An explicit asymptotic preserving low Froude scheme for the multilayer shallow water model with density stratification

François Couderc, A Duran, Jean-Paul Vila

An explicit asymptotic preserving low Froude scheme for the multilayer shallow water model with density stratification

F. Couderc[a], A. Duran[a], J.-P. Vila[a,*]

[a]*Institut de Mathématiques de Toulouse; UMR5219, Université de Toulouse; CNRS, INSA, F-31077 Toulouse, France.*

Abstract

We present an explicit scheme for a two-dimensional multilayer shallow water model with density stratification, for general meshes and collocated variables. The proposed strategy is based on a regularized model where the transport velocity in the advective fluxes is shifted proportionally to the pressure potential gradient. Using a similar strategy for the potential forces, we show the stability of the method, in the sense of a discrete dissipation of the mechanical energy, in the general multilayer and non-linear frame. Based on a linear analysis, and with the objective of minimizing the diffusive losses in realistic contexts, sufficient conditions are exhibited on the regularizing terms to ensure linear stability. These results are subsequently validated by numerical investigations. The other main result stands in the consistency with respect to the asymptotics reached at small and large time scales in low-Froude regimes, which governs large scale oceanic circulation. Additionally, robustness and well balanced results for motionless steady states are also ensured. These stability properties tend to provide a very robust and efficient approach, easy to implement and particularly well suited for large scale simulations. Two numerical experiments are proposed to highlight the scheme efficiency: a first experiment of fast gravitational modes and a second of slow Rossby modes simulating the displacement of a baroclinic vortex subject to the Coriolis force.

Keywords: multilayer shallow water, asymptotic preserving scheme, non linear stability, energy dissipation.

1. Introduction

Mathematical modelling of geophysical phenomena involves three-dimensional and turbulent free surface flows with complex geometries. The capability to perform such simulations in a direct way is still unachievable for supercomputers, even with a maximized parallel computing efficiency, especially since most simulation platforms devoted to prevision need to be oriented toward real time resolutions. Based on this assessment, many models have been derived during the past decades to reduce the complexity of the original primitive equations by integrating/averaging the vertical dimension, notably under the classical *hydrostatic* and *shallow water* assumptions. Moreover, the density stratification, mainly related to the temperature and salinity variations, can profoundly affect the water flow dynamics and must be taken into account in many cases such for example oceanic or estuary flows. Taking these aspects under consideration, the inviscid multilayer shallow water model, which involves an arbitrary number of superposed immiscible layers, offers a simple way to integrate the vertical density distribution with a satisfactory time computation request.

The model presented in this work corresponds to a vertical discretization of the primitive equations based on a piecewize constant density distribution along the layer interfaces, as detailed in [40], and shown in Fig. 1. Nevertheless, it has to be mentioned that the advection model on which is based our numerical approach involves a more general formulation of potential forces, allowing to extend the applicability range of the resulting scheme to other physical contexts, possibly decorrelated from large scale oceanic circulation. Let us mention the issue associated with the simple case of a single layer, with specific one or two dimensional applications to hydraulic or coastal engineering for instance, notably with the support of natural positivity and well-balanced properties. Of course, many other physical contexts sharing the same general structure can be considered, such as the Euler equations for gas dynamics.

*Corresponding author

Email addresses: couderc@math.univ-toulouse.fr (F. Couderc), aduran@math.univ-toulouse.fr (A. Duran), vila@insa-toulouse.fr (J.-P. Vila)

From a numerical point of view, a large range of approaches devoted to the single layer case are available in the literature, with the handling of complex geometries and rugged topography using unstructured environments, robust treatment of friction forces with wetting and drying, and allowing high order resolutions. However, the quantity of advances concerning the multilayer system is far less plentiful. This may be due to the complex nature of the model, that brings a greater difficulty level. Indeed, in addition of non linearities, it is a known fact that the multilayer shallow water equations exhibit particular structural properties, making the system theoretically and numerically more demanding than in the single layer case. The first point concerns hyperbolicity, which can be violated if the velocity shear between two layers is too high in relation to the respective heights and densities, possibly leading to Kelvin-Helmotz instabilities. Preventing the apparition of complex eigenvalues is a quite complicated task, and this possible local hyberbolicity loss can significantly reduce the application range of the numerical schemes. These stability conditions are rigorously characterized in [29], where a general criterion of hyperbolicity and local well-posedness is given, under a particular asymptotic regime and weak stratification assumptions of the densities and the velocities. A similar study has been realized in [19] in the limit of small density contrast. It is shown that, under reasonable conditions on the flow, the system is well-posed on a large time interval. The second difficulty arises from the potential forces resulting from the hydrostatic pressure assumption, introducing a non conservative coupling between the layers.

As a consequence, classical finite volume methods based on approximate Riemann solvers cannot be directly applied to the multilayer shallow water equations. Nevertheless, when the number of layers is restricted to two, several techniques have been proposed on the basis of classical non-linear stability criteria, generally borrowed from the advances made on the single layer system. Thus, as concerns the two layers approximations, one can note for instance the *Q-scheme* proposed in [13], the recent relaxation approach [4] able to guarantee the preservation of motionless steady states, or the so called central-upwind scheme in [23]. Other splitting and upwind schemes can be found, with for instance [18] (see also its extension to three layers proposed in [14] with a study of the hyperbolicity range), the *f-wave* propagation finite volume method in [28] handling dry states or the *well-balancing* and positivity-preserving results established in [8] within a splitting approach. That being so, and although a first relevant approximation for ocean modelling may be provided by a bi-fluid stratification, the number of layers involved in most of current oceanic flow simulations with modern operational softwares is much more important in practise, in the order of several tens. Unfortunately, it turns out that exporting the single layer approaches to the multilayer general case is quite difficult to achieve, and most of them are not specially conceived to preserve the asymptotics observed in low Froude number regimes, a specificity of oceanic flows where the velocities magnitude are very moderate compared to the gravity wave speed far from the coast. Considering the integration time of realistic simulations, this limitation is also due to the paramount importance of the dissipation of the mechanical energy, which has to be guaranteed in all situations in order to produce physically acceptable solutions. At last, numerical methods for one-dimensional multilayer shallow water models with mass exchange are also proposed without density stratification in [5] and with in [6]. The approach is quite different since the layer depths are not independent variables and only the free surface is treated, and also because a part of the coupling terms are treated as a source term.

Adapting the choices made to express the distribution of the pressure law, which is also generally formulated, in some sense, by mean of staggered discretizations of the vertical direction, the multilayer equations formulated in this work are closely connected to those used in the majority of operational oceanic simulation softwares like HYCOM [9], ROMS [35] or NEMO [27], in isopycnal coordinates. These codes have been developed on staggered grids, sharing an Arakawa C-grid type as a general basis with orthogonal curvilinear coordinates to take into account irregular lateral boundaries. This kind of horizontal space discretization allows to prevent from well known spurious computational modes observed in low Froude number regimes. The barotropic and baroclinic modes are resolved with a time splitting technique allowing to use different time steps, as the barotropic wave speed is much higher than the larger baroclinic one, and this allows to save time computation. The barotropic continuity equation is often resolved with a FCT (Flux Corrected Transport) scheme and the momentum equations discretized with centered schemes of order two or four. As concerns time integration, Leapfrog-type schemes are usually employed, coupled with stabilization procedures using a Robert-Asselin filter in order to minimize the dissipation. A detailed report outlining the stability aspects related to oceanic modelling is available in [24]. If this kind of staggered numerical approaches has been largely successfully applied, it can exhibit some weaknesses for some practical implications. The global stability of the numerical methods is not always guaranteed, with one special consequence of a tricky wetting and drying treatment and the difficulty to handle boundary conditions.

The constant willingness to increase the quality and the versatility of numerical resolutions brought

progressively many interests for unstructured geometries during the past decade. The use of such environments may appear of major interest for many practical applications, and notably for oceanic circulation, for which geometrical flexibility allows to describe complex shaped shoreline coastlines and many different scales. Thus, other projects started to make the choice of unstructured meshes, despite numerical and implementation issues that have not been overcomed yet, and still subject of ongoing research. In this connexion, a quite complete review of the most recent results oriented toward ocean modelling can be found in [16]. The SLIM [3] and FVCOM [2] projects can be cited as examples. Among the available works, mention can be made of [34] with the study of Finite Element methods stability applied to the rotating shallow water equations. It is concluded that all the numerical schemes considered are, at some point, concerned with spurious solutions. Some pioneering works devoted to the derivation of numerical schemes for the single layer rotating shallow water equations using unstructured meshes can be cited [7], [15], [21], [33], [36], [38] and [39].

The numerical approach proposed in the present work is a collocated Finite Volume like method applicable for general meshes. Let us recall that this kind of method allows to naturally ensure conservation properties like the water volume conservation. Based on the numerical constraints discussed above, several guidelines are to be followed, principally based on two particular stability criterions. The first one is related to the dissipation of the mechanical energy. This point appears as fully relevant in the frame of geophysical flows, since an inappropriate discretization of the system may bring non physical terms in the energy production and break the stability of the system in large times. Such considerations of physically admissible solutions are studied in the numerical approach [11] for the one-dimensional model, where a semi-discrete entropy inequality is established in addition to the well-balancing property, treating the non-conservative coupling part as a source term. A stronger result is obtained in the two layers case with a fully discrete version [10]. The second essential point, that has so far not been rigorously addressed in the general multilayer case, concerns the capability to capture the low Froude asymptotics. In these regimes, most Godunov-type schemes bring too much dissipation and totally miss the problem physics, and some consistency results are unavoidable to describe properly the flow dynamics. As stated in [17] in the context of Euler equations, these asymptotic behaviours are principally governed by the gradient pressure treatment, for which centred approaches should be favoured. Regarding the available numerical strategies employed to control the discrete mechanical energy, some techniques can be found in [22], in the context of a compressible multifluid model. Inspired from the ideas of the AUSM methods for gas dynamics, see [26] and[25], the formalism implies a modified velocity transport, shifted proportionally to the potential gradient, whose goal is to provide a control on the energy budget at the continuous level. On this basis, a simple and efficient Finite Volume like scheme is derived, designed to provide a discrete mimetism of this result. More recently, a general extension has been proposed with the semi-implicit CPR scheme for the two-dimensional multilayer shallow water model with density stratification [30, 31]. Note that in addition, the mentioned approaches have the common feature of being asymptotic-preserving with respect to low Froude number regimes, notably thanks to a centred discretization of the potential gradient, as discussed above. Following these lines, we propose an extension of the semi-implicit scheme introduced in [31], formulating a totally explicit version. In this environement, the use of a shifted velocity transport is not sufficient to ensure a mechanical energy control, and a correction term is also needed on the potential forces. It may also be shown that this adjustment, expressed in terms of discharge divergence, has also regularizing virtues on the energy budget at the continuous level. From a practical point of view, the advantages of an explicit formulation stand in the exemption of resolving the non-linear system arising from for the continuity equation, an easier implementation of boundary conditions and high order extensions in space and time can be more easily derived. The price to pay is the time step restriction, subject to a classical explicit CFL condition based on the barotropic gravitational wave speed. It is however expected that this drawback should be compensated by the computational savings induced by a rapid resolution of the continuity equation and the performances of fully explicit high order resolutions.

The outline of this paper is organized as follows. In §2, we first recall the multilayer shallow water equations with density stratification numerically resolved in this work. We subsequently propose a regularization of the model that allows a better control of the mechanical energy production. We finally give the formulation of the explicit scheme, designed to provide a discrete mimetism to this formalism, i.e. that allows the decrease of the mechanical energy. The §3 is devoted to stability issues. Well balanced and robustness properties are addressed first. We then show a control on the mechanical energy production, and put it in correlation with our investigations in the linear case. Asymptotic preserving properties are established in a semi-continuous context in §4. A last step of numerical validation is finally proposed to assess the scheme abilities for large scale simulations, with a first experiment of fast gravity modes and

a second one implying slow Rossby modes.

2. Preliminaries

2.1. Physical model

The present work describes a numerical strategy devoted to approximate the solutions of the two-dimensional multilayer shallow water system with a density stratification. Denoting L the number of layers involved in the description of the flow, t and $\mathbf{x} = (x, y)$ the time and space variables, the dynamics is governed by a general conservation law which consists of a set of $3 \times L$ equations linking the mass in each layer $H_i(t, \mathbf{x}) \geq 0$ to the velocity $\mathbf{u}_i(t, \mathbf{x})$. The system is submitted to gravitational forces through the scalar potential $\Phi_i(\boldsymbol{H}, \mathbf{x})$, where $\boldsymbol{H} = {}^t(H_1, \cdots, H_L)$:

$$\left\{ \begin{array}{lcll} \partial_t H_i & + & div\,(H_i \mathbf{u}_i) & = & 0 \\ \partial_t (H_i \mathbf{u}_i) & + & div\,(H_i \mathbf{u}_i \otimes \mathbf{u}_i) & = & -H_i \nabla \Phi_i / \varepsilon^2 \end{array} \right. . \qquad (M_{t,\varepsilon})$$

In the above equations the parameter ε is introduced to account for the scale factor between inertial and potential forces. The potential and kinetic energies attached to the general system are defined by $\partial_{H_i} \mathcal{E} = \Phi_i$ and $\mathcal{K}_i = \dfrac{1}{2} H_i \left\| \mathbf{u}_i \right\|^2$. We recall the conservation law satisfied by the mechanical energy $E = \mathcal{E}/\varepsilon^2 + \displaystyle\sum_{i=1}^{L} \mathcal{K}_i$ for regular solutions:

$$\partial_t E + \sum_{i=1}^{L} div\Big(\left(H_i \Phi_i / \varepsilon^2 + \mathcal{K}_i \right) \, \mathbf{u}_i \Big) = 0 \, . \qquad (1)$$

In the spirit of the semi-implicit schemes proposed in [22, 30, 31], the stability of the numerical approach is based on advective fluxes where the transport velocity is shifted proportionally to the potential gradient. To get a better picture of the formalism, it's worthwhile to recall that the strategy can be interpreted at the continuous level as a discrete form of the following regularized model:

$$\left\{ \begin{array}{lcll} \partial_t H_i & + & div\,(H_i\,(\mathbf{u}_i - \delta \mathbf{u}_i)) & = & 0 \\ \partial_t (H_i \mathbf{u}_i) & + & div\,(H_i \mathbf{u}_i \otimes (\mathbf{u}_i - \delta \mathbf{u}_i)) & = & -H_i \nabla \Phi_i / \varepsilon^2 \end{array} \right. , \qquad (M^r_{t,\varepsilon})$$

$\delta \mathbf{u}_i$ standing for a generic perturbation on the velocity. This modification has the following impact on the energy conservation (1):

$$\partial_t E + \sum_{i=1}^{L} div\Big(\left(H_i \Phi_i / \varepsilon^2 + \mathcal{K}_i \right) \, (\mathbf{u}_i - \delta \mathbf{u}_i) \Big) = -\sum_{i=1}^{L} \delta \mathbf{u}_i . \nabla \Phi_i / \varepsilon^2 \, , \qquad (2)$$

which formally justifies a calibration of $\delta \mathbf{u}_i$ in terms of the potential forces gradient, to ensure a global decrease of the mechanical energy. In this work, we aim at proposing a discrete equivalent of (2), in a fully explicit context. From a general point of view, it is important to recall here that the conservation law $(M_{t,\varepsilon})$ enjoys a large range of applicability and the present approach is not only limited to large scale oceanic circulation. Indeed many other physical systems fall within the present formalism, provided several regularity hypothesis on the potential forces:

2.1. *Regularity assumptions on the potential forces*

- *The potential $\mathcal{E}$ is a regular and convex function of the mass, which means that the Hessian $\boldsymbol{\mathcal{H}}$ given by $\boldsymbol{\mathcal{H}}_{ij} = \partial^2_{H_i H_j} \mathcal{E} = \partial_{H_j} \Phi_i$ is positive-definite.*
- *The potential is a symmetric and linear function of the mass, that is $\Phi = \boldsymbol{\mathcal{H}}.\boldsymbol{H}$ and $\boldsymbol{\mathcal{H}}$ symmetric.*
- *also assume that the L^2 norm of $\boldsymbol{\mathcal{H}}$ is uniformly bounded with respect to space and time, more precisely:*

$$|||\boldsymbol{\mathcal{H}}(\boldsymbol{H}, \mathbf{x})|||_{L^2} \leq C_{\boldsymbol{\mathcal{H}}} \, . \qquad (3)$$

In the case of the multilayer shallow water system, and assuming a constant density ρ_i for each layer i, the effective mass corresponds to $H_i = \rho_i h_i$, h_i standing for the layer thickness. Then, denoting by

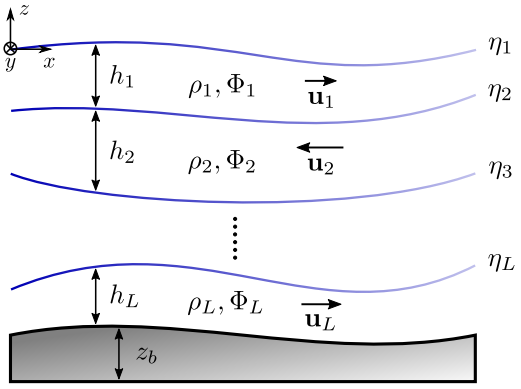

Figure 1: Multilayer shallow water model with density stratification sketch ($\eta_i = \sum_{k=i}^{L} h_k$). All the model variables are collocated along the z coordinate.

z_b the bottom topography, the potential pressure is given by $\Phi_i = g\left(z_b + \sum_{j=1}^{L} \frac{\rho_j}{\rho_{\max(i,j)}} h_j\right)$ (see Fig. 1). The Hessian $\boldsymbol{\mathcal{H}}(\boldsymbol{H}, \mathbf{x})$ associated with the system is thus constant in space and time:

$$\boldsymbol{\mathcal{H}}_{i,j} = g/\rho_{max(i,j)}\,,$$

and the requirements listed in Hypothesis 2.1 are trivially satisfied. Note also that this independence automatically brings the conservation of the total momentum, as shown in [31]. However, this is not sufficient to guarantee the well-posedness of the problem: some conditions can be found in [29], regarding $\boldsymbol{\mathcal{H}}$ as a natural symmetrizer of the system. These conditions are based on smallness assumptions on the velocity shear and are sufficient to ensure that the system is hyperbolic.

2.2. Notations, numerical approach

We consider in this work a tesselation $\mathbb{T}$ of the computational domain $\Omega \subset \mathbb{R}^2$. We will denote m_K the area and $m_{\partial K}$ the perimeter of a cell $K \in \mathbb{T}$. The boundary of K will be denoted ∂K , and for any edge $e \in \partial K$, m_e the length of the corresponding boundary interface and $\mathbf{n}_{e,K}$ the outward normal to e pointing to the neighbour K_e (see Fig. 2).

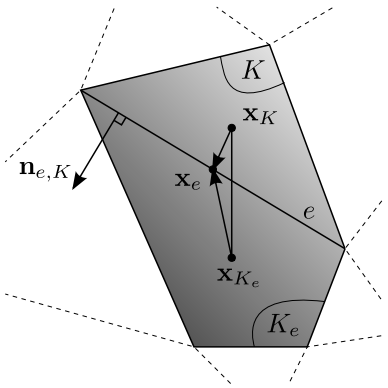

Figure 2: Geometric settings - Focus on the interface $e \in \partial K \cap \partial K_e$.

Let's now introduce some useful notations. For a scalar piecewize constant function w we define:

$$\overline{w}_e = \frac{1}{2}\left(w_{K_e} + w_K\right) \quad , \quad \boldsymbol{\delta} w_e = \frac{1}{2}\left(w_{K_e} - w_K\right)\mathbf{n}_{e,K}\,,$$

and similary, for a piecewize constant vectorial function $\mathbf{w}$:

$$\overline{\mathbf{w}}_e = \frac{1}{2}\left(\mathbf{w}_{K_e} + \mathbf{w}_K\right) \quad , \quad \delta \mathbf{w}_e = \frac{1}{2}\left(\mathbf{w}_{K_e} - \mathbf{w}_K\right).\mathbf{n}_{e,K}\,.$$

We also set: $w^{\pm} = \dfrac{1}{2}(w \pm |w|)$ the positive and negative parts of a scalar function w.
The numerical scheme we consider is the following:

$$\left\{ \begin{array}{rcll} H_{K,i}^{n+1} & = & H_{K,i}^{n} & - \dfrac{\Delta t}{m_K} \displaystyle\sum_{e \in \partial K} \mathcal{F}_{e,i}^{n} . \mathbf{n}_{e,K} m_e \\ H_{K,i}^{n+1} \mathbf{u}_{K,i}^{n+1} & = & H_{K,i}^{n} \mathbf{u}_{K,i}^{n} & - \dfrac{\Delta t}{m_K} H_{K,i}^{n} \displaystyle\sum_{e \in \partial K} \dfrac{\Phi_{e,i}^{n,*}}{\varepsilon^2} \mathbf{n}_{e,K} m_e \\ & & & - \dfrac{\Delta t}{m_K} \displaystyle\sum_{e \in \partial K} \Big(\mathbf{u}_{K,i}^{n} \big(\mathcal{F}_{e,i}^{n} . \mathbf{n}_{e,K} \big)^{+} + \mathbf{u}_{K_e,i}^{n} \big(\mathcal{F}_{e,i}^{n} . \mathbf{n}_{e,K} \big)^{-} \Big) m_e \end{array} \right. , \tag{4}$$

where we have set:

$$\Phi_{e,i}^{n,*} = \overline{\Phi}_{e,i}^{n} - \Lambda_{e,i}^{n} = \left(\frac{\Phi_{K,i}^{n} + \Phi_{K_e,i}^{n}}{2} \right) - \Lambda_{e,i}^{n} \, , \tag{5}$$

$$\mathcal{F}_{e,i}^{n} = \overline{H \mathbf{u}}_{e,i}^{n} - \Pi_{e,i}^{n} = \left(\frac{H_{K,i}^{n} \mathbf{u}_{K,i}^{n} + H_{K_e,i}^{n} \mathbf{u}_{K_e,i}^{n}}{2} \right) - \Pi_{e,i}^{n} \, . \tag{6}$$

The quantities $\Lambda_{e,i}^{n}$ and $\Pi_{e,i}^{n}$ introduced above stand for the perturbations respectively assigned to the potential forces and numerical fluxes, designed to ensure the stability of the method. They are defined as follows:

$$\Lambda_{e,i}^{n} = \alpha \Delta t C_{\boldsymbol{\mathcal{H}}} \mu_e \delta (H \mathbf{u})_{e,i}^{n} \quad , \quad \alpha > 0 \, , \tag{7}$$

$$\Pi_{e,i}^{n} = \gamma \Delta t (\overline{H \mu})_{e,i}^{n+1/2} \frac{\boldsymbol{\delta} \Phi_{e,i}^{n}}{\varepsilon^2} \quad , \quad \gamma > 0 \, , \tag{8}$$

with the geometric constant:

$$\mu_e = \frac{1}{2} \left(\mu_K + \mu_{K_e} \right) = \frac{1}{2} \left(\frac{m_{\partial K}}{m_K} + \frac{m_{\partial K_e}}{m_{K_e}} \right) ,$$

and the weighted average:

$$(\overline{H \mu})_{e,i}^{n+1/2} = \frac{1}{2} \left((H \mu)_{K,i}^{n+1/2} + (H \mu)_{K_e,i}^{n+1/2} \right) = \frac{1}{2} \left(\frac{(H_{K,i}^{n})^2}{H_{K,i}^{n+1}} \frac{m_{\partial K}}{m_K} + \frac{(H_{K_e,i}^{n})^2}{H_{K_e,i}^{n+1}} \frac{m_{\partial K_e}}{m_{K_e}} \right) . \tag{9}$$

We refer to Theorem (3.1) and subsequent Remarks (3.4,3.5) for the calibration of the stabilization constants α and γ. Let us finally remark that the numerical scheme satisfied by the velocity is:

$$\mathbf{u}_{K,i}^{n+1} = \mathbf{u}_{K,i}^{n} - \frac{\Delta t}{m_K} \sum_{e \in \partial K} \frac{\mathbf{u}_{K_e,i}^{n} - \mathbf{u}_{K,i}^{n}}{H_{K,i}^{n+1}} \big(\mathcal{F}_{e,i}^{n} . \mathbf{n}_{e,K} \big)^{-} m_e - \frac{\Delta t}{m_K} \frac{H_{K,i}^{n}}{H_{K,i}^{n+1}} \sum_{e \in \partial K} \frac{\Phi_{e,i}^{n,*}}{\varepsilon^2} \mathbf{n}_{e,K} m_e \, , \tag{10}$$

and note that:

$$\sum_{e \in \partial K} \Phi_{e,i}^{n,*} \mathbf{n}_{e,K} m_e = \sum_{e \in \partial K} \boldsymbol{\delta} \Phi_{e,i}^{n} m_e - \sum_{e \in \partial K} \Lambda_{e,i}^{n} \mathbf{n}_{e,K} m_e \, , \tag{11}$$

since the main term of (5) involves a centred discretization of the potential. We finally recall the explicit CFL condition on which are usually based Godunov-type schemes:

$$\left(|\overline{\mathbf{u}}_{e,i}^{n} . \mathbf{n}_{e,K}| + \frac{c_{e,i}^{n}}{\varepsilon} \right) \Delta t \max \left(\frac{m_e}{m_K}, \frac{m_e}{m_{K_e}} \right) \leq \tau_{CFL} \, , \tag{12}$$

where $c_{e,i}^{n}$ stands for an estimation of the potential wave celerity at the interface e for the $i-th$ layer.

3. Stability issues

In this section we focus on crucial non-linear stability criterion that are the preservation of motionless steady states, preservation of the water height positivity, and mechanical energy dissipation. These are

essential points in the description of realistic situations and hence they need absolutely to be integrated in the construction of numerical schemes expected to respond to applicative issues. Traditionally, providing a numerical approach able to account simultaneously for all these aspects remains a quite complicated task, especially in the context of general geometries and stratified multiscale models. Nevertheless, the formalism employed here allows a quite simple treatment of well balancing and robustness properties. This allows to channel our efforts on the control of the energy production, which may also seem quite natural, since the strategy can be interpreted as a discrete derivation of the stability results previously exhibited at the continuous level.

3.1. Well Balancing

As a first stability criterion we study the problem of steady states preservation. From a general point of view, regarding the difficulty to explicitly describe and handle numerically the full set of steady states observed in most of realistic evolution processes, it is classical to consider the case of rest states only, that is implying zero velocity. In our formalism, this leads to the following equilibriums for $(M_{t,\varepsilon})$:

$$\mathbf{u}_{K,i} = 0 \quad , \quad \Phi_{K,i} = \Phi_i \, ,$$

for all volume control K and layer i. In the present context, the interest for such configurations is not only due to their simple derivation and formulation, to the extent that they appear as fundamental for many practical applications. The previous configuration is nothing but the generalization to the multilayer case of the classical *lake at rest* solution in the $L = 1$ case:

$$\mathbf{u} = 0 \quad , \quad h + z_b = 0 \, ,$$

which has indeed to be exactly preserved to avoid the apparition of non physical perturbations in the vicinity of flat free surface configurations. That being so, the capability to preserve these particular steady states already stands for a discriminating property, even in the one layer case, notably with the increasing interest of unstructured meshes and high order resolutions in the modern approaches. In spite of these difficulties, the proposed discretization allows their exact preservation in a very simple way, even at high order in space, and without the need of any correction term. The following approach is thus intrinsically adapted to the preservation of such equilibriums, allowing to avoid the traditional reconstructions that usually tend to alterate the scheme's accuracy.

3.1. *Well Balancing*
The scheme (4) equipped with the numerical fluxes (6) and discrete potential (5) preserves the steady states at rest defined by $\boldsymbol{u}^n_{K,i} = 0$ *and* $\Phi^n_{K,i} = \Phi_i$.

Proof. Since the perturbation $\Pi^n_{e,i}$ (8) is expressed in terms of $\boldsymbol{\delta}\Phi^n_{e,i}$, we immediately have $\mathcal{F}^n_{e,i} = 0$ and $H^{n+1}_{K,i} = H^n_{K,i}$. Then, since $\Lambda^n_{e,i} = 0$, the momentum equation reduces to:

$$H^{n+1}_{K,i}\mathbf{u}^{n+1}_{K,i} = -\frac{\Delta t}{m_K} H^n_{K,i} \sum_{e\in\partial K} \frac{\Phi_i}{\varepsilon^2} \mathbf{n}_{e,K} m_e = 0 \, , \tag{13}$$

which allows to conclude. □

3.2. Robustness

We investigate here the problem of robustness by proposing a CFL condition allowing to obtain the preservation of the water height positivity. The problem is addressed with the explicit version of the numerical fluxes, that is where $H^n_{K,i}$ is substituted to $\left(H^n_{K,i}\right)^2 / H^{n+1}_{K,i}$ in (9). From a practical point of view (and this is the solution adopted in our operational contexts), if the stability of the scheme is ensured with (8) in terms of dissipation of the mechanical energy, replacing $\left(H^n_{K,i}\right)^2 / H^{n+1}_{K,i}$ by $H^n_{K,i}$ is also widely sufficient to preserve stability, especially in the Low-Froude regimes where the variation of potential is very low. This simplified choice appears as the most relevant to guarantee the stability of the numerical method while remaining in a totally explicit environment. In this connexion, the explicit alternative of the perturbed fluxes discussed in Remark (3.5), possibly more diffusive, can also be considered, leading to very similar results.

3.2. *We consider the explicit version of the advective fluxes in the mass scheme (4), that is:*

$$\mathcal{F}^n_{e,i} = \overline{H\boldsymbol{u}}^n_{e,i} - \Pi^n_{e,i} = \overline{H\boldsymbol{u}}^n_{e,i} - \gamma\Delta t(\overline{H\mu})^n_{e,i} \frac{\boldsymbol{\delta}\Phi^n_{e,i}}{\varepsilon^2} \, , \tag{14}$$

where $(\overline{H\mu})^n_{e,i} = \frac{1}{2}\left((H\mu)^n_{K,i} + (H\mu)^n_{K_e,i}\right) = \frac{1}{2}\left(H^n_{K,i}\left(\frac{m_{\partial K}}{m_K}\right) + H^n_{K_e,i}\left(\frac{m_{\partial K_e}}{m_{K_e}}\right)\right)$. *Assume a CFL condition of the type:*

$$\Delta t \max\left(\frac{m_{\partial K}}{m_K}, \frac{m_{\partial K_e}}{m_{K_e}}\right)\left(|\overline{\boldsymbol{u}}^n_{e,i}.\boldsymbol{n}_{e,K}| + \sqrt{\gamma}\sqrt{\frac{|\boldsymbol{\delta}\Phi^n_{e,i}|}{\varepsilon^2}}\right) \leq \left(\frac{\beta}{\beta+1}\right)\frac{\min\left(H^n_{K,i}, H^n_{K_e,i}\right)}{\max\left(H^n_{K,i}, H^n_{K_e,i}\right)}, \tag{15}$$

where $0 < \beta \leq 1$. *Then:*

$$H^{n+1}_{K,i} \geq \frac{1}{\beta}\frac{\Delta t}{m_K}\sum_{e\in\partial K} -\left(\mathcal{F}^n_{e,i}.\boldsymbol{n}_{e,K}\right)^- m_e \geq 0\,. \tag{16}$$

Proof. Gathering

$$\frac{\Delta t}{m_K}\sum_{e\in\partial K} -\left(\mathcal{F}^n_{e,i}.\mathbf{n}_{e,K}\right)^- m_e \leq \frac{\Delta t}{m_K}\sum_{e\in\partial K}|\mathcal{F}^n_{e,i}.\mathbf{n}_{e,K}|m_e\,,$$

and

$$H^{n+1}_{K,i} \geq H^n_{K,i} - \frac{\Delta t}{m_K}\sum_{e\in\partial K}|\mathcal{F}^n_{e,i}.\mathbf{n}_{e,K}|m_e\,,$$

we get:

$$\begin{aligned}\beta H^{n+1}_{K,i} - \frac{\Delta t}{m_K}\sum_{e\in\partial K} -\left(\mathcal{F}^n_{e,i}.\mathbf{n}_{e,K}\right)^- m_e \geq \beta H^n_{K,i} &- (1+\beta)\frac{\Delta t}{m_K}\sum_{e\in\partial K}|\overline{H\mathbf{u}}^n_{e,i}.\mathbf{n}_{e,K}|m_e \\ &- (1+\beta)\frac{\Delta t}{m_K}\gamma\Delta t\sum_{e\in\partial K}(\overline{H\mu})^n_{e,i}\frac{|\boldsymbol{\delta}\Phi^n_{e,i}.\mathbf{n}_{e,K}|}{\varepsilon^2}m_e\,.\end{aligned}$$

From this, a sufficient condition to obtain (16) can be expressed locally as:

$$(1+\beta)\frac{\Delta t}{m_K}|\overline{H\mathbf{u}}^n_{e,i}.\mathbf{n}_{e,K}| + (1+\beta)\gamma\Delta t\frac{\Delta t}{m_K}(\overline{H\mu})^n_{e,i}\frac{|\boldsymbol{\delta}\Phi^n_{e,i}.\mathbf{n}_{e,K}|}{\varepsilon^2} \leq \beta\frac{H^n_{K,i}}{m_{\partial K}}\,,$$

This leads to:

$$\mu|\overline{\mathbf{u}}^n_{e,i}.\mathbf{n}_{e,K}| + \mu^2\gamma\frac{|\boldsymbol{\delta}\Phi^n_{e,i}.\mathbf{n}_{e,K}|}{\varepsilon^2} \leq \left(\frac{\beta}{1+\beta}\right)\frac{\min\left(H^n_{K,i}, H^n_{K_e,i}\right)}{\max\left(H^n_{K,i}, H^n_{K_e,i}\right)}, \tag{17}$$

where $\mu = \Delta t \max\left(\frac{m_{\partial K}}{m_K}, \frac{m_{\partial K_e}}{m_{K_e}}\right)$. Since the right member of the previous inequality is lower than 1, we conclude that (17) is ensured under (15).

3.1. *The CFL condition expressed above can be interpreted as the one based on the original fluxes (8) subject to a* $\mathcal{O}(\Delta t)$ *perturbation.*

□

3.2. *The quantity* $\boldsymbol{\delta}\Phi^n_{e,i}$ *being in the order of the mesh size, the CFL condition (15) is far less restrictive than a time step restriction of the form (12) in practise. From now, taking these aspects under consideration, we will assume that for all* $\beta > 0$ *the following positivity result holds under the CFL constraint (12):*

$$\frac{\Delta t}{m_K}\sum_{e\in\partial K} -\left(\mathcal{F}^n_{e,i}.\boldsymbol{n}_{e,K}\right)^- m_e \leq \beta H^{n+1}_{K,i}\,. \tag{18}$$

3.3. Energy dissipation

The main result of the current section concerns the dissipation of the mechanical energy at the discrete level. More precisely, we have the following result:

3.1. *Control of the mechanical energy*
We consider the numerical scheme (4), together with the corrected potential (5):

$$\Phi^{n,*}_{e,i} = \overline{\Phi}^n_{e,i} - \Lambda^n_{e,i} \quad , \quad \Lambda^n_{e,i} = \alpha\Delta t C_{\boldsymbol{\mathcal{H}}}\mu_e\delta(H\boldsymbol{u})^n_{e,i}\,,$$

and numerical fluxes (6):

$$\mathcal{F}_{e,i}^n = \overline{H\boldsymbol{u}}_{e,i}^n - \Pi_{e,i}^n \quad , \quad \Pi_{e,i}^n = \gamma \Delta t (\overline{H\mu})_{e,i}^{n+1/2} \frac{\boldsymbol{\delta\Phi}_{e,i}^n}{\varepsilon^2} \ ,$$

Assume that the time step is governed by an explicit CFL condition (12). Then, with the following calibration of the stabilization constants:

$$\alpha = \beta = 2 \, ,$$

we have the following control on the production of mechanical energy:

$$E^{n+1} - E^n \leq 0 \, . \tag{19}$$

To establish the announced result, we first give an estimate for the kinetic and potential energy productions, and finally show that the choice $\alpha = \beta = 2$ in (7) and (8) allows a global control of these contributions. The proof is organized around the following steps:

A Estimation of the kinetic energy production (**Proposition 3.3**).

B Estimation of the potential energy production (**Proposition 3.4**).

C Control of the mechanical energy: we gather the two inequalities resulting from $\#A$ and $\#B$ to deduce a sufficient condition on the stabilization constants γ, α present in the correction terms (8, 7) (**Proof of Theorem 3.1**).

3.3.1. Kinetic energy

We begin by the kinetic energy, and set:

$$\mathcal{K}_{K,i}^n = \frac{1}{2} H_{K,i}^n \left\| \mathbf{u}_{K,i}^n \right\|^2 \, .$$

We have the following result:

3.3. *Estimation of the kinetic energy production*

$$\mathcal{K}_{K,i}^{n+1} - \mathcal{K}_{K,i}^n + \frac{\Delta t}{m_K} \sum_{e \in \partial K} (\mathcal{G}_{\mathcal{K},e,i}^n . \boldsymbol{n}_{e,K}) m_e + \mathcal{Q}_{\mathcal{K},K,i} \leq \mathcal{R}_{\mathcal{K},K,i} + \mathcal{H}_{\mathcal{K},K,i} - \mathcal{A}_{\mathcal{K},K,i} + \tilde{\mathcal{A}}_{\mathcal{K},K,i} \, ,$$

with

$$\mathcal{G}_{\mathcal{K},e,i}^n . \boldsymbol{n}_{e,K} = \frac{1}{2} \left\| \boldsymbol{u}_{K,i}^n \right\|^2 \left(\mathcal{F}_{e,i}^n . \boldsymbol{n}_{e,K} \right)^+ + \frac{1}{2} \left\| \boldsymbol{u}_{K_e,i}^n \right\|^2 \left(\mathcal{F}_{e,i}^n . \boldsymbol{n}_{e,K} \right)^- \, ,$$

$$\mathcal{Q}_{\mathcal{K},K,i} = \frac{\Delta t}{m_K} H_{K,i}^n \boldsymbol{u}_{K,i}^n . \sum_{e \in \partial K} \frac{\boldsymbol{\delta\Phi}_{e,i}^n}{\varepsilon^2} m_e \, , \tag{20}$$

$$\mathcal{H}_{\mathcal{K},K,i} = \frac{\Delta t}{m_K} \sum_{e \in \partial K} \overline{H\boldsymbol{u}}_{e,i}^n . \frac{\Lambda_{e,i}^n}{\varepsilon^2} \boldsymbol{n}_{e,K} m_e \, ,$$

$$\mathcal{A}_{\mathcal{K},K,i} = \frac{\Delta t}{m_K} \sum_{e \in \partial K} \frac{\Lambda_{e,i}^n}{\varepsilon^2} \frac{1}{2} (H_{K_e,i}^n \boldsymbol{u}_{K_e,i}^n - H_{K,i}^n \boldsymbol{u}_{K,i}^n) . \boldsymbol{n}_{e,K} m_e \, , \tag{21}$$

$$\tilde{\mathcal{A}}_{\mathcal{K},K,i} = 2 \left(\frac{\Delta t}{m_K} \right)^2 \frac{(H_{K,i}^n)^2}{H_{K,i}^{n+1}} m_{\partial K} \sum_{e \in \partial K} \left(\frac{\Lambda_{e,i}^n}{\varepsilon^2} \right)^2 m_e \, , \tag{22}$$

$$\mathcal{R}_{\mathcal{K},K,i} = \left(\frac{\Delta t}{m_K} \right)^2 \frac{(H_{K,i}^n)^2}{H_{K,i}^{n+1}} m_{\partial K} \sum_{e \in \partial K} \left\| \frac{\boldsymbol{\delta\Phi}_{e,i}^n}{\varepsilon^2} \right\|^2 m_e \, . \tag{23}$$

Proof. We first use the equation on $\mathbf{u}$ (10):

$$H_{K,i}^{n+1}(\mathbf{u}_{K,i}^{n+1}-\mathbf{u}_{K,i}^{n}).\mathbf{u}_{K,i}^{n} = -\frac{\Delta t}{m_K}\sum_{e\in\partial K}(\mathbf{u}_{K_e,i}^{n}-\mathbf{u}_{K,i}^{n}).\mathbf{u}_{K,i}^{n}\big(\mathcal{F}_{e,i}^{n}.\mathbf{n}_{e,K}\big)^{-}m_e - \frac{\Delta t}{m_K}H_{K,i}^{n}\mathbf{u}_{K,i}^{n}.\sum_{e\in\partial K}\frac{\Phi_{e,i}^{n,*}}{\varepsilon^2}\mathbf{n}_{e,K}m_e\,.$$

Then, using the relation $(\mathbf{a}-\mathbf{b}).\mathbf{b}=\frac{1}{2}\|\mathbf{a}\|^2-\frac{1}{2}\|\mathbf{b}\|^2-\frac{1}{2}\|\mathbf{a}-\mathbf{b}\|^2$:

$$\begin{aligned}H_{K,i}^{n+1}\Big(\frac{1}{2}\Big\|\mathbf{u}_{K,i}^{n+1}\Big\|^2-\frac{1}{2}\|\mathbf{u}_{K,i}^{n}\|^2-\frac{1}{2}\Big\|\mathbf{u}_{K,i}^{n+1}-\mathbf{u}_{K,i}^{n}\Big\|^2\Big)&\\ = -\frac{\Delta t}{m_K}\sum_{e\in\partial K}\Big(\frac{1}{2}\|\mathbf{u}_{K_e,i}^{n}\|^2-\frac{1}{2}\|\mathbf{u}_{K,i}^{n}\|^2-\frac{1}{2}\|\mathbf{u}_{K_e,i}^{n}-\mathbf{u}_{K,i}^{n}\|^2\Big)\big(\mathcal{F}_{e,i}^{n}.\mathbf{n}_{e,K}\big)^{-}m_e&\\ -\frac{\Delta t}{m_K}H_{K,i}^{n}\mathbf{u}_{K,i}^{n}.\sum_{e\in\partial K}\frac{\Phi_{e,i}^{n,*}}{\varepsilon^2}\mathbf{n}_{e,K}m_e\,.&\end{aligned}$$

the previous equality and invoking the mass equation (4), we have:

$$\begin{aligned}\tilde{\mathcal{K}}_{K,i}^{n+1}-\tilde{\mathcal{K}}_{K,i}^{n} =& -\frac{\Delta t}{m_K}\sum_{e\in\partial K}\Big(\frac{1}{2}\|\mathbf{u}_{K,i}^{n}\|^2\big(\mathcal{F}_{e,i}^{n}.\mathbf{n}_{e,K}\big)^{+}+\frac{1}{2}\|\mathbf{u}_{K_e,i}^{n}\|^2\big(\mathcal{F}_{e,i}^{n}.\mathbf{n}_{e,K}\big)^{-}\Big)m_e\\ &+\frac{1}{2}H_{K,i}^{n+1}\Big\|\mathbf{u}_{K,i}^{n+1}-\mathbf{u}_{K,i}^{n}\Big\|^2+\frac{\Delta t}{m_K}\sum_{e\in\partial K}\frac{1}{2}\|\mathbf{u}_{K_e,i}^{n}-\mathbf{u}_{K,i}^{n}\|^2\big(\mathcal{F}_{e,i}^{n}.\mathbf{n}_{e,K}\big)^{-}m_e\\ &-\frac{\Delta t}{m_K}H_{K,i}^{n}\mathbf{u}_{K,i}^{n}.\sum_{e\in\partial K}\frac{\Phi_{e,i}^{n,*}}{\varepsilon^2}\mathbf{n}_{e,K}m_e\,.\end{aligned}\tag{24}$$

We now denote:

$$S_{K,i}=\frac{1}{2}H_{K,i}^{n+1}\Big\|\mathbf{u}_{K,i}^{n+1}-\mathbf{u}_{K,i}^{n}\Big\|^2+\frac{\Delta t}{m_K}\sum_{e\in\partial K}\frac{1}{2}\|\mathbf{u}_{K_e,i}^{n}-\mathbf{u}_{K,i}^{n}\|^2\big(\mathcal{F}_{e,i}^{n}.\mathbf{n}_{e,K}\big)^{-}m_e\,,$$

focus on the first term of $S_{K,i}$. We first use Jensen's inequality with the weights $1/4,1/2,1/4$ to obtain a control of the form:

$$\begin{aligned}\frac{1}{2}H_{K,i}^{n+1}\Big\|\mathbf{u}_{K,i}^{n+1}-\mathbf{u}_{K,i}^{n}\Big\|^2 \le\ & \frac{\big(H_{K,i}^{n}\big)^2}{H_{K,i}^{n+1}}\left(\frac{\Delta t}{m_K}\right)^2\left\|\sum_{e\in\partial K}\frac{\boldsymbol{\delta}\Phi_{e,i}^{n}}{\varepsilon^2}\mathbf{n}_{e,K}m_e\right\|^2\\ &+2\frac{\big(H_{K,i}^{n}\big)^2}{H_{K,i}^{n+1}}\left(\frac{\Delta t}{m_K}\right)^2\left\|\sum_{e\in\partial K}\frac{\Lambda_{e,i}^{n}}{\varepsilon^2}\mathbf{n}_{e,K}m_e\right\|^2\\ &+\frac{2}{H_{K,i}^{n+1}}\left(\frac{\Delta t}{m_K}\right)^2\left\|\sum_{e\in\partial K}\big(\mathbf{u}_{K_e,i}^{n}-\mathbf{u}_{K,i}^{n}\big)\big(\mathcal{F}_{e,i}^{n}.\mathbf{n}_{e,K}\big)^{-}m_e\right\|^2.\end{aligned}$$

We now carry on a separate analysis of each of the resulting terms. Using again Jensen's inequality:

$$\begin{aligned}\left\|\sum_{e\in\partial K}\frac{\overline{\Phi}_{e,i}^{n}}{\varepsilon^2}\mathbf{n}_{e,K}m_e\right\|^2 &\le m_{\partial K}\left(\sum_{e\in\partial K}\left\|\frac{\boldsymbol{\delta}\Phi_{e,i}^{n}}{\varepsilon^2}\right\|^2 m_e\right),\\ \left\|\sum_{e\in\partial K}\frac{\Lambda_{e,i}^{n}}{\varepsilon^2}\mathbf{n}_{e,K}m_e\right\|^2 &\le m_{\partial K}\left(\sum_{e\in\partial K}\left(\frac{\Lambda_{e,i}^{n}}{\varepsilon^2}\right)^2 m_e\right).\end{aligned}\tag{25}$$

On the other hand, the Cauchy-Schwarz inequality gives:

$$\left\|\sum_{e\in\partial K}(\mathbf{u}_{K_e,i}^{n}-\mathbf{u}_{K,i}^{n})\big(\mathcal{F}_{e,i}^{n}.\mathbf{n}_{e,K}\big)^{-}m_e\right\|^2\le\Big(\sum_{e\in\partial K}\|\mathbf{u}_{K_e,i}^{n}-\mathbf{u}_{K,i}^{n}\|^2\big(\mathcal{F}_{e,i}^{n}.\mathbf{n}_{e,K}\big)^{-}m_e\Big)\Big(\sum_{e\in\partial K}\big(\mathcal{F}_{e,i}^{n}.\mathbf{n}_{e,K}\big)^{-}m_e\Big)\,.$$

Thus:

$$S_{K,i} \leq \left(\frac{\Delta t}{m_K}\right)^2 \frac{(H^n_{K,i})^2}{H^{n+1}_{K,i}} m_{\partial K} \left(\sum_{e\in\partial K} \left\| \frac{\boldsymbol{\delta}\Phi^n_{e,i}}{\varepsilon^2} \right\|^2 m_e \right) + 2 \left(\frac{\Delta t}{m_K}\right)^2 \frac{(H^n_{K,i})^2}{H^{n+1}_{K,i}} m_{\partial K} \left(\sum_{e\in\partial K} \left(\frac{\Lambda^n_{e,i}}{\varepsilon^2} \right)^2 m_e \right)$$
$$+ \frac{1}{2} \frac{\Delta t}{m_K} \sum_{e\in\partial K} \left\| \mathbf{u}^n_{K_e,i} - \mathbf{u}^n_{K,i} \right\|^2 \left(\mathcal{F}^n_{e,i}.\mathbf{n}_{e,K} \right)^- m_e \times \left[1 - 4 \frac{\Delta t}{m_K} \sum_{e\in\partial K} \frac{-\left(\mathcal{F}^n_{e,i}.\mathbf{n}_{e,K} \right)^-}{H^{n+1}_{K,i}} m_e \right].$$

The third term being assumed negative according to Remark 3.2 (condition (18) with $\beta = 1/4$), this yields the remainder $\mathcal{R}_{\mathcal{K},K,i}$ (23) and the contribution $\tilde{\mathcal{A}}_{\mathcal{K},K,i}$ (22). Finally, using again (11) the term involving the potential forces in (24) is rewritten as:

$$\frac{\Delta t}{m_K} H^n_{K,i} \mathbf{u}^n_{K,i} . \sum_{e\in\partial K} \frac{\Phi^{n,*}_{e,i}}{\varepsilon^2} \mathbf{n}_{e,K} m_e = \frac{\Delta t}{m_K} H^n_{K,i} \mathbf{u}^n_{K,i} . \sum_{e\in\partial K} \frac{\boldsymbol{\delta}\Phi^n_{e,i}}{\varepsilon^2} m_e - \frac{\Delta t}{m_K} H^n_{K,i} \mathbf{u}^n_{K,i} . \sum_{e\in\partial K} \frac{\Lambda^n_{e,i}}{\varepsilon^2} \mathbf{n}_{e,K} m_e .$$

We use the relation $H^n_{K,i}\mathbf{u}^n_{K,i} = \overline{H\mathbf{u}}^n_{e,i} + \frac{1}{2}(H^n_{K,i}\mathbf{u}^n_{K,i} - H^n_{K_e,i}\mathbf{u}^n_{K_e,i})$ on the second member of the right hand side in the previous equality, to finally obtain $\mathcal{Q}_{\mathcal{K},K,i}$, $\mathcal{H}_{\mathcal{K},K,i}$ and $\mathcal{A}_{\mathcal{K},K,i}$. □

3.3.2. Potential energy

We now turn to the potential part, and denote $\mathcal{E}^n_K$ the potential energy on the cell K at time n. We have the following result:

3.4. *Estimation of the potential energy production:*

$$\mathcal{E}^{n+1}_K - \mathcal{E}^n_K + \frac{\Delta t}{m_K} \sum_{i=1}^{L} \sum_{e\in\partial K} \left(\mathcal{G}^n_{\mathcal{E},e,i}.\boldsymbol{n}_{e,K} \right) m_e - \mathcal{Q}_{\mathcal{E},K} \leq -\mathcal{R}_{\mathcal{E},K} + \mathcal{H}_{\mathcal{E},K} + \mathcal{A}_{\mathcal{E},K} + \tilde{\mathcal{R}}_{\mathcal{E},K} ,$$

with

$$\mathcal{G}^n_{\mathcal{E},e,i}.\boldsymbol{n}_{e,K} = \overline{\Phi}^n_{e,i} \mathcal{F}^n_{e,i}.\boldsymbol{n}_{e,K} ,$$
$$\mathcal{Q}_{\mathcal{E},K} = \frac{\Delta t}{m_K} \sum_{i=1}^{L} H^n_{K,i} \boldsymbol{u}^n_{K,i} . \sum_{e\in\partial K} \boldsymbol{\delta}\Phi^n_{e,i} m_e , \tag{26}$$
$$\mathcal{R}_{\mathcal{E},K} = \frac{\Delta t}{m_K} \sum_{i=1}^{L} \sum_{e\in\partial K} \Pi^n_{e,i} . \boldsymbol{\delta}\Phi^n_{e,i} m_e , \tag{27}$$
$$\mathcal{H}_{\mathcal{E},K} = \frac{\Delta t}{m_K} \sum_{i=1}^{L} \sum_{e\in\partial K} \left(\frac{H^n_{K_e,i}\boldsymbol{u}^n_{K_e,i} - H^n_{K,i}\boldsymbol{u}^n_{K,i}}{2} \right) . \boldsymbol{\delta}\Phi^n_{e,i} m_e ,$$

and the Taylor's residuals:

$$\tilde{\mathcal{R}}_{\mathcal{E},K} = C_{\boldsymbol{\mathcal{H}}} \left(\frac{\Delta t}{m_K} \right)^2 m_{\partial K} \sum_{i=1}^{L} \sum_{e\in\partial K} (\Pi^n_{e,i}.\boldsymbol{n}_{e,K})^2 m_e , \tag{28}$$
$$\mathcal{A}_{\mathcal{E},K} = C_{\boldsymbol{\mathcal{H}}} \left(\frac{\Delta t}{m_K} \right)^2 m_{\partial K} \sum_{i=1}^{L} \sum_{e\in\partial K} (\delta(H\boldsymbol{u})^n_{e,i})^2 m_e . \tag{29}$$

Proof. Using Taylor's formula between time steps n and $n+1$, we have for a certain $s \in [0,1]$:

$$\mathcal{E}^{n+1}_K - \mathcal{E}^n_K = -\frac{\Delta t}{m_K} \sum_{i=1}^{L} \sum_{e\in\partial K} \Phi^n_{K,i} \mathcal{F}^n_{e,i}.\mathbf{n}_{e,K} m_e + \frac{1}{2} \sum_{i=1}^{L} \sum_{j=1}^{L} (H^{n+1}_{K,i} - H^n_{K,i}) \boldsymbol{\mathcal{H}}^{n+s}_{ij,K} (H^{n+1}_{K,j} - H^n_{K,j}) ,$$

where $\boldsymbol{\mathcal{H}}^{n+s}_{ij,K} = \boldsymbol{\mathcal{H}}_{ij} \left(s\boldsymbol{H}^{n+1}_K + (1-s)\boldsymbol{H}^n_K, \mathbf{x}_K \right)$, where we recall that $\boldsymbol{H}^n_K = {}^t\left(H^n_{K,1}, \cdots, H^n_{K,L} \right)$. Then we call the following decomposition:

$$\Phi^n_{K,i} \mathcal{F}^n_{e,i}.\mathbf{n}_{e,K} m_e$$

Expanding $\overline{H\mathbf{u}}^n_{e,i} = H^n_{K,i}\mathbf{u}^n_{K,i} + \left(\dfrac{H^n_{K_e,i}\mathbf{u}^n_{K_e,i} - H^n_{K,i}\mathbf{u}^n_{K,i}}{2}\right)$ we recover the symmetric fluxes $\mathcal{G}^n_{\mathcal{E},e,i}.\mathbf{n}_{e,K}$ and the residuals $\mathcal{Q}_{\mathcal{E},K}$, $\mathcal{R}_{\mathcal{E},K}$, $\mathcal{H}_{\mathcal{E},K}$. Concerning now the Taylor's residual, we have, according to (3):

$$\mathcal{W}_{\mathcal{E},K} := \frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{L}(H^{n+1}_{K,i} - H^n_{K,i})\boldsymbol{\mathcal{H}}^{n+s}_{ij,K}(H^{n+1}_{K,j} - H^n_{K,j}) \leq \frac{1}{2}C_{\boldsymbol{\mathcal{H}}}\sum_{i=1}^{L}\left(H^{n+1}_{K,i} - H^n_{K,i}\right)^2 . \tag{30}$$

We reformulate the mass scheme:

$$\begin{aligned} H^{n+1}_{K,i} - H^n_{K,i} &= -\frac{\Delta t}{m_K}\sum_{e\in\partial K}\mathcal{F}^n_{e,i}.\mathbf{n}_{e,K}m_e = -\frac{\Delta t}{m_K}\sum_{e\in\partial K}\overline{H\mathbf{u}}^n_{e,i}.\mathbf{n}_{e,K}m_e + \frac{\Delta t}{m_K}\sum_{e\in\partial K}\Pi^n_{e,i}.\mathbf{n}_{e,K}m_e\,, \\ &= -\frac{\Delta t}{m_K}\sum_{e\in\partial K}\delta(H\mathbf{u})^n_{e,i}m_e + \frac{\Delta t}{m_K}\sum_{e\in\partial K}\Pi^n_{e,i}.\mathbf{n}_{e,K}m_e\,, \end{aligned}$$

where we recall that $\delta(H\mathbf{u})^n_{e,i} = \dfrac{1}{2}(H^n_{K_e,i}\mathbf{u}^n_{K_e,i} - H^n_{K,i}\mathbf{u}^n_{K,i}).\mathbf{n}_{e,K}$. Injecting this in (30), we use Jensen's inequality to obtain:

$$\begin{aligned} \mathcal{W}_{\mathcal{E},K} \leq & C_{\boldsymbol{\mathcal{H}}}\sum_{i=1}^{L}\left(\frac{\Delta t}{m_K}\sum_{e\in\partial K}\delta(H\mathbf{u})^n_{e,i}m_e\right)^2 + C_{\boldsymbol{\mathcal{H}}}\sum_{i=1}^{L}\left(\frac{\Delta t}{m_K}\sum_{e\in\partial K}\Pi^n_{e,i}.\mathbf{n}_{e,K}m_e\right)^2 \\ \leq & C_{\boldsymbol{\mathcal{H}}}\left(\frac{\Delta t}{m_K}\right)^2 m_{\partial K}\sum_{i=1}^{L}\sum_{e\in\partial K}(\delta(H\mathbf{u})^n_{e,i})^2m_e + C_{\boldsymbol{\mathcal{H}}}\left(\frac{\Delta t}{m_K}\right)^2 m_{\partial K}\sum_{i=1}^{L}\sum_{e\in\partial K}(\Pi^n_{e,i}.\mathbf{n}_{e,K})^2m_e\,, \end{aligned} \tag{31}$$

and fall on the two remaining terms of the estimation. □

3.3.3. Mechanical Energy (Proof of Theorem 3.1)

Let's now denote $E^n = \sum_{K\in\mathbb{T}} m_K\left(\mathcal{E}^n_K/\varepsilon^2 + \sum_{i=1}^{L}\mathcal{K}^n_{K,i}\right)$ the discrete mechanical energy, and focus on the non-antisymmetric terms. We first observe an exact balance between the terms (20) and (26) arising from the kinetic and potential parts. In consequence the effort is put on a simultaneous control of the terms $\mathcal{R}$ and $\mathcal{A}$ appearing in the kinetic and potential energy budgets.

Terms in $\mathcal{R}$:

We gather the contributions issuing from the estimations on the kinetic and potential discrete energies, i.e. (23) and (27,28) respectively:

$$\begin{aligned} m_K\sum_{i=1}^{L}\mathcal{R}_{\mathcal{K},K,i} &= (\Delta t)^2\sum_{i=1}^{L}\left(\frac{(H^n_{K,i})^2}{H^{n+1}_{K,i}}\frac{m_{\partial K}}{m_K}\right)\sum_{e\in\partial K}\left\|\frac{\boldsymbol{\delta}\Phi^n_{e,i}}{\varepsilon^2}\right\|^2 m_e\,, \\ -m_K\mathcal{R}_{\mathcal{E},K}/\varepsilon^2 &= -\Delta t\sum_{i=1}^{L}\sum_{e\in\partial K}\Pi^n_{e,i}.\frac{\boldsymbol{\delta}\Phi^n_{e,i}}{\varepsilon^2}m_e\,, \\ m_K\tilde{\mathcal{R}}_{\mathcal{E},K}/\varepsilon^2 &= (\Delta t)^2 C_{\boldsymbol{\mathcal{H}}}\left(\frac{m_{\partial K}}{m_K}\right)\sum_{i=1}^{L}\sum_{e\in\partial K}\left(\frac{\Pi^n_{e,i}.\mathbf{n}_{e,K}}{\varepsilon}\right)^2 m_e\,. \end{aligned}$$

As a preliminary step we split the first contribution $m_K\sum_{i=1}^{L}\mathcal{R}_{\mathcal{K},K,i}$ in a sum of symmetric and antisymmetric parts:

$$\begin{aligned} m_K\sum_{i=1}^{L}\mathcal{R}_{\mathcal{K},K,i} = & (\Delta t)^2\sum_{i=1}^{L}\left(\frac{(H\mu)^{n+1/2}_{K,i} + (H\mu)^{n+1/2}_{K_e,i}}{2}\right)\sum_{e\in\partial K}\left\|\frac{\boldsymbol{\delta}\Phi^n_{e,i}}{\varepsilon^2}\right\|^2 m_e \\ & + (\Delta t)^2\sum_{i=1}^{L}\left(\frac{(H\mu)^{n+1/2}_{K,i} - (H\mu)^{n+1/2}_{K_e,i}}{2}\right)\sum_{e\in\partial K}\left\|\frac{\boldsymbol{\delta}\Phi^n_{e,i}}{\varepsilon^2}\right\|^2 m_e\,. \end{aligned}$$

where we have introduced the notation $(H\mu)^{n+1/2}_{K,i} = \left(\dfrac{(H^n_{K,i})^2}{H^{n+1}_{K,i}}\dfrac{m_{\partial K}}{m_K}\right)$.
In a similar way, with $\mu_K = \dfrac{m_{\partial K}}{m_K}$, the term $m_K\tilde{\mathcal{R}}_{\mathcal{E},K}/\varepsilon^2$ reads:

$$m_K\tilde{\mathcal{R}}_{\mathcal{E},K}/\varepsilon^2 = (\Delta t)^2\, C_{\boldsymbol{\mathcal{H}}}\left(\frac{\mu_K+\mu_{K_e}}{2}\right)\sum_{i=1}^{L}\sum_{e\in\partial K}\left(\frac{\Pi^n_{e,i}.\mathbf{n}_{e,K}}{\varepsilon}\right)^2 m_e$$
$$+ (\Delta t)^2\, C_{\boldsymbol{\mathcal{H}}}\left(\frac{\mu_K-\mu_{K_e}}{2}\right)\sum_{i=1}^{L}\sum_{e\in\partial K}\left(\frac{\Pi^n_{e,i}.\mathbf{n}_{e,K}}{\varepsilon}\right)^2 m_e\,.$$

Dropping the antisymmetric terms, which vanish after global summation, we use (8):

$$\Pi^n_{e,i} = \gamma\Delta t(\overline{H\mu})^{n+1/2}_{e,i}\frac{\boldsymbol{\delta}\Phi^n_{e,i}}{\varepsilon^2} = \gamma\Delta t\left(\frac{(H\mu)^{n+1/2}_{K,i}+(H\mu)^{n+1/2}_{K_e,i}}{2}\right)\frac{\boldsymbol{\delta}\Phi^n_{e,i}}{\varepsilon^2}\qquad,\qquad \gamma>0\,,$$

to write the total contribution as:

$$\sum_{K\in\mathbb{T}} m_K\left(\sum_{i=1}^{L}\mathcal{R}_{\mathcal{K},K,i}\tilde{\mathcal{R}}_{\mathcal{E},K}/\varepsilon^2 - \mathcal{R}_{\mathcal{E},K}/\varepsilon^2 + \tilde{\mathcal{R}}_{\mathcal{E},K}/\varepsilon^2\right)$$
$$= (\Delta t)^2\sum_{K\in\mathbb{T}}\sum_{i=1}^{L}\sum_{e\in\partial K}\left[1+\gamma^2\left(\frac{(\Delta t)^2}{\varepsilon^2}C_{\boldsymbol{\mathcal{H}}}(\overline{H\mu})^{n+1/2}_{e,i}\left(\frac{\mu_K+\mu_{K_e}}{2}\right)\right)-\gamma\right](\overline{H\mu})^{n+1/2}_{e,i}\left\|\frac{\boldsymbol{\delta}\Phi^n_{e,i}}{\varepsilon^2}\right\|^2 m_e\,. \tag{32}$$

Defining the quantity ρ_ε such that:

$$\rho^2_\varepsilon = C_{\boldsymbol{\mathcal{H}}}\frac{(\Delta t)^2}{\varepsilon^2}(\overline{H\mu})^{n+1/2}_{e,i}\left(\frac{\mu_K+\mu_{K_e}}{2}\right), \tag{33}$$

the negativity of (32) reduces to:

$$p(\gamma) = \rho^2_\varepsilon\gamma^2 - \gamma + 1 \le 0\,. \tag{34}$$

Based on the positivity of the discriminant (that is $\rho_\varepsilon \le \dfrac{1}{2}$) and the roots of p: $\gamma^{\pm} = \dfrac{1\pm\sqrt{1-4\rho^2_\varepsilon}}{2\rho^2_\varepsilon}$, one can establish that the value $\gamma = 2$ allows to ensure the negativity of p.

Terms in $\mathcal{A}$:
We consider the three remaining terms involved in the energy budget (21, 22 and 29):

$$-m_K\sum_{i=1}^{L}\mathcal{A}_{\mathcal{K},K,i} = -\Delta t\sum_{i=1}^{L}\sum_{e\in\partial K}\frac{\Lambda^n_{e,i}}{\varepsilon^2}\delta(H\mathbf{u})^n_{e,i}m_e\,,$$
$$m_K\sum_{i=1}^{L}\tilde{\mathcal{A}}_{\mathcal{K},K,i} = 2\,(\Delta t)^2\sum_{i=1}^{L}\left(\frac{(H^n_{K,i})^2}{H^{n+1}_{K,i}}\frac{m_{\partial K}}{m_K}\right)\sum_{e\in\partial K}\left(\frac{\Lambda^n_{e,i}}{\varepsilon^2}\right)^2 m_e\,,$$
$$m_K\mathcal{A}_{\mathcal{E},K}/\varepsilon^2 = (\Delta t)^2\, C_{\boldsymbol{\mathcal{H}}}\left(\frac{m_{\partial K}}{m_K}\right)\sum_{i=1}^{L}\sum_{e\in\partial K}(\delta(H\mathbf{u})^n_{e,i}/\varepsilon)^2 m_e\,.$$

In the spirit of the previous analysis we decompose $m_K \sum_{i=1}^L \tilde{\mathcal{A}}_{\mathcal{K},K,i}$ and $m_K \sum_{i=1}^L \mathcal{A}_{\mathcal{E},K}$:

$$
\begin{aligned}
m_K \sum_{i=1}^{L} \tilde{\mathcal{A}}_{\mathcal{K},K,i} &= 2\,(\Delta t)^2 \sum_{i=1}^{L} \left(\frac{(H\mu)_{K,i}^{n+1/2} + (H\mu)_{K_e,i}^{n+1/2}}{2} \right) \sum_{e\in\partial K} \left(\frac{\Lambda_{e,i}^n}{\varepsilon^2} \right)^2 m_e \\
&\quad + 2\,(\Delta t)^2 \sum_{i=1}^{L} \left(\frac{(H\mu)_{K,i}^{n+1/2} - (H\mu)_{K_e,i}^{n+1/2}}{2} \right) \sum_{e\in\partial K} \left(\frac{\Lambda_{e,i}^n}{\varepsilon^2} \right)^2 m_e \\
m_K \mathcal{A}_{\mathcal{E},K}/\varepsilon^2 &= (\Delta t)^2\, C_{\boldsymbol{\mathcal{H}}} \left(\frac{\mu_K + \mu_{K_e}}{2} \right) \sum_{i=1}^{L} \sum_{e\in\partial K} (\delta(H\mathbf{u})_{e,i}^n/\varepsilon)^2 m_e \\
&\quad + (\Delta t)^2\, C_{\boldsymbol{\mathcal{H}}} \left(\frac{\mu_K - \mu_{K_e}}{2} \right) \sum_{i=1}^{L} \sum_{e\in\partial K} (\delta(H\mathbf{u})_{e,i}^n/\varepsilon)^2 m_e\,.
\end{aligned}
$$

Again we neglect the antisymmetric terms, and consider (7):

$$
\Lambda_{e,i}^n = \alpha C_{\boldsymbol{\mathcal{H}}} \Delta t \mu_e \delta(H\mathbf{u})_{e,i}^n \quad , \quad \mu_e = \frac{1}{2}\left(\mu_K + \mu_{K_e}\right) = \frac{1}{2}\left(\frac{m_{\partial K}}{m_K} + \frac{m_{\partial K_e}}{m_{K_e}} \right) \quad , \quad \alpha > 0\,.
$$

The total contribution attached to these terms becomes:

$$
\begin{aligned}
&\sum_{K\in\mathbb{T}} \sum_{e\in\partial K} m_K \left(-\sum_{i=1}^{L} \tilde{\mathcal{A}}_{\mathcal{K},K,i} + \sum_{i=1}^{L} \tilde{\mathcal{A}}_{\mathcal{K},K,i} + \mathcal{A}_{\mathcal{E},K}/\varepsilon^2 \right) \\
&= (\Delta t)^2 \sum_{K\in\mathbb{T}} \sum_{i=1}^{L} \sum_{e\in\partial K} \left[-\alpha + \alpha^2 \left(2\frac{(\Delta t)^2}{\varepsilon^2} C_{\boldsymbol{\mathcal{H}}} (\overline{H\mu})_{e,i}^{n+1/2} \right) + 1 \right] C_{\boldsymbol{\mathcal{H}}} \mu_e \left(\frac{\delta(H\mathbf{u})_{e,i}^n}{\varepsilon} \right)^2 m_e\,.
\end{aligned}
\tag{35}
$$

Using the same notations as previously, we are this time left with the study of the second order polynomial:

$$
q(\alpha) = 2\rho_\varepsilon^2 \alpha^2 - \alpha + 1 \leq 0\,. \tag{36}
$$

Supposing that $\rho_\varepsilon \leq \dfrac{1}{2\sqrt{2}}$, the real roots are $\alpha^{\pm} = \dfrac{1 \pm \sqrt{1 - 8\rho_\varepsilon^2}}{4\rho_\varepsilon^2}$, from which we extract the value $\alpha = 2$ again.

3.3. *In one dimension and in the single layer case, the quantity ρ_ε simply reduces to $2\dfrac{\Delta t}{\Delta x}\dfrac{c}{\varepsilon}$ so that the smallness assumptions made on ρ_ε are satisfied under a classical explicit CFL condition. This is also the case for the general L layers case in 2d, where the conditions required on ρ_ε are always satisfied with a time constraint of the form (12) in practise.*

3.4. *Although they ensure a global decrease of the mechanical energy, numerical experiments also shown that the values $\alpha = \beta = 2$ brought too many diffusion in practise, especially in the Low-Froude regimes. As a matter of fact the optimality of the current approach has been lost within the Jensen's inequalities used during the estimations of the kinetic and potential energies (formulas (25) and (31) respectively). If a natural choice has been made on the weights for the sake of clarity, a more general result can be established introducing a general set of constants in these two inequalities. Playing with these parameters one can significantly relax the conditions on the stabilization constants. In the single layer case and one dimensional context for instance, the condition on γ becomes:*

$$
\gamma \in \left[\gamma^-, \gamma^+\right] \quad , \quad \textit{with} \quad \gamma^{\pm} = \frac{1 \pm \sqrt{1 - \rho_\varepsilon^2}}{\rho_\varepsilon^2}\,, \tag{37}
$$

where we recall that $\rho_\varepsilon = 2\dfrac{\Delta t}{\Delta x}\dfrac{c}{\varepsilon}$. Similar results can be obtained for α. Thus, when ρ_ε (or equivalently the CFL) decreases, it offers a more important latitude regarding the choice of γ and α, as illustrated in Fig. 3. When ρ_ε tends to zero, one recovers the critical value $\gamma = \alpha = \dfrac{1}{2}$, which also appears as a strict threshold both in the linear case and in our one dimensional numerical experiments with a first order time scheme.

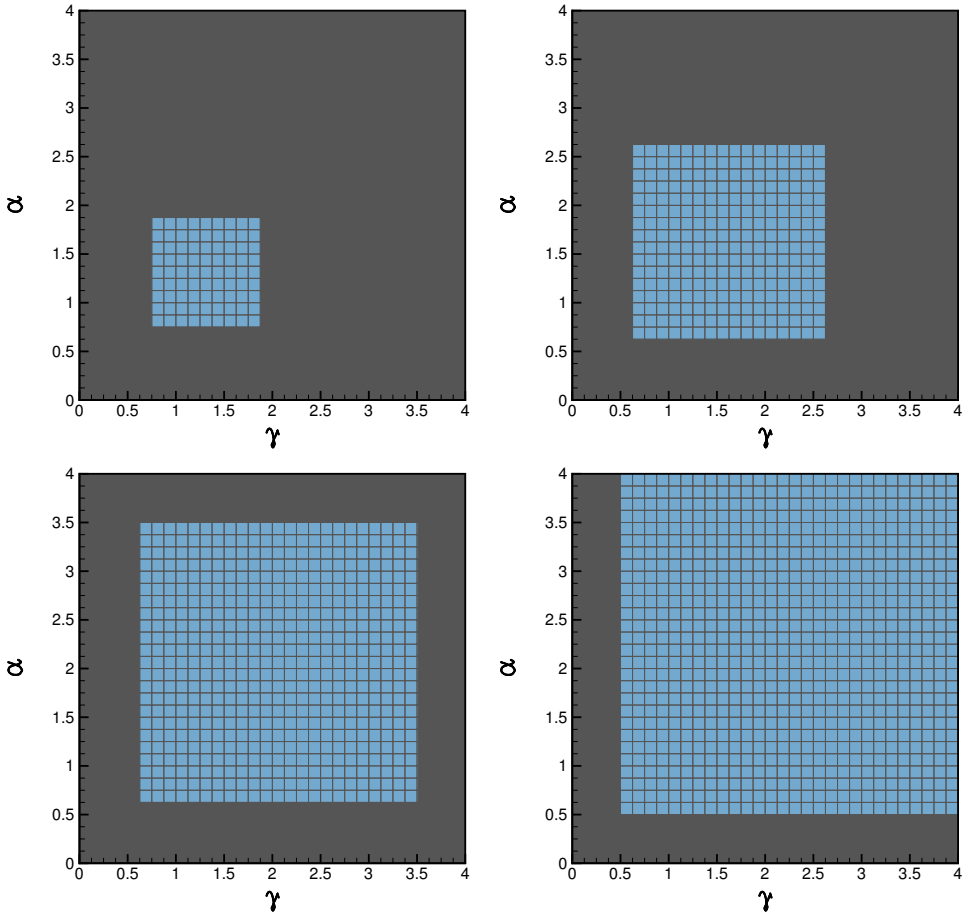

Figure 3: Non linear discrete analysis : (γ, α)-cartography at first order in the single layer case for decreasing CFL ($\frac{\Delta t}{\Delta x} =$ 0.45,0.40, 0.35 and 0.01 respectively). Stability regions appear in black.

It results that in practise, and at first order in space and time, one can get stability taking α and β in the vicinity of 1/2, even in the general case of arbitrary stratifications. As it will be discussed later, less restrictive conditions will be extracted from the linear stability analysis (see §3.4), where second order in space and time is also considered. The stabilizing effects of the second order time scheme allow to considerably relax the stabilization constants, and are in conformity with our numerical observations. Nevertheless, we have to point out here that the first order stability conditions that have been established here, based on a strict dissipation of the mechanical energy, are always included in those arising from the linear analysis, even when increasing the time and space order.

3.5. *It also has to be noticed that the formulation of $\Pi^n_{e,i}$ (8) is not totally explicit, since $(\overline{H\mu})^{n+1/2}_{e,i}$ depends on $\left(H^{n+1}_{K,i}, H^{n+1}_{K_e,i}\right)$, and therefore on the set of perturbations $\left\{\Pi^n_{e,i}\right\}_{e\in\partial K\cup\partial K_e}$ computed on the surrounding edges of K and K_e. As it this has been discussed in Remark 3.1, replacing $(H^n_{K,i})^2/H^{n+1}_{K,i}$ by $H^n_{K,i}$ only introduces a $\mathcal{O}(\Delta t)$ perturbation with respect to the original choice. As a consequence, from a practical point of view (and this is the solution adopted in our numerical experiments), replacing $(H^n_{K,i})^2/H^{n+1}_{K,i}$ by $H^n_{K,i}$ is quite sufficient to guarantee stability, especially in the Low-Froude regimes where the variation of potential is very low. However, at the price of being slightly more restrictive, a fully explicit condition can be exhibited. The strategy implies a global calibration of the stabilization parameters, to reduce to the study of a polynomial of order 3 at the level of each element.*

3.6. *As it has been discussed above, the domain of negativity of the polynomials p and q defined in (34) and (36) respectively can be enlarged by diminishing the CFL. One of the consequences is that the control (19) announced in the main result can be extended to a more general estimation. More precisely, let consider a small parameter $\delta > 0$, and a combination of values $(\Delta t, \alpha, \gamma)$ satisfying (34) and (36). Going back to the definition of the main coefficient of p and q, one easily obtains $p(\gamma) < -\delta$ and $q(\alpha) < -\delta$ with*

a space step Δt subject to a $\mathcal{O}(\delta)$ perturbation. Then, gathering (32) and (35), we obtain:

$$E^{n+1} - E^n \leq -\delta\,(\Delta t)^2 \sum_K \sum_{i=1}^{L} \sum_{e\in\partial K} (\overline{H\mu})_{e,i}^{n+1/2} \left\| \frac{\boldsymbol{\delta}\Phi_{e,i}^n}{\varepsilon^2} \right\|^2 m_e$$
$$-\delta\,(\Delta t)^2 \sum_K \sum_{i=1}^{L} \sum_{e\in\partial K} C_{\boldsymbol{\mathcal{H}}}\mu_e \left(\frac{\delta(H\boldsymbol{u})_{e,i}^n}{\varepsilon} \right)^2 m_e\,.$$

These estimates give a control of $L^1(0,T,H^1_w(\Omega))(u)$ with some adhoc weighted semi-norm on H^1_w. They insure validity of Lax Wendroff type theorem for weak consistency of conservative terms (in divergence form) in mass, momentum and energy equations. We refer to [20] and also [41] for further details concerning the use of such estimates to study consistency and convergence of the methods.

3.4. Discussion on linear stability

In this part we aim at assessing the relevance of the previous energetic considerations through linear stability arguments. For the sake of clarity the results of the current section are given in one dimension for the $L = 1$ case. The elements will be indexed by k and we denote $\mathcal{F}^n_{k+1/2}$ the numerical flux interface between the elements k and $k+1$. Let us take the example of negative fluxes. Using the definition $\phi^{\pm} = \dfrac{\phi \pm |\phi|}{2}$ we have:

$$\left(\mathcal{F}^n_{k+1/2}\right)^- = \mathcal{F}^n_{k+1/2} \quad \text{and} \quad \left(\mathcal{F}^n_{k+1/2}\right)^+ = 0\,.$$

In that context the mass (4) and velocity (10) schemes degenerate as follows:

$$\begin{aligned} H_k^{n+1} &= H_k^n - \frac{\Delta t}{\Delta x}\left[\mathcal{F}^n_{k+1/2} - \mathcal{F}^n_{k-1/2}\right], \\ u_k^{n+1} &= u_k^n - \frac{\Delta t}{\Delta x}\left(\Phi^{n,*}_{k+1/2} - \Phi^{n,*}_{k-1/2}\right) - \frac{\Delta t}{\Delta x}\left[\frac{u^n_{k+1}-u^n_k}{H^n_k}\mathcal{F}^n_{k+1/2}\right]. \end{aligned} \tag{38}$$

with the following numerical fluxes:

$$\mathcal{F}^n_{k+1/2} = \frac{H^n_k u^n_k + H^n_{k+1}u^n_{k+1}}{2} - 2\gamma\frac{\Delta t}{\Delta x}\left(\frac{H^n_k + H^n_{k+1}}{2}\right)\left(\frac{\Phi^n_{k+1}-\Phi^n_k}{2}\right),$$

and a corrected potential of the form:

$$\Phi^{n,*}_{k+1/2} = \left(\frac{\Phi^n_{k+1}+\Phi^n_k}{2}\right) - 2\alpha\frac{\Delta t}{\Delta x}C_{\boldsymbol{\mathcal{H}}}\left(\frac{H^n_{k+1}u^n_{k+1} - H^n_k u^n_k}{2}\right).$$

The scheme (38) is linearized around the constant state $\bar{w} = \left(\bar{H}, \bar{u}\right)$. Introducing a generic perturbation $\tilde{w}^n_k = \left(\tilde{H}^n_k, \tilde{u}^n_k\right)$ on the flow, we write:

$$H^n_k = \bar{H} + \tilde{H}^n_k \quad u^n_k = \bar{u} + \tilde{u}^n_k\,,$$

to obtain the following linearized system:

$$\begin{aligned} \tilde{H}_k^{n+1} &= \tilde{H}_k^n - \frac{\Delta t}{\Delta x}\left[\bar{H}\delta^n[\tilde{u}_k] + \bar{u}\delta^n[\tilde{H}_k] - 2\gamma\bar{\Phi}_H\frac{\Delta t}{\Delta x}\bar{H}\Delta^n[\tilde{H}_k]\right], \\ \tilde{u}_k^{n+1} &= \tilde{u}_k^n - \frac{\Delta t}{\Delta x}\left[\bar{\Phi}_H\delta^n[\tilde{H}_k] + \bar{u}d^n_+[\tilde{u}^n_k] - 2\alpha C_{\boldsymbol{\mathcal{H}}}\frac{\Delta t}{\Delta x}\left(\bar{H}\Delta^n[\tilde{u}_k] + \bar{u}\Delta^n[\tilde{H}_k]\right)\right], \end{aligned} \tag{39}$$

where we have set $\bar{\Phi}_H = \partial_H \Phi_{|\bar{H}}$, and with the following discrete operators:

$$\delta^n[f] = \frac{f^n_{k+1} - f^n_{k-1}}{2} \quad , \quad \Delta^n[f] = \frac{f^n_{k+1} + f^n_{k-1} - 2f^n_k}{2} \quad , \quad d^n_+[f] = f^n_{k+1} - f^n_k\,.$$

classically for solutions of the form $w_k^n = \widehat{w}^n exp\,(ik\Delta x)$ to the system (39), we obtain the following amplification matrix:

$$\widehat{w}^{n+1} = \begin{pmatrix} 1 - \tau\bar{u}isin(\Delta x) + 2\gamma\tau^2\bar{\Phi}_H\bar{H}\,(cos(\Delta x) - 1)) & -\tau\bar{H}isin(\Delta x) \\ -\tau\bar{\Phi}_H isin(\Delta x) + 2\alpha\tau^2 C_{\boldsymbol{\mathcal{H}}}\bar{u}\,(cos(\Delta x) - 1) & 1 - \tau\bar{u}\left(e^{i\Delta x} - 1\right) + 2\alpha\tau^2 C_{\boldsymbol{\mathcal{H}}}\bar{H}\,(cos(\Delta x) - 1) \end{pmatrix} \widehat{w}^n\,,$$

where we have set $\tau = \dfrac{\Delta t}{\Delta x}$. The amplification factor issuing from the previous system induces a relation between the CFL (i.e. τ) and the stabilization parameters (γ, α). To illustrate these dependencies, we propose several series of analysis in the one dimensional shallow water case (corresponding to $\bar{\Phi}_H = C_{\boldsymbol{\mathcal{H}}} = g$), allowing to draw up a cartography of the stability domain.

Concerning the influence of γ on the time step, Fig. 4 (*right*) proposes a pattern of the stability criteria attached to the linear system, developed around $\bar{c} = \left(g\bar{H}\right)^{1/2} = 1$ and $\bar{u} = 0$, for $\alpha = 0.5$. The admissible areas are filled in black. Using the same colour code, the results are put in balance with the optimized stability criteria issuing from the non-linear study (Fig. 4 (*left*)), that is the one dimensional relaxed condition (37) discussed in Remark 3.4. As expected, the study conducted in §3.3, based on a strict energy dissipation criteria, is more restrictive and fully embedded in the linear analysis.

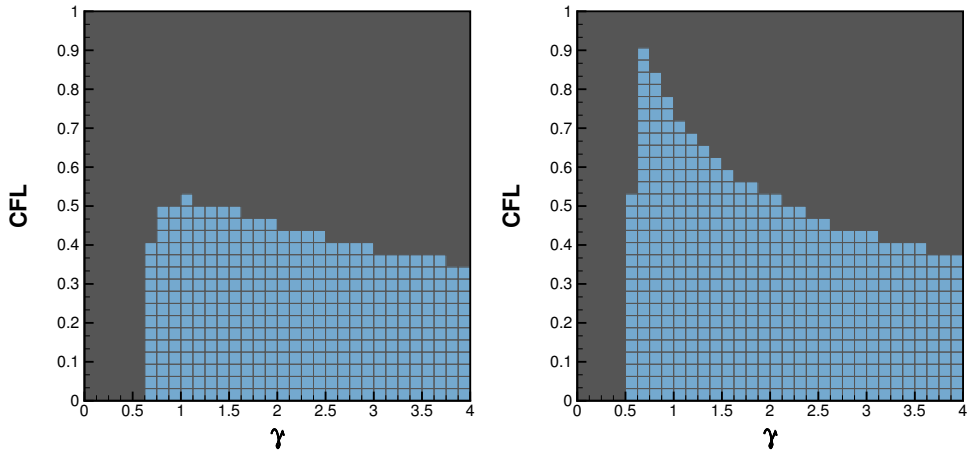

Figure 4: (γ, CFL)-cartography at first order in space in time. Energy dissipation (*left*) - Linear analysis (*right*).

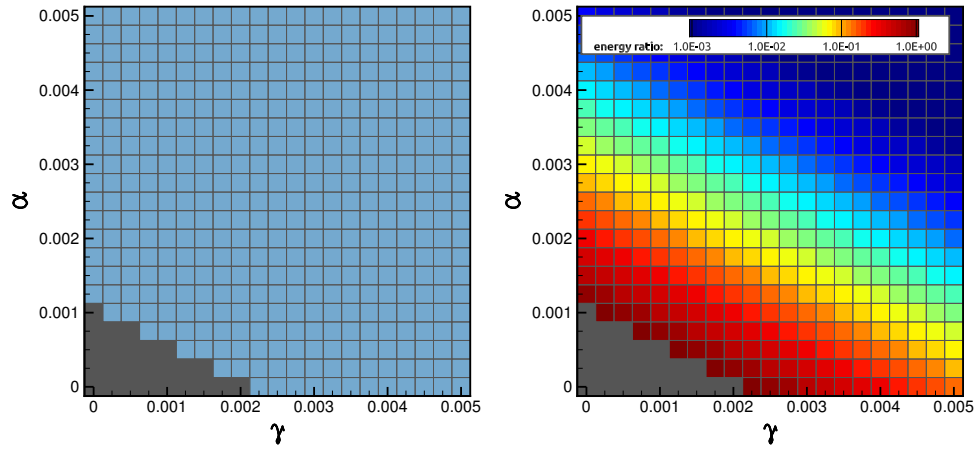

Figure 5: (γ, α)-cartography at first order in space and RK2 in time. Linear analysis (*left*) - Numerical experience (*right*). The CFL is equal to 0.5 in both tests.

As concerns the increase of time and space accuracy, if it is difficult to exhibit conclusions based on the fully discrete model, some interesting results can be established in the linear case. Other series of tests were made integrating a second order MUSCL reconstruction in space, together with a Heun time scheme (see Appendix 7.1.1 for implementations issues). From a general point of view, the improvement of time order comes with significant enhancements. In particular, the CFL can be increased while reducing the

required values for γ and α (see Fig. 5 (*left*) and 6 (*left*)). These results are also in accordance with those provided by our simulations in linear regimes (see Fig. 5 (*right*) and Fig. 6 (*right*)). These conclusions are of major interest from the extent that minimizing the diffusive losses is essential in our applicative contexts. Note finally that when α and γ are small, the amplification factor tends to a function of $2\alpha+\gamma$ only. Naturally, this has repercussions on the (γ,α)-cartography, even when using high order schemes in time and space. This behaviour can clearly be observed on Fig. 5 and Fig. 6, and matches with our numerical investigations.

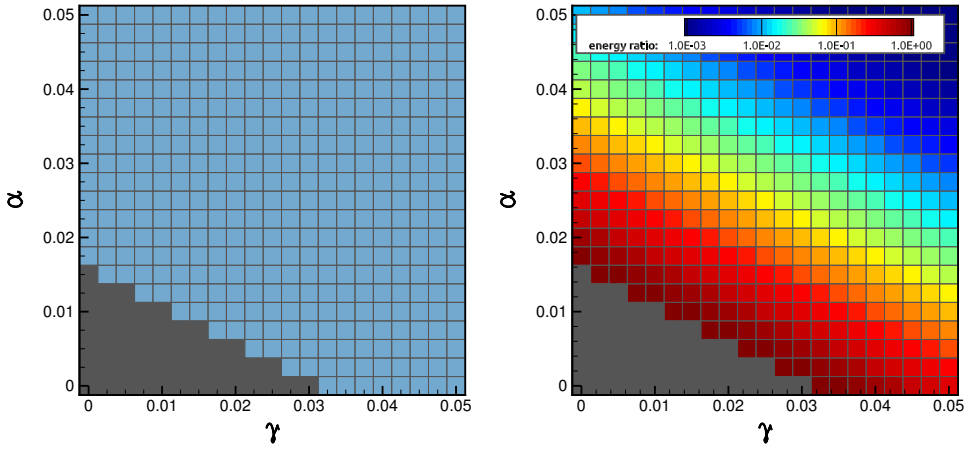

Figure 6: (γ,α)-cartography with MUSCL scheme in space and RK2 in time. Linear analysis (*left*) - Numerical experience (*right*). The CFL is equal to 0.5 in both tests.

4. Asymptotic regimes

We show in this part the asymptotic preserving features of the current approach. In the one dimensional frame, for a given time step Δt and space step Δx, the numerical scheme (4) can be interpreted at the semi-discrete level as follows:

$$\left\{\begin{array}{rcl} H_i^{n+1}-H_i^n & = & -\Delta t\partial_x(Hu)_i^n+(\Delta t)^2\gamma\partial_x\left(H_i\dfrac{\partial_x\Phi_i}{\varepsilon^2}\right)^n \\ (Hu)_i^{n+1}-(Hu)_i^n & = & -\Delta t\left(\partial_x\left(\bar{u}_i(Hu)_i^*\right)\right)^n \\ & & -\Delta t\left(H_i\dfrac{\partial_x\Phi_i}{\varepsilon^2}\right)^n+(\Delta t)^2\alpha\left(H_i\dfrac{\partial_{xx}(Hu)_i}{\varepsilon^2}\right)^n \end{array}\right. , \tag{40}$$

where $(Hu)_i^*=(Hu)_i-\Delta t\gamma\left(H_i\dfrac{\partial_x\Phi_i}{\varepsilon^2}\right)$, and $\bar{u}_i$ stands for the velocity u_i perturbed with a $\mathcal{O}(\Delta x)$ viscosity term resulting from the upwind strategy on the momentum equations. Note that the space step is submitted to a classical explicit CFL condition of the form:

$$\frac{\Delta t}{\Delta x}\left(u+\frac{c}{\varepsilon}\right)\leq cte\,. \tag{41}$$

In this section, for the sake of clarity, the results will be established in the continuous frame. We refer to [31] for some developments in a fully discrete context.

4.1. Fine time scale

For small time scale $t=\varepsilon\tau$ the model $(M_{t,\varepsilon})$ degenerates toward a system of wave equations (see [37], [12]):

$$\partial^2_{\tau\tau}H_i-div(H_i\nabla\Phi_i)=0\,. \tag{$M_{\tau,0}$}$$

4.1. *Consistency with the wave equations $(M_{\tau,0})$:*
Consider the time step scaling $\Delta t=\varepsilon\Delta\tau$. The semi-discrete model (40) furnishes an approximation of the wave equations $(M_{\tau,0})$ with an error in the order of $\mathcal{O}(\Delta\tau)$.

Proof. Using the mass equation of (40) at times n and $n+1$:

$$\begin{array}{rcrcrcl} H_i^{n+1} & - & H_i^n & = & -\varepsilon\Delta\tau\partial_x(Hu)_i^n & + & (\Delta\tau)^2\gamma\partial_x\left(H_i\partial_x\Phi_i\right)^n, \\ H_i^n & - & H_i^{n-1} & = & -\varepsilon\Delta\tau\partial_x(Hu)_i^{n-1} & + & (\Delta\tau)^2\gamma\partial_x\left(H_i\partial_x\Phi_i\right)^{n-1}, \end{array}$$

we write:

$$\frac{H_i^{n+1}-2H_i^n+H_i^{n-1}}{(\Delta\tau)^2} = -\frac{\varepsilon}{\Delta\tau}\left[\partial_x\left((Hu)_i^n-(Hu)_i^{n-1}\right)\right]+\gamma\left[\partial_x\left(\left(h_i\partial_x\Phi_i\right)^n-\left(h_i\partial_x\Phi_i\right)^{n-1}\right)\right]. \quad (42)$$

Consider now the momentum equations appearing in (40), and multiply by $\dfrac{\varepsilon}{\Delta\tau}$:

$$\begin{aligned} \frac{\varepsilon}{\Delta\tau}\left((Hu)_i^n-(Hu)_i^{n-1}\right) = &-\varepsilon^2\left(\partial_x\left(\bar{u}_i(Hu)_i^*\right)\right)^{n-1} \\ &-\varepsilon^2\left(H_i\frac{\partial_x\Phi_i}{\varepsilon^2}\right)^{n-1}+\varepsilon\Delta\tau\alpha\left(H_i\partial_{xx}\left((Hu)_i\right)\right)^{n-1}. \end{aligned} \quad (43)$$

Going back to the definition of $(Hu)_i^*$ we write:

$$\varepsilon^2(Hu)_i^* = \varepsilon^2\left((Hu)_i-\varepsilon\Delta\tau\gamma\partial_x\left(H_i\frac{\partial_x\Phi_i}{\varepsilon^2}\right)\right) = \underset{\varepsilon\to0}{\mathcal{O}}(\varepsilon^2)+\mathcal{O}(\Delta\tau).$$

Since $\bar{u}_i$ is $\mathcal{O}(1)$, we have as a direct consequence:

$$\varepsilon^2\left(\partial_x\left(\bar{u}_i(Hu)_i^*\right)\right) = \underset{\varepsilon\to0}{\mathcal{O}}(\varepsilon^2)+\mathcal{O}(\Delta\tau).$$

Finally, injecting (43) in (42) we obtain:

$$\frac{H_i^{n+1}-2H_i^n+H_i^{n-1}}{(\Delta\tau)^2} = \partial_x\left(H_i\partial_x\Phi_i\right)^n+\underset{\varepsilon\to0}{\mathcal{O}}(\varepsilon^2)+\mathcal{O}(\Delta\tau), \quad (44)$$

that is the one dimensional equivalent of $(M_{\tau,0})$ with an error in the order of $\Delta\tau$ and a second order perturbation. □

4.2. Large time scale

Assuming the Hessian $\boldsymbol{\mathcal{H}}$ well conditioned with respect to ε, that is the condition number of $\boldsymbol{\mathcal{H}}$ is $\underset{\varepsilon\to0}{\mathcal{O}}(1)$, the asymptotic regime associated with large time scales $t = \underset{\varepsilon\to0}{\mathcal{O}}(1)$ can be derived as a divergence-free model:

$$\begin{cases} div\left(H\mathbf{u}_i\right) = 0, \\ \partial_t\mathbf{u}_i+\left(\mathbf{u}_i.\nabla\right)\mathbf{u}_i = -\nabla\Phi_i \end{cases}. \quad (M_{t,0})$$

4.2. *Consistency with the divergence free model $(M_{t,0})$:*
Consider the time step scaling $\Delta t = \underset{\varepsilon\to0}{\mathcal{O}}(1)$, and assume that the spatial perturbation of the potential is in the order of ε^2:

$$\Phi_i(t,x) = \bar{\Phi}_i(t)+\varepsilon^2\hat{\Phi}_i(t,x). \quad (45)$$

Then then semi-discrete model (40) furnishes an approximation of the wave equations $(M_{t,0})$ with an error in the order of $\mathcal{O}(\Delta t)$ and $\mathcal{O}(\Delta t,\Delta x)$ respectively.

Proof. Note that with (45), and based on the regularity assumptions made on the potential forces (2.1), we also have:

$$\partial_t\Phi_i = \underset{\varepsilon\to0}{\mathcal{O}}(\varepsilon^2) \qquad \text{and} \qquad \partial_t H_i = \underset{\varepsilon\to0}{\mathcal{O}}(\varepsilon^2).$$

We directly obtain from the mass equation of (40):

$$\partial_x\left(Hu_i^*\right)^n = \partial_x\left(Hu_i\right)^n-\Delta t\gamma\partial_x\left(H_i\partial_x\hat{\Phi}_i\right)^n = \underset{\varepsilon\to0}{\mathcal{O}}(\varepsilon^2), \quad (46)$$

that is the divergence-free condition with an error in $\mathcal{O}(\Delta t)$ and a second order perturbation. Using the

relation:

$$(Hu)_i^{n+1} - (Hu)_i^n = \left(H_i^{n+1} - H_i^n\right) u_i^{n+1} + H_i^n \left(u_i^{n+1} - u_i^n\right) = H_i^n \left(u_i^{n+1} - u_i^n\right) + \underset{\varepsilon \to 0}{\mathcal{O}}(\varepsilon^2)\,,$$

together with the momentum equation available in (40) we write:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = -\frac{1}{H_i^n}\left(\partial_x\left(\bar{u}_i (Hu)_i^*\right) + \left(H_i \partial_x \hat{\Phi}\right)\right)^n + \Delta t \alpha \left(\frac{\partial_{xx}\left((Hu)_i\right)^n}{\varepsilon^2}\right) + \underset{\varepsilon \to 0}{\mathcal{O}}(\varepsilon^2)\,. \tag{47}$$

For any time n, we first note that:

$$\frac{1}{H_i}\partial_x\left(\bar{u}_i (Hu)_i^*\right) = u_i \partial_x u_i + \underset{\varepsilon \to 0}{\mathcal{O}}(\varepsilon^2) + \mathcal{O}(\Delta t, \Delta x) \tag{48}$$

going back to the semi-discrete divergence free relation (46), one has:

$$\partial_{xx}\left((Hu)_i\right) = \Delta t \gamma \partial_{xx}\left(H_i \partial_x \hat{\Phi}_i\right) + \underset{\varepsilon \to 0}{\mathcal{O}}(\varepsilon^2)\,, \tag{49}$$

and hence:

$$\Delta t \alpha \left(\frac{\partial_{xx}\left((Hu)_i\right)}{\varepsilon^2}\right) = \left(\frac{\Delta t}{\varepsilon}\right)^2 \alpha \gamma \partial_{xx}\left(H \partial_x \hat{\Phi}_i\right) + \mathcal{O}(\Delta t)\,. \tag{50}$$

Under the explicit CFL (41), the first term of the right hand side is $\mathcal{O}(\Delta x^2)$. At last this gives:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = -u_i \partial_x u_i - \partial_x \hat{\Phi}_i + \underset{\varepsilon \to 0}{\mathcal{O}}(\varepsilon^2) + \mathcal{O}(\Delta t, \Delta x)\,. \tag{51}$$

□

5. Numerical test cases

This part is dedicated to the survey of the numerical scheme's global efficiency at first and second order, with a particular focus on low Froude regimes. Theoretical and numerical investigations involving other specific mechanisms such as hydraulic jumps and/or evolution of wet/dry fronts, considering a non-trivial topography and a complex management of the layers, are left for future works. For the sake of completeness, the second order extension in space and time, the adaptative time step used and the time stepping scheme to incorporate the Coriolis force are given in the Appendix 7.

It should be underlined that it is difficult to carry on qualitative comparisons on the different numerical approaches in case of multiple layers. This is mainly due to the very low quantity of such academic test cases available in the literature, as it is difficult to derive analytical solutions. Some reference solutions for the multilayer shallow water model with the Coriolis force are of course provided by more sophisticated operational softwares like HYCOM [9], ROMS [35] or NEMO [27], but with the inconvenience of not being necessary exactly based on the same physical model as that considered here.

A first academic test case is considered, involving two-dimensional oscillating layers around a steady state in the linear small amplitude limit. It is investigated for this test the necessary inequality conditions on the two stabilization coefficients γ and α for the first and second order schemes, ones ensuring the linear stability and ones ensuring the strict decrease of the mechanical energy, with a minimum of dissipation and dispersion. A more advanced test case is then studied, the so called baroclinic vortex, that can be found in the COMODO benchmark [1], a test suite set-up by the international oceanographic community to evaluate and compare the numerical solvers efficiency.

5.1. Linear waves

In the present test case we investigate the two-dimensional simulation of oscillating layers around a steady state of flat layers in the presence of a flat bottom. In case of small amplitude of the waves, an approximate analytical solution can be derived from the linear wave theory. Considering only one layer, in the limit of small amplitude variation around the layer depth at rest η_0, the deviation ζ_1 ($\eta_1 = \eta_0 + \zeta_1$) is solution of the wave equation in two-dimensions with the associated dispersive relation $\omega^2 = c^2\left(k_x^2 + k_y^2\right)$, where k_x and k_y are the wave numbers in the x and y direction respectively. Considering now the same problem for the L layers shallow water model, we obtain L coupled linear wave equations,

$$\forall i \in [\![1, L]\!], \quad \frac{\partial^2 \zeta_i}{\partial t^2} - c_i^2 \sum_{j=1}^{L} \frac{min(\rho_i, \rho_j)}{\rho_i} \triangle \zeta_j = 0 \,.$$

By denoting $\tilde{c}_i$ the eigenvalues of the matrix $A_{ij} = \left(c_i^2 \dfrac{\min(\rho_i, \rho_j)}{\rho_i} \right)$, the above coupled system of wave equations can be rewritten in L uncoupled linear wave equations,

$$\forall i \in [\![1, L]\!], \quad \frac{\partial^2 \tilde{\zeta}_i}{\partial t^2} - \tilde{c}_i^2 \triangle \tilde{\zeta}_i = 0 \,,$$

where $\tilde{\zeta}$ is the projection of ζ onto the diagonal basis using the left eigenvectors matrix. Simulations are initialized in a 100 km square box with closed lateral walls, a sea surface at rest $\eta_0 = 5000$ m, five evenly spaced layers $h_i = 1000$ m with densities following a linear law $\rho_i = 1000 + 50\,(i-1)$ and a gravitational constant $g = 10$ $m.s^{-2}$. Note that the density ratios considered here, large in comparison with those encountered in more realistic contexts like oceans, have the effect of reducing the wave phase speed differences, allowing to consider a smaller time integration to capture the layers interaction. Considering a deviation $\zeta_1 = \cos(k_x x)\cos(k_y y)$ only for the first layer with one wavelength in each direction, approximatively 11 wave periods can be observed with a simulation time of one hour for a maximum wave velocity $\max(\tilde{c}_i) \approx \sqrt{2gh_0} \approx 316$ $m.s^{-1}$. Note that the associated regime is giving a very low Froude solution.

5.1.1. Stability issues - searching for optimal stabilization parameters

A preliminary goal for this test case is to search the range of the stabilization coefficients γ and α in order to achieve linear stability as a strict decrease of mechanical energy. To address that question, a first stopping criterion has been established, based on a clear exponential growing of the mechanical energy. A second stopping criterion, more restrictive, is to test if the mechanical energy decrease is violated between each time step. In order to find by numerical experiments the optimum values minimizing the dissipation without loosing these two stability criteria, thousands of simulations have been performed changing the two coefficients γ and α values with a step of 0.025. This experiment was carried out with the first and second order schemes (scheme (57 - 58) in Appendix 7) using a fixed CFL number of 0.5 with the adaptative time step (64). A 41×41 mesh is used at first order and a 11×11 mesh for the second order scheme, remain in the same order of mechanical energy diffusion. All the results are summarized for the first order scheme in Fig. 7 and for the second order scheme in Fig. 8.

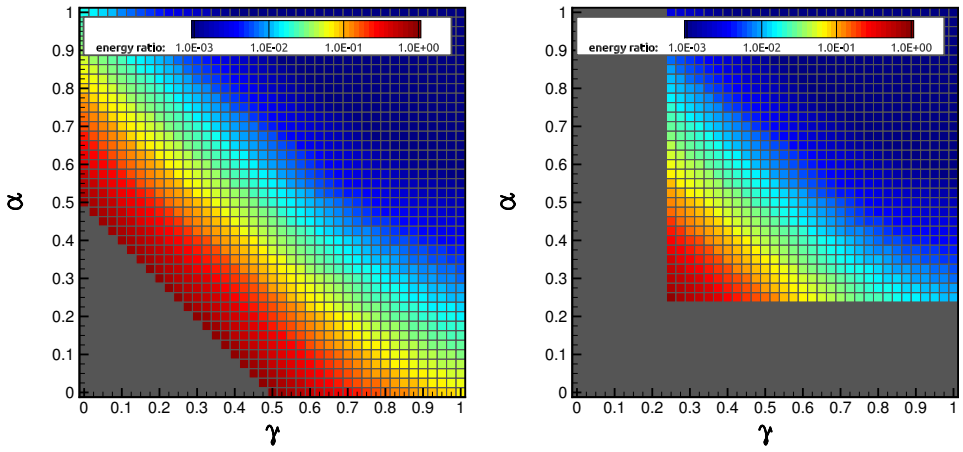

Figure 7: Mechanical energy ratio of dissipation function of γ and α with a fixed CFL number of 0.5 using the first order explicit scheme on a 41×41 mesh; (left) gray zone corresponds to an exponential growing of energy; (right) gray zone corresponds to a mechanical energy greater than at previous time step.

For the first criterion, which is corresponding to the linear stability of the scheme, it can be clearly observed for the first order scheme that the sum of the two coefficients γ and α must be greater than a minimum value of 0.5. For this sum value, it is found a minimum of dissipation. Notice that one of the two coefficients can be taken to zero. One must note that greater sum values introduce quickly and very nearly proportionally large amounts of dissipation. For the second order scheme, the same general

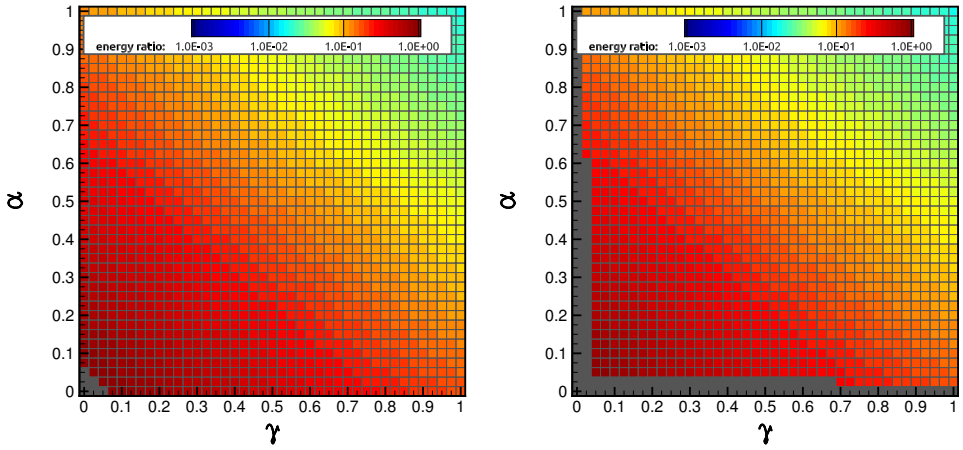

Figure 8: Mechanical energy ratio of dissipation function of γ and α with a fixed CFL number of 0.5 using the second order explicit scheme on a 11×11 mesh; (left) gray zone corresponds to an exponential growing of energy; (right) gray zone corresponds to a mechanical energy greater than at previous time step.

behaviour is observed, except that the minimum sum value found is now 0.075, really very lower than for the first order case. This result may be perceived unintuitive because MUSCL reconstructions tends to reduce the value of the stabilization term in the mass flux and for the pressure term for very regular solutions. We have verified that this is the time stepping scheme which mainly explains this reduction, changing profoundly the diffusion terms nature, as it was presented for the one dimensional linear analysis in the §3.4. The dissipation effect for greater sum values is also much more limited compared to the first order scheme. If we consider now the strict decrease of mechanical energy, the two coefficients must be both greater than a minimum value of 0.25 for the first order scheme, and a minimum value of 0.05 for the second order scheme except for too high inefficient values exhibiting more dissipation. This result confirms that the two stabilization coefficients are always both necessary in order to find a strict mechanical energy decrease.

The stability condition inequalities found for this test case of fast gravitational waves are summarized in Tab.1. It is found optimal values for the stabilization constants which are $\gamma = 0.25$ and $\alpha = 0.25$ for the first order scheme and $\gamma = 0.05$ and $\alpha = 0.05$ for the second order scheme if the CFL number is fixed to 0.5. Many other simulations were run in other contexts, without bringing any significant variability on these conditions. Finally note that these conditions are fully embedded in the results issuing from the one dimensional linear analysis presented in §3.4.

first order scheme		**second order scheme**	
linear stability	mechanical energy dissipation	linear stability	mechanical energy dissipation
$\gamma + \alpha \geq 0.5$	$\gamma \geq 0.25$ and $\alpha \geq 0.25$	$\gamma + \alpha \geq 0.075$	$\gamma \geq 0.05$ and $\alpha \geq 0.05$

Table 1: Stability inequalities conditions found by numerical experiments (two-dimensional gravity waves for one fixed wave-length in each direction) for the first and second order scheme.

A similar study has been performed considering varying values of γ and CFL numbers. second stabilization parameter α is now fixed to 0.25 for the first order scheme and 0.05 the second order scheme, corresponding to the optimal values found above for a CFL number of 0.5. The results are summarized in Fig. 9. For the first order scheme, it is found the same associated optimal stabilization coefficient $\gamma = 0.25$, regardless of the CFL number. For greater values, the eligible range of CFL numbers ensuring linear stability is reduced as γ increases, but it cannot be retrieved the same dissipation except for too low inefficient CFL numbers. For the second order scheme, the dependence with the CFL is more complicated and it is not clear how to extract an optimal stabilization parameter. An optimal value of 0.05 can be obtained for example with an optimal CFL number of 0.6. The behaviour seems to indicate that an optimal efficient CFL value is found for a stabilization coefficient $\gamma = 0.375$, but it can be checked

that the mechanical energy is not dissipated in this case.
At last, a general robustness property of the scheme is that taken lesser CFL numbers for given stabilization coefficients always reduces the dissipation, at first and second order.

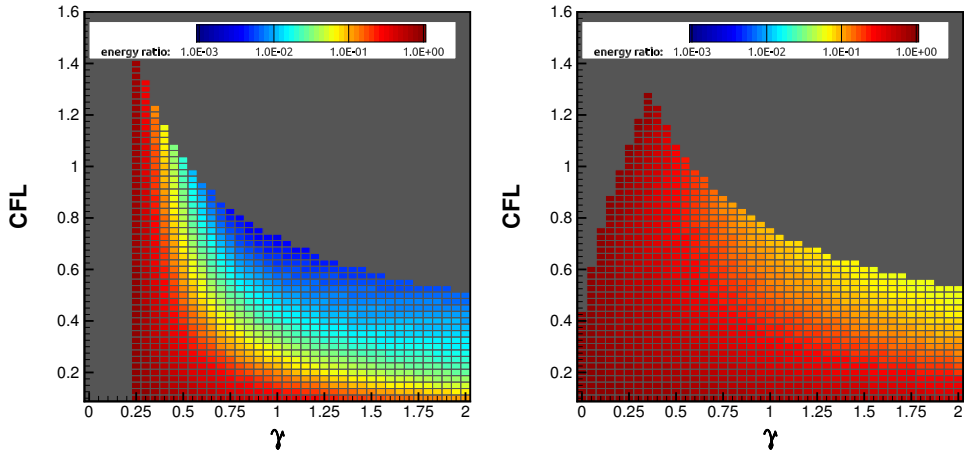

Figure 9: Mechanical energy ratio of dissipation function of γ and the CFL. Gray zone is corresponding to an exponential growing of energy; (left) using the first order explicit scheme with $\alpha = 0.25$ and a 41×41 mesh; (right) using the second order explicit scheme with $\alpha = 0.05$ and a 11×11 mesh.

5.1.2. Comparison with analytical solution

In Fig. 10 we propose the time evolution of the five surface layers using the first order scheme with $\gamma = 0.25$ and $\alpha = 0.25$ and the second order scheme with $\gamma = 0.05$ and $\alpha = 0.05$, both with a CFL number of 0.5, following the optimal values found in the previous section. The dispersive nature of the scheme can clearly be observed because of the obvious phase changes, although this effect is reduced by the second order scheme. Nevertheless, the scheme at first and second order is reproducing qualitatively very well the multiple interactions between the layers in light of the 11×11 coarse mesh used. For this resolution and stabilization constants, the second order scheme does exhibit a minimum of dissipation, only the dispersive effects can be clearly distinguished.

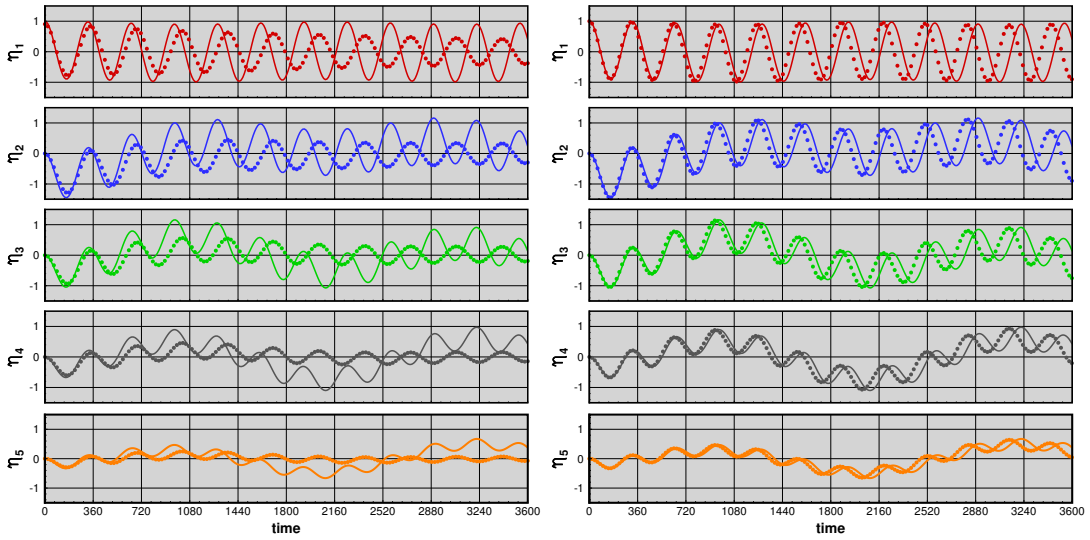

Figure 10: Evolution of the five surface layers at the box center for a 11×11 mesh with analytical solution in continuous line and numerical solution in dotted line; (left) using the first order scheme with $\gamma = 0.25$, $\alpha = 0.25$ and a CFL number of 0.5; (right) using the second order scheme with $\gamma = 0.05$, $\alpha = 0.05$ and a CFL number of 0.5.

5.2. Baroclinic vortex

on the COMODO benchmark [1], we study here an idealized axisymmetric and anticyclonic baroclinic vortex initially centered, propagating south-westward due to a β-plane approximation, following

the numerical experiment proposed in [32]. The vortex is expected to approximatively retains its axisymmetric shape with a progressive decrease of energy along its trajectory, mainly in the wave of emissions of weak-amplitude Rossby-waves. The present test represents a good indicator of the schemes accuracy, the principal difficulty lying here in the capability to describe precisely the vortex evolution. Indeed, the numerical diffusion and dispersion induced by unsuitable schemes or coarse resolutions can quickly break the axisymmetric nature of the vortex as the cyclostrophic initial balance and subsequently deteriorate the vortex trajectory.

5.2.1. Initialization

A vortex is placed at the center of the box $[-900\ km, 900\ km]^2$ with boundary walls with an axisymmetric Gaussian pressure profile:

$$\eta_1 = \frac{P_0}{g\rho_0} e^{-r^2/2\lambda^2} \quad , \tag{52}$$

where $\rho_0 = 1024.4\ kg.m^{-3}$ is the density at sea surface, $g = 9.81\ m.s^{-2}$ is the gravitational acceleration, $\lambda = 60\ km$ and $P_0 = \rho_0 f_0 u_{max} \lambda \sqrt{e}$ is a pressure defined from a maximum velocity $u_{max} = 0.8\ m.s^{-1}$, giving an anticyclonic vortex. In each layer i, the vortex at cyclostrophic equilibrium respects an axisymmetric balance between centripetal acceleration $v_{i,\theta}$, pressure p_i and Coriolis force:

$$-\frac{v_{i,\theta}^2}{r} - f v_{i,\theta} + \frac{dp_i}{dr} = 0\,.$$

Eliminating the unphysical solution of the above second order equation, we obtain the final expression of the velocity in each layer as a function of the layer pressure gradient in cylindrical coordinates:

$$v_{i,\theta} = -\frac{fr}{2}\left(1 - \sqrt{1 + \frac{4\frac{dp_i}{dr}}{rf^2}}\right)\,.$$

With simulations initialized with a velocity at geostrophic equilibrium $v_{i,\theta} = -\frac{1}{f}\frac{dp_i}{dr}$ as prescribed in the original test case [1], it has been found too much parasitic currents from initial numerical unbalance implying improper convergence. As a consequence, the value u_{max} considered here is smaller than in the original test case to ensure the positivity of the term in the square root.
A β-plane approximation is made for the Coriolis force:

$$f = f_0 + \beta y \quad , \tag{53}$$

with a latitude $\theta = 38.5°$, giving the two constants $f_0 = 2\Omega \sin(\theta) \simeq 9,054\ 10^{-5}$ and $\beta = 2\Omega \cos(\theta)/R_{earth} \simeq 1,788\ 10^{-11}$. The density distribution involves 10 layers at rest, evenly sized, following the linear law:

$$\rho_i = \rho_0\left(1 - \frac{N^2}{g} z_i\right) \quad \text{with} \quad z_i = \frac{h_0\left(i - \frac{1}{2}\right)}{N}\,, \tag{54}$$

where $N = 3.10^{-3}\ s^{-1}$ is the Brunt-Väisälä frequency and $h_0 = 5000\ m$ is the sea height. No motion is prescribed under a level $h_1 = 2500\ m$ in order to prevent from fast barotropic modes. It is derived here a formal way to nullify the velocity starting from the sixth layer. For a L layers system, the potential in the layer i can be written:

$$\Phi_i = \frac{g}{\rho_i}\left(\rho_1\eta_1 + \sum_{k=2}^{i}(\rho_k - \rho_{k-1})\,\eta_k\right)\,.$$

If we suppose $\nabla\Phi_i = 0$, it is found that:

$$\rho_1\nabla\eta_1 + \sum_{k=2}^{i}(\rho_k - \rho_{k-1})\,\nabla\eta_k = 0\,.$$

If $\nabla\eta_k = 0\ ,\ \forall k > i$, then we have also $\nabla\Phi_k = 0\ ,\ \forall k > i$. Suppose now that $\nabla\eta_{k+1} = \alpha\nabla\eta_k$. Then:

$$\rho_1 \nabla\eta_1 + \nabla\eta_2 \sum_{k=2}^{i} (\rho_k - \rho_{k-1})\, \alpha^{k-2} = 0\,.$$

This gives the final expression for the surface level gradient:

$$\nabla\eta_i = \frac{-\rho_1 \alpha^{i-2} \nabla\eta_1}{\sum_{k=2}^{i} (\rho_k - \rho_{k-1})\, \alpha^{k-2}}\,,$$

from which we extract the free surface distribution, adding the layer level at rest as a constant. Let us notice the inverse sign of the internal layer gradients compared to the sea surface gradient $\nabla\eta_1$ (since $\rho_i > \rho_{i+1}$), implying a pressure gradient decrease. Finally, we recall that we do not consider any viscosity or bottom friction effects in this test.

5.2.2. Simulations

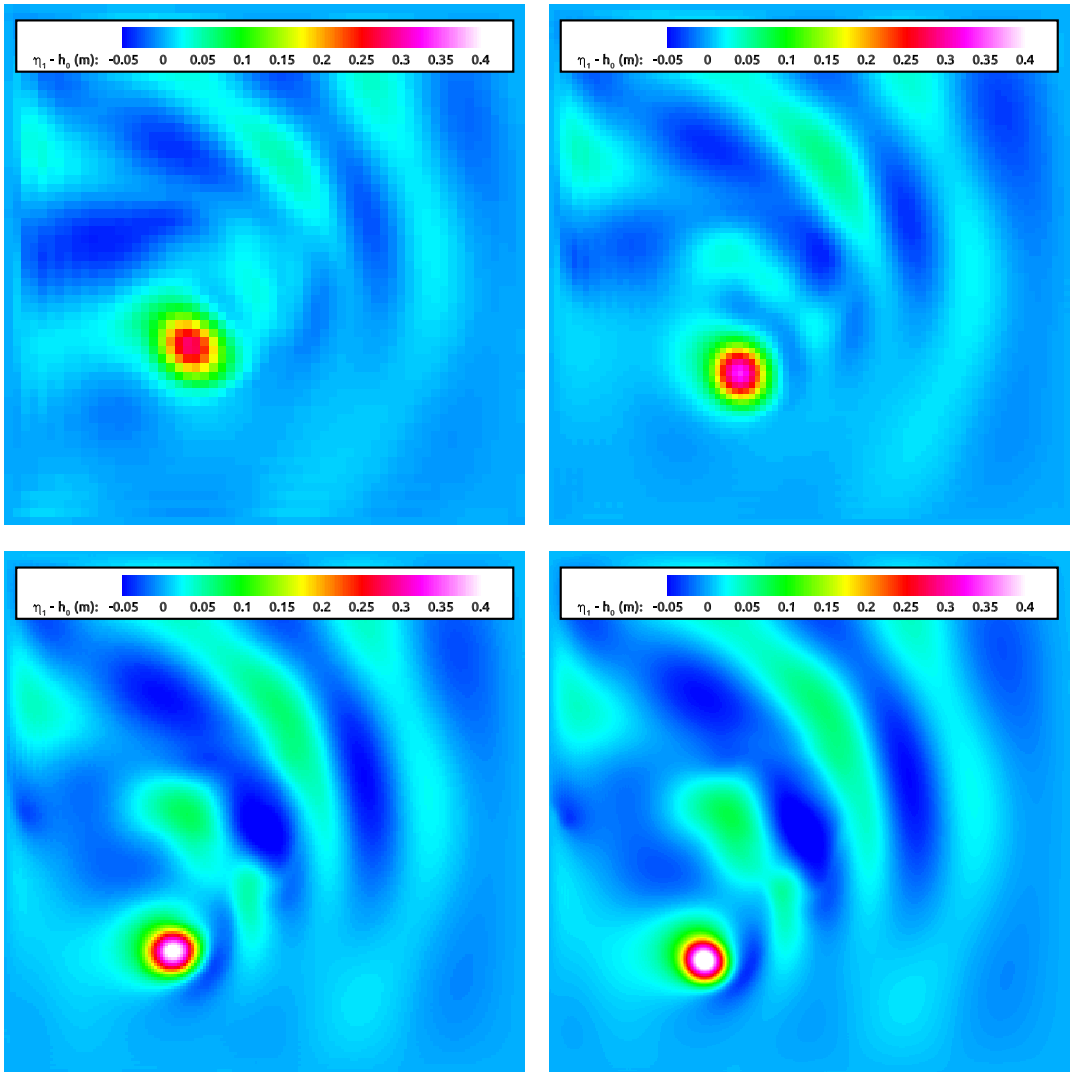

Figure 11: Sea surface levels for the baroclinic vortex test case. After 100 days of simulation, the vortex, initially centered, has moved to the southwest and small amplitudes Rossby waves emission can be observed in the trajectory wake; (top-left) $\Delta x = 30\ km$; (top-right) $\Delta x = 20\ km$; (upper-left) $\Delta x = 10\ km$; (upper-right) $\Delta x = 5\ km$.

Simulations have been performed using the second order scheme presented in the Appendix 7.1.2 with a time integration period of 100 days with five space resolutions $\Delta x = 30\,km$, $20\,km$, $10\,km$, $5\,km$ and $2\,km$ coresponding respectively to discretizations of space domain with $60 \times 60 \times 10$, $90 \times 90 \times 10$, $180 \times 180 \times$

10, $360\times360\times10$ and $900\times900\times10$ cells and layers. It has been chosen for the two stability coefficients a value of 0.1 for γ and 0 for α, with a CFL number of 0.5. We have seen before that this set of parameters is sufficient to ensure the linear stability of the numerical scheme. Some additional comments on this point will be discussed later on.

The surface levels for the first four resolutions are given in Fig. 11. It can roughly be observed a relative rapid convergence since the solutions for the $10\,km$ and $5\,km$ resolutions are already very close. The vortex final shape, as well as the position and amplitude of the Rossby waves in the trajectory wake, are very similar excepted maybe for very fine structures. For the lower resolutions of $30\,km$ and $20\,km$, the final axisymmetric vortex shape has not been completely broken, resulting to qualitative acceptable simulations in comparison to the converged result. The large structures of the emitted Rossby waves are correctly captured, especially the two bands in the northeast. However, the vortex has clearly lost an important energy as the maximum amplitude is lower than for the finer resolutions.

Going further in the convergence analysis, it is given in Fig. 12 the time evolution of the vortex y-deviation (computed from the maximum amplitude with bilinear interpolation), the vortex maximum amplitude, the kinetic energy and the total mechanical energy. Substracting to the potential energy the unperturbed state contribution, the mechanical energy has been rescaled to the initial value. First, the overall results for the $5\,km$ and $2\,km$ are sufficiently close to consider that the convergence has been reached. For the $10\,km$ resolution, all the curves are in very good agreement with the converged solution, giving a very satisfactory simulation for this time integration period. Towards the end of the simulation, the kinetic energy loss starts to move away the vortex trajectory from the converged one. It also results a lower maximum amplitude as these two quantities are completely correlated. For the two lower resolutions, the kinetic energy is lost directly at the beginning of the simulation because of a bad initial numerical balance between centripetal acceleration, pressure and Coriolis forces. Despite this initial unbalance, a lower decrease can be observed afterwards, highlighting a good accuracy for long time simulations.

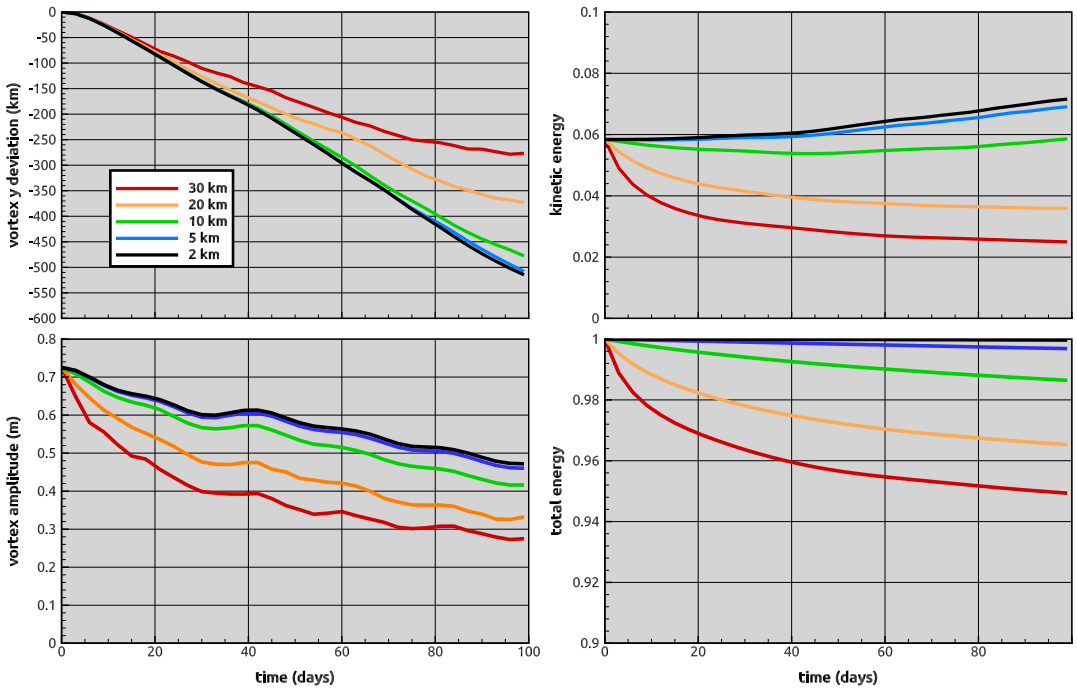

Figure 12: Time evolution of some revelant diagnostic quantities for the baroclinic vortex test case. Simulation coefficients are $\gamma = 0.1$ and $\alpha = 0$ with the explicit scheme.

From a numerical point of view, it can be observed for all the resolutions a strict decrease of the mechanical energy for the two chosen stabilization coefficients γ and α. It appears that in practise the parameter α is not always required to ensure a global control of the total mechanical energy, although the simulated flow is very complex. Since this parameter acts on a term proportional to the velocity divergence, it could be explained by the fact that the simulated flow is always close to the incompressible condition. We also have performed another series of simulations for the $10\,km$ resolution, with $\alpha = 0.025$, 0.050, 0.075 and 0.100 keeping the same other simulation parameters. The results given in Fig.

13 show a quick deterioration for increasing values of α. All the diagnostic quantities are approximately in a range of the results obtained with the $20\,km$ and $30\,km$ resolutions eliminating this stabilization term. The explanation is that the pressure term is impacted by a more important initial unbalance between centripetal acceleration, pressure and Coriolis forces. It can be easily checked through the initial decrease of the kinetic energy or the vortex maximum amplitude.

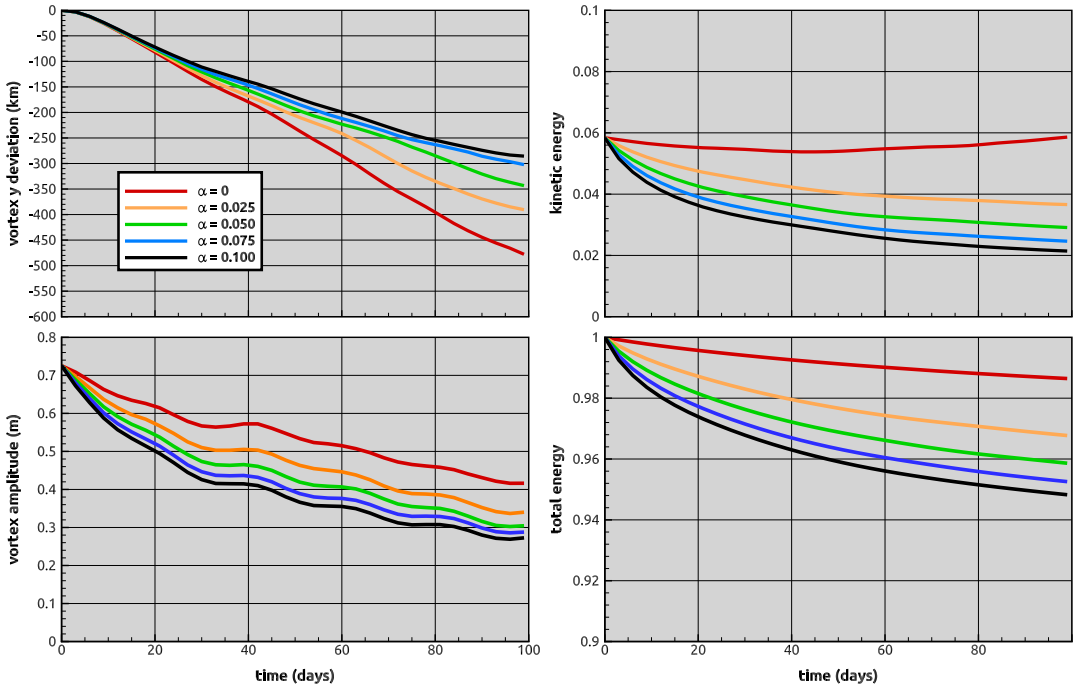

Figure 13: Influence of the α parameter on the same relevant diagnostic quantities.

6. Conclusion

In this paper we have introduced an explicit numerical scheme for the two-dimensional multilayer shallow water system with density stratification on unstructured meshes. The main characteristic of the numerical approach stands in its ability to deal with non conservative terms with strong stability properties, and without the need of evaluating the eigenvalues of the system. The formalism is particularly adapted to deal with well-balancing issues, and a positivity result is also exhibited. Assuming a classical explicit CFL condition, the dissipation of the mechanical energy has been demonstrated under sufficient inequality conditions on two stabilization coefficients, as well as the consistency with respect to the low-Froude regimes at different time scales, which stand for two fundamental and challenging criterion in the context of large scale oceanic or estuary flows. The stability analysis has been complemented through a linear study and put in balance with the numerical experiments. From a numerical point of view, it has been observed that the calibration of the stabilization constants could be significantly relaxed at second order with the use of an appropriate time scheme. The increase of space and time accuracy being generally not accompanied with a gain of stability, this behaviour may appear as counter-intuitive. That being so, the practical consequences are undeniable since it allows to considerably limit the diffusive losses in the numerical simulations. In view of these results, a more advanced high order space and time analysis is currently in progress, including an eventual extension to a general finite elements frame. As it is still confirmed by our numerical experiments, these stability properties make the approach particularly well suited to large scale oceanic circulation, and competitive with the other simulation platforms developed within the oceanographic community.

In addition to high order space and time extensions, many other perspectives are driven by the present developments. First, this explicit scheme must be balanced with its semi-implicit version, which accepts eventually bigger time steps, but at the price of a more important computational cost due to the presence of a nonlinear system on the implicit part and the difficulty to derive high order time and space extensions. Thus to date, the time benefits brought by the semi-implicit version are not so clear, especially since the

use of bigger time steps tends to rapidly deteriorate the accuracy and appropriate high order schemes need to be used in order to limit this drawback. The global stability analysis of the numerical scheme taking into account the Coriolis force with or without time stepping also needs to be performed. In addition, and in view of very promising preliminary results, the present approach is currently oriented toward other crucial operational contexts such as river flows or coastal applications, with the necessity to handle hydraulic jumps and wetting as in case of drying areas with management of disappearing layers or emerging topographies.

Acknowledgements

This work was granted access to the HPC resources of CALMIP supercomputing center under the allocation 2016-P1234.

7. Appendix

This appendix presents some technical aspects for implementation purposes, including the MUSCL reconstruction scheme, treatment of Coriolis force and the fully explicit formula used for the time step selection.

7.1. Second order scheme

7.1.1. MUSCL reconstructions

We consider in this work a monoslope second order MUSCL scheme, which consists of local linear reconstructions by computing a vectorial slope $[\nabla \mathbf{W}_K]_m$ in each cell K and for each primitive variable m, such that the two reconstructed primitive variables vectors $\mathbf{W}_{e,K}$ and $\mathbf{W}_{e,K_e}$ are evaluated at each side of edge e by:

$$\begin{array}{l} \mathbf{W}_{e,K} = \mathbf{W}_K + \nabla \mathbf{W}_K . \mathbf{x}_K \, \mathbf{x}_e \\ \mathbf{W}_{e,K_e} = \mathbf{W}_{K_e} + \nabla \mathbf{W}_{K_e} . \mathbf{x}_{K_e} \mathbf{x}_e \end{array} . \tag{55}$$

These quantities are intended to replace the primitive variables in the original first order scheme (Eqs. (4), (5) and (6)) to evaluate the numerical flux $\mathcal{F}_e^n$ and the pressure $\Phi_e^{n,*}$ at the edge e. Classically, with such a linear reconstruction, one can expect a scheme with a second-order accuracy in space for sufficient regular solutions. To this end, a least square method is employed to compute the vectorial slopes for each primitive variable h_K^n, u_K^n and v_K^n. More explicitly, the following sums of squares

$$E_m\left([\nabla \mathbf{W}_K]_m\right) = \sum_{e \in \partial K} \left([\mathbf{W}_{K_e}]_m - \left([\mathbf{W}_K]_m + [\nabla \mathbf{W}_K]_m . \mathbf{x}_K \mathbf{x}_{K_e}\right)\right)^2 , \tag{56}$$

are minimized by setting the gradients to zero solution of simple 2 x 2 linear systems. This method represents a good alternative among others to find the hyperplane because of its accuracy and robustness, independently from the number of neighbours. No limitation is imposed to the computed vectorial slope because all the numerical solutions considered in this work are largely sufficiently regular and far from wet/dry conditions to ensure numerical stability.

7.1.2. Second order scheme

With the two reconstructed primitive variables vectors $\mathbf{W}_{e,K}^n$ and $\mathbf{W}_{e,K_e}^n$ at each side of the edge e, interface terms are simply replaced in the original first order scheme. In the general L layer case, and omitting the subscript i referring to the layer numbering for the sake of clarity, this leads to the scheme:

$$\left\{ \begin{array}{lcll} h_K^{n+1} & = & h_K^n & - \dfrac{\Delta t}{m_K} \displaystyle\sum_{e \in \partial K} \left(\mathcal{F}_e^n . \mathbf{n}_{e,K}\right) m_e \\ h_K^{n+1} \mathbf{u}_K^{n+1} & = & h_K^n \mathbf{u}_K^n & - \dfrac{\Delta t}{m_K} h_K^n \displaystyle\sum_{e \in \partial K} \left(\Phi_e^{n,*} \, \mathbf{n}_{e,K}\right) m_e \\ & & & - \dfrac{\Delta t}{m_K} \displaystyle\sum_{e \in \partial K} \left(\mathbf{u}_{e,K}^n \left(\mathcal{F}_e^n . \mathbf{n}_{e,K}\right)^+ + \mathbf{u}_{e,K_e}^n \left(\mathcal{F}_e^n . \mathbf{n}_{e,K}\right)^- \right) m_e \end{array} \right. , \tag{57}$$

with

$$\left\{\begin{array}{l} \mathcal{F}_e^n = \dfrac{1}{2}\left(h_{e,K}^n \mathbf{u}_{e,K}^n + h_{e,K_e}^n \mathbf{u}_{e,K_e}^n\right) - \gamma \dfrac{\Delta t}{2}\left(h_{e,K}^n \dfrac{m_{\partial K}}{m_K} + h_{e,K_e}^n \dfrac{m_{\partial K_e}}{m_{K_e}}\right)\left(\dfrac{\Phi_{e,K_e}^n - \Phi_{e,K}^n}{2}\right)\mathbf{n}_{e,K} \\ \\ \Phi_e^{n,*} = \dfrac{1}{2}\left(\Phi_{e,K}^n + \Phi_{e,K_e}^n\right) - \alpha \dfrac{\Delta t}{2} gL\left(\dfrac{m_{\partial K}}{m_K} + \dfrac{m_{\partial K_e}}{m_{K_e}}\right)\left(\dfrac{h_{e,K_e}^n \mathbf{u}_{e,K_e}^n - h_{e,K}^n \mathbf{u}_{e,K}^n}{2}\right).\mathbf{n}_{e,K} \end{array}\right. . \quad (58)$$

For $\mathcal{F}_e^n$ we use the fully explicit version of the numerical fluxes, following comments of §3.2 and Remark 3.5. As concerns the corrected potential, $\Phi_e^{n,*}$, the constant $C_{\boldsymbol{\mathcal{H}}}$ is roughly estimated by gL/ρ, ρ standing for the density of the considered layer. All the vectorial slopes are first computed and the reconstructed primitive variables $h_{e,K}^n$, $\mathbf{u}_{e,K}^n$ are subsequently extracted at each edge side. The numerical scheme can afterwards be supplemented by a Heun scheme for time integration in order to derive a full second order scheme is space and time, stable under a classical CFL number.

7.2. Time stepping for Coriolis force

It has been demonstrated that under inequalities conditions on γ and α, the first order scheme given by Eqs. (4), (5) and (6) dissipates mechanical energy. This property has also been highlighted for the second order scheme (57) and (58), at least numerically, in §5.1. The proposed approach to incorporate the Coriolis force is designed to preserve at best these stability properties. From this perspective, a time stepping scheme is considered to integrate the following ordinary differential equations :

$$\frac{\partial}{\partial t}\begin{pmatrix} u \\ v \end{pmatrix} = f\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}. \quad (59)$$

Among the desired stability properties, one asks the numerical approach to be a symplectic integrator and to preserve kinetic energy, i.e. $\|\mathbf{u}\|^{n+1} = \|\mathbf{u}\|^n$. A first way to proceed is to consider the exact integration of the previous ordinary differential equations (59), resulting to the scheme:

$$\left\{\begin{array}{rcl} u^{n+1} &=& \cos\left(f\Delta t^n\, u^n\right) + \sin\left(f\Delta t^n\, v^n\right) \\ \\ v^{n+1} &=& \cos\left(f\Delta t^n\, v^n\right) - \sin\left(f\Delta t^n\, u^n\right) \end{array}\right. . \quad (60)$$

Another way is to consider the Crank-Nicolson scheme:

$$\left\{\begin{array}{rcl} u^{n+1} &=& \dfrac{f\Delta t^n}{2}\left(v^n + v^{n+1}\right) \\ \\ v^{n+1} &=& -\dfrac{f\Delta t^n}{2}\left(u^n + u^{n+1}\right) \end{array}\right. . \quad (61)$$

It has been found by numerical experience that the last scheme (61) with an IMEX time steeping scheme H-CN(2,2,2) defined below in Tab. 2 by his Butcher tableau is globally dissipative for long time simulations.

0	0	0
1	1	0
	1/2	1/2

0	0	0
1	1/2	1/2
	1/2	1/2

Table 2: second order IMEX scheme H-CN(2,2,2) with an explicit Heun scheme for the model without Coriolis force and a Crank-Nicolson scheme for the Coriolis force.

The above IMEX time steeping can be written for numerical implementation purpose as follows:

$$\begin{array}{rcl} \mathbf{U}_K^{(1)} &=& \mathbf{U}_K^n + \Delta t^n\, \mathcal{L}(\mathbf{U}_K^{(1)}) \\ \mathbf{U}_K^{(2)} &=& \mathbf{U}_K^{(1)} + \dfrac{\Delta t^n}{2}\mathcal{C}(\mathbf{U}_K^n) + \dfrac{\Delta t^n}{2}\mathcal{C}(\mathbf{U}_K^{(2)}) \\ \mathbf{U}_K^{(3)} &=& \mathbf{U}_K^{(2)} + \Delta t^n\, \mathcal{L}(\mathbf{U}_K^{(2)}) \\ \mathbf{U}_K^{n+1} &=& \dfrac{1}{2}\left(\mathbf{U}_K^n - \mathbf{U}_K^{(1)} + \mathbf{U}_K^{(2)} + \mathbf{U}_K^{(3)}\right) \end{array}, \quad (62)$$

where $\mathcal{L}$ is the numerical space integration of the homogeneous model (corresponding to Eqs. (57) and (58)) and $\mathcal{C}$ is the operator corresponding to the Coriolis force:

$$\mathcal{C}(\mathbf{U}_i) = \begin{bmatrix} 0 \\ f\, h_i u_i \\ -f\, h_i v_i \end{bmatrix}. \tag{63}$$

As it can be observed in Fig. 12 for the long time simulations of the baroclinic vortex, the mechanical energy is effectively dissipated using this time stepping scheme. These energy losses gradually become less important as the mesh resolution increases.

7.3. Time step

Based on (12), the numerical CFL-like condition for the time step Δt^n for all the two-dimensional simulations presented in this article is:

$$\Delta t^n = \tau_{CFL} \min_{K\in\Omega} \left(\frac{2\, m_K}{m_{\partial K} \left(\|\bar{\mathbf{u}}_K^n\| + \sqrt{g \bar{h}_K^n} \right)} \right), \tag{64}$$

where τ_{CFL} is the CFL number, $\bar{h}_K^n$ is the total water depth and $\|\bar{\mathbf{u}}_K^n\|$ is the mean velocity, computed from:

$$\begin{cases} \bar{h}_K^n = \displaystyle\sum_{i=1}^{L} h_{K,i}^n \\ \\ \|\bar{\mathbf{u}}_K^n\| = \dfrac{1}{\bar{h}_K^n} \sqrt{\left(\displaystyle\sum_{i=1}^{L} h_{K,i}^n u_{K,i}^n \right)^2 + \left(\displaystyle\sum_{i=1}^{L} h_{K,i}^n v_{K,i}^n \right)^2} \end{cases}. \tag{65}$$

The time step is thus calibrated on the barotropic gravity wave.

References

[1] COMODO benchmark, `http://indi.imag.fr/wordpress/`.

[2] FVCOM: The Unstructured Grid Finite Volume Community Ocean Model, `http://fvcom.smast.umassd.edu/fvcom/`.

[3] SLIM: Second-generation Louvain-la-Neuve Ice-ocean Model, `http://sites.uclouvain.be/slim/`.

[4] R. Abgrall, S. Karni, Two-layer shallow water system: a relaxation approach, SIAM Journal on Scientific Computing 31 (3) (2009) 1603 – 1627.

[5] E. Audusse, F. Benkhaldoun, S. Sari, M. Seaid, P. Tassi, A fast finite volume solver for multi-layered shallow water flows with mass exchange, Journal of Computational Physics 272 (2014) 23–45.

[6] E. Audusse, M.-O. Bristeau, M. Pelanti, J. Sainte-Marie, Approximation of the hydrostatic Navier–Stokes system for density stratified flows by a multilayer model: Kinetic interpretation and numerical solution, Journal of Computational Physics 230 (9) (2011) 3453 – 3478.

[7] A. Beljadid, A. Mohammadian, H. M. Qiblawey, An unstructured finite volume method for large-scale shallow flows using the fourth-order Adams scheme, Computers & Fluids 88 (2013) 579 – 589.

[8] C. Berthon, F. Foucher, T. Morales, An efficient splitting technique for two layer shallow water model, Numerical Methods for Partial Differential Equations 31 (5) (2015) 1396 – 1423.

[9] R. Bleck, An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates, Ocean Modelling 4 (1) (2002) 55 – 88.

[10] F. Bouchut, T. M. de Luna, An entropy satisfying scheme for two-layer shallow water equations with uncoupled treatment, ESAIM: Mathematical Modelling and Numerical Analysis 42 (4) (2008) 683 – 698.

[11] F. Bouchut, V. Zeitlin, A robust well-balanced scheme for multi-layer shallow water equations, Discrete and Continuous Dynamical Systems-Series B 13 (4) (2010) 739 – 758.

[12] D. Bresch, R. Klein, C. Lucas, Multiscale analyses for the Shallow Water equations, in: Computational Science and High Performance Computing IV, vol. 115 of Notes on Numerical Fluid Mechanics and Multidisciplinary Design, 2011, pp. 149 – 164.

[13] M. Castro, J. Macías, C. Parés, A Q-scheme for a class of systems of coupled conservation laws with source term. Application to a two-layer 1-D shallow water system, ESAIM: Mathematical Modelling and Numerical Analysis 35 (01) (2001) 107 – 127.

[14] A. Chertock, A. Kurganov, Z. Qu, T. Wu, Three-Layer Approximation of Two-Layer Shallow Water Equations, Mathematical Modelling and Analysis 18 (2013) 675 – 693.

[15] C. J. Cotter, J. Thuburn, A finite element exterior calculus framework for the rotating shallow-water equations, Journal of Computational Physics 257, Part B (2014) 1506 – 1526, physics-compatible numerical methods.

[16] S. Danilov, Ocean modelling on unstructured meshes, Ocean Modelling 69 (2013) 195 – 210.

[17] S. Dellacherie, Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number, Journal of Computational Physics (2010) 978 – 1016.

[18] M. J. C. Díaz, Y. Cheng, A. Chertock, A. Kurganov, Solving two-mode shallow water equations using finite volume methods, Communications in Computational Physics 16 (5) (2014) 1323 – 1354.

[19] V. Duchêne, The multilayer shallow water system in the limit of small density contrast, Asymptotic Analysis 98 (3) (2016) 189 – 235.

[20] R. Eymard, T. Gallouët, R. Herbin, Finite volume methods, Handbook of Numerical Analysis 7 (2000) 713–1018.

[21] A. Gassmann, A global hexagonal C-grid non-hydrostatic dynamical core (ICON-IAP) designed for energetic consistency, Quarterly Journal of the Royal Meteorological Society 139 (2012) 152 – 175.

[22] N. Grenier, J.-P. Vila, P.Villedieu, An accurate low-Mach scheme for a compressible two-fluid model applied to free-surface flows, Journal of Computational Physics 252 (2013) 1–19.

[23] A. Kurganov, G. Petrova, Central-Upwind Schemes for Two-Layer Shallow Water Equations, SIAM Journal on Scientific Computing 31 (3) (2009) 1742 – 1773.

[24] F. Lemarié, L. Debreu, G. Madec, J. Demange, J. M. Molines, M. Honnorat, Stability constraints for oceanic numerical models: implications for the formulation of time and space discretizations, Ocean Modelling 92 (2015) 124 – 148.

[25] M.-S. Liou, A sequel to ausm, part ii: Ausm+-up for all speeds, Journal of Computational Physics 214 (2006) 137 – 170.

[26] M.-S. Liou, C. J. Steffen, A new flux splitting scheme, Journal of Computational Physics 107 (1993) 23 – 39.

[27] G. Madec, and the NEMO team, NEMO ocean engine, Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27, ISSN, No 1288-1619 (2008).

[28] K. T. Mandli, A numerical method for the two layer shallow water equations with dry states, Ocean Modelling 72 (2013) 80–91.

[29] R. Monjarret, The multi-layer shallow water model with free surface. Numerical treatment of the open boundaries, Ph.D. thesis, Institut National Polytechnique de Toulouse, Université de Toulouse (2014).

[30] M. Parisot, J.-P. Vila, Numerical scheme for multilayer shallow-water model in the low Froude number regime, Comptes Rendus Mathematique 352 (2014) 953 – 957.

[31] M. Parisot, J.-P. Vila, Centered-potential regularization of advection upstream splitting method : Application to the multilayer shallow water model in the low Froude number regime, to appear in SIAM Journal on Numerical Analysis.

[32] P. Penven, L. Debreu, P. Marchesiello, J. C. McWilliams, Evaluation and application of the ROMS 1-way embedding procedure to the central california upwelling system, Ocean Modelling 12 (2006) 157 – 187.

[33] T. D. Ringler, J. Thuburn, J. B. Klemp, W. C. Skamarock, A unified approach to energy conservation and potential vorticity dynamics for arbitrarily-structured C-grids, Journal of Computational Physics 229 (9) (2010) 3065 – 3090.

[34] D. L. Roux, Spurious inertial oscillations in shallow water models, Journal of Computational Physics 231 (2012) 7959 – 7987.

[35] A. F. Shchepetkin, J. C. McWilliams, The regional oceanic modeling system (roms): a split-explicit, free-surface, topography-following-coordinate oceanic model, Ocean Modelling 9 (2005) 347 – 404.

[36] A. L. Stewart, P. J. Dellar, An energy and potential enstrophy conserving numerical scheme for the multi-layer shallow water equations with complete Coriolis force, Journal of Computational Physics 313 (2016) 99 – 120.

[37] W. A. Strauss, Partial Differential Equations : An Introduction, John Wiley, 1992.

[38] J. Szmelter, P. Smolarkiewicz, An edge-based unstructured mesh discretization in geospherical framework, Journal of Computational Physics 229 (2010) 4980 – 4995.

[39] J. Thuburn, T. Ringler, J. Klemp, W. Skamarock, Numerical representation of geostrophic modes on arbitrarily structured C-grids, Journal of Computational Physics 228 (2009) 8321 – 8335.

[40] G. K. Vallis, Atmospheric and Oceanic Fluid Dynamics, Cambridge University Press, Cambridge, U.K., 2006.

[41] J.-P. Vila, P. Villedieu, Convergence of an explicit finite volume scheme for first order symmetric systems, Numerische Mathematik 94 (2003) 573 – 602.