

An unconditional energy stable data assimilation scheme for Navier-Stokes-Cahn-Hilliard equations with local discretized observed data

Xin Song^a, Qing Xia^a, Junseok Kim^b, Yibao Li^{a,*}

^a School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China

^b Department of Mathematics, Korea University, Seoul 02841, Republic of Korea

ARTICLE INFO

Keywords:

Date assimilation
Navier-Stokes-Cahn-Hilliard equations
Second-order accuracy
Unconditionally stability

ABSTRACT

In this paper, we introduce the modified Navier-Stokes-Cahn-Hilliard equations with a data assimilation term to utilize the information of the observed data. Based on the idea of feedback control, this term nudges the solutions to the observed data sampled from the reference process. By utilizing the Crank-Nicolson formula and the scalar auxiliary variable approach, we introduce an efficient numerical scheme for the modified Navier-Stokes-Cahn-Hilliard equations. Properties of the mass conservation and unconditional energy stability are proved. We explore the robustness and efficiency of the proposed scheme with various experiments.

1. Introduction

Multi-phase flows can be simulated by coupling the Navier-Stokes equations with the phase-field model, which governs the interfacial evolution of multi-phase fluid materials implicitly [1–3]. This coupled system models the dynamics of multi-phase fluids and preserves the total mass due to the properties of the CH equation [4,5]. However, in practical applications, the initial conditions and parameters may be inaccurate. The advancement of data acquisition equipment allows the utilization of observed data obtained from experiments or real processes to alleviate these issues. Data assimilation is a crucial approach to integrating observed data with a physical model. These methods have been proposed to estimate weather and ocean variables. Recently, data assimilation methods have found applications in numerical simulations of geophysical flows [6] and hemodynamics [7].

Existing data assimilation methods can be categorized into three classes: Bayesian-based methods, deep learning-based methods, and the nudging approach. The Bayesian-based method is an approach based on Bayesian statistics, including the Kalman Filter [8] and the Ensemble Kalman Filter [9]. It employs the forecast-correction step to incorporate observed data into the simulation process. Yamamura et al. [10] considered the flow of liquids and solid motion during dendrite growth and proposed a data assimilation algorithm based on the Ensemble Kalman Filter method. Similarly, Sasaki et al. [11] estimated unknown initial values and boundary conditions based on a data assimilation method with Ensemble Kalman Filter. By utilizing the observed data, they successfully estimated the true phase-field mobility during the simulation of the isothermal austenite-to-ferrite transformation in a Fe–C–Mn alloy. With the advancement of deep learning technology, deep learning-based data assimilation methods have garnered significant attention. Fablet et al. [12] focused on reconstructing sea surface dynamics and proposed an end-to-end inversion scheme based on the LSTM neural network. This scheme combines observed satellite data with the prior physical model and achieves significant progress in fine-scale sea surface dynamics reconstruction. To enhance the accuracy of simulation solutions, Pawar et al. [13] introduced a recurrent neural network to estimate the assimilation term. Arcucci et al. [14] proposed a framework named Deep Data Assimilation, which follows the idea of Bayesian approximation and employs a neural network to learn the assimilation process. By substituting the assimilation process, this method overcomes the limitations of observed data, such as zero error covariances, making it better suited for complex problems. Although the methods mentioned above have been widely applied, they often incur significant computational costs and require specific modifications for implementation.

To address the challenges of data assimilation, nudging has emerged as a promising method. By incorporating observed data into the physical model through an additional feedback control term, nudging allows the system variables to converge towards the observed data. Since this kind of method just adds a new term in the equations, it has high flexibility and tractability. The nudging method was proposed to estimate the initial

* Corresponding author.

E-mail address: yibaoli@xjtu.edu.cn (Y. Li).

<https://doi.org/10.1016/j.camwa.2024.03.018>

Received 4 August 2023; Received in revised form 20 January 2024; Accepted 11 March 2024

conditions of hurricane models using observed data [15]. Nudging has been employed to solve dissipative nonlinear partial differential equations (PDEs) including the two-dimensional Navier-Stokes equations [16,17], the two-dimensional surface quasi-geostrophic equation [18], the three-dimensional simplified Bardina model [19], the three-dimensional Brinkman-Forchheimer-extended Darcy model [20], and the three-dimensional Navier-Stokes equation [21]. Azouani et al. [16] derived conditions for the two-dimensional Navier-Stokes equations, ensuring that the simulated solution converges to the reference solution over time in finite-dimensional space. These conclusions have been extended to other scenarios, including the two-dimensional surface quasi-geostrophic equation [18], the three-dimensional simplified Bardina model [19], the three-dimensional Brinkman-Forchheimer-extended Darcy model [20], and the three-dimensional Navier-Stokes equations [21]. Foias et al. [22] demonstrated that with a sufficient number of densely distributed points, the solution of the governing equation can be determined by values at these points. They also introduced an algorithm that works with measurement data obtained discretely in time [23]. Carlson et al. [24] developed an algorithm that proves, theoretically, that the large-time error between the simulated solution and reference solution is controlled by the error in viscosity.

Up to now, various energy-stable numerical schemes have been proposed to preserve the energy dissipation property of the discretized scheme [25–29]. Shen et al. [30] proposed a semi-implicit unconditionally energy-stable scheme for CH equations and conducted an error analysis for this scheme. Zhao et al. [31] utilized a constraint gradient flow form to reformulate the CHNS system. Their method decoupled the variables and achieved second-order accuracy while preserving the discrete energy laws. Shen et al. [32] proposed the scalar auxiliary variable (SAV) approach. It is built upon invariant energy quadratization approaches and is computationally efficient, as only decoupled equations with constant parameters need to be solved at each time step.

In this study, we modify the NSCH equations by adding a data assimilation term. Thus, the observed data can be utilized during solving the equations. Next, we propose an efficient numerical scheme for the modified NSCH equations based on the Crank-Nicolson formula. To ensure the unconditional energy stability of the proposed scheme, we adopt the SAV approach and provide a proof of this property. Additionally, we investigate the influence of various superparameters and demonstrate the efficiency of our scheme.

The organization of this paper is as follows. In Section 2, we describe the modified NSCH equations with a data assimilation term and prove that the modified model also preserves the mass conservation and energy stability. In Section 3, we construct a discretized scheme with Crank-Nicolson formula and SAV approach and prove the proposed scheme is energy stability. In Section 4, provides various numerical results to demonstrate the performance of our scheme. The conclusions are given in Section 5.

2. NSCH equation with data assimilation term

In this section, we introduce the modified NSCH equation with a data assimilation term. Let $\Omega \in \mathbb{R}^2$ be a smooth domain and \mathbf{x} be the coordinate within this domain. The entire space is occupied by two immiscible and incompressible materials. The distribution of these materials is represented by a phase-field function $\phi(\mathbf{x}, t)$. We set $\phi = 1$ to represent one material and $\phi = -1$ to represent the other material. To ensure continuity of the phase field function, a thin interfacial layer is added between the two materials. The thickness of this interfacial layer is determined by the parameter ϵ . We use $\mathbf{u}(\mathbf{x}, t)$, $\mathbf{p}(\mathbf{x}, t)$, and $\rho(\mathbf{x}, t)$ to represent the velocity, pressure, and density fields respectively. The original NSCH equations are given as:

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\nabla p + \frac{1}{Re} \nabla \cdot (\nabla \mathbf{u}) - \frac{1}{We} \phi \nabla \mu, \quad (1a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (1b)$$

$$\phi_t = \frac{1}{Pe} \Delta \mu - \nabla \cdot (\mathbf{u} \phi), \quad (1c)$$

$$\mu = -\epsilon^2 \Delta \phi + F'(\phi), \quad (1d)$$

where $F(\phi) = 0.25(\phi^2 - 1)^2$ is the nonlinear potential, Re is the Reynolds number, Pe is the Peclet number, and We is the Weber number. To incorporate observed data into the model, we follow the approach of [16] and consider situations where the velocity and material distributions can be measured. To make the proposed scheme preserve the energy dissipation law, data assimilation terms related to the observed quantities $\bar{\mathbf{u}}(\mathbf{x}, t)$ and $\bar{\phi}$ are added to Eq. (1a) and Eq. (1d), respectively. As a result, the modified NSCH equations can be expressed as follows:

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\nabla p + \frac{1}{Re} \nabla \cdot (\nabla \mathbf{u}) - \frac{1}{We} \phi \nabla \mu - \lambda_1 I(\mathbf{x}) \left(\frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\bar{\mathbf{u}}}{|\bar{\mathbf{u}}|} \right), \quad (2a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (2b)$$

$$\phi_t = \frac{1}{Pe} \Delta \mu - \nabla \cdot (\mathbf{u} \phi), \quad (2c)$$

$$\mu = -\epsilon^2 \Delta \phi + F'(\phi) + \lambda_2 I(\mathbf{x}) (\phi - \bar{\phi}), \quad (2d)$$

where $|\cdot|$ represents modulus of a vector. λ_1 and λ_2 are nudging parameters and determine the influence of the observed data. $\bar{\mathbf{u}}$ and $\bar{\phi}$ represent the observed value of velocity \mathbf{u} and phase-field variable ϕ , respectively. Note that the measurement values may not be obtained for every point in domain, we assume the missing values are substituted by zeros and introduce a map $I(\mathbf{x})$ as:

$$I(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \Omega_b \\ 0, & \mathbf{x} \notin \Omega_b \end{cases} \quad (3)$$

where the $\Omega_b \subset \Omega$ represents the locations of the observed values. This map restricts the effect of data assimilation terms on the points with measurement data. According to the definition, we can obtain $I(\mathbf{x}) = \sqrt{I(\mathbf{x})}$. Based on the classical self-consistent mean field theory [33], a Ginzburg–Landau type of Helmholtz free energy function can be given as the following form:

$$E(\phi, \mathbf{u}) = \int_{\Omega} \left(\frac{\epsilon^2}{2} |\nabla \phi(\mathbf{x}, t)|^2 + F(\phi(\mathbf{x}, t)) + \frac{\lambda_2}{2} |\sqrt{I(\mathbf{x})}(\phi(\mathbf{x}, t) - \bar{\phi}(\mathbf{x}, t))|^2 + \frac{We}{2} |\mathbf{u}(\mathbf{x}, t)|^2 \right) d\mathbf{x}. \quad (4)$$

Here, the periodic boundary condition is used. We introduce some notations and then prove the mass conservation property and energy dissipation law in continuous situation. Let $f(\mathbf{x}), g(\mathbf{x}) \in L^2(\Omega)$, and their inner product is denoted as $(f(\mathbf{x}), g(\mathbf{x})) := \int_{\Omega} f(\mathbf{x})g(\mathbf{x})d\mathbf{x}$ and the L_2 -norm of f is defined as $\|f\| = \sqrt{(f, f)}$.

Theorem 2.1. *The modified NSCH equations Eq. (2) preserve the mass conservation property.*

Proof. By performing inner product of Eq. (2c) with the unity element $\mathbf{1}$, we obtain:

$$\begin{aligned} \frac{\partial}{\partial t} \int_{\Omega} \phi d\mathbf{x} &= \int_{\Omega} \phi_t d\mathbf{x} = \int_{\Omega} \frac{1}{Pe} \Delta \mu d\mathbf{x} - \int_{\Omega} \nabla \cdot (\mathbf{u}\phi) d\mathbf{x} \\ &= \frac{1}{Pe} \int_{\partial\Omega} \nabla \mu \cdot \mathbf{n} d\mathbf{x} - \int_{\partial\Omega} \mathbf{u}\phi d\mathbf{x} = 0. \end{aligned} \quad (5)$$

The proof is completed. \square

Theorem 2.2. *The modified NSCH equations Eq. (2) preserve the energy dissipation law.*

Proof. By performing inner product of Eq. (2c) with $\frac{1}{We} \mu$, we obtain

$$-\frac{1}{We} (\phi_t, \mu) = \frac{1}{We \cdot Pe} \|\nabla \mu\|^2 + \frac{1}{We} (\nabla \cdot (\mathbf{u}\phi), \mu). \quad (6)$$

By performing inner product of Eq. (2d) with $\frac{1}{We} \phi_t$, we obtain

$$\frac{1}{We} (\mu, \phi_t) = \frac{d}{dt} \left(\frac{\epsilon^2}{2We} \|\nabla \phi\|^2 + \frac{1}{We} (F(\phi), \mathbf{1}) + \frac{\lambda_2}{2We} \|\sqrt{I(\mathbf{x})}(\phi - \bar{\phi})\|^2 \right). \quad (7)$$

By performing inner product of Eq. (2a) with \mathbf{u} and considering the Eq. (2b), we obtain

$$\frac{d}{dt} \int_{\Omega} \frac{1}{2} |\mathbf{u}|^2 d\mathbf{x} + \frac{1}{Re} \|\nabla \mathbf{u}\|^2 + \frac{1}{We} (\phi \nabla \mu, \mathbf{u}) + ((\mathbf{u} \cdot \nabla) \mathbf{u}, \mathbf{u}) + (\nabla p, \mathbf{u}) + \frac{\lambda_1}{2} \|\sqrt{I(\mathbf{x})}(\frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\bar{\mathbf{u}}}{|\bar{\mathbf{u}}|})\|^2 = 0. \quad (8)$$

Due to the boundary conditions and Eq. (2b), the terms $(\mathbf{u} \cdot \nabla) \mathbf{u} \cdot \mathbf{u}$ and $\nabla p \cdot \mathbf{u}$ are equal to zero. Thus, combining the Eqs. (6)-(8), we obtain:

$$\begin{aligned} \frac{1}{We} \frac{dE(\phi, \mathbf{u})}{dt} &= \frac{1}{We} \frac{d}{dt} \int_{\Omega} \frac{\epsilon^2}{2} \|\nabla \phi\|^2 + F(\phi) + \frac{\lambda_2}{2} \|\sqrt{I(\mathbf{x})}(\phi - \bar{\phi})\|^2 + \frac{We}{2} |\mathbf{u}|^2 d\mathbf{x} \\ &= -\frac{1}{We \cdot Pe} \|\nabla \mu\|^2 - \frac{1}{Re} \|\nabla \mathbf{u}\|^2 - \frac{\lambda_1}{2} \|\sqrt{I(\mathbf{x})}(\frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\bar{\mathbf{u}}}{|\bar{\mathbf{u}}|})\|^2 \leq 0. \end{aligned} \quad (9)$$

The proof is completed. \square

To preserve the energy dissipation law during the discretized evolving process, the scalar auxiliary variable methods are applied to the modified equation system:

$$\begin{aligned} \frac{\partial \mathbf{u}(\mathbf{x}, t)}{\partial t} + Q(\mathbf{u} \cdot \nabla) \mathbf{u} &= -\nabla p + \frac{1}{Re} \nabla \cdot (\nabla \mathbf{u}) - \frac{Q}{We} \phi \nabla \mu - \lambda_1 I(\mathbf{x}) \left(\frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\bar{\mathbf{u}}}{|\bar{\mathbf{u}}|} \right), \\ \nabla \cdot \mathbf{u} &= 0, \\ \phi_t &= \frac{1}{Pe} \Delta \mu - Q \nabla \cdot (\mathbf{u}\phi), \\ \mu &= -\epsilon^2 \Delta \phi + \frac{RF'(\phi)}{\sqrt{\int_{\Omega} F(\phi) d\mathbf{x}}} + \lambda_2 I(\mathbf{x}) (\phi - \bar{\phi}), \\ \frac{dR}{dt} &= \frac{1}{2} \int_{\Omega} \frac{F'(\phi)}{\sqrt{\int_{\Omega} F(\phi) d\mathbf{x}}} \phi_t d\mathbf{x}, \\ \frac{dQ}{dt} &= \frac{1}{2} \int_{\Omega} (\nabla \cdot (\mathbf{u}\phi) \mu + \frac{1}{We} \phi \nabla \mu \cdot \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} \cdot \mathbf{u}) d\mathbf{x}, \end{aligned} \quad (10)$$

where R and Q are two time dependent auxiliary variables.

3. Numerical scheme and energy dissipation law

In this section, we consider the discretized situation. We first introduce some notations and utilize the Crank–Nicolson formula to discretize the above equation system in time. Then we demonstrate the energy dissipation property of the proposed numerical scheme. Let $\Omega = (a, b) \times (c, d)$ be the computational domain. The number of mesh cells along the x -axis and y -axis are denoted as $N_x, N_y \in \mathbb{Z}^+$. The uniform mesh size is defined as $h = (b-a)/N_x = (d-c)/N_y$. Then, the set of mesh cell centers is represented as $\Omega_d = \{(x_i, y_j) : x_i = (i-0.5)h, y_j = (j-0.5)h, 1 \leq i \leq N_x, 1 \leq j \leq N_y\}$. Let $f, g \in l^2(\Omega_d)$ and $f_{i,j}^k$ represent the approximation of $f(x_i, y_j, k\Delta t)$. The discretized inner product and form are defined as:

$$(f, g)_d \triangleq \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} f_{i,j} g_{i,j} \quad \text{and} \quad \|f\|_d \triangleq \sqrt{(f, f)_d}. \quad (11)$$

We apply the Crank–Nicolson discretization to Eqs. (10) and obtain the scheme as:

$$\frac{\phi^{k+1} - \phi^k}{\Delta t} = \frac{1}{Pe} \Delta_d \mu^{k+\frac{1}{2}} - Q^{k+\frac{1}{2}} \nabla_d \cdot (\tilde{\mathbf{u}}^{k+\frac{1}{2}} \tilde{\phi}^{k+\frac{1}{2}}), \quad (12a)$$

$$\mu^{k+\frac{1}{2}} = -\frac{\epsilon^2}{2} \Delta_d (\phi^{k+1} + \phi^k) + \frac{R^{k+\frac{1}{2}} F'(\tilde{\phi}^{k+\frac{1}{2}})}{\sqrt{(F(\tilde{\phi}^{k+\frac{1}{2}}), \mathbf{1})_d}} + \lambda_2 \mathcal{I}_d(\mathbf{x})(\phi^{k+\frac{1}{2}} - \tilde{\phi}), \quad (12b)$$

$$\frac{\hat{R}^{k+1} - R^k}{\Delta t} = \frac{1}{2} \left(\frac{F'(\tilde{\phi}^{k+\frac{1}{2}})}{\sqrt{(F(\tilde{\phi}^{k+\frac{1}{2}}), \mathbf{1})_d}} \frac{\phi^{k+1} - \phi^k}{\Delta t}, \mathbf{1} \right)_d, \quad (12c)$$

$$\begin{aligned} \frac{\mathbf{u}^* - \mathbf{u}^k}{\Delta t} = & -\nabla p^k - Q^{k+\frac{1}{2}} (\tilde{\mathbf{u}}^{k+\frac{1}{2}} \cdot \nabla_d) \tilde{\mathbf{u}}^{k+\frac{1}{2}} + \frac{1}{Re} \nabla_d \cdot (\nabla_d \frac{\mathbf{u}^* + \mathbf{u}^k}{2}) - \frac{Q^{k+\frac{1}{2}}}{We} \phi^{k+\frac{1}{2}} \nabla_d \mu^{k+\frac{1}{2}} \\ & - \lambda_1 \mathcal{I}_d(\mathbf{x}) \left(\frac{\mathbf{u}^* + \mathbf{u}^k}{|\mathbf{u}^* + \mathbf{u}^k|} - \frac{\tilde{\mathbf{u}}}{|\tilde{\mathbf{u}}|} \right), \end{aligned} \quad (12d)$$

$$\frac{\mathbf{u}^{k+1} - \mathbf{u}^*}{\Delta t} = -\frac{\nabla_d p^{k+1} - \nabla_d p^k}{2}, \quad (12e)$$

$$\frac{\hat{Q}^{k+1} - Q^k}{\Delta t} = \left(\nabla_d \cdot (\tilde{\mathbf{u}}^{k+1} \tilde{\phi}^{k+\frac{1}{2}}) \mu^{n+\frac{1}{2}} + \frac{1}{We} \phi^{k+\frac{1}{2}} \nabla_d \mu^{k+\frac{1}{2}} \cdot (\mathbf{u}^* + \mathbf{u}^k) \right. \quad (12f)$$

$$\left. + \tilde{\mathbf{u}}^{k+\frac{1}{2}} \cdot \nabla_d \tilde{\mathbf{u}}^{k+\frac{1}{2}} \cdot (\mathbf{u}^* + \mathbf{u}^k), \mathbf{1} \right)_d, \quad (12g)$$

where $(\cdot)^{k+\frac{1}{2}} = \frac{1}{2}((\cdot)^{k+1} + (\cdot)^k)$ and $(\tilde{\cdot})^{k+\frac{1}{2}} = 3/2(\cdot)^k - 1/2(\cdot)^{k-1}$. $\mathcal{I}_d(\mathbf{x})$ represents the discretization of map $\mathcal{I}(\mathbf{x})$ and is a matrix composing of zero and one. $\mathbf{1}$ represents a constant function which is equal to one identically. Note that in continuous situation, $R = \sqrt{\int_{\Omega} F(\phi) dx}$ and $Q = 1$. The scheme with the scalar auxiliary variable method should be equal to the original one. However, after time discretization, the calculated \hat{R}^{k+1} will be not equal to the discretized $\sqrt{\int_{\Omega} F(\phi^{k+1}) dx}$ and \hat{Q}^{k+1} will deviate away from 1. In order to make a correction, we utilize the following relaxation step inspired by the idea in [34]. With computed \hat{R}^{k+1} , R^k , ϕ^{k+1} , \hat{Q}^{k+1} , Q^k and $\mu^{k+\frac{1}{2}}$, we suppose the corrected value R^{k+1} is obtained by:

$$R^{k+1} = \eta_1 \hat{R}^{k+1} + (1 - \eta_1) R(\phi^{k+1}) \quad \text{and} \quad R(\phi^{k+1}) = \sqrt{(F(\phi^{k+1}), \mathbf{1})_d}, \quad (13)$$

where η_1 is the solution of following optimization problem:

$$\eta_1 = \min_{\eta \in [0,1]} |R^{k+1}|^2 - |\hat{R}^{k+1}|^2 \leq \frac{\Delta t}{2Pe} \|\nabla \mu^{k+\frac{1}{2}}\|_d^2. \quad (14)$$

The above inequality can be simplified as:

$$|(\hat{R}^{k+1} - R(\phi^{k+1}))\eta_1 + R(\phi^{k+1})|^2 \leq |\hat{R}^{k+1}|^2 + \frac{\Delta t}{2Pe} \|\nabla \mu^{k+\frac{1}{2}}\|_d^2. \quad (15)$$

The feasible set V_R of η in inequality (15) is not empty. Thus, the update of R^{k+1} is

$$R^{k+1} = \max \left\{ 0, \frac{-R(\phi^{k+1}) - \sqrt{|\hat{R}^{k+1}|^2 + \frac{\Delta t}{2Pe} \|\nabla \mu^{k+\frac{1}{2}}\|_d^2}}{\hat{R}^{k+1} - R(\phi^{k+1})} \right\}. \quad (16)$$

Similarly, the corrected value Q^{k+1} is obtained by:

$$Q^{k+1} = \eta_2 \hat{Q}^{k+1} + (1 - \eta_2), \quad (17)$$

where η_2 is the solution of the following problem:

$$\eta_2 = \min_{\eta \in [0,1]} |Q^{k+1}|^2 - |\hat{Q}^{k+1}|^2 \leq \frac{\Delta t}{2Pe} \|\nabla \mu^{k+\frac{1}{2}}\|_d^2. \quad (18)$$

The above inequality can be simplified as:

$$|(\hat{Q}^{k+1} - 1)\eta_2 + 1|^2 \leq |\hat{Q}^{k+1}|^2 + \frac{\Delta t}{2Pe} \|\nabla \mu^{k+\frac{1}{2}}\|_d^2. \quad (19)$$

The feasible set V_Q of η in inequality (19) is not empty. Thus, the update of Q^{k+1} is defined as

$$Q^{k+1} = \max \left\{ 0, \frac{-1 - \sqrt{|\hat{Q}^{k+1}|^2 + \frac{\Delta t}{2Pe} \|\nabla \mu^{k+\frac{1}{2}}\|_d^2}}{\hat{Q}^{k+1} - 1} \right\}. \quad (20)$$

Then, we define the modified discretized energy as:

$$\begin{aligned} E_d(\phi^k, \mathbf{u}^k) = & \frac{\epsilon^2}{2} \|\nabla_d \phi^k\|_d^2 + |R^k|^2 + \frac{\lambda_2}{2} \|\sqrt{\mathcal{I}_d(\mathbf{x})} \phi^k\|_d^2 - \frac{\lambda_2}{2} (\mathcal{I}_d(\mathbf{x}) \phi^k, 2\tilde{\phi})_d \\ & + \|\mathbf{u}^k\|_d^2 + \frac{\Delta t^2}{4} \|\nabla_d p^k\|_d^2 + |Q^k|^2. \end{aligned} \quad (21)$$

Based on the modified discretized energy, we now prove the unconditional stability of the proposed scheme Eq. (12).

Remark 3.1. In literature, lots of works have been done to justify whether the proposed numerical scheme is uniquely solvable. Zhang et al. [35] proved the unconditional unique solvability for the Functionalized Cahn-Hilliard equation utilizing a stabilized scalar auxiliary variable method. Wang et al. [36] theoretically substantiated the unique solvability for the square phase field crystal equation with a scalar auxiliary variable method. Shen et al. [32] demonstrated the unique solvability for a Cahn-Hilliard equation employing a similar method. Further, Yang et al. [37] examined the uniquely solvable property of the Allen-Cahn-Navier-Stokes system. In the context of the Cahn-Hilliard-Navier-Stokes system, Chen et al. [38] proposed a numerical scheme using the invariant energy quadratization approach, proving its unique solvability for each time step. In this article, the proposed numerical scheme Eqs. (12) based on a modified Cahn-Hilliard-Navier-Stokes system incorporating only a regularization term to the original system. Consequently, the proof of unique solvability for the proposed scheme parallels the methodologies articulated in the aforementioned studies. For readers interested in a more comprehensive understanding of this topic, additional resources including [39–42] can provide more detail.

Theorem 3.1. The proposed scheme (12a)-(12g) is unconditional energy stable, i.e. $E_d(\phi^{k+1}, \mathbf{u}^{k+1}) \leq E_d(\phi^k, \mathbf{u}^k)$

Proof. By performing inner product of Eq. (12a) with $\mu^{k+\frac{1}{2}}$, we obtain:

$$(\phi^{k+1} - \phi^k, \mu^{k+\frac{1}{2}})_d = -\frac{\Delta t}{Pe} \|\nabla_d \mu^{k+\frac{1}{2}}\|_d^2 - Q^{k+\frac{1}{2}} \Delta t (\nabla_d \cdot \tilde{\mathbf{u}}^{k+\frac{1}{2}} \tilde{\phi}^{k+\frac{1}{2}}, \mu^{k+\frac{1}{2}})_d. \quad (22)$$

By performing inner product of Eq. (12b) with $\phi^{k+1} - \phi^k$, we obtain:

$$\begin{aligned} (\mu^{k+\frac{1}{2}}, \phi^{k+1} - \phi^k)_d &= \frac{\epsilon^2}{2} (\|\nabla_d \phi^{k+1}\|_d^2 - \|\nabla_d \phi^k\|_d^2) + \left(\frac{R^{k+\frac{1}{2}} F'(\tilde{\phi}^{k+\frac{1}{2}})}{(F(\tilde{\phi}^{k+\frac{1}{2}}), \mathbf{1})_d}, \phi^{k+1} - \phi^k \right)_d \\ &\quad + \frac{\lambda_2}{2} (\|\sqrt{I_d(\mathbf{x})} \phi^{k+1}\|_d^2 - \|\sqrt{I_d(\mathbf{x})} \phi^k\|_d^2 - (\phi^{k+1} - \phi^k, 2I_d(\mathbf{x})\tilde{\phi})_d). \end{aligned} \quad (23)$$

By multiplying Eq. (12c) with $\hat{R}^{k+1} + R^k$, we obtain:

$$(\hat{R}^{k+1} - R^k, \hat{R}^{k+1} + R^k) = |\hat{R}^{k+1}|^2 - |R^k|^2 = \left(\frac{R^{k+\frac{1}{2}} F'(\tilde{\phi}^{k+\frac{1}{2}})}{(F(\tilde{\phi}^{k+\frac{1}{2}}), \mathbf{1})_d}, \phi^{k+1} - \phi^k \right)_d. \quad (24)$$

Combining Eqs. (22)-(24), we obtain:

$$\begin{aligned} &\frac{\epsilon^2}{2} (\|\nabla_d \phi^{k+1}\|_d^2 - \|\nabla_d \phi^k\|_d^2) + |\hat{R}^{k+1}|^2 - |R^k|^2 + \frac{\lambda_2}{2} (\|\sqrt{I_d(\mathbf{x})} \phi^{k+1}\|_d^2 - \|\sqrt{I_d(\mathbf{x})} \phi^k\|_d^2 - (\phi^{k+1} - \phi^k, 2I_d(\mathbf{x})\tilde{\phi})_d) \\ &= -\frac{\Delta t}{Pe} \|\nabla_d \mu^{k+\frac{1}{2}}\|_d^2 - Q^{k+\frac{1}{2}} \Delta t (\nabla_d \cdot \tilde{\mathbf{u}}^{k+\frac{1}{2}} \tilde{\phi}^{k+\frac{1}{2}}, \mu^{k+\frac{1}{2}})_d. \end{aligned} \quad (25)$$

By performing inner product of Eq. (12d) with $\Delta t(\mathbf{u}^* + \mathbf{u}^k)$, we obtain:

$$\begin{aligned} \|\mathbf{u}^*\|_d^2 - \|\mathbf{u}^k\|_d^2 &= -\Delta t (\nabla_d p^k, \mathbf{u}^*)_d - \Delta t Q^{k+\frac{1}{2}} ((\tilde{\mathbf{u}}^{k+\frac{1}{2}} \cdot \nabla_d) \tilde{\mathbf{u}}^{k+\frac{1}{2}}, \mathbf{u}^* + \mathbf{u}^k)_d - \frac{\Delta t}{Re} \|\nabla_d \frac{\mathbf{u}^* + \mathbf{u}^k}{2}\|_d^2 \\ &\quad - \frac{Q^{k+\frac{1}{2}} \Delta t}{We} (\phi^{k+\frac{1}{2}} \nabla_d \mu^{k+\frac{1}{2}}, \mathbf{u}^* + \mathbf{u}^k)_d - \frac{1}{2} \lambda_1 \Delta t \|\mathbf{u}^* + \mathbf{u}^k\|_d \|\sqrt{I_d(\mathbf{x})} (\frac{\mathbf{u}^* + \mathbf{u}^k}{|\mathbf{u}^* + \mathbf{u}^k|} - \frac{\tilde{\mathbf{u}}}{|\tilde{\mathbf{u}}|})\|_d^2. \end{aligned} \quad (26)$$

For Eq. (12e), we have:

$$\frac{2}{\Delta t} \mathbf{u}^{k+1} + \nabla_d p^{k+1} = \frac{2}{\Delta t} \mathbf{u}^* + \nabla_d p^k. \quad (27)$$

By squaring the above equality and performing the integral, we have:

$$(\mathbf{u}^*, \nabla_d p^k)_d = \frac{1}{\Delta t} (\|\mathbf{u}^{k+1}\|_d^2 - \|\mathbf{u}^*\|_d^2) + \frac{\Delta t}{4} (\|\nabla_d p^{k+1}\|_d^2 - \|\nabla_d p^k\|_d^2). \quad (28)$$

By multiplying Eq. (12g) with $Q^{k+\frac{1}{2}}$, we obtain:

$$\begin{aligned} (|\hat{Q}^{k+1}|^2 - |Q^k|^2) &= \Delta t \left(Q^{k+\frac{1}{2}} (\tilde{\mathbf{u}}^{k+\frac{1}{2}} \cdot \nabla_d \tilde{\mathbf{u}}^{k+\frac{1}{2}}, \mathbf{u}^* + \mathbf{u}^k)_d + \frac{Q^{k+\frac{1}{2}}}{We} (\phi^{k+\frac{1}{2}} \nabla_d \mu^{k+\frac{1}{2}}, \mathbf{u}^* + \mathbf{u}^k)_d \right. \\ &\quad \left. + Q^{k+\frac{1}{2}} (\nabla_d \cdot \tilde{\mathbf{u}}^{k+\frac{1}{2}} \tilde{\phi}^{k+\frac{1}{2}}, \mu^{k+\frac{1}{2}})_d \right). \end{aligned} \quad (29)$$

Consider the conditions Eqs. (14) and (18) and combining Eqs. (25) and (26)-(29), we have:

$$\begin{aligned} &\|\mathbf{u}^{k+1}\|_d^2 - \|\mathbf{u}^k\|_d^2 + \frac{\Delta t^2}{4} (\|\nabla_d p^{k+1}\|_d^2 - \|\nabla_d p^k\|_d^2) + |Q^{k+1}|^2 - |Q^k|^2 + \frac{\epsilon^2}{2} (\|\nabla \phi^{k+1}\|_d^2 - \|\nabla \phi^k\|_d^2) \\ &\quad + |R^{k+1}|^2 - |R^k|^2 + \frac{\lambda_2}{2} (\|\sqrt{I_d(\mathbf{x})} \phi^{k+1}\|_d^2 - \|\sqrt{I_d(\mathbf{x})} \phi^k\|_d^2 - (\phi^{k+1} - \phi^k, 2I_d(\mathbf{x})\tilde{\phi})_d) \\ &= -\frac{\Delta t}{Pe} \|\nabla_d \mu^{k+\frac{1}{2}}\|_d^2 + |Q^{k+1}|^2 - |\hat{Q}^{k+1}|^2 + |R^{k+1}|^2 - |\hat{R}^{k+1}|^2 - \frac{\Delta t}{Re} \|\nabla_d \frac{\mathbf{u}^* + \mathbf{u}^k}{2}\|_d^2 \\ &\quad - \frac{1}{2} \lambda_1 Q^{k+\frac{1}{2}} \Delta t \|\mathbf{u}^* + \mathbf{u}^k\|_d \|\sqrt{I_d(\mathbf{x})} (\frac{\mathbf{u}^* + \mathbf{u}^k}{|\mathbf{u}^* + \mathbf{u}^k|} - \frac{\tilde{\mathbf{u}}}{|\tilde{\mathbf{u}}|})\|_d^2 \\ &\leq 0. \quad \square \end{aligned} \quad (30)$$

Thus, the proposed scheme is energy stability. Note that phase-field variable ϕ^{k+1} , intermediate velocity field \mathbf{u}^{k+1} , and auxiliary variables \mathbf{R}^{k+1} and \mathbf{Q}^{k+1} are coupled. To obtain the solution, we employ the Picard iteration to achieve totally decoupled computation.

Remark 3.2. In literature, there have been quite a few existing works of convergence and error analysis for various Cahn-Hilliard-fluid models. Feng et al. [43] presented a convergence analysis for the Functionalized Cahn-Hilliard equation. Cheng et al. [44] introduced an energy-stable numerical scheme for the Cahn-Hilliard equation and provide an optimal rate convergence analysis, leveraging the global bound for the second-order functional derivatives. Similarly, Chen et al. [45] offered an optimal rate convergence analysis for the Cahn-Hilliard-Hele-Shaw equations. Additionally, convergence analysis and error estimates for scalar auxiliary variable numerical schemes have been extensively reported. Zhang [35] investigated the Cahn-Hilliard equation employing a stabilized scalar auxiliary variable method and establishes the optimal error estimates for the proposed numerical scheme. Additionally, Li et al. [46] conducted a rigorous error analysis for the Navier-Stokes equations utilizing a scalar auxiliary variable approach. The scheme proposed in this article adapts the Cahn-Hilliard-Navier-Stokes system, incorporating an additional data assimilation term into the equations. Therefore, the methodologies for establishing convergence analysis and error estimation parallel those delineated in the mentioned studies. For a more extensive exploration of this topic, readers are encouraged to consult additional resources such as [47–50].

4. Experiments

In this section, we conduct various experiments to test our scheme. As the proposed scheme requires observed data, we perform twin experiments [11] for each case. Each case involves two types of processes: the reference process, which assumes the real physical process and is obtained by solving the NSCH models without the observed data (i.e., $\lambda_1 = \lambda_2 = 0$), and the simulated process, where observed data is sampled from the reference process and used.

4.1. Performance demonstration

This section demonstrates the performance of the proposed scheme with the spinodal decomposition phenomenon. Spinodal decomposition occurs when small perturbations are imposed on initially homogeneous mixtures. This leads to an increase in the amplitude of the order parameter ϕ over time, resulting in the formation of distinct two-phase states. The computational domain is $\Omega = (0, 2\pi) \times (0, 2\pi)$ with a 256×256 mesh grid, and the initial conditions of the reference process are defined as:

$$\begin{aligned}\phi(x, y, 0) &= 0.3 \tanh(x - \pi) + 0.001 \text{rand}(x, y), \\ u(x, y, 0) &= -\sin^2(x) \sin(y), \\ v(x, y, 0) &= \sin(2x) \sin^2(y), \\ p(x, y, 0) &= 0.\end{aligned}\tag{31}$$

Other parameters are $\Delta t = 10^{-5}$, $Pe = We = Re = 1$, and $\epsilon = 0.05$. Then, the observed data $\bar{\phi}$ and $\bar{\mathbf{u}}$ are sampled from the reference process with a sample interval $200\Delta t$. The initial conditions of the simulated process are set as:

$$\begin{aligned}\phi(x, y, 0) &= 0.001 \text{rand}(x, y), \\ u(x, y, 0) &= v(x, y, 0) = 10^{-6} \text{rand}(x, y), \\ p(x, y, 0) &= 0,\end{aligned}\tag{32}$$

where function $\text{rand}(x, y)$ generates a random number sampled uniformly from the interval $[-1, 1]$. To avoid the division formula of $\frac{\mathbf{u}}{|\mathbf{u}|}$ making no sense, we add small perturbations on the initial values of \mathbf{u} .

Fig. 1 illustrates the results of reference process and the results of simulated processes with different λ_1 and λ_2 . The first and second rows show snapshots of variable ϕ at time 0.02 and 0.1, respectively. The third row displays the snapshots of variable \mathbf{u} at time 0.1. Column (a) presents the results of the situation without the observed data ($\lambda_1 = 0, \lambda_2 = 0$). Columns (b), (c), and (d) represent the results of the simulated process with the observed data $\bar{\mathbf{u}}$ ($\lambda_1 = 5000, \lambda_2 = 0$), $\bar{\phi}$ ($\lambda_1 = 0, \lambda_2 = 1$), and $\bar{\mathbf{u}}$ and $\bar{\phi}$ ($\lambda_1 = 5000, \lambda_2 = 1$), respectively. Note that the values of λ_1 and λ_2 are selected to ensure that impact of the data assimilation terms is proportional to that of the other terms. Column (e) corresponds to the results of the reference process. Comparing these results, we observe that the observed values of $\bar{\mathbf{u}}$ and $\bar{\phi}$ have different effects on the results. Specifically, the observed value of $\bar{\mathbf{u}}$ exhibits a slower influence on the state of ϕ compared to the observed value of $\bar{\phi}$. The third row shows the state of \mathbf{u} at time $t = 0.1$. The black arrows represent the direction and magnitude of u and the blue line represents the interface of two different materials. These results show that the observed values $\bar{\mathbf{u}}$ significantly influence the state of \mathbf{u} , compared to the observed value $\bar{\phi}$. Furthermore, upon comparing the results in columns (b), (c), and (d), it is evident that when both observed data $\bar{\phi}$ and $\bar{\mathbf{u}}$ are available, the simulated process accurately approximates the reference process. Consequently, in subsequent experiments, we assume that the data $\bar{\phi}$ and $\bar{\mathbf{u}}$ can be obtained.

4.2. Energy dissipation and mass conservation

In this section, we numerically prove the energy dissipation and mass conservation properties of the proposed scheme. We consider the situation where two droplets merge together in $\Omega_d = (0, 1) \times (0, 1)$ with a 256×256 mesh grid. The initial conditions of the reference process are set as follows:

$$\begin{aligned}\phi(x, y, 0) &= \max \left\{ \tanh\left(\frac{0.15 - \sqrt{(x - 0.65)^2 + (y - 0.5)^2}}{\sqrt{2}\epsilon}\right), \tanh\left(\frac{0.15 - \sqrt{(x - 0.35)^2 + (y - 0.5)^2}}{\sqrt{2}\epsilon}\right) \right\} \\ u(x, y, 0) &= v(x, y, 0) = 10^{-6} \text{rand}(x, y), \\ p(x, y, 0) &= 0.\end{aligned}\tag{33}$$

The initial conditions of the simulated process are

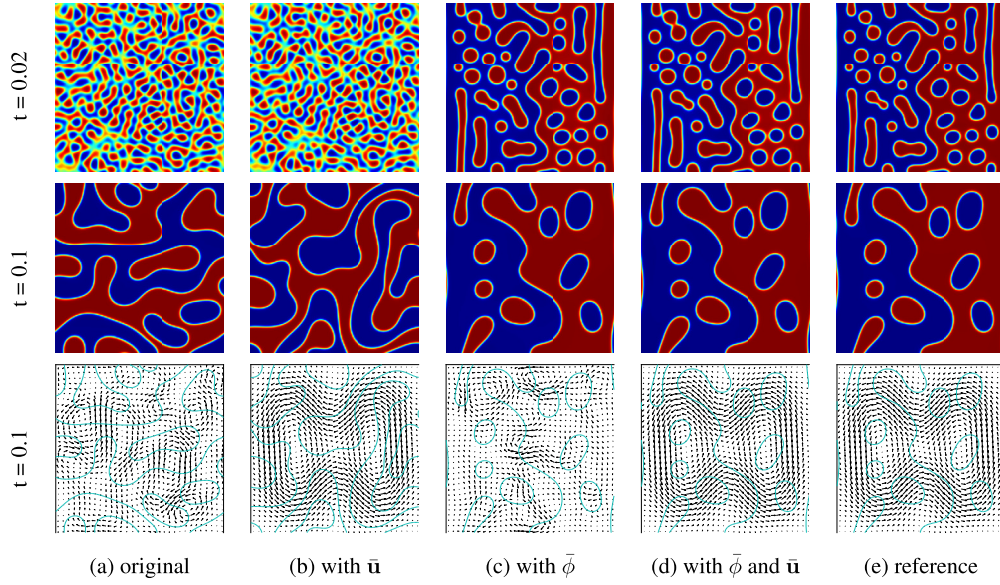


Fig. 1. Performance demonstration of the proposed scheme with spinodal decomposition test. Columns (a) and (e) are the results of the process with no observed data and the reference process. Columns (b)-(d) show the results with observed data \bar{u} , $\bar{\phi}$, and $\bar{\phi}$ and \bar{u} , respectively.

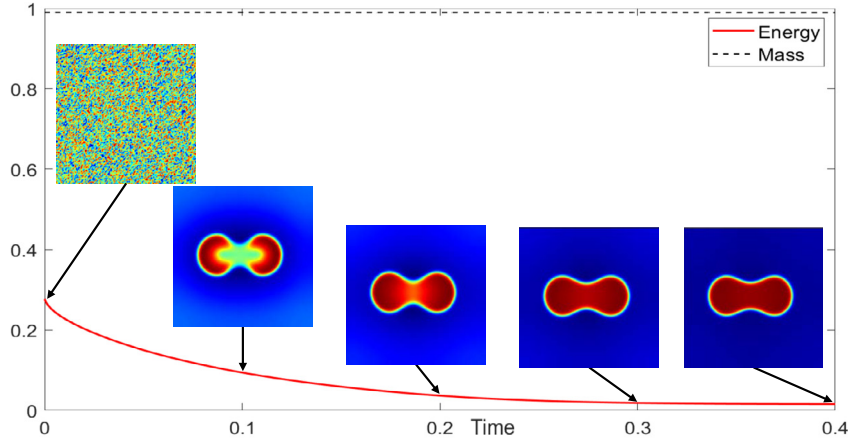


Fig. 2. The variation of mass and energy with time evolving. Red line represents the variation of energy and black dashed line represents the variation of mass variation.

$$\begin{aligned}
 \phi(x, y, 0) &= -0.734 + 10^{-3} \text{rand}(x, y) \\
 u(x, y, 0) &= v(x, y, 0) = 10^{-6} \text{rand}(x, y), \\
 p(x, y, 0) &= 0,
 \end{aligned} \tag{34}$$

where the initial value of ϕ ensure the simulated process has the same mass with the reference process. The mass of the system is measured using $M(t) = \frac{1}{N_d M_0} \sum_{i,j \in \Omega_d} \phi_{i,j}(t)$, where N_d is the number of mesh grid points, and $M_0 = \frac{1}{N_d} \sum_{i,j \in \Omega_d} \phi_{i,j}(0)$ is the average mass of the initial conditions. The energy is measured using the definition provided in Eq. (21). The variation processes of mass and discretized energy are shown in Fig. 2. The results demonstrate that the proposed scheme maintains mass conservation and energy stability. Additionally, five snapshots at different moments are presented in Fig. 2, indicating that as energy dissipates, the pattern of the state variable ϕ converges to a single drop.

4.3. Effects with different initial conditions

In this section, we investigate the situation where the initial conditions of the reference and simulated processes are different. We aim to explore whether the simulated process can converge to the reference process with the help of the observed values. As an example, we study the Ostwald ripening dynamics, which describes the thermodynamic phenomenon where smaller particles in a solution aggregate into larger particles, leading the system to a more stable state. The computational domain is $\Omega = (0, 1) \times (0, 1)$ with a mesh grid of size 256×256 . The initial values of velocity are set to zero, and the initial values of the state variable ϕ represent several drops with different radii. The initial conditions of the simulated process are set as the transpose of the reference process. The illustrations are shown in Fig. 3(a). Other parameters are set as $\Delta t = 10^{-4}$, $We = 10$, $Re = 50$, and $Pe = 1/\epsilon$. The data assimilation parameters are set as $\lambda_1 = 10000$ and $\lambda_2 = 1$. Fig. 3 (b)-(e) illustrates the results of these processes. The first and second rows show the snapshots of reference process and simulated process at time $t = 0.1, 0.3, 0.5, 1$, respectively. These results indicate that by incorporating the observed data, processes with different initial conditions can converge to the same results.

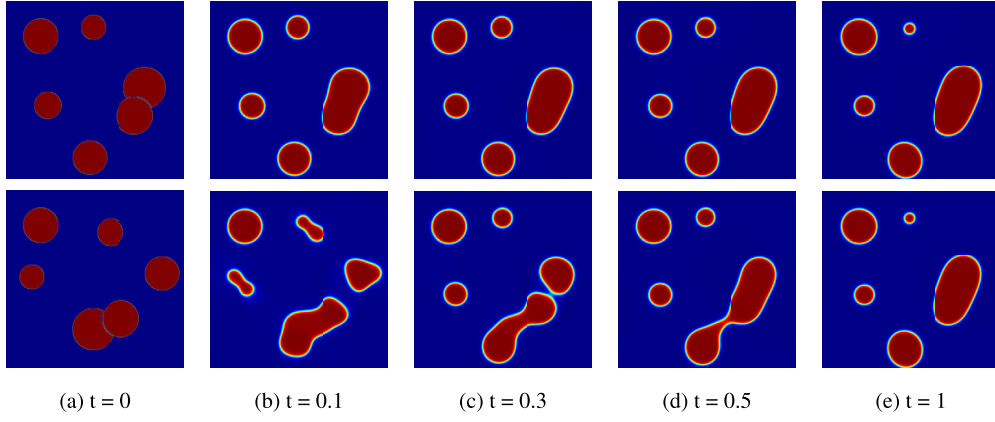


Fig. 3. Illustrations of density variable ϕ at time $t = 0.1, 0.3, 0.5, 1$. The first row is the reference process and the second row is the simulated process.

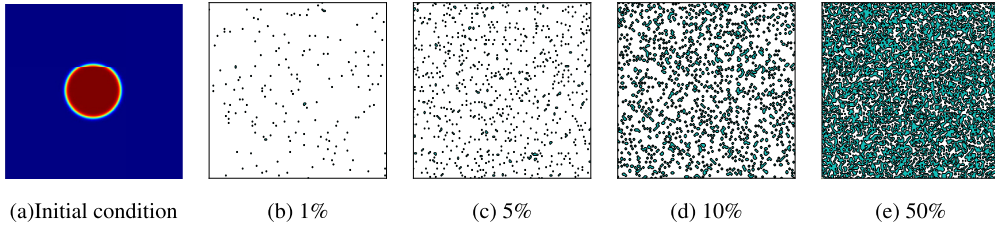


Fig. 4. Illustrations of initial conditions of the reference process and the simulated with different sample methods. (a) represents the initial condition of the reference process. (b)–(e) are illustrations of the observed location with sample rates 1%, 5%, 10% and 50%, respectively.

4.4. Effects of local observed data in space dimension

In certain real situations, obtaining global measurement data may be unfeasible due to the cost of devices. Therefore, in this subsection, we consider a scenario where only local data can be observed. We focus on the deformation of a drop under shear flow on the domain $\Omega = (0, 1) \times (0, 1)$ with a 256×256 mesh grid. The initial conditions of the reference process are defined as:

$$\begin{aligned} \phi(x, y, 0) &= -\tanh\left(\frac{\sqrt{(x-0.5)^2 + (y-0.5)^2} - 0.15}{\sqrt{2}\epsilon}\right) \\ u(x, y, 0) &= 10^{-6} \text{rand}(x, y), \quad v(x, y, 0) = 3 \sin(2\pi y) \\ p(x, y, 0) &= 0. \end{aligned} \quad (35)$$

Other parameters of the real process are set as $Re = 150$, $We = Pe = 1$, $\Delta t = 10^{-7}$, and $T = 0.0003$. To explore the effects of different types of the observed data, we adopt a randomly sampling method to obtain the observed data. Here, we introduce the sample rate r in spatial dimension, which is defined as the ratio of the observed data points to the total data points in the domain. Illustrations of sample results with sample rates 1%, 10% and 50% are shown in Fig. 4(b)–(e), where the void area represents the missing values and the blue area represents the observed values. We also use the local observed data as the initial conditions of the simulated process. For the missing values in the initial conditions, we use the bilinear interpolation method to make an approximation. Note that the observed values are sampled from the reference process at intervals of $20\Delta t$. This means that for every 20 time steps of the reference process, one solution is sampled as the observed values.

The simulated processes use observed data with sample rates of 1%, 5%, 10%, 50%, and 100%. Fig. 5 illustrates the state of density variable ϕ . It can be seen that the initial conditions have difference due to the different sample rates. However, these differences are eliminated as time evolves.

4.5. Effects of the sample interval in time dimension

In this section, we investigate the effect of time interval between two sample values. On the domain $\Omega = (0, 2\pi) \times (0, 2\pi)$ with a mesh grid 256×256 , the phenomenon of flow-coupled phase separation with different average concentrations is simulated in this case. The initial conditions of the reference process are defined to be

$$\begin{aligned} \phi(x, y, 0) &= \alpha \frac{|y - \pi|}{2\pi} + 0.001 \text{rand}(x, y) \\ u(x, y, 0) &= v(x, y, 0) = 10^{-6} \text{rand}(x, y), \\ p(x, y, 0) &= 0, \end{aligned} \quad (36)$$

where the parameter $\alpha = 1$ controls the value of average concentration. Other parameters are set as $Re = We = Pe = 1$ and $\Delta t = 10^{-6}$. As for the simulated processes, the initial conditions are sampled from the reference process with the 30% sample rate. A linear interpolation is also used to approximate the missing values. In the time dimension, we sample the data using different time intervals $N\Delta t$, meaning that for every N time steps, one solution of the reference process is sampled as the observed values. To incorporate the observed data into the simulation process, the data

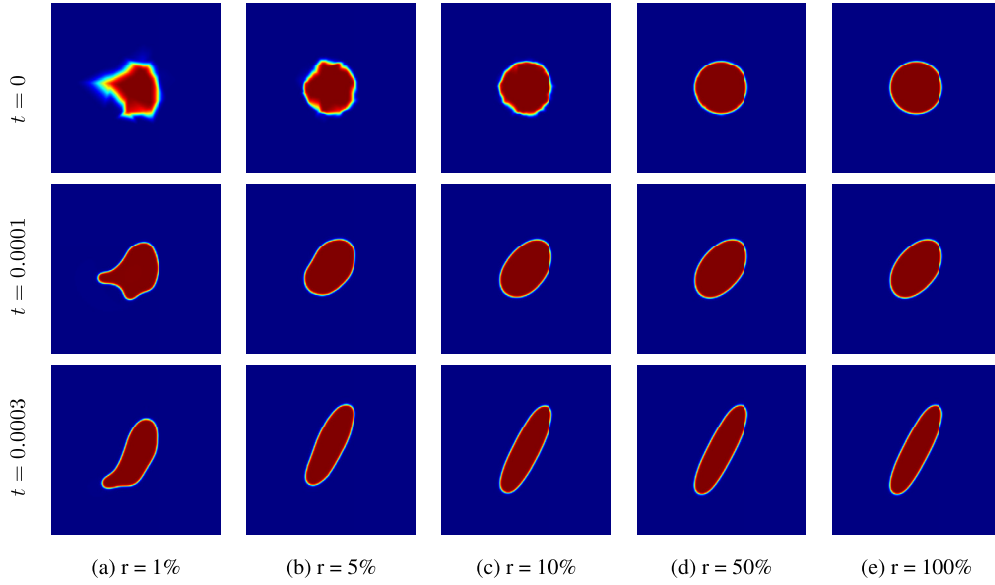


Fig. 5. Illustrations of the simulation results of density variable ϕ at time $t = 0, 0.0001, 0.0003$ with different and sample rates 1%, 5%, 10%, 50%, 100%.

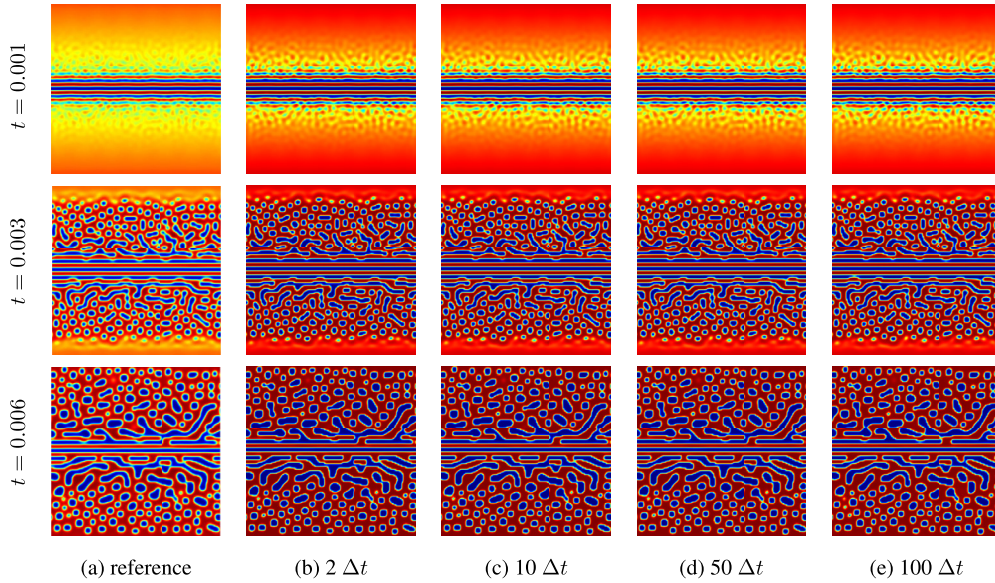


Fig. 6. Illustrations of the simulation results of density variable ϕ at time $t = 0.001, 0.003$, and 0.006 . Columns (b)-(e) represent the results with sample intervals $2\Delta t, 10\Delta t, 50\Delta t$, and $100\Delta t$, where Δt is the time step.

assimilation coefficients λ_1 and λ_2 are set as 6000 and 1, respectively. Fig. 6 shows the results with different time intervals 2, 10, 50, and 100 at time $T = 0.001, 0.003$, and 0.006 . To quantify these differences, we introduce measurements as:

$$D_\phi(\phi, \phi_r, t) = \frac{1}{N_p} \sum_{\mathbf{x} \in \Omega_d} |\phi(\mathbf{x}, t) - \phi_r(\mathbf{x}, t)|, \quad (37)$$

$$D_u(\mathbf{u}, \mathbf{u}_r, t) = \frac{1}{N_p} \sum_{\mathbf{x} \in \Omega_d} \|\mathbf{u}(\mathbf{x}, t) - \mathbf{u}_r(\mathbf{x}, t)\|_2, \quad (38)$$

where ϕ_r and \mathbf{u}_r are the variable of reference process. N_p is the number of mesh points. From these results, we can see that the smaller sample interval makes a better reconstruction of the reference process. Fig. 7 also indicates that too large sample interval may cause a disturbance for the simulated process, since the huge difference between the current state and the observed data. Meanwhile, the sample interval has more influence on the variable ϕ .

4.6. The observed data with noisy

In this section, we investigate the impact of different levels of noise in the observed data, which is common in real-world situations due to the errors in data acquisition equipment. For this purpose, we simulate the single vortex deformation of a circle on the square domain Ω_d using a 256×256 meshgrid. The initial conditions of the reference process are set as:

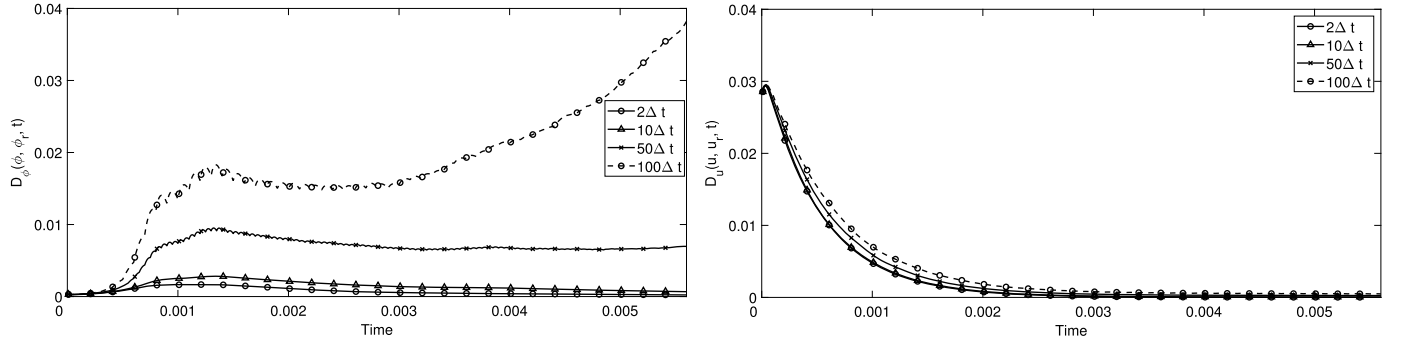


Fig. 7. Differences D_ϕ , D_u between the simulated process and reference process with different sample intervals $2\Delta t$, $10\Delta t$, $50\Delta t$, and $100\Delta t$.

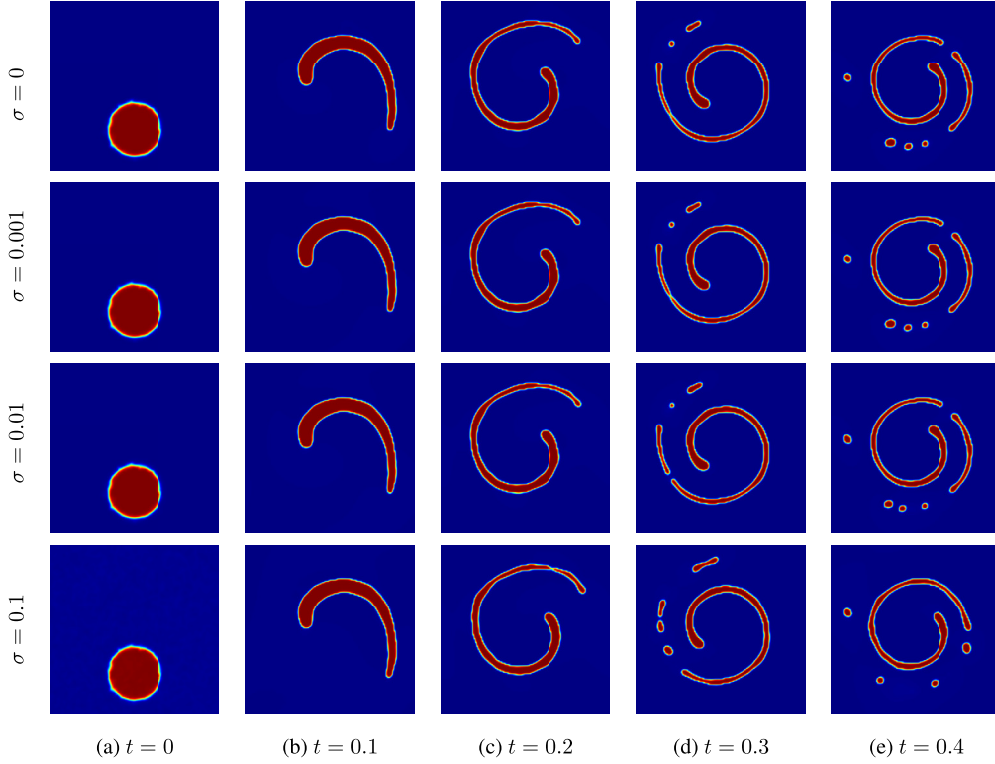


Fig. 8. Illustrations of the simulation results of density variable ϕ at time $t = 0, 0.1, 0.2, 0.3$, and 0.4 with different level noises σ .

$$\begin{aligned} \phi(x, y, 0) &= \tanh((0.15 - \sqrt{(x-0.5)^2 + (y-0.75)^2})/2\sqrt{2\epsilon}) \\ u(x, y, 0) &= -\sin^2(\pi x) \sin(2\pi y), \quad v(x, y, 0) = \sin(2\pi x) \sin^2(\pi y), \\ p(x, y, 0) &= 0. \end{aligned} \quad (39)$$

Other parameters are set as $Re = 50$, $We = 10$, $Pe = 100$, $\Delta t = 5 \times 10^{-5}$, and $\epsilon = 0.005$. Then, the observed data are sampled with the 20% sample rate and the $100\Delta t$ time interval from the reference process. To simulate the observed error of the data acquisition equipment, random noise is added to each observed data point. At $t = 0$, the observed data is used as the initial condition for the simulated process. Any missing values in the initial condition are interpolated using a linear interpolation method. The data assimilation parameters are set to $\lambda_1 = 1000$ and $\lambda_2 = 0.1$. The results of the simulated process with different levels of noise ($\sigma = 10^{-1}$, 10^{-2} , 10^{-3} , and 0) are shown in Fig. 8. Here, $\sigma = 10^{-1}$ means that we add a $10^{-1} \text{rand}(\mathbf{x})$ noise to the observed data. $\sigma = 0$ means that the observed data is the same as the reference process. The results demonstrate that the proposed method is capable of reconstructing the process with noisy observed data.

4.7. Dripping droplet

In this subsection, we investigate the performance of the proposed scheme through the evolution of a 2D dripping droplet phenomenon. In the presence of gravity, the droplet deforms and descends over time. For simplicity, we consider the case with a small density ratio, allowing us to adopt the Boussinesq approximation and recast the momentum equation as follows:

$$\rho^* \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) = -\nabla p + \frac{1}{Re} \Delta \mathbf{u} - \frac{1}{We} \phi \nabla \mu + (\rho(\phi) - \rho^*) \mathbf{g} - \lambda \left(\frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\bar{\mathbf{u}}}{|\bar{\mathbf{u}}|} \right), \quad (40)$$

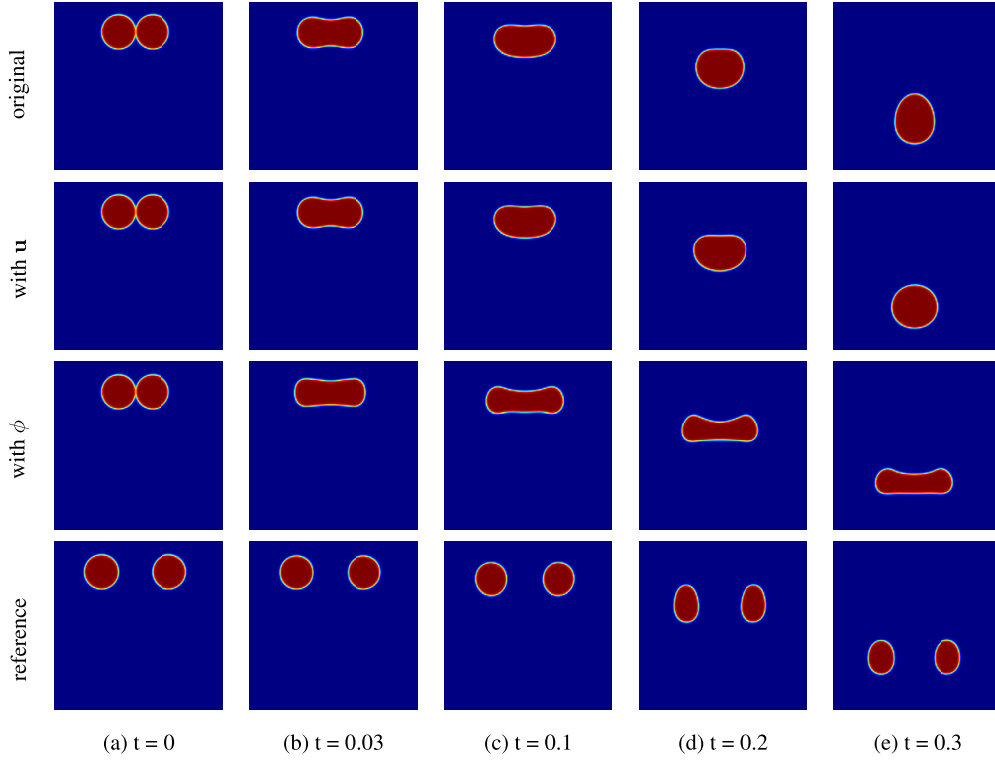


Fig. 9. Illustrations of the dripping droplets process at time $t = 0, 0.03, 0.1, 0.2$, and 0.3 , respectively.

where ρ^* is the background density and we let $\rho^* = \rho_2$. The density function is $\rho(\phi) = \rho_1(1 + \phi)/2 + \rho_2(1 - \phi)/2$. The gravitational acceleration is $\mathbf{g} = (0, -1)$. In heavy and light fluids, we set $\rho_1 = 1.5$ and $\rho_2 = 1$. The computational domain is $\Omega = (0, 1) \times (0, 1)$ with 256×256 meshgrid. The initial conditions of the simulated processes are defined to be

$$\begin{aligned} \phi(x, y, 0) &= \max\left\{\tanh\left(\frac{0.1 - \sqrt{(x-0.6)^2 + (y-0.2)^2}}{\sqrt{2}\epsilon}\right), \tanh\left(\frac{0.1 - \sqrt{(x-0.4)^2 + (y-0.2)^2}}{\sqrt{2}\epsilon}\right)\right\} \\ u(x, y, 0) &= v(x, y, 0) = 10^{-6}\text{rand}(x, y), \\ p(x, y, 0) &= 0. \end{aligned} \quad (41)$$

As shown in Fig. 9, with this initial conditions, two droplets will drop and blend together, if there is no observed data. Then, the observed data are sampled from the reference process with a time sample interval $200\Delta t$. The initial conditions of the reference process are:

$$\begin{aligned} \phi(x, y, 0) &= \max\left\{\tanh\left(\frac{0.1 - \sqrt{(x-0.7)^2 + (y-0.2)^2}}{\sqrt{2}\epsilon}\right), \tanh\left(\frac{0.1 - \sqrt{(x-0.3)^2 + (y-0.2)^2}}{\sqrt{2}\epsilon}\right)\right\} \\ u(x, y, 0) &= v(x, y, 0) = 10^{-6}\text{rand}(x, y), \\ p(x, y, 0) &= 0. \end{aligned} \quad (42)$$

Due to the larger initial interval between the droplets, they will drip without blending. Next, we simulate the processes with the data assimilation term. The second and third row of Fig. 9 represent the processes with the observed \mathbf{u} ($\lambda_1 = 5, \lambda_2 = 0$) and the observed ϕ ($\lambda_1 = 0, \lambda_2 = 0.5$), respectively. It can be observed that the assimilation term $\phi - \bar{\phi}$ changes the evolving direction of these two droplets, while the assimilation term $\frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\bar{\mathbf{u}}}{|\bar{\mathbf{u}}|}$ has a slower and smaller influence on the state evolving.

4.8. Simulation of the multi-component flow

In this section, the performance of the proposed scheme is evaluated within the context of multi-component flow dynamics. This investigation focuses on the equilibrium phase interface, specifically considering a prescribed contact angle. Within the multi-component flow environment, the governing equation assumes the following form:

$$\rho(\phi) \frac{\partial \mathbf{u}}{\partial t} = -\rho(\phi) \mathbf{u} \cdot \nabla \mathbf{u} - \nabla p + \frac{1}{Re} \nabla \cdot (\nabla \mathbf{u}) - \frac{1}{We} \mathbf{S}\mathbf{F} - \lambda_1 \mathbf{I}(\mathbf{x}) \left(\frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\bar{\mathbf{u}}}{|\bar{\mathbf{u}}|} \right), \quad (43a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (43b)$$

$$\frac{\partial \phi_k}{\partial t} = \frac{1}{Pe} \Delta \mu_k - \mathbf{u} \cdot \nabla \phi_k, \quad (43c)$$

$$\mu_k = -\epsilon^2 \Delta \phi_k + F'(\phi_k) + \lambda_2 \mathbf{I}(\mathbf{x}) (\phi - \bar{\phi}). \quad (43d)$$

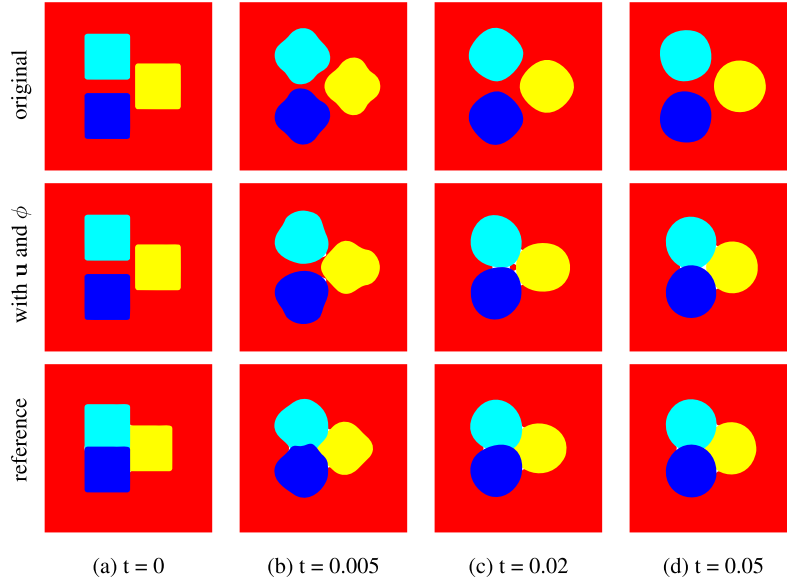


Fig. 10. Illustrations of the multi-component process at time $t = 0, 0.005, 0.02$, and 0.05 , respectively.

Within this framework, the density of a multiphase incompressible fluid is represented as $\rho(\phi) = \sum_{k=1}^K \phi_k \rho_k$, where each ϕ_k corresponds to the volume fraction of the k -th fluid. The term $F'(\phi_k) = \phi_k^2(1 - \phi_k)^2/4$ describes the free energy density associated with each phase. A crucial element in this model is the surface tension force, denoted as $\mathbf{SF} = \sum_{k=1}^K \kappa(\phi_k) \mathbf{n}(\phi_k) \delta(\phi_k)$. This force is pivotal in managing multi-component systems, particularly during significant topological alterations of the interface. Here, $\kappa(\phi_k) = \nabla \cdot (\nabla \phi_k / |\nabla \phi_k|)$ represents the mean curvature, $\mathbf{n}(\phi_k) = (\nabla \phi_k / |\nabla \phi_k|)$ is the unit normal vector, and $\delta(\phi_k) = |\nabla \phi_k|$ signifies the smoothed Dirac delta function corresponding to the interface of the k -th fluid.

In this experiment, we choose the computational domain as $\Omega = [0, 1] \times [0, 1]$ with a 256×256 mesh grid and Δt are set as 10^{-5} . Other parameters are set as $\lambda_1 = 0.5$, $\lambda_2 = 5000$, $\epsilon = 5h/2\sqrt{2} \tanh(0.9)$, $Re = 200$, $We = 0.1$, and $Pe = 1/\epsilon$. The initial conditions of the reference process are set as:

$$\begin{aligned}
 \phi_1(x, y, 0) &= 1 - \tanh(|x - 0.25| + |x - 0.50| + |y - 0.50| + |y - 0.75| - 0.5) / (2\sqrt{2\epsilon}), \\
 \phi_2(x, y, 0) &= 1 - \tanh(|x - 0.25| + |x - 0.50| + |y - 0.50| + |y - 0.25| - 0.5) / (2\sqrt{2\epsilon}), \\
 \phi_3(x, y, 0) &= 1 - \tanh(|x - 0.50| + |x - 0.75| + |y - 0.375| + |y - 0.625| - 0.5) / (2\sqrt{2\epsilon}), \\
 \phi_4(x, y, 0) &= 1 - \phi_1(x, y, 0) - \phi_2(x, y, 0) - \phi_3(x, y, 0), \\
 u(x, y, 0) &= v(x, y, 0) = 10^{-6} \text{rand}(x, y), \\
 p(x, y, 0) &= 0.
 \end{aligned} \tag{44}$$

The observed data are sampled from the reference process with a time interval $100\Delta t$. The initial conditions of the simulated processes are defined to be:

$$\begin{aligned}
 \phi_1(x, y, 0) &= 1 - \tanh(|x - 0.25| + |x - 0.50| + |y - 0.55| + |y - 0.80| - 0.5) / (2\sqrt{2\epsilon}), \\
 \phi_2(x, y, 0) &= 1 - \tanh(|x - 0.25| + |x - 0.50| + |y - 0.45| + |y - 0.20| - 0.5) / (2\sqrt{2\epsilon}), \\
 \phi_3(x, y, 0) &= 1 - \tanh(|x - 0.55| + |x - 0.85| + |y - 0.35| + |y - 0.65| - 0.6) / (2\sqrt{2\epsilon}), \\
 \phi_4(x, y, 0) &= 1 - \phi_1(x, y, 0) - \phi_2(x, y, 0) - \phi_3(x, y, 0), \\
 u(x, y, 0) &= v(x, y, 0) = 10^{-6} \text{rand}(x, y), \\
 p(x, y, 0) &= 0.
 \end{aligned} \tag{45}$$

The results are illustrated in Fig. 10. The first row of Fig. 10 depicts the original process, characterized by $\lambda_1 = \lambda_2 = 0$. In this scenario, when an initial condition presents a distinct interval between different phases, they remain separate and do not blend. Conversely, the third row demonstrates the reference process, wherein no interval exists between the three phases, leading to their amalgamation. Utilizing observed data sampled from the reference process, the outcomes of the simulated process are displayed in the second row of Fig. 10. Notably, it can be observed that the data assimilation term significantly alters the evolving direction of the process.

5. Conclusion

In this paper, we introduced a modification to the NSCH equation, incorporating a new data assimilation term to effectively utilize observed data during the equation-solving process. To solve the modified equation efficiently, we proposed a numerical scheme based on the Crank-Nicolson formula and the SAV approach. The unconditional energy stability of our proposed scheme has been both theoretically and experimentally proven. In the experimental section, we validated the mass conservation property of our proposed scheme and examined its performance with various initial

conditions. Furthermore, we have conducted investigations into the effects of different parameters, such as the data assimilation coefficients λ_1 and λ_2 , sample interval, and sample rate. We explored scenarios where the observed data contain different levels of noise. We also applied our scheme to a classical droplet experiment to demonstrate the efficiency of our approach.

Data availability

No data was used for the research described in the article.

Acknowledgement

This work is supported by National Natural Science Foundation of China (No. 12271430). The authors would like to thank the reviewers for their constructive and helpful comments regarding the revision of this article.

References

- [1] Q. Xia, J. Yang, Y. Li, On the conservative phase-field method with the N-component incompressible flows, *Phys. Fluids* 35 (2023) 012120.
- [2] Q. Xia, G. Sun, J. Kim, Y. Li, Thermal–fluid topology optimization with unconditional energy stability and second-order accuracy via phase-field model, *Commun. Nonlinear Sci. Numer. Simul.* 116 (2023) 106782.
- [3] Q. Xia, J. Kim, Y. Li, Modeling and simulation of multi-component immiscible flows based on a modified Cahn–Hilliard equation, *Eur. J. Mech. B, Fluids* 95 (2022) 194–204.
- [4] Y. Li, J. Choi, J. Kim, Multi-component Cahn–Hilliard system with different boundary conditions in complex domains, *J. Comput. Phys.* 323 (2016) 1–16.
- [5] Y. Li, Y. Choi, J. Kim, Computationally efficient adaptive timestep method for the Cahn–Hilliard equation, *Comput. Math. Appl.* 73 (2017) 1855–1864.
- [6] J. Blum, F.-X. Le Dimet, I.M. Navon, Data assimilation for geophysical fluids, *Handb. Numer. Anal.* 14 (2009) 385–441.
- [7] M. D’Elia, M. Perego, A. Veneziani, A variational data assimilation procedure for the incompressible Navier–Stokes equations in hemodynamics, *SIAM J. Sci. Comput.* 52 (2012) 340–359.
- [8] R.E. Kalman, A new approach to linear filtering and prediction problems, *J. Basic Eng.* 82 (1960) 35–45.
- [9] D.T. Kelly, K.J. Law, A.M. Stuart, Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time, *Nonlinearity* 27 (10) (2014) 2579.
- [10] A. Yamamura, S. Sakane, M. Ohno, H. Yasuda, T. Takaki, Data assimilation with phase-field lattice Boltzmann method for dendrite growth with liquid flow and solid motion, *Comput. Mater. Sci.* 215 (2022) 111776.
- [11] K. Sasaki, A. Yamanaka, S.-i. Ito, H. Nagao, Data assimilation for phase-field models based on the ensemble Kalman filter, *Comput. Mater. Sci.* 141 (2018) 141–152.
- [12] R. Fablet, Q. Febvre, B. Chapron, Multimodal 4DVarNets for the reconstruction of sea surface dynamics from SST–SSH synergies, *IEEE Trans. Geosci. Remote Sens.* (2023).
- [13] S. Pawar, S.E. Ahmed, O. San, A. Rasheed, I.M. Navon, Long short-term memory embedded nudging schemes for nonlinear data assimilation of geophysical flows, *Phys. Fluids* 32 (7) (2020) 076606.
- [14] R. Arcucci, J. Zhu, S. Hu, Y.-K. Guo, Deep data assimilation: integrating deep learning with data assimilation, *Appl. Sci.* 11 (3) (2021) 1114.
- [15] R.A. Anthes, Data assimilation and initialization of hurricane prediction models, *J. Atmos. Sci.* 31 (3) (1974) 702–719.
- [16] A. Azouani, E. Olson, E.S. Titi, Continuous data assimilation using general interpolant observables, *J. Nonlinear Sci.* 24 (2) (2014) 277–304.
- [17] A. Azouani, E.S. Titi, Feedback control of nonlinear dissipative systems by finite determining parameters—a reaction–diffusion paradigm, *arXiv preprint*, arXiv:1301.6992, 2013.
- [18] M.S. Jolly, V.R. Martinez, E.S. Titi, A data assimilation algorithm for the subcritical surface quasi-geostrophic equation, *Adv. Nonlinear Stud.* 17 (1) (2017) 167–192.
- [19] D.A. Albanese, M.J. Benvenuti, Continuous data assimilation algorithm for simplified Bardina model, *Evol. Equ. Control Theory* 7 (1) (2018) 33.
- [20] P.A. Markowich, E.S. Titi, S. Trabelsi, Continuous data assimilation for the three-dimensional Brinkman–Forchheimer–extended Darcy model, *Nonlinearity* 29 (4) (2016) 1292.
- [21] D.A. Albanese, H.J. Nussenzweig Lopes, E.S. Titi, Continuous data assimilation for the three-dimensional Navier–Stokes– α model, *Asymptot. Anal.* 97 (1–2) (2016) 139–164.
- [22] C. Foias, R. Temam, Determination of the solutions of the Navier–Stokes equations by a set of nodal values, *Math. Comput.* 43 (167) (1984) 117–133.
- [23] C. Foias, C.F. Mondaini, E.S. Titi, A discrete data assimilation scheme for the solutions of the two-dimensional Navier–Stokes equations and their statistics, *SIAM J. Appl. Dyn. Syst.* 15 (4) (2016) 2109–2142.
- [24] E. Carlson, J. Hudson, A. Larios, Parameter recovery for the 2 dimensional Navier–Stokes equations via continuous data assimilation, *SIAM J. Sci. Comput.* 42 (1) (2020) A250–A270.
- [25] G. Zhu, H. Chen, S. Sun, Efficient energy-stable schemes for the hydrodynamics coupled phase-field model, *Appl. Math. Model.* 70 (2019) 82–108.
- [26] G. Zhu, J. Kou, J. Yao, A. Li, S. Sun, A phase-field moving contact line model with soluble surfactants, *J. Comput. Phys.* 70 (2020) 109170.
- [27] Y. Li, Q. Yu, W. Fang, B. Xia, J. Kim, A stable second-order BDF scheme for the three-dimensional Cahn–Hilliard–Hele–Shaw system, *Adv. Comput. Math.* 47 (2021) 3.
- [28] Y. Li, J. Kim, N. Wang, An unconditionally energy-stable second-order time-accurate scheme for the Cahn–Hilliard equation on surfaces, *Commun. Nonlinear Sci. Numer. Simul.* 53 (2017) 213–227.
- [29] Y. Li, A. Yun, D. Lee, J. Shin, D. Jeong, J. Kim, Three-dimensional volume-conserving immersed boundary model for two-phase fluids flow, *Comput. Methods Appl. Math.* 257 (2013) 36–46.
- [30] J. Shen, X. Yang, Numerical approximations of Allen–Cahn and Cahn–Hilliard equations, *Discrete Contin. Dyn. Syst.* 28 (4) (2010) 1669–1691.
- [31] J. Zhao, D. Han, Second-order decoupled energy-stable schemes for Cahn–Hilliard–Navier–Stokes equations, *J. Comput. Phys.* 443 (2021) 110536.
- [32] J. Shen, J. Xu, J. Yang, The scalar auxiliary variable (SAV) approach for gradient flows, *J. Comput. Phys.* 353 (2018) 407–416.
- [33] P.M. Chaikin, T.C. Lubensky, T.A. Witten, *Principles of Condensed Matter Physics*, vol. 10, Cambridge University, 1995.
- [34] M. Jiang, Z. Zhang, J. Zhao, Improving the accuracy and consistency of the scalar auxiliary variable (SAV) method with relaxation, *J. Comput. Phys.* 456 (2022) 110954.
- [35] C. Zhang, J. Ouyang, C. Wang, M. Steven, Numerical comparison of modified-energy stable SAV-type schemes and classical BDF methods on benchmark problems for the functionalized Cahn–Hilliard equation, *J. Comput. Phys.* 423 (2020) 109772.
- [36] M. Wang, Q. Huang, C. Wang, A second order accurate scalar auxiliary variable (SAV) numerical method for the square phase field crystal equation, *J. Sci. Comput.* 88 (2021) 33.
- [37] X. Yang, A novel fully-decoupled, second-order and energy stable numerical scheme of the conserved Allen–Cahn type flow-coupled binary surfactant model, *Comput. Methods Appl. Mech. Eng.* 373 (2021) 113502.
- [38] L. Chen, J. Zhao, A novel second-order linear scheme for the Cahn–Hilliard–Navier–Stokes equations, *J. Comput. Phys.* 423 (2020) 109782.
- [39] Y. Qian, C. Wang, S. Zhou, Convergence analysis on a structure-preserving numerical scheme for the Poisson–Nernst–Planck–Cahn–Hilliard system, *CSIAM Trans. Appl. Math.* 4 (2023) 345–380.
- [40] X. Li, Z. Qiao, C. Wang, Stabilization parameter analysis of a second-order linear numerical scheme for the nonlocal Cahn–Hilliard equation, *IMA J. Numer. Anal.* 43 (2023) 1089–1114.
- [41] C. Liu, C. Wang, M. Steven, S. Zhou, An iteration solver for the Poisson–Nernst–Planck system and its convergence analysis, *J. Comput. Appl. Math.* 406 (2022) 114017.
- [42] Y. Cai, J. Chen, C. Wang, C. Xie, A second-order numerical method for Landau–Lifshitz–Gilbert equation with large damping parameters, *J. Comput. Phys.* 451 (2022) 110831.
- [43] W. Feng, Z. Guan, J. Lowengrub, C. Wang, M. Steven, Y. Chen, A uniquely solvable, energy stable numerical scheme for the functionalized Cahn–Hilliard equation and its convergence analysis, *J. Sci. Comput.* 76 (2018) 1938–1967.
- [44] K. Cheng, C. Wang, M. Steven, A weakly nonlinear, energy stable scheme for the strongly anisotropic Cahn–Hilliard equation and its convergence analysis, *J. Comput. Phys.* 405 (2020) 109109.
- [45] W. Chen, W. Feng, Y. Liu, C. Wang, M. Steven, A second order energy stable scheme for the Cahn–Hilliard–Hele–Shaw equations, *Discrete Contin. Dyn. Syst., Ser. B* 85 (2016) 2231–2257.
- [46] X. Li, J. Shen, Error analysis of the SAV–MAC scheme for the Navier–Stokes equations, *SIAM J. Numer. Anal.* 58 (2020) 2465–2491.
- [47] Y. He, H. Chen, Efficient algorithm and convergence analysis of conservative SAV compact difference scheme for Boussinesq Paradigm equation, *Comput. Math. Appl.* 125 (2022) 34–50.
- [48] Y. He, H. Chen, Efficient and conservative compact difference scheme for the coupled Schrödinger–Boussinesq equations, *Appl. Numer. Math.* 182 (2022) 285–307.
- [49] L. Ju, X. Li, Z. Qiao, Generalized SAV–exponential integrator schemes for Allen–Cahn type gradient flows, *SIAM J. Numer. Anal.* 60 (2022) 1905–1931.
- [50] X. Li, J. Shen, H. Rui, Energy stability and convergence of SAV block-centered finite difference method for gradient flows, *Math. Comput.* 88 (2019) 2047–2068.