NUMERICAL SCHEMES FOR A MOVING-BOUNDARY CONVECTION-DIFFUSION-REACTION MODEL OF SEQUENCING BATCH REACTORS

RAIMUND BÜRGER[1], JULIO CAREAGA[2], STEFAN DIEHL[3] AND ROMEL PINEDA[1,*]

Abstract. Sequencing batch reactors (SBRs) are devices widely used in wastewater treatment, chemical engineering, and other areas. They allow for the sedimentation and compression of solid particles of biomass simultaneously with biochemical reactions with nutrients dissolved in the liquid. The kinetics of these reactions may be given by one of the established activated sludge models (ASMx). An SBR is operated in various stages and is equipped with a movable extraction and fill device and a discharge opening. A one-dimensional model of this unit can be formulated as a moving-boundary problem for a degenerating system of convection-diffusion-reaction equations whose unknowns are the concentrations of the components forming the solid and liquid phases, respectively. This model is transformed to a fixed computational domain and is discretized by an explicit monotone scheme along with an alternative semi-implicit variant. The semi-implicit variant is based on solving, during each time step, a system of nonlinear equations for the total solids concentration followed by the solution of linear systems for the solid component percentages and liquid component concentrations. It is demonstrated that the semi-implicit scheme is well posed and that both variants produce approximations that satisfy an invariant region principle: solids concentrations are nonnegative and less or equal to a set maximal one, percentages are nonnegative and sum up to one, and substrate concentrations are nonnegative. These properties are achieved under a Courant-Friedrichs-Lewy (CFL) condition that is less restrictive for the semi-implicit than for the explicit variant. Numerical examples with realistic parameters illustrate that as a consequence, the semi-implicit variant is more efficient than the explicit one.

1. INTRODUCTION

1.1. Scope

Sequencing batch reactors (SBRs) are widely used in the pharmaceutical [41, 44], petrochemical [20], and chemical [33] industries as well as in wastewater treatment [1, 42]. An SBR is a tank designed for the sedimen-

[1] CI^{2}MA and Departamento de Ingeniería Matemática, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Casilla 160-C, Concepción, Chile.

[2] Department of Mathematics, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands.

[3] Centre for Mathematical Sciences, Lund University, P.O. Box 118, 221 00 Lund, Sweden.

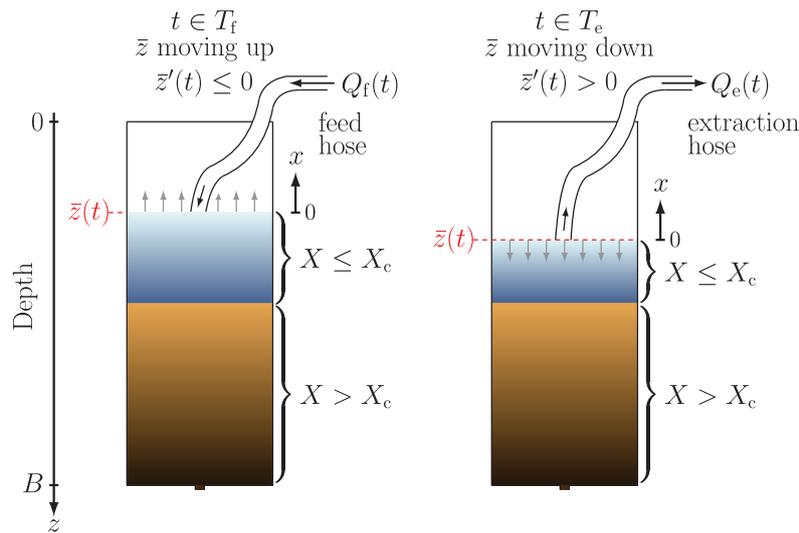*Corresponding author: `rpineda@ing-mat.udec.cl`

FIGURE 1. Illustration of an SBR in the two stages that involve flow through the hose: (left) feed and (right) extraction. The boundary $\bar{z}$ moves upward at time t when $\bar{z}'(t) < 0$ and downward when $\bar{z}'(t) > 0$, and T_f and T_e are the feed and extraction time intervals, respectively. The blue-shaded and brown regions symbolically illustrate regions where $X \leq X_c$ and $X > X_c$, respectively, where X is the total solids (biomass) concentration and $X_c > 0$ is a critical concentration beyond which the flocculent particles form a porous network that is able to bear solid stress. The interface between both regions marks the sludge blanket level $X = X_c$.

tation of a suspension composed of solid particles of biomass that react with substrates (nutrients) dissolved in the liquid. This process is a fundamental stage of purification in wastewater treatment plants. 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. The simultaneous process of sedimentation and reactions, called reactive settling, occurs both in plants with continuously operated secondary settling tank (SST) and in SBRs when such are used. We study here a general model of an SBR, which is operated in a number of sequential stages [21, 26, 40]. Depending on the stage, the suspension is continuously fed or extracted from the top of its surface by a floating device; see Figure 1, and concentrated mixture can be extracted from the bottom of the vessel. The cycle of an SBR takes into account five principal stages: fill, react, settle, draw and idle. The bulk flows at the different stages lead to a moving boundary; see Figure 2. The description with five principal stages is sufficient for the understanding of the process for the development of numerical schemes of the model. In a real SBR, there are usually more time events than between the stages, where mixing can occur during part of the fill stage, aeration may occur during parts of the mixing time intervals, etc. Our model, and hence the numerical schemes presented, can handle any combination of such events.

A nonlinear, strongly degenerating convection-diffusion-reaction model of an SBR was introduced in [15] and a numerical scheme defined on a fixed spatial grid in [16]. That scheme will here be referred to as SBR2. Within that approach one needs to carefully track the upper boundary as it moves through the computational grid for the numerical solution. It is the purpose of the present work to introduce an alternative explicit scheme that after a suitable transformation of the time-dependent spatial domain is defined on a fixed computational grid. Moreover, a semi-implicit variant is introduced that allows for more efficient simulations due to a more favourable Courant-Friedrichs-Lewy (CFL) condition.

The moving boundary, denoted by $\bar{z} := \bar{z}(t)$, where $t \geq 0$ is time, is a known function given by the bulk flows acting on the SBR. The curve $t \mapsto \bar{z}(t)$ defines a spatial change of variables, where in the new variable ξ, the

top and bottom boundaries are located at $\xi = 0$ and $\xi = 1$, respectively. The sought vector unknowns, in the variable ξ, are the solid phase and liquid substrate components

$$\boldsymbol{C} := \boldsymbol{C}(\xi, t) := \left(C^{(1)}(\xi, t), \dots, C^{(k_{\boldsymbol{C}})}(\xi, t) \right)^{\mathrm{T}}, \quad \boldsymbol{S} := \boldsymbol{S}(\xi, t) := \left(S^{(1)}(\xi, t), \dots, S^{(k_{\boldsymbol{S}})}(\xi, t) \right)^{\mathrm{T}}.$$

The model equations for $0 < \xi < 1$ are given by

$$\partial_t \boldsymbol{C} + \partial_\xi \left(U_{\boldsymbol{C}}(X, \xi, t) \boldsymbol{C} \right) = \partial_\xi \left(\gamma(\xi) \beta(t)^2 \partial_\xi \mathcal{D}(X) \boldsymbol{C} \right)$$
$$+ \beta(t) \bar{z}'(t) \boldsymbol{C} + \delta(\xi) \beta(t) q_{\mathrm{f}}(t) \boldsymbol{C}_{\mathrm{f}}(t) + \gamma(\xi) \boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{C}, \boldsymbol{S}), \tag{1.1a}$$

$$\partial_t \boldsymbol{S} + \partial_\xi \left(U_{\boldsymbol{S}}(X, \xi, t) \boldsymbol{S} \right) = -\partial_\xi \left(\frac{\gamma(\xi) \beta(t)^2}{\rho_X - X} \partial_\xi \mathcal{D}(X) \boldsymbol{S} \right)$$
$$+ \beta(t) \bar{z}'(t) \boldsymbol{S} + \delta(\xi) \beta(t) q_{\mathrm{f}}(t) \boldsymbol{S}_{\mathrm{f}}(t) + \gamma(\xi) \boldsymbol{R}_{\boldsymbol{S}}(\boldsymbol{C}, \boldsymbol{S}), \tag{1.1b}$$

where $U_{\boldsymbol{C}}$ and $U_{\boldsymbol{S}}$ are velocity functions related to the velocities of the solid and liquid phases, respectively, $X := (C^{(1)} + \cdots + C^{(k_{\boldsymbol{C}})})/c$ is the total solids concentration, c is a conversion factor, $\beta(t) = 1/(B - \bar{z}(t))$ where $z = B$ is the bottom of the tank, ρ_X is the (constant) mass density of solids, and $\mathcal{D}(X)$ is an integrated diffusion coefficient describing compressibility of the sediment. The particles are not rigid spheres, but flocs (or flocculent particles) that form a porous network capable of undergoing further compression (consolidation), with release of water, under their own weight, whenever the concentration X is higher than a so-called critical concentration $X_{\mathrm{c}} > 0$, which marks the transition between the hindered settling and compression regimes. It is assumed that $\mathcal{D}(X) = 0$ for $X \leq X_{\mathrm{c}}$, so the system (1.1) is of first order wherever $X \leq X_{\mathrm{c}}$, and therefore is strongly degenerate. The second term in the right-hand side of both equations is due to the change of variables in the spatial partial derivative, while the third term contains the feed concentrations $\boldsymbol{C}_{\mathrm{f}}$ and $\boldsymbol{S}_{\mathrm{f}}$ for the solids and substrates, respectively, the feed velocity q_{f}, and the Dirac symbol δ. The last terms model the chemical reactions between the components of the solid and liquid phases, where $\boldsymbol{R}_{\boldsymbol{C}}$ and $\boldsymbol{R}_{\boldsymbol{S}}$ are vectors of reaction terms, and γ is the characteristic function that equals one inside the mixture and zero outside. The system (1.1) is coupled to transport equations modelling the concentrations in the extraction pipe connected at $\xi = 0$. (The present explanations provide the rough structure of the model, and the variables and functions arising in (1.1) are defined precisely in later parts of the paper.)

During certain intervals of an SBR cycle, mixing occurs by an impeller or aeration. Then all concentrations are assumed homogeneous in the tank. By averaging the partial differential equations (PDEs) (1.1) over depth one obtains a system of ordinary differential equations without the terms with velocity and compression functions. The resulting system governs the mixing periods, which in the examples here occur precisely during the react stage. Moreover, the traditional notion of "react stage" only refers to that most of the reactions occur during that period. In reality, and which is one of the points of our model, reactions may occur all the time during an SBR cycle.

Finally, we recall that since for $X \leq X_{\mathrm{c}}$, (1.1) is a system of first-order balance equations with nonlinear convective flux, its solutions $\boldsymbol{C} = \boldsymbol{C}(\xi, t)$ and $\boldsymbol{S} = \boldsymbol{S}(\xi, t)$ are in general discontinuous. As a consequence, solutions of (1.1) must be understood as weak solutions. Moreover, the coefficients of both equations are discontinuous with respect to the coordinate ξ, as becomes evident through the discontinuous dependence of $U_{\boldsymbol{C}}$, $U_{\boldsymbol{S}}$, and γ on ξ. The model is posed on the whole real line for ξ without any explicit boundary conditions. In fact, the solution behaviour across $\xi = 0$ and $\xi = 1$ is dictated by the continuity of flux (of each species) in conjunction with the discontinuous algebraic definition of the flux on either side. This particular solution concept makes it also unnecessary to specify a derivative of γ.

1.2. Related work

Several works posed for SBRs have mostly considered the fill and react stages, where a homogeneously mixed tank is assumed and the biochemical reactions are modelled by ODEs from some of the established activated sludge models (ASM1, ASM2, ASM2d, or ASM3; in short, AMSx) [30, 39]. We refer to handbooks
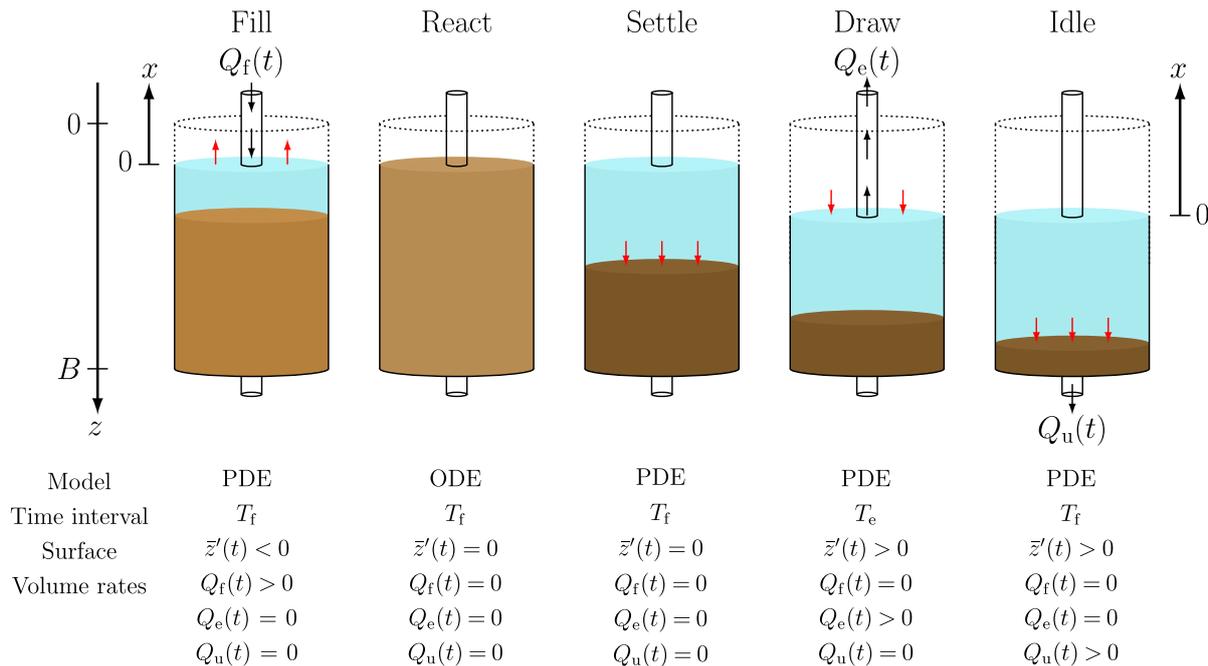
FIGURE 2. Schematic of the five stages of an SBR. The red arrows show the movement of the surface or the sediment level during each stage. The black arrows show the direction of the bulk flows Q_f (feed), Q_e (extraction) and Q_u (underflow). Colours are symbolic; the darker brown region corresponds to the sediment, where $X > X_c$).

(*e.g.*, [21, 26, 40]) for a broader introduction to the background of wastewater treatment. PDE models for the reactive-settling process have recently been developed [13–16]. One of the first PDE-based models combining the sedimentation-consolidation process for batch denitrification in wastewater treatment was presented in [12]. That model consists of five unknowns, two solid components and three substrates, and includes nonlinear terms for the reactions between the solid and liquid components. An extension to the case of continuous sedimentation was introduced in [13], where the model was based on the percentages of the concentrations of the solid and liquid phases. A reliable numerical scheme and simulations of the denitrification process were also included. The present work also utilizes the concept of percentages for sedimentation models, which was introduced in [25]. Here we emphasize that percentages are the local fractions of two or more components of the solid phase. While the various phases undergo relative movement, the components of each phase always move with their phase. In [14], the authors presented an alternative reactive settling model which also included non-cylindrical vessels. A numerical scheme based on a method-of-lines formulation was developed and the numerical results are compared with the ones of [13]. Recently, a slightly modified version of that model including extra diffusion terms modelling mixing near the feed inlet was calibrated to experiments from a pilot plant in [18].

The present SBR model is the one in [15], which is a development of the one in [14] to take into account the moving boundary due to bulk flows of the SBR stages. The settling of particles is allowed in the model in all five stages with the exception of the react one. A conservative and positivity-preserving numerical scheme developed on a fixed mesh grid was introduced in [16]. The scheme of [16] employs careful mass distributions in cells near the moving boundary to ensure the conservation of mass. Simulations of the activated sludge model no. 1 (ASM1) [30] were included in [16] for the case of a cylindrical vessel.

The trajectory of $\bar{z}(t)$ is given, and not part of the solution of the problem, so the governing PDE model (1.1) is a moving boundary problem, but not a free boundary problem. Nevertheless, extensions of the SBR model

to a free boundary problem are conceivable. For one scalar PDE related models of filtration have been studied in [7, 8].

The numerical schemes formulated herein (the explicit and semi-implicit versions) are based on conservative difference methods for systems of nonlinear conservation laws or more generally, conservative convection-diffusion equations, together with additional terms to handle feed sources and reactions. These schemes can therefore be understood as a special case of finite volume schemes. Well-known introductions to these schemes include [29, 31, 38], while related methods for convection-diffusion-reaction problems are treated, for instance, in [34].

Finally, we mention that the formulation and partial analysis of the semi-implicit scheme presented herein is based on analyses of schemes of that type for degenerate parabolic PDE in [10, 28].

1.3. Outline of the paper

The remainder of this work is organized as follows. In Section 2, we review the SBR model, starting with assumptions on the tank and the one-dimensional model, and explaining the relation between the moving boundary and bulk flows (Sect. 2.1). The one-dimensional model of the reactive sedimentation process in the tank is obtained from the balance laws of mass of all the particulate and soluble components. The details of the formulation of the model are outlined in Section 2.2, starting from the solid and liquid components and reactions within the biokinetic reaction model (as an example, we here employ the model ASM1) followed by a summary of the sedimentation-compression model that has also been employed in previous works. The balance laws for the SBR are coupled with transport equations for the pipe. During the react stage with full mixing, the PDEs are replaced by a system of ordinary differential equations (ODEs). While these model ingredients are reviewed from our previous treatment [15, 16], one of the three decisive novelties of the present approach, namely the transformation of the moving boundary model to a fixed computational domain, is formulated in Section 2.3. The second novelty is the reformulation of the governing model in terms of the vector of percentages $\boldsymbol{p}$ (of the solid components, with respect to X) done in Section 2.4.

An explicit numerical scheme for the solution of the transformed governing model is introduced in Section 3. This scheme is based on a standard discretization in space and time of the computational domain (Sect. 3.1). The explicit numerical scheme, presented in Section 3.2, is based on the percentage and fixed-domain formulations of Sections 2.3 and 2.4 and combines upwind discretizations for transport terms, the Engquist-Osher numerical flux [27] for nonlinear flux terms, and a central difference formula for nonlinear degenerating diffusion terms arising in the governing models, combined with appropriate discretizations of the reaction terms. The scheme is complemented by a suitable discretization of the ODEs describing the mixing stage (Sect. 3.3). It is assumed that the spatial mesh width $\Delta\xi$ and the time step τ are related by a CFL condition outlined in Section 3.4, and we prove in Section 3.5 that this scheme is montone, with the consequence that an invariant-region principle holds, *i.e.*, solids concentrations are nonnegative and less or equal a maximal one, percentages are nonnegative and sum up to one, and substrate concentrations are nonnegative. The CFL condition for the explicit scheme essentially bounds $\tau/\Delta\xi^2$. Here we use the term "CFL condition" in the usual mathematical sense as the limitation of τ for given $\Delta\xi$ to achieve stability of the scheme; this concept generalizes that of the original definition [23] that strictly involves a velocity and is essentially applicable to hyperbolic problems only. In our case, the presence of $\tau/\Delta\xi^2$ within the CFL condition arises from the explicit discretization of the degenerating diffusion term in (1.1a).

A more favorable CFL condition that only bounds $\tau/\Delta\xi$ is associated with a semi-implicit scheme for the governing PDE model, which is the third novelty introduced in Section 4. At each time step, it consists of a nonlinear semi-implicit scheme for the PDE for the total solids concentration X, which is described in Section 4.1, where we also demonstrate that the nonlinear equations are well posed, *i.e.*, possess a unique solution that depends continuously on data. The solution of the nonlinear equations is achieved by Newton-Raphson method (Sect. 4.2). Once the scalar function X has been updated, one proceeds to update the vectors $\boldsymbol{p}$ and $\boldsymbol{S}$. This is done by an implicit version of the corresponding $\boldsymbol{p}$- and $\boldsymbol{S}$-schemes of Section 3. This version is, however, linearly implicit and only requires the solution of one linear system per time step, see Sections 4.3 and 4.4. In those sections, it is proved that the semi-implicit scheme has the same monotonicity and invariant-region properties as

its explicit counterpart (*cf.* Sect. 3.5) but does so under the more favorable CFL condition. Numerical examples that illustrate the performance of the numerical schemes of Sections 3 and 4 are presented in Section 5. In particular, it is demonstrated that the semi-implicit scheme indeed leads to the expected gain in efficiency. Finally, conclusions are collected in Section 6.

2. A model of a sequencing batch reactor (SBR)

We here review the assumptions of the general SBR model in [15] and refer to [16] for a full description of the numerical scheme SBR2. That model will then be slightly reformulated, which allows for a semi-implicit numerical scheme to be developed. The sedimentation and compression properties of the flocculated sludge are namely assumed to depend on the total concentration X and not on the concentrations of the individual components since these are flocculated to larger particles. Each flocculated particle consists of several individual components, which can be described as percentages of the total concentration X. For simplicity of writing, we confine here to a constant cross-sectional area A, which is the most common case in the applications.

2.1. Assumptions on the tank and the 1D model

The reactor tank is assumed to be a cylindrical vessel with constant horizontal cross-sectional area A; see Figure 2. We place a fixed z-axis indicating depth from the top ($z = 0$) to the bottom at $z = B$. At the surface of the mixture, located at $z = \bar{z}(t)$, a floating device connected to a pipe allows one to feed the tank at given volumetric feed flow $Q_{\mathrm{f}}(t)$ [$\mathrm{m^3\,s^{-1}}$] and feed concentrations $\boldsymbol{C}_{\mathrm{f}}(t)$ and $\boldsymbol{S}_{\mathrm{f}}(t)$. The floating device can alternatively extract mixture at a given volume rate $Q_{\mathrm{e}}(t)$ during the draw stage. One cannot fill and draw simultaneously. The extraction pipe is modelled by a half-line and the flow through it by a linear advection PDE. The concentrations in the pipe are denoted by $\boldsymbol{C}_{\mathrm{e}}(t)$ and $\boldsymbol{S}_{\mathrm{e}}(t)$. At the bottom, $z = B$, one can withdraw mixture at a given volume rate $Q_{\mathrm{u}}(t) \geq 0$, and the corresponding output concentrations there are denoted by $\boldsymbol{C}_{\mathrm{u}}(t)$ and $\boldsymbol{S}_{\mathrm{u}}(t)$. We define the bulk velocities

$$q_{\mathrm{u}}(t) := Q_{\mathrm{u}}(t)/A, \qquad q_{\mathrm{e}} := Q_{\mathrm{e}}(t)/A, \qquad q_{\mathrm{f}} := Q_{\mathrm{f}}(t)/A.$$

If $[0, T]$ denotes the total time interval of modelling (and simulation in Sect. 3), we assume that $T := T_{\mathrm{e}} \cup T_{\mathrm{f}}$, where

$$T_{\mathrm{e}} := \big\{ t \in \mathbb{R}_+ : Q_{\mathrm{e}}(t) > 0, Q_{\mathrm{f}}(t) = 0 \big\}, \qquad T_{\mathrm{f}} := \big\{ t \in \mathbb{R}_+ : Q_{\mathrm{e}}(t) = 0, Q_{\mathrm{f}}(t) \geq 0 \big\}$$

(such that $T_{\mathrm{e}} \cap T_{\mathrm{f}} = \varnothing$). The volume of the mixture at time t is

$$\bar{V}(t) := A\big(B - \bar{z}(t)\big). \tag{2.1}$$

The surface location $\bar{z}(t)$ is determined by the given volumetric flows, since by differentiation of (2.1),

$$\bar{z}'(t) = -\frac{\bar{V}'(t)}{A} = \frac{Q_{\mathrm{u}}(t) - \bar{Q}(t)}{A}, \quad \text{where} \quad \bar{Q}(t) := \begin{cases} -Q_{\mathrm{e}}(t) < 0 & \text{if } t \in T_{\mathrm{e}}, \\ Q_{\mathrm{f}}(t) \geq 0 & \text{if } t \in T_{\mathrm{f}}. \end{cases} \tag{2.2}$$

We assume that $Q_{\mathrm{f}}(t)$ is controlled in such a way that overflow never occurs.

Clearly, $\bar{z}$ is given at any time, so the model under consideration is a *moving* boundary problem (for which the time-dependent position of the boundary is given externally and therefore known at any moment), but not a *free* boundary problem (where the trajectory of the boundary is unknown *a priori* and is part of the solution of the problem). That said, we emphasize that because of the volumetric flows, no boundary *conditions* need to be imposed since the conservation law implies natural output concentrations when reactions and sedimentation are assumed to occur inside the tank only. The latter property is ensured by the discontinuous coefficient $\gamma(\xi)$ in (1.1).

2.2. A model of reactive settling with moving boundary

2.2.1. Biochemical reaction model and solid and liquid components

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. Any biochemical reaction model can be used such as one of the standard ASMx activated sludge models. Within the ASMx models concentrations are usually expressed in terms of more easily measurable units such as chemical oxygen demand (COD) (*cf.* Tab. A.1 in Appendix A), wherefore conversion factors have to be used to obtain the mass concentrations. We here use the model ASM1 (see Appendix A), in which the particulate concentrations are (in ASM1 units)

$$X_I, X_S, X_{B,H}, X_{B,A}, X_P, X_{ND},$$

and the corresponding definition of the total suspended solids concentration is

$$X := c(X_I + X_S + X_{B,H} + X_{B,A} + X_P), \quad \text{where } c = 0.75 \, \text{g/(g COD)}. \tag{2.3}$$

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

$$\boldsymbol{C} := \left(X_I, X_{S-ND}, X_{B,H}, X_{B,A}, X_P, X_{ND}\right)^T \quad (i.e., k_{\boldsymbol{C}} = 6),$$
$$\boldsymbol{S} := \left(S_I, S_S, S_O, S_{NO}, S_{NH}, S_{ND}\right)^T \quad (i.e., k_{\boldsymbol{S}} = 6).$$

Moreover, we define

$$X := c\left(C^{(1)} + \cdots + C^{(k_{\boldsymbol{C}})}\right), \quad L := W + S^{(1)} + \cdots + S^{(k_{\boldsymbol{S}})} \tag{2.4}$$

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

$$\boldsymbol{p} := \frac{c}{X}\boldsymbol{C}. \tag{2.5}$$

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; however, 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.

2.2.2. Reaction terms

The nonlinear reaction terms are given by

$$\boldsymbol{\Lambda_C R_C}(\boldsymbol{C}, \boldsymbol{S}), \quad \text{where} \quad \boldsymbol{R_C}(\boldsymbol{C}, \boldsymbol{S}) := \boldsymbol{\sigma_C r}(\boldsymbol{C}, \boldsymbol{S}),$$
$$\boldsymbol{\Lambda_S R_S}(\boldsymbol{C}, \boldsymbol{S}), \quad \text{where} \quad \boldsymbol{R_S}(\boldsymbol{C}, \boldsymbol{S}) := \boldsymbol{\sigma_S r}(\boldsymbol{C}, \boldsymbol{S}),$$

which model the increase in (COD) concentration per time unit; see Table A.1. Here, $\boldsymbol{\Lambda_C} = \text{diag}(c, c, c, c, c, 1)$ and $\boldsymbol{\Lambda_S}$ are diagonal matrices with conversion factors, $\boldsymbol{\sigma_C}$ and $\boldsymbol{\sigma_S}$ are constant stoichiometric matrices, and $\boldsymbol{r}(\boldsymbol{C}, \boldsymbol{S}) \geq \boldsymbol{0}$ is a vector of nonlinear functions modelling the reaction processes; see Appendix A.

We assume that if a solid component is not present, $p^{(k)} = 0$, then no more such can vanish, *i.e.* $R_{\boldsymbol{C}}^{(k)}(\boldsymbol{C}, \boldsymbol{S})|_{p^{(k)}=0} = 0$, where $R_{\boldsymbol{C}}^{(k)}$ denotes the k-th component of $\boldsymbol{R_C}$. Finally, to be able to establish an invariant-region property for the numerical solution, we make some additional technical assumptions. To ensure that the numerical solution for the solids does not exceed the maximum total solids concentration $\hat{X}$ (see (2.10)), we assume that

$$\text{there exists an } \varepsilon > 0 \text{ such that } \boldsymbol{R_C}(\boldsymbol{C}, \boldsymbol{S}) = \boldsymbol{0} \text{ for all } X \geq \hat{X} - \varepsilon. \tag{2.6}$$

This condition means that when the concentration X is near the maximal one $\hat{X}$, biomass cannot grow any more. To obtain positivity of component k of the concentration vector $\boldsymbol{C}$, we let

$$I_{\boldsymbol{C},k}^- := \left\{ l \in \mathbb{N} : \sigma_{\boldsymbol{C}}^{(k,l)} < 0 \right\} \quad \text{and} \quad I_{\boldsymbol{C},k}^+ := \left\{ l \in \mathbb{N} : \sigma_{\boldsymbol{C}}^{(k,l)} > 0 \right\} \tag{2.7}$$

denote the sets of indices l that have negative and positive stoichiometric coefficients, respectively, and assume that

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

The latter assumption on the reactions is necessary to have a physically correct model with nonnegative concentrations. The assumption implies that if consumption of a component occurs ($\sigma_{\boldsymbol{C}}^{(k,l)} < 0$) and its concentration reaches zero, then no more consumption is physically possible; see [15].

2.2.3. Bulk velocities and constitutive equations

The structure of the appearance of the constitutive assumptions in the sedimentation-compression model has been used frequently and are here briefly reviewed. We refer the reader to [9, 11] for detailed information. A feature of this structure is that it leaves two constitutive functions to be defined by the user: the hindered-settling function $v_{\text{hs}}(X)$ and the effective solids stress function $\sigma_{\text{e}}(X)$.

We define the characteristic function $\chi_\omega = 1$ if the statement ω is true, otherwise $\chi_\omega = 0$. The characteristic function for the mixture is thus $\gamma(z,t) = \chi_{\{\bar{z}(t) < z < B\}}(z)$. The bulk velocity of the mixture in the tank and below it in the underflow pipe is defined as $q(z,t) := q_{\text{u}}(t)\chi_{\{z > \bar{z}(t)\}}$ and the excess velocity (relative to the bulk velocity) due to sedimentation and compression is

$$v := v(X, \partial_z X, z, t) := \gamma(z,t)v_{\text{hs}}(X)\left(1 - \frac{\rho_X \sigma_{\text{e}}'(X)}{Xg\Delta\rho}\partial_z X\right) = \gamma(z,t)\big(v_{\text{hs}}(X) - \partial_z D(X)\big) \tag{2.8}$$

where

$$D(X) := \int_{X_c}^X d(s)\,\mathrm{d}s, \qquad d(X) := v_{\text{hs}}(X)\frac{\rho_X \sigma_{\text{e}}'(X)}{gX\Delta\rho}. \tag{2.9}$$

Here, $\Delta\rho := \rho_X - \rho_L$ is the density difference of the flocculated particles and the liquid, g is the acceleration of gravity, and $v_{\text{hs}}(X)$ is the hindered-settling velocity function, which is assumed to satisfy

$$v_{\text{hs}}(X) \begin{cases} > 0 & \text{for } X \in [0, \hat{X}), \\ = 0 & \text{for } X \geq \hat{X}, \end{cases} \tag{2.10}$$

where $\hat{X} < \rho_X$ is a maximum total solids concentration. By explicitly setting v_{hs} to zero for $X \geq \hat{X}$ we can ensure that the numerical solution for X indeed assumes values between zero and $\hat{X}$ only (see Lem. 3.3). The second constitutive function is the effective solids stress $\sigma_{\text{e}}(X)$, which satisfies

$$\sigma_{\text{e}}'(X) \begin{cases} = 0 & \text{for } X \leq X_{\text{c}}, \\ > 0 & \text{for } X > X_{\text{c}}, \end{cases}$$

where X_{c} is the critical concentration (mentioned in Sect. 1.1) above which the flocculent particles form a porous network that can bear a certain stress.

2.2.4. Balance laws

The balance law of each material component gives the system of PDEs

$$\begin{aligned}
\partial_t \boldsymbol{C} + \partial_z\big(\mathcal{V}_{\boldsymbol{C}}(X, \partial_z X, z, t)\boldsymbol{C}\big) &= \delta\big(z - \bar{z}(t)\big)q_{\text{f}}(t)\boldsymbol{C}_{\text{f}} + \gamma(z,t)\boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{C},\boldsymbol{S}), \\
\partial_t \boldsymbol{S} + \partial_z\big(\mathcal{V}_{\boldsymbol{S}}(X, \partial_z X, z, t)\boldsymbol{S}\big) &= \delta\big(z - \bar{z}(t)\big)q_{\text{f}}(t)\boldsymbol{S}_{\text{f}} + \gamma(z,t)\boldsymbol{R}_{\boldsymbol{S}}(\boldsymbol{C},\boldsymbol{S}),
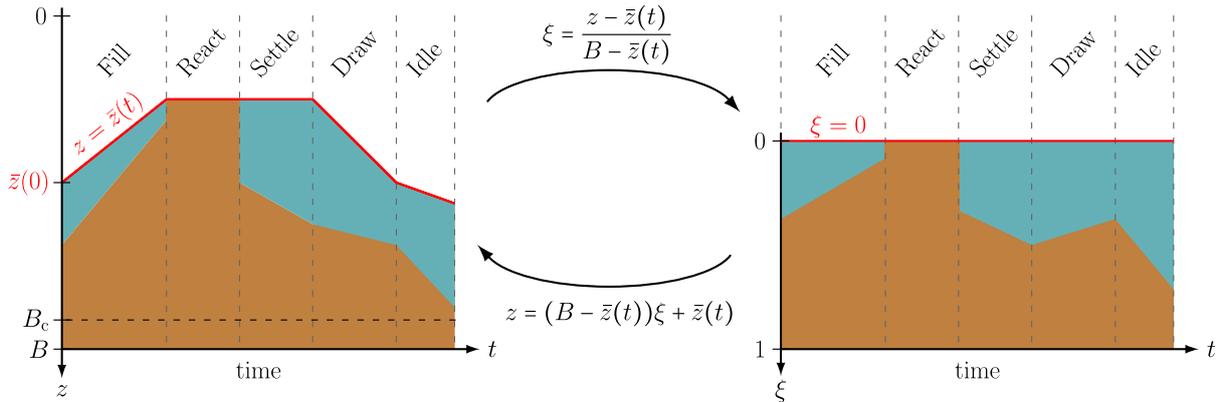\end{aligned} \tag{2.11}$$

FIGURE 3. Evolution of a concentration profile varying with respect to time at the five SBR stages in the space variable z (left) and new variable ξ (right). The moving boundary $z = \bar{z}(t)$ (red line) is mapped to the constant line $\xi = 0$. The mappings $z = z(\xi, t)$ and $\xi = \xi(z, t)$ are shown in the middle.

modelling reactive settling for $z \in \mathbb{R}$, where we have divided away the COD factors c etc. in each equation, $\delta(\cdot)$ [m^{-1}] is the delta symbol, and the total velocities are

$$\mathcal{V}_{\boldsymbol{C}} = \mathcal{V}_{\boldsymbol{C}}(X, \partial_z X, z, t) := q(z, t) + \gamma(z, t)\big(v_{\text{hs}}(X) - \partial_z D(X)\big),$$
$$\mathcal{V}_{\boldsymbol{S}} = \mathcal{V}_{\boldsymbol{S}}(X, \partial_z X, z, t) := q(z, t) - \gamma(z, t)\frac{(v_{\text{hs}}(X) - \partial_z D(X))X}{\rho_X - X},$$

where $\mathcal{V}_{\boldsymbol{C}}$ is the sum of the bulk velocity q and the excess velocity v in (2.8), whereas for the derivation of $\mathcal{V}_{\boldsymbol{S}}$ we refer to [15].

The pipe of extraction is modelled as a half line $x \geq 0$ (upwards) with $x = 0$ coupled at $z = \bar{z}(t)$. Any concentration is assumed to follow the advection equations

$$\partial_t \boldsymbol{C}_{\text{pipe}} + q_{\text{e}}(t)\partial_x \boldsymbol{C}_{\text{pipe}} = \boldsymbol{0}, \quad \partial_t \boldsymbol{S}_{\text{pipe}} + q_{\text{e}}(t)\partial_x \boldsymbol{S}_{\text{pipe}} = \boldsymbol{0}. \tag{2.12}$$

The extraction concentrations in the pipe are given by complicated formulas due to the moving boundary; see [15]. With the variable change suggested below, these will be obtained more easily.

2.2.5. Equations during mixing

During periods of mixing, the system of PDEs (2.11) reduces to the system of ODEs

$$\bar{V}(t)\boldsymbol{C}'(t) = \big(Q_{\text{u}}(t) - \bar{Q}(t)\big)\boldsymbol{C} + Q_{\text{f}}(t)\boldsymbol{C}_{\text{f}}(t) + \bar{V}(t)\boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{C}, \boldsymbol{S}),$$
$$\bar{V}(t)\boldsymbol{S}'(t) = \big(Q_{\text{u}}(t) - \bar{Q}(t)\big)\boldsymbol{S} + Q_{\text{f}}(t)\boldsymbol{S}_{\text{f}}(t) + \bar{V}(t)\boldsymbol{R}_{\boldsymbol{S}}(\boldsymbol{C}, \boldsymbol{S})$$

for the homogeneous concentrations in $\bar{z}(t) < z < B$, where all concentrations depend only on time since they are averages in the tank. Mixing occurs usually during the react stage when there is no in- or outflow; however, our model is general and allows mixing also during such operations. In the region $z < \bar{z}(t)$ all concentrations are zero, whereas the outlet concentrations are $\boldsymbol{C}_{\text{u}}(t) = \boldsymbol{C}(t)$ and $\boldsymbol{C}_{\text{e}}(t) = \boldsymbol{C}(t)$ (if $Q_{\text{e}}(t) > 0$) (analogously for $\boldsymbol{S}$).

2.3. Model equations on a fixed domain

To solve the model equations (2.11)–(2.12) without complicated formulas for the outlet concentrations, we transform the time-varying interval $[\bar{z}(t), B]$ to the fixed domain $[0, 1]$ for all $t \geq 0$ by introducing the space

variable, for all $z \in \mathbb{R}$,

$$\xi = \xi(z,t) := \frac{z - \bar{z}(t)}{B - \bar{z}(t)} \qquad \Leftrightarrow \qquad z = (B - \bar{z}(t))\xi + \bar{z}(t), \tag{2.13}$$

where it is assumed that $\bar{z}(t) \geq B_c > 0$ for all $t \geq 0$ for some constant B_c. We define the unknowns of the model in the new variable ξ by $\tilde{X}(\xi(z,t),t) := X(z,t)$ (analogously for the rest of the unknowns and space dependent functions). The partial derivatives are

$$\partial_t \xi = -\bar{z}'(t) \frac{1 - \xi}{B - \bar{z}(t)} =: \alpha(\xi,t), \qquad \partial_z \xi = \frac{1}{B - \bar{z}(t)} =: \beta(t), \tag{2.14}$$

Clearly, the sign of α depends uniquely on the slope $\bar{z}'$ and therefore on t, while $\beta(t) > 0$ for all $t > 0$. Then the time and space partial derivatives are transformed as

$$\partial_t X = \partial_\xi \tilde{X} \partial_t \xi + \partial_t \tilde{X} = \alpha(\xi,t) \partial_\xi \tilde{X} + \partial_t \tilde{X}, \quad \partial_z X = \partial_\xi \tilde{X} \partial_z \xi = \beta(t) \partial_\xi \tilde{X} = \partial_\xi \big(\beta(t) \tilde{X} \big).$$

It will be convenient to rewrite the term

$$\alpha(\xi,t) \partial_\xi \tilde{X} = \partial_\xi \big(\alpha(\xi,t) \tilde{X} \big) - \partial_\xi \alpha(\xi,t) \tilde{X} = \partial_\xi \big(\alpha(\xi,t) \tilde{X} \big) - \beta(t) \bar{z}'(t) \tilde{X}$$

(anticipating the transformation of $\partial_t \boldsymbol{C}$ in (2.11) etc.). The characteristic function becomes

$$\gamma(z,t) = \chi_{\{\bar{z}(t) < z < B\}}(z) = \tilde{\chi}_{\{0 < \xi < 1\}}(\xi) =: \tilde{\gamma}(\xi).$$

The delta symbol in (2.11) is formally transformed *via* the Heaviside function H as

$$H\big(z - \bar{z}(t)\big) = H\big(\xi(B - \bar{z}(t))\big) =: \tilde{H}(\xi),$$
$$\delta\big(z - \bar{z}(t)\big) = H'\big(z - \bar{z}(t)\big) = d/dz\, \tilde{H}(\xi) = \tilde{H}'(\xi)\beta(t) = \tilde{\delta}(\xi)\beta(t).$$

The bulk velocity becomes $q(z,t) = q_u(t)\chi_{\{z > \bar{z}(t)\}}(z) = q_u(t)\tilde{H}(\xi)$. Equation (2.11) can be written in the new variable ξ as the system

$$\begin{aligned}
\partial_t \boldsymbol{C} + \partial_\xi \big(\tilde{\mathcal{V}}_{\boldsymbol{C}}(X, \partial_\xi X, \xi, t)\boldsymbol{C} \big) &= \beta(t)\bar{z}'(t)\boldsymbol{C} + \delta(\xi)\beta(t)q_f(t)\boldsymbol{C}_f + \gamma(\xi)\boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{C}, \boldsymbol{S}), \\
\partial_t \boldsymbol{S} + \partial_\xi \big(\tilde{\mathcal{V}}_{\boldsymbol{S}}(X, \partial_\xi X, \xi, t)\boldsymbol{S} \big) &= \beta(t)\bar{z}'(t)\boldsymbol{S} + \delta(\xi)\beta(t)q_f(t)\boldsymbol{S}_f + \gamma(\xi)\boldsymbol{R}_{\boldsymbol{S}}(\boldsymbol{C}, \boldsymbol{S}),
\end{aligned} \tag{2.15}$$

where we directly have removed the tilde above $\boldsymbol{C}$, $\boldsymbol{S}$, X, γ, H, and δ, and

$$\tilde{\mathcal{V}}_{\boldsymbol{C}}(X, \partial_\xi X, \xi, t) := \tilde{q}(\xi, t) + \gamma(\xi)\beta(t)\big(v_{hs}(X) - \beta(t)\partial_\xi D(X)\big),$$
$$\tilde{\mathcal{V}}_{\boldsymbol{S}}(X, \partial_\xi X, \xi, t) := \tilde{q}(\xi, t) - \gamma(\xi)\beta(t)\frac{f(X) - \beta(t)\partial_\xi \mathcal{D}(X)}{\rho_X - X},$$

where $\tilde{q}(\xi, t) := \alpha(\xi, t) + \beta(t)q_u(t)\tilde{H}(\xi)$ (temporary definition),

$$f(X) := v_{hs}(X)X, \tag{2.16}$$

and

$$\mathcal{D}(X) := \int_{X_c}^X a(s)\, ds, \quad \text{where} \quad a(s) := sd(s). \tag{2.17}$$

Clearly, the governing equation (2.15) can be expressed in the form (1.1) if we define

$$U_{\boldsymbol{C}}(X, \xi, t) := \tilde{q}(\xi, t) + \gamma(\xi)\beta(t)v_{hs}(X), \quad U_{\boldsymbol{S}}(X, \xi, t) := \tilde{q}(\xi, t) - \gamma(\xi)\beta(t)\frac{f(X)}{\rho_X - X}. \tag{2.18}$$

Equations (2.15) hold for $\xi \in \mathbb{R}$ when $t \in T_f$ if it is assumed that all concentrations are zero above the surface $\xi < 0$. In particular, for the underflow zone $z > B \Leftrightarrow \xi > 1$, the equations are

$$\partial_t \boldsymbol{C} + \big(\alpha(\xi,t) + \beta(t)q_u(t)\big)\partial_\xi \boldsymbol{C} = \beta(t)\bar{z}'(t)\boldsymbol{C}, \quad \partial_t \boldsymbol{S} + \big(\alpha(\xi,t) + \beta(t)q_u(t)\big)\partial_\xi \boldsymbol{S} = \beta(t)\bar{z}'(t)\boldsymbol{S}.$$

To ensure that fluxes have correct units across the surface during extraction ($t \in T_e$), we also need to transform the extraction pipe. The pipe is originally modelled by the upwards-pointing x-axis with bulk flow upwards $q_e(t)$ and coupled to the z-axis by $-x = z - \bar{z}(t)$. Consequently, the transformation for the extraction pipe is

$$\xi(x,t) := \frac{-x}{B - \bar{z}(t)}, \tag{2.19}$$

and $\boldsymbol{C}_{\text{pipe}}(x,t) = \tilde{\boldsymbol{C}}(\xi(x,t),t)$. Then we get

$$\partial_t \boldsymbol{C}_{\text{pipe}} = -\partial_\xi \tilde{\boldsymbol{C}}\xi\bar{z}'(t)\beta(t) + \partial_t \tilde{\boldsymbol{C}} = -\big(\partial_\xi(\xi\tilde{\boldsymbol{C}}) - \tilde{\boldsymbol{C}}\big)\bar{z}'(t)\beta(t) + \partial_t \tilde{\boldsymbol{C}}, \quad \partial_x \boldsymbol{C}_{\text{pipe}} = -\partial_\xi \tilde{\boldsymbol{C}}\beta(t).$$

Equation (2.12) are transformed to (we immediately remove the tildes)

$$\partial_t \boldsymbol{C} - \partial_\xi \big(\beta(t)(\xi\bar{z}'(t) + q_e(t))\boldsymbol{C}\big) = -\bar{z}'(t)\beta(t)\boldsymbol{C},$$
$$\partial_t \boldsymbol{S} - \partial_\xi \big(\beta(t)(\xi\bar{z}'(t) + q_e(t))\boldsymbol{S}\big) = -\bar{z}'(t)\beta(t)\boldsymbol{S}, \quad \xi < 0.$$

(In comparison to (2.15), there is a minus sign on the right-hand side here.) With the bulk velocity redefined as

$$\tilde{q}(\xi,t) := \begin{cases} 0 & \text{if } \xi < 0 \text{ and } q_e(t) = 0, \\ -\beta(t)(\xi\bar{z}'(t) + q_e(t)) = -\beta(t)(\xi(q_u(t) + q_e(t)) + q_e(t)) & \text{if } \xi < 0 \text{ and } q_e(t) > 0, \\ \alpha(\xi,t) + \beta(t)q_u(t) = \beta(t)(\xi\bar{z}'(t) + q_u(t) - \bar{z}'(t)) & \text{if } \xi > 0, \end{cases}$$

where we recall that $\xi < 0$ corresponds to the pipe and $\xi > 0$ to the interior of the mixture in the tank, we thus get the governing equations

$$\partial_t \boldsymbol{C} + \partial_\xi \big(\tilde{\mathcal{V}}_{\boldsymbol{C}}(X, \partial_\xi X, \xi, t)\boldsymbol{C}\big) = \text{sgn}(\xi)\bar{z}'(t)\beta(t)\boldsymbol{C} + \delta(\xi)\beta(t)q_f(t)\boldsymbol{C}_f + \gamma(\xi)\boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{C}, \boldsymbol{S}),$$
$$\partial_t \boldsymbol{S} + \partial_\xi \big(\tilde{\mathcal{V}}_{\boldsymbol{S}}(X, \partial_\xi X, \xi, t)\boldsymbol{S}\big) = \text{sgn}(\xi)\bar{z}'(t)\beta(t)\boldsymbol{S} + \delta(\xi)\beta(t)q_f(t)\boldsymbol{S}_f + \gamma(\xi)\boldsymbol{R}_{\boldsymbol{S}}(\boldsymbol{C}, \boldsymbol{S}) \tag{2.20}$$

for $\xi \in \mathbb{R}$ and $t \in T_e$. These equations hold for all $t > 0$ if we for $t \in T_f$ define all concentrations in $\xi < 0$ to be zero; then the system is reduced to (2.15). The salient point of these transformations is that the outlet concentrations are now simply defined by

$$\begin{aligned} \boldsymbol{C}_e(t) &= \boldsymbol{C}(0^-, t), & \boldsymbol{S}_e(t) &= \boldsymbol{S}(0^-, t) & \text{if } t \in T_e, \\ \boldsymbol{C}_u(t) &= \boldsymbol{C}(1^+, t), & \boldsymbol{S}_u(t) &= \boldsymbol{S}(1^+, t) & \text{if } t \in T_f \text{ and } Q_u(t) > 0. \end{aligned} \tag{2.21}$$

For other times the outlet concentrations are defined to be zero.

During periods of mixing and $t \in T_f$, the following ODEs for time-dependent concentrations $\boldsymbol{C}(t)$ and $\boldsymbol{S}(t)$ are obtained by averaging equation (2.15), *i.e.*, integrating over the interval $[0^-, 1)$ when the convective and diffusive terms are zero:

$$\begin{aligned} \boldsymbol{C}'(t) &= \beta(t)\bar{z}'(t)\boldsymbol{C} + \beta(t)q_f(t)\boldsymbol{C}_f(t) + \boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{C}, \boldsymbol{S}), \\ \boldsymbol{S}'(t) &= \beta(t)\bar{z}'(t)\boldsymbol{S} + \beta(t)q_f(t)\boldsymbol{S}_f(t) + \boldsymbol{R}_{\boldsymbol{S}}(\boldsymbol{C}, \boldsymbol{S}). \end{aligned} \tag{2.22}$$

The same ODEs are indeed obtained for $t \in T_e$ (then $\boldsymbol{C}_f(t) = \boldsymbol{S}_f(t) = 0$).

2.4. Model equations with percentages

The restrictive part of the explicit CFL condition of an explicit scheme is due to the second-order derivative term containing the function $D(X)$, which depends only on the scalar variable X. The governing model will be rewritten so that the convective and diffusive terms are clearly seen. Furthermore, to establish boundedness on the total particulate concentration, $0 \leq X \leq \hat{X}$, we will rewrite the model in terms of percentages $\boldsymbol{p}$; see (2.5). We define the flux due to bulk movement and sedimentation, and the reaction term of the total solids concentration X by

$$F(\xi, t, X) := \tilde{q}(\xi, t)X + \gamma(\xi)\beta(t)f(X), \qquad R(\boldsymbol{C}, \boldsymbol{S}) := c\sum_{k=1}^{k_{\boldsymbol{C}}} R_{\boldsymbol{C}}^{(k)}(\boldsymbol{C}, \boldsymbol{S}). \tag{2.23}$$

Then we can write

$$\tilde{\mathcal{V}}_{\boldsymbol{C}}X = F(\xi, t, X) - \gamma(\xi)\beta(t)^2\partial_\xi\mathcal{D}(X), \quad \tilde{\mathcal{V}}_{\boldsymbol{S}} = \frac{\tilde{q}(\xi, t)\rho_X - F(\xi, t, X) + \gamma(\xi)\beta(t)^2\partial_\xi\mathcal{D}(X)}{\rho_X - X}.$$

By first multiplying the first equation of (2.20) by c and adding the vector components corresponding to (2.4), one obtains the scalar equation

$$\begin{aligned}
\partial_t X + \partial_\xi F(\xi, t, X) &= \partial_\xi\big(\gamma(\xi)\beta(t)^2\partial_\xi\mathcal{D}(X)\big) \\
&\quad + \beta(t)\,\mathrm{sgn}(\xi)\bar{z}'(t)X + \beta(t)\delta(\xi)q_{\mathrm{f}}(t)X_{\mathrm{f}} + \gamma(\xi)R(\boldsymbol{p}X/c, \boldsymbol{S})
\end{aligned} \tag{2.24}$$

for X. The concentrations $\boldsymbol{C} = \boldsymbol{p}X/c$ and $\boldsymbol{S}$ are given by the system (2.20), which we now can rewrite with the unknowns X, $\boldsymbol{p}$ and $\boldsymbol{S}$:

$$\begin{aligned}
\partial_t(\boldsymbol{p}X) + \partial_\xi(F(\xi, t, X)\boldsymbol{p}) &= \partial_\xi\big(\boldsymbol{p}\gamma(\xi)\beta(t)^2\partial_\xi\mathcal{D}(X)\big) + \beta(t)\,\mathrm{sgn}(\xi)\bar{z}'(t)\boldsymbol{p}X \\
&\quad + \beta(t)\delta(\xi)q_{\mathrm{f}}(t)\boldsymbol{p}_{\mathrm{f}}X_{\mathrm{f}} + \gamma(\xi)c\boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{p}X/c, \boldsymbol{S}),
\end{aligned} \tag{2.25}$$

$$\partial_t\boldsymbol{S} + \partial_\xi\big(\tilde{\mathcal{V}}_{\boldsymbol{S}}\boldsymbol{S}\big) = \beta(t)\,\mathrm{sgn}(\xi)\bar{z}'(t)\boldsymbol{S} + \beta(t)\delta(\xi)q_{\mathrm{f}}(t)\boldsymbol{S}_{\mathrm{f}} + \gamma(\xi)\boldsymbol{R}_{\boldsymbol{S}}(\boldsymbol{p}X/c, \boldsymbol{S}). \tag{2.26}$$

Not all equations in (2.25) need to be solved; one may solve only the first $k_{\boldsymbol{C}} - 1$ ones and set $p^{(k_{\boldsymbol{C}})} = 1 - (p^{(1)} + \cdots + p^{(k_{\boldsymbol{C}}-1)})$. The output concentrations for $\boldsymbol{p}$ and X are defined as in (2.21). The mixing ODEs (2.22) are converted analogously:

$$\begin{aligned}
X'(t) &= \beta(t)\bar{z}'(t)X + \beta(t)q_{\mathrm{f}}(t)X_{\mathrm{f}}(t) + R(\boldsymbol{p}X/c, \boldsymbol{S}), \\
(\boldsymbol{p}X)'(t) &= \beta(t)\bar{z}'(t)\boldsymbol{p}X + \beta(t)q_{\mathrm{f}}(t)\boldsymbol{p}_{\mathrm{f}}(t)X_{\mathrm{f}}(t) + c\boldsymbol{R}_{\boldsymbol{C}}(\boldsymbol{p}X/c, \boldsymbol{S}), \\
\boldsymbol{S}'(t) &= \beta(t)\bar{z}'(t)\boldsymbol{S} + \beta(t)q_{\mathrm{f}}(t)\boldsymbol{S}_{\mathrm{f}}(t) + \boldsymbol{R}_{\boldsymbol{S}}(\boldsymbol{p}X/c, \boldsymbol{S}).
\end{aligned} \tag{2.27}$$

Finally, we comment that by the representation using percentages, we have replaced the system of $k_{\boldsymbol{C}}$ coupled nonlinear, degenerate convection-diffusion PDEs (1.1a) by one nonlinear scalar PDE of that type for X, namely (2.24), plus $k_{\boldsymbol{C}}$ PDEs for the components of $\boldsymbol{p}$, (2.25). The salient point is that if X is known, then the convection-diffusion part of (2.25) is linear. This observation suggests a numerical scheme in which in each time step, one first updates X by solving numerically the scalar equation (2.24), and then uses the updated X-value to solve numerically the system (2.25), which then becomes a linear convection-reaction equation. This procedure is more efficient than solving the coupled system (1.1a) in each time step, and is pursued in Section 3.

3. Explicit numerical scheme

3.1. Discretization in space and time

To discretize the PDE system (2.24)–(2.26), we define $\Delta\xi := 1/(N + 1/2)$ for an integer N, $\xi_j := j\Delta\xi$, and let cell j be the interval $I_j := [\xi_{j-1/2}, \xi_{j+1/2}]$; see Figure 4.
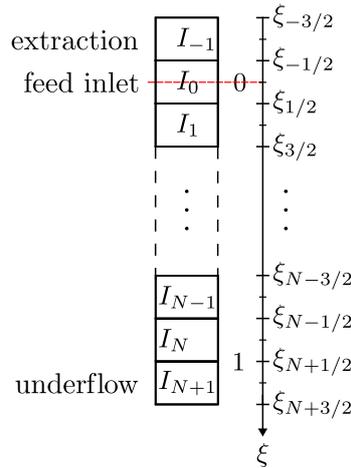
FIGURE 4. Schematic of the division of the computational domain into cells.

Thus, the feed inlet is located in the middle of I_0, which makes the numerical fluxes at the cell boundaries easy to define. The bottom of the tank is located at $\xi = \xi_{N+1/2} = (N + 1/2)\Delta\xi = 1$. Given a total simulation time T and the total number of discrete time points N_T, we define the time step $\tau := T/N_T$, which is supposed to satisfy a suitable CFL condition, and the discrete time points $t^n = n\tau$, $n = 0, 1, \ldots, N_T$. The total particulate concentration is denoted by $X_j^n \approx X(\xi_j, t^n)$, and analogous notation is used for other concentrations. The underflow concentration is captured by an additional cell $j = N+1$ below the tank. The extraction concentrations can be obtained in an analogous way in an additional cell I_{-1}. For both the explicit and implicit schemes the outlet concentrations are defined by $\boldsymbol{p}_{\mathrm{e}} = \boldsymbol{p}_{-1}^n$ and $\boldsymbol{p}_{\mathrm{u}} = \boldsymbol{p}_{N+1}^n$, and similarly for X and $\boldsymbol{S}$.

3.2. Explicit scheme

For ease of notation, we introduce $a^+ := \max\{0, a\}$, $a^- := \min\{0, a\}$, and the upwind and and numerical divergence operators

$$\mathrm{Upw}(a; b, c) := \max\{a, 0\}b + \min\{a, 0\}c = a^+ b + a^- c \qquad \text{for } a, b, c \in \mathbb{R},$$
$$[\Delta\boldsymbol{\Phi}]_j^n := \boldsymbol{\Phi}_{j+1/2}^n - \boldsymbol{\Phi}_{j-1/2}^n.$$

(The upwind discretization is applied to discretize the linear convective terms within (2.24)–(2.26); the nonlinear convective terms are discretized by the Engquist-Osher (generalized upwind) flux defined in (3.3).) We define the Kronecker delta and the characteristic function for the mixture in the tank by

$$\delta_{j,0} := \int_{I_j} \delta(\xi) \, \mathrm{d}\xi = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j \neq 0, \end{cases}$$

$$\gamma_j := \begin{cases} \gamma(\xi_j) & \text{if } j \neq 0, \\ \frac{1}{2} & \text{if } j = 0 \end{cases} = \begin{cases} 0 & \text{if } j < 0 \text{ or } j > N, \\ \frac{1}{2} & \text{if } j = 0, \\ 1 & \text{if } j = 1, \ldots, N, \end{cases} \tag{3.1}$$

$$\gamma_{j+1/2} := \gamma(\xi_{j+1/2}) = \begin{cases} 0 & \text{if } j \leq 0 \text{ or } j \geq N, \\ 1 & \text{if } j = 1, \ldots, N-1. \end{cases}$$

The term $\mathrm{sgn}(\xi)\beta(t)\bar{z}'(t)X(\xi,t)$ in (2.24) is approximated by the average

$$\frac{1}{\Delta\xi}\int_{\xi_{-1/2}}^{\xi_{1/2}}\mathrm{sgn}(\xi)\beta^n(\bar{z}')^n X(\xi,t^n)\,\mathrm{d}\xi = \begin{cases} 0 & \text{if } q_\mathrm{e}^n > 0, \\ \frac{1}{2}\beta^n(\bar{z}')^n X_0^n = \frac{1}{2}\beta^n(q_\mathrm{u}^n - q_\mathrm{f}^n)X_0^n & \text{if } q_\mathrm{e}^n = 0 \end{cases}$$

for $j = 0$, where we recall that $X(\xi,t) = 0$ for $\xi < 0$ if $q_\mathrm{e}(t) = 0$, and analogously for $\boldsymbol{S}$ instead of X. For the numerical update formulas, we define

$$\kappa_j^n := \begin{cases} 1 - \tau\beta^n(\bar{z}')^n = 1 - \tau\beta^n(q_\mathrm{u}^n + q_\mathrm{e}^n) & \text{if } j < 0 \text{ and } q_\mathrm{e}^n > 0, \\ 1 & \text{if } j < 0 \text{ and } q_\mathrm{e}^n = 0 \text{ or } j = 0 \text{ and } q_\mathrm{e}^n > 0, \\ 1 - \frac{\tau}{2}\beta^n(q_\mathrm{u}^n - q_\mathrm{f}^n) & \text{if } j = 0 \text{ and } q_\mathrm{e}^n = 0, \\ 1 + \tau\beta^n(\bar{z}')^n & \text{if } j > 0. \end{cases} \tag{3.2}$$

Finally, we define

$$\tilde{q}_{j+1/2}^n := \tilde{q}(\xi_{j+1/2},t^n) = \begin{cases} 0 & \text{if } j = -2,-1 \text{ and } q_\mathrm{e}^n = 0, \\ -\beta^n(\xi_{j+1/2}(q_\mathrm{u}^n + q_\mathrm{e}^n) + q_\mathrm{e}^n) & \text{if } j = -2,-1 \text{ and } q_\mathrm{e}^n > 0, \\ \alpha_{j+1/2}^n + \beta^n q_\mathrm{u}^n & \text{if } j = 0,\ldots,N+1. \end{cases}$$

An explicit approximation of the system (2.24)–(2.26) is now obtained by following ideas from [13, 14]. We assume that the function f defined by (2.16) has a unique maximum at $X^* \in (0,\hat{X})$. The Engquist-Osher numerical flux [27]

$$\mathcal{E}_{j+1/2}^n := \gamma_{j+1/2}\left(f(0) + \int_0^{X_j^n}\max\{0,f'(s)\}\,\mathrm{d}s + \int_0^{X_{j+1}^n}\min\{0,f'(s)\}\,\mathrm{d}s\right) \tag{3.3}$$

is used, which for a unimodal flux function f is

$$\mathcal{E}_{j+1/2}^n(X_j^n, X_{j+1}^n) = \gamma_{j+1/2}\begin{cases} f(X_j^n) & \text{if } X_j^n, X_{j+1}^n \leq X^*, \\ f(X^*) & \text{if } X_{j+1}^n \leq X^* < X_j^n, \\ -f(X^*) + f(X_j^n) + f(X_{j+1}^n) & \text{if } X_j^n \leq X^* < X_{j+1}^n, \\ f(X_{j+1}^n) & \text{if } X^* < X_j^n, X_{j+1}^n. \end{cases}$$

The diffusive term and the other fluxes are discretized by

$$\mathcal{J}_{j+1/2}^n := \gamma_{j+1/2}\frac{(\beta^n)^2}{\Delta\xi}\big(\mathcal{D}(X_{j+1}^n) - \mathcal{D}(X_j^n)\big), \quad j = -2,\ldots,N+1, \tag{3.4}$$

$$\mathcal{B}_{j+1/2}^n := \mathrm{Upw}(\tilde{q}_{j+1/2}^n; X_j^n, X_{j+1}^n), \tag{3.5}$$

$$\begin{aligned} \mathcal{F}_{j+1/2}^n &:= \mathcal{B}_{j+1/2}^n + \beta^n\mathcal{E}_{j+1/2}^n \\ &= \begin{cases} 0 & \text{if } j = -2,-1 \text{ and } q_\mathrm{e}^n = 0, \\ -\beta^n\big(\xi_{j+1/2}(q_\mathrm{u}^n + q_\mathrm{e}^n) + q_\mathrm{e}^n\big)X_{j+1}^n & \text{if } j = -2,-1 \text{ and } q_\mathrm{e}^n > 0, \\ \mathrm{Upw}(\alpha_{j+1/2}^n + \beta^n q_\mathrm{u}^n; X_j^n, X_{j+1}^n) + \beta^n\mathcal{E}_{j+1/2}^n & \text{if } j = 0,\ldots,N-1, \\ (\alpha_{j+1/2}^n + \beta^n q_\mathrm{u}^n)X_j^n & \text{if } j = N, N+1, \end{cases} \end{aligned} \tag{3.6}$$

$$\Phi_{j+1/2}^n := \mathcal{F}_{j+1/2}^n - \mathcal{J}_{j+1/2}^n. \tag{3.7}$$

We have $\alpha_{N+1/2} = 0$, and when $q_\mathrm{e} > 0$, we assume that $\Delta\xi$ is sufficiently small, so that all fluxes at the top $\xi = \xi_{-1/2}$ and bottom $\xi_{N+1/2}$ are directed out of the tank, which means that no boundary values are needed. The numerical fluxes are then defined, for $j = -2,\ldots,N+1$, by

$$\begin{aligned} \boldsymbol{\Phi}_{\boldsymbol{p},j+1/2}^n &:= \mathrm{Upw}(\Phi_{j+1/2}^n; \boldsymbol{p}_j^n, \boldsymbol{p}_{j+1}^n), \\ \boldsymbol{\Phi}_{\boldsymbol{S},j+1/2}^n &:= \mathrm{Upw}\left(\rho_X\tilde{q}_{j+1/2}^n - \Phi_{j+1/2}^n; \frac{\boldsymbol{S}_j^n}{\rho_X - X_j^n}, \frac{\boldsymbol{S}_{j+1}^n}{\rho_X - X_{j+1}^n}\right). \end{aligned} \tag{3.8}$$

With an Euler time step, $\lambda := \tau/\Delta\xi$, and $\mu := \tau/\Delta\xi^2$, we can formulate the explicit scheme as follows. For the boundary layers $j = -1$ and $j = N+1$, we set

$$X_{-1}^{n+1} = 0, \quad \boldsymbol{p}_{-1}^{n+1} = \boldsymbol{p}_0^{n+1}, \quad \boldsymbol{S}_{-1}^{n+1} = \boldsymbol{0} \quad \text{if } q_{\mathrm{e}}^n = 0, \tag{3.9}$$

$$X_{N+1}^{n+1} = 0, \quad \boldsymbol{p}_{N+1}^{n+1} = \boldsymbol{p}_N^{n+1}, \quad \boldsymbol{S}_{N+1}^{n+1} = \boldsymbol{0} \quad \text{if } q_{\mathrm{u}}^n = 0. \tag{3.10}$$

Otherwise, we have for $j = -1, \dots, N+1$,

$$X_j^{n+1} = \kappa_j^n X_j^n - \lambda[\Delta\mathcal{F}]_j^n + \lambda[\Delta\mathcal{J}]_j^n + \lambda\delta_{j,0}\beta^n q_{\mathrm{f}}^n X_{\mathrm{f}}^n + \tau\gamma_j R(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n), \tag{3.11}$$

$$\boldsymbol{p}_j^{n+1} X_j^{n+1} = \kappa_j^n \boldsymbol{p}_j^n X_j^n - \lambda[\Delta\boldsymbol{\Phi_p}]_j^n + \lambda\delta_{j,0}\beta^n q_{\mathrm{f}}^n \boldsymbol{p}_{\mathrm{f}}^n X_{\mathrm{f}}^n + \tau\gamma_j c \boldsymbol{R_C}(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n), \tag{3.12}$$

$$\boldsymbol{S}_j^{n+1} = \kappa_j^n \boldsymbol{S}_j^n - \lambda[\Delta\boldsymbol{\Phi_S}]_j^n + \lambda\delta_{j,0}\beta^n q_{\mathrm{f}}^n \boldsymbol{S}_{\mathrm{f}}^n + \tau\gamma_j \boldsymbol{R_S}(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n). \tag{3.13}$$

Scheme (3.11) is solved first, then the others. If $X_j^{n+1} = 0$, then the value of $\boldsymbol{p}_j^{n+1}$ is irrelevant, and can be set to $\boldsymbol{p}_j^{n+1} := \boldsymbol{p}_j^n$. For the cells outside the mixture, the scheme for X is reduced to the following. If $q_{\mathrm{e}}^n = 0$, then (3.9) handles $j = -1$; otherwise, we define, for simplicity of writing,

$$q_{\mathrm{out}}^n := q_{\mathrm{u}}^n + q_{\mathrm{e}}^n$$

and utilize

$$X_{-1}^{n+1} = \left(1 - \tau\beta^n q_{\mathrm{out}}^n\right)X_{-1}^n + \lambda\beta^n\left((\xi_{-1/2}q_{\mathrm{out}}^n + q_{\mathrm{e}}^n)X_0^n - (\xi_{-3/2}q_{\mathrm{out}}^n + q_{\mathrm{e}}^n)X_{-1}^n\right). \tag{3.14}$$

On the other hand, if $q_{\mathrm{u}}^n = 0$, then (3.10) is in effect for $j = N+1$; otherwise,

$$X_{N+1}^{n+1} = (1 + \tau\beta^n q_{\mathrm{out}}^n)X_{N+1}^n - \lambda\left((\alpha_{N+3/2}^n + \beta^n q_{\mathrm{u}}^n)X_{N+1}^n - \beta^n q_{\mathrm{u}}^n X_N^n\right).$$

Similar update formulas hold for $\boldsymbol{p}$ and $\boldsymbol{S}$. The resulting approximate concentrations are transformed back to the original z- and x-coordinates *via* (2.13) and (2.19), respectively.

3.3. Numerics during mixing

Suppose a (PDE or numerical) solution $X(\xi, T_0)$ (or $p^{(k)}(\xi, T_0)$ or $S^{(k)}(\xi, T_0)$) is known at $t = T_0 = t^{n_0}$ when a period of complete mixing starts. The initial concentrations for the ODEs (2.27) are defined as the averages (analogously for $p^{(k)}$ and $S^{(k)}$)

$$X(T_0) := \int_0^1 X(\xi, T_0)\,\mathrm{d}\xi \approx \Delta\xi\left(\frac{X_0^{n_0}}{2} + X_1^{n_0} + \cdots + X_N^{n_0}\right).$$

The ODE system (2.27) can then be integrated in time with an Euler step. If an ODE mixing period ends at $t = t^n$ with the values $X(t^n)$, and the PDE model is to be simulated thereafter, then the value of each component k is allocated to all cells in the tank; $X_j^n := X(t^n)$, $j = 0, \dots, N$.

3.4. CFL condition

We denote the solution variables by $\boldsymbol{\mathcal{U}} := (X, \boldsymbol{p}^{\mathrm{T}}, \boldsymbol{S}^{\mathrm{T}})^{\mathrm{T}}$ and define the constants

$$\|f\| := \max_{0 \le X \le \hat{X}} |f(X)|, \qquad\qquad M_R := \frac{1}{c}\sup_{\substack{\boldsymbol{\mathcal{U}} \in \Omega \\ 1 \le k \le k_{\boldsymbol{C}}}} \left|\frac{\partial R}{\partial C^{(k)}}\right|,$$

$$M_{q1} := \max_{0 \le t \le T}\left(\max\{q_{\mathrm{u}}(t) + q_{\mathrm{e}}(t), q_{\mathrm{f}}(t)\}\right), \qquad C_1 := \zeta\left(M_{q2} + \|f'\|\right), \quad C_2 := \zeta^2\|a\|,$$

$$M_{q2} := \max_{0 \le t \le T}\left(\max\{q_{\mathrm{f}}(t), q_{\mathrm{e}}(t)\} + 2q_{\mathrm{u}}(t)\right), \quad M_{\boldsymbol{C}} := \sup_{\substack{\boldsymbol{\mathcal{U}} \in \Omega, \\ 1 \le k \le k_{\boldsymbol{C}}}} \sum_{l \in I_{\boldsymbol{C},k}^-} |\sigma_{\boldsymbol{C}}^{(k,l)}| \bar{r}_{\boldsymbol{C}}^{(l)}(\boldsymbol{C}, \boldsymbol{S}),$$

$$\zeta := \frac{1}{B - B_{\mathrm{c}}}, \qquad\qquad M_{\boldsymbol{S}} := \sup_{\substack{\boldsymbol{\mathcal{U}} \in \Omega, \\ 1 \le k \le k_{\boldsymbol{S}}}} \sum_{l \in I_{\boldsymbol{S},k}^-} |\sigma_{\boldsymbol{S}}^{(k,l)}| \bar{r}_{\boldsymbol{S}}^{(l)}(\boldsymbol{C}, \boldsymbol{S}).$$

The time step τ and the spatial mesh width $\Delta\xi$ should satisfy the CFL condition

$$\tau\left(\zeta M_{q1} + \max\{M_R, M_C, M_S\} + \frac{2}{\Delta\xi}\max\left\{C_1 + \frac{C_2}{\Delta\xi}, \frac{1}{\rho_X - \hat{X}}\left(\zeta\rho_X + C_1\hat{X} + \frac{C_2\hat{X}}{\Delta\xi}\right)\right\}\right) \leq 1. \qquad \text{(CFL)}$$

Condition (CFL) expresses a limitation of the time step τ for a given spatial discretization $\Delta\xi$ that ensures that the whole scheme is monotone and an invariant region principle holds (see Lems. 3.3 to 3.6 and their proofs). Satisfaction of these properties requires all ingredients of the governing PDEs be taken into account.

3.5. Monotonicity and invariant region property

In what follows, we demonstrate that the solution variables $\boldsymbol{\mathcal{U}} = (X, \boldsymbol{p}^{\mathrm{T}}, \boldsymbol{S}^{\mathrm{T}})^{\mathrm{T}}$ produced by the explicit numerical scheme stay in the set

$$\Omega := \left\{\boldsymbol{\mathcal{U}} \in \mathbb{R}^{1+k_C+k_S} : 0 \leq X \leq \hat{X}, \ \boldsymbol{p} \geq \boldsymbol{0}, \ p^{(1)} + \cdots + p^{(k_C)} = 1, \ \boldsymbol{S} \geq \boldsymbol{0}\right\}$$

provided that this property holds for the initial values. Our proofs rely on the following lemma, which follows directly from the definition (3.3).

Lemma 3.1. *Assume that $0 \leq X_j \leq \hat{X}$ for all j. Then the Engquist-Osher flux $\mathcal{E}_{j+1/2} = \mathcal{E}_{j+1/2}(X_j, X_{j+1})$ applied on the unimodal function $0 \leq f \in C^1$ satisfies*

$$-\|f'\| \leq \frac{\partial\mathcal{E}_{j+1/2}}{\partial X_{j+1}} \leq 0 \leq \frac{\partial\mathcal{E}_{j+1/2}}{\partial X_j} \leq \|f'\|, \quad \left|\frac{\partial[\Delta\mathcal{E}]_j}{\partial X_j}\right| \leq \|f'\|, \quad \frac{\mathcal{E}_{j+1/2}}{X_j} \leq \|f'\|, \quad \frac{\mathcal{E}_{j+1/2}}{X_{j+1}} \leq \|f'\|.$$

Lemma 3.2. *If $\boldsymbol{\mathcal{U}}_j^n \in \Omega$ for all j, then the following estimates hold for $j = -1, \ldots, N+1$:*

$$\kappa_j^n \geq 1 - \zeta\tau M_{q1}, \tag{3.15}$$

$$\left|\tilde{q}_{j+1/2}^n\right| \leq \zeta\left(\max\{q_{\mathrm{f}}^n, q_{\mathrm{e}}^n\} + 2q_{\mathrm{u}}^n\right) \leq \zeta M_{q2}, \tag{3.16}$$

$$\frac{\partial[\Delta\mathcal{B}]_j^n}{\partial X_k^n}\begin{cases} = -\tilde{q}_{j-1/2}^{n,+} \leq 0 & \text{if } k = j-1, \\ \leq \zeta M_{q2}, & \text{if } k = j, \\ = \tilde{q}_{j+1/2}^{n,-} \leq 0, & \text{if } k = j+1, \\ = 0 & \text{otherwise,} \end{cases} \tag{3.17}$$

$$\left|\frac{\partial}{\partial X_k}R(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n)\right|\begin{cases} \leq M_R & \text{if } k = j, \\ = 0 & \text{if } k \neq j, \end{cases} \tag{3.18}$$

$$\max\left\{\Phi_{j+1/2}^{n,+}, -\Phi_{j-1/2}^{n,-}\right\} \leq \left(\zeta(M_{q2} + \|f'\|) + \frac{\zeta^2\|a\|}{\Delta\xi}\right)X_j^n. \tag{3.19}$$

Proof. By the definition of the transformation of variables,

$$-q_{\mathrm{f}}(t) \leq \bar{z}'(t) \leq q_{\mathrm{u}}(t) + q_{\mathrm{e}}(t), \qquad -\zeta(q_{\mathrm{u}}(t) + q_{\mathrm{e}}(t)) \leq \alpha(\xi, t) \leq \zeta q_{\mathrm{f}}(t), \qquad 0 < \beta(t) \leq \zeta.$$

These inclusions directly imply (3.15) since

$$\kappa_j^n \geq 1 - \tau\zeta\max\{q_{\mathrm{u}}^n + q_{\mathrm{e}}^n, q_{\mathrm{f}}^n\} \geq 1 - \tau\zeta M_{q1}.$$

To prove (3.16) we observe that when $q_{\mathrm{e}}^n = 0$, then $\tilde{q}_{-1/2}^n = \tilde{q}_{-3/2}^n = 0$, and when $q_{\mathrm{e}}^n > 0$ and $j = -1, -2$,

$$\tilde{q}_{j+1/2}^n = -\beta^n\left(\xi_{j+1/2}(q_{\mathrm{u}}^n + q_{\mathrm{e}}^n) + q_{\mathrm{e}}^n\right) \geq -\zeta q_{\mathrm{e}}^n. \tag{3.20}$$

For $j = 0, \ldots, N$, (we recall that if $q_\mathrm{f}^n > 0$, then $q_\mathrm{e}^n = 0$ and vice versa)

$$\tilde{q}_{j+1/2}^n = \alpha_{j+1/2}^n + \beta^n q_\mathrm{u}^n = \beta^n \left(-(\bar{z}')^n(1 - \xi_{j+1/2}) + q_\mathrm{u}^n\right) \begin{cases} \leq \zeta(q_\mathrm{f}^n + q_\mathrm{u}^n) & \text{if } q_\mathrm{e}^n = 0, \\ \geq -\zeta(q_\mathrm{u}^n + q_\mathrm{e}^n) & \text{if } q_\mathrm{e}^n > 0, \end{cases} \qquad (3.21)$$

and for $j = N + 1$,

$$\tilde{q}_{N+3/2}^n = \frac{(\bar{z}')^n \Delta\xi + q_\mathrm{u}^n}{B - \bar{z}^n} \leq \zeta(q_\mathrm{out}^n \Delta\xi + q_\mathrm{u}^n) \leq \zeta\left((1 + \Delta\xi)q_\mathrm{u}^n + q_\mathrm{e}^n\right). \qquad (3.22)$$

From (3.20) to (3.22) we now deduce (3.16). Next, computing the difference of the flux

$$\mathcal{B}_{j+1/2}^n = \tilde{q}_{j+1/2}^{n,+} X_j^n + \tilde{q}_{j+1/2}^{n,-} X_{j+1}^n$$

and differentiating this expression with respect to X_k^n we obtain

$$\frac{\partial [\Delta\mathcal{B}]_j^n}{\partial X_k^n} = \frac{\partial}{\partial X_k^n} \left(\tilde{q}_{j+1/2}^{n,+} X_j^n + \tilde{q}_{j+1/2}^{n,-} X_{j+1}^n - \tilde{q}_{j-1/2}^{n,+} X_{j-1}^n - \tilde{q}_{j-1/2}^{n,-} X_j^n\right)$$

$$= \begin{cases} -\tilde{q}_{j-1/2}^{n,+} \leq 0 & \text{if } k = j - 1, \\ \tilde{q}_{j+1/2}^{n,+} - \tilde{q}_{j-1/2}^{n,-} \leq \zeta\left((1 + \Delta\xi)q_\mathrm{u}^n + \max\{q_\mathrm{e}^n, q_\mathrm{f}^n\}\right) \leq \zeta M_{q2} & \text{if } k = j, \\ \tilde{q}_{j+1/2}^{n,-} \leq 0 & \text{if } k = j + 1, \\ 0 & \text{otherwise.} \end{cases}$$

This proves (3.17). For the reaction term, the cases $k \neq j$ are trivial. Assuming that $k = j$ and differentiating we obtain

$$\left|\frac{\partial}{\partial X_j} R(\boldsymbol{p}_j^n X_j^n / c, \boldsymbol{S}_j^n)\right| = \frac{1}{c} \left|(\boldsymbol{p}_j^n)^\mathrm{T} \nabla_{\boldsymbol{C}} R\right| \leq \frac{1}{c} \sum_{i=1}^{k_{\boldsymbol{C}}} p_j^{(i),n} \left|\frac{\partial R}{\partial C_j^{(i),n}}\right| = M_R \sum_{i=1}^{k_{\boldsymbol{C}}} p_j^{(i),n} = M_R,$$

which implies (3.18). Finally, (3.19) follows from

$$\Phi_{j+1/2}^{n,+} = \left(\mathcal{B}_{j+1/2}^n + \beta^n \mathcal{E}_{j+1/2}^n - \mathcal{J}_{j+1/2}^n\right)^+ \leq \tilde{q}_{j+1/2}^{n,+} X_j^n + \zeta\|f'\| X_j^n + \gamma_{j+1/2} \frac{(\beta^n)^2}{\Delta\xi} \mathcal{D}(X_j^n)$$

$$\leq \left(\zeta\left(M_{q2} + \|f'\|\right) + \zeta^2 \frac{\|a\|}{\Delta\xi}\right) X_j^n,$$

$$-\Phi_{j-1/2}^{n,-} = -\left(\mathcal{B}_{j-1/2}^n + \beta^n \mathcal{E}_{j-1/2}^n - \mathcal{J}_{j-1/2}^n\right)^- \leq \tilde{q}_{j-1/2}^{n,+} X_j^n + \zeta\|f'\| X_j^n + \gamma_{j+1/2} \frac{(\beta^n)^2}{\Delta\xi} \mathcal{D}(X_j^n)$$

$$\leq \left(\zeta\left(M_{q2} + \|f'\|\right) + \frac{\zeta^2 \|a\|}{\Delta\xi}\right) X_j^n.$$

$$\square$$

Lemma 3.3. *If $\mathcal{U}_j^n \in \Omega$ for all j and (CFL) is in effect, then $0 \leq X_j^{n+1} \leq \hat{X}$ for all j.*

Proof. We write the update formula (3.11) for $j = -1, \ldots, N + 1$ as

$$X_j^{n+1} = \mathcal{H}_X^n\left(X_{j-1}^n, X_j^n, X_{j+1}^n\right)$$

and we shall show that $\mathcal{H}_X^n$ is a monotone function in each of its variables. We recall that

$$\lambda[\Delta\mathcal{J}]_j^n = \mu(\beta^n)^2 \left(\gamma_{j+1/2}\left(\mathcal{D}(X_{j+1}^{n+1}) - \mathcal{D}(X_j^{n+1})\right) - \gamma_{j-1/2}\left(\mathcal{D}(X_j^{n+1}) - \mathcal{D}(X_{j-1}^{n+1})\right)\right),$$

$a = \mathcal{D}'$, and we differentiate to obtain, by means of (CFL) and Lems. 3.1 and 3.2,

$$\frac{\partial X_j^{n+1}}{\partial X_{j-1}^n} = -\lambda \frac{\partial [\Delta \mathcal{B}]_j^n}{\partial X_{j-1}^n} - \lambda \beta^n \frac{\partial [\Delta \mathcal{E}]_j^n}{\partial X_{j-1}^n} + \lambda \frac{\partial [\Delta \mathcal{J}]_j^n}{\partial X_{j-1}^n}$$

$$= \lambda \tilde{q}_{j-1/2}^{n,+} + \lambda \beta^n \frac{\partial \mathcal{E}_{j-1/2}}{\partial X_{j-1}^n} + \mu(\beta^n)^2 \gamma_{j-1/2} a(X_{j-1}^n) \geq 0,$$

$$\frac{\partial X_j^{n+1}}{\partial X_j^n} = \kappa^n - \lambda \frac{\partial [\Delta \mathcal{B}]_j^n}{\partial X_j^n} - \lambda \beta^n \frac{\partial [\Delta \mathcal{E}]_j^n}{\partial X_j^n} + \lambda \frac{\partial [\Delta \mathcal{J}]_j^n}{\partial X_j^n} + \frac{\partial}{\partial X_j} R(\boldsymbol{p}_j^n X_j^n / c, \boldsymbol{S}_j^n)$$

$$\geq 1 - \tau \zeta M_{q1} - \lambda \zeta M_{q2} - \zeta \lambda \|f'\| - \mu \zeta^2 2\|a\| - \tau M_R \geq 0,$$

$$\frac{\partial X_j^{n+1}}{\partial X_{j+1}^n} = -\lambda \frac{\partial [\Delta \mathcal{B}]_j^n}{\partial X_{j+1}^n} - \lambda \beta^n \frac{\partial [\Delta \mathcal{E}]_j^n}{\partial X_{j+1}^n} = -\lambda \tilde{q}_{j+1/2}^{n,-} - \lambda \beta^n \frac{\partial \mathcal{E}_{j+1/2}}{\partial X_{j+1}^n} \geq 0.$$

The proved monotonicity of $\mathcal{H}_X^n$ and the assumptions (2.6) imply that for $j \neq 0$,

$$0 = \mathcal{H}_X^n(0,0,0) \leq X_j^{n+1} = \mathcal{H}_X^n(X_{j-1}^n, X_j^n, X_{j+1}^n) \leq \mathcal{H}_X^n(\hat{X}, \hat{X}, \hat{X})$$

$$= \kappa_j^n \hat{X} - \lambda(\tilde{q}_{j+1/2}^n - \tilde{q}_{j-1/2}^n)\hat{X}$$

$$= \begin{cases} \hat{X}\big(1 - \tau \beta^n q_{\text{out}}^n - \lambda \beta^n(-(\xi_{-1/2} q_{\text{out}}^n + q_{\text{e}}^n) + (\xi_{-3/2} q_{\text{out}}^n + q_{\text{e}}^n))\big) \\ \quad = \hat{X}\big(1 - \tau \beta^n q_{\text{out}}^n - \lambda \beta^n(\frac{\Delta \xi}{2} q_{\text{out}}^n - \frac{3\Delta \xi}{2} q_{\text{out}}^n)\big) = \hat{X} & \text{if } j = -1 \text{ and } q_{\text{e}}^n > 0, \\ \hat{X} & \text{if } j = -1 \text{ and } q_{\text{e}}^n = 0, \\ \hat{X}\big(1 + \tau \beta^n (\bar{z}')^n \\ \quad - \lambda \beta^n(-(\bar{z}')^n(1 - \xi_{j+1/2}) + (\bar{z}')^n(1 - \xi_{j-1/2}))\big) = \hat{X} & \text{if } j \geq 1. \end{cases}$$

For $j = 0$ we have, since $\alpha(\xi_{1/2}, t) = -\bar{z}'(t)(1 - \Delta \xi/2)\beta(t)$, and assuming $X_{\text{f}}^n \leq \hat{X}$,

$$0 \leq \lambda \beta^n X_{\text{f}}^n q_{\text{f}}^n = \mathcal{H}_X^n(0,0,0) \leq X_0^{n+1} = \mathcal{H}_X^n(X_{-1}^n, X_0^n, X_1^n) \leq \mathcal{H}_X^n(\hat{X}, \hat{X}, \hat{X})$$

$$= \begin{cases} \kappa_0^n \hat{X} - \lambda\big(\alpha_{1/2}^n + \beta^n q_{\text{u}}^n + \beta^n(-\frac{\Delta \xi}{2} q_{\text{out}}^n + q_{\text{e}}^n)\big)\hat{X} \\ \quad = \hat{X}\big(1 - \tau \beta^n q_{\text{out}}^n - \lambda(-q_{\text{out}}^n(1 - \frac{\Delta \xi}{2})\beta^n + \beta^n q_{\text{u}}^n + \beta^n(-\frac{\Delta \xi}{2} q_{\text{out}}^n + q_{\text{e}}^n))\big) \\ \quad = \hat{X}(1 - \tau \beta^n q_{\text{out}}^n) \leq \hat{X} & \text{if } q_{\text{e}}^n > 0, \\ \kappa_0^n \hat{X} - \lambda\big(\alpha_{1/2}^n + \beta^n q_{\text{u}}^n)\big)\hat{X} + \lambda \beta^n X_{\text{f}}^n q_{\text{f}}^n \\ \quad = (1 - \frac{\tau}{2}\beta^n q_{\text{out}}^n)\hat{X} - \lambda\big(-(q_{\text{u}}^n - q_{\text{f}}^n)(1 - \frac{\Delta \xi}{2})\beta^n + \beta^n q_{\text{u}}^n)\big)\hat{X} + \lambda \beta^n X_{\text{f}}^n q_{\text{f}}^n \\ \quad = \hat{X} - \lambda\big(-(q_{\text{u}}^n - q_{\text{f}}^n)\beta^n + \beta^n q_{\text{u}}^n\big)\hat{X} + \lambda \beta^n X_{\text{f}}^n q_{\text{f}}^n \\ \quad = \hat{X} - \lambda \beta^n(\hat{X} - X_{\text{f}}^n)q_{\text{f}}^n \leq \hat{X} & \text{if } q_{\text{e}}^n = 0. \end{cases}$$

<div align="right">□</div>

Lemma 3.4. *If* $\boldsymbol{\mathcal{U}}_j^n \in \Omega$ *for all* j *and* (CFL) *holds, then*

$$p_j^{(k),n+1} \geq 0 \quad \text{for all } k = 1, \ldots, k_{\boldsymbol{C}} \text{ and all } j. \tag{3.23}$$

Proof. If $X_j^{n+1} = 0$, we define

$$p_j^{(k),n+1} := p_j^{(k),n} \in [0,1] \quad \text{for all } k = 1, \ldots, k_{\boldsymbol{C}}. \tag{3.24}$$

If $X_j^{n+1} > 0$, we have for each $k \in \{1, \dots, k_{\boldsymbol{C}}\}$

$$
\begin{aligned}
p_j^{(k),n+1} X_j^{n+1} &= \kappa_j^n p_j^{(k),n} X_j^n - \lambda\big(\Phi_{j+1/2}^{n,+} p_j^{(k),n} + \Phi_{j+1/2}^{n,-} p_{j+1}^{(k),n} - \Phi_{j-1/2}^{n,+} p_{j-1}^{(k),n} - \Phi_{j-1/2}^{n,-} p_j^{(k),n}\big) \\
&\quad + \lambda\delta_{j,0}\beta^n q_{\mathrm{f}}^n p_{\mathrm{f}}^{(k),n} X_{\mathrm{f}}^n + \tau\gamma_j c R_j^{(k),n}(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n) \\
&\geq (1 - \zeta\tau M_{q1})\, p_j^{(k),n} X_j^n - 2\lambda\left(\zeta(M_{q2} + \|f'\|) + \frac{\zeta^2\|a\|}{\Delta\xi}\right) p_j^{(k),n} X_j^n \\
&\quad + \tau c \sum_{l \in I_{\boldsymbol{C},k}^-} \sigma_{\boldsymbol{C}}^{(k,l)} \bar{r}^{(l)}\big(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n\big) p_j^{(k),n} X_j^n/c \\
&\geq \left(1 - \zeta(\tau M_{q1}) - \frac{2\tau}{\Delta\xi}\left(\zeta(M_{q2} + \|f'\|) + \frac{\zeta^2\|a\|}{\Delta\xi}\right) + \tau M_{\boldsymbol{C}}\right) p_j^{(k),n} X_j^n \geq 0.
\end{aligned}
$$

This implies (3.23). $\qquad\square$

Lemma 3.5. *If $\mathcal{U}_j^n \in \Omega$ for all j and* (CFL) *holds, then*

$$
p_j^{(1),n+1} + \cdots + p_j^{(k_{\boldsymbol{C}}),n+1} = 1 \quad \text{for all } j. \tag{3.25}
$$

Proof. If $X_j^{n+1} = 0$, then by (3.24),

$$
p_j^{(1),n+1} + \cdots + p_j^{(k_{\boldsymbol{C}}),n+1} = p_j^{(1),n} + \cdots + p_j^{(k_{\boldsymbol{C}}),n} = 1,
$$

so let us assume that $X_j^{n+1} > 0$. We sum up all equations in (3.12) and utilize that

$$
p_j^{(1),n} + \cdots + p_j^{(k_{\boldsymbol{C}},n)} = 1
$$

along with

$$
\begin{aligned}
\sum_{k=1}^{k_{\boldsymbol{C}}} [\Delta\Phi_{\boldsymbol{p}}^{(k)}]_j^n &= \sum_{k=1}^{k_{\boldsymbol{C}}} \big(\mathrm{Upw}(\Phi_{\boldsymbol{p},j+1/2}^{(k),n}, p_j^{(k),n}, p_{j+1}^{(k),n}) - \mathrm{Upw}(\Phi_{\boldsymbol{p},j-1/2}^{(k),n}, p_j^{(k),n}, p_{j+1}^{(k),n})\big) \\
&= \sum_{k=1}^{k_{\boldsymbol{C}}} \big(\Phi_{j+1/2}^{n,+} p_j^{(k),n} + \Phi_{j+1/2}^{n,-} p_{j+1}^{(k),n} - \Phi_{j-1/2}^{n,+} p_{j-1}^{(k),n} - \Phi_{j-1/2}^{n,-} p_j^{(k),n}\big) \\
&= \Phi_{j+1/2}^{n,+} + \Phi_{j+1/2}^{n,-} - \Phi_{j-1/2}^{n,+} - \Phi_{j-1/2}^{n,-} = [\Delta\Phi]_j^n.
\end{aligned}
$$

Then the sum of the equations in (3.12) is

$$
X_j^{n+1} \sum_{k=1}^{k_{\boldsymbol{C}}} p_j^{(k),n+1} = \kappa_j^n X_j^n - \lambda[\Delta\Phi]_j^n + \lambda\delta_{j,0}\beta^n q_{\mathrm{f}}^n X_{\mathrm{f}}^n + \tau\gamma_j R(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n). \tag{3.26}
$$

The right-hand side is identical to that of (3.11). Hence, subtracting (3.26) from (3.11) we get

$$
X_j^{n+1}\big(1 - (p_j^{(1),n+1} + \cdots + p_j^{(k_{\boldsymbol{C}}),n+1})\big) = 0,
$$

which proves the desired result (3.25). $\qquad\square$

Lemma 3.6. *If $\mathcal{U}_j^n \in \Omega$ for all j and* (CFL) *holds, then*

$$
S_j^{(k),n+1} \geq 0 \quad \text{for all } k = 1, \dots, k_{\boldsymbol{S}} \text{ and all } j.
$$

Proof. From the update formula for a component $S_j^{(k),n}$ of $\boldsymbol{S}_j^n$ we get

$$
\begin{aligned}
S_j^{(k),n+1} = {}& \kappa_j^n S_j^{(k),n} - \lambda\bigg(\frac{S_j^{(k),n}}{\rho_X - X_j^n}(\rho_X \tilde{q}_{j+1/2}^n - \Phi_{j+1/2}^n)^+ + \frac{S_{j+1}^{(k),n}}{\rho_X - X_{j+1}^n}(\rho_X \tilde{q}_{j+1/2}^n - \Phi_{j+1/2}^n)^- \\
& - \frac{S_{j-1}^{(k),n}}{\rho_X - X_{j-1}^n}(\rho_X \tilde{q}_{j-1/2}^n - \Phi_{j-1/2}^n)^+ - \frac{S_j^{(k),n}}{\rho_X - X_j^n}(\rho_X \tilde{q}_{j-1/2}^n - \Phi_{j-1/2}^n)^- \bigg) \\
& + \lambda\delta_{j,0}\beta^n q_{\mathrm{f}}^n S_{\mathrm{f}}^{(k),n} + \tau\gamma_j R_{\boldsymbol{S},j}^{(k),n}(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n) \\
\geq {}& (1 - \tau\zeta M_{q1})\, S_j^{(k),n} - \frac{\lambda S_j^{(k),n}}{\rho_X - X_j^n}\big((\rho_X \tilde{q}_{j+1/2}^n - \Phi_{j+1/2}^n)^+ - (\rho_X \tilde{q}_{j-1/2}^n - \Phi_{j-1/2}^n)^-\big) \\
& + \tau \sum_{l \in I_{\boldsymbol{S},k}^-} \sigma_{\boldsymbol{S}}^{(k,l)} \bar{r}^{(l)}(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n) S_j^{(k),n} \\
\geq {}& \bigg(1 - \tau\zeta M_{q1} - \frac{2\tau\zeta}{\Delta\xi(\rho_X - \hat{X})}\bigg((\rho_X + \hat{X})M_{q2} + \|f'\|\hat{X} + \frac{\zeta\|a\|\hat{X}}{\Delta\xi}\bigg) + \tau M_C\bigg) S_j^{(k),n} \geq 0.
\end{aligned}
$$

$\square$

4. A SEMI-IMPLICIT SCHEME

4.1. Semi-implicit scheme for the update of X

To obtain a semi-implicit scheme, we write out several terms in (3.11)–(3.13) and evaluate those containing the coefficient $\mu = \tau/\Delta\xi^2$ at time t^{n+1}. Then (3.11) becomes

$$
\begin{aligned}
X_j^{n+1} = {}& \kappa_j^n X_j^n - \lambda[\Delta\mathcal{F}]_j^n \\
& + (\beta^n)^2 \mu\big(\gamma_{j+1/2}(\mathcal{D}(X_{j+1}^{n+1}) - \mathcal{D}(X_j^{n+1})) - \gamma_{j-1/2}(\mathcal{D}(X_j^{n+1}) - \mathcal{D}(X_{j-1}^{n+1}))\big) \\
& + \lambda\delta_{j,0}\beta^n q_{\mathrm{f}}^n X_{\mathrm{f}}^n + \tau\gamma_j R(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n).
\end{aligned}
\tag{4.1}
$$

For $j = -1$ and $j = N + 1$, many terms are zero and this formula is in fact explicit; see (3.14) and (3.2). For $j = 0, \ldots, N$, one has to solve a system of $(N + 1)$ nonlinear equations. For the further analysis, it is useful to rewrite (4.1) as a two-step implicit-explicit scheme:

(1) Given X_j^n for $j = -1, \ldots, N+1$, calculate $\tilde{X}_j^{n+1}$ from

$$
\tilde{X}_j^{n+1} = \kappa_j^n X_j^n - \lambda[\Delta\mathcal{F}]_j^n + \lambda\delta_{j,0}\beta^n q_{\mathrm{f}}^n X_{\mathrm{f}}^n + \tau\gamma_j R(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n), \quad j = -1, \ldots, N+1.
\tag{4.2}
$$

(2) Let

$$
X_{-1}^{n+1} = \tilde{X}_{-1}^{n+1}, \quad X_{N+1}^{n+1} = \tilde{X}_{N+1}^{n+1},
\tag{4.3}
$$

and compute $\boldsymbol{X}^{n+1} = (X_0^{n+1}, \ldots, X_N^{n+1})^{\mathrm{T}}$ by solving the nonlinear system of equations

$$
\boldsymbol{X}^{n+1} + (\beta^n)^2 \mu\boldsymbol{T} \begin{pmatrix} \mathcal{D}(X_0^{n+1}) \\ \mathcal{D}(X_1^{n+1}) \\ \vdots \\ \mathcal{D}(X_N^{n+1}) \end{pmatrix} = \tilde{\boldsymbol{X}}^{n+1}, \quad \boldsymbol{T} := \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix},
\tag{4.4}
$$

where $\tilde{\boldsymbol{X}}^{n+1} := (\tilde{X}_0^{n+1}, \ldots, \tilde{X}_N^{n+1})^{\mathrm{T}}$.

In what follows, we assume that the CFL condition for the semi-implicit scheme

$$\tau\left(\zeta M_{q1} + \max\{M_R, M_C, M_S\} + \frac{2}{\Delta\xi}\max\left\{C_1, \frac{\zeta\rho_X + C_1\hat{X}}{\rho_X - \hat{X}}\right\}\right) \leq 1 \qquad \text{(CFL-SI)}$$

is in effect. Notice that (CFL-SI) arises from (CFL) by setting $C_2 = 0$, and therefore (CFL-SI) stipulates a bound just on $\tau/\Delta\xi$ as $\Delta\xi \to 0$, but in contrast to (CFL), not on $\tau/\Delta\xi^2$. Thus, the semi-implicit scheme allows one to utilize a much larger time step τ (for a given spatial discretization $\Delta\xi$) than the explicit scheme. As we will see in Example 2 of Section 5, this property even implies a gain of efficiency.

To prove that the system (4.4) has a unique solution at all, we follow a strategy similar to that of [10]: we first assume that (4.4) has a solution and show that the scheme is monotone and satisfies an invariant region property. We then invoke a topological degree argument to show that (4.4) indeed has a solution, and show that it depends Lipschitz continuously on the solution at the previous time step. As a consequence, the whole scheme (4.2)–(4.4) is well defined.

Lemma 4.1. *Assume that τ and $\Delta\xi$ satisfy (CFL-SI). Then the scheme (4.2)–(4.4) is monotone, i.e. there exist functions $\mathcal{K}_j^n$, $j = -1, \ldots, N+1$, such that*

$$X_j^{n+1} = \mathcal{K}_j^n\big(X_{-1}^n, X_0^n, X_1^n, \ldots, X_N^n, X_{N+1}^n, t^n\big), \quad j = -1, \ldots, N+1,$$

that are monotone in each X-argument.

Proof. By repeating the monotonicity part of the proof of Lemma 3.3 we see that under condition (CFL-SI), the scheme (4.2) is monotone, *i.e.*, $\partial\tilde{X}_j^{n+1}/\partial X_k^n \geq 0$ for all $-1 \leq j, k \leq N+1$, so by (4.3) the statement immediately holds for $j = -1$ and $j = N+1$. In what follows, we define for a vector $\boldsymbol{X} = (X_0, \ldots, X_N)^{\mathrm{T}}$ the Jacobian matrix of the left-hand side of (4.4), namely

$$\boldsymbol{J}(\boldsymbol{X}) := \boldsymbol{I}_{N+1} + (\beta^n)^2\mu\boldsymbol{T}\,\text{diag}\big(a(X_0), \ldots, a(X_N)\big). \qquad (4.5)$$

We wish to show that

$$\frac{\partial X_j^{n+1}}{\partial X_k^n} \geq 0 \quad \text{for all } 1 \leq k, j \leq n.$$

To this end we introduce for $k = 0, \ldots, N$ the vectors

$$\frac{\partial \boldsymbol{X}^{n+1}}{\partial X_k^n} := \left(\frac{\partial X_0^{n+1}}{\partial X_k^n}, \ldots, \frac{\partial X_N^{n+1}}{\partial X_k^n}\right)^{\mathrm{T}}, \quad \frac{\partial \tilde{\boldsymbol{X}}^n}{\partial X_k^n} := \left(\frac{\partial \tilde{X}_0^n}{\partial X_k^n}, \ldots, \frac{\partial \tilde{X}_N^n}{\partial X_k^n}\right)^{\mathrm{T}}$$

Assume that for given $\boldsymbol{X}^n$, the vector $\boldsymbol{X}^{n+1}$ is a solution to (4.1). Then

$$\boldsymbol{J}(\boldsymbol{X}^{n+1})\frac{\partial \boldsymbol{X}^{n+1}}{\partial X_k^n} = \frac{\partial \tilde{\boldsymbol{X}}^n}{\partial X_k^n}, \quad k = 0, \ldots, N.$$

We already know that $\partial\tilde{\boldsymbol{X}}^n/\partial X_k^n \geq \boldsymbol{0}$. On the other hand, for any $\boldsymbol{X}$, the matrix $(\boldsymbol{J}(\boldsymbol{X}))^{\mathrm{T}}$ is a strictly diagonally dominant L-matrix and therefore an M-matrix; in particular, $(\boldsymbol{J}(\boldsymbol{X}))^{\mathrm{T}}$ has a nonnegative inverse, and therefore also $(\boldsymbol{J}(\boldsymbol{X}))^{-1}$ is nonnegative, hence

$$\frac{\partial \boldsymbol{X}^{n+1}}{\partial X_k^n} = \big(\boldsymbol{J}(\boldsymbol{X}^{n+1})\big)^{-1}\frac{\partial \tilde{\boldsymbol{X}}^n}{\partial X_k^n} \geq \boldsymbol{0}, \quad k = 0, \ldots, N.$$

$\square$

Lemma 4.2. *If* $\boldsymbol{\mathcal{U}}_j^n \in \Omega$ *for all* j *and* (CFL-SI) *is in effect, then*

$$0 \le X_j^{n+1} \le \hat{X} \quad \text{for all } j = -1, \ldots, N+1. \tag{4.6}$$

Proof. Repeating the second part of the proof of Lemma 3.3 under the assumption $a \equiv 0$ we see that under condition (CFL-SI),

$$0 \le \tilde{X}_j^{n+1} \le \hat{X} \quad \text{for all } j = -1, \ldots, N+1. \tag{4.7}$$

This directly proves (4.6) for $j = -1$ and $j = N+1$. Furthermore, we define

$$\varpi_{j+1/2}^n := \begin{cases} (\mathcal{D}(X_{j+1}^n) - \mathcal{D}(X_j^n))/(X_{j+1}^n - X_j^n) & \text{if } X_{j+1}^n \neq X_j^n, \\ 0 & \text{otherwise,} \end{cases} \quad j = 0, \ldots, N-1,$$

and write the nonlinear scheme (4.4) as $\boldsymbol{M}(\boldsymbol{X}^{n+1})\boldsymbol{X}^{n+1} = \tilde{\boldsymbol{X}}^{n+1}$, where the entries $(m_{ij})_{0 \le i,j \le N}$ of the tridiagonal matrix $\boldsymbol{M} = \boldsymbol{M}(\boldsymbol{X}^{n+1})$ are given by

$$m_{j,j-1} = -(\beta^n)^2 \mu \varpi_{j-1/2}^{n+1}, \quad j = 1, \ldots, N,$$
$$m_{jj} = 1 + (\beta^n)^2 \mu \big(\varpi_{j-1/2}^{n+1} + \varpi_{j+1/2}^{n+1}\big), \quad j = 1, \ldots, N-1,$$
$$m_{j,j+1} = -(\beta^n)^2 \mu \varpi_{j+1/2}^{n+1}, \quad j = 0, \ldots, N-1,$$
$$m_{00} = 1 + (\beta^n)^2 \mu \varpi_{1/2}^{n+1}, \quad m_{NN} = 1 + (\beta^n)^2 \mu \varpi_{N-1/2}^{n+1}.$$

Since $\varpi_{j+1/2}^{n+1} \ge 0$ for all j, $\boldsymbol{M}$ is a strictly diagonally dominant L-matrix and therefore an M-matrix, that is $\boldsymbol{M}^{-1}$ exists and $\boldsymbol{M}^{-1} \ge \boldsymbol{0}$, *i.e.*, if we write $\boldsymbol{M}^{-1} = (\bar{m}_{jk})_{0 \le j,k \le N}$, then $\bar{m}_{jk} \ge 0$. Since

$$X_j^{n+1} = \bar{m}_{j,0} \tilde{X}_0^{n+1} + \bar{m}_{j,1} \tilde{X}_1^{n+1} + \cdots + \bar{m}_{j,N} \tilde{X}_N^{n+1}, \tag{4.8}$$

this property implies that $X_j^{n+1} \ge 0$. On the other hand, since

$$m_{jj} - \sum_{\substack{k=0 \\ k \neq j}}^{N} |m_{jk}| \ge 1 \quad \text{and} \quad \sum_{k=0}^{N} m_{jk} = 1 \quad \text{for all } j = 0, \ldots, N,$$

we have $\boldsymbol{M}\boldsymbol{1} = \boldsymbol{1}$ with $\boldsymbol{1} := (1, \ldots, 1)^{\mathrm{T}}$, hence $\boldsymbol{M}^{-1}\boldsymbol{1} = \boldsymbol{1}$, that is $\bar{m}_{j,0} + \bar{m}_{j,1} + \cdots + \bar{m}_{j,N} = 1$. In view of the upper bound in (4.7) we then deduce from (4.8) that $X_j^{n+1} \le \hat{X}$ for $j = 0, \ldots, N$. $\qquad \square$

In what follows, we define $\boldsymbol{\mathcal{X}} := (X_{-1}, X_0, X_1, \ldots, X_N, X_{N+1})^{\mathrm{T}} \in \mathbb{R}^{N+3}$. The following lemma and its proof closely follow [10], Lemma 3.3, part (a).

Lemma 4.3. *Assume that condition* (CFL-SI) *is in effect and that* $\boldsymbol{\mathcal{X}}^n \in [0, \hat{X}]^{N+3}$. *Then the scheme* (4.1), *or equivalently,* (4.2)–(4.4), *admits a solution* $\boldsymbol{\mathcal{X}}^{n+1} \in [0, \hat{X}]^{N+3}$.

Proof. The existence of $\boldsymbol{\mathcal{X}}^{n+1}$ follows by adopting an argument used in[29] to prove the existence of a solution of implicit schemes for hyperbolic equations based on topological degree theory [24]. To this end, let us write the scheme in the form

$$\boldsymbol{\mathcal{X}}^{n+1} - \boldsymbol{\mathcal{E}}(\boldsymbol{\mathcal{X}}^{n+1}, t^n) = \tilde{\boldsymbol{\mathcal{X}}}^{n+1}(\boldsymbol{\mathcal{X}}^n, t^n). \tag{4.9}$$

Here $\tilde{\boldsymbol{\mathcal{X}}}^{n+1}(\cdot, t^n) : \mathbb{R}^{N+3} \to \mathbb{R}^{N+3}$ is a continuous function defined by (4.2), and $\boldsymbol{\mathcal{E}}(\cdot, t^n) : \mathbb{R}^{N+3} \to \mathbb{R}^{N+3}$ is another continuous function defined in an obvious way by (4.3), (4.4). By Lemma 4.2, if $\boldsymbol{\mathcal{X}}^{n+1}$ satisfies (4.9),

then $\boldsymbol{\mathcal{X}}^{n+1} \in [0, \hat{X}]^{N+3}$. On the other hand we know (and have used that) if $a \equiv 0$ and hence $\mathcal{D} \equiv 0$, then the explicit scheme for the hyperbolic case

$$\boldsymbol{\mathcal{X}}^{n+1} = \tilde{\boldsymbol{\mathcal{X}}}^{n+1}(\boldsymbol{\mathcal{X}}^n, t^n), \tag{4.10}$$

which corresponds just to the scheme (4.2), satisfies the same bound, *i.e.*, $\tilde{\boldsymbol{\mathcal{X}}}^{n+1}(\boldsymbol{\mathcal{X}}^n, t^n) \in [0, \hat{X}]^{N+3}$. Consequently, if $B_R \subset \mathbb{R}^{N+3}$ is a ball with center $\mathbf{0}$ and sufficiently large radius R, then (4.9) has no solution on the boundary of B_R, and one can define the topological degree of the mapping $\mathrm{Id} - \boldsymbol{\mathcal{E}}$ associated with the set B_R and the point $\tilde{\boldsymbol{\mathcal{X}}}^{n+1}(\boldsymbol{\mathcal{X}}^n, t^n)$, that is, $\deg(\mathrm{Id} - \boldsymbol{\mathcal{E}}, B_R, \tilde{\boldsymbol{\mathcal{X}}}^{n+1}(\boldsymbol{\mathcal{X}}^n, t^n))$. Furthermore, if $\alpha \in [0, 1]$, then the same argument allows us to define $\deg(\mathrm{Id} - \alpha\boldsymbol{\mathcal{E}}, B_R, \tilde{\boldsymbol{\mathcal{X}}}^{n+1}(\boldsymbol{\mathcal{X}}^n, t^n))$. The property of invariance of degree under continuous transformations then asserts that the latter quantity does not depend on α, hence

$$\deg\big(\mathrm{Id} - \boldsymbol{\mathcal{E}}, B_R, \tilde{\boldsymbol{\mathcal{X}}}^{n+1}(\boldsymbol{\mathcal{X}}^n, t^n)\big) = \deg\big(\mathrm{Id}, B_R, \tilde{\boldsymbol{\mathcal{X}}}^{n+1}(\boldsymbol{\mathcal{X}}^n, t^n)\big) = 1,$$

where the equality for $\alpha = 0$ holds since we can solve the scheme (4.10) in a unique way. This proves that (4.9) has a solution in B_R, where we already have proved that the solution belongs to $[0, \hat{X}]^{N+3}$. $\square$

Lemma 4.4. *Assume that condition* (CFL-SI) *is in effect and that* $\boldsymbol{\mathcal{X}}^n, \boldsymbol{\mathcal{Y}}^n \in [0, \hat{X}]^{N+3}$, *and that* $\boldsymbol{\mathcal{X}}^{n+1}$ *and* $\boldsymbol{\mathcal{Y}}^{n+1}$ *are both computed by scheme* (4.1), *or equivalently,* (4.2)–(4.4)*. Then there exists a constant* $C > 0$ *such that*

$$\|\boldsymbol{\mathcal{X}}^{n+1} - \boldsymbol{\mathcal{Y}}^{n+1}\|_1 \le (1 + C\Delta t)\|\boldsymbol{\mathcal{X}}^n - \boldsymbol{\mathcal{Y}}^n\|_1. \tag{4.11}$$

This means that the solution of (4.2)–(4.4) *depends Lipschitz continuously on* $\boldsymbol{\mathcal{X}}^n$, *and in particular, setting* $\boldsymbol{\mathcal{X}}^n = \boldsymbol{\mathcal{Y}}^n$, *we obtain uniqueness.*

Proof. We define $\vartheta_j^n := Y_j^n - X_j^n$ for $j = -1, \dots, N+1$ and the quantities

$$\theta_j^{n+1} := \begin{cases} \mu(\beta^n)^2(\mathcal{D}(Y_j^{n+1}) - \mathcal{D}(X_j^{n+1}))/\vartheta_j^{n+1} \ge 0 & \text{if } \vartheta_j^{n+1} \neq 0, \\ 0 & \text{otherwise,} \end{cases} \quad j = -1, \dots, N+1.$$

Furthermore, we let $\mathcal{E}_{j+1/2}(\boldsymbol{X}^n)$ and $\mathcal{E}_{j+1/2}(\boldsymbol{Y}^n)$ be the Engquist-Osher numerical fluxes (3.3) applied to $\boldsymbol{X}^n$ and $\boldsymbol{Y}^n$, and similarly for $\mathcal{B}_{j+1/2}^n$ and $\mathcal{F}_{j+1/2}^n$. Clearly,

$$\begin{aligned} &\mathcal{F}_{j+1/2}^n(\boldsymbol{Y}^n) - \mathcal{F}_{j+1/2}^n(\boldsymbol{X}^n) \\ &= \mathcal{B}_{j+1/2}^n(\boldsymbol{Y}^n) - \mathcal{B}_{j+1/2}^n(\boldsymbol{X}^n) + \beta^n\big(\mathcal{E}_{j+1/2}^n(\boldsymbol{Y}^n) - \mathcal{E}_{j+1/2}^n(\boldsymbol{X}^n)\big) \\ &= \mathrm{Upw}\big(\tilde{q}_{j+1/2}^n; \vartheta_j^n, \vartheta_{j+1}^n\big) + \beta^n\gamma_{j+1/2}\bigg(\int_{X_j^n}^{Y_j^n}(f')^+(s)\,\mathrm{d}s + \int_{X_{j+1}^n}^{Y_{j+1}^n}(f')^-(s)\,\mathrm{d}s\bigg) \\ &= \eta_{j+1/2}^n\vartheta_j^n - \nu_{j+1/2}^n\vartheta_{j+1}^n, \end{aligned}$$

where we define

$$\eta_{j+1/2}^n := (\tilde{q}_{j+1/2}^n)^+ + \gamma_{j+1/2}\int_0^1 (f')^+\big(X_j^n + \sigma(Y_j^n - X_j^n)\big)\,\mathrm{d}\sigma \ge 0,$$

$$\nu_{j+1/2}^n := -\bigg((\tilde{q}_{j+1/2}^n)^- + \gamma_{j+1/2}\int_0^1 (f')^-\big(X_{j+1}^n + \sigma(Y_{j+1}^n - X_{j+1}^n)\big)\,\mathrm{d}\sigma\bigg) \ge 0,$$

hence

$$-\lambda\big([\Delta\mathcal{F}(\boldsymbol{Y})]_j^n - [\Delta\mathcal{F}(\boldsymbol{X})]_j^n\big) = -\lambda\big(\eta_{j+1/2}^n + \nu_{j-1/2}^n\big)\vartheta_j^n + \lambda\eta_{j-1/2}^n\vartheta_{j-1}^n + \lambda\nu_{j+1/2}^n\vartheta_{j+1}^n.$$

With this in mind, we obtain from (4.1) and the corresponding scheme for Y_j^n

$$
\begin{aligned}
\vartheta_{-1}^{n+1} &= \big(\kappa_{-1}^n - \lambda(\eta_{-1/2}^n + \nu_{-3/2}^n)\big)\vartheta_{-1}^n + \lambda\nu_{-1/2}^n\vartheta_0^n, \\
\big(1+\theta_0^{n+1}\big)\vartheta_0^{n+1} &= \big(\kappa_0^n - \lambda(\eta_{1/2}^n + \nu_{-1/2}^n)\big)\vartheta_0^n + \lambda\nu_{1/2}^n\vartheta_1^n + \lambda\eta_{-1/2}^n\vartheta_{-1}^n + \theta_1^{n+1}\vartheta_1^{n+1} \\
&\quad + \tau\gamma_0\big(R(\boldsymbol{p}_0^n Y_0^n/c, \boldsymbol{S}_0^n) - R(\boldsymbol{p}_0^n X_0^n/c, \boldsymbol{S}_0^n)\big), \\
\big(1+2\theta_j^{n+1}\big)\vartheta_j^{n+1} &= \big(\kappa_j^n - \lambda(\eta_{j+1/2}^n + \nu_{j-1/2}^n)\big)\vartheta_j^n + \lambda\nu_{j+1/2}^n\vartheta_{j+1}^n + \lambda\eta_{j-1/2}^n\vartheta_{j-1}^n \\
&\quad + \theta_{j+1}^{n+1}\vartheta_{j+1}^{n+1} + \theta_{j-1}^{n+1}\vartheta_{j-1}^{n+1} \\
&\quad + \tau\gamma_j\big(R(\boldsymbol{p}_j^n Y_j^n/c, \boldsymbol{S}_j^n) - R(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n)\big), \quad j=1,\dots,N-1, \\
\big(1+\theta_N^{n+1}\big)\vartheta_N^{n+1} &= \big(\kappa_N^n - \lambda(\eta_{N+1/2}^n + \nu_{N-1/2}^n)\big)\vartheta_N^n + \lambda\nu_{N+1/2}^n\vartheta_{N+1}^n + \lambda\eta_{N-1/2}^n\vartheta_{N-1}^n \\
&\quad + \theta_{N-1}^{n+1}\vartheta_{N-1}^{n+1} + \tau\gamma_N\big(R(\boldsymbol{p}_N^n Y_N^n/c, \boldsymbol{S}_N^n) - R(\boldsymbol{p}_N^n X_N^n/c, \boldsymbol{S}_N^n)\big), \\
\vartheta_{N+1}^{n+1} &= \big(\kappa_{N+1}^n - \lambda(\eta_{N+3/2}^n + \nu_{N+1/2}^n)\big)\vartheta_{N+1}^n + \lambda\eta_{N+1/2}^n\vartheta_N^n.
\end{aligned}
\tag{4.12}
$$

Finally, we define

$$
\rho_j^n := \int_0^1 \partial_\sigma R\Big((\boldsymbol{p}_j^n/c)\big(X_j^n + \sigma(Y_j^n - X_j^n)\big), \boldsymbol{S}_j^n\Big)\,\mathrm{d}\sigma, \quad j=0,\dots,N,
$$

such that

$$
\tau\gamma_j\big(R(\boldsymbol{p}_j^n Y_j^n/c, \boldsymbol{S}_j^n) - R(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n)\big) = \tau\gamma_j\rho_j^n\vartheta_j^n, \quad j=0,\dots,N.
$$

Consquently, we may write (4.12) as

$$
\begin{aligned}
\vartheta_{-1}^{n+1} &= \big(\kappa_{-1}^n - \lambda(\eta_{-1/2}^n + \nu_{-3/2}^n)\big)\vartheta_{-1}^n + \lambda\nu_{-1/2}^n\vartheta_0^n, \\
\big(1+\theta_0^{n+1}\big)\vartheta_0^{n+1} &= \big(\kappa_0^n - \lambda(\eta_{1/2}^n + \nu_{-1/2}^n) + \tau\gamma_0\rho_0^n\big)\vartheta_0^n + \lambda\nu_{1/2}^n\vartheta_1^n + \lambda\eta_{-1/2}^n\vartheta_{-1}^n + \theta_1^{n+1}\vartheta_1^{n+1}, \\
\big(1+2\theta_j^{n+1}\big)\vartheta_j^{n+1} &= \big(\kappa_j^n - \lambda(\eta_{j+1/2}^n + \nu_{j-1/2}^n) + \tau\gamma_j\rho_j^n\big)\vartheta_j^n + \lambda\nu_{j+1/2}^n\vartheta_{j+1}^n \\
&\quad + \lambda\eta_{j-1/2}^n\vartheta_{j-1}^n + \theta_{j+1}^{n+1}\vartheta_{j+1}^{n+1} + \theta_{j-1}^{n+1}\vartheta_{j-1}^{n+1}, \quad j=1,\dots,N-1, \\
\big(1+\theta_N^{n+1}\big)\vartheta_N^{n+1} &= \big(\kappa_N^n - \lambda(\eta_{N+1/2}^n + \nu_{N-1/2}^n) + \tau\gamma_N\rho_N^N\big)\vartheta_N^n + \lambda\eta_{N-1/2}^n\vartheta_{N-1}^n \\
&\quad + \lambda\nu_{N+1/2}^n\vartheta_{N+1}^n + \theta_{N-1}^{n+1}\vartheta_{N-1}^{n+1}, \\
\vartheta_{N+1}^{n+1} &= \big(\kappa_{N+1}^n - \lambda(\eta_{N+3/2}^n + \nu_{N+1/2}^n)\big)\vartheta_{N+1}^n + \lambda\eta_{N+1/2}^n\vartheta_N^n.
\end{aligned}
$$

Since by (CFL-SI), all coefficients in these equations are nonnegative, we get

$$
\begin{aligned}
\big|\vartheta_{-1}^{n+1}\big| &\le \big(\kappa_{-1}^n - \lambda(\eta_{-1/2}^n + \nu_{-3/2}^n)\big)|\vartheta_{-1}^n| + \lambda\nu_{-1/2}^n|\vartheta_0^n|, \\
\big(1+\theta_0^{n+1}\big)\big|\vartheta_0^{n+1}\big| &\le \big(\kappa_0^n - \lambda(\eta_{1/2}^n + \nu_{-1/2}^n) + \tau\gamma_0\rho_0^n\big)|\vartheta_0^n| + \lambda\nu_{1/2}^n|\vartheta_1^n| + \lambda\eta_{-1/2}^n|\vartheta_{-1}^n| + \theta_1^{n+1}\big|\vartheta_1^{n+1}\big|, \\
\big(1+2\theta_j^{n+1}\big)\big|\vartheta_j^{n+1}\big| &\le \big(\kappa_j^n - \lambda(\eta_{j+1/2}^n + \nu_{j-1/2}^n) + \tau\gamma_j\rho_j^n\big)\big|\vartheta_j^n\big| + \lambda\nu_{j+1/2}^n\big|\vartheta_{j+1}^n\big| \\
&\quad + \lambda\eta_{j-1/2}^n\big|\vartheta_{j-1}^n\big| + \theta_{j+1}^{n+1}\big|\vartheta_{j+1}^{n+1}\big| + \theta_{j-1}^{n+1}\big|\vartheta_{j-1}^{n+1}\big|, \quad j=1,\dots,N-1, \\
\big(1+\theta_N^{n+1}\big)\big|\vartheta_N^{n+1}\big| &\le \big(\kappa_N^n - \lambda(\eta_{N+1/2}^n + \nu_{N-1/2}^n) + \tau\gamma_N\rho_N^N\big)|\vartheta_N^n| + \lambda\eta_{N-1/2}^n\big|\vartheta_{N-1}^n\big| \\
&\quad + \lambda\nu_{N+1/2}^n\big|\vartheta_{N+1}^n\big| + \theta_{N-1}^{n+1}\big|\vartheta_{N-1}^{n+1}\big|, \\
\big|\vartheta_{N+1}^{n+1}\big| &\le \big(\kappa_{N+1}^n - \lambda(\eta_{N+3/2}^n + \nu_{N+1/2}^n)\big)\big|\vartheta_{N+1}^n\big| + \lambda\eta_{N+1/2}^n|\vartheta_N^n|.
\end{aligned}
$$

Summing over these inqualities, cancelling terms, rewriting the result again in terms of $\{Y_j^n\}$ and $\{X_j^n\}$, and taking into account that $\nu_{-3/2}^n \geq 0$ and $\eta_{N+3/2}^n$ we obtain the inequality

$$
\begin{aligned}
\sum_{j=-1}^{N+1} |Y_j^{n+1} - X_j^{n+1}| &\leq \sum_{j=-1}^{N+1} (\kappa_j^n + \tau\gamma_j\rho_j^n)|Y_j^n - X_j^n| \\
&\quad - \lambda\nu_{-3/2}^n|Y_{-1}^n - X_{-1}^n| - \lambda\eta_{N+3/2}^n|Y_{N+1}^n - X_{N+1}^n| \\
&\leq \sum_{j=-1}^{N+1} (1 + C\tau)|Y_j^n - X_j^n|,
\end{aligned}
$$

that is, (4.11), where we take into account that $\kappa_j^n \leq 1 + C\tau$ with a suitable constant $C > 0$ (see (3.2)) and that the quantities ρ_j^n are uniformly bounded. $\qquad\square$

4.2. Numerical solution of the nonlinear system

System (4.4) can be written as $\boldsymbol{\varphi}(\boldsymbol{X}^{n+1}) = \boldsymbol{0}$, where $\boldsymbol{\varphi}$ is a nonlinear vector function and $\boldsymbol{\mathcal{J}_\varphi} \in \mathbb{R}^{(N+1)\times(N+1)}$ is its associated Jacobian matrix. The Newton-Raphson method applied to (4.4) reads

$$
\boldsymbol{\mathcal{J}_\varphi}(\boldsymbol{u}^{[k]})(\boldsymbol{u}^{[k+1]} - \boldsymbol{u}^{[k]}) = -\boldsymbol{\varphi}(\boldsymbol{u}^{[k]}), \quad k = 0, 1, \ldots, \tag{4.13}
$$

with the Jacobian matrix $\boldsymbol{\mathcal{J}_\varphi}(\boldsymbol{u})$ given by (4.5) (recall that $a = \mathcal{D}'$). The iteration starts with the initial vector $\boldsymbol{u}^{[0]} := \boldsymbol{X}^n$ and evolves formally according to (4.13) until the termination criterion

$$
\frac{\|\boldsymbol{u}^{[k+1]} - \boldsymbol{u}^{[k]}\|_1}{\|\boldsymbol{u}^{[k]}\|_1} < \varepsilon_{\text{tol}}
$$

is reached, where $\varepsilon_{\text{tol}} > 0$ is the tolerance and $\|\cdot\|_1$ is the ℓ_1-norm. After convergence, we set $\boldsymbol{X}^{n+1} := \boldsymbol{u}^{[k+1]}$. Since the matrix $\boldsymbol{\mathcal{J}_\varphi}(\boldsymbol{u})$ is strictly diagonally dominant by columns for all $\boldsymbol{u}$, it is invertible and the iteration (4.13) is well defined.

4.3. Update of the percentage vector $\boldsymbol{p}_j^n$

An inspection of the proof of Lemma 3.4 reveals that although the update formula for the percentages (3.12) is an explicit upwind scheme, it is still associated with a CFL condition that imposes a bound on $\tau/\Delta\xi^2$ due to the presence of differences of $\mathcal{D}$-values divided by $\Delta\xi$ in the convective flux, *cf.* (3.4), (3.7) and (3.8). Consequently, to remove this shortcoming so that the whole semi-implicit scheme (and not just the update formula for X) is associated with a CFL bound on $\tau/\Delta\xi$ only, we need to resort to an implicit difference scheme.

We write out all terms in (3.12) and evaluate those containing μ at time t^{n+1}:

$$
\begin{aligned}
\boldsymbol{p}_j^{n+1}X_j^{n+1} &= \kappa_j^n\boldsymbol{p}_j^nX_j^n - \lambda\big(\Phi_{j+1/2}^{n,n+1,+}\boldsymbol{p}_j^{n+1} + \Phi_{j+1/2}^{n,n+1,-}\boldsymbol{p}_{j+1}^{n+1} - \Phi_{j-1/2}^{n,n+1,+}\boldsymbol{p}_{j-1}^{n+1} - \Phi_{j-1/2}^{n,n+1,-}\boldsymbol{p}_j^{n+1}\big) \\
&\quad + \lambda\delta_{j,0}\beta^nq_{\text{f}}^n\boldsymbol{p}_{\text{f}}^nX_{\text{f}}^n + c\tau\gamma_j\boldsymbol{R_C}(\boldsymbol{p}_j^nX_j^n/c, \boldsymbol{S}_j^n),
\end{aligned} \tag{4.14}
$$

where

$$
\Phi_{j+1/2}^{n,n+1} := \mathcal{F}_{j+1/2}^n - \mathcal{J}_{j+1/2}^{n+1}.
$$

For the cells outside the mixture, this reduces to (3.9) if $q_{\text{e}} = 0$ and $j = -1$; otherwise,

$$
\begin{aligned}
\boldsymbol{p}_{-1}^{n+1}X_{-1}^{n+1} &= (1 - \tau\beta^nq_{\text{out}}^n)\boldsymbol{p}_{-1}^nX_{-1}^n + \lambda\beta^n\big((\xi_{-1/2}q_{\text{out}}^n + q_{\text{e}}^n)\boldsymbol{p}_0^nX_0^n - (\xi_{-3/2}q_{\text{out}}^n + q_{\text{e}}^n)\boldsymbol{p}_{-1}^nX_{-1}^n\big), \\
\boldsymbol{p}_{N+1}^{n+1}X_{N+1}^{n+1} &= (1 + \tau\beta^nq_{\text{out}}^n)\boldsymbol{p}_{N+1}^nX_{N+1}^n - \lambda\big((\alpha_{N+3/2}^n + \beta^nq_{\text{u}}^n)\boldsymbol{p}_{N+1}^nX_{N+1}^n - \beta^nq_{\text{u}}^n\boldsymbol{p}_N^nX_N^n\big),
\end{aligned}
$$

where we recall that $q_{\text{out}}^n := q_{\text{u}}^n + q_{\text{e}}^n$. Let us focus the cells $j = 0, \ldots, N$ and put all the unknowns in a matrix (cells in rows and solid components in columns):

$$
\boldsymbol{P}^n := \begin{bmatrix} (\boldsymbol{p}_0^n)^{\text{T}} \\ (\boldsymbol{p}_1^n)^{\text{T}} \\ \vdots \\ (\boldsymbol{p}_N^n)^{\text{T}} \end{bmatrix}, \qquad \boldsymbol{W}^n := \begin{bmatrix} (\lambda\beta^n q_{\text{f}}^n \boldsymbol{p}_{\text{f}}^n X_{\text{f}}^n + c\frac{\tau}{2}\boldsymbol{R_C}(\boldsymbol{p}_0^n X_0^n/c, \boldsymbol{S}_0^n))^{\text{T}} \\ c\tau\boldsymbol{R_C}(\boldsymbol{p}_1^n X_1^n/c, \boldsymbol{S}_1^n)^{\text{T}} \\ \vdots \\ c\tau\boldsymbol{R_C}(\boldsymbol{p}_N^n X_N^n/c, \boldsymbol{S}_N^n)^{\text{T}} \end{bmatrix},
$$

$$
\boldsymbol{M}(\boldsymbol{\Phi}, \boldsymbol{X}) := \operatorname{diag}(\boldsymbol{X})
$$

$$
+ \lambda \begin{bmatrix} \Phi_{1/2}^+ - \Phi_{-1/2}^- & \Phi_{1/2}^- & 0 & \cdots & & 0 \\ -\Phi_{1/2}^+ & \Phi_{3/2}^+ - \Phi_{1/2}^- & \ddots & & \ddots & \vdots \\ 0 & -\Phi_{3/2}^+ & \ddots & & \ddots & 0 \\ \vdots & & \ddots & \ddots & \Phi_{N-1/2}^+ - \Phi_{N-3/2}^- & \Phi_{N-1/2}^- \\ 0 & \cdots & & 0 & -\Phi_{N-1/2}^+ & \Phi_{N+1/2}^+ - \Phi_{N-1/2}^- \end{bmatrix},
$$

where for a vector $\boldsymbol{X}$ we define $\operatorname{diag}(\boldsymbol{X}) := \operatorname{diag}(X_0, \ldots, X_N)$. With $\boldsymbol{\kappa} := (\kappa_0, \ldots, \kappa_N)^{\text{T}}$, we get the linear system

$$
\boldsymbol{M}(\boldsymbol{\Phi}^{n,n+1}, \boldsymbol{X}^{n+1})\boldsymbol{P}^{n+1} = \operatorname{diag}(\boldsymbol{\kappa}^n)\operatorname{diag}(\boldsymbol{X}^n)\boldsymbol{P}^n + \boldsymbol{W}^n =: \boldsymbol{\Theta}^n. \tag{4.15}
$$

In the case $X_j^{n+1} = 0$ the percentage vector $\boldsymbol{p}_j^{n+1}$ is irrelevant and one can define $\boldsymbol{p}_j^{n+1} := \boldsymbol{p}_j^n$. Then Equation (4.15) should be modified in the following way. Row j is removed in $\boldsymbol{p}^{n+1}$, $\boldsymbol{p}^n$ and $\boldsymbol{W}^n$, and both row j and column j should be removed in the matrices $\boldsymbol{M}(\boldsymbol{\Phi}^{n,n+1}, \boldsymbol{X}^{n+1})$, $\operatorname{diag}(\boldsymbol{\kappa}^n)$ and $\operatorname{diag}(\boldsymbol{X}^n)$. Then, one verifies that $\boldsymbol{M}(\boldsymbol{\Phi}^{n,n+1}, \boldsymbol{X}^{n+1})$ is strictly diagonally dominant by columns, and therefore invertible without any restrictions. Thus, the implicit scheme (4.15) is well defined. Furthermore, $\boldsymbol{M}(\boldsymbol{\Phi}^{n,n+1}, \boldsymbol{X}^{n+1})^{\text{T}}$ is an M-matrix and hence has a nonnegative inverse.

Lemma 4.5. *If $\boldsymbol{\mathcal{U}}_j^n \in \Omega$ for all j and* (CFL-SI) *holds, then*

$$
p_j^{(k),n+1} \geq 0 \quad \text{for all } k = 1, \ldots, k_{\boldsymbol{C}} \text{ and all } j.
$$

Proof. Since $(\boldsymbol{M}(\boldsymbol{\Phi}^{n,n+1}, \boldsymbol{X}^{n+1}))^{-1} \geq \boldsymbol{0}$, we estimate each entry of $\boldsymbol{\Theta}^n =: (\Theta_{jk}^n)$. By Lemma 3.2,

$$
\begin{aligned}
\Theta_{0,k}^n &= \kappa^n X_0^n p_0^{(k),n} + \lambda\beta^n q_{\text{f}}^n \boldsymbol{p}_{\text{f}}^n X_{\text{f}}^n + c\frac{\tau}{2}R_{\boldsymbol{C}}^{(k)}(\boldsymbol{p}_0^n X_0^n/c, \boldsymbol{S}_0^n) \\
&\geq (1 - \tau\zeta M_{q1})\, X_0^n p_0^{(k),n} + 0 + c\frac{\tau}{2}\sum_{l \in I_{\boldsymbol{C},k}^-} \sigma_{\boldsymbol{C}}^{(k,l)}\bar{r}^{(l)}\big(\boldsymbol{p}_0^n X_0^n/c, \boldsymbol{S}_0^n\big)p_0^{(k),n} X_0^n/c \\
&\geq (1 - \tau\zeta M_{q1} - \tau M_{\boldsymbol{C}}/2)\, p_0^{(k),n} X_0^n \geq 0,
\end{aligned}
$$

whereas for all the other $j \neq 0$,

$$
\Theta_{j,k}^n = \kappa^n X_j^n p_j^{(k),n} + c\tau R_{\boldsymbol{C}}^{(k)}(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n) \geq (1 - \tau\zeta M_{q1} - \tau M_{\boldsymbol{C}})\, p_j^{(k),n} X_j^n \geq 0.
$$

$\square$

Lemma 4.6. *If $\boldsymbol{\mathcal{U}}_j^n \in \Omega$ for all j and* (CFL-SI) *holds, then*

$$
p_j^{(1),n+1} + \cdots + p_j^{(k_{\boldsymbol{C}}),n+1} = 1 \quad \text{for all } j. \tag{4.16}
$$

Proof. If $X_j^{n+1} = 0$, then by definition,

$$p_j^{(1),n+1} + \cdots + p_j^{(k_C),n+1} = p_j^{(1),n} + \cdots + p_j^{(k_C),n} = 1,$$

so let us assume that $X_j^{n+1} > 0$. We sum up all equations in (4.14) and use the notation

$$\mathcal{P}_j^{n+1} := p_j^{(1),n+1} + \cdots + p_j^{(k_C),n+1}$$

to obtain

$$\begin{aligned}
X_j^{n+1}\mathcal{P}_j^{n+1} = \kappa_j^n X_j^n \\
- \lambda\big(\Phi_{j+1/2}^{n,n+1,+}\mathcal{P}_j^{(k),n+1} + \Phi_{j+1/2}^{n,n+1,-}\mathcal{P}_{j+1}^{(k),n+1} - \Phi_{j-1/2}^{n,n+1,+}\mathcal{P}_{j-1}^{(k),n+1} - \Phi_{j-1/2}^{n,n+1,-}\mathcal{P}_j^{(k),n+1}\big) \\
+ \lambda\delta_{j,0}\beta^n q_{\mathrm{f}}^n X_{\mathrm{f}}^n + \tau\gamma_j R(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n).
\end{aligned}$$

We subtract component j of equation (4.2), let $y_j^{n+1} := \mathcal{P}_j^{n+1} - 1$ and obtain

$$X_j^{n+1}y_{j+1}^{n+1} = -\lambda\big(\Phi_{j+1/2}^{n,n+1,+}y_j^{n+1} + \Phi_{j+1/2}^{n,n+1,-}y_{j+1}^{n+1} - \Phi_{j-1/2}^{n,n+1,+}y_{j-1}^{n+1} - \Phi_{j-1/2}^{n,n+1,-}y_j^{n+1}\big).$$

Thus, with $\boldsymbol{y}_{j+1}^{n+1} := (y_0^{n+1}, \ldots, y_N^{n+1})^{\mathrm{T}}$, we get $\boldsymbol{M}^{n,n+1}\boldsymbol{y}_{j+1}^{n+1} = \boldsymbol{0}$, which implies $\boldsymbol{y}_{j+1}^{n+1} = \boldsymbol{0}$, *i.e.* (4.16). □

4.4. Update of the soluble concentrations $\boldsymbol{S}_j^n$

For equation (2.26) we have

$$\begin{aligned}
\boldsymbol{S}_j^{n+1} = \kappa_j^n \boldsymbol{S}_j^n - \lambda\Bigg(&\frac{(\rho_X\tilde{q}^n - \mathcal{F}^n + \mathcal{J}^{n+1})_{j+1/2}^+}{\rho_X - X_j^{n+1}}\boldsymbol{S}_j^{n+1} + \frac{(\rho_X\tilde{q}^n - \mathcal{F}^n + \mathcal{J}^{n+1})_{j+1/2}^-}{\rho_X - X_{j+1}^{n+1}}\boldsymbol{S}_{j+1}^{n+1} \\
&- \frac{(\rho_X\tilde{q}^n - \mathcal{F}^n + \mathcal{J}^{n+1})_{j-1/2}^+}{\rho_X - X_{j-1}^{n+1}}\boldsymbol{S}_{j-1}^{n+1} - \frac{(\rho_X\tilde{q}^n - \mathcal{F}^n + \mathcal{J}^{n+1})_{j-1/2}^-}{\rho_X - X_j^{n+1}}\boldsymbol{S}_j^{n+1}\Bigg) \\
&+ \lambda\delta_{j,0}\beta^n q_{\mathrm{f}}^n \boldsymbol{S}_{\mathrm{f}}^n + \tau\gamma_j \boldsymbol{R_S}(\boldsymbol{p}_j^n X_j^n/c, \boldsymbol{S}_j^n).
\end{aligned}$$

For $j = -1, N+1$, this scheme is explicit and analogous to those above. For $j = 0, \ldots, N$, we write the formula in matrix form as follows. Define

$$\theta_{j+1/2}^{n,n+1} := (\rho_X\tilde{q}^n - \mathcal{F}^n + \mathcal{J}^{n+1})_{j+1/2}, \qquad y_j^n := \frac{1}{\rho_X - X_j^n},$$

$$\boldsymbol{S}^n := \begin{bmatrix} (\boldsymbol{S}_0^n)^{\mathrm{T}} \\ (\boldsymbol{S}_1^n)^{\mathrm{T}} \\ \vdots \\ (\boldsymbol{S}_N^n)^{\mathrm{T}} \end{bmatrix}, \qquad \boldsymbol{\mathcal{W}}^n := \begin{bmatrix} (\lambda\beta^n q_{\mathrm{f}}^n \boldsymbol{S}_{\mathrm{f}}^n + \frac{\tau}{2}\boldsymbol{R_S}(\boldsymbol{p}_0^n X_0^n/c, \boldsymbol{S}_0^n))^{\mathrm{T}} \\ \tau\boldsymbol{R_S}(\boldsymbol{p}_1^n X_1^n/c, \boldsymbol{S}_1^n)^{\mathrm{T}} \\ \vdots \\ \tau\boldsymbol{R_S}(\boldsymbol{p}_N^n X_N^n/c, \boldsymbol{S}_N^n)^{\mathrm{T}} \end{bmatrix}.$$

Since $\tilde{q}_{-1/2}^n \leq 0$ and $\tilde{q}_{N+1/2}^n \geq 0$, we have

$$\theta_{-1/2}^{n,n+1,+} = \rho_X\tilde{q}_{-1/2}^{n+1,+} = 0, \quad \theta_{N+1/2}^{n,n+1,-} = \rho_X\tilde{q}_{N+1/2}^{n+1,-} = 0.$$

Then we form the tridiagonal matrix

$$
\boldsymbol{M_S}(\boldsymbol{\theta}, \boldsymbol{y}) := \boldsymbol{I}_{N+1}
$$

$$
+ \lambda
\begin{bmatrix}
(\theta_{1/2}^+ - \theta_{-1/2}^-)y_0 & \theta_{1/2}^- y_1 & 0 & \cdots & & 0 \\
-\theta_{1/2}^+ y_0 & (\theta_{3/2}^+ - \theta_{1/2}^-)y_1 & \ddots & & \ddots & \vdots \\
0 & -\theta_{3/2}^+ y_1 & \ddots & & \ddots & 0 \\
\vdots & & \ddots & \ddots & (\theta_{N-1/2}^+ - \theta_{N-3/2}^-)y_{N-1} & \theta_{N-1/2}^- y_N \\
0 & & \cdots & 0 & -\theta_{N-1/2}^+ y_{N-1} & (\theta_{N+1/2}^+ - \theta_{N-1/2}^-)y_N
\end{bmatrix}.
$$

Then we get the linear system

$$
\boldsymbol{M_S}(\boldsymbol{\theta}^{n,n+1}, \boldsymbol{y}^{n+1})\boldsymbol{S}^{n+1} = \mathrm{diag}(\boldsymbol{\kappa}^n)\boldsymbol{S}^n + \boldsymbol{W}^n. \tag{4.17}
$$

The matrix $\boldsymbol{M_S}(\boldsymbol{\theta}^{n,n+1}, \boldsymbol{y}^{n+1})$ is diagonally dominant by columns; hence its transpose is an M-matrix and invertible with a nonnegative inverse, so that (4.17) defines $\boldsymbol{S}^{n+1}$.

Lemma 4.7. *If* $\boldsymbol{\mathcal{U}}_j^n \in \Omega$ *for all* j *and* (CFL-SI) *holds, then*

$$
S_j^{(k),n+1} \geq 0 \quad \text{for all } k = 1, \ldots, k_{\boldsymbol{S}} \text{ and all } j.
$$

Proof. This is similar to the proof of Lemma 4.5. The element (j, k), $j \neq 0$, of the matrix on the right-hand side of (4.17) is estimated by

$$
\kappa^n S_j^{(k),n} + \tau R_{\boldsymbol{S}}^{(k)}(\boldsymbol{p}_j^n X_j^n / c, \boldsymbol{S}_j^n) \geq (1 - \tau \zeta M_{q1} - \tau M_{\boldsymbol{S}}) S_j^{(k),n} \geq 0.
$$

$\square$

5. Numerical simulations

For all examples we consider a cylindrical SBR of depth $B = 3\,\mathrm{m}$ and cross-sectional area $A = 400\,\mathrm{m}^2$, and the reactive settling process of an activated sludge described by the ASM1 model [30]; see Table A.1. The constitutive functions used in the examples are

$$
v_{\mathrm{hs}}(X) := \frac{v_0}{1 + (X/\check{X})^\eta}, \quad \sigma_{\mathrm{e}}(X) := \begin{cases} 0 & \text{if } X < X_{\mathrm{c}}, \\ \sigma_0(X - X_{\mathrm{c}}) & \text{if } X \geq X_{\mathrm{c}} \end{cases} \tag{5.1}
$$

with $v_0 = 1.76 \times 10^{-3}\,\mathrm{m\,s^{-1}}$, $\check{X} = 3.87\,\mathrm{kg\,m^{-3}}$, $\eta = 3.58$, $X_{\mathrm{c}} = 5\,\mathrm{kg\,m^{-3}}$, and $\sigma_0 = 0.2\,\mathrm{m^2\,s^{-2}}$. Other parameters are $\rho_X = 1050\,\mathrm{kg\,m^{-3}}$, $\rho_L = 998\,\mathrm{kg\,m^{-3}}$, $g = 9.81\,\mathrm{m\,s^{-2}}$, and $B_{\mathrm{c}} = 2\,\mathrm{m}$. To satisfy (2.10), one could multiply $v_{\mathrm{hs}}(X)$ by a function that is one for most concentrations but tends to zero smoothly as $X \to \hat{X}^-$, where the maximum concentration $\hat{X}$ is set to a large number; *e.g.* $\hat{X} = 30\,\mathrm{kg\,m^{-3}}$ for activated sludge that we simulate here. When $\sigma_0 > 0$, the second-order derivative compression term will balance the convective flux and the particulate concentration X never reaches $\hat{X}$ in simulations. To be able to run the scheme with $\sigma_0 = 0$, an alternative is to set a concentration X^{t} from which we redefine and extend the settling velocity function by its tangent as

$$
v_{\mathrm{hs}}(X) := \begin{cases} v_0/(1 + (X/\check{X})^\eta) & \text{if } 0 \leq X \leq X^{\mathrm{t}}, \\ v_{\mathrm{hs}}(X^{\mathrm{t}}) + v_{\mathrm{hs}}'(X^{\mathrm{t}})(X - X^{\mathrm{t}}) & \text{if } X^{\mathrm{t}} < X \leq \hat{X}, \\ 0 & \text{if } X \geq \hat{X}, \end{cases}
$$

TABLE 1. Example 1: time functions for the simulation.

Stage	Time period [h]	$X_f(t)[\mathrm{kg\,m^{-3}}]$	$Q_f(t)[\mathrm{m^3\,h^{-1}}]$	$Q_u(t)[\mathrm{m^3\,h^{-1}}]$	$Q_e(t)[\mathrm{m^3\,h^{-1}}]$	Model
Fill	$0 \le t < 1$	5	790	0	0	PDE
React	$1 \le t < 3$	0	0	0	0	ODE
Settle	$3 \le t < 5$	0	0	0	0	PDE
Draw	$5 \le t < 5.5$	0	0	0	1570	PDE
Idle	$5.5 \le t < 6$	0	0	10	0	PDE

where $\hat{X}$ is given by the intersection of the tangent with the X-axis (zero velocity), *i.e.*,

$$\hat{X} := X^{\mathrm{t}} - v_{\mathrm{hs}}(X^{\mathrm{t}})/v'_{\mathrm{hs}}(X^{\mathrm{t}}).$$

We here utilize $X^{\mathrm{t}} = 25\,\mathrm{kg/m^3}$, such that $\hat{X} = 31.992\,\mathrm{kg/m^3}$.

The initial concentrations have been chosen as

$$\boldsymbol{C}^0(z) = \begin{cases} \boldsymbol{0} & \text{if } z < 2.0\,\mathrm{m}, \\ (0.8889, 0.0295, 1.4503, 0.0904, 0.7371, 0.0025)^{\mathrm{T}} & \text{if } z \ge 2.0\,\mathrm{m}, \end{cases}$$

$$\boldsymbol{S}^0(z) = \begin{cases} \boldsymbol{0} & \text{if } z < 2.0\,\mathrm{m}, \\ (0.04, 0.0026, 0.0, 0.0333, 0.0004, 0.0009)^{\mathrm{T}} & \text{if } z \ge 2.0\,\mathrm{m} \end{cases}$$

(with units as in Tab. A.1) while the feed concentrations are [30]

$$\boldsymbol{C}_{\mathrm{f}}(t) = \frac{X_{\mathrm{f}}(t)}{(0.04 + 0.16 + 0.096 + 1 \cdot 10^{-6})c}(0.04,\ 0.16 - 0.01828,\ 0.096,\ 1 \cdot 10^{-6},\ 0,\ 0.01828)^{\mathrm{T}},$$

$$\boldsymbol{S}_{\mathrm{f}}(t) = (0.04,\ 0.064,\ 0.0,\ 0.001,\ 0.0125,\ 0.0101)^{\mathrm{T}},$$

where the total solids feed concentration $X_{\mathrm{f}}(t)$ is given by Table 1. When plotting the particulate concentrations, we prefer to not plot $C^{(2)} = X_{\mathrm{S-ND}}$, but rather $X_{\mathrm{S}} = C^{(2)} + X_{\mathrm{ND}} = C^{(2)} + C^{(6)}$ and $X_{\mathrm{ND}} = C^{(6)}$. All results are shown after transforming back to the original coordinates.

5.1. Example 1

It is the purpose of this example to illustrate the SBR model as a whole. (The performance of the three numerical schemes SBR2 in [16], and the explicit and semi-implicit ones here in terms of errors and efficiency is studied in Examples 2 and 3.) We simulated the five stages of an SBR as outlined in Table 1. The duration of the whole cycle of stages during a couple of hours is realistic. The results are illustrated in Figures 5 to 8, which depict the concentration profiles of total suspended solids, particulate, and soluble components within the reactor vessel, respectively.

With respect to the numerical results for the react stage, we comment that we assume that the mixing during the react stage is accompanied by aeration so that a controller maintains the set-point value of dissolved oxygen $S_{\mathrm{O}}^{\mathrm{sp}} = 8\,\mathrm{g/m^3}$. This assumption is also considered in the numerical examples of [16], and it explains why the oxygen is not consumed during mixing (*i.e.*, react stage) in the present example. After this time interval oxygen is quickly consumed, as can be seen in Figure 7b.

5.2. Example 2

The prime motivation of this example is to study the numerical errors of the two new schemes defined for the PDE model and the scheme SBR2 of [16]. To this end we consider the scenario defined by Table 2, which
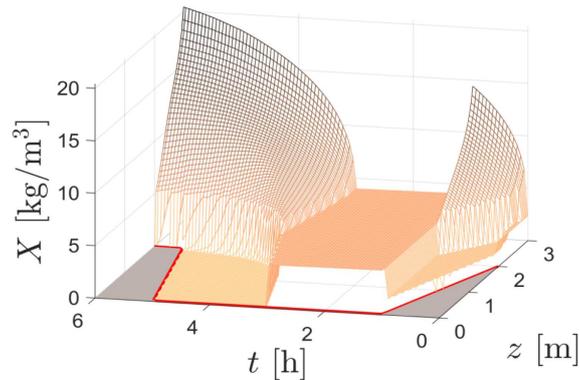
FIGURE 5. Example 1: simulated concentration (semi-implicit scheme, $N = 100$) of total suspended solids.

refers to a shorter period of total simulated time ($T = 1\,\mathrm{h}$) and does not include a react stage (for which the original SBR model, as studied in Example 1, stipulates a model given by a system of ODEs). We calculated a reference solution, using the same feed conditions as in Example 1 with $N = N_{\mathrm{ref}} := 4800$. This solution was obtained by the explicit scheme of Section 3.2. The relative approximate numerical error

$$e_N^{\mathrm{rel}}(t) := \sum_{k=1}^{k_C} \frac{\|C_N^{(k)} - C_{N_{\mathrm{ref}}}^{(k)}(\cdot, t)\|_{L^1(0,B)}}{\|C_{N_{\mathrm{ref}}}^{(k)}(\cdot, t)\|_{L^1(0,B)}} + \sum_{k=1}^{k_S} \frac{\|S_N^{(k)} - S_{N_{\mathrm{ref}}}^{(k)}(\cdot, t)\|_{L^1(0,B)}}{\|S_{N_{\mathrm{ref}}}^{(k)}(\cdot, t)\|_{L^1(0,B)}}. \tag{5.2}$$

compares the approximate solution to the reference solution at a given time point t for a specific number of cells N.

In Figures 9 to 11, we demonstrate the convergence of the numerical solutions, all produced with the semi-implicit scheme, to the reference solution. Before discussing the observed convergence properties, numerical errors, and efficiency, we mention that while Figure 9 illustrates that the total solids concentration X increases toward the bottom forming a moderate spatial gradient, the proportions of X_{I} and X_{S} increase rapidly closely to the bottom (see Figs. 9 (a) and (b)) while those of $X_{\mathrm{B,H}}$, $X_{\mathrm{B,A}}$, X_{P}, and X_{ND} decrease.

The results shown in Figures 9 to 11 are based on a tolerance $\varepsilon_{\mathrm{tol}} = 10^{-8}$. That value was chosen by previous experience. The effect of the tolerance parameter $\varepsilon_{\mathrm{tol}}$ itself is studied in Table 3 for simulations done with $N = 50$, 100, and 200, in terms of the average number of iterations during the simulation, the relative error, and CPU time. It turns out that the relative error depends only marginally on the choice of $\varepsilon_{\mathrm{tol}}$. As one should expect, the average number of iterations (of the Newton-Raphson scheme, per time step), as well as the CPU time for the whole simulation, consistently increase with decreasing $\varepsilon_{\mathrm{tol}}$. Overall, it appears that the relative error and CPU time, and therefore efficiency, do not depend critically on $\varepsilon_{\mathrm{tol}}$ (at least for the model functions and parameters chosen here).

For the same configuration, we compare in Table 4 and Figure 12 the relative errors and CPU times obtained by the three schemes SBR2 of [16], and the explicit and semi-implicit schemes of Sections 3 and 4, respectively. For both explicit and semi-implicit schemes, simulations were performed for $N = 25, \dots, 1600$ using the Engquist-Osher flux (3.3). The semi-implicit scheme was calculated with the tolerance $\varepsilon_{\mathrm{tol}} = 10^{-8}$. It turns out that with the exception of very coarse discretizations, the semi-implicit scheme is significantly more efficient in error reduction per CPU time than its explicit counterpart due to its more favorable CFL condition, and is more efficient also than SBR2 for sufficiently fine discretizations. It is noteworthy that all schemes converge to the same solution.
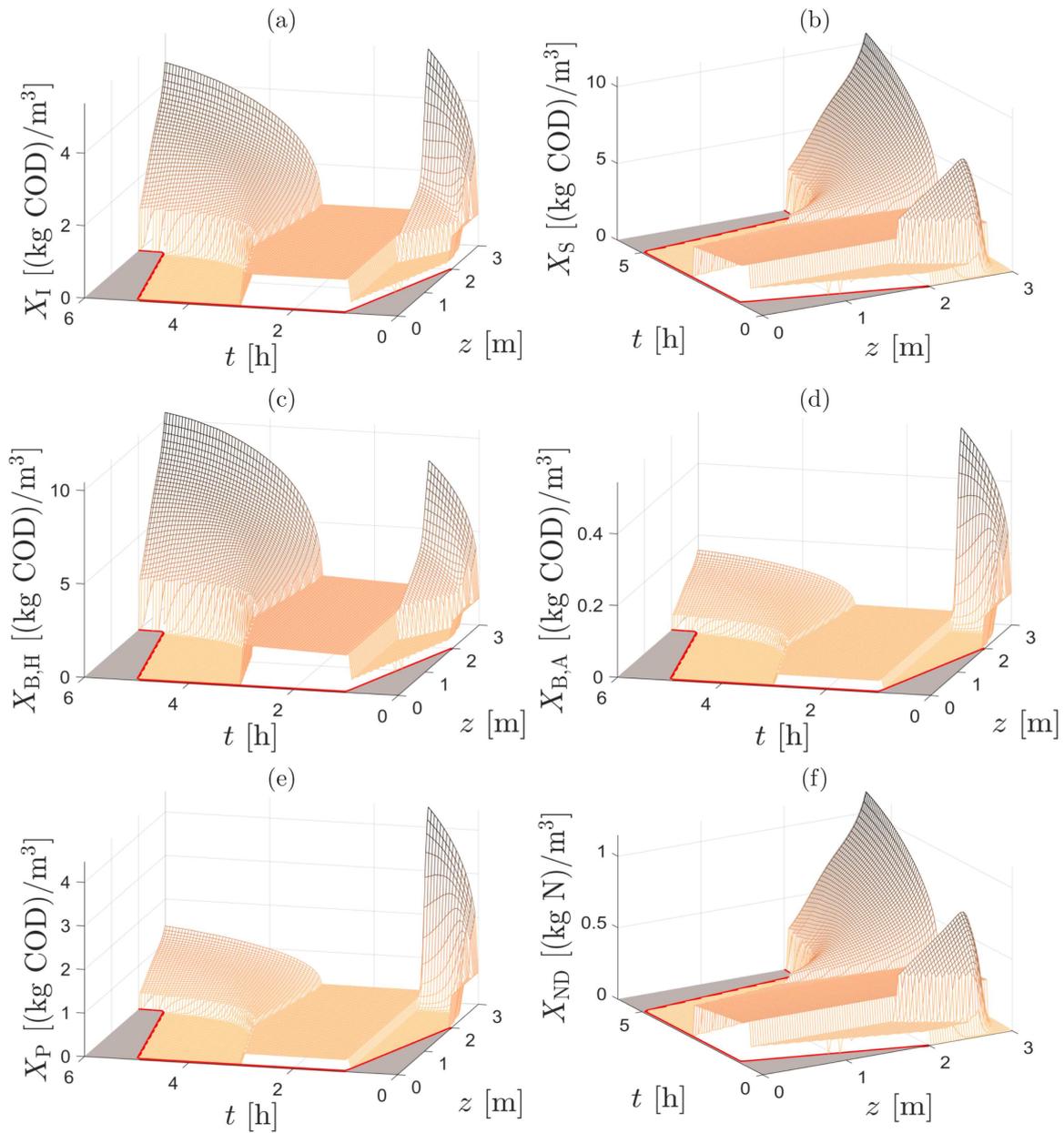
FIGURE 6. Example 1: simulated concentrations of solid components (semi-implicit scheme, $N = 100$): (a) particulate inert organic matter, (b) slowly biodegradable substrate, (c) active heterotrophic biomass, (d) active autotrophic biomass, (e) particle products from biomass decay, (f) particulate biodegradable organic nitrogen.
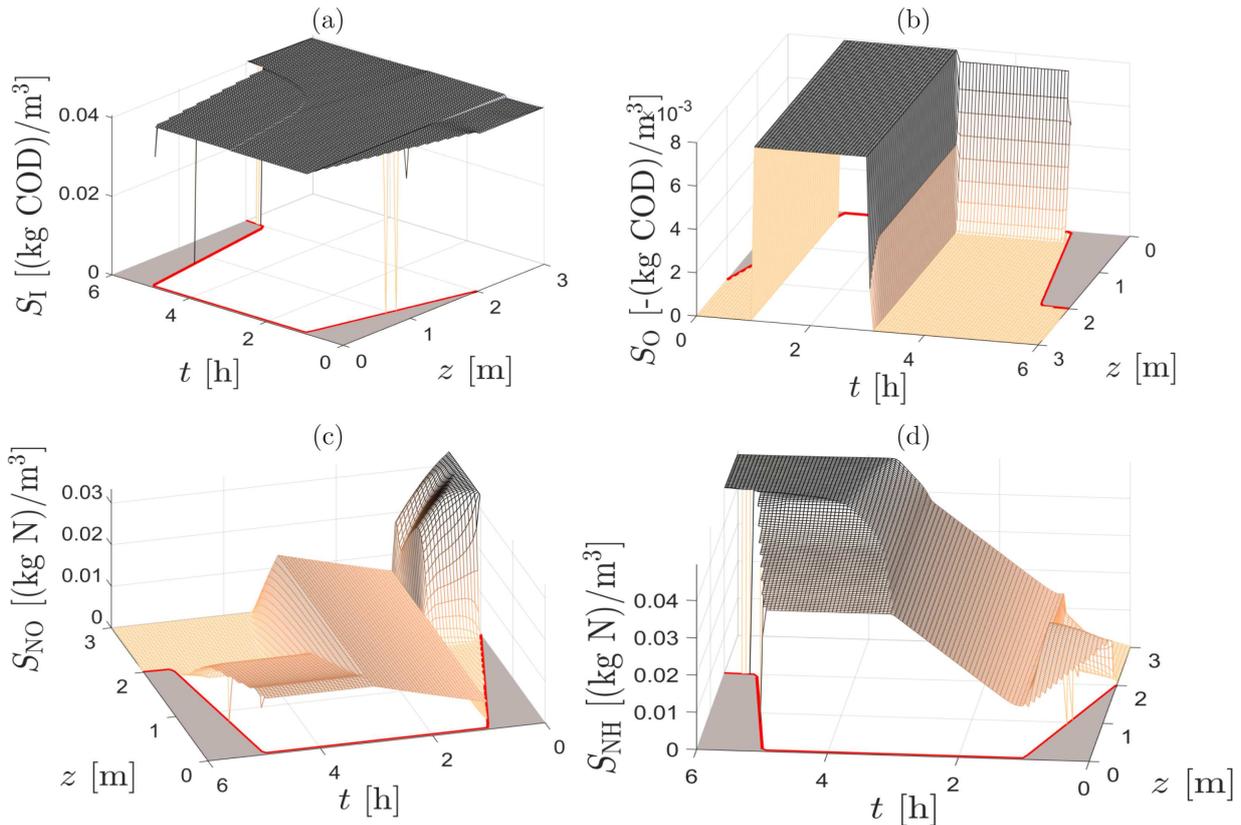
FIGURE 7. Example 1: simulated densities (semi-implicit scheme, $N = 100$) of soluble components (part 1): (a) soluble inert organic matter, (b) oxygen, (c) nitrate and nitrite nitrogen, (d) $NH_4^+ + NH_3$ nitrogen.

5.3. Example 3

We investigate the impact of a moving mesh with three different spatial grids, compared to a fixed grid; see Figure 13. The simulation conditions are given in Table 5. The semi-implicit scheme is used with the Engquist-Osher flux and $\varepsilon_{tol} = 10^{-8}$. The data in Table 5, which contemplate no solids feed or discharge but a moving boundary, have been chosen in such a way that the solution converges quickly in time to a stationary one with a layer of sediment at the bottom produced by the settling of the initially homogeneous suspension that is not affected by the moving boundary. Figures 13a, c, and e show that the numerical solution of the SBR2 scheme reproduces this property at any of the chosen discretizations $N = 100$, $N = 200$ or $N = 400$ while the solution produced by the semi-implicit scheme is not stationary due to the effect of the moving mesh. The deviation from a stationary solution seems, however, to decrease with increasing N.

6. CONCLUSIONS

The numerical results demonstrate that the two versions of the numerical scheme presented herein, the explicit one of Section 3.2 and the semi-implicit one of Section 4, generate physically relevant solutions (as is expressed by the invariant region principle) and are working alternatives to the scheme SBR2 in [16] of the PDE model advanced in [15]. Moreover, as Example 2 clearly illustrates, there is a substantial gain in CPU times
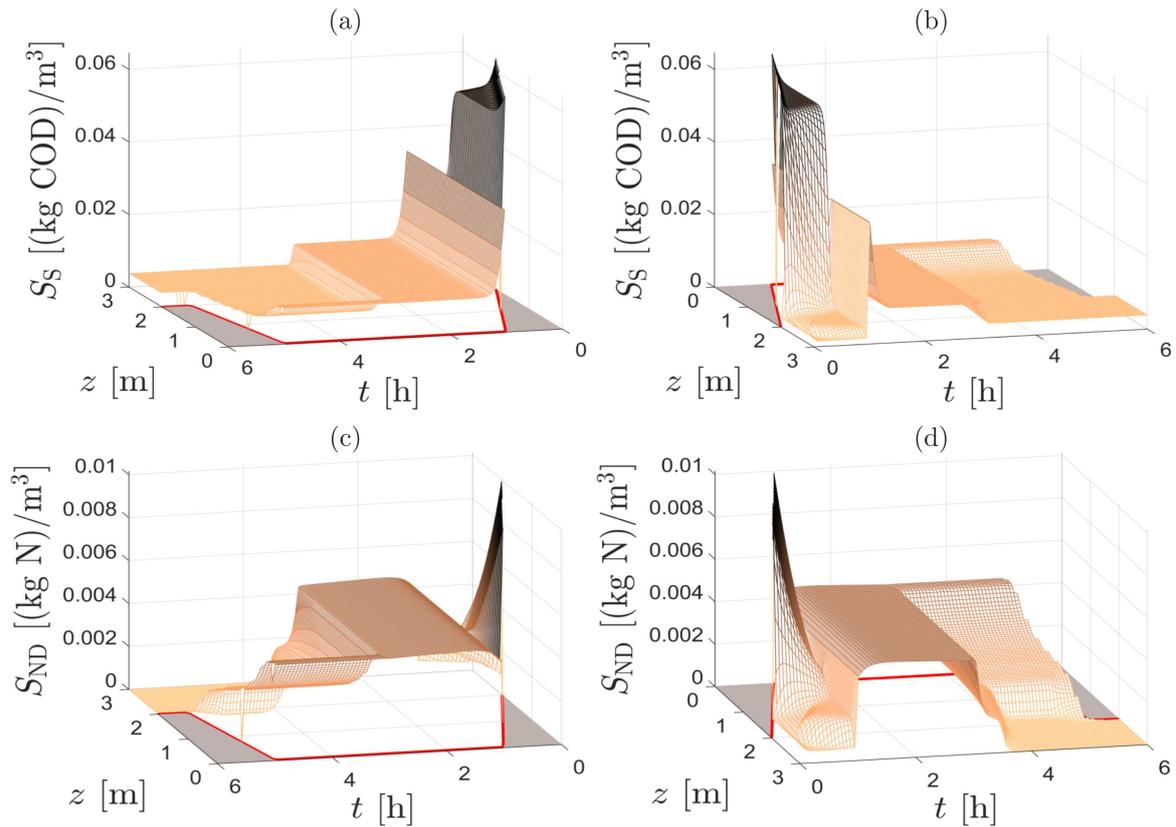
FIGURE 8. Example 1: simulated densities (semi-implicit scheme, $N = 100$) of soluble components (part 2): (a, b) readily biodegradable substrate (two different views), (c, d) soluble biodegradable organic nitrogen (two different views).
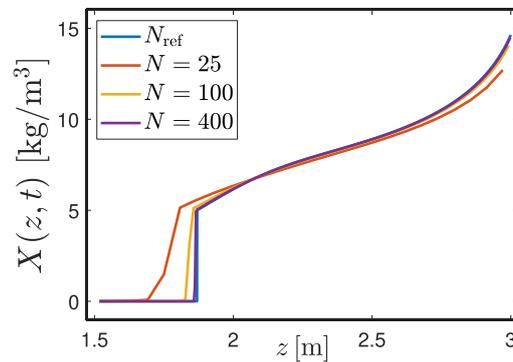
FIGURE 9. Example 2: simulated total solids concentration (semi-implicit scheme) at $T = 1\,\mathrm{h}$.

FIGURE 10. Example 2: simulated concentrations (semi-implicit scheme) at $T = 1\,\mathrm{h}$ of solid components:(a) particulate inert organic matter, (b) slowly biodegradable substrate, (c) active heterotrophic biomass, (d) active autotrophic biomass, (e) particle products from biomass decay, (f) particulate biodegradable organic nitrogen.
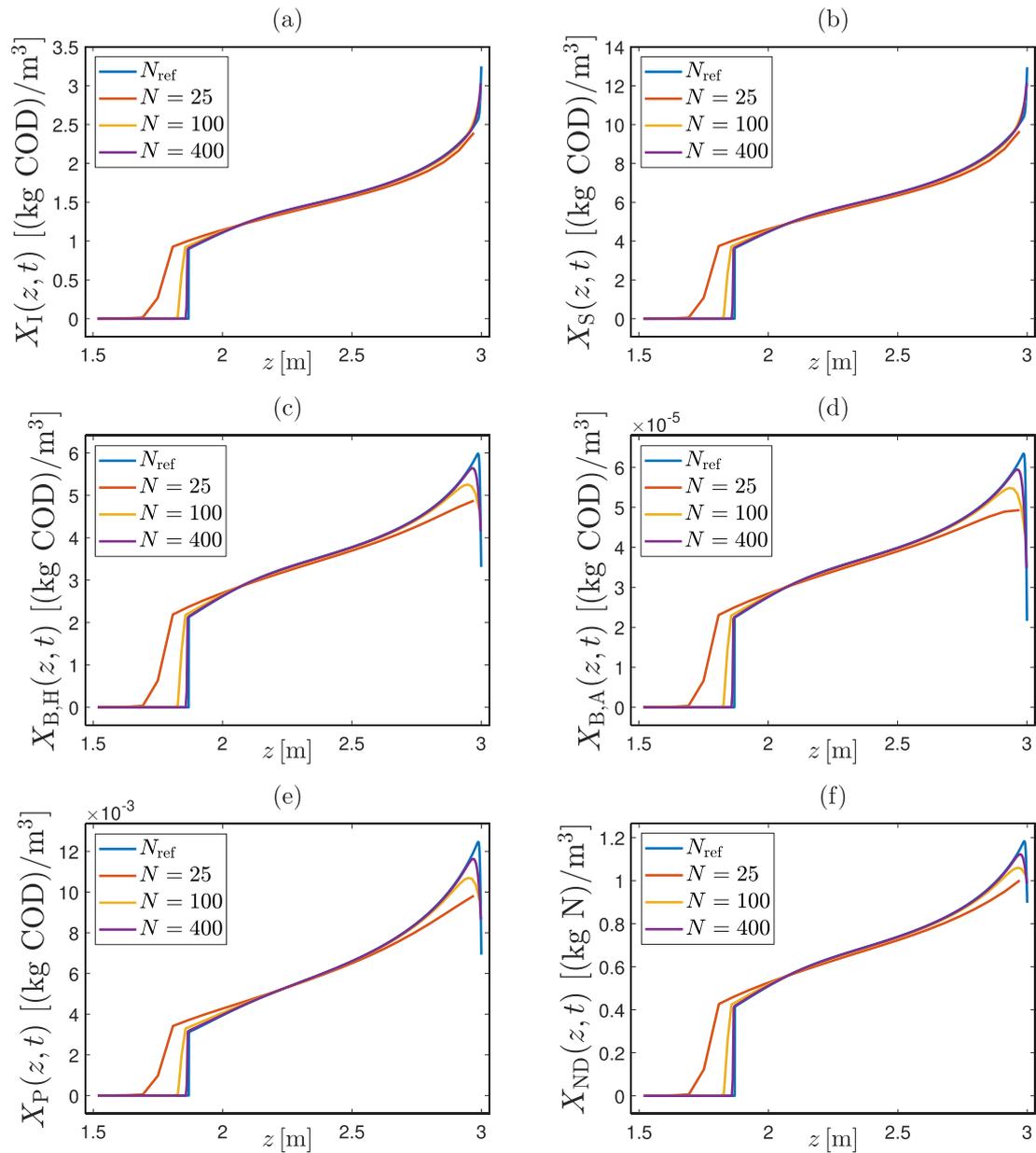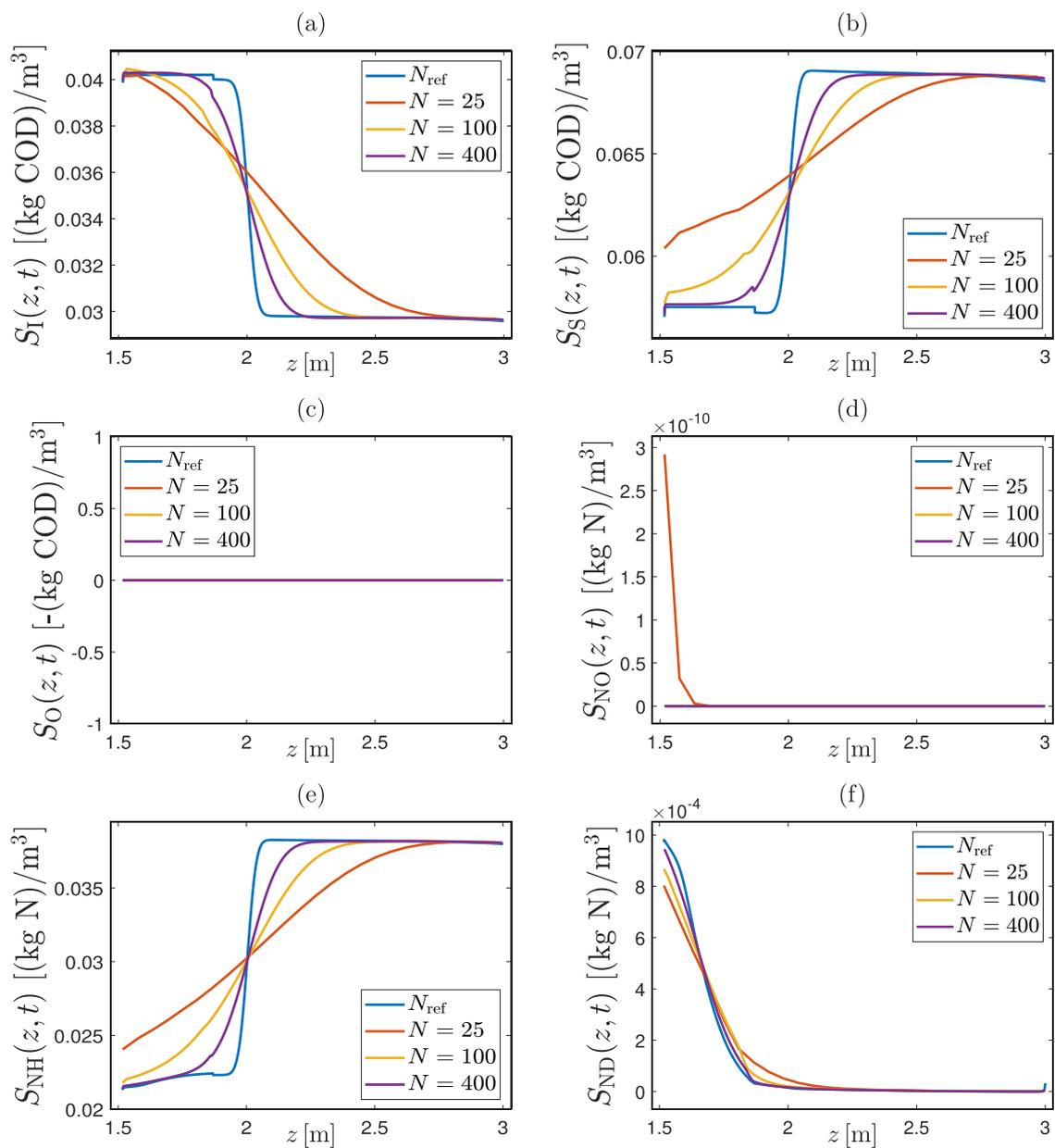
FIGURE 11. Example 2: simulated concentrations (semi-implicit scheme) at $T = 1\,\mathrm{h}$ of soluble components: (a) soluble inert organic matter, (b) readily biodegradable substrate, (c) oxygen, (d) nitrate and nitrite nitrogen, (e) $\mathrm{NH_4^+}$ + $\mathrm{NH_3}$ nitrogen, (f) soluble biodegradable organic nitrogen.

TABLE 2. Example 2: time functions for the simulation of the reference solution.

Time period [h]	$X_f(t)[\text{kg m}^{-3}]$	$Q_f(t)[\text{m}^3\,\text{h}^{-1}]$	$Q_u(t)[\text{m}^3\,\text{h}^{-1}]$	$Q_e(t)[\text{m}^3\,\text{h}^{-1}]$	Model
$0 \le t < 0.3$	5	2660	0	0	PDE
$0.3 \le t < 0.85$	0	0	0	0	PDE
$0.85 \le t < 0.95$	0	0	0	6000	PDE
$0.95 \le t < 1$	0	0	100	0	PDE

TABLE 3. Example 2: effect of the tolerance ε_{tol} on the average number of iterations in the Newton-Raphson method, the errors e_N^{rel} and CPU times during a simulation of $T = 1\,\text{h}$. The errors are computed with the reference solution obtained by the explicit scheme with $N = 4800$ and the Godunov numerical flux.

|	ε_{tol}	Avg. iterations	$e_N^{\text{rel}}(t)$	CPU [s]
	1E-1	1.00	0.0497756500328602	0.1250
	1E-2	1.03	0.0497756500328602	0.1406
	1E-3	1.60	0.0497679356628912	0.1797
	1E-4	2.00	0.0497674399512266	0.1953
	1E-5	2.01	0.0497674389897926	0.1875
$N = 50$	1E-6	2.04	0.0497674143163951	0.1797
	1E-7	2.12	0.0497674123749513	0.1953
	1E-8	2.77	0.0497674121051164	0.2109
	1E-9	2.92	0.0497674120899535	0.2344
	1E-10	2.96	0.0497674120878119	0.2188
	1E-11	3.07	0.0497674120875128	0.2422
	1E-12	3.69	0.0497674120874927	0.2560
	1E-1	1.00	0.0244467141865713	0.3984
	1E-2	1.01	0.0244467141865713	0.4062
	1E-3	1.21	0.0244432794902836	0.4453
	1E-4	2.00	0.0244427209345141	0.6719
	1E-5	2.01	0.0244427236620445	0.5938
$N = 100$	1E-6	2.02	0.0244427043886650	0.6172
	1E-7	2.06	0.0244427032349412	0.6250
	1E-8	2.17	0.0244427031258883	0.7188
	1E-9	2.87	0.0244427030983648	0.7656
	1E-10	2.97	0.0244427030958998	0.7812
	1E-11	3.01	0.0244427030956931	0.7734
	1E-12	3.11	0.0244427030956539	0.8672
	1E-1	1.00	0.0115032639159754	1.4609
	1E-2	1.00	0.0115032639159754	1.4766
	1E-3	1.07	0.0115029463554388	1.6406
	1E-4	1.86	0.0115009448742330	2.0234
	1E-5	2.01	0.0115008746422637	2.2578
$N = 200$	1E-6	2.02	0.0115008746587289	2.2188
	1E-7	2.03	0.0115008741995477	2.1250
	1E-8	2.08	0.0115008740687287	2.2656
	1E-9	2.24	0.0115008740547202	2.3281
	1E-10	2.91	0.0115008740528361	2.7812
	1E-11	3.00	0.0115008740526303	2.7812
	1E-12	3.03	0.0115008740526086	2.9609
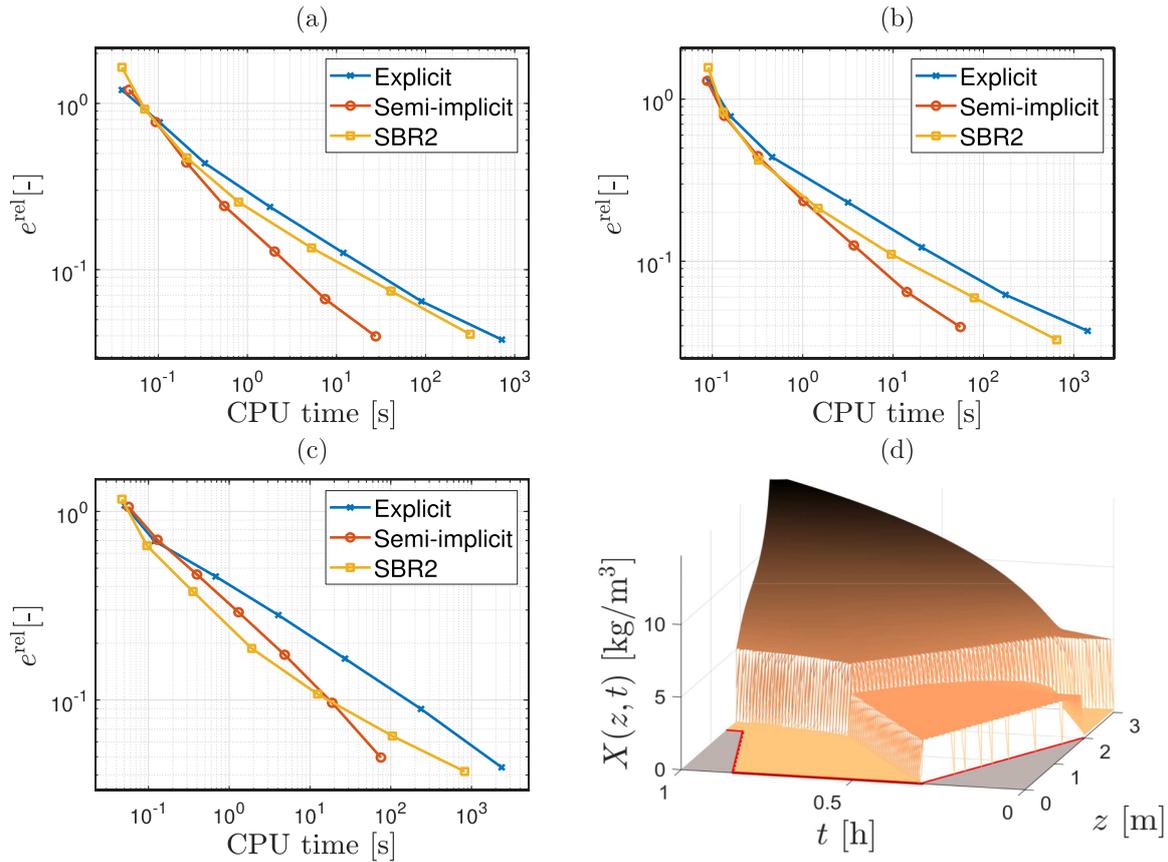
FIGURE 12. Example 2: error *versus* CPU time for various values of N (Tab. 4) at simula times (a) $T = 0.4$ h, (b) $T = 0.75$ h and (c) $T = 1$ h; (d) simulated total suspended solids concentration ($N = 4800$) during 1 h (reference solution).

and efficiency of the semi-implicit version if compared with the explicit version. This gain is of course due to the more favorable CFL condition (CFL-SI) (for the semi-implicit scheme) that permits a larger time step than (CFL) (for the explicit scheme). The degree of this advantage depends, of course, on the constitutive functions and parameters used, so it is important to emphasize that it appears for choices of these constitutive ingredients that are largely considered typical for activated sludge [12, 18, 37, 43].

The SBR2 scheme of [16] is based on a fixed grid with respect to z, and the moving interface $\bar{z}(t)$ needs to be tracked explicitly. This movement gives rise to a number of cases of boundary cells of time-dependent size that need to be handled separately. That treatment is fairly involved and the approach presented in the present work is in general easier to implement, and for the semi-implicit variant turns out to be more efficient. That said, we mention that the fixed-grid approach of [16] has certain advantages in situations when the method is required to leave a consolidated bed invariant under movements of $\bar{z}(t)$ above it (as is illustrated by Example 3), and can be extended more easily to additional inlets or outlets in the reactor. It is therefore interesting to note that the observed convergence of the three schemes discussed to the same solution, as evidenced by Table 4 and Figure 12a, supports that the treatment in [16] is correct.

With respect to limitations, we mention that the approach of [16] involves a vessel with variable cross-sectional area; this model ingredient has been left out here for notational convenience and is not expected to make the
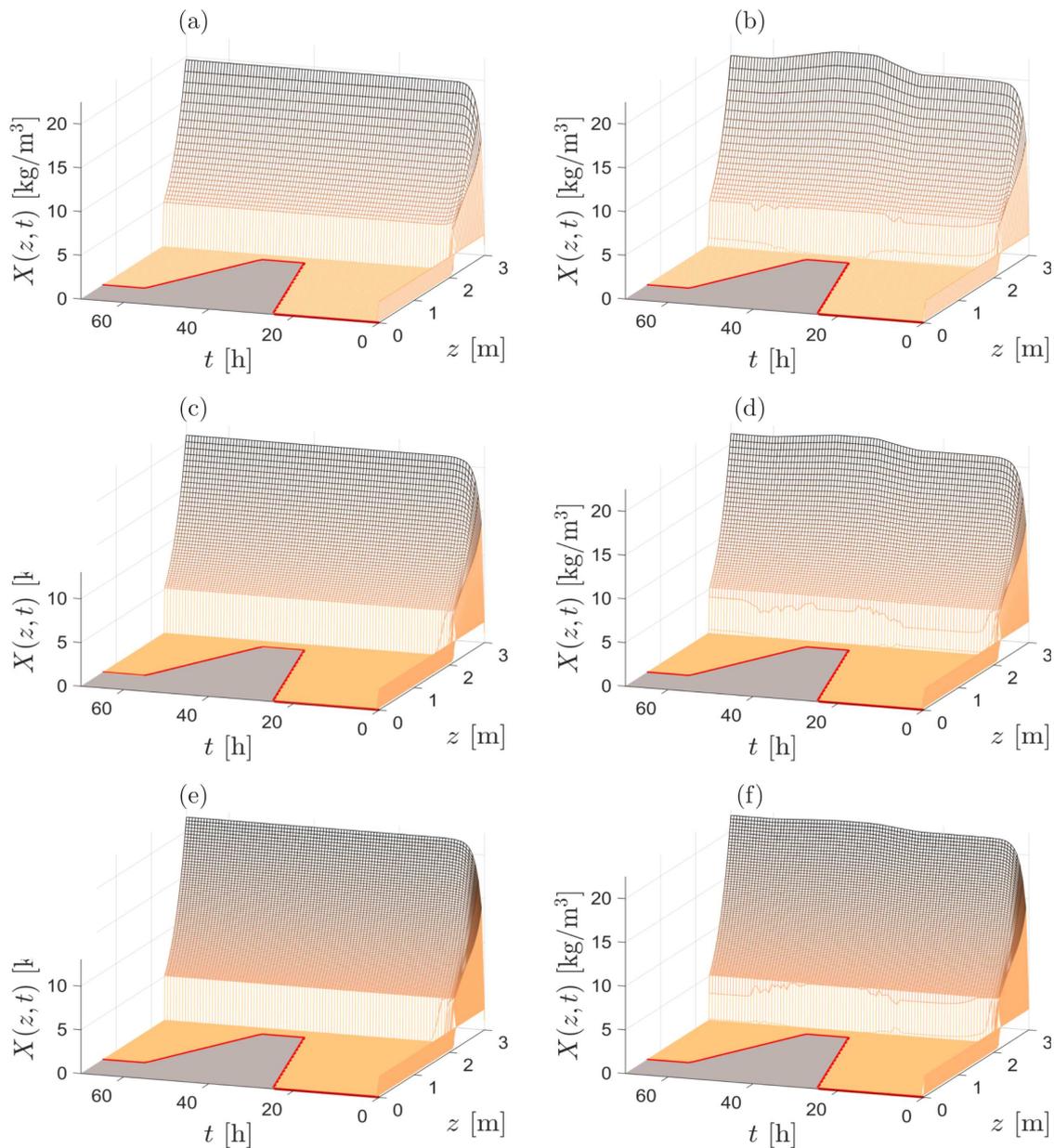
FIGURE 13. Example 3: simulations of the scenario of Table 5 with the indicated schemes and values of N. The fixed-mesh scheme SBR2 reproduces a steady-state profile while numerical solutions by scheme SIP are affected by the moving mesh. (a) SBR2, $N = 100$. (b) SIP, $N = 100$. (c) SBR2, $N = 200$. (d) SIP, $N = 200$. (e) SBR2, $N = 400$. (f) SIP, $N = 400$.

TABLE 4. Example 2: errors e_N^{rel} and CPU times at simulated time $T = 1\,\mathrm{h}$. The errors have been computed with the reference solution obtained by explicit scheme with $N = 4800$ and the Godunov numerical flux. The abbreviation SBR2 refers to the scheme in [16] (without variable transformation).

		SBR2		Explicit		Semi-implicit	
	N	$e_N^{\mathrm{rel}}(t)$	CPU [s]	$e_N^{\mathrm{rel}}(t)$	CPU [s]	$e_N^{\mathrm{rel}}(t)$	CPU [s]
	25	1.6483	0.0391	1.2053	0.0391	1.2099	0.0469
	50	0.9230	0.0703	0.7688	0.1035	0.7732	0.0938
	100	0.4704	0.2090	0.4368	0.3340	0.4414	0.2070
$T = 0.4\,\mathrm{h}$	200	0.2551	0.7969	0.2384	1.7969	0.2416	0.5508
	400	0.1352	5.2441	0.1261	11.9395	0.1286	2.0234
	800	0.0743	40.9023	0.0645	89.9941	0.0665	7.4453
	1600	0.0409	316.4102	0.0379	717.4668	0.0397	27.5156
	25	1.5691	0.0900	1.2901	0.0920	1.2959	0.0880
	50	0.8327	0.1320	0.7835	0.1590	0.7880	0.1360
	100	0.4204	0.3240	0.4392	0.4590	0.4451	0.3190
$T = 0.75\,\mathrm{h}$	200	0.2125	1.4766	0.2304	3.1797	0.2345	1.0234
	400	0.1103	9.5469	0.1218	20.7266	0.1250	3.6797
	800	0.0595	79.1719	0.0620	175.6562	0.0644	14.2188
	1600	0.0327	646.8438	0.0371	1421.2190	0.0392	55.6470
	25	1.1617	0.0469	1.0747	0.0508	1.0573	0.0566
	50	0.6580	0.0957	0.6966	0.1191	0.7078	0.1289
	100	0.3760	0.3555	0.4519	0.6797	0.4627	0.3984
$T = 1\,\mathrm{h}$	200	0.1877	1.9004	0.2821	4.0605	0.2919	1.3066
	400	0.1075	12.4375	0.1658	27.0586	0.1737	4.8535
	800	0.0644	104.6543	0.0896	237.3340	0.0966	18.8379
	1600	0.0418	822.3809	0.0439	2362.8570	0.0495	75.5801

TABLE 5. Example 3: Time functions for the simulation.

Time period [h]	$X_{\mathrm{f}}(t)[\mathrm{kg\,m^{-3}}]$	$Q_{\mathrm{f}}(t)[\mathrm{m^3\,h^{-1}}]$	$Q_{\mathrm{u}}(t)[\mathrm{m^3\,h^{-1}}]$	$Q_{\mathrm{e}}(t)[\mathrm{m^3\,h^{-1}}]$	Model
$0 \le t < 25$	0	0	0	0	PDE
$25 \le t < 35$	0	0	0	84	PDE
$35 \le t < 45$	0	0	0	0	PDE
$45 \le t < 60$	0	40	0	0	PDE
$60 \le t < 70$	0	0	0	0	PDE

formulation and analysis significantly more difficult. Another limitation of the proposed model is that all solid particles are assumed to be of the same size, which is an approximation of the mean of a real size distribution. Assuming more than one size of particles with different settling behaviour would lead to a more complicated system of model equations (*cf.*, *e.g.*, [2]). From an application point of view, it would be valuable to model both fast and slowly settling particles.

We also note that while it has been proven that the schemes are monotone and obey an invariant-region principle and numerical evidence of convergence has been presented, a rigorous convergence proof is still missing but can possibly be obtained by combining analyses of weakly coupled degenerate parabolic systems [32], numerical methods for zero-flux boundary value problems [10, 36], and studies of triangular systems of conservation laws and degenerate convection-diffusion equations [17, 19, 22]. (In the latter reference, convergence proofs decisively

depend on usage of the Engquist-Osher flux with its separable upwind and downwind contributions; this has partly motivated usage of that particular numerical flux in the present work.)

Several extensions, improvements, and applications of the model and numerical methods are conceivable. For instance, both scheme versions are only first-order accurate in space and time; this order can possibly be improved by applying more sophisticated variable extrapolation or maximum-principle-preserving weighted essentially non-oscillatory (WENO) techniques (*cf.*, *e.g.*, [35] and references cited in that work) combined with changing between the implicit and explicit steps in a more involved manner as is done in implicit-explicit (IMEX) schemes, *cf.*, *e.g.*, [3]. Such a procedure could be obtained in a natural way from the two-stage formulation (steps (1) and (2)) of the present semi-implicit formulation of Section 4.1. Even further savings and gains in efficiency could arise from a fully implicit scheme (without any bounds on $\tau/\Delta\xi$ or even on τ at all). Such a scheme would require the solution of large systems of nonlinear equations, considering that while the present semi-implicit approach requires solving $N + 1$ nonlinear equations per time step, namely (4.4), to update X, the fully implicit approach would require solving a nonlinear system of $(k_{\boldsymbol{C}} + k_{\boldsymbol{S}})(N + 1)$ equations per time step. Furthermore, additional nonlinearities to those related to the degenerating diffusion terms in the present approach, namely those that correspond to the nonlinear convection and reaction terms, would have to be evaluated "implicitly" (with unknowns at the new time step). Moreover, while there is a theory of implicit monotone schemes for nonlinear conservation laws (see, *e.g.*, [4–6]), it is not clear whether such schemes can straightforwardly be extended to conservation laws with discontinuous flux, which is the situation at hand due to the flux discontinuities at $\xi = 0$ and $\xi = 1$. Thus, by theoretical and practical considerations it seems that improvements of efficiency at acceptable cost of implementation are more likely to come from IMEX rather than fully implicit approaches.

Furthermore, one can think of an SBR whose bottom works as a filter or a filtration cell with chemical reactions. In such applications, the trajectory of $\bar{z}(t)$ is no longer prescibed but arises from an applied pressure, and is balanced by the hydraulic resistance of the filter medium plus that of the forming sediment (filter cake). The composition of the latter is part of the solution, so that possible extension forms a free-boundary problem [7, 8].

We emphasize that within the present approach soluble components are supposed to be transported passively with the fluid and are subject to the reaction terms. In some of the figures this property causes fairly sharp concentration profiles when one expects that the evolution of solute concentrations should be subject to additional effects such as dispersion/diffusion. In fact, diffusion driven by the gradient of the respective concentration was the unique mechanism of spatial propagation of solutes considered in our first effort of modelling reactive settling by PDEs [12]. It would not be a problem to include diffusion to the present approach as an additional mechanism of solute transport, with the effect to be an overall blurring of profiles.

APPENDIX A. A MODIFIED ASM1 MODEL

The model ASM1 is described, for example, in [30]. For the reason of reformulating the PDE model to include percentages of the particulate concentrations, we have redefined the second component from X_{S} to $X_{\mathrm{S-ND}} := X_{\mathrm{S}} - X_{\mathrm{ND}}$. Then the variables are

$$\boldsymbol{C} = \left(X_{\mathrm{I}}, X_{\mathrm{S-ND}}, X_{\mathrm{B,H}}, X_{\mathrm{B,A}}, X_{\mathrm{P}}, X_{\mathrm{ND}}\right)^{\mathrm{T}}, \quad \boldsymbol{S} = \left(S_{\mathrm{I}}, S_{\mathrm{S}}, S_{\mathrm{O}}, S_{\mathrm{NO}}, S_{\mathrm{NH}}, S_{\mathrm{ND}}\right)^{\mathrm{T}},$$

with units given in Table A.1. The stoichiometric matrices of the modified ASM1 are given by

$$\boldsymbol{\sigma}_{\boldsymbol{C}} := \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 - f_{\mathrm{P}}(1 + i_{\mathrm{XP}}) - i_{\mathrm{XB}} & 1 - f_{\mathrm{P}}(1 + i_{\mathrm{XP}}) - i_{\mathrm{XB}} & 0 & -1 & 1 \\ 1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{\mathrm{P}} & f_{\mathrm{P}} & 0 & 0 & 0 \\ 0 & 0 & 0 & i_{\mathrm{XB}} - f_{\mathrm{P}}i_{\mathrm{XP}} & i_{\mathrm{XB}} - f_{\mathrm{P}}i_{\mathrm{XP}} & 0 & 0 & -1 \end{bmatrix}$$

TABLE A.1. The variables in the modified ASM1 model.

Material	Notation	Unit
Particulate inert organic matter	X_{I}	$(\mathrm{kg\,COD})\,\mathrm{m}^{-3}$
Slowly biodegradable substrate less part. organic nitrogen	$X_{\mathrm{S-ND}}$	$(\mathrm{kg\,COD})\,\mathrm{m}^{-3}$
Active heterotrophic biomass	$X_{\mathrm{B,H}}$	$(\mathrm{kg\,COD})\,\mathrm{m}^{-3}$
Active autotrophic biomass	$X_{\mathrm{B,A}}$	$(\mathrm{kg\,COD})\,\mathrm{m}^{-3}$
Particulate products arising from biomass decay	X_{P}	$(\mathrm{kg\,COD})\,\mathrm{m}^{-3}$
Particulate biodegradable organic nitrogen	X_{ND}	$(\mathrm{kg\,N})\,\mathrm{m}^{-3}$
Soluble inert organic matter	S_{I}	$(\mathrm{kg\,COD})\,\mathrm{m}^{-3}$
Readily biodegradable substrate	S_{S}	$(\mathrm{kg\,COD})\,\mathrm{m}^{-3}$
Oxygen	S_{O}	$-(\mathrm{kg\,COD})\,\mathrm{m}^{-3}$
Nitrate and nitrite nitrogen	S_{NO}	$(\mathrm{kg\,N})\,\mathrm{m}^{-3}$
$\mathrm{NH_4^+ + NH_3}$ nitrogen	S_{NH}	$(\mathrm{kg\,N})\,\mathrm{m}^{-3}$
Soluble biodegradable organic nitrogen	S_{ND}	$(\mathrm{kg\,N})\,\mathrm{m}^{-3}$

TABLE A.2. Stoichiometric and kinetic parameters.

Symbol	Name	Value	Unit
Y_{A}	Yield for autotrophic biomass	0.24	$(\mathrm{g\,COD})(\mathrm{g\,N})^{-1}$
Y_{H}	Yield for heterotrophic biomass	0.67	$(\mathrm{g\,COD})(\mathrm{g\,COD})^{-1}$
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	$(\mathrm{g\,N})(\mathrm{g\,COD})^{-1}$
i_{XP}	Mass of nitrogen per mass of COD in products from biomass	0.06	$(\mathrm{g\,N})(\mathrm{g\,COD})^{-1}$
μ_{H}	Maximum specific growth rate for heterotrophic biomass	6.0	d^{-1}
K_{S}	Half-saturation coefficient for heterotrophic biomass	20.0	$(\mathrm{g\,COD})\,\mathrm{m}^{-3}$
$K_{\mathrm{O,H}}$	Oxygen half-saturation coefficient for heterotrophic biomass	0.2	$-(\mathrm{g\,COD})\,\mathrm{m}^{-3}$
K_{NO}	Nitrate half-saturation coefficient for denitrifying heterotrophic biomass	0.5	$(\mathrm{g\,NO_3\text{-}N})\,\mathrm{m}^{-3}$
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	$(\mathrm{g\,COD})\,(\mathrm{g\,COD})^{-1}\mathrm{d}^{-1}$
K_{X}	Half-saturation coefficient for hydrolysis of slowly biodegradable substrate	0.03	$(\mathrm{g\,COD})(\mathrm{g\,COD})^{-1}$
μ_{A}	Maximum specific growth rate for autotrophic biomass	0.8	d^{-1}
$\bar{K}_{\mathrm{NH}}$	Ammonia half-saturation coefficient for aerobic and anaerobic growth of heterotrophs	0.05	$(\mathrm{g\,NH_3\text{-}N})\,\mathrm{m}^{-3}$
K_{NH}	Ammonia half-saturation coefficient for autotrophic biomass	1.0	$(\mathrm{g\,NH_3\text{-}N})\,\mathrm{m}^{-3}$
b_{A}	Decay coefficient for autotrophic biomass	0.15	d^{-1}
$K_{\mathrm{O,A}}$	Oxygen half-saturation coefficient for autotrophic biomass	0.4	$-(\mathrm{g\,COD})\,\mathrm{m}^{-3}$
k_{a}	Ammonification rate	0.08	$\mathrm{m}^3(\mathrm{g\,COD})^{-1}\mathrm{d}^{-1}$

for the solid components and

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

for the substrates, where the constants are given in Table A.2 and the vector of reaction rates is

$$
\boldsymbol{r}(\boldsymbol{C}, \boldsymbol{S}) := \begin{pmatrix}
\mu_{\mathrm{H}}\mu(S_{\mathrm{NH}}, \bar{K}_{\mathrm{NH}})\mu(S_{\mathrm{S}}, K_{\mathrm{S}})\mu(S_{\mathrm{O}}, K_{\mathrm{O,H}})X_{\mathrm{B,H}} \\
\mu_{\mathrm{H}}\mu(S_{\mathrm{NH}}, \bar{K}_{\mathrm{NH}})\mu(S_{\mathrm{S}}, K_{\mathrm{S}})\mu(K_{\mathrm{O,H}}, S_{\mathrm{O}})\mu(S_{\mathrm{NO}}, K_{\mathrm{NO}})\eta_{\mathrm{g}}X_{\mathrm{B,H}} \\
\mu_{\mathrm{A}}\mu(S_{\mathrm{NH}}, K_{\mathrm{NH}})\mu(S_{\mathrm{O}}, K_{\mathrm{O,A}})X_{\mathrm{B,A}} \\
b_{\mathrm{H}}X_{\mathrm{B,H}} \\
b_{\mathrm{A}}X_{\mathrm{B,A}} \\
k_{\mathrm{a}}S_{\mathrm{ND}}X_{\mathrm{B,H}} \\
k_{\mathrm{h}}\mu_7(X_{\mathrm{S}}, X_{\mathrm{B,H}})\big(\mu(S_{\mathrm{O}}, K_{\mathrm{O,H}}) + \eta_{\mathrm{h}}\mu(K_{\mathrm{O,H}}, S_{\mathrm{O}})\mu(S_{\mathrm{NO}}, K_{\mathrm{NO}})\big) \\
k_{\mathrm{h}}\mu_8(X_{\mathrm{B,H}}, X_{\mathrm{ND}})\big(\mu(S_{\mathrm{O}}, K_{\mathrm{O,H}}) + \eta_{\mathrm{h}}\mu(K_{\mathrm{O,H}}, S_{\mathrm{O}})\mu(S_{\mathrm{NO}}, K_{\mathrm{NO}})\big)
\end{pmatrix}.
$$

Here we define the Monod expression $\mu(A, B) := A/(A + B)$ and

$$
\mu_7(X_{\mathrm{S}}, X_{\mathrm{B,H}}) := \begin{cases} 0 & \text{if } X_{\mathrm{S}} = 0 \text{ and } X_{\mathrm{B,H}} = 0, \\ \dfrac{X_{\mathrm{S}}X_{\mathrm{B,H}}}{K_{\mathrm{X}}X_{\mathrm{B,H}} + X_{\mathrm{S}}} & \text{otherwise,} \end{cases}
$$

$$
\mu_8(X_{\mathrm{B,H}}, X_{\mathrm{ND}}) := \begin{cases} 0 & \text{if } X_{\mathrm{S}} = 0 \text{ and } X_{\mathrm{B,H}} = 0, \\ \dfrac{X_{\mathrm{B,H}}X_{\mathrm{ND}}}{K_{\mathrm{X}}X_{\mathrm{B,H}} + X_{\mathrm{S}}} & \text{otherwise.} \end{cases}
$$

APPENDIX B. LIST OF SYMBOLS

The variables in the ASM1 model are given in Table A.1 and the stoichiometric and kinetic parameters in Table A.2. The notation '(k)' in $C^{(k)}$ means component no. k of a vector $\boldsymbol{C}$. The tilde symbol $\tilde{\square}$ is used temporarily after the spatial transformation of the model. The rest of the symbols used are listed below, except for local variables used in the proofs and similar derivations.

Latin symbols

A	cross-sectional area of tank $[\mathrm{m}^2]$
$a(X)$	compression function defined in (2.17) $[\mathrm{m}^2\,\mathrm{s}^{-1}]$
B	depth of tank $[\mathrm{m}]$
B_{c}	critical depth of surface $[\mathrm{m}]$
$\mathcal{B}_{j+1/2}^n$	numerical flux of bulk defined by (3.5) $[\mathrm{kg}\,(\mathrm{m}^{-3}\,\mathrm{s}^{-1})]$
C_1, C_2	coefficients in CFL condition (CFL)
$\boldsymbol{C} = (C^{(1)}, \dots, C^{(k_{\boldsymbol{C}})})^{\mathrm{T}}$	vector of solid phase concentrations $[\mathrm{kg}\,\mathrm{m}^{-3}]$
c	unit conversion factor defined in (2.3) $[\mathrm{g}/(\mathrm{g}\ \mathrm{COD})]$
$D(X)$	integrated compression function defined in (2.9) $[\mathrm{m}^2\,\mathrm{s}^{-1}]$
$\mathcal{D}(X)$	integrated compression function defined by (2.17) $[\mathrm{kg}\,(\mathrm{m}^{-1}\,\mathrm{s}^{-1})]$
$d(X)$	compression function defined in (2.9) $[\mathrm{m}^5/(\mathrm{kg}\ \mathrm{s})]$
$\mathcal{E}_{j+1/2}^n$	Engquist-Osher numerical flux defined by (3.3) $[\mathrm{kg}\,(\mathrm{m}^{-2}\,\mathrm{s}^{-1})]$
$e_N^{\mathrm{rel}}(t)$	relative approximate numerical error defined by (5.2) $[-]$
$F(\xi, t, X)$	flux function defined in (2.23) $[\mathrm{kg}\,(\mathrm{m}^{-3}\,\mathrm{s}{-}1)]$
$\mathcal{F}_{j+1/2}^n$	numerical flux of bulk and sedimentation defined by (3.6) $[\mathrm{kg}\,(\mathrm{m}^{-3}\,\mathrm{s}^{-1})]$
$f(X)$	flux function of hindered sedimentation defined in (2.16) $[\mathrm{kg}\,(\mathrm{m}^{-2}\,\mathrm{s}^{-1})]$
H	Heaviside function $[-]$
$I_{\boldsymbol{C},k}^-, I_{\boldsymbol{C},k}^+$	index sets of negative/positive stoichiometric coefficients, *cf.* (2.7)
I_j	interval of cell for numerical schemes $[-]$
$\mathcal{J}_{j+1/2}^n$	numerical flux of compression defined by (3.4) $[\mathrm{kg}\,(\mathrm{m}^{-3}\,\mathrm{s}^{-1})]$
$\boldsymbol{\mathcal{J}}(\boldsymbol{X})$	Jacobian matrix of left-hand side of (4.4) $[-]$

$k_{\boldsymbol{C}}, k_{\boldsymbol{S}}$	number of solid/substrate phase components [–]
$M_{q1}, M_{q2}, M_R, M_{\boldsymbol{C}}, M_{\boldsymbol{S}}$	coefficients in CFL condition (CFL)
N	number of spacial cells in numerical schemes covering the mixture [–]
$\boldsymbol{p}$	vector of percentages defined in (2.5) [–]
$\bar{Q}(t)$	volumetric flow in effluent pipe [$\mathrm{m}^3\,\mathrm{s}^{-1}$]
q	bulk velocity [$\mathrm{m}\,\mathrm{s}^{-1}$]
q_{out}^n	$= q_{\mathrm{u}}^n + q_{\mathrm{e}}^n$ [s^{-1}]
$\tilde{q}(\xi, t)$	transformed bulk velocity [s^{-1}]
$R(\boldsymbol{C}, \boldsymbol{S})$	reaction term of total solids concentration defined in (2.23)
$\boldsymbol{R_C}, \boldsymbol{R_S}$	vectors of reaction terms [$\mathrm{kg}\,(\mathrm{m}^{-3}\,\mathrm{s}{-1})$]
$\boldsymbol{r}$	vector of reaction processes [$\mathrm{kg}\,(\mathrm{m}^{-3}\,\mathrm{s}{-1})$]
$\boldsymbol{S} = (S^{(1)}, \ldots, S^{(k_{\boldsymbol{S}})})^{\mathrm{T}}$	vector of substrate concentrations [$\mathrm{kg}\,\mathrm{m}^{-3}$]
T	total time of modelling or simulation [s]
$\boldsymbol{T}$	Toeplitz matrix defined in (4.4)
t	time [s]
$\mathcal{U}$	collection of solution variables
$U_{\boldsymbol{C}}, U_{\boldsymbol{S}}$	total velocity functions after transformation, defined in (2.18) [s^{-1}]
$\mathcal{V}_{\boldsymbol{C}}, \mathcal{V}_{\boldsymbol{S}}$	total velocity functions in model (2.11) [$\mathrm{m}\,\mathrm{s}^{-1}$]
$\bar{V}(t)$	volume of mixture at time t [m^3]
v_0	constant in settling-velocity function (5.1) [$\mathrm{m}\,\mathrm{s}^{-1}$]
$v_{\mathrm{hs}}(X)$	constitutive function for hindered sedimentation [$\mathrm{m}\,\mathrm{s}^{-1}$]
X	total solids concentration defined in (2.4) [$\mathrm{kg}\,\mathrm{m}^{-3}$]
X^{t}	concentration from which tangent of v_{hs} is defined [$\mathrm{kg}\,\mathrm{m}^{-3}$]
$\check{X}$	constant in settling-velocity function (5.1) [$\mathrm{kg}\,\mathrm{m}^{-3}$]
$\boldsymbol{X}$	vector of numerical concentrations for cells covering mixture [$\mathrm{kg}\,\mathrm{m}^{-3}$]
$\boldsymbol{\mathcal{X}}$	vector of numerical concentrations for all cells [$\mathrm{kg}\,\mathrm{m}^{-3}$]
X_{c}	critical concentration above which solids are in contact [$\mathrm{kg}\,\mathrm{m}^{-3}$]
X^*	maximizer of f [$\mathrm{kg}\,\mathrm{m}^{-3}$]
$\hat{X}$	maximum total solids concentration [$\mathrm{kg}\,\mathrm{m}^{-3}$]
z	depth from top of tank [m]
$\bar{z}(t)$	depth of moving surface given by (2.2) [$\mathrm{m}\,\mathrm{s}^{-1}$]

Greek symbols

$\alpha(\xi, t)$	coefficient in transformed model, defined in (2.14) [s^{-1}]
$\beta(t)$	coefficient in transformed model, defined in (2.14) [m^{-1}]
$\gamma(z, t), \gamma(\xi)$	characteristic functions for mixture [–]
$\Delta\xi$	spatial mesh width for numerical scheme [–]
$\Delta\rho$	$= \rho_X - \rho_L$ [$\mathrm{kg}\,\mathrm{m}^{-3}$]
$[\Delta\square]_j^n$	numerical divergence operator [–]
$\delta(z), \delta(\xi)$	Dirac delta functions [m^{-1}], [–]
$\delta_{j,0}$	Kronecker delta defined in (3.1)
ε	small number in technical assumption (2.6)
$\varepsilon_{\mathrm{tol}}$	tolerance in Newton iterations in Section 4.2 [–]
ζ	$= 1/(B - B_{\mathrm{c}})$ [m^{-1}]
η	constant in settling-velocity function (5.1) [–]
κ_j^n	coefficient defined by (3.2)
$\boldsymbol{\Lambda_C}, \boldsymbol{\Lambda_S}$	diagonal matrices of solid/substrate conversion factors [–]
λ	$= \tau/\Delta\xi$ [s]
μ	$= \tau/\Delta\xi^2$ [s]

ξ	space coordinate in transformed system, defined in (2.13) [–]
ρ_X, ρ_L	density of solids/liquid [kg m^{-3}]
$\boldsymbol{\sigma_C}, \boldsymbol{\sigma_S}$	stoichiometric matrices [–]
σ_0	constant in effective-solids-stress function (5.1) [kg (m^{-2} s-1)]
$\sigma_e(X)$	constitutive function for effective solids stress [kg (m^{-2} s-1)]
τ	time step in numerical schemes [s]
$\Phi_{j+1/2}^n$	numerical flux of total concentration defined by (3.7) [kg (m^{-3} s-1)]
χ_ω	characteristic function; $\chi_\omega = 1$ if ω is true, otherwise $\chi_\omega = 0$ [–]

Indices

$\square_c$	critical
$\square_e$	effluent, extraction
$\square_f$	feed
$\square_{hs}$	hindered sedimentation
$\square_{out}$	bulk velocity out of the tank
$\square_{pipe}$	effluent pipe
$\square_u$	underflow
$\square_j^n$	numerical approximation of a variable in cell j at time $n\tau$

References

[1] M.M. Amin, M.H. Khiadani (Hajian), A. Fatehizadeh and E. Taheri, Validation of linear and non-linear kinetic modeling of saline wastewater treatment by sequencing batch reactor with adapted and non-adapted consortiums. *Desalination* **344** (2014) 228–235.

[2] S. Berres, R. Bürger, K.H. Karlsen and E.M. Tory, Strongly degenerate parabolic-hyperbolic systems modeling polydisperse sedimentation with compression. *SIAM J. Appl. Math.* **64** (2003) 41–80.

[3] S. Boscarino, R. Bürger, P. Mulet, G. Russo and L.M. Villada, Linearly implicit IMEX Runge-Kutta methods for a class of degenerate convection-diffusion problems. *SIAM J. Sci. Comput.* **37** (2015) B305–B331.

[4] M. Breuß, A theory of implicit methods for scalar conservation laws. In *Hyperbolic Problems: Theory, Numerics, Applications*. Springer, Berlin (2003) 377–386.

[5] M. Breuß, The implicit upwind method for 1-D scalar conservation laws with continuous fluxes. *SIAM J. Numer. Anal.* **43** (2005) 970–986.

[6] M. Breuß and A. Kleefeld, Implicit monotone difference methods for scalar conservation laws with source terms. *Acta Math. Vietnam.* **45** (2020) 709–738.

[7] R. Bürger, F. Concha and K.H. Karlsen, Phenomenological model of filtration processes: 1. Cake formation and expression. *Chem. Eng. Sci.* **56** (2001) 4537–4553.

[8] R. Bürger, H. Frid and K.H. Karlsen, On a free boundary problem for a strongly degenerate quasi-linear parabolic equation with an application to a model of pressure filtration. *SIAM J. Math. Anal.* **34** (2002) 611–635.

[9] R. Bürger, K.H. Karlsen and J.D. Towers, A model of continuous sedimentation of flocculated suspensions in clarifier-thickener units. *SIAM J. Appl. Math.* **65** (2005) 882–940.

[10] R. Bürger, A. Coronel and M. Sepúlveda, 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** (2006) 91–112.

[11] R. Bürger, S. Diehl and I. Nopens, A consistent modelling methodology for secondary settling tanks in wastewater treatment. *Water Res.* **45** (2011) 2247–2260.

[12] R. Bürger, J. Careaga, S. Diehl, C. Mejías, I. Nopens, E. Torfs and P.A. Vanrolleghem, Simulations of reactive settling of activated sludge with a reduced biokinetic model. *Comput. Chem. Eng.* **92** (2016) 216–229.

[13] R. Bürger, S. Diehl and C. Mejías, A difference scheme for a degenerating convection-diffusion-reaction system modelling continuous sedimentation. *ESAIM: Math. Model. Numer. Anal.* **52** (2018) 365–392.

[14] R. Bürger, J. Careaga and S. Diehl, A method-of-lines formulation for a model of reactive settling in tanks with varying cross-sectional area. *IMA J. Appl. Math.* **86** (2021) 514–546.

[15] R. Bürger, J. Careaga, S. Diehl and R. Pineda, A moving-boundary model of reactive settling in wastewater treatment. Part 1: Governing equations. *Appl. Math. Model.* **106** (2022) 390–401.

[16] R. Bürger, J. Careaga, S. Diehl and R. Pineda, A moving-boundary model of reactive settling in wastewater treatment. Part 2: Numerical scheme. *Appl. Math. Model.* **111** (2022) 247–269.

[17] R. Bürger, S. Diehl, M.C. Martí and Y. Vásquez, A degenerating convection-diffusion system modelling froth flotation with drainage. *IMA J. Appl. Math.* **87** (2022) 1151–1190.

[18] R. Bürger, J. Careaga, S. Diehl and R. Pineda, A model of reactive settling of activated sludge: Comparison with experimental data. *Chem. Eng. Sci.* **267** (2023) 118244.

[19] R. Bürger, S. Diehl, M.C. Martí and Y. Vásquez, A difference scheme for a triangular system of conservation laws with discontinuous flux modeling three-phase flows. *Netw. Heterog. Media* **18** (2023) 140–190.

[20] M. Caluwé, D. Daens, R. Blust, L. Geuens and J. Dries, The sequencing batch reactor as an excellent configuration to treat wastewater from the petrochemical industry. *Water Sci. Tech.* **75** (2016) 793–801.

[21] G. Chen, M.C.M. van Loosdrecht, G.A. Ekama and D. Brdjanovic, Biological Wastewater Treatment, 2nd edition. IWA Publishing, London, UK (2020).

[22] G.M. Coclite, S. Mishra and N.H. Risebro, Convergence of an Engquist-Osher scheme for a multi-dimensional triangular system of conservation laws. *Math. Comput.* **79** (2010) 71–94.

[23] R. Courant, K. Friedrichs and H. Lewy, Über die partiellen Differenzengleichungen der mathematischen Physik. *Math. Ann.* **100** (1928) 32–74.

[24] K. Deimling, Nonlinear Functional Analysis. Springer-Verlag, Berlin (1985).

[25] S. Diehl, Continuous sedimentation of multi-component particles. *Math. Meth. Appl. Sci.* **20** (1997) 1345–1364.

[26] R. Droste and R. Gear, Theory and Practice of Water and Wastewater Treatment, 2nd edition. Wiley, Hoboken, NJ, USA (2019).

[27] B. Engquist and S. Osher, One-sided difference approximations for nonlinear conservation laws. *Math. Comput.* **36** (1981) 321–351.

[28] S. Evje and K.H. Karlsen, Degenerate convection-diffusion equations and implicit monotone difference schemes. In *Hyperbolic Problems: Theory, Numerics, Applications, Vol. I (Zürich, 1998)*, Vol. 129 of *Internat. Ser. Numer. Math.* Birkhäuser, Basel (1999) 285–294.

[29] R. Eymard, T. Gallouët and R. Herbin, Finite volume methods. In *Handbook of Numerical Analysis, Vol. VII*, Handb. Numer. Anal., VII, North-Holland, Amsterdam (2000) 713–1020.

[30] M. Henze, W. Gujer, T. Mino and M.C.M. van Loosdrecht, Activated Sludge Models ASM1, ASM2, ASM2d and ASM3. IWA Scientific and Technical Report No. 9. IWA Publishing, London, UK (2000).

[31] J.S. Hesthaven, Numerical methods for conservation Laws, In Vol. 18 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2018).

[32] H. Holden, K.H. Karlsen and N.H. Risebro, On uniqueness and existence of entropy solutions of weakly coupled systems of nonlinear degenerate parabolic equations. *Electron. J. Differ. Equ.* **2003** (2003) 46.

[33] Z. Hu, R.A. Ferraina, J.F. Ericson, A.A. MacKay and B.F. Smets, Biomass characteristics in three sequencing batch reactors treating a wastewater containing synthetic organic chemicals. *Water Res.* **39** (2005) 710–720.

[34] W. Hundsdorfer and J. Verwer, Numerical solution of time-dependent advection-diffusion-reaction equations, In Vol. 33 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin (2003).

[35] Y. Jiang and Z. Xu, Parametrized maximum principle preserving limiter for finite difference WENO schemes solving convection-dominated diffusion equations. *SIAM J. Sci. Comput.* **35** (2013) A2524–A2553.

[36] K.H. Karlsen and J.D. Towers, Convergence of monotone schemes for conservation laws with zero-flux boundary conditions. *Adv. Appl. Math. Mech.* **9** (2017) 515–542.

[37] G. Kirim, E. Torfs and P.A. Vanrolleghem, An improved 1d reactive BürgerDiehl settler model for secondary settling tank denitrification. *Water Environ. Res.* **94** (2022) e10825.

[38] R.J. LeVeque, Finite Volume Methods for Hyperbolic Problems. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge (2002).

[39] J. Makinia and E. Zaborowska, Mathematical Modelling and Computer Simulation of Activated Sludge Systems, 2nd edition. IWA Publishing, London, UK (2020).

[40] L. Metcalf and H.P. Eddy, Wastewater Engineering. Treatment and Resource Recovery, 5th edition. McGraw-Hill, New York, USA (2014).

[41] T. Popple, J. Williams, E. May, G. Mills and R. Oliver, Evaluation of a sequencing batch reactor sewage treatment rig for investigating the fate of radioactively labelled pharmaceuticals: Case study of propranolol. *Water Res.* **88** (2016) 83–92.

[42] B. Song, Z. Tian, R. van der Weijden, C. Buisman and J. Weijma, High-rate biological selenate reduction in a sequencing batch reactor for recovery of hexagonal selenium. *Water Res.* **193** (2021) 116855.

[43] E. Torfs, S. Balemans, F. Locatelli, S. Diehl, R. Bürger, J. Laurent, P. François and I. Nopens, On constitutive functions for hindered settling velocity in 1-d settler models: Selection of appropriate model structure. *Water Res.* **110** (2017) 38–47.

[44] S. Wang and C.K. Gunsch, Effects of selected pharmaceutically active compounds on treatment performance in sequencing batch reactors mimicking wastewater treatment plants operations. *Water Res.* **45** (2011) 3398–3406.