Journal of Computational Physics **165,** 89–125 (2000)
doi:10.1006/jcph.2000.6603, available online at http://www.idealibrary.com on IDEAL®

Flux Difference Splitting and the Balancing of Source Terms and Flux Gradients

M. E. Hubbard*,[1] and P. Garcia-Navarro†

*DAMTP, University of Cambridge, Silver Street, Cambridge CB3 9EW, United Kingdom, and †Area de Mecanica de Fluidos, Centro Politecnico Superior, C/Maria de Luna 3, 50015 Zaragoza, Spain
E-mail: M.E.Hubbard@damtp.cam.ac.uk

Received July 13, 1999; revised August 10, 2000; published online November 3, 2000

Flux difference splitting methods are widely used for the numerical approximation of homogeneous conservation laws where the flux depends only on the conservative variables. However, in many practical situations this is not the case. Not only are source terms commonly part of the mathematical model, but also the flux can vary spatially even when the conservative variables do not. It is the discretisation of the additional terms arising from these two situations which is addressed in this work, given that a specific flux difference splitting method has been used to approximate the underlying conservation law. The discretisation is constructed in a manner which retains an exact balance between the flux gradients and the source terms when this is appropriate. The effectiveness of these new techniques, in both one and two dimensions, is illustrated using the shallow water equations, in which the additional terms arise from the modelling of bed slope and, in one dimension, breadth variation. Roe's scheme is chosen for the approximation of the conservation laws and appropriate discrete forms are constructed for the additional terms, not only in the first-order case but also in the presence of flux- and slope-limited high-resolution corrections. The method is then extended to two-dimensional flow where it can be applied on both quadrilateral and triangular grids. © 2000 Academic Press

Key Words: source terms; finite volume schemes; upwind schemes; flux limiters; slope limiters; conservation laws.

1. INTRODUCTION

There has been much research in CFD into the accurate and efficient solution of homogeneous systems of conservation laws. More recently, as numerical models become more

[1] After 18th September 18, 2000, address is School of Computing, University of Leeds, Leeds, LS2 9JT, United Kingdom.

0021-9991/00 $35.00
Copyright © 2000 by Academic Press
All rights of reproduction in any form reserved.

complicated and the areas of application of these methods widens, it has become important that other aspects of the discretisation be given due attention. This is certainly true in the field of computational hydraulics where modelling can be dominated by the effects not only of source terms, but also of quantities which vary spatially but independent of the flow variables. In this paper methods will be constructed which address both of these issues in a range of situations.

It can be argued that the presence of these effects warrants the construction of new numerical schemes which are appropriate to the nature of the equations, rather than the use of one of the many which have been constructed for the simple, homogeneous case. However, the emphasis of this work is on how the additional terms should be discretised, given that a specific conservative finite volume scheme (Roe's scheme is used here to illustrate the ideas) has been used to approximate the flux terms, and the main focus is on how the numerical scheme can be modified so that it maintains any balance between flux and source terms which exists as part of the underlying mathematical model, e.g., at the steady state. This approach has been taken previously by a number of authors and applied in a variety of different situations. For example, Smolarkiewicz has adapted his own MPDATA scheme to solve inhomogeneous equations arising from geophysical flows [19]; LeVeque has incorporated the modelling of source terms for shallow water flows within his wave-propagation algorithm [14]; Jenny and Muller used ideas similar to those presented here to approximate source terms within their characteristic-based scheme [12]; and Roe's scheme [16] has been modified by a number of authors to include source terms, the research of Glaister [7], Vázquez-Cendón [21], Bermúdez and Vázquez [1], and Bermúdez *et al.* [2] being of particular relevance to this work.

In each of the aforementioned papers discussing Roe's scheme, the discrete form of the source terms has been deliberately constructed along lines similar to the numerical fluxes. This is done to ensure that equilibria which occur in the mathematical model are retained by the numerical model, and that in the absence of additional terms, the conservative fluxes are retrieved for accurate modelling of discontinuous solutions. The intention of this paper is to concentrate on finite volume schemes based on the concept of a numerical flux and how the source term should be discretised within such a framework to correctly balance the resulting flux differences. In particular, it aims to show how these ideas apply to higher order total variation diminishing (TVD) versions of Roe's scheme (using both flux- and slope-limiting techniques) and to describe a source term approximation which has each of the above properties on all types of regular and irregular grids in any number of dimensions. Furthermore, following on from [5], a new formulation is presented for the discretisation of the flux in the case where it depends on a spatially varying quantity which is independent of the solution.

The shallow water equations have been chosen to demonstrate the effectiveness of these new techniques in one and two dimensions, by modelling the effects of a sloping bed and, in one dimension only, the inclusion of breadth variation in an open channel. The one-dimensional discretisation is described first, for a scalar equation in Section 2 and then, in Section 3, for a general system of conservation laws, followed by its application to the shallow water equations and a wide selection of results to show its accuracy. In Section 4 the generalisation to two dimensions, illustrated using unstructured triangular grids, is presented and again applied to the shallow water equations. The final section contains some brief conclusions obtained from the work.

2. ONE-DIMENSIONAL SCALAR EQUATIONS

The basic ideas underlying the following work can be illustrated by examining a scalar equation,

$$u_t + f(u,x)_x = s(u,x), \tag{2.1}$$

modelling the evolution of a single variable u in one space dimension (f is a flux and s a source, both of which depend on x as well as u), and looking for (1) a way of incorporating the x-dependence of the flux into the numerical scheme and (2) a discretisation which maintains equilibria of the form $f_x \equiv s$ (steady state). In essence this requires the flux derivatives and the source terms to be discretised in a similar manner. However, the precise form this implies for the discrete source term is not always obvious from the flux approximation.

As an example, consider the first-order upwind scheme with forward Euler time-stepping for the homogeneous equation ($s \equiv 0$ in (2.1)) which, in fluctuation-signal form [17], can be written as

$$u_i^{n+1} \;=\; u_i^n - \frac{\Delta t}{\Delta x_i}\left(\Delta f_{i-1/2}^{+} + \Delta f_{i+1/2}^{-}\right). \tag{2.2}$$

Note that i here is a cell index, not a node index, so this is not the traditional fluctuation-signal framework, but instead a device which will be used to illustrate the form that the discrete source terms should take when the equivalent finite volume approach is used to approximate the fluxes. In the case where f depends on both u and x (case 1 above),

$$\Delta f_{i+1/2} = \left(\widetilde{\frac{\partial f}{\partial u}}\right)_{i+1/2} \Delta u_{i+1/2} + \left(\widetilde{\frac{\partial f}{\partial x}}\right)_{i+1/2} \Delta x_{i+1/2} = \tilde{\lambda}_{i+1/2}\Delta u_{i+1/2} + \tilde{V}_{i+1/2}, \tag{2.3}$$

in which λ is the advection velocity and V is the extra term due to the independent spatial variation, so the scheme (2.2) is completed by setting

$$\Delta f_{i+1/2}^{\pm} = \tilde{\lambda}_{i+1/2}^{\pm}\Delta u_{i+1/2} + \frac{1}{2}\left(1 \pm \operatorname{sgn}\left(\tilde{\lambda}_{i+1/2}\right)\right)\tilde{V}_{i+1/2}, \tag{2.4}$$

where $\tilde{\lambda}^{\pm} = \frac{1}{2}(\tilde{\lambda} \pm |\tilde{\lambda}|)$. In both equations $\sim$ represents a Roe-averaged quantity evaluated at a cell interface which is defined to ensure that (2.4) is satisfied. Straightforward algebraic manipulation converts (2.2) to an equivalent flux-based finite volume scheme,

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x_i}\left(f_{i+1/2}^{*} - f_{i-1/2}^{*}\right), \tag{2.5}$$

in which the numerical flux (denoted by an asterisk) for first-order upwinding is given by

$$\begin{aligned} f_{i+1/2}^{*} &= \frac{1}{2}(f_{i+1} + f_i) - \frac{1}{2}\operatorname{sgn}\left(\tilde{\lambda}_{i+1/2}\right)(f_{i+1} - f_i) \\ &= \frac{1}{2}(f_{i+1} + f_i) - \frac{1}{2}\left(\left|\tilde{\lambda}_{i+1/2}\right|\Delta u_{i+1/2} + \operatorname{sgn}\left(\tilde{\lambda}_{i+1/2}\right)\tilde{V}_{i+1/2}\right), \end{aligned} \tag{2.6}$$

with a similar expression for $f_{i-1/2}^{*}$. The dependence of the flux on x is incorporated within the evaluation of the numerical flux. Note that throughout this paper the asterisk indicates an approximation of the flux across a cell face or the source over a cell volume.

In the case when $s \not\equiv 0$ (case 2 above), the numerical scheme (2.2) becomes

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x_i}\left(\Delta f_{i-1/2}^+ + \Delta f_{i+1/2}^-\right) + \frac{\Delta t}{\Delta x_i}\mathbf{s}_i^*, \tag{2.7}$$

but it remains to define $\mathbf{s}_i^*$. The bold type used here (and later) indicates that it is the integrated source term which is being approximated, so a factor of the cell length/area is incorporated within it. In this work the discretisation is constructed so that (2.7) maintains a balance of the form $f_x \equiv s$; i.e., in such situations the discrete form gives

$$\mathbf{s}_i^* \equiv \Delta f_{i-1/2}^+ + \Delta f_{i+1/2}^-. \tag{2.8}$$

Now, given this equilibrium, at an interface

$$\Delta f_{i+1/2} = \tilde{\lambda}_{i+1/2}\Delta u_{i+1/2} + \tilde{V}_{i+1/2} = \tilde{\mathbf{s}}_{i+1/2}, \tag{2.9}$$

as long as the flux and source terms are evaluated in the same way, and at the same interface state (something which should be borne in mind throughout this paper). Comparing (2.4) and (2.9) leads to expressions for components of the source term associated with the local wave speed, which can be written

$$\Delta f_{i+1/2}^\pm = \frac{1}{2}\left(1 \pm \operatorname{sgn}\left(\tilde{\lambda}_{i+1/2}\right)\right)\tilde{\mathbf{s}}_{i+1/2} = \tilde{\mathbf{s}}_{i+1/2}^\pm, \tag{2.10}$$

and it becomes obvious from the requirement that (2.8) should be satisfied that

$$\begin{aligned}\mathbf{s}_i^* &= \tilde{\mathbf{s}}_{i-1/2}^+ + \tilde{\mathbf{s}}_{i+1/2}^- \\ &= \frac{1}{2}\left(\left[1 + \operatorname{sgn}\left(\tilde{\lambda}_{i-1/2}\right)\right]\tilde{\mathbf{s}}_{i-1/2}\right) + \frac{1}{2}\left(\left[1 - \operatorname{sgn}\left(\tilde{\lambda}_{i+1/2}\right)\right]\tilde{\mathbf{s}}_{i+1/2}\right)\end{aligned} \tag{2.11}$$

is the form that should be substituted into the flux-based form of the complete scheme

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x_i}\left(f_{i+1/2}^* - f_{i-1/2}^*\right) + \frac{\Delta t}{\Delta x_i}\mathbf{s}_i^*. \tag{2.12}$$

Note that $\mathbf{s}_i^*$ is an approximation to the integral of the source term over cell i, so $\mathbf{s}_i^*/\Delta x_i$ represents some cell-averaged value. This allows Δx_i to appear in both the flux and the source discretisations so that the balance achieved here can be maintained if irregular grids are used. Furthermore, $\tilde{\mathbf{s}}_{i+1/2} = \tilde{\mathbf{s}}_{i+1/2}^- + \tilde{\mathbf{s}}_{i+1/2}^+$ can be thought of as approximating an integral across the $i + 1/2$ cell edge (i.e., over a dual cell rather than a grid cell) so that its discrete form can be related to, and balanced with, the flux difference across this edge. The integrated quantity then contributes to the cell updates on either side of the edge according to the local wave speeds, as indicated in Fig. 1.

This type of discretisation can be used for any type of source term, but it is designed specifically for equations such as

$$u_t + \left(\frac{u^2}{2}\right)_x = a u_x, \tag{2.13}$$

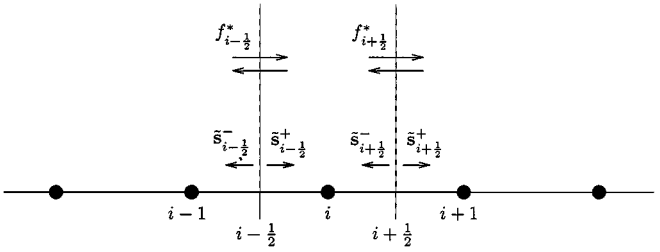

FIG. 1. Scalar flux and source distribution within a control volume.

which clearly has a balance when $a \equiv u$. The extra terms which model variable topography in the shallow water equations give an example of this type of equation. It is less obvious that this treatment would be beneficial to the approximation of, for example,

$$u_t + (\lambda u)_x = u^3, \tag{2.14}$$

which has no spatial derivative on the right hand side, and hence no obvious equilibrium with the flux derivative. Bed friction terms are more akin to this form and, while an upwind discretisation should still be beneficial, a full investigation is left to future work.

It should be noted that even when a in Eq. (2.13) varies with x and is supplied analytically, it should be evaluated at cell interfaces using a Roe-average form of the cell centre data and not the exact values there, otherwise the discrete balance $\tilde{a} \equiv \tilde{u}$ would be disturbed and the equilibrium destroyed.

The above analysis is not restricted to the first-order upwind scheme. For example, the cell-based fluctuation-signal form of the Lax–Wendroff scheme,

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x_i}\left(\frac{1}{2}\left[\left(1+\nu_{i-1/2}\right)\Delta f_{i-1/2}\right] + \frac{1}{2}\left[\left(1-\nu_{i+1/2}\right)\Delta f_{i+1/2}\right]\right) + \frac{\Delta t}{\Delta x_i}\mathbf{s}_i^*, \tag{2.15}$$

in which $\nu = \lambda \Delta x/\Delta t$ is the local CFL number and Δf is as in (2.3), leads to the corresponding numerical flux

$$\begin{aligned} f^*_{i+1/2} &= \frac{1}{2}(f_{i+1} + f_i) - \frac{\nu}{2}(f_{i+1} - f_i) \\ &= \frac{1}{2}(f_{i+1} + f_i) - \frac{1}{2}\operatorname{sgn}\left(\tilde{\lambda}_{i+1/2}\right)\left[1 - \left(1 - \left|\nu_{i+1/2}\right|\right)\right](f_{i+1} - f_i) \end{aligned} \tag{2.16}$$

and subsequently to the numerical source

$$\begin{aligned} \mathbf{s}_i^* = \frac{1}{2}&\left[\left(1 + \operatorname{sgn}\left(\tilde{\lambda}_{i-1/2}\right)\left[1 - \left(1 - \left|\nu_{i-1/2}\right|\right)\right]\right)\tilde{\mathbf{s}}_{i-1/2}\right] \\ &+ \frac{1}{2}\left[\left(1 - \operatorname{sgn}\left(\tilde{\lambda}_{i+1/2}\right)\left[1 - \left(1 - |\nu|_{i+1/2}\right)\right]\right)\tilde{\mathbf{s}}_{i+1/2}\right]. \end{aligned} \tag{2.17}$$

Approximations now follow automatically for flux-limited TVD schemes, since

$$f^*_{\mathrm{TVD}} = f^*_{\mathrm{UP}} + L(r)(f^*_{\mathrm{LW}} - f^*_{\mathrm{UP}}) \tag{2.18}$$

(where the subscripts have their obvious interpretations), which implies that

$$\mathbf{s}^*_{\mathrm{TVD}} = \mathbf{s}^*_{\mathrm{UP}} + L(r)(\mathbf{s}^*_{\mathrm{LW}} - \mathbf{s}^*_{\mathrm{UP}}). \tag{2.19}$$

In the above equations $L(r)$ represents a nonlinear flux-limiter function, as described in [13, 20], and

$$r = \frac{\Delta u^{\mathrm{upwind}}}{\Delta u^{\mathrm{local}}}, \tag{2.20}$$

where $\Delta u^{\mathrm{local}}$ is the change in u across the current cell edge and $\Delta u^{\mathrm{upwind}}$ is the change across the upwind edge according to the local advection velocity $\tilde{\lambda}$. The important properties of L are that for smooth flow the accuracy of the scheme is second order (when $L(r) = 1$ it gives the Lax–Wendroff scheme) but near discontinuities and other local extrema of the solution where $r < 0$ the scheme reduces to first-order accuracy in order to avoid the introduction of unwanted oscillations into the solution.

The limited numerical flux is

$$f^*_{i+1/2} = \frac{1}{2}(f_{i+1} + f_i) - \frac{1}{2}\operatorname{sgn}\left(\tilde{\lambda}_{i+1/2}\right)\left(1 - L\left(r_{i+1/2}\right)\left(1 - \left|\nu_{i+1/2}\right|\right)\right)(f_{i+1} - f_i) \tag{2.21}$$

so the corresponding discrete source term should be

$$\begin{aligned}\mathbf{s}^*_i = {} & \frac{1}{2}\left[\left(1 + \operatorname{sgn}\left(\tilde{\lambda}_{i-1/2}\right)\left[1 - L\left(r_{i-1/2}\right)\left(1 - \left|\nu_{i-1/2}\right|\right)\right]\right)\tilde{\mathbf{s}}_{i-1/2}\right] \\ & + \frac{1}{2}\left[\left(1 - \operatorname{sgn}\left(\tilde{\lambda}_{i+1/2}\right)\left[1 - L\left(r_{i+1/2}\right)\left(1 - \left|\nu_{i+1/2}\right|\right)\right]\right)\tilde{\mathbf{s}}_{i+1/2}\right].\end{aligned} \tag{2.22}$$

In the slope-limited case, the numerical sources are (a) evaluated as in (2.11) but using the reconstructed solution values at the cell edge and (b) augmented by a correction term which arises due to the linear variation of the solution over the cell. The latter modification is necessary because whereas previously

$$\Delta f_{i+1/2} + \Delta f_{i-1/2} = (f_{i+1} + f_i) + (f_i + f_{i-1}), \tag{2.23}$$

it is now only true to say

$$\Delta f_{i+1/2} + \Delta f_{i-1/2} = \left(f^{\mathrm{R}}_{i+1/2} + f^{\mathrm{L}}_{i+1/2}\right) - \left(f^{\mathrm{R}}_{i-1/2} + f^{\mathrm{L}}_{i-1/2}\right) - 2\left(f^{\mathrm{L}}_{i+1/2} - f^{\mathrm{R}}_{i-1/2}\right), \tag{2.24}$$

where the superscripts R and L represent evaluation on, respectively, the right and left hand sides of the interface indicated by the associated subscript, so the numerical source term is defined to be

$$\mathbf{s}^*_i = \left(\tilde{\mathbf{s}}^-_{i+1/2} + \tilde{\mathbf{s}}^+_{i-1/2}\right) - \tilde{\mathbf{s}}\left(u^{\mathrm{L}}_{i+1/2}, u^{\mathrm{R}}_{i-1/2}\right). \tag{2.25}$$

The first term on the right hand side is evaluated precisely as in (2.11) except that the interface values are now those of the MUSCL reconstruction of the solution within each

cell. The second term is evaluated at the Roe-average state over the cell (indicated again by $\sim$), which ensures that

$$\tilde{\mathbf{s}}\left(u^{\mathrm{L}}_{i+1/2}, u^{\mathrm{R}}_{i-1/2}\right) = f^{\mathrm{L}}_{i+1/2} - f^{\mathrm{R}}_{i-1/2} \tag{2.26}$$

at equilibrium (as was done in (2.9)), so the balance is maintained with the final term in (2.24) within cell i. As will be noted later, the correction term could be improved away from the chosen equilibrium.

It should be noted that the TVD condition which the flux and slope limiters have been constructed to satisfy applies to homogeneous equations, and the inclusion of source terms means that spurious oscillations (i.e., those which appear simply as numerical artifacts, not as part of the actual solution in the presence of source terms) may appear in the final solution. This problem has yet to be fully addressed.

These ideas are discussed more thoroughly in the following sections, in the practical context of the shallow water equations, where it is shown how they extend to nonlinear systems of equations in one and more space dimensions on arbitrary grids, although the emphasis is placed on steady-state and nearly steady problems.

3. ONE DIMENSION

The one-dimensional equations representing a system of conservation laws with source terms may be written

$$\underline{U}_t + \underline{F}_x = \underline{S}, \tag{3.1}$$

where $\underline{U}$ is the vector of conservative variables, $\underline{F}$ is the conservative flux vector, and $\underline{S}$ includes all of the source terms. In this section it is assumed that $\underline{F} = \underline{F}(\underline{U})$; in Section 3.2 the flux will be assumed to depend not only on the conservative variables but also another independent, spatially varying quantity, i.e., $\underline{F} = \underline{F}(\underline{U}, b(x))$.

Using the standard finite volume approximation of the flux terms in (3.1), combined with a simple, forward Euler discretisation of the time derivative leads to a difference scheme which can be written

$$\underline{U}^{n+1}_i = \underline{U}^n_i - \frac{\Delta t}{\Delta x_i}\left(\underline{F}^*_{i+1/2} - \underline{F}^*_{i-1/2}\right) + \frac{\Delta t}{\Delta x_i}\underline{\mathbf{S}}^*_i, \tag{3.2}$$

in which $\underline{F}^*$ represents a numerical flux evaluated at an interface between control volumes and $\underline{\mathbf{S}}^* \approx \int \underline{S}\,\mathrm{d}x$ is a numerical source integral over the control volume, which has yet to be approximated. For convenience, a cell centre scheme in which the control volumes coincide with the mesh cells has been considered throughout this work, although the ideas may be applied to other types of schemes in a similar manner.

At first sight, the second term on the right hand side of (3.2) looks like a discrete flux derivative for cell i. However, for the purposes of this work it is more convenient to consider it from the point of view of the numerical fluxes being constructed from an approximation to the integral of the flux derivatives over dual cells and providing contributions to the cell updates (Δx_i comes from the integration of the original equations over the control volume).

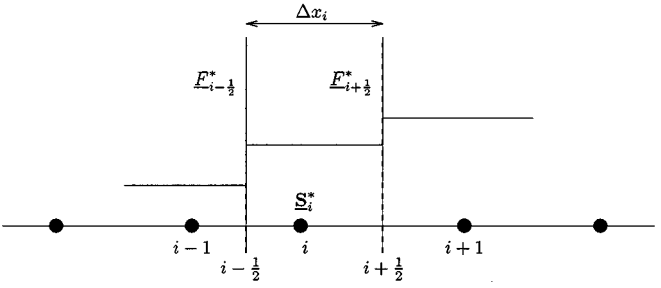

FIG. 2. Numerical fluxes and sources for the cell centre scheme.

Commonly, $\underline{\mathbf{S}}_i^*$ is evaluated pointwise, taking the value $\Delta x_i\, \underline{S}(\underline{U}_i)$, or split symmetrically, giving an expression of the form

$$\underline{\mathbf{S}}_i^* = \frac{\Delta x_i}{2}\left(\underline{S}\left(\underline{U}_{i-1/2}\right) + \underline{S}\left(\underline{U}_{i+1/2}\right)\right), \tag{3.3}$$

but a more sophisticated approach is sought here, based on the approach of Glaister [7], which accounts for the form of the numerical fluxes.

Note that in the absence of source terms the scheme given by (3.2) reduces to a conservative discretisation of the homogeneous system. Also, (3.2) has been written with irregular grids in mind, and as a consequence the mesh spacing $\Delta x_i = x_{i+1/2} - x_{i-1/2}$ relates to the cells, not the nodes (see Fig. 2).

Roe's scheme [16] is one of the most commonly used examples of the conservative finite volume method mentioned above. This is an upwind scheme which uses an approximate Riemann solver to decompose the flux terms into characteristic components by diagonalisation of the homogeneous part of a linearised form of the system (3.1), which is

$$\underline{U}_t + \tilde{\mathbf{A}}\,\underline{U}_x = \underline{0}, \tag{3.4}$$

where $\tilde{\mathbf{A}} \approx \partial \underline{F}/\partial \underline{U}$ is the linearised flux Jacobian of the system. The Riemann problems arise at the interfaces between the control volumes (the mesh nodes in this case) where discontinuities occur in the discrete representation of the solution.

Application of Roe's Riemann solver results in a decoupling of the linearised equations that splits the flux difference so that it can be written in a number of equivalent forms; i.e., at an interface

$$\Delta\,\underline{F}_{i+1/2} = (\tilde{\mathbf{A}}\Delta\,\underline{U})_{i+1/2} = (\tilde{\mathbf{R}}\,\tilde{\mathbf{\Lambda}}\tilde{\mathbf{R}}^{-1}\Delta\,\underline{U})_{i+1/2} = \left(\sum_{k=1}^{N_w} \tilde{\alpha}_k \tilde{\lambda}_k \underline{\tilde{r}}_k\right)_{i+1/2}, \tag{3.5}$$

in which $\Delta\,\underline{F}$ represents the jump in $\underline{F}$ across the edge of a control volume, $\tilde{\mathbf{R}}$ is the matrix whose columns are the right eigenvectors $\underline{\tilde{r}}_k$ of $\tilde{\mathbf{A}}$, $\tilde{\mathbf{\Lambda}}$ is the diagonal matrix of eigenvalues $\tilde{\lambda}_k$ of $\tilde{\mathbf{A}}$, and the components of $\tilde{\mathbf{R}}^{-1}\Delta\,\underline{U}(=\Delta\underline{W})$ are the strengths $\tilde{\alpha}_k$ associated with each component of the decomposition ($\underline{W}$ being the vector of characteristic variables of the system). The final expression indicates how the flux difference is decomposed into N_w characteristic components (or waves of the Riemann problem), where N_w is the number of

equations of the system (3.4). In both (3.4) and (3.5) (and throughout the rest of this paper), $\sim$ denotes the evaluation of a quantity at its Roe-average state [6, 16]. This is a special average state of the flow variables which is constructed so that (3.5) is always satisfied for the given system.

Having obtained the decomposition (3.5), we construct Roe's scheme for a homogeneous system of equations from (3.2) by taking the numerical fluxes to be

$$\underline{F}^*_{i+1/2} = \frac{1}{2}\left(\underline{F}_{i+1} + \underline{F}_i\right) - \frac{1}{2}(\tilde{\mathbf{R}}|\tilde{\mathbf{\Lambda}}|\tilde{\mathbf{R}}^{-1}\Delta\underline{U})_{i+1/2}, \tag{3.6}$$

where $|\tilde{\mathbf{\Lambda}}| = \mathrm{diag}(|\tilde{\lambda}_k|)$; the source terms have been temporarily ignored. A similar expression can be written down for $\underline{F}^*_{i-1/2}$.

Choosing the Roe-average state (represented by $\sim$) to satisfy (3.5) means that the resulting approximate Riemann solver is an exact solver for this local linearisation of the Riemann problem. More importantly, in the context of this work, when (3.5) is combined with (3.6) the nodal update scheme given by (3.2) is equivalent to the fluctuation-signal style scheme [17] mentioned earlier and given by

$$\underline{U}^{n+1}_i = \underline{U}^n_i - \frac{\Delta t}{\Delta x_i}\left[(\tilde{\mathbf{R}}\tilde{\mathbf{\Lambda}}^-\tilde{\mathbf{R}}^{-1}\Delta\underline{U})_{i+1/2} + (\tilde{\mathbf{R}}\tilde{\mathbf{\Lambda}}^+\tilde{\mathbf{R}}^{-1}\Delta\underline{U})_{i-1/2}\right] + \frac{\Delta t}{\Delta x_i}\underline{\mathbf{S}}^*_i, \tag{3.7}$$

in which

$$\tilde{\mathbf{\Lambda}}^\pm = \frac{1}{2}(\tilde{\mathbf{\Lambda}} \pm |\tilde{\mathbf{\Lambda}}|). \tag{3.8}$$

This splits the update into contributions related to right-going (+) and left-going (−) characteristics in the decomposition. It follows that the solution is updated using only contributions from the wave perturbations of the Riemann problems at the nodes which enter the cell under consideration, as illustrated in Fig. 3. It remains to choose an appropriate form for the numerical source term integral $\underline{\mathbf{S}}^*$.

3.1. *Source Terms*

This work follows much recent research into source term discretisation, see for example [1, 5–7], which has concentrated on the use of a characteristic decomposition of the type shown in (3.5). This similarly projects the source term integral onto the eigenvectors of the

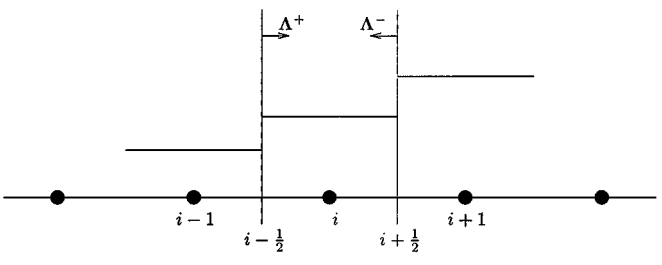

FIG. 3. Wave propagation directions in a control volume.

flux Jacobian $\tilde{\mathbf{A}}$, so that in its linearised form it can be expressed as

$$\int_{x_i}^{x_{i+1}} \underline{S}\,\mathrm{d}x \approx \tilde{\underline{\mathbf{S}}}_{i+1/2} = (\tilde{\mathbf{R}}\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}})_{i+1/2} = \left(\sum_{k=1}^{N_w} \tilde{\beta}_k \tilde{\underline{r}}_k\right)_{i+1/2}, \tag{3.9}$$

where $\tilde{\beta}_k$, the coefficients of the decomposition, are the components of the vector $\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}}$. Note that the integral approximated in (3.9) is over a dual cell of the mesh (associated with the interface $i + 1/2$) and can be easily incorporated within the fluctuation-signal form of the finite volume scheme given by (3.7). $\underline{\mathbf{S}}_i^*$ will be constructed out of contributions from both ends of the cell, with consistency assured as long as the whole of each dual cell integral (3.9) is distributed. Note that the only extra expense incurred by the decomposition is the evaluation of the coefficients $\tilde{\beta}_k$ in (3.9); all other quantities are already evaluated for discretising the flux terms.

The terms on the right hand side of (3.9) may be called upon to balance components of the flux difference $\Delta\,\underline{F}$ (3.5) so they must be linearised in the same way to ensure that, for the chosen equilibrium state,

$$\underline{F}_x - \underline{S} \equiv \underline{0} \Rightarrow \Delta\,\underline{F}_{i+1/2} - \tilde{\underline{\mathbf{S}}}_{i+1/2} = \underline{0}, \tag{3.10}$$

throughout the domain. This follows because at this equilibrium the decompositions (3.5) and (3.9) have been constructed to give $\tilde{\mathbf{\Lambda}}\tilde{\mathbf{R}}^{-1}\Delta\,\underline{U} = \tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}}$ (or alternatively $\tilde{\alpha}_k\tilde{\lambda}_k = \tilde{\beta}_k$). Hence $\sim$ still represents the evaluation of a quantity at the Roe-average state. In fact it is important that the numerical fluxes and sources are evaluated at the same state in order to maintain the balance which the discretisation has been constructed to satisfy.

As a result of the characteristic decomposition (3.9), the source terms may be discretised in an "upwind" manner (although, since none of the components has an inherent upwind direction, this must be taken from the corresponding flux component) and the analysis of the scalar equation presented in Section 2 can now be repeated for this system. This leads straightforwardly to an appropriate upwind fluctuation-signal formulation for the first-order scheme (3.7) with source terms, given by

$$\underline{U}_i^{n+1} = \underline{U}_i^n - \frac{\Delta t}{\Delta x_i}\left[(\tilde{\mathbf{R}}(\tilde{\mathbf{\Lambda}}^-\tilde{\mathbf{R}}^{-1}\Delta\,\underline{U} - \mathbf{I}^-\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}}))_{i+1/2} + (\tilde{\mathbf{R}}(\tilde{\mathbf{\Lambda}}^+\tilde{\mathbf{R}}^{-1}\Delta\,\underline{U} - \mathbf{I}^+\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}}))_{i-1/2}\right], \tag{3.11}$$

in which $\mathbf{I}^\pm = \tilde{\mathbf{\Lambda}}^{-1}\tilde{\mathbf{\Lambda}}^\pm$. The correct balance follows immediately from (3.10).

It is not immediately clear though, how the discretisation of the source term implied by (3.11) can be converted into a numerical source integral $\underline{\mathbf{S}}_i^*$ so that the same balance can be achieved within the flux-based form of the scheme (3.2), particularly when it is extended to a higher order.

The difficulties which arise (and the solution to the problem) can be highlighted by following the transformation of (3.11) into an equation corresponding to (3.2). With a small amount of algebraic manipulation, similar to that carried out in Section 2, (3.11) becomes

$$\begin{aligned}\underline{U}_i^{n+1} = \underline{U}_i^n &- \frac{\Delta t}{2\Delta x_i}\left[(\tilde{\mathbf{R}}(\tilde{\mathbf{\Lambda}}\tilde{\mathbf{R}}^{-1}\Delta\,\underline{U} - \tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}}))_{i+1/2} + (\tilde{\mathbf{R}}(\tilde{\mathbf{\Lambda}}\tilde{\mathbf{R}}^{-1}\Delta\,\underline{U} - \tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}}))_{i-1/2}\right] \\ &+ \frac{\Delta t}{2\Delta x_i}\left[(\tilde{\mathbf{R}}(|\tilde{\mathbf{\Lambda}}|\tilde{\mathbf{R}}^{-1}\Delta\,\underline{U} - \mathrm{sgn}(\mathbf{I})\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}}))_{i+1/2}\right. \\ &\left. - (\tilde{\mathbf{R}}(|\tilde{\mathbf{\Lambda}}|\tilde{\mathbf{R}}^{-1}\Delta\,\underline{U} - \mathrm{sgn}(\mathbf{I})\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}}))_{i-1/2}\right],\end{aligned} \tag{3.12}$$

in which $\mathrm{sgn}(\mathbf{I}) = \tilde{\mathbf{\Lambda}}^{-1}|\tilde{\mathbf{\Lambda}}|$. Since (3.5) and (3.9) hold, and

$$\Delta \underline{F}_{i+1/2} + \Delta \underline{F}_{i-1/2} = (\underline{F}_{i+1} + \underline{F}_i) - (\underline{F}_i + \underline{F}_{i-1}), \tag{3.13}$$

it follows that the scheme (3.12) can be simplified to

$$\underline{U}_i^{n+1} = \underline{U}_i^n - \frac{\Delta t}{\Delta x_i}\left(\underline{F}^*_{i+1/2} - \underline{F}^*_{i-1/2}\right) + \frac{\Delta t}{\Delta x_i}\left(\tilde{\underline{\mathbf{S}}}^-_{i+1/2} + \tilde{\underline{\mathbf{S}}}^+_{i-1/2}\right). \tag{3.14}$$

From the above expressions and (3.5), the numerical fluxes, $\underline{F}^*_{i+1/2}$ and $\underline{F}^*_{i-1/2}$, are precisely those defined by (3.6) and the numerical source term integral of (3.2) is given by

$$\underline{\mathbf{S}}^*_i = \tilde{\underline{\mathbf{S}}}^-_{i+1/2} + \tilde{\underline{\mathbf{S}}}^+_{i-1/2}, \tag{3.15}$$

where

$$\tilde{\underline{\mathbf{S}}}^-_{i+1/2} = \frac{1}{2}(\tilde{\mathbf{R}}(\mathbf{I} - \mathrm{sgn}(\mathbf{I}))\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}})_{i+1/2} = (\tilde{\mathbf{R}}\mathbf{I}^-\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}})_{i+1/2} \tag{3.16}$$

and

$$\tilde{\underline{\mathbf{S}}}^+_{i-1/2} = \frac{1}{2}(\tilde{\mathbf{R}}(\mathbf{I} + \mathrm{sgn}(\mathbf{I}))\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}})_{i-1/2} = (\tilde{\mathbf{R}}\mathbf{I}^+\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}})_{i-1/2}. \tag{3.17}$$

Thus, following the procedure outlined in Section 2, a source term discretisation has been found which, by construction, will ensure that the appropriate equilibria of the underlying mathematical model are maintained by the numerical scheme. The method is closely related to that proposed by LeVeque [14]. However, he constructs new states for the conservative variables $\underline{U}$ within each cell whose differences balance the source term internally before the Riemann solver and any higher order corrections are applied. Here, the source term is decomposed and explicit expressions are constructed for its evaluation in terms of the flow variables. In this manner, any underlying balance in the equations being modelled is automatically passed on to the discrete form. It is even more closely related to the work of Jenny and Muller [12] who also propose the consideration of source terms as edge-based quantities which should be distributed to the cells. However, they considered a characteristic-based scheme whereas this work applies to flux-based finite volume schemes and addresses a specific problem of maintaining a precise balance between flux and source terms.

Note that because the numerical source integral cannot, in general, be written as a difference, nothing similar to (3.13) can be applied to it to allow it to be included within the numerical flux (3.6). This means that the balance which is sought between flux derivatives and sources in the flux-based scheme can only be obtained locally by balancing non-zero fluxes through the edges of a control volume, and not by setting each edge flux to zero. One important consequence of this is that the most sensible method of applying the boundary conditions to the numerical scheme is through the addition of ghost cells, since this requires no further correction to maintain the balance which is sought. The distribution of the numerical fluxes and source term components is shown in Fig. 4.

It is of course possible to overcome the above problem when the source term takes the form of a derivative. If this is the case the source simply augments the conservative flux in the scheme (3.2); i.e., given $\underline{S} = \underline{G}_x$,

$$\underline{F}^* \to \underline{F}^* - \underline{G}^*, \tag{3.18}$$

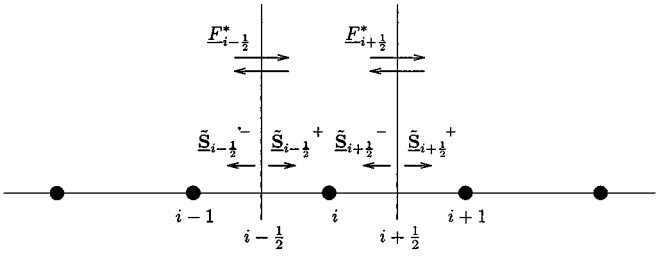

FIG. 4. System flux and source distribution within a control volume.

and $\underline{\mathbf{S}}^*$ becomes obsolete. In some cases it may also be possible to incorporate some part of the source term that can be expressed as a derivative within the numerical flux, and then apply an appropriate discretisation to the remaining component of the source.

3.1.1. Flux-limited schemes. The approach presented in the previous section is no different from the standard upwind technique for approximating source terms when a first-order upwind flux discretisation is being used [6]. The only new aspect is the way it has been written, splitting the dual cell source integral into two parts. Usually though, accuracy of higher than first order is required for practical calculations.

The accuracy of Roe's scheme is improved, without introducing spurious oscillations into the solution, by the application of flux-limiting techniques [13, 20]. These ensure second-order accuracy in smooth regions of the flow, while enforcing a total variation diminishing (TVD) property. It is achieved by including a high-order correction term in the numerical flux, which becomes [18]

$$\underline{F}^*_{i+1/2} = \frac{1}{2}(\underline{F}_{i+1} + \underline{F}_i) - \frac{1}{2}(\tilde{\mathbf{R}}|\tilde{\mathbf{\Lambda}}|\mathbf{L}\tilde{\mathbf{R}}^{-1}\Delta\underline{U})_{i+1/2}, \tag{3.19}$$

in which $\mathbf{L} = \mathrm{diag}(1 - L(r_k)(1 - |\nu_k|))$, where $\nu_k = \tilde{\lambda}_k \Delta t/\Delta x$ is the Courant number associated with the kth component of the decomposition, L is a nonlinear flux-limiter function, as described in [13, 20], and

$$r_k = \frac{\tilde{\alpha}_k^{\text{upwind}}}{\tilde{\alpha}_k^{\text{local}}}. \tag{3.20}$$

Within (3.19), $\tilde{\alpha}_k^{\text{local}}$ would be evaluated at the $i + 1/2$ edge, while $\tilde{\alpha}_k^{\text{upwind}}$ is found at the upwind edge according to the velocity $\tilde{\lambda}_k$ at edge $i + 1/2$.

It is clear that if a balance is to be maintained a corresponding high-order correction must also be made to the source term approximation, and its form can be derived simply by comparing the numerical sources of (3.16, 3.17) with the numerical fluxes in (3.6), all of which have been split into two parts which are balanced separately. The flux limiter is only applied to the second part of the numerical flux, so the flux-limited numerical source which maintains the balance achieved by the first-order discretisation takes the form

$$\tilde{\underline{\mathbf{S}}}^-_{i+1/2} = \frac{1}{2}(\tilde{\mathbf{R}}(\mathbf{I} - \mathrm{sgn}(\mathbf{I})\mathbf{L})\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}})_{i+1/2}, \tag{3.21}$$

with a similar expression for $\tilde{\underline{\mathbf{S}}}^+_{i-1/2}$ in (3.15). Note that this work is concerned more with maintaining equilibria between the flux and source terms than with constructing higher order

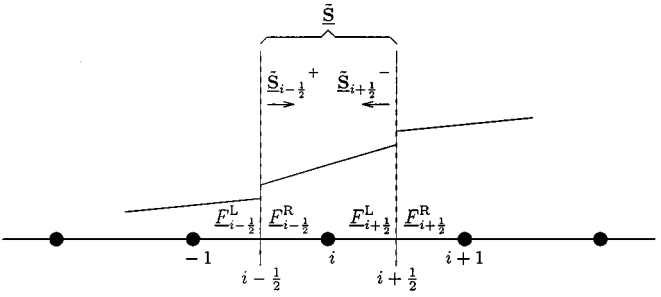

FIG. 5. Flux and source evaluation for the MUSCL scheme.

approximations of the source terms on their own so that the *overall* scheme gives better numerical solutions. Also, since (3.21) is an edge-based quantity, it is simple to evaluate with the fluxes and include within the numerical model with only a small increase in overall expense.

3.1.2. Slope-limited schemes. The same balance is slightly more difficult to achieve when the high-resolution scheme is constructed using a MUSCL-type slope-limiting approach [22]. This is because the underlying representation of the solution is now taken to be linear within each cell so that (3.13) is no longer true. It can, though, be replaced by the more general expression

$$\Delta \underline{F}_{i+1/2} + \Delta \underline{F}_{i-1/2} = \left(\underline{F}^{R}_{i+1/2} + \underline{F}^{L}_{i+1/2}\right) - \left(\underline{F}^{R}_{i-1/2} + \underline{F}^{L}_{i-1/2}\right) - 2\left(\underline{F}^{L}_{i+1/2} - \underline{F}^{R}_{i-1/2}\right), \tag{3.22}$$

where the superscripts R and L represent evaluation on, respectively, the right and left hand sides of the interface indicated by the associated subscript (as shown in Fig. 5). The corresponding numerical flux is

$$\underline{F}^{*}_{i+1/2} = \frac{1}{2}\left(\underline{F}^{R}_{i+1/2} + \underline{F}^{L}_{i+1/2}\right) - \frac{1}{2}(\tilde{\mathbf{R}}|\tilde{\mathbf{\Lambda}}|\tilde{\mathbf{R}}^{-1}\Delta \underline{U})_{i+1/2}, \tag{3.23}$$

in which the Roe averages are now evaluated from the reconstructed piecewise linear solution. An appropriate correction must therefore be made to the numerical source within each cell, and this leads to

$$\underline{\mathbf{S}}^{*}_{i} = \left(\tilde{\underline{\mathbf{S}}}^{-}_{i+1/2} + \tilde{\underline{\mathbf{S}}}^{+}_{i-1/2}\right) - \tilde{\underline{\mathbf{S}}}\left(\underline{U}^{L}_{i+1/2}, \underline{U}^{R}_{i-1/2}\right). \tag{3.24}$$

The first term on the right hand side is evaluated precisely as before, in (3.15), except that the interface values are now those of the MUSCL reconstruction of the solution within each cell. The final term $\tilde{\underline{\mathbf{S}}}$ is simply the source term integral approximated over the mesh cell (cf. (3.9)); it is indicated by $\sim$ because it is evaluated at the Roe average of the left and right states of the linear reconstruction of the solution within the cell. In terms of the approximation (3.10) the extra term can be thought of as a correction to the integral of the source term over the dual cell arising from the linear variation of the approximation.

3.2. *Spatially Dependent Fluxes*

In some situations the flux may depend on quantities other than the flow variables, and the numerical scheme needs to be modified appropriately. Only one extra spatially

varying quantity will be considered here, but the approach is easily extended to any number. Returning to Eq. (3.1) and taking $\underline{F} = \underline{F}(\underline{U}, b(x))$, where b varies independent of $\underline{U}$, requires a modification to the characteristic decomposition, so that (3.5) becomes

$$\begin{aligned}\Delta \underline{F}_{i+1/2} &= (\tilde{\mathbf{A}}\,\Delta \underline{U} + \tilde{\underline{V}})_{i+1/2} \\ &= (\tilde{\mathbf{R}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{R}}^{-1}\Delta \underline{U} + \tilde{\mathbf{R}}\tilde{\mathbf{R}}^{-1}\tilde{\underline{V}})_{i+1/2} \\ &= \left(\sum_{k=1}^{N_w} \tilde{\alpha}_k \tilde{\lambda}_k \tilde{\underline{r}}_k + \sum_{k=1}^{N_w} \tilde{\gamma}_k \tilde{\underline{r}}_k\right)_{i+1/2}, \end{aligned} \tag{3.25}$$

where $\tilde{\underline{V}} \approx \frac{\partial \underline{F}}{\partial b}\Delta b$ and $\tilde{\gamma}_k$, the coefficients of the decomposition of this extra term, are the components of $\tilde{\mathbf{R}}^{-1}\tilde{\underline{V}}$. Equation (3.25) gives a set of N_w equations in $N_w + 1$ unknowns, which are taken to be a set of consistent Roe-averaged independent variables from which $\tilde{\underline{U}}$ and $\tilde{b}$ (and all other variables) can be evaluated. This leaves one degree of freedom which can be used by enforcing $\Delta \underline{F} - \tilde{\underline{S}} = \underline{0}$ at an appropriate state of equilibrium.

Following the same steps as earlier in Section 3 to transform the fluctuation-signal scheme to the flux-based scheme, but including this extra term in the flux difference, leads to precisely the same form for the scheme as when approximating the homogeneous system as shown in (3.2), but with new expressions for the numerical fluxes, given by

$$\underline{F}^*_{i+1/2} = \frac{1}{2}(\underline{F}_{i+1} + \underline{F}_i) - \frac{1}{2}(\tilde{\mathbf{R}}|\tilde{\mathbf{\Lambda}}|\tilde{\mathbf{R}}^{-1}\,\Delta \underline{U} + \tilde{\mathbf{R}}\,\mathrm{sgn}(\mathbf{I})\tilde{\mathbf{R}}^{-1}\tilde{\underline{V}})_{i+1/2}, \tag{3.26}$$

in the first-order case,

$$\underline{F}^*_{i+1/2} = \frac{1}{2}(\underline{F}_{i+1} + \underline{F}_i) - \frac{1}{2}(\tilde{\mathbf{R}}|\tilde{\mathbf{\Lambda}}|\mathbf{L}\tilde{\mathbf{R}}^{-1}\Delta \underline{U} + \tilde{\mathbf{R}}\,\mathrm{sgn}(\mathbf{I})\mathbf{L}\tilde{\mathbf{R}}^{-1}\tilde{\underline{V}})_{i+1/2}, \tag{3.27}$$

when the flux-limited high-resolution scheme is being used, or

$$\underline{F}^*_{i+1/2} = \frac{1}{2}\left(\underline{F}^{\mathrm{R}}_{i+1/2} + \underline{F}^{\mathrm{L}}_{i+1/2}\right) - \frac{1}{2}(\tilde{\mathbf{R}}|\tilde{\mathbf{\Lambda}}|\tilde{\mathbf{R}}^{-1}\Delta \underline{U} + \tilde{\mathbf{R}}\,\mathrm{sgn}(\mathbf{I})\tilde{\mathbf{R}}^{-1}\tilde{\underline{V}})_{i+1/2}, \tag{3.28}$$

for the MUSCL scheme, where the averages are now calculated from the linearly reconstructed solution. By including this extra term in the numerical flux it is possible to avoid altering the form of the source term, as was suggested in [15].

3.3. *Shallow Water Flows*

The shallow water equations have been chosen as the system of equations to illustrate the use of these new techniques. In one dimension, shallow water flows through a rectangular open channel of varying breadth and bed slope are modelled. The effects of bed friction may also be included and, as described in Section 3.1, are simple to treat within the new framework without disturbing the balance between the other source terms and the flux derivatives. However, since it is this balance which the new discretisation has been constructed to maintain, friction is not included in the following discussion. The remaining system can be modelled by the equations [3]

$$\underline{U}_t + \underline{F}_x = \underline{S}, \tag{3.29}$$

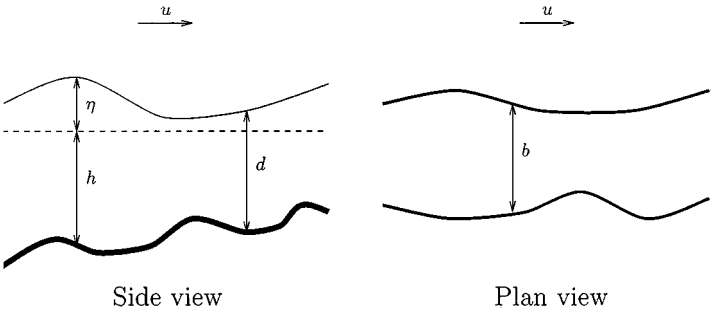

FIG. 6. The shallow water flow variables.

where

$$\underline{U} = \begin{pmatrix} bd \\ bdu \end{pmatrix}, \quad \underline{F} = \begin{pmatrix} bdu \\ bdu^2 + \frac{1}{2}gbd^2 \end{pmatrix}, \quad \underline{S} = \begin{pmatrix} 0 \\ \frac{1}{2}gd^2 b_x + gbdh_x \end{pmatrix}, \tag{3.30}$$

which leads to

$$\frac{\partial \underline{F}}{\partial \underline{U}} = \begin{pmatrix} 0 & 1 \\ gd - u^2 & 2u \end{pmatrix}, \quad \frac{\partial \underline{F}}{\partial b} = \begin{pmatrix} 0 \\ -\frac{1}{2}gd^2 \end{pmatrix}. \tag{3.31}$$

In these equations d is the depth of the flow, h is the depth of the bed below a nominal still water level, $b = b(x)$ is the channel breadth, u is the flow velocity, and g is the acceleration due to gravity. These quantities are depicted in Fig. 6, along with η, the height of the free surface above still water.

Equation (3.30) provides an example which includes source terms and a spatial dependence on channel breadth which is independent of the flow. Furthermore, the balance which has been sought in previous sections is illustrated by the steady state represented by still water ($d \equiv h$ and $u \equiv 0$), in which case the system (3.30) reduces to

$$\left(\frac{1}{2}gbd^2\right)_x = \frac{1}{2}gd^2 b_x + gbdh_x. \tag{3.32}$$

This exhibits precisely the type of balance which has been focused on in this work.

The characteristic decomposition (3.25) for the one-dimensional shallow water equations (3.30) and (3.31) is completely defined by

$$\begin{aligned} \tilde{\alpha}_1 &= \frac{\Delta(bd)}{2} + \frac{1}{2\tilde{c}}(\Delta(bdu) - \tilde{u}\Delta(bd)), & \tilde{\alpha}_2 &= \frac{\Delta(bd)}{2} - \frac{1}{2\tilde{c}}(\Delta(bdu) - \tilde{u}\,\Delta(bd)) \\ \tilde{\lambda}_1 &= \tilde{u} + \tilde{c}, & \tilde{\lambda}_2 &= \tilde{u} - \tilde{c} \\ \underline{\tilde{r}}_1 &= \begin{pmatrix} 1 \\ \tilde{u} + \tilde{c} \end{pmatrix}, & \underline{\tilde{r}}_2 &= \begin{pmatrix} 1 \\ \tilde{u} - \tilde{c} \end{pmatrix} \\ \tilde{\gamma}_1 &= -\frac{1}{4g}\tilde{c}^3 \Delta b, & \tilde{\gamma}_2 &= \frac{1}{4g}\tilde{c}^3 \Delta b, \end{aligned} \tag{3.33}$$

and it is easily shown that (3.25) is satisfied exactly when

$$\tilde{u} = \frac{\sqrt{b^{\mathrm{R}}d^{\mathrm{R}}}u^{\mathrm{R}} + \sqrt{b^{\mathrm{L}}d^{\mathrm{L}}}u^{\mathrm{L}}}{\sqrt{b^{\mathrm{R}}d^{\mathrm{R}}} + \sqrt{b^{\mathrm{L}}d^{\mathrm{L}}}}, \quad \tilde{c}^2 = g\left(\frac{\sqrt{b^{\mathrm{R}}}d^{\mathrm{R}} + \sqrt{b^{\mathrm{L}}}d^{\mathrm{L}}}{\sqrt{b^{\mathrm{R}}} + \sqrt{b^{\mathrm{L}}}}\right), \tag{3.34}$$

which reduce to the Roe averages for the one-dimensional shallow water flow described in [6] in the absence of breadth variation (i.e., when $b^{\mathrm{R}} = b^{\mathrm{L}}$). The corresponding decomposition of the source terms (3.9) then leads to

$$\tilde{\beta}_1 = \frac{1}{4g}\tilde{c}^3\Delta b + \frac{1}{2}\tilde{b}\tilde{c}\Delta h = -\tilde{\beta}_2 \tag{3.35}$$

In order for (3.16) and (3.17) to maintain the correct balance, i.e.,

$$\tilde{\alpha}_k\tilde{\lambda}_k + \tilde{\gamma}_k - \tilde{\beta}_k = 0 \qquad \forall k, \tag{3.36}$$

or equivalently,

$$\tilde{\mathbf{R}}(\tilde{\mathbf{\Lambda}}\tilde{\mathbf{R}}^{-1}\,\Delta\,\underline{U} + \tilde{\mathbf{R}}^{-1}\underline{\tilde{Y}} - \tilde{\mathbf{R}}^{-1}\underline{\tilde{\mathbf{S}}}) = \underline{0} \tag{3.37}$$

when the flow is quiescent, $\tilde{b}$ is constructed so that it satisfies

$$\tilde{b}\Delta h = \Delta(bh) - \tilde{h}\Delta b, \tag{3.38}$$

where $\tilde{h}$ is evaluated in a manner similar to that in which $\tilde{d}$ is evaluated

$$\tilde{h} = \frac{\sqrt{b^{\mathrm{R}}}h^{\mathrm{R}} + \sqrt{b^{\mathrm{L}}}h^{\mathrm{L}}}{\sqrt{b^{\mathrm{R}}} + \sqrt{b^{\mathrm{L}}}}, \tag{3.39}$$

so that $d \equiv h \Rightarrow \tilde{d} = \tilde{h}$ throughout the domain. Note that this also requires that d and h are reconstructed in the same manner if the MUSCL high-resolution scheme is used.

3.3.1. Numerical results. The results presented in this section have been chosen to illustrate the improvement in the approximation using the new techniques by focusing on the following:

- The ability to maintain quiescent flow,
- The accuracy of approximations to both continuous and discontinuous steady state solutions, and
- The accuracy of simple time-dependent approximations.

These have been studied using a variety of channel geometries. The following one-dimensional results are presented on regular grids but all of the tests have been run on randomly perturbed grids with very similar outcomes.

The geometry for the first test case was proposed by the Working Group on Dam-Break Modelling [4], and the bed and breadth variation of the channel (of length 1500) are depicted in Fig. 7. The upwind source term treatment described in this paper is compared with a much simpler pointwise discretisation in Fig. 8 (using a uniform 600-cell grid, so that $\Delta x = 2.5$), which shows graphs of water surface level and unit discharge for the numerical steady states which result from quiescent initial conditions ($\eta = d - h = 12.0$ and $u = 0.0$) and simple non-reflecting boundary conditions. In this case the initial (still water) conditions should be maintained indefinitely by the numerical scheme.

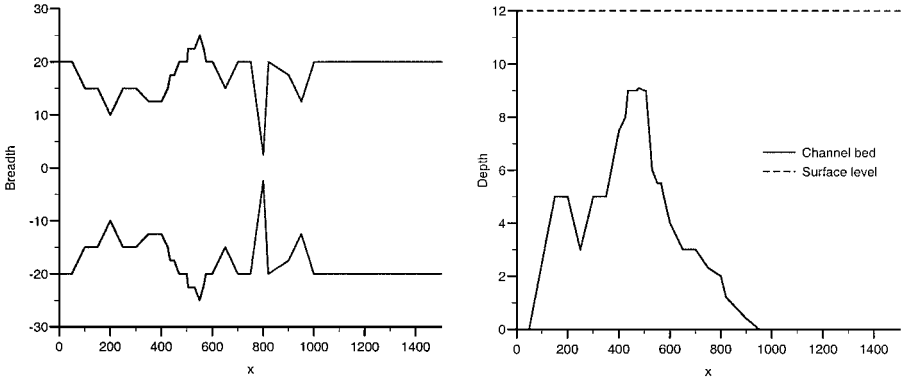

FIG. 7. Breadth and bed variation for the tidal flow test case.

A comparison is made between first-order, slope-limited, and flux-limited schemes combined with pointwise and upwind source term discretisations: in all high-resolution cases the Minmod limiter [20] has been applied. The upwind source term discretisations always produce the correct steady-state solution, exactly to machine accuracy and indistinguishable from the exact solution in the graphs (even on irregular grids). This is true not only for the first-order scheme (which has been achieved previously) but also for the high-resolution TVD schemes using any flux or slope limiter on any grid in the presence of bed slope and breadth variations. The pointwise discretisations show small discrepancies (a central discretisation of the source term was also tried but produced even worse results than the pointwise approximation and is not presented here), most notably in the unit discharge, a quantity which depends on the flow velocity. In each case the method described in Section 3.2 is used to discretise the fluxes where the channel breadth varies.

The second channel geometry used in this work (shown in Fig. 9) is defined over the interval [0.0, 3.0] and has a smoothly varying depth and breadth, given by

$$b(x) = \begin{cases} 1.0 - (1.0 - b_{\min}) \cos^2(\pi(x - 1.5)) & \text{for } |x - 1.5| \le 0.5 \\ 1.0 & \text{otherwise,} \end{cases} \tag{3.40}$$

where $b_{\min}$ is the minimum channel breadth, and

$$h(x) = \begin{cases} 1.0 - z_{\max} \cos^2(\pi(x - 1.5)) & \text{for } |x - 1.5| \le 0.5 \\ 1.0 & \text{otherwise,} \end{cases} \tag{3.41}$$

in which $z_{\max}$ is the maximum height of the bed above the level $\eta = 0.0$. This has been chosen as a simple channel geometry for which exact steady-state solutions to the one-dimensional shallow water equations are available for comparison [8]. The parameters chosen to define the channel here are $z_{\max} = 0.1$ and $b_{\min} = 0.9$. A uniform 150-cell grid has been used for each computation.

Three flows (each defined by a local Froude number F and a depth d for the flow) are compared:

- $F_\infty = 0.5$, $d_\infty = 1.0$, giving purely subcritical flow which is symmetric about the throat of the constriction (the most narrow point, $x = 1.5$),
- $F_\infty = 0.6$, $d_\infty = 1.0$, giving transcritical flow with a stationary hydraulic jump downstream of the throat and a critical point at the throat, and

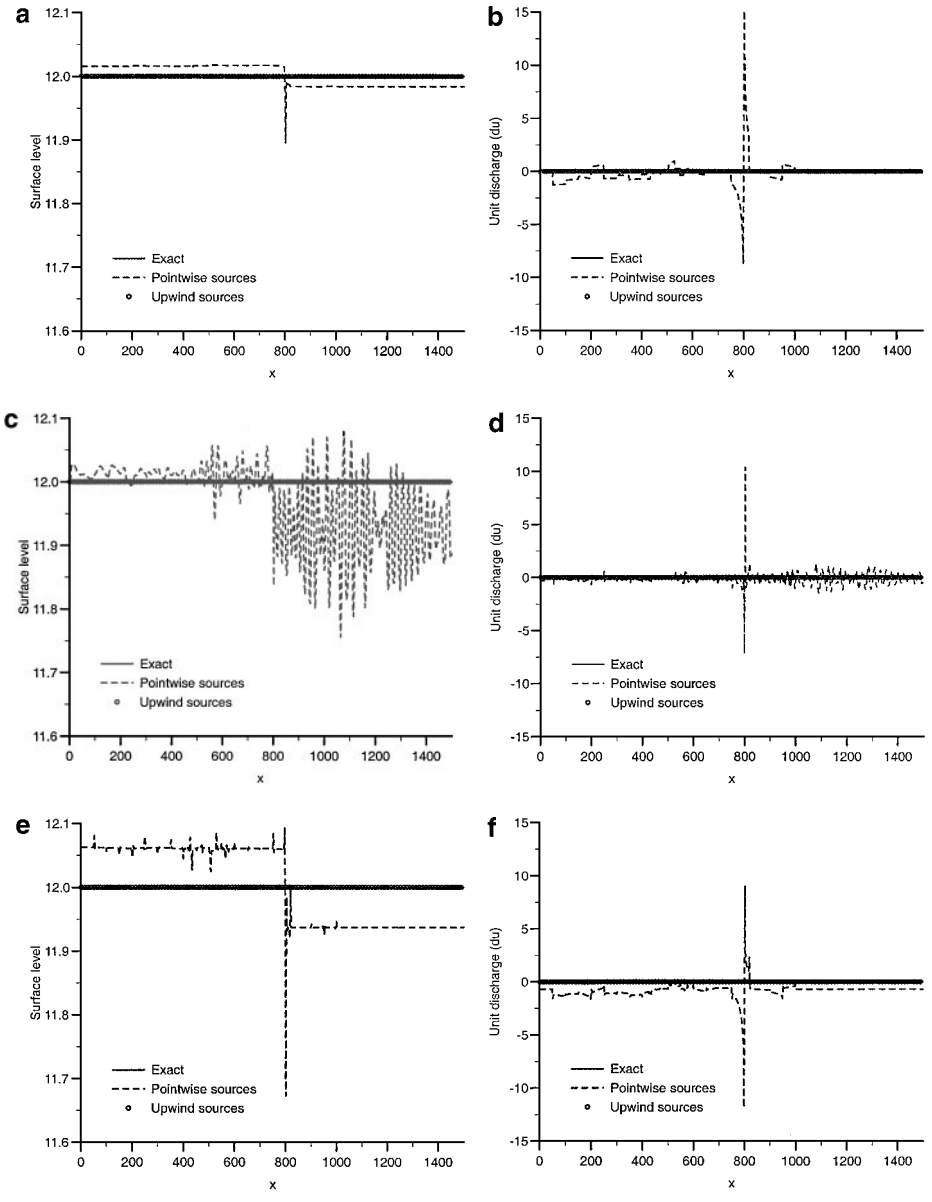

FIG. 8. Water surface level and unit discharge for quiescent flow in a channel with variable bed and breadth, see Fig. 7, for first-order (a, b) and high-resolution slope-limited (c, d) and flux-limited (e, f) schemes ($t = 1000$).

- $F_\infty = 1.7$, $d_\infty = 1.0$, giving purely supercritical flow which is symmetric about the throat.

The subscript ∞ represents the freestream flow values at infinity which are used in the application of simple characteristic boundary conditions at inflow and outflow. The results of the comparisons for each of the schemes are shown in Figs. 10–12. The graphs show the variation of total discharge $Q = bdu$ and depth d through the channel. Discharge is a quantity that should remain constant at steady state and therefore gives an easily visualised indication of the accuracy of the approximation.

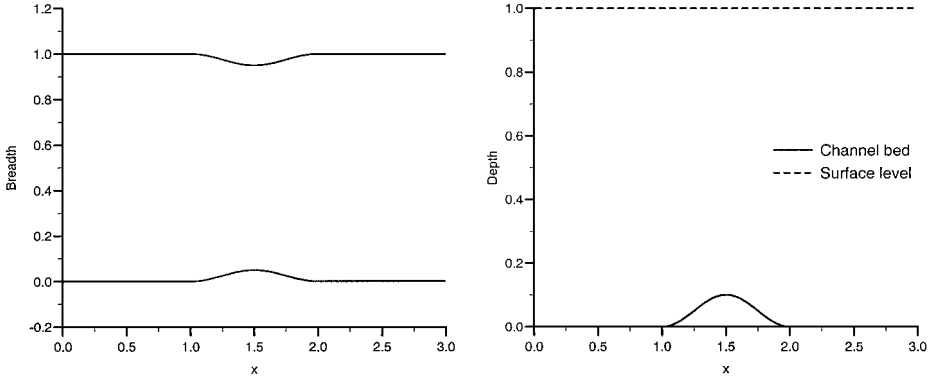

FIG. 9. Breadth and bed variation for the sine channel test case.

In each case the upwinded source terms can be seen to model the solution better than the pointwise evaluation. This is particularly noticeable in the subcritical case where the latter is unable to attain a symmetric solution. The position and strength of the hydraulic jump is predicted accurately by all of the schemes, although there is a small discrepancy in the discharge at the discontinuity in every case. Note that in the second-order case small oscillations appear in the upwinded, slope-limited solution. These are not prohibited by enforcing the TVD condition because this only applies to the homogeneous equations, although they appear in neither the first-order nor the flux-limited results. This indicates that the correction term of (3.24) may require modification away from the still water steady state.

One might expect that using the decomposed source term with a first-order upwind discretisation to balance would produce a good balance when combined with a TVD discretisation of the flux terms. However, Fig. 13 shows that this is clearly not the case and that the TVD correction needs to be applied to the discrete source term as well. Further evidence of the improved modelling of the new flux-limited scheme can be found in [11] where it is compared with a number of other methods.

In the subcritical case, a series of runs has been carried out to indicate the accuracy of the scheme. The l_2 errors in these results are plotted against mesh size for the various source term treatments and shown in Fig. 14. Only the first-order and flux-limited schemes are compared, because the slope-limited results did not converge to a steady state. This seems to be because the reconstruction procedure creates discrepancies between the discrete flux and source terms away from the still water steady state. The upwinding of the source term produces indistinguishable results in both the first, and higher order cases (suggesting that the upwind discretisation is actually second-order accurate at the steady state). The differences are only seen in the time-dependent results when the higher order schemes show their advantages. In all cases the accuracy is dramatically improved by the upwinding (and the order of accuracy is always increased to second).

The speed of convergence to the steady state is indicated in Fig. 15 and it can be seen that the rates of convergence achieved with the pointwise and upwind evaluations of the source term are very similar (the slope-limited results did not converge). This was seen for all of the channel flow test cases run.

In [21] it is shown that for a short channel (of length L, taken here to be 1500) and low-speed flow, given the initial conditions

$$d(x,0) = h(x), \quad q(x,0) = 0, \tag{3.42}$$

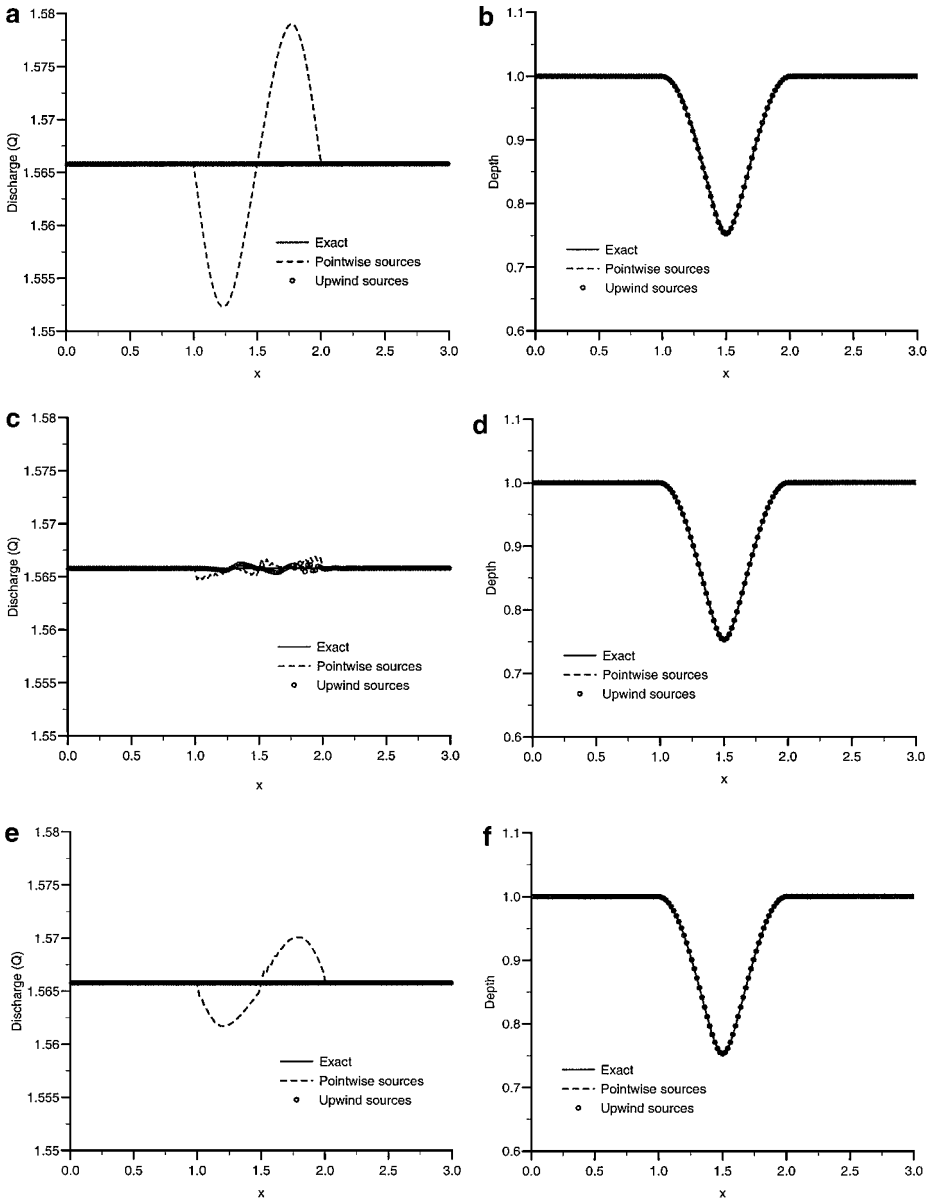

FIG. 10. Discharge and depth for the steady, subcritical, symmetric constricted channel test case for first-order (a, b) and high-resolution slope-limited (c, d) and flux-limited (e, f) schemes.

where $q = du$ is the unit discharge and $h(x)$ is indicated in Fig. 7, and the boundary conditions

$$d(0, t) = h(0) + \phi(t), \quad q(L, t) = \psi(t), \tag{3.43}$$

a first-order approximate solution to Eq. (3.30) can be expressed as

$$\begin{aligned} d(x, t) &= h(x) + \phi(t) \\ q(x, t) &= \psi(t) + \frac{\phi'(t)}{b(x)} \int_x^L b(s)\, \mathrm{d}s. \end{aligned} \tag{3.44}$$

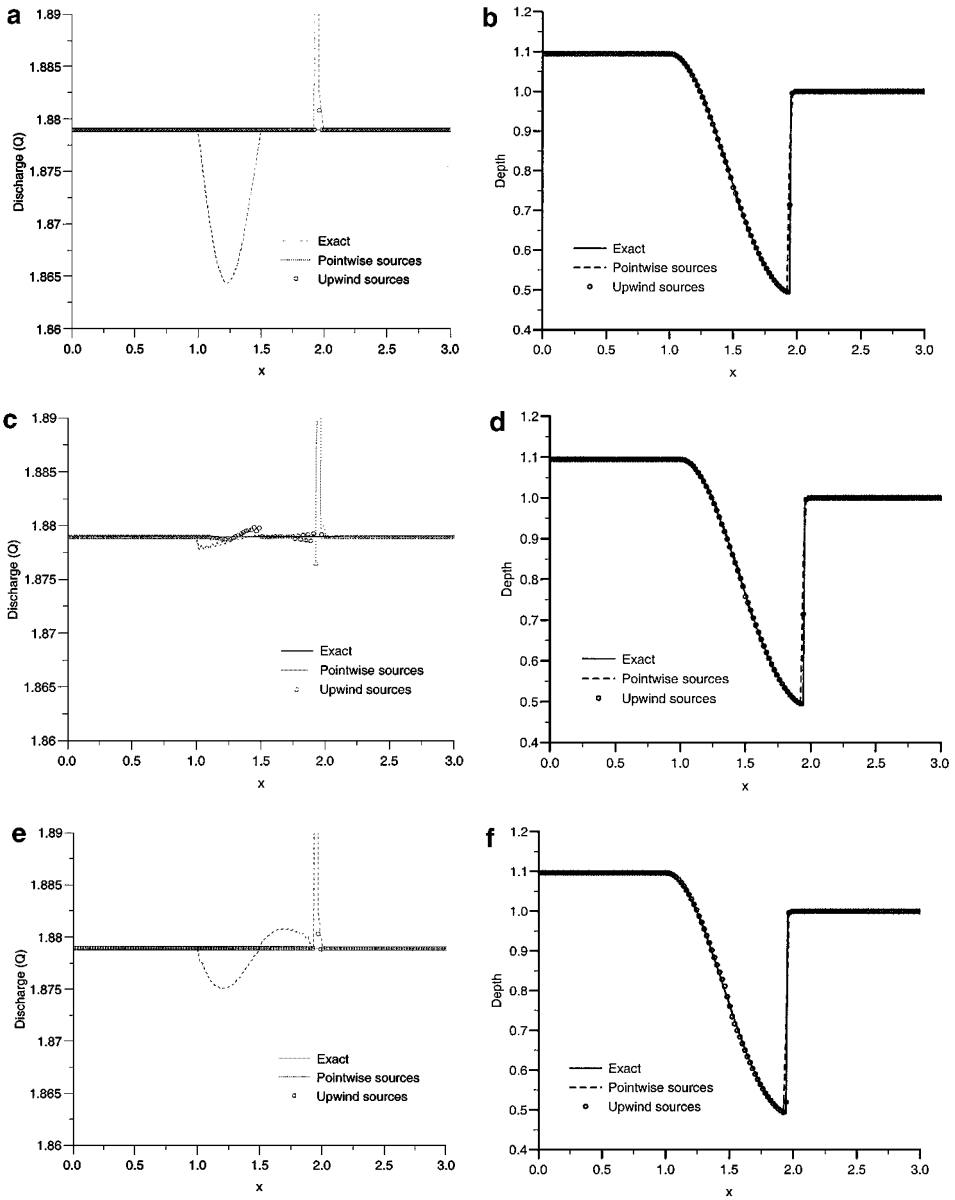

FIG. 11. Discharge and depth for the steady, transcritical, symmetric constricted channel test case for first-order (a, b) and high-resolution slope-limited (c, d) and flux-limited (e, f) schemes.

The quiescent flow case considered earlier corresponds to taking $\phi(t) \equiv \psi(t) \equiv 0$.

A time-dependent tidal flow test case was suggested in [21] for which

$$\phi(t) \;=\; 4 + 4\sin\left(\frac{(t-10{,}800)\pi}{21{,}600}\right) \tag{3.45}$$

and $\psi(t) \equiv 0$, the asymptotically exact solution being given by (3.44). The exact and numerical solutions (all computed on the same regular 600-cell grid) to this problem when $t = 10{,}800$ are compared in Fig. 16. The surface level of the flow should be nearly constant

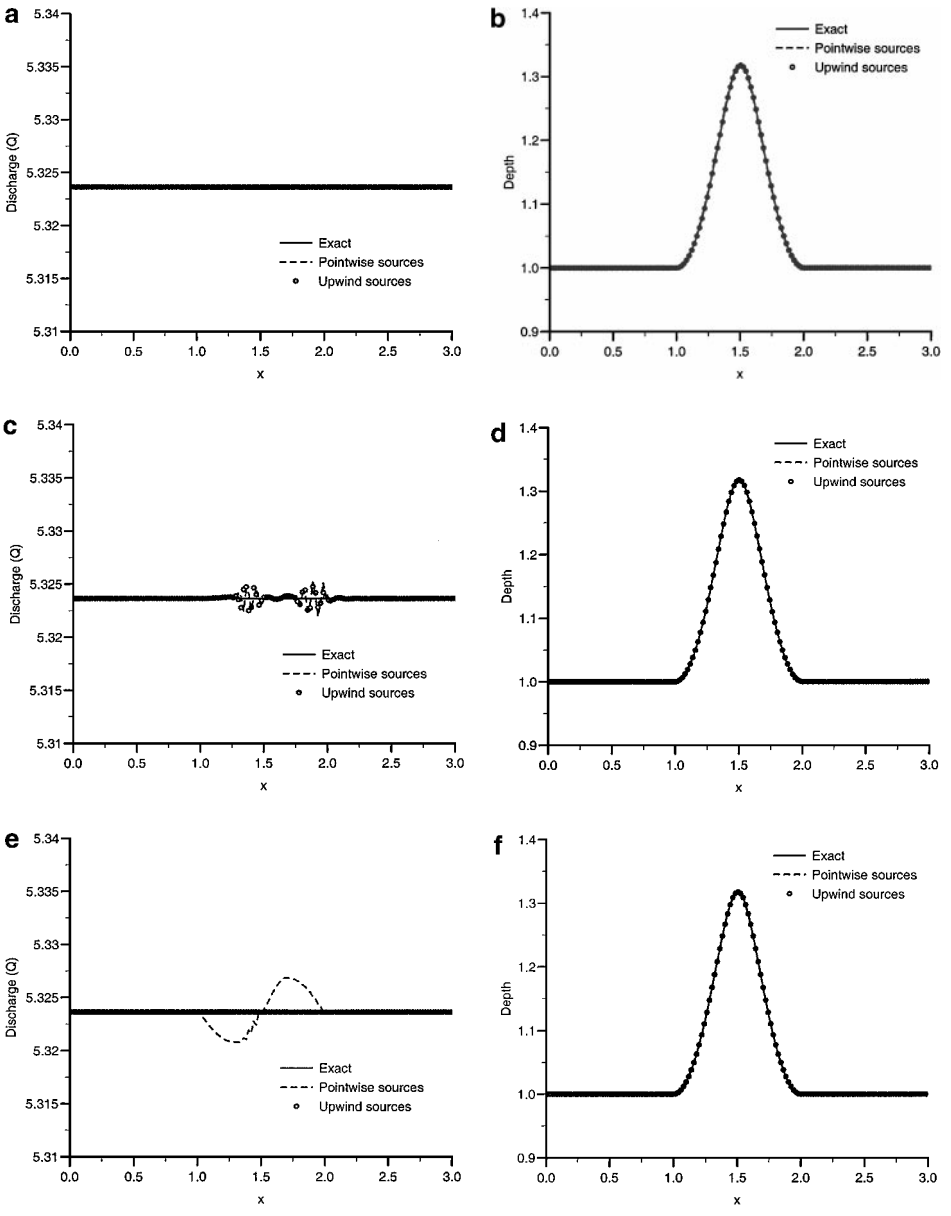

FIG. 12. Discharge and depth for the steady, supercritical, symmetric constricted channel test case for first-order (a, b) and high-resolution slope-limited (c, d) and flux-limited (e, f) schemes.

at any time due to its low velocity, but this is the time when the inflow velocity is at its maximum. This problem is particularly difficult because of the huge running time for the calculation (involving >100,000 time-steps) so even the smallest oscillations introduced numerically are amplified and distort the final solution.

The agreement is very close for the first-order and both of the higher order schemes when the upwind source discretisation is used. The only problem is seen when the slope-limited scheme is used, since it requires a much stricter practical bound on the CFL number for the solution to remain free of unwanted oscillations (a value of 0.1 was used compared

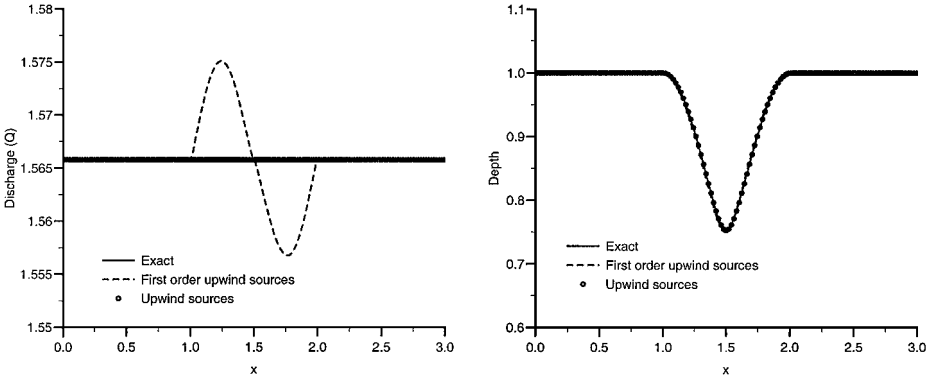

FIG. 13. Discharge and depth for the steady, subcritical, symmetric constricted channel test case for high-resolution flux-limited scheme with first-order upwind source terms.

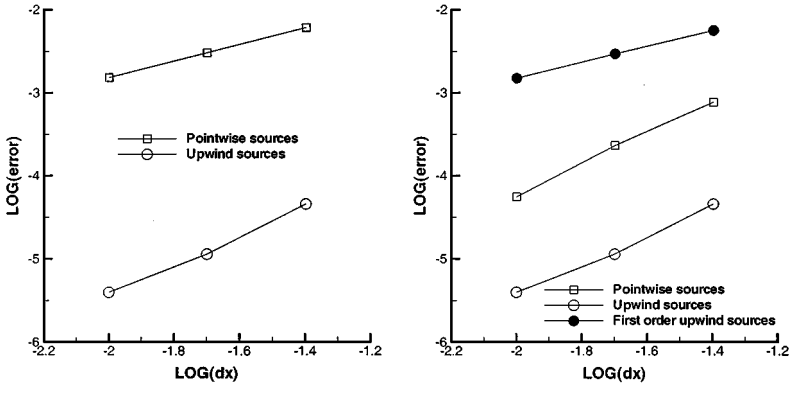

FIG. 14. Graphs of l_2 error against mesh size for the steady, subcritical, symmetric channel test case for the first-order (left) and high-resolution TVD (right) schemes.

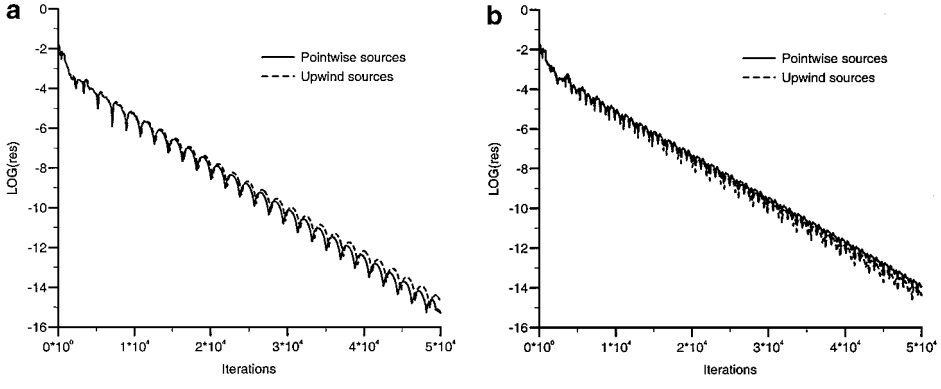

FIG. 15. Convergence histories for the steady, subcritical, symmetric channel test case for the first-order (a) and high-resolution TVD (b) schemes.

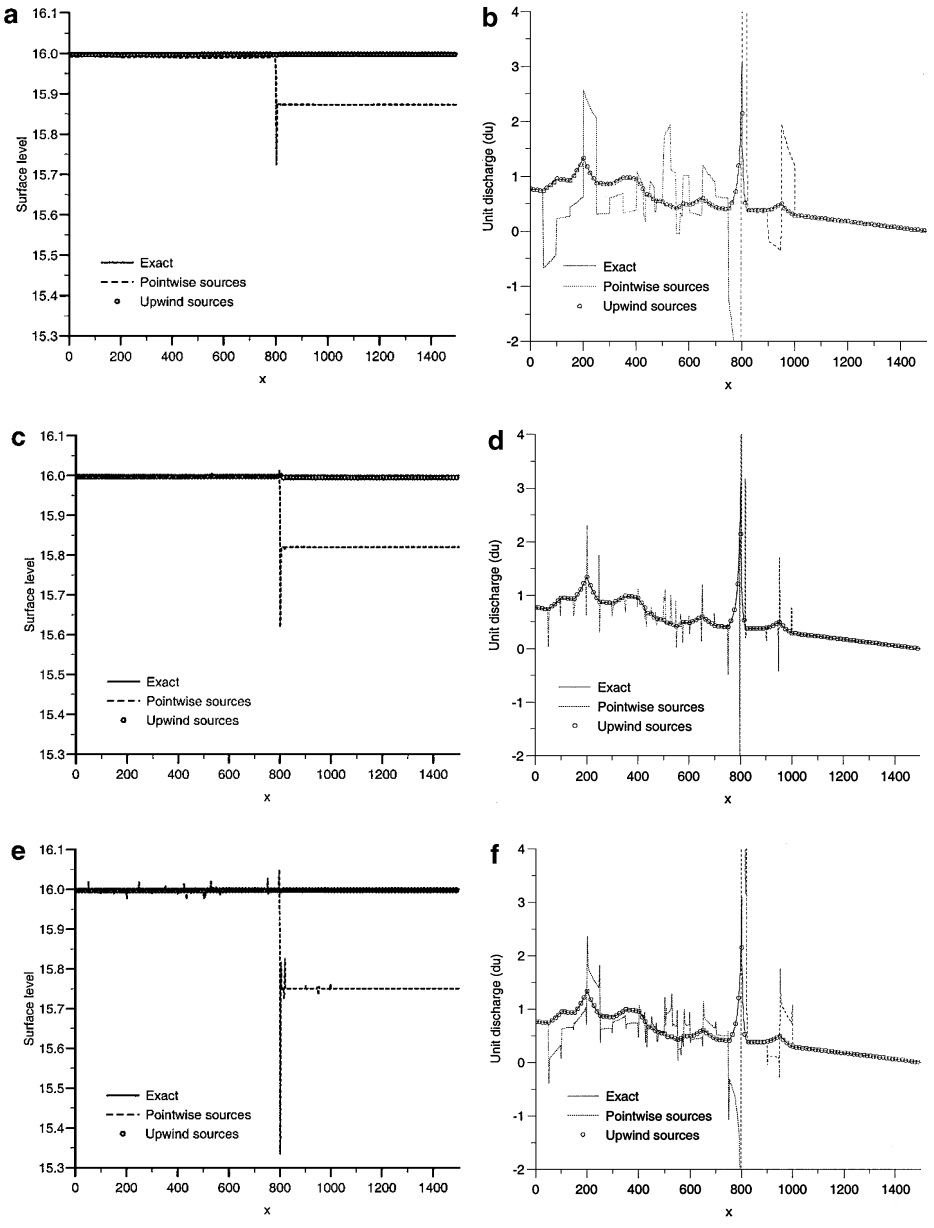

FIG. 16. Water surface level and unit discharge for the tidal flow test case for first-order (a, b) and high-resolution slope-limited (c, d) and flux-limited (e, f) schemes.

with 0.8 for the other schemes, a severe practical restriction). At a CFL number of 0.8 the slope-limited scheme exhibits similar accuracy in this case, whether combined with the upwind or the pointwise source discretisation. At shorter experiment times, though, upwinding the source terms proves to be considerably more accurate. Furthermore, the accuracy of the slope-limited scheme depends on the variables to which the slope limiting is applied. Minor discrepancies appear in the balance due to the method by which the solution is reconstructed, and these cause small oscillations which amplify over time. The problem does not seem to be due to any reduction in the stability range of the scheme caused by the

presence of the source terms since the solution remains bounded even for a CFL number of 0.8 (although the sources are likely to have some effect on the scheme's stability, an issue which has been discussed elsewhere; see for example [9]). In the steady-state test cases shown earlier this extra restriction does not apply. This problem is also affected by the fact that the TVD condition is designed to apply in the absence of source terms and that here we are only ensuring that the flux terms, when considered in isolation from the sources, are being discretised in a TVD manner. These problems do not arise in the first-order and flux-limited cases. As in the still water test, even though the pointwise source discretisation gives a reasonable approximation to the depth, it is very poor at predicting the flow velocity.

One final test case, suggested in [14], is used here to show that the method also works well in the modelling of wave propagation problems. The channel geometry is similar to that of the second case above, but it is defined on a channel with no breadth variation and of length 1.0, where the bed topography is defined by

$$h(x) = \begin{cases} 1.0 - 0.25(\cos(\pi(x-0.5)/0.1) + 1.0) & \text{for } |x-0.5| \le 0.1 \\ 1.0 & \text{otherwise.} \end{cases} \tag{3.46}$$

The data are initially stationary ($u \equiv 0$), with a surface profile defined by

$$d(x) = \begin{cases} h(x) + \epsilon & \text{for } 0.1 < x < 0.2 \\ h(x) & \text{otherwise.} \end{cases} \tag{3.47}$$

Two cases have been run: $\epsilon = 0.2$ and $\epsilon = 0.01$. Both sets of initial conditions are shown in Fig. 17. The solutions obtained using upwinded source terms with the flux-limited (CFL = 0.8) and slope-limited (CFL = 0.2) schemes agree closely with each other and with the solution obtained using the flux-limited scheme on a much finer grid. The water surface levels at time $t = 0.7$ are shown in Fig. 18, obtained on a 600 cell uniform grid with simple transmissive boundary conditions, and compared with a 2400 cell solution. Most importantly, even in the case of the smaller disturbance, the noise is negligible and the wave has propagated with no visible distortion.

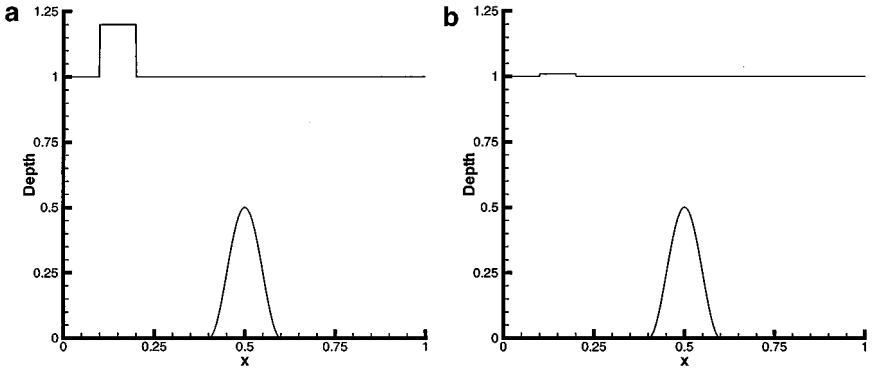

FIG. 17. Initial water surface level and bed variation for the wave propagation test cases: $\epsilon = 0.2$ (a) and $\epsilon = 0.01$ (b).

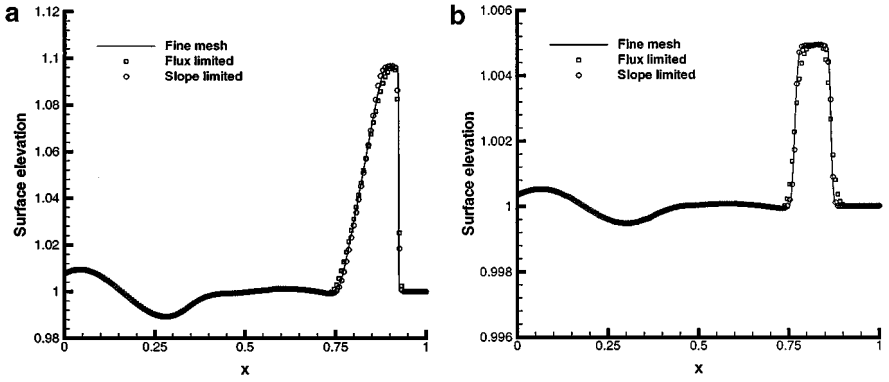

FIG. 18. Water surface levels obtained using the high-resolution flux- and slope-limited schemes, both with upwinded source terms, at $t = 0.7$ for the wave propagation test cases: $\epsilon = 0.2$ (a) and $\epsilon = 0.01$ (b).

4. HIGHER DIMENSIONS

The following analysis is presented for the two-dimensional case but can be applied simply in three dimensions as well. The conservative form of a system of conservation laws with additional source terms is expressed as

$$\underline{U}_t + \underline{F}_x + \underline{G}_y = \underline{S}, \tag{4.1}$$

in which there are now two flux vectors, denoted by $\underline{F} = \underline{F}(\underline{U})$ and $\underline{G} = \underline{G}(\underline{U})$. The case where the fluxes depend on a quantity other than the flow variables is not presented here, having no obvious application to two-dimensional shallow water flows, but can be dealt with in a similar manner to the one-dimensional case presented in Section 3.2.

A combination of a standard finite volume approximation of the flux terms on an arbitrary polygonal mesh (although only triangular and quadrilateral meshes will be considered in the results) and a forward Euler discretisation of the time derivative leads to the conservative difference scheme,

$$\underline{U}_i^{n+1} = \underline{U}_i^n - \frac{\Delta t}{V_i} \sum_{l=1}^{N_e} \delta_{il}(\underline{F}_{il}^*, \underline{G}_{il}^*) \cdot \hat{\hat{n}}_{il} + \frac{\Delta t}{V_i} \underline{\mathbf{S}}_i^* \tag{4.2}$$

where V_i is the area of the chosen control volume, N_e is the number of edges it has, $\hat{\hat{n}}_{il}$ is the outward pointing unit normal to the edge common to cells i and l (where l represents a generic neighbouring cell) and δ_{il} is the length of that edge (as shown for a triangular mesh cell in Fig. 19). $\underline{\mathbf{S}}^* \approx \iint_{\text{cell}} \underline{S}\, \mathrm{d}x\, \mathrm{d}y$ is once more a numerical approximation to the source integral over the control volume. For simplicity we will again assume the scheme to be a cell centre discretisation in which the control volumes coincide with the mesh cells, although the techniques may also be applied to other types of schemes. The following analysis is similar to that presented in previous sections for the one-dimensional case.

4.1. *The First-Order Scheme*

The numerical fluxes which lead to the first-order Roe's scheme in two dimensions are given by

$$(\underline{F}_{il}^*, \underline{G}_{il}^*) \cdot \hat{\hat{n}}_{il} = \frac{1}{2}(\underline{F}_i + \underline{F}_l, \underline{G}_i + \underline{G}_l) \cdot \hat{\hat{n}}_{il} - \frac{1}{2}(\tilde{\mathbf{R}}|\tilde{\mathbf{\Lambda}}|\tilde{\mathbf{R}}^{-1}\Delta \underline{U})_{il}, \tag{4.3}$$

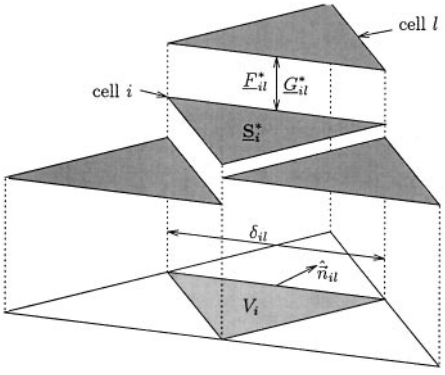

FIG. 19. Numerical fluxes and sources for the cell centre scheme.

in which the eigenvectors and eigenvalues needed to construct $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{\Lambda}}$ are now those of the matrix $\tilde{\mathbf{C}}_n = (\tilde{\mathbf{A}}, \tilde{\mathbf{B}}) \cdot \hat{\vec{n}}$, where

$$\tilde{\mathbf{A}} \approx \frac{\partial \underline{F}}{\partial \underline{U}} \quad \text{and} \quad \tilde{\mathbf{B}} \approx \frac{\partial \underline{G}}{\partial \underline{U}} \tag{4.4}$$

are the linearised flux Jacobians. It can be seen that the numerical flux is similar in form to that used in one dimension (3.6). In particular, $\sim$ again denotes the evaluation of a quantity at its Roe-average state.

Since the two-dimensional scheme is based on Riemann solvers oriented perpendicularly to the edges of the grid cells, the decomposition also bears a strong resemblance to the one-dimensional case. Once more, as long as the quantities denoted $\sim$ are evaluated at the appropriate Roe-average state [16], the flux differences can be written in the decomposed form

$$\Delta(\underline{F}, \underline{G}) \cdot \hat{\vec{n}} = \tilde{\mathbf{C}}_n \Delta \underline{U} = \tilde{\mathbf{R}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{R}}^{-1} \Delta \underline{U} = \sum_{k=1}^{N_w} \tilde{\alpha}_k \tilde{\lambda}_k \tilde{\underline{r}}_k, \tag{4.5}$$

from which it follows in much the same way as in one dimension that the scheme (4.2) is equivalent to

$$\underline{U}_i^{n+1} = \underline{U}_i^n - \frac{\Delta t}{V_i} \sum_{l=1}^{N_e} \delta_{il} (\tilde{\mathbf{R}}\tilde{\mathbf{\Lambda}}^- \tilde{\mathbf{R}}^{-1} \Delta \underline{U})_{il} + \frac{\Delta t}{V_i} \underline{\mathbf{S}}_i^*, \tag{4.6}$$

where the superscript $-$ now indicates the incoming characteristics at the appropriate edge of the control volume (see Fig. 20). It is easily seen that this reduces to (3.7) when restricted to one dimension.

The terms within (4.5) may again be required to balance the flux difference, so the same Roe linearisation is used in their evaluation, and it follows that

$$\underline{F}_x + \underline{G}_y - \underline{S} \equiv \underline{0} \Rightarrow (\Delta(\underline{F}, \underline{G}) \cdot \vec{n} - \underline{\tilde{\mathbf{S}}})_{il} = \underline{0}, \tag{4.7}$$

throughout the domain: $\diamond_{il}$ is the edge-cell corresponding to the edge between cells i and l, as shown in Fig. 20. The three-dimensional case is similar, with all the approximations being carried out over a face-cell with the solution being assumed constant on either side.

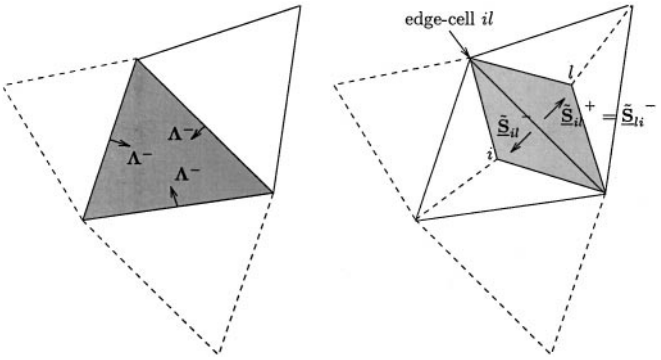

FIG. 20. Wave propagation directions (left) and source distribution (right) within a triangular cell in two dimensions.

The two-dimensional source term can now be written as a characteristic decomposition similar to that of the flux difference (4.5); i.e., its linearisation can take the form

$$\tilde{\underline{\mathbf{S}}}_{il} = (\tilde{\mathbf{R}}\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}})_{il} = \delta_{il}\left(\sum_{k=1}^{N_w}\tilde{\beta}_k\tilde{\underline{r}}_k\right)_{il}. \tag{4.8}$$

Evaluating this at the same Roe-average state as the flux difference means that the correct balance is attained because, at equilibrium, the decompositions give $\delta\mathbf{\Lambda}\mathbf{R}^{-1}\Delta\underline{U} = \mathbf{R}^{-1}\underline{\mathbf{S}}$. $\underline{\mathbf{S}}_i^*$ will be constructed out of contributions from each edge of the cell, with consistency ensured as long as the whole of each edge-cell integral (4.8) is distributed.

The decomposition has been carried out so that, when (4.6) is combined with (4.8) to give

$$\underline{U}_i^{n+1} = \underline{U}_i^n - \frac{\Delta t}{V_i}\sum_{l=1}^{N_e}(\tilde{\mathbf{R}}(\delta\tilde{\mathbf{\Lambda}}^-\tilde{\mathbf{R}}^{-1}\Delta\underline{U} - \mathbf{I}^-\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}}))_{il}, \tag{4.9}$$

a precise balance can be achieved when one is sought between the sources and the flux gradients.

The relationship between the two forms of the finite volume scheme, (4.2) and (4.6), can now be exploited. Substituting for $\mathbf{I}^-$ in (4.9) gives

$$\begin{aligned}\underline{U}_i^{n+1} = \underline{U}_i^n &- \frac{\Delta t}{2V_i}\sum_{l=1}^{N_e}(\tilde{\mathbf{R}}(\delta\tilde{\mathbf{\Lambda}}\tilde{\mathbf{R}}^{-1}\Delta\underline{U} - \tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}}))_{il}\\ &- \frac{\Delta t}{2V_i}\sum_{l=1}^{N_e}(\tilde{\mathbf{R}}(\delta|\tilde{\mathbf{\Lambda}}|\tilde{\mathbf{R}}^{-1}\Delta\underline{U} - \mathrm{sgn}(\mathbf{I})\tilde{\mathbf{R}}^{-1}\tilde{\underline{\mathbf{S}}}))_{il}.\end{aligned} \tag{4.10}$$

In addition, it is easily shown that

$$\sum_{l=1}^{N_e}\Delta(\underline{F}_{il}, \underline{G}_{il})\cdot\vec{n}_{il} = \sum_{l=1}^{N_e}(\underline{F}_i + \underline{F}_l, \underline{G}_i + \underline{G}_l)\cdot\vec{n}_{il}, \tag{4.11}$$

in which $\Delta\underline{F}_{il} = \underline{F}_l - \underline{F}_i$ is the jump in $\underline{F}$ across the lth edge of cell i (and similarly for

$\underline{G}$). Therefore, since $\sim$ indicates evaluation at the Roe-average state, (4.5) holds and (4.10) can be rewritten as

$$\underline{U}_i^{n+1} = \underline{U}_i^n - \frac{\Delta t}{V_i} \sum_{l=1}^{N_e} (\underline{F}_{il}^*, \underline{G}_{il}^*) \cdot \vec{n}_{il} + \frac{\Delta t}{V_i} \underline{\mathbf{S}}_i^*, \tag{4.12}$$

in which the numerical fluxes are given by (4.3) and the numerical source is

$$\underline{\mathbf{S}}_i^* = \sum_{l=1}^{N_e} \underline{\tilde{\mathbf{S}}}_{il}^-, \tag{4.13}$$

where

$$\underline{\tilde{\mathbf{S}}}_{il}^- = \frac{1}{2}(\tilde{\mathbf{R}}(\mathbf{I} - \operatorname{sgn}(\mathbf{I}))\tilde{\mathbf{R}}^{-1}\underline{\tilde{\mathbf{S}}})_{il} = (\tilde{\mathbf{R}}\mathbf{I}^-\tilde{\mathbf{R}}^{-1}\underline{\tilde{\mathbf{S}}})_{il}. \tag{4.14}$$

These expressions bear a close resemblance to the numerical fluxes and can be incorporated into the flux-based scheme in a similar manner at little extra expense. The special case of a structured rectangular mesh gives simplified expressions, which are described and applied in a separate publication [10]. As in one dimension, it is not possible to combine the source term completely with the numerical fluxes.

4.2. *High-Resolution Schemes*

When the accuracy of the scheme is increased by the use of a flux-limiting, technique, the numerical flux takes the form

$$(\underline{F}_{il}^*, \underline{G}_{il}^*) \cdot \hat{\vec{n}}_{il} = \frac{1}{2}(\underline{F}_i + \underline{F}_l, \underline{G}_i + \underline{G}_l) \cdot \hat{\vec{n}}_{il} - \frac{1}{2}(\tilde{\mathbf{R}}\tilde{\mathbf{\Lambda}}\mathbf{L}\tilde{\mathbf{R}}^{-1}\,\Delta\underline{U})_{il}, \tag{4.15}$$

and the appropriate discretisation of the source term can be shown to be

$$\underline{\tilde{\mathbf{S}}}_{il}^- = \frac{1}{2}(\tilde{\mathbf{R}}(\mathbf{I} - \operatorname{sgn}(\mathbf{I})\mathbf{L})\tilde{\mathbf{R}}^{-1}\underline{\tilde{\mathbf{S}}})_{il}, \tag{4.16}$$

by arguments similar to those used in one dimension.

For a MUSCL-type slope-limited higher order numerical scheme, the numerical fluxes take the form

$$(\underline{F}_{il}^*, \underline{G}_{il}^*) \cdot \hat{\vec{n}}_{il} = \frac{1}{2}(\underline{F}_{Il} + \underline{F}_{iL}, \underline{G}_{Il} + \underline{G}_{iL}) \cdot \hat{\vec{n}}_{il} - \frac{1}{2}(\tilde{\mathbf{R}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{R}}^{-1}\,\Delta\underline{U})_{il}, \tag{4.17}$$

in which the subscripts Il and iL represent evaluation of the piecewise linear reconstruction of the solution on, respectively, the inside and the outside of the edge between cells i and l, relative to cell i (indicated in Fig. 21), giving new values from which the Roe averages at the interface are calculated. Now, instead of (4.11) the flux differences satisfy the more general expression

$$\sum_{l=1}^{N_e} \Delta(\underline{F}_{il}, \underline{G}_{il}) \cdot \vec{n}_{il} = \sum_{l=1}^{N_e} (\underline{F}_{Il} + \underline{F}_{iL}, \underline{G}_{Il} + \underline{G}_{iL}) \cdot \vec{n}_{il} - 2\sum_{l=1}^{N_e} (\underline{F}_{Il} - \underline{F}_i, \underline{G}_{Il} - \underline{G}_i) \cdot \vec{n}_{il}. \tag{4.18}$$

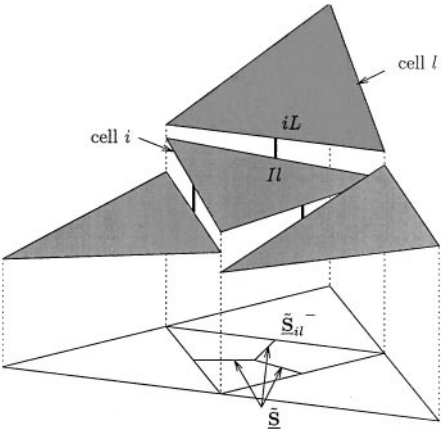

FIG. 21. Flux and source evaluations for a two-dimensional MUSCL-type scheme on triangles.

Consequently, the numerical source term appropriate to this type of scheme is given by

$$\underline{\mathbf{S}}_i^* = \sum_{l=1}^{N_e} \left(\tilde{\underline{\mathbf{S}}}_{il}^{-} - \tilde{\underline{\mathbf{S}}}(\underline{U}_{Il}, \underline{U}_i) \right), \tag{4.19}$$

where $\sim$ indicates the evaluation of the source term integral at the Roe average of the specified conservative variables (taken from the linear reconstruction at the midpoints of the cell edges) and $\tilde{\underline{\mathbf{S}}}_{il}^{-}$ is taken directly from (4.14). This can again be considered as applying a higher order correction to the integral of the source term over the edge-cell.

4.3. *Shallow Water Flows*

In two dimensions the shallow water equations including the effects of varying bed slope are obtained by substituting

$$\underline{U} = \begin{pmatrix} d \\ du \\ dv \end{pmatrix}, \quad \underline{F} = \begin{pmatrix} du \\ du^2 + \frac{gd^2}{2} \\ duv \end{pmatrix}, \quad \underline{G} = \begin{pmatrix} dv \\ duv \\ dv^2 + \frac{gd^2}{2} \end{pmatrix}, \tag{4.20}$$

where v is the flow velocity in the y-direction, in addition to the variables defined for (3.30), and

$$\underline{S} = \begin{pmatrix} 0 \\ gdh_x \\ gdh_y \end{pmatrix}, \tag{4.21}$$

into (4.1). The matrix $\mathbf{C}_n$ can be calculated simply from these for any edge orientation.

When $d \equiv h$ and $u \equiv v \equiv 0$ (quiescent flow in two dimensions) the desired balance is given by the equations

$$\left(\frac{gd^2}{2}\right)_x = gdh_x, \quad \left(\frac{gd^2}{2}\right)_y = gdh_y. \tag{4.22}$$

The discretisation should satisfy (4.22) exactly in this special case.

The characteristic decomposition is now carried out on the eigenvectors of the matrix $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}) \cdot \hat{\vec{n}}$, which are

$$\underline{\tilde{r}}_1 = \begin{pmatrix} 1 \\ \tilde{u} + \tilde{c}n^x \\ \tilde{v} + \tilde{c}n^y \end{pmatrix}, \quad \underline{\tilde{r}}_2 = \begin{pmatrix} 0 \\ -\tilde{c}n^y \\ \tilde{c}n^x \end{pmatrix}, \quad \underline{\tilde{r}}_3 = \begin{pmatrix} 1 \\ \tilde{u} - \tilde{c}n^x \\ \tilde{v} - \tilde{c}n^y \end{pmatrix}, \tag{4.23}$$

in which $(n^x, n^y) = \hat{\vec{n}}$ and

$$\tilde{c} = \sqrt{\frac{g(d^{\mathrm{R}} + d^{\mathrm{L}})}{2}}, \quad \tilde{u} = \frac{\sqrt{d^{\mathrm{R}}}u^{\mathrm{R}} + \sqrt{d^{\mathrm{L}}}u^{\mathrm{L}}}{\sqrt{d^{\mathrm{R}}} + \sqrt{d^{\mathrm{L}}}}, \quad \tilde{v} = \frac{\sqrt{d^{\mathrm{R}}}v^{\mathrm{R}} + \sqrt{d^{\mathrm{L}}}v^{\mathrm{L}}}{\sqrt{d^{\mathrm{R}}} + \sqrt{d^{\mathrm{L}}}}. \tag{4.24}$$

The superscripts R and L indicate here the evaluation of a quantity on either side of a cell edge, at its midpoint. The corresponding expressions for the eigenvalues (wave speeds) are

$$\lambda_1 = \tilde{u}n^x + \tilde{v}n^y + \tilde{c}, \quad \lambda_2 = \tilde{u}n^x + \tilde{v}n^y, \quad \lambda_3 = \tilde{u}n^x + \tilde{v}n^y - \tilde{c}, \tag{4.25}$$

and the wave strengths,

$$\begin{aligned} \tilde{\alpha}_1 &= \frac{\Delta d}{2} + \frac{1}{2\tilde{c}}(\Delta(du)n^x + \Delta(dv)n^y - (\tilde{u}n^x + \tilde{v}n^y)\Delta d) \\ \tilde{\alpha}_2 &= \frac{1}{\tilde{c}}((\Delta(dv) - \tilde{v}\Delta d)n^x - (\Delta(du) - \tilde{u}\Delta d)n^y) \\ \tilde{\alpha}_3 &= \frac{\Delta d}{2} - \frac{1}{2\tilde{c}}(\Delta(du)n^x + \Delta(dv)n^y - (\tilde{u}n^x + \tilde{v}n^y)\Delta d), \end{aligned} \tag{4.26}$$

complete the decomposition (4.5).

In this case, in order to provide the desired balance, the source term is written

$$\underline{S} = \begin{pmatrix} 0 \\ gd \\ 0 \end{pmatrix} \vec{\nabla} \cdot (h, 0) + \begin{pmatrix} 0 \\ 0 \\ gd \end{pmatrix} \vec{\nabla} \cdot (0, h). \tag{4.27}$$

At first glance this seems counterproductive, but it immediately allows the source term integral over an edge-cell to be approximated in a manner which will allow the discrete balance with the flux integral. This leads to

$$\underline{\tilde{\mathbf{S}}}_{il} = \delta_{il} \begin{pmatrix} 0 \\ g\tilde{d}\Delta h n^x \\ g\tilde{d}\Delta h n^y \end{pmatrix}, \tag{4.28}$$

which is used to obtain the coefficients used in the characteristic decomposition (4.8). In this case these are

$$\tilde{\beta}_1 = \frac{1}{2}\tilde{c}\Delta h, \quad \tilde{\beta}_2 = 0, \quad \tilde{\beta}_3 = -\frac{1}{2}\tilde{c}\Delta h. \tag{4.29}$$

By construction, it follows that $\tilde{\alpha}_k\tilde{\lambda}_k - \tilde{\beta}_k = 0$ for each k; i.e.,

$$\tilde{\mathbf{R}}(\delta\tilde{\mathbf{\Lambda}}\tilde{\mathbf{R}}^{-1}\,\Delta\underline{U} - \tilde{\mathbf{R}}^{-1}\underline{\tilde{\mathbf{S}}}) = \underline{0} \tag{4.30}$$

when the flow is quiescent, and the numerical balance is ensured.

4.4. *Numerical Results*

The test cases presented in this section are all related to those described in Section 3.3 for the one-dimensional schemes, but applied to the two-dimensional shallow water equations. For the purposes of presentation, comparisons will often be made between breadth-averaged solutions for channel flows and exact solutions to the corresponding one-dimensional problem. These will obviously differ slightly in the non-quiescent cases due to the simplifications inherent in the one-dimensional model, but they still provide an accurate guide when the crossflow velocity is small, as it is in the results presented. The Minmod limiter [20] is again used in all high-resolution calculations.

The ability of the new techniques to maintain the still water steady state is illustrated using the geometry of Fig. 7 and a triangular grid with 4854 cells and 2738 nodes (giving about 300 cells along the channel, or roughly half the one-dimensional grid resolution). As in one dimension, the upwind source term discretisation maintains still, flat water indefinitely to machine accuracy in both the first-order and the high-resolution cases, see Fig. 22. This is true of any geometry and bed topography and each of the schemes described earlier in the text. The pointwise evaluation of the source term is clearly unable to match this.

Results for the tidal flow test case described in Section 3.3.1 are shown in Fig. 23 for the same triangular grid. Again, the advantage of using the upwind source term discretisation is clearly visible and here, unlike in one dimension, the CFL number used to obtain the results is still 0.8. When the source terms are upwinded the results from the high-resolution scheme are almost oscillation-free. The averaging across the channel breadth does produce a small

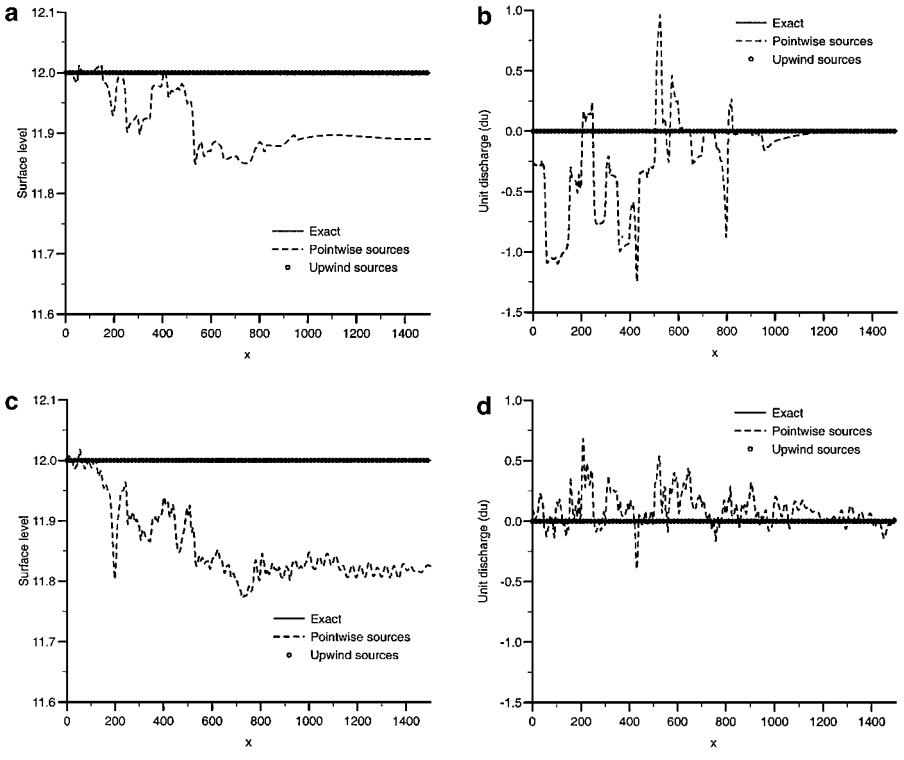

FIG. 22. Breadth-averaged water surface level and unit discharge for the two-dimensional still water test case for first-order (a, b) and high-resolution slope-limited (c, d) schemes ($t = 1000$).

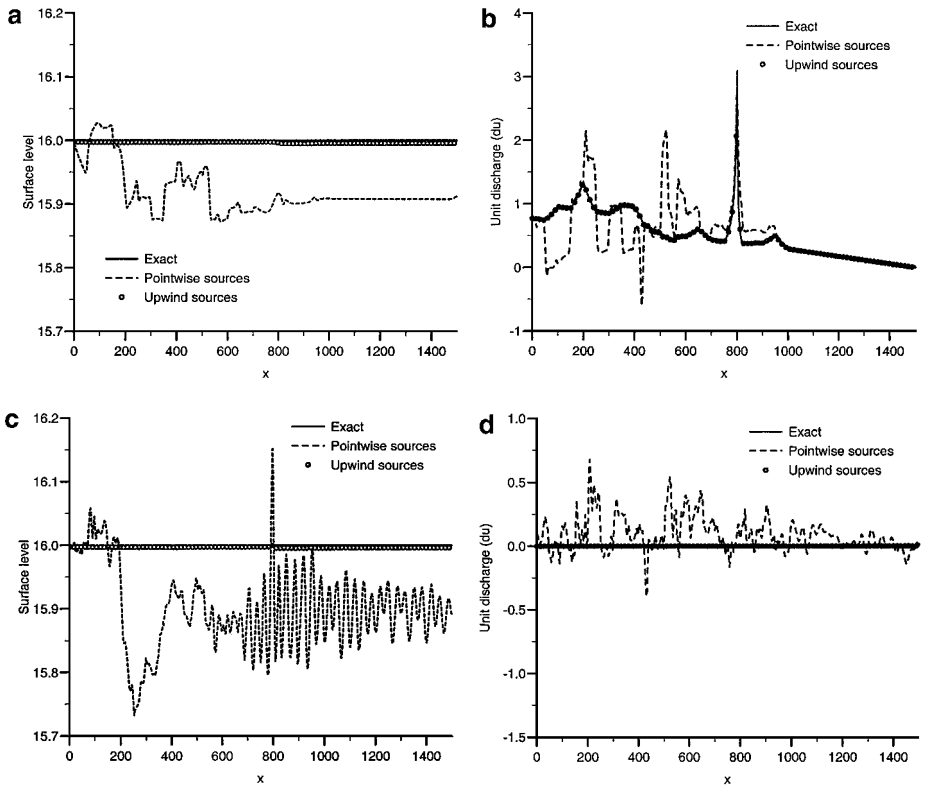

FIG. 23. Breadth-averaged water surface level and unit discharge for the two-dimensional tidal flow test case for first-order (a, b) and high-resolution slope-limited (c, d) schemes.

amount of smoothing, but it can be seen from Fig. 24, which shows a scatter plot of the water surface level from side on, that this smoothing is not significant and that oscillations do not appear in the solution when the source terms are upwinded. Even the slope-limited scheme gives good results at a CFL number of 0.8. This appears to be because the numerical cross-diffusion which occurs on the triangular grid smooths out the small oscillations which can appear and counteracts their growth over long times, rather than the absence of the source terms representing breadth variation.

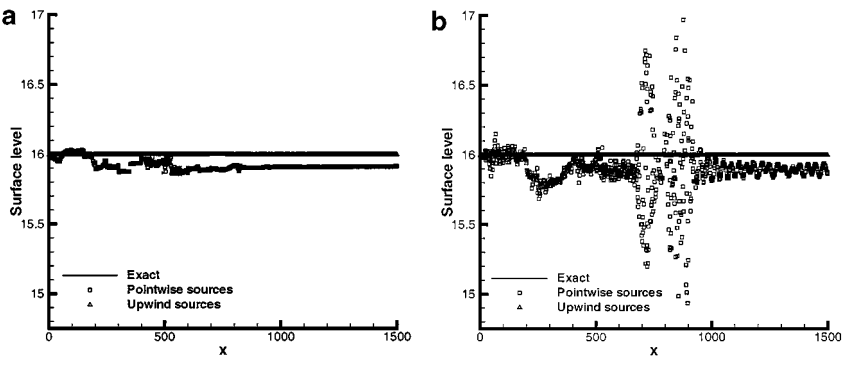

FIG. 24. Scatter plots of water surface level for the tidal flow test case for first-order (a) and high-resolution slope-limited (b) schemes.

The last test case presented is again taken from [14] and demonstrates the ability of the new method to model wave propagation problems. The computational domain is the unit square $[0:1, 0:1]$ and the bed topography is defined by

$$h(x, y) = 1.0 - 0.5 \exp(-50.0[(x - 0.5)^2 + (y - 0.5)^2]). \tag{4.31}$$

The data are initially stationary ($u \equiv v \equiv 0$), with a surface profile defined by

$$d(x, y) = \begin{cases} h(x, y) + 0.01 & \text{for } 0.1 < x < 0.2, 0.0 \le y \le 1.0 \\ h(x, y) & \text{otherwise.} \end{cases} \tag{4.32}$$

The water surface levels obtained using the flux-limited (CFL $= 0.8$) scheme on a uniform 100×100 square grid, and the slope-limited (also CFL $= 0.8$) scheme on a fairly uniform 5218 node, 10170 cell triangular grid, are shown in Fig. 25 at a time $t = 0.7$. They both show that this small disturbance propagates over the variable bed without any interference from numerically induced noise and they agree with the results presented in [14].

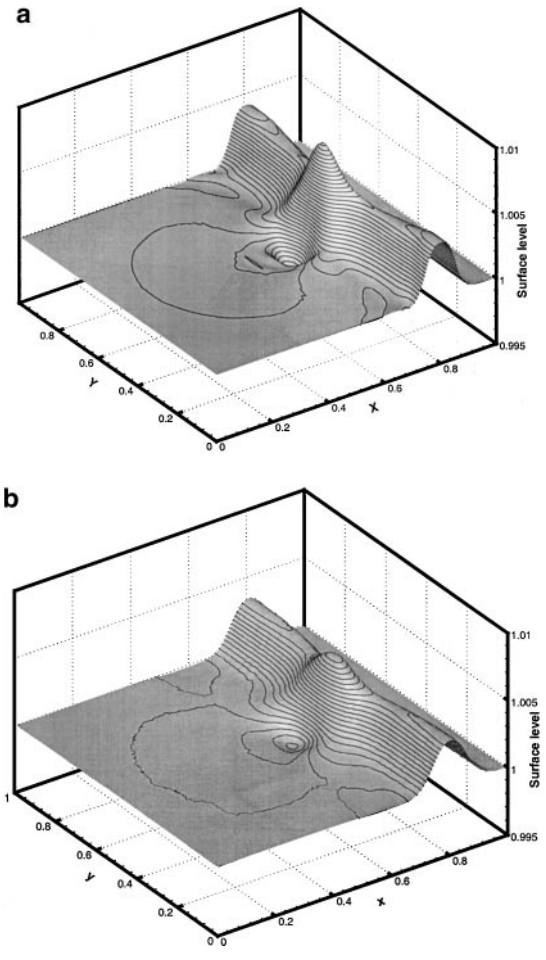

FIG. 25. Water surface levels for the two-dimensional wave propagation test case at $t = 0.7$ using upwinded source terms with a flux-limited scheme on quadrilaterals (a) and a slope-limited scheme on triangles (b).

Generally, it has been seen that the properties exhibited by the schemes in one dimension are carried over into higher dimensions. Extensive results obtained for the two-dimensional flux-limited schemes on rectangular meshes, which confirm these observations, can be found in [10].

5. CONCLUSIONS

In this paper a new method has been presented for the discretisation of source terms when they appear as part of a nonlinear system of conservation laws. Specifically, the correct approximation to the source terms is sought, given that a particular finite volume scheme has been used for the discretisation of the flux terms. Roe's scheme has been chosen here as the underlying numerical scheme, but the philosophy behind the source term approximation (that the source terms must, in some sense, be discretised in the same manner as the flux derivatives) may also be applied to other finite volume methods. The discretisation builds on the work of many previous authors [1, 5–7], who approximated their source terms in a manner which took into account the flux discretisation and, as a consequence, allowed the numerical model to maintain specific equilibria which are satisfied by the mathematical model. The new aspect of this work is the generalisation of these techniques to high-order TVD versions of Roe's scheme (using both flux limiters and slope limiters) and to arbitrary polygonal meshes in any number of dimensions. The methods have been designed specifically for source terms which provide some sort of balance with the flux derivatives. Even so, the same techniques can easily be applied to other source terms (such as those which model bed friction in the shallow water equations) which do not exhibit a precise balance, but the advantages over the simple pointwise discretisation are less obvious.

The effectiveness of these techniques has been illustrated using the one- and two-dimensional shallow water equations (the extension to three-dimensional systems of equations is straightforward, though not described here in detail), in which source terms are used to model variations in the bed topography and (in one dimension) channel breadth. Particular attention has been paid to the special case of still water, and the schemes have been constructed so that they maintain this state. In fact, the improved accuracy of the new upwind discretisation of the source terms is also shown in the approximation of other steady-state solutions, particularly in one dimension when flux limiters have been used, and to a great extent by time-dependent test cases as well. These improvements were also seen on non-uniform grids. The improvement is less marked for slope-limited schemes, indicating that a more sophisticated approximation to the source term may be necessary away from the still water steady state. This has been shown by comparison with a selection of test cases for which exact solutions are available. The advantages over the commonly used pointwise discretisations are particularly apparent when quantities depending on the flow velocity are compared. In the high-resolution schemes the new source term approximation also gives a great improvement over the first-order decomposed source term.

At this stage of the research, the main problem with the new technique (a problem which also applies to many previous methods) is in the modelling of time-dependent problems. Here, in order to avoid spurious oscillations in the slope-limited results, a low CFL number has to be imposed (0.1 in the cases tested here), and in some cases the unphysical oscillations cannot be removed completely. These oscillations are very sensitive to the variables to which the reconstruction of the solution is applied since changing the reconstruction can alter the balance between terms. Even so, over short time scales the source term correction applied

in the slope-limiting case still improves significantly on the pointwise discretisations. This problem is also related to the TVD condition satisfied by the scheme, which is only really intended for the homogeneous equations. It is encouraging that these problems have not arisen in the flux-limited cases. The possible construction of a TVD condition in the presence of source terms is a topic for future research. In the meantime it may prove beneficial in some cases to apply a flux-corrected transport approach since it is clear from the techniques presented in this paper how the source terms should be treated for both upwind and Lax–Wendroff schemes, and the first-order upwind scheme appears to be robust enough to eradicate the unwanted oscillations.

Finally, an alternative method has also been proposed for the discretisation of the flux term in the case where it varies spatially but independent of the flow variables (as with one-dimensional models of shallow water flow through a channel of variable breadth). It has been shown that, in combination with the source term approximation, the method produces accurate solutions for a wide variety of steady-state and time-dependent test cases.

ACKNOWLEDGMENTS

The first author thanks the EPSRC for providing the funding for his research, while the second author thanks CICYT for her funding under Research Project HID98-1099-C02-01. The authors also thank the referees, whose comments helped greatly to improve the presentation of this paper.

REFERENCES

1. A. Bermúdez and M. E. Vázquez, Upwind methods for hyperbolic conservation laws with source terms, *Comput. Fluids* **23**(8), 1049 (1994).
2. A. Bermúdez, A. Dervieux, J.-A. Desideri, and M. E. Vázquez, Upwind schemes for the two-dimensional shallow water equations with variable depth using unstructured meshes, *Comput. Methods Appl. Mech. Eng.* **155**, 49 (1998).
3. J. A. Cunge, F. M. Holly, and A. Verwey, *Practical Aspects of Computational River Hydraulics* (Pitman, London, 1980).
4. N. Goutal and F. Maurel, *Proceedings of the 2nd Workshop on Dam-Break Wave Simulation*, Technical Report HE-43/97/016/A (Electricité de France, Département Laboratoire National d'Hydraulique, Groupe Hydraulique Fluviale, 1997).
5. P. Garcia-Navarro and M. E. Vazquez-Cendon, *Some Considerations and Improvements on the Performance of Roe's Scheme for 1d Irregular Geometries*, Internal Report 23 (Departamento de Matemática Aplicada, Universidade de Santiago do Compostela, 1997).
6. P. Glaister, *Difference Schemes for the Shallow Water Equations*, Numerical Analysis Report 9/87 (Department of Mathematics, University of Reading, 1987).
7. P. Glaister, Prediction of supercritical flow in open channels, *Comput. Math. Appl.* **24**(7), 69 (1992).
8. M. E. Hubbard, *On the Accuracy of One-Dimensional Models of Steady Converging/Diverging Open Channel Flows*, Numerical Analysis Report 1/99 (Department of Mathematics, University of Reading, 1999).
9. J. Huang and C. S. Song, Stability of dynamic flood routing schemes, *J. Hydraul. Eng. ASCE* **111**(12) (1985).
10. M. E. Hubbard and N. Dodd, A 2d numerical model of wave run-up, overtopping and regeneration, in preparation.
11. J. Hudson, *Numerical Techniques for the Shallow Water Equations*, Numerical Analysis Report 2/99 (Department of Mathematics, University of Reading, 1999).
12. P. Jenny and B. Muller, Rankine–Hugoniot–Riemann solver considering source terms and multidimensional effects, *J. Comput. Phys.* **145**(2), 575 (1998).
13. R. J. LeVeque, *Numerical Methods for Conservation Laws* (Birkhäuser, Basel, 1992).

14. R. J. LeVeque, Balancing source terms and flux gradients in high-resolution Godunov methods: The quasi-steady wave-propagation algorithm, *J. Comput. Phys.* **146**(1), 346 (1998).
15. A. Priestley, *Roe-Type Schemes for Super-Critical Flows in Rivers*, Numerical Analysis Report 13/89 (Department of Mathematics, University of Reading, 1989).
16. P. L. Roe, Approximate Riemann solvers, parameter vectors, and difference schemes, *J. Comput. Phys.* **43**(2), 357 (1981).
17. P. L. Roe, Fluctuations and signals—A framework for numerical evolution problems, in *Numerical Methods for Fluid Dynamics*, edited by K. W. Morton (Oxford Univ. Press, London, 1982), 219–257.
18. P. L. Roe, Characteristic-based schemes for the Euler equations, *Ann. Rev. Fluid Mech.* **18**, 337 (1986).
19. P. K. Smolarkiewicz and L. G. Margolin, MPDATA: A finite-difference solver for geophysical flows, *J. Comput. Phys.* **140**, 459 (1998).
20. P. K. Sweby, High resolution schemes using flux limiters for hyperbolic conservation laws, *SIAM J. Numer. Anal.* **21**, 995 (1984).
21. M. E. Vázquez-Cendón, Improved treatment of source terms in upwind schemes for the shallow water equations in channels with irregular geometry, *J. Comput. Phys.* **148**(2), 497 (1999).
22. B. van Leer, Towards the ultimate conservative difference scheme V. A second order sequel to Godunov's method, *J. Comput. Phys.* **32**, 101 (1979).