Journal of Computational Physics 229 (2010) 6362–6380

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Upwind-biased FORCE schemes with applications to free-surface shallow flows

Guglielmo Stecca *, Annunziato Siviglia, Eleuterio F. Toro

Department of Civil and Environmental Engineering, University of Trento, Via Mesiano 77, I-38100 Trento, Italy

ARTICLE INFO

Article history:
Received 14 December 2009
Received in revised form 16 April 2010
Accepted 3 May 2010
Available online 19 May 2010

Keywords:
Conservative hyperbolic systems
Centred schemes
Structured meshes
Numerical fluxes
Shallow water equations
FORCE
Upwind-biased

ABSTRACT

In this paper we develop numerical fluxes of the centred type for one-step schemes in conservative form for solving general systems of conservation laws in multiple-space dimensions on structured meshes. The proposed method is an extension of the multidimensional FORCE flux developed by Toro et al. (2009) [14]. Here we introduce upwind bias by modifying the shape of the staggered mesh of the original FORCE method. The upwind bias is evaluated using an estimate of the largest eigenvalue, which in any case is needed for selecting a time step. The resulting basic flux is first-order accurate and monotone. For the linear advection equation, the proposed UFORCE method reproduces exactly the upwind Godunov method. Extension to non-linear systems has been done empirically via the two-dimensional inviscid shallow water equations. Second order of accuracy in space and time on structured meshes is obtained in the framework of finite volume methods. The proposed method improves the accuracy of the solution for small Courant numbers and intermediate waves associated with linearly degenerate fields (contact discontinuities, shear waves and material interfaces). It achieves comparable accuracy to that of upwind methods with approximate Riemann solvers, though retaining the simplicity and efficiency of centred methods. The performance of the schemes is assessed on a suite of test problems for the two-dimensional shallow water equations.

1. Introduction

There are essentially two approaches for designing non-oscillatory numerical fluxes for computing approximate solutions to systems of hyperbolic equations. The first approach is the *upwind approach*, represented by Godunov's method [3] while the second is the *centred approach*, typically represented by the Lax–Friedrichs flux and variations of it [10]. The upwind approach is more accurate than the centred approach, if used with suitable Riemann solvers, the disadvantage being its complexity and computational expense. The centred approach is more general and simpler to apply to complicated set of equations. The difference in accuracy between the upwind and the centred approach is evident in two special cases, namely, for calculations with small Courant numbers and for computing intermediate waves associated with linearly degenerate fields (contact discontinuities, shear waves, material interfaces and vortical flows). For a comprehensive presentation of upwind, and also some centred methods, see for example [13] and references therein.

This paper is about a centred scheme with upwind bias. The scheme 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 [9], using an adaptive staggered mesh. Their scheme is based on a modification of the centred scheme of Nessyahu and Tadmor [11], where the staggered mesh is fixed. The scheme of Nessyahu and Tadmor [11] has been extended

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

0021-9991/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2010.05.001

to multi-dimensions by Jiang and Tadmor [4] and by Arminjon and his collaborator [1]. Multidimensional extensions of the scheme of Kurganov and Tadmor have been presented by [6] (Cartesian version) and [8] (unstructured version), while a modified version of the scheme optimised for treating contact discontinuities, which makes use of the partial characteristic decomposition, has been presented by Kurganov and Petrova [7].

Our scheme is a modification of the FORCE centred method, first put forward for one-dimensional systems in [17]. See also [2], where convergence of the scheme is established for two pairs of hyperbolic systems. A multidimensional version of FORCE for unstructured meshes in two and three space dimensions has recently been proposed in [14]. 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 present scheme puts forward the idea of replacing the barycentres of each cell in the primary mesh by a moving point, whose position is determined so as to reduce numerical dissipation. For the linear advection equation in two and three space dimensions the approach reproduces identically the Godunov scheme constructed by solving Riemann problems normal to each interface. In fact, as proved here, this is the most accurate monotone scheme that can be constructed in the given stencil. For non-linear systems the extension is empirical and makes explicit use of estimates for the eigenvalues of the relevant system.

The method is fully analysed for the linear advection equation in multiple-space dimensions. Then the method is implemented for solving the two-dimensional non-linear shallow water equations on Cartesian meshes, in first-order mode and in second-order mode using a simple MUSCL-Hancock type extension. The schemes are thoroughly assessed on a number of well-established test problems for the non-linear shallow water equations, some of them with exact solutions. Significant improvements in accuracy are observed with respect to the conventional FORCE scheme without upwind information. As already mentioned, there are two critical situations in which classical centred schemes are inaccurate: small Courant numbers and intermediate waves. Our numerical experiments show very significant improvements in both classes of problems. The present approach offers a practical tool for solving problems associated with real applications in geophysical flows.

The rest of this paper proceeds as follows. In Section 2 we set the background by reviewing the FORCE method in multiple-space dimensions. In Section 3 we present our UFORCE first-order accurate numerical method on structured meshes. The optimal choice for the upwind bias is discussed in Section 3.2 for the linear case and in Section 3.3 for non-linear systems. In Section 4 we extend the UFORCE approach and construct numerical fluxes to second order using the MUSCL-Hancock approach. In Section 5 we assess the performance of the numerical scheme for shocked and smoothed flows via a suite of test problems for the shallow water equations. Conclusions are drawn in Section 6.

2. Background

In this section we review the multidimensional FORCE flux on Cartesian meshes in the framework of finite volume methods.

We consider a two-dimensional system of m non-linear hyperbolic equations written in differential conservation-law form:

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_y \mathbf{G}(\mathbf{Q}) = 0, \tag{1}$$

where $\mathbf{Q}$ is the vector of conserved variables, $\mathbf{F}(\mathbf{Q})$, and $\mathbf{G}(\mathbf{Q})$ are the flux vectors in the x and y directions respectively.

Finite volume schemes for (1) have the form:

$$\mathbf{Q}_{i,j}^{n+1} = \mathbf{Q}_{i,j}^{n} - \frac{\Delta t}{\Delta x}\left[\mathbf{F}_{i+\frac{1}{2}j} - \mathbf{F}_{i-\frac{1}{2}j}\right] - \frac{\Delta t}{\Delta y}\left[\mathbf{G}_{ij+\frac{1}{2}} - \mathbf{G}_{ij-\frac{1}{2}}\right]. \tag{2}$$

The construction of the FORCE flux in multi-dimensions has been proposed by Toro et al. [14]. It requires the adoption of two different meshes: the *primary mesh*, where the numerical scheme provides cell averages, and a *staggered mesh*, used to define numerical fluxes for the conservative form of the scheme.

The *primary mesh* is the mesh chosen for the discretization of the domain of interest, while the *staggered mesh* is obtained using edge-based meshes. Fig. 1 depicts the situation for the two-dimensional case. The chosen Cartesian mesh determines the computing cells denoted by $C_{i,j}$. Each cell $C_{i,j}$ possesses four edges (or faces or inter-cell boundaries). We adopt the conventional anticlockwise orientation so that the cell $C_{i,j}$ lies on the left of edge k while the neighbouring cell that shares edge k with cell $C_{i,j}$ lies on the right of edge k. Each edge k is associated to one element of the staggered mesh, which is generated by joining the barycentre of $C_{i,j}$ and the barycentre of its neighbour across edge k, with the vertices of edge k. This results in a quadrilateral straddling face k, as depicted in Fig. 1. The four edge-based secondary control volumes and the corresponding edges of cell $C_{i,j}$ are defined as follows:

$$\left.\begin{aligned} S_{i+\frac{1}{2}j} &= S^{-}_{i+\frac{1}{2}j} \bigcup S^{+}_{i+\frac{1}{2}j}, \\ S_{ij+\frac{1}{2}} &= S^{-}_{ij+\frac{1}{2}} \bigcup S^{+}_{ij+\frac{1}{2}}, \\ S_{i-\frac{1}{2}j} &= S^{-}_{i-\frac{1}{2}j} \bigcup S^{+}_{i-\frac{1}{2}j}, \\ S_{ij-\frac{1}{2}} &= S^{-}_{ij-\frac{1}{2}} \bigcup S^{+}_{ij-\frac{1}{2}}. \end{aligned}\right\} \tag{3}$$

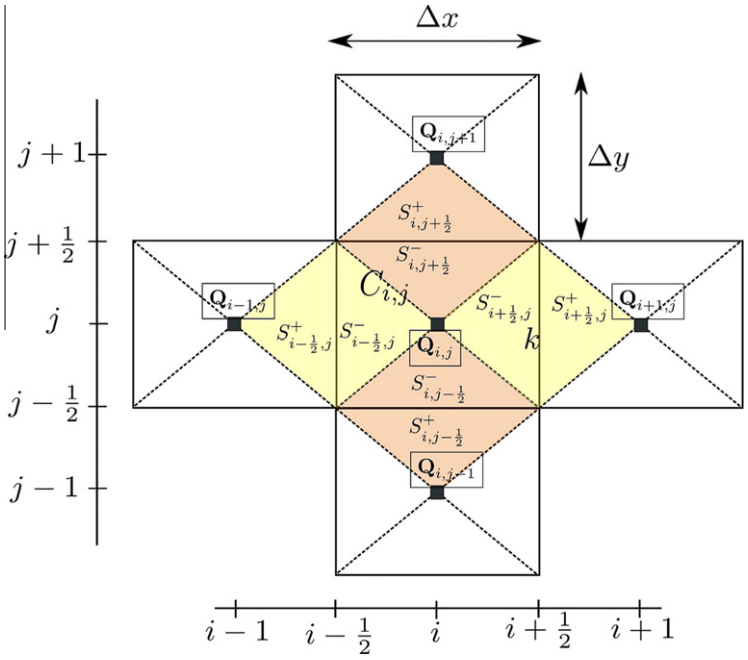

Fig. 1. Notation for a general configuration on a Cartesian mesh for the FORCE scheme.

In (3), $S^-_{i\pm\frac{1}{2}j}$, $S^-_{ij\pm\frac{1}{2}}$ denote the portion of the corresponding control volumes of the staggered mesh $S_{i\pm\frac{1}{2}j}$, $S_{ij\pm\frac{1}{2}}$ that lay inside C_{ij} and $S^+_{i\pm\frac{1}{2}j}$, $S^+_{ij\pm\frac{1}{2}}$ denote the portion of the corresponding control volumes that lay outside (see Fig. 1).

In order to derive the FORCE scheme, we recall the integral form of the conservation laws (1), applied for a time interval δt for a generic control volume S_k, assumed to be a polygon of n_k edges:

$$\mathbf{Q}_k^1 = \mathbf{Q}_k^0 - \frac{\delta t}{|S_k|}\sum_{l=1}^{n_k}\int_{v_l}^{v_{l+1}}[\cos(\theta_l)\mathbf{F}(\mathbf{Q}) + \sin(\theta_l)\mathbf{G}(\mathbf{Q})]ds. \tag{4}$$

Here $\mathbf{Q}_k^0$ is the initial condition at time $t=0$, $\mathbf{Q}_k^1$ is the integral average at time $t=\delta t$, θ_l is the angle between the x-axis and the normal vector to face l joining vertices v_l and v_{l+1}, with $v_{n_k+1}\equiv v_1$, $|S_k|$ is the area of control volume S_k, $\mathbf{n}_k=(\cos(\theta_l),\sin(\theta_l))$ is the outward unit normal vector to edge l of S_k.

Keeping in mind (4), the construction of the FORCE scheme is achieved in three stages:

(i) Averages on staggered mesh.
In the first stage we assume initial condition $\mathbf{Q}^n_{i,j}$ for the conservation laws at time t^n in cells $C_{i,j}$. Note that initial conditions in cell $C_{i,j}$ and all its immediate neighbours determine initial conditions for each of the control volumes (3) of the staggered mesh associated with the boundaries of $C_{i,j}$. Then we evolve the solution for a time interval $\frac{1}{2}\Delta t$ in each control volume in (3). The solution is obtained by applying the integral form of the conservation laws (4) with $\delta t=\frac{1}{2}\Delta t$. Application of (4) to each of the secondary control volumes in (3) yields:

$$\left.\begin{aligned}
\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2}j} &= \tfrac{1}{2}\left(\mathbf{Q}^n_{ij}+\mathbf{Q}^n_{i+1j}\right) - \tfrac{\Delta t}{\Delta x}\left[\mathbf{F}(\mathbf{Q}^n_{i+1j})-\mathbf{F}(\mathbf{Q}^n_{ij})\right],\\
\mathbf{Q}^{n+\frac{1}{2}}_{ij+\frac{1}{2}} &= \tfrac{1}{2}\left(\mathbf{Q}^n_{ij}+\mathbf{Q}^n_{ij+1}\right) - \tfrac{\Delta t}{\Delta y}\left[\mathbf{G}(\mathbf{Q}^n_{ij+1})-\mathbf{G}(\mathbf{Q}^n_{ij})\right],\\
\mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2}j} &= \tfrac{1}{2}\left(\mathbf{Q}^n_{i-1j}+\mathbf{Q}^n_{ij}\right) - \tfrac{\Delta t}{\Delta x}\left[\mathbf{F}(\mathbf{Q}^n_{ij})-\mathbf{F}(\mathbf{Q}^n_{i-1j})\right],\\
\mathbf{Q}^{n+\frac{1}{2}}_{ij-\frac{1}{2}} &= \tfrac{1}{2}\left(\mathbf{Q}^n_{ij-1}+\mathbf{Q}^n_{ij}\right) - \tfrac{\Delta t}{\Delta y}\left[\mathbf{G}(\mathbf{Q}^n_{ij})-\mathbf{G}(\mathbf{Q}^n_{ij-1})\right].
\end{aligned}\right\} \tag{5}$$

(ii) Averages on primary mesh.
In the second stage of the construction of the schemes, we evolve the solution by another time step $\frac{1}{2}\Delta t$ within each cell $C_{i,j}$, having assumed (5) at time $t=t^n+\frac{1}{2}\Delta t$ as initial condition. By applying again the integral form of the conservation laws (4), we obtain:

$$\mathbf{Q}^{n+1}_{ij} = \frac{1}{4}\left[\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2}j}+\mathbf{Q}^{n+\frac{1}{2}}_{ij+\frac{1}{2}}+\mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2}j}+\mathbf{Q}^{n+\frac{1}{2}}_{ij-\frac{1}{2}}\right] - \frac{1}{2}\frac{\Delta t}{\Delta x}\left[\mathbf{F}(\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2}j})-\mathbf{F}\left(\mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2}j}\right)\right] - \frac{1}{2}\frac{\Delta t}{\Delta y}\left[\mathbf{G}(\mathbf{Q}^{n+\frac{1}{2}}_{ij+\frac{1}{2}})-\mathbf{G}\left(\mathbf{Q}^{n+\frac{1}{2}}_{ij-\frac{1}{2}}\right)\right]. \tag{6}$$

(iii) Conservative form and the FORCE flux.
In the final stage of the derivation of the scheme we perform algebraic manipulations on (6) so as to reproduce the conservative formula (2). The resulting FORCE numerical flux is:

$$\mathbf{F}_{i+\frac{1}{2}j} = \frac{1}{2}\left[\mathbf{F}\left(\mathbf{Q}_{i+\frac{1}{2}j}^{n+\frac{1}{2}}\right) + \frac{1}{2}\left(\mathbf{F}(\mathbf{Q}_{ij}^{n}) + \mathbf{F}(\mathbf{Q}_{i+1j}^{n})\right) - \frac{1}{4}\frac{\Delta x}{\Delta t}\left(\mathbf{Q}_{i+1j}^{n} - \mathbf{Q}_{ij}^{n}\right)\right], \tag{7}$$

with analogous formulae for the other fluxes. It is possible to rewrite the inter-cell flux (7) as the arithmetic average of two fluxes, namely

$$\mathbf{F}_{i+\frac{1}{2}j} = \frac{1}{2}\left(\mathbf{F}_{i+\frac{1}{2}j}^{LW2} + \mathbf{F}_{i+\frac{1}{2}j}^{LF2}\right). \tag{8}$$

Here

$$\mathbf{F}_{i+\frac{1}{2}j}^{LW2} = \mathbf{F}(\mathbf{Q}_{i+\frac{1}{2}j}^{LW2}) \tag{9}$$

with

$$\mathbf{Q}_{i+\frac{1}{2}j}^{LW2} = \frac{1}{2}(\mathbf{Q}_{ij}^{n} + \mathbf{Q}_{i+1j}^{n}) - \frac{1}{2}\left(\frac{2\Delta t}{\Delta x}\right)[\mathbf{F}(\mathbf{Q}_{i+1j}^{n}) - \mathbf{F}(\mathbf{Q}_{ij}^{n})]. \tag{10}$$

The second flux term in (8) is

$$\mathbf{F}_{i+\frac{1}{2}j}^{LF2} = \frac{1}{2}[\mathbf{F}(\mathbf{Q}_{i+1j}^{n}) + \mathbf{F}(\mathbf{Q}_{ij}^{n})] - \frac{1}{2}\left(\frac{\Delta x}{2\Delta t}\right)(\mathbf{Q}_{i+1j}^{n} - \mathbf{Q}_{ij}^{n}). \tag{11}$$

Fluxes (9) and (11) may be regarded as generalisations of the Lax–Wendroff (LW) and Lax–Friedrichs (LF) flux, respectively.

Generalising the FORCE method in α dimensions [14], where α denotes the number of spatial dimensions, we can write:

$$\mathbf{F}_{i+\frac{1}{2}}^{FORCE\alpha} = \frac{1}{2}(\mathbf{F}_{i+\frac{1}{2}}^{LW\alpha} + \mathbf{F}_{i+\frac{1}{2}}^{LF\alpha}), \tag{12}$$

with

$$\mathbf{F}_{i+\frac{1}{2}}^{LW\alpha} = \mathbf{F}\left(\mathbf{Q}_{i+\frac{1}{2}}^{LW\alpha}\right) \tag{13}$$

and

$$\mathbf{Q}_{i+\frac{1}{2}}^{LW\alpha} = \frac{1}{2}(\mathbf{Q}_{i}^{n} + \mathbf{Q}_{i+1}^{n}) - \frac{1}{2}\left(\frac{\alpha\Delta t}{\Delta x}\right)[\mathbf{F}(\mathbf{Q}_{i+1}^{n}) - \mathbf{F}(\mathbf{Q}_{i}^{n})]. \tag{14}$$

The Lax–Friedrichs type flux is

$$\mathbf{F}_{i+\frac{1}{2}}^{LF\alpha} = \frac{1}{2}[\mathbf{F}(\mathbf{Q}_{i+1}^{n}) + \mathbf{F}(\mathbf{Q}_{i}^{n})] - \frac{1}{2}\left(\frac{\Delta x}{\alpha\Delta t}\right)(\mathbf{Q}_{i+1}^{n} - \mathbf{Q}_{i}^{n}). \tag{15}$$

A linear analysis shows that the FORCE scheme is monotone under the following conditions (see [14]):

$$\begin{aligned} &c_x^2 + c_y^2 \leqslant \frac{1}{2} \quad \text{in two space dimensions } \alpha = 2, \\ &c_x^2 + c_y^2 + c_z^2 \leqslant \frac{1}{3} \quad \text{in three space dimensions } \alpha = 3, \end{aligned} \tag{16}$$

where

$$c_x = \lambda_x\frac{\Delta t}{\Delta x}, \quad c_y = \lambda_y\frac{\Delta t}{\Delta y}, \quad c_z = \lambda_z\frac{\Delta t}{\Delta z} \tag{17}$$

represent the Courant numbers in x, y and z directions and λ_x, λ_y, λ_z are the corresponding characteristic speeds.

3. Upwind-biased FORCE scheme in multiple-space dimensions

In this section we construct an upwind-biased variation of the FORCE method in multi-dimensions for a system of hyperbolic equations. This is achieved by modifying the shape of the staggered mesh in the FORCE scheme [14], accordingly to *upwind* information, which only requires the knowledge of the eigenvalues of the system.

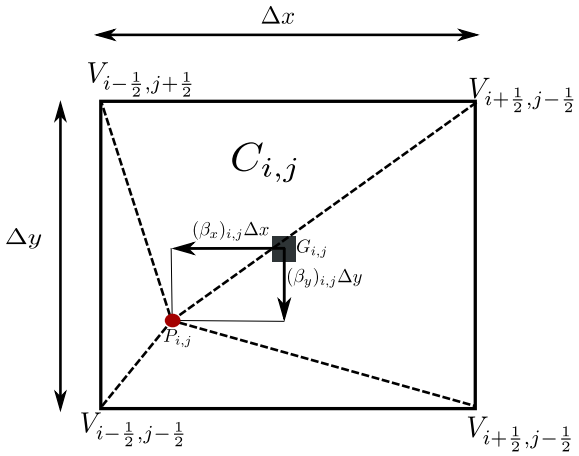

Fig. 2. Notation for a general configuration on an Cartesian mesh.

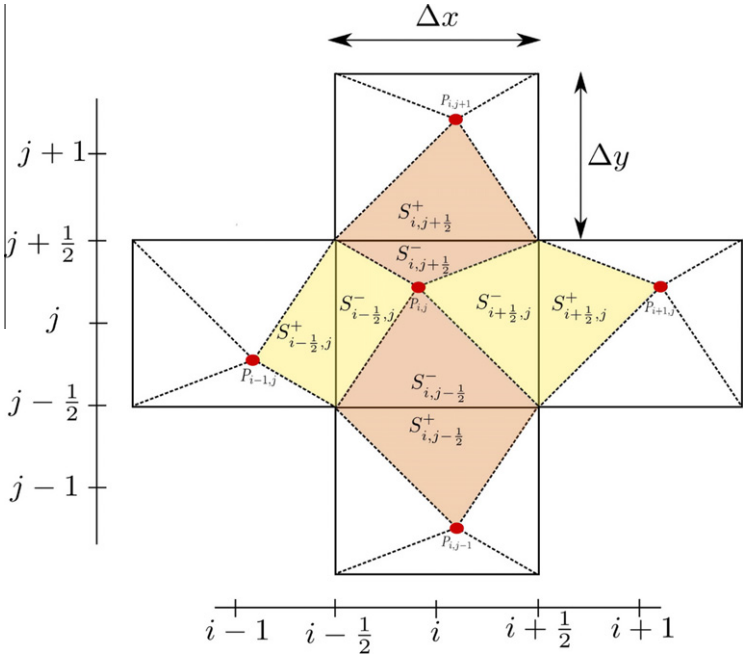

Fig. 3. Notation for a general configuration on an Cartesian mesh for the UFORCE scheme.

3.1. Derivation of the numerical method

We first consider the two-dimensional case (2). As for the classical FORCE scheme [14], the upwind-biased version of the FORCE scheme (UFORCE) requires the adoption of a primary mesh for computing cell averages, and a staggered mesh used to define numerical fluxes for the conservative form of the scheme. The primary mesh, which determines the computing cells $C_{i,j}$, is Cartesian in this paper, while the staggered mesh is obtained joining the four vertices ($V_{i-\frac{1}{2}j-\frac{1}{2}}$, $V_{i+\frac{1}{2}j-\frac{1}{2}}$, $V_{i+\frac{1}{2}j+\frac{1}{2}}$, $V_{i-\frac{1}{2}j+\frac{1}{2}}$) of cell $C_{i,j}$ with a generic point $P_{i,j}$ lying inside the cell as shown in Fig. 2. This results in a non-Cartesian convex quadrilateral straddling face k, as depicted in Fig. 3. In order to construct the control volumes of the staggered grid, for each cell $C_{i,j}$ we consider its barycentre $G_{i,j}$ having local coordinates $G_{i,j} \equiv (0,0)$. The bias is measured as the distance between $P_{i,j} \equiv (-(\beta_x)_{i,j}\Delta x, -(\beta_y)_{i,j}\Delta y)$ and $G_{i,j}$, where the directional bias $(\beta_x)_{i,j}$ and $(\beta_y)_{i,j}$ along the x and y directions, are taken positive in the $i-1$ and $j-1$ directions. The situation is described in Fig. 2, where positive values for the bias are depicted. The range of variation for the upwind bias is $-\frac{1}{2} \leqslant (\beta_x)_{i,j},\ (\beta_y)_{i,j} \leqslant \frac{1}{2}$, where the case $((\beta_x)_{i,j}, (\beta_y)_{i,j}) \equiv 0$ reproduces the standard FORCE scheme.

By defining a triangle of vertices P_1, P_2, P_3 as $\mathcal{T}\{P_1, P_2, P_3\}$, we can identify the triangles of the secondary mesh as:

$$\left.\begin{aligned} S^-_{i\pm\frac{1}{2}j} &= \mathcal{T}\left(V_{i\pm\frac{1}{2}j-\frac{1}{2}}, V_{i\pm\frac{1}{2}j+\frac{1}{2}}, P_{ij}\right), \\ S^+_{i+\frac{1}{2}j} &= \mathcal{T}\left(V_{i+\frac{1}{2}j-\frac{1}{2}}, V_{i+\frac{1}{2}j+\frac{1}{2}}, P_{i+1j}\right), \\ S^+_{i-\frac{1}{2}j} &= \mathcal{T}\left(V_{i-\frac{1}{2}j+\frac{1}{2}}, V_{i-\frac{1}{2}j-\frac{1}{2}}, P_{i-1j}\right) \end{aligned}\right\} \tag{18}$$

while analogous expressions are obtained for $S^-_{ij\pm\frac{1}{2}}$, $S^+_{ij-\frac{1}{2}}$ and $S^+_{ij+\frac{1}{2}}$ (see Fig. 3).

Then, the areas associated with the secondary mesh are given by:

$$\left.\begin{aligned} &\left\{|S^-_{i+\frac{1}{2}j}| = \tfrac{\Delta y\Delta x}{2}\left[\tfrac{1}{2} + (\beta_x)_{ij}\right]\right\} + \left\{|S^+_{i+\frac{1}{2}j}| = \tfrac{\Delta y\Delta x}{2}\left[\tfrac{1}{2} - (\beta_x)_{i+1j}\right]\right\}, \\ &\left\{|S^-_{i-\frac{1}{2}j}| = \tfrac{\Delta y\Delta x}{2}\left[\tfrac{1}{2} - (\beta_x)_{ij}\right]\right\} + \left\{|S^+_{i-\frac{1}{2}j}| = \tfrac{\Delta y\Delta x}{2}\left[\tfrac{1}{2} + (\beta_x)_{i-1j}\right]\right\}. \end{aligned}\right\} \tag{19}$$

Analogous expressions can be written for $|S^-_{ij+\frac{1}{2}}|$, $|S^+_{ij+\frac{1}{2}}|$, $|S^-_{ij-\frac{1}{2}}|$, $|S^+_{ij-\frac{1}{2}}|$. Then, the UFORCE scheme is achieved in three stages:

(i) Averages on staggered mesh. Assuming initial condition $\mathbf{Q}^n_{ij}$ in cells C_{ij} at time t^n, cell averages are evolved for a time step $\frac{1}{2}\Delta t$ in each control volume in (19) of the staggered mesh associated with the edge of C_{ij}. The solution is obtained by applying the integral form of the conservation laws (4) with $\delta t = \frac{1}{2}\Delta t$. Application of (4) to each of the secondary control volumes in (19) yields the following expression for intermediate states:

$$\left.\begin{aligned} \mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2}j} &= \frac{1}{\left|S^-_{i+\frac{1}{2}j}\right| + \left|S^+_{i+\frac{1}{2}j}\right|}\left\{\left|S^-_{i+\frac{1}{2}j}\right|\mathbf{Q}^n_{ij} + \left|S^+_{i+\frac{1}{2}j}\right|\mathbf{Q}^n_{i+1j} - \tfrac{\Delta t\Delta y}{2}\left[\mathbf{F}(\mathbf{Q}^n_{i+1j}) - \mathbf{F}(\mathbf{Q}^n_{ij})\right]\right\} \\ \mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2}j} &= \frac{1}{\left|S^-_{i-\frac{1}{2}j}\right| + \left|S^+_{i-\frac{1}{2}j}\right|}\left\{\left|S^+_{i-\frac{1}{2}j}\right|\mathbf{Q}^n_{i-1j} + \left|S^-_{i-\frac{1}{2}j}\right|\mathbf{Q}^n_{ij} - \tfrac{\Delta t\Delta y}{2}\left[\mathbf{F}(\mathbf{Q}^n_{ij}) - \mathbf{F}(\mathbf{Q}^n_{i-1j})\right]\right\} \end{aligned}\right\} \tag{20}$$

while the expressions for $\mathbf{Q}^{n+\frac{1}{2}}_{ij+\frac{1}{2}}, \mathbf{Q}^{n+\frac{1}{2}}_{ij-\frac{1}{2}}$ are obtained from (20) by exchanging indexes i and j and replacing $\mathbf{F}$ with $\mathbf{G}$.

(ii) Averages on primary mesh. In the second stage of the construction of the scheme, we evolve the solution by another time step $\frac{1}{2}\Delta t$ within each cell C_{ij}, having assumed (20) at time $t = t^n + \frac{1}{2}\Delta t$ as initial condition. By applying again the integral form of the conservation laws (4), we obtain:

$$\begin{aligned} \mathbf{Q}^{n+1}_{ij} = {} & \frac{1}{\Delta x\Delta y}\left[\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2}j}\left|S^-_{i+\frac{1}{2}j}\right| + \mathbf{Q}^{n+\frac{1}{2}}_{ij+\frac{1}{2}}\left|S^-_{ij+\frac{1}{2}}\right| + \mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2}j}\left|S^-_{i-\frac{1}{2}j}\right| + \mathbf{Q}^{n+\frac{1}{2}}_{ij-\frac{1}{2}}\left|S^-_{ij-\frac{1}{2}}\right|\right] - \frac{1}{2}\frac{\Delta t}{\Delta x}\left[\mathbf{F}(\mathbf{Q}^{n+\frac{1}{2}}_{i+\frac{1}{2}j}) - \mathbf{F}(\mathbf{Q}^{n+\frac{1}{2}}_{i-\frac{1}{2}j})\right] \\ & - \frac{1}{2}\frac{\Delta t}{\Delta y}\left[\mathbf{G}\left(\mathbf{Q}^{n+\frac{1}{2}}_{ij+\frac{1}{2}}\right) - \mathbf{G}\left(\mathbf{Q}^{n+\frac{1}{2}}_{ij-\frac{1}{2}}\right)\right]. \end{aligned} \tag{21}$$

(iii) Conservative form and the upwind-biased FORCE flux. In this stage we perform algebraic manipulations on (21) so as to reproduce the conservative formula (2). As for the standard FORCE method the intercell flux can be rewritten as the arithmetic average of two fluxes, namely

$$\mathbf{F}^{UFORCE2}_{i+\frac{1}{2}j} = \frac{1}{2}\left(\mathbf{F}^{uLW2}_{i+\frac{1}{2}j} + \mathbf{F}^{uLF2}_{i+\frac{1}{2}j}\right). \tag{22}$$

Here $\mathbf{F}^{uLW2}_{i+\frac{1}{2}j}$ and $\mathbf{F}^{uLF2}_{i+\frac{1}{2}j}$ represent upwind-biased versions of the Lax-Wendroff (9), (10) and Lax–Friedrichs (11) fluxes, which read:

$$\mathbf{F}^{uLW2}_{i+\frac{1}{2}j} = \mathbf{F}\left(\mathbf{Q}^{uLW2}_{i+\frac{1}{2}j}\right), \tag{23}$$

$$\mathbf{Q}^{uLW2}_{i+\frac{1}{2}j} = \frac{1}{\left|S^+_{i+\frac{1}{2}j}\right| + \left|S^-_{i+\frac{1}{2}j}\right|}\left\{\left[\left|S^-_{i+\frac{1}{2}j}\right|\mathbf{Q}^n_{ij} + \left|S^+_{i+\frac{1}{2}j}\right|\mathbf{Q}^n_{i+1j}\right] - \frac{1}{2}\Delta t\Delta y\left[\mathbf{F}\left(\mathbf{Q}^n_{i+1j}\right) - \mathbf{F}\left(\mathbf{Q}^n_{ij}\right)\right]\right\}, \tag{24}$$

$$\mathbf{F}^{uLF2}_{i+\frac{1}{2}j} = \frac{1}{\left|S^+_{i+\frac{1}{2}j}\right| + \left|S^-_{i+\frac{1}{2}j}\right|}\left\{\left[\left|S^-_{i+\frac{1}{2}j}\right|\mathbf{F}(\mathbf{Q}^n_{i+1j}) + \left|S^+_{i+\frac{1}{2}j}\right|\mathbf{F}(\mathbf{Q}^n_{ij})\right] - 2\left(\frac{1}{\Delta t\Delta y}\right)\left|S^-_{i+\frac{1}{2}j}\right|\left|S^+_{i+\frac{1}{2}j}\right|\left(\mathbf{Q}^n_{i+1j} - \mathbf{Q}^n_{ij}\right)\right\}. \tag{25}$$

Finally, the generalisation of the flux in the x direction of the UFORCE method on a arbitrary number α of dimensions can be written as function of the upwind bias in the following way:

$$\mathbf{F}^{UFORCE\alpha}_{i+\frac{1}{2}} = \frac{1}{2}(\mathbf{F}^{uLW\alpha}_{i+\frac{1}{2}} + \mathbf{F}^{uLF\alpha}_{i+\frac{1}{2}}). \tag{26}$$

Here

$$\mathbf{F}^{uLW\alpha}_{i+\frac{1}{2}} = \mathbf{F}\left(\mathbf{Q}^{uLW\alpha}_{i+\frac{1}{2}}\right) \tag{27}$$

with

$$\mathbf{Q}_{i+\frac{1}{2}}^{uLW\alpha} = \frac{1}{2\left[1-(\beta_x)_{i+1}+(\beta_x)_i\right]}\left\{(1+2(\beta_x)_i)\mathbf{Q}_i^n + (1-2(\beta_x)_{i+1})\mathbf{Q}_{i+1}^n - \left(\frac{\alpha\Delta t}{\Delta x}\right)[\mathbf{F}(\mathbf{Q}_{i+1}^n) - \mathbf{F}(\mathbf{Q}_i^n)]\right\}. \tag{28}$$

The second flux term in (26) is

$$\mathbf{F}_{i+\frac{1}{2}}^{uLF\alpha} = \frac{1}{2\left[1-(\beta_x)_{i+1}+(\beta_x)_i\right]}\left\{(1+2(\beta_x)_i)\mathbf{F}(\mathbf{Q}_{i+1}^n) + (1-2(\beta_x)_{i+1})\mathbf{F}(\mathbf{Q}_i^n) - \left(\frac{\Delta x}{\alpha\Delta t}\right)(1+2(\beta_x)_i)(1-2(\beta_x)_{i+1})(\mathbf{Q}_{i+1,j}^n - \mathbf{Q}_{i,j}^n)\right\}. \tag{29}$$

In (26)–(29) we make an exception to the two-dimensional Cartesian notation used in the rest of this paper. Here x indicates the current direction and the subscript i is the index along the x direction.

The method derived in this section depends directly on the upwind bias inside the computational cell.

Even though the derivation of the UFORCE method is rather complex, its implementation is quite simple. In fact, in two-space dimension it requires to update the solution at time level $n+1$ using formula (2) with fluxes $\mathbf{F}$ given by (22) and fluxes $\mathbf{G}$ obtained in analogous manner.

In the next section we shall derive optimal values for the upwind bias.

3.2. Optimal choice for the upwind bias: the linear case

In this section we analyse how to determine the optimal bias for the UFORCE scheme taking into account information from the eigenvalues of the system.

The study is based on the two-dimensional linear advection equation with constant coefficients, in two space dimensions:

$$\partial_t q + \partial_x f(q) + \partial_y g(q) = 0, \tag{30}$$

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

3.2.1. Accuracy of the two-dimensional Godunov upwind method

Numerical schemes for (30) are written as:

$$q_{ij}^{n+1} = q_{ij}^n - \frac{\Delta t}{\Delta x}\left(f_{i+\frac{1}{2}j} - f_{i-\frac{1}{2}j}\right) - \frac{\Delta t}{\Delta y}\left(g_{ij+\frac{1}{2}} - g_{ij-\frac{1}{2}}\right). \tag{31}$$

We restrict our analysis to five-point schemes, in which the stencil consists of the central point (i,j) and the 4 neighbours $(i,j-1)$, $(i+1,j)$, $(i,j+1)$, $(i-1,j)$. This implies that the scheme (31) can be rewritten as:

$$q_{ij}^{n+1} = \sum_{l=-1}^{1}\sum_{m=-1}^{1} b_{l,m} q_{i+l,j+m}^n \quad (b_{\pm1,\pm1} = 0). \tag{32}$$

To this class belong all the numerical methods discussed in this paper.

Consider the Godunov upwind method, whose flux in the one-dimensional case is given by:

$$\mathbf{F}_{i+\frac{1}{2}}^{God} = \mathbf{F}\left(\mathbf{Q}_{i+\frac{1}{2}}(0)\right), \tag{33}$$

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

$$\left.\begin{aligned} &\text{PDEs}: \partial_t\mathbf{Q} + \partial_x\mathbf{F}(\mathbf{Q}) = \mathbf{0},\\ &\text{IC}: \mathbf{Q}(x,0) = \begin{cases}\mathbf{Q}_i^n & \text{if } x<0,\\ \mathbf{Q}_{i+1}^n & \text{if } x>0,\end{cases}\end{aligned}\right\} \tag{34}$$

Here we consider a two-dimensional extension of the Godunov method in which fluxes are calculated by solving one-dimensional Riemann problems (34) orthogonally to the current edge.

Remark. The coefficients in (32) of the Godunov upwind method in two space dimensions read:

$$b_{-1,0} = \frac{1}{2}(c_x + |c_x|), \quad b_{1,0} = \frac{1}{2}(|c_x| - c_x), \quad b_{0,-1} = \frac{1}{2}(c_y + |c_y|), \quad b_{0,1} = \frac{1}{2}(|c_y| - c_y), \quad b_{0,0} = 1 - |c_x| - |c_y|, \tag{35}$$

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 (32), (35) proves to be stable if the following condition holds:

$$|c_x| + |c_y| \leqslant 1. \tag{36}$$

Proposition 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 (32).*

Proof. We consider the generic five-points scheme (32). Coefficients $b_{l,m}$, $-1 \leqslant l, m \leqslant 1$ are constant. Assume the scheme to be at least first-order, that is consistent. Then from Roe's accuracy lemma (see [13]) we have the following three equations relating the five coefficients:

$$\begin{cases} \sum_{l=-1}^{1} \sum_{m=-1}^{1} b_{l,m} = 1 \quad (b_{\pm 1,\pm 1} = 0), \\ b_{-1,0} - b_{1,0} = c_x, \\ b_{0,-1} - b_{0,1} = c_y. \end{cases} \tag{37}$$

System (37) gives a two-parameter family of solutions. We set:

$$\begin{aligned} b_{-1,0} + b_{1,0} &= \hat{q}_x, \\ b_{0,-1} + b_{0,1} &= \hat{q}_y \end{aligned} \tag{38}$$

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

$$b_{-1,0} = \frac{1}{2}(\hat{q}_x + c_x) \quad b_{0,-1} = \frac{1}{2}(\hat{q}_y + c_y), \quad b_{1,0} = \frac{1}{2}(\hat{q}_x - c_x) \quad b_{0,1} = \frac{1}{2}(\hat{q}_y - c_y), \quad b_{0,0} = 1 - (\hat{q}_x + \hat{q}_y). \tag{39}$$

By performing a truncation error analysis, both cross contributions to numerical viscosity (proportional to $\partial_x \partial_y q$) and normal contributions (proportional to $\partial_x^2 q, \partial_y^2 q$) are found. The cross coefficient of viscosity in the x, y direction:

$$\mu_{x,y} = -\frac{\Delta x \Delta y}{\Delta t} c_x c_y \tag{40}$$

does not depend on $\hat{q}_x$ and $\hat{q}_y$. The normal coefficients of viscosity in the x and y direction read:

$$\begin{aligned} \mu_{x,x} &= \frac{\Delta x^2}{2\Delta t}(\hat{q}_x - c_x^2), \\ \mu_{y,y} &= \frac{\Delta y^2}{2\Delta t}\left(\hat{q}_y - c_y^2\right). \end{aligned} \tag{41}$$

In order to minimise the truncation error, the normal contributions (41) must be minimised.

Monotonicity requires positivity (non-negativity) of all coefficients (39). This leads to the conditions:

$$\begin{cases} |c_x| \leqslant \hat{q}_x \leqslant 1 - \hat{q}_y, \\ |c_y| \leqslant \hat{q}_y \leqslant 1 - \hat{q}_x. \end{cases} \tag{42}$$

From (42), being $|c_x| \geqslant 0$ and $|c_y| \geqslant 0$, a wider set of inequalities can be written as:

$$\begin{cases} 0 \leqslant |c_x| \leqslant \hat{q}_x \leqslant 1 - \hat{q}_y \leqslant 1, \\ 0 \leqslant |c_y| \leqslant \hat{q}_y \leqslant 1 - \hat{q}_x \leqslant 1 \end{cases} \tag{43}$$

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

$$\begin{cases} 0 \leqslant |c_x| \leqslant 1, \\ 0 \leqslant |c_y| \leqslant 1, \end{cases} \tag{44}$$

$$\begin{cases} |c_x| \leqslant \hat{q}_x \leqslant 1, \\ |c_y| \leqslant \hat{q}_y \leqslant 1. \end{cases} \tag{45}$$

For any given value of $|c_x|$, $|c_y|$ in the range of (44), taking into account (45), and noting that $\mu_{x,x}$ and $\mu_{y,y}$ in (41) are linearly increasing functions of $\hat{q}_x, \hat{q}_y$, we can write:

$$\begin{aligned} \min_{|c_x| \leqslant \hat{q}_x \leqslant 1} (\mu_{x,x}) &= \mu_{x,x}\left(\min_{|c_x| \leqslant \hat{q}_x \leqslant 1} (\hat{q}_x)\right) = \mu_{x,x}(\hat{q}_x = |c_x|), \\ \min_{|c_y| \leqslant \hat{q}_y \leqslant 1} (\mu_{y,y}) &= \mu_{y,y}\left(\min_{|c_y| \leqslant \hat{q}_y \leqslant 1} (\hat{q}_y)\right) = \mu_{y,y}(\hat{q}_y = |c_y|). \end{aligned} \tag{46}$$

Therefore, the values of $\hat{q}_x$ and $\hat{q}_y$ which minimise $\mu_{x,x}$ and $\mu_{y,y}$ in (41), also satisfying (45), are

$$\begin{cases} \hat{q}_x = |c_x|, \\ \hat{q}_y = |c_y|. \end{cases} \tag{47}$$

In order to fully satisfy the monotonitcity requirements (42), by substituting (47) into (42) a further condition is found:

$$|c_x| + |c_y| \leqslant 1. \tag{48}$$

Table 1
Coefficients for selected schemes.

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

Substitution of (47) into (39) gives the Godunov upwind coefficients (35), while (48) turns out to be the Godunov upwind stability condition (36). □

This proof can be easily extended in an arbitrary number α of spatial dimensions.

3.2.2. The optimal upwind bias in UFORCE

All the numerical methods discussed in this paper, say UFORCE (flux is given in (22)), FORCE (flux is given in (7)), Lax–Friedrichs (flux is given in (11)) Lax–Wendroff (flux is given in (9) and (10)) and Godunov upwind (flux is given in (33) and (34)) methods belong to the class of five-point schemes (32). The coefficients of such methods to be inserted in (32) are given in Table 1. For the purpose at hand we have assumed upwind bias β_x and β_y constant inside each cell and indicated with $c_x=\frac{\lambda_x\Delta t}{\Delta x}$ and $c_y=\frac{\lambda_y\Delta t}{\Delta y}$ the directional CFL numbers. It is worth noticing that all methods given in Table 1 are directionally split since coefficients $b_{-1,0}$ and $b_{1,0}$ depend only on parameters involving quantities evaluated along the x direction (i.e. c_x and/or β_x) while $b_{0,-1}$ and $b_{0,1}$ depend only on parameters evaluated using quantities evaluated along the y direction (c_y and/or β_y).

We begin our analysis noting that the UFORCE method exactly reproduces different numerical methods provided suitable values for the upwind bias β_x and β_y are chosen. In fact, comparing the coefficients in Table 1 one can verify that setting $\beta_x=\beta_y=0$, the classical FORCE method (7) is exactly reproduced. Moreover setting

$$\beta_x=\pm\sqrt{\frac{1}{4}-c_x^2},\quad \beta_y=\pm\sqrt{\frac{1}{4}-c_y^2}, \tag{49}$$

the Lax–Wendroff method (9) and (10) is reproduced, while setting:

$$\beta_x^2-c_x^2+\frac{1}{4}=0,\quad \beta_y^2-c_y^2+\frac{1}{4}=0, \tag{50}$$

the Lax–Friedrichs scheme (11) is obtained.

Amongst all the possibilities available, the optimal choice for the upwind bias is the one that allows us to obtain the first-order monotone scheme with the smallest truncation error. In Section 3.2.1 it is proved that assuming the five-point stencil method (32) the first-order monotone scheme with the smallest truncation error is the Godunov upwind method whose coefficients are given in Table 1. By forcing the coefficients of the UFORCE (see Table 1) method to be those of two-dimensional Godunov method (Table 1), we obtain the following solutions for the upwind-bias:

$$\beta_x=\pm\frac{1}{2}(1-2|c_x|),\quad \beta_y=\pm\frac{1}{2}(1-2|c_y|), \tag{51}$$

which allows scheme (32) to exactly reproduce the Godunov upwind method in two dimensions.

The reader can also easily verify that upwind bias (51) represents also the limit for the monotonicity region of the proposed UFORCE method, i.e.:

$$|\beta_x|\leqslant\frac{1}{2}(1-2|c_x|)\quad\text{and}\quad|\beta_y|\leqslant\frac{1}{2}(1-2|c_y|). \tag{52}$$

Considering the optimal choice for the upwind bias (51), the Godunov upwind stability region (36) is found. The possible choices for the upwind bias are summarised in Fig. 4. Here the relationship between the upwind bias and the directional Courant number in both x and y directions is shown. The directionally split property of the analysed methods, allows one to plot the same graphs in the (c_x,β_x) and (c_y,β_y) plane, independently. Moreover, if we restrict the stability condition assuming $|c_x|\leqslant\frac{1}{2}$ and $|c_y|\leqslant\frac{1}{2}$, 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 but $|c_x|=\frac{1}{2}$ ($|c_y|=\frac{1}{2}$);
- any choice of β_x (β_y) lying within the quadrilateral of Godunov method leads to stable and monotone methods.

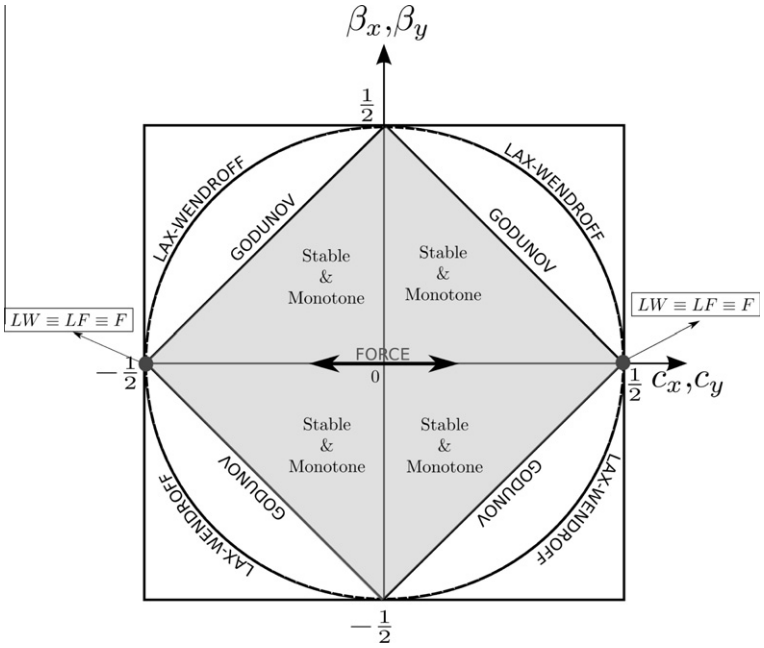

Fig. 4. UFORCE method applied to the linear advection Eq. (30): monotonicity and stability region for different choices of the directional bias in the (β_x, c_x) and (β_y, c_y) plane. Relationships between upwind-bias and directional Courant number which allow UFORCE to reproduce existing methods.

The numerical viscosity of methods in Fig. 4 is given by

$$\mu_{xx} = \frac{1}{4}\frac{\Delta x^2}{\Delta t}(1 - 4\beta_x^2), \quad \left(\mu_{yy} = \frac{1}{4}\frac{\Delta y^2}{\Delta t}\left(1 - 4\beta_y^2\right)\right). \tag{53}$$

The scheme possessing the maximum viscosity is the FORCE method ($\beta_x = 0$), while increasing the bias to $\beta_x = \pm\frac{1}{2}$ ($\beta_y = \pm\frac{1}{2}$), the numerical viscosity decrease to 0, thus representing second-order methods. The monotone scheme possessing the smallest numerical viscosity is the Godunov method. Exceeding the threshold imposed by the Godunov method, the numerical viscosity decreases, but oscillatory methods are obtained. Eventually, if we set $c_x = \pm\frac{1}{2}$ and $\beta_x = 0$ all the methods given in Table 1 coincide.

The analysis presented in this section can be easily extended in more than two dimensions. For α dimensions the relationship for the optimal choice of the upwind bias can be generalised as:

$$\beta_x = \pm\frac{1}{2}(1 - \alpha|c_x|), \tag{54}$$

where x denotes the current direction and c_x is the directional Courant number. In the next section we give a procedure for the non-linear case.

3.3. Optimal choice for the upwind bias: non-linear case

The theory developed in the previous section for the linear case has been extended empirically to non-linear systems.

For the purpose at hand we consider the two-dimensional non-linear inviscid shallow water equations. They have 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, respectively. Here u and v represent the particle velocity along x and y respectively, while $a = \sqrt{gh}$ is celerity, where g is acceleration due to gravity and h water depth (see Section 5 for details).

First, for the linear case, we note the following: (i) any choice for the sign inside expression (51) gives rise to identical results, while this is not true for the non-linear case, (ii) the local 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. The open questions for non-linear systems is thus, finding the correct sign $\mathcal{S}$ and the most suitable estimation for the eigenvalue λ_S to use in Eq. (51).

After numerical investigations we found that the expression for the upwind bias in the x direction (analogous considerations can be made for the y direction) can be written as:

$$(\beta_x)_{ij} = \mathcal{S}\left\{\frac{1}{2} - \lambda_S\left[\frac{\Delta t}{(\Delta x)_{ij}}\right]\right\}, \tag{55}$$

where

$$\mathcal{S} = \begin{cases} \text{sign}(u_{i,j}) & \text{if } (u_{i,j} \neq 0), \\ \text{sign}\left[\left(\lambda_x^{(3)}\right)_{i-1,j} + \left(\lambda_x^{(1)}\right)_{i+1,j}\right] & \text{if } (u_{i,j} = 0) \quad \text{and} \quad \left(\lambda_x^{(3)}\right)_{i-1,j} + \left(\lambda_x^{(1)}\right)_{i+1,j} \neq 0, \\ 0 & \text{if } (u_{i,j} = 0) \quad \text{and} \quad \left(\lambda_x^{(3)}\right)_{i-1,j} + \left(\lambda_x^{(1)}\right)_{i+1,j} = 0 \end{cases} \tag{56}$$

and the sign function is 1 if its argument is positive and -1 if it is negative. This means that the sign $\mathcal{S}$ in front the expression of the bias is chosen, in an upwind fashion, accordingly to the sign of the particle velocity inside the cell. If the velocity inside the cell is 0, 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 $\mathcal{S} = 0 \Rightarrow (\beta_x)_i = 0$, reproducing the standard FORCE scheme.

λ_S is an estimate for the maximum local characteristic speed in absolute value. Our analyses on the two-dimensional Burgers equation (having one eigenvalue $\lambda = u$) and the two-dimensional inviscid shallow water equations show that λ_S can computed as:

$$\lambda_S = \max_{-a_0 \leqslant a \leqslant a_0} \left(\max_k \left| \left(\lambda_x^{(k)}\right)_{i+a,j} \right| \right). \tag{57}$$

Relationship (57) takes into account the absolute value of the velocity of the fastest wave, for any $-a_0 \leqslant a \leqslant a_0$ where a_0 takes values 0 or 1. If $a_0 = 1$, λ_S is evaluated with the eigenvalues in cell (i,j) and its neighbours $(i-1,j)$ and $(i+1,j)$, while if $a_0 = 0$ λ_S is evaluated with the eigenvalues in cell (i,j) only. Our numerical experiments show that setting $a_0 = 0$ gives rise to sharper profiles, which are more likely to be affected by oscillations in presence of large spatial gradients, while the choice $a_0 = 1$ gives oscillation-free results, with increased numerical diffusion. The choice of a_0 is done empirically, according to the behaviour of the hyperbolic system considered. In our applications we found that for the inviscid Burgers equation the choice $a_0 = 1$ is required in order to avoid spurious oscillations, while for the shallow water equations oscillation-free solutions were found by setting $a_0 = 0$. So for the shallow water equations we recommend:

$$\lambda_S = \max_{k=1,2,3} \left(\left| \left(\lambda_x^{(k)}\right)_{i,j} \right| \right), \tag{58}$$

which comes from (57) having set $a_0 = 0$. Finally, the generalisation of the optimal bias in the x direction of the UFORCE method on an arbitrary number α of dimensions can be written in the following way:

$$(\beta_x)_i = \mathcal{S}\frac{1}{2}\left\{1 - \alpha\lambda_S\left[\frac{\Delta t}{(\Delta x)_i}\right]\right\}. \tag{59}$$

It is worth mentioning that the analysis conducted in this section can be straightforward extended to other systems of equations characterised by eigenvalues of the form $u \pm a$ as the Euler equations, while it is also extended to other hyperbolic systems characterised by a different expression for the eigenvalues if an estimation of the maximum eigenvalue and criteria for the $\mathcal{S}$ function are provided.

4. Second order non-oscillatory extension

In this section we extend the first-order UFORCE method to second order in space and time using the MUSCL-Hancock approach [18] for the two-dimensional case. We remark that strictly speaking the MUSCL technique applies to regularly spaced grids. This method proceeds as follows:

(I) MUSCL reconstruction. The cell averages $\mathbf{Q}_{i,j}^n$ are reconstructed independently in the x and y directions by selecting respective slope vectors $\overline{\Delta}_i$ and $\overline{\Delta}_j$, where the bar indicates a limited slope (difference) so as to avoid spurious oscillations near large gradients of the solution.
Boundary extrapolated values are

$$\begin{aligned} \mathbf{Q}_{i,j}^{-x} &= \mathbf{Q}_{i,j}^n - \frac{1}{2}\overline{\Delta}_i; \quad \mathbf{Q}_{i,j}^{+x} = \mathbf{Q}_{i,j}^n + \frac{1}{2}\overline{\Delta}_i, \\ \mathbf{Q}_{i,j}^{-y} &= \mathbf{Q}_{i,j}^n - \frac{1}{2}\overline{\Delta}_j; \quad \mathbf{Q}_{i,j}^{+y} = \mathbf{Q}_{i,j}^n + \frac{1}{2}\overline{\Delta}_j. \end{aligned} \tag{60}$$

(II) Evolution of these states by time $\frac{1}{2}\Delta t$ as follows:

$$\bar{\mathbf{Q}}_{i,j}^l = \mathbf{Q}_{i,j}^l + \frac{\Delta t}{2\Delta x}\left[\mathbf{F}(\mathbf{Q}_{i,j}^{-x}) - \mathbf{F}(\mathbf{Q}_{i,j}^{+x})\right] + \frac{\Delta t}{2\Delta y}\left[\mathbf{G}(\mathbf{Q}_{i,j}^{-y}) - \mathbf{G}(\mathbf{Q}_{i,j}^{+y})\right] \tag{61}$$

for $l = -x, +x, -y, +y$.

Table 2
Convergence rates study for the two-dimensional UFORCE scheme, (c_0 = 0.1 m, q_0 = 1 m, T = 10 s, L = 400 m).

N	Variable Q			
	L_1	$\mathcal{O}(L_1)$	L_∞	$\mathcal{O}(L_\infty)$
80	5.103E−02	–	8.029E−02	–
160	1.393E−02	1.873	2.188E−02	1.876
320	3.060E−03	2.187	4.806E−03	2.187
640	7.229E−04	2.082	1.136E−03	2.081
1280	1.777E−04	2.024	2.792E−04	2.024
2560	4.426E−05	2.006	6.953E−05	2.006
5120	1.104E−05	2.003	1.734E−05	2.003

(III) Computation of UFORCE flux. This is done by evaluation of intermediate states applying the UFORCE scheme (22) with:

$$\mathbf{F}^{uLW2}_{i+\frac{1}{2}j} = \mathbf{F}\left(\mathbf{Q}^{uLW2}_{i+\frac{1}{2}j}\right) \tag{62}$$

and

$$\mathbf{Q}^{uLW2}_{i+\frac{1}{2}j} = \frac{1}{2\left[1-(\beta_x)_{i+1\,j}+(\beta_x)_{i\,j}\right]}\left\{(1+2(\beta_x)_{i\,j})\bar{\mathbf{Q}}^{-x}_{i+1\,j}+(1-2(\beta_x)_{i+1\,j})\bar{\mathbf{Q}}^{+x}_{i\,j}-\left(\frac{2\Delta t}{\Delta x}\right)[\mathbf{F}(\bar{\mathbf{Q}}^{-x}_{i+1\,j})-\mathbf{F}(\bar{\mathbf{Q}}^{+x}_{i\,j})]\right\}. \tag{63}$$

The second flux term in (22) is evaluated as:

$$\mathbf{F}^{uLF2}_{i+\frac{1}{2}j} = \frac{1}{2\left[1-(\beta_x)_{i+1\,j}+(\beta_x)_{i\,j}\right]}\left\{(1+2(\beta_x)_{i\,j})\mathbf{F}(\bar{\mathbf{Q}}^{-x}_{i+1\,j})+(1-2(\beta_x)_{i+1\,j})\mathbf{F}(\bar{\mathbf{Q}}^{+x}_{i\,j})\right] - \left(\frac{\Delta x}{2\Delta t}\right)(1+2(\beta_x)_{i\,j})(1-2(\beta_x)_{i+1\,j})(\bar{\mathbf{Q}}^{-x}_{i+1\,j}-\bar{\mathbf{Q}}^{+x}_{i\,j})\Big\}. \tag{64}$$

In order to avoid oscillations near large gradients we limit the slope (difference) $\overline{\Delta}_i$ following the ENO approach [5], namely:

$$\overline{\Delta}_i = minmax\left[(\mathbf{Q}^n_{i+1\,j}-\mathbf{Q}^n_{i\,j}),(\mathbf{Q}^n_{i\,j}-\mathbf{Q}^n_{i-1\,j})\right], \tag{65}$$

where the minmax function

$$minmax[a,b] = \begin{cases} a & \text{if } |a| \leqslant |b|, \\ b & \text{otherwise} \end{cases} \tag{66}$$

is applied componentwise to the left and right difference vectors $(\mathbf{Q}^n_{i+1\,j}-\mathbf{Q}^n_{i\,j}),(\mathbf{Q}^n_{i\,j}-\mathbf{Q}^n_{i-1\,j})$.

4.1. Numerical convergence study

Here we compute the order of accuracy of the scheme to verify that the expected theoretical order is achieved. We consider the one-dimensional linear advection equation, but computing the solution in a fully two-dimensional setting. In order to validate the order of accuracy an exact solution is constructed by prescribing a function for $q(x,y,t)$, which satisfies exactly (30). It reads

$$q(x,y,t) = q_0 - \frac{c_0}{k}\sin(kx-\omega t), \tag{67}$$

where $k=\frac{2\pi}{L}$, $\omega=\frac{2\pi}{T}$, $\lambda_x=\frac{\omega}{k}$, L and T being the wave length and the period of the sinusoidal oscillation, respectively.

We solve the equations using a MUSCL-Hancock extension of the UFORCE method without making use of slope limiters. We use a domain of 800 × 100 m, imposing periodic boundary conditions. Table 2 shows the errors, quantified through the standard norms L_1, L_∞, and the relative convergence rates for variable q at time t = 10 s with q_0 = 1 m, c_0 = 0.1 m, T=10 s and L=400 m. The expected orders of accuracy are achieved with each norm.

5. Applications to the two-dimensional shallow water equations

Here we apply the second order version of the UFORCE method proposed in this paper to 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:

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \partial_y \mathbf{G}(\mathbf{Q}) = \mathbf{0}, \tag{68}$$

where the vector of conserved variables $\mathbf{Q}$ and the fluxes along the x and y direction $\mathbf{F}(\mathbf{Q})$, $\mathbf{G}(\mathbf{Q})$ in Eq. (68), can be written as:

$$\mathbf{Q} = \begin{bmatrix} h \\ hu \\ hv \\ hC \end{bmatrix}, \quad \mathbf{F}(Q) = \begin{bmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \\ huv \\ huC \end{bmatrix}, \quad \mathbf{G}(Q) = \begin{bmatrix} hv \\ huv \\ hv^2 + \frac{1}{2}gh^2 \\ hvC \end{bmatrix}. \tag{69}$$

Here $u(x,y,t)$ and $v(x,y,t)$ are the x and y components of velocity, $h(x,y,t)$ is water depth, $C(x,y,t)$ is the passive scalar concentration and $g = 9.81\ \mathrm{ms}^{-1}$ is the acceleration due to gravity. The Jacobians of fluxes $\mathbf{F}(\mathbf{Q})$, $\mathbf{G}(\mathbf{Q})$ admit three distinct real eigenvalues in each direction, namely

$$\begin{bmatrix} \lambda_x^{(1)} \\ \lambda_x^{(2)} \\ \lambda_x^{(3)} \end{bmatrix} = \begin{bmatrix} u - a \\ u \\ u + a \end{bmatrix}, \quad \begin{bmatrix} \lambda_y^{(1)} \\ \lambda_y^{(2)} \\ \lambda_y^{(3)} \end{bmatrix} = \begin{bmatrix} v - a \\ v \\ v + a \end{bmatrix}, \tag{70}$$

where $a = \sqrt{gh}$ is celerity and $\lambda_x^{(2)}$ and $\lambda_y^{(2)}$ have 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.

Results have been obtained enforcing the following CFL condition:

$$\Delta t \leqslant \frac{\mathrm{CFL}}{\max\left(\frac{\max\limits_{k=1,2,3} \left| \left(\lambda_x^{(k)}\right)_{i,j} \right|}{\Delta x}, \frac{\max\limits_{k=1,2,3} \left| \left(\lambda_y^{(k)}\right)_{i,j} \right|}{\Delta y} \right)} \tag{71}$$

with $0 < \mathrm{CFL} < \frac{1}{2}$, where $\max_{k=1,2,3} |(\lambda_x^{(k)})_{i,j}| = |u_{i,j}| + a_{i,j}$ and $\max_{k=1,2,3} |(\lambda_y^{(k)})_{i,j}| = |v_{i,j}| + a_{i,j}$ are the maximum wave propagation speeds in absolute value in the x and y directions. Four numerical methods have been used in addition to the UFORCE method: one centred method (the FORCE method) and three Godunov-type upwind methods. Among the latter category, we used a Godunov method coupled with an exact Riemann solver (since now, referred as Godunov-exact), a Godunov method coupled with the HLL Riemann solver (referred as Godunov-HLL) and a Godunov method based on the Rusanov flux (referred as Rusanov). All the results presented here are second-order accurate obtained using the MUSCL-Hancock ENO reconstruction described in Section 4.

Three test problems have been solved in order to assess the behaviour of the UFORCE method: namely the collapse of a circular dam, the collapse of a transversal dam and the collapse of a circular dam solved on a variably-spaced grid.

5.1. Test 1: collapse of a circular dam

This test case consists of the instantaneous breaking of a cylindrical tank initially filled with 2.5 m deep water at rest. When the column of water 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 0.5 m.

We solve (68) and (69) together with initial conditions:

$$\begin{cases} h(x,y,0) = 2.5\ \mathrm{m} & \text{if } x^2 + y^2 \leqslant R^2, \\ h(x,y,0) = 0.5\ \mathrm{m} & \text{if } x^2 + y^2 > R^2, \\ u(x,y,0) = 0 & \forall\, x,y, \\ v(x,y,0) = 0 & \forall\, x,y, \end{cases} \tag{72}$$

being $R = 2.5$ m the tank radius. We use a coarse mesh of 101×101 cells in the square computational domain $([-20,20] \times [-20,20])$ m with transmissive boundary conditions. Solution is computed at time $t = 1.4$ s.

We provide an accurate reference solution, which was obtained by turning the problem (68), (69), (72) into a one-dimensional problem in the radial direction (see [16]):

$$\partial_t \begin{bmatrix} h \\ hu_r \end{bmatrix} + \partial_r \begin{bmatrix} hu_r \\ hu_r^2 + \frac{1}{2}gh^2 \end{bmatrix} = -\frac{1}{r} \begin{bmatrix} hu_r \\ hu_r^2 \end{bmatrix}, \tag{73}$$

where r is the radial coordinate and $u_r(r,t)$ the radial velocity. The initial conditions (72) in the radial coordinate system read:

$$\begin{cases} h(r,0) = 2.5\ \mathrm{m} & \text{if } r \leqslant R, \\ h(r,0) = 0.5\ \mathrm{m} & \text{if } r > R, \\ u(r,0) = 0 & \forall\, r. \end{cases} \tag{74}$$

System (73) and (74) is solved numerically on a fine mesh of 1000 cells using the WAF method in conjunction with the HLLC approximate Riemann solver [16]. The CFL number is set to 0.9 and the limiter used is SUPERBEE [12].

The reference solution exhibits an outer circular shock and a circular rarefaction following the shock. With this test we aim to assess the ability of the UFORCE 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 wave 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.

Results for this test are displayed in Figs. 5–7. In Figs. 5 and 6 we show numerical results for the depth $h(x,y,t)$. The solutions obtained with five numerical methods (symbols) are presented in terms of slices along the x-axis ($y = 0$) and compared with the reference radial solution (full line). The numerical results in Fig. 5 have been obtained setting the CFL to 0.1. At a low value of the CFL number the solution obtained with the centred method (FORCE) significantly differs from the solution obtained with genuinely upwind methods (Godunov-exact, Godunov-HLL and Rusanov), being excessively smoothed and smeared both in the shock and in the rarefaction zones. Among the upwind methods, slight differences are found in the rarefaction zones, depending on the kind of Riemann solver used, while the left and right facing shocks are solved almost to the

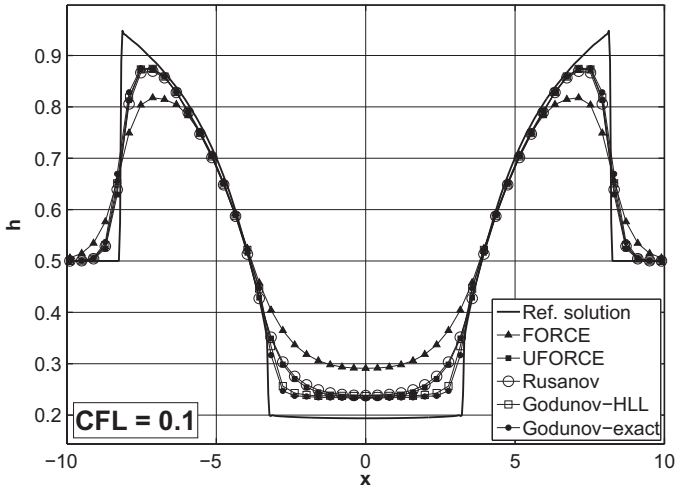

Fig. 5. Test 1: collapse of a circular dam. Numerical results for water depth h of FORCE, UFORCE, Rusanov, Godunov-HLL and Godunov-exact numerical methods (symbols) are compared with the reference radial solution (full line) at time $t = 1.4$ s. The numerical solution profiles are sliced on the x-axis. The mesh used is 101×101 cells and CFL is set to 0.1.

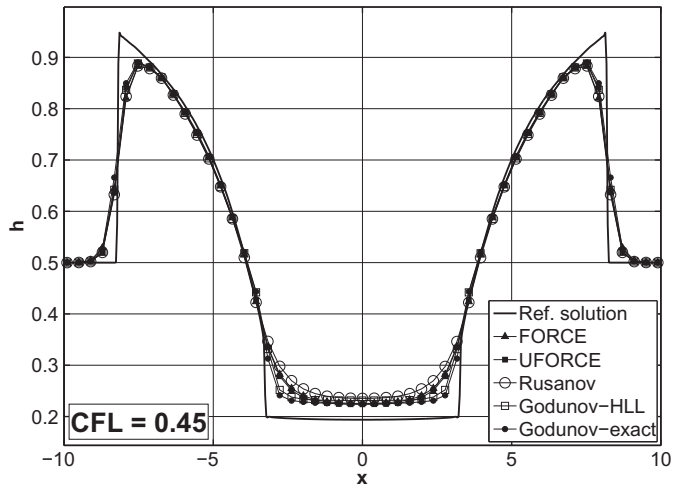

Fig. 6. Test 1: collapse of a circular dam. Numerical results for water depth h of FORCE, UFORCE, Rusanov, Godunov-HLL and Godunov-exact numerical methods (symbols) are compared with the reference radial solution (full line) at time $t = 1.4$ s. The numerical solution profiles are sliced on the x-axis. The mesh used is 101×101 cells and CFL is set to 0.45.

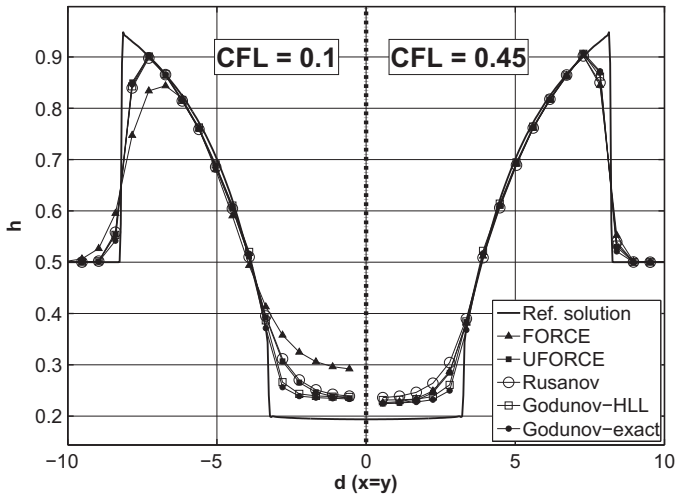

Fig. 7. Test 1: collapse of a circular dam. Numerical results for water depth h of FORCE, UFORCE, Rusanov, Godunov-HLL and Godunov-exact numerical methods (symbols) are compared with the reference radial solution (full line) at time $t = 1.4$ s. The numerical solution profiles are sliced on the diagonal $x = y$. Results obtained with CFL = 0.1 and CFL = 0.45 are shown, for $x < 0$ and $x > 0$, respectively. The mesh used is 101×101 cells.

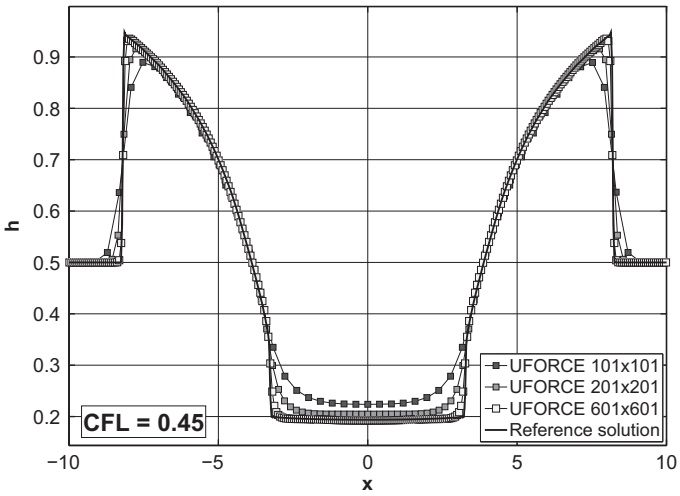

Fig. 8. Test 1: collapse of a circular dam. Numerical results for water depth h of the UFORCE methods (symbols) obtained at different grid resolution (101×101,201×201,601×601 cells) are compared with the reference radial solution (full line) at time $t = 1.4$ s. The numerical solution profiles are sliced on the x-axis. CFL is set to 0.45.

same accuracy by all the three upwind methods. The UFORCE method turns out to improve significantly accuracy compared to FORCE both in the shock and in the rarefaction zones, reaching the same degree of accuracy as the Rusanov method.

The numerical results shown in Fig. 6, which have been obtained setting CFL = 0.45, show that at a larger value of the CFL number the behavioural differences among centred and upwind methods tend to vanish. In fact, the FORCE method recovers accuracy, reaching almost the same resolution of the other methods. Still, some differences remain in the description of the rarefaction region, where both the FORCE and UFORCE methods outperform the Rusanov method.

In Fig. 7 the results obtained using CFL = 0.1 and CFL = 0.45 are plotted in terms of slices along the diagonal $x = y$. Fig. 7 shows that loss of accuracy at low CFL numbers is significant for the FORCE (centred) method, while the UFORCE is less dissipative, performing analogously to the Rusanov method.

Finally, in Fig. 8 it is shown that the proposed UFORCE method converges to the correct solution as the grid is refined.

5.2. Test 2: collapse of a transversal dam

This test consists of the collapse of a dam positioned along the diagonal direction $x = y$. Across the wall the water depth h initially exhibits a discontinuity, being 1 m on the top-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 in the top-left side and of a shock wave on the other side, travelling 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 (68) and (69) with initial conditions

$$\begin{cases} \begin{cases} h(x,y,0) = 1\ \text{m} \\ C(x,y,0) = 1 \end{cases} \text{if } x \leqslant y \quad \text{and} \begin{cases} h(x,y,0) = 0.5\ \text{m} \\ C(x,y,0) = 0 \end{cases} \text{if } x > y \\ u(x,y,0) = v(x,y,0) = 0 \quad \forall\, x,y. \end{cases} \tag{75}$$

We use a coarse mesh of 101×101 cells in the square computational domain ($[-25,25] \times [-25,25]$) with transmissive boundary conditions. Solution is computed at time $t = 5$ s.

An exact solution for this problem can be computed by solving a one-dimensional dam-break problem in the transversal direction ($y = -x$) 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 [15] for an accurate description). We focus our attention to the contact discontinuity and discuss the results in terms of concentration 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, the upwind schemes based on the HLL Riemann solver behave like centred methods for linear fields [13].

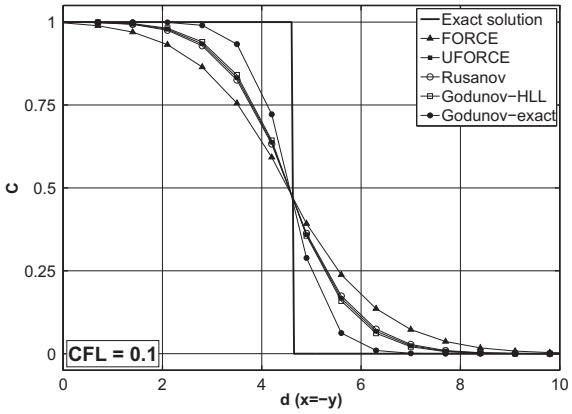

Fig. 9. Test 2: collapse of a transversal dam. Numerical results for concentration C of FORCE, UFORCE, Rusanov, Godunov-HLL and Godunov-exact numerical methods (symbols) are compared with the exact solution (full line) at time $t = 5$ s. The numerical solution profiles are sliced on the diagonal $x = -y$. The mesh used is 101×101 cells and CFL is set to 0.1.

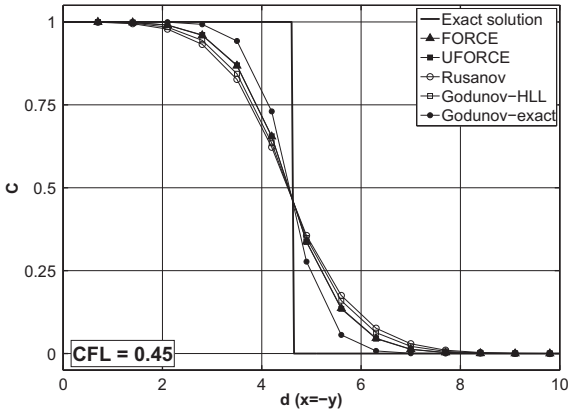

Fig. 10. Test 2: collapse of a transversal dam. Numerical results for concentration C of five numerical methods (symbols) are compared with the exact solution (full line) at time $t = 5$ s. The numerical solution profiles are sliced on the diagonal $x = -y$. The mesh used is 101×101 cells and CFL is set to 0.45.

Results for this test are displayed in Figs. 9 and 10. The solution for variable $C(x,y,t)$ is represented in terms of slices in the transversal direction ($y = -x$). In Fig. 9 results computed at CFL = 0.1 are given. It is seen that the Godunov-exact method gives rise to the sharpest resolution of this wave, describing the concentration discontinuity using 8 cells, while the FORCE method excessively smears the solution, requiring 15 cells for the discontinuity description. The Rusanov and Godunov-HLL methods give very similar results, requiring 12 cells for the description of the contact discontinuity. The UFORCE method significantly outperforms the FORCE method giving results that are similar to those obtained using the Godunov methods with approximate Riemann solvers.

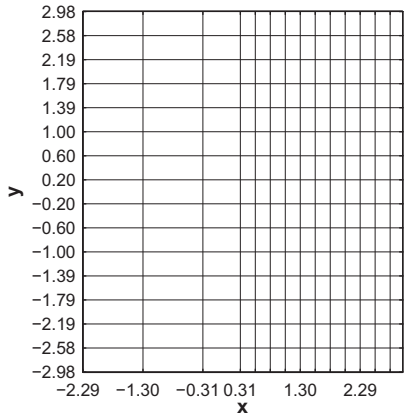

Fig. 11. Test 3: collapse of a circular dam on a variably space grid. Sketch of the grid.

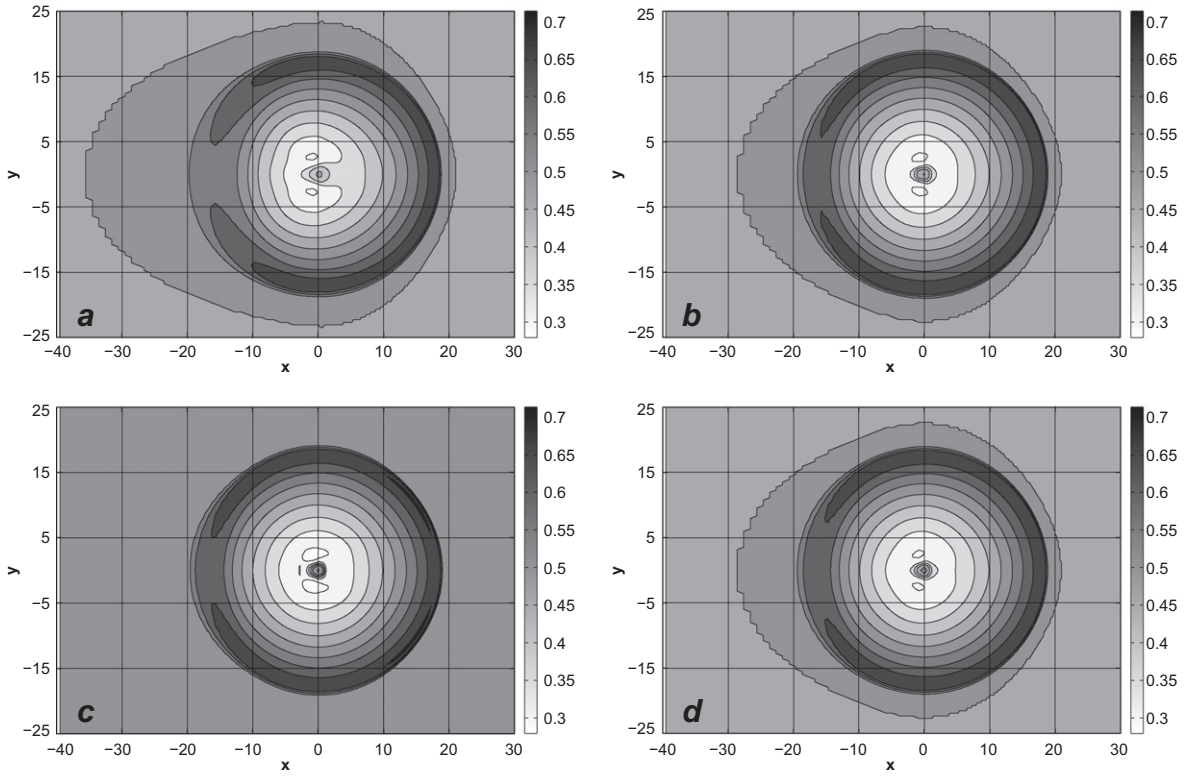

Fig. 12. Test 3: collapse of a circular dam on a variably space grid. Numerical results for water depth h for FORCE (a), UFORCE (b), Godunov-exact (c), Godunov-HLL method (d) are presented at time $t = 4.7$ s in terms of contourplots. The mesh used is 201 equally spaced cells in the y direction and 201 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 0.45.

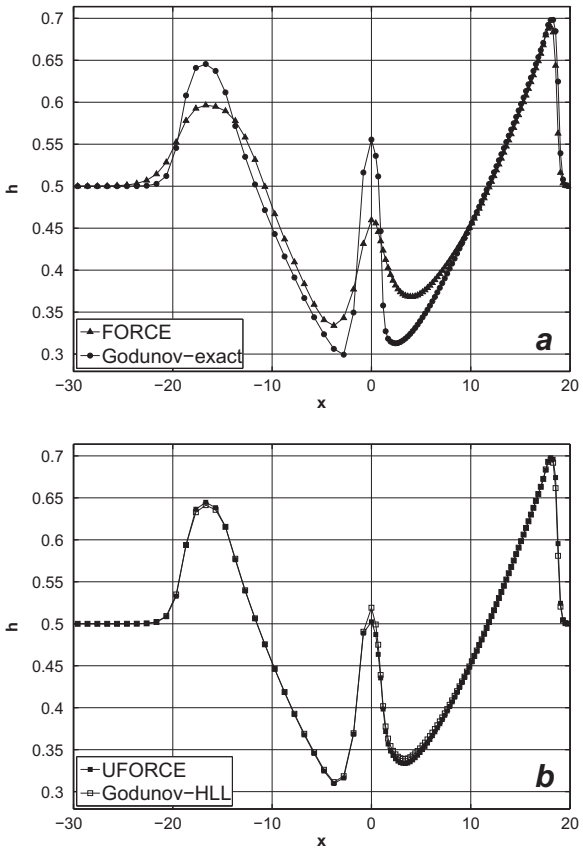

Fig. 13. Test 3: collapse of a circular dam on a variably space grid. Numerical results for water depth h for FORCE and Godunov-exact methods (a) and for UFORCE and Godunov-HLL (b) methods are presented at time $t = 4.7$. The numerical solution profiles are sliced on the x-axis. The mesh used is 201 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 0.45.

For CFL = 0.45 (Fig. 10) both centred methods (FORCE, UFORCE) and upwind methods based on approximate Riemann Solvers (Godunov-HLL, Rusanov) give rise to comparable results, while Godunov-exact outperforms all the other methods presented.

5.3. Test 3: collapse of a circular dam on a variably-spaced grid

As it was shown in the previous sections, an attractive feature of the UFORCE method rely on its ability to perform consistently in 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 (68), (69) and (72). However, we use a different mesh from that in (5.1). We use 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, namely $\Delta x = 0.99$ 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. 11.

We impose transmissive boundary conditions. Solution is computed at time $t = 4.7$ s using CFL = 0.45.

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 overexpansion of the flow caused by the reflection of the interior rarefaction from the centre of the dam (see [16] 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.

In our test case however 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 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. So, the interaction between a highly variable x-spacing and the CFL condition results in much more

accurate solution in the fine mesh part of the domain ($x > 0$) compared to the coarse mesh part ($x < 0$). In this test case, preserving symmetry along the y-axis is the challenge.

Results to this test are displayed in Fig. 12 in terms of contourplots, while in Fig. 13a-b are given results in terms of slices along the x-axis ($y = 0$). In Fig. 12a we show that the FORCE method does not preserve symmetry because of an excessive smearing at low local CFL number in the coarse mesh side of the domain, while the Godunov-exact solution (Fig. 12c) gives rise to optimal results preserving symmetry in all the directions. The UFORCE solution (Fig. 12b) shows a significant improvement compared to the FORCE method, possessing the same degree of symmetry as the Godunov-HLL numerical solution (Fig. 12d). Similar conclusions can be drawn analysing Fig. 13a-b. 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. 13a) are affected by a severe numerical diffusion on the coarse mesh side of the domain if compared with the Godunov-exact method, 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. 13b), even though the Godunov-HLL method, due to its genuinely upwind nature, has a slight advantage over the UFORCE method in the solution of the shock reflection around $x = 0$, but turns out to be slightly less accurate in the description of the rarefactions (see Fig. 13b).

6. Conclusions

An upwind-biased version of FORCE flux in multiple-space dimensions for solving hyperbolic equations in conservation-law form on structured meshes has been presented. Monotonicity, linear stability and numerical viscosity of the schemes have been analysed in two space dimensions. We have then extended the new UFORCE fluxes to second order of accuracy in space and time in the framework of finite volume methods. Results of the second-order version of the proposed method on structured meshes have been shown. For the shallow water equations the performance of the numerical scheme has been assessed by solving some well-established test problems on structured meshes. The solution obtained have been compared with the classical centred FORCE scheme and with the Godunov upwind method in conjunction with exact and approximate Riemann solvers.

The main feature of the proposed scheme is simplicity; it only requires partial knowledge of the eigenstructure of the system of equations, i.e. an estimate of the largest eigenvalue and it does not require the availability of a Riemann solver. It improves the accuracy of the solution for small Courant numbers and intermediate waves associated with linearly degenerate fields (contact discontinuities, shear waves and material interfaces). It achieves comparable accuracy to that of upwind methods used in conjunction with approximate Riemann solvers, though retaining the simplicity and efficiency of centred methods. This is attractive for applications in geophysical flows in which contact discontinuities play an important role as well as the dynamics evolve experiencing different ranges of CFL numbers.

Future developments of UFORCE concerns the extension to high order of accuracy in space and time and extension to unstructured grids for both conservative and non-conservative systems.

References

[1] P. Arminjon, A. St-Cyr, Nessyahu–Tadmor-type central finite volume methods without predictor for 3D Cartesian and unstructured tetrahedral grids, Applied Numerical Mathematics 46 (2) (2003) 135–155.
[2] G. Chen, E. Toro, Centered difference schemes for nonlinear hyperbolic equations, Journal of Hyperbolic Differential Equations 1 (2004) 531–566.
[3] S.K. Godunov, Finite Difference Methods for the Computation of Discontinuous Solutions of the Equations of Fluid Dynamics, Mat. Sb. 47 (1959) 271–306.
[4] Guang-Shan Jiang, E. Tadmor, Nonoscillatory central schemes for multidimensional hyperbolic conservation laws, SIAM Journal on Scientific Computing 19 (6) (1998) 1892–1917.
[5] A. Harten, B. Engquist, S. Osher, S. Chakravarthy, Uniformly high order accuracy essentially non-oscillatory schemes iii, Journal of Computational Physics (1987).
[6] A. Kurganov, S. Noelle, G. Petrova, Semidiscrete central-upwind schemes for hyperbolic conservation laws and Hamilton–Jacobi equations, SIAM Journal on Scientific Computing 23 (3) (2001) 707–740.
[7] A. Kurganov, G. Petrova, Central schemes and contact discontinuities, Esaim – Mathematical Modelling and Numerical Analysis – Modelisation Mathematique et Analyse Numerique 34 (6) (2000) 1259–1275.
[8] A. Kurganov, G. Petrova, Central-upwind schemes on triangular grids for hyperbolic systems of conservation laws, Numerica Methods For Partial Differential Equations 21 (3) (2005) 536–552.
[9] A. Kurganov, E. Tadmor, New high-resolution central schemes for nonlinear conservation laws and convection–diffusion equations, Journal of Computational Physics 160 (1) (2000) 241–282.
[10] P. Lax, Weak Solutions of Nonlinear Hyperbolic Equations and Their Numerical Computation, Communications on Pure and Applied Mathematics VII (1954) 159–193.
[11] H. Nessyahu, E. Tadmor, Non-oscillatory central differencing for hyperbolic conservation-laws, Journal of Computational Physics 87 (2) (1990) 408–463.
[12] P.L. Roe, Some contributions to the modelling of discontinuous flows, in: Proceedings of the SIAM/AMS Seminar, 1983.
[13] E. Toro, Riemann Solvers and Numerical Methods for Fluid Dynamics, third ed., Springer-Verlag, 2009.
[14] E. Toro, A. Hidalgo, M. Dumbser, FORCE schemes on unstructured meshes I: Conservative hyperbolic systems, Journal of Computational Physics 228 (9) (2009) 3368–3389.
[15] E. Toro, A. Siviglia, Price: primitive centred schemes for hyperbolic systems, International Journal for Numerical Methods in Fluids (2003) 1263–1291.
[16] E.F. Toro, Shock-Capturing Methods for Free-Surface Shallow Flows, Wiley and Sons Ltd., 2001.
[17] E.F. Toro, S.J. Billet, Centered TVD schemes for hyperbolic conservation laws, IMA Journal of Numerical Analysis 20 (2000) 44–79.
[18] B. van Leer, Towards the ultimate conservative difference Scheme V. A second order sequel to Godunov's method, Journal of Computational Physics 32 (2) (1979) 101–136.