Commun. Comput. Phys.
doi: 10.4208/cicp.180511.071211a

Vol. **12**, No. 4, pp. 1183-1214
October 2012

A Finite Volume Upwind-Biased Centred Scheme for Hyperbolic Systems of Conservation Laws: Application to Shallow Water Equations

Guglielmo Stecca[1,*], Annunziato Siviglia[1] and Eleuterio F. Toro[2]

[1] *Department of Civil and Environmental Engineering, University of Trento, Via Mesiano 77, I-38100 Trento, Italy.*
[2] *Laboratory of Applied Mathematics, University of Trento, Via Mesiano 77, I-38100 Trento, Italy.*

Received 18 May 2011; Accepted (in revised version) 7 December 2011

Communicated by Song Jiang

Available online 17 April 2012

Abstract. We construct a new first-order central-upwind numerical method for solving systems of hyperbolic equations in conservative form. It applies in multidimensional structured and unstructured meshes. The proposed method is an extension of the UFORCE method developed by Stecca, Siviglia and Toro [25], in which the upwind bias for the modification of the staggered mesh is evaluated taking into account the smallest and largest wave of the entire Riemann fan. The proposed first-order method is shown to be identical to the Godunov upwind method in applications to a 2×2 linear hyperbolic system. The method is then extended to non-linear systems and its performance is assessed by solving the two-dimensional inviscid shallow water equations. Extension to second-order accuracy is carried out using an ADER-WENO approach in the finite volume framework on unstructured meshes. Finally, numerical comparison with current competing numerical methods enables us to identify the salient features of the proposed method.

AMS subject classifications: 65M08, 76M12

Key words: Conservative hyperbolic systems, centred schemes, unstructured meshes, numerical fluxes, shallow water equations, FORCE, upwind-biased.

1 Introduction

1.1 Preliminaries

We consider a general system of non-linear conservation laws in α space dimensions:

*Corresponding author. *Email addresses:* guglielmo.stecca@ing.unitn.it (G. Stecca), nunzio.siviglia@ing.unitn.it (A. Siviglia), toroe@ing.unitn.it (E. F. Toro)

$$\partial_t \mathbf{Q} + \mathrm{div}\left(\underline{\underline{\mathbf{F}}}(\mathbf{Q})\right) = \mathbf{0}, \tag{1.1}$$

where $\underline{\underline{\mathbf{F}}}(\mathbf{Q})$ is the flux tensor.

We assume a conforming tessellation $\mathcal{T}_\Omega$ of the computational domain $\Omega \subset \mathbb{R}^\alpha$ by n_e elements T_i such that:

$$\mathcal{T}_\Omega = \bigcup_{i=1}^{n_e} T_i. \tag{1.2}$$

Each element T_i has n_f plane interfaces S_j of size $|S_j|$, with associated outward pointing face normal vectors $\vec{n}_j$. Element T_i, having size $|T_i|$, 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 as V_j^+. Fig. 1 illustrates the above definitions and notation for the two-dimensional case. Note that the intersection of V_j^- and V_j^+ gives the interface S_j of the element T_i. With reference to Fig. 1 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 $j=1,2,3$.

Finite volume schemes are obtained by integration of the conservation law (1.1) over a space-time control volume $T_i \times \left[t^n, t^{n+1}\right]$, yielding:

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{\Delta t}{|T_i|} \sum_{j=1}^{n_f} \int_{S_j} \underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}} \left(\mathbf{Q}_i^n, \mathbf{Q}_j^n\right) \cdot \vec{n}_j \mathrm{d}\vec{x}, \tag{1.3}$$

where $\mathbf{Q}_i^n$ is the cell average at time level n and $\Delta t = t^{n+1} - t^n$ is the time step. Two different approaches are available for determining $\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}$. The first approach is the *upwind approach*, represented by Godunov's method [9] and the second is the *centred approach*, typically represented by the Lax-Friedrichs flux and variations of it [18]. For a comprehensive presentation of upwind, and also some centred methods, see for example [27] and references therein.

In this paper we derive a central-upwind method which partially uses 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 [17], using an adaptive staggered mesh. Their scheme is based on a modification of the centred scheme of Nessyahu and Tadmor [21], where the staggered mesh is fixed. Extensions to multidimensions of the scheme of Nessyahu and Tadmor [21] has been obtained by Jiang and Tadmor [13] and by Arminjon and collaborators [1]. Multi-dimensional extensions of the scheme of Kurganov and Tadmor have been presented in [14] (Cartesian version) and [16] (unstructured version), while a modified version of the scheme optimised for treating contact discontinuities, which makes use of partial characteristic decomposition, has been presented in [15].

Our scheme is strictly related to the UFORCE central-upwind method developed by Stecca *et al.* [25], which is an upwind-biased version of the FORCE scheme for Cartesian meshes. The present scheme fully exploits the idea of varying adaptively the secondary mesh size, which was first introduced in the FORCE framework in [25].

For a 2×2 linear system in two and three space dimensions the proposed scheme reproduces identically the Godunov scheme constructed by solving exactly the Riemann problems normal to each interface. For non-linear systems the extension is empirical and makes explicit use of estimates for the largest and smallest wave speeds of the entire Riemann fan.

1.2 The FORCE scheme on general meshes

We review the construction of the multidimensional FORCE scheme originally proposed by Toro et al. [29]. The multi-dimensional FORCE flux on unstructured meshes is constructed as follows:

1. 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 (1.1) over the *secondary elements*:

$$\mathbf{Q}_{j+\frac{1}{2}}^{n+\frac{1}{2}}=\frac{\mathbf{Q}_i^n\left|V_j^-\right|+\mathbf{Q}_j^n\left|V_j^+\right|}{\left|V_j^-\right|+\left|V_j^+\right|}-\frac{1}{2}\frac{\Delta t\left|S_j\right|}{\left|V_j^-\right|+\left|V_j^+\right|}\left(\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_j^n\right)-\underline{\underline{\mathbf{F}}}(\mathbf{Q}_i^n)\right)\cdot\vec{n}_j, \tag{1.4}$$

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

2. Then, with initial condition at time $t^{n+\frac{1}{2}}$ given by (1.4), integration of the conservation law (1.1) over the *primary elements* $T_i\times\left[t^{n+\frac{1}{2}},t^{n+1}\right]$ yields averages at time $t^{n+1}=t^n+\Delta t$, namely:

$$\mathbf{Q}_i^{n+1}=\frac{1}{|T_i|}\sum_{j=1}^{n_f}\left(\mathbf{Q}_{j+\frac{1}{2}}^{n+\frac{1}{2}}\left|V_j^-\right|-\frac{1}{2}\Delta t\left|S_j\right|\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_{j+\frac{1}{2}}^{n+\frac{1}{2}}\right)\cdot\vec{n}_j\right). \tag{1.5}$$

Eqs. (1.4) and (1.5) constitute a first-order accurate, explicit two-step method for solving (1.1) on a staggered mesh. Finally, following the FORCE approach [28] the scheme can now 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 (1.4), (1.5) is recast into the sought one-step form:

$$\mathbf{Q}_i^{n+1}=\mathbf{Q}_i^n-\frac{\Delta t}{|T_i|}\sum_{j=1}^{n_f}\left|S_j\right|\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{FORCE}\cdot\vec{n}_j, \tag{1.6}$$

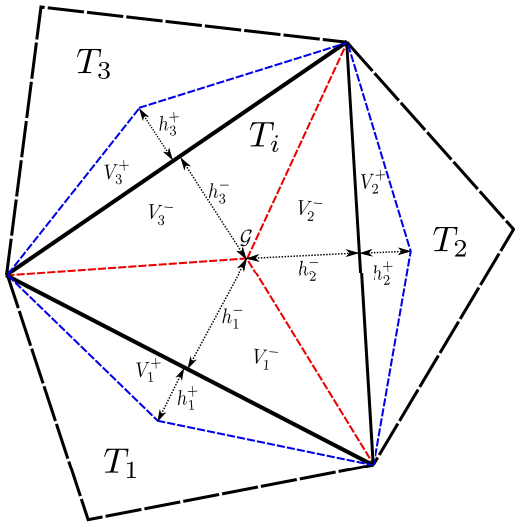

Figure 1: Sketch of the primary and secondary mesh for the FORCE method: 2D triangular case.

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

$$\underline{\underline{\mathbf{F}}}^{FORCE}_{j+\frac{1}{2}} = \frac{1}{2}\left(\underline{\underline{\mathbf{F}}}^{LW}_{j+\frac{1}{2}}\left(\mathbf{Q}^n_i,\mathbf{Q}^n_j\right)+\underline{\underline{\mathbf{F}}}^{LF}_{j+\frac{1}{2}}\left(\mathbf{Q}^n_i,\mathbf{Q}^n_j\right)\right). \tag{1.7}$$

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 (1.4):

$$\underline{\underline{\mathbf{F}}}^{LW}_{j+\frac{1}{2}} = \underline{\underline{\mathbf{F}}}\left(\mathbf{Q}^{LW}_{j+\frac{1}{2}}\right), \tag{1.8}$$

$$\mathbf{Q}^{LW}_{j+\frac{1}{2}} = \frac{\mathbf{Q}^n_i\left|V^-_j\right|+\mathbf{Q}^n_j\left|V^+_j\right|}{\left|V^-_j\right|+\left|V^+_j\right|} - \frac{1}{2}\frac{\Delta t\left|S_j\right|}{\left|V^-_j\right|+\left|V^+_j\right|}\left(\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}^n_j\right)-\underline{\underline{\mathbf{F}}}(\mathbf{Q}^n_i)\right)\cdot\vec{n}_j, \tag{1.9}$$

while the Lax-Friedrichs-type flux for general meshes in multiple space dimensions is given by

$$\underline{\underline{\mathbf{F}}}^{LF}_{j+\frac{1}{2}} = \frac{\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}^n_j\right)\left|V^-_j\right|+\underline{\underline{\mathbf{F}}}(\mathbf{Q}^n_i)\left|V^+_j\right|}{\left|V^-_j\right|+\left|V^+_j\right|} - \frac{\left|V^-_j\right|\left|V^+_j\right|}{\left|V^-_j\right|+\left|V^+_j\right|}\frac{2}{\Delta t\left|S_j\right|}\left(\mathbf{Q}^n_j-\mathbf{Q}^n_i\right)\vec{n}^T_j, \tag{1.10}$$

where $\vec{n}^T_j$ denotes the transpose of n_j.

It is worth mentioning that different FORCE-type methods suitable to be applied to non-conservative systems are available [2,3,6].

1.3 The FORCE-α method on general meshes

In this section we generalize the FORCE-α formulation developed by Toro *et al.* [29] for Cartesian meshes to general unstructured meshes. This is helpful for deriving the proposed method in the next section. 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 3D ($\alpha=3$). In this paper we indicate with $\left|V_j^-\right|$ and $\left|V_j^+\right|$ the size of these subvolumes, which dimensionally corresponds to their surface area in 2D and their volume in 3D, given by

$$\left|V_j^+\right|=\frac{h_j^+\left|S_j\right|}{\alpha},\qquad \left|V_j^-\right|=\frac{h_j^-\left|S_j\right|}{\alpha}, \tag{1.11}$$

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 area of the triangle base or pyramid base surface (see Fig. 1). Substitution of (1.11) into the FORCE flux formulae (1.7)-(1.10) gives

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{FORCE-\alpha}=\frac{1}{2}\left(\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LW-\alpha}\left(\mathbf{Q}_i^n,\mathbf{Q}_j^n\right)+\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LF-\alpha}\left(\mathbf{Q}_i^n,\mathbf{Q}_j^n\right)\right), \tag{1.12}$$

with

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LW-\alpha}=\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_{j+\frac{1}{2}}^{LW-\alpha}\right), \tag{1.13}$$

$$\mathbf{Q}_{j+\frac{1}{2}}^{LW-\alpha}=\frac{\mathbf{Q}_i^n h_j^-+\mathbf{Q}_j^n h_j^+}{h_j^-+h_j^+}-\frac{1}{2}\frac{\alpha\Delta t}{h_j^-+h_j^+}\left(\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_j^n\right)-\underline{\underline{\mathbf{F}}}(\mathbf{Q}_i^n)\right)\cdot\vec{n}_j, \tag{1.14}$$

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{LF-\alpha}=\frac{\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_j^n\right)h_j^-+\underline{\underline{\mathbf{F}}}(\mathbf{Q}_i^n)h_j^+}{h_j^-+h_j^+}-\frac{h_j^-h_j^+}{h_j^-+h_j^+}\frac{2}{\alpha\Delta t}\left(\mathbf{Q}_j^n-\mathbf{Q}_i^n\right)\vec{n}_j^T. \tag{1.15}$$

It is worth mentioning that the above formulation applies to any kind of mesh. The method requires the knowledge of altitudes $h_j^\pm$, which in the case of triangular and tetrahedral meshes are given by

$$h_j^-=\frac{\alpha\left|T_i\right|}{n_f\left|S_j\right|},\qquad h_j^+=\frac{\alpha\left|T_j\right|}{n_f\left|S_j\right|}, \tag{1.16}$$

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

Finally, the FORCE-α flux represents the starting point for the development of a multidimensional upwind biased FORCE flux to be applied to general meshes. This is the main purpose of the paper and the object of the next sections.

2 The UFORCE-δ method

The purpose of this section is to design a monotone FORCE-type method characterised by reduced numerical dissipation. The sought flux shall be of the centred upwind-biased (or central-upwind) type and applicable to general meshes in multidimensions. We propose a modification of the FORCE-α flux (1.10)-(1.12) which consists in varying adaptively the two secondary subvolumes adjacent to each interface in the primary mesh. The resulting flux will contain two upwind bias parameters, i.e. δ_j^+ and δ_j^-, which control the shape of the secondary mesh at the same time level.

2.1 Derivation of the UFORCE-δ flux

In analogy to the multi-dimensional FORCE scheme, the derivation of the UFORCE-δ method requires the adoption of a primary mesh for computing cell averages and a staggered secondary mesh used to define numerical fluxes for the conservative form of the scheme. The primary mesh corresponds to a conforming tessellation (see Fig. 2), while the secondary mesh is staggered with respect to the primary mesh. Each cell of the secondary mesh is composed by two subvolumes V_j^- and V_j^+, the former laying within cell T_i, the latter laying outside. The intersection of V_j^- and V_j^+ gives the interface S_j (see Fig. 2, where the two-dimensional triangular case is shown).

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. Subvolumes V_j^- are generated independently from each other by connecting the vertices of interface S_j with one point P_j associated with S_j laying within T_i.

For the proposed scheme we impose that each subvolume V_j^- cannot be greater than the corresponding V_j^- in the FORCE method. This condition also ensures that the primary mesh subvolumes V_j^+ have smaller size than their counterparts in the FORCE method. Since the amount of numerical dissipation associated to the averaging procedure described in (1.4) and (1.5) 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 (1.11) we obtain

$$\left|V_j^-\right| = \frac{\delta_j^- h_j^- \left|S_j\right|}{\alpha}, \qquad \left|V_j^+\right| = \frac{\delta_j^+ h_j^+ \left|S_j\right|}{\alpha}, \tag{2.1}$$

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

$$0 \le \delta_j^- \le 1, \qquad 0 \le \delta_j^+ \le 1. \tag{2.2}$$

Once the secondary mesh (2.1) is defined, the derivation of the UFORCE-δ numerical flux proceeds from equations (1.4) and (1.5), giving the following flux

$$\underline{\mathbf{F}}_{j+\frac{1}{2}}^{UFORCE-\delta} = \frac{1}{2}\left\{\underline{\mathbf{F}}_{j+\frac{1}{2}}^{ULW-\delta}\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n\right) + \underline{\mathbf{F}}_{j+\frac{1}{2}}^{ULF-\delta}\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n\right)\right\}, \tag{2.3}$$

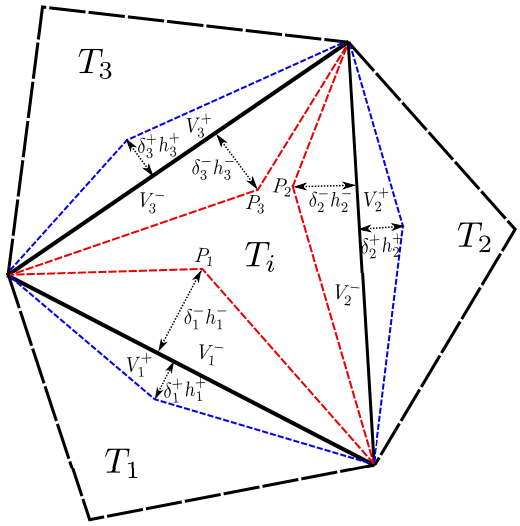

Figure 2: Sketch of the primary and secondary mesh for the UFORCE-δ method: 2D triangular case.

with

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{ULW-\delta} = \underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_{j+\frac{1}{2}}^{ULW-\delta}\right), \tag{2.4}$$

$$\mathbf{Q}_{j+\frac{1}{2}}^{ULW-\delta} = \frac{1}{\left(\delta_j^- h_j^- + \delta_j^+ h_j^+ + \epsilon\right)} \left\{ \left(\mathbf{Q}_i^n \delta_j^- h_j^- + \mathbf{Q}_j^n \delta_j^+ h_j^+\right) - \frac{1}{2}\alpha\Delta t \left(\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_j^n\right) - \underline{\underline{\mathbf{F}}}(\mathbf{Q}_i^n)\right)\cdot\vec{n}_j \right\}, \tag{2.5}$$

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{ULF-\delta} = \frac{1}{\left(\delta_j^- h_j^- + \delta_j^+ h_j^+ + \epsilon\right)} \left\{ \left(\underline{\underline{\mathbf{F}}}\left(\mathbf{Q}_j^n\right)\delta_j^- h_j^- + \underline{\underline{\mathbf{F}}}(\mathbf{Q}_i^n)\delta_j^+ h_j^+\right) - 2\frac{\delta_j^- h_j^- \delta_j^+ h_j^+}{\alpha\Delta t}\left(\mathbf{Q}_j^n - \mathbf{Q}_i^n\right)\vec{n}_j^T \right\}, \tag{2.6}$$

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 is a small positive real number, e.g. $\epsilon = 10^{-10}$.

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.

2.2 The optimal upwind bias

Different choices for $\delta_j^\pm$ in (2.3)-(2.6) give different numerical methods. Here we concentrate on an adaptive choice of the upwind bias parameters, i.e. a relationship 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 adopted by Stecca *et al.* [25]. The steps we follow 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.

For sake of generality, we perform our analysis on a three-dimensional test problem on an unstructured mesh. In order to evaluate the two bias δ_j^+ and δ_j^- we make use of the following linear system:

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_y \mathbf{G}(\mathbf{Q}) + \partial_z \mathbf{H}(\mathbf{Q}) = \mathbf{0}, \tag{2.7}$$

where $\mathbf{Q} = [q_1, q_2]^T$ 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

$$\mathbf{F}(\mathbf{Q}) = \mathbf{A}_x \mathbf{Q}, \qquad \mathbf{G}(\mathbf{Q}) = \mathbf{A}_y \mathbf{Q}, \qquad \mathbf{H}(\mathbf{Q}) = \mathbf{A}_z \mathbf{Q}, \tag{2.8}$$

where $\mathbf{A}_x$, $\mathbf{A}_y$, $\mathbf{A}_z$ are 2×2 hyperbolic matrices with constant entries. Hyperbolicity of these matrices ensures that each of them possesses two real eigenvalues. Therefore (2.7) presents two waves in each space direction that can be used for evaluating the two bias δ_j^+ and δ_j^-. Let us consider a cell T_i of the considered unstructured mesh, whose boundaries are S_j. Let $\vec{n}_j$ be the outward pointing normal unit vector and T_j the neighbouring cell associated to the current boundary and the initial condition (time $t = t^n$, local time $\tau = 0$) given by piecewise constant data, namely $\mathbf{Q}(x,y,z \in T_i) = \mathbf{Q}_i^n = [q_{1i}, q_{2i}]^T$ and $\mathbf{Q}(x,y,z \in T_j) = \mathbf{Q}_j^n = [q_{1j}, q_{2j}]^T$.

The sought reference method shall be monotone and characterised by minimum numerical dissipation. Stecca *et al.* [25] prove that the Godunov upwind scheme on structured meshes in two space dimensions for the linear advection equation 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 the result by Stecca *et al.* [25] and adopt the Godunov upwind method as reference. The Godunov upwind flux at interface S_j is given by

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{Godunov}\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n\right) = \underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}\left(\mathbf{Q}\left(\hat{n}=0, \tau=0^+\right)\right), \tag{2.9}$$

where $\mathbf{Q}(\hat{n}=0, \tau=0^+)$ 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 have to equate the projected fluxes

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{UFORCE-\delta}\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n\right) \cdot \vec{n}_j = \underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{Godunov}\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n\right) \cdot \vec{n}_j. \tag{2.10}$$

At this stage we apply both numerical fluxes to the system (2.7)-(2.8).

The Godunov upwind numerical flux (2.9) is obtained by solving the following Riemann problem:

$$\begin{cases} \partial_t \mathbf{Q} + \partial_{\hat{n}} (\mathbf{A}_{\hat{n}} \mathbf{Q}) = 0, \\ \begin{cases} \mathbf{Q}(\tau = 0, \hat{n}) = \mathbf{Q}_i^n, & \text{if } \hat{n} < 0, \\ \mathbf{Q}(\tau = 0, \hat{n}) = \mathbf{Q}_j^n, & \text{if } \hat{n} > 0, \end{cases} \end{cases} \tag{2.11}$$

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

$$\mathbf{A}_{\hat{n}} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \left(\mathbf{A}_x, \mathbf{A}_y, \mathbf{A}_z\right) \cdot \vec{n}_j . \tag{2.12}$$

Hyperbolicity of all matrices in (2.8) guarantees that the matrix $\mathbf{A}_{\hat{n}}$ is itself hyperbolic. Therefore $\mathbf{A}_{\hat{n}}$ possesses two real eigenvalues $\lambda_{\hat{n}}^{(1)}$ and $\lambda_{\hat{n}}^{(2)}$ (sorted in increasing order), defined as:

$$\lambda_{\hat{n}}^{(1)} = \frac{1}{2}(a_{11} + a_{22} - R), \qquad \lambda_{\hat{n}}^{(2)} = \frac{1}{2}(a_{11} + a_{22} + R), \tag{2.13}$$

with $R = \sqrt{(a_{11} - a_{22})^2 + 4 a_{12} a_{21}}$. The exact solution at cell interface for problem (2.11) is given by

$$\begin{aligned} &\mathbf{Q}\left(\hat{n} = 0, \tau = 0^+\right) \\ &= \begin{cases} \mathbf{Q}_i^n, & \text{if } \lambda_{\hat{n}}^{(1)} > 0, \\ \mathbf{Q}_j^n, & \text{if } \lambda_{\hat{n}}^{(2)} < 0, \\ \begin{bmatrix} \frac{R + a_{11} - a_{22}}{2R^+} q_{1i} + \frac{R - a_{11} + a_{22}}{2R^+} q_{1j} - \frac{a_{12}}{R^+}\left(q_{2j} - q_{2i}\right) \\ -\frac{a_{21}}{R^+}\left(q_{1j} - q_{1i}\right) + \frac{R - a_{11} + a_{22}}{2R^+} q_{2i} + \frac{R + a_{11} - a_{22}}{2R^+} q_{2j} \end{bmatrix}, & \text{otherwise}, \end{cases} \end{aligned} \tag{2.14}$$

where $R^+ = (R + \epsilon)$ allows to handle the case of two vanishing projected eigenvalues. The three-dimensional Godunov flux is obtained applying the flux operators (2.8) and (2.14) into (2.9), from which we obtain:

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{Godunov}\left(\mathbf{Q}_i^n, \mathbf{Q}_j^n\right) \cdot \vec{n}_j = \mathbf{A}_{\hat{n}} \mathbf{Q}\left(\hat{n} = 0, \tau = 0^+\right). \tag{2.15}$$

The three-dimensional UFORCE-δ flux for (2.7)-(2.8) is obtained by substitution of (2.8) into (2.3)-(2.6). The projected UFORCE-δ flux reads

$$\begin{aligned} &\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{UFORCE-\delta} \cdot \vec{n}_j \\ &= \frac{1}{2}\left\{ \mathbf{A}_{\hat{n}} - \frac{1}{\left(\delta_j^- h_j^- + \delta_j^+ h_j^+ + \epsilon\right)} \left(\frac{\alpha \Delta t}{2} \mathbf{A}_{\hat{n}}^2 + \frac{2\left(\delta_j^- h_j^- \delta_j^+ h_j^+\right)}{\alpha \Delta t} \mathbf{I} \right) \right\} \left(\mathbf{Q}_j^n - \mathbf{Q}_i^n\right), \end{aligned} \tag{2.16}$$

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

Equating the UFORCE-δ flux (2.16) to the Godunov flux (2.15), after algebraic manipulations, two solutions for the optimal upwind bias can be found:

$$\delta_j^- = \frac{\left|\lambda_{\hat{n}}^{(1)}\right|\alpha\Delta t}{2h_j^-}, \qquad \delta_j^+ = \frac{\left|\lambda_{\hat{n}}^{(2)}\right|\alpha\Delta t}{2h_j^+}, \tag{2.17}$$

$$\delta_j^- = \frac{\left|\lambda_{\hat{n}}^{(2)}\right|\alpha\Delta t}{2h_j^-}, \qquad \delta_j^+ = \frac{\left|\lambda_{\hat{n}}^{(1)}\right|\alpha\Delta t}{2h_j^+}. \tag{2.18}$$

In order to satisfy constraint (2.2), the following condition must hold:

$$\frac{\max\left(\left|\lambda_{\hat{n}}^{(1)}\right|,\left|\lambda_{\hat{n}}^{(2)}\right|\right)\Delta t}{2h_j^-} \leq \frac{1}{\alpha}, \tag{2.19}$$

for each h_j in the domain. Therefore the time step Δt will be chosen according to the following relationship:

$$\Delta t = \frac{2}{\alpha} CFL \min_{1\leq i\leq n_e}\left(\min_{1\leq j\leq n_f}\left(\frac{h_j^-}{\max\left(\left|\lambda_{\hat{n}}^{(1)}\right|,\left|\lambda_{\hat{n}}^{(2)}\right|\right)}\right)\right)_{T_i}, \tag{2.20}$$

being CFL the Courant-Friedrichs-Lewy coefficient ($0 < CFL \leq 1$).

It is worth mentioning that for the linear case the two solutions (2.17) and (2.18) are equivalent giving the same numerical results.

We are now ready to write our UFORCE-δ flux for the considered problem by inserting (2.17) or (2.18) into (2.16):

$$\begin{aligned}&\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{UFORCE-\delta}\cdot\vec{n}_j \\ &= \frac{1}{2}\left\{\mathbf{A}_{\hat{n}} - \frac{1}{\max\left(\left|\lambda_{\hat{n}}^{(1)}\right|+\left|\lambda_{\hat{n}}^{(2)}\right|+\epsilon\right)}\left(\mathbf{A}_{\hat{n}}^2+\left|\lambda_{\hat{n}}^{(1)}\right|\left|\lambda_{\hat{n}}^{(2)}\right|\mathbf{I}\right)\right\}\left(\mathbf{Q}_j^n-\mathbf{Q}_i^n\right).\end{aligned} \tag{2.21}$$

Comparing (2.16) and (2.21) we observe that the original cell subvolume altitudes $h_j^{\pm}$ have been replaced by adaptive subvolumes whose size is controlled by 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 the characteristic speeds. We remark that given the optimal upwind bias (2.17) or (2.18), the proposed UFORCE-δ method identically reproduces the results of the Godunov upwind method in applications to the linear system (2.7)-(2.8). In Section 4 we will experimentally prove this statement. In the next section we explain how to extend the proposed flux to general non-linear hyperbolic systems.

2.3 Extension to non-linear hyperbolic systems of PDEs

The aim of this section is to find a suitable modification of relationships (2.17) and (2.18) 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 $m \geq 2$:

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_y \mathbf{G}(\mathbf{Q}) + \partial_z \mathbf{H}(\mathbf{Q}) = \mathbf{0}, \tag{2.22}$$

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 flux vectors $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$ are defined as:

$$\mathbf{A}_x(\mathbf{Q}) = \left[\frac{\partial \mathbf{F}}{\partial \mathbf{Q}}\right], \qquad \mathbf{A}_y(\mathbf{Q}) = \left[\frac{\partial \mathbf{G}}{\partial \mathbf{Q}}\right], \qquad \mathbf{A}_z(\mathbf{Q}) = \left[\frac{\partial \mathbf{H}}{\partial \mathbf{Q}}\right]. \tag{2.23}$$

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

$$\mathbf{A}_{\hat{n}}(\mathbf{Q}) = \left(\mathbf{A}_x(\mathbf{Q}), \mathbf{A}_y(\mathbf{Q}), \mathbf{A}_z(\mathbf{Q})\right) \cdot \vec{n}_j. \tag{2.24}$$

Unlike in the linear case, Jacobians (2.23) 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 $\mathbf{Q}\left(x,y,z \in T_j\right) = \mathbf{Q}_j^n$, we have to consider two projected Jacobian matrices:

$$\mathbf{A}_{i\hat{n}} = \mathbf{A}_{\hat{n}}\left(\mathbf{Q}_i^n\right), \qquad \mathbf{A}_{j\hat{n}} = \mathbf{A}_{\hat{n}}\left(\mathbf{Q}_j^n\right). \tag{2.25}$$

Note that the two projected Jacobians (2.25) have been obtained with different data, but using the same unit vector $\vec{n}_j$ (outward pointing from cell T_i). Each of these Jacobian matrices possesses m real eigenvalues (sorted in increasing order), namely $\lambda_{i\hat{n}}^{(1)}, \cdots, \lambda_{i\hat{n}}^{(m)}$ and $\lambda_{j\hat{n}}^{(1)}, \cdots, \lambda_{j\hat{n}}^{(m)}$ 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 speed can be estimated in practice in order to use an expression based on (2.17) or (2.18) for the upwind bias?

Concerning the first question we consider a two-wave approach similar to HLL [11] or to the central-upwind method developed by Kurganov, Noelle and Petrova (in the following KNP) [14]. Following this approach we conclude that for system possessing more than two waves we have to consider the smallest and largest characteristic speed of the entire Riemann fan. In the following we denote these wave speeds as $s_n^{(1)}$ and $s_n^{(m)}$. Therefore the optimal upwind bias can be written as

$$\delta_j^- = \frac{\left|s_{\hat{n}}^{(1)}\right| \alpha \Delta t}{2h_j^-}, \qquad \delta_j^+ = \frac{\left|s_{\hat{n}}^{(m)}\right| \alpha \Delta t}{2h_j^+}, \tag{2.26}$$

$$\delta_j^- = \frac{\left|s_{\hat{n}}^{(m)}\right| \alpha \Delta t}{2h_j^-}, \qquad \delta_j^+ = \frac{\left|s_{\hat{n}}^{(1)}\right| \alpha \Delta t}{2h_j^+}. \tag{2.27}$$

The associated CFL condition reads

$$\Delta t = \frac{2}{\alpha} CFL \min_{1\le i\le n_e} \left(\min_{1\le j\le n_f} \left(\frac{h_j^-}{\max\left(\left|s_{\hat{n}}^{(1)}\right|, \left|s_{\hat{n}}^{(m)}\right| \right)} \right) \right)_{T_i}, \tag{2.28}$$

with $0 < CFL \le 1$.

The second question requires more discussion and different approaches are viable. Hereby we present two possible choices leading to genuinely centred method:

1. The first approach considers the smallest and largest wave speeds of the Riemann fan, which are obtained using directly the eigenvalues of the projected Jacobian matrices (2.25) as linearised wave speed estimates, i.e.:

$$\left|s_{\hat{n}}^{(1)}\right| = \max\left(\left|\lambda_{i\hat{n}}^{(1)}\right|, \left|\lambda_{j\hat{n}}^{(1)}\right| \right), \qquad \left|s_{\hat{n}}^{(m)}\right| = \max\left(\left|\lambda_{i\hat{n}}^{(m)}\right|, \left|\lambda_{j\hat{n}}^{(m)}\right| \right). \tag{2.29}$$

   The upwind bias is obtained inserting (2.29) into (2.27). Since these eigenvalues are in any case needed for selecting a time step, no computational effort is added to the global method. Based on our experience with the two-dimensional inviscid shallow water equations, in order to avoid spurious oscillations we recommend use of (2.27). The resulting method is genuinely centred since the system eigenstructure does not have to be known in details, 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 (1.7)-(1.10), 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:

$$\left|s_{\hat{n}}^{(1)}\right| = \left|s_{\hat{n}}^{(m)}\right| = \max\left(\max_{1\le l\le m}\left(\left|\lambda_{i\hat{n}}^{(l)}\right| \right), \max_{1\le l\le m}\left(\left|\lambda_{j\hat{n}}^{(l)}\right| \right) \right). \tag{2.30}$$

   In this case use of (2.26) or (2.27) leads to 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 given by (2.29), but still more accurate than the FORCE method.

Finally, the UFORCE-δ method (two-wave version) can be summarised:

- computation of two sets of eigenvalues of the projected Jacobians (2.25) across each boundary S_j ($\lambda_{i\hat{n}}^{(1)},\cdots,\lambda_{i\hat{n}}^{(m)}$ and $\lambda_{j\hat{n}}^{(1)},\cdots,\lambda_{j\hat{n}}^{(m)}$);
- selection of a time step Δt using (2.28);
- computation of the optimal upwind bias (2.27) using linearised wave speeds (2.29);
- application of the update formula (1.3).

2.4 The UFORCE-δ flux on Cartesian meshes

In this section we derive a Cartesian formulation for the UFORCE-δ flux (2.3)-(2.6). The formulation here proposed is also suitable for one-dimensional applications. We focus again on the non-linear system (2.22) and we consider only fluxes in the x-direction, while the other fluxes can be written in analogous manner. We use a Cartesian-type indexation of cells: therefore 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 such as $\mathbf{Q}_i^n$ and $\mathbf{Q}_{i+1}^n$. The UFORCE-δ flux at interface $S_{i+\frac{1}{2}}$ is given by

$$\mathbf{F}_{i+\frac{1}{2}}^{UFORCE-\delta}=\frac{1}{2}\left\{\mathbf{F}_{i+\frac{1}{2}}^{ULW-\delta}\left(\mathbf{Q}_i^n,\mathbf{Q}_{i+1}^n\right)+\mathbf{F}_{i+\frac{1}{2}}^{ULF-\delta}\left(\mathbf{Q}_i^n,\mathbf{Q}_{i+1}^n\right)\right\}, \tag{2.31}$$

with

$$\mathbf{F}_{i+\frac{1}{2}}^{ULW-\delta}=\mathbf{F}\left(\mathbf{Q}_{i+\frac{1}{2}}^{ULW-\delta}\right), \tag{2.32}$$

$$\mathbf{Q}_{i+\frac{1}{2}}^{ULW-\delta}=\frac{1}{\left(\delta_{i+\frac{1}{2}}^{-}+\delta_{i+\frac{1}{2}}^{+}+\epsilon\right)}\left\{\left(\mathbf{Q}_i^n\delta_{i+\frac{1}{2}}^{-}+\mathbf{Q}_{i+1}^n\delta_{i+\frac{1}{2}}^{+}\right)-\frac{\alpha\Delta t}{\Delta x}\left(\mathbf{F}\left(\mathbf{Q}_{i+1}^n\right)-\mathbf{F}\left(\mathbf{Q}_i^n\right)\right)\right\}, \tag{2.33}$$

$$\begin{aligned}\mathbf{F}_{i+\frac{1}{2}}^{ULF-\delta}=\frac{1}{\left(\delta_{i+\frac{1}{2}}^{-}+\delta_{i+\frac{1}{2}}^{+}+\epsilon\right)}\Bigg\{&\left(\mathbf{F}\left(\mathbf{Q}_{i+1}^n\right)\delta_{i+\frac{1}{2}}^{-}+\mathbf{F}\left(\mathbf{Q}_i^n\right)\delta_{i+\frac{1}{2}}^{+}\right)\\&-\delta_{i+\frac{1}{2}}^{-}\delta_{i+\frac{1}{2}}^{+}\frac{\Delta x}{\alpha\Delta t}\left(\mathbf{Q}_{i+1}^n-\mathbf{Q}_i^n\right)\Bigg\},\end{aligned} \tag{2.34}$$

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 (2.3)-(2.6) holds. Note that imposing $\delta^{\pm}=1$ the FORCE-α method on Cartesian meshes [29] is obtained. The optimal upwind bias are given by

$$\delta_{i+\frac{1}{2}}^{-}=\frac{\left|s_x^{(m)}\right|\alpha\Delta t}{\Delta x},\qquad \delta_{i+\frac{1}{2}}^{+}=\frac{\left|s_x^{(1)}\right|\alpha\Delta t}{\Delta x}, \tag{2.35}$$

where $\left|s_x^{(1)}\right|$, and $\left|s_x^{(m)}\right|$ 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 $\mathbf{A}_x$ defined in (2.23). Across the current interface $S_{i+\frac{1}{2}}$ we have two Jacobians, namely:

$$\mathbf{A}_i = \mathbf{A}_x\left(\mathbf{Q}_i^n\right), \qquad \mathbf{A}_{i+1} = \mathbf{A}_x\left(\mathbf{Q}_{i+1}^n\right), \tag{2.36}$$

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

$$\left|s_x^{(1)}\right| = \max\left(\left|\lambda_i^{(1)}\right|,\left|\lambda_{i+1}^{(1)}\right|\right), \qquad \left|s_x^{(m)}\right| = \max\left(\left|\lambda_i^{(m)}\right|,\left|\lambda_{i+1}^{(m)}\right|\right). \tag{2.37}$$

3 Second-order extension

The first-order UFORCE-δ flux (2.3)-(2.6) 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 i) a non-oscillatory polynomial reconstruction of degree $p-1$ and ii) a temporal evolution technique. Concerning the reconstruction technique, several different procedures are available. Total Variation Diminishing (TVD) schemes are most frequently used for second-order accuracy (see e.g. [20,33]), while essentially non-oscillatory (ENO) schemes (see e.g. [4,10]) or weighted essentially non-oscillatory (WENO) schemes (see e.g. [7,8,12,19]) are used for accuracy $p\geq 2$. Moreover, concerning the temporal evolution of reconstructed polynomials, several different techniques are available, including the method-of-lines based on Runge-Kutta time stepping (see e.g. [24]), the one-step ADER technique based on the semi-analytical Cauchy-Kowalewski procedure (see e.g. [30,32]). Recently, the discontinuous Galerkin predictor has been successfully applied to design fully-numeric Cauchy-Kowalewski free one-step schemes, see e.g. [5].

In this section we briefly review the one-step ADER-WENO technique used in this paper. For the applications presented in Section 4 we specialise our review on the achievement of second-order accuracy in the solution of two-dimensional systems having the form

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

to be solved on unstructured triangular meshes. We remark however that there is no theoretical barrier for the implementation of methods based on the UFORCE-δ flux up to any desired order of accuracy up to three space dimensions using the same standard techniques (see [7,8] and references therein for details).

3.1 Nonlinear reconstruction technique

Here we briefly discuss the proposed nonlinear weighted essentially non-oscillatory (WENO) reconstruction procedure to reconstruct polynomial data within each spatial

cell at time t^n from the given cell averages. We emphasize that the reconstruction procedure is nonlinear and depends strongly on the input data. Thus, the resulting numerical scheme, even when applied to a completely linear PDE, will be nonlinear and thus it will not be possible to give a closed expression of the scheme. The reconstruction procedure follows directly from the guidelines given in [7,8] for general unstructured two- and three-dimensional meshes. Since applications will be carried out in the two-dimensional case, here we refer only to the case of two-dimensional triangular meshes. The procedure reconstructs entire polynomials, as the original ENO approach proposed by Harten *et al.* [10]. However, we formally write our method like a WENO scheme [12, 19] with a particularly simple choice for the linear weights. For each stencil $\mathcal{S}_i^s = \bigcup T_k$ [7], we require integral conservation for the reconstruction polynomial $\mathbf{w}_i^s$:

$$\frac{1}{|T_k|}\int_{T_k} \mathbf{w}_i^s(\vec{x},t^n)\,\mathrm{d}\vec{x} = \mathbf{Q}_k^n, \qquad \forall T_k \in \mathcal{S}_i^s. \tag{3.2}$$

The reconstruction equations (3.2) are solved using a constrained least-squares method in order to guarantee that (3.2) is exactly satisfied at least inside element T_i [7]. 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 following transformation:

$$x = X_1 + (X_2 - X_1)\xi + (X_3 - X_1)\eta, \qquad y = Y_1 + (Y_2 - Y_1)\xi + (Y_3 - Y_1)\eta, \tag{3.3}$$

whose Jacobian matrix is defined as

$$\mathbf{J} = \left[\frac{\partial \vec{x}}{\partial \vec{\xi}}\right]. \tag{3.4}$$

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 $\mathbf{w}_i^s\left(\vec{\xi},t^n\right)$:

$$\mathbf{w}_i\left(\vec{\xi},t^n\right) = \sum_{s=1}^{7} \omega_s \mathbf{w}_i^s\left(\vec{\xi},t^n\right). \tag{3.5}$$

It is worth noting that seven stencils are required to be used in (3.5): one centred stencil ($s=1$), three forward sector stencils ($s=2,3,4$) and three reverse sector stencils ($s=5,6,7$) [7]. The nonlinear WENO weights ω_s are computed as follows

$$\omega_s = \frac{\widetilde{\omega}_s}{\sum_{k=1}^{7}\widetilde{\omega}_k}, \qquad \widetilde{\omega}_s = \frac{\lambda_s}{(\sigma_s + \epsilon^W)^r}, \qquad \lambda_s = \begin{cases} 10^2 \div 10^5, & \text{if } s=1, \\ 1, & \text{otherwise,} \end{cases} \tag{3.6}$$

with the oscillation indicators σ_s defined in [7], $r=4$ and $\epsilon^W = 10^{-5}$.

3.2 Second-order accurate one-step time discretisation

The result of the reconstruction procedure is a non-oscillatory spatial polynomial $\mathbf{w}_i(\vec{\xi},t^n)$ defined at time t^n inside each spatial element T_i. However, we still 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 our final second-order accurate one-step finite volume scheme. Second-order accuracy is obtained using the ADER approach [32]. The key idea is to solve high-order Riemann problems at the element boundaries. This is achieved by using a Taylor series expansion in time, use of the Cauchy-Kowalewski procedure and 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 $\mathbf{Q}_i(\vec{\xi},t)$ of the PDE in each cell in a space-time Taylor series with respect to the barycentre element $(\xi_i,\eta_i)=\left(\frac{1}{3},\frac{1}{3}\right)$

$$\mathbf{Q}_i(\xi,\eta,t)=\mathbf{Q}(\xi_i,\eta_i,t^n)+(\xi-\xi_i)\partial_\xi\mathbf{Q}+(\eta-\eta_i)\partial_\eta\mathbf{Q}+(t-t^n)\partial_t\mathbf{Q}+\mathcal{O}\left(\xi^2,\eta^2,t^2\right). \tag{3.7}$$

Then we use the Cauchy-Kowalewski procedure in order to substitute time derivatives with space derivatives in (3.7). To this aim we rewrite the governing PDE system (3.1) in the transformed coordinate space

$$\partial_t\mathbf{Q}+\partial_\xi\mathbf{F}'(\mathbf{Q})+\partial_\eta\mathbf{G}'(\mathbf{Q})=\mathbf{0}, \tag{3.8}$$

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

$$\mathbf{F}'(\mathbf{Q})=\mathbf{F}(\mathbf{Q})\partial_x\xi+\mathbf{G}(\mathbf{Q})\partial_y\xi,\qquad \mathbf{G}'(\mathbf{Q})=\mathbf{F}(\mathbf{Q})\partial_x\eta+\mathbf{G}(\mathbf{Q})\partial_y\eta, \tag{3.9}$$

being $\partial_x\xi,\cdots,\partial_y\eta$ the (constant) entries of the inverse of the transformation Jacobian (3.4). For second-order accuracy it suffices to obtain the first time derivative by differentiating (3.8) as follows:

$$\partial_t\mathbf{Q}=-\left(\mathbf{A}'_\xi\partial_\xi\mathbf{Q}+\mathbf{A}'_\eta\partial_\eta\mathbf{Q}\right), \tag{3.10}$$

where $\mathbf{A}'_\xi$ and $\mathbf{A}'_\eta$ are the Jacobian matrices of fluxes (3.9) in the transformed space

$$\mathbf{A}'_\xi=\left[\frac{\partial\mathbf{F}'}{\partial\mathbf{Q}}\right],\qquad \mathbf{A}'_\eta=\left[\frac{\partial\mathbf{G}'}{\partial\mathbf{Q}}\right]. \tag{3.11}$$

The value of $\mathbf{Q}(\xi_i,\eta_i,t^n)$ and its spatial derivatives are obtained from the WENO reconstruction polynomial $\mathbf{w}_i(\vec{\xi},t^n)$. For an efficient implementation up to any order of accuracy in space and time we refer the reader to [7,8] and references therein.

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

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

follows:

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{\Delta t}{|T_i|} \sum_{j=1}^{3} |S_j| \left(\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}} \cdot \vec{n}_j \right), \tag{3.12}$$

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

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}} = \frac{1}{\left|S_j'\right|} \int_{S_j'} \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{UFORCE-\delta} \left(\mathbf{Q}_i \left(\vec{\xi}, t \right), \mathbf{Q}_j \left(\vec{\xi}, t \right) \right) \mathrm{d}t \mathrm{d}\vec{\xi}, \tag{3.13}$$

where S_j' is the counterpart of edge S_j in the transformed coordinate system, $\left|S_j'\right|$ represents its length, $\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{UFORCE-\delta}$ is the UFORCE-δ flux given by (2.3)-(2.6) 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 (3.13) 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 integration step. Therefore in practice we use

$$\underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}} = \underline{\underline{\mathbf{F}}}_{j+\frac{1}{2}}^{UFORCE-\delta} \left(\mathbf{Q}_i \left(\vec{\xi}_M, t^{n+\frac{1}{2}} \right), \mathbf{Q}_j \left(\vec{\xi}_M, t^{n+\frac{1}{2}} \right) \right), \tag{3.14}$$

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 extension to higher order, where quadratures may get computationally heavy, this approach can be modified using fully-analytical procedures, see [7] for details.

4 Numerical applications

The purpose of this section is to assess the performance of the proposed UFORCE-δ method comparing numerical results with both exact and numerical solutions obtained using well-established centred and upwind finite volume methods.

4.1 Applications to the two-dimensional inviscid shallow water equations

Here we apply the UFORCE-δ method (2.3)-(2.6) to well-established test problems for the two-dimensional non-linear inviscid shallow water equations augmented by an equation for a passive scalar. The system written in conservative form reads:

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_y \mathbf{G}(\mathbf{Q}) = \mathbf{0}, \tag{4.1}$$

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

$$\mathbf{Q}=\begin{bmatrix} D \\ Du \\ Dv \\ DC \end{bmatrix}, \qquad \mathbf{F}(\mathbf{Q})=\begin{bmatrix} Du \\ Du^2+\frac{1}{2}gD^2 \\ Duv \\ DuC \end{bmatrix}, \qquad \mathbf{G}(\mathbf{Q})=\begin{bmatrix} Dv \\ Duv \\ Dv^2+\frac{1}{2}gD^2 \\ DvC \end{bmatrix}. \tag{4.2}$$

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 $g=9.81ms^{-2}$ is the acceleration due to gravity. Jacobian matrices of fluxes $\mathbf{F}(\mathbf{Q})$ and $\mathbf{G}(\mathbf{Q})$ have three distinct eigenvalues, namely

$$\begin{bmatrix} \lambda_x^{(1)} \\ \lambda_x^{(2)} \\ \lambda_x^{(3)} \end{bmatrix}=\begin{bmatrix} u-a \\ u \\ u+a \end{bmatrix}, \qquad \begin{bmatrix} \lambda_y^{(1)} \\ \lambda_y^{(2)} \\ \lambda_y^{(3)} \end{bmatrix}=\begin{bmatrix} v-a \\ v \\ v+a \end{bmatrix}, \tag{4.3}$$

where $a=\sqrt{gD}$ is celerity. Given a unit vector $\vec{n}_j$, the projected Jacobian matrix, defined by (2.24), has three distinct real eigenvalues, namely

$$\begin{bmatrix} \lambda_{\hat{n}}^{(1)} \\ \lambda_{\hat{n}}^{(2)} \\ \lambda_{\hat{n}}^{(3)} \end{bmatrix}=\begin{bmatrix} u_{\hat{n}}-a \\ u_{\hat{n}} \\ u_{\hat{n}}+a \end{bmatrix}, \tag{4.4}$$

where $u_{\hat{n}}=(u,v)\cdot\vec{n}_j$ is the velocity projection. $\lambda_{\hat{n}}^{(2)}$ has multiplicity 2. The equation for transport of a passive scalar has been introduced in order to analyse the performance of our method in presence of contact waves, which usually poses difficulties to all centred methods.

Four test problems have been chosen in order to assess the behaviour of the UFORCE-δ method. Two of them (the collapse of a circular dam and the propagation of a passive scalar discontinuity) have been solved using first-order accurate methods on Cartesian meshes, while the latter two tests (namely the advection of a potential vortex and the collapse of a circular dam solved on a variably-spaced grid) have been performed on triangular unstructured meshes using second-order extension of methods.

4.1.1 Collapse of a circular dam

This test case consists of the instantaneous breaking of a cylindrical tank initially filled with 2.5 meter deep water 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. The wave generated by the breaking of the tank propagates into still water with an initial depth of $0.5m$.

We solve (4.1), (4.2) together with initial conditions

$$\begin{cases} D(x,y,0)=2.5m, & \text{if } x^2+y^2\leq R^2, \\ D(x,y,0)=0.5m, & \text{if } x^2+y^2>R^2, \\ u(x,y,0)=v(x,y,0)=0, & \forall x,y, \end{cases} \tag{4.5}$$

being $R=2.5m$ the tank radius. With this test we aim to assess the ability of the UFORCE-δ method in 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. Time step has been selected using the following CFL condition:

$$\Delta t=\frac{CFL}{2}\min\left(\min_{i,j}\frac{\Delta x}{(|u|+a)_{i,j}},\min_{i,j}\frac{\Delta y}{(|v|+a)_{i,j}}\right), \tag{4.6}$$

which comes from (2.28) with the eigenvalues given by (4.3). Results obtained with the proposed method, which uses information from the largest and smallest wave of the Riemann fan (two-wave method), have been compared with those obtained using one centred method (the FORCE method), two central-upwind methods (the UFORCE method [25] and the KNP method [14]) 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 approximated Riemann solver (Godunov-HLL). These methods can be classified considering the number of waves which are taken in 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). This comparison has been carried out with the first order version of the above mentioned numerical methods.

We provide an accurate reference solution, which was obtained by turning the problem (4.1, 4.2, 4.5) into a one-dimensional problem in the radial direction (see [26] for details):

$$\partial_t\begin{bmatrix} D \\ Du_r \end{bmatrix}+\partial_r\begin{bmatrix} Du_r \\ Du_r^2+\frac{1}{2}gD^2 \end{bmatrix}=-\frac{1}{r}\begin{bmatrix} Du_r \\ Du_r^2 \end{bmatrix}, \tag{4.7}$$

where r is the radial coordinate and $u_r(r,t)$ the radial velocity. The initial conditions (4.5) in the radial coordinate system read:

$$\begin{cases} D(r,0)=2.5m, & \text{if } r\leq R, \\ D(r,0)=0.5m, & \text{if } r>R, \\ u(r,0)=0, & \forall r. \end{cases} \tag{4.8}$$

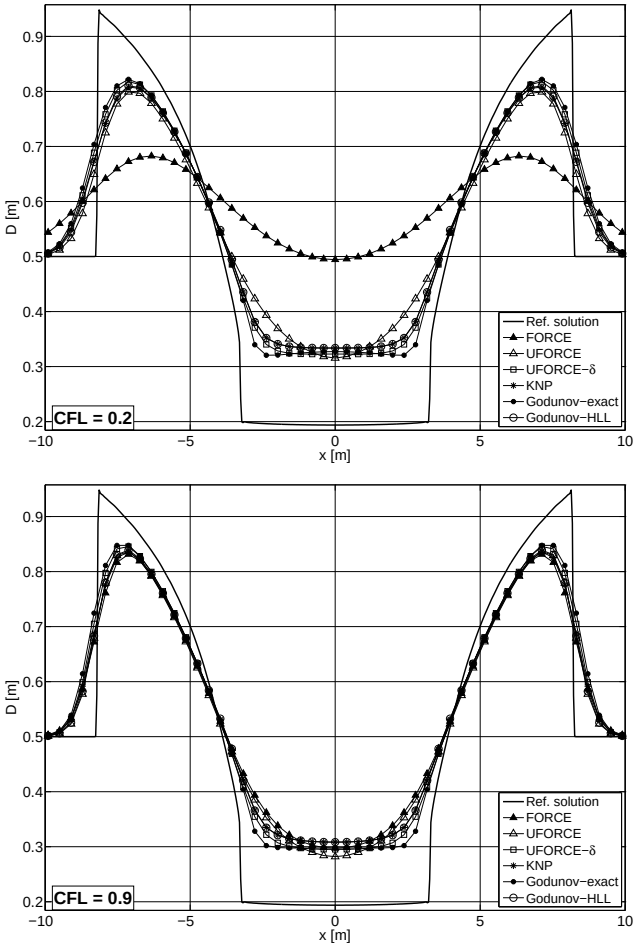

Figure 3: Collapse of a circular dam. Numerical results for water depth D of six numerical methods (symbols) are compared with the reference radial solution (full line) at time $t=1.4s$. The numerical solution profiles are sliced on the x-axis. The mesh used is 101×101 cells and CFL is set to 0.2 (top panel) and 0.9 (bottom panel).

The reference solution is obtained solving numerically system (4.7), (4.8) on a fine mesh of 1000 cells using the WAF method in conjunction with the HLLC approximate Riemann solver [26]. The CFL number is set to 0.9 and the limiter used is SUPERBEE [23].

Results for water depth $D(x,y,t)$ are displayed in Fig. 3 at time $t=1.4s$ and in Fig. 4 at time $t=3.5s$. Here, the numerical solution (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 and 4 the Godunov-exact method, based on exact evaluation of all the three waves, is found to be the least dissipative among all method and to have a consistent performance at low and high CFL numbers. In contrast, accuracy of the FORCE centred method significantly depends on the CFL number. While the solution obtained

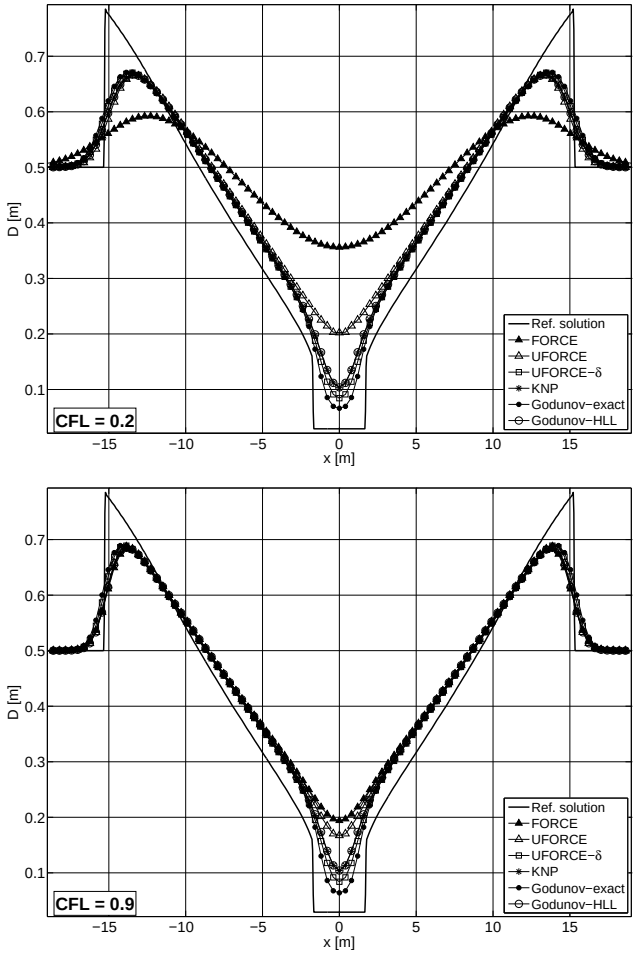

Figure 4: Collapse of a circular dam. Numerical results for water depth D of six numerical methods (symbols) are compared with the reference radial solution (full line) at time $t=3.5s$. The numerical solution profiles are sliced on the x-axis. The mesh used is 101×101 cells and CFL is set to 0.2 (top panel) and 0.9 (bottom panel).

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 at $CFL=0.9$ FORCE fails in the accurate description of inflections in the free-surface profile (Fig. 4 around $x=0$). The one-wave UFORCE method increases significantly the accuracy of the solution 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 at $x=\pm3$). 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 the CFL number. The UFORCE-δ solution profile is always bounded between that of Godunov-exact and that of KNP and HLL.

We speculate that improved accuracy compared to HLL and KNP is due to improved resolution of contact waves, i.e. the shear wave in this case. This feature of UFORCE-δ is proved in the next section.

4.1.2 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 $1m$ on the left side of the domain and $0.5m$ 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 (4.1), (4.2) with initial conditions

$$\begin{cases} \begin{cases} D(x,y,0)=1m, \\ C(x,y,0)=1, \end{cases} & \text{if } x\leq 0, \quad \begin{cases} D(x,y,0)=0.5m, \\ C(x,y,0)=0, \end{cases} \quad \text{if } x>0, \\ u(x,y,0)=v(x,y,0)=0, & \forall x,y. \end{cases} \tag{4.9}$$

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 $t=5s$.

The exact solution for this problem can be computed by solving a one-dimensional dam-break problem in the x-direction using an exact Riemann solver. The exact solution contains a left rarefaction, a right-facing shock wave and a contact discontinuity in the middle, across which the concentration C varies discontinuously (see [31] for an accurate description). We focus our attention on the contact discontinuity and discuss the results in terms of C. In general, 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 (see [27]). 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. Results for this test, obtained with $CFL=0.2$ and $CFL=0.9$, are displayed in Fig. 5 (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 this wave, outperforming all the other

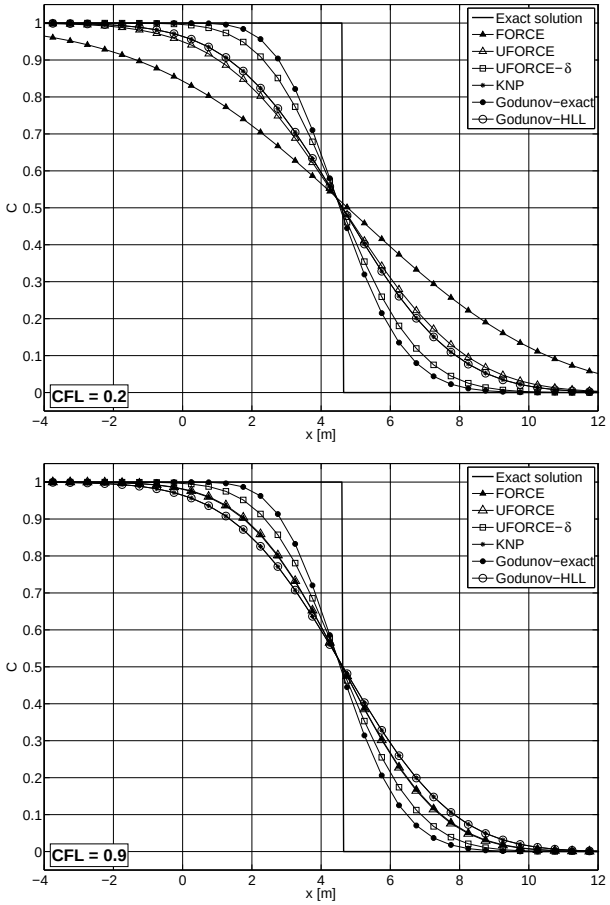

Figure 5: Propagation of a passive scalar discontinuity. Numerical results for concentration C of six numerical methods (symbols) are compared with the exact solution (full line) at time $t=5s$. The numerical solution profiles are along on x-axis. The mesh used is 100×100 cells and CFL is set to 0.2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

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. 5 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.

4.1.3 Vortex advection: assessment of second-order accuracy

In the present test, originally proposed by Ricchiuto and Bollermann [22], we examine the case of a vortex travelling at mean velocity $\mathbf{U}_\infty=(u_\infty,v_\infty)$ 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

$$\mathbf{U}=\mathbf{U}'+\mathbf{U}_\infty, \tag{4.10}$$

where $\mathbf{U}'$ in cylindrical coordinates reads:

$$\mathbf{U}'=\left(u_r',u_\theta'\right)=\left(0,u_\theta'\right) \tag{4.11}$$

being u_r' and u_θ' relative radial and tangential velocity respectively. Then, the first equation in (4.1)-(4.2) becomes

$$\partial_t D+\mathbf{U}_\infty\cdot\nabla D=0, \tag{4.12}$$

which admits the following exact solution:

$$D(x,y,t)=D_0(\vec{\gamma}), \tag{4.13}$$

being $\vec{\gamma}=(x-u_\infty t,y-v_\infty t)$ and $D_0(x,y)$ the initial condition for water depth distribution. Substituting (4.10), (4.11) into the second and third equation in (4.1)-(4.2) we obtain

$$\partial_t\mathbf{U}'+(\mathbf{U}_\infty\cdot\nabla)\mathbf{U}'+\left(\mathbf{U}'\cdot\nabla\right)\mathbf{U}'+g\nabla D_0(\vec{\gamma})=\mathbf{0}, \tag{4.14}$$

which admits an exact solution of the form

$$\mathbf{U}'(x,y,t)=\mathbf{U}_0'(\vec{\gamma}), \tag{4.15}$$

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

$$\mathbf{U}_0'(r_c)=\begin{cases}\Gamma(1+\cos(\omega r_c))(y_c-y,x-x_c), & \text{if } \omega r_c\leq\pi,\\ (0,0), & \text{otherwise,}\end{cases} \tag{4.16}$$

being Γ the vortex intensity, (x_c,y_c) 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 (4.14) along the radial direction yields the initial conditions for the water surface profile

$$D_0(r_c)=D_\infty+\begin{cases}\frac{1}{g}\left(\frac{\Gamma}{\omega}\right)^2(\phi(\omega r_c)-\phi(\pi)), & \text{if } \omega r_c\leq\pi,\\ 0, & \text{otherwise,}\end{cases} \tag{4.17}$$

Table 1: 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_∞	$\mathcal{O}(L_\infty)$	L_1	$\mathcal{O}(L_1)$	L_2	$\mathcal{O}(L_2)$
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 2: 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_∞	$\mathcal{O}(L_\infty)$	L_1	$\mathcal{O}(L_1)$	L_2	$\mathcal{O}(L_2)$
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: 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_∞	$\mathcal{O}(L_\infty)$	L_1	$\mathcal{O}(L_1)$	L_2	$\mathcal{O}(L_2)$
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

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

Following [22], the parameters used in computations are: $\Gamma=15$, $\omega=4\pi$, $\mathbf{U}_\infty=(6,0)$, $D_\infty=5$, $g=1$. 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 output time 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 (2.28) is applied, using projected eigenvalues (4.4), $\alpha=2$ and $CFL=0.9$. We solve the problem using the second-order ADER-WENO method together with the FORCE, UFORCE-δ and HLL flux.

In Table 1 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 Tables 2 and 3 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 Tables 1 and 2 we assess the great improvement of UFORCE-δ over FORCE (the norms of UFORCE-δ are about half of these of FORCE) and comparing the norms in Tables 1 and 3 we observe a slight but significant improvement with respect to HLL.

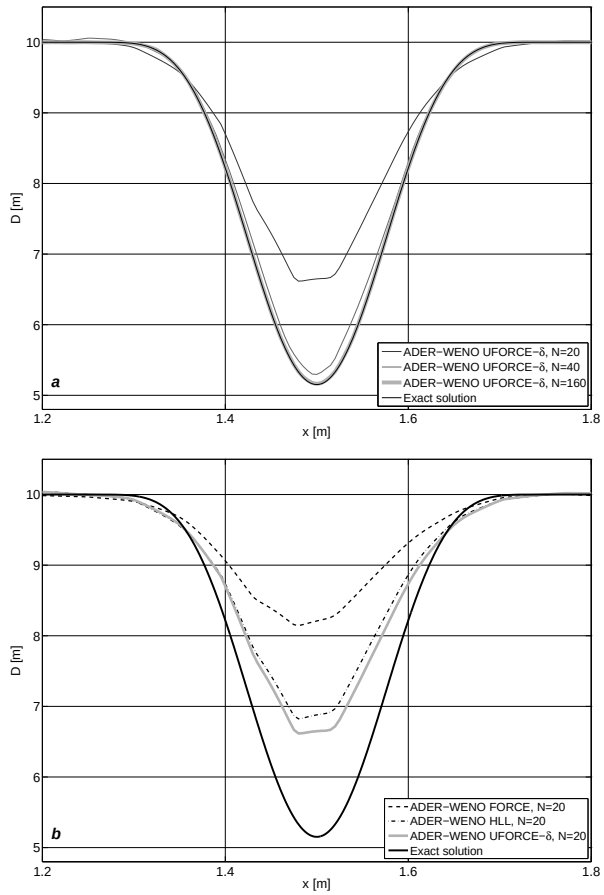

Figure 6: Vortex advection. Solution profiles are sliced along the x-axis. a) Numerical results for water depth D of the second-order ADER-WENO UFORCE-δ method obtained with different grid resolution (grey lines) are compared with the exact solution (black line) at time $t=1/6s$. N denotes the reciprocal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

In Fig. 6(a) we show the convergence of the proposed UFORCE-δ 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. 6(b) 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.

4.1.4 Collapse of a circular dam on a variably-spaced grid

As it was shown in the tests performed on Cartesian meshes, an attractive feature of the UFORCE-δ method compared to FORCE relies on its ability to perform consistently in

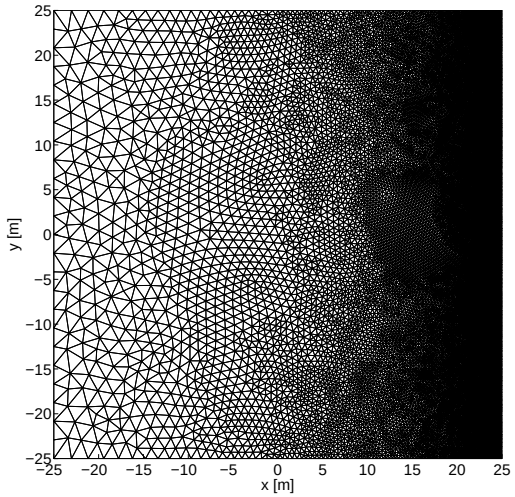

Figure 7: Collapse of a circular dam on a variably-spaced grid: variably-spaced mesh of 34753 triangles. Mesh length ranges from 2.08 on the left boundary to 0.15 on the right boundary.

the full range of stable CFL numbers. This fact has important consequences in practical applications when the shallow water equations are solved over irregular domains where a wide range of CFL numbers from small to large is generated. In order to highlight this behaviour we solve again the problem defined by (4.1), (4.2), (4.5) in the square computational domain $[-25,25]\times[-25,25]$. We use an irregular triangular mesh of 34753 cells as depicted in Fig. 7, whose length ranges from 2.08 on the left boundary to 0.15 on the right boundary, and impose transmissive boundary conditions. The solution is computed at time $t=4.7$ s setting $CFL=0.9$.

The solution of 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 flow caused by the reflection of the interior rarefaction from the centre of the dam (see [26] for an accurate description). The exact reproduction of the complicated wave pattern in the shock reflection would be challenging itself even on a fine regularly spaced grid.

However, in this test case, due to irregular grid spacing we provide an additional difficulty to numerical methods. In fact, being the test problem symmetrical along the x-axis ($x=0$), the CFL condition (2.28) is enforced where h_j^- 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.

The results for water depth D obtained using the second-order ADER-WENO method in conjunction with the FORCE and UFORCE-δ flux are presented in Fig. 8 in terms of slices along the x-axis. The numerical profiles are compared to a refined reference solu-

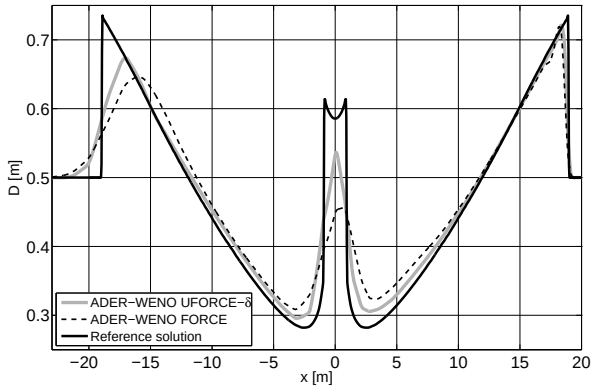

Figure 8: Collapse of a circular dam on a variably-spaced grid. Numerical results for water depth D of the second-order ADER-WENO FORCE and UFORCE-δ numerical methods are compared with the reference radial solution at time $t=4.7s$. The numerical solution profiles are sliced on the x-axis. The mesh used is depicted in Fig. 7.

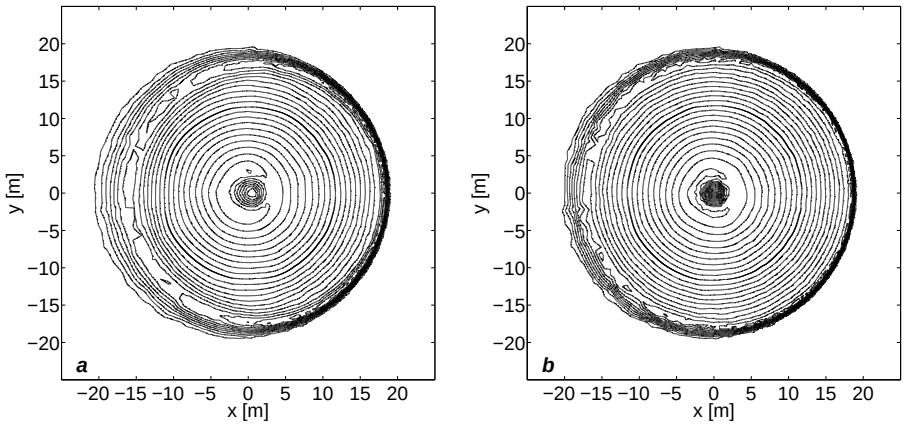

Figure 9: Collapse of a circular dam on a variably-spaced grid. Numerical results for water depth D of the second-order ADER-WENO FORCE and UFORCE-δ methods are presented in terms of contourplots at time $t=4.7s$. a) ADER-WENO FORCE method. b) ADER-WENO UFORCE-δ method.

tion obtained solving (4.7)-(4.8) as explained in Section 4.1.1. 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, while the right-facing shock ($x=18$) is solved almost to the same accuracy by both methods. This behaviour 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. 9 in terms of contourplots. Comparing Fig. 9(a) (FORCE) with Fig. 9(b) (UFORCE-δ) the same conclusion as for Fig. 8 about resolution of the left-facing shock and overall degree of symmetry can be drawn.

4.2 Applications to a three-dimensional linear systems

The aim of this section is twofold: first to demonstrate that the proposed UFORCE-δ method is identical to the Godunov upwind method in the linear case; second that the formulation developed in this paper applies to three-dimensional unstructured meshes. In order to prove the above statements we solve the three-dimensional linear system (2.7)-(2.8) where matrices (2.8) are set to

$$\mathbf{A}_x = \begin{bmatrix} 10 & 6 \\ 2 & -1 \end{bmatrix}, \qquad \mathbf{A}_y = \mathbf{A}_z = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{4.18}$$

where $\mathbf{A}_x$ has eigenvalues

$$\begin{bmatrix} \lambda_x^{(1)} \\ \lambda_x^{(2)} \end{bmatrix} = \begin{bmatrix} -2 \\ 11 \end{bmatrix}. \tag{4.19}$$

Initial condition is represented by a discontinuity located at $x = 0$ for both conserved variables q_1 and q_2, namely:

$$\begin{cases} q_1 = 1, \\ q_2 = 1, \end{cases} \quad \text{if } x < 0, \qquad \begin{cases} q_1 = -1, \\ q_2 = -1, \end{cases} \quad \text{otherwise}. \tag{4.20}$$

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. 10) and solutions are displayed at the final time $3s$. Stability is imposed enforcing the CFL condition (2.20) with $\alpha = 3$ and CFL $= 0.9$. Numerical results have been obtained applying the first-order version of the proposed numerical method. In Fig. 11 we compare the results of the UFORCE-δ method and the Godunov upwind method (2.9), (2.14) together with the exact solution. Numerical profiles in Fig. 11 have been obtained slicing the numerical solution parallel to the x-axis at

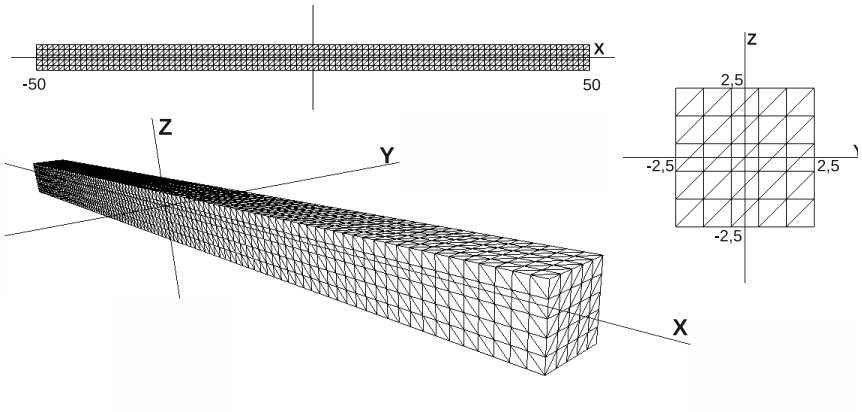

Figure 10: Applications to the linear system (2.7)-(2.8): three-dimensional unstructured tetrahedral mesh of 15000 cells.

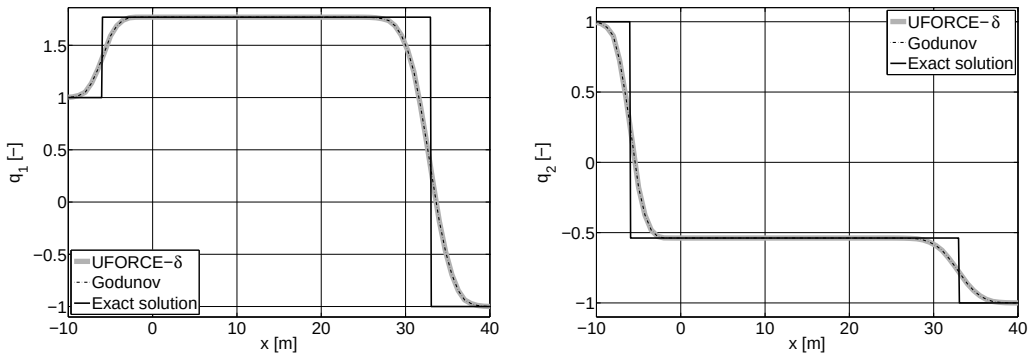

Figure 11: Applications to the linear system (2.7)-(2.8): the solution obtained using two numerical methods (Godunov upwind and UFORCE-δ) is compared to the exact solution at $t=3s$. Numerical solution profiles are sliced along the x-axis at $y=z=0$.

$y=z=0$, while the exact solution has been obtained using a one-dimensional exact Riemann solver. In this case we have verified that the results of the UFORCE-δ method and of the Godunov upwind method are identical (see Fig. 11 for a visual confirmation).

5 Conclusions

An upwind-biased version of the multi-dimensional FORCE flux on unstructured meshes for solving hyperbolic systems of PDEs in conservation-law form has been presented. The proposed first order UFORCE-δ method is genuinely centred since the use of Riemann solvers either exact or approximate is not required. To be implemented, the method requires only knowledge of the eigenvalues evaluated at current time t^n which are needed in any case for selecting the time step for time integration. We demonstrate that for the linear case UFORCE-δ is identical to the Godunov upwind method and then we extend the validity of our method to non-linear hyperbolic systems of PDEs. Second-order accuracy in space and time has been obtained in the framework of finite volume methods using an ADER-WENO approach.

Numerical performance of our method 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 three different centred methods and two upwind methods. It is shown that the UFORCE-δ method outperforms all the centred methods we tested. Moreover the accuracy of the solution for small Courant numbers and intermediate waves associated with linearly degenerate fields (contact discontinuities, shear waves and material interfaces) is improved and comparable accuracy to that of upwind methods used in conjunction with the HLL Riemann solver is achieved.

Finally, the main features of the proposed method are simplicity (due to its centred nature), accuracy compared to classical centred methods, generality of the formulation

since it applies to structured and unstructured meshes in one, two or three space dimensions.

Future developments of UFORCE-δ will concern the extension to non-conservative hyperbolic systems of PDEs in the path-conservative framework.

References

[1] P. Arminjon and A. St-Cyr, Nessyahu-Tadmor-type central finite volume methods without predictor for 3D Cartesian and unstructured tetrahedral grids, Appl. Numer. Math., 46(2):135–155, 2003.
[2] A. Canestrelli, M. Dumbser, A. Siviglia and E. F. Toro, Well-balanced high-order centered schemes on unstructured meshes for shallow water equations with fixed and mobile bed, Adv. Water Resour., 33(3):291–303, 2010.
[3] 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, Adv. Water Resour., 32(6):834–844, 2009.
[4] J. Casper, H. L. Atkins, A finite-volume high-order ENO scheme for two-dimensional hyperbolic systems, J. Comput. Phys., 106:62–76, 1993.
[5] M. Dumbser, C. Enaux and E. F. Toro, Finite volume schemes of very high order for stiff hyperbolic balance laws, J. Comput. Phys., 227(8):3971–4001, 2008.
[6] M. Dumbser, A. Hidalgo, M. Castro, C. Pares and E. F. Toro, FORCE schemes on unstructured meshes II: Non-conservative hyperbolic systems, Comput. Method. Appl. M., 199(9–12):625–647, 2010.
[7] M. Dumbser and M. Käser, Arbitrary high order non-oscillatory finite volume schemes on unstructured meshes for linear hyperbolic systems, J. Comput. Phys., 221:693–723, 2007.
[8] M. Dumbser, M. Käser, V. A. Titarev and E. F. Toro, Quadrature-free non-oscillatory finite volume schemes on unstructured meshes for nonlinear hyperbolic systems, J. Comput. Phys., 226:204–243, 2007.
[9] S. K. Godunov, Finite difference methods for the computation of discontinuous solutions of the equations of fluid dynamics, Mat. Sb., 47:271–306, 1959.
[10] A. Harten, B. Engquist, S. Osher and S. Chakravarthy, Uniformly high order essentially non-oscillatory schemes, III, J. Comput. Phys., 71:231–303, 1987.
[11] A. Harten, P. D. Lax and B. van Leer, On upstream differencing and Godunov-type schemes, SIAM Rev., 25(1):35–61, 1983.
[12] G. S. Jiang and C. W. Shu, Efficient implementation of weighted ENO schemes, J. Comput. Phys., 126:202–228, 1996.
[13] G. S. Jiang and E. Tadmor, Nonoscillatory central schemes for multidimensional hyperbolic conservation laws, SIAM J. Sci. Comput., 19(6):1892–917, 1998.
[14] A. Kurganov, S. Noelle and G. Petrova, Semidiscrete central-upwind schemes for hyperbolic conservation laws and Hamilton-Jacobi equations, SIAM J. Sci. Comput., 23(3):707–740, 2001.
[15] A. Kurganov and G. Petrova, Central schemes and contact discontinuities, ESAIM-Math. Model. Num., 34(6):1259–1275, 2000.
[16] A. Kurganov and G. Petrova, Central-upwind schemes on triangular grids for hyperbolic systems of conservation laws, Numer. Meth. Part. D. E., 21(3):536–552, 2005.

[17] A. Kurganov and E. Tadmor, New high-resolution central schemes for nonlinear conservation laws and convection-diffusion equations, J. Comput. Phys., 160(1):241–282, 2000.
[18] P. D. Lax, Weak solutions of nonlinear hyperbolic equations and their numerical computation, Comm. Pure. Appl. Math., VII:159–193, 1954.
[19] X. D. Liu, S. Osher and T. Chan, Weighted essentially non-oscillatory schemes, J. Comput. Phys., 115:200–212, 1994.
[20] C. D. Munz, On the numerical dissipation of high resolution schemes for hyperbolic conservation laws, J. Comput. Phys., 77:18-39,1998.
[21] H. Nessyahu and E. Tadmor, Non-oscillatory central differencing for hyperbolic conservation-laws, J. Comput. Phys., 87(2):408–463, 1990.
[22] M. Ricchiuto and A. Bollermann, Stabilized residual distribution for shallow water simulations, J. Comput. Phys., 228(4):1071–1115, 2009.
[23] P. L. Roe, Some contributions to the modelling of discontinuous flows, in: Proceedings of the SIAM/AMS Seminar, 1983.
[24] C. W. Shu and S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes, J. Comput. Phys., 77:439–471, 1988.
[25] G. Stecca, A. Siviglia and E. F. Toro, Upwind-biased FORCE schemes with applications to free-surface shallow flows, J. Comput. Phys., 229(18):6362–6380, 2010.
[26] E. F. Toro, Shock-Capturing Methods for Free-Surface Shallow Flows, Wiley and Sons Ltd, 2001.
[27] E. F. Toro, Riemann Solvers and Numerical Methods for Fluid Dynamics, Third Edition. Springer-Verlag, 2009.
[28] E. F. Toro and S. J. Billett, Centred TVD schemes for hyperbolic conservation laws, IMA J. Numer. Anal., 20:47–79, 2000.
[29] E. F. Toro, A. Hidalgo and M. Dumbser, FORCE schemes on unstructured meshes I: Conservative hyperbolic systems, J. Comput. Phys., 228(9):3368–3389, 2009.
[30] E. F. Toro, R. C. Millington and L. A. M. Nejad, Towards very high order Godunov schemes, in: E. F. Toro (Ed.), Godunov Methods: Theory and Applications, Kluwer/Plenum Academic Publishers, 907–940, 2001.
[31] E. F. Toro and A. Siviglia, PRICE: primitive centred schemes for hyperbolic systems, Int. J. Numer. Methods Fluids, 42(12):1263–1291, 2003.
[32] E. F. Toro and V. A. Titarev, Solution of the generalized Riemann problem for advection-reaction equations, Proc. R. Soc. A-Math. Phys. Eng. Sci., 458:271–281, 2002.
[33] B. van Leer, Towards the ultimate conservative difference scheme II: Monotonicity and conservation combined in a second order scheme, J. Comput. Phys., 14:361–370, 1974.