Authors' response to Reviewer 1'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

26 February 2024

The paper discusses a very interesting and pertinent problem and proposes a solution that is certainly worth looking at further. The continued use of quadratic error functional is somewhat anachronistic given that the main motivation for using it are simplicity of analytic computations and an assumption of Gaussian errors. The first argument carries less weight in the age of supercomputers, and the second was always known to be wrong except in (important) special cases. I cannot, however, recommend the publication of the paper in its present form and believe that the paper requires a major revision. This is due to the following major concerns (see below for a few more minor concerns)

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) I find the first section very confusing. The authors are discussing problems related to the so-called double penalty error and the non-overlap of functions (or distributions) that appear in data assimilation. It is not clear however on what level the authors are working. More specifically, it seems first that the authors want to work on the level of probability distributions for state variables. Later however it turns out that they want to work on the level of state variables directly, yet focus on those that represent meteorological fields which essentially have the character of distributions. This, however, has to be clear from the very beginning.

It has been made clear in the manuscript that the theory applies to fields, not probability distributions. For instance in section 1.4:
In the context of this paper, it is critical to be aware that the use of OT in practical DA focused so far on applying OT to the pdf of a single variable. Quite often, OT is applied to the pdf of a single random variable because... This is very different from our context and objective where the objects dealt with by OT are field states, not the pdf of one of their variables.

Or, in section 1.5, Objective and Outline, we state:
At least within the perimeter of this paper, some restrictions apply compared to traditional DA. Firstly, the physical fields considered in the DA problem are non-negative (concentration of tracer, pollutants, water vapour, hydrometeors, chemical and biogeochemical species, etc.). However, as opposed to Feyeux (2016), the methods of this paper do not require the (possibly noisy) background state $\mathbf{y}^{\mathrm{b}}$ and observation $\mathbf{y}^{\mathrm{o}}$ to be non-negative. Secondly, the observation operator $\mathbf{H}$ is assumed to be linear. This is only meant for convenience and to obtain a rigorous correspondence between the primal and dual cost functions of the 3D-Var. Finally, we stress once again that the states of our DA problem are physical fields onto which OT is applied and are not meant to be pdf of a random variable.

That said, we fully agree with the reviewer that the second motivational argument (overlapping of fields) could generate a lot of confusion and could even be misleading, especially so early in the manuscript. We explain below with the third point raised how we addressed this issue.

(2) I understand what the authors label as the first (of two) weaknesses of classical DA which is often termed the double penalty error. This problem however is ultimately an issue resulting from a mismatch between the employed distance and the smoothness of meteorological fields. If the correct metric is selected depending on the smoothness of the meteorological fields, there is no double penalty problem. Labelling this as a problem of "classical DA" however implies that classical DA only uses the mean square error which is not correct (error covariances are important part of the error functional and have a strong influence on whether the double penalty problem occurs or not).

In the light of your comments, we have decided to renounce calling the main issue (double penalty error) a weakness of classical DA since it is instead a weakness in the choice of the metrics which, as a result, impacts data assimilation. We hope the revision will be clearer in that regard.

However, we would like to stress that the double-penalty issue, which comes from a (model) location error of fields, does not disappear by a smart choice of a weighted Euclidean distance (Mahalanobis distance). It is true that inflating the correlation length of the covariance matrices is a known trick to mitigate the double penalty error but it does not solve it, since it cannot correct for the location of fields (e.g., Hoffman et al., 1995). The weighted Euclidean distance only measures amplitude and smoothness mismatches, not the full *distorsion*. This was a key incentive for the stream of work mentioned in the introduction (e.g.. Hoffman et al., 1995; Hoffman and Grassotti, 1996; Ravela et al., 2007; Ning et al., 2014; Feyeux et al., 2018). Here is a thought experiment to stress this: one can imagine inflating the correlation length of the covariance matrix of a weighted Euclidean distance of two mislocated peaked fields which are far apart. But it could become a blunt tool with poor separation power. Indeed, it is not difficult to check that in the very large correlation length limit, such norm only compares the mean of the two fields, i.e. only one degree of freedom.

This is now clarified/further discussed in the motivational section 1.1 of the introduction.

(3) Related to the previous question, I do NOT understand what the authors label as the second weakness of classical DA. In fact, the last paragraph of Sec. 1.1 hardly makes any sense when it talks about "overlap in space and time" between background and observations. The material seems to draw on intuition coming from the Bayes rule but that applies to probability densities; the 3D-Var analysis is an operator-convex combination of meteorological fields which is something completely different.

We agree that the motivational argument is too confusing, especially at the very beginning of the manuscript. Hence, in the light of your comments, we have decided to withdraw it. Hence section 1.1, now entitled *The double-penalty issue and data assimilation* has been rewritten and is now mainly focused on the double-penalty error. Thank you for pointing out the lack of clarity of the original section 1.1.

(4) As far as I can see, the technique can only assimilate meteorological fields that essentially behave like distributions, which is clearly a major restriction. I believe the assumption that the fields be positive is not enough (after all, by choice of origin any meteorological field can be assumed to have non-negative values). All the examples mentioned by the authors are extensive quantities. Is there an issue with applying the approach to an intensive quantity such as, say, the temperature?

Yes, the technique has a major limitation in that it is meant for non-negative fields. We have been very upfront about this. Note, however, that we have introduced a way to address the unbalanced case where the mass of the origin and target fields are not equal. That is why we speak of non-negative fields, not distributions (which is too restrictive).

Yes, the assumption that the fields are non-negative is sufficient, at least mathematically. If one wishes, the origin of fields could be redefined to make them non-negative and our theory could be applied. But of course such trick would introduce major biases that would force to reconsider how to apply data assimilation. And what would be the origin of fields for wind components? That is why we prefer to emphasise potential applications to chemical species fields, biogeochemical fields, water vapour, etc., where location errors can be significant.

No, there is no issue applying OTDA to temperature. But of course, temperature fields are known to be much smoother; they do not exhibit strong location errors, compared to, say, the relative or specific humidity (which are intensive fields as well). Hence, it is a priori more interesting to apply OTDA to such humidity fields where, for this reason, the double-penalty error is more prominent, and perhaps for fields with exact or approximate conservation laws. But these are not necessary conditions for the applicability of the tool.

(a) It is not clear how the authors deal with comparing fields that do not have the same mass. Also, it is not clear what the cumbersome result in Fig. 5 has to do with the assumption whether or not the two fields have the same mass.

The reviewer refers to Figure 5, where the solution that we proposed has not been introduced yet. At this stage in the paper, we only state the nature of the problem, and explain that it was deemed so far unclear how to compare fields that do not have the same mass. Our solution is only introduced later on, starting from section 2.3. We stress this in the revised manuscript.

As for the second comment, the assumption that $\mathbf{y}^{\mathrm{b}}$ and $\mathbf{y}^{\mathrm{o}}$ have the same mass was merely for convenience; this is not crucial to the argument. We have clarified the corresponding statement and sentence. (But the fact that OT – not OTDA –conserves mass by definition is of course critical to the derivation we provide.)

(b) The concept of the entropy regularisation is not clear. It is not even clear why this renders the problem convex or at least uniquely solvable.

We follow a classical technique used by the applied mathematicians in the field of numerical optimal transport. As clearly mentioned when introducing the technique, we refer to the nice and accessible textbook of Peyré and Cuturi (2019):
Hence, entropic regularisation is used to render the problem strictly convex. A comprehensive justification is given by Peyré and Cuturi (2019).

We have extended the sentence referencing their book and now precisely points to where the proof of convexity is given if it can help. That said, since the proof is simple and enlightening, we can give it in this response and in the revised manuscript. Indeed, the regularised problem has a cost function of the form

$$\varepsilon\mathcal{K}(\mathbf{p}|\mathbf{q}) + \mathbf{c}^{\mathsf{T}}\mathbf{p} = \varepsilon\sum_i p_i \ln\left(\frac{p_i}{q_i} - p_i\right) + q_i + \sum_i c_i p_i \tag{1}$$

to be minimised over the p_i, and where $\varepsilon > 0$. The Hessian of this cost function is readily computed and it is diagonal with entries

$$\varepsilon/p_i > 0, \tag{2}$$

which, since $0 \leq p_i \leq 1$, shows that the optimisation problem is ε–strongly convex and in particular strictly convex as long as $\varepsilon > 0$.

(c) There are other ways to measure distances between meteorological fields that avoid or alleviate the double penalty error (depending on the smoothness of the fields). Have the authors compared their Wasserstein approach with other metrics, also given that the entropy regularised Wasserstein distances are not easy to calculate and optimise?

We did not, since the main focus of the paper was to show how to merge classical data assimilation and the optimal transport. One of the reason why we chose optimal transport over other approaches is that there is a large community of applied mathematicians that built solid foundations for this tool, and recently made progress on the numerical implementation (Peyré and Cuturi, 2019; Flamary et al., 2021), so that computing the Wasserstein distance is not as challenging as it used to be. Further, in order to make a meaningful comparison, we would need a concrete, practical case with real data, which is beyond the scope of this paper.

(d) On pg 6 the authors claim that "our problem is not subject to the curse of dimensionality". Although this is true, the curse of dimensionality (in the sense implied by the discussion here) is not actually a concern in 3D-Var either which is what the method should be compared with. So this remark is misleading.

The remark was just intended for readers that use optimal transport theory and apply it to probability distributions in the field of geosciences (which includes some of us) and may have thought that the present theory is impacted by the curse of dimensionality as well. Nothing more was intended, especially in the scope of this introduction.

References

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.

Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., K., F., Fournier, N., Gautheron, L., Gayraud, N. T. H., Janati, H., Rakotomamonjy, A., Redko, I., Rolet, A., Schutz, A., Seguy, V., Sutherland, D. J., Tavenard, R., Tong, A., and Vayer, T.: POT: Python Optimal Transport, Journal of Machine Learning Research, 22, 1–8, http://jmlr.org/papers/v22/20-451.html, 2021.

Hoffman, R. N. and Grassotti, C.: A Technique for Assimilating SSM/I Observations of Marine Atmospheric Storms: Tests with ECMWF Analyses, Journal of Applied Meteorology and Climatology, 35, 1177–1188, https://doi.org/10.1175/1520-0450(1996)035¡1177:ATFASO¿2.0.CO;2, 1996.

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.

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.

Peyré, G. and Cuturi, M.: Computational Optimal Transport: With Applications to Data Science, Foundations and Trends® in Machine Learning, 11, 355–607, https://doi.org/10.1561/2200000073, 2019.

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.