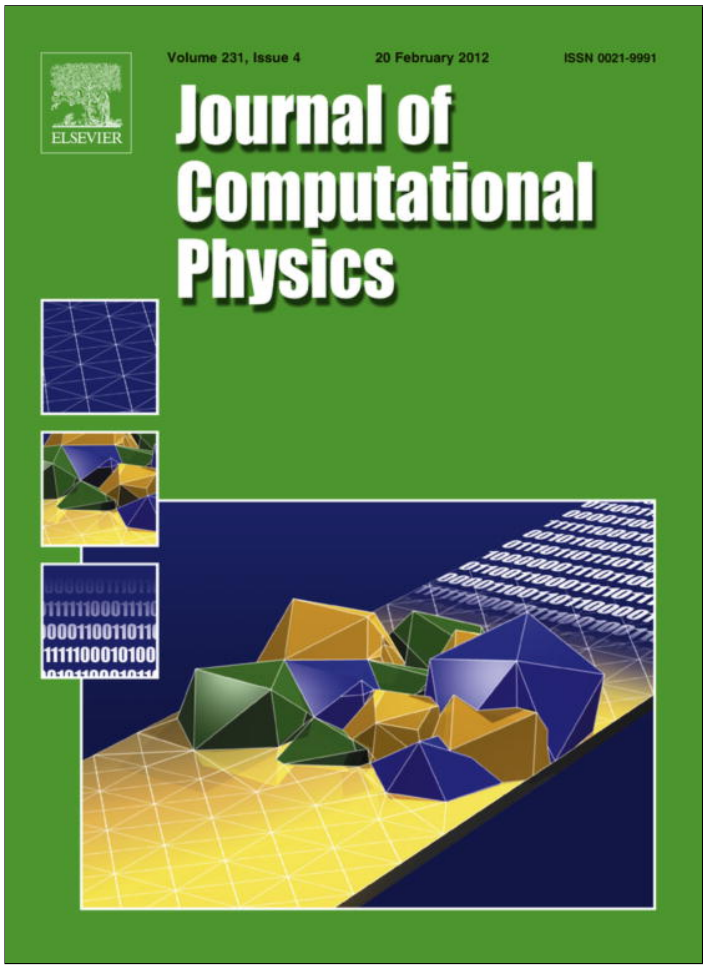

Journal of Computational Physics 231 (2012) 1387–1425

Contents lists available at SciVerse ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

The MAST-edge centred lumped scheme for the flow simulation in variably saturated heterogeneous porous media

Costanza Aricò *, Marco Sinagra, Tullio Tucciarelli

Dipartimento di Ingegneria Civile, Ambientale ed Aerospaziale, Università di Palermo, Viale delle Scienze, 90128 Palermo, Italy

ARTICLE INFO

Article history:
Received 16 October 2010
Received in revised form 4 October 2011
Accepted 12 October 2011
Available online 19 October 2011

Keywords:
Variably saturated porous medium
Numerical model
Finite element
Lumped scheme
Mass conservation
Unstructured mesh
Analytical solution

ABSTRACT

A novel methodology is proposed for the solution of the flow equation in a variably saturated heterogeneous porous medium. The computational domain is descretized using triangular meshes and the governing PDEs are discretized using a lumped in the edge centres numerical technique. The dependent unknown variable of the problem is the piezometric head. A fractional time step methodology is applied for the solution of the original system, solving consecutively a prediction and a correction problem. A scalar potential of the flow field exists and in the prediction step a MArching in Space and Time (MAST) formulation is applied for the sequential solution of the Ordinary Differential Equation of the cells, ordered according to their potential value computed at the beginning of the time step. In the correction step, the solution of a large linear system with order equal to the number of edges is required. A semi-analytical procedure is also proposed for the solution of the prediction step. The computational performance, the order of convergence and the mass balance error have been estimated in several tests and compared with the results of other literature models.

1. Introduction

Many environmental problems are strongly related to the flow dynamic occurring in variably saturated soils, affecting the heat and/or chemical transport and the resulting groundwater quality.

Flow in variably saturated porous media is mathematically described by the Darcy equation of fluid motion and by the fluid mass conservation equation. The resulting governing Partial Differential Equation (PDE), called Richards equation [40], is characterized by the strongly non linear relationship between the water content (or the saturation rate) and the water piezometric depth and between the relative hydraulic conductivity and the water content (or saturation rate), as explained in details in the following.

Due to the strong non-linearity, analytical solutions are restricted to very simple cases and numerical methods are almost regularly required. Because of the highly nonlinearity of the governing equation, linearization techniques, that maintain both the accuracy and the mass conservation property, are commonly applied. Typical linearization techniques used are Newton and Picard methods. A comparison of the two techniques in the solution of variably saturated multidimensional porous media can be found in [35].

The dependent variables of the process are the water piezometric depth and the volumetric water content (or the saturation rate). To close the mathematical model, a constitutive relationship between the piezometric depth and the water content functions is needed to convert one variable to the other (and vice versa). A second relationship between the water

* Corresponding author.
E-mail addresses: arico@idra.unipa.it (C. Aricò), sinagra@idra.unipa.it (M. Sinagra), tucciar@idra.unipa.it (T. Tucciarelli).

0021-9991/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2011.10.012

content (or the saturation rate) and the relative hydraulic conductivity is additionally required. In order to reduce the dimensionality of the problem, it is possible to merge one set of equations (the momentum equations) in the other (the continuity equations) solving the problem in the piezometric depth unknown only or in the water content unknown only. Depending on such a choice, different modelling approaches result which are mathematically equivalent in the continuous formulation, but different in their discrete analogous.

Three forms of the variably saturated flow equation can be derived: (1) the pressure based – form, where the primary variable is the piezometric depth (or the piezometric head), (2) the water content – form, where the volumetric water content is chosen as the primary variable, and (3) the mixed – form, where both variables are employed and, in solving the discrete equation system, the piezometric depth is actually used as the primary variable.

Each of the three different forms has its own advantages and drawbacks. The water content – form is restricted to unsaturated flow conditions because the water content variable is unique for saturated regions and the piezometric depth – water content relationship no longer exists. The use of this formulation also requires the assumption that the parameters of the capillary pressure relation are spatially invariable (see for example discussion in [28]).

The pressure based – form can be used for both saturated and unsaturated soils. The piezometric depth variable is unique and continuous. Models of this type have been extensively used in various applications [17,18,23,31,34,35,41]. It has been shown [12] that, under very dry conditions, where the piezometric depth – water content curve is strongly non-linear, this approach shows convergence difficulties and the resulting approximations could produce significant global mass balance errors, unless very small time steps are used. Tocci et al. [42] have shown that a differential algebraic approximation (DAE) of the method of lines (MOL) gives solutions of the pressure based – form of the RE that are accurate, with good mass balance properties, specially if higher order methods in times are used.

These difficulties can be avoided if the mixed form is used. Celia et al. [12] solve the mixed form by a modified Picard scheme, where the piezometric depth is assumed as primary variable for the solution at each new iteration step. The authors expanded each element of the water content at the new iteration step in Taylor series truncated to the linear term. They obtained the so-called delta form of the Picard scheme. Manzini and Ferraris [30] applied the same procedure in a 2D Finite Volume code. According to Celia et al. [12] and Manzini and Ferraris [30], it is the term related to the time change of the water content in the Picard scheme that guarantees a more accurate mass balance with respect to the piezometric depth formulation, where this term does not appear.

Forsyth et al. [16] introduced from multiphase flow modelling the idea of the "primary variable substitution" or "primary variable switching" technique. In this approach, a full Newton method is used where the different primary variables, namely saturation and pressure, are switched in different regions depending on the prevailing saturation conditions at each node of a mesh. This technique was found to yield rapid convergence in both the unsaturated and saturated zones compared to the pressure based – formulation.

In the past decades, both Finite Volumes (FV) and Finite Elements (FE) methods have been widely applied for the solution of flow fields in variably saturated porous media. FE schemes guarantee both global and local mass conservation (for local mass conservation see the discussion provided in [22]). Unlike FV first order and second order methods, the potential continuity is guaranteed across the elements, but velocity is calculated by differentiation of the potentials inside the elements. This implies that the discrete normal fluxes are discontinuous across inter-element boundaries. The control volume where mass balance is enforced is properly identified [37]. For example, in the Galerkin technique applied on 2D triangulation, the subdomains can be defined as the Voronoi (or Thiessen) polygons. Application of this approach in 3D is very complicated and provides a strong increment of the computational time. On the other hand, the FV approach is locally mass conservative because the subdomain where the mass balance is applied is the same mesh element. The Mixed Finite Element (MFE) methods provide and attractive framework for these types of problems: by simultaneously approximating the potential and the normal fluxes, the computed normal fluxes are continuous across inter-element edges and the local and global mass balance are automatically achieved in the case of constant parameters. MFE methods have been extensively used for the solution of parabolic problem in many application fields (groundwater flow problems, petroleum reservoir problems, potential flow problems, ...), but in elliptic problems (i.e. steady state problems) the matrix of the system becomes ill-conditioned, leading to saddle-point problem [9,20]. Mixed Hybrid Finite Element (MHFE) methods represent a way to solve this problem. In the MHFE method, piezometric heads at element edges are assumed as additional variables to overcome this problem. The final linear algebraic systems are always symmetric and positive definite. In engineering applications concerning porous media, the lowest order (zero order [39]) RT_0-$P0$ is the most common formulation of MHFE.

MFE method leads to the solution of a symmetric positive definite linear system in approximating pure transient parabolic problems. In the case of triangular elements, the MHFE needs to invert a (3×3) order matrix for each element. Flat triangles could blow up the conditioning of the corresponding (3×3) matrices, so the way these matrices are inverted is important [20]. Hoteit et al. [20] proved that in heterogeneous media, the conditioning of the resulting linear system for the MHFE grows up linearly according to the ratio between the highest and the lowest values of the hydraulic conductivity of adjacent elements and the algorithm could accumulate numerical errors if large jumps in the conductivity values take place.

Putti and Sartoretto [38] compared the conforming linear Galerkin FE scheme with the RT_0-$P0$ MHFE for the solution of 2D diffusion problems, like saturated porous media flow, discretized on unstructured triangular meshes. They found that both methods are mass conservative with constant parameters if the proper control volume is chosen: the Galerkin linear FE is conservative on the Voronoi cell, while the RT_0-$P0$ MHFE is conservative on both triangle and Voronoi cell. The authors

compared also the streamlines obtained from the two numerical vector fields on both homogeneous and heterogeneous media. The Galerkin FE velocities are not parallel to the impervious boundaries, while the RT_0-$P0$ do are because of the continuity of normal component of the velocity vectors at element edges. In heterogeneous media the streamlines computed by means of the FE scheme show unfeasible behaviour also where a strong change of the hydraulic conductivity occurs, since the mass balance control volumes (Voronoi cells) across the discontinuity.

In a work by Brezzi et al. [10] it is shown that MHFE methods applied to time-independent advection-diffusion problems do not obey the discrete maximum principle. The maximum principle problem states that the solution cannot have a maximum or a minimum within the interior of the domain [36]. Hoteit et al. [19] have shown that the MHFE scheme conditionally satisfies the maximum principle for parabolic time dependent problems, like groundwater problems. A mass-lumping technique is one of the alternative approaches suggested by the authors to prevent unphysical oscillations.

Recently, an edge centred lumped formulation of the MHFE schemes has been proposed [44]. The flow equations are spatially discretized in a set of continuity equations across all the edges of the mesh, using the average potentials along these edges as unknowns. The formulation leads to final linear algebraic systems that are always symmetric and positive definite. Numerical experiments have shown that the LMHFE, compared to the standard MHFE, avoids oscillations for acute triangulations, strongly reduces oscillations with rectangular or quadrangular meshes and does not create additional numerical errors [19,44].

In the present paper, a numerical methodology for the simulation of the 2D flow field (in the vertical plane x–z) in variably saturated homogeneous and heterogeneous porous media is proposed. The governing equations are solved applying a fractional time step procedure, solving consecutively a prediction and a correction problem. The non-linear components of the problem are concentrated in the prediction step, while the correction step leads to the solution of a linear system, with order equal to the number of computational cells. The governing equations are written in pressure – form and the unknown of the problem is the piezometric head.

In the proposed approach, the governing PDEs are discretized in space using triangular meshes, according to a scheme equivalent to the LMHFE schemes, enforcing the mass balance in the computational cells located in the middle of each edge. A linear variation of the piezometric head is assumed inside each element of the computational mesh and the volume of each triangle is concentrated in the midpoint of each of the three sides, in the measure of one third of the area of each triangle (one or two) sharing the edge of the cell.

The prediction step is solved applying the MArching in Space and Time (MAST) methodology, recently proposed for the solution of advection dominated problems [2,7], fully dynamic shallow waters equations [3,4], diffusive form of the shallow waters equations [6], as well as transport problems in saturated porous media with variable density [5]. The requirement for the application of the MAST methodology is the existence of an exact or approximated scalar potential for the flow field. In the present physical problem, an exact scalar potential of the flow field exists and is the piezometric head. At the beginning of each time step, computational cells are ordered according the their potential values. MAST solves a sequence of Ordinary Differential Equations (ODEs), one for each computational cell, from the highest to the lowest potential value. MAST has shown unconditional stability with regard to the time step size, also for Courant–Friedrichs–Levy (CFL) number much greater than 1.

The paper is organized as follows: in Section 2 we show the governing PDEs along with the adopted closure relationships; in Section 3 the prediction and the correction problems are defined in the framework of the fractional time step methodology. In Section 4 the MAST procedure for an efficient solution of the two problems defined in Section 3 and the corresponding flux discretization are presented. In the same Section 4 a semi-analytical solution of the prediction problem in each cell is proposed. In Section 5 the application of the proposed model to a number of literature tests is presented. The tests are aimed to investigate the computational efficiency, the convergence ratio and the mass conservation capability of the proposed scheme.

2. Physical model and governing equations system

The governing PDEs are the mass conservation equation and the momentum equation [8]. The momentum equation, neglecting the inertial terms as well as the interfacial drag terms (Forchheimer term) and the deviatoric component of the stress tensor (Brinkmann term), becomes the Darcy expression for the velocity [8]. Assuming isotropic medium, the governing PDEs system is:

$$S_0 \frac{\theta(\psi)}{\varepsilon} \frac{\partial \psi}{\partial t} + \frac{\partial \theta(\psi)}{\partial t} + \nabla \cdot \mathbf{q} = Q, \tag{1}$$

$$\mathbf{q} = -K_s K_r(\theta(\psi)) \cdot \nabla H, \tag{2}$$

where ε is the porosity (greater than zero), S_0 is the specific storage due to fluid and medium compressibility ($S_0 = \varepsilon\gamma + (1-\varepsilon)\Gamma$, with γ and Γ coefficients of fluid and skeleton compressibility respectively), ψ is the piezometric depth, θ is the volumetric water content further specified, $s(\psi) = \theta(\psi)/\varepsilon$ is the saturation rate, $\mathbf{q}$ is the Darcy flux vector, K_s is the hydraulic conductivity for saturated medium, K_r is the relative hydraulic conductivity ($0 < K_r \leqslant 1$; $K_r = 1$ for saturated medium), H is the piezometric head (or hydraulic head or potential), $H = z + \psi$, where z is the topographic elevation and Q is a

source term. In the present approach, the air pressure in the unsaturated soil is assumed constant and equal to the atmospheric pressure.

Mass conservation Eq. (1) is written in pressure - form, assuming:

$$\frac{\partial\theta(\psi)}{\partial t}=\frac{\partial\theta}{\partial\psi}\frac{\partial\psi}{\partial t}, \tag{3}$$

where $\partial\theta/\partial\psi$ is referred to as specific capacity. Since $\partial H/\partial t = \partial\psi/\partial t$ (because $\partial z/\partial t = 0$), Eq. (1) can be written as:

$$\frac{\partial H}{\partial t}\left(S_0\frac{\theta(\psi)}{\varepsilon}+\frac{\partial\theta}{\partial\psi}\right)+\nabla\cdot\mathbf{q}=Q. \tag{4}$$

Initial and boundary conditions have to be specified to make the problem well posed. Boundary conditions may be of Dirichlet (prescribed piezometric head or piezometric depth) or Neumann (prescribed flux) type. If Ω is the spatial domain where problem (4) is defined, initial and boundary conditions can be written as:

$$\begin{aligned}&\psi(\mathbf{x},t)=\psi_D(\mathbf{x},t)\quad\text{or}\quad H(\mathbf{x},t)=H_D(\mathbf{x},t),\quad \mathbf{x}\in\Gamma_D,\\&\mathbf{q}(\mathbf{x},t)\cdot\mathbf{n}=g_N(\mathbf{x},t),\quad \mathbf{x}\in\Gamma_N,\\&\psi(\mathbf{x},0)=\psi_0\quad\text{or}\quad H(\mathbf{x},0)=H_0,\quad \mathbf{x}\in\Omega,\end{aligned} \tag{5}$$

where $\Gamma=\Gamma_D\cup\Gamma_N$ is the boundary of Ω, Γ_D and Γ_N are the portions of Γ where Dirichlet and Neumann boundary conditions respectively hold, H_D and ψ_D are the assigned Dirichlet values for H and ψ, g_N is the assigned Neumann flux, $\mathbf{n}$ is the unit outward normal to the boundary and the subscript 0 marks the initial state in the domain.

To solve the PDEs governing system, closure relationships linking the volumetric water content and the relative hydraulic conductivity to the piezometric depth are needed. Neglecting hysteretic phenomena, the following empirical relationships are used:

(a) the van Genucthen model [43]:

$$\theta^*=\frac{1}{[1+|\mu\psi|^{\lambda}]^m}\quad\text{if }\psi<\psi_a;\quad\theta^*=1\quad\text{if }\psi\geqslant\psi_a, \tag{6a}$$

$$K_r(\theta^*)=\sqrt{\theta^*}\left[1-\left(1-\sqrt[m]{\theta^*}\right)^m\right]^2, \tag{6b}$$

(b) the Brooks–Corey model [11]:

$$\theta^*=\left(\frac{\psi_b}{\psi}\right)^{\lambda}=\left(\frac{|\psi_b|}{|\psi|}\right)^{\lambda}\quad\text{if }\psi<\psi_b;\theta^*=1\quad\text{if }\psi\geqslant\psi_b, \tag{7a}$$

$$K_r(\theta^*)=(\theta^*)^{p*}, \tag{7b}$$

where the effective water content θ^* is defined as:

$$\theta^*=\frac{\theta-\theta_r}{\theta_s-\theta_r}, \tag{8}$$

where $\theta_r\leqslant\theta\leqslant\theta_s$, θ_s is the volumetric water content at saturation, θ_r is the residual water content, $0<\theta^*\leqslant 1$, ψ_a is the "air-entry pressure head: ($\psi_a\leqslant 0$), ψ_b is the "bubble pressure" ($\psi_b<0$), λ is the "pore size distribution index", $m=1-1/\lambda$, μ is a curve fitting parameter, and $p^*=3+2/\lambda$.

By merging Eqs. (2) and (4) you get the final governing PDE:

$$\alpha(\psi)\frac{\partial H}{\partial t}-\nabla\cdot(K_sK_r(\theta(\psi))\cdot\nabla H)=Q,\quad\text{with}\quad\alpha(\psi)=\left(S_0\frac{\theta(\psi)}{\varepsilon}+\frac{\partial\theta}{\partial\psi}\right). \tag{9}$$

The present model does not account for possible hysteresis in soil hydraulic properties, although this phenomenon can be incorporated using, for example, the approach developed by Kool and Parker [26], or Huang et al. [21].

3. The fractional time step methodology

Eq. (9) can be solved in the H unknown by means of a fractional time step approach by solving consecutively a prediction and a correction problem.

Assume a general system of balance laws:

$$\frac{\partial\mathbf{U}}{\partial t}+\nabla\cdot\mathbf{F}(\mathbf{U})=\mathbf{B}(\mathbf{U}), \tag{10}$$

where $\mathbf{U}$ is the vector of the unknown variables, $\mathbf{F}(\mathbf{U})$ is the flux vector and $\mathbf{B}(\mathbf{U})$ is a source term. Applying a fractional time step procedure, we set:

$$\mathbf{F}(\mathbf{U}) = \mathbf{F^p}(\mathbf{U}) + (\mathbf{F}(\mathbf{U}) - \mathbf{F^p}(\mathbf{U})), \quad \mathbf{B}(\mathbf{U}) = \mathbf{B^p}(\mathbf{U}) + (\mathbf{B}(\mathbf{U}) - \mathbf{B^p}(\mathbf{U})), \tag{11}$$

where $\mathbf{F}^p(\mathbf{U})$ and $\mathbf{B}^p(\mathbf{U})$ are respectively a suitable numerical flux and a source term, further defined. After integration in time, system (10) can be split in the two following ones:

$$\mathbf{U}^{k+1/2} - \mathbf{U}^k + \nabla \cdot \int_0^{\Delta t} \mathbf{F}^p \, dt = \int_0^{\Delta t} \mathbf{B}^p \, dt, \tag{12a}$$

$$\mathbf{U}^{k+1} - \mathbf{U}^{k+1/2} + \nabla \cdot \int_0^{\Delta t} \mathbf{F} \, dt - \nabla \cdot \overline{\mathbf{F}}^p \Delta t = \int_0^{\Delta t} \mathbf{B} \, dt - \overline{\mathbf{B}}^p \Delta t, \tag{12b}$$

where $\overline{\mathbf{F}}^p$ and $\overline{\mathbf{B}}^p$ are the mean in time values of the numerical flux and source terms computed along the prediction step, $\mathbf{U}^{k+1/2}$ and$\mathbf{U}^{k+1}$ are the unknown variables computed respectively at the end of the prediction and the correction phase. $\overline{\mathbf{F}}^p$ and $\overline{\mathbf{B}}^p$ will be estimated "a posteriori" after the solution of the prediction problem, according to the procedure explained in the next section. We call systems (12a) and (12b) prediction and correction systems respectively. Observe that summing systems (12a) and (12b), the integral of the original system (10) is formally obtained. The difference between $\mathbf{U}^{k+1}$ and $\mathbf{U}^{k+1/2}$ in Eq. (12b) is close to zero as far as the difference between the predicted and mean in time values of the fluxes and source terms is either small or time-independent. The advantage of using formulations (12) instead of (10) is that, with a suitable choice of the prediction terms $\mathbf{F}^p(\mathbf{U})$ and $\mathbf{B}^p(\mathbf{U})$, each of the two systems (12a) and (12b) can be much easier to solve than the original system (10).

To keep the same structure of Eqs. (10), (9) can be rewritten as:

$$\frac{\partial H}{\partial t} - \frac{1}{\alpha(\psi)} \nabla \cdot (K_s K_r(\theta(\psi)) \cdot \nabla H) = \frac{Q}{\alpha(\psi)}. \tag{13a}$$

Observe that formulation (13a) fails if the porous medium becomes incompressible and saturated ($\alpha(\psi) = 0$). See in Section 4.3.1 the treatment of saturated condition and the transition from unsaturated to saturated condition with negligible specific storage coefficient S_0.

By setting:

$$K'_r(\theta(\psi)) = \frac{K_r(\theta(\psi))}{\alpha(\psi)}, \quad Q' = \frac{Q}{\alpha(\psi)}, \tag{13b}$$

in the present case we have:

$$\mathbf{U} = H, \quad \mathbf{F} = K_s K'_r(\theta(\psi)) \nabla H, \quad \mathbf{B} = Q'. \tag{14}$$

We set:

$$\mathbf{F}^p = K_s K'_r(\theta(\psi))(\nabla H)^k, \tag{15}$$

$$\mathbf{B}^p = \mathbf{B}, \tag{16}$$

where index k marks the beginning of the time step (time level t^k). Observe that the flux formulation of the prediction step differs from the original one (Eq. (13a)) in the time level of the gradients of H. In the prediction step, spatial gradients of the piezometric head are assumed constant in time and equal to the values computed at the end of the previous time step.

The prediction problem is solved in its integral form, as shown in the following, while the correction problem is solved in its differential linearized form:

$$\frac{\partial H}{\partial t} - (K_s \overline{K'}_r(\theta(\psi)))\nabla^2 H = -\nabla \cdot [(K_s \overline{K'}_r(\theta(\psi))) \cdot (\nabla H)^k], \tag{17a}$$

with

$$\overline{K'}_r(\theta(\psi)) = \frac{\overline{K}_r(\theta(\psi))}{\bar{\alpha}(\psi)}, \quad \bar{\alpha}(\psi) = \left(S_0 \frac{\bar{\theta}(\psi)}{\varepsilon} + \overline{\frac{\partial \theta}{\partial \psi}} \right), \tag{17b}$$

where $\bar{(\cdot)}$ is the mean in time operator of the solution of the prediction problem. A simple first order quadrature formula is applied.

Observe that Eq. (17a) can be seen as a PDE with a source term given by its right hand side, but also as a PDE with a zero source term and a flux F given by:

$$F = -K_S \overline{K'}_r(\theta(\psi))(\nabla H - \nabla H^k). \tag{17c}$$

This implies that fluxes, in the correction step, go to zero along with the time step size or the solution time variability. In the same conditions, the error associated with the solution of the correction component becomes small with respect to the error associated to the solution of the prediction component.

After some simple manipulations, it can be shown that the quasi-linear differential form of the prediction problem is kinematic, with only one bi-characteristic line passing through each $(\mathbf{x},t)$ point. The prediction PDE system is equivalent to a single non-linear convection equation, function of the gradient of the piezometric head at time level t^k. The correction system has the functional characteristics of a pure diffusive process. For these reasons we call the prediction and the correction systems respectively convective prediction system and diffusive correction system.

4. The MAST procedure

Spatial discretization of the governing PDE (13a) is based on a generally unstructured triangular mesh. Let $\Omega \subset \Re^2$ be a bounded domain, Ω_h a polygonal approximation of Ω and T_h an unstructured Delaunay-type non overlapping triangulation of Ω_h. The triangulation T_h is called *basic mesh*, N_T is the number of triangles of T_h, T_i, $i = 1,\ldots,N_T$ is the generic triangle of T_h (black dashed lines in Fig. 1(a)) and $|T_i|$ is the area of T_i. Let $P_h = \{P_i, i = 1,\ldots,L\}$ be the set of the midpoints of the edges $e^j_{T_i}$ $(j = 1,2,3)$ of all $T_i \in T_h$ and L the number of the sides in the mesh. A *dual mesh* $E_h = \{e_i, i = 1,\ldots,L\}$ is constructed over the basic mesh. The dual finite volume e_i associated with the midpoint P_i, $i = 1,\ldots,L$, is the closed polygon given by the union of the sub-triangles sharing side i and bounded by the orthogonals to each side passing through the opposite node in the basic mesh (blue polygons in Fig. 1(a)). In the following of the paper the dual volumes e_i are called also cells or control volumes and the edge midpoints, computational nodes. The storage capacity is assumed to be concentrated only in the cells, in the measure of 1/3 of the area of each triangle (one or two) sharing the cell edge.

A linear variation of the piezometric head inside each triangle is assumed on the base of the values at midpoints of the three sides. Piezometric head is also assumed continuous at midpoints. According to the linear variation of H, it is possible to compute the flux from node i to node j (as well as from node i to node k) of the same triangle as the flux through the common side of the two corresponding sub-triangles (see Fig. 1(b)).

After integration of the prediction equations in space, the differential form of the prediction system is:

$$A_i \frac{dH_i}{dt} + \sum_j Fl^{out}_{i,j} = A_i Q'_i + \sum_l Fl^{in}_{i,l}, \quad i = 1,\ldots,L \tag{18a}$$

with

$$A_i = \frac{1}{3}\sum_{N_T} |T_m|\delta_{im}, \tag{18b}$$

where A_i is the area of the ith computational cell, δ_{im} is the Kronecker delta and it is equal to 1 if side i belongs to triangle T_m, 0 otherwise. $Fl^{out}_{i,j}$ is the flux leaving from cell i to any neighbouring cell j with lower potential value at time level t^k, $Fl^{in}_{i,l}$ is the flux entering in cell i from any neighbouring cell l with higher potential value at the same time level.

Solution of the prediction problem (18) can be strongly simplified if we change the second term on the r.h.s. of each equation with its mean value along the given time step, according to:

$$A_i \frac{dH_i}{dt} + \sum_j Fl^{out}_{i,j} = A_i \overline{Q'}_i + \sum_l \overline{Fl}^{in}_{i,l}, \quad i = 1,\ldots,L, \tag{19}$$

where $\overline{Fl}^{in}_{i,l}$ is the mean in time value of the incoming flux $Fl^{in}_{i,l}$, known from the solution of the previously solved cells, as further specified and $\overline{Q}'_i$ is the mean in time value of Q'_i. In this case, the system can be solved using the marching in space and time (MAST) procedure: the cells are ordered according to their potential values at the beginning of the time step and Eq. (19) are then solved sequentially, one after the others. After the ODE in cell i is solved, the mean total flux $\overline{Fl}^{out}_i$ leaving from cell i along the time step is computed from the local mass balance, that is:

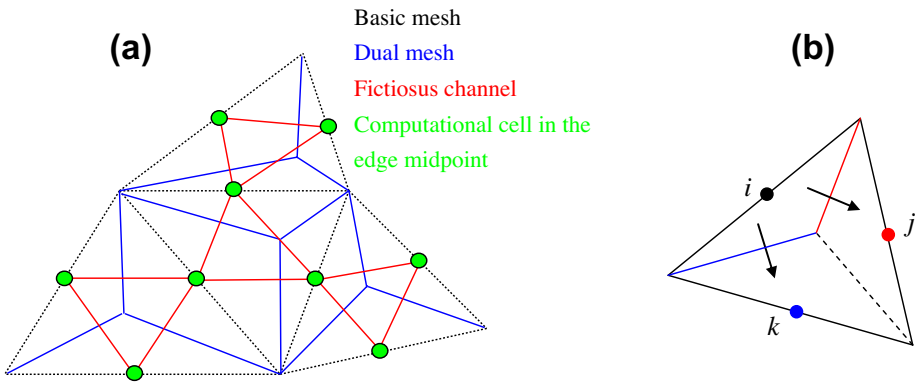

Fig. 1. (a) The computational mesh, (b) flux between cell i and j (k) inside the same triangle.

$$\overline{Fl}_i^{out} = \sum_l \overline{Fl}_{i,l}^{in} - A_i \frac{\theta_i^{k+1/2} - \theta_i^k}{\Delta t} + A_i \overline{Q'}_i, \tag{20}$$

where θ is the volumetric water content and index $k + 1/2$ is assigned to the values computed at the end of the prediction step.

Once the total mean leaving flux is computed, the mean flux $\overline{Fl}_{i,j}^{out}$ going from cell i to the neighbouring cell j with $H_i^k > H_j^k$ can be estimated by partitioning $\overline{Fl}_i^{out}$ according to the ratio between the flux $Fl_{i,j}^{out}$ and the sum of the leaving fluxes at the end of the time step, that is:

$$\overline{Fl}_{i,j}^{out} = \frac{\left(Fl_{i,j}^{out}\right)^{k+1/2}}{\sum_j \left(Fl_{i,j}^{out}\right)^{k+1/2}} \overline{Fl}_i^{out}. \tag{21}$$

Finally you set:

$$\overline{Fl}_{j,i}^{in} = \overline{Fl}_{i,j}^{out} \tag{22}$$

for all the neighbouring cells with $H_i^k > H_j^k$ and you can proceed to solve Eq. (19) for the next cell, that has among the unsolved ones the maximum potential value smaller than or equal to H_i^k.

The basic idea of the proposed numerical technique is to compute the solution, within a given time step, by marching in space along the velocity direction throughout the computational domain, computing the solution for each cell, one after the other according to the potential scale. MAST scheme can be classified as "explicit", because the solution in each cell depends only on the initial state in the cell and on the incoming information (i. e. the flux) from the upstream (in the potential scale) cells, previously solved.

A major property of the MAST solution is to guarantee the local mass balance even if the relationship between the potential and the water content is strongly non-linear, because the mean (along Δt) fluxes leaving from each cell are computed according to the initial and final state inside the cell and not by numerical integration.

All the soil parameters of the closure relationships (6) or (7), as well as the saturation conductivity are assumed constant in space inside each triangle of the mesh.

Define Y_h the finite element approximate space of the linear functions over each triangle T_j, $j = 1,\ldots,N_T$, continuous at midpoint B_i of each side i, $1 \leqslant i \leqslant L$. $\omega_{ih}(\mathbf{x})$ are piecewise linear basis functions of Y_h, globally defined by the values assumed at midpoints of the triangle sides, so that $\omega_{ih}(B_j) = \delta_{ij}$, where i and j are indexes of sides ($i, j = 1,\ldots,L$) and δ_{ij} is the Kronecker delta (see for example [13]). Piezometric head $H \in Y_h$.

After discretization also in time, Eqs. (17) become:

$$A_i \frac{H_i^{k+1} - H_i^{k+1/2}}{\Delta t} + \sum_{T_j} \int_{T_j} \left(\left(K_s^j \overline{K'}_r(\theta(\psi)) \right) \nabla H \nabla \omega_{ih} \right) dT_j = \sum_{T_j} \int_{T_j} \left(\left(K_s^j \overline{K'_r}(\theta(\psi)) \right) (\nabla H)^k \nabla \omega_{ih} \right) dT_j, \tag{23}$$

where K_s^j is the saturated conductivity inside each triangle T_j. The mean in time value $\overline{K'}_r(\theta(\psi))$ is computed via first order quadrature formula of the values of $K'_r(\theta(\psi))$.

A fully implicit time discretization has been chosen for the solution of system (23). This guarantees unconditional stability with regard to the time step size [29]. Fully implicit time discretization provides an approximation error that is proportional to the size of the correction and becomes small along this last one.

Observe that the difference between the two integrals in Eq. (23), as well as the piezometric head correction, goes to zero along with the time step size. Even during abrupt potential changes, the potential correction will be small with respect to the predicted change. This implies that the absolute error in the estimation of the piezometric head correction will only weakly affect the piezometric final value computed at time level $k + 1$.

4.1. Flux computation in acute triangles

Call P_j^m the jth node of triangle m and $\overrightarrow{l_j^m}$ the edge vector following in counterclockwise direction. Call P_m the orthocentre, intersection of the three lines normal to each edge $\overrightarrow{l_j^m}$ and passing through P_{jm}^m, the node opposite to $\overrightarrow{l_j^m}$ (see Fig. 2).

Because the storage capacity is assumed to be entirely concentrated in the computational nodes, it is possible to compute the flux through each side $\overrightarrow{l_j^m}$ as the sum of the fluxes trough the $P_{jp}^m - P_m$ and the $P_m - P_j^m$ subintervals, named respectively $Fl_{i,ip}^{out}$ and $Fl_{i,im}^{out}$. Because the line through cells i and im is orthogonal to the subinterval $P_m - P_j^m$ the crossing flux $Fl_{i,im}^{out}$ going from i to im is proportional to the difference between the two potentials:

$$Fl_{i,im}^{out} = -rc_{j,jm}^m K'_r(\theta(\psi_i)) \left(H_{im}^k - H_i^k \right), \tag{24a}$$

with

$$K'_r(\theta(\psi_i)) = \frac{K_r(\theta(\psi_i))}{\alpha(\psi_i)} \tag{24b}$$

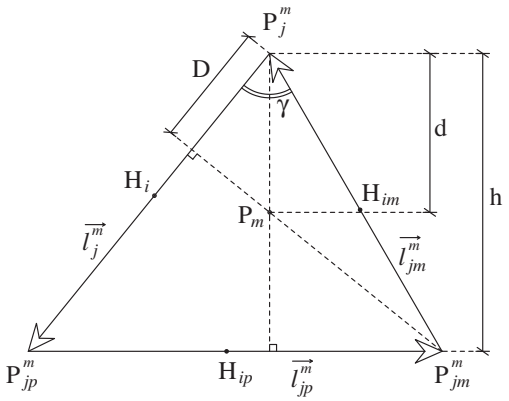

Fig. 2. A triangular element – notations.

and

$$rc^m_{j,jm} = -\frac{\overrightarrow{l^m_j}\cdot\overrightarrow{l^m_{jm}}}{A_m}K^m_s, \tag{24c}$$

where jm is the index of the node preceding j in counterclockwise direction, the scalar product in Eq. (24c) is between the edge vectors sharing node P^m_j, upper index m is assigned to the quantities related to triangle m, $H_{im(i)}$ is the piezometric head in cell im (i), located in the midpoint of edge $\overrightarrow{l^m_{jm}}\left(\overrightarrow{l^m_j}\right)$, and $\cdot$ is the symbol of scalar product. In the prediction problem the piezometric head difference on the r.h.s. of Eq. (24a) is kept constant and equal to the value computed at the end of the previous time step.

The coefficient $rc^m_{j,jm}$ in Eq. (24c) can be computed according to the following equalities (see Fig. 2):

$$rc^m_{j,jm} = -\frac{\overrightarrow{l^m_j}\cdot\overrightarrow{l^m_{jm}}}{A_m}K^m_s = \frac{2Dl^m_j}{l^m_{jp}h}K^m_s = \frac{2Dl^m_j}{l^m_{jp}l^m_j\cos\gamma}K^m_s = \frac{2D}{l^m_{jp}\cos\gamma}K^m_s = \frac{2d}{l^m_{jp}}K^m_s. \tag{25}$$

Observe that the flux between nodes i and im is computed in Eq. (24a) as function of the soil parameters in the upstream node i. According to the upwind estimation of the relative conductivity, the K'_r value is adjusted during the time step in the solution of the prediction problem as function of the $\psi(H)$ value. This flux spatial discretization can be viewed as the reduction of the original 2D problem to a 1D network of cells (the blue polygons in Fig. 1(a)) located in the midpoint of each edge (green points in Fig. 1(a)) and linked by fictitious channels (red lines in Fig. 1(a)).

Observe that $rc^m_{j,jm}$ coefficient in Eqs. (24) is always positive, unless the jth angle of the triangle m is obtuse. In this case, the oriented fluxes given by Eq. (24a) inside the triangle m can form a loop. This is inconsistent with the irrotationality of the velocity field and hampers the sequential solution of the ODEs in each cell (see an example in Fig. 3). To avoid this inconvenient, a special treatment of the obtuse triangles is required, as better described in the following section.

4.2. Flux computation in obtuse triangles

Assume j to be the vertex of the obtuse triangle m corresponding to the maximum angle, jp and jm respectively the following and the preceding one in couterclockwise direction. Call i, ip and im the index of the corresponding cells. Define two auxiliary internal sides $\overrightarrow{l^m_{j_1}}$ and $\overrightarrow{l^m_{jm_1}}$, respectively orthogonal to sides $\overrightarrow{l^m_{jm}}$ and $\overrightarrow{l^m_j}$, sharing vertex j (see Fig. 4). In the case shown in Fig. 4, the coefficient $rc^m_{j,jm}$ is negative, while the other two coefficients are positive. Assume $rc^m_{j,jm} = 0$, so that the fluxes through the edges $\overrightarrow{l^m_j}$ and $\overrightarrow{l^m_{jm}}$ are proportional only to the difference $\left(H^k_{ip} - H^k_i\right)$ and $\left(H^k_{ip} - H^k_{im}\right)$ respectively. The fluxes $Fl^m_{j,jp1}$ and $Fl^m_{j,jm1}$ through sides $\overrightarrow{l^m_{j1}}$ and $\overrightarrow{l^m_{jm1}}$ are given by:

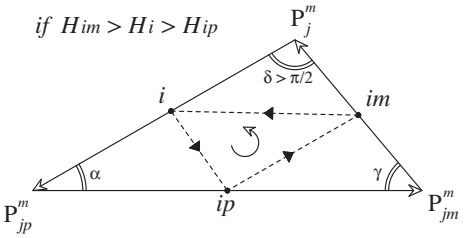

Fig. 3. Example of obtuse triangle with oriented fluxes inconsistent with the velocity irrotationality.

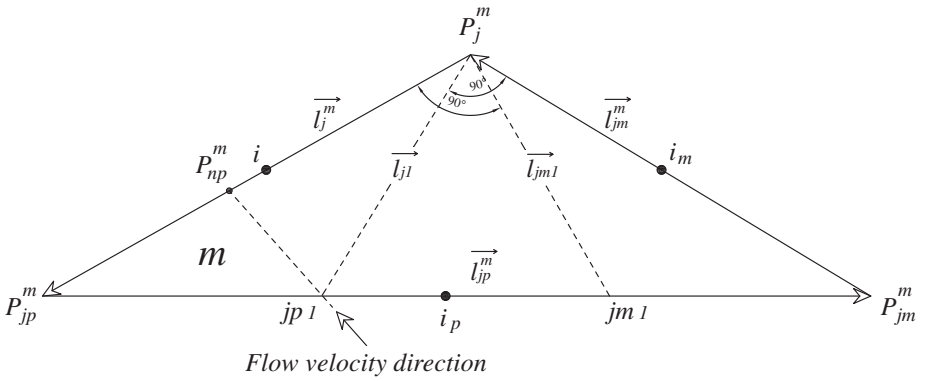

Fig. 4. Triangular element scheme and notations for obtuse triangle.

$$Fl_{j,jp_1}^m = -r'_{j,jp_1}\left(H_{ip}^k - H_i^k\right), \quad r'_{j,jp_1} = \frac{2l_{j,jp_1}^m}{l_{jm}^m}K_s^m, \tag{26a}$$

$$Fl_{j,jm_1}^m = -r'_{j,jm_1}\left(H_{ip}^k - H_{im}^k\right), \quad r'_{j,jm_1} = \frac{2l_{j,jm_1}^m}{l_j^m}K_s^m. \tag{26b}$$

On the other hand, since a linear variation of the piezometric heads has been assumed inside each triangle m, flow velocity is constant along all side $\overrightarrow{l_j^m}$. For instance, given the flow velocity direction shown in Fig. 4, call P_{np}^m the intersection point between the velocity vector drawn from point jp_1 and side $\overrightarrow{l_j^m}$, l_j^m the distance of P_{jp}^m from P_j^m and l_{np}^m the distance of P_{np}^m from P_j^m measured in counterclockwise direction. Position of point P_{np}^m depends on the direction of the element velocity and is updated after each time step. Flux through side $\overrightarrow{l_{j,np}^m}$ (between points P_j^m and P_{np}^m) is equal to Fl_{j,jp_1}^m. Because velocity is constant along the all side $\overrightarrow{l_j^m}$, the total flux $Fl_{i,ip}^{out}$ can be computed as:

$$Fl_{i,ip}^{out} = Fl_{j,jp_1}^m \frac{l_j^m}{l_{j,np}^m} = -\left(H_{ip}^k - H_i^k\right)\frac{2l_{j,jp_1}^m}{l_{jm}^m}K_s^m\frac{l_j^m}{l_{j,np}^m} = -r''_{j,jp_1}\left(H_{ip}^k - H_i^k\right), \tag{27a}$$

$$\text{where} \quad r''_{j,jp_1} = r'_{j,jp_1}\frac{l_j^m}{l_{j,np}^m}. \tag{27b}$$

Coefficient r''_{j,jp_1} tends to infinity as point P_{np}^m approaches node j and takes a negative value if the velocity direction is between the directions $\overrightarrow{l_j^m}$ and $\overrightarrow{l_{jp_1}^m}$. In this case, a flux estimation error (and a corresponding reduction of the solution accuracy) is unavoidable using a fixed mesh. To guarantee the computational efficiency and the r''_{j,jp_1} positive sign, the following approximation is made:

$$l_{j,np}^m \approx \max\left(\varepsilon_0, \left|l_{j,np}^m\right|\right), \tag{28}$$

where the two arguments of the max function are a small positive number and the absolute value of the $l_{j,np}^m$ distance.

4.3. Numerical solution of the prediction and correction steps

The ODE (19) in the prediction step is integrated along the original time step using a 5th order Runge–Kutta method with adaptive step-size control [32]. According to the flux definitions given in Eqs. (24) and (25), the ODE for cell i is written as:

$$A_i\frac{dH_i}{dt} + \sum_{m=1,\ldots,N_T}\left[\sum_{ja=jm,jp}\left[rc_{j,ja}^m K'_r(\theta(\psi_i))\left(H_i^k - H_{ia}^k\right)\delta_{i,ia}^m\right]\right] = A_i\overline{Q'}_i + \overline{Fl}_i^{in}, \tag{29}$$

where j is the local index of triangle m corresponding to cell i, ia is the cell index corresponding to the local index ja, $\delta_{i,ia}^m$ is equal to 1 if cells i and ia belong to triangle m and $H_i^k > H_{ia}^k$, 0 otherwise.

After the ODE for cell i is solved along the time step, the mean in time flux values from cell i to cell ia is estimated as explained in Eqs. (20) and (21).

The solution of a linear system associated to each Eq. (23) is required for the correction step. The order of the system is equal to the number L of mesh sides. A preconditioned conjugate gradient method is used to solve the system.

According to the fluxes definitions given in Eqs. (24) and (23) can be written as:

$$A_i \frac{H_i^{k+1} - H_i^{k+1/2}}{\Delta t} + \sum_{m=1,\dots,N_T} \left[\sum_{ja=jm,jp} \left[\overline{K'}_r(\theta(\psi_s)) rc_{j,ja}^m \left(H_i^{k+1} - H_{ia}^{k+1} \right) \delta_i^m \right] \right]$$
$$= \sum_{m=1,\dots,N_T} \left[\sum_{ja=jm,jp} \left[\overline{K'}_r(\theta(\psi_s)) rc_{j,ja}^m \left(H_i^{k} - H_{ia}^{k} \right) \delta_i^m \right] \right], \quad i = 1, \dots, L, \tag{30}$$

$$s = \begin{cases} i & \text{if } H_i^k > H_{ia}^k \\ ia & \text{if } H_i^k < H_{ia}^k \end{cases}$$

where δ_i^m is equal to 1 or 0 according if cell i belongs to element m or not and the other symbols have been previously defined. System matrix is symmetric with 2 or 4 extra-diagonal elements, respectively for boundary or internal cells.

The numerical structure of the fluxes is the same in the prediction and in the correction steps. The piezometric head difference on the r. h. s. of Eqs. (24a) and (25) is computed at time level k in the prediction step, at time level $k+1$ in the correction step (where they are the unknowns of the problem). The relative conductivity K_r is computed in both Eqs. (29) and (30) as function of the $\theta(\psi)$ value in the upstream cell. This implies that the solution of Eq. (30), in steady-state conditions, vanishes for any possible value of the adopted time step.

In the correction problem, matrix system is positive definite and has the so called *M*-property (positive diagonal elements and negative or null extra-diagonal elements), because both the mean permeability $\overline{K'}_r$ and $rc_{j,ja}^m$ coefficients are always positive. The *M*-property guarantees the existence of an asymptotic monotonic solution at any point with zero forcing function [44]. The procedure proposed in the previous Section 4.2 guarantees the *M*-property of the matrix system also in the case of obtuse triangles.

4.3.1. Saturated condition and transition from unsaturated to saturated condition in the prediction and correction steps

In most cases, the specific storage coefficient S_0 is small with respect to the water content derivative in Eq. (1). When the soil is saturated during all the time step, both time derivatives in the same Eq. (1) are negligible and the problem becomes locally elliptic.

In this case the solution of the prediction problem is trivial, because in each cell the total leaving flux is equal to the total incoming flux. A more detailed procedure is required when the cell remains saturated only during a fraction of the total time step, that is when only one between the initial and the final pressure is greater than the minimum saturation value.

If the cell is unsaturated at the beginning of the time step, solution of the ODE (29) is carried out as described in the previous sections, from time t^k to the time $\tilde{t}$ when saturation is attained. The average incoming flux $\overline{Fl}_i^{in} \cdot ((t^k + \Delta t - \tilde{t})/\Delta t)$, left out in the prediction step, is then added to the source term of the system in the corrective step in order to save the local mass balance inside the cell. Equation of this cell in the correction system preserves the same structure as in Eq. (30), plus the additional source term.

If the cell is saturated at the beginning of the time step, the problem is solved as follows. In the prediction step, if the potential of all or some of the neighbouring cell is lower than the potential of the considered cell, the incoming flux is transferred to the lower potential cells as described in the previous section, neglecting the storage term in Eq. (20). If all the neighbouring cells have higher potential, the incoming flux is added to the source term in the solution of the correction system. Storage term is zero in the equation of this cell in the correction system.

4.3.2. A semi-analytical procedure for the solution of the prediction problem

As previously stated, solution of the ODE (19) can be easily found using a numerical approach. If the specific storage S_0 is negligible with respect to the specific capacity $\partial\theta/\partial\psi$ and the Brooks–Corey model (Eqs. (7)) is adopted as closure relationship, an approximated semi-analytical solution can also be found.

According to Eqs. (7a) and (8) we can write:

$$\frac{\partial\theta}{\partial\psi} = (\theta_s - \theta_r)\frac{\partial\theta^*}{\partial\psi} = (\theta_s - \theta_r)\lambda \left(\frac{|\psi_b|}{|\psi|} \right)^{\lambda} \frac{1}{|\psi|} \tag{31}$$

and, after simple manipulations, the convective step in Eq. (29) can be written as:

$$A_i \frac{dH_i}{dt} \left((\theta_s - \theta_r)\lambda \left(\frac{|\psi_b|}{|\psi_i|} \right)^{\lambda} \frac{1}{|\psi_i|} \right) + \sum_{m=1,\dots,N_T} \left[\sum_{ja=jm,jp} \left[rc_{j,ja}^m K'_r(\theta(\psi_i)) \left(H_i^k - H_{ia}^k \right) \delta_{i,ia}^m \right] \right] = \overline{Fl'}_i^{in}, \tag{32}$$

where

$$\overline{Fl'}_i^{in} = A_i \overline{Q'}_i + \overline{Fl}_i^{in}. \tag{33}$$

Eq. (32) can be written as:

$$k_1|\psi_i|^{\alpha}\frac{dH_i}{dt}+k_2|\psi_i|^{\beta}=\overline{Fl'}_i^{in}, \tag{34}$$

where

$$k_1=A_i\lambda(\theta_s-\theta_r)|\psi_b|^{\lambda},\quad k_2=K_{is}|\psi_b|^{\lambda p^*}, \tag{35a}$$

$$K_{is}=\sum_{m=1,\dots,N_T}\left[\sum_{ja=jm,jp}\left[rc_{j,ja}^{m}\left(H_i^k-H_{ia}^k\right)\delta_{i,ia}^{m}\right]\right], \tag{35b}$$

$$\alpha=-(\lambda+1),\quad \beta=-\lambda p^*. \tag{35c}$$

Observe that both exponents α and β are negative.

Call ψ_i^k the value of ψ at the beginning of the time step in cell i and ψ_i^{kf} its asymptotic value calculated according to Eq. (34), that is:

$$\left|\psi_i^{kf}\right|=\left(\frac{\overline{Fl'}_i^{in}}{k_2}\right)^{\frac{1}{\beta}}. \tag{36}$$

Eq. (34) can be written in dimensionless form as:

$$\xi^{\alpha}\frac{d\xi}{d\tau}=\xi^{\beta}-\xi_f^{\beta}\quad \text{if } \psi_i^{kf}>\psi_i^k, \tag{37a}$$

with

$$\tau=\frac{t\cdot\overline{Fl'}_i^{in}}{k_1\cdot\left|\psi_i^k\right|^{\alpha+1}}\left(\frac{\left|\psi_i^{kf}\right|}{\left|\psi_i^k\right|}\right)^{\beta}\quad \text{if } \psi_i^{kf}>\psi_i^k, \tag{37b}$$

and with

$$\xi=\frac{|\psi_i|}{\left|\psi_i^k\right|}\quad \text{if } \psi_i^{kf}>\psi_i^k, \tag{37c}$$

or

$$\xi^{\alpha}\frac{d\xi}{d\tau}=\xi^{\beta}-1\quad \text{if } \psi_i^{kf}<\psi_i^k \tag{38a}$$

with

$$\tau=\frac{t\cdot\overline{Fl'}_i^{in}}{k_1\cdot\left|\psi_i^{kf}\right|^{\alpha+1}}\quad \text{if } \psi_i^{kf}<\psi_i^k \tag{38b}$$

and with

$$\xi=\frac{|\psi_i|}{\left|\psi_i^{kf}\right|}\quad \text{if } \psi_i^{kf}<\psi_i^k, \tag{38c}$$

where sub-index f in Eq. (37a) marks the asymptotic value $\xi_f=\left|\psi_i^{kf}\right|/\left|\psi_i^k\right|$. The qualitative solution of the two ODEs (37a) and (38a) are represented in Figs. 5 and 6.

Analytical solutions of Eqs. (37a) and (38a) do not exist, due to the non-integer and negative values of the power exponents α and β. A very computationally expensive implicit transhendent series solution is possible, but a good approximation of Eqs. (37a) and (38a) can also be found with a smaller computational time by setting:

$$\xi=1+(\xi_f-1)\frac{\exp(c_1\tau)+c_2}{\exp(c_1\tau)+c_3}\quad \text{if } \psi_i^{kf}>\psi_i^k, \tag{39}$$

$$\xi=\frac{\exp(c_1\tau)+c_2}{\exp(c_1\tau)+c_3}\quad \text{if } \psi_i^{kf}<\psi_i^k \tag{40}$$

with a proper choice of the c_1, c_2 and c_3 coefficients. Using any c_3 value, it is possible to match the initial value ξ_0 and its first derivative ξ_0' by setting:

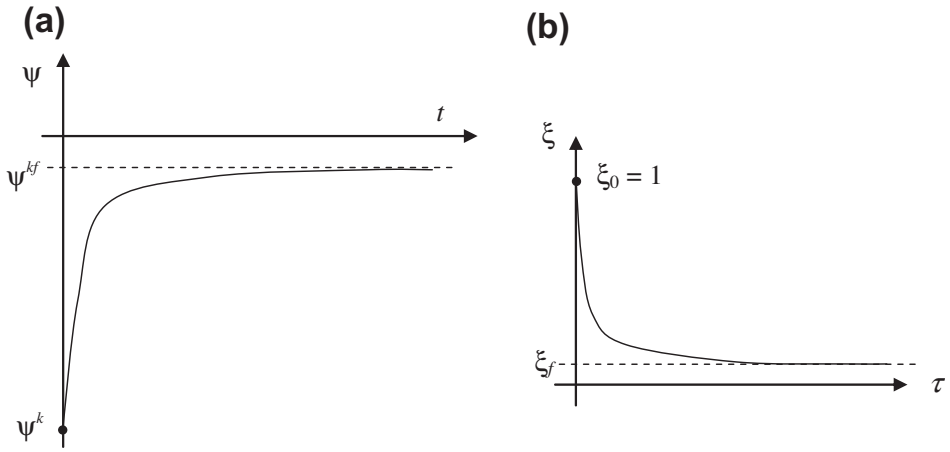

Fig. 5. The case $\psi_i^{kf} > \psi_i^k$.

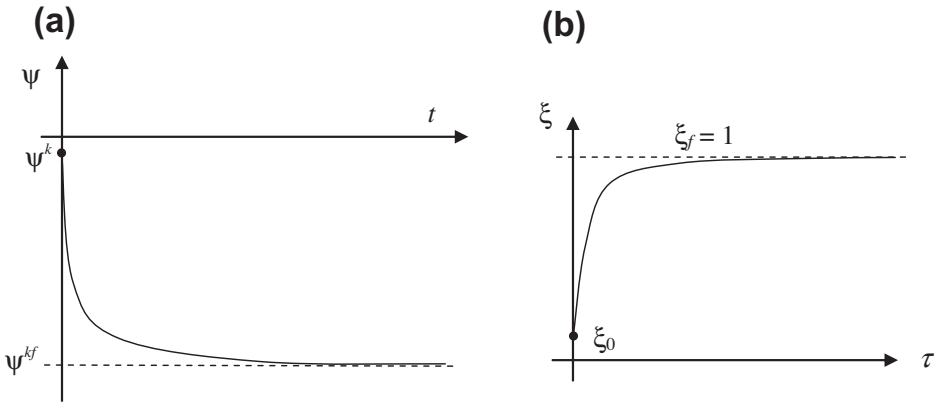

Fig. 6. The case $\psi_i^{kf} < \psi_i^k$.

$$c_2 = \frac{(\xi_0 - 1)(1 + c_3)}{(\xi_f - 1)} - 1, \quad c_1 = \xi_0' \frac{(1 + c_3)^2}{(c_3 - c_2)} \frac{1}{(\xi_f - 1)} \quad \text{if } \psi_i^{kf} > \psi_i^k, \tag{41}$$

$$c_2 = \xi_0(1 + c_3) - 1, \quad c_1 = \xi_0' \frac{(1 + c_3)^2}{(c_3 - c_2)} \quad \text{if } \psi_i^{kf} < \psi_i^k. \tag{42}$$

The first derivative ξ_0' can be computed as the r. h. s. of Eqs. (37a) or (38a), for $\xi = \xi_0$. This choice guarantees a second order reduction of the error around the $\tau = 0$ value. Because in the MAST approach the time step is not restricted by the CFL number condition, it is also important to select a c_3 coefficient that provides a small error for any possible τ value.

The research of the optimal c_3 coefficient values can be carried out, for given ξ_f if $\psi_i^{kf} > \psi_i^k$ or ξ_0 if $\psi_i^{kf} < \psi_i^k$, according to an iterative procedure.

Divide the interval $0.0 \leqslant \xi \leqslant 1.0$ in np equal parts $\Delta\xi$. Call η the difference between the semi-analytical and a reference solution and η_{max} a fixed maximum admissible value of η. The reference solution is numerically computed once for ever using a very small time step.

The reference solution is computed for different parameters ξ_f^q (if $\psi_i^{kf} > \psi_i^k$), or ξ_0^q (if $\psi_i^{kf} < \psi_i^k$), where q is an integer ranging from 1 to np and $\xi_f^q = 1 - \sum_{j=1,q}\Delta\xi_j$, $\xi_0^q = \sum_{j=1,q-1}\Delta\xi_j$. For each q value the coefficient c_3^q which maximizes the τ_n^q interval, measured from $\tau = 0$, where $\eta \leqslant \eta_{\max}$, is found via numerical optimization. Compute also the dimensionless time value τ_x^q such that $\eta \leqslant \eta_{\max}$ for any $\tau \geqslant \tau_x^q$ (see in Fig. 7 the case $\psi_i^{kf} > \psi_i^k$).

Start from the initial value ξ_0 at $\tau = 0$. According to Eqs. (37b) or (38b), compute $\tau_{\Delta t}$ (the τ value corresponding to Δt). According to the asymptotic value ξ_f (if $\psi_i^{kf} > \psi_i^k$), or to the initial value ξ_0 (if $\psi_i^{kf} < \psi_i^k$), select the two curves with parameters ξ_f^q and ξ_f^{q+1} (or ξ_0^q and ξ_0^{q+1}), such that $\xi_f^{q+1} \leqslant \xi_f \leqslant \xi_f^q$ (or $\xi_0^q \leqslant \xi_0 \leqslant \xi_0^{q+1}$) (see Fig. 8(a) and (b)). Call $\bar{\tau}_n = \min\left(\tau_n^q, \tau_n^{q+1}\right)$ as well as $\bar{\tau}_x = \max\left(\tau_x^q, \tau_x^{q+1}\right)$. Compute the c_3 coefficient corresponding to the curve parameter $\xi_{f(0)}$, applying a cubic-spline interpolation between c_3^q and c_3^{q+1}. If $\tau_{\Delta t} > \bar{\tau}_n$ and $\tau_{\Delta t} < \bar{\tau}_x$ compute the solution ξ' at $\tau\prime = \bar{\tau}_n$ by means of Eqs. (39)–(42). The corresponding piezometric depth is the new initial conditions for the next iteration. At each iteration, compute $|\psi'|$ according to ξ', set $|\psi'| = |\psi_0|$, $\xi' = \xi_0$, $\tau' = 0$, compute the new curve parameter $\xi_{f(0)}$ and $\tau_{\Delta t}$ according to ξ' as well as the new couples ξ_f^q and

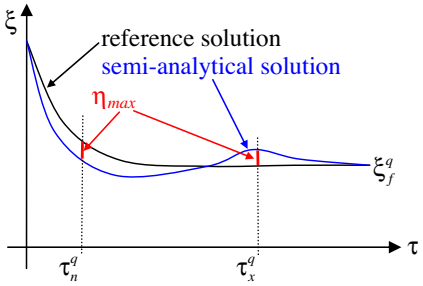

Fig. 7. Estimation of τ_n^q and τ_x^q in the case $\psi_i^{kf} > \psi_i^k$.

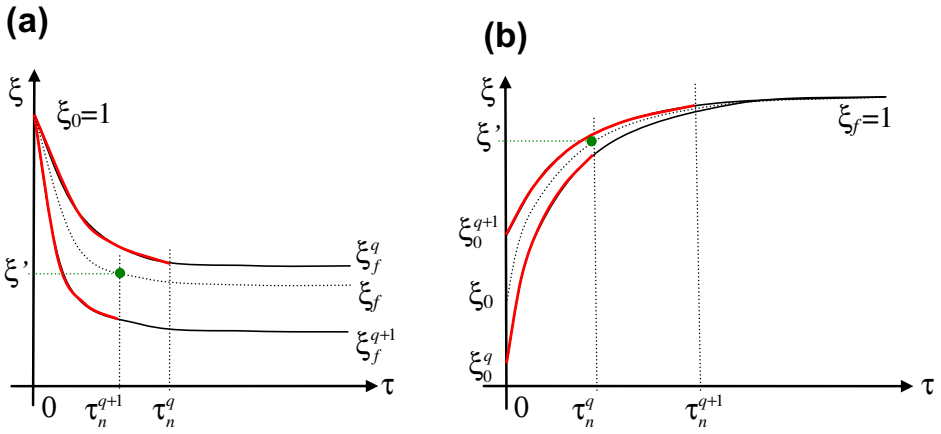

Fig. 8. The iterative procedure in the case $\psi_i^{kf} > \psi_i^k$ (a) and $\psi_i^{kf} < \psi_i^k$ (b).

ξ_f^{q+1} (or ξ_0^q and ξ_0^{q+1}), and repeat the above procedure until $\tau_{\Delta t} \leqslant \Delta\bar{\tau}$ or $\tau_{\Delta t} \geqslant \bar{\tau}_x$. After solution of the ODEs (37a) or (38a), the mean in time value of the flux from cell i to any connected cell im with $H_{im}^k < H_i^k$ is computed as in Eqs. (20) and (21).

The semi-analytical approach can be applied only if saturation is not reached during the time step. If the initial pressure is lower than the saturation pressure and the final computed water content is greater than the saturation value, the τ_{sat} value corresponding to ξ_{sat} can be obtained by integrating Eq. (37a) written as:

$$d\tau = \frac{\xi^\alpha}{\xi^\beta - \xi_f^\beta}\, d\xi, \tag{43}$$

in the τ unknown, along the interval from $\xi = \xi_0$ to $\xi = \xi_{sat}$, with initial value $\tau = 0$.

Eq. (43) is numerically solved. Because only a limited number of cells is affected during each time step by the transition from the unsaturated to the saturated condition, this has no significant effect on the total computational burden. In the interval from τ_{sat} to $\tau_{\Delta t}$ the storage capacity does not change any more, since cell i is saturated and the incoming flux is equal to the leaving one. The leaving flux to any j cell is computed according to Eqs. (20) and (21), where Δt is replaced by the fraction of time step corresponding to τ_{sat}.

4.4. Boundary conditions

4.4.1. Boundary conditions for the prediction problem

In the prediction step the Neumann flux is included in the term $\overline{Fl}_i^{in}$ representing the incoming flux in the cell. No boundary conditions are assigned to the part of the boundary where first type (Dirichlet) conditions apply. Observe that first type and second type (Neumann) boundary conditions are assigned at the beginning of the computation only in the part of the boundary that is below the level of the open water (portion Γ_D as described in Eq. (5)) or in the part of the boundary that is impervious (portion Γ_N as described in Eq. (5)). The boundary conditions in each cell of the remaining boundary, facing the open air, are updated at the beginning of each time step according to the pressure value computed at end of the previous time step. If the pressure value in the boundary cell is greater than the external atmospheric pressure, the storage term is neglected and the average flux leaving the domain is given by the difference between the incoming flux and the average flux going to the cells with lower potential. If the value of the pressure in the cell is lower than the external atmospheric value, the boundary is assumed as impervious.

4.4.2. Boundary conditions in the correction problem

Boundary conditions in the correction problem are zero corrective flux passing through the boundary cells where entering fluxes are assigned in the prediction step. In the cell i where the piezometric head is assigned, the Dirichlet condition is:

$$H_i^{k+1} = H_D \quad \text{or} \quad H_i^{k+1} = \psi_D + z_i, \tag{44}$$

where H_D (ψ_D) is the Dirichlet value of the piezometric head (depth) (see also Eq. (5)). Diffusive boundary fluxes are computed a posteriori by means of Eq. (30).

4.5. Similarities of numerical fluxes in LMHFE, linear (P1) nonconforming FEM and MAST correction problem

The linear ($P1$) nonconforming FEM is described in the Appendix A for a time-dependent diffusion problem, with piecewise constant in space storage coefficient term and space dependent symmetric positive definite tensor of the flux term. It is easy to show, after simple algebraic manipulations, that the structure of the fluxes proportional to coefficients $rc_{j,im}^m$ given in Eqs. (24) and (25) is similar to the structure of the fluxes given in the $P1$ nonconforming FEM and this is equivalent to the LMHFE flux discretization (as shown in the Appendix A). The difference is that, while in the $P1$ nonconforming FEM the flux coefficient is assumed constant inside each triangle, and equal to its mean value inside the element, in the MAST procedure the same coefficient is function of the parameters (namely the relative conductivity K_r) of the cell with higher potential inside the element. If the parameters of two neighbouring cells i and im in the same element are equal, in the case of unsaturated medium, the flux between cells i and im is null, since this implies that $H_i^k = H_{im}^k$; in the case of saturated medium the computed flux is exactly the same as the one given by the $P1$ nonconforming FEM.

5. Numerical tests

According to Celia et al. [12], the mass conservation capability of a model can be measured by the Mass Balance Ratio (MBR), defined as:

$$\text{MBR} = \frac{\text{total additional mass in the domain}}{\text{total net flux into the domain}}, \tag{45}$$

where the "total additional mass in the domain" is the difference between the mass measured at any simulation time t and the initial mass in the domain, the "total net flux into the domain" is the flux integrated in time up to time t. In a "perfect" model MBR is equal to 1. According to Celia et al. [12], MBR is function of the simulation time step Δt, and if a pressure based formulation is adopted in the governing equations, it decreases for increasing Δt. Using a water content form or a mixed form for the governing equations, MBR value is very close to 1 for all the investigated Δt range. Similar behaviour has been observed by Manzini and Ferraris [30]. Celia et al. [12] ascribed the poor mass balance of the pressure-based form to the computation of the local time derivative term of θ: while $\partial\,\theta/\partial t$ and $\partial\theta/\partial\psi \cdot\ \partial\psi/\partial t$ are equivalent in the continuous PDE, their discrete analogous are not. The difference between the discrete forms is exacerbated by the high non-linearity of the specific capacity $\partial\theta/\partial\psi$ term. This non-linearity leads to significant mass balance errors in the pressure based formulation because the discrete value of $\partial\theta/\partial t$, approximating the change in mass per unit volume and unit time, is computed using a first order Taylor series expansion of the product $\partial\theta/\partial\psi \cdot \partial\psi/\partial t$.

Five numerical tests are presented in the following. The first three tests are aimed to evaluate the algorithm ability to preserve the mass conservation and to cope with sharp soil heterogeneities. In the second test a mesh refinement has been also carried out in order to investigate the stability of the proposed model with respect to the mesh size. In the third test the properties of the computed velocity field are presented and analyzed.

In the fourth test, computational (CPU) times required by the different steps included in the model (cells ordering, managing of the obtuse triangles, prediction and correction problems) are evaluated using three different structured and unstructured meshes. The CPU times are compared also to the ones obtained using the semi-analytical solution for the prediction step, in order to estimate the computational cost abatement. Finally, the last test is presented in order to asses the order of convergence of the proposed methodology.

The local CFL number is defined as:

$$CFL = \frac{|\mathbf{q}_i|\Delta t}{\sqrt{A_i}}, \tag{46}$$

where $\mathbf{q}_i$ is the velocity vector in cell i. The software Argus one [1] has been used as mesh generator.

5.1. Test 1. Perched water table problem

The problem was originally presented by Kirkland et al. [25] and then adapted by Forsyth et al. [16] and Diersch and Perrochet [14]. It is a 2D problem of a developing perched water table surrounded by very dry unsaturated conditions. The problem is described in Fig. 9. Water infiltrates with a very large rate into a dry soil at initial piezometric depth $\psi_0 = -500$ m and encounters a clay barrier which allows for the formation of a perched water table. All boundaries are impervious (no flow)

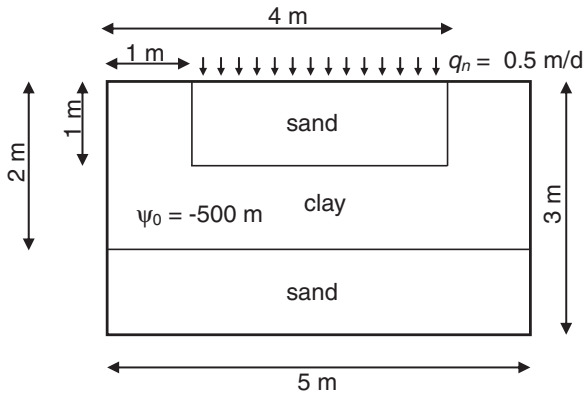

Fig. 9. Test 1 – computational domain, initial and boundary conditions.

except the portion of the upper side where the infiltration is imposed. The material properties of the problem are summarized in Table 1 for the Van Genuchten parametric model. The residual saturation rate s_r corresponds to the residual water content θ_r.

An unstructured mesh with 3412 triangles, 1788 nodes and 5199 edges and a $\Delta t = 100$ s have been used for the simulations of the present model (see Fig. 10). Mesh is not locally refined at material interfaces. In Figs. 11a and 11b the computed piezometric depth and saturation contours are shown for the final simulation time of 1 day (864 time iterations). The maximum value of the CFL number attained during the simulation is 1.85. In Figs. 12a and 12b the piezometric depth contours computed by Diersch and Perrochet [14] and Kirkland et al. [25] are shown, while in Figs. 13a and 13b the saturation contours computed by Diersch and Perrochet [14] and by Forsyth et al. [16] are shown for the same simulation time. Kirkland et al. [25] imposed the initial condition $\psi_0 = -4000$ kPa. Diersch and Perrochet [14] proposed a variable switching procedure for finite elements methods. The technique is incorporated in both an adaptive error-controlled predictor-corrector one-step Newton (PCOSN) iteration strategy and a target-based full Newton (TBFN) iteration scheme. Forsyth et al. [16] presented one-phase and two-phase (active air-phase) Finite Element variable switching model. Kirkland et al. [25] presented a Finite Difference scheme allowing to change the unknown variables in saturated and unsaturated medium. In the reference literature models the authors discretize the half domain with 50×60 quadrilateral elements (3111 nodes). The number of time iterations of the reference literature models ranges from 500 to 3000.

It can be observed in the figures that the MAST computed results and the computed fully saturated volume are dramatically different from the ones computed by the other two literature models. The number of unknowns in the MAST scheme (number of edges) is close to those of the literature models (number of nodes of the quadrilateral elements) and the number of total time iterations of the MAST simulation is close to that of the reference models, eventhough in the model by Diersch and Perrochet [14] a control procedure of the iteration process with a variable time step size is activated in order to change the time step during the simulation.

The MBR at time 1 day computed by the proposed model is about $1 + 1.5 \times 10^{-16}$, with a total mass balance errors $O(10^{-16})$ of the same order of the machine truncation error. Diersch and Perrochet [14] listed a total mass balance errors (TMBE) ranging from $O(10^{-5})$ to $O(10^{-4})$.

A much more fine mesh has also been used for MAST simulations, with 12,090 triangles, 6195 nodes and 18,284 edges and a time step ranging from 0.864 s to 10 s. Results obtained with the different time steps are very similar and the corresponding maximum CFL values range from 0.11 to 2.56. The MBR and the total mass balance error are very similar to the ones obtained for the coarse mesh. Observe in Figs. 14a and 14b the computed piezometric depth and saturation contours at 1 d using $\Delta t = 2$ s (43,200 time iterations). Results obtained with the coarse mesh (see Figs. 11a and 11b) can be considered qualitatively very close to the ones computed with the fine mesh.

The difference between results of the proposed scheme and those obtained by the literature models probably relays in the conservation property of the proposed mass scheme, that is very likely to result also in a more accurate estimation of the piezometric depths. MAST computed piezometric depth contours show indeed sharp discontinuities at zone interfaces, while the corresponding literature contours are surprisingly almost continuous throughout the sand – clay interface in the upper part of the domain.

Table 1
Test 1 – Van Genuchten parameters.

Material	K_s [m/s]	ε [−]	s_r [−]	μ [1/m]	λ [−]	ψ_a [m]
Sand	6.262d−05	0.3658	0.07818	2.8	2.239	0
Clay	1.516d−06	0.4686	0.2262	1.04	1.3954	0

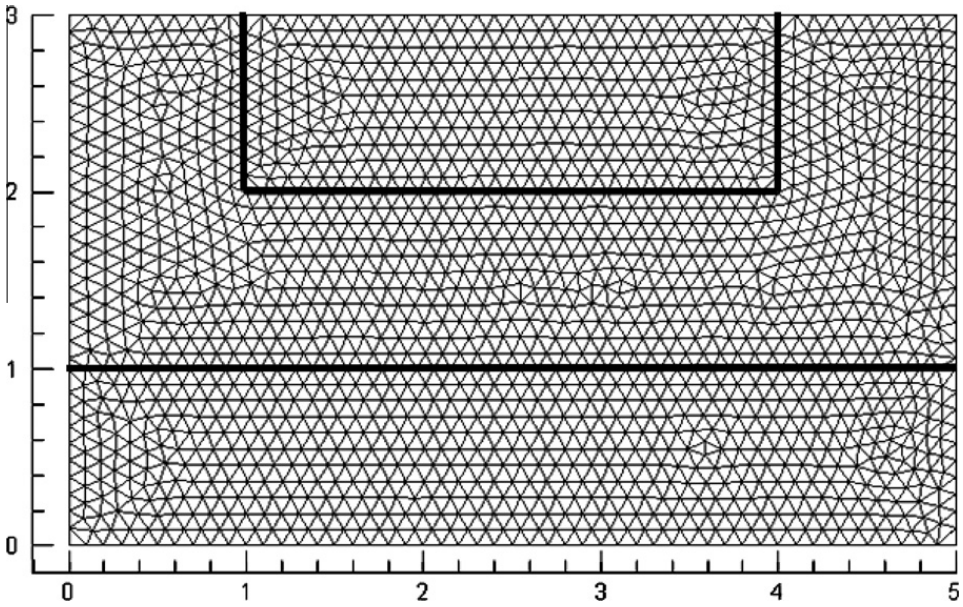

Fig. 10. Test 1 – computational mesh.

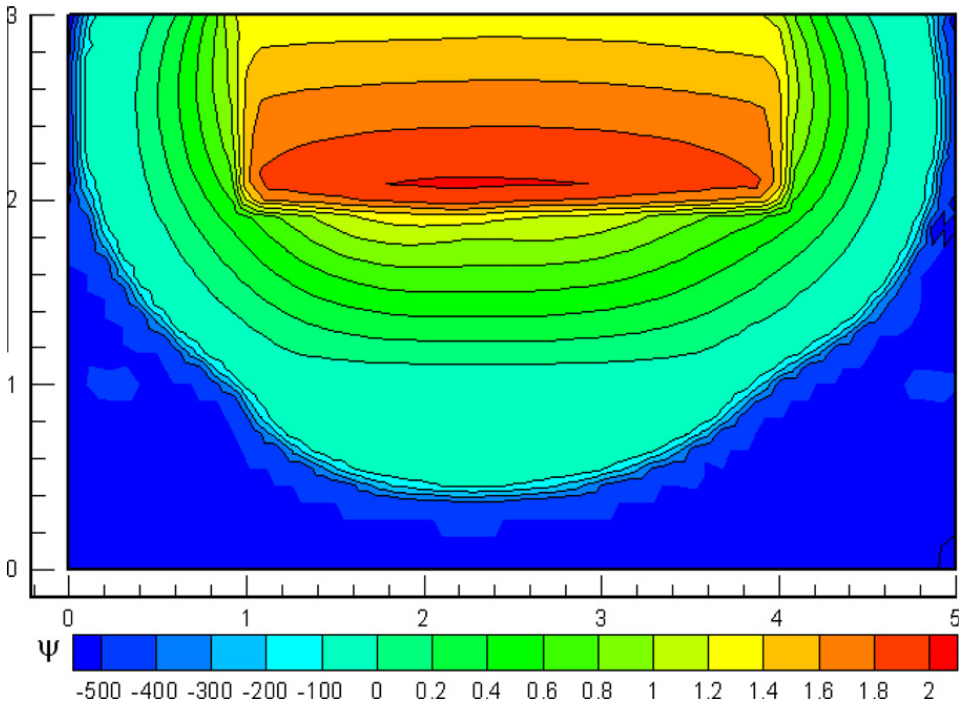

Fig. 11a. Test 1 – MAST computed piezometric depth contours.

5.2. Test 2. Infiltration in a large caisson

5.2.1. Test 2a: Forsyth et al.'s problem [16]

The infiltration process in a large caisson consisting of heterogeneous materials at dry initial conditions has been thoroughly studied by Forsyth et al. [16]. Fig. 15 presents a schematic view of the 2D cross-sectional problem. All boundaries are impervious except the infiltration boundary section on the top side. Two initial piezometric depth conditions of $\psi_0 = -7.34$ m and $\psi_0 = -100$ m are simulated. Table 2 lists the material properties for the different zones of the domain. A mesh with 1148 triangles, 620 nodes and 1767 edges, with no local refinement at material interface, has been used (see Fig. 16). Time step Δt is 3240 s. In Figs. 17a–17c the computed piezometric head, piezometric depth and saturation profiles are shown in the case $\psi_0 = -7.34$ m, for the final simulation time 30 d (800 total time iterations). In Fig. 18, the MAST

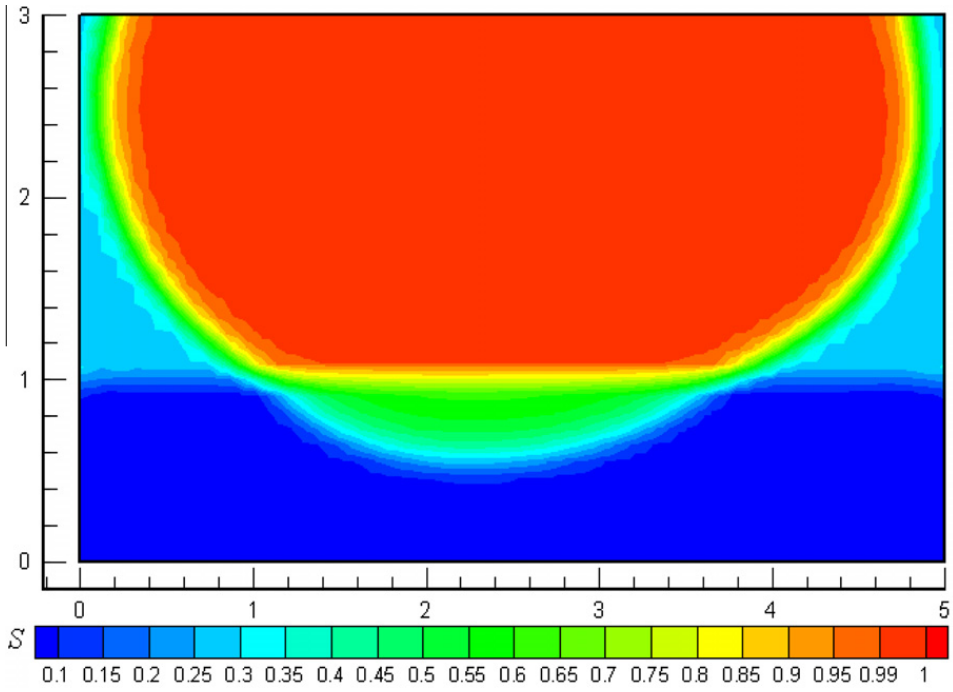

Fig. 11b. Test 1 – MAST computed saturation contours.

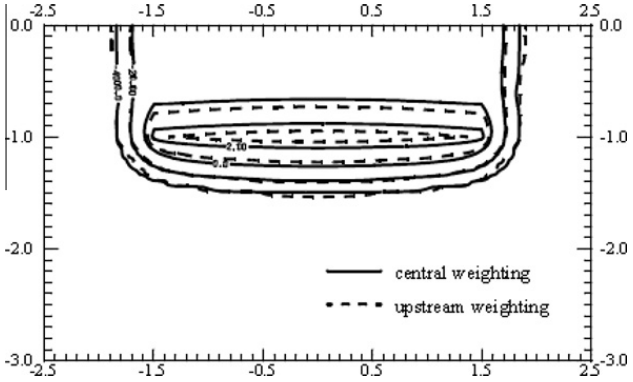

Fig. 12a. Test 1 – computed piezometric depth contours by Diersch and Perrochet [14].

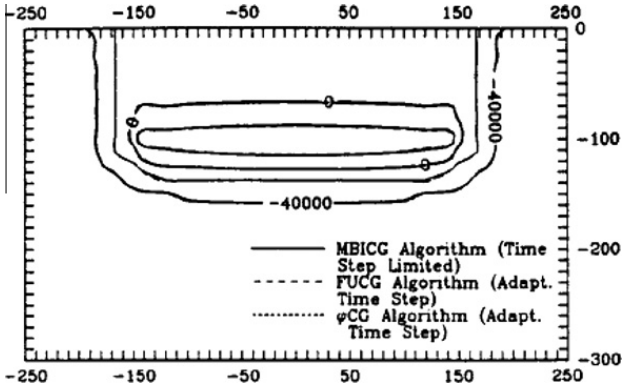

Fig. 12b. Test 1 – computed piezometric depth contours by Kirkland et al. [25] (measures in cm).

computed saturation contours corresponding to the initial condition $\psi_0 = -100$ m are shown for the same mesh, time step size and simulation time. Maximum CFL value is 2.45 and 2.56, respectively for the case $\psi_0 = -7.34$ m and $\psi_0 = -100$ m. The MBR after 30 d is $1 + 1.2 \times 10^{-16}$, with a total mass balance error $O(10^{-16})$.

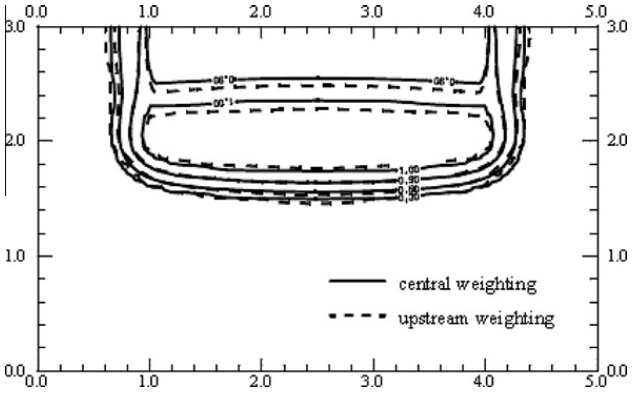

Fig. 13a. Test 1 – computed saturation contours by Diersch and Perrochet [14].

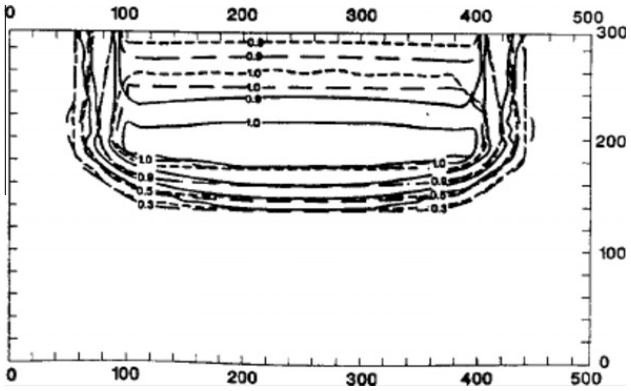

Fig. 13b. Test 1 – computed saturation contours by Forsyth et al. [16] (measures in cm).

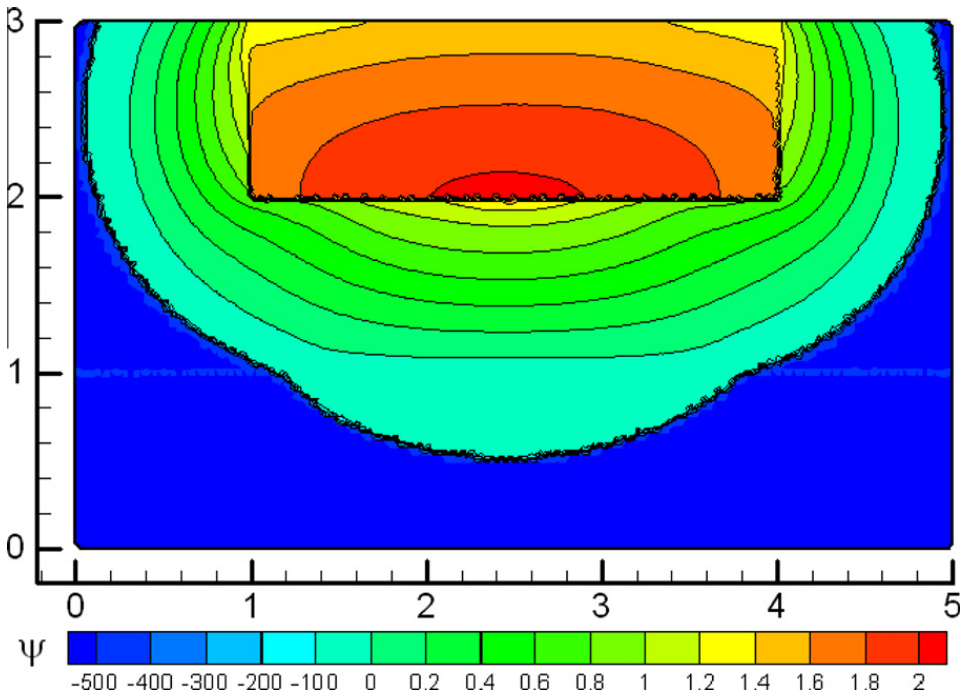

Fig. 14a. Test 1 – MAST computed piezometric depth contours (fine mesh).

The computed saturation contours are compared with the ones given by Diersch and Perrochet [14] and by Forsyth et al. [16] in the two cases $\psi_0 = -7.34$ m and $\psi_0 = -100$ m, for the same simulation time, shown in Figs. 19a–20b respectively. In both reference literature models, the authors used a quadrilateral mesh with 90×21 quadrilateral elements (1890 nodes).

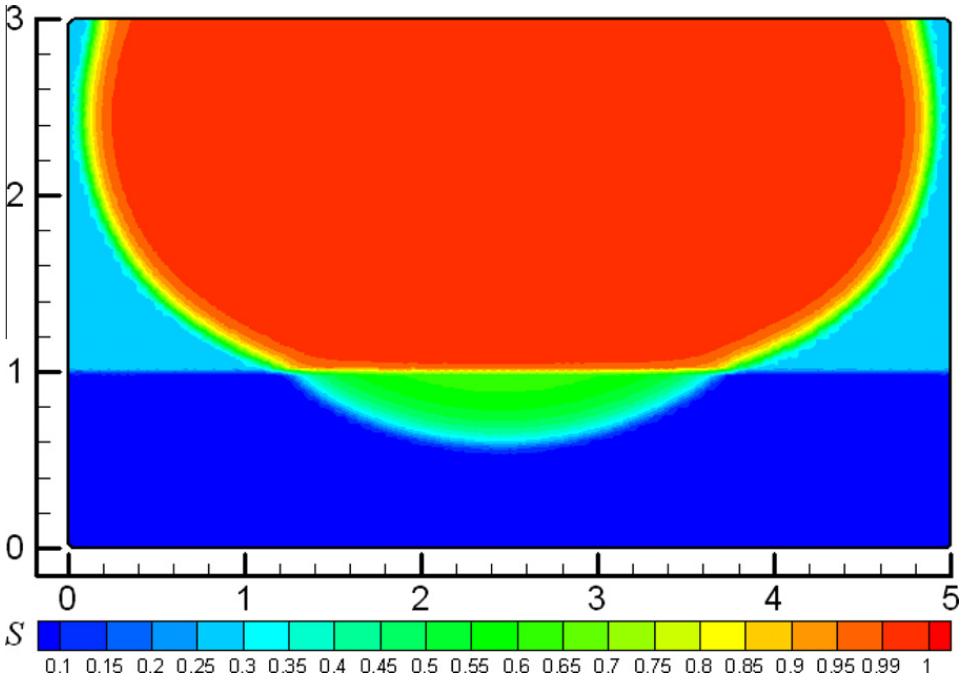

Fig. 14b. Test 1 – MAST computed saturation contours (fine mesh).

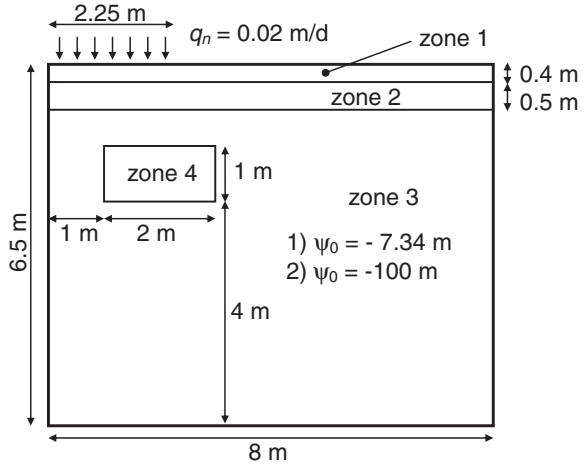

Fig. 15. Test 2 – computational domain, initial and boundary conditions.

Table 2
Test 2 – Van Genuchten parameters.

Zone	K_s [m/s]	ε [−]	s_r [−]	μ [1/m]	λ [−]	ψ_a [m]
1	9.153d−05	0.368	0.2771	3.34	1.982	0
2	5.445d−06	0.351	0.2806	3.63	1.632	0
3	4.805d−05	0.325	0.2643	3.45	1.573	0
4	4.805d−04	0.325	0.2643	3.45	1.573	0

The initial time step used by Diersch and Perrochet [14] is 86.4 s and they carried out 200–300 time iterations. The number of unknowns in the MAST scheme is close to the one in the literature models, while the number of time iterations of the MAST scheme is higher, but Diersch and Perrochet [14] apply a control procedure at each time step to change the Δt size. If the same number of time steps is used with the proposed model (Δt = 6480 s), a numerical diffusion effect can be observed. The computed saturation profiles are much smoother than the previous ones (for simplicity are not shown here), but no oscillations occur at material interfaces. The MBR and the total mass balance error are very similar to the previous ones obtained with Δt = 3240 s.

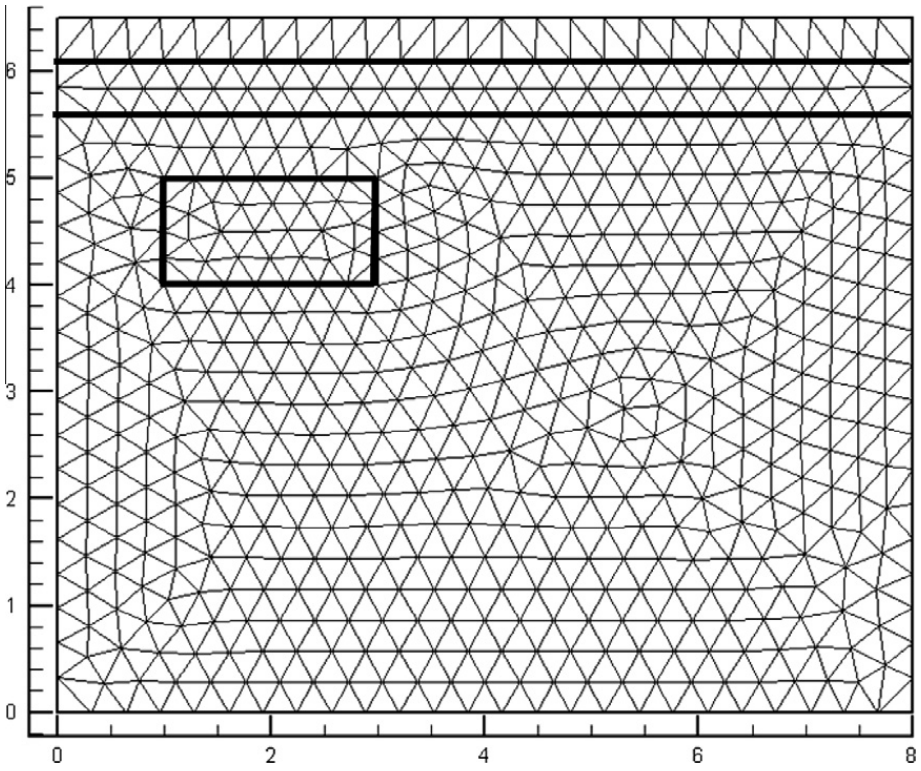

Fig. 16. Test 2 – computational mesh.

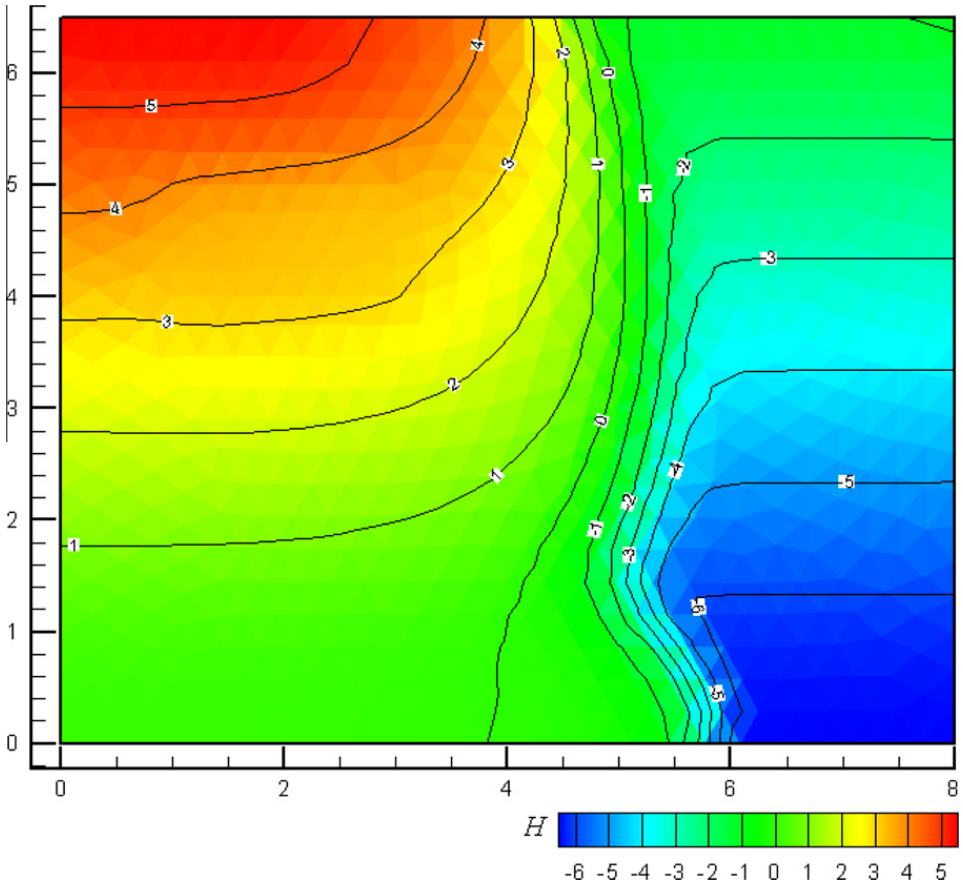

Fig. 17a. Test 2a – MAST computed piezometric head contours ($\psi_0 = -7.34$ m).

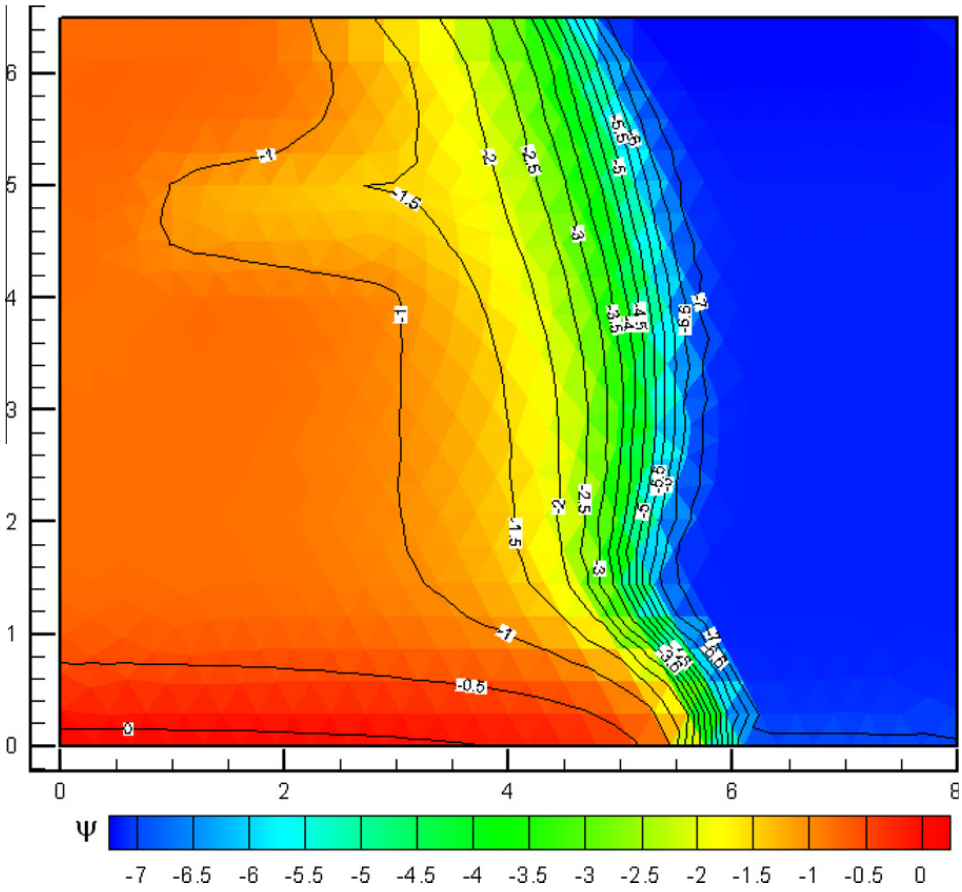

Fig. 17b. Test 2a – MAST computed piezometric depth contours ($\psi_0 = -7.34$ m).

Diersch and Perrochet [14] obtained a total mass balance errors (TMBE) ranging from $O(10^{-5})$ to $O(10^{-4})$.

Observe that also in the present test the saturation zone computed by the MAST scheme is more extended than the ones computed by the literature models.

As in the previous test, a much more fine mesh has been used to test the effect of mesh density on the results in Figs. 17a–18. The fine mesh has 12,162 triangles, 6227 nodes and 18,388 sides. Time step Δt is 86.4 s (30,000 total time iterations). Results are shown in Figs. 21a–21c (case $\psi_0 = -7.34$ m) and in Fig. 22 (case $\psi_0 = -100$ m). As for test 1, MAST results obtained with the coarse mesh can be assumed qualitatively very close to the ones computed with the fine mesh. The MBR and the total mass balance error computed using the fine mesh are very similar to the ones computed with the coarse mesh.

The differences between MAST results and those provided by the other literature models can be explained as for test 1.

In Fig. 23 the computed saturation contours obtained by Diersch and Perrochet for different meshes are shown in the case of $\psi_0 = -100$ m. The dense mesh has 56,960 quadrilateral elements. Mesh size effects are not negligible.

5.2.2. Test 2b: Forsyth and Kropinski's problem [15]

Forsyth and Kropinski [15] modified the above infiltration problem shown in Fig. 15 by increasing the pore size distribution λ index to 5 for the zones 3 and 4. The other parameters are the same as the ones in Table 2. This increment of λ makes the capillary pressure curve very flat at intermediate saturation values and spurious local maxima and minima result in the simulations carried out by Diersch and Perrochet [14] for coarse meshes. For MAST numerical simulations, the same coarse mesh as in test 2a (1148 triangles, 620 nodes and 1767 edges, here marked as mesh 0) and the same $\Delta t = 3240$ s have been used. The present mesh has been refined by dividing each element in four equal triangles, connecting the midpoints of the three sides of the triangle. In order to limit the growth of the CFL number value, the time step size has been halved at each refinement. Three refinement levels have been considered (mesh 1, mesh 2 and mesh 3). In Figs. 24a–24d the computed saturation contours are shown for the simulation time 30 d. The initial condition is $\psi_0 = -100$ m. CFL ranges from 2.77 to 2.89. Values of the MBR and the total mass balance error are very similar to the ones computed for test 2a.

No oscillations occur in the computed profiles, especially in the ones obtained using the coarsest mesh and mesh effects can be considered modest on the computed results.

In Figs. 25a and 25b the saturation profiles computed by Diersch and Perrochet [14] using quadrilateral meshes, respectively with 21 x 90 and 28,917 nodes, are shown. Similar results have been obtained by Forsyth and Kropinski [15] using a

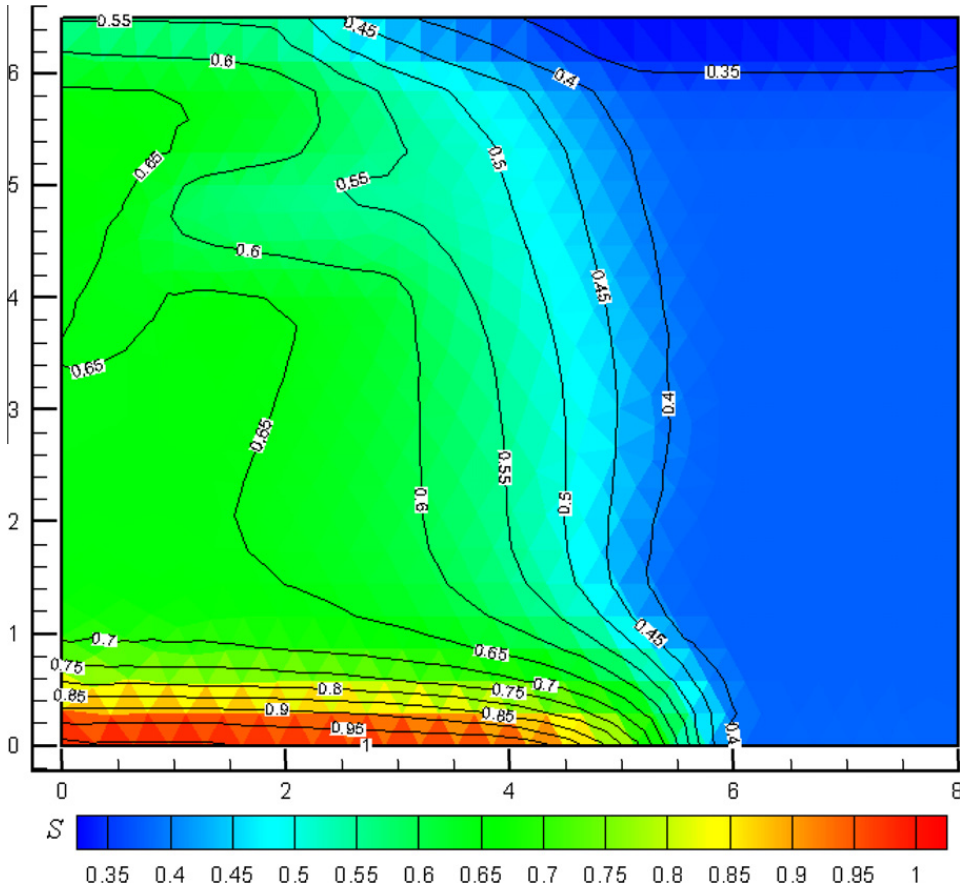

Fig. 17c. Test 2a – MAST computed saturation contours ($\psi_0 = -7.34$ m).

quadrilateral meshes with 159 × 51 nodes, not shown here for brevity. Literature results are in this case much close to the ones computed by the proposed model.

5.3. Test 3. Computation of the velocity fields in strongly saturation conductivity contrast medium

The square domain 1 m × 1 m shown in Fig. 26(a) is used, with boundary conditions and permeability distribution graphically shown. The domain is discretized with two structured triangular mesh with 200 rectangular isosceles triangles, 121 nodes and 320 sides; the two meshes have the triangle hypotenuses orthogonal to each other (see Fig. 26(a)). The time step is $\Delta t = 2$ s. An initial piezometric depth $\psi_0 = -3$ m has been assigned in the domain. Table 3 reports the van Genuchten parameters. Observe the strong saturation conductivity contrast for the three zones. This test is similar to the one carried out by Hoteit et al. [20], but to obtain physical and computational mesh symmetry, we rotated the computational domain as shown in Fig. 26(b). For brevity, piezometric head and depth contours are not shown. Observe in Figs. 27a and 27b the computed velocity vectors for mesh (a) and (b). In both cases results are perfectly symmetric. The x and z velocity components have been calculated by computing the x and z piezometric head spatial gradient components in each triangle on the base of the H values at the edge centres. In the computation of the vector modules, the relative conductivity has been assumed equal to 1. Velocity values are very close to zero in the zones with the lowest conductivity value (vectors disappear in the figures) and velocity vectors tend to become parallel to the boundary of these zones. Vector norms increase in the highest conductivity zones, while the high values of velocity in the low-central area of the domain are due to the sharp gradients of the piezometric head.

5.4. Test 4. Computational cost investigation

A unitary square domain 1 m × 1 m has been used for this test, shown in Fig. 28 along with the boundary conditions. The domain is discretized using three different meshes. The first two are unstructured meshes, shown in Fig. 29 (cases (a) and (b)) [24]. These two meshes have 128 triangular elements, 81 nodes and 208 sides. The first mesh is mildly unstructured while the second one is highly unstructured and characterized by the presence of several obtuse triangles. The third mesh is structured with 153 equilateral triangles, 95 nodes and 247 sides. In the three cases an initial piezometric head $H_0 = -10$ m has been assigned in the domain. The Dirichlet condition imposed on the right boundary of the domain is $H_D = -0.75$ m and

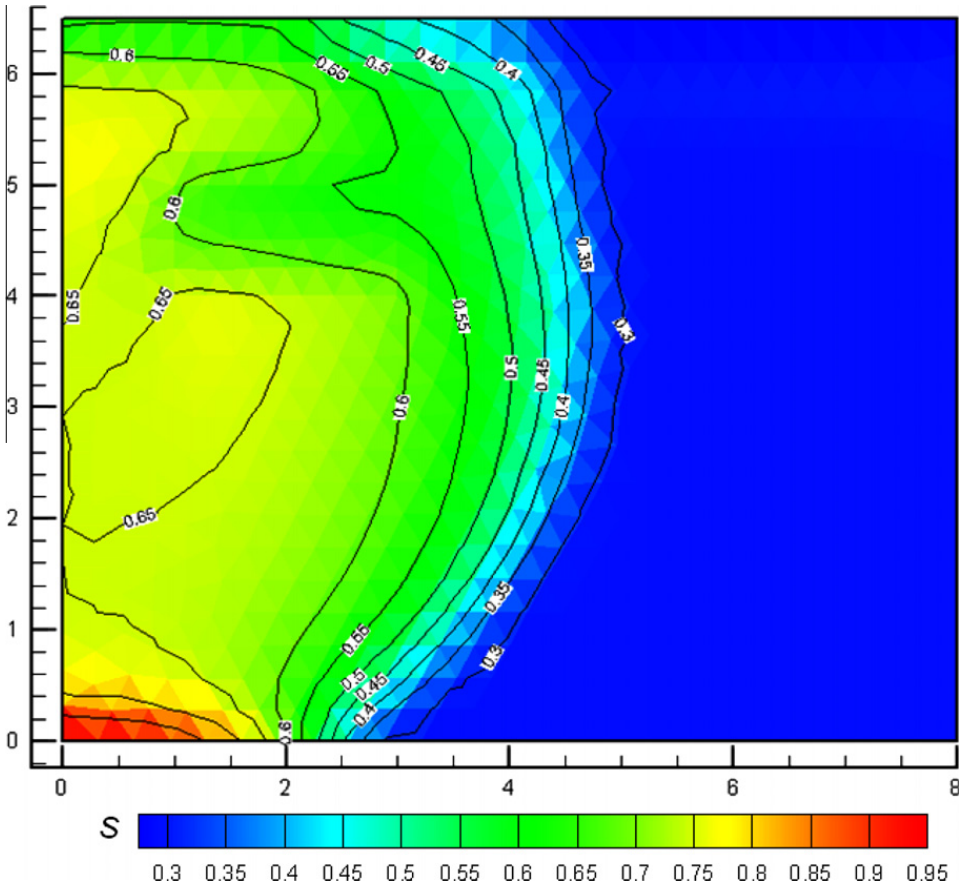

Fig. 18. Test 2a – MAST computed saturation contours ($\psi_0 = -100$ m).

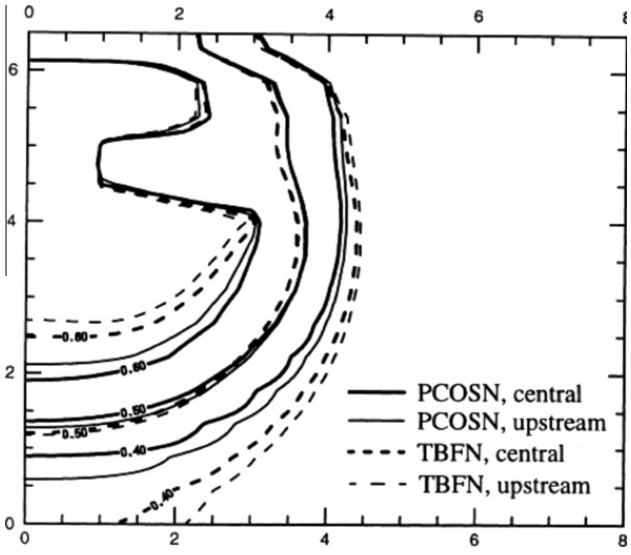

Fig. 19a. Test 2a – computed saturation contours by Diersch and Perrochet [14] ($\psi_0 = -7.34$ m).

the time step is $\Delta t = 25$ s for the three meshes. Table 4 shows the Brooks–Corey parameters, along with the assigned Neumann fluxes per element along the top and left domain sides.

Starting from the three meshes before described, a mesh refinement has been carried out for each of them, as described before. Three refinement levels have been obtained. CFL ranges from 0.86 to 1.59 from level 0 to level 3 for the mesh type A and very close variation ranges have been observed for the other two meshes.

In Tables 5a–5c the mean computational time (CPU), per iteration and per single cell, for the different model phases – managing of obtuse triangles, cells ordering, convective step and diffusive step – are reported. A single processor Intel T

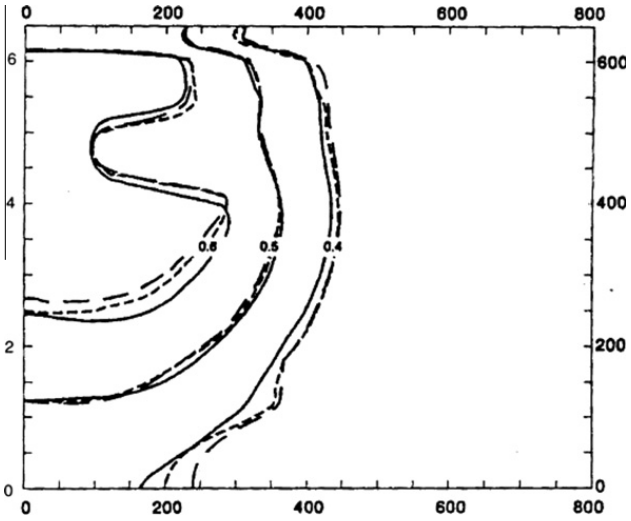

Fig. 19b. Test 2a – computed saturation contours by Forsyth et al. [16] ($\psi_0 = -7.34$ m) (measures in cm).

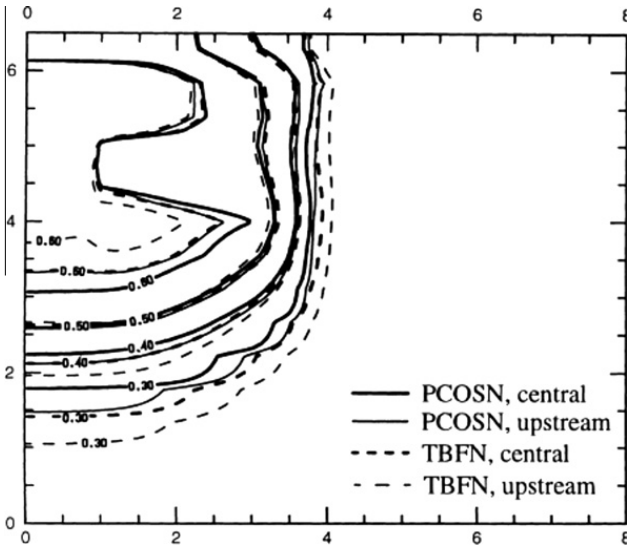

Fig. 20a. Test 2a – computed saturation contours by Diersch and Perrochet [14] ($\psi_0 = -100$ m).

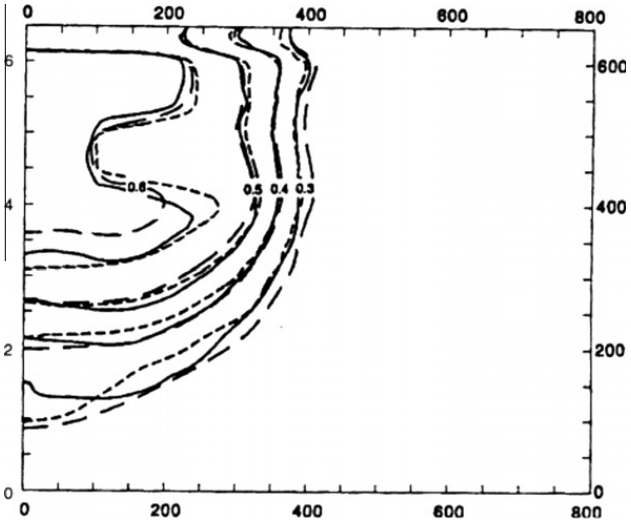

Fig. 20b. Test 2a – computed saturation contours by Forsyth et al. [16] ($\psi_0 = -100$ m).

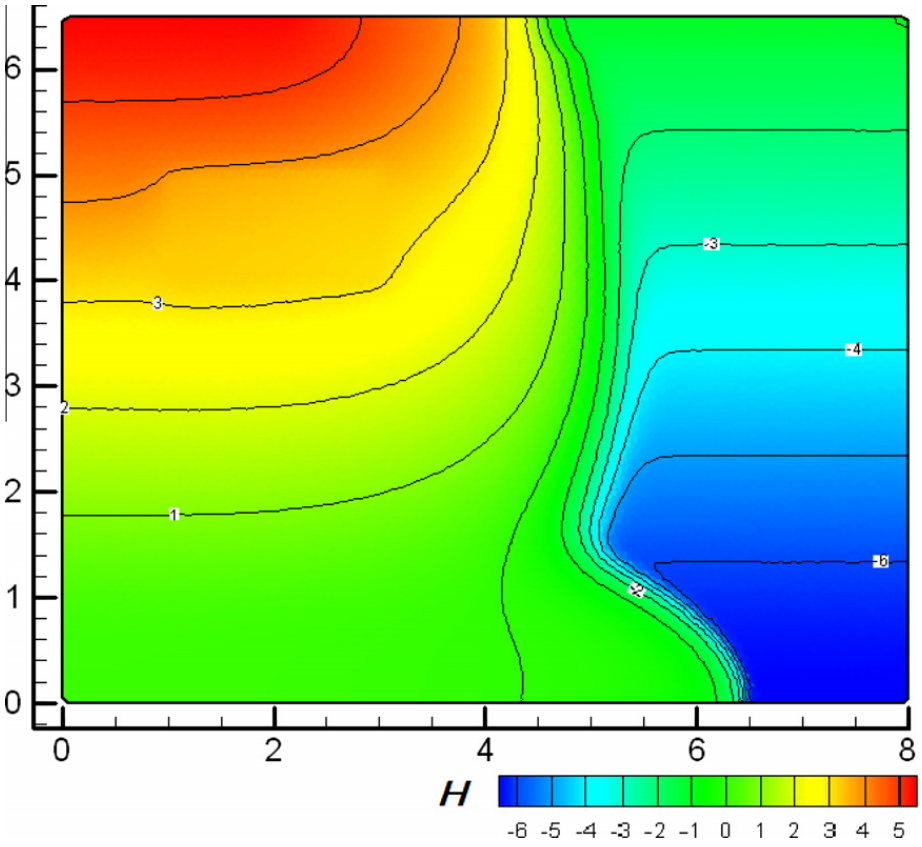

Fig. 21a. Test 2a – MAST computed piezometric head contours (fine mesh, $\psi_0 = -7.34$ m).

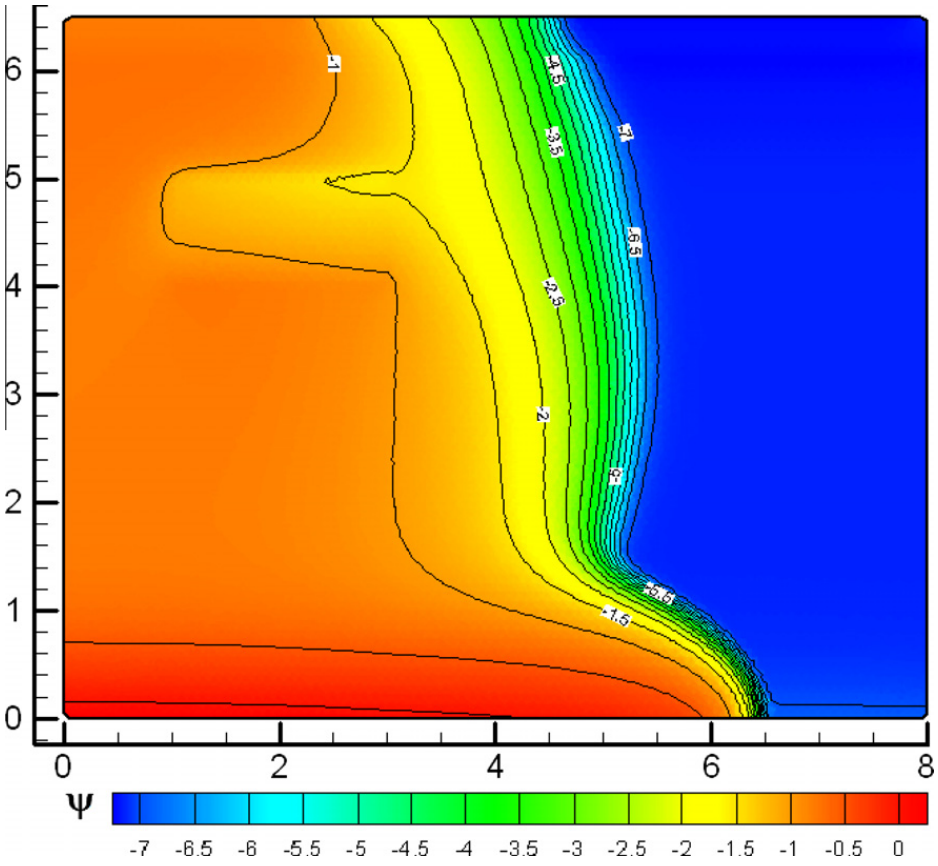

Fig. 21b. Test 2a – MAST computed piezometric depth contours (fine mesh, $\psi_0 = -7.34$ m).

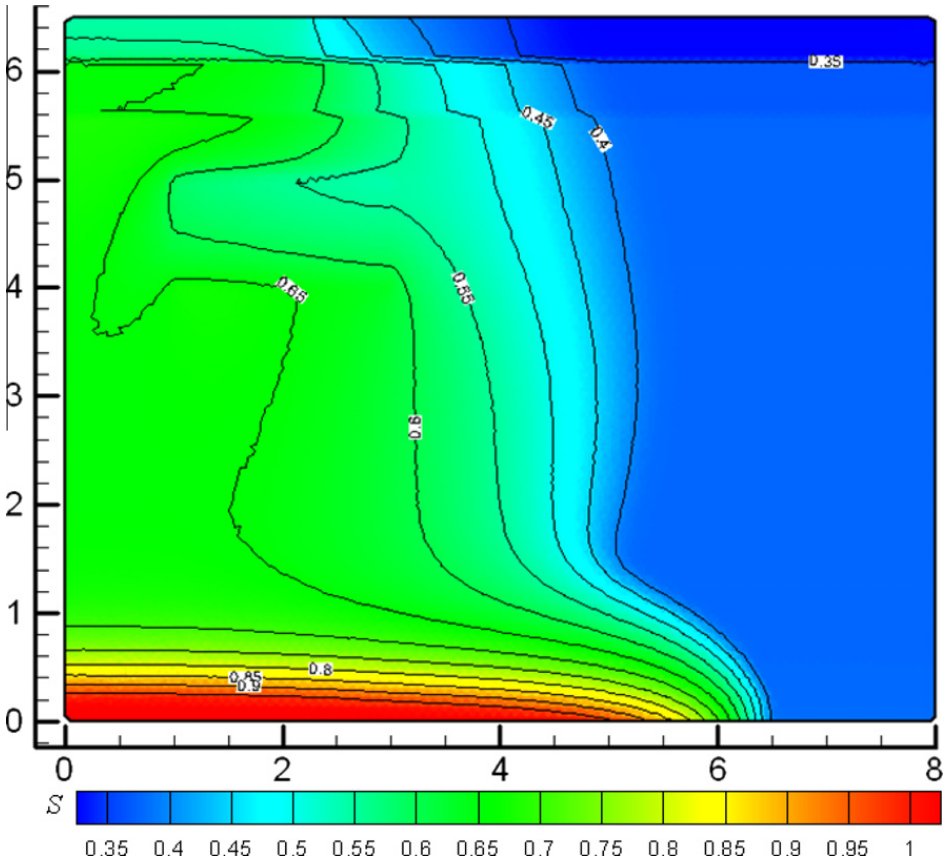

Fig. 21c. Test 2a – MAST computed saturation contours (fine mesh, $\psi_0 = -7.34$ m).

9400, 2.53 GHz has been used. The computation of the convective prediction step is the most demanding one, while the solution of the diffusive system needs computational time approximately one magnitude order less than the convective step.

The CPU time relative to the convective step is almost independent on the mesh elements number, since this represents the "explicit" component of the method. The small decrement of the average CPU time of the convective component can be related to the increasing CFL numbers obtained by partitioning and to the best aptitude of the algorithm to work with CFL numbers greater than one [3–6]. The CPU time per single cell required for the solution of the diffusive linear system increases with the element number. In fact, this step, representing the "non explicit" component of the algorithm, requires the solution of a large linear system of the order of the edges number.

As expected, the cost of managing obtuse triangulation (equal to zero for the mesh with equilateral triangles) per single cell is independent on the number of elements, while the ordering step requires a CPU time per single cell that grows with the cell number much less than linearly and is two magnitude order less than the CPU time required by the diffusive step.

The growth rate β, measured as the exponent of the relationship:

$$\overline{CP} = L^{\beta} \;\Rightarrow\; \log(\overline{CP}) = \log(L)\cdot\beta + c, \tag{47}$$

where L is the number of edges, $\overline{CP}$ is the mean CPU time per iteration and c is an arbitrary constant, has been investigated for the four components of the algorithm, see in Figs. 30a–30c that β ranges from 1.047 to 1.16 for the diffusive step and from 1.14 to 1.46 for the cells ordering step, while it is almost one for the convective and the cost of managing obtuse triangles steps.

In Tables 6a–6c the mean CPU time per single cell and per iteration are shown for the same previous tests; the only difference is that the prediction step is computed by applying the semi-analytical procedure described in Section 4.3.2. For this set of simulations the interval $0 \leqslant \xi \leqslant 1$ (see Section 4.3.2) has been divided in 20 parts. The maximum admissible error η_{max} (see Section 4.3.2) is 1.d−05. Observe that the CPU times for the solution of the convective step are approximately halved with respect to the previous ones obtained using the numerical solution. The ranges of the growth rates β are almost the same as the ones shown in Figs. 30a–30c, and for simplicity not reported here. Very similar values have been obtained using η_{max} = 1.d−04 and η_{max} = 1.d−06.

5.5. Test 5. Estimation of the order of convergence

An analytical solution is compared to its numerical approximation on a sequence of refined meshes in order to evaluate the rate of convergence. The analytical solution is arbitrarily given for Eq. (13a), where the sink term on its r.h.s. is computed

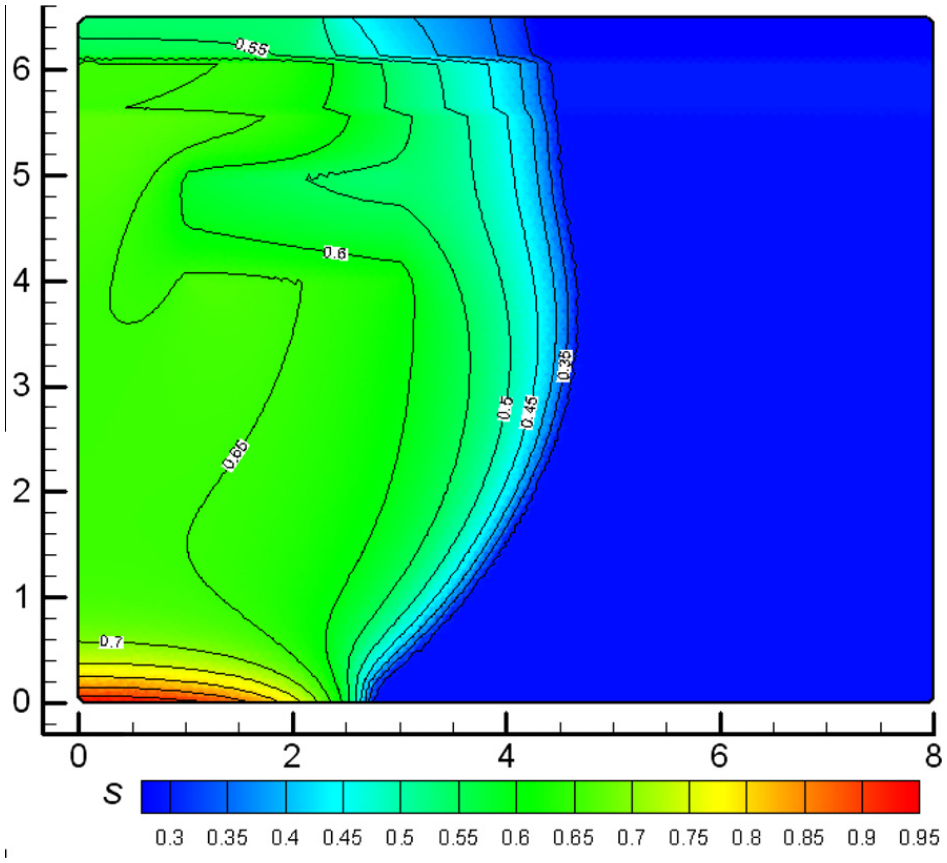

Fig. 22. Test 2a – MAST computed saturation contours (fine mesh, $\psi_0 = -100$ m).

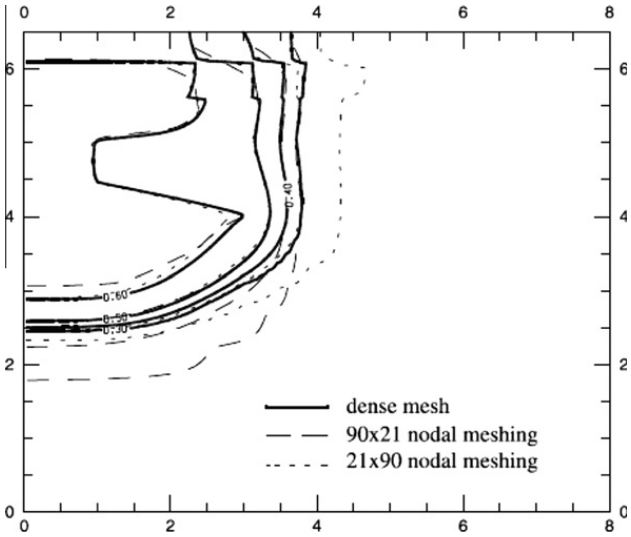

Fig. 23. Test 2a – computed saturation contours by Diersch and Perrochet [14] ($\psi_0 = -100$ m).

by time and space differentiation of the same solution on the l.h.s. of the same Eq. (13a). The assigned analytical solution is [30]:

$$\psi(x,z,t) = -\frac{1}{2} - x^2(2-x)e^{x(2-x)} \cdot z(1-z)e^{z} \cdot (1-e^{-t/T}), \tag{48}$$

where T is the characteristic time of the process, equal to 1 day. In Fig. 31(a) the analytical solution for $t/T \to \infty$ is shown.

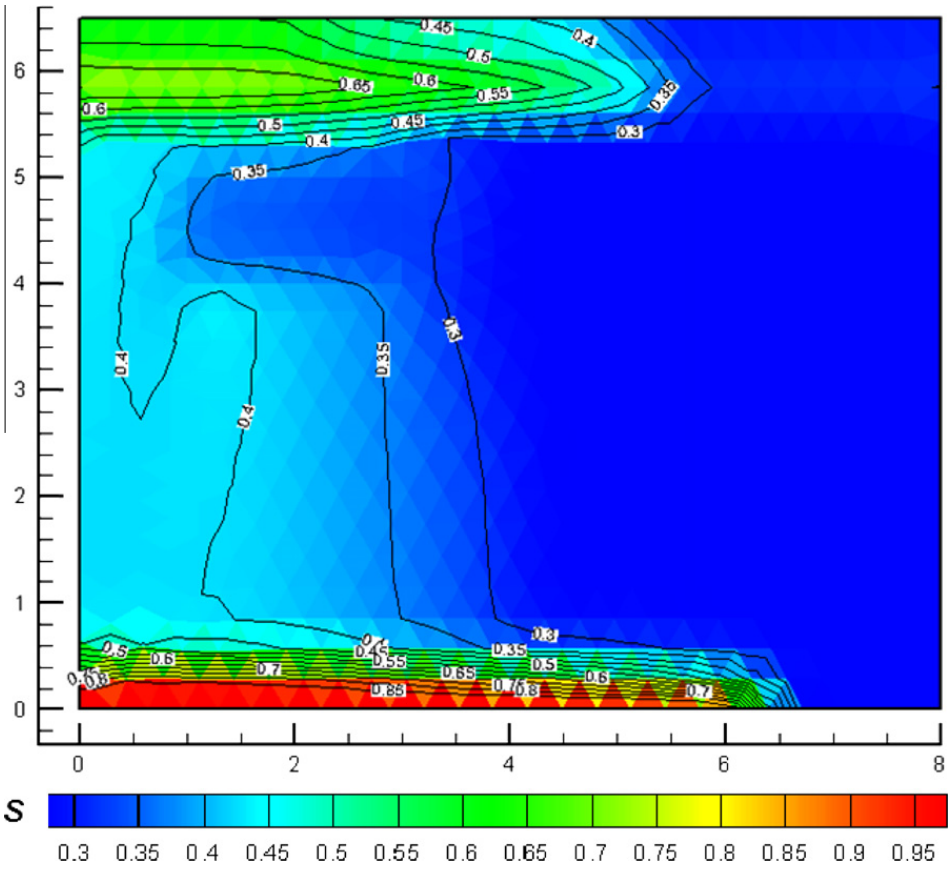

Fig. 24a. Test 2b – MAST computed saturation contours (mesh 0, $\psi_0 = -100$ m).

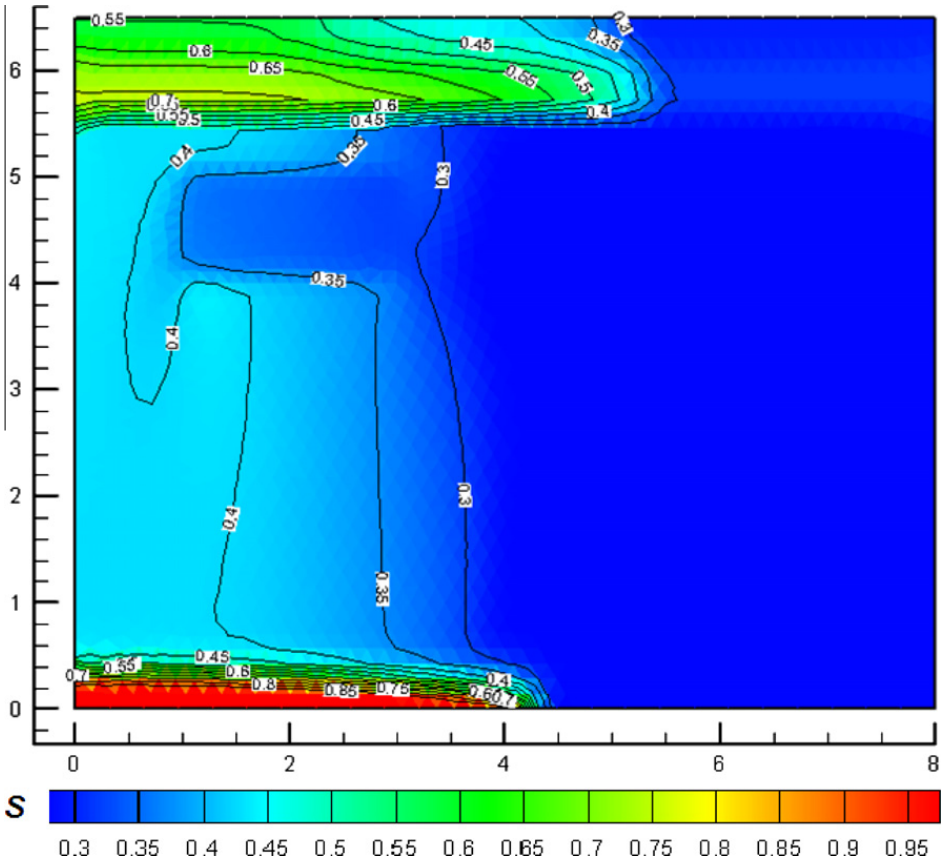

Fig. 24b. Test 2b – MAST computed saturation contours (mesh 1, $\psi_0 = -100$ m).

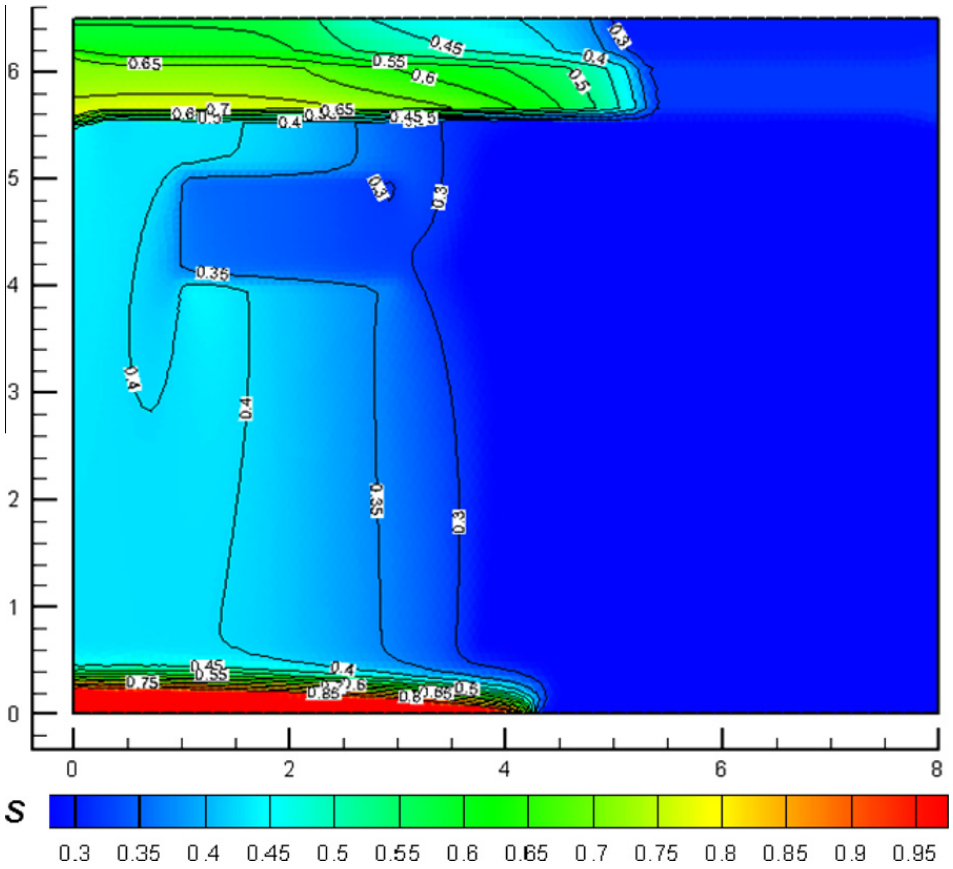

Fig. 24c. Test 2b – MAST computed saturation contours (mesh 2, $\psi_0 = -100$ m).

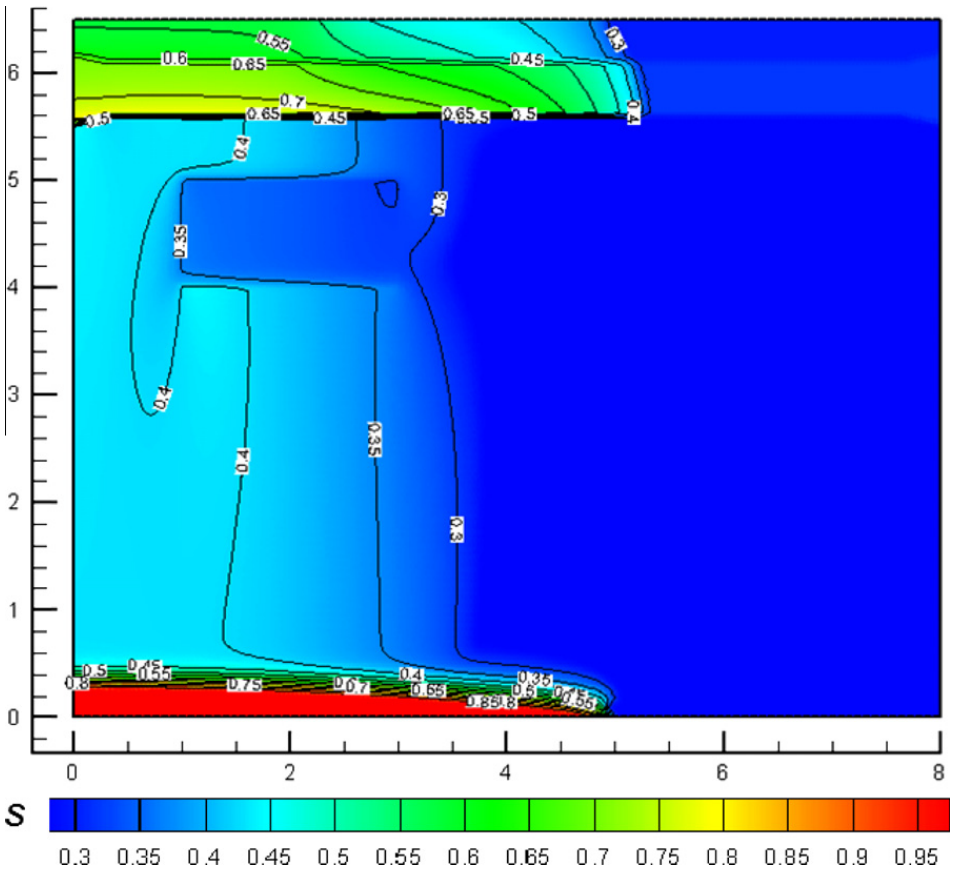

Fig. 24d. Test 2b – MAST computed saturation contours (mesh 3, $\psi_0 = -100$ m).

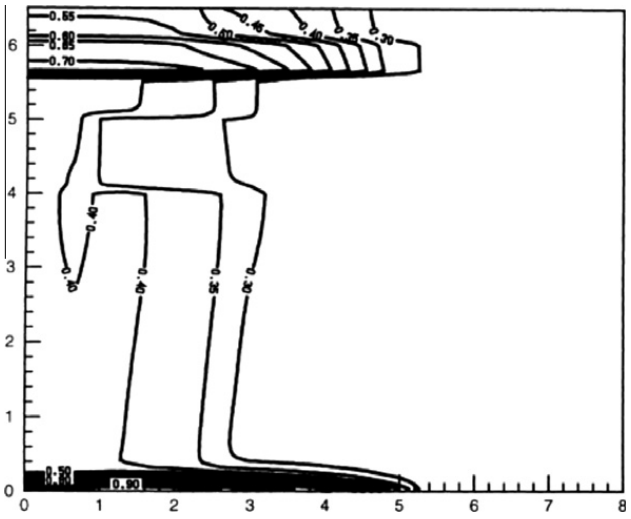

Fig. 25a. Test 2b – computed saturation contours by Diersch and Perrochet [14] ($\psi_0 = -100$ m, 21 × 90 nodes).

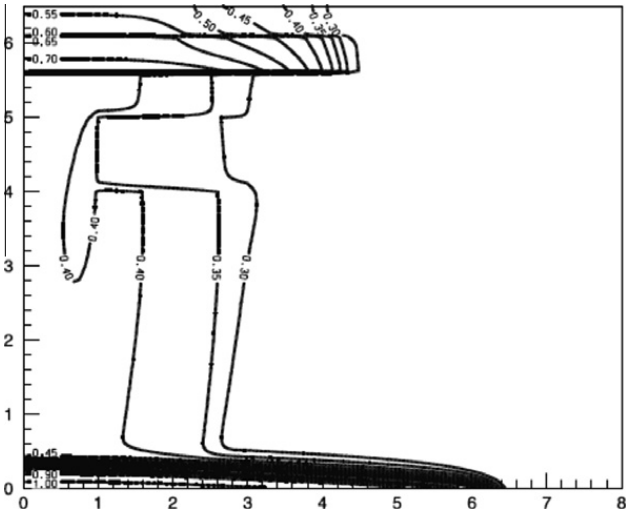

Fig. 25b. Test 2b – computed saturation contours by Diersch and Perrochet [14] ($\psi_0 = -100$ m, 28,917 nodes).

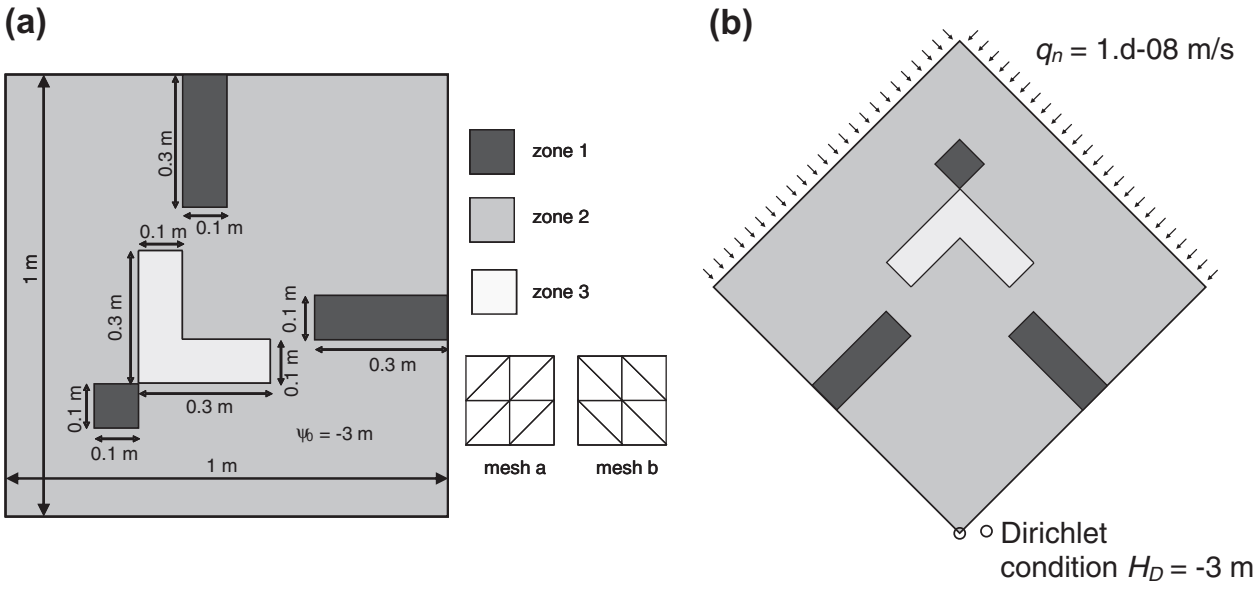

Fig. 26. Test 3 – (a) computational domain with permeability zones and mesh types; (b) rotation of the domain.

Table 3
Test 3 – Van Genuchten parameters.

Zone	K_s [m/s]	θ_s [−]	θ_r [−]	μ [1/m]	λ [−]	ψ_a [m]
1	1.889d−09	0.47	0.08	3.35	2	0
2	1.889d−05	0.47	0.08	3.35	2	0
3	1.889d−01	0.47	0.08	3.35	2	0

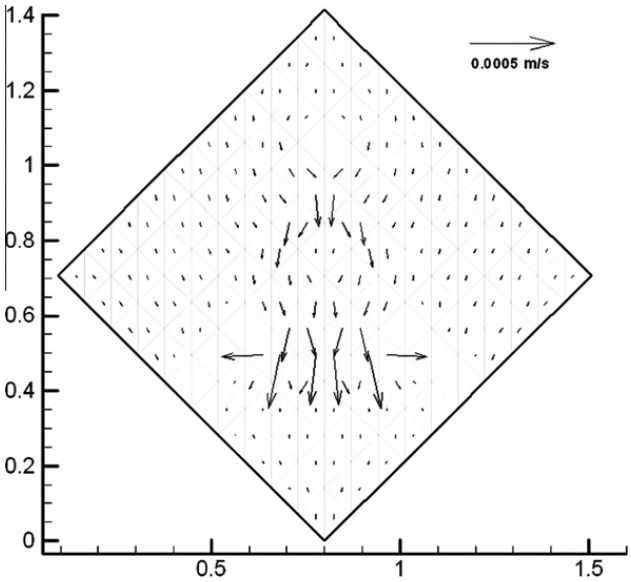

Fig. 27a. Test 3 – MAST computed velocity vectors (mesh a).

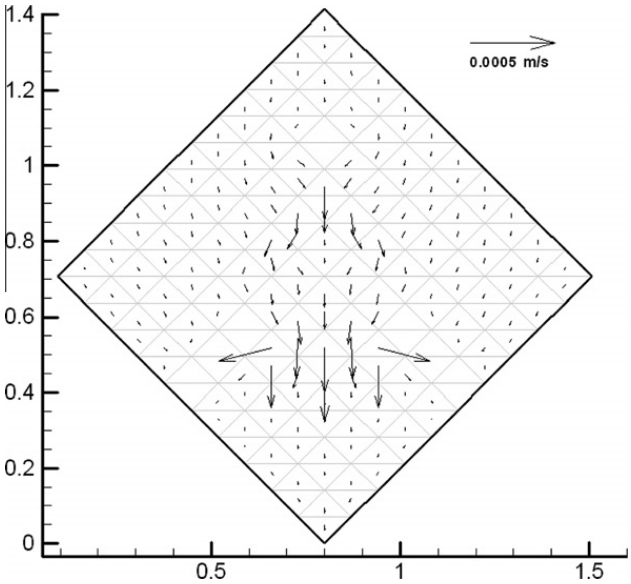

Fig. 27b. Test 3 – MAST computed velocity vectors (mesh b).

The 2D sink term simulates the roots uptake. Because of the symmetry of the analytical solution along x direction, only half domain 1 m × 1 m has been solved. Table 7 shows the van Genuchten soil parameters; the hydraulic conductivity is given by Manzini and Ferraris [30]:

$$K(\psi) = K_s(z)K_r(\psi), \tag{49}$$

$$K_s = K_s'' + (K_s' - K_s'')(1 - e^{-(1-z/L_z)}) \quad \text{and } K_r(\psi) = e^{\psi/\psi_0}, \tag{50}$$

where constants L_z = 1 m and ψ_0 = 0.0333 m. Values of K_s' and K_s'' are listed in Table 7. The two scalars provide different heterogeneity degrees between the top and bottom hydraulic conductivity [30]. A basic coarse mesh (refinement level l = 0) with

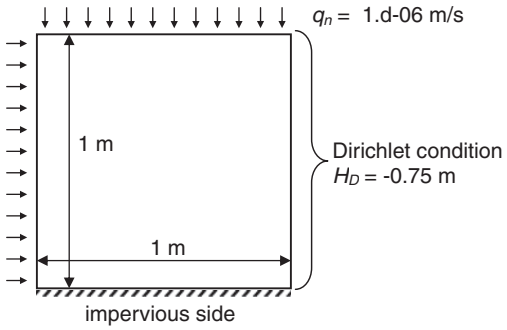

Fig. 28. Test 4 – computational domain and boundary conditions.

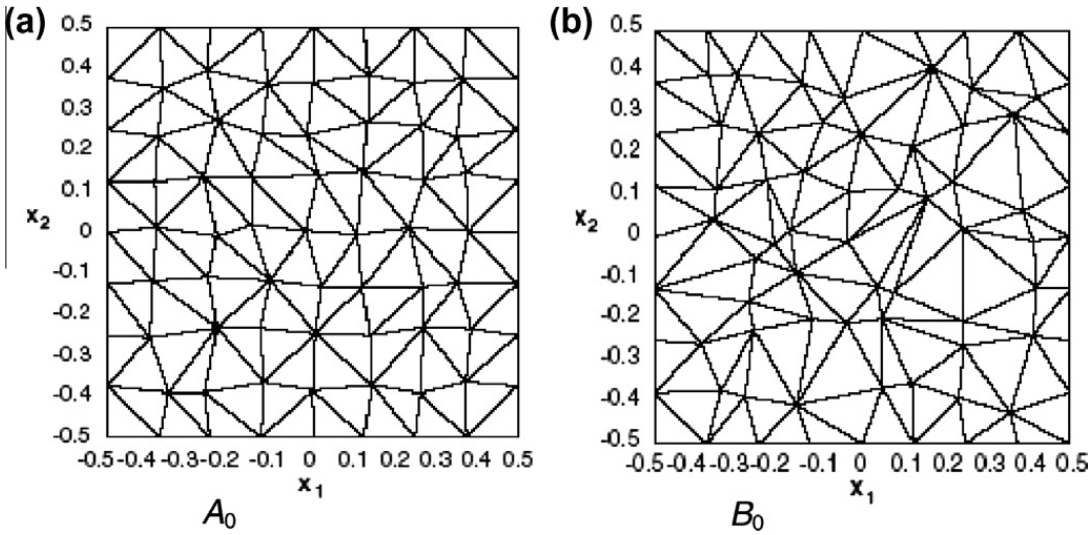

Fig. 29. Test 4 – (a) computational mesh A (level 0), (b) computational mesh B (level 0).

Table 4
Test 4 – Brooks-Corey parameters and incoming Neumann fluxes.

K_s [m/s]	θ_s [−]	θ_r [−]	λ [−]	ψ_b [m]	q_n (mesh A) [m^2/s]	q_n (mesh B) [m^2/s]	q_n (mesh equil.) [m^2/s]
1.889d−06	0.47	0.08	0.2857	−0.3	1.25d−07	1.25d−07	1.111d−07

Table 5a
Test 4 – mean CPU time for equilateral mesh (numerical solution of the convective step).

Refinement level	N_T	L	CPU_obt [s]	CPU_ord [s]	CPU_conv [s]	CPU_diff [s]
0	153	247	0.00E+00	0.00E+00	1.38E−05	3.28E−06
1	612	953	0.00E+00	5.46E−09	1.46E−05	2.77E−06
2	2448	3742	0.00E+00	1.39E−08	1.46E−05	3.31E−06
3	9792	14,828	0.00E+00	1.96E−08	1.42E−05	3.84E−06

Table 5b
Test 4 – mean CPU time for mesh *A* (numerical solution of the convective step).

Refinement level	N_T	L	CPU_obt [s]	CPU_ord [s]	CPU_conv [s]	CPU_diff [s]
0	128	208	2.00E−07	0.00E+00	1.45E−05	1.90E−06
1	512	800	1.95E−07	1.30E−08	1.53E−05	2.31E−06
2	2048	3136	2.06E−07	1.24E−08	1.47E−05	3.04E−06
3	8192	12,416	2.05E−07	2.38E−08	1.47E−05	3.61E−06

Table 5c
Test 4 – mean CPU time for mesh *B* (numerical solution of the convective step).

Refinement level	N_T	L	CPU_obt [s]	CPU_ord [s]	CPU_conv [s]	CPU_diff [s]
0	128	208	4.00E−07	0.00E+00	1.45E−05	2.70E−06
1	512	800	3.90E−07	1.30E−08	1.46E−05	2.83E−06
2	2048	3136	2.98E−07	1.58E−08	1.46E−05	3.06E−06
3	8192	12,416	3.24E−07	1.91E−08	1.45E−05	3.64E−06

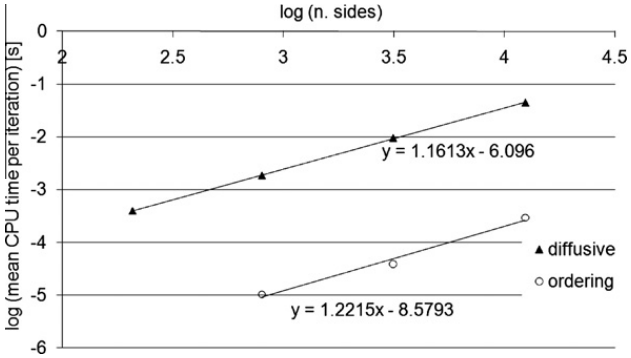

Fig. 30a. Test 4 – mean CPU times per iteration for the different steps of the MAST algorithm (mesh A – numerical solution of the convective step).

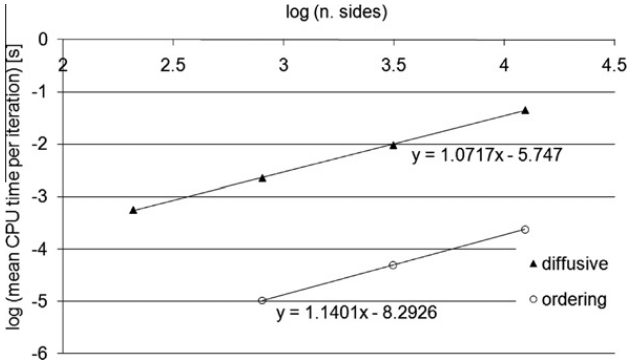

Fig. 30b. Test 4 – mean CPU times per iteration for the different steps of the MAST algorithm (mesh B – numerical solution of the convective step).

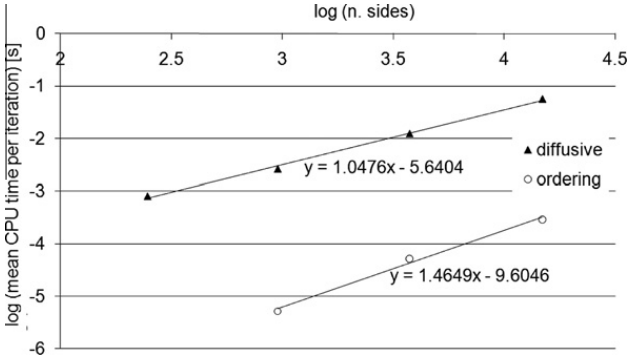

Fig. 30c. Test 4 – mean CPU times per iteration for the different steps of the MAST algorithm (equilateral triangles mesh – numerical solution of the convective step).

164 triangles, 99 nodes and 262 edges has been used for the numerical simulations (see Fig. 31(b)) and 3 mesh refinements have been carried out (refinements levels $l = 1,2,3$). In Tables 8a–8d the relative errors for the potential values and the x and z velocity components are reported for $K_s'' = 0.1$ cm h^{-1}. Very similar results have been obtained using the other K_s'' values and for simplicity are not reported here. Relative errors corresponding to refinement mesh level lth are computed at edges mid-points for piezometric heads and at centres of mass of the triangles for velocity components and are defined as [9]:

Table 6a
Test 4 – mean CPU time for equilateral mesh (semi-analytical solution of the convective step).

Refinement level	N_T	L	CPU_obt [s]	CPU_ord [s]	CPU_conv [s]	CPU_diff [s]
0	153	247	0.00E+00	0.00E+00	7.88E−06	3.55E−06
1	612	953	0.00E+00	2.60E−08	7.53E−06	3.48E−06
2	2448	3742	0.00E+00	3.41E−08	7.08E−06	4.11E−06
3	9792	14,828	0.00E+00	8.00E−08	6.97E−06	4.93E−06

Table 6b
Test 4 – mean CPU time for mesh *A* (semi-analytical solution of the convective step).

Refinement level	N_T	L	CPU_obt [s]	CPU_ord [s]	CPU_conv [s]	CPU_diff [s]
0	128	208	2.00E−07	0.00E+00	6.35E−06	2.70E−06
1	512	800	2.66E−07	0.00E+00	6.55E−06	2.86E−06
2	2048	3136	2.76E−07	2.90E−08	6.38E−06	3.15E−06
3	8192	12,416	2.53E−07	6.84E−08	6.42E−06	3.74E−06

Table 6c
Test 4 – mean CPU time for mesh *B* (semi-analytical solution of the convective step).

Refinement level	N_T	L	CPU_obt [s]	CPU_ord [s]	CPU_conv [s]	CPU_diff [s]
0	128	208	3.50E−07	0.00E+00	7.07E−06	2.25E−06
1	512	800	3.58E−07	2.60E−08	6.59E−06	3.00E−06
2	2048	3136	3.67E−07	3.15E−08	6.41E−06	3.32E−06
3	8192	12,416	4.11E−07	6.62E−08	6.41E−06	3.81E−06

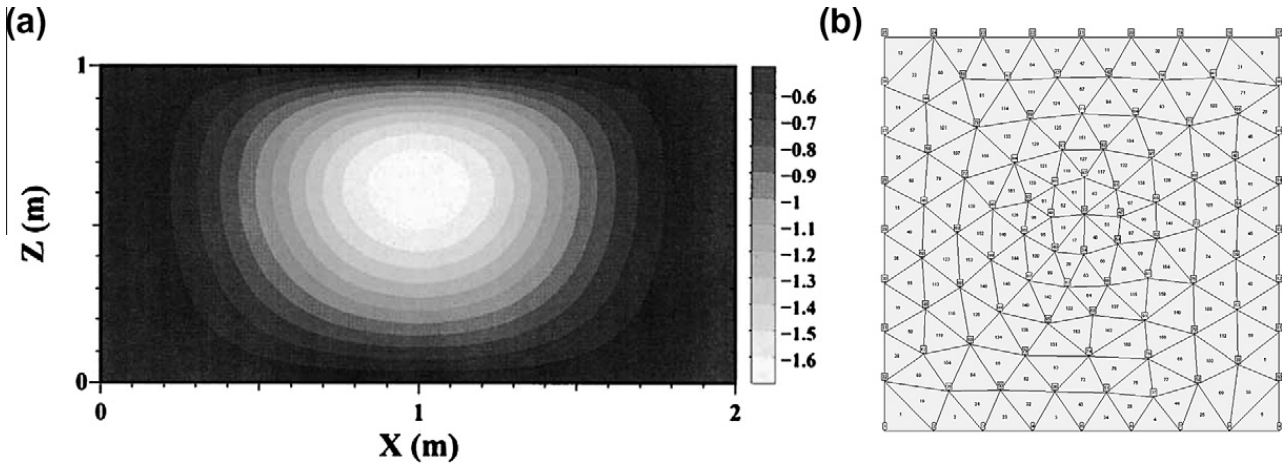

Fig. 31. Test 5 – (a) exact steady state solution (from [30]), (b) – the coarse mesh.

Table 7
Test 5 – Van Genuchten parameters.

K_s' [cm/h]	K_s'' [cm/h]	θ_s [−]	θ_r [−]	μ [1/m]	λ [−]
10.d−00	{1., 0.1, 1.d−01, 1.d−02}	0.368	0.102	3.35	2

Table 8a
Test 5 – errors and convergence orders for piezometric heads and velocity components – $t/T = 1$ and $K_s'' = 0.1$ cm/h.

Refinement level	N_T	L	err_H^l	$r_{c,H}$	$err_{q_x}^l$	r_{c,q_x}	$err_{q_z}^l$	r_{c,q_z}
0	164	262	9.16E−08		2.55E−02		2.52E−02	
1	656	1016	2.24E−08	2.03E+00	1.22E−02	1.06E+00	1.20E−02	1.07E+00
2	2624	4000	5.55E−09	2.01E+00	6.03E−03	1.02E+00	5.90E−03	1.02E+00
3	10,496	15,872	1.37E−09	2.02E+00	3.01E−03	1.00E+00	2.90E−03	1.02E+00

Table 8b
Test 5 – errors and convergence orders for piezometric heads and velocity components – $t/T = 2$ and $K_s'' = 0.1$ cm/h.

Refinement level	N_T	L	err_H^l	$r_{c,H}$	$err_{q_x}^l$	r_{c,q_x}	$err_{q_z}^l$	r_{c,q_z}
0	164	262	5.46E−08		2.39E−02		2.36E−02	
1	656	1016	1.32E−08	2.05E+00	1.20E−02	9.94E−01	1.14E−02	1.05E+00
2	2624	4000	3.25E−09	2.02E+00	6.02E−03	9.95E−01	5.80E−03	9.75E−01
3	10,496	15,872	8.07E−10	2.01E+00	3.00E−03	1.01E+00	2.77E−03	1.07E+00

Table 8c
Test 5 – errors and convergence orders for piezometric heads and velocity components – $t/T = 5$ and $K_s'' = 0.1$ cm/h.

Refinement level	N_T	L	err_H^l	$r_{c,H}$	$err_{q_x}^l$	r_{c,q_x}	$err_{q_z}^l$	r_{c,q_z}
0	164	262	2.36E−08		2.29E−02		2.28E−02	
1	656	1016	5.69E−09	2.05E+00	1.08E−02	1.08E+00	1.06E−02	1.10E+00
2	2624	4000	1.39E−09	2.03E+00	5.10E−03	1.09E+00	5.08E−03	1.06E+00
3	10,496	15,872	3.46E−10	2.01E+00	2.55E−03	9.99E−01	2.39E−03	1.09E+00

Table 8d
Test 5 – errors and convergence orders for piezometric heads and velocity components – $t/T = \infty$ and $K_s'' = 0.1$ cm/h.

Refinement level	N_T	L	err_H^l	$r_{c,H}$	$err_{q_x}^l$	r_{c,q_x}	$err_{q_z}^l$	r_{c,q_z}
0	164	262	4.45E−09		2.10E−02		2.08E−02	
1	656	1016	1.07E−09	2.06E+00	1.05E−02	9.99E−01	1.03E−02	1.01E+00
2	2624	4000	2.62E−10	2.03E+00	4.93E−03	1.09E+00	4.88E−03	1.08E+00
3	10,496	15,872	6.51E−11	2.01E+00	2.35E−03	1.07E+00	2.28E−03	1.10E+00

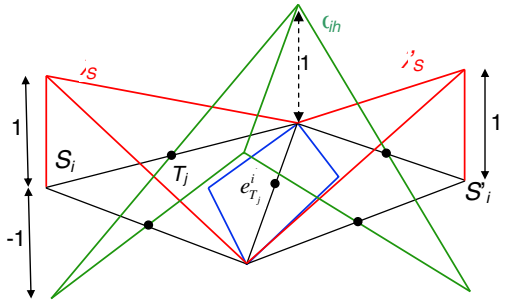

Fig. 32. The nodal basis function λ_i and the basis function of the Y_h space ω_{ih} with its support in the $P1$ nonconforming FEM.

$$err_H^l = \frac{\sqrt{\sum_{i=1,L}\left(H_i - H_i^{ex}\right)^2}}{\sqrt{\sum_{i=1,L}\left(H_i^{ex}\right)^2}}, \quad err_{q_{x(z)}}^l = \frac{\sqrt{\sum_{i=1,N_T}\left(q_{x(z),i} - q_{x(z),i}^{ex}\right)^2}}{\sqrt{\sum_{i=1,N_T}\left(q_{x(z),i}^{ex}\right)^2}}, \tag{51}$$

where H_i and H_i^{ex} are respectively the numerical and the exact solution at midpoint of side i, $q_{x(z),\,i}$ and $q_{x(z),i}^{ex}$ are the numerical and the exact $x(z)$ velocity components in the centre of triangle i. Numerical velocity components have been computed as described in Section 5.3, while the exact ones have been obtained by analytically computing in the centre of mass of each triangle the gradients of H according to the exact solution given by Eq. (48). The relative hydraulic conductivity is assumed equal to one in the computation of $err_{q_{x(z)}}^l$.

The rate of convergence is defined by comparing the relative errors of two consecutive mesh levels. Assuming the relative error obtained for mesh level l proportional to a power of the linear size of the area of the mean triangle in the mesh,

$$err_l = \left(\sqrt{A_l}\right)^{r_c}, \tag{52}$$

where A_l is the area of the mean triangle in the mesh refinement level l and $\sqrt{A_l}$ represents a measure of its linear size, the rate of convergence r_c is computed by comparing the relative errors of two successive refinement levels l and $l+1$:

$$r_c = \frac{\log\left(\frac{err_l}{err_{l+1}}\right)}{\log(2)}. \tag{53}$$

According to the values reported in Tables 8a–8d, the convergence rate (5th, 7th and 9th column) is quadratic for H, as for the standard formulation of the MHFE schemes, while it is linear for velocity components, as expected, since the velocity is not an unknown of the problem.

6. Conclusions

A novel methodology has been presented for the solution of the flow equation in a variably saturated heterogeneous porous medium. The methodology is based on the use of an unstructured triangular mesh and a MArching in Space and Time (MAST) approach. Hysteretic phenomena have not been considered in the present version of the model, but they can be easily included applying literature approaches [21,26]. The methodology can be extended to 3D case, adopting a similar flux discretization and by using the triangular tetrahedral faces (four per each element) instead of the edges. The flux correction procedure should be applied in this case to all the elements that have one or more obtuse angles between their faces.

One of the main advantage of the methodology is that its prediction step is inherently locally mass conservative, even if strong nonlinearities exist between the pressure and the equation parameters. A very small mass balance error can be found in the correction step, due to the non linearity existing between the pressure and the storage parameters. In the tested examples this error is very small, sometimes many order of magnitude smaller with respect to the error shown by other algorithms in literature tests. Despite other literature models, mesh size effects can be assumed negligible and similar results for the same tests have been computed with in a wide range of mesh and time step size.

A special treatment is applied to the obtuse triangles in order to avoid flux loops, that are inconsistent with the irrotationality of the velocity field and hamper the sequential solution of the ordinary differential equations in each cells. This procedure guarantees the positive definite property and *M*-property of the matrix system in the correction step also in the case of obtuse triangles.

The computational burden has been estimated for each different component of the proposed algorithm: cell ordering, treatment of the obtuse triangles, prediction step and correction step. The mean CPU time per cell required for the solution of the prediction step is one magnitude order higher than the one required by the correction step and is almost independent on the mesh elements number, while the CPU time required for the solution of the correction system increases with the element number only a bit more than linearly. A dramatic cost abatement of the prediction step is obtained if a semi-analytical solution is applied.

The order of convergence has been estimated using an exact reference solution and it is quadratic for the piezometric head and linear for the velocity components.

Appendix A. The LMHFE and the linear (P1) nonconforming FEMs

A.1. Preliminaries and notations

Let $\Omega \subset \Re^2$ be a bounded domain, Γ its boundary and T_h an unstructured Delaunay-type triangulation over Ω (as reported in Section 4). The meaning of the symbols used in this Appendix A is the same as specified in the previous sections.

Let $H^1(\Omega)$ be the Sobolev space of square-integrable functions with square-integrable first order derivatives and assume

$$H_0^1(\Omega) = \{v \in H^1(\Omega); v|_{\Gamma_D} = 0\}. \tag{A.1}$$

Given the following diffusive time-dependent problem with variable coefficients:

$$\begin{cases} \zeta \frac{\partial u}{\partial t} - \nabla \cdot \mathbf{b} \nabla u = f & \text{in } \Omega \times [0,T] \\ u = u_D & \text{in } \Gamma_D \\ u(\mathbf{x},0) = u_0(\mathbf{x}) & \text{in } \Omega \quad \text{at } t = 0,\ \mathbf{x} \in \Omega \end{cases}, \tag{A.2}$$

where $u(\mathbf{x},t)$ is the scalar dependent variable, ζ is a non-negative piecewise constant function, $\mathbf{b}(\mathbf{x})$ is a symmetric positive definite matrix function, u_D is a fixed Dirichlet value on the portion Γ_D of Γ, u_0 is the initial state and $f \in L_2(\Omega)$ is a source term. Since for the solution of non linear problems, like the Richards equation, numerical methods (e.g Newton or Picard techniques) are applied to linearized governing equations, in problem (A.2) coefficient ζ and tensor $\mathbf{b}$ are assumed to be non-dependent on variable u.

Let $a(u,v)$ be a bilinear form on $H^1(\Omega) \times H^1(\Omega)$, defined as:

$$a(u,v) = \int_\Omega \mathbf{b} \nabla u \nabla v dx + \int_\Omega \zeta \frac{\partial u}{\partial t} v dx. \tag{A.3}$$

The variational formulation of problem (A.2) is:

$$\begin{cases} \text{find } u \in H^1(\Omega) \quad \text{such that :} \\ a(u,v) = (f,v), \quad v \in H_0^1(\Omega) \\ u = u_D \quad \text{on } \Gamma_D \\ (u(\mathbf{x},0), v) = (u_0(\mathbf{x}), v) \end{cases}, \tag{A.4}$$

where $(\cdot,\cdot)$ denotes the $L_2(\Omega)$ inner product.

A.2. The linear (P1) nonconforming FEM

Call L the total number of sides and N the number of the internal sides plus the boundary sides where Dirichlet conditions do not hold. Let Y_h be the $P1$ nonconforming finite element approximate space of $H^1(\Omega)$:

$$Y_h = \{v_h|_{T_j} \in P1(T_j) : v_h \quad \text{continuous at midpoints } B_i, \quad 1 \leqslant i \leqslant L\}, \tag{A.5}$$

where T_j is the generic triangle of T_h, $j = 1,\ldots,N_T$. We define

$$Y_{h0} = \{v_h \in Y_h : v_h = 0 \quad \text{at } B_i, \quad N+1 \leqslant i \leqslant L\}. \tag{A.6}$$

Observe that $Y_h \not\subset H^1(\Omega)$ and $Y_{h0} \not\subset H_0^1(\Omega)$.

Let λ_S be the nodal linear basis function associated with vertex S_i opposite to side $e^i_{T_j}$ ($i = 1,2,3$) of T_j. λ_S is equal to one at node S_i and zero at the other two nodes (see the red function in Fig. 32). $\omega_{ih}(\mathbf{x})$, $1 \leqslant i \leqslant L$, are piecewise linear basis functions of Y_h, defined by the values of the midpoints of the triangle sides, so that:

$$\omega_{ih}(B_j) = \delta_{ij} \quad i,j = 1,\ldots,L, \tag{A.7}$$

where i and j are indexes of sides and δ_{ij} is the Kronecker delta. Then the set of functions

$$\omega_{ih}(B_j) = \delta_{ij}, \quad i,j = 1,\ldots,N, \tag{A.8}$$

forms the basis of Y_{h0}. Locally, in the triangle T_j, $\omega_{ih}(\mathbf{x})$ is (see the green function in Fig. 32):

$$\omega_{ih}(\mathbf{x}) = 1 - 2\lambda_S(\mathbf{x}) \quad \forall \mathbf{x} \in T_j. \tag{A.9}$$

According to the definitions given in Eqs. (A.7)–(A.9), functions $\omega_{ih}(\mathbf{x})$ are orthogonal and the following relation holds [27]:

$$\int_\Omega \omega_{ih}(\mathbf{x})\omega_{jh}(\mathbf{x})\,d\Omega = \frac{\sigma_i}{3}\delta_{ij}, \tag{A.10}$$

where σ_i is the area of the local support of function $\omega_{ih}(\mathbf{x})$ (the blue quadrilateral in Fig. 32 represents $\sigma_i/3$). It can be easily shown that function ω_i has the following constant gradient over T_j (see also [13]):

$$\nabla\omega_{ih}(\mathbf{x}) = \frac{|e^i_{T_j}|}{|T_j|}\mathbf{n}_i, \tag{A.11}$$

where $\left|e^i_{T_j}\right|$ is the length of side $e^i_{T_j}$ and $\mathbf{n}_i$ the unitary outward normal vector with respect to the same side.

The local representation over triangle T_j of any function $p_h(\mathbf{x}) \in Y_h$ is:

$$p_h(\mathbf{x})|_{T_j} = \sum_{i=1,3} p_i\omega_{ih}(\mathbf{x}). \tag{A.12}$$

Define the lowest order continuous in normal components Raviart–Thomas space [39] as:

$$\mathbf{V}_h = \{\mathbf{g}_h \in H(div;\Omega) : \mathbf{g}_h|_{T_j} \in RT_0(T_j) \quad \forall T_j \in T_h\}, \tag{A.13}$$

where the spaces $H(div;\Omega) = \{\mathbf{w} : \mathbf{w} \in (L_2(\Omega))^2, \nabla\cdot\mathbf{w} \in L_2(\Omega)\}$ and $RT_0(T_j) = \{\mathbf{g} = (g_1,g_2)^T : g_1 = a + bx_1, g_2 = c + bx_2|_{T_j}\}$, with a, b and c real coefficients. Let $w_i^{T_j}$ be the basis functions of $RT_0(T_j)$ [39]:

$$w_i^{T_j}(\mathbf{x}) = \frac{1}{2|T_j|}(\mathbf{x} - \mathbf{x}_i^{T_j}), \quad i = 1,2,3, \quad \forall \mathbf{x} \in T_j, \tag{A.14}$$

where $\left(\mathbf{x}_i^{T_j}\right)$ are the coordinates of the T_j nodes. According to Eq. (A.14), the following relations hold:

$$\frac{1}{2|T_j|}\left(\mathbf{x} - \mathbf{x}_G^{T_j}\right) = \frac{1}{3}\sum_{i=1,3}\frac{1}{2|T_j|}\left(\mathbf{x} - \mathbf{x}_i^{T_j}\right) \tag{A.15a}$$

and

$$w_i^{T_j}(\mathbf{x})\cdot\mathbf{n}_i = \begin{cases} 1\Big/\left|e^i_{T_j}\right| & \forall \mathbf{x} \in e^i_{T_j}, \\ 0 & \text{otherwise}, \end{cases} \tag{A.15b}$$

where $\left(\mathbf{x}_G^{T_j}\right)$ are the coordinates of the centre of mass of T_j.

Assume a linear variation of the unknown variable u inside each triangle according to the values in the midpoints of each side. An approximate solution u_h is sought for in the form of a linear combination of finite elements:

$$u_h(\mathbf{x},t)=\sum_{i=1,L}u_i(t)\omega_{ih}(\mathbf{x}), \tag{A.16}$$

where $u_i(t)$ are the weight factors to be computed. In determining the approximate solution of the original problem (A.2), integral equalities have to be satisfied $\forall v\in H_0^1(\Omega)$, but the basis functions $\omega_{ih}(\mathbf{x})$ have discontinuities at their support boundaries, so the $P1$ nonconforming FE discretization of problem (A.2) is:

$$\begin{cases}\text{find } u_h\in Y_h \quad \text{such that :}\\ a_h(u_h,v_h)=(f,v_h)\quad \forall v_h\in Y_{h0},\\ u_h(B_i)=u_{hD},\quad N+1\leqslant i\leqslant L,\\ u_h(B_i)=u_{h0}(B_i),\quad 1\leqslant i\leqslant L,\end{cases} \tag{A.17}$$

where $u_{hD}\in Y_h$ is chosen so that $u_{hD}(B_i)=u_D(B_i)$, $i=N+1\leqslant i\leqslant L$. In Eq. (A.17) the bilinear form $a_h(u_h,v_h)$ on $Y_h\times Y_h$ is defined as:

$$a_h(u_h,v_h)=\sum_{T_j}\int_{T_j}(\widehat{\mathbf{b}}\,\nabla u_h\cdot\nabla v_h)dT_j+\sum_{T_j}\int_{T_j}\widehat{\zeta}\,\frac{\partial u_h}{\partial t}v_h dT_j, \tag{A.18}$$

where the $\widehat{(\cdot)}$ symbol in Eq. (A.18) marks the mean spatial value in the triangle T_j, $\widehat{\zeta}=\frac{1}{|T_j|}\int_{T_j}\zeta dx$ and $\widehat{\mathbf{b}}=\frac{1}{|T_j|}\int_{T_j}\mathbf{b}dx$ [13].

A.3. Flux construction in the P1 nonconforming FEM

Assume now T_R and T_L the two triangles sharing side $e^i_{T_j}$, call m the midpoint of $e^i_{T_j}$ and let $\omega_{mh}\in Y_h$ be its global basis function, as defined in Eq. (A.7). By setting in Eqs. (A.17) and (A.18) $v_h=\omega_{mh}$, one gets [13]:

$$(\widehat{\mathbf{b}}\,\nabla u_h,\nabla\omega_{mh})_{T_L}-(\widehat{\mathbf{b}}\,\nabla u_h,\nabla\omega_{mh})_{T_R}+\int_{T_L\cup T_R}\widehat{\zeta}\,\frac{\partial u_h(m)}{\partial t}\omega_{mh}\,dx=\int_{T_L\cup T_R}f\omega_{mh}\,dx. \tag{A.19}$$

Using the Green's formula and the property of ω_{mh} to vanish at midpoints of the boundary of $T_L\cup T_R$, according to Eqs. (A.11) and (A.12), one gets:

$$|e_i|[(\widehat{\mathbf{b}}\,\nabla u_h\cdot\mathbf{n}_L)_{T_L}+(\widehat{\mathbf{b}}\,\nabla u_h\cdot\mathbf{n}_R)_{T_R}]+\frac{|T_L|\widehat{\zeta}_{k_L}+|T_R|\widehat{\zeta}_{k_R}}{3}\frac{\partial u_h(m)}{\partial t}=\int_{T_L\cup T_R}f\omega_{mh}\,dx, \tag{A.20}$$

with $\mathbf{n}_L$ and $\mathbf{n}_R$ the unitary vectors orthogonal to side $e^i_{T_j}$, respectively outward from T_L and T_R.

Call $\mathbf{q}_T\in\mathbf{V}_h$ ($\mathbf{V}_h$ defined in Eq. (A.13)) the approximation of Darcy flux $\mathbf{q}$ over element T_j. In the case of stationary problem and without sink/source term, it can be easily shown (see for example [13]) that $\tilde{\mathbf{q}}_T(\mathbf{x})=-\widehat{\mathbf{b}}\,\nabla u_h|_{T_j}$, where the $\tilde{(\cdot)}$ sign marks the flux in the stationary problem without sink/source term. Finally, Eq. (A.20) can be written at midpoint m as:

$$-|e_i|(\tilde{\mathbf{q}}_{T_L}(m)\cdot\mathbf{n}_L+\tilde{\mathbf{q}}_{T_R}(m)\cdot\mathbf{n}_R)+\frac{|T_L|\widehat{\zeta}_{T_L}+|T_R|\widehat{\zeta}_{T_R}}{3}\frac{\partial u_h(m)}{\partial t}=\int_{T_L\cup T_R}f\omega_{mh}\,dx. \tag{A.21}$$

According to Eqs. (A.15b) and (A.21) can be written as:

$$\mathbf{q}_{T_L}(m)\cdot\mathbf{n}_L+\mathbf{q}_{T_R}(m)\cdot\mathbf{n}_R=0, \tag{A.22a}$$

where [13,44]

$$\mathbf{q}_{T_{L(R)}}(m)=\tilde{\mathbf{q}}_{T_{L(R)}}(m)+\frac{\widehat{\mathbf{f}}_{T_{L(R)}}}{3}\sum_{i=1,3}\frac{1}{2}\left(\mathbf{x}-\mathbf{x}_i^{T_{L(R)}}\right)-\frac{\widehat{\zeta}_{T_{L(R)}}}{3}\sum_{i=1,3}\left[\frac{1}{2}\left(\mathbf{x}-\mathbf{x}_i^{T_{L(R)}}\right)\frac{\partial u_h(m)}{\partial t}\right], \tag{A.22b}$$

where $\widehat{\mathbf{f}}_{T_{L(T)}}$ is the mean value of f over element $T_{L(R)}$. Formulations (A.22) represent a balance law related to a control volume stored in the midpoint of $e^i_{T_j}$, which area $(|T_L|+|T_R|)/3$ is given by the sum of 1/3 of the area of each triangle sharing $e^i_{T_j}$. Eqs. (A.22) guarantee the continuity of the flux $\mathbf{q}$ at midpoint of $e^i_{T_j}$. Area $(|T_L|+|T_R|)/3$ is equal to $\sigma_i/3$ (see Eq. (A.10)) and plays the role of the lumping region in the standard conforming FE schemes [33]. This implies that Eqs. (A.22) are equivalent to a lumped MHFE scheme [44].

References

[1] Argus Holdings Ltd. <http://www.argusint.com>.

[2] C. Aricò, T. Tucciarelli, MAST solution of advection problems in irrotational flow fields, Adv. Water Resour. 30 (2007) 665–685, doi:10.1016/j.advwatres.2006.03.007.

[3] C. Aricò, T. Tucciarelli, A Marching in space and time (MAST) solver of the shallow water equations. Part I: the 1D case, Adv. Water Resour. 30 (2007) 1236–1252, doi:10.1016/j.advwatres.2006.11.003.

[4] C. Aricò, C. Nasello, T. Tucciarelli, A Marching in space and time (MAST) solver of the shallow water equations. Part II: the 2D case, Adv. Water Resour. 30 (2007) 1253–1271, doi:10.1016/j.advwatres.2006.11.004.

[5] C. Aricò, T. Tucciarelli, The MAST-FV/FE scheme for the simulation of two-dimensional thermohaline processes in variable density saturated porous media, J. Comput. Phys. 228 (4) (2009) 1234–1274, doi:10.1016/j.jcp.2008.10.015.
[6] C. Aricò, M. Sinagra, L. Begnudelli, T. Tucciarelli, MAST-2D diffusive model for flood prediction on domains with triangular Delaunay unstructured meshes, Adv. Water Resour. 34 (2011) 1427–1449, doi:10.1016/j.advwatres.2011.08.002.
[7] A. Bascià, T. Tucciarelli, An explicit unconditionally stable numerical solution of the advection problem in irrotational flow fields, Water Resour. Res. 40 (6) (2004) W06501, doi:10.1029/2003WR002646.
[8] J. Bear, Hydraulics of Groundwater, McGraw-Hill, New York, 1979.
[9] L. Bergamaschi, M. Putti, Mixed finite elements and Newton-type linearization for the solution of Richards' equation, Int. J. Numer. Methods Eng. 45 (1999) 1025–1046.
[10] F. Brezzi, L. Marini, P. Pietra, Two-dimensional exponential fitting and applications to drift–diffusion models, SIAM J. Numer. Anal. 26 (6) (1989) 1342–1355.
[11] R.J. Brooks, A.T. Corey, Hydraulic properties of porous media, Hydrological paper 3, Colorado State University, Fort Collins, 1964.
[12] M.A. Celia, E.T. Bouloutas, R.L. Zarba, A general mass-conservative numerical solution for the unsaturated flow equation, Water Resour. Res. 26 (1990) 1483–1496.
[13] S.H. Chou, S. Tang, Conservative *P*1 conforming and nonconforming Galerkin FEMS: effective flux evaluation via a nonmixed method approach, SIAM J. Numer. Anal. 28 (2) (2000) 660–680.
[14] H.-J.G. Diersch, P. Perrochet, On the primary variable switching technique for simulating unsaturated–saturated flows, Adv. Water Resour. 23 (1999) 271–301.
[15] P.A. Forsyth, M.C. Kropinski, Monotonicity considerations for saturated–unsaturated subsurface flow, SIAM J. Sci. Comput. 18 (5) (1997) 1254–1328.
[16] P.A. Forsyth, Y.S. Wu, K. Pruess, Robust numerical methods for saturated–unsaturated flow with dry initial conditions in heterogeneous media, Adv. Water Resour. 18 (1995) 25–38.
[17] E.O. Frind, M. Verge, Three-dimensional modeling of groundwater flow systems, Water Resour. Res. 14 (5) (1978) 844–856.
[18] R. Haverkamp, M. Vauclin, L. Touma, P.J. Wierenga, G. Vachaud, Comparison of numerical simulation models for one-dimensional infiltration, Soil Sci. Soc. Am. J. 41 (1977) 285–294.
[19] H. Hoteit, R. Mosé, B. Philippe, P. Ackerer, J. Erhel, About the maximum principle violations of the mixed-hybrid finite element method applied to diffusion equations, Int. J. Numer. Methods Eng. 55 (12) (2002) 1373–1390.
[20] H. Hoteit, J. Erhel, R. Mosé, B. Philippe, P. Ackerer, Numerical reliability for mixed methods applied to flow problems in porous media, Comput. Geosci. 6 (2002) 161–194.
[21] H.-C. Huang, Y.-C. Tan, C.-W. Liu, C.-H. Chen, A novel hysteresis model in unsaturated soil, Hydrol. Process. 19 (2005) 1653–1665.
[22] T.J.R. Hughes, G. Engel, L. Mazzei, M.G. Larson, The Continuous Galerkin method is locally conservative, J. Comp. Phys. 163 (2000) 467–488.
[23] P.S. Huyakorn, S.D. Thomas, B.M. Thompson, Techniques for making finite elements competitive in modeling flow in variably saturated media, Water Resour. Res. 20 (1984) 1099–10115.
[24] M. Kaser, A. Iske, ADER schemes on adaptive triangular meshes for scalar conservation laws, J. Comput. Phys. 205 (2005) 486–508.
[25] M.R. Kirkland, R.G. Hills, P.J. Wierenga, Algorithms for solving Richards' equation for variably saturated soils, Water Resour. Res. 28 (8) (1992) 2049–2058.
[26] J.B. Kool, J.C. Parker, Development and evaluation of closed-form expressions for hysteretic soil hydraulic properties, Water Resour. Res. 23 (1987) 105–114.
[27] V.I. Kuzin, V.V. Kravtchenko, Application of nonconforming finite elements to solving advection-diffusion problems, Numer. Anal. Appl. 3 (1) (2010) 39–51.
[28] E. M LaBolle, V. Clausnitzer, Comment on Russo [1991], Serrano [1990,1998], and other applications of the water-content-based form of Richards' equation to heterogeneous solid, Water Resour. Res. 35 (2) (1999) 605–607.
[29] L. Lapidus, G.F. Pinder, Numerical Solution of Partial Differential Equations in Science and Engineering, John Wiley, New York, 1982.
[30] G. Manzini, S. Ferraris, Mass conservative finite volume method on 2D unstructured grids for the Richards' equation, Adv. Water Resour. 27 (2004) 1199–1215.
[31] C.T. Miller, G.A. Williams, C.T. Kelley, M.D. Tocci, Robust solution of Richards' equation for nonuniform porous media, Water Resour. Res. 34 (10) (1998) 2599–2610.
[32] Nag Library Manual. <http://www.nag.co.uk/numeric/fl/manual/html/mark22.html>, 2010.
[33] K. Ohmori, T. Ushijima, A technique of upstream type applied to a linear nonconforming finite element approximation of convection diffusion equations, Rairo Anal. Numer. 18 (3) (1984) 309–332.
[34] C.M. Oldenburg, K. Pruess, On numerical modeling of capillary barriers, Water Resour. Res. 29 (4) (1993) 1045–1056.
[35] C. Paniconi, M. Putti, A comparison of Picard and Newton iteration in the numerical solution of multidimensional variably saturated flow problems, Water Resour. Res. 30 (12) (1994) 3357–3374.
[36] M. Protter, H. Weinberger, Maximum Principles in Differential Equations, Partial Differential Equations Series, Prentice-Hall, New Jersey, 1967.
[37] M. Putti, C. Cordes, Finite element approximation of the diffusion operator on tetrahedral, SIAM J. Sci. Comput. 19 (4) (1998) 1154–1168.
[38] M. Putti, F. Sartoretto, Linear Galerkin vs mixed finite element 2D flow fields, Int. J. Numer. Methods Fluids 60 (2009) 1011–1031.
[39] P.A. Raviart, J.A. Thomas, A mixed finite element method for second order elliptic problems, Mathematical Aspects of Finite Element Method, Lecture Notes in Mathematics, vol. 606, Springer, New York, 1977.
[40] L.A. Richards, Capillary conduction of liquids through porous mediums, Physics 1 (5) (1931) 318–333, doi:10.1063/1.1745010.
[41] N. Romano, B. Brunone, A. Santini, Numerical analysis of one-dimensional unsaturated flow in layered soils, Adv. Water Resour. 21 (4) (1998) 315–324.
[42] M.D. Tocci, C.T. Kelley, C.T. Miller, Accurate and economical solution of the pressure-head form of Richards' equation by the method of lines, Adv. Water Resour. 20 (1) (1997) 1–14.
[43] M. Van Genuchten, A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, Soil Sci. Soc. Am. J. 44 (5) (1980) 892–898.
[44] A. Younes, P. Ackerer, F. Lehmann, A new mass lumping for the mixed hybrid finite element method, Int. J. Numer. Metods Eng. 67 (2006) 89–107.