

# Unsupervised Machine Learning Based on Non-Negative Tensor Factorization for Analyzing Reactive-Mixing

V. V. Vesselinov<sup>1,\*</sup>, M. K. Mudunuru<sup>1</sup>, S. Karra<sup>1</sup>, D. O'Malley<sup>1</sup>, and B. S. Alexandrov<sup>2</sup>

<sup>1</sup>Computational Earth Science Group, Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545.

<sup>2</sup>Physics and Chemistry of Materials Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545.

## Abstract

Analysis of reactive-diffusion simulations representing complex mixing processes requires a large number of independent model runs. For each high-fidelity model simulation, the model inputs are varied and the predicted mixing behavior is represented by temporal and spatial changes in species concentration. It is then required to discern how the model inputs (such as diffusivity, dispersion, anisotropy, and velocity field properties) impact the mixing process. This task is challenging and typically involves interpretation of large model outputs representing temporal and spatial changes of species concentration within the model domain. However, the task can be automated and substantially simplified by applying Machine Learning (ML) methods. In this paper, we present an application of an unsupervised ML method (called NTF $k$ ) using Non-negative Tensor Factorization (NTF) coupled with a custom clustering procedure based on  $k$ -means to reveal the temporal and spatial features in product concentrations. An attractive and unique aspect of the proposed ML method is that it ensures the extracted features are non-negative, which are important to obtain a meaningful deconstruction of the mixing processes. The ML methodology is applied to a large set of high-resolution finite-element model simulations representing anisotropic reaction-diffusion processes in perturbed vortex-based velocity fields. The applied finite-element method ensures that spatial and temporal species concentration are always non-negative, even in the case of high anisotropic contrasts. The simulated reaction is a fast irreversible bimolecular reaction  $A + B \rightarrow C$ , where species  $A$  and  $B$  react to form species  $C$ . The reactive-diffusion model input parameters that control mixing include properties of velocity field (such as vortex structures), anisotropic dispersion, and molecular diffusion. We demonstrate the applicability of the ML feature extraction method to produce a meaningful deconstruction of model outputs to discriminate between different physical processes impacting the reactants, their mixing, and the spatial distribution of the product  $C$ . The presented ML analysis allowed us to identify additive temporal and spatial features that characterize mixing behavior. The application of the proposed NTF $k$  approach is not limited to reactive-mixing. NTF $k$  can be readily applied to any observed or simulated datasets that can be represented as tensors and have separable latent signatures or features.

**Keywords:** Non-negative tensor factorization; Tucker decompositions; Structure-preserving feature extraction; Unsupervised machine learning; Reaction-diffusion equations; Anisotropic dispersion; Reactive-mixing; Non-negative solutions; Fast irreversible bimolecular reactions

### Highlights of the paper

- NTF $k$  provide an elegant way to extract hidden features in irreversible fast bimolecular reaction-diffusion systems.
- Features extracted using NTF $k$  are non-negative and physically meaningful, which are an important constraint to understand mixing process.
- Another added benefit of our proposed NTF $k$  method is data compression, which is achieved with minimal loss of information. In our case, the compression ratio is in the  $\mathcal{O}(10^{-4})$

## 1. Introduction

Mixing and reaction of species is an important physical phenomena crucial in many applied research areas related not only to flow in porous/fractured media but also to turbulent flows at various scales in groundwater, rivers, lakes, oceans, and the atmosphere [1–10]. A problem of predominant interest is groundwater contaminant remediation. For several decades in the past and for the foreseeable future, aquifer contamination has been and is expected to be one of the most important application in the hydrogeological sciences [11–13]. Research has been driven by substantial scientific and engineering challenges associated with prediction and remediation of contaminant plumes in natural environments. Many of these challenges are due to geochemical complexities associated with contaminant fate and transport in the subsurface, which is affected by numerous physical processes. One of the most important factors is fluid flow. The flow controls how the geochemical species of interest (e.g., contaminants and remedial agents) move and interact in the subsurface porous/fractured media. Typically, water in an aquifer is a mixture of different groundwater types with different origins and geochemical signatures [14]. For example, groundwater might be originating from different recharge sources with contrasting geochemical properties. Also, groundwater may have been flowing through different rock types which may have altered the composition by means of geochemical reactions and ion exchanges. Furthermore, some of the groundwater recharge sources might be associated with contamination sources with different geochemical signatures. In the case of active contaminant remediation, additional complexity occurs due to injection of water with species which are expected to alter the groundwater composition and eliminate the contaminant species (for example, through reduction and/or precipitation). Understanding how different groundwater types mix and how some of the species in these mixtures react is a challenging but very important task. This task is typically performed using complex inverse models representing groundwater mixing processes in the aquifer [15–23].

Typically, mixing analyses are performed in the laboratory based on microfluidics experiments [24]. Another option is to perform high-resolution numerical simulations which are capable of representing the governing processes. Typically, these simulations involve solving anisotropic reaction-diffusion equations using numerical methods like the pseudo-spectral and finite element methods [25–28]. From these simulations, one then infers scaling laws related to quantities of interest such as species decay rates, product yield, degree of mixing, extent of spreading, etc [27–29]. Since the reactive-mixing simulations are dependent on a series of model input parameters and produce

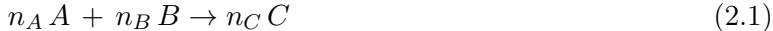
large model outputs of state variables, the interpretation of the subsequent modeling results is challenging. For instance, one can be interested in how the model input parameters impact the aforementioned quantities of interest over time. This interpretation can be automated and substantially simplified by applying unsupervised Machine Learning (ML) methods such as Non-negative Matrix Factorization (NMF) and Non-negative Tensor Factorization (NTF) [30]. For example, NMF/NTF-based ML methods have been successfully used for analysis of Monte Carlo simulated fission chamber’s output signals [31], for compression of scientific simulation data [32], and for a variety of other different applications [30]. The goal of these unsupervised ML methods (such as NMF and NTF) is to separate the dependence of a quantity of interest into independent physical/chemical processes. Through such decomposition, NMF/NTF methods not only provide rapid data analysis and data reduction but also provide insight into the physics of the underlying system.

Conventional numerical formulations for anisotropic reaction-diffusion equations can produce negative and unphysical solutions for concentration of chemical species [33–36], and this is exacerbated when anisotropy dominates [6, 25, 26]. However, it is absurd for a physical species to have a negative concentration. To overcome this first challenge, we employ a novel non-negative finite element method [25, 26, 29] to produce high-fidelity non-negative numerical solutions. This non-negative finite element method provides physically meaningful concentrations for reaction-diffusion equations even under high anisotropy. Then, in order to discover hidden features in the high-fidelity data, we present a new unsupervised ML methodology by combining NTF and  $k$ -means clustering. The proposed ML method is called NTF $k$ . NTF $k$  ensures that the discovered features are additive and are in accordance with the non-negativity of underlying numerical solutions. For this reason, NTF $k$  is very attractive for our problem as it is a structure-preserving feature extraction approach, which other unsupervised ML methods that does not include non-negativity constraints do not guarantee.

The paper is organized as follows. In Section 2, we describe briefly a mathematical model for reaction-diffusion mixing simulations. Then, we present our NTF $k$  method with the goal of discovering hidden non-negative features present in the high-fidelity model outputs. High-fidelity datasets for NTF $k$  are generated by performing a series of model runs using a parallel non-negative finite element solver with different model input parameters representing uncertainties in the underlying reaction-diffusion processes. The obtained results are presented in Section 3 followed by conclusions in Section 4.

## 2. Methodology

**2.1. Reaction-Diffusion Mixing Simulations.** Let  $\Omega \subset \mathbb{R}^n$  be a domain with  $n$  spatial dimensions and a piecewise smooth boundary  $\partial\Omega$ . Let  $\bar{\Omega}$  be the set closure of  $\Omega$  and  $\mathbf{x} \in \bar{\Omega}$  be a spatial point. The divergence and gradient operators with respect to  $\mathbf{x}$  are denoted by  $\text{div}[\bullet]$  and  $\text{grad}[\bullet]$ , respectively. Let  $\mathbf{n}(\mathbf{x})$  be the unit outward normal to  $\partial\Omega$ . Let  $t \in (0, \mathcal{I})$  denote the time, where  $\mathcal{I}$  is the length of time of interest. Consider a fast bimolecular reaction where species  $A$  and  $B$  react irreversibly to produce species  $C$ :



where  $n_A$ ,  $n_B$ , and  $n_C$  are (positive) stoichiometric coefficients.

The governing equations for the reaction in Eq. (2.1) without any sources/sinks are given by:

$$\frac{\partial c_A}{\partial t} - \operatorname{div}[\mathbf{D}(\mathbf{x}, t) \operatorname{grad}[c_A]] = -n_A k_{AB} c_A c_B \quad \text{in } \Omega \times ]0, \mathcal{I}[ \quad (2.2a)$$

$$\frac{\partial c_B}{\partial t} - \operatorname{div}[\mathbf{D}(\mathbf{x}, t) \operatorname{grad}[c_B]] = -n_B k_{AB} c_A c_B \quad \text{in } \Omega \times ]0, \mathcal{I}[ \quad (2.2b)$$

$$\frac{\partial c_C}{\partial t} - \operatorname{div}[\mathbf{D}(\mathbf{x}, t) \operatorname{grad}[c_C]] = +n_C k_{AB} c_A c_B \quad \text{in } \Omega \times ]0, \mathcal{I}[ \quad (2.2c)$$

$$c_i(\mathbf{x}, t) = c_i^p(\mathbf{x}, t) \quad \text{on } \Gamma_i^D \times ]0, \mathcal{I}[ \quad (i = A, B, C) \quad (2.2d)$$

$$(-\mathbf{D}(\mathbf{x}, t) \operatorname{grad}[c_i]) \bullet \mathbf{n}(\mathbf{x}) = h_i^p(\mathbf{x}, t) \quad \text{on } \Gamma_i^N \times ]0, \mathcal{I}[ \quad (i = A, B, C) \quad (2.2e)$$

$$c_i(\mathbf{x}, t = 0) = c_i^0(\mathbf{x}) \quad \text{in } \Omega \quad (i = A, B, C) \quad (2.2f)$$

where  $c_i$  is the molar concentration of  $i$ -th chemical species,  $\mathbf{D}(\mathbf{x}, t)$  is the anisotropic dispersion tensor, and  $k_{AB}$  is the bilinear reaction rate coefficient.  $c_i^p(\mathbf{x}, t)$  and  $h_i^p(\mathbf{x}, t)$  are the prescribed molar concentration and flux on Dirichlet and Neumann boundaries  $\Gamma_i^D$  and  $\Gamma_i^N$ , respectively. Additionally,  $c_i^0(\mathbf{x})$  is the initial concentration of  $i$ -th chemical species.

Using the non-negative linear transformations [25, 29],

$$c_F := c_A + \left(\frac{n_A}{n_C}\right) c_C \quad (2.3a)$$

$$c_G := c_B + \left(\frac{n_B}{n_C}\right) c_C \quad (2.3b)$$

Eqs. (2.2a)–(2.2f) can be transformed to

$$\frac{\partial c_F}{\partial t} - \operatorname{div}[\mathbf{D}(\mathbf{x}, t) \operatorname{grad}[c_F]] = 0 \quad \text{in } \Omega \times ]0, \mathcal{I}[ \quad (2.4a)$$

$$c_F(\mathbf{x}, t) = c_F^p(\mathbf{x}, t) := c_A^p(\mathbf{x}, t) + \left(\frac{n_A}{n_C}\right) c_C^p(\mathbf{x}, t) \quad \text{on } \Gamma^D \times ]0, \mathcal{I}[ \quad (2.4b)$$

$$(-\mathbf{D}(\mathbf{x}, t) \operatorname{grad}[c_F]) \bullet \mathbf{n}(\mathbf{x}) = h_F^p(\mathbf{x}, t) := h_A^p(\mathbf{x}, t) + \left(\frac{n_A}{n_C}\right) h_C^p(\mathbf{x}, t) \quad \text{on } \Gamma^N \times ]0, \mathcal{I}[ \quad (2.4c)$$

$$c_F(\mathbf{x}, t = 0) = c_F^0(\mathbf{x}) := c_A^0(\mathbf{x}) + \left(\frac{n_A}{n_C}\right) c_C^0(\mathbf{x}) \quad \text{in } \Omega \quad (2.4d)$$

and

$$\frac{\partial c_G}{\partial t} - \operatorname{div}[\mathbf{D}(\mathbf{x}, t) \operatorname{grad}[c_G]] = 0 \quad \text{in } \Omega \times ]0, \mathcal{I}[ \quad (2.5a)$$

$$c_G(\mathbf{x}, t) = c_G^p(\mathbf{x}, t) := c_B^p(\mathbf{x}, t) + \left(\frac{n_B}{n_C}\right) c_C^p(\mathbf{x}, t) \quad \text{on } \Gamma^D \times ]0, \mathcal{I}[ \quad (2.5b)$$

$$(-\mathbf{D}(\mathbf{x}, t) \operatorname{grad}[c_G]) \bullet \mathbf{n}(\mathbf{x}) = h_G^p(\mathbf{x}, t) := h_B^p(\mathbf{x}, t) + \left(\frac{n_B}{n_C}\right) h_C^p(\mathbf{x}, t) \quad \text{on } \Gamma^N \times ]0, \mathcal{I}[ \quad (2.5c)$$

$$c_G(\mathbf{x}, t = 0) = c_G^0(\mathbf{x}) := c_B^0(\mathbf{x}) + \left(\frac{n_B}{n_C}\right) c_C^0(\mathbf{x}) \quad \text{in } \Omega \quad (2.5d)$$

Since fast bimolecular reactions are considered, it is assumed that the reactants  $A$  and  $B$  cannot co-exist at the same location  $\mathbf{x}$  at any instant of time. Based on this assumption, quantities  $c_F(\mathbf{x}, t)$



and  $c_G(\mathbf{x}, t)$  can be used to evaluate  $c_A(\mathbf{x}, t)$ ,  $c_B(\mathbf{x}, t)$  and,  $c_C(\mathbf{x}, t)$  via:

$$c_A(\mathbf{x}, t) = \max \left[ c_F(\mathbf{x}, t) - \left( \frac{n_A}{n_B} \right) c_G(\mathbf{x}, t), 0 \right] \quad (2.6a)$$

$$c_B(\mathbf{x}, t) = \left( \frac{n_B}{n_A} \right) \max \left[ -c_F(\mathbf{x}, t) + \left( \frac{n_A}{n_B} \right) c_G(\mathbf{x}, t), 0 \right] \quad (2.6b)$$

$$c_C(\mathbf{x}, t) = \left( \frac{n_C}{n_A} \right) (c_F(\mathbf{x}, t) - c_A(\mathbf{x}, t)) \quad (2.6c)$$

The uncoupled governing equations given by Eqs. (2.4a)–(2.5d) are solved using the lower-order non-negative finite element methods [25, 26, 29] to produce non-negative solutions. Once we obtain the non-negative quantities  $c_F$  and  $c_G$ , Eqs. (2.6a)–(2.6c) provide the concentrations of  $A$ ,  $B$ , and  $C$ .

Fig. 1 provides a pictorial description of the initial boundary value problem. The computational domain is a square with  $L = 1$  and zero flux boundary conditions. For all the chemical species, the non-reactive volumetric source  $f_i(\mathbf{x}, t)$  is equal to zero. Initially, species  $A$  and  $B$  are separated;  $A$  is placed in the left half of the domain and  $B$  is placed in the right half. The stoichiometric coefficients are taken as  $n_A = 1$ ,  $n_B = 1$ , and  $n_C = 1$ . The total time of interest is taken as  $\mathcal{I} = 1$ . The dispersion tensor follows the formulation of [37]:

$$\mathbf{D}_{\text{subsurface}}(\mathbf{x}) = D_m \mathbf{I} + \alpha_T \|\mathbf{v}\| \mathbf{I} + \frac{\alpha_L - \alpha_T}{\|\mathbf{v}\|} \mathbf{v} \otimes \mathbf{v} \quad (2.7)$$

where  $D_m$  is the molecular diffusivity,  $\alpha_L$  is the longitudinal dispersivity,  $\alpha_T$  is the transverse dispersivity,  $\mathbf{I}$  is the identity tensor,  $\otimes$  is the tensor product,  $\mathbf{v}$  is the velocity vector field, and  $\|\bullet\|$  is the Frobenius norm. Synthetic model velocity field is used to define the anisotropic dispersion tensor through the following stream function [27–29]:

$$\psi(\mathbf{x}, t) = \begin{cases} \frac{1}{2\pi\kappa_f} (\sin(2\pi\kappa_f x) - \sin(2\pi\kappa_f y) + v_o \cos(2\pi\kappa_f y)) & \text{if } \nu T \leq t < (\nu + \frac{1}{2}) T \\ \frac{1}{2\pi\kappa_f} (\sin(2\pi\kappa_f x) - \sin(2\pi\kappa_f y) - v_o \cos(2\pi\kappa_f x)) & \text{if } (\nu + \frac{1}{2}) T \leq t < (\nu + 1) T \end{cases} \quad (2.8)$$

where  $\nu = 0, 1, 2, 3, \dots$  is an integer and  $v_o$  is the perturbation parameter.  $\kappa_f L$  and  $T$  are characteristic spatial and time-scales of the vortex-based velocity field. Using Eq. (2.8), the divergence free velocity field components are given as follows:

$$v_x(\mathbf{x}, t) = -\frac{\partial \psi}{\partial y} = \begin{cases} \cos(2\pi\kappa_f y) + v_o \sin(2\pi\kappa_f y) & \text{if } \nu T \leq t < (\nu + \frac{1}{2}) T \\ \cos(2\pi\kappa_f y) & \text{if } (\nu + \frac{1}{2}) T \leq t < (\nu + 1) T \end{cases} \quad (2.9)$$

$$v_y(\mathbf{x}, t) = +\frac{\partial \psi}{\partial x} = \begin{cases} \cos(2\pi\kappa_f x) & \text{if } \nu T \leq t < (\nu + \frac{1}{2}) T \\ \cos(2\pi\kappa_f x) + v_o \sin(2\pi\kappa_f x) & \text{if } (\nu + \frac{1}{2}) T \leq t < (\nu + 1) T \end{cases} \quad (2.10)$$

Fig. 2 shows the streamlines of the vortex-based velocity field given by Eqs. (2.9)–(2.10). Small-scale and large-scale vortices are observed when  $\kappa_f L$  is high and low, respectively. The number of vortices increases for larger values of  $\kappa_f L$  and as a result there are more small-scale features in the velocity field. Fig. 2 also shows that the perturbation parameter  $v_o$  does not substantially impact the location of the vortices.

**2.2. Non-negative Tensor Factorization.** In this subsection, we present a novel structure-preserving feature extraction method to extract hidden features from high-resolution reaction-diffusion model outputs. These outputs are generated by varying input parameters in Eqs. (2.7)–(2.10). The proposed methodology is based on NTF, which is an emerging research field in the area of data analytics and data compression. In addition to extracting hidden features that are buried in large high-dimensional datasets [30], NTF-based methods are also used in source separation. For example, Blind Source Separation (BSS) techniques based on matrix factorization methods such as Principle Component Analysis (PCA) [38], Independent Component Analysis (ICA) [39], and Non-negative Matrix/Tensor Factorization (NMF/NTF) [40], form a class of unsupervised machine learning (ML) methods that are instrumental in model-free feature extraction and dimensionality reduction. When a BSS technique is applied in signal processing, the extracted features are the unique original signals represented in the mixtures of signals that are recorded by a set of spatially distributed sensors (e.g., the voices of several speakers recorded by multiple microphones placed at different locations in a ball room [41]). However, the matrix factorization methods are inherently deficient for examining multidimensional datasets (i.e., tensor datasets), which are the natural extensions of the matrix datasets. Numerical model outputs are typically multidimensional and discrete. They often represent one or more state variables at a discrete set of locations in space and time, and, as a result, are ideal for NTF analyses. There are two main tensor factorization methods: Canonical Polyadic (CP) decomposition (CANDECOMP/PARAFAC) [42–44] and Tucker decomposition [45, 46]. Both can be computationally effective for interpreting multidimensional datasets [47].

Herein, we analyze three-dimensional data that present the evolution in time and space of concentration of product  $C(t, x, y)$ . These concentrations are generated by non-negative finite element numerical simulations of the reaction-diffusion equation described in Section 2.1. Fig. 3 and Algorithm 1 summarizes the proposed tensor-factorization methodology. For a given set of parameters, the output of our simulations is a data-tensor  $X_{ijl}$  in three dimensions:  $(t, x, y) \Rightarrow (1000 \times 81 \times 81)$ .  $X(t, x, y)$  corresponds to concentrations of non-negative invariants, which are  $c_F$  and  $c_G$ . In each cell of this non-negative tensor, we have the concentration  $X(t_i, x_j, y_l)$  at a specific time  $t_i$  and at a spatial point whose coordinates are  $(x_j, y_l)$ . Due to inherent non-linearity in the problem, first, hidden features are extracted from each non-negative invariant. Then, using the non-linear transformations given by Eqs. (2.6a)–(2.6c), we obtain the corresponding hidden features of product  $C$ .

To analyze such a three-dimensional simulation dataset, we utilize a sparse non-negative Tucker-3 decomposition model [48]. Our choice for non-negative constraints is motivated by (i) the fact that the considered data-tensors are inherently non-negative and (ii) our goal to relate the extracted features to easily interpretable quantities without introducing any a priori assumptions. Indeed, a meaningful interpretation of the obtained results requires the extracted features to be *parts* of the original data [49]. The non-negative constraints in the tensor factorization process lead to extraction of strictly additive components, which are parts of the original data [50]. Thus, the non-negative tensor factorization has the ability to identify readily understandable structure-preserving features that enables the discovery of new causal structures and unknown mechanisms hidden in the data [30].

The non-negative Tucker-3 decomposition of the three dimensional data-tensor representing the temporal evolution of  $X(t, x, y)$  on a two-dimensional  $(x, y)$  non-negative finite element grid is,

$$X(t, x, y) = G \otimes W(t) \otimes H(x) \otimes V(y) + \epsilon(t, x, y) \quad (2.11)$$

where  $\otimes$  denotes the tensor product. The decomposition of the non-negative tensor  $X(t, x, y)$  ( $X \in \mathbb{R}_{\geq 0}^{K \times M \times N}$ ) can be expressed by components:

$$X_{ijl} = \sum_{p=1}^k \sum_{q=1}^m \sum_{r=1}^n G_{pqr} W_{ip} H_{jq} V_{lr} + \epsilon_{ijl} \quad \forall i, j, l \quad (2.12)$$

where all the elements of  $X$ ,  $G$ ,  $W$ ,  $H$ , and  $V$  are non-negative,

$$X_{ijl}, G_{pqr}, W_{ip}, H_{jq}, V_{lr} \geq 0 \quad \forall i, j, l, p, q, r \quad (2.13)$$

Eq. 2.12 can be also represented as:

$$X_{ijl} = \tilde{X}_{ijl} + \epsilon_{ijl} \quad (2.14)$$

where  $\tilde{X}$  ( $\tilde{X} \in \mathbb{R}_{\geq 0}^{K \times M \times N}$ ) is the Tucker-3 estimate of  $X$ :

$$\tilde{X}_{ijl} = \sum_{p=1}^k \sum_{q=1}^m \sum_{r=1}^n G_{pqr} W_{ip} H_{jq} V_{lr} \quad (2.15)$$

The Tucker-3 decomposition includes (i) an unknown core-tensor  $G$  ( $G \in \mathbb{R}_{\geq 0}^{k \times m \times n}$ ) that represents the interactions between the  $t$ ,  $x$ , and  $y$  components of  $W(t)$ ,  $H(x)$ , and  $V(y)$ ; (ii) an unknown matrix  $W$  ( $W \in \mathbb{R}_{\geq 0}^{K \times k}$ ) representing the changes of the  $X$  in time (the time-component); (iii) an unknown matrix  $H$  ( $H \in \mathbb{R}_{\geq 0}^{M \times m}$ ) representing the changes of  $X$  in x-direction (the x-component), and (iv) an unknown matrix  $V$  ( $V \in \mathbb{R}_{\geq 0}^{N \times n}$ ) representing the changes of  $X$  in y-direction (the y-component). Here,  $\mathbb{R}_{\geq 0}$  denotes the set of non-negative real numbers  $\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} \mid x \geq 0\}$ . Additionally,  $\epsilon$  ( $\epsilon \in \mathbb{R}^{K \times M \times N}$ ) denotes the unknown discrepancy between the original data  $X(t, x, y)$  and estimate  $\tilde{X}(t, x, y)$ ; the discrepancy is caused by presence of a noise or errors in the measurements. If the Tucker decomposition is successful,  $\epsilon$  should represent a white noise. An important part of our NTF $k$  reconstruction analyses is to make sure that  $\epsilon$  characterizes a white noise and there additional signals that are not extracted by our procedure.

Mathematically, the solution of the non-negative Tucker-3 tensor decomposition is a solution of a complex multi-dimensional optimization with non-negative constraints given by:

$$\min_{G, W, H, V \geq 0} \|X - G \otimes W \otimes H \otimes V\|^2 \quad (2.16)$$

To extract the unknown core-tensor  $G$ , and the factor matrices  $W$ ,  $H$ , and  $V$ , we utilize the block coordinate descent method for regularized multi-convex optimization algorithms introduced by Xu et.al. [51, 52] and programmed in Julia [53]. In the analyses below, we focus on the temporal features  $W$ , i.e., we estimate the robustness of the factorization related to the temporal factor  $W$ . To identify the optimal number of features (i.e. the optimal number of columns of  $W$ ), we perform ‘ $N$ ’ NTF runs with random initial guesses for the unknown parameters, and then cluster the resulting sets of columns of the factors  $W_1^k, W_2^k, \dots, W_N^k$ , where  $k$  is the number of features used in this set of simulations. We use a custom clustering algorithm for this purpose, which is based on  $k$ -means clustering and described by [54] and [55].  $k$ -means is one of the simplest unsupervised learning algorithms that solve the clustering problem. The procedure classifies every single dataset entry to

belong to a certain number of clusters fixed apriori based on minimization of a user-defined objective function (in our case, the objective functions is based on cosine distances; [54], [55]). The optimal number of features is evaluated by comparison of the quality of the reconstruction,  $R$ , based on the Frobenius norm (Eq. 2.16):  $R = \|X - \tilde{X}\|/\|X\|$  and the quality of the derived clusters for different number of features,  $k$ , estimated by their average Silhouettes [56]. Silhouette is a measure of how similar a given single dataset entry is to its own cluster compared to other clusters. The silhouette ranges from 1 to +1, where values close to +1 indicates that the dataset entry is well matched to its own cluster and poorly matched to neighboring clusters. High average Silhouette values suggest that the clustering configuration is appropriate. Low average Silhouette values suggest that the clustering configuration is not optimal; potentially, there are too many or too few clusters.

### 3. Results

The NTF $k$  methodology was applied to analyze more than 2000 simulations predicting the concentrations of product  $C$  using the methodology outlined in Section 2.1. In our case,  $K = 1000$ ,  $M = 81$ , and  $N = 81$ . For  $\epsilon \approx \mathcal{O}(10^{-9})$ , the size of the obtained core tensor  $G$ , which is given in terms of  $k$ ,  $m$ , and  $n$ , vary from 2 to 4, 2 to 14, and 3 to 15, respectively. Each numerical simulation is obtained for a different set of reaction-diffusion model input parameters. The varied input parameters are: perturbation parameter of the underlying vortex-based velocity field  $v_o$ , longitudinal-to-transverse anisotropic dispersion ratio  $\frac{\alpha_L}{\alpha_T}$ , molecular diffusivity  $D_m$ , and velocity field characteristic scales  $\kappa_f L$  and  $T$ . The corresponding values for the varied input parameters are:  $v_o = [1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}]$ ,  $\frac{\alpha_L}{\alpha_T} = [1, 10^1, 10^2, 10^3, 10^4]$ ,  $D_m = [10^{-8}, 10^{-1}, 10^{-2}, 10^{-3}]$ ,  $\kappa_f L = [2, 3, 4, 5]$ , and  $T = [1 \times 10^{-4}, 2 \times 10^{-4}, 3 \times 10^{-4}, 4 \times 10^{-4}, 5 \times 10^{-4}]$ . In our simulations,  $\alpha_L = 1$  and  $\alpha_T$  is varied accordingly. There is no existing methodology or closed-form mathematical solutions that can predict how the variation of these input parameters impacts the concentration of product  $C$ . The goal of the proposed unsupervised ML analyses of the simulation outputs is to provide insights about the influences of the input parameters on species mixing, decay of reactants, and formation of the reaction product.

The output from each numerical simulation is a tensor with dimensions  $1000 \times 81 \times 81$  representing the concentration of product  $C$  in time and space. That is, there are 1000 time steps for computational times ranging from 0.001 to 1.0 with a uniform time step equal to 0.001. The two-dimensional computational domain is discretized by low-order structured triangular finite elements with 81 nodes on each side. The resulting finite element mesh is of size  $81 \times 81$ . Fig. 4 shows the concentration contours of product  $C$  at various times for a case when anisotropic dispersion is very high and molecular diffusivity is very low. The input parameter that is varied is  $\kappa_f L$  and other parameters such as  $\frac{\alpha_L}{\alpha_T}$ ,  $v_o$ ,  $T$ , and  $D_m$  are kept constant. Small values of  $\kappa_f L$  correspond to large-scale vortex structures and large values of  $\kappa_f L$  correspond to small-scale vortex structures in the velocity field. From this figure, one can infer that the small-scale vortex structures present in the velocity field enhance mixing even when the value of molecular diffusivity is small and anisotropic dispersivity is large.

Example results obtained from the NTF $k$  methodology are presented in Figs. 5–12. Fig. 5 provides a comparison of the ground truth concentrations with that estimated by the NTF $k$  methodology. Ground truth is a high-fidelity numerical simulation with input parameters as:  $v_o = 10^{-3}$ ,  $\frac{\alpha_L}{\alpha_T} = 10^4$ ,  $D_m = 10^{-3}$ ,  $\kappa_f L = 3$ , and  $T = 1 \times 10^{-4}$ . It is clear that the solution obtained from

---

**Algorithm 1** Overview of NTFk-based structure-preserving feature extraction framework for reactive-mixing
 

---

- 1: Input: Time step  $\Delta t$ ; total time of interest  $\mathcal{I}$ ; stoichiometric coefficients; initial and boundary conditions for the chemical species  $A$ ,  $B$ , and  $C$ ; finite element mesh parameters; model parameters for anisotropic reaction-diffusion equation.
- 2: Calculate initial and boundary conditions for the non-negative invariants  $c_F$  and  $c_G$  using Eqs. (2.3a)–(2.3b).
- 3: Solve for invariant concentrations  $c_F$  and  $c_G$  for all times.
- 4: **for**  $n = 0, 1, \dots, (\text{NumTimeSteps} - 1)$  **do**
- 5: Call optimization-based diffusion with decay solver to obtain  $\mathbf{c}_F^{(n+1)}$ :

$$\begin{aligned} & \underset{\mathbf{c}_F^{(n+1)} \in \mathbb{R}^{ndofs}}{\text{minimize}} && \frac{1}{2} \langle \mathbf{c}_F^{(n+1)}; \mathbf{K} \mathbf{c}_F^{(n+1)} \rangle - \langle \mathbf{c}_F^{(n+1)}; \mathbf{f}_F^{(n+1)} \rangle - \frac{1}{\Delta t} \langle \mathbf{c}_F^{(n+1)}; \mathbf{c}_F^{(n)} \rangle \\ & \text{subject to} && \mathbf{c}_F^{\min} \mathbf{1} \preceq \mathbf{c}_F^{(n+1)} \preceq \mathbf{c}_F^{\max} \mathbf{1} \end{aligned}$$

- 6: Call optimization-based diffusion with decay solver to obtain  $\mathbf{c}_G^{(n+1)}$ :

$$\begin{aligned} & \underset{\mathbf{c}_G^{(n+1)} \in \mathbb{R}^{ndofs}}{\text{minimize}} && \frac{1}{2} \langle \mathbf{c}_G^{(n+1)}; \mathbf{K} \mathbf{c}_G^{(n+1)} \rangle - \langle \mathbf{c}_G^{(n+1)}; \mathbf{f}_G^{(n+1)} \rangle - \frac{1}{\Delta t} \langle \mathbf{c}_G^{(n+1)}; \mathbf{c}_G^{(n)} \rangle \\ & \text{subject to} && \mathbf{c}_G^{\min} \mathbf{1} \preceq \mathbf{c}_G^{(n+1)} \preceq \mathbf{c}_G^{\max} \mathbf{1} \end{aligned}$$

- 7: where  $\mathbf{K}$  is the symmetric positive definite coefficient matrix,  $ndofs$  is the number of degrees of freedom,  $\langle \cdot; \cdot \rangle$  represents the standard inner-product on Euclidean spaces,  $\mathbf{1}$  denotes a vector of ones of size  $ndofs \times 1$ , and the symbol  $\preceq$  represents the component-wise inequality for vectors.

8: **end for**

- 9: Get non-negative solutions for  $c_A$ ,  $c_B$ , and  $c_C$  using Eqs. (2.6a)–(2.6c).

- 10: Decompose  $c_F$  and  $c_G$  using Eq. (2.16).

$$\begin{aligned} & \bullet c_F(\mathbf{x}, t) = c_F(\mathbf{x}, t)|_{T1} + c_F(\mathbf{x}, t)|_{T2} + c_F(\mathbf{x}, t)|_{T3} \quad \forall (\mathbf{x}, t) \in \bar{\Omega} \times [0, \mathcal{I}] \\ & \bullet c_G(\mathbf{x}, t) = c_G(\mathbf{x}, t)|_{T1} + c_G(\mathbf{x}, t)|_{T2} + c_G(\mathbf{x}, t)|_{T3} \quad \forall (\mathbf{x}, t) \in \bar{\Omega} \times [0, \mathcal{I}] \end{aligned}$$

where T1, T2, and T3 are temporal features identified by NTFk method.

- 11: Using non-linear transformations given by Eqs. (2.6a)–(2.6c), obtain the hidden features of  $c_C$ .

Based on Eq. (2.6a), as  $c_A \leq c_F \quad \forall t \in [0, \mathcal{I}]$ , we have

$$\begin{aligned} & \bullet c_A(\mathbf{x}, t) = \underbrace{\max \left[ c_F - \left( \frac{n_A}{n_B} \right) c_G, 0 \right]}_{c_A|_{T1}} \Big|_{T1} + \underbrace{\max \left[ c_F - \left( \frac{n_A}{n_B} \right) c_G, 0 \right]}_{c_A|_{T2}} \Big|_{T2} + \underbrace{\max \left[ c_F - \left( \frac{n_A}{n_B} \right) c_G, 0 \right]}_{c_A|_{T3}} \Big|_{T3} \\ & \bullet c_C(\mathbf{x}, t) = \left( \frac{n_C}{n_A} \right) ((c_F - c_A)|_{T1} + (c_F - c_A)|_{T2} + (c_F - c_A)|_{T3}) \quad \forall (\mathbf{x}, t) \in \bar{\Omega} \times [0, \mathcal{I}] \end{aligned}$$

- 12: Physically, these hidden features correspond to formation of  $c_C$  due to longitudinal dispersion coefficient, transverse dispersion coefficient, and molecular diffusivity. That is,

- T1  $\Rightarrow$  mixing of reactants due to longitudinal dispersion coefficient.
- T2  $\Rightarrow$  mixing of reactants due to transverse dispersion coefficient.
- T3  $\Rightarrow$  mixing of reactants due to molecular diffusivity.

- 13: Compare the decomposed features with the ground truth (for example, see Fig. 5).
-

the proposed method compares well with that of the high-fidelity finite element numerical simulation. Fig. 5 also shows a linear decomposition of concentration of product  $C$  at various stages in time. One can see that at earlier stages in time ( $t = 0.1$ ), streamline mixing and mixing due to anisotropic dispersion are dominant. Mixing due to molecular diffusion is close to zero, which is expected. At later stages in time ( $t = 0.5$  and  $t = 1.0$ ), one can see that concentration of product  $C$  due to streamline mixing and anisotropic dispersion gradually decrease to zero and mixing due to molecular diffusion dominates.

Fig. 6 shows another validation of the proposed NTFk. It compares the true mean concentration of product  $C$  obtained from the finite element simulation to that of the estimated mean concentration based on NTFk. From Fig. 6, it is clear that estimated mean agrees well with true mean concentration. The figure also shows the mean concentrations associated with the NTFk estimated temporal components in the dataset. The sum of the mean temporal-components concentrations over time reproduces the total NTFk estimated mean concentrations (the brown curves in Fig. 6).

As discussed in Section 2, the NTFk predicted concentration are obtained based on products of the core tensor and the matrix factors (Eq.2.12; Fig.3). The temporal matrix components of  $W$  are visualized in Fig. 7. The  $W$  components drive the transients in the NTFk predicted concentrations. As expected, the overall shapes of the curves in Fig. 7 are similar to those in Fig. 6.

Figs. 8–12 show the maximum concentration of product  $C$  predicted by each temporal feature identified by our methodology for each time snapshot (there are 1000 snapshots for times between 0.001 to 1.0). It is important to note that the ML-predicted maximum concentration in some cases exceeds 0.5 (the true maximum concentration of product  $C$  in the input data; for example, see Fig. 8a) due to estimation errors. This estimation error can be reduced by imposing another constraint that the maximum concentration of product  $C$  cannot exceed 0.5. In most cases, the proposed NTFk methodology identified three temporal features. These are labeled as T1, T2, and T3 in Figs. 8–12. However, in one case, there are only two features (see Fig. 11a) while in another there are four features (8d). The first temporal feature T1 is dominated by species mixing due to longitudinal dispersion coefficient. The process of reactive-mixing due to longitudinal dispersion is much faster than mixing of reactants due to transverse dispersion and molecular diffusion. As a result, we see a sharp increase in the concentration of product  $C$  at early times (for example, see Fig. 8a). The second temporal feature T2 corresponds to species mixing due to transverse dispersion coefficient. Finally, the third temporal feature T3 represents species mixing due to molecular diffusion. The shape of the T3 curve is close to linear when plotted as a square root of time which is consistent with pure isotropic molecular diffusion process. The above inferred observations are valid for most of the results discussed below.

Fig. 8 demonstrates how the perturbations in vortex structures (as dictated by  $v_o$  values) impacts the identified temporal features. Figs. 8a–8d depict changes in the maximum concentration in product  $C$  due to varying  $v_o$  by keeping other parameters fixed, which are  $T = 1 \times 10^{-4}$ ,  $\frac{\alpha_L}{\alpha_T} = 10^4$ ,  $D_m = 10^{-3}$ , and  $\kappa_f L = 3$ .  $\frac{\alpha_L}{\alpha_T} = 10^4$  corresponds to high anisotropic dispersion,  $D_m = 10^{-3}$  corresponds to low molecular diffusivity, and  $\kappa_f L = 3$  corresponds to medium-scale vortex structures present in the velocity field. As evident from Eqs. (2.9) and (2.10),  $v_o = 10^{-4}$  corresponds to low deviation and  $v_o = 1$  corresponds to high deviation from the symmetric vortex structure of the velocity field. For low values of  $v_o$  ( $10^{-4}$ ,  $10^{-3}$ , and  $10^{-2}$ ), one can observe that the maximum concentration of temporal feature T1 rises and decays sharply. However, for high values of  $v_o$  (for example,  $v_o = 1$ ), maximum concentration of temporal feature T1 decays slowly. This is because



temporal feature T1 corresponds to formation of product  $C$  along the vortex streamlines (due to longitudinal anisotropic dispersion coefficient). The non-symmetric vortex structures (caused due to higher values of  $v_o$ ) enhance mixing of reactants resulting in formation of product  $C$  even at longer times. This is not the case for lower values of  $v_o$  because of the symmetric vortex structures present in the velocity field. For low values of  $v_o$  and based on temporal feature T2, one can observe that mixing across the streamlines starts to decay once the maximum concentration of the first temporal feature T1 reaches zero. This is due to the fact that the anisotropic contrast is very high, which hinders reactants  $A$  and  $B$  from dispersing across the streamlines. As a result, for temporal feature T2, we see a decrease in the maximum concentration of product  $C$  as time progresses. For higher values of  $v_o$ , we also observe decay in the maximum concentration of temporal feature T2. But at larger times (in feature T2), the maximum concentration of product  $C$  is still non-zero, which is not the case for lower values of  $v_o$ . This is because the non-symmetric nature of the vortex-based velocity field brings the unused reactants near to each other when compared to the symmetric case. As a result, for longer times (even under high anisotropy) one can see formation of product  $C$  for higher values of  $v_o$ . The comparison of Figs. 8a–8c demonstrates that similar temporal features are identified regardless of the differences in  $v_o$ . This observation fails only at high values of  $v_o$  as shown in Fig. 8d. In this case, the proposed NTFk methodology identifies a fourth temporal feature T4. This fourth temporal feature (T4) is caused by boundary effects, which are brought on by high values of  $v_o$ .

Fig. 9 shows the impact of transverse anisotropic dispersion on the identified temporal features. On comparison with the results for  $v_o$  (see Fig. 8), the identified temporal features obtained for different values of  $\frac{\alpha L}{\alpha T}$  are substantially different. With the increase in anisotropic dispersion ratio, it is clear that both the first and second temporal feature (T1 & T2) become less important at later times, and the third temporal feature (T3) becomes more dominant. This observation is in accordance with the anisotropic reaction-diffusion process as high anisotropic contrast hinders the transverse dispersion resulting in incomplete mixing of reactants. At later times, molecular diffusion becomes the major mechanism to enhance mixing between reactants  $A$  and  $B$  resulting in higher yield of product  $C$ .

Fig. 10 shows the effects of the molecular diffusivity  $D_m$  on the identified temporal features. From this figure, it is clear that the increase in the values of  $D_m$  shows an increase in the maximum concentration of T3 at later times. This is expected because the temporal feature T3 is dominated by the molecular diffusion.

Fig. 11 presents the importance of the spatial characteristic scale  $\kappa_f L$  on the identified temporal features. This parameter provides us information on small-scale and large-scale nature of vortex structures present in the velocity field. For small values of  $\kappa_f L$ , we have large-scale vortex structures and for bigger values of  $\kappa_f L$ , we have small-scale vortex structures. As  $\kappa_f L$  increases, the number of vortices present in the velocity field increases resulting in enhance mixing between reactant  $A$  and  $B$ . Moreover, it also affects the number of identified temporal features. In the case for  $\kappa_f L = 2$ , the NTFk methodology identifies only two temporal features (see Fig. 11a). The proposed method, could not separate the reactive-mixing occurring due to vortex structures and anisotropic dispersion. For this value of  $\kappa_f L$ , product formation due to mixing by longitudinal dispersion and transverse dispersion are lumped together as a single feature T1. However, for values of  $\kappa_f L > 2$  (Figs. 11b and 11c), both of these aspects are distinguished. It is important to note that as  $\kappa_f L$  increases, the first temporal features T1 remains dominant throughout the entire simulation time. Moreover,

T1 does not diminish as compared to other cases presented here (for example, see Figs. 10, 8, and 9) even when the anisotropic contrast is very high. This is because of the high number of vortex structures present in the velocity field, which increases the reactive-mixing even at the later times.

Fig. 12 shows the impact of flipping the vortex-based velocity field from clockwise direction to anti-clockwise direction (see Eq. 2.8) on the trend observed in formation of product  $C$ . Flipping the vortex-based velocity field can be associated with stirring rate [2]. Low values of  $T$  correspond to fast flipping of the velocity field from clockwise direction to anti-clockwise direction while high values of  $T$  correspond to slow flipping. For low values of  $T$  ( $1 \times 10^{-4}$  and  $2 \times 10^{-4}$ ), we observe formation of product  $C$  even at later times (see Figs. 12a and 12b). This is due to reactive-mixing along the streamlines due to longitudinal dispersion, which is not observed at higher values of  $T$  (see Figs. 12c and 12d). The reason is that fast flipping brings more reactants together resulting in enhanced mixing. The proposed NTF $k$  methodology is able to identify and delineate the above aspects. For example, the identified temporal features between the scenarios when  $T \leq 2 \times 10^{-4}$  (see Figs. 12a and 12b) and  $T > 2 \times 10^{-4}$  (see Figs. 12c and 12d) are distinct. A comparison of the results in Fig. 12 demonstrates that for  $T > 2 \times 10^{-4}$  the temporal feature T1 is dominant only at very early times where as temporal features T2 & T3 become more dominant at intermediate and late times. This is not the case when  $T \leq 2 \times 10^{-4}$  due the fast flipping of vortex-based velocity field.

All the high-resolution numerical simulations performed using the non-negative finite element method are about 200 GB of data. The tensor decomposition of these datasets using the proposed NTF $k$  methodology has produced a compressed dataset of about 70 MB. This corresponds to a compression ratio close to  $4 \times 10^{-4}$ . In addition to identifying hidden temporal features, the data compression is an added benefit of Tucker-3 NTF $k$  model. To summarize, the proposed NTF $k$  method substantially reduces the data storage requirements with minimal loss of information and provides accelerated data processing.

## 4. Conclusions

Our results demonstrate the applicability of the NTF $k$  approach to analyze complex large-scale model outputs of anisotropic reaction-diffusion equations. NTF $k$  is based on a Non-negative Tensor Factorization (NTF) combined with a custom semi-supervised clustering [54, 55]. NTF $k$  allows for the deconstruction of tensor datasets (multi-dimensional arrays) into additive components. The separation is based on the temporal and spatial variability of the data. In this paper, we applied NTF $k$  to identify between 2 and 4 temporal features (components) in model simulation outputs representing the impact of species mixing in a fast irreversible bimolecular reaction  $A + B \rightarrow C$ . Five model parameters representing physical processes such as anisotropic dispersion contrast, molecular diffusivity, perturbations in velocity field as well as characteristic spatial and time-scales of the velocity field, are varied in our finite element numerical simulations to evaluate the effect of these parameters on formation of product  $C$  in space and time. The features identified by NTF $k$  have physical meaning and enable the interpretation of the governing process impacting the model outputs. The identified features appear to be dominated by vortex structure, anisotropic dispersion, molecular diffusion, and boundary effects. The analyzed species mixing problem is challenging but important for various real world applications (e.g., contaminant remediation, material design, etc.). These ML results can play a critical role in improving our conceptual understanding of these reaction-diffusion processes. Our NTF $k$  work also acts as data compression of the numerical



simulations and could streamline the development of reduced-order models that can be applied to predict the system behavior in a more computationally efficient manner.

The application of the proposed NTF $k$  approach is not limited to reactive-mixing. NTF $k$  can readily be applied to any observed or simulated datasets that can be represented as a tensor (multi-dimensional array) and have separable latent signatures or features. Frequently, these signatures/features are related to underlying physical processes governing the spatial and temporal variability in the data. Extracting these signatures and structure-preserving features can help the development of conceptual models, reduced-order models, and simplified closed-form mathematical expressions, which can then be used to predict the system behavior. Furthermore, NTF $k$  analyses can be applied to detect anomalies or disruptions in the data, which might be caused by additional phenomena not detected in past. Lastly, unsupervised machine learning methods such as NTF $k$  can be coupled with supervised machine learning methods to perform deep learning.

### Acknowledgments

This research was funded by the Environmental Programs Directorate of the Los Alamos National Laboratory. In addition, Velimir V. Vesselinov and Daniel O'Malley were supported by the DiaMonD project (An Integrated Multifaceted Approach to Mathematics at the Interfaces of Data, Models, and Decisions, U.S. Department of Energy Office of Science, Grant #11145687). Maruti Kumar Mudunuru and Satish Karra also thank the support of the LANL Laboratory Directed Research and Development Early Career Award 20150693ECR. Maruti Kumar Mudunuru also thanks the support of LANL Chick-Keller Postdoctoral Fellowship through Center for Space and Earth Sciences (CSES). Velimir V. Vesselinov, Boian S. Alexandrov, and Daniel O'Malley were also supported by LANL LDRD grant 20180060. This research used resources provided by the Los Alamos National Laboratory Institutional Computing Program, which is supported by the U.S. Department of Energy National Nuclear Security Administration under Contract No. DE-AC52-06NA25396.

### References

- [1] C. duP. Donaldson and G. R. Hilst. Effect of inhomogeneous mixing on atmospheric photochemical reactions. *Environmental Science and Technology*, 6:812–816, 1972.
- [2] J. Ottino. Mixing, chaotic advection, and turbulence. *Annual Review of Fluid Mechanics*, 22:207–253, 1990.
- [3] G. N. Ivey and J. Imberger. On the nature of turbulence in a stratified fluid. Part I: The energetics of mixing. *Journal of Physical Oceanography*, 21:650–658, 1991.
- [4] D. M. Imboden and A. Wüest. Mixing mechanisms in lakes. In *Physics and Chemistry of Lakes*, pages 83–138. Springer Berlin Heidelberg, Berlin, Heidelberg, 1995.
- [5] R. F. Ismagilov, A. D. Stroock, P. J. A. Kenis, G. Whitesides, and H. A. Stone. Experimental and theoretical scaling laws for transverse diffusive broadening in two-phase laminar flows in microchannels. *Applied Physics Letters*, 76:2376–2378, 2000.
- [6] M. K. Mudunuru, M. Shabouei, and K. B. Nakshatrala. On local and global species conservation errors for nonlinear ecological models and chemical reacting flows. In *Proceedings of ASME 2015 International Mechanical Engineering Congress and Exposition*, pages V009T12A018–V009T12A018, 2015.
- [7] M. K. Mudunuru and K. B. Nakshatrala. On mesh restrictions to satisfy comparison principles, maximum principles, and the non-negative constraint: Recent developments and new results. *Mechanics of Advanced Materials and Structures*, 24:556–590, 2017.
- [8] V. Hessel, H. Löwe, and F. Schönfeld. Micromixers—A review on passive and active mixing principles. *Chemical Engineering Science*, 60:2479–2501, 2005.

- [9] T. W. Willingham, C. J. Werth, and A. J. Valocchi. Evaluation of the effects of porous media structure on mixing-controlled reactions using pore-scale modeling and micromodel experiments. *Environmental Science and Technology*, 42:3185–3193, 2008.
- [10] W. Stumm and J. J. Morgan. *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters*. Wiley, New Jersey, USA, third edition, 2012.
- [11] L. W. Gelhar. *Stochastic Subsurface Hydrology*. Prentice-Hall, 1993.
- [12] C. W. Fetter, T. Boving, and D. Kremer. *Contaminant Hydrogeology*. Prentice-Hall, Illinois, USA, third edition, 2017.
- [13] A. Vengosh, R. B. Jackson, N. Warner, T. H. Darrah, and A. Kondash. A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States. *Environmental Science & Technology*, 48:8334–8348, 2014.
- [14] W. J. Deutsch and R. Siegel. *Groundwater Geochemistry: Fundamentals and Applications to Contamination*. CRC Press, Florida, USA, 1997.
- [15] B. J. Wagner. Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modelling. *Journal of Hydrology*, 135:275–303, 1992.
- [16] R. M. Neupauer, B. Borchers, and J. L. Wilson. Comparison of inverse methods for reconstructing the release history of a groundwater contamination source. *Water Resources Research*, 36:2469–2475, 2000.
- [17] J. Atmadja and A. C. Bagtzoglou. Pollution source identification in heterogeneous porous media. *Water Resources Research*, 37:2113–2125, 2001.
- [18] A. M. Michalak and P. K. Kitanidis. Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling. *Water Resources Research*, 40, W08302, doi:10.1029/2004WR003214, 2004.
- [19] J. Guan, M. M. Aral, M. L. Maslia, and W. M. Grayman. Identification of contaminant sources in water distribution systems using simulation-optimization method: Case study. *Journal of Water Resources Planning and Management*, 132:252–262, 2006.
- [20] A. V. Mamonov and Y. H. R. Tsai. Point source identification in nonlinear advection-diffusion-reaction systems. *Inverse Problems*, 29:035009, 2013.
- [21] A. Hamdi and I. Mahfoudhi. Inverse source problem in a one-dimensional evolution linear transport equation with spatially varying coefficients: Application to surface water pollution. *Inverse Problems in Science and Engineering*, 21:1007–1031, 2013.
- [22] J. M.-Bruce and P. L. Dragotti. Spatio-temporal sampling and reconstruction of diffusion fields induced by point sources. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 31–35, 2014.
- [23] V. T. Borukhov and G. M. Zayats. Identification of a time-dependent source term in nonlinear hyperbolic or parabolic heat equation. *International Journal of Heat and Mass Transfer*, 91:1106–1113, 2015.
- [24] J. M. Ottino and S. Wiggins. Introduction: Mixing in microfluidics. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, pages 923–935, 2004.
- [25] K. B. Nakshatrala, M. K. Mudunuru, and A. J. Valocchi. A numerical framework for diffusion-controlled bimolecular-reactive systems to enforce maximum principles and non-negative constraint. *Journal of Computational Physics*, 253:278–307, 2013.
- [26] J. Chang, S. Karra, and K. B. Nakshatrala. Large-scale optimization-based non-negative computational framework for diffusion equations: parallel implementation and performance studies. *Journal of Scientific Computing*, 70(1):243–271, 2017.
- [27] A. Adrover, S. Cerbelli, and M. Giona. A spectral approach to reaction/diffusion kinetics in chaotic flows. *Computers & Chemical Engineering*, 26:125–139, 2002.
- [28] Y. K. Tsang. Predicting the evolution of fast chemical reactions in chaotic flows. *Physical Review E*, 80:026305(8), 2009.
- [29] M. K. Mudunuru and K. B. Nakshatrala. On enforcing maximum principles and achieving element-wise species balance for advection-diffusion-reaction equations under the finite element method. *Journal of Computational Physics*, 305:448–493, 2016.

- [30] A. Cichocki, R. Zdunek, A. H. Phan, and S.-I. Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. John Wiley & Sons, West Sussex, UK, 2009.
- [31] M. Laassiri, E.-M. Hamzaoui, and R. C. E. Moursli. Application of Nonnegative Tensor Factorization for neutron-gamma discrimination of Monte Carlo simulated fission chamber’s output signals. *Results in Physics*, 7:1422–1426, 2017.
- [32] W. Austin, G. Ballard, and T. G. Kolda. Parallel tensor compression for large-scale scientific data. In *Parallel and Distributed Processing Symposium, IEEE International*, pages 912–922, 2016.
- [33] H.-G. Roos, M. Stynes, and L. Tobiska. *Numerical Methods for Singularly Perturbed Differential Equations*. Springer-Verlag, New York, USA, 1996.
- [34] M. Stynes. Numerical methods for convection–diffusion problems or the 30 years war. *arXiv:1306.5172*, 2013.
- [35] P. M. Gresho and R. L. Sani. *Incompressible Flow and the Finite Element Method: Advection-Diffusion*, volume 1. John Wiley & Sons, Inc., Chichester, UK, 2000.
- [36] Z. Mei. *Numerical Bifurcation Analysis for Reaction-Diffusion Equations*. Springer-Verlag, New York, USA, 2000.
- [37] G. F. Pinder and M. A. Celia. *Subsurface Hydrology*. John Wiley & Sons, Inc., New Jersey, USA, 2006.
- [38] I. T. Jolliffe. Principal component analysis and factor analysis. In *Principal Component Analysis*, Springer Series in Statistics, pages 115–128. Springer Science + Business Media, LLC, New York, USA, 1986.
- [39] S.-I. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems*, 8:757–763, 1996.
- [40] P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5:111–126, 1994.
- [41] S. Haykin and Z. Chen. The cocktail party problem. *Neural Computation*, 17:1875–1902, 2005.
- [42] F. L. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Studies in Applied Mathematics*, 6:164–189, 1927.
- [43] R. A. Harshman and M. E. Lundy. PARAFAC: Parallel factor analysis. *Computational Statistics & Data Analysis*, 18:39–72, 1994.
- [44] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21:1253–1278, 2000.
- [45] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.
- [46] C. A. Andersson and R. Bro. The N-way toolbox for MATLAB. *Chemometrics and Intelligent Laboratory Systems*, 52:1–4, 2000.
- [47] A. Gredilla, S. F.-O. De Vallejuelo, A. de Diego, J. M. Madariaga, and J. M. Amigo. Unsupervised pattern-recognition techniques to investigate metal pollution in estuaries. *TrAC Trends in Analytical Chemistry*, 46:59–69, 2013.
- [48] M. Mørup, L. K. Hansen, and S. M. Arnfred. Algorithms for sparse nonnegative Tucker decompositions. *Neural Computation*, 20:2112–2131, 2008.
- [49] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.
- [50] D. A. Ross and R. S. Zemel. Learning parts-based representations of data. *Journal of Machine Learning Research*, 7:2369–2397, 2006.
- [51] Y. Xu and W. Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 6:1758–1789, 2013.
- [52] Y. Xu. Alternating proximal gradient method for sparse nonnegative Tucker decomposition. *Mathematical Programming Computation*, 7:39–70, 2015.
- [53] J. Bezanson, S. Karpinski, V. B. Shah, and A. Edelman. Julia: A fast dynamic language for technical computing. *arXiv preprint arXiv:1209.5145*, 2012.
- [54] B. S. Alexandrov and V. V. Vesselinov. Blind source separation for groundwater pressure analysis based on nonnegative matrix factorization. *Water Resources Research*, 50:7332–7347, 2014.
- [55] V. V. Vesselinov, B. S. Alexandrov, and D. O’Malley. Contaminant source identification using semi-supervised machine learning. *Journal of Contaminant Hydrology*, 2017.
- [56] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

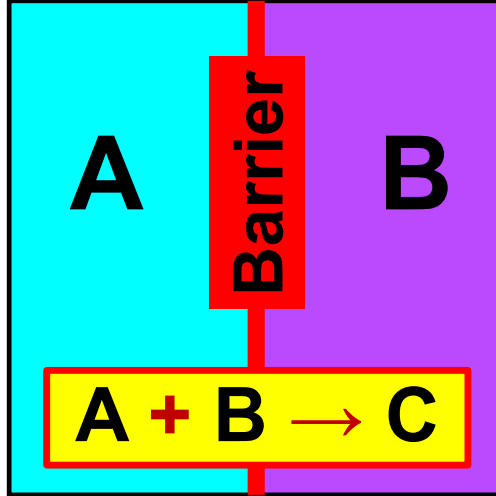


FIGURE 1. Schematic representation of the initial boundary value problem with zero flux boundary conditions along the domain boundaries. Initially species  $A$  and  $B$  are in the left and right part of the domain, respectively, and they are separated by a barrier at  $t = 0$ ;  $A$  and  $B$  are allowed to mix for time  $t > 0$  and form  $C$  driven by synthetic vortex-based velocity field defined in Eq. (2.9)–(2.10). The initial concentrations of  $A$  and  $B$  are equal to 1.0. The initial concentration of  $C$  is 0.0. The maximum plausible concentration of  $C$  is 0.5. There are no sources/sinks.

\*CORRESPONDING AUTHOR: DR. VELIMIR VALENTINOV VESSELINOV, COMPUTATIONAL EARTH SCIENCE GROUP (EES-16), MS T003 EARTH AND ENVIRONMENTAL SCIENCES DIVISION, LOS ALAMOS NATIONAL LABORATORY, LOS ALAMOS, NM 87545., E-MAIL ADDRESS: [vvv@lanl.gov](mailto:vvv@lanl.gov)

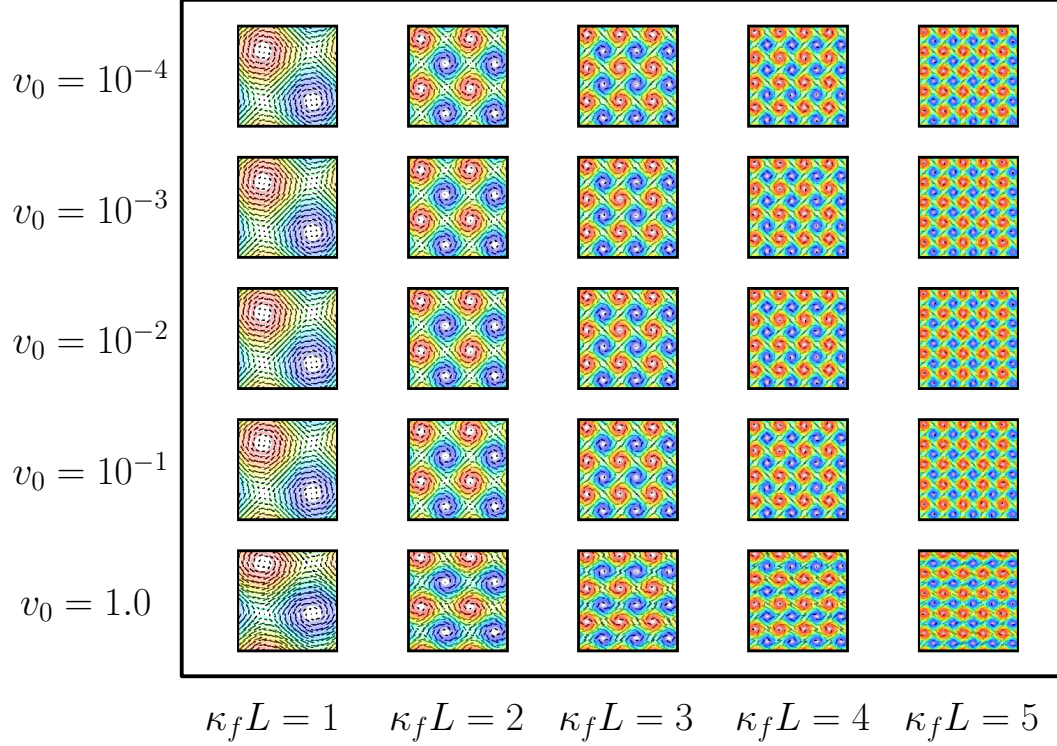


FIGURE 2. Streamlines and associated velocity vectors for different values of  $v_o$  and  $\kappa_f L$  representing large-scale and small-scale vortex structures present in the velocity field for  $\nu T \leq t < (\nu + \frac{1}{2}) T$ . The number of vortices increases for larger values of  $\kappa_f L$  and as a result there are more small-scale features present in the velocity field. Note that the parameter  $v_o$  does not substantially impact the location of vortices.

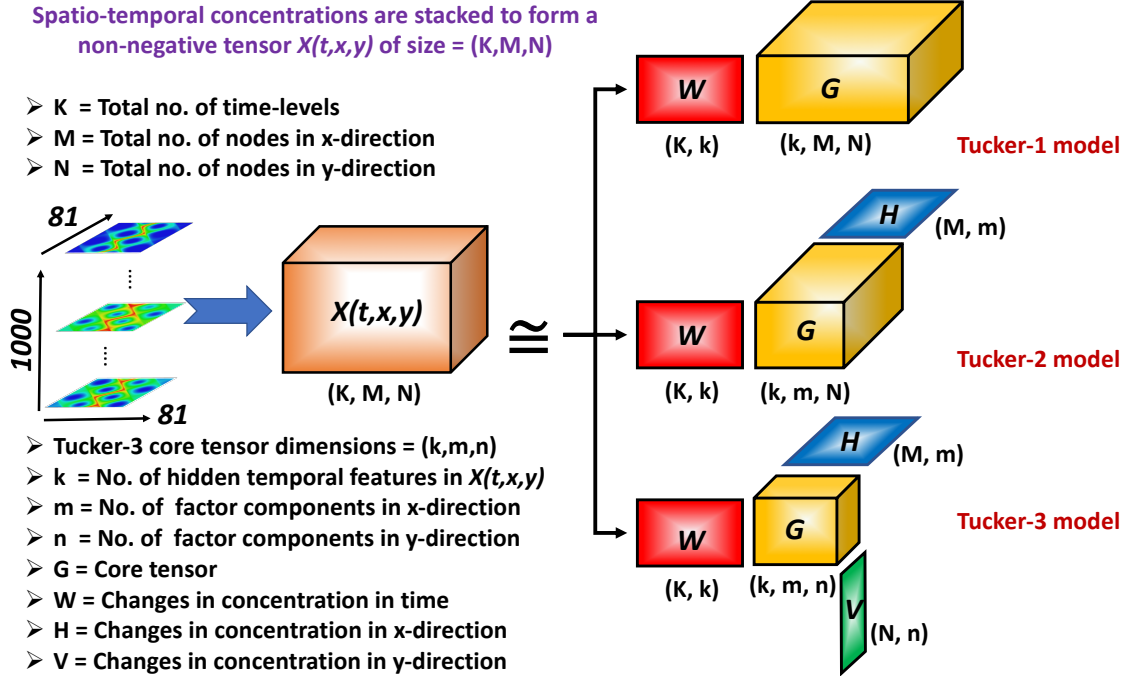


FIGURE 3. Schematic representation of various Tucker-based tensor factorization models. Herein, we employ Tucker-3 model to decompose the non-negative concentration tensor  $X(t, x, y)$  into a core tensor  $G$  and factors  $W$ ,  $V$ , and  $H$ .

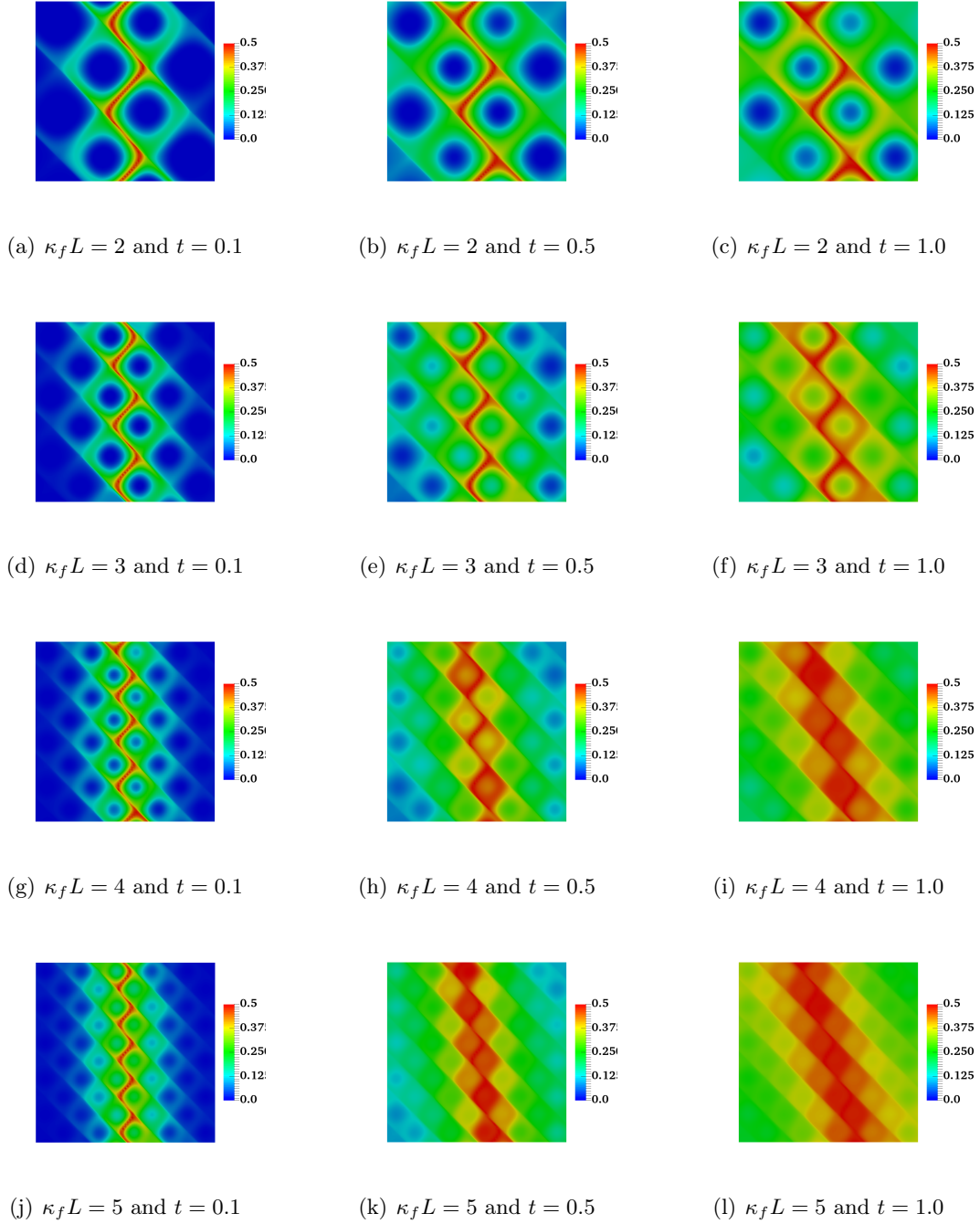


FIGURE 4. Concentration contours of product  $C$  at times  $t = 0.1, 0.5,$  and  $1.0$  for different  $\kappa_f L$  values. Other input parameters are  $\frac{\alpha L}{\alpha T} = 10^4$ ,  $v_o = 10^{-1}$ ,  $T = 1 \times 10^{-4}$ , and  $D_m = 10^{-3}$ . At early times and for smaller values of  $\kappa_f L$ , concentration of product  $C$  is zero in the vortex centers. At late times and as  $\kappa_f L$  increases, the regions with zero concentration of product  $C$  decrease.

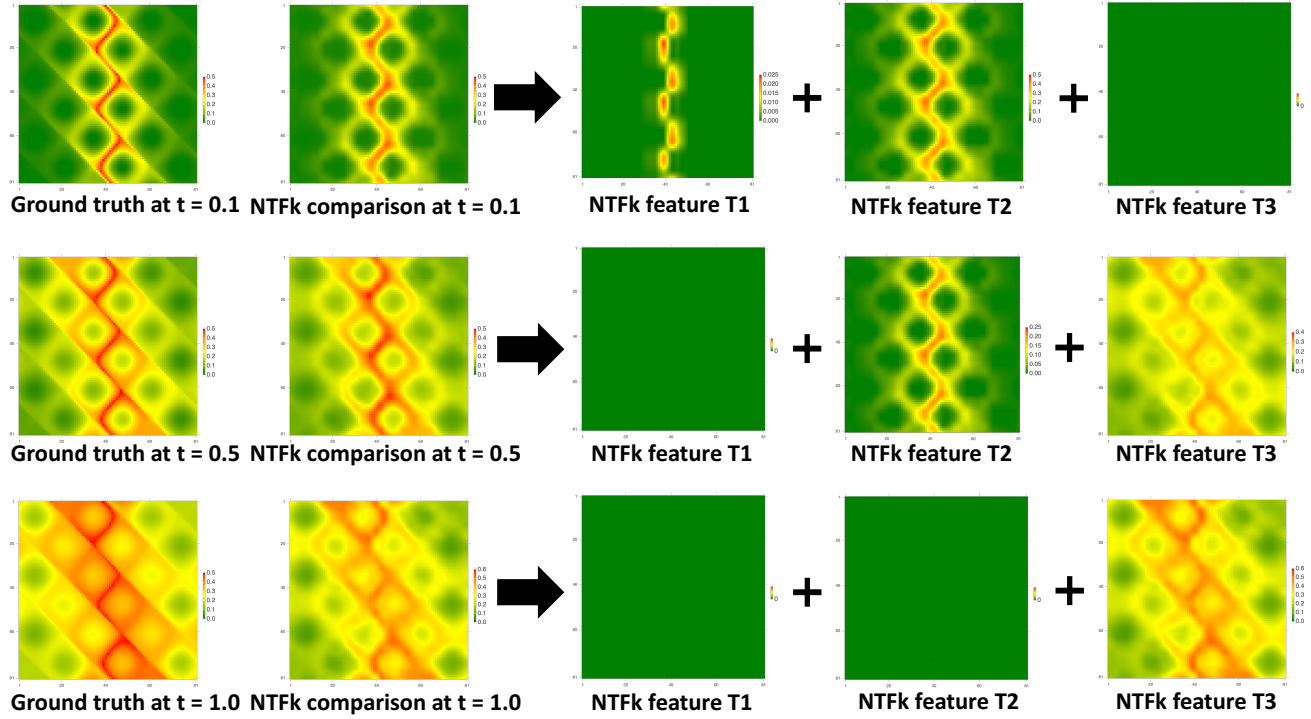
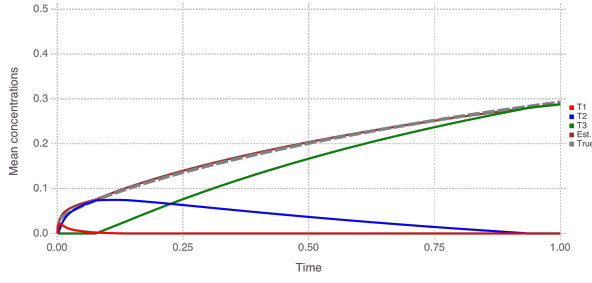
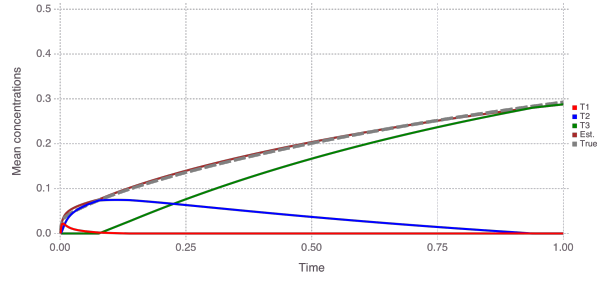


FIGURE 5. Comparison of ground truth and NTFk results at various times for product  $C$ . This figure also shows the decomposition of NTFk result into features T1, T2, and T3, which represent reactive-mixing due to longitudinal dispersion, transverse dispersion, and molecular diffusion, respectively. The input parameters corresponding to this particular case are:  $v_o = 10^{-3}$ ,  $\frac{\alpha_L}{\alpha_T} = 10^4$ ,  $D_m = 10^{-3}$ ,  $\kappa_f L = 3$ , and  $T = 1 \times 10^{-4}$ .  $v_o = 10^{-3}$  corresponds to small perturbations in underlying vortex-based velocity field.  $T = 1 \times 10^{-4}$  corresponds to fast flipping of vortex structures from clockwise direction to anti-clockwise direction.  $\frac{\alpha_L}{\alpha_T} = 10^4$  and  $D_m = 10^{-3}$  corresponds to high anisotropic dispersion and low molecular diffusivity.  $\kappa_f L = 3$  corresponds to medium-scale vortex structures present in the velocity field.

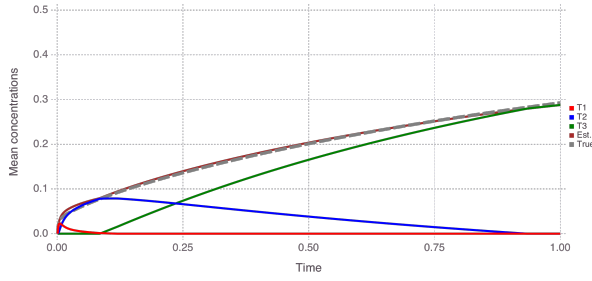




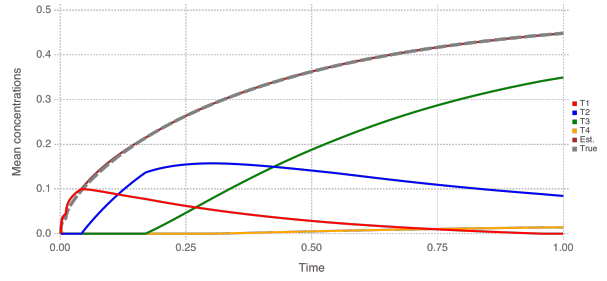
(a)  $v_o = 10^{-4}$



(b)  $v_o = 10^{-3}$

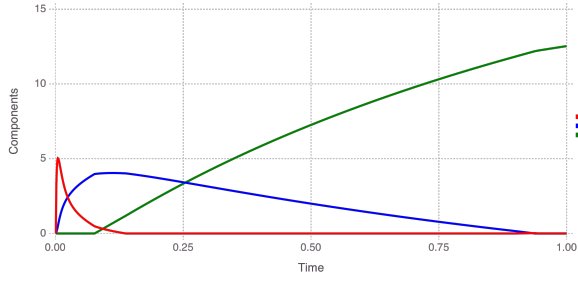


(c)  $v_o = 10^{-2}$

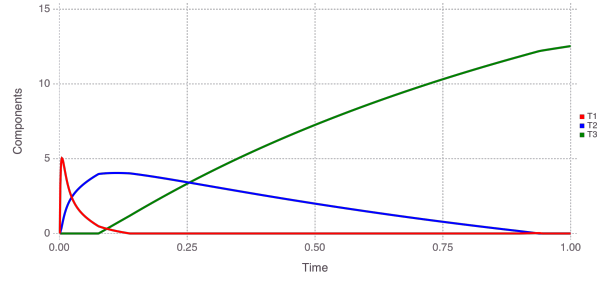


(d)  $v_o = 10^{-1}$

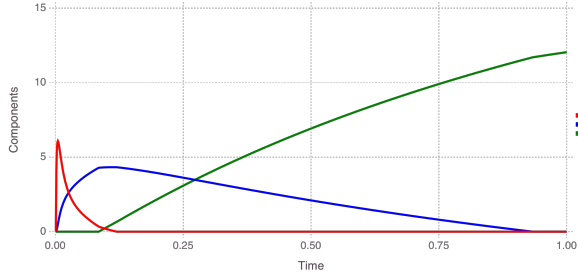
FIGURE 6. Transients in the estimated mean concentrations associated with the identified temporal features (T1, T2, T3 and T4) for a series of simulations representing varying  $v_o$  and keeping other parameters fixed ( $T = 1 \times 10^{-4}$ ,  $\frac{\alpha_L}{\alpha_T} = 10^4$ ,  $D_m = 10^{-3}$ ,  $\kappa_f L = 3$ ). The figures also show the estimated and the true mean concentrations throughout the spatial domain. The estimated mean concentrations are a sum of the mean concentrations associated with each temporal feature.



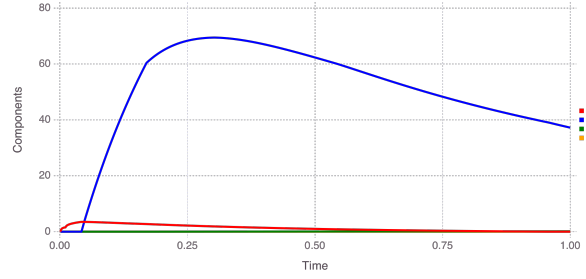
(a)  $v_o = 10^{-4}$



(b)  $v_o = 10^{-3}$

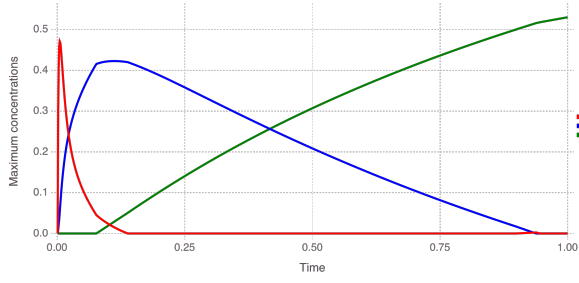


(c)  $v_o = 10^{-2}$

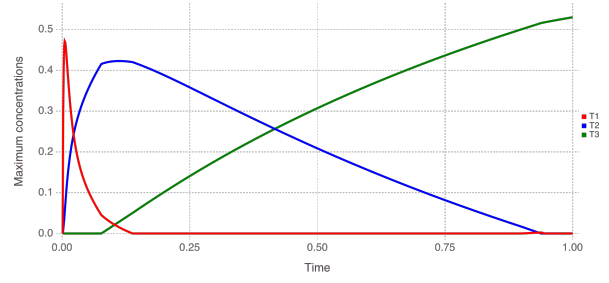


(d)  $v_o = 1$

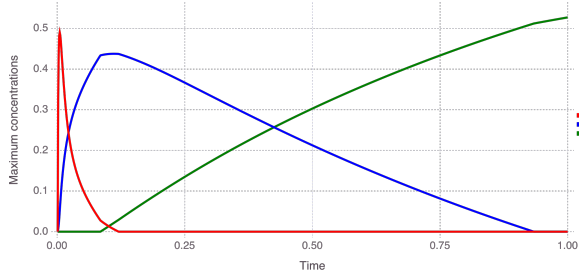
FIGURE 7. Components of the temporal matrix  $W$  in Eq.2.12. For consistency, the horizontal axis is presented in terms of time where row indices of  $W$  are mapped to the respective simulation times. The number of  $W$  columns define the number of temporal features. The temporal features (T1, T2, T3, and T4) are identified for a series of finite element simulations representing varying  $v_o$  and keeping other parameters fixed ( $T = 1 \times 10^{-4}$ ,  $\frac{\alpha_L}{\alpha_T} = 10^4$ ,  $D_m = 10^{-3}$ ,  $\kappa_f L = 3$ ).



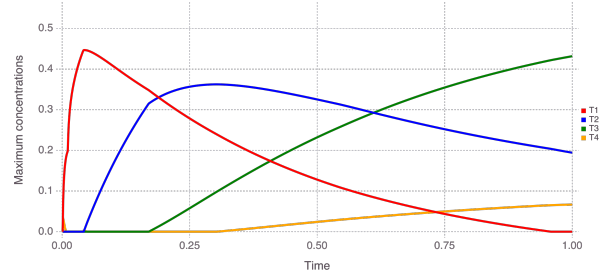
(a)  $v_o = 10^{-4}$



(b)  $v_o = 10^{-3}$



(c)  $v_o = 10^{-2}$



(d)  $v_o = 1$

FIGURE 8. Transients in the estimated maximum concentrations associated with the identified temporal features (T1, T2, T3, and T4) for a series of simulations representing varying  $v_o$  and keeping other parameters fixed ( $T = 1 \times 10^{-4}$ ,  $\frac{\alpha_L}{\alpha_T} = 10^4$ ,  $D_m = 10^{-3}$ ,  $\kappa_f L = 3$ ).  $T = 1 \times 10^{-4}$  corresponds to fast flipping of vortex structures as given by Eqs. (2.9)–(2.10).  $\frac{\alpha_L}{\alpha_T} = 10^4$  and  $D_m = 10^{-3}$  corresponds to high anisotropic dispersion and low molecular diffusivity.  $\kappa_f L = 3$  corresponds to medium-scale vortex structures present in the velocity field.

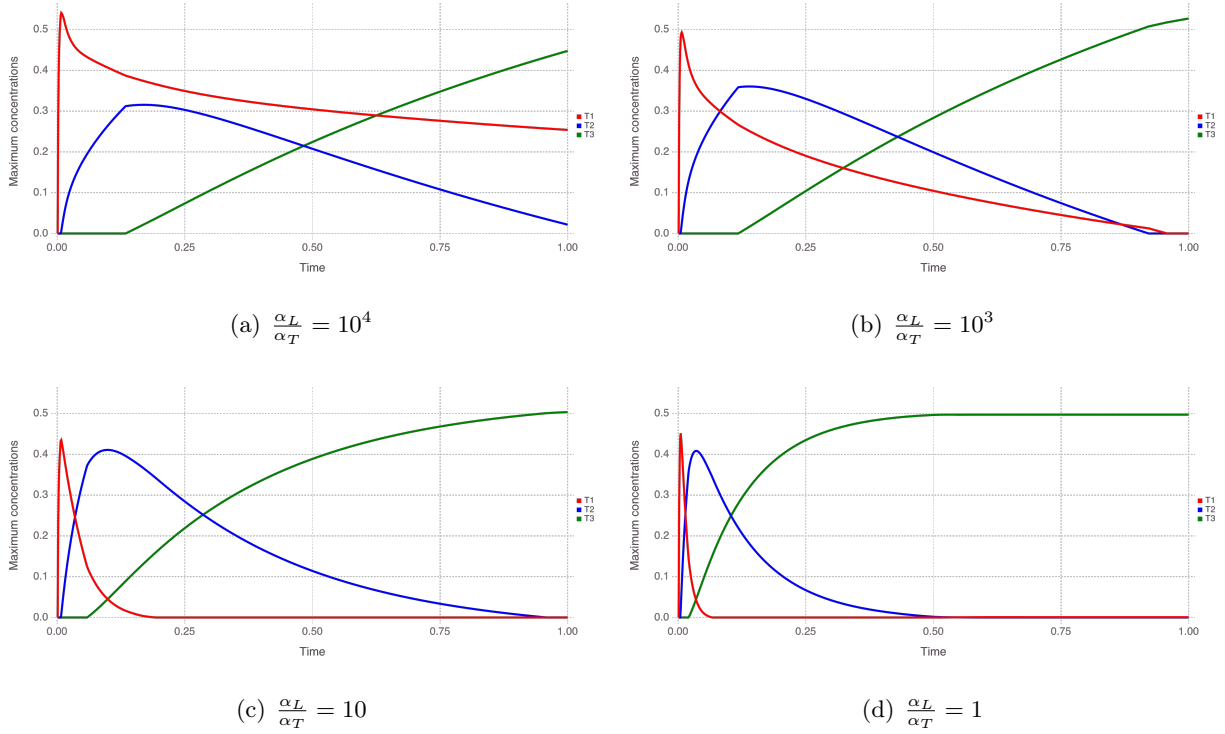
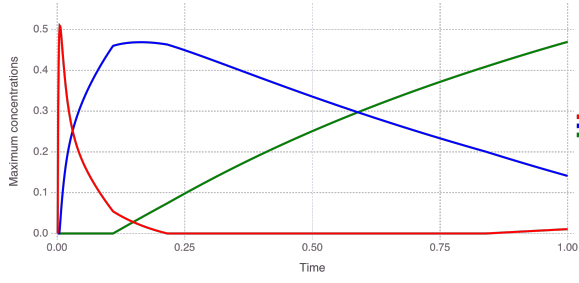
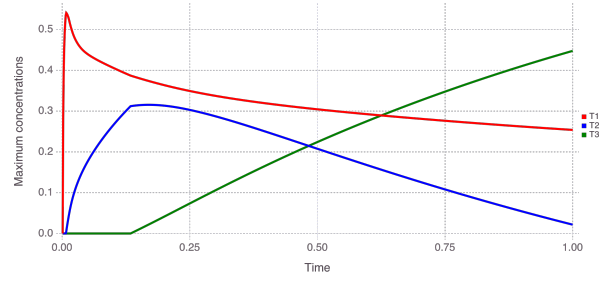


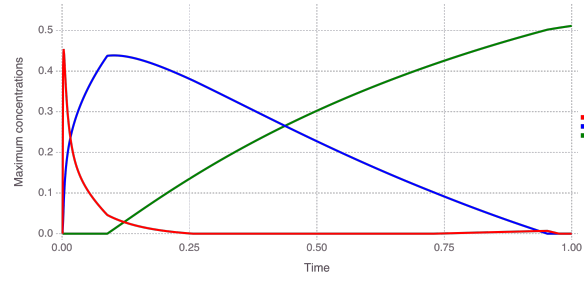
FIGURE 9. Transients in the estimated maximum concentrations associated with the identified temporal features (T1, T2, and T3) for a series of simulations representing varying anisotropic dispersion contrast  $\frac{\alpha_L}{\alpha_T}$  and keeping other parameters fixed ( $T = 1 \times 10^{-4}$ ,  $v_o = 10^{-1}$ ,  $D_m = 10^{-3}$ ,  $\kappa_f L = 3$ ).  $v_o = 10^{-3}$  corresponds to small perturbations in underlying vortex-based velocity field.  $\frac{\alpha_L}{\alpha_T} = 10^4$  and  $\frac{\alpha_L}{\alpha_T} = 1$  corresponds to high and low anisotropic dispersion.



(a)  $D_m = 10^{-8}$

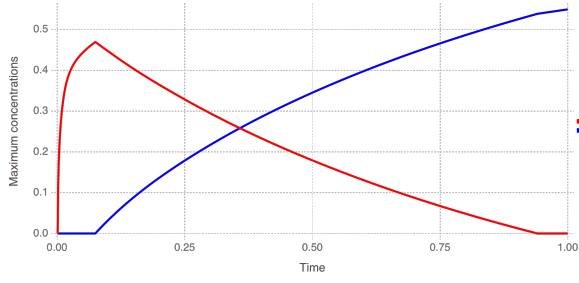


(b)  $D_m = 10^{-3}$

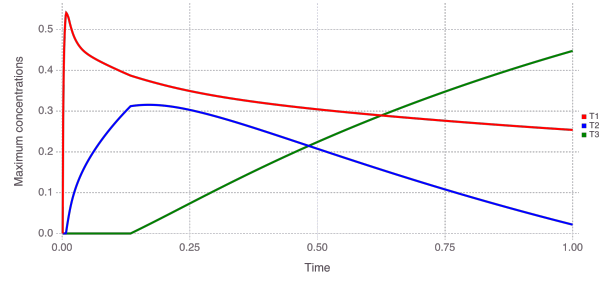


(c)  $D_m = 10^{-2}$

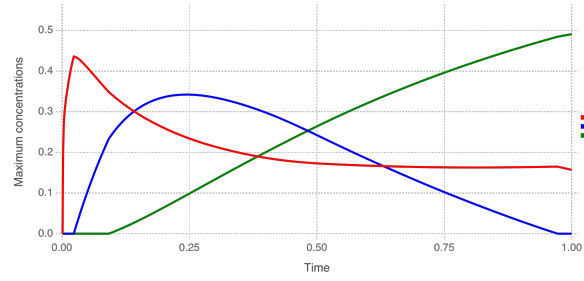
FIGURE 10. Transients in the estimated maximum concentrations associated with the identified temporal features (T1, T2, and T3). Analysis is performed for a series of simulations representing varying molecular diffusivity  $D_m$  and keeping other input parameters fixed ( $T = 1 \times 10^{-4}$ ,  $\frac{\alpha_L}{\alpha_T} = 10^4$ ,  $v_o = 10^{-1}$ ,  $\kappa_f L = 3$ ).



(a)  $\kappa_f L = 2$

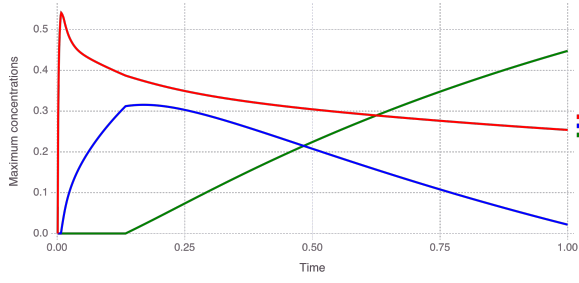


(b)  $\kappa_f L = 3$

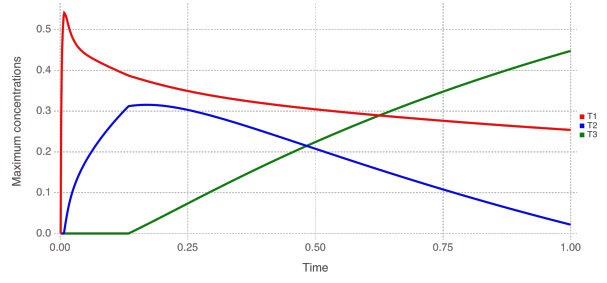


(c)  $\kappa_f L = 5$

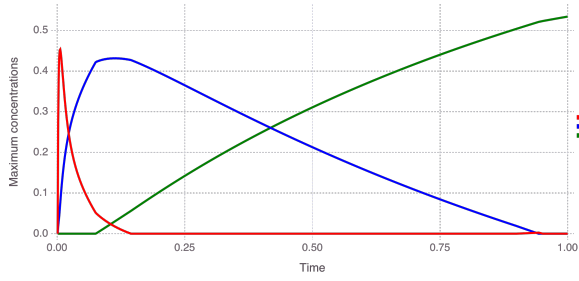
FIGURE 11. Transients in the estimated maximum concentrations associated with the identified temporal features (T1, T2, and T3). Analysis is performed for a series of simulations corresponding to varying spatial scales of underlying vortex-structures  $\kappa_f L$  and keeping other parameters fixed ( $T = 1 \times 10^{-4}$ ,  $\frac{\alpha L}{\alpha_T} = 10^4$ ,  $v_o = 10^{-1}$ ,  $D_m = 10^{-3}$ ).  $\kappa_f L = 2$ ,  $\kappa_f L = 3$ , and  $\kappa_f L = 4$  corresponds to large-scale, medium-scale, and small-scale vortex structures present in the synthetic velocity field.



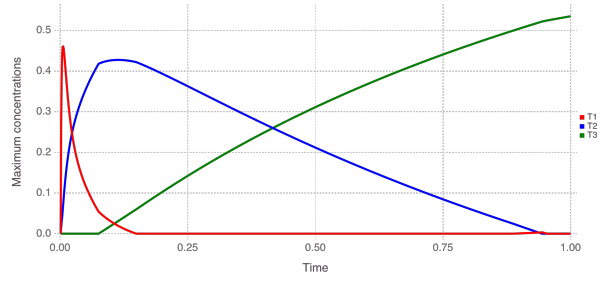
(a)  $T = 1 \times 10^{-4}$



(b)  $T = 2 \times 10^{-4}$



(c)  $T = 3 \times 10^{-4}$



(d)  $T = 4 \times 10^{-4}$

FIGURE 12. Transients in the estimated maximum concentrations associated with the identified temporal features for a series of simulations representing varying  $T$  and keeping other parameters fixed ( $\frac{\alpha L}{\alpha T} = 10^4$ ,  $v_o = 10^{-1}$ ,  $D_m = 10^{-3}$ ,  $\kappa_f L = 3$ ).  $T = 1 \times 10^{-4}$  and  $T = 4 \times 10^{-4}$  corresponds to fast flipping and slow flipping of vortex structures from clockwise direction to anti-clockwise direction.