Review response

To Christopher Watson, August, 26 2020

We thank Christopher Watson for his very helpful and comprehensive comments and his profound corrections. His remarks pointed out several vital aspects, which were so far missing or insufficiently covered in the manuscript. In response to his comments, we added more precise explanations, specified technical details and discussed the results much more thoroughly. Inclusion of these aspects greatly helps to enhance the manuscript and makes it more comprehensible and complete.

General note: We changed the Envisat mission data version from V2.1 to V3 (for the ALES altimetry data). This influences all associated dataset combination (ALES-PSMSL-250km, ALES-GESLA-250km and ALES-GESLA-ZOI). Because the statistics are not significantly altered, the key messages of the study remain. All statistics and plots are updated accordingly (i.e. Figure 1,2,4,5,6,7 and table 2).

We use italic formatting to answer the comments. Existing text is marked in blue and changes in the text are highlighted in red. All line numbers refer to the originally submitted version.

This manuscript by Oelsmann et al provides a contribution to the improved determination of vertical land motion (VLM) at tide gauge locations using satellite altimetry. The work advocates for a refined approach to selecting valid altimeter data in a variable 'zone of influence' around a specific tide gauge to improve the accuracy and precision of VLM estimates. The team make incremental advances using a coastally retracked altimeter dataset and tide gauge data with various temporal resolutions. The focus of the paper has a rich history in the literature and this technical contribution provides a worthy contribution to progressing the method and ultimately achieving improved VLM estimates around the coast. The demand for improved VLM is clear – VLM is a vexing issue with manifold impacts across the scientific and the broader community. The manuscript is well presented, well crafted and generally presents a compelling and comprehensive case worthy of publication in this outlet. Noting my support for this manuscript, I have a few substantive issues that I feel would benefit from consideration and discussion/revision in the paper. I present these issues below, followed by a list of technical clarifications and minor suggestions/corrections that may benefit the manuscript.

Main Issues:
Central to the manuscript is the comparison of various altimeter minus tide gauge differences against a single 'ground truth' GPS solution. Significant differences in the solutions derived from different variants (e.g. Fig 4b) are evident, highlighting the sensitivity of the problem. This suggests that fundamental issues remain to be understood in one or more of the constituent components used (i.e. any one or all of ALT/TG/GPS). Relating to this point and notwithstanding my remarks above regarding the relevance of this work, my sense from reading the manuscript is that the authors pay too little attention to discussing a few key issues that I argue should otherwise appear in a paper such as this:

1) Despite the technique itself originally emerging at the time investigators were using tide gauges (and thus VLM) to validate the stability of the altimeter measurement system, the authors do not seem to acknowledge the potential for regionally correlated error within the altimeter – brief mention of the importance of this point in the heritage (and current use) of the technique is also surprisingly lacking. Rather, the authors make the bold assumption of 'no instrumental drifts' (pg 3, ln 75) as the sole

mention of possible systematic error in the altimeter record. I note the authors exclude TOPEX/Poseidon from the coastally retracked dataset given waveforms are presently unavailable, however, it is included in the AVISO product. The well-known issues associated with TOPEX (brought to light using this technique and later confirmed) put
the issue into focus. Further, inspection of solely orbit differences in a regional context places even greater emphasis on the issue (e.g. Couhert et al, ASR (2015) who report on the significant challenge of achieving 1 mm/yr orbit stability at regional scales). The use of a consistent orbit across all missions in this current paper (which I assume are the GFZ Ver13 SLCCI orbits?), will assist to mitigate this effect but it is likely that mission-specific, spatially-correlated 'whole-of-system' drift will remain in the residuals and undoubtedly make at least a contribution to the differences observed in the current paper. I contend that 'no instrumental drifts' is somewhat of a misnomer in the context of whole-of-system altimetry at regional scales. Not mentioning these broad issues nor the heritage of the ALT-TG technique in this regard would be disappointing for a paper such as this.

We highly appreciate this comment, because it holds valuable and necessary information concerning fundamental underlying errors of the ALT-TG technique, which we had not covered so far in the pa-per. We make several adjustments in the text to integrate these issues:

__Systematic errors (regionally correlated errors, altimeter drifts):__

Introduction (we add short mentioning of systematic errors) in L75:

'The accuracy of this technique is, among numerous other factors, strongly dependent on the stability of the altimeter measurement system. Major systematic errors stem from limitations in the realization of the reference frame, as well as from limitations in the long-term stability of altimeter instruments and corrections (e.g. Couhert et al. (2015); Wöppelmann and Marcos (2016); Watson et al. (2015)). Altimetry records can be affected by regionally varying drifts, which were found to be most pronounced on TOPEX altimeter side-A. The calibration of altimetry with tide gauges is influenced by the applied VLM correction (Watson et al. (2015). Conversely, SAT-TG VLM estimates are affected by the altimeter drift.

Methods (we add more information on our cross-calibration analysis in L170):

'To reduce radial errors in the different missions, the tailored coastal altimetry product is cross-calibrated using the global multi-mission crossover analysis (MMXO) ~~global calibration~~ (Bosch and Savcenko, 2007; Bosch et al., 2014). The MMXO minimizes a large set of globally distributed single- and dual sea surface height crossover differences by least-squares adjustment. The estimated radial errors are used to correct each individual sea surface height measurement. In this way, we not only reduce orbit inconsistencies, but also those originating from the range and from applied corrections. Since we estimate a radial correction for each observation, we minimize intermission drift differences as well as regionally correlated errors. Note that this approach is a relative calibration and provides range bias corrections with respect to NASA/CNES reference missions. Any remaining absolute drift of these reference missions (with respect to TGs) still influence the drift of the whole altimeter solution.'

Results (L385):

'For both combinations, the ~~absolute~~ median of the VLM ~~bias of trend~~ differences (ALES-PSMSL-250km: ~~-0.51~~ -0.87 mm/year AVISO-PSMSL-250km: 0.56 mm/year) ~~exceeds~~ deviates from values shown in previous studies [WM16: -0.25 mm/year and, Kleinherenbrink et al. (2018): -0.06 mm/year]. In contrast to these previous estimates, we use different spatial selection scales of SLAs, smaller numbers of TG-GNSS pairs and deviating record lengths, which impedes a direct comparison.

Moreover, the altimetry datasets might be affected by instrumental drifts. In this respect, differences among the datasets may be caused not only by different techniques applied to reduce intermission biases (e.g., the MMXO approach for ALES), but also by different missions incorporated in the records. Note that in contrast to ALES, AVISO contains TOPEX, which has also been shown to be affected by a strong drift (Watson, 2015).'

Even though the ALES dataset does not contain any Topex data (in contrast to AVISO), this dataset can however be affected by erroneous drifts present in Topex, because this is a reference mission of the MMXO. In our dataset the cal1 correction has been subtracted from the original MGDR data. Thus, it is not applied to the TOPEX-A data.

Discussion (numerous sources of systematic errors) from L524:

'5.2 Systematic errors

VLM estimates from different datasets (e.g. AVISO-PSMSL-250km and ALES-GESLA-ZOI) are biased compared to trends inferred from GNSS observations. Based on Monte-Carlo simulations (see appendix Figure B1) we argue that these biases are significant for most of the dataset combinations (ALES-PSMSL-250km and AVISO-PSMSL-250km, ALES-GESLA-ZOI). In the following, potential sources for these biases will be discussed.

Next to the record-length (see section 5.1), systematic errors critically affect the accuracy of the SAT-TG technique and can have strong systematic effects on the trend differences. Limiting factors for VLM determination from both SAT-TG and GNSS observations are the accuracy and uncertainty of origin and scale of the reference frame (see WM16, Collilieux and Woppelmann (2009), Santamaría-Gómez et al. (2012)), which cannot be realized yet at the required accuracy level (Bloßfeld et al, 2018; Seitz et al., submitted).

Moreover, as mentioned before, the intermission calibration applied for ALES (MMXO) reduces intermission biases, but does not feature a calibration against TG. The median bias identified for ALES-GESLA-ZOI could be affected by a drift of the mission used as reference. In contrast, the AVISO dataset does not include time-dependent intermission biases and might therefore be additionally influenced by systematic effects of e.g. Envisat or Sentinel-3a (Dettmering and Schwatke, 2019). ...'

Bloßfeld M., Angermann D., Seitz M.: DGFI-TUM analysis and scale investigations of the latest Terrestrial Reference Frame realizations. In: (Eds.), International Association of Geodesy Symposia, 10.1007/1345_2018_47, 2018

Dettmering D., Schwatke C.: Multi-Mission Cross-Calibration of Satellite Altimeters - Systematic Differences between Sentinel-3A and Jason-3. International Association of Geodesy Symposia, 10.1007/1345_2019_58, 2019

Seitz M., M. Bloßfeld, D Angermann, M. Gerstl, F. Seitz: DTRF2014: The first secular ITRS realization considering non-tidal station loading. Journal of Geodesy, submitted.

Orbits:

Because we use ITRF2008 (as our GNSS solutions are based on this system), we don't use GFZ Ver13 SLCCI orbits that are in ITRF2014. For Jason3 and Saral we use the CNES GDR-E orbits and for ERS2 the GFZ VER11 orbits. The original orbits based on the GDR data are used for Envisat, Jason1 and Jason2. The reference system is consistent for all orbits (ITRF2008). Small differences due to smaller processing differences are accounted for in MMXO.

We specify the orbits further in lines (L156):

2) A second issue of less significance relates to the treatment of the tides in the high frequency tide gauge data. The authors attenuate the tide with a 40-hour loess filter (pg 8, ln 217). Clarification and further defence of the filter strategy would be beneficial here (e.g. both are likely small, but I wonder what % of *non*-tidal variance is attenuated by the filter, and what % of *tidal* variance at longer periods is retained (but removed from the altimeter with its model-based correction)). The comment regarding the temporal resolution of the DAC used for the high rate TG data felt ambiguous here (ln _219), and should be clarified with respect to consistency with the filtering approach. The magnitude of the difference in tides between (coastal) altimetry that uses a global model and at the tide gauge is an important issue that warrants at least a mention in the discussion.

Removal of TG tides: To remove the tidal signal from high frequency TGs, we followed the approach by Saraceno (2008), who used the same 40h-loess filter. We did not use a global tide model to correct for tides, because the quality of the modelled signals in shallow and coastal waters (i.e. near the Tide gauges) deteriorates. As an example, Piccioni (2018) showed that, regardless of the improvements of using dedicated coastal altimetry, the variance of residual tidal components (after subtracting tidal signal based on FES2014) can still remain in the order of several cm.

We add more information on the impact of our filtering approach on several tidal constituents in a qualitative way. For our considered GESLA TG dataset (consisting of 58 stations) the tidal filtering reduces the median standard deviation of the data by 27.5%. Note that such a reduction is strongly location-dependent. The following plots illustrate the reduction in tidal variance at different periods for selected TGs:

A)

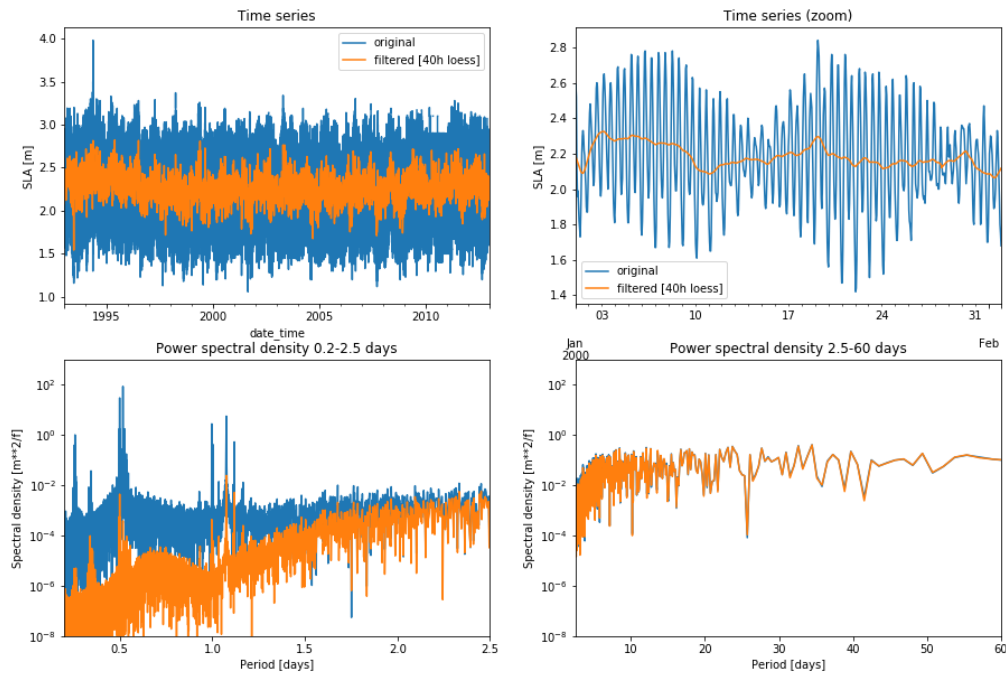TG: roscoff-roscoff-france-refmar, [48.7184, -3.9657]

B)
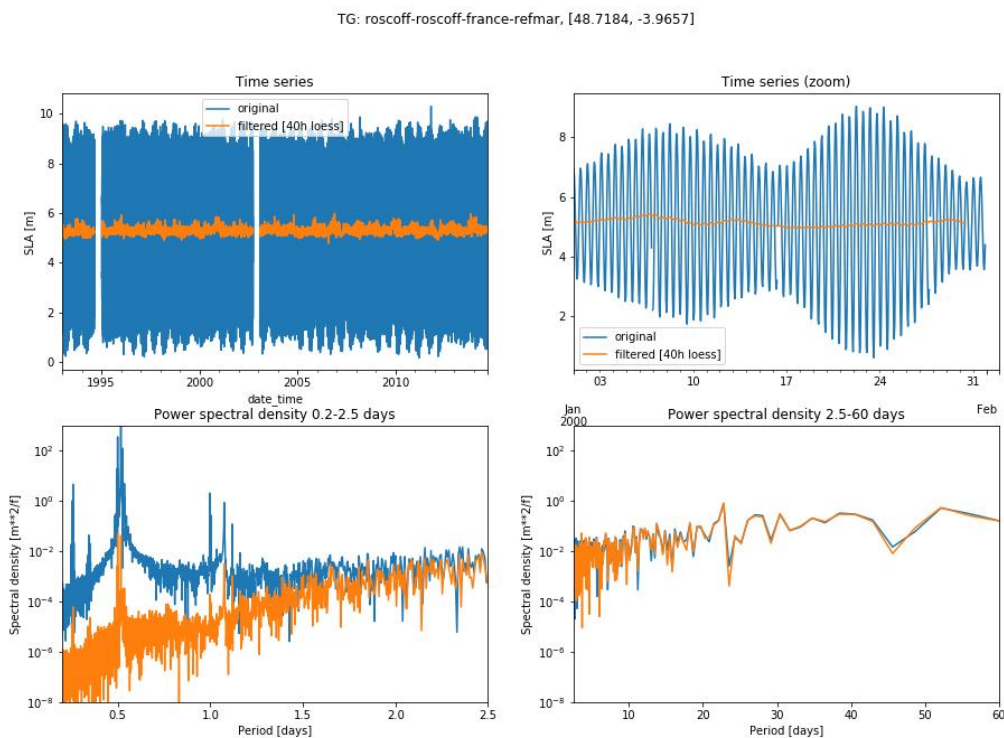
Figure R1: Time series (upper rows) and spectral densities (lower rows) for two different TG stations (A) France and B) Brazil). The original time series is shown in blue and the filtered one in orange. Spectral density estimates are obtained from using Welch's method and are based on time series with at least 2 years of data. Data gaps in the time series are filled by linear interpolation, with a maximum allowed data gap of 10 hours.

The plots show that for time-scales up to ~1 day spectral power is reduced by about two orders of magnitude. For time scales larger than 2.5 days, the filter does not reduce any of the longer periods. At the tide gauge in France (Figure R1 A) the long period tidal components Mm and Mf have amplitudes of 0.889 cm and 1.143 cm, respectively (based on the TICON dataset, Piccioni et al., 2019). At the second TG in Brazil (Figure R1 B) the amplitudes are 0.475 cm (Mm) and 1.288 cm (Mf). The plot below illustrates the amplitudes of these long period tidal components. At Mf frequency, tidal corrections applied to TGs or altimetry can deviate from each other with an amplitude up to ~2.5 cm.
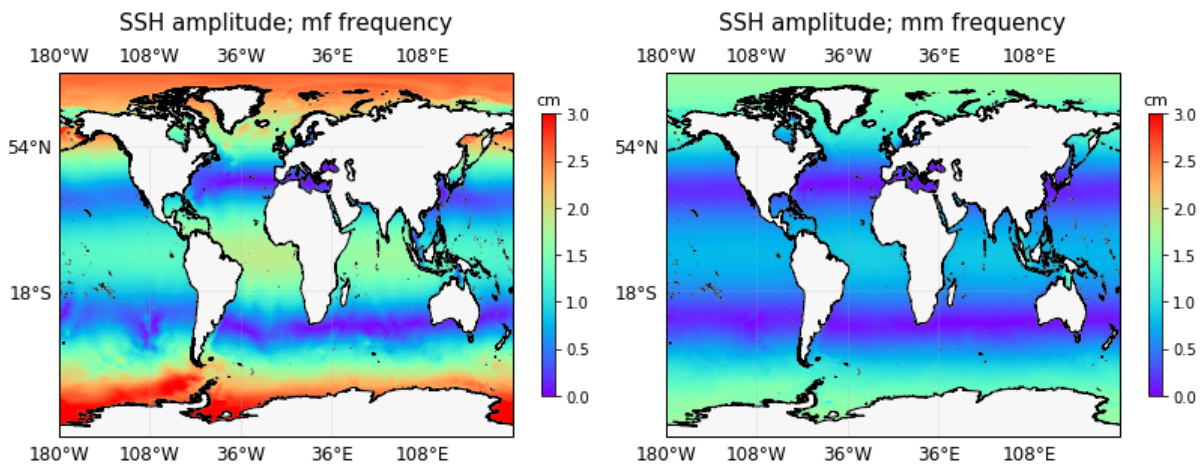

Figure R2: Modelled (FES2014) SSH amplitude of two fortnightly and monthly tidal constituents: Mf (left, period 13.66 days) and Mm (right, period 27.55 days)

Piccioni, G, Dettmering, D, Bosch, W, Seitz, F. TICON: TIdal CONstants based on GESLA sea-level records from globally located tide gauges. Geosci. Data J. 2019; 6: 97– 104.
https://doi.org/10.1002/gdj3.72

We are grateful to the reviewer for pointing this out and we reckon that we could further improve the comparability between TG and altimetry by accounting for longer period tides, for example performing a tidal harmonic analysis directly on each TG record. We keep this as an interesting development for our further studies and for now we clearly state in the text the caveat of our current methodology. Therefore, we added in L218:

'This filtering approach most effectively reduces tidal variance at periods lower than ~2 days (e.g. reduction by more than two orders of magnitudes at daily periods). However, tidal variability at periods larger than 2 days is not significantly attenuated by the filter. Therefore, one caveat of this approach is that there remains residual tidal variance at longer periods between TGs and altimetry, given that the latter features a model-based adjustment for longer tides. We do, however, not apply the same tidal model to the TGs, due to known issues related to decreased model performance in shallow water (Piccioni et al., 2018). '

Piccioni, G., Dettmering, D., Passaro, M., Schwatke, C., Bosch, W., and Seitz, F.: Coastal Improvements for Tide Models: The Impact of ALES Retracker, Remote Sensing, 10, 700, https://doi.org/10.3390/rs10050700, 2018.

Regarding the 6h-resolution of DAC (and interpolation on hourly TG data): As in Saraceno et al. (2008) we also use a 6h-resolution DAC correction and apply it after filtering the TG data. As the best compromise, we decided to interpolate (cubic) this correction onto the hourly time step of the TGs (L220). Overall, this approach ensures consistency between the DAC correction for TGs and altimetry, which requires the same temporal resolution as subsequently TG and SAT time series and therefore, the DAC correction are subtracted from each other.

3) Finally, the improvement observed between along-track altimetry v gridded altimetry is presented in the context of 'coastal altimetry' driving the improvement. I'm very pleased to see these results however I feel that conclusion needs some further defence. From my understanding of Fig 5d, the lower quartile of the data commences at just under 30 km from the coast (and increases from there). Discussion of Fig 5b in the text (pg 22, ln 470) refers to a much smaller subset in which the mean distance was 33 km. Given the benefit of retracking is 'most pronounced in the last 20 km' the authors should further evidence what % of their data is within this threshold and thus further elucidate whether it is simply the influence of along-track data versus its retracked equivalent. To be clear, I am not asking for the former to investigated in the same way – the benefits of the retracking in general terms are clear – I just wish to ensure the conclusions are appropriately elucidated.

As the reviewer noted, the mean distance of 33 km referred to a smaller sub-set comprising all ZOIs within the 0.8 – 0.975 levels. To clarify how much of the data is near the coast at the global optimum, we now refer to the 80th level (based on the correlation metric.)

We adjust the text as follows:

At the global optimum itself (0.8, based on correlations), the median distance to coast (of all SLA measurements in a ZOI) is 39.4 km. 25 % of the altimeter observations are within a range of 20 km to the coast, i.e. the region with the most pronounced coastal advancements of the along-track dataset (Passaro et al., 2015). ~~*At these high thresholds, the mean distance to coast averaged over all SLA measurements in a ZOI is 33 km. Hence, a large proportion of these SLAs heavily profits from demonstrated coastal advancements of the along-track dataset, which are most pronounced in the last 20 km off the coast (Passaro et al., 2015).*~~

Technical Issues:

It is unclear throughout if the RMS statistics computed against the GPS time series take into account the GPS uncertainties? I expected to see WRMS mentioned rather than RMS.

The RMS statistics as used in the study is the unweighted RMS. To account for GPS and SAT-TG formal uncertainties we added the weighted RMS in table 2 (pg 17.) and adjusted paragraph 3.4. Comparing the weighted and un-weighted RMS, we find that not in all cases accounting for formal uncertainties is associated with better accuracy.

Because the weighted RMS supports the use of ALES-GESLA over ALES-PSMSL as well, we add (L402):
' … Although trend uncertainties are higher for the ALES-GESLA-250km configuration, we choose this set-up to investigate the impact of the ZOI. This dataset not only provides better results concerning trend accuracy (weighted or unweighted RMS) and has a lower median bias. …'

In section 2.1, I assume the GFZ Ver 13 SLCCI orbits are used homogeneously for all missions, but this is only inferred. I suggest adding detail to Table 1 for clarity and citing the relevant Rudenko et al paper.

Please refer to the answer to issue #1 (more details on the applied orbits)

I was pleased to see the data has been cross calibrated using the MMXO approach (pg 6, ln 165). Some additional detail here would be beneficial: I assume this provides a location-specific and mission-specific radial bias, computed using the *identical* dataset (Table 1). Some further detail on this would be useful here. Further, some comment on the relative trends observed (and applied?) between the Jason-series and ERS/Envisat missions (using the MMXO approach) is warranted here.

As the reviewer noted, the approach uses a very large set of location-specific, single- and dual-satellite crossover differences to reduce the radial errors of the missions. Therefore, the technique enables us to minimize the geographically correlated mean errors. This approach is applied to the *identical* dataset (based on all corrections and adjustments as listed in Table 1).

Please refer to the additional information given as a response to the first comment or changes from L170.

To illustrate the magnitude of the MMXO correction, we show the estimated relative range bias (10 day means) for Envisat V3 w.r.t. Jason-1 (Fig. R3) indicating the temporal systematics which are accounted for by the MMXO. Fig. R4 shows the geographically correlated mean errors (offset subtracted) between Envisat and Jason-1. Both effects are removed from the Envisat data by applying the estimated radial corrections for each measurement.
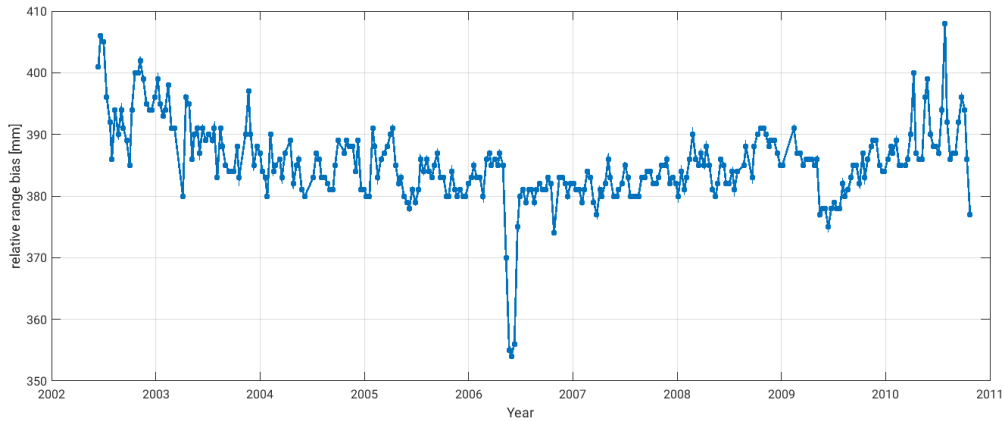
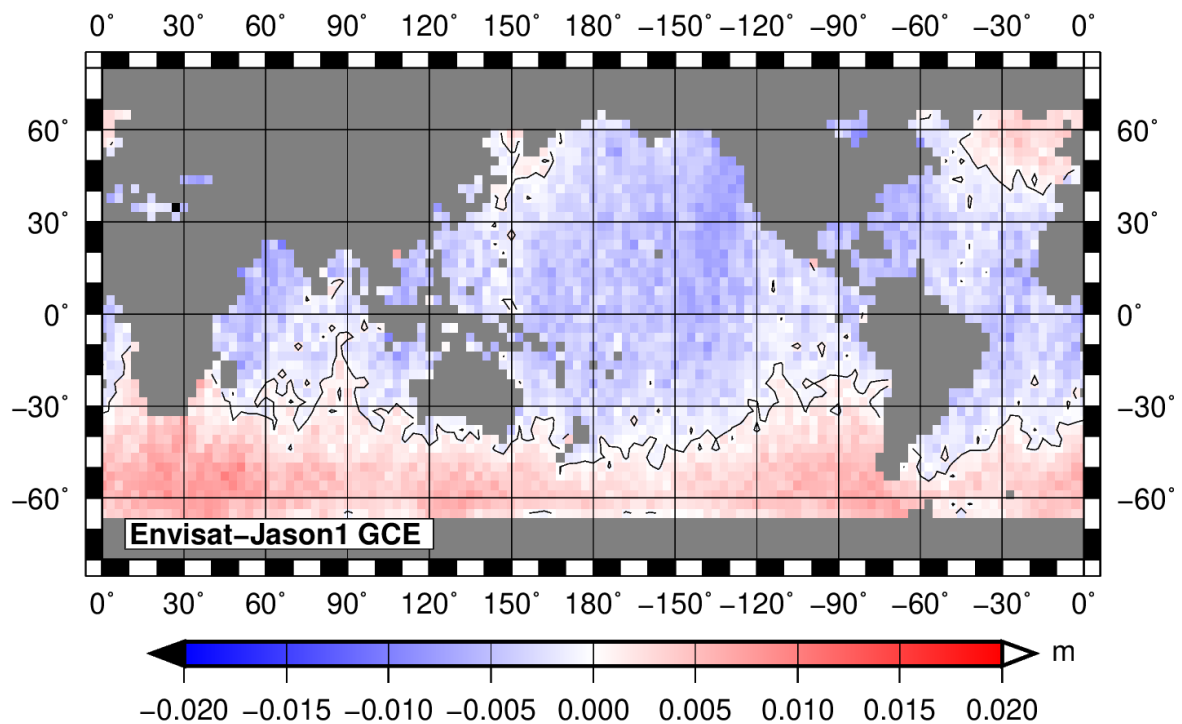*Figure R3: Envisat relative range bias per 10-day cycle (w.r.t. Jason-1) [mm]*

Figure R4: Geographically correlated mean errors of Envisat w.r.t. Jason-1 [mm]

Regarding the pole tide in Table 1, could the authors confirm they only apply the solid Earth component of the pole tide (noting the pole tide that is present in the GDR combines the ocean plus solid earth component, hence needs to be multiplied by h2/(1+k2).

We are using our own software to correct for solid earth and pole tides following the IERS conventions (2010) instead of relying to the corrections provided in GDR. Thus, we can confirm that point.

Regarding the ocean tide in Table 1, it might be worth adding a paragraph to further discuss the long-period tides that are/are not considered in the altimeter data. The current treatment of the tide in the high frequency dataset at least should retain any long period constituents – confirmation that is the case with the altimetry would also be beneficial.

Please refer to the answer of the second comment.

Regarding the cubic interpolation to the hourly tide gauge data, I'm assuming you don't use this across outages larger than some threshold? Worth clarifying.
In Figure 2, this may be more beneficial to show the data in panel 3 over the data in panel 2. Then, in the now vacant third panel, show the same but with the annual and semi-annual terms estimated and removed. Note the caption reports "SAT minus TG" but the y-axes report "TG-ALT".

Concerning the combination of hourly TG data and altimetry data, we only match values (or interpolate values) with data gaps of maximal 3 hours. This is added in the text in L235: '' … by cubic interpolation of the latter and a maximum allowed time-lag of 3 hours between the measurements.'

Figure 2: We changed the Figure and its caption accordingly and corrected the axes descriptions.

In Figure 4, I found Fig4b and d difficult to interpret. I feel they need to be introduced earlier and/or more comprehensively explained as I felt it took me a few reads to get the point.

We assume that the reviewer refers to Figure 5 (Given that there would not be differences between Figure 4a-d). We added more explanations to better introduce the figure and to better understand the purpose of why we performed this analysis. We also clarified our interpretations of Figure 5b:

We changed this paragraph (in which Figure 5 is explained) (from line 460-478):

'The results presented in Figure 5 and Table 2 denote ~~average~~ metrics and performances derived from the global TG-GNSS dataset for ALES-GESLA-ZOI ~~performances for the globally distributed TG-GNSS station pairs for ALES-GESLA-ZOI~~ and support an optimal threshold at 20%. It remains to be investigated, ~~It is however unclear,~~ whether the described optimum global threshold ~~for this 'global' selection~~ also reflects the best choice at every ~~considered~~ coastal site considered. Therefore, we investigate at which relative levels individual VLM SAT-TG and VLM GNSS trends estimates yield the smallest absolute deviations. Postulating that the actual VLM at the TG location is linear and perfectly detected by the GNSS station, these thresholds denote the 'local' optimal levels. With this analysis, we aim to better understand the spread of individual optimal ZOIs and what would be the best theoretically achievable RMS_ΔVLM. This analysis also provides a basis to motivate future investigations, in particular to identify systematic factors, which may lead to local different extents of the ZOI and to improve the accuracy of trend estimates.

Figure 5b displays the distribution of local optimal thresholds for TG-GNSS stations for the ALES-GESLA-ZOI dataset. Note that these estimates are not independent as they are based on prior knowledge of the ground truth VLM from GNSS. … An associated ideal selection of trends, based on optimal individual levels shown in Figure 5d would largely reduce to RMS_VLM to ~~0.9~~ 0.89 mm/year. We emphasize that this constitutes the best RMS, which could theoretically be achieved with our dataset combination, if all of the local optimal levels could be systematically explained. This demonstrates that, albeit there might be room for minor improvements, there is still a strong limitation remaining in bringing the RMS below 1 mm/year.

Minor Issues:

The word 'frequent' is used throughout in the context: 'high frequent data'. I suggest this should be 'high frequency'.

Corrected

The terms accuracies and uncertainties are used throughout when singular is arguably more appropriate. E.g. ln 11 in the abstract.

We changed the terms to singular, when it is more appropriate.

Suggest a search and replace for units – some (e.g. km) do not have a space beforehand after the quantity.

Corrected, Note that for dataset names, we intentionally did not use space e.g. ALES-PSMSL-250km

L15 p1. Perhaps increasing instead of progressing? Perhaps : : :the spatial scales of the coherency in coastal sea level trends increase.

Both changed

L20, p1. : : :as the sea level rise signal itself: : :

Added

L21, p1. : : :can regionally account for a large fraction of the: : :

Corrected

L35, p2. I had expected TG instability to at least be mentioned here. I know it comes later, but consider a mention here.

In the paragraph the reviewer refers to, we have listed the possible causes of VLM and not the issues related to the kind of measurements. Therefore, we assume that with this comment the reviewer is referring to non-linear short-time-scale vertical movements which affect the TG records and would be seen as instabilities when trying to estimate a linear relative sea level trend, just as when a record changes its reference datum. Therefore we added this sentence in L36: *'These and other non-linear processes at very short time scales are particularly challenging in the estimation of a linear rate of long-term VLM from TGs, since they would appear as instabilities in the record (similarly to a change in datum).'*

L40, p2. "global" trends range 1 to 3 mm/yr. Do you really mean "global"? If you do then I feel this could be made more specific to a time period of interest as it feels too vague.

We changed L40:

'Given that global absolute sea level trends *during the altimeter era is in the order of 3 mm/year (3.1+-0.1 mm/year from 1995 to 2018 as reported in Cazenave et al., 2018)* *one prerequisite for VLM estimation is that associated trend uncertainties should be at least one order of magnitude lower than those subtle signals (Wöppelmann and Marcos, 2016).'*

Cazenave, A., Palanisamy, H., and Ablain, M.: Contemporary sea level changes from satellite altimetry: What have we learned? What are the new challenges?, Advances in Space Research, 62, 1639 – 1653, https://doi.org/https://doi.org/10.1016/j.asr.2018.07.017, http://www.sciencedirect.com/science/article/pii/S0273117718305799, 2018.

L54, pg2. : : :solutions against those using measurements: : :

Added

L59, pg3. "considerable" isn't the right word here, consider change.

Changed ~~considerable~~ to *comparably*

P3 – See main issue #1.

Please refer to answers to #1

L80, p3. "synergistic applications" isn't quite right, consider change.

We changed ~~synergistic application~~ to 'their combined use'.

L87, p3. : : :further develop: : : instead of "carry on" the latest progress? Suggest characterisation and quantification instead of detection?

Changed carry on to 'further develop'
Changed detection to 'characterization and quantification'

L96, p4. "were achieved" c.f. what?

We changed this line to
'Using this dataset, WM16 also obtained the most precise VLM estimates … '

L108, p4. suggest: : : :however, they reported insignificant improvements of: : :

Adopted

L126, p4. no comma needed after processes.

deleted

P6 – See main issues 2 and 3.

Please refer to answers to these main issues #2 and #3

L190, p7. The PSMSL constitutes: : :

Corrected

L191, p7. : : :for most sea-level research.

Corrected

L222, p8. low latitude maybe ambiguous for some readers. Do you mean nearer to the equator or poles? I note you use high-latitude later.

We changed 'in low-latitude regions' *to* 'nearer to the equator'; *Later we added* '… higher latitudes (i.e. towards the poles) …'

L240, p8. I'm unsure that "matching" is the correct word. "Differences formed using" perhaps?

Changed 'matching' *to* 'differences formed using …'

L252, p9. : : :the capability to compare: : : no comma after altimetry.

Corrected

L264, p9. duplicate as

deleted

L310, p12. performance not performant

Corrected

L330, p14. "next to SLA correction" could be significantly improved.

We rephrased these line as follows:
' ~~Such sources include next to SLA correction or adjustment errors most predominantly sea level dynamics, which do not reflect the tide gauge observations.~~ *The SLA computation is affected by the instrumental errors of the range estimation and of each of the geophysical corrections (Ablain et al 2009). Such errors, as well as the measurement error of the TG itself, show up as residuals in the differenced time series. Moreover, sea level dynamics that do not reflect the variability observed at the tide gauge locations also contribute to the SAT-TG differences.* Therefore, … '

M. Ablain, A. Cazenave, G. Valladeau, S. Guinehut. A new assessment of the error budget of global mean sea level rate estimated by satellite altimetry over 1993-2008. Ocean Science, European Geosciences Union, 2009, 5 (2), pp.193-201.

L_340, p14. Some comment from the authors regarding whether they consider PL+WN more appropriate than GGM for example would be useful.

We added the following lines to the end of this paragraph L340:
'*We emphasize that for individual regions other noise models could be more appropriate than the implemented PL+WN model and would thus yield more realistic uncertainty estimates. An advanced regional spectral analysis to identify the most suitable models is however beyond the scope of this study.*'

L349, p14. needs "ellipsoid" inserted after TOPEX/Poseidon.

Inserted

Figure 3: The distribution is unavoidably clustered. Does residual VLM show any spatial patterns?

We observe a slight tendency of higher GNSS than SAT-TG trends in Northern Europe and North America as well as lower GNSS trends toward equatorial regions and Australia. We add the associated trends and geographical distributions in Figure R5 below. Note that the causes of the spatial pattern of trend differences in Figure R5 requires further investigations.
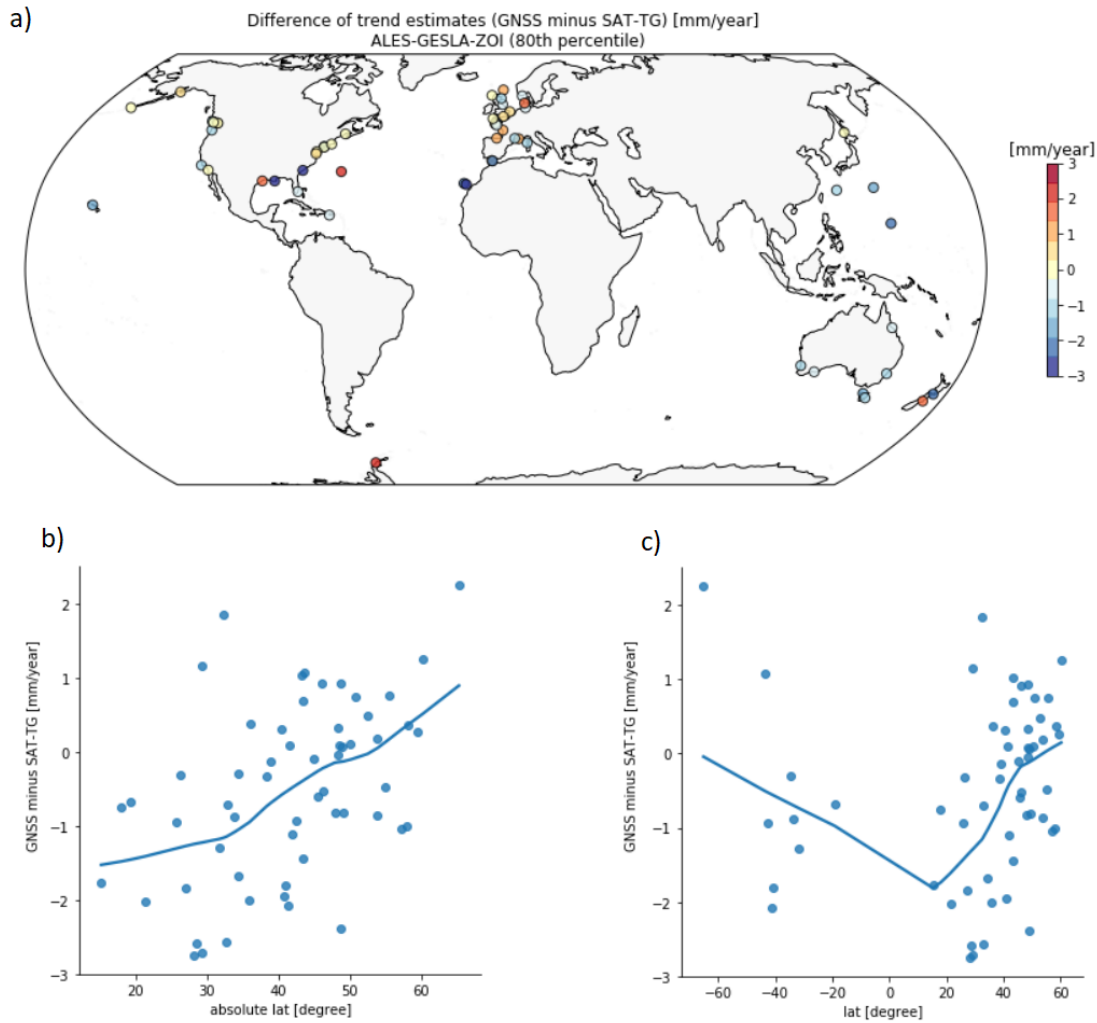

Figure R5: Trend differences of GNSS minus SAT-TG VLM [mm/year] for ALES-GESLA-ZOI (80th percentile) (a) Map of trend differences; (b and c) Trend differences as a function of absolute latitude and latitude. Solid lines represent locally weighted regressions (Cleveland et al., 1988).

We also discuss the possible (and neglected) influence of contemporary mass distribution on GNSS and SAT-TG trend differences in the discussion.

Second part of added discussion paragraph from L524:

'… Next to altimeter bias drift, nonlinear VLM from contemporary mass redistribution (CMR) changes were shown to cause differences between VLM_SAT-TG and VLM_GPS, due to the different time periods covered (e.g. Kleinherenbrink et al. (2018)). Using GRACE (Gravity Recovery and Climate Experiment) observations Frederikse et al. (2019) demonstrated that associated deformations can cause VLM trends in the order of 1 mm/year. Therefore, they introduced a new method to reduce VLM_GPS by GIA and CMR signals to minimize their associated induced extrapolation biases. Kleinherenbrink et al. (2018) incorporated nonlinear VLM from CMR to assess the corresponding trend

differences between VLM_SAT-TG and VLM_GPS. They exposed that VLM_SAT-TG estimates are lower than VLM_GPS in many parts of North America and Europe and higher in subtropical/tropical regions as well as Australia and New Zealand (refer to Figure 9 in Kleinherenbrink et al. (2018)). Because northerly regions, for instance, are affected by stronger recent uplift, GNSS observations which cover shorter and more recent time spans than satellite altimetry detect more positive trends. For a set of 155 TG-GNSS pairs, integration of these signals slightly reduced the median bias from -0.14 mm/year to -0.07 mm/year, but had no significant effect on RMS. Given that most of the TG-GNSS stations used in this study are located in Europe, North America and Australia, CMR might as well alleviate the negative trend bias of ALES-GESLA-ZOI. Therefore, extending the validation platform, not only by using other homogeneous GNSS observations, but also GRACE and GIA estimates would strongly support identification and mitigation of such systematic errors.'

L370, p15.is "also" needed?

L370: Deleted ~~also~~

L399 p16. Suggest improve "which stronger deviate".

Changed '… which stronger deviate … ' to '… with stronger deviations from the TG records …'

L402, p16. This is the opposite sign however: : :

Due to the change of Envisat data and thus a change in the bias of trend differences (of ALES-GESLA-250km) we deleted this line.

L435 p18. assuming issues with e.g. tide model errors, the method presented will likely have a optimum threshold that still enables sufficient spatial averaging to mitigate that systematic effect as much as possible. See main issue 2.

We agree with the reviewer that the benefit of our method is dependent on thresholds and that these can be optimized either globally (as we do in the paper) or locally (which is a challenge to which currently we can only answer "a posteriori" using the comparison with the "truth" at the GNSS stations, as shown in Figure 5). In the text we highlight the influence of the sampling size on the global optimal threshold, because at very high levels the sample size decreases. We agree that there are many various factors which additionally contribute to these optimal levels (such as tidal model errors). For the discussion concerning the main issue 2, the reviewer is referred above to the corresponding answer.

We add *from L429*:

'RMS_VLM and trend uncertainties level off at very high thresholds and ultimately increase when only 5% of the data is used (Figure 5a and 5c). We argue that this is mainly related to a decrease in sampling-density of the time series included in the selection: At the 95th percentile, the median sample size (i.e. number of monthly averages in a time series) is 20% smaller that the sample size at the 80th percentile. Robust trend estimates require a minimum of samples, hence, using a reduced number of along-track data time series, even when they show a maximum degree of comparability, yields on a global average decreased trend accuracies (RMS_VLM). ~~Indeed, one would expect the highest comparable (for instance expressed by correlations or RMS) or even the closest altimetry measurement point to result in most accurate VLMSAT-TG trends. This is, however, not the case for this SAT and TG combination on a global average.~~ We thus argue that the optimum threshold identified at about the 80th percentile (of the data sorted by RMS) represents a compromise between data-comparability, as well as sampling-density of altimetry data. We emphasize that there are numerous factors, other than the time period covered, which may contribute to a lack of comparability of SAT-TG and GNSS trends. We further elaborate those in the subsequent discussion section 5.'

L439, p18. Is bisector the correct term? 1:1 line?

Changed to *1:1 line*

L441, p18. Unclear what you mean by outstrips improvements induced by: : :. Clarify.

We changed the sentence:

This result proves the importance of using such a refined selection procedure (ZOI), as this approach outstrips the improvements induced by the different altimeter or tide gauge data combinations.

To

'These results underpin that a refined selection procedure (ZOI) represents the dominant advancement, as this approach outstrips the improvements (in terms of trend accuracy and uncertainty) which are obtained from using different altimeter or tide gauge data combinations.'

Figure 4 – Comment error bars are 1 sigma?

The error-bars are 1 sigma, we added that to the figure caption.

L447, p20. The residual annual cycle criterion: : :

corrected

L450, p20. I tend to agree with the statement, but what about the bias in the accuracy as arguably the low frequency component may lead to larger biases in deltaVLM (as well as scatter). A comment maybe useful here.
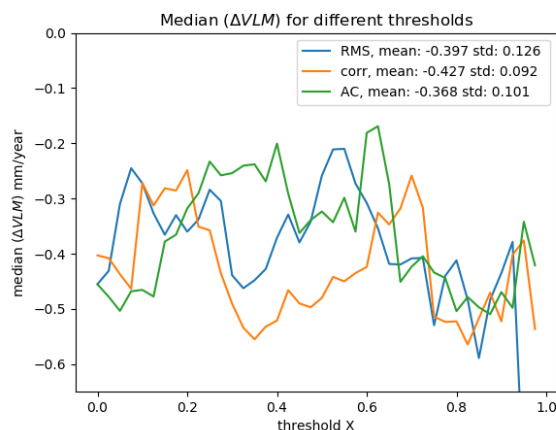

Figure R6: Shown is the evolution of the bias (median of trend differences) depending on different criteria and thresholds for ALES-GESLA-ZOI (as in Figure 5)

We give some additional information on the evolution of the trend bias in Figure R6 above. We do not find a similar convergence of the biases to a global optimum as for accuracy or uncertainty. The values exhibit fluctuations in the order of 0.1 mm/year. There is also no significant difference between using different criteria. Thus, we add to L450:

'Considering improvements in the bias of trend differences, we find no significant differences in using different thresholds. In contrast to the improvements in accuracy (as shown in Figure 5), the median

ΔVLM does not converge to a global optimum. Therefore, we discuss the contribution of other factors affecting the comparability of SAT and GNSS in section 5.2'

Figure 5. In the legend, unclear what AM is. I assume something to do with Annual, but needs clarification in the caption.

We changed AM to AC, and added a description in the caption

L455, p22. estimates do not always

Added

L458, p22. See main issue #1.

As given in the response to main issue #1 we added a section discussing systematic errors from L524.

L460, p22. Is average the best word here. I understand your point, but statistically speaking? Perhaps denote metrics of performances derived from the global dataset?

We changed 'The results presented in Figure 5 and Table 2 denote average performances for the globally distributed TG-GNSS station pairs for ALES-GESLA-ZOI and support …' to

'The results presented in Figure 5 and Table 2 denote metrics and performances derived from the global TG-GNSS dataset for ALES-GESLA-ZOI …'

L463, p22. at every coastal site considered.

Changed

L469, p22. "large" in inadequately defended. See issue #3.

Changed. See previous response to #3

Section 5.1 was challenging to read, Fig 4b and d were difficult to interpret. I suggest some further elaboration.

Please refer to the changes made in response to a previous comment (improvements of section 5.1.)

L494, p23, no comma needed after we note.

deleted

L503, p24. I understand your point here but I feel more emphasis should be placed on the sensitivity of the ALT and TG to the same phenomena. If both were sensing 100% the same signal, the time span is not as critical. Again, I come back to main issue #1 here.

In this chapter, we investigated how the accuracy of the estimates changed with different time series lengths. We found a dependence of an optimal spatial selection of SLAs on time scale. This dependency indicated that with increasing time scales also the spatial scales, which would yield the best accuracy of the estimates, would increase. We argued that with increasing record lengths SL-trends would emerge from the shorter term variability and also be similar to TG trends at greater distance.

As far as we understand, the reviewer points to the possible influence of systematic drift biases in causing overall biases when comparing ALT-TG and GNSS trend estimates. We agree that such an effect would be less time scale dependent. The finding that accuracy does not improve as strongly as uncertainty could also indicate limitations set by systematic errors such as altimeter drift biases.

In our understanding, in the context of the analysis presented in the chapter, a drift bias would cause a constant offset in the trend residuals and would thus not influence the time-scale dependencies found in this analysis.

However, given that the bias itself reduces the accuracy of our estimates by about 10% (when simply comparing the STD of the trend differences against the RMS), we further elaborated associated issues at several positions in the text. (refer to previous responses, e.g. L385 and L524)

L535, p26. Quantify "much of" and "close vicinity to the coast". See main issue 3.

Quantified, see response to #3

L538, p26. : : :not be unequivocally

corrected

L560, p26. Are these statistics medians?

Yes, we added 'median formal uncertainties …'

L565, p26. : : :and dedicated coastal altimetry: : : confine ZOIs and increase: : : : : :the global coastline which improved uncertainty.

Corrected (we assume it should be 'with improved uncertainty')