Review of "Regionalizing the Sea-level Budget With Machine Learning Techniques" by Camargo *et al* (2022) for Ocean Science

Review by Dr Sam Royston, University of Bristol

Summary

This work discusses the sea-level budget from observations at different spatial scales. While the global-mean sea-level budget closes for the period of observations from 1993—2016, there remain differences in the budget at smaller scales, which are important to understand. The approach here is to use two unsupervised machine learning methods to define smaller, sub-ocean-basin scale regions with covarying sea level and test the sea-level trend budget at a range of scales. It is an important topic and the paper is well written and clear. I do have a reservation about the use of reanalyses data that dominate both the steric (pre-Argo) and manometric DSL components but the authors have provided sensitivity analysis and comparison with observations (GRACE) that imply the main conclusions of the paper are still valid. There are a few general remarks that I feel the authors should address before publication, but overall it is a substantive piece of work of excellent quality and worthy of publication.

General Remarks

The authors do note in Appendix B that there is some circularity in their sea-level (trend) budget. They use reanalysis data for the manometric dynamic SLA and in the steric SLA ensemble, but these reanalyses mostly assimilate altimetric SSHA data; which the authors are then comparing against. This issue is most concerning pre-Argo, as the steric SLA ensemble becomes heavily weighted to the reanalyses products. I don't feel it appropriate to ask the authors to repeat the trends and analysis for the Argo + GRACE period (since ~2005). The authors have done some work to investigate the difference between GRACE observations and the reanalyses DSL (Fig A1) and to sensitivity test using different data sets, including Argo-only steric data sets (Section 4.3). I would like to see the authors move the comment from the Appendix B into the main text, in Section 2.1, with a specific reference to the sensitivity tests they do and the period of data they are choosing to use (for the main result and if some of the sensitivity tests are applied over shorter durations).

The abstract could be more precise, in particular to quantify how 'well' the sea-level trend budget matches. Line 8 says the authors can close the SL(T)B on [some] scales but then line 11 says some regions the SL(T)B does not close – it might be clearer to state that the SLTB closes in 100% of the 18 sub-basin regions defined using SOM, but on smaller scales the SLTB can fail to close.

The authors appear to discuss both a SLB (time series) and SLTB and the manuscript would benefit from clarity between these two metrics. (Line 111 states your SLB is actually a SLTB but Figures 4,5 suggest you are also comparing time series.)

Terminology. Reading the Appendix A is a bit confusing! The authors should clarify if I understand correctly, and perhaps change the terminology if needed for clarity. My understanding is the authors are replicating absolute SSHA (observed with IB-corrected altimetry) with a sum of what they call "steric SLA", "GRD" and a manometric "dynamic SLA" term. The GRD term here, is a combination of the GRD terms applicable to absolute sea-level and the barystatic SLA (i.e. it includes the global-mean manometric change at each time step). I appreciate the citation but I think it could be clearer exactly what this 'GRD' includes. And the "dynamic SLA" term in the main paper is the residual of modelled sterodynamic SLA with the steric contribution removed (assuming IB-corrected is consistent with the Gregory et al, 2019, definition). If this is correct then I think the authors should simplify the description in Appendix A.

Review of "Regionalizing the Sea-level Budget With Machine Learning Techniques" by Camargo *et al* (2022) for Ocean Science

Not a necessity, but it would be easier to digest some of the text as Tables or Figures. In Section 2.1 you could tabulate the data sources for each component, description, temporal and spatial resolution, and citations. In Section 4.1, you state that the residuals are improved with scale. I would like to see a scatter plot of the residual (altimetry – sum of components) for the \delta-maps and SOM (i.e. from Fig. 3 c,e) against the area that each of those regions cover to see if there is a simple relationship with scale.

Line by Line Comments

Figure titles: What are the error bars shown (1 or 2 standard errors or standard deviations?).

Abstract Line 7: "besides indicating" can be simplified to "indicate" ("The extracted domains provide … and indicate …")

Abstract Line 8: Suggest you be specific as most readers will skim the abstract. State what period you can close the observational sea level budget for. State within what error it closes (1 or 2 standard errors?). Do your time series SLB also close within 1 or 2 standard errors or is it the trends.

Abstract Line 9: Suggest replace "transport" with "exchange".

Lines 40-45: Novi et al (2021) mostly discuss SST so citation should be moved to preceding sentence.

Lines 68-71 etc: What temporal resolution are you using for steric, DSL, GRD and total SSH (altimetry)?

Line 80 etc: Can you state the temporal and spatial resolution of the Camargo (2020) ensemble steric SL data set, the Camargo et al (2022) GRD data set and the ensemble DSL reanalyses?

Line 101: Can you clarify, do you mean you take the steric SLA from the total SLA for each model in turn.

Line 119: How do you combine 'uncertainty' in the spatial averages in the SOM / delta-map regions?

Section 2.3: Sorry if I missed it, but it isn't clear which data set(s) you use for the clustering analysis (Satellite altimetry \eta_{total} or the reanalysis products you also use; what temporal extent and sampling of the data is used). This is particularly important when you discuss possible mechanisms for the cross-correlation, in Section 3.

Line 135: Do you mean a Gaussian filter with 300 km **half-**width (i.e. the power is 50% at 300 km radial distance), akin to GRACE resolution?

Lines to 172: Out of interest what value of \delta did you use for the \delta-maps threshold? i.e. what is the Figure 3 "uncertainty", is it 1-sigma standard error?

Section 3: This is an interesting discussion but it would be worth noting that the SOM and \delta-maps don't account for auto-correlation in time, i.e. the time lag in the progression of a signal across the ocean basin. So signals that are rapidly propagating compared with the time-sample of your data (monthly?) correlate – typically barotropic, manometric signals - but slower-propagating signals such as the first baroclinic mode will lose correlation in space. So baroclinic signals near the equator, which can propagate faster than those at high latitudes, appear 'better' correlated. i.e. the temporal sampling of the observations that you provide these algorithms dictates which processes appear 'coherent'.

Figure 4 c,d: There is a lot going on there, you need to label solid and dashed lines or just present the time series for one ML region set.

Section 4.3 is very useful to see the sensitivity of the SL(T)B to the size and clustering of regions and the data sets used, but omitting components from the SLB isn't really a sensitivity? Fig 5 relating to the components isn't that informative (sorry!) because the dominance of the steric signal relates to your choice of sampling / data (monthly data that retains the seasonal cycle). But the comparison of the different domain scales is informative. (Actually the box plots in Fig 5a,b ste+dyn+GRD show the 'improvement' for the larger SOM maps that I thought would be interesting to scatter plot with area.)

Line 299: As the authors have shown in Figs 4d,e the seasonal cycle is predominantly steric, which gives rise to a 'better' correlation and lower RMSE since most of the variance in monthly SSHA is the seasonal signal. Line 299, to reduce the apparent dominance of the steric signal in your analysis, you would need to deseason steric SLA and SSHA at each location (not just remove a global mean at each time step).

Line 303: Additionally the more samples you average, the smaller the standard error.

Line 306: Just a note, that measurement errors between altimetry and the sum of components "average out" to zero only if the errors are uncorrelated in space, i.e. they are random, at the scale you are averaging over.

Lines 319-324 and Fig 6: This is a really useful analysis and allays concerns about the data choice that detract from your main conclusions.

Conclusion: You could also add that coherent total sea surface height change might not be same coherency as component parts (steric, manometric dynamic and GRD), so depending it could be 'better' to isolate manometric-dominated variability from steric dominated variability first and then cluster them separately.

Appendix A: Line 377: The "as a result of steric changes" is confusing here and I think unnecessary. I think what the authors are doing is replicating absolute SSHA with a sum of steric SLA, GRD (which here the authors define to include barystatic SLC) and a manometric dynamic SLA term. The latter is, in line with the Gregory et al (2019) definition, the residual of modelled sterodynamic SLA with the steric contribution removed. So it includes mass exchange at any point; changes to the ocean circulation and atmospheric redistribution effects (by mass redistribution, by wind stress and by non-linear interaction due to density changes). (And if the model output weren't IB-corrected, the local atmospheric pressure changes.)

Appendix A: Lines 409-411: There is a strong difference between GRACE and ocean reanalysis in the South Atlantic and you are using a period with a good coverage of Argo float and ship-based in-situ data to characterise the steric component in the reanalysis. You conclude that a source of dynamic SLC may be poorly parameterised in the reanalyses, and perhaps the difference between different reanalyses on different resolutions could be interesting. But my counter-argument would be that the difference covers a large spatial area pointing towards the GRACE spherical harmonic solutions and low-degree corrections.