Simulated plumes freed from meteorology biases using smarter metrics? – response to the reviewers

P. Vanderbecken, J. Dumont Le Brazidec, A. Farchi, M. Bocquet, Y. Roustan, E. Potier, and G. Broquet

December 2022

1 Answer to referee #2

1. *The paper proposes a new measure for forecast performance that accounts for displacement methods. Overall, the methodology seems sound, but per comment 1 below, it is unclear if this approach is really new, or what the new contribution is. Moreover, it is unclear if the added complexity of the approach over the displacement methods (e.g., as discussed in doi: 10.1175/MWR-D-19-0256.1) adds enough value to warrant its use. Therefore, I recommend acceptance after the authors consider the comments below.*
→ First, we would like to thank you for your review and the relevance of your remarks and suggestions. This paper presented new metrics to compare pollutant plumes. Even though their description takes a significant part of this paper, the main goal is to see how the proposed metrics handle changes in meteorological conditions. To make this point clearer from the start, we propose a new title. We share your concerns about the benefit of using those metrics against the usual ones. Yet, for us, only the inversion results should be the judge of it. This will be discussed in a future paper.

2. *The method proposed is very similar to several of the field deformation approaches described by the cited Gilleland et al. paper and several since that time: e.g., see doi: 10.1016/S0022-1694(00)00343-7, doi: 10.1175/2010WAF2222365.1, doi: 10.5065/D62805JJ, doi: 10.1002/2012GL053964, and doi: 10.1175/2010WAF2222351.1 to name just a few. In particular, doi: 10.3402/tellusb.v68.31682 uses the Wasserstein distance. A thorough literature review and comparison of the differences and added utility of the present approach is necessary to put this work into the greater context of these deformation methods. As it is, it is not clear what the new contribution is over these other works.*
→ First of all, the different metrics proposed in this paper aim at handling position errors in a better way. This is indeed a similar goal to the field deformation approach or warping technique. But, as far as we know, the rotation that we consider as an orientation error,

which was suggested to us by practitioners, is usually not included in the position error and remains included within the shape error. The plane transform used in both d_F and w_F will conserve the shape of the plume which is not the case with the warping function used in the literature.

It is true that the Wasserstein metric has already been proposed by Farchi et al. (2016) to perform plume comparisons. The novelties here are that (i) we propose additional metrics beyond the Wasserstein distance, (ii) we provide a more systematic evaluation of these metrics, and (iii) we use a different algorithm to compute the Wasserstein distance (namely, the Sinkhorn algorithm).

In the revised manuscript, we have reformulated the introduction to explain these elements (L. 56 to 67), with hopefully more pertinent references to the literature.

3. *How does this approach address the issues outlined in doi: 10.5065/4px3-5a05 ?*
→ In (doi: 10.5065/4px3-5a05), the authors share the analytical issues proposed by Davis, C et al. in (doi: 10.1175/2009WAF2222241.1). They correspond to pathological situations that occur when comparing features in images. We thank you for pointing out this relevant reference since any of our 10,000 analytical cases can be seen as a combination of these pathological situations. We add it in the manuscript (L. 309). Our metrics address these issues as described in the results of 10,000 analytic cases (section 4). To be more specific, both translation or rotation displacement are solved by d_F and w_F whether the plumes were initially overlapping or not, which is what we were aiming for in our case.

4. *The authors make reference to the measure's being fairer, but it is unclear what they mean by fair in the general concept of a fair verification measure.*
→ Indeed, the word 'fairer' is subjective and should be avoided. We meant here that the translation error is linearly penalised by the Wasserstein distance while the L2-metric will reach a maximum when the plumes do not overlap. This has been explained in the revised manuscript (L. 211). Thank you for pointing this out.