Authors' response to Reviewer 2' s comments for "Bridging classical data assimilation and optimal transport"

Preprint egusphere-2023-2755

M. Bocquet , P. J. Vanderbecken, A. Farchi, J. Dumont Le Brazidec, and Y. Roustan

13 March 2024

The paper proposed a different objective function for the data assimilation problem, which is a mixed sum between the classic cost function and the Wasserstein metrics. With assumptions on strictly convexity and linearity, the problem is convex. The paper then derives the optimization conditions for the new objective function by combining convex optimization analysis and the duality concerned about the Wasserstein metric with entropy regularization. It is further supported by several 1D and 2D data assimilation problems.

We appreciate the reviewer's comments and suggestions. In the following, we discuss the raised concerns and what we have changed or will revise in the manuscript.

(1) My biggest complaint is the paper title. I think the title does not match the content of the paper. Based on the title, it sounds like a providing a theoretical connection between classical data assimilation and OT. However, it is more like introducing an application of OT theory for data assimilation problems. I suggest the authors change the title of this paper to a more descriptive one.

We have presented this work to many colleagues (in geophysical data assimilation with some knowledge of optimal transport) who did not feel this issue. But we can certainly make the title of the paper more specific. We intend to choose "**Bridging classical data assimilation and optimal transport: The 3D-Var case**" since we are proposing the OTDA 3D-Var and a way to carry out its analysis for both first-order and second-order moments.

(2) The assumption in equation (11) is rather strong. H is linear, and all the costs are convex. That means the paper only deals with log-concave distributions, which does not apply to the multi-modal distributions that are challenging to handle. The baseline DA (17) is a strictly convex problem that can be solved easily. Even if one is worried about noise overfitting, many existing good methods exist to handle this.

The fact that $\mathbf{H}$ is linear is for convenience because it can make the dual optimal problem formally equivalent. This could be alleviated, by for instance sticking to a primal formulation. In any case, it must be kept in mind that assuming $\mathbf{H}$ linear is a common practice (with good reasons) in classical data assimilation through explicit linearisation, or implicitly as in the EnKF. For instance in one of the seminal paper of data assimilation (Courtier, 1997) about the dual formulation of 4D-Var (relevant to and cited in the manuscript), the authors assume linearisation of $\mathbf{H}$ from the very beginning of the paper owing to the so-called *incremental* formulation of 4D-Var.

So this assumption is actually not as shocking as it may seem. But we agree that as such, this is a strong assumption, out of the data assimilation context. Following your concern, we have added a bit a context in the revised manuscript to support the assumption.

However, the fact that the problem is convex, or can be considered so, is a completely standard assumption to classical geophysical data assimilation. Even though the standard methods DA applied to high-dimensional systems (3D-Var, 4D-Var, EnKF) are not rigorously convex, they are in practice assumed to be so with one single outstanding minimum. Almost all the methods in operation rely on a linearisation and implicit Gaussian assumptions (see, e.g., Carrassi et al., 2018, and references within). Moreover, all the incremental variational formulations of 3D–Var and 4D–Var use a truly strongly convex cost function in their inner loop so as to leverage a conjugate gradient optimisation scheme. They would fail in presence of true multi-modality. With the exception of the linearity of $\mathbf{H}$, the statistical assumptions in Eq. (11) are actually more general that those used in classical 3D-Var (Daley, 1991).

We note that the problem of strong nonlinearity, non-Gaussianity, and multi-modal distributions, a problem the authors are not unfamiliar with (e.g., Bocquet et al., 2010; Farchi and Bocquet, 2018), is really far apart from the goal of this paper which is merely to replace local metrics with nonlocal ones in classical DA.

Finally, to answer the end of the comment, the objective of the paper is not to come back to the well-known (numerical) solution of the classical 3D-Var, but to address the extended hybrid OTDA problem.

(3) Between (17) and (18), there are new introductions of $\mathbf{x}^{\mathrm{b}}$ *and* $\mathbf{x}^{\mathrm{o}}$ *with Wasserstein metrics, turning the problem from strongly convex to a mixed problem. The motivation is not very clearly stated. The significantly increased computational cost associated with (18) has to be supported by very strong reasons. For example, what properties can we achieve by combining these two different cost functions?*

This *strongly convex problem* is the basic 3D-Var cost function of classical data assimilation, which is very common in geophysical data assimilation, which we indeed turn into a mixed problem to tackle the double-penalty, location error issues.

The main motivation for introducing non-local metrics into classical data assimilation is that of model error, and especially model error based on location errors, which is ubiquitous in the geosciences and leads to the so-called *double-penalty error*. Following your comment and that of reviewer 1, we have clarified the introduction (which is key to understanding why one would move from Eq. (17) to Eq. (18)), especially its beginning which could have been misleading in that regard. Note that this motivation was put forward in a long stream of works mentioned in the introduction, e.g., Hoffman et al. (1995); Ravela et al. (2007); Ning et al. (2014); Feyeux (2016) who also motivate the introduction of non-local metrics in the fields of verification and data assimilation. We are to a very large extent making this motivation ours.

(4) Section 2.4 has too many details about the derivation that are standard steps in convex optimization. I suggest putting many in an appendix and only stating the main formula.

The reviewer may have a strong mathematical/computer science background, which is not very common to many experts in geophysical data assimilation. But we agree that the manuscript could benefit from moving a few derivations in an appendix, which we did in the revised manuscript following your advice.

(5) The numerical examples in Sections 3 & 4 are a bit too simple. Of course, 1D and 2D OT are not so costly. However, when the dimension becomes large, the extra two terms in (18) become increasingly cumbersome, and computational cost is forbidden. This work is for geoscience applications with often high-dimensional state space.

We disagree with the reviewer. Firstly, the example that we provide are numerically significantly more challenging than those provided earlier on the same attempt to bridge classical data assimilation with optimal transport (Feyeux, 2016; Feyeux et al., 2018). Secondly and most importantly, the reviewer is mistaking the dimension of the vector space stemming from the discretised physical space for the dimension of the physical space. **Although the physical space is here**

of dimensions 1 or 2, the discretised state space dimension goes here up to 10^4 (to be multiplied by the number of control variable types, which is 3 here). Note that in geophysical data assimilation, *high-dimensional* refers to the dimension of the discretised vector space, not that of the underlying physical space (which is 1, 2, or 3, possibly but rarely 4).

Overall, I feel the paper title is too big of a summary for the paper, and the numerical examples, on the other hand, are elementary. While I can relate the formulation from (6) to (7), which is very neat and has a clear mathematical intuition, the hybrid sum in (18) seems to be a "cocktail" of two different metrics. Further understanding is necessary even if the authors don't plan on proving any mathematical properties.

Following your recommendation, we have changed the title. We believe that it is now really faithful to what is developed in the paper.

We understand your feeling as to the *cocktail of metrics*. But, please understand that our formalism is rather rigorous for the geophysical data assimilation and is actually shown to improve upon first attempts by Feyeux (2016). It also has the very pleasant property that classical data assimilation was shown to be embedded in the OTDA formalism, as opposed to the theories previously advocated. To us, this already makes a strong case for this cocktail.

Finally, to strengthen this approach, we have developed a consistent Bayesian and probabilistic view on the OTDA formalism, a view often supported in classical data assimilation to back the methods (Lorenc, 1986). However, we had the feeling it would have been too long and technical to be reported here, and only marginally relevant to *Nonlinear Processes in Geophysics* so that is was not included in the manuscript.

References

Bocquet, M., Pires, C. A., and Wu, L.: Beyond Gaussian statistical modeling in geophysical data assimilation, Mon. Wea. Rev., 138, 2997–3023, https://doi.org/10.1175/2010MWR3164.1, 2010.

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data Assimilation in the Geosciences: An overview on methods, issues, and perspectives, WIREs Climate Change, 9, e535, https://doi.org/10.1002/wcc.535, 2018.

Courtier, P.: Dual formulation of four-dimensional variational assimilation, Q. J. R. Meteorol. Soc., 123, 2449–2461, https://doi.org/10.1002/qj.49712354414, 1997.

Daley, R.: Atmospheric Data Analysis, Cambridge University Press, New-York, 1991.

Farchi, A. and Bocquet, M.: Review article: Comparison of local particle filters and new implementations, Nonlin. Processes Geophys., 25, 765–807, https://doi.org/10.5194/npg-25-765-2018, 2018.

Feyeux, N.: Transport optimal pour l'assimilation de données images, Ph.D. thesis, Université Grenoble Alpes, `https://inria.hal.science/tel-01480695`, 2016.

Feyeux, N., Vidard, A., and Nodet, M.: Optimal transport for variational data assimilation, Nonlin. Processes Geophys., 25, 55–66, https://doi.org/10.5194/npg-25-55-2018, 2018.

Hoffman, R. N., Liu, Z., Louis, J.-F., and Grassoti, C.: Distortion representation of forecast errors, Mon. Wea. Rev., 123, 2758–2770, https://doi.org/10.1175/1520-0493(1995)123¡2758:DROFE¿2.0.CO;2, 1995.

Lorenc, A. C.: Analysis methods for numerical weather prediction, Q. J. R. Meteorol. Soc., 112, 1177–1194, https://doi.org/10.1002/qj.49711247414, 1986.

Ning, L.and Carli, F. P., Ebtehaj, A. M., Foufoula-Georgiou, E., and Georgiou, T. T.: Coping with model error in variational data assimilation using optimal mass transport, Water Resources Research, 50, 5817–5830, https://doi.org/https://doi.org/10.1002/2013WR014966, 2014.

Ravela, S., Emanuel, K., and McLaughlin, D.: Data assimilation by field alignement, Physica D, 230, 127–145, https://doi.org/10.1016/j.physd.2006.09.035, 2007.