Reply to Editor and Reviewer Comments

Jared C. Magyar & Malcolm Sambridge

Editor Comments

The discussion of the paper is very insightful and interesting to read. I commend the reviewers for a job well done. I concur with the reviewers that the paper is in good shape already. From your replies to the comment it is obvious that you have a clear idea about the best way to address the comments. I encourage you to do so. I believe that minor revisions will suffice.

As suggested, we have made the suggested minor revisions for the resubmitted manuscript. We should note to the editor that the title has changed, along with an initial for a co-author. These adjustments have been made in the submission of the revised manuscript.

Reviewer #1 Comments

We thank the reviewer for their helpful feedback. Our replies to these comments and adjustments in the revised manuscript are displayed in italics.

Title: The paper discusses how the Wasserstein distance can be used i) as an objective function and ii) to construct ensemble representatives, but only the first aspect is reflected in the title. I suggest changing the title such that is contains both aspects.

The authors agree with this point. We have adjusted the title to ***'Hydrological objective functions and ensemble averaging with the Wasserstein distance'*** *to better reflect the contents of the paper.*

Use of the word "metric": In the article, the authors use the term "metric" to refer to distance measures in general, not to the strict definition of a metric being a distance measure with the following properties:

- d(a,b)>=0
- d(a,b)=0 then a identical b
- d(a,b)+d(b,c)<=d(a,c) (triangle inequality).

As the authors go quite deep into the derivation of the Wasserstein distance, and the discussion of its properties, I suggest they should i) restrict usage of the term "metric" to true metrics only, and ii) mention if Wasserstein distance is a true metric or not.

The Wasserstein distance is a true metric on the space of density functions which is ***now stated explicitly on L130 with appropriate references****.*

Similarly, it will be helpful for the reader to discuss if the Wasserstein distance is a symmetrical or non-symmetrical distance function, i.e. if d(a,b) = d(b,a) or not. In other words, does Wasserstein compare the distance of some "model" to a "reference truth", or the distance between to objects on equal terms.

The Wasserstein distance is indeed symmetric, and is ***now clearly stated as such on L130*** *of the revised manuscript.*

A comment (no changes requested): Apart from the illustrative results from the synthetic test cases, for the hydrological community it would be very useful to see some applications to realworld, and long, hydrographs. These will be the cases that are relevant from a practical point of view, but where the limitations of the Wasserstein "mass-problem" will become most obvious. I do not request extra work here, as the authors mention in lines 78-83 that the main purpose of this paper is to introduce the concept and show some illustrative examples, and I think this is enough to justify the paper. Nevertheless it would be helpful to provide such applications at least in a follow-up paper. This will greatly increase the chances that the method will be picked up by the community.

The authors agree with this comment, and also agree that it is beyond the scope and purpose of this particular contribution.

Sect. 3.1: The authors apply the Wasserstein- and RMSE-optimization to the system with the rainfall timing errors. It would be helpful to initially mention that when using the true rainfall input, both Wasserstein- and RMSE-optimization would perfectly identify the true model parameters (which I assume should be the case).

As with RMSE, the Wasserstein distance will only be zero (global minimum) when the inputs are identical. Under the assumption that the solution is unique, errorless data would indeed give us the true model parameters when the Wasserstein distance is optimised. ***This is now stated on L313-314.***

Eq. 1: at this point, it is unclear why two different sets of data **x** and **y** are required for functions f and g. Either explain or replace **y** by **x**.

Thank you, this was an error in the equation and ***was fixed as suggested****. Indeed, it is important that the data points are the same for both f and g.*

Reviewer #2 Comments

We thank the reviewer for their thoughtful comments on our manuscript. Replies to the comments and adjustments in the revised manuscript are displayed in italics.

The manuscript presents an excellent introduction to the Wasserstein distance and explores its relevance for hydrological modelling. The paper is well structured, the figures are relevant and of high quality and the writing is clear. I have no reservation to recommend the paper to be published. I do have a few minor comments as detailed below.

1. References to hydrological literature. The literature review is comprehensive and relevant, but it would benefit from some references to hydrological literature on objective functions. Below are a couple of references I believe to be relevant. I leave it to the authors to decide if they want to include them (never include a reference just because a reviewer suggests it).

The authors agree that some additional reference to hydrological literature would be suitable. Priority thus far has been placed on deterministic objective functions, as significant discussion of Bayesian approaches may confuse the reader on the proposed use of the Wasserstein distance. We also do not propose any way of building a likelihood function from the Wasserstein distance, so we believe any additions referring to likelihood functions should be kept brief.

However, ***we add two sentences in L77-80 about the need of a likelihood function for Bayesian inference, but this being distinct from an objective function*** *as it is dependent on understanding the noise statistics of the data. It is worth noting however that likelihood functions such as*

Schoups & Vrugt (2010) and Vrugt et al. (2022) which account for temporally correlated errors operate directly on the point-wise residuals, while we use a different definition of residual (see Fig. 7c, where residuals are between the inverse cumulative distributions). It is therefore not straightforward to design a Wasserstein-based likelihood function and beyond the scope of this manuscript.

2. The Wasserstein distance appears to have concepts in common with flow anamorphosis (http://www.stat.boogaart.de/Publications/g1901.pdf). Can you explore or comment on how the Wasserstein distance compares to flow anamorphosis?

This is an interesting concept, and the authors thank you for bringing it to our attention. While there are similarities in the transport-based methods, the emphasis for the given flow anamorphosis method is the mapping between distributions, rather than quantifying the work required to do so as is the case with the Wasserstein distance. While van den Boogaart et al. (2015) do use a form of transport map via Lagrangian trajectories, whether this is the optimal transport map from which the Wasserstein distance is defined in Eq. 6 is not immediately clear.

We should note the potential use of the optimal transport map for this application. In particular, the Sliced Wasserstein distance (Rabin et al. 2011) may find use in mapping points in high dimensions to a multivariate normal to allow sampling of arbitrary distributions through the inverse mapping (Magyar, unpublished work).

While there are interesting links and applications here, much of it is beyond the scope of this manuscript, so ***there are two additions that we have made in the revised manuscript:***

1. *When introducing the optimal transport map,* ***we now note the similarity to sampling transformations (it is the same as the inverse transform sampling method in 1D, and similar in concept to flow anamorphosis for higher dimensions) in L118-120****. We also make very clear that the Wasserstein distance is defined for the optimal (work-minimising) map, not just any map between the distributions.* ***We again draw the link to the inverse sampling method in L200-201.***
2. *In the conclusions****, we now mention on L425-428 the untapped potential of the optimal transport map (rather than just the Wasserstein distance) for a two-way mapping between collected data and a reference distribution*** *(e.g. multivariate Gaussian) in hydrological applications. Further details however are beyond the scope of this contribution.* ***The link to flow anamorphosis is again made here.***

References:

Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Research, 46, W10531. https://doi.org/10.1029/2009WR008933

Vrugt, J. A., de Oliveira, D. Y., Schoups, G., & Diks, C. G. H. (2022). On the use of distribution-adaptive likelihood functions: Generalized and universal likelihood functions, scoring rules and multi-criteria ranking. Journal of Hydrology, 615, 128542. https://doi.org/10.1016/j.jhydrol.2022.128542

Rabin, J., Peyré, G., Delon, J., and Bernot, M.: Wasserstein Barycenter and Its Application to Texture Mixing, in: International Conference on Scale Space and Variational Methods in Computer Vision, pp. 435–446, 2011.