Seasonal statistical-dynamical prediction of the North Atlantic Oscillation by probabilistic post-processing and its evaluation

André Düsterhus

January 10, 2020

1 General

Thanks a lot for these important suggestions. I answer them bellow point-by-point. I have also added the annotations in the plots again, which I had forgotten in the last resubmission.

2 Editor

Comments to the Author: Dear André Düsterhus,

I agree with the positive reports of the three reviewers. In light of the reviewer's comments and the revised version of the manuscript, I would like to suggest a few additional minor revisions.

- page 2, line 35: I suggest to change "climate science" to "atmospheric science", "weather prediction" or similar as all references are from the NWP literature.

Changed to "atmospheric science".

- page 2, line 50: Is the model resolutions of 40 km over the ocean correct? I would have expected a lower resolution compared to the T63 model.

Yes that is correct. Of course in the relevant regions there will never be a resolution this coarse, as the numerical North pole will be placed in a way to have a finer resolution where it matters (e. g. the North Atlantic). That is why these numbers of degree and km are so misleading and in my opinion we should stick generally only to the technical resolutions. Nevertheless, as many have no understanding of these terms it might be useful (and at the same time confusing) for some readers to include the equivalents.

- page 3, line 64-65: The revised version of this sentence is more clear (see comments by Anonymous Referee #3), however, I suggest to replace "shown" by "indicated" or similar

to clarify further.

Changed to "indicated".

- Section 3.3: Personally, I was very happy to see proper divergence measures used for evaluation. However, as evident from all review reports, this concept seems to be not very well known in the community. Therefore, I suggest to make the difference between IQD and CRPS more clear, for example by extending the discussion in lines 111-112 on page 4 by stating more explicitly that the IQD extends the CRPS towards observations that are not deterministic, but "distributions".

I clarified it further. Added the sentence: "As a consequence, while CRPS needs to have a point observation, the IQD can take into account the full uncertainty distribution of an observation."

- page 5, line 119-121: The sentence starting with "EMD is a metric..." will be very hard to understand for readers not familiar with Thorarinsdottir et al. (2013). Perhaps it can be rephrased in a more generic way without referring to the concepts of "k-proper" or "asymptotically proper"? Further, I am not sure that I understand the reasoning behind the statement in the subsequent sentence claiming that EMD specifically prefers underdispersed model simulations. Why does this follow from being only asymptotically proper, but not a k-proper divergence measure? In my view, this only implies that EMD may prefer incorrect models over the correct one, but may those incorrect models not as well be overdispersed?

The argument have been taken from Thorarinsdottir et al. (2013) (THO13) and there are reasons why I omitted it in the original submissions. Generally, it makes the paper much more complex in the statistical sense, but as the reviewer have mentioned that seems to be ok. Taking only the IQD will not prevent me from explaining the problematics (as squaring is a very unintuitive choice for most), so I included both. I have tried to simplify the sentence to "While EMD is a metric measuring the distance between the pdf it is in contrast to IQD not a proper score." I would like to prevent to go into more detail here, because the difference between the two measures given by THO13 with "(iv) the divergence function ought to be "mathematically well behaved and well understood."" is quite hard to formulate in a correct and at the same time understandable way for non-verification-statisticians. Concerning the "underdispersed" statement, I have taken it directly from THO13 "..., thereby suggesting that d_{AV} encourages underdispersed model simulations." While I see your questions as valid, it follows, as far as I understand the paper, out of MC-simulations (p 528 bottom). In this part of the manuscript, I just point to THO13 and their mathematical argumentation and do not see any argument, which contradicts their findings (and it would be overstretching this article to redo the mathematical background entirely). As I made clear in the manuscript that the IQD is to be preferred out of theoretical argumentations (even when EMD is much more intuitive), I think it is a fair way to handle it. Also as I stated in my last reply: I think it is important to get theoretical results out of the mathematical literature step by step into the more applied literature to make it more approachable for everybody. At the IMSC in Toulouse last year it was apparent in the Verification session that handling observations with uncertainties is an important next step. So this manuscript should show that choices matter, and makeing this way reders aware of the problematics we face when we approach this new horizon. Showing both metrics and their differences in this manuscript is therefore in my view a valid approach to achieve this aim.

- page 9, line 237-238: For evaluating point forecasts the RMSE can be used without an (implicit or explicit) assumption of Gaussianity, see, for example Gneiting (2011, https://doi.org/10.1198/jas The statement in the sentence starting with "An important advantage ..." thus seems a bit misleading and I suggest to rephrase accordingly.

I have taken the sentence out, as it is unnecessary at this point of the storyline. The problem with non-normal distributions of variables for correlations is discussed some paragraphs below.

- page 17: The caption of Figure 5 mentions a "dark green" color which is not present. Should this read "red" instead?

Is changed.

Best regards, Sebastian Lerch