Averaging Atmospheric Gas Concentration Data using Wasserstein Barycenters

Mathieu Barré[1], Clément Giron[2], Matthieu Mazzolini[2], Alexandre d'Aspremont[1]

[1]*INRIA, Département d'informatique de l'ENS, École normale supérieure, CNRS, PSL Research University, Paris, France.* [2]*Kayrros SAS, Paris, France.*

Reporting greenhouse gas emissions

Hyperspectral satellite images report greenhouse gas concentrations worldwide on a daily basis.
We focus on methane (CH_4) emissions, a very potent greenhouse gas with global warming potential 85 times higher than that of CO_2. Mostly two sources of anthropogenic methane emissions sources:

- **Oil and gas industry.** ⟵ our focus
- **Agriculture and waste management.**

Oil and gas emissions are concentrated in a few dense clusters around shale basins (e.g. Permian US, pipelines and major fields in Turkmenistan or Algeria...)
Short list of key emitters, which means that oil and gas sector is both a **low cost** and **high short term impact** greenhouse gas emissions mitigation target.

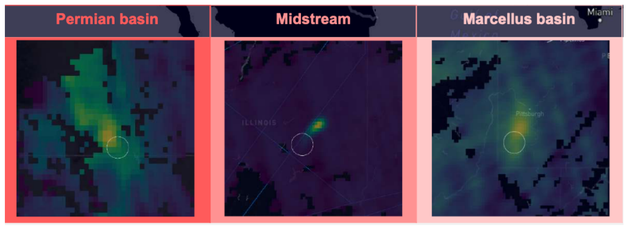


Figure 1:Observed plumes due to methane emissions in several shale basins in the US.

One cannot always observe clean plumes as in Figure 1, especially when the studied area covers a multitude of emitting sources. In this work we aim at averaging the daily concentration observations, in order to concentrate the mass around **significant emitting sources**.

Due to **atmospheric transport** simple arithmetic averages of emissions data fail to pinpoint the sources of emissions. We try using Wasserstein barycenter to address this problem.

Wasserstein Barycenter

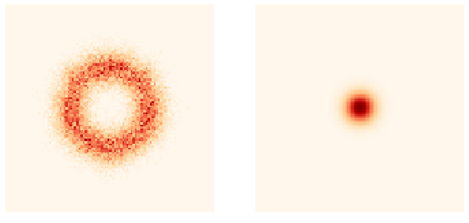


Figure 2:Data points are gaussian clouds rotating around the center. Left: Arithmetic mean. Right: Wasserstein Barycenter (with euclidean cost).

Notations $h(\cdot)$ the discrete entropy and $KL(\cdot|\cdot)$ the KL divergence.

Given a set of images $(g^k)_{k\in[1,N]} \in (\mathbb{R}_+^{n\times n})^N$ and a set of cost tensors $(C^k)_{k\in[1,N]} \in (\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n})^N$ the Wasserstein barycenter of the (g^k) is defined as

$$\bar{g} \coloneqq \operatorname*{argmin}_{g\in\mathbb{R}_+^{n\times n}} \sum_{k=1}^{N} \alpha_k W_{C^k}^2(g^k, g) \tag{1}$$

$$W_C^2(g_1,g_2) \coloneqq \min_{\Pi\in\mathbb{R}_+^{n\times n}\times\mathbb{R}_+^{n\times n}} \langle\Pi, C\rangle + \lambda h(\Pi) + \mu KL(\Pi\mathbf{1}|g_1) + \mu KL(\mathbf{1}\Pi|g_2)$$

where $\lambda,\mu\in\mathbb{R}_+$ are regularization parameters, α_k's are positive weights, $\Pi\mathbf{1}\in\mathbb{R}^{n\times n}$ (resp. $\mathbf{1}\Pi\in\mathbb{R}^{n\times n}$) with $(\Pi\mathbf{1})_{ij} = \Sigma_{k,l}\,\Pi_{i,j,k,l}$ (resp. $(\mathbf{1}\Pi)_{kl} = \Sigma_{i,j}\,\Pi_{i,j,k,l}$).

Total mass is fluctuating in the different images (emissions, missing pixels, diffusion, ...). Here we consider **unbalanced** transport with marginal relaxations (Chizat et al., 2018a).
Balanced and unbalanced optimal transport problems are closely related with fluid mechanics problems (Benamou and Brenier, 2000),(Chizat et al., 2018b). It seems well suited for our atmospheric transport problem.

Given a cost tensor C, the coefficients $C_{i,j,k,l} = c(x_{i,j}, x_{k,l})$ where $x_{i,j}$ are the coordinates in $\mathbb{R}^2$ of the center of pixel (i,j) in the concentration images, and the cost c can be defined as follow

$$\text{Eucl.} : c(x,y) = \|x-y\|^2 \tag{L_2}$$

$$\text{Eucl.-Wind} : t>0, w_k\in\mathbb{R}^2 \text{ the mean wind vector of image } g^k$$
$$c(x,y) = \|x-y\|^2 - t\langle w_k, \mathrm{Diag}(\mathrm{abs}(x-y))(x-y)\rangle \tag{L_2+W}$$

$$\text{WFR} : \delta>0,\ c(x,y) = -\log\left(\cos^2\left(\frac{\|x-y\|}{2\delta}\wedge\frac{\pi}{2}\right)\right) \tag{WFR}$$

Costs (L_2) and (L_2+W) are separable and lead to faster computations and lower memory storage. (WFR) is not, but enjoys a nice fluid mechanics interpretation (Chizat et al., 2018b).

Simulated Emissions

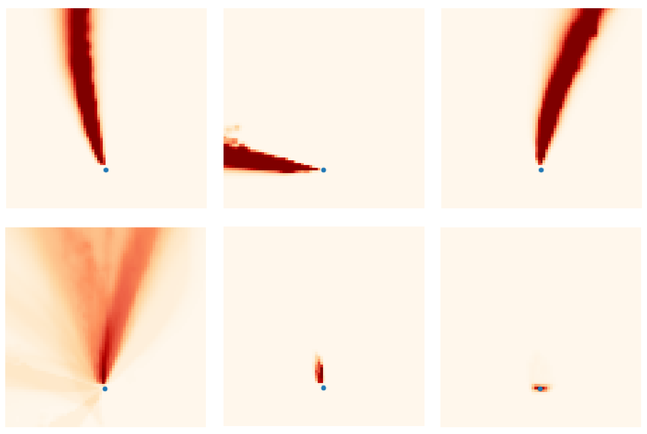
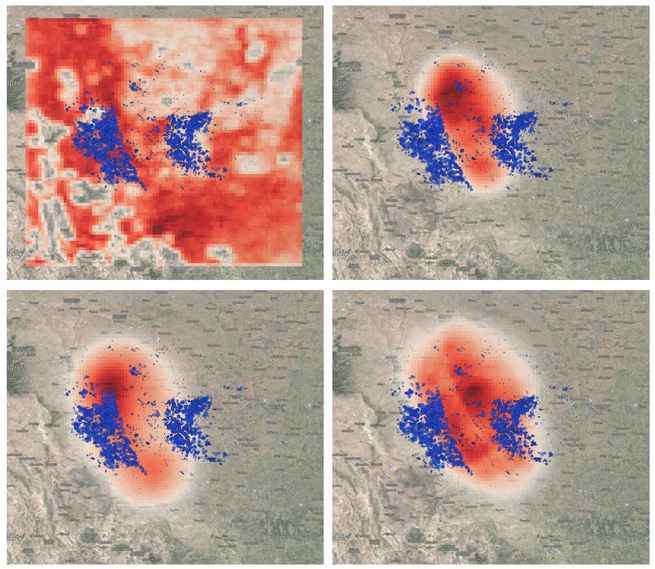


Figure 3:Top: Simulated emissions from HYSPLIT with blue dot as source. Bottom: Left- arithmetic mean, Middle- (1) with cost (L_2), Right- (1) with cost (L_2+W).

Experiments on Real Data

Permian basin



Iran-Iraq-Kuwait region

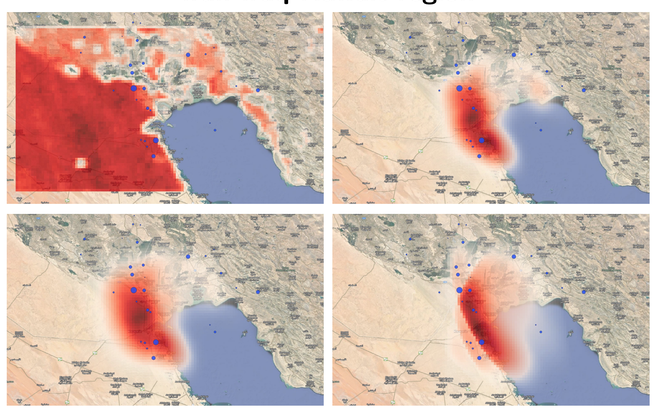


Figure 4:Blue dots are locations of recently completed oil wells (Permian) or oil fields (Iran-Iraq-Kuwait) (source: Kayrros analysis). The size of the dots represents production level. Top Left: Arithmetic mean, Top Right: (1) with (L_2), Bottom Left: (1) with (L_2+W), Bottom Right: (1) with (WFR).

References

- Chizat et al., Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, 2018a.
- Chizat et al., Unbalanced optimal transport: Dynamic and kantorovich formulations. *Journal of Functional Analysis*, 2018b.
- Benamou and Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 2000.