Reconstruction of hourly coastal water levels and counterfactuals without sea level rise for impact attribution – Responses to Reviewer Comments

RC1

GENERAL COMMENTS

The submitted manuscript presents a compelling study on global sea level, considering various and diverse datasets. Additionally, the authors have generated and provided a counterfactual dataset that excludes the sea level trend. The work is highly interesting and relevant to the journal to which it has been submitted. I recommend its acceptance and publication following a minor review.

We thank the reviewer for taking time to review the manuscript and for the specific feedback that helped us to improve the manuscript. Please find below our responses in *blue italic text*. Pasted text from the revised manuscript is in *italic*.

SPECIFIC COMMENTS

1- The introductory chapter is fine, but lines 94-99 should be relocated to the methods section.

Thank you. We removed the lines from the introduction and merged them into the beginning of the methods section. Due to the several other comments on the method section structure we revised the section completely.

2- It would be beneficial to provide direct links to the datasets used in this study, including CoDEC, HR, VLM, GESLA-2, and altimetry. Currently, the Data and Code Availability section only mentions codes and counterfactual data. Could you please clarify if HR is still available upon request by contacting Sonke Dangendorf? Additionally, it would be helpful to explain the process for accessing all the datasets easily.

Thank you for pointing out the need for greater clarity and accessibility regarding our datasets. The HR and CoDEC datasets are now archived on Zenodo for easier access https://doi.org/10.5281/zenodo.8322750. We updated the Data and Code Availability section with direct links to all datasets used in our study see lines 604-620:

The source code (v1.1) underlying the analysis and producing the figures presented in the paper is archived at https://doi.org/10.5281/zenodo.8400878 (Treu, 2023). All code is open to use under the Creative Commons Attribution 4.0 International license.
We provide direct links to all datasets used in this study.

- *CoDEC and HR: https://doi.org/10.5281/zenodo.8322750*
- *VLM: https://doi.org/10.5281/zenodo.8308347*
- *PSMSL tide gauge data, retrieved 26 Sep 2022 from https://psmsl.org/data/obtaining/. An archived version of this dataset can be downloaded from https://psmsl.org/data/obtaining/year_end/2022/.*
- *GESLA-3 tide gauge data, retrieved 23 Aug 2023 from https://gesla.org*
- *We use two different sources of satellite altimetry*
    - *To separate the seasonal cycle from HR, we use the same satellite altimetry dataset that was employed in the production of HR. The dataset is a merged product of TOPEX/Poseidon and Jason altimeter missions and was at that time distributed by AVISO. This dataset, including the corrections described in Dangendorf et al. 2019 is archived together with the HR dataset at https://doi.org/10.5281/zenodo.8322750*
    - *To reference our dataset to the WGS 84 geoid, we use the CMEMS mean dynamic topography dataset SEALEVEL_GLO_PHY_MDT_008_063, available at https://doi.org/10.48670/moi-00150*

The presented counterfactual climate dataset at tide gauge locations is archived at https://doi.org/10.5281/zenodo.8400053 (Treu et al., 2023) and based on v1.1 of the source code.

3- Line 110 is unclear to me regarding the meaning of "adjust it for residual VLM" I would appreciate further clarification on how the global geocentric sea level was precisely obtained. Is the procedure explained in lines 175-181? Please provide more information to clarify this point.

Thank you for this remark. We indeed explained this procedure in lines 175-181. We rephrased the description of the procedure and shifted it to line 197-204 for clarity. The passage now reads:
We remove VLM contributions from HR to represent the evolution of water levels relative to the geoid (geocentric). HR contains VLM from long-term glacial isostatic adjustment (GIA) since the glacial maximum 21,000 years ago, and from short-term crustal responses to present-day ice melt since 1900 (Pfeffer et al., 2017; Spada, 2017; Riva et al., 2017). GIA is explicitly modeled in HR and can thus be readily taken out. The crustal responses to present-day ice melt is implicitly contained in HR through cryostatic fingerprints that are fitted to tide gauges. We use the annual reconstructions of the crustal responses to present-day ice melt from glaciers, the Antarctic ice sheet and the Greenland ice sheet from Frederikse et al. (2020) to remove this contribution. For the HCC dataset we deseasonalize the geocentric version of HR by removing the monthly average climatology over the years 1993-2015.

4- Could you please provide an explanation and justification for using the GESLA-2 dataset in this study, despite the availability of GESLA-3 since November 2021? Is there any potential impact on the results due to this choice? Furthermore, there appears to be a discrepancy in the references provided for the GESLA dataset. Initially, Muis et al. 2020 is mentioned (line 128), while later, Woodworth et al. 2016 is referenced (line 160) for the same dataset. Could you clarify why two different references are used for the same dataset? Lastly, it would be beneficial to include the expanded form of the GESLA acronym, similar to how you later provide the expanded form for PSMSL.

Thank you for this remark. We updated our work to the GESLA-3 dataset and accordingly updated Figures 1, 3, 4, and 5. We also updated the numbers in Table 1. While the specific values changed, they do not change our overall results and conclusions. It is worth noting that the HCC dataset itself is independent of the choice between GESLA-2 or GESLA-3 since they are only used for evaluation. We now reference only Woodworth et al 2016 for GESLA, which is the correct reference. We also write GESLA out in full at its first occasion. In line 177 we still refer to GESLA-2, as we there report results from Muis et al. (2020).

5- Line 165: "with at least one year of observations in the interval 1993-2015" I'm still not convinced. Is this truly sufficient to allow for a valid alignment? Please provide further information on this matter.

Sorry, we have been imprecise here and also updated the requirements for data availability, as we now use the GESLA-3 dataset. We implement three slightly different requirements for data availability for the different types of analysis performed in this study. We also added a test for the uncertainty in the vertical alignment to the Supplementary Material. We now describe this in the main text in the 'Tide Gauge datasets' paragraph of the Materials and Methods section, see lines 234-269.

In summary we use

- *663 PSMSL records of at least 20 year length and with at least 30 percent data coverage in the 1993-2012 period.*
- *One subset of GESLA-3 with at least 30 percent of valid years between 1979 and 2015 to analyze annual sea level anomalies. Here we only consider years with at least 11 months of valid observations*
- *Another subset of GESLA-3 to analyze the 99th percentile of surge levels. Here we require at least 90 percent data coverage in this period which leaves 999 stations for the analysis.*

6- Lines 183-184: This is repetitive. Please double-check and remove any duplicated statements in the Materials and Methods chapter, including line 199 and 207.

We have refined the Materials and Methods chapter to achieve a more rigorous structure. Processing steps specific to particular datasets have been relocated to their respective subchapters for clarity. Moreover, the 'Factual Dataset' section has been reorganized for enhanced readability.

7- Line 185: Is it truly appropriate to extrapolate VLMs all the way back to 1900? Wouldn't this introduce a considerable margin of error and oversimplify the analysis? While this approach may be suitable for GIA and geological components, VLMs entail more intricate considerations. How did you address the substantial non-linear effects (both spatially and temporally) caused by underground fluid exploitation from WWII to the 1990s?

This is an absolutely valid point. As we want to provide the impact-relevant metric of relative sea level rise, and VLM is a key component of relative rise, it is however difficult to justify not providing an estimate of this component. The alternative would be the assumption of zero VLM, which we know is wrong, or to only include long-term GIA, which would ignore the recent advances in the field.

It is true that given the sparsity of direct observational data, particularly prior to the 1990s, extrapolating VLMs back to the early 20th century makes our estimates error prone and that anthropogenic influence post WWII adds additional difficulties.

We have tried to address this issue as rigorously as possible with a state-of-the art method. We employed a probabilistic approach to reconstruct VLM. This reconstruction is derived from all presently available observations (GNSS, tide gauges, and altimetry) from 1995-2020 and provides estimates of linear trends, present-day VLM variability, and the uncertainties associated with these components. We only extend the linear trend to the past. In regions affected by tectonic activity we use the interseismic trend only.

It is difficult to exclude anthropogenic influences like fluid extraction from the trend estimation. However this only affects specific regions. Oelsmann et al. 2023 provide uncertainties of the linear trend estimate that can serve as an indicator which regions are problematic. This is under the simplified assumption that the uncertainties, which are informed by the variability observed over the recent period (1995-2020), are representative for unobserved effects before that period. We added the following text to our Discussion section in lines 524-529:

Given the sparsity of direct observational data, particularly prior to the 1990s, extrapolating VLM back to the early 20th century makes our estimates error prone and anthropogenic influence post WWII adds additional difficulties. It is difficult to exclude anthropogenic influences like fluid extraction from the trend estimation. Further advances in this field should be incorporated in future sea level assessments. By providing a geocentric version of the HCC dataset it is possible to combine our water levels with other reconstructions of VLM.

8- Line 193: "We thus expect that it depends on the location which product performs better", Could you please provide further clarification or explanation regarding the basis for this expectation?

The extended explanation in lines 125-136 reads now:

Barotropic water level changes due to wind and atmospheric pressure on time scales longer or equal to one month are covered in both, the HR, and the CoDEC dataset. We use frequency *filtering to prevent double counting. CoDEC models these water level changes explicitly whereas HR is based on a statistical reconstruction method based on sparse observations. However, HR is not restricted to barotropic variations alone and covers the full spectrum of intra-annual and longer sea level variations (including sterodynamic and barystatic processes). The dominating process at different time scales depends on local conditions (Zhu et al., 2021; Dangendorf et al., 2014). We thus expect that it depends on the location which product performs better. Our method's cutoff frequency determines which scale of variations is captured by each product. CoDEC captures variations with higher frequencies than the cutoff, while HR captures variations with lower frequencies. We have tested for different cutoff frequencies and how it affects the performance of the combined product when compared to tide gauges. We varied the cutoff frequency for values of 1, 2, 3, 4, 5, 6, and 12 months and found an optimal cutoff frequency of 3 months (90 days).*

9- Lines 208-210: This sentence appears to be quite complex. Would it be beneficial to use acronyms for each byproduct or perhaps express the information in the form of a mathematical equation?

We have refined the Materials and Methods chapter to achieve a more rigorous structure. Processing steps specific to particular datasets have been relocated to their respective subchapters for clarity. Moreover, the 'Factual Dataset' section has been reorganized for enhanced readability.

10- The provided link (2) directs to the AVISO website homepage, which does not provide information about a specific altimetry product. Furthermore, the link (3) leads to a product that is no longer accessible on the CMEMS website. It is unclear which altimetry product has been utilized for this analysis. If a DUACS L4 product from CMEMS or C3S was employed, it should be clearly indicated throughout the text instead of referencing AVISO, as AVISO is no longer responsible for distribution.

Thank you for highlighting this oversight. We've updated the links and have enhanced our data availability section for clarity. The section now explicitly specifies the altimetry products used for our analysis in lines 612-618.

We use two different sources of satellite altimetry

- *To separate the seasonal cycle from HR, we use the same satellite altimetry dataset that was employed in the production of HR. The dataset is a merged product of TOPEX/Poseidon and Jason altimeter missions and was at that time distributed by AVISO. This dataset, including the corrections described in Dangendorf et al. 2019 is archived together with the HR dataset at https://doi.org/10.5281/zenodo.8322750*
- *To reference our dataset to the WGS 84 geoid, we use the CMEMS mean dynamic topography dataset SEALEVEL_GLO_PHY_MDT_008_063, available at https://doi.org/10.48670/moi-00150*

11- Counterfactual water levels: I would appreciate further elaboration on the rationale behind choosing a quadratic fit for detrending the series. Could you provide more details on why this specific model was selected and whether it effectively removes the trend and accounts for nonlinearities? Furthermore, in line 220, it is mentioned that the effects of nonlinearity "can be expected to be minor." Was any testing conducted to assess the accuracy of trend removal after applying the quadratic fit? Lastly, in line 398, you clarify that the quadratic trend estimates do not account for the non-linear behavior discussed in Dangendorf et al. 2019 (but also Slangen et al. 2016 and Frederikse et al. 2020 should be taken into account and referenced). Given this, is your statement "we expect our trend estimation to largely exclude natural variability…" sufficiently strong and reliable?

We employ a quadratic trend model to account for an accelerated trend in sea level rise. We now clarified this in the manuscript in lines 321 to 325:
There is increased evidence for an acceleration in global sea level rise (Church and White, 2011; Hay et al., 2015; Frederikse et al., 2020; Dangendorf et al., 2017, 2019). To account for the accelerated trend in sea level rise, we employ a quadratic trend model. For each location, we estimate this trend using linear regression on the annual mean time series, setting the intercept to the average sea level from 1900-1905. After removing this long-term trend, we obtain the counterfactual time series.

Concerning the robustness of the trend estimation we agree with the reviewer that additional statistical tests were necessary. We therefore added a robustness test based on moving block bootstrapping to the supplementary materials (Text S2 and Fig. S2). We also added the following lines to the manuscript in lines 325-330:
We evaluate the robustness of this trend estimate using the moving block bootstrapping algorithm as described by (Mudelsee, 2019). We find that the total sea level rise derived from the trend estimate varies depending on the location with standard deviations ranging from 3.4 mm to 16.7 mm and an average standard deviation of 8.1 mm (See supplementary materials Text S2 and Fig. S2). While uncertainties exist, this demonstrates that the quadratic model provides a robust representation of the long-term trend in sea level rise. Importantly, the uncertainty associated with our model is relatively small when compared to the uncertainties in the factual dataset itself.

Furthermore, in relation to your queries about non-linearity as discussed in Dangendorf et al. 2019, we acknowledge the necessity of referencing additional studies such as Slangen et al. 2016 and Frederikse et al. 2020. While our findings indicate that estimation based on a quadratic trend is sufficiently robust for this application we acknowledge that future work could employ a more complex trend model, accounting for the main modes of climate variability. We updated lines 573-577 in the Discussion section accordingly.
The quadratic model provides robust trend estimates, given the fact that we do not extrapolate the trend into the future, which would increase uncertainties. To further increase the robustness of the trend estimate, future studies should include predictors for the main modes of climate variability as for example in (Menéndez and Woodworth, 2010; Marcos and Woodworth, 2017; Wang et al., 2021).

12- Line 238: In addition to referencing "Fig. 2a" and before "Fig. 2c," it is necessary to include and briefly explain Fig. 2b.

We agree the respective figure was not well introduced in the text. We changed the respective sentence in lines 355-359 to:
Our dataset reflects well the different trends in different world regions (Fig. 2a) and shows a comparable root mean squared error (RMSE) against observations. Figure 2b shows the root mean squared error (RMSE) of the monthly HCC dataset and the HR dataset against observations. Figure 2c depicts latitudes for all tide gauge stations from the PSMSL record that were used to analyze monthly water levels, aligned by ocean basins.

13- Lines 308-322: It is advisable, in my opinion, to transfer a substantial portion of the information in this paragraph to the methods section. By doing so, the procedural details can be appropriately placed, and the density of this paragraph can be reduced.

This is a good suggestion and we make the text largely part of our refined Materials and Methods chapter specifically in the 'Tide Gauge datasets' paragraph.

14- Lines 319-321: Could you please provide more explanation as to why specifically the 18 highest values were selected? Do these values correspond to the >99th percentile? It would be beneficial to present and clarify this information in a more explicit manner.

We clarified the respective paragraph and describe now explicitly how we selected the 99th percentile of maximum surge levels. The text in lines 463-477 reads now:
We therefore analyze the 99th percentile of daily maximum surge levels. To that end, we pick the one percent highest daily maximum surge levels from the observational data in the years 2011-2015 and compare those to our factual and counterfactual dataset. We focus solely on sea level anomalies, to level out differences in surge height between different stations, caused by permanent differences to the geoid. We calculate sea level anomalies for tide gauge records by removing the mean sea level over time steps with valid observations in the period 2011-2015. To reduce the alignment error, we calculate the mean sea level for the factual dataset only for time steps with valid observations in the tide gauge record. From that we derive sea level anomalies for the factual and counterfactual by removing the factual mean sea level from both datasets. This maintains the difference between the factual and counterfactual datasets resulting from long-term sea level rise.

15- Lines 355-358: It seems that this paragraph is redundant and lacks a clear purpose. It partially repeats information from the methods section and contains elements that would be more suitable for the discussion or conclusions. It might be necessary to either remove this paragraph or rephrase it to ensure its coherence and relevance within the context of the text.

We remove that paragraph and also remove the orange markers in Figure 5b representing the total sea level rise in the counterfactual. It is by design of the counterfactuals that there is zero relative sea level rise (except from some noise) and it is therefore not necessary to show it in a Figure.

16- Lines 369-375: This paragraph appears to be misplaced in the discussion section. It would be more suitable to relocate it to either the methods section or the introduction, where the issue of VLM is first introduced and explained.

We agree and shift the paragraph to the revised Materials and Methods section, where VLM is discussed in more detail in lines 223-230.

17- Lines 385-389: This paragraph seems to duplicate information that has already been discussed in the methods section. It is unclear why it is repeated in this particular location.

We removed the paragraph to avoid duplication.

18- Figure 1: Due to the current resolution, it is challenging to discern the three distinct symbols. To enhance clarity, it is recommended to either enlarge the map or incorporate a color code in addition to the symbology.

We have updated Figure 1 and now use distinct symbols and varied color schemes to distinguish the different subsets of tide gauges.

19- Figure 2: The meaning of the numbers on the x-axis is not apparent. Could you please specify whether they represent a progressive pure number of TGs considered? It would be beneficial to provide this clarification either in the figure itself or in the caption. The same clarification should also be provided for figures 2 and 5 in relation to the x-axis.

The numbers on the x-axis indeed represent a progressive pure number of TGs considered. We added to the figure captions of figures 2, 4 and 5 :
A progressive integer of the considered tide gauge is plotted on the x-axis.

20- Figure 3: The use of blue, black, pink, and red colors together is not considered colorblind friendly, which can pose accessibility challenges. To ensure inclusivity, it is advisable to utilize a different color palette or employ different line styles (such as solid, dash-dotted, dotted) to differentiate the time series. It is worth noting that the HR time series already meets the criteria for colorblind-friendly representation.

We have incorporated different line styles (solid, dash, dotted, etc.) in Figure 3 to distinguish between the time series and enhance their differentiation. This adjustment should aid readers with color blindness in interpreting the figure.

Regarding the color palette, we fully recognize the importance of making our graphics accessible to all readers, including those with color vision deficiencies. However, due to the sheer number of datasets we're representing, finding a color scheme that is both distinct for each dataset and universally colorblind-friendly has proven challenging. We did our best to strike a balance and will consider alternate solutions in future publications to enhance inclusivity further.

TECHNICAL CORRECTIONS

- The usage of terms such as 'sea-level rise' and 'sea level rise' is inconsistent throughout the manuscript. The same inconsistency applies to terms like 'sea level change,' 'sea level variability,' and 'GESLA2.' I recommend using hyphens consistently in all of these terms. Further, there is inconsistency among 'water levels' and 'waterlevels', and 'Figure' with 'Fig.'. Please ensure the homogeneity of the terms used.

We corrected the inconsistencies and are now using 'sea level' and 'water levels' without a hyphen, GESLA-2 and GESLA-3 with a hyphen. The term 'Figure' is only used at the beginning of a sentence otherwise we use the abbreviation Fig.

- Line 36: remove a round bracket.

Done

- double space found at line 136, 194, 201, 346, 366

Done

- line 180: "gauges.We" space needed.

Done

- RSL has never been specified, do that at the first term appearance (should be line 34).

Done

- line 268: Tanzania

Done.

RC2

GENERAL COMMENTS

The ms shows a new analysis of global sea level using different data sets. The core of the study is the counterfactual analysis to exclude long-term trends in sea level. In my opinion, the study is interesting and valuable of publication after major revision. See comments below.

We thank the reviewer for taking time to review our manuscript and for making specific suggestions on how we can improve it. Please find below our responses in *blue italic text*. Pasted text from the revised manuscript is in *italic*.

SPECIFIC COMMENTS

1 – **Introduction**. Lines 95-101. I suggest the Authors to provide link to data sets used in the study to facilitate readers in understanding the work done and possibly to replicate the analysis.

Thank you for pointing this out. We updated the Data and Code Availability section with direct links to all datasets used in our study see lines 604-620:
The source code (v1.1) underlying the analysis and producing the figures presented in the paper is archived at https://doi.org/10.5281/zenodo.8400878 (Treu, 2023). All code is open to use under the Creative Commons Attribution 4.0 International license.
We provide direct links to all datasets used in this study.

- *CoDEC and HR: https://doi.org/10.5281/zenodo.8322750*
- *VLM: https://doi.org/10.5281/zenodo.8308347*
- *PSMSL tide gauge data, retrieved 26 Sep 2022 from https://psmsl.org/data/obtaining/. An archived version of this dataset can be downloaded from https://psmsl.org/data/obtaining/year_end/2022/.*
- *GESLA-3 tide gauge data, retrieved 23 Aug 2023 from https://gesla.org*
- *We use two different sources of satellite altimetry*
    - *To separate the seasonal cycle from HR, we use the same satellite altimetry dataset that was employed in the production of HR. The dataset is a merged product of TOPEX/Poseidon and Jason altimeter missions and was at that time distributed by AVISO. This dataset, including the corrections described in Dangendorf et al. 2019 is archived together with the HR dataset at https://doi.org/10.5281/zenodo.8322750*
    - *To reference our dataset to the WGS 84 geoid, we use the CMEMS mean dynamic topography dataset SEALEVEL_GLO_PHY_MDT_008_063, available at https://doi.org/10.48670/moi-00150*

The presented counterfactual climate dataset at tide gauge locations is archived at https://doi.org/10.5281/zenodo.8400053 (Treu et al., 2023) and based on v1.1 of the source code.

2 – **Material and methods**. Line 111. Specifies how geocentric water levels are calculated with the probabilistic VLM. This is a relevant point for the analysis and may affect results.

We agree that there was an unclarity how VLM has been removed and added, and due to the several comments on the structure on the Materials and Methods chapter we decided to restructure it completely. We now describe in lines 197-204 in more detail how geocentric water levels are derived from HR by removing all VLM contributions.

We remove VLM contributions from HR to represent the evolution of water levels relative to the geoid (geocentric). HR contains VLM from long-term glacial isostatic adjustment (GIA) since the glacial maximum 21,000 years ago, and from short-term crustal responses to present-day ice melt since 1900 (Pfeffer et al., 2017; Spada, 2017; Riva et al., 2017). GIA is explicitly modeled in HR and can thus be readily taken out. The crustal responses to present-day ice melt is implicitly contained in HR through cryostatic fingerprints that are fitted to tide gauges. We use the annual reconstructions of the crustal responses to present-day ice melt from glaciers, the Antarctic ice sheet and the Greenland ice sheet from Frederikse et al. (2020) to remove this contribution.

We also added the following sentence in lines 112-113 to make it more clear that relative water levels are derived by combining the geocentric water levels with the probabilistic VLM construction from Oelsmann et al (2023).
[...] Derived from a geocentric, deseasonalized version of the hybrid reconstructions (HR) dataset from Dangendorf et al. (2019) and a probabilistic VLM reconstruction from Oelsmann et al. (2023).

3 – **Tide gauge data set**. Line 163. I think it is risky to interpolate the monthly averages of tide records to hourly resolution. The authors should provide an example of this interpolation with relative uncertainties to make the reader understand the reliability of the method used.

The description was misleading here and is now also updated to GESLA-3, which contains only records with hourly resolution or higher sampling frequency. We therefore did not interpolate monthly averages of tide gauge records to hourly resolution but rather aggregated records with sub-hourly resolution to hourly resolution. The respective text in lines 247-250 reads now:

To evaluate the higher frequencies shorter than a month we use the tide gauge data provided by the GESLA-3 database (Woodworth et al., 2016; Haigh et al., 2023; Caldwell et al., 2015), which is provided on an hourly or sub-hourly sampling frequency. We aggregated all records with higher sampling frequency to hourly resolution by taking the average over all timesamples that fall within plus or minus 30 min of a specific hour.

Line 165. Probably “at least one year of observations in the interval 1993-2015" is not enough to get a valid alignment. Again, authors should provide an example of this interpolation with relative uncertainties.

Sorry, we have been imprecise here and also updated the requirements for data availability, as we now use the GESLA-3 dataset. We implement three slightly different requirements for data availability for the different types of analysis performed in this study. We also added a test for the uncertainty in the vertical alignment to the Supplementary Material. We now describe this in *the main text in the 'Tide Gauge datasets' paragraph of the Materials and Methods section, see lines 234-269.*

In summary we use

- *663 PSMSL records of at least 20 year length and with at least 30 percent data coverage in the 1993-2012 period.*
- *One subset of GESLA-3 with at least 30 percent of valid years between 1979 and 2015 to analyze annual sea level anomalies. Here we only consider years with at least 11 months of valid observations*
- *Another subset of GESLA-3 to analyze the 99th percentile of surge levels. Here we require at least 90 percent data coverage in this period which leaves 999 stations for the analysis.*

4 – **Factual water levels**. Line 185-181. Which model did the author adopt to estimate the long-term (secular) effect of GIA?

The work of Oelsmann et al. (2023), which we use to estimate VLM, incorporates long-term secular VLM based on present day observations of the combined effects of GIA and various other VLM processes. In regions that are dominated by GIA, such as the Baltic Sea and the NE-US coast, the trends of the VLM reconstruction align well with a GIA model by Caron et al. 2018 as can be seen in Oelsmann et al. 2023, Fig. 4C). We added a sentence in lines 208-211 to provide this information.

Their [Oelsmann et al. 2023] approach incorporates long-term secular VLM based on present day observations of the combined effects of GIA and various other VLM processes. In regions that are dominated by GIA, such as the Baltic Sea and the NE-US coast, the trends of the VLM reconstruction align well with a GIA model by Caron et al. 2018 (as can be seen in Oelsmann et al. 2023, Fig. 4C).

Concerning vertical land motion, did the authors evaluate the effects of large earthquakes on the sea level? (e.g., Indonesia 2004; Japan, 2011, etc.).

We did not evaluate the effect of large earthquakes, but the manuscript describing the VLM component of our dataset (Oelsmann et al. 2023) does. Their method has been designed to robustly account also for effects from large earthquakes. Only linear, interseismic trends are used for extension beyond the data coverage period in tectonically active regions. See Oelsmann et al. 2023, lines 318ff and their Figure 2 for details.

5 – Lines 187-210. Text is hardly understandable. I suggest the authors to make it more linear also through the use of a figure with a flowchart.

We agree that this was not well written. We restructured the Materials and Methods section completely to make the text more understandable and the flow more linear. This in our opinion now is understandable without an additional figure.

6 - Line 220, the effects of nonlinearity "can be expected to be minor." What do you mean by "minor"?

We agree with the reviewer that additional statistical tests were necessary to assess the robustness of the trend estimation. We therefore added a statistical test based on moving block bootstrapping to the supplementary materials (Text S2 and Fig. S2). We removed line 220 and added the following lines to the manuscript in lines 225-230:
We evaluate the robustness of this trend estimate using the moving block bootstrapping algorithm as described by (Mudelsee, 2019). We find that the total sea level rise derived from the trend estimate varies depending on the location with standard deviations ranging from 3.4 mm to 16.7 mm and an average standard deviation of 8.1 mm (See supplementary materials Text S2 and Fig. S2). While uncertainties exist, this demonstrates that the quadratic model provides a robust representation of the long-term trend in sea level rise. Importantly, the uncertainty associated with our model is relatively small when compared to the uncertainties in the factual dataset itself.

7 – **Long term sea level trends**. Line 246-247. RMSE might have a bias, I'd suggest a statistical test comparing the distributions of the samples. The difference is just 0.23 cm: please explain if there is a real improvement from this analysis.

Thank you for the constructive suggestion. To address this, we employed the Wilcoxon signed-rank test, a non-parametric test suitable for comparing two paired samples. The hypotheses for the test were defined as:

H0 (null hypothesis): The distribution of RMSE values for HCC and PSMSL is the same as or greater than the distribution of RMSE values for HR and PSMSL.

H1 (alternative hypothesis): The distribution of RMSE values for HCC and PSMSL is less than the distribution of RMSE values for HR and PSMSL.

Our analysis produced a p-value of 5.36×10^{-10}, which is significantly below the commonly accepted threshold of 0.05. Such a low p-value offers robust evidence against the null hypothesis, suggesting that the RMSE values for HCC and PSMSL are significantly less than those for HR and PSMSL.

While the statistical evidence is compelling, we acknowledge your observation about the small absolute difference of 0.26 cm (this is the changed value after the updated analysis with adjusted requirements on data availability, see comment 3). Even though the difference is statistically significant, its magnitude relative to the total errors is indeed modest. We wish to emphasize that our central aim was not merely to enhance precision over the HR dataset. Our primary objective was to present a consistent high-frequency dataset detailing long-term trends, complemented by a counterfactual. We updated the text in lines 370-372 which reads now:

We performed a Wilcoxon signed-rank test to compare the RMSE samples and found that the improvement is statistically significant with a p-value of 5.36×10^{-10}. However, the improvement of 0.26 cm is only modest when compared to the total error magnitudes.

8 – **Intra-annual variability**. Line 264. I suggest doing a spectral analysis alongside fig.3 to see any anomalies.

We thank the reviewer for this suggestion. A spectral analysis would indeed provide deeper insights into the dominant frequencies of water level change at the five chosen tide gauges, as well as their representation in CoDEC and HR. However, the five stations were chosen as illustrative examples, and the dominant frequencies of water level change are expected to vary across different locations. To comprehensively capture these variations, a more detailed regional analysis would be required, which goes beyond the primary objective of this study: to develop a consistent high-frequency dataset complemented by a counterfactual. We added to our discussion in lines 352-355:

Future work could explore using region-specific cutoff frequencies to better harness the individual strengths of HR and CoDEC. This approach, however, would necessitate a comprehensive analysis of the dominant time scales of variability within each dataset for different regions.

9- Lines 319-321: how were the 18 points selected?

We agree that the description of selecting extreme water levels was not clear. We improved the paragraph to outline how we select the 99th percentile of daily maximum surge levels. The respective text in lines 463-470 reads now:

We therefore analyze the 99th percentile of daily maximum surge levels. To that end, we pick the one percent highest daily maximum surge levels from the observational data in the years 2011-2015 and compare those to our factual and counterfactual dataset. We focus solely on sea level anomalies, to level out differences in surge height between different stations, caused by permanent differences to the geoid.

10 – **Extreme water levels**. Line 321. Since the average value has been removed, the 1% can be high. What value does the surge component correspond to?

The average value over the period 2011-2015 has been removed to neglect permanent difference to the geoid between different stations. We have refined the description text in lines 470-477, which reads now:

We calculate sea level anomalies for tide gauge records by removing the mean sea level over time steps with valid observations in the period 2011-2015. To reduce the alignment error, we calculate the mean sea level for the factual dataset only for time steps with valid observations in the tide gauge record. From that we derive sea level anomalies for the factual and counterfactual by removing the factual mean sea level from both datasets. This maintains the difference between the factual and counterfactual datasets resulting from long-term sea level rise.

11 – **Discussion**. Line 363-371. It is unclear here whether the VLM calculated from the geodetic data has been added to the GIA value. The GNSS stations record the VLM which also includes the GIA, obviously.

Exactly, the observation based VLM reconstruction contains the effect of GIA as well. This VLM reconstruction is the only source of GIA in the HCC dataset. This is clarified now in lines 208-209:

Their approach incorporates long-term secular VLM based on present day observations of the combined effects of GIA and various other VLM processes.

We shortened Lines 363-371 to clearly distinguish our approach from other approaches that explicitly model GIA. We also moved those lines to Materials and Methods section to lines 223-226:

We incorporated this VLM dataset, which is directly derived from observations, as the most independent source for such data. Alternatives to this approach exist and were already used in earlier datasets. One possibility is to only account for VLM that is caused by GIA which can directly be modeled, or implicitly through cryostatic fingerprints in the case of responses to present day ice melt (Dangendorf et al., 2019).

FIGURES

Figure 2, 4 and 5. Hardly readable. I suggest using thicker strokes and larger dots.

Thank you for your feedback. We acknowledge the challenges in readability due to the fine strokes and small data points. However, the stroke thickness is a reflection of the sheer volume of data points we're presenting in these figures.

Our intention in choosing this presentation style was to effectively highlight the region-specific aspects of our analysis. Given the highly uneven global distribution of tide gauge data, maps might not effectively capture the nuances of our findings. This approach, while dense, provides a more granular view of the data, which we believe is vital. When we aggregate the data regionally, the results can be substantially skewed due to variations in the number and quality of tide gauge records.

We believe that offering this detailed perspective complements our aggregated analysis, ensuring a comprehensive understanding of the data. While we understand and appreciate the concerns about clarity, we feel that retaining these visuals is imperative for the depth and precision they provide to our study.

Figure 2 and 5. Please, explain the meaning of the numbers on the x-axis.

The numbers on the x-axis represent an index for the tide gauges considered. We changed the figure captions of Figures 2 and 5 accordingly.

Figure 3. How do you explain the difference in the data for Miami?

If you are referencing the discrepancy observed for Miami in October and November 2001, it's important to note that this variation falls within the uncertainty margins of the HR dataset. Given that the HR represents a global statistical reconstruction of water levels, it's expected that it won't perfectly match observed time series in every instance. However, it's noteworthy that the HCC dataset aligns more closely with the observations compared to CoDEC. Therefore, even for Miami, integrating HR enhances the accuracy of the water level reconstruction. We discuss this now in the manuscript in lines 411-414.

Notably, also after including HR, the HCC water levels don't perfectly match the observations. discrepancies remain, e.g. in Miami, USA between October and November 2001. However this falls well within the uncertainty margins of the HR dataset, given that it represents a global statistical reconstruction of water levels.