Authors' Response to Reviewer 1

General Comment. 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.

Response:

Dear Dr. Royston,

Thank you for your feedback and positive review. We understand your reservation, and hope that the answers to your comments have clarified the issues. We have carefully addressed all the issues item by item as follows.

Kind regards,

Carolina Camargo, on behalf of the authors

Comment 1

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)

Response: Indeed, pior to Argo, the steric datasets rely more on reanalyses products, but note that the our ensemble steric estimate is composed of 15 data sets: 5 based on Argo, 5 based on multiple in-situ observations, and 5 based on ocean reanalyses. So while prior to 2002 the ensemble relies more on ocean reanalysis, it does not become heavily weighted by it, as we still have 5 other in-situ data sets. Nonetheless, there is some circularity in using ocean reanalysis' sea-surface height, which incorporates satellite altimetry, to estimate the dynamic component, and then use this dynamic estimate to compare with satellite altimetry. We have moved the comment about the circularity and the validation with GRACE to the main text:

> η_{DSL} is computed from the sea-surface height of 5 ocean reanalyses (Table **??**), by first removing the time-varying global mean from the sea-surface height, and then by removing the local steric anomaly. This procedure is done in each ocean

reanalysis individually, and we then combine the 5 estimates into an ensemble. We acknowledge that this method introduces some circularity to the budget analysis: the reanalysis, used to obtain η_{DSL}, assimilate satellite sea-surface height, and in the budget analysis we compare this estimate with satellite sea-surface height (η_{total}). Compared with the η_{DSL} estimated from Gravity Recovery and Climate Experiment Satellite (GRACE, Tapley et al., 2004), η_{DSL} sea-level trends from 2005-2015 agree on large scale patterns and magnitude of dynamic changes (Figure A1). Note that our budget components do not incorporate GRACE mass changes over the oceans, hence it is an independent estimate for validation.

Comment 2

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.

Response: We have modified the abstract to be more precise:

Using these domains we can close, within 1-sigma uncertainty, the sub-basin regional sea-level budget from 1993-2016 in 100% and 83% of the SOM and δ-MAPS regions, respectively.

Comment 3

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.)

Response: Thank you for your comment, which we agree. We modified line 111 to clarify that the budget analysis included both time series and trends:

> Our sea-level budget includes the comparison of sea-level time series, trends and associated uncertainties.

We also modified the title of section 4.1 to clearly state that we discuss trends in that section.

Comment 4

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.

Response: That is correct: we compare the absolute SSHA from altimetry with the sum of steric, GRD and (manometric-)dynamic SLA. We modified Appendix A to be clearer:

> GRACE measures total mass changes, which can be used to derive estimates of manometric sea-level change over the oceans, that is the sea-level change in response to both the dynamic ocean mass redistribution (η_{DSL}) and to mass redistribution due to the land-ocean mass exchange (η_{GRD}) (Chambers et al., 2004; Royston et al., 2020).

And we clarified it in the main text:

> The dynamic component (η_{DSL}, Figure 1e) refers to mass changes driven by bottom pressure changes, that is, the redistribution of mass that was already in the oceans. Note that, by our definition, the dynamic sea-level change (η_{DSL}) is part of the ocean dynamic sea-level change (Δ_ζ, Gregory et al., 2019), the latter also including the effect of local steric anomalies (η'_{SSL}). That is, the dynamic term here is the residual of the sterodynamic sea-level change with the steric contribution removed (Gregory et al., 2019)

Comment 5

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 δ-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.

Response: Thank you for the suggestion. We have added a table at the end of Section 2.1, summarizing the components description, temporal and spatial resolution and data sources:

dynamics. A summary of the budget components and data sets sources is given on Table 1. Note that all the used data sets have monthly temporal resolution and a 1°x1° spatial resolution.

Table 1. Summary of the sea-level budget components and data sources used in this manuscript.

Symbol	Name	Description	Reference
η_{total}	Observed change	Total sea-level change from satellite altimetry	Ensemble of CMEMS (CMEMS, 2022), JPL MEaSUREs (Zlotnicki et al., 2019), SLcci (SLcci, 2022) and CSIRO (CSIRO, 2022)
η_{SSL}	Steric expansion	Full depth density-driven sea-level change due to ocean temperature and salinity variations	Camargo et al. (2020) and Purkey and Johnson (2010)
η_{GRD}	Mass change	Contemporary ocean mass redistribution due to the land-ocean mass exchange	Camargo et al. (2022)
η_{DSL}	Dynamic change	Mass redistribution due to purely ocean dynamics	Ensemble of SODA (Carton et al., 2018), C-GLORS (Storto and Masina, 2016), GLORYS (Garric and Parent, 2017), FOAM-GloSea (Blockley et al., 2014; Maclachlan et al., 2015) and ORAS (Zuo et al., 2019)

We have also made a scatter plot of the residual against the area of each region (Figure 1). This figure shows how the residual values are much larger when smaller regions are considered, confirming that the residuals have improved with scale. We have added it to Appendix B, as it illustrates the conclusions drawn in Section 4.

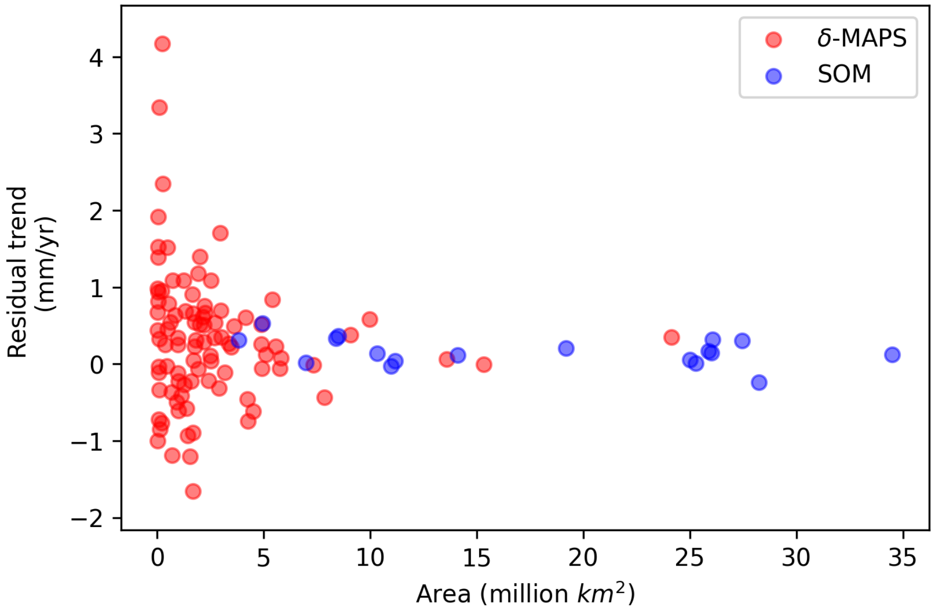

Figure 1: Scatter plot of the residuals (i.e., altimetry minus sum of components) against region size for δ-maps (red) and SOM (blue).

Line by Line Comments

Comment 1

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

Response: The error bars show the 1-sigma uncertainty. We have added this information to the caption of Figure 3 and 4. On Figure 5, the whiskers (not error bars) show the full distribution, while the box shows the quantiles of the data.

Comment 2

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

Response: We have modified it accordingly.

Comment 3

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.

Response: We have modified the abstract to be more specific.

> Using these domains we can close, within 1-sigma uncertainty, the sub-basin regional sea-level budget from 1993-2016 in 100% and 83% of the SOM and δ-MAPS regions, respectively.

Comment 4

Abstract Line 9: Suggest replace “transport” with “exchange”.

Response: We have modified it accordingly.

Comment 5

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

Response: We moved the citation to the preceding sentence, as suggested.

Comment 6

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

Response: We used a monthly temporal resolution for all estimates. We now clarify this when presenting the satellite altimetry data sets and the other contributions:

> We use multi-mission gridded Level-4 data from 4 distribution centers: CMEMS (CMEMS, 2022), JPL MEaSUREs (Zlotnicki et al., 2019), SLcci (SLcci, 2022) and CSIRO (CSIRO, 2022). All of these products use the same reference ellipsoid model (GRS80/WGS). All data sets have a monthly temporal resolution, except for JPL MEaSUREs time series which provides sea surface height data every 5 days and was averaged into monthly means. All data is regridded to 1x1 degree, selected within 66°S to 66°N of latitude, and combined into an ensemble mean, to avoid systematic errors. All the following data sets have the same spatio-temporal characteristics: monthly mean values on a 1°x1° map.

Comment 7

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?

Response: Following the previous reply, we have added this information to the text when the first data sets are introduced. This information is also reinforced when table 1 is introduced (see reply to General Comment 5).

Comment 8

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

Response: We do it model by model, and then combine the 5 estimates into an ensemble. We have clarified the sentence accordingly:

> η_{DSL} is computed from the sea-surface height of 5 ocean reanalyses (Table 1), by first removing the time-varying global mean from the sea-surface height, and then by removing the local steric anomaly. This procedure is done in each ocean reanalysis individually, and we then combine the 5 estimates into an ensemble.

Comment 9

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

Response: We took the area-weighted value of the uncertainties within a SOM/delta-map region. We have added this information to the main text to have it clear for the readers

> For each SOM and δ-MAPS region we take the area-weighted spatial average of the time series, trend and uncertainties.

Comment 10

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 η_{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.

Response: For the clustering analysis we used satellite altimetry from 1993 to 2019 (η_{total}), as stated in Line 132. We added now the information about the temporal and spatial resolution:

> For both clustering techniques we use 1°x1° monthly satellite altimetry time-series (CMEMS, 2022), for 1993-2019, as input.

Comment 11

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?

Response: Yes, that is what was meant. We modified it to 'half-width'.

Comment 12

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

Response: The δ value used in δ-MAPS is not related to the uncertainty shown in Figure 3. In Figure 3 we show the 1-sigma uncertainty, as mentioned in Section 2.2 (line 113) (see also reply to Line-by-line Comment 1). For δ-maps, we actually do not choose the δ parameter but the α parameter, which defines the significance level of the homogeneity test between grid cells. We set α to 0.01, meaning that every cell which is included in a domain has a similarity with the other cells of 99% significance level.

Comment 13

Section 3: This is an interesting discussion but it would be worth noting that the SOM and δ-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'.

Response: Thank you for this interesting comment. We think this is a relevant information to be added to the the discussion. Indeed, the time resolution of data used in this study is not appropriate to adequately identify the propagation of fast signals. However, using SOM one can infer dynamical propagation of signals, as long as the input data has a high frequency temporal resolution. For instance in Liu et al. (2016), it was explored the penetration of the Loop Current into the Gulf of Mexico through the combined SOM analysis in the time and space domain with wavelets, based on daily sea-surface height data. That is, you can analyze how the identified SOM regions correlate in time, for instance computing the cross correlation (or cross-wavelets) between the temporal patterns. We have added this discussion to the first paragraph of Section 3 as follows:

It is also important to note that these clustering methods do not account for auto-correlation in time, that is the time lag in the progression of a signal across the ocean basin. Since we use monthly data, signals that propagate faster than a month (typically barotropic) will be more clearly correlated in our clustering. On the other hand, slower propagating signals, such as the first baroclinic mode, will lose correlation in space, and will not be represented in the identified domains.

Comment 14

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.

Response: Thank you for pointing this out. We chose to reformat the figure, in a way that panels c and d take the entire width of the figure. Since the color of the dashed lines represents the same as the solid lines, we chose to just add the difference between the solid and dashed lines in the caption.

Comment 15

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.)

Response: We agree that omitting components of the SLB is not a real sensitivity analysis, but it does show how the budget improves as we include more components (as expected), and highlights the importance of including both the deep ocean and the dynamic component in the budget. Thus, we decided to keep Figure 5 how it is, but we changed the title of Section 4.3 to "Sea-level Budget Performance", and do not say we are doing a sensitivity analysis, but instead just investigating how the closure of the budget changes when considering different components, area, and datasets:

> ### 4.3. Sea-Level Budget Performance
>
> Here, we investigate the closure of the budget considering (i) the components included in the budget, (ii) the size of the domains and the clustering method, and (iii) the data sets used for each component.

However, it is true that the correlation and RMSE will improve with the addition of the steric contribution, because of its dominance on the seasonal cycle of the altimetry time series. We added this information to Line 298:

> While we get a poorer performance when only considering the dynamic or the GRD component, the budget with only steric already performs relatively well. The improved correlation and lower RMSE with the steric component is probably a result of the seasonal cycle being predominantly steric.

Comment 16

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).

Response: Indeed, not only removing the global mean but also removing the seasonal cycle would be important to reduce the dominance of the steric signal. We have, however, removed this sentence from the manuscript.

Comment 17

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

Response: Thank you for highlight this point. Indeed, the more samples are averaged, the smaller the error. We added this information to the manuscript.

Comment 18

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.

Response: Thank you for the comment. We added this note together with the previous comment:

> Additionally the more samples are averaged the smaller the standard error. However, the measurement errors between altimetry and the sum of components will only compensate each other if the errors are uncorrelated in space, i.e., if they are random at the scale being analyzed.

And indeed, as mentioned in General Comment 5, the scatter plot of the residuals with the domain area confirms the improvement of the budget performance as shown in Figure 5. We have added the scatter plot to Appendix B.

Comment 19

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.

Response: Thank you.

Comment 20

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.

Response: Thank you for your suggestion. That is indeed the case, and we tested SOM on different budget components, and it does give different patterns, specially when GRD patterns are used as input. We now comment on this in the second paragraph of the conclusions:

... we applied a neural network approach, SOM, and a deep-network detection method, δ-MAPS, to identify domains of coherent sea-level variability (Figure 2). Note however, that the coherent patterns will be different whether total sea surface height or the individual components (steric, dynamic, GRD) are considered. Hence, depending on the purpose of the study, one should first remove the unwanted components from total sea-surface height, and then perform the clustering.

Comment 21

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.)

Response: The "as a result of steric changes" had been added as a complement only to the last part of that sentence (ocean bottom pressure changes), to differentiate from the GRD-barystatic sea-level change, which is also a response to ocean bottom pressure changes. But we can see how this can be misinterpreted. Hence, we modified the definition as suggested:

The dynamic redistribution of mass due to ocean circulation and atmospheric redistribution effects is known as dynamic sea-level change (η_{DSL} Gregory et al., 2019; Landerer et al., 2007). η_{DSL} refers to mass changes driven by bottom pressure changes, that is, the redistribution of mass that was already in the oceans, and includes mass exchange at any point by mass redistribution, by wind stress and by non-linear interaction due to density changes.

Comment 22

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.

Response: Thank you for pointing this out. Indeed, this difference could also be due to issues in the GRACE harmonic solutions. We have added a comment to the text:

> Another strong divergence is seen in the South Atlantic, where the positive trends of GRACE are not represented in the reanalysis, possibly suggesting that a source of dynamic sea-level change is not well parameterized in the reanalysis. Alternatively, this divergence might also be an artefact of the GRACE spherical harmonic solutions and low-degree corrections.

References

Chambers, D. P., Wahr, J., & Nerem, R. S. (2004). Preliminary observations of global ocean mass variations with GRACE. *Geophysical Research Letters*, *31*(13), 1–4. https://doi.org/10.1029/2004GL020461

CMEMS. (2022). GLOBAL OCEAN GRIDDED L4 SEA SURFACE HEIGHTS AND DERIVED VARIABLES REPROCESSED (1993-ONGOING) SEALEVEL_GLO_PHY_L4_MY_0 https://doi.org/https://doi.org/10.48670/moi-00148

CSIRO. (2022). Combined TOPEX Poseidon, Jason-1, Jason-2 OSTM, Jason-3 near-global gridded monthly-average sea level product. https://doi.org/http://www.cmar.csiro.au/sealevel/sl_data_cmar.html

Gregory, J. M., Griffies, S. M., Hughes, C. W., Lowe, J. A., Church, J. A., Fukimori, I., Gomez, N., Kopp, R. E., Landerer, F., Cozannet, G. L., Ponte, R. M., Stammer, D., Tamisiea, M. E., & van de Wal, R. S. (2019). Concepts and Terminology for Sea Level: Mean, Variability and Change, Both Local and Global. *Surveys in Geophysics*, *40*(6), 1251–1289. https://doi.org/10.1007/s10712-019-09525-z

Landerer, F. W., Jungclaus, J. H., & Marotzke, J. (2007). Ocean bottom pressure changes lead to a decreasing length-of-day in a warming climate. *Geophys. Res. Lett.*, *34*. http://dx.doi.org/10.1029/2006GL029106

Liu, Y., Weisberg, R. H., Vignudelli, S., & Mitchum, G. T. (2016). Patterns of the loop current system and regions of sea surface height variability in the eastern Gulf of Mexico revealed by the self-organizing maps. *Journal of Geophysical Research: Oceans*, *121*(5), 2347–2366. https://doi.org/10.1002/2015JC011493.Received

Royston, S., Vishwakarma, B. D., Westaway, R., Rougier, J., Sha, Z., & Bamber, J. (2020). Can We Resolve the Basin-Scale Sea Level Trend Budget From GRACE Ocean Mass? *Journal of Geophysical Research: Oceans*, *125*(1), 1–16. https://doi.org/10.1029/2019JC015535

SLcci. (2022). Time series of gridded sea level anomalies. https://doi.org/10.5270/esa-sea

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The Gravity Recovery and Climate Experiment: Mission Overview and Early Results. *Geophysical Research Letters*, *31*(L09607), 1–4. https://doi.org/10.1029/2004GL019920

Zlotnicki, V., Qu, Z., Willis, J. K., Ray, R., & Hausman, J. (2019). *JPL MEASURES Gridded Sea Surface Height Anomalies Version JPL1812* (tech. rep.). PO.DAAC. CA, USA. https://doi.org/doi.org/10.5067/SLREF-CDRV2