

Application of optimal transport to data assimilation

Nelson Feyeux, Maëlle Nodet, Arthur Vidard

► **To cite this version:**

Nelson Feyeux, Maëlle Nodet, Arthur Vidard. Application of optimal transport to data assimilation. Special semester on new trends in calculus of variation, Dec 2014, Linz, Austria. hal-01097190

HAL Id: hal-01097190

<https://hal.inria.fr/hal-01097190>

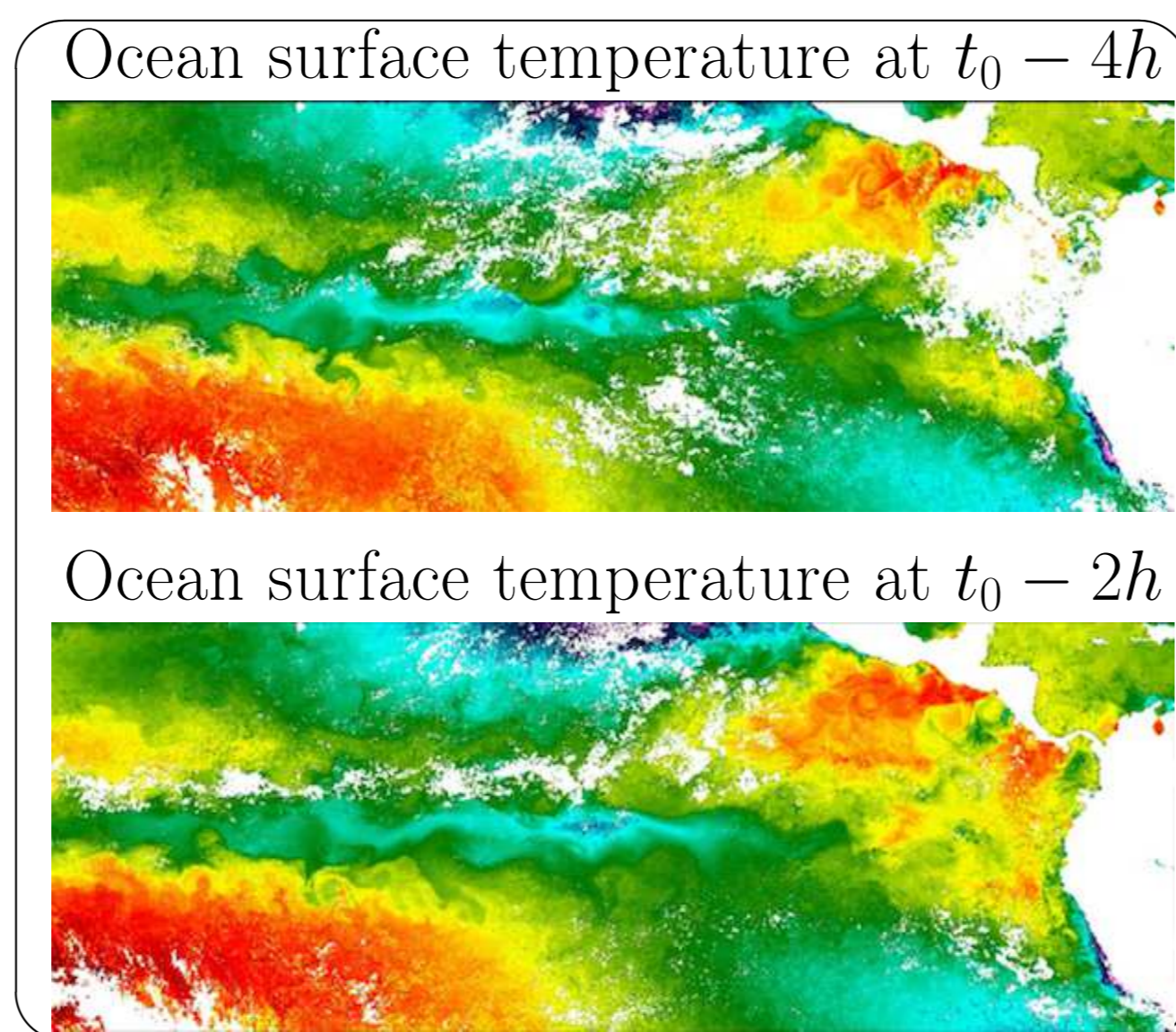
Submitted on 19 Dec 2014

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.

EXAMPLE OF OCEANOGRAPHIC DATA ASSIMILATION

Given a state \mathbf{x} whose evolution is simulated through a model \mathcal{M} , **data assimilation** is an inverse problem aiming at reconstructing an initial condition \mathbf{x}_0 of \mathbf{x} given only several partial observations $(\mathbf{y}_i^o)_i$ of \mathbf{x} . For example, the state $\mathbf{x}(t, x, y, z)$ of the ocean gathers :



- Temperature $T(t, x, y, z)$
- Velocities $\mathbf{u}(t, x, y, z)$
- Salinity $S(t, x, y, z)$
- Water height $h(t, x, y)$

The question is then: What is the complete state of the ocean at $t = 0$, $\mathbf{x}_0(x, y, z)$?

In variational data assimilation this is done by minimizing a cost function \mathcal{J} defined as

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2} \sum_i d(\mathcal{H}_i(\mathbf{x}_0), \mathbf{y}_i^o)^2 + \frac{\gamma}{2} d(\mathbf{x}_0, \mathbf{x}_0^b)^2, \quad (1)$$

$$\approx d(\mathbf{x}_0, \mathbf{x}_0^t)^2$$

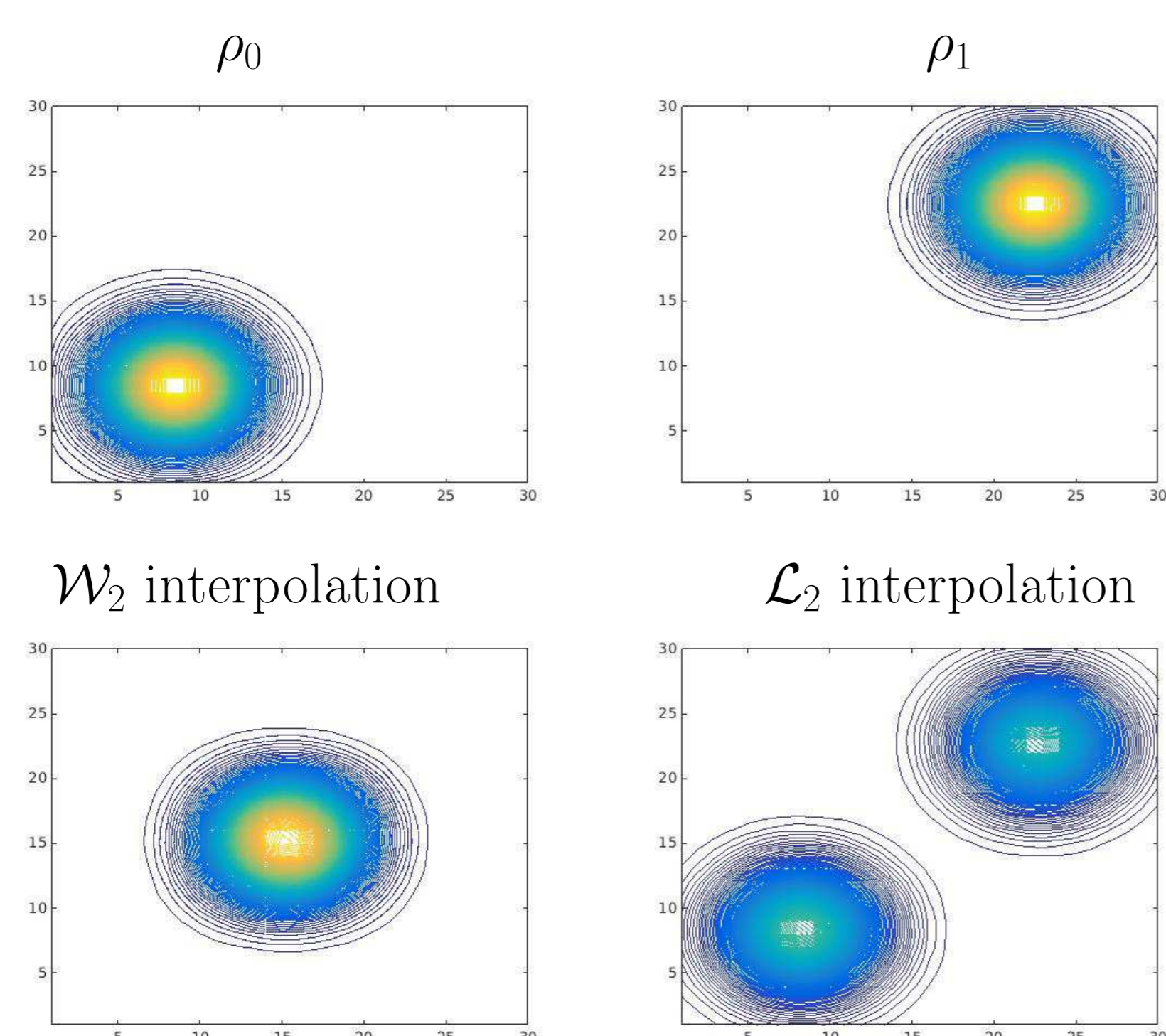
It is common for the distance d to be a weighted \mathcal{L}_2 distance. Yet, with this \mathcal{L}_2 distance, variational data assimilation cannot well cope with displacement errors. By choosing distances d that take into account the data more in its entirety, we hope get a more realistic variational data assimilation. That is why we are interested in the Wasserstein distance $d = \mathcal{W}_2$.

ADVANTAGES OF OPTIMAL TRANSPORTATION IN CASE OF DISPLACEMENT ERROR

Knowing two densities $\rho_0(x)$ and $\rho_1(x)$, the **Wasserstein distance** $\mathcal{W}_2(\rho_0, \rho_1)$ is defined as

$$\mathcal{W}_2^2(\rho_0, \rho_1) := \inf_{\substack{(\rho(t, x), v(t, x)) \\ \partial_t \rho + \operatorname{div}(\rho v) = 0 \\ \rho(0, x) = \rho_0(x), \rho(1, x) = \rho_1(x)}} \int \int_{[0,1] \times \Omega} \rho(t, x) |v(t, x)|^2 dt dx.$$

Example of interpolation using Wasserstein and \mathcal{L}_2 distances. (Interpolation in the sense of minimizing $d(\rho, \rho_0)^2 + d(\rho, \rho_1)^2$).



For Wasserstein distance to be defined, one needs $\rho_0 \geq 0$, $\rho_1 \geq 0$ and $\int_{\Omega} \rho_0 = \int_{\Omega} \rho_1$.

GRADIENT DESCENT WITH THE WASSERSTEIN DISTANCE

Assuming \mathbf{x}_0 , $\mathcal{H}_i(\mathbf{x}_0)$ and \mathbf{y}_i^o are probability measures, the cost function \mathcal{J}_W is

$$\mathcal{J}_W(\mathbf{x}_0) = \frac{1}{2} \sum_i \mathcal{W}_2^2(\mathcal{H}_i(\mathbf{x}_0), \mathbf{y}_i^o)$$

The gradient descent algorithm for minimizing \mathcal{J}_W consists in using the following iterative algorithm

$$\mathbf{x}_0^{n+1} = \mathbf{x}_0^n - \alpha \operatorname{grad} \mathcal{J}_W(\mathbf{x}_0^n)$$

so that $\mathcal{J}_W(\mathbf{x}_0^{n+1}) < \mathcal{J}_W(\mathbf{x}_0^n)$. To get the gradient, the **differential** must be computed and also an **inner product**. Indeed, $\operatorname{grad} \mathcal{J}(\mathbf{x}_0)$ is such that for all η , $(\eta, \operatorname{grad} \mathcal{J}(\mathbf{x}_0)) = D\mathcal{J}[\mathbf{x}_0].\eta$.

The differential:

Let's differentiate \mathcal{J}_W by computing $\mathcal{J}_W(\mathbf{x}_0 + \epsilon \eta)$:

- $\mathcal{H}_i(\mathbf{x}_0 + \epsilon \eta) = \mathcal{H}_i(\mathbf{x}_0) + \epsilon L[t_i, \mathbf{x}_0].\eta$ with L the tangent model.
- $\frac{1}{2} \mathcal{W}_2^2(\mathbf{y} + \epsilon \mu, \mathbf{y}') = \epsilon \langle \mu, \psi \rangle_2$ with ψ the **Kantorovich potential** of optimal transportation between \mathbf{y} and \mathbf{y}' .

Then,

$$D\mathcal{J}_W[\mathbf{x}_0].\eta = \sum_i \left\langle L[t_i, \mathbf{x}_0].\eta, \psi_i \right\rangle_2$$

with ψ_i the Kantorovich potential of the optimal transportation between $\mathcal{H}_i(\mathbf{x}_0)$ and \mathbf{y}_i^o .

The inner product:

For convergence reason, instead of using the \mathcal{L}_2 inner product, we use the \mathcal{W}_2 inner product defined in \mathbf{x}_0 by

$$(\eta, \eta') = \int_{\Omega} \mathbf{x}_0 \nabla \Phi \cdot \nabla \Phi',$$

with Φ, Φ' s.t. $\eta = -\operatorname{div}(\mathbf{x}_0 \nabla \Phi)$, $\eta' = -\operatorname{div}(\mathbf{x}_0 \nabla \Phi')$

Finally, the gradient of \mathcal{J}_W w.r.t. \mathcal{W}_2 inner product is, (L^* is the **adjoint model**),

$$\operatorname{grad} \mathcal{J}_W(\mathbf{x}_0) = -\operatorname{div} \left(\mathbf{x}_0 \nabla \left(\sum_i L^*[t_i, \mathbf{x}_0].\psi_i \right) \right).$$

ASSIMILATION OF NON-PROBABILITY MEASURE VARIABLES

In the case where $\mathcal{H}_i(\mathbf{x}_0)$ and \mathbf{y}_0 are probability measures, but *not* \mathbf{x}_0 , it is not possible to write the background term \mathcal{J}^b with \mathcal{W}_2 . For example, for the Shallow-water system

$$\begin{cases} \partial_t h + \operatorname{div}(hu) = 0 \\ \partial_t u + (u \cdot \nabla)u + g \nabla h = \mu \Delta u \end{cases} \quad (2)$$

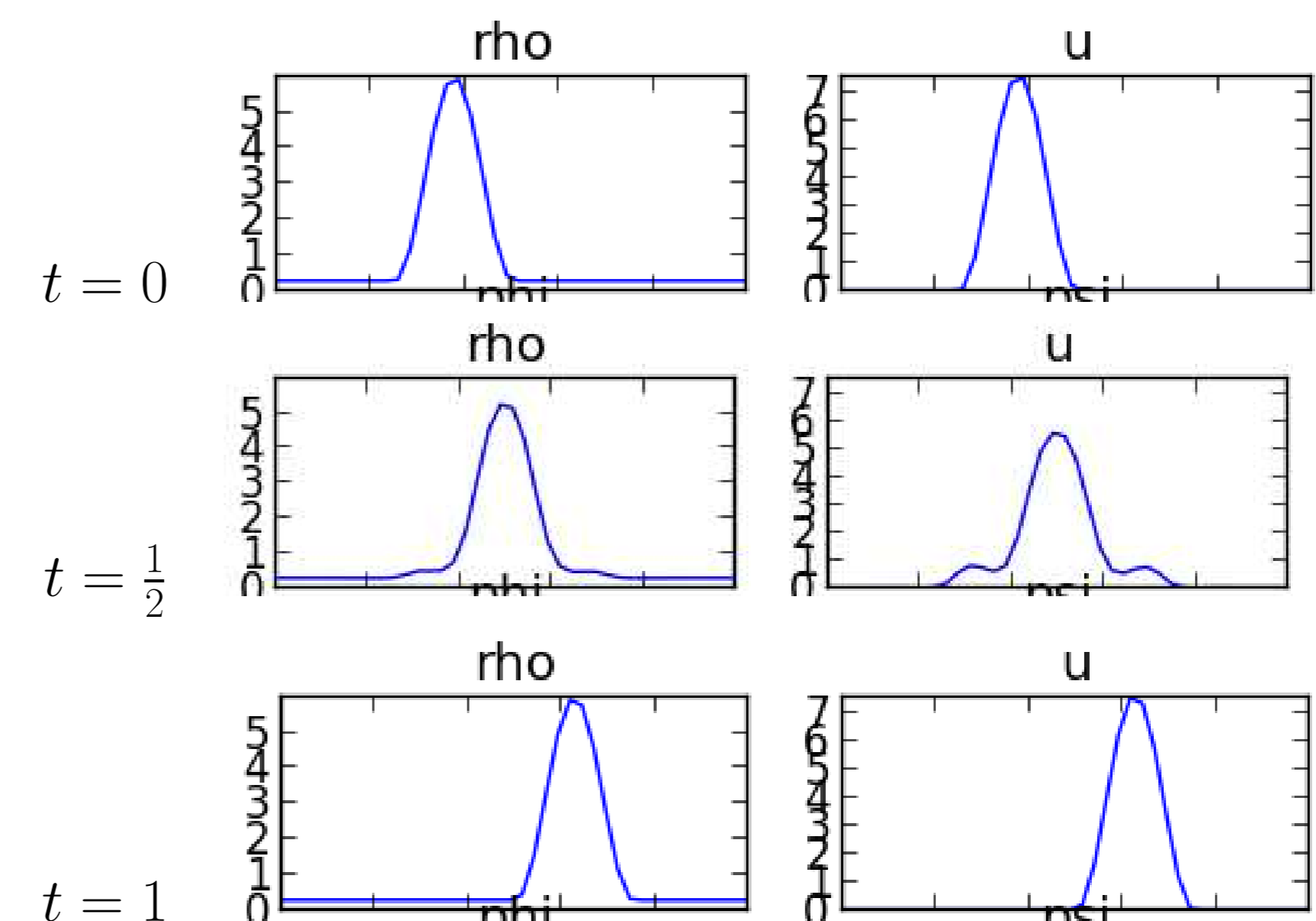
where (h_0, u_0) are to be assimilated using observations of h only. The cost function writes

$$\mathcal{J}(h_0, u_0) = \frac{1}{2} \sum_i \mathcal{W}_2^2(\mathcal{H}_i(h_0, u_0), h_i^o) + \gamma \mathcal{J}^b(h_0, u_0).$$

As it is impossible to write $\mathcal{W}_2^2(u_0, u_0^b)$, we rather use the following background term

$$\mathcal{J}^b(h_0, u_0) = \inf_{\substack{\partial_t \rho + \operatorname{div}(\rho v) = 0 \\ \partial_t u + \operatorname{div}(\rho w) = 0 \\ \rho(0, x) = h_0, \rho(1, x) = h_0^b \\ u(0, x) = u_0, u(1, x) = u_0^b}} \int \int_{[0,1] \times \Omega} (|v|^2 + |w|^2) \rho dt dx, \quad (3)$$

Using this, the interpolation of two shifted (h_0, u_0) and (h_1, u_1) is in-between:



The inner product to be chosen for computing the gradient will be

$$\left(\begin{pmatrix} \eta \\ v \end{pmatrix}, \begin{pmatrix} \eta' \\ v' \end{pmatrix} \right) = \int_{\Omega} h_0 \nabla \Phi \cdot \nabla \Phi' + \int_{\Omega} h_0 \nabla \Psi \cdot \nabla \Psi'$$

with Φ, Φ', Ψ, Ψ' such that

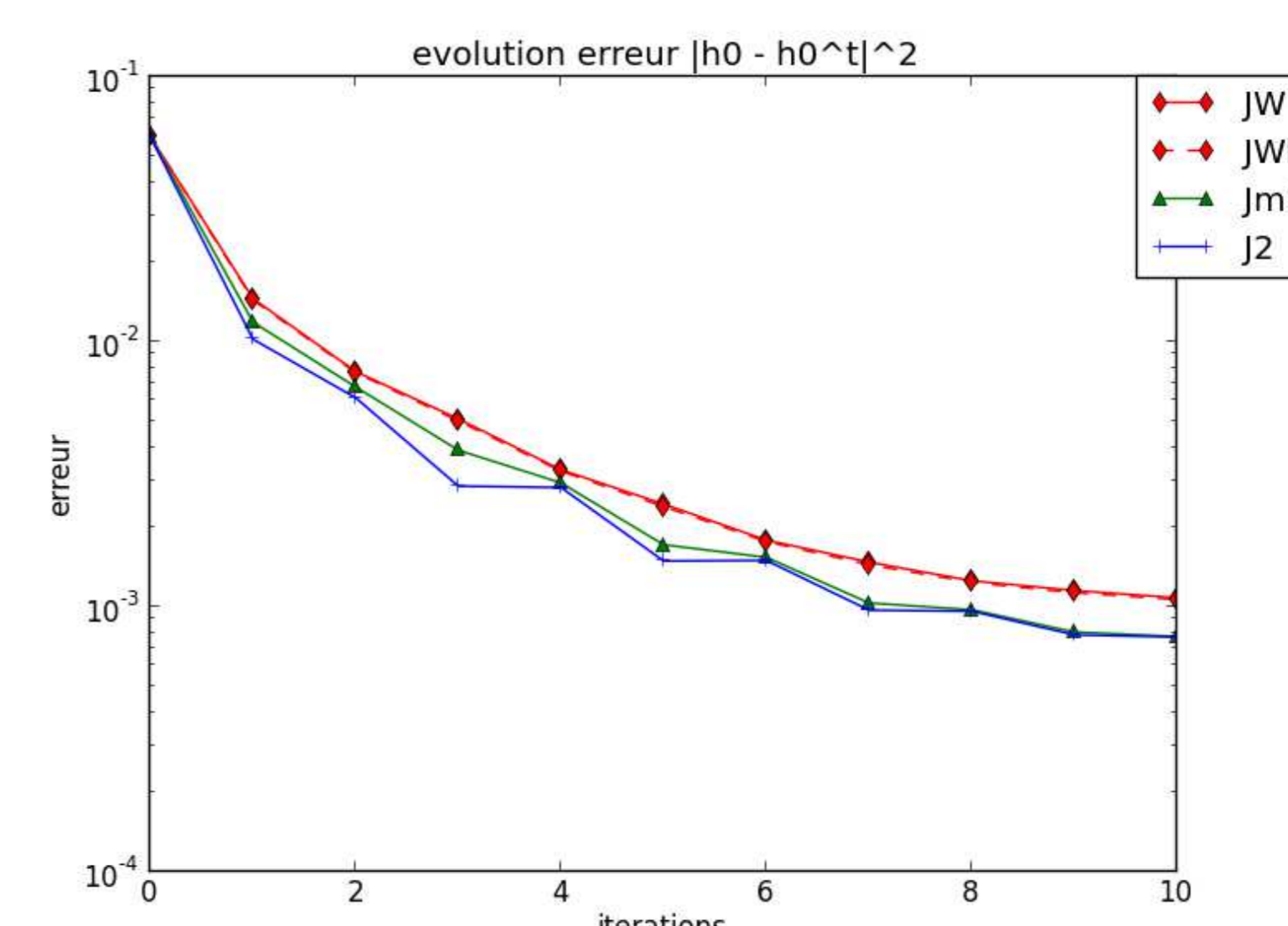
$$\begin{aligned} \eta &= -\operatorname{div}(h_0 \nabla \Phi), & \eta' &= -\operatorname{div}(h_0 \nabla \Phi') \\ v &= -\operatorname{div}(h_0 \nabla \Psi), & v' &= -\operatorname{div}(h_0 \nabla \Psi'). \end{aligned}$$

DIFFICULTIES OF USING THE WASSERSTEIN DISTANCE

- The Wasserstein distance is only defined for probability measures.
- When $\rho_0, \rho_1 \approx 1$, the \mathcal{W}_2 interpolation looks like \mathcal{L}_2 interpolation...
- When $\mathcal{J}(h_0^n) \rightarrow \min_{h_0} \mathcal{J}(h_0)$, one only has $h_0^n \rightarrow \arg \min_{h_0} \mathcal{J}(h_0)$.
- The computing time is larger for \mathcal{W}_2 than for \mathcal{L}_2 [Peyré, Papadakis, Oudet, 2013].

RESULTS AND PROPECTS

With some tests on the assimilation of (2), we compare the error $\|h_0 - h_0^b\|_2^2$ by using $d = \mathcal{L}_2$ and $d = \mathcal{W}_2$ in the cost function. The behaviors seem correct.



The background term (3) is still to be implemented.