealing with bedload transport, allows adoption of algebraic closure relations for sediment transport, avoiding use of differential relations for adaptation. The second hypothesis allows us to evaluate q_s through one of the existing one-dimensional relations for bedload transport capacity. Plenty of different formulae for sediment transport capacity are available in the literature. Most of them are given in the form

$$q_s = \frac{\sqrt{g\Delta D_s^3}}{1-\lambda_p} \Phi(\theta) \ , \tag{2.16}$$

where $\Phi(\theta)$ $[-]$ is the Einstein parameter. Relations for $\Phi(\theta)$ of three bedload transport formulae are given in Tab. 2.1. In (2.16) Δ is reduced density, defined as

$$\Delta = \frac{\rho_s - \rho_w}{\rho_w} \ , \tag{2.17}$$

being ρ_s and ρ_w sediment and water density, and θ is the Shields parameter representing dimensionless shear stress, given by

$$\theta = \frac{\tau}{(\rho_s - \rho_w) g D_s} \ , \tag{2.18}$$

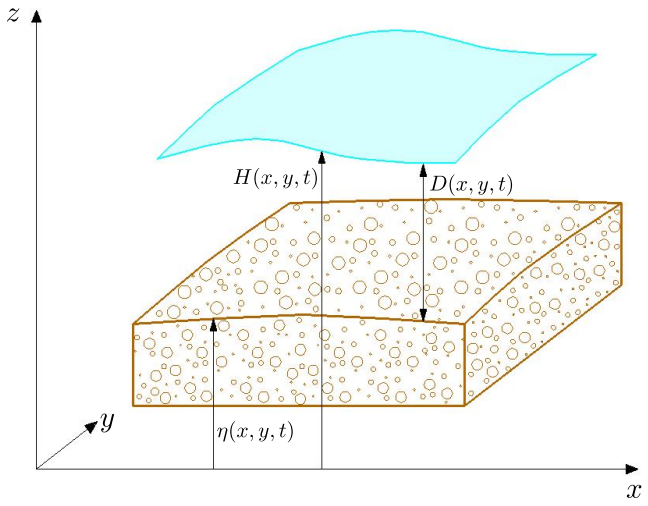

Figure 2.1: Sketch of free surface and river bed with notation.

Author	**Formula**	**Remarks**
Parker [96]	$\Phi = G(\xi)\,0.00218\theta^{1.5}$ with $\xi = \theta/0.0386$ and	
	$G = \xi^{14.2}$	for $\xi < 1$
	$G = \exp\left(14.2\,(\xi-1) - 9.28\,(\xi-1)^2\right)$	for $1 \leq \xi \leq 1.59$
	$G = 5474\,(1-0.853/\xi)^{4.5}$	for $\xi \geq 1.59$
Meyer-Peter and Müller [85]	$\Phi = 8\max\left((\theta-\theta_{cr}),\,0\right)^{\frac{3}{2}}$	$\theta_{cr} = 0.047$. Bed slope ≤ 0.02. Bedload transport.
Wong and Parker [135]	$\Phi = 4.93\max\left((\theta-\theta_{cr}),\,0\right)^{1.6}$	$\theta_{cr} = 0.047$

Table 2.1: Einstein parameter $\Phi(\theta)$ of three sediment transport formulae.

where $\tau = \rho_w g\,(H-\eta)\,S_f$ is the dimensional shear stress at the interface between fluid and bed. Depending on the friction law adopted in the evaluation of τ, (2.18) specialises as

$$\theta = \frac{n^2 q^2}{\Delta D_s\,(H-\eta)^{\frac{7}{3}}} \tag{2.19}$$

if the Manning law (2.10) is used and as

$$\theta = \frac{q^2}{g\Delta D_s C_h^2\,(H-\eta)^2} \tag{2.20}$$

if the Chézy law (2.12) is adopted. In Tab. 2.1 $\theta_{cr} = 0.047$ represents the critical value of the Shields stress. Notice that the formula proposed by Wong and Parker [135] has the same structure as that of Meyer-Peter and Müller [85], but different coefficients.
Unlike these transport formulae, where q_s is given as a function of θ in the form (2.16), relations of more theoretical use relate bedload transport capacity to local values of depth and of the magnitude of discharge. Among them, the Grass formula [54] reads

$$q_s = A_G \left(\frac{q}{H-\eta}\right)^{m_G} \tag{2.21}$$

with $A_G \in [0,\,1]$, $m_G \in [1,\,4]$.

Having defined the magnitude of bedload transport q_s using one of the above formulae, two-dimensional projection is achieved setting

$$\vec{q}_s = (q_{sx}, q_{sy}) = q_s(\cos\gamma, \sin\gamma)\ , \tag{2.22}$$

where γ represents the angle between the direction defined by vector $\vec{q}_s$ and the positive direction of the x axis. According to theoretical and experimental studies, e.g. Ikeda *et al.* [66] and Talmon *et al.* [116], γ can be computed as the sum of two contributions

$$\gamma = \gamma_q + \gamma_g\ . \tag{2.23}$$

See Fig. 2.2, where these definitions and the notation are shown. γ_q represents the direction defined by the shear stress, which is parallel to $\vec{q}$:

$$\cos\gamma_q = \frac{q_x}{q} \qquad , \qquad \sin\gamma_q = \frac{q_y}{q}\ , \tag{2.24}$$

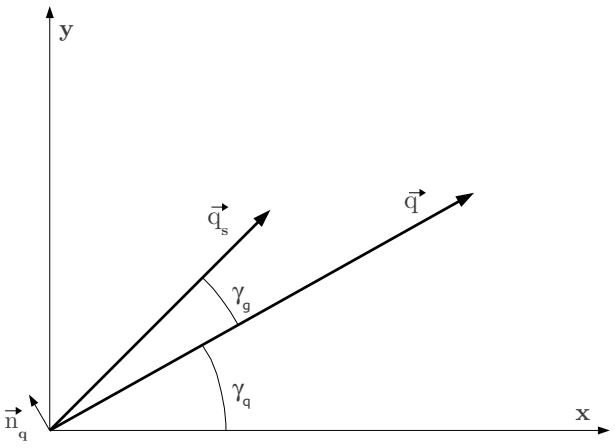

Figure 2.2: Sketch of vectors $\vec{q}$ and $\vec{q}_s$.

while γ_g represents a correction to the direction defined by $\vec{q}$, which accounts for the action of gravity on sediments due to local lateral bed slope. According to Ikeda [64] the latter contribution can be computed as

$$\tan\gamma_g = -\frac{r_{ik}}{\sqrt{\theta}}\frac{\partial \eta}{\partial \vec{n}_q} = -\frac{r_{ik}}{\sqrt{\theta}}\left(\frac{\partial \eta}{\partial x}, \frac{\partial \eta}{\partial y}\right)\cdot \vec{n}_q \,, \tag{2.25}$$

where $\vec{n}_q = (-\sin\gamma_q, \cos\gamma_q) = \frac{(-q_y, q_x)}{q}$ is the unit vector normal to $\vec{q}$ and r_{ik} is a dimensionless empirical parameter ranging between 0.3 and 0.6 [64, 116].
Two-dimensional projection of the Grass formula (2.21) is typically achieved neglecting the role of bed lateral slope in (2.22), i.e. assuming $\vec{q}_s$ to be parallel to $\vec{q}$ ($\gamma = \gamma_q$):

$$q_{sx} = A_G \frac{q^{m_G-1}}{(H-\eta)^{m_G}} q_x \qquad , \qquad q_{sy} = A_G \frac{q^{m_G-1}}{(H-\eta)^{m_G}} q_y \,. \tag{2.26}$$

Finally, the mathematical model is given by system (2.6), together with the algebraic closure relations for friction and bedload transport presented in this section. The model has four unknowns, namely H, q_x, q_y, η.

2.2 The two-dimensional inviscid shallow water equations

In this section we introduce a model describing two-dimensional free-surface shallow flows in absence of friction and sediment transport over horizontal bed, namely the two-dimensional inviscid

shallow water equations. This model is unsuitable for addressing the morphodynamic evolution of river systems since it is unable to predict changes in the riverbed configuration. Moreover, the assumption of inviscid flow restricts its applications to the study of inertial hydrodynamic processes, i.e. the processes dominated by advection. However, the study of free-surface flows over non-movable bed in idealised conditions has strong relevance for the development of the numerical hydro-morphodynamic solver and for the morphodynamic applications which are the final goal of this thesis. There are three main reasons for considering the shallow water model.

- The classical numerical solution strategy for hydro-morphodynamic problems consists of applying a numerical solver for the shallow water equations with fixed bed and subsequently updating riverbed elevation by solving the Exner equation (*uncoupled solution strategy*, see sections 2.5 and 5.1).

- The general mathematical structure of advection problems, describing wave propagation across the domain, can be much more easily studied with reference to this system than using the full shallow water-Exner model. The shallow water equations are a remarkable example of hyperbolic system of PDEs: in section 2.3, we will use this case in order to introduce the class of hyperbolic systems of PDEs and review their general properties.

- Part of this thesis (chapters 3 and 4) is devoted to the development of numerical schemes for hyperbolic systems of PDEs, whose performance will be assessed in a suite of tests for the inviscid shallow water equations on horizontal bed.

The model is derived from the mass conservation equation (2.1) and the momentum equations (2.2)-(2.3), in which we neglect the frictional source terms ($S_{fx} = S_{fy} = 0$) and riverbed elevation gradients ($\partial_x \eta = \partial_y \eta = 0$). It reads

$$\begin{cases} \partial_t D + \partial_x (Du) + \partial_y (Dv) = 0 \\ \partial_t (Du) + \partial_x \left(Du^2 + \frac{1}{2} g D^2 \right) + \partial_y (Duv) = 0 \\ \partial_t (Dv) + \partial_x (Duv) + \partial_y \left(Dv^2 + \frac{1}{2} g D^2 \right) = 0 \end{cases} . \tag{2.27}$$

Alternatively, system (2.27) can be equivalently written in terms of the variables H, q_x, q_y. Notice that, as a consequence of the assumption of inviscid flow and horizontal bed configuration, the equations do not include any source term, i.e. the system is homogeneous.

2.3 Hyperbolicity of the shallow water model

In this section, following Toro [117], we introduce the class of hyperbolic systems of partial differential equations (PDEs) and present their fundamental properties with reference to the two-dimensional inviscid shallow water equations (2.27). The motivation for this review is twofold. First, one of the main objectives of the present thesis is to solve the shallow water-Exner model (2.6), which, as we will prove in section 2.4, is hyperbolic under some hypotheses. Furthermore, part of the thesis is devoted to the derivation of numerical schemes for the solution of hyperbolic systems of PDEs: an introduction to their mathematical properties and features is required. For a comprehensive presentation of this subject we refer the reader to [118] and references therein.

The section outlines as follows. First in section 2.3.1 we give some preliminary notions and definitions. The case of conservative hyperbolic systems is treated in section 2.3.2, where we briefly recall the wave relations, we define rarefaction, shear and shock waves and introduce the concept of weak solutions of the system. Finally in section 2.3.3 we focus on hyperbolic systems in non-conservative form and we introduce the concept of weak solution of these systems in the path-conservative framework.

2.3.1 Preliminaries

We consider two-dimensional homogeneous systems of m first-order partial differential equations having the form

$$\partial_t \mathrm{q}_k + \sum_{l=1}^{m} a_{1kl}\left(x,y,t,\mathrm{q}_1,...,\mathrm{q}_m\right)\partial_x \mathrm{q}_l + \sum_{l=1}^{m} a_{2kl}\left(x,y,t,\mathrm{q}_1,...,\mathrm{q}_m\right)\partial_y \mathrm{q}_l = 0\,, \tag{2.28}$$

with $1 \leq k,l \leq m$, where $\mathrm{q}_1,...,\mathrm{q}_m$ are the conserved variables.
An example is given by the two-dimensional inviscid shallow water equations over horizontal bed (2.27).

Let us recast system (2.28) in quasi-linear or non-conservative form

$$\partial_t \mathbf{Q} + \mathbf{A}_1 \partial_x \mathbf{Q} + \mathbf{A}_2 \partial_y \mathbf{Q} = \mathbf{0}\,, \tag{2.29}$$

where $\mathbf{Q}$ is the vector of unknowns, $\mathbf{A}_1$ and $\mathbf{A}_2$ are the system matrices. For the shallow water equations (2.27) they read

$$\mathbf{Q} = \begin{bmatrix} D \\ uD \\ vD \end{bmatrix} \quad , \quad \mathbf{A}_1 = \begin{bmatrix} 0 & 1 & 0 \\ gD - u^2 & 2u & 0 \\ -uv & v & u \end{bmatrix} \quad , \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 0 & 1 \\ -uv & v & u \\ gD - v^2 & 0 & 2v \end{bmatrix} . \tag{2.30}$$

Definition 2.3.1: Eigenvalues. The eigenvalues of a matrix $\mathbf{A}$ having size $m \times m$ are given by the m roots of the characteristic polynomial:

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}) = 0 , \tag{2.31}$$

where $\mathbf{I}$ is the $m \times m$ identity matrix. We denote these eigenvalues, sorted in increasing order, with $\lambda^{(1)}, ..., \lambda^{(m)}$. For the shallow water equations the eigenvalues of $\mathbf{A}_1$ and $\mathbf{A}_2$ (2.30) are respectively given by:

$$\lambda_x^{(1)} = u - a \quad , \quad \lambda_x^{(2)} = u \quad , \quad \lambda_x^{(3)} = u + a , \tag{2.32}$$

$$\lambda_y^{(1)} = v - a \quad , \quad \lambda_y^{(2)} = v \quad , \quad \lambda_y^{(3)} = v + a , \tag{2.33}$$

being $a = \sqrt{gD}$.

Consider a non-zero vector $\vec{n} = (n_1, n_2)$. Without loss of generality, let us normalise $\vec{n}$, i.e. consider a unit vector, whose norm is $|\vec{n}| = \sqrt{n_1^2 + n_2^2} = 1$. We define a matrix $\mathbf{A}_{\hat{n}}$ that is the projection of matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ (2.29) in the direction of $\vec{n}$, i.e. the linear combination

$$\mathbf{A}_{\hat{n}} = (\mathbf{A}_1, \mathbf{A}_2) \cdot \vec{n} = n_1 \mathbf{A}_1 + n_2 \mathbf{A}_2 . \tag{2.34}$$

Projection of the system matrices in (2.34) allows generalisation of the theory from the one-dimensional to the two-dimensional case.

Definition 2.3.2: Hyperbolic systems of PDEs. A system of PDEs in quasi-linear form (2.29) having matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ is said to be hyperbolic if matrix $\mathbf{A}_{\hat{n}}$ (2.34) has m real eigenvalues. In this case we also call matrices $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_{\hat{n}}$ hyperbolic matrices. Furthermore, the system is called strictly hyperbolic if in addition the eigenvalues are all distinct.

For the two-dimensional inviscid shallow water equations matrix $\mathbf{A}_{\hat{n}}$ reads

$$\mathbf{A}_{\hat{n}}(\mathbf{Q}) = \begin{bmatrix} 0 & n_1 & n_2 \\ \left(gD-u^2\right)n_1 - uvn_2 & 2un_1+vn_2 & un_2 \\ -uvn_1 + \left(gD-v^2\right)n_2 & vn_1 & un_1+2vn_2 \end{bmatrix}, \tag{2.35}$$

having eigenvalues

$$\lambda_{\hat{n}}^{(1)} = un_1+vn_2-a \quad , \quad \lambda_{\hat{n}}^{(2)} = un_1+vn_2 \quad , \quad \lambda_{\hat{n}}^{(3)} = un_1+vn_2+a\,. \tag{2.36}$$

Since the eigenvalues (2.36) are all real, the shallow water equations prove to be a hyperbolic system of PDEs . Moreover, for wet bed, i.e. provided $a>0$, the system is strictly hyperbolic.

Definition 2.3.3: Right eigenvectors. The right eigenvector of a matrix $\mathbf{A}$ (having size $m\times m$) associated to one of its eigenvalues $\lambda^{(l)}$ is the column vector $\mathbf{R}^{(l)} = [r_1,...,r_m]^T$ such that

$$\mathbf{A}\mathbf{R}^{(l)} = \lambda^{(l)}\mathbf{R}^{(l)}\,. \tag{2.37}$$

For the two-dimensional shallow water equations, the eigenvectors of $\mathbf{A}_1$ and $\mathbf{A}_2$ (2.30) respectively read

$$\mathbf{R}_x^{(1)} = \alpha_x^{(1)}\begin{bmatrix}1\\u-a\\v\end{bmatrix} \quad , \quad \mathbf{R}_x^{(2)} = \alpha_x^{(2)}\begin{bmatrix}0\\0\\1\end{bmatrix} \quad , \quad \mathbf{R}_x^{(3)} = \alpha_x^{(3)}\begin{bmatrix}1\\u+a\\v\end{bmatrix} \tag{2.38}$$

and

$$\mathbf{R}_y^{(1)} = \alpha_y^{(1)}\begin{bmatrix}1\\u\\v-a\end{bmatrix} \quad , \quad \mathbf{R}_y^{(2)} = \alpha_y^{(2)}\begin{bmatrix}0\\1\\0\end{bmatrix} \quad , \quad \mathbf{R}_y^{(3)} = \alpha_y^{(3)}\begin{bmatrix}1\\u\\v+a\end{bmatrix}, \tag{2.39}$$

being $\alpha_x^{(1)}$, $\alpha_x^{(2)}$, $\alpha_x^{(3)}$ and $\alpha_y^{(1)}$, $\alpha_y^{(2)}$, $\alpha_x^{(3)}$ scaling factors. The eigenvectors of the projected matrix $\mathbf{A}_{\hat{n}}$ (2.35) read:

$$\mathbf{R}_{\hat{n}}^{(1)} = \alpha_{\hat{n}}^{(1)}\begin{bmatrix}1\\u-an_1\\v-an_2\end{bmatrix} \quad , \quad \mathbf{R}_{\hat{n}}^{(2)} = \alpha_{\hat{n}}^{(2)}\begin{bmatrix}0\\n_2\\-n_1\end{bmatrix} \quad , \quad \mathbf{R}_{\hat{n}}^{(3)} = \alpha_{\hat{n}}^{(3)}\begin{bmatrix}1\\u+an_1\\v+an_2\end{bmatrix}, \tag{2.40}$$

with $\alpha_{\hat{n}}^{(1)}$, $\alpha_{\hat{n}}^{(2)}$, $\alpha_{\hat{n}}^{(3)}$ scaling factors.

Consider a hyperbolic system (2.29) with real eigenvalues $\lambda_{\hat{n}}^{(l)}(\mathbf{Q})$ and real eigenvectors $\mathbf{R}_{\hat{n}}^{(l)}(\mathbf{Q})$ for $1 \leq l \leq m$. The characteristic speed $\lambda_{\hat{n}}^{(l)}(\mathbf{Q})$ defines a characteristic field, the $\lambda_{\hat{n}}^{(l)}$ field or simply the l field.

Definition 2.3.4: Linearly degenerate fields. A characteristic field is said to be linearly degenerate if

$$\nabla\lambda_{\hat{n}}^{(l)}(\mathbf{Q}) \cdot \mathbf{R}_{\hat{n}}^{(l)}(\mathbf{Q}) = 0 \qquad , \qquad \forall \mathbf{Q} \in \mathrm{R}^m \,, \tag{2.41}$$

where R^m is the set of real-valued vectors of m components, called the phase space.

Definition 2.3.5: Genuinely non-linear fields. A characteristic field is said to be genuinely non-linear if

$$\nabla\lambda_{\hat{n}}^{(l)}(\mathbf{Q}) \cdot \mathbf{R}_{\hat{n}}^{(l)}(\mathbf{Q}) \neq 0 \qquad , \qquad \forall \mathbf{Q} \in \mathrm{R}^m \,. \tag{2.42}$$

In equations (2.41) and (2.42) we recall that the gradient of an eigenvalue $\lambda^{(l)}$ is a column vector given by

$$\nabla\lambda^{(l)}(\mathbf{Q}) = \left[\frac{\partial\lambda^{(l)}}{\partial\mathbf{Q}}\right] \,. \tag{2.43}$$

For the two-dimensional inviscid shallow water equations we have that the eigenvalues $\lambda_{\hat{n}}^{(1)}$ and $\lambda_{\hat{n}}^{(3)}$ define genuinely non-linear fields, while the eigenvalue $\lambda_{\hat{n}}^{(2)}$ defines a linearly degenerate field. The proof follows trivially from the definitions (2.41) and (2.42). Notice that the same definitions apply to the eigenvalues $\lambda_x^{(l)}$ and $\lambda_y^{(1)}$ of the x-split and y-split systems respectively. This is readily obtained setting $\vec{n} = (1,0)$ or $\vec{n} = (0,1)$ in the projection.

2.3.2 Conservative formulation

Definition 2.3.6: Conservative systems. A two-dimensional system of PDEs (2.28) is said to be conservative if it can be recast in the form

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_x \mathbf{G}(\mathbf{Q}) = \mathbf{0} \,, \tag{2.44}$$

where $\mathbf{F}(\mathbf{Q})$ and $\mathbf{G}(\mathbf{Q})$ are the x and y flux vectors. Alternatively, (2.44) can be called system of conservation laws or system in conservation-law form.

If a system admits a conservative form, it can still be written in quasi-linear form (2.29), based on

the relation

$$\mathbf{A}_1(\mathbf{Q}) = \left[\frac{\partial \mathbf{F}}{\partial \mathbf{Q}}\right] \quad , \quad \mathbf{A}_2(\mathbf{Q}) = \left[\frac{\partial \mathbf{G}}{\partial \mathbf{Q}}\right] . \tag{2.45}$$

In this case matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are called Jacobian matrices of the flux vectors.

The two-dimensional inviscid shallow water equations over horizontal bed (2.27) can be written in conservation-law form (2.44), with flux vectors given by

$$\mathbf{F}(\mathbf{Q}) = \begin{bmatrix} Du \\ Du^2 + \frac{1}{2}gD^2 \\ Duv \end{bmatrix} \quad , \quad \mathbf{G}(\mathbf{Q}) = \begin{bmatrix} Dv \\ Duv \\ Dv^2 + \frac{1}{2}gD^2 \end{bmatrix} , \tag{2.46}$$

while the vector of unknowns $\mathbf{Q}$ is given in (2.30).

We shall now move to the definition of the Riemann problem and its elementary wave solutions. For sake of simplicity, let us consider the x-split two-dimensional homogeneous shallow water equations, having the form

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) = \mathbf{0} \tag{2.47}$$

with fluxes $\mathbf{F}$ given in (2.46) and unknowns $\mathbf{Q}$ given in (2.30).

Definition 2.3.7: Riemann problem. The Riemann problem is an initial value problem (IVP) having the form

$$\left.\begin{array}{ll} \text{PDEs:} & \partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) = \mathbf{0} \\ \text{IC:} & \mathbf{Q}(x,0) = \left\{ \begin{array}{ll} \mathbf{Q}_L & \text{if } x < 0 \\ \mathbf{Q}_R & \text{if } x > 0 \end{array} \right. \end{array}\right\} \tag{2.48}$$

where the initial condition

$$\mathbf{Q}_L = \begin{bmatrix} D_L \\ D_L u_L \\ D_L v_L \end{bmatrix} \quad , \quad \mathbf{Q}_R = \begin{bmatrix} D_R \\ D_R u_R \\ D_R v_R \end{bmatrix} \tag{2.49}$$

is discontinuous in $x = 0$. For the shallow water equations the Riemann problem generalises the (physical) dam-break problem, in which a dam or gate separating two basins filled with still water at rest having different free-surface elevation is suddenly removed. In contrast to the dam-break problem, in the Riemann problem (2.48) initial velocity may assume any value.

At time $t = 0^+$, the problem (2.48) results in the generation of waves propagating from the

initial discontinuity located at $x = 0$. In the present case, four different wave patterns may arise [117]. In general, extrema of these waves pattern are shock waves or rarefaction waves: thus we may have two shocks or two rarefactions or two combinations of one shock and one rarefaction. In any of these cases, the middle wave is always a shear wave, across which the tangential velocity v changes discontinuously. Thus, there are three wave families, which are associated with the eigenvalues $\lambda_x^{(1)}$, $\lambda_x^{(2)}$, $\lambda_x^{(3)}$ respectively. Next, we will consider the much simpler case in which the initial data states (2.49) are connected by one single wave.

2.3.2.1 Wave relations

Before proceeding to the study of waves, we recall the notion of Riemann invariants and the Rankine-Hugoniot condition.

Let us turn our x-split system (2.47) in quasi-linear form

$$\partial_t \mathbf{Q} + \mathbf{A}_1(\mathbf{Q})\,\partial_x \mathbf{Q} = \mathbf{0}\,, \tag{2.50}$$

with vector of unknowns given by

$$\mathbf{Q} = [\mathrm{q}_1, \mathrm{q}_2, ..., \mathrm{q}_m]^T \tag{2.51}$$

and consider the wave associated with the $\lambda_x^{(l)}(\mathbf{Q})$ characteristic field, with eigenvalue $\lambda_x^{(l)}$ and corresponding right eigenvector

$$\mathbf{R}_x^{(l)} = \left[r_1^{(l)}, r_2^{(l)}, ..., r_m^{(l)}\right]^T\,. \tag{2.52}$$

Definition 2.3.8: Generalised Riemann invariants. The generalised Riemann invariants are relations that hold across the wave structure (for certain types of waves) and lead to the following $m-1$ ordinary differential equations:

$$\frac{\mathrm{dq}_1}{r_1^{(l)}} = \frac{\mathrm{dq}_2}{r_2^{(l)}} = ... = \frac{\mathrm{dq}_\mathrm{m}}{r_1^{(m)}}\,. \tag{2.53}$$

They relate rates of change $\mathrm{dq_k}$ of a quantity q_k to the respective components $r_k^{(l)}$ of the right eigenvector $\mathbf{R}_x^{(l)}$ corresponding to a $\lambda_x^{(l)}$ wave family.

An important concept which is applicable to discontinuous solutions of hyperbolic systems *in conservation-law form* (2.47) is the *Rankine-Hugoniot jump condition* or simply Rankine-Hugoniot

condition. It applies to a discontinuous wave travelling with speed s_x, which is related to jumps in the conserved variables $\mathbf{Q}$ and fluxes $\mathbf{F}(\mathbf{Q})$ across the wave as follows:

$$\mathbf{F}(\mathbf{Q}_{ahead}) - \mathbf{F}(\mathbf{Q}_{behind}) = \mathrm{s}_x(\mathbf{Q}_{ahead} - \mathbf{Q}_{behind}) \ . \tag{2.54}$$

Here subscript *ahead* denotes the state immediately ahead of the discontinuity while *behind* denotes the state immediately behind the discontinuity. We remark that the Rankine-Hugoniot condition holds only for systems in conservation-law form.
We will now turn to the study of elementary wave solutions of the Riemann problem (2.48) for the shallow water equations.

2.3.2.2 Rarefaction waves

Here we are interested in the situation in which the two data states are connected, through a smooth transition, in a genuinely non-linear field, say, a rarefaction wave. A rarefaction wave is a smooth wave, i.e. all flow quantities vary continuously across the wave, at any fixed time. Across the bounding characteristic corresponding to the wave head and tail, however, all flow variables have a discontinuity in the x derivative. For the shallow water equations, rarefactions are associated with the genuinely non-linear field $l = 1$ and $l = 3$. They satisfy the following conditions:

- constancy of generalised Riemann invariants across the wave (2.53);
- divergence of characteristics:

$$\lambda_x^{(l)}(\mathbf{Q}_L) < \lambda_x^{(l)}(\mathbf{Q}_R) \ . \tag{2.55}$$

Condition (2.55) says that the corresponding eigenvalue increases monotonically as the wave is crossed from left to right.

2.3.2.3 Shear waves

Shear waves are discontinuous solutions across which the tangential velocity component v jumps discontinuously. Suppose that the two states on the left and right sides of the wave are respectively denoted by $\mathbf{Q}_{*L}$ and $\mathbf{Q}_{*R}$. The states on either side of the discontinuity are connected through a single jump discontinuity of speed $\mathrm{s}_x^{(l)}$, in a linearly degenerate field. This field is that of the eigenvalue $\lambda_x^{(2)}$ for the two-dimensional x-split shallow water equations. The following conditions apply:

- the Rankine-Hugoniot condition

$$\mathbf{F}(\mathbf{Q}_{*R}) - \mathbf{F}(\mathbf{Q}_{*L}) = \mathrm{s}_x^{(2)}(\mathbf{Q}_{*R} - \mathbf{Q}_{*L}) \ ; \tag{2.56}$$

- constancy of the generalised Riemann invariants across the wave (2.53);
- the parallel characteristic condition:

$$\lambda_x^{(2)}(\mathbf{Q}_{*L}) = \lambda_x^{(2)}(\mathbf{Q}_{*R}) = \mathrm{s}_x^{(2)} \ . \tag{2.57}$$

From the above relations we obtain that the water depth and the normal component of velocity u are constant across the wave, while the tangential component v changes discontinuously. Contact waves or discontinuities arising from pollutant transport models behave identically to shear waves. See section 3.5 of chapter 3 for more details about contact waves.

2.3.2.4 Shock waves

Here we assume that the solution of the Riemann problem (2.48) consists of an isolated shock of wave speed $\mathrm{s}_x^{(l)}$. The two constant data states $\mathbf{Q}_L$ and $\mathbf{Q}_R$ are connected through a single jump discontinuity in a genuinely non-linear field $l = 1$ or $l = 3$ and the following conditions apply:

- the Rankine-Hugoniot condition

$$\mathbf{F}(\mathbf{Q}_R) - \mathbf{F}(\mathbf{Q}_L) = \mathrm{s}_x^{(l)}(\mathbf{Q}_R - \mathbf{Q}_L) \ ; \tag{2.58}$$

- the entropy condition

$$\lambda_x^{(l)}(\mathbf{Q}_L) > \mathrm{s}_x^{(l)} > \lambda_x^{(l)}(\mathbf{Q}_R) \ . \tag{2.59}$$

The latter condition says that characteristics of both sides shall run into the shock wave. The entropy condition is the criterion for selecting the physically meaningful solution, in case of non-unique solutions. Hyperbolic conservation laws admit, for instance, rarefaction shocks and compression shocks, but only the latter ones are physically acceptable.

2.3.2.5 Weak solutions and convergence of conservative methods

We have been focusing on hyperbolic systems of PDEs expressed in differential form (2.28). An implicit assumption when dealing with conservation laws in differential form is that their solution is smooth enough, i.e. it admits the definition of the derivatives contained in the mathematical problem. However, hyperbolic systems of conservation laws may develop discontinuities in the solution even starting from purely continuous initial conditions, thus making the differential formulation inadequate.

Here we concern about expression of the solutions of conservation laws in weak (or integral)

form. Consider the two-dimensional x-split shallow water equations and a one-dimensional space control volume $[x_L, x_R]$. Integration of the system of PDEs over the selected control volume gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{x_L}^{x_R}\mathbf{Q}(x,t) = \mathbf{F}(\mathbf{Q}(x_L,t)) - \mathbf{F}(\mathbf{Q}(x_R,t)) \ . \tag{2.60}$$

Further integration in time of (2.60) in the set $\left[t^1, t^2\right]$ gives

$$\int_{x_L}^{x_R}\mathbf{Q}\left(x,t^2\right)\mathrm{d}x = \int_{x_L}^{x_R}\mathbf{Q}\left(x,t^1\right)\mathrm{d}x + \int_{t^1}^{t^2}\mathbf{F}(\mathbf{Q}(x_L,t))\,\mathrm{d}t - \int_{t^1}^{t^2}\mathbf{F}(\mathbf{Q}(x_R,t))\,\mathrm{d}t \ . \tag{2.61}$$

There are important reasons for considering the integral forms of conservation laws. First, the derivation of the governing equations for a wide range of physical problems is based on physical conservation principles originally expressed as integral relations on control volumes. Subsequent expression of these laws in differential form relies on certain hypotheses over regularity of the considered functions, which makes differential forms less general. Thus the definition of the integral form indicates a strategy for designing numerical methods able to capture the discontinuous solutions arising from non-linear hyperbolic systems.
Furthermore the concept of weak solutions is used for assessing convergence of numerical schemes. A basic requirement for numerical schemes solving hyperbolic systems of PDEs is the following: if the approximations produced by the scheme converge to some function as the mesh is refined, then this function must be a weak solution of the system. In the case of systems of conservation laws, the classical Lax-Wendroff theorem [78] ensures that conservative numerical methods have this property.

2.3.3 Non-conservative formulation

Definition 2.3.9: Non-conservative systems. A system of PDEs is said to be non-conservative if it cannot be recast in conservation-law form (2.44).

For this class of systems only the quasi-linear formulation (2.29) is available. Conversely, conservative systems can always be expressed in non-conservative form using (2.45).
Non-conservative hyperbolic systems arise in a wide range of applications, which makes their theoretical study and numerical approximation a very important topic. To this category, for instance, belongs the coupled shallow water-Exner model (2.6) governing the hydro-morphodynamic evolution of gravel-bed rivers, as we will show in section 2.5. In one space dimension, they assume the form

$$\partial_t\mathbf{Q} + \mathbf{A}\partial_x\mathbf{Q} = \mathbf{0} \ . \tag{2.62}$$

The major problem in solving system (2.62) comes from the presence of the non-conservative product $\mathbf{A}\partial_x\mathbf{Q}$ which poses difficulties even in the definition of weak solutions. Hou and LeFloch [59] proved a theorem which states that non-conservative methods converge towards wrong solutions if shock waves are present. The challenge is thus to generalise the notion of weak solutions, so as to give a correct definition of shock waves. Next we focus on this topic, which is currently matter of debate in the literature.

2.3.3.1 The path-conservative framework

Recently, starting from the pioneering work of Dal Maso *et al.* [42], a rigorous definition of weak solutions has been given provided a family of Lipschitz continuous paths $\mathbf{\Psi}(\mathbf{Q}_L, \mathbf{Q}_R, s)$ in the phase space is prescribed. These paths, connecting two states $\mathbf{Q}_L$ and $\mathbf{Q}_R$, with $s \in [0,\, 1]$, must satisfy some natural regularity conditions, namely:

$$\mathbf{\Psi}(\mathbf{Q}_L, \mathbf{Q}_R, 0) = \mathbf{Q}_L \qquad , \qquad \mathbf{\Psi}(\mathbf{Q}_L, \mathbf{Q}_R, 1) = \mathbf{Q}_R \qquad , \qquad \mathbf{\Psi}(\mathbf{Q}_L, \mathbf{Q}_L, s) = \mathbf{Q}_L \,. \tag{2.63}$$

Following Castro *et al.* [30], it is convenient to assume some particular structure on the family of paths:

- given an integral curve $\gamma^{(l)}$ of a linearly degenerate l field and $\mathbf{Q}_L, \mathbf{Q}_R \in \gamma^{(l)}$, the path $\mathbf{\Psi}(\mathbf{Q}_L, \mathbf{Q}_R, s)$ is a parametrisation of the arc of $\gamma^{(l)}$ connecting $\mathbf{Q}_L$ and $\mathbf{Q}_R$;
- given an integral curve $\gamma^{(l)}$ of a genuinely non-linear l field and $\mathbf{Q}_L, \mathbf{Q}_R \in \gamma^{(l)}$, with $\lambda_x^{(l)}(\mathbf{Q}_L) < \lambda_x^{(l)}(\mathbf{Q}_R)$, the path $\mathbf{\Psi}(\mathbf{Q}_L, \mathbf{Q}_R, s)$ is a parametrisation of the arc of $\gamma^{(l)}$ connecting $\mathbf{Q}_L$ and $\mathbf{Q}_R$;
- given a Riemann problem

$$\left.\begin{array}{ll} \text{PDEs:} & \partial_t\mathbf{Q} + \mathbf{A}\partial_x\mathbf{Q} = \mathbf{0} \\ \text{IC:} & \mathbf{Q}(x,0) = \left\{\begin{array}{ll} \mathbf{Q}_L & \text{if } x < 0 \\ \mathbf{Q}_R & \text{if } x > 0 \end{array}\right. \end{array}\right\} \tag{2.64}$$

  having a unique solution consisting of m waves, the curve described by the path $\mathbf{\Psi}(\mathbf{Q}_L, \mathbf{Q}_R, s)$ is equal to the union of the curves corresponding to the paths connecting the constant states across each wave.

Once defined a suitable family of paths, the Rankine-Hugoniot condition can be generalised in the form

$$\mathrm{s}_x(\mathbf{Q}_R - \mathbf{Q}_L) = \int_0^1 \mathbf{A}(\mathbf{\Psi}(\mathbf{Q}_L, \mathbf{Q}_R, s)) \frac{\partial \mathbf{\Psi}}{\partial s}(\mathbf{Q}_L, \mathbf{Q}_R, s)\, \mathrm{d}s \,, \tag{2.65}$$

being s_x the shock speed. Such *extended Rankine-Hugoniot condition* (2.65) allows an acceptable definition of shock waves even for systems expressed in non-conservative form. Notice that if

matrix **A** is the Jacobian of some flux vector, then (2.65) reduces to the classical Rankine-Hugoniot condition for conservative methods (2.54). However, in contrast to the classical Rankine-Hugoniot condition, here the shock speed depends also on the path employed. This represents a major problem while designing shock-capturing numerical schemes based on the path-conservative framework.

2.3.3.2 Convergence of path-conservative schemes

The introduction of the theory of Dal Maso *et al.* [42] gives a way to properly define weak solutions to non-conservative systems (2.62) and indicates a strategy for designing shock-capturing numerical methods for these systems [95]. This framework is based on the concept of path-conservative numerical schemes, which is a generalisation of the usual concept of conservative method for systems of conservation laws.
However, assessment of convergence for general path-conservative schemes is still matter of debate in the literature. We recall that the basic requirement concerning the convergence of a numerical scheme is the following: if the approximations produced by the scheme converge to some function as the mesh is refined, then this function should be a weak solution of the system. For conservative systems this is ensured by the Lax-Wendroff theorem [78]. Castro *et al.* [30] have extended the classical Lax-Wendroff theorem to non-conservative systems proving that the approximations produced by a path-conservative method converge uniformly in the sense of graphs. However, in general this notion of convergence is too strong and cannot be recovered for finite-difference type methods. Moreover failure of convergence of non-conservative schemes has been experimentally reported in the literature (see e.g. Abgrall and Karni [1]).

2.4 Hyperbolicity of the shallow water-Exner model

This section is devoted to determining the nature of the shallow water-Exner model (2.6) together with the closure relations in section 2.1.2. The question is whether or not, or for which range of flow data, the considered system of partial differential equations is hyperbolic. The answer to this question is crucial. For a few classes of systems of PDEs, including hyperbolic systems, a mature mathematical theory assessing well-posedness of the initial value problem (IVP) exists. Moving away from these problems, no mathematical certainty of existence and uniqueness of the IVP solution is available, thus making the numerical approximation of these solutions meaningless. Hence, we aim at proving that the considered mathematical model is hyperbolic, at least in the range of data which are commonly found in applications to gravel-bed rivers.

Unfortunately, in the case of the shallow water-Exner model, unlike for the shallow water equations, the problem eigenstructure is not readily available in explicit form, at least if complex relations

based on the Shields parameter are applied in the evaluation of bedload transport. For this reason, here we are not able to carry out a detailed analysis like that presented in the previous section for the shallow water equations. In spite of these difficulties, the problem at hand has been recently addressed by Cordier *et al.* [35], who have developed simple criteria to determine the hyperbolicity domain of the shallow water-Exner model in the one-dimensional inviscid case, avoiding direct computation of eigenvalues. These methods allow us to assess whether the system is hyperbolic for any given pair of hydraulic variables u, D (velocity and depth). In this section we review and extend their analysis.

The section outlines as follows. First in section 2.4.1 we introduce the one-dimensional inviscid shallow water-Exner model. Then in section 2.4.2 we present and extend the method of Cordier *et al.* [35]. Last in section 2.4.3 we apply these criteria in order to draw the hyperbolicity domains in certain cases. Based on these results, in section 2.4.4 we conclude that the shallow water-Exner model is hyperbolic in the range of data which are of interest in practical applications in real rivers.

2.4.1 The one-dimensional mathematical model

We consider a one-dimensional system of PDEs having three equations and unknowns, describing free-surface shallow flow in absence of friction over erodible bed. The system, representing the one-dimensional inviscid version of the two-dimensional shallow water-Exner system (2.6), reads

$$\begin{cases} \partial_t D + \partial_x (Du) = 0 \\ \partial_t (Du) + \partial_x \left(Du^2 + \frac{1}{2} g D^2 \right) = -gD\partial_x \eta \\ \partial_t \eta + \partial_x q_{sx} = 0 \end{cases} . \tag{2.66}$$

The notation is explained in section 2.1.1.

Algebraic closure relations are needed for the evaluation of bedload transport discharge q_{sx} in (2.66). As we have found in the two-dimensional case (2.16), most bedload transport formulae are given as function of the Shields stress θ in the form

$$q_{sx} = \mathrm{sign}(q_x) \frac{\sqrt{g\Delta D_s^3}}{1-\lambda_p} \Phi(\theta) , \tag{2.67}$$

where reduced density Δ is defined in (2.17) and $\Phi(\theta)$ depends on the transport formula adopted (see Tab. 2.1). The Shields stress θ is given by (2.18) where we set $q = |q_x|$. Evaluation of θ requires the adoption of a closure relation for friction such as the Manning law (2.10) or the Chézy law (2.12), giving (2.19) or (2.20) respectively. Cordier *et al.* [35] consider the Darcy-Weisbach

friction law, having the form

$$S_f = \frac{\chi q^2}{8g\left(H-\eta\right)^3}\,, \tag{2.68}$$

where χ is Darcy-Weisbach's coefficient. This friction formula has identical structure as the Chézy law (2.12) with constant C_h coefficient and reduces to (2.12) imposing $C_h = \sqrt{\frac{8}{\chi}}$.

Coming back to bedload transport formulae, we finally consider the one-dimensional Grass formula, in which q_{sx} does not depend on θ, thus avoiding use of a friction model:

$$q_{sx} = A_G u\left|u\right|^{m_G-1}\,. \tag{2.69}$$

System (2.66) written in quasi-linear form reads:

$$\partial_t \mathbf{Q} + \mathbf{A}\partial_x \mathbf{Q} = \mathbf{0}\,, \tag{2.70}$$

where the vector of conserved variables $\mathbf{Q}$ and the system matrix $\mathbf{A}$ are given by:

$$\mathbf{Q} = \begin{bmatrix} D \\ uD \\ \eta \end{bmatrix} \quad , \quad \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ gD-u^2 & 2u & gD \\ \frac{\partial q_{sx}}{\partial D} & \frac{\partial q_{sx}}{\partial q_x} & 0 \end{bmatrix}\,. \tag{2.71}$$

The third line in matrix $\mathbf{A}$ contains the derivatives of bedload discharge with respect to the conserved variables. For formulae of the type (2.67) these derivatives read:

$$\frac{\partial q_{sx}}{\partial D} = \mathrm{sign}\left(q_x\right)\frac{\sqrt{g\Delta D_s^3}}{1-\lambda_p}\frac{\partial\theta}{\partial D}\frac{\mathrm{d}\Phi}{\mathrm{d}\theta} \quad , \quad \frac{\partial q_{sx}}{\partial q_x} = \mathrm{sign}\left(q_x\right)\frac{\sqrt{g\Delta D_s^3}}{1-\lambda_p}\frac{\partial\theta}{\partial q_x}\frac{\mathrm{d}\Phi}{\mathrm{d}\theta}\,. \tag{2.72}$$

In (2.72) adoption of the Manning law (2.19) in the evaluation of θ yields the following expressions for its derivatives:

$$\frac{\partial\theta}{\partial D} = -\frac{7}{3}\frac{u^2n^2}{\Delta D_s D^{\frac{4}{3}}} \quad , \quad \frac{\partial\theta}{\partial q_x} = \frac{2un^2}{\Delta D_s D^{\frac{4}{3}}}\,, \tag{2.73}$$

while adoption of the Chézy law (2.20) gives

$$\frac{\partial\theta}{\partial D} = -\frac{2u^2}{\Delta D_s g D C_h^2}\left(1+\frac{D}{C_h}\frac{\partial C_h}{\partial D}\right) \quad , \quad \frac{\partial\theta}{\partial q_x} = \frac{2u}{\Delta D_s g D C_h^2}\,. \tag{2.74}$$

Notice that $\frac{\mathrm{d}\Phi}{\mathrm{d}\theta}$ in (2.72) depends on the bedload transport formula in use (see Tab. 2.1).

2.4.2 Criteria for assessing hyperbolicity of the shallow water-Exner model

Here, following Cordier *et al.* [35] we present two different criteria which can be used for assessing hyperbolicity of the one-dimensional inviscid shallow water-Exner system (2.66) by checking the values of the hydrodynamic variables u and D. The first condition, derived in section 2.4.2.1, is the *necessary and sufficient condition* for hyperbolicity, i.e. the condition representing the actual hyperbolicity domain. Moreover, in section 2.4.2.2 we turn the necessary and sufficient condition in dimensionless form, in order to highlight the role of the flow control parameters. Finally, a simplified *sufficient* condition, resulting in reduced hyperbolicity domains, which can be easily applied analytically, is presented in section 2.4.2.3.

2.4.2.1 The necessary and sufficient condition

By *Definition 2.3.2* (see section 2.3.1), the *necessary and sufficient* condition for hyperbolicity of system (2.70) is that the system matrix $\mathbf{A}$ (2.71) possesses three real eigenvalues, i.e. that the three roots of its characteristic polynomial

$$p_{\mathbf{A}}(\lambda) = -\lambda\left((u-\lambda)^2 - gD\right) + gD\left(\frac{\partial q_{sx}}{\partial D}\lambda + \frac{\partial q_{sx}}{\partial q_x}\right) \tag{2.75}$$

are real. It can be shown [35] that $p_{\mathbf{A}}(\lambda)$ is the difference between a linear function

$$d(\lambda) = gD\left(\frac{\partial q_{sx}}{\partial q_x}\lambda + \frac{\partial q_{sx}}{\partial D}\right), \tag{2.76}$$

containing the derivatives of bedload discharge, and a cubic polynomial

$$f(\lambda) = \lambda\left((u-\lambda)^2 - gD\right), \tag{2.77}$$

related to the characteristic polynomial of the fixed-bed model. Therefore the necessary and sufficient condition can be geometrically interpreted as the existence of three real intersection points between these two curves. This condition can be equivalently written as

$$\alpha_- < k_f u < \alpha_+, \tag{2.78}$$

where $\alpha_\pm$ is given by

$$\alpha_\pm = \lambda_\pm - \frac{f(\lambda_\pm)}{gD\frac{\partial q_{sx}}{\partial q_x}} \qquad \text{with} \qquad \lambda_\pm = \frac{2u \pm \sqrt{u^2 + 3gD\left(1 + \frac{\partial q_{sx}}{\partial q_x}\right)}}{3} \tag{2.79}$$

and $f(\lambda_{\pm})$ given in (2.77). Coefficient k_f $[-]$ in (2.78), defined as

$$k_f = -\frac{\partial q_{sx}}{\partial D}\left(u\frac{\partial q_{sx}}{\partial q_x}\right)^{-1} , \tag{2.80}$$

depends on the bedload transport formula and on the friction law adopted for evaluation of Shields stress (if required by the transport formula). In detail:

- the Grass transport formula (2.69) is characterised by $k_f = 1$;
- for the bedload transport formulae based on Shields stress (2.67), use of the Manning law (2.73) gives $k_f = \frac{7}{6}$, while use of the Chézy law (2.74) gives $k_f = 1 + \frac{D}{C_h}\frac{\mathrm{d}C_h}{\mathrm{d}D}$. The latter relation specialises as $k_f = 1$ if the Chézy coefficient is constant and $k_f = 1 + \frac{2.5}{C_h}$ if logarithmic dependence on depth (2.13) is assumed.

2.4.2.2 The necessary and sufficient condition in dimensionless form

We can turn (2.78) into dimensionless form in order to highlight the role of dimensionless flow parameters. Thus, the necessary and sufficient condition for hyperbolicity reads:

$$\widetilde{\alpha}_- < k_f F_r < \widetilde{\alpha}_+ , \tag{2.81}$$

where $F_r = \frac{u}{\sqrt{gD}}$ is the flow Froude number and $\widetilde{\alpha}_{\pm} = \frac{\alpha_{\pm}}{\sqrt{gD}}$ $[-]$ is given by

$$\widetilde{\alpha}_{\pm} = \widetilde{\lambda}_{\pm} - \widetilde{f}\left(\widetilde{\lambda}_{\pm}\right)\left(\frac{\partial q_{sx}}{\partial q_x}\right)^{-1} , \tag{2.82}$$

with

$$\widetilde{\lambda}_{\pm} = \frac{2F_r \pm \sqrt{F_r^2 + 3\left(1 + \frac{\partial q_{sx}}{\partial q_x}\right)}}{3} \qquad , \qquad \widetilde{f}\left(\widetilde{\lambda}\right) = \widetilde{\lambda}\left(\left(F_r - \widetilde{\lambda}\right)^2 - 1\right) . \tag{2.83}$$

In (2.82) and (2.83), for the transport formulae based on the Shields stress (2.67), $\frac{\partial q_{sx}}{\partial q_x}$ can be expressed as

$$\frac{\partial q_{sx}}{\partial q_x} = \frac{2\sqrt{\Delta}}{1-\lambda_p} d_s^{\frac{3}{2}} \theta |F_r|^{-1} \frac{\mathrm{d}\Phi}{\mathrm{d}\theta}(\theta) , \tag{2.84}$$

where $d_s = \frac{D_s}{D}$ is the dimensionless grain size. Equation (2.84) is based on the assumption that the Shields parameter depends on the square of discharge ($\theta \propto q^2$) as it is found for the Chézy (2.20) and Manning (2.19) laws. In equations (2.81)-(2.84) dependence on three dimensionless

flow parameters is found, namely F_r, θ and d_s. We remark that θ itself may weakly depend on d_s if we use the logarithmic Chézy law (2.13).

2.4.2.3 A simplified sufficient condition

Furthermore, Cordier *et al.* [35] propose a *sufficient* condition for hyperbolicity of the system (2.70). A sufficient condition is by definition non-unique and defines a reduced hyperbolicity domain, which may be excessively limited compared to the actual domain defined by the necessary and sufficient condition (2.78). However the sought condition will be of practical interest due to very simple analytical application. Since it can be proved that $\alpha_- < u - \sqrt{gD} < k_f u < u + \sqrt{gD} < \alpha_+$, from (2.78) the sought sufficient condition reads:

$$(k_f - 1)\,|u| < \sqrt{gD} \qquad \text{i.e.} \qquad |F_r| < \frac{1}{k_f - 1}\,. \tag{2.85}$$

We observe that, unlike the necessary and sufficient condition (2.78), the proposed sufficient condition does not require evaluation of $\frac{\partial q_{sx}}{\partial q_x}$ and therefore can be checked analytically. From equation (2.85) some conclusions can be easily drawn:

- use of the Grass formula, characterised by $k_f = 1$, guarantees unconditional hyperbolicity;
- use of the transport formulae based on Shields stress (2.67) coupled with the Chézy friction law with constant C_h coefficient (or equivalently on the Darcy-Weisbach law), characterised by $k_f = 1$, results in unconditional hyperbolicity;
- use of the transport formulae (2.67) coupled with the Manning law or with the logarithmic Chézy law, having $k_f > 1$, results in specific requirements on data, namely:

  $$|u| < 6\sqrt{gD} \qquad \text{i.e.} \qquad |F_r| < 6 \tag{2.86}$$

  for the Manning law and

  $$|u| < \frac{C_h}{2.5}\sqrt{gD} \qquad \text{i.e.} \qquad |F_r| < 2.4 + \ln\left(\frac{D}{k_{C_h} D_s}\right) \tag{2.87}$$

  for the Chézy law.

2.4.3 Applications

In order to test the necessary and sufficient condition (2.78) and the sufficient condition (2.85) in cases of practical relevance we have developed a code implementing the bisection method. The

code runs on different values of water depth in a given interval and identifies the corresponding threshold for hyperbolicity in terms of velocity. Moreover it converts the results into the space of the Froude (F_r) and Shields (θ) dimensionless numbers. The code may run with different bedload transport formulae (2.67) and may consider the Manning law or the logarithmic Chézy law for evaluation of θ. Here we present the results of two set of runs, namely Test 1 and Test 2.

2.4.3.1 Test 1

In the first set of runs, we assess the hyperbolicity domain associated to the Meyer-Peter and Müller formula with the Manning law. We set different reasonable values for the Manning coefficient n (or, equivalently, for the Strickler coefficient $Ks = n^{-1}$) and use grain size $D_s = 0.01\,m$. Results are given in Fig. 2.3 in terms of dimensional variables (top panel) and in terms of dimensionless variables (bottom panel). Together with the *actual* hyperbolicity domains associated to different values of n, which are represented by coloured lines, in Fig. 2.3 we depict with a black line the *reduced* hyperbolicity domain associated to the sufficient condition (2.86). Since the formulation of the sufficient condition does not include any friction law, thus being insensitive to the value of the friction coefficient n, the threshold for the associated hyperbolicity domain is unique in the plot.

In the top profile we observe that the lines representing the actual hyperbolicity domains are tangent to that of the reduced domain and that the actual domains are wider than the domain given by (2.86), as expected. The position and shape of the non-hyperbolicity region strongly depends on the friction coefficient. In the bottom profile, where results are plotted in dimensionless form, we observe that the curves obtained for low values of n present a vertical asymptote corresponding to $\theta = \theta_{cr}$. This is related to the fact that under the critical Shields stress the considered shallow water-Exner model reduces to the shallow water model, which is unconditionally hyperbolic as we have seen in section 2.3.1. However, the threshold curves associated to higher values of n do not present any vertical asymptote. It is also seen that, accordingly with the sufficient condition (2.86), hyperbolicity is always verified for $F_r < 6$.

2.4.3.2 Test 2

In the second set of runs we have assessed the hyperbolicity domains of the shallow water-Exner model using two different bedload formulae (Meyer-Peter and Müller and Wong and Parker) in combination with two friction laws (Manning and logarithmic Chézy). In the friction laws we assume sediment diameter $D_s = 0.01\,m$ and compute the Manning coefficient $n = 0.022$ from the Strickler formula (2.11). In the logarithmic Chézy law we set $k_{C_h} = 2.5$. We remark that strictly

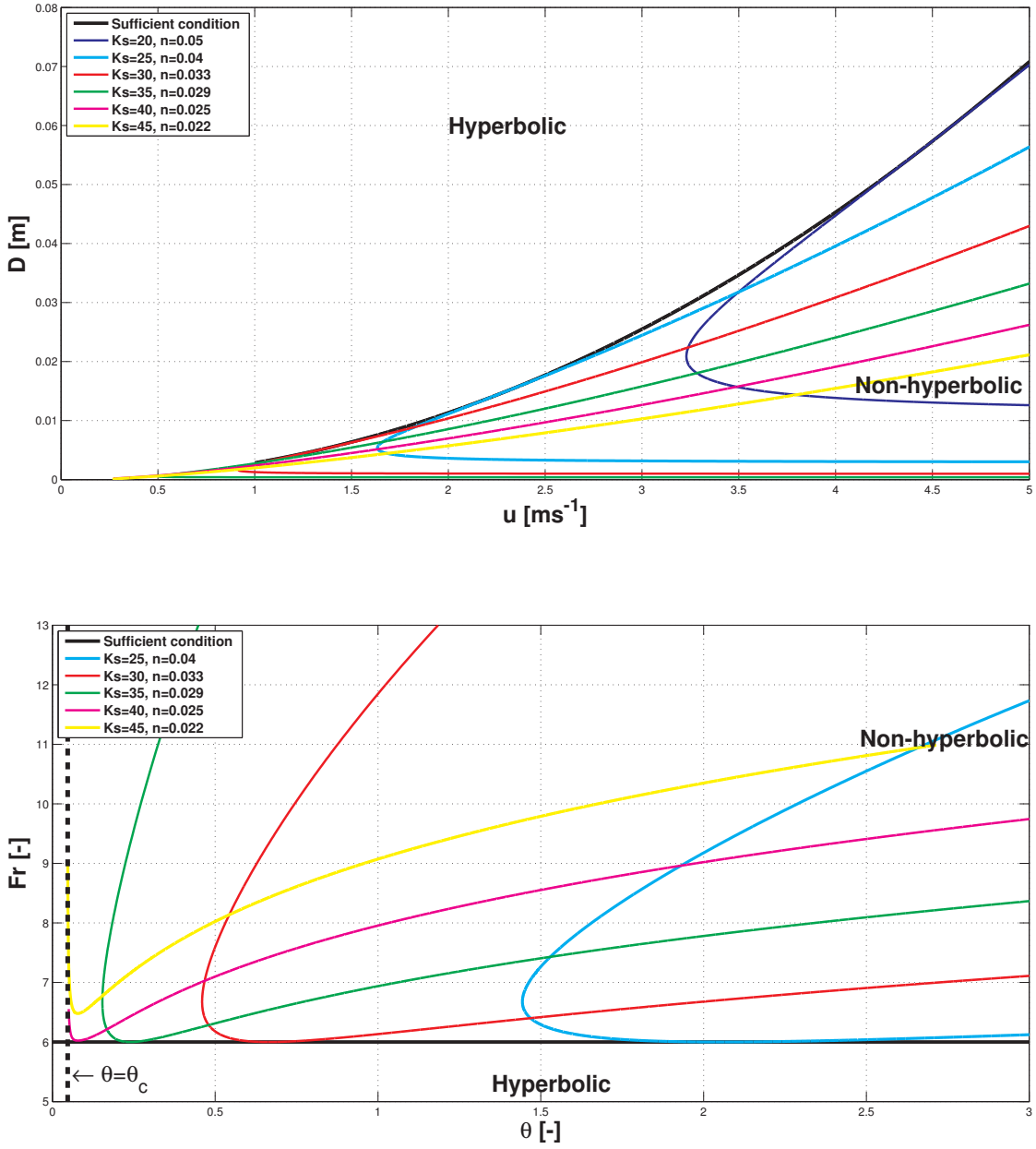

Figure 2.3: The actual hyperbolicity domains obtained testing the necessary and sufficient condition (2.78) numerically (coloured lines) are plotted together with the reduced domain of the sufficient condition (2.86) (black line). The Meyer-Peter and Müller formula in conjunction with the Manning friction law is applied, with different values of n. Sediment diameter is $D_s = 0.01\,m$. Results are presented in terms of water velocity u and depth D (top panel) and of the dimensionless parameters θ and F_r (bottom panel).

speaking the Chézy law applies only for $D \geq k_{C_h} D_s$, i.e. for $D \geq 0.025\,m$ in this case. Results are given in Fig. 2.4 in terms of dimensional variables (top panel) and in terms of dimensionless variables (bottom panel). The hyperbolicity domains obtained using the Manning law are represented by full lines, while those of the Chézy law are represented by dashed lines. Dark-grey lines are associated to the Wong and Parker formula and light-grey lines are associated to the Meyer-Peter and Müller formula. Moreover with black lines we represent the reduced hyperbolicity domains of the sufficient condition.

Comparison of the actual domains of the two transport formulae suggests that the Wong and Parker formula, which predicts lower sediment transport than the Meyer-Peter and Müller formula, returns wider hyperbolicity domains.
Comparing the actual domains of the Chézy law to these of the Manning law, we observe that the Chézy law gives rise to smaller hyperbolicity domains, as a result of higher prediction of friction and bedload transport on low depth. This is evident concerning the sufficient conditions (black lines) obtained using the two friction laws, where the threshold value of the Froude number in the hyperbolicity domain is equal to $F_r = 2.4$ for the Chézy law when $D = k_{C_h} D_s$, while it is $F_r = 6$ for the Manning law (see the bottom panel).

2.4.4 Conclusions

In this section we have reported and extended the analysis of Cordier *et al.* [35] concerning assessment of the hyperbolicity domains of the one-dimensional inviscid shallow water-Exner model describing the hydro-morphodynamic evolution of open channels with erodible bottom in absence of friction. We have presented two criteria for assessing hyperbolicity of this system: the necessary and sufficient condition, associated to the actual domain of hyperbolicity, and a simplified sufficient condition, returning reduced hyperbolicity domains. These conditions, originally presented in terms of inequalities on the flow data (depth and velocity) have been here conveniently expressed in terms of dimensionless parameters. Finally, we have applied these criteria to draw the hyperbolicity domains in practical cases.

In general, the shallow water-Exner model is *not* unconditionally hyperbolic. Unconditional hyperbolicity is obtained using the Grass transport formula or other transport formulae depending on the Shields parameter used in conjunction with the Chézy law with constant C_h coefficient. However, in many other cases of practical relevance, the request of hyperbolicity results in specific requirements on the flow data or dimensionless parameters.

The question at hand is now how strong these requirements are, i.e. whether in practical ap-

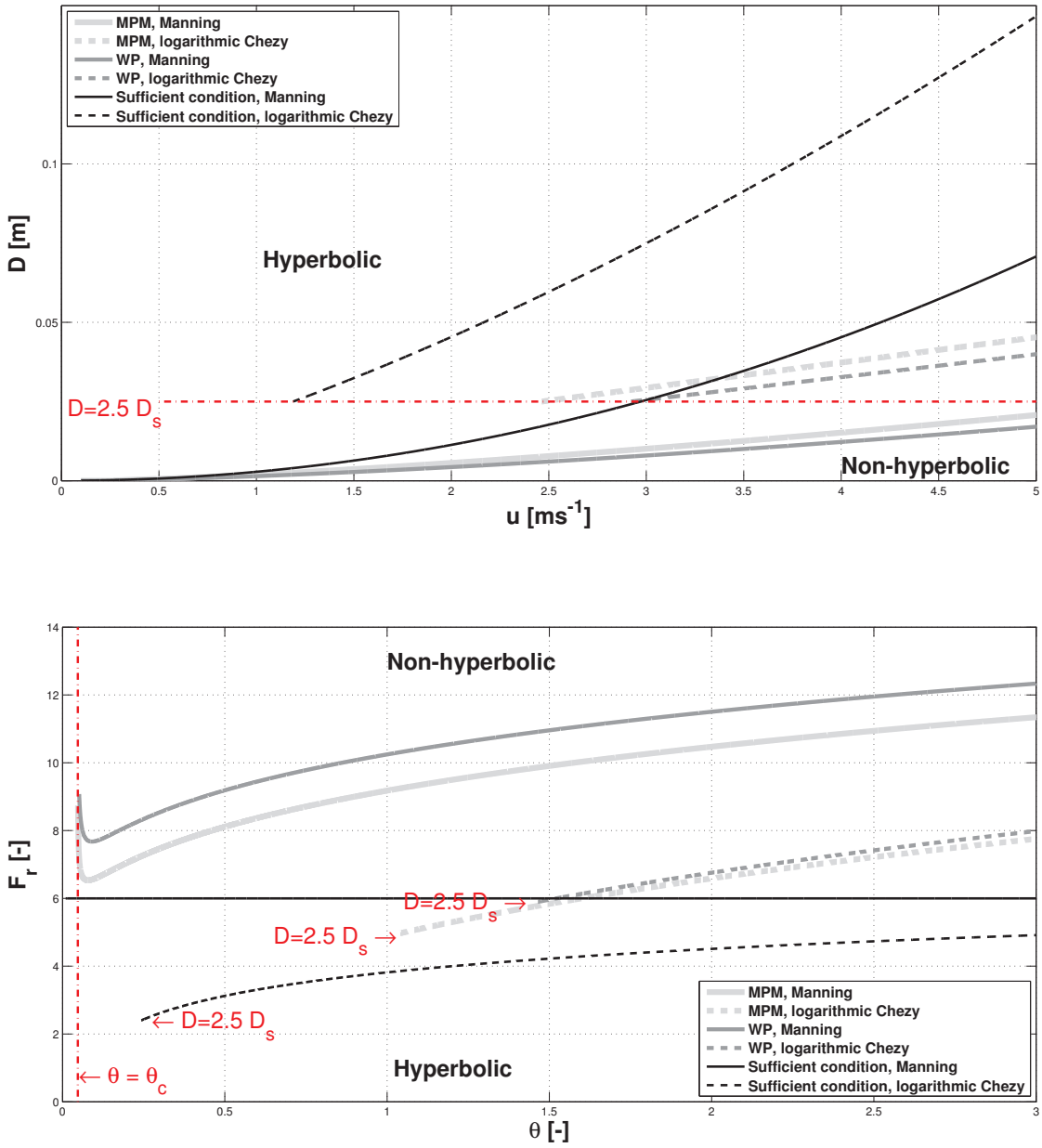

Figure 2.4: The actual hyperbolicity domains obtained testing the necessary and sufficient condition (2.78) numerically (grey lines) are plotted together with the reduced domains of the sufficient conditions (2.86) and (2.87) (black lines). The Meyer-Peter and Müller (MPM) and the Wong and Parker (WP) formulae in conjunction with the Manning and the logarithmic Chézy friction law are applied. Sediment diameter is $D_s = 0.01\,m$, n in the Manning law is set to 0.022. Results are presented in terms of water velocity u and depth D (top panel) and in terms of the dimensionless parameters θ and F_r (bottom panel).

plications concerning real rivers we can safely assume the system to be hyperbolic. The answer to the latter question is yes. As we have seen, if we adopt the Manning law for evaluation of the Shields number, the threshold value of the Froude number for hyperbolicity, according to the sufficient condition, is $F_r = 6$. Since such value in nature is very hardly exceeded, we can confidently assume in this case that the shallow water-Exner model is hyperbolic. Some concerns about the use of the logarithmic Chézy law instead may arise, since the threshold associated to the sufficient condition may be as low as $F_r = 2.4$ on very shallow depth (see Fig. 2.4), which might be locally exceeded in a gravel-bed river. However, from Fig. 2.4 we observe that the *actual* hyperbolicity domains of the Chézy law (grey dashed lines) are much wider than these predicted by the sufficient condition: using the Meyer-Peter and Müller formula we get $F_r \lesssim 5$. Thus even in this case we conclude that in practice the model is hyperbolic.

The present analysis has been carried out using the inviscid model, i.e. neglecting friction in the momentum equation. The adoption of such idealised approach makes us even more safe in claiming that the system applications in natural rivers must be hyperbolic. The reason is that friction increases with decreasing depth, thus preventing the occurrence of excessive values of water velocity in shallow regions, i.e. limiting the range of Froude numbers which are reasonably found in natural flows.

We extend these conclusions to the two-dimensional case in a straightforward manner. These results ensure that the shallow water-Exner model representing the hydro-morphodynamic problem in open channels is mathematically well-posed and pave the way for the numerical approximation of the solutions of the system. Moreover, they motivate the use of schemes for hyperbolic systems of PDEs in the construction of a numerical hydro-morphodynamic solver, as we will do in section 5.2.2 of chapter 5.

2.5 Non-conservativity of the shallow water-Exner model

From the previous section we can safely conclude that the shallow water-Exner model (2.6), together with the closure relations presented in section 2.1.2, is hyperbolic in the applications concerning real rivers. The question at hand is whether this hyperbolic system of PDEs is conservative, i.e. whether it can be recast in conservation-law form. The answer is intimately related to the numerical solution strategy to be adopted and requires some preliminary discussion on the available solution approaches.

Essentially two approaches are available in the literature to solve the system (2.6). In the *un-*

coupled approach the hydrodynamic problem governed by the shallow water equations is solved separately from the Exner equation: the solution of the hydrodynamic unknowns is then passed to the morphodynamic module (and vice-versa) at the end of each time step (asynchronous solution strategy), see e.g. [41, 45, 136]. Uncoupling allows the use of a conservative formulation, relying on the fact that the shallow water equations possess a conservation-law form with source terms. More subtle forms of uncoupled strategies can be designed by uncoupling just single terms in the equations, while maintaining a synchronous one-step algorithm. This is for instance the case of the models presented in [19, 23, 61, 24, 90, 137]. The system of governing equations (2.6) is recast in conservation-law form with source terms as follows:

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_y \mathbf{G}(\mathbf{Q}) = \mathbf{S}(\mathbf{Q}) \ , \tag{2.88}$$

where the vector of unknowns $\mathbf{Q}$, the flux vectors $\mathbf{F}(\mathbf{Q})$ and $\mathbf{G}(\mathbf{Q})$ and the source term vector $\mathbf{S}(\mathbf{Q})$ are given by

$$\mathbf{Q} = \begin{bmatrix} H \\ q_x \\ q_y \\ \eta \end{bmatrix} \quad , \quad \mathbf{F}(\mathbf{Q}) = \begin{bmatrix} q_x + q_{sx} \\ \frac{q_x^2}{H-\eta} + \frac{1}{2} g H^2 - g H \eta \\ \frac{q_x q_y}{H-\eta} \\ q_{sx} \end{bmatrix}$$

$$\mathbf{G}(\mathbf{Q}) = \begin{bmatrix} q_y + q_{sy} \\ \frac{q_x q_y}{H-\eta} \\ \frac{q_y^2}{H-\eta} + \frac{1}{2} g H^2 - g H \eta \\ q_{sy} \end{bmatrix} \quad , \quad \mathbf{S}(\mathbf{Q}) = \begin{bmatrix} 0 \\ -gH\partial_x \eta - g(H-\eta) S_{fx} \\ -gH\partial_y \eta - g(H-\eta) S_{fy} \\ 0 \end{bmatrix} . \tag{2.89}$$

This formulation ultimately relies on the ambiguous role of riverbed elevation η, which provides topographic momentum source terms through its gradients $\partial_y \eta$ and $\partial_y \eta$ in $\mathbf{S}(\mathbf{Q})$, while being a problem unknown in $\mathbf{Q}$. Since the source terms are updated only at the end of the integration time step and do not take part in the definition of the problem eigenstructure, from a mathematical point of view this approximation results in uncoupling the hydrodynamic and morphodynamic part [110], even though models based on this formulation are often claimed to be coupled in the literature.

Conversely, the *coupled* approach provides a synchronous solution for all the unknowns, which are jointly updated within the same time step [83, 58, 74, 36, 37, 110, 22, 21, 111]. Adoption of a fully-coupled formulation results in a non-conservative hyperbolic system of PDEs having the form

$$\partial_t \mathbf{Q} + \mathbf{A}_1 \partial_x \mathbf{Q} + \mathbf{A}_2 \partial_y \mathbf{Q} = \mathbf{S}(\mathbf{Q}) \ , \tag{2.90}$$

where the vector of unknowns $\mathbf{Q}$, the coefficient matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ and the vector of source terms

$\mathbf{S}(\mathbf{Q})$ are given by

$$\mathbf{Q} = \begin{bmatrix} H \\ q_x \\ q_y \\ \eta \end{bmatrix} \quad , \quad \mathbf{A}_1 = \begin{bmatrix} \frac{\partial q_{sx}}{\partial H} & 1+\frac{\partial q_{sx}}{\partial q_x} & \frac{\partial q_{sx}}{\partial q_y} & \frac{\partial q_{sx}}{\partial \eta} \\ g(H-\eta) - \frac{q_x^2}{(H-\eta)^2} & \frac{2q_x}{H-\eta} & 0 & \frac{q_x^2}{(H-\eta)^2} \\ -\frac{q_x q_y}{(H-\eta)^2} & \frac{q_y}{H-\eta} & \frac{q_x}{H-\eta} & \frac{q_x q_y}{(H-\eta)^2} \\ \frac{\partial q_{sx}}{\partial H} & \frac{\partial q_{sx}}{\partial q_x} & \frac{\partial q_{sx}}{\partial q_y} & \frac{\partial q_{sx}}{\partial \eta} \end{bmatrix},$$

$$\mathbf{A}_2 = \begin{bmatrix} \frac{\partial q_{sy}}{\partial H} & \frac{\partial q_{sy}}{\partial q_x} & 1+\frac{\partial q_{sy}}{\partial q_y} & \frac{\partial q_{sy}}{\partial \eta} \\ -\frac{q_x q_y}{(H-\eta)^2} & \frac{q_y}{H-\eta} & \frac{q_x}{H-\eta} & \frac{q_x q_y}{(H-\eta)^2} \\ g(H-\eta) - \frac{q_y^2}{(H-\eta)^2} & 0 & \frac{2q_y}{H-\eta} & \frac{q_y^2}{(H-\eta)^2} \\ \frac{\partial q_{sy}}{\partial H} & \frac{\partial q_{sy}}{\partial q_x} & \frac{\partial q_{sy}}{\partial q_y} & \frac{\partial q_{sy}}{\partial \eta} \end{bmatrix}, \tag{2.91}$$

$$\mathbf{S}(\mathbf{Q}) = \begin{bmatrix} 0 \\ -g(H-\eta) S_{fx} \\ -g(H-\eta) S_{fy} \\ 0 \end{bmatrix}. \tag{2.92}$$

In the development of our hydro-morphodynamic numerical model we will implement the latter coupled formulation, solving the non-conservative problem (2.90)-(2.92) by means of a path-conservative scheme. As a consequence of this choice, the model will be able to correctly account for sediment transport in Froude trans-critical flow conditions. This will be discussed and motivated in section 5.1 of chapter 5.

We shall now move to the presentation of our original achievements in the field of applied mathematics. In the next two chapters we will design novel numerical schemes of the centred upwind-biased type for the solution of hyperbolic systems of PDEs. In detail, in chapter 3 we will focus on the solution of hyperbolic systems of conservation laws, while in chapter 4 we will extend the centred upwind-biased approach to the solution of non-conservative hyperbolic systems in the path-conservative framework.
Later, in chapter 5 we will come back to the solution of the initial problem, i.e. addressing the morphodynamic evolution of open channels with erodible bed, by developing a two-dimensional hydro-morphodynamic solver based on the numerical techniques in chapter 4.

3 Upwind-biased centred schemes for hyperbolic systems in conservation-law form

In this thesis we aim at numerically solving the non-conservative non-homogeneous hyperbolic system of PDEs (2.90)-(2.92) arising from the coupled shallow water-Exner model (2.6), governing the morphodynamic evolution of gravel-bed rivers. The hydro-morphodynamic problem can be split into a purely advective problem, represented by a homogeneous system of PDEs, describing the propagation of waves in absence of friction, and an ordinary differential problem related to the presence of frictional source terms. Focusing on the advection problem, this chapter and the following one are devoted to the derivation of numerical schemes for the approximation of homogeneous hyperbolic systems of PDEs, with applications to the inviscid shallow water equations on horizontal bed.

Among the numerous techniques available for the solution of homogeneous systems (finite difference, finite element, spectral methods and many others) in this thesis we concentrate on numerical schemes based on the finite volume framework. The main reason for this choice is the simplicity of the finite volume approach, which relies on the discretisation of the weak form of systems. This leads to a formulation of the numerical problem which is immediately conservative, provided the underlying system of PDEs can be recast in conservation-law form. Conservativity of the numerical scheme has fundamental consequences: the theorem of Lax and Wendroff [78] guarantees in this case that the numerical method will converge to the weak solution of the problem even in the presence of shock waves, while the complementary theorem of Hou and LeFloch [59] states that a non-conservative method in general will not. The latter statement represents a major difficulty in the case of the considered hydro-morphodynamic problem since, as we know from the previous chapter, the coupled shallow water-Exner model is non-conservative. However, as we have seen in section 2.3.3, the path-conservative framework [42] allows to recover an acceptable definition of weak solutions and the development of shock-capturing methods even in this case.
Initially, we will focus on the development of numerical schemes for the solution of hyperbolic systems of conservation laws, thus getting rid of the uncertainties and difficulties of the path-

conservative framework. This is the matter of the present chapter. The solution techniques here developed will then be extended to path-conservative schemes in chapter 4. Finally, source terms will be re-introduced into the model in chapter 5: the scheme in chapter 5 will be able to handle all the features of the shallow water-Exner model.

In the framework of finite volume methods, there are essentially two approaches for designing non-oscillatory numerical schemes for the numerical solution of hyperbolic systems PDEs. The first approach is the *upwind approach*, represented by Godunov's method [53] and the other one is the *centred approach*, typically represented by the Lax-Friedrichs method [77]. The upwind approach, based on the application of Riemann solvers, provides more accurate solutions than the centred approach especially in the computations carried out at small Courant numbers and in the resolution of intermediate waves, the disadvantage being its complexity and computational expense. The centred approach is more general and simpler to apply to complicated sets of equations. A centred strategy is required by many physical and engineering problems of practical interest for which the solution of the Riemann problem is not easily achieved; our coupled shallow water-Exner model belongs to this category [21].

In this chapter two *centred upwind-biased* schemes are developed. The schemes partially use upwind information, while retaining the simplicity and efficiency of a centred scheme. Kurganov and Tadmor put forward an analogous idea in their *central-upwind* approach [73], using an adaptive staggered mesh. Multi-dimensional extensions of the scheme of Kurganov and Tadmor have been presented in [70] (Cartesian version) and [72] (unstructured version), while a modified version of the scheme, optimised for treating contact discontinuities, has been presented in [71].
Our schemes are modifications of the FORCE centred method, first put forward for one-dimensional systems in [119]. A multidimensional version of FORCE for unstructured meshes has recently been proposed by Toro *et al.* [120]. Their approach is of the predictor-corrector type and has conservative form, with a numerical flux defined on a secondary mesh. This secondary mesh is *edge-based* and is typically defined by joining the barycentre to the vertices of the primary mesh, in which conservative variables are defined. The FORCE-type upwind biased schemes presented in this chapter, regarded as UFORCE and UFORCE-δ, put forward the idea of varying the shape of the secondary mesh in an adaptive manner in order to achieve fine tuning numerical dissipation. In the UFORCE scheme for Cartesian meshes [112] the secondary mesh is redefined using one moving point which does not have to lay in the barycentres of the primary cells. In the UFORCE-δ method [113] we fully exploit this idea, allowing each subvolume composing the secondary mesh to be resized independently from the other ones, which makes the scheme more accurate and suitable for implementation on general unstructured meshes in multi-dimensions.

The proposed methods are linearly analysed in order to determine the optimal upwind bias, i.e. the secondary mesh setup resulting in the least dissipative monotone method. Our schemes in linear applications identically reproduce the Godunov upwind method constructed by solving Riemann problems normal to each interface. In fact, as proved here in the Cartesian case, the Godunov method is the most accurate scheme that can be constructed for the given stencil. Extension of UFORCE and UFORCE-δ to non-linear hyperbolic systems of PDEs is empirical and makes use of estimates of the eigenvalues of the system, which are in any case needed for selecting a time step. The basic first-order accurate UFORCE and UFORCE-δ methods are extended to second-order accuracy using standard techniques in the framework of finite volumes and implemented for solving the two-dimensional inviscid shallow water equations. Performance of the schemes is thoroughly assessed on a number of well-established test problems, some of them with exact solutions. Significant improvements in accuracy are observed with respect to the conventional FORCE scheme without upwind information.

The rest of this chapter proceeds as follows. In section 3.1 we set the background by reviewing the FORCE method in multiple-space dimensions, in the Cartesian and unstructured case. The first-order accurate UFORCE and UFORCE-δ methods are presented in section 3.2 and 3.3 respectively. In section 3.4 we extend the first-order UFORCE and UFORCE-δ fluxes to second order accuracy. Finally, numerical applications are presented in section 3.5.

3.1 Background

In this section we first briefly introduce the finite volume framework for the solution of systems of conservation laws (section 3.1.1). The rest of the section is devoted to reviewing the multi-dimensional FORCE numerical flux [120] in the Cartesian (section 3.1.2) and unstructured mesh configuration (section 3.1.3). For the latter scheme, in section 3.1.4 following Stecca *et al.* [113] we develop an alternative formulation which is explicit in terms of the number of space dimensions. This formulation will be useful in the development of the UFORCE-δ scheme in section 3.3.

3.1.1 Finite volume numerical methods

Finite volume numerical methods are based on the discretisation of weak forms of systems of PDEs. This approach, in the case of schemes for the solution of conservative systems, naturally leads to the development of conservative numerical schemes, whose convergence properties even in presence of shock waves are guaranteed by the Lax-Wendroff theorem [78] (see section 2.3.2.5 in the previous chapter). Here we introduce the finite volume approach in the one-dimensional case. For a more detailed presentation of this subject we refer the reader to Toro [118].

Consider a one-dimensional homogeneous hyperbolic system of conservation laws in differential form

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) = \mathbf{0} \tag{3.1}$$

where $\mathbf{Q}(x,t)$ is the vector of unknowns (or conserved variables) and $\mathbf{F}(\mathbf{Q})$ is the flux vector.
The development of finite volume schemes requires suitable domain discretisation. The one-dimensional domain for the problem at hand is discretised using spatial volumes having extent Δx, which we denote with T_i. Each cell T_i has barycentre in $x_i = i\Delta x$ and lays between $x_{i-\frac{1}{2}} = \left(i-\frac{1}{2}\right)\Delta x$ and $x_{i+\frac{1}{2}} = \left(i+\frac{1}{2}\right)\Delta x$, being i an integer index. Furthermore, we assume a discretisation of the time variable by time step Δt. With t^n we denote the time level at which the initial condition is provided, while with $t^{n+1} = t^n + \Delta t$ we indicate the time level at which the solution is sought.

The sought weak formulation of (3.1), is obtained integrating the system in the space-time control volume $\left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}\right] \times \left[t^n, t^{n+1}\right]$. It reads

$$\begin{aligned} &\frac{1}{\Delta t}\left\{\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathbf{Q}\left(x,t^{n+1}\right)\mathrm{d}x - \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathbf{Q}\left(x,t^{n}\right)\mathrm{d}x\right\} = \\ -&\frac{1}{\Delta x}\left\{\int_{t^n}^{t^n+1} \mathbf{F}\left(\mathbf{Q}\left(x_{i+\frac{1}{2}},t\right)\right)\mathrm{d}t - \int_{t^n}^{t^n+1} \mathbf{F}\left(\mathbf{Q}\left(x_{i-\frac{1}{2}},t\right)\right)\mathrm{d}t\right\} \end{aligned} \tag{3.2}$$

and is conveniently recast as

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{\Delta t}{\Delta x}\left(\mathbf{F}_{i+\frac{1}{2}} - \mathbf{F}_{i-\frac{1}{2}}\right) . \tag{3.3}$$

Relation (3.3) is *exact*, i.e. it is satisfied by the exact solution of (3.1), provided the following definitions hold:

$$\mathbf{Q}_i^n = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathbf{Q}(x,t^n)\,\mathrm{d}x \qquad , \qquad \mathbf{Q}_i^{n+1} = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathbf{Q}\left(x,t^{n+1}\right)\mathrm{d}x\,, \tag{3.4}$$

$$\mathbf{F}_{i-\frac{1}{2}} = \frac{1}{\Delta t}\int_{t^n}^{t^n+1} \mathbf{F}\left(\mathbf{Q}\left(x_{i-\frac{1}{2}},t\right)\right)\mathrm{d}t \qquad , \qquad \mathbf{F}_{i+\frac{1}{2}} = \frac{1}{\Delta t}\int_{t^n}^{t^n+1} \mathbf{F}\left(\mathbf{Q}\left(x_{i+\frac{1}{2}},t\right)\right)\mathrm{d}t\,. \tag{3.5}$$

Notice that in (3.4) $\mathbf{Q}_i^n$ and $\mathbf{Q}_i^{n+1}$ denote cell averages of the conserved variables at time t^n and $t^n + 1$ respectively.

The finite volume approach is based on the use of numerical approximations to the time inte-

grals of fluxes in (3.5):

$$\mathbf{F}_{i-\frac{1}{2}} \approx \frac{1}{\Delta t}\int_{t^n}^{t^n+1} \mathbf{F}\left(\mathbf{Q}\left(x_{i-\frac{1}{2}},t\right)\right)\mathrm{d}t \qquad , \qquad \mathbf{F}_{i+\frac{1}{2}} \approx \frac{1}{\Delta t}\int_{t^n}^{t^n+1} \mathbf{F}\left(\mathbf{Q}\left(x_{i+\frac{1}{2}},t\right)\right)\mathrm{d}t \tag{3.6}$$

The (cell-averaged) solution is updated applying the *update formula* (3.3) where the numerical approximated fluxes (3.6) are inserted. Finite volume schemes differ among each other by the relations and procedures which are needed for the computation of these fluxes.

3.1.2 The FORCE scheme on Cartesian meshes

Here we review the derivation of the FORCE scheme [120] on two-dimensional Cartesian meshes. We consider a two-dimensional system of m non-linear hyperbolic equations written in differential conservation-law form:

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_y \mathbf{G}(\mathbf{Q}) = \mathbf{0}\,, \tag{3.7}$$

where $\mathbf{Q}$ is the vector of conserved variables and $\mathbf{F}(\mathbf{Q})$ and $\mathbf{G}(\mathbf{Q})$ are the flux vectors in the x and y direction respectively. We adopt a two-dimensional Cartesian mesh by rectangular elements (or cells) $T_{i,j}$. Each cell $T_{i,j}$ has barycentre $\vec{G}_{i,j} = (i\Delta x, j\Delta y)$, and vertices $\vec{K}_{i\pm\frac{1}{2},j\pm\frac{1}{2}}$ given by:

$$\begin{aligned} \vec{K}_{i+\frac{1}{2},j-\frac{1}{2}} &= \left(\left(i+\tfrac{1}{2}\right)\Delta x, \left(j-\tfrac{1}{2}\right)\Delta y\right) \quad , \quad \vec{K}_{i+\frac{1}{2},j+\frac{1}{2}} = \left(\left(i+\tfrac{1}{2}\right)\Delta x, \left(j+\tfrac{1}{2}\right)\Delta y\right)\,, \\ \vec{K}_{i-\frac{1}{2},j+\frac{1}{2}} &= \left(\left(i-\tfrac{1}{2}\right)\Delta x, \left(j+\tfrac{1}{2}\right)\Delta y\right) \quad , \quad \vec{K}_{i-\frac{1}{2},j-\frac{1}{2}} = \left(\left(i-\tfrac{1}{2}\right)\Delta x, \left(j-\tfrac{1}{2}\right)\Delta y\right)\,, \end{aligned} \tag{3.8}$$

being i and j the Cartesian indices and Δx and Δy mesh spacing.

By analogy with the one-dimensional case previously reviewed, finite volume schemes on two-dimensional Cartesian meshes are obtained by integration of (3.7) over the control volume $T_{i,j}\times[t^n,t^{n+1}]$, yielding:

$$\mathbf{Q}_{i,j}^{n+1} = \mathbf{Q}_{i,j}^{n} - \frac{\Delta t}{\Delta x}\left(\mathbf{F}_{i+\frac{1}{2},j} - \mathbf{F}_{i-\frac{1}{2},j}\right) - \frac{\Delta t}{\Delta y}\left(\mathbf{G}_{i,j+\frac{1}{2}} - \mathbf{G}_{i,j-\frac{1}{2}}\right)\,, \tag{3.9}$$

where $\Delta t = t^{n+1} - t^n$ is time step, $\mathbf{Q}_{i,j}^n$ is the cell average at time level t^n, $\mathbf{Q}_{i,j}^{n+1}$ denotes the sought numerical solution at time t^{n+1} and $\mathbf{F}_{i+\frac{1}{2},j}$, $\mathbf{F}_{i-\frac{1}{2},j}$ and $\mathbf{G}_{i,j+\frac{1}{2}}$, $\mathbf{G}_{i,j-\frac{1}{2}}$ represent numerical approximations to fluxes.

The derivation of the FORCE flux [120] requires the adoption of two different meshes: the *primary mesh*, where the numerical solution is sought, and a staggered *secondary mesh*, used to define numerical fluxes. Fig. 3.1 depicts the primary and secondary mesh setup for the two-dimensional

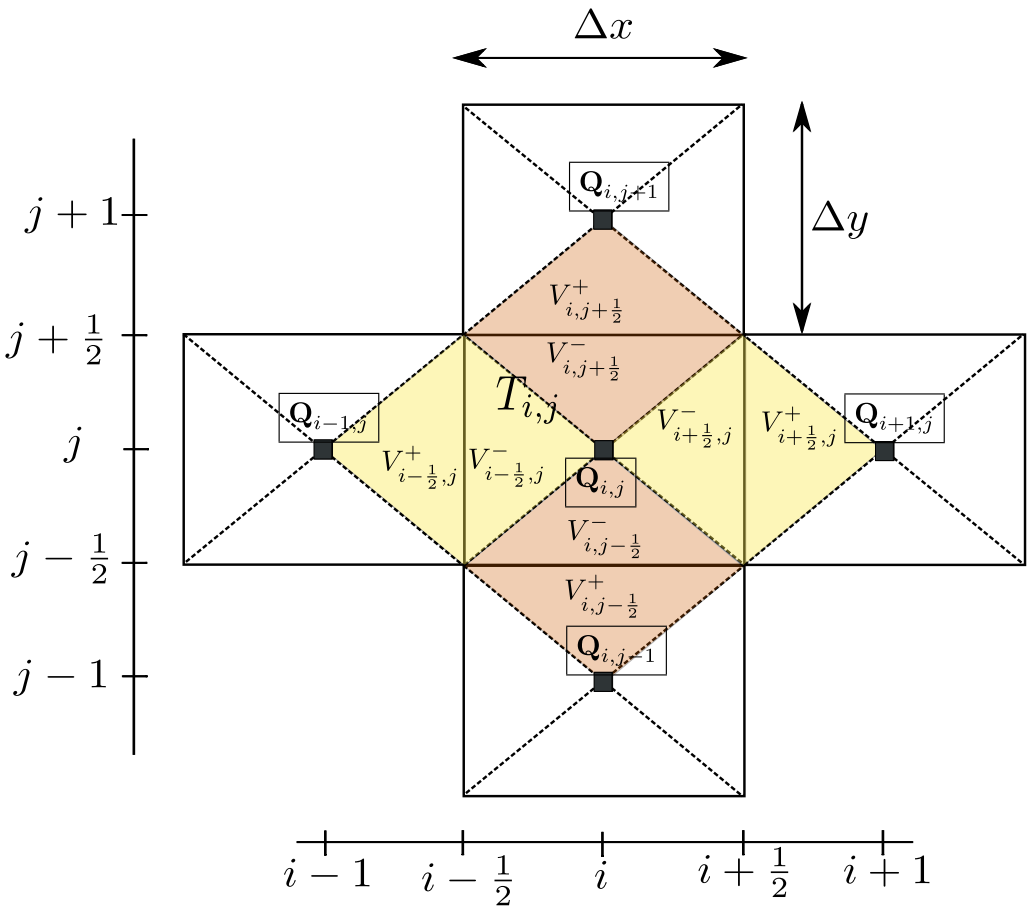

Figure 3.1: Sketch of the primary and secondary mesh for the FORCE method set on two-dimensional Cartesian meshes.

case. The primary mesh is the Cartesian mesh by elements $T_{i,j}$ chosen for domain discretisation. Each primary cell $T_{i,j}$ possesses four edges (or interfaces or inter-cell boundaries). To each of these edges one element (or volume) of the secondary mesh is associated. These secondary elements are generated by joining the barycentre of $T_{i,j}$ and that of its neighbour across the current edge with the edge vertices. This results in a quadrilateral straddling each edge, as depicted in Fig. 3.1. The four edge-based secondary control volumes and the corresponding edges of cell $T_{i,j}$ are defined as

$$\begin{aligned} V_{i+\frac{1}{2},j} &= V^-_{i+\frac{1}{2},j} \bigcup V^+_{i+\frac{1}{2},j} \quad , \quad V_{i,j+\frac{1}{2}} = V^-_{i,j+\frac{1}{2}} \bigcup V^+_{i,j+\frac{1}{2}} \, , \\ V_{i-\frac{1}{2},j} &= V^-_{i-\frac{1}{2},j} \bigcup V^+_{i-\frac{1}{2},j} \quad , \quad V_{i,j-\frac{1}{2}} = V^-_{i,j-\frac{1}{2}} \bigcup V^+_{i,j-\frac{1}{2}} \, , \end{aligned} \tag{3.10}$$

where $V^-_{i\pm\frac{1}{2},j}$, $V^-_{i,j\pm\frac{1}{2}}$ denote the portions of the corresponding secondary volumes of the staggered mesh that lay inside $T_{i,j}$ and $V^+_{i\pm\frac{1}{2},j}$, $V^+_{i,j\pm\frac{1}{2}}$ denote the portions of these volumes that lay outside (see Fig. 3.1). In the following, these portions will be regarded as *subvolumes*.

The FORCE method is derived as a two-step method of the predictor-corrector type and then recast as a one-step conservative method on a non-staggered mesh. Assuming $\mathbf{Q}^n_{i,j}$ as initial condition at time t^n in each cell $T_{i,j}$ of the primary mesh, the construction of the FORCE scheme is then achieved in three stages.

- The initial condition $\mathbf{Q}^n_{i,j}$ in the current cell $T_{i,j}$ and in all its immediate neighbours determines initial conditions for each of the four volumes (3.10) of the staggered mesh associated with the boundaries of $T_{i,j}$. Starting from these initial conditions, in the first stage we define intermediate states at time $t^{n+\frac{1}{2}} = t^n + \frac{1}{2}\Delta t$ in each control volume of the staggered mesh. This is achieved evolving the solution for a time interval $\frac{1}{2}\Delta t$ by integrating the conservation law (3.7) in each volume (3.10), thus obtaining

$$\begin{aligned}
\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2},j} &= \tfrac{1}{2}\left(\mathbf{Q}^n_{i,j} + \mathbf{Q}^n_{i+1,j}\right) - \tfrac{\Delta t}{\Delta x}\left\{\mathbf{F}(\mathbf{Q}^n_{i+1,j}) - \mathbf{F}(\mathbf{Q}^n_{i,j})\right\}, \\
\mathbf{Q}^{n+\frac{1}{2}}_{i,j+\frac{1}{2}} &= \tfrac{1}{2}\left(\mathbf{Q}^n_{i,j} + \mathbf{Q}^n_{i,j+1}\right) - \tfrac{\Delta t}{\Delta y}\left\{\mathbf{G}(\mathbf{Q}^n_{i,j+1}) - \mathbf{G}(\mathbf{Q}^n_{i,j})\right\}, \\
\mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2},j} &= \tfrac{1}{2}\left(\mathbf{Q}^n_{i-1,j} + \mathbf{Q}^n_{i,j}\right) - \tfrac{\Delta t}{\Delta x}\left\{\mathbf{F}(\mathbf{Q}^n_{i,j}) - \mathbf{F}(\mathbf{Q}^n_{i-1,j})\right\}, \\
\mathbf{Q}^{n+\frac{1}{2}}_{i,j-\frac{1}{2}} &= \tfrac{1}{2}\left(\mathbf{Q}^n_{i,j-1} + \mathbf{Q}^n_{i,j}\right) - \tfrac{\Delta t}{\Delta y}\left\{\mathbf{G}(\mathbf{Q}^n_{i,j}) - \mathbf{G}(\mathbf{Q}^n_{i,j-1})\right\}.
\end{aligned} \tag{3.11}$$

- In the second stage, with initial condition at time $t^{n+\frac{1}{2}}$ given by the intermediate states (3.11) we further integrate the conservation law (3.7) within each cell $T_{i,j}$ by a time step $\frac{1}{2}\Delta t$, obtaining averages at time $t^{n+1} = t^n + \Delta t$ in the form

$$\begin{aligned}
\mathbf{Q}^{n+1}_{i,j} &= \tfrac{1}{4}\left(\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2},j} + \mathbf{Q}^{n+\frac{1}{2}}_{i,j+\frac{1}{2}} + \mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2},j} + \mathbf{Q}^{n+\frac{1}{2}}_{i,j-\frac{1}{2}}\right) \\
&- \tfrac{1}{2}\tfrac{\Delta t}{\Delta x}\left\{\mathbf{F}(\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2},j}) - \mathbf{F}(\mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2},j})\right\} \\
&- \tfrac{1}{2}\tfrac{\Delta t}{\Delta y}\left\{\mathbf{G}(\mathbf{Q}^{n+\frac{1}{2}}_{i,j+\frac{1}{2}}) - \mathbf{G}(\mathbf{Q}^{n+\frac{1}{2}}_{i,j-\frac{1}{2}})\right\}.
\end{aligned} \tag{3.12}$$

- In the final stage of the derivation of the scheme we perform algebraic manipulations on (3.12) so as to reproduce the conservative formula (3.9). The resulting FORCE numerical

flux in the two-dimensional case (FORCE-2) is

$$\mathbf{F}^{FORCE-2}_{i+\frac{1}{2},j} = \frac{1}{2}\left\{\mathbf{F}\left(\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2},j}\right) + \frac{1}{2}\left(\mathbf{F}\left(\mathbf{Q}^{n}_{i,j}\right) + \mathbf{F}\left(\mathbf{Q}^{n}_{i+1,j}\right)\right) - \frac{1}{4}\frac{\Delta x}{\Delta t}\left(\mathbf{Q}^{n}_{i+1,j} - \mathbf{Q}^{n}_{i,j}\right)\right\}, \tag{3.13}$$

with analogous formulae for the other fluxes. It is possible to rewrite the intercell flux (3.13) as the arithmetic average of two fluxes:

$$\mathbf{F}^{FORCE-2}_{i+\frac{1}{2},j} = \frac{1}{2}\left(\mathbf{F}^{LW-2}_{i+\frac{1}{2},j} + \mathbf{F}^{LF-2}_{i+\frac{1}{2},j}\right). \tag{3.14}$$

These two fluxes may be regarded as a two-dimensional generalisation of the Lax-Wendroff (LW-2) and Lax-Friedrichs (LF-2) flux respectively. The Lax-Wendroff type flux is given by the physical flux $\mathbf{F}$ evaluated at the intermediate state obtained from the first averaging procedure (3.11):

$$\mathbf{F}^{LW-2}_{i+\frac{1}{2},j} = \mathbf{F}\left(\mathbf{Q}^{LW-2}_{i+\frac{1}{2},j}\right) \tag{3.15}$$

with

$$\mathbf{Q}^{LW-2}_{i+\frac{1}{2},j} = \frac{1}{2}\left(\mathbf{Q}^{n}_{i,j} + \mathbf{Q}^{n}_{i+1,j}\right) - \frac{1}{2}\left(\frac{2\Delta t}{\Delta x}\right)\left(\mathbf{F}\left(\mathbf{Q}^{n}_{i+1,j}\right) - \mathbf{F}\left(\mathbf{Q}^{n}_{i,j}\right)\right). \tag{3.16}$$

The Lax-Friedrichs type flux is given by

$$\mathbf{F}^{LF-2}_{i+\frac{1}{2},j} = \frac{1}{2}\left(\mathbf{F}\left(\mathbf{Q}^{n}_{i+1,j}\right) + \mathbf{F}\left(\mathbf{Q}^{n}_{i,j}\right)\right) - \frac{1}{2}\left(\frac{\Delta x}{2\Delta t}\right)\left(\mathbf{Q}^{n}_{i+1,j} - \mathbf{Q}^{n}_{i,j}\right). \tag{3.17}$$

The factor $\alpha = 2$ multiplying the time step Δt in (3.16) and (3.17) denotes the number of spatial dimensions.

Toro *et al.* [120] generalise the two-dimensional flux (3.14)-(3.17), thus obtaining a formulation which applies for α space dimensions on Cartesian meshes:

$$\mathbf{F}^{FORCE-\alpha}_{i+\frac{1}{2}} = \frac{1}{2}\left(\mathbf{F}^{LW-\alpha}_{i+\frac{1}{2}} + \mathbf{F}^{LF-\alpha}_{i+\frac{1}{2}}\right). \tag{3.18}$$

In (3.18), the Lax-Wendroff type flux is given by

$$\mathbf{F}^{LW-\alpha}_{i+\frac{1}{2}} = \mathbf{F}\left(\mathbf{Q}^{LW-\alpha}_{i+\frac{1}{2}}\right) \tag{3.19}$$

with

$$\mathbf{Q}^{LW-\alpha}_{i+\frac{1}{2}} = \frac{1}{2}\left(\mathbf{Q}^{n}_{i} + \mathbf{Q}^{n}_{i+1}\right) - \frac{1}{2}\left(\frac{\alpha\Delta t}{\Delta x}\right)\left(\mathbf{F}\left(\mathbf{Q}^{n}_{i+1}\right) - \mathbf{F}\left(\mathbf{Q}^{n}_{i}\right)\right), \tag{3.20}$$

while the Lax-Friedrichs type flux is

$$\mathbf{F}^{LF-\alpha}_{i+\frac{1}{2}} = \frac{1}{2}\left(\mathbf{F}\left(\mathbf{Q}^n_{i+1}\right) + \mathbf{F}\left(\mathbf{Q}^n_i\right)\right) - \frac{1}{2}\left(\frac{\Delta x}{\alpha \Delta t}\right)\left(\mathbf{Q}^n_{i+1} - \mathbf{Q}^n_i\right) . \tag{3.21}$$

In equations (3.18)-(3.21), x indicates the current direction and subscript i is the index along the x direction.

A linear analysis shows that the FORCE scheme is monotone under the following conditions:

$$\begin{array}{rll} c_x^2 + c_y^2 & \leq \frac{1}{2} & \text{in two space dimensions} \\ c_x^2 + c_y^2 + c_z^2 & \leq \frac{1}{3} & \text{in three space dimensions} \end{array} \tag{3.22}$$

where

$$c_x = \lambda_x \frac{\Delta t}{\Delta x} \quad , \quad c_y = \lambda_y \frac{\Delta t}{\Delta y} \quad , \quad c_z = \lambda_z \frac{\Delta t}{\Delta z} \tag{3.23}$$

represent the Courant numbers in x, y and z directions and $\lambda_x, \lambda_y, \lambda_z$ are the corresponding characteristic speeds.

3.1.3 The FORCE scheme on multi-dimensional unstructured meshes

Here we review the derivation of the FORCE scheme in the more general case of multi-dimensional unstructured meshes. We consider a multi-dimensional system of non-linear conservation laws having the form

$$\partial_t \mathbf{Q} + \nabla \cdot \left(\underline{\underline{\mathbf{F}}}(\mathbf{Q})\right) = \mathbf{0} , \tag{3.24}$$

where $\underline{\underline{\mathbf{F}}}(\mathbf{Q})$ is the flux tensor. The discretisation of the computational domain $\Omega \subset \mathrm{R}^{\alpha}$, being α the number of spatial dimensions, is achieved using a conforming tessellation $\mathcal{T}_\Omega$ by n_e elements T_i such that

$$\mathcal{T}_\Omega = \bigcup_{i=1}^{n_e} T_i . \tag{3.25}$$

Each element T_i, having size $|T_i|$, has n_f plane interfaces S_j of size $|S_j|$, with associated outward-pointing normal unit vectors $\vec{n}_j$. The neighbouring element of the primary mesh sharing interface S_j with T_i is denoted with T_j.

By analogy with the Cartesian case, the method derivation makes use of two meshes: a primary mesh (3.25) and a secondary, edge-based staggered mesh. The setup is illustrated in Fig. 3.2 for the case of triangular meshes in two space dimensions. In order to define the secondary mesh, each primary element T_i is sub-divided into subvolumes V_j^- generated by connecting the barycentre

of T_i with the vertices of S_j. The corresponding adjacent subvolume in the neighbouring element that shares face S_j with element T_i is denoted with V_j^+. With reference to Fig. 3.2 we distinguish two kinds of elements: *primary elements* T_i, at which the solution is sought at each time step, and *secondary elements* formed by $V_j^- \bigcup V_j^+$, for $1 \leq j \leq n_f$.

Finite volume schemes on unstructured meshes are obtained by integration of the conservation law (3.24) over the space-time control volume $T_i \times \left[t^n, t^{n+1}\right]$, yielding

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{\Delta t}{|T_i|} \sum_{j=1}^{n_f} \int_{S_j} \underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}} \left(\mathbf{Q}_i^n, \mathbf{Q}_j^n\right) \cdot \vec{n}_j \, \mathrm{d}S \,, \tag{3.26}$$

where $\mathbf{Q}_i^n$ is the cell average at time t^n and $\mathbf{Q}_i^{n+1}$ is the numerical solution at time t^{n+1}, being $\Delta t = t^{n+1} - t^n$ the time step. The multi-dimensional FORCE flux on unstructured meshes is constructed in three stages as follows.

- First, assuming averages in each primary element at time $t = t^n$, an intermediate state for each interface S_j is defined at the half-time level $t^{n+\frac{1}{2}} = t^n + \frac{1}{2}\Delta t$ by integrating the conservation law (3.24) by $\frac{1}{2}\Delta t$ over the secondary elements:

$$\mathbf{Q}_{j+\frac{1}{2}}^{n+\frac{1}{2}} = \frac{\mathbf{Q}_i^n \left|V_j^-\right| + \mathbf{Q}_j^n \left|V_j^+\right|}{\left|V_j^-\right| + \left|V_j^+\right|} - \frac{1}{2} \frac{\Delta t \left|S_j\right|}{\left|V_j^-\right| + \left|V_j^+\right|} \left(\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_j^n\right) - \underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_i^n\right)\right) \cdot \vec{n}_j \,, \tag{3.27}$$

  where $\left|V_j^-\right|$ and $\left|V_j^+\right|$ indicate the size of subvolumes V_j^- and V_j^+ respectively, namely their length in 1D, surface area in 2D and volume in 3D.

- Then, with initial condition at time $t^{n+\frac{1}{2}}$ given by (3.27), integration of the conservation law (3.24) over the *primary elements* T_i by $\frac{1}{2}\Delta t$ yields averages at time t^{n+1}, namely

$$\mathbf{Q}_i^{n+1} = \frac{1}{|T_i|} \sum_{j=1}^{n_f} \left(\mathbf{Q}_{j+\frac{1}{2}}^{n+\frac{1}{2}} \left|V_j^-\right| - \frac{1}{2}\Delta t \left|S_j\right| \underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_{j+\frac{1}{2}}^{n+\frac{1}{2}}\right) \cdot \vec{n}_j \right) . \tag{3.28}$$

  Equations (3.27) and (3.28) constitute a first-order accurate, explicit two-step method for solving (3.24) on a staggered mesh.

- Finally, following the FORCE approach [119] the scheme can be written as a one-step scheme in conservative form on a non-staggered mesh, with a corresponding numerical flux. After some algebraic manipulations involving the Gauss theorem ($\sum_j S_j \vec{n}_j = \vec{0}$) and normalizing

the face-normal vectors ($\vec{n}_j^2 = 1$) the scheme is recast into the sought conservative form:

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{\Delta t}{|T_i|} \sum_{j=1}^{n_f} |S_j| \, \underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{FORCE} \cdot \vec{n}_j \, , \tag{3.29}$$

where the multidimensional FORCE flux on general meshes $\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{FORCE}$ is given by

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{FORCE} = \frac{1}{2} \left(\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LW} \left(\mathbf{Q}_i^n, \mathbf{Q}_j^n \right) + \underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LF} \left(\mathbf{Q}_i^n, \mathbf{Q}_j^n \right) \right) . \tag{3.30}$$

The FORCE flux is then the arithmetic average of two fluxes: a two-point flux of the Lax-Wendroff type and a two-point flux of the Lax-Friedrichs type. The Lax-Wendroff type flux is given by the physical flux $\underline{\underline{\mathbf{F}}}$ evaluated at the intermediate state obtained from the first averaging procedure (3.27):

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LW} = \underline{\underline{\mathbf{F}}} \left(\mathbf{Q}_{j+\frac{1}{2}}^{LW} \right) , \tag{3.31}$$

$$\mathbf{Q}_{j+\frac{1}{2}}^{LW} = \frac{\mathbf{Q}_i^n \left| V_j^- \right| + \mathbf{Q}_j^n \left| V_j^+ \right|}{\left| V_j^- \right| + \left| V_j^+ \right|} - \frac{1}{2} \frac{\Delta t \left| S_j \right|}{\left| V_j^- \right| + \left| V_j^+ \right|} \left(\underline{\underline{\mathbf{F}}} \left(\mathbf{Q}_j^n \right) - \underline{\underline{\mathbf{F}}} \left(\mathbf{Q}_i^n \right) \right) \cdot \vec{n}_j \, , \tag{3.32}$$

while the Lax-Friedrichs type flux is given by

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LF} = \frac{\underline{\underline{\mathbf{F}}} \left(\mathbf{Q}_j^n \right) \left| V_j^- \right| + \underline{\underline{\mathbf{F}}} \left(\mathbf{Q}_i^n \right) \left| V_j^+ \right|}{\left| V_j^- \right| + \left| V_j^+ \right|} - \frac{\left| V_j^- \right| \left| V_j^+ \right|}{\left| V_j^- \right| + \left| V_j^+ \right|} \frac{2}{\Delta t \left| S_j \right|} \left(\mathbf{Q}_j^n - \mathbf{Q}_i^n \right) \vec{n}_j^T \, . \tag{3.33}$$

3.1.4 The FORCE-α scheme on general meshes

Although the FORCE flux on unstructured meshes [120] generally applies in multi-dimensions, the original formulation (3.30)-(3.33) is not explicit in terms of the number of spatial dimensions α. An explicit formulation of this type is instead available in the Cartesian case (3.18)-(3.21). In this section, following Stecca *et al.* [113], we fill in this gap by developing an alternative α formulation of the FORCE flux on unstructured meshes, which generalises (3.18)-(3.21). The proposed FORCE-α formulation, providing much insight on the role of the size of subvolumes $V_j^{\pm}$ in controlling numerical dissipation, will be useful in the derivation of the UFORCE-δ method in section 3.3.

We recall that the secondary mesh cells in the multidimensional FORCE method are composed of two subvolumes V_j^- and V_j^+ having triangular shape in 2D ($\alpha = 2$) and pyramidal shape in

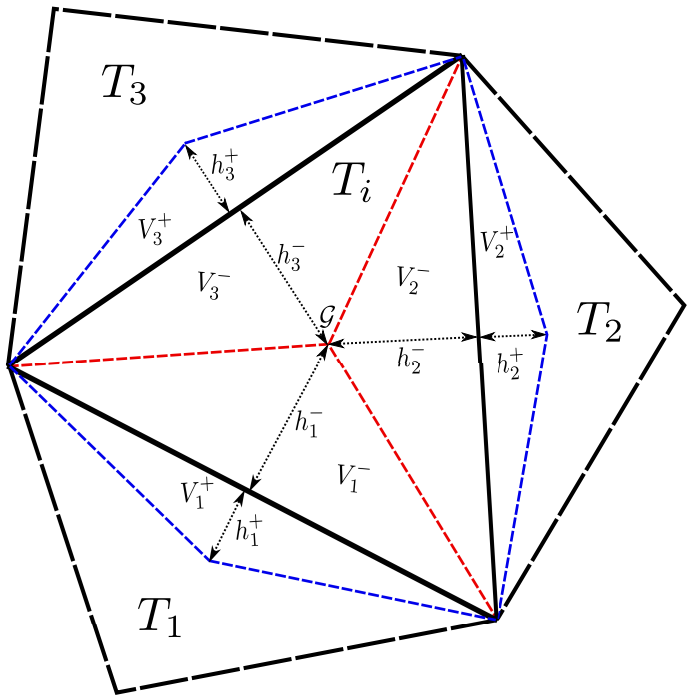

Figure 3.2: Sketch of the primary and secondary mesh for the FORCE method set on two-dimensional unstructured triangular meshes.

3D ($\alpha = 3$). With $\left|V_j^-\right|$ and $\left|V_j^+\right|$ we indicate the size of these subvolumes, which dimensionally corresponds to their surface area in 2D and their volume in 3D, given by

$$\left|V_j^+\right| = \frac{h_j^+ \left|S_j\right|}{\alpha} \quad , \quad \left|V_j^-\right| = \frac{h_j^- \left|S_j\right|}{\alpha} \, , \tag{3.34}$$

where h_j^+ and h_j^- are the altitudes of $\left|V_j^+\right|$ and $\left|V_j^-\right|$ respectively and $\left|S_j\right|$ represents the length of the triangle base or the area of the pyramid base surface (see Fig. 3.2). Substitution of (3.34) into the FORCE flux formulae (3.30)-(3.33) gives

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{FORCE-\alpha} = \frac{1}{2}\left(\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LW-\alpha}\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n\right) + \underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LF-\alpha}\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n\right)\right) \, , \tag{3.35}$$

with

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LW-\alpha} = \underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_{j+\frac{1}{2}}^{LW-\alpha}\right) \, , \tag{3.36}$$

$$\mathbf{Q}_{j+\frac{1}{2}}^{LW-\alpha} = \frac{\mathbf{Q}_i^n h_j^- + \mathbf{Q}_j^n h_j^+}{h_j^- + h_j^+} - \frac{1}{2}\frac{\alpha \Delta t}{h_j^- + h_j^+}\left(\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_j^n\right) - \underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_i^n\right)\right) \cdot \vec{n}_j \tag{3.37}$$

and

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LF-\alpha} = \frac{\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_j^n\right) h_j^- + \underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_i^n\right) h_j^+}{h_j^- + h_j^+} - \frac{h_j^- h_j^+}{h_j^- + h_j^+} \frac{2}{\alpha \Delta t} \left(\mathbf{Q}_j^n - \mathbf{Q}_i^n\right) \vec{n}_j^T \ . \tag{3.38}$$

The above formulation applies to any kind of mesh and reduces to (3.18)-(3.21) in the Cartesian case.

The altitudes $h_j^{\pm}$, which are required by the method, in the case of triangular and tetrahedral meshes are simply given by

$$h_j^- = \frac{\alpha |T_i|}{n_f |S_j|} \qquad , \qquad h_j^+ = \frac{\alpha |T_j|}{n_f |S_j|} \ , \tag{3.39}$$

where n_f represents the number of cell boundaries ($n_f = 3$ on triangular meshes and $n_f = 4$ on tetrahedral meshes).

3.2 The UFORCE scheme on Cartesian meshes

In this section we develop an upwind-biased variation of the FORCE method on Cartesian meshes. This is achieved by modifying the shape of the staggered mesh in the FORCE scheme accordingly with upwind information represented by the system eigenvalues. The resulting scheme, first presented by Stecca *et al.* [112], is regarded as UFORCE. The method is first derived in the two-dimensional case and then generalised on α space dimensions. We proceed as follows: first in section 3.2.1 we construct our upwind-biased version of the FORCE method, in which the secondary mesh setup is determined by two parameters per cell, still to be determined. Then in section 3.2.2 we analytically determine the optimal setup for these parameters in applications to the two-dimensional linear advection equation with constant coefficients. Last in section 3.2.3 the recipe for the optimal upwind bias is extended to the case of non-linear hyperbolic systems of PDEs.

3.2.1 Derivation

As for the classical FORCE scheme reviewed in section 3.1.2, the derivation of UFORCE requires the adoption of a primary mesh for computing cell averages and of a staggered mesh used to define numerical fluxes for the conservative form of the scheme. We initially focus on the two-dimensional case. The primary mesh, by elements $T_{i,j}$, is Cartesian and is defined in identical manner as in section 3.1.2. The staggered mesh is now defined with reference to a generic point $\vec{P}_{i,j}$ laying inside each primary cell $T_{i,j}$, which does not have to coincide with the barycentre $\vec{G}_{i,j}$ of $T_{i,j}$. This is shown in Fig. 3.3. The subvolumes composing the secondary staggered mesh are

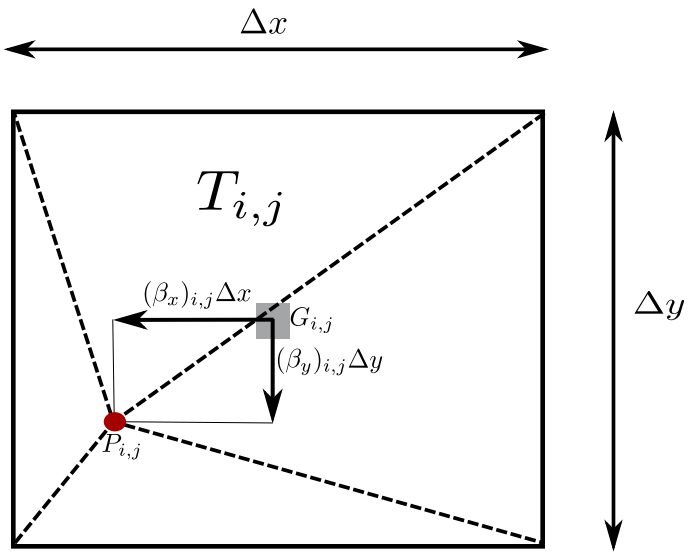

Figure 3.3: The upwind bias parameters in the UFORCE scheme.

then obtained by joining the four vertices of cell $T_{i,j}$ with $\vec{P}_{i,j}$. The resulting primary and secondary mesh setup are illustrated in Fig. 3.4.

With *upwind bias* we define a set of two dimensionless parameters $-\frac{1}{2} \leq (\beta_x)_{i,j} \leq \frac{1}{2}$ and $-\frac{1}{2} \leq (\beta_y)_{i,j} \leq \frac{1}{2}$ for each primary cell controlling the distance between $\vec{P}_{i,j}$ and the cell barycentre:

$$\vec{P}_{i,j} - \vec{G}_{i,j} \equiv \left(-(\beta_x)_{i,j}\Delta x, -(\beta_y)_{i,j}\Delta y \right) . \tag{3.40}$$

Positive values of the directional bias are assumed in the $i-1$ and $j-1$ direction. See Fig. 3.3, where positive values for the bias are depicted. Notice that the case $\left((\beta_x)_{i,j}, (\beta_y)_{i,j}\right) = \vec{0}$ reproduces the standard FORCE scheme. The areas of volumes in the secondary mesh are then given as function of the upwind bias (3.40) in the form

$$\begin{aligned}
\left|V_{i+\frac{1}{2},j}\right| &= \left\{ \left|V^-_{i+\frac{1}{2},j}\right| = \tfrac{\Delta y \Delta x}{2}\left(\tfrac{1}{2} + (\beta_x)_{i,j}\right) \right\} + \left\{ \left|V^+_{i+\frac{1}{2},j}\right| = \tfrac{\Delta y \Delta x}{2}\left(\tfrac{1}{2} - (\beta_x)_{i+1,j}\right) \right\} , \\
\left|V_{i-\frac{1}{2},j}\right| &= \left\{ \left|V^-_{i-\frac{1}{2},j}\right| = \tfrac{\Delta y \Delta x}{2}\left(\tfrac{1}{2} - (\beta_x)_{i,j}\right) \right\} + \left\{ \left|V^+_{i-\frac{1}{2},j}\right| = \tfrac{\Delta y \Delta x}{2}\left(\tfrac{1}{2} + (\beta_x)_{i-1,j}\right) \right\} , \\
\left|V_{i,j+\frac{1}{2}}\right| &= \left\{ \left|V^-_{i,j+\frac{1}{2}}\right| = \tfrac{\Delta x \Delta y}{2}\left(\tfrac{1}{2} + (\beta_y)_{i,j}\right) \right\} + \left\{ \left|V^+_{i,j+\frac{1}{2}}\right| = \tfrac{\Delta x \Delta y}{2}\left(\tfrac{1}{2} - (\beta_y)_{i,j+1}\right) \right\} , \\
\left|V_{i,j-\frac{1}{2}}\right| &= \left\{ \left|V^-_{i,j-\frac{1}{2}}\right| = \tfrac{\Delta x \Delta y}{2}\left(\tfrac{1}{2} - (\beta_y)_{i,j}\right) \right\} + \left\{ \left|V^+_{i,j-\frac{1}{2}}\right| = \tfrac{\Delta x \Delta y}{2}\left(\tfrac{1}{2} + (\beta_y)_{i,j-1}\right) \right\} .
\end{aligned} \tag{3.41}$$

Following an identical procedure as that used in the derivation of the FORCE method (see section 3.1.2) the derivation of UFORCE is achieved in three stages:

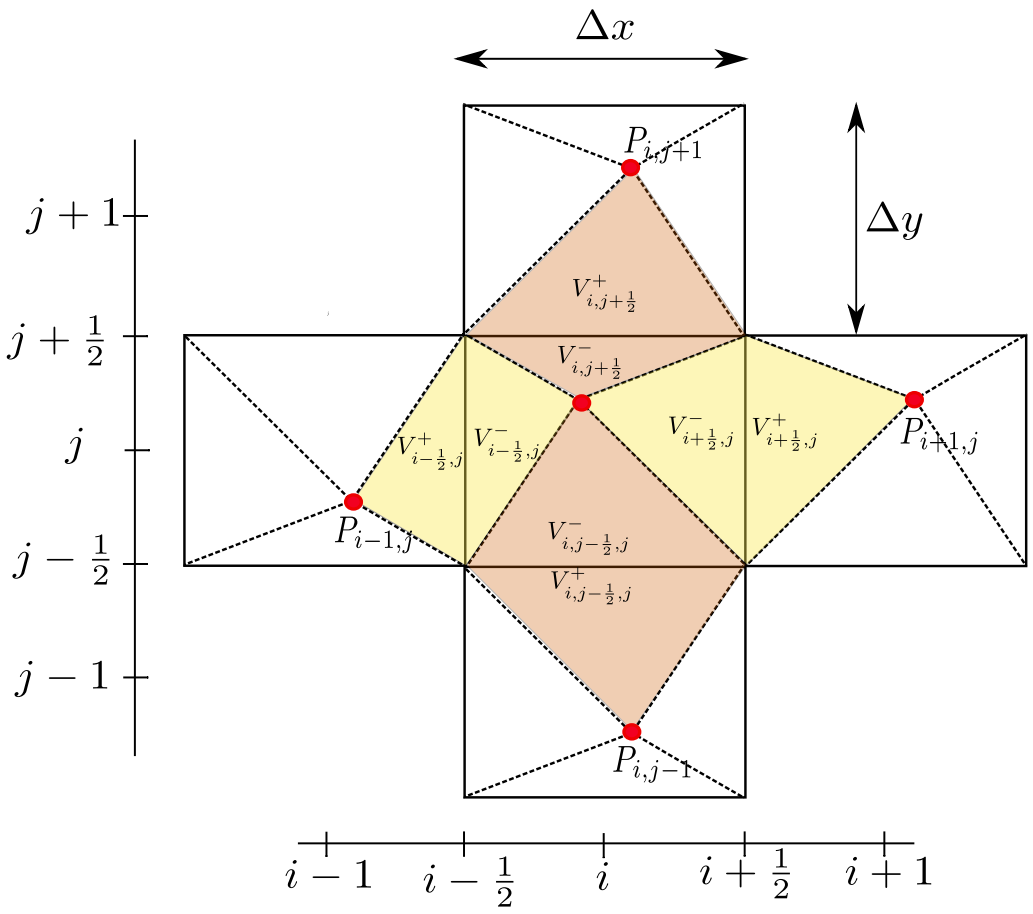

Figure 3.4: Sketch of the primary and secondary mesh for the UFORCE method set on two-dimensional Cartesian meshes.

- In the first stage, assuming initial condition $\mathbf{Q}^n_{i,j}$ in cells $T_{i,j}$ at time t^n, cell averages are evolved for a time step $\frac{1}{2}\Delta t$ in the control volumes of the staggered mesh associated with each edge of $T_{i,j}$. The resulting intermediate states read

$$
\begin{aligned}
\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2},j} &= \frac{1}{\left|V^-_{i+\frac{1}{2},j}\right|+\left|V^+_{i+\frac{1}{2},j}\right|} \quad \left\{\left|V^-_{i+\frac{1}{2},j}\right|\mathbf{Q}^n_{i,j}+\left|V^+_{i+\frac{1}{2},j}\right|\mathbf{Q}^n_{i+1,j}\right. \\
&\qquad \left. -\frac{\Delta t\Delta y}{2}\left(\mathbf{F}(\mathbf{Q}^n_{i+1,j})-\mathbf{F}(\mathbf{Q}^n_{i,j})\right)\right\}, \\
\mathbf{Q}^{n+\frac{1}{2}}_{i,j+\frac{1}{2}} &= \frac{1}{\left|V^-_{i,j+\frac{1}{2}}\right|+\left|V^+_{i,j+\frac{1}{2}}\right|} \quad \left\{\left|V^-_{i,j+\frac{1}{2}}\right|\mathbf{Q}^n_{i,j}+\left|V^+_{i,j+\frac{1}{2}}\right|\mathbf{Q}^n_{i,j+1}\right. \\
&\qquad \left. -\frac{\Delta t\Delta x}{2}\left(\mathbf{G}(\mathbf{Q}^n_{i,j+1})-\mathbf{G}(\mathbf{Q}^n_{i,j})\right)\right\}, \\
\mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2},j} &= \frac{1}{\left|V^-_{i-\frac{1}{2},j}\right|+\left|V^+_{i-\frac{1}{2},j}\right|} \quad \left\{\left|V^+_{i-\frac{1}{2},j}\right|\mathbf{Q}^n_{i-1,j}+\left|V^-_{i-\frac{1}{2},j}\right|\mathbf{Q}^n_{i,j}\right. \\
&\qquad \left. -\frac{\Delta t\Delta y}{2}\left(\mathbf{F}(\mathbf{Q}^n_{i,j})-\mathbf{F}(\mathbf{Q}^n_{i-1,j})\right)\right\}, \\
\mathbf{Q}^{n+\frac{1}{2}}_{i,j-\frac{1}{2}} &= \frac{1}{\left|V^-_{i,j-\frac{1}{2}}\right|+\left|V^+_{i,j-\frac{1}{2}}\right|} \quad \left\{\left|V^-_{i,j-\frac{1}{2}}\right|\mathbf{Q}^n_{i,j}+\left|V^+_{i,j-\frac{1}{2}}\right|\mathbf{Q}^n_{i,j-1}\right. \\
&\qquad \left. -\frac{\Delta t\Delta x}{2}\left(\mathbf{G}(\mathbf{Q}^n_{i,j})-\mathbf{G}(\mathbf{Q}^n_{i,j-1})\right)\right\}.
\end{aligned}
\tag{3.42}
$$

- In the second stage of the scheme derivation the solution is evolved by another time step $\frac{1}{2}\Delta t$ within each element $T_{i,j}$, assuming (3.42) at time $t = t^n + \frac{1}{2}\Delta t$ as initial condition. We obtain

$$
\begin{aligned}
\mathbf{Q}^{n+1}_{i,j} &= \frac{1}{\Delta x\Delta y}\left(\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2},j}\left|V^-_{i+\frac{1}{2},j}\right|+\mathbf{Q}^{n+\frac{1}{2}}_{i,j+\frac{1}{2}}\left|V^-_{i,j+\frac{1}{2}}\right|+\mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2},j}\left|V^-_{i-\frac{1}{2},j}\right|+\mathbf{Q}^{n+\frac{1}{2}}_{i,j-\frac{1}{2}}\left|V^-_{i,j-\frac{1}{2}}\right|\right) \\
&- \frac{1}{2}\frac{\Delta t}{\Delta x}\left\{\mathbf{F}\left(\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2},j}\right)-\mathbf{F}\left(\mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2},j}\right)\right\} \\
&- \frac{1}{2}\frac{\Delta t}{\Delta y}\left\{\mathbf{G}\left(\mathbf{Q}^{n+\frac{1}{2}}_{i,j+\frac{1}{2}}\right)-\mathbf{G}\left(\mathbf{Q}^{n+\frac{1}{2}}_{i,j-\frac{1}{2}}\right)\right\}.
\end{aligned}
\tag{3.43}
$$

- Finally we perform algebraic manipulations on (3.43) so as to reproduce the conservative formula (3.9). As for the classical FORCE method the intercell flux can be rewritten as the arithmetic average of two fluxes, namely

$$
\mathbf{F}^{UFORCE-2}_{i+\frac{1}{2},j} = \frac{1}{2}\left(\mathbf{F}^{ULW-2}_{i+\frac{1}{2},j}+\mathbf{F}^{ULF-2}_{i+\frac{1}{2},j}\right). \tag{3.44}
$$

Here $\mathbf{F}^{ULW-2}_{i+\frac{1}{2},j}$ and $\mathbf{F}^{ULF-2}_{i+\frac{1}{2},j}$ represent two-dimensional upwind-biased versions of the Lax-Wendroff (3.15)-(3.16) and Lax-Friedrichs (3.17) fluxes. They read

$$
\mathbf{F}^{ULW-2}_{i+\frac{1}{2},j} = \mathbf{F}\left(\mathbf{Q}^{ULW-2}_{i+\frac{1}{2},j}\right) \tag{3.45}
$$

with

$$\mathbf{Q}_{i+\frac{1}{2},j}^{ULW-2} = \frac{1}{\left|V_{i+\frac{1}{2},j}^{+}\right|+\left|V_{i+\frac{1}{2},j}^{-}\right|}\left\{\left|V_{i+\frac{1}{2},j}^{-}\right|\mathbf{Q}_{i,j}^{n}+\left|V_{i+\frac{1}{2},j}^{+}\right|\mathbf{Q}_{i+1,j}^{n} - \frac{1}{2}\Delta t\Delta y\left(\mathbf{F}\left(\mathbf{Q}_{i+1,j}^{n}\right)-\mathbf{F}\left(\mathbf{Q}_{i,j}^{n}\right)\right)\right\} \tag{3.46}$$

and

$$\mathbf{F}_{i+\frac{1}{2},j}^{ULF-2} = \frac{1}{\left|V_{i+\frac{1}{2},j}^{+}\right|+\left|V_{i+\frac{1}{2},j}^{-}\right|}\left\{\left|V_{i+\frac{1}{2},j}^{-}\right|\mathbf{F}(\mathbf{Q}_{i+1,j}^{n})+\left|V_{i+\frac{1}{2},j}^{+}\right|\mathbf{F}(\mathbf{Q}_{i,j}^{n}) -2\left(\frac{1}{\Delta t\Delta y}\right)\left|V_{i+\frac{1}{2},j}^{-}\right|\left|V_{i+\frac{1}{2},j}^{+}\right|\left(\mathbf{Q}_{i+1,j}^{n}-\mathbf{Q}_{i,j}^{n}\right)\right\}. \tag{3.47}$$

Last, we generalise the UFORCE flux for α space dimension, obtaining

$$\mathbf{F}_{i+\frac{1}{2}}^{UFORCE-\alpha} = \frac{1}{2}\left(\mathbf{F}_{i+\frac{1}{2}}^{ULW-\alpha}+\mathbf{F}_{i+\frac{1}{2}}^{ULF-\alpha}\right). \tag{3.48}$$

Here

$$\mathbf{F}_{i+\frac{1}{2}}^{ULW-\alpha} = \mathbf{F}\left(\mathbf{Q}_{i+\frac{1}{2}}^{ULW-\alpha}\right) \tag{3.49}$$

with

$$\mathbf{Q}_{i+\frac{1}{2}}^{ULW-\alpha} = \frac{1}{2\left(1-(\beta_x)_{i+1}+(\beta_x)_i\right)}\left\{\left(1+2\left(\beta_x\right)_i\right)\mathbf{Q}_i^n+\left(1-2\left(\beta_x\right)_{i+1}\right)\mathbf{Q}_{i+1}^n - \left(\frac{\alpha\Delta t}{\Delta x}\right)\left(\mathbf{F}\left(\mathbf{Q}_{i+1}^n\right)-\mathbf{F}\left(\mathbf{Q}_i^n\right)\right)\right\} \tag{3.50}$$

and

$$\mathbf{F}_{i+\frac{1}{2}}^{ULF-\alpha} = \frac{1}{2\left(1-(\beta_x)_{i+1}+(\beta_x)_i\right)}\left\{\left(1+2\left(\beta_x\right)_i\right)\mathbf{F}\left(\mathbf{Q}_{i+1}^n\right)+\left(1-2\left(\beta_x\right)_{i+1}\right)\mathbf{F}\left(\mathbf{Q}_i^n\right) - \left(\frac{\Delta x}{\alpha\Delta t}\right)\left(1+2\left(\beta_x\right)_i\right)\left(1-2\left(\beta_x\right)_{i+1}\right)\left(\mathbf{Q}_{i+1,j}^n-\mathbf{Q}_{i,j}^n\right)\right\}. \tag{3.51}$$

In equations (3.48)-(3.51) x indicates the current direction and subscript i is the index along the x direction.

The UFORCE flux is a function of the upwind bias parameters and its performance is expected to depend on their setup. Next, we shall derive optimal values for the upwind bias, taking into account upwind information based on the system eigenvalues.

3.2.2 Optimal choice for the upwind bias: the linear case

Here we analyse how to determine the optimal bias for the UFORCE scheme taking into account information from the eigenvalues of the system. The analysis is based on the two-dimensional

linear advection equation with constant coefficients, which reads

$$\partial_t \mathrm{q} + \partial_x f(\mathrm{q}) + \partial_y g(\mathrm{q}) = 0\,, \tag{3.52}$$

having fluxes $f(\mathrm{q}) = \lambda_x \mathrm{q}$ and $g(\mathrm{q}) = \lambda_y \mathrm{q}$, where λ_x and λ_y are the (constant) characteristic speeds in the x and y direction.

Application of the optimal upwind bias into the UFORCE flux (3.44)-(3.47) shall result in the least dissipative monotone method in the solution of (3.52). We proceed as follows: in section (3.2.2.1) we demonstrate that the Godunov upwind method is the least dissipative monotone method among five-point schemes on two-dimensional Cartesian meshes and then in section (3.2.2.2) we identify the optimal upwind bias, which allows UFORCE to identically reproduce the Godunov method in applications to (3.52).

3.2.2.1 Accuracy of the two-dimensional Godunov upwind method

Finite volume schemes for (3.52) are written as

$$\mathrm{q}_{i,j}^{n+1} = \mathrm{q}_{i,j}^{n} - \frac{\Delta t}{\Delta x}\left(f_{i+\frac{1}{2},j} - f_{i-\frac{1}{2},j}\right) - \frac{\Delta t}{\Delta y}\left(g_{i,j+\frac{1}{2}} - g_{i,j-\frac{1}{2}}\right)\,. \tag{3.53}$$

We restrict our analysis to five-point schemes on two-dimensional Cartesian meshes, i.e. the schemes in which the stencil includes the current cell $T_{i,j}$ and its four neighbours $T_{i,j-1}$, $T_{i+1,j}$, $T_{i,j+1}$, $T_{i-1,j}$. This implies that the scheme (3.53) can be rewritten as

$$\mathrm{q}_{i,j}^{n+1} = \sum_{l=-1}^{1}\sum_{m=-1}^{1} b_{l,m}\mathrm{q}_{i+l,j+m}^{n} \qquad \text{with} \qquad b_{\pm1,\pm1} = 0\,. \tag{3.54}$$

All the schemes considered in this section belong to this class.

We recall the Godunov upwind method, whose flux in the one-dimensional case is given by

$$\mathbf{F}_{i+\frac{1}{2}}^{Godunov} = \mathbf{F}\left(\mathbf{Q}_{i+\frac{1}{2}}\left(0^{+}\right)\right)\,, \tag{3.55}$$

where $\mathbf{Q}_{i+\frac{1}{2}}(0^{+})$ is the solution of the Riemann problem

$$\left.\begin{array}{rl} \text{PDEs:} & \partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) = \mathbf{0}\,, \\ \text{IC:} & \mathbf{Q}(x,0) = \left\{\begin{array}{lll} \mathbf{Q}_i^n & \text{if} & x < x_{i+\frac{1}{2}} \\ \mathbf{Q}_{i+1}^n & \text{if} & x > x_{i+\frac{1}{2}} \end{array}\right. \end{array}\right\} \tag{3.56}$$

obtained at local time $\tau = 0^+$ at intercell $x_{i+\frac{1}{2}}$. Here we consider a two-dimensional extension of the Godunov method in which fluxes are calculated by solving one-dimensional Riemann problems (3.56) orthogonally to the current edge.

The coefficients in (3.54) of the Godunov upwind method in two space dimensions read

$$\begin{array}{lll} b_{-1,0} = \frac{1}{2}\left(c_x + |c_x|\right) & , & b_{1,0} = \frac{1}{2}\left(|c_x| - c_x\right) \ , \\ b_{0,-1} = \frac{1}{2}\left(c_y + |c_y|\right) & , & b_{0,1} = \frac{1}{2}\left(|c_y| - c_y\right) \ , \\ b_{0,0} = 1 - |c_x| - |c_y| \ , & & \end{array} \tag{3.57}$$

where $c_x = \lambda_x \frac{\Delta t}{\Delta x}$ and $c_y = \lambda_y \frac{\Delta t}{\Delta y}$ are the directional CFL numbers. The numerical scheme (3.54), (3.57) proves to be stable if the following condition holds:

$$|c_x| + |c_y| \leq 1 \ . \tag{3.58}$$

Proposition 3.2.1. The Godunov upwind scheme in two space dimensions for the linear advection equation is the monotone scheme with the smallest truncation error among all the five-point schemes (3.54).

Proof. We consider the generic five-point scheme (3.54). Coefficients $b_{l,m}$, $-1 \leq l,m \leq 1$ are constant. We assume the scheme to be at least first-order, that is consistent. Then from Roe's accuracy lemma (see [118]) we have the following three equations relating the five coefficients:

$$\left\{ \begin{array}{ll} \sum_{l=-1}^{1} \sum_{m=-1}^{1} b_{l,m} = 1 & (b_{\pm 1,\pm 1} = 0) \\ b_{-1,0} - b_{1,0} = c_x & \\ b_{0,-1} - b_{0,1} = c_y & \end{array} \right. . \tag{3.59}$$

System (3.59) gives a two-parameter family of solutions. We set

$$b_{-1,0} + b_{1,0} = \hat{\mathrm{q}}_x \quad , \quad b_{0,-1} + b_{0,1} = \hat{\mathrm{q}}_y \tag{3.60}$$

and solve the complete system in terms of the arbitrary parameters $\hat{\mathrm{q}}_x$ and $\hat{\mathrm{q}}_y$:

$$\begin{array}{lll} b_{-1,0} = & \frac{1}{2}\left(\hat{\mathrm{q}}_x + c_x\right) & , \quad b_{0,-1} = \frac{1}{2}\left(\hat{\mathrm{q}}_y + c_y\right) \ , \\ b_{1,0} = & \frac{1}{2}\left(\hat{\mathrm{q}}_x - c_x\right) & , \quad b_{0,1} = \frac{1}{2}\left(\hat{\mathrm{q}}_y - c_y\right) \ , \\ b_{0,0} = & 1 - \left(\hat{\mathrm{q}}_x + \hat{\mathrm{q}}_y\right) \ . & \end{array} \tag{3.61}$$

By performing a truncation error analysis, both cross contributions to numerical viscosity (proportional to $\partial_x \partial_y \mathrm{q}$) and normal contributions (proportional to $\partial_x^2 \mathrm{q}$, $\partial_y^2 \mathrm{q}$) are found. The cross

coefficients of viscosity, reading

$$\mu_{x,y} = -\frac{\Delta x \Delta y}{\Delta t} c_x c_y \tag{3.62}$$

are independent from $\hat{\mathrm{q}}_x$ and $\hat{\mathrm{q}}_y$. Instead, the normal coefficients of viscosity in the x and y direction, reading

$$\mu_{x,x} = \frac{\Delta x^2}{2\Delta t}\left(\hat{\mathrm{q}}_x - c_x^2\right) \qquad , \qquad \mu_{y,y} = \frac{\Delta y^2}{2\Delta t}\left(\hat{\mathrm{q}}_y - c_y^2\right) \tag{3.63}$$

depend on the parameters $\hat{\mathrm{q}}_x$ and $\hat{\mathrm{q}}_y$. In order to minimise the truncation error, the normal contributions (3.63) must be minimised.

Monotonicity requires positivity (non-negativity) of all coefficients (3.61). This leads to the conditions

$$\begin{cases} |c_x| \leq \hat{\mathrm{q}}_x \leq 1 - \hat{\mathrm{q}}_y \\ |c_y| \leq \hat{\mathrm{q}}_y \leq 1 - \hat{\mathrm{q}}_x \end{cases} . \tag{3.64}$$

From (3.64), being $|c_x| \geq 0$ and $|c_y| \geq 0$, a wider set of inequalities can be written as

$$\begin{cases} 0 \leq |c_x| \leq \hat{\mathrm{q}}_x \leq 1 - \hat{\mathrm{q}}_y \leq 1 \\ 0 \leq |c_y| \leq \hat{\mathrm{q}}_y \leq 1 - \hat{\mathrm{q}}_x \leq 1 \end{cases} , \tag{3.65}$$

from which the widest admissible range of variation for $|c_x|$, $|c_y|$ and $\hat{\mathrm{q}}_x$, $\hat{\mathrm{q}}_y$ can be found:

$$\begin{cases} 0 \leq |c_x| \leq 1 \\ 0 \leq |c_y| \leq 1 \end{cases} \tag{3.66}$$

and

$$\begin{cases} |c_x| \leq \hat{\mathrm{q}}_x \leq 1 \\ |c_y| \leq \hat{\mathrm{q}}_y \leq 1 \end{cases} . \tag{3.67}$$

For any given value of $|c_x|$, $|c_y|$ in the range (3.66), taking into account (3.67) and noting that $\mu_{x,x}$ and $\mu_{y,y}$ in (3.63) are linearly increasing functions of $\hat{\mathrm{q}}_x$, $\hat{\mathrm{q}}_y$, we can write:

$$\begin{aligned} \min_{|c_x| \leq \hat{\mathrm{q}}_x \leq 1}\left(\mu_{x,x}\right) &= \mu_{x,x}\left(\min_{|c_x| \leq \hat{\mathrm{q}}_x \leq 1}\left(\hat{\mathrm{q}}_x\right)\right) = \mu_{x,x}\big|_{\hat{\mathrm{q}}_x = |c_x|} , \\ \min_{|c_y| \leq \hat{\mathrm{q}}_y \leq 1}\left(\mu_{y,y}\right) &= \mu_{y,y}\left(\min_{|c_y| \leq \hat{\mathrm{q}}_y \leq 1}\left(\hat{\mathrm{q}}_y\right)\right) = \mu_{y,y}\big|_{\hat{\mathrm{q}}_y = |c_y|} . \end{aligned} \tag{3.68}$$

Therefore, the values of $\hat{\mathrm{q}}_x$ and $\hat{\mathrm{q}}_y$ which minimise $\mu_{x,x}$ and $\mu_{y,y}$ in (3.63), also satisfying (3.67), are

$$\begin{cases} \hat{\mathrm{q}}_x = |c_x| \\ \hat{\mathrm{q}}_y = |c_y| \end{cases} . \tag{3.69}$$

In order to fully satisfy the monotonicity requirements (3.64), by substituting (3.69) into (3.64) a further condition is found:

$$|c_x| + |c_y| \leq 1 . \tag{3.70}$$

Substitution of (3.69) into (3.61) gives the Godunov upwind coefficients (3.57), while (3.70) turns out to be the Godunov upwind stability condition (3.58). □

This proof can be easily extended in three space dimensions.

3.2.2.2 The optimal upwind bias in UFORCE

All the numerical methods considered in this section, i.e. the UFORCE (3.44)-(3.47), FORCE (3.13), Lax-Friedrichs (3.17), Lax-Wendroff (3.15)-(3.16) and Godunov upwind (3.55)-(3.56) methods belong to the class of five-point schemes. The coefficients of these methods to be inserted in (3.54) are given in Tab. 3.1. In the considered linear case we assume upwind bias β_x and β_y constant all over the domain and indicate with $c_x = \frac{\lambda_x \Delta t}{\Delta x}$ and $c_y = \frac{\lambda_y \Delta t}{\Delta y}$ the directional CFL numbers. All the methods given in Tab. 3.1 are directionally split, i.e. the coefficients $b_{-1,0}$ and $b_{1,0}$ depend only on parameters evaluated along the x direction (c_x and/or β_x) while $b_{0,-1}$ and $b_{0,1}$ depend on quantities along the y-direction (c_y and/or β_y).

The UFORCE method is able to exactly reproduce other numerical methods provided suitable values for the upwind bias β_x and β_y are chosen. From Tab. 3.1 we observe that setting $\beta_x = \beta_y = 0$,

Table 3.1: Coefficients of two-dimensional five-point schemes.

	UFORCE	FORCE	Lax-Friedrichs	Godunov	Lax-Wendroff
$b_{-1,0}$	$\frac{1}{8} - \frac{1}{2}\beta_x^2 + \frac{1}{2}c_x^2 + \frac{1}{2}c_x$	$\frac{1}{8} + \frac{1}{2}c_x^2 + \frac{1}{2}c_x$	$\frac{1}{2}\left(\frac{1}{2} + c_x\right)$	$\frac{1}{2}\left(c_x + \lvert c_x\rvert\right)$	$c_x\left(\frac{1}{2} + c_x\right)$
$b_{1,0}$	$\frac{1}{8} - \frac{1}{2}\beta_x^2 + \frac{1}{2}c_x^2 - \frac{1}{2}c_x$	$\frac{1}{8} + \frac{1}{2}c_x^2 - \frac{1}{2}c_x$	$\frac{1}{2}\left(\frac{1}{2} - c_x\right)$	$\frac{1}{2}\left(\lvert c_x\rvert - c_x\right)$	$-c_x\left(\frac{1}{2} - c_x\right)$
$b_{0,0}$	$\frac{1}{2} + \beta_x^2 + \beta_y^2 - c_x^2 - c_y^2$	$\frac{1}{2} - c_x^2 - c_y^2$	0	$1 - \lvert c_x\rvert - \lvert c_y\rvert$	$1 - 2c_x^2 - 2c_y^2$
$b_{0,-1}$	$\frac{1}{8} - \frac{1}{2}\beta_y^2 + \frac{1}{2}c_y^2 + \frac{1}{2}c_y$	$\frac{1}{8} + \frac{1}{2}c_y^2 + \frac{1}{2}c_y$	$\frac{1}{2}\left(\frac{1}{2} + c_y\right)$	$\frac{1}{2}\left(c_y + \lvert c_y\rvert\right)$	$c_y\left(\frac{1}{2} + c_y\right)$
$b_{0,1}$	$\frac{1}{8} - \frac{1}{2}\beta_y^2 + \frac{1}{2}c_y^2 - \frac{1}{2}c_y$	$\frac{1}{8} + \frac{1}{2}c_y^2 - \frac{1}{2}c_y$	$\frac{1}{2}\left(\frac{1}{2} - c_y\right)$	$\frac{1}{2}\left(\lvert c_y\rvert - c_y\right)$	$-c_y\left(\frac{1}{2} - c_y\right)$

the classical FORCE method (3.13) is exactly recovered, while setting

$$\beta_x = \pm\sqrt{\frac{1}{4} - c_x^2} \qquad , \qquad \beta_y = \pm\sqrt{\frac{1}{4} - c_y^2} \tag{3.71}$$

the Lax-Wendroff method (3.15)-(3.16) is reproduced and assuming

$$\beta_x^2 - c_x^2 + \frac{1}{4} = 0 \qquad , \qquad \beta_y^2 - c_y^2 + \frac{1}{4} = 0 \tag{3.72}$$

the Lax-Friedrichs scheme (3.17) is obtained.
Amongst all the possibilities, the optimal choice for the upwind bias is the choice providing the first-order monotone scheme with the smallest truncation error. In section 3.2.2.1 we have proved that among five-point schemes (3.54) the first-order monotone method with the smallest truncation error is the Godunov upwind method. By forcing the coefficients of the UFORCE method to be equal to those of the Godunov method (see Tab. 3.1), we obtain

$$\beta_x = \pm\frac{1}{2}\left(1 - 2\left|c_x\right|\right) \qquad , \qquad \beta_y = \pm\frac{1}{2}\left(1 - 2\left|c_y\right|\right) . \tag{3.73}$$

Thus equation (3.73) denotes the *optimal upwind bias* in UFORCE.
The optimal upwind bias (3.73) also defines the monotonicity region of UFORCE: it is easy to prove that the UFORCE method is monotone provided

$$|\beta_x| \leq \frac{1}{2}(1 - 2\left|c_x\right|) \qquad \text{and} \qquad |\beta_y| \leq \frac{1}{2}(1 - 2\left|c_y\right|) . \tag{3.74}$$

Furthermore, the optimal upwind bias (3.73) gives the same stability region as in the Godunov upwind method (3.58).
In Fig. 3.5 we represent the above choices by plotting these relationships between the upwind bias and the directional Courant number. The directionally-split property of methods allows us to draw the same relations in the $c_x - \beta_x$ or $c_y - \beta_y$ plane, independently.

Finally, if we restrict the stability condition assuming $|c_x| \leq \frac{1}{2}$ and $|c_y| \leq \frac{1}{2}$, the following considerations can be drawn:

- the FORCE scheme is reproduced for any c_x (c_y) assuming $\beta_x = 0$ ($\beta_y = 0$);
- the Godunov upwind method is represented by a quadrilateral with vertices ($c_x = \pm\frac{1}{2}, \beta_x = 0$), ($c_x = 0, \beta_x = \pm\frac{1}{2}$) which is inscribed in a circle (radius= $\frac{1}{2}$) representing the Lax-Wendroff method;
- no choice of β_x (β_y) exists which can reproduce the Lax-Friedrichs method except for $|c_x| = \frac{1}{2}$

($|c_y| = \frac{1}{2}$);

- any choice of β_x (β_y) lying within the quadrilateral representing the Godunov upwind method gives stable and monotone methods.

The normal numerical viscosity coefficients of the methods in Fig. 3.5 are

$$\mu_{xx} = \frac{1}{4}\frac{\Delta x^2}{\Delta t}\left(1-4\beta_x^2\right) \qquad , \qquad \mu_{yy} = \frac{1}{4}\frac{\Delta y^2}{\Delta t}\left(1-4\beta_y^2\right) \ . \tag{3.75}$$

The scheme possessing the maximum viscosity is the FORCE method ($\beta_x = 0$, $\beta_y = 0$), while increasing the bias to $\beta_x = \pm\frac{1}{2}$ ($\beta_y = \pm\frac{1}{2}$), numerical viscosity vanishes, thus giving second order methods. The monotone scheme possessing the smallest numerical viscosity is the Godunov method. Exceeding the threshold imposed by the Godunov method, numerical viscosity decreases, but oscillatory methods are found. Finally, setting $c_x = \pm\frac{1}{2}$ and $\beta_x = 0$, all the methods given in Tab. 3.1 coincide.

The analysis presented in this section can be easily generalised to α space dimensions. The optimal upwind bias is then given by:

$$\beta_x = \pm\frac{1}{2}\left(1-\alpha|c_x|\right) \ , \tag{3.76}$$

where x denotes the current direction and c_x is the directional Courant number.

3.2.3 Optimal choice of the upwind bias: the non-linear case

Here we empirically extend the theory developed in the previous section for the linear case to non-linear homogeneous hyperbolic systems. For the purpose at hand we consider the two-dimensional non-linear inviscid shallow water equations over horizontal bed (2.27), having real eigenvalues $\lambda_x^{(1)} = u-a$, $\lambda_x^{(2)} = u$, $\lambda_x^{(3)} = u+a$ along the x direction and $\lambda_y^{(1)} = v-a$, $\lambda_y^{(2)} = v$, $\lambda_y^{(3)} = v+a$ along the y direction (see section 2.3.1). Here, u and v represent velocity along x and y and $a = \sqrt{gD}$ is celerity, being g the acceleration due to gravity and D water depth. With respect to the linear case previously considered, we observe the following differences:

- in the linear case, any choice for the sign in the optimal upwind bias (3.73) gives rise to identical results, while this is not true for the non-linear case;

- in applications to the scalar linear equation (3.52) the local directional Courant number is precisely defined since there is only one possible choice for the eigenvalue, while this is not the case for non-linear systems.

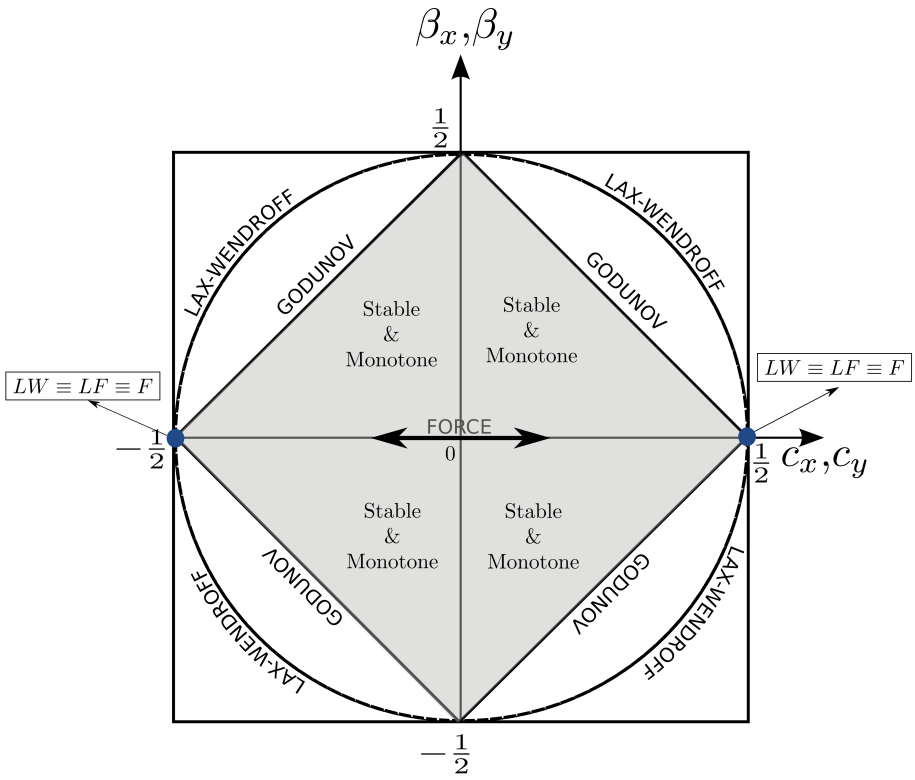

Figure 3.5: The UFORCE method applied to the linear advection equation (3.52): monotonicity and stability regions for different choices of the directional bias in the $\beta_x - c_x$ and in the $\beta_y - c_y$ plane.

The open questions for non-linear systems are thus finding the correct sign s_i and the most suitable estimate for the eigenvalue λ^S to be used in equation (3.73). Based on our numerical experiments, we extend the linear upwind bias (3.73) to the non-linear case in the form

$$(\beta_x)_{i,j} = s_i \left\{ \frac{1}{2} - \lambda^S \left(\frac{\Delta t}{(\Delta x)_{i,j}} \right) \right\} , \tag{3.77}$$

with

$$s_i = \begin{cases} \text{sign}\,(u_{i,j}) & \text{if } (u_{i,j} \neq 0) \\ \text{sign}\left(\left(\lambda_x^{(3)}\right)_{i-1,j} + \left(\lambda_x^{(1)}\right)_{i+1,j} \right) & \text{if } (u_{i,j} = 0) \text{ and } \left(\lambda_x^{(3)}\right)_{i-1,j} + \left(\lambda_x^{(1)}\right)_{i+1,j} \neq 0 \\ 0 & \text{if } (u_{i,j} = 0) \text{ and } \left(\lambda_x^{(3)}\right)_{i-1,j} + \left(\lambda_x^{(1)}\right)_{i+1,j} = 0 \end{cases} . \tag{3.78}$$

In (3.78) the sign function assumes the value of 1 if its argument is positive and -1 if it is negative. This means that in general the sign s_i in (3.77) is chosen, in an upwind fashion, accordingly with the sign of the particle velocity inside the cell. If velocity inside the current cell vanishes, then the choice for the sign is made using upwind information coming from the neighbouring cells, i.e. assuming the sign accordingly to the summation between the maximum eigenvalue coming from the left and minimum eigenvalue coming from the right. Otherwise we choose $s_i = 0$ and end up with $(\beta_x)_{i,j} = 0$, thus recovering the standard FORCE scheme.
In (3.77), λ^S is an estimate for the maximum local characteristic speed in absolute value. A suitable expression in the case of the shallow water equations is

$$\lambda^S = \max_{k=1,2,3} \left| \left(\lambda_x^{(k)} \right)_{i,j} \right| = |u| + a \, . \tag{3.79}$$

Based on the directionally-split property, analogous expressions are found in the y direction.

A generalisation of the non-linear upwind bias (3.77) in α space dimensions is readily achieved in the form

$$(\beta_x)_i = \frac{1}{2} s_i \left\{ 1 - \alpha \lambda^S \left(\frac{\Delta t}{(\Delta x)_i} \right) \right\} \, . \tag{3.80}$$

3.3 The UFORCE-δ scheme

The purpose of this section is to design another upwind-biased FORCE-type method, fully exploiting the idea of adaptively varying the secondary mesh setup first introduced in the derivation of the UFORCE method (section 3.2). With respect to UFORCE, here we remove the constraint that the subvolumes composing the secondary mesh must join in one unique point $\vec{P}$ laying inside the current cell of the primary mesh. In the sought method we will use for the definition of subvolumes as many moving points as the number of edges in the primary cell, in order to allow finer adaptation to local upwind information. The resulting flux, regarded as UFORCE-δ by Stecca *et al.* [113], will contain two upwind bias parameters δ_j^+ and δ_j^- controlling the shape of the secondary mesh. The method derivation will be carried out in the light of the theory for the optimal upwind bias developed in the previous section.
A key requirement for the UFORCE-δ flux is that it shall be generally applicable to any kind of meshes in multi-dimensions, i.e. it shall be natively derived on unstructured meshes. Thus the starting point for the construction of the method will be the multi-dimensional unstructured FORCE-α flux (3.35)-(3.38) [113] presented in section 3.1.4.

3.3.1 Derivation

As for the multi-dimensional FORCE scheme [120], the derivation of the UFORCE-δ method requires the adoption of a primary mesh for computing cell averages and of a staggered secondary mesh used to define numerical fluxes for the conservative form of the scheme. The primary mesh corresponds to a conforming tassellation of the computational domain by n_e elements T_i in α space dimensions (3.25). Each cell T_i has n_f plane interfaces S_j of size $\left|S_j\right|$, with associated outward-pointing normal unit vectors $\vec{n}_j$. The secondary mesh is staggered with respect to the primary mesh. The primary and secondary mesh setup for UFORCE is depicted in Fig. 3.6 for the two-dimensional case on triangular meshes. Each cell of the secondary mesh is composed by two subvolumes V_j^- and V_j^+ associated to interface S_j of the current primary cell T_i, the former laying within cell T_i, the latter laying outside.

In deriving the proposed numerical method we allow the vertex of each subvolume V_j^- to not necessarily join in the barycentre of T_i. Each subvolume is generated independently from the other ones by connecting the vertices of interface S_j to one point $\vec{P}_j$ associated with S_j laying within T_i. Unlike in the UFORCE method developed in the previous section, where a unique moving point per cell $\vec{P}$ was considered in the definition of the secondary mesh, here we use as many moving points $\vec{P}_j$ as the number of edges (compare Figs. 3.6 and 3.4).
For the proposed scheme we impose that each volume V_j^- cannot be greater than the corresponding V_j^- in the FORCE method. This condition ensures that the primary mesh subvolumes V_j^+ have smaller size than their counterparts in the FORCE method. As a consequence, unlike in FORCE and UFORCE, in UFORCE-δ the subvolumes composing the secondary mesh will not entirely cover the computational domain. Since the amount of numerical dissipation associated to the averaging procedure described in (3.27) and (3.28) increases with subvolume size $\left|V_j^{\pm}\right|$, this constraint guarantees that the proposed method will be less dissipative than FORCE.

Adopting an explicit formulation in terms of α as in (3.34) we obtain

$$\left|V_j^-\right| = \frac{\delta_j^- h_j^- \left|S_j\right|}{\alpha} \qquad , \qquad \left|V_j^+\right| = \frac{\delta_j^+ h_j^+ \left|S_j\right|}{\alpha} \, , \tag{3.81}$$

where δ_j^- and δ_j^+ are the upwind bias parameters associated to S_j, which must satisfy the constraint

$$0 \leq \delta_j^- \leq 1 \qquad , \qquad 0 \leq \delta_j^+ \leq 1 \, . \tag{3.82}$$

Once the secondary mesh (3.81) is defined, the UFORCE-δ numerical flux is obtained from (3.30)-(3.33) in the form

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{UFORCE-\delta} = \frac{1}{2}\left\{\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{ULW-\delta}\left(\mathbf{Q}_i^n,\mathbf{Q}_j^n\right)+\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{ULF-\delta}\left(\mathbf{Q}_i^n,\mathbf{Q}_j^n\right)\right\}, \tag{3.83}$$

with

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{ULW-\delta} = \underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_{j+\frac{1}{2}}^{ULW-\delta}\right), \tag{3.84}$$

$$\mathbf{Q}_{j+\frac{1}{2}}^{ULW-\delta} = \frac{1}{\left(\delta_j^- h_j^- + \delta_j^+ h_j^+ + \varepsilon_c\right)}\left\{\left(\mathbf{Q}_i^n\delta_j^- h_j^- + \mathbf{Q}_j^n\delta_j^+ h_j^+\right) - \frac{1}{2}\alpha\Delta t\left(\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_j^n\right)-\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_i^n\right)\right)\cdot\vec{n}_j\right\} \tag{3.85}$$

and

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{ULF-\delta} = \frac{1}{\left(\delta_j^- h_j^- + \delta_j^+ h_j^+ + \varepsilon_c\right)}\left\{\left(\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_j^n\right)\delta_j^- h_j^- + \underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_i^n\right)\delta_j^+ h_j^+\right) - 2\frac{\delta_j^- h_j^-\delta_j^+ h_j^+}{\alpha\Delta t}\left(\mathbf{Q}_j^n-\mathbf{Q}_i^n\right)\vec{n}_j^T\right\}, \tag{3.86}$$

where a slight correction in the denominator has been introduced in order to handle the case of both vanishing δ_j^- and δ_j^+. Here ε_c is a small positive real number, e.g. $\varepsilon_c \sim 10^{-10}$.
At this stage the UFORCE-δ flux is expressed as a function of the upwind bias parameters $\delta_j^\pm$, still to be determined. In the next section we shall derive optimal values for these parameters.

3.3.2 Optimal choice of the upwind bias: the linear case

Different choices for $\delta_j^\pm$ in (3.83)-(3.86) give different numerical methods. Here we concentrate on an adaptive choice of the upwind bias parameters, i.e. a relation governing their variation in space and (if the problem is non-linear) in time.
The purpose of this section is to determine the optimal upwind bias, i.e. the choice of δ_j^+ and δ_j^- providing the least dissipative monotone first order flux. To this aim we adopt an approach which is analogous to that used in the derivation of the optimal upwind bias for UFORCE (see section 3.2.2). The steps are:

- choice of the appropriate linear test problem;
- identification of an existing upwind numerical method to be used as reference;
- evaluation of the optimal upwind bias for the UFORCE-δ method by equating its flux to the reference method flux for the selected test equation.

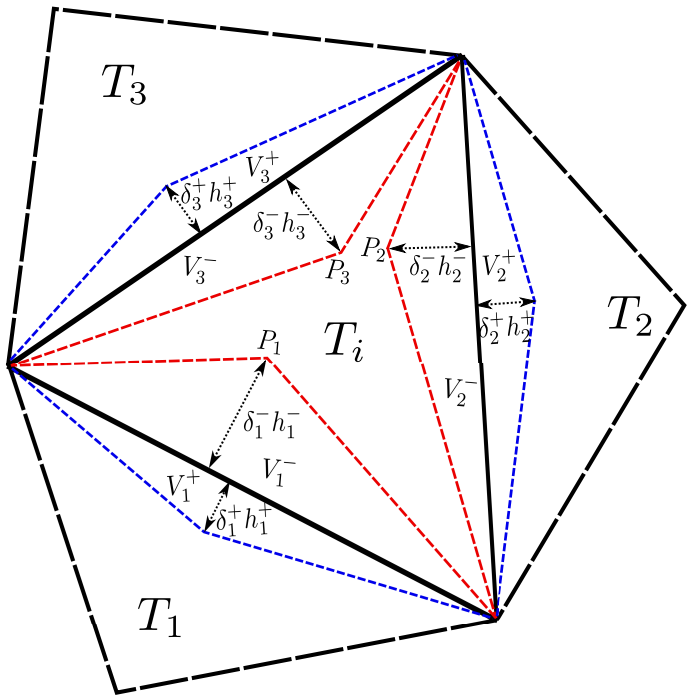

Figure 3.6: Sketch of the primary and secondary mesh for the UFORCE-δ method set on two-dimensional unstructured triangular meshes.

For sake of generality, we perform our analysis on a three-dimensional test problem on an unstructured mesh using the following linear system:

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_y \mathbf{G}(\mathbf{Q}) + \partial_z \mathbf{H}(\mathbf{Q}) = \mathbf{0}\,, \tag{3.87}$$

where $\mathbf{Q} = [\mathrm{q}_1, \mathrm{q}_2]^T$ is the vector of conserved variables and $\mathbf{F}(\mathbf{Q})$, $\mathbf{G}(\mathbf{Q})$, $\mathbf{H}(\mathbf{Q})$ are the x, y, z fluxes, given by

$$\mathbf{F}(\mathbf{Q}) = \mathbf{A}_1\mathbf{Q} \quad , \quad \mathbf{G}(\mathbf{Q}) = \mathbf{A}_2\mathbf{Q} \quad , \quad \mathbf{H}(\mathbf{Q}) = \mathbf{A}_3\mathbf{Q}\,. \tag{3.88}$$

Here, $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ are 2×2 hyperbolic matrices with constant entries. Hyperbolicity of these matrices ensures that each of them possesses two real eigenvalues. Therefore (3.87) presents two waves in each space direction that can be used for evaluating the two bias δ_j^+ and δ_j^-.
We focus on one cell T_i of the considered unstructured mesh, whose boundaries are denoted with S_j. Let $\vec{n}_j$ be the outward-pointing normal unit vector and T_j the neighbouring cell associated to the current boundary. The initial condition at time $t = t^n$ (local time $\tau = 0$) is given by piecewise constant data $\mathbf{Q}(x,y,z \in T_i) = \mathbf{Q}_i^n = [\mathrm{q}_{1i}, \mathrm{q}_{2i}]^T$ and $\mathbf{Q}(x,y,z \in T_j) = \mathbf{Q}_j^n = [\mathrm{q}_{1j}, \mathrm{q}_{2j}]^T$.

The sought reference method shall be monotone and characterised by minimum numerical dissipation. In section 3.2.2.1 for the linear advection equation we have proved that the Godunov upwind scheme on two-dimensional Cartesian meshes is the monotone scheme with the smallest truncation error among all the five-point schemes. Since the same proof is not viable on general unstructured meshes, here we assume a straightforward extension of this result and adopt the Godunov upwind method as reference. The Godunov flux at interface S_j is given by

$$\underline{\underline{\mathbf{F}}}^{Godunov}_{j+\frac{1}{2}}\left(\mathbf{Q}^n_i,\mathbf{Q}^n_j\right)=\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}\left(\mathbf{Q}\left(\hat{n}=0,\tau=0^+\right)\right)\ , \tag{3.89}$$

where $\mathbf{Q}\left(\hat{n}=0,\tau=0^+\right)$ is the solution of a classical one-dimensional Riemann problem projected orthogonally to cell interface S_j and $\hat{n}$ denotes a local normal coordinate defined by $\vec{n}_j$ with origin at S_j.

In order to obtain identical numerical methods we proceed by equating the projected fluxes of the UFORCE-δ and Godunov methods

$$\underline{\underline{\mathbf{F}}}^{UFORCE-\delta}_{j+\frac{1}{2}}\left(\mathbf{Q}^n_i,\mathbf{Q}^n_j\right)\cdot\vec{n}_j=\underline{\underline{\mathbf{F}}}^{Godunov}_{j+\frac{1}{2}}\left(\mathbf{Q}^n_i,\mathbf{Q}^n_j\right)\cdot\vec{n}_j\ , \tag{3.90}$$

in applications to the system (3.87)-(3.88). The Godunov flux for this system is obtained by the solution of the following Riemann problem:

$$\left.\begin{array}{rl} \text{PDEs:} & \partial_t\mathbf{Q}+\partial_{\hat{n}}\left(\mathbf{A}_{\hat{n}}\mathbf{Q}\right)=0\ , \\ \text{IC:} & \mathbf{Q}\left(\hat{n},\tau=0\right)=\left\{\begin{array}{l}\mathbf{Q}^n_i \text{ if } \hat{n}<0\ , \\ \mathbf{Q}^n_j \text{ if } \hat{n}>0\end{array}\right. \end{array}\right\}\ , \tag{3.91}$$

where $\mathbf{A}_{\hat{n}}$ is the projected flux Jacobian matrix:

$$\mathbf{A}_{\hat{n}}=\left[\begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22}\end{array}\right]=\left(\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3\right)\cdot\vec{n}_j\ . \tag{3.92}$$

Hyperbolicity of all the matrices in (3.88) guarantees that the matrix $\mathbf{A}_{\hat{n}}$ is itself hyperbolic. Therefore $\mathbf{A}_{\hat{n}}$ possesses two real eigenvalues $\lambda^{(1)}_{\hat{n}}$ and $\lambda^{(2)}_{\hat{n}}$ (sorted in increasing order), defined as

$$\lambda^{(1)}_{\hat{n}}=\frac{1}{2}\left(a_{11}+a_{22}-R\right)\qquad,\qquad\lambda^{(2)}_{\hat{n}}=\frac{1}{2}\left(a_{11}+a_{22}+R\right)\ , \tag{3.93}$$

with $R=\sqrt{(a_{11}-a_{22})^2+4a_{12}a_{21}}$. The exact solution at cell interface for the Riemann problem (3.91) is

$$\mathbf{Q}\left(\hat{n}=0,\tau=0^+\right)=\begin{cases}\mathbf{Q}_i^n & \text{if } \lambda_{\hat{n}}^{(1)}>0\,,\\ \mathbf{Q}_j^n & \text{if } \lambda_{\hat{n}}^{(2)}<0\,,\\ \begin{bmatrix}\frac{R+a_{11}-a_{22}}{2R^+}q_{1i}+\frac{R-a_{11}+a_{22}}{2R^+}q_{1j}-\frac{a_{12}}{R^+}\left(q_{2j}-q_{2i}\right)\\ -\frac{a_{21}}{R^+}\left(q_{1j}-q_{1i}\right)+\frac{R-a_{11}+a_{22}}{2R^+}q_{2i}+\frac{R+a_{11}-a_{22}}{2R^+}q_{2j}\end{bmatrix} & \text{otherwise}\,,\end{cases} \tag{3.94}$$

where $R^+=(R+\varepsilon_c)$ allows to handle the case of two vanishing projected eigenvalues. Finally, recalling (3.92), the projected Godunov upwind flux reads

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{Godunov}\left(\mathbf{Q}_i^n,\mathbf{Q}_j^n\right)\cdot\vec{n}_j=\mathbf{A}_{\hat{n}}\mathbf{Q}\left(\hat{n}=0,\tau=0^+\right)\,. \tag{3.95}$$

The three-dimensional UFORCE-δ flux for the linear problem (3.87)-(3.88) is obtained by substitution of (3.88) into (3.83)-(3.86). The projected UFORCE-δ flux reads

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{UFORCE-\delta}\cdot\vec{n}_j=\frac{1}{2}\left\{\mathbf{A}_{\hat{n}}-\frac{1}{\left(\delta_j^-h_j^-+\delta_j^+h_j^+ +\varepsilon_c\right)}\left(\frac{\alpha\Delta t}{2}\mathbf{A}_{\hat{n}}^2+\frac{2\left(\delta_j^-h_j^-\delta_j^+h_j^+\right)}{\alpha\Delta t}\mathbf{I}\right)\right\}\left(\mathbf{Q}_j^n-\mathbf{Q}_i^n\right)\,, \tag{3.96}$$

where $\mathbf{I}$ is the 2×2 identity matrix.

Equating the UFORCE-δ flux (3.96) to the Godunov flux (3.95), after algebraic manipulations, two solutions for the optimal upwind bias can be found:

$$\delta_j^-=\frac{\left|\lambda_{\hat{n}}^{(1)}\right|\alpha\Delta t}{2h_j^-}\qquad,\qquad\delta_j^+=\frac{\left|\lambda_{\hat{n}}^{(2)}\right|\alpha\Delta t}{2h_j^+}\,, \tag{3.97}$$

$$\delta_j^-=\frac{\left|\lambda_{\hat{n}}^{(2)}\right|\alpha\Delta t}{2h_j^-}\qquad,\qquad\delta_j^+=\frac{\left|\lambda_{\hat{n}}^{(1)}\right|\alpha\Delta t}{2h_j^+}\,. \tag{3.98}$$

For the linear case the two solutions (3.97) and (3.98) are equivalent, both allowing the UFORCE-δ flux to reproduce the Godunov upwind flux.
In order to satisfy constraint (3.82), the following condition must hold:

$$\frac{\max\left(\left|\lambda_{\hat{n}}^{(1)}\right|,\left|\lambda_{\hat{n}}^{(2)}\right|\right)\Delta t}{2h_j^-}\leq\frac{1}{\alpha}\,, \tag{3.99}$$

for each h_j in the domain. Therefore the time step Δt will be chosen accordingly with the CFL condition

$$\Delta t = \frac{2}{\alpha} CFL \min_{1\leq i\leq n_e} \left(\min_{1\leq j\leq n_f} \left(\frac{h_j^-}{\max\left(\left|\lambda_{\hat{n}}^{(1)}\right|, \left|\lambda_{\hat{n}}^{(2)}\right| \right)} \right) \right)_{T_i} , \tag{3.100}$$

being CFL the Courant-Friedrichs-Lewy coefficient ($0 < CFL \leq 1$).
We are now able to write our UFORCE-δ flux for the considered problem by inserting (3.97) or (3.98) into (3.96):

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{UFORCE-\delta} \cdot \vec{n}_j = \frac{1}{2} \left\{ \mathbf{A}_{\hat{n}} - \frac{1}{\max\left(\left|\lambda_{\hat{n}}^{(1)}\right| + \left|\lambda_{\hat{n}}^{(2)}\right| + \varepsilon_c \right)} \left(\mathbf{A}_{\hat{n}}^2 + \left|\lambda_{\hat{n}}^{(1)}\right| \left|\lambda_{\hat{n}}^{(2)}\right| \mathbf{I} \right) \right\} \left(\mathbf{Q}_j^n - \mathbf{Q}_i^n \right) . \tag{3.101}$$

Comparing (3.96) and (3.101) we observe that the original cell subvolume altitudes $h_j^{\pm}$ have been replaced by adaptive subvolumes whose size is controlled by the local characteristic speeds in absolute value. The amount of numerical dissipation, which is related to the size of secondary subvolumes, is now controlled by local parameters related to characteristic speeds. We remark that given the optimal upwind bias (3.97) or (3.98), the proposed UFORCE-δ method identically reproduces the results of the Godunov upwind method in applications to the linear system (3.87)-(3.88). Next, we will experimentally prove this statement.

3.3.3 Application to a three-dimensional linear system

The aim of the present application is twofold: first to prove that the proposed UFORCE-δ method is identical to the Godunov upwind method in the linear case; second to show that the considered upwind-biased formulation applies to three-dimensional unstructured meshes. In order to prove the above statements we solve the three-dimensional linear system (3.87)-(3.88) where matrices (3.88) are set to

$$\mathbf{A}_1 = \begin{bmatrix} 10 & 6 \\ 2 & -1 \end{bmatrix} \quad , \quad \mathbf{A}_2 = \mathbf{A}_3 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} . \tag{3.102}$$

Matrix $\mathbf{A}_1$ has eigenvalues $\lambda_x^{(1)} = -2$ and $\lambda_x^{(2)} = 11$. The initial condition is represented by a discontinuity located at $x = 0$ for both conserved variables q_1 and q_2, namely:

$$q_1 = q_2 = \begin{cases} 1 & \text{if } x < 0 \\ -1 & \text{otherwise} \end{cases} . \tag{3.103}$$

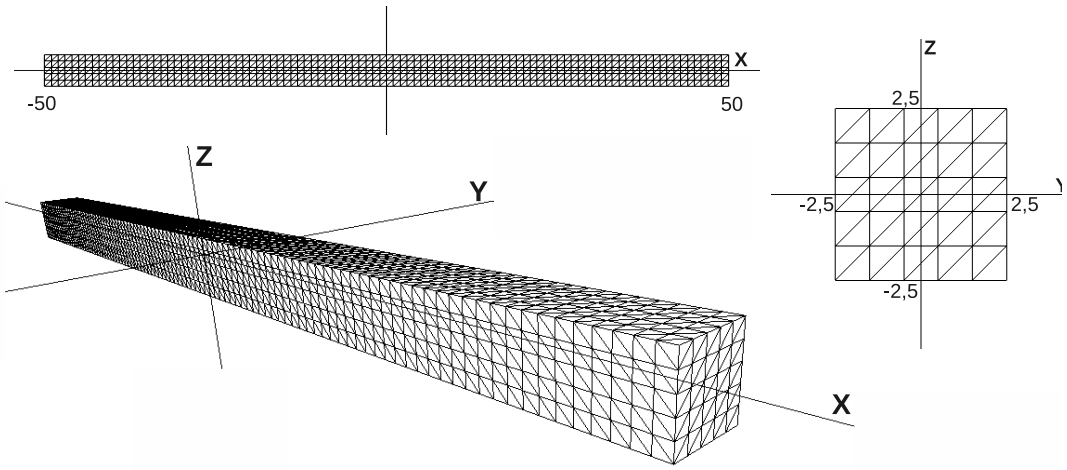

Figure 3.7: Application to the linear system (3.87)-(3.88). Three-dimensional unstructured tetrahedral mesh of 15000 cells.

The system is solved in the domain $[-50,50]\times[-2.5,2.5]\times[-2.5,2.5]$ on a tetrahedral mesh composed of 15000 cells (see Fig. 3.7) and solutions are displayed at final time $t = 3\,s$. Stability is imposed enforcing the CFL condition (3.100) with $\alpha = 3$ and $CFL = 0.9$. In Fig. 3.8 we compare the results of the UFORCE-δ method and of the Godunov upwind method (3.89), (3.94) together with the exact solution. The numerical profiles in Fig. 3.8 have been obtained slicing the numerical solutions along the x axis at $y = z = 0$, while the exact solution has been obtained using the one-dimensional exact Riemann solver (3.94). In this case we have verified that the results of the UFORCE-δ method and of the Godunov upwind method are identical (see Fig. 3.8 for a visual confirmation).

3.3.4 Optimal choice of the upwind bias: the non-linear case

The aim of this section is to find a suitable modification of relations (3.97), (3.98) in order to express the upwind bias as a function of available wave speed estimates for general non-linear hyperbolic systems.

Let us consider the following three-dimensional hyperbolic system of m equations and m unknowns, with $m \geq 2$:

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_y \mathbf{G}(\mathbf{Q}) + \partial_z \mathbf{H}(\mathbf{Q}) = \mathbf{0}\,, \tag{3.104}$$

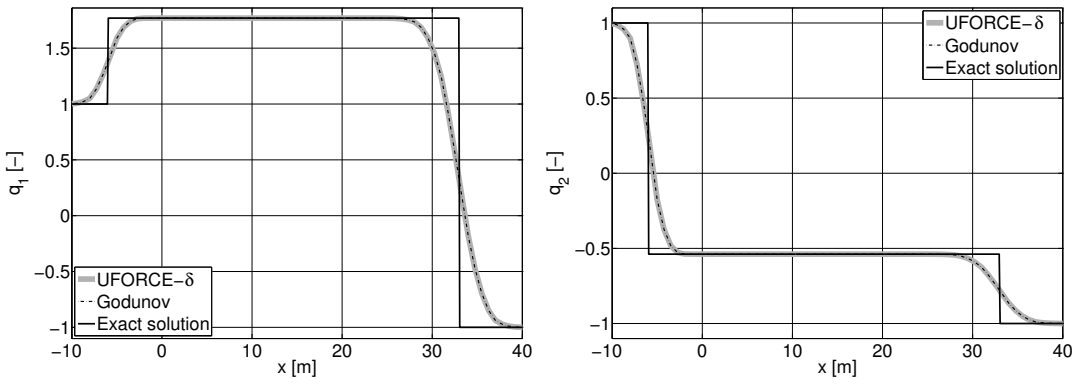

Figure 3.8: Applications to the linear system (3.87)-(3.88). The solution obtained using two numerical methods (Godunov upwind and UFORCE-δ) is compared to the exact solution at $t = 3\,s$. Numerical solution profiles are sliced along the x axis at $y = z = 0$.

where $\mathbf{F}(\mathbf{Q})$, $\mathbf{G}(\mathbf{Q})$ and $\mathbf{H}(\mathbf{Q})$ are the flux vectors in the x, y and z direction respectively. The Jacobian matrices of these flux vectors are

$$\mathbf{A}_1(\mathbf{Q}) = \left[\frac{\partial \mathbf{F}}{\partial \mathbf{Q}}\right] \quad , \quad \mathbf{A}_2(\mathbf{Q}) = \left[\frac{\partial \mathbf{G}}{\partial \mathbf{Q}}\right] \quad , \quad \mathbf{A}_3(\mathbf{Q}) = \left[\frac{\partial \mathbf{H}}{\partial \mathbf{Q}}\right] . \tag{3.105}$$

In order to evaluate the flux associated to interface S_j of cell T_i we consider the projected Jacobian matrix:

$$\mathbf{A}_{\hat{n}}(\mathbf{Q}) = (\mathbf{A}_1(\mathbf{Q}), \mathbf{A}_2(\mathbf{Q}), \mathbf{A}_3(\mathbf{Q})) \cdot \vec{n}_j . \tag{3.106}$$

Unlike in the linear case, Jacobians (3.105) are generally data-dependent. Therefore, given piecewise constant initial data presenting a discontinuity at interface S_j, namely $\mathbf{Q}(x,y,z \in T_i) = \mathbf{Q}_i^n$ and $\mathbf{Q}(x,y,z \in T_j) = \mathbf{Q}_j^n$, we have to consider two projected Jacobian matrices:

$$\mathbf{A}_{i\hat{n}} = \mathbf{A}_{\hat{n}}(\mathbf{Q}_i^n) \quad , \quad \mathbf{A}_{j\hat{n}} = \mathbf{A}_{\hat{n}}\left(\mathbf{Q}_j^n\right) . \tag{3.107}$$

Notice that the two projected Jacobians (3.107) have been obtained with different data, but using the same unit vector $\vec{n}_j$ (outward-pointing from cell T_i). Each of these Jacobians possesses m real eigenvalues (sorted in increasing order), namely $\lambda_{i\hat{n}}^{(1)}, \ldots, \lambda_{i\hat{n}}^{(m)}$ and $\lambda_{j\hat{n}}^{(1)}, \ldots, \lambda_{j\hat{n}}^{(m)}$ respectively.
Compared to the linear case with $m = 2$ we have to address the following questions:

- which waves should be taken into account when dealing with a system possessing more than two waves?
- how the wave speeds can be estimated in practice in order to use an expression based on (3.97) or (3.98) for the upwind bias?

Concerning the first question we consider a two-wave approach similar to HLL [57] or to the central-upwind method developed by Kurganov, Noelle and Petrova (in the following KNP) [70]. Following this approach we conclude that for systems possessing more than two waves we have to consider the smallest and largest characteristic speed of the entire Riemann fan. In the following we will denote these wave speeds as $\mathrm{s}_n^{(1)}$ and $\mathrm{s}_n^{(m)}$. Therefore the optimal upwind bias can be written as

$$\delta_j^- = \frac{\left|\mathrm{s}_{\hat{n}}^{(1)}\right| \alpha \Delta t}{2h_j^-} \qquad , \qquad \delta_j^+ = \frac{\left|\mathrm{s}_{\hat{n}}^{(m)}\right| \alpha \Delta t}{2h_j^+} \tag{3.108}$$

or

$$\delta_j^- = \frac{\left|\mathrm{s}_{\hat{n}}^{(m)}\right| \alpha \Delta t}{2h_j^-} \qquad , \qquad \delta_j^+ = \frac{\left|\mathrm{s}_{\hat{n}}^{(1)}\right| \alpha \Delta t}{2h_j^+} \, . \tag{3.109}$$

The associated CFL condition reads

$$\Delta t = \frac{2}{\alpha} CFL \min_{1 \le i \le n_e} \left(\min_{1 \le j \le n_f} \left(\frac{h_j^-}{\max\left(\left|\mathrm{s}_{\hat{n}}^{(1)}\right|, \left|\mathrm{s}_{\hat{n}}^{(m)}\right| \right)} \right) \right)_{T_i} , \tag{3.110}$$

with $0 < CFL \le 1$.

The second question requires more discussion and different solutions can be obtained using different approaches. Hereby we present two possible choices leading to genuinely centred methods:

1. The first approach considers the smallest and largest wave speeds of the Riemann fan, which are obtained using the eigenvalues of the projected Jacobian matrices (3.107) as linearised wave speed estimates, i.e.

$$\left|\mathrm{s}_{\hat{n}}^{(1)}\right| = \max\left(\left|\lambda_{i\hat{n}}^{(1)}\right|, \left|\lambda_{j\hat{n}}^{(1)}\right| \right) \qquad , \qquad \left|\mathrm{s}_{\hat{n}}^{(m)}\right| = \max\left(\left|\lambda_{i\hat{n}}^{(m)}\right|, \left|\lambda_{j\hat{n}}^{(m)}\right| \right) . \tag{3.111}$$

   The upwind bias is obtained inserting (3.111) into (3.109). Since these eigenvalues are in any case needed for selecting a time step, no computational effort is added. Based on our experience with the two-dimensional inviscid shallow water equations, in order to avoid spurious oscillations we recommend the use of (3.109). The resulting method is genuinely centred since no detailed knowledge of the system eigenstructure is required, therefore being very general and suitable to be applied to systems for which the solution of the Riemann problem is not viable. Moreover, compared to classical centred methods like FORCE (3.30)-(3.33), it will be characterised by reduced numerical dissipation.

2. A second approach is based on a one-wave framework, where both wave speeds are estimated

based on the maximum eigenvalue in absolute value:

$$\left|\mathrm{s}_{\hat{n}}^{(1)}\right| = \left|\mathrm{s}_{\hat{n}}^{(m)}\right| = \max\left(\max_{1\leq l\leq m}\left(\left|\lambda_{i\hat{n}}^{(l)}\right|\right), \max_{1\leq l\leq m}\left(\left|\lambda_{j\hat{n}}^{(l)}\right|\right)\right) . \tag{3.112}$$

In this case use of (3.108) or (3.109) returns identical methods. Use of a one-wave method is mandatory when dealing with scalar equations, possessing only one wave ($m = 1$). In applications to systems of PDEs this choice may be convenient whether only an estimate of the maximum eigenvalue in absolute value is available. The resulting one-wave UFORCE-δ method is expected to be more dissipative than the two-wave UFORCE-δ method given by (3.111), but still more accurate than FORCE.

3.3.5 The UFORCE-δ flux on Cartesian meshes

In this section we derive a Cartesian formulation for the UFORCE-δ flux (3.83)-(3.86). The formulation here proposed is also suitable for one-dimensional applications. We consider again the non-linear system (3.104) and focus only on the fluxes in the x direction, while the other fluxes will be written in analogous manner. Here we use a multi-dimensional Cartesian-type indexation of cells: we indicate with T_i the current cell and with T_{i+1} its right neighbour in the x direction. We assume piecewise constant initial data $\mathbf{Q}_i^n$ and $\mathbf{Q}_{i+1}^n$. The UFORCE-δ flux at interface $S_{i+\frac{1}{2}}$ is given by

$$\mathbf{F}_{i+\frac{1}{2}}^{UFORCE-\delta} = \frac{1}{2}\left\{\mathbf{F}_{i+\frac{1}{2}}^{ULW-\delta}\left(\mathbf{Q}_i^n, \mathbf{Q}_{i+1}^n\right) + \mathbf{F}_{i+\frac{1}{2}}^{ULF-\delta}\left(\mathbf{Q}_i^n, \mathbf{Q}_{i+1}^n\right)\right\} , \tag{3.113}$$

with

$$\mathbf{F}_{i+\frac{1}{2}}^{ULW-\delta} = \mathbf{F}\left(\mathbf{Q}_{i+\frac{1}{2}}^{ULW-\delta}\right) , \tag{3.114}$$

$$\mathbf{Q}_{i+\frac{1}{2}}^{ULW-\delta} = \frac{1}{\left(\delta_{i+\frac{1}{2}}^{-} + \delta_{i+\frac{1}{2}}^{+} + \varepsilon_c\right)}\left\{\left(\mathbf{Q}_i^n\delta_{i+\frac{1}{2}}^{-} + \mathbf{Q}_{i+1}^n\delta_{i+\frac{1}{2}}^{+}\right) - \frac{\alpha\Delta t}{\Delta x}\left(\mathbf{F}\left(\mathbf{Q}_{i+1}^n\right) - \mathbf{F}\left(\mathbf{Q}_i^n\right)\right)\right\} \tag{3.115}$$

and

$$\mathbf{F}_{i+\frac{1}{2}}^{ULF-\delta} = \frac{1}{\left(\delta_{i+\frac{1}{2}}^{-} + \delta_{i+\frac{1}{2}}^{+} + \varepsilon_c\right)}\left\{\left(\mathbf{F}\left(\mathbf{Q}_{i+1}^n\right)\delta_{i+\frac{1}{2}}^{-} + \mathbf{F}\left(\mathbf{Q}_i^n\right)\delta_{i+\frac{1}{2}}^{+}\right) - \delta_{i+\frac{1}{2}}^{-}\delta_{i+\frac{1}{2}}^{+}\frac{\Delta x}{\alpha\Delta t}\left(\mathbf{Q}_{i+1}^n - \mathbf{Q}_i^n\right)\right\} , \tag{3.116}$$

where Δx is mesh spacing in the x direction, assumed as constant. In contrast we remark that if the mesh is variably-spaced only the general formulation (3.83)-(3.86) holds. Note that imposing $\delta^{\pm} = 1$ the FORCE-α method on Cartesian meshes (3.18)-(3.21) is recovered.

The optimal upwind bias is given by

$$\delta^-_{i+\frac{1}{2}} = \frac{\left|s_x^{(m)}\right| \alpha \Delta t}{\Delta x} \qquad , \qquad \delta^+_{i+\frac{1}{2}} = \frac{\left|s_x^{(1)}\right| \alpha \Delta t}{\Delta x} \ , \tag{3.117}$$

where $s_x^{(1)}$, and $s_x^{(m)}$ are the extrema of the Riemann fan in the current x direction, which can be approximated using both of the strategies already presented for the general case. In particular, seeking for a two-wave method, we consider the eigenvalues of the Jacobian matrix $\mathbf{A}_1$ defined in (3.105). Across the current interface $S_{i+\frac{1}{2}}$ we have two Jacobians

$$\mathbf{A}_i = \mathbf{A}_1\left(\mathbf{Q}_i^n\right) \qquad , \qquad \mathbf{A}_{i+1} = \mathbf{A}_1\left(\mathbf{Q}_{i+1}^n\right) \ , \tag{3.118}$$

giving rise to two sets of eigenvalues (sorted in increasing order) $\lambda_i^{(1)}, ..., \lambda_i^{(m)}$ and $\lambda_{i+1}^{(1)}, ..., \lambda_{i+1}^{(m)}$. We use these eigenvalues as wave speed estimates in the form

$$\left|s_x^{(1)}\right| = \max\left(\left|\lambda_i^{(1)}\right|, \left|\lambda_{i+1}^{(1)}\right|\right) \qquad , \qquad \left|s_x^{(m)}\right| = \max\left(\left|\lambda_i^{(m)}\right|, \left|\lambda_{i+1}^{(m)}\right|\right) \ . \tag{3.119}$$

3.3.6 Comparison of UFORCE and UFORCE-δ

As a concluding remark, here we highlight the similarities and differences between the UFORCE and the UFORCE-δ method. UFORCE represents our first attempt to develop an upwind-bias method, in the simple case of Cartesian meshes. The simplicity of the Cartesian setup has allowed us to establish the theoretical framework presented in section 3.2.2, on which the more refined and general UFORCE-δ method for unstructured meshes is based.

Both methods are of the upwind-biased type, derived from FORCE. The upwind bias is inserted into the schemes by varying the secondary mesh setup of the classical FORCE method. The size of subvolumes composing the secondary mesh is determined by the position of moving points, varying in space and time. The upwind bias parameters ($(\beta_x)_{i,j}, (\beta_y)_{i,j}$ in the case of UFORCE and δ_j^-, δ_j^+ in the case of UFORCE-δ) determine the secondary mesh setup and tune the amount of numerical dissipation in the resulting method. These parameters are determined so as to reduce numerical dissipation, still retaining monotonicity. This result is achieved in both cases by equating the parametric fluxes to the Godunov upwind flux in linear applications. Thus, our upwind-biased FORCE methods in linear applications are identical to the Godunov-upwind method. Extension of the optimal upwind bias to non-linear systems is achieved for both methods using a centred approach, i.e. considering the system eigenvalues as wave speed estimates, without application of any Riemann solver.

The secondary mesh setup in UFORCE and UFORCE-δ is different. While in UFORCE the subvolumes composing the secondary mesh join in one single point per cell $\vec{P}$, in UFORCE-δ each subvolume V_j^-, associated to each edge in the primary cells, is independently determined using one point $\vec{P}_j$ and is by construction smaller than its counterpart in FORCE. Thus, while in UFORCE the secondary mesh subvolumes entirely cover the computational domain, in UFORCE-δ they do not. In UFORCE, the geometrical constraint relating subvolumes within the primary cell makes the problem of determining the optimal upwind bias to be directionally split and very easily solved on Cartesian meshes. In contrast, in UFORCE-δ the problem of determining the optimal upwind bias is edge-based (no geometrical constraints link the size of different subvolumes within the same primary cell) and this makes the formulation suitable for applications on unstructured meshes.
Finally, in UFORCE one upwind bias parameter in each Cartesian direction per cell is considered, while in UFORCE-δ two upwind bias parameters per edge are used. This feature allows us to consider two waves travelling normal to the current edge, i.e. extrema of the Riemann fan, in the flux computation in UFORCE-δ, while in UFORCE only one wave speed per Cartesian direction (the maximum directional eigenvalue in absolute value) is used in the evaluation of the upwind bias. Therefore, we have used different linear problems to determine the optimal upwind bias in the linear case. Concerning UFORCE, we have considered the scalar linear advection equation, describing the propagation of one wave, while for UFORCE-δ we have used a linear system having two equations and two unknowns, providing two different characteristic speeds.

3.4 Second order extension

First-order monotone fluxes such as UFORCE and UFORCE-δ can be used as a basic building block for high-order extension. The key ingredients for extending a first-order flux to order of accuracy $p > 1$ in the Finite Volume framework are the availability of a non-oscillatory polynomial reconstruction of degree $p-1$ and of a temporal evolution technique.
Concerning the reconstruction technique, Total Variation Diminishing (TVD) schemes are most frequently used for second-order accuracy (see e.g. [118, 130, 89, 17]), while essentially non-oscillatory (ENO) schemes (see e.g. [26, 56]) or weighted essentially non-oscillatory (WENO) schemes (see e.g. [67, 82, 48, 49]) are used for accuracy higher than two.
For the temporal evolution of reconstructed polynomials, several techniques are available. The one-step MUSCL-Hancock approach [131, 118, 17] is conveniently applied for achieving second-order accuracy. The method-of-lines based on Runge-Kutta time stepping (see e.g. [108]) allows achievement of higher order accuracy, but is hardly extended beyond fourth order accuracy due to the so called Butcher barriers [18], which cause the number of intermediate Runge-Kutta stages to become larger than the formal order of accuracy. In contrast, the one-step ADER technique (see

e.g. [121, 123]), based on space and time Taylor expansions and on the semi-analytical Cauchy-Kowalewski procedure for replacing time with space derivatives, has no theoretical accuracy barrier, even though the algebraic manipulations contained in the Cauchy-Kowalewski procedure may be very cumbersome for high-order accuracy. Finally, a new class of one-step high-order methods has been developed based on a fully numeric alternative to the Cauchy-Kowalewski procedure, in which local space-time finite element methods are used in order to compute the polynomial [47].

In this section we apply standard techniques in the finite volume framework to extend the UFORCE and UFORCE-δ fluxes to second-order accuracy. For the UFORCE method, we use the MUSCL-Hancock procedure associated to the ENO reconstruction on Cartesian meshes. Concerning the UFORCE-δ method, we apply the ADER technique in conjunction with the WENO reconstruction on unstructured triangular meshes. The review is specialised for the two-dimensional applications which will be presented in section 3.5.

3.4.1 The MUSCL-Hancock approach with ENO reconstruction on Cartesian meshes

Here we extend the first order UFORCE method to second order in space and time using the MUSCL-Hancock approach [131] with ENO reconstruction [56] for the two-dimensional case.

- *ENO reconstruction.*
  The cell averages $\mathbf{Q}^n_{i,j}$ are reconstructed independently in the x and y direction by selecting limited slope (difference) vectors $\bar{\Delta}_i$ and $\bar{\Delta}_j$ respectively. Limited slope vectors are obtained using the ENO technique in the form

$$\bar{\Delta}_i = \text{minmax}\left(\left(\mathbf{Q}^n_{i+1,j} - \mathbf{Q}^n_{i,j}\right), \left(\mathbf{Q}^n_{i,j} - \mathbf{Q}^n_{i-1,j}\right)\right) , \tag{3.120}$$

  where the minmax function

$$\text{minmax}\,(a,b) = \begin{cases} a & \text{if } |a| \leq |b| \\ b & \text{otherwise} \end{cases} \tag{3.121}$$

  is applied component-wise to the left and right difference vectors. Boundary-extrapolated values of the reconstructed variables are given by

$$\mathbf{Q}^{-x}_{i,j} = \mathbf{Q}^n_{i,j} - \frac{1}{2}\bar{\Delta}_i \qquad , \qquad \mathbf{Q}^{+x}_{i,j} = \mathbf{Q}^n_{i,j} + \frac{1}{2}\bar{\Delta}_i \, , \tag{3.122}$$

$$\mathbf{Q}^{-y}_{i,j} = \mathbf{Q}^n_{i,j} - \frac{1}{2}\bar{\Delta}_j \qquad , \qquad \mathbf{Q}^{+y}_{i,j} = \mathbf{Q}^n_{i,j} + \frac{1}{2}\bar{\Delta}_j \, . \tag{3.123}$$

- *Evolution of these states by* $\frac{1}{2}\Delta t$.
  Evolution of the extrapolated values is achieved in the MUSCL-Hancock framework as follows:

$$\bar{\mathbf{Q}}^{l}_{i,j} = \mathbf{Q}^{l}_{i,j} + \frac{\Delta t}{2\Delta x}\left\{\mathbf{F}\left(\mathbf{Q}^{-x}_{i,j}\right) - \mathbf{F}\left(\mathbf{Q}^{+x}_{i,j}\right)\right\} + \frac{\Delta t}{2\Delta y}\left\{\mathbf{G}\left(\mathbf{Q}^{-y}_{i,j}\right) - \mathbf{G}\left(\mathbf{Q}^{+y}_{i,j}\right)\right\}, \tag{3.124}$$

  for $l = -x, +x, -y, +y$. Notice that the evolution stage is cell-based, i.e. the extrapolated values are evolved within the cells without any contribution from the cell boundaries.

- *Computation of the UFORCE flux.*
  This is achieved by application of the UFORCE scheme (3.44) to the boundary-extrapolated evolved values (3.124). The resulting flux is

$$\mathbf{F}^{UFORCE-2}_{i+\frac{1}{2},j} = \mathbf{F}^{ULW-2}_{i+\frac{1}{2},j} + \mathbf{F}^{ULF-2}_{i+\frac{1}{2},j}, \tag{3.125}$$

  with

$$\mathbf{F}^{ULW-2}_{i+\frac{1}{2},j} = \mathbf{F}\left(\mathbf{Q}^{ULW-2}_{i+\frac{1}{2},j}\right), \tag{3.126}$$

$$\mathbf{Q}^{ULW-2}_{i+\frac{1}{2},j} = \frac{1}{2\left(1-(\beta_x)_{i+1,j}+(\beta_x)_{i,j}\right)} \begin{array}{l} \left\{\left(1+2\,(\beta_x)_{i,j}\right)\bar{\mathbf{Q}}^{-x}_{i+1,j} + \left(1-2\,(\beta_x)_{i+1,j}\right)\bar{\mathbf{Q}}^{+x}_{i,j}\right. \\ \left. -\left(\frac{2\Delta t}{\Delta x}\right)\left(\mathbf{F}\left(\bar{\mathbf{Q}}^{-x}_{i+1,j}\right) - \mathbf{F}\left(\bar{\mathbf{Q}}^{+x}_{i,j}\right)\right)\right\} \end{array} \tag{3.127}$$

  and

$$\begin{aligned} \mathbf{F}^{ULF-2}_{i+\frac{1}{2},j} = \; & \frac{1}{2\left(1-(\beta_x)_{i+1,j}+(\beta_x)_{i,j}\right)}\left\{\left(1+2\,(\beta_x)_{i,j}\right)\mathbf{F}\left(\bar{\mathbf{Q}}^{-x}_{i+1,j}\right)\right. \\ & + \left(1-2\,(\beta_x)_{i+1,j}\right)\mathbf{F}\left(\bar{\mathbf{Q}}^{+x}_{i,j}\right) \\ & \left. -\left(\frac{\Delta x}{2\Delta t}\right)\left(1+2\,(\beta_x)_{i,j}\right)\left(1-2\,(\beta_x)_{i+1,j}\right)\left(\bar{\mathbf{Q}}^{-x}_{i+1,j} - \bar{\mathbf{Q}}^{+x}_{i,j}\right)\right\}. \end{aligned} \tag{3.128}$$

For empirical assessment of second-order accuracy of the present implementation we refer the reader to [112].

3.4.2 The ADER approach with WENO reconstruction on triangular meshes

In this section we review the one-step ADER-WENO technique used for extending the UFORCE-δ flux to second-order accuracy in space and time on two-dimensional triangular meshes.

3.4.2.1 Non-linear reconstruction technique

Here we present a non-linear weighted essentially non-oscillatory (WENO) reconstruction procedure to reconstruct polynomial data within each spatial cell at time t^n from the given cell

averages. The reconstruction procedure, which is non-linear and depends strongly on the input data, follows directly from the guidelines given in [48, 49] for general unstructured two- and three-dimensional meshes. We point out that, even though UFORCE-δ flux has been natively developed to be applied to any kind of unstructured mesh, the proposed reconstruction method applies only to triangular or tetrahedral meshes in two or three space dimensions respectively [48]. This restricts our two-dimensional applications to triangular meshes. The procedure reconstructs entire polynomials, as the original ENO approach [56]. However, we formally write the method like a WENO scheme [67, 82] with a particularly simple choice for the linear weights. For each stencil $\mathcal{S}_i^s = \bigcup T_k$, having defined with $\mathbf{w}_i^s$ the reconstruction polynomial, we require integral conservation:

$$\frac{1}{|T_k|}\int_{T_k} \mathbf{w}_i^s\left(\vec{x},t^n\right)\mathrm{d}\vec{x} = \mathbf{Q}_k^n \qquad \forall T_k \in \mathcal{S}_i^s \,. \tag{3.129}$$

The reconstruction equations (3.129) are solved using a constrained least squares method in order to guarantee that (3.129) is exactly satisfied at least inside element T_i. This procedure is performed in a transformed coordinate space $\vec{\xi} \equiv (\xi,\eta)$ in order to avoid ill-conditioning due to scaling effects. In practice cell T_i of vertices (X_1,Y_1), (X_2,Y_2), (X_3,Y_3) is mapped into a reference triangle with vertices in $(0,0)$, $(1,0)$, $(0,1)$ by applying the transformation:

$$x = X_1 + (X_2 - X_1)\,\xi + (X_3 - X_1)\,\eta \qquad , \qquad y = Y_1 + (Y_2 - Y_1)\,\xi + (Y_3 - Y_1)\,\eta \,, \tag{3.130}$$

whose Jacobian matrix is defined as

$$\vec{J} = \left[\frac{\partial \vec{x}}{\partial \vec{\xi}}\right] \,. \tag{3.131}$$

Therefore polynomials $\mathbf{w}_i^s$ are naturally expressed in terms of transformed coordinates as $\mathbf{w}_i^s\left(\vec{\xi},t^n\right)$. The WENO reconstruction polynomial is obtained by a weighted combination of the polynomials $\mathbf{w}_i^s\left(\vec{\xi},t^n\right)$ in the form

$$\mathbf{w}_i\left(\vec{\xi},t^n\right) = \sum_{s=1}^{7} \omega_s \mathbf{w}_i^s\left(\vec{\xi},t^n\right) \,. \tag{3.132}$$

Seven stencils are required to be used in (3.132): one centred stencil ($s=1$), three forward sector stencils ($s=2,3,4$) and three reverse sector stencils ($s=5,6,7$) [48]. The non-linear WENO weights ω_s are computed as follows:

$$\omega_s = \frac{\widetilde{\omega}_s}{\sum_{k=1}^{7}\widetilde{\omega}_k} \,, \qquad \widetilde{\omega}_s = \frac{\lambda_s}{\left(\sigma_s + \varepsilon_W\right)^r} \,, \qquad \lambda_s = \begin{cases} 10^2 \div 10^5 & \text{if } s = 1, \\ 1 & \text{otherwise} \end{cases} \tag{3.133}$$

with the oscillation indicators σ_s defined in [48], $r=4$ and $\varepsilon_W = 10^{-5}$.

3.4.2.2 Second-order accurate one-step time discretisation

The result of the reconstruction procedure is a non-oscillatory spatial polynomial $\mathbf{w}_i\left(\vec{\xi},t^n\right)$ defined at time t^n inside each spatial element T_i. Next, we need to compute the temporal evolution of these polynomials inside each space-time element $T_i \times \left[t^n, t^{n+1}\right]$ in order to be able to construct the final second-order accurate one-step finite volume scheme.
We consider a two-dimensional system of PDEs having the form

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_y \mathbf{G}(\mathbf{Q}) = \mathbf{0}\,. \tag{3.134}$$

Second-order accuracy is obtained using the ADER approach [123]. The key idea is to solve high-order Riemann problems at the element boundaries. This is achieved by use of a Taylor series expansion in time, of the Cauchy-Kowalewski procedure and of solutions of classical Riemann problems for the state variables and their spatial derivatives. Here, consistently with the reconstruction polynomial procedure, we apply the ADER technique in the transformed coordinate system $\vec{\xi} \equiv (\xi,\eta)$. We adopt the following strategy: we expand the local solution $\mathbf{Q}_i\left(\vec{\xi},t\right)$ of the PDE in each cell in a space-time Taylor series with respect to the element barycentre $(\xi_i,\eta_i) = \left(\frac{1}{3},\frac{1}{3}\right)$

$$\mathbf{Q}_i(\xi,\eta,t) = \mathbf{Q}(\xi_i,\eta_i,t^n) + (\xi-\xi_i)\,\partial_\xi \mathbf{Q} + (\eta-\eta_i)\,\partial_\eta \mathbf{Q} + (t-t^n)\,\partial_t \mathbf{Q} + \mathbb{O}\left(\xi^2,\eta^2,t^2\right)\,. \tag{3.135}$$

Then we use the Cauchy-Kowalewski procedure in order to substitute time derivatives with space derivatives in (3.135). To this aim we rewrite the governing PDE system (3.134) in the transformed coordinate space

$$\partial_t \mathbf{Q} + \partial_\xi \mathbf{F}'(\mathbf{Q}) + \partial_\eta \mathbf{G}'(\mathbf{Q}) = \mathbf{0}\,, \tag{3.136}$$

where $\mathbf{F}'(\mathbf{Q})$ and $\mathbf{G}'(\mathbf{Q})$ are given by

$$\mathbf{F}'(\mathbf{Q}) = \mathbf{F}(\mathbf{Q})\,\partial_x\xi + \mathbf{G}(\mathbf{Q})\,\partial_y\xi \qquad , \qquad \mathbf{G}'(\mathbf{Q}) = \mathbf{F}(\mathbf{Q})\,\partial_x\eta + \mathbf{G}(\mathbf{Q})\,\partial_y\eta\,, \tag{3.137}$$

being $\partial_x\xi$,...,$\partial_y\eta$ the (constant) entries of the inverse of the transformation Jacobian (3.131). For second-order accuracy it suffices to obtain the first time derivative by differentiating (3.136) as follows:

$$\partial_t \mathbf{Q} = -\left(\mathbf{A}'_\xi \partial_\xi \mathbf{Q} + \mathbf{A}'_\eta \partial_\eta \mathbf{Q}\right)\,, \tag{3.138}$$

where $\mathbf{A}'_\xi$ and $\mathbf{A}'_\eta$ are the Jacobian matrices of fluxes (3.137) in the transformed space

$$\mathbf{A}'_\xi = \left[\frac{\partial \mathbf{F}'}{\partial \mathbf{Q}}\right] \qquad , \qquad \mathbf{A}'_\eta = \left[\frac{\partial \mathbf{G}'}{\partial \mathbf{Q}}\right]\,. \tag{3.139}$$

The value of $\mathbf{Q}(\xi_i, \eta_i, t^n)$ and its spatial derivatives are obtained from the WENO reconstruction polynomial $\mathbf{w}_i\left(\vec{\xi}, t\right)$.

3.4.2.3 The fully-discrete second-order accurate one-step scheme

Once the WENO reconstruction and the Cauchy-Kowalewski procedure have been performed for each cell, the final high-order accurate one-step scheme can be written as follows:

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{\Delta t}{|T_i|} \sum_{j=1}^{3} \left|S_j\right| \left(\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}} \cdot \vec{n}_j\right) , \tag{3.140}$$

where $\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}$ is given by

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}} = \frac{1}{\left|S_j'\right|} \int_{S_j'} \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \underline{\underline{\mathbf{F}}}^{UFORCE-\delta}_{j+\frac{1}{2}} \left(\mathbf{Q}_i^n\left(\vec{\xi}, t\right), \mathbf{Q}_j^n\left(\vec{\xi}, t\right)\right) \mathrm{d}t \, \mathrm{d}\vec{\xi} , \tag{3.141}$$

where S_j' is the counterpart of edge S_j in the transformed coordinate system, $\left|S_j'\right|$ represents its length, $\underline{\underline{\mathbf{F}}}^{UFORCE-\delta}_{j+\frac{1}{2}}$ is the UFORCE-δ flux given by (3.83)-(3.86) and $\mathbf{Q}_i\left(\vec{\xi}, t\right)$ and $\mathbf{Q}_j\left(\vec{\xi}, t\right)$ are space-time polynomials in cells T_i and T_j obtained applying the Cauchy-Kowalewski procedure to the WENO reconstruction polynomials. Space and time integrals in (3.141) can be approximated using Gaussian quadratures. For second-order accuracy we use a very compact and efficient midpoint quadrature rule both in time and space, resulting in one single flux evaluation per edge at each evolution step. Therefore in practice we use

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}} = \underline{\underline{\mathbf{F}}}^{UFORCE-\delta}_{j+\frac{1}{2}} \left(\mathbf{Q}_i^n\left(\vec{\xi}_M, t^{n+\frac{1}{2}}\right), \mathbf{Q}_j^n\left(\vec{\xi}_M, t^{n+\frac{1}{2}}\right)\right) , \tag{3.142}$$

being $t^{n+\frac{1}{2}} = t^n + \frac{1}{2}\Delta t$ and $\vec{\xi}_M$ equal to $\left(\frac{1}{2}, 0\right)$, $\left(\frac{1}{2}, \frac{1}{2}\right)$ and $\left(0, \frac{1}{2}\right)$ for first, second and third edge respectively. For higher order extension, where quadratures may get computationally heavy, this approach can be modified using fully-analytical procedures [48].

3.5 Applications to the shallow water equations

The purpose of this section is to assess the performance of the proposed UFORCE and UFORCE-δ method comparing their numerical results with the numerical solutions of well-established centred and upwind finite volume methods and exact solutions. A suite of standard tests for the two-dimensional inviscid shallow water equations, for Cartesian and unstructured meshes, is presented. The material covered in this section is published in [112, 113].

We consider the two-dimensional non-linear inviscid shallow water equations augmented by an equation for a passive scalar. The system written in conservative form reads

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_y \mathbf{G}(\mathbf{Q}) = \mathbf{0}\,, \tag{3.143}$$

where the vector of conserved variables $\mathbf{Q}$ and the flux vectors along the x and y direction $\mathbf{F}(\mathbf{Q})$, $\mathbf{G}(\mathbf{Q})$ are given by

$$\mathbf{Q} = \begin{bmatrix} D \\ Du \\ Dv \\ DC \end{bmatrix} \quad , \quad \mathbf{F}(\mathbf{Q}) = \begin{bmatrix} Du \\ Du^2 + \frac{1}{2}gD^2 \\ Duv \\ DuC \end{bmatrix} \quad , \quad \mathbf{G}(\mathbf{Q}) = \begin{bmatrix} Dv \\ Duv \\ Dv^2 + \frac{1}{2}gD^2 \\ DvC \end{bmatrix} . \tag{3.144}$$

Here $u(x,y,t)$ and $v(x,y,t)$ are the x and y components of velocity, $D(x,y,t)$ is water depth, $C(x,y,t)$ is the passive scalar concentration and $g = 9.81\,ms^{-2}$ is the acceleration due to gravity. The Jacobian matrices of fluxes $\mathbf{F}(\mathbf{Q})$ and $\mathbf{G}(\mathbf{Q})$ have three distinct eigenvalues,

$$\begin{aligned} \lambda_x^{(1)} &= u - a \quad , \quad \lambda_x^{(2)} = u \quad , \quad \lambda_x^{(3)} = u + a\,, \\ \lambda_y^{(1)} &= v - a \quad , \quad \lambda_y^{(2)} = v \quad , \quad \lambda_y^{(3)} = v + a\,, \end{aligned} \tag{3.145}$$

where $a = \sqrt{gD}$ is celerity. Given a unit vector $\vec{n}_j$, the projected Jacobian matrix, defined by (3.106), has three distinct real eigenvalues

$$\lambda_{\hat{n}}^{(1)} = u_{\hat{n}} - a \quad , \quad \lambda_{\hat{n}}^{(2)} = u_{\hat{n}} \quad , \quad \lambda_{\hat{n}}^{(3)} = u_{\hat{n}} + a\,, \tag{3.146}$$

where $u_{\hat{n}} = (u,v)\cdot\vec{n}_j$ is projected velocity. The eigenvalue $\lambda_{\hat{n}}^{(2)}$ (and equivalently $\lambda_x^{(2)}$ and $\lambda_y^{(2)}$), defining a linearly degenerate field as we have seen in section 2.3.2 of chapter 2, has multiplicity 2. This is the consequence of the inclusion of the equation for passive-scalar transport, which provides a contact wave associated to $\lambda_{\hat{n}}^{(2)} = u_{\hat{n}}$. The computation of this wave provides additional difficulties to numerical methods, as we shall see in section 3.5.2.

Results have been obtained enforcing the CFL condition (3.110) with $0 < CFL \leq 1$, which for the considered two-dimensional system specialises as

$$\Delta t = \frac{1}{2}\,CFL \min_{i,j}\left(\min\left(\frac{\Delta x}{(|u|+a)_{i,j}} + \frac{\Delta y}{(|v|+a)_{i,j}}\right)\right) \tag{3.147}$$

on Cartesian meshes and as

$$\Delta t = CFL \min_{1\leq i\leq n_e} \left(\min_{1\leq j\leq 3} \left(\frac{h_j^-}{\max\left((|u|+a)_i, (|u|+a)_j \right)} \right) \right)_{T_i}, \tag{3.148}$$

on unstructured meshes.

Four test problems have been chosen in order to assess the behaviour of the UFORCE and UFORCE-δ methods. Two of them, namely the collapse of a circular dam (section 3.5.1) and the propagation of a passive scalar discontinuous wave (section 3.5.2) have been solved using first-order accurate methods on Cartesian meshes, while the advection of a potential vortex (section 3.5.3) has been performed on triangular unstructured meshes using second-order extensions of numerical methods. Furthermore this test, possessing a smooth exact solution, is used for assessing second-order convergence of the ADER-WENO extension of UFORCE-δ. The last test, namely the collapse of a circular dam solved on a variably-spaced grid, is solved using second-order accurate schemes both on Cartesian meshes (section 3.5.4) and unstructured meshes (section 3.5.5).

The results obtained with the proposed UFORCE and UFORCE-δ methods have been compared with those obtained using one centred method (the FORCE method), one centred upwind-biased method (the KNP method [70]) and two purely upwind methods. Among upwind methods we used the Godunov method coupled with an exact Riemann solver (Godunov-exact) and the Godunov method coupled with the HLL approximate Riemann solver (Godunov-HLL).
These numerical methods can be classified considering the number of waves taken into account for the flux evaluation. Thus, we have one zero-wave method (FORCE), one one-wave method (UFORCE), three two-wave methods (HLL, KNP, UFORCE-δ), one complete method (Godunov-exact).

3.5.1 Collapse of a circular dam

This test case consists of the instantaneous breaking of a cylindrical tank initially filled with 2.5 meter deep still water at rest, surrounded by a square, half-meter deep lake at rest. When the water column is released, the shock wave results in a dramatic increase of water depth in the lower depth region, propagating in the radial direction.

We solve (3.143)-(3.144) with initial condition

$$D(x,y,0) = \begin{cases} 2.5\,m & \text{if } x^2+y^2 \le R^2 \\ 0.5\,m & \text{if } x^2+y^2 > R^2 \end{cases} \quad , \qquad u(x,y,0) = v(x,y,0) = 0 \quad \forall x,y\,. \tag{3.149}$$

being $R = 2.5\,m$ the tank radius. With this test we aim to assess the ability of the UFORCE-δ method of accurately reproducing shock and rarefaction waves. Shock waves are discontinuous waves associated with the genuinely non-linear fields $\lambda_x^{(1),(3)} = u \pm a$, $\lambda_y^{(1),(3)} = v \pm a$. These waves require correct speed of propagation, sharp resolution of the transition zone and absence of spurious oscillations around the shock. Rarefaction waves are smooth waves and numerical methods should be able to resolve these features accurately, especially their heads and tails, which contain discontinuities in space derivatives.

Numerical solutions have been obtained using a coarse mesh of 101×101 cells in the square computational domain $[-20,20] \times [-20,20]$ m with transmissive boundary conditions. This comparison has been carried out with the first order version of the above mentioned numerical methods.

We provide an accurate reference solution, which has been obtained by turning the problem (3.143)-(3.144)-(3.149) into a one-dimensional problem in the radial direction [117]:

$$\partial_t \begin{bmatrix} D \\ Du_r \end{bmatrix} + \partial_r \begin{bmatrix} Du_r \\ Du_r^2 + \frac{1}{2} g D^2 \end{bmatrix} = -\frac{1}{r} \begin{bmatrix} Du_r \\ Du_r^2 \end{bmatrix} , \tag{3.150}$$

where r is the radial coordinate and $u_r(r,t)$ the radial velocity. The initial conditions (3.149) in the radial coordinate system read:

$$\begin{cases} D(r,0) = 2.5\,m & \text{if } r \le R\,, \\ D(r,0) = 0.5\,m & \text{if } r > R \end{cases} \quad , \qquad u(r,0) = 0 \quad \forall r\,. \tag{3.151}$$

The reference solution is obtained solving numerically system (3.150), (3.151) on a fine mesh of 1000 cells using the WAF method in conjunction with the HLLC approximate Riemann solver [117]. The CFL number is set to 0.9 and the limiter used is SUPERBEE [101].

Results for water depth $D(x,y,t)$ for this test are displayed in Fig. 3.9 at time $t = 1.4\,s$ and in Fig. 3.10 at time $t = 3.5\,s$. Here, numerical solutions (symbols) are presented in terms of slices along the x axis ($y = 0$) and compared with the reference radial solution (full line). Each figure presents the results obtained setting $CFL = 0.9$ (top panel) and $CFL = 0.2$ (bottom panel). From both Figs. 3.9 and 3.10 the Godunov-exact method, based on exact evaluation of all the three waves,

is found to be the least dissipative among all methods and to have a consistent performance with low and high CFL numbers. In contrast, the accuracy of the FORCE centred method significantly depends on the CFL number. While the solution obtained with FORCE at $CFL = 0.2$ is excessively smoothed and smeared both in the shock and in the rarefaction zones, for larger values of the CFL number the behavioural differences among centred and upwind methods tend to disappear. However even with $CFL = 0.9$ FORCE fails in the accurate description of inflections in the free-surface profile (Fig. 3.10 around $x = 0$).
The proposed one-wave UFORCE method significantly improves accuracy especially at low CFL numbers, but still the solution is slightly more diffused than that of two-wave solvers like HLL and KNP (see Fig. 3.9 at $x = \pm 3$). These two-wave methods give indistinguishable results. The results of the proposed UFORCE-δ method are significantly more accurate than those of KNP and HLL for both values of CFL. The UFORCE-δ solution profile is always bounded between that of Godunov-exact and that of KNP and HLL.
We speculate that improved accuracy of UFORCE compared to FORCE and of UFORCE-δ over all the other methods except Godunov-exact is due to improved resolution of contact waves, i.e. the shear wave in this case. This feature of UFORCE and UFORCE-δ is proved in the next section.

Finally, in Fig. 3.11 it is shown that the proposed UFORCE method, here extended to second-order accuracy using the MUSCL-Hancock ENO technique, converges to the correct solution as the grid is refined.

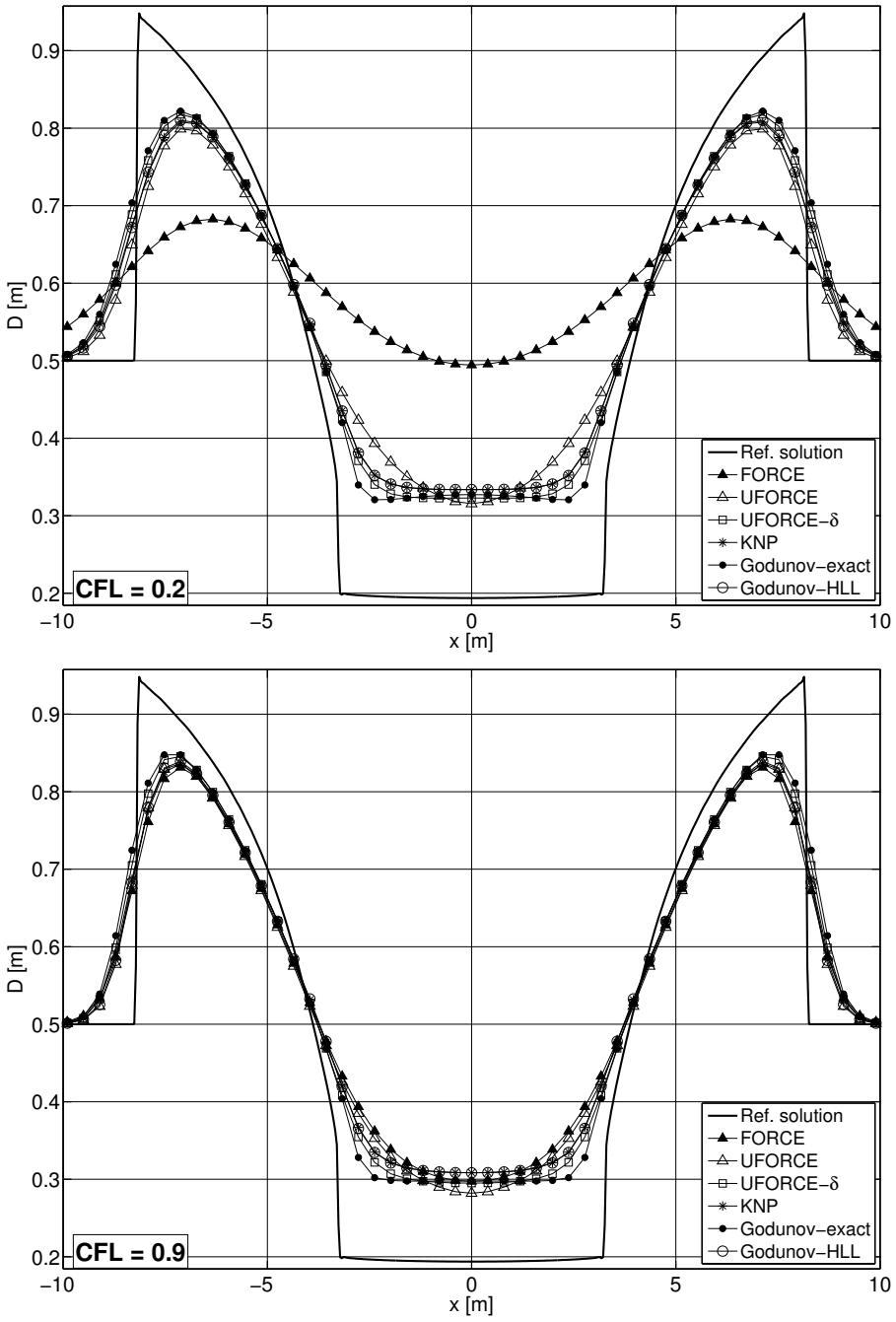

Figure 3.9: Collapse of a circular dam. The numerical results for water depth D of six numerical methods (symbols) are compared with the reference radial solution (full line) at time $t = 1.4\,s$. The numerical solution profiles are sliced on the x axis. The mesh used is 101×101 cells and CFL is set to 0.2 (top profile) and 0.9 (bottom profile).

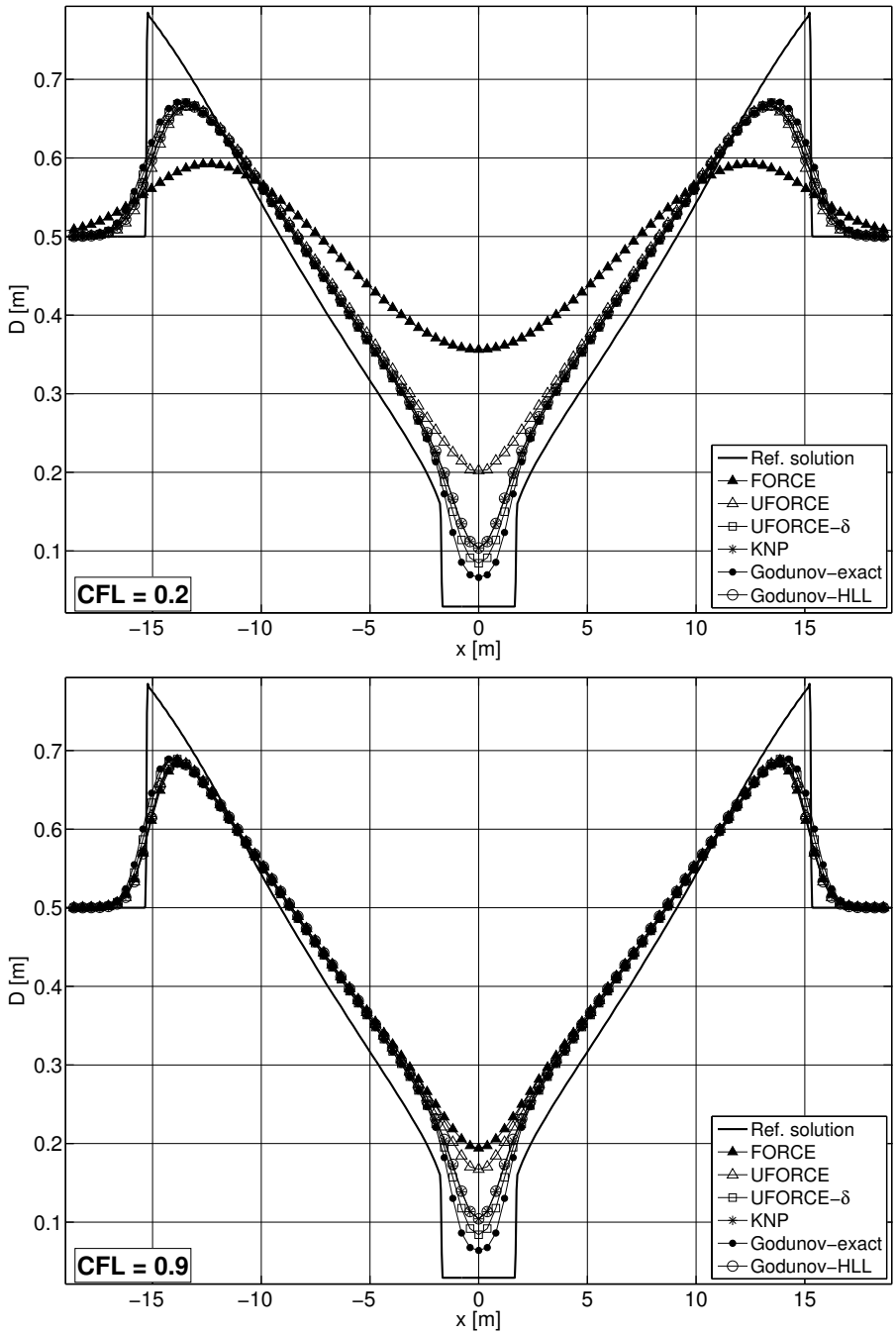

Figure 3.10: Collapse of a circular dam. The numerical results for water depth D of six numerical methods (symbols) are compared with the reference radial solution (full line) at time $t = 3.5\,s$. The numerical solution profiles are sliced on the x axis. The mesh used is 101×101 cells and CFL is set to 0.2 (top profile) and 0.9 (bottom profile).

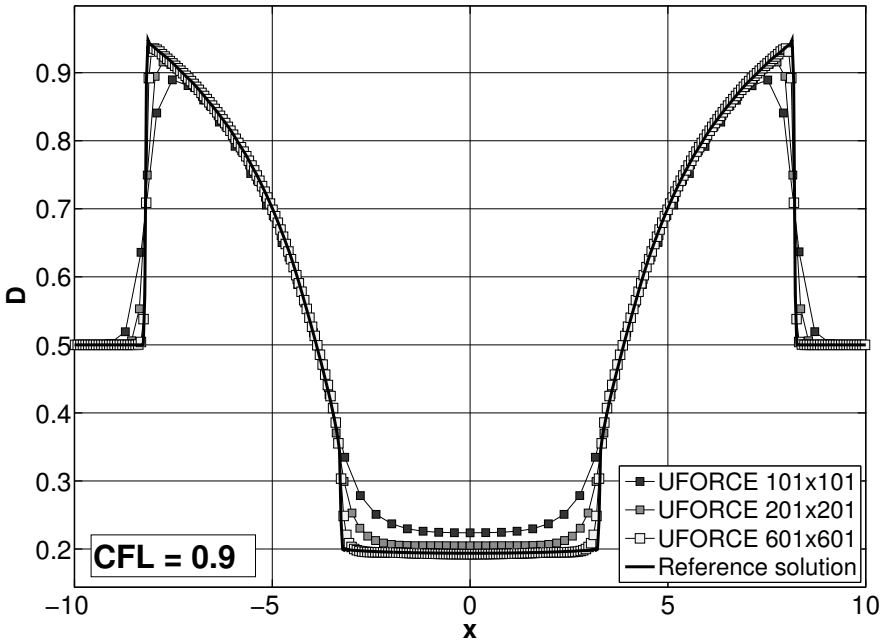

Figure 3.11: Collapse of a circular dam. The numerical results for water depth D of the second-order MUSCL-Hancock ENO UFORCE scheme (symbols) obtained at different grid resolution (101×101, 201×201, 601×601 cells) are compared with the reference radial solution (full line) at time $t = 1.4\,s$. The numerical solution profiles are sliced on the x axis. CFL is set to 0.9.

3.5.2 Propagation of a passive scalar discontinuity

The aim of this test is to assess the accuracy of the UFORCE-δ method when dealing with contact waves. The flow field results from the collapse of a dam initially placed at $x = 0$. Across the wall the water depth D initially exhibits a discontinuity, being $1\,m$ on the left side of the domain and $0.5\,m$ on the other side. Also the concentration field C is discontinuous across the dam, while water is initially at rest all over the domain. The dam removal causes the propagation of a rarefaction wave orthogonally to the dam towards the left side of the domain and of a shock wave on the other side. This shock wave travels faster than the water particles. An intermediate wave for the concentration discontinuity, passively transported at a speed equal to water velocity, is also produced.

We solve (3.143)-(3.144) with initial condition

$$\begin{cases} \begin{cases} D(x,y,0) = 1\,m \\ C(x,y,0) = 1 \end{cases} & \text{if } x \leq 0 \\ u(x,y,0) = v(x,y,0) = 0 & \forall\, x,y\,. \end{cases} \quad , \quad \begin{cases} D(x,y,0) = 0.5\,m \\ C(x,y,0) = 0 \end{cases} \quad \text{if } x > 0 \tag{3.152}$$

We use a Cartesian mesh of 100×100 cells in the square computational domain $[-25,25] \times [-25,25]$ with transmissive boundary conditions. The solution is computed at time $t = 5\,s$.

The exact solution for this problem is obtained by solving a one-dimensional dam-break problem in the x direction using an exact Riemann solver. It contains a left rarefaction, a right-facing shock wave and a contact discontinuity in the middle, across which the concentration $C(x,y,t)$ varies discontinuously (see [122] for an accurate description). We focus our attention to the contact discontinuity and discuss the results in terms of $C(x,y,t)$. In general, the computation of contact waves, associated with the linearly degenerate fields ($\lambda_x^{(2)} = u$, $\lambda_y^{(2)} = v$) is very challenging. One main difficulty is to preserve sharpness in the resolution of these waves in time evolution problems. Upwind methods are distinctly better than centred methods on this task; however, only the upwind methods based on complete Riemann solvers (in our case, Godunov-exact) explicitly include the contact wave in the flux computation. In contrast, schemes based on the HLL Riemann solver behave like centred methods for linear fields [118]. Similarly, refined centred schemes like UFORCE, UFORCE-δ and KNP do not include any upwind bias related to linear fields.

We compare the results of the first-order version of the proposed method with the first-order version of the same six numerical methods used in the previous section. The results for this test, obtained with $CFL = 0.2$ and $CFL = 0.9$, are displayed in Fig. 3.12 (top and bottom panel respectively). The solution for variable C is represented in terms of slices along the x axis. As expected, for both values of the CFL number, the Godunov-exact method gives rise to the sharpest resolution of the contact wave, outperforming all the other methods, while HLL and KNP are found to perform in analogous manner. These two-wave methods perform consistently over the entire range of CFL numbers considered, but the solution profile is always found to be more smeared than that of Godunov-exact. In contrast, the behaviour of the genuinely centred FORCE method is deeply influenced by the CFL number: at $CFL = 0.2$ the solution is affected by excessive dissipation, while at $CFL = 0.9$ the solution is found to be slightly more accurate than that of KNP and HLL. The one-wave UFORCE method represents a significant improvement with respect to the FORCE method for low values of the CFL parameter, while the improvement in accuracy over FORCE vanishes at high CFL numbers. Finally, let us focus on the proposed UFORCE-δ method. From Fig. 3.12 it is clear that the solution obtained using UFORCE-δ is affected by small values of

numerical dissipation and outperforms the results obtained using all the other methods except the Godunov-exact method, as it was expected.

3.5.3 Vortex advection: convergence test

In the present test, originally proposed by Ricchiuto and Bollermann [99], we examine the case of a vortex travelling at mean velocity $\vec{u}_\infty = (u_\infty, v_\infty)$ while maintaining its properties (water surface level and velocity field). Since an exact solution of this problem is available, we use this test case for assessing the second-order convergence of our ADER-WENO UFORCE-δ method.

In order to derive the exact solution we apply the following decomposition to flow field

$$\vec{u} = \vec{u}' + \vec{u}_\infty \,, \tag{3.153}$$

where $\vec{u}'$ in cylindrical coordinates reads:

$$\vec{u}' = \left(u'_r, u'_\theta\right) = \left(0, u'_\theta\right) \tag{3.154}$$

being u'_r and u'_θ the relative radial and tangential velocity respectively. Then, the first equation in (3.143)-(3.144) becomes

$$\partial_t D + \vec{u}_\infty \cdot \nabla D = 0 \,, \tag{3.155}$$

which admits the following exact solution:

$$D\left(x,y,t\right) = D_0\left(\gamma\right) \,, \tag{3.156}$$

being $\gamma = (x - u_\infty t, y - v_\infty t)$ and $D_0\left(x,y\right)$ the initial condition for water depth distribution. Substituting (3.153)-(3.154) into the second and third equation in (3.143)-(3.144) we obtain

$$\partial_t \vec{u}' + \left(\vec{u}_\infty \cdot \nabla\right) \vec{u}' + \left(\vec{u}' \cdot \nabla\right) \vec{u}' + g \nabla D_0\left(\gamma\right) = \vec{0} \,, \tag{3.157}$$

which admits an exact solution of the form

$$\vec{u}'\left(x,y,t\right) = \vec{u}'_0\left(\gamma\right) \,, \tag{3.158}$$

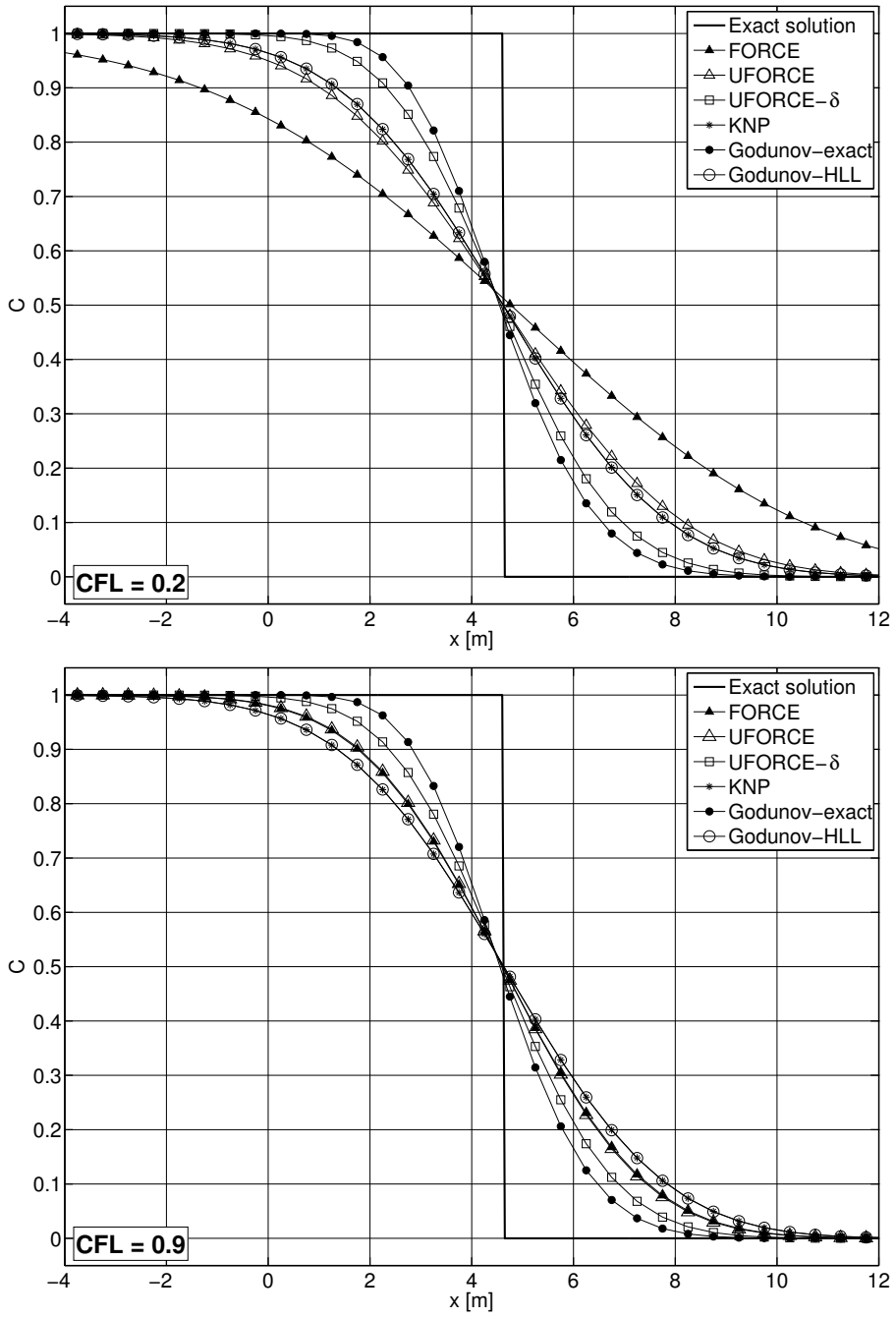

Figure 3.12: Propagation of a passive scalar discontinuity. The numerical results for concentration C of six numerical methods (symbols) are compared with the exact solution (full line) at time $t = 5\ s$. The numerical solution profiles are obtained slicing the solution along the x axis. The mesh used is 100×100 cells and CFL is set to 0.2 (top panel) and 0.9 (bottom panel).

by which initial conditions are advected over the spatial domain, as in the case of the linear advection equation. Appropriate initial conditions for the velocity field are

$$\vec{u}_0'(r_c) = \begin{cases} \Gamma(1+\cos(\omega r_c))(y_c - y, x - x_c) & \text{if } \omega r_c \leq \pi \\ (0,0) & \text{otherwise} \end{cases}, \tag{3.159}$$

where Γ is vortex intensity, (x_c, y_c) are the coordinates of the vortex centre at initial time, r_c is the distance to the vortex core and ω the angular wave frequency determining the vortex width. Integration of (3.157) along the radial direction yields the initial conditions for the water surface profile

$$D_0(r_c) = D_\infty + \begin{cases} \frac{1}{g}\left(\frac{\Gamma}{\omega}\right)^2 \left(\phi^R(\omega r_c) - \phi^R(\pi)\right) & \text{if } \omega r_c \leq \pi \\ 0 & \text{otherwise} \end{cases}, \tag{3.160}$$

where $\phi^R(a) = 2\cos(a) + 2a\sin(a) + \frac{1}{8}\cos(2a) + \frac{1}{4}\sin(2a) + \frac{3}{4}a^2$ and D_∞ is water depth outside the vortex.

Following Ricchiuto and Bollermann [99], the parameters used in computations are $\Gamma = 15$, $\omega = 4\pi$, $\vec{u}_\infty = (6,0)$, $D_\infty = 5$, $g = 1$. We solve the problem in the rectangular computational domain $[0,1] \times [0,2]$ with weak far field conditions prescribed at the four boundaries. The initial position of vortex centre is $(x_c, y_c) = \left(\frac{1}{2}, \frac{1}{2}\right)$. Having set timeout equal to $t = \frac{1}{6}s$, the vortex centre is expected to be located at $\left(\frac{3}{2}, \frac{1}{2}\right)$ at the end of computations.
We use a sequence of regularly-refined triangular meshes characterised by N (reciprocal of mesh length) equal to 10, 20, 40, 80, 160. The CFL condition (3.148) is applied, using $CFL = 0.9$. We solve the problem using the second-order ADER-WENO method together with the FORCE, UFORCE-δ and HLL flux.

In Tab. 3.2 we present error norms and resulting order of accuracy for the ADER-WENO UFORCE-δ method for variable D. Expected second-order accuracy is achieved in each norm. Moreover, in Tabs. 3.3 and 3.4 we present the error norms and order of accuracy obtained using the ADER-WENO method together with the FORCE and HLL flux respectively. Comparing the norms in Tabs. 3.2 and 3.3 we assess the great improvement of UFORCE-δ over FORCE (the norms of UFORCE-δ are about half of those of FORCE) and comparing the norms in Tabs. 3.2 and 3.4 we observe a slight but significant improvement with respect to HLL.

In Fig. 3.13a we show the convergence of the proposed UFORCE-δ method to the exact solution. Solution profiles are sliced along the x axis at $y = \frac{1}{2}$. We observe that the numerical profile obtained with $N = 160$ is almost indistinguishable from the exact solution. In Fig. 3.13b we

compare the results of UFORCE-δ to those of HLL and FORCE obtained on the same coarse mesh ($N = 20$). The UFORCE-δ method in this condition is seen to be more accurate than FORCE and HLL. In detail, the improvement of UFORCE-δ with respect to the standard FORCE method is very significant. The reason for this can be identified in improved resolution of contact waves, i.e. the shear wave associated to tangential velocity in this case, which is expected to play an important role in determining overall accuracy in the present vortical problem.

Table 3.2: Vortex advection. Convergence rate study for the second-order ADER-WENO UFORCE-δ method for variable D. N denotes the reciprocal of mesh length, $N_0 = 10$.

N/N_0	L_∞	$\mathbb{O}(L_\infty)$	L_1	$\mathbb{O}(L_1)$	L_2	$\mathbb{O}(L_2)$
2	1.101E+00	1.69	3.645E-02	1.51	1.086E-01	1.63
4	2.679E-01	2.04	7.748E-03	2.23	2.143E-02	2.34
8	5.114E-02	2.39	1.747E-03	2.15	4.243E-03	2.34
16	1.355E-02	1.92	4.388E-04	1.99	9.965E-04	2.09

Table 3.3: Vortex advection. Convergence rate study for the second-order ADER-WENO FORCE method for variable D. N denotes the reciprocal of mesh length, $N_0 = 10$.

N/N_0	L_∞	$\mathbb{O}(L_\infty)$	L_1	$\mathbb{O}(L_1)$	L_2	$\mathbb{O}(L_2)$
2	2.669E+00	0.67	6.893E-02	0.84	2.456E-01	0.76
4	5.332E-01	2.32	1.643E-02	2.07	5.275E-02	2.22
8	1.336E-01	2.00	3.484E-03	2.24	1.083E-02	2.28
16	3.021E-02	2.14	7.976E-04	2.13	2.290E-03	2.24

Table 3.4: Vortex advection. Convergence rate study for the second-order ADER-WENO HLL method for variable D. N denotes the reciprocal of mesh length, $N_0 = 10$.

N/N_0	L_∞	$\mathbb{O}(L_\infty)$	L_1	$\mathbb{O}(L_1)$	L_2	$\mathbb{O}(L_2)$
2	1.301E+00	1.51	4.118E-02	1.43	1.292E-01	1.45
4	2.831E-01	2.20	8.399E-03	2.29	2.428E-02	2.41
8	5.644E-02	2.33	1.837E-03	2.19	4.648E-03	2.39
16	1.356E-02	2.06	4.455E-04	2.04	1.033E-03	2.17

3.5.4 Collapse of a circular dam on a Cartesian variably-spaced grid

As it was shown in the previous sections, an attractive feature of the proposed methods relies on their ability to perform consistently throughout the range of stable CFL numbers. This fact has important consequences in practical applications when the shallow water equations are solved in irregular domains, where a wide range of CFL numbers from small to large is generated.

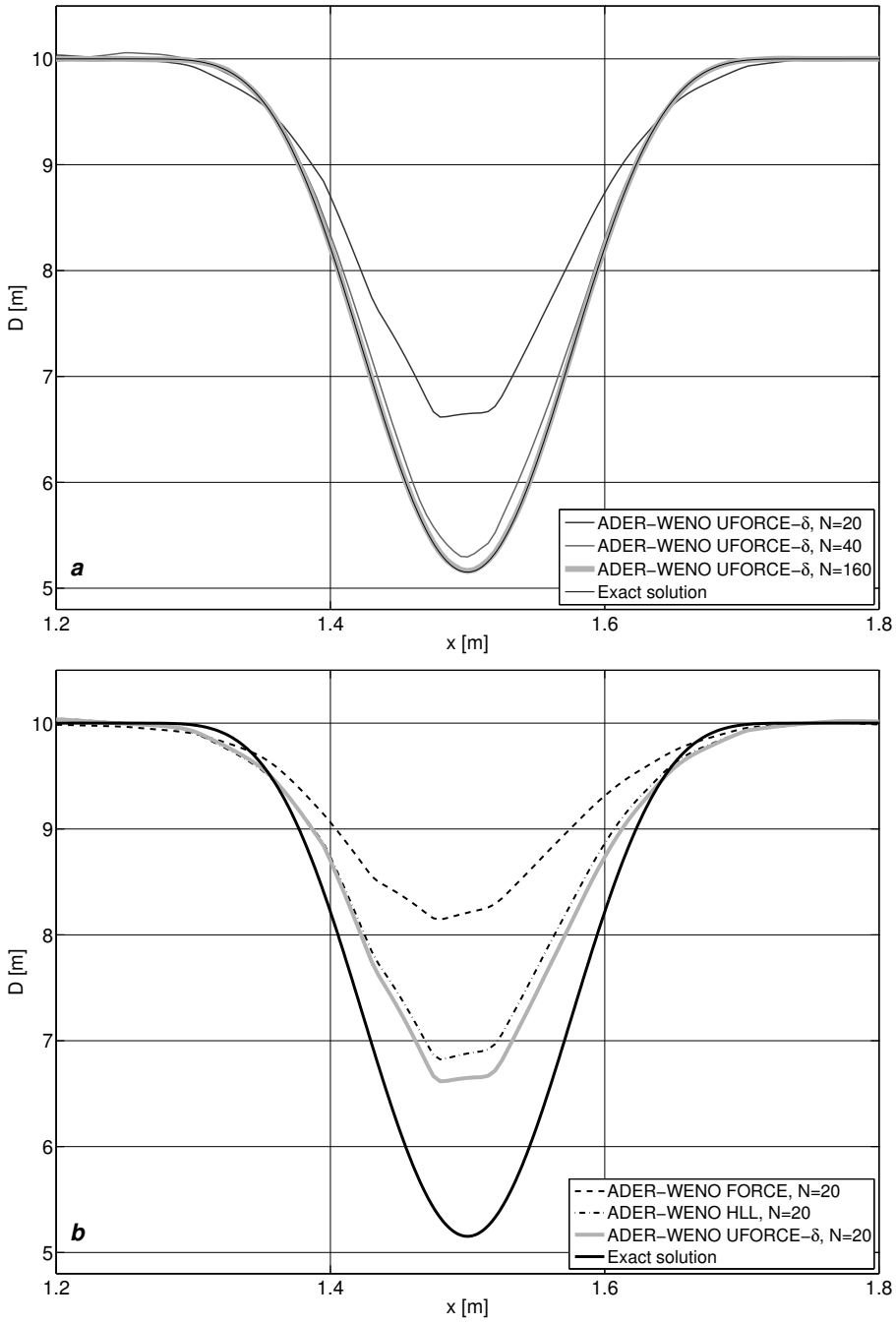

Figure 3.13: Vortex advection. The solution profiles are sliced along the x axis. a) The numerical results for water depth D of the second-order ADER-WENO UFORCE-δ method obtained with different grid resolution (grey lines) are compared with the exact solution (black line) at time $t = 1/6\ s$. N denotes the reciprocal of mesh length. b) Numerical results for water depth D obtained using the second-order ADER-WENO method with the FORCE, UFORCE-δ and HLL numerical fluxes on the same coarse mesh ($N = 20$) are compared with the exact solution (black line) at time $t = 1/6\ s$.

In order to highlight this behaviour of the UFORCE method we solve again the problem defined by (3.143)-(3.144)-(3.149), regarded as the collapse of a circular dam in section 3.5.1. However, here we use a different mesh from that previously adopted. We set the square computational domain $[-40,40]\times[-40,40]$. Along the y direction we adopt a regular grid spacing $\Delta y = 0.398\ m$ corresponding to 201 computational cells. Along the x direction we adopt an irregular grid spacing, i.e. $\Delta x = 0.99$ for $x < 0$, $\Delta x = 0.24$ for $x > 0$ and $\Delta x = 0.62$ for the cell centred in $x = 0$, corresponding to 201 variably-spaced computational cells. The grid for this test is displayed in Fig. 3.14. We impose transmissive boundary conditions. The solution is computed at time $t = 4.7\ s$ using $CFL = 0.9$.

The solution to this problem is expected to exhibit an outer-facing shock, a circular rarefaction following the shock and an inner shock which has been formed by the over-expansion of the flow caused by the reflection of the interior rarefaction from the centre of the dam [117]. The exact reproduction of the complicated wave pattern in the shock reflection would be challenging itself even on a fine regularly spaced grid.
However, here, due to irregular grid spacing we provide an additional difficulty to numerical methods. In fact, being the test problem symmetrical along the y axis ($x = 0$), the CFL condition is enforced where Δx reaches its minimum value, that is, within the fine grid side of the domain. Being the time step Δt common to all the cells in the domain, in the coarse mesh side low local values of the CFL number will be found, causing a poor performance of numerical methods in terms of accuracy. Thus, in this test case, preserving symmetry along the y axis is the challenge.

Results, obtained using second-order MUSCL-Hancock ENO extensions of numerical methods, are displayed in Figs. 3.15a-3.15d in terms of contourplots, while in Figs. 3.16a-3.16b they are given in terms of slices along the x axis ($y = 0$). In Fig. 3.15a we show that the FORCE method does not preserve symmetry in the solution because of excessive smearing at low local CFL in the coarse mesh side of the domain, while the Godunov-exact solution (Fig. 3.15c) gives rise to optimal results preserving symmetry in all the directions. The UFORCE solution (Fig. 3.15b) shows a significant improvement compared that of the FORCE method, possessing the same degree of symmetry as the Godunov-HLL numerical solution (Fig. 3.15d). Similar conclusions can be drawn analysing Figs. 3.16a-3.16b. All methods turn out to solve to the same degree of accuracy in the finer mesh region of the domain, while significant differences can be found in the coarser mesh side. Results obtained with the centred FORCE method (Fig. 3.16a) are affected by severe numerical diffusion on the coarse mesh side of the domain if compared with the solution of Godunov-exact, which outperforms all the other methods. The centred UFORCE method and the upwind Godunov-HLL method give rise to very close results through all the domain (see Fig. 3.16b). Even though the Godunov-HLL method has a slight advantage over the UFORCE method in the solution of the

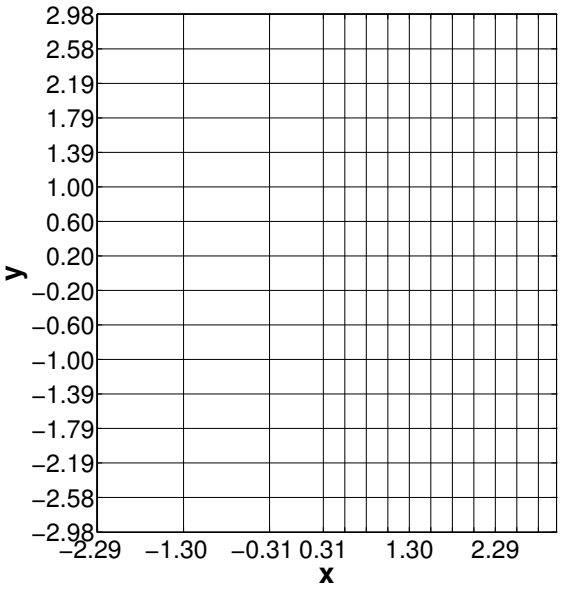

Figure 3.14: Collapse of a circular dam on a Cartesian variably-spaced grid. Sketch of the grid.

shock reflection around $x = 0$, it turns out to be slightly less accurate in the description of the rarefactions (see Fig. 3.16b).

3.5.5 Collapse of a circular dam on an unstructured variably-spaced grid

Aim of this test is to assess the improvement in accuracy of the UFORCE-δ method on unstructured meshes when dealing with a great variety of local CFL numbers. We propose again the test case presented in the Cartesian case in section 3.5.4, namely the collapse of a circular dam defined by equations (3.143)-(3.144)-(3.149) solved on a variably-spaced grid. Here we use an irregular triangular mesh of 34753 cells as depicted in Fig. 3.17, whose length ranges from 2.08 on the left boundary to 0.15 on the right boundary, covering the square computational domain $[-25,25] \times [-25,25]$. We impose transmissive boundary conditions. The solution is computed at time $t = 4.7$ s setting $CFL = 0.9$. For a description of the features of this test problem we refer the reader to the previous section.

The results for water depth D obtained using the second-order ADER-WENO method in conjunction with the FORCE and UFORCE-δ fluxes are presented in Fig. 3.18 in terms of slices along the x axis. The numerical profiles are compared to a refined reference solution obtained solving (3.150)-(3.151) as explained in section 3.5.1. It is seen that the UFORCE-δ method solves the left-facing shock ($x = -18$) with a higher degree of accuracy compared to FORCE method,

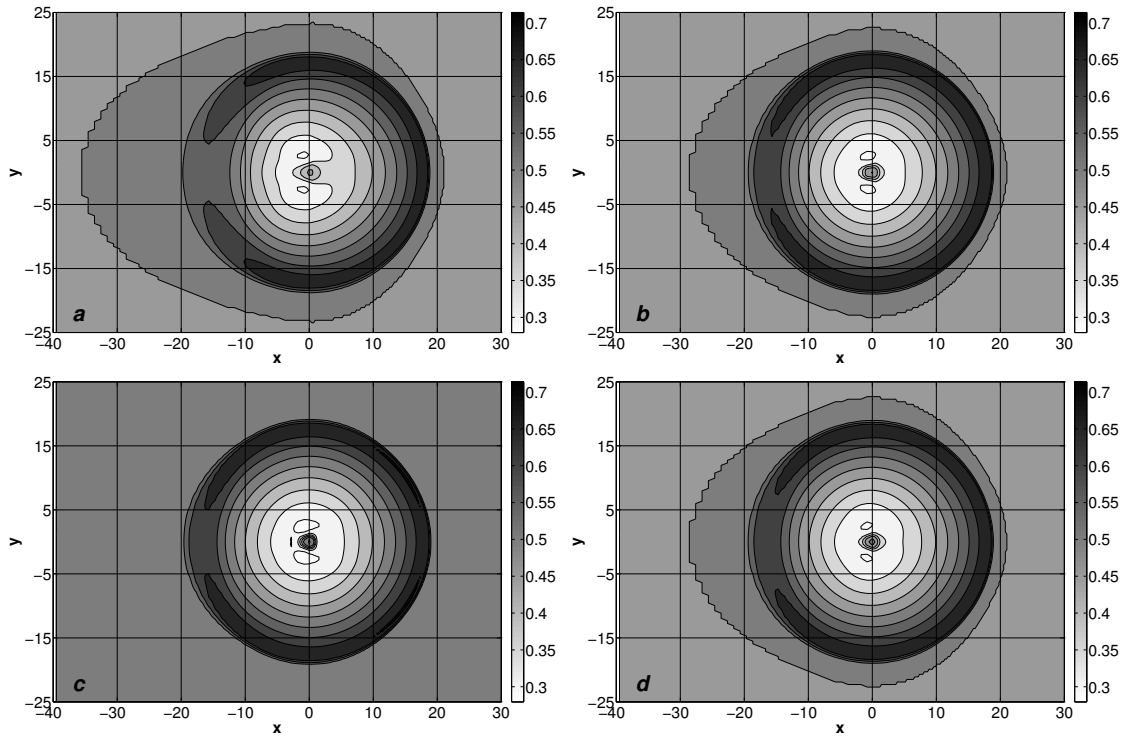

Figure 3.15: Collapse of a circular dam on a Cartesian variably-spaced grid. Numerical results for water depth D for the second-order MUSCL-Hancock ENO extension of the FORCE (a), UFORCE (b), Godunov-exact (c), Godunov-HLL (d) methods are presented at time $t = 4.7\ s$ in terms of contourplots. The mesh used has 201 equally-spaced cells in the y direction and 201 variably-spaced cells in the x direction ($\Delta x = 0.99$ for $x < 0$, $\Delta x = 0.24$ for $x > 0$ and $\Delta x = 0.62$ for the cell centred in $x = 0$). CFL is set to 0.9.

while the right-facing shock ($x = 18$) is solved almost to the same accuracy by both methods. This behaviour of UFORCE-δ gives rise to a more symmetric solution. Moreover, the influence of the upwind-bias is dominant in the rarefaction zones ($x = \pm 3$), and in the central reflected shock ($x = 0$) where UFORCE-δ outperforms FORCE.

The results of the UFORCE-δ and FORCE methods are then presented in Fig. 3.19 in terms of contourplots. Comparing Fig. 3.19a (FORCE) with Fig. 3.19b (UFORCE-δ) the same conclusion as for Fig. 3.18 about resolution of the left-facing shock and overall degree of symmetry can be drawn.

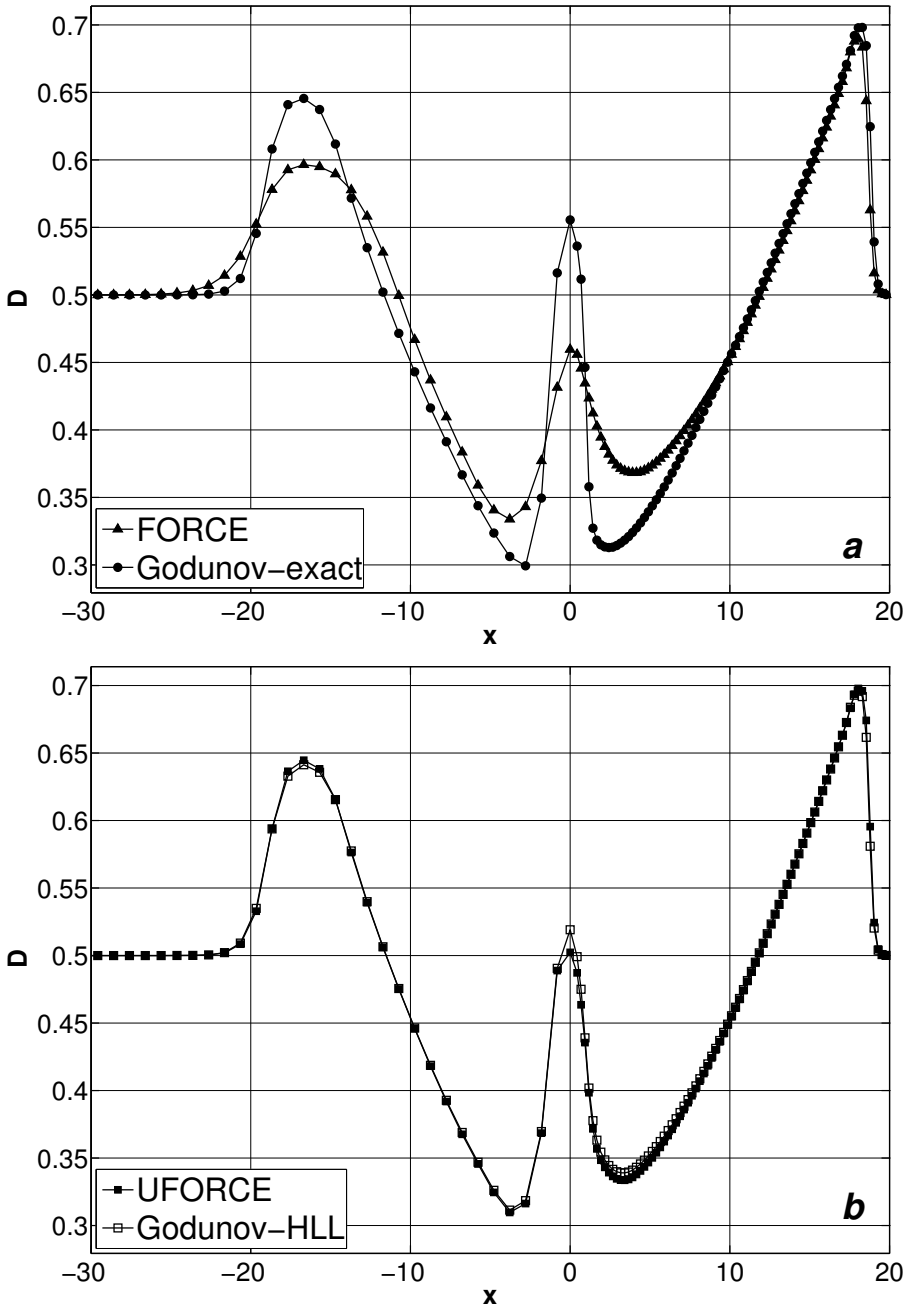

Figure 3.16: Collapse of a circular dam on a Cartesian variably-spaced grid. Numerical results for water depth D for the second-order MUSCL-Hancock ENO extension of the FORCE and Godunov-exact methods (a) and of the UFORCE and Godunov-HLL methods (b) are presented at time $t = 4.7\ s$. The numerical solution profiles are sliced along the x axis. The mesh used has 201 equally-spaced cells in the y direction and 201 variably-spaced cells in the x direction ($\Delta x = 0.99$ for $x < 0$, $\Delta x = 0.24$ for $x > 0$ and $\Delta x = 0.62$ for the cell centred in $x = 0$). CFL is set to 0.9.

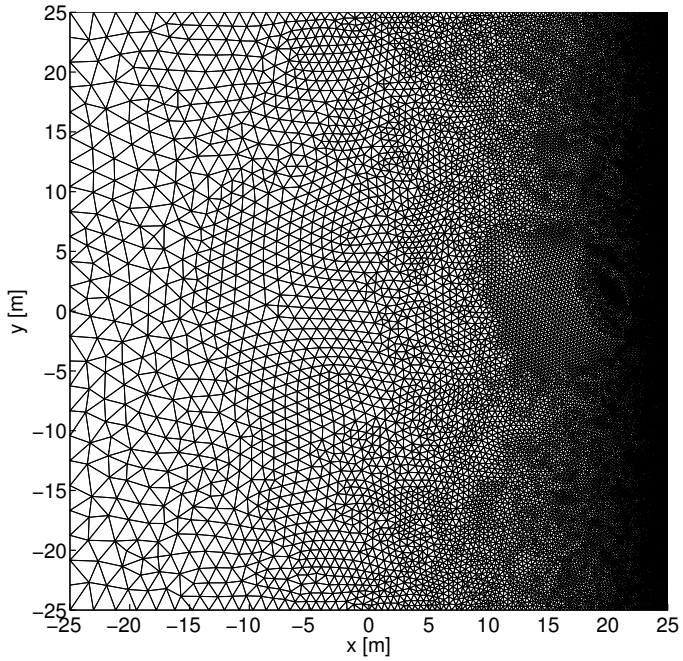

Figure 3.17: Collapse of a circular dam on an unstructured variably-spaced grid: variably-spaced mesh of 34753 triangles. Mesh length ranges from 2.08 on the left boundary to 0.15 on the right boundary.

3.6 Conclusions

Two upwind-biased versions of the multi-dimensional FORCE flux for solving hyperbolic systems of PDEs in conservation-law form have been presented. The proposed UFORCE flux applies to multidimensional Cartesian meshes, while the proposed UFORCE-δ flux applies to general meshes in multi-dimensions. The first-order accurate UFORCE and UFORCE-δ methods are genuinely centred since the use of Riemann solvers either exact or approximate is not required. To be implemented, the methods require only knowledge of the eigenvalues evaluated at current time t^n which are needed in any case for selecting the integration time step. In particular, the one-wave UFORCE method includes only the maximum eigenvalue in absolute value in the flux computation, while the two-wave UFORCE-δ method considers the maximum and minimum eigenvalues of the entire set. We demonstrate that for the linear case the UFORCE and UFORCE-δ are identical to the Godunov upwind method and then we empirically extend the validity of our methods to non-linear hyperbolic systems of PDEs. Extension to second-order accuracy in space and time has been obtained using standard techniques (the MUSCL-Hancock approach with ENO reconstruction for UFORCE and the ADER approach with WENO reconstruction for UFORCE-δ).

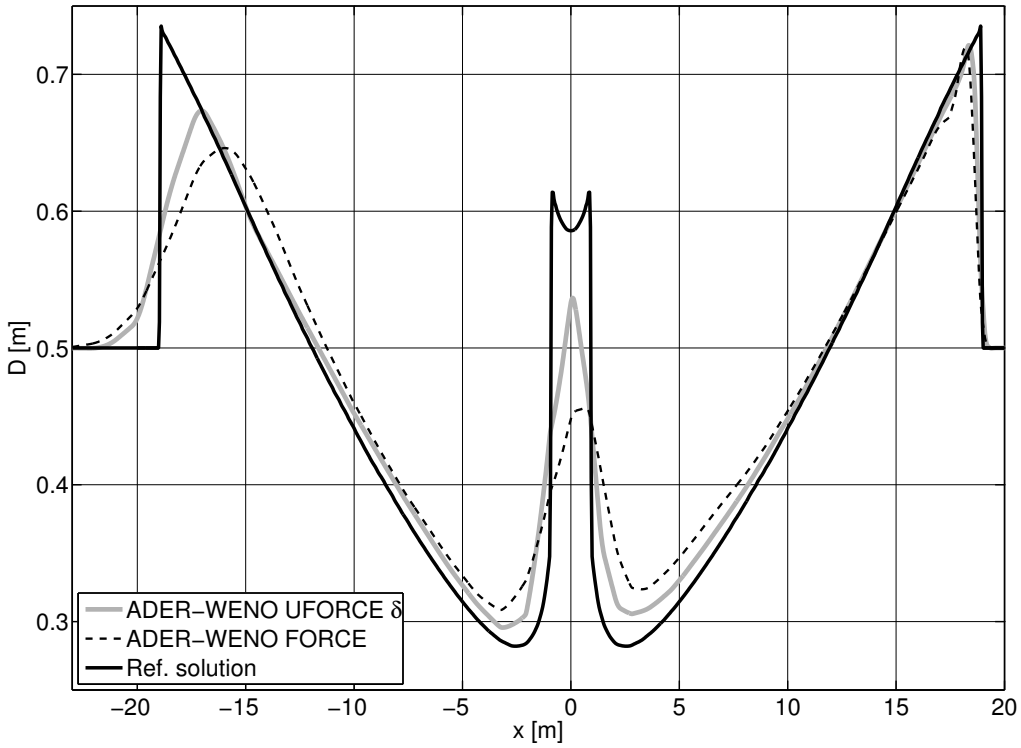

Figure 3.18: Collapse of a circular dam on an unstructured variably-spaced grid. Numerical results for water depth D of the second-order ADER-WENO FORCE and UFORCE-δ numerical methods (symbols) are compared with the reference radial solution (full line) at time $t = 4.7\,s$. The numerical solution profiles are sliced on the x axis. The mesh used is depicted in Fig. 3.17.

Numerical performance of our methods has been assessed by solving the two-dimensional shallow water equations on structured and unstructured meshes. Four different test problem have been solved and the numerical results have been compared with those obtained using two well-established centred methods (the classical centred FORCE scheme [120] and the central-upwind scheme developed by Kurganov, Noelle and Petrova [70]) and two upwind methods (the Godunov upwind method in conjunction with the exact Riemann solver and the approximate HLL Riemann solver). It is shown that the UFORCE method provides a great improvement compared to FORCE, while the UFORCE-δ method outperforms all the other methods except Godunov-exact. Moreover the proposed methods improve the accuracy in the solution for small Courant numbers and intermediate waves associated with linearly degenerate fields (contact discontinuities and shear waves), thus achieving comparable accuracy to that of upwind methods based on the HLL Riemann solver.

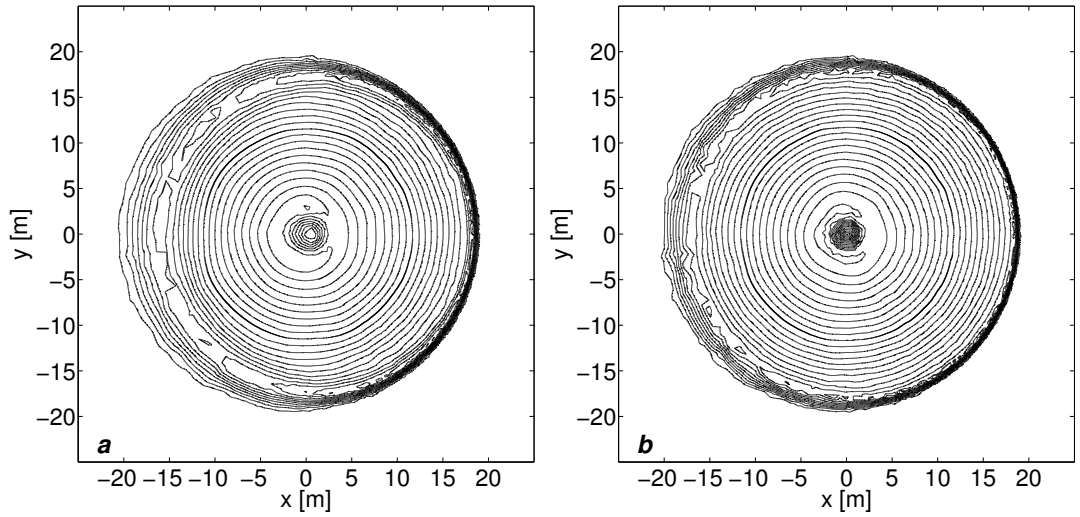

Figure 3.19: Collapse of a circular dam on an unstructured variably-spaced grid. Numerical results for water depth D of the second-order ADER-WENO FORCE (a) and ADER-WENO UFORCE-δ (b) methods are presented in terms of contourplots at time $t = 4.7\,s$.

Finally, the main features of the proposed methods are simplicity and generality (due to their centred nature) and improved accuracy compared to classical centred methods. These characteristics turn out to be very attractive when dealing with hyperbolic systems for which the application of the upwind approach is not viable due to difficulties in the solution of the Riemann problem. This is for instance the case of the two-dimensional hyperbolic system of PDEs governing the hydro-morphodynamic evolution of gravel-bed rivers (2.90)-(2.92), arising from the coupled shallow water-Exner model (2.6). Finding an accurate numerical solution to this system of PDEs is one of the main goals of the present thesis. Since this system is non-conservative, the next step is to extend the UFORCE-δ method to non-conservative hyperbolic systems in the framework of path-conservative methods [42]. This is the object of the next chapter.

4 An upwind-biased centred scheme for hyperbolic systems in non-conservative form

In this chapter we derive a novel upwind-biased centred scheme for the solution of two-dimensional non-conservative homogeneous hyperbolic systems of PDEs in the path-conservative framework. Our scheme, referred to as UPRICE2-Cδ, comes as a modification of the PRICE2-C scheme recently developed by Canestrelli *et al.* [21], representing the path-conservative version of the FORCE centred method [120]. The upwind bias is inserted into PRICE2-C by suitably varying the shape and size of the secondary mesh subvolumes in agreement with upwind information based on the system eigenvalues, in order to achieve fine local control on numerical dissipation.

A key feature of path-conservative schemes of the PRICE-C type is simplicity and generality due to their centred nature (Riemann solvers are not required) and to fully-numerical integration of the Roe matrix, for which no detailed knowledge of the problem eigenstructure is needed. The proposed UPRICE2-Cδ scheme retains these characteristics, while significantly improving accuracy in the solution of the waves associated to linearly degenerate fields (as we show by one test for the shallow water equations in this chapter) and in the computations carried out at low CFL (as we will show in the next chapter) due to insertion of the upwind bias. As we shall see in the next chapter, these features make the proposed scheme very attractive when solving the non-conservative hyperbolic problem arising from the coupled shallow water-Exner model, for which the upwind approach instead is not easily applicable [21] due to the difficulties in the solution of the Riemann problem.

The procedure adopted in the derivation of UPRICE2-Cδ is identical to that presented in the previous chapter concerning the development of the upwind-biased UFORCE-δ method. First the method is derived in the linear case by identifying the optimal upwind-bias, i.e. the relations which allow the UPRICE2-Cδ method to linearly reproduce the Godunov upwind method. Since in the linear case UPRICE2-Cδ is identical to UFORCE-δ, these relations come in a straightforward manner from the framework established in the previous chapter for the conservative case. Then

we develop a formulation of the upwind bias suitable to be applied in the solution of non-linear non-conservative hyperbolic problems and we extend the method to second-order accuracy in the ADER-TVD framework. Finally we apply the scheme to the vortex advection test [99] for the two-dimensional inviscid shallow water equations in order to assess the improvement in the solution of shear waves and empirically prove second-order convergence. The material presented in this chapter is partially covered in a paper in preparation, see Siviglia *et al.* [111].

The chapter outlines as follows. In section 4.1 we review the PRICE2-C centred method. Then in section 4.2 we develop our UPRICE2-Cδ method and in section 4.3 we extend it to second-order accuracy. Finally, the vortex advection test for the shallow water equations is presented in section 4.4.

4.1 The two-dimensional PRICE-C scheme

In this section we review the PRICE2-C method by Canestrelli *et al.* [21] for non-conservative systems, recently developed as the path-conservative version of the FORCE method [120] in the two-dimensional case.

We consider a homogeneous non-conservative hyperbolic system of m partial differential equations having the form

$$\partial_t \mathbf{Q} + \mathbf{A}_1 \partial_x \mathbf{Q} + \mathbf{A}_2 \partial_y \mathbf{Q} = \mathbf{0} \,, \tag{4.1}$$

where $\mathbf{Q} = [\mathrm{q}_1, ..., \mathrm{q}_\mathrm{m}]^T$ denotes the vector of conserved variables and $\mathbf{A}_1$ and $\mathbf{A}_2$ are hyperbolic matrices. Given a unit vector $\vec{n}_j$, the projected system matrix

$$\mathbf{A}_{\hat{n}} = (\mathbf{A}_1, \mathbf{A}_2) \cdot \vec{n}_j \tag{4.2}$$

is also hyperbolic, therefore admitting m real eigenvalues $\lambda_{\hat{n}}^{(1)}, ..., \lambda_{\hat{n}}^{(m)}$, sorted in increasing order.

The mesh is represented by a conforming tassellation by n_e triangular elements T_i of the computational domain

$$\mathcal{T}_\Omega = \bigcup_{i=1}^{n_e} T_i \,. \tag{4.3}$$

Each element T_i has three edges S_j of length $\left|S_j\right|$, with associated outward-pointing face-normal vectors $\vec{n}_j$. Data are represented by cell averages $\mathbf{Q}_i^n$ and the numerical solution, sought at time $t^{n+1} = t^n + \Delta t$, is denoted with $\mathbf{Q}_i^{n+1}$.
By analogy with the FORCE method, the derivation of PRICE2-C makes use of two meshes: a

primary mesh, at which the solution is sought at each time step, i.e. the triangulation (4.3), and a secondary staggered mesh. Each element T_i of the primary mesh, having size $|T_i|$, is sub-divided into subvolumes V_j^- generated by connecting the barycentre $\vec{G}_i$ of T_i with the vertices of S_j. The corresponding adjacent subvolume in the neighbouring element T_j that shares edge S_j with element T_i is denoted with V_j^+. The intersection of V_j^- and V_j^+ gives edge S_j of the element T_i. With h_j^+ and h_j^- we denote the altitudes of triangular subvolumes V_j^+ and V_j^- with respect to S_j. Fig. 3.2 illustrates the above definitions and notation.
For the two-dimensional case the area of subvolumes is given by

$$\left|V_j^-\right| = \frac{h_j^- \left|S_j\right|}{2} \qquad , \qquad \left|V_j^+\right| = \frac{h_j^+ \left|S_j\right|}{2} \, . \tag{4.4}$$

The one-step PRICE2-C scheme is based on the following update formula:

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{\Delta t}{|T_i|} \sum_{j=1}^{3} S_j \mathbf{A}^-_{j+\frac{1}{2}} \left(\mathbf{Q}_j^n - \mathbf{Q}_i^n\right) \, , \tag{4.5}$$

where $\mathbf{A}^-_{j+\frac{1}{2}}$ is

$$\mathbf{A}^-_{j+\frac{1}{2}} = \frac{1}{2}\hat{\mathbf{A}}_{j+\frac{1}{2}} - \frac{V_j^+ V_j^-}{V_j^+ + V_j^-} \frac{1}{\Delta t S_j} \mathbf{I}_m - \frac{1}{4} \frac{\Delta t S_j}{V_j^+ + V_j^-} \hat{\mathbf{A}}^2_{j+\frac{1}{2}} \, . \tag{4.6}$$

In equation (4.6) $\mathbf{I}_m$ is the modified identity matrix, i.e. a hyperbolic matrix resulting from possible modification of some entries in the $m \times m$ identity matrix representing the Lax-Friedrichs type numerical diffusion. The definition of a modified identity matrix is a problem-specific issue (it depends on the considered system of PDEs and on the considered test case) and may be motivated

- by the achievement of well-balancing of the scheme with respect to certain steady states, i.e. in order to exactly reproduce these steady states in the numerical solution;
- in order to avoid or limit spurious diffusion on certain conserved variables, especially those related to intermediate and steady waves.

About these issues we refer the reader to section 5.2.2 in the next chapter in the applications to the shallow water-Exner model (2.6).

Matrix $\hat{\mathbf{A}}_{j+\frac{1}{2}}$ in (4.6) is a function of two states $\mathbf{Q}_i^n$ and $\mathbf{Q}_j^n$. The system (4.1) contains non-conservative products, which, in general, does not make sense in the classical framework of the theory of distributions. However, as we have seen in section 2.3.3 of chapter 2, with the theory developed by Dal Maso *et al.* [42] a rigorous definition of weak solutions can be given using a family of paths $\mathbf{\Psi} = \mathbf{\Psi}\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n, s, \vec{n}_j\right)$ connecting two states $\mathbf{Q}_i^n$ and $\mathbf{Q}_j^n$ across a discontinuity with

$s \in [0,1]$. With this insight $\hat{\mathbf{A}}_{j+\frac{1}{2}}$ in (4.6) is evaluated as a Roe-type matrix

$$\hat{\mathbf{A}}_{j+\frac{1}{2}} = \mathbf{A}_{\boldsymbol{\Psi}}\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n, \vec{n}_j\right) = \int_0^1 \underline{\underline{\mathbf{A}}}\left(\boldsymbol{\Psi}\left(s, \mathbf{Q}_i^n, \mathbf{Q}_j^n\right)\right) \cdot \vec{n}_j \, \mathrm{ds} \tag{4.7}$$

where $\boldsymbol{\Psi}$ denotes a path connecting states $\mathbf{Q}_i^n$ and $\mathbf{Q}_j^n$. Canestrelli *et al.* [22, 21] propose a simple segment path

$$\boldsymbol{\Psi} = \mathbf{Q}_i^n + s\left(\mathbf{Q}_j^n - \mathbf{Q}_i^n\right) \tag{4.8}$$

connecting the states represented by the conserved variables $\mathbf{Q}_i^n$ and $\mathbf{Q}_j^n$. Consistently with the underlying centred philosophy, integration of the Roe matrix (4.7) over the path (4.8) is carried out numerically, by means of a M-point Gaussian quadrature rule:

$$\mathbf{A}_{\boldsymbol{\Psi}}\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n, \vec{n}_j\right) \approx \mathbf{A}_{\boldsymbol{\Psi}}^M\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n, \vec{n}_j\right) = \sum_{q=1}^{M} \omega_q \underline{\underline{\mathbf{A}}}\left(\boldsymbol{\Psi}\left(s_q, \mathbf{Q}_i^n, \mathbf{Q}_j^n\right)\right) \cdot \vec{n}_j \,, \tag{4.9}$$

where s_q are the Gaussian abscissas and ω_q are the Gaussian weights. Conservation can be mantained up to the desired order of accuracy by using a suitable number M of Gaussian points. In our applications we use three-point quadratures.

The centred nature of the PRICE-C method, where no Riemann solver is applied and the Roe matrix is evaluated numerically (4.9) makes the resulting scheme very simple and general. Canstrelli *et al.* [22, 21] apply the PRICE-C scheme to the coupled shallow water-Exner model, for which the solution of the Riemann problem is not generally viable if complex empirically-derived formulae for sediment transport are used.
However, from the previous chapter we are aware that the solutions obtained using a genuinely centred approach are likely to be affected by excessive numerical dissipation in some specific cases, namely for computations carried out at low local CFL and intermediate waves. One possible remedy is to develop an upwind-biased version of the present method, combining the simplicity and generality of centred methods with the accuracy of upwind methods: this is achieved in the next section.

4.2 The two-dimensional UPRICE-Cδ scheme

The purpose of this section is to develop an upwind-biased variation of the PRICE2-C method (4.5)-(4.7) which will be regarded as UPRICE2-Cδ. The sought method shall result from straightforward application in the path-conservative framework of the theory presented in the previous chapter concerning the derivation of the UFORCE-δ method for conservative hyperbolic systems.

As we have done in that case, here we insert an upwind bias into the PRICE-C scheme by suitably varying the size of secondary volumes $V_j^{\pm}$, in order to achieve fine local control on numerical dissipation. This is obtained by removing the constraint that subvolumes V_j^- must join in the barycentre of primary cells T_i and introducing the constraint that each subvolume must have lower size than its own counterpart in PRICE2-C (4.4). See section 3.3.1 of chapter 3 for details. The secondary subvolume size is now given by

$$\left|V_j^-\right| = \frac{\delta_j^- h_j^- \left|S_j\right|}{2} \qquad , \qquad \left|V_j^+\right| = \frac{\delta_j^+ h_j^+ \left|S_j\right|}{2} \, . \tag{4.10}$$

where δ_j^- and δ_j^+ are the upwind-bias parameters associated to S_j. The above geometrical constraints give the following range of variation for the upwind bias parameters:

$$0 \leq \delta_j^- \leq 1 \qquad , \qquad 0 \leq \delta_j^+ \leq 1 \, . \tag{4.11}$$

Notice that the value $\delta_j^{\pm} = 1$ inserted in (4.10) recovers the secondary mesh setup of the centred method (4.4). In Fig. 3.4 the mesh setup for the upwind-biased method in the two-dimensional case is illustrated.

The sought method is given by the update formula (4.5), where in matrix $\mathbf{A}^-_{j+\frac{1}{2}}$ (4.6) we use the subvolume size given by (4.10), obtaining

$$\mathbf{A}^-_{j+\frac{1}{2}} = \frac{1}{2}\left(\hat{\mathbf{A}}_{j+\frac{1}{2}} - \frac{h_j^+ \delta_j^+ h_j^- \delta_j^-}{h_j^+ \delta_j^+ + h_j^- \delta_j^-}\frac{1}{\Delta t}\mathbf{I}_m - \frac{\Delta t}{h_j^+ \delta_j^+ + h_j^- \delta_j^-}\hat{\mathbf{A}}^2_{j+\frac{1}{2}}\right) \, . \tag{4.12}$$

At this stage the values of the upwind bias $\delta_j^{\pm}$ to be inserted in (4.12) are still undetermined. Next, we shall determine the optimal value for the upwind bias, i.e. adaptive functions varying in time and space providing the least-diffusive linearly non-oscillatory method.

In the previous chapter in section 3.3.2 we have addressed an analogous problem for the UFORCE-δ method. The derivation has been carried out in applications to a linear hyperbolic system having two equations and unknowns (3.87)-(3.88), representing the propagation of two waves. System (3.87)-(3.88), turned for the purpose at hand in non-conservative form, reads

$$\partial_t \mathbf{Q} + \mathbf{A}_1 \partial_x \mathbf{Q} + \mathbf{A}_2 \partial_y \mathbf{Q} = \mathbf{0} \, , \tag{4.13}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are 2×2 hyperbolic matrices with constant entries. Given a unit vector $\vec{n}_j$, the projected system matrix $\mathbf{A}_{\hat{n}}$ (4.2) is also hyperbolic, therefore admitting two eigenvalues $\lambda^{(1)}_{\hat{n}}$ and

$\lambda_{\hat{n}}^{(2)}$.

Since the PRICE-Cδ and UFORCE-δ methods are *identical* in the solution of linear systems of PDEs, the analysis presented in section 3.3.2, to which we refer the reader, applies in a straightforward manner in the present case. Thus, we can borrow the optimal upwind bias formulation for our UPRICE2-Cδ method for the linear case, obtaining

$$\delta_j^- = \frac{\left|\lambda_{\hat{n}}^{(1)}\right| \Delta t}{h_j^-} \qquad , \qquad \delta_j^+ = \frac{\left|\lambda_{\hat{n}}^{(2)}\right| \Delta t}{h_j^+} \tag{4.14}$$

or

$$\delta_j^- = \frac{\left|\lambda_{\hat{n}}^{(2)}\right| \Delta t}{h_j^-} \qquad , \qquad \delta_j^+ = \frac{\left|\lambda_{\hat{n}}^{(1)}\right| \Delta t}{h_j^+} \, . \tag{4.15}$$

Last, we shall extend the formulation (4.14) or (4.15) to non-linear hyperbolic systems of PDEs. The sought formulation shall be suitable for applications to general non-conservative non-linear two-dimensional systems of PDEs composed of m equations and unknowns having the form (4.1). We apply again the procedures developed for the UFORCE-δ method in the previous chapter (see section 3.3.4). Consider edge S_j laying between cells T_i and T_j and the projected system matrix $\mathbf{A}_{\hat{n}}$ (4.2), which is in general data-dependent. Given piecewise constant initial data $\mathbf{Q}(x, y \in T_i) = \mathbf{Q}_i^n$, and $\mathbf{Q}(x, y \in T_j) = \mathbf{Q}_j^n$, two choices of the projected matrix can be made:

$$\mathbf{A}_{i\hat{n}} = \mathbf{A}_{\hat{n}}\left(\mathbf{Q}_i^n\right) \qquad , \qquad \mathbf{A}_{j\hat{n}} = \mathbf{A}_{\hat{n}}\left(\mathbf{Q}_j^n\right) \, . \tag{4.16}$$

Each of these matrices possesses m real eigenvalues $\lambda_{i\hat{n}}^{(1)}, ..., \lambda_{i\hat{n}}^{(m)}$, $\lambda_{j\hat{n}}^{(1)}, ..., \lambda_{j\hat{n}}^{(m)}$ (sorted in increasing order). The optimal upwind bias is:

$$\delta_j^- = \frac{\left|\mathrm{s}_{\hat{n}}^{(1)}\right| \Delta t}{h_j^-} \qquad , \qquad \delta_j^+ = \frac{\left|\mathrm{s}_{\hat{n}}^{(m)}\right| \Delta t}{h_j^+} \tag{4.17}$$

or

$$\delta_j^- = \frac{\left|\mathrm{s}_{\hat{n}}^{(m)}\right| \Delta t}{h_j^-} \qquad , \qquad \delta_j^+ = \frac{\left|\mathrm{s}_{\hat{n}}^{(1)}\right| \Delta t}{h_j^+} \, . \tag{4.18}$$

where $\mathrm{s}_{\hat{n}}^{(m)}$ and $\mathrm{s}_{\hat{n}}^{(1)}$ are the fastest and slowest wave travelling normal to interface S_j in the direction given by $\vec{n}_j$. These wave speeds can be easily estimated with reference to the system eigenvalues computed at initial time, which are in any case needed for selecting a time step, in the form

$$\left|\mathrm{s}_{\hat{n}}^{(1)}\right| = \max\left(\left|\lambda_{i\hat{n}}^{(1)}\right|, \left|\lambda_{j\hat{n}}^{(1)}\right|\right) \qquad , \qquad \left|\mathrm{s}_{\hat{n}}^{(m)}\right| = \max\left(\left|\lambda_{i\hat{n}}^{(m)}\right|, \left|\lambda_{j\hat{n}}^{(m)}\right|\right) \, . \tag{4.19}$$

Equations (4.19) introduce the speed of the slowest and fastest wave (extrema of the Riemann fan) in the flux computation, giving a HLL flavour to the PRICE2-C scheme, thus reducing numerical dissipation compared to the standard PRICE2-C. However the method is still centred (not requiring detailed knowledge of the problem eigenstructure), therefore retaining simplicity and generality. Substitution of (4.17) or (4.18) into (4.12) gives the following relation for $\mathbf{A}^-_{j+\frac{1}{2}}$:

$$\mathbf{A}^-_{j+\frac{1}{2}} = \frac{1}{2}\left(\hat{\mathbf{A}}_{j+\frac{1}{2}} - \frac{\left|\mathrm{s}^{(1)}_{\hat{n}}\right|\left|\mathrm{s}^{(m)}_{\hat{n}}\right|}{\left|\mathrm{s}^{(1)}_{\hat{n}}\right| + \left|\mathrm{s}^{(m)}_{\hat{n}}\right|}\mathbf{I}_m - \frac{1}{\left|s^{(1)}_{\hat{n}}\right| + \left|s^{(m)}_{\hat{n}}\right|}\hat{\mathbf{A}}^2_{j+\frac{1}{2}}\right) . \tag{4.20}$$

Notice that in this case, unlike in the conservative case, we do not have to chose whether to apply (4.17) or (4.18) since they both give (4.20).
Finally, the CFL condition satisfying the constraint (4.11) reads:

$$\Delta t = CFL \min_{1\le i\le n_e}\left(\min_{1\le j\le 3}\left(\frac{h^-_j}{\max\left(\left|\mathrm{s}^{(1)}_{\hat{n}}\right|, \left|\mathrm{s}^{(m)}_{\hat{n}}\right|\right)}\right)\right)_{T_i} , \tag{4.21}$$

with $0 < CFL \le 1$.

Summarising, application of the UPRICE2-Cδ method to non-linear systems of PDEs is given by:

- computation of two sets of eigenvalues $\lambda^{(1)}_{i\hat{n}},...,\lambda^{(m)}_{i\hat{n}}$, $\lambda^{(1)}_{j\hat{n}},...,\lambda^{(m)}_{j\hat{n}}$ for each interface S_j;
- evaluation of wave speeds using (4.19);
- selection of the integration time step Δt using (4.21);
- evaluation of the Roe-type matrix $\hat{\mathbf{A}}^M_\Psi$ for each interface S_j using (4.9)
- application of the update formula (4.5) with the matrix $\mathbf{A}^-_{j+\frac{1}{2}}$ given by (4.20).

4.3 Second-order extension

The first-order UPRICE2-Cδ method can be used as a building block for the development of a high-order accurate upwind-biased centred scheme. For a presentation of some of the available techniques for achieving high-order accuracy in the framework of finite volume methods see section 3.4 and references therein. Here we briefly review the one-step ADER-TVD technique used for extending the UPRICE2-Cδ scheme to second-order accuracy. The key ingredients in order to obtain second-order accuracy in the ADER framework [123] are

- a non-oscillatory reconstruction polynomial;
- a space-time predictor based on the Cauchy-Kowalewski procedure;
- the solution of local derivative Riemann problems for the state variables and their derivatives.

4.3.1 The TVD reconstruction technique

We seek for a reconstruction polynomial of degree one in the TVD framework. For each element T_i we identify a stencil of four cells composed of T_i itself and of its three neighbouring cells T_j, with $j = 1,2,3$:

$$\mathcal{S}_i = T_i \bigcup T_{j=1} \bigcup T_{j=2} \bigcup T_{j=3} \, . \tag{4.22}$$

With $\mathbf{w}_i = \mathbf{w}_i^0 + \mathbf{a}_{i1}(x - x_i) + \mathbf{a}_{i2}(y - y_i)$ we define the linear reconstruction polynomial in cell T_i, being $\mathbf{a}_{i1}$ and $\mathbf{a}_{i2}$ the slopes of the reconstructed variables and (x_i, y_i) are the coordinates of the cell barycentre $\vec{G}_i$. We initially compute the values of $\mathbf{w}_i^0$, $\mathbf{a}_{i1}$ and $\mathbf{a}_{i2}$ by requiring integral conservation all over $\mathcal{S}_i$:

$$\frac{1}{|T_k|} \int_{T_k} \mathbf{w}_i(\vec{x}) \, \mathrm{d}\vec{x} = \mathbf{Q}_k^n \qquad \forall T_k \in \mathcal{S}_i \, . \tag{4.23}$$

The reconstruction equation (4.23) is solved using a constrained least-squares method in order to guarantee that it is exactly satisfied at least inside element T_i. For sake of simplicity in our implementation the procedure is performed in a transformed coordinate space (3.130). See section 3.4.2.1 and references [48, 49] for details.
In order to avoid spurious oscillations in the vicinity of large gradients, for each conserved variable we apply a TVD limiter to the linear reconstruction. The following limited reconstruction polynomial is used:

$$\hat{\mathbf{w}}_i = \mathbf{Q}_i^n + \hat{\mathbf{a}}_{i1}(x - x_i) + \hat{\mathbf{a}}_{i2}(y - y_i) \, , \tag{4.24}$$

where $\hat{\mathbf{a}}_{i1} = \boldsymbol{\phi}_i \mathbf{a}_{i1}$ and $\hat{\mathbf{a}}_{i2} = \boldsymbol{\phi}_i \mathbf{a}_{i2}$ are the limited reconstruction slopes. Following Buffard and Clain [17], we adopt a vector of monoslope limiters $\boldsymbol{\phi}_i$, whose components ϕ_i, associated to each variable q_i^n in $\mathbf{Q}_i^n$, are given by

$$\phi_i = \min_{j=1,2,3} (\phi_{i,j}) \, , \tag{4.25}$$

with

$$\phi_{i,j} = \begin{cases} \max\left(0, f_j \frac{|\mathrm{q}_j^n - \mathrm{q}_i^n|}{\vec{a}_i \cdot \vec{s}_j}\right) & \text{if} \quad \vec{a}_i \cdot \vec{s}_j \neq 0 \\ 1 & \text{otherwise} \end{cases} \, , \tag{4.26}$$

where $\vec{a}_i$ denotes the components of the reconstruction slope vector $\vec{\mathbf{a}}_i = (\mathbf{a}_{i1}, \mathbf{a}_{i2})$ and $\vec{s}_j$ and f_j are given by

$$\vec{s}_j = \text{sign}\left(\mathrm{q}_j^n - \mathrm{q}_i^n\right)\left(\vec{R}_j - \vec{G}_i\right) \qquad , \qquad f_j = \frac{\left|\vec{R}_j - \vec{G}_i\right|}{\left|\vec{G}_j - \vec{G}_i\right|} , \tag{4.27}$$

being $\vec{R}_j$ the intersection of the segment connecting $\vec{G}_i$ and $\vec{G}_j$ with the straight line containing edge S_j.

4.3.2 Second-order accurate one-step time discretisation

Next we compute the temporal evolution of the reconstructed polynomial (4.24) in order to construct the final second-order accurate one-step finite volume scheme. The key idea in the ADER approach is to solve high-order Riemann problems at the element boundaries. This is done by using a Taylor series expansion in space and time:

$$\mathbf{Q}_i(x,y,t) = \mathbf{Q}(x_i,y_i,t^n) + (x-x_i)\,\partial_x\mathbf{Q} + (y-y_i)\,\partial_y\mathbf{Q} + (t-t^n)\,\partial_t\mathbf{Q} + \mathbb{O}\left(x^2,y^2,t^2\right) . \tag{4.28}$$

Then, the Cauchy-Kowalewski procedure is applied in order to substitute time derivatives with space derivatives in (4.28). For second-order accuracy it suffices to rewrite system (4.1) as follows:

$$\partial_t\mathbf{Q} = -\left(\mathbf{A}_1\partial_x\mathbf{Q} + \mathbf{A}_2\partial_y\mathbf{Q}\right) . \tag{4.29}$$

The value of $\mathbf{Q}(x_i,y_i,t^n)$ and its spatial derivatives are obtained from the reconstruction polynomial (4.24).

4.3.3 The fully discrete second-order accurate one-step scheme

Once the TVD reconstruction and the Cauchy-Kowalewski procedure have been performed for each cell, giving the space-time polynomial $\mathbf{Q}_i(x,y,t^n)$, the final second-order accurate one-step scheme can be written as:

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{1}{|T_i|}\int_{t^n}^{t^{n+1}}\left(\int_{T_i}\underline{\underline{\mathbf{A}}}(\mathbf{Q}_i)\cdot\nabla\mathbf{Q}_i\,\mathrm{d}\vec{x} + \sum_{j=1}^{3}\int_{S_j}\mathbf{D}^-_{j+\frac{1}{2}}\mathrm{d}\vec{x}\right)\mathrm{d}t , \tag{4.30}$$

with:

$$\mathbf{D}^-_{j+\frac{1}{2}} = \mathbf{A}^-_{j+\frac{1}{2}}\left(\mathbf{Q}_i^+, \mathbf{Q}_j^-, \vec{n}_j\right)\left(\mathbf{Q}_j^- - \mathbf{Q}_i^+\right) , \tag{4.31}$$

where $\mathbf{A}^-_{j+\frac{1}{2}}$ is given by (4.20), $\mathbf{Q}_i^+$ denotes the boundary-extrapolated data from the element T_i and $\mathbf{Q}_j^-$ indicates the boundary-extrapolated data from the neighbouring cell T_j. In our implementation

space and time integrals in (4.30) are approximated using Gaussian quadratures of suitable order.

4.4 Applications to the two-dimensional shallow water equations: vortex advection

In this section we apply the UPRICE2-Cδ scheme in one test case for the two-dimensional inviscid shallow water equations. Further detailed assessment of the performance of the proposed scheme will be found in section 5.5 of the next chapter in applications to the two-dimensional shallow water-Exner model.

The aim of this section is twofold: first to prove second-order convergence of the ADER schemes presented in the previous section, then to show that the upwind-biased formulation of the PRICE2-C method presents a significant improvement in accuracy with respect to the centred formulation of PRICE2-C in the computation of waves associated to linearly degenerate fields. We consider the test problem concerning advection of a potential vortex presented in section 3.5.3. Since this test presents an exact smooth solution, it is suitable for rate convergence study. We recall that in section 3.5.3, comparing the solution profiles obtained by the FORCE (centred) method, the UFORCE-δ (upwind-biased) method and the Godunov-HLL (upwind) method, we have claimed that the degree of accuracy in the solution of shear waves, associated to the linearly degenerate field $\lambda_{\hat{n}}^{(2)} = u_{\hat{n}}$, in the case of vortical flows plays a fundamental role in determining overall accuracy.

The shallow water equations written in non-conservative form read

$$\partial_t \mathbf{Q} + \mathbf{A}_1 \partial_x \mathbf{Q} + \mathbf{A}_2 \partial_y \mathbf{Q} = \mathbf{0} \,, \tag{4.32}$$

where the vector of unknowns $\mathbf{Q}$ and the Jacobian matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are given by

$$\mathbf{Q} = \begin{bmatrix} D \\ uD \\ vD \end{bmatrix} \quad , \quad \mathbf{A}_1 = \begin{bmatrix} 0 & 1 & 0 \\ gD - u^2 & 2u & 0 \\ -uv & v & u \end{bmatrix} \quad , \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 0 & 1 \\ -uv & v & u \\ gD - v^2 & 0 & 2v \end{bmatrix} , \tag{4.33}$$

being D water depth and u and v the x and y components of velocity. The eigenvalues of these Jacobians are given by (3.145) and (3.146).

We solve the system (4.32)-(4.33) with initial condition given by (3.160)-(3.159) and test parameters given in section 3.5.3. In contrast to that solution however, here we use the computational domain $[0,1] \times [0,1]$ with weak far field conditions prescribed at the boundaries $y = 0$ and $y = 1$

and periodic boundary conditions prescribed at the boundaries $x = 0$ and $x = 1$. Since the vortex core is initially located at $\left(\frac{1}{2}, \frac{1}{2}\right)$ and travels with speed equal to $(6,0)$, having set timeout $t = \frac{1}{6}\,s$ and imposed periodic boundary conditions, it is expected to be located again at $\left(\frac{1}{2}, \frac{1}{2}\right)$ at the end of computations.
We use a sequence of regularly-refined triangular meshes characterised by N (reciprocal of mesh length) equal to 10, 20, 40, 80, 160. In order to select a time step we apply the CFL condition (4.21) setting $CFL = 0.9$. We solve the problem using the second-order ADER UPRICE2-Cδ and ADER PRICE2-C schemes with unlimited reconstruction polynomials, i.e. setting $\phi_i = 1$ in (4.25). In both schemes, the modified identity matrix $\mathbf{I}_m$ in (4.7) is simply set as the 3×3 identity matrix.

In Tab. 4.1 and Tab. 4.2 we present the error norms and resulting order of accuracy for variable D of ADER UPRICE2-Cδ and ADER PRICE2-C respectively. Expected second-order accuracy is achieved in each norm. Moreover comparing the norms of the two methods, we notice that those of the centred PRICE2-C method are significantly higher than those of the centred upwind-biased UPRICE2-Cδ method, as expected. This is a clear indication of the improvement in accuracy given by insertion of an upwind bias into the scheme.

Table 4.1: Vortex advection. Convergence rate study for the second-order ADER UPRICE2-Cδ method for variable D. N denotes the reciprocal of mesh length, $N_0 = 10$.

N/N_0	L_∞	$\mathbb{O}(L_\infty)$	L_1	$\mathbb{O}(L_1)$	L_2	$\mathbb{O}(L_2)$
2	1.487E-01	1.35	7.233E-03	1.40	1.874E-02	1.43
4	3.403E-02	2.13	1.932E-03	1.90	4.145E-03	2.18
8	6.345E-03	2.42	4.186E-04	2.21	7.818E-04	2.41
16	1.524E-03	2.06	9.756E-05	2.10	1.680E-04	2.22

Table 4.2: Vortex advection. Convergence rate study for the second-order centred ADER PRICE2-C method for variable D. N denotes the reciprocal of mesh length, $N_0 = 10$.

N/N_0	L_∞	$\mathbb{O}(L_\infty)$	L_1	$\mathbb{O}(L_1)$	L_2	$\mathbb{O}(L_2)$
2	2.415E-01	0.77	8.501E-03	1.02	2.890E-02	0.92
4	6.597E-02	1.87	2.606E-03	1.71	7.628E-03	1.92
8	1.246E-02	2.40	5.677E-04	2.20	1.474E-03	2.37
16	3.234E-03	1.95	1.266E-04	2.17	2.899E-04	2.35

4.5 Conclusions

In this chapter we have applied in the framework of path-conservative schemes for non-conservative hyperbolic systems of PDEs the upwind-biased approach developed in chapter 3 for

conservative methods. We have derived an upwind-biased modification of the PRICE2-C scheme [21] by suitably varying the secondary mesh setup in the genuinely centred PRICE2-C method. The resulting UPRICE2-Cδ scheme in linear applications is identical to the UFORCE-δ scheme [113] presented in the previous chapter: therefore the development of UPRICE2-Cδ simply consists in straightforward application of the mathematical theory underlying UFORCE-δ. The proposed scheme identically reproduces the Godunov upwind method in the solution of a linear hyperbolic system having two equations and unknowns, provided the optimal upwind bias is used. Extension of UPRICE-Cδ to non-linear hyperbolic systems of PDEs is obtained empirically following a centred approach, i.e. using the system eigenvalues as wave speed estimates. Finally, the scheme has been extended to second-order accuracy in the ADER-TVD framework and second-order convergence has been assessed using the vortex advection test for the two-dimensional inviscid shallow water equations. In this test, comparison with the results of the standard PRICE2-C method highlights the improvement in accuracy in the resolution of intermediate waves given by insertion of the upwind bias.

Application of the UPRICE2-Cδ scheme requires only estimates of the system eigenvalues, which are in any case needed for selecting a time step: no detailed knowledge of the system eigenstructure is required. The main feature of the scheme, beyond improved accuracy, is generality, due to its centred nature and to the use of a numerical procedure for the integration of the Roe matrix. These characteristics make the proposed scheme very attractive for the numerical solution of the non-conservative system of PDEs arising from the coupled shallow water-Exner model (2.6), for which the system eigenstructure in general is not easily available. The construction of a two-dimensional hydro-morphodynamic solver including the ADER-TVD UPRICE2-Cδ scheme is achieved in the next chapter.

5 GIAMT2D: a two-dimensional hydro-morphodynamic solver

In this chapter we design and test our two-dimensional hydro-morphodynamic numerical model, which we regard as the GIAMT2D[1] model. GIAMT2D solves the two-dimensional shallow water-Exner model (2.6) written in non-conservative form (2.90)-(2.92) on unstructured triangular meshes. The non-conservative character of the system arises from the adoption of a fully-coupled formulation of the hydro-morphodynamic problem, i.e. from seeking for a numerical solution to be obtained in a synchronous fashion for all the unknowns. The advantage of this approach is that the model is correctly able to deal with sediment transport near Froude-critical conditions. As a consequence of this choice, a path-conservative scheme has to be applied in the solution. GIAMT2D embeds the path-conservative UPRICE2-Cδ upwind-biased centred scheme developed in the previous chapter, able to combine accuracy comparable to that of upwind schemes with the simplicity and generality of centred schemes. A key feature of the scheme is the ability to include any existing algebraic bedload transport formula in computations, regardless of its mathematical complexity. The UPRICE2-Cδ scheme is extended to second-order accuracy in the ADER-TVD framework and included in a robust second-order preserving splitting technique for the treatment of frictional source terms. A novel wetting-and-drying procedure enables the GIAMT2D model to deal with partially dry domains in an adaptive and robust manner. For sake of computational efficiency, the code is parallelised using the OpenMP protocol on shared memory. The performance of the numerical model is carefully assessed in well-established test cases by checking the numerical solutions against analytical solutions and laboratory experiments. The material presented in this chapter is partially covered in a paper in preparation, see Siviglia *et al.* [111]

The chapter outlines as follows. In section 5.1 we address some preliminary problems in order to identify the most suitable solution technique. In section 5.2 we present the splitting technique

[1]The GIAMT2D model has been developed within my research activity in the *Gruppo di Idraulica Ambientale e Morfologia, Trento - GIAMT* (research Group in Environmental Hydraulics and Morphology, Trento), from which the name comes.

adopted for treating the frictional source terms, which embeds the solution of a homogeneous hyperbolic system of PDEs in non-conservative form obtained using the ADER-TVD UPRICE2-Cδ scheme. Our original recipe for the treatment of wetting-and-drying is presented in section 5.3 and the code parallelisation is introduced and tested in section 5.4. Finally in section 5.5 we employ a suite of classical tests with fixed and movable bed to demonstrate the capabilities of the GIAMT2D model.

5.1 Preliminaries

Before designing our numerical solver for the shallow water-Exner model, three questions must be addressed, concerning

- the mathematical nature of the system of PDEs: is the system hyperbolic?
- the coupling or uncoupling of the hydrodynamic and morphodynamic part of the problem: shall we seek for a coupled or uncoupled solution?
- the choice of the numerical scheme: do we need a centred or an upwind scheme?

The first question is preliminary with respect to the other ones since it deals with assessing well-posedness of the initial value problem given by system (2.6) together with appropriate initial conditions. In section 2.4 of chapter 2, following the analysis carried out by Cordier *et al.* [35] we have proved in the one-dimensional inviscid case that the considered mathematical model is hyperbolic in the range of flow data which are of interest in practical applications in rivers. The *sufficient* condition for hyperbolicity (2.86), namely that the local Froude number does not exceed the threshold value of 6 if the Manning friction law is adopted, is satisfied in all the hydro-morphodynamic applications which will be presented in the following of this thesis.

Concerning the second question, we recall from section 2.5 of chapter 2 that essentially two approaches are available in the literature to solve this system of equations. The first one is the *uncoupled* approach, according to which the shallow water equations are solved separately from the Exner equation: the solution of the hydrodynamic problem is then passed to the morphodynamic module (and vice-versa) at the end of each time step, see e.g. [41, 45, 136]. Alternatively, the uncoupling may concern only some terms in the partial differential system, such as the bed elevation gradients in the momentum equations [19, 23, 61, 24, 90, 137] (see section 2.5): although one-step synchronous solution strategies can be adopted in this case, mathematically this formulation is still uncoupled. Conversely, the second approach is referred to as *coupled* and relies on full coupling of the governing equations within each time step, see e.g. [83, 58, 74, 36, 37, 110, 22, 21, 111].

The choice of the solution strategy ultimately relies on the relative order of magnitude of the celerities associated with the characteristic curves of the hyperbolic system. The uncoupled strategy is usually justified by the different time scales characterising flow and sediment transport and the inherent inaccuracies introduced by the use of empirical formulae for friction and sediment transport capacity. It is well known that far from the critical flow conditions the bed evolves on a time scale which is considerably greater than that on which the surface water evolves, i.e. the bed interacts only weakly with the water surface, thus justifying an approach in which the equations that govern the dynamics of the liquid phase are solved separately from those governing the solid phase.
However, a Froude trans-critical region exists ($0.8 \lesssim F_r \lesssim 1.2$) in which the celerity of the wave associated to bed variations has the same order of magnitude as that of one of the non-linear waves associated to the hydrodynamic variables [109]. In this case, the free surface strongly interacts with the bed [84] and the coupled approach is mandatory. Moreover, Saiedi [104], Cao *et al.* [25] and Cordier *et al.* [35] have compared the numerical stability of coupled and uncoupled models, finding that coupled models are more stable, especially in the case of rapid variation of bottom elevation.
From these considerations and recalling the importance of correctly solving trans-critical flows in gravel-bed rivers, in GIAMT2D we use a coupled strategy for the solution of the hydrodynamic and morphodynamic problem. This results in the non-conservative problem (2.90)-(2.92), which requires the adoption of a path-conservative method.

Concerning the last question, from chapter 3 we know that in the framework of finite volume methods two different approaches exist for the solution of hyperbolic systems of PDEs: the upwind approach and the centred approach.
Examples of the application of both strategies for the solution of hydro-morphodynamic problems in one and two space dimensions are available in the literature. Upwind schemes have been applied to the case of high sediment transport in the one-dimensional case [102] and in the two-dimensional case on Cartesian meshes [2], while examples of two-dimensional hydro-morphodynamic models on unstructured meshes based on upwind schemes are given in [46, 81, 7, 27]. The centred approach has been used in the solution of one-dimensional hydro-morphodynamic problems in [61, 19, 22] and two-dimensional problems on unstructured meshes in [21].
Upwind methods are generally recognised as more accurate than centred methods, the disadvantage being however that they are of more difficult implementation and less general due to the use of Riemann solvers. We refer the reader to chapters 3 and 4, where a detailed comparison of the features of these two approaches has been carried out. For the problem at hand, i.e. solving the shallow water-Exner model, following Canestrelli *et al.* [22, 21], we claim that the upwind approach has an evident drawback, since the solution of the Riemann problem for this hyperbolic

system of PDEs is not generally available, at least if complex empirical formulae for sediment transport (2.16), which are of common use in the engineering practice, are employed. The PRICE-C centred approach [22, 21] instead, which does not require the adoption of Riemann solvers and relies on fully-numerical integration of the Roe matrix (4.9), does not pose any limitation on the bedload transport formulae to be used. For these reasons in the GIAMT2D numerical model we choose a refined PRICE-C type *centred upwind-biased* numerical scheme, including some upwind information based on estimates of the system eigenvalues, still avoiding the use of Riemann solvers. Thus the GIAMT2D model will be able to apply any bedload transport formula, obtaining solutions of accuracy comparable to that of upwind methods. We remark that, since the system eigenvalues are in any case needed to select a time step using the CFL condition, the application of an upwind bias does not add any computational effort with respect to the genuinely centred scheme.

Finally, the GIAMT2D solver shall be applied easily in the case of complex polygonal computational domains, which may occur in simulations aiming at reproducing the hydro-morphodynamic evolution of natural and regulated river reaches characterised by complex planforms. Therefore the problem shall be conveniently discretised using unstructured meshes. The UPRICE2-Cδ numerical scheme developed in chapter 4 possesses all the characteristics and requirements previously listed, i.e.:

- being path-conservative, is able to solve the non-conservative problem (2.90)-(2.92) arising from the adoption of a coupled formulation for the mathematical model (2.6), thus being able to correctly deal with bedload transport in trans-critical flow conditions;
- being centred and using a numerical procedure for the evaluation of the Roe matrix, can handle any bedload transport formula required in the engineering practice;
- being upwind-biased, it is characterised by improved accuracy with respect to the genuinely centred PRICE2-C scheme and may reach comparable accuracy to that of upwind methods based on incomplete Riemann solvers;
- has been natively developed on unstructured triangular meshes.

5.2 Splitting technique

We aim at solving the two-dimensional non-conservative hyperbolic system (2.90)-(2.92). We discretise the domain assuming a conforming triangulation $\mathcal{T}_\Omega$ of the computational domain $\Omega \subset \mathrm{R}^2$ by n_e elements T_i such that

$$\mathcal{T}_\Omega = \bigcup_{i=1}^{n_e} T_i \, . \tag{5.1}$$

Each element T_i has 3 edges S_j of length $\left|S_j\right|$, with associated outward-pointing face-normal vectors $\vec{n}_j = (n_{jx}, n_{jy})$. Data are represented by cell averages $\mathbf{Q}_i^n$ at time t^n and the numerical solution sought at time $t^{n+1} = t^n + \Delta t$ is denoted with $\mathbf{Q}_i^{n+1}$.

The system (2.90)-(2.92) is non-homogeneous due to the presence of the frictional source terms (2.92). We handle inhomogeneity adopting a robust splitting technique which formally preserves second-order accuracy [117]. We rewrite the considered system in the form

$$\mathcal{C}(\mathbf{Q}) = \mathcal{S}(\mathbf{Q}) \ , \tag{5.2}$$

where $\mathcal{C}(\mathbf{Q})$ represents the advective operator

$$\mathcal{C}(\mathbf{Q}) = \partial_t \mathbf{Q} + \mathbf{A}_1 \partial_x \mathbf{Q} + \mathbf{A}_2 \partial_y \mathbf{Q} \tag{5.3}$$

and $\mathcal{S}(\mathbf{Q})$ represents the frictional source term operator. Consider the full initial value problem (IVP):

$$\left.\begin{array}{lrcl} \text{PDEs:} & \mathcal{C}(\mathbf{Q}) & = & \mathcal{S}(\mathbf{Q}) \\ \text{IC:} & \mathbf{Q}(x,y,0) & = & \mathbf{Q}_i^n \end{array}\right\} \text{IVP} \ . \tag{5.4}$$

In the present splitting technique, the complete IVP (5.4) is split into three different IVPs as

$$\left.\begin{array}{lrcl} \text{ODEs:} & \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}t} & = & \mathcal{S}(\mathbf{Q}) \\ \text{IC:} & \mathbf{Q}(x,y,0) & = & \mathbf{Q}_i^n \end{array}\right\} \overset{\frac{1}{2}\Delta t}{\Longrightarrow} \overline{\mathbf{Q}}_i \quad \text{IVP1} \ , \tag{5.5}$$

$$\left.\begin{array}{lrcl} \text{PDEs:} & \mathcal{C}(\mathbf{Q}) & = & 0 \\ \text{IC:} & \mathbf{Q}(x,y,0) & = & \overline{\mathbf{Q}}_i \end{array}\right\} \overset{\Delta t}{\Longrightarrow} \overline{\overline{\mathbf{Q}}}_i \quad \text{IVP2} \ , \tag{5.6}$$

$$\left.\begin{array}{lrcl} \text{ODEs:} & \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}t} & = & \mathcal{S}(\mathbf{Q}) \\ \text{IC:} & \mathbf{Q}(x,y,0) & = & \overline{\overline{\mathbf{Q}}}_i \end{array}\right\} \overset{\frac{1}{2}\Delta t}{\Longrightarrow} \mathbf{Q}_i^{n+1} \quad \text{IVP3} \ . \tag{5.7}$$

IVP1 (5.5) and IVP3 (5.7) contain systems of ODEs related to the frictional source terms (2.92) while IVP2 (5.6) contains a homogeneous non-conservative hyperbolic system of PDEs representing the advective part of the hydro-morphodynamic problem (2.90)-(2.92). The proposed splitting procedure formally preserves second-order accuracy for the solution to (5.4), provided the solutions of IVP1, IVP2 and IVP3 are themselves second-order accurate [117]. The time step Δt is obtained imposing a CFL condition of the type (4.21) for IVP2; then IVP1 and IVP3 are integrated by $\frac{1}{2}\Delta t$

while IVP2 is integrated by Δt.
We emphasise that the technique here described concerns the uncoupling of the frictional source term only, but does *not* affect the coupling of the hydrodynamic and morphodynamic part in the solution of the advective problem IVP2 (5.6).

The splitting technique allows the adoption of different numerical schemes for different types of problems: for IVP1 and IVP3 we use the second-order implicit Runge-Kutta method, while for IVP2 we apply the second-order ADER-TVD extension of the UPRICE2-Cδ scheme presented in chapter 4.

5.2.1 The implicit Runge-Kutta method

In order to integrate the systems of ODEs in IVP1 (5.5) and IVP3 (5.7) by a time step $\frac{1}{2}\Delta t$ each, we use a second-order accurate implicit discretisation via the Runge-Kutta method. Implicitness allows the scheme to handle possibly stiff source terms (2.92) under arbitrarily large time steps. At this stage, the integration procedure is cell-based, i.e. the solution is updated in time within each cell T_i with no contribution from the boundaries. In the following we illustrate the solution for IVP1. An analogous procedure is implemented for solving IVP3.

The right-hand side of the ODEs in (5.5) is evaluated at time $t = \frac{1}{2}\Delta t$ making use of a time Taylor expansion. After straightforward manipulations, the system is discretised in the form

$$\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}t} = \left(\mathbf{I} - \frac{1}{2}\Delta t \left[\frac{\partial \mathbf{S}}{\partial \mathbf{Q}} \right]_{\mathbf{Q}_i^n} \right)^{-1} \mathbf{S}\left(\mathbf{Q}_i^n\right) \,, \tag{5.8}$$

where $\mathbf{I}$ represents the 4×4 identity matrix and $\left[\frac{\partial \mathbf{S}}{\partial \mathbf{Q}}\right]$ is the matrix containing the derivatives of the source terms (2.92) with respect to the unknowns, which reads:

$$\left[\frac{\partial \mathbf{S}}{\partial \mathbf{Q}}\right] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \frac{7}{3}gn^2 \frac{q_x q}{(H-\eta)^{\frac{10}{3}}} & -gn^2 \frac{q+\frac{q_x^2}{q}}{(H-\eta)^{\frac{7}{3}}} & -gn^2 \frac{q_x q_y}{q(H-\eta)^{\frac{7}{3}}} & -\frac{7}{3}gn^2 \frac{q_x q}{(H-\eta)^{\frac{10}{3}}} \\ \frac{7}{3}gn^2 \frac{q_y q}{(H-\eta)^{\frac{10}{3}}} & -gn^2 \frac{q_x q_y}{q(H-\eta)^{\frac{7}{3}}} & -gn^2 \frac{q+\frac{q_y^2}{q}}{(H-\eta)^{\frac{7}{3}}} & -\frac{7}{3}gn^2 \frac{q_y q}{(H-\eta)^{\frac{10}{3}}} \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{5.9}$$

if the Manning law is adopted.

Equation (5.8) is suitable for time integration using the well-established second-order accurate Runge-Kutta scheme, assuming as initial condition $\mathbf{Q}_i^n$ at time t^n:

$$\overline{\mathbf{Q}}_i = \mathbf{Q}_i^n + \frac{1}{2}(\mathbf{K}_1 + \mathbf{K}_2) \,, \tag{5.10}$$

with

$$\mathbf{K}_1 = \frac{1}{2}\Delta t \left(\mathbf{I} - \frac{1}{2}\Delta t \left[\frac{\partial \mathbf{S}}{\partial \mathbf{Q}} \right]_{\mathbf{Q}_i^n} \right)^{-1} \mathbf{S}(\mathbf{Q}_i^n) \,, \tag{5.11}$$

and

$$\mathbf{K}_2 = \frac{1}{2}\Delta t \left(\mathbf{I} - \frac{1}{2}\Delta t \left[\frac{\partial \mathbf{S}}{\partial \mathbf{Q}} \right]_{\mathbf{Q}_i^n + \mathbf{K}_1} \right)^{-1} \mathbf{S}(\mathbf{Q}_i^n + \mathbf{K}_1) \,. \tag{5.12}$$

5.2.2 The advection scheme

This section concerns the solution to the advective part of the hydro-morphodynamic problem IVP2 (5.6) to be performed after the solution to IVP1 (5.5) has been computed. As discussed and motivated in section 5.1, for IVP2 we adopt a coupled approach, relying on synchronous treatment of the hydrodynamic and morphodynamic part of the problem. This strategy results in the solution of a non-conservative homogeneous problem having the form

$$\partial_t \mathbf{Q} + \mathbf{A}_1 \partial_x \mathbf{Q} + \mathbf{A}_2 \partial_y \mathbf{Q} = \mathbf{0} \,, \tag{5.13}$$

with vector of unknowns $\mathbf{Q}$ and system matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ given in (2.91), to be obtained by means of a path-conservative method. With initial condition given by $\overline{\mathbf{Q}_i}$ (5.10), we seek for a numerical solution $\overline{\overline{\mathbf{Q}_i}}$ of (5.13) applying the centred upwind-biased UPRICE2-Cδ method presented in the previous chapter extended to second-order accuracy in the ADER-TVD framework (4.30)-(4.31). Next, we shall tailor the ADER-TVD UPRICE2-Cδ scheme for the solution of the considered problem.

Application of the UPRICE2-Cδ method requires knowledge of approximations of the system eigenvalues, to be used as estimates of wave speed celerities in the form (4.19) for the selection of the integration time step using the CFL condition (4.21) and for local fine tuning of numerical dissipation through (4.20). Here we employ an approximation of the eigenvalues of the shallow water-Exner model obtained in the one-dimensional case by Lyn and Altinakar [84]. We apply this approximation using the flow data projected normal to the current cell interface S_j, i.e. considering the velocity projection

$$u_{\hat{n}} = (u, v) \cdot \vec{n}_j \,, \tag{5.14}$$

which allows us to define a projected Froude number $F_{r\hat{n}} = \frac{|u_{\hat{n}}|}{\sqrt{g(H-\eta)}}$.
The one-dimensional model possesses three eigenvalues $\lambda_{\hat{n}}^{(1)} < \lambda_{\hat{n}}^{(2)} < \lambda_{\hat{n}}^{(3)}$. Two distinct cases must be considered, namely the case of sub-critical ($F_{r\hat{n}} \lesssim 0.8$) or super-critical flows ($F_{r\hat{n}} \gtrsim 1.2$) and the case of trans-critical flows ($0.8 \lesssim F_{r\hat{n}} \lesssim 1.2$):

- in sub-critical or super-critical flows the extrema of the eigenvalue set are readily identified with the eigenvalues of the fixed-bed shallow water model

$$\lambda_{\hat{n}}^{(1)} = u_{\hat{n}} - \sqrt{g(H-\eta)} \qquad , \qquad \lambda_{\hat{n}}^{(3)} = u_{\hat{n}} + \sqrt{g(H-\eta)}\,, \tag{5.15}$$

  while the intermediate eigenvalue, associated to bedload transport, reads

$$\lambda_{\hat{n}}^{(2)} = \left[\frac{\zeta}{1-F_{r\hat{n}}^2}\right] u_{\hat{n}}\,, \tag{5.16}$$

  being $\zeta = \frac{1}{(1-\lambda_p)(H-\eta)}\frac{\partial q_{s\hat{n}}}{\partial u_{\hat{n}}}$, with $q_{s\hat{n}} = (q_{sx}, q_{sy})\cdot\vec{n}_j$;

- in trans-critical flows, eigenvalues are given by

$$\begin{aligned} \lambda_{\hat{n}}^{(1)} &= \left\{\tfrac{1}{4}\left(1-\tfrac{1}{F_{r\hat{n}}^2}\right) - \sqrt{\left(1-\tfrac{1}{F_{r\hat{n}}^2}\right)^2 + \tfrac{8\zeta}{F_{r\hat{n}}^2}}\right\} u_{\hat{n}}\,, \\ \lambda_{\hat{n}}^{(2)} = \left\{\tfrac{1}{4}\left(1-\tfrac{1}{F_{r\hat{n}}^2}\right) + \sqrt{\left(1-\tfrac{1}{F_{r\hat{n}}^2}\right)^2 + \tfrac{8\zeta}{F_{r\hat{n}}^2}}\right\} u_{\hat{n}} &\qquad , \qquad \lambda_{\hat{n}}^{(3)} = \left(\tfrac{3}{2} + \tfrac{1}{2F_{r\hat{n}}^2}\right) u_{\hat{n}} \end{aligned} \tag{5.17}$$

  if $u_{\hat{n}} > 0$ and

$$\begin{aligned} \lambda_{\hat{n}}^{(1)} = \left(\tfrac{3}{2} + \tfrac{1}{2F_{r\hat{n}}^2}\right) u_{\hat{n}} &\qquad , \qquad \lambda_{\hat{n}}^{(2)} = \left\{\tfrac{1}{4}\left(1-\tfrac{1}{F_{r\hat{n}}^2}\right) + \sqrt{\left(1-\tfrac{1}{F_{r\hat{n}}^2}\right)^2 + \tfrac{8\zeta}{F_{r\hat{n}}^2}}\right\} u_{\hat{n}}\,, \\ \lambda_{\hat{n}}^{(3)} &= \left\{\tfrac{1}{4}\left(1-\tfrac{1}{F_{r\hat{n}}^2}\right) - \sqrt{\left(1-\tfrac{1}{F_{r\hat{n}}^2}\right)^2 + \tfrac{8\zeta}{F_{r\hat{n}}^2}}\right\} u_{\hat{n}} \end{aligned} \tag{5.18}$$

  if $u_{\hat{n}} < 0$.

Following Canestrelli *et al.* [21], in the definition of the matrix $\mathbf{A}^{-}_{j+\frac{1}{2}}$ (4.20) we use the following modified identity matrix:

$$\mathbf{I}_m = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \varepsilon^b \end{bmatrix}\,, \tag{5.19}$$

where ε^b represents the ratio between the speed of waves associated to bed elevation and the maximum wave speed in absolute value of the entire Riemann fan. The formulation (5.19) not only improves accuracy in the solution of the waves associated to bed variations, but is also crucial for avoiding undesirable diffusion of bed elevation gradients with quiescent water: when water is locally at rest, ε^b shall be set to zero. We estimate ε^b using the above eigenvalues in the form

$$\varepsilon^b = \max\left(\frac{\left|\lambda^{(2)}_{i\hat{n}}\right|}{\max\left(\left|\lambda^{(1)}_{i\hat{n}}\right|, \left|\lambda^{(3)}_{i\hat{n}}\right|\right)}, \frac{\left|\lambda^{(2)}_{j\hat{n}}\right|}{\max\left(\left|\lambda^{(1)}_{j\hat{n}}\right|, \left|\lambda^{(3)}_{j\hat{n}}\right|\right)} \right) , \tag{5.20}$$

where subscripts i and j indicate that eigenvalues are computed with the data $\overline{\mathbf{Q}}_i$ and $\overline{\mathbf{Q}}_j$ associated to cells T_i and T_j respectively, still using the normal unit vector $\vec{n}_j$ outward-pointing from T_i. Notice that elimination of spurious diffusion of bed gradients with quiescent water is immediately achieved in our implementation: if the projected velocity $u_{\hat{n}}$ vanishes in both cells T_i and T_j sharing edge S_j, application of (5.16) gives $\varepsilon^b = 0$ in (5.20).

5.3 Treatment of wetting-and-drying fronts

In applications aiming at reproducing the behaviour of real gravel-bed rivers, the case of partially dry domains is likely to occur. Since the location and extent of these dry regions may change during simulations, numerical models must be able to deal with these features in an adaptive and robust manner. From the point of view of numerical analysis, several issues are required by the solution of system (2.90)-(2.92) over possibly dry domains, including satisfaction of the extended C-property [28], preservation of depth positivity and correct prediction of the celerity of wet-dry fronts propagating over flat and sloping domains. Since these requirements are not generally satisfied by standard numerical methods developed for fully-wet domains, increasingly refined numerical solution strategies have been developed. Most of these schemes apply only to fixed-bed problems. Among them, Brufau *et al.* [16] and Castro *et al.* [29] locally modify the Roe upwind method in order to achieve a better description of wet-dry fronts, while robust depth-positivity preserving schemes have been built using the hydrostatic reconstruction procedure [3, 4, 31].

The mathematical model composed by the system of PDEs (2.6) along with the closure relations for friction and bedload transport presented in section 2.1.2 is inadequate for straightforward numerical treatment in possibly dry domains. On very shallow depth ($D \to 0$), the relationship linking velocity, discharge and depth

$$u = \frac{q_x}{D} \qquad , \qquad v = \frac{q_y}{D} \tag{5.21}$$

becomes singular since, from a physical point of view, all these quantities have to tend to zero. In general on vanishing depth the computations involving evaluation of the inverse of D might be source of numerical instability. This is likely to happen when applying the Manning (2.10) or Chézy (2.12) friction laws for the evaluation of the dimensionless friction term $\vec{S}_f$. In detail, difficulties are found in the computation of

- the second and third line of the system matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ (2.91), containing velocities (inertial terms);
- the derivatives of bedload transport in the first and fourth line of matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ (2.91), for which formulae based on the Shields stress θ (2.16), requiring the adoption of the above friction models, are commonly applied;
- of the vector of source terms (2.92) and of its derivatives with respect to the conserved variables, employing the above mentioned friction formulae.

Notice that, in our solution procedure based on the splitting technique, the first two difficulties concern the solution of the advection problem IVP2 (5.6), i.e. the evaluation of the system matrices (2.91), while the last problem is encountered in the solution of IVP1 (5.5) and IVP3 (5.7) for the numerical treatment of the frictional source terms (2.92). We observe that over-prediction of these terms on vanishing depth may result in a friction-dominated, stiff problem: this is the motivation for the adoption of a splitting technique, in which the frictional source terms are taken into account using an implicit integration scheme, able to handle possible stiffness as claimed in section 5.2.1.

In practice, most of the existing algorithms embedded in hydro-morphodynamic models use a small preset value of minimum depth to detect dry cells and stabilise the propagation of wet-dry fronts, see e.g. [15, 80, 91, 103]. Our wetting-and drying procedure is based on values for threshold depth as well, which are used in order to smooth the inertial and bedload transport terms in the system matrices and the frictional source terms on vanishing depth. While in the wet core of the domain we use a second-order accurate method, at the transition between wet and dry regions we employ a first-order monotone scheme, which in this case is more robust due to higher numerical dissipation and absence of spurious oscillations. We define three different thresholds:

- ε^D is minimum depth used for detecting dry cells;
- $\varepsilon^F > \varepsilon^D$ is minimum depth under which the discharge values are reduced in the computation of the system matrices (2.91);
- $\varepsilon^{1o} > \varepsilon^D$ is used for detecting cells to be solved using a first-order scheme.

Next, we give the modifications in the solution procedure which are needed to handle wetting-and-drying fronts.

- *Friction modifications.*
  The present modifications apply in solution to (5.5) and (5.7) carried out in the cells having local depth lower than ε^D (dry cells). They are required for the evaluation of the right-hand side of (5.8), containing the frictional source terms and their derivatives with respect to the conserved variables. Before performing Runge-Kutta time integration, each cell T_i is checked for minimum depth ε^D. If local depth is lower than ε^D, the Runge-Kutta procedure is skipped and in its output the discharge terms are simply set to zero; conversely in the wet cells (depth higher than ε^D) the Runge-Kutta procedure (5.10)-(5.12) is performed normally.

- *Momentum and Shields parameter modifications.*
  These modifications aim at stabilising the computation of inertial terms (velocities) in the second and third line in the system matrices (2.91) and of the derivatives of bedload transport in the first and fourth line of the matrices. Stabilisation is achieved by reducing the momentum terms q_x and q_y, i.e. multiplying these values by a reduction coefficient r^F having the form

$$r^F = \begin{cases} 1 & \text{if } D > \varepsilon^F \\ \frac{D-\varepsilon^D}{\varepsilon^F-\varepsilon^D} & \text{if } \varepsilon^D < D \leq \varepsilon^F \\ 0 & \text{if } D \leq \varepsilon^D \end{cases} , \tag{5.22}$$

  $D = H - \eta$ being the associated value of depth in the matrix evaluation. Thus, such reduction of q_x and q_y is effective only for $D < \varepsilon^F$. The reduced momentum values are also applied in the computation of the Shields number using (2.19) or (2.20) and of its derivatives with respect to the unknowns, which ensures that bedload transport tends to vanish on shallow depth.

- *Renounce to second-order accuracy.*
  First-order schemes prove to be more robust in the treatment of wet and dry fronts. Thus, whereas in general we use a second-order ADER-TVD scheme throughout the wet portions of the domain, we adopt a first-order scheme on vanishing depth. For this purpose we use a preset of depth ε^{1o}: if depth is lower than ε^{1o} at least one cell in the stencil associated to the current cell T_i (4.22), i.e. T_i and its three neighbours, the ADER-TVD procedure for T_i is simply skipped, thus recovering a first-order method.

- *Treatment of emerging topography.*
  Consider the situation in which the current cell T_i is wet ($D_i > \varepsilon^D$), the neighbouring cell T_j across interface S_j is dry ($D_j < \varepsilon^D$) and water elevation in T_i is lower than bed elevation

in T_j ($H_i < \eta_j$), i.e. dry topography emerges in cell T_j. Notice that in the dry cell free-surface elevation coincides with bed elevation ($H_j = \eta_j$), thus being greater than free-surface elevation in the wet cell ($H_j > H_i$). Ordinary numerical schemes for fully-wet domains will fail in this case: they would predict a momentum flux from T_j to T_i due to such gradient in free-surface elevation, regardless of the fact that T_j is dry [16, 29]. Thus, some modifications of the scheme are required. We treat this case by locally imposing a reflective boundary condition for the wet cell T_i, either inviscid (slip condition) or viscous (no-slip condition), located at interface S_j. When updating the solution in the dry cell in the corresponding edge, we simply assume a dry-dry interface.

The setup of the three threshold parameters ε^D, ε^F, ε^{1o} must be done empirically. Some guidelines based on our experience may be helpful in this respect.

- ε^D is known to control the celerity of wet-dry fronts, i.e. the speed of these fronts is better predicted setting this parameter to low values. Our approach is to set ε^D as some fraction of sediment diameter D_s, i.e. $\varepsilon^D \sim \frac{1}{10} \div 1\, D_s$.
- ε^F is crucial for the stability of the numerical method, especially in movable-bed computation. We propose the range $\varepsilon^F \sim 1 \div 5\, D_s$.
- For ε^{1o} we propose the range $\varepsilon^{1o} \sim 2 \div 10\, D_s$.

5.4 Parallelisation

In order to increase the GIAMT2D model performance in terms of computational efficiency, we have parallelised the code using the OpenMP[2] protocol on shared memory. OpenMP allows the programmer to create parallelised regions of the code by invoking specific directives. We have parallelised only the regions which have proved to be heavy in terms of the computational time required. For complete explanation of the method and its advantages and disadvantages we refer the reader to [32].

In this section we wish to test the increase in computational performance given by the parallelisation. We have performed a series of numerical runs varying the number of cores in order to quantify such increase. Tests have been performed on a domain of 33372 cells. Runs have been executed using 1 to 10 cores on a workstation having 12 cores: not all the available cores have been employed in order to avoid the interaction with the operating system, requiring at least one core to

[2] www.openmp.org

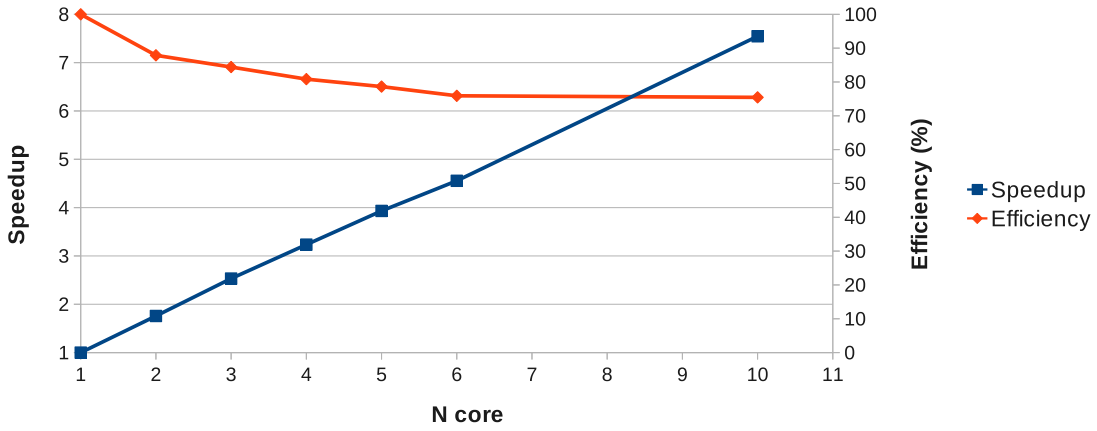

Figure 5.1: Speedup and efficiency of the parallelised GIAMT2D code.

run. To evaluate the parallelisation performance two different parameters have been used, namely Speedup and Efficiency, defined as

$$\text{Speedup}_{n_c} = \frac{\text{time}_1}{\text{time}_{n_c}} \quad , \quad \text{Efficiency}_{n_c}(\%) = 100\frac{\text{Speedup}_{n_c}}{n_c} \, , \tag{5.23}$$

where subscript n_c is the number of cores used and time is real simulation time. The theoretical optimal Speedup is equal to n_c, but in practical implementations this value is never reached. The Efficiency parameter represents the fraction cores effectively employed and assumes the value of 100% only in the ideal case (Speedup$= n_c$).

The results obtained in the test for these two parameters are given in Fig. 5.1. The Speedup is found almost linear with n_c even with 10 cores: hence the code implementation has a good scalability also over 10 cores. Efficiency is always over 70% and its curve is almost flat, i.e. the decrease in efficiency almost ceases between 6 and 10 cores.

5.5 Applications

In this section we apply the GIAMT2D numerical model to classical numerical test cases with fixed and movable bed in order to test our implementation and show the model capabilities. We compare the numerical solution of the model with analytical solutions and experimental results. The selected test cases aim at highlighting some characteristics of the model and motivate the choices we have made in designing the numerical solution strategy.

First, we devote some attention to assessing the performance of the numerical scheme in tests conducted in fully-wet domains. The first test, presented in section 5.5.1, concerning the evolution of a conical dune of sediments, highlights the advantage in accuracy connected to insertion of an upwind bias into the centred scheme and extension to second order, thus motivating the research in applied mathematics carried out in the previous chapters. The second test (section 5.5.2) consists of the modelling of sediment transport due to a hydraulic jump over erodible-bed: since the Froude critical state is crossed, by this test we prove the ability of the model in dealing with sediment transport in trans-critical flow conditions.
Then, we focus on assessing the performance of our original approach for the treatment of wetting-and-drying fronts presented in section 5.3. To this aim we first prove in section 5.5.3 that the model satisfies the *extended* C-property, i.e. exactly reproduces still water at rest in partially dry domains characterised by emerging topography. Then in section 5.5.4 we move to the simulation of dam-break fronts over dry domains with horizontal or sloping fixed bed in order to test the numerical model against analytical solutions for the front position and speed. Finally in section 5.5.5 we conduct numerical simulations of dam-break flow in a dry erodible channel and compare our numerical results with laboratory experimental data.

5.5.1 Evolution of a conical dune

In this two-dimensional test proposed by Hudson [60] we consider the evolution of a conical dune of sediments under almost uniform sub-critical flow. During the simulation, the sediment bump propagates downstream and evolves towards a star-shaped pattern expanding in time with a certain angle. For the spread angle an approximate analytical solution has been obtained by de Vriend [43]. Due to very long simulation (the timeout is set equal to $100\,h$) and very low sediment transport, this test highlights the performance of numerical schemes in terms of amount of spurious numerical diffusion in the solution of the morphodynamic problem: excessive numerical dissipation causes overestimation of the spread angle and underestimation of maximum dune height. Thus this case is commonly used in the accuracy assessment of hydro-morphodynamic solvers, see e.g. [27, 7, 6, 21]. In detail, here we aim at assessing the advantage in accuracy connected to the insertion of an upwind bias into the centred PRICE2-C scheme and to the extension to second-order accuracy.

We solve the problem in the square computational domain $[0,\,1000]\times[0,\,1000]\,m$. The initial condition for bed elevation is given by flat horizontal bed with a sediment bump

$$\eta(x,y,0)=\begin{cases} 0.1+\sin^2\left(\frac{\pi(x-300)}{200}\right)\sin^2\left(\frac{\pi(y-400)}{200}\right) & \text{if } x\in[300,500] \text{ and } y\in[400,600] \\ 0.1 & \text{otherwise} \end{cases} \quad (5.24)$$

The initial condition for the hydrodynamic variables is

$$H(x,y,0) = 10.1\,m \qquad , \qquad q_x(x,y,0) = 10\,m^2 s^{-1} \qquad , \qquad q_y(x,y,0) = 0\,, \tag{5.25}$$

representing uniform sub-critical flow in the x direction. Flow is inviscid in this test. At the upstream boundary $x = 0$ we prescribe constant discharge $q_x = 10\,m^2 s^{-1}$. A free outflow condition is set at the downstream boundary $x = 1000\,m$. At the side boundaries $y = 0$ and $y = 1000\,m$ we impose a reflective slip condition (inviscid bank). Timeout is $t = 100\,h$ and the CFL number is set equal to $CFL = 0.8$. The initial condition for the hydrodynamic variables in the movable-bed simulation is obtained by preliminarily running fixed-bed computations until steady flow state. We solve the problem on two different mesh setups applying the first-order (centred) PRICE2-C method, the (upwind-biased) UPRICE2-Cδ method and the second-order ADER extension of the latter scheme. Bedload transport is computed with the two-dimensional Grass formula (2.26).

The analytical solution for the spread angle [43] relies on the hypothesis of weak interaction between sediment layer and fluid. Recalling the Grass formula (2.26), having parameters $A_G \in [0,\,1]$ and $m_G \in [1,\,4]$, this is ensured by setting $A_G < 0.01$. The spread angle α^S is then given by

$$\alpha^S = \arctan\left(\frac{3\sqrt{3}\,(m_G - 1)}{9m_G - 1}\right)\,. \tag{5.26}$$

In our numerical simulations we choose $m_G = 3$, which gives analytical spread angle $\alpha^S = 21.787^\circ$, and $A_G = 0.00167$.

5.5.1.1 Evolution of a conical dune on a uniformly-spaced grid

In this test we wish to assess the advantage in accuracy connected to the use of a second-order extension of the scheme. Thus we present the numerical results of the upwind-biased UPRICE2-Cδ method both in the basic first-order configuration and in the second-order ADER extension. Being the solution smooth in all variables, we do not apply a TVD limiter in this case. Comparison of numerical results is carried out in terms of spread angle and maximum bed elevation. We discretise the domain with a triangular mesh of 32406 cells having approximately uniform size.

In Fig. 5.2 we represent the solution profiles of the first-order and second-order upwind-biased schemes (red and blue line respectively) sliced along the x axis at $y = 500\,m$. Improved resolution of the bed profile connected to second-order accuracy is evident. Maximum bed elevation is equal to $0.58\,m$ and $0.81\,m$ for the first and second-order scheme, respectively. The result of the second-order upwind-biased scheme in terms of maximum elevation is similar to that obtained by Castro *et al.*

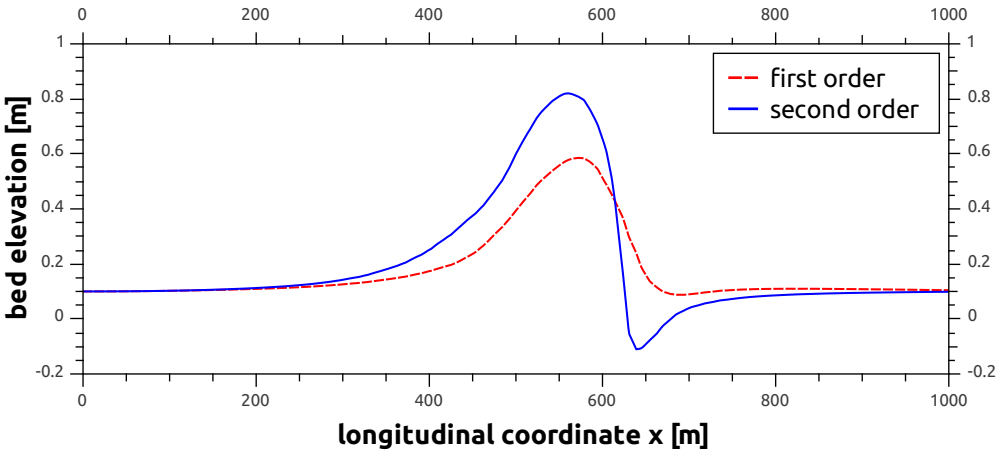

Figure 5.2: Evolution of a conical dune on a uniformly-spaced grid. Longitudinal bed profile obtained slicing the solution at $y = 500\,m$ and $t = 100\,h$ for the first-order UPRICE2-Cδ scheme and the second-order ADER UPRICE2-Cδ scheme.

[27] with a second-order upwind scheme.
Concerning the spread angle, the advantage of using a second-order scheme is not so evident: we measure $\alpha^S = 23.133°$ for the first-order scheme and $\alpha^S = 22.972°$ for the second order scheme, both quite close to the theoretical value $\alpha^S = 21.787°$.

5.5.1.2 Evolution of a conical dune on a variably-spaced grid

Here we perform another set of simulations in order to highlight the advantage given by insertion of an upwind bias into the scheme. From the numerical tests for the two-dimensional shallow water equations with fixed bed carried out in chapter 3 we know that genuinely centred methods may fail in the accurate description of the solution when a wide variety of local CFL numbers is produced. In the present test the flow field is almost homogeneous since the disturbance given by the dune is very weak. Thus, if uniformly-spaced grids are used, the distribution of CFL numbers is expected to be homogeneous as well.
We artificially introduce inhomogeneity by solving the problem on a variably-spaced grid, as we have done for the circular dam-break problem in sections 3.5.4 and 3.5.5. The mesh adopted, composed of 22554 triangular elements, is coarser in the upper side of the domain ($y > 500\,m$) than in the lower side. See Fig. 5.3, where a sketch of the domain discretisation across the centerline $y = 500\,m$ is depicted. The solution of any numerical method is expected to be more smeared on the coarse side of the domain ($y > 500\,m$) than on the other side. Thus, preserving symmetry in the transverse profile is challenging and highlights the ability of numerical methods in performing

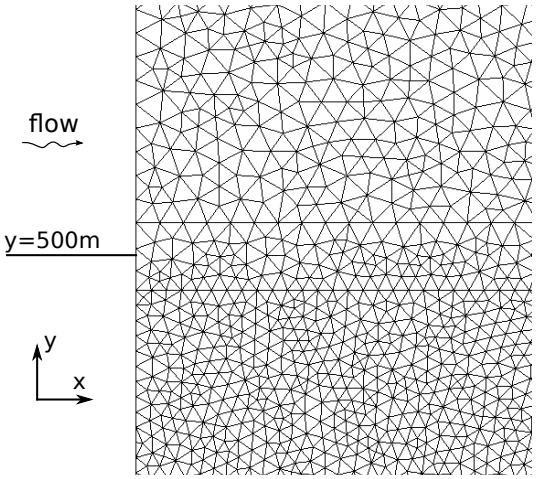

Figure 5.3: Evolution of a conical dune on a variably-spaced grid. Sketch of the computational mesh.

consistently over a wide range of CFL numbers. We solve the problem using second-order ADER extensions of the UPRICE2-Cδ (upwind-biased) method and of the PRICE2-C (centred) method.

The results of the two numerical methods are given in Fig. 5.4 in terms of transversal profiles, obtained slicing the solution along the y axis at $x = 565\,m$, and in Fig. 5.5 in terms of contourplots. In Fig 5.4 the profile of the centred method (blue dashed line) is found to be much more smeared than that of the upwind-biased method (red line) as a consequence of higher numerical dissipation. Moreover, as expected, the solution of the upwind-biased method is much more symmetrical, thus indicating more consistent performance over the entire range of stable CFL numbers.
From Fig. 5.5, where the solutions of the two methods are represented in terms of contourplots, analogous conclusions can be drawn. The solution of the centred method (left panel) is deeply influenced by inhomogeneity in the mesh concerning the degree of symmetry and sharpness, while these problems are not observed for the upwind-biased method.

We conclude that the increase in accuracy connected to the introduction of an upwind bias into the scheme, which had been carefully assessed in the case of conservative methods in chapter 3, is evident even in the solution of the shallow water- Exner model by means of path-conservative schemes.

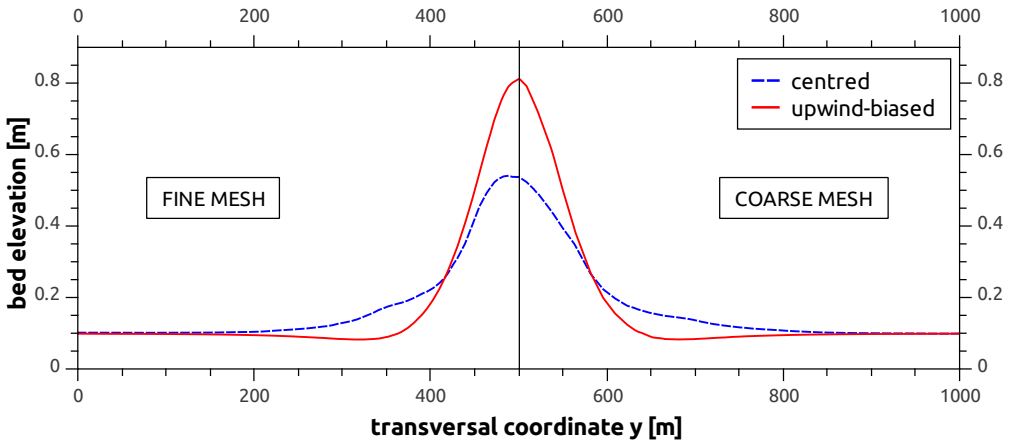

Figure 5.4: Evolution of a conical dune on a variably-spaced grid. Transversal bed profile sliced at $x = 565\,m$, obtained at $t = 100\,h$ for the second-order ADER PRICE2-C centred scheme (blue dashed line) and ADER UPRICE2-Cδ upwind-biased scheme (red line).

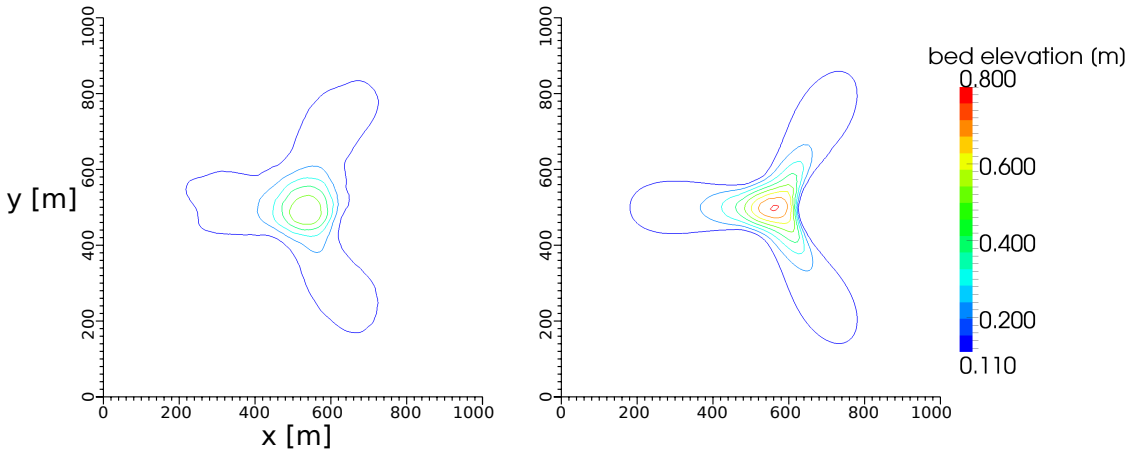

Figure 5.5: Evolution of a conical dune on a variably-spaced grid. The solution for bed elevation of two methods obtained at $t = 100\,h$ is represented in terms of contour lines. Left panel: second-order ADER PRICE2-C centred scheme. Right panel: second-order ADER UPRICE2-Cδ upwind-biased scheme.

5.5.2 Propagation of a sediment bore

This numerical tests concerns the reproduction of the experimental data obtained by Bellal *et al.* [5], concerning the propagation of a sediment bore in a steep erodible channel under the hydrodynamic action of a hydraulic jump. Their experimental setup consists of a 7.6 meter long, half meter wide glass-walled channel with adjustable slope, linked to a water recirculating system, an upstream sand feeder supplying dry sediments at a constant rate, and a downstream settling tank. The flume bed is covered by a layer of quasi-uniform coarse sand having mean diameter 1.65 *mm* and porosity 0.42. The authors indicate a bed roughness Manning coefficients $n = 0.0165\, m^{-\frac{1}{3}} s$.

As described by Bellal *et al.* [5], the experimental runs start from an equilibrium configuration, corresponding to uniform super-critical flow over movable bed fully developed throughout the channel length. Upstream sediment supply is in equilibrium with the hydrodynamic conditions. At time $t = 0$, the equilibrium condition is perturbed by the rapid raise of a submerged weir having height $H_w = 15.75\, cm$ at the downstream end of the flume, which raises the local water level up to $20.93\, cm$, thus imposing local sub-critical flow. This generates a hydraulic jump propagating in the upstream direction. The discontinuity in the hydraulic conditions induces longitudinal variations of sediment transport capacity, thus triggering riverbed morphological evolution. While the upstream-migrating hydraulic jump progressively slows down, a bore of sediments is generated and grows underneath. Subsequently, the bed further interacts with the hydraulic jump. Continuous deposition at the front of the sediment bore makes it grow in amplitude as it migrates downstream. Finally, the hydraulic jump noticeable at the free surface decreases in amplitude and progressively vanishes as the sediment front takes over the transition from supercritical to sub-critical flow.

Since Froude trans-critical flow conditions are encountered in correspondence to the hydraulic jump, the numerical reproduction of this experiment severely tests the ability of the numerical model in correctly addressing bedload transport in these conditions. In the case of GIAMT2D, as we have claimed, this is guaranteed by the adoption of a coupled formulation of the hydrodynamic and morphodynamic problem: here we aim at proving this statement. We solve the problem in the computational domain $[0, 7.6] \times [-0.25, 0.25]\ m$ discretised using 25002 triangular cells. The initial bed setup is given by the slope vector $\vec{S} = (0.0302, 0)$. Flow is in the x direction. At the lateral boundaries ($y = \pm 0.25\, m$) we use inviscid reflective conditions, while at the upstream boundary $x = 0$ we prescribe inlet flow with discharge $q = 0.024\, m^2 s^{-1}$ and at the downstream boundary $x = 7.6\, m$ we simulate the presence of the submerged weir by imposing water elevation $H(7.6, y, t) = 0.2093\, m$ above the reference level. We solve the problem setting timeout $t = 400\, s$ and imposing $CFL = 0.95$. Sediment transport is accounted by a calibrated Meyer-Peter and

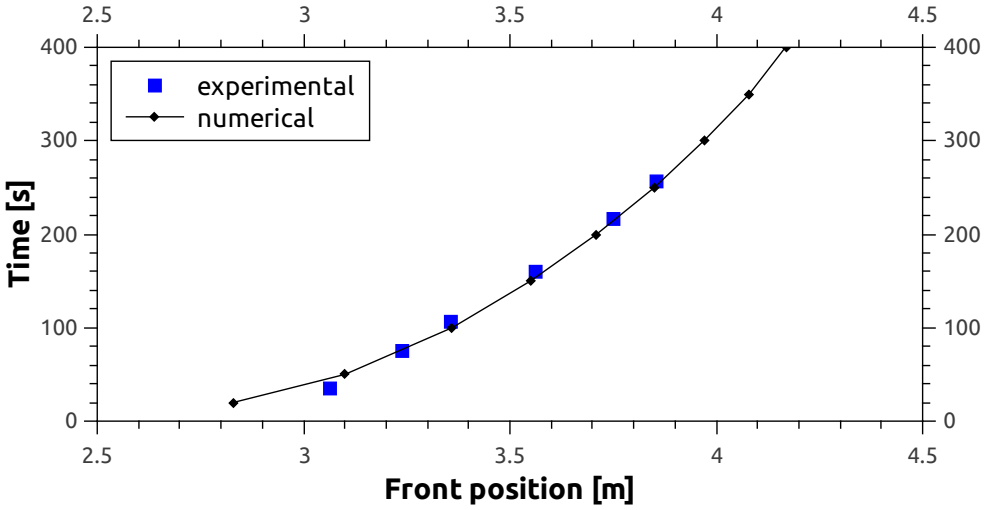

Figure 5.6: Propagation of a sediment bore. The position of the sediment bore front, obtained by numerical simulations using the GIAMT2D model embedding the second-order upwind-biased UPRICE2-Cδ method, is plotted in the $x-t$ plane (black symbols and full line) against the experimental results of Bellal *et al.* [5] (blue symbols).

Müller-type formula, reading

$$\Phi(\theta) = 5.5 \max\left((\theta - \theta_{cr}), 0\right)^{\frac{3}{2}} , \tag{5.27}$$

with $\theta_{cr} = 0.047$. Numerical simulations are conducted using the GIAMT2D model, embedding the second-order upwind-biased ADER-TVD UPRICE2-Cδ path-conservative scheme.

In our numerical results we track the position of the sediment bore front. This is reported in Fig. 5.6 , where the front position is plotted in the $x-t$ plane and compared with the experimental data of Bellal *et al.* [5]. Good agreement is found, confirming that the GIAMT2D model is able to predict the front advancement by correctly taking into account sediment transport in quasi-critical conditions.

In Fig. 5.7 in the top panel we plot a profile for water and bed elevation obtained slicing the numerical solution along the x axis at the channel centerline $y = 0$ at $t = 150\,s$. The sediment bore and hydraulic jump are located at $x \sim 3.6\,m$. On the bottom panel, we plot the corresponding Froude number, which is seen to decrease from $F_r \sim 1.4$ upstream of the hydraulic jump to $F_r \sim 0.2$ downstream of the jump. The trans-critical Froude region is thus crossed by the numerical profile. Cordier *et al.* [35] report examples of failure of uncoupled numerical methods in solving analogous

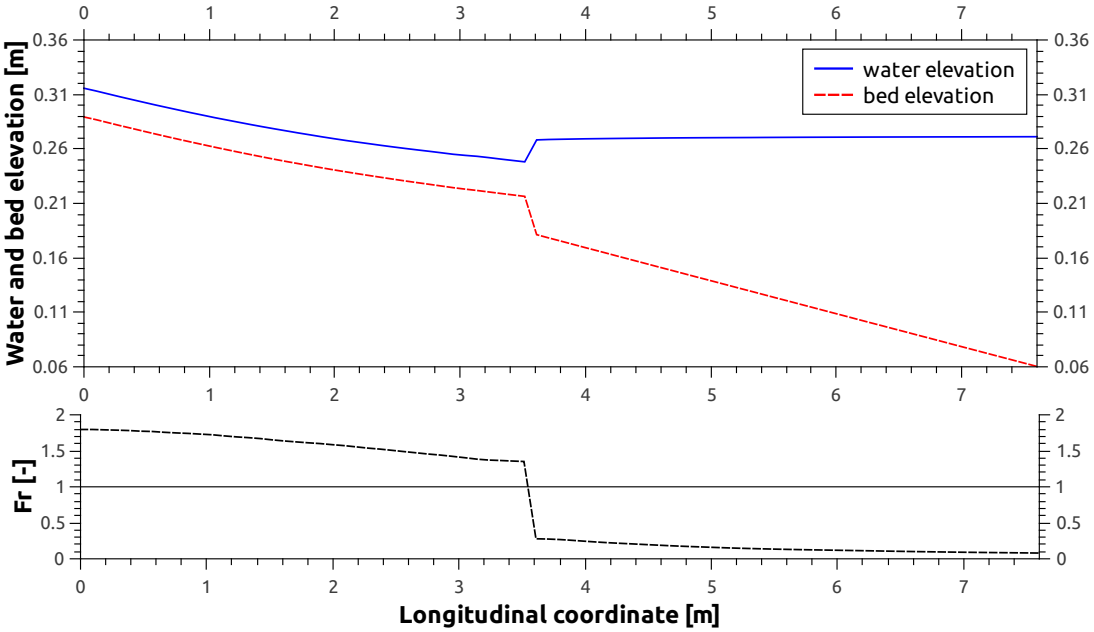

Figure 5.7: Propagation of a sediment bore. Top panel: profile of the numerical solution for free-surface and riverbed elevation, obtained slicing the solution of the GIAMT2D model at $t = 150\,s$ along the x axis at $y = 0$. Bottom panel: representation of the solution profile in terms of the Froude number.

hydro-morphodynamic problems conducted in trans-critical conditions: they observe that the numerical solution may get unstable, initially in the riverbed profile and then in the water elevation profile, until finally the code crashes. In the case depicted in Fig. 5.7, instead, the GIAMT2D model proves to be able to produce a correct, stable solution for the hydro-morphodynamic problem due to the coupled formulation adopted.

5.5.3 Verification of the extended C-property

Bermúdez and Vázquez Cendón [8] have introduced a well-balancedness condition called *conservation property* or *C-property*, which prevents the appearance of non-physical oscillations due to the discretisation of the source terms in the shallow water equations over sloping bed.

Definition 5.5.1: C-property. A numerical scheme is said to verify the C-property if it solves exactly the steady-state solutions corresponding to water at rest with flat free surface over non-movable bed $\eta(x,y)$, i.e.

$$q_x(x,y,t) = 0 \quad , \quad q_y(x,y,t) = 0 \quad , \quad H(x,y,t) = H_0 \qquad \forall x,y,t \tag{5.28}$$

where H_0 is a constant such that $H_0 > \max(\eta(x, y \in \Omega))$.

Verification of the C-property is a key issue for a numerical method. There is wide consensus in the literature that well-balancing with respect to water at rest crucially contributes to improving the solution of steady and transient flows in any condition.
Concerning schemes able to deal with wet-dry fronts, Castro *et al.* [28] generalise the above definition introducing the *extended* C-property:

Definition 5.5.2: Extended C-property. A numerical scheme is said to verify the extended C-property if it solves exactly the steady-state solutions of the form

$$\begin{cases} q_x(x,y,t) = q_y(x,y,t) = 0 \\ H(x,y,t) = \max(H_0, \eta(x,y)) \end{cases} \qquad \forall x, y, t \ . \tag{5.29}$$

Notice that the extended C-property (5.29) includes the C-property (5.28), which must hold in the wet portions of the domain. Moreover it poses some requirements over the solution in the dry portions of the domain and the treatment of wet-dry transitions. In particular, in these transition zones no spurious momentum flux shall be generated by the presence of dry emerging topography nearby a wet cell.

In this section we analytically and numerically test the ability of the GIAMT2D model to satisfy the extended C-property. Notice that the initial data (5.29), which also give the exact solution at any time, representing steady water, do not give rise to frictional terms nor to bedload discharge. Thus strictly speaking here we test only the solution of the the homogeneous problem (5.6) in fixed-bed computations.

5.5.3.1 Analytical verification of the C-property on wet domains

Canestrelli *et al.* [21] have analytically proved that high-order schemes based on PRICE-C solving the shallow water equations in wet domains are well-balanced as a consequence of the choice of a segment path (4.8) in the integration of the Roe matrix and of the adoption of a modified identity matrix (5.19) with $\varepsilon^b = 0$ for quiescent water. Here we report their analysis, showing that it applies to our upwind-biased scheme as well.

Consider the following initial condition in two neighbouring cells T_i and T_j, given in terms of the

vectors of conserved variables $\mathbf{Q}_i^n$ and $\mathbf{Q}_j^n$:

$$\mathbf{Q}_i^n = \begin{bmatrix} H_0 \\ 0 \\ 0 \\ \eta_i \end{bmatrix} \quad , \quad \mathbf{Q}_j^n = \begin{bmatrix} H_0 \\ 0 \\ 0 \\ \eta_j \end{bmatrix} , \tag{5.30}$$

with $\eta_i \neq \eta_j$ and $H_0 > \max(\eta_i, \eta_j)$. It follows from linearity of the integration path (4.8) that the Roe matrix and its square are

$$\mathbf{A}_{\boldsymbol{\Psi}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ g\bar{D}(n_{jx}+n_{jy}) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad , \quad \mathbf{A}_{\boldsymbol{\Psi}}^2 = \begin{bmatrix} g\bar{D}(n_{jx}+n_{jy}) & 0 & 0 & 0 \\ 0 & g\bar{D}(n_{jx}+n_{jy}) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} , \tag{5.31}$$

where $\bar{D}$ is given by

$$\bar{D} = H_0 - \int_0^1 \eta_i + s(\eta_j - \eta_i)\,\mathrm{d}s = H_0 - \frac{1}{2}(\eta_i + \eta_j) = \frac{1}{2}(D_i + D_j) \ . \tag{5.32}$$

Using the modified identity matrix (5.19) with $\varepsilon^b = 0$, it follows trivially from (4.20) for the upwind-biased method or from (4.12) for the centred method that $\mathbf{A}^-_{j+\frac{1}{2}}\left(\mathbf{Q}_j^n - \mathbf{Q}_i^n\right) = \mathbf{0}$. Therefore first-order schemes (4.5) satisfy the C-property. Concerning second-order schemes, the term $\int_{T_i} \underline{\underline{\mathbf{A}}}(\mathbf{Q}_i) \cdot \nabla \mathbf{Q}_i \,\mathrm{d}\vec{x}$ (4.30) clearly vanishes, thus preserving well-balancing even in this case.

5.5.3.2 Numerical verification of the extended C-property

Once theoretically verified the C-property in fully-wet domains, we empirically assess that the GIAMT2D model satisfies the extended C-property. This test case not only proves the well-balancedness of the numerical scheme in the wet case, but also it tests the strategy presented in section 5.3 for the treatment of wet-dry transitions with emerging topography.

We solve the problem in the square computational domain $[-5,5] \times [-5,5]$ *m* imposing reflective boundary conditions. We discretise the domain using a mesh having 860 triangular cells. We use $CFL = 0.95$ and set timeout $t = 60\,s$. Bed elevation is prescribed by

$$\eta(x,y) = \exp\left(-\frac{1}{4}\left(x^2 + y^2\right)\right) \tag{5.33}$$

Table 5.1: Numerical verification of the extended C-property. Error norms at time $t = 60\,s$ for $q = \sqrt{q_x^2 + q_y^2}$ of four schemes used in conjunction with the wetting-and-drying procedure presented in section 5.3.

	first-order			second-order ADER-TVD		
	L_∞	L_1	L_2	L_∞	L_1	L_2
UPRICE2-Cδ	0.000E+00	0.000E+00	0.000E+00	9.800E-16	2.203E-16	2.659E-16
PRICE2-C	0.000E+00	0.000E+00	0.000E+00	9.044E-16	2.438E-16	3.078E-16

and the initial condition for the flow variables is given by

$$\begin{cases} q_x(x,y,0) = q_y(x,y,0) = 0 \\ H(x,y,0) = \max(H_0, \eta(x,y)) \end{cases} \qquad \forall x,y\,, \tag{5.34}$$

with $H_0 = 0.5\,m$. Notice that (5.34) also gives the exact solution to this problem at any time. We solve this test applying first- and second-order schemes based on the centred PRICE2-C and on the upwind-biased UPRICE2-Cδ method in conjunction with the wetting-and-drying procedure presented in section 5.3.

In Tab. 5.1 we present the error norms obtained comparing the module of discharge $q = \sqrt{q_x^2 + q_y^2}$ computed by the numerical schemes with the exact solution $q = 0\ \forall x,y$. We observe that the results of first-order numerical methods are exact, while those of second-order methods are exact up to machine precision due to round-off errors in the TVD reconstruction procedure. These errors arise in the wet part of the domain. Comparing the results of centred and upwind schemes we do not find significant differences.

In Fig. 5.8 we plot the results for water and bed elevation of the second-order ADER-TVD UPRICE2-Cδ scheme. From Fig. 5.8 we obtain a visual confirmation that the model exactly reproduces flat water at rest over sloping bed in the wet portions of the domain (up to machine precision), correctly deals with the dry portions of the domain and exactly handles the transition between the two, avoiding prediction of spurious momentum fluxes at wet-dry interfaces with emerging topography.

5.5.4 Modelling of dam-break flow over fixed dry bed

In this section we numerically solve two inviscid dam-break problems over partially dry domains, resulting in the propagation of wet-dry fronts in fixed-bed channels. The aim of these tests is to assess the accuracy of the GIAMT2D model, based on the second-order ADER-TVD UPRICE2-Cδ scheme and on the wetting-and-drying procedure presented in section 5.3, in predicting the celerity of wet-dry fronts.

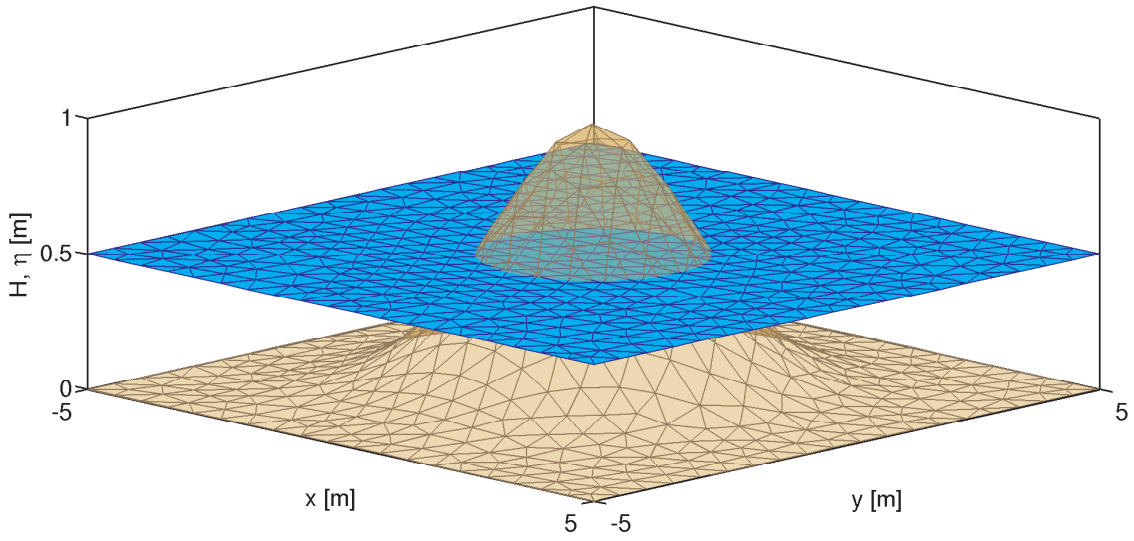

Figure 5.8: Numerical verification of the extended C-property. Solution of the second-order ADER-TVD UPRICE2-Cδ scheme used in conjunction with the wetting-and-drying procedure in section 5.3 for water elevation H (blue) and bed elevation η (brown) computed at $t = 60\,s$.

In general, a key requirement for numerical methods is that they should be able to describe fronts arising from initial discontinuities by correctly predicting their propagation speed in agreement with exact analytical solutions of the Riemann problem. Concerning dambreak problems in fully-wet domains, this requirement is commonly satisfied by the existing shock-capturing numerical methods. For conservative methods, convergence towards the exact weak solution, guaranteed by the Lax-Wendroff theorem [78], implies that the numerical solution satisfies this requirement. For path-conservative methods, even though theoretical assessment of convergence is still a matter of debate (see section 2.3.3), encouraging results concerning convergence to the exact weak solution in discontinuous problems have been experimentally obtained for the shallow water equations [22]. However, moving to the numerical modelling of front propagation over dry domains, experimental failure of convergence towards the exact solution due to under-prediction of front celerity is widely reported. Even the use of refined solutions of the Riemann problem with a dry state in upwind schemes guarantees only moderate improvement in this respect [29].

For the two tests presented in this section we consider a thirty-meter long fixed-bed channel whose axis is aligned with the x axis, having width equal to $W = 1\,m$. The computational domain is given in the range $[-15, 15] \times [0.5, 0.5]$ m and discretised with 16446 triangular cells. We apply inviscid reflective conditions at the side boundaries ($y = \pm 0.5\,m$) and transmissive conditions at the

upstream and downstream boundaries ($x = \pm 15\,m$). The channel bed is flat, i.e. it is characterised by constant slope vector $\vec{S} = (S_x, 0)$. Thus the bed setup is given by

$$\eta(x) = -S_x x\,. \tag{5.35}$$

Our two tests differ in the value of x-slope S_x. In the first test we use a channel with horizontal bed ($S_x = 0$), while in the second one we consider a sloping-bed channel characterised by $S_x = -\tan 3^\circ$.

A dam is initially placed in the cross-section corresponding to half channel length, namely $x = 0$. On the left side of this gate ($x < 0$), a layer of water at rest having horizontal free-surface is found, while on the other side the channel is dry. The initial condition is thus given by

$$H(x,y,0) = \begin{cases} 1\,m & \text{if } x < 0 \\ \eta(x) & \text{otherwise} \end{cases} \quad , \quad q_x(x,y,0) = q_y(x,y,0) = 0 \qquad \forall\, x,y\,. \tag{5.36}$$

Notice that such initial condition combined with the bed setup (5.35) gives a discontinuity by one meter in water depth and elevation in the dam cross-section:

$$H\left(0^-,y,0\right) = D_0 = 1\,m \quad , \quad H\left(0^+,y,0\right) = 0\,, \tag{5.37}$$

where with D_0 we denote initial water depth just upstream of the dam (at $x = 0^-$). At initial time we remove the dam and observe the propagation of waves arising from the discontinuity (5.37) in the positive and negative direction of the x axis. The waves in the right side of the domain ($x > 0$) propagate on initially dry bed. As reported by Castro *et al.* [29], the position $x_f(t)$ of the wet-dry front is known by exact analytical solutions of the dam-break problem in the form

$$x_f(t) = 2t\sqrt{gD_0\cos\gamma_D} - \frac{1}{2}gt^2\tan\gamma_D\,, \tag{5.38}$$

being $\gamma_D = -\arctan S_x$. The exact solution for the front position (5.38), together with the complete solution profile, was first computed by Ritter [100] in the horizontal-bed case ($\gamma_D = 0$). We use these solutions as benchmarks for our GIAMT2D model embedding the second-order ADER-TVD UPRICE2-Cδ method and our wetting-and-drying procedure described in section 5.3. We remark that use of a second-order method may improve accuracy in the computations in the wet part of the domain, but does not directly affect the solution of the front since on vanishing depth the scheme always reduces to first-order accuracy. We use $CFL = 0.95$ and set timeout $t = 2\,s$.

The first test is conducted in a horizontal channel. Starting from the initial condition (5.36) we observe the propagation of a rarefaction wave towards the left side ($x < 0$) and of a wet-dry

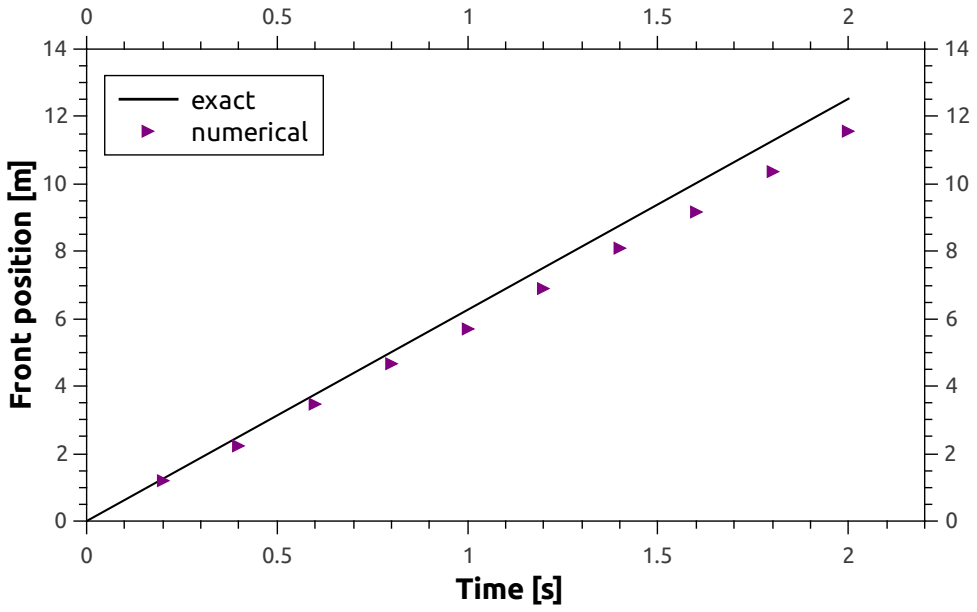

Figure 5.9: Modelling of dam-break flow over dry horizontal bed. The position of the wet-dry front of the numerical solution (symbols) is compared to the exact solution. The numerical solution is obtained with the GIAMT2D model embedding the ADER-TVD UPRICE2-Cδ scheme and the wetting-and-drying procedure described in section 5.3.

front in the right side of the domain ($x > 0$). The results are represented in Figs. 5.9 and 5.10.

In Fig. 5.9 we plot the position of the wet-dry front every $0.2\,s$ (symbols) and compare it with the exact solution (5.38) represented by a full line. It is seen that the front celerity is underestimated by the numerical model. This causes a moderate but evident lag between the exact and the numerical solution.

In Fig. 5.10 in the top panel we plot the profile of the numerical solution sliced on the x axis along the channel centerline ($y = 0$) at $t = 2\,s$. The numerical solution is compared to the exact analytical solution. Good overall agreement is found between the two solutions from this perspective. The numerical solution seems to be slightly inaccurate only in the description of the tail of the rarefaction at $x = -6\,m$, being locally excessively smeared.

However if we zoom the view focusing on the solution carried out at the wet-dry interface, as we do in the bottom panel of Fig. 5.10, we observe a discrepancy in the front position between the exact and numerical solution, causing inaccurate description of the profile on very shallow water depth. Analogous results are widely reported in the numerical literature. The results here presented are in line with those obtained by the numerical techniques presently available. For instance, the front position computed by GIAMT2D well compares to that obtained by Castro *et al.* [29] using a

refined solution of the Riemann problem with a dry state embedded in Roe's upwind Q-scheme.

Then, we consider the case of a sloping channel with $\gamma_D = 3^\circ$, i.e. characterised by adverse slope with respect to the wet-dry front propagation. In this test, while the wet-dry interface moves towards positive values of the x direction, water climbs over the steps generated by emerging topography. The treatment of emerging topography is expected to play a key role in determining the wave propagation speed. Thus by this test we assess the performance of our simple recipe presented in section 5.3, in which emerging topography is modelled imposing a local reflective condition at the wet-dry interface. As we have explained, in the GIAMT2D model we are allowed to choose whether to impose a no-slip or slip condition. In the present run we use a slip condition (inviscid bank), which is in agreement with the inviscid character of the considered dam-break problem and maximises the front celerity.
In Fig. 5.11 we plot the position of the wet-dry front obtained by numerical modelling every $0.2\,s$ (symbols) and the exact solution (5.38) (full line). Due to the influence of the emerging topography, this test is found more challenging than the previous one carried out over horizontal bed. The spread between the numerical and the exact solution is higher than in the previous case (Fig. 5.9) due to significant under-prediction of the front speed. Still however our results are similar to those of Castro *et al.* [29] obtained with a modified Roe's upwind Q-scheme.

Finally, we wish to give a comment on the role of the parameters of the wetting-and-drying procedure in determining the speed of fronts. The values adopted in the present tests are reported in Tab. 5.2. We have used the lowest possible values of parameters, in order to maximise the front speed while retaining numerical stability. Determining the optimal values in this respect is matter of empirical calibration, depending on the current test. From Tab. 5.2 we notice that the values in use are extremely low, thus being possibly unsafe from the point of view of numerical stability when moving to more complex, movable-bed simulations.
From our experience, ε^D, used for detecting the dry cells, plays the most important role in determining the front celerity in the flat-bed case: high values of this parameter result in significant slowing down of the wet-dry front. The values for the test conducted on a sloping bed are higher than those for the horizontal-bed test: this is an indication of the fact that the numerical simulation of front propagation over adverse slope is more challenging due to the presence of emerging topography modelled by a reflective condition. In the case of adversely-sloping bed, the reflective condition is found to significantly influence the front celerity: in the calibration process, at some stage further reduction of ε^D is found to be ineffective and only mesh refinement, reducing the height of emerging topography in the first dry cell, is able to improve the prediction.

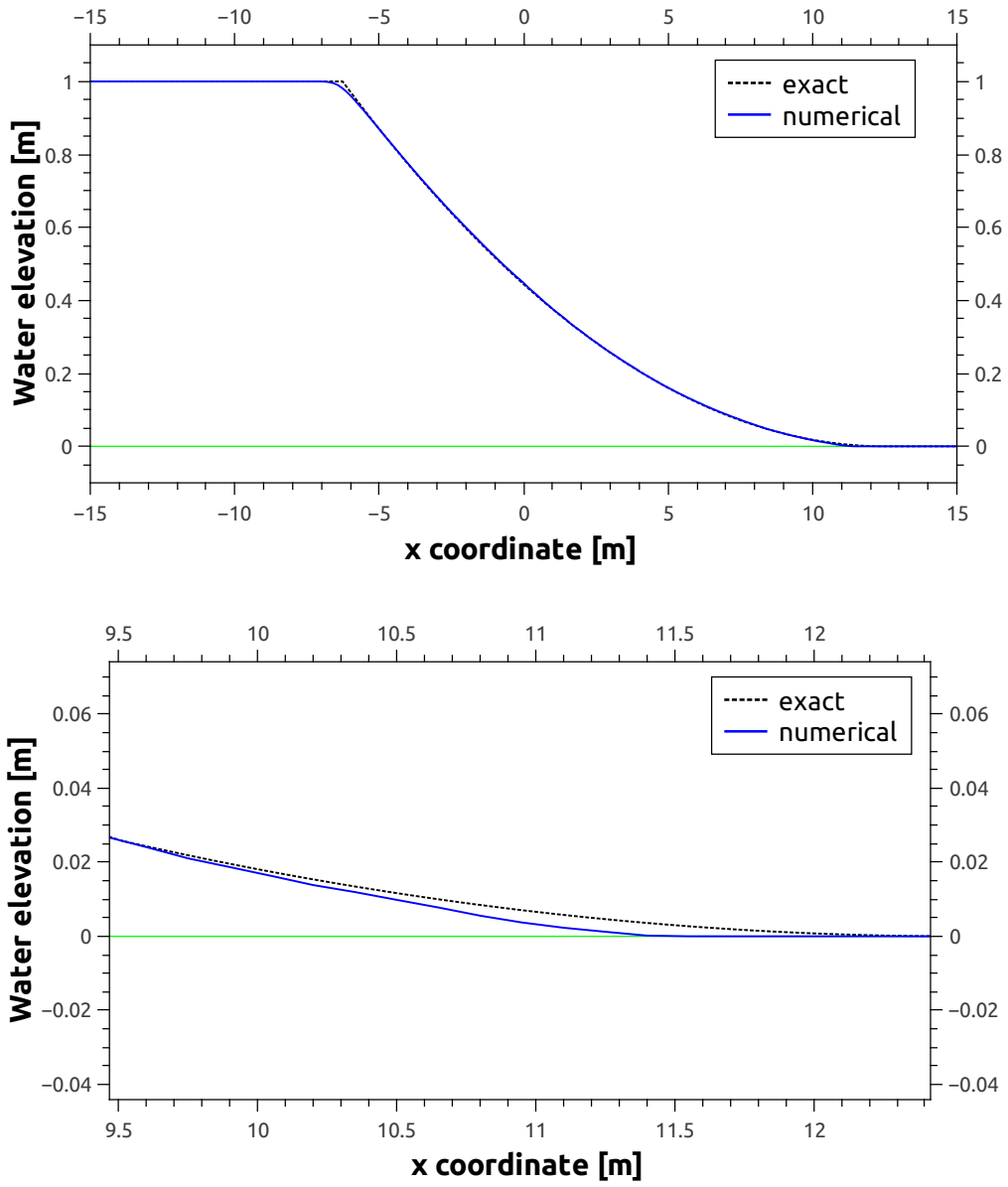

Figure 5.10: Modelling of dam-break flow over dry horizontal bed. The numerical solution for water elevation of the GIAMT2D model embedding the ADER-TVD UPRICE2-Cδ scheme and the wetting-and-drying procedure described in section 5.3 is sliced along the x axis at $y = 0$ and compared to the exact solution. Top panel: the full solution is presented. Bottom panel: zoom of the wet-dry front.

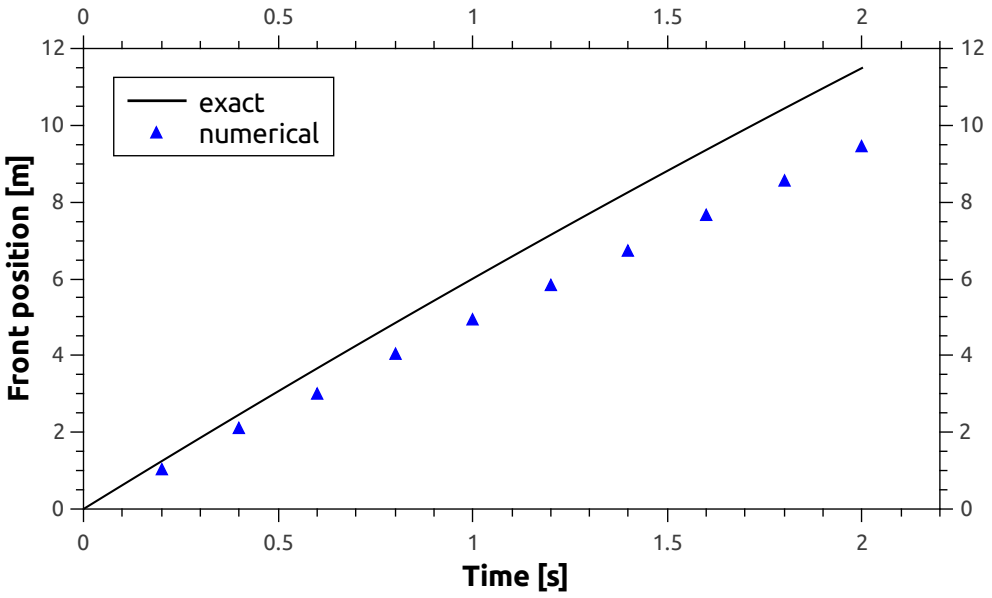

Figure 5.11: Modelling of dam-break flow over dry sloping bed. Bed slope $S_x = -\tan 3^\circ$ is adverse to the wet-dry front propagation direction. The position of the wet-dry front of the numerical solution (symbols) is compared to the exact solution. The numerical solution is obtained with the GIAMT2D model embedding the ADER-TVD UPRICE2-Cδ scheme and the wetting-and-drying procedure described in section 5.3.

Table 5.2: Parameters of the wetting-and-drying procedure used in the modelling of dam-break flow over dry bed.

	ε^D $[m]$	ε^F $[m]$	ε^{1o} $[m]$
horizontal bed ($S_x = 0$)	8.000E-06	1.000E-05	1.000E-05
sloping bed ($S_x = -\tan 3^\circ$)	4.000E-05	2.200E-04	2.200E-04

The tests presented in this section highlight the capabilities and limitations of the proposed simple technique for the treatment of wetting-and-drying. The recipes adopted prove to be robust (i.e. the code does not crash, even using very low values of the wetting-and-drying parameters) but the speed of fronts describing the transition between wet and dry portions of the domain is under-predicted. These results however are in line with those contained in the recent literature about numerical modelling of shallow flows in partially dry domains. Under-prediction of the front speed with respect to the exact solution emphasises the need of further fundamental numerical research on this topic. This however is out of the scope of the present thesis.

5.5.5 Simulation of dam-break flow in a movable-bed channel with a sudden enlargement

This section is devoted to the numerical reproduction of the laboratory experiments conducted at the Civil Engineering Laboratory of the Université Catholique de Louvain, Belgium, by Palumbo *et al.* [94], concerning dam-break flow over erodible bed. Experiments have been conducted in a flume with a non-symmetrical sudden enlargement as shown in Fig. 5.12. The flume bed is composed of uniform coarse sand having median diameter $D_s = 1.82\,mm$ and density $\rho_s = 2680\,kgm^{-3}$, deposited with bulk concentration 53%, and is initially covered by a 0.1 meter thick layer of fully saturated sand.
A dam is initially placed in the cross-section located at $x = 3\,m$. Upstream of this cross-section ($x < 3$), the flume is initially filled by $0.25\,m$ deep still water, while in the downstream region ($x > 3\,m$) the bed is dry. At $t = 0$ the dam is suddenly removed. The propagation of hydrodynamic waves associated to dam-break flow induces variations in riverbed elevation. As reported by Xia *et al.* [137], due to relatively coarse bed material and moderate velocity of the dam-break flow, the bed evolution can be assumed to be determined by bedload transport only.

We numerically reproduce this test case using the GIAMT2D model, including the second-order ADER-TVD UPRICE2-Cδ scheme and the wetting-and-drying procedure presented in section 5.3. We discretise the domain using an unstructured triangular mesh having 23070 cells. We apply reflective inviscid conditions at all side walls, while free outflow boundary conditions are used at the downstream outlet. We use the Manning friction law with $n = 0.0167\,m^{-\frac{1}{3}}s$. Timeout for the simulation is $t = 12\,s$ and the CFL number is $CFL = 0.95$. We compare the output of the GIAMT2D model with the time series for water elevation surveyed in six different locations $P1, ..., P6$ (see [94] for details) and with the final bed elevation measured along two cross-sections, namely $CS1$ (located at $x = 4.1\,m$) and $CS2$ ($x = 4.4\,m$). See Fig. 5.12, where a sketch of the domain and of the survey points and cross-sections is reported. This experimental dataset has been already used for testing other numerical models, see e.g. [94, 138, 63, 137].

In our simulations we have found that the final bed configuration is significantly influenced by the bedload transport formula employed. Therefore we have calibrated a Meyer-Peter and Müller-type formula (see Tab. 2.1) to achieve the best fitting of experimental data, obtaining

$$\Phi(\theta) = 12 \max\left((\theta - \theta_{cr}), 0\right)^{\frac{3}{2}} \tag{5.39}$$

with $\theta_{cr} = 0.047$. The flexibility of the UPRICE-Cδ scheme due to numerical integration of the Roe matrix (4.9) allows us to easily change the bedload transport formula in order to best fit the

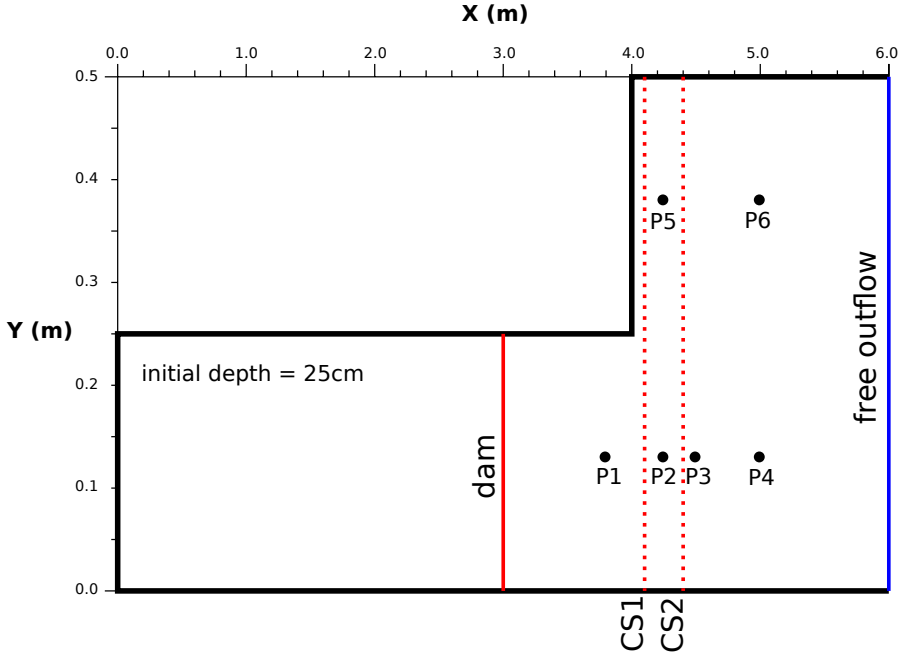

Figure 5.12: Simulation of dam-break flow in a movable-bed channel with a sudden enlargement. Sketch of the experimental setup and survey points.

experimental data.

In Fig. 5.13 we plot the numerical time series of water elevation in $P1, \ldots, P6$ (blue line) compared to the experimental results (symbols). Numerical results show good agreement with experimental data.

In Fig. 5.14 we show the comparison between the final experimental and numerical cross-sectional profiles of bed elevation obtained at time $t = 12\,s$. The profiles of the numerical simulation fairly agree with these surveyed experimentally: the numerical model is reasonably able to predict the location of scour and deposit regions along the cross-sections. In detail, in $CS1$ the position and extent of the scour is well predicted by the numerical model, while some mismatch between the computed and surveyed profile is instead found concerning the position of the right deposit region. Concerning $CS2$, the model is correctly able to locate the deposit region along the cross-section, but overestimates the peak value. These mistakes, which are found close to the bank, may be explained by locally three-dimensional flow induced by the sudden enlargement, which cannot be correctly handled in a depth-averaged approach [137], as well as by the difficulties of friction and sediment transport models in accurately describing the complex dynamics of dam-break front

waves. Anyway, our results well compare to those obtained by Xia *et al.* [137] in the numerical simulation of the same test case.

5.6 Conclusions

In this chapter we have built and tested a two-dimensional hydro-morphodynamic numerical model, referred to as GIAMT2D, solving the shallow water equations coupled with the Exner equation. The model has been designed in order to be able to reproduce two-dimensional flow and transport in gravel-bed rivers. Its main features are:

- adoption of a *coupled solution approach* in order to correctly deal with sediment transport in trans-critical flow conditions;
- use of an *upwind-biased centred scheme* (the second-order ADER-TVD UPRICE2-Cδ scheme developed in chapter 4) able to embed any bedload transport formula due to its centred nature, still retaining accuracy comparable to that of upwind methods;
- inclusion of an original *algorithm for wetting-and-drying* able to treat wet-dry fronts over fixed and dry bed in an adaptive and robust manner;
- implementation on *triangular unstructured* meshes in order to be able to deal with domains characterised by complex boundaries;
- parallelisation using the OpenMP protocol on shared memory for improving computational efficiency.

The performance of the model has been carefully assessed on a suite of classical tests with fixed and movable bed, using benchmark analytical solutions and experimental results. These tests allow us to claim that the GIAMT2D model correctly solves the equations describing flow over erodible bed and is able to simulate the flow and transport processes arising in the idealised physical configurations of laboratory conditions.

We shall now move to the second part of this thesis, concerning the applications of the GIAMT2D model aiming at reproducing the morphodynamic evolution of gravel-bed rivers. Our applications will be initially devoted to reproducing the dynamics of basic morphodynamic processes in rivers such as bar and bifurcation instability. By these applications we will show that the model not only satisfactorily performs in classical tests as we have seen in this chapter, but is also able to address the evolution of morphodynamic units (bars, bifurcations) in agreement with analytical morphodynamic theories and laboratory experiments. Finally, we will apply the GIAMT2D model

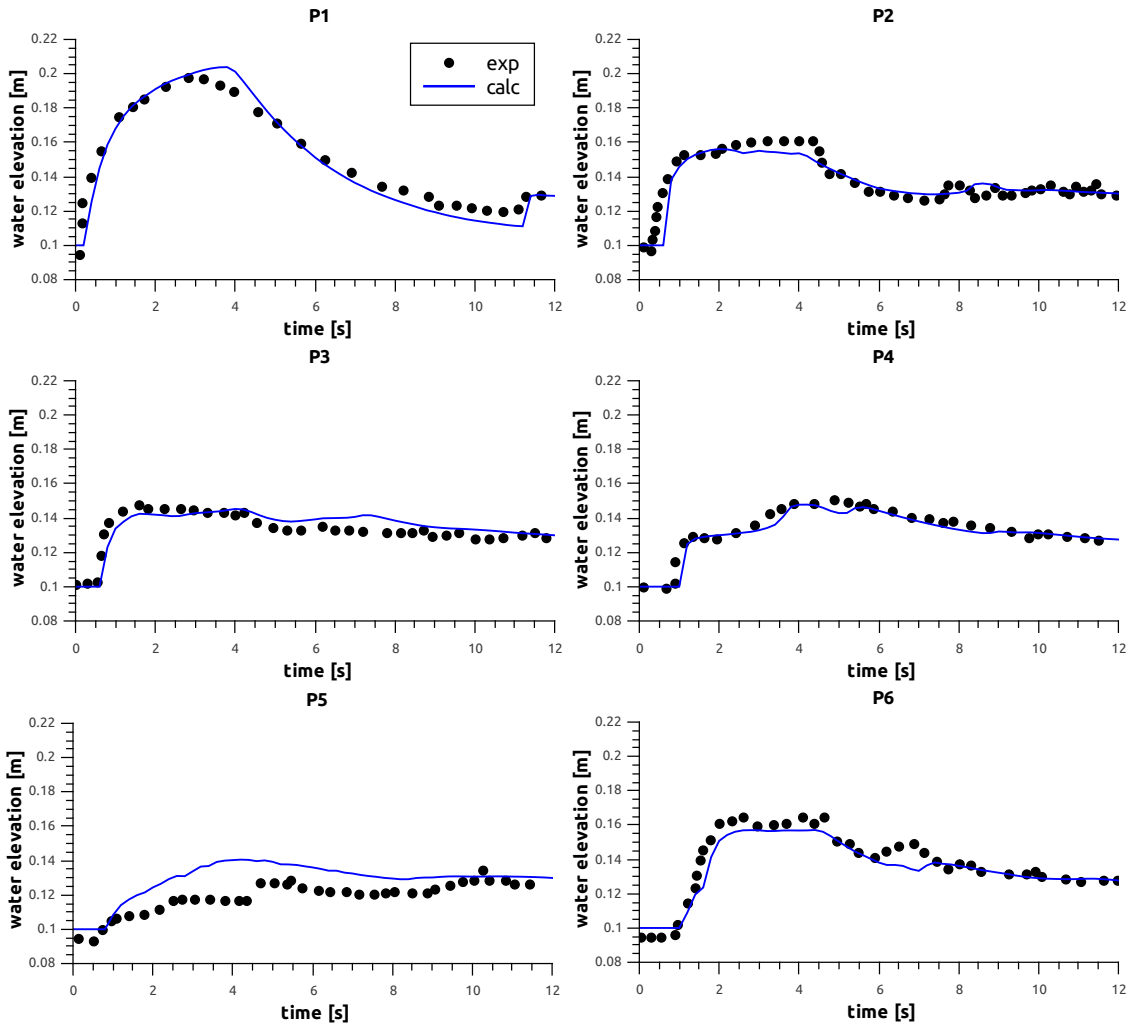

Figure 5.13: Simulation of dam-break flow in a movable-bed channel with a sudden enlargement. The time series for water elevation of the GIAMT2D model (blue line) is compared to the experimental data of Palumbo *et al.* [94] (symbols) collected in six survey points $P1, ..., P6$.

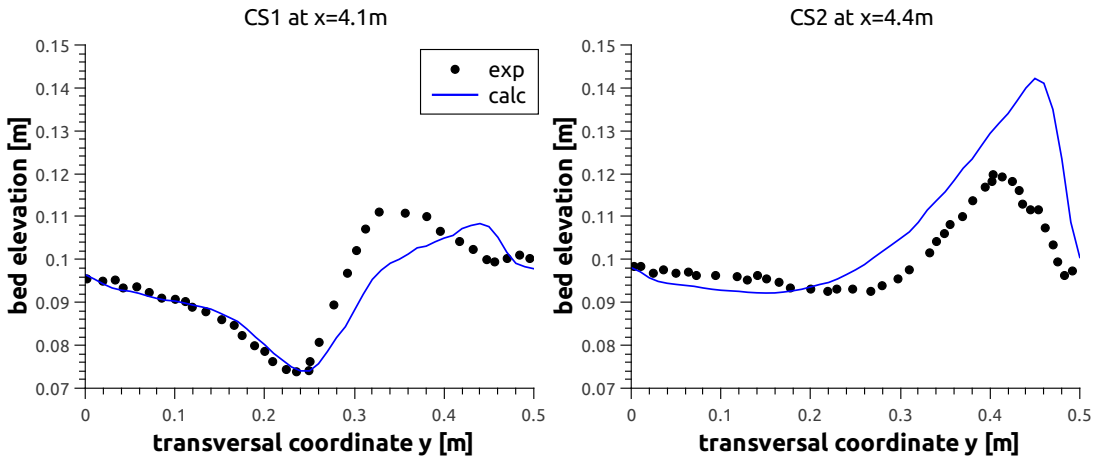

Figure 5.14: Simulation of dam-break flow in a movable-bed channel with a sudden enlargement. The bed elevation profiles computed by the GIAMT2D numerical model (blue line) at $t = 12\,s$ are compared with the experimental data of Palumbo *et al.* [94] (symbols) along the cross-sections $CS1$ and $CS2$.

in an original study concerning the interaction of free and forced bars in straight channels, together with theoretical speculations and experimental observations.

The next chapter is introductory to these applications: we will review some of the main achievements in river morphodynamics obtained by analytical an physical modelling, concerning the dynamics of free and forced bars and of channel bifurcations.

6 Bars and bifurcations in gravel-bed rivers: theoretical and experimental state of art

The second part of the present thesis, namely this chapter together with chapters 7 and 8, is devoted to applications of the GIAMT2D numerical model to the morphodynamics of gravel-bed rivers. In detail, in chapter 7 we will test the ability of GIAMT2D of reproducing the behaviour of basic morphodynamic processes such as bars and bifurcations, while in chapter 8 we will present novel achievements on the interaction of free and forced bars obtained by integrating the application of GIAMT2D with analytical theories and observations. The present chapter represents an introduction to the physical processes which will be subsequently investigated, namely the dynamics of free and forced bars and of channel bifurcations.

River bars are often regarded as the basic morphodynamic process at the scale of channel width: the morphological behaviour of movable bed rivers is essentially governed by the interaction between "free" bars, spontaneously developing in almost-straight channels as the result of inherent instability of the flow-erodible bed system, and "forced" bars produced by physical constraints, such as curvature, width variations, backwater effects and local persistent discontinuities in channel geometry [106]. In many cases it has been observed that, provided the forcing effect is large enough, free migrating bars are suppressed in favour of a steady bar pattern which in turn may affect the channel planform evolution. The case of planimetric forcing produced by channel curvature has received much attention [68, 127, 134] because of its inherent association with the development of river meanders. In the same way that curvature-forced bars can be seen as the fundamental units of meandering rivers, central bars developing in channels with spatial width variations can be seen as the fundamental bedform units of braided rivers [98]. The development and abandonment of channels in braided networks is governed by the dynamics of river bifurcations, to which much work has been recently devoted [14, 86, 125, 11]. In braided networks, evidence has been provided of intriguing legacies between the characteristic length of channels and the central anabranch width, in analogy to the consistent scaling of the length of free bars with channel width [62].

The above considerations highlight the role of bars and bifurcations as basic morphodynamic units of single and multi-thread natural rivers. The key idea beyond the second part of this thesis is that computational tools that aim at reproducing the morphodynamic evolution of real river systems should be carefully tested against the established dynamics of these basic features in simple and controlled conditions, such as those emerging from well-known and consolidated analytical theories, besides the outcomes of laboratory flume experiments. Thus, the dynamics of bars and bifurcations can be regarded as a fundamental morphodynamic benchmark for any numerical hydro-morphodynamic model.

Mathematical modelling of bar and bifurcation dynamics has been tackled by means of simplified analytical solutions. These solutions have the advantage of capturing the salient features and behaviour of the considered physical problems and identifying the controlling parameters. The main drawbacks are that simplified geometrical configurations are considered in order to keep the problem amenable to analytical treatment and that the interaction of different processes often cannot be taken into account. In addition to analytical modelling, physical modelling at laboratory scale allows the reproduction of morphodynamic processes in controlled conditions and provides quantitative benchmarks to test analytical theories.

In this chapter we will review the most remarkable and established results on the dynamics of bars and bifurcations in gravel-bed rivers. The chapter outlines as follows. In section 6.1 we introduce the main dimensionless parameters governing the evolution of morphodynamic units according to analytical theories. In section 6.2 we present the outcomes of analytical and physical models of free and forced bars. The linear bar theory, prescribing the conditions of formation of free bars in straight channels, and the weakly non-linear theory of Colombini *et al.* [33], giving the equilibrium amplitude of free bars, are reviewed in section 6.2.1. The linear theory of spatial bars, originally developed by De Vriend and Struiksma [44] and Olesen [92] and further refined by Seminara and Tubino [107] and Zolezzi and Seminara [141] is reviewed in section 6.2.2. The available theories concerning the interaction of free an forced bars in regular channel configurations are reviewed in section 6.2.3 together with related experimental results. The knowledge gap related to such interaction in the basic straight channel configuration will be the object of a novel application of the GIAMT2D model in chapter 8.
Section 6.3 concerns about channel bifurcations. The analytical model of Bolla Pittaluga *et al.* [14] is reviewed in section 6.3.1 and applied in section 6.3.2 . The analytical results of Bertoldi and Tubino [11] are presented in section 6.3.3 and compared to the outcomes of the analytical model.

6.1 Governing parameters

The mathematical modelling of bar morphodynamics in single-thread channels is commonly achieved through approximate solutions of the two-dimensional shallow water-Exner model (2.6), along with appropriate closure relations that relate sediment transport rate and friction to the local flow properties (see section 2.1.2). Concerning river bifurcations, more simplified models are usually adopted, in which flow in channels is assumed to be one-dimensional and a quasi two-dimensional model is employed only at the bifurcation node.
Analytical models are commonly based on a series of simplifying assumptions with the aim of focusing on the fundamental morphodynamic processes, while keeping the mathematical problem amenable to analytical treatment. It is often assumed that the system is fed by a representative constant value of flow discharge, the grain size is uniform and sediment transport mainly occurs as bedload at a rate that is in equilibrium with the stream transport capacity. Concerning bar models, the effects of flow unsteadiness [124], grain size heterogeneity [75] and suspended load [126, 52] have also been investigated, although their presentation is out of the scope of this chapter.

Mathematical models of physical processes can be conveniently cast in dimensionless form, which allows generality in the analysis and easy identification of the controlling parameters. The physical scale of morphodynamic processes is commonly given by the reference uniform flow, occurring in a straight channel carrying the same discharge and having the same channel width, longitudinal slope and sediment size. The key dimensionless parameters arising from the theories are:

- aspect ratio β (half-width-to-depth ratio) of the channel, defined as

$$\beta = \frac{W}{2D_0} , \tag{6.1}$$

  where W is channel width and D_0 is water depth of the reference uniform flow;

- dimensionless grain size d_s, expressing relative bed roughness

$$d_s = \frac{D_s}{D_0} , \tag{6.2}$$

  D_s being sediment diameter;

- the Shields parameter θ, reading

$$\theta = \frac{\tau}{(\rho_s - \rho_w) g D_s} , \tag{6.3}$$

  which represents a dimensionless measure of relative sediment mobility, expressed through

the near-bed shear stress τ and the relative sediment submerged weight, ρ_s and ρ_w being sediment and water density and g the acceleration due to gravity. Equation (6.3) specialises as (2.19) if the Manning law is adopted for the evaluation of τ, while adoption of the Chézy law gives (2.20). Notice that in the latter case θ may weakly depend on the inverse of d_s (6.2) through (2.13).

The three above parameters apply in the description of the morphodynamic evolution of channels at any scale, from laboratory flumes to natural channels. Once these parameters have been set, the reference uniform flow is defined through the closure relations in section 2.1.2.

6.2 Free and forced bars in straight channels

The classification of free and forced bars has a precise physical and mathematical meaning [106]. Free bars are repetitive sequences of scour holes and depositional diagonal fronts, with vertical scale of the order of the flow depth. They develop spontaneously and migrate in almost-straight channels with constant width, provided the width-to-depth ratio exceeds a threshold value which depends on the flow and sediment characteristics. Free bars display a variety of possible topographic patterns depending upon the most unstable transverse mode (alternate, central or multiple-row bars). Theoretical analyses and experimental results suggest that, when sediments are mainly transported as bedload, the alternate bar configuration is dominant for the typical channel aspect ratio values of single-thread alluvial gravel-bed rivers.
Forced bars develop in response to an external forcing and keep fixed with respect to the planform itself. Forcing effects can be both distributed along river systems and localised in space. To the former category belong longitudinal variations of channel curvature [13] or width [97, 98], whereas local perturbations are typically represented by groynes, bridge piers or other local river structures, as well as by any other abrupt planform change in space, like confluences or bifurcations. Two typical examples of localised disturbance are the transition between a bend of constant radius and a straight reach [114], causing a discontinuity in channel curvature, and a bifurcation located downstream of a straight channel [11]. Localised forcing effects are able to produce steady bars also in straight river reaches [93]. Recent numerical and experimental research [40] suggests that the presence of steady bars in straight channels can be justified even in the absence of any detectable planform discontinuity.

6.2.1 Free bar formation: a perturbative analysis

The formation of free bars is connected with intrinsic altimetric instability of the system composed by water flowing over an erodible bed. Linear stability of free bars has been increasingly

investigated in the seventies starting from the work of Callander [20]. Following Colombini *et al.* [33] in this section we present a brief overview of the analytical theories of free bars.
These theories mostly refer to a straight and regular channel with rectangular cross-section having width W. For sake of simplicity in the presentation we assume the x axis to be oriented in the direction given by the longitudinal coordinate of the channel. Positive values of x are in the downstream direction. The y axis is transverse to the considered channel. The channel centreline is associated to $y = 0$, the left bank to $y = \frac{W}{2}$, the right bank to $y = -\frac{W}{2}$. The initial bed configuration $\eta_0(x)$ is flat, i.e. the associated bed slope vector $\vec{S} = (S_x, 0)$, having magnitude S, is constant everywhere.
The system evolution in terms of variables H, q_x, q_y, η is governed by the two-dimensional shallow water-Exner model (2.6) together with appropriate closure relations for friction and bedload transport (see section 2.1.2), provided appropriate boundary conditions for water and sediment discharge have been set. System (2.6) admits equilibrium solutions. The simplest of these solution is uniform flow, which we denote with

$$\begin{cases} H(x,y,t) = H_0(x) = D_0 + \eta_0(x) \\ q_x(x,y,t) = q_{x0} \\ q_y(x,y,t) = 0 \\ \eta(x,y,t) = \eta_0(x) \end{cases} . \tag{6.4}$$

D_0 being water depth associated to q_{x0} in uniform flow condition. Notice that the values D_0 and q_{x0} are linked together by the friction formula (2.10) or (2.12) where we impose $S_f = S$. We assume uniform flow (6.4) as initial condition.
Free bar development is typically studied by means of linear stability analyses. The solutions of system (2.6) can be found by expanding each unknown as follows:

$$\begin{cases} H(x,y,t) = H_0 + H_1(x,y,t) \\ q_x(x,y,t) = q_{x0} + q_{x1}(x,y,t) \\ q_y(x,y,t) = 0 + q_{y1}(x,y,t) \\ \eta(x,y,t) = \eta_0(x) + \eta_1(x,y,t) \end{cases} , \tag{6.5}$$

$H_1(x,y,t)$, $q_{x1}(x,y,t)$, $q_{y1}(x,y,t)$, $\eta_1(x,y,t)$ being the perturbation of the uniform flow steady-state solution (6.4) for each variable. The assumed indefinite length and the problem linearity suggests to impose a time-dependent periodic perturbation having the form

$$\eta_1(x,y,t) = \varepsilon^P \exp(\Omega t + I\lambda^* x) \sin\left(\frac{m\pi y}{W}\right) + \text{c. c.} , \tag{6.6}$$

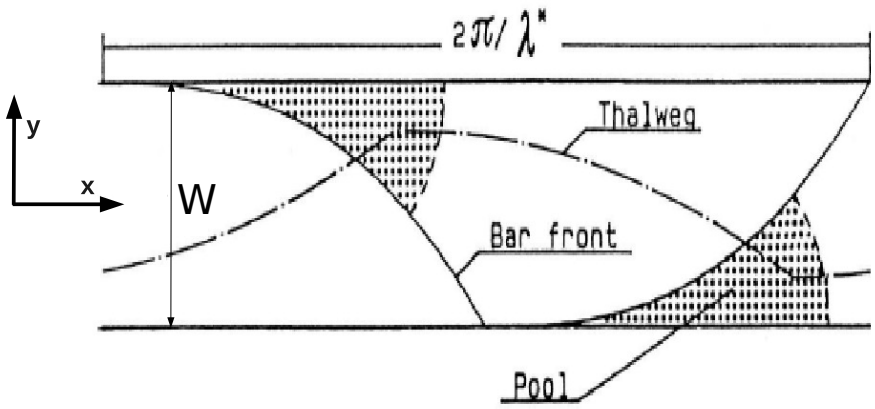

Figure 6.1: Definitions and notation for the analytical free bar theory.

while analogous structure is assumed for the wave perturbation of the other variables. In (6.6) ε^P is wave amplitude, assumed to be negligible compared to D_0 in order to perform a linear analysis, I is the imaginary unit, m is the bar mode, expressing the number of bar rows in the channel ($m = 1$: alternate bars; $m = 2$: central bars ...), c. c. stands for complex conjugate, Ω is the signal growth rate and λ^* is the dimensional wavenumber of the perturbation, given by

$$\lambda^* = \frac{2\pi}{L} , \tag{6.7}$$

where L is the bar wavelength. Fig. 6.1 illustrates these definitions and the notation.
The wavenumber λ^* can be non-dimensionalised in the form

$$\lambda = \lambda^* \frac{W}{2} = \frac{\pi W}{L} . \tag{6.8}$$

As a general rule, both the growth rate Ω and the wavenumber λ are complex:

$$\Omega = \Omega_r + I\Omega_i \qquad , \qquad \lambda = \lambda_r + I\lambda_i . \tag{6.9}$$

The theory of free bars assumes λ to be real, i.e. $\lambda = \lambda_r$ and $\lambda_i = 0$, while this assumption will be relaxed in the theory of spatial bars (see section 6.2.2). In (6.9) the real part of growth rate Ω_r represents the amplification rate in time, whereas the imaginary part Ω_i carries information about the signal migration direction.
Both Ω_r and Ω_i are functions of the undisturbed flow parameters θ, β and d_s computed in uniform flow conditions (6.4), of the bar wavelength λ and lateral mode m. Moreover they depend on the bedload transport formula adopted. This is obtained by inserting the perturbation relation (6.6) for

all variables into the system of governing equations (2.6) and linearising. A homogeneous algebraic system is found, which poses a solvability condition that results in the eigenrelation

$$\Omega = \Omega\left(\theta, \beta, d_s, \lambda, m\right) \ . \tag{6.10}$$

Time-dependent bar amplitude $A\left(t\right)$ is an exponential function of Ω_r, having the form

$$A\left(t\right) = \exp\left(\Omega_r t\right) \ . \tag{6.11}$$

Thus from (6.11) we identify two distinct asymptotic regime configurations for riverbed elevation:

- if $\Omega_r < 0$ then $A\left(t\right) \to 0$, i.e. we find a stable regime in which the initial bar wave is damped and uniform flow (6.4) represents a stable solution;
- otherwise we find an unstable regime in which the wave amplitude tends to grow, which physically corresponds to the condition for bar formation.

In the latter case we obtain a bar having spatial periodic structure, growing in time and migrating in space along the x direction. Concerning migration, the sign of the imaginary part of the growth rate coefficient (6.9) Ω_i discriminates between two cases:

- if $\Omega_i < 0$, the wave signal migrates downstream;
- otherwise, the wave signal migrates *upstream*.

The latter behaviour may seem unrealistic concerning finite-amplitude bars of real rivers; we remark however that it results from the adoption of a linearised approach, which strictly applies only for perturbations of infinitesimal amplitude.

In Fig. 6.2 we give a visual representation of the outcomes of the linear theory. Here we plot the predicted growth rate Ω_r (red line) and migration rate Ω_i (blue line) obtained as functions of λ for a given set of parameters $\theta = 0.06$, $d_s = 0.02$ and $\beta = 10$. The bedload formula of Wong and Parker [135] (see Tab. 2.1) has been adopted. From the chart we observe that a central range of wavenumbers exists for which the growth rate Ω_r is positive, i.e. perturbations of riverbed elevation tend to amplify. Even though most of this region is characterised by negative values of Ω_i (downstream-propagating bars) a small subset with $\Omega_i > 0$ exists, giving unstable and upstream-migrating bars. However, we notice that the highest value of Ω_r corresponds to downstream-migrating bars, thus suggesting that the instability process tends to select downstream-migrating wavenumbers. Bars having excessive or too small wavenumber, falling outside the unstable range of λ, tend to be suppressed.

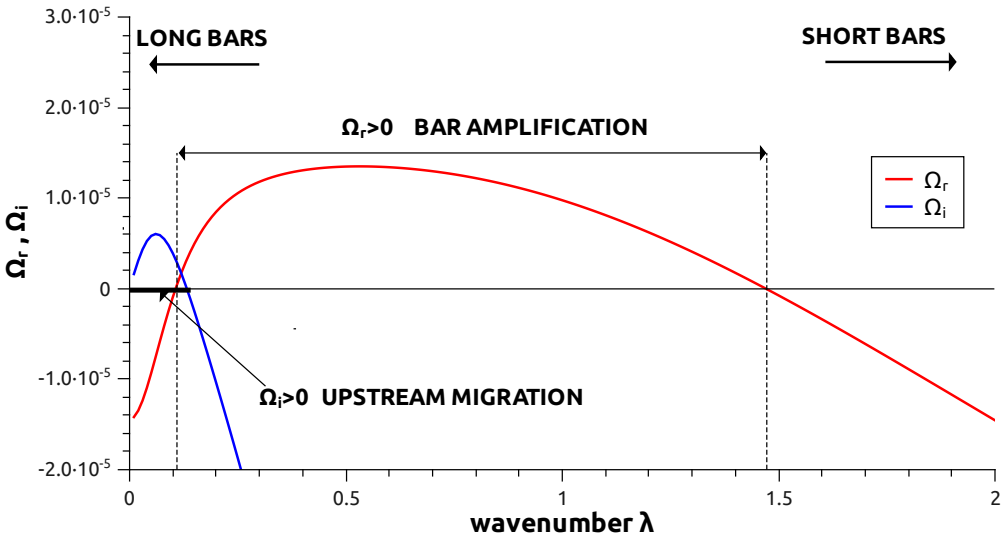

Figure 6.2: The real part Ω_r and the imaginary part Ω_i of the amplification rate parameter (6.11) are plotted as functions of wavenumber (6.8), with $\theta = 0.06$, $d_s = 0.02$ and $\beta = 10$, adopting the bedload transport formula of Wong and Parker [135].

The key controlling parameter for bar instability emerges when considering the $\lambda - \beta$ plane, plotting the curves associated to $\Omega_r = 0$ and $\Omega_i = 0$ respectively, still using constant values of d_s and θ. These two curves, which are commonly regarded as the *marginal curves*, represent the loci of states characterised by neutral stability, i.e. the hydrodynamic and transport conditions for which bars neither amplify nor get damped ($\Omega_r = 0$) and do not propagate ($\Omega_i = 0$). In Fig. 6.3 we depict the marginal curves associated to $\theta = 0.1$ and $d_s = 0.01$ computed using the bedload transport formula of Wong and Parker [135]. The marginal curve associated to $\Omega_r = 0$ is represented by a black full line. The region above such curve represents the states expressing intrinsic instability of the system, giving bar amplification, while the region below corresponds to the states for which any disturbance of uniform flow is damped. The minimum β value of the black marginal curve in Fig. 6.3 corresponds to a critical value of aspect ratio β_{cr}, i.e. *the minimum aspect ratio for which amplification is possible*. β_{cr} depends on d_s, θ and on the bedload transport formula adopted. Physically, this parameter carries information of crucial importance: the theory predicts that alternate bars can develop only in channels that are wide and shallow enough in formative conditions, i.e. only if $\beta > \beta_{cr}(\theta, d_s)$. With λ_{cr} we denote the wavenumber corresponding to β_{cr} on the marginal curve. Notice that, while a precise value β_{cr} of aspect ratio exists above which free bars amplify, the amplification mechanism does not clearly indicate the bar wavenumber likely to be selected, since the marginal curve is rather smooth.

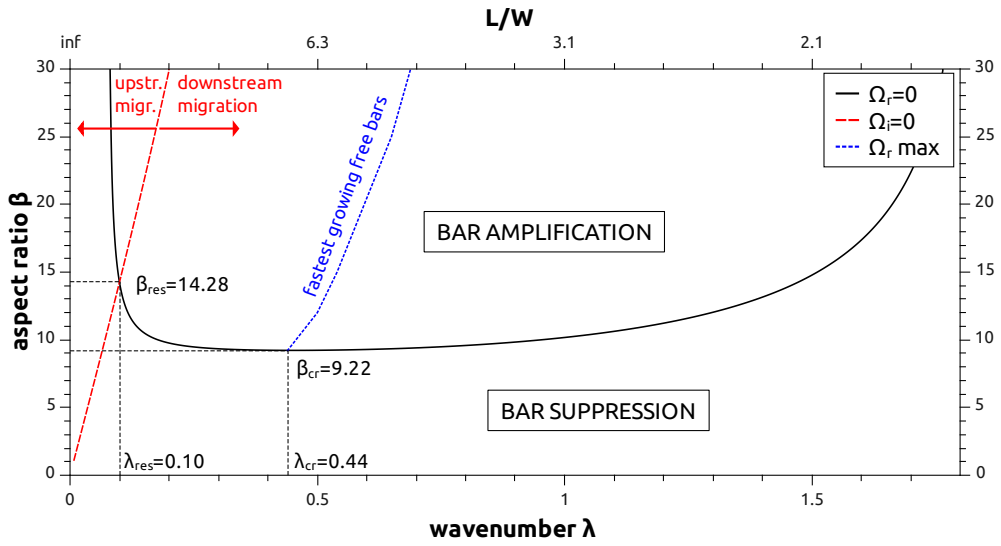

Figure 6.3: Marginal curves obtained with $\theta = 0.1$ and $d_s = 0.01$ using the bedload transport formula of Wong and Parker [135].

The blue dashed line in Fig. 6.3 represents the wavenumber associated to the maximum value of Ω_r for given β, i.e. the signal which is more likely amplified for any value of β. This curve intersects the marginal curve ($\Omega_r = 0$) in the critical point $(\lambda_{cr}, \beta_{cr})$. The critical wavenumber is usually found in the range $\lambda_{cr} \sim 0.35 \div 0.45$, which in terms of dimensional variables gives L in the range of $7 \div 9\,W$. This is consistent with the values of bar length reported from experimental observations, which have consistently provided a validation of the linear theory.
The marginal curve associated to $\Omega_i = 0$ is depicted in red in Fig. 6.3. On the right of this curve the states associated to downstream migration of free bars are represented, while on the left the states associated to upstream migration are found.
The value of β for which $\Omega_r = 0$ and $\Omega_i = 0$, i.e. the aspect ratio associated to the intersection point of the marginal curves, identifies the *resonant* value of aspect ratio β_{res}: above this threshold linearly unstable signals may migrate upstream (provided λ is sufficiently small), while below β_{res} only downstream migration occurs. β_{res} depends on θ, d_s and on the bedload transport formula adopted. The threshold β_{res} is named resonant because it corresponds to theoretically infinite growth rate of sinusoidal channels having $\lambda = \lambda_{res}$ [13]. Such concept will be further explored in section 6.2.2.

The linear theory focuses on the conditions of incipient formation of free bars due to the linear base flow instability, but is not able to provide information on the growth of finite-amplitude free bars. This topic has been investigated experimentally by Ikeda [65] and by Colombini *et al.*

[33] by means of a weakly non-linear perturbation analysis. Colombini *et al.* [33] found that the exponential bar growth predicted by (6.11) is damped by non-linear effects, eventually leading to an equilibrium configuration in time. Equilibrium bar amplitude A_E is then given by

$$A_E = D_0 \left\{ b_1 \left(\frac{\beta - \beta_{cr}}{\beta_{cr}} \right)^{\frac{1}{2}} + b_2 \left(\frac{\beta - \beta_{cr}}{\beta_{cr}} \right) \right\} \qquad \text{for} \qquad \beta < 2\beta_{cr} \,, \tag{6.12}$$

b_1 and b_2 being functions of θ and d_s. From equation (6.12) we observe that bar equilibrium amplitude (from maximum scour to maximum deposition) scales with uniform flow depth D_0. Moreover, equation (6.12) highlights the role of the relative distance of aspect ratio from the critical value $\frac{\beta-\beta_{cr}}{\beta_{cr}}$ as a measure of the strength of free bars: for increasing values of this parameter, the amplitude of free bars increases. This result will be found again in section 6.2.3 concerning the coexistence of free and forced bars. We will use the same concept in chapter 8 in order to interpret our novel numerical results on the interaction between free and forced bars.

6.2.2 Forced bars

On the basis of theoretical analyses [93, 107] we recall the distinctive features of forced bars in straight channels:

- *forced origin:* unlike free bars, forced bars are generated by a forcing effect, often represented by localised or distributed variation in channel planform;
- *steadiness:* forced bars do not migrate, but instead they keep adjoint to the forcing disturbance;
- *length:* steady bars are typically longer than free bars ($L \sim 15\,W$);
- *spatial damping:* spatial bars due to localised planform disturbances decay with increasing distance from the forcing itself.

Steadiness of forced bars implies that they can be theoretically investigated by means of a linear pertubative approach in asymptotic steady conditions, i.e. neglecting time derivatives in the shallow water-Exner model (2.6). Consider a straight channel as described in section 6.2.1. Zolezzi and Seminara [141] use a perturbative approach analogous to that already presented for free bars (6.5)-(6.6). However, here the perturbation of infinitesimal amplitude has only spatial structure:

$$\eta = \eta_0(x) + \eta_1(x,y) \,. \tag{6.13}$$

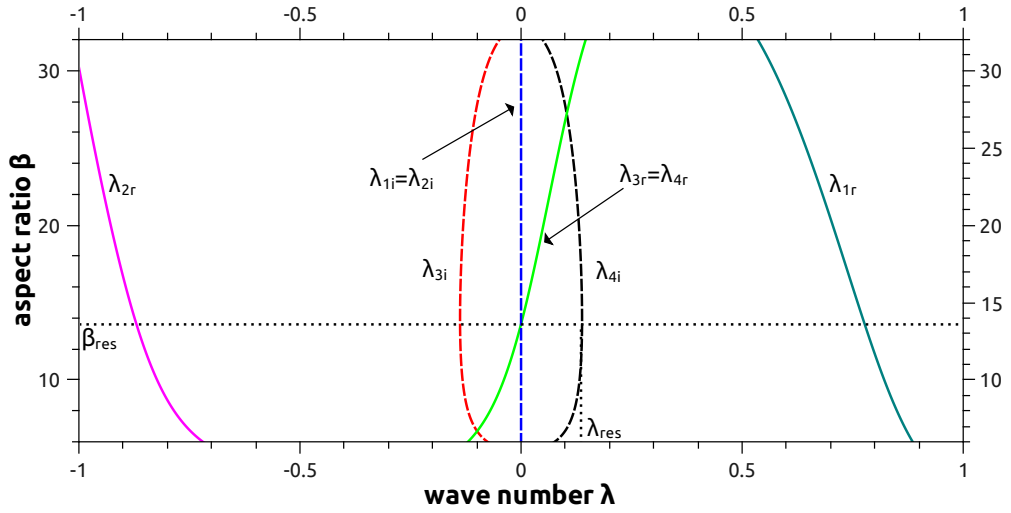

Figure 6.4: Characteristic wavenumbers $\lambda_1, ..., \lambda_4$ of spatial bars versus aspect ratio β, obtained with $\theta = 0.08$ and $d_s = 0.02$ using the bedload transport formula of Wong and Parker [135].

In (6.13) $\eta_0(x)$ is initial bed elevation in uniform flow (6.4) and $\eta_1(x,y)$ is the spatial perturbation, reading

$$\eta_1(x,y) = \varepsilon^P \exp(\lambda^* x) \sin\left(\frac{m\pi y}{W}\right) + \text{c. c.}\,. \tag{6.14}$$

In equation (6.14) the amplitude ε^P is assumed to be negligible with respect to uniform flow depth D_0 as requested by the linear approach, while the perturbation wavenumber λ^*, made dimensionless using (6.8), is assumed as complex:

$$\lambda = \lambda_r + I\lambda_i\,, \tag{6.15}$$

λ_r and λ_i being the damping rate and spatial oscillation wavenumber respectively. Insertion of (6.13)-(6.15) into (2.6) and linearisation gives four solutions of the perturbed homogeneous system in terms of the four complex wavenumbers $\lambda_1, ..., \lambda_4$.
These solutions depend on β, θ, d_s, on the bedload transport formula adopted and on the lateral bar mode m. In Fig. 6.4 we show the solutions as functions of β having set $d_s = 0.02$, $\theta = 0.08$ and adopting the bedload transport formula of Wong and Parker [135]. Here, $\lambda_{1i}, ..., \lambda_{4i}$ denote the imaginary part of $\lambda_1, ..., \lambda_4$ respectively, representing the signal spatial wavelength, while $\lambda_{1r}, ..., \lambda_{4r}$ denote the real part associated to the spatial damping rate of steady bars. Positive values of the real part indicate that the signal decays in the upstream direction, while negative values indicate downstream damping. Next we describe in detail all the four solutions.

- λ_1 is a real positive number, corresponding to a non-alternate bar which rapidly decays

upstream.

- λ_2 describes an analogous configuration as λ_1 but with opposite sign, namely a non-oscillating signal that quickly decays downstream.

- λ_3 represents a wave of considerable length which decays either upstream or downstream of the forcing disturbance. The imaginary part $\lambda_{3i} \sim 0.15 \div 0.2$ corresponds to wavelength $L \sim 15 \div 20\,W$ (i.e. approximately double than that of free bars). The direction of decay is governed by the sign of the real part λ_{3r}: for small values of β the signal is damped in the downstream direction, while for high values of β the signal decays upstream;

- λ_4 has the same real part as λ_3 and the same absolute value for the imaginary part, i.e. it is the complex conjugate of λ_3. It is associated to an identical signal.

From Fig. 6.4 we notice that the transition between upstream and downstream damping (i.e. positive and negative values of λ_{3r}) is found in correspondence to the *resonant* value of aspect ratio β_{res} already defined in section 6.2.1: if $\beta < \beta_{res}$ forced bars are expected to decay downstream of the forcing discontinuity while if $\beta > \beta_{res}$ forced bars are expected to decay upstream. We define *sub-resonant* the former and *super-resonant* the latter behaviour. Notice that the characteristic distance of decay is proportional to the inverse of $|\lambda_{3r}|$. Thus in resonant condition $\beta = \beta_{res}$, where $|\lambda_{3r}| = 0$, no damping of forced bars is predicted by the linear theory, i.e. the influence of the forcing disturbance is theoretically felt at infinite distance. Identical considerations apply to λ_4. For sake of simplicity hereinafter with λ_{si} we denote the spatial wavenumber of forced steady bars λ_{3i} and λ_{4i} and with λ_{sr} we indicate the spatial damping rate $\lambda_{3r} = \lambda_{4r}$.

In Fig. 6.5 we compare the outcomes of the linear theories of forced bars [141] (left panel) and of free bars [33] (right panel), plotted using the same set of parameters. We observe that the concept of resonance deeply interlaces the two theories, so that the resonant value of aspect ratio β_{res} is the same in both panels. This is readily explained recalling the meaning of resonance according to the considered theories. The resonant condition corresponds both (left panel) to steady bars whose decay distance tends to infinity ($\lambda_{sr} = 0$) and (right panel) to non-migrating free bars ($\Omega_i = 0$) which do not amplify and neither get damped in time ($\Omega_r = 0$) nor in space. Notice that the wavenumber of resonant free bars λ_{res} coincides with the imaginary part λ_{si} of the complex wavenumber of steady bars. Thus the resonant condition identifies the only state which is common between the two plots in Fig. 6.5, describing the same steady bar.

The theory of spatial bars prescribes the direction of damping of a forced bar triggered by a localised variation of the planimetric configuration, which shall happen in the downstream or

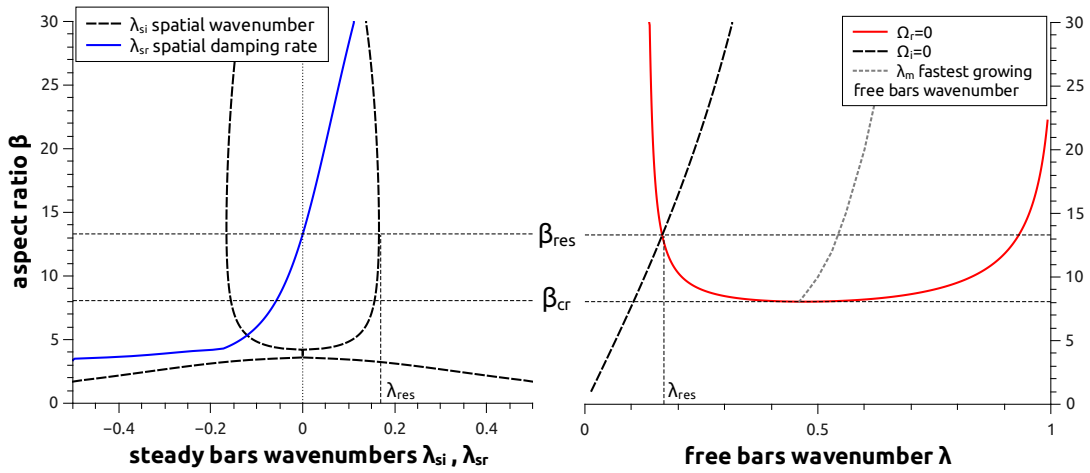

Figure 6.5: Left panel: spatial wavenumber and spatial damping rate of steady bars. Right panel: marginal curves for free bar stability (red line), migration (black dashed line) and curve representing the fastest-growing bar wavenumber (light-blue dotted line). Parameters are $\theta = 0.1$ and $d_s = 0.07$. The bedload transport formula of Wong and Parker [135] has been used.

upstream direction if undisturbed flow is sub-resonant or super-resonant respectively. Zolezzi and Seminara [141] have explicitly linked the direction of *two-dimensional morphodynamic influence* of the local planform disturbance to the direction of bar damping: forced bars are expected to develop only in the direction of damping (decay) prescribed by the theory. The rationale of this link can be easily understood thinking of a straight channel with a localized disturbance in its middle section. If the channel is long enough, it is physically reasonable to assume the occurrence of uniform flow conditions over flat bed at its downstream and upstream ends, provided these are located far enough from the disturbance itself. On the contrary, if a forced bar pattern could spatially develop in the opposite direction with respect to the damping direction, it would amplify towards infinite amplitude, which would not match the above uniform flow-bed boundary conditions. In other words, the presence of the localized disturbance must be felt only upstream under super-resonant conditions and only downstream under sub-resonant condition. Thus, the super-resonant and sub-resonant regimes are associated to upstream and downstream morphodynamic influence respectively.
Concerning numerical modelling, to our knowledge a correct reproduction of the sub-resonant and super-resonant flow regime has been achieved only very recently by van der Meer *et al.* [129] using the Delft3D model [79] and by Siviglia *et al.* [111] using the GIAMT2D model, as we will show in chapter 7. Finally, Crosato *et al.* [40] investigate both experimentally and numerically the possibility of steady bar formation in straight channels without planimetric forcing occurring at very long time scales.

6.2.3 Interaction between free and forced bars

Free and forced bars have been intensively investigated in the last decades by means of theoretical analysis carried out in simple geometrical configurations as well as laboratory experiments. Even though the fundamental mechanisms and the governing parameters have been identified, a comprehensive description of the actual bar dynamics occurring in real rivers has not been achieved so far. Both kinds of bars can be present in real river systems. Theoretical and experimental evidence has been provided that free and forced bars interact through non-linear processes: the bar configuration actually observed in a stream should therefore be fundamentally determined by such interaction. This emerges from experimental and theoretical studies in slowly-varying planform geometries characterised by spatial periodicity, like small-amplitude meanders [68, 127], large-amplitude meanders [134] and channels with longitudinal width oscillations [97, 98]. Here we review the case of weakly meandering channels and the case of straight channels characterised by periodic width variations, for which the interaction has been studied in more detail.

The interactions of free bars with forced bars induced by curvature in small-amplitude meanders has been first experimentally investigated by Kinoshita and Miwa [68]. They observed the coexistence of free bars with the forced topography driven by curvature at low sinuosity and the suppression of migrating bars for larger values of channel curvature. Their results suggest that the interaction between migrating free and steady forced bars is responsible for the suppression of the former perturbations. This has been also documented in subsequent laboratory experiments [34]. Starting from the representation of finite-amplitude free bars given in [33], Tubino and Seminara [127] give a theoretical interpretation of the suppression process by means of a weakly non-linear analysis. The solution is expanded in terms of a small parameter $\frac{\beta-\beta_{cr}}{\beta_{cr}}$ measuring the distance from the marginal conditions for free bar formation. A finite-amplitude representation of forced bars is also used in terms of a dimensionless parameter $\nu = \frac{W}{4R_m}$, R_m being the meander radius at the bend apex, measuring curvature effects. The results of the model of Tubino and Seminara [127] highlight the role of curvature in damping free bars and modifying their propagation celerity. In particular, a threshold value $\nu_{c1}(\theta, d_s, \lambda_m, \beta)$ of the curvature ratio is found, where λ_m denotes the meander dimensionless wavenumber: if $\nu < \nu_{c1}$ (low sinuosity, weak forcing disturbance) free bars are damped and slowed down, but still present, otherwise (high sinuosity, strong forcing disturbance) free bars are suppressed. Such threshold value essentially depends on the amplitude of bars which would develop in the channel in absence of the forcing effect of the curvature (6.12):

$$\nu_{c1} \propto \sqrt{\frac{\beta-\beta_c}{\beta_c}}\,. \tag{6.16}$$

Repetto and Tubino [97] develop a linear stability theory for free bar formation in a straight channel with sinusoidal width variations of small amplitude. They employ a new dimensionless parameter δ^W expressing the amplitude of sinusoidal width variations with respect to the reach-averaged channel width. Their solution results in a linear correction to the growth rate of free bars as predicted in the case of constant width channels (6.9) governed by aspect ratio β [33]. The effect of width variations is found to be invariably stabilizing. In particular, free bars are suppressed provided δ^W exceeds some threshold value. Such threshold is found to be an increasing function of β, i.e. of distance from the critical value for free bar formation $\beta - \beta_{cr}$ expressing the strength of free bars. Below this threshold, width variations are unable to suppress free bars, even though they significantly affect the selection process of the most unstable free bar depending on the perturbation wavelength.

In both configurations, the interaction of free and forced bars is found to be invariably destructive, i.e. free bars are suppressed provided the forcing disturbance is strong enough compared to the strength of free bars. The strength of free bars is found to be a function of the distance from the critical aspect ratio $\beta - \beta_{cr}$ or of its relative value $\frac{\beta-\beta_{cr}}{\beta_{cr}}$ as it was found in the case of a straight channel characterised by free bars only (6.12).
In spite of so much analytical work, however, no theory or predictive model is available at present to describe the interaction of free and forced bars in the basic straight channel configuration having one localised forcing disturbance. In chapter 8 we will present the original results of a numerical study conducted using the GIAMT2D model, concerning free-forced bar interaction in such a planform configuration [132, 142]. While interpreting our numerical results in chapter 8 we will take advantage of the results of analytical modelling, arguing that the distance from the critical conditions $\beta - \beta_{cr}$ and the distance from the resonant conditions $\beta - \beta_{res}$ can represent a measure of the strength of free and forced bars in their non-linear interaction.

6.3 Channel bifurcations

Channel bifurcations are the key morphological units of braided rivers, governing the evolution of gravel-bed braided networks. One of the most relevant field evidences is that channel bifurcations often show markedly unbalanced configurations: downstream branches are generally asymmetrical in terms of discharge, width and topographic configuration [86]. Theoretical and experimental work on the morphodynamics of river bifurcations has been developed only relatively recently (Wang *et al.* [133], Bolla Pittaluga *et al.* [14], Bertoldi [10], Tubino and Bertoldi [125], Kleinhans *et al.* [69]) and has been focusing on the following key questions.

- *equilibrium*: does an equilibrium configuration exist?

Figure 6.6: Bifurcation in a braided river (Tagliamento River, Italy).

- *stability of equilibrium*: is such equilibrium configuration stable?
- *role of different processes*: what is the role of external processes (bar migration, planimetric variations, unsteadiness of flow regime, back-water effects) and intrinsic processes on equilibrium and stability?

Modelling of river bifurcations has been tackled by means of simplified quasi-two-dimensional analytical theories and laboratory experiments; following the summary of Tubino and Bertoldi [125], next we present an introduction to the theoretical approach and experimental observations.

6.3.1 Analytical modelling of river bifurcations

Although several contributions have been proposed, here we focus on reviewing the analytical model of river bifurcations developed by Bolla Pittaluga *et al.* [14], that will be used as a reference theory in the presentation of the numerical results of GIAMT2D in chapter 7. In the model of Bolla Pittaluga *et al.* [14] bifurcations are schematised as a system having one upstream channel, denoted hereinafter with subscript a, and two downstream channels, denoted with b and c respectively (see Fig. 6.7). The geometrical configuration consists of a Y-shaped bifurcation having the upstream channel aligned with the x axis and two symmetrical downstream branches. All the channels are straight and have rectangular cross-section with constant width and bed slope. For each channel we define aspect ratio β_l, Shields stress θ_l and relative roughness d_{sl} (where $l = a,\ b,\ c$) and we solve a one-dimensional hydro-morphodynamic problem. Use of a one-dimensional scheme requires three boundary conditions for each channel: two conditions at the upstream end and one downstream. At the inlet section of channel a water and sediment discharges are prescribed, while a constant water level is set at the downstream sections of channels a, b, c. Concerning the node, Bolla

Pittaluga *et al.* [14] adopt a quasi two-dimensional formulation in order to take into account solid discharge partition between channels b and c. As highlighted in Fig. 6.7, the final reach of channel a is subdivided into two contiguous cells having length $\alpha^B W_a$, where α^B is related to the spatial extension of the upstream branch that is affected by the bifurcation. Five nodal conditions must be set to ensure the correct closure of the problem, namely:

- water mass conservation
$$Q_a = Q_b + Q_c \,, \tag{6.17}$$

- constancy of water level in the node
$$H_l = H_a \qquad \text{for } \; l = b, \, c \,, \tag{6.18}$$

  where two additional conditions expressing the relationship between depth and discharge in the downstream channels using uniform flow approximation are employed:
$$Q_l = W_l C_{hl} D_l \sqrt{g R_{hl} S_l} \qquad \text{for } \; l = b, \, c \,, \tag{6.19}$$

  Q_l being water discharge, W_l channel width, C_{hl} the Chézy coefficient, D_l water depth, R_{hl} the hydraulic radius and S_l bed slope;

- sediment mass conservation equation applied to both cells of the final reach of channel a (see Fig. 6.7):
$$\frac{1}{2}\left(1-\lambda_p\right)\left(1+\frac{W_a}{W_b+W_c}\right)\frac{\mathrm{d}\eta_l}{\mathrm{d}t}+\frac{q_{sl}-q_{sa}\left(\frac{W_a}{W_b+W_c}\right)}{\alpha^B W_a}\mp\frac{q_{sy}}{W_l}=0 \qquad \text{for } \; l = b, \, c \,, \tag{6.20}$$

λ_p being porosity, η bed elevation and q_{sl} solid discharge per unit width in respective channels. In (6.20), q_{sy} denotes the lateral swap of solid discharge between the two cells due to water discharge partition and to local gravitational effects caused by transversal bed deformation. Generalising the relation (2.25) due to Ikeda *et al.* [66] we obtain:
$$q_{sy} = q_{sa}\left\{\frac{Q_y D_a}{q_a \alpha^B \bar{D}} - \frac{2 r_{ik}\left(\eta_b - \eta_c\right)}{\sqrt{\theta_a}\left(W_b + W_c\right)}\right\} \,, \tag{6.21}$$

where $\bar{D}$ is average depth in the contiguous cells, given by
$$\bar{D} = \frac{1}{2}\left(\frac{D_b + D_c}{2} + D_a\right) \tag{6.22}$$

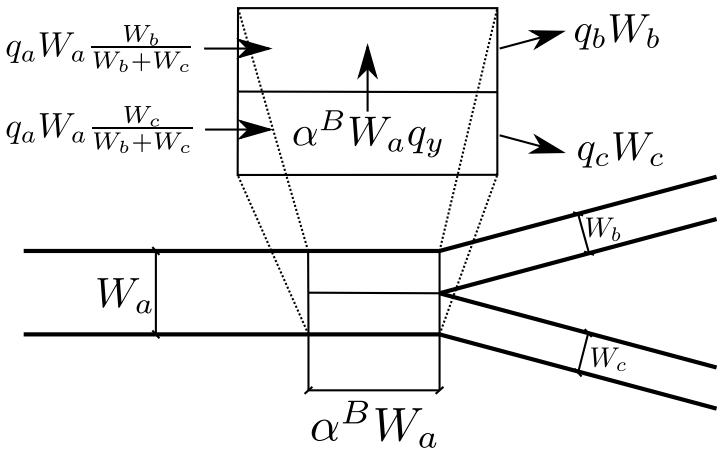

Figure 6.7: Analytical model of Bolla Pittaluga *et al.* [14]. Scheme of the nodal point relation.

and Q_y is water discharge swapped by the cells, computed imposing mass balance applied to each cell:

$$Q_y = \frac{1}{2}\left(Q_b - Q_c - Q_a \frac{W_b - W_c}{W_b + W_c}\right) . \tag{6.23}$$

The swap of solid discharge (6.21) proves to be the key mechanism governing the evolution of channel bifurcations: when one branch decreases its activity in terms of bedload transport, the swap mechanism forces a diversion of solid discharge into the second branch.

The present approach has been extended to the case of erodible banks by Miori *et al.* [86]. However in order to introduce the numerical simulations which will be presented in the next chapter we focus on the fixed-bank case only. In the next section we review the main achievements of Bolla Pittaluga *et al.* [14].

6.3.2 Equilibrium configurations

We consider the geometrical configuration shown in Fig. 6.7, composed by one upstream channel a having width W_a joining two downstream channels b and c having the same width $W_b = W_c = \frac{1}{2}W_a$. Different choices for the model parameters d_s, r_{ik} and α^B produce quantitatively different results. For all the results presented in this section, if not stated otherwise, we set $W_a = 10\,m$, $D_s = 0.01\,m$, $S_a = 0.005$, $r_{ik} = 0.5$, $\alpha^B = 2$ and we apply the bedload transport formula of Wong and Parker [135] (see Tab. 2.1) together with the logarithmic Chézy law for friction

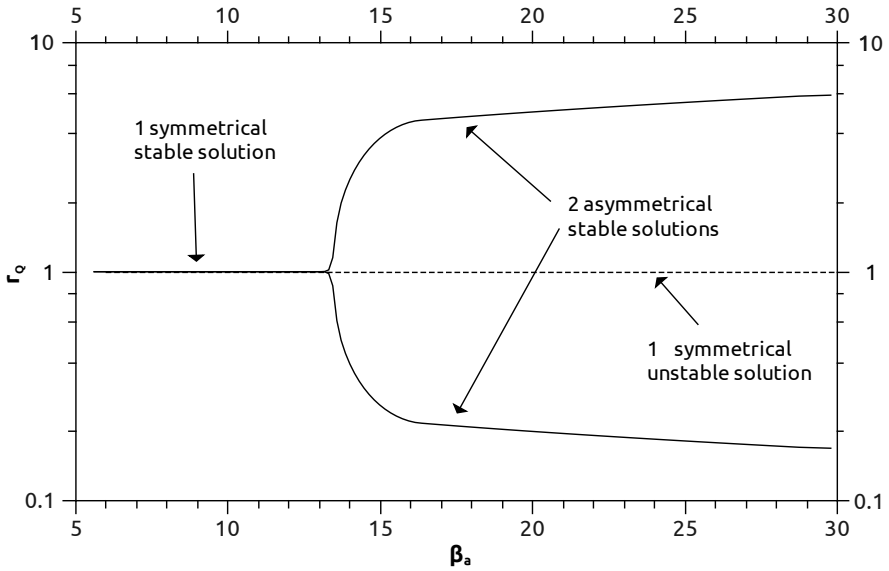

Figure 6.8: Analytical model of Bolla Pittaluga *et al.* [14]. Stable and unstable equilibrium solutions, plotted in terms of discharge ratio (6.24) versus aspect ratio β_a.

(2.20)-(2.13). With the discharge ratio parameter $r_Q \in [0, 1]$, defined as

$$r_Q = \min(r_{Qb}, r_{Qc}) \qquad \text{with} \qquad r_{Qb} = \frac{Q_b}{Q_c} \quad , \quad r_{Qc} = \frac{Q_c}{Q_b}, \tag{6.24}$$

we express the degree of symmetry in flow partition. The symmetrical configuration is associated to the value $r_Q = 1$, whereas if one downstream branch of the bifurcation tends to close, we have $r_Q \to 0$. In non-symmetrical configurations, the downstream branch carrying most of discharge has larger bedload, which generates an increment in bed erosion; on the other hand the channel with lower discharge experiences bed aggradation. The absolute difference in bed elevation between the two channels measured at the inlet section is called *inlet step* $\Delta\eta$, given by

$$\Delta\eta = |\eta_b - \eta_c| = |(H_b - D_b) - (H_c - D_c)| = |D_c - D_b| \ , \tag{6.25}$$

under the hypothesis that water elevation measured just downstream of the node is equal in both branches ($H_b = H_c$).

From Fig. 6.8 we observe that two regions exist depending on the aspect ratio of the upstream channel β_a:

- for low values of β_a, corresponding to high values of Q_a, an unique equilibrium solution

exists, i.e. the symmetric configuration ($r_Q = 1$), which is stable;

- increasing the aspect ratio a threshold is reached: hereinafter the model gives three possible solutions. One of them, representing the symmetrical configuration, is unstable, while the others give non-symmetrical equilibrium, diverting most of discharge in channel b ($Q_b > Q_c$) or c ($Q_c > Q_b$) respectively. Both stable configurations give $r_Q < 1$ (6.24).

It is worth noticing that if a bifurcation reaches non-symmetrical stable equilibrium diverting most of the discharge to one downstream channel (say, channel b), a switch towards the opposite stable configuration is possible at any time due to intrinsic or external causes. Among external causes we mention the interference given by propagation of free bars from the inlet channel a into the bifurcation. The existence of two opposite non-symmetrical equilibrium configurations motivates the use of the minimum function in the definition of r_Q (6.24).

The same considerations about existence of symmetrical equilibrium solutions can be drawn from Fig. 6.9, where the discharge ratio is plotted versus discharge Q_a in the upstream channel. High values of Q_a (low β_a) give symmetrical equilibrium, whereas low Q_a (high β_a) gives non-symmetrical equilibrium. In Fig. 6.10 these two different kinds of behaviour can be observed in terms of inlet step $\Delta\eta$. In detail, in the top panel the model output is represented in terms of inlet step $\Delta\eta$ and depth in the inlet channel D_a, while in the bottom panel we plot the *dimensionless inlet step* given by their ratio $\frac{\Delta\eta}{D_a}$. We observe that with decreasing discharge the inlet step increases and becomes of the same order of magnitude as the average upstream water depth.

The transition between configurations characterised by symmetrical or non-symmetrical equilibrium can be represented by threshold curves in the plane $\beta_a - \theta_a$, (Fig. 6.11). Here, for each curve we have used one constant value of d_{sa} by varying the dimensional sediment size D_s accordingly with D_a. Thus, the Chézy coefficient C_{ha} computed using the logarithmic relation (2.13) is constant within each curve as well. Depending on the Shields stress in the upstream channel, for high values of θ_a the bifurcation tends to keep both branches open in a stable symmetrical configuration, while for low values of θ_a the symmetrical configuration is unstable. This is in agreement with the field observations of Zolezzi *et al.* [139]. In the latter case the system may reach a stable non-symmetrical equilibrium configuration or may even tend to close one branch. Furthermore, external processes such as free bar migration in channel a may generate a repeated switch between the two non-symmetrical equilibrium configurations, which may result in general unstable and possibly cyclic behaviour [12].

The threshold curves in Fig. 6.11 show moderate sensitivity to variations of C_{ha}, r_{ik} and S_a,

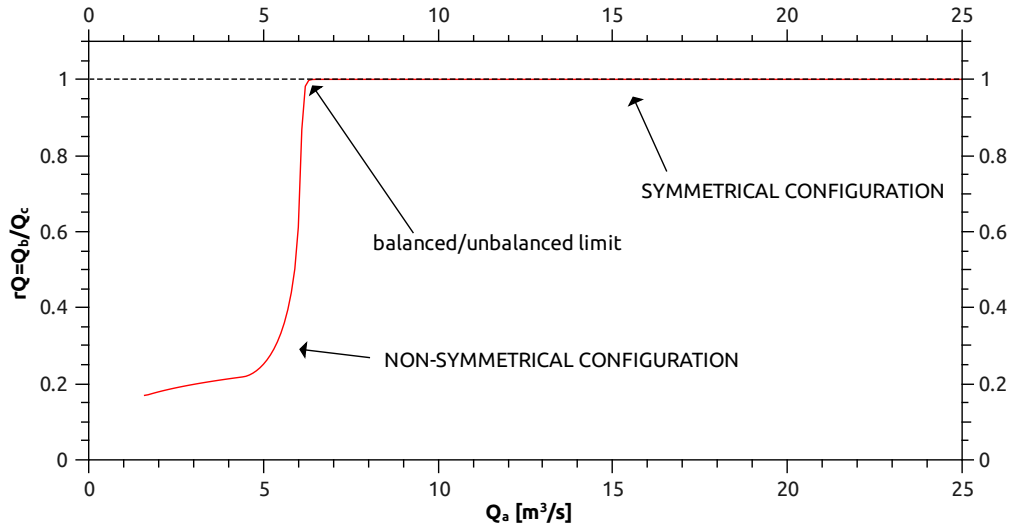

Figure 6.9: Analytical model of Bolla Pittaluga *et al.* [14]. Discharge ratio (6.24) versus upstream discharge Q_a.

especially for low values of β and θ. On the other hand, different values of the α^B parameter generate quantitatively different curves. Hence an open issue is to find the value of α^B which best matches the model predictions with experimental observations. This question has been experimentally addressed by Bertoldi and Tubino [11]. In the next section we will present an overview of the experimental results available, compared with the theoretical predictions.

6.3.3 Overview of experimental observations

Experimental observations of channel bifurcations have been made in several studies, as reported by Tubino and Bertoldi [125]. Laboratory experiments have been performed in a single bifurcation system by Bertoldi and Tubino [11]. Several works also focus on bifurcations in a braided network, in order to better understand their role in overall braided rivers evolution (see e.g. Federici and Paola [50]). From all these experimental works the role of the Shields parameter and of the aspect ratio in the upstream channel as key controlling parameters emerges, in agreement with the outcomes of analytical modelling. Here we concisely review the experimental results obtained by Bertoldi and co-workers [11] in a comprehensive set of experiments at the Department of Civil and Environmental Engineering of the Università di Trento.

In Figs. 6.12 and 6.13 we plot the discharge ratio r_Q and the dimensionless inlet step $\frac{\Delta\eta}{D_a}$ as function of θ_a (Figs. 6.12a and 6.13a) and β_a (Figs. 6.12b and 6.13b) reported in [11] and observe

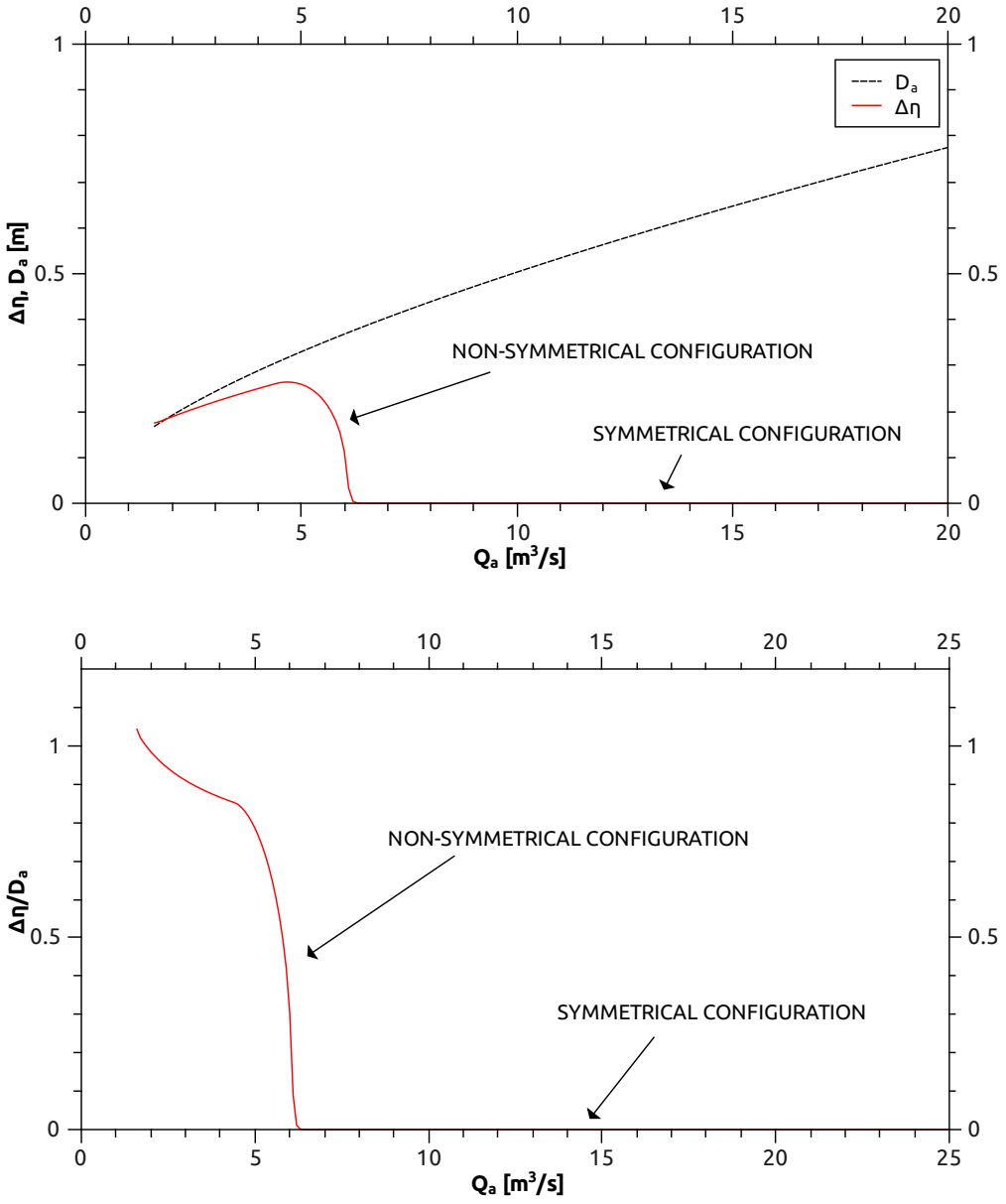

Figure 6.10: Analytical model of Bolla Pittaluga *et al.* [14]. The inlet step $\Delta\eta$ and water depth D_a (top panel) and their ratio (bottom panel) are plotted versus discharge in the upstream branch Q_a.

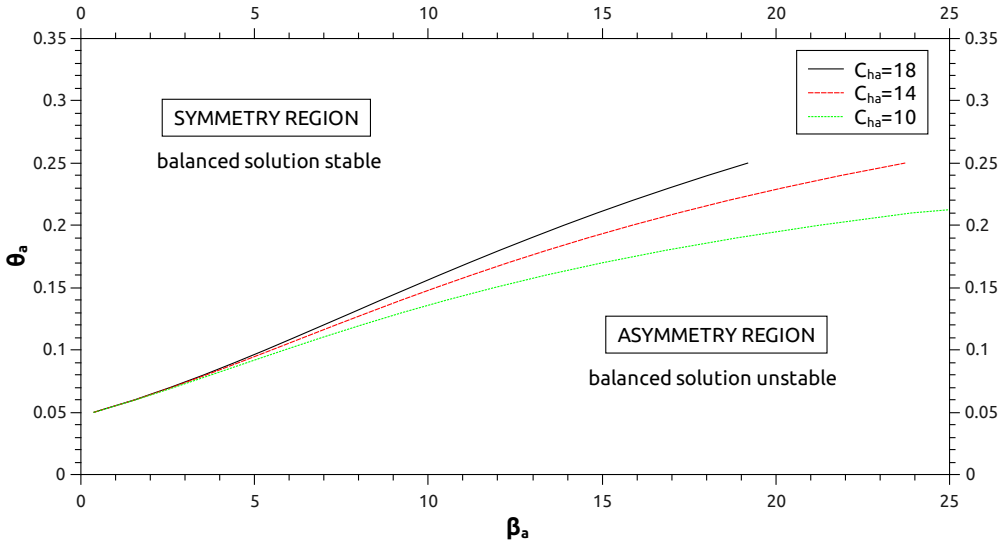

Figure 6.11: Analytical model of Bolla Pittaluga *et al.* [14]. The threshold curve for stability of the symmetrical configuration is plotted in the plane $\beta_a - \theta_a$ for three values of the Chézy coefficient in the upstream branch C_{ha}.

that quasi-symmetrical configurations ($r_Q \sim 1$, $\frac{\Delta\eta}{D_a} \to 0$) are likely to occur for high θ_a and low β_a, in agreement with the analytical model [14]. In contrast, low values of discharge tend to destabilise balanced equilibrium, finally giving non-symmetrical configurations.

In Fig. 6.14 all the experimental runs of Bertoldi and Tubino [11] are reported in the $\beta_a - \theta_a$ plane together with the threshold curve separating the region of symmetrical equilibrium configurations to that of non-symmetrical equilibrium according to the theoretical model of Bolla Pittaluga *et al.* [14]. The latter curve has been obtained using a set of parameters which well represents the average hydraulic conditions in the experiments. Good agreement is found between experimental data and the output of the analytical model.

Finally, we introduce a dimensionless time variable allowing us to compare the time development of different experiments by dropping the dependence on bedload intensity:

$$T = \frac{t}{T_M} \,. \tag{6.26}$$

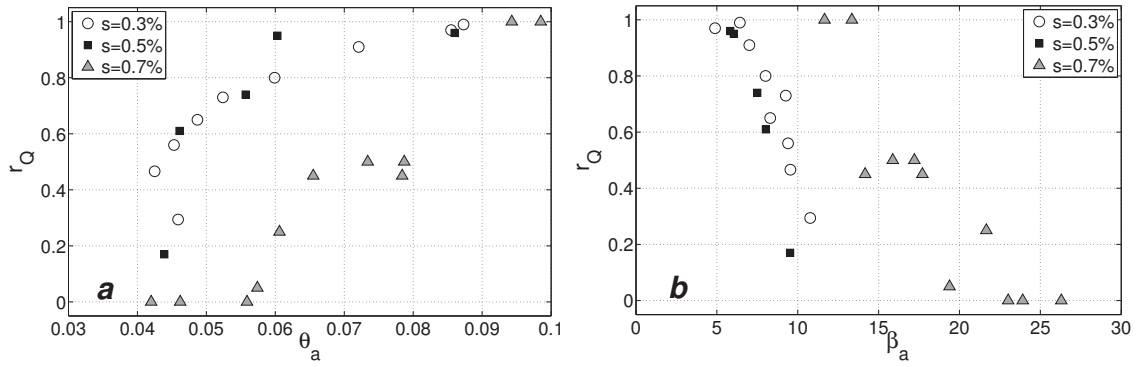

Figure 6.12: Experimental results of Bertoldi and Tubino [11]. Equilibrium values of the discharge ratio r_Q of the downstream branches as a function of the Shields stress θ_a (a) and of the aspect ratio in the upstream channel β_a (b).

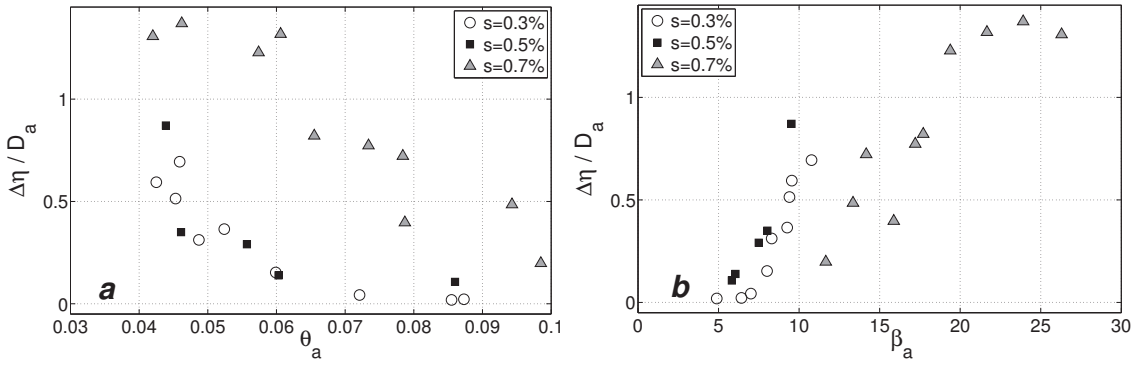

Figure 6.13: Experimental results of Bertoldi and Tubino [11]. Equilibrium values of the dimensionless inlet step $\frac{\Delta\eta}{D_a}$ as a function of the Shields stress θ_a (a) and the aspect ratio in the upstream channel β_a (b).

The morphological scale T_M expresses the duration of morphological processes and is defined with reference to the Exner equation (2.4)

$$T_M = \frac{W_a D_a}{\sqrt{g\Delta D_{sa}^3}\Phi(\theta_a)}\,, \tag{6.27}$$

where the meaning of symbols has been introduced in section 2.1.2.

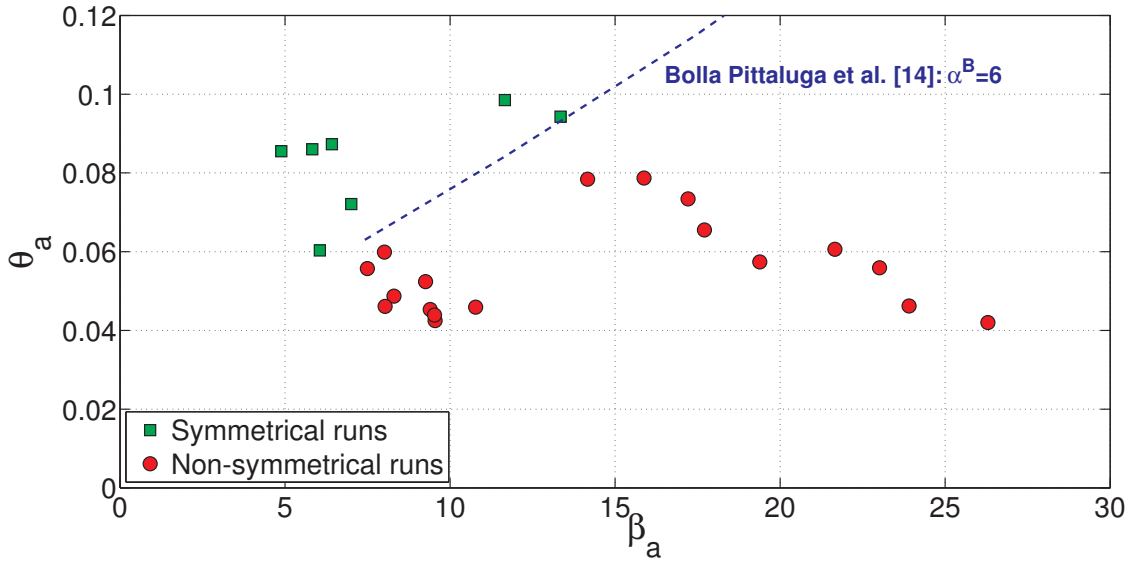

Figure 6.14: Occurrence of symmetrical and non-symmetrical configurations, as predicted by the theoretical model of Bolla Pittaluga *et al.* [14] (dashed line) and experimentally observed by Bertoldi and Tubino [11] (symbols).

7 Morphodynamic benchmarks: numerical applications

In this chapter we apply the GIAMT2D numerical model to three well-established morphodynamic benchmarks, namely the development of free alternate bars in straight channels, the assessment of bifurcation stability and the assessment of the direction of morphodynamic influence in a straight channel with a local planimetric disturbance. We compare the numerical results of GIAMT2D with those of analytical modelling and laboratory experiments. With these applications we wish to test the ability of the numerical model in reproducing well-known morphodynamic behaviours in controlled conditions.
Concerning the development of free bars, we consider the theory of Colombini *et al.* [33], while for bifurcations we use the experimental results of Bertoldi and Tubino [11] and the analytical model of Bolla Pittaluga *et al* [14]. Concerning morphodynamic influence, we check the numerical results with the linear theory developed by Zolezzi *et al.* [141]. For an overview of these theories and experimental achievements we refer the reader to chapter 6.
With these tests we aim at assessing the ability of GIAMT2D in reproducing the behaviour of migrating and steady bars and bifurcations in regular channel geometry, which we propose as the basic morphodynamic units having strong significance for the evolution of complex river systems. Thus, these tests are preliminary for the application of the numerical model as a predictive tool in order to carry out original research in morphodynamics, as we will do in chapter 8, and for future upscaling in order to simulate the dynamics of complex gravel-bed river systems. We emphasise that testing a numerical model against the major outcomes of analytical and physical modelling in morphodynamics seems quite a novel approach in the literature, since few systematic applications of this kind can be presently found. The material here presented is partially covered in [111].

The present chapter outlines as follows. In section 7.1 we present our results concerning the development of free bars in straight channels. In section 7.2 we focus on reproducing the morphodynamic evolution of river bifurcations. Finally in section 7.3 we prove that the GIAMT2D model is able to correctly account for the direction of morphodynamic influence.

Q	$[m^3 s^{-1}]$	500	Fr	$[-]$	1.1
W	$[m]$	60	r_{ik}	$[-]$	0.5
Ks	$[m^{\frac{1}{3}} s^{-1}]$	30	d_s	$[-]$	0.067
D_s	$[m]$	0.12	β	$[-]$	16.7
D_0	$[m]$	1.79	β_{cr}	$[-]$	8.1
S_x	$[-]$	0.011	β_{res}	$[-]$	13.3
λ_p	$[-]$	0.4	λ_{cr}	$[-]$	0.46
Δ	$[-]$	1.65	λ_{res}	$[-]$	0.17
θ_0	$[-]$	0.100	T_M	$[s]$	14425

Table 7.1: Time evolution of free bars. Parameters used in the simulation.

7.1 Free bar formation

Alternate bars appear due to intrinsic instability of the system composed by water flowing over erodible bottom in almost-straight channels. This section concerns the numerical simulation of free bar dynamics in straight computational domains. We propose two tests. In the first test we observe the development of a train of bars from an initially almost-flat riverbed configuration. In the second test, we assess the ability of GIAMT2D in correctly predicting the stability properties of selected bars depending on flow parameters. We compare our results to those of the analytical bar theories presented in section 6.2.1.

7.1.1 Time evolution of free bars

In this test we generate a train of free migrating bars from an almost-flat initial riverbed configuration and assess the time evolution of bar amplitude. We set a straight, rectangular channel having width $W = 60\,m$ in the computational domain $[0,6000] \times [0,60]$ m discretised with 33372 triangular cells. Flow is in the x direction. The lateral reflective boundary conditions are slip. Concerning the upstream condition, we prescribe constant liquid inlet flow discharge Q and solid feed in equilibrium with the hydrodynamic conditions, while a free outflow condition is applied at the downstream end. We use the bedload transport formula of Wong and Parker [135] (see Tab. 2.1). The values of parameters for this test are given in Tab. 7.1.

Numerical generation of free bars has to be triggered by introducing some localised or diffused perturbation. One approach consists of prescribing a small hump or scour in the river bed near the inlet section. Such local perturbation triggers instability, which results in the growth of alternate bed forms. Examples of this procedure are given in Bernini *et al.* [9] and Defina [45]. Another

approach consists of introducing a small random perturbation throughout the domain of flat bed. Here we choose the latter approach, which proves to be effective in triggering the development of bar instability, and impose the following initial riverbed configuration:

$$\eta(x,y,0) = \eta_0(x) + \eta^R(x,y) \tag{7.1}$$

being $\eta_0(x)$ the flat configuration associated to the slope vector $\vec{S} = (S_x, 0)$ and $\eta^R(x,y)$ the random perturbation, whose amplitude is set 5% of uniform flow depth. The numerical generation of free bars requires to be constantly fed by periodically introducing a small perturbation of flat riverbed elevation close to the upstream end. Otherwise, a train of bars would develop and propagate downstream, finally leaving a flat configuration [9, 45]. This can be related to the convective nature of free bar instability assessed by Federici and Seminara [51].

From the analytical theories of bars we recall that the key parameter controlling the formation of free bars is aspect ratio β, to be compared with its critical value β_{cr}. For the set of data in the present test (Tab. 7.1) we have $\beta > \beta_{cr}$, thus free bars of finite amplitude are expected to develop from the slightly perturbed initial configuration (7.1). Notice that the wavenumber of these bars is spontaneously selected by the amplification mechanism and, according to the linear theory, is equal to the wavenumber of the fastest-growing bar, i.e. slightly higher than $\lambda_{cr} = 0.46$.

Starting from the initial riverbed configuration (7.1) and conducting a movable-bed simulation we observe the development of a train of alternate bars and measure their amplitude. The survey is performed in the central part of the domain, far from the upstream and downstream boundary conditions, and the final representative value of bar amplitude is obtained by averaging the amplitude of each bar. The temporal trend obtained is represented in dimensionless form in Fig. 7.1. Bar amplitude is scaled with the reference uniform depth ($\frac{A}{D_0}$) while dimensionless time T (6.26) is computed with reference to the morphological time scale $T_M = 14425\,s$ (6.27). Markers represent the output of the numerical model, while the full red line is a best fit of the numerical results obtained using a sigmoid Boltzmann curve in order to highlight the general trend. This curve well qualitatively compares with the output of analytical bar theories. From these results we observe that initially, say for $T < 3$, the trend well approximates exponential growth prescribed by the linear theory (6.11), while later non-linear effects damp the exponential growth, eventually allowing the achievement of equilibrium amplitude (6.12), in agreement with the weakly non-linear theory [33]. The wavenumber obtained at equilibrium by averaging the wavelength of bars is $\lambda = 0.47$. This value is in good agreement with that prescribed by the linear theory ($\lambda_{cr} = 0.46$) and with the outcomes of weakly non-linear analyses [33], which suggest that bar wavelength selection is only

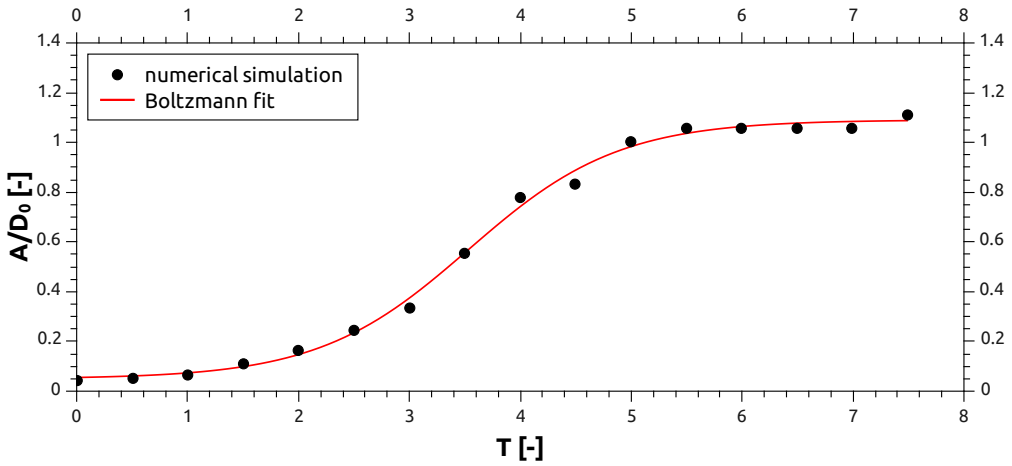

Figure 7.1: Time evolution of free bars. The dimensionless amplitude of free alternate bars $\frac{A}{D_0}$ is plotted against dimensionless time T (6.26)-(6.27). The numerical results, represented by symbols, are fitted with a Boltzmann sigmoid curve.

weakly affected by non-linear effects.

7.1.2 Numerical reproduction of the marginal curve

According to the linear bar theory, the evolution of a monochromatic perturbation having wavenumber λ and infinitesimal amplitude is fully determined by the dimensionless parameters β, d_s and θ of the undisturbed uniform flow. Such perturbation may either get damped or amplify and migrate and in the latter case propagate in the downstream or upstream direction. The marginal curves reported in Fig. 6.3, drawn in the $\lambda - \beta$ plane for given values of θ and d_s, allow easy identification of the states associated to these three kinds of behaviour, as we have explained in section 6.2.1.

In this test we wish to prove the ability of the GIAMT2D numerical model in predicting the fate of bar perturbations of selected wavenumber in agreement with the linear bar theory. We remark that, unlike the previous test, here we wish to assess the response of the numerical model to a monochromatic perturbation initially imposed having wavelength λ. Thus we assign the following initial condition for bed elevation, given by a wave perturbation of the flat bed configuration $\eta_0(x) = -Sx$:

$$\eta(x,y,0) = \eta_0(x) + \varepsilon^P \sin\left(\frac{\pi y}{W}\right) \cos\left(\frac{2\lambda}{W}x\right) . \tag{7.2}$$

#	λ	β
1	0.05	16.3
2	0.15	16.3
3	0.3	16.3
4	0.08	10.1
5	0.4	7.9
6	0.6	5.6
7	0.8	8.4
8	1	11.2
9	0.2	6.0
10	0.5	16.3
11	0.9	16.3
12	0.4	23.0
13	0.15	23.0

Table 7.2: Parameters for numerical reproduction of the marginal curve.

Perturbation amplitude ε^P must be small in order to comply with the hypotheses of the linear bar theory. In practice we set $\varepsilon^P = 0.05\,D_0$, D_0 being the uniform flow depth.

We have performed 13 runs setting different values of λ and β, as reported in Tab. 7.2. In all the numerical runs we have been using the same values of the dimensionless parameters $\theta = 0.1$, $d_s = 0.061$, with slope $S = 0.01$, relative density $\Delta = 1.65$ and porosity of the river bed $\lambda_p = 0.4$. We have applied the bedload transport formula of Wong and Parker [135] and used $r_{ik} = 0.3$. For the given set of data, the linear theory prescribes $\beta_{cr} = 6.28$, $\lambda_{cr} = 0.42$, $\beta_{res} = 10.37$, $\lambda_{res} = 0.13$. For all the runs, the domain is $[0,\,300] \times [0,\,1]\ m$, representing a channel having width $W = 1\,m$, discretised with 71476 triangular cells. Dimensional data such as discharge Q, uniform flow depth D_0 and sediment diameter D_s vary in each run and can be computed from the non-dimensional parameters.
Numerical runs are conducted just for a few seconds: in the few initial integration steps the perturbation amplitude is seen to decay or amplify and, in the latter case, to propagate upstream or downstream. The survey of numerical results is performed *before* amplifying waves could reach to finite amplitude, i.e. still within the hypotheses of the linear bar theory.

The results of two runs, namely run 8 and run 12, are depicted in Figs. 7.2 and 7.3 respectively in terms of longitudinal profiles of riverbed elevation computed at the left bank (analogous

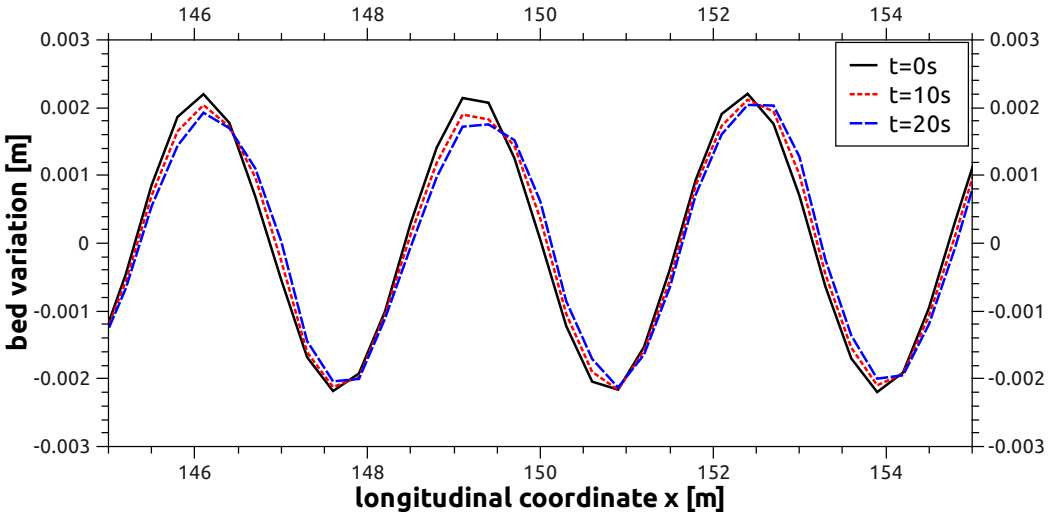

Figure 7.2: Numerical reproduction of the marginal curve. Bar stability (run 8). Longitudinal profile of riverbed elevation obtained slicing the numerical solution along the left bank plotted at time $t = 0,\ 10,\ 20\,s$.

results are obtained on the other bank).

Let us focus on run 8 (Fig. 7.2). According to the linear theory, the combination of parameters in this run should give rise to perturbation damping. Such behaviour is due to the extremely high wavenumber ($\lambda = 1$, representing a very short bar), even though the channel aspect ratio for this run is super-critical. Our numerical results well agree with the prediction of the linear theory, as we observe from Fig. 7.2, where wave amplitude is seen to decay in time.

Let us consider then run 12 (Fig. 7.3). The linear theory in this case predicts amplification and downstream migration. In detail, the value $\lambda = 0.4$ is quite close to that associated to maximum amplification rate and the channel aspect ratio for this run is super-critical ($\beta = 23.0 > \beta_{cr}$). From Fig. 7.3 we observe that bar amplification for this run is correctly reproduced by the numerical model. While amplifying, the bars migrate downstream, accordingly with the theory.

Concerning the runs associated to small wavenumbers, e.g. run 1 and run 13, the theory predicts upstream migration and amplification: our numerical results confirm this prediction showing that wave peaks grow and slightly move upstream with respect to the initial position.

The results of all the numerical runs are presented in Fig. 7.4. Here we plot the marginal curves of the linear theory (full lines) together with the results of numerical runs represented by symbols. Three kinds of behaviour are observed, namely suppression (damping) of riverbed perturbation (red symbols), amplification and downstream migration (triangular blue symbols), amplification

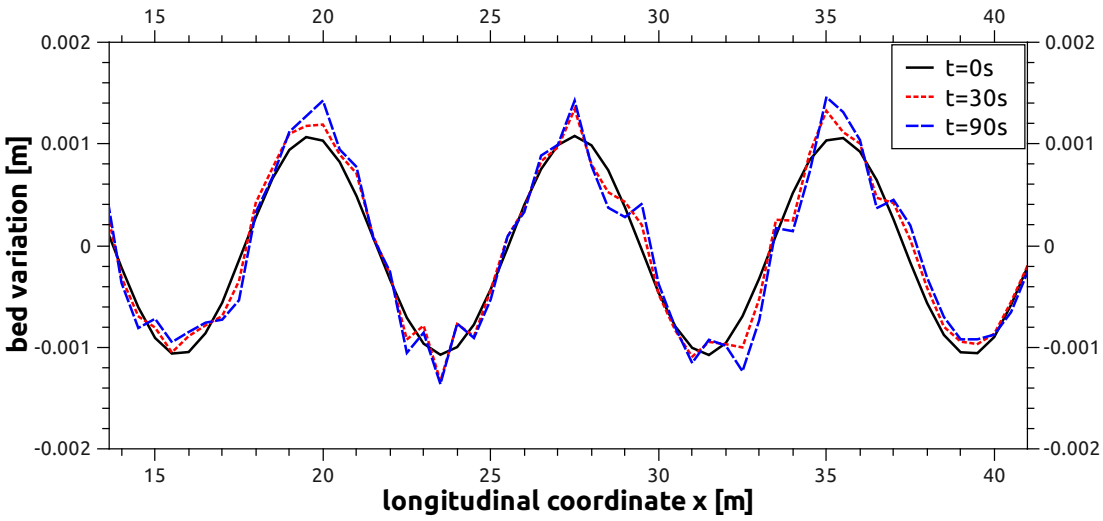

Figure 7.3: Numerical reproduction of the marginal curve. Bar instability (run 12). Longitudinal profile of riverbed elevation obtained slicing the numerical solution along the left bank plotted at time $t = 0, 30, 90\,s$.

and upstream migration (circular light-blue symbols). The numerical results of GIAMT2D closely match the linear theory.

The two tests presented in this section prove that the GIAMT2D numerical model is able to correctly simulate the bar generation mechanism, both in terms of time evolution and wavelength selection.

7.2 Numerical simulation of bifurcation stability

In this section we apply the GIAMT2D model to study the evolution of channel bifurcations. Our benchmarks are the experimental observations of Bertoldi and Tubino [11] and the theoretical model of Bolla Pittaluga *et al.* [14], which have been reviewed in section 6.3. The aim of the present tests is to assess the capability of the numerical model to correctly predict the threshold between the symmetrical and asymmetrical configurations depending on the characteristics of the inlet flow.

Numerical simulations are performed on a Y-shaped domain coherently with the analytical model [14]. The planimetric configuration is symmetrical and the downstream branches (b and c) draw an angle of 15° with the longitudinal direction of the inlet channel a. Channels b and c have the

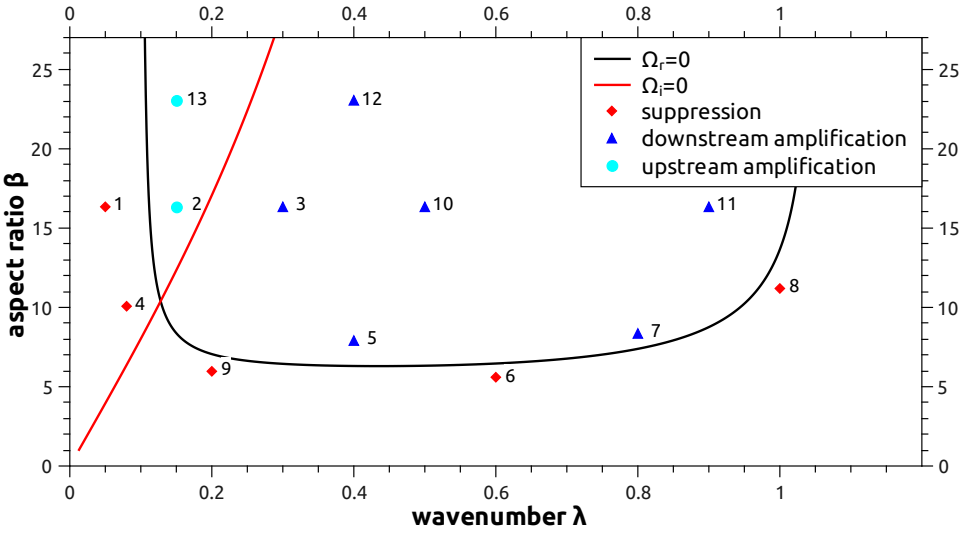

Figure 7.4: Numerical reproduction of the marginal curve. The stability properties of all numerical runs (symbols) are plotted together with the marginal curves of the linear bar theory (full lines).

same width, the sum of which is equal to 1.3 times the width of channel a, as suggested by Griffiths [55] in the case of symmetrical configurations. Such planimetric configuration is analogous to that used by Bertoldi and Tubino [11]. We have performed 18 numerical runs. The dimensional size of the domain changes in different runs in order to obtain the required values of the dimensionless parameters β, θ, d_s. This approach relies on the assumption that the morphodynamic processes governing the evolution of bifurcations depend on dimensionless parameters and not on the absolute size of channels, as suggested by the analytical model. In Fig. 7.5 a sketch of one computational domain, discretised with a triangular mesh, is shown. Tab. 7.3 reports the parameters associated to reference uniform flow in the upstream channel a for each run. The initial bed setup is given by

$$\eta(x,y,0) = \eta_0(x) + \eta^R(x,y) \ , \tag{7.3}$$

where $\eta_0(x)$ represents the initial setup associated to the slope vector $\vec{S} = (S_x, 0)$, i.e. a flat plane aligned with the direction of channel a, while $\eta^R(x,y)$ is a random disturbance having amplitude equal to 5% of undisturbed flow depth in channel a. For all runs we use $S_x = 0.005$. We apply the disturbance throughout channel a, except close to the upstream boundary conditions. Use of a random disturbance shall trigger instability, provided the considered configuration is physically unstable. Runs are conducted using the bedload transport formula of Wong and Parker [135] and imposing $r_{ik} = 0.5$ to quantify the effect of the lateral bed slope.

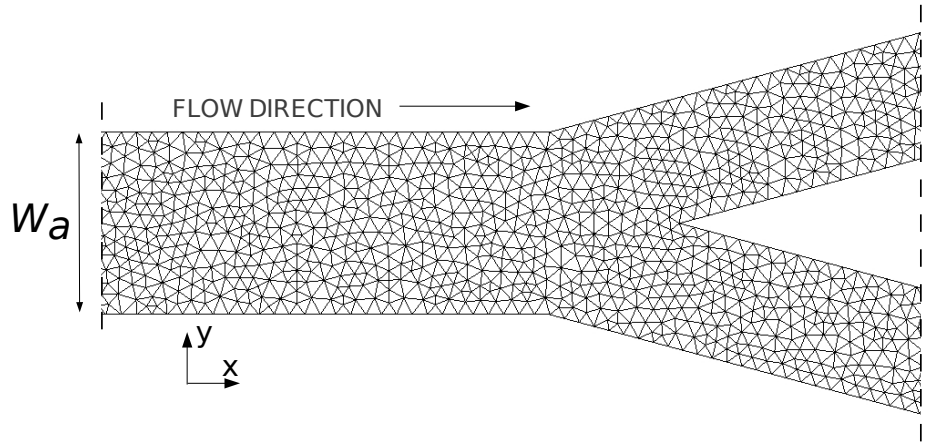

Figure 7.5: Numerical simulation of bifurcation stability. Sketch of the computational domain.

Simulations are first conducted in fixed-bed mode until hydrodynamic equilibrium is achieved. At this stage, the bed being perfectly flat in the transverse direction at the downstream end of channel a, flow partition between channels b and c is symmetrical, which gives $r_Q = 1$ (6.24).
Afterwards, we allow bedload transport and observe changes in riverbed elevation throughout the domain. A key role in bifurcation evolution is played by alternate bars: if they form in channel a and migrate downstream they may interact with the bifurcation, possibly causing alternate discharge swaps around the equilibrium configuration. The random disturbance imposed in the initial condition (7.3) may initially trigger the development of free bars; however, in order to limit their influence, their formation is not further sustained.

We use the discharge ratio parameter r_Q (6.24) as an indicator for the degree of symmetry in the topographic configuration. Measuring r_Q at different time stages during the simulation allows us to detect the evolution trend. In order to compare different simulations, characterised by different bedload intensity, we conveniently recast time in dimensionless form T (6.26) using the morphological time scale T_M (6.27). In Fig. 7.6 we report the computed patterns of discharge ratio r_Q versus dimensionless time T for runs 1, 2 and 9. Run 1 (black dashed line) and 2 (red line) represent the evolution of bifurcations towards non-symmetrical configurations, while the bifurcation in run 9 (green line) has symmetrical equilibrium. Our results show that the characteristic time scale for destabilisation is of the order of 10 times the morphological scale T_M (6.27).

#	β_a	θ_a	d_{sa}
1	19.7	0.090	0.034
2	12.2	0.090	0.034
3	24.0	0.090	0.034
4	7.7	0.080	0.038
5	10.2	0.110	0.028
6	6.7	0.160	0.019
7	10.2	0.200	0.015
8	14.5	0.150	0.020
9	4.2	0.110	0.028
10	14.1	0.060	0.051
11	18.2	0.060	0.051
12	15.1	0.150	0.020
13	20.2	0.070	0.043
14	3.5	0.080	0.038
15	5.1	0.100	0.030
16	15.1	0.080	0.038
17	17.2	0.120	0.025
18	14.5	0.120	0.025

Table 7.3: Numerical simulation of bifurcation stability: non-dimensional parameters β_a, θ_a and d_{sa} of each run.

In order to prove that the final configuration is *not* determined by the balanced initial condition (7.3) we have performed again some runs imposing initial markedly imbalanced riverbed configurations. The final configuration obtained with this setup is identical to that obtained using (7.3), which confirms that asymptotic equilibrium only depends on the parameters of inlet flow, as predicted by the analytical model [14].

In Fig. 7.7 we give a comprehensive presentation of our runs in terms of the discharge ratio r_Q measured at final time $t = 10\,T_M$. Results are presented in terms of the dimensionless parameters β_a and θ_a related to the characteristics of undisturbed inlet flow. The runs ending with non-symmetrical discharge partition ($r_Q < 1$) are represented by square symbols in the $\beta_a - \theta_a$ plane, while those ending with symmetrical equilibrium ($r_Q \sim 1$) are represented by full triangular symbols. Two distinct regions are clearly identified: stable symmetrical configurations are associated to low values of β_a and high values of θ_a, while unstable symmetrical configurations are found for higher β_a and lower θ_a. This general behaviour qualitatively agrees with the outcomes of analytical and physical modelling (see section 6.3).
The numerical results can be interpreted in the light of the analytical model of Bolla Pittaluga *et al.* [14] obtaining good qualitative agreement. We evaluate the parameter α^B in the theoretical

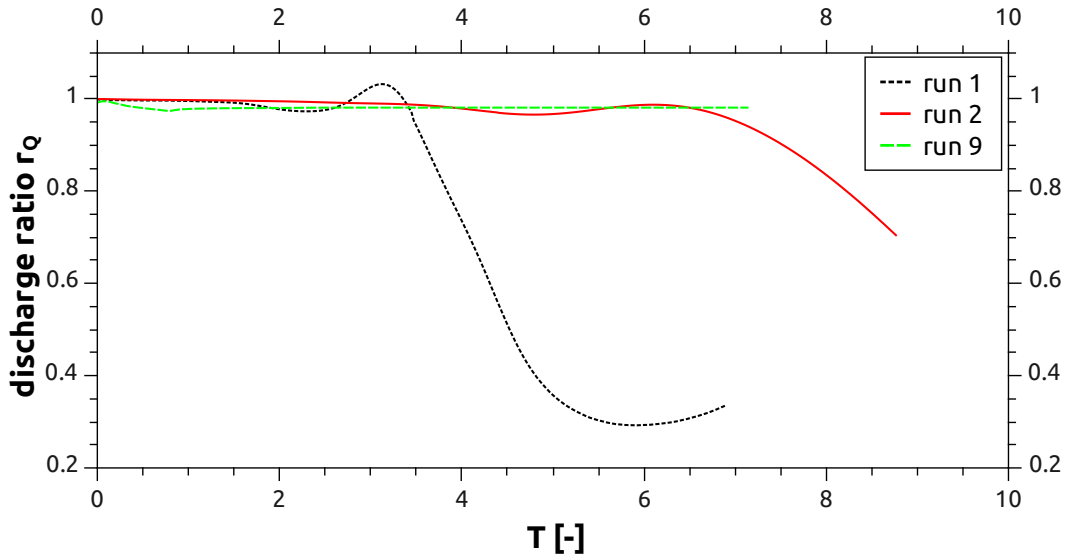

Figure 7.6: Numerical simulation of bifurcation stability. Discharge ratio r_Q versus dimensionless time T (6.26)-(6.27) for three numerical runs.

model which best fits the numerical results. As we show in Fig. 7.7, where the thresholds given by the analytical model are represented by full lines, the optimal values of α^B are found in the range $\alpha^B = 2.5 \div 3$.

In Fig. 7.8 we plot our numerical results in the $\theta_a - \beta_a$ plane together with the experimental results of Bertoldi and Tubino [11]. Numerical results are represented by blue symbols, experimental data by red symbols. Triangular symbols are associated to symmetrical final configurations, square symbols to non-symmetrical final configurations. Good agreement between the numerical results and the experimental results is found, suggesting that the GIAMT2D model is able to correctly reproduce the dynamics of bifurcations highlighted by physical modelling.
Moreover, with full lines we represent the threshold given by the theoretical model [14] plotted for three different values of parameter α^B. We use $\alpha^B = 2.5$ and $\alpha^B = 3$ (red and blue line) which best fit the numerical results and $\alpha^B = 6$ (green line) which best fits the experimental results. Both values fall within the suitable range $\alpha^B = 1 \div 7$ suggested by Bolla Pittaluga *et al.* [14]. The reasons for the slight mismatch may be attributed to the approximations of mathematical and numerical modelling and could be the matter for further analysis in future work.

The results of the numerical runs concerning channel bifurcations reported in this section confirm that the GIAMT2D model correctly identifies the threshold in the inlet flow parameters separating

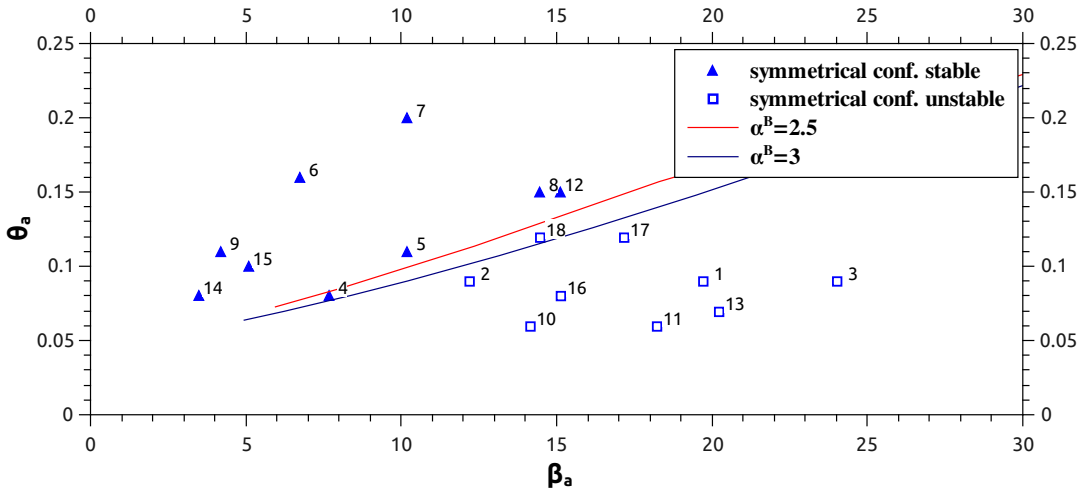

Figure 7.7: Numerical simulation of bifurcation stability. The final configuration for all the numerical runs is represented in the $\beta_a - \theta_a$ plane (symbols) together with the thresholds of the analytical model [14] for $\alpha^B = 2.5$ and $\alpha^B = 3$ (full lines).

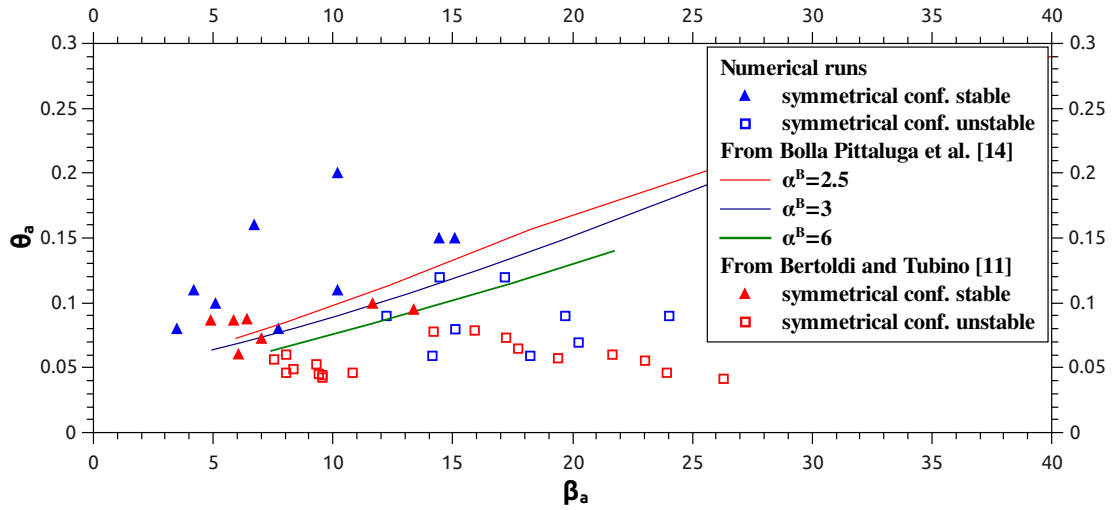

Figure 7.8: Numerical simulation of bifurcation stability. The final configuration for all the numerical runs is represented in the $\beta_a - \theta_a$ plane (blue symbols) together with the experimental results of Bertoldi and Tubino [11] (red symbols) and the thresholds of the analytical model [14] for different values of α^B (full lines).

the asymptotically symmetric and non-symmetric configuration, as described by the analytical model of Bolla Pittaluga *et al.* [14]. Good agreement is found between the numerical results of GIAMT2D and the experimental data of Bertoldi and Tubino [11].

7.3 Numerical simulation of morphodynamic influence

As we have seen in section 6.2.2, the morphodynamic influence of a localised planimetric forcing perturbation is felt in the upstream or downstream direction depending on the channel aspect ratio. According to the theory of Zolezzi and Seminara [141], forced bars shall develop only and decay in the downstream direction if flow is sub-resonant ($\beta < \beta_{res}$) or in the upstream direction if flow is super-resonant ($\beta > \beta_{res}$). This picture is supported by experimental observations by Zolezzi *et al.* [141] on an U-shaped flume, where a 180° bend determines the forcing discontinuities triggering steady patterns. Recently, Crosato *et al.* [38] have performed experimental runs in a straight flume having a localised half-width constriction at the inlet section in order to investigate the formation of steady bars downstream of the constriction.

In this section we aim at proving that the GIAMT2D model is able to correctly reproduce the morphodynamic influence of a localised forcing disturbance by developing steady bar patterns in agreement with the theory of morphodynamic influence. We conduct our numerical simulations in a straight domain with a localised constriction, analogous to that in the laboratory experiments of Crosato *et al.* [38]. The obstacle size is half the channel width and is located halfway from the upstream and downstream end of the channel in order to observe the development of steady bars in both directions. A zoom of the computational domain close to the constriction is given in Fig. 7.9.

We present the results of two runs: one is conducted in sub-resonant condition, the other one in super-resonant condition. In both cases, the parameters defining the reference uniform flow are $\theta = 0.100$ and $d_s = 0.067$ and the bedload transport formula of Wong and Parker [135] is applied. The critical and resonant values of aspect ratio predicted by the linear theories for these values of d_s and θ are $\beta_{cr} = 8.06$ and $\beta_{res} = 13.28$. In the sub-resonant run we use $\beta = 10.42$ while in the super-resonant test we use $\beta = 16.72$. In both cases we set channel slope $S_x = 0.011$ and $r_{ik} = 0.5$.

Starting from an initially flat bed configuration $\eta_0(x) = -S_x x$, we follow the time evolution of the riverbed configuration. At the beginning of the simulation a pronounced alternate perturbation appears downstream of the obstacle due to local scour both in the sub-resonant and super-resonant case. This alternate perturbation quickly amplifies and migrates downstream while elongating. At the same time, albeit at slower pace, in the upstream channel free bar formation and downstream

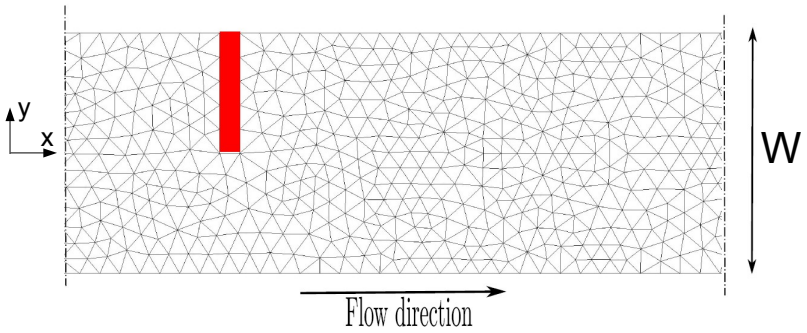

Figure 7.9: Numerical simulation of morphodynamic influence. Sketch of the straight computational domain with a localised constriction.

migration is observed. This is particularly evident in the super-resonant run. These migrating bars are able to pass through the constricted section and interact also with the downstream forced bars. Later, only in the super-resonant case, we observe that migrating bars coming from the upstream reach tend to slow down and eventually stop as they reach the constricted section. These steady bars are longer than typical free alternate bars.
During all the simulation, we observe bed aggradation associated with one-dimensional non-uniformity in bedload transport capacity due to the constriction. The characteristic time of aggradation is well described by the morphological scale T_M derived in (6.26), from the Exner one-dimensional balance. We remark however that the longitudinal adaptation is too weak to alter the values of θ, d_s and β appreciably and therefore does not seem to significantly affect two-dimensional morphodynamic influence. For both configurations, the final almost-steady topographic configuration is reached at about $T \sim 10\,T_M$.

The difference between the final riverbed configuration in the sub- and super-resonant case is reported in Fig. 7.10. Here we represent in terms of contourplots the variation in riverbed elevation $\Delta\eta$ with respect to the initial setup obtained at dimensionless time $T = 12.3$ (6.27) in the sub-resonant (SUB) and $T = 13.1$ in the super-resonant (SUPER) run. The bed elevation variation has been preliminarily filtered of one-dimensional variations and is made dimensionless with the reference to uniform flow depth D_0. While in both runs a steady bar pattern is found downstream of the constriction, only in the super-resonant case a steady pattern is observed *upstream* of the constriction.

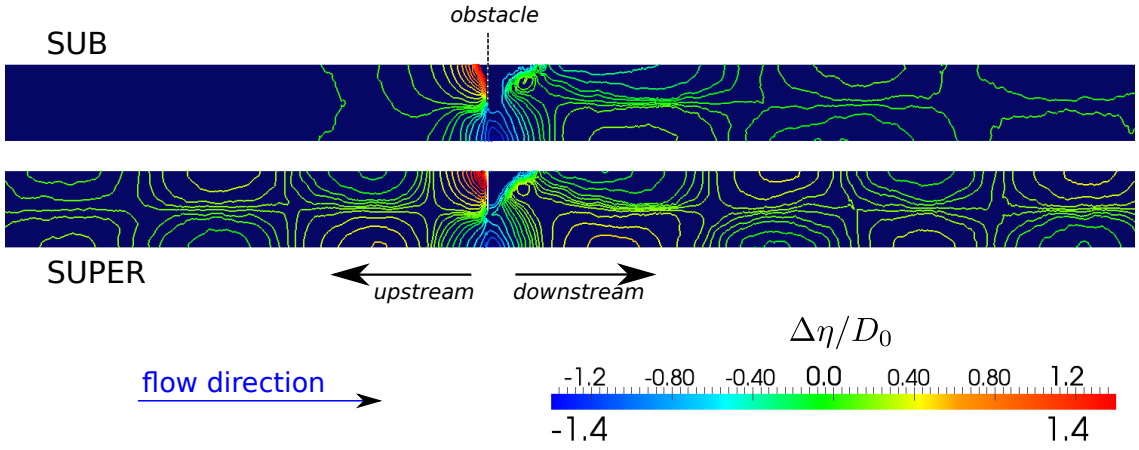

Figure 7.10: Numerical simulation of morphodynamic influence. The bed variation with respect to the flat initial condition scaled with uniform flow depth $\Delta\eta/D_0$ is computed at dimensionless time $T = 12.3$ (6.26)-(6.27) in the sub-resonant (SUB) and $T = 13.1$ in the super-resonant (SUPER) run. Planimetric axes are distorted.

In Fig. 7.11 we plot the riverbed configuration obtained in the super-resonant case at dimensionless time $T = 13.1$. The solution is represented by two longitudinal profiles taken close to the right, unperturbed straight bank (blue line) and to the left bank with the obstacle (red line) respectively. The alternate steady bar pattern is evident both downstream and upstream of the constriction. These results agree with the theoretical suggestions concerning morphodynamic influence from the linear theory of Zolezzi and Seminara [141] and the experimental observations of Zolezzi *et al.* [140], suggesting that the GIAMT2D model correctly predicts the direction of morphodynamic influence.

Besides representing an important morphodynamic test, this is a preliminary outcome on a topic that deserves further investigations. A key question is related to the nature of the steady bars we have observed. Are they purely forced bars in the sense of the spatial bars in Zolezzi and Seminara [141] or instead free bars which have progressively ceased their migration due to the effect of the constriction on sediment transport? A possible way to answer is to measure their length. From the theory we know indeed that the wavelength of forced bars is $L \sim 15 \div 20\,W$, corresponding to wavenumber $\lambda_s = 0.15 \div 0.2$, i.e. roughly twice as long as free migrating bars. For the set of data used in the present simulations, the theory predicts $\lambda_s = 0.17$ corresponding to $L = 18\,W$ and $\lambda_{cr} = 0.46$, corresponding to $L = 6.8\,W$. Concerning our numerical results, in the sub-resonant case we measure $L \sim 16\,W$, while in the super-resonant case (Fig. 7.11) we obtain $L \sim 11\,W$ in the downstream reach and $L \sim 9\,W$ in the upstream reach, corresponding to $\lambda = 0.28$ and $\lambda = 0.35$

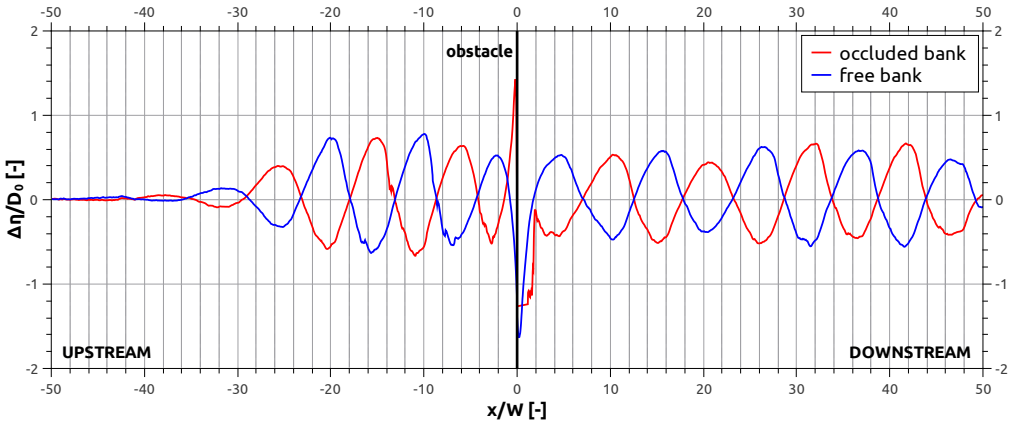

Figure 7.11: Numerical simulation of morphodynamic influence. The dimensionless bed variation with respect to the initial configuration $\Delta\eta/D_0$ obtained in the super-resonant case, computed at dimensionless time $T = 13.1$ (6.26)-(6.27), is sliced along the right and left channel bank. The longitudinal coordinate is $x/W = 0$ in the constricted section.

respectively.

These bars are too long to be considered free bars, but too short to be purely forced bars. A possible interpretation of this discrepancy is that they result from the non-linear interaction between purely free and forced bar components. Such interaction combines the natural tendency to the formation of free bars (notice that both simulations are conducted in β super-critical conditions) with the tendency to the development of forced bars. This seems reasonable because both types of bars are theoretically predicted to occur in the straight reaches separated by the constriction.

No theory or modelling approach has been developed so far to predict the dynamics of such interaction in the basic straight channel configuration. To test the interaction hypothesis, in the next chapter the GIAMT2D model is integrated with the linear theories and remotely-sensed field observations.

7.4 Conclusions

In this chapter we have presented applications of the GIAMT2D numerical model to well-established problems in river morphodynamics, namely the development of free bars in straight channels, the dynamics of river bifurcations and the assessment of the direction of two-dimensional morphodynamic influence associated to a localised disturbance. These cases, which had already been studied by classical morphodynamic approaches (analytical models and laboratory experiments) provide morphodynamic benchmarks for our numerical model. From the results presented

in this chapter we can state that the numerical model well reproduces the basic two-dimensional mechanisms responsible for the morphodynamic evolution of single-thread and bifurcating channels. Bars and bifurcations constitute basic morphodynamic units composing real river systems: their evolution and interaction plays a key role in determining the morphodynamic changes in natural rivers.

8 Non-linear interaction between free and forced bars in straight channels

In the previous chapter we have performed extensive validation of the GIAMT2D numerical model against well-established results of analytical theories and laboratory experiments in river morphodynamics. Once ascertained that GIAMT2D is able to correctly reproduce the salient features of basic morphodynamic processes such as bars and bifurcations and to predict the direction of morphodynamic influence associated to a localised disturbance, we can confidently apply it in order to carry out an original study in morphodynamics.

We focus on the steady bar patterns generated by a local disturbance in a straight channel, obtained by numerical simulations in section 7.3 of chapter 7. The starting point for the present study is the left-open question concerning the nature of these bars: they have been found longer than free bars, but too short compared with the forced bars in the linear theory of Zolezzi *et al.* [141]. In this chapter we will observe the same behaviour in the steady bar pattern in a channelised reach of the Rhine river bordering Switzerland and Lichtenstein. These observations will be integrated by a theoretical analysis of the potential interaction dynamics between free and forced bars and their consequences in determining the final steady bar pattern associated to a localised disturbance.
Although existing morphodynamic theories provide solutions for free bars or forced bars as individual concepts, or even for their interaction in periodic geometrical configurations, understanding these interactions in general channel configurations is still out of the capabilities of analytical modelling (see section 6.2.3 of chapter 6 and references therein). In the present chapter, merging the outcomes of the free bar theory of Colombini *et al.* [33] with those of the theory of forced bars [141], we argue that two parameters control the interaction, namely the channel aspect ratio β and the distance between the resonant and critical threshold $\beta_{res} - \beta_{cr}$ or its relative value $\frac{\beta_{res}-\beta_{cr}}{\beta_{res}}$. While the role of β in controlling the development of free and forced bars is widely accepted, the role of the distance $\beta_{res} - \beta_{cr}$ in favouring the free or forced components in the interaction is quite a novel and interesting concept. We further validate our theoretical speculations by means of numerical modelling, using the fully non-linear GIAMT2D solver. The research hypothesis developed on the

basis of the existing theories are satisfactorily confirmed by the numerical simulations. Finally, we check our results against the field data of the Rhine river obtaining good agreement. These results have been first presented in [132]. A full journal paper is in preparation [142].

The chapter outlines as follows. In section 8.1 we introduce the physical problem moving from the case of the Rhine river and highlight the need for a theory of free-forced bar interaction, which is presented in section 8.2. Finally in section 8.3 we prove our theoretical speculations by means of numerical modelling and comparison with the field data.

8.1 Interaction between free and forced bars in the Rhine river

As common to many morphologically regulated rivers worldwide, the considered channelised reach of the Rhine river has been designed as a sequence of straight reaches of varying length and constant channel width. In Fig. 8.1 the reach under investigation is represented. Consider the mildly curved bend with uniform radius highlighted with a yellow ellipse in the figure. Upstream of that bend, the river is almost straight for about $9.5\,km$, thus reasonably allowing for complete development of free migrating bars. Downstream of the planar discontinuity, a bar pattern appears again. A couple of kilometres downstream of the considered bend, flow discharge abruptly increases due to a lateral tributary and channel width decreases, thus decreasing the channel aspect ratio and suppressing free bar instability.
Fig. 8.2 illustrates the initial and final alternate bar configuration over a nine-year time span. The background represents the pattern observed in October 2009, while that in December 2000 is represented by contour lines. Notice that between 2000 and 2009 the bars in reach U (upstream of the bend) have migrated for about $300\,m$ while the bars in reach D (downstream of the bend) have remained fixed.

These bar configurations allow us to draw some hypotheses about the underlying morphodynamics in the Rhine river. The presence of free migrating bars in the upstream reach suggests that channel aspect ratio in formative conditions might be super-critical ($\beta > \beta_{cr}$). The observed steady bar pattern in reach D could be triggered by the planimetric disturbance represented by the localised curvature discontinuity at the downstream end of the bend. Since the forced bars appear only downstream of the bend and not upstream, recalling the concept of morphodynamic influence [141] we hypothesise that flow is in sub-resonant condition ($\beta < \beta_{res}$).
These hypotheses are tested by computing the channel aspect ratio and its critical and resonant value in formative conditions. We assume channel width $W \sim 100\,m$ from aerial geo-referenced images. Channel slope is estimated as $S \sim 0.003$. Flow rate statistics available at the Swiss Federal Office

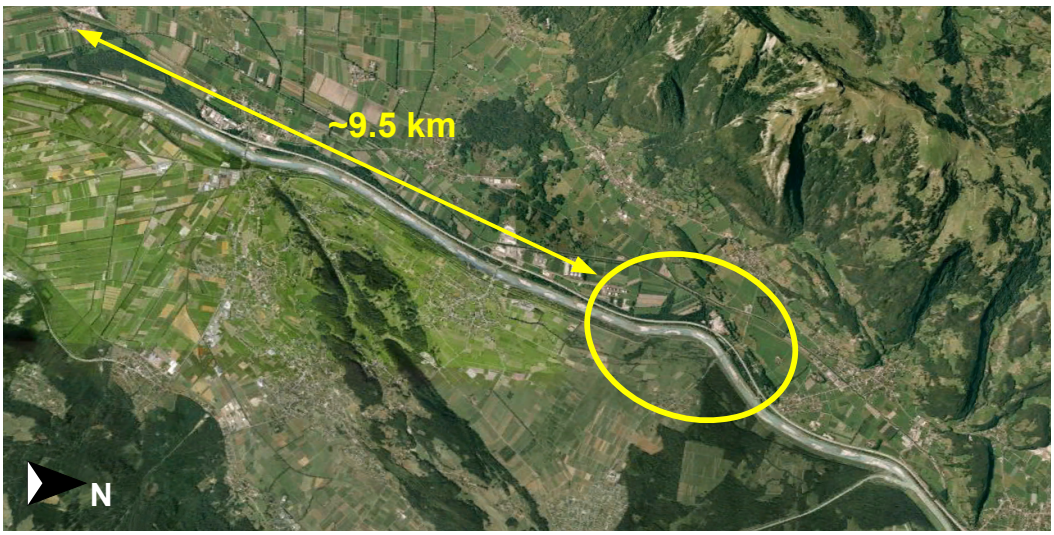

Figure 8.1: Wide aerial photography of a reach of the Rhine river bordering Switzerland and Lichtenstein.

for the Environment[1] allow us to calculate the formative discharge as $1730\,m^3s^{-1}$. An estimate for the mean sediment diameter is $D_s \sim 0.05\,m$ (courtesy of Dr. Jaeggi[2]), which gives the Manning and Strickler friction parameters $n^{-1} = Ks = 35\,m^{1/3}s^{-1}$ using relation (2.11). From this set of data we compute uniform flow water depth $D_0 \sim 3.7\,m$, which gives channel aspect ratio $\beta = 13.5$. Moreover, from the dimensional data we obtain $\theta = 0.13$ and $d_s = 0.014$, for which the linear free bar theory predicts $\beta_{cr} = 11.9$ and $\beta_{res} = 21.0$ adopting the bedload transport formula of Wong and Parker [135]. Thus, the considered reach in formative flow condition is in bar super-critical and sub-resonant regime, which confirms the initial hypotheses made on the basis of the multi-temporal image analysis.

We shall now compare the theoretically predicted bar wavenumbers to those obtained by field observations. For the above values of parameters, the linear theory gives $\lambda_{cr} = 0.41$ and $\lambda_{res} = 0.13$. From the aerial geo-referenced snapshot in Fig. 8.2 we measure bar wavelength $L = 855\,m$ in reach U and $L = 1530\,m$ in reach D, which correspond to $\lambda = 0.40$ and $\lambda = 0.22$ respectively. Concerning the migrating bars in reach U, very good agreement between the predicted wavelength of free bars

[1] http://www.bafu.admin.ch/hydrologie/01834/02049/index.html?lang=de
[2] http://www.rivers.ch

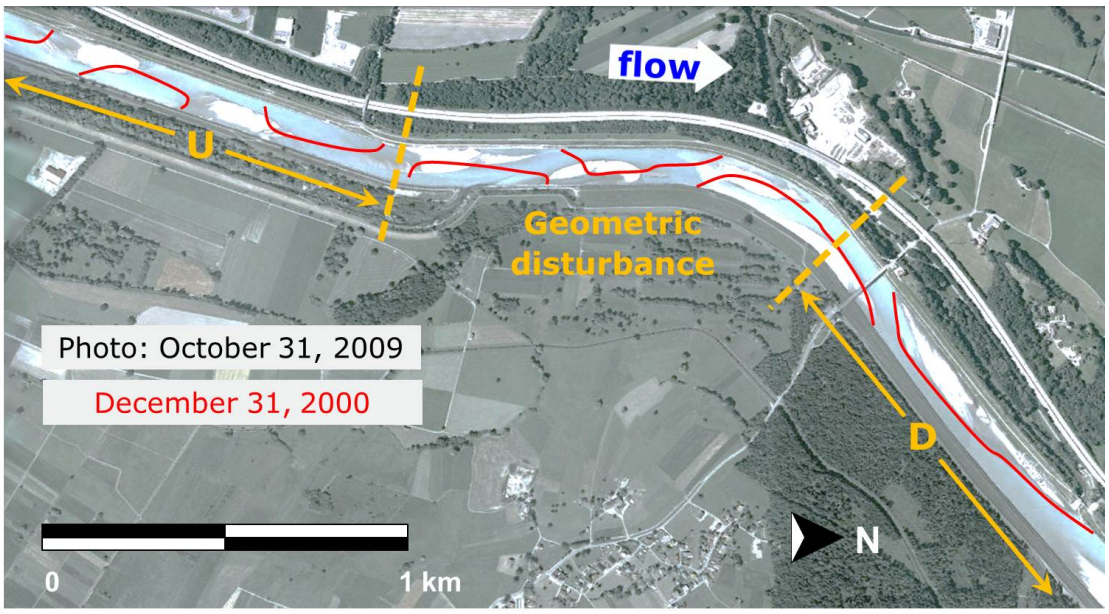

Figure 8.2: Location of alternate bars in the considered reach of the Rhine River. Red contours indicate the bar pattern observed in December 2000 and the background image refers to the pattern in October 2009. During this period, in the upstream straight segment U shorter alternate bars have been migrating downstream, while longer alternate bars in the downstream straight segment D have been keeping fixed, both being subjected to the same formative conditions.

and the measured wavelength is found. This is coherent with the finding that the reach is under sub-resonant conditions, indicating that migrating bars should be the only type of bars to be found upstream of the bend.

Let us then focus on the steady bar pattern in reach D, where theoretically both steady and migrating bars can occur. The nature of these bars poses a problem analogous to that discussed at the end of section 7.3, in relation to morphodynamic influence. Our field observations show that the bars in reach D present intermediate wavelength between that of free and forced bars, being longer than free bars, but shorter than the purely forced bars described by the theory of Zolezzi and Seminara [141]. Thus, analogously to our numerical results on two-dimensional morphodynamic influence (section 7.3), we hypothesise that the steady bars observed on the Rhine river arise from the non-linear interaction of free and forced bars and that their final wavelength is controlled by the relative "strength" of these components in the natural competition.

This field example highlights the need for a non-linear theory able to account for these interactions in order to explain the actual bar dynamics in real gravel-bed rivers. However, as we have

seen in section 6.2.2 of chapter 6, a morphodynamic theory able to account for these processes in the basic straight-channel configuration does not exist. We choose to revisit the outcomes of the existing linear theories, in order to develop a unified interpretation framework. This is achieved in the next section.

8.2 Re-examination of linear theories

Here we re-examine the linear theories of free bars and forced bars exposed in sections 6.2.1 and 6.2.2 of chapter 6 in order to merge their most relevant outcomes and develop a framework useful for interpreting free and forced bar interaction in straight reaches. Herein we argue that the interaction process can be viewed as a competition between free and forced bars. We therefore revisit these theories in order to assess how the relative "strength" of free and forced bars can depend on the controlling parameters.

We recall from the two theories the role of channel aspect ratio β compared to the critical and resonant thresholds β_{cr} and β_{res} in controlling the bar dynamics.
Free bars are able to develop only if β exceeds β_{cr}; their amplitude from the theory of Colombini *et al.* [33] increases with the distance $\beta - \beta_{cr}$ accordingly with equation (6.12). We claim that the strength of free bars can be represented by their amplitude, i.e. free bars producing larger scour-deposition sequences can increasingly dominate the competition with steady bars. Therefore, the distance $\beta - \beta_{cr}$ is here identified as the controlling parameter for the strength of free bars.
Measuring the strength of steady bars requires a different approach, because their amplitude is related to the type of forcing disturbance. Because these bars are damped in space, a convenient inherent measure is their spatial damping rate λ_{sr}. Aspect ratio β compared to its resonant value β_{res} is the parameter controlling characteristic damping length and direction. In the sub-resonant region ($\beta < \beta_{res}$) bar amplitude shall decay in the downstream direction, while in the super-resonant region ($\beta > \beta_{res}$) it shall decay in the upstream direction. When aspect ratio approaches the resonant value, the characteristic damping length tends to infinity, i.e. the morphodynamic influence of a disturbance does not get spatially damped, thus maximising the spatial occurrence of forced bars and therefore their strength in the interaction with free bars. Therefore we assume the distance $|\beta - \beta_{res}|$ as inverse measure of the strength of forced bars, no matter in this case whether in sub- or super-resonant condition.

What does the combination of the two theories tell us concerning the *interaction* of free and forced bars? We preliminarily observe that β_{res} is always higher than β_{cr}. Consider Fig. 8.3, where the outcomes of the theory of forced bars (left panel) and of the theory of free bars (right panel) are

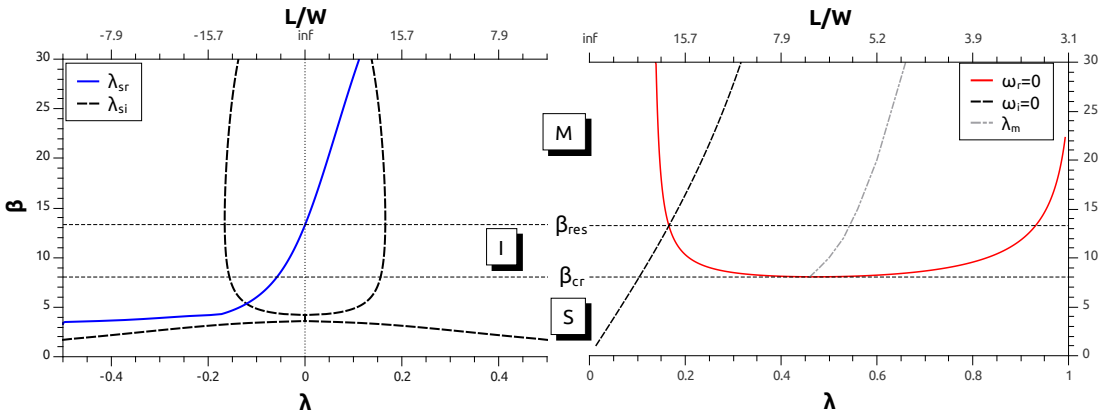

Figure 8.3: Variation of expected steady and migrating bar wavelength with channel aspect ratio according to the linear theory for steady forced bars (left panel) and free migrating bars (right panel). Three regions are identified. *S: Steady region* $\beta < \beta_{cr}$: steady bars are expected to grow and develop in presence of a local disturbance. *I: Interaction region* $\beta_{cr} < \beta < \beta_{res}$: free and forced bars can interact at non-linear level. *M: Migrating bar region* $\beta > \beta_{res}$: free migrating bars tend to overcome possible steady bars.

jointly represented. On this basis, in Fig. 8.3 we propose to distinguish three distinct regions in which we expect different interaction dynamics:

- *S: Steady bar region*, $\beta < \beta_{cr}$. No migrating bars occur for $\beta < \beta_{cr}$ and therefore only steady bars can develop in this region.

- *I: Interaction region*, $\beta_{cr} < \beta < \beta_{res}$. The extent of this region, defined as the distance $\beta_{res} - \beta_{cr}$, is expected to control the relative importance of migrating bars and forced bars in the interaction. In detail, if such extent is low, migrating bars will play a relatively minor role in the interaction, while if it is high they are expect to govern the interaction. This is readily explained starting from the above considerations about the distances $\beta - \beta_{cr}$ and $|\beta - \beta_{res}|$. We have assumed that the strength of free bars increases with the distance $\beta - \beta_{cr}$, while the strength of forced bars increases with proximity to resonance, i.e. with decreasing $|\beta - \beta_{res}|$. Thus if the width of the *Interaction region* $\beta_{res} - \beta_{cr}$ is low, forced bars are favoured by the low values of $|\beta - \beta_{res}|$, while the low values of $\beta - \beta_{cr}$ are unfavourable for free bars. Conversely, if the *Interaction region* is wide, free bars are expected to prevail in the interaction.

- *M: Migrating bar region*, $\beta > \beta_{res}$. For increasing values of β larger than β_{res}, the amplitude of free migrating bars increases along with the spatial damping of steady bars; in this region migrating bars are expected to increasingly dominate the interaction as β increases.

We therefore hypothesise that the main morphodynamic parameters controlling steady-migrating bar interaction are aspect ratio β and the difference $\beta_{res} - \beta_{cr}$ computed under bar formative conditions. The fully non-linear GIAMT2D numerical model is a suitable tool for studying the considered non-linear processes. Thus in the next section we will test our hypotheses by means of numerical modelling. Moreover, we will use the field case presented in section 8.1 for the validation of our study.

8.3 Steady-migrating bar interaction in a straight channel with a localised constriction

8.3.1 Simulation setup

We consider again the channel configuration adopted in section 7.3. The domain is represented by a straight channel with a localised constriction obstructing half of the channel width (see Fig. 7.9). We perform three sets of numerical simulations, which we denote with T1, T2 and T3. For all sets, we use the bedload transport formula of Wong and Parker [135] (see Tab. 2.1). Each of these sets of runs is characterised by a unique combination of θ and d_s, which gives a unique value for the critical and resonant aspect ratio (β_{cr} and β_{res}) per set. We wish to investigate whether the interaction of free and forced bars is controlled by the distance between the resonant and critical aspect ratio $\beta_{res} - \beta_{cr}$, as we have hypothesised in the previous section. Thus the combinations of θ and d_s are chosen so as to obtain different values for such distance. The non-dimensional parameters for the three sets and the related critical and resonant threshold are summarised in Tab. 8.1.
Within the same set, several runs have been performed varying the value of aspect ratio β. These values are reported in Tab. 8.2. This procedure allows us to explore the three regions defined by the two thresholds β_{cr} and β_{res}. Notice that all runs are in β super-critical conditions ($\beta > \beta_{cr}$), i.e. in all runs free bars are expected to play a role in determining the final steady bar wavenumber. In order to keep d_s and θ constant within each set of runs, even with varying uniform flow depth, the absolute sediment size D_s changes in each run.

Furthermore in Tab. 8.1 we report the uniform flow Froude number F_r of the three considered configurations, given by

$$F_r = C_h \sqrt{\theta \Delta d_s} = C_h \sqrt{S}\,, \tag{8.1}$$

where the Chézy coefficient $C_h(d_s)$ has been estimated using the logarithmic relation (2.13). Notice that all the three sets of simulations are characterised by trans-critical values of the Froude number $0.8 \lesssim Fr \lesssim 1.2$. As we have explained in section 5.1 of chapter 5, the numerical solution of

hydro-morphodynamic problems in trans-critical flow conditions requires a coupled approach. Thus, the present application once again motivates the choice of a coupled solution strategy in the GIAMT2D numerical solver.

parameters	*T1*	*T2*	*T3*
θ	0.10	0.06	0.15
d_s	0.067	0.111	0.044
r_{ik}	0.5	0.5	0.5
S	0.011	0.011	0.011
F_r	1.10	0.97	1.20
β_{cr}	8.06	4.63	9.64
β_{res}	13.28	6.24	19.33
λ_{cr}	0.46	0.47	0.43
λ_{res}	0.17	0.15	0.15

Table 8.1: Steady-migrating bar interaction in a straight channel with a localised constriction. Parameters and thresholds of the sets T1, T2 and T3.

8.3.2 Results and discussion

In all runs we have observed similar evolution dynamics. Migrating bars develop in the reach located upstream of the obstacle and are initially able to pass through the constricted section. The development of the steady bar pattern is usually slower. At some stage in the downstream reach the migrating bars slow down and tend to elongate while in the upstream reach the migrating bars tend to be stopped by the obstacle. Both in the upstream and downstream reach the length of the resulting bars changes in different runs, depending on the reference flow aspect ratio β.

In Fig. 8.4 we observe time evolution of the steady bar wavenumber measured downstream

run	*T1*	*T2*	*T3*
1	10.42	6.52	14.29
2	12.16	7.46	18.06
3	14.56	8.45	22.53
4	18.06	9.89	23.98
5	20.03	-	26.55

Table 8.2: Steady-migrating bar interaction in a straight channel with a localised constriction. Values of aspect ratio β in the simulations in sets T1, T2 and T3.

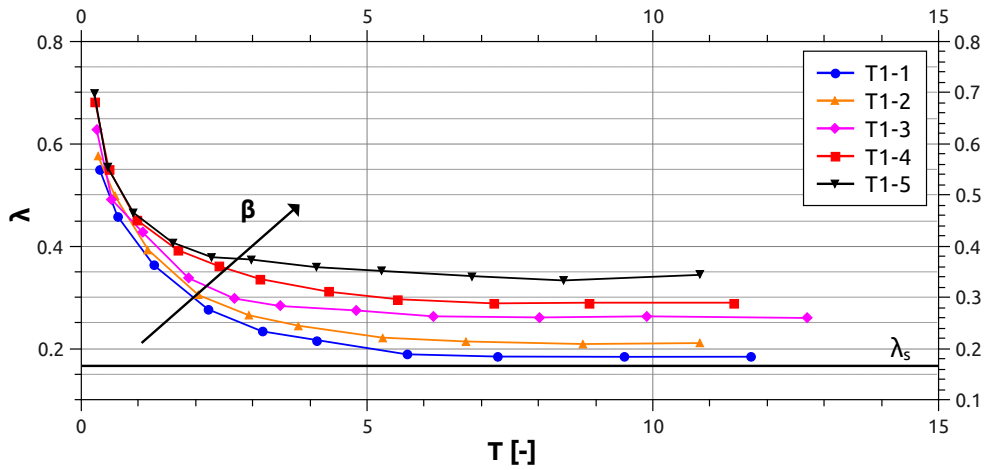

Figure 8.4: Steady-migrating bar interaction in a straight channel with a localised constriction. Time evolution of the first downstream bar wavenumber for the runs of set T1.

of the constriction for the runs in set T1. Bars are initially very short ($\lambda \sim 0.55 \div 0.7$) and tend to elongate in time, finally reaching a steady state. The time needed to reach a steady bar pattern is found around ten times the morphological time scale (6.27). The final wavenumber is found to be an increasing function of the reference uniform flow aspect ratio β of each run (see Tab. 8.2): for higher values of β shorter bars are found, having wavelength closer to that of free bars. This suggests that, for increasing β, free bars tend to play a progressively dominant role, in agreement with our speculations presented in the previous section. Analogous results are obtained for sets T2 and T3.

We shall now interpret our numerical results by analysing the values of these equilibrium wavenumbers. In Fig. 8.5 we report in the $\beta - \lambda$ plane the steady-state values of bar wavenumber for the five runs in set T1 by circular symbols, together with the marginal curves of the free-bar theory (red and dashed black line), the curve corresponding to the fastest-growing free bars (light-blue dashed line) and the wavenumber of forced bars (blue dashed line). Full symbols represent the wavenumber of the bars observed downstream of the constriction, while empty symbols represent the bars observed upstream (if present, i.e. only in super-resonant runs).
As we have observed in Fig. 8.4 concerning steady-state wavenumbers, a consistent trend relating wavenumbers and aspect ratio can be detected. The bar wavenumber progressively increases with increasing values of β, tending to the most unstable free bar wavenumber λ_m predicted by the linear theory. Such behaviour can be interpreted recalling that the amplitude of free bars, i.e. their strength in the interaction with the forced component, increases with the difference $\beta - \beta_{cr}$. Larger

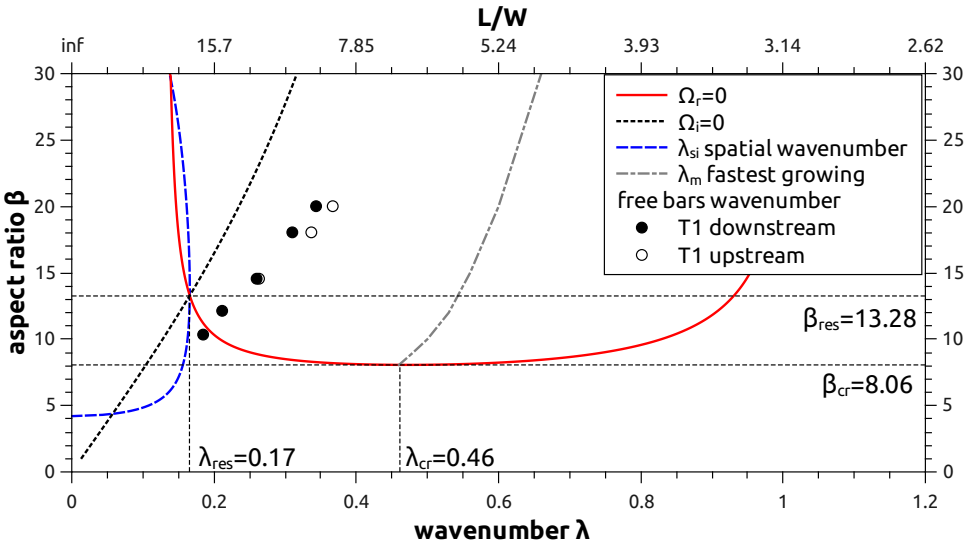

Figure 8.5: Steady-migrating bar interaction in a straight channel with a localised constriction. The wavenumber of steady bars measured in the runs of set T1 is plotted as function of aspect ratio of uniform flow β (symbols), together with the curves indicating the spatial bar wavenumber λ_{si}, the fastest-growing bar wavenumber λ_m and the marginal curves of the free bar theory. Full symbols are associated to the bars developed downstream of the constriction, empty symbols to those upstream of the constriction.

amplitudes reasonably imply that free bars progressively dominate in the competition, thus shifting the wavenumber of the equilibrium bar pattern towards the wavenumber of the fastest growing free bar. Conversely, when the aspect ratio decreases, the wavenumber of the resulting bar configuration progressively tends to that of the steady bars.
This trend is detected for the bar wavelengths measured both upstream and downstream of the constriction. However, the strength of the free-bar component is found to be higher in the upstream case compared to the downstream case. This may be due to stronger interaction of the free bars in the upstream reach, possibly because in this case free bars are allowed to directly collide with the constricted section, while in the downstream case the propagation of free bars from the upstream reach is partially prevented by the constriction.

In Fig 8.5 we notice that the theoretical wavenumber of steady bars $\lambda_{si}(\beta,\theta,d_s)$, represented by a blue dashed line, is very well approximated, for all the values of β corresponding to the numerical runs (symbols), by the resonant wavenumber $\lambda_{res}(\theta,d_s)$, which takes the value $\lambda_{res}=0.17$ in set T1. In the following, for sake of clarity in the graphical presentation of our results, we will use λ_{res} in place of λ_{si}, thus dropping the (very weak) dependence of the steady bar wavenumber on the

aspect ratio value of each run.

Next, we consider the role of the second control parameter, i.e. the distance between β_{res} and β_{cr}, focusing on the bar wavenumbers measured downstream of the constriction for the three sets T1, T2 and T3. We recall that these sets are characterised by different values of θ and d_s, giving different values of β_{cr} and β_{res}. Results are represented in Fig. 8.6 in terms of wavenumber λ and relative distance of aspect ratio from the resonant value $\frac{\beta-\beta_{res}}{\beta_{res}}$. In Fig. 8.6 we highlight by horizontal full lines the values of $\frac{\beta_{cr}-\beta_{res}}{\beta_{res}}$ for each set of runs, related to the amplitude of the *Interaction regions* I_1, I_2 and I_3. Moreover we plot by dashed lines the theoretical wavenumber of steady bars λ_{res} and of free bars λ_{cr}. Preliminarily we notice that all the sets show a consistent trend with β, analogous to that obtained for T1 in Fig. 8.5. Then, we focus on the range of wavenumbers associated to each set.
We observe that the runs of set T2 (red symbols) give a wavenumber close to λ_{res}, while those of T3 (blue symbols) give a wavenumber closer to λ_{cr} and T1 gives intermediate wavenumbers between T2 and T3. Such behaviour can be interpreted considering the value of the relative distance $\frac{\beta_{cr}-\beta_{res}}{\beta_{res}}$ associated to each set and related to the amplitude of the *Interaction region*. We have argued that this is a good measure of the strength of free and forced bars in their mutual interaction (see section 8.2). Set T3 is associated to the widest *Interaction region*, set T2 to the narrowest.
Higher wavenumbers, closer to those of free bars, are found for sets of simulations with wider *Interaction region*, thus suggesting that the larger the *Interaction region* is, the stronger free bars are in the non-linear mix. This substantiates the theoretical speculations in section 8.2.

In Fig. 8.7 we plot the same wavenumbers as in Fig. 8.6, measured downstream of the constriction (full symbols), together with the wavenumbers of the bars observed upstream (empty symbols). Due to intense interaction with free bars upstream of the constriction, it is more difficult to identify a clear trend concerning the simulations carried out in super-resonant conditions. Generally the upstream wavenumbers are higher than the downstream wavenumbers, thus supporting the idea that the signal competition might be stronger in the upstream case than in the downstream case.

Finally, we aim at re-plotting our results in terms of the uniform flow parameters β and F_r. The latter parameter can be obtained from the Shields number θ, relative grain size d_s and the Chézy coefficient C_h in the form (8.1), where we assume $C_h = C_h(d_s)$ through the logarithmic relation (2.13). In the following, we take d_s as constant, thus obtaining from equation (8.1) that $F_r = F_r(\theta)$ in the form

$$F_r \propto \sqrt{\theta}\,. \tag{8.2}$$

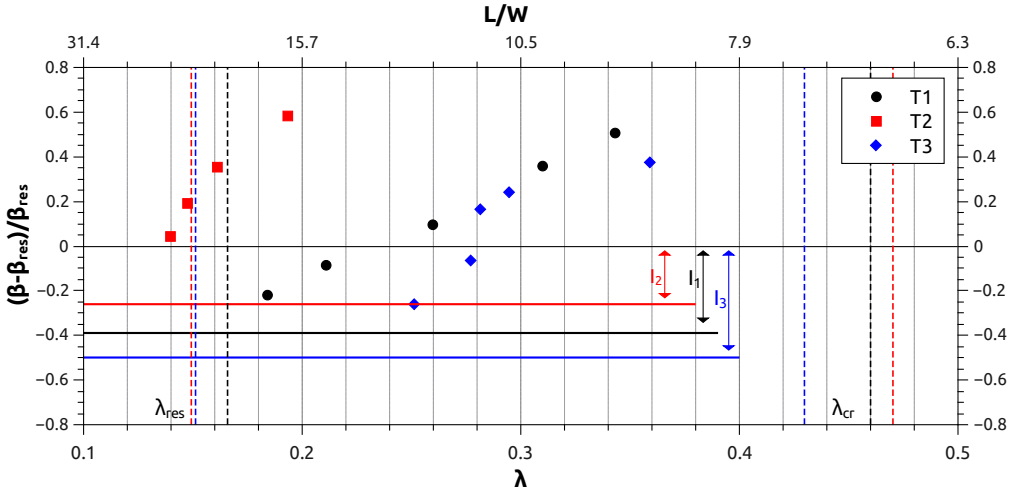

Figure 8.6: Steady-migrating bar interaction in a straight channel with a localised constriction. The wavenumbers of three sets of runs T1, T2 and T3 measured downstream of the constriction are plotted together with the theoretical wavenumber of spatial bars (approximately represented by λ_{res}) and of free bars (approximately represented by λ_{cr}) (dashed lines) in the plane $\lambda - \frac{\beta-\beta_{res}}{\beta_{res}}$. The horizontal full lines denote the parameter $\frac{\beta_{cr}-\beta_{res}}{\beta_{res}}$ of each set, related to the amplitude of the *Interaction region* I_1, I_2 and I_3.

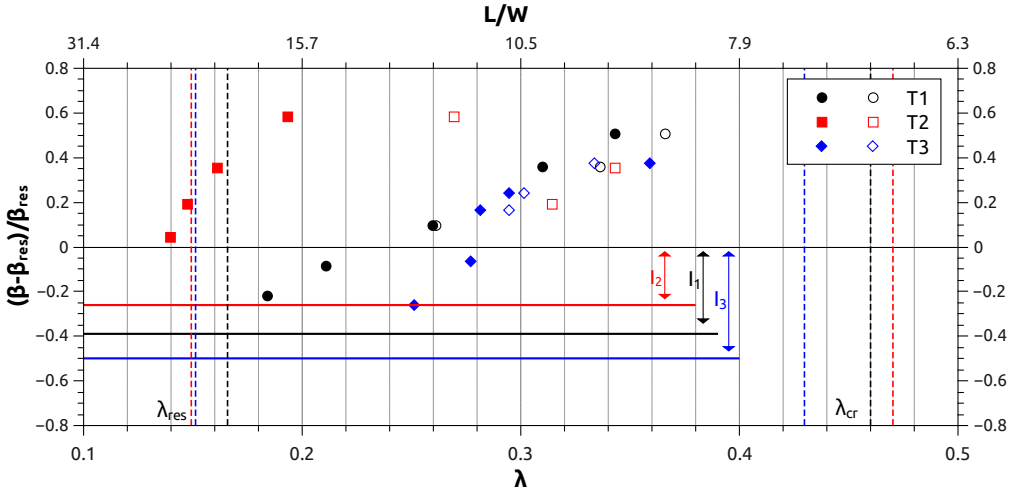

Figure 8.7: Steady-migrating bar interaction in a straight channel with a localised constriction. The wavenumbers of the three sets of runs T1, T2 and T3 measured downstream (full symbols) and upstream of the constriction (empty symbols) are plotted together with λ_{res} and λ_{cr} (dashed lines) in the plane $\lambda - \frac{\beta-\beta_{res}}{\beta_{res}}$. The horizontal full lines denote the parameter $\frac{\beta_{cr}-\beta_{res}}{\beta_{res}}$ of each set, related to the amplitude of the *Interaction region* I_1, I_2 and I_3.

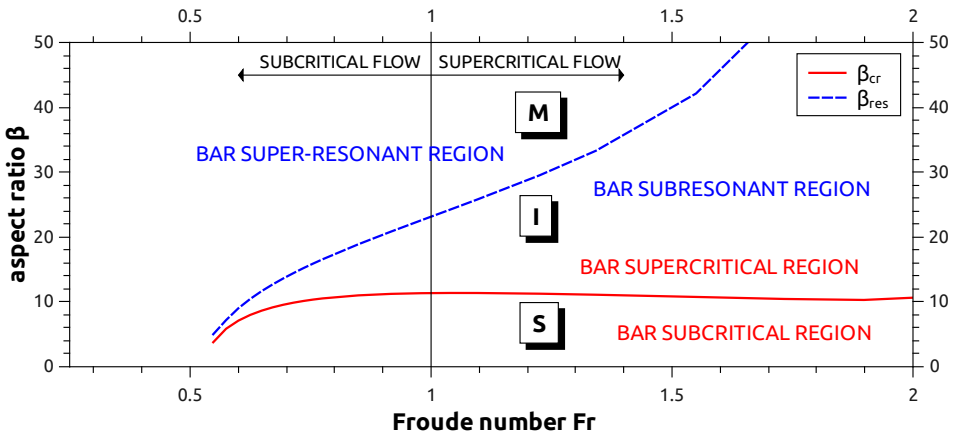

Figure 8.8: Critical and resonant aspect ratio β_{cr} and β_{res} as functions of the Froude number $F_r(\theta)$ (8.2). Three regions are identified: *S: Steady bar region*, *I: Interaction region*, *M: Migrating bar region*.

In general, critical and resonant aspect ratio β_{cr} and β_{res} depend on θ and d_s. With constant d_s, we are allowed to drop their dependence on this parameter, which enables us to draw the curves representing $\beta_{cr}(\theta)$ and $\beta_{res}(\theta)$ as functions of $F_r(\theta)$ (8.2) in the $F_r - \beta$ plane. This is achieved in Fig. 8.8, using $d_s = 0.02$. The critical threshold β_{cr} (red line) discriminates between the sub-critical and super-critical bar region, while the resonant threshold β_{res} (blue line) separates the sub- and super-resonant bar region. Thus, the three regions introduced in section 8.2, namely the *Steady bar region S* ($\beta < \beta_{cr}$), the *Interaction region I* ($\beta_{cr} < \beta < \beta_{res}$) and the *Migrating bar region M* ($\beta > \beta_{res}$) are easily identified in this representation. Fig. 8.8 clearly shows that the amplitude of the *Interaction region* decreases moving towards small Froude numbers and increases for high values of Froude.

In Fig. 8.9 our numerical results of three sets of runs (symbols) are plotted together with the critical and resonant thresholds (full and dashed line respectively). Notice that three sets of critical and resonant thresholds are represented, matching the dimensionless grain size of each set of runs (see Tab. 8.1). This representation does not carry any information about the bar wavenumbers, but gives immediate visualisation of the expected behaviour in terms of the free and forced bar interaction. It suggests that the influence of migrating bars in set T2 should be weak because the *Interaction region* is rather narrow for these runs. Set T3 instead, for which the *Interaction region* is much wider, is expected to show greater influence of free bars.

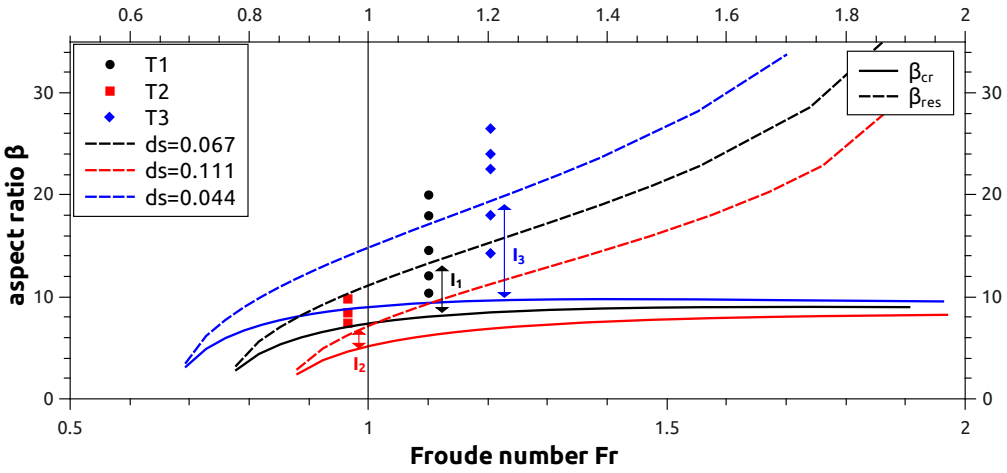

Figure 8.9: Steady-migrating bar interaction in a straight channel with a localised constriction. The numerical results of the T1, T2 and T3 sets are plotted in the $\beta - F_r$ plane together with the resonant and critical threshold of each set. I_1, I_2 and I_3 denote the amplitude of the *Interaction region* of the three sets.

8.3.3 The case of the Rhine river

Finally, we come back to the case of the Rhine river introduced in section 8.1 and compare these field observations with our theoretical-numerical framework. We focus on the wavelength of the steady bars observed downstream of the bend in Fig. 8.2. In section 8.1 we have observed that these bars, having wavenumber $\lambda = 0.22$, are longer than linear free bars ($\lambda_{cr} = 0.41$), but shorter than purely forced bars ($\lambda_{res} = 0.13$). This has suggested that they may result from their non-linear interaction.

Using the parameters given in section 8.1, we represent the Rhine river by a green star-shaped symbol in Fig. 8.10. With a green line we also represent the relative distance between the critical and resonant threshold $\frac{\beta_{cr}-\beta_{res}}{\beta_{res}}$ for the Rhine, representing the amplitude of the Rhine *Interaction region* I_{Rhine}. Moreover we plot the output of the numerical runs T1, T2 and T3 previously described and the associated lines representing the relative distance of the critical and resonant threshold for the numerical simulations. Here, concerning the numerical results we restrict our attention to those obtained downstream of the constriction. From the chart we notice that the case of the Rhine river is intermediate between that of runs T1 and T3 in terms of the distance $\frac{\beta_{cr}-\beta_{res}}{\beta_{res}}$, i.e. amplitude of the *Interaction region*. Moreover we observe that the wavenumber of the bars in the Rhine river approximately falls in the range of the chart between the results of sets T1 and T3, thus providing good agreement of the observations with the numerical results. This represents a confirmation of

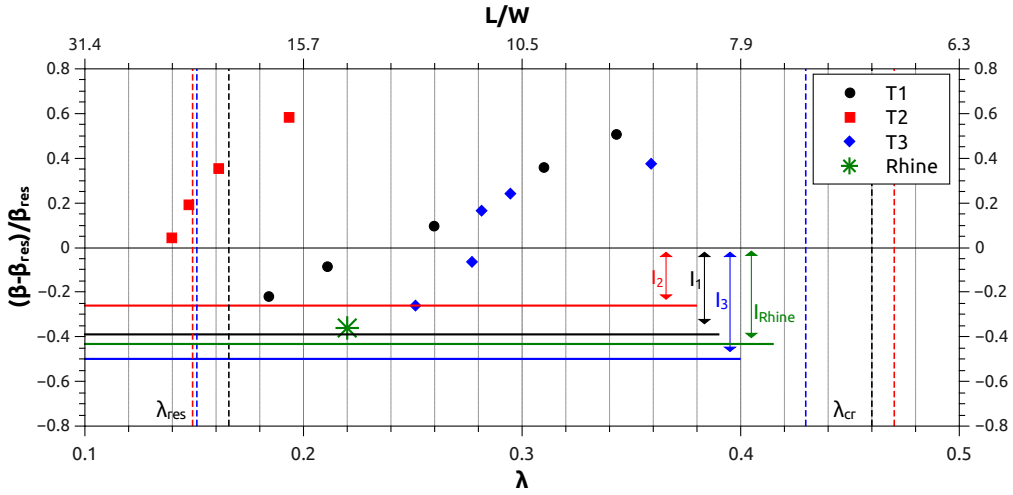

Figure 8.10: Steady-migrating bar interaction in a straight channel with a localised constriction. The wavenumber of the steady bar pattern of the Rhine river is plotted together with the wavenumbers of the three sets of numerical runs T1, T2 and T3 measured downstream of the constriction. The horizontal full lines denote the parameter $\frac{\beta_{cr}-\beta_{res}}{\beta_{res}}$ for the Rhine river (green line) and for each set of runs, related to the amplitude of the *Interaction region* $I_{\text{Rhine}}, I_1, I_2, I_3$.

the role of the parameter $\frac{\beta_{res}-\beta_{cr}}{\beta_{res}}$ in controlling the interaction of free and forced bars in nature, together with aspect ratio β. This field application provides further support to the suitability of the GIAMT2D numerical model to simulate the morphodynamics of gravel-bed rivers.

8.4 Conclusions

In this chapter we have conducted an original study in river morphodynamics by integrating numerical modelling, existing morphodynamic theories and remotely-sensed field data. The aim of this study was to address the non-linear interaction between free and forced bars in straight channels. We have hypothesised that the steady bar pattern arising from a forcing disturbance in a straight channel results from the interaction of a free and forced bar component. This was suggested by our previous observations on the wavenumber of the steady bar pattern generated by the GIAMT2D model in a straight channel with a localised constriction (section 7.3), is coherent with existing theoretical analyses and matches our observations of a steady bar pattern in an almost-straight channelised reach in the Rhine river from remotely-sensed field data.

Reviewing the bar theories, we argued that the channel aspect ratio β and the distance between the resonant and critical thresholds $\beta_{res} - \beta_{cr}$, or its relative value $\frac{\beta_{res}-\beta_{cr}}{\beta_{res}}$, are the two parameters con-

trolling the non-linear interaction between free and forced components in steady bars. For different values of aspect ratio three regions can be identified, namely the *Steady bar region* ($\beta < \beta_{cr}$), the *Interaction region* ($\beta_{cr} < \beta < \beta_{res}$) and the *Migrating bar region* ($\beta > \beta_{cr}$).

We have conducted three sets of numerical runs in a straight channel with a localised constriction and studied the resulting steady bars in order to highlight the relationship between their wavenumber and the above controlling parameters. Furthermore we have reconsidered the case of the Rhine river in light of the theoretical framework. The numerical and field results are fairly consistent with our theoretical speculations.

The applications presented in this chapter prove that the GIAMT2D model is a tool able to reproduce the highly non-linear morphodynamic behaviour of gravel-bed rivers and can also be used for the development and testing of theories in river morphodynamics.

9 Conclusions

This thesis concerns the development, testing and application of a two-dimensional numerical model, regarded as GIAMT2D, able to address the hydro-morphodynamic evolution of gravel-bed rivers. GIAMT2D solves the two-dimensional shallow water-Exner model, describing free-surface flow in open channels with erodible bed, where sediment transport mainly occurs as bedload. The model shall be generally applicable to natural and regulated gravel-bed river systems: thus the solution strategy has been developed having in mind the main distinctive features of gravel-bed rivers.
More specifically, the model is able to account for the two-dimensional (x, y) character of the flow and transport condition, which reflects the typical shallowness encountered in gravel-bed rivers, by employing a two-dimensional formulation of the mathematical problem. Planform complexity, resulting in irregularly-shaped computational domains, is handled using domain discretisation by unstructured meshes. An adaptive and robust wetting-and-drying procedure allows GIAMT2D to deal with the presence of possibly dry regions in the domain, whose extent and location may change in time accordingly with the unsteady flow regime and the hydro-morphodynamic evolution of the river system. Bedload transport and related morphodynamics in Froude trans-critical flow conditions, which are likely to occur under formative conditions in many gravel-bed rivers, is correctly taken into account by adopting a coupled formulation of the hydrodynamic and morphodynamic problem.

The coupled shallow water-Exner model results in a non-conservative system of PDEs, which is hyperbolic within the range of flow data typically found in rivers, to be solved by a shock-capturing path-conservative method in the framework of finite volumes. A key characteristic of the considered problem is that the solution of the Riemann problem is not readily available if complex, empirically-derived bedload transport formulae of common use in the engineering practice are applied in computations, thus making the upwind solution approach unsuitable: application of the centred approach, not requiring detailed knowledge of the system eigenstructure, is mandatory. The GIAMT2D model is based on a refined finite volume scheme of the centred type, namely UPRICE2-Cδ, in which an upwind bias, consisting of upwind information based on the system

eigenvalues, is included in computations. Absence of Riemann solvers and numerical integration of the Roe matrix (centred philosophy) allows the use of any bedload transport formula, while the insertion of an upwind bias significantly improves accuracy in the solution of contact waves and in computations carried out at low CFL number. The UPRICE-Cδ scheme results from original research in applied mathematics, concerning the development of upwind-biased centred schemes for conservative and non-conservative homogeneous hyperbolic systems of PDEs, whose performance is thoroughly assessed in a suite of classical numerical tests for the shallow water equations. These numerical applications show that the proposed schemes are able to improve accuracy with respect to standard centred methods in the solution of contact waves (shear waves, contact discontinuities), in the computations performed at low CFL and in general on irregular meshes, where a wide range of CFL number is induced by mesh variability. In these cases, the solution of upwind-biased methods is comparable to those obtained by upwind methods based on incomplete Riemann solvers.

The model, based on the above path-conservative upwind-biased scheme extended to second-order accuracy in the ADER framework, is first tested employing well-established test cases with fixed and movable bed conducted in idealised configurations. These applications highlight the capabilities of the model to correctly and accurately solve the equations in various cases of most interest for gravel-bed streams, e.g. in computations at low local CFL, in the solution of wet-dry fronts with fixed and movable bed and in the prediction of sediment transport in Froude trans-critical conditions.
After proving that the GIAMT2D model correctly solves the equations, we have assessed its ability in reproducing the basic processes which are responsible for the morphodynamic evolution of gravel-bed rivers, namely the development of free and forced bars and the stability of bifurcations, for which analytical solutions and laboratory data, obtained in simplified configurations, exist. The approach of systematically testing the numerical model performance against these "morphodynamic benchmarks" has seldom been employed in numerical morphodynamic modelling and is herein proposed as a preliminary step with respect to the model application to real cases. These applications show that GIAMT2D correctly reproduces most of the relevant dynamics of the unit morphodynamic processes.
With the support of the above testing, the model has been employed to study the non-linear interactions between free and forced bars in straight channels, responsible for the development of the actual bar pattern observed in rivers, for which a mature analytical theory is not available. Focusing on the simple configuration given by a straight channel having one localised constriction, the aim of this analysis was to identify the parameters controlling the relative weight of free and forced bars in their non-linear competition. This exercise has resulted in an original modelling analysis in the field of river morphodynamics, conducted by integrating remotely-sensed field

observations on alternate bars dynamics in the upper Rhine River with the outcomes of the existing linear bar theories and numerical results. Critical re-examination of the bar theories has allowed identification of two controlling parameters for the interaction, namely the channel aspect ratio β and the distance between the resonant and critical aspect ratio $\beta_{res} - \beta_{cr}$, or its relative value $\frac{\beta_{res}-\beta_{cr}}{\beta_{res}}$. These physically-based research hypotheses have been extensively validated by performing numerical runs using the GIAMT2D model and by comparison with the field data.

The present Ph.D. research has allowed to develop the GIAMT2D numerical model, whose innovation relates both to the discipline of applied mathematics and of river morphodynamics. Further work is needed to develop applications to real cases and to improve the existing algorithms.
Preliminary applications to the morphodynamics of real gravel-bed rivers are currently in progress. During the last part of this Ph.D. activity, a collaboration with Dr. M. Hicks, NIWA, Christchurch (New Zealand) has been developed in order to study the morphodynamic evolution of the Selwyn River (New Zealand), a reference river system for alluvial gravel-bed rivers [76], for which high-resolution, multi-temporal bed topography data provide ideal information of morphological changes due to bar- and channel-forming events.
The morphodynamic applications in chapters 7 and 8, together with the ongoing research activity on the Selwyn river, suggest the need for the inclusion of the following modelling ingredients:

- a two-dimensional parametrisation of secondary flows associated with streamline curvature, which is essential to achieve general applicability to configurations in which curvature plays an important role, such as bends and meanders;
- a multiple grain size formulation of the sediment transport module, in order to study selection and sorting processes, which can play a key morphodynamic role in gravel-bed rivers especially in braided systems;
- a physically-based bank erosion and accretion sub-model, which is required to study river planform change.

From the point of view of the code implementation, we have in mind to fully exploit the power of parallel computing by using GPU acceleration. This would dramatically increase the computational performance in term of time efficiency, which is likely to be required for upscaling the model in applications to real rivers over long spatial and temporal scales.

Bibliography

[1] R. Abgrall and S. Karni. A comment on the computation of non-conservative products. *Journal of Computational Physics*, 229:2759–2763, 2010.

[2] A. Armanini, L. Fraccarollo, and G. Rosatti. Two-dimensional simulation of debris flows in erodible channels. *Computers and Geosciences*, 35:993–1006, 2009.

[3] E. Audusse, F. Bouchut, M. O. Bristeau, R. Klein, and B. Perthame. A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water equations. *SIAM Journal on Scientific Computing*, 25(6):2050–2065, 2004.

[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*, 206:311–333, 2005.

[5] M. Bellal, B. Spinewine, C. Savary, and Y. Zech. Morphological evolution of steep-sloped river beds in the presence of a hydraulic jump. In *Proceedings 30th IAHR Congress*, volume C–II, pages 133–140, 2003.

[6] F. Benkhaldoun, S. Sahmim, and M. Seaid. Solution of the sediment transport equations using a finite volume method based on sign matrix. *SIAM Journal on Scientific Computing*, 31:2866–2889, 2009.

[7] F. Benkhaldoun, S. Sahmim, and M. Seaid. A two-dimensional finite volume morphodynamic model on unstructured triangular grids. *International Journal for Numerical Methods in Fluids*, 63(11):1296–1327, 2010.

[8] A. Bermúdez and M. E. Vázquez-Cendón. Upwind methods for hyperbolic conservation laws with source terms. *Computers and Fluids*, 23(8):1049–1071, 1994.

[9] A. Bernini, V. Caleffi, and A. Valiani. *Braided rivers: process, deposits, ecology and management*, chapter Numerical modelling of alternate bars, pages 153–175. Special

Publication 36 of the International Association of Sedimentologists. Blackwell Publishing, Oxford (UK), 2006.

[10] W. Bertoldi. *River Bifurcations*. PhD thesis, Università degli studi di Trento, 2004.

[11] W. Bertoldi and M. Tubino. River bifurcations: experimental observations on equilibrium configurations. *Water Resources Research*, 43(W10437), 2007.

[12] W. Bertoldi, L. Zanoni, S. Miori, R. Repetto, and M. Tubino. Interaction between migrating bars and bifurcations in gravel bed rivers. *Water Resources Research*, 45(W10437), 2009.

[13] P. Blondeaux and G. Seminara. A unified bar bend theory of river meanders. *Journal of Fluid Mechanics*, 157:449–470, 1985.

[14] M. Bolla Pittaluga, R. Repetto, and M. Tubino. Channel bifurcations in braided rivers: equilibrium configurations and stability. *Water Resources Research*, 39(3):1046, 2003.

[15] S. F. Bradford and B. F. Sanders. Finite-volume model for shallow water flooding on arbitrary topography. *Journal of Hydraulic Engineering, ASCE*, 128(3):289–298, 2002.

[16] P. Brufau, M. E. Vázquez-Cendón, and P. García Navarro. A numerical method for the flooding and drying of irregular domains. *International Journal for Numerical Methods in Fluids*, 39:247–275, 2002.

[17] T. Buffard and S. Clain. Monoslope and multislope MUSCL methods for unstructured meshes. *Journal of Computational Physics*, 229:3745–3776, 2010.

[18] J. C. Butcher. *The numerical analysis of ordinary differential equations: Runge-Kutta and general linear methods*. Wiley and Sons Ltd, 1987.

[19] V. Caleffi, A. Valiani, and B. Bernini. High-order balanced CWENO scheme for movable bed shallow water equations. *Advances in Water Resources*, 30:730–741, 2007.

[20] R. Callander. Instability and river channels. *Journal of Fluid Mechanics*, 36:465–480, 1969.

[21] A. Canestrelli, M. Dumbser, A. Siviglia, and E. F. Toro. Well-balanced high-order centred schemes on unstructured meshes for shallow water equations with fixed and mobile bed. *Advances in Water Resources*, 33:291–303, 2010.

[22] A. Canestrelli, A. Siviglia, M. Dumbser, and E. F. Toro. Well-balanced high-order centred schemes for non-conservative hyperbolic systems. Applications to shallow water equations with fixed and mobile bed. *Advances in Water Resources*, 32(6):834–844, 2009.

[23] Z. Cao and G. Pender. Numerical modelling of alluvial rivers subject to interactive sediment mining and feeding. *Advances in Water Resources*, 27:533–546, 2004.

[24] Z. Cao, G. Pender, and P. Carling. Shallow water hydrodynamic models for hyperconcentrated sediment-laden floods over erodible bed. *Advances in Water Resources*, 29:546–557, 2006.

[25] Z. Cao, R. Day, and A. Egashira. Coupled and decoupled numerical modeling of flow and morphological evolution in alluvial rivers. *Journal of Hydraulic Engineering, ASCE*, 128(3):306–321, 2002.

[26] J. Casper and H. L. Atkins. A finite-volume high-order ENO scheme for two-dimensional hyperbolic systems. *Journal of Computational Physics*, 106:62–76, 1993.

[27] M. J. Castro, E. D. Fernández-Nieto, A. M. Ferreiro Ferreiro, and C. Parés. Two-dimensional sediment transport models in shallow water equations. A second-order finite volume approach on unstructured meshes. *Computer Methods in Applied Mechanics and Engineering*, 198(33-36):2520–2538, 2009.

[28] M. J. Castro, A. M. Ferreiro Ferreiro, J. A. García-Rodríguez, J. M. González-Vida, J. Macías, C. Parés, and M. E. Vázquez-Cendón. The numerical treatment of wet/dry fronts in shallow flows: application to one-layer and two-layer systems. *Mathematical and Computer Modelling*, 42:419–439, 2005.

[29] M. J. Castro, J. M. González-Vida, and C. Parés. Numerical treatment of wet/dry fronts in shallow flows with a modified Roe scheme. *Mathematical Models and Methods in Applied Sciences*, 16(6):97–931, 2006.

[30] M. J. Castro, P. G. LeFloch, M. L. Muñoz-Ruiz, and C. Parés. Why many theories of shock waves are necessary: convergence error in formally path-consistent schemes. *Journal of Computational Physics*, 227:8107–8129, 2008.

[31] M. J. Castro, A. Pardo, and C. Parés. Well-balanced numerical schemes based on a generalized hydrostatic reconstruction technique. *Mathematical models and methods in applied sciences*, 17(12):2055–2113, 2007.

[32] B. Chapman, G. Jost, and R. van der Pas. *Using OpenMP. Portable shared memory parallel programming*. Cambridge: the MIT press, 2007.

[33] M. Colombini, G. Seminara, and M. Tubino. Finite-amplitude alternate bars. *Journal of Fluid Mechanics*, 181:213–232, 1987.

[34] M. Colombini, M. Tubino, and P. J. Whiting. *Dynamics of gravel-bed rivers*, chapter Topographic expression of bars in meandering channels, pages 457–474. Wiley and Sons Ltd, 1992.

[35] S. Cordier, M. H. Le, and T. Morales de Luna. Bedload transport in shallow water models: why splitting (may) fail, how hyperbolicity (can) help. *Advances in Water Resources*, 34:980–989, 2011.

[36] L. R. P. Correia, B. G. Krishnappan, and W. H. Graf. Fully coupled unsteady mobile boundary flow model. *Journal of Hydraulic Engineering, ASCE*, 118(3):476–494, 1992.

[37] N. Crnjaric-Zic, S. Vukovic, and L. Sopta. Extension of ENO and WENO schemes to one-dimensional bed-load sediment transport equations. *Computers and Fluids*, 33(1):31–56, 2003.

[38] A. Crosato. *Analysis and modelling of river meandering*. PhD thesis, Delft University of Technology, 2008.

[39] A. Crosato and E. Mosselman. Simple physics-based predictor for the number of river bars and the transition between meandering and braiding. *Water Resources Research*, 45(W03424), 2009.

[40] A. Crosato, E. Mosselman, F. Desta, and W. Uijttewaal. Experimental and numerical evidence for intrinsic nonmigrating bars in alluvial channels. *Water Resources Research*, 47:191–219, 2011.

[41] J. A. Cunge and N. Perdreau. Mobile bed fluvial mathematical models. *La Houille Blanche*, 7:561–580, 1973.

[42] G. Dal Maso, P. G. LeFloch, and F. Murat. Definition and weak stability of nonconservative products. *Journal de Mathematiques pures et appliquées*, 74(6):483–548, 1995.

[43] H. J. de Vriend. 2DH mathematical modelling of morphological evolutions in shallow water. *Coastal Engineering*, 11(1):1 – 27, 1987.

[44] H. J. de Vriend and N. Struiksma. Flow and bed deformation in river bends. In *Proceedings of the Conference River '83*, 1983. New Orleans, Louisiana.

[45] A. Defina. Numerical experiments on bar growth. *Water Resources Research*, 39(4), 2003.

[46] A. Delis and I. Papoglou. Relaxation approximation to bed-load sediment transport. *Journal of Computational and Applied Mathematics*, 213:521–546, 2008.

[47] M. Dumbser, C. Enaux, and E. F. Toro. Finite volume schemes of very high order of accuracy for stiff hyperbolic balance laws. *Journal of Computational Physics*, 227:3971–4001, 2008.

[48] M. Dumbser and M. Käser. Arbitrary high order non-oscillatory finite volume schemes on unstructured meshes for linear hyperbolic systems. *Journal of Computational Physics*, 221:693–723, 2007.

[49] M. Dumbser, M. Käser, V. Titarev, and E. Toro. Quadrature-free non-oscillatory finite volume schemes on unstructured meshes for nonlinear hyperbolic systems. *Journal of Computational Physics*, 226:204–243, 2007.

[50] B. Federici and C. Paola. Dynamics of channel bifurcations in noncohesive sediments. *Water Resources Research*, 39(6):1162, 2003.

[51] B. Federici and G. Seminara. On the convective nature of bar instability. *Journal of Fluid Mechanics*, 487:125–145, 2003.

[52] B. Federici and G. Seminara. Effect of suspended load on sandbar instability. *Water Resources Research*, 42(7), 2006.

[53] S. K. Godunov. Finite difference methods for the computation of discontinuous solutions of the equations of fluid dynamics. *Mat. Sb.*, 47:271–306, 1959.

[54] A. J. Grass. Sediment transport by waves and currents. Technical Report FL29, SERC London Centre for Marine Technology, 1981.

[55] G. Griffiths. Stable-channel design in gravel-bed rivers. *Journal of Hydrology*, 52(3–4):291–305, 1981.

[56] A. Harten, B. Engquist, S. Osher, and S. Chakravarthy. Uniformly high order essentially non-oscillatory schemes, III. *Journal of Computational Physics*, 71:231–303, 1987.

[57] A. Harten, P. D. Lax, and B. van Leer. On upstream differencing and Godunov-type schemes. *SIAM Review*, 25(1):35–61, 1983.

[58] F. Holly and J. Rahuel. New numerical/physical framework for mobile bed modelling, part 1: Numerical and physical principles. *Journal of Hydraulic Research*, 28(4):401–416, 1990.

[59] T. Y. Hou and P. G. LeFloch. Why non-conservative schemes converge to the wrong solutions: error analysis. *Mathematics of Computation*, 62:497–530, 1994.

[60] J. Hudson. *Numerical techniques for morphodynamic modelling*. PhD thesis, University of Reading, 2001.

[61] J. Hudson and P. Sweby. A high-resolution scheme for the equations governing 2D bed-load sediment transport. *International Journal for Numerical Methods in Fluids*, 47:1085–1091, 2005.

[62] E. J. Hundey and P. E. Ashmore. Length scale of braided river morphology. *Water Resources Research*, 45, 2009.

[63] M. Iervolino, A. Leopardi, S. Soares-Frazão, C. Swartenbroekx, and Y. Zech. 2D-H numerical simulation of dam-break flow on mobile bed with sudden enlargement. In Dittrich, Koll, Aberle, and Geisenhainer, editors, *River flow 2010*. Bundesanstalt für Wasserbau (Ferderal Waterways Engineering and Research Institute, Germany), 2010.

[64] S. Ikeda. Lateral bedload transport on side slopes. *Journal of Hydraulic Engineering*, 108:1369–1373, 1982.

[65] S. Ikeda. Prediction of alternate bar wavelength and height. *Report of Department of Foundation Engineering and Department of Construction Engineering, Saitama University, Japan*, 12:23–45, 1982.

[66] S. Ikeda, G. Parker, and K. Sawai. Bend theory of river meanders. Part I-linear development. *Journal of Fluid Mechanics*, 112:363–377, 1981.

[67] G. S. Jiang and C. W. Shu. Efficient implementation of weighted ENO schemes. *Journal of Computational Physics*, 126:202–228, 1996.

[68] R. Kinoshita and H. Miwa. River channel formation which prevents downstream translation of transverse bar. *Shinsabo*, 94:12–17, 1974. In Japanese.

[69] M. G. Kleinhans, H. R. A. Jagers, E. Mosselman, and C. J. Sloff. Bifurcation dynamics and avulsion duration in meandering rivers by one-dimensional and three-dimensional models. *Water Resources Research*, 44(W08454), 2008.

[70] A. Kurganov, S. Noelle, and G. Petrova. Semidiscrete central-upwind schemes for hyperbolic conservation laws and Hamilton-Jacobi equations. *SIAM Journal on Scientific Computing*, 23(3):707–740, 2001.

[71] A. Kurganov and G. Petrova. Central schemes and contact discontinuities. *ESAIM-Mathematical Modelling and Numerical Analysis-Modelisation Mathematique et Analyse Numerique*, 34(6):1259–1275, 2000.

[72] A. Kurganov and G. Petrova. Central-upwind schemes on triangular grids for hyperbolic systems of conservation laws. *Numerical Methods For Partial Differential Equations*, 21(3):536–552, 2005.

[73] A. Kurganov and E. Tadmor. New high-resolution central schemes for nonlinear conservation laws and convection-diffusion equations. *Journal of Computational Physics*, 160(1):241–282, 2000.

[74] C. Lai. Modeling alluvial-channel flow by multimode characteristics method. *Journal of Engineering Mechanics, ASCE*, 117(1):32–53, 1991.

[75] S. Lanzoni and M. Tubino. Grain sorting and bar instability. *Journal of Fluid Mechanics*, 393:149–174, 1999.

[76] S. T. Larned, D. M. Hicks, J. Schmidt, A. J. H. Davey, K. D. amd M. Scarsbrook, D. B. Arscott, and R. A. Woods. The Selwyn River of New Zealand: a benchmark system for alluvial plain rivers. *River research and applications*, 24:1–21, 2008.

[77] P. D. Lax. Weak solutions of nonlinear hyperbolic equations and their numerical computation. *Communications on Pure and Applied Mathematics*, 7(1):159–193, 1954.

[78] P. D. Lax and B. Wendroff. Systems of conservation laws. *Communications on Pure and Applied Mathematics*, 13:217–237, 1960.

[79] G. R. Lesser, J. A. Roelvink, J. A. T. M. Van Kester, and G. S. Stelling. Development and validation of a three-dimensional morphological model. *Coastal Engineering*, 51(8–9):883–915, 2004.

[80] C. B. Liao, M. S. Wu, and S. J. Liang. Numerical simulation of dam break for an actual river environment. *Hydrological Processes*, 21(4):447–460, 2007.

[81] X. Liu, B. J. Landry, and M. H. García. Two-dimensional scour simulations based on coupled model of shallow water equations and sediment transport on unstructured meshes. *Coastal Engineering*, 55(10):800–810, 2008.

[82] X. Liu, S. Osher, and T. Chan. Weighted essentially non-oscillatory schemes. *Journal of Computational Physics*, 115:200–212, 1994.

[83] D. A. Lyn. Unsteady sediment transport modeling. *Journal of Hydraulic Engineering, ASCE*, 113(1):1–15, 1987.

[84] D. A. Lyn and M. Altinakar. St. Venant-Exner equations for near-critical and transcritical flows. *Journal of Hydraulic Engineering, ASCE*, 128(6):579–587, 2002.

[85] E. Meyer-Peter and R. Müller. Formulas for bed-load transport. In *Proceedings of the Second IAHR meeting*, pages 39–64, 1948.

[86] S. Miori, R. Repetto, and M. Tubino. A one-dimensional model of bifurcations in gravel bed channels with erodible banks. *Water Resources Research*, 42(W11413), 2006.

[87] E. Mosselman. *Braided rivers: process, deposits, ecology and management*, chapter Bank protection and river training along the braided Brahmaputra-Jamuna River, Bangladesh, pages 277–287. Special Publication 36 of the International Association of Sedimentologists. Blackwell Publishing, Oxford (UK), 2006.

[88] E. Mosselman. Modelling sediment transport and morphodynamics of gravel-bed rivers. In *Gravel bed rivers: processes, tools, environments*, pages 101–115, 2012.

[89] C. D. Munz. On the numerical dissipation of high resolution schemes for hyperbolic conservation laws. *Journal of Computational Physics*, 77:18–39, 1998.

[90] J. Murillo and P. García-Navarro. An Exner-based coupled model for two-dimensional transient flow over erodible bed. *Journal of Computational Physics*, 229:8704–8732, 2010.

[91] I. K. Nikolos and A. I. Delis. An unstructured node-centred 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.

[92] K. W. Olesen. Alternate bars and meandering of alluvial rivers. Communications on Hydraulics 83–1, Department of Civil Engineering, Delft University of Technology, Delft, the Netherlands, 1983.

[93] K. W. Olesen. Steady flow in alluvial channels - discussion. *Journal of Waterway Port Coastal and Ocean Engineering, ASCE*, 110(1):99–100, 1984.

[94] A. Palumbo, S. Soares-Frazão, L. Goutiere, D. Pianese, and Y. Zech. Dam-break flow on mobile bed in a channel with a sudden enlargement. In M. S. Altinakar, M. A. Kokpinar, I. Aydin, S. Cokgor, and S. Kirkgoz, editors, *River flow 2008*, pages 645–654. Kubaba Congress Department and Travel Services, Ankara.

[95] C. Parés. Numerical methods for non-conservative hyperbolic systems: a theoretical framework. *SIAM Journal of Numerical Analysis*, 44:300–321, 2006.

[96] G. Parker. Surface-based bedload transport relation for gravel rivers. *Journal of Hydraulic Research*, 28(4):417–436, 1990.

[97] R. Repetto and M. Tubino. Transition from migrating alternate bars to steady central bars in channels with variable width. In *River, Coastal and Estuarine Morphodynamics: RCEM 1999*, 1999.

[98] R. Repetto, M. Tubino, and C. Paola. Planimetric instability of channels with variable width. *Journal of Fluid Mechanics*, 457:79–109, 2002.

[99] M. Ricchiuto and A. Bollermann. Stabilized residual distribution for shallow water simulations. *Journal of Computational Physics*, 228(4):1071–1115, 2009.

[100] A. Ritter. Die fortpflanzung der wasserwellen. *Zeitschrift Verein Deutscher Ingenieure*, 36:947–954, 1892. In German.

[101] P. L. Roe. Some contributions to the modelling of discontinuous flows. In *Proceedings of the SIAM/AMS Seminar*, 1983.

[102] G. Rosatti and L. Fraccarollo. A well-balanced approach for flows over mobile-bed with high sediment-transport. *Journal of Computational Physics*, 220(1):312–338, 2006.

[103] S. R. Sabbagh-Yazdi and M. Zounemat-Kermani. Numerical solution of tidal currents at marine waterways using wet and dry technique on Galerkin finite volume algorithm. *Computers and Fluids*, 38(10):1876–1886, 2009.

[104] S. Saiedi. Coupled modeling of alluvial flows. *Journal of Hydraulic Engineering*, 123(5):440–446, 1997.

[105] G. Seminara. Stability and morphodynamics. *Meccanica*, 33(1):59–99, 1998.

[106] G. Seminara and M. Tubino. *River Meandering*, chapter Alternate bars and meandering: Free, forced and mixed interactions, pages 267–320. AGU Water Resources Monographs, 1989.

[107] G. Seminara and M. Tubino. Weakly nonlinear theory of regular meanders. *Journal of Fluid Mechanics*, 244:257–288, 1992.

[108] C. W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77(2):439–471, 1988.

[109] J. Sieben. A theoretical analysis of discontinuous flow with mobile bed. *Journal of Hydraulic Research*, 37(2):199–212, 1999.

[110] A. Siviglia, G. Nobile, and M. Colombini. Quasi-conservative formulation of the one dimensional Saint Venant-Exner model. *Journal of Hydraulic Engineering*, 134(10):1521–1526, 2008.

[111] A. Siviglia, G. Stecca, D. Vanzo, G. Zolezzi, E. F. Toro, and M. Tubino. Numerical modelling of two-dimensional morphodynamcs in gravel-bed rivers. In preparation.

[112] G. Stecca, A. Siviglia, and E. F. Toro. Upwind-biased FORCE schemes with applications to free-surface shallow flows. *Journal of Computational Physics*, 229(18):6362–6380, 2010.

[113] G. Stecca, A. Siviglia, and E. F. Toro. A finite volume upwind-biased centred scheme for hyperbolic systems of conservation laws. Application to shallow water equations. *Communications in Computational Physics*, 2012. To appear.

[114] N. Struiksma, K. Olesen, C. Flokstra, and H. J. de Vriend. Bed deformation in curved alluvial channels. *Journal of Hydraulic Research*, 23(1):57–79, 1985.

[115] J. P. M. Syvitski, R. L. Slingerland, P. Burgess, E. Meiburg, A. B. Murray, P. Wiberg, G. Tucker, and A. A. Voinov. Morphodynamic models: an overview. In *River, Coastal and Estuarine Morphodynamics: RCEM 2009*. Taylor and Francis Group, London, 2010.

[116] A. Talmon, M. C. L. M. van Mierlo, and N. Struiskma. Laboratory measurements of direction of sediment transport on transverse alluvial-bed slopes. *Journal of Hydraulic Research*, 33(4):519–543, 1995.

[117] E. F. Toro. *Shock-Capturing Methods for Free-Surface Shallow Flows*. Wiley and Sons Ltd, 2001.

[118] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics*. Springer–Verlag, 2009.

[119] E. F. Toro and S. Billett. Centred TVD Schemes for Hyperbolic Conservation Laws. *IMA Journal of Numerical Analysis*, 20:47–79, 2000.

[120] E. F. Toro, A. Hidalgo, and M. Dumbser. FORCE schemes on unstructured meshes I: Conservative hyperbolic systems. *Journal of Computational Physics*, 228(9):3368–3389, 2009.

[121] E. F. Toro, R. C. Millington, and L. A. M. Nejad. *Godunov Methods: Theory and Applications*, pages 905–938. Kluwer Academic / Plenum Publishers, 2001.

[122] E. F. Toro and A. Siviglia. PRICE: primitive centred schemes for hyperbolic systems. *International Journal for Numerical Methods in Fluids*, 42:1263–1291, 2003.

[123] E. F. Toro and V. A. Titarev. Solution of the generalized Riemann problem for advection-reaction equations. *Proceedings of The Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 458(2018):271–281, 2002.

[124] M. Tubino. Growth of alternate bars in unsteady flow. *Water Resources Research*, 27(1):37–52, 1991.

[125] M. Tubino and W. Bertoldi. *Gravel Bed River 6- From process understanding to river restoration*, chapter Bifurcations in gravel-bed streams. Amsterdam, Elsevier Science, 2008.

[126] M. Tubino, R. Repetto, and G. Zolezzi. Free bars in rivers. *Journal of Hydraulic Research*, 37(6):759–775, 1999.

[127] M. Tubino and G. Seminara. Free forced interactions in developing meanders and suppression of free bars. *Journal of Fluid Mechanics*, 214:131–159, 1990.

[128] M. Tubino, A. Siviglia, G. Stecca, and M. Ciolli. Modelli matematici per l'evoluzione morfologica dell'assetto del fondo della biforcazione sul Torrente Armea in corrispondenza delle proprietà della Betonfin s.n.c. Technical report, Department of Civil and Environmental Engineering, University of Trento, Trento, Italy, 2010. In Italian.

[129] C. van der Meer, E. Mosselman, K. C. J. Sloff, H. R. A. Jagers, G. Zolezzi, and M. Tubino. Numerical simulations of upstream and downstream overdeepening. In *River, Coastal and Estuarine Morphodynamics: RCEM 2011*. Tsinghua University Press, Beijing, 2011.

[130] B. van Leer. Towards the ultimate conservative difference scheme II. Monotonicity and conservation combined in a second order scheme. *Journal of Computational Physics*, 14:361–370, 1974.

[131] B. van Leer. Towards the ultimate conservative difference scheme V. A second order sequel to Godunov's method. *Journal of Computational Physics*, 32(2):101–136, 1979.

[132] D. Vanzo, A. Siviglia, G. Zolezzi, G. Stecca, and M. Tubino. Interaction between steady and migrating bars in straight channels. In *River, Coastal and Estuarine Morphodynamics: RCEM 2011*. Tsinghua University Press, Beijing, 2011.

[133] Z. B. Wang, R. J. Fokkink, M. de Vries, and A. Langerak. Stability of river bifurcations in 1D morphodynamic models. *Journal of Hydraulic Research*, 33(6):739–750, 1995.

[134] P. J. Whiting and W. E. Dietrich. Experimental studies of bed topography and flow patterns in large-amplitude meanders. 1. Observations. *Water Resources Research*, 29(11):3605–3614, 1993.

[135] M. Wong and G. Parker. Reanalysis and correction of bed-load relation of Meyer-Peter and Müller using their own database. *Journal of Hydraulic Engineering*, 132(11):1159–1168, 2006.

[136] W. Wu, D. Vieira, and S. Wang. Onedimensional numerical model for nonuniform sediment transport under unsteady flows in channel networks. *Journal of Hydraulic Engineering, ASCE*, 130(9):914–923, 2004.

[137] J. Xia, B. Lin, R. A. Falconer, and G. Wang. Modelling dam-break flows over mobile beds using a 2D coupled approach. *Advances in Water Resources*, 33(2):171–183, 2010.

[138] Y. Zech, S. Soares-Frazão, S. Spinewine, and N. Le Grelle. Dam-break induced sediment movement: Experimental approaches and numerical modelling. *Journal of Hydraulic Research*, 46(2):176–190, 2008.

[139] G. Zolezzi, W. Bertoldi, and M. Tubino. *Braided rivers: process, deposits, ecology and Management*, chapter Morphological analysis and prediction of channel bifurcations, pages 233–256. Special Publication 36 of the International Association of Sedimentologists. Blackwell Publishing, Oxford (UK), 2006.

[140] G. Zolezzi, M. Guala, D. Termini, and G. Seminara. Experimental observations of upstream overdeepening. *Journal of Fluid Mechanics*, 531:191–219, 2005.

[141] G. Zolezzi and G. Seminara. Downstream and upstream influence in river meandering. Part 1. General theory and application to overdeepening. *Journal of Fluid Mechanics*, 438:183–211, 2001.

[142] G. Zolezzi, D. Vanzo, M. Tubino, A. Siviglia, and G. Stecca. Nonlinear interaction between steady and migrating bars in straight gravel-bed river reaches. In preparation.

Acknowledgements

This section is in Italian, my mother tongue. I apologise with those who will not be able to understand it, but I believe this is the only way I had to express my deepest feelings.

Scrivere una pagina di ringraziamenti, che nel mio caso viene buona ultima dopo tante pagine dedicate a quella che prova ad essere scienza, è un esercizio pericoloso: l'abuso di retorica, la ricostruzione edulcorata, il "volemose bene", il "raggiungimento di un grande traguardo umano scientifico e professionale" sono rischi concreti. Provo allora a fare sì che questa sezione ne sia priva.

Il primo pensiero è per la mia famiglia: per mia madre e mio padre e per mio fratello Giovanni. A Giovanni, la cui saggezza io non riuscirò mai lontanamente ad avvicinare, va una menzione particolare, per avermi aiutato nell'ultima rilettura delle bozze di questa tesi.

Il mio dottorato avrebbe preso una direzione radicalmente diversa e questa tesi sarebbe decisamente meno corposa senza l'impegno profuso da Davide nello svolgimento del suo lavoro di tesi magistrale e nei primi mesi del suo dottorato. Davide ha ridisegnato i grafici delle soluzioni numeriche fino alla notte prima della consegna. Grazie infinite, collega, e *ad maiora*.

La versione finale di questa tesi beneficia della precisa rilettura e dei puntuali suggerimenti del Prof. Mosselman (TU Delft, Olanda): per l'attenzione dedicata al mio lavoro gli sono personalmente grato.

Una menzione speciale va ai miei colleghi di ventura Sebastiano e Matilde: a Sebastiano per la serenità che gli è propria e regala a chi lo circonda, a Matilde per la grande, profonda amicizia che ci lega.

Un pensiero è per Murray e per la squisita gentilezza che mi ha dimostrato durante la mia permanenza in Nuova Zelanda. Peccato non si sia riusciti, nonostante gli sforzi, a concludere il lavoro in tempo per includerlo nella tesi. Non disperiamo, comunque, la questione è soltanto rimandata.

Un sentito ringraziamento va ai miei numerosi supervisori (formalmente tre, ma all'atto pratico quattro, con l'inclusione della figura del Supervisore Occulto) per avermi permesso di sviluppare le mie conoscenze in due campi, l'analisi numerica e la morfodinamica fluviale, di particolare interesse per me. La loro azione si è esplicata in interazioni di carattere marcatamente non-lineare, caratterizzate da elevata variabilità spazio-temporale: un raro caso di identità fra le caratteristiche della supervisione e quelle dell'oggetto dello studio. Allo stato, non è disponibile alcun modello o teoria analitica capace di spiegare queste interazioni nel caso generale. Eppure, il problema ha risvolti pratici di grande rilevanza: dalla possibilità di prevedere questi fenomeni, la qualità della vita dottorale trarrebbe significativo beneficio.

Un abbraccio va a Luca perché c'è sempre, da venticinque anni a questa parte; a Clarissa per i bei viaggi tra Spagna e Portogallo; ad Ilaria per le sciate di fondo; a Mike, perché un tuffo da un aereo in volo tra Fox Glacier e il Mare di Tasmania è cosa che lega in maniera profonda; a Laura, perché quando canta "Oh my sweet Carolina" la Carolina diventa un luogo dell'anima; a Maralva perché ha una parola sola e ne è orgogliosa; a Max e i Tennessee River per tanta birra, risate, country e rock-'n'-roll; a Tullia per il caffè condito di scurrilità e facezie irriferibili; a Giulio, perché dimostra in pratica che la vita riserva una possibilità umana anche a chi ha studiato ingegneria; a Simone per la buona cucina; a Damiana per le discussioni mai scontate; ad Errico per la ventata di pensiero libero che la sua frequentazione mi porta e a Paolo perché è scemo oltre misura.

Un ricordo affettuoso va a Gianni, maestro di tutto, indimenticato amico.

Desidero infine menzionare le numerose entità, fisiche e metafisiche, che hanno accompagnato il mio dottorato costituendone parte integrante, trasformandolo a tratti in un teatro dei sogni. Ricordo, a questo proposito, in ordine rigorosamente alfabetico, A^2, gli Abominidi, l'Acqua del vicino, Alfred Pass, l'Amico di Copertura, l'Amico ritrovato, l'Amplissimo preside Toobusy, Annibale, sua Assenza, l'Assenzio ovvero Eau de Absence (il vino regalato da sua Assenza), l'Atletico Grifone, Axel Wayn, la Badante, Bad Man, Big Al, il Beppe, il Bivacco Bailoni, il Brahmaputra-Gemona, il Braiding-storming, i Brunos, il Calpo, il Cammello, il Caffè Corretto ed il Caffè Shakerato (*sensu Tulliae*), the Chincut River, Claude LeMans e la Terrazza con Piscina, il Cloaker, la Colata Levica, Cristochiesa Categorica, il vasto e ben assortito Collegio Supervisionale, le Cooperative di Sceriffi, le Distribuzioni Bimodali, la Dona Bionda, il Dottor Cappelletti, il Duo dell'Ascensore, l'Elefante, l'Espansione verso il Basso, i Fonomontaggi, la Formula di Braulins, il Fuso orario GIAMT+1, il Galletto al Teroldego, il Gatto e la Volpe, Giacomo Daniele, il Grifone Geostazionario, HITTIGER!, the Hole Driller, Holy Manoly, l'Ingegnere di Guardia, l'Ingegnere Zen, Iperroaz (altrimenti detto Piccolo Grande Roaz, SeBaikal, SeBike, SeBalkan), l'Ipervisore Guado (aspettando Guidot), Lalla

minore, Lani, il Lupo, Melone, Mesiano Beach, il Metodo FORST (predictor-corrector), i Metodi P.A.T.-conservative, i Metodi TVB, lo yoghurt Meyer-Peter Müller (con croccanti sedimenti log-normalmente assortiti), il Mezzo Pomeriggio, il noto modello morfodinamico "Mezzo Pomeriggio 2D", Mick Jagger, la Modella 2D, Natalie 'N', Nostra Signora della Marca da Bollo (patrona degli Amministrativi), Obi Wan Kenobi e il metodo FORCE, il protocollo OpenCL, il Paiolo, la Panda Millenove, le Pappardelle Selvagge, Partenope (i, j), Pecos Bill, Ponte Avances, il Pozzo di Scienza, Pretty (Scotland), i Progetti Europei, il ProgrammA^2 (DirettivA^2), R^2, la Retata di Passeggiatrici, Short John, gli Smarties, Snunzio (ν), Spiridione l'Africano, la Squadra di Reperibilità Morfodinamica, la Studentella, Tilde (Hermione, Feldmaresciallo), il Topolino, la celebre Tradizione Orale dei Modelli del Dipartimento, la Vegetazione Cariparia e tanti, tanti altri enti che il pudore e la decenza ci impongono non menzionare esplicitamente.