



HAL
open science

Optimal Transportation for Data Assimilation

Nelson Feyeux, Maëlle Nodet, Arthur Vidard

► **To cite this version:**

Nelson Feyeux, Maëlle Nodet, Arthur Vidard. Optimal Transportation for Data Assimilation. 5th International Symposium for Data Assimilation (ISDA 2016), Jul 2016, Reading, United Kingdom. , 2016. hal-01349637

HAL Id: hal-01349637

<https://hal.science/hal-01349637v1>

Submitted on 28 Jul 2016

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.

Optimal Transportation for Data Assimilation

Nelson FEYEUX, Maëlle NODET, Arthur VIDARD

Inria & Université Grenoble-Alpes



DATA ASSIMILATION

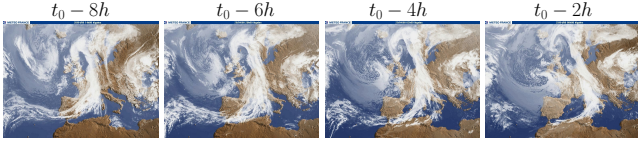
Given

- a physical system and its state $\mathbf{x}(t, x)$;
- **partial** observations of the system (\mathbf{y}_i^o);
- a (numerical) model \mathcal{M} simulating the evolution of \mathbf{x} ;

E.g., for the atmosphere, the state $\mathbf{x}(t, x, y)$ gathers the different variables

- humidity $H(t, x, y)$;
- velocities $\mathbf{u}(t, x, y)$;
- temperature $T(t, x, y)$;
- pressure $p(t, x, y)$.

Can we estimate the initial condition \mathbf{x}_0 of the system?



Variational data assimilation consists in retrieving \mathbf{x}_0 by minimizing

$$\mathcal{J}(\mathbf{x}_0) := \sum_i d\left(\underbrace{\mathcal{H}_i \mathcal{M}_i(\mathbf{x}_0)}_{\text{Mapping of } \mathbf{x}_0 \text{ on the space of } \mathbf{y}_i^o}, \mathbf{y}_i^o\right)^2 + \omega d(\mathbf{x}_0, \mathbf{x}_0^b)^2. \quad (1)$$

It is common for the distance d to be a weighted \mathcal{L}^2 distance. Our main goal to use the Wasserstein distance \mathcal{W}_2 instead, which seems very interesting when dealing with dense data (see right panel). The Wasserstein cost function writes

$$\mathcal{J}_W(\mathbf{x}_0) := \sum_i \mathcal{W}_2(\mathcal{H}_i \mathcal{M}_i(\mathbf{x}_0), \mathbf{y}_i^o)^2 + \omega \mathcal{W}_2(\mathbf{x}_0, \mathbf{x}_0^b)^2. \quad (2)$$

OPTIMAL TRANSPORTATION AND THE WASSERSTEIN DISTANCE

For two functions $\rho_0(x)$ and $\rho_1(x)$, the square of the **Wasserstein distance** $\mathcal{W}_2(\rho_0, \rho_1)$ is defined as the minimal kinetic energy necessary to transport ρ_0 to ρ_1 ,

$$\mathcal{W}_2(\rho_0, \rho_1)^2 := \inf_{\rho(t, x), \mathbf{v}(t, x)} \frac{1}{2} \iint_{[0,1] \times \Omega} \rho |\mathbf{v}|^2 dt dx.$$

$$\partial_t \rho + \text{div}(\rho \mathbf{v}) = 0$$

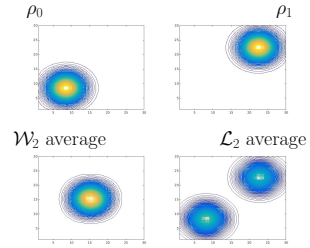
$$\rho(0, x) = \rho_0(x), \rho(1, x) = \rho_1(x)$$

For the Wasserstein distance to be well-defined, one needs

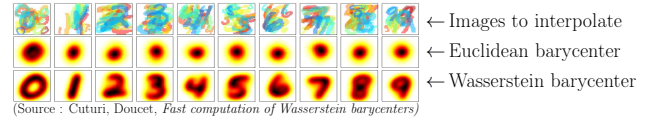
$$\rho_0 \geq 0, \rho_1 \geq 0 \text{ and } \int_{\Omega} \rho_0 = \int_{\Omega} \rho_1 = 1.$$

Average w.r.t the Wasserstein distance

The average, or barycenter, minimizes $\mathcal{W}_2(\rho, \rho_0)^2 + \mathcal{W}_2(\rho, \rho_1)^2$. It is also the optimal ρ in the definition of $\mathcal{W}_2(\rho_0, \rho_1)^2$ at time $t = 1/2$.



Example of use of the Wasserstein distance

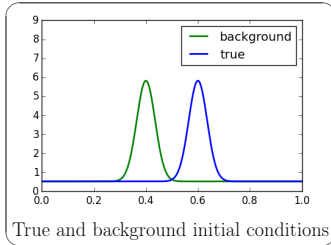
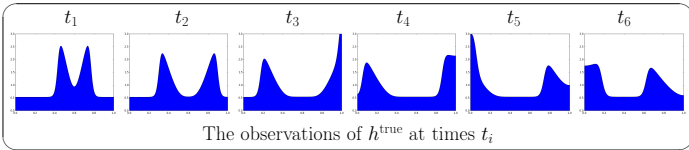


RESULTS ON A SHALLOW-WATER EQUATION

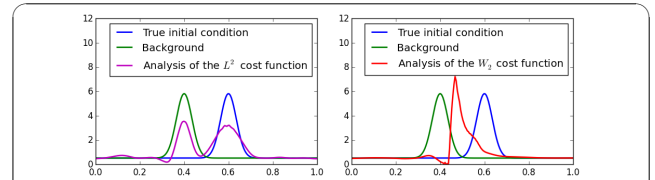
Let the model \mathcal{M} be a Shallow-Water equation, with initial condition (h_0, u_0) ,

$$\mathcal{M}: \begin{cases} \frac{\partial h}{\partial t} + \text{div}(\rho u) = 0 \\ \frac{\partial u}{\partial t} + u \cdot \nabla u = -g \nabla h. \end{cases}$$

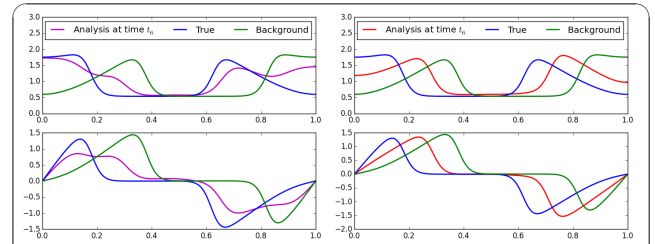
We control the initial condition h_0 only, thanks to the Wasserstein cost function \mathcal{J}_W . We set $u_0 = 0$.



Results :



Analysis h_0 of the assimilation when using the Euclidean (\mathcal{L}^2) or the Wasserstein (\mathcal{W}_2) distance.



Values of h and u for the background and true states, as well as analysis for Euclidean and Wasserstein distances, at time $t = t_6$

SPECIFICITIES ON USING THE WASSERSTEIN DISTANCE

- The Wasserstein distance is only defined for probability measures, *i.e.* $\rho \geq 0$ s.t.

$$\rho \geq 0 \text{ and } \int_{\Omega} \rho = 1$$

Relaxations of the latter constraint are possible, however complex;

- the \mathcal{W}_2 interpolation works well if ρ_0 and ρ_1 are of distinct support;
- when $\mathcal{J}(\rho_0^n) \rightarrow \min_{\rho_0} \mathcal{J}(\rho_0)$, then there is only **weak convergence** of ρ_0^n to ρ_0^{opt} : oscillations or diracs can occur!
- Computing the Wasserstein distance is expensive [Peyré, Papadakis, Oudet, 2013];

- The minimization of \mathcal{J}_W is performed through a gradient descent, using the Wasserstein gradient, arising from the use of the following Wasserstein scalar product depending on ρ_0 ,

$$\text{For } \eta, \eta' \text{ s.t. } \int_{\Omega} \eta = \int_{\Omega} \eta' = 0$$

$$\text{Let } \Phi, \Phi' \text{ s.t. } -\text{div}(\rho_0 \nabla \Phi) = \eta \text{ (with Neumann BC)}$$

$$-\text{div}(\rho_0 \nabla \Phi') = \eta'$$

$$\text{Then } \langle \eta, \eta' \rangle_W = \int_{\Omega} \rho_0 \nabla \Phi \cdot \nabla \Phi' dx.$$

