Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

E.D. de Goede

A time splitting method for the three-dimensional shallow water equations

Department of Numerical Mathematics Report NM-R9013 August

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O.).

A Time Splitting Method for the Three-Dimensional Shallow Water Equations

Erik D. de Goede
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

In this paper we describe a time splitting method for the three-dimensional shallow water equations. The stability of this method neither depends on the vertical diffusion term nor on the terms describing the propagation of the surface waves. The method consists of two stages and requires the solution of a sequence of linear systems. For the solution of these systems we apply a Jacobi-type iteration method and a Conjugate Gradient iteration method. The performance of both methods is accelerated by a technique based on smoothing. The resulting method is mass conservative and efficient on vector and parallel computers. The accuracy, stability and computational efficiency of this method is demonstrated for wind induced problems in a rectangular basin.

1980 Mathematics Subject Classification: 65M10, 76D99.
Key Words & Phrases: three-dimensional shallow water equations, method of lines, time integrators, stability, vector and parallel computers.
Note: This work was supported by the Ministry of Transport and Public Works (RWS).
Note: This paper will be submitted for publication elsewhere.

1. Introduction

In this paper a time splitting method for the three-dimensional shallow water equations (SWEs) will be described. The aim of splitting methods is always to split the solution of a large and complicated system, which arises when applying fully implicit methods to multi-dimensional problems, into a few less complicated systems. Well-known splitting methods are alternating direction implicit (ADI) methods, locally one-dimensional (LOD) methods and Hopscotch methods [6].

For the two-dimensional shallow water equations several of the existing numerical methods have been based on the ADI method (see e.g., [10,12]). These ADI methods are unconditionally stable and therefore allow the use of large time steps. However, for large time steps these methods suffer from inaccuracies when dealing with complex geometries [13]. In [15] a two-stage time-splitting method has been developed in which these inaccuracies are absent, even for large time steps.

In this paper we will present a two-stage time-splitting method for the three-dimensional shallow water equations which has a strong resemblance with the method in [15]. We will use a model for the shallow water equations in which the advective terms have been omitted. The stability of a numerical method for this model depends on the conditions imposed by the vertical diffusion term and by the terms describing the propagation of the surface waves (the CFL condition). In two-dimensional models many methods are known in which the terms describing the propagation of the surface waves are treated implicitly (see e.g., [1,2,15]). In addition to that, in three-dimensional models, where a vertical diffusion term is involved, we have to treat the vertical diffusion term implicitly to avoid the maximally stable time

step becoming too small. (see e.g., [3]). In this paper we will develop a two-stage method in which the vertical diffusion is treated implicitly at the first stage, whereas the terms concerning the propagation of surface waves are treated implicitly at the second stage. It will be shown that the stability of this time splitting method neither depends on the vertical diffusion nor on the propagation of the surface waves. Due to computational efficiency, the Coriolis term will be treated in a semi-implicit way. The Coriolis term hardly affects the stability, which justifies this simplification.

At the first stage our time integration method requires the solution of a large number of tridiagonal systems, all of the same dimension. Since the tridiagonal systems are independent of each other, the solution of these systems can be computed in parallel [4].

At the second stage a linearization process is used to solve iteratively the nonlinear system. The linearization is done in such a way that conservation of mass remains guaranteed. Then at each iteration step a linear, symmetric, positive definite system has to be solved. In literature, a large number of iteration methods have been proposed for such systems. In this paper we apply a Jacobi-type iteration method and a Conjugate Gradient iteration method for the solution of this system. Both iteration methods will be accelerated by a technique based on smoothing. Application of the smoothing matrices reduces the number of iterations considerably. Moreover, the smoothing matrices are very simple to implement and are highly suited for vector and parallel computers.

In [15] a two-stage time-splitting method has been developed for the two-dimensional shallow water equations. It was reported that this time-splitting method is feasible for practical computations. For this method, a major part of the computation is involved in the nonlinear system at the second stage. Since in our time-splitting method the water elevation is the only unknown in the system at the second stage, this system is of the same (two-dimensional) structure and thus of the same computational complexity for both two-dimensional and three-dimensional models. The computation time required by the other parts of our method, viz., the computation of the three-dimensional velocity components, is proportional to the number of grid layers in the vertical direction. Therefore, the efficiency of the time-splitting method developed in this paper is relatively even higher for three-dimensional models than for two-dimensional models.

The accuracy, stability and computational efficiency of our time-splitting method will be illustrated in the numerical experiments.

2. Mathematical model

In this section we will describe a mathematical model for the three-dimensional shallow water equations. The following symbols are used:

C	Chezy coefficient
d	undisturbed depth of water
f	Coriolis term
F_b	bottom stress in x-direction
F_s	surface stress in x-direction
g	acceleration due to gravity
G_b	bottom stress in y-direction

G_s	surface stress in y-direction
h	total depth (= d + ζ)
N	vertical diffusion coefficient
t	time
u,v	velocity components in x- and y-direction
x,y,σ	a left-handed set of coordinates
W_f	wind stress
ρ	density
φ	angle between wind direction and the positive x-axis
ζ	elevation above undisturbed depth.

We will use a three-dimensional model in sigma coordinates in which the advective terms have been omitted. In this paper we focus on stability conditions imposed by the vertical diffusion term and by the terms describing the propagation of the surface waves. In future we will develop a numerical method for a mathematical model in which the advective terms are present.

The mathematical model used in this paper is described by

$$\frac{\partial u}{\partial t} = fv - g\frac{\partial \zeta}{\partial x} + \frac{1}{\rho h^2}\frac{\partial\left(N\frac{\partial u}{\partial \sigma}\right)}{\partial \sigma} \tag{2.1}$$

$$\frac{\partial v}{\partial t} = -fu - g\frac{\partial \zeta}{\partial y} + \frac{1}{\rho h^2}\frac{\partial\left(N\frac{\partial v}{\partial \sigma}\right)}{\partial \sigma} \tag{2.2}$$

$$\frac{\partial \zeta}{\partial t} = -\frac{\partial}{\partial x}\left(h\int_0^1 u\,d\sigma\right) - \frac{\partial}{\partial y}\left(h\int_0^1 v\,d\sigma\right), \tag{2.3}$$

with boundaries

$$0 \le x \le L$$
$$0 \le y \le B$$
$$1 \ge \sigma \ge 0 .$$

Thus, the domain is a rectangular basin. Due to the sigma transformation in the vertical, the domain is constant in time [3,5]. We have the closed boundary conditions

$$u(0,y,\sigma,t) = 0, \quad u(L,y,\sigma,t) = 0,$$
$$v(x,0,\sigma,t) = 0, \quad v(x,B,\sigma,t) = 0.$$

The boundary conditions at the sea surface ($\sigma = 0$) are given by

$$-\left(N\frac{\partial u}{\partial \sigma}\right)_0 = h F_s , \qquad -\left(N\frac{\partial v}{\partial \sigma}\right)_0 = h G_s ,$$

and at the bottom ($\sigma = 1$)

$$-\left(N\frac{\partial u}{\partial \sigma}\right)_1 = h F_b , \qquad -\left(N\frac{\partial v}{\partial \sigma}\right)_1 = h G_b .$$

The bottom stress is parametrized using a linear law of bottom friction, which is of the form

$$F_b = g\rho u_d / C^2, \qquad G_b = g\rho v_d / C^2,$$

with u_d and v_d the components of the velocity at some depth near the bottom. The surface stresses are expressed as

$$F_s = W_f \cos\varphi, \qquad G_s = W_f \sin\varphi .$$

3. Space discretization

For the space discretization of the equations (2.1)-(2.3) the computational domain is covered by an nx·ny·ns rectangular staggered grid. Due to the sigma transformation, we have a constant number of grid layers in the vertical direction. In what follows **U** is a grid function whose components $U_{i,j,k}$ approximate the velocity u. The components $U_{i,j,k}$ are numbered lexicographically. Likewise **V**, **Z**, **D** and **H** are grid functions for v, ζ, d and h, respectively. Figure 1 shows the horizontal grid spacing. Note that **D**, **H** and **Z** are only computed at the upper layer. Furthermore,

$\Lambda_{\sigma\sigma}$ is a tridiagonal matrix approximating the vertical diffusion term,

Θ_1 is an (nx·ny·ns)·(nx·ny) matrix (a row of ns diagonal matrices of order $(nx\cdot ny)^2$ with $\Delta\sigma_k$ on the diagonal of the kth submatrix),

Θ_2 is an (nx·ny)·(nx·ny·ns) matrix (a column of ns identity matrices of order $(nx\cdot ny)^2$),

F is a four-diagonal matrix (due to the grid staggering) of order $(nx\cdot ny\cdot ns)^2$, approximating the Coriolis term,

D_x and D_y are bidiagonal matrices (one diagonal and one lower diagonal) of order $(nx\cdot ny)^2$, approximating the differential operators $\partial/\partial x$ and $\partial/\partial y$, respectively,

E_x and E_y are bidiagonal matrices (one diagonal and one upper diagonal) with $E_x = -D_x^T$ and $E_y = -D_y^T$.

The matrices D_x and E_x differ because of the grid staggering. We remark that $\Lambda_{\sigma\sigma}$ also contains the discretization of the term $1/(\rho h^2)$.

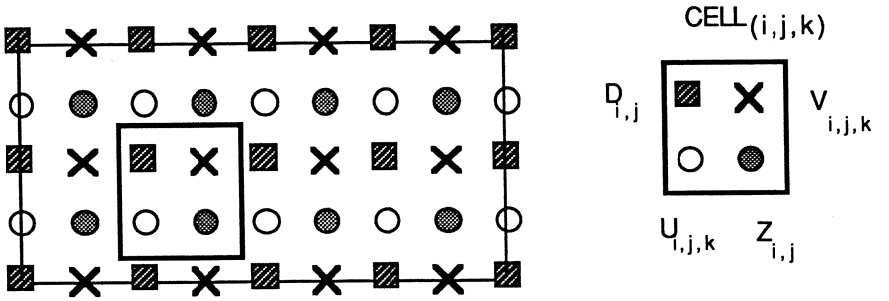

Figure 1. The staggered grid in (x,y)-plane

For the approximation of the spatial derivatives, second-order central finite differences are used in both the horizontal and vertical direction. Now, the semi-discretized system can be written in the form

$$\frac{d}{dt}\mathbf{W} = \mathbf{F}(\mathbf{W}) = \begin{pmatrix} \Lambda_{\sigma\sigma} & F & -\Theta_2 g D_x \\ -F & \Lambda_{\sigma\sigma} & -\Theta_2 g E_y \\ -\Theta_1 H E_x & -\Theta_1 H D_y & 0 \end{pmatrix} \mathbf{W} + \begin{pmatrix} F_u \\ F_v \\ 0 \end{pmatrix}, \tag{3.1}$$

where $\mathbf{W}=(U,V,Z)^T$ and $(F_u,F_v,0)^T$ contains the components of the wind stress. Note that the integrals in (2.3) are approximated by $\Theta_1 U$ and $\Theta_1 V$, respectively.

4. Time integration

In this section we develop a time integration method for the semi-discretized system (3.1). We apply a two-stage time splitting method of the form

$$\begin{aligned} \mathbf{W}^{n+1/2} &= \mathbf{W}^n + \tfrac{1}{2}\tau \{ \mathbf{F}^1(\mathbf{W}^{n+1/2}) + \mathbf{G}^1(\mathbf{W}^n) + \mathbf{C}^{n+1/2} \} \\ \mathbf{W}^{n+1} &= \mathbf{W}^{n+1/2} + \tfrac{1}{2}\tau \{ \mathbf{F}^2(\mathbf{W}^{n+1/2}) + \mathbf{G}^2(\mathbf{W}^{n+1}) + \mathbf{C}^{n+1/2} \}, \end{aligned} \tag{4.1}$$

where $\mathbf{C}=(F_u,F_v,0)^T$, τ denotes the time step and $\mathbf{W}^n$ is a numerical approximation to $\mathbf{W}(t)$ of (3.1) at $t=n\tau$. Several well-known splitting methods, e.g., ADI methods, can be written in this form. In this paper, we choose

$$\begin{aligned} \mathbf{F}^1(\mathbf{W}^{n+1/2}) &= \begin{pmatrix} \Lambda_{\sigma\sigma} & 0 & 0 \\ -F & \Lambda_{\sigma\sigma} & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{W}^{n+1/2}, \quad \mathbf{G}^1(\mathbf{W}^n) = \begin{pmatrix} 0 & F & -\Theta_2 g D_x \\ 0 & 0 & -\Theta_2 g E_y \\ -\Theta_1 H^n E_x & -\Theta_1 H^n D_y & 0 \end{pmatrix} \mathbf{W}^n, \\ \mathbf{F}^2(\mathbf{W}^{n+1/2}) &= \begin{pmatrix} \Lambda_{\sigma\sigma} & F & 0 \\ -F & \Lambda_{\sigma\sigma} & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{W}^{n+1/2} \text{ and } \mathbf{G}^2(\mathbf{W}^{n+1}) = \begin{pmatrix} 0 & 0 & -\Theta_2 g D_x \\ 0 & 0 & -\Theta_2 g E_y \\ -\Theta_1 H^{n+1} E_x & -\Theta_1 H^{n+1} D_y & 0 \end{pmatrix} \mathbf{W}^{n+1}. \end{aligned} \tag{4.2}$$

Apart from the Coriolis term F, all terms are treated in a symmetrical way. When we neglect the Coriolis term, the time splitting method (4.1)-(4.2) is second order accurate in time.

The structure of the resulting systems at both stages determines the efficiency of this time splitting method. At the first stage we have to solve the system

$$\begin{pmatrix} I - \frac{1}{2}\tau\Lambda_{\sigma\sigma} & 0 & 0 \\ \frac{1}{2}\tau F & I - \frac{1}{2}\tau\Lambda_{\sigma\sigma} & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} U^{n+1/2} \\ V^{n+1/2} \\ Z^{n+1/2} \end{pmatrix} = \mathbf{B}^n, \tag{4.3}$$

where $\mathbf{B}^n$ contains the discretizations at time level $t=n\tau$. It is evident that the **Z**-component can be computed straightforwardly. For the **U**- and **V**-component, the implicit treatment of the vertical diffusion term requires the

solution of nx·ny tridiagonal systems of order ns (see [4]). Due to computational efficiency, the Coriolis term is treated in a semi-implicit way. Although an implicit treatment of the Coriolis term for the V-component (see (4.3)) prohibits the U- and V-component from being computed in parallel, we prefer this choice because of accuracy considerations. The results are more accurate than in the case of a fully explicit treatment of the Coriolis term, especially when large time steps are used.

At the second stage the terms describing the propagation of the surface waves are treated implicitly. This system reads

$$\begin{pmatrix} I & 0 & \frac{1}{2}\tau\Theta_2 g D_x \\ 0 & I & \frac{1}{2}\tau\Theta_2 g E_y \\ \frac{1}{2}\tau\Theta_1 H^{n+1} E_x & \frac{1}{2}\tau\Theta_1 H^{n+1} D_y & I \end{pmatrix} \begin{pmatrix} U^{n+1} \\ V^{n+1} \\ Z^{n+1} \end{pmatrix} = B^{n+1/2}, \tag{4.4}$$

where $B^{n+1/2}$ contains the discretizations at time level $t=(n+1/2)\tau$. The equations for the U- and V-component are linear and are not coupled with each other. They are only coupled with the equation for the Z-component. Therefore, the components U^{n+1} and V^{n+1} can easily be eliminated from (4.4) and a system in the unknown Z^{n+1} results. Thus, at the second stage the continuity equation (2.3) and the water elevation gradient in the momentum equations (2.1)-(2.2) are treated implicitly. This approach was originally proposed in [8] and has been applied by many others (e.g., [1,2,15]).

We now describe this system for each cell (i,j) of component Z. The grid sizes in x- and y-direction are denoted by Δx and Δy, respectively. Then, the system for $Z_{i,j}$ reads

$$\begin{aligned} Z_{i,j}^{n+1} &- \frac{\tau^2 g}{4(\Delta x)^2}\left\{ \bar{H}_{i+1,j}^{n+1}\left(Z_{i+1,j}^{n+1} - Z_{i,j}^{n+1} \right) - \bar{H}_{i,j}^{n+1}\left(Z_{i,j}^{n+1} - Z_{i-1,j}^{n+1} \right) \right\} \\ &- \frac{\tau^2 g}{4(\Delta y)^2}\left\{ \tilde{H}_{i,j}^{n+1}\left(Z_{i,j+1}^{n+1} - Z_{i,j}^{n+1} \right) - \tilde{H}_{i,j-1}^{n+1}\left(Z_{i,j}^{n+1} - Z_{i,j-1}^{n+1} \right) \right\} = B_{i,j}^{n+1/2}, \quad \text{for } \begin{matrix} i=1,\ldots,nx \\ j=1,\ldots,ny \end{matrix}, \end{aligned} \tag{4.5}$$

where

$$\bar{H}_{i,j}^{n} = Z_{i,j}^{n} + \frac{1}{2}\left(D_{i,j} + D_{i,j-1} \right) \quad \text{and} \quad \tilde{H}_{i,j}^{n} = Z_{i,j}^{n} + \frac{1}{2}\left(D_{i,j} + D_{i+1,j} \right).$$

Note that $\bar{H}$ and $\tilde{H}$ differ because of the grid staggering. System (4.5) is a nonlinear equation, because $H_{i,j}$ contains the component $Z_{i,j}$. When system (4.5) has been solved, the values for the components U^{n+1} and V^{n+1} can be computed by back substitution.

System (4.5) can be written in the form

$$A(Z^{n+1})\, Z^{n+1} = B_z^{n+1/2}, \tag{4.6}$$

where $B_z^{n+1/2}$ contains the discretizations at $t=(n+1/2)\tau$ for the Z-component. For its linearization we introduce the process

$$A(Z^{(q)})\, Z^{(q+1)} = B_z^{n+1/2}\,, \tag{4.7}$$

with $Z^{(0)}=Z^{n+1/2}$. In (4.7) the upper index (q) denotes the iteration index. The matrix $A(Z^{(q)})$ is a symmetric and strictly diagonal dominant matrix with positive values on the main diagonal and negative ones elsewhere because we require that $D+Z^{(q)}$ $(=H^{(q)})$ >0. Thus, system (4.7) is positive definite. In Section 6 we will discuss iteration methods for the solution of system (4.7).

It should be noted that the water elevation is the only unknown in system (4.7). So, this system is of the same (two-dimensional) structure and thus of the same computational complexity for both two-dimensional and three-dimensional models. This is an important feature of the time integration method (4.1)-(4.2), because for two-dimensional problems a major part of the computation is required for the solution of this system.

The linearization process (4.7) has first been used by Leendertse [10]. Conservation of mass remains guaranteed by this process. A slightly different linearization process has been introduced in [15]. In our numerical experiments (see Section 7) we obtained comparable results for both linearization processes. The linearization process will be topic of further research.

5. Stability

We now analyze the stability of method (4.1)-(4.2) with the matrix method. In this section we omit the Coriolis force and the inhomogeneous term. It is well-known that the Coriolis force hardly affects the stability. In this section we make plausible that the simplified method is unconditionally stable. The stability analysis used here is similar to the one described in [14]. That paper was devoted to a study of the stability and convergence properties of the Peaceman-Rachford ADI method when applied to initial-boundary value problems, including nonlinear ones.

Since we have omitted the Coriolis force and the inhomogeneous term, we have that (cf. (4.2))

$$F^1(W^n) = F^2(W^n) = A^n\, W^n \quad \text{and} \quad G^1(W^n) = G^2(W^n) = B^n\, W^n\,,$$

where

$$A^n = \begin{pmatrix} \Lambda_{\sigma\sigma} & 0 & 0 \\ 0 & \Lambda_{\sigma\sigma} & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad B^n = \begin{pmatrix} 0 & 0 & -\Theta_2 g D_x \\ 0 & 0 & -\Theta_2 g E_y \\ -\Theta_1 H^n E_x & -\Theta_1 H^n D_y & 0 \end{pmatrix}.$$

Then, method (4.1)-(4.2) can be written in the form

$$W^{n+1} = (I - \tfrac{1}{2}\tau\, B^{n+1})^{-1}\ (I + \tfrac{1}{2}\tau\, A^{n+1/2})\ (I - \tfrac{1}{2}\tau\, A^{n+1/2})^{-1}\ (I + \tfrac{1}{2}\tau\, B^n)\, W^n\,. \tag{5.1}$$

Let us now define the amplification matrix

$$C^n = (I - \tfrac{1}{2}\tau\, B^{n+1})^{-1}\ (I + \tfrac{1}{2}\tau\, A^{n+1/2})\ (I - \tfrac{1}{2}\tau\, A^{n+1/2})^{-1}\ (I + \tfrac{1}{2}\tau\, B^n)\,.$$

In order to guarantee that method (5.1) is stable we have to require $\| \prod_{i=0}^{n-1} C^i \|$ to remain uniformly bounded for all values of n and τ, such that

$$\| \prod_{i=0}^{n-1} C^i \| < K, \quad \text{for } 0 < n\tau < T \text{ and } K \text{ constant [9].} \tag{5.2}$$

Let us now verify this condition for the numerical method (5.1). We have that

$$\| \prod_{i=0}^{n-1} C^i \| \le \| (I - \tfrac{1}{2}\tau B^n)^{-1} \| \; \| (I + \tfrac{1}{2}\tau A^{n-1/2}) (I - \tfrac{1}{2}\tau A^{n-1/2})^{-1} \| \prod_{i=0}^{n-2} \| (I + \tfrac{1}{2}\tau B^{i+1}) (I - \tfrac{1}{2}\tau B^{i+1})^{-1} \; (I + \tfrac{1}{2}\tau A^{i+1/2}) (I - \tfrac{1}{2}\tau A^{i+1/2})^{-1} \| \; \| I + \tfrac{1}{2}\tau B^0 \| .$$

It can be verified that both the matrix A^n and the matrix B^n have their eigenvalues in the left half plane. The eigenvalues of the matrix A^n are even real nonpositive. Therefore, we have that

$$\| (I + \tfrac{1}{2}\tau B^{i+1}) (I - \tfrac{1}{2}\tau B^{i+1})^{-1} \| \le 1 \quad \text{and} \quad \| (I + \tfrac{1}{2}\tau A^{i+1/2}) (I - \tfrac{1}{2}\tau A^{i+1/2})^{-1} \| \le 1 , \quad \text{for } i=0,\ldots,n-1 .$$

Using these relations, we obtain that

$$\| \prod_{i=0}^{n-1} C^i \| \le \| (I - \tfrac{1}{2}\tau B^n)^{-1} (I + \tfrac{1}{2}\tau B^0) \| .$$

It is evident that only the explicit part $(I + \frac{1}{2}\tau B^0)$ may cause problems. In general, it is not possible to find an upper bound for $\| (I + \frac{1}{2}\tau B^0) \|$. In such a situation we may stabilize our integration method by computing the first approximation $\mathbf{W}^1$ by the backward Euler-LOD method, and apply method (5.1) for $n \ge 1$. This technique has been proposed in [14]. We thus consider the method with first time step

$$\mathbf{W}^1 = (I - \tfrac{1}{2}\tau B^1)^{-1} (I - \tfrac{1}{2}\tau A^{1/2})^{-1} \mathbf{W}^0 , \tag{5.3a}$$

and for $n \ge 1$

$$\mathbf{W}^{n+1} = (I - \tfrac{1}{2}\tau B^{n+1})^{-1} \; (I + \tfrac{1}{2}\tau A^{n+1/2}) \; (I - \tfrac{1}{2}\tau A^{n+1/2})^{-1} \; (I + \tfrac{1}{2}\tau B^n) \, \mathbf{W}^n . \tag{5.3b}$$

On a fixed space grid the LOD method (5.3a) is only first order accurate in time, but since we only perform one LOD step, method (5.3) is still second-order accurate on fixed space grids. Using method (5.3), we obtain that

$$\| \prod_{i=0}^{n-1} C^i \| \le \| (I - \tfrac{1}{2}\tau B^n)^{-1} \| \quad \| (I + \tfrac{1}{2}\tau A^{1/2})^{-1} \| .$$

Now, condition (5.2) is satisfied. We have no practical experience with method (5.3). In the numerical experiments no large errors were found in the original method (5.1), also for very large time steps. So, there was no need for stabilization. In [14] the authors advise to a use one or more LOD steps in situations where the initial values contain large errors. This might occur when experimental data with significant errors are used as initial values.

6. Solving the linear systems

In this section we describe how the linear systems at both stages, viz., system (4.3) and system (4.7), are solved. At the first stage we apply the Gaussian Elimination (double sweep) method for the solution of the tridiagonal systems. Since this is a recursive method, it seems to be unattractive on vector and parallel computers. However, we make use of the fact that a large number of tridiagonal systems of the same dimension have to be solved. In [4], the computational efficiency of this approach has been demonstrated on vector and parallel computers. Moreover, this method requires a minimal number of operations.

At the second stage we have to solve the linear, symmetric system (4.7). In literature, a large number of iterative methods have been proposed for such systems. Here, we will apply a Jacobi-type method and a Conjugate Gradient (CG) method. Both methods will be accelerated by a preconditioning technique. Before discussing the iteration methods, we first consider the preconditioning.

The essence of preconditioning is the determination of a matrix S such that the system

$$SA\,\mathbf{Z}^{q+1} = SB\,,$$

has a much smaller condition number than the original system $A\mathbf{Z}^{q+1}=\mathbf{B}$. For the preconditioning of system (4.7) we will use a smoothing matrix S of the form S=P(D) where P(z) is a polynomial and D is some matrix. The matrix D will be a difference matrix of which the eigenvalues are assumed to be in the interval [-1,0] (see [7]). The polynomial P(z) will be chosen such that P(0)=1 and the eigenvalues of S are in the interval [0,1]. Firstly, we discuss the choice of the matrix D. In our *theoretical* considerations we assume that D is equal to the normalized matrix A, i.e.,

$$D = \frac{A}{\rho(A)}\,, \tag{6.1}$$

where $\rho(\,)$ denotes the spectral radius. We emphasize that in *practice*, it is generally not attractive to choose D according to (6.1), and we shall employ some cheap approximation to the normalized matrix. If D is defined according to (6.1), then the eigenvalues of SA=P(D)A are given by $\rho(A)zP(z)$, where z runs through the spectrum of D. Now we are looking for a polynomial such that the magnitude of zP(z) on [-1,0] is sufficiently small. In this paper, we choose the polynomial [16]

$$P_{2^q-1}(z) = \prod_{k=1}^{q} \left(1 + \mu \frac{T_{2k-1}(1+2z) - 1}{2} \right)\,,\quad T_p(x) = \cos(p \arccos(x))\,. \tag{6.2}$$

In [7] it has been proved that the spectral radius of the matrix SA is minimized by the polynomial (6.2) when the matrix A has real nonpositive eigenvalues and $\mu=1$. For $\mu=1$ we have that (6.2) is equal to

$$P_{2^q-1}(z) = \frac{T_{2^q}(1+2z) - 1}{2z}\,\frac{1}{4^q}\,. \tag{6.3}$$

The factorization in (6.2) makes it possible to implement the smoothing operator S in a very efficient way. This is stated in the following theorem:

Theorem 6.1. *Let* S=P(D) *with* P(z) *defined by* (6.2) *and let the factor matrices* F_j *be defined by*

$$F_j = I + \mu D_j, \quad \textit{where} \quad D_{j+1} = 4\, D_j\, (I + D_j) \quad \textit{with} \quad D_1 = D \quad \textit{and} \quad j \geq 1 . \tag{6.4}$$

Then, $S = P_2{}^q{}_{-1}(D)$ *can be factorized according to*

$$S = F_q F_{q-1} \cdots F_1 \cdot \; [] \tag{6.5}$$

Thus, the smoothing matrix S consists of q factor matrices. For a proof of Theorem 6.1 we refer to [16].

The most efficient implementation of S is based on the factorization property in Theorem 6.1. However, in two or more dimensions the precomputation of the factor matrices F_j defined by (6.4) is not attractive. Therefore, we consider an alternative smoothing matrix S which only consists of one-dimensional operators. For our two-dimensional problem (4.7) we apply one-dimensional smoothing in x- and y-direction, successively. An extra advantage of the splitting in one-dimensional operators is that the application of the smoothing operator S is now hardly complicated when the domain is irregular [5].

As mentioned before, in practice we shall choose D equal to some cheap approximation of (6.1). Since the smoothing matrix S consists of one-dimensional operators, we choose

$$D = (1/4)\begin{pmatrix} -1 & 1 & & & 0 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & -2 \;\; 1 \\ 0 & & & 1 & -1 \end{pmatrix}. \tag{6.6}$$

If the factor matrices of (6.5) are computed in advance, then the evaluation of P(D) only requires q matrix-vector operations. Moreover, the factor matrices exhibit a regular pattern which can be exploited for an efficient implementation. E.g., applying Theorem 6.1 for matrix D in (6.6), we find the factor matrices

$$F_1 = (1/4)\begin{pmatrix} 3 & 1 & & & 0 \\ 1 & 2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & 2 \;\; 1 \\ 0 & & & 1 & 3 \end{pmatrix}, \qquad F_2 = (1/4)\begin{pmatrix} 2 & 1 & 1 & & & & 0 \\ 1 & 2 & 0 & 1 & & & \\ 1 & 0 & 2 & 0 & 1 & & \\ & 1 & 0 & 2 & 0 & 1 & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \text{ etc.} \tag{6.7}$$

Evidently, the matrix-vector multiplications with these essentially three-diagonal factor matrices is extremely cheap, especially on vector computers.

We shall now discuss the application of the Jacobi-type iteration method and the CG-type iteration method to system (4.7).

6.1. The smoothed Jacobi method

For the solution of system (4.7), we apply the smoothed Jacobi method [7]

$$\mathbf{Z}_{k+1} = \mathbf{Z}_k + \omega S \{ \mathbf{B} - A\, \mathbf{Z}_k \} , \qquad k=1,2,3,..., \tag{6.8}$$

where Z_k is the kth iterate, ω is a relaxation parameter and S is the smoothing matrix described in Theorem 6.1 with D as in (6.6). For the smoothed Jacobi method we choose $\mu=1$ (see (6.4)). As mentioned in the previous section, the smoothing matrix S consists of q smoothing factors. In [7] it has been demonstrated that one should not iterate with a fixed value of q. Therefore, we choose the number of smoothing factors at the kth iteration step equal to k modulo(q+1), which yields the cyclic sequence of 1,2,...,q,0,1,2,...,q,0,1,2,...,q,.. smoothing factors.
Let us now examine how the relaxation parameter ω should be chosen. For the spectral radius of A we have

$$\rho(A) \approx 1 + gH_{max}\left\{\frac{\tau^2}{(\Delta x)^2}+\frac{\tau^2}{(\Delta y)^2}\right\}, \quad \text{with } H_{max} = \max_{\substack{1\le i\le nx\\ 1\le j\le ny}} \{H_{i,j}\}.$$

Similarly, for the spectral radius of SA we have that

$$\rho(SA) \approx 1 + \frac{1}{4^q} gH_{max}\left\{\frac{\tau^2}{(\Delta x)^2}+\frac{\tau^2}{(\Delta y)^2}\right\}.$$

Following the analysis in [7], we obtain

$$\omega = \frac{2}{\rho(SA)}. \tag{6.9}$$

However, in our case, we do not choose ω fixed for each component $Z_{i,j}$. We make ω dependent on the local depth, viz.,

$$\omega_{i,j} = \frac{2}{1 + \frac{1}{4^q} gH_{i,j}\left\{\frac{\tau^2}{(\Delta x)^2}+\frac{\tau^2}{(\Delta y)^2}\right\}}. \tag{6.10}$$

In the case of a fixed relaxation parameter, we observed in our experiments that (6.9) was the optimum relaxation parameter. However, we obtained much better results with the relaxation parameter in (6.10) when an irregular bottom topography was used.

6.2. The smoothed CG method

The second iteration method that we applied for the solution of system (4.7), is a preconditioned CG method. The preconditioned CG method can be formulated as follows:

Let $\mathbf{Z}_0$ be an initial guess for $\mathbf{Z}^{(q+1)}$ and
$\mathbf{R}_0 = \mathbf{B} - A\mathbf{Z}_0$, $\mathbf{P}_0 = S\mathbf{R}_0$
For k=0,1,2,...., until convergence

$$\begin{aligned}
\alpha_k &= \frac{\mathbf{R}_k^T(S\mathbf{R}_k)}{\mathbf{P}_k^T(A\mathbf{P}_k)}\\
\mathbf{Z}_{k+1} &= \mathbf{Z}_k + \alpha_k \mathbf{P}_k\\
\mathbf{R}_{k+1} &= \mathbf{R}_k - \alpha_k A\mathbf{P}_k\\
\beta_k &= \frac{\mathbf{R}_{k+1}^T(S\mathbf{R}_{k+1})}{\mathbf{R}_k^T(S\mathbf{R}_k)}\\
\mathbf{P}_{k+1} &= S\mathbf{R}_{k+1} + \beta_k \mathbf{P}_k
\end{aligned} \tag{6.11}$$

In (6.11) the matrix S denotes the preconditioning matrix. It is well-known that the unpreconditioned CG method can be implemented efficiently on vector and parallel computers, but in general the preconditioned version is much more troublesome. In literature, various techniques for the construction of a suitable preconditioning matrix have been proposed (see [11] for a survey). Here, we choose a positive definite matrix S of the form S=P(D), where D is the difference matrix in (6.6) and P(z) the polynomial (6.2). By choosing $\mu \in [0,1)$ we obtain that S is positive definite. This preconditioning matrix can be implemented efficiently on vector and parallel computers, because only matrix times vector operations are involved. The convergence is improved by this preconditioning matrix since the condition number of SA is much smaller than that of A. It should be noted that this preconditioning matrix S is independent of A, whereas in general the preconditioning matrix S is some approximation to the inverse of A.

7. Numerical experiments

In this section we illustrate for a number of test problems the accuracy and the computational aspects of the time integration method (4.1)-(4.2). In the test problems the water is initially at rest and the motion in the closed basin is generated by a periodic wind stress. Thus, a wind driven circulation is gradually developed.
The following parameter values are used in all experiments:

$f = 0.44/3600 = 1.22\ e\text{-}4$
$g = 9.81\ [m/s^2]$
$N = 0.065/\rho$
$\rho = 1025\ [kg/m^3]$
$\varphi = 45$ [deg.] (north-eastern wind).

The experiments have been carried out on an ALLIANT FX/4. This mini-supercomputer consists of four vector processors. In all experiments we have used both the vector optimization and the parallel optimization.
At the end of the integration process the numerical solution was compared with a reference solution computed on the same grid with τ=60 seconds. The reference solution may be considered as an almost exact solution of our semi-discretized system (3.1). So, the accuracy results listed in this section represent the error due to the time integration.
In the experiments we used a rectangular basin of 400 by 800 km with different bottom topographies. For the grid sizes we have chosen Δx=10 km, Δy=10 km and $\Delta\sigma$=0.2 m. Thus, the computations have been performed on a grid with nx=41, ny=81 and ns=5. We integrated over a period of five days with the time-dependent wind stress

$$W_f = 1.5 * \left(1 + 0.5 * \sin \frac{2\pi t}{24 * 3600} \right). \qquad (7.1)$$

Thus, we have a periodically varying (north-eastern) wind with a period of 24 hours. To measure the obtained accuracy, we define

ERR-x : maximal global error for component x, with $x=\zeta$, u, or v,
compared with the reference solution at the end point T = 120 hours. (7.2)

In the first experiment we have a plane bottom with a depth of 65 m., except for a deeper channel in a diagonal direction (depth 85 m.). This is shown in Figure 2. In the second experiment we use a basin with an inclined bottom of a depth of 20 m. at one end and 340 m. at the other end (see Fig. 3).

In Table 7.1 we list the maximal global errors for the test problem with a channel in a diagonal direction.

Table 7.1. Test problem with a channel in a diagonal direction.

τ (sec.)	ERR-ζ (m.)	ERR-u (m/s.)	ERR-v (m/s.)
600	0.038	0.006	0.018
1200	0.087	0.014	0.041
2400	0.212	0.035	0.103
4800	0.549	0.088	0.269

In this experiment the maximal values for ζ, u and v are about 2.6 m, 0.4 m/s. and 1.1 m/s., respectively. We observed that after a few days the solution became periodic with a period of 24 hours for any time step τ. For the largest time steps the accuracy results seem to be unacceptable. However, a careful examination of the integration process shows that, even in the case of large time steps, the maximal and minimal values of the periodic numerical solution are very close to the extreme values of the reference solution. The differences are in the size of a few centimetres. Thus, our integration method hardly introduces a dissipation error. However, for the large time steps, errors in the phase of the periodic solution appear. E.g., in the case of τ=4800 sec. the phase error is about one hour. When we compare the numerical solution computed with τ=4800 sec. at T=121.33 hours with the reference solution at T=120 hours, the maximal global errors for the three components are 0.124 m, 0.050 m/s. and 0.054 m/s., respectively. This is significantly less than in Table 7.1.

We now discuss the computational efficiency of the time integration method (4.1)-(4.2). To represent the results, we use the following notation:

q	: number of smoothing factors (see (6.5))
μ	: smoothing coefficient (see (6.2))
TOTAL	: total computation time
ITER	: computation time for the iteration process
PREC	: computation time for the preconditioning
#ITER	: number of iterations averaged over the integration steps
CONV	: convergence factor averaged over the integration steps

At each integration step the convergence factor is defined by $(r(k))^{1/k}$, where k is the smallest value for which the residue (cf. (4.6))

$$r(k) = \| B_z^{n+1/2} - AZ_k \|_\infty$$

drops below a certain tolerance. In the experiments we required that $r(k) \leq 10^{-3}$.

In Table 7.2 we list the computation times and the convergence results for the time integration method (4.1)-(4.2) in which either the smoothed Jacobi (SJAC) method or the smoothed CG (SCG) method has been applied. For both iteration methods we vary the number of smoothing factors. The case q=0 corresponds to the unpreconditioned case. For the parameter μ in the preconditioning matrix of the SCG method, we experimentally derived an optimum value. As mentioned in Section 6.1, for the SJAC method we have $\mu=1$.

Table 7.2. Computation times for the channel problem.

τ (sec.)	method	q	μ	TOTAL (sec.)	ITER (sec.)	PREC (sec.)	#ITER	CONV
600	SCG	0		395.6	70.3	0.	3.0	0.23
	SJAC	1		473.0	148.8	21.5	8.4	0.56
1200	SCG	0		233.5	69.8	0.	9.2	0.60
	SJAC	2		318.2	154.7	47.7	17.0	0.72
2400	SCG	0		169.5	87.5	0.	26.0	0.81
		1	0.9	156.8	75.1	25.4	14.1	0.68
		2	0.75	161.3	79.5	37.9	11.4	0.61
	SJAC	3		220.1	137.7	50.9	26.3	0.81
4800	SCG	0		155.3	114.7	0.	75.4	0.91
		1	0.925	118.9	77.9	28.2	31.2	0.80
		2	0.85	109.9	69.5	35.8	21.9	0.74
	SJAC	0		827.1	786.7	0.	869.6	0.99
		1		517.3	477.3	100.2	297.2	0.98
		2		279.0	238.2	75.0	115.8	0.94
		3		193.1	151.8	64.1	55.0	0.88
		4		184.2	143.8	62.1	48.6	0.86

In the case of a time step of 4800 sec., we have listed the results for various values of q. When no preconditioning is applied, the Jacobi method converges extremely slow in this case. However, by applying four smoothing factors, the number of iterations is reduced by a factor 18, whereas the computation time for the iteration process is reduced by a factor 5.5.

When no preconditioning is applied, the CG method has a much better convergence behaviour than the Jacobi method. For the CG method it is even better to apply no preconditioning in the case of small time steps, since the number of iterations is already very limited. However, for large time steps both the number of iterations and the computation time is reduced when the preconditioning is applied.

In Table 7.2 we have listed the optimum values for μ. For values in the neighbourhood of the optimum value the number of iterations hardly increases. So, the choice of the parameter μ in the preconditioning matrix S of the SCG method is not critical. In this experiment the SCG method requires less computation time than the SJAC method.

In Table 7.3 we list the maximal global errors for the test problem with the inclined bottom (see Fig. 3). In this experiment the maximal values for ζ, u and v are about 1.2 m, 0.7 m/s. and 1.4 m/s., respectively. The results are comparable with the results in the first experiment. After a few days the numerical solution also became periodic with a period of 24 hours for any time step τ. However, in this experiment the phase errors are much smaller.

Table 7.3. Test problem with an inclined bottom.

τ (sec.)	ERR-ζ (m.)	ERR-u (m/s.)	ERR-v (m/s.)
600	0.016	0.005	0.003
1200	0.018	0.008	0.005
1800	0.035	0.012	0.009
3600	0.122	0.032	0.034

The computational results in this experiment, which are listed in Table 7.4, are also comparable with the results of the first experiment. Both the number of iterations and the computation time for the iteration process is reduced when the preconditioning is applied. As in the first experiment the SCG method requires less computation time than the SJAC method.

Table 7.4. Computation times for the problem with an inclined bottom.

τ (sec.)	method	q	μ	TOTAL (sec.)	ITER (sec.)	PREC (sec.)	#ITER	CONV
600	SCG	0		432.2	108.1	0.	7.4	0.49
	SJAC	1		570.4	241.2	41.4	15.8	0.68
1200	SCG	0		326.6	163.9	0.	24.4	0.75
	SJAC	2		413.5	248.3	71.9	28.2	0.77
1800	SCG	0		271.5	160.0	0.	39.7	0.84
		1	0.85	253.7	143.9	48.1	22.7	0.73
	SJAC	3		357.5	247.9	100.3	38.4	0.81
3600	SCG	0		276.1	221.3	0.	109.9	0.92
		1	0.9	193.8	141.2	51.6	47.3	0.82
		2	0.8	240.2	186.4	91.9	45.0	0.82
	SJAC	4		257.5	203.0	91.4	48.6	0.82

In the experiments we used both the vector and the parallel optimization of the ALLIANT FX/4. For both iteration methods the computation time was reduced by about a factor 3 due to the vectorization, and also by a factor 3 due to the parallel optimization. However, not only the computation time for both iteration methods, but also the computation time for our integration method (4.1)-(4.2) was reduced by the above-mentioned factors. This shows that our integration method (4.1)-(4.2), in which either the SCG method or the SJAC method has been applied, can be implemented efficiently on vector and parallel computers.

We now carry out an experiment in which we vary the number of layers in the vertical direction. Our aim is to illustrate the efficiency of the time integration method (4.1)-(4.2) for three-dimensional shallow water problems. We have chosen the bottom topography of the first experiment (i.e., a plane bottom with a deeper diagonal channel) and a time step of 4800 sec. The SCG method is used with q=1 and μ=0.925 (cf. Table 7.2). Table 7.5 presents the computation times and the convergence results for a different number of grid layers in the vertical direction. The number of vertical grid layers is denoted by ns.

Table 7.5. Computation times for a different number of vertical layers.

ns	TOTAL (sec.)	ITER (sec.)	PREC (sec.)	#ITER	CONV
1	84.7	76.5	26.0	28.5	0.79
2	94.2	77.1	27.2	30.2	0.80
5	118.9	77.9	28.2	31.2	0.80
10	160.1	80.5	28.5	31.6	0.80
25	278.5	81.3	29.0	31.9	0.80

Since the system that we have to solve at the second stage, is of the same computational complexity for both two-dimensional and three-dimensional problems, the results in the last four columns are more or less constant. So, the computation time required for the solution of this system is independent of the number of vertical grid layers. In the two-dimensional case (i.e., ns=1) a major part of the computation time is required for the solution of the system at the second stage (about 90%). However, for three-dimensional experiments the computation time for the solution of the system at the second stage becomes relatively less. For example, in the case of ten vertical grid layers, about half the computation time is required for the solution of this system. This percentage depends on the time step used. In this experiment we used a rather large time step. For smaller time steps the percentage of computation time required for the solution of the system, is significantly less. In conclusion, the time integration method (4.1)-(4.2) is very suited for three-dimensional problems, especially when large time steps are used.

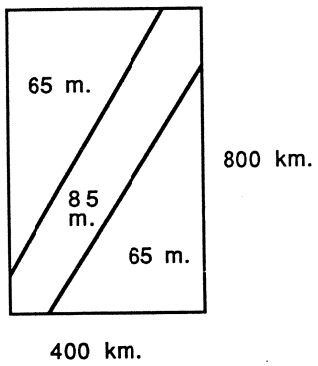

Figure 2.
plane bottom with a channel

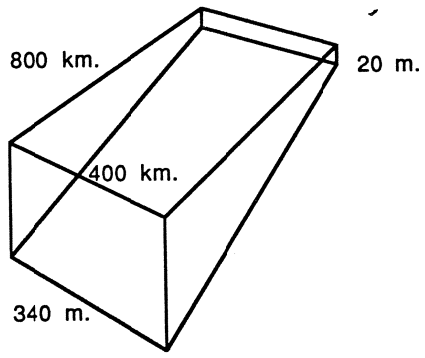

Figure 3.
inclined bottom

8. Conclusions

In this paper we have presented a two-stage time splitting method for the three-dimensional shallow water equations. The method has been developed in such a way that the stability of this method neither depends on the vertical diffusion term nor on the terms describing the propagation of the surface waves. At the first stage a large number of tridiagonal systems of the same dimension have to be solved. At the second stage the system to be solved is symmetric, five-diagonal and positive definite. For the solution of the latter system we have developed a smoothed Jacobi (SJAC) method and a smoothed CG (SCG) method. Both methods have been accelerated by a technique based on smoothing. The smoothing matrices have been chosen in such a way that the number of iterations was moderate in all cases. Moreover, the smoothing matrices can be implemented efficiently on vector and parallel computers, because only matrix times vector operations are involved. It should be noted that the smoothing matrices for the CG method are independent of the system to be solved. In the experiments the SCG method required less computation time than the SJAC method.

It has been shown that the time integration method presented in this paper is very suited for three-dimensional problems. When we apply our method to two-dimensional problems, the system to be solved at the second stage is the most time consuming part. In three-dimensional models the same amount of computation time is required, because this system is independent of the number of grid layers in the vertical direction. The computation time for the other parts of the method is proportional to the number of vertical grid layers. Therefore, the time splitting method is relatively more efficient for three-dimensional problems than for two-dimensional problems. In [15] it was reported that a time-splitting method of this form is already feasible for practical computations of two-dimensional problems.

Finally, the method is mass conservative and can be implemented efficiently on vector and parallel computers.

9. References

[1] J.O. Backhaus, A semi-implicit scheme for the shallow water equations for application to shelf sea modelling, *Continental Shelf Research*, **2**, 243-254 (1983).

[2] V. Casulli, Semi-implicit finite difference methods for the two dimensional shallow water equations, *J. Comp. Phys.*, **86**, 56-74 (1990).

[3] A.M. Davies, Application of the DuFort-Frankel and Saul'ev methods with time splitting to the formulation of a three dimensional hydrodynamic sea model, *Int. J. Numer. Methods Fluids*, **5**, 405-425 (1985).

[4] E.D. de Goede, A computational model for the three-dimensional shallow water flows on the ALLIANT FX/4, *Supercomputer*, **32**, 43-49 (1988).

[5] E.D. de Goede, Stabilization of a time integrator for the 3D shallow water equations by smoothing techniques, *Int. J. Numer. Methods Fluids*, to appear.

[6] A.R. Gourlay, Splitting methods for time dependent partial differential equations, The State of the Art in Numerical Analysis, D. Jacobs (ed.), Academic press, London - New York - San Francisco, 757-791 (1977).

[7] P.J. van der Houwen, C. Boon and F.W. Wubs, Analysis of smoothing matrices for the preconditioning of elliptic difference equations, *Z. Angew. Math. Mech.*, **68**, 3-10 (1988).

[8] Y. Kurihara, On the use of implicit and iterative methods for the time integration of the wave equation, *Month. Weather Rev.*, **93**, 33-46 (1965).

[9] P.D. Lax and R.D. Richtmyer, Survey of the stability of linear finite difference equations, *Comm. Pure Appl. Math.*, **17**, 267-293 (1956).

[10] J.J. Leendertse, Aspects of a computational model for long period water wave propagation, *Memorandum RM-5294-PR*, Rand Corp., Santa Monica, California, 1967.

[11] J. Ortega and R. Voigt, Solution of partial differential equations on vector and parallel computers, *SIAM Review*, **27**, 149-240 (1985).

[12] G.S. Stelling, On the construction of computational methods for shallow water flow problems, *Ph.D. Thesis*, Delft University, 1983.

[13] G.S. Stelling, A.K. Wiersma and J.B.T.M. Willemse, Practical aspects of accurate tidal computations, *J. Hydr. Eng., ASCE*, **112**, 802-817 (1986).

[14] J. Verwer and W.H. Hundsdorfer, Stability and convergence of the Peaceman-Rachford ADI method for initial-boundary problems, *Math. Comp.*, **53**, 81-101 (1989).

[15] P. Wilders, Th.L.van Stijn, G.S. Stelling and G.A. Fokkema, A fully implicit splitting method for accurate tidal computations, *Int. J. Numer. Methods Eng.*, **26**, 2707-2721 (1988).

[16] F.W. Wubs, Numerical solution of the shallow-water equations, *Ph.D. Thesis*, University of Amsterdam, Amsterdam, 1987.