

FINITE ELEMENT SOLUTION OF THE SHALLOW WATER EQUATIONS BY A QUASI-DIRECT DECOMPOSITION PROCEDURE

J. GOUSSEBAILE, F. HECHT, G. LABADIE AND L. REINHART

INRIA, Domaine de Voluceau, Rocquencourt, B P 105, 78153 Le Chesnay Cedex, France

SUMMARY

We describe in this paper some new methods for solving the shallow water flow problem; this problem is composed of two coupled non-linear equations. First a fractional step method is used to decouple the difficulties due to the convection and the propagation. The diffusion propagation step amounts to solving two coupled linear equations on the water depth and the fluxes. We used then some quasi-direct decomposition techniques to decouple the unknowns, leading to symmetric systems only. In addition some non-trivial boundary conditions on the flux are examined. Particularly we are interested to get decoupled systems on the two components of the flux. Numerical tests are presented to enlighten the behaviour and the capabilities of the method. Finally an industrial example is treated on the Dunkirk harbour.

1. INTRODUCTION

Mathematical models of shallow water flows have various applications concerning the transport processes in harbours and estuaries: for example the thermic impact of the construction of a central power station on the seaside or the estimate of the effects of a seaside construction on the sedimentology to optimize the shape of the harbour. Let us now describe these equations.

1.1. *Shallow water equations*

The shallow water equations are derived from the three dimensional Navier–Stokes problem by integrating over the water depth and assuming hydrostatic pressure. They are described by the momentum and continuity equations.

Let $\mathbf{u} = (u_i)_{i=1,2}$ be the stream velocity, h the depth of water and $\mathbf{Q} = (\mathbf{u}h)$ the flux per unit length; h , \mathbf{Q} and \mathbf{u} are solutions of

$$\frac{\partial h}{\partial t} + \operatorname{div} \mathbf{Q} = 0, \quad (\text{conservation law}) \quad (1)$$

$$\frac{\partial \mathbf{Q}}{\partial t} + \nabla \mathbf{Q} \cdot \mathbf{u} + gh \nabla h - K \Delta \mathbf{Q} = \mathbf{f}, \quad (\text{momentum equations}) \quad (2)$$

where g is the gravity acceleration, K is the eddy viscosity, and \mathbf{f} takes into account the Coriolis force and other forces.

The unknowns \mathbf{u} , \mathbf{Q} and h are defined on an open set Ω in the \mathbb{R}^2 space and for a time interval $[0, T]$. The boundary conditions can be of several types among which are:

Dirichlet conditions

$$\mathbf{Q} = \mathbf{Q}_0 \text{ given on the boundary } \Gamma \text{ or a part } \Gamma_0 \text{ of } \Gamma \tag{3}$$

Friction condition

$$\left. \begin{aligned} \mathbf{Q} \cdot \mathbf{n} &= 0 \quad (\text{impermeability condition}) \\ \lambda \mathbf{Q} \cdot \boldsymbol{\tau} + K \frac{\partial \mathbf{Q}}{\partial n} \cdot \boldsymbol{\tau} &\text{ given} \end{aligned} \right\} \tag{4}$$

\mathbf{n} being the outward unit normal vector and $\boldsymbol{\tau}$ a unitary vector tangential to the boundary; this condition (4) is similar to a wall law for the Navier–Stokes case, the quantity $(\partial \mathbf{Q} / \partial n) \cdot \boldsymbol{\tau}$ being the tangential friction force per unit length.¹

Wave condition

$$c\mathbf{Q} + K \frac{\partial \mathbf{Q}}{\partial n} - c^2 z \mathbf{n} \text{ given} \tag{5}$$

where $c^2 = gh$ and $z = z_F + h$, z_F being the bottom depth. This condition simulates the propagation of a wave through sea boundaries and is called the incident wave condition.²

1.2. A frictional step algorithm

Equations (1) and (2) couple the difficulties due to the convection term and the propagation problem. To overcome these difficulties we use a frictional step method together with a quasi-implicit discretization in time.

Let Δt be the time step, \mathbf{Q}^n , \mathbf{u}^n and h^n denote approximations of \mathbf{Q} , \mathbf{u} and h , respectively, at time $n\Delta t$; the chosen time discretization is as follows:

$$\frac{\hat{Q}_i^n - Q_i^n}{\Delta t} + \nabla \cdot (\hat{Q}_i^n \mathbf{u}^n) = 0, \quad \text{for } i = 1, 2 \text{ (advection step)} \tag{6}$$

$$\left. \begin{aligned} \frac{h^{n+1} - h^n}{\Delta t} + \nabla \cdot \mathbf{Q}^{n+1} &= 0 \\ \frac{Q_i^{n+1} - \hat{Q}_i^n}{\Delta t} - K \Delta Q_i^{n+1} + g\bar{h} \frac{\partial h^{n+1}}{\partial x_i} &= f_i + g(\bar{h} - h^n) \frac{\partial h^n}{\partial x_i}, \quad \text{for } i = 1, 2 \end{aligned} \right\} \tag{7}$$

where \bar{h} is a function of the spatial co-ordinates only and is usually taken equal to the mean value of h in time (given *a priori*).

The diffusion propagation step (7) is detailed in the second paragraph. The main advantage of the fractional step method is to lead to symmetric systems and to decouple the difficulties.

1.3. Solution of the advection step (6)

The auxilliary flux \hat{Q}_i is the solution of the discrete convection equation (6). We propose a method of characteristics to solve this equation. This method has been already used for the pure transport equation^{3,4} and for the Navier–Stokes equation.⁵

This method leads to ordinary differential equations, and their discrete solutions can be carried out by an explicit Runge–Kutta method of order 2 or 4. For details on this method see Reference 6.

In fact, the divergence of \mathbf{u}^n not being equal to zero, equation (6) can be replaced by

$$h \frac{\partial u_i}{\partial t} + u_i \frac{\partial h}{\partial t} + hu_i \nabla \cdot \mathbf{u} + \mathbf{u} \cdot (h \nabla u_i + u_i \nabla h) = 0 \tag{8}$$

and using the continuity equation and dividing by h :

$$\frac{\partial \hat{u}_i}{\partial t} + \mathbf{u} \cdot \nabla \hat{u}_i = 0, \quad \text{for } i = 1, 2 \tag{9}$$

The solution \hat{u}_i is constant along the characteristic lines and the auxiliary flux \hat{Q}_i is obtained by multiplying the solution \hat{u}_i by h^n .

Although explicit, the method of characteristics is unconditionally stable, and the time step is not strictly limited by the Courant–Friedrichs–Levy criterion.

1.4. Possible choices on the time discretization scheme

In the previous paragraphs we have chosen an implicit discretization in time. Numerical experiments show that it seems better to use so called θ -schemes. We have studied these time discretization schemes from the phase error and the numerical damping points of view.

In one dimensional space, the propagation step reduces to

$$\left. \begin{aligned} \frac{\partial Q}{\partial t} + c^2 \frac{\partial h}{\partial x} - K \frac{\partial^2 Q}{\partial x^2} &= f \\ \frac{\partial h}{\partial t} + \frac{\partial Q}{\partial x} &= 0 \end{aligned} \right\} \tag{10}$$

with some initial conditions on h and Q .

For θ and θ' belonging to $[0, 1]$, the associated θ -scheme can be written as follows:

$$\left. \begin{aligned} \frac{Q^{n+1} - Q^n}{\Delta t} - K \frac{\partial^2}{\partial x^2} (\theta Q^{n+1} + (1 - \theta)Q^n) \\ + c^2 \frac{\partial}{\partial x} (\theta' h^{n+1} + (1 - \theta')h^n) &= 0 \\ \frac{h^{n+1} - h^n}{\Delta t} + \frac{\partial}{\partial x} (\theta Q^{n+1} + (1 - \theta)Q^n) &= 0 \end{aligned} \right\} \tag{11}$$

If $\theta = \theta' = 1$, the scheme is implicit, if $\theta = \theta' = 0$, we obtain the explicit scheme and for $\theta = \theta' = \frac{1}{2}$, the Crank–Nicholson scheme.

It is easy to eliminate Q in the second equation and to get the following system:

$$\left. \begin{aligned} \frac{1}{(\Delta t)^2} (h^{n+1} - h^n) - \frac{K}{\Delta t} \frac{\partial^2}{\partial x^2} (h^{n+1} - h^n) - c^2 \frac{\partial^2}{\partial x^2} (\theta' h^{n+1} + (1 - \theta')h^n) &= -\frac{1}{\Delta t \theta} \frac{\partial Q^n}{\partial x} \\ \frac{1}{\Delta t} (Q^{n+1} - Q^n) - K \frac{\partial^2}{\partial x^2} (\theta Q^{n+1} + (1 - \theta)Q^n) + gh \frac{\partial}{\partial x} (\theta' h^{n+1} + (1 - \theta')h^n) &= 0 \end{aligned} \right\} \tag{12}$$

A spectral method is then used to study the behaviour of the phase shift and the numerical damping.

The numerical tests we did show that the implicit scheme is very dissipative whereas the Crank–Nicholson scheme ($\theta = \theta' = 0.5$) spares the phase together with the amplitude of the wave. However

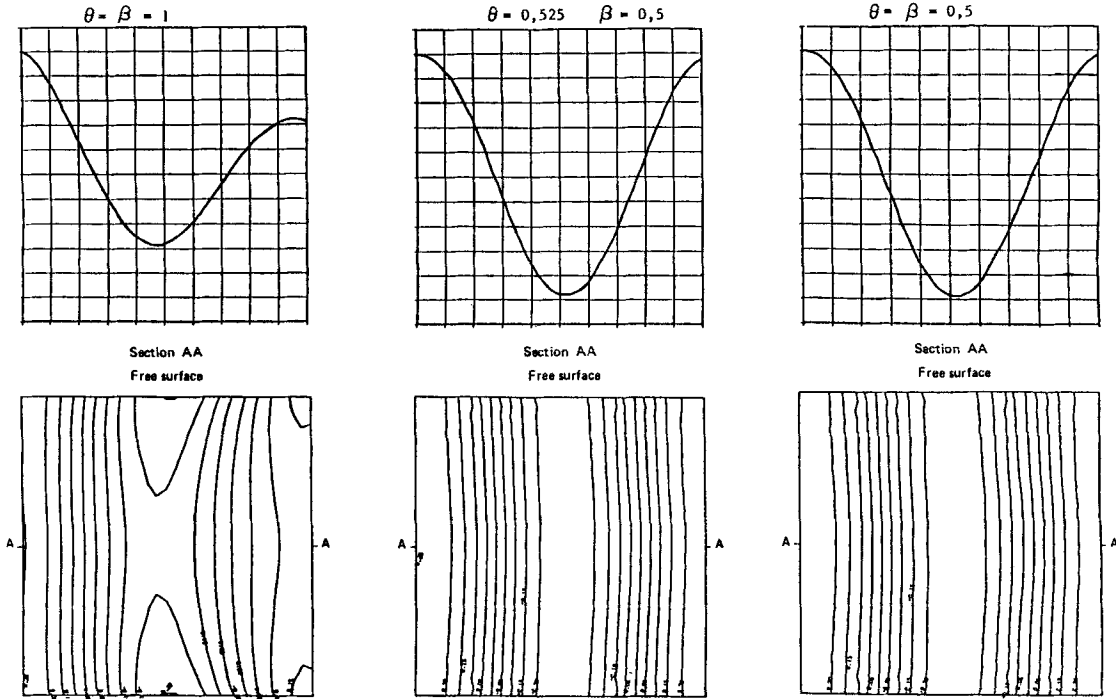


Figure 1. Propagation of a wave from left to right: 17 points of discretization per wave length

numerical instabilities can occur. The choice $\theta = 0.525$ improves the phase shift and maintains a correct amplitude of the wave.

The propagation of a wave has been simulated and the results are plotted in Figure 1 after a period. Seventeen discretization nodes per wavelength have been used.

2. SOLUTION OF THE PROPAGATION STEP

For the sake of simplicity we restrict ourselves to Dirichlet boundary conditions. Other conditions will be treated in the Section 3. The time discretized boundary value problem can be written as follows:

$$\alpha h + \nabla \cdot \mathbf{Q} = \alpha h^n, \quad \text{in } \Omega \tag{13}$$

$$\alpha \mathbf{Q} - K \Delta \mathbf{Q} + c^2 \nabla h = \alpha \mathbf{Q}^n + \mathbf{S}, \quad \text{in } \Omega \tag{14}$$

where $\alpha = 1/\Delta t$, $c^2 = g\bar{h}$ and \mathbf{S} takes into account the exterior forces and the explicit part of the terms; we add to (13), (14) the following boundary conditions:

$$\mathbf{Q} = \mathbf{Q}_0, \quad \text{given on the boundary } \Gamma \tag{15}$$

Although symmetric when the bottom is flat ($c^2 = \text{constant}$), this system after discretization in space leads to a large linear system. In industrial tests, the number of unknowns is too large to directly solve the system.

The more natural way to reduce the size of this system is to eliminate one of the unknowns h or \mathbf{Q} as a function of the other one. The first possibility is to eliminate h in (14) with the help of equation

(13). This method leads to a non-symmetric system; in addition the two components of the flux are coupled in Ω .

Applying the divergence operator to equation (14) and plugging equation (13) into it, we then get the equation on h only:

$$\alpha^2 h - \nabla \cdot [(K\alpha + c^2)\nabla h] = \alpha^2 h^n - K\alpha \nabla h^n - \nabla \cdot (\alpha \mathbf{Q}^n + \mathbf{S}^n), \quad \text{in } \Omega \tag{16}$$

$$= w^n$$

This equation is elliptic and leads to a linear symmetric system, but its solution requires the value of h on the boundary. On the other hand when the boundary conditions are known for h , the solution of (14)–(16) requires three Dirichlet problems (one for h , and one for each component of the flux).

2.1. *A quasi-direct decomposition method*

This method has been already used by Glowinski and Pironneau⁷ to solve the first biharmonic equation; we refer to this paper for more details.

The method is built on the observation that the continuity equation (13) with equations (14) and (16) is equivalent to a boundary condition on the solution of an elliptic problem.

If λ is a given function on the boundary, let h_λ , \mathbf{Q}_λ and ψ_λ be the solutions of the following elliptic problems, if $D = K\alpha + c^2$:

$$\left. \begin{aligned} \alpha^2 h_\lambda - \nabla \cdot (D\nabla h_\lambda) &= \alpha^2 h^n - K\alpha \Delta h^n - \nabla \cdot (\alpha \mathbf{Q}^n + \mathbf{S}^n), \quad \text{in } \Omega \\ h_{\lambda|\Gamma} &= \lambda \end{aligned} \right\} \tag{17}$$

$$\left. \begin{aligned} \alpha \mathbf{Q}_\lambda - K\Delta \mathbf{Q}_\lambda + c^2 \nabla h_\lambda &= \alpha \mathbf{Q}^n + \mathbf{S}^n, \quad \text{in } \Omega \\ \mathbf{Q}_\lambda &= \mathbf{Q}_0, \quad \text{on } \Gamma \end{aligned} \right\} \tag{18}$$

$$\left. \begin{aligned} \alpha^2 \psi_\lambda - \nabla \cdot (D\nabla \psi_\lambda) &= c^2(\alpha h_\lambda + \text{div } \mathbf{Q}_\lambda - \alpha h^n), \quad \text{in } \Omega \\ \psi_{\lambda|\Gamma} &= 0 \end{aligned} \right\} \tag{19}$$

Under the stability condition

$$\Delta t < (\|\nabla^t(c) \cdot \nabla(c)\|_{L^\infty})^{-1/2} \tag{20}$$

the continuity equation (18) is equivalent to the boundary relation

$$D \frac{\partial \psi_\lambda}{\partial n} = 0 \tag{21}$$

For the proof of this result see Reference 2.

Remark. The sufficient condition (20) shows that the time step has to be small enough when the gradient of the bottom step increases. This restrictive condition disappears when the bottom is flat.

The boundary operator which associates to any function λ on the boundary the quantity $D(\partial\psi_\lambda/\partial n)$ is clearly an affine operator; splitting up the linear part A and the constant part b of this operator, we then have

$$D \frac{\partial \psi_\lambda}{\partial n} = A\lambda - b$$

The method consists of constructing operator A and right hand side b and solving the linear boundary equation $A\lambda = b$, the result of which, λ , is the wanted boundary condition $h|_\Gamma$ for the equation (16). Operator A and right side b are constructed as follows.

For each function λ of the boundary, solve successively the three following problems:

$$\left. \begin{array}{l} \text{Find } h_\lambda \text{ in } H^1(\Omega) \text{ such that} \\ \int_{\Omega} (\alpha^2 h_\lambda \pi + D \nabla h_\lambda \nabla \pi) dx = 0, \forall \pi \in H_0^1(\Omega) \\ h_\lambda = \lambda \text{ on } \Gamma \end{array} \right\} \quad (22)$$

$$\left. \begin{array}{l} \text{Find } \mathbf{Q}_\lambda \text{ in } (H_0^1(\Omega))^2 \text{ such that} \\ \int_{\Omega} (\alpha \mathbf{Q}_\lambda \phi + K \nabla \mathbf{Q}_\lambda \nabla \phi) dx = - \int_{\Omega} c^2 \nabla h_\lambda \phi dx, \forall \phi \in (H_0^1(\Omega))^2 \end{array} \right\} \quad (23)$$

and

$$\left. \begin{array}{l} \text{Find } \psi_\lambda \in H_0^1(\Omega), \text{ solution of} \\ \int_{\Omega} (\alpha^2 \psi_\lambda \pi + D \nabla \psi_\lambda \nabla \pi) dx = \int_{\Omega} c^2 (\alpha h_\lambda + \operatorname{div} \mathbf{Q}_\lambda) \pi dx, \forall \pi \in H_0^1(\Omega) \end{array} \right\} \quad (24)$$

and define for each function μ in $H^{1/2}(\Gamma)$

$$\langle A\lambda, \mu \rangle = - \int_{\Gamma} D \frac{\partial \psi_\lambda}{\partial n} \mu d\Gamma \quad (25)$$

where $\langle \cdot, \cdot \rangle$ is the duality pairing between $H^{1/2}(\Gamma)$ and $H^{-1/2}(\Gamma)$.

Then, under the stability condition (20), the operator A is an isomorphism from $H^{1/2}(\Gamma)$ onto $H^{-1/2}(\Gamma)$. Furthermore, if the bottom is flat (i.e. $\bar{h} = \text{constant}$), the operator A is self-adjoint.

In a similar way, the right hand side b is obtained by solving the following set of Dirichlet problems:

$$\left. \begin{array}{l} \text{Find } h_0 \text{ in } H_0^1(\Omega), \text{ such that} \\ \int_{\Omega} (\alpha^2 h_0 \pi + D \nabla h_0 \nabla \pi) dx = \int_{\Omega} (\alpha^2 h^n - K \alpha \Delta h^n - \nabla \cdot (\alpha \hat{\mathbf{Q}}^n + \mathbf{S}^n)) \pi, \forall \pi \in H_0^1(\Omega) \end{array} \right\} \quad (26)$$

Find \mathbf{Q}_0 in $H^1(\Omega)$, solution of

$$\left. \int_{\Omega} (\alpha \mathbf{Q}_0 \phi + K \nabla \mathbf{Q}_0 \nabla \phi) dx = - \int_{\Omega} c^2 \nabla h_0 \phi dx + \int_{\Omega} (\alpha \mathbf{Q}^n + \mathbf{S}^n) \phi dx, \forall \phi \in (H_0^1(\Omega))^2 \right\} \quad (27)$$

\mathbf{Q}_0 given on Γ

$$\left. \begin{array}{l} \text{Find } \psi_0 \text{ in } H_0^1(\Omega), \text{ such that} \\ \int_{\Omega} (\alpha^2 \psi_0 \pi + D \nabla \psi_0 \nabla \pi) dx = \int_{\Omega} c^2 (\alpha h_0 - \alpha h^n + \operatorname{div} \mathbf{Q}_0) \pi dx, \forall \pi \in H_0^1(\Omega) \end{array} \right\} \quad (28)$$

and set

$$b = D \frac{\partial \psi_0}{\partial n}, \text{ on } \Gamma \quad (29)$$

It is quite obvious that, if $\tilde{\lambda}$ is the unique solution of the system

$$A\tilde{\lambda} = b$$

then $h = h_{\tilde{\lambda}} + h_0$ and $\mathbf{Q} = \mathbf{Q}_{\tilde{\lambda}} + \mathbf{Q}_0$ are the solutions of the weak formulation of (16) and (14). In

addition, the function ψ defined by

$$\psi = \psi_{\bar{\lambda}} + \psi_0$$

is the solution of (19) (via the solution of (17), (18)) and satisfies

$$\frac{\partial \psi}{\partial n} = 0$$

which ensures that the continuity equation is satisfied.

2.2. Approximation by a finite element method

In this section we shall consider only a polygonal domain Ω of \mathbb{R}^2 , but what follows can be easily extended to curved boundaries using isoparametric finite elements.⁸ We show in this section how to constrict the discrete linear equation to approximate the boundary equation $A\lambda = b$.

Let \mathcal{T}_h be a triangulation of Ω , satisfying usual uniformity assumptions. Let P_k be the space of polynomials of two variables of degree less or equal to k ; we introduce the following spaces:

$$V_h^{(k)} = \{v_h \in \mathcal{C}^0(\Omega); v_{h|T} \in P_k, \quad \forall T \in \mathcal{T}_h\} \tag{30}$$

$$V_{0,h}^{(k)} = V_h \cap H_0^1(\Omega) \tag{31}$$

$$\mathcal{M}_h^{(k)} \text{ is a complementary space of } V_{0h} \text{ in } V_h \tag{32}$$

i.e. $\mathcal{M}_h^{(k)} \subset V_h^{(k)}$ and $V_h^{(k)} = \mathcal{M}_h^{(k)} \oplus V_{0,h}^{(k)}$.

Equations (14) and (16) are approximated in their variational form by

$$\left. \begin{aligned} &\text{Find } h_h \text{ in } V_h^{(k)} \text{ such that} \\ &\forall \pi_h \in V_{0,h}^{(k)}, \int_{\Omega} (\alpha^2 h_h \pi_h + D \nabla h_h \nabla \pi_h) dx = \int_{\Omega} w_h^n \pi_h dx \end{aligned} \right\} \tag{33}$$

and

$$\left. \begin{aligned} &\text{Find } \mathbf{Q}_h \text{ in } (V_h^{(k')})^2 \text{ such that} \\ &\forall \phi_h \in (V_{0,h}^{(k')})^2, \int_{\Omega} (\alpha \mathbf{Q}_h \phi_h + K \nabla \mathbf{Q}_h \nabla \phi_h) = - \int_{\Omega} c^2 \nabla h_h \phi_h dx + \int_{\Omega} (\alpha \hat{\mathbf{Q}}_h^n + \mathbf{S}_h^n) \phi_h dx \end{aligned} \right\} \tag{34}$$

in (33) and (34) w_h^n and S_h^n are approximations of w^n and S^n , respectively, which belong to V_h .

It is interesting to notice that it is possible to take here k equal to k' contrary to the flow–pressure formulation of the Navier–Stokes equations. This property is due to the presence of the time derivative in the continuity equation. In the following we shall use this remark and take $k = k'$.

We define then the approximate operator A_h by

$$\begin{aligned} \forall \lambda, \mu \in \mathcal{M}_h^{(k)}, a_h(\lambda, \mu) &= \langle A_h \lambda, \mu \rangle = \int_{\Omega} D \frac{\partial \psi_{h,\lambda}}{\partial n} \mu d\Gamma \\ &= - \int_{\Omega} (\alpha^2 \psi_{h,\lambda} \mu + D \nabla \psi_{h,\lambda} \nabla \mu) dx + \int_{\Omega} c^2 (\alpha h_{h,\lambda} + \text{div} \mathbf{Q}_{h,\lambda}) \mu dx \end{aligned} \tag{35}$$

where $h_{h,\lambda}$ (resp. $\mathbf{Q}_{h,\lambda}$ and $\psi_{h,\lambda}$) is the discrete solution of (22) (resp. (23) and (24) and where the test functions are taken in $V_{0,h}$ (resp. $(V_{0,h})^2$).

This construction leads to a matrix, which is symmetric when the bottom is flat (for details see Reference 2). It is quite easy to estimate the error bound between A and A_h , since the operators

which arise in (22)–(24) are elliptic and the right hand sides of these equations are linear functions of $h_{h,\lambda}$ and $Q_{h,\lambda}$.

If $k = k'$, we get the following estimate:

$$\|A - A_h\|_{(H^{1/2}(\Gamma), H^{-1/2}(\Gamma))} \leq Ch^k \quad (36)$$

where C is a constant which depends only on the domain Ω , and on the coefficients (α, K, \dots) of the problem. In a similar way, we construct the discrete right hand side b_h and show that

$$\|b - b_h\|_{H^{1/2}(\Gamma)} \leq Ch^k \quad (37)$$

and, consequently

$$\|\lambda - \lambda_h\|_{H^{1/2}(\Gamma)} \leq Ch^k \quad (38)$$

The approximate solutions h_h and Q_h defined by

$$\begin{aligned} h_h &= h_{\lambda_h} + h_{0,h} \\ Q_h &= Q_{\lambda_h} + Q_{0,h} \end{aligned} \quad (39)$$

satisfy the following error estimates:

$$\begin{aligned} \|h_h - h\|_{1,\Omega} &\leq Ch^k \\ \|Q_h - Q\|_{1,\Omega} &\leq Ch^k \end{aligned} \quad (40)$$

under the regularity assumption that h (resp Q) belongs to the Sobolev space $H^k(\Omega)$ (resp $(H^k(\Omega))^2$).

2.3. Estimate of computational effort—choice of \mathcal{M}_h

The space \mathcal{M}_h should be chosen so that the computations of the operator A_h and the right hand side b_h are easy. Therefore the basis functions w_i should have a small support. It seems from References 9 and 10 that a good choice is as follows:

$$\begin{aligned} \mathcal{M}_h &\text{ is complementary of } V_{0h} \text{ in } V_h, \text{ and} \\ v_h \in \mathcal{M}_h &\Rightarrow v_{h|T} = 0 \text{ for any triangle } T \text{ of } \mathcal{T}_h \end{aligned} \quad (41)$$

such that $T \cap \Gamma = \emptyset$.

With Lagrangian finite elements, \mathcal{M}_h is the space of those functions which vanish at every node of \mathcal{T}_h which does not belong to the boundary Γ .

Then $N_h = \dim(\mathcal{M}_h) = \text{card}(\Sigma_h)$, where $\Sigma_h = \{P \in \Gamma; P \text{ node of } \mathcal{T}_h\}$, i.e. N_h is the number of boundary nodes. A good choice for the basis of \mathcal{M}_h is the canonical basis $(w_I(P_J) = \delta_{IJ})$, where δ_{IJ} is the Kronecher symbol). Thus, to find the j th column of A_h , it is necessary to solve the three Dirichlet problems (22)–(24), so that A_h can be computed by solving $3 \times N_h$ Dirichlet problems.

In addition, when A_h is known, it will be factorized by a Gauss method (we can use a Cholesky method when the matrix is symmetric) to get the decomposition

$$A_h = L_h U_h \quad (42)$$

where L_h is a lower triangular matrix and U_h an upper triangular matrix.

However, these computations can be done once and for all, and at each time step, the evaluation of the solution h_h and Q_h requires then 3 Dirichlet problems to find the right hand side b_h , two linear triangular systems to compute λ_h and two Dirichlet problems to compute h_h and Q_h .

Thus a total of five Dirichlet problems and two triangular systems are required. It is important to notice that the system for the flow Q is decoupled, at least for the Dirichlet conditions, into two

subsystems on the components of \mathbf{Q} . The order of its matrix is then equal to the number of nodes for the discretization for each components of \mathbf{Q} .

2.4. Conjugate gradient method for solving the boundary system

The decomposition method described in the previous sections lies on the computation of the matrix A_h . This matrix depends closely on the variable quantity $\bar{h}\nabla h$, resulting from the linearization of the quantity $h\nabla h$; when the depth of water has some significant changes in time, it is necessary to re-evaluate the quantity \bar{h} during the computations; this imposes in term to re-evaluate the boundary matrix and to factorize it again. For these reasons we have tested an iterative version of the method.

When the bottom is flat, the matrix A_h is symmetric and it is natural to solve the boundary system $A_h\lambda = b_h$, by a conjugate gradient method. This algorithm only requires the products of A_h by some known vectors x , and the coefficients of the matrix A_h do not need to be explicitly known. When the gradient of \bar{h} does not vanish, it is also possible to use conjugate gradient techniques, by solving the so called 'normal equation'

$${}^tA_h A_h x = {}^tA_h^b$$

or, if the condition number becomes too bad, we can use some biconjugate gradient techniques.¹¹

These iterative techniques are performed when the product of the number of iterations N_i of conjugate gradient and the number of time iterations N_t , where \bar{h} is constant is less than the number of boundary nodes. Otherwise, and when the number of boundary nodes is not too large, it is cheaper to use the direct method. Anyway for both techniques we have to evaluate the product of tA_h by some known vectors y . For this purpose we have to find the adjoint system, associated with problems (22)–(24).

Computation of ${}^tA_h y$, where $y \in H^{+1/2}(\Gamma)$. After some considerations on the Lagrangian of the system it appears that the evaluation of ${}^tA_h y$ can be made by the following cascade of Dirichlet problems:

$$\left. \begin{aligned} &\text{Find } h_y^* \text{ in } V_h \text{ such that} \\ &\forall \omega_i \in V_h^0, \int_{\Omega} (\alpha^2 h_y^* \omega_i + D\nabla h_y^* \nabla \omega_i) dx = 0, \quad h_y^* = y \text{ on } \Gamma \end{aligned} \right\} \quad (43)$$

$$\left. \begin{aligned} &\text{Find } \mathbf{Q}_y^* \text{ in } (V_h^0)^2, \text{ solution of} \\ &\forall \omega_i \in (V_h^0)^2, \int_{\Omega} (\alpha \mathbf{Q}_y^* \omega_i + K \nabla \mathbf{Q}_y^* \nabla \omega_i) dx = - \int_{\Omega} \nabla (c^2 h_y^*) \omega_i dx \end{aligned} \right\} \quad (44)$$

$$\begin{aligned} &\text{Find } \psi_h^* \text{ in } V_h^0 \text{ such that} \\ &\forall \omega_i \in V_h^0, \int_{\Omega} (\alpha^2 \psi_y^* \omega_i + D\nabla \psi_y^* \nabla \omega_i) dx = \int_{\Omega} (c^2 \alpha h_y^* + \text{div}(c^2 \mathbf{Q}_y^*)) \cdot \omega_i dx \end{aligned}$$

and set

$$A_h' y = -D \frac{\partial \psi_y^*}{\partial n}, \quad \text{on } \Gamma$$

It is easy to check that A_h' is the transpose operator of A_h . Conversely it appears that it is as easy to compute ${}^tA_h x$ than $A_h x$ and it requires the same volume of computations.

2.5. Numerical tests

The method was tested on the well-known example of the Massachusetts bay for which many measurements and computation results are available (see for instance Reference 12). In this first experiment, emphasis has been laid on the general behaviour of the method rather than on the comparison with actual data. Thus the friction and Coriolis forces were not simulated. The domain was approximated by a $37 \text{ km} \times 103 \text{ km}$ rectangle and divided into 864 triangles with 481 nodes for a linear approximation. On the land boundaries the two components of the flux are taken to be zero and on the ocean limit, the inflow was assumed to be normal to the boundary and uniform; its variation in time simulates a tidal motion. Only the diffusion—propagation problem has been solved. The bottom is supposed first to be flat and then in a second experiment a Gaussian lump is assumed in the middle of the domain Figures (2 and 3) show the velocity field when the inflow is

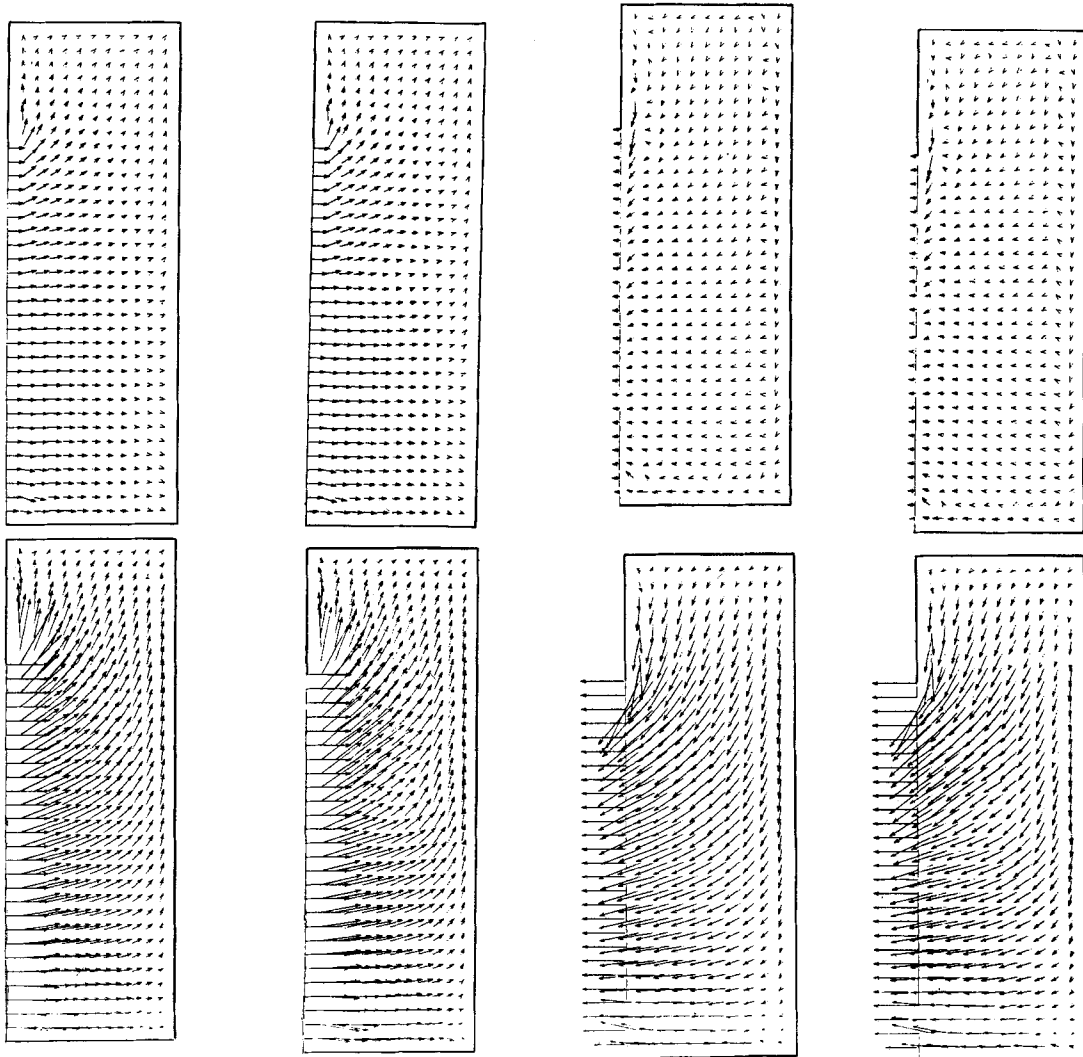


Figure 2. Velocity field. Top: after 100s, low tide.
Bottom: after 11,000s (0.25 cycle)

Figure 3. Velocity field. Top: after 22,000s (0.5 cycle), high tide. Bottom: after 33,000s (0.75 cycle)

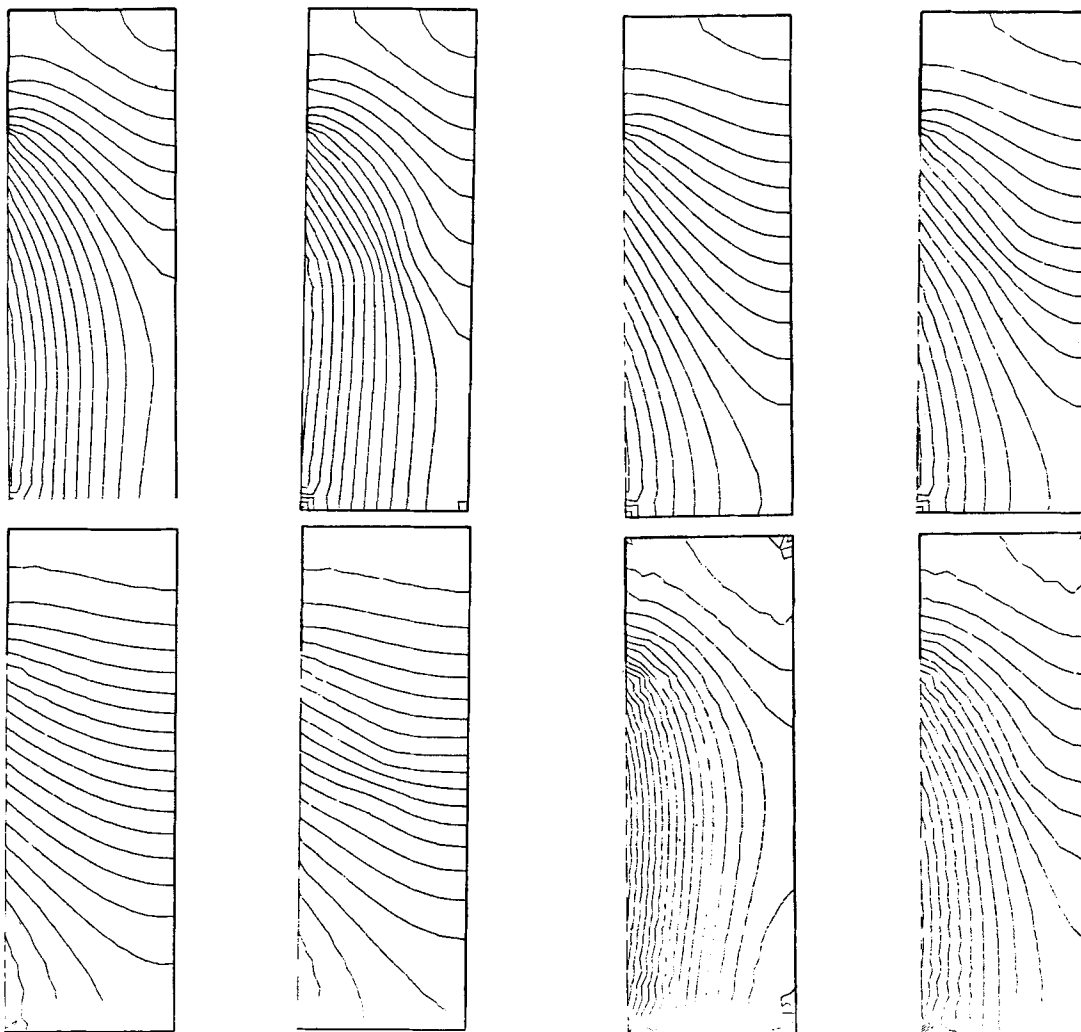


Figure 4. Surface contour lines. Top: after 100s, low tide. Bottom: after 11,000s (0.25 cycle)

Figure 5. Surface contour lines. Top: after 23,000s (0.54 cycle), high tide. Bottom: after 33,000s (0.75 cycle)

maximum in both configurations of the bottom and in Figures 4 and 5 are plotted the surface contour lines at different times of the tide. When the bottom is not flat, the influence of the lump on the contour lines and the speeding down of the flux above the lump can be observed.

We also solved the boundary system with a conjugate gradient method when the bottom was flat. The number of nodes on the boundary was equal to 96. The number of conjugate gradient iterations was varied between 4 and 9, leading to the same solution as in the direct resolution.

3. OTHER BOUNDARY CONDITIONS

The Dirichlet (or the Neumann) boundary conditions have the advantage of decoupling the problem on the flow into two problems for each component of Q . However, these boundary conditions do not take into account all physical phenomena. In this section we examine some other

boundary conditions for which the decomposition method of section 2 can be applied with some changes.

We shall divide the boundary Γ into two parts: $\Gamma = \Gamma_1 \oplus \Gamma_2$. On the ocean boundary (part Γ_1 of the boundary), we shall apply an incident wave condition and along the coast (part Γ_2) we impose an impermeability condition coupled with a friction condition.

3.1. Incident wave condition

The reader can refer to Reference 2 for a more complete justification. This condition takes into account waves coming into and going out of the domain Ω . The normal waves coming out of the domain have to cross the sea boundary without reflection; in addition, it may be possible to generate a wave coming into the domain in a given direction \mathbf{v} . It can then be written as follows:

$$-c^2 h \mathbf{n} + K \frac{\partial \mathbf{Q}}{\partial n} + c \mathbf{Q} = \mathbf{F} \quad (45)$$

with

$$\mathbf{F} = \begin{cases} (\mathbf{n} \cdot \mathbf{v} - 1) c^2 A \cos\left(\frac{\omega}{c}(\mathbf{x} \cdot \mathbf{v} - ct)\right) \\ \boldsymbol{\tau} \cdot \mathbf{v} c^2 A \cos\left(\frac{\omega}{c}(\mathbf{x} \cdot \mathbf{v} - ct)\right) \end{cases}$$

where \mathbf{n} is the exterior normal at the boundary Γ_1 , $\boldsymbol{\tau}$ the associate tangential vector, A and ω denote the amplitude and the pulsation of the wave.

For sake of simplicity we suppose that Dirichlet conditions are given along the coastal boundary Γ_2 .

A variational formulation of the propagation problem (13), (14) with the condition (45) can be written as follows:

Let $H_{\Gamma_2} = \{v \in (H^1(\Omega))^2; v = 0 \text{ on } \Gamma_2\}$; find h in $H^1(\Omega)$, \mathbf{Q} in H_{Γ_2} such that

$$\forall \omega_i \in H_0^1(\Omega), \int_{\Omega} (\alpha^2 h \omega_i + (K\alpha + c^2) \nabla h \nabla \omega_i) dx = \int_{\Omega} w^n \omega_i dx \quad (46)$$

$$\begin{aligned} \forall \omega_i \in H_{\Gamma_2}, \int_{\Omega} (\alpha \mathbf{Q} \omega_i + K \nabla \mathbf{Q} \nabla \omega_i) dx + \int_{\Gamma_1} c \mathbf{Q} \omega_i d\Gamma \\ = - \int_{\Omega} c^2 \nabla \mathbf{h} \cdot \boldsymbol{\omega}_i dx + \int_{\Gamma_1} c^2 h \boldsymbol{\omega}_i \cdot \mathbf{n} d\Gamma + \int_{\Omega} \mathbf{S}^n \omega_i dx + \int_{\Gamma_1} \mathbf{F} \boldsymbol{\omega}_i d\Gamma \end{aligned} \quad (47)$$

Using then a Green formula we get

$$\begin{aligned} \forall \omega_i \in H_{\Gamma_2}, \int_{\Omega} (\alpha \mathbf{Q} \omega_i + K \nabla \mathbf{Q} \nabla \omega_i) dx + \int_{\Gamma_1} c \mathbf{Q} \omega_i d\Gamma \\ = \int_{\Omega} (h \nabla (c^2 \omega_i) + \mathbf{S}^n \omega_i) dx + \int_{\Gamma_1} \mathbf{F} \boldsymbol{\omega}_i d\Gamma \end{aligned} \quad (48)$$

In this formulation the two components of \mathbf{Q} are decoupled and the decomposition method for the solution of h at the nodes of Γ_2 can be applied as in Section 2.

3.2. Wall condition on the flow

This condition coming from the physics expresses on the one hand the impermeability of the

coast:

$$\mathbf{Q} \cdot \mathbf{n} = 0 \tag{49}$$

and, on the other hand, simulates friction along the coasts:

$$\lambda \mathbf{Q} \cdot \boldsymbol{\tau} + K \frac{\partial \mathbf{Q}}{\partial n} \cdot \boldsymbol{\tau} = g, \quad \text{given on } \Gamma_2 \tag{50}$$

The coefficient λ may depend on the spatial co-ordinates; when λ varies from 0 to infinity, the condition varies from a free slip to a no slip condition.

The main difficulty lies in the fact that the two components of \mathbf{Q} are coupled on the boundary, and the solution of (14) with the boundary conditions (49) and (50) leads to a system of size $2N \times 2N$, if N is the number of nodes for each component. We show in this section how this difficulty can be overcome.

We set

$$V^0 = \{ \boldsymbol{\phi} \in (H^1(\Omega))^2; \boldsymbol{\phi} \cdot \mathbf{n} = 0 \text{ on } \Gamma_2, \boldsymbol{\phi} = 0 \text{ on } \Gamma_1 \}, \tag{51}$$

and, if Dirichlet conditions are fixed on Γ_1 , a variational formulation may be written as:

Find \mathbf{Q} in V^0 , solution of:

$$\left. \begin{aligned} \forall \boldsymbol{\omega}_i \in V^0, & \int_{\Omega} (\alpha \mathbf{Q} \boldsymbol{\omega}_i + K \nabla \mathbf{Q} \nabla \boldsymbol{\omega}_i) \, dx + \int_{\Gamma_2} \lambda \mathbf{Q} \boldsymbol{\omega}_i \, d\Gamma \\ & = \int_{\Omega} \mathbf{f} \boldsymbol{\omega}_i \, dx + \int_{\Gamma_2} \mathbf{g} \boldsymbol{\omega}_i \, d\Gamma \end{aligned} \right\} \tag{52}$$

In this form, it is easy to show that problem (52) has one and only one solution in V^0 .

As a direct discretization of V^0 is difficult, the more natural idea is then to find decoupled conditions for which we get the same solutions as the original problem. The most simple way is to use Dirichlet conditions.

Let $\gamma V^0 = \{ \boldsymbol{\phi} \in (H^{1/2}(\Gamma))^2; \boldsymbol{\phi} \cdot \mathbf{n} = 0 \text{ on } \Gamma_2, \boldsymbol{\phi} = 0 \text{ on } \Gamma_1 \}$

$$\left. \begin{aligned} \pi(H^{1/2}(\Gamma))^2 & \rightarrow \gamma V^0 \\ \mathbf{v} & \rightarrow \pi \mathbf{v} = \mathbf{v} - \mathbf{v} \cdot \mathbf{n} \mathbf{n} \end{aligned} \right\} \tag{53}$$

Let $\mathbf{q} \in (H^{1/2}(\Gamma))^2$ be given and $\mathbf{q}^* = \pi \mathbf{q}$, we solve:

Find \mathbf{Q} in $(H^1(\Omega))^2$ satisfying:

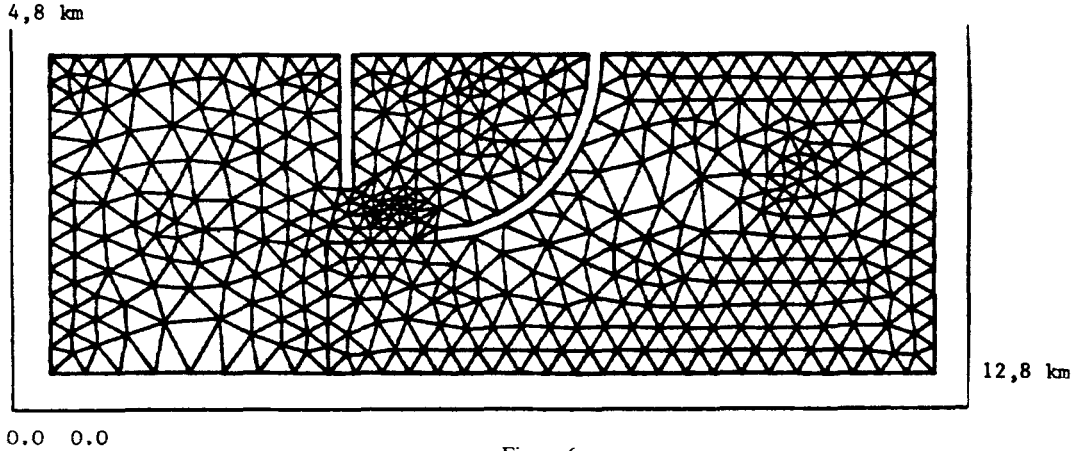
$$\left. \begin{aligned} \int_{\Omega} (\alpha \mathbf{Q} \boldsymbol{\omega}_i + K \nabla \mathbf{Q} \nabla \boldsymbol{\omega}_i) \, dx & = \int_{\Omega} \mathbf{f} \boldsymbol{\omega}_i \, dx, \forall \boldsymbol{\omega}_i \in (H_0^1(\Omega))^2 \\ \mathbf{Q} & = \mathbf{q}^* \text{ on } \Gamma_2 \end{aligned} \right\} \tag{54}$$

It is then equivalent to solve (52) and to minimize the functional \mathcal{J} defined by

$$\mathcal{J}(\mathbf{q}) = \int_{\Omega} (\frac{1}{2}(\alpha \mathbf{Q}^2 + K(\nabla \mathbf{Q})^2) - \mathbf{f} \mathbf{Q}) \, dx + \int_{\Gamma_2} (\frac{1}{2} \lambda (q^*)^2 - \mathbf{g} \cdot \mathbf{q}^*) \, d\Gamma \tag{55}$$

The functional \mathcal{J} is symmetric and coercive; it can then be minimized by a conjugate gradient method.

When a discretization in space is done, and for polygonal domains we have to define an approximation \mathbf{n}_h of \mathbf{n} at each boundary node, we define \mathbf{n}_h as the solution of the following linear



variational problem:

$$\left. \begin{aligned} \mathbf{n}_h \in \gamma V_h &= \{ \boldsymbol{\mu}_h; \boldsymbol{\mu}_h = \mathbf{v}_{h|\Gamma}, \mathbf{v}_h \in V_h^2 \} \\ \int_{\Gamma} \mathbf{n}_h \cdot \boldsymbol{\mu}_h \, d\Gamma &= \int_{\Gamma} \mathbf{n} \cdot \boldsymbol{\mu}_h \, d\Gamma \end{aligned} \right\} \quad (56)$$

whose matrix is symmetric positive definite; the approximate operator π_h is defined by:

$$\begin{aligned} \pi_h: \gamma V_h &\rightarrow \gamma V_h^0 \\ \boldsymbol{\mu}_h &\rightarrow \pi_h \boldsymbol{\mu}_h = \boldsymbol{\mu}_h - \boldsymbol{\mu}_h \cdot \mathbf{n}_h \mathbf{n}_h \end{aligned}$$

This method is easy to implement and the number of unknowns \mathbf{q}_h^* is equal to N_h , the number of nodes on the part Γ_2 of the boundary. The iterative process to find \mathbf{q}_h^* has to be inserted inside the iterative algorithm for the boundary value of the water depth λ .

We have tested this method for the propagation problem (13), (14) with the boundary conditions (49), (50), where Ω was a rectangular domain with two jetties representing a harbour during a tidal period. The triangulation of the domain Ω is plotted in Figure 6. There are 505 nodes and 892 triangles.

The boundary conditions are of Dirichlet type on the sea boundary (horizontal lower and vertical boundaries) and simulate a tidal current:

$$\left. \begin{aligned} Q_x &= 2.5 \sin \left(\frac{2\pi}{T_m} (t - 0.72x) \right) \exp(-10^{-4}y) \\ Q_y &= 0 \end{aligned} \right\}$$

where T_m is the period of the tide, $T_m = 43,200$ s.

The boundary conditions on the seashore follow the 'wall law' described in this section:

$$\left. \begin{aligned} \mathbf{Q} \cdot \mathbf{n} &= 0 \\ \lambda \mathbf{Q} \cdot \boldsymbol{\tau} + K \frac{\partial \mathbf{Q}}{\partial \mathbf{n}} \cdot \boldsymbol{\tau} &= 0 \end{aligned} \right\}$$

with $\lambda = 10^{-4}K$.

In Figure 7 are plotted the surface contour lines during a tidal period at mid tide, high and low

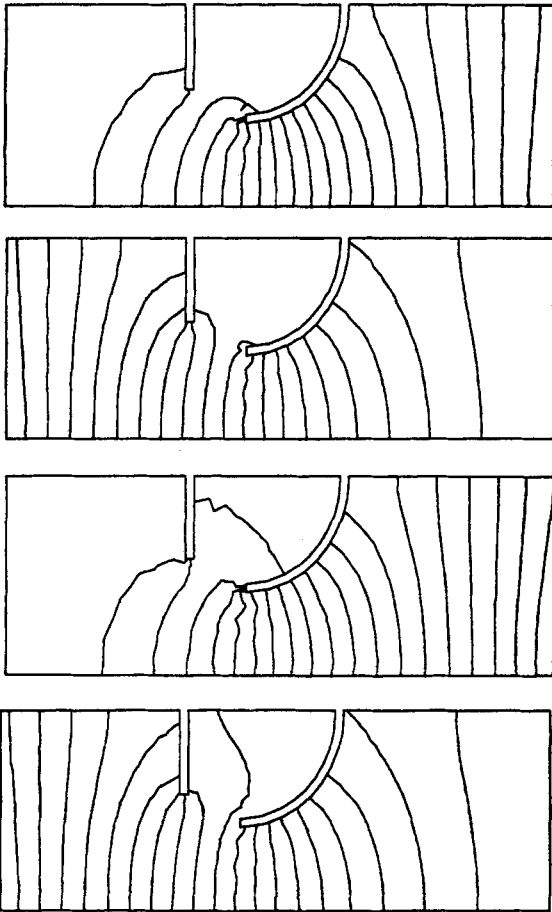


Figure 7. Surface contour lines: (a) after 10,800s (0.25 cycle); (b) after 21,600s (0.5 cycle), high tide; (c) after 32,400s (0.75 cycle); (d) after 43,200s (1 cycle), low tide

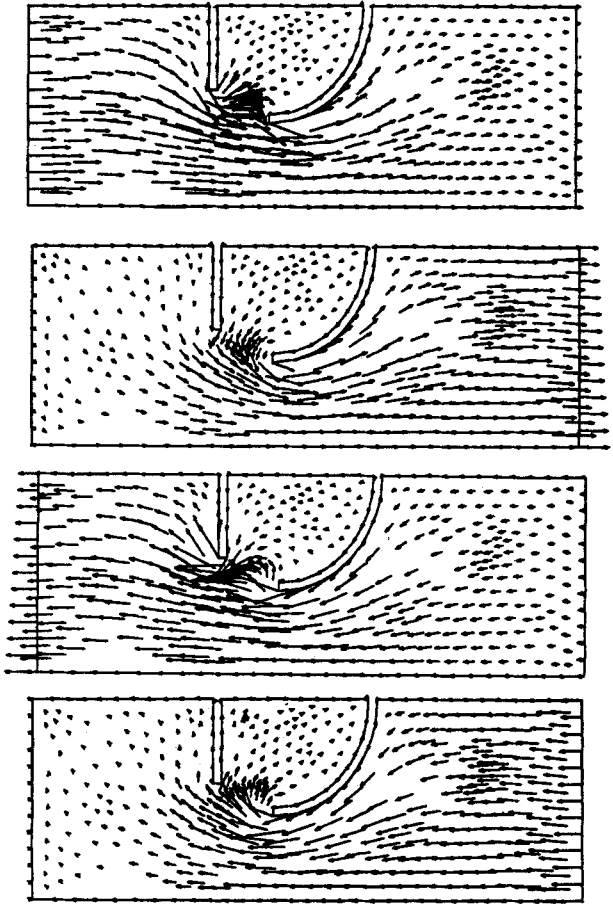


Figure 8. Velocity field: (a) after 10,800s (0.25 cycle); (b) after 21,600s (0.5 cycle), high tide; (c) after 32,400s (0.75 cycle); (d) after 43,200s (1 cycle), low tide

tide; in Figures 8 and 9 the velocity field is plotted for the same time step in the whole domain and at the entrance of the harbour.

Both boundary problems have been solved simultaneously by the iterative conjugate gradient method, described in Section 2.4 and in this section. Initial guesses for these algorithms play an important role: they were taken equal to the solution of the corresponding problems at the previous time step.

For the boundary problem on the depth of water $\lambda = h_{\Gamma}$, the number of conjugate gradient iterations was varied between 2 and 7 to reach the error

$$e_h = \frac{\|A_h \lambda - b_h\|}{\|b_h\|}$$

less than $\varepsilon_h = 10^{-5}$.

For the minimization problem (55) the number of conjugate gradient iterations was varied between 12 (at the beginning of the computations) and 4 (when initialized with the solution at the

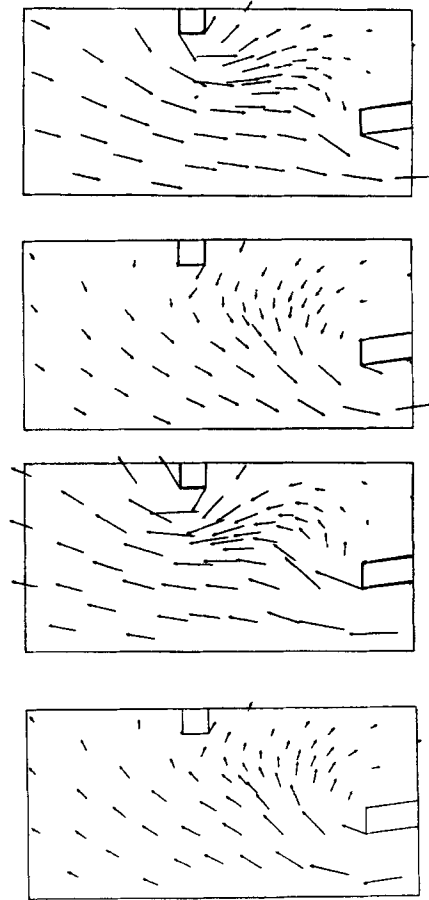


Figure 9. Velocity Field at the entrance of the harbour: (a) after 10,800s (0.25 cycle); (b) after 21,600s (0.5 cycle), high tide; (c) after 32,400s (0.75 cycle); (d) after 43,200s (1 cycle), low tide

previous time step) to reach the relative error

$$e_q = \frac{J(\mathbf{q})}{J(\mathbf{q}_0)}$$

less than 10^{-5} .

4. INDUSTRIAL APPLICATION

4.1. *The outer harbour of Dunkirk*

We present here a computation of the tidal currents in the neighbourhood of the outer harbour of Dunkirk. For this case, a physical model has been studied and we can compare then the physical observations with our computational results. The computations have been made at E.D.F. (Chatou).

The finite element grid is made of 892 triangular elements. Both the water level h and the fluxes Q

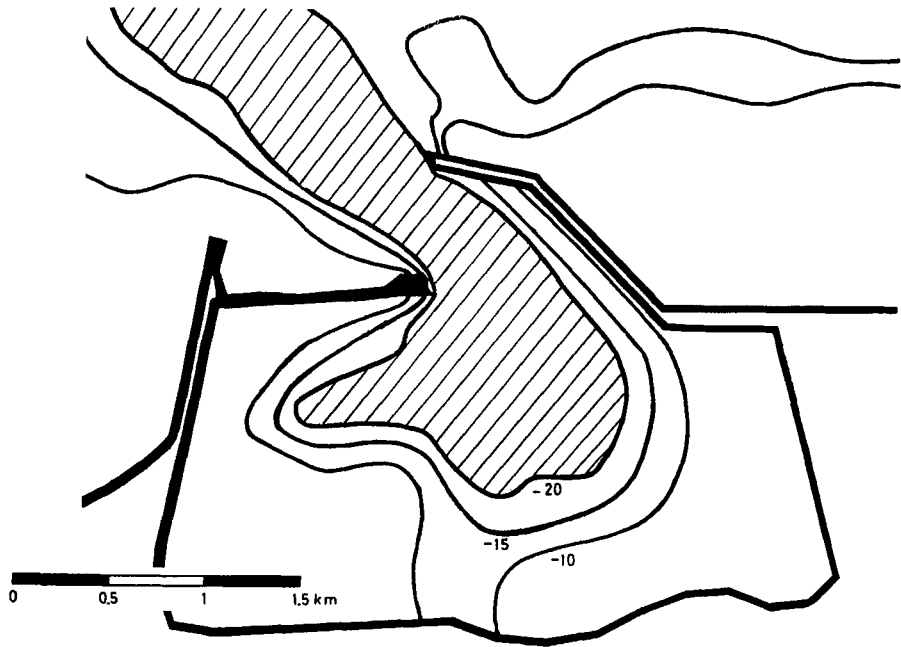


Figure 10. The outer harbour of Dunkirk—topography of the bottom

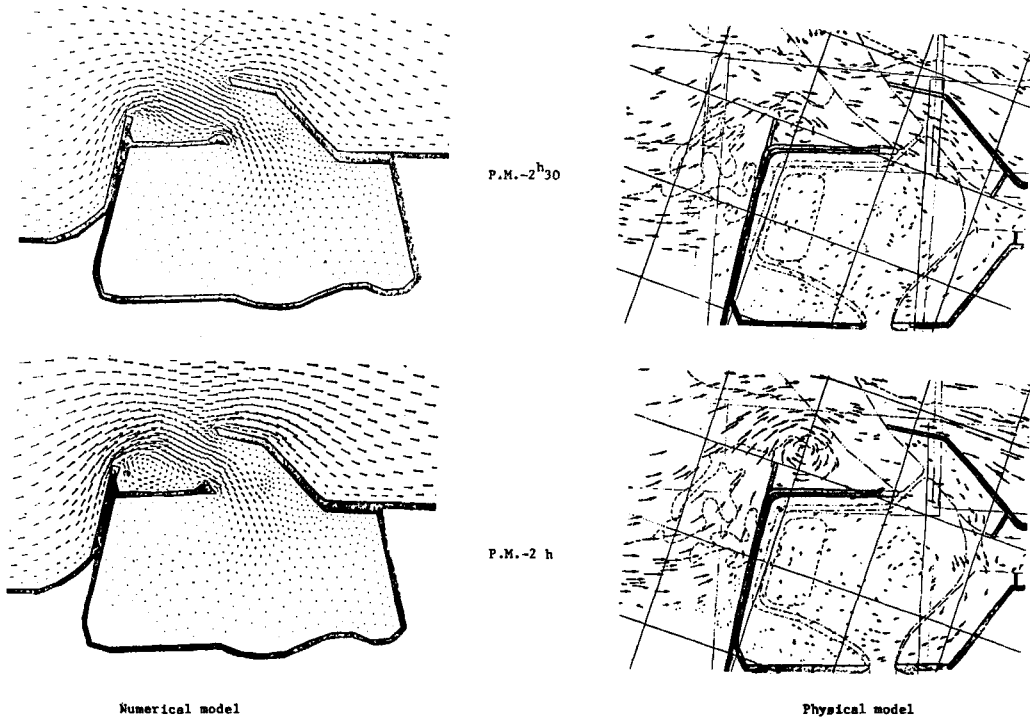


Figure 11

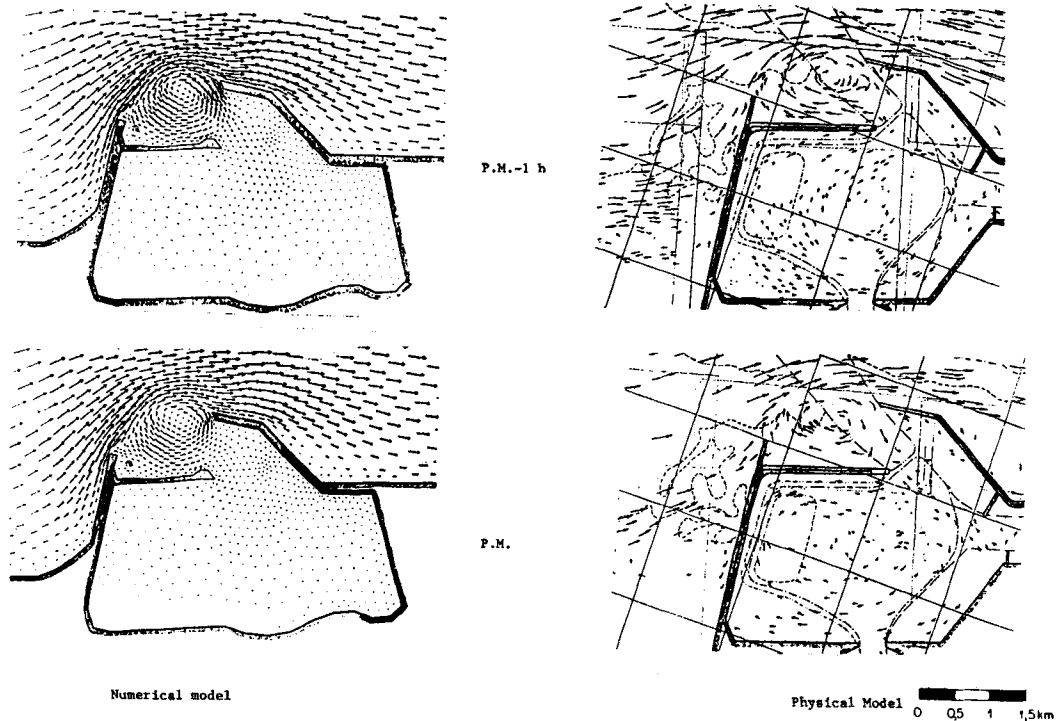


Figure 12

have been approximated by quadratic piecewise functions. There are 1909 degrees of freedom for each unknown function with 232 boundary nodes, leading to a global system of 5727 unknown.

The right hand sides of the dynamic equations include bottom friction forces and the topography of the bottom takes into account a navigation channel at the entrance of the harbour, shown in Figure 10. The slopes attain 10 per cent and this is a severe condition for the shallow water equations.

The boundary conditions are of Dirichlet types on the fluxes: they have been interpolated from results of a previous computation made on a larger domain with a coarse mesh. The time step is 60s leading to a Courant number from 1 to 5. The Chezy coefficient is equal to $60 \sqrt{\text{m/s}}$.

The computations have been carried out during more than a tidal period and are compared in Figures 11 and 12 with the data obtained by photography of floats on the physical model.

In Figure 11, the plotted currents show how the harbour fills in with the flood and a vortex develops behind the west jetty in both models.

Later on (Figure 12) the vortex size increases and occupies the whole entrance of the harbour.

The global error during one tide cycle on the mass conservation attains 3 per cent of the tidal range. The probable cause is the important variation of the bottom, since this error was less than 1 per cent in a first schematic run where the bottom was flat.

4.2. The channel

The purpose of such a computation is to obtain storm surges induced by meteorological conditions. For this application, the pressure field upon the sea surface and the stresses due to friction of wind have been introduced in the model. Because of the horizontal extension of the

domain Coriolis effects cannot be neglected. They have been implemented through an explicit formulation.

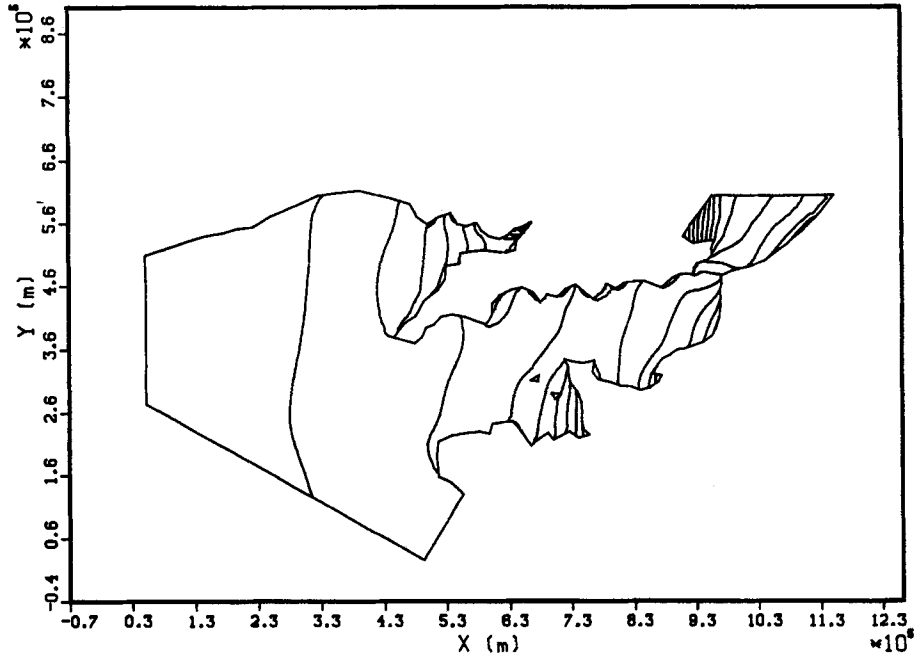


Figure 13. Model of the channel—surface contour lines at time $t = 8 \times 10^4$ s

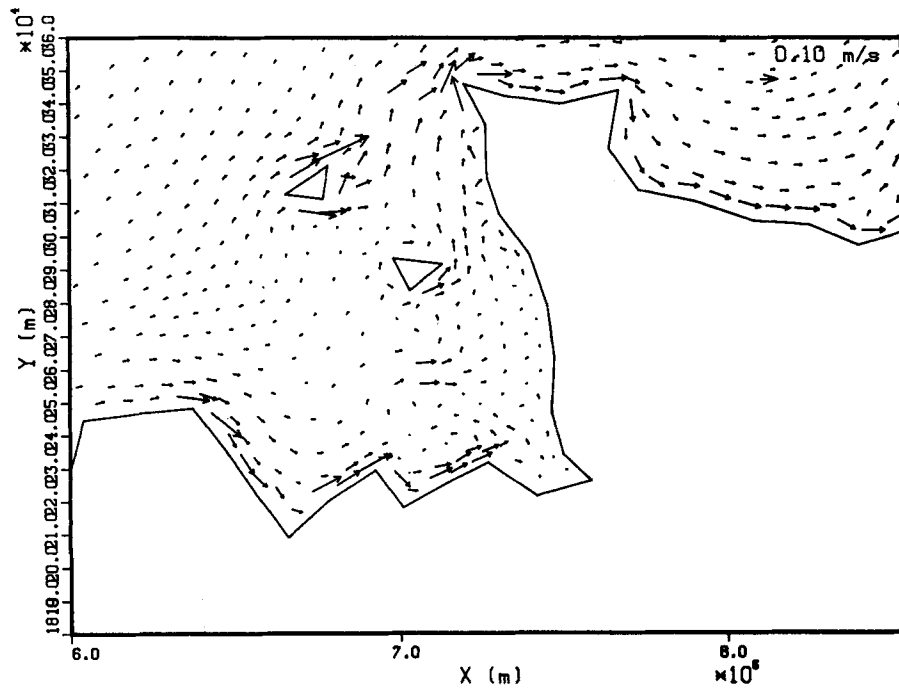


Figure 14. Model of the channel—velocity field at time $t = 8 \cdot 10^4$ s

In a first step, in order to simplify the problem, the computation of storm surges is supposed decoupled from tidal effects (i.e. one computes only variations of sea-surface elevation and of currents). In these conditions, the boundary conditions are:

- (i) $Q_{\Gamma_2} = 0$ along the coasts,
- (ii) Incident wave condition (see Section 3.1) without incoming wave on the sea sides.

Figures 13 and 14 present the results obtained near steady state, by blowing a constant West wind (20 m/s) on a initial rest state. For this computation the Courant number varies from 1 to 3 in the shallow part of the domain.

In Figure 13 are plotted the isoelevation lines of the sea surface, and the velocity field in the neighbourhood of France is presented in Figure 14. In that application non-zero velocities are obtained at steady state because of the non-uniform bathymetry of the problem.

At present real simulation of storm surges with observed pressure and wind fields are being studied.

5. CONCLUSION

We have discussed in this paper a new method for solving the diffusion-propagation step of the shallow water equations, where we have added some physical boundary conditions: incident wave condition, impermeability and friction constraints. For each type of boundary condition we have shown that the global system can be decoupled into a problem for the water elevation and a problem for the fluxes. In addition, for this last problem the two components of the fluxes can be treated as independent unknowns.

These methods are well adapted to approximations by finite element methods which allow more complicated geometry than with finite difference approximations together with local refinement of the mesh. With this work (decoupling and finite element) industrial simulations are possible.

REFERENCES

1. J. Goussebaile, 'Modélisation d'écoulements et transfert de chaleur par une méthode de différences finies en mailles non orthogonales', *Rapport EDF HE41/81/27*.
2. S. Dalsecco, G. Labadie and B. Latteux, 'Résolution des équations de St Venant par une méthode d'éléments finis', *Rapport EDF, HE/41/82.15 et HE/42/82.34*.
3. G. Labadie, J. P. Benque and J. Ronat, 'Une nouvelle méthode d'éléments finis pour les équations de Navier–Stokes couplées avec une équation thermique', *Rapport EDF HE41/83/10*, soumis à *International Journal for Numerical Methods in Fluids*.
4. B. Ibler, 'Résolution des équations de Navier–Stokes par une méthode d'éléments finis', *Thèse de Doctorat de l'Université Paris Sud*, Janvier 1981.
5. M. Bercovier, O. Pironneau and V. Sastri, 'Finite elements and characteristics for some parabolic-hyperbolic problems', *Applied Mathematical Modelling*, **7**, April (1983).
6. J. P. Benque and J. Ronat, 'Quelques difficultés des modèles numériques en hydraulique', *Fifth International Symposium on Computing Methods in Applied Sciences and Engineering*, INRIA Versailles, France, December 1981.
7. R. Glowinski and O. Pironneau, 'Numerical methods for the first biharmonic equation and for the two dimensional Stokes problem', *SIAM Review*, **21**, (2), 1979.
8. P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, 1978.
9. P. G. Ciarlet and R. Glowinski, 'Dual iterative techniques for solving a finite element approximation of the biharmonic equation', *Comp. Methods Applied Mech. Engng*, **5**, 277–295 (1977).
10. R. Glowinski, J. L. Lions and R. Tremolieres, *Analysis Numérique des Inéquations variationnelles*, Vols. 1 et 2, Dunod-Bordas, Paris, 1976.
11. P. Joly, 'Résolution de systèmes linéaires non symétriques par des méthodes de gradient conjugué', *Rapport Université Pierre et Marie Curie*,
12. J. J. Connor and C. A. Brebbia, *Finite Element Techniques for Fluid Flow*, Newnes-Butterworth, 1976.
13. J. P. Benque, B. Ibler, A. Keramsi and G. Labadie, 'A finite element method for the Navier Stokes equations', *3rd International Conference on Finite Elements in Flow Problems*, Banff, Alberta, Canada 10–13 June, 1980.
14. O. Pironneau, 'On the transport-diffusion algorithm and its application to Navier Stokes equations', *Num. Math.*, **38**, 309–332 (1982).