# Optimal Transportation for Data Assimilation

Nelson Feyeux, Maëlle Nodet, Arthur Vidard

# Optimal Transportation for Data Assimilation

Nelson FEYEUX, Maëlle NODET, Arthur VIDARD

Inria & Université Grenoble-Alpes

## Data assimilation

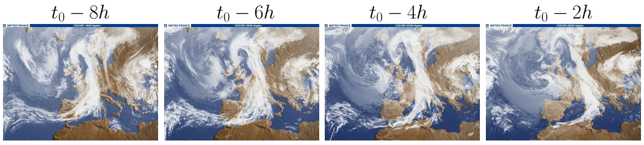
Given

- a physical system and its state $\mathbf{x}(t,x)$;
- partial observations of the system $(\mathbf{y}_i^o)_i$;
- 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

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

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

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

## 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 $\rho_0$ to $\rho_1$,

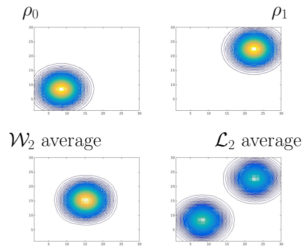
$$\mathcal{W}_2(\rho_0,\rho_1)^2 := \inf_{\substack{(\rho(t,x),\mathbf{v}(t,x)) \\ \partial_t\rho + \operatorname{div}(\rho\mathbf{v}) = 0 \\ \rho(0,x)=\rho_0(x),\, \rho(1,x)=\rho_1(x)}} \frac{1}{2}\iint_{[0,1]\times\Omega} \rho|\mathbf{v}|^2\,\mathrm{d}t\mathrm{d}x.$$

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

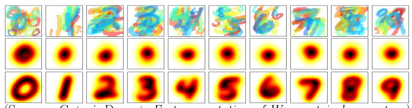
$$\rho_0 \geq 0, \rho_1 \geq 0 \text{ and } \int_\Omega \rho_0 = \int_\Omega \rho_1 = 1.$$

### 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 $\rho$ in the definition of $\mathcal{W}_2(\rho_0,\rho_1)^2$ at time $t=1/2$.




### Example of use of the Wasserstein distance



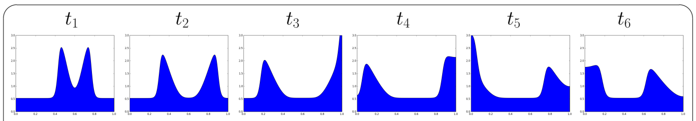

(Source : Cuturi, Doucet, *Fast computation of Wasserstein barycenters*)

## Results on a Shallow-water equation

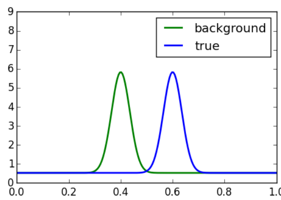
Let the model $\mathcal{M}$ be a Shallow-Water equation, with initial condition $(h_0,u_0)$,

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

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

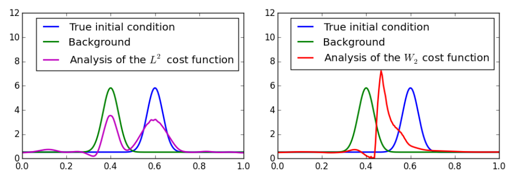



The observations of $h^{\text{true}}$ at times $t_i$

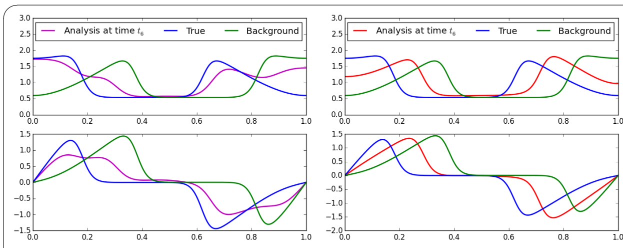



True and background initial conditions

**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 $t = t_6$

## Specifities on using the Wasserstein distance

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

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

  Relaxations of the latter constraint are possible, however complex;
- the $\mathcal{W}_2$ interpolation works well if $\rho_0$ and $\rho_1$ are of distinct support;
- when $\mathcal{J}(\rho_0^n) \to \min_{\rho_0}\mathcal{J}(\rho_0)$, then there is only **weak convergence** of $\rho_0^n$ to $\rho_0^{\text{opt}}$ : 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 $\rho_0$,

$$\begin{aligned} &\text{For } \eta,\eta' \text{ s.t. } \int_\Omega \eta = \int_\Omega \eta' = 0 \\ &\text{Let } \Phi,\Phi' \text{ s.t. } \quad -\operatorname{div}(\rho_0\nabla\Phi) = \eta \ \text{ (with Neumann BC)} \\ &\qquad\qquad\qquad -\operatorname{div}(\rho_0\nabla\Phi') = \eta' \\ &\text{Then } \langle\eta,\eta'\rangle_W = \int_\Omega \rho_0\nabla\Phi\cdot\nabla\Phi'\,\mathrm{d}x. \end{aligned}$$

[Reference] Nelson Feyeux, Maëlle Nodet, Arthur Vidard. Optimal Transport for Data Assimilation. 2016. <hal-01342193>

*This work has been supported by the region Rhône-Alpes.*