



HAL
open science

Topological data assimilation using Wasserstein distance

Long Li, Arthur Vidard, François-Xavier Le Dimet, Jianwei Ma

► **To cite this version:**

Long Li, Arthur Vidard, François-Xavier Le Dimet, Jianwei Ma. Topological data assimilation using Wasserstein distance. *Inverse Problems*, 2019, 35 (1), pp.015006. 10.1088/1361-6420/aae993. hal-01960206

HAL Id: hal-01960206

<https://inria.hal.science/hal-01960206>

Submitted on 5 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Topological data assimilation using Wasserstein distance

Long Li^{1,2} Arthur Vidard² François-Xavier Le Dimet² Jianwei Ma¹

1,Department of Mathematics and Center of Geophysics, Harbin Institute of Technology, 150001 Harbin, China

2,Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

E-mail: jma@hit.edu.cn

July 2018

Abstract. This work combines a level-set approach and the optimal transport-based Wasserstein distance in a data assimilation framework. The primary motivation of this work is to reduce assimilation artifacts resulted from the position and observation error in the tracking and forecast of pollutants present on the surface of oceans or lakes. Both errors lead to spurious effect on the forecast that need to be corrected. In general, the geometric contour of such pollution can be retrieved from observation while more detailed characteristics such as concentration remain unknown. Herein, level sets are tools of choice to model such contours and the dynamical evolution of their topology structures. They are compared with contours extracted from observation using the Wasserstein distance. This allows to better capture position mismatches between both sources compared with the more classical Euclidean distance. Finally, the viability of this approach is demonstrated through academic test cases and its numerical performance is discussed.

1. Introduction

Numerical prediction of geophysical fluids requires a good knowledge of initial and boundary conditions, external forces, and internal physical parameters. Such information is indirectly accessible from observation and previous or external numerical simulations; however, they are partial and uncertain. Data assimilation (DA) is an efficient method for calibrating the state of dynamical equations by combining all the available heterogeneous information (i.e., mathematical models based on physical laws, observation, and *a priori* knowledge). Variational data assimilation (VDA) [1, 2] and filtering DA [3, 4, 5] are the two most common methods. Their applications include

numerical weather prediction (NWP), marine and pollution monitoring [6], hydrologic forecasting, soil moisture monitoring, land surface flux and wave speed estimation [7]. The problem of determining the unknown variables called control variables is considered as a nonlinear least-squares problem in the framework of VDA and is solved using the adjoint method [2]. This is the preferred approach in numerical weather prediction, in its incremental formulation [8].

For the pollution problem, reliable prediction from the concentration transport model is not possible when the initial concentration information is limited, observations are scarce, and velocity fields from the ocean and wind are inaccurate (e.g., oil spill). In this scenario, the level-set method can be considered. It was pioneered by Osher and Sethian, aiming at computing and analyzing the motion of an interface in two or three dimensions [9, 10, 11, 12, 13]. Topological merging and breaking are well defined and easily performed by the method of implicit representation of a time-evolving contour. At present, the level-set method has been successfully applied to tumor growth [14, 15], flame propagation [13], and wildland fire propagation [16]. However, owing to the uncertainty in the model such as the parameters of the indirect measurement and inaccurate initial contour position, several data-driven methods have gained increased attention. The interface evolution can be corrected and become reliable by combining the model variables with observation and statistical methods. Among these, the level-set methods based on VDA [2] have been used for motion estimation and structure tracking [17, 18, 19, 20].

In [17], the theoretical framework for assimilating the contour information in ocean image observation was illustrated and a shape-fitting functional used in image processing was considered. The continuous cost function and adjoint-based gradient were given, and the results showed its potential for oil spill and pollutant transport. In addition, there have been attempts to use the level-set structure to indirectly retrieve the flow fields in fluids [19, 20]. Moreover, filtering-based methods have been proposed for front tracking the wildfire spread by assimilating observation [21, 22].

However in DA, because of the displacement and obstacle in the remote-sensing observer, position and shape errors usually arise in observation data (e.g., image-type sequence) compared with true state and background simulation. To avoid this effect, in traditional approaches, a step of data alignment is usually included before using DA for correcting the position error [23, 24, 25]. A shape similarity measure based on the Chan–Vese contour fitting functional was studied for DA [26]. This novel method could more efficiently compare the position information of the level-set structure

from observation and model variables for the front-tracking problem. However, the background information could not be made full use in these DA methods. One possible solution would be to use optimal transport-based metrics.

The optimal transport theory has been significantly developed and applied to wide fields since it was introduced by Monge [27]. In [28], the transportation problem between two probability measures was described as the minimization of the kinetic energy. The optimal “geodesics” maintained the mass conservation over time, driven by the flow in a continuity equation. A series of numerical methods have been proposed for obtaining the optimal transportation such as the primal-dual method [29], Fisher information regularization method [30] and parallel method [31]. Recently, multi-physics optimal transportation, in which a constraint term is added to the cost function based on the physical criterion [32], has been introduced. In addition, a generalized metric interpolating the optimal transport and Fisher–Rao was proposed and found to be beneficial for different mass transportations [33]. Recent techniques have demonstrated impressive performance in geophysical inverse problems such as full waveform inversion (FWI) [34] and in the field of image processing: image registration [35], video reconstruction [36], color transfer [37], image segmentation [38].

Motivated by the advantages of the Wasserstein metric in the optimal transport theory, in [39], it was used to compare the air pollutant dispersion from observation and model simulation. This non-local metric can avoid the “double penalty” effect in classical point-wise indicators when there is a misplacement. At present, a learning-type Wasserstein metric [40] has been introduced for noise reduction [41]. The model error in DA based on the Wasserstein measure was investigated in [42]. Viewing the 2-Wasserstein metric as a displacement interpolation [43, 44, 45], recently, an optimal transport-based DA [46, 47] considering the position error in the observation and background was proposed. The information from the observation and background was sufficiently used and promising results for shape preservation and position correction were obtained.

Owing to its tremendous effect on real weather and pollution forecasting (e.g., the scope), generally, the position and observation error are considered for DA. Moreover, the concentration observation is limited or even impossible to obtain (e.g., oil spill). To deal with these two key problems, in this work, our main contribution is the utilization of the Wasserstein measure to misfit the discrepancy and a combination with level-set method in the DA framework. First, a distance regularization-based method is established. Then, based on the Wasserstein metric instead of the Euclidean measure,

two novel methods called original and splitting topological data assimilation (OTDA and STDA) are proposed. The proposed methods are suitable for an event lacking concentration information and different from previous approaches of dealing with the position error such as performing a concentration information alignment step before DA [23, 24, 25]. In new methods, contours in observations can be assimilated for dispersion prediction and simultaneously, the shape and position information between contours from observations and background can be integrated when there are position and shape errors (e.g., partially missing observation because of cloud cover and a displaced remote-sensing observer). The proposed method reduces the non-Gaussian type of error, position error, to some extent and optimizes the shape.

The remainder of the article is organized as follows. Section 2 presents a brief introduction on the classical and distance regularized level-set-based DA, including the contour data-fitting cost function and gradient. Section 3 details the optimal transport theory, Wasserstein distance, and topological data assimilation (OTDA and STDA) using the Wasserstein distance. The numerical experiment framework and results of the application of the novel level-set-based DA approach to a model test problem are discussed in Section 4. Finally, the main results are summarized and future work is outlined in Section 5.

2. Level-set-based assimilation using Euclidean distance

2.1. Level-set function and model

A given shape (e.g., pollutant spot C) can be represented by sub-domain $\Omega_c \subset \mathbb{R}^2$, whose boundary $\partial\Omega_c$ can be defined by the zero level set of a mapping called level-set function $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$\forall \mathbf{x} \in \mathbb{R}^2, \quad \begin{cases} \phi(\mathbf{x}) < 0, & \mathbf{x} \in \Omega_c \\ \phi(\mathbf{x}) = 0, & \mathbf{x} \in \partial\Omega_c \\ \phi(\mathbf{x}) > 0, & \mathbf{x} \notin \bar{\Omega}_c \end{cases} \quad (2.1)$$

where function $\phi(\mathbf{x})$ is generally the signed distance from \mathbf{x} to $\partial\Omega_c = \{\mathbf{x} \in \Omega | \phi(\mathbf{x}) = 0\}$. The motion of interface $\partial\Omega_c(t)$ over time is considered as the evolution of level-set function ϕ , which is modeled by a two-dimensional (2D) nonlinear partial differential equation called the level-set model expressed as

$$\begin{cases} \frac{\partial\phi}{\partial t} + \mathbf{V} \cdot \nabla\phi + a\|\nabla\phi\| - \sigma\kappa\|\nabla\phi\| = 0, & \text{on } \Omega \times [0, T] \\ \phi(0) = \phi_0, \end{cases} \quad (2.2)$$

$$\kappa = \operatorname{div}\left(\frac{\nabla\phi}{\|\nabla\phi\|}\right) = \frac{\left(\frac{\partial\phi}{\partial x}\right)^2\frac{\partial^2\phi}{\partial y^2} - 2\frac{\partial\phi}{\partial x}\frac{\partial\phi}{\partial y}\frac{\partial^2\phi}{\partial x\partial y} + \left(\frac{\partial\phi}{\partial y}\right)^2\frac{\partial^2\phi}{\partial x^2}}{\|\nabla\phi\|^3} \quad (2.3)$$

where ϕ_0 is the initial value, a signed distance function (SDF) of contour $\partial\Omega_c(0)$. $\partial\Omega_c$ denotes the boundary of the polluted domain, ∇ is the gradient operator in the spatial domain, $\mathbf{V} = (u, v)$ represents the external velocity field from the flow, constant values a and σ denote the diffusion coefficients, and κ is the mean curvature. The modeling domain is Ω and simulation time is equal to T .

In the level-set model, level-set function $\phi(t)$ is driven by external velocity field \mathbf{V} , and over time, the diffusion is along the normal direction by constant value a and in proportion to mean curvature κ .

Regarding the physics in the level-set model, in the presence of term $\sigma\kappa\|\nabla\phi\|$, the closed curves of an arbitrary shape will collapse into a circle and then disappear, whereas under term $a\|\nabla\phi\|$, a relatively closed smooth curve will contain the sharp edges, and the topological structures of the curve may change. Both diffusion coefficients σ and a depend on the physical characteristics of the pollutant. The numerical periodic or homogenous Neumann boundary condition can be used for this model. To maintain the stability of the level-set function in the numerical simulation, the re-initialization approach [11] can be adopted. Fig. 1 gives an example of the level-set method applied to two spots of pollutant.

2.2. Contour-fitting observation operators

To assimilate the active contours from the observation, two observation operators suitable to both metrics, namely the Euclidean metric and Wasserstein distance, are introduced. These types of operators can combine level-set function ϕ with the corresponding contours so that the Euclidean or Wasserstein distance can be used to efficiently measure the discrepancy between them. Let us define the ‘‘observation to structure’’ operator as

$$t \in [0, T], C^S(\mathbf{x}, t) = \mathcal{H}_{\mathcal{O} \rightarrow \mathcal{S}}(C_{obs}(\mathbf{x}, t)) = \begin{cases} 1, & \mathbf{x} \in \bar{\Omega}_c(t) \\ 0, & \mathbf{x} \in \Omega \setminus \bar{\Omega}_c(t) \end{cases} \quad (2.4)$$

where $C^S(t)$ belongs to the new observation space.

Another observation operator called the ‘‘model to structure’’ operator is defined as

$$\mathcal{H}_{\phi \rightarrow \mathcal{S}}(\phi) = \begin{cases} 1, & \phi \leq 0 \\ 0, & \phi > 0 \end{cases} \quad (2.5)$$

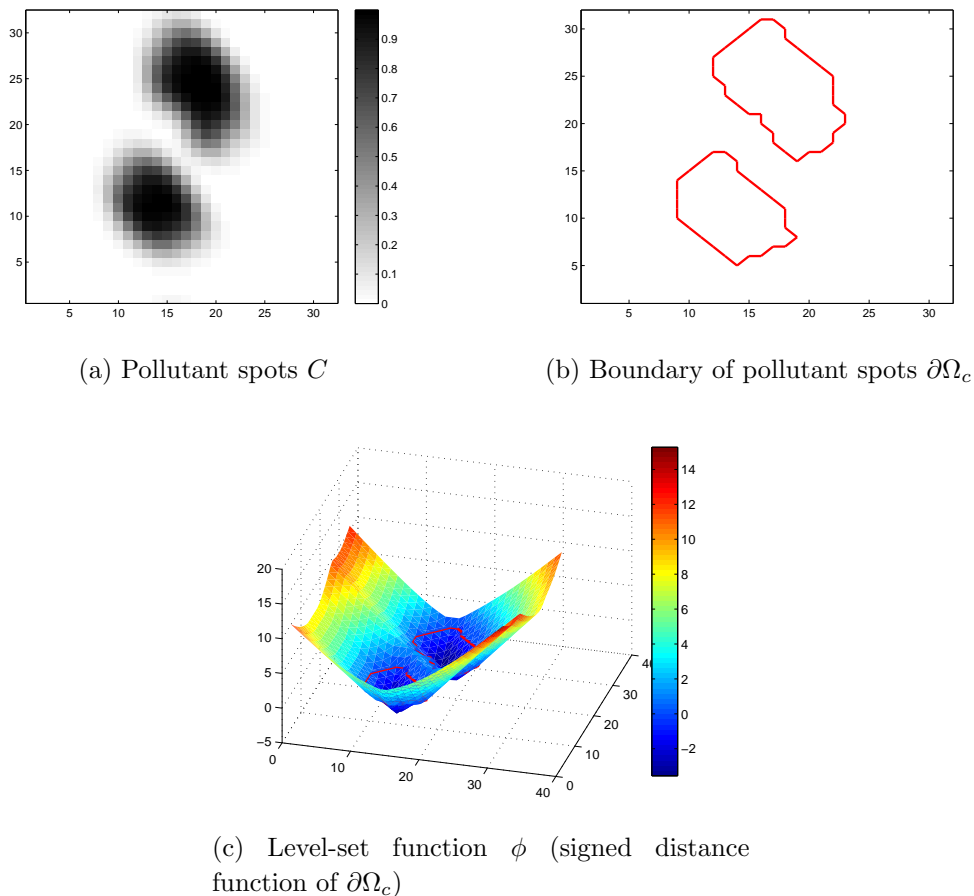


Figure 1: Level-set method. (a): Concentration distribution of the pollutant spots at one moment; (b): Boundary of pollutant spots $\partial\Omega_c$; (c): Level-set function ϕ (signed distance function based on $\partial\Omega_c$ where $\partial\Omega_c = \{\mathbf{x} \in \Omega | \phi(\mathbf{x}) = 0\}$).

where ϕ denotes the level set function. $\mathcal{H}_{\phi \rightarrow \mathcal{S}}(\phi)$ is a Heaviside-like function, which approximates to the following differentiable form:

$$\mathcal{H}_{\phi \rightarrow \mathcal{S}}(\phi) = \frac{1}{2}(1 - \text{sgn}(\phi)) \approx \frac{1}{2}\left(1 - \frac{\phi}{\sqrt{\phi^2 + \varepsilon}}\right) \quad (2.6)$$

where ε is a small constant. This operator maps level-set function ϕ to the new observation space.

2.3. Contour-fitting cost function with Euclidean distance

Generally, the optimal control theory is involved in the framework of traditional four-dimensional variational data assimilation (4D-Var). The cost function estimates the residual between the control variable from the governing model and observation over

time. For the level-set model, the control variable is initial condition ϕ_0 . In [17], the optimal control problem aimed at obtaining the minimizer of the following cost function:

$$J(\phi_0) = J^o + J^b = \frac{\delta_o}{2} \sum_{i=1}^N \|\text{sgn}[G_i(\phi_0)] - \tilde{C}^S(t_i)\|_{L^2(\Omega)}^2 + \frac{\delta_b}{2} \|\phi_0 - \phi_0^b\|_{L^2(\Omega)}^2 \quad (2.7)$$

where $G_i(\cdot)$ denotes the forward operator of the level-set model from time 0 to t_i and sgn is a smoothing sign operator. \tilde{C}^S is the corresponding structural image from the observation, and ϕ_0^b represents the background, which is an SDF. The zero level set of ϕ_0^b is the contour of the first guess. When weight δ_b is small, the initial value is corrected by the observation. Due to the inconsistent way of comparison, information contained in both the observation and background cannot be used efficiently. In this case, the cost function can be modified by a mass conservation-type functional,

$$\begin{aligned} J(\phi_0, a, \sigma) &= J^o + J^b + J_{\text{reg}}^a \\ &= \frac{\delta_o}{2} \sum_{i=1}^N \|\mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[G_i(\phi_0, a, \sigma)]) - \mathcal{N}(C^S(t_i))\|_{L^2(\Omega)}^2 \\ &\quad + \frac{\delta_b}{2} \|\mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[\phi_0]) - \mathcal{N}(B^S)\|_{L^2(\Omega)}^2 + J_{\text{init}}(\phi_0) \end{aligned} \quad (2.8)$$

where

$$B^S = \mathcal{H}_{\phi \rightarrow \mathcal{S}}(\phi_0^b) \quad (2.9)$$

and $J_{\text{init}}(\phi_0) = \lambda \int_{\Omega} \mathcal{R}(\phi_0)$,

$$\mathcal{R}(\phi_0) = \begin{cases} \frac{1}{2}(|\nabla\phi_0| - 1)^2, & |\nabla\phi_0| \geq 1 \\ \frac{1}{(2\pi)^2}(1 - \cos(2\pi|\nabla\phi_0|)), & |\nabla\phi_0| < 1 \end{cases} \quad (2.10)$$

$\mathcal{H}_{\phi \rightarrow \mathcal{S}}[\cdot]$ is defined as (2.6), and B^S is the structure from the background. $\mathcal{N}(\cdot)$ is the normalization operator defined as $\mathcal{N}(f) = \frac{f}{\int_{\Omega} f}$. The minimum problem is ill-posed when only the contour information from the observation and background is assimilated. Therefore, a stable function called distance-based regularization term J_{reg}^a used in image segmentation [48] is added. This term contains an SDF-fitting function making ϕ_0 an SDF. λ denotes the weight.

In the following, we will focus on the computation of the gradient of the above optimization problem. The gradient of cost function J^o with respect to the initial value can be obtained by solving an adjoint equation backward. For the term related to the background and distance-based regularization, the gradient is, respectively, expressed as

$$\text{grad}J^b(\phi_0) = \delta_b(\mathbf{H}_{\phi \rightarrow \mathcal{S}})^*(\mathcal{H}_{\phi \rightarrow \mathcal{S}}(\phi_0) - B^S) \quad (2.11)$$

and

$$\text{grad}J_{\text{reg}}^a(\phi_0) = \begin{cases} -\lambda \text{div}\left(\left(1 - \frac{1}{|\nabla\phi_0|}\right)\nabla\phi_0\right), & |\nabla\phi_0| \geq 1 \\ -\lambda \text{div}(\text{sinc}(2|\nabla\phi_0|)\nabla\phi_0), & |\nabla\phi_0| < 1 \end{cases} \quad (2.12)$$

where $*$ denotes the adjoint of a linear operator and $\text{sinc}(x) = \frac{\sin(2\pi x)}{2\pi x}$.

3. Topological data assimilation

3.1. Optimal transport theory and Wasserstein distance

In [28], a fluid mechanics formulation of the optimal transport was introduced with the objective of obtaining optimal geodesic $\rho(t, x)$ that changed over time between two densities. The mechanics formulation was described as determining the minimizer of the following kinetic energy:

$$\int \int_{[0,1] \times \Omega} \rho(t, x) |\nu(t, x)|^2 dt dx \quad (3.13)$$

where densities $\rho(t, x)$ should belong to mass function space $\mathcal{P}(\Omega)$ defined as

$$\mathcal{P}(\Omega) := \left\{ \rho \geq 0 : \int_{\Omega} \rho(x) dx = 1 \right\} \quad (3.14)$$

and the density and velocity (ρ, ν) satisfy a continuity equation,

$$\begin{aligned} \partial_t \rho + \text{div}(\rho \nu) &= 0, \\ \mathcal{C}(\rho_0, \rho_1) &:= \left\{ (\rho, \nu) \text{ s.t. } \rho(t=0) = \rho_0, \rho(t=1) = \rho_1, \right\} \\ \rho \nu \cdot \mathbf{n} &= 0 \quad \text{on } \partial\Omega \end{aligned} \quad (3.15)$$

ρ_0 and ρ_1 are given as the densities at the initial and final moment, respectively. This type of optimal transport satisfies mass conservation. At present, there are numerous numerical methods for solving the optimal transport problem [29, 30].

Wasserstein distance \mathcal{W} between mass functions ρ_0 and ρ_1 is defined as

$$\mathcal{W}(\rho_0, \rho_1) = \sqrt{\min_{(\rho, \nu) \in \mathcal{C}(\rho_0, \rho_1)} \int \int_{[0,1] \times \Omega} \rho(t, x) |\nu(t, x)|^2 dt dx} \quad (3.16)$$

Wasserstein metric \mathcal{W} can be considered as a displacement interpolation between two densities functions. Hence, in the next section, we will mainly focus on the level-set-based DA with the Wasserstein metric-type contour-fitting functional to tackle the effect of the position and shape errors in the data.

3.2. Topological data assimilation using Wasserstein distance (TDA)

Both position and shape errors in the observation and background affect the quality of the assimilation. Motivated by the optimal transport theory, the Wasserstein distance could be used to simulate the residual instead of the l_2 -norm to make full use of their information and optimize the shape. Thus, the contour-fitting functional can be modified by

$$\begin{aligned} J_{\mathcal{W}}(\phi_0) &= J_{\mathcal{W}}^o + J_{\mathcal{W}}^b + J_{\text{reg}}^a \\ &= \frac{\delta_o}{2} \sum_{i=1}^N \mathcal{W}(\mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[G_i(\phi_0)]), \mathcal{N}(C^{\mathcal{S}}(t_i)))^2 \\ &\quad + \frac{\delta_b}{2} \mathcal{W}(\mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[\phi_0]), \mathcal{N}(B^{\mathcal{S}}))^2 + J_{\text{init}}(\phi_0) \end{aligned} \quad (3.17)$$

where $\mathcal{W}(\cdot, \cdot)$ represents the Wasserstein distance. The normalization operator makes the mass of transport equal to 1, i.e., their masses are the same. $\mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[G_i(\phi_0)])$, $\mathcal{N}(C^{\mathcal{S}}(t_i))$, $\mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[\phi_0])$, and $\mathcal{N}(B^{\mathcal{S}})$ belong to mass function space $\mathcal{P}(\Omega)$, which satisfies the optimal transport theory. Alternatively, the Fisher–Rao metric-based approach [33] could be used to solve the problem of different masses in optimal transport. δ_o and δ_b are the weights and N is the number of observations. $J_{\text{init}}(\phi_0)$ is the SDF-fitting term.

To obtain the minimizer of the variational model (3.17), we propose splitting method-based topological data assimilation (STDA), which is summarized in Algorithm 3.2. The cost function is split into two parts, with one term being associated with the observation and background ($J_{\mathcal{W}}^{o,b}$) and the other being SDF-fitting term $J_{\text{init}}(\phi_0)$. First, the gradient of $J_{\mathcal{W}}^{o,b}$ is obtained and the solution is updated by the descent algorithm. Then, the updated solution is re-initialized to an SDF, taken as the optimal solution in the next iteration. To accelerate the convergence, in this work, the re-initialization process [11] is performed by

$$\begin{cases} \frac{\partial \psi}{\partial \tau} = \text{sgn}(\phi_0)(1 - \|\nabla \psi\|), & \text{on } \Omega \times [0, \tau_0] \\ \psi(0) = \phi_0 \end{cases} \quad (3.18)$$

As time $t \rightarrow \infty$, ψ gradually satisfies $\|\nabla \psi\| = 1$, which approximates an SDF. Alternatively, the TDA method based on the SDF-fitting function in (2.8) is called the original approach (OTDA), which is achieved by Algorithm 3.2. Next, we concentrate on the computation of the gradient of new cost function $J_{\mathcal{W}}^{o,b}$ with respect to initial value ϕ_0 .

In [46], the Wasserstein distance-based variational model proved to be differentiable, and by choosing different representations of the inner products and metric methods in

the tangent space, the gradient of the Wasserstein-based cost function changed according to the definition of the Gâteaux derivative expressed as

$$\forall \eta \in T_{\phi_0} \mathcal{P}, \quad \lim_{\epsilon \rightarrow 0} \frac{J_{\mathcal{W}}^{\alpha,b}(\phi_0 + \epsilon \eta) - J_{\mathcal{W}}^{\alpha,b}(\phi_0)}{\epsilon} = \langle \text{grad}_{\star} J_{\mathcal{W}}^{\alpha,b}(\phi_0), \eta \rangle_{\star} \quad (3.19)$$

where $\langle \cdot, \cdot \rangle_{\star}$ represents a type of inner product and grad_{\star} represents the gradient of the cost function using such inner product. $T_{\phi_0} \mathcal{P}$ is called the tangent space. The following theorem provides the gradient of the proposed cost function.

Theorem 3.1 *If the Euclidean scalar product and l_2 -norm are used in tangent space $T_{\phi_0} \mathcal{P}$, then the gradient of $J_{\mathcal{W}}^{\alpha,b}$ contained in the objective function (3.17) with respect to initial value ϕ_0 becomes*

$$\text{grad}_2 J_{\mathcal{W}}^{\alpha,b}(\phi_0) = \delta_o \sum_{i=1}^N (\mathbf{G}_i)^* (\mathbf{H}_{\phi \rightarrow \mathcal{S}})^* \mathbf{N}^* \Psi^i + \delta_b (\mathbf{H}_{\phi \rightarrow \mathcal{S}})^* \mathbf{N}^* \Psi^b \quad (3.20)$$

where $\Psi^i, i = 1, \dots, N$ and Ψ^b are the Kantorovich potentials of the transport between $\mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[G_i(\phi_0)])$ and $\mathcal{N}(C^{\mathcal{S}}(t_i))$, $\mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[\phi_0])$ and $\mathcal{N}(B^{\mathcal{S}})$, respectively. \mathbf{G}_i , $\mathbf{H}_{\phi \rightarrow \mathcal{S}}$ and \mathbf{N} are the Jacobian of operators $G_i(\cdot)$, $\mathcal{H}_{\phi \rightarrow \mathcal{S}}[\cdot]$, and $\mathcal{N}(\cdot)$. The definitions of the Euclidean scalar product and l_2 -norm are described by

$$\forall \xi, \zeta \in T_{\phi_0} \mathcal{P}, \quad \langle \xi, \zeta \rangle_2 = \int_{\Omega} \xi \zeta dx \quad \text{and} \quad \|\xi\|_{L^2(\Omega)}^2 = \int_{\Omega} |\xi|^2 dx \quad (3.21)$$

The brief proof is given in the Appendix. The method of computing the Kantorovich potential numerically used in the gradient (3.20) is illustrated in Remark 3.2 and 3.3. Then, nonlinear descent algorithms such as the gradient descent and conjugate gradient-like method can be used to obtain the optimal solution of the cost function. To maintain the stability of the optimal value, at each iteration, the re-initialization method should be adopted to make initial contour ϕ_0 an SDF. The re-initialization method can accelerate the convergence during the shape optimization (Remark 3.4). A schematic of proposed methods is presented in Fig. 2.

Remark 3.2 *For the optimal transport problem (3.13), the optimal velocity fields $\tilde{\nu}(t, x)$ have the property [28, 46, 47],*

$$\nu(t, x) = \nabla \Phi(t, x)$$

where the function $\Phi(t, x)$ follows a Hamilton-Jacobi equation

$$\frac{\partial \Phi}{\partial t} + \frac{|\nabla \Phi|^2}{2} = 0$$

In addition, the so-called Kantorovich potential $\Psi(x)$ is defined by

$$\Psi(x) := -\Phi(t = 0, x)$$

So the following representation can be obtained

$$\tilde{v}(t = 0, x) = -\nabla\Psi(x)$$

where $\tilde{v}(0)^i$ is the initial velocity fields for the minimizer of (3.13).

Therefore, Kantorovich potentials Ψ^i and Ψ^b can be determined by solving a series of 2D Poisson equations

$$\begin{cases} \Delta\Psi^i &= -\operatorname{div}(\tilde{v}(0)^i), \\ \frac{\partial\Psi^i}{\partial\mathbf{n}} &= 0 \quad \text{on} \quad \partial\Omega \end{cases}$$

where $\tilde{v}(0)^i$ and $\tilde{v}(0)^b$ are the initial velocities in the optimal transport problems, in which the available mass functions are described by $\rho_1^i = \mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[G_i(\phi_0)])$ and $\rho_2^i = \mathcal{N}(C^{\mathcal{S}}(t_i))$, $i = 1, \dots, N$, $\rho_1^b = \mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[\phi_0])$ and $\rho_2^b = \mathcal{N}(B^{\mathcal{S}})$. $\tilde{v}(0)^i$ and $\tilde{v}(0)^b$ can be obtained by the iterative algorithm in [29].

Remark 3.3 For normalization operator $\mathcal{N}(\cdot)$, the following condition is well satisfied:

$$\mathbf{N}^*(\Psi + C) = \mathbf{N}^*\Psi$$

where C belongs to the constant space. Thus, the effect of the non-unique solution of the above Poisson equation can be avoided. A brief illustration is given in the Appendix.

Remark 3.4 In [46], two types of gradients with respect to the initial value were defined by using different inner products in tangent space $T_{\phi_0}\mathcal{P}$, called as the \mathcal{L}^2 gradient and \mathcal{W} gradient. \mathcal{W} gradient shows a better performance than \mathcal{L}^2 gradient for convergence but under the assumption of a non-flux at the boundary of the \mathcal{L}^2 gradient. Our proposed method also has a rapid convergence, mainly owing to the re-initialization process when the \mathcal{L}^2 gradient is utilized.

Algorithm 3.2 : Splitting method-based TDA Algorithm (STDA)

Input: Observations $C_{obs}(t_i)$, $i = 1, \dots, N$, background $\phi_0^0 = \phi_0^b$, weight parameters δ_o and δ_b , stopping criterion τ .

Output: Analysis ϕ_0^a .

- 1: while $|J(\phi_0^k) - J(\phi_0^{k-1})| > \tau$ do
 - 2: Compute the gradient of $J_{\mathcal{W}}^{o,b}$ with respect to ϕ_0^k :

$$\text{grad}_2 J_{\mathcal{W}}^{o,b}(\phi_0^k) = \delta_o \sum_{i=1}^N (\mathbf{G}_i)^* (\mathbf{H}_{\phi \rightarrow \mathcal{S}})^* \mathbf{N}^* \Psi^{i^k} + \delta_b (\mathbf{H}_{\phi \rightarrow \mathcal{S}})^* \mathbf{N}^* \Psi^{b^k}.$$
 - 3: Iterate ϕ_0^k by the descent algorithm (e.g., steepest descent or conjugate gradient method) $\phi_0^{k+1} = \phi_0^k - \mu^k \Theta(\text{grad}_2 J_{\mathcal{W}}^{o,b}(\phi_0^k))$ where μ^k is the optimal step size in the k iteration and $\Theta(\text{grad}_2 J_{\mathcal{W}}^{o,b}(\phi_0^k))$ represents a kind of combination of $\text{grad}_2 J_{\mathcal{W}}^{o,b}(\phi_0^k)$.
 - 4: Re-initialize ϕ_0^{k+1} to a signed distance function $\phi_0^{\text{init}^k}$ and set $\phi_0^{\text{init}^{k+1}}$ as ϕ_0^{k+1} .
 - 5: $k = k + 1$.
 - 6: end while
-

Algorithm 3.2 : Original TDA Algorithm (OTDA)

Input: Observations $C_{obs}(t_i)$, $i = 1, \dots, N$, background $\phi_0^0 = \phi_0^b$ (a signed distance function), weight parameters δ_o and δ_b , stopping criterion τ .

Output: Analysis ϕ_0^a .

- 1: while $|J(\phi_0^k) - J(\phi_0^{k-1})| > \tau$ do
 - 2: Compute the gradient of $J_{\mathcal{W}}$ with respect to ϕ_0^k :

$$\text{grad}_2 J_{\mathcal{W}}(\phi_0^k) = \delta_o \sum_{i=1}^N (\mathbf{G}_i)^* (\mathbf{H}_{\phi \rightarrow \mathcal{S}})^* \mathbf{N}^* \Psi^{i^k} + \delta_b (\mathbf{H}_{\phi \rightarrow \mathcal{S}})^* \mathbf{N}^* \Psi^{b^k} + \text{grad} J_{\text{init}}(\phi_0^k),$$
where $\text{grad} J_{\text{init}}(\cdot)$ is given in (2.12).
 - 3: Iterate ϕ_0^k by the descent algorithm (e.g., steepest descent or conjugate gradient method) $\phi_0^{k+1} = \phi_0^k - \mu^k \Theta(\text{grad}_2 J_{\mathcal{W}}(\phi_0^k))$ where μ^k is the optimal step size in the k iteration and $\Theta(\text{grad}_2 J_{\mathcal{W}}(\phi_0^k))$ represents a kind of combination of $\text{grad}_2 J_{\mathcal{W}}(\phi_0^k)$.
 - 4: $k = k + 1$.
 - 5: end while
-

4. Numerical examples

4.1. Numerical setting

To perform the numerical simulation, the finite difference scheme is adopted for the forward level-set model (2.2). The upwind scheme is used for discretizing the convection term, and the central difference method is employed for the diffusion terms. For the

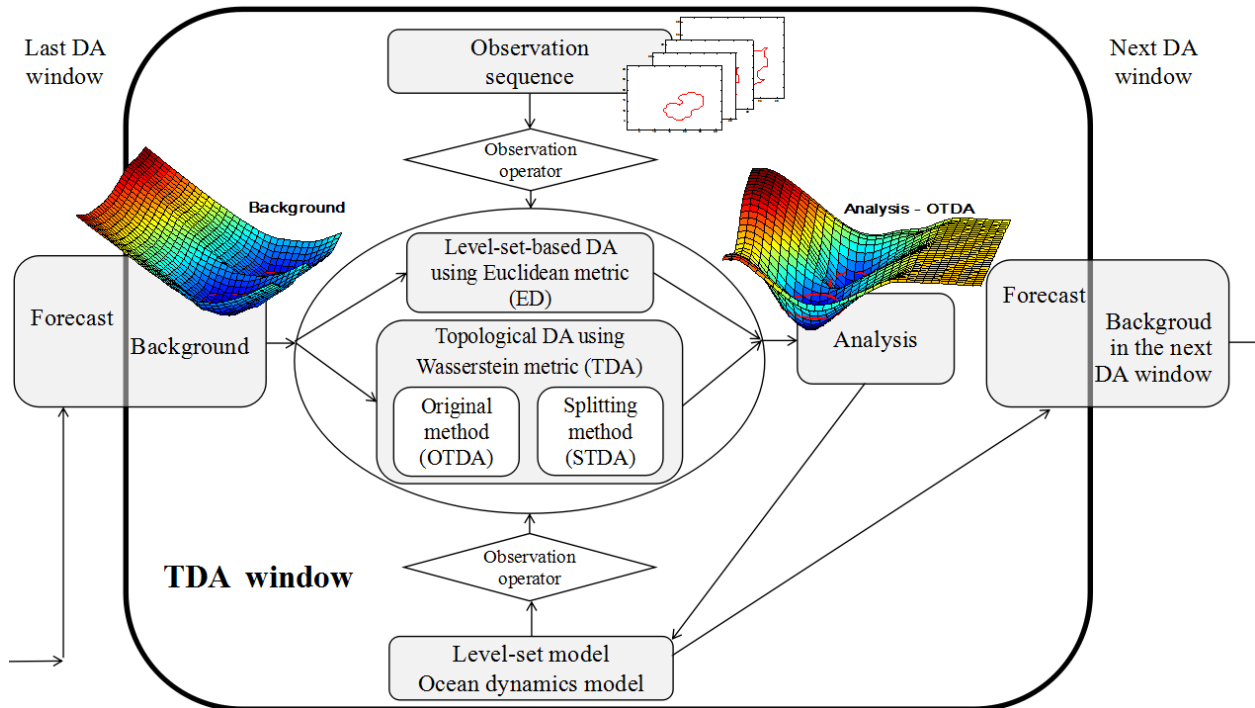


Figure 2: Flowchart of topological data assimilation (OTDA: Original TDA; STDA: Splitting method-based TDA).

sake of simplicity, time integration is performed using the Euler scheme. The model parameters used in the numerical test are set as follows: Modeling domain Ω is defined by $\Omega = [0, L_x] \times [0, L_y]$, where $L_x = L_y = 31\text{m}$. The spatial discretization mesh is $(M, M) = (32, 32)$. The time step is set to $\Delta t = 0.5\text{s}$. The homogeneous Neumann boundary condition is utilized in this test. In this section, two groups of tests are provided, whose parameters are described as follows:

Parameters in the first group of DA test:

The total simulation time is equal to 6s. The assimilation window is $[0, T]$ with $T = 3\text{s}$, and the moment of prediction is 6s. Heaviside-like parameters ε used in (2.8) and (3.17) are defined as $(\Delta x)^2$ and $(2\Delta x)^2$, respectively. The model (2.2) has the following parameters: $a = 0.05$ and $\sigma = 1e - 5$. The velocities are known in the test with $u = v = 1\text{ms}^{-1}$.

The control variable is initial value ϕ_0 only. The observations are generated by the forward level-set model using a given initial value ϕ_0^{obs} , whose zero level set is shown as a pink solid line and named as Observation in Figs. 4a, 4d, and 4g. In the Euclidean distance-based (ED) method and original topology data assimilation (OTDA), the number of observations is $N = 5$ and weight is $\lambda = 0.3$. The zero

level sets of the corresponding backgrounds with position or shape errors are shown as black dashed lines and named as Background in Figs. 4a, 4d, and 4g. Time τ_0 in the re-initialization (3.18) is set to 9.5s. The nonlinear conjugate gradient algorithm is iterated to search for the optimal solution.

Parameters in the second group of DA test:

A simple 2D concentration transport model is used to test the efficiency of the proposed method. The model is described by

$$\begin{cases} \frac{\partial C}{\partial t} + \mathbf{V} \cdot \nabla C = \mu \Delta C, & \text{on } \Omega \times [0, T] \\ C(0) = C_0, \\ \frac{\partial C}{\partial \mathbf{n}} = 0 & \text{on } \partial\Omega \end{cases} \quad (4.22)$$

where C_0 denotes the concentration of the initial pollutant spots. $\mathbf{V} = (u, v)$ represents the velocity, which is the same as that in the model (2.2), and μ is the diffusion coefficient. The homogenous Neumann boundary condition is used. In the test, $\mu = 0.01$ and $\mathbf{V} = (u, v)$ is obtained by a 2D Euler-type fluid dynamics model. The general form currents of the 2D turbulence are displayed in Fig. 5a in which the vortex can be seen clearly. Numerical solutions $C(t)$ after thresholding are taken as the observation images of the pollutants (shown partly in Fig. 5a), which can be expressed as

$$\Omega_c(t, \mathbf{x}) = \{\mathbf{x}(t) \in \Omega | C(t, \mathbf{x}) > \delta\}, \quad (4.23)$$

$$C_{obs}(t) = \begin{cases} C(t), & C(t, \mathbf{x}) > \delta \\ 0, & C(t, \mathbf{x}) \leq \delta \end{cases} \quad (4.24)$$

here, δ represents the threshold above which the spots of the pollutants are visible. In this test, δ is set to 0.2. The contour can be extracted from the observation by the wavelet-based method [49].

The total simulation time is equal to 20s. The assimilation window is $[0, T]$, with $T = 8.5$ s. Heaviside-like parameters ε used in (2.8) and (3.17) are defined as $(1.3\Delta x)^2$. Parameters σ and a are unknown depending on the pollutants and need to be retrieved in the model (2.2), which may occur in a real event. In our examples, σ is known ($\sigma = 0$) for simplicity.

The control variables are parameter a and initial value ϕ_0 . First, parameter a should be identified with the weight couple ($\delta_o = 30$, $\delta_b = \lambda = 0$) by using the variational model (2.8) under the assumption that a reliable background is provided. Number of concentration observations $N = 9$. For the initial value inversion, background with position or shape errors is given, whose zero level sets are shown as black dashed lines and named as Background in Figs. 6a and 8b. Here, in the ED method, number of concentration observations $N = 5$ and weight $\lambda = 0.3$.

Remark 4.1 *In a real event, first, unknown diffusion parameters a and σ in the level-set model have to be identified using available observation data, which highly depend on the characteristics of the corresponding pollutants. Then using the identified parameter couples, the first guess or background, and observed contours, the initial value can be optimized via our proposed method.*

4.2. Numerical test and results

In this section, the results of the comparison between the ED method and Wasserstein distance-based method (WD method) are presented. By using normalization operator $\mathcal{N}(f) = \frac{f}{\int_{\Omega} f}$, the mass of the observation in one moment and background will remain unchanged, satisfying the theory of optimal transport. Motivated by this, the proposed WD method can deal with the problem that the observation and background not only have position errors but also include different shapes. First, we show the results of test cases where the observation sequence is acquired from the level-set method and *the parameters in the first group of DA test* described in Section 4.1 are used. Let us consider the optimal solution of the following cost function (time-independent):

$$\begin{aligned} J_{\mathcal{W}}(\phi) &= \frac{\delta_x}{2} \mathcal{W}(\mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[\phi]), \mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[\phi_x]))^2 \\ &+ \frac{\delta_y}{2} \mathcal{W}(\mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[\phi]), \mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[\phi_y]))^2 + J_{\text{init}}(\phi) \end{aligned} \quad (4.25)$$

where ϕ_x and ϕ_y are the given level-set functions that have the same shape but different positions (shown in Fig. 3a). $\delta_x = \delta_y = 1e + 3$. Zero level sets $\{\mathbf{x} \in \Omega | \phi(\mathbf{x}, t) = 0\}$ of optimal value ϕ obtained by using the two methods are displayed in Fig. 3b. It can be seen that the WD method can maintain the shape, which is different from the results obtained by the ED method. The changes in the cost function and l_2 -norm of the gradient with increasing number of iterations are presented in Figs. 3c and 3d. The fast convergence shows the efficiency of the proposed method.

For the time-dependent data assimilation problem (3.17), two groups of DA test cases are formed.

4.2.1. The first group of DA test These are three tests (Test 1-3) in the first group of DA test. In test 1, observation with one simple spot and background with multiple interests are provided. The boundaries of the initial observation and zero level sets of the background are shown in Fig. 4a (shown as pink solid lines and black dashed lines). Here, we assume that the background is in the right position and observation provides the right shape information. The zero level set of the optimal value called analysis

ϕ_0^a with weights couples a and b (Table 1) are shown in Figs. 4b and 4c respectively (ED method shown as blue dashed lines and named as Analysis-E; WD method shown as red solid lines and named as Analysis-W). The WD method makes full use of the information between the observation and background, which can be illustrated as the optimal interpolation by using Wasserstein distance $\mathcal{W}(\cdot, \cdot)$ when the background and observation contain position errors. However, compared with the WD method, the ED method, only the linear interpolation is used in by the l_2 -norm. Once position errors occur, a larger shape will be obtained, the so-called “double penalty” effect, where one effect is from the observation and the other from the background. Therefore, compared with the ED method, the proposed WD method provides position and shape correction for the results only from observation and the relatively accurate shape information when weights δ_o and δ_b are uncertain. In this regard, the ED method will be ineffective. In addition, with weight $\delta_o \gg \delta_b$, the results obtained by both the methods are similar and are closer to those obtained only from the observation (presented in Fig. 4c).

Table 1: Weights in the first group of DA test

DA test case: group 1	No. of weight couples	Background weight δ_b (ED)	Background weight δ_b (WD-OTDA)	Observation weight δ_o
Test 1	a	1e+4	3e+2	$\frac{\delta_b}{N}$
	b	1e+3	30	$\frac{25\delta_b}{N}$
Test 2	c	1e+4	3e+2	$\frac{2\delta_b}{N}$
	d	1e+3	30	$\frac{25\delta_b}{N}$
Test 3	e	1e+4	3e+2	$\frac{2\delta_b}{N}$
	f	1e+3	30	$\frac{25\delta_b}{N}$

In order to further verify the validity of the proposed approach, another two tests (Test 2 and 3) have been done in which both the observation and background simulation contain complex multiple structures of interest (two and three spots, respectively). Figs. 4d and 4g display the boundaries of the initial observation and zero level sets of the background (shown as pink solid lines and black dashed lines). Similarly, some reasonable results of the test cases are presented in Figs. 4e and 4f with weights couples c and d (Table 1), and Figs. 4h and 4i with weights couples e and f (Table 1). As shown, the WD method preserves the shape of the contours compared with the ED method. Because the weights $\delta_o = \frac{2\delta_b}{N}$ used in the test is a little larger than that in test 1 ($\delta_o = \frac{\delta_b}{N}$), the position of the contours could be closer to the true state rather than

in the middle (shown in Figs. 4e and 4h). But from the results it can be seen that the proposed method will smooth the edge of the contours and if two contours are too close to each other, they will merge together easily (shown in Figs. 4h and 4i). It may lie in the regularization method, SDF-fitting method and the inaccurate computation for the minimizer of the cost function. Nevertheless, it can still indicate that the novel method applies equally to multiple structures of interest problems. In addition, for the multiple structures of interest problem, numerical tests show that SDF-fitting function $J_{\text{init}}(\cdot)$ in (2.8) is used for the stable convergence although it is slower.

4.2.2. The second group of DA test For a more realistic experiment, the full and partially missing concentration observation from the 2D transport model and currents from the Euler equation are provided to test the efficiency of the proposed method. Here *the parameters in the second group of DA test* described in Section 4.1 are used. The concentration images and currents at four times are presented in Fig. 5a. Before the correction of the initial contours, constant diffusion parameter a in the level-set model is retrieved by using $N = 9$ concentration observation images C_{obs} , with σ set to 0 for simplicity. In reality, the diffusion parameters depend on the pollutant. The change in a with the iteration number is shown in Fig. 5b, with first guess $a = 0$. It converges to $a = 0.1507$ after 10 iterations in this test.

By using identified parameter a and a given background, the tests for assimilating the contours of the pollutants from the full concentration observation images and background with position and shape errors and three weights couples g-i (Table 2) are implemented, the results of which are presented in Figs. 6b–6d. For comparison, Fig. 6a displays the contours of the pollutants in the first observation image (shown as a pink solid line), which can be missing in reality. Similar to the test above, under the assumption that the background is in the right position and observation provides the right shape information, an obvious modification in the position and optimal shape can be found in the results of the analyses with weights couples g (Table 2) as well as in the results of prediction at moment $t = 10\text{s}$ (shown in Fig. 7).

Another test is performed with a missing observation in the spatial domain (shown in Fig. 8a). Differing from previous assumptions, in this test, we assume that the missing observation is in the right position, and shape of the first guess is correctly simulated but has the wrong position. The analyses obtained by both methods with weights couples j (Table 2) are compared in Fig. 8c. It can be concluded that using the proposed method optimally integrates the shape and position information and makes

Table 2: Weights in the second group of DA test

DA test case: group 2	No. of weight couples	Background weight δ_b (ED)	Background weight δ_b (WD-STDA)	Observation weight δ_o
Test 4	g	1e+4	1e+3	$\frac{2\delta_b}{N}$
	h	2e+3	2e+2	$\frac{15\delta_b}{N}$
	i	2.5e+4	2.5e+3	$\frac{0.05\delta_b}{N}$
Test 5	j	1e+4	1e+3	$\frac{3\delta_b}{N}$

full use of them when the missing observation is available. The forecast at moment $t = 20s$ (shown as a red solid line in Fig. 8e) is closer to that by using the true state, which can assist in optimizing rescue operations for real accidents to some extent.

5. Conclusion

In this paper, we proposed a novel data assimilation strategy, namely topological data assimilation, which integrates the level-set method with the Wasserstein metric in an optimal way. Moreover, several numerical tests are presented. First, a distance regularization-based approach is introduced. Then based on it, the Wasserstein distance is used to measure the discrepancy between the model variables, observation, and background, rather than using the Euclidean distance. Combined with the level-set method, the contour information can be assimilated, which is suitable for the observation provided with only the edges of the pollutants, such as an oil spill. Once the misplaced missing observation and background are available and there is uncertainty in the errors, the proposed approach considered as a type of shape optimization can overcome the “double penalty” effect in the local Euclidean metric method. An analysis after the optimal shape integration and modification in the position between the observation and background is obtained. The new method is particularly capable of dealing with obstacles (e.g., cloud cover) in a remote-sensing observer, which occur in a real pollutant forecast. However, the results show that the proposed method will smooth the analysis. In future, we will consider some edge-preserving regularization methods, SDF-fitting methods and extend the proposed method to possible real applications such as oil spill, wildfire propagation, tumor growth, and/or typhoons.

6. Acknowledgement

The authors would like to thank Dr. Nelson Feyeux for his active discussion and helpful suggestions. The work is supported in part by National Key Research and Development Program of China under Grant 2017YFB0202902, and NSFC under Grant 41625017 and Grant 91730306, and China Scholarship Council.

7. Appendix

In this section, we provide a brief proof of Theorem 3.1 and mathematical illustration of Remark 3.3.

Proof 7.1 Here, we only give the short proof for the case that is not included in [46].

Owing to the use of normalization operator $\mathcal{N}(\cdot)$ and observation operators $\mathcal{H}_{\phi \rightarrow \mathcal{S}}(\cdot)$, the following condition can be satisfied simultaneously

$$\forall \text{ small } \epsilon \in \mathbb{R}, \quad \eta \in T_{\phi_0} \mathcal{P}, \quad \mathcal{M}_i(\phi_0) \in \mathcal{P}(\Omega) \quad \text{and} \quad \mathcal{M}_i(\phi_0 + \epsilon \eta) \in \mathcal{P}(\Omega)$$

where $\mathcal{M}_i(\cdot) = \mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[G_i(\cdot)])$. We have

$$\begin{aligned} \mathcal{J}_{\mathcal{W}}^{o,b}(\phi_0 + \epsilon \eta) &= \frac{\delta_o}{2} \sum_{i=1}^N \mathcal{W}(\mathcal{M}_i(\phi_0 + \epsilon \eta), \mathcal{N}(C^{\mathcal{S}}(t_i)))^2 + \frac{\delta_b}{2} \mathcal{W}(\mathcal{T}(\phi_0 + \epsilon \eta), \mathcal{N}(B^{\mathcal{S}}))^2 \\ &= \frac{\delta_o}{2} \sum_{i=1}^N \mathcal{W}(\mathcal{M}_i(\phi_0) + \epsilon \mathbf{M}_i[\phi_0] \eta + o(\epsilon), \mathcal{N}(C^{\mathcal{S}}(t_i)))^2 \\ &\quad + \frac{\delta_b}{2} \mathcal{W}(\mathcal{T}(\phi_0) + \epsilon \mathbf{T}[\phi_0] \eta + o(\epsilon), \mathcal{N}(B^{\mathcal{S}}))^2 \end{aligned}$$

where \mathbf{M}_i and \mathbf{T} are the Jacobian of the operators $\mathcal{M}_i(\cdot)$ and $\mathcal{T}(\cdot)$, which are $\mathbf{N}\mathbf{H}_{\phi \rightarrow \mathcal{S}}\mathbf{G}_i$ and $\mathbf{N}\mathbf{H}_{\phi \rightarrow \mathcal{S}}$. Based on the theory (Theorem 8.13 in [44]), we can obtain

$$\frac{1}{2} \mathcal{W}(\mathcal{M}_i(\phi_0) + \epsilon \mathbf{M}_i[\phi_0] \eta + o(\epsilon), \mathcal{N}(C^{\mathcal{S}}(t_i)))^2 = \frac{1}{2} \mathcal{W}(\mathcal{M}(\phi_0), \mathcal{N}(C^{\mathcal{S}}(t_i)))^2 + \epsilon \langle \eta, \Psi^i \rangle + o(\epsilon)$$

as well as the term regarding the background, where Ψ^i denotes the Kantorovich potential of the transport between $\mathcal{N}(\mathcal{H}_{\phi \rightarrow \mathcal{S}}[G_i(\phi_0)])$ and $\mathcal{N}(C^{\mathcal{S}}(t_i))$. Hence,

$$\begin{aligned} \mathcal{J}_{\mathcal{W}}^{o,b}(\phi_0 + \epsilon \eta) &= \frac{\delta_o}{2} \sum_{i=1}^N \mathcal{W}(\mathcal{M}_i(\phi_0), \mathcal{N}(C^{\mathcal{S}}(t_i)))^2 + \epsilon \delta_o \sum_{i=1}^N \langle \Psi^i, \mathbf{M}_i[\phi_0] \eta \rangle \\ &\quad + \frac{\delta_b}{2} \mathcal{W}(\mathcal{T}(\phi_0), \mathcal{N}(B^{\mathcal{S}}))^2 + \epsilon \delta_b \langle \Psi^b, \mathbf{T}[\phi_0] \eta \rangle + o(\epsilon) \end{aligned}$$

Finally, according to the definition of the Gâteaux derivative, the gradient is obtained as follows

$$\text{grad}_2 J_{\mathcal{W}}^{o,b}(\phi_0) = \delta_o \sum_{i=1}^N (\mathbf{G}_i)^* (\mathbf{H}_{\phi \rightarrow \mathcal{S}})^* \mathbf{N}^* \Psi^i + \delta_b (\mathbf{H}_{\phi \rightarrow \mathcal{S}})^* \mathbf{N}^* \Psi^b$$

where $*$ denotes the transpose of the linear operator.

For Remark 3.3, first, we give the derivative of the normalization operator.

Suppose $f \in L^2(\Omega)$, then according to definition $\mathcal{N}(f) = \frac{f}{\int_{\Omega} f}$, we have

$$\forall \epsilon > 0, \eta \in L^2(\Omega), \quad \mathcal{N}(f + \epsilon\eta) = \frac{f + \epsilon\eta}{\int_{\Omega} f + \epsilon\eta} = \frac{\epsilon\eta}{\int_{\Omega} f} + \frac{f}{\int_{\Omega} f} - \epsilon \frac{f \int_{\Omega} \eta}{(\int_{\Omega} f)^2} + o(\epsilon)$$

Based on the definition of the derivative of operator $\mathcal{G}(\cdot)$ described by

$$\langle \mathbf{G}(f), \eta \rangle = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{G}(f + \epsilon\eta) - \mathcal{G}(f)}{\epsilon}$$

the following can be obtained

$$\langle \mathbf{N}(f), \eta \rangle = \frac{\eta \int_{\Omega} f - f \int_{\Omega} \eta}{(\int_{\Omega} f)^2}$$

Next, we compute the continuous adjoint operator of \mathbf{N} . Because

$$\begin{aligned} \langle \mathbf{N}\eta, \varphi \rangle &= \int_{\Omega} \mathbf{N}\eta\varphi = \int_{\Omega} \frac{\eta m - f \int_{\Omega} \eta}{m^2} \varphi = \frac{1}{m} \int_{\Omega} \eta\varphi - \int_{\Omega} \frac{f\varphi(\int_{\Omega} \eta)}{m^2} \\ &= \frac{1}{m} \int_{\Omega} \eta\varphi - \int_{\Omega} \eta \left(\int_{\Omega} \frac{f\varphi}{m^2} \right) \\ &= \left\langle \eta, \frac{\varphi}{m} - \int_{\Omega} \frac{f\varphi}{m^2} \right\rangle \end{aligned}$$

where $m = \int_{\Omega} f$, we have,

$$\mathbf{N}^*\varphi = \frac{\varphi}{\int_{\Omega} f} - \frac{1}{(\int_{\Omega} f)^2} \int_{\Omega} f\varphi$$

Thus, if C belongs to constant space,

$$\mathbf{N}^*C = 0$$

8. Reference

- [1] Sasaki Y 1970 Some Basic formalisms in numerical variational analysis *Monthly Weather Review* **98** 875-883
- [2] Le Dimet F X and Talagrand O 1986 Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects *Tellus A: Dynamic Meteorology and Oceanography* **38** 97-110
- [3] Ghil M, Cohn S, Tavantzis J, Bube K and Isaacson E 1981 Applications of estimation theory to numerical weather prediction *In Dynamic Meteorology: Data Assimilation Methods* Springer **139-224**
- [4] Evensen G 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte

- Carlo methods to forecast error statistics *Journal of Geophysical Research: Oceans* **99** 10143-10162
- [5] van Leeuwen P J 2010 Nonlinear data assimilation in geosciences: an extremely efficient particle filter *Quarterly Journal of the Royal Meteorological Society* **136** 1991-1999
- [6] Nodet M 2006 Variational assimilation of Lagrangian data in oceanography *Inverse problems* **22** 245
- [7] Fehrenbach J, Oudry J and Sandrin L 2010 Variational data assimilation to estimate the velocity in the wave equation *Inverse Problems* **26** 115005
- [8] Courtier P, Thépaut J N and Hollingsworth A 1994 A strategy for operational implementation of 4D-Var, using an incremental approach *Quarterly Journal of the Royal Meteorological Society* **120** 1367-1387
- [9] Osher S and Sethian J A 1988 Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations *Journal of Computational Physics* **79** 12-49
- [10] Osher S and Fedkiw R P 2001 Level set methods: an review and some recent results *Journal of Computational Physics* **169** 463-502
- [11] Sethian J A and Smereka P 2003 Level set methods for fluid interfaces *Annual Review of Fluid Mechanics* **35** 341-372
- [12] Gibou F, Fedkiw R and Osher S 2018 A review of level-set methods and some recent applications *Journal of Computational Physics* **353** 82-109
- [13] Sethian J A 1999 *Level Set Methods and Fast Marching Methods* Cambridge University Press
- [14] Martins M, Ferreira Jr S and Vilela M 2007 Multiscale models for the growth of avascular tumors *Physics of Life Reviews* **4** 128-156
- [15] Macklin P and Lowengrub J 2006 An improved geometry-aware curvature discretization for level set methods: application to tumor growth *Journal of Computational Physics* **215** 392-401
- [16] Mallet V, Keyes D and Fendell F 2009 Modeling wildland fire propagation with level set methods *Computers and Mathematics with Applications* **57** 1089-1101
- [17] Li L, Le Dimet F X, Ma J and Vidard A 2017 A level-set-based image assimilation method: potential applications for predicting the movement of oil spills *IEEE Transactions on Geoscience and Remote Sensing* **55** 6330-6343
- [18] Papadakis N and Mémin E 2008 An variational technique for time consistent tracking of curves and motion *Journal of Mathematical Imaging and Vision* **31** 81-103
- [19] Lepoittevin Y, Béréziat D, Herlin I and Mercier N 2013 Continuous tracking of structures from an image sequence *In VISAPP-8th International Conference on Computer Vision Theory and Applications* Springer Verlag **386-389**
- [20] Lepoittevin Y, Herlin I and Béréziat D 2013 Object's tracking by advection of a distance map *In ICIP-20th IEEE International Conference on Image Processing* **3612-3616**
- [21] Rochoux M C, Ricci S, Lucor D, Cuenot B and Trouvé A 2014 Towards predictive data-driven simulations of wildfire spread-Part 1: reduced-cost ensemble Kalman filter based on a polynomial chaos surrogate model for parameter estimation *Natural Hazards and Earth System Sciences* **14** 2951-2973
- [22] Rochoux M C, Emery C, Ricci S, Cuenot B and Trouvé A 2015 Towards predictive data-driven simulations of wildfire spread-Part 2: ensemble Kalman filter for the state estimation of a front-tracking simulator of wildfire spread *Natural Hazards and Earth System Sciences* **15** 1721-1739
- [23] Hoffman R N and Grassotti C 1996 A technique for assimilating SSM/I observations of marine

- atmospheric storms: tests with ECMWF analyses *Journal of Applied Meteorology* **35** 1177-1188
- [24] Ravela S, Emanuel K and McLaughlin D 2007 Data assimilation by field alignment *Physica D: Nonlinear Phenomena* **230** 127-145
- [25] Nehr Korn T, Woods B K, Hoffman R N and Auligné T 2015 Correcting for position errors in variational data assimilation *Monthly Weather Review* **143** 1368-1381
- [26] Rochoux M C, Collin A, Zhang C, Trouvé A, Lucor D and Moireau P 2017 Front shape similarity measure for shape-oriented sensitivity analysis and data assimilation for Eikonal equation *ESAIM: Proceeding and Surveys* **1-22**
- [27] Monge G 1781 Mémoire sur la théorie des déblais et des remblais *De l'Imprimerie Royale*
- [28] Benamou J D and Brenier Y 2000 A computational fluid mechanics solution to the Monge-Kantorovitch mass transfer problem *Numerische Mathematik* **84** 375-393
- [29] Papadakis N, Peyré G and Oudet E 2014 Optimal transport with proximal splitting *SIAM Journal on Imaging Sciences* **7** 212-238
- [30] Li W, Yin P and Osher S 2018 Computations of optimal transport distance with Fisher information regularization *Journal of Scientific Computing* **75** 1581-1595
- [31] Li W, Ryu E K, Osher S, Yin W and Gangbo W 2018 A parallel method for Earth Mover's distance *Journal of Scientific Computing* **75** 182-197
- [32] Bouharguane A, Maitre E, Oudet E and Papadakis N 2012 Multiphysics optimal transportation and image analysis
- [33] Chizat L, Peyré G, Schmitzer B and Vialard F X 2018 An interpolating distance between optimal transport and Fisher-Rao metrics *Foundations of Computational Mathematics* **18** 1-44
- [34] Métivier L, Brossier R, Merigot Q, Oudet E and Virieux J 2016 An optimal transport approach for seismic tomography: application to 3D full waveform inversion *Inverse Problems* **32** 115008
- [35] Haker S, Zhu L, Tannenbaum A and Angenent S 2004 Optimal mass transport for registration and warping *International Journal of Computer Vision* **60** 225-240
- [36] Delon J 2006 Movie and video scale-time equalization application to flicker reduction *IEEE Transactions on Image Processing* **15** 241-248
- [37] Papadakis N, Provenzi E and Caselles V 2011 A variational model for histogram transfer of color images *IEEE Transactions on Image Processing* **20** 1682-1695
- [38] Ni K, Bresson X, Chan T and Esedoglu S 2009 Local histogram based segmentation using the Wasserstein distance *International Journal of Computer Vision* **84** 97-111
- [39] Farchi A, Bocquet M, Roustan Y, Mathieu A and Quérel A 2016 Using the Wasserstein distance to compare fields of pollutants: application to the radionuclide atmospheric dispersion of the Fukushima-Daiichi accident *Tellus B: Chemical and Physical Meteorology* **68** 31682
- [40] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V and Courville A C 2017 Improved training of Wasserstein GANs *Advances in Neural Information Processing Systems* **5767-5777**
- [41] Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra K, Zhang Y, Sun L and Wang G 2018 Low dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss *IEEE Transactions on Medical Imaging* **37** 1348-1357
- [42] Ning L, Carli F P, Ebtehaj A M, Foufoula-Georgiou E and Georgiou T T 2014 Coping with model error in variational data assimilation using optimal mass transport *Water Resources Research* **50** 5817-5830
- [43] McCann, R J 1997 A convexity principle for interacting gases *Advances in Mathematics* **128** 153-179

- [44] Villani C 2003 *Topics in optimal transportation (Graduate studies in mathematics)* American Mathematical Society **58**
- [45] Bonneel N, Van De Panne M, Paris S and Heidrich W 2011 Displacement interpolation using Lagrangian mass transport *ACM Transactions on Graphics* **30 158**
- [46] Feyeux N 2016 *Transport optimal pour l'assimilation de données images PhD thesis*, Communauté Université Grenoble Alpes
- [47] Feyeux N, Vidard A and Nodet M 2018 Optimal transport for variational data assimilation *Nonlinear Processes in Geophysics* **25 55-66**
- [48] Li C, Xu C, Gui C and Fox M D 2010 Distance regularized level set evolution and its application to image segmentation *IEEE Transactions on Image Processing* **19 3243-3254**
- [49] Mallat S 1991 Zero-crossings of a wavelet transform *IEEE Transactions on Information theory* **37 1019-1033**

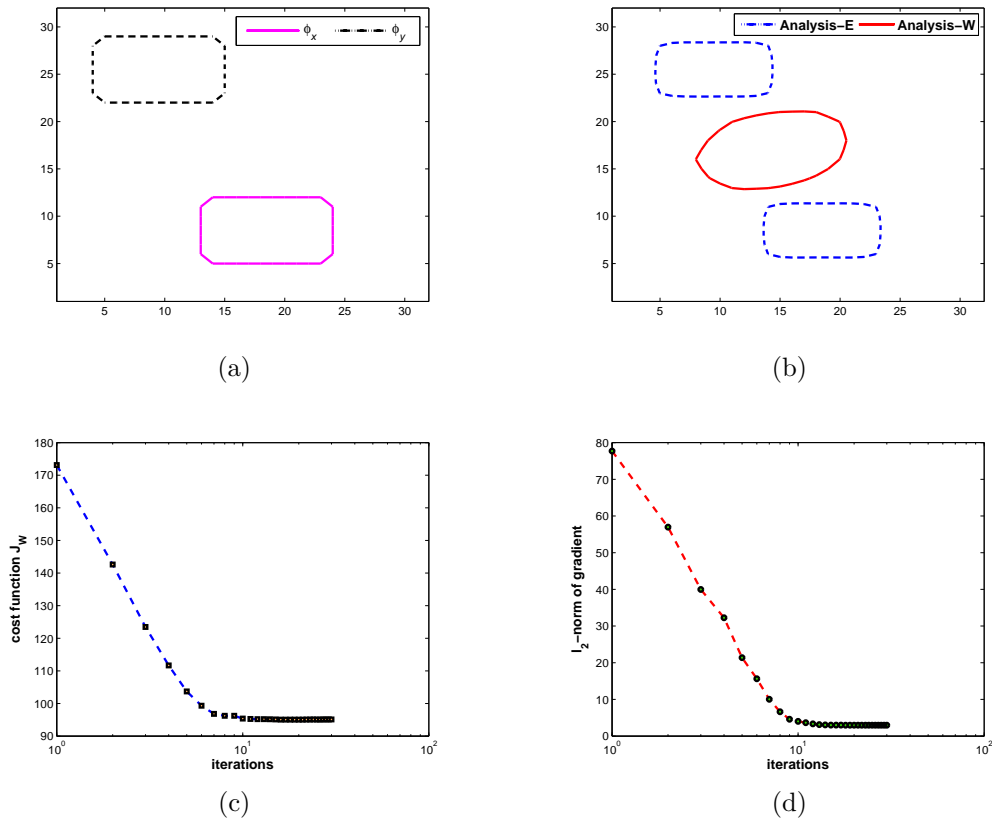


Figure 3: The results of the proposed WD method (STDA) compared with ED method, with observations and background containing position and shape errors. (a): Boundary of the initial observation and zero level set of the background (shown as a pink solid line and black dashed line); (b): Zero level sets of the corresponding analyses (ED method shown as a blue dashed line and named as Analysis-E; WD method shown as a red solid line and named as Analysis-W); (c): Cost function J_W ; (d): l_2 -norm of the gradient.

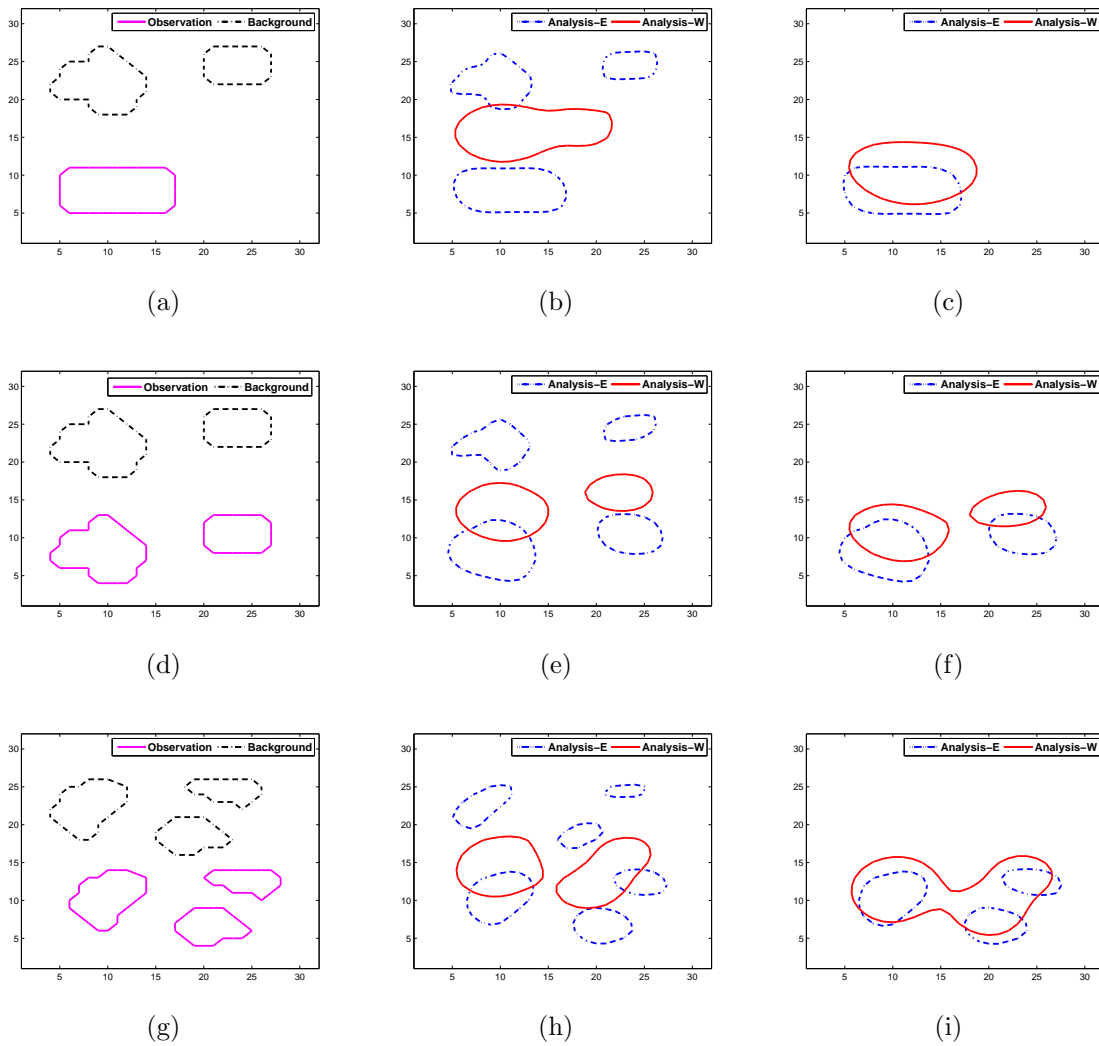
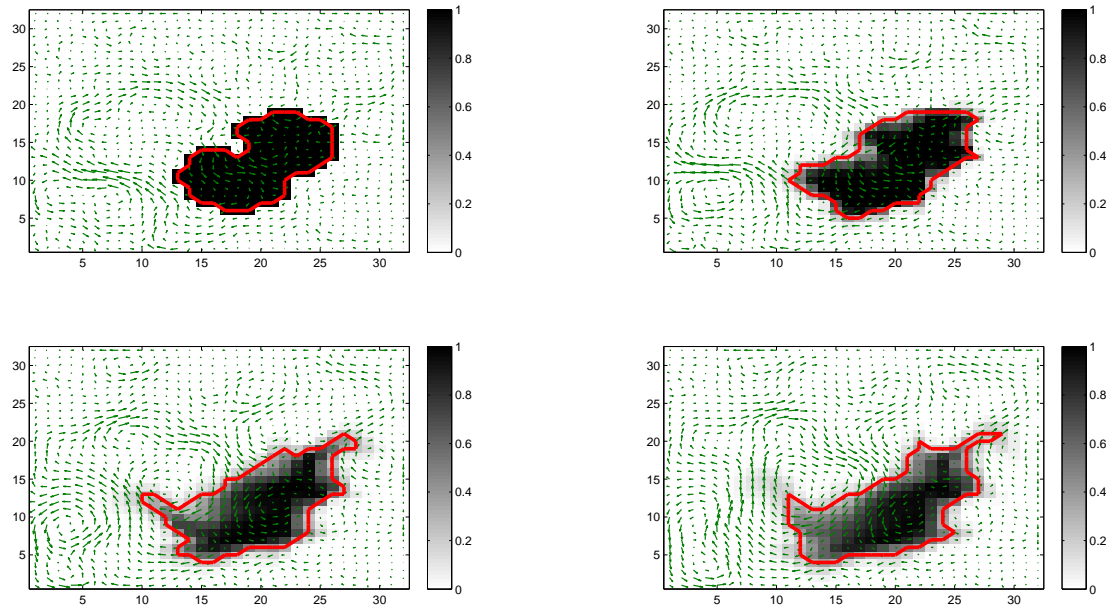
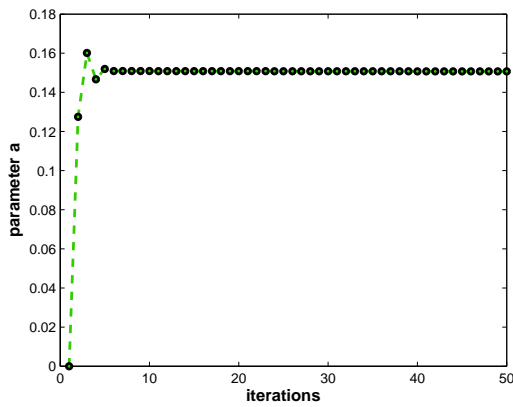


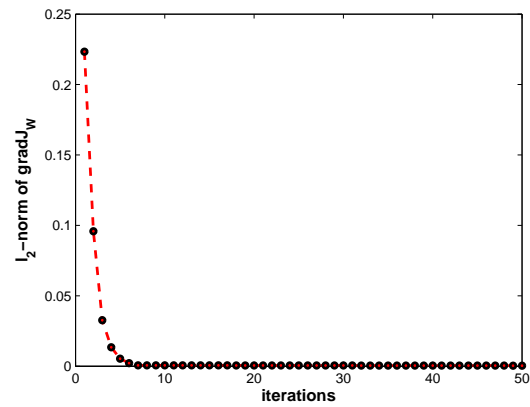
Figure 4: The results of the proposed WD method (OTDA) compared with ED method, with observations and background of multiple structures containing position and shape errors. (a), (d) and (g): Boundary of the initial observation and zero level set of the background (shown as a pink solid line and black dashed line); (b), (c), (e), (f), (h) and (i): Zero level sets of the corresponding analyses with weights $\delta_o = \frac{\delta_b}{N}$, $\delta_o = \frac{25\delta_b}{N}$, $\delta_o = \frac{2\delta_b}{N}$, $\delta_o = \frac{25\delta_b}{N}$, $\delta_o = \frac{2\delta_b}{N}$ and $\delta_o = \frac{25\delta_b}{N}$ (ED method shown as a blue dashed line and named as Analysis-E; WD method shown as a red solid line and named as Analysis-W).



(a)



(b)



(c)

Figure 5: Identification of parameter a in the level-set model using the contour information in concentration observations. (a): Partial observation sequences $C_{obs}(t)$ from the 2D concentration model, currents and boundary of the pollutant (Pollutant spots transportation over time); (b): Retrieved parameter a in the level-set model using such data; (c): l_2 -norm of the gradient.

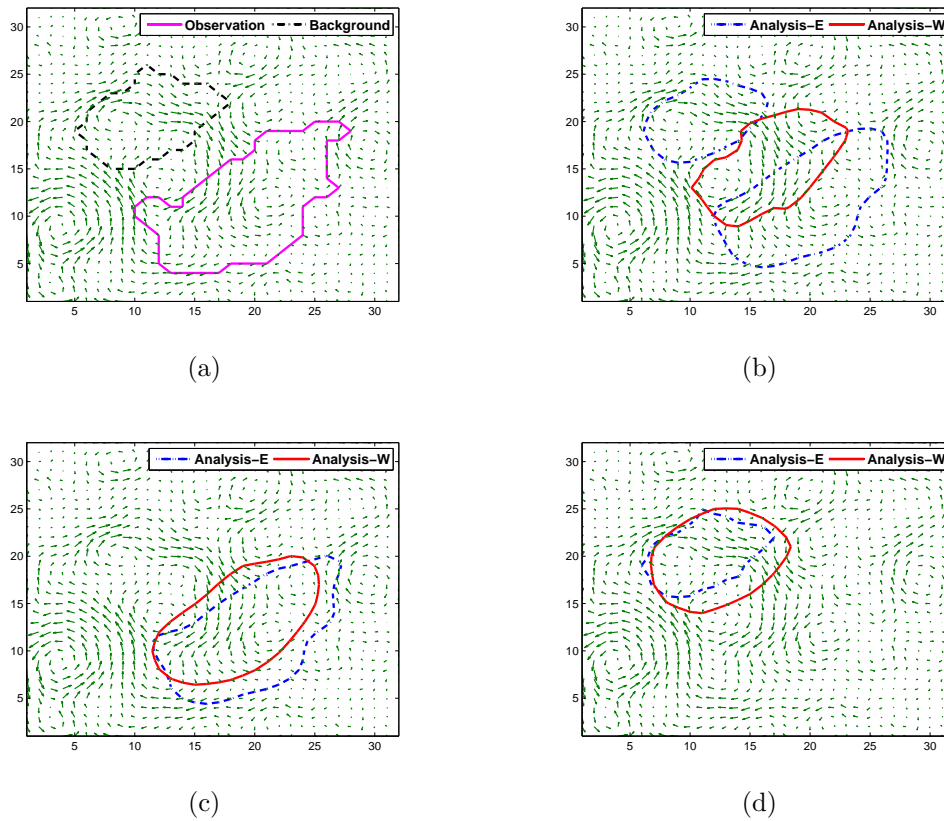


Figure 6: The results of the proposed WD method (STDA) compared with ED method, with observations from the 2D concentration model containing position errors, under the assumption that the background is in the right position but has inaccurate shape and the observations are with accurate shape but in the wrong position. (a): Boundary of the initial observation and zero level set of the background (shown as a pink solid line and black dashed line); (b)-(d): Zero level sets of the corresponding analyses with weights $\delta_o = \frac{2\delta_b}{N}$, $\delta_o = \frac{15\delta_b}{N}$ and $\delta_o = \frac{0.05\delta_b}{N}$ (ED method shown as a blue dashed line and named as Analysis-E; WD method shown as a red solid line and named as Analysis-W).

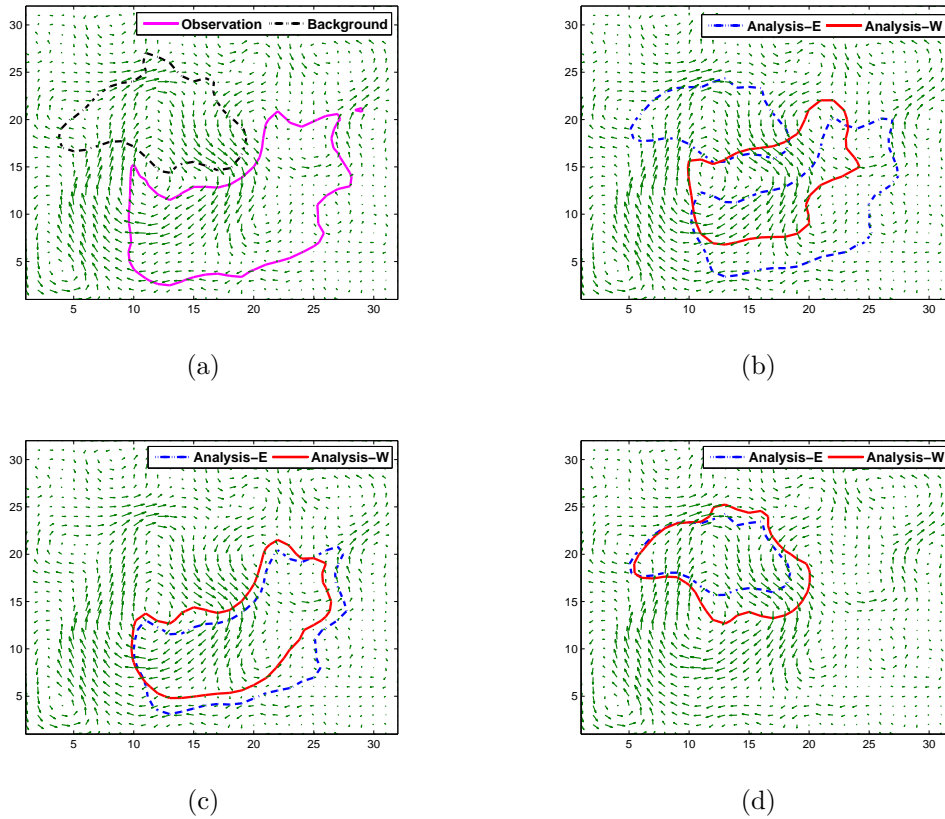


Figure 7: The results of prediction. (a): The results of prediction using the boundary of the initial observation and background (shown as a pink solid line and black dashed line); (b)-(d): The results of prediction using the corresponding analyses with weights $\delta_o = \frac{2\delta_b}{N}$, $\delta_o = \frac{15\delta_b}{N}$ and $\delta_o = \frac{0.05\delta_b}{N}$ (ED method shown as a blue dashed line and named as Analysis-E; WD method shown as a red solid line and named as Analysis-W).

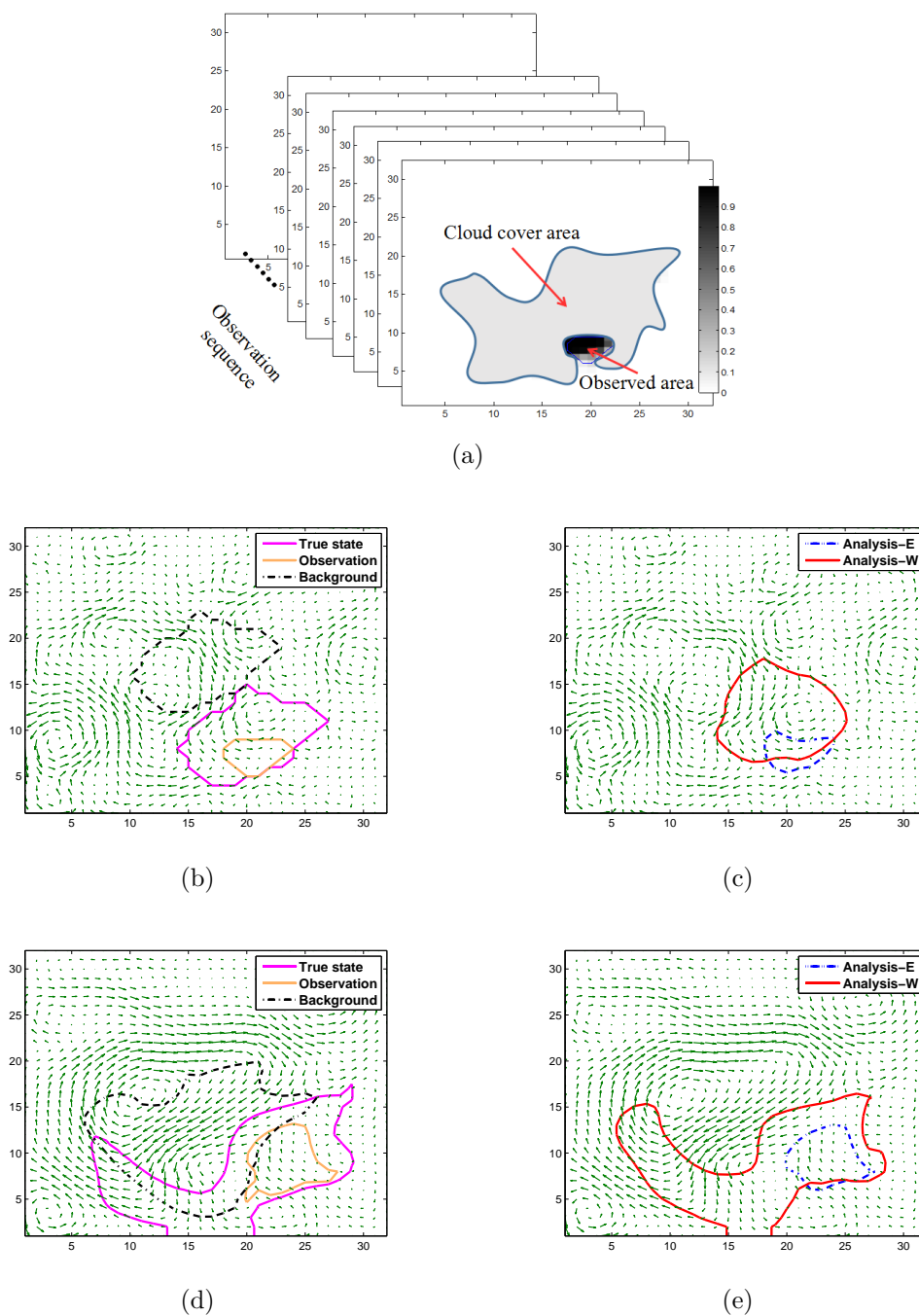


Figure 8: The results of the proposed WD method (STDA) compared with ED method, with partially missing observations (in spatial domain) and background containing position errors, under the assumption that the background is in the wrong position but has accurate shape. (a): Observation sequence covered by cloud; (b): Boundary of the true state (reference), boundary of the initial observation and zero level set of the background (shown as a pink solid line, black dashed line, and orange solid line); (c): Zero level sets of the corresponding analyses with weights $\delta_o = \frac{3\delta_b}{N}$ (ED method shown as a blue dashed line and named as Analysis-E; WD method shown as a red solid line and named as Analysis-W); (d): The results of prediction using the boundary of the true state, boundary of the initial observation and background; (e): The results of prediction using the corresponding analyses.