

## 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

MANUEL CASTRO<sup>1</sup>, JORGE MACÍAS<sup>1</sup> AND CARLOS PARÉS<sup>1</sup>

**Abstract.** The goal of this paper is to construct a first-order upwind scheme for solving the system of partial differential equations governing the one-dimensional flow of two superposed immiscible layers of shallow water fluids. This is done by generalizing a numerical scheme presented by Bermúdez and Vázquez-Cendón [3, 26, 27] for solving one-layer shallow water equations, consisting in a  $Q$ -scheme with a suitable treatment of the source terms. The difficulty in the two layer system comes from the coupling terms involving some derivatives of the unknowns. Due to these terms, a numerical scheme obtained by performing the upwinding of each layer, independently from the other one, can be unconditionally unstable. In order to define a suitable numerical scheme with global upwinding, we first consider an abstract system that generalizes the problem under study. This system is not a system of conservation laws but, nevertheless, Roe's method can be applied to obtain an upwind scheme based on Approximate Riemann State Solvers. Following this, we present some numerical tests to validate the resulting schemes and to highlight the fact that, in general, numerical schemes obtained by applying a  $Q$ -scheme to each separate conservation law of the system do not yield good results. First, a simple system of coupled Burgers' equations is considered. Then, the  $Q$ -scheme obtained is applied to the two-layer shallow water system.

**Mathematics Subject Classification.** 65M99, 76B55, 76B70.

Received: March 7, 2000. Revised: November 3, 2000.

### 1. INTRODUCTION

The problem that gave rise to the questions addressed in this work was the discretization, by means of finite volume schemes, of the system of partial differential equations governing the one-dimensional flow of two

---

*Keywords and phrases.*  $Q$ -schemes, coupled conservation laws, source terms, 1D shallow water equations, two-layer flows, hyperbolic systems.

<sup>1</sup> Dpto. Análisis Matemático. Facultad de Ciencias. Universidad de Málaga, Campus de Teatinos s/n, 29080 Málaga, Spain.  
e-mail: grupo@anamat.cie.uma.es

superposed immiscible layers of shallow water fluids:

$$\left\{ \begin{array}{l} \frac{\partial h_1}{\partial t} + \frac{\partial q_1}{\partial x} = 0, \\ \frac{\partial q_1}{\partial t} + \frac{\partial}{\partial x} \left( \frac{q_1^2}{h_1} + \frac{g}{2} h_1^2 \right) = gh_1 \frac{dH}{dx} - gh_1 \frac{\partial h_2}{\partial x}, \\ \frac{\partial h_2}{\partial t} + \frac{\partial q_2}{\partial x} = 0, \\ \frac{\partial q_2}{\partial t} + \frac{\partial}{\partial x} \left( \frac{q_2^2}{h_2} + \frac{g}{2} h_2^2 \right) = gh_2 \frac{dH}{dx} - \frac{\rho_1}{\rho_2} gh_2 \frac{\partial h_1}{\partial x}. \end{array} \right. \quad (1.1)$$

In these equations, index 1 makes reference to the upper layer and index 2 to the lower one. The fluid is assumed to occupy a straight channel with a constant rectangular cross-section. The coordinate  $x$  refers to the axis of the channel and  $t$  is time. Each layer is assumed to have a constant density,  $\rho_i$ ,  $i = 1, 2$  ( $\rho_1 < \rho_2$ ). The unknowns  $q_i(x, t)$  and  $h_i(x, t)$  represent the mass-flow and the thickness of the  $i$ th layer at the section of coordinate  $x$  at time  $t$ ;  $g$  is gravity and  $H(x)$  is the depth at  $x$  from a fixed reference level.

System (1.1) can be written as follows:

$$\left\{ \begin{array}{l} \frac{\partial W_1}{\partial t}(x, t) + \frac{\partial F}{\partial x}(W_1(x, t)) = \phi(x, W_1(x, t)) + B_{1,2}(W_1(x, t)) \frac{\partial W_2}{\partial x}(x, t), \\ \frac{\partial W_2}{\partial t}(x, t) + \frac{\partial F}{\partial x}(W_2(x, t)) = \phi(x, W_2(x, t)) + B_{2,1}(W_2(x, t)) \frac{\partial W_1}{\partial x}(x, t), \end{array} \right. \quad (1.2)$$

where

$$W_j(x, t) = \begin{bmatrix} h_j(x, t) \\ q_j(x, t) \end{bmatrix}, \quad j = 1, 2, \quad (1.3)$$

and, given  $W^T = [h, q]$  with  $h \neq 0$ :

$$F(W) = \begin{bmatrix} q \\ \frac{q^2}{h} + \frac{g}{2} h^2 \end{bmatrix}, \quad (1.4)$$

$$B_{1,2}(W) = \begin{bmatrix} 0 & 0 \\ -gh & 0 \end{bmatrix}, \quad (1.5)$$

$$B_{2,1}(W) = \begin{bmatrix} 0 & 0 \\ -grh & 0 \end{bmatrix}, \quad (1.6)$$

$$\phi(x, W) = \begin{bmatrix} 0 \\ gh \frac{dH}{dx}(x) \end{bmatrix}, \quad (1.7)$$

with

$$r = \frac{\rho_1}{\rho_2}.$$

Observe that  $F$  and  $\phi$  are the usual flux and source terms appearing in shallow water problems. *Bermúdez* and *Vázquez-Cendón* [3, 26, 27] have presented some numerical upwind schemes for solving the one-layer shallow

water equations consisting in a  $Q$ -scheme with a suitable treatment of the source terms. Nevertheless, in (1.2) new kinds of source terms, involving derivatives of the unknowns, appear. These source terms affect the nature of the system which can even lose its hyperbolic character. In effect, if the two layers were uncoupled, the eigenvalues of each separate system would be those of the Jacobian matrix of the flux:

$$A(W) = \frac{\partial F}{\partial W} = \begin{bmatrix} 0 & 1 \\ -\frac{q^2}{h^2} + gh & 2\frac{q}{h} \end{bmatrix}. \quad (1.8)$$

These eigenvalues are:

$$u_i \pm \sqrt{gh_i}, \quad i = 1, 2, \quad (1.9)$$

where  $u_i = q_i/h_i$  is the averaged velocity of the  $i$ th layer. On the other hand, if the system (1.2) is written under the form:

$$\mathbf{W}_t + \mathcal{A}(\mathbf{W})\mathbf{W}_x = \Phi(x, \mathbf{W}), \quad (1.10)$$

where

$$\mathbf{W} = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}, \quad (1.11)$$

$$\Phi(x, \mathbf{W}) = \begin{bmatrix} \phi(x, W_1) \\ \phi(x, W_2) \end{bmatrix}, \quad (1.12)$$

$$\mathcal{A}(\mathbf{W}) = \begin{bmatrix} A(W_1) & -B_{1,2}(W_1) \\ -B_{2,1}(W_2) & A(W_2) \end{bmatrix}, \quad (1.13)$$

it is clear that the eigenvalues of the system are now those of the matrix  $\mathcal{A}(\mathbf{W})$ , whose characteristic equation is:

$$(\lambda^2 - 2u_1\lambda + u_1^2 - gh_1)(\lambda^2 - 2u_2\lambda + u_2^2 - gh_2) = rg^2h_1h_2. \quad (1.14)$$

Observe that, when  $r = 0$ , the eigenvalues are those corresponding to each layer separately, so that for  $r \ll 1$ , the eigenvalues of the system (1.1) approach (1.9). In this situation, the coupling terms do not affect the nature of the system in any essential manner.

Nevertheless, we are mainly interested in the case  $r \cong 1$ , as this is the situation occurring in many oceanographical flows. For example, our interest focuses on the Strait of Gibraltar where, in first approximation, two superposed layers of water of very close densities can be distinguished: the Atlantic and the Mediterranean waters, with densities of 1027 kg/m<sup>3</sup> and 1029 kg/m<sup>3</sup>, respectively.

In the case  $r \cong 1$ , first-order approximation of the eigenvalues were given in [23]:

$$\lambda_{\text{ext}}^{\pm} = U_{\text{con}} \pm (g(h_1 + h_2))^{\frac{1}{2}}, \quad (1.15)$$

$$\lambda_{\text{int}}^{\pm} = U_{\text{con}} \pm \left( g' \frac{h_1 h_2}{(h_1 + h_2)} \left[ 1 - \frac{(u_1 - u_2)^2}{g'(h_1 + h_2)} \right] \right)^{\frac{1}{2}}. \quad (1.16)$$

In the former expression,  $g'$  is the *reduced gravity*:

$$g' = (1 - r)g,$$

and  $U_{\text{con}}$ , the so-called *convective velocity*, given by:

$$U_{\text{con}} = \frac{u_1 h_2 + u_2 h_1}{h}, \quad (1.17)$$

where  $h = h_1 + h_2$  is the total thickness.

The two *external eigenvalues*,  $\lambda_{\text{ext}}^{\pm}$ , are related to the *barotropic component* of the flow and the two *internal eigenvalues*,  $\lambda_{\text{int}}^{\pm}$ , to the *baroclinic component*. These latter eigenvalues depend on the reduced gravity  $g'$  and  $g' \ll 1$ . As a consequence, the velocity of internal waves, relative to  $U_{\text{con}}$ , is much lower than that of waves at the surface (see [12, 17]).

It is easy to check that, for a two-layer flow with free surface, the condition under which one of the eigenvalues vanishes is:

$$G^2 = F_1^2 + F_2^2 - (1-r)F_1^2 F_2^2 = 1, \quad (1.18)$$

where  $G$  is the so-called *composite Froude number*, and  $F_i$  for  $i = 1, 2$  are the internal Froude numbers for each layer ( $F_i^2 = \frac{u_i^2}{g'h_i}$ ). When this condition is achieved at a section of coordinate  $x$ , the flow is said to be *critical* at this point and the section  $x$  is called a *control*. When  $G^2 < 1$ , the flow is *subcritical*. Finally, when  $G^2 > 1$ , the flow is *supercritical*.

From equation (1.16), we can observe that the internal eigenvalues may become complex. This situation occurs when they verify, approximately, the following inequality:

$$\frac{(u_1 - u_2)^2}{g'(h_1 + h_2)} > 1. \quad (1.19)$$

In this case, the system loses its hyperbolic character. These situations are related with the appearance of the so-called *Kelvin-Helmholtz instabilities*. These instabilities appear at the interface and lead, in real flows, to intense mixing of the two layers. While, in practice, this mixture partially dissipates the energy, in numerical experiments, these interface disturbances grow and overwhelm the solution. Clearly, we cannot expect to simulate these phenomena with a two-immiscible-layer model. Therefore, the inequality (1.19) in fact gives the range of validity of a model based on the equations (1.1), if viscosity effects are neglected.

The fact that some of the source terms affect the eigenvalues of the system makes clear that a correct upwinding of the fluxes and source terms must use the information given by the global matrix  $\mathcal{A}$ . Therefore, the goal of this paper is to generalize to system (1.1) the numerical scheme presented by *Bermúdez* and *Vázquez-Cendón* for solving one-layer shallow water equations in such a way that the upwinding will be performed in a global manner.

In order to define a numerical scheme with these properties, we consider an abstract system that generalizes the problem under study. The abstract problem consists of several conservation laws with source terms that couple every law to each other, and it can be written as a quasi-linear non-conservative hyperbolic system. The interpretation of the non-conservative terms appearing in this system and, related to this, the notion of an entropy weak solution, create serious difficulties from a mathematical point of view. These topics have been extensively studied in [4, 18, 19]. As we are only concerned by the definition of a numerical scheme, we will assume the existence and uniqueness of an entropy solution, regular enough so that the calculations performed make sense.

In defining the scheme, we will treat the difficulties progressively. First we consider a system where source terms of the form (1.12) do not appear. This system is not a system of conservation laws but, nevertheless, Roe's method [20] can still be applied to obtain an upwind scheme based on Approximate Riemann State Solvers.

Then, source terms of the form (1.12) are added. The numerical treatment proposed in [26, 27] is applied, but using again the global matrix  $\mathcal{A}(\mathbf{W})$  to correctly upwind.

An alternative scheme can be obtained by applying an upwind scheme to each separate system. This procedure leads to a numerical scheme easier to implement and computationally less expensive. Unfortunately, in practice, this scheme is useless because it may be unconditionally unstable. This point will be further discussed in Section 2.

In [7], a similar problem, arising from the simulation of two-phase liquid-vapour flows, has been considered. Model equations are also a system of coupled conservation laws with source term. Nevertheless, this problem does not fit into the abstract framework proposed here in Section 2, as their coupling terms also involve time derivatives of the unknowns. These authors introduce a splitting technique for treating these coupling terms that allows the use of approximate Riemann solvers to each separate system.

In Section 3, some numerical tests that validate the schemes defined in Section 2 are presented. First, a simple system of coupled Burgers' equations with known solution is considered. Then, the numerical solution of system (1.1) with flat bottom is studied. In order to make clear the fact that numerical schemes with uncoupled upwinding may be unstable, we apply two different numerical schemes, one with global and another with uncoupled upwinding, and compare the results.

At our knowledge, exact solutions of the system (1.1) are not known. Nevertheless, in [1] some approximate steady-state solutions are presented, under the "rigid-lid" hypothesis. In Section 3 we also compare these approximate steady solutions with the numerical solutions obtained with the scheme presented here, by iterating in time until convergence. Finally, we present a numerical experiment where a periodic barotropic forcing is imposed to the two-layer flow over a sill. This represents a first approximation to the study of tidal effects on such a geometry.

## 2. DEFINITION OF A Q-SCHEME FOR A GENERAL PROBLEM

Let us consider the following system of P.D.E.

$$\frac{\partial W_j}{\partial t} + \frac{\partial F_j}{\partial x}(W_j) = \phi_j(x, W_1, \dots, W_K) + \sum_{k \neq j} B_{j,k}(x, W_1, \dots, W_K) \cdot \frac{\partial W_k}{\partial x}, \quad j = 1, \dots, K, \quad (2.20)$$

where

$$W_j(x, t) = \begin{bmatrix} w_1^j(x, t) \\ w_2^j(x, t) \\ \vdots \\ w_{N_j}^j(x, t) \end{bmatrix} \in \mathbb{R}^{N_j},$$

$F_j$  is a continuous function from a domain  $D_j$  of  $\mathbb{R}^{N_j}$  to  $\mathbb{R}^{N_j}$ ;  $\phi_j$  is a continuous function from  $\mathbf{D} = D_0 \times D_1 \times \dots \times D_K$ , where  $D_0$  is an open interval of  $\mathbb{R}$ , to  $\mathbb{R}^{N_j}$ . Finally,  $B_{j,k}$  is a continuous matrix function from  $\mathbf{D}$  to  $\mathcal{M}_{N_j \times N_k}$ . The first-order P.D.E. system (2.20) contains  $N = N_1 + N_2 + \dots + N_K$  unknowns and two independent variables  $x$  and  $t$ . The  $j$ th equation of this system can be thought as a conservation law with source terms for the set of variables  $W_j$ .

The following global notation will be used:

$$\mathbf{W}_t + \mathbf{F}(\mathbf{W})_x = \Phi(x, \mathbf{W}) + \mathbf{G}(x, \mathbf{W}, \mathbf{W}_x), \quad (2.21)$$

where

$$\begin{aligned}\mathbf{W} &= \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_K \end{bmatrix}, \\ \mathbf{F}(\mathbf{W}) &= \begin{bmatrix} F_1(W_1) \\ F_2(W_2) \\ \vdots \\ F_K(W_K) \end{bmatrix}, \\ \Phi(x, \mathbf{W}) &= \begin{bmatrix} \phi_1(x, \mathbf{W}) \\ \phi_2(x, \mathbf{W}) \\ \vdots \\ \phi_K(x, \mathbf{W}) \end{bmatrix}, \\ \mathbf{G}(x, \mathbf{W}, \mathbf{W}_x) &= \mathbf{B}(x, \mathbf{W}) \cdot \mathbf{W}_x.\end{aligned}$$

In the latter equation,  $\mathbf{B}$  represents the  $N \times N$  matrix with the following block structure:

$$\mathbf{B}(x, \mathbf{W}) = \begin{bmatrix} 0 & B_{1,2}(x, \mathbf{W}) & \dots & B_{1,K}(x, \mathbf{W}) \\ B_{2,1}(x, \mathbf{W}) & 0 & \dots & B_{2,K}(x, \mathbf{W}) \\ \vdots & \vdots & \ddots & \vdots \\ B_{K,1}(x, \mathbf{W}) & B_{K,2}(x, \mathbf{W}) & \dots & 0 \end{bmatrix}. \quad (2.22)$$

A further hypothesis concerning the hyperbolic nature of the system (2.24) is required. But first, let us introduce some notation. For  $1 \leq j \leq K$ , given  $W_j \in D_j \subset \mathbb{R}^{N_j}$ ,  $A_j(W_j)$  represents the Jacobian of the  $j$ th flux  $F_j$  in  $W_j$ :

$$A_j(W_j) = \frac{\partial F_j}{\partial W_j}(W_j).$$

Clearly, given  $\mathbf{W} \in D_1 \times \dots \times D_K$ , one has:

$$\frac{\partial \mathbf{F}}{\partial \mathbf{W}} = \mathbf{A}(\mathbf{W}) = \begin{bmatrix} A_1(W_1) & 0 & \dots & 0 \\ 0 & A_2(W_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_K(W_K) \end{bmatrix}.$$

Let  $\mathcal{A}(x, \mathbf{W})$  be the matrix:

$$\mathcal{A}(x, \mathbf{W}) = \mathbf{A}(\mathbf{W}) - \mathbf{B}(x, \mathbf{W}) = \begin{bmatrix} A_1(W_1) & -B_{1,2}(x, \mathbf{W}) & \dots & -B_{1,K}(x, \mathbf{W}) \\ -B_{2,1}(x, \mathbf{W}) & A_2(W_2) & \dots & -B_{2,K}(x, \mathbf{W}) \\ \vdots & \vdots & \ddots & \vdots \\ -B_{K,1}(x, \mathbf{W}) & -B_{K,2}(x, \mathbf{W}) & \dots & A_K(W_K) \end{bmatrix}.$$

Using these matrices, (2.21) can be written as follows:

$$\mathbf{W}_t + \mathcal{A}(x, \mathbf{W}) \cdot \mathbf{W}_x = \Phi(x, \mathbf{W}). \quad (2.23)$$

We assume that this system is **hyperbolic**, *i.e.*, for all  $(x, \mathbf{W}) \in \mathbf{D}$  the matrix  $\mathcal{A}(x, \mathbf{W})$  has real eigenvalues and a complete set of linearly independent eigenvectors.

The difficulties related to the numerical treatment of the two kinds of source terms appearing in the equations are very different. On the one hand, the source terms involving derivatives of the unknown have to be taken into account to calculate the eigenvalues of the system and, due to this, the upwinding of these terms is strongly related to the upwinding of the flux term and has to be performed at the same time. This will be done in Section 2.1 where, in order to separate the difficulties, we will consider  $\Phi = 0$ . On the other hand, the difficulties related to the source terms of the form  $\Phi(x, \mathbf{W})$  are similar to those appearing in a standard system of conservation laws with source terms. The numerical treatment proposed here is a straightforward adaptation of the techniques developed in [26, 27], and will be presented in Section 2.2.

### 2.1. A Q-scheme with global upwinding

In this section we consider problem (2.21) with  $\Phi = 0$ , *i.e.*,

$$\mathbf{W}_t + \mathbf{F}(\mathbf{W})_x = \mathbf{G}(x, \mathbf{W}, \mathbf{W}_x). \quad (2.24)$$

An important remark is that, in the case where the matrices  $B_{j,k}$  are constants, (2.24) can be interpreted as a standard system of conservation laws:

$$\mathbf{W}_t + \mathbf{F}_{\mathbf{B}}(\mathbf{W})_x = 0, \quad (2.25)$$

with

$$\mathbf{F}_{\mathbf{B}}(\mathbf{W}) = \mathbf{F}(\mathbf{W}) - \mathbf{B} \cdot \mathbf{W}.$$

Hence, the system can be numerically solved using a standard scheme.

In this paragraph, we adapt Roe's method to define a Q-scheme for (2.24). As this equation is not a system of conservation laws, this procedure cannot be directly applied, but it can be adapted. The idea is the following: in Roe's method, the numerical fluxes at intercells are calculated by solving a Riemann Problem related to a linearized system. But taking into account the remark above, these linearized problems can be considered as standard system of conservation laws by including the source terms in the flux. Roughly speaking, the source terms can be **locally** viewed as flux terms. Once these local systems of conservation laws are solved, the deduction of the scheme closely follows Roe's method.

Suppose that the system (2.24) has to be solved in the spatial domain  $[0, L]$  and the temporal domain  $[0, T]$  with initial conditions:

$$W_j(x, 0) = W_j^0(x), \quad x \in [0, L], \quad j = 1, \dots, K. \quad (2.26)$$

To discretize the equations,  $M$  computing cells  $I_i = [x_{i-1/2}, x_{i+1/2}]$ ,  $i = 1, \dots, M$  are considered. For the sake of simplicity, we assume that these cells have a constant size:

$$\Delta x = L/M,$$

and that

$$x_{i+1/2} = i\Delta x, \quad i = 0, \dots, M.$$

$x_i = (i - 1/2)\Delta x$  is the center of the cell  $I_i$ . Let  $\Delta t$  be the time step and  $t^n = n\Delta t$ .

As usual, we define the cell averages of the exact solution as follows:

$$\mathbf{W}_i^n = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathbf{W}(x, t^n) dx.$$

Integrating by parts the system in the control volume  $I_i \times [t^n, t^{n+1}]$ , we formally obtain:

$$\mathbf{W}_i^{n+1} = \mathbf{W}_i^n + \frac{\Delta t}{\Delta x} (\mathbf{F}_{i-1/2} - \mathbf{F}_{i+1/2}) + \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{t^n}^{t^{n+1}} \mathbf{G}(x, \mathbf{W}, \mathbf{W}_x) dt dx, \quad (2.27)$$

where

$$\mathbf{F}_{i+1/2} = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \mathbf{F}(\mathbf{W}(x_{i+1/2}, t)) dt,$$

is the averaged flux through the intercell  $x_{i+1/2}$ . Clearly,  $\mathbf{F}_{i+1/2}$  depends also on the time level, nevertheless, in order to avoid an excess of notation, we will not write 0 the dependency on time for quantities that are computed at time  $t^n$  without any possible ambiguity. We write (2.27) as follows:

$$\begin{aligned} \mathbf{W}_i^{n+1} = \mathbf{W}_i^n &+ \frac{\Delta t}{\Delta x} \left( \mathbf{F}_{i-1/2} + \frac{1}{\Delta t} \int_{x_{i-1/2}}^{x_i} \int_{t^n}^{t^{n+1}} \mathbf{G}(x, \mathbf{W}, \mathbf{W}_x) dt dx \right) \\ &- \frac{\Delta t}{\Delta x} \left( \mathbf{F}_{i+1/2} - \frac{1}{\Delta t} \int_{x_i}^{x_{i+1/2}} \int_{t^n}^{t^{n+1}} \mathbf{G}(x, \mathbf{W}, \mathbf{W}_x) dt dx \right). \end{aligned} \quad (2.28)$$

As usual in ‘‘Godunov-type’’ methods, we consider the Riemann problems:

$$\begin{cases} \widetilde{\mathbf{W}}_t + \mathbf{F}(\widetilde{\mathbf{W}})_x = \mathbf{G}(x, \widetilde{\mathbf{W}}, \widetilde{\mathbf{W}}_x), \\ \widetilde{\mathbf{W}}(x, t^n) = \begin{cases} \mathbf{W}_i^n & \text{if } x < x_{i+1/2}, \\ \mathbf{W}_{i+1}^n & \text{if } x > x_{i+1/2}, \end{cases} \end{cases} \quad (2.29)$$

and, assuming that each of these problems has a unique ‘‘entropic’’ solution  $\widetilde{\mathbf{W}}^{i+1/2}$ , for  $i = 1, \dots, M-1$  the parenthesis appearing on the right-hand side of (2.28) are approximated, respectively, by:

$$\begin{aligned} \mathbf{g}_{i-1/2}^+ &= \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \mathbf{F}(\widetilde{\mathbf{W}}^{i-1/2}(x_{i-1/2}, t)) dt + \frac{1}{\Delta t} \int_{x_{i-1/2}}^{x_i} \int_{t^n}^{t^{n+1}} \mathbf{G}(x, \widetilde{\mathbf{W}}^{i-1/2}, \widetilde{\mathbf{W}}_x^{i-1/2}) dt dx, \\ \mathbf{g}_{i+1/2}^- &= \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \mathbf{F}(\widetilde{\mathbf{W}}^{i+1/2}(x_{i+1/2}, t)) dt - \frac{1}{\Delta t} \int_{x_i}^{x_{i+1/2}} \int_{t^n}^{t^{n+1}} \mathbf{G}(x, \widetilde{\mathbf{W}}^{i+1/2}, \widetilde{\mathbf{W}}_x^{i+1/2}) dt dx. \end{aligned}$$

We assume also that the time step  $\Delta t$  has been chosen small enough so that the following relations hold:

$$x_i \leq x_{i+1/2} + \Delta t S_L^{i+1/2}, \quad x_{i+1} \geq x_{i+1/2} + \Delta t S_R^{i+1/2}, \quad i = 1, \dots, M-1, \quad (2.30)$$

where  $S_L^{i+1/2}$  and  $S_R^{i+1/2}$  are the fastest velocities to the left and to the right of the perturbations arising at  $x_{i+1/2}$  at time  $t = t^n$  in problem (2.29).

Under this hypothesis, integrating equation (2.29) on the control volume  $[x_i, x_{i+1/2}] \times [t^n, t^{n+1}]$ , we obtain:

$$\mathbf{g}_{i+1/2}^- = \mathbf{F}(\mathbf{W}_i^n) - S_L^{i+1/2} \mathbf{W}_i^n - \frac{1}{\Delta t} \int_{x_{i+1/2} + \Delta t S_L^{i+1/2}}^{x_{i+1/2}} \widetilde{\mathbf{W}}^{i+1/2}(x, t^{n+1}) dx. \quad (2.31)$$



Analogously,

$$\mathbf{g}_{i-1/2}^+ = \mathbf{F}(\mathbf{W}_i^n) - S_R^{i-1/2} \mathbf{W}_i^n + \frac{1}{\Delta t} \int_{x_{i-1/2}}^{x_{i-1/2} + \Delta t} S_R^{i-1/2} \widetilde{\mathbf{W}}^{i-1/2}(x, t^{n+1}) dx. \quad (2.32)$$

In order to approximate the integral expressions appearing on the right-hand side of (2.31) and (2.32), we introduce linear approximations of the Riemann problems (2.29). To do this, we consider the matrices:

$$\begin{aligned} \widetilde{\mathbf{A}}_{i+1/2} &= \mathbf{A}(\widetilde{\mathbf{W}}_{i+1/2}^n), \quad 1 \leq i \leq M-1, \\ \widetilde{\mathbf{B}}_{i+1/2} &= \mathbf{B}(x_{i+1/2}, \widetilde{\mathbf{W}}_{i+1/2}^n), \quad 1 \leq i \leq M-1, \end{aligned}$$

where  $\widetilde{\mathbf{W}}_{i+1/2}^n$  is an “intermediate state” between  $\mathbf{W}_i^n$  and  $\mathbf{W}_{i+1}^n$  to be chosen.

Then the approximate linear Riemann problems considered are:

$$\begin{cases} \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{A}}_{i+1/2} \cdot \widetilde{\mathbf{U}}_x = \widetilde{\mathbf{B}}_{i+1/2} \cdot \widetilde{\mathbf{U}}_x, \\ \widetilde{\mathbf{U}}(x, t^n) = \begin{cases} \mathbf{W}_i^n & \text{if } x < x_{i+1/2}, \\ \mathbf{W}_{i+1}^n & \text{if } x > x_{i+1/2}. \end{cases} \end{cases} \quad (2.33)$$

Now, the equations of these problems can be written under the form of standard linear systems of conservation laws as follows:

$$\widetilde{\mathbf{U}}_t + \widetilde{\mathcal{A}}_{i+1/2} \cdot \widetilde{\mathbf{U}}_x = 0, \quad (2.34)$$

where

$$\widetilde{\mathcal{A}}_{i+1/2} = \widetilde{\mathbf{A}}_{i+1/2} - \widetilde{\mathbf{B}}_{i+1/2}.$$

Let us introduce some notation:  $\{\lambda_l^{i+1/2}\}_{l=1}^N$  are the eigenvalues of  $\widetilde{\mathcal{A}}_{i+1/2}$ , assumed to be ordered as:

$$\lambda_1^{i+1/2} < \lambda_2^{i+1/2} < \dots < \lambda_N^{i+1/2},$$

and  $\{\mathbf{K}_l^{i+1/2}\}_{l=1}^N$  is a complete set of linearly independent eigenvectors.  $\mathbf{K}_{i+1/2}$  is the  $N \times N$  matrix whose columns are the eigenvector  $\mathbf{K}_l^{i+1/2}$ , and  $\Lambda_{i+1/2}$  is the diagonal matrix whose diagonal coefficients are the eigenvalues  $\lambda_1^{i+1/2}, \dots, \lambda_N^{i+1/2}$ . Clearly,

$$\widetilde{\mathcal{A}}_{i+1/2} = \mathbf{K}_{i+1/2} \Lambda_{i+1/2} \mathbf{K}_{i+1/2}^{-1}.$$

We also introduce the matrices  $\Lambda_{i+1/2}^+$ ,  $\Lambda_{i+1/2}^-$ ,  $\widetilde{\mathcal{A}}_{i+1/2}^+$ ,  $\widetilde{\mathcal{A}}_{i+1/2}^-$  as usual

$$\Lambda_{i+1/2}^\pm = \begin{bmatrix} (\lambda_1^{i+1/2})^\pm & & 0 \\ & \ddots & \\ 0 & & (\lambda_N^{i+1/2})^\pm \end{bmatrix}, \quad \widetilde{\mathcal{A}}_{i+1/2}^\pm = \mathbf{K}_{i+1/2} \Lambda_{i+1/2}^\pm \mathbf{K}_{i+1/2}^{-1}.$$

Assuming that

$$S_L^{i+1/2} \leq \lambda_1^{i+1/2}, \quad \lambda_N^{i+1/2} \leq S_R^{i+1/2}, \quad 1 \leq i \leq M-1; \quad (2.35)$$

and repeating the same procedure of integration applied to the solution of (2.29), we obtain

$$\begin{aligned} \frac{1}{\Delta t} \int_{x_{i+1/2} + \Delta t S_L^{i+1/2}}^{x_{i+1/2}} \tilde{\mathbf{U}}^{i+1/2}(x, t^{n+1}) dx &= -S_L^{i+1/2} \mathbf{W}_i^n - \tilde{\mathcal{A}}_{i+1/2}^- \cdot (\mathbf{W}_{i+1}^n - \mathbf{W}_i^n); \\ \frac{1}{\Delta t} \int_{x_{i-1/2}}^{x_{i-1/2} + \Delta t S_R^{i-1/2}} \tilde{\mathbf{U}}^{i-1/2}(x, t^{n+1}) dx &= S_R^{i-1/2} \mathbf{W}_i^n - \tilde{\mathcal{A}}_{i-1/2}^+ \cdot (\mathbf{W}_i^n - \mathbf{W}_{i-1}^n), \end{aligned}$$

where the explicit form of the solution of (2.33) has been used (see [25]). Using the values of these integrals to approximate the integrals appearing in (2.31) and (2.32) we obtain

$$\begin{aligned} \mathbf{g}_{i-1/2}^+ &\cong \tilde{\mathbf{g}}_{i-1/2}^+ = \mathbf{F}(\mathbf{W}_i^n) - \tilde{\mathcal{A}}_{i-1/2}^+ \cdot (\mathbf{W}_i^n - \mathbf{W}_{i-1}^n), \\ \mathbf{g}_{i+1/2}^+ &\cong \tilde{\mathbf{g}}_{i+1/2}^+ = \mathbf{F}(\mathbf{W}_i^n) + \tilde{\mathcal{A}}_{i+1/2}^- \cdot (\mathbf{W}_{i+1}^n - \mathbf{W}_i^n). \end{aligned}$$

Finally, using these approximations in (2.28) the following numerical scheme is obtained:

$$\mathbf{W}_i^{n+1} = \mathbf{W}_i^n + \frac{\Delta t}{\Delta x} (\tilde{\mathbf{g}}_{i-1/2}^+ - \tilde{\mathbf{g}}_{i+1/2}^-), \quad 2 \leq i \leq M-1, \quad (2.36)$$

(the expressions for  $i = 1$  and  $i = M$  depend on the boundary conditions imposed). In order to avoid the introduction of unnecessary notation, for the approximate value of  $\mathbf{W}(x_i, t^n)$  given by the scheme we use the same symbol,  $\mathbf{W}_i^n$ , already used for the cell averages of the exact solution.

If the intermediate states  $\tilde{\mathbf{W}}_{i+1/2}^n$  chosen are those of Roe's method (see [20]), the following equalities hold:

$$\mathbf{F}(\mathbf{W}_{i+1}^n) - \mathbf{F}(\mathbf{W}_i^n) = \tilde{\mathbf{A}}_{i+1/2} \cdot (\mathbf{W}_{i+1}^n - \mathbf{W}_i^n), \quad 1 \leq i \leq M-1.$$

In this case, introducing the matrices

$$|\tilde{\mathcal{A}}_{i+1/2}| = \tilde{\mathcal{A}}_{i+1/2}^+ - \tilde{\mathcal{A}}_{i+1/2}^-,$$

a new form of the scheme can be obtained:

$$\begin{aligned} \mathbf{W}_i^{n+1} &= \mathbf{W}_i^n + \frac{\Delta t}{\Delta x} (\mathbf{F}_{i-1/2} - \mathbf{F}_{i+1/2}) \\ &+ \frac{\Delta t}{2\Delta x} \left( \tilde{\mathbf{B}}_{i-1/2} \cdot (\mathbf{W}_i^n - \mathbf{W}_{i-1}^n) + \tilde{\mathbf{B}}_{i+1/2} \cdot (\mathbf{W}_{i+1}^n - \mathbf{W}_i^n) \right), \end{aligned} \quad (2.37)$$

with

$$\mathbf{F}_{i+1/2} = \frac{1}{2} (\mathbf{F}(\mathbf{W}_i^n) + \mathbf{F}(\mathbf{W}_{i+1}^n)) - \frac{1}{2} |\tilde{\mathcal{A}}_{i+1/2}| \cdot (\mathbf{W}_{i+1}^n - \mathbf{W}_i^n). \quad (2.38)$$

This scheme is thus the natural extension of the  $Q$ -scheme of Roe to system (2.24). We will also consider the extension of the  $Q$ -scheme of van Leer, whose expression is again (2.37)-(2.38) but where the matrix  $\mathcal{A}_{i+1/2}$  is now calculated using the intermediate state

$$\tilde{\mathbf{W}}_{i+1/2}^n = \frac{\mathbf{W}_i^n + \mathbf{W}_{i+1}^n}{2}. \quad (2.39)$$

A straightforward calculation shows that both Roe and van Leer schemes are consistent.

At first glance it may seem that a centered approximation of the derivatives appearing in the source terms is performed in (2.37) but, in fact, these derivatives are upwinded with the non-diagonal blocks of the matrix appearing in the numerical viscosity term, while the flux is upwinded with the diagonal blocks.

Finally, observe that in the case where the matrices  $B_{j,k}$  are constant, the scheme obtained coincides with the application of a standard  $Q$ -scheme to the system of conservation laws (2.25). Moreover, if the flux is also linear, the scheme coincides with the  $CIR$  scheme. Therefore, the scheme is linearly stable, under the usual  $CFL$  condition.

## 2.2. A numerical scheme for the complete system

In this section we consider the general problem (2.20). For solving this system we propose a  $Q$ -scheme with upwinding of the source term  $\Phi$  using the techniques introduced in [26, 27]. Such a scheme reads as follows:

$$\begin{aligned} \mathbf{W}_i^{n+1} &= \mathbf{W}_i^n + \frac{\Delta t}{\Delta x} (\mathbf{F}_{i-1/2} - \mathbf{F}_{i+1/2}) \\ &+ \frac{\Delta t}{2\Delta x} \left( \tilde{\mathbf{B}}_{i-1/2} \cdot (\mathbf{W}_i^n - \mathbf{W}_{i-1}^n) + \tilde{\mathbf{B}}_{i+1/2} \cdot (\mathbf{W}_{i+1}^n - \mathbf{W}_i^n) \right) \\ &+ \frac{\Delta t}{\Delta x} \left( \mathbf{P}_{i-1/2}^+ \tilde{\Phi}_{i-1/2} + \mathbf{P}_{i+1/2}^- \tilde{\Phi}_{i+1/2} \right), \end{aligned} \quad (2.40)$$

with

$$\begin{aligned} \mathbf{F}_{i+1/2} &= \frac{1}{2} (\mathbf{F}(\mathbf{W}_i^n) + \mathbf{F}(\mathbf{W}_{i+1}^n)) - \frac{1}{2} \left| \tilde{\mathcal{A}}_{i+1/2} \right| \cdot (\mathbf{W}_{i+1}^n - \mathbf{W}_i^n), \\ \tilde{\Phi}_{i+1/2} &= \Phi \left( x_{i+1/2}, \tilde{\mathbf{W}}_{i+1/2}^n \right), \end{aligned} \quad (2.41)$$

being  $\tilde{\mathbf{W}}_{i+1/2}^n$  the intermediate state corresponding to the particular scheme, and

$$\mathbf{P}_{i+1/2}^\pm = \frac{1}{2} \mathbf{K}_{i+1/2} (\mathbf{Id} \pm \text{sgn}(\Lambda_{i+1/2})) \mathbf{K}_{i+1/2}^{-1}, \quad (2.42)$$

where

$$\text{sgn}(\Lambda_{i+1/2}) = \begin{bmatrix} \text{sgn}(\lambda_{1,i}) & & 0 \\ & \ddots & \\ 0 & & \text{sgn}(\lambda_{N,i}) \end{bmatrix}.$$

Observe that the projection matrices used to upwind the source terms are obtained from the global matrix  $\tilde{\mathcal{A}}_{i+1/2}$  so that the eigenvalues of the system are taken into account.

## 2.3. The $CFL$ condition and Harten regularization

Observe that in the deduction of the schemes a  $CFL$ -like requirement (2.30) has been imposed. In practice, we propose the following condition:

$$\frac{\Delta x}{\Delta t} \leq \gamma \max \left\{ \left| \lambda_l^{i+1/2} \right|, \quad 1 \leq l \leq N, \quad 1 \leq i \leq M-1 \right\}, \quad (2.43)$$

with  $0 < \gamma \leq 1$ .

On the other hand, to prevent the numerical viscosity of the  $Q$ -schemes from vanishing when any of the eigenvalues of the matrices  $|\tilde{\mathcal{A}}_{i+1/2}|$  are zero, we apply the Harten regularization [14]. We define

$$\mathbf{F}_{i+1/2} = \frac{1}{2}(\mathbf{F}(\mathbf{W}_i^n) + \mathbf{F}(\mathbf{W}_{i+1}^n)) - \frac{1}{2}|\tilde{\mathcal{A}}_{i+1/2}|_\epsilon \cdot (\mathbf{W}_{i+1}^n - \mathbf{W}_i^n),$$

where

$$|\tilde{\mathcal{A}}_{i+1/2}|_\epsilon = \mathbf{K}_{i+1/2} |\Lambda_{i+1/2}|_\epsilon \mathbf{K}_{i+1/2}^{-1}.$$

Here,

$$|\Lambda_{i+1/2}|_\epsilon = \begin{bmatrix} |\lambda_1^{i+1/2}|_\epsilon & & 0 \\ & \ddots & \\ 0 & & |\lambda_N^{i+1/2}|_\epsilon \end{bmatrix},$$

with the notation:

$$|\lambda|_\epsilon = |\lambda| + 0.5 \left\{ \left(1 + \operatorname{sgn}(\epsilon - |\lambda|)\right) \left(\frac{\lambda^2 + \epsilon^2}{2\epsilon} - |\lambda|\right) \right\}, \quad (2.44)$$

being  $\epsilon$  a small parameter.

#### 2.4. A $Q$ -scheme with uncoupled upwinding

An alternative scheme can be obtained by applying a  $Q$ -scheme to each equation of (2.20). If source terms are treated numerically by using the techniques described in [3, 26, 27] the following scheme is obtained:

$$\begin{aligned} \mathbf{W}_i^{n+1} &= \mathbf{W}_i^n + \frac{\Delta t}{\Delta x} (\mathbf{F}_{i-1/2}^* - \mathbf{F}_{i+1/2}^*) \\ &+ \frac{\Delta t}{\Delta x} (\mathbf{P}_{i-1/2}^{*+} \tilde{\Phi}_{i-1/2} + \mathbf{P}_{i+1/2}^{*-} \tilde{\Phi}_{i+1/2}) \\ &+ \frac{\Delta t}{\Delta x} (\mathbf{P}_{i-1/2}^{*+} \cdot \tilde{\mathbf{B}}_{i-1/2} \cdot (\mathbf{W}_i^n - \mathbf{W}_{i-1}^n) + \mathbf{P}_{i+1/2}^{*-} \tilde{\mathbf{B}}_{i+1/2} \cdot (\mathbf{W}_{i+1}^n - \mathbf{W}_i^n)), \end{aligned} \quad (2.45)$$

with

$$\mathbf{F}_{i+1/2}^* = \frac{1}{2}(\mathbf{F}(\mathbf{W}_i^n) + \mathbf{F}(\mathbf{W}_{i+1}^n)) - \frac{1}{2}|\mathbf{A}_{i+1/2}| \cdot (\mathbf{W}_{i+1}^n - \mathbf{W}_i^n), \quad (2.46)$$

where

$$|\mathbf{A}_{i+1/2}| = \begin{bmatrix} |\tilde{A}_1^{i+1/2}| & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & |\tilde{A}_K^{i+1/2}| \end{bmatrix}. \quad (2.47)$$

In the former formula,  $\tilde{A}_j^{i+1/2}$  represents the  $j$ th diagonal block of the matrix  $\tilde{\mathbf{A}}_{i+1/2}$ . Let  $\Lambda_j^{i+1/2}$  be the  $N_j \times N_j$  diagonal matrix whose coefficients are the eigenvalues of  $\tilde{A}_j^{i+1/2}$ , and  $Q_j^{i+1/2}$  a  $N_j \times N_j$  a regular matrix such that

$$\tilde{A}_j^{i+1/2} = Q_j^{i+1/2} \Lambda_j^{i+1/2} (Q_j^{i+1/2})^{-1}.$$

With these notations, the projection matrices appearing in (2.45) are defined as

$$\mathbf{P}_{i+1/2}^{*\pm} = \begin{bmatrix} P_{i+1/2,1}^{*\pm} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_{i+1/2,K}^{*\pm} \end{bmatrix}, \quad (2.48)$$

where

$$P_{i+1/2,j}^{*\pm} = \frac{1}{2} Q_j^{i+1/2} \left( \mathbf{Id}_{N_j} \pm \text{sgn}(\Lambda^{i+1/2})_j \right) (Q_j^{i+1/2})^{-1}.$$

This scheme, must solve  $K$  spectral problems of order  $N_1, \dots, N_K$  at each intercell, instead of a unique spectral problem of order  $N$ , as in (2.40). Therefore, its practical implementation can be easier and computationally less expensive. Unfortunately, in practice this scheme is useless because it may be unconditionally unstable. This situation occurs, for instance, if all the eigenvalues of the Jacobians of the fluxes have the same sign and at least one of the eigenvalues of the global matrix has the opposite sign. To see this, let us apply the scheme to a linear problem with  $\Phi = 0$ , so that all the matrices in (2.40) do not depend on  $i + 1/2$  and the last term of the right-hand side vanishes. Let us suppose that all the eigenvalues of the matrices  $A_j$  have the same sign, say positive. In this case, the following equalities hold:

$$|\mathbf{A}| = \mathbf{A}, \quad \mathbf{P}^{*+} = I, \quad \mathbf{P}^{*-} = 0.$$

Using these equalities, some straightforward calculation, show that the scheme can be written as follows:

$$\mathbf{W}_i^{n+1} = \mathbf{W}_i^n - \frac{\Delta t}{\Delta x} \mathcal{A} \cdot (\mathbf{W}_i^n - \mathbf{W}_{i-1}^n),$$

where, as usual,  $\mathcal{A} = \mathbf{A} - \mathbf{B}$ . Now, we pass to characteristic variables, *i.e.*, we consider the matrix  $\mathbf{K}$  whose columns are the eigenvectors of  $\mathcal{A}$  and we make the change of basis:

$$\mathbf{V} = \mathbf{K}^{-1} \cdot \mathbf{W}.$$

In the new coordinates the scheme reads as follows:

$$\mathbf{V}_i^{n+1} = \mathbf{V}_i^n - \frac{\Delta t}{\Delta x} \Lambda \cdot (\mathbf{V}_i^n - \mathbf{V}_{i-1}^n),$$

where  $\Lambda$  is the diagonal matrix of the eigenvalues of  $\mathcal{A}$ . The  $j$ th component of  $\mathbf{V}_i^{n+1}$  is then obtained with the formula:

$$v_{j,i}^{n+1} = v_{j,i}^n - \frac{\Delta t}{\Delta x} \lambda^j (v_{j,i}^n - v_{j,i-1}^n).$$

Clearly, the scheme is unconditionally unstable whenever one of the eigenvalues  $\lambda^j$  is negative, as the upwinded is performed in the wrong sense.

### 3. NUMERICAL TESTS

#### 3.1. Test problem 1

The goal of this test is to validate the numerical scheme obtained in Section 2.1 in a simple system of the form (2.24) with known solution. We consider a system composed by two coupled Burgers equations:

$$\begin{cases} \frac{\partial u_1}{\partial t} + \frac{1}{2} \frac{\partial u_1^2}{\partial x} = -u_1 \frac{\partial u_2}{\partial x} \\ \frac{\partial u_2}{\partial t} + \frac{1}{2} \frac{\partial u_2^2}{\partial x} = -u_2 \frac{\partial u_1}{\partial x}, \end{cases} \quad (3.49)$$

which is of the form (2.24) with:

$$\begin{aligned} F_1(u) = F_2(u) &= \frac{1}{2}u^2, \\ B_{1,2} &= -u_1, \quad B_{2,1} = -u_2. \end{aligned}$$

The eigenvalues of the system are:

$$\lambda_1 = 0, \quad \lambda_2 = u_1 + u_2.$$

In the domain

$$\{(u_1, u_2), u_1 + u_2 > 0\},$$

the system is hyperbolic, with a characteristic field linearly degenerated and the other genuinely nonlinear [13].

Given  $x_0, (a_1, a_2), (b_1, b_2)$  such that

$$a_1 + a_2 = b_1 + b_2 > 0,$$

the function

$$\mathbf{W}_{x_0}(x, t) = \begin{cases} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} & \text{if } x < x_0, \\ \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} & \text{if } x > x_0, \end{cases} \quad (3.50)$$

is a weak solution of (3.49), representing a steady contact discontinuity placed at  $x = x_0$  connecting two constant states.

The  $Q$ -scheme of van Leer (2.37), (2.38) and (2.39) is applied to (3.49) with the initial condition

$$\mathbf{W}_{50}(x, 0) = \begin{cases} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} & \text{if } x < 50, \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \text{if } x > 50. \end{cases}$$

In this example,  $\Delta x = 1.0$  and  $CFL = 1.0$ . Figure 1 depicts the numerical solution obtained at any time  $t^n$ : the exact solution is captured.

### 3.2. Test problem 2

The goal of this test is to illustrate the instability of the scheme with uncoupled upwinding in the case of the two-layer shallow-water system (1.1). Let us consider system (1.1) with constant depth  $H$ . This problem can be written under the formulation (2.24) with  $K = 2$ ,  $N_1 = N_2 = 2$ ,  $N = 4$  and with  $\mathbf{W}$  given by (1.11). Finally,

$$\mathbf{G}(x, \mathbf{W}, \mathbf{W}_x) = \mathbf{B}(x, \mathbf{W}) \cdot \mathbf{W}_x,$$

with

$$\mathbf{B}(x, \mathbf{W}) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -gh_1 & 0 \\ 0 & 0 & 0 & 0 \\ -rgh_2 & 0 & 0 & 0 \end{bmatrix}.$$

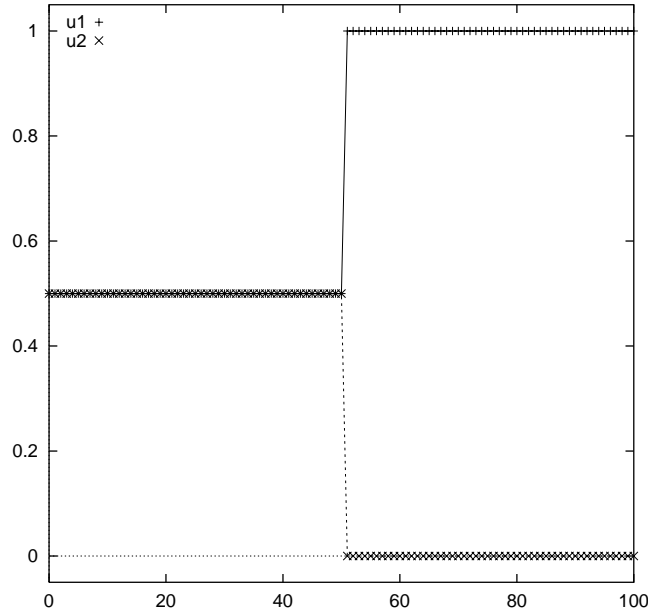


FIGURE 1. Numerical solution obtained with the  $Q$ -scheme of Van-Leer applied to (3.49) with initial data  $\mathbf{W}_{50}(x, 0)$ .

We have applied the  $Q$ -scheme of van Leer (2.40), (2.41) and (2.39) to (1.1). The calculation of the matrices  $|\mathcal{A}_{i+1/2}|$  demands the resolution of the algebraic equation (1.14). This has been done as follows: first, Newton’s method is applied to (1.14) with (1.15) as initial guess. Once the first two roots are found, the fourth-degree polynomial is deflected. The two roots of the resulting second-degree polynomial (if they are real) are used again as an initial guess for Newton’s method applied to (1.14). With this choice of initial guess, Newton’s method converges rapidly (one or two iterations are needed in the numerical tests performed). Once the eigenvalues of (1.13) are approximated, the calculation of their associated eigenvectors is performed. The matrix–vector and inverse of matrix computations are performed by using the C++-library *newmat09*.

The following initial conditions are considered:

$$h_1(x, 0) = \begin{cases} 0.5 & \text{if } x < 50 \\ 0.55 & \text{if } x > 50 \end{cases}, \quad q_1(x, 0) = \begin{cases} 1.25 & \text{if } x < 50 \\ 1.375 & \text{if } x > 50 \end{cases}; \quad (3.51)$$

$$h_2(x, 0) = \begin{cases} 0.5 & \text{if } x < 50 \\ 0.45 & \text{if } x > 50 \end{cases}, \quad q_2(x, 0) = \begin{cases} 1.25 & \text{if } x < 50 \\ 1.125 & \text{if } x > 50 \end{cases}. \quad (3.52)$$

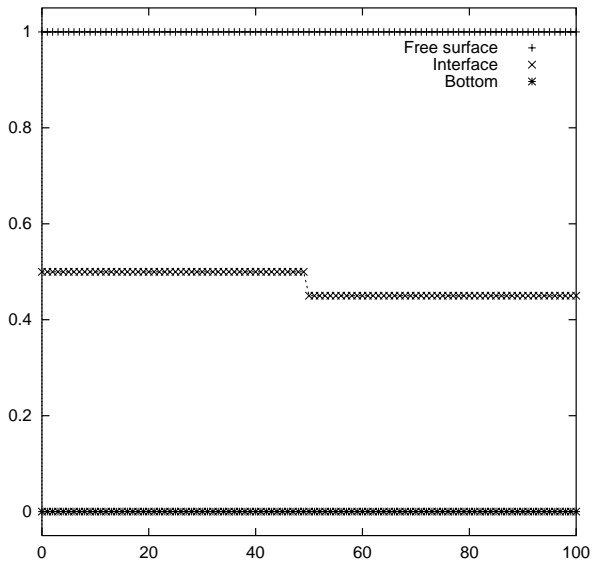
At the boundaries two fictitious states are introduced:

$$\mathbf{W}_0^n = \mathbf{W}_1^n, \quad \mathbf{W}_{M+1}^n = \mathbf{W}_M^n,$$

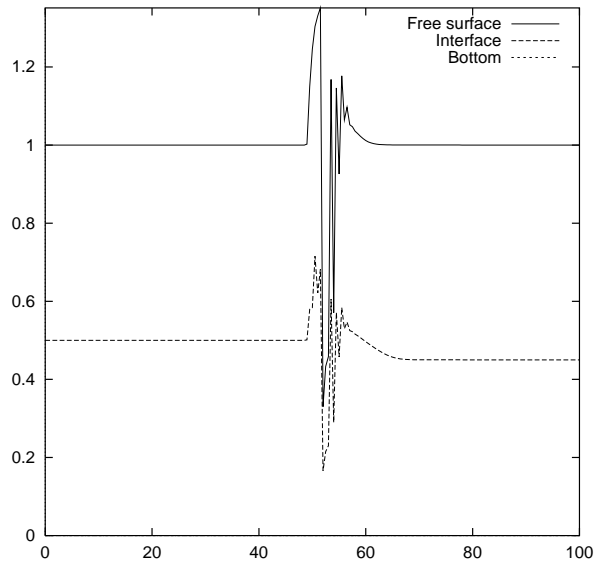
in order to calculate the average states  $\widetilde{\mathbf{W}}_{1/2}^n$  and  $\widetilde{\mathbf{W}}_{M+1/2}^n$  (see [25]).

In the simulations presented here  $r = 0.98$ ,  $\Delta x = 1.0$  and  $CFL = 1.0$  (i.e.,  $\gamma = 1.0$  in (2.43)). Figure 2(a) shows the free surface and the interface corresponding to the initial state. Figures 2(c) and 2(d) depict the interface at time  $t = 5$  and  $t = 10$ , respectively.

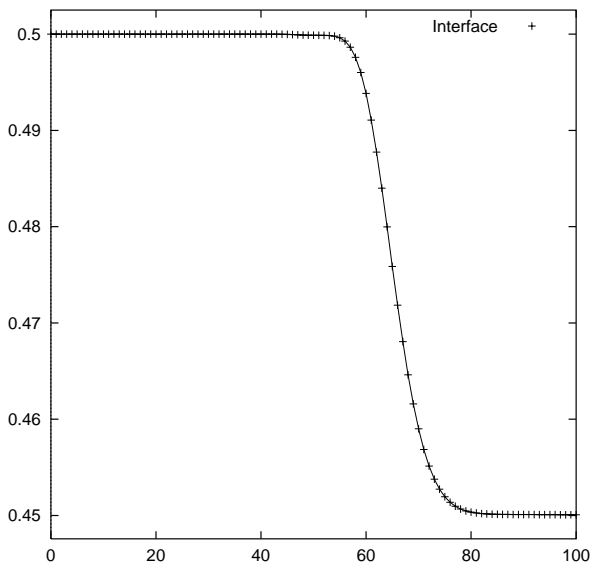
Next, we have applied the scheme (2.45)-(2.46) to the same problem. The choice of the initial condition has been done in order to have the situation discussed in Section 2.4 above: the eigenvalues of the Jacobians of the



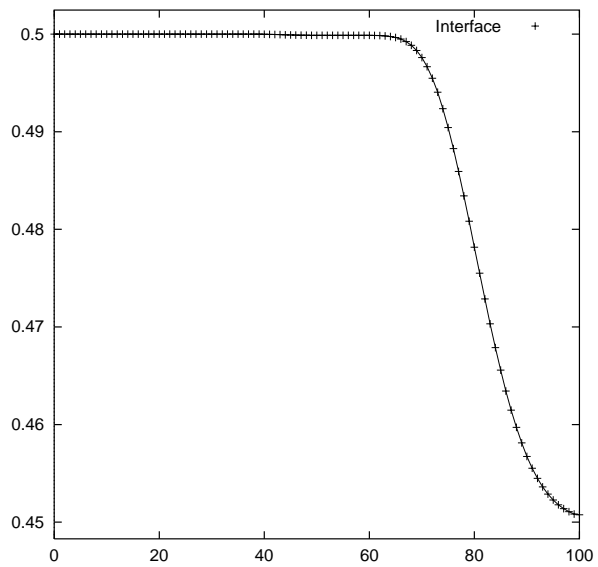
(a) Free surface and interface at  $t = 0$ .



(b) Free surface and interface at  $t = 5$  using the scheme given by (2.45)-(2.46).



(c) Interface at  $t = 5$  using the scheme given by (2.40)-(2.41).



(d) Interface at  $t = 10$  using the scheme given by (2.40)-(2.41).

FIGURE 2. Numerical solution for equation (1.1) with initial conditions (3.51), (3.52), using schemes (2.45)-(2.46) and (2.40)-(2.41).



fluxes of both layers are all of them positives, but the global matrix has a negative eigenvalue. Furthermore, we have applied the scheme (2.45)-(2.46) for a variety of  $CFL$  conditions ranging from 1 to  $10^{-4}$ . In all the cases instabilities appear at the free surface and interface. In Figure 2(b) the free surface and interface at time  $t = 5$  obtained with  $CFL = 10^{-4}$  are shown.

### 3.3. Test problem 3: Farmer and Armi steady solutions

Armi and Farmer studied the stationary solutions of two-layer exchange flows through contractions (in [1] and [2]) and over sills (in [1] and [9]). In [9] they also considered flows through the combination of a sill and contraction. Their model is based on the Bernoulli equations under the assumption of rigid lid. They parameterize the flows in terms of the internal Froude numbers for each layer,  $F_i$ , and possible solutions to the steady two-layer exchange problem are shown as curves in the Froude-number plane  $(F_1^2, F_2^2)$ . This parameterization is done by the deduction of a dimensionless Bernoulli equation expressed in terms of  $F_1^2$  and  $F_2^2$  that, for the case of flows over a sill, can be written as:

$$\Delta H''(q_1')^{-\frac{2}{3}} = F_1^{-\frac{2}{3}} + \frac{1}{2}F_1^{\frac{4}{3}} - \frac{1}{2}q_r^{-\frac{2}{3}}F_2^{\frac{4}{3}}, \quad (3.53)$$

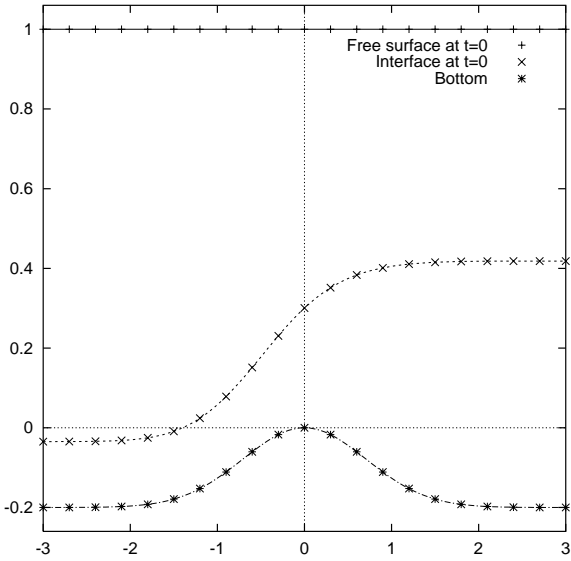
where  $q_r = q_1/q_2$  is the ratio of flow rates at each layer and  $q_1'$  is the dimensionless first layer flow (that is constant, as only steady solutions are considered by these authors). The term  $\Delta H''$  is related to the dimensionless energy difference between the two layers. In the absence of friction, mixing or any other losses of energy, this quantity, and consequently  $\Delta H''(q_1')^{-\frac{2}{3}}$ , is conserved. Therefore, the curves verifying equation (3.53) with constant  $\Delta H = \Delta H''(q_1')^{-\frac{2}{3}}$  represent the solutions of the model in the  $(F_1^2, F_2^2)$  plane (where dissipative processes are not considered). These solutions haven been plotted and discussed in detail by Farmer and Armi in [1] and [9]. The aim of this section is to compare the numerical solutions obtained using the scheme given by (2.40)-(2.41) for flows over a sill with the approximate steady solutions given by Farmer and Armi (F&A hereafter). The numerical solutions are obtained by iterating in time until reaching a steady state. The procedure has been the following: a F&A steady solution corresponding to a given value of  $\Delta H$ , say  $\Delta H_0$ , is taken as initial condition. Then boundary conditions corresponding to another steady state solution are imposed. More precisely, the input flow at the right end of the channel for the lower layer and the input flow at the left side for the upper layer are imposed. The numerical solution is expected to reach a steady state close to the corresponding F&A solution. Once this steady state is reached, the procedure is repeated to obtain a new stationary state. This procedure has been done for different values of  $\Delta H$  and  $q_r$ . In all the cases the degree of agreement between the numerical and the F&A solutions is excellent. As an example, we show a numerical experiment consisting in taking as initial conditions F&A's steady solution for  $q_r = 1$  and  $\Delta H = 2$ . The boundary conditions imposed the flow corresponding to the  $\Delta H = 1.8$  F&A's steady solution, being the new steady state rapidly reached.

Figure 3(a) depicts the free surface and interface for  $\Delta H = 2$  F&A solution (initial conditions). Figure 3(b) shows the final stationary state reached in the experiment presented here compared with the stationary approximate solution obtained by F&A and corresponding to the constant  $\Delta H = 1.8$  level. It can be observed that both solutions superposed.

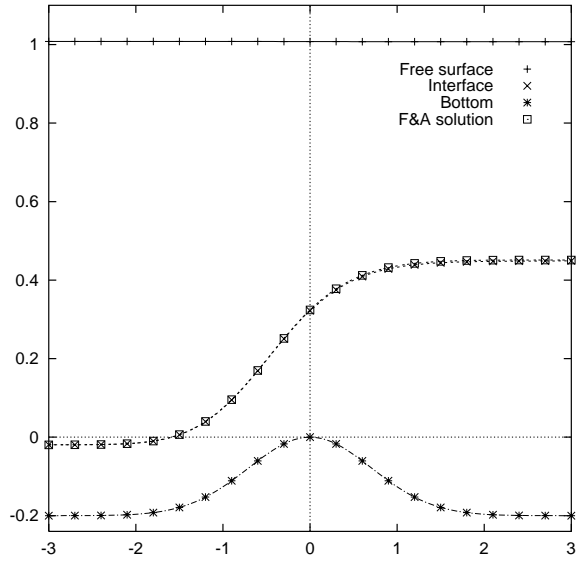
Figure 4(a) depicts the composite Froude number  $G^2$  along the channel at the final stationary state reached. This figure shows that the flow is supercritical to the left of the sill, critical at the sill and subcritical to the right. Figure 4(b) shows the four eigenvalues for the stationary solution. It can be observed how one of these eigenvalues becomes zero at the sill, being two eigenvalues positives and two negatives in the subcritical region, and three negative and one positive in the supercritical region. Note that the two internal eigenvalues are almost equal in the interval  $(-3, -1)$ , so that the flow is very close to conditions where Kelvin-Helmholtz instabilities can appear. Nevertheless, no problem of numerical instability has been observed.

### 3.4. Test problem 4: periodic barotropic forcing

In this section, the model is used to simulate flows over a sill forced by a periodic barotropic flow. A similar experiment was performed in [16], where a model for the study of time-dependent two-layer hydraulic flows

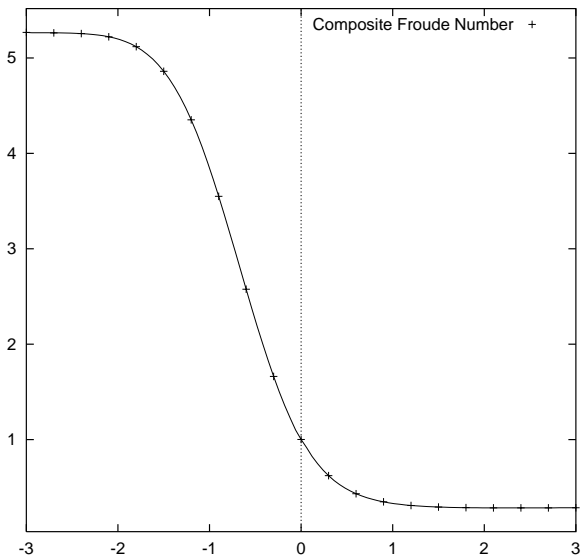


(a) Initial condition: F&A  $\Delta H = 2.0$  steady solution.

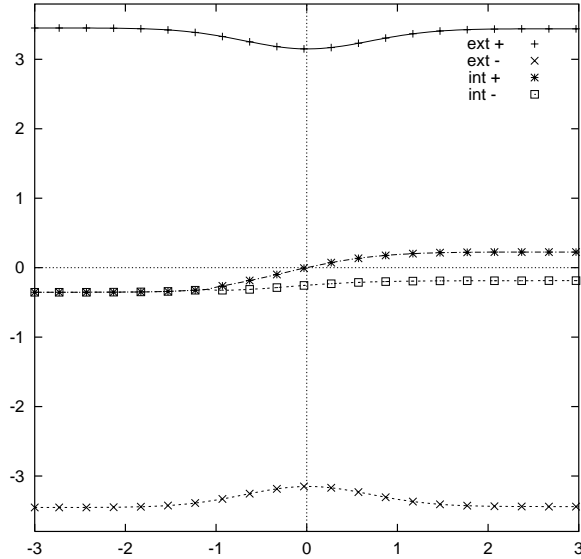


(b) Stationary state and F&A  $\Delta H = 1.8$  solution.

FIGURE 3. Transition from  $\Delta H = 2.0$  to  $\Delta H = 1.8$  stationary states and comparison with the solution found by F&A for  $\Delta H = 1.8$ .



(a) Composite Froude numbers along the channel.



(b) Eigenvalues along the channel.

FIGURE 4. Final stationary state reached ( $\Delta H = 1.8$  solution).

through straits is presented. Nevertheless, this author presents results for exchange flows through a contraction and through a combination of a sill and contraction, but not for flows over a single sill.

As initial conditions the steady state corresponding to the  $\Delta H = 2.0$  F&A steady solution is considered. At the boundaries,  $q_1$  and  $q_2$  are imposed only on the right border. The barotropic transport considered is periodic with zero time mean and given by  $q_b(t) = q_{b_0} \sin(2\pi t/T)$ , where  $q_{b_0} = 1 \cdot 10^{-2}$  is the barotropic transport amplitude and  $T = 800$  is the period.

Figure 5(a) depicts the interface along the channel at four points through the forcing period. The interface moves forth-downwards and back-upwards, with the barotropic flow but still similar in shape to the steady solutions. The periodic solution is obtained from the initial state without any prior adjustment period. In Figure 5(b) the same four transient solutions are shown in the Froude number space  $(F_1^2, F_2^2)$  (compare this with, for example, Fig. 11(b) in [1]).

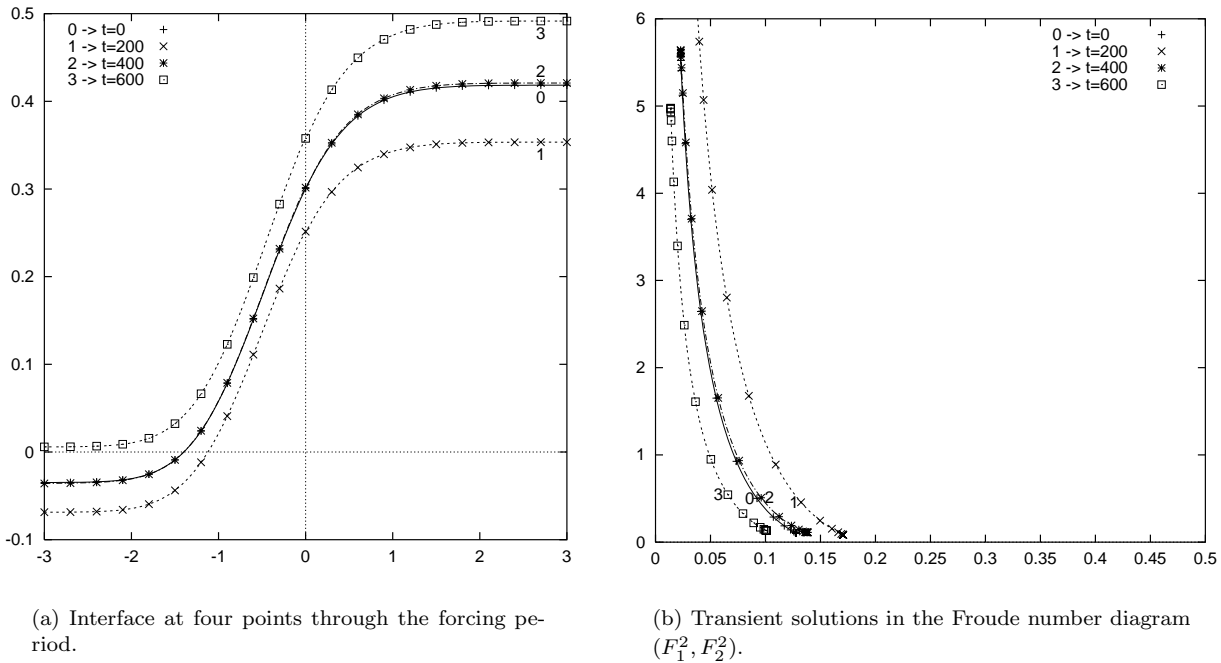


FIGURE 5. Exchange flow forced by a periodic barotropic flow.

#### 4. CONCLUDING REMARKS

The main contribution of this work is the improved discretization of source terms of systems of conservation laws that involve derivatives of the unknowns, as in the two-layer shallow water system. The efficiency of the proposed scheme is proved. It has also been pointed out that, in general, numerical schemes achieved by applying a  $Q$ -scheme to each separate conservation law of the system do not yield good results.

The present work is the first stage of a project whose final goal is to obtain a numerical model based on finite volume techniques well suited to analyze the relationship between the water exchange through the Strait of Gibraltar and the generation of internal waves induced by tidal effects.

In order to obtain the model we are carrying out, in collaboration with E. Vázquez-Cendón (University of Santiago de Compostela), the following research activities:

- Dealing with channels with irregular geometry. In [22] the model studied here has been generalized in order to include breadth variations. This introduces new source terms in the equations that are of a similar nature that those written as  $\phi(x, \mathbf{W})$  in the present work. The case of a single contraction has already been undertaken in [5] and the results compared with the steady solutions given in [1] and [2], and also with the periodic solutions obtained by [16]. In [6], we analyze model results for flows through a combination of a sill and contraction. The following step will consist in considering a simplified geometry of the Strait of Gibraltar with two sills and one narrows (as in [8], for example). We are also working in adapting the model to include non-rectangular cross-sections.
- Implementation of a model consisting in two superposed 2D layers.
- Implementation of adaptive mesh techniques.

*Acknowledgements.* This research has been partially supported by the C.I.C.Y.T. (project MAR97-1055-CO2-01). We thank M. Elena Vázquez and Jesús García Lafuente for stimulating discussions.

## REFERENCES

- [1] L. Armi, The hydraulics of two flowing layers with different densities. *J. Fluid Mech.* **163** (1986) 27–58.
- [2] L. Armi and D. Farmer, Maximal two-layer exchange through a contraction with barotropic net flow. *J. Fluid Mech.* **164** (1986) 27–51.
- [3] A. Bermúdez and M.E. Vázquez, Upwind methods for hyperbolic conservation laws with source terms. *Computers and Fluids* **23** (1994) 1049–1071.
- [4] C. Berthon and F. Coquel, Travelling wave solutions of a convective diffusive system with first and second order terms in nonconservation form, in *Hyperbolic Problems: Theory, Numerics, Applications*, Vol. I of *Internat. Ser. Numer. Math.* **129**, Birkhäuser (1999) 47–54.
- [5] M.J. Castro, J. Macías and C. Parés, Simulation of two-layer exchange flows through a contraction with a finite volume shallow water model, in *Actas de las II Jornadas de Análisis de Variables y Simulación Numérica del Intercambio de Masas de Agua a través del Estrecho de Gibraltar*, Cádiz (2000) 205–221.
- [6] M.J. Castro, J. Macías and C. Parés, *Simulation of two-layer exchange flows through the combination of a sill and contraction with a finite volume shallow-water model*. Internal Journal 1610, group on “Differential Equations, Numerical Analysis and Applications”, University of Málaga (2000).
- [7] F. Coquel, K. El Amine, E. Godlewski, B. Perthame and P. Rascle, Une méthode numérique décentrée pour la résolution d’écoulements diphasiques. *C. R. Acad. Sci. Paris, Sér. I* **324** (1997) 717–723.
- [8] S.B. Dalziel, *Two-layer Hydraulics Maximal Exchange Flows*. Ph.D. thesis, University of Cambridge (1988).
- [9] D. Farmer and L. Armi, Maximal two-layer exchange over a sill and through a combination of a sill and contraction with barotropic flow. *J. Fluid Mech.* **164** (1986) 53–76.
- [10] P. García-Navarro and F. Alcrudo, Implicit and explicit TVD methods for discontinuous open channel flows, in *Proc. of the 2nd Int. Conf. on Hydraulic and Environmental Modelling of Coastal, Estuarine and River Waters*, R.A. Falconer, K. Shiono, and R.G.S. Matthew, Eds. **2** Ashgate (1992).
- [11] P. García-Navarro, F. Alcrudo and J.M. Savirón, 1D open channel flow simulation using TVD McCormack scheme. *J. Hydraul. Eng.* **118** (1992) 1359–1373.
- [12] A.E. Gill, *Atmosphere-Ocean Dynamics*, *Int. Geophys. Series* **30**, Springer-Verlag, San Diego (1982) 662 p.
- [13] E. Godlewski and P.A. Raviart, Numerical Approximation of Hyperbolic Systems of Conservation Laws, *Appl. Math. Sc.* **118**, Springer-Verlag, New York (1996).
- [14] A. Harten, On a class of high resolution total-variation-stable finite-difference schemes. *SIAM J. Numer. Anal.* **21** (1984) 1–23.
- [15] A. Harten, P. Lax and A. van Leer, On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Rev.* **25** (1983) 35–61.
- [16] K.R. Helfrich, Time-dependent two-layer hydraulic exchange flows. *J. Phys. Oceanogr.* **25** (1995) 359–373.
- [17] P.K. Kundu, *Fluid Mechanics*. Academic Press Inc., San Diego (1990) 638 p.
- [18] P.G. Lefloch and A.E. Tzavaras, Existence theory for the Riemann problem for non-conservative hyperbolic systems. *C. R. Acad. Sci. Paris, Sér. I* **323** (1996) 347–352.
- [19] P.G. Lefloch and A.E. Tzavaras, Representation of weak limits and definition of nonconservative products. *SIAM J. Math. Anal.* **30** (1999) 1309–1342.
- [20] P.L. Roe, Approximate Riemann solvers, parameter vectors and difference schemes. *J. Comput. Phys.* **43** (1981) 357–371.

- [21] P.L. Roe, Upwinding differenced schemes for hyperbolic conservation laws with source terms, in *Proc. of the Conference on Hyperbolic Problems*, C. Carasso, P.-A. Raviart, and D. Serre, Eds., Springer-Verlag, Berlin (1986) 41–51.
- [22] J.A. Rubal and M.E. Vázquez, Aplicación del método de volúmenes finitos y esquemas tipo Godunov a un modelo bicapa, in *Actas de las II Jornadas de Análisis de Variables y Simulación Numérica del Intercambio de Masas de Agua a través del Estrecho de Gibraltar*, Cádiz (2000) 223–239.
- [23] J.B. Schijf and J.C. Schonfeld, Theoretical considerations on the motion of salt and fresh water, in *Proc. of the Minn. Int. Hydraulics Conv.* Joint meeting IAHR and Hyd. Div. ASCE., Sept. 1953 (1953) 321–333.
- [24] J.J. Stoker, *Water Waves*. Interscience, New York (1957).
- [25] E.F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics. A Practical Introduction*. Springer-Verlag, Berlin (1997).
- [26] M.E. Vázquez-Cendón, *Estudio de Esquemas Descentrados para su Aplicación a las leyes de Conservación Hiperbólicas con Términos Fuente*. PhD thesis, Universidad de Santiago de Compostela (1994).
- [27] M.E. Vázquez-Cendón, Improved treatment of source terms in upwind schemes for the shallow water equations in channels with irregular geometry. *J. Comp. Physics* **148** (1999) 497–526.