INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2001; **37**: 209–248 (DOI: 10.1002/fld.175)

Efficient construction of high-resolution TVD conservative schemes for equations with source terms: application to shallow water flows

J. Burguete and P. García-Navarro*

Fluid Mechanics, CPS, University of Zaragoza, Zaragoza, Spain

SUMMARY

High-resolution total variation diminishing (TVD) schemes are widely used for the numerical approximation of hyperbolic conservation laws. Their extension to equations with source terms involving spatial derivatives is not obvious. In this work, efficient ways of constructing conservative schemes from the conservative, non-conservative or characteristic form of the equations are described in detail. An upwind, as opposed to a pointwise, treatment of the source terms is adopted here, and a new technique is proposed in which source terms are included in the flux limiter functions to get a complete second-order compact scheme. A new correction to fix the entropy problem is also presented and a robust treatment of the boundary conditions according to the discretization used is stated. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS: boundaries; conservative schemes; entropy; flux limiter; high resolution; shallow water; source terms; TVD

1. INTRODUCTION

For many practical applications it is accepted that the unsteady flow of water in a one-dimensional approach is governed by the shallow water or St. Venant equations. These represent the conservation of mass and momentum along the direction of the main flow. In the field of computational hydraulics, where modelling can be dominated by the effects not only of source terms but also of quantities that vary spatially but independently of the flow variables, it has traditionally been difficult to have only one method able to reproduce automatically any general situation. The numerical modelling of unsteady flow in rivers is a complicated task and the difficulties grow as the pretensions to obtain better quality or more general solutions do.

* Correspondence to: Area de Mecanica de Fluidos, CPS, Universidad de Zaragoza, C/. Maria de Luna 3, 50015 Zaragoza, Spain.

Received September 2000
Accepted January 2001

Copyright © 2001 John Wiley & Sons, Ltd.

There exist good methods developed to deal with gas dynamics problems (Euler equations), able to cope with complex systems of discontinuities and shock waves [1,2]. Among them, flux difference splitting methods are widely used for the numerical approximation of homogeneous conservation laws where the flux depends only on the conservative variables. In many practical situations common in hydraulics, not only are source terms part of the mathematical model, but the flux can vary spatially, even when the conservative variables do not. The extension of the basic techniques to cope with these two situations keeping their original properties is not straightforward and is addressed in this work.

Many upwind schemes have been reported successful for flow in channels [3,4]. However, their application to river flow is not so common in the literature and their adaptation to river hydraulics is hindered by the irregular topography. The presence of extreme slopes, high roughness and strong changes in the irregular geometry represent a difficulty that can lead to important numerical errors presumably arising from the source terms of the equations.

In this paper, the main focus is on the construction of methods that address the issue of solving conservation laws with source terms and a general flux function that can vary independently of the conserved variables. The discretization is constructed in a manner that retains the conservative character of the schemes seeking an exact balance between the flux gradient and the source terms discretization so that any balance between flux and source terms which exists as part of the underlying mathematical model (e.g. at the steady state) is maintained. This has been already addressed and applied in a variety of different situations. For example, Smolarkiewicz [5] has adapted his own MPDATA schemes to solve inhomogeneous equations arising from geophysical flows; LeVeque [6] has incorporated the modelling of source terms for shallow water flows within his wave-propagation algorithm; and Roe's scheme [7] has been modified by a number of authors to include source terms, the research of Glaister [3], Vázquez-Cendón [8], Bermúdez and Vázquez [9] and Garcia-Navarro and Vazquez Cendon [10] being of particular relevance to this work. In the papers discussing Roe's scheme, the discretization of the source terms has been constructed according to that of the numerical fluxes to ensure that in the absence of source terms, the conservative fluxes are retrieved for accurate modelling of discontinuous solutions.

The intention of this paper is to accomplish the same but replacing the numerical flux concept, usual in finite volume formulations, by a different and better adapted point of view. Due to the fact that the numerical source integral cannot, in general, be written as a difference, it is impossible to allow it to be included within the numerical flux and the balance which is sought between flux derivatives and sources in the flux-based schemes can only be obtained locally by balancing numerical flux differences through the edges of a control volume, and integrals of the source terms. The idea is presented not only in the first-order case but also in the presence of flux limited high-resolution corrections. Traditionally, total variation diminishing (TVD) schemes have been constructed under certain condition derived from the homogeneous system of conservation laws. The TVD property can then by violated in the presence of important source terms. This work is also concerned with constructing high resolution schemes for the full equations so that, although the TVD property is not guaranteed in the presence of source terms, with an upwind discretization of these, their limitation is required in order to maintain a balance with the fluxes. Furthermore, the introduction of the source term in the definition of the limiting functions is necessary to preserve second-order of accuracy. On the

Copyright © 2001 John Wiley & Sons, Ltd.

other hand, the mathematical formulation of the shallow water equations is revisited and the optimal form of the equations is presented for the case where the flux of the conserved variables depends on a spatially varying quantity which is independent of the solution.

1.1. The equations

We are interested in solving as efficiently as possible one-dimensional hyperbolic systems with source terms. In a general conservative form

$$\frac{\partial \vec{u}(x,t)}{\partial t} + \frac{\mathrm{d}\vec{F}(x,\vec{u})}{\mathrm{d}x} = \vec{H}(x,\vec{u}) \tag{1.1}$$

where $\vec{u}$ is the vector of conserved variables, $\vec{F}$ the vector of fluxes and $\vec{H}$ that of source terms. Our interest is led by the numerical modelling of one-dimensional shallow water flows of practical application in hydraulics such as river flows. In that case

$$\vec{u} = \begin{pmatrix} A \\ Q \end{pmatrix}, \qquad \vec{F} = \begin{bmatrix} Q \\ \dfrac{Q^2}{A} + gI_1 \end{bmatrix}, \qquad \vec{H} = \begin{pmatrix} 0 \\ g[I_2 + A(S_0 - S_{\mathrm{f}})] \end{pmatrix}$$

where Q is the discharge, A is the wetted cross section, g is the acceleration of gravity and S_0 is the bed slope, I_1 and I_2 account for pressure forces

$$I_1(x,A) = \int_0^{h(x,A)} [h(x,A) - z]\sigma(x,z)\,\mathrm{d}z, \qquad I_2(x,A) = \int_0^{h(x,A)} [h(x,A) - z]\frac{\partial \sigma(x,z)}{\partial x}\,\mathrm{d}z$$

(h the water depth and σ the channel width at a position z from the bottom) and S_{f} is associated with bed friction and represented by the empirical Manning law

$$S_{\mathrm{f}} = \frac{n^2 Q^2 P^{4/3}}{A^{10/3}}$$

where n is the coefficient of bed roughness and P the wetted perimeter.

It is important to remark that in the conservative form (1.1) the total derivative $\mathrm{d}\vec{F}/\mathrm{d}x$ is used to represent the increments due to the pure spatial variations in x and those due to the variations of the conserved variable $\vec{u}$, whereas the partial derivative is reserved to represent only the variation due to the x with $\vec{u}$ constant. The difference between the variations due to both $\vec{u}$ and x from those due to only one of the variables is subtle but significant. When the dependence of a general function is $f = f(x, u)$, the relation between total and partial derivatives is

$$\frac{\mathrm{d}f(x,u)}{\mathrm{d}x} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u}\frac{\partial u}{\partial x}$$

Copyright © 2001 John Wiley & Sons, Ltd.

and the discrete increments $\delta f/\delta x$ approach actually the total derivative $\mathrm{d}f/\mathrm{d}x$, and not the partial derivative $\partial f/\partial x$. For this reason, partial spatial derivatives of this type of functions should be avoided in the formulations. This is an important detail not stressed in the related literature that has proved decisive in the development of the schemes described in this work.

From the equations in conservative form (1.1), it is possible to pass to an associated non-conservative form using

$$\frac{\mathrm{d}\vec{F}(x,\vec{u})}{\mathrm{d}x}=\frac{\partial\vec{F}(x,\vec{u})}{\partial x}+\frac{\partial\vec{F}(x,\vec{u})}{\partial\vec{u}}\frac{\partial\vec{u}}{\partial x}=\frac{\partial\vec{F}(x,\vec{u})}{\partial x}+\mathbf{J}(x,\vec{u})\frac{\partial\vec{u}}{\partial x}$$

where

$$\mathbf{J}=\frac{\partial\vec{F}}{\partial\vec{u}}$$

is the Jacobian matrix of the original system. Redefining the source term as

$$\vec{H}'(x,\vec{u})=\vec{H}(x,\vec{u})-\frac{\partial\vec{F}(x,\vec{u})}{\partial x}$$

the non-conservative form is obtained

$$\frac{\partial\vec{u}(x,t)}{\partial t}+\mathbf{J}(x,\vec{u})\frac{\partial\vec{u}(x,t)}{\partial x}=\vec{H}'(x,\vec{u}) \tag{1.2}$$

In the shallow water system of equations, the following holds:

$$\frac{\mathrm{d}h}{\mathrm{d}x}=\frac{\partial h}{\partial x}+\frac{1}{B}\frac{\partial A}{\partial x}$$

$$\frac{\mathrm{d}I_1}{\mathrm{d}x}=\frac{\partial I_1}{\partial x}+\frac{\partial I_1}{\partial A}\frac{\partial A}{\partial x}=I_2+A\frac{\partial h}{\partial x}+\frac{A}{B}\frac{\partial A}{\partial x}=I_2+A\frac{\mathrm{d}h}{\mathrm{d}x}$$

and the Jacobian and new source terms of the non-conservative formulation (1.2) are

$$J=\begin{pmatrix}0 & 1\\ c^2-v^2 & 2v\end{pmatrix},\qquad \vec{H}'=\begin{pmatrix}0\\ gA[S_0-S_\mathrm{f}-(\mathrm{d}h/\mathrm{d}x)+(1/B)(\mathrm{d}A/\mathrm{d}x)]\end{pmatrix} \tag{1.3}$$

with B the width at the free surface, $c=\sqrt{g(A/B)}$ the celerity of the infinitesimal surface wave and $v=Q/A$ the mean fluid velocity. Note that the standard formulations of previous works based on these equations do not contain the $(1/B)(\mathrm{d}A/\mathrm{d}x)$ term. It appears as a consequence of the necessity of avoiding partial x derivatives in order to ensure the success of the discretization, as pointed above.

Copyright © 2001 John Wiley & Sons, Ltd.

It is now convenient to develop the characteristic form of the equations, important for the correct formulation of upwind schemes and boundary conditions. This form is obtained from a diagonalization of the Jacobian in (1.2). Calling $\mathbf{P}$ and $\mathbf{P}^{-1}$ the matrices that make diagonal $\mathbf{J}$, and $\mathbf{\Lambda}$ the resulting diagonal matrix

$$\mathbf{J} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}, \qquad \mathbf{\Lambda} = \mathbf{P}^{-1}\mathbf{J}\mathbf{P}$$

The matrix $\mathbf{\Lambda}$ is formed by the eigenvalues of $\mathbf{J}$, and $\mathbf{P}$ is constructed with its eigenvectors. Let $\vec{w}$ be the set of variables (characteristics variables) that verify

$$\mathrm{d}\vec{u} = \mathbf{P}\,\mathrm{d}\vec{w}, \qquad \mathrm{d}\vec{w} = \mathbf{P}^{-1}\,\mathrm{d}\vec{u}$$

Then

$$\frac{\partial \vec{w}(x, t)}{\partial t} + \mathbf{\Lambda}(x, \vec{w})\frac{\partial \vec{w}(x, t)}{\partial x} = \mathbf{P}^{-1}(x, \vec{w})\vec{H}'(x, \vec{w}) \tag{1.4}$$

For the shallow water equations, the above matrices are

$$\mathbf{P} = \begin{pmatrix} 1 & 1 \\ v + c & v - c \end{pmatrix}, \qquad \mathbf{P}^{-1} = \frac{1}{2c}\begin{pmatrix} c - v & 1 \\ c + v & -1 \end{pmatrix}, \qquad \mathbf{\Lambda} = \begin{pmatrix} v + c & 0 \\ 0 & v - c \end{pmatrix}$$

2. CONSERVATIVE SCHEMES

The conservation law (1.1) contains an important physical meaning. A spatial integration

$$\int_0^L \left(\frac{\partial \vec{u}}{\partial t} + \frac{\mathrm{d}\vec{F}}{\mathrm{d}x}\right)\mathrm{d}x = \int_0^L \vec{H}\,\mathrm{d}x \;\Rightarrow\; \int_0^L \frac{\partial \vec{u}}{\partial t}\,\mathrm{d}x = \vec{F}_0 - \vec{F}_L + \int_0^L \vec{H}\,\mathrm{d}x \tag{2.1}$$

expresses that the time variation of the conserved variable in a given volume is equal to the difference between the incoming and the outgoing fluxes plus the contribution of the source term. When discretizing a conservation law of this kind, incorrect numerical approximations can lead to bad behaviour in the solution and unacceptable errors which make them useless techniques. Schemes properly approximating the conservation Equation (2.1) are called conservative schemes. Some ways of defining them are presented next.

2.1. Conservative schemes using a conservative formulation

The most common definition of a conservative scheme follows the structure

$$\frac{\Delta \vec{u}_i^n}{\Delta t} = \vec{H}_i^* - \frac{1}{\delta x}(\vec{F}^*_{i+(1/2)} - \vec{F}^*_{i-(1/2)}) \tag{2.2}$$

Copyright © 2001 John Wiley & Sons, Ltd.

where $\vec{H}^*$ is the numerical source of cell i and $\vec{F}^*$ is the numerical flux at cell interface $i+(1/2)$. They represent a suitable approximation to the true source and flux term in the equation so that the scheme gets the properties required (Figure 1). Schemes so defined will be conservative since they produce a good approximation of (2.1) cancelling the contributions of the flux at the grid interface, being the global variation of the conserved variable due only to the source terms and to the flux at the boundaries. Δ will be used for time increments $\Delta f_i^n = f_i^{n+1} - f_i^n$, and δ for spatial increments $\delta f_{i+(1/2)}^n = f_{i+1}^n - f_i^n$.

$$\sum_{i=1}^{N-1} \frac{\Delta \vec{u}_i^n}{\Delta t}\,\delta x = \frac{\Delta}{\Delta t} \sum_{i=1}^{N-1} \vec{u}_i^n \delta x \approx \frac{\partial}{\partial t} \int_{x_{1/2}}^{x_{N-(1/2)}} \vec{u}\,\mathrm{d}x$$

$$\sum_{i=1}^{N-1} [\vec{H}_i^* \delta x - (\vec{F}_{i+(1/2)}^* - \vec{F}_{i-(1/2)}^*)] = \sum_{i=1}^{N-1} [\vec{H}_i^* \delta x] + \vec{F}_{1/2}^* - \vec{F}_{N-(1/2)}^*$$

$$\approx \vec{F}_{1/2} - \vec{F}_{N-(1/2)} + \int_{x_{1/2}}^{x_{N-(1/2)}} \vec{H}\,\mathrm{d}x$$

A numerical flux $\vec{F}_i^T$ can also be defined at the grid nodes. The difference in this flux between two nodes can be decomposed into parts affecting the nodes on the left and right. Schemes so built follow

$$\delta \vec{F}_{i+(1/2)}^T = \vec{F}_{i+1}^T - \vec{F}_i^T = \delta \vec{F}_{i-(1/2)}^{\mathrm{R}} + \delta \vec{F}_{i+(1/2)}^{\mathrm{L}}$$

$$\frac{\Delta \vec{u}_i^n}{\Delta t} = \vec{H}_i^* - \frac{1}{\delta x} (\delta \vec{F}_{i+(1/2)}^{\mathrm{R}} + \delta \vec{F}_{i-(1/2)}^{\mathrm{L}}) \tag{2.3}$$

This also leads to conservative schemes since this form is equivalent to (2.2) Figure 2. It is possible to define a function $\vec{\Phi}$ so that

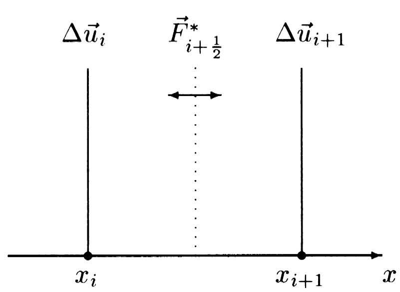

Figure 1. Adding up all the variable increments $\Delta \vec{u}_i$, the numerical flux contributions at the interfaces $\vec{F}_{i+(1/2)}^*$ cancel producing a conservative scheme.

Copyright © 2001 John Wiley & Sons, Ltd.

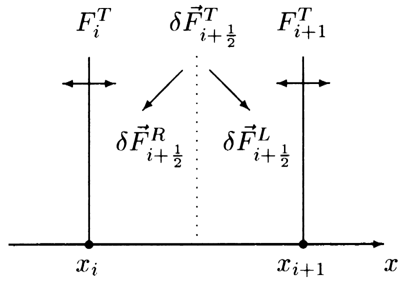

Figure 2. Adding up all the flux difference parts $\delta\vec{F}^R_{i+(1/2)}$ and $\delta\vec{F}^L_{i+(1/2)}$ the flux contributions in the grid nodes $\vec{F}^T_i$ cancel producing a conservative scheme.

$$\delta\vec{F}^{\mathrm{L}}_{i+(1/2)} = \frac{1}{2}\,\delta\vec{F}^T_{i+(1/2)} + \vec{\Phi}_{i+(1/2)}, \qquad \delta\vec{F}^{\mathrm{R}}_{i+(1/2)} = \frac{1}{2}\,\delta\vec{F}^T_{i+(1/2)} - \vec{\Phi}_{i+(1/2)}$$

hence from (2.3)

$$\frac{\Delta\vec{u}^n_i}{\Delta t} = \vec{H}^*_i - \frac{1}{\delta x}\left[\frac{1}{2}(\vec{F}^T_{i+1} - \vec{F}^T_i) - \vec{\Phi}_{i+1/2} + \frac{1}{2}(\vec{F}^T_i - \vec{F}^T_{i-1}) + \vec{\Phi}_{i-1/2}\right]$$

and the interface numerical flux can be defined

$$\vec{F}^*_{i+(1/2)} = \frac{1}{2}(\vec{F}^T_{i+1} + \vec{F}^T_i) - \vec{\Phi}_{i+(1/2)} = \vec{F}^T_i + \delta\vec{F}^{\mathrm{R}}_{i+(1/2)} = \vec{F}^T_{i+1} - \delta\vec{F}^{\mathrm{L}}_{i+(1/2)} \tag{2.4}$$

In addition, and following evidence from previous works [8,10,11], we consider a non-centred contribution of the source terms

$$\vec{H}^T_{i+(1/2)} = \vec{H}^{\mathrm{R}}_{i+(1/2)} + \vec{H}^{\mathrm{L}}_{i+(1/2)}$$

so that the following formulation for the conservative scheme is proposed:

$$\frac{\Delta\vec{u}^n_i}{\Delta t} = \left(\vec{H} - \frac{\delta\vec{F}}{\delta x}\right)^{\mathrm{L}}_{i-(1/2)} + \left(\vec{H} - \frac{\delta\vec{F}}{\delta x}\right)^{\mathrm{R}}_{i+(1/2)} \tag{2.5}$$

As an example, a first-order upwind discretization admits a decomposition like

$$\vec{F}^T_i = \vec{F}^n_i, \qquad \delta\vec{F}^{\mathrm{L}}_{i+(1/2)} = (\delta\vec{F}^+)^n_{i+(1/2)}, \qquad \delta\vec{F}^{\mathrm{R}}_{i+(1/2)} = (\delta\vec{F}^-)^n_{i+(1/2)}$$

where $\delta\vec{F}^+$ and $\delta\vec{F}^-$ are associated to positive and negative propagation velocities respectively, whereas second-order centred discretization admits a decomposition like

Copyright © 2001 John Wiley & Sons, Ltd.

$$\vec{F}_i^T = \vec{F}_i^n, \qquad \delta\vec{F}^{\mathrm{L}}_{i+(1/2)} = \delta\vec{F}^{\mathrm{R}}_{i+(1/2)} = \frac{1}{2}\,\delta\vec{F}^n_{i+(1/2)}$$

2.2. Conservative schemes using a non-conservative formulation

Conservative schemes can also be derived from the non-conservative formulation of the Equations (1.2). The advantage is that the latter form tends to be simpler to deal with than the conservative form. We need to establish the conditions under which schemes derived in this way are equivalent to Schemes (2.2) and (2.5) derived from the conservative equations. First of all, the following equality must hold at the discrete level:

$$\vec{G}_{i+(1/2)} \equiv \left(\vec{H} - \frac{\delta\vec{F}}{\delta x}\right)_{i+(1/2)} = \left(\vec{H}' - \mathbf{J}\,\frac{\delta\vec{u}}{\delta x}\right)_{i+(1/2)} \tag{2.6}$$

Note that this equality requires a non-pointwise treatment of source terms and is an extension of the Roe's average [2,7]. From (2.6) it follows that two equivalent forms of building conservative schemes with de-centred source terms are possible. Defining $\vec{G}$ as

$$\vec{G}_{i+(1/2)} \equiv \left(\vec{H} - \frac{\delta\vec{F}}{\delta x}\right)_{i+(1/2)} \tag{2.7}$$

the forms (2.2) and (2.5) are achieved, whereas defining this term like

$$\vec{G}_{i+(1/2)} \equiv \left(\vec{H}' - \mathbf{J}\,\frac{\delta\vec{u}}{\delta x}\right)_{i+(1/2)} \tag{2.8}$$

with the restriction (2.6), a conservative scheme is also obtained. In any case

$$\frac{\Delta\vec{u}_i^n}{\Delta t} = \vec{G}^{\mathrm{L}}_{i-(1/2)} + \vec{G}^{\mathrm{R}}_{i+(1/2)} \tag{2.9}$$

where the decomposition in left and right parts has to be defined in every numerical scheme.

In the particular case of the shallow water equations, condition (2.6) imposes the following equality:

$$\begin{pmatrix} 0 \\ g[I_2 + A(S_0 - S_{\mathrm{f}})] \end{pmatrix}_{i+(1/2)} - \frac{\delta}{\delta x}\begin{pmatrix} Q \\ (Q^2/A) + gI_1 \end{pmatrix}_{i+(1/2)}$$
$$= \begin{pmatrix} 0 \\ gA(S_0 - S_{\mathrm{f}} + (1/B)(\mathrm{d}A/\mathrm{d}x - \mathrm{d}h/\mathrm{d}x)) \end{pmatrix}_{i+(1/2)} - \begin{pmatrix} 0 & 1 \\ c^2 - v^2 & 2v \end{pmatrix}_{i+(1/2)} \frac{\delta}{\delta x}\begin{pmatrix} A \\ Q \end{pmatrix}_{i+(1/2)} \tag{2.10}$$

which considering that

Copyright © 2001 John Wiley & Sons, Ltd.

$$\left(\frac{\mathrm{d}h}{\mathrm{d}x}\right)_{i+(1/2)} \approx \left(\frac{\delta h}{\delta x}\right)_{i+(1/2)}, \qquad \left(\frac{1}{B}\frac{\mathrm{d}A}{\mathrm{d}x}\right)_{i+(1/2)} \approx \left(\frac{1}{B}\frac{\delta A}{\delta x}\right)_{i+(1/2)}$$

$$(I_2)_{i+(1/2)} \approx \left(\frac{\delta I_1}{\delta x} - A\frac{\delta h}{\delta x}\right)_{i+(1/2)}$$

reduces (2.10) to

$$-\frac{\delta}{\delta x}\left(\frac{Q^2}{A}\right)_{i+(1/2)} = \left[g\frac{A}{B}\frac{\delta A}{\delta x} - (c^2 - v^2)\frac{\delta A}{\delta x} - 2v\frac{\delta Q}{\delta x}\right]_{i+(1/2)}$$

and this is satisfied if the following average values are defined:

$$c_{i+(1/2)} = \sqrt{g\frac{A_{i+(1/2)}}{B_{i+(1/2)}}}, \qquad v_{i+(1/2)} = \frac{Q_{i+1}/\sqrt{A_{i+1}} + Q_i/\sqrt{A_i}}{\sqrt{A_{i+1}} + \sqrt{A_i}} \tag{2.11}$$

The choice of the discrete averages $A_{i+(1/2)}$, $B_{i+(1/2)}$, $(S_0)_{i+(1/2)}$ and $(S_f)_{i+(1/2)}$ is open but, in our work, this has not proved significant. The simplest arithmetic average has been applied.

2.3. Conservative schemes using a characteristic formulation

Conservative schemes based in the characteristic form of the equations are the basis of the wave decomposition in upwind schemes. From (1.4) it is possible to rewrite

$$\frac{\partial \vec{w}}{\partial t} = \mathbf{P}^{-1}\left(\vec{H}' - \mathbf{J}\frac{\partial \vec{u}}{\partial x}\right) = \mathbf{P}^{-1}\vec{G}$$

Then a discrete wave decomposition into left and right moving contributions can be done

$$(\mathbf{P}^{-1}\vec{G})_{i+(1/2)} = (\mathbf{\Omega}^{\mathrm{L}}\mathbf{P}^{-1}\vec{G})_{i+(1/2)} + (\mathbf{\Omega}^{\mathrm{R}}\mathbf{P}^{-1}\vec{G})_{i+(1/2)}$$

being $\mathbf{\Omega}^{\mathrm{L}}$ and $\mathbf{\Omega}^{\mathrm{R}}$ diagonal matrices to be defined in each numerical scheme. In order to ensure the conservative character of the scheme

$$(\mathbf{\Omega}^{\mathrm{L}} + \mathbf{\Omega}^{\mathrm{R}})_{i+(1/2)} = \mathbf{I} \tag{2.12}$$

Multiplication by $\mathbf{P}$ gives the final form for the discretization in terms of the conserved quantities

$$\frac{\Delta \vec{u}_i^n}{\Delta t} = (\mathbf{P}\mathbf{\Omega}^{\mathrm{L}}\mathbf{P}^{-1}\vec{G})_{i-(1/2)} + (\mathbf{P}\mathbf{\Omega}^{\mathrm{R}}\mathbf{P}^{-1}\vec{G})_{i+(1/2)} \tag{2.13}$$

Note that this discretization requires again a non-centred formulation of source terms, being equally possible (2.7) or (2.8) for $\vec{G}$. A pointwise treatment of the source terms can also be used, changing (2.13) to

Copyright © 2001 John Wiley & Sons, Ltd.

$$\frac{\Delta\vec{u}_i^n}{\Delta t} = \vec{H}_i^n - \left(\mathbf{P}\boldsymbol{\Omega}^{\mathrm{L}}\mathbf{P}^{-1}\frac{\delta\vec{F}}{\delta x}\right)_{i-(1/2)} - \left(\mathbf{P}\boldsymbol{\Omega}^{\mathrm{R}}\mathbf{P}^{-1}\frac{\delta\vec{F}}{\delta x}\right)_{i+(1/2)} \tag{2.14}$$

3. FIRST-ORDER UPWIND SCHEME

Upwind schemes are based on the idea of approximating the spatial derivatives by non-centred differences biased in the sense of propagation of information in the physical problem. In order to construct a general first-order scheme, the following is usually written

$$\Delta\vec{u}_i^n = \Delta t\left[\vec{H}_i^* - \left(\frac{\delta\vec{F}^+}{\delta x}\right)_{i-(1/2)}^n - \left(\frac{\delta\vec{F}^-}{\delta x}\right)_{i+(1/2)}^n\right] \tag{3.1}$$

where $\delta\vec{F}^-$ is associated to negative velocities and $\delta\vec{F}^+$ to positive velocities. A linear analysis of the homogeneous equations shows that the stability condition is Courant–Friedrich–Lewy number (CFL) ≤ 1 and it will be dissipative provided that CFL < 1, with

$$\mathrm{CFL} = \max|\alpha_k|\frac{\Delta t}{\delta x}$$

and α_k being the eigenvalues of the Jacobian.

When the source terms are dominant, it may be necessary to introduce a semi-implicit treatment for them in order to stabilize the scheme. One way to proceed is

$$\vec{H}_i^* = \theta\vec{H}_i^{n+1} + (1-\theta)\vec{H}_i^n \approx \vec{H}_i^n + \theta\left(\frac{\partial\vec{H}}{\partial\vec{u}}\frac{\partial\vec{u}}{\partial t}\right)_i^n\Delta t = \vec{H}_i^n + \theta\mathbf{K}_i^n\Delta\vec{u}_i^n$$

where $\mathbf{K} = \partial\vec{H}/\partial\vec{u}$ is the Jacobian of the source term. The conservative form (2.13) can then be expressed

$$(1-\theta\mathbf{K}_i^n\Delta t)\Delta\vec{u}_i^n = \Delta t\left[\vec{H}_i^n - \left(\mathbf{P}\boldsymbol{\Omega}^{\mathrm{L}}\mathbf{P}^{-1}\frac{\delta\vec{F}}{\delta x}\right)_{i-(1/2)}^n - \left(\mathbf{P}\boldsymbol{\Omega}^{\mathrm{R}}\mathbf{P}^{-1}\frac{\delta\vec{F}}{\delta x}\right)_{i+(1/2)}^n\right] \tag{3.2}$$

The following wave decomposition is assumed in this scheme in order to select the appropriate influence region in every case:

$$\boldsymbol{\Omega}^{\mathrm{L}} = \boldsymbol{\Omega}^+ = \frac{1}{2}[\mathbf{I} + \mathrm{sign}(\boldsymbol{\Lambda})], \qquad \boldsymbol{\Omega}^{\mathrm{R}} = \boldsymbol{\Omega}^- = \frac{1}{2}[\mathbf{I} - \mathrm{sign}(\boldsymbol{\Lambda})], \qquad \delta\vec{F}^{\pm} = \mathbf{P}\boldsymbol{\Omega}^{\pm}\mathbf{P}^{-1}\delta\vec{F} \tag{3.3}$$

An upwind treatment of the source terms is also possible using (2.13) instead of (2.14). Using a semi-implicit treatment of source terms and the wave decomposition (3.3)

Copyright © 2001 John Wiley & Sons, Ltd.

$$(1-\theta \mathbf{K}_i^n \Delta t)\Delta \vec{u}_i^n = \Delta t[(\mathbf{P\Omega}^+\mathbf{P}^{-1}\vec{G})_{i-(1/2)}^n + (\mathbf{P\Omega}^-\mathbf{P}^{-1}\vec{G})_{i+(1/2)}^n]$$

With the notation

$$\vec{G}^{\pm} = \mathbf{P\Omega}^{\pm}\mathbf{P}^{-1}\vec{G} \tag{3.4}$$

the scheme reduces to a simpler form

$$(1-\theta \mathbf{K}_i^n \Delta t)\Delta \vec{u}_i^n = \Delta t[(\vec{G}^+)_{i-(1/2)}^n + (\vec{G}^-)_{i+(1/2)}^n] \tag{3.5}$$

In the presence of transcritical flow, this numerical scheme does not produce good results due to the change of sign from negative to positive in the advection velocities. The numerical scheme is not able to interpret the transition as smooth and gives rise to non-physical shocks (entropy problem). The situation is associated to a local lack of numerical dissipation that can be corrected as outlined in Appendix A.

It is worth stressing here that this first-order upwind scheme produces second-order accuracy in space for steady cases [8]. Assuming for simplicity a positive propagation velocity in (3.5) at steady state

$$\left(\vec{H} - \frac{\delta \vec{F}}{\delta x}\right)_{i-(1/2)}^n = 0$$

which means that

$$\vec{F}_i^n = \vec{F}_{i-1}^n + \delta x \vec{H}_{i-(1/2)}^n$$

This is the mean point integration rule, an approximation of second-order. This gain in accuracy is the main advantage of the upwinding of the source terms. Actually, taking a pointwise approach

$$\vec{F}_i^n = \vec{F}_{i-1}^n + \delta x \vec{H}_i^n$$

which is the Euler integration rule, a first-order approximation.

The conservative character of this scheme is proved by the existence of a nodal numerical flux and a wave decomposition like

$$\vec{F}_i^T = \vec{F}_i^n, \qquad \delta \vec{F}_{i+(1/2)}^{\mathrm{L}} = (\delta \vec{F}^+)_{i+(1/2)}^n, \qquad \delta \vec{F}_{i+(1/2)}^{\mathrm{R}} = (\delta \vec{F}^-)_{i+(1/2)}^n$$

and an interface numerical flux

$$\vec{F}_{i+(1/2)}^* = \frac{1}{2}\{\vec{F}_i^n + \vec{F}_{i+1}^n - [\mathbf{P}\,\mathrm{sign}(\mathbf{\Lambda})\mathbf{P}^{-1}\delta\vec{F}]_{i+(1/2)}^n\}$$

Copyright © 2001 John Wiley & Sons, Ltd.

4. SPATIALLY SECOND-ORDER TVD SCHEME

4.1. Scalar case

As a preliminary step, the TVD conditions will be outlined for the scalar case. Assuming for simplicity a homogeneous scalar conservation law

$$\frac{\partial u(x, t)}{\partial t} + \frac{\partial F(u)}{\partial x} = 0 \;\Rightarrow\; \frac{\partial u(x, t)}{\partial t} + \alpha(u)\frac{\partial u(x, t)}{\partial x} = 0$$

with $\alpha = \partial F/\partial u$. The total variation (TV) of the discrete solution is defined as

$$\mathrm{TV}^n = \sum_i |u_{i+1}^n - u_i^n|$$

Then a scheme will be TVD if it satisfies the following property that prevents oscillatory solutions

$$\mathrm{TV}^{n+1} = \sum_i |u_{i+1}^{n+1} - u_i^{n+1}| \leq \sum_i |u_{i+1}^n - u_i^n| = \mathrm{TV}^n \tag{4.1}$$

A general explicit scheme that can be written in the form

$$\Delta u_i^n = -\lambda(\delta F_{i-(1/2)}^+ + \delta F_{i+(1/2)}^-)$$

with $\lambda = \Delta t/\delta x$, admits the definition of the following coefficients

$$\delta F_{i+(1/2)}^+ = C_{i+(1/2)}^+ \delta u_{i+(1/2)}^n, \qquad \delta F_{i+(1/2)}^- = C_{i+(1/2)}^- \delta u_{i+(1/2)}^n$$

The scheme will be TVD if (4.1) is used so that the coefficients (see [1] or [14] for more detail)

$$C_{i+(1/2)}^- \leq 0, \qquad C_{i+(1/2)}^+ \geq 0, \qquad \lambda(C_{i+(1/2)}^+ - C_{i+(1/2)}^-) \leq 1 \tag{4.2}$$

These conditions are automatically fulfilled by the first-order upwind scheme with the CFL stability condition.

The TVD condition for higher-order schemes involves scalar flux limiter functions [1]

$$\Delta \vec{u}_i^n = -\lambda\Big\{(\delta F^+)_{i-(1/2)}^n + (\delta F^-)_{i+(1/2)}^n + \frac{1}{2}[(\Psi^+\delta F^+)_{i-(1/2)}^n - (\Psi^+\delta F^+)_{i-(3/2)}^n + (\Psi^-\delta F^-)_{i+(1/2)}^n - (\Psi^-\delta F^-)_{i+(3/2)}^n]\Big\}$$

Copyright © 2001 John Wiley & Sons, Ltd.

The flux limiter functions are here defined combining second-order spatial centred and upwind schemes, for preserving the second-order in space, and according to (4.2) for avoiding numerical oscillations. In order to preserve the second-order of accuracy, the dependence of the flux limiter functions is defined as

$$\Psi^{+}_{i+(1/2)} = \Psi(r^{+}_{i+(1/2)}) = \Psi\left(\frac{\delta F^{+}_{i+(3/2)}}{\delta F^{+}_{i+(1/2)}}\right), \qquad \Psi^{-}_{i+(1/2)} = \Psi(r^{-}_{i+(1/2)}) = \Psi\left(\frac{\delta F^{-}_{i-(1/2)}}{\delta F^{-}_{i+(1/2)}}\right)$$

With this dependence, it can be easily seen that $\Psi(r)=1$, $\forall r$ produces the second-order in space upwind scheme (unstable), and that $\Psi(r)=r$, $\forall r$ reduces to the central Lax's scheme second-order accurate in space (unstable). Infinite spatially second-order schemes can be created for intermediate values of $\Psi(r)$. The zone of second-order accuracy in space is shaded in Figure 3.

Applying TVD conditions (4.2) the flux limiter will be a positive function so that [1]

$$\Psi(r)=0, \qquad \forall r<0; \qquad \Psi(r)\leq 2r, \qquad \forall r>0$$

and it produces the following stability condition:

$$\mathrm{CFL} \leq \frac{1}{1+\frac{1}{2}\max(\Psi)} \tag{4.3}$$

It is usual to establish the restriction $\Psi(r)\leq 2$ in order to be able to work up to $\mathrm{CFL}\leq\frac{1}{2}$. The intersection between the second-order region and the TVD region for the flux limiter functions is represented in Figure 4. Many particular flux limiter functions are defined in the literature. We use the most usual [1,15]

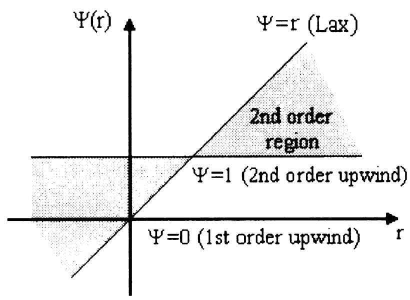

Figure 3. Second-order of approximation region for the flux limiter function.

Copyright © 2001 John Wiley & Sons, Ltd.

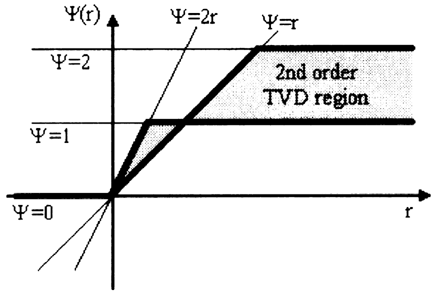

Figure 4. Second-order TVD region for the flux limiter functions.

Superbee: $\Psi(r) = \max[0, \min(1, 2r),\ \min(2, r)]$

Van-Lear: $\Psi(r) = \dfrac{r + |r|}{1 + |r|}$

Van-Albada: $\Psi(r) = \begin{cases} \dfrac{r + r^2}{1 + r^2} & \forall r > 0 \\ 0 & \forall r \leq 0 \end{cases}$

Minmod: $\Psi(r) = \max[0, \min(1, r)]$

4.2. Extension to systems with source terms

The second-order central Lax's scheme and the second-order upwind approximation to the space derivative, although unstable, are the basis for the construction of the second-order in space TVD schemes. By means of an adequate limitation of the spatial second-order terms, second-order accuracy can be preserved whilst oscillations are avoided. Once the conditions on the flux limiter functions have been established so that the desired properties are achieved on the solution to a scalar equation, a generalization to systems with source terms is desirable. With a semi-implicit and pointwise treatment of the source terms, the second-order in space TVD scheme is

$$(1 - \theta \mathbf{K}_i^n \Delta t)\Delta \vec{u}_i^n = \Delta t \Bigg\{ \vec{H}_i^n - \left(\frac{\delta \vec{F}^+}{\delta x}\right)_{i-(1/2)}^n - \left(\frac{\delta \vec{F}}{\delta x}\right)_{i+(1/2)}^n + \frac{1}{2}\Bigg[\left(\mathbf{\Psi}^+ \frac{\delta \vec{F}^+}{\delta x}\right)_{i-(3/2)}^n + \left(\mathbf{\Psi}^- \frac{\delta \vec{F}^-}{\delta x}\right)_{i+(3/2)}^n - \left(\mathbf{\Psi}^+ \frac{\delta \vec{F}^+}{\delta x}\right)_{i-(1/2)}^n - \left(\mathbf{\Psi}^- \frac{\delta \vec{F}^-}{\delta x}\right)_{i+(1/2)}^n \Bigg]\Bigg\} \quad (4.4)$$

where $\mathbf{\Psi}^+$ and $\mathbf{\Psi}^-$ are in general matrices. One of the simplest ways to define the flux limiting matrices $\mathbf{\Psi}^+$ and $\mathbf{\Psi}^-$ for a flux $\vec{F} = (F^1, \ldots, F^k)$ is

$$\mathbf{\Psi}^{\pm}_{i+(1/2)} = \begin{bmatrix} \Psi\left(\dfrac{(\delta F^1)^{\pm}_{i+(1/2)\pm 1}}{(\delta F^1)^{\pm}_{i+(1/2)}}\right) & & \\ & \ddots & \\ & & \Psi\left(\dfrac{(\delta F^k)^{\pm}_{i+(1/2)\pm 1}}{(\delta F^k)^{\pm}_{i+(1/2)}}\right) \end{bmatrix} \tag{4.5}$$

Another form is shown in reference [4].

This TVD second-order in space scheme is conservative since it admits the following wave decomposition:

$$\vec{F}^T_i = \vec{F}^n_i$$

$$\delta\vec{F}^{\mathrm{L}}_{i+(1/2)} = (\delta\vec{F}^+)^n_{i+(1/2)} - \frac{1}{2}(\mathbf{\Psi}^+\delta\vec{F}^+)^n_{i-(1/2)} + \frac{1}{2}(\mathbf{\Psi}^-\delta\vec{F}^-)^n_{i+(3/2)}$$

$$\delta\vec{F}^{\mathrm{R}}_{i+(1/2)} = (\delta\vec{F}^-)^n_{i+(1/2)} - \frac{1}{2}(\mathbf{\Psi}^+\delta\vec{F}^+)^n_{i-(1/2)} + \frac{1}{2}(\mathbf{\Psi}^-\delta\vec{F}^-)^n_{i+(3/2)}$$

and the following numerical flux:

$$\vec{F}^*_{i+(1/2)} = \frac{1}{2}\{\vec{F}^n_i + \vec{F}^n_{i+1} - [\mathbf{P}\,\mathrm{sign}(\mathbf{\Lambda})\mathbf{P}^{-1}\delta\vec{F}]^n_{i+(1/2)} + (\mathbf{\Psi}^+\delta\vec{F}^+)^n_{i-(1/2)} - (\mathbf{\Psi}^-\delta\vec{F}^-)^n_{i+(3/2)}\}$$

With an interest focussed on the correct treatment of problems involving source terms we will work with the already defined generalized flux $\vec{G}_{i+(1/2)}$. With this notation, the conservative TVD second-order in space scheme with upwind treatment of source terms can be written

$$(1 - \theta\mathbf{K}^n_i\Delta t)\Delta\vec{u}^n_i = \Delta t\Big\{(\vec{G}^+)^n_{i-(1/2)} + (\vec{G}^-)^n_{i+(1/2)} + \frac{1}{2}[(\mathbf{\Psi}^+\vec{G}^+)^n_{i-(3/2)} + (\mathbf{\Psi}^-\vec{G}^-)^n_{i+(1/2)} - (\mathbf{\Psi}^+\vec{G}^+)^n_{i-(1/2)} - (\mathbf{\Psi}^-\vec{G}^-)^n_{i+(3/2)}]\Big\} \tag{4.6}$$

The unstable Lax's scheme with non-pointwise source term is

$$(1 - \theta\mathbf{K}^n_i\Delta t)\Delta\vec{u}^n_i = \frac{\Delta t}{2}[(\vec{G}^+)^n_{i-(1/2)} + (\vec{G}^-)^n_{i-(1/2)} + (\vec{G}^+)^n_{i+(1/2)} + (\vec{G}^-)^n_{i+(1/2)}]$$

In order to ensure this limit when $\Psi(r) = r$, hence preserving the second-order of accuracy, the limiting diagonal matrices $\mathbf{\Psi}^+$ and $\mathbf{\Psi}^-$, for $\vec{G} = (G^1 \ldots G^k)$, must be

Copyright © 2001 John Wiley & Sons, Ltd.

$$\mathbf{\Psi}^{\pm}_{i+(1/2)} = \begin{bmatrix} \Psi\left(\dfrac{(G^1)^{\pm}_{i+(1/2)\pm 1}}{(G^1)^{\pm}_{i+(1/2)}}\right) & & \\ & \ddots & \\ & & \Psi\left(\dfrac{(G^k)^{\pm}_{i+(1/2)\pm 1}}{(G^k)^{\pm}_{i+(1/2)}}\right) \end{bmatrix} \tag{4.7}$$

It is worth stressing that a rigorous upwind treatment of the source terms in the TVD scheme imposes not only the limitation of the source terms but also their involvement in the definition of the limitation function itself. However, it is possible to neglect the contributions of the source terms in the limiter functions using the limiter matrix (4.5) instead of (4.7). Proceeding this way, although the limitation of the flux is ensured and the scheme is free from numerical oscillations, second-order is not ensured and a loss of accuracy is produced. On the other hand, including the source terms in the definition of the limiters proposed in (4.7) is not hard work since the vector components G^k must be equally calculated.

5. SECOND-ORDER IN SPACE AND TIME TVD SCHEME

To develop the schemes to second-order in time the approximation of the time derivatives to second-order has to be included, which, with reference to system (1.1)

$$\Delta\vec{u}^n_i = \left(\frac{\partial\vec{u}}{\partial t}\right)^n_i \Delta t + \frac{1}{2}\left(\frac{\partial^2\vec{u}}{\partial t^2}\right)^n_i \Delta t^2 + O(\Delta t^3) = \left(\vec{H} - \frac{\partial\vec{F}}{\partial x}\right)^n_i \Delta t + \frac{1}{2}\frac{\partial}{\partial t}\left(\vec{H} - \frac{\partial\vec{F}}{\partial x}\right)^n_i \Delta t^2 + O(\Delta t^3)$$

For the time derivative the following has to be taken into account:

$$\frac{\partial}{\partial t}\left(\vec{H} - \frac{\partial\vec{F}}{\partial x}\right) = \frac{\partial\vec{H}}{\partial\vec{u}}\frac{\partial\vec{u}}{\partial t} - \frac{\partial}{\partial x}\left(\frac{\partial\vec{F}}{\partial\vec{u}}\frac{\partial\vec{u}}{\partial t}\right) = \mathbf{K}\frac{\partial\vec{u}}{\partial t} - \frac{\partial}{\partial x}\left[\mathbf{J}\left(\vec{H} - \frac{\partial\vec{F}}{\partial x}\right)\right]$$

where the Jacobians of the flux and source term have been used. Note that in this approximation the $\mathbf{K}(\partial\vec{u}/\partial t)$ and $\partial(\mathbf{J}\vec{H})/\partial x$ terms appear due to the presence of the source term in the equation. These terms are obviated in previous works but are relevant to the quality of the results. With a central approximation of the spatial derivatives

$$\frac{\partial}{\partial t}\left(\vec{H} - \frac{\partial\vec{F}}{\partial x}\right)^n_i = \mathbf{K}^n_i \frac{\Delta\vec{u}^n_i}{\Delta t} - \frac{1}{\delta x}\left[(\mathbf{J}\vec{G})^n_{i+(1/2)} - (\mathbf{J}\vec{G})^n_{i-(1/2)}\right] + O(\delta x, \Delta t)$$

and a central and semi-implicit discretization of the source terms, the Lax–Wendroff scheme is obtained

$$\left(1 - \mathbf{K}^n_i \frac{\Delta t}{2}\right)\Delta\vec{u}^n_i = \Delta t\left(\vec{H}^n_i - \frac{\vec{F}^n_{i+1} - \vec{F}^n_i}{2\delta x}\right) - \frac{\Delta t^2}{2\delta x}\left[(\mathbf{J}\vec{G})^n_{i+(1/2)} - (\mathbf{J}\vec{G})^n_{i-(1/2)}\right] \tag{5.1}$$

Copyright © 2001 John Wiley & Sons, Ltd.

The second-order in space and time scheme is stable provided that CFL ≤ 1 and dissipative if CFL < 1. Written as a sum of the first-order upwind plus second-order correction terms

$$\left(1-\mathbf{K}_i^n\frac{\Delta t}{2}\right)\Delta\vec{u}_i^n=\left[\vec{H}_i^n-\left(\frac{\partial\vec{F}^+}{\delta x}\right)_{i-(1/2)}^n-\left(\frac{\delta\vec{F}^-}{\delta x}\right)_{i+(1/2)}^n\right]$$
$$+\frac{\Delta t}{2\delta x}[(\delta\vec{F}^- -\Delta t\mathbf{J}^-\vec{G}^-)_{i+(1/2)}^n-(\delta\vec{F}^- -\Delta t\mathbf{J}^-\vec{G}^-)_{i-(1/2)}^n$$
$$-(\delta\vec{F}^+ +\Delta t\mathbf{J}^+\vec{G}^+)_{i+(1/2)}^n+(\delta\vec{F}^+ +\Delta t\mathbf{J}^+\vec{G}^+)_{i-(1/2)}^n]$$

Similarly, an upwind second-order in space and time approximation can be derived. The starting point is the same but the space derivatives are approximated in a non-centred form

$$\left(1-\mathbf{K}_i^n\frac{\Delta t}{2}\right)\Delta\vec{u}_i^n=\Delta t\left\{\vec{H}_i^n-\frac{3}{2}\left[\left(\frac{\delta\vec{F}^+}{\delta x}\right)_{i-(1/2)}^n+\left(\frac{\delta\vec{F}^-}{\delta x}\right)_{i+(1/2)}^n\right]\right.$$
$$+\frac{1}{2}\left[\left(\frac{\delta\vec{F}^+}{\delta x}\right)_{i-(3/2)}^n+\left(\frac{\delta\vec{F}^-}{\delta x}\right)_{i+(3/2)}^n\right]-\frac{\Delta t^2}{2\delta x}[(\mathbf{J}^-\vec{G}^-)_{i+(3/2)}^n$$
$$\left.-(\mathbf{J}^-\vec{G}^-)_{i+(1/2)}^n+(\mathbf{J}^+\vec{G}^+)_{i-(1/2)}^n-(\mathbf{J}^+\vec{G}^+)_{i-(3/2)}^n]\right\} \tag{5.2}$$

The stability condition for this scheme is CFL ≤ 2 and it is dissipative if CFL < 2.

A TVD scheme second-order accurate in space and time can be built from (5.2) as a sum of the first-order upwind plus limited second-order correction terms, as

$$\left(1-\mathbf{K}_i^n\frac{\Delta t}{2}\right)\Delta\vec{u}_i^n=\Delta t\left[\vec{H}_i^n-\left(\frac{\delta\vec{F}^+}{\delta x}\right)_{i-(1/2)}^n-\left(\frac{\delta\vec{F}^-}{\delta x}\right)_{i+(1/2)}^n\right]$$
$$+\frac{\Delta t}{2\delta x}\{[\mathbf{\Psi}^-(\delta\vec{F}-\Delta t\mathbf{J}\vec{G})^-]_{i+(3/2)}^n-[\mathbf{\Psi}^-(\delta\vec{F}-\Delta t\mathbf{J}\vec{G})^-]_{i+(1/2)}^n$$
$$-[\mathbf{\Psi}^+(\delta\vec{F}+\Delta t\mathbf{J}\vec{G})^+]_{i-(1/2)}^n+[\mathbf{\Psi}^+(\delta\vec{F}+\Delta t\mathbf{J}\vec{G})^+]_{i-(3/2)}^n\} \tag{5.3}$$

The flux limiter functions $\Psi(r)$ have to preserve the second-order, therefore

$$\Psi(r)=1,\ \forall r\ \Rightarrow\ \text{second-order in space and time upwind scheme}$$
$$\Psi(r)=r,\ \forall r\ \Rightarrow\ \text{Lax–Wendroff} \tag{5.4}$$

Infinite schemes can be created again that would be intermediate between the fully upwind or fully central second-order approximations. For the second-order in space and time scheme the TVD second-order region is the same as that of Figure 4, so that all the limiting functions defined in Section 4 are also suitable for this scheme. The limiting functions $\mathbf{\Psi}^+$ and $\mathbf{\Psi}^-$ will be defined in terms of

Copyright © 2001 John Wiley & Sons, Ltd.

$$\vec{R}^{\pm} = (\delta\vec{F} \pm \Delta t \mathbf{J}\vec{G})^{\pm}$$

with $\vec{R} = (R^1, \ldots, R^k)$, matrices of the form

$$\mathbf{\Psi}^{\pm}_{i+(1/2)} = \begin{bmatrix} \Psi\left(\dfrac{(R^1)^{\pm}_{i+(1/2)\pm 1}}{(R^1)^{\pm}_{i+(1/2)}}\right) & & \\ & \ddots & \\ & & \Psi\left(\dfrac{(R^k)^{\pm}_{i+(1/2)\pm 1}}{(R^k)^{\pm}_{i+(1/2)}}\right) \end{bmatrix}$$

The desired second-order properties are then achieved and, although the TVD condition is only strictly ensured for the characteristic variables, it produces minimum oscillations in practice when dealing with conservative variables. The stability condition is $\mathrm{CFL} \leq 1$ loosing, by limiting the second-order upwind scheme, the property of remaining stable up to $\mathrm{CFL} = 2$.

Scheme (5.3) is defined conservative since it admits a wave decomposition like

$$\vec{F}^{T}_{i} = \vec{F}^{n}_{i}$$

$$\delta\vec{F}^{\mathrm{L}}_{i+(1/2)} = (\delta\vec{F}^{+})^{n}_{i+(1/2)} - \frac{1}{2}(\mathbf{\Psi}^{+}\vec{R}^{+})^{n}_{i-(1/2)} + \frac{1}{2}(\mathbf{\Psi}^{-}\vec{R}^{-})^{n}_{i+(3/2)}$$

$$\delta\vec{F}^{\mathrm{R}}_{i+(1/2)} = (\delta\vec{F}^{-})^{n}_{i+(1/2)} + \frac{1}{2}(\mathbf{\Psi}^{+}\vec{R}^{+})^{n}_{i-(1/2)} - \frac{1}{2}(\mathbf{\Psi}^{-}\vec{R}^{-})^{n}_{i+(3/2)}$$

and an interface numerical flux

$$\vec{F}^{*}_{i+(1/2)} = \frac{1}{2}\{\vec{F}^{n}_{i} + \vec{F}^{n}_{i+1} - [\mathbf{P}\,\mathrm{sign}(\mathbf{\Lambda})\mathbf{P}^{-1}\delta\vec{F}]^{n}_{i+(1/2)} + (\mathbf{\Psi}^{+}\vec{R}^{+})^{n}_{i-(1/2)} - (\mathbf{\Psi}^{-}\vec{R}^{-})^{n}_{i+(3/2)}\}$$

Note that vector $\vec{R}$ includes source terms, therefore a coupling of flux and source terms appears as a consequence of the second-order in time.

In order to complete the construction of the scheme, an upwind treatment of the source terms will be applied now. For that purpose, and using again the notation based on variable $\vec{G}_{i+(1/2)}$, the upwind TVD second-order in space and time scheme can be expressed as a sum of the first-order upwind plus limited corrections to second-order as

$$\begin{aligned}\left(1 - \mathbf{K}^{n}_{i}\frac{\Delta t}{2}\right)\Delta\vec{u}^{n}_{i} = {} & \Delta t[(\vec{G}^{+})^{n}_{i-(1/2)} + (\vec{G}^{-})^{n}_{i+(1/2)}] \\ & + \frac{\Delta t}{2}\left\{\left[\mathbf{\Psi}^{+}\left(1 - \frac{\Delta t}{\delta x}\mathbf{J}^{+}\right)\vec{G}^{+}\right]^{n}_{i-(1/2)} - \left[\mathbf{\Psi}^{+}\left(1 - \frac{\Delta t}{\delta x}\mathbf{J}^{+}\right)\vec{G}^{+}\right]^{n}_{i-(3/2)}\right. \\ & \left. + \left[\mathbf{\Psi}^{-}\left(1 + \frac{\Delta t}{\delta x}\mathbf{J}^{-}\right)\vec{G}^{-}\right]^{n}_{i+(1/2)} - \left[\mathbf{\Psi}^{-}\left(1 + \frac{\Delta t}{\delta x}\mathbf{J}^{-}\right)\vec{G}^{-}\right]^{n}_{i+(3/2)}\right\}\end{aligned} \tag{5.5}$$

Copyright © 2001 John Wiley & Sons, Ltd.

And, Lax–Wendroff scheme, on the other hand, with upwind source term can be expressed

$$\left(1-\mathbf{K}_i^n\frac{\Delta t}{2}\right)\Delta\vec{u}_i^n=\Delta t[(\vec{G}^+)_{i-(1/2)}^n+(\vec{G}^-)_{i+(1/2)}^n]$$
$$+\frac{\Delta t}{2}\left\{\left[\left(1-\frac{\Delta t}{\delta x}\mathbf{J}^+\right)\vec{G}^+\right]_{i+(1/2)}^n-\left[\left(1-\frac{\Delta t}{\delta x}\mathbf{J}^+\right)\vec{G}^+\right]_{i-(1/2)}^n\right.$$
$$\left.+\left[\left(1+\frac{\Delta t}{\delta x}\mathbf{J}^-\right)\vec{G}^-\right]_{i-(1/2)}^n-\left[\left(1+\frac{\Delta t}{\delta x}\mathbf{J}^-\right)\vec{G}^-\right]_{i+(1/2)}^n\right\} \quad (5.6)$$

Defining now

$$\vec{S}^+=\left(1-\frac{\Delta t}{\delta x}\mathbf{J}^+\right)\vec{G}^+,\qquad \vec{S}^-=\left(1+\frac{\Delta t}{\delta x}\mathbf{J}^-\right)\vec{G}^-$$

In order to recover with $\Psi(r)=r$ the Lax–Wendroff scheme (5.6), hence preserving the second-order to approximation, the diagonal limiting matrices have to be

$$\mathbf{\Psi}_{i+(1/2)}^{\pm}=\begin{bmatrix}\Psi\left(\dfrac{(S^1)_{i+(1/2)\pm 1}^{\pm}}{(S^1)_{i+(1/2)}^{\pm}}\right) & & \\ & \ddots & \\ & & \Psi\left(\dfrac{(S^k)_{i+(1/2)\pm 1}^{\pm}}{(S^k)_{i+(1/2)}^{\pm}}\right)\end{bmatrix}$$

6. NUMERICAL RESULTS

A set of tests has been selected to illustrate the performance of some of the techniques described in the paper. In all the examples, the schemes have been applied to the system of shallow water equations. In all of them, the non-conservative formulation of the equations has been used with the discretization given by (2.8). When source terms are present, the semi-implicit upwind treatment has been implemented with $\theta=0.5$. The numerical treatment at the boundaries follows Appendix B and the entropy correlation used in all the schemes is described in Appendix A. The CFL number used is always 90% of the maximum for stability. In spatial second-order TVD schemes this is 0.45 with the Superbee flux limiter and 0.6 with the Minmod flux limiter, this being 0.9 in the other schemes.

6.1. Dam-break flow

This classical test case is considered a benchmark for comparison of the performance of numerical schemes specially designed for discontinuous transient flow. Although defined by the system of homogeneous shallow water equations, it is widely considered a standard test case for validation of schemes. Starting from initial conditions given by still water and two different water levels separated by a dam, the theory of characteristics supplies an exact

Copyright © 2001 John Wiley & Sons, Ltd.

evolution solution [12] that can be used as a reference. In the example presented, two ratios of initial water depths $h_L/h_R = 10$ and $h_L/h_R = 100$ are used. The solution is displayed in Figures 5–10 for $t = 20$ s. A space interval of $\Delta x = 1$ m is used in the mesh. The entropy correction produces remarkable results, the typical '*dog-leg*' effect being negligible. It is also remarkable that the Lax–Wendroff scheme only with entropy correction, although displaying numerical oscillations, is able to solve strong shocks without a TVD correction.

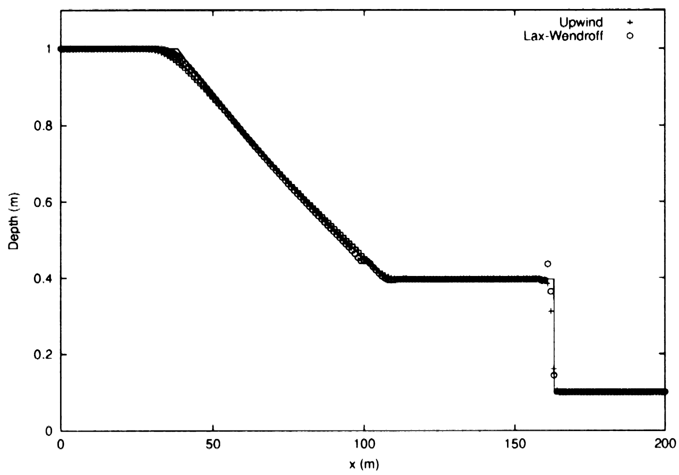

Figure 5. Ideal dam-break (mild shock) with first-order upwind and Lax–Wendroff schemes.

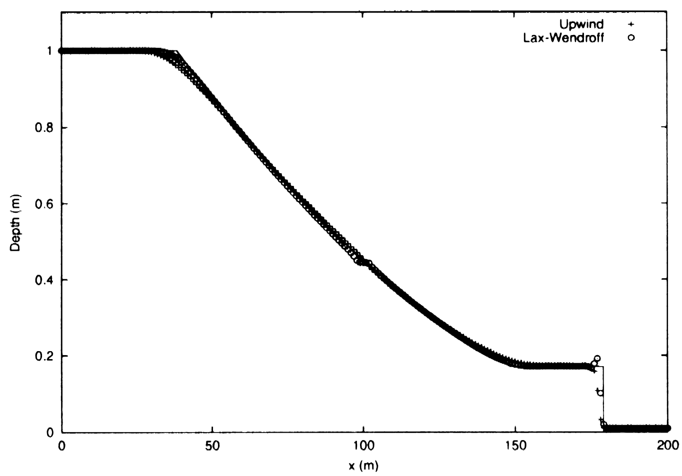

Figure 6. Ideal dam-break (strong shock) with first-order upwind and Lax–Wendroff schemes.

Copyright © 2001 John Wiley & Sons, Ltd.

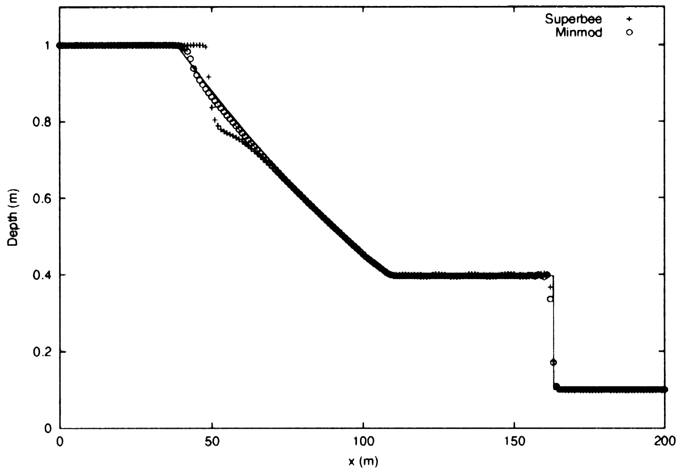

Figure 7. Ideal dam-break (mild shock) with second-order in space TVD scheme.

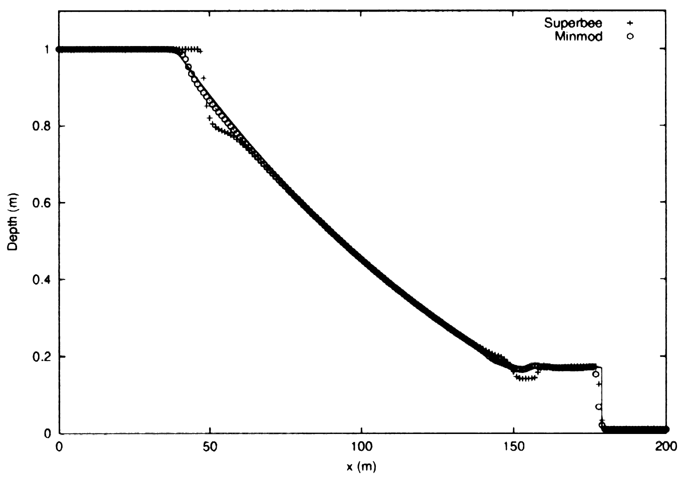

Figure 8. Ideal dam-break (strong shock) with second-order in space TVD scheme.

The first-order upwind scheme provides a reasonably good result with a slight numerical diffusion. The second-order in space TVD scheme tends to produce antidiffusive solutions, this being excessive with the Superbee flux limiter. Nevertheless with the Minmod flux limiter this is less noticeable providing a slight improvement with regard to the first-order scheme. Second-order in space and time improves the numerical solution being the most accurate scheme as expected.

Copyright © 2001 John Wiley & Sons, Ltd.

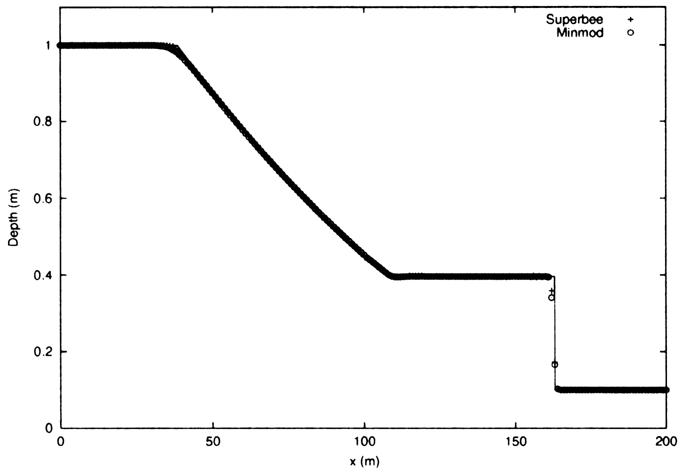

Figure 9. Ideal dam-break (mild shock) with second-order in space and time TVD scheme.

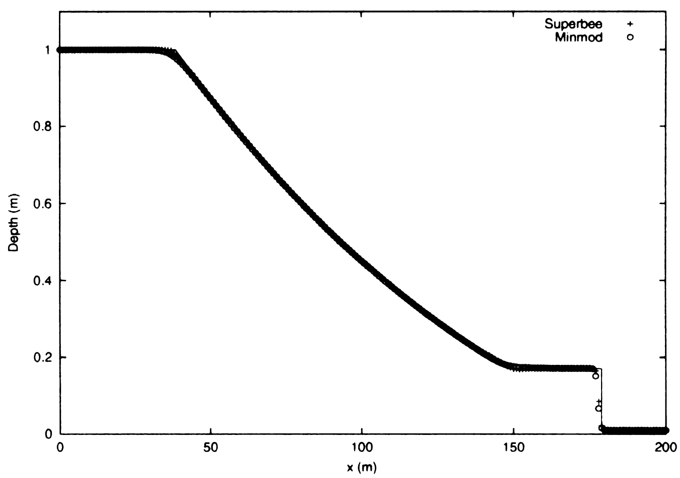

Figure 10. Ideal dam-break (strong shock) with second-order in space and time TVD scheme.

6.2. Steady flow in channels

When the shallow water equations are used to model hydraulic problems involving bed slope changes and bed friction, the system is no longer homogeneous and the source terms have to be taken into account. On the other hand, this renders it more difficult and often impossible to find reference exact solutions. McDonald [13] proposed a set of test cases based on steady

Copyright © 2001 John Wiley & Sons, Ltd.

flow in channels of varying bed slope and/or breadth by calculating the analytical slope and breadth functions compatible with constant discharge conditions given an analytical water depth function. Among them, we have chosen an example consisting of a 650 m long trapezoidal channel with a bed variation given by a slope function of x and a roughness coefficient $n = 0.03$. The constant discharge $Q = 20$ m^3 s^{-1} is imposed upstream and supercritical flow is enforced downstream. There are some points of transcritical flow. The same CFL number of the dam-break problem is used in this case. The space interval of the mesh is $\Delta x = 6.5$ m. The most important detail to note on the results is the perfect coincidence of all the upwind based schemes with upwind treatment of the source terms and the superiority of limitation proposed in the TVD schemes versus the results given by the Lax–Wendroff scheme (Figures 11–14) This is an expected behaviour since it is a steady flow problem in which the upwind treatment of the source terms produces second-order spatial accuracy even in the first-order upwind scheme as mentioned before. In Figure 16 the '*traditional*' form of the TVD scheme (pointwise and no limited treatment of the source terms) performance is shown making evident the superiority of an upwind limited treatment of the source terms. In Figure 15 two different forms of limiting the source terms are compared. A slight loss of accuracy excluding the source terms of the flux limiter is patent in the discontinuities because it is the place where the flux limiter plays a determining role.

6.3. Unsteady flow in rivers

In order to show the application to a practical case, an example of unsteady flow in a river is presented now. It is a 9000 m long reach of the upstream part of the River Neila in Spain. Being a mountain river, it is characterized by strong irregularities in the cross section, by a rather steep part in the first kilometres and by a low base discharge (1 m^3 s^{-1}) which, altogether, produce a high velocity basic flow, transcritical in some parts. In order to check the conservation properties of the schemes applied, and the absence of oscillations in the TVD schemes, a sudden increase in discharge up to 40 m^3 s^{-1} and a critical depth is imposed at the upstream end. This step hydrograph propagates into the river. The same CFL number as the steady flow cases and an interval of $\Delta x = 22.5$ m in the mesh is used. Figures 17–20 show that the discharge wave propagates with almost a perfectly constant value at times $t = 500$, 1000 and 1500 s. In Figure 21 the oscillations produced by non-limiting the source terms are displayed. Figure 22 shows the detail of the front wave where the advantages of using higher-order approaches are noticeable, this is not so clear when reproducing steady states. In Figure 23 the strong gradient in the bed slope of River Neila can be seen. Figures 24–28 show some other variables as calculated with the second-order in space and time TVD scheme with the Superbee limiter (the most used scheme) and the strong irregularities of the river are evident.

7. CONCLUSIONS

In this work a study of different one-step explicit schemes is presented. The one-step schemes, although slightly more complex than two-step methods (e.g. McCormack [1]) of second-order

Copyright © 2001 John Wiley & Sons, Ltd.

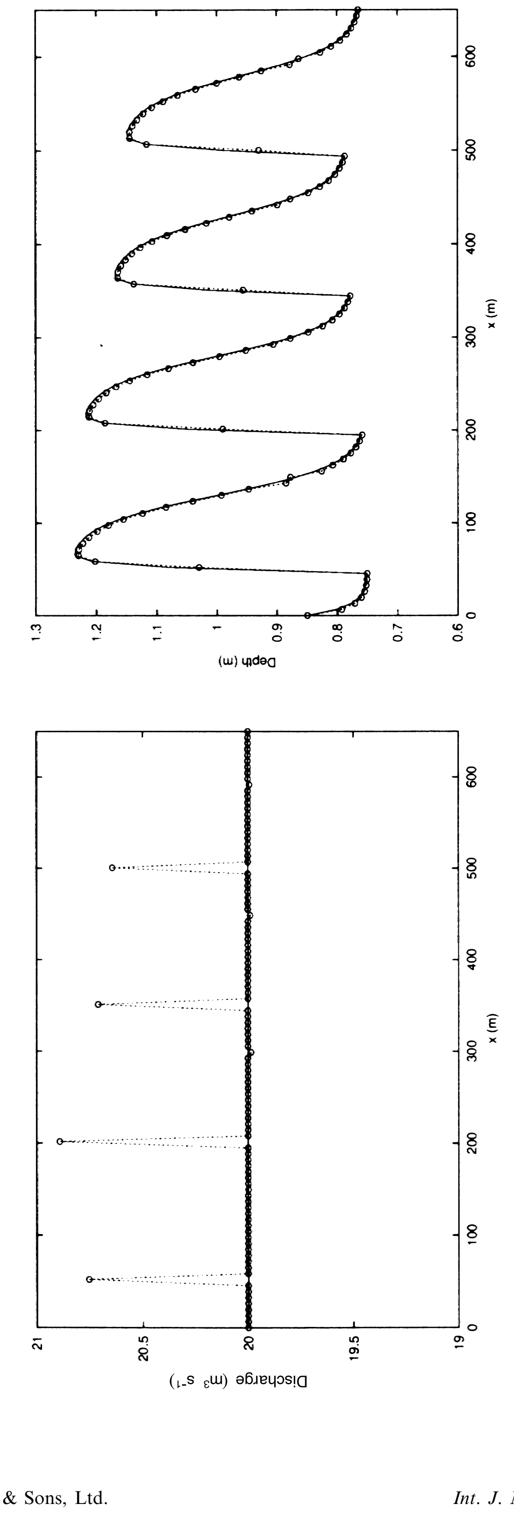

Figure 11. Discharge and depth with first-order upwind scheme.

Copyright © 2001 John Wiley & Sons, Ltd.

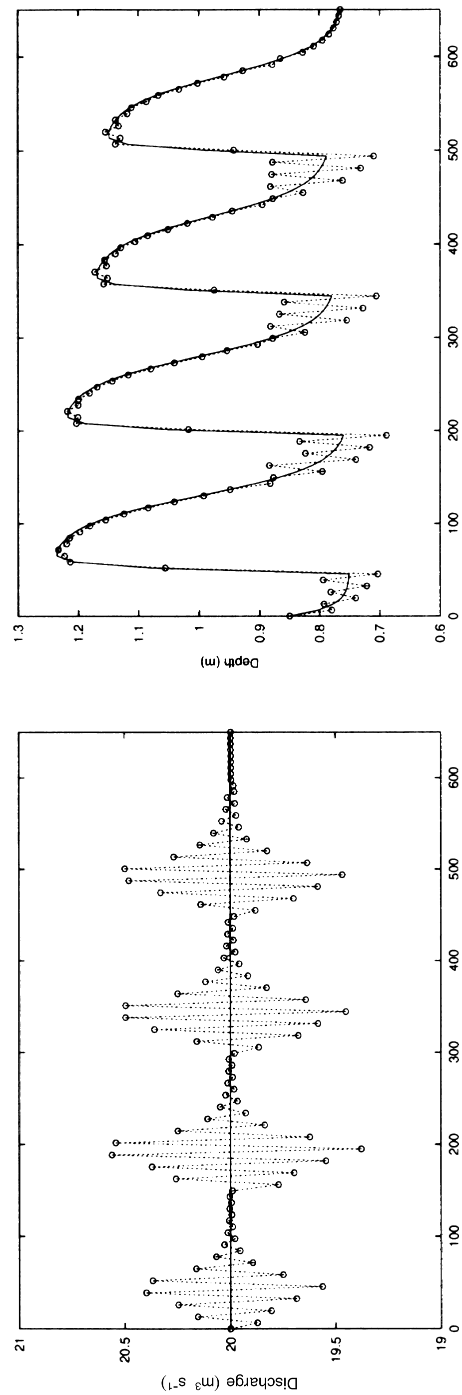

Figure 12. Discharge and depth with Lax–Wendroff scheme with entropy correction and non-pointwise treatment of source terms.

Copyright © 2001 John Wiley & Sons, Ltd.

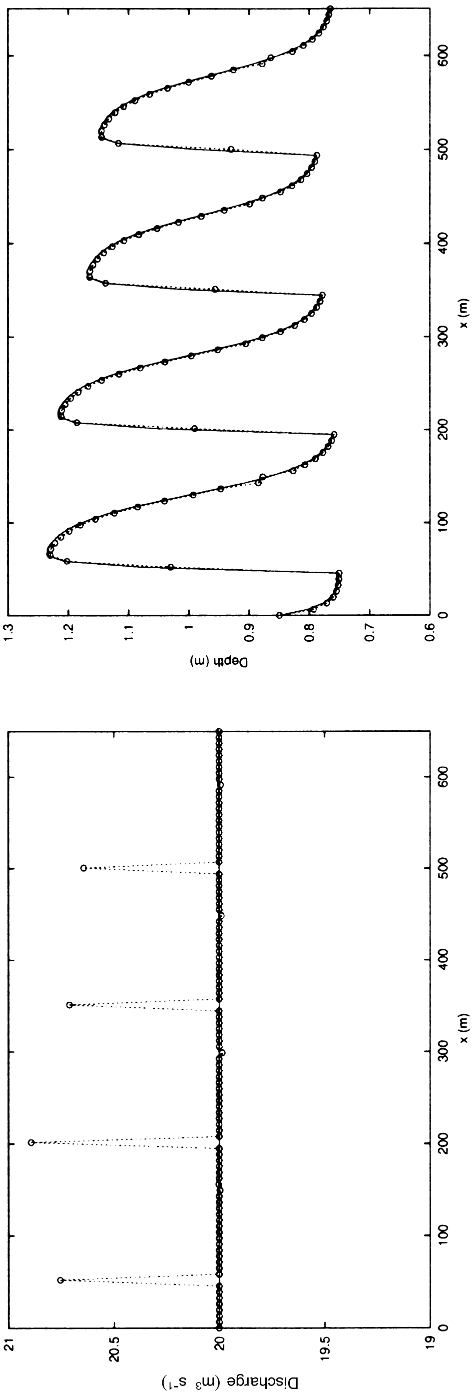

Figure 13. Discharge and depth with second-order in space TVD scheme with upwind treatment and the limitation proposed of source terms.

Copyright © 2001 John Wiley & Sons, Ltd.

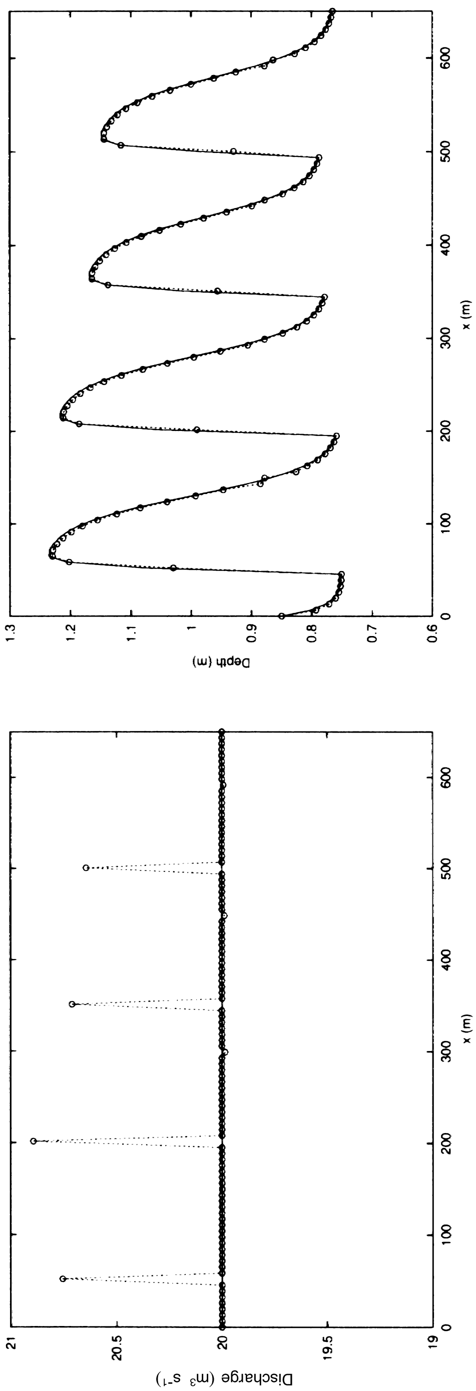

Figure 14. Discharge and depth with second-order in space and time TVD scheme with upwind treatment and the limitation proposed of source terms.

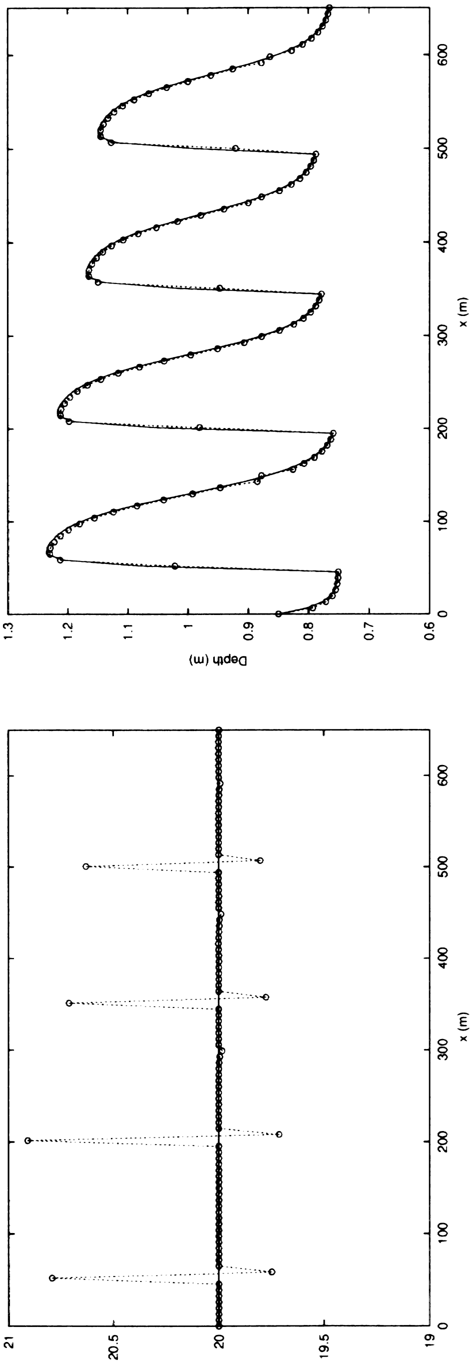

Figure 15. Discharge and depth with second-order in space TVD scheme with upwind and limited source terms without introducing they in the flux limiter (using the limiter matrix (4.7)).

Copyright © 2001 John Wiley & Sons, Ltd.

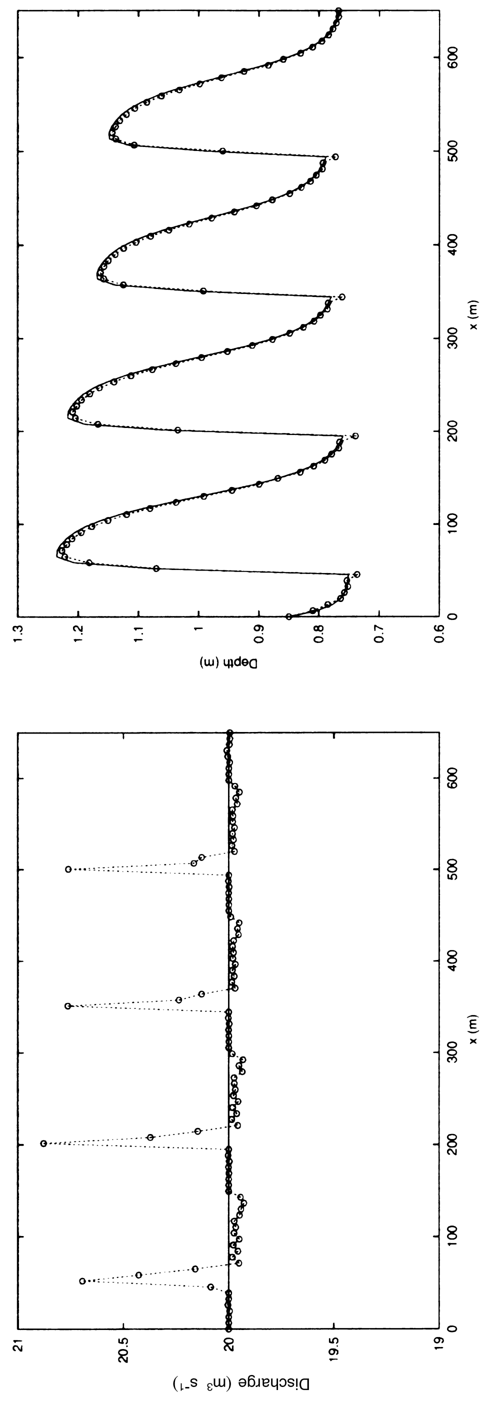

Figure 16. Discharge and depth with the second-order in space TVD scheme with a pointwise treatment of source terms.

Copyright © 2001 John Wiley & Sons, Ltd.

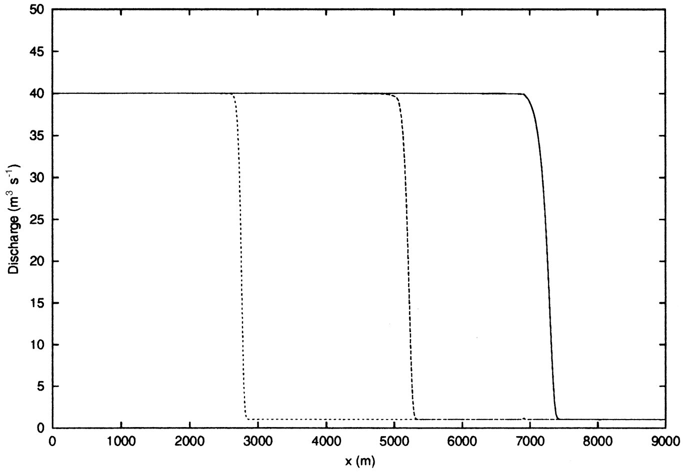

Figure 17. Evolution in the discharge with first-order upwind scheme.

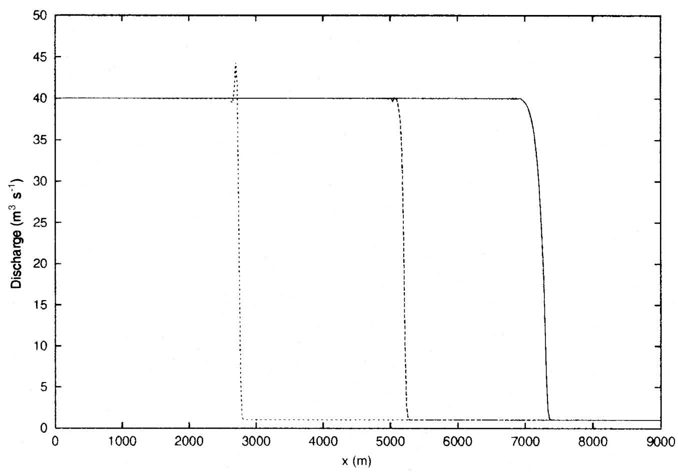

Figure 18. Evolution in the discharge with Lax–Wendroff scheme.

of approximation, are faster for resolving shallow water equations in rivers and irregular channels because the time elapsed calculating the sectional parameters of the channel is much greater than the time elapsed resolving the numerical schemes. Predictor–corrector schemes calculate twice the parameters every time step requiring double CPU time versus one-step schemes. Moreover, one-step methods admit a semi-implicit treatment of source terms, necessary for stabilizing the simulations when these become dominant (rivers and irregular

Copyright © 2001 John Wiley & Sons, Ltd.

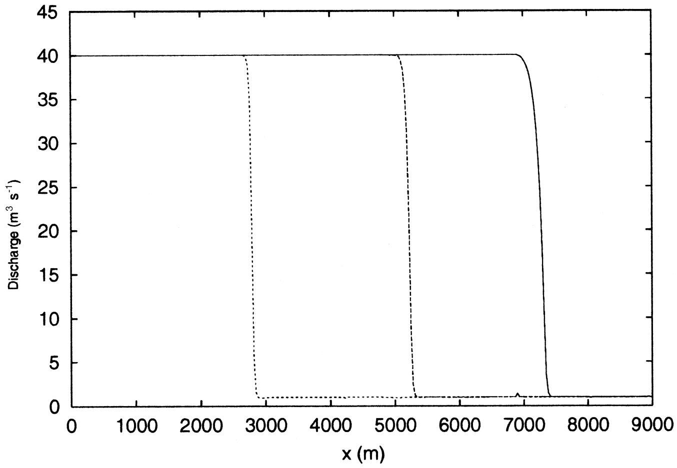

Figure 19. Evolution in the discharge with second-order in space TVD scheme with Minmod limiter.

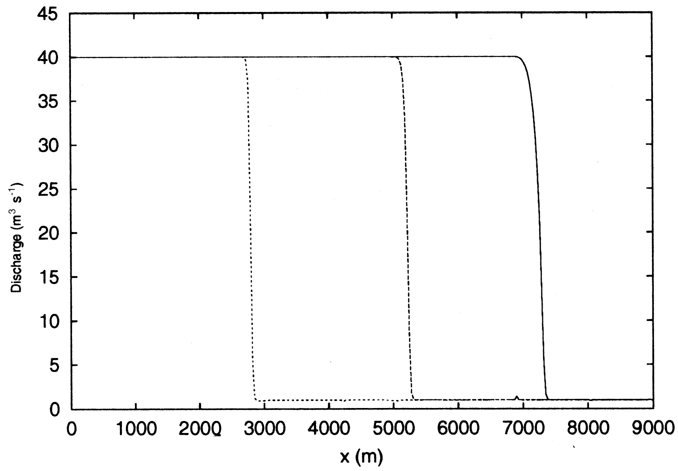

Figure 20. Evolution in the discharge with second-order in space and time TVD scheme with Superbee limiter.

channels), whereas two-step schemes loose the second-order of approximation in time property when a semi-implicit treatment of source terms is applied.

The applications presented are based on conservative schemes in the non-conservative formulation of the equations. This form takes advantage of the simplicity of the non-conservative form of the equations and produces faster and simpler schemes without loosing the accuracy of conservative schemes.

Copyright © 2001 John Wiley & Sons, Ltd.

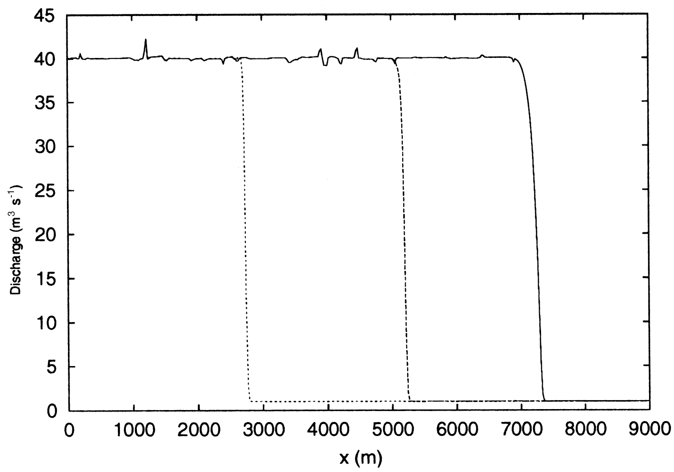

Figure 21. Evolution in the discharge with second-order in space and time TVD scheme with Superbee limiter with pointwise treatment and without limiting the source terms.

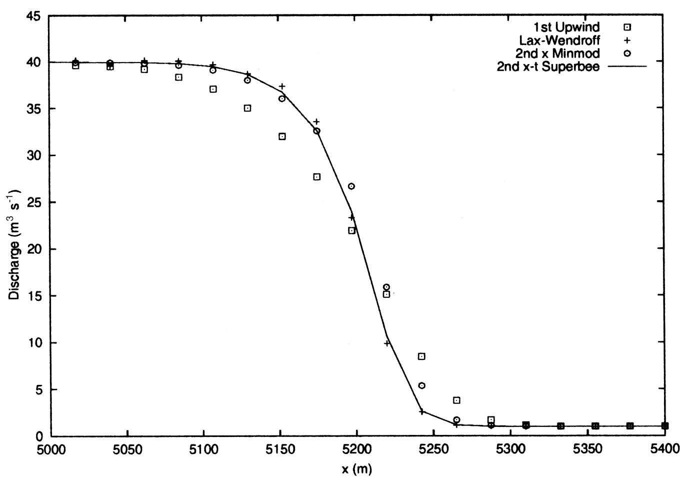

Figure 22. Zoom view of the shock at $t = 1000$ s with all the considered schemes.

The first-order upwind scheme with upwind and semi-implicit treatment of the source terms, at this moment, is one of the best schemes simulating shallow water equations because, although this is a first-order scheme, it is robust, reasonably simple, fast and it produces second-order solutions in steady and quasi-steady problems. Therefore, it is the preferential explicit scheme for steady flow. On the other hand, it is very accurate solving strong shocks

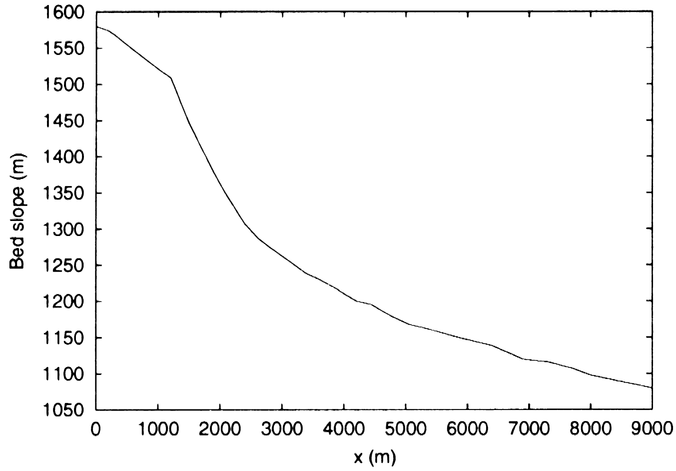

Figure 23. Bed slope of River Neila. The strong gradient can be seen.

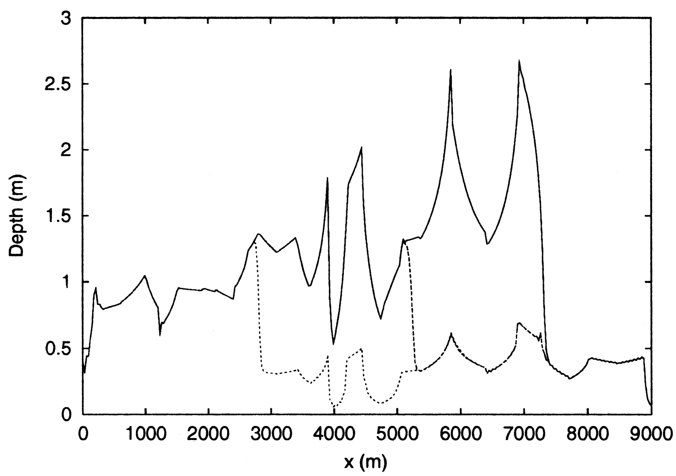

Figure 24. Evolution in the depth with second-order in space and time TVD scheme with Superbee limiter.

which many other schemes cannot simulate. The conservation errors are very small and there is an absence of oscillations. Nevertheless, the results are less accurate in unsteady flows.

The TVD second-order in space scheme is equally robust, more accurate, but slightly more complex than the first-order upwind scheme and is the simplest high-order TVD scheme. It is recommended to use the Minmod flux limiter with this scheme since best results are achieved

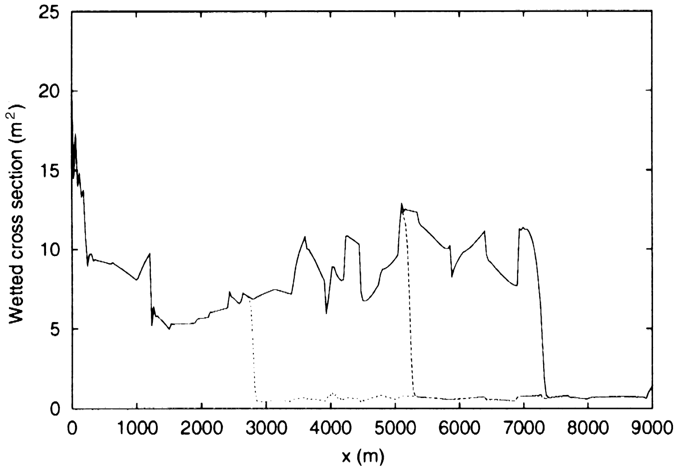

Figure 25. Evolution in the wetted cross-section with second-order in space and time TVD scheme with Superbee limiter.

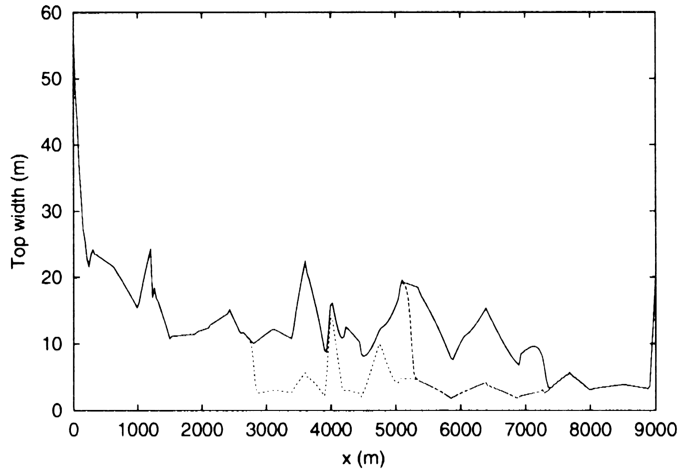

Figure 26. Evolution in the top width with second-order in space and time TVD scheme with Superbee limiter.

and is less restrictive in the CFL condition (CFL $< \frac{2}{3}$ versus CFL $< \frac{1}{2}$). However, this limitation in the CFL number is compensated by an increment of 50% in the CPU time compared with all other methods used in this work.

Copyright © 2001 John Wiley & Sons, Ltd.

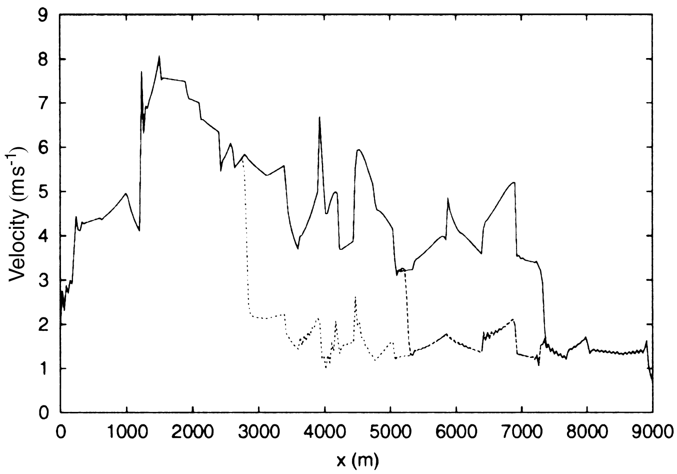

Figure 27. Evolution in the fluid velocity with second-order in space and time TVD scheme with Superbee limiter.

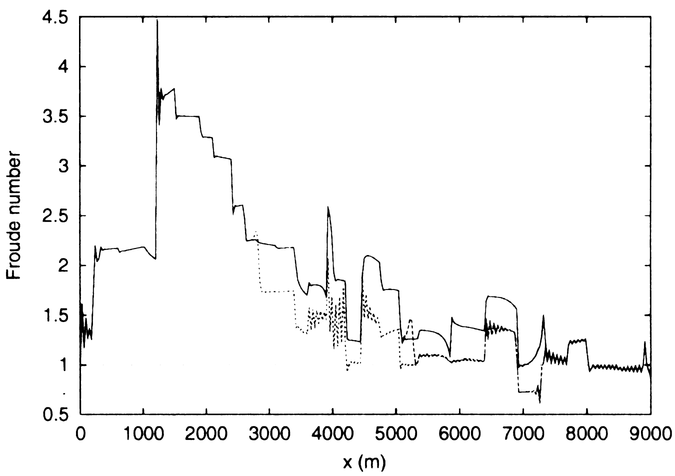

Figure 28. Evolution in the Froude number with second-order in space and time TVD scheme with Superbee limiter.

Though a little more complex, the TVD second-order in space and time scheme is faster (the CPU time is similar to the first-order schemes) than the TVD second-order in space scheme. With this scheme the Superbee and Van-Leer flux limiters produce more accurate solutions representing a small improvement with regard to the TVD second-order in space scheme.

Copyright © 2001 John Wiley & Sons, Ltd.

Therefore, it is the best of the analysed schemes presenting the best performance in unsteady flows. Nevertheless, this is at the cost of greater complexity. The results on shallow water equations are better than the first-order upwind schemes for highly unsteady flows but identical for steady flows.

A new entropy correction has been proposed. It can be applied both to upwind or central schemes, independently of TVD corrections, and it always improves the solutions. The *dog-leg* effect is negligible as is shown in the dam-break tests. The Lax–Wendroff scheme in particular has been enhanced with the entropy correction. If maximal accuracy in unsteady flows is required, one of the second-order TVD schemes is recommended. In these schemes, upwind and limiting treatment of source terms produces better results than pointwise treatment of these terms which leads to oscillations and conservation properties. Furthermore, a slight improvement is achieved incorporating the source terms into the flux limiter.

It must finally be stressed that the non-pointwise treatment of the source terms represents a remarkable improvement in all the numerical schemes either upwind or not since it allows a good balance between flux and source terms in all cases. Furthermore, when using flux limited schemes, this balance has to be maintained by limiting in an analogous form the fluxes and the sources.

APPENDIX A. SOLUTION TO THE ENTROPY PROBLEM WITH ARTIFICIAL VISCOSITY

A.1. Scalar case

For the resolution of scalar propagation equation like

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0, \quad a = a(x, u)$$

If there is a transition from subcritical ($\alpha < 0$) to supercritical ($\alpha > 0$) flow between two nodal points in the grid, the way in which the situation will evolve in one Δt can be studied. This is illustrated in Figure 29.

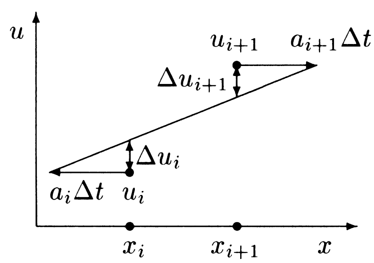

Figure 29. Evolution of a propagation wave in a cell of the mesh with transcritical flux at an infinitesimal time Δt.

Copyright © 2001 John Wiley & Sons, Ltd.

The value of the increments in the variable can be calculated by linear interpolation

$$\Delta u_i = -\frac{a_i \Delta t}{\delta x + (a_{i+1} - a_i)\Delta t}\,\delta u_{i+(1/2)} = -\frac{\sigma_i \delta u_{i+(1/2)}}{1 + \delta\sigma_{i+(1/2)}}$$

$$\Delta u_{i+1} = -\frac{a_{i+1} \Delta t}{\delta x + (a_{i+1} - a_i)\Delta t}\,\delta u_{i+(1/2)} = -\frac{\sigma_{i+1} \delta u_{i+(1/2)}}{1 + \delta\sigma_{i+(1/2)}}$$

with $\sigma = \alpha(\Delta t/\delta x)$. These will be increments in the first-order. Then, the viscosity necessary in a numerical scheme to solve properly this kind of transition can be found. Concentrating on the grid cell where the transition takes place, the conservative numerical schemes in general will follow a wave decomposition like

$$\delta F^{\mathrm{L}}_{i+(1/2)} = a^{\mathrm{L}}_{i+(1/2)}\delta u_{i+(1/2)}, \qquad \delta F^{\mathrm{R}}_{i+(1/2)} = a^{\mathrm{R}}_{i+(1/2)}\delta u_{i+(1/2)}$$

$$\Delta u_{i+1} = -\sigma^{\mathrm{L}}_{i+(1/2)} = \delta u_{i+(1/2)}, \qquad \Delta u_i = -\sigma^{\mathrm{R}}_{i+(1/2)} = \delta u_{i+(1/2)}$$

Adding artificial viscosity to fix the entropy problem

$$\Delta u_{i+1} = -(\sigma^{\mathrm{L}}_{i+(1/2)} + v_{i+(1/2)})\delta u_{i+(1/2)}, \qquad \Delta u_i = -(\sigma^{\mathrm{R}}_{i+(1/2)} - v_{i+(1/2)})\delta u_{i+(1/2)}$$

Identifying these increments with those obtained by interpolation two possible values are found. To avoid problems, the maximum is chosen

$$v_{i+(1/2)} = \max\left(\sigma^{\mathrm{R}}_{i+(1/2)} - \frac{\sigma_i}{1 + \delta\sigma_{i+(1/2)}}, \frac{\sigma_{i+1}}{1 + \delta\sigma_{i+(1/2)}} - \sigma^{\mathrm{L}}_{i+(1/2)}\right)$$

This viscosity, having been obtained from a linear interpolation, gives a correction in the first-order, then it is valid for first-order schemes. Furthermore, considering that there will be a finite number of transitions like this, it can be acceptable also for second-order schemes. As the only first-order scheme considered with entropy problems is the first-order upwind, and the second-order schemes can be expressed in terms of that first-order scheme plus second-order corrections, the previous solution is written as

$$v_{i+(1/2)} = \max\left(\sigma^{-}_{i+(1/2)} - \frac{\sigma_i}{1 + \delta\sigma_{i+(1/2)}}, \frac{\sigma_{i+1}}{1 + \delta\sigma_{i+(1/2)}} - \sigma^{+}_{i+(1/2)}\right) \tag{A.1}$$

Other forms of artificial viscosity to fix the entropy problem are described in Reference [1].

A.2. Systems of equations

When dealing with homogeneous systems of equations, in order to study separately the behaviour of every wave, the system is first formulated in characteristic variables

Copyright © 2001 John Wiley & Sons, Ltd.

$$\frac{\partial w^k}{\partial t} + a^k \frac{\partial w^k}{\partial x} = 0$$

The entropy problem can be fixed by analogy to the scalar case acting over every k component in the decoupled system. The artificial viscosity is then defined as a diagonal matrix. For the particular case of having a transition sub-super between nodes i and $i+1$ in the k component

$$(\mathbf{V}^{kkk})_{i+(1/2)} = \max\left((\sigma^k)^-_{i+(1/2)} - \frac{(\sigma^k)_i}{1+\delta(\sigma^k)_{i+(1/2)}}, \frac{(\sigma^k)_{i+1}}{1+\delta(\sigma^k)_{i+(1/2)}} - (\sigma^k)^+_{i+(1/2)}\right) \tag{A.2}$$

with

$$\sigma^k = \frac{\Delta t}{\delta x} \alpha^k$$

The characteristic system then is transformed to

$$\frac{\partial \vec{w}}{\partial t} + \mathbf{\Lambda} \frac{\partial \vec{w}}{\partial x} = \mathbf{V} \frac{\partial^2 \vec{w}}{\partial x^2}$$

Returning to the physical variables by means of $\mathbf{P}$ and $\mathbf{P}^{-1}$, the matrices that diagonalize the Jacobian

$$\mathbf{P}\left(\frac{\partial \vec{w}}{\partial t} + \mathbf{\Lambda} \frac{\partial \vec{w}}{\partial x} = \mathbf{V} \frac{\partial^2 \vec{w}}{\partial x^2}\right) \Rightarrow \frac{\partial \vec{u}}{\partial t} + \mathbf{J} \frac{\partial \vec{u}}{\partial x} = \mathbf{PVP}^{-1} \frac{\partial^2 \vec{u}}{\partial x^2}$$

and the artificial viscosity is defined as

$$v_{i+(1/2)} = (\mathbf{PVP}^{-1})_{i+(1/2)} \tag{A.3}$$

with $\mathbf{V}$ defined in (A.2). As a last remark, the entropy problem does not only affect a transition in fluxes and is not modified by the presence of source terms. Hence, in the presence of source terms, (A.3) will still be used.

APPENDIX B. BOUNDARY CONDITIONS

For the treatment at the boundaries, the distinction between numerical boundary conditions and external (imposed) boundary conditions has been exploited. The use of the characteristic variables to obtain the correct region of dependence of a point is a suitable method to produce numerical boundary conditions [1]. In this work, however, a similar way to generate numerical boundary conditions for any conservative scheme keeping the degree of accuracy of the scheme has been developed. All the conservative schemes admit a decomposition in sum of contributions from left and right like

Copyright © 2001 John Wiley & Sons, Ltd.

$$\Delta\vec{u}_i^n = \Delta t(\vec{G}^{\mathrm{R}}_{i+(1/2)} + \vec{G}^{\mathrm{L}}_{i-(1/2)}) \tag{B.1}$$

A decentralized treatment of source terms is supposed but a pointwise treatment is also possible. At the boundaries, the following information can be used:

$$\text{upstream: } \Delta\vec{u}_i^n = \Delta t\vec{G}^{\mathrm{R}}_{i+(1/2)}, \qquad \text{downstream: } \Delta\vec{u}_i^n = \Delta t\vec{G}^{\mathrm{L}}_{i-(1/2)}$$

Therefore, the numerical boundary conditions can be worked out from the scheme itself whenever the domain of dependence of the boundary points is inside the calculation grid. Otherwise, physical or imposed external information must be used. Let us call $\Delta\vec{u}_i^{\mathrm{N}}$ the numerical increments of the variable obtained using (B.1), and $\Delta\vec{u}_i^{\mathrm{F}}$ the physical increments, that is, the final values that we want to determine. Then, if w^k is the characteristic variable associated to a propagation velocity α^k that requires a numerical boundary condition, the following will be used.

$$(\Delta w^k)_i^{\mathrm{F}} = (\Delta w^k)_i^{\mathrm{N}} \tag{B.2}$$

The procedure to implement this technique in the shallow water equations requires first us to know the kind of flow at every boundary. Then, (B.1) is used to get the numerical increments $\Delta\vec{u}_i^{\mathrm{N}}$. Once calculated, they are used in (B.2) to obtain two possible numerical boundary conditions, one associated to the velocity $v+c$ and the other associated to the velocity $v-c$. The characteristic variables in this problem can be defined from

$$\Delta\vec{w} = \frac{1}{2c}\begin{pmatrix}(c-v)\Delta A + \Delta Q\\(c+v)\Delta A - Q)\end{pmatrix}$$

And there are four possibilities:

1. Subcritical inlet: one physical and one numerical boundary condition

$$(c+v)_i^n\Delta A_i^{\mathrm{F}} - \Delta Q_i^{\mathrm{F}} = (c+v)_i^n\Delta A_i^{\mathrm{N}} - \Delta Q_i^{\mathrm{N}}$$

2. Supercritical inlet: two physical boundary conditions

$$\Delta A_i^{\mathrm{F}} = \text{datum}, \quad \Delta Q_i^{\mathrm{F}} = \text{datum}$$

3. Subcritical outlet: one physical and one numerical boundary condition

$$(c-v)_i^n\Delta A_i^{\mathrm{F}} + \Delta Q_i^{\mathrm{F}} = (c-v)_i^n\Delta A_i^{\mathrm{N}} + \Delta Q_i^{\mathrm{N}}$$

4. Supercritical outlet: two numerical boundary conditions

$$\Delta A_i^{\mathrm{F}} = \Delta A_i^{\mathrm{N}}, \quad \Delta Q_i^{\mathrm{F}} = \Delta A_i^{\mathrm{N}}$$

Copyright © 2001 John Wiley & Sons, Ltd.

REFERENCES

1. Hirsch C. *Computational Methods for Inviscid and Viscous Flows: Numerical Computation of Internal and External Flows*, vol. 2. Wiley: New York, 1990.
2. Toro E. *Riemann Solves and Numerical Methods for Fluid Dynamics: A Practical Introduction*. Springer: Berlin, 1997.
3. Glaister P. Prediction of supercritical flow in open channels. *Computing in Mathematics Applications* 1992; **24**(7): 69–75.
4. Alcrudo F, García-Navarro P, Saviron JM. Flux difference splitting for 1D open channel flow equations. *International Journal for Numerical Methods in Fluids* 1991; **13**: 85–95.
5. Smolarkiewicz PK, Margolin LG. MPDATA: a finite-difference solver for geophysical flows. *Journal of Computational Physics* 1998; **140**: 459–480.
6. LeVeque RJ. Balancing source terms and flux gradients in high-resolution Godunov methods: the quasi-steady wave-propagation algorithm. *Journal of Computational Physics* 1998; **146**(1): 346–365.
7. Roe PL. Approximate Riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics* 1981; **43**(2): 357–372.
8. Vázquez-Cendón ME. Improved treatment of source terms in upwind schemes for the shallow water equations in channels with irregular geometry. *Journal of Computational Physics* 1999; **148**(2): 497–526.
9. Bermúdez A, Vázquez ME. Upwind methods for hyperbolic conservation laws with source terms. *Computers and Fluids* 1994; **23**(8): 1049–1071.
10. García Navarro P, Vázquez Cendón ME. Some considerations and improvements on the performance of Roe's schemes for 1D irregular geometries on numerical treatment of the source terms in the shallow water equations. *Computers and Fluids* 2000; **126**: 26–40.
11. García Navarro P, Hubbard ME. Flux difference splitting and the balancing of source terms and flux gradients. Internal Report 3/99, Department of Mathematics, University of Reading, 1999.
12. Stoker JJ. *Water Waves*. Interscience: New York, 1957.
13. McDonald I. Analysis and computation of steady open channel flow. PhD thesis, University of Reading, 1996.
14. LeVeque RJ. *Numerical Methods for Conservation Laws*. Birkhäuser: Basel, 1992.
15. Sweby PK. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM Journal of Numerical Analysis* 1984; **21**: 995–1011.

Copyright © 2001 John Wiley & Sons, Ltd.