SIMULATION OF COMPLEX FREE SURFACE FLOWS

Krisztian Benyo[1], Ayoub Charhabil[2], Mohamed-Ali Debyaoui[3] and Yohan Penel[4]

Abstract. We study the Serre–Green-Naghdi system under a non-hydrostatic formulation, modelling incompressible free surface flows in shallow water regimes. This system, unlike the well-known (nonlinear) Saint-Venant equations, takes into account the effects of the non-hydrostatic pressure term as well as dispersive phenomena. Two numerical schemes are designed, based on a finite volume – finite difference type splitting scheme and iterative correction algorithms. The methods are compared by means of simulations concerning the propagation of solitary wave solutions. The model is also assessed with experimental data concerning the Favre secondary wave experiments [12].

Résumé. L'objet de l'étude est le système de Serre–Green-Naghdi afin de modéliser des écoulements incompressibles à surface libre en eaux peu profondes, tout en tenant compte les effets de la pression non-hydrostatique et les phénomènes dispersifs qui jouent un rôle important en particulier en océanographie côtière. Plusieurs schémas numériques sont présentés, basés sur une méthode de splitting, en séparant la partie hyperbolique et la partie non-hydrostatique. Les algorithmes sont comparés à travers des simulations avec des ondes solitaires. Le modèle est également validé avec des données expérimentales concernant les ondes de Favre [12].

Introduction

The simulation of incompressible, homogeneous, inviscid free surface fluid flows as a tool to describe and examine the propagation of surface water waves has been a major field of research oriented towards oceanographic applications for the past several decades. In general, such a physical system is governed by the free surface Euler equations, which, to this day, remains an analytically and numerically challenging topic.

For coastal, or large scale applications, it is commonly approximated by reduced-complexity models based on the shallowness of the fluid domain. Diverse models based on the (nonlinear) shallow water [or Saint Venant] system [14] have been proposed and analysed in the past (see for instance [4, 5, 8, 11, 23, 24, 26]). Although this model provides a reasonable tool for capturing the nonlinear transformation of waves, it fails to represent dispersive effects because of the underlying hydrostatic assumption.

[1] Laboratoire d'Hydraulique Saint-Venant, EDF R&D, ENPC
[2] Laboratoire Analyse, Géométrie et Applications, Université Paris 13
[3] IMATH – Institut de Mathématiques de Toulon
[4] INRIA Paris – Sorbonne Université – CNRS (LJLL), team ANGE, 2 rue Simone Iff, CS 42112, 75589 Paris cedex 12

The model investigated in the present article is the Serre–Green-Naghdi system (originally appeared in [18, 28]), also known as the fully nonlinear Boussinesq system and referred to as $\text{LDNH}_2(1)$ in [17]. This model is derived from the free surface Euler equations and reads [16, 25] in d-dimension for $d \in \{1, 2\}$

$$
\begin{cases}
\partial_t h + \nabla_{\boldsymbol{x}} \cdot (h\boldsymbol{u}) = 0, & \text{(1a)} \\[4pt]
\partial_t (h\boldsymbol{u}) + \nabla_{\boldsymbol{x}} \cdot (h\boldsymbol{u} \otimes \boldsymbol{u}) + \nabla_{\boldsymbol{x}}(hq) + q_b \nabla_{\boldsymbol{x}} z_b = -gh\nabla_{\boldsymbol{x}}(h + z_b) - h\nabla_{\boldsymbol{x}} p_{atm}, & \text{(1b)} \\[4pt]
\partial_t (hw) + \nabla_{\boldsymbol{x}}(hw\boldsymbol{u}) - q_b = 0, & \text{(1c)} \\[4pt]
\partial_t (h\sigma) + \nabla_{\boldsymbol{x}}(h\sigma\boldsymbol{u}) - 2\sqrt{3}\left(q - \dfrac{q_b}{2}\right) = 0, & \text{(1d)} \\[4pt]
w - \boldsymbol{u} \cdot \nabla_{\boldsymbol{x}} z_b - \sqrt{3}\sigma = 0, & \text{(1e)} \\[4pt]
2\sqrt{3}\sigma + h\nabla_{\boldsymbol{x}} \cdot \boldsymbol{u} = 0, & \text{(1f)}
\end{cases}
$$

where $\boldsymbol{x} = x$, $\boldsymbol{u} = u$ if $d = 1$ or $\boldsymbol{x} = (x, y)$, $\boldsymbol{u} = (u, v)$ if $d = 2$. The equations are set in a smooth bounded domain $\Omega \subset \mathbb{R}^d$.

Unknowns. Here $h : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}^+$ is the fluid height; $\boldsymbol{u} : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}^d$ is the vertically averaged horizontal velocity of the fluid, $w : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$ is the vertically averaged vertical velocity and $\sigma : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$ is the vertical correction to w. Finally $q : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$ is the vertically averaged hydrodynamic pressure and $q_b : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$ is the hydrodynamic pressure at the bottom.

Data. $z_b : \Omega \to \mathbb{R}$ is the time-independent bottom topography such that

$$
\nabla_{\boldsymbol{x}} z_b \cdot \boldsymbol{n} = 0 \text{ over } \partial\Omega, \tag{2}
$$

where $\boldsymbol{n}$ is the outward normal unit vector and $g = 9.81$ is the standard acceleration due to gravity. The system is complemented with multiple possible boundary conditions, depending on the physical setting modeled.

Boundary conditions. The boundary conditions applied to velocity unknowns are detailed in the numerical part of the paper. As for the pressure, we impose

$$
\nabla_{\boldsymbol{x}}(hq) \cdot \boldsymbol{n} = \chi \text{ over } \Gamma_{in}, \tag{3}
$$

$$
q = 0 \text{ over } \Gamma_{out}, \tag{4}
$$

for some partition of $\partial\Omega = \Gamma_{in} \cup \Gamma_{out}$ to be specified later.

The main purpose of the paper is to analyze some of the properties of the system in any dimension and to examine some 1D numerical algorithms. Developing robust and efficient algorithms for the numerical resolution of the Serre–Green-Naghdi system has drawn much attention lately, partially due to the knowledge from theoretical analyses and the surge in computational power (see for instance [6, 13, 15, 20–22, 25]).

Our approach relies on a finite volume based splitting scheme that has shown promising results [1, 2] when applied to the non-hydrostatic shallow water model introduced in [10].

The outline of the document is as follows. In Section 1, we present some properties of the model (equivalent formulations, energy balance, . . .) and we prove that the corresponding projection step is well-posed.

In Section 2, we elaborate the finite volume – finite difference type splitting scheme. It consists in solving the hyperbolic part first, then the dispersive part. The hyperbolic prediction step involves the numerical resolution of the classic shallow water equations with a non-flat bottom topography. As for the resolution of the system arising from the non-hydrostatic correction step – which is the core of this work – we present two iterative algorithms: the Uzawa algorithm applied to the mixed velocity-pressure problem and the Gauss-Seidel method for the projection elliptic step. The different choices for the boundary conditions depending on the physical situation at hand, as well as their consequences on the discretization procedures, are also presented.

In Section [3], we describe some numerical test cases in order to assess the aforementioned algorithms. Simulations concerning the propagation of exact solitary waves over a flat bottom topography are run in order to compare the two algorithms, mainly in terms of computational time. It is followed by simulations of a dispersive dam-break problem. Finally, the numerical validation against the experimental data obtained from the Favre secondary wave experiment [12] is presented and compared with previous results [9].

1. The Serre – Green-Naghdi model

1.1. **Specific operators**

Given the vectors

$$\mathbf{Q} = \begin{pmatrix} q \\ q_b \end{pmatrix} \in \mathbb{R}^2 \quad \text{and} \quad \mathbf{X} = \begin{pmatrix} \boldsymbol{u} \\ w \\ \sigma \end{pmatrix} \in \mathbb{R}^{d+2},$$

we define the dispersive divergence and gradient operators

$$\nabla_{SGN}\mathbf{Q} = \begin{pmatrix} \nabla_{\boldsymbol{x}}(hq) + q_b\nabla_{\boldsymbol{x}}z_b \\ -q_b \\ -2\sqrt{3}\left(q - \frac{q_b}{2}\right) \end{pmatrix} \in \mathbb{R}^{d+2} \quad \text{and} \quad \nabla_{SGN} \cdot \mathbf{X} = \begin{pmatrix} 2\sqrt{3}\sigma + h\nabla_{\boldsymbol{x}} \cdot \boldsymbol{u} \\ w - \boldsymbol{u} \cdot \nabla_{\boldsymbol{x}}z_b - \sqrt{3}\sigma \end{pmatrix} \in \mathbb{R}^2, \qquad (5)$$

and the source term vector

$$S(h, z_b) = (-gh\nabla_{\boldsymbol{x}}(h + z_b) - h\nabla_{\boldsymbol{x}}p_{atm}, 0, 0)^T.$$

Hence System [1] reads under the compact non-conservative form

$$\begin{cases} \partial_t h + \nabla_{\boldsymbol{x}} \cdot (h\boldsymbol{u}) = 0, & (6a) \\ \partial_t(h\mathbf{X}) + \nabla_{\boldsymbol{x}} \cdot (\mathbf{X} \otimes h\boldsymbol{u}) + \nabla_{SGN}\mathbf{Q} = S(h, z_b), & (6b) \\ \nabla_{SGN} \cdot \mathbf{X} = 0. & (6c) \end{cases}$$

Lemma 1.1. *The dispersive operators defined by* [5] *satisfy the Green-type formula:*

$$\mathbf{Q} \cdot (\nabla_{SGN} \cdot \mathbf{X}) = \nabla_{\boldsymbol{x}} \cdot (hq\boldsymbol{u}) - \mathbf{X} \cdot (\nabla_{SGN}\mathbf{Q}). \qquad (7)$$

1.2. **Reformulations of the system**

Let us first mention why we refer to System [6] as the Serre–Green-Naghdi model. Indeed:

Remark 1.2. System [6] can be rewritten under the Boussinesq formulation

$$\begin{cases} \partial_t h + \nabla_{\boldsymbol{x}} \cdot (h\boldsymbol{u}) = 0, \\ (1 + \mathcal{T}[h, z_b])(\partial_t\boldsymbol{u} + (\boldsymbol{u} \cdot \nabla_{\boldsymbol{x}})\boldsymbol{u}) + g\nabla_{\boldsymbol{x}}(h + z_b) + \mathcal{Q}[h, z_b]\boldsymbol{u} = 0, \end{cases}$$

where we denote for $d = 1$

$$\mathcal{T}[h, z_b]v = \mathcal{R}_1[h, z_b](\partial_x v) + \mathcal{R}_2[h, z_b](v\partial_x z_b),$$

$$\mathcal{Q}[h, z_b]v = -2\mathcal{R}_1[h, z_b]\left((\partial_x v)^2\right) + \mathcal{R}_2[h, z_b](v^2\partial_{xx}^2 z_b),$$

$$\mathcal{R}_1[h, z_b]w = -\frac{1}{3h}\partial_x(h^3 w) - \frac{h}{2}w\partial_x z_b,$$

$$\mathcal{R}_2[h, z_b]w = \frac{1}{2h}\partial_x(h^2 w) + w\partial_x z_b.$$

and for $d = 2$

$$\mathcal{T}[h, z_b]\boldsymbol{v} = \mathcal{R}_1[h, z_b](\nabla_{\boldsymbol{x}} \cdot \boldsymbol{v}) + \mathcal{R}_2[h, z_b](\nabla_{\boldsymbol{x}} z_b \cdot \boldsymbol{v}),$$

$$\mathcal{Q}[h, z_b]\boldsymbol{v} = -2\mathcal{R}_1[h, z_b]\left(\partial_x \boldsymbol{v} \cdot \partial_y \boldsymbol{v}^{\perp} + (\nabla_{\boldsymbol{x}} \cdot \boldsymbol{v})^2\right) + \mathcal{R}_2[h, z_b]\left(\boldsymbol{v} \cdot (\boldsymbol{v} \cdot \nabla_{\boldsymbol{x}})\nabla_{\boldsymbol{x}} z_b\right),$$

$$\mathcal{R}_1[h, z_b]f = -\frac{1}{3h}\nabla_{\boldsymbol{x}}(h^3 f) - \frac{h}{2}f\nabla_{\boldsymbol{x}} z_b,$$

$$\mathcal{R}_2[h, z_b]g = \frac{1}{2h}\nabla_{\boldsymbol{x}}(h^2 g) + g\nabla_{\boldsymbol{x}} z_b.$$

Here we made use of the notation $\boldsymbol{v}^{\perp} = (-v_2, v_1)^T$ for $\boldsymbol{v} = (v_1, v_2)^T$.

System (1) has a similar structure to pressure-based problems:

Proposition 1.3. *For smooth solutions, Equations* (6) *can be rewritten as*

$$\begin{cases} \partial_t h + \nabla_{\boldsymbol{x}} \cdot (h\boldsymbol{u}) = 0, & \text{(8a)} \\ h\left[\partial_t \mathbf{X} + (\boldsymbol{u} \cdot \nabla_{\boldsymbol{x}})\mathbf{X}\right] + \nabla_{SGN}\mathbf{Q} = S(h, z_b), & \text{(8b)} \\ -\nabla_{SGN} \cdot \left(\frac{\nabla_{SGN}\mathbf{Q}}{h}\right) = -\nabla_{SGN} \cdot \left(\frac{S(h, z_b)}{h}\right) + \nabla_{SGN} \cdot \left((\boldsymbol{u} \cdot \nabla_{\boldsymbol{x}})\mathbf{X}\right). & \text{(8c)} \end{cases}$$

Remark 1.4. Equation (8c) expands as

$$12\frac{q}{h} - h\nabla_{\boldsymbol{x}} \cdot \left(\frac{\nabla_{\boldsymbol{x}}(hq)}{h}\right) - 6\frac{q_b}{h} - h\nabla_{\boldsymbol{x}} \cdot \left(\frac{q_b}{h}\nabla_{\boldsymbol{x}} z_b\right) = h\Delta_{\boldsymbol{x}}(g(h + z_b) + p_{atm}) + 2h\left[\partial_x \boldsymbol{u} \cdot \partial_y \boldsymbol{u}^{\perp} + (\nabla_{\boldsymbol{x}} \cdot \boldsymbol{u})^2\right],$$

$$\left(4 + |\nabla_{\boldsymbol{x}} z_b|^2\right)\frac{q_b}{h} + \nabla_{\boldsymbol{x}} z_b \cdot \frac{\nabla_{\boldsymbol{x}}(hq)}{h} - 6\frac{q}{h} = (\boldsymbol{u} \cdot \nabla_{\boldsymbol{x}})^2 z_b - \nabla_{\boldsymbol{x}} z_b \cdot \nabla_{\boldsymbol{x}}(g(h + z_b) + p_{atm}).$$

1.3. Energy

The d-dimensional Serre–Green-Naghdi equations (6) admit an energy balance identity.

Proposition 1.5. *Smooth solutions to System* (6) *satisfy the following identity:*

$$\partial_t \left[h\left(\frac{1}{2}|\mathbf{X}|^2 + g\left(\frac{h}{2} + z_b\right) + p_{atm}\right)\right] + \nabla_{\boldsymbol{x}} \cdot \left[h\left(\frac{1}{2}|\mathbf{X}|^2 + g(h + z_b) + p_{atm} + q\right)\boldsymbol{u}\right] = 0.$$

Proof. Let us multiply (1b) by $\boldsymbol{u}$, (1c) by w and (1d) by σ, and sum the results. This leads to

$$\partial_t \left(\frac{1}{2}h|\mathbf{X}|^2\right) + \nabla_{\boldsymbol{x}} \cdot \left(\frac{1}{2}h|\mathbf{X}|^2\boldsymbol{u}\right) + \frac{1}{2}|\mathbf{X}|^2[\partial_t h + \nabla_{\boldsymbol{x}} \cdot (h\boldsymbol{u})] + \nabla_{SGN}\mathbf{Q} \cdot \mathbf{X} + h\boldsymbol{u} \cdot \nabla_{\boldsymbol{x}}(g(h + z_b) + p_{atm}) = 0.$$

By making use of Equation (1a), of the duality relation (7) and since $\mathbf{X}$ satisfies Equations (1e-1f), we obtain the expected result. $\square$

1.4. Variational formulation

Let us focus in this part on the well-posedness of "elliptic" problems of type

$$-\nabla_{SGN} \cdot \left(\frac{\nabla_{SGN}\mathbf{Q}}{h}\right) = \mathbf{f}, \tag{9}$$

where operators are defined in (5), $\mathbf{f}$ is a given source term assumed in $\mathrm{L}^2(\Omega)$ and h is given under the hypothesis

$$\exists\, H_{min}, H_{max} \in \mathbb{R}_+^*, \ \forall\, (t,x) \in \mathbb{R}_+ \times \Omega, \ H_{min} \leq h(t,x) \leq H_{max}. \tag{H}$$

System (9) is supplemented with boundary conditions (3-4). This kind of problem occurs at the continuous level in (8) and in the numerical strategy (see (14) below).

Let us define the following functional space.

Definition 1.6. For $h \in H^1(\Omega)$, let us define the following function space

$$\mathcal{H}_h^1(\Omega) := \{f \in H^1(\Omega) \,:\, f = 0 \text{ over } \Gamma_{out}, \ \|f\|_h < \infty\},$$

endowed with the inner product

$$\langle f, g \rangle_h = \langle f, g \rangle_{L^2(\Omega)} + \langle \nabla_{\boldsymbol{x}}(hf), \nabla_{\boldsymbol{x}}(hg) \rangle_{L^2(\Omega)},$$

and the corresponding norm

$$\|f\|_h^2 := \langle f, f \rangle_h = \|f\|_{L^2(\Omega)}^2 + \|\nabla(hf)\|_{L^2(\Omega)}^2.$$

Lemma 1.7. *Under* (H), *the h-norm is a norm and the space* $\mathcal{H}_h^1(\Omega)$ *is a Hilbert space.*

Due to the duality relation (7) and the boundary conditions (3-4), the variational problem associated to (9) reads:

Find $\mathbf{Q} \in \mathcal{H}_h^1(\Omega) \times \mathrm{L}^2(\Omega)$ such that

$$\forall\, \widetilde{\mathbf{Q}} \in \mathcal{H}_h^1(\Omega) \times \mathrm{L}^2(\Omega), \ \mathcal{A}(\mathbf{Q}, \widetilde{\mathbf{Q}}) := \int_\Omega \frac{1}{h} \nabla_{SGN} \mathbf{Q} \cdot \nabla_{SGN} \widetilde{\mathbf{Q}} \, \mathrm{d}\boldsymbol{x} = \int_\Omega \widetilde{\mathbf{Q}} \cdot \mathbf{f} \, \mathrm{d}\boldsymbol{x} + \int_{\Gamma_{in}} \chi \widetilde{q} \, \mathrm{d}\varsigma. \tag{10}$$

More precisely,

$$\mathcal{A}(\mathbf{Q}, \widetilde{\mathbf{Q}}) = \int_\Omega \frac{1}{h} \left[(\nabla_{\boldsymbol{x}}(hq) + q_b \nabla_{\boldsymbol{x}} z_b) \cdot (\nabla_{\boldsymbol{x}}(h\widetilde{q}) + \widetilde{q}_b \nabla_{\boldsymbol{x}} z_b) + q_b \widetilde{q}_b + 12 \left(q - \frac{q_b}{2} \right) \left(\widetilde{q} - \frac{\widetilde{q}_b}{2} \right) \right] \mathrm{d}\boldsymbol{x}.$$

The *Cauchy-Schwarz* inequality directly shows that $\mathcal{A}$ is continuous over $\mathcal{H}_h^1(\Omega) \times \mathrm{L}^2(\Omega)$. Moreover:

Proposition 1.8. *Under Hypothesis* (H) *and for* $z_b \in \mathcal{C}^1(\Omega)$ *satisfying* (2), *the operator* $\mathcal{A}$ *is coercive over* $\mathcal{H}_h^1(\Omega) \times \mathrm{L}^2(\Omega)$. *Consequently, the variational formulation* (10) *is well-posed.*

Proof. For $\epsilon > 0$ and $\eta > 0$, we observe that

$$|\nabla_{\boldsymbol{x}}(hq) + q_b \nabla_{\boldsymbol{x}} z_b|^2 + q_b^2 + 12 \left(q - \frac{q_b}{2} \right)^2$$

$$= (1 - \epsilon^2) |\nabla_{\boldsymbol{x}}(hq)|^2 + 12(1 - \eta^2)q^2 + \left[4 - \frac{3}{\eta^2} + |\nabla_{\boldsymbol{x}} z_b|^2 \left(1 - \frac{1}{\epsilon^2} \right) \right] q_b^2$$

$$+ \left| \epsilon \nabla(hq) + \frac{1}{\epsilon} q_b \nabla_{\boldsymbol{x}} z_b \right|^2 + 12 \left(\eta q - \frac{1}{\eta} \frac{q_b}{2} \right)^2. \tag{11}$$

Given some η such that

$$\frac{3}{4} < \eta^2 < 1,$$

we set $\gamma = 4 - \frac{3}{\eta^2} > 0$ and impose that

$$\frac{1}{1 + \frac{\gamma}{\sup_\Omega |\nabla z_b|^2}} < \epsilon^2 < 1.$$

Hence all terms in (11) are positive and

$$\mathcal{A}(\mathbf{Q}, \mathbf{Q}) \geq \frac{1}{H_{max}} \min\left\{ 1 - \epsilon^2, 12(1 - \eta^2), 4 - \frac{3}{\eta^2} + \sup_\Omega |\nabla_{\boldsymbol{x}} z_b|^2 \left(1 - \frac{1}{\epsilon^2}\right) \right\} \|\mathbf{Q}\|_{\mathcal{H}_h^1 \times L^2}.$$

Consequently, the *Lax-Milgram* theorem ensures that the variational formulation (10) is well-posed.   □

2. Numerical strategy

2.1. **The splitting scheme**

Let us take $n \in \mathbb{N}$ and discretize the system in time. Given a series of timesteps Δt^k, $k \in \mathbb{N}$ (defined later by an appropriate CFL condition) giving rise to the times $t^n = \sum_{k \leq n} \Delta t^k$, one has that for the step $n + 1$, the system can be split into two parts at the semi-discrete (in time) level:

(1) A prediction step for the hyperbolic part of the system:

$$\frac{h^{n+1/2} - h^n}{\Delta t^{n+1}} + \nabla_{\boldsymbol{x}} \cdot (h^n \boldsymbol{u}^n) = 0,$$
$$\frac{(h\mathbf{X})^{n+1/2} - (h\mathbf{X})^n}{\Delta t^{n+1}} + \nabla_{\boldsymbol{x}} \cdot (\mathbf{X}^n \otimes h^n \boldsymbol{u}^n) = S(h^n, z_b), \tag{12}$$

for variables h and $h\mathbf{X}$. Notice that this system is essentially the *shallow water* equations with uneven bottom topography: its resolution is classic and well-known. In the numerical simulations to follow (see Section 3), we chose a first order finite volume approach, a Rusanov solver with hydrostatic reconstruction for the scheme to be well-balanced [3].

(2) A projection / correction step involving the non-hydrostatic (dispersive) part of the system:

$$\begin{cases} \dfrac{h^{n+1} - h^{n+1/2}}{\Delta t^{n+1}} = 0, & \text{(13a)} \\[2mm] \dfrac{(h\mathbf{X})^{n+1} - (h\mathbf{X})^{n+1/2}}{\Delta t^{n+1}} + \nabla_{SGN} \mathbf{Q}^{n+1} = 0, & \text{(13b)} \\[2mm] \nabla_{SGN} \cdot \mathbf{X}^{n+1} = 0, & \text{(13c)} \end{cases}$$

for variables $\mathbf{X}$ and $\mathbf{Q}$. The novelty of the problem lies in the efficient resolution of the latter nonlinear problem. In what follows, we detail two techniques of discretization, both of them make use of some particular (structural) properties of the system and both of them are iterative methods for computational efficiency reasons.

Notice that Equations (13b-13c) imply

$$-\nabla_{SGN} \cdot \left(\frac{\nabla_{SGN}\mathbf{Q}}{h^{n+1/2}}\right) = -\frac{1}{\Delta t^{n+1}} \nabla_{SGN} \cdot \mathbf{X}^{n+1/2}. \tag{14}$$
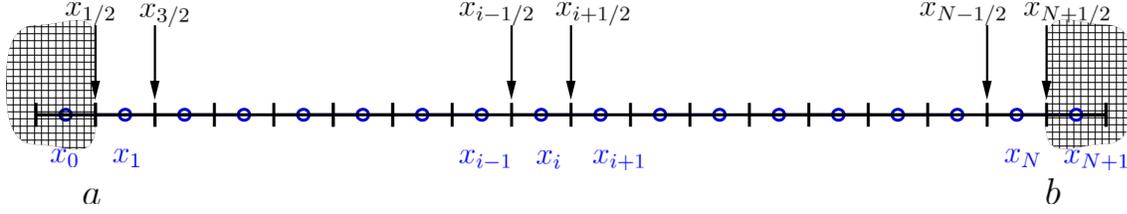
FIGURE 1. Schematic representation of the staggered grid

2.2. **Spatial discretization in 1D**

Let us give some details on the discretization in space in the case $d = 1$. We consider a (bounded) interval $I = [a, b]$. The mesh is defined as a homogeneous Cartesian grid on I, with $N \in \mathbb{N}^*$ cells of constant size $\Delta x = (b - a)/N$.

In the collocated framework, all variables are defined at the cell centers x_i ($i \in \{0, \ldots, N + 1\}$) with x_0 and x_{N+1} corresponding to ghost cells at the boundaries (see Figure 1). The use of a collocated scheme (especially for the pressure variables (q, q_b)) is justified by the special form of the dispersive gradient $\nabla_{SGN} \mathbf{Q}$.

Hyperbolic step: System (12) consists of the classic $1D$ shallow water equations for (h, hu) and two transport equations for $(hw, h\sigma)$. As it was already mentioned, a first order finite volume scheme was implemented:

$$\frac{(h\mathbf{X}^*)_i^{n+1/2} - (h\mathbf{X}^*)_i^n}{\Delta t^{n+1}} + \frac{\mathcal{F}_{i+1/2}^n - \mathcal{F}_{i-1/2}^n}{\Delta x} = S(h_{i-1}^n, h_i^n, h_{i+1}^n, z_b),$$

where $\mathbf{X}^* = (1, \mathbf{X}^T)^T$. The numerical fluxes $\mathcal{F}_{i+1/2}^n$ ($i \in \{0, \ldots, N\}$) approximate the physical fluxes at the interfaces; for the simulations, a *Rusanov* solver was chosen as an approximation.

Such a numerical scheme is stable, provided the following *Courant–Friedrichs–Lewy* (CFL) condition holds:

$$\Delta t^{n+1} \leq c_{\mathcal{F}} \frac{\Delta x}{\max\limits_{i \in \{1, \ldots, N\}} (|u_i^n| + \sqrt{gh_i^n})},$$

where $c_{\mathcal{F}}$ is a constant depending on the chosen numerical solver. For the treatment of boundary conditions, we refer to Section 2.5.

Velocity-pressure system: System (13) is a Stokes-type system for which the water height is given:

$$h_i^{n+1} = h_i^{n+1/2}.$$

Let us denote by $\mathcal{H}^{n+1} \in \mathcal{M}_{N,N}(\mathbb{R})$ the diagonal $N \times N$ matrix with the corresponding diagonal elements h_i^{n+1}, and by $\overline{\mathcal{H}}^{n+1} \in \mathcal{M}_{3N,3N}(\mathbb{R})$ the $3N \times 3N$ block-diagonal matrix with block entries $\mathcal{H}^{n+1}$.

The dispersive gradient $\nabla_{SGN} Q^{n+1}$ is discretized as $B\mathbf{Q}^{n+1}$, as part of a finite difference type discretization scheme on the mesh of the interval I. Here, and in what follows throughout this section, boldfaced vectors refer to the discretized values over all cells ($i \in \{1, \ldots, N\}$), such as for instance

$$\mathbf{Q}^{n+1} = (q_1^{n+1}, q_2^{n+1}, \ldots, q_N^{n+1}, q_{b,1}^{n+1}, q_{b,2}^{n+1}, \ldots, q_{b,N}^{n+1})^T.$$

The matrix $B \in \mathcal{M}_{3N,2N}(\mathbb{R})$ has a relatively simple structure:

$$B = \begin{pmatrix} B_{11} & B_{12} \\ 0_{N,N} & -\operatorname{Id}_N \\ -2\sqrt{3}\operatorname{Id}_N & \sqrt{3}\operatorname{Id}_N \end{pmatrix},$$

where the matrix $B_{12} \in \mathcal{M}_{N,N}(\mathbb{R})$ is diagonal, with elements $(\partial_x z_b)_i$, and $B_{11} \in \mathcal{M}_{N,N}(\mathbb{R})$ is an appropriate discretization for the term $\partial_x(hq)$. We consider two different finite difference schemes for this

(1) <u>Second order central scheme on mesh points</u>: for any interior cell ($i \in \{2, \ldots, N-1\}$) we have that

$$(\partial_x(hq))_i \simeq \frac{1}{\Delta x}\left(h_{i+1/2}q_{i+1/2} - h_{i-1/2}q_{i-1/2}\right) = \frac{(h_{i+1}+h_i)(q_{i+1}+q_i) - (h_i+h_{i-1})(q_i+q_{i-1})}{4\Delta x}. \tag{15}$$

Here, cell boundaries are $x_{i+1/2}$ ($i \in \{0, \ldots, N\}$) and the corresponding indices refer to the averaged values

$$h_{i+1/2} = \frac{h_i + h_{i+1}}{2}.$$

For this discretization, the corresponding tridiagonal matrix B_{11}^1 has

$$\frac{1}{2\Delta x}\left(\ldots \quad -h_{i-1/2} \quad (h_{i+1/2} - h_{i-1/2}) \quad h_{i+1/2} \quad \ldots \right)$$

as the ith row.

(2) <u>Second order central scheme on adjacent cells</u>: for any interior cell ($i \in \{2, \ldots, N-1\}$) we have that

$$(\partial_x(hq))_i \simeq \frac{h_{i+1}q_{i+1} - h_{i-1}q_{i-1}}{2\Delta x}. \tag{16}$$

For this discretization, the corresponding tridiagonal matrix B_{11}^2 has

$$\frac{1}{2\Delta x}\left(\ldots \quad -h_{i-1} \quad 0 \quad h_{i+1} \quad \ldots \right)$$

as the ith row.

When not specified, matrices B_{11} and B can refer to either of these discretization schemes. Notice that the first and last rows of B_{11} may contain only the truncated discretized formulae; the missing terms in the discretization correspond to information on the two ghost cells. With the establishment of the boundary conditions, the values on the ghost cells call for corrections on the first and last elements of the diagonals, up to an error term $\widehat{0}$. The exact corrections to be made depend on the boundary conditions chosen, they will be specified in Section 2.5.

With this at our disposal, Equation (13b) is discretized as

$$\frac{\overline{\mathcal{H}}^{n+1}\mathbf{X}^{n+1} - \overline{\mathcal{H}}^{n+1}\mathbf{X}^{n+1/2}}{\Delta t^{n+1}} + B\mathbf{Q}^{n+1} = \widehat{\mathbf{0}},$$

where $\widehat{\mathbf{0}}$ is the error terms (independent from the variables $\mathbf{X}$ and $\mathbf{Q}$) due to the correction associated to the boundary conditions.

Owing to the duality relation (5), Equation (13c) admits a natural discretization: $B^T\mathbf{X}^{n+1} = \widetilde{\mathbf{0}}$ since B^T is a suitable discretization of the dispersive divergence, up to a boundary term correction induced error term of $\widetilde{0}$ (independent from variables $\mathbf{X}$ and $\mathbf{Q}$).

To sum up, we are left with the resolution of the following linear system:

$$
\begin{pmatrix} \frac{1}{\Delta t^{n+1}} \overline{\mathcal{H}}^{n+1} & B \\ B^T & 0_{2N,2N} \end{pmatrix} \begin{pmatrix} \mathbf{X}^{n+1} \\ \mathbf{Q}^{n+1} \end{pmatrix} = \begin{pmatrix} \frac{1}{\Delta t^{n+1}} \overline{\mathcal{H}}^{n+1} \mathbf{X}^{n+1/2} + \widehat{\mathbf{0}} \\ \widetilde{\mathbf{0}} \end{pmatrix}.
\tag{17}
$$

In the next part, we detail two iterative algorithms for the resolution of this system, both of them using the symmetric structure of (17). We chose the iterative approach to circumvent solving directly the system of $5N$ equations.

2.3. **Iterative methods: the Uzawa algorithm**

Given the symmetric nature of System (17), a natural alternative to directly solving the mixed velocity-pressure problem is to consider the Uzawa algorithm. This consists in the following consecutive iterations:

$$
\begin{cases} \mathbf{X}^{n+1,p+1} = \mathbf{X}^{n+1/2} - \Delta t^{n+1} \left(\overline{\mathcal{H}}^{n+1} \right)^{-1} B\mathbf{Q}^{n+1,p} + \Delta t^{n+1} \left(\overline{\mathcal{H}}^{n+1} \right)^{-1} \widehat{\mathbf{0}}, \\ \mathbf{Q}^{n+1,p+1} = \mathbf{Q}^{n+1,p} + \alpha(B^T \mathbf{X}^{n+1,p+1} - \widetilde{\mathbf{0}}), \end{cases}
\tag{18}
$$

where α is an iterative parameter and $p \in \mathbb{N}$ is the current iteration index. We remark that the matrix $\overline{\mathbb{H}}$ is diagonal, its inversion induces no additional computational cost. As the initialization of the iterative process, $\mathbf{Q}^{n+1,0} = \mathbf{Q}^n$ is taken (since $\mathbf{Q}$ is not involved in the prediction step of the algorithm).

One additional concern about the initialization has to be adressed, which is the very first time step. Contrary to the variables in $\mathbf{X}$, which are natural, physically easily measurable quantities, *a priori* knowledge on the hydrodynamic pressure components Q is rarely available. Throughout the implementation of the algorithm, an (arbitrary) initial quess of 0 was considered for these variables. This can obviously result in a high number of iterations during the first timestep, which can be spread out on multiple timesteps by imposing a maximal number of iterations $p \leq p_{max}$ (with p_{max} being constant).

Apart from the upper bound on the iterations, the Uzawa algorithm is run until the error on the correction is reasonably small, namely an error threshold of ϵ. The error itself is measured as the discrete L^∞-norm of the difference between two consecutive iterations, that is:

$$
\|\mathbf{Q}^{n+1,p+1} - \mathbf{Q}^{n+1,p}\|_{L^\infty} \leq \epsilon.
$$

Notice that System (18) can be simplified into an iterative algorithm on $\mathbf{Q}$ only:

$$
\mathbf{Q}^{n+1,p+1} = \left(\mathrm{Id}_{2N} - \alpha \Delta t^{n+1} B^T (\overline{\mathcal{H}}^{n+1})^{-1} B \right) \mathbf{Q}^{n+1,p} + \alpha(B^T \mathbf{X}^{n+1/2} + \Delta t^{n+1} B^T (\overline{\mathcal{H}}^{n+1})^{-1} \widehat{\mathbf{0}} - \widetilde{\mathbf{0}}),
\tag{19}
$$

which gives direct information on the convergence of the Uzawa algorithm [19] as a function of the eigenvalues of the matrix $\Delta t^{n+1} C := \Delta t^{n+1} B^T (\overline{\mathcal{H}}^{n+1})^{-1} B$.

Lemma 2.1. *The algorithm converges iff the parameter α satisfies*

$$
0 < \alpha < \frac{2}{\Lambda_{max}},
$$

where Λ_{max} is the largest eigenvalue of the aforementioned matrix $\Delta t^{n+1} C$. The optimal iterative parameter is given by

$$
\alpha_{opt} = \frac{2}{\Lambda_{min} + \Lambda_{max}},
$$

Λ_{min} being the smallest eigenvalue of $\Delta t^{n+1} C$.

However, obtaining information on the eigenvalues of this matrix is highly nontrivial, except for very particular cases (for example a lake at rest situation with flat bottom topography).

2.4. **Iterative methods: the Gauss-Seidel approach**

Multiplying the first equation of (17) by $\Delta t^{n+1} B^T (\overline{\mathcal{H}})^{-1}$ and substituting the second equation in it yields

$$\Delta t^{n+1} C \mathbf{Q}^{n+1} \left(= \Delta t^{n+1} B^T (\overline{\mathcal{H}}^{n+1})^{-1} B \mathbf{Q}^{n+1} \right) = B^T \mathbf{X}^{n+1/2} + \Delta t^{n+1} B^T (\overline{\mathcal{H}}^{n+1})^{-1} \widehat{\mathbf{0}} - \widetilde{\mathbf{0}} =: \mathbf{f}, \qquad (20)$$

an equation similar to (19), with $\mathbf{f} = (\mathbf{f}_q^T, \mathbf{f}_{q_b}^T)^T$ as the right hand side.

Remark 2.2. In fact, this corresponds to a direct discretization of System (14).

The symmetric matrix C has a particular block structure of the following form

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{12}^T & C_{22} \end{pmatrix},$$

with C_{11} and C_{22} being symmetric and positive definite $N \times N$ matrices.

Direct resolution of (20), much like the full velocity pressure system (17), may be time consuming, even though the matrix possesses a pleasant "quasi-diagonal" blockwise structure. Since C_{22} is, in fact, a diagonal matrix, one could once again implement an Uzawa type iterative algorithm for the resolution of the discretized elliptic problem.

However, we present instead a block Gauss-Seidel type method. This basically entails in performing a simple block-LU decomposition to separate C_{12} for the explicit part of the algorithm and keeping the rest as an implicit part. For the iterative step $k \in \mathbb{N}$ we have the following two subsequent iterations to perform:

$$\begin{cases} C_{11} \mathbf{q}^{n+1,p+1} = \frac{1}{\Delta t^{n+1}} \mathbf{f} - C_{12} \mathbf{q}_b^{n+1,p}, \\ C_{22} \mathbf{q}_b^{n+1,p+1} = \frac{1}{\Delta t^{n+1}} \mathbf{f}_b - C_{12}^T \mathbf{q}^{n+1,p+1}, \end{cases} \qquad (21)$$

up to a maximal iteration of $p \leq p_{max}$, or sufficiently small error term, the error being measured by the norm of the relative difference between two subsequent iterations, exactly the same way as in the Uzawa algorithm. Once again, we initialize by taking $\mathbf{q}_b^{n+1,0} = \mathbf{q}_b^n$.

Owing to the positiveness of matrices C_{11} and C_{22} [27], one can show that

Lemma 2.3. *The block Gauss-Seidel type algorithm converges (with respect to the norm induced by the matrix C).*

2.5. **Boundary conditions**

As far as boundary conditions are concerned, the three important variables are h, u (or equivalently the flux hu) and q (or its conservative counterpart hq). In what follows, we analyze different boundary conditions: the solid wall boundaries (mixed Dirichlet and Neumann boundary conditions) and a fluvial type inflow-outflow setting. We remark that boundary conditions on the boundary pressure term q_b have no actual influence on the algorithm since boundary terms for q_b do not appear in the scheme at any given moment.

In what follows we specify not only the appropriate boundary conditions but the calculations concerning the related error terms for both discretization schemes: $(\star)$ the central scheme on mesh points and $(\star\star)$ the central scheme on adjacent cells.

2.5.1. *Solid wall boundaries*

We assume that the boundary is a mirror, meaning that the spatial derivative (normal derivative) of quantities vanishes, except for the horizontal velocity, which simply vanishes at the boundary. In particular, on the left hand side boundary, we have that

$$(\partial_x h)_{i=1/2} = 0, \quad u_{i=1/2} = 0, \quad (\partial_x w)_{i=1/2} = 0, \quad \text{and } (\partial_x \sigma)_{i=1/2} = 0,$$

that is, on the ghost cells $i = 0$ these quantities are defined through basic discretization as

$$h_0 = h_1, \quad u_0 = -u_1, \quad w_0 = w_1, \quad \text{and } \sigma_0 = \sigma_1.$$

For the Dirichlet boundary condition on the velocity u, we impose Neumann boundary condition on the pressure q, as well as on q_b, therefore

$$(\partial_x q)_{i=1/2} = 0, \quad (\partial_x q_b)_{i=1/2} = 0 \quad \Longrightarrow \quad q_0 = q_1, \quad q_{b0} = q_{b1}.$$

Similarly, one has on the right hand side boundary:

$$h_{N+1} = h_N, \quad u_{N+1} = -u_N, \quad w_{N+1} = w_N, \quad \sigma_{N+1} = \sigma_N, \quad q_{N+1} = q_N, \quad \text{and } q_{bN+1} = q_{bN}.$$

$(\star)$ With the boundary conditions at our disposal, we can write the discretization of the momentum equation for the first cell:

$$\frac{h_1 u_1^{n+1} - h_1 u_1^{n+1/2}}{\Delta t} + \frac{1}{2\Delta x}(h_{3/2}(q_1 + q_2) - h_{1/2}(q_0 + q_1)) + q_{b1}(\partial_x z_b)_1 = 0.$$

Since $q_0 = q_1$, this means that

$$(B^1 \mathbf{Q})_1 = \frac{1}{2\Delta x}(h_{3/2} - 2h_{1/2})q_1 + \frac{1}{2\Delta x}h_{3/2}q_2 + q_{b1}(\partial_x z_b)_1,$$

and that no correction is required, $\widehat{0}_1 = 0$.

So, given the matrix B^1, we can compute the approximation for the divergence

$$((B^1)^T \mathbf{X})_1 = \frac{1}{2\Delta x}(h_{3/2} - 2h_{1/2})u_1 - \frac{1}{2\Delta x}h_{3/2}u_2 - 2\sqrt{3}\sigma_1$$

$$= -\frac{1}{2\Delta x}(h_{3/2}(u_2 - u_1) + h_{1/2}(u_1 + u_1)) - 2\sqrt{3}\sigma_1.$$

Given that $u_1 = -u_0$, we obtain that

$$((B^1)^T \mathbf{X})_1 = -\frac{1}{2}\left(h_{3/2}\frac{u_2 - u_1}{\Delta x} + h_{1/2}\frac{u_1 - u_0}{\Delta x}\right) - 2\sqrt{3}\sigma_1,$$

which is exactly a first order discretization of $-(h\partial_x u + 2\sqrt{3}\sigma)$. Therefore no corrections are necessary either; $\widetilde{0}_1 = 0$.

With the same reasoning, we have that

$$(B^1 \mathbf{Q})_N = -\frac{1}{2\Delta x}h_{N-1/2}q_{N-1} + \frac{1}{2\Delta x}(2h_{N+1/2} - h_{N-1/2})q_N + q_{bN}(\partial_x z_b)_N,$$

and that corrections are not necessary neither for the gradient nor for the divergence: $\widehat{0}_N = 0, \widetilde{0}_N = 0$.

($\star\star$) With the boundary conditions at our disposal, we can write the discretization of the momentum equation for the first cell:

$$\frac{h_1 u_1^{n+1} - h_1 u_1^{n+1/2}}{\Delta t} + \frac{1}{2\Delta x}(h_2 q_2 - h_0 q_0) + q_{b1}(\partial_x z_b)_1 = 0.$$

Since $q_0 = q_1$, and that $h_0 = h_1$, this means that

$$(B^2 \mathbf{Q})_1 = -\frac{1}{2\Delta x} h_1 q_1 + \frac{1}{2\Delta x} h_2 q_2 + q_{b1}(\partial_x z_b)_1,$$

and that no correction is required, $\widehat{0}_1 = 0$.

So, given the matrix B^2, we can compute the approximation for the divergence

$$((B^2)^T \mathbf{X})_1 = -\frac{1}{2\Delta x} h_1 u_1 - \frac{1}{2\Delta x} h_1 u_2 - 2\sqrt{3}\sigma_1.$$

Given that $u_1 = -u_0$ we obtain that

$$((B^2)^T \mathbf{X})_1 = -h_1 \frac{u_2 - u_0}{2\Delta x} - 2\sqrt{3}\sigma_1,$$

which is exactly a first order discretization of $-(h\partial_x u + 2\sqrt{3}\sigma)$. Therefore no corrections are necessary either; $\widetilde{0}_1 = 0$.

With the same reasoning, we have that

$$(B^2 \mathbf{Q})_N = -\frac{1}{2\Delta x} h_{N-1} q_{N-1} + \frac{1}{2\Delta x} h_N q_N + q_{bN}(\partial_x z_b)_N,$$

and that corrections are not necessary neither for the gradient nor for the divergence: $\widehat{0}_N = 0$, $\widetilde{0}_N = 0$.

2.5.2. *Fluvial inflow-outflow*

For this case, we suppose that we are in a fluvial regime, meaning that $|u| < \sqrt{gh}$. In this case, it is enough to prescribe either the fluid height or the flow flux at each of the boundaries. In our case we will prescribe a constant inflow flux $Q_{inc} > 0$ on the left hand side boundary and a constant height H_{out} on the right hand side boundary. The missing variables (h or hu respectively) can be recovered due to the underlying hyperbolic problem and the conservation of the associated Riemann-invariant on the outgoing characteristic.

Incoming flux on the left hand side: We prescribe a Dirichlet type condition on the flux, that is $(hu)_{i=1/2} = Q_{inc}$. This means that

$$\frac{h_0 u_0 + h_1 u_1}{2} = Q_{inc} \implies h_0 u_0 = 2Q_{inc} - h_1 u_1.$$

Since we are in a fluvial regime, the outgoing characteristic $u + \sqrt{gh}$ is always positive, therefore the corresponding Riemann-invariant is preserved. This yields a second equation on h_0 and u_0:

$$u_0 + 2\sqrt{gh_0} = u_1 + 2\sqrt{gh_1}.$$

Expressing u_0 from one equation and substituting it in the other one leads to a third order polynomial in $\sqrt{h_0}$:

$$2\sqrt{g}(\sqrt{h_0})^3 - (u_1 + 2\sqrt{gh_1})h_0 + (2Q_{inc} - h_1 u_1) = 0.$$

Given that it could be expensive to solve this nonlinear equation, one can also take an extrapolation for h_0, namely $h_0 = h_1$ or $h_0 = 2h_1 - h_2$.

Once h_0 is determined, one can express u_0 from the inflow flux condition, so we have that

$$u_0 = \frac{2Q_{inc} - h_1 u_1}{h_0}.$$

For the vertical compontents of the velocity, we impose Neumann boundary condition, just as before, hence

$$(\partial_x w)_{i=1/2} = 0, \quad (\partial_x \sigma)_{i=1/2} = 0 \quad \Longrightarrow \quad w_0 = w_1, \quad \sigma_0 = \sigma_1.$$

Since we imposed a Dirichlet type boundary condition on the velocity u, we prescribe a Neumann type boundary condition on the conservative pressure term hq and q_b:

$$(\partial_x (hq))_{i=1/2} = 0, \quad (\partial_x q_b)_{i=1/2} = 0 \quad \Longrightarrow \quad q_0 = \frac{h_1}{h_0} q_1, \quad q_{b0} = q_{b1}.$$

($\star$) With the boundary conditions at our disposal, we can write the discretization of the momentum equation for the first cell:

$$\frac{h_1 u_1^{n+1} - h_1 u_1^{n+1/2}}{\Delta t} + \frac{1}{2\Delta x}(h_{3/2}(q_1 + q_2) - h_{1/2}(q_0 + q_1)) + q_{b1}(\partial_x z_b)_1 = 0.$$

Since $q_0 = \frac{h_1}{h_0} q_1$, this means that

$$(B^1 \mathbf{Q})_1 = \frac{1}{2\Delta x}\left(h_{3/2} - h_{1/2}\left(1 + \frac{h_1}{h_0}\right)\right) q_1 + \frac{1}{2\Delta x}h_{3/2}q_2 + q_{b1}(\partial_x z_b)_1,$$

and that no correction is required, $\widehat{0}_1 = 0$.

So, given the matrix B^1, we can compute the approximation for the divergence

$$((B^1)^T \mathbf{X})_1 = \frac{1}{2\Delta x}\left(h_{3/2} - h_{1/2}\left(1 + \frac{h_1}{h_0}\right)\right) u_1 - \frac{1}{2\Delta x}h_{3/2}u_2 - 2\sqrt{3}\sigma_1$$

$$= -\frac{1}{2\Delta x}\left(h_{3/2}(u_2 - u_1) + h_{1/2}\left(1 + \frac{h_1}{h_0}\right) u_1\right) - 2\sqrt{3}\sigma_1.$$

Given that $u_1 = \frac{2Q_{inc} - h_0 u_0}{h_1}$, we obtain that

$$((B^1)^T \mathbf{X})_1 = -\frac{1}{2}\left(h_{3/2}\frac{u_2 - u_1}{\Delta x} + h_{1/2}\frac{u_1 - u_0}{\Delta x}\right) - \frac{1}{\Delta x}\frac{h_{1/2}}{h_0}Q_{inc} - 2\sqrt{3}\sigma_1,$$

which is a first order discretization of $-(h\partial_x u + 2\sqrt{3}\sigma)$ up to a correction term $\widetilde{0}_1 = \frac{1}{\Delta x}\frac{h_{1/2}}{h_0}Q_{inc}$. It is at this point that we need that h_0 is not a function of u_1.

($\star\star$) Here we only treat the special case when $h_0 = h_1$. With the boundary conditions at our disposal, we can write the discretization of the momentum equation for the first cell:

$$\frac{h_1 u_1^{n+1} - h_1 u_1^{n+1/2}}{\Delta t} + \frac{1}{2\Delta x}(h_2 q_2 - h_0 q_0) + q_{b1}(\partial_x z_b)_1 = 0.$$

Since $q_0 = q_1$, and our additional hypothesis on h_0, this means that

$$(B^2 \mathbf{Q})_1 = -\frac{1}{2\Delta x}h_1 q_1 + \frac{1}{2\Delta x}h_2 q_2 + q_{b1}(\partial_x z_b)_1,$$

and that no correction is required, $\widehat{0}_1 = 0$.

So, given the matrix B^2, we can compute the approximation for the divergence

$$((B^2)^T \mathbf{X})_1 = -\frac{1}{2\Delta x} h_1 u_1 - \frac{1}{2\Delta x} h_1 u_2 - 2\sqrt{3}\sigma_1.$$

Given that $u_1 = \frac{2Q_{inc} - h_0 u_0}{h_1}$ and that $h_0 = h_1$, we obtain that

$$((B^2)^T \mathbf{X})_1 = -h_1 \frac{u_2 - u_0}{2\Delta x} - \frac{h_1 Q_{inc}}{\Delta x} - 2\sqrt{3}\sigma_1,$$

which is a first order discretization of $-(h\partial_x u + 2\sqrt{3}\sigma)$ up to the error term $\widetilde{0}_1 = \frac{h_1 Q_{inc}}{\Delta x}$.

Remark 2.4. Here we consider a Neumann boundary condition for w and σ. If we rather impose a Dirichlet boudnary condition with a vertical component of the incoming flux Q_{vert}, then one has to take into account the divergence constraint. This means that, while one imposes the vertical flux on hw, $h\sigma$ on the boundary has to verify the corresponding equation, yielding:

$$w_0 = \frac{2Q_{vert} - h_1 w_1}{h_0}, \qquad \sigma_0 = \frac{2(Q_{vert} - Q_{inc} \cdot (\partial_x z_b)|_{i=1/2}) - \sqrt{3}h_1\sigma_1}{\sqrt{3}h_0}.$$

Outgoing flow with prescribed height: For the outgoing flow, we impose a Dirichlet type boundary condition for the fluid height:

$$h_{i=N+1/2} = H_{out} \implies h_{N+1} = 2H_{out} - h_N.$$

Once again, the outgoing characteristic of the hyperbolic problem preserves the corresponding Riemann-invariant, so

$$u_N + 2\sqrt{gh_N} = u_{N+1} + 2\sqrt{gh_{N+1}}.$$

This means that the velocity on the right hand side phantom cell is given by

$$u_{N+1} = u_N + 2\sqrt{g}(\sqrt{2H_{out} - h_N} - \sqrt{h_N}).$$

For the vertical compontents of the velocity, we impose Neumann boundary condition, just as before, hence

$$(\partial_x w)_{i=N+1/2} = 0, \quad (\partial_x \sigma)_{i=N+1/2} = 0 \implies w_{N+1} = w_N, \quad \sigma_{N+1} = \sigma_N.$$

Since we imposed a Neumann type boundary condition on the velocity u, we prescribe a Dirichlet type boundary condition on the pressure terms. However, we have to distinguish the two discretizations when defining whether we impose Dirichlet type boundary conditions on the conservative variable hq or the non-conservative variable q to ensure the compatibility for the transposed quantities.

($\star$) In this case, we impose the Dirichlet type boundary condition on the nonconservative variable q:

$$q_{i=N+1/2} = 0, \quad q_{b,i=N+1/2} = 0 \implies q_{N+1} = -q_N, \quad q_{b,N+1} = -q_{b,N}.$$

With the boundary conditions at our disposal, we can write the discretization of the momentum equation for the first cell:

$$\frac{h_N u_N^{n+1} - h_N u_N^{n+1/2}}{\Delta t} + \frac{1}{2\Delta x}\big(h_{N+1/2}(q_N + q_{N+1}) - h_{N-1/2}(q_{N-1} + q_N)\big) + q_{b,N}(\partial_x z_b)_N = 0.$$

Since $q_{N+1} = -q_N$, this means that

$$(B^1 \mathbf{Q})_N = -\frac{1}{2\Delta x} h_{N-1/2} q_{N-1} - \frac{1}{2\Delta x} h_{N-1/2} q_N + q_{b,N} (\partial_x z_b)_N,$$

and that no correction is required, $\widehat{0}_N = 0$.

So, given the matrix B^1, we can compute the approximation for the divergence

$$((B^1)^T \mathbf{X})_N = \frac{1}{2\Delta x} h_{N-1/2} u_{N-1} - \frac{1}{2\Delta x} h_{N-1/2} u_N - 2\sqrt{3}\sigma_N$$

$$= \frac{1}{2\Delta x} \left(-h_{N-1/2}(u_N - u_{N-1}) \right) - 2\sqrt{3}\sigma_N.$$

Given that $u_N = u_{N+1} - 2\sqrt{g}(\sqrt{2H_{out} - h_N} - \sqrt{h_N})$, we obtain that

$$((B^1)^T \mathbf{X})_N = -\frac{1}{2} \left(h_{N+1/2} \frac{\frac{h_N}{h_{N+1}} u_{N+1} - u_N}{\Delta x} + h_{N-1/2} \frac{u_N - u_{N-1}}{\Delta x} \right)$$

$$- 2\sqrt{3}\sigma_1 - \frac{1}{\Delta x} \sqrt{g}(\sqrt{2H_{out} - h_N} - \sqrt{h_N}),$$

which is a first order discretization of $-(h\partial_x u + 2\sqrt{3}\sigma)$. The corresponding correction term is given by $\widetilde{0}_N = -\frac{\sqrt{g}(\sqrt{2H_{out} - h_N} - \sqrt{h_N})}{\Delta x}$.

$(\star\star)$ In this case, we impose the Dirichlet type boundary condition on the conservative variable hq:

$$(hq)_{i=N+1/2} = 0, \quad q_{bi=N+1/2} = 0 \implies q_{N+1} = -\frac{h_N}{h_{N+1}} q_N, \quad q_{bN+1} = q_{bN}.$$

With the boundary conditions at our disposal, we can write the discretization of the momentum equation for the first cell:

$$\frac{h_N u_N^{n+1} - h_N u_N^{n+1/2}}{\Delta t} + \frac{1}{2\Delta x}(h_{N+1}q_{N+1} - h_{N-1}q_{N-1}) + q_{bN}(\partial_x z_b)_N = 0.$$

Since $q_{N+1} = -\frac{h_N}{h_{N+1}} q_N$ this means that

$$(B^2 \mathbf{Q})_N = -\frac{1}{2\Delta x} h_{N-1} q_{N-1} - \frac{1}{2\Delta x} h_N q_N + q_{bN}(\partial_x z_b)_N,$$

and that no correction is required, $\widehat{0}_N = 0$.

So, given the matrix B^2, we can compute the approximation for the divergence

$$((B^2)^T \mathbf{X})_N = \frac{1}{2\Delta x} h_N u_{N-1} - \frac{1}{2\Delta x} h_N u_N - 2\sqrt{3}\sigma_N.$$

Given that $u_N = u_{N+1} - 2\sqrt{g}(\sqrt{2H_{out} - h_N} - \sqrt{h_N})$ we obtain that

$$((B^2)^T \mathbf{X})_N = -h_N \frac{u_{N+1} - u_{N-1}}{2\Delta x} + \frac{\sqrt{g}(\sqrt{2H_{out} - h_N} - \sqrt{h_N})}{\Delta x} - 2\sqrt{3}\sigma_1,$$

which is a first order discretization of $-(h\partial_x u + 2\sqrt{3}\sigma)$ up to the error term $\widetilde{0}_N = -\frac{\sqrt{g}(\sqrt{2H - h_N} - \sqrt{h_N})}{\Delta x}$.

3. Numerical results

To compare the numerical strategies, three cases are considered:

- An exact solution (solitary wave) where the three numerical methods are compared;
- A dispersive shock wave from a dam break simulated with the Uzawa algorithm;
- Comparisons with experimental data (Favre waves) and the numerical results obtained with the Uzawa algorithm.

In all the following simulations, the threshold used for the stopping criterion is set to $\epsilon = 10^{-3}$.

3.1. **Solitary wave solution**

It is a well known fact that the system admits solitary wave solutions in the case of constant atmospheric pressure and flat bottom topography.

Let us fix $p_{atm} \equiv cst$ and $z_b \equiv cst$. For $a > 0$ and $H_0 > 0$, we set $c = \sqrt{g(H_0 + a)}$ and

$$\mathcal{H}(\xi) = H_0 + a \operatorname{sech}^2\left(\frac{\sqrt{3a}}{2H_0\sqrt{H_0 + a}}\xi\right)$$

so that

$$\begin{cases} h_c(t, x) = \mathcal{H}(x - ct), \\ u_c(t, x) = c\left(1 - \frac{H_0}{\mathcal{H}(x - ct)}\right), \\ w_c(t, x) = -\frac{cH_0}{2}\frac{\mathcal{H}'(x - ct)}{\mathcal{H}(x - ct)}, \qquad\qquad \sigma_c(t, x) = -\frac{cH_0}{2\sqrt{3}}\frac{\mathcal{H}'(x - ct)}{\mathcal{H}(x - ct)}, \\ q_c(t, x) = \frac{c^2 H_0^2}{3\mathcal{H}^2(x - ct)}\left[\mathcal{H}(x - ct)\mathcal{H}''(x - ct) - \mathcal{H}'(x - ct)^2\right], \quad q_{bc}(t, x) = \frac{3}{2}q_c(t, x), \end{cases} \quad (22)$$

are solutions to System (6) in dimension 1.

For this test case, we consider the propagation of a solitary wave from the left to the right, initially centered at $x_0 = 10\ m$, of relative amplitude $a = 0.2\ m$, over a constant water depth $H_0 = 1\ m$. The computational domain is $100\ m$ long and discretized with 1280 cells.

Since the solitary wave is initially far from boundaries, the boundary conditions do not affect the computation, so we have chosen to impose free boundary conditions at the downstream and upstream ends for the sake of simplicity. We compare the water surface profile of our numerical solutions to the exact profile given by (22) at several times using two iterative methods, namely Uzawa (18) and Gauss-Seidel (21), and a direct resolution of (20).

Figure 2 shows that all schemes manage to capture the correct wave velocity but the Uzawa method seems to be less diffusive than other methods. The explanation may be related to a faster convergence of the iterative procedure with Uzawa than with Gauss-Seidel.

Table 1 shows the computational time and the relative error between the numerical solutions and the exact one given by (22):

$$E_{L^2}(h) = \frac{\|h_{num} - h_{ex}\|_2}{\|h_{ex}\|_2} \quad ; \quad E_{L^2}(u) = \frac{\|u_{num} - u_{ex}\|_2}{\|u_{ex}\|_2}.$$

To obtain these results, we set a time $T = 5\ s$ and compute the numerical solution by increasing the number of cells from $N = 80$ to $N = 1280$.
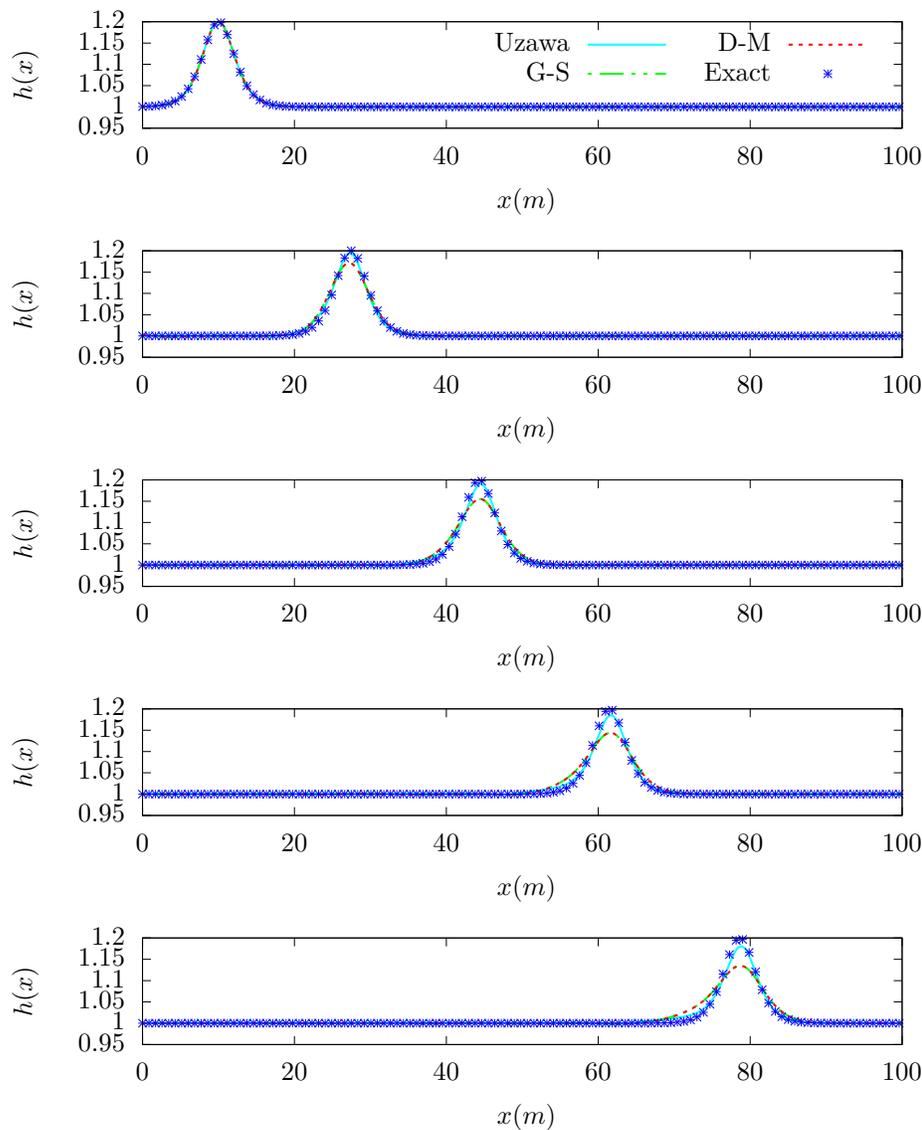
FIGURE 2. Propagation of solitary wave over a flat bottom: water surface profiles at $t = 5$, 10, 15 and 20 s. Comparison of the numerical and exact solutions for the Uzawa, Gauss-Seidel and a direct method

Remark 3.1. Note that the Gauss-Seidel method and the direct method were implemented using Python, whereas the Uzawa method was implemented using Fortran 95. This may explain the delay of computational times (CPU) between these methods.

From now on, all tests cases are run with the Uzawa method.

N	Uzawa method			Gauss-Seidel method		
	$E_{L^2}(h)$	$E_{L^2}(u)$	CPU	$E_{L^2}(h)$	$E_{L^2}(u)$	CPU
80	1.2×10^{-2}	4.3×10^{-1}	$0.018s$	2.19×10^{-2}	1.34×10^{-1}	$0.197s$
160	8.4×10^{-3}	2.8×10^{-1}	$0.075s$	1.7×10^{-2}	1.4×10^{-1}	$0.395s$
320	5.4×10^{-3}	1.8×10^{-1}	$0.318s$	1.2×10^{-2}	1.94×10^{-1}	$1.326s$
640	3.4×10^{-3}	1.1×10^{-1}	$1.230s$	7.78×10^{-3}	1.54×10^{-1}	$12.34s$
1280	2.1×10^{-3}	6.9×10^{-2}	$4.202s$	4.73×10^{-3}	1.58×10^{-1}	$157.25s$

TABLE 1. Relative L^2-error table for the variables h and u and and time execution

3.2. **Dam-break**

We consider the dam-break problem in order to study the ability of our numerical scheme to capture dispersives shocks. It is well known that discontinuous initial data of this type generate dispersive shock waves due to dispersive effects [6]. This issue for non-linear and dispersive shallow water equations was also investigated in [6, 13, 21]. We consider the following initial data:

$$u(0, x) = w(0, x) = \sigma(0, x) = 0, \qquad \eta(0, x) = \begin{cases} 1.8, & \text{if } x \leq 0, \\ 1, & \text{otherwise.} \end{cases}$$

The computational domain is the interval $(-300, 300)$ and discretized with 8000 cells.

Figure 3 shows the dispersive effects generated by System (1) when using discontinuous initial data. Here, we chose more cells (8000) because it is claimed in [7, 21] that dispersive effects are hard to capture. It is recommended that high-order numerical schemes be used but that is not the topic of this paper. A third-order method is applied in [16].

3.3. **Favre waves**

In the following section, we compare our numerical scheme based on the Uzawa algorithm with experimental data. A good example of a physical situation in which the non-hydrostatic pressure terms play a significant role is the so-called Favre wave experiment. Favre waves are secondary free surface waves, undular bores that appear typically in a channel following a sudden change in the flow discharge.

Such a situation occurs in the case when a sluice gate is suddenly closed, realised during the physical experiments carried out at EDF (Électricité de France) [12]. A long and thin wave tank is considered with a small constant slope and open uphill entry and downhill exit as well as vertical walls (see Figure 4(a)). A stationary water flow is established in the basin up until a time t_0 when a gate rapidly descends at the exit level, blocking the flow and generating an upstream moving series of undulating waves.

In the experiments, a wave tank of length 40 m was taken; the lower 30 m section plays a significant role. The tank is of width 0.4 m, the walls on the side are vertical and the bottom is of a constant slope of

$$\partial_x z_b = -0.0004.$$

At the lower end of the basin, a vertical solid gate is placed perpendicular above the water flow. Additionally, three sensors were also placed over the basin, each one measuring the water height at its horizontal position. The first sensor is situated 2 m from the exit of the tank, the second sensor is 15 m from the lower end, and finally the third sensor is at a distance of 30 m from the gate (see Figure 4(b)).
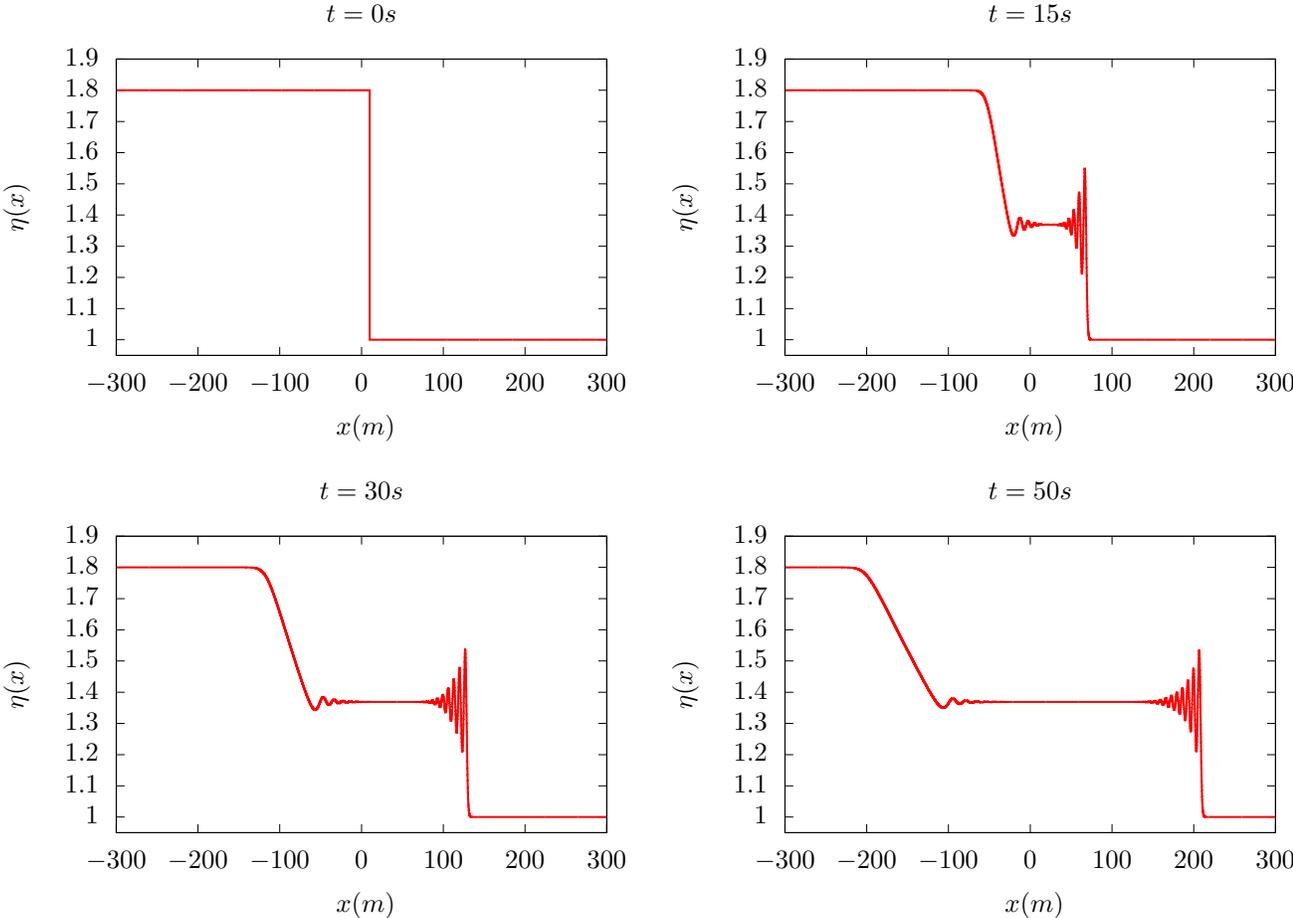
FIGURE 3. Dam-break: water surface at several times

In the experiment chosen for validation, a stationary gravity flow is generated in the tank by a constant inflow (and outflow) volumic flux of

$$Q_{in} = 0.0351.$$

The associated stationary water height is observed to be approximately $0.2\ m$.

As it was mentioned before, the gate is closed at time t_0 (when the flow has reached a stationary regime with the above given physical parameters). The closure takes place rapidly, in a time frame of $T = 0.07s$, when the gate completely blocks the exit, forming a solid wall boundary for the gravity flow. Measurements on the water height are available in the timeframe of $[t_0 + 1.31, t_0 + 5.35]$ for the first sensor, in the timeframe of $[t_0 + 10, t_0 + 16.9]$ for the second sensor, as well as in the timeframe of $[t_0 + 21, t_0 + 28.92]$ for the third sensor.

In the context of an extended Saint-Venant system, in which non-hydrostatic pressure terms enrich the classic shallow water equations, first and second order finite volume solvers have already been developed and validated against this particular experimental case [9].

In our case, the simulations were implemented with a one horizontal dimensional system corresponding to the experiments (see Figure 4(b)). The implemented physical parameters were the same, except for the the flux,

(A) The wave tank of the experiment

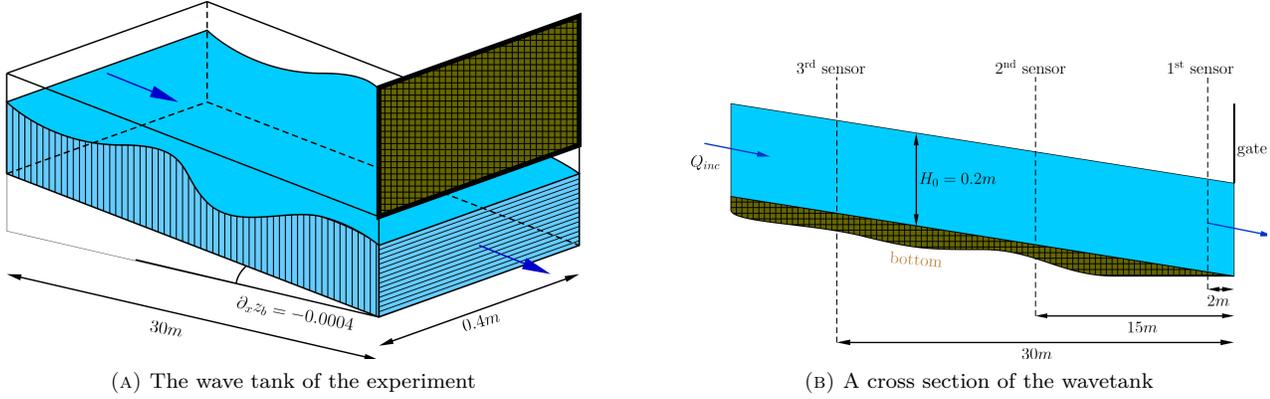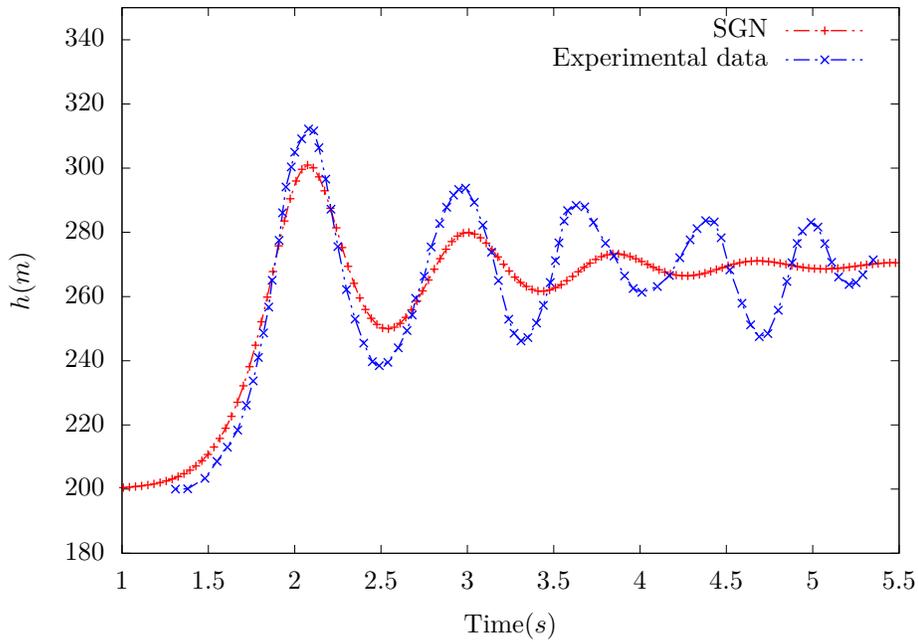(B) A cross section of the wavetank

FIGURE 4. Schematic representation of the Favre wave experiment

FIGURE 5. Comparison between the numerical solution (obtained with $N = 3000$ nodes) and the experimental data for Sensor 1 $(2\ m)$.

which had to be recomputed for its one dimensional equivalent:

$$Q_{inc} = \frac{1}{0.4} Q_{in} = 0.08775.$$

Close to the gate $(2\ m)$, Figure 5 shows that the first dispersive wave is correctly captured by the scheme while the following ones are damped. The standard shallow water equations only provide a shock wave with the correct velocity but nor the shape neither the other waves. As expected, the Serre–Green-Naghdi equations
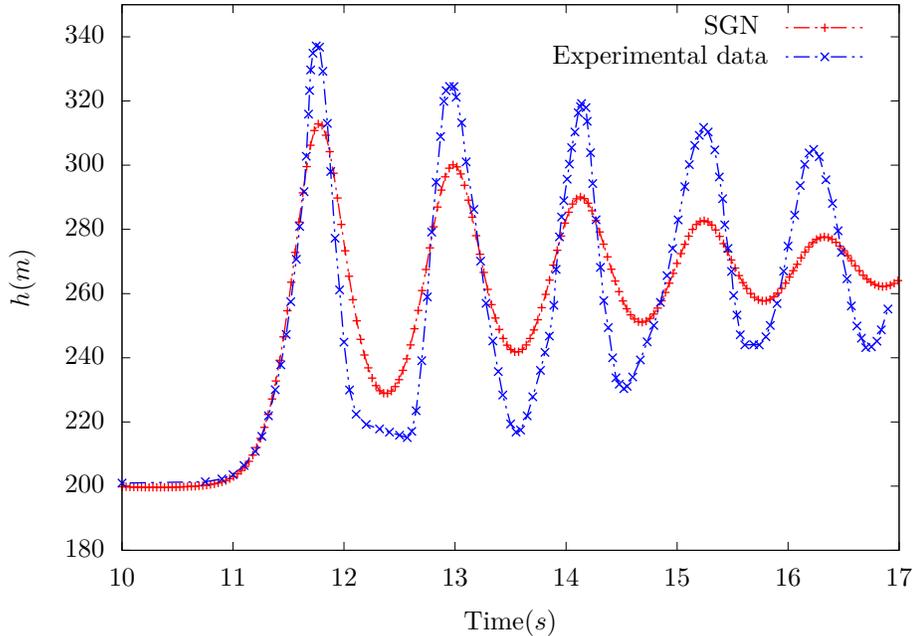
FIGURE 6. Comparison between the numerical solution (obtained with $N = 3000$ nodes) and the experimental data for Sensor 2 (15 m).

provide a solution with a more relevant structure. Further (15 m), the damping is still observed but the period seems to be preserved. Once again, a higher-order method should be considered.

CONCLUSION

The non-hydrostatic formulation of the well-known Serre–Green-Naghdi equations is investigated in this paper. Unlike its Boussinesq formulation, only first order derivatives are involved in the present one. The consequence is a larger number of unknowns. The physical content of this model is richer than the shallow water equations but the computational cost is higher: at each time step, in addition to the resolution of a shallow water system, an elliptic equation must be solved involving non-standard operators. The latter one is proved to be well-posed in any dimension $d \in \{1, 2\}$.

We designed some 1D numerical schemes based on the algebraic structure of the underlying linear system which mimics the properties of the continuous equation. They provide reliable results whether it be for analytical solutions or experimental solutions. It is a preliminary work which is extended in [16] where more numerical tests are carried out with a third-order method; in addition an extension of the numerical strategy to the multilayer extension of the Serre–Green-Naghdi model [17] will be proposed.

References

[1] Nora Aïssiouene, Marie-Odile Bristeau, Edwige Godlewski, and Jacques Sainte-Marie. A combined finite volume - finite element scheme for a dispersive shallow water system. *Networks and Heterogeneous Media*, 11(1):1–27, 2016.

[2] Nora Aïssiouene, Marie-Odile Bristeau, Edwige Godlewski, and Jacques Sainte-Marie. A robust and stable numerical scheme for a depth-averaged Euler system. hal-01162109, submitted.

[3] Emmanuel Audusse, François Bouchut, Marie-Odile Bristeau, Rupert Klein, and Benoît Perthame. A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. *SIAM J. Sci. Comput.*, 25(6):2050–2065, 2004.

[4] Jerry L. Bona, Thomas B. Benjamin, and John J. Mahony. Model equations for long waves in nonlinear dispersive systems. *Philos. Trans. Royal Soc. London Series A*, 272:47–78, 1972.

[5] Jerry L. Bona, Min Chen, and Jean-Claude Saut. Boussinesq equations and other systems for small-amplitude long waves in nonlinear dispersive media. I: Derivation and linear theory. *Journal of Nonlinear Sciences*, 12(4):283–318, 2002.

[6] Philippe Bonneton, Florent Chazel, David Lannes, Fabien Marche, and Marion Tissier. A splitting approach for the fully nonlinear and weakly dispersive Green–Naghdi model. *J. Comput. Phys.*, 230(4):1479–1498, 2011.

[7] Christian Bourdarias, Stéphane Gerbi, and Ralph Lteif. A numerical scheme for an improved Green–Naghdi model in the Camassa–Holm regime for the propagation of internal waves. *Computers & Fluids*, 156:283–304, 2017.

[8] Joseph V. Boussinesq. Théorie des ondes et des remous qui se propagent le long d'un canal rectangulaire horizontal, en communiquant au liquide contenu dans ce canal des vitesses sensiblement pareilles de la surface au fond. *J. Math. Pures Appl.*, pages 55–108, 1872.

[9] Marie-Odile Bristeau, Nicole Goutal, and Jacques Sainte-Marie. Numerical simulations of a non-hydrostatic shallow water model. *Computers & fluids*, 47(1):51–64, 2011.

[10] Marie-Odile Bristeau, Anne Mangeney, Jacques Sainte-Marie, and Nicolas Seguin. An energy-consistent depth-averaged Euler system: derivation and properties. *Discrete and Continuous Dynamical Systems - Series B*, 20(4):961–988, 2015.

[11] Roberto Camassa, Darryl D. Holm, and C. David Levermore. Long-time effects of bottom topography in shallow water. *Phys. D: Nonlinear phenomena in ocean dynamics*, 98(2-4):258–286, 1996.

[12] L Caudron. Contribution à l'étude des ondes de favre. Technical report, Technical Report HC. 032-E230, Laboratoire National d'Hydraulique . . . , 1968.

[13] Florent Chazel, David Lannes, and Fabien Marche. Numerical simulation of strongly nonlinear and dispersive waves using a green–naghdi model. *Journal of Scientific Computing*, 48(1-3):105–116, 2011.

[14] Adhémar Barré de Saint-Venant. Théorie du mouvement non permanent des eaux, avec application aux crues des rivières et à l'introduction des marées dans leurs lits. *C. R. Acad. Sci.*, 73:237–240, 1871.

[15] Mohamed Ali Debyaoui and Mehmet Ersoy. Generalised serre-green-naghdi equations for open channel and for natural river hydraulics. 2020.

[16] Cipriano Escalante Sanchez, Enrique D. Fernandez Nieto, Tomas Morales de Luna, Yohan Penel, and Jacques Sainte-Marie. Numerical simulations of Serre – Green-Naghdi type models for dispersive free surface flows. 2019.

[17] Enrique D. Fernández Nieto, Martin Parisot, Yohan Penel, and Jacques Sainte-Marie. A hierarchy of dispersive layer-averaged approximations of Euler equations for free surface flows. *Communications in Mathematical Sciences*, 16(05):1169–1202, 2018.

[18] Albert E. Green and Paul M. Naghdi. A derivation of equations for wave propagation in water of variable depth. *J. Fluid Mech.*, 78(02):237–246, 1976.

[19] Ulrich Langer and Werner Queck. On the convergence factor of Uzawa's algorithm. *Journal of Computational and Applied Mathematics*, 15:191–202, 1986.

[20] David Lannes and Fabien Marche. A new class of fully nonlinear and weakly dispersive Green-Naghdi models for efficient 2D simulations. *Journal of Computational Physics*, 282:238–268, 2015.

[21] Olivier Le Métayer, Sergey Gavrilyuk, and Sarah Hank. A numerical scheme for the Green–Naghdi model. *Journal of Computational Physics*, 229(6):2034–2045, 2010.

[22] Maojun Li, Philippe Guyenne, Fengyan Li, and Liwei Xu. High order well-balanced CDG-FE methods for shallow water waves by a Green-Naghdi model. *Journal of Comutational Physics*, 257:169–192, 2014.

[23] Per A. Madsen, Richard Murray, and Ole R. Sørensen. A new form of the Boussinesq equations with improved linear dispersion characteristics. *Coastal Eng.*, 15(4):371–388, 1991.

[24] Okey Nwogu. Alternative form of Boussinesq equations for nearshore wave propagation. *J. Waterway, Port, Coastal, Ocean Eng.*, 119(6):618–638, 1993.

[25] Martin Parisot. Entropy-satisfying scheme for a hierarchy of dispersive reduced models of free surface flow. *International Journal for Numerical Methods in Fluids*, 91(10):509–531, 2019.

[26] D. Howell Peregrine. Long waves on a beach. *J. Fluid Mech.*, 27(04):815–827, 1967.

[27] Aleksandr A. Samarskii and Evgenii S. Nikolaev. *Numerical methods for grid equations II: Iterative methods*. Birkhäuser, 1989. Translated from Russian by Stephen G. Nash.

[28] François Serre. Contribution à l'étude des écoulements permanents et variables dans les canaux. *La Houille Blanche*, (6):830–872, 1953.