Review of « Improving Statistical Projections of Ocean Dynamic Sea-level Change Using Pattern Recognition Techniques" by Malagon-Santos et al.

The paper investigates the benefit of using pattern recognition approaches to assess statistical regional sea level projections from coupled climate model outputs. The study shows that using EOF pattern recognition and low-frequency component analysis significantly reduce errors in pattern scaling of regional ocean dynamic sea level change. The authors apply those two methods on the large ensemble MPI-GE simulations. Each member has different initial conditions. Therefore, it is possible to assess the impact of ocean initial conditions on projected dynamic sea level change. The presented results highlight the need to apply such a pattern recognition methods to reduce errors in regional emulation tools of ocean dynamic sea level change especially when a few realizations are available because of the huge computation cost.

The topic of the paper is interesting as the future generation of AOGCM will increase both atmosphere and oceans spatial resolutions. Thus, a few simulation integrations will be preferred from large ensembles because of the computational cost. Therefore, this technique may be relevant for future sea level change investigations.

Thank you for the positive comments.

I find the paper well written. It is well organized. However, the methodology part could be improved as the methodology is not easy to understand especially for a non-expert in pattern filtering. I think the authors can provide more explanations to help the reader.

We agree the methodology section may be difficult to follow for a non-expert. We thought it was not needed to further expand the methodological steps, as interested readers are referred to appropriate literature where methods are thoroughly explained already. We will revise the methodology to make it more consistent with the rest of the manuscript, simplifying some of the most technical explanations. We may also consider adding a flowchart, if we feel this would help the methods to be more easily understood. We believe these changes will make the methods sections more accessible to non-experts.

Overall, the paper is well supported but some parts are unclear. For instance, I struggle to fully understand and interpret Fig1 as it lacks of explanation in the caption (see my comment below).

I find the paper very technical and I wander if Ocean Science is the right journal to publish this piece of work. I recommend a major revision for the manuscript before a possible publication.

Ocean Science was chosen based on the specific dedication of the journal, which is on all aspects of ocean science including experimental, theoretical, and laboratory studies. We believe the topic of manuscript falls within different fields covered by Ocean Sciences, especially ocean physics and ocean models. Moreover, since our paper aims to simplify complex global climate (or related) models by using statistical approximations, the methods and results presented here may be also tested in other oceanic processes, such as regional changes in temperature. This made us feel Ocean Science was a suitable journal for this manuscript. We hope the amended methods sections will make the paper seem less technical and more accessible to a broader audience in Ocean Science.

Major comments

- When using EOF decomposition, one strong assumption is that all the modes are independent (i.e., they are orthogonal to each other). Is it really the case especially at global scale? This might be discussed in the conclusion as a limitation of the approach.

  By construction, the EOFs patterns and principal components are orthogonal. It is this orthogonality constraint what inhibits a physical interpretation of EOFs (as noted in Line 308 of the original manuscript).

- What do you mean by 'well separated'? (L143) How is it performed? Are you sure the initial conditions are totally different and independent? Please, clarify.

  Single-model initial condition large ensembles (SMILES) are designed to assess a range of outcomes due to the presence of unpredictable internal climate variability. This is usually achieved by running a number of simulations with the same model and identical forcing, only differing in their initial conditions. Simulations are independent as long as the memory of the initial conditions is lost, which ensures each ensemble member have a unique climate trajectory (Deser, 2020). There are two main procedures to achieve this: 1) by inducing small round-off level differences in their atmospheric initial conditions (micro-initialization); 2) by branching simulations at different times in the control simulation (macro-initialization). Both micro and macro initialization procedures are useful to characterize unpredictable internal variability within a model. Macro-initialization, however, provides larger differences in the initial states in both the atmosphere and ocean. Since we are assessing ocean processes here (i.e., ocean memory is important for our analysis), we deemed a macro-initialized ensemble more suitable for the purpose of this study.

  MPI-GE simulations assume a stationary and volcano free 1850 climate, and are macro-initialized on the first of January in different years of the control simulation (Table 1 in Maher et al., 2019). The branching separation between realizations varies along the pre-industrial control, ranging from 6 to 24 years and with a median of 16 years. As already noted in Line 144, the branching times of macro ensembles are designed to sample large scale aspects of the climate systems (atmosphere, land, and ocean). Nonetheless, we will include further information on the MPI-GE and emphasize its value for assessing internal climate variability within the model. We will also provide an improved description of micro vs macro and why we chose the latter.

Deser, C. (2020). Certain uncertainty: The role of internal climate variability in projections of regional climate change and risk management. Earth's Future, 8(12), e2020EF001854.

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., ... & Marotzke, J. (2019). The Max Planck Institute Grand Ensemble: enabling the exploration of climate system variability. Journal of Advances in Modeling Earth Systems, 11(7), 2050-2069.

- As GMTSLR is removed, the underline hypothesis is that the model conserves volume instead of mass. Is that right? If so, this is due to the Boussinesq's approximation. This should be clearly stated to avoid any misunderstanding.

We would like to clarify GMTSLR is not removed, we simply do not use it. Dynamic sea level (zos) and GMTSLR (zostoga) are usually provided separately in AOGCMs.

Almost all CMIP6 and CMIP5 models use the Boussinesq approximation (Irving et al., 2021) which implies, as mentioned by the reviewer, that volume is conserved instead of mass. This means steric processes are represented by a change in density, from which a change in mass can be inferred (the so-called Boussinesq ocean mass). That is why GMTSLR (or zostoga) is inferred separately.

Irving, D., Hobbs, W., Church, J., & Zika, J. (2021). A mass and energy conservation analysis of drift in the CMIP6 ensemble. Journal of Climate, 34(8), 3157-3170.

- MPI-GE description is too succinct. Please, provide more insights. There is no mention on the spatial resolution of the MPI-GE simulations especially for the ocean part. I assume that the ocean spatial resolution is about 1° meaning that the oceans have laminar flows. If so, what is the consequence when assessing the internal variability? Are not you underestimated it? Some studies have estimated the ocean-based internal variability from a large ensemble of forced OGCM. When increasing the spatial ocean resolution, the ocean-based internal variability increases in space and time. We can expect the same behavior for the coupled internal variability. I would appreciate some discussion on this specific point in the discussion's section.

The model is indeed course resolution: T63L47/GR15L40 ("LR" - Low Resolution). Nonetheless, Suarez-Gutierrez et al. (2021) show that MPI-GE well samples observed ocean variability in all regions except for the Southern Ocean. Below we provide more details about the model's resolution:

- Atmosphere: approximate horizontal resolution of 200 km (1.875 degrees) at 47 layers (up to 0.01 hPa / 80 km in height)
- Land biosphere (interactive vegetation): same horizontal resolution as atmosphere.
- Ocean including biogeochemistry: horizontal resolution varies from 12 to 150 km at 40 layers.

We will expand the MPI-GE description in the revised manuscript.

It is true that variability tends to become larger at higher model resolutions (Penduff, 2010). However, since the goal of our study is to remove internal variability and isolate the forced response to improve its statistical modelling, how well internal variability is represented in a model is not our concern.

Penduff, T., Juza, M., Brodeau, L., Smith, G. C., Barnier, B., Molines, J. M., ... & Madec, G. (2010). Impact of global ocean model resolution on sea-level variability with emphasis on interannual time scales. Ocean Science, 6(1), 269-284.

Suarez-Gutierrez, L., Milinski, S., & Maher, N. (2021). Exploiting large ensembles for a better yet simpler climate model evaluation. Climate Dynamics, 57(9-10), 2557-2580.

Minor comments

L54-63: When describing the drivers of regional sea level changes, one might want to know the associated time scales of each processes. Please, clarify. This would help the reader.

We agree with the reviewer that this could help the reader notice the differences between the different regional contributions and highlight the importance of ocean dynamics as a significant driver of variability in sea-level change projections. This will be clarified in the revised manuscript by referring to appropriate literature, for example Fig. 1 in Durand et al. (2022).

Durand, G., van den Broeke, M. R., Le Cozannet, G., Edwards, T. L., Holland, P. R., Jourdain, N. C., ... & Chapuis, A. (2022). Sea-level rise: From global perspectives to local services. Frontiers in marine science, 8, 2088.

L66: What do you mean by natural variability? Could you define this concept? This would help the readers.

As also pointed out by the other reviewer, we use 'natural variability' here when we meant to refer to 'internal climate variability'. We realized we committed a mistake when using those terms interchangeable throughout the paper, which can lead to confusion. Climate variability is defined as variations in the mean state and other statistics (e.g., extremes) of the climate (Mason-Delmotte et al., 2018). Climate variability can be caused by natural internal processes (internal variability) or by variations in natural or anthropogenic external forcing (external variability). In this study, we address internal climate variability, defined as naturally occurring climatic variations controlled by interactions between different components of the Earth system (Hasselmann, 1976; Schwarzwald et al., 2022). We will check the paper and make terminology consistent to avoid ambiguity, including a definition of internal climate variability as suggested.

IPCC, 2018: Annex I: Glossary [Matthews, J.B.R. (ed.)]. In: *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 541-562, doi:10.1017/9781009157940.008.

Hasselmann, K. (1976). Stochastic climate models part I. Theory. tellus, 28(6), 473-485.

Schwarzwald, Kevin, and Nathan Lenssen. "The importance of internal climate variability in climate impact projections." Proceedings of the National Academy of Sciences 119.42 (2022): e2208095119.

L75: What do you mean by 'regional emulation tools'? Please, define any new terminology.

Emulation is a method consisting of parameterizing process-based models so that their output is estimated at significantly reduced computational cost (Thomas and Lin, 2018). Regional emulation follows the same principle and aims to estimate a spatiotemporal varying variable by mimicking computationally expensive approaches, such as process based GCMs, using less computationally extensive approaches, such as statistical models. We will include this definition here and provide appropriate references and examples when doing so.

Thomas, M. A., & Lin, T. (2018). A dual model for emulation of thermosteric and dynamic sea-level change. Climatic Change, 148(1-2), 311-324.

L109-110: How many members do you need to completely cancel out the internal variability?

There is no straightforward answer for this question. Strictly speaking, we would need an infinite number of realizations to completely cancel out variability. The number of members needed to robustly characterize internal variability depends on the question to address and acceptable error, as explained by Milinski et al. (2020). What we have observed in this study regarding dynamic sea-level variability is that internal variability associated to well-known modes of climatic oscillations leading to coherent spatial structures may be easy to define using a few ensemble members, whereas higher variability (e.g., related to eddy dynamics) is much more difficult identify.

Milinski et al 2020: How large does a large ensemble need to me.
https://doi.org/10.5194/esd-11-885-2020

L297: What do you mean by 'conventional approaches'? please, clarify.

We refer to simpler but less efficient approaches that have been widely used to remove internal variability, such as ensemble averaging. This has been referred to in other sentences, (e.g., 27, 110, 333, 365, 502) but will also include it here to increase readability.

L323-324: '...that appear to be linked to volcanic eruptions'. Can you bring extra explanation here or a suitable reference?

The time evolution (in standard deviation) of these patterns is rather stable except for specific points in time where aerosol forcing was significantly altered in the atmosphere due to volcanic eruptions. Aerosols from volcano eruptions can change temperatures in the atmosphere (Wills et al., 2018), which in turn also affects sea level. As an example, the peaks that can be seen in the figure below (a, c, e) coincide with eruptions from Krakatoa, Agung, El Chinchón, and Pinatubo, suggesting that those eruptions indeed exerted a change in ocean dynamic sea level. We will include this explanation in this in text.

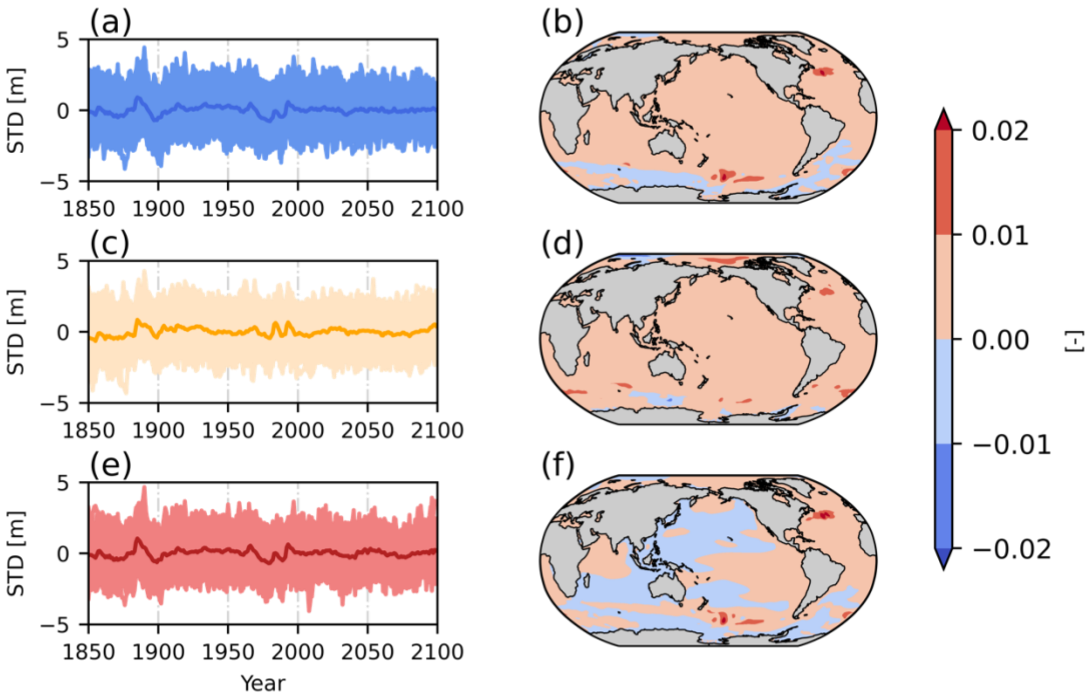

Figure1: I do not fully understand this plot. Why Sk is decreasing when pattern number is increasing? Please, clarify it and maybe extend the caption.

In the pattern filtering approaches tested here, signal to noise (S/N) patterns are sorted in terms of their signal fractions (Sk), and that is why Sk is maximal for pattern 1 and decreases with pattern number. This is already briefly mentioned in line 233. The caption of this figure will emphasize this in the amended manuscript.

Figure2: Please, change SD by standard deviation. This would help the reader.

SD will be changed in the revised manuscript.

Figure 4: Are the results consistent when considering RCP 4.5 an RCP 8.5? It would be interesting to add them into the supplementary materials.

As noted in Sentence 371, we found no significant differences between scenarios when it comes to RMSE reduction provided by pattern filtering techniques. As filtering techniques remove internal variability, and the latter does not significantly change between scenarios up to 2100, results are similar for different scenarios. However, we did found contrast in the slope difference between unfiltered and filtered results for different RCP scenarios. Slope differences are highest for the lowest emission scenarios, and decrease as radiative forcing increasing. Since lower radiative forcing leads to a lower signal/noise ratio, noise (internal variability) can drive large differences in slopes between filtered and unfiltered results. On the contrary, a higher emission scenario is characterized by a higher signal/noise ratio, as noise exerts less control on slope differences. We will include a plot showing differences in slopes for distinct RCPs considered here, explaining why slope differences differ between scenarios while model RMSE does not.