The authors propose the Wasserstein barycenter for use in multi-model ensemble forecasting and test its relative performance against the more traditional L2 barycenter. Overall, the study is well-written and presented in a way that was digestible by a general audience, despite some of the more complicated mathematical constructs employed. While the results indicate that the Wasserstein barycenter does provide some advantages, is it not universally better than the traditional approaches. Nonetheless, this work is important for exploring how other options for ensemble forecasting could provide additional insights. I recommend the paper for publication subject to minor clarifications and revisions below.

We thank anonymous reviewer #2 for their overall assessment, the insightful comments and encouraging review. Indeed, in no way do we claim that the Wasserstein barycenter is universally advantageous, but rather argue that it is of interest as an alternative tool to explore and manipulate multi-model ensembles. Please find below our point-by-point responses:

Clarifications:

1. Line 164: “Note that they additionally assumed the signal to be stationary and periodic which we did not do.” Can you clarify the implications of this choice?

The assumption that the signal is stationary and periodic allows them to simplify the computation of the Wasserstein mapping and barycenter. These hypotheses were too strong to be applied to our data set. The sentence will be modified as follows: “*Note that they additionally assumed the signal to be stationary and periodic to simplify the computations, however, these hypotheses are not adapted to our dataset*”.

2. Line 201: Since the members of the Wasserstein barycenter isn’t drawn from the set of possible simulated states, is it possible that it is not actually physically realizable? Namely, that there’s no physically consistent way to connect the final and initial states?

This is an interesting question. The members of the L2 barycenter are taken from the initial ensembles and are thus physically consistent (according to the numerical weather model they came from as you noted in point 4 below). While the W2 barycenter allows for more flexibility in the sense that its members are not coinciding with the input ensembles, it indeed does not guarantee their physical consistency. However, the W2 barycenter does not treat the variables independently, their covariances are taken into account. This is explicitly done in the computation of the Gaussian Wasserstein barycenter (see equation l170).

In our case study, the barycenters are applied to several lead times of the same parameter (2m temperature, 10m wind speed or geopotential height). Only looking at one parameter, it is difficult to check the physical consistency between several weekly values.

3. Line 247: “Due to model errors, forecasts tend to drift away from the observed climate toward the model climatology as lead time increases.” Isn’t the drift also caused by initial observational uncertainty?

Forecast errors are indeed due to both model error and uncertainty on the initial condition. However, here, we were referring to the drift toward the model climatology, which is independent from the initial states and has a large impact at sub-seasonal scales.

Thank you for the question. We will clarify this in text as follows :“*Ensemble forecasts suffer from systematic errors (e.g. mean bias) due to uncertainties in the ensemble initialization and model formulation. In particular, forecasts tend to drift away from the observed climate toward the model climatology as lead time increases (Takaya et al. 2019), making statistical correction essential at extended-range time scales*”.

4. Perhaps the fact that the W2 barycenter could give results which are not part of the ensemble is an advantage: ultimately every model cannot represent physical co-variances correctly and have their own biases (all modeled physical processes are approximations). By choosing a trajectory that is not present in the ensemble you are no longer bound to those biases.

The Wasserstein barycenter also depends on the variable covariances because they are derived from the model outputs in the GaussW2 barycenter. However it is true that the covariances are modified/merged by it and this allows having samples that could have not been generated by the individual models.

Thank you for raising these points (2 and 4), we will modify section 2.2.3 (including l 201) to reflect this discussion.

Minor comments:

1. Section numbers referred to in the manuscript seem to be rendered incorrectly by the latex template.

Thank you for pointing it out. The section numbers will be corrected in the revised version.

2. Line 143: Should the variables inside the curly brackets be a_1 and a_2 instead of a and b?

You are correct. Thank you for pointing it out, it will be corrected in the revised version.

3. Line 163: Parenthetical citation should be used for Flamary et al. (2020)

Parentheses will be added to this citation in the revised version.

4. Line 182: hypothesis hypothesis

This typo will be corrected in the revised version.

5. Line 428: multi-model

This will be corrected in the revised version.