

Optimal transport and Data Assimilation

Dans cette note nous reprenons ce qui existe en terme d'utilisation de l'OT pour l'assimilation de données. Cela se base sur différents articles et thèse. Tout d'abord l'article de *Bocquet et al. 2023* qui présente l'OTDA comme des approches interpolante en déplacement dans l'état physique tandis que les approches classiques vont interpoler dans l'espace des intensités.

Il y a ensuite les travaux de Feyeux pour qui utilise une distance de Wasserstein pour l'assimilation de données issues d'images.

Egalement les travaux de Rivella 2007 peuvent trouver une résonnance dans une approche OT via une étape d'alignement de champs.

Un travail de W.Steven Rosenthal Displacement data assimilation et al. 2017 trouve de similarités avec notre problématique de Vortex.

Bocquet, Marc, Pierre Vanderbecken, Alban Farchi, Joffrey Dumont Le Brazidec, et Yelva Roustan. Bridging classical data assimilation and optimal transport, 2023.

<https://doi.org/10.5194/egusphere-2023-2755>.

Il rappelle tout d'abord que DA est critique pour des systèmes chaotiques pour mettre à jour les conditions initiales, estimer les paramètres physiques et statistique du modèle et permet de réanalyser l'histoire passée (smoothing). Ils présentent deux faiblesses à l'AD outre les hypothèses sur les distributions d'erreur. - *the double-penalty error*, La DA classique en sciences géophysiques présente une faiblesse majeure appelée "erreur de double pénalité". Cela se produit lorsque les erreurs du modèle et des données observationnelles sont surpénalisées, compromettant l'équilibre nécessaire. Par exemple, un léger déplacement de polluants peut entraîner des prédictions élevées là où aucun polluant n'est observé (voir image), créant des difficultés d'évaluation du modèle. Cette erreur est répandue dans les géosciences, affectant la prévision météorologique, la chimie atmosphérique, la prévision océanique, etc. L'utilisation de l'erreur quadratique moyenne aggrave ce problème, entravant l'apprentissage efficace des modèles. C'est une composante dominante à l'erreur de représentativité.

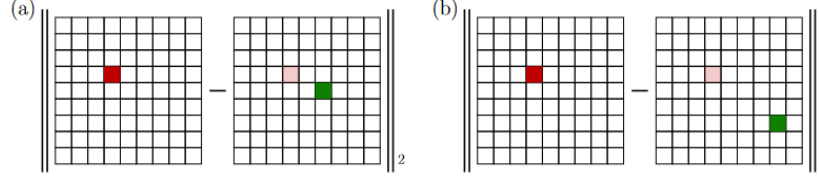


Figure 1. These two panels schematise the computation of the RMSE of two analysis increments. Those increments are the difference between the truth (left mesh within both Euclidean norm), concentrated here in the red grid-cell and the analysis located in the green grid-cells (right mesh within both Euclidean norm). The increment on panel (a) is the outcome of a better analysis spatially closer to the truth, as compared to the one in panel (b), and yet both increments yield the same RMSE. This verification metrics is hence impacted by the double-penalty error and does not help in discriminating location errors.

- La mise à jour de la DA classique peut être limitée à l'espace des valeurs des champs, ce qui entraîne une analyse confinée aux limites de l'état de fond et des observations. Ceci devient une lacune majeure lorsque le désaccord entre les observations et l'état de fond résulte d'une erreur de localisation des champs ou d'une mauvaise spécification générale de ces champs. La Figure 2 illustre des expériences de DA classique avec des analyses bénéfiques et inutiles.

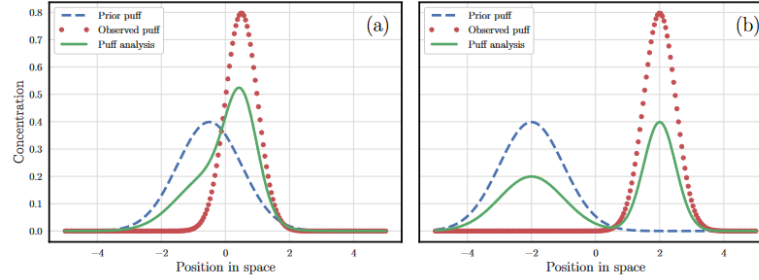


Figure 2. These two panels illustrate a reasonably beneficial and a probably useless classical DA update. It is assumed that the observations (red dots) and first guess (dashed blue curve), which represent one-dimensional puffs of pollutant, are subject to a location mismatch. With a significant overlap of these fields, panel (a) displays a consistent analysis. By contrast, in panel (b), the analysis state (green curve) fails to propose a state with a significant presence in between the observations and first guess, just as expected since classical DA is based on an interpolation in field values, but neither in space, nor in time.

On peut considérer que l'erreur provient du choix du RMSE comme mesure d'erreur, qui est une mesure localisée à une position. Une mesure plus intelligente consiste en la combinaison d'une carte de déplacement suivi de la norme classique RMSE (Hoffman et al., 1995; Keil and Craig, 2009). Des métriques basées sur des approches on pu être proposées. Une approche élégante consiste à appliquer la théorie du *transport optimal* qui est associée à la *distance de Wasserstein*. Des exemples d'utilisation d'OT peuvent être trouvés chez Farchi et al. (2016) et Vanderbecken et al. (2023).

Posons quelques notations pour le transport optimal. Le problème consiste à trouver le plan au coût minimal de transporter la mesure ρ_0 à ρ_1 . Mesures donc non négative et d'intégrale à 1. Un choix d'intégral peut être aussi écrit. De

plus, des généralisations pour des masses différentes ont également été faites.

On définit un coût défini pour tout couple de point $(x, y) \in \Omega^2$ généralement fonction de la distance, on présente le coût quadratique

$$\mathcal{C}_{bo}(x, y) = \|x - y\|_2^2.$$

On introduit l'ensemble des plans de transformations différentiables admissibles T

$$\mathcal{U}_{bo} = \{T : \Omega \mapsto \Omega, \quad \rho_0 = |\partial_x T| \rho_b \circ T\},$$

où $|\partial_x T|$ est le déterminant de la jacobienne de T qui prend compte de la déformation de la mesure par la conservation totale de la masse.

On peut alors définir le carré de la distance de *Wasserstein* comme

$$\mathcal{W}_{\mathcal{C}_{li}}^2(\rho_0, \rho_1) = \min_{T \in \mathcal{U}_{bo}} \int_{\Omega} \mathcal{C}_{bo}(x, T(x)) \rho_0(x) dx,$$

Dont le but est de minimiser le coût total de transport d'une mesure à l'autre. D'un point de vue pratique, il n'est pas possible de résoudre le problème formulé avec un plan de transport déterministe comme énoncé par Monge. Kantorovitch introduit une formulation probabiliste au XX^e siècle. Le plan de transport est remplacé par une mesure de probabilité π définie sur $\Omega \times \Omega$.

L'espace des transformations admissibles devient

$$\mathcal{V}_{bo} = \{\pi : \Omega \times \Omega \mapsto \mathbb{R}^+, \text{quad} \rho_o(x) = \int_{\Omega} \rho_b(y) \pi(x, y) dy, \quad \rho_b(y) = \int_{\Omega} \rho_o(x) \pi(x, y) dx\}.$$

La distance de Wasserstein au carré devient alors

$$\mathcal{W}_{\mathcal{C}_{li}}^2(\rho_0, \rho_1) = \min_{T \in \mathcal{V}_{bo}} \int_{\Omega \times \Omega} \mathcal{C}_{bo}(x, y) \pi(x, y) dx dy.$$

Non local, multiscale metrics and DA

On trouve des métriques sur l'erreur de déplacement ont pu être introduites par Ravela et al. (2007); Plu (2013). De même des approches multiéchelles Ying (2019); Ying et al. (2023). On trouve des formes proches de la distance de Wasserstein développées pour transporter la distribution du prior vers le posterior (El Moselhy and Marzouk, 2012; Oliver, 2014; Marzouk et al., 2017; Farchi and Bocquet, 2018; Tamang et al., 2020). On peut l'utiliser aussi pour ajuster les poids du filtre particulaire (Tamang et al., 2021, 2022). Elle a également été utilisée pour comparer des ensembles de forecast (Le Coz et al., 2023; Lledó

et al., 110 2023), or precipitation (Duc and Sawada, 2023). Aujourd’hui on ne traite que des cas de pdf de variables aléatoires scalaire. A cause du problème du fléau de la dimension. Dans notre cas, on traite des états qui sont des champs, pas la pdf d’une seule variable. Dans les cas de champs de dimension 2 ou 3.

Le papier actuel se focalise plutôt en l’utilisation de la distance de Wasserstein lors de la formulation variationnelle de l’analyse.

On part ici de la fonction de coût du 3D-Var (avec la distance Euclidienne)

$$\mathcal{J}_{cl}(\mathbf{x}^a) = \|\mathbf{y}^b - \mathbf{x}^a\|_2^2 + \|\mathbf{y}^o - \mathbf{H}\mathbf{x}^a\|_2^2,$$

où \mathbf{y}^b est la prédiction du forecast, où \mathbf{y}^o est le vecteur des observations. En substituant la norme Euclidienne par la norme de Wasserstein on obtient la fonction de coût suivante

$$\mathcal{J}_w(\mathbf{x}^a) = \mathcal{W}_2^2(\mathbf{y}^b, \mathbf{x}^a) + \mathcal{W}_2^2(\mathbf{y}^o, \mathbf{H}\mathbf{x}^a).$$

Il est nécessaire d’équilibrer les deux distances. On appelle l’état d’analyse le *barycentre de Wasserstein*.

On remarque dans les travaux de Feyeux que cette formulation peut être inconsistente. Dans le cas où l’état est complètement observé (\mathbf{H} est l’identité), l’assimilation se déroule sans encombre. Une inconsistance s’observe lorsque l’état est partiellement observé. (Note: Dans notre cas, je pense que si l’observation est homogène alors tout va bien).

La principale raison à cela est de supposer que les densités d’origine et objectif ont la même masse. Ce qui s’applique pour les deux termes de calcul de distance de Wasserstein, à la fois que $m(\mathbf{x}^a) = m(\mathbf{y}^b)$ et $m(\mathbf{H}\mathbf{x}^a) = m(\mathbf{y}^o)$. En supposant finalement que $m(\mathbf{y}^b) = m(\mathbf{y}^o)$, l’égalité $m(\mathbf{x}^a) = m(\mathbf{H}\mathbf{x}^a)$ est difficilement atteignable pour $\mathbf{x}^a \in \mathbb{R}^{N_a}$ *textker*(\mathbf{H}). Il faut pour cela proposer une méthode qui puissent s’accomoder d’avoir des densités de masses différentes. Une résolution a été proposée par *Chizat et al. (2018)*. Ici une solution orientée DA.

Proposition

Hypothèses: champs physiques non-négatifs, mais pas nécessairement l’ébauche et l’observation et H est supposé linéaire.

Sur la partie 2.5

Ils ont introduit l’équation (18) dans leur papier qui est une nouvelle fonction coût

$$L_w(\mathbf{x}^a) = \min_{\mathbf{x}^b \in \mathcal{O}_{N_b}^+, \mathbf{x}^o \in \mathcal{O}_{N_o}^+} \{\zeta_b(\mathbf{y}^b - \mathbf{x}^b) + \zeta_o(\mathbf{y}^o - \mathbf{H}\mathbf{x}^o) + W_{C_{ba}}(\mathbf{x}^b, \mathbf{x}^a) + W_{C_{oa}}(\mathbf{x}^o, \mathbf{x}^a)\}.$$

Pour résoudre le problème, on utilise une version simplifiée. On ne se base que sur deux transferts différents $\mathbf{P}^{ba}, \mathbf{P}^{oa}$. On régularise pour obtenir

$$\begin{aligned} & \min_{\mathbf{x}^b \in \mathcal{O}_{N_b}^+, \mathbf{x}^o \in \mathcal{O}_{N_o}^+, \mathbf{x}^a \in \mathcal{O}_{N_a}^+} [\zeta_b(\mathbf{y}^b - \mathbf{x}^b) + \zeta_o(\mathbf{y}^o - \mathbf{H}\mathbf{x}^o) + W_{C_{ba}}(\mathbf{x}^b, \mathbf{x}^a) + W_{C_{oa}}(\mathbf{x}^o, \mathbf{x}^a) \\ & + \min_{\mathbf{P}^{ba} \in \mathcal{U}_{ba}, \mathbf{P}^{oa} \in \mathcal{U}_{oa}} \{\varepsilon \mathcal{K}(\mathbf{P}^{oa} | \mathbf{v}^{oa}) + \varepsilon \mathcal{K}(\mathbf{P}^{ba} | \mathbf{v}^{ba}) + \text{Tr}(\mathbf{C}_{oa}^T \mathbf{P}^{oa}) + \text{Tr}(\mathbf{C}_{ba}^T \mathbf{P}^{ba})\}]. \end{aligned}$$