Reply to reviewer #2

We wish to thank the reviewer for her/his constructive comments. We reply to each comment below (original comments *in italics*, and our response in regular font).

1. General remarks
The paper treats a very actual problem: ensembles technique choice for decadal simulations. The work frame is well defined for the purpose of the paper, considering non-initialised simulations in order to directly compare ensemble spread when using two initial perturbation methods: lagged initial conditions and bred vectors. The paper is original, as it treats the problem at decadal scale, extending techniques already tested for seasonal predictions. The conclusions are very useful for decadal prediction field.

I do recommend the paper to be published (minor revisions are suggested bellow).

2. Specific comments
The text is very clear and the methods and conclusions fairly presented and targeted to the main aim of the paper. Figures are clear and support the results and conclusions.

3. Technical comments
General
****
**) knowing that the error growth is a function of the perturbation size, it would be interesting for the conclusions, to see if the comparison here, compares indeed same initial perturbation size: Here are compared 1-day lag perturbations against normalized monthly, so are they, after normalization comparable as to make more clear the comparison ?*

This is an interesting comment, and would require the calculation of error growth. In the present setup, however, the results are unlikely to be immediately of additional value: on the one hand, we have comprehensively tested the breeding implementation and the identified where the implementation is most sensitive (see next point). On the other hand, the perturbation size for the lagged initialization is from a practical standpoint fixed (longer periods would considerably limit the ensemble size). We agree that for further comparison against other ensemble generation methods the computation of error growth (and saturation) can potentially be very informative.

**) also would be interesting to mention if there is impact known (tests) of using different norms for temperature and salinity ? may improve regions of Eckman convergence?*

The main sensitivity that we have found was on the length of the breeding cycle and on the extent of the vertical profile for the norm. Breeding temperature and salinity only at individual levels (for example 50m or 200m) or restricting the breeding to the upper ocean or 500-1500 m resulted in an underrepresentation of the spread. We did not specifically conduct any tests using separate norms for temperature and salinity, but we did conduct experiments were the norm was restricted to the upper ocean heat content, and salinity was not bred. Here, little sensitivity in the bred vectors was found. Overall, none of these additional experiments yielded additional insights to the results presented here. In a model with higher spatial resolution than the one presented here, we agree that it might be worth to analyze the regional impacts of different norms further.

**) par20,page6: "here we allow for a period of 2 years with monthly normalization" : how has been chosen this period? is it a model error growth feature ?*

The period was chosen so that the structure of the initial perturbations is no longer apparent in the perturbations generated by the bred vectors. As a few extra breeding cycles do not deteriorate the result, this period is well beyond the time at which the bred vectors have lost their memory of the initial perturbations. Hence, the period cannot be indicative of error growth in the model.

**) the integration length during the breeding cycles iterations is function of new perturbation size ? (after the first normalization): the saturation time may be a function of it, hence accounting that could provide a better approach for the growth slope.*

We are unsure whether we entirely understand this comment. Therefore, we think it might be useful to clarify that the length of the breeding cycle depends on the dynamical mode that should be isolated. It is the size of the norm that influences the amplitude of the perturbations. Hence, the length of the breeding cycle and the amplitude of the perturbations are not directly linked.

****

Core text

****

**) par25, page6: "size of the rescaling norm" while the norm definition is clear for a vector space, I would suggest to define what is here termed by "the size of the norm"*

We now specifically refer to step 4 in the breeding implementation here (were we explain the rescaling).

**) how do EOF for other levels (deeper ocean) compare ?*

The similarity between the unperturbed experiment and the bred experiment extends to 3500m for the first EOF, both with respect to explained variance and structure.

**) par10, page11: "The spread error ratio is then the ratio of this spread and the difference .." : possibly you used "absolute difference" not "difference" ?*

Corrected to ‘root mean square difference’.

**) par20, page11 typing error " After four months ...lagged initialized ensemble " 2 times.*

Corrected.

**) par5,page13 typing error "comparable .. for the ocean temperature case"*

We were not able to identify this specific error at the indicated location.

**) par15,page 14 and conclusion 2: not clear if there was any sensitivity to rescaling norm experiment conducted here.*

We have removed the reference to ‘better’ representation from the conclusions.