UNIVERSIDAD DE CONCEPCIÓN

Centro de Investigación en Ingeniería Matemática (CI^{2}MA)

A semi-implicit method for a degenerating convection-diffusion-reaction problem modeling secondary settling tanks

Carlos D. Acosta, Raimund Bürger,
Julio Careaga, Stefan Diehl,
Romel Pineda, Daniel Támara

PREPRINT 2024-13

SERIE DE PRE-PUBLICACIONES

A semi-implicit method for a degenerating convection-diffusion-reaction problem modeling secondary settling tanks

Carlos D. Acosta[0000−0002−6477−8984], Raimund Bürger[0000−0001−9298−8981], Julio Careaga[0000−0003−4466−6173], Stefan Diehl, Romel Pineda[0000−0002−5509−795X], and Daniel Támara[0009−0003−6910−1491]

Abstract A one-dimensional model of reactive sedimentation in wastewater treatment is formulated. The model combines three main elements: the Activated Sludge Model No.1 (ASM1) for the reactive part, a secondary sedimentation tank (SST) with a variable cross-sectional area, and a description with percentages of the solid phase. The final form of the model is a system of partial differential equations of the convection-diffusion-reaction type. Two new numerical methods are developed and compared for the proposed model, which differ in the discretization of the temporal variable. One method is explicit (method XPE) and the other is semi-implicit (method XPSI). In both cases, the Engquist-Osher numerical flux is used to approximate the nonlinear term of the convective flux of the solid particles. The numerical results demonstrate that the methods exhibit comparable accuracy in error estimation and that the semi-implicit scheme is more efficient in terms of computational time.

Carlos D. Acosta
Departamento de Matemáticas y Estadística, Universidad Nacional de Colombia sede Manizales, Manizales, Caldas 170003, Colombia, e-mail: cdacostam@unal.edu.co

Raimund Bürger
CI^{2}MA and Departamento de Ingeniería Matemática, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Casilla 160-C, Concepción, Chile, e-mail: rburger@ing-mat.udec.cl

Julio Careaga
Departamento de Matemática, Universidad del Bío-Bío, Concepción, Chile,
e-mail: jcareaga@ubiobio.cl

Stefan Diehl
Centre for Mathematical Sciences, Lund University, P.O. Box 118, S-221 00 Lund, Sweden,
e-mail: stefan.diehl@math.lth.se

Romel Pineda
Facultad de Ciencias, Universidad Central del Ecuador, Quito, Ecuador,
e-mail: rtpineda@uce.edu.ec

Daniel Támara
Departamento de Matemáticas y Estadística, Universidad Nacional de Colombia sede Manizales, Manizales, Caldas 170003, Colombia, e-mail: detamaral@unal.edu.co

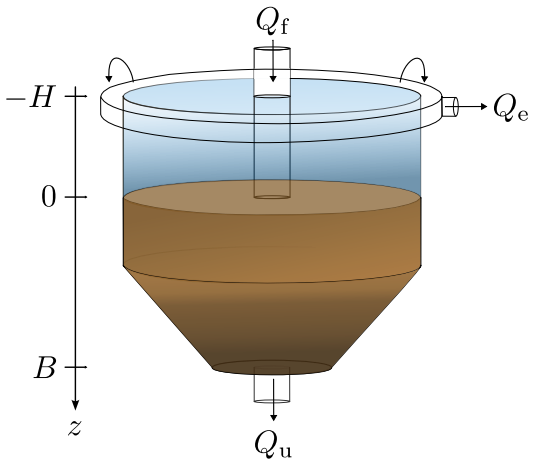

Fig. 1 Schematic of an axisymmetric secondary settling tank (SST). We assume a quasi-1D model of the sedimentation tank by letting the cross-sectional area $A = A(z)$ depend on depth z. The volumetric flows of the feed Q_{f}, effluent Q_{e} and underflow Q_{u} are shown

1 Introduction

Reactive settling is the combined process of sedimentation of small solid particles, each consisting of several components, dispersed in a viscous fluid with simultaneous reactions between the solids and soluble fluid components. This process is of particular importance in secondary settling tanks (SSTs) in wastewater treatment plants, more recently termed water resource recovery facilities (WRRFs). The primary purpose of an SST (see Figure 1) is to allow the biomass (essentially, bacteria) to settle out from the process effluent of a bioreactor. The overflow produced by the SST should ideally be water, while most of the sediment (activated sludge) leaves the unit through the underflow and is recycled to the bioreactor. The use of SSTs within wastewater treatment is described in detail e.g. in [11, 13, 22, 23]. Due to the living biomass (activated sludge; bacteria), biochemical reactions always occur. In particular, these reactions are the basis of the well-known activated sludge process in wastewater treatment [17, 18, 21]. Reactive settling occurs both in plants with continuously operated SSTs and in so-called sequencing batch reactors (SBRs) [2, 3, 5].

Mathematical models that are able to capture reactive settling are urgently needed for the simulation of operational scenarios. In a series of papers that includes [1, 4, 7–10, 12] some of the authors have contributed to the formulation of models of an SST, with various degrees of complexity with respect to model ingredients such as sediment compressibility, reaction kinetics and variability of the cross-sectional area. All of them give rise to spatially one-dimensional, nonlinear convection-diffusion or convection-diffusion-reaction partial differential equations (PDEs) with nonstandard ingredients such as discontinuous flux and partly degenerate diffusion. Sediment compressibility is usually modeled by a degenerating diffusion term. Available numerical methods for these models have, however, been based on explicit discretizations. If Δt and Δz denote the time step and spatial meshwidth of the numerical scheme, then the Courant-Friedrich-Lewy (CFL) condition

essentially compels a uniform bound of $\Delta t/\Delta z^2$, which makes simulations with fine discretizations or over long time intervals unacceptably slow. If the diffusion term is discretized implicitly then only $\Delta t/\Delta z$ needs to be bounded. The resulting scheme combines explicit and implicit discretizations of various terms and is therefore called semi-implicit; see [6]. We recently used this observation to define an efficient semi-implicit scheme for reactive settling in an SBR [5]. The novelty of this contribution is an analogue semi-implicit scheme for reactive settling in an SST.

2 Reactive sedimentation model

In this section, we present a convection-diffusion-reaction system of PDEs that models reactive settling in an SST with variable cross-sectional area. The system is described under two approaches: one in which the unknowns are the concentrations of the solid and liquid particles comprising a suspension, and another in which the solid phase is described in terms of percentages. For the first case, we follow the ideas of [1, 3, 4], and for the second, those of [5, 8, 12].

2.1 Reactive model with concentration vectors

The governing model can be written as the following one-dimensional system of convection-diffusion-reaction equations, where $z \in \mathbb{R}$ and $t \geq 0$ denote the spatial coordinate and time, respectively:

$$\begin{aligned} A(z)\partial_t \boldsymbol{C} + \partial_z\big(A(z)\boldsymbol{\mathcal{U}_C}(X,z,t)\boldsymbol{C}\big) &= \partial_z\big(\gamma(z)A(z)\partial_z(D(X)\boldsymbol{C})\big) \\ &\quad + \delta(z)\boldsymbol{C}_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\boldsymbol{R_C}(\boldsymbol{C},\boldsymbol{S}), \quad (1) \\ A(z)\partial_t \boldsymbol{S} + \partial_z\left(A(z)\boldsymbol{\mathcal{U}_S}(X,z,t)\boldsymbol{S}\right) &= \delta(z)\boldsymbol{S}_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\boldsymbol{R_S}(\boldsymbol{C},\boldsymbol{S}). \end{aligned}$$

The unknowns are the concentrations of solid and soluble particles, stored in the vectors $\boldsymbol{C} = \boldsymbol{C}(z,t)$ and $\boldsymbol{S} = \boldsymbol{S}(z,t)$ respectively. The total concentration of suspended solids is denoted by X, while the cross-sectional area of the tank is represented by $A = A(z)$, which depends on the depth z, where $z = 0$ indicates the top of the tank and $z = B$ the bottom. The characteristic function γ is equal to one inside the tank and zero outside it, i.e., $\gamma(z) := 1$ if $-H < z < B$ and $\gamma(z) := 0$ if $z \leq -H$ or $z \geq B$. Additionally, the velocities $\boldsymbol{\mathcal{U}_C}$ and $\boldsymbol{\mathcal{U}_S}$ depend nonlinearly on X and represent the velocities of the solid and liquid phases in the tank. The diffusive term involving D models the compressibility of the sediment. The terms involving $\delta = \delta(z)$ model the feed of the tank with suspension at volumetric flow rate $Q_{\mathrm{f}} = Q_{\mathrm{f}}(t)$. The last term in each equation contains the reaction rates (local increase in mass per unit time and volume) $\boldsymbol{R_C}$ and $\boldsymbol{R_S}$ of the solid and liquid components, respectively. The system (1) is complemented with appropriate initial conditions; boundary conditions are not necessary.

2.1.1 Description of the solid and liquid phases and phase velocities

The subsequent description of the solid and liquid phases follows that of [5, Sect. 2.2.1]. Two constitutive functions describe the sedimentation-compression process of the flocculated particles that consist of several components. These functions are stated in terms of the solids in suspension X. This quantity equals the sum of either all or of most of the particulate concentrations; the precise definition of X depends on the specific reaction model. Within the standard ASMx models, concentrations are usually expressed in terms of easily measurable units such as chemical oxygen demand (COD) (see Table 3 in Appendix A); conversion factors have to be used to obtain the mass concentrations. We here use the model ASM1 (see Appendix A) with the particulate concentrations (in ASM1 units) X_I, X_S, $X_\mathrm{B,H}$, $X_\mathrm{B,A}$, X_P, and X_ND, and the corresponding definition of the total suspended solids concentration is

$$X \coloneqq c(X_\mathrm{I} + X_\mathrm{S} + X_\mathrm{B,H} + X_\mathrm{B,A} + X_\mathrm{P}), \quad \text{where } c = 0.75\,\mathrm{g}/(\mathrm{g\,COD}). \tag{2}$$

The conversion factor c ensures that X has the proper mass unit, since gravity acts on the total solids in the model.

The concentration X_ND is not a summand in (2) since X_ND represents the nitrogen that is already part of X_S. To ensure that the total solids concentration X equals the sum of all particulate components (for mathematical reasons), we replace the variable X_S by $X_\mathrm{S-ND} \coloneqq X_\mathrm{S} - X_\mathrm{ND}$, and define (in ASM1 units)

$$\begin{aligned} \boldsymbol{C} &\coloneqq (X_\mathrm{I}, X_\mathrm{S-ND}, X_\mathrm{B,H}, X_\mathrm{B,A}, X_\mathrm{P}, X_\mathrm{ND})^\mathrm{T} \quad (k_{\boldsymbol{C}} = 6), \\ \boldsymbol{S} &\coloneqq (S_\mathrm{I}, S_\mathrm{S}, S_\mathrm{O}, S_\mathrm{NO}, S_\mathrm{NH}, S_\mathrm{ND})^\mathrm{T} \quad (k_{\boldsymbol{S}} = 6). \end{aligned} \tag{3}$$

We define

$$X \coloneqq c\big(C^{(1)} + \cdots + C^{(k_{\boldsymbol{C}})}\big) \quad \text{and} \quad L \coloneqq W + S^{(1)} + \cdots + S^{(k_{\boldsymbol{S}})}, \tag{4}$$

and when $X > 0$, the vector of percentages

$$\boldsymbol{p} \coloneqq (c/X)\boldsymbol{C}. \tag{5}$$

Clearly, these definitions imply that

$$p^{(1)} + \cdots + p^{(k_{\boldsymbol{C}})} = 1. \tag{6}$$

If $X = 0$, the values of $\boldsymbol{p}$ are irrelevant; however, they must always satisfy (6).

The total liquid concentration is L and W is the concentration of water. All these liquid components are assumed to have the constant density ρ_L. Conversion factors similar to c appear for the soluble concentrations, but we will divide these factors away directly, since the left-hand sides of the governing equations to be presented are linear in $\boldsymbol{C}$ and $\boldsymbol{S}$ apart from the coefficients, which are nonlinear functions of X.

All concentrations depend on depth z and time t. It is assumed that all solids have the same density ρ_X, which is considered greater than the maximum solid

concentration $X_{\max}$. The density of the liquid phase is $\rho_L < \rho_X$. Since the liquid phase in the feed consists mainly of water, typically ρ_L is taken as the density of water, regardless of the concentrations of the other soluble components.

It is assumed that the initial concentrations $X(z,0)$ and $L(z,0)$, and those in the feed, $X_{\mathrm{f}}(t)$ and $L_{\mathrm{f}}(t)$, are consistent with the definition of X, i.e., satisfy (4)–(6).

The volumetric flows in the feed, discharge and effluent are functions of t and satisfy $Q_{\mathrm{f}} \geq Q_{\mathrm{u}} > 0$ and that $Q_{\mathrm{e}} = Q_{\mathrm{f}} - Q_{\mathrm{u}}$. The volume for which the model is formulated does not change, in marked contrast to the operation of a so-called sequencing batch reactor (SBR), where the level of the mixture surface varies depending on the stage being executed (see [2, 3, 5]).

The velocities of the solid and liquid phases are denoted by $v_X = v_X(z,t)$ and $v_L = v_L(z,t)$ respectively, and as derived in [1, 3, 4], they are given by

$$v_X \coloneqq q + v, \quad v_L \coloneqq q - \frac{X/\rho_X}{1 - X/\rho_X} v, \tag{7}$$

where the average volumetric velocity of the mixture $q = q(z,t)$ satisfies

$$A(z)q(z,t) = \begin{cases} -Q_{\mathrm{e}}(t) = Q_{\mathrm{u}}(t) - Q_{\mathrm{f}}(t) & \text{if } z \leq 0, \\ Q_{\mathrm{u}}(t) & \text{if } z > 0. \end{cases}$$

The velocity v, associated with hindrance and compression phenomena, is given by

$$v = v(X, \partial_z X, z) \coloneqq \gamma(z) v_{\mathrm{hs}}(X) \left(1 - \frac{\rho_X \sigma_{\mathrm{e}}'(X)}{X g \Delta\rho} \partial_z X\right). \tag{8}$$

Here, $\Delta\rho \coloneqq \rho_X - \rho_L$ denotes the density difference between the solid and liquid phases, g is the acceleration due to gravity, and $v_{\mathrm{hs}} = v_{\mathrm{hs}}(X)$ is the hindered settling velocity [20], which is assumed to be a decreasing function that satisfies $v_{\mathrm{hs}}(X) > 0$ if $X \in [0, X_{\max})$ and $v_{\mathrm{hs}}(X) = 0$ if $X \geq X_{\max}$. Additionally, $\sigma_{\mathrm{e}} = \sigma_{\mathrm{e}}(X)$ is the effective solid stress function that is assumed to satisfy

$$\sigma_{\mathrm{e}}'(X) \coloneqq \frac{\mathrm{d}\sigma_{\mathrm{e}}}{\mathrm{d}X} = \begin{cases} = 0 & \text{if } X \leq X_{\mathrm{c}}, \\ > 0 & \text{if } X > X_{\mathrm{c}}, \end{cases}$$

where X_{c} is a critical concentration above which particles come into contact with each other and sediment compression occurs (see [7, 10]).

2.1.2 Reaction terms

The reaction terms modeling the increase of the $k_{\boldsymbol{C}}$ bacteria and the $k_{\boldsymbol{S}}$ soluble components are stored in the vectors

$$\boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{C}, \boldsymbol{S}) \coloneqq \boldsymbol{\sigma}_{\boldsymbol{C}} \boldsymbol{R}(\boldsymbol{C}, \boldsymbol{S}) \in \mathbb{R}^{k_{\boldsymbol{C}}} \quad \text{and} \quad \boldsymbol{R}_{\boldsymbol{S}}(\boldsymbol{C}, \boldsymbol{S}) \coloneqq \boldsymbol{\sigma}_{\boldsymbol{S}} \boldsymbol{R}(\boldsymbol{C}, \boldsymbol{S}) \in \mathbb{R}^{k_{\boldsymbol{S}}}, \tag{9}$$

respectively, where $\boldsymbol{\sigma_C}$ and $\boldsymbol{\sigma_S}$ are stoichiometric matrices of constant coefficients and $\boldsymbol{R}(\boldsymbol{C},\boldsymbol{S}) \geq \boldsymbol{0}$ is a vector containing the reaction rates, which are assumed to be Lipschitz continuous and bounded functions. It is also assumed that the concentration of water W is neither influenced by nor influences any reaction. The net productions (growth minus decay) of biomass and substrates are given by

$$\tilde{R}_{\boldsymbol{C}}(\boldsymbol{C},\boldsymbol{S}) \coloneqq \sum_{i=1}^{k_{\boldsymbol{C}}} R_{\boldsymbol{C}}^{(i)}(\boldsymbol{C},\boldsymbol{S}) \quad \text{and} \quad \tilde{R}_{\boldsymbol{S}}(\boldsymbol{C},\boldsymbol{S}) \coloneqq \sum_{j=1}^{k_{\boldsymbol{S}}} R_{\boldsymbol{S}}^{(j)}(\boldsymbol{C},\boldsymbol{S}), \tag{10}$$

respectively. To prevent the numerical solution of the solid particles from exceeding the maximum concentration $X_{\max}$, it is assumed that there exists $\varepsilon > 0$ such that $\boldsymbol{R_C}(\boldsymbol{C},\boldsymbol{S}) = \boldsymbol{0}$ if $X \geq X_{\max} - \varepsilon$, hence when X is close to $X_{\max}$, the biomass cannot grow further. Additionally, it is assumed that there is no bacterial growth when they are not present, that is, $\boldsymbol{R_C}(\boldsymbol{0},\boldsymbol{S}) = \boldsymbol{0}$. However, due to bacterial decay, growth of soluble components is allowed even in their absence, that is, $\boldsymbol{R_S}(\boldsymbol{C},\boldsymbol{0}) \geq \boldsymbol{0}$. To ensure positivity of the k-th component of the concentration vector $\boldsymbol{C}$, the sets

$$I_{\boldsymbol{C},k}^{-} \coloneqq \left\{ l \in \mathbb{N} : \sigma_{\boldsymbol{C}}^{(k,l)} < 0 \right\} \quad \text{and} \quad I_{\boldsymbol{C},k}^{+} \coloneqq \left\{ l \in \mathbb{N} : \sigma_{\boldsymbol{C}}^{(k,l)} > 0 \right\} \tag{11}$$

are defined, which denote the indices l with negative and positive stoichiometric coefficients respectively, and it is assumed that

$$\text{if } l \in I_{\boldsymbol{C},k}^{-}, \text{ then } r^{(l)}(\boldsymbol{C},\boldsymbol{S}) = \bar{r}^{(l)}(\boldsymbol{C},\boldsymbol{S}) C^{(k)} \text{ with } \bar{r}^{(l)} \text{ bounded.} \tag{12}$$

The last assumption about the reactions is necessary to have a physically correct model with non-negative concentrations. The assumption implies that if a component is consumed ($\sigma_{\boldsymbol{C}}^{(k,l)} < 0$) and its concentration reaches zero, then further consumption is not physically possible. Analogous sets and assumptions to those given in (11) and (12) are also considered for each component of the soluble concentrations vector $\boldsymbol{S}$. For more details on positivity of concentrations we refer to [16].

2.1.3 Balance equations and model equations in final form

The mass balance for each solid and soluble component leads to the system of PDEs

$$\begin{aligned}
&A(z)\partial_t C^{(i)} + \partial_z\left(A(z) v_X C^{(i)}\right) \\
&= \delta(z) C_{\mathrm{f}}^{(i)}(t) Q_{\mathrm{f}}(t) + \gamma(z) A(z) R_{\boldsymbol{C}}^{(i)}(\boldsymbol{C},\boldsymbol{S}), \quad i = 1,\ldots,k_{\boldsymbol{C}}, \\
&A(z)\partial_t S^{(j)} + \partial_z\left(A(z) v_L S^{(j)}\right) \\
&= \delta(z) S_{\mathrm{f}}^{(j)}(t) Q_{\mathrm{f}}(t) + \gamma(z) A(z) R_{\boldsymbol{S}}^{(j)}(\boldsymbol{C},\boldsymbol{S}), \quad j = 1,\ldots,k_{\boldsymbol{S}}.
\end{aligned} \tag{13}$$

Here v_X and v_L are the phase velocities defined in (7). The first term on the right-hand side represents the feed mechanism and the second term describes the reactions. These are assumed to occur only inside the tank, as is expressed by the factor $\gamma(z)$.

The term inside the spatial derivative represents the total flux of particles. The nonlinearity of the system arises from the function v appearing in the phase velocities, and more specifically, from the functional form (8).

Using the concentration vectors $\boldsymbol{C}$ and $\boldsymbol{S}$ and the reaction vectors $\boldsymbol{R_C}(\boldsymbol{C},\boldsymbol{S})$ and $\boldsymbol{R_S}(\boldsymbol{C},\boldsymbol{S})$ from (9), we may rewrite (13) as

$$A(z)\partial_t \boldsymbol{C} + \partial_z\big(A(z)v_X\boldsymbol{C}\big) = \delta(z)\boldsymbol{C}_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\boldsymbol{R_C}(\boldsymbol{C},\boldsymbol{S}), \tag{14a}$$
$$A(z)\partial_t \boldsymbol{S} + \partial_z\big(A(z)v_L\boldsymbol{S}\big) = \delta(z)\boldsymbol{S}_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\boldsymbol{R_S}(\boldsymbol{C},\boldsymbol{S}). \tag{14b}$$

Defining the total fluxes

$$\begin{aligned}\boldsymbol{\Phi_C} &= \boldsymbol{\Phi_C}(\boldsymbol{C}, X, \partial_z X, z, t) \coloneqq A(z)v_X\boldsymbol{C} \quad \text{and}\\ \boldsymbol{\Phi_S} &= \boldsymbol{\Phi_S}(\boldsymbol{S}, X, \partial_z X, z, t) \coloneqq A(z)v_L\boldsymbol{S},\end{aligned}$$

we may rewrite the model (14) as

$$\begin{aligned}A(z)\partial_t \boldsymbol{C} + \partial_z\boldsymbol{\Phi_C} &= \delta(z)\boldsymbol{C}_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\boldsymbol{R_C}(\boldsymbol{C},\boldsymbol{S}),\\ A(z)\partial_t \boldsymbol{S} + \partial_z\boldsymbol{\Phi_S} &= \delta(z)\boldsymbol{S}_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\boldsymbol{R_S}(\boldsymbol{C},\boldsymbol{S}).\end{aligned} \tag{15}$$

Let us comment on the model (15) and its ingredients. First of all, if we define

$$d(X) \coloneqq \frac{\rho_X v_{\mathrm{hs}}(X)\sigma_{\mathrm{e}}'(X)}{Xg\Delta\rho} \quad \text{and} \quad D(X) \coloneqq \int_{X_{\mathrm{c}}}^{X} d(\xi)\,\mathrm{d}\xi, \tag{16}$$

then from (7) and (8) the phase velocities are obtained as

$$\begin{aligned}v_X &= v_X(X, \partial_z X, z, t) = q(z,t) + \gamma(z)\big(v_{\mathrm{hs}}(X) - \partial_z D(X)\big) \quad \text{and}\\ v_L &= v_L(X, \partial_z X, z, t) = q(z,t) - \frac{X}{\rho_X - X}\gamma(z)\big(v_{\mathrm{hs}}(X) - \partial_z D(X)\big).\end{aligned}$$

Furthermore, the model (15) can be expressed in the form (1) if we define the convective fluxes of the solid and liquid phases, respectively, as

$$\begin{aligned}\boldsymbol{\mathcal{U}_C}(X,z,t) &\coloneqq q(z,t) + \gamma(z)v_{\mathrm{hs}}(X) \quad \text{and}\\ \boldsymbol{\mathcal{U}_S}(X,z,t) &\coloneqq \big(\rho_X q(z,t) - (q(z,t) + \gamma(z)v_{\mathrm{hs}}(X))\,X\big)/(\rho_X - X).\end{aligned}$$

Since $D = 0$ on $[0, X_{\mathrm{c}}]$, the model (15) is first-order hyperbolic if $X \le X_{\mathrm{c}}$ and second-order parabolic if $X_{\mathrm{c}} < X < X_{\max}$. This makes the model termed as *strongly degenerate*, and its solutions are in general discontinous. On the other hand, in comparison with the reactive models of [1, 4], the model (15) does not include hydrodynamic dispersion or mixing near the feed inlet. However, the inclusion of such effects in the model did not significantly improve its predictive capability [4].

2.2 Reactive model with percentage vector

The aim now is to use the ideas from [5, 8, 12] to describe the solid phase using percentage vectors. The description of the liquid phase, on the other hand, is the same as given in the previous section. The unknowns for the solid particles are the total concentration of solids X and the percentage vector $\boldsymbol{p}$. The mass balance for the solid phase, now described in terms of percentages, leads to the system of PDEs

$$\begin{aligned} &\partial_t\big(A(z)\boldsymbol{p}X\big) + \partial_z\big(A(z)\boldsymbol{p}Xv_X\big) \\ &= \delta(z)\boldsymbol{p}_{\mathrm{f}}(t)X_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{p}X, \boldsymbol{S}), \end{aligned} \tag{17}$$

where the components of $\boldsymbol{p}$ satisfy (6). Finally, we require that the percentage vectors of the initial data, $\boldsymbol{p}(z, 0)$, and of the feed, $\boldsymbol{p}_{\mathrm{f}}(t)$, also satisfy (5) and (6).

Lemma 1 *The conservation equation* (14a) *for the solid phase is equivalent to Equations* (17) *and* (6)*.*

Proof. The result follows from (5) and its application to the feed flow. □

If we define the flux

$$F_X = F_X(X, \partial_z X, z, t) \coloneqq Xv_X, \tag{18}$$

then Equation (17) becomes

$$\partial_t\big(A(z)\boldsymbol{p}X\big) + \partial_z\big(A(z)\boldsymbol{p}F_X\big) = \delta(z)\boldsymbol{p}_{\mathrm{f}}(t)X_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{p}X, \boldsymbol{S}). \tag{19}$$

Lemma 2 *Equations* (19) *and* (6) *are equivalent to* (19) *and*

$$A(z)\partial_t X + \partial_z\big(A(z)F_X\big) = \delta(z)X_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\tilde{R}_{\boldsymbol{C}}(\boldsymbol{p}X, \boldsymbol{S}),$$

where $\tilde{R}_{\boldsymbol{C}}$ *is the net biomass production defined by* (10)*.*

Proof. See the proof of [8, Lemma 2.2]. □

We are now ready to state in final form the governing equations of the reactive model with percentage vector.

Theorem 1 *The model* (14) *is equivalent to the system*

$$A(z)\partial_t X + \partial_z\big(A(z)F_X\big) = \delta(z)X_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\tilde{R}_{\boldsymbol{C}}(\boldsymbol{p}X, \boldsymbol{S}), \tag{20a}$$

$$\begin{aligned} \partial_t\big(A(z)\boldsymbol{p}X\big) + \partial_z\big(A(z)\boldsymbol{p}F_X\big) &= \delta(z)\boldsymbol{p}_{\mathrm{f}}(t)X_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) \\ &\quad + \gamma(z)A(z)\boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{p}X, \boldsymbol{S}), \end{aligned} \tag{20b}$$

$$A(z)\partial_t \boldsymbol{S} + \partial_z\big(A(z)v_L\boldsymbol{S}\big) = \delta(z)\boldsymbol{S}_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\boldsymbol{R}_{\boldsymbol{S}}(\boldsymbol{p}X, \boldsymbol{S}) \tag{20c}$$

defined for $z \in \mathbb{R}$ *and* $t \geq 0$*.*

Proof. All that needs to be verified is the equivalence between (14a) and (20a) plus (20b), which follows directly from Lemmas 1 and 2. □

We comment that if we define

$$f_{\mathrm{bk}}(X) \coloneqq X v_{\mathrm{hs}}(X), \quad d(X) \coloneqq \rho_X v_{\mathrm{hs}}(X)\sigma_{\mathrm{e}}'(X)/(g\Delta\rho), \tag{21}$$

and D as in (16), the mass flux of the solid phase given in (18) becomes

$$F_X(X, \partial_z X, z, t) = Xq(z,t) + \gamma(z)\big(f_{\mathrm{bk}}(X) - \partial_z D(X)\big). \tag{22}$$

If we define the convective flux of solid particles $\mathcal{F}(X,z,t) \coloneqq Xq(z,t)+\gamma(z)f_{\mathrm{bk}}(X)$, then (20a) and (20b) can be written as the system of PDEs

$$\begin{aligned}
&A(z)\partial_t X + \partial_z\big(A(z)\mathcal{F}(X,z,t)\big)\\
&= \partial_z\left(\gamma(z)A(z)\partial_z D(X)\right) + \delta(z)X_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\tilde{R}_{\boldsymbol{C}}(\boldsymbol{p}X, \boldsymbol{S}),\\
&A(z)\partial_t(\boldsymbol{p}X) + \partial_z\big(A(z)\boldsymbol{p}\mathcal{F}(X,z,t)\big)\\
&= \partial_z\left(\gamma(z)A(z)\boldsymbol{p}\partial_z D(X)\right) + \delta(z)\boldsymbol{p}_{\mathrm{f}}(t)X_{\mathrm{f}}(t)Q_{\mathrm{f}}(t) + \gamma(z)A(z)\boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{p}X, \boldsymbol{S}).
\end{aligned}$$

Furthermore, we mention that not all equations in (20b) need to be solved. The first $k_{\boldsymbol{C}} - 1$ equations could be solved followed by setting $p^{(k_{\boldsymbol{C}})} = 1 - \sum_{i=1}^{k_{\boldsymbol{C}}-1} p^{(i)}$.

Finally, and as mentioned in [5], by using the model (20) we transition from a system containing $k_{\boldsymbol{C}}$ coupled nonlinear and strongly degenerate equations, model (14a), to one that has a single nonlinear scalar equation for X, namely (20a), plus $k_{\boldsymbol{C}}$ equations for the components of $\boldsymbol{p}$ (20b). The key point here is that if X is known, then the convection-diffusion part of (20b) is linear. This observation suggests a numerical scheme in which, at each time step, X is first updated by numerically solving (20a), and then the updated value of X is used to numerically solve (20b).

3 Numerical schemes

In this section, two numerical methods are developed to solve the model (20): methods XPE and XPSI. The nomenclature indicates that the solid phase is described in terms of percentages (XP). The difference lies in the fact that the former is explicit (E) and the latter is semi-implicit (SI).

3.1 Spatial discretization

The SST is divided into N internal layers, or cells, with depth $\Delta z = (B+H)/N$. The midpoint of layer j has coordinate z_j, thus, the layer is the interval $[z_{j-\frac{1}{2}}, z_{j+\frac{1}{2}}]$. Layer 1, located at the top of the clarification zone, corresponds to the interval

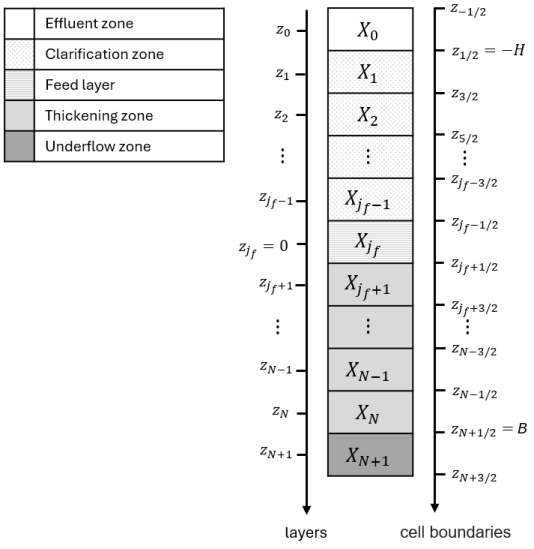

Fig. 2 Discretization of the computational domain into cells. We define j_f as the smallest integer greater than or equal to $H/\Delta z$, i.e., $j_\mathrm{f} \coloneqq \lceil H/\Delta z \rceil$. The feed inlet, $z = 0$, is located in layer j_f (the feed layer). An additional layer is added at each end of the tank for effluent and discharge concentrations: layers 0 and $N + 1$, respectively

$[z_{\frac{1}{2}}, z_{\frac{3}{2}}] = [-H, -H + \Delta z]$. Layer N, situated at the bottom of the thickening zone, corresponds to the interval $[z_{N-\frac{1}{2}}, z_{N+\frac{1}{2}}] = [B - \Delta z, B]$. Thus, the computational domain consists of $N + 2$ layers and $N + 3$ interfaces (boundaries between cells), on which it is necessary to define the numerical fluxes, see Figure 2.

The average values of the unknowns in layer j are denoted by $X_j = X_j(t)$, $\boldsymbol{p}_j = \boldsymbol{p}_j(t)$ and $\boldsymbol{S}_j = \boldsymbol{S}_j(t)$, which are approximations of $X(z_j, t)$, $\boldsymbol{p}(z_j, t)$ and $\boldsymbol{S}(z_j, t)$ respectively. In particular, the concentrations in the effluent and discharge are defined by $X_\mathrm{e}(t) \coloneqq X_0(t)$ and $X_\mathrm{u}(t) \coloneqq X_{N+1}(t)$, and similarly for $\boldsymbol{p}$ and $\boldsymbol{S}$. Two external variables appear in the formulas of the numerical schemes; however, their values are irrelevant since they do not influence any value in any other cell. Therefore, we take $X_{-1} = 0$, $X_{N+2} = 0$, and similarly for $\boldsymbol{p}$ and $\boldsymbol{S}$.

The cross-sectional area of the SST is approximated by

$$A_{j+\frac{1}{2}} \coloneqq \frac{1}{\Delta z} \int_{z_j}^{z_{j+1}} A(\xi)\, \mathrm{d}\xi \quad \text{and} \quad A_j \coloneqq \frac{1}{\Delta z} \int_{z_{j-\frac{1}{2}}}^{z_{j+\frac{1}{2}}} A(\xi)\, \mathrm{d}\xi.$$

Finally, we define

$$\begin{aligned} \delta_{j,j_\mathrm{f}} &\coloneqq \begin{cases} 0 & \text{if } j \neq j_\mathrm{f}, \\ 1 & \text{if } j = j_\mathrm{f}, \end{cases} \qquad \gamma_j \coloneqq \begin{cases} 0 & \text{if } j = 0 \text{ or } j = N+1, \\ 1 & \text{if } j = 1, \ldots, N, \end{cases} \\ \gamma_{j+\frac{1}{2}} &\coloneqq \begin{cases} 0 & \text{if } j \in \{-1, 0, N, N+1\}, \\ 1 & \text{if } j = 1, \ldots, N-1. \end{cases} \end{aligned} \tag{23}$$

3.2 Numerical fluxes and method-of-lines formulation

The total mass fluxes in (20) are approximated at each cell boundary $z_{j+\frac{1}{2}}$, $j \in \mathcal{I}_{\mathrm{i}} \coloneqq \{-1, \dots, N+1\}$. To do this, the flux of the solid phase F_X that appears in (22) will first be approximated using the functions f_{bk} and d given in (21). To simplify the notation, we define $a^+ \coloneqq \max\{a, 0\}$, $a^- \coloneqq \min\{a, 0\}$ and the upwind operator $\mathrm{upw}(a, b, c) \coloneqq \max\{a, 0\}b + \min\{a, 0\}c = a^+ b + a^- c$ for $a, b, c \in \mathbb{R}$.

The notation $\gamma_{j+\frac{1}{2}} \coloneqq \gamma(z_{j+\frac{1}{2}})$ will be used (analogously for the rest of the variables). The volume average velocity q is approximated by

$$q_{j+\frac{1}{2}}(t) \coloneqq \begin{cases} (Q_{\mathrm{u}}(t) - Q_{\mathrm{f}}(t))/A_{j+\frac{1}{2}} & \text{if } j \le j_{\mathrm{f}}, \\ Q_{\mathrm{u}}(t)/A_{j+\frac{1}{2}} & \text{if } j > j_{\mathrm{f}}. \end{cases}$$

For the linear term qX the upwind operator is used: $\mathcal{B}_{j+\frac{1}{2}} \coloneqq \mathrm{upw}(q_{j+\frac{1}{2}}, X_j, X_{j+1})$, and for the nonlinear term $f_{\mathrm{bk}}(X)$ the Engquist-Osher numerical flux [14]

$$\mathcal{E}_{j+\frac{1}{2}} \coloneqq \gamma_{j+\frac{1}{2}} \left(f_{\mathrm{bk}}(0) + \int_0^{X_j} \max\{0, f_{\mathrm{bk}}'(s)\} \, \mathrm{d}s + \int_0^{X_{j+1}} \min\{0, f_{\mathrm{bk}}'(s)\} \, \mathrm{d}s \right) \quad (24)$$

is employed. If f_{bk} has a unique maximum at $X^* \in (0, X_{\max})$, then

$$\mathcal{E}_{j+\frac{1}{2}} = \gamma_{j+\frac{1}{2}} \begin{cases} f_{\mathrm{bk}}(X_j) & \text{if } X_j, X_{j+1} \le X^*, \\ f_{\mathrm{bk}}(X^*) & \text{if } X_{j+1} \le X^* < X_j, \\ f_{\mathrm{bk}}(X_j) + f_{\mathrm{bk}}(X_{j+1}) - f_{\mathrm{bk}}(X^*) & \text{if } X_j \le X^* < X_{j+1}, \\ f_{\mathrm{bk}}(X_{j+1}) & \text{if } X^* < X_j, X_{j+1}. \end{cases}$$

Therefore, the convective flux $\mathcal{F}(X, z, t)$ is approximated by $\mathcal{F}_{j+\frac{1}{2}} \coloneqq \mathcal{B}_{j+\frac{1}{2}} + \mathcal{E}_{j+\frac{1}{2}}$. The diffusive term, on the other hand, is approximated by

$$\mathcal{J}_{j+\frac{1}{2}} \coloneqq \gamma_{j+\frac{1}{2}} \big(D(X_{j+1}) - D(X_j) \big) / \Delta z. \quad (25)$$

For the numerical implementation of $D(X_j)$, see [9]. The mass flux of solid particles F_X (22) is now given by $F^X_{j+\frac{1}{2}} \coloneqq \mathcal{F}_{j+\frac{1}{2}} - \mathcal{J}_{j+\frac{1}{2}}$. In this way, the total mass fluxes in (20) are approximated at each cell boundary $z_{j+\frac{1}{2}}$, with $j \in \mathcal{I}_{\mathrm{i}}$, by

$$\begin{aligned} \Phi^X_{j+\frac{1}{2}} &\coloneqq (A(z)F_X)_{j+\frac{1}{2}} \coloneqq A_{j+\frac{1}{2}} F^X_{j+\frac{1}{2}}, \\ \Phi^{\boldsymbol{p}}_{j+\frac{1}{2}} &\coloneqq (A(z)\boldsymbol{p}F_X)_{j+\frac{1}{2}} \coloneqq A_{j+\frac{1}{2}} \, \mathrm{upw}(F^X_{j+\frac{1}{2}}, \boldsymbol{p}_j, \boldsymbol{p}_{j+1}), \\ \Phi^{\boldsymbol{S}}_{j+\frac{1}{2}} &\coloneqq (A(z)v_L\boldsymbol{S})_{j+\frac{1}{2}} \coloneqq A_{j+\frac{1}{2}} \, \mathrm{upw} \left(\rho_X q_{j+\frac{1}{2}} - F^X_{j+\frac{1}{2}}, \frac{\boldsymbol{S}_j}{\rho_X - X_j}, \frac{\boldsymbol{S}_{j+1}}{\rho_X - X_{j+1}} \right). \end{aligned}$$

With the notation $[\Delta\Phi]_j \coloneqq \Phi_{j+\frac{1}{2}} - \Phi_{j-\frac{1}{2}}$, the mass conservation law applied to layer $j \in \mathcal{I}_{\mathrm{c}} \coloneqq \{0, \dots, N+1\}$, leads to the method-of-lines equations

$$\begin{aligned}
\frac{\mathrm{d}X_j}{\mathrm{d}t} &= -\frac{[\Delta\Phi_X]_j}{A_j\Delta z} + \delta_{j,j_\mathrm{f}}\frac{X_\mathrm{f}Q_\mathrm{f}}{A_j\Delta z} + \gamma_j\tilde{\boldsymbol{R}}_{\boldsymbol{C},j},\\
\frac{\mathrm{d}(\boldsymbol{p}_jX_j)}{\mathrm{d}t} &= -\frac{[\Delta\Phi_{\boldsymbol{p}}]_j}{A_j\Delta z} + \delta_{j,j_\mathrm{f}}\frac{\boldsymbol{p}_\mathrm{f}X_\mathrm{f}Q_\mathrm{f}}{A_j\Delta z} + \gamma_j\boldsymbol{R}_{\boldsymbol{C},j},\\
\frac{\mathrm{d}\boldsymbol{S}_j}{\mathrm{d}t} &= -\frac{[\Delta\Phi_{\boldsymbol{S}}]_j}{A_j\Delta z} + \delta_{j,j_\mathrm{f}}\frac{\boldsymbol{S}_\mathrm{f}Q_\mathrm{f}}{A_j\Delta z} + \gamma_j\boldsymbol{R}_{\boldsymbol{S},j}.
\end{aligned} \tag{26}$$

3.3 Explicit method

Let T be the simulation time, $t_n = n\Delta t$, $n = 1,\dots,N_T$, and $\Delta t \coloneqq T/N_T$ the time step size, which must satisfy a certain CFL condition according to the chosen temporal integration method. The value of a variable at time t_n is denoted by a superscript, e.g., $X_j(t_n) =: X_j^n$. Approximating the time derivatives in (26) by

$$\frac{\mathrm{d}X_j}{\mathrm{d}t}(t_n) \approx \frac{X_j^{n+1}-X_j^n}{\Delta t}, \quad \frac{\mathrm{d}(\boldsymbol{p}_jX_j)}{\mathrm{d}t}(t_n) \approx \frac{\boldsymbol{p}_j^{n+1}X_j^{n+1}-\boldsymbol{p}_j^nX_j^n}{\Delta t}, \tag{27}$$

and similarly for $\boldsymbol{S}_j$, substituting (27) into (26), denoting $\lambda \coloneqq \Delta t/\Delta z$, evaluating the right-hand sides at $t = t_n$ (as corresponding to an explicit Euler discretization in time) and rearranging terms, we obtain for $j \in \mathcal{I}_\mathrm{c}$

$$X_j^{n+1} = X_j^n - (\lambda/A_j)\big([\Delta(A\mathcal{F})]_j^n - [\Delta(A\mathcal{J})]_j^n - \delta_{j,j_\mathrm{f}}X_\mathrm{f}^nQ_\mathrm{f}^n\big) + \gamma_j\Delta t\tilde{\boldsymbol{R}}_{\boldsymbol{C},j}^n, \tag{28a}$$

$$\boldsymbol{p}_j^{n+1}X_j^{n+1} = \boldsymbol{p}_j^nX_j^n - (\lambda/A_j)[\Delta\Phi_{\boldsymbol{p}}]_j^n + (\lambda/A_j)\delta_{j,j_\mathrm{f}}\boldsymbol{p}_\mathrm{f}^nX_\mathrm{f}^nQ_\mathrm{f}^n + \gamma_j\Delta t\boldsymbol{R}_{\boldsymbol{C},j}^n, \tag{28b}$$

$$\boldsymbol{S}_j^{n+1} = \boldsymbol{S}_j^n - (\lambda/A_j)[\Delta\Phi_{\boldsymbol{S}}]_j^n + (\lambda/A_j)\delta_{j,j_\mathrm{f}}\boldsymbol{S}_\mathrm{f}^nQ_\mathrm{f}^n + \gamma_j\Delta t\boldsymbol{R}_{\boldsymbol{S},j}^n. \tag{28c}$$

The method (28) is applied sequentially. First, (28a) is solved to find X_j^{n+1} and then (28b) and (28c) are used to calculate $\boldsymbol{p}_j^{n+1}$ and $\boldsymbol{S}_j^{n+1}$, respectively. If $X_j^{n+1} = 0$ (i.e., there are no solid particles in layer j), then the value of $\boldsymbol{p}_j^{n+1}$ is irrelevant; in this case, $\boldsymbol{p}_j^{n+1}$ can be set to $\boldsymbol{p}_j^n$. Since these equations calculate the value of X, $\boldsymbol{p}$ and $\boldsymbol{S}$ at time t_{n+1} from known quantities at time t^n, the method (28) is explicit.

3.4 CFL condition for the explicit method

For the statement of the CFL conditions we define the set

$$\Omega \coloneqq \left\{\boldsymbol{\mathcal{U}} = (X,\boldsymbol{p}^\mathrm{T},\boldsymbol{S}^\mathrm{T})^\mathrm{T} \in \mathbb{R}^{1+k_{\boldsymbol{C}}+k_{\boldsymbol{S}}} : 0 \le X \le X_\mathrm{max},\ \boldsymbol{p} \ge \boldsymbol{0},\ \sum_{l=1}^{k_{\boldsymbol{C}}} p^{(l)} = 1,\ \boldsymbol{S} \ge \boldsymbol{0}\right\}.$$

For the explicit method (28), the CFL condition is

$$\Delta t \left(\max\left\{ \frac{1}{A_{\min}}, \kappa_1 M_A \right\} \frac{\|Q\|}{\Delta z} + \max\{1, \kappa_2\} \left(\|f'_{\mathrm{bk}}\| + \frac{\|d\|}{\Delta z} \right) \frac{M_A}{\Delta z} + \hat{M} \right) \le 1, \quad \text{(CFL)}$$

with the constants (here, $\hat{\xi} = f'_{\mathrm{bk}}$ or $\hat{\xi} = d$, $I_{\boldsymbol{C}} := \{1, \dots, k_{\boldsymbol{C}}\}$, and $I_{\boldsymbol{S}} := \{1, \dots, k_{\boldsymbol{S}}\}$)

$$\hat{M} := \max\left\{ \frac{1}{c} \sup_{\substack{\boldsymbol{U}\in\Omega \\ k\in I_{\boldsymbol{C}}}} \left| \frac{\partial \tilde{R}_{\boldsymbol{C}}}{\partial C^{(k)}} \right|, \frac{1}{c} \sup_{\substack{\boldsymbol{U}\in\Omega \\ k\in I_{\boldsymbol{C}}}} \sum_{l\in I^-_{\boldsymbol{C},k}} |\sigma^{(k,l)}_{\boldsymbol{C}}| \bar{r}^{(l)}_{\boldsymbol{C}}, \sup_{\substack{\boldsymbol{U}\in\Omega \\ k\in I_{\boldsymbol{S}}}} \sum_{l\in I^-_{\boldsymbol{S},k}} |\sigma^{(k,l)}_{\boldsymbol{S}}| \bar{r}^{(l)}_{\boldsymbol{S}} \right\},$$

$$M_A := \max_{j=1,\dots,N} \left\{ \frac{A_{j+\frac{1}{2}} + A_{j-\frac{1}{2}}}{A_j} \right\}, \quad \|\hat{\xi}\| := \max_{0\le X\le X_{\max}} |\hat{\xi}(X)|, \quad \|Q\| := \max_{0\le t\le T} Q_{\mathrm{f}}(t),$$

$$A_{\min} := \min\{A_1, \dots, A_N\}, \quad \kappa_1 := \frac{\rho_X + X_{\max}}{\rho_X - X_{\max}}, \quad \kappa_2 := \frac{X_{\max}}{\rho_X - X_{\max}}.$$

This condition establishes an upper limit for the time step size Δt for a given Δz. The condition (CFL) guarantees that the complete scheme is monotone and satisfies an invariant region property, which implies, respectively, that the method is stable and allows obtaining physically relevant solutions, that is, non-negative concentrations and bounded solid concentrations. Specifically, under the condition (CFL), the solution vector $\boldsymbol{\mathcal{U}} := (X, \boldsymbol{p}^{\mathrm{T}}, \boldsymbol{S}^{\mathrm{T}})^{\mathrm{T}}$ generated by the explicit scheme (28) remains in as long as the initial data is also in Ω. The proofs of these properties are not presented here, but they are similar to the proofs carried out in [5].

3.5 Semi-implicit method

To obtain a semi-implicit method, we evaluate some terms of the right-hand side of (28) at time t_{n+1}, i.e., certain terms are treated implicitly, while others are still computed explicitly. Clearly, three systems of equations need to be solved: a nonlinear one for updating X and two linear ones for updating the vectors $\boldsymbol{p}$ and $\boldsymbol{S}$. For the first case, we will use Newton's method, and for the second one, the Thomas algorithm. (The well-known Thomas algorithm [15, 24], also known as tridiagonal matrix algorithm (TDMA), is a numerical technique used to solve systems of linear equations where the system matrix is tridiagonal. This method leverages the special structure of the tridiagonal matrix to simplify the Gaussian elimination process, significantly reducing the number of operations required to obtain the solution.)

We begin with the update of X. By explicitly writing out some terms of (28a) and evaluating those with the coefficient $\mu := \lambda/\Delta z = \Delta t/\Delta z^2$ at time t_{n+1}, we get

$$\begin{aligned} X_j^{n+1} = X_j^n - (\lambda/A_j)[\Delta(A\mathcal{F})]_j^n + (\lambda/A_j)\delta_{j,j_{\mathrm{f}}} X_{\mathrm{f}}^n Q_{\mathrm{f}}^n + \gamma_j \Delta t \tilde{R}^n_{\boldsymbol{C},j} \\ + (\mu/A_j)\big(A_{j+\frac{1}{2}} \gamma_{j+\frac{1}{2}} \big(D(X_{j+1}^{n+1}) - D(X_j^{n+1}) \big) \\ - A_{j-\frac{1}{2}} \gamma_{j-\frac{1}{2}} \big(D(X_j^{n+1}) - D(X_{j-1}^{n+1}) \big) \big), \quad j \in \mathcal{I}_{\mathrm{c}}. \end{aligned} \tag{29}$$

Since some terms on the right-hand side of (29) are evaluated at time t_{n+1} and others at time t^n, the formula is semi-implicit. If we denote

$$\tilde{X}_j^{n+1} \coloneqq X_j^n - (\lambda/A_j)[\Delta(A\mathcal{F})]_j^n + (\lambda/A_j)\delta_{j,j_\mathrm{f}} X_\mathrm{f}^n Q_\mathrm{f}^n + \gamma_j \Delta t \tilde{R}_{\boldsymbol{C},j}^n, \quad j \in \mathcal{I}_\mathrm{c}, \tag{30}$$

it follows from (29) that

$$\begin{aligned} X_j^{n+1} = \tilde{X}_j^{n+1} + (\mu/A_j)\big(&A_{j+\frac{1}{2}}\gamma_{j+\frac{1}{2}}\big(D(X_{j+1}^{n+1}) - D(X_j^{n+1})\big) \\ &- A_{j-\frac{1}{2}}\gamma_{j-\frac{1}{2}}\big(D(X_j^{n+1}) - D(X_{j-1}^{n+1})\big)\big), \quad j \in \mathcal{I}_\mathrm{c}. \end{aligned} \tag{31}$$

In the outer layers $j = 0$ and $j = N+1$, the characteristic function γ is equal to zero, see (23), therefore, X_0^{n+1} and X_{N+1}^{n+1} can be explicitly calculated from (31):

$$X_0^{n+1} = \tilde{X}_0^{n+1} \quad \text{and} \quad X_{N+1}^{n+1} = \tilde{X}_{N+1}^{n+1}. \tag{32}$$

In the inner layers ($j = 1, \ldots, N$), the values of $X_1^{n+1}, \ldots, X_N^{n+1}$ are found by solving the system of N nonlinear equations

$$\boldsymbol{X}^{n+1} = \tilde{\boldsymbol{X}}^{n+1} + \mu \boldsymbol{M} \boldsymbol{v}(\boldsymbol{X}^{n+1}), \quad \text{where} \tag{33}$$

$$\boldsymbol{X}^{n+1} \coloneqq \begin{pmatrix} X_1^{n+1} \\ \vdots \\ X_N^{n+1} \end{pmatrix}, \quad \tilde{\boldsymbol{X}}^{n+1} \coloneqq \begin{pmatrix} \tilde{X}_1^{n+1} \\ \vdots \\ \tilde{X}_N^{n+1} \end{pmatrix}, \quad \boldsymbol{v}(\boldsymbol{X}^{n+1}) \coloneqq \begin{pmatrix} D(X_1^{n+1}) \\ \vdots \\ D(X_N^{n+1}) \end{pmatrix}, \quad \text{and}$$

$$\boldsymbol{M} \coloneqq \begin{bmatrix} -A_{\frac{3}{2}}/A_1 & A_{\frac{3}{2}}/A_1 & 0 & \cdots & 0 \\ A_{\frac{3}{2}}/A_2 & -(A_{\frac{5}{2}} + A_{\frac{3}{2}})/A_2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & -(A_{N-\frac{1}{2}} + A_{N-\frac{3}{2}})/A_{N-1} & A_{N-\frac{1}{2}}/A_{N-1} \\ 0 & \cdots & 0 & A_{N-\frac{1}{2}}/A_N & -A_{N-\frac{1}{2}}/A_N \end{bmatrix}.$$

In summary, the method (29) can be rewritten as the following two-step scheme:

1. Given X_j^n with $j \in \mathcal{I}_\mathrm{c}$, calculate $\tilde{X}_j^{n+1}$, for $j \in \mathcal{I}_\mathrm{c}$, from (30).
2. Let X_0^{n+1} and X_{N+1}^{n+1} be as in (32). Find $\boldsymbol{X}^{n+1} \coloneqq \big(X_1^{n+1}, \ldots, X_N^{n+1}\big)^\mathrm{T}$ by solving (33).

To elucidate the numerical solution of (33) by the Newton-Raphson method, consider

$$\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_N)^\mathrm{T} : \mathbb{R}^N \longrightarrow \mathbb{R}^N, \quad \boldsymbol{\varphi}(\boldsymbol{X}^{n+1}) \coloneqq \boldsymbol{X}^{n+1} - \tilde{\boldsymbol{X}}^{n+1} - \mu \boldsymbol{M} \boldsymbol{v}(\boldsymbol{X}^{n+1}),$$

where each component $\varphi_j : \mathbb{R}^N \longrightarrow \mathbb{R}$, $j = 1, \ldots, N$, is such that

$$\begin{aligned} \varphi_j(\boldsymbol{X}^{n+1}) = X_j^{n+1} - \tilde{X}_j^{n+1} &- (\mu/A_j)A_{j+\frac{1}{2}}\gamma_{j+\frac{1}{2}}\big(D(X_{j+1}^{n+1}) - D(X_j^{n+1})\big) \\ &+ (\mu/A_j)A_{j-\frac{1}{2}}\gamma_{j-\frac{1}{2}}\big(D(X_j^{n+1}) - D(X_{j-1}^{n+1})\big). \end{aligned} \tag{34}$$

Then solving the nonlinear system (33) is equivalent to solving $\boldsymbol{\varphi}(\boldsymbol{X}^{n+1}) = \mathbf{0}$.

Let $\mathcal{J}_{\boldsymbol{\varphi}}(\boldsymbol{X}^{n+1}) \in \mathbb{R}^{N\times N}$ be the Jacobian matrix of $\boldsymbol{\varphi}$ at $\boldsymbol{X}^{n+1}$, that is,

$$\mathcal{J}_{\boldsymbol{\varphi}}(\boldsymbol{X}^{n+1}) = (\partial\varphi_i/\partial X_j^{n+1})_{1\le i,j\le N}, \tag{35}$$

which by (34) is tridiagonal. Here, the partial derivatives of each component function are evaluated at $\boldsymbol{X}^{n+1}$. Furthermore, since $D'(X) = d(X)$, it follows that

$$\begin{aligned}
\partial\varphi_j/\partial X_{j+1}^{n+1} &= (\mu/A_j)A_{j+\frac{1}{2}}\gamma_{j+\frac{1}{2}}d\big(X_{j+1}^{n+1}\big), \quad j = 1,\dots,N-1,\\
\partial\varphi_j/\partial X_j^{n+1} &= 1-(\mu/A_j)\big(A_{j+\frac{1}{2}}\gamma_{j+\frac{1}{2}} - A_{j-\frac{1}{2}}\gamma_{j-\frac{1}{2}}\big)d\big(X_j^{n+1}\big), \quad j = 1,\dots,N,\\
\partial f_j/\partial X_{j-1}^{n+1} &= (\mu/A_j)A_{j-\frac{1}{2}}\gamma_{j-\frac{1}{2}}d\big(X_{j-1}^{n+1}\big), \quad j = 2,\dots,N.
\end{aligned}$$

Therefore, by substituting these equations into (35), we obtain the Jacobian matrix

$$\mathcal{J}_{\boldsymbol{\varphi}}(\boldsymbol{X}^{n+1}) = \boldsymbol{I}_N + \mu\boldsymbol{M}\mathrm{diag}\big(d(X_1^{n+1}),\dots,d(X_N^{n+1})\big).$$

The Newton-Raphson method applied to the equation $\boldsymbol{\varphi}(\boldsymbol{X}^{n+1}) = \mathbf{0}$ corresponds to first setting $\boldsymbol{u}^{(0)} := \boldsymbol{X}^n$ and then solving the linear system

$$\mathcal{J}_{\boldsymbol{\varphi}}\big(\boldsymbol{u}^{(k)}\big)\boldsymbol{h}^{(k)} = -\boldsymbol{\varphi}\big(\boldsymbol{u}^{(k)}\big), \quad \boldsymbol{u}^{(k+1)} := \boldsymbol{u}^{(k)} + \boldsymbol{h}^{(k)}, \quad k = 0,1,\dots.$$

The iteration starts from $\boldsymbol{u}^{(0)} := \boldsymbol{X}^n$ and continues until a stopping criterion

$$\|\boldsymbol{u}^{(k+1)} - \boldsymbol{u}^{(k)}\|_1/\|\boldsymbol{u}^{(k)}\|_1 < \varepsilon_{\mathrm{tol}}$$

is reached, where $\varepsilon_{\mathrm{tol}} > 0$ is a given tolerance and $\|\cdot\|_1$ denotes the l_1 norm. After convergence, we take $\boldsymbol{X}^{n+1} := \boldsymbol{u}^{(k+1)}$.

We now discuss the update of the percentage vector $\boldsymbol{p}$. A semi-implicit version of (28b) is obtained by evaluating the terms with coefficient μ at time t_{n+1}. This yields

$$\begin{aligned}
\boldsymbol{p}_j^{n+1}X_j^{n+1} &= \boldsymbol{p}_j^nX_j^n + (\lambda/A_j)\delta_{j,j_{\mathrm{f}}}\boldsymbol{p}_{\mathrm{f}}^nX_{\mathrm{f}}^nQ_{\mathrm{f}}^n + \gamma_j\Delta t\boldsymbol{R}_{\boldsymbol{C},j}^n\\
&\quad - (\lambda/A_j)A_{j+\frac{1}{2}}\big(\big(\mathcal{F}_{j+\frac{1}{2}}^n - \mathcal{J}_{j+\frac{1}{2}}^{n+1}\big)^+\boldsymbol{p}_j^{n+1} + \big(\mathcal{F}_{j+\frac{1}{2}}^n - \mathcal{J}_{j+\frac{1}{2}}^{n+1}\big)^-\boldsymbol{p}_{j+1}^{n+1}\big)\\
&\quad + (\lambda/A_j)A_{j-\frac{1}{2}}\big(\big(\mathcal{F}_{j-\frac{1}{2}}^n - \mathcal{J}_{j-\frac{1}{2}}^{n+1}\big)^+\boldsymbol{p}_{j-1}^{n+1} + \big(\mathcal{F}_{j-\frac{1}{2}}^n - \mathcal{J}_{j-\frac{1}{2}}^{n+1}\big)^-\boldsymbol{p}_j^{n+1}\big), \ j\in\mathcal{I}_{\mathrm{c}}.
\end{aligned}$$

If we define

$$\Phi_{j+\frac{1}{2}}^{n,n+1} := \mathcal{F}_{j+\frac{1}{2}}^n - \mathcal{J}_{j+\frac{1}{2}}^{n+1}, \tag{36}$$

then the previous equation becomes

$$\begin{aligned}
\boldsymbol{p}_j^{n+1}X_j^{n+1} &= \boldsymbol{p}_j^nX_j^n + (\lambda/A_j)\delta_{j,j_{\mathrm{f}}}\boldsymbol{p}_{\mathrm{f}}^nX_{\mathrm{f}}^nQ_{\mathrm{f}}^n + \gamma_j\Delta t\boldsymbol{R}_{\boldsymbol{C},j}^n\\
&\quad - (\lambda/A_j)\big(A_{j+\frac{1}{2}}\big(\Phi_{j+\frac{1}{2}}^{n,n+1,+}\boldsymbol{p}_j^{n+1} + \Phi_{j+\frac{1}{2}}^{n,n+1,-}\boldsymbol{p}_{j+1}^{n+1}\big)\\
&\quad\quad - A_{j-\frac{1}{2}}\big(\Phi_{j-\frac{1}{2}}^{n,n+1,+}\boldsymbol{p}_{j-1}^{n+1} + \Phi_{j-\frac{1}{2}}^{n,n+1,-}\boldsymbol{p}_j^{n+1}\big)\big), \quad j\in\mathcal{I}_{\mathrm{c}}.
\end{aligned} \tag{37}$$

Lemma 3 *If $j = 0$, then the flux defined in* (36) *satisfies*

$$\Phi^+_{j+\frac{1}{2}} = 0, \quad \Phi^-_{j+\frac{1}{2}} = -\frac{Q_{\mathrm{e}}(t)}{A_{\frac{1}{2}}} X_1, \quad \Phi^+_{j-\frac{1}{2}} = 0, \quad \text{and} \quad \Phi^-_{j-\frac{1}{2}} = -\frac{Q_{\mathrm{e}}(t)}{A_{\frac{1}{2}}} X_0. \tag{38}$$

For $j = N + 1$ there holds

$$\Phi^+_{j+\frac{1}{2}} = \frac{Q_{\mathrm{u}}(t)}{A_{N+\frac{3}{2}}} X_{N+1}, \quad \Phi^-_{j+\frac{1}{2}} = 0, \quad \Phi^+_{j-\frac{1}{2}} = \frac{Q_{\mathrm{u}}(t)}{A_{N+\frac{1}{2}}} X_N, \quad \text{and} \quad \Phi^-_{j-\frac{1}{2}} = 0. \tag{39}$$

Proof. If $j = 0$, from (23) it follows that $\gamma_{j+\frac{1}{2}} = 0$. Then, from (24) and (25) it follows that $\Phi_{j+\frac{1}{2}} = \Phi_{\frac{1}{2}} = \mathcal{B}_{\frac{1}{2}}$. Therefore,

$$\begin{aligned}\Phi_{j+\frac{1}{2}} &= \mathrm{upw}(q_{\frac{1}{2}}, X_0, X_1) = \max\{q_{\frac{1}{2}}, 0\} X_0 + \min\{q_{\frac{1}{2}}, 0\} X_1 \\ &= \max\big\{-Q_{\mathrm{e}}(t)/A_{\frac{1}{2}}, 0\big\} X_0 + \min\big\{-Q_{\mathrm{e}}(t)/A_{\frac{1}{2}}, 0\big\} X_1 = -\big(Q_{\mathrm{e}}(t)/A_{\frac{1}{2}}\big) X_1,\end{aligned}$$

hence the proof of the first and second identities in (38) follows from

$$\begin{aligned}\Phi^+_{j+\frac{1}{2}} &= \max\big\{-\big(Q_{\mathrm{e}}(t)/A_{\frac{1}{2}}\big) X_1, 0\big\} = 0, \\ \Phi^-_{j+\frac{1}{2}} &= \min\big\{-\big(Q_{\mathrm{e}}(t)/A_{\frac{1}{2}}\big) X_1, 0\big\} = -\big(Q_{\mathrm{e}}(t)/A_{\frac{1}{2}}\big) X_1.\end{aligned} \tag{40}$$

The third and the fourth are demonstrated analogously. If $j = N + 1$, then $\Phi_{j+\frac{1}{2}} = \Phi_{N+\frac{3}{2}} = \mathcal{B}_{N+\frac{3}{2}}$, meaning that

$$\Phi_{j+\frac{1}{2}} = \mathrm{upw}(q_{N+\frac{3}{2}}, X_{N+1}, X_{N+2}) = \max\{q_{N+\frac{3}{2}}, 0\} X_{N+1} + \min\{q_{N+\frac{3}{2}}, 0\} X_{N+2}.$$

Therefore, analogously to (40) we get

$$\begin{aligned}\Phi^+_{j+\frac{1}{2}} &= \max\big\{\big(Q_{\mathrm{u}}(t)/A_{N+\frac{3}{2}}\big) X_{N+1}, 0\big\} = \big(Q_{\mathrm{u}}(t)/A_{N+\frac{3}{2}}\big) X_{N+1}, \\ \Phi^-_{j+\frac{1}{2}} &= \min\big\{\big(Q_{\mathrm{u}}(t)/A_{N+\frac{3}{2}}\big) X_{N+1}, 0\big\} = 0.\end{aligned}$$

This concludes the proof of the first and the second identity in (39), respectively. The proof of the third and the fourth is similar. □

In light of Lemma 3, for $j = 0$ and $j = N + 1$ formula (37) reduces to

$$\begin{aligned}\boldsymbol{p}_0^{n+1} X_0^{n+1} &= \big(1 - (\lambda/A_0) Q_{\mathrm{e}}^n\big) \boldsymbol{p}_0^n X_0^n + (\lambda/A_0) Q_{\mathrm{e}}^n \boldsymbol{p}_1^n X_1^n, \\ \boldsymbol{p}_{N+1}^{n+1} X_{N+1}^{n+1} &= \big(1 - (\lambda/A_{N+1}) Q_{\mathrm{u}}^n\big) \boldsymbol{p}_{N+1}^n X_{N+1}^n + (\lambda/A_{N+1}) Q_{\mathrm{u}}^n \boldsymbol{p}_N^n X_N^n,\end{aligned}$$

that is, the percentages $\boldsymbol{p}_0^{n+1}$ and $\boldsymbol{p}_{N+1}^{n+1}$ are calculated explicitly. For the inner layers ($j = 1, \ldots, N$), if we define

$$\boldsymbol{P}^n \coloneqq \begin{bmatrix} (\boldsymbol{p}_1^n)^{\mathrm{T}} \\ \vdots \\ (\boldsymbol{p}_{j_{\mathrm{f}}}^n)^{\mathrm{T}} \\ \vdots \\ (\boldsymbol{p}_N^n)^{\mathrm{T}} \end{bmatrix}_{N\times k_{\boldsymbol{C}}}, \quad \boldsymbol{W}_{\boldsymbol{p}}^n \coloneqq \begin{bmatrix} \Delta t(\boldsymbol{R}_{\boldsymbol{C},1}^n)^{\mathrm{T}} \\ \vdots \\ \left((\lambda/A_{j_{\mathrm{f}}})\boldsymbol{p}_{\mathrm{f}}^n X_{\mathrm{f}}^n Q_{\mathrm{f}}^n + \Delta t \boldsymbol{R}_{\boldsymbol{C},j_{\mathrm{f}}}^n\right)^{\mathrm{T}} \\ \vdots \\ \Delta t(\boldsymbol{R}_{\boldsymbol{C},N}^n)^{\mathrm{T}} \end{bmatrix}_{N\times k_{\boldsymbol{C}}}$$

and the matrix $\boldsymbol{T}_{\boldsymbol{p}}(\boldsymbol{\Phi}) \in \mathbb{R}^{N\times N}$ by

$$\boldsymbol{T}_{\boldsymbol{p}}(\boldsymbol{\Phi}) = \begin{bmatrix} \frac{A_{\frac{3}{2}}\Phi^+_{\frac{3}{2}} - A_{\frac{1}{2}}\Phi^-_{\frac{1}{2}}}{A_1} & \frac{A_{\frac{3}{2}}}{A_1}\Phi^-_{\frac{3}{2}} & 0 & \cdots & 0 \\ -\frac{A_{\frac{3}{2}}}{A_2}\Phi^+_{\frac{3}{2}} & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \frac{A_{N-\frac{1}{2}}}{A_{N-1}}\Phi^-_{N-\frac{1}{2}} \\ 0 & \cdots & 0 & -\frac{A_{N-\frac{1}{2}}}{A_N}\Phi^+_{N-\frac{1}{2}} & \frac{A_{N+\frac{1}{2}}\Phi^+_{N+\frac{1}{2}} - A_{N-\frac{1}{2}}\Phi^-_{N-\frac{1}{2}}}{A_N} \end{bmatrix},$$

then the problem of finding the percentages $\boldsymbol{p}_1^{n+1}, \ldots, \boldsymbol{p}_N^{n+1}$ translates to solving

$$\boldsymbol{M}_{\boldsymbol{p}}(\boldsymbol{\Phi}^{n,n+1}, \boldsymbol{X}^{n+1})\boldsymbol{P}^{n+1} = \operatorname{diag}(\boldsymbol{X}^n)\boldsymbol{P}^n + \boldsymbol{W}_{\boldsymbol{p}}^n, \tag{41}$$

which is a linear system with $\boldsymbol{M}_{\boldsymbol{p}}(\boldsymbol{\Phi}^{n,n+1}, \boldsymbol{X}^{n+1}) \coloneqq \operatorname{diag}(\boldsymbol{X}^{n+1}) + \lambda \boldsymbol{T}_{\boldsymbol{p}}(\boldsymbol{\Phi}^{n,n+1})$. Here, for a vector $\boldsymbol{X}$, we define $\operatorname{diag}(\boldsymbol{X}) \coloneqq \operatorname{diag}(X_1, \ldots, X_N)$. If $X_j^{n+1} = 0$, the percentage vector $\boldsymbol{p}_j^{n+1}$ is irrelevant and, in this case, it can be defined as: $\boldsymbol{p}_j^{n+1} = \boldsymbol{p}_j^n$.

Finally, we deal with the update of the soluble concentration vector $\boldsymbol{S}$. Similarly to the previous cases, by explicitly writing all the terms of (28c) and evaluating those containing the coefficient μ at time t_{n+1}, we obtain for $j \in \mathcal{I}_{\mathrm{c}}$

$$\begin{aligned} \boldsymbol{S}_j^{n+1} = \boldsymbol{S}_j^n &+ \frac{\lambda}{A_j}\delta_{j,j_{\mathrm{f}}}\boldsymbol{S}_{\mathrm{f}}^n Q_{\mathrm{f}}^n + \gamma_j \Delta t \boldsymbol{R}_{\boldsymbol{S},j}^n \\ &- \frac{\lambda A_{j+\frac{1}{2}}}{A_j}\left(\frac{(\rho_X q^n - \mathcal{F}^n + \mathcal{J}^{n+1})^+_{j+\frac{1}{2}}}{\rho_X - X_j^{n+1}}\boldsymbol{S}_j^{n+1} + \frac{(\rho_X q^n - \mathcal{F}^n + \mathcal{J}^{n+1})^-_{j+\frac{1}{2}}}{\rho_X - X_{j+1}^{n+1}}\boldsymbol{S}_{j+1}^{n+1}\right) \\ &+ \frac{\lambda A_{j-\frac{1}{2}}}{A_j}\left(\frac{(\rho_X q^n - \mathcal{F}^n + \mathcal{J}^{n+1})^+_{j-\frac{1}{2}}}{\rho_X - X_{j-1}^{n+1}}\boldsymbol{S}_{j-1}^{n+1} + \frac{(\rho_X q^n - \mathcal{F}^n + \mathcal{J}^{n+1})^-_{j-\frac{1}{2}}}{\rho_X - X_j^{n+1}}\boldsymbol{S}_j^{n+1}\right). \end{aligned}$$

Denoting

$$\theta_{j+\frac{1}{2}}^{n,n+1} \coloneqq (\rho_X q^n - \mathcal{F}^n + \mathcal{J}^{n+1})_{j+\frac{1}{2}} \quad \text{and} \quad y_j^n \coloneqq 1/\left(\rho_X - X_j^n\right), \tag{42}$$

we may rewrite the above equation as

$$\begin{aligned}\boldsymbol{S}_j^{n+1} &= \boldsymbol{S}_j^n + (\lambda/A_j)\delta_{j,j_\mathrm{f}}\boldsymbol{S}_\mathrm{f}^n Q_\mathrm{f}^n + \gamma_j \Delta t \boldsymbol{R}_{\boldsymbol{S},j}^n \\ &\quad - (\lambda A_{j+\frac{1}{2}}/A_j)\big(\theta_{j+\frac{1}{2}}^{n,n+1,+} y_j^{n+1}\boldsymbol{S}_j^{n+1} + \theta_{j+\frac{1}{2}}^{n,n+1,-} y_{j+1}^{n+1}\boldsymbol{S}_{j+1}^{n+1}\big) \\ &\quad + (\lambda A_{j-\frac{1}{2}}/A_j)\big(\theta_{j-\frac{1}{2}}^{n,n+1,+} y_{j-1}^{n+1}\boldsymbol{S}_{j-1}^{n+1} + \theta_{j-\frac{1}{2}}^{n,n+1,-} y_j^{n+1}\boldsymbol{S}_j^{n+1}\big), \quad j \in \mathcal{I}_\mathrm{c}.\end{aligned} \tag{43}$$

Lemma 4 *If $j = 0$, then the flux defined in* (42) *satisfies*

$$\theta_{j+\frac{1}{2}}^+ = 0,\ \theta_{j+\frac{1}{2}}^- = (X_1 - \rho_X)\frac{Q_\mathrm{e}(t)}{A_{\frac{1}{2}}},\ \theta_{j-\frac{1}{2}}^+ = 0,\ \text{and } \theta_{j-\frac{1}{2}}^- = (X_0 - \rho_X)\frac{Q_\mathrm{e}(t)}{A_{-\frac{1}{2}}}. \tag{44}$$

Furthermore, if $j = N + 1$, then there holds

$$\theta_{j+\frac{1}{2}}^+ = (\rho_X - X_{N+1})\frac{Q_\mathrm{u}(t)}{A_{N+\frac{3}{2}}},\ \theta_{j+\frac{1}{2}}^- = 0,\ \theta_{j-\frac{1}{2}}^+ = (\rho_X - X_N)\frac{Q_\mathrm{u}(t)}{A_{N+\frac{1}{2}}},\ \text{and } \theta_{j-\frac{1}{2}}^- = 0. \tag{45}$$

Proof. If $j = 0$, then $\theta_{j+\frac{1}{2}} = \theta_{\frac{1}{2}} = \rho_X q_{\frac{1}{2}} - \mathcal{B}_{\frac{1}{2}}$. Therefore,

$$\begin{aligned}\theta_{j+\frac{1}{2}} &= \rho_X q_{\frac{1}{2}} - \mathrm{upw}(q_{\frac{1}{2}}, X_0, X_1) = \rho_X q_{\frac{1}{2}} - \max\{q_{\frac{1}{2}}, 0\}X_0 - \min\{q_{\frac{1}{2}}, 0\}X_1 \\ &= -\big(Q_\mathrm{e}(t)/A_{\frac{1}{2}}\big)\rho_X - \max\big\{-Q_\mathrm{e}(t)/A_{\frac{1}{2}}, 0\big\}X_0 - \min\big\{-Q_\mathrm{e}(t)/A_{\frac{1}{2}}, 0\big\}X_1 \\ &= -\big(Q_\mathrm{e}(t)/A_{\frac{1}{2}}\big)\rho_X + \big(Q_\mathrm{e}(t)/A_{\frac{1}{2}}\big)X_1 = (X_1 - \rho_X)Q_\mathrm{e}(t)/A_{\frac{1}{2}}.\end{aligned}$$

Now, since $\rho_X > X_{\max} \geq X_1$, for $j = 0$ we have

$$\begin{aligned}\theta_{j+\frac{1}{2}}^+ &= \max\big\{(X_1 - \rho_X)Q_\mathrm{e}(t)/A_{\frac{1}{2}}, 0\big\} = 0, \\ \theta_{j+\frac{1}{2}}^- &= \min\big\{(X_1 - \rho_X)Q_\mathrm{e}(t)/A_{\frac{1}{2}}, 0\big\} = (X_1 - \rho_X)Q_\mathrm{e}(t)/A_{\frac{1}{2}}.\end{aligned}$$

This proves the first and the second identity in (44). The third and the fourth are proven similarly. If $j = N + 1$, then $\theta_{j+\frac{1}{2}} = \theta_{N+\frac{3}{2}} = \rho_X q_{N+\frac{3}{2}} - \mathcal{B}_{N+\frac{3}{2}}$, hence

$$\begin{aligned}\theta_{j+\frac{1}{2}} &= \rho_X q_{N+\frac{3}{2}} - \mathrm{upw}(q_{N+\frac{3}{2}}, X_{N+1}, X_{N+2}) \\ &= \rho_X q_{N+\frac{3}{2}} - \max\{q_{N+\frac{3}{2}}, 0\}X_{N+1} - \min\{q_{N+\frac{3}{2}}, 0\}X_{N+2} \\ &= (Q_\mathrm{u}(t)/A_{N+\frac{3}{2}})\rho_X - \max\big\{Q_\mathrm{u}(t)/A_{N+\frac{3}{2}}, 0\big\}X_{N+1} \\ &\quad - \min\big\{Q_\mathrm{u}(t)/A_{N+\frac{3}{2}}, 0\big\}X_{N+2} \\ &= \big(Q_\mathrm{u}(t)/A_{N+\frac{3}{2}}\big)\rho_X - \big(Q_\mathrm{u}(t)/A_{N+\frac{3}{2}}\big)X_{N+1} = (\rho_X - X_{N+1})Q_\mathrm{u}(t)/A_{N+\frac{3}{2}}.\end{aligned}$$

Then, as $\rho_X > X_{\max} \geq X_{N+1}$, for $j = N + 1$ we have that

$$\theta_{j+\frac{1}{2}}^+ = \max\big\{(\rho_X - X_{N+1})Q_\mathrm{u}(t)/A_{N+\frac{3}{2}}, 0\big\} = (\rho_X - X_{N+1})Q_\mathrm{u}(t)/A_{N+\frac{3}{2}},$$

$$\theta^-_{j+\frac{1}{2}} = \min\left\{(\rho_X - X_{N+1})Q_{\mathrm{u}}(t)/A_{N+\frac{3}{2}}, 0\right\} = 0.$$

This proves the first and the second identity in (45), respectively. Again, the third and the fourth are proven similarly. □

Lemma 4 implies that in the outer layers $j = 0$ and $j = N + 1$, formula (43) is explicit and reduces to

$$\boldsymbol{S}_0^{n+1} = \left(1 - \frac{\lambda}{A_0}Q_{\mathrm{e}}^n\right)\boldsymbol{S}_0^n + \frac{\lambda}{A_0}Q_{\mathrm{e}}^n\boldsymbol{S}_1^n,\ \boldsymbol{S}_{N+1}^{n+1} = \left(1 - \frac{\lambda}{A_{N+1}}Q_{\mathrm{u}}^n\right)\boldsymbol{S}_{N+1}^n + \frac{\lambda}{A_{N+1}}Q_{\mathrm{u}}^n\boldsymbol{S}_N^n.$$

For the inner layers ($j = 1, \ldots, N$), if we define

$$\boldsymbol{S}^n \coloneqq \begin{bmatrix} (\boldsymbol{S}_1^n)^{\mathrm{T}} \\ \vdots \\ (\boldsymbol{S}_{j_{\mathrm{f}}}^n)^{\mathrm{T}} \\ \vdots \\ (\boldsymbol{S}_N^n)^{\mathrm{T}} \end{bmatrix}_{N\times k_{\boldsymbol{S}}}, \quad \boldsymbol{y}^n \coloneqq \begin{pmatrix} y_1^n \\ \vdots \\ y_{j_{\mathrm{f}}}^n \\ \vdots \\ y_N^n \end{pmatrix}, \quad \boldsymbol{W}_{\boldsymbol{S}}^n \coloneqq \begin{bmatrix} \Delta t(\boldsymbol{R}_{\boldsymbol{S},1}^n)^{\mathrm{T}} \\ \vdots \\ \left((\lambda/A_{j_{\mathrm{f}}})\boldsymbol{S}_{\mathrm{f}}^n Q_{\mathrm{f}}^n + \Delta t\boldsymbol{R}_{\boldsymbol{S},j_{\mathrm{f}}}^n\right)^{\mathrm{T}} \\ \vdots \\ \Delta t(\boldsymbol{R}_{\boldsymbol{S},N}^n)^{\mathrm{T}} \end{bmatrix}_{N\times k_{\boldsymbol{S}}}$$

and the matrix $\boldsymbol{T}_{\boldsymbol{S}}(\boldsymbol{\theta}, \boldsymbol{y}) \in \mathbb{R}^{N\times N}$ by

$$\boldsymbol{T}_{\boldsymbol{S}}(\boldsymbol{\theta}, \boldsymbol{y}) \coloneqq \begin{bmatrix} \frac{A_{\frac{3}{2}}\theta^+_{\frac{3}{2}} - A_{\frac{1}{2}}\theta^-_{\frac{1}{2}}}{A_1}y_1 & \frac{A_{\frac{3}{2}}\theta^-_{\frac{3}{2}}}{A_1}y_2 & 0 & \cdots & 0 \\ -\frac{A_{\frac{3}{2}}\theta^+_{\frac{3}{2}}}{A_2}y_1 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \frac{A_{N-\frac{1}{2}}}{A_{N-1}}\theta^-_{N-\frac{1}{2}}y_N \\ 0 & \cdots & 0 & -\frac{A_{N-\frac{1}{2}}\theta^+_{N-\frac{1}{2}}}{A_N}y_{N-1} & \frac{A_{N+\frac{1}{2}}\theta^+_{N+\frac{1}{2}} - A_{N-\frac{1}{2}}\theta^-_{N-\frac{1}{2}}}{A_N}y_N \end{bmatrix},$$

then the soluble concentrations $\boldsymbol{S}_1^{n+1}, \ldots, \boldsymbol{S}_N^{n+1}$ are obtained by solving

$$\boldsymbol{M}_{\boldsymbol{S}}\left(\boldsymbol{\theta}^{n,n+1}, \boldsymbol{y}^{n+1}\right)\boldsymbol{S}^{n+1} = \boldsymbol{S}^n + \boldsymbol{W}_{\boldsymbol{S}}^n, \tag{46}$$

which is a linear system with $\boldsymbol{M}_{\boldsymbol{S}}\left(\boldsymbol{\theta}^{n,n+1}, \boldsymbol{y}^{n+1}\right) \coloneqq \boldsymbol{I}_N + \lambda\boldsymbol{T}_{\boldsymbol{S}}(\boldsymbol{\theta}^{n,n+1}, \boldsymbol{y}^{n+1})$.

3.6 CFL condition for semi-implicit method

The CFL condition for the semi-implicit method (30)–(33), (41), (46) is

$$\Delta t\left(\frac{\|Q\|}{\Delta z A_{\min}}+\frac{M_A}{\Delta z}\|f'_{\mathrm{bk}}\|+\hat{M}\right)\leq 1. \tag{CFL-SI}$$

As a result, the semi-implicit scheme remains stable as Δz approaches 0, if $\lambda=\Delta t/\Delta z$ is chosen in such a way that condition (CFL-SI) is satisfied. Then, computations can be performed more rapidly with the semi-implicit scheme compared to the explicit scheme, since Δt only needs to be proportional to Δz, not Δz^2 as required by the explicit scheme. Just like in the explicit method (28), the condition (CFL-SI) was obtained from the requirement that the scheme be monotone and satisfy an invariant region property.

4 Numerical examples

For the numerical examples, we employ an SST with variable cross-sectional area whose geometry is given in [4] and a modified ASM1 model (detailed in Appendix A). The unknowns are the concentration vectors $\boldsymbol{C}$ and $\boldsymbol{S}$ specified in (3). The constitutive functions and parameters used are those given in [4], namely

$$v_{\mathrm{hs}}(X):=\frac{v_0}{1+(X/\bar{X})^{\eta}},\quad \sigma_{\mathrm{e}}(X):=\begin{cases}0 & \text{for } X\leq X_{\mathrm{c}},\\ \alpha(X-X_{\mathrm{c}}) & \text{for } X>X_{\mathrm{c}},\end{cases}$$

with $v_0=5.85\,\mathrm{m/h}$, $\bar{X}=1.59\,\mathrm{kg/m^3}$, $\eta=2.19$, $\alpha=261901.56\,\mathrm{m^2/h^2}$ and $X_{\mathrm{c}}=3.2\,\mathrm{kg/m^3}$. These functions are plotted in Figure 3. Other parameters are $\rho_X=1050\,\mathrm{kg/m^3}$, $\Delta\rho=52\,\mathrm{kg/m^3}$ and $g=9.8\,\mathrm{m/s^2}$. We have used $X_{\max}=30\,\mathrm{kg/m^3}$, a value that is never reached in our simulations. The initial and feed concentrations are

$$\boldsymbol{C}_{\mathrm{ini}}:=\begin{pmatrix}650\\ 150-100\\ 800\\ 150\\ 700\\ 100\end{pmatrix},\ \boldsymbol{S}_{\mathrm{ini}}:=\begin{pmatrix}30\\ 12\\ 0.4\\ 6\\ 7.5\\ 5\\ 2.83\end{pmatrix},\ \boldsymbol{C}_{\mathrm{feed}}:=\begin{pmatrix}914.08\\ 40.2-3.3\\ 1489.41\\ 93.45\\ 757.08\\ 3.3\end{pmatrix},\ \boldsymbol{S}_{\mathrm{feed}}:=\begin{pmatrix}17\\ 0.01\\ 5.2\\ 7\\ 0.01\\ 0.01\\ 22\end{pmatrix},$$

whose units are given in Table 2. The flows used correspond to scenario M from [19], namely, $Q_{\mathrm{f}}(t)=0.65\,\mathrm{m^3/h}$ and $Q_{\mathrm{u}}(t)=0.15\,\mathrm{m^3/h}$.

4.1 Example 1: Dynamics of the solid and soluble concentrations

The objective of this example is to use the semi-implicit method to illustrate the variation of the concentrations of solid and soluble components as a function of depth z in a period of $T=24$ hours, time in which the system is approximately

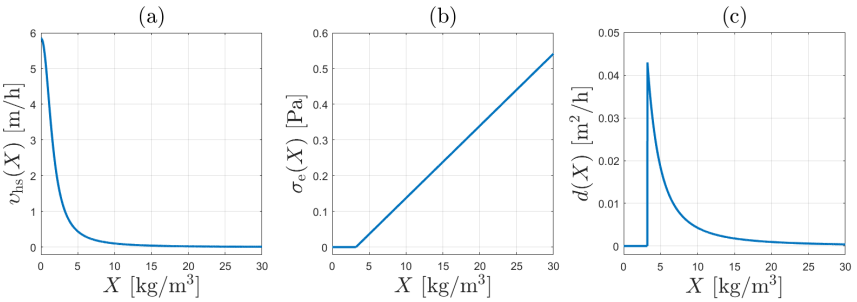

Fig. 3 Graphs of the constitutive functions: (a) hindered settling velocity, (b) effective solids stress, (c) compression (21).

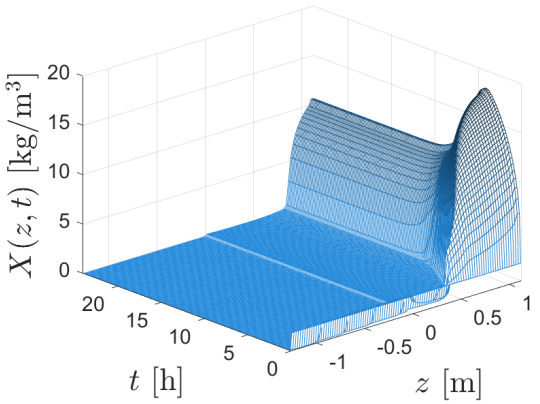

Fig. 4 Example 1: Simulation of the total concentration of suspended solids using the semi-implicit method with $N = 100$ layers and $T = 24$ h

in a steady state. The simulation results obtained by using $N = 100$ layers and consistently with (CFL-SI), $\Delta t = 0.00117$ h, are shown in Figures 4 to 7, which represent the concentration profiles of the total suspended solids, solid particles, soluble components and the concentrations in the underflow zone respectively.

Figure 4 shows that the total solid material quickly settles to the bottom of the tank, reaching a maximum concentration of approximately 17 kg/m^3, which decreases continuously for several hours until reaching a steady concentration of approximately 10.7 kg/m^3. Each of the components comprising the solid phase presents similar dynamics (Figure 5). In Figure 7, it is observed that at approximately $t = 15$ hours, the concentrations in the underflow zone reach a steady state.

4.2 Example 2: Convergence of the semi-implicit method

In this case, the simulations allow us to observe how the numerical solutions obtained by the semi-implicit method converge towards a reference solution as the number of layers that subdivide the tank increases. We consider a subdivision of the tank into

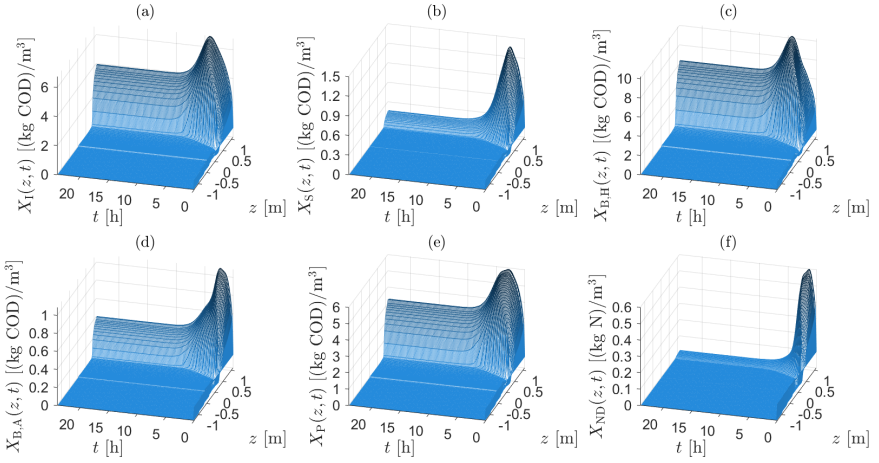

Fig. 5 Example 1: Simulation of the concentrations of solid components using the semi-implicit method with $N = 100$ layers and $T = 24$ h: (a) particulate inert organic matter, (b) slowly biodegradable substrate, (c) active heterotrophic biomass, (d) active autotrophic biomass, (e) particle products arising from biomass decay, (f) particulate biodegradable organic nitrogen

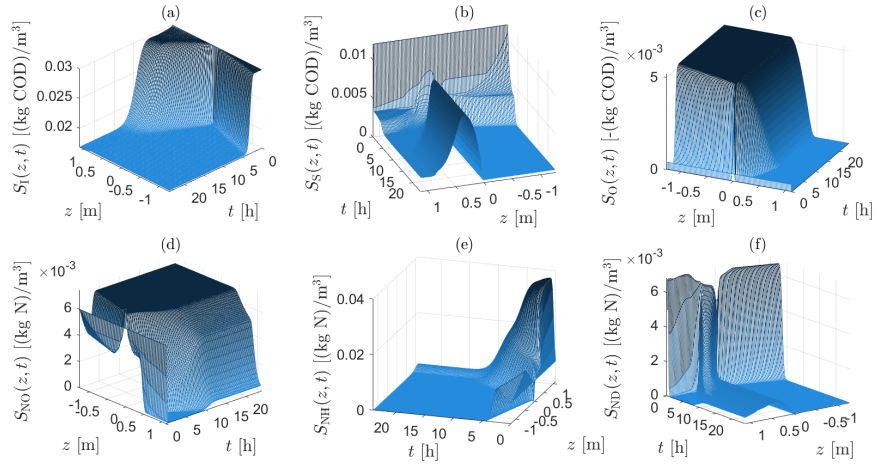

Fig. 6 Example 1: Simulation of the concentrations of soluble components using the semi-implicit method with $N = 100$ layers and $T = 24$ h: (a) soluble inert organic matter, (b) readily biodegradable substrate, (c) oxygen, (d) nitrate and nitrite nitrogen, (e) NH_4^+ + NH_3 nitrogen, (f) soluble biodegradable organic nitrogen

$N = 10$, 30, and 90 layers along with $\Delta t = 0.01107$ h, 0.00408 h, and 0.00131 h, such that (CFL-SI) is satisfied in each case. The reference solution was calculated using the explicit method developed in [1], method CS, with $N = N_{\text{ref}} = 1620$ layers during $T = 1$ h. The results shown in Figures 8 to 10 were obtained with a tolerance of $\varepsilon_{\text{tol}} = 10^{-8}$ in the Newton-Raphson method, a value that was also used in [5]. These

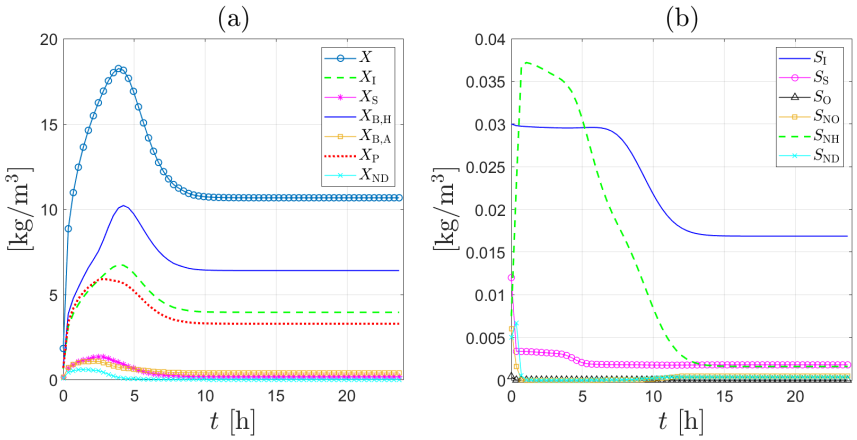

Fig. 7 Example 1: Simulation of concentrations in the underflow zone using the semi-implicit method with $N = 100$ layers and $T = 24$ h: (a) total suspended solids and solid particles, (b) soluble components

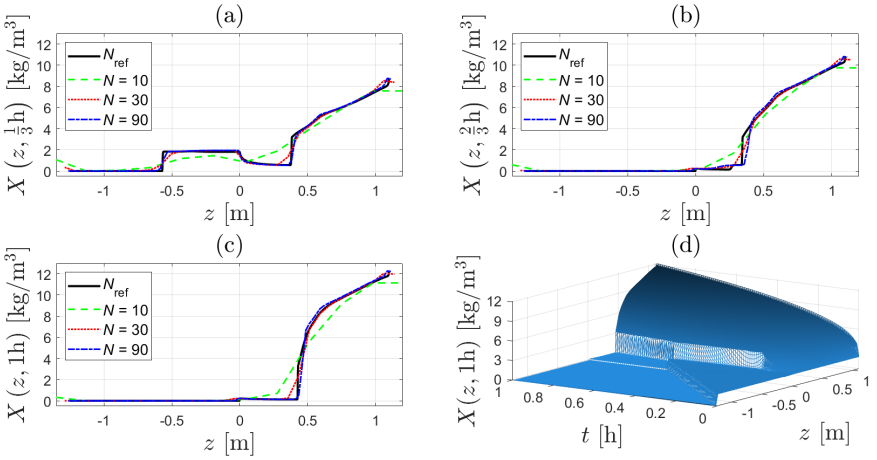

Fig. 8 Example 2: Convergence of the numerical solutions generated with the semi-implicit method ($N = 10, 30, 90$) towards the reference solution ($N = N_{\text{ref}} = 1620$) calculated with the method CS [1]. The graphs depict the concentration of total suspended solids at: (a) $T = \frac{1}{3}$ h, (b) $T = \frac{2}{3}$ h, (c) $T = 1$ h. Plot (d) shows the reference solution for the total suspended solids concentration at $T = 1$ h

graphs show that the numerical solution generated with $N = 10$ layers is far from the reference solution, but when the number of layers increases the approximations clearly converge to it.

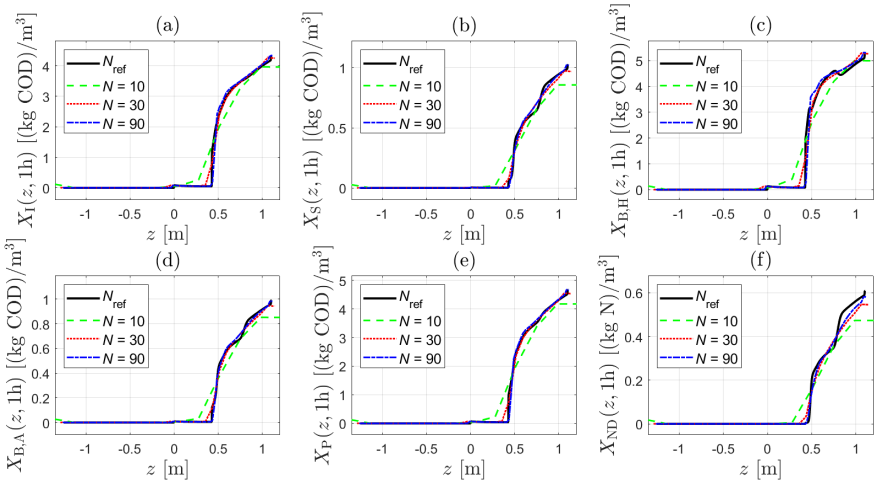

Fig. 9 Example 2: Convergence of the numerical solutions generated with the semi-implicit method ($N = 10, 30, 90$) towards the reference solution ($N = N_{ref} = 1620$) calculated with the method CS [1]. The graphs depict the concentration of each solid particle at $T = 1$ h

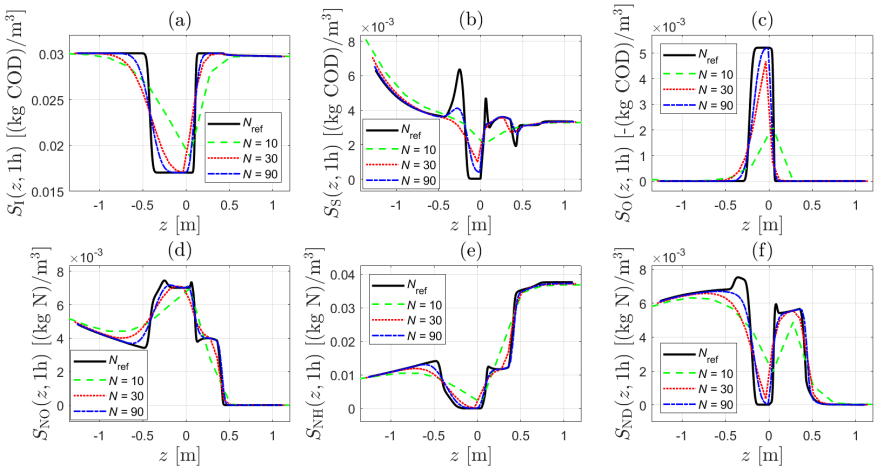

Fig. 10 Example 2: Convergence of the numerical solutions generated with the semi-implicit method ($N = 10, 30, 90$) towards the reference solution ($N = N_{ref} = 1620$) calculated with the method CS [1]. The graphs depict the concentration of each soluble species at $T = 1$ h

4.3 Example 3: Comparison of the explicit and semi-implicit methods

The main purpose of this example is to compare the efficiency of the explicit and semi-implicit methods in terms of computational time (speed) and the relative numerical errors (accuracy) obtained for different values of the spatial discretization parameter N. Accuracy is quantified by measuring the error of the numerical solu-

Table 1 Example 3: CPU times and numerical relative errors

			Method XPE			Method XPSI		
time t	N	Δz [m]	Δt [h]	CPU time [s]	$e_N^{\mathrm{rel}}(t)$	Δt [h]	CPU time [s]	$e_N^{\mathrm{rel}}(t)$
$\frac{1}{3}$ h	10	0.2350	0.01084	0.0904	0.6577	0.01107	0.1290	0.6102
	30	0.0783	0.00385	0.2139	0.1231	0.00408	0.1921	0.1215
	90	0.0261	0.00113	0.9052	0.0223	0.00131	0.6590	0.0232
	180	0.0131	0.00048	6.4898	0.0081	0.00064	4.5239	0.0086
	270	0.0087	0.00028	28.6836	0.0043	0.00042	15.0752	0.0047
	405	0.0058	0.00016	133.0598	0.0023	0.00028	53.0967	0.0026
$\frac{2}{3}$ h	10	0.2350	0.01084	0.1646	0.6263	0.01107	0.2063	0.5908
	30	0.0783	0.00385	0.4237	0.1141	0.00408	0.3095	0.1131
	90	0.0261	0.00113	1.7622	0.0237	0.00131	1.3037	0.0256
	180	0.0131	0.00048	13.3206	0.0099	0.00064	9.2676	0.0110
	270	0.0087	0.00028	63.1219	0.0060	0.00042	30.3464	0.0068
	405	0.0058	0.00016	217.3334	0.0037	0.00028	109.2461	0.0042
1 h	10	0.2350	0.01084	0.1918	0.6687	0.01107	0.2812	0.6445
	30	0.0783	0.00385	0.5187	0.1118	0.00408	0.4175	0.1059
	90	0.0261	0.00113	2.9370	0.0206	0.00131	2.3002	0.0214
	180	0.0131	0.00048	17.4146	0.0081	0.00064	13.8376	0.0088
	270	0.0087	0.00028	85.0397	0.0048	0.00042	45.1463	0.0054
	405	0.0058	0.00016	392.6393	0.0029	0.00028	166.6455	0.0033

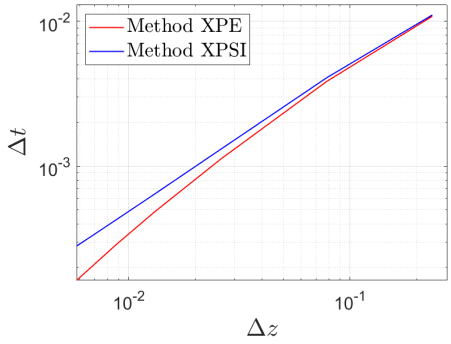

Fig. 11 Example 3: Comparison of the time step size Δt as a function of the parameter Δz for the XPE and XPSI methods

tion compared to the exact solution of the PDE. However, in general, obtaining an exact solution for a dynamic scenario is not feasible, and instead, a high-resolution numerical solution must be used as a reference solution. In our case, as mentioned in Example 2, we will use the explicit method proposed in [1], method CS, to calculate a reference solution with $N = N_{\mathrm{ref}} = 1620$ layers. The approximate relative numerical error, $e_N^{\mathrm{rel}}(t)$, of a numerical solution at a fixed time t is defined by

$$e_N^{\mathrm{rel}}(t) := \sum_{j=1}^{k_{\boldsymbol{C}}} \frac{\|(\boldsymbol{C}_N^{(j)} - \boldsymbol{C}_{N_{\mathrm{ref}}}^{(j)})(\cdot,t)\|_{L^1(-H,B)}}{\|\boldsymbol{C}_{N_{\mathrm{ref}}}^{(j)}(\cdot,t)\|_{L^1(-H,B)}} + \sum_{j=1}^{k_{\boldsymbol{S}}} \frac{\|(\boldsymbol{S}_N^{(j)} - \boldsymbol{S}_{N_{\mathrm{ref}}}^{(j)})(\cdot,t)\|_{L^1(-H,B)}}{\|\boldsymbol{S}_{N_{\mathrm{ref}}}^{(j)}(\cdot,t)\|_{L^1(-H,B)}},$$

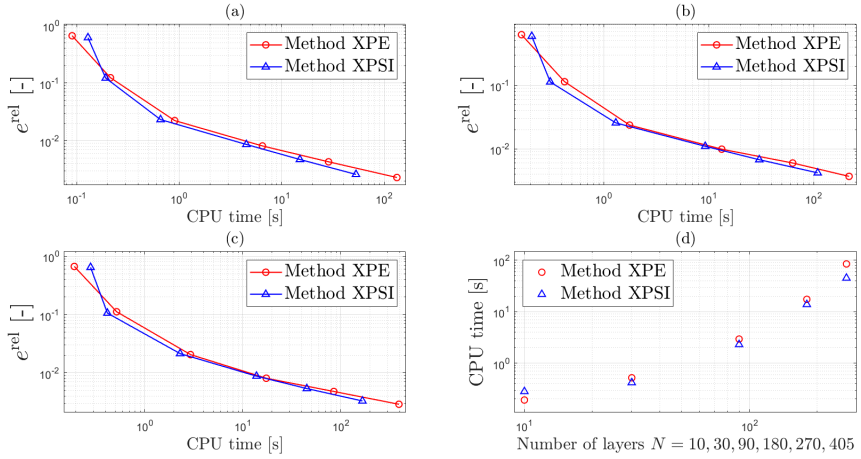

Fig. 12 Example 3: Numerical relative errors versus CPU time for various values of N and different simulation times (Table 1): (a) $T = \frac{1}{3}$ h, (b) $T = \frac{2}{3}$ h, (c) $T = 1$ h. Plot (d) displays the CPU time versus the number of layers N for $T = 1$ hour

where $\boldsymbol{C}_N$ and $\boldsymbol{S}_N$ are the approximate solutions obtained with N internal layers, and $\boldsymbol{C}_{N_{\mathrm{ref}}}$ and $\boldsymbol{S}_{N_{\mathrm{ref}}}$ are the reference solutions.

In Table 1, the errors and CPU times obtained with XPE and XPSI methods are compared using different spatial discretizations ($N = 10, 30, 90, 180, 270, 405$) at three different time points: $T = \frac{1}{3}$ h, $T = \frac{2}{3}$ h and $T = 1$ h. As before, the semi-implicit method was implemented with the tolerance $\epsilon_{\mathrm{tol}} = 10^{-8}$.

Regarding CPU time, the results show that the method XPSI (semi-implicit) is faster than the method XPE (explicit), except for the coarsest discretization ($N = 10$). The observed trend suggests that as resolution increases, (i.e., the larger the value of N), the efficiency of the method XPSI in terms of computational time is greater than that of the method XPE. For example, for a value of $N = 270$, the semi-implicit method is approximately two times faster than its explicit counterpart. This is clearly due to the CFL condition of each method; specifically, given a Δz, the condition (CFL-SI) allows the semi-implicit method to achieve larger time step sizes, see Figure 11. On the other hand, the results also show that the semi-implicit method is more accurate for coarse discretizations ($N = 10, 30$); however, as the spatial domain is refined, the explicit method achieves smaller errors. In any case, it is important to mention that both methods exhibited comparable accuracy in error estimation.

Figures 12 (a), (b), and (c) show that the semi-implicit method achieves an accuracy similar to that of the explicit scheme in less time, so the semi-implicit version is more efficient. In plot (d), the simulation time for each method is plotted as a function of the number of layers. Again, results favor the semi-implicit method.

5 Conclusions

In this work, we have formulated a one-dimensional reactive settling model that simultaneously considers an SST with a variable cross-sectional area and a description with percentages of the solid phase, see (20). Furthermore, two new numerical methods that differ in temporal discretization have been developed and compared: method XPE (explicit) and method XPSI (semi-implicit). In both cases, the Engquist-Osher flux was used to approximate the nonlinear term of the convective flux of solid particles. The numerical results allow us to conclude that for the considered constitutive functions, parameters and initial conditions, the concentrations of solid and soluble particles reach a steady state. Furthermore, the numerical solutions obtained by both methods converge towards the reference solution as the number of layers that subdivide the tank increases, and finally the approximate relative numerical errors made by each method are very similar (Table 1). However, for fine discretizations, the semi-implicit scheme is faster and requires less time to achieve an accuracy similar to that of its explicit counterpart (Figure 12). For $N = 270$ layers, for instance, the semi-implicit method is nearly twice as fast as the explicit scheme. This gain is clearly due to the CFL condition (CFL-SI) of the semi-implicit method, which is more favorable than the CFL condition (CFL) of the explicit method by allowing larger time step sizes. As mentioned in [5], the degree of this gain also depends on the constitutive functions and parameters used, which, in this case, correspond to data fitted to experimental results from a pilot plant [4].

On the other hand, future research should focus on the analysis of the well-posedness of the semi-implicit method, which means establishing results of existence and uniqueness of solutions for the systems of equations associated with the computation of X, $\boldsymbol{p}$, and $\boldsymbol{S}$. For the case of the nonlinear system (33), the well-posedness could be determined using a topological degree argument (as in [10]). Furthermore, although Example 2 demonstrates the numerical convergence of the semi-implicit method towards a reference solution, a rigorous proof of the convergence of the method is still lacking.

Appendix A. A modified ASM1 model

The Activated Sludge Model No.1, known as ASM1 [18], is a mathematical model developed by the International Water Association (IWA) to represent and simulate the biological processes occurring in wastewater treatment plants, specifically in the activated sludge process. The model considers a total of 13 state variables (6 solid and 7 soluble), 8 processes (3 growth, 2 decay, 2 hydrolysis and 1 ammonification) and 19 parameters (5 stoichiometric and 14 kinetic). However, we herein consider a modified ASM1 model, in which only 6 soluble components are taken into account (alkalinity is excluded). Furthermore, due to the reformulation of the PDE model to one that includes percentages for the concentrations of solid particles, the second solid component was redefined from X_{S} to $X_{\mathrm{S-ND}} \coloneqq X_{\mathrm{S}} - X_{\mathrm{ND}}$ (see Tables 2 and 3).

Table 2 State variables in the modified ASM1 model

Material	Symbol	Unit
Particulate inert organic matter	X_{I}	$(\mathrm{g\ COD})\mathrm{m}^{-3}$
Slowly biodegradable substrate	$X_{\mathrm{S-ND}}$	$(\mathrm{g\ COD})\mathrm{m}^{-3}$
Active heterotrophic biomass	$X_{\mathrm{B,H}}$	$(\mathrm{g\ COD})\mathrm{m}^{-3}$
Active autotrophic biomass	$X_{\mathrm{B,A}}$	$(\mathrm{g\ COD})\mathrm{m}^{-3}$
Particulate products of biomass decay	X_{P}	$(\mathrm{g\ COD})\mathrm{m}^{-3}$
Biodegradable organic nitrogen in particles	X_{ND}	$(\mathrm{g\ N})\mathrm{m}^{-3}$
Soluble inert organic matter	S_{I}	$(\mathrm{g\ COD})\mathrm{m}^{-3}$
Easily biodegradable substrate	S_{S}	$(\mathrm{g\ COD})\mathrm{m}^{-3}$
Oxygen	S_{O}	$-(\mathrm{g\ COD})\mathrm{m}^{-3}$
Nitrate and nitrite nitrogen	S_{NO}	$(\mathrm{g\ N})\mathrm{m}^{-3}$
$\mathrm{NH}_4^+ + \mathrm{NH}_3$ nitrogen	S_{NH}	$(\mathrm{g\ N})\mathrm{m}^{-3}$
Soluble biodegradable organic nitrogen	S_{ND}	$(\mathrm{g\ N})\mathrm{m}^{-3}$

With the values of Table 3, the stoichiometric matrices $\boldsymbol{\sigma_C}$ and $\boldsymbol{\sigma_S}$ are given by

$$\boldsymbol{\sigma_C} \coloneqq \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-f_{\mathrm{P}}(1+i_{\mathrm{XP}})-i_{\mathrm{XB}} & 1-f_{\mathrm{P}}(1+i_{\mathrm{XP}})-i_{\mathrm{XB}} & 0 & -1 & 1 \\ 1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{\mathrm{P}} & f_{\mathrm{P}} & 0 & 0 & 0 \\ 0 & 0 & 0 & i_{\mathrm{XB}}-f_{\mathrm{P}}i_{\mathrm{XP}} & i_{\mathrm{XB}}-f_{\mathrm{P}}i_{\mathrm{XP}} & 0 & 0 & -1 \end{bmatrix} \quad \text{and}$$

$$\boldsymbol{\sigma_S} \coloneqq \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1/Y_{\mathrm{H}} & -1/Y_{\mathrm{H}} & 0 & 0 & 0 & 0 & 1 & 0 \\ -(1-Y_{\mathrm{H}})/Y_{\mathrm{H}} & 0 & -(4.57-Y_{\mathrm{A}})/Y_{\mathrm{A}} & 0 & 0 & 0 & 0 & 0 \\ 0 & -(1-Y_{\mathrm{H}})/(2.86Y_{\mathrm{H}}) & 1/Y_{\mathrm{A}} & 0 & 0 & 0 & 0 & 0 \\ -i_{\mathrm{XB}} & -i_{\mathrm{XB}} & -i_{\mathrm{XB}}-1/Y_{\mathrm{A}} & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}.$$

The reaction rates involve the Monod expression $\mu(A,B) \coloneqq \frac{A}{A+B}$ and the functions μ_7 and μ_8 defined by $\mu_7(0,0) \coloneqq 0$, $\mu_8(0,0) \coloneqq 0$ if $X_{\mathrm{S}} = 0$ and $X_{\mathrm{B,H}} = 0$ and

$$\mu_7(X_{\mathrm{S}}, X_{\mathrm{B,H}}) \coloneqq \frac{X_{\mathrm{S}} X_{\mathrm{B,H}}}{K_{\mathrm{X}} X_{\mathrm{B,H}} + X_{\mathrm{S}}}, \quad \mu_8(X_{\mathrm{B,H}}, X_{\mathrm{ND}}) \coloneqq \frac{X_{\mathrm{B,H}} X_{\mathrm{ND}}}{K_{\mathrm{X}} X_{\mathrm{B,H}} + X_{\mathrm{S}}} \quad \text{otherwise.}$$

The reaction vector $\boldsymbol{R}(\boldsymbol{C}, \boldsymbol{S})$ then becomes

Table 3 Stoichiometric and kinetic parameters at 26°C

Symbol	Name	Value	Unit
Y_A	Yield for autotrophic biomass	0.24	$(\text{g COD})(\text{g N})^{-1}$
Y_H	Yield for heterotrophic biomass	0.67	$(\text{g COD})(\text{g COD})^{-1}$
f_P	Fraction of biomass leading to particulate products	0.08	dimensionless
i_{XB}	Mass of nitrogen per mass of COD in biomass	0.086	$(\text{g N})(\text{g COD})^{-1}$
i_{XP}	Mass of nitrogen per mass of COD in products from biomass	0.06	$(\text{g N})(\text{g COD})^{-1}$
μ_H	Maximum specific growth rate for heterotrophic biomass	6.0	d^{-1}
K_S	Half-saturation coefficient for heterotrophic biomass	20.0	$(\text{g COD})\text{m}^{-3}$
$K_{O,H}$	Oxygen half-saturation coefficient for heterotrophic biomass	0.2	$-(\text{g COD})\text{m}^{-3}$
K_{NO}	Nitrate half-saturation coefficient for denitrifying heterotrophic biomass	0.5	$(\text{g } NO_3\text{-N})\text{m}^{-3}$
b_H	Decay coefficient for heterotrophic biomass	0.62	d^{-1}
η_g	Correction factor for μ_H under anoxic conditions	0.8	dimensionless
η_h	Correction factor for hydrolysis under anoxic conditions	0.4	dimensionless
k_h	Maximum specific hydrolysis rate	3.0	$(\text{g COD})(\text{g COD})^{-1}\text{d}^{-1}$
K_X	Half-saturation coefficient for hydrolysis of slowly biodegradable substrate	0.03	$(\text{g COD})(\text{g COD})^{-1}$
μ_A	Maximum specific growth rate for autotrophic biomass	0.8	d^{-1}
$\bar{K}_{NH}$	Ammonia half-saturation coefficient for aerobic and anaerobic growth of heterotrophs	0.05	$(\text{g } NH_3\text{-N})\text{m}^{-3}$
K_{NH}	Ammonia half-saturation coefficient for autotrophic biomass	1.0	$(\text{g } NH_3\text{-N})\text{m}^{-3}$
b_A	Decay coefficient for autotrophic biomass	0.15	d^{-1}
$K_{O,A}$	Oxygen half-saturation coefficient for autotrophic biomass	0.4	$-(\text{g COD})\text{m}^{-3}$
k_a	Ammonification rate	0.08	$\text{m}^3(\text{g COD})^{-1}\text{d}^{-1}$

$$\boldsymbol{R}(\boldsymbol{C},\boldsymbol{S}) \coloneqq \begin{pmatrix} \mu_H \mu(S_{NH},\bar{K}_{NH})\mu(S_S,K_S)\mu(S_O,K_{O,H})X_{B,H} \\ \mu_H \mu(S_{NH},\bar{K}_{NH})\mu(S_S,K_S)\mu(K_{O,H},S_O)\mu(S_{NO},K_{NO})\eta_g X_{B,H} \\ \mu_A \mu(S_{NH},K_{NH})\mu(S_O,K_{O,A})X_{B,A} \\ b_H X_{B,H} \\ b_A X_{B,A} \\ k_a S_{ND} X_{B,H} \\ k_h \mu_7(X_S,X_{B,H})\left(\mu(S_O,K_{O,H}) + \eta_h \mu(K_{O,H},S_O)\mu(S_{NO},K_{NO})\right) \\ k_h \mu_8(X_{B,H},X_{ND})\left(\mu(S_O,K_{O,H}) + \eta_h \mu(K_{O,H},S_O)\mu(S_{NO},K_{NO})\right) \end{pmatrix}.$$

Acknowledgements R.B. is supported by ANID (Chile) through Fondecyt project 1210610; project Anillo ANID/ACT210030; Centro de Modelamiento Matemático (CMM), project FB210005 of BASAL funds for Centers of Excellence; and CRHIAM, projects ANID/FONDAP/15130015 and ANID/FONDAP/1523A0001. J.C. is supported by ANID through Fondecyt project 3230553. S.D. acknowledges support from the Swedish Research Council (Vetenskapsrådet, 2019-04601).

References

1. Bürger, R., Careaga, J., Diehl, S.: A method-of-lines formulation for a model of reactive settling in tanks with varying cross-sectional area. IMA J. Appl. Math. **86**, 514–546 (2021)
2. Bürger, R., Careaga, J., Diehl, S., Pineda, R.: A moving-boundary model of reactive settling in wastewater treatment. Part 1: Governing equations. Appl. Math. Modelling **106**, 390–401 (2022)
3. Bürger, R., Careaga, J., Diehl, S., Pineda, R.: A moving-boundary model of reactive settling in wastewater treatment. Part 2: Numerical scheme. Appl. Math. Modelling **111**, 247–269 (2022)
4. Bürger, R., Careaga, J., Diehl, S., Pineda, R.: A model of reactive settling of activated sludge: Comparison with experimental data. Chem. Eng. Sci. **267**, article 118244 (2023)
5. Bürger, R., Careaga, J., Diehl, S., Pineda, R.: Numerical schemes for a moving-boundary convection-diffusion-reaction model of sequencing batch reactors. ESAIM: Math. Model. Numer. Anal. **57**, 2931–2976 (2023)
6. Bürger, R., Coronel, A., Sepúlveda, M.: A semi-implicit monotone difference scheme for an initial-boundary value problem of a strongly degenerate parabolic equation modeling sedimentation-consolidation processes. Math. Comp. **75**, 91–112 (2006)
7. Bürger, R., Damasceno, J. J. R., Karlsen, K. H.: A mathematical model for batch and continuous thickening of flocculated suspensions in vessels with varying cross-section. Int. J. Mineral Process. **73**, 183–208 (2004)
8. Bürger, R., Diehl, S., Mejías, C.: A difference scheme for a degenerating convection-diffusion-reaction system modelling continuous sedimentation. ESAIM: Math. Model. Numer. Anal. **52**, 365–392 (2018)
9. Bürger, R., Diehl, S., Nopens, I.: A consistent modelling methodology for secondary settling tanks in wastewater treatment. Water Res. **45**, 2247–2260 (2011)
10. Bürger, R., Karlsen, K. H., Towers, J. D.: A model of continuous sedimentation of flocculated suspensions in clarifier-thickener units. SIAM J. Appl. Math. **65**, 882–940 (2005)
11. Chen, G., van Loosdrecht, M. C. M., Ekama, G. A.; Brdjaniovic, D.: Biological Wastewater Treatment, 2nd edition. IWA Publishing, London (2020)
12. Diehl, S.: Continuous sedimentation of multicomponent particles. Math. Meth. Appl. Sci. **20**, 1345–1364 (1997)
13. Droste, R., Gear, R.: Theory and Practice of Water and Wastewater Treatment, 2nd edition. Wiley, Hoboken, NJ, USA (2019)
14. Engquist, B., Osher, S.: One-sided difference approximations for nonlinear conservation laws. Math. Comp. **36**, 321–351 (1981)
15. Ford, W.: Numerical Linear Algebra with Applications: Using MATLAB. Academic Press, San Diego (2015)
16. Formaggia, L., Scotti, A.: Positivity and conservation properties of some integration schemes for mass action kinetics. SIAM J. Numer. Anal. **49**, 1267–1288 (2011)
17. Gray, N. F.: Activated Sludge: Developments and Sustainable Solutions. World Scientific, New Jersey (2023)
18. Henze, M., Gujer, W., Mino, T., van Loosdrecht, M. C.: Activated Sludge Models ASM1, ASM2, ASM2d and ASM3. IWA Publishing, London (2000)
19. Kirim, G., Torfs, E., Vanrolleghem, P.A.: An improved 1d reactive Bürger-Diehl settler model for secondary settling tank denitrification. Water Environ. Res. **94**, article e10825 (2022)
20. Kynch, G. J.: A theory of sedimentation. Trans. Farad. Soc. **48**, 166–176 (1952)
21. Makinia, J., Zaborowska, E.: Mathematical Modelling and Computer Simulation of Activated Sludge Systems, 2nd edition. IWA Publishing, London (2020)
22. Metcalf, L., Eddy, H. P.: Wastewater Engineering. Treatment and Resource Recovery, 5th edition. McGraw-Hill, New York (2014)
23. Riffat, R., Husnain, T.: Fundamentals of Wastewater Treatment and Engineering, second edition. CRC Press, Abingdon, Oxon, UK (2022)
24. Thomas, L. H.: Elliptic Problems in Linear Difference Equations over a Network. Watson Sci. Comput. Lab. Rept., Columbia University, New York (1949)

Centro de Investigación en Ingeniería Matemática (CI^{2}MA)

PRE-PUBLICACIONES 2024

2024-02 Julio Careaga, Gabriel N. Gatica, Cristian Inzunza, Ricardo Ruiz-Baier: *New Banach spaces-based mixed finite element methods for the coupled poroelasticity and heat equations*

2024-03 Harold D. Contreras, Paola Goatin, Luis M. Villada: *A two-lane bidirectional nonlocal traffic model*

2024-04 Rommel Bustinza, Matteo Cicuttin, Ariel Lombardi: *A Hybrid High-Order method for the mixed Steklov eigenvalue problem*

2024-05 Isaac Bermudez, Jaime Manríquez, Manuel Solano: *A hybridizable discontinuous Galerkin method for Stokes/Darcy coupling in dissimilar meshes*

2024-06 Thomas Führer, Diego Paredes: *Robust hybrid finite element methods for reaction-dominated diffusion problems*

2024-07 Raimund Bürger, Enrique D. Fernández Nieto, Jorge Moya: *A multilayer shallow water model for tsunamis and coastal forest interaction*

2024-08 Fernando Betancourt, Raimund Bürger, Stefan Diehl, María Carmen Martí, Yolanda Vásquez: *A degenerating convection-diffusion model of a flotation column: theory, numerics and applications*

2024-09 Fernando Betancourt, Raimund Bürger, Julio Careaga, Lucas Romero: *Coupled finite volume methods for settling in inclined vessels with natural convection*

2024-10 Kaïs Ammari, Vilmos Komornik, Mauricio Sepúlveda, Octavio Vera: *Stability of the Rao-Nakra sandwich beam with a dissipation of fractional derivative type: theoretical and numerical study*

2024-11 Lady Angelo, Jessika Camaño, Sergio Caucao: *A skew-symmetric-based mixed FEM for stationary MHD ows in highly porous media*

2024-12 Gabriel N. Gatica: *A note on the generalized Babuska-Brezzi theory: revisiting the proof of the associated Strang error estimates*

2024-13 Carlos D. Acosta, Raimund Bürger, Julio Careaga, Stefan Diehl, Romel Pineda, Daniel Támara: *A semi-implicit method for a degenerating convection-diffusion-reaction problem modeling secondary settling tanks*

Para obtener copias de las Pre-Publicaciones, escribir o llamar a: Director, Centro de Investigación en Ingeniería Matemática, Universidad de Concepción, Casilla 160-C, Concepción, Chile, Tel.: 41-2661324, o bien, visitar la página web del centro: `http://www.ci2ma.udec.cl`







Centro de Investigación en
Ingeniería Matemática (CI²MA)
Universidad de Concepción

Casilla 160-C, Concepción, Chile
Tel.: 56-41-2661324/2661554/2661316
http://www.ci2ma.udec.cl