Authors' response to the Editor's comments on the revised manuscript "Bridging classical data assimilation and optimal transport: The 3D-Var case"

Preprint egusphere-2023-2755.R1

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

17 May 2024

We appreciate the Editor's commitment in reviewing the manuscript, his comments and suggestions. In the following, we provide answers to the questions and suggestions and describe what has been changed accordingly in the manuscript. The pdf document highlighting the differences between the original and the revised manuscript is provided.

(1) Figure 14. Contrary to what the authors write, no error bars are visible on panel (a). The panel is actually identical with panel (a) of Fig. 11.

By error bars we meant the shades encompassing each curve and whose boundaries correspond to plus and minus the standard deviations about each curve. These are the key additions compared to Fig. 11 (they are based on the same test configuration but are not identical). We modified the caption to be clearer:

... but with the addition of shaded regions delineated by plus and minus the standard deviations about the estimates for $\mathbf{x}^{\mathrm{a}}, \mathbf{x}^{\mathrm{a}}_{\mathrm{cl}}, \mathbf{x}^{\mathrm{b}}, \mathbf{x}^{\mathrm{o}}$. These standard deviations are computed from the diagonal of the diagnosed posterior error covariance matrices associated with $\mathbf{x}^{\mathrm{a}}$, $\mathbf{x}^{\mathrm{a}}_{\mathrm{cl}}$, $\mathbf{x}^{\mathrm{b}}$, and $\mathbf{x}^{\mathrm{o}}$.

(2) The authors write in subsection 2.5 (ll. 362-363) ... entropic regularisation is enforced via ..., but say nothing about entropic regularization in subsection 2.4. Does that mean that entropic regularization is not implemented in the algorithm described in subsection 2.4 ?

You are absolutely right. We did not mention entropic regularisation in Section 2.4 because it would have made the equations lengthier, and more obscure. However, in practice in numerics, we do implement entropic regularisation. By contrast, it is critical to account for the entropic regularisation in Section 2.5, since the section provides the final formulas used in the numerical experiments and because the few tricks used in the derivation deal with the entropic regularisation. We have added a sentence at the end of Subsection 2.4.1 to clarify this point. Thank you for spotting this.

Note that in Section 2.4, we did not add the entropic regularisation to the cost functions for the sake of conciseness and because it does not play a role in the key ideas developed in this section; it would however be added and employed in numerical applications.

(3) Ll. 335 and 597, what is a Monge map (a scheme of the form of Fig. 3) ?

The Monge map is the optimal deterministic map T introduced in the deterministic formalism of optimal transport by Monge, and defined by Eqs. (2,3) of the manuscript. This has been made explicit in the revised manuscript, just after Eq. (3):

... whose purpose is to minimise the total transport cost between ρ_{o} and ρ_{b}, and whose optimal map T is often referred to as the *Monge map*.

(4) Ll. 469- 471, I find the sentence starting If $\mathbf{y}^{\mathrm{b}}$ and $\mathbf{y}^{\mathrm{o}}$ were... rather vague. What does it mean that the classical analysis would be as good as can be, while the OTDA solution may be too safe ? That the sharp transitions visible on panel 12 (b) would not be there, and that the OTDA solution would be too smooth ? Or something else ?

Yes, exactly. Following your questions, we elaborated on the sentence in the revised manuscript:

If $\mathbf{y}^{\mathrm{b}}$ and $\mathbf{y}^{\mathrm{o}}$ were consistently obtained from a truth perturbed with errors with short-range correlation, i.e. if they were drawn from the true prior distribution and in the absence of mislocation errors, then the classical analysis would be as good as can be, while the OTDA solution may be too safe, i.e. too smooth.

(5) L. 478, What does it mean that the first guess is weak ? That it is numerically small so as to be negligible, or erroneous, unreliable, or what ?

We meant, using rigorous terminology, uncertain (unreliable) as compared to the observations uncertainty. We modified the text to clarify this point. Thank you for the suggestion.

... because the first guess in that region is very uncertain.

(6) Caption of Fig. 12, what is a heatmap (I suspect a typo) ?

Heatmap is another popular name for *density plot*. To avoid introducing new names for the same object, we replaced *heatmaps* in the revised manuscript with *concentration maps* as in the first sentence of the caption.

(7) L. 155, the kernel of $\mathbf{H}$ is non-trivial. May be confusing. I suggest you simply write The main caveat of Eq. (7) comes from the fact that the system is only partially observed.

We followed your suggestion and changed the sentence accordingly. Thank you for the suggestion.

(8) L. 537, ... conformally to the spaces ... $\longrightarrow$ ... onto the spaces ...

Corrected. Thank you for the suggestion.

(9) L. 285 ... the same tensor index is present ...

Corrected.

(10) From what I can judge, the English is correct and perfectly understandable. A few corrections will however have to be made here and there. The paper will be copy-edited anyway, but I mention one point. The authors repeatedly write associated to (l. 288 for instance). That should be associated with (incidentally, that is a great classic of French authors when writing English).

All the occurrences of *associated to* have been corrected. Thank you for the suggestion.

Authors' response to Reviewer 1's comments on the revised manuscript "Bridging classical data assimilation and optimal transport: The 3D-Var case"

Preprint egusphere-2023-2755.R1

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

17 May 2024

We appreciate the reviewer's comments and suggestions. In the following, we discuss the raised concerns and what we have changed in the manuscript. The pdf document highlighting the differences between the original and the revised manuscript is provided.

(1) Related to my previous comment 1 and the authors' response, I agree that the manuscript ultimately clarifies that the theory applies to fields and not probability distributions. But in the original version at least, this is not stated until Sec. 1.4 on pg 5. The new MS makes this somewhat clearer although I personally believe it should be said earlier and more prominently.

We indeed had already modified the text to clarify this point following your initial comment. We do not believe that the *revised* manuscript is confusing or misleading in that regard. The title and the abstract of the revised manuscript were quite clear on the topic: (i) 3D-Var (in the revised title) is mostly focused on states and not pdfs, (ii) in the revised abstract, we wrote: *As such it provides a very attractive metric for non-negative, sharp fields comparison — the Wasserstein distance — which could further be used in data assimilation for the geosciences.* And then later on in the revised abstract: *The resulting OTDA framework accounts for both the classical source of prior errors, background and observation, together with a Wasserstein barycentre in between states that stand for these background and observation.* Note that, in the introduction specifically, we cannot alert the reader about this before section 1.4, because we did not introduce the specific problem of the paper yet. Adding a warning on a subject not yet introduced would be confusing.

We would like to mention that, closely connected to this point, we have slightly improved the paragraph where the curse of dimensionality is mentioned.

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 independently to the pdf of each single scalar variable. Quite often, OT is applied to the pdf of a single random variable because

- OT in one dimension (the space of the values taken by this random variable) together with the quadratic cost has a very simple solution that only relies on the cumulative distribution functions of the origin and target measures (see e.g., Remark 2.30 in Peyré and Cuturi, 2019), a technique known in statistics as quantile matching,
- increasing the number of random variables is subject to the curse of dimensionality, necessitating an exponential increase in computational resources, when increasing the resolution of the discretised fields.

This is very different from our context and objective where the objects dealt with by OT are (non-negative) physical field states, not the pdf of one of their scalar variables.

Note that we have completely removed the last sentence of the former version of this paragraph about OTDA not being subject to the curse of dimensionality, which is was unnecessary and indeed potentially confusing.

(2) Related to my previous comments 2 and 4c and what types of distances are able to "cope" with "distortions", I believe it depends on what one means with "to cope" and with "distortions",

and I also believe that there is no simple binary answer. The authors do not analyse this important issue in any depth which I feel is the main shortcoming of this paper.

We do not make claim of novelty on this topic; we do not add more in the revised manuscript and rely entirely on the literature where the need for non-local metrics was advocated. The novelty is in a consistent and practical set of techniques to embed OT into classical DA. The justification of why we would need that set of methods, what *distorsion* is and why classical methods have poor discriminating power (to cope with) also very much depend on the application of the potential user of the method. For us, the authors, our physical motivation is rooted in atmospheric chemistry and dispersion and is discussed at length in Farchi et al. (2016); Vanderbecken et al. (2023). We also found other references in the field of meteorological verification where instead of *distorsion*, the authors speak of *amplitude*, *structure* and *location*, (e.g., Wernli et al., 2008), pointing again to the fact that traditional local metrics are non-discriminative for fields such as precipitation. This reference was added to the newly revised manuscript.

(3) Related to previous comment (4a), it is still not clear to me how the authors deal with comparing fields that do not have the same mass.

It is likely that we did not understand your question, and in particular on which part of the manuscript it is related to. If your question is related to the general idea of the method, the unbalance is addressed in the cost function by the error terms (hence those of classical DA), which do not require the fields $\mathbf{y}^{\mathrm{b}}$ and $\mathbf{x}^{\mathrm{b}}$ on the one hand, and $\mathbf{y}^{\mathrm{o}}$ and $\mathbf{x}^{\mathrm{o}}$ on the other hand to have the same mass, as opposed to the OT terms where the origin and target fields must have the same mass. If your question is related to the thought experiment where we showed that the previous proposal on combining OT and DA is flawed and where we assumed that $\mathbf{y}^{\mathrm{b}}$ and $\mathbf{y}^{\mathrm{o}}$ have the same mass, then we merely proved that the previous proposal is flawed in the specific case where $\mathbb{m}(\mathbf{y}^{\mathrm{b}}) = \mathbb{m}(\mathbf{y}^{\mathrm{o}})$. It is hence flawed in general.

(4) Related to previous comment (4b), I now see that the regularised problem is convex but as a matter of fact, the original problem is a linear program and thus convex, albeit potentially not strictly convex. This should be clarified. In addition, it cannot hurt to comment on why the KL approach is appropriate, given that it strongly penalises fields that are rather localised with respect to the measure ν. Dealing with strongly localised fields was, as far as I understand, a motivation for the authors to propose the optimal transport methodology in the first place.

You are absolutely right. We were wrong in assuming that the original problem was not necessarily convex. Further, as you explained, it is often not strictly convex (which we know from the theory and by our own experience). Thank you for insisting on the topic. We have amended the manuscript accordingly. So the KL regularisation makes the problem strictly convex whenever $\varepsilon > 0$. It also has the key property that it lifts the positivity constraint of the fields. It is so obvious that it is easy to forget insisting on this key virtue.

The optimisation problem Eq. (14) is a *linear program* which is convex (Peyré and Cuturi, 2019, and references therein). Yet, it is not generally strictly convex, and hence does not necessarily exhibit a single minimum. Adding to the difficulty, its cost function Eq. (14a) is constrained. *Entropic regularisation* addresses these issues and is used here to lift the constraints and to render the problem strictly convex. It will in particular force any state vector which is solution of the problem to be positive.

The KL term is scaled by $\varepsilon > 0$ which can nonetheless be small. In practice, even with rather strongly localised fields, we did not see the issue with the 1D and 2D examples (many more than those reported in the manuscript). However, we agree that with fields with multiple relevant scales with, e.g., a large scale together with highly localised features, getting the optimal ε could be a real subject of investigation.

References

Farchi, A., Bocquet, M., Roustan, Y., Mathieu, A., and Quérel, A.: Using the Wasserstein distance to compare fields of pollutants: application to the radionuclide atmospheric dispersion of the Fukushima-Daiichi accident, Tellus B, 68, 31 682, https://doi.org/10.3402/tellusb.v68.31682, 2016.

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.

Vanderbecken, P. J., Dumont Le Brazidec, J., Farchi, A., Bocquet, M., Roustan, Y., Potier, E., and Broquet, G.: Accounting for meteorological biases in simulated plumes using smarter metrics, Atmos. Meas. Tech., 16, 1745–1766, https://doi.org/10.5194/amt-16-1745-2023, 2023.

Wernli, H., Paulat, M., Hagen, M., and Frei, C.: SAL–A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts, Mon. Wea. Rev., 136, 4470–4487, https://doi.org/10.1175/2008MWR2415.1, 2008.