

Two-dimensional Sediment Transport models in Shallow Water equations. A second order finite volume approach on unstructured meshes. *

M. J. Castro Díaz[†], E. D. Fernández-Nieto[‡], A.M. Ferreiro[§], C. Parés[¶]

Abstract

In this paper we study the numerical approximation of bedload sediment transport due to shallow layer flows. The hydrodynamical component is modeled by a 2D shallow water system and the morphodynamical component by a solid transport discharge formula that depends on the hydrodynamical variables. The coupled system can be written as a nonconservative hyperbolic system. To discretize it, first we consider a Roe-type first order scheme as well as a variant based on the use of flux limiters. These first order schemes are then extended to second order accuracy by means of a new MUSCL-type reconstruction operator on unstructured meshes. Finally, some numerical tests are presented.

Short title : Second order discretization of 2D sediment transport models.

Keywords : Finite Volume Method, shallow water, bedload sediment transport, nonconservative hyperbolic system, high-order schemes, unstructured meshes, well-balanced.

Subject Classifications : AMS (MOS) : 65N06, 76B15, 76M20, 76N99.

1 Introduction

In this work we consider the bedload transport caused in a sediment layer by the flow of a shallow layer of fluid. The model considered here is a coupled model composed by a hydrodynamical and a morphodynamical component. The hydrodynamical component is given by a 2D shallow water system and the morphodynamical one by a solid transport discharge formula that depends on the hydrodynamical variables. Several expressions for this formula has been proposed by different authors: Grass [14], Meyer-Peter&Müller [25], Nielsen [26], Van Rjin [39], Fowler et al. [13], etc. In general, these formulae are obtained from empirical considerations.

Both the hydrodynamical and the morphodynamical components define a P.D.E. system with four unknowns: $h(\mathbf{x}, t)$ the thickness of the fluid; $q_1(\mathbf{x}, t)$, $q_2(\mathbf{x}, t)$, the discharge of the fluid in the horizontal directions, and $z_b(\mathbf{x}, t)$, the thickness of the sediment layer. This kind of systems, which are known as Saint-Venant-Exner models (see for example [22] and the corresponding references), can be written as a first order nonconservative system of the form:

*This research has been partially supported by the Spanish Government Research projects MTM2006-08075, MTM2006-01275 and P06-RNM-01594. The numerical computations have been performed at the Laboratory of Numerical Methods of the University of Málaga.

[†]Dpto. Análisis Matemático, Universidad de Málaga, Campus Teatinos s/n, Málaga, Spain (castro@anamat.cie.uma.es).

[‡]Departamento de Matemática Aplicada I, Universidad de Sevilla. 41012 Sevilla, Spain (edofner@us.es).

[§]Dpto. Matemáticas, Universidad de A Coruña, Campus de Elviña S/N, A Coruña(aferreiro@udc.es).

[¶]Dpto. Análisis Matemático, Universidad de Málaga, Campus Teatinos s/n, Málaga, Spain (castro@anamat.cie.uma.es).

$$\frac{\partial W}{\partial t} + \mathcal{A}_1(W) \frac{\partial W}{\partial x_1} + \mathcal{A}_2(W) \frac{\partial W}{\partial x_2} = 0. \quad (1)$$

The presence of nonconservative products add some important difficulties to both the theoretical and the numerical analysis of the system. In particular for discontinuous solutions these products do not make sense in general within the framework of distributions. Here, we follow the theory developed by Dal Maso, LeFloch and Murat in [24] to give a sense to these products as Borel measures. This theory, which is based on the choice of a family of paths, not only gives a way to properly define the concept of weak solutions of nonconservative systems but it also gives a theoretical framework for the design of finite volume methods for first order nonconservative hyperbolic systems: see [28]. This framework is based on the concept of *path-conservative* numerical scheme, which is a generalization of the usual concept of conservative method for conservation laws.

One of the main difficulties in designing good numerical methods for the particular case of bedload sediment transport models is related to the fact that the interaction between the fluid and the sediment layers is usually very weak (see [4]). Due to this, first order numerical schemes are, in general, too diffusive and numerical methods which are at least second order accurate are thus needed.

The goal of this work is to design a robust second order finite volume numerical scheme for Saint-Venant-Exner models. We consider unstructured meshes in order to make easier the adaptation to complex geometries. To design such a method we follow the general framework described in [5]: first, a Roe-type scheme is obtained based on the notion of Roe linearization introduced by Toumi in [36]. This notion also depends on the choice of a family of paths: here, the family of straight segments is considered. We also present some new variants of these Roe schemes based on a generalized flux limiter technique. Next, the accuracy of the schemes is increased by using a second order reconstruction operator. The main difficulty to define such an operator on an unstructured mesh is that, in general, huge complex stencils are needed what increases dramatically the computational cost. In [17] two high order schemes based on WENO reconstructions for 2D hyperbolic conservative problems on unstructured meshes composed by triangles have been presented: a third order and a fourth order operator based, respectively, on a combination of linear and quadratic polynomials. In [30] a third order non-oscillatory reconstruction for unstructured quadrilateral meshes is proposed which is based on a bi-hyperbolic reconstruction. In [9], [10] and [11] a family of non-oscillatory finite volume and discontinuous Galerkin schemes of arbitrary accuracy in space and time for solving hyperbolic systems on unstructured triangular and tetrahedral grids using the ADER approach has been introduced. These authors propose a new WENO reconstruction technique that can be easily evaluated and differentiated at any point. In [34] a discontinuous Galerkin finite element for river bed evolution has also been presented. In this work, we present a new second order reconstruction operator for unstructured meshes of MUSCL type (see [1]).

This paper is organized as follows: in Section 2 the system of equations of the Saint-Venant-Exner models considered here are presented. In Section 3 the main ingredients to design the numerical schemes are presented: first the general form of a Roe's scheme for (1) is recalled and a generalized flux limiter technique is also introduced to reduce the numerical diffusion. Then, the extension to higher order by means of a reconstruction operator is discussed. Section 4 is devoted to the definition of the MUSCL type reconstruction used here, achieving second order accuracy. In Section 5 the numerical schemes are particularized for the considered Saint-Venant-Exner models. In particular, the Roe matrices for the different models are presented. Finally, several numerical tests are shown in Section 6 to validate the methods and some conclusions are drawn.

2 Sediment transport models in shallow water

In this section, the equations of the hydrodynamical and the morphodynamical models are first presented. Next, the formulation of the coupled system in the form (1) is derived.

2.1 Hydrodynamical model: shallow-water system

Let us consider the one layer shallow-water system

$$\begin{cases} \frac{\partial h}{\partial t} + \frac{\partial q_1}{\partial x_1} + \frac{\partial q_2}{\partial x_2} = 0, \\ \frac{\partial q_1}{\partial t} + \frac{\partial}{\partial x_1} \left(\frac{q_1^2}{h} + \frac{1}{2}gh^2 \right) + \frac{\partial}{\partial x_2} \left(\frac{q_1 q_2}{h} \right) = gh \frac{\partial H}{\partial x_1} - S_{f,1}, \\ \frac{\partial q_2}{\partial t} + \frac{\partial}{\partial x_1} \left(\frac{q_1 q_2}{h} \right) + \frac{\partial}{\partial x_2} \left(\frac{q_2^2}{h} + \frac{1}{2}gh^2 \right) = gh \frac{\partial H}{\partial x_2} - S_{f,2}, \end{cases} \quad (2)$$

which are the equations governing the flow of a shallow layer of homogeneous inviscid fluid in a two dimensional domain $D \subset \mathbb{R}^2$. The points of D will be represented by $\mathbf{x} = (x_1, x_2)$. In the equations, $H(\mathbf{x}, t)$ represents the bottom depth measured from a fixed level of reference A_R . Notice that, due to the motion of the sediment layer, this function may depend on the time. $h(\mathbf{x}, t)$ represents the thickness of the layer; g , the gravity ; and

$$\mathbf{q}(\mathbf{x}, t) = (q_1(\mathbf{x}, t), q_2(\mathbf{x}, t)),$$

the mass-flow. These quantities are related to the vertical averaged velocity $\mathbf{u} = (u_1(\mathbf{x}, t), u_2(\mathbf{x}, t))$ by the relations:

$$q_j(\mathbf{x}, t) = u_j(\mathbf{x}, t)h(\mathbf{x}, t), \quad j = 1, 2.$$

$S_f = (S_{f,1}, S_{f,2})$ represents the bed friction forces which are modeled here by the Manning formula:

$$S_{f,1} = \frac{ghn^2 \|\mathbf{u}\| u_1}{h^{4/3}}, \quad S_{f,2} = \frac{ghn^2 \|\mathbf{u}\| u_2}{h^{4/3}}, \quad (3)$$

where n is the Manning coefficient. Finally the water surface elevation, which is denoted by η , is given by the formula $\eta = h - H$.

2.2 Morphodynamical model

Let us consider that the bottom is composed by a sediment layer whose thickness is given by $z_b(\mathbf{x}, t)$ laying on a non-erodible bottom whose depth measured from the level of reference is given by the function $\tilde{H}(\mathbf{x})$. Therefore the bottom depth function is given by the formula:

$$H(\mathbf{x}, t) = \tilde{H}(\mathbf{x}) - z_b(\mathbf{x}, t).$$

(See Figure 1). Notice that the thickness of the sediment layer may vanish in a part of the bottom.

The sediment layer motion due to bedload transport is modelled here by the formula:

$$\frac{\partial z_b}{\partial t} + \xi \frac{\partial q_{b,1}}{\partial x_1} + \xi \frac{\partial q_{b,2}}{\partial x_2} = 0. \quad (4)$$

where $\xi = 1/(1 - \rho_0)$ and ρ_0 is the porosity of the sediment layer. $\mathbf{q}_b = (q_{b,1}(h, \mathbf{q}), q_{b,2}(h, \mathbf{q}))$ represents the solid transport discharge, which is assumed to depend on the hydraulic variables.

Let us describe the formulae considered in this work for the solid transport discharge \mathbf{q}_b for granular and non-cohesive sediments. Even if these formulae are usually obtained for stationary flux in rivers, they can also be applied to tidal or coastal currents, as the time of response of the sediment is very small in comparison with the period of tides or waves. Notice that no pressure terms and nor deposition effects are included in the formulae listed below.

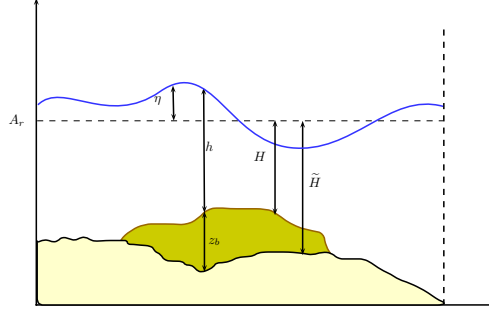


Figure 1: A shallow channel with a sediment layer.

1. grass model: Grass (see [14]) proposed the following formula for the solid transport discharge,

$$\begin{aligned} q_{b,1} &= A_g u_1 (u_1^2 + u_2^2), \\ q_{b,2} &= A_g u_2 (u_1^2 + u_2^2). \end{aligned} \quad (5)$$

The constant A_g (s^2/m), which is usually obtained from experimental data, takes into account the grain diameter and the kinematic viscosity. This coefficient takes values between 0 and 1: the closer to 0 the weaker the interaction between the sediment and the fluid. Notice that, according to Grass formula, the bedload sediment transport begins automatically when the fluid starts to move.

2. Meyer-Peter&Müller model: the formula proposed by Meyer-Peter&Müller in [25]) – MP&M in what follows –, which is among the most frequently used, is based on the median grain diameter d_{50} (see [33]). Chien in [7] proved that the original formula can be reduced to the following expression,

$$\begin{aligned} q_{b,1} &= 8\sqrt{(G-1)gd_i^3} \frac{u_1}{\sqrt{u_1^2 + u_2^2}} \max(\tau_* - \tau_{*,c}, 0)^{3/2}, \\ q_{b,2} &= 8\sqrt{(G-1)gd_i^3} \frac{u_2}{\sqrt{u_1^2 + u_2^2}} \max(\tau_* - \tau_{*,c}, 0)^{3/2}, \end{aligned} \quad (6)$$

where:

$$\tau_* = \frac{\gamma n^2 (u_1^2 + u_2^2)^{\frac{3}{2}}}{(\mathcal{P}_s - \mathcal{P}) d_i h^{1/3}}.$$

Here \mathcal{P} denotes the specific weight of the fluid, $\mathcal{P} = g\rho$, ρ being the water density; \mathcal{P}_s , the specific weight of the sediment, $\mathcal{P}_s = g\rho_s$, ρ_s being the sediment density; d_i , the sediment grain size (diameter); n , the Manning coefficient; G , the relative density, $G = \rho_s/\rho$; $\tau_{*,c}$, the non-dimensional critical shear stress, that for MP&M is equals to 0.047.

According to this formula, the motion of the granular sediment only begins when the non-dimensional shear stress τ_* is bigger that the non-dimensional critical shear stress $\tau_{*,c} = 0.047$.

This formula is usually applied to rivers and channels whose slope is below to 2%, $0.4 \leq d_i \leq 29$ mm, and $1.25 \leq G \leq 4.2$.

2.3 Coupled model

The expression of the complete system is as follows:

$$\left\{ \begin{array}{l} \frac{\partial h}{\partial t} + \frac{\partial q_1}{\partial x_1} + \frac{\partial q_2}{\partial x_2} = 0, \\ \frac{\partial q_1}{\partial t} + \frac{\partial}{\partial x_1} \left(\frac{q_1^2}{h} + \frac{1}{2}gh^2 \right) + \frac{\partial}{\partial x_2} \left(\frac{q_1 q_2}{h} \right) = -gh \frac{\partial z_b}{\partial x_1} + gh \frac{\partial \tilde{H}}{\partial x_1} - S_{f,1}, \\ \frac{\partial q_2}{\partial t} + \frac{\partial}{\partial x_1} \left(\frac{q_1 q_2}{h} \right) + \frac{\partial}{\partial x_2} \left(\frac{q_2^2}{h} + \frac{1}{2}gh^2 \right) = -gh \frac{\partial z_b}{\partial x_2} + gh \frac{\partial \tilde{H}}{\partial x_2} - S_{f,2}, \\ \frac{\partial z_b}{\partial t} + \xi \frac{\partial q_{b,1}}{\partial x_1} + \xi \frac{\partial q_{b,2}}{\partial x_2} = 0. \end{array} \right. \quad (7)$$

If the depth function H is used the system can be rewritten as follows:

$$\left\{ \begin{array}{l} \frac{\partial h}{\partial t} + \frac{\partial q_1}{\partial x_1} + \frac{\partial q_2}{\partial x_2} = 0, \\ \frac{\partial q_1}{\partial t} + \frac{\partial}{\partial x_1} \left(\frac{q_1^2}{h} + \frac{1}{2}gh^2 \right) + \frac{\partial}{\partial x_2} \left(\frac{q_1 q_2}{h} \right) = gh \frac{\partial H}{\partial x_1} - S_{f,1}, \\ \frac{\partial q_2}{\partial t} + \frac{\partial}{\partial x_1} \left(\frac{q_1 q_2}{h} \right) + \frac{\partial}{\partial x_2} \left(\frac{q_2^2}{h} + \frac{1}{2}gh^2 \right) = gh \frac{\partial H}{\partial x_2} - S_{f,2}, \\ \frac{\partial H}{\partial t} - \xi \frac{\partial q_{b,1}}{\partial x_1} - \xi \frac{\partial q_{b,2}}{\partial x_2} = 0, \end{array} \right. \quad (8)$$

or equivalently:

$$\frac{\partial W}{\partial t} + \frac{\partial}{\partial x_1} F_1(W) + \frac{\partial}{\partial x_2} F_2(W) = B_1(W) \frac{\partial W}{\partial x_1} + B_2(W) \frac{\partial W}{\partial x_2} + S_F, \quad (9)$$

where,

$$W = \begin{bmatrix} h \\ q_1 \\ q_2 \\ H \end{bmatrix}, \quad F_1 = \begin{bmatrix} q_1 \\ \frac{q_1^2}{h} + \frac{1}{2}gh^2 \\ \frac{q_1 q_2}{h} \\ -\xi q_{b,1} \end{bmatrix}, \quad F_2 = \begin{bmatrix} q_2 \\ \frac{q_1 q_2}{h} \\ \frac{q_2^2}{h} + \frac{1}{2}gh^2 \\ -\xi q_{b,2} \end{bmatrix},$$

$$B_1(W) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & gh \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad B_2(W) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & gh \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad S_F(W) = \begin{bmatrix} 0 \\ -S_{f,1} \\ -S_{f,2} \\ 0 \end{bmatrix}. \quad (10)$$

The friction term S_F will be discretized in a semi-implicit manner. For the sake of simplicity, in the presentation of the numerical schemes this term is supposed to vanish, so that the system reduces to:

$$\frac{\partial W}{\partial t} + \frac{\partial}{\partial x_1} F_1(W) + \frac{\partial}{\partial x_2} F_2(W) = B_1(W) \frac{\partial W}{\partial x_1} + B_2(W) \frac{\partial W}{\partial x_2}, \quad (11)$$

which can be also written in the nonconservative form (1):

$$\frac{\partial W}{\partial t} + \mathcal{A}_1(W) \frac{\partial W}{\partial x_1} + \mathcal{A}_2(W) \frac{\partial W}{\partial x_2} = 0. \quad (12)$$

where

$$\mathcal{A}_k(W) = J_k(W) - B_k(W), \quad k = 1, 2,$$

with:

$$J_1(W) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\frac{q_1^2}{h^2} + gh & 2\frac{q_1}{h} & 0 & 0 \\ -\frac{q_1 q_2}{h^2} & \frac{q_2}{h} & \frac{q_1}{h} & 0 \\ \frac{\partial q_{b,1}}{\partial h} & \frac{\partial q_{b,1}}{\partial q_1} & \frac{\partial q_{b,1}}{\partial q_2} & 0 \end{bmatrix}, \quad J_2(W) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ -\frac{q_1 q_2}{h^2} & \frac{q_2}{h} & \frac{q_1}{h} & 0 \\ -\frac{q_2^2}{h^2} + gh & 0 & 2\frac{q_2}{h} & 0 \\ \frac{\partial q_{b,2}}{\partial h} & \frac{\partial q_{b,2}}{\partial q_1} & \frac{\partial q_{b,2}}{\partial q_2} & 0 \end{bmatrix}.$$

3 High order finite volume schemes

3.1 Roe methods

Let us consider a 2D nonconservative hyperbolic system

$$W_t + \mathcal{A}_1(W)W_{x_1} + \mathcal{A}_2(W)W_{x_2} = 0, \quad \mathbf{x} = (x_1, x_2) \in D \subset \mathbb{R}^2, t \in (0, T), \quad (13)$$

where $W(\mathbf{x}, t)$ takes values on a convex domain Ω of \mathbb{R}^N and \mathcal{A}_i , $i = 1, 2$ are two smooth and locally bounded matrix-valued functions from Ω to $\mathcal{M}_{N \times N}(\mathbb{R})$.

Given an unitary vector $\eta = (\eta_1, \eta_2) \in \mathbb{R}^2$, we define the matrix

$$\mathcal{A}(W, \eta) = \mathcal{A}_1(W)\eta_1 + \mathcal{A}_2(W)\eta_2.$$

We assume that (13) is strictly hyperbolic, i.e. for all $W \in \Omega$ and $\forall \eta \in \mathbb{R}^2$, the matrix $\mathcal{A}(W, \eta)$ has N real and distinct eigenvalues

$$\lambda_1(W, \eta) < \dots < \lambda_N(W, \eta).$$

$\mathcal{A}(W, \eta)$ is thus diagonalizable:

$$\mathcal{A}(W, \eta) = \mathcal{K}(W, \eta) \cdot \Lambda(W, \eta) \cdot \mathcal{K}(W, \eta)^{-1},$$

where $\Lambda(W, \eta)$ is the diagonal matrix whose coefficients are the eigenvalues of $\mathcal{A}(W, \eta)$ and $\mathcal{K}(W, \eta)$ is a matrix whose j -th column is an eigenvector $R_j(W, \eta)$ associated to the eigenvalue $\lambda_j(W, \eta)$, $j = 1, \dots, N$.

In order to discretize (13), first the computational domain D is decomposed into subsets with a simple geometry, called cells or finite volumes, $V_i \subset \mathbb{R}^2$. We assume here that the cells are closed convex polygons whose intersections are either empty, a complete edge or a vertex. We will denote by \mathcal{T} the mesh, i.e. the set of cells, and by NV the number of cells.

The following notation is considered: given a finite volume V_i , $|V_i|$ represents its area; $N_i \in \mathbb{R}^2$, its center; \mathcal{N}_i , the set of indexes j such that V_j is a neighbor of V_i ; E_{ij} , the common edge to two neighbor cells V_i and V_j , and $|E_{ij}|$ its length; d_{ij} , the distance from N_i to E_{ij} ; $\eta_{ij} = (\eta_{ij,1}, \eta_{ij,2})$, the normal unit vector of the edge E_{ij} pointing towards the cell V_j (see Figure 2); Δ , the maximum of the diameters of the cells; and W_i^n , the constant approximation of the averaged solution in the cell V_i at time t^n provided by the numerical scheme:

$$W_i^n \cong \frac{1}{|V_i|} \int_{V_i} W(\mathbf{x}, t^n) dx.$$

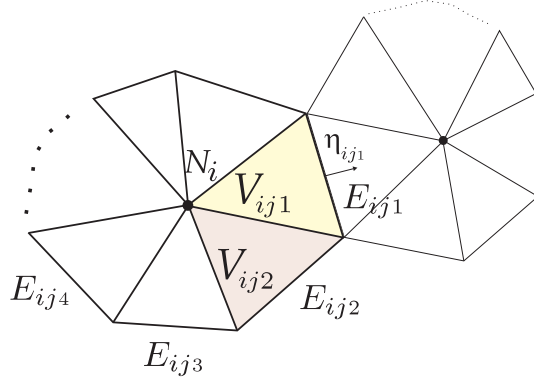


Figure 2: Finite Volumes.

Next, we consider a Roe linearization of (13) based on a *family of paths* in $\Omega \subset \mathbb{R}^N$, i.e., a locally Lipschitz map

$$\Psi: [0, 1] \times \Omega \times \Omega \times \mathcal{S}^1 \rightarrow \Omega.$$

where $\mathcal{S}^1 \subset \mathbb{R}^2$ denotes the unit sphere, that satisfies some natural properties (see [5] for details):

1. $\Psi(0; W_L, W_R, \eta) = W_L$ and $\Psi(1; W_L, W_R, \eta) = W_R$, for any $W_L, W_R \in \Omega$, $\eta \in \mathcal{S}^1$.
2. $\Psi(s; W_L, W_R, \eta) = \Psi(1 - s; W_R, W_L, -\eta)$, for any $W_L, W_R \in \Omega$, $s \in [0, 1]$, $\eta \in \mathcal{S}^1$.
3. $\Psi(s, W, W, \eta) = W$, for any $W \in \Omega$, $s \in [0, 1]$, $\eta \in \mathcal{S}^1$.

Once a family of paths has been chosen, a Roe linearization of the system (13) is considered, i.e., a function $\mathcal{A}_\Psi: \Omega \times \Omega \times \mathcal{S}^1 \rightarrow \mathcal{M}_{N \times N}(\mathbb{R})$ satisfying the following properties:

1. For each $W_L, W_R \in \Omega$ and $\eta \in \mathcal{S}^1$, $\mathcal{A}_\Psi(W_L, W_R, \eta)$ has N distinct real eigenvalues:

$$\lambda_1(W_L, W_R, \eta) < \lambda_2(W_L, W_R, \eta) < \dots < \lambda_N(W_L, W_R, \eta).$$

2. $\mathcal{A}_\Psi(W, W, \eta) = \mathcal{A}(W, \eta)$, for every $W \in \Omega$, $\eta \in \mathcal{S}^1$.
3. For any $W_L, W_R \in \Omega$, $\eta \in \mathcal{S}^1$:

$$\mathcal{A}_\Psi(W_L, W_R, \eta) \cdot (W_R - W_L) = \int_0^1 \mathcal{A}(\Psi(s; W_L, W_R, \eta), \eta) \frac{\partial \Psi}{\partial s}(s; W_L, W_R, \eta) ds. \quad (14)$$

Let us denote by $\Lambda_\Psi(W_L, W_R, \eta)$ the diagonal matrix whose coefficients are the eigenvalues $\lambda_1(W_L, W_R, \eta), \dots, \lambda_N(W_L, W_R, \eta)$ and $\mathcal{K}_\Psi(W_L, W_R, \eta)$ a $N \times N$ matrix whose columns are associated eigenvectors. The following notation will be used:

$$\begin{aligned} \mathcal{A}_\Psi^-(W_L, W_R, \eta) &= \mathcal{K}_\Psi(W_L, W_R, \eta) \cdot \Lambda_\Psi^-(W_L, W_R, \eta) \cdot \mathcal{K}_\Psi(W_L, W_R, \eta)^{-1}, \\ |\mathcal{A}_\Psi|(W_L, W_R, \eta) &= \mathcal{K}_\Psi(W_L, W_R, \eta) \cdot |\Lambda_\Psi|(W_L, W_R, \eta) \cdot \mathcal{K}_\Psi(W_L, W_R, \eta)^{-1}, \end{aligned}$$

where $\Lambda_\Psi^-(W_L, W_R, \eta)$ and $|\Lambda_\Psi|(W_L, W_R, \eta)$ are respectively the diagonal matrices whose coefficients are the negative part and the absolute value of the eigenvalues $\lambda_1(W_L, W_R, \eta), \dots, \lambda_N(W_L, W_R, \eta)$. The following identity holds:

$$\mathcal{A}_\Psi^- = \frac{1}{2}(\mathcal{A}_\Psi - |\mathcal{A}_\Psi|). \quad (15)$$

Notice that if $\mathcal{A}_k(W)$, $k = 1, 2$ are the Jacobian matrices of two smooth flux functions $F_k(W)$, $k = 1, 2$, (14) is independent of the family of paths and it reduces to the usual Roe property:

$$\mathcal{A}_\Psi(W_L, W_R, \eta) \cdot (W_R - W_L) = F_\eta(W_R) - F_\eta(W_L), \quad (16)$$

for any $\eta = (\eta_1, \eta_2) \in \mathcal{S}^1$, where

$$F_\eta(W) = \eta_1 F_1(W) + \eta_2 F_2(W) \quad (17)$$

represents the flux along the η direction.

The general expression of a Roe's scheme in upwind form for (13) is given by (see [5] for details):

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \mathcal{A}_{ij}^- \cdot (W_j^n - W_i^n), \quad (18)$$

where

$$\mathcal{A}_{ij}^- = \mathcal{A}_\Psi^-(W_i^n, W_j^n, \eta_{ij}).$$

A CLF condition has to be imposed to ensure the stability. We consider here the condition:

$$\max \left\{ \frac{|\lambda_{ij,k}|}{d_{ij}} : i = 1, \dots, NV, j \in \mathcal{N}_i, k = 1, \dots, N \right\} \cdot \Delta t = \delta, \quad (19)$$

with $0 < \delta \leq 1$.

As in the case of systems of conservation laws, when sonic rarefaction waves appear it is necessary to modify the numerical scheme in order to obtain entropy-satisfying solutions. The Harten-Hyman Entropy Fix technique (see [15]), for instance, can be easily adapted to this case.

In [5] some general results concerning the consistency and well-balanced properties of these Roe schemes have been presented.

3.2 Flux limiters for nonconservative systems

Taking into account (15), the scheme (18) can also be written in the form:

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \left(\frac{1}{2} \mathcal{A}_{ij} \cdot (W_j^n - W_i^n) - \frac{1}{2} |\mathcal{A}_{ij}| \cdot |W_j^n - W_i^n| \right), \quad (20)$$

where

$$|\mathcal{A}_{ij}| = |\mathcal{A}_\Psi|(W_i^n, W_j^n, \eta_{ij}).$$

This expression of the numerical scheme can be interpreted as a viscosity form: the first summand within the parenthesis in (20) corresponds to the centered part and the second one to the numerical viscosity.

We propose here to consider the more general family of schemes:

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \left(\frac{1}{2} \mathcal{A}_{ij} \cdot (W_j^n - W_i^n) - \frac{1}{2} \mathcal{Q}_{ij} \cdot (W_j^n - W_i^n) \right). \quad (21)$$

where

$$\mathcal{Q}_{ij} = \mathcal{Q}_\Psi(W_i^n, W_j^n, \eta_{ij}) \quad (22)$$

is a viscosity matrix.

In the particular case of Roe schemes

$$\mathcal{Q}_\Psi = |\mathcal{A}_\Psi| \quad (23)$$

but some other choices are possible. For instance, the choice

$$\mathcal{Q}_\Psi = \|A_\Psi\|_\infty \mathcal{I},$$

where \mathcal{I} is the identity matrix gives an extension of the usual Rusanov numerical method to nonconservative systems. But this choice of viscosity matrix increases the numerical diffusion and our goal is to reduce it. To do it, let us consider that the mesh is composed by closed convex polygons with an even number $2m$ of pairwise parallel edges (in the meshes considered here $m = 2$). We propose here the numerical scheme (21) with:

$$\mathcal{Q}_{ij} = |\mathcal{A}_{ij}| - \mathcal{K}_{ij} \cdot \mathcal{L}_{\phi,ij} \cdot \mathcal{K}_{ij}^{-1}, \quad (24)$$

where:

$$\mathcal{K}_{ij} = \mathcal{K}_\Psi(W_i^n, W_j^n, \eta_{ij})$$

and

$$\mathcal{L}_{\phi,ij} = \text{diag} \left\{ \left(\text{sgn}(\lambda_{ij,k}) - m \nu_{ij} \lambda_{ij,k} \right) \lambda_{ij,k} \phi(r_{ij,k}), k = 1, \dots, N \right\},$$

with $\lambda_{ij,k} = \lambda_k(W_i^n, W_j^n, \eta_{ij})$, $\nu_{ij} = \frac{\Delta t}{|E_{ij}|}$, ϕ is a flux-limiter function, and

$$r_{ij,k} = \begin{cases} \frac{[\mathcal{K}_{ij}^{-1}(W_j^* - W_j)]_k}{[\mathcal{K}_{ij}^{-1}(W_j - W_i)]_k} & \text{if } \text{sgn}(\lambda_{ij}) < 0, \\ \frac{[\mathcal{K}_{ij}^{-1}(W_i - W_i^*)]_k}{[\mathcal{K}_{ij}^{-1}(W_j - W_i)]_k} & \text{if } \text{sgn}(\lambda_{ij,k}) > 0, \\ 1 & \text{if } \lambda_{ij,k} = 0, \end{cases}$$

where,

$$\begin{aligned} W_i^* &= W_i - \nabla W_i \cdot \overrightarrow{N_i N_j}, \\ W_j^* &= W_j + \nabla W_j \cdot \overrightarrow{N_i N_j}. \end{aligned}$$

Here ∇W_i (respectively ∇W_j) represents an approximation of $\nabla W(\mathbf{x})$ in the finite volume V_i (respectively V_j). In practice, we use the approximation of gradients defined in Section 4.

Finally, an example of flux-limiter function is

$$\phi(r) = \max(0, \min(1, \beta r), \min(\beta, r)),$$

where if $\beta = 1$ we obtain the well-known *minmod* flux-limiter, while $\beta = 2$ corresponds to the *superbee* flux-limiter.

Remark 1 Notice that \mathcal{Q}_{ij} defined by (24) coincides with $|\mathcal{A}_{ij}|$ in regions where $W(x)$ is discontinuous while in regions where $W(x)$ is regular is close to $m \nu_{ij} \mathcal{A}_{ij}^2$, which is the viscosity matrix of a Lax-Wendroff type method. As a consequence, even if it is also first order, it is more accurate than a Roe scheme in regions where the solution is smooth.

Finally, concerning the stability of the numerical schemes (21) the following result can be proved for 2D linear systems ****(see [?])

Theorem 1 Let $\lambda_{ij,k}$ and $\lambda_{ij,k}^\mathcal{Q}$, $k = 1, \dots, N$ be the eigenvalues of \mathcal{A}_{ij} and \mathcal{Q}_{ij} , respectively. If \mathcal{Q}_{ij} and \mathcal{A}_{ij} have the same eigenvectors and their eigenvalues verify:

$$\nu_i = \frac{\Delta t |E_{ij}|}{|V_i|}, \quad m \nu_i^2 (\lambda_{ij,k})^2 \leq \nu_i \left(\lambda_{ij,k}^\mathcal{Q} \right) \leq \frac{1}{m}, \quad k = 1, \dots, N, \quad i \in \mathbb{Z}, \quad j \in \mathcal{N}_i, \quad (25)$$

then the numerical schemes (21) is L^2 stable.

It can be easily shown that the CFL condition (19) implies the L^2 linear stability of both schemes (20) and (21)-(24) for meshes composed by regular polygons of $2m$ edges.

3.3 Higher order extension

In order to extend to high order the schemes of the form (21) we consider a reconstruction operator, i.e. an operator that associates to a given family $\{W_i\}_{i=1}^{NV}$ of values at the cells two families of functions defined at the edges:

$$\gamma \in E_{ij} \rightarrow W_{ij}^\pm(\gamma),$$

in such a way that, whenever

$$W_i = \frac{1}{|V_i|} \int_{V_i} W(\mathbf{x}) d\mathbf{x} \quad (26)$$

for some smooth function W , then

$$W_{ij}^\pm(\gamma) = W(\gamma) + O(\Delta^p), \quad \forall \gamma \in E_{ij}.$$

We will assume that the reconstructions are calculated as follows: given the family $\{W_i\}_{i=1}^{NV}$ of values at the cells, first an approximation function is constructed at every cell V_i , based on the values of W_j at some of the cells close to V_i (the *stencil*):

$$P_i(\mathbf{x}) = P_i(\mathbf{x}; \{W_j\}_{j \in \mathcal{B}_i}),$$

for some set of indexes \mathcal{B}_i . If, for instance, the reconstruction only depends on the neighbor cells of V_i , then $\mathcal{B}_i = \mathcal{N}_i \cup \{i\}$. These approximation functions are calculated usually by means of an interpolation or approximation procedure. Once these functions have been constructed, the reconstructions at $\gamma \in E_{ij}$ are defined as follows:

$$W_{ij}^-(\gamma) = \lim_{\mathbf{x} \rightarrow \gamma} P_i(\mathbf{x}), \quad W_{ij}^+(\gamma) = \lim_{\mathbf{x} \rightarrow \gamma} P_j(\mathbf{x}). \quad (27)$$

Clearly, for any $\gamma \in E_{ij}$ the following equalities are satisfied:

$$W_{ij}^-(\gamma) = W_{ji}^+(\gamma); \quad W_{ij}^+(\gamma) = W_{ji}^-(\gamma).$$

We suppose that the reconstruction operator satisfies the following properties:

(HP1) It is conservative, i.e. the following equality holds for any cell V_i :

$$W_i = \frac{1}{|V_i|} \int_{V_i} P_i(\mathbf{x}) d\mathbf{x}. \quad (28)$$

(HP2) It is of order p , verifying $W(\gamma) - W_{ij}^\pm(\gamma) = \Delta^p g_{ij}(\gamma) + O(\Delta^{p+1})$, for any $\gamma \in E_{ij}$, being g_{ij} a regular function.

(HP3) It is of order q in the interior of the cells, i.e. if the operator is applied to a sequence $\{W_i\}$ satisfying (26) for some smooth function $W(\mathbf{x})$, then:

$$P_i(\mathbf{x}) = W(\mathbf{x}) + O(\Delta^q), \quad \forall \mathbf{x} \in \text{int}(V_i). \quad (29)$$

(HP4) The gradient of P_i provides an approximation of order m of the gradient of W :

$$\nabla P_i(\mathbf{x}) = \nabla W(\mathbf{x}) + O(\Delta^m), \quad \forall \mathbf{x} \in \text{int}(V_i). \quad (30)$$

Once the reconstruction operator has been chosen, the general expression of a semi-discrete scheme extending to higher order a first order scheme of the form (21) is the following:

$$\begin{aligned}
W_i'(t) &= -\frac{1}{|V_i|} \left[\sum_{j \in \mathcal{N}_i} \int_{E_{ij}} \frac{1}{2} (\mathcal{A}_{ij}(\gamma, t) - \mathcal{Q}_{ij}(\gamma, t)) \cdot (W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)) d\gamma \right. \\
&\quad \left. + \int_{V_i} \left(\mathcal{A}_1(P_i^t(x)) \frac{\partial P_i^t}{\partial x_1}(x) + \mathcal{A}_2(P_i^t(x)) \frac{\partial P_i^t}{\partial x_2}(x) \right) dx \right], \tag{31}
\end{aligned}$$

where P_i^t are the approximation functions corresponding to the cell values $W_i(t)$, i.e.

$$P_i^t(\mathbf{x}) = P_i(\mathbf{x}; \{W_j(t)\}_{j \in \mathcal{B}_i}),$$

$W_{ij}^\pm(\gamma, t)$ are given by

$$W_{ij}^-(\gamma, t) = \lim_{\mathbf{x} \rightarrow \gamma} P_i^t(\mathbf{x}), \quad W_{ij}^+(\gamma, t) = \lim_{\mathbf{x} \rightarrow \gamma} P_j^t(\mathbf{x}), \tag{32}$$

and

$$\begin{aligned}
\mathcal{A}_{ij}(\gamma, t) &= \mathcal{A}_\Psi(W_{ij}^-(\gamma, t), W_{ij}^+(\gamma, t), \eta_{ij}), \\
\mathcal{Q}_{ij}(\gamma, t) &= \mathcal{Q}_\Psi(W_{ij}^-(\gamma, t), W_{ij}^+(\gamma, t), \eta_{ij}).
\end{aligned}$$

The following result can be proved (see [5]):

Theorem 1 *Let us assume that \mathcal{A}_1 and \mathcal{A}_2 are of class \mathcal{C}^2 with bounded derivatives and \mathcal{A}_Ψ and \mathcal{Q}_Ψ are bounded. Let us also suppose that the reconstruction operator satisfies the hypothesis (HP1)-(HP4). Then (31) is an approximation of order at least $\alpha = \min(p, q, m)$.*

Remark 2 *This result is rather pessimistic: the order of the observed error is usually $\alpha = \min(p, q, m+1)$: see [5] for more details.*

In practice, the integral terms in (31) are numerically approached. In this case, a 1d quadrature formula of order \bar{r} has to be chosen to calculate the line integrals:

$$\int_a^b f(s) ds = (b-a) \left(\sum_{l=1}^{n(\bar{r})} \omega_l f(x_l) \right) + O(\Delta^{\bar{r}}), \tag{33}$$

where $n(\bar{r})$ denotes the number of points, ω_l are the weights, and $x_l = a + s_l(b-a)$ with $s_l \in [0, 1]$, represent the quadrature points. A quadrature formula of order \bar{s} is also needed to calculate the volume integrals:

$$\int_{V_i} f(\mathbf{x}) d\mathbf{x} = |V_i| \sum_{l=1}^{n(\bar{s})} \alpha_l f(\mathbf{x}_l^i) + \mathcal{O}(|V_i|^{\bar{s}}). \tag{34}$$

In order to preserve the order of the numerical scheme, it is necessary to have $\bar{r} \geq \alpha$ and $\bar{s} \geq \alpha$. The numerical scheme writes then as follows:

$$\begin{aligned}
W_i'(t) &= -\frac{1}{|V_i|} \left[\sum_{j \in \mathcal{N}_i} |E_{ij}| \sum_{l=1}^{n(\bar{r})} \frac{\omega_l}{2} (\mathcal{A}_{ij,l}(t) - \mathcal{Q}_{ij,l}(t)) \cdot (W_{ij,l}^+(t) - W_{ij,l}^-(t)) \right. \\
&\quad \left. + |V_i| \sum_{l=1}^{n(\bar{s})} \alpha_l \left(\mathcal{A}_1(P_i^t(\mathbf{x}_l^i)) \frac{\partial P_i^t}{\partial x_1}(\mathbf{x}_l^i) + \mathcal{A}_2(P_i^t(\mathbf{x}_l^i)) \frac{\partial P_i^t}{\partial x_2}(\mathbf{x}_l^i) \right) \right], \tag{35}
\end{aligned}$$

where

$$W_{ij,l}^\pm(t) = W_{ij}^\pm(a_{ij} + s_l(b_{ij} - a_{ij}), t),$$

and

$$\begin{aligned}\mathcal{A}_{ij,l}(t) &= \mathcal{A}_\Psi \left(W_{ij,l}^-(t), W_{ij,l}^+(t), \eta_{ij} \right), \\ \mathcal{Q}_{ij}(\gamma, t) &= \mathcal{Q}_\Psi \left(W_{ij,l}^-(t), W_{ij,l}^+(t), \eta_{ij} \right).\end{aligned}$$

Remark 3 *An interesting technique avoiding the explicit computation of $\nabla P_i(\mathbf{x})$ has been introduced in [27] making thus the expected order of accuracy equal to $\min(p, q)$. The extension to 2D problems of this technique, which is based on the use of the trapezoidal rule and Romberg extrapolation for the numerical integration is straightforward for structured meshes. For unstructured meshes a Romberg extrapolation formula for triangles could be used (see [41]).*

In [5] the well-balanced properties of the schemes (31) or (35) are analyzed.

3.4 Application to systems of the form (11)

The explicit calculation of Roe matrices for general nonconservative systems may be a difficult task. Nevertheless, when the system (13) comes from the reformulation of a system of the form (11) this calculation may if, given the family of paths Ψ , any unit vector η , and two states W_L, W_R , it is possible to obtain:

- A matrix $\mathcal{J}(W_L, W_R, \eta)$ such that:

$$\mathcal{J}(W_L, W_R, \eta)(W_R - W_L) = F_\eta(W_R) - F_\eta(W_L), \quad (36)$$

i.e. a Roe matrix in the usual sense for the flux function F_η .

- A matrix $B_\Psi(W_L, W_R, \eta)$ satisfying:

$$B_\Psi(W_L, W_R, \eta)(W_R - W_L) = \int_0^1 B(\Psi(s; W_L, W_R, \eta), \eta) \frac{\partial \Psi}{\partial s}(s; W_L, W_R, \eta) ds; \quad (37)$$

where $B(W, \eta) = B_1(W)\eta_1 + B_2(W)\eta_2$.

Then, it can be easily verified that the matrix:

$$\mathcal{A}_\Psi(W_L, W_R, \eta) = \mathcal{J}(W_L, W_R, \eta) - B_\Psi(W_L, W_R, \eta), \quad (38)$$

satisfies (14). Therefore, It is a Roe linearization provided that it has N real different eigenvalues (see [29] and [5]). In this case, using the divergence theorem, the semi-discrete numerical scheme (31) can be also rewritten as follows:

$$\begin{aligned}W_i'(t) &= -\frac{1}{|V_i|} \left[\sum_{j \in \mathcal{N}_i} \int_{E_{ij}} \frac{1}{2} (\mathcal{A}_{ij}(\gamma, t) - \mathcal{Q}_{ij}(\gamma, t)) \cdot (W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)) d\gamma \right. \\ &\quad \left. + \int_{E_{ij}} F_{\eta_{ij}}(W_{ij}^-(\gamma, t)) ds - \int_{V_i} \left(B_1(P_i^t(\mathbf{x})) \frac{\partial P_i^t}{\partial x_1}(\mathbf{x}) + B_2(P_i^t(\mathbf{x})) \frac{\partial P_i^t}{\partial x_2}(\mathbf{x}) \right) d\mathbf{x} \right].\end{aligned} \quad (39)$$

Taking into account the form of the Roe matrix and the Roe property (36) the numerical scheme can be rewritten in the form:

$$\begin{aligned}W_i'(t) &= -\frac{1}{|V_i|} \left[\int_{E_{ij}} \left(\mathcal{F}(W_{ij}^-(\gamma, t), W_{ij}^+(\gamma, t), \eta_{ij}) - \frac{1}{2} B_{ij}(\gamma, t) \cdot (W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)) \right) d\gamma \right. \\ &\quad \left. - \int_{V_i} \left(B_1(P_i^t(\mathbf{x})) \frac{\partial P_i^t}{\partial x_1}(\mathbf{x}) + B_2(P_i^t(\mathbf{x})) \frac{\partial P_i^t}{\partial x_2}(\mathbf{x}) \right) d\mathbf{x} \right],\end{aligned} \quad (40)$$

where

$$\begin{aligned}\mathcal{F}(W_{ij}^-(\gamma, t), W_{ij}^+(\gamma, t), \eta_{ij}) &= \frac{1}{2} (F_{\eta_{ij}}(W_{ij}^-(\gamma, t)) + F_{\eta_{ij}}(W_{ij}^+(\gamma, t))) \\ &\quad - \frac{1}{2} \mathcal{Q}_{ij}(\gamma, t)(W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)),\end{aligned}$$

and

$$B_{ij}(\gamma, t) = B_{\Psi}(W_{ij}^-(\gamma, t), W_{ij}^+(\gamma, t), \eta_{ij}).$$

Finally, if the integrals are numerically approached, the expression of the numerical scheme is the following:

$$\begin{aligned}W_i'(t) &= -\frac{1}{|V_i|} \left[\sum_{j \in \mathcal{N}_i} |E_{ij}| \sum_{l=1}^{n(\bar{r})} w_l \left(\mathcal{F}(W_{ij,l}^-(t), W_{ij,l}^+(t), \eta_{ij}) - \frac{1}{2} B_{ij,l}(t) \cdot (W_{ij,l}^+(t) - W_{ij,l}^-(t)) \right) \right. \\ &\quad \left. - |V_i| \sum_{l=1}^{n(\bar{s})} \alpha_l \left(B_1(P_i^t(\mathbf{x}_l^i)) \frac{\partial P_i^t}{\partial x_1}(\mathbf{x}_l^i) + B_2(P_i^t(\mathbf{x}_l^i)) \frac{\partial P_i^t}{\partial x_2}(\mathbf{x}_l^i) \right) \right],\end{aligned}\tag{41}$$

with the obvious notation.

For the time discretization, we consider high order TVD Runge-Kutta method like those described in [32]. In particular, in this work we use a second order MUSCL reconstruction operator in space and a second order TVD Runge-Kutta method. The resulting numerical scheme is thus second order accurate both in space and time.

Observe that if a first order reconstruction operator is chosen and the Euler scheme is applied to (40), the following Roe-type scheme is obtained:

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \left(\mathcal{F}(W_i^n, W_j^n, \eta_{ij}) - \frac{1}{2} B_{ij} \cdot (W_j^n - W_i^n) \right).\tag{42}$$

Before concluding this section, let us discuss the difficulties arising when a Roe linearization is not available. More precisely, let us suppose that, given two states W_L, W_R , it is not easy to obtain a matrix $\mathcal{J}(W_L, W_R, \eta)$ satisfying (36) (which is the case for some of the Saint-Venant-Exner models considered here) and/or a matrix $B_{\Psi}(W_L, W_R, \eta)$ satisfying (37). In this case, an intermediate matrix

$$\mathcal{A}_{ij}(\gamma, t) = \mathcal{J}_{ij}(\gamma, t) + B_{ij}(\gamma, t)$$

can always be calculated: for instance, it can be obtained by evaluating $\mathcal{A}(W, \eta_{ij})$ at some intermediate state $\widehat{W}_{ij}(\gamma, t)$ between $W_{ij}^-(\gamma, t)$ and $W_{ij}^+(\gamma, t)$. The viscosity matrix can be then obtained by applying (23) or (24) to this intermediate matrix. But if the matrix $\mathcal{J}(W_L, W_R, \eta)$ does not satisfy (36) the corresponding numerical scheme (39) has a serious drawback: if the system has a conservative subsystem (which is the case for Saint-Venant-Exner systems (7) in which the first and the fourth equations are pure conservation laws) the numerical scheme is not conservative for that subsystem. In particular, the numerical scheme is not conservative when it is applied to a conservative system, i.e. when B vanishes. Nevertheless, this drawback may be overcome if the numerical scheme is written in the form (40) (or (41)). In that case, the numerical scheme is conservative for any conservative subsystem and it reduces to a conservative scheme when B vanishes. Moreover, it has the same accuracy than those obtained on the basis of a Roe matrix and the linear stability is the same in both cases, as the two approaches coincide when the system is linear. Nevertheless, we have observed that, for nonlinear systems, a reduction of the CFL parameter can be necessary to guarantee the stability when the intermediate matrix is not a Roe matrix.

In order to explain the reason of this phenomenon, let us consider the particular case of the first order scheme (42). When this scheme is based on a Roe linearization it can be rewritten in the form:

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{|V_i|} \sum_{j \in \mathcal{N}_i} \frac{|E_{ij}|}{2} (\mathcal{A}_{ij} - \mathcal{Q}_{ij}) \cdot (W_j^n - W_i^n). \quad (43)$$

The matrices \mathcal{A}_{ij} and \mathcal{Q}_{ij} have the same eigenvectors and thus Theorem 1 can be locally applied to the scheme (provided that the mesh satisfies its hypothesis). Due to this, the *CFL* condition (25) is still valid for nonlinear systems. On the other hand, when the intermediate matrix \mathcal{A}_{ij} used to compute B_{ij} and \mathcal{Q}_{ij} is not a Roe matrix, the numerical scheme can still be written in a form similar to (43) but replacing \mathcal{A}_{ij} by:

$$\mathcal{A}_{ij}^* = \mathcal{J}_{ij}^* - B_{ij},$$

where \mathcal{J}_{ij}^* is any matrix satisfying:

$$\mathcal{J}_{ij}^* \cdot (W_j^n - W_i^n) = F_{\eta_{ij}}(W_j^n) - F_{\eta_{ij}}(W_i^n).$$

Such a matrix can be theoretically obtained by applying for instance the mean value theorem to every component. But Notice that in this case \mathcal{A}_{ij}^* and \mathcal{Q}_{ij} can have different eigenvectors (moreover, \mathcal{A}_{ij}^* could be not even diagonalizable). As a consequence, Theorem 1 cannot be applied and we have observed in practice that condition (25) may be not enough to guarantee the stability for nonlinear systems.

4 A second order reconstruction operator

The definition of reconstruction operators for 2D domains is in general a difficult task. In the case of structured meshes composed by rectangles whose edges are parallel to the axes, the problem usually reduces to consider a standard 1D reconstruction in both coordinates such as ENO [31], WENO ([19], [21]), Hyperbolic Reconstructions [23], etc. Nevertheless the development of a reconstruction operator on unstructured meshes is a more difficult problem, that may require a higher computational cost (see [17], for example). In this section we propose a MUSCL-type second order reconstruction operator for unstructured meshes of edge type (see Figure 3).

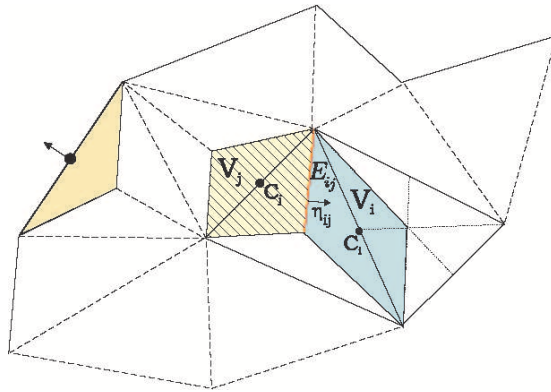


Figure 3: Finite volume of edge type.

Let us assume that the finite volume has been constructed on the basis of a Finite Element triangulation of the domain. Then, given an edge E_i of the triangulation, the vertices of the corresponding finite volume V_i are given by the extremes of E_i and the barycenters of the two triangles containing E_i (see Figure 3). The middle point of the edge E_i is represented by \mathbf{C}_i .

Let us consider the decomposition $V_i = V_{i1} \cup V_{i2} \cup V_{i3} \cup V_{i4}$, where V_{ij} , $j = 1, 2, 3, 4$, are the triangles defined by \mathbf{C}_i and the four edges of the finite volume V_i (see Figure 4). We also denote by b_{ij} , $j = 1, 2, 3, 4$, the corresponding barycenters of the triangles V_{ij} .

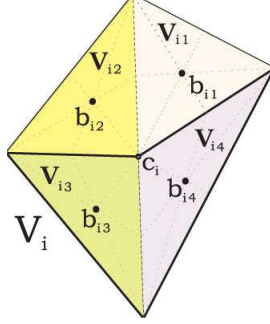


Figure 4: Subcells V_{ij} (triangles) on V_i .

Given a smooth function W , by applying the quadrature formula of the barycenter, we obtain:

$$\frac{1}{|V_{ij}|} \int_{V_{ij}} W(\mathbf{x}) d\mathbf{x} = W(b_{ij}) + O(\Delta^2), \quad j = 1, 2, 3, 4,$$

and thus the following equality holds:

$$\bar{W}_i = \frac{1}{|V_i|} \int_{V_i} W(\mathbf{x}) d\mathbf{x} = \sum_{j=1}^4 \frac{|V_{ij}|}{|V_i|} W(b_{ij}) + O(\Delta^2). \quad (44)$$

Moreover, using Taylor expansion, it can be easily verified that given a, b, c, d such that $a+b+c+d = 1$, on has:

$$aW(\mathbf{x}_1) + bW(\mathbf{x}_2) + cW(\mathbf{x}_3) + dW(\mathbf{x}_4) = W(a\mathbf{x}_1 + b\mathbf{x}_2 + c\mathbf{x}_3 + d\mathbf{x}_4) + O(\Delta^2).$$

Therefore, the following equality also holds:

$$\bar{W}_i = W(N_i) + O(\Delta^2), \quad (45)$$

where

$$N_i = \sum_{j=1}^4 \frac{|V_{ij}|}{|V_i|} b_{ij}. \quad (46)$$

In what follows we denote by N_i the point of V_i defined by (46), that is, the center of mass of the cell V_i . We will look for a linear approximation function in V_i of the form:

$$P_i(\mathbf{x}) = \bar{W}_i + \nabla W_i \cdot (\mathbf{x} - N_i)^T, \quad (47)$$

where ∇W_i represents a constant approximation of the gradient of $W(\mathbf{x})$ in V_i .

Remark 4 A reconstruction operator based on approximation functions P_i of the form (47) is conservative, i.e. it satisfies:

$$\frac{1}{|V_i|} \int_{V_i} P_i(\mathbf{x}) d\mathbf{x} = \bar{W}_i.$$

In effect, by using the definition of N_i (46), we obtain:

$$\begin{aligned} \frac{1}{|V_i|} \int_{V_i} \nabla W_i \cdot (\mathbf{x} - N_i)^T d\mathbf{x} &= \frac{\nabla W_i}{|V_i|} \left(\int_{V_{i1}} \mathbf{x} d\mathbf{x} + \int_{V_{i2}} \mathbf{x} d\mathbf{x} + \int_{V_{i3}} \mathbf{x} d\mathbf{x} + \int_{V_{i4}} \mathbf{x} d\mathbf{x} - |V_i| N_i \right) \\ &= \frac{\nabla W_i}{|V_i|} \left(\sum_{j=1}^4 |V_{ij}| b_{ij} - |V_i| N_i \right) = 0. \end{aligned}$$

In order to prove that the state reconstruction operator (47) provides a second order approximation, the following result is used:

Theorem 2 *Let $f, g \in C^p(\Omega)$, being $\Omega \subset \mathbb{R}^n$ an open convex set. If there exists a set of p points $\{\mathbf{x}_0, \dots, \mathbf{x}_{p-1}\} \subset \Omega$ such that $\nabla^k(f-g)(\mathbf{x}_k) = O(m(\Omega)^{p-k})$ then $(f-g)(\mathbf{x}) = O(m(\Omega)^p)$, where $m(\Omega)$ is the measure of Ω .*

Therefore, taking into account that $P_i(N_i) = \overline{W}_i = W(N_i) + O(\Delta^2)$, it is enough to obtain a first order approximation ∇W_i of $\nabla W(x)$ in V_i .

4.1 Gradient approximation

For the sake of simplicity, let us suppose that V_i is an interior cell and let us denote by $V_{i,1}, \dots, V_{i,4}$ its four neighbors. Let us consider the points $N_{i,j}$, $j = 1, \dots, 4$, associated to $V_{i,j}$, $j = 1, \dots, 4$, respectively, given by (46). Let us also consider the four triangles T_1, \dots, T_4 shown in Figure 5: the vertices of T_j , $j = 1, \dots, 4$, are $\{N_i, N_{i,j}, N_{i,ip(j)}\}$, where $ip(1) = 2$, $ip(2) = 3$, $ip(3) = 4$, and $ip(4) = 1$. We consider in T_j a linear approximation of the gradient $\nabla W|_{T_j}$ using the values $W_i, W_{i,1}, \dots, W_{i,4}$ which are second order approximation of $W(N_i), W(N_{i,1}), \dots, W(N_{i,4})$, respectively. This linear approximation is given by:

$$\nabla W|_{T_j} = W_i \nabla \lambda_j^0 + W_{i,j} \nabla \lambda_j^j + W_{i,ip(j)} \nabla \lambda_j^{ip(j)}, \quad (48)$$

where $\lambda_j^0, \lambda_j^j, \lambda_j^{ip(j)}$ are the barycentric coordinates associated to the vertices.

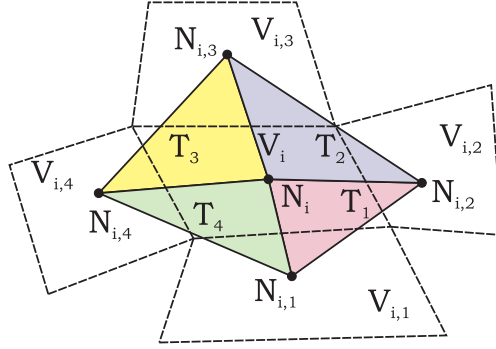


Figure 5: Triangles T_1, T_2, T_3, T_4 , used to approximate the gradient of $W(\mathbf{x})$ in V_i .

Finally, the following approximation of $\nabla W(\mathbf{x})$ in V_i is considered:

$$\nabla W(\mathbf{x})|_{V_i} \approx \nabla W_i = \frac{\sum_{j=1}^4 |T_j| \nabla W|_{T_j}}{\sum_{j=1}^4 |T_j|}. \quad (49)$$

The following result holds:

Theorem 2 For a given regular solution W , (49) is a first order approximation of the gradient of W in V_i .

The proof is detailed in Appendix A.

Observe that, due to the small size of the stencils and to the easy expression of the coefficients (49) used to approximate the gradients, the computational cost involved by the calculation of this reconstruction operator is low.

4.2 MUSCL reconstructions

As it is well known the solutions of nonlinear hyperbolic systems are likely to develop discontinuities even for very smooth initial conditions. It is thus necessary to modify the reconstruction operator (47) by introducing a slope limiter in order to avoid spurious oscillations near the discontinuities. We use the following notation:

$$W_i = \begin{bmatrix} W_{i,1} \\ W_{i,2} \\ \vdots \\ W_{i,N} \end{bmatrix}, \quad \nabla W_i = \begin{bmatrix} \nabla W_{i,1} \\ \nabla W_{i,2} \\ \vdots \\ \nabla W_{i,N} \end{bmatrix},$$

where ∇W_i is the approximation of the gradient given by (49),

We define here a modified reconstruction operator of the form:

$$P_i(\mathbf{x}) = W_i + \widetilde{\nabla W}_i \cdot (\mathbf{x} - N_i)^T, \quad (50)$$

where $\widetilde{\nabla W}_i$ is given by:

$$\widetilde{\nabla W}_i = \varphi_i : \nabla W_i = \begin{bmatrix} \varphi_{i,1} \nabla W_{i,1} \\ \varphi_{i,2} \nabla W_{i,2} \\ \vdots \\ \varphi_{i,N} \nabla W_{i,N} \end{bmatrix}. \quad (51)$$

where $\varphi_i = (\varphi_{i,1}, \dots, \varphi_{i,N})$ is a slope limiter function associated to the control volume V_i . An operator of this type is known as a MUSCL reconstruction. See [1] and [2] for some possible definitions of the limiter function.

Remark 5 In [38], the following historical remark can be found: "It has been pointed out to me by Dr. Vladimir Sabelnikov, formerly of TsAGI, the Central Aerodynamical National Laboratory near Moscow, that a scheme closely resembling MUSCL (including limiting) was developed in this laboratory by V. P. Kolgan (1972). Kolgan died young; his work apparently received little notice outside TsAGI."

Indeed, although the reference [37] is usually cited for MUSCL reconstructions, the paper of Kolgan [20] appeared seven years before. We found this remark in [12].

The slope limiter in [1] is defined by:

$$\varphi_{i,l} = \min_{j \in \mathcal{N}_i} \{\varphi_{ij,l}\}, \quad l = 1, \dots, N, \quad (52)$$

where

$$\varphi_{ij,l} = \max\{0, \min\{\beta r_{ij,l}, 1\}, \min\{r_{ij,l}, \beta\}\}, \quad \beta > 0. \quad (53)$$

Here, $r_{ij,l}$ is given by:

$$r_{ij,l} = \begin{cases} \frac{W_{i,l}^{max} - W_{i,l}}{W_{ij,l}^* - W_{i,l}} & \text{if } W_{ij,l}^* - W_{i,l} > 0, \\ \frac{W_{i,l}^{min} - W_{i,l}}{W_{ij,l}^* - W_{i,l}} & \text{if } W_{ij,l}^* - W_{i,l} < 0, \\ 1 & \text{if } W_{ij,l}^* = W_{i,l}, \end{cases} \quad l = 1, \dots, N, \quad (54)$$

where

$$W_{i,l}^{min} = \min\{W_{i,l}, \min_{j \in \mathcal{N}_i}\{W_{j,l}\}\}, \quad W_{i,l}^{max} = \max\{W_{i,l}, \max_{j \in \mathcal{N}_i}\{W_{j,l}\}\},$$

and

$$W_{ij}^* = W_i + \nabla W_i \cdot (\mathbf{c}_{ij} - N_i)^T,$$

where \mathbf{c}_{ij} is the middle point of the edge E_{ij} . In the numerical tests we use $\beta = 1$ that corresponds to the *minmod* limiter.

The slope limiter (53)-(54) is commonly used in MUSCL reconstructions for 2D systems and, in general, it provides good results. Nevertheless, in the numerical tests we have observed some oscillations near the shocks in the simulations obtained with the Grass model for medium or high interactions between the fluid and the sediment, or when the MP&M model is used. In order to avoid these oscillations, we introduce a new slope limiter using a greater number of slopes in its definition. The idea is to use the slope limiter which is usually considered in 1D problems along the normal direction at the middle point of every edge. This increment of the information used to define the limiter allows us in particular to improve the reconstruction when the sign of some of the derivatives changes.

The new limiter $\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,N})$ is defined as follows:

$$\alpha_{i,l} = \min_{j \in \mathcal{N}_i}\{\alpha_{ij,l}\}, \quad l = 1, \dots, N \quad (55)$$

where

$$\alpha_{ij,l} = \min\{\alpha_{ij,l}^1, \alpha_{ij,l}^2\}, \quad (56)$$

with

$$\begin{aligned} \alpha_{ij,l}^1 &= \max(0, \min(1, r_{ij,l}^1)), \\ \alpha_{ij,l}^2 &= \max(0, \min(1, r_{ij,l}^2)), \end{aligned}$$

and $r_{ij,l}^1, r_{ij,l}^2, l = 1, \dots, N$ are defined by,

$$r_{ij,l}^1 = \begin{cases} \frac{W_{i,l} - W_{i,l}^*}{W_{j,l} - W_{i,l}} & \text{if } W_{j,l} - W_{i,l} \neq 0, \\ 1 & \text{if } W_{j,l} - W_{i,l} = 0, \end{cases}$$

and

$$r_{ij,l}^2 = \begin{cases} \frac{W_{j,l}^* - W_{j,l}}{W_{j,l} - W_{i,l}} & \text{if } W_{j,l} - W_{i,l} \neq 0, \\ 1 & \text{if } W_{j,l} - W_{i,l} = 0; \end{cases}$$

being,

$$\begin{aligned} W_i^* &= W_i - \nabla W_i \cdot \overrightarrow{N_i N_j}, \\ W_j^* &= W_j + \nabla W_j \cdot \overrightarrow{N_i N_j}. \end{aligned}$$

Notice that $\alpha_{ij,l} = \alpha_{ji,l}$.

Remark 6 Observe that in the definition of $r_{ij,l}^1$ and $r_{ij,l}^2$, $l = 1, \dots, N$ three approximations of the directional derivative along $\overrightarrow{N_i N_j}$ are used:

$$\frac{W_j - W_i}{d(N_i, N_j)}, \quad \nabla W_i \frac{N_j - N_i}{d(N_i, N_j)} \quad \text{and} \quad \nabla W_j \frac{N_j - N_i}{(d_{N_i, N_j})}.$$

Even if this new slope limiter gets rid of the oscillations mentioned above, it is too diffusive in practice. Therefore we propose to consider a linear combination of both limiters, so that the approximation functions used in practice are as follows;

$$P_i(\mathbf{x}) = W_i + \widehat{\nabla W_i} \cdot (\mathbf{x} - N_i)^T, \quad (57)$$

where

$$\widehat{\nabla W_i} = \vartheta_i : \nabla W_i, \quad (58)$$

and $\vartheta_i = (\vartheta_{i,1}, \dots, \vartheta_{i,N})$ is given by

$$\vartheta_{i,l} = \min_{j \in N_i} \{\vartheta_{ij,l}\}, \quad l = 1, \dots, N, \quad (59)$$

being

$$\vartheta_{ij,l} = \theta \varphi_{ij,l} + (1 - \theta) \alpha_{ij,l}, \quad l = 1, \dots, N. \quad (60)$$

The reconstruction operator (50) is obtained if $\theta = 1$. In practice the choice $\theta = \frac{1}{2}$ gives good results. In Section 6.1 a comparison between the reconstruction operators (57) corresponding to $\theta = 1$ and $\theta = \frac{1}{2}$ is presented.

5 Application to Saint-Venant-Exner models

5.1 Roe matrices

Given a unit vector η , the expressions of W , F_η , and $B(W, \eta)$ for the particular case of Saint-Venant-Exner models (8) are the following:

$$W = \begin{bmatrix} U \\ H \end{bmatrix}, \quad U = \begin{bmatrix} h \\ q_1 \\ q_2 \end{bmatrix},$$

$$F_\eta(W) = \begin{bmatrix} F_\eta^{sw}(U) \\ F_\eta^b(U) \end{bmatrix}, \quad F_\eta^{sw}(U) = \begin{bmatrix} \mathbf{q} \cdot \eta \\ \frac{q_1}{h} \mathbf{q} \cdot \eta + \frac{1}{2} g h^2 \eta_1 \\ \frac{q_2}{h} \mathbf{q} \cdot \eta + \frac{1}{2} g h^2 \eta_2 \end{bmatrix}, \quad F_\eta^b(U) = -\xi \mathbf{q}_b \cdot \eta, \quad (61)$$

$$B(W, \eta) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & g h \eta_1 \\ 0 & 0 & 0 & g h \eta_2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

In order to construct a Roe linearization, first a family of paths has to be chosen. Here, we consider the family of straight segments:

$$\Psi(s; W_L, W_R, \eta) = s W_R + (1 - s) W_L, \quad s \in [0, 1].$$

Next, following the indications in Section 3.4, given two states W_L and W_R whose components are denoted by:

$$W_L = \begin{bmatrix} U_L \\ H_l \end{bmatrix}, \quad U_L = \begin{bmatrix} h_l \\ q_{1,l} \\ q_{2,l} \end{bmatrix}; \quad W_R = \begin{bmatrix} U_R \\ H_r \end{bmatrix}, \quad U_R = \begin{bmatrix} h_r \\ q_{1,r} \\ q_{2,r} \end{bmatrix};$$

we have to look for two matrices $\mathcal{J}(W_L, W_R, \eta)$ and $B_\Psi(W_L, W_R, \eta)$ satisfying (36) and (37) respectively. Due to the linear expression of the coefficients of B and of the family of paths, this latter matrix can be easily obtained:

$$B_\Psi(W_L, W_R, \eta) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & g\tilde{h}\eta_1 \\ 0 & 0 & 0 & g\tilde{h}\eta_2 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (62)$$

where

$$\tilde{h} = \frac{h_r + h_l}{2}.$$

On the other hand a Roe matrix is available for the flux function corresponding to the shallow water model F_η^{sw} :

$$\mathcal{J}^{sw}(U_L, U_R, \eta) = \begin{bmatrix} 0 & \eta_1 & \eta_2 \\ -\tilde{u}_1^2\eta_1 + \tilde{c}^2\eta_1 - \tilde{u}_1\tilde{u}_2\eta_2 & 2\tilde{u}_1\eta_1 + \tilde{u}_2\eta_2 & \tilde{u}_1\eta_2 \\ -\tilde{u}_2^2\eta_2 + \tilde{c}^2\eta_2 - \tilde{u}_1\tilde{u}_2\eta_1 & \tilde{u}_2\eta_1 & \tilde{u}_1\eta_1 + 2\tilde{u}_2\eta_2 \end{bmatrix} \quad (63)$$

where,

$$\tilde{c} = \sqrt{g\tilde{h}},$$

$$\tilde{u}_i = \frac{\sqrt{h_r}u_{i,r} + \sqrt{h_l}u_{i,l}}{\sqrt{h_r} + \sqrt{h_l}}, \quad u_{i,l} = \frac{q_{i,l}}{h}, \quad u_{i,r} = \frac{q_{i,r}}{h}, \quad i = 1, 2.$$

If we find a 1×3 vector $J^b(U_L, U_R, \eta)$ such that:

$$J^b(U_L, U_R, \eta) \cdot (U_R - U_L) = F_\eta^b(U_R) - F_\eta^b(U_L), \quad (64)$$

then the 4×4 matrix whose block structure is given by:

$$\mathcal{J}(W_L, W_R, \eta) = \left[\begin{array}{c|c} \mathcal{J}^{sw}(U_L, U_R, \eta) & \begin{matrix} 0 \\ 0 \end{matrix} \\ \hline J^b(U_L, U_R, \eta) & 0 \end{array} \right] \quad (65)$$

trivially satisfies (36). As a consequence, the intermediate matrix:

$$A_\Psi(W_L, W_R, \eta) = \mathcal{J}(W_L, W_R, \eta) - B_\Psi(W_L, W_R, \eta) \quad (66)$$

satisfies (14). It is thus a Roe matrix if it has N different real eigenvalues.

Even if it is always possible to calculate a vector $J^b(U_L, U_R, \eta)$ satisfying (64) (by applying for instance the mean value theorem), its explicit calculation may be difficult due to the complex expression of the solid transport formulae. In the particular case of the Grass model, it can explicitly calculated:

$$J^b(U_L, U_R, \eta) = [\tilde{u}_1\tilde{f} + \tilde{u}_2\tilde{i}, -\tilde{f}, -\tilde{i}] \quad (67)$$

with

$$\begin{aligned}
\tilde{f} &= \frac{A_g \xi (\sqrt{h_r} + \sqrt{h_l}) \left(u_{1,r}^2 + u_{1,r} u_{1,l} + u_{1,l}^2 + \tilde{u}_2^2 \right) \eta_1}{h_r \sqrt{h_l} + h_l \sqrt{h_r}} \\
&+ \frac{A_g \xi \left(2\sqrt{h_l} \sqrt{h_r} (u_{1,r} u_{2,r} + u_{1,l} u_{2,l}) + (h_l u_{2,r} + h_r u_{2,l}) (u_{1,r} + u_{1,l}) \right) \eta_2}{(h_r \sqrt{h_l} + h_l \sqrt{h_r}) (\sqrt{h_l} + \sqrt{h_r})} \\
\tilde{i} &= \frac{A_g \xi (\sqrt{h_r} \sqrt{h_l}) \left(u_{2,r}^2 + u_{2,r} u_{2,l} + u_{2,l}^2 + \tilde{u}_1^2 \right) \eta_2}{h_r \sqrt{h_l} + h_l \sqrt{h_r}} \\
&+ \frac{A_g \xi \left(2\sqrt{h_l} \sqrt{h_r} (u_{2,r} u_{1,r} + u_{2,l} u_{1,l}) + (h_l u_{1,r} + h_r u_{1,l}) (u_{2,r} + u_{2,l}) \right) \eta_1}{(h_r \sqrt{h_l} + h_l \sqrt{h_r}) (\sqrt{h_l} + \sqrt{h_r})}.
\end{aligned}$$

The expression of corresponding Roe matrix (66) is:

$$\mathcal{A}_{\Psi}^G(W_L, W_R, \eta) = \begin{bmatrix} 0 & \eta_1 & \eta_2 & 0 \\ -\tilde{u}_1^2 \eta_1 + \tilde{c}^2 \eta_1 - \tilde{u}_1 \tilde{u}_2 \eta_2 & 2\tilde{u}_1 \eta_1 + \tilde{u}_2 \eta_2 & \tilde{u}_1 \eta_2 & -\tilde{c}^2 \eta_1 \\ -\tilde{u}_2^2 \eta_2 + \tilde{c}^2 \eta_2 - \tilde{u}_1 \tilde{u}_2 \eta_1 & \tilde{u}_2 \eta_1 & \tilde{u}_1 \eta_1 + 2\tilde{u}_2 \eta_2 & -\tilde{c}^2 \eta_2 \\ \tilde{u}_1 \tilde{f} + \tilde{u}_2 \tilde{i} & -\tilde{f} & -\tilde{i} & 0 \end{bmatrix}. \quad (68)$$

In the particular case of $\eta = [0, 1]^t$ or $\eta = [1, 0]^t$, this matrix coincides with that obtained by Hudson in [18].

If the MP&M model is used, even if it is possible to obtain the explicit expression of a vector $J^b(U_L, U_R, \eta)$ satisfying (64), due to its very complex expression, it is not easy to check that the corresponding intermediate matrix has N different real eigenvalues and thus it may be not a Roe matrix. Instead, the following approximation is proposed here:

$$J^b(U_L, U_R, \eta) = -\xi \left[\frac{\partial \mathbf{q}_b}{\partial h}(\tilde{h}, \tilde{\mathbf{q}}) \cdot \eta, \frac{\partial \mathbf{q}_b}{\partial q_1}(\tilde{h}, \tilde{\mathbf{q}}) \cdot \eta, \frac{\partial \mathbf{q}_b}{\partial q_2}(\tilde{h}, \tilde{\mathbf{q}}) \cdot \eta \right] \quad (69)$$

where

$$\tilde{\mathbf{q}} = (\tilde{h} \tilde{u}_1, \tilde{h} \tilde{u}_2).$$

The expression of the corresponding intermediate matrix is the following:

$$\mathcal{A}_{\Psi}^{MPM}(W_R, W_R, \eta) = \begin{bmatrix} 0 & \eta_1 & \eta_2 & 0 \\ -\tilde{u}_1^2 \eta_1 + \tilde{c}^2 \eta_1 - \tilde{u}_1 \tilde{u}_2 \eta_2 & 2\tilde{u}_1 \eta_1 + \tilde{u}_2 \eta_2 & \tilde{u}_1 \eta_2 & -\tilde{c}^2 \eta_1 \\ -\tilde{u}_2^2 \eta_2 + \tilde{c}^2 \eta_2 - \tilde{u}_1 \tilde{u}_2 \eta_1 & \tilde{u}_2 \eta_1 & \tilde{u}_1 \eta_1 + 2\tilde{u}_2 \eta_2 & -\tilde{c}^2 \eta_2 \\ -\xi \frac{\partial q_b}{\partial h}(\tilde{h}, \tilde{u}) & -\xi \frac{\partial q_b}{\partial q_1}(\tilde{h}, \tilde{u}) & -\xi \frac{\partial q_b}{\partial q_2}(\tilde{h}, \tilde{u}) & 0 \end{bmatrix}. \quad (70)$$

This is not a Roe matrix, but according to the discussion at the end of Section 3.4, it allows us to obtain a numerical scheme of the form (40) or (41). Moreover, the possible stability restrictions discussed in that Section have not been noticed in this particular case.

Let us briefly discuss the eigenstructure of the matrices (68) and (70) (see [18] for more details). Observe that only the entries (4,1), (4,2), (4,3) of both matrices are different. Let us consider a general 4×4 matrix \mathcal{A} such that all its entries $\mathcal{A}_{i,j}$ with $i \neq 4$ or $j = 4$ coincide with both matrices. The eigenvalues of such a matrix are

$$\lambda_1 = \tilde{u}_1 \eta_1 + \tilde{u}_2 \eta_2,$$

and the roots $\lambda_2, \lambda_3, \lambda_4$ of the following polynomial:

$$P(\lambda) = \lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 \quad (71)$$

which coefficients are given by:

$$a_1 = -2(\tilde{u}_1\eta_1 + \tilde{u}_2\eta_2), \quad a_2 = (\tilde{u}_1\eta_1 + \tilde{u}_2\eta_2)^2 - \tilde{c}^2(1 + \eta_2\mathcal{A}_{4,3} + \eta_1\mathcal{A}_{4,2}),$$

$$a_3 = \tilde{c}^2 \left(-\mathcal{A}_{4,1} + \mathcal{A}_{4,2}(-\tilde{u}_1\eta_2^2 + \tilde{u}_2\eta_1\eta_2) + \mathcal{A}_{4,3}(-\tilde{u}_2\eta_1^2 + \tilde{u}_1\eta_1\eta_2) \right).$$

Let us define $Q = (3a_2 - a_1^2)/9$, $R = (9a_1a_2 - 27a_3 - 2a_1^3)/54$. If $Q^3 + R^2 < 0$ all the roots are real and they can be obtained by the the Cardano-Vieta formula:

$$\lambda_2 = 2\sqrt{-Q} \cos(\theta/3) - a_1/3, \quad (72)$$

$$\lambda_3 = 2\sqrt{-Q} \cos((\theta + 2\pi)/3) - a_1/3, \quad (73)$$

$$\lambda_4 = 2\sqrt{-Q} \cos((\theta - 2\pi)/3) - a_1/3, \quad (74)$$

where $\theta = \arccos(R/\sqrt{-Q^3})$. Moreover, these three eigenvalues are always different.

Concerning the eigenvectors, if λ_1 is simple, an associated eigenvector is:

$$R_1 = \begin{bmatrix} 1 \\ \frac{\tilde{u}_1\eta_1\mathcal{A}_{4,3} + \eta_2(\tilde{u}_2\mathcal{A}_{4,3} + \tilde{u}_1\eta_1 + \tilde{u}_2\eta_2 + \mathcal{A}_{4,1})}{\eta_1\mathcal{A}_{4,3} - \eta_2\mathcal{A}_{4,2}} \\ \frac{-\tilde{u}_2\eta_2\mathcal{A}_{4,2} - \eta_1(\tilde{u}_1\mathcal{A}_{4,2} + \tilde{u}_2\eta_2 + \tilde{u}_1\eta_1 + \mathcal{A}_{4,1})}{\eta_1\mathcal{A}_{4,3} - \eta_2\mathcal{A}_{4,2}} \\ 1 \end{bmatrix};$$

and otherwise:

$$R_1 = \begin{pmatrix} 0 \\ 0 \\ -\eta_2 \\ \eta_1 \end{pmatrix}.$$

The eigenvectors R_i , $i = 2, 3, 4$ associated to λ_i , $i = 2, 3, 4$ are defined by

$$R_i = \begin{pmatrix} 1 \\ \lambda_i\eta_1 + \tilde{u}_1\eta_2^2 - \tilde{u}_2\eta_1\eta_2 \\ \lambda_i\eta_2 + \tilde{u}_2\eta_1^2 - \tilde{u}_1\eta_1\eta_2 \\ 1 - \frac{(\tilde{u}_1\eta_1 + \tilde{u}_2\eta_2 - \lambda_i)^2}{\tilde{c}^2} \end{pmatrix}, \quad i = 2, 3, 4.$$

While in the case of the matrix (68) it can be proved that the condition $Q^3 + R^2 < 0$ is always satisfied provided that $\tilde{h} > 0$, we have not been able to check it for the matrix (70) due to its complex expression. Nevertheless, in all the numerical tests performed the expressions (72)-(74) gave always three different real eigenvalues.

5.2 Approximation of the integrals.

The following criteria must be followed to choose the quadrature formulae used in numerical schemes (35) or (41):

- a) The order of the quadrature formulae must be greater than that of the reconstruction operator to preserve the accuracy of the scheme.
- b) The numerical scheme has to be well-balanced for the stationary solutions corresponding to water and sediments at rest.

The third order Gauss formula for the line integrals and the barycenter quadrature formula for the volume integrals satisfy both criteria.

5.3 Implementation of the numerical schemes

In this section we clarify how the numerical schemes are implemented. Let us suppose that the approximations at time t^k , W_i^k and Δt^k have been yet calculated. To advance in time, let us consider, for example, the numerical scheme resulting to combine the second order scheme (41) based on a Roe scheme with a second order TVD Runge-Kutta method:

$$\begin{aligned} W_i^{k+1/2} &= W_i^k - \frac{\Delta t^k}{|V_i|} \left[\sum_{j \in \mathcal{N}_i} |E_{ij}| \sum_{l=1}^{n(\bar{r})} w_l \left(\mathcal{F}(W_{ij,l}^{k,-}, W_{ij,l}^{k,+}, \eta_{ij}) - \frac{1}{2} B_{ij,l}^k \cdot (W_{ij,l}^{k,+} - W_{ij,l}^{k,-}) \right) \right. \\ &\quad \left. - |V_i| \sum_{l=1}^{n(\bar{s})} \alpha_l \left(B_1(P_i^k(\mathbf{x}_l^i)) \frac{\partial P_i^k}{\partial x_1}(\mathbf{x}_l^i) + B_2(P_i^k(\mathbf{x}_l^i)) \frac{\partial P_i^k}{\partial x_2}(\mathbf{x}_l^i) \right) \right], \end{aligned}$$

and

$$\begin{aligned} W_i^{k+1} &= \frac{W_i^k + W_i^{k+1/2}}{2} - \frac{1}{2} \frac{\Delta t^k}{|V_i|} \left[\sum_{j \in \mathcal{N}_i} |E_{ij}| \sum_{l=1}^{n(\bar{r})} w_l \left(\mathcal{F}(W_{ij,l}^{k+1/2,-}, W_{ij,l}^{k+1/2,+}, \eta_{ij}) \right. \right. \\ &\quad \left. \left. - \frac{1}{2} B_{ij,l}^{k+1/2} \cdot (W_{ij,l}^{k+1/2,+} - W_{ij,l}^{k+1/2,-}) \right) \right. \\ &\quad \left. - |V_i| \sum_{l=1}^{n(\bar{s})} \alpha_l \left(B_1(P_i^{k+1/2}(\mathbf{x}_l^i)) \frac{\partial P_i^{k+1/2}}{\partial x_1}(\mathbf{x}_l^i) + B_2(P_i^{k+1/2}(\mathbf{x}_l^i)) \frac{\partial P_i^{k+1/2}}{\partial x_2}(\mathbf{x}_l^i) \right) \right], \end{aligned}$$

where $P_i^\alpha(\mathbf{x})$, $\mathbf{x} \in V_i$, $\alpha = k, k + 1/2$, $i = 1, \dots, NV$, is the reconstruction operator defined by (57); $W_{ij,l}^{\alpha,\pm}$ are defined by

$$W_{ij,l}^{\alpha,-} = \lim_{\mathbf{x} \rightarrow \mathbf{x}_{ij,l}} P_i^\alpha(\mathbf{x}), \quad W_{ij,l}^{\alpha,+} = \lim_{\mathbf{x} \rightarrow \mathbf{x}_{ij,l}} P_j^\alpha(\mathbf{x}), \quad l = 1, \dots, n(\bar{r}),$$

where $\mathbf{x}_{ij,l}$ are the quadrature points over the common edge E_{ij} to the volumes V_i and V_j respectively;

$$\begin{aligned} \mathcal{F}(W_{ij,l}^{\alpha,-}, W_{ij,l}^{\alpha,+}, \eta_{ij}) &= \frac{1}{2} \left(F_{\eta_{ij}}(W_{ij,l}^{\alpha,-}) + F_{\eta_{ij}}(W_{ij,l}^{\alpha,+}) \right) \\ &\quad - \frac{1}{2} \mathcal{Q}_{ij,l}^\alpha (W_{ij,l}^{\alpha,+} - W_{ij,l}^{\alpha,-}) \end{aligned}$$

where $F_{\eta_{ij}}(\cdot)$ is defined by (61) and $\mathcal{Q}_{ij,l}^\alpha = |\mathcal{A}_{ij,l}^\alpha|$, being $\mathcal{A}_{ij,l}^\alpha = \mathcal{A}_\Psi^\beta(W_{ij,l}^{\alpha,-}, W_{ij,l}^{\alpha,+}, \eta_{ij})$, $\beta = G, MPM$, given by (68) and (70), respectively; $B_{ij,l}^\alpha = B_\Psi(W_{ij,l}^{\alpha,-}, W_{ij,l}^{\alpha,+}, \eta_{ij})$ given by (62); finally $B_1(\cdot)$ and $B_2(\cdot)$ are given by (10).

Concerning the practical implementation, let us say that in order to obtain W_i^α , $\alpha = k, k + 1/2$, the following procedure is followed:

- For each volume V_i , the reconstruction operator $P_i^\alpha(\mathbf{x})$ is computed and stored. Next the term

$$|V_i| \sum_{l=1}^{n(\bar{s})} \alpha_l \left(B_1(P_i^\alpha(\mathbf{x}_l^i)) \frac{\partial P_i^\alpha}{\partial x_1}(\mathbf{x}_l^i) + B_2(P_i^\alpha(\mathbf{x}_l^i)) \frac{\partial P_i^\alpha}{\partial x_2}(\mathbf{x}_l^i) \right)$$

is computed.

- For each edge E_{ij} of the finite volume mesh, the term

$$\sum_{l=1}^{n(\bar{r})} w_l \left(\mathcal{F}(W_{ij,l}^{\alpha,-}, W_{ij,l}^{\alpha,+}, \eta_{ij}) - \frac{1}{2} B_{ij,l}^\alpha \cdot (W_{ij,l}^{\alpha,+} - W_{ij,l}^{\alpha,-}) \right)$$

is computed and stored into the cells V_i and V_j respectively, taking into account that η_{ji} is the normal unit vector of the edge E_{ij} pointing towards the cell V_i and $\eta_{ji} = -\eta_{ij}$. Note that $W_{ij,l}^{\alpha,\pm}$ are defined evaluating the reconstruction operator P_i^α (respectively P_j^α) previously computed at the quadrature points $\mathbf{x}_{ij,l}$.

- Finally, all the contributions are added at each cell and the final expression W_i^α is computed using the previous formulae. Δt^{k+1} is also estimated if $\alpha = k + 1/2$.

6 Numerical experiments

In this section three numerical tests are presented. The first one was also used in [4] to validate the 1d versions of the numerical schemes presented here. The goal of this test is to validate the ability of the 2D models to capture solutions which are essentially 1D when using unstructured meshes: for structured meshes the 2D numerical schemes presented here reduce to their 1d counterparts introduced and tested in [4]. The second one is a purely 2D test for which only an analytical approach of the spreading angle of the sediment layer is available. In this case the second order accuracy of the numerical schemes plays a fundamental role: as it is a large time simulation, first order numerical schemes are not able to correctly capture the spread angle due to the excess of numerical diffusion. Finally, we present a simulation of the sediment layer evolution in an L-shaped channel whose non-erodible bottom is flat. This is a hard test for the numerical schemes as the thickness of the sediment layer vanishes in parts of the domain. We have designed this test by its similarity with a lot of real situations in which bends in rivers or channels play an important role in the sediment transport. The right-angle bend introduces some extra difficulties to the test.

In [4] the numerical results provided by a 1D second order finite volume method have been compared with some experimental data corresponding to an 1D experiment performed by our collaborators of the 'Escuela Superior de Ingenieros de Caminos, Canales y Puertos'(A Coruña University) . We have also compared the 2D numerical solutions with these data. As expected, the results are similar to those obtained with the 1D versions of the numerical schemes.

Concerning the numerical schemes, four different schemes are compared in this section:

- First order numerical schemes of the form (42) whose viscosity matrix is given by (23) based on the intermediate matrices defined in Section 5.1. For simplicity these schemes will be referred to as **ROE**, even if in the case of the MP&M model they are not Roe schemes in the strict sense.

- First order numerical schemes of the form (42) whose viscosity matrix is given by (24) also based on the intermediate matrices defined in Section 5.1. They will be referred to as **FL**.
- Second order schemes based on a Roe scheme and on the reconstruction operator (57) with $\theta = 1/2$. They will be referred to as **HOS2-Rec1**.
- Second order schemes based on a Roe scheme and on the reconstruction operator (57) with $\theta = 1$. They will be referred to as **HOS2-Rec2**.

Concerning the computational cost, ROE is the cheapest one. The computational cost of FL is about $3/2$ times greater, as it requires the calculations of the reconstruction operator and the extra term in the viscosity matrix (which is basically the intermediate matrix to the square). As expected, the most expensive schemes are HOS2-Rec1 and HOS2-Rec2, whose computational cost is about three times that of a ROE: due to the time stepping, two reconstructions have to be calculated at every time step.

There is a fifth possibility which is not taken into account here: to consider a second order extension of a flux limiter scheme. We have verified in practice that the numerical results provided by such a scheme are hardly distinguishable of those given by the second order extension of a Roe scheme while the computational cost is greater, due to the calculation of the extra term in the viscosity matrix.

6.1 Two-dimensional bedload transport of a sediment layer with lintel form

We consider first an essentially 1D test case and in order to compare ROE, FL, HOS2-Rec1, and HOS2-Rec2 for a Saint-Venant-Exner model based on Grass formula (5) with $A_g = 0.01$ (median-high interaction) and sediment porosity $\rho_0 = 0.4$

We consider a rectangular channel $D = [0, 1000 \text{ m}] \times [0, 100 \text{ m}]$ with a flat non-erodible bottom. The computational mesh is composed of 12220 finite volumes of edge type. The CFL parameter is set to 0.8.

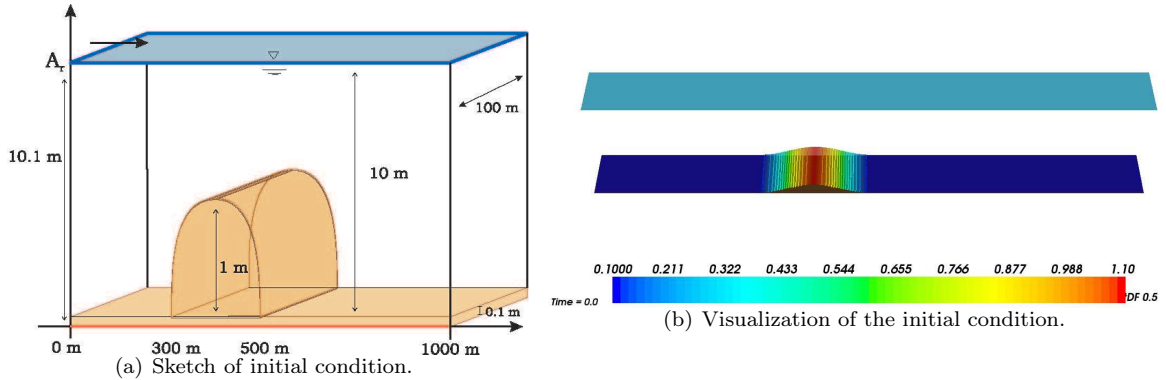


Figure 6: Bedload transport of a sediment layer with lintel form: Initial condition.

As initial condition we consider (see Figures 6(a) and 6(b)):

$$\begin{aligned}
 h(x_1, x_2, 0) &= 10.1 - z_b(x_1, x_2, 0), \quad q_1(x_1, x_2, 0) = 10, \quad q_2(x_1, x_2, 0) = 0, \\
 z_b(x_1, x_2, 0) &= \begin{cases} 0.1 + \sin^2\left(\frac{\pi(x_1 - 300)}{200}\right) & \text{if } 300 \leq x_1 \leq 500, \\ 0.1 & \text{otherwise.} \end{cases} \quad (75)
 \end{aligned}$$

The discharge $\mathbf{q} = (10, 0)$ and the sediment thickness $z_b = 0.1$ are imposed at $x_1 = 0$ and a free boundary condition is imposed at $x_1 = 1000$. Finally, slip boundary conditions are imposed at the lateral walls $x_2 = 0, x_2 = 100$.

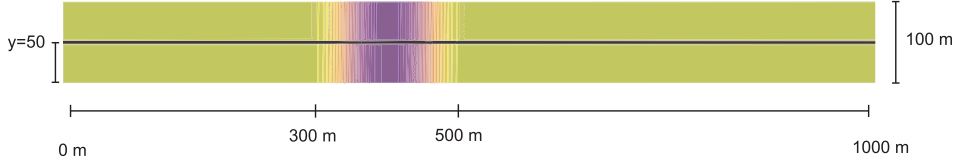


Figure 7: Bedload transport of a sediment layer with lintel form: Section used to make the comparison with 1d solution.

A 1D numerical solution obtained with a second order WENO2-Roe scheme has been also obtained using a mesh with 200 cells (see [4] for details). Furthermore in [18] an asymptotic solution is proposed.

Figure 8 shows a comparison at time $t = 50000$ s of the 1D numerical solution (continuous line), the 1D asymptotic solution (dashed line), the 2D numerical solutions at the section $x_2 = 50$ (see Figure 7) obtained with ROE (continuous line with pentagons), and FL (continuous line with crosses). As expected, ROE is more diffusive than FL, although both of them have first order accuracy.

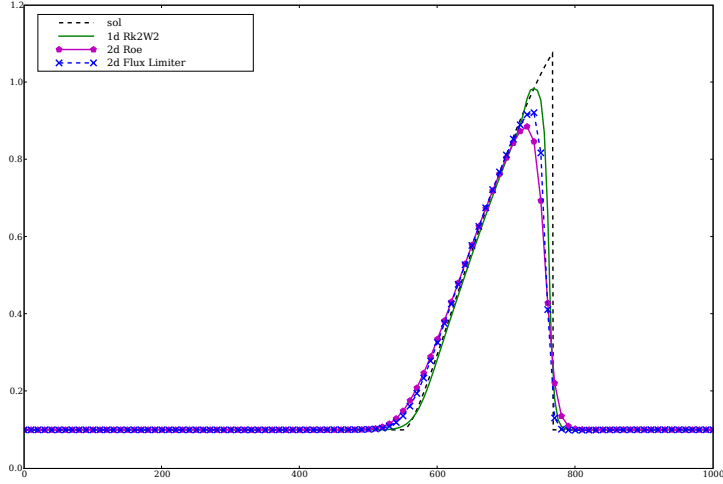


Figure 8: Bedload transport of a sediment layer with lintel form: the 1D asymptotic solution (dashed line) is compared with the numerical solutions obtained with the following schemes at time $t = 50000$ s: 1D WENO2-Roe numerical scheme (continuous line), ROE (continuous line with pentagons), FL (dashed line with crosses).

Figure 9 shows a comparison at time $t = 50000$ s of the 1D asymptotic solution (dashed line) with the numerical solutions given by the 1D second order numerical solution (continuous line), and the 2D numerical solutions at $x_2 = 50$ obtained with FL (dashed line with crosses) and HOS2-Rec1 (continuous line with circles). It can be observed that the 2D numerical solution obtained using HOS2-Rec1 is almost equal to the one dimensional one.

Figure 10 shows a comparison at time $t = 50000$ s of the 1D asymptotic solution (dashed line) with the 1D second order numerical solution (continuous line), and the 2D numerical solutions at $x_2 = 50$ obtained with HOS2-Rec1 (continuous line with circles) and HOS2-Rec2 (continuous line with triangles). It can be observed that the height of the sediment is maximal (even higher than the 1D second order numerical solution) if HOS2-Rec2 numerical is used. Nevertheless, if a higher

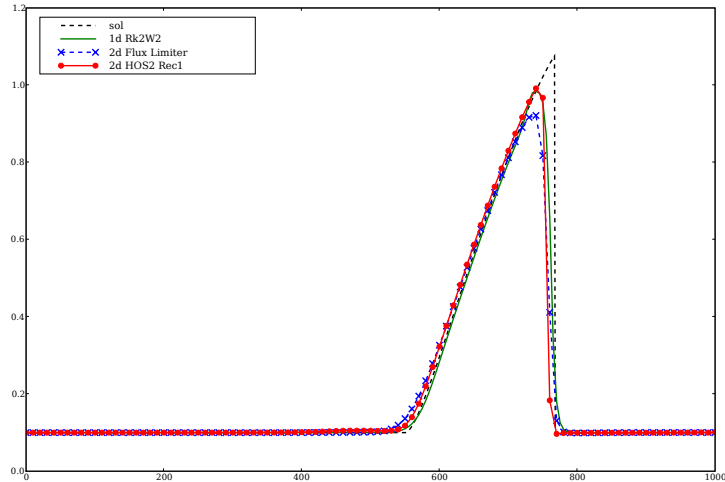


Figure 9: Bedload transport of a sediment layer with lintel form: the 1D asymptotic solution (dashed line) is compared with the numerical solutions obtained with the following schemes at time $t = 50000$ s: 1D WENO2-Roe numerical scheme (continuous line), FL (dashed line with crosses) and HOS2-Rec1 (continuous line with circles).

interaction constant is used, the peak produced by HOS2-Rec2 can degenerate into oscillations.

7 Two-dimensional simulation of the evolution of a conical dune of sand

In this purely two-dimensional test proposed in [18], the evolution of a conical dune in a channel with a non-erodible flat bottom, whose dimensions are $1000\text{ m} \times 1000\text{ m}$ is considered. The initial conditions are (see Figure 11):

$$h(x_1, x_2, 0) = 10.1 - z_b(x_1, x_2, 0), \quad q_1(x_1, x_2, 0) = 10, \quad q_2(x_1, x_2, 0) = 0,$$

and the initial form of the sediment layer is given by the function:

$$z_b(x_1, x_2, 0) = \begin{cases} 0.1 + \sin^2\left(\frac{\pi(x_1 - 300)}{200}\right) \sin^2\left(\frac{\pi(x_2 - 400)}{200}\right) & \text{if } \begin{matrix} 300 \leq x_1 \leq 500, \\ 400 \leq x_2 \leq 600, \end{matrix} \\ 0.1 & \text{otherwise.} \end{cases}$$

We consider again Grass formula. In this case, the sediment layer evolves towards a star-shaped pattern expanding along the time with a given spreading angle. De Vrien obtained in [8] the following analytical approximation of this angle, assuming that the interaction between the sediment layer and fluid is small, that is $A_g < 0.01$:

$$\tan \alpha = \frac{3\sqrt{3}(m_g - 1)}{9m_g - 1}.$$

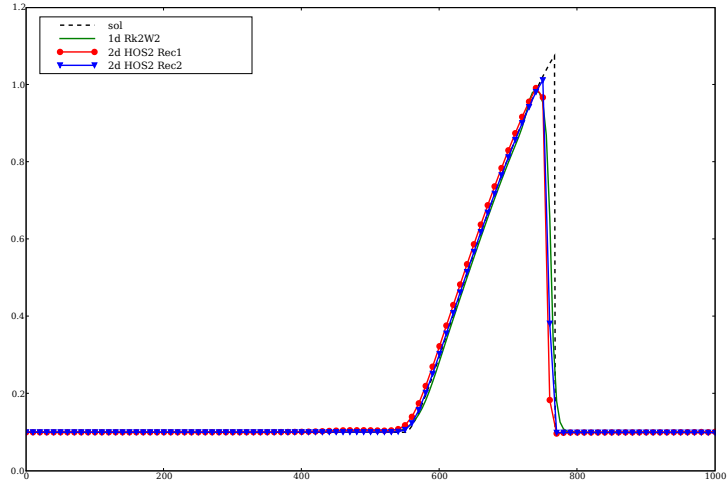


Figure 10: Bedload transport of a sediment layer with lintel form: the 1D asymptotic solution (dashed line) is compared with the numerical solutions obtained with the following schemes at time $t = 50000$ s: 1D WENO2-Roe numerical scheme (continuous line), HOS2-Rec1 (continuous line with circles), and HOS2-Rec2 (discontinuous line with triangles).

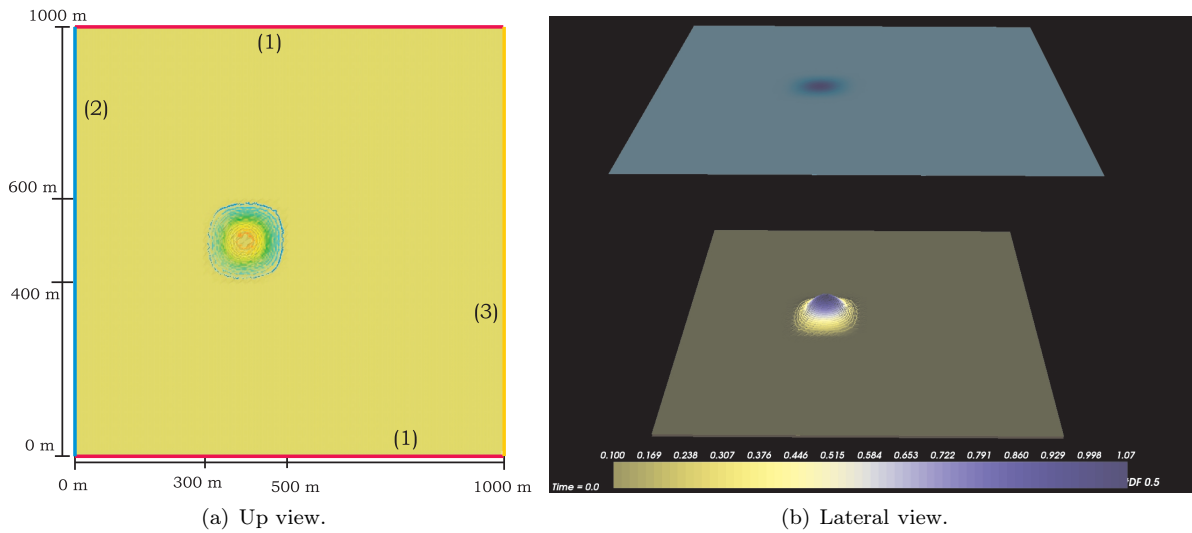


Figure 11: Evolution of a conical dune of sand: Initial condition.

In particular, for $m_g = 3$ the value of the spreading angle is: $\alpha = \tan^{-1} \left(\frac{3\sqrt{3}}{13} \right) = 21.786789^\circ$.

We use ROE and HOS2-Rec1 to solve the model with $A_g = 0.001$ and $\rho_0 = 0.4$. The CFL is set to 0.8. We impose the flux $\mathbf{q} = (10, 0)$ and the sediment layer height $z_b = 0.1$ at $x_1 = 0$ and free boundary conditions at $x_2 = 1000$. At the lateral walls we impose the slip condition $\mathbf{q} \cdot \boldsymbol{\eta} = 0$. We have considered three unstructured finite volume meshes of edge type with increasing number of cells: Mesh 1 (1240 volumes); Mesh 2 (4880 volumes) and Mesh 3 (19360 volumes). All meshes are symmetric with respect to the axis $x_2 = 500$.

In Table 1 we show the spread angle and the maximum height of the sediment layer obtained for each mesh using both schemes. It can be observed that as the mesh is refined, the spread angle tends to the analytical one. Furthermore, note that ROE is more diffusive than HOS2-Rec1, as expected. In Figure 12 we show the estimates of the spreading angle (comparing different times of the sand dune evolution) obtained with ROE and HOS2-Rec1 for Mesh 3.

	Mesh 1		Mesh 2		Mesh 3	
	Roe	HOS2-Rec1	Roe	HOS2-Rec1	Roe	HOS2-Rec1
Spread angle	39°	30°	36°	25°	35°	24.5°
$\max(z_b)$ at $t = 100\text{h}$.	0.368	0.495	0.488	0.714	0.602	0.804

Table 1: Evolution of a conical dune of sand: Spread angle and maximum height of the sediment layer at $t = 100$ h, obtained by ROE and HOS2-Rec1 using three meshes.

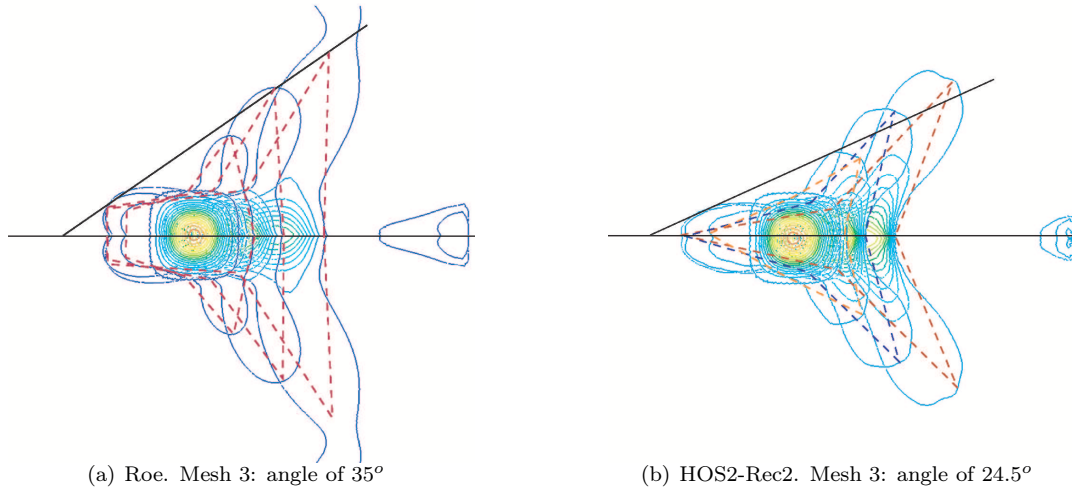


Figure 12: Evolution of a conical dune of sand: Estimation of the spreading angle (using 15 iso-levels). ROE (left). HOS2-Rec1 (right).

Finally, Figure 13 shows the numerical solution obtained by ROE and HOS2-Rec1 with the finest mesh at time $t = 100\text{h}$.

7.1 Sediment evolution in a L-shaped channel.

In this test, we study the evolution of the sediments in a L-shaped channel whose non-erodible bottom is flat, using the MP&M formula (6). A sketch of the channel is shown in Figure (14(a)). The lower left vertex of the channel is placed at the origin. An unstructured finite volume mesh of edge type with 20880 volumes has been constructed. As initial condition, we impose

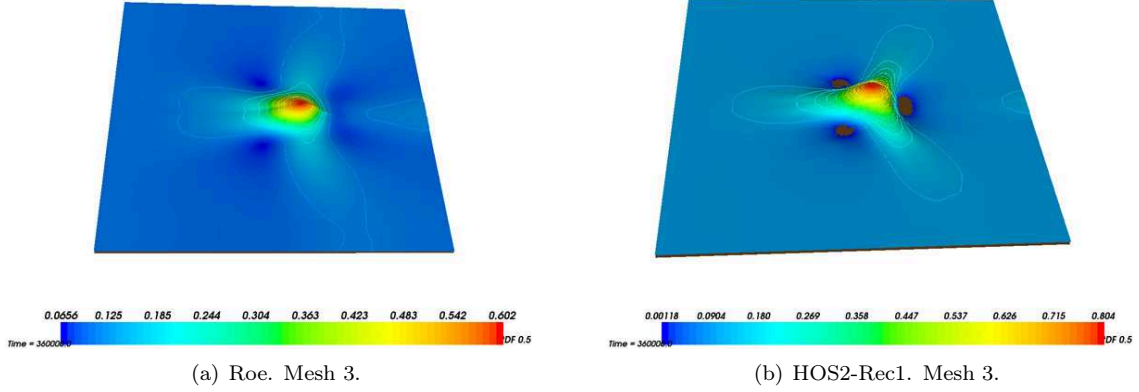


Figure 13: Evolution of a conical dune of sand: Sediment layer after 100 hours.

$$h(x_1, x_2, 0) = 1.0 - z_b(x_1, x_2, 0), \quad q_1(x_1, x_2, 0) = 0, \quad q_2(x_1, x_2, 0) = 0,$$

and

$$z_b(x_1, x_2, 0) = 0.1 + 0.2e^{-((x_1-2)^2+(x_2-2)^2)}.$$

Figure 14(b) shows the initial form of the sediment layer. As boundary conditions, a constant velocity profile is imposed at Γ_U : $u_1(x_1, x_2, t) = 0$ m/s and $u_2(x_1, x_2, t) = -0.5$ m/s together with $z_b(x_1, x_2, 0) = 0.1$ m. A free boundary condition is imposed at Γ_D . At the lateral walls Γ_L the slip boundary condition $\mathbf{q} \cdot \boldsymbol{\eta} = 0$ is imposed. The water density is $\rho = 1000$ kg/m³; the sediment density, $\rho_s = 2600$ kg/m³; the sediment grain size, $d_i = 10^{-3}$ m; the Manning coefficient, $n = 0.0196$; the non-dimensional critical shear stress, $\tau_{*,c} = 0.047$; and the sediment porosity, $\rho_0 = 0.4$. Finally, the *CFL* parameter is set to 0.9. HOS2-Rec1 scheme is used.

Figure 15 shows the evolution of the sediment layer (left column) as well as the velocity field (right column). As expected, the bedload transport is more intense where the velocities are bigger, and some regions in which the sediment layer vanishes can be observed. In regions with small velocities no sediment transport is produced.

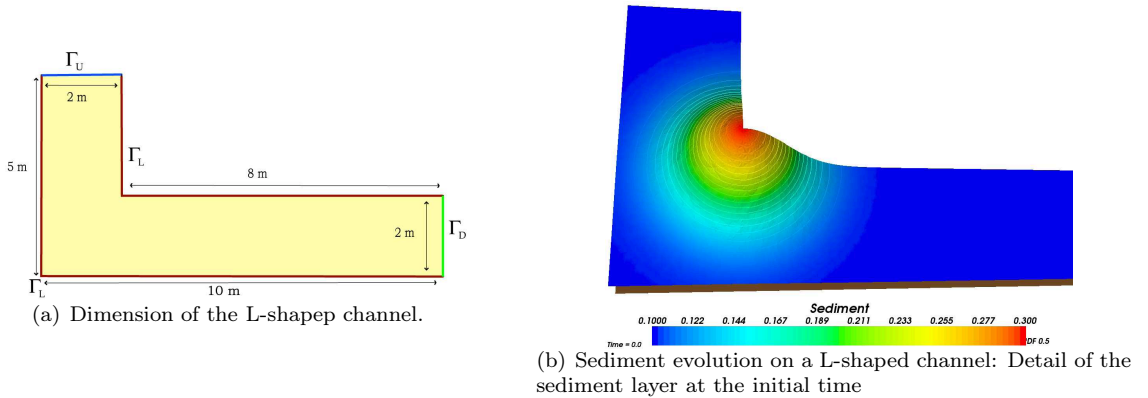


Figure 14: Sediment evolution on L-shaped channel: Dimensions of the channel (left) and initial sediment layer depth (right).

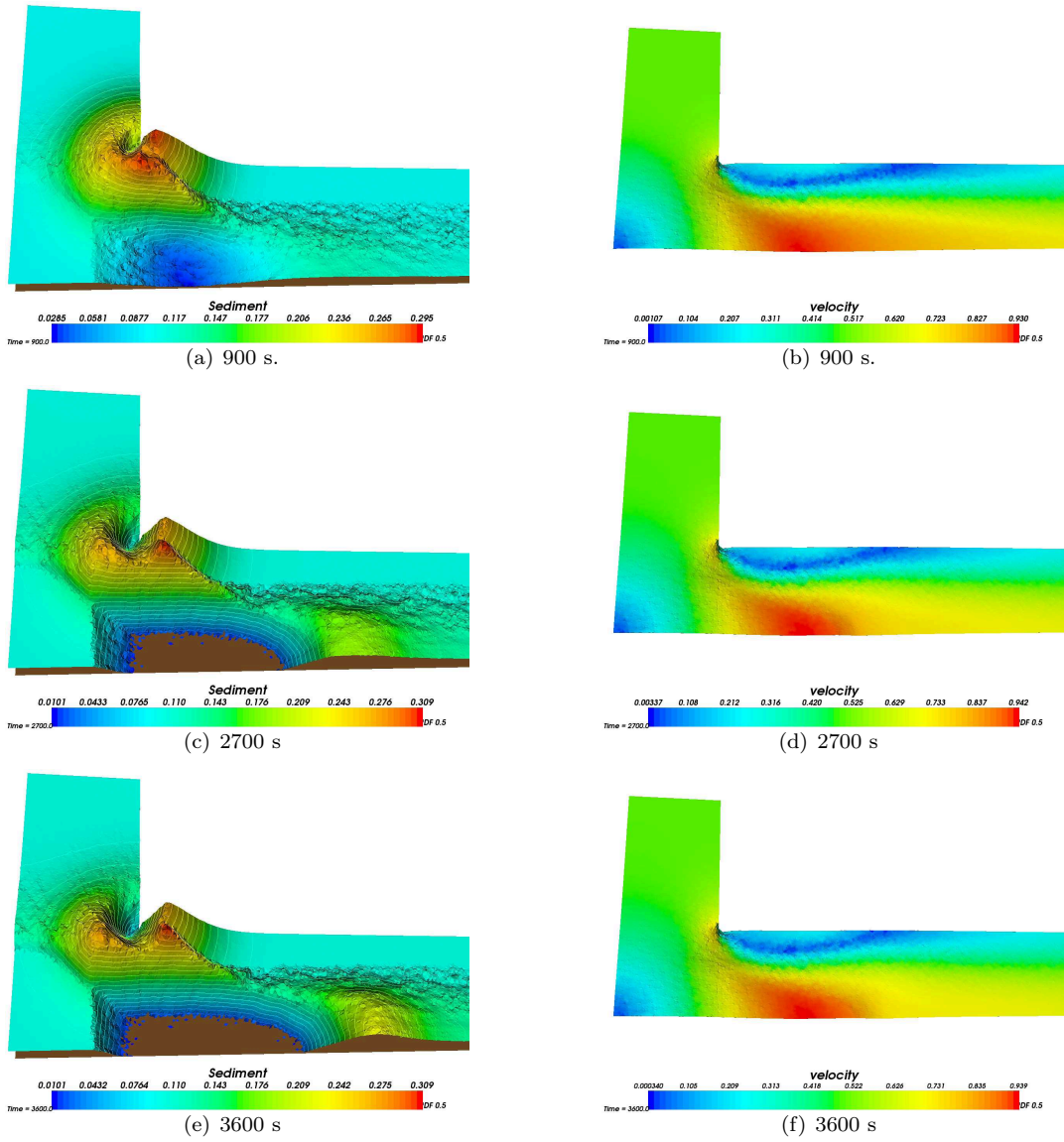


Figure 15: Sediment evolution on a L-shaped channel: Sediment layer evolution (left column) , and water velocity field (right column).

8 Conclusions

In this paper the numerical approximation of Saint-Venant-Exner models by means of finite volume methods on unstructured meshes has been considered. The models studied here consist of a hydrodynamical component that is modeled by a 2D shallow water system and a morphodynamical component given by a solid transport discharge formula that depends on the hydrodynamical variables. Two different formulae for the solid transport discharge have been taken into account: the Grass and the MP&M models. A first difficulty involved by the numerical solution of these systems is due to the presence of nonconservative products. On the one hand, it has been shown in [6] that, even if a standard numerical flux is chosen to discretize the flux term of the system, the scheme may become unconditionally unstable if the nonconservative products are not discretized in an adequate manner: roughly speaking a correct treatment requires to discretize the system as a whole. To deal with this difficulty we consider first order schemes based on linear approximate Riemann solvers. The key point is that the linearization of the systems considered at every inter-cell satisfies a generalized Roe property based on the choice of a family of paths (see [36], [29], and [5]). The use of these generalized Roe schemes presents the following advantages:

1. The family of paths in which the Roe matrices are based allows one to give a precise sense to the nonconservative products that are ambiguous from the mathematical point of view (see [24]).
2. The generalized Roe property ensures that, if the system has a conservative subsystem (which is the case for the subsystem composed by the mass conservation equation of the fluid and the solid transport discharge in Saint-Venant-Exner models) the numerical scheme is conservative (in the usual sense) for the subsystem.
3. The Roe linearization takes into account the information of the whole system, including the conservative and the nonconservative terms. This global information is used to correctly upwind the characteristics variables. As a consequence, the numerical scheme is stable under a standard CFL condition.

Nevertheless, some difficulties arise in the practical implementation of these Roe schemes. On the one hand, the choice of the family of paths should be related to the physics of the system: This notion is based on a family of paths in the phases space, whose selection is important, as it determines the speed of propagation of the discontinuities: a motivation for the selection of the family of paths when a physical regularization by diffusion, dispersion, etc. can be found in [16]. Unfortunately, the calculation of such a family of paths can be very difficult in practice. As a first approximation, we have considered here the simplest choice which is given by the family of straight segments. On the other hand, once the family of paths has been chosen, the calculation of intermediate matrices that satisfy exactly the generalized Roe property can also be difficult. Here, we have presented the explicit calculation of a Roe matrix for the model corresponding to the Grass formula but not for the MP&M formula: its complex expression makes difficult this task even for the simple family of straight segments. To solve this difficulty, we have considered a general family of first order schemes in viscosity form that contain Roe schemes as a particular case in which the viscosity matrix is given by the absolute value of the intermediate matrix. The numerical schemes of this more general family have still the advantages 1 and 2 above but their stability properties can be worse than those of a Roe method.

A second important difficulty of Saint-Venant-Exner models is related to the different characteristic speeds of the water and the sediment layers when the interaction is weak. Due to this, first order numerical schemes are, in general, too diffusive to correctly capture the waves related to the sediment layer motion. In order to solve this difficulty, we have first derived a first order scheme in viscosity form in which a flux limiter technique has been used to reduce the numerical diffusion of Roe methods. Next, second order extensions of the first order schemes have been introduced on the basis of a reconstruction technique: a new MUSCL type reconstruction operator for unstructured meshes has been defined. The main advantages of this operator are the small size of the stencils and the

simplicity of the coefficients of the linear approximation functions at the cells. Two different slope limiters have been introduced to avoid the appearance of oscillations near a discontinuity: the first one is frequently used for MUSCL schemes on unstructured meshes and the second one mimics at every edge the slope limiters that are commonly used for 1D problems. Once the two slope limiters techniques have been introduced, a more general slope limiter is introduced consisting of a convex linear combination of both of them: a parameter $\theta \in [0, 1]$ has to be chosen, the choices $\theta = 0$ and $\theta = 1$ corresponding to one of the two slope limiters initially considered. A second order TVD-RK scheme is finally used for the time stepping.

We have presented three test cases, the first two ones using the Grass formula and the third one using the MP&M formula. First, an essentially 1D test case has been considered in which an asymptotic solution is known. The numerical schemes have been compared with this solution and with the numerical solutions provided by a second order 1D numerical schemes. Next, the 2D evolution of a conical dune of sand has been simulated. In this case, the available analytical approximation of the spreading angle has been used to validate the schemes. For the 2D model based on the MP&M model, we do not know any test case in which exact or approximate analytical solutions are available. To test the schemes, we have simulated the evolution of the sediments in a L-shaped channel whose non-erodible bottom is flat. This is a hard test as the thickness of the sediment layer vanishes in parts of the domain.

The numerical tests show that:

- Roe schemes are too diffusive, as expected.
- The second order schemes are able to correctly simulate the sediment layer motion and to properly capture the shocks. The choice $\theta = 1/2$ gives a good compromise between the numerical diffusion added by the slope limiter and the elimination of oscillations near the discontinuities. The computational cost is about three times the corresponding to a Roe scheme.
- In the case of medium-high interaction between the water and the sediment layers, the first order scheme based on the flux limiter technique is an interesting alternative to second order schemes, as they provide good numerical results with a significant reduction of the computational cost: its cost is about 3/2 times the corresponding to a Roe scheme.

This paper represents the first stage of a conjoint work with geologists of the Spanish Institute of Oceanography (I.E.O.) to study the sediment transport in the continental shelf near the mouth of a river. The in-situ data will be used in the next future to calibrate and to test the numerical models. The test cases shown here, specially the third one, shows that the second order numerical schemes presented here can be used to develop realistic and useful models.

Appendix A

PROOF OF THEOREM 2

In order to avoid an excess of indexes, let us use the notation (x, y) for the coordinates instead of (x_1, x_2) . Let us prove that, given $\bar{N} = (\bar{x}, \bar{y}) \in V_i$, one has $\nabla W_i = \nabla W(\bar{N}) + O(\Delta)$, where ∇W_i is given by (49)

Let us consider the following notation: $V_{i,l}$, $l = 0, \dots, 4$ are the cells defining the stencil of the reconstruction, where by $V_{i,0}$ we denote the cell V_i and by $V_{i,l}$, $l = 1, \dots, 4$ its four neighbors. Let us define $\bar{W}_{i,l}$, $l = 0, \dots, 4$ by

$$\bar{W}_{i,l} = \frac{1}{|V_{i,l}|} \int_{V_{i,l}} W(\mathbf{x}) d\mathbf{x}, \quad l = 0, \dots, 4.$$

$N_{i,l} = (x_{i,l}, y_{i,l})$, $l = 0, \dots, 4$ represent the centers of mass of $V_{i,l}$ defined by (46). Let us also define $\alpha_{i,l} = (x_{i,l} - \bar{x})$ and $\beta_{i,l} = (y_{i,l} - \bar{y})$, $l = 0, \dots, 4$.

A Taylor expansion of $W(\mathbf{x})$ at $\bar{N} = (\bar{x}, \bar{y})$ gives the approximation:

$$W(\mathbf{x}) = W(\bar{N}) + W_x(\bar{N})(x - \bar{x}) + W_y(\bar{N})(y - \bar{y}) + O(\Delta^2).$$

Then, by averaging the previous expression at cells $V_{i,l}$, we obtain

$$\bar{W}_{i,l} = \frac{1}{|V_{i,l}|} \int_{V_{i,l}} W(\mathbf{x}) d\mathbf{x} = W(\bar{N}) + W_x(\bar{N})(x_{i,l} - \bar{x}) + W_y(\bar{N})(y_{i,l} - \bar{y}) + O(\Delta^2).$$

Taking into account the definition of $\alpha_{i,l}$ and $\beta_{i,l}$, the last equation can be rewritten as:

$$\bar{W}_{i,l} = W(\bar{N}) + W_x(\bar{N})\alpha_{i,l} + W_y(\bar{N})\beta_{i,l} + O(\Delta^2). \quad (76)$$

As ∇W_i is a linear combination of $\bar{W}_{i,l}$ it can be written as follows:

$$\nabla W_i = \left(\sum_{l=0}^4 \mu_l^1 \bar{W}_{i,l}, \sum_{l=0}^4 \mu_l^2 \bar{W}_{i,l} \right).$$

Moreover, by using (76), we have:

$$\sum_{l=0}^4 \mu_l^k \bar{W}_{i,l} = \left(\sum_{l=0}^4 \mu_l^k \right) W(\bar{N}) + W_x(\bar{N}) \left(\sum_{l=0}^4 \alpha_{i,l} \mu_l^k \right) + W_y(\bar{N}) \left(\sum_{l=0}^4 \beta_{i,l} \mu_l^k \right) + \sum_{l=0}^4 \mu_l^k O(\Delta^2), \quad k = 1, 2. \quad (77)$$

Therefore, if we prove that:

$$\begin{aligned} a) \sum_{l=0}^4 \mu_l^1 &= 0, & b) \sum_{l=0}^4 \alpha_{i,l} \mu_l^1 &= 1, & c) \sum_{l=0}^4 \beta_{i,l} \mu_l^1 &= 0, \\ d) \sum_{l=0}^4 \mu_l^2 &= 0, & e) \sum_{l=0}^4 \alpha_{i,l} \mu_l^2 &= 0, & f) \sum_{l=0}^4 \beta_{i,l} \mu_l^2 &= 1, \end{aligned} \quad (78)$$

and

$$\mu_l^k = O(\Delta^{-1}),$$

the proof is finished.

Using the same arguments than in [30], it is easy to prove that $\mu_l^k = O(\Delta^{-1})$, $l = 0, \dots, 4$, $k = 1, 2$. Let us prove only a), b) and c) (the proof of d), e) and f) is similar).

The proof is divided in three parts:

1) Let us prove that $\sum_{l=0}^4 \mu_l^1 = 0$.

We consider $T = T_1 \cup T_2 \cup T_3 \cup T_4$ (see Figure 16) . The gradient approximation (49) verifies,

$$\begin{aligned} \nabla W_i &= \frac{|T_1|}{|T|} (\bar{W}_{i,0} \nabla \lambda_1^0 + \bar{W}_{i,1} \nabla \lambda_1^1 + \bar{W}_{i,2} \nabla \lambda_1^2) + \frac{|T_2|}{|T|} (\bar{W}_{i,0} \nabla \lambda_2^0 + \bar{W}_{i,2} \nabla \lambda_2^2 + \bar{W}_{i,3} \nabla \lambda_2^3) \\ &+ \frac{|T_3|}{|T|} (\bar{W}_{i,0} \nabla \lambda_3^0 + \bar{W}_{i,3} \nabla \lambda_3^3 + \bar{W}_{i,4} \nabla \lambda_3^4) + \frac{|T_4|}{|T|} (\bar{W}_{i,0} \nabla \lambda_4^0 + \bar{W}_{i,4} \nabla \lambda_4^4 + \bar{W}_{i,1} \nabla \lambda_4^1) \\ &= \frac{\bar{W}_{i,0}}{|T|} (|T_1| \nabla \lambda_1^0 + |T_2| \nabla \lambda_2^0 + |T_3| \nabla \lambda_3^0 + |T_4| \nabla \lambda_4^0) \\ &+ \sum_{j=1}^4 \frac{\bar{W}_{i,ip(j)}}{|T|} (|T_{ip(j)}| \nabla \lambda_{ip(j)}^{ip(j)} + |T_j| \nabla \lambda_j^{ip(j)}), \end{aligned}$$

where $ip(j)$, $j = \{1, 2, 3, 4\}$ takes values in the ordered set $\{2, 3, 4, 1\}$, and λ_k^l is the barycentric coordinate associated to the node $N_{i,l}$ in the triangle T_k .

Remark 7 Note that the following equalities hold for a triangle T of vertices v_1, v_2 and v_3

$$\nabla \lambda_j = \frac{-1}{2|T|} \eta_j, \quad j = 1, 2, 3,$$

where λ_j is the barycentric coordinate associated to the vertex v_j and η_j is the outer normal vector to the opposite edge whose modulus is equal to the length of that edge.

Taking into account Remark 7 and using the notation introduced in Figure 16, we deduce

$$\begin{aligned} \mu_0^1 &= \frac{-1}{2|T|} (\eta_{12} + \eta_{23} + \eta_{34} + \eta_{41})_1, \\ \mu_1^1 &= \frac{-1}{2|T|} (\eta_{02} - \eta_{04})_1, \quad \mu_2^1 = \frac{-1}{2|T|} (\eta_{03} - \eta_{01})_1, \\ \mu_3^1 &= \frac{-1}{2|T|} (-\eta_{02} + \eta_{04})_1, \quad \mu_4^1 = \frac{-1}{2|T|} (-\eta_{03} + \eta_{01})_1. \end{aligned} \quad (79)$$

Trivially, $\mu_0^1 = 0$, because it is the sum of the outer normals to a closed polygon, and therefore $\sum_{l=0}^4 \mu_l^1 = 0$.

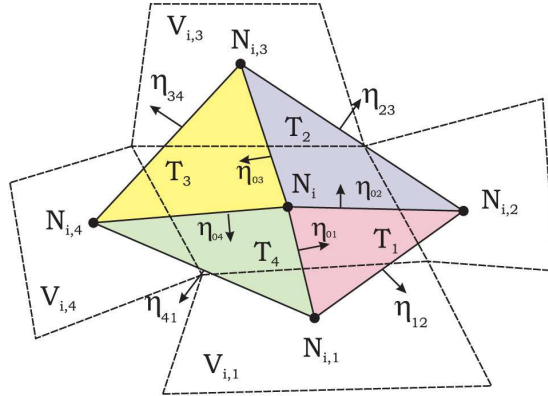


Figure 16: Building the triangles T_1, T_2, T_3, T_4 , to approximate the gradient of the solution.

2) Let us prove that $\sum_{l=0}^4 \alpha_{i,l} \mu_l^1 = 1$.

Taking into account the definition of $\alpha_{i,l}$ and the previous result, we have

$$\sum_{l=0}^4 \alpha_{i,l} \mu_l^1 = \sum_{l=0}^4 x_{i,l} \mu_l^1 - \bar{x} \sum_{l=0}^4 \mu_l^1 = \sum_{l=0}^4 x_{i,l} \mu_l^1.$$

Therefore, it is enough to prove that $\sum_{l=0}^4 x_{i,l} \mu_l^1 = 1$. Focussing for example on T_1 , we have that

$$|T_1| = \int_{T_1} \operatorname{div}(x, 0) dx dy = \left(\frac{x_{i,1} + x_{i,2}}{2} \eta_{12} + \frac{x_{i,2} + x_{i,0}}{2} \eta_{02} - \frac{x_{i,0} + x_{i,1}}{2} \eta_{01} \right)_1.$$

As $\eta_{12} + \eta_{02} - \eta_{01} = 0$ we can subtract $(x_{i,0} + x_{i,1} + x_{i,2})(\eta_{12} + \eta_{02} - \eta_{01})_1/2$ from the previous expression, obtaining finally:

$$|T_1| = \frac{-1}{2}(x_{i,0}\eta_{12} + x_{i,1}\eta_{02} - x_{i,2}\eta_{01})_1.$$

As a consequence:

$$\begin{aligned} \sum_{l=0}^4 x_{i,l}\mu_l^1 &= \frac{-1}{2|T|}(x_{i,0}\eta_{12} + x_{i,1}\eta_{02} - x_{i,2}\eta_{01})_1 + \frac{-1}{2|T|}(x_{i,0}\eta_{23} - x_{i,3}\eta_{02} + x_{i,2}\eta_{03})_1 \\ &+ \frac{-1}{2|T|}(x_{i,0}\eta_{34} + x_{i,3}\eta_{04} - x_{i,4}\eta_{03})_1 + \frac{-1}{2|T|}(x_{i,0}\eta_{41} + x_{i,4}\eta_{01} - x_{i,1}\eta_{04})_1 \\ &= \frac{1}{|T|}|T_1| + \frac{1}{|T|}|T_2| + \frac{1}{|T|}|T_3| + \frac{1}{|T|}|T_4| = 1. \end{aligned}$$

3) Let us prove that $\sum_{l=0}^4 \beta_{i,l}\mu_l^1 = 0$.

Taking into account again the definition $\beta_{i,l} = y_{i,l} - \bar{y}$ and a), it is straightforward to verify that

$$\sum_{l=0}^4 \beta_{i,l}\mu_l^1 = \sum_{l=0}^4 (y_{i,l} - \bar{y})\mu_l^1 = \sum_{l=0}^4 y_{i,l}\mu_l^1 - \sum_{l=0}^4 \bar{y}\mu_l^1 = \sum_{l=0}^4 y_{i,l}\mu_l^1.$$

Therefore, it is enough to prove that $\sum_{l=0}^4 y_{i,l}\mu_l^1 = 0$:

$$\begin{aligned} \sum_{l=0}^4 y_{i,l}\mu_l^1 &= \frac{-1}{2|T|}(y_{i,1}(-(y_{i,2} - y_{i,0}) + (y_{i,4} - y_{i,0})) + y_{i,2}(-(y_{i,3} - y_{i,0}) + (y_{i,1} - y_{i,0})) \\ &+ y_{i,3}((y_{i,2} - y_{i,0}) - (y_{i,4} - y_{i,0})) + y_{i,4}((y_{i,3} - y_{i,0}) - (y_{i,1} - y_{i,0}))) = 0. \end{aligned}$$

We conclude that

$$\sum_{l=0}^4 \mu_l^1 \bar{W}_{i,l} = W_x(\bar{N}) + O(\Delta), \quad \forall \bar{N} \in V_i.$$

Using similar arguments, it is straightforward to prove that the approximation of $W_y(\bar{N})$ is also first order accurate and, consequently, the reconstruction (47) is second order accurate in V_i , $i \in \mathbb{Z}$.

References

- [1] K. Anastasiou, C. T. Chan. *Solution of the 2D shallow water equations using the finite volume method on unstructured triangular meshes*. Num. Methods Fluids, 24: 1225–1245, 1997.
- [2] T. Barth, D. Jespersen *The design and application of upwind schemes on unstructured meshes*. AIAA Paper 89-0366, 1989.
- [3] M.J. Castro, José M. Gallardo, Carlos Parés. *High order finite volume schemes based on reconstruction of states for solving hyperbolic systems with nonconservative products. Applications to shallow-water systems*. Math. Comp. 75(255): 1103–1134, 2006.

- [4] M.J. Castro Díaz, E.D. Fernández Nieto, A.M. Ferreiro. *Sediment transport models in Shallow Water equations and numerical approach by high order finite volume methods*. Computers and Fluids, 37(3): 299–316, 2008.
- [5] M.J. Castro Díaz, E.D. Fernández Nieto, A.M. Ferreiro, A. García Rodríguez, C. Parés. *High order extension of Roe schemes for two dimensional nonconservative hyperbolic systems*. accepted on J. Sci. Comp. DOI 10.1007/s10915-008-9250-4.
- [6] M.J. Castro, J. Macía, C. Parés. *A Q-scheme for a class of systems of coupled conservation laws with source term. Application to a two-layer 1-D shallow water system*. ESAIM-Math. Model. Num. 35 (1): 107–127, 2001.
- [7] N. Chien. *The present status of research on sediment transport*. Trans. ASCE, 121: 833–68, 1956.
- [8] H. J. De Vrien. *2DH Mathematical Modelling of Morphological Evolutions in Shallow Water*. Coastal Engineering, 11: 1-27, 1987.
- [9] M. Dumbser, M. Käser. *Arbitrary high order non-oscillatory finite volume schemes on unstructured meshes for linear hyperbolic systems*. J. Comput. Phys., 221(2): 693–723, 2007.
- [10] M. Dumbser, M. Käser, V. A. Titarev, E. F. Toro. *Quadrature-free non-oscillatory finite volume schemes on unstructured meshes for nonlinear hyperbolic systems*. J. Comput. Phys, 226: 204–243, 2007.
- [11] M. Dumbser, D.S. Balsara, E.F. Toro, C.D. Munz. *A unified framework for the construction of one-step finite-volume and discontinuous Galerkin schemes on unstructured meshes*. J Comput. Phys., 227: 8209–8253, 2008.
- [12] Denys Dutykh. *Modélisation mathématique des tsunamis*. Ph.D. Thesis, École Normale Supérieure de Cachan, 2007.
- [13] A.C. Fowler, N. Kopteva, C. Oakley. *The formation of river channels*. SIAM J. Appl. math. Vol. 67, No. 4: 1016–1040, 2007.
- [14] A. J. Grass. *Sediments transport by waves and currents*. SERC London Cent. Mar. Technol., Report No. FL29, 1981.
- [15] A. Harten, J.M. Hyman. *Self-adjusting grid methods for one-dimensional hyperbolic conservation laws*. J. Comp. Phys. 50: 235–269, 1983.
- [16] P.G. LEFLOCH. *Shock waves for nonlinear hyperbolic systems in nonconservative form*, Institute for Math. and its Appl., Minneapolis, Preprint 593, 1989.
- [17] Changqing Hu, Chi-Wang Shu. *Weighted essentially non-oscillatory schemes on triangular meshes*. J. Comput. Phys. 150: 97–127, 1999.
- [18] J. Hudson. *Numerical technics for morphodynamic modelling*. Tesis doctoral. University of Whiteknights, 2001.
- [19] G. Jiang, C.-W. Shu. *Efficient implementation of wieghted ENO schemes*. J. Comput. Phys. 126: 202–228, 1996.
- [20] N.E. Kolgan. *Application of the minimum-derivative principle in the construction of finite-difference schemes for numerical analysis of discontinuous solutions in gas dynamics*. Uchenye Zapiski Tsagi [Sci. Notes Central Inst. Aerodyn], 3(6):68-77, 1972.
- [21] X.D. Liu, S. Osher, T. Chan. *Weighted essentially nonoscillatory schemes*. J. Comput. Phys. 115: 200–212, 1994.

- [22] D.A. Lyn, M. Altinakar. *St. Venant-Exner equations for near-critical and transcritical flows*. J. Hydraulic Engineering, 128 (6): 579–587, 2002.
- [23] A. Marquina. *Local piecewise hyperbolic reconstructions for nonlinear scalar conservation laws*. SIAM J. Sci. Comput. 15: 892–915, 1994.
- [24] G. Dal Masso, P.G. LeFloch, F. Murat. *Definition and weak stability of nonconservative products*. J. Math. Pures Appl. 74: 483–548, 1995.
- [25] E. Meyer-Peter, R. Müller. *Formulas for bed-load transport*. Rep. 2nd Meet. Int. Assoc. Hydraul. Struct. Res., Stockholm: 39–64, 1948.
- [26] P. Nielsen. *Coastal Bottom Boundary Layers and Sediment Transport*. World Scientific Publishing, Singapore, Advanced Series on Ocean Engineering, 4, 1992.
- [27] S. Noelle, N. Pankratz, G. Puppo, J. Natvig. *Well-balanced finite volume schemes of arbitrary order of accuracy for shallow water flows*. J. Comput. Phys., 213: 474–499, 2006.
- [28] C. Parés: *Numerical methods for nonconservative hyperbolic systems: a theoretical framework*, SIAM J. Num. Anal., 44: 300– 321 , 2006.
- [29] C. Parés, M.J. Castro. *On the well-balance property of Roe’s method for non conservative hiperbolic systems. Applicatons to shallow-water systems*. ESAIM: M2AN, 38(5): 821–852, 2004.
- [30] H. Joachim Schroll, Fredrik Svensson. *A bi-hyperbolic finite volume method on quadrilateral meshes*. J. Sci. Comp., 26 (2): 237-260(24), 2006.
- [31] C.W. Shu. *Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws*. ICASE Report: 97–65, 1997.
- [32] C.W. Shu, S. Osher. *Efficient implementation of essentially non-oscillatory shock capturing schemes*. J. Comput. Phys.,77: 439–71, 1998.
- [33] Richard Soulsby. *Dynamics of marine sands. A manual for practical applications*. Published by Thomas Telford Publications, Thomas Telford Services Ltd, 1997.
- [34] P.A. Tassi, S. Rhebergena, C.A. Vionnetb and O. Bokhovea. *A discontinuous Galerkin finite element model for river bed evolution under shallow flows*. CMAME, 197(33-40):2930–2947, 2008.
- [35] E.F. Toro. *Shock-Capturing Methods for Free-Surface Shallow Flows*. Wiley, 2001.
- [36] I. Toumi. *A weak formulation of Roe’s approximate Riemann Solver*. J. Comp. Phys. 102(2): 360–373, 1992.
- [37] B. Van Leer. *Towards the ultimate conservative difference scheme v: a second order sequel to Godunov’ method*. J. Comput. Phys., 32:101-136, 1979.
- [38] B. Van Leer. *Upwind and high-resolution methods for compressible flow: From donor cell to residual-distribution schemes*. Communications in Computational Physics, 1:192-206, 2006.
- [39] L. C. Van Rijn. *Sediment transport (I): bed load transport*. J. Hydraul. Div., Proc. ASCE, 110: 1431–56, 1984.
- [40] A. I. Volpert, *Spaces BV and quasilinear equations*, Math. USSR Sbornik, 73 (1967), pp. 255–302.
- [41] G. Walz, *Romberg Type Cubature over Arbitrary Triangles*. Mannheimer Mathem. Manuskripte Nr.225, Mannheim, 1997.