Review response

To Alvaro Santamaría-Gómez, August, 26 2020

We thank Alvaro Santamaría-Gómez for his very constructive and positive comments, which are extremely valuable for improving the manuscript. In response to the comments, we performed additional computations and integrated the associated results. In order to refine some of our messages and to make the manuscript clearer, we reformulated explanations or interpretations of the data at several positions in the text.

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 paper addresses the methodology of estimating vertical land motion (VLM) from the combination of satellite altimetry (SAT) and tide gauge (TG) observations. The work by J. Oelsmann, M. Passaro et al. builds upon earlier studies concerning the selection of the most suitable SAT observations that show high temporal correlation with high-frequency TG observations. In addition, and contrary to past studies, they use dedicated coastal "retracked" along-track observations to reduce the VLM differences with respect to co-located GNSS VLM estimates, which are taken as ground truth.
My feeling is that this paper is a significant technical contribution to the estimation of coastal VLM from altimeters and tide gauges, which is a relevant topic for the journal. I generally agree with the authors that advances in this field call for better consistency of the sea-level observations from tide gauges and altimeters. To reach this goal the authors focus on using altimeter observations closer to the coast and high-frequency tide gauge data. The authors show that areas of high consistency between both datasets, which they call "zone of influence" or ZOI, can be defined based on different statistical criteria. The comparison of SAT-TG VLM estimates against GNSS VLM estimates is improved, but the typical differences between both VLM estimates are still much larger than their respective formal errors. This indicates that there are still missing pieces to be accounted for in the VLM estimates from SAT-TG or GNSS or, likely, in both.
Below a few minor comments that hopefully will improve the quality of the paper:

Abstract: ZOI should be defined in the abstract, if space allows it.

We changed L7-8 to:
'To improve the coupling-procedure, a so-called 'Zone of Influence' (ZOI) is defined, which confines ~~to identify~~ coherent zones of sea level variability on the basis of relative levels of comparability between tide gauge and altimetry observations.'

L24&59: Many thanks for citing my 2017 paper, but there is no need to add it twice to the reference list. 2017a/b should be 2017.

Fixed!

L62: change accuracy by precision

Changed!

L271-273: *(comment separated by author) part1:* **"To confine the ZOI, we select subsets of the data containing the best performing statistics (i.e. highest correlation, lowest RMS SAT-TG or residual annual cycle) above the Xth-percentile according to the distribution of the statistic S in a 300 km radius around the tide gauges." » It is not clear whether the TG and SAT series were detrended and de-seasoned before comparing them. If the TG and SAT series were not de-seasoned and the seasonal variation is prominent in the series, then there is the risk that the three metrics are telling us almost the same thing, that is, the impact of the amplitude and phase differences between the seasonal signals in both series. The correlation and the RMS of the differences may be more representative from de-seasoned series, as done in past studies.**

We only detrended the data (and removed an offset) before matching. We agree with the reviewer that without de-seasoning the independency of the metrics is reduced as they are all also influenced by the consistency of annual cycle signal of both time series. We did not de-season the data in the first place, as we use high-frequency data and assume that the spatial coherency (and thus the extent of the ZOI) would still be dominated by the similarity of high-frequency processes.

Accordingly, we assumed that the influence of the annual cycle on the relative distribution (i.e. the relative change of a metric in space around a TG) on RMS and correlation would be minor. However, we did not quantify the contribution of the annual cycle to those metrics nor evaluated our assumption. Hence, in response to the reviewer we repeated the analyses by first de-trending and de-seasoning the data (before computation of the statistics.)

In the following plot we compare the impact of de-seasoning on the metrics RMS, correlation and residual annual cycle. In this plot, we also use weighted RMS or standard-deviations of the trend differences (SAT-TG minus GPS) as suggested in a subsequent comment:

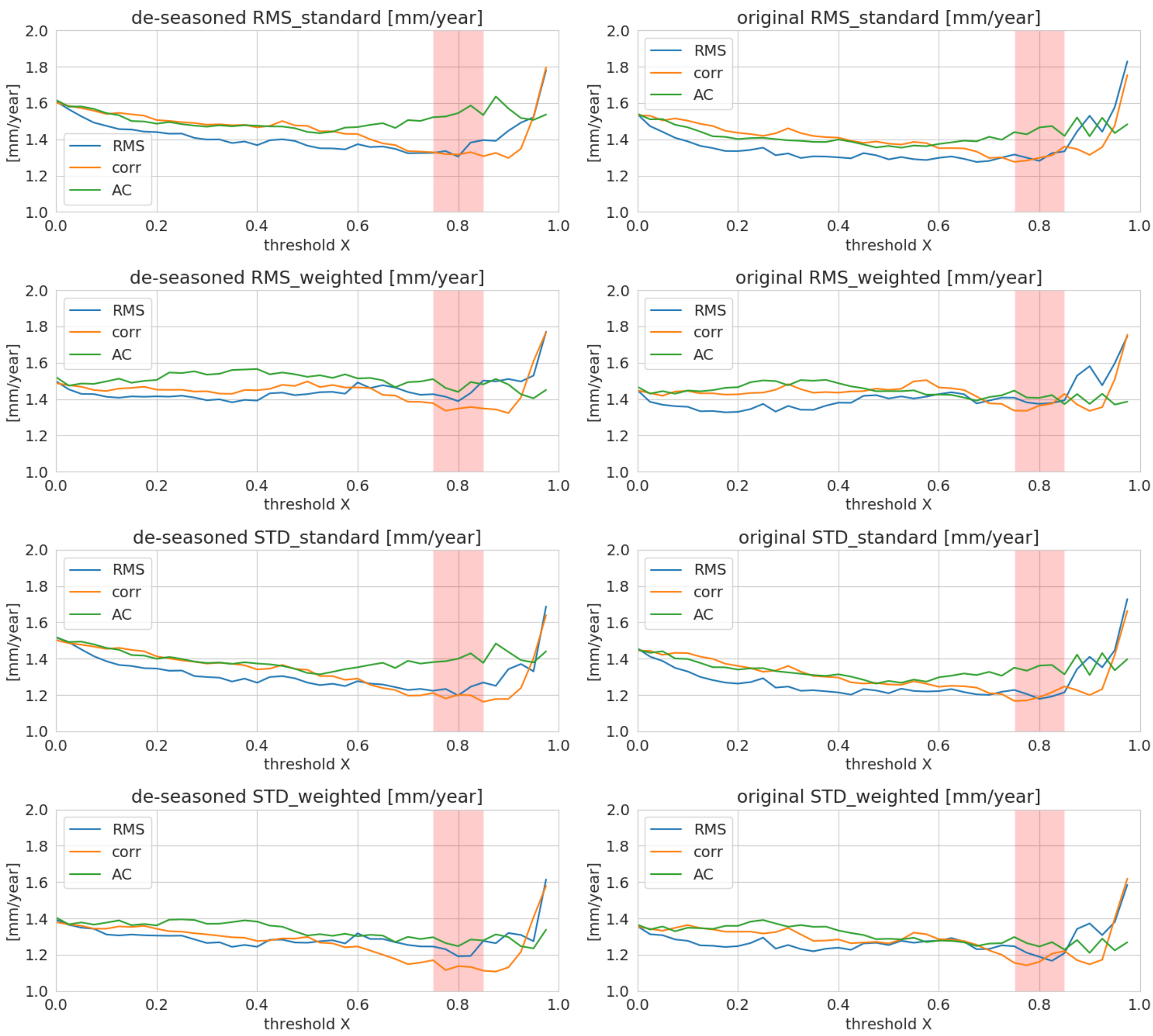
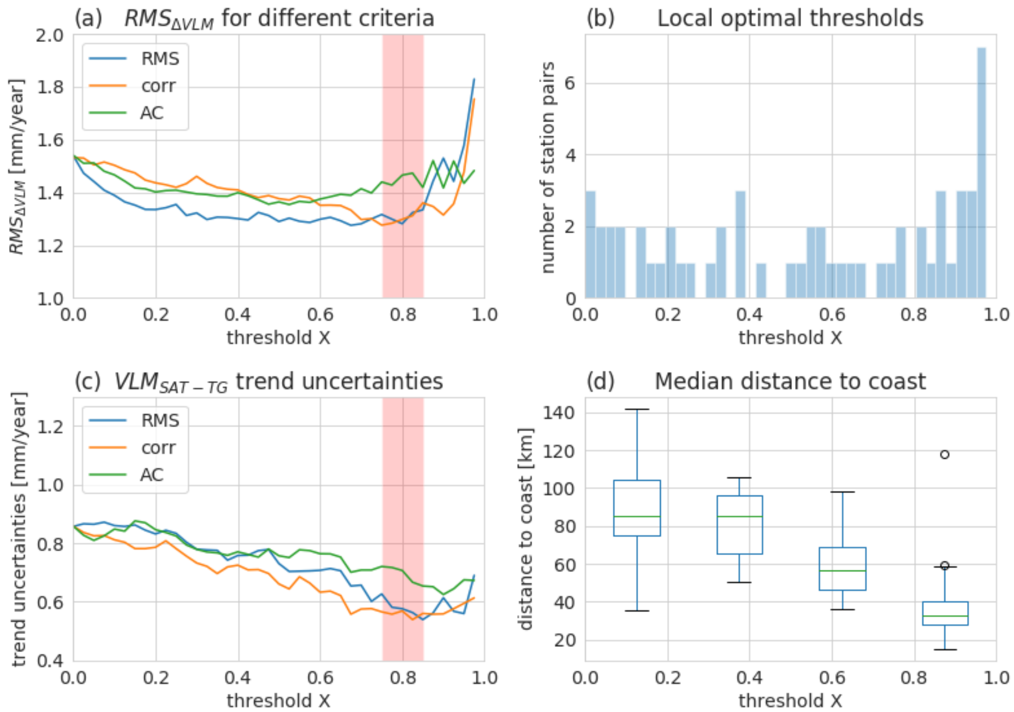


Figure R1: Comparison of statistics when the data was de-seasoned (left) or not (right) before matching. First row shows the unweighted RMS, the second row shows the weighted RMS of trend differences, the third row shows the un-weighted standard deviation and the last column shows the weighted standard deviation. The red bar marks the level range, which had been identified as the global optimum based on the un-weighted RMS.

In the original version (detrended and not de-seasoned) the metrics RMS and correlation were shown to provide very similar results (in terms of accuracy and uncertainties). When we de-season the data (prior to computation of the metrics), we do not find significant improvements (in accuracy) or effects on the use of correlations or RMS, as we also concluded in the study. To assess the influence of de-seasoning on uncertainties, we reconstructed Figure 4, now with results based on metrics derived from detrended and de-seasoned time series:

A. Original results



B. Results after de-seasoning and detrending

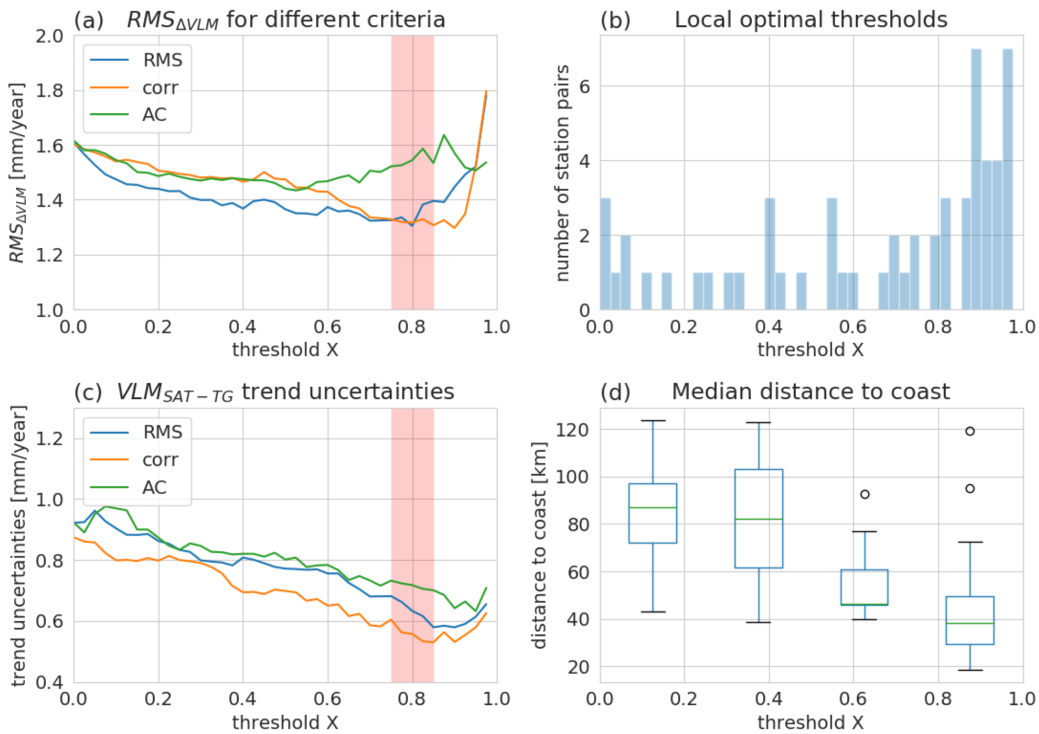


Figure R2: Shown are A) statistics when data was only detrended but not de-seasoned; B) statistics based on detrended and de-seasoned data. Figure captions are as in Figure 5: Performance of VLMSAT-TG trend estimates for ALES-GESLA-ZOI. a) RMSΔVLM for different relative thresholds (step size 2.5%) and different selection criteria: RMS SAT-TG (blue), correlation (red) and residual annual cycle (AC, green); c) same as (a) but for median uncertainties. b) Distribution of best performing relative thresholds for individual stations. The local optimal threshold is defined at the minimum of the absolute difference of VLMSAT-TG and GNSS trends. d) Boxplot shows the distribution of the mean distances to coast for the individual optimum ZOI's as denoted in b). The distances refer to the distributions within the 0-25%, 25-50%, etc. levels, respectively.

Hence, in our application de-seasoning of the data does not significantly alter the choice of the statistics (on which the ZOI is based on). To better justify our choice, we added the following lines in the manuscript L270:

'The statistics are based on de-trended data. Thus, all the metrics may be influenced by the similarity of the annual cycle. However, by repeating this analysis using de-trended and de-seasoned data (not shown), no significant differences were identified.'

L271-273: part2: In addition the authors fit the seasonal variation together with the linear trend in the SAT-TG series, i.e., the residual seasonal variation may play a minor role in the estimated VLM and its uncertainty. I'm not sure if this is what the authors intended. It may not have a significant impact on the selected ZOI areas, but the authors would be at least using more independent criteria.

To assess the trend components of the SAT-TG VLM time series we followed the standard approach to estimate also the seasonality as done in previous studies (e.g. Wöppelmann and Marcos, 2016). In case that the residual annual cycle is caused by differences in observed SLA variability, we still parameterize such variations, since we are interested in the long-term changes and the uncertainties arising from long-term variations. The annual cycle signal, which we assume to be constant over time, should not contribute to uncertainties associated with changes on times scales longer than one year.. Another factor, which however only has a minor impact, is that at some locations the SAT-TG series might even contain annual signals that exist due to actual seasonality of the local VLM, but not due to the residual annual cycle between the SAT-TG measurements. Such signal should then be modelled; otherwise, it would increase the trend uncertainties and it would be inconsistent with the GNSS trend estimates (where seasonality was also taken into account).

L367: The authors take the GNSS VLM as ground truth, and that is fine, but are the formal VLM errors similar among the TG and GNSS stations, respectively? Formal errors can provide valuable information for the VLM validation and this should be accounted for when assessing the VLM differences (WRMS instead of RMS for instance).

Formal errors are still much lower for the GNSS VLM estimates than for the SAT-TG estimates. To take these into account, we re-computed the weighted RMS as well as the weighted STD of the trend deviations as follows (also as a response to one of the subsequent comments):

$$RMS_{weighted} = \sqrt{\sum_{i=0}^{n} w_i (GPS_i - SATTG_i)^2}$$

$$STD_{weighted} = \sqrt{\sum_{i=0}^{n} w_i ((\overline{GPS - SATTG}) - (GPS_i - SATTG_i))^2}$$

With weights $w_i = \frac{\sqrt{(GPSuncertainty_i^2 + SATTGuncertainty_i^2)^{-1}}}{\sum_{i=0}^{n} \sqrt{(GPSuncertainty_i^2 + SATTGuncertainty_i^2)^{-1}}}$

Table 1: Comparison of SAT-TG minus GPS trends: first column: RMS not weighted as in manuscript, 2nd: weighted RMS and 3rd: weighted standard deviation.

	RMS-normal	RMS-weighted	STD-weighted
ALES_GESLA_250km	1.51	1.47	1.39
ALES_PSMSL_250km	1.68	1.57	1.46
ALES_GESLA_ZOI	1.28	1.37	1.19
AVISO_PSMSL_250km	1.50	1.48	1.32

Using a weighted RMS most strongly improves the ALES_PSMSL_250km configuration, but it has a smaller effect on the other data sets. This is an interesting finding, which shows that lower formal uncertainties are not in every case associated with more accurate trend estimates. Such de-coupling of accuracies and uncertainties was also addressed in the discussion and points towards other undetected error sources, which limit the comparability of SAT-TG and GNSS.

We added the formulation of the weighted RMS in section 3.4 and associated results in table 2.

A weighted STD improves all of the datasets, because here the mean bias of the differences does not have an impact on the performances anymore.

We decide not to add the (weighted) STD in the table, but add a more thorough discussion on causes of trend biases in the section 5.2. Systematic errors (for more details please refer to next comment).

L377-380: the differences between the SLA trends in ALES and AVISO are quite significant. From Table 2 it appears that the median SLA value from AVISO is 1 mm/year higher than that from ALES (also seen in Fig. 4). Is this correct? If so, how would you explain this difference?

We agree that there is a strong difference between the median (or as we call it the bias) of the two dataset combinations. Also comparing the bias of AVISO-PSMSL with the result from Wöppelmann and Marcos 2016 we obtain a much larger value. We briefly addressed this issue by pointing out the differences in the settings of the different studies (250km range instead of 1° average, other time periods and TGs (numbers and locations)). Also in response to the second reviewer, we integrated the discussion of the impact of possible mission drifts, which could generate systematic trend biases. We added some further lines in the introduction as well as in the results section:

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.'

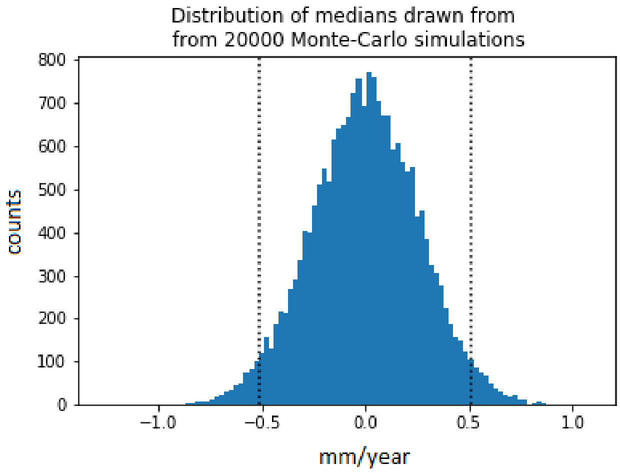
In the results section, we add following lines to the paragraph from L385:

'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).'

We added some additional statistical analysis in the supplemental material (from L581):

'To gain a better understanding why the VLM SAT-TG and VLM GNSS difference distributions are significantly biased, we create a Monte-Carlo experiment to check the H0 hypothesis: 'the median of the distribution is not significantly different from zero' (with alpha=0.025). Therefore, we generated a bootstrapped distribution of random medians, which are derived from 20,000 individual sub-sets of size 52 (the number of TG of our dataset), which are randomly drawn from normally distributed values with a standard deviation of 1.5 mm/year (according to the RMS of AVISO-PSMSL-250km) and zero mean.

Figure 1 shows that the biases of the datasets ALES-PSMSL-250km and AVISO-PSMSL-250km exceed the 2.5 and 97.5 percentiles of the sampled distribution (average of absolute bounds: 0.512 mm/year). This means that in less than 5 out of 100 cases, we would obtain such biases by chance, which supports the significance of these biases. We highlight that this is a purely statistical analysis, which cannot account for any of the errors from corrections, adjustments, drifts etc. introduced in the altimeter and GNSS.'



*'**Figure B1:** Histogram of median values of randomly sampled sub-sets. The sub-sets consists of 52 samples (according to the number of TGs in AVISO-PSMSL-250km) and are randomly drawn from normal distributed values with zero mean and a standard deviation of 1.5 mm/year (according to the RMS of AVISO-PSMSL-250km). Dashed lines mark the 2.5 and 97.5 percentiles of the distribution.'*

Given these results, we dedicated another sub-section to trend biases and systematic errors in the discussion (section systematic errors, after 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).

Next to altimeter bias drift, non-linear 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 non-linear 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 reduces 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 support the identification and mitigation of such systematic errors.'

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.

L386: "absolute median bias of trend differences" can be confused with the median of absolute VLM differences. I suggest changing this by "median of the VLM differences" or similar here and elsewhere.

Corrected to: '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].'*

L400-401: Some comments on results shown in Table 2: the unweighted RMS from GESLA is smaller than that from PSMSL, but this is probably because the median value is closer to zero. The weighted standard deviation or any other measure of dispersion (interquartile range) that does not include the mean/median value would be more appropriate here. Also the formal VLM rate uncertainties are higher with GESLA than with PSMSL, even with a spectral index slightly closer to zero. This means the noise (especially the power-law variance) of the residual series (trend and seasonal variations removed) in the GESLA VLM series is larger than that from PSMSL or that the GESLA series are significantly shorter or less complete. In that case, the choice of using GESLA instead of PSMSL would need better argumentation. There may be also a TG trend bias between GESLA and PSMSL of around 0.3 mm/year. This is probably not significant, but it may be worth discussing.

In a previous comment, we added the statistics weighted RMS as well as weighted STD, which confirm that ALES-GESLA-250km still outperforms ALES-PSMSL-250km in terms of accuracy. As mentioned by the reviewer, larger trend uncertainties for the GESLA configuration can be a result of larger power-law variance. We found that, the median driving noise of ALES_GESLA_250km is by 5% larger than for ALES_PSMSL_250km. Thus, we further discuss such potential causes of the trend uncertainty differences by adding to L402:

'Compared to ALES-PSMSL-250km, we find increased trend uncertainties for ALES-GESLA-250km, which can be partially explained by higher power-law variance of this GESLA -based configuration. '

! Due to the update of Envisat data (see first response), the mentioned 0.3 mm/year difference of trend biases (ALES-GESLA-250km vs. ALES-PSMSL-250km) increased to 0.48 mm/year !

Relating to our previous response, showing the probability of occurrence of a median, we argue that for such a sample size a trend bias of 0.48 mm/year is not significant. Please also refer here to the discussion of the impact of systematic errors on trend biases (section 5.2). The general question raised by the reviewer, which requests for a better justification of the use of one TG dataset over another needs to be better assessed. Therefore, we add following lines to the previous corrections (L402):

'… variance of this GESLA -based configuration. 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 provides better results concerning trend accuracy (weighted or unweighted RMS) and has a lower median bias. Moreover, using the high-frequency data, we are able to couple SAT and TG observations at much higher temporal resolution than it would be the case when using monthly PSMSL data. ~~Given the strong improvement in the bias,~~ Therefore, the ALES-GESLA coupling is further developed based on a better definition of the ZOI in the next section. '

L421-422: the power-law variance may have also changed, maybe producing a significant improvement of the VLM formal errors.

We added: 'stems from the reduction of the power law and white noise amplitudes'

L429-436: The discussion in this paragraph is not very clear and would require improvements. We only need a single SLA series to estimate VLM from SAT-TG. This SLA series can be obtained using different strategies as the authors have discussed: spatial averaging/filtering of SLA data, the single most correlated SLA series, the single

closest SLA series, etc. The more similar the selected SLA series is to the TG series, i.e., the smaller the SAT-TG differences (again excluding the seasonal variations that are captured by the model fitted to the SAT-TG series), the more precise VLM will be obtained in terms of formal error. This is a metric very easy to interpret. Note that interpreting how the SAT-TG VLM values compare to the GNSS VLM is much more complex and is, in general, not a strong criterion given the large differences between both. The smaller quantity of averaged SLA should not be blamed if they represent increasingly consistent SLA series with respect to the TG series. A different explanation for the bad results with >80% thresholds could be that the RMS metric is not telling us whether the SAT and TG series are more similar, especially if the seasonal signals were included in the RMS as per my comments above. In addition, the RMS alone is not directly tied to the autocorrelation of the SAT-TG series (i.e., the spectral index), which is another important metric to assess the consistency of the SLA and TG series.

In this paragraph (L429-436), we discuss why the RMS (of SAT-TG and GNSS trend differences, i.e. the accuracy of trend estimates) increases when we select very highly comparable data, or smaller sub-sets of altimetry data. Overall, we still argue that at very high levels (which can also mean less selected tracks) a mere decrease in sample size of the time series is the major reason for decreased accuracies of the SAT-TG trends. As an example, a 95% level-ZOI selection (based on RMS) would only hold 80% of the samples (i.e. number of monthly averaged observations) which we would obtain at the 80% level-ZOI selection. Considering the subsequent analysis in the discussion (e.g. dependence of accuracies on the length of the covered time period) this is in our understanding the most obvious explanation for decreased accuracies, when we strongly decrease the sampling density at high levels of comparability.

In general, we fully agree that there is a large range of error sources which influence the comparability of SAT-TG and GNSS trends. However, when we only adjust the amount of selected SAT observations, we keep much of those error sources constant (mission drift biases, nonlinear VLM, some of the errors in applied corrections …). Thus, because we reduce the number of samples e.g. at a 95% level, compared to a 80% level we came to this conclusion.

Therefore, to better clarify our explanations we modify the paragraph as follows (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 sam-pling-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 compa-rable (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 com-bination 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.'

The reviewer argues that, because we model the annual cycle this might be one reason why our metric, the RMS of SAT-TG differences, is not telling us whether SAT and TGs are more similar to each other and thus inadequately expresses deviations between SAT and TG time series. We note, however, that this metric indeed also captures differences in the annual cycle between both time series (see previous discussion, because the RMS is computed based on the differences of the *detrended but not de-seasoned time series). We then use this metric to confine the ZOI, to average data from which we again compute SAT-TG time series. Concerning the spectral index, we agree that the RMS is not directly related to this metric.*

L466-478: the discussion here is interesting, but an important point should be stated more clearly. The optimal ZOI in Fig. 5 a&c was retained by assessing the consistency between SAT and TG series. On the other hand, the local optimal ZOIs in Fig. 5b are defined by comparing SAT-TG and GNSS VLM estimates. It is therefore not surprising that different optimal local ZOIs are obtained and that there is not an optimal threshold that fits well all sites. In addition, imposing the GNSS VLM as a criterion to the computation of the SAT-TG VLM would remove its independent nature.

We change this paragraph to better clarify the meaning of the results shown in Figure 5 as well as the aim of our interpretations of Figure 5 (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 is however unclear, whether the described optimum 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.

L487-488: I guess you derived the 3 to 4 inflation factor from the color range in Fig. 6, but Table 2 actually shows that using different SLA data does increase VLM uncertainties by less than 10% (comparing ALES-PSMSL to AVISO-PSMSL).

Exactly, we better specify these lines (L487-488):

These examples show, that at individual locations ~~The~~ the use of less comparable SLAs can increase uncertainty by a factor of three to four.