The Zone of Influence: Matching sea level variability from coastal altimetry and tide gauges for vertical land motion estimation

Julius Oelsmann[1], Marcello Passaro[1], Denise Dettmering[1], Christian Schwatke[1], Laura Sanchez[1], and Florian Seitz[1]

[1]Deutsches Geodätisches Forschungsinstitut der Technischen Universität München, Arcisstraße 21, 80333 Munich, Germany

Correspondence: Julius Oelsmann (julius.oelsmann@tum.de)

Abstract.

Vertical land motion (VLM) at the coast is a substantial contributor to relative sea level change. In this work, we present a refined method for its determination, which is based on the combination of absolute satellite altimetry (SAT) sea level measurements and relative sea level changes recorded by tide gauges (TG). These measurements complement VLM estimates from GNSS (Global Navigation Satellite System) by increasing their spatial coverage. Trend estimates from SAT and TG combination are particularly sensitive to the quality and resolution of applied altimetry data as well as to the coupling procedure of altimetry and tide gauges. Hence, a multi-mission, dedicated coastal along-track altimetry dataset is coupled with high-frequency tide gauge measurements at 58 stations. 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. Selecting 20% of the most representative absolute sea level observations in a 300 km radius around the tide gauges results in the best VLM-estimates in terms of accuracy~~ies~~ and uncertainty~~ies~~. At this threshold, $VLM_{SAT\text{-}TG}$ estimates have median formal uncertainties of ~~0.59~~0.58 mm/year. Validation against GNSS VLM estimates yields a root-mean-square ($RMS_{\Delta VLM}$) of $VLM_{SAT\text{-}TG}$ and VLM_{GNSS} differences of 1.28 mm/year, demonstrating the level of accuracy of our approach. Compared to a reference 250 km radius selection, the 300 km Zone of Influence improves trend accuracies by ~~12~~15% and uncertainties by ~~28~~35%. With ~~progressing~~ increasing record lengths, the spatial scales of the coherency in coastal sea level trends ~~coherency~~ increase. Therefore the relevance of the ZOI for improving $VLM_{SAT\text{-}TG}$ accuracy decreases. Further individual Zone of Influence adaptations offer the prospect of bringing the accuracy of the estimates below 1 mm/year.

1 Introduction

Coastal vertical land motion (VLM) significantly contributes to relative sea level change (SLC). VLM is in many places of the same order of magnitude (1-10 mm/year) as the sea level rise signal itself and displays significant spatial variations (Santamaría-Gómez et al., 2012). Consequently, VLM affects coastal impacts of climate-sensitive processes and can regionally account~~s~~ for large fractions of the observed and projected coastal SLC signal (Wöppelmann and Marcos, 2016; Slangen et al., 2014). Thus, the accurate estimation of VLM is vital, not only to disentangle climatic and geodynamic SLC signatures, but also to obtain more robust estimates of past and future relative SLC and its associated uncertainty (Church et al., 2013; Santamaría-Gómez

et al., 2017). In this work, we present a novel approach of VLM estimation using coastal satellite altimetry, tide gauges and the Global Navigation Satellite System (GNSS).

VLM is caused by the superimposition of natural processes and anthropogenic influences in the Earth System and operate on manifold spatial and temporal scales (Pugh and Woodworth, 2014). Mechanisms such as the Glacial Isostatic Adjustment (GIA), the postglacial rebound of the Earth to changing ice and water load, cause distinct large-scale VLM, which can be assumed to be uniform on centennial timescales (Peltier, 2004). Recent acceleration of land ice mass loss was shown to additionally enhance deformation rates, posing new challenges for sea-level studies due to its time-varying signal (Riva et al., 2017). Surface mass changes are also caused by terrestrial freshwater storage changes and can have small-scale effects on VLM. Groundwater pumping for instance, contributes not only to local small-scale VLM and gravity changes, but also modifies sea level rise in distant areas (e.g. Wada et al. (2012); Veit and Conrad (2016)). Other small-scale VLM effects such as erosion or tectonic movements can be locally confined to several kilometers with more subtle, not necessarily linear temporal behaviour (Brooks et al., 2007; Kolker et al., 2011; Poitevin et al., 2019). 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).

In response to the substantial impact on relative sea level and the large spectrum of VLM sources, several strategies have been developed to estimate VLM. The ability to capture the diversity of VLM processes, however, strongly depends on the method and geodetic technique used in the VLM estimation. Furthermore, the coverage and associated accuracy of VLM estimates differ across the methods. Given that global absolute sea level trends during the altimeter era is in the order of ~~1 to 3 mm/year~~ 3 mm/year (3.1 $\pm$0.1 mm/year from 1995 to 2018 as reported in (Cazenave et al., 2018)), one prerequisite for VLM estimation is that the associated trend uncertainty should be at least one order of magnitude lower than those subtle signals (Wöppelmann and Marcos, 2016). Hence, dense and accurate VLM estimates are required to complement modelled or measured rates of absolute sea level rates, which is ultimately crucial for coastal planning. Improving the reliability of VLM estimates and their associated uncertainties is thus one major concern of this study. In the following, we briefly contrast the three major approaches of deriving coastal VLM globally.

1.1 Estimating Coastal Vertical Land Motions

The majority of global sea-level-studies utilized geodynamic GIA-models to correct, for example, tide gauge records for secular land motion trends or to extrapolate future relative SLC based on climate-projections (e.g. Church and White (2011); Hay et al. (1990); Carson et al. (2016)). GIA still represents the only long-term geological process for which VLM can be modeled on a global scale. However, one caveat is that GIA-VLM models were shown to be still biased by imperfect assumptions of ice history and the Earth's structure (King et al., 2012) and are thus model-dependent (Jevrejeva et al., 2014). Another foreseeable disadvantage is that the sole application of GIA-models neglects other sources of VLM (e.g. tectonics, erosion or anthropogenic impacts, Wöppelmann and Marcos (2016)). This led, for instance, to discrepancies in estimated rates of historical global mean sea level (GMSL) change, when comparing model-based solutions against those using measurements from GNSS (Hamlington et al., 2016).

For more than a decade, these direct geodetic estimates (GNSS, such as GPS, GLONASS or GALILEO) have been exploited to determine vertical velocities (Wöppelmann et al., 2007; Snay et al., 2007; Mazzotti et al., 2008). GNSS measurements denote the most precise source of VLM detection and are well established in local to global scale studies (e.g. Bouin and Wöppelmann (2010); Fenoglio et al. (2012); Santamaría-Gómez et al. (2017)). Wöppelmann and Marcos (2016) identified ~~considerable~~ comparably low formal errors of GNSS-VLM rates (0.21 mm/year) when auto-correlation was taken into account. Santamaría-Gómez et al. (2012) estimated an ~~precision~~ accuracy of 0.6 mm/year of GNSS-based VLM (from at least three years of continuous data), by comparing 36 globally distributed co-located GNSS velocity estimates. Thus, because of its considerable accuracy, vertical GNSS velocities frequently served as benchmark estimates for many sea-level applications, e.g. for GIA-model evaluation or local VLM-corrections of TG records (Sánchez L., 2009; Sanli and Blewitt, 2001).

For the latter use a necessary working hypothesis is that GNSS-VLM represents the same movements as experienced at the tide gauge (Wöppelmann and Marcos, 2016). Because VLM is shown to potentially possess high spatial variability even on small scales (tens of kilometers), GNSS-stations should be ideally very close to the tide gauge. This requirement, however, reduces the number of available co-located stations (130 GNSS stations within a 1 km range of GLOSS (Global Sea Level Observing System) tide gauges, Wöppelmann et al. (2019)) and thus confine the global coastal coverage to mostly Europe, Japan and North America.

To extend the number of VLM estimates, several studies advanced the application of combining satellite altimetry (SAT) and tide gauge (TG) observations (Cazenave et al., 1999; Nerem and Mitchum, 2003; Kuo et al., 2004; Pfeffer and Allemand, 2016; Wöppelmann and Marcos, 2016; Kleinherenbrink et al., 2018). The principle of this approach is to subtract the absolute SLC gathered by the altimeter from relative SLC observations at the tide gauge. Ideally, the differenced time series (SAT minus TG) yields the vertical displacement of the tide gauge with respect to the reference of the altimeter. 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. Due to the availability of global and continuous absolute sea level measurements, this method not only provides a complementary source to GNSS measurements, but also improves the geographical distribution of the data, as virtually every valid TG is usable.

While all of these three sources of information, GIA-models, GNSS and 'satellite altimetry minus tide gauge' (SAT-TG) techniques have individual merits, ~~synergetic applications are~~ their combined use is valuable to further substantiate VLM estimates. GNSS-observations are necessary to validate both GIA-models and the SAT-TG approach (Santamaría-Gómez et al., 2012; Wöppelmann and Marcos, 2016; Kleinherenbrink et al., 2018). Recent studies combined all three approaches to reconstruct GMSL (Dangendorf et al., 2017), or to densify the estimation of contemporary rates of vertical land motions (Pfeffer et al., 2017) or relative sea level change (Hawkins et al., 2019). Any advancement in these individual approaches, therefore, supports developments of the others and improves the global assessment of coastal VLM estimates.

In this study, we focus on enhancing the application of SAT-TG difference for VLM detection. Our investigations not only further develop ~~carry on~~ the latest progress of the method, but also gain a new perspective on sea level trend and uncertainty characterization and quantification ~~detection~~ in coastal zones. The next section recapitulates the latest state of $VLM_{SAT\text{-}TG}$ estimation on which we base our innovations.

1.2 Progress in VLM estimation by satellite altimetry and tide gauge difference

The combination of SAT and TG observations for VLM determination was steadily improved over the last two decades and is elaborated in the latest review by Wöppelmann and Marcos (2016), hereinafter WM16. WM16 investigated performances of different gridded and along-track altimetry products (e.g. AVISO (Archiving, Validation, and Interpretation of Satellite Oceanographic data) and GSFC (Goddard Space Flight Center). They combined sea-level anomalies (SLAs) as 1°-radius averages with monthly-mean tide gauge records from PSMSL (Permanent Service for Mean Sea Level). Among all datasets, the gridded AVISO product revealed the best correlations and residuals between altimetry and tide gauge records. Using this dataset, WM16 also obtained the most precise VLM estimates ~~were achieved~~, with median formal uncertainties of 0.8 mm/year. Validation against GNSS-based trends from ULR5 (Université de La Rochelle, Institut Géographique National analysis) at 113 colocated stations resulted in an $RMS_{\Delta VLM}$ of 1.47 mm/year of $VLM_{SAT\text{-}TG}$ and VLM_{GNSS} trend differences, providing the highest accuracies among the datasets.

Notwithstanding the weaker performance of the along-track product (from GSFC) achieved in WM16, Kleinherenbrink et al. (2018) made great progress in using along-track altimetry (Topex, Jason1 and 2, from the Radar Altimeter Database, RADS (Scharroo et al., 2012)) to estimate VLM. Their approach aimed to overcome the spatial downsampling and associated loss of information in gridded products such as AVISO. They also advanced the procedure of combining altimetry and tide gauge data. Instead of taking 1°-averages around the tide gauges, they selected altimetry data according to different absolute correlation thresholds and implemented correlation-weighting of time series. Generally, the thresholding strategy functioned as a filter to remove stations of low comparability: With varying correlations between 0.0 and 0.7, they obtained $RMS_{\Delta VLM}$ errors from 2.1 to 1.20 mm/year (at 155 stations), which significantly improved WM16's results. For a consistent set of stations, they found a slight sensitivity of the $RMS_{\Delta VLM}$ to variations of the prescribed minimum correlation, however, they reported insignificant improvements of ~~with insignificant improvements~~ of a few percent. Because they derived GNSS vertical velocities from the Nevada Geodetic Laboratory (NGL) database by taking the median of available estimates within a 50 km range to the tide gauge, they increased the number at which $VLM_{SAT\text{-}TG}$-trends could be validated to 155.

Based on Kleinherenbrink et al. (2018) and WM16, we identify two essential factors which are vital for the quality of trend estimation by SAT-TG difference. Advancements with respect to both factors not only potentially led to improved VLM estimates in Kleinherenbrink et al. (2018), but also motivate for further innovations:

1 **Data quality**: In coastal regions accuracy of altimetry measurements is affected by the local departure of the radar signal from the known ocean response (due to inhomogeneities of the illuminated area) and by the inaccuracy of the standard routinely applied corrections and tidal models. Developments for the solution of both issues led to rapid improvements

in the recent years by e.g. application of coastal-retracking and advanced geophysical corrections (e.g. Cipollini et al. (2017); Passaro et al. (2014); Fernandes et al. (2015)). Dedicated coastal altimetry datasets (e.g. COASTALT, ALES, PISTACH) might thus outperform previously applied products (e.g. AVISO), which do not yet benefit from these implementations.

2 **Data selection**: Next to issues concerning data quality, the second factor defining trend uncertainty is the sensitivity of $VLM_{SAT\text{-}TG}$ estimates to the spatial selection of altimeter data in the vicinity of the tide gauge. WM16 showed, that averaging SLA in a radius of 1° around the tide gauge resulted in higher correlations than using the best correlated or the closest grid point to the tide gauge. Kleinherenbrink et al. (2018) found a small influence of variations of absolute correlation thresholds on the trend estimates. Therefore, considering the diversity of processes~~,~~ which drive coastal sea level variability, such as waves, winds or coastal and bathymetric properties, an advanced adaptation of the choice of altimetry SLA might improve representation of the signal captured by the tide gauge.

These reasons motivate for further improvements of both components, quality of the data and practise of combining altimetry and tide gauge data. We aim to understand how dedicated along-track coastal altimetry can outperform standard-gridded products. We also seek to generalize an optimal selection of SLA's, underpinned by the local dynamical features of measured sea level variability.

In this work, we present a new approach of combining SAT and TG observations to improve VLM estimates. In contrast to previous attempts, we exploit TG and SAT data at the highest available temporal and spatial scale for globally distributed stations. We couple advanced coastal altimetry data with high frequency tide gauge records from the Global Extreme Sea Level Analysis (GESLA). Implementation of these high frequency tide gauge records constitutes a further innovation for VLM estimation. So far such data has only been applied in local studies (Idžanović et al., 2019) and monthly tide gauge data were commonly exploited in this regard. We show that precision and accuracy of the trend estimates can be optimized, when using refined spatial selection criteria of altimetry sea level anomalies. With this approach we identify coherent zones of sea level variability, which best represent the coastal in situ measurements. Our method is generally transferable to analysis of coastal sea level trend determination.

Sections 2 and 3 describe the individual datasets, applied processing steps and the optimization of combining altimetry and tide gauge data. Section 4 presents performance of trend estimates, i.e. estimated uncertainties and validation against GNSS data (in this study all GNSS data are based on the Global Positioning System (GPS)). Finally, we contrast our results and methods with previous work and discuss the impact of the interconnection of time and space-scales on the evolution of coherency of sea level in coastal regions (section 5).

2 Data

We use different altimetry products, in order to assess the impact of special coastal products on associated $VLM_{SAT\text{-}TG}$ trend estimates. We compare the coast-dedicated retracker ALES (Adaptive Leading Edge Subwaveform retracker, Passaro et al.

Table 1. Applied models and geophysical corrections for estimating sea level anomalies.

Parameter	Model	reference
Range and Sea State Bias	ALES	(Passaro et al., 2014)
Inverse barometer	DAC-ERA	(Carrère et al., 2016)
Wet troposphere	GPD+	(Fernandes et al., 2015)
Dry troposphere	VMF3	(Landskron and Böhm, 2018)
Ionosphere	NIC09	(Scharroo and Smith, 2010)
Ocean and Load tide	FES2014	(Carrère et al., 2015)
Solid Earth and Pole tide	IERS 2010	(Petit and Luzum, 2010)
Mean Sea surface	DTU18MSS	(Andersen et al., 2018)

(2014)) along-track product against the interpolated AVISO dataset (sections 2.1 and 2.2). Altimetry data are combined with tide gauge observations from the monthly mean PSMSL and the high frequency GESLA data base, which are described in sections 2.3 and 2.4. We develop a new coupling strategy of high-rate altimetry and tide gauge records in section 3.2.

2.1 Coastal along-track altimetry - ALES

The coastal altimetry product is constructed from 1-Hz multi-mission altimetry measurements processed by DGFI-TUM (OpenADB, https://openadb.dgfi.tum.de). We combine data from the missions ERS-2, Envisat, Saral, Jason1-Jason3 and their extended missions, which provide continuous altimetry time series of 23 years (1995-2018). For all missions, satellite orbits in ITRF2008 are used, mostly processed by CNES (GDR-E). For ERS-2, GFZ VER11 orbits are applied. The SGDR data are re-processed with the ALES retracker (Passaro et al., 2014) and an improved sea state bias correction scheme (Passaro et al., 2018). The geophysical corrections are summarized in Table 1. They are consistent with those incorporated in the latest development of the empirical ocean tide model (EOT19p) by Piccioni et al. (2019). If available, the Dynamic Atmopsheric Correction consists of the ECMWF ERA-Interim reanalysis (DAC-ERA, Carrère et al. (2016)). This product especially reduces along-track sea level errors in the earlier missions (in this study ERS-2). Because this product is unavailable for the very recent missions, we implement the DAC (Carrère and Lyard, 2003) based on ECMWF for the last cycles of Jason-2 (and its extended mission) and the full Jason-3 and Saral missions. 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.

We map all altimetry records on 1-Hz nominal tracks consistent with the CTOH nominal paths (Center for Topographic studies of the Ocean and Hydrosphere, www.ctoh.legos.obs-mip.fr) of the individual missions, using nearest-neighbour interpolation. Then, we scan the data for outliers along the tracks, to hinder spurious extreme values to propagate in time series. This scheme features:

- Absolute thresholds: Any absolute SLA exceeding ±2 m is excluded.
- Running median test: If the absolute difference of the data and its running median (centered, over 20 points) is greater than ± 12 cm, data are excluded.
- Consecutive difference test: Outliers are detected when the difference of consecutive points exceeds ±8 cm. The test identifies the outliers according to the differences of the other neighbouring values

The absolute thresholds (±12 cm, ±8 cm) correspond to 2-σ of the median running variability and 2-σ of absolute consecutive differences based on the analysis of different tracks of Jason-2 and ERS-2.

SLAs along the same track and cycle are then averaged over predefined areas as described in sections 2.5 and 2.6. We built a time series by considering all averaged SLAs from the along-track multi-mission dataset for the study period. To check for outliers in each SLA time series, we exclude values exceeding absolute values of ±3-σ of the data. This cleaned 1 Hz coastal altimeter dataset is hereinafter called ALES and used for the combination with the TG-datasets described in section 3.1.

2.2 Gridded altimetry data - AVISO

The gridded Ssalto/Duacs altimeter product was produced and distributed by the Copernicus Marine Environment Monitoring Service (CMEMS, http://marine.copernicus.eu) and is hereinafter called AVISO as it was previously distributed by CNES AVISO+. We use monthly sea level anomalies, which are resolved on a 0.25° Cartesian grid and cover the period from 1992-2019. The product already includes the DAC (Carrère and Lyard, 2003) comprising the dynamical barotropic ocean response to atmospheric forcing (modelled with MOG2D-G), as well as the inverse barometer (IB) response. Consistent with the along-track dataset ALES, FES2014 (Carrère et al., 2015) is implemented to correct for tidal signals. Other corrections and pre-processing steps are documented by CMEMS.

2.3 Monthly tide gauge data - PSMSL

We use monthly mean tide gauge data from the datum controlled PSMSL (Holgate et al., 2013) database. The PSMSL constitutes the primary source of tide gauge data for most ~~of~~ sea-level research, or for the assessment of long-term trends of VLM based on SAT and TGs. The service undertakes quality control of the data including checks for consistency of the annual cycle, outlier detection or intercomparisons with neighbouring stations, which enhances the reliability of the data. Among all available stations, we select those which contain at least 180 months (15 years) of valid measurements during the altimetric era (1993 - present), resulting in a total number of 627 stations. We apply the same monthly-averaged DAC-correction as used

for the AVISO data (Carrère and Lyard, 2003). To match the DAC-correction with the tide gauge records, we select among the 9-closest grid-points of the solution, the one which results in the highest variance reduction.

2.4 High frequency tide gauge data - GESLA

In addition to monthly-mean PSMSL TG-data, we exploit the GESLA dataset (Woodworth et al., 2016), which contains a large global collection of high-frequency TG records with sampling rates ranging from hours down to 6 minutes. The latest version GESLA 2 contains in total 1355 station records and was assembled from a variety of international and national databanks (e.g. UHSLC (University of Hawaii Sea Level Center) and GLOSS) or independent sources. It thus also shares many stations with the monthly PSMSL database, the preferred dataset for $VLM_{SAT\text{-}TG}$ computation. Unfortunately, at this time, GESLA holds only data until 2015. Therefore, we also restrict the considered period for all dataset combinations (see section 3.1) to before 2015. As for the PSMSL data, we select stations with at least 180 months of valid data.

In contrast to PSMSL data used in WM16, GESLA tide gauges feature no rigorous outlier rejection by default in except that of the primary data providers (Woodworth et al., 2016). Extreme values from strong signals like tsunamis or station shifts and other irregularities are still present in the data. Some of those issues are adressed on the GESLA-webpage, however, for the sake of long-term trend evaluation, we perform a further global outlier analysis. Therefore, we check all tide gauge time series manually for irregularities: Station shifts from seasonal to interannual timescales are either handled by dismissing certain sections of the time series or completely excluding the tide gauge from the analysis. Single extreme events from hourly to monthly timescales are only excluded, when they deviate from the measurements by several meters, because we want to maintain as much data as possible. If such events are present, we flag any values beyond the upper/lower 0.999 quantiles of a fitted normal distribution of the data. Occasionally we apply this quantile-outlier exclusion recursively.

To obtain a uniform temporal resolution, we resample this outlier-free tide gauge set to hourly records by cubic interpolation. The records are then corrected for the tidal signal as well as for the ocean response to atmospheric wind and pressure forcing. The tidal variability is suppressed by using a 40-h loess (locally estimated scatterplot smoothing) filter (Cleveland and Devlin, 1988) as in Saraceno et al. (2008). 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). In accordance with PSMSL-tide gauge data, we implement the same Dynamic Atmospheric Correction (Carrère and Lyard, 2003). This solution features a 6h sampling frequency, which is therefore down-sampled to hourly anomalies by cubic interpolation. For the global dataset, we obtain a mean variance reduction of 37.8% and a mean correlation of 0.6. As in WM16 and Ponte (2006), we find a distinct latitude dependence of correlations and variance reduction, with decreasing performance nearer to the equator ~~in low-latitude regions~~. We note, that the total variance reduction, which we apply on the high-rate tide gauge data is naturally less than in WM16, who corrected monthly mean, detrended and deseasoned data.

3 Methods

3.1 Dataset combinations

To understand the sensitivity of the VLM estimations on (1) quality and resolution of the data and (2) the selection procedure, we analyse the performances of four different dataset combinations: ALES-PSMSL-250km, ALES-GESLA-250km, AVISO-PSMSL-250km and ALES-GESLA-ZOI.

The first three combinations are constructed to compare the performances of the along-track (ALES) against the gridded altimetry product (AVISO) combined with monthly tide gauge observations. With ALES-GESLA-250km we also investigate the possible advantage of using the GESLA high rate tide gauge product. For all these experimental sets, SLA time series are merged or averaged within a 250 km radius around the tide gauges, which is thus a selection procedure independent of the actual comparability of SLAs.

To produce the ALES-GESLA-250km dataset, we derive differences of the merged, non-uniformly sampled SLAs and the hourly-sampled GESLA tide gauge records, by cubic interpolation of the latter and a maximum allowed time-lag of 3 hours between the measurements. We down-sample these high-rate differenced time series to monthly means. For ALES-PSMSL-250km on the other hand, we first compute monthly-means from SLAs and subsequently subtract these monthly SLAs from the monthly-sampled relative SLAs from PSMSL. Finally, we directly compute the differenced SAT-TG time series from the averaged monthly AVISO and the PSMSL data, which yields the AVISO-PSMSL-250km dataset.

Using these combinations, we investigate the mere changes from differences formed using ~~matching~~ along-track data at high or at low frequency (ALES-GESLA-250km and ALES-PSMSL-250km), or using monthly gridded data (AVISO-PSMSL-250km). Here, 'high-frequency' refers to daily time scales of variability and 'low-frequency' to monthly time scales. The dataset ALES-GESLA-ZOI incorporates further SLA-selection-schemes, which are explained in the following section.

3.2 The Zone of Influence

We aim to develop a new SLA-selection scheme, which accounts for the observed coherency of sea level variability. However, due to the diversity of the underlying physical mechanisms and their complex interplay with the coast, the spatial coherency of sea level dynamics is highly variable in coastal regions (Woodworth et al., 2019). Coastally trapped waves, for instance, were argued to establish long-range correlations along the continental slopes (Hughes and Meredith, 2006) and to mediate the influence of the open ocean (Hughes et al., 2019) on the coast. While some signals, such as interannual modes of climate variability, generate high spatial coherence, other local features, such as the presence of a coastal current, can significantly modify the sea level variability within few kms of the coast, as shown in the case of the seasonal signal of the Norwegian Coastal Current in Passaro et al. (2015). Accordingly, the capability of compare TG-based sea level variability with altimetry~~,~~ utterly depends on which time and length scales are resolved by the data.

The key concept of our approach is to capture the extent to which coastal altimetry measurements are similar to the in situ tide gauge observations. To do so, we extend the methodology proposed by Santamaría-Gómez et al. (2014), who looked for the altimetry grid point mostly correlated with the tide gauge, and Kleinherenbrink et al. (2018), who considered a larger set

of points based on absolute thresholds of correlation. In contrast to these previous studies, we assess the influence of using relative thresholds of comparability on both the accuracy and the uncertainty of the trends.

We exploit combinations of along-track ALES data and high frequency GESLA records, to identify regions of sea level variability that show maximum coherency with tide gauge observations, which we hereinafter call the Zone of Influence (ZOI). With this approach, our objective is to decrease noise of the differenced, high frequency $VLM_{SAT\text{-}TG}$ time series using the ZOI to hone trends and uncertainty estimates.

To define the ZOI, we investigate different statistical criteria S, which provide a measure of similarity of sea level variability between TG and SAT observations. Here, we use the Pearson correlation coefficient, the $RMS_{SAT\text{-}TG}$ as well as ~~as~~ the amplitude of the residual annual cycle between both TG and SAT records. We compute each of those measures for every point of the 1 Hz along-track data (ALES) in combination with the TG records from GESLA. As for ALES-GESLA-250km, tide gauge data are interpolated onto the time step of the altimetry records. Correlations and $RMS_{SAT\text{-}TG}$ are computed from the detrended TG and SAT time series. The amplitude of the residual annual cycle is obtained from the remaining seasonal signal of the difference of the time series (SAT-TG). We acquire a dataset, holding information of the performance of multi-mission along-track data in the vicinity of every GESLA TG. 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 are identified.

To confine the ZOI, we select sub-sets of the data containing the best-performing statistics (i.e. highest correlation, lowest $RMS_{SAT\text{-}TG}$ 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. Every sub-set (X, S) represents an individual ZOI, in which we average SLAs in accordance with the steps involved in the aforementioned 250 km-radius-selection (ALES-GESLA-250km, section 3.1). The high-rate SLA time series (ALES) are then again subtracted from GESLA, providing the ALES-GESLA-ZOI dataset for VLM estimation (section 3.3).

Note that in contrast to the 250 km selection, we extend the range in which SLAs are taken into account to 300 km to define the ZOI. The previous 250 km selection is, as in Kleinherenbrink et al. (2018), based on the space auto-correlation scales of SLAs, which reflect characteristic eddy length scales (Stammer and Böning, 1992; Ducet et al., 2000). These scales decrease towards higher latitudes (i.e. towards the poles) with changing internal Rossby radius. However, several studies found much larger correlation length scales of SLAs, in particular along shorelines (Calafat et al., 2018; Hughes and Meredith, 2006). Other mechanisms than mesoscale eddy activity were investigated to account for these coherent changes. One example is given by Calafat et al. (2018), who analysed the driving factors of sea-level variability at the south-eastern coast of the US. Using altimetry and three different ocean-models, they found coherent changes of the annual amplitude of SLAs over length scales of thousands of kilometers along the coast from the Yucatan Penisula to Cape Hatteras. While the annual cycle signal itself was dominated by steric changes, with likewise large-scale correlations at the continental slope, changes of the annual amplitude were argued to be dominated by boundary waves exerted by incident Rossby waves. Because we similarly find correlations beyond the 250 km length-scale, in particular along elongated coastal regions (Figure 1 a and b), we justify the larger 300 km radius.

We identify coherent zones of sea-level variability represented by different selection-criteria in Figure 1. The statistics S are computed based on individual along-track SLA time series (ALES) and GESLA tide gauges. We show different maps of these along-track statistics for (a) the Australian Coast, (b) Californian Coast and (c) Chichijima island (Japan). The contour in the first column exemplifies the extend of a ZOIs, which represents a sub-set of 20% of the best correlated data.

The obtained coherent structures reveal notable dependencies on the local bathymetric and coastal properties. Figure 1 (a), for instance, shows far-reaching alongshore correlations, which is supported by all of the analysed selection criteria. In this example, the separation of the region of the coastal shelf-sea dynamics from region of offshore variability is in good agreement with the underlying bathymetric gradients. Kurapov et al. (2016) found similarly pronounced SLA coherency along for the Californian coast, as shown in Figure 1b). Based on model data and tide gauge observations, they explained the large-scale along-shore correlation pattern in part with the propagation of coastal trapped waves. In other locations such as in Chichijima island (see Figure 1c), coastal and bathymetric control of SL is reduced and different structures of coherency evolve. Consequently, the ZOI can strongly vary in shape depending on the local coastal features and drivers of coastal variability.

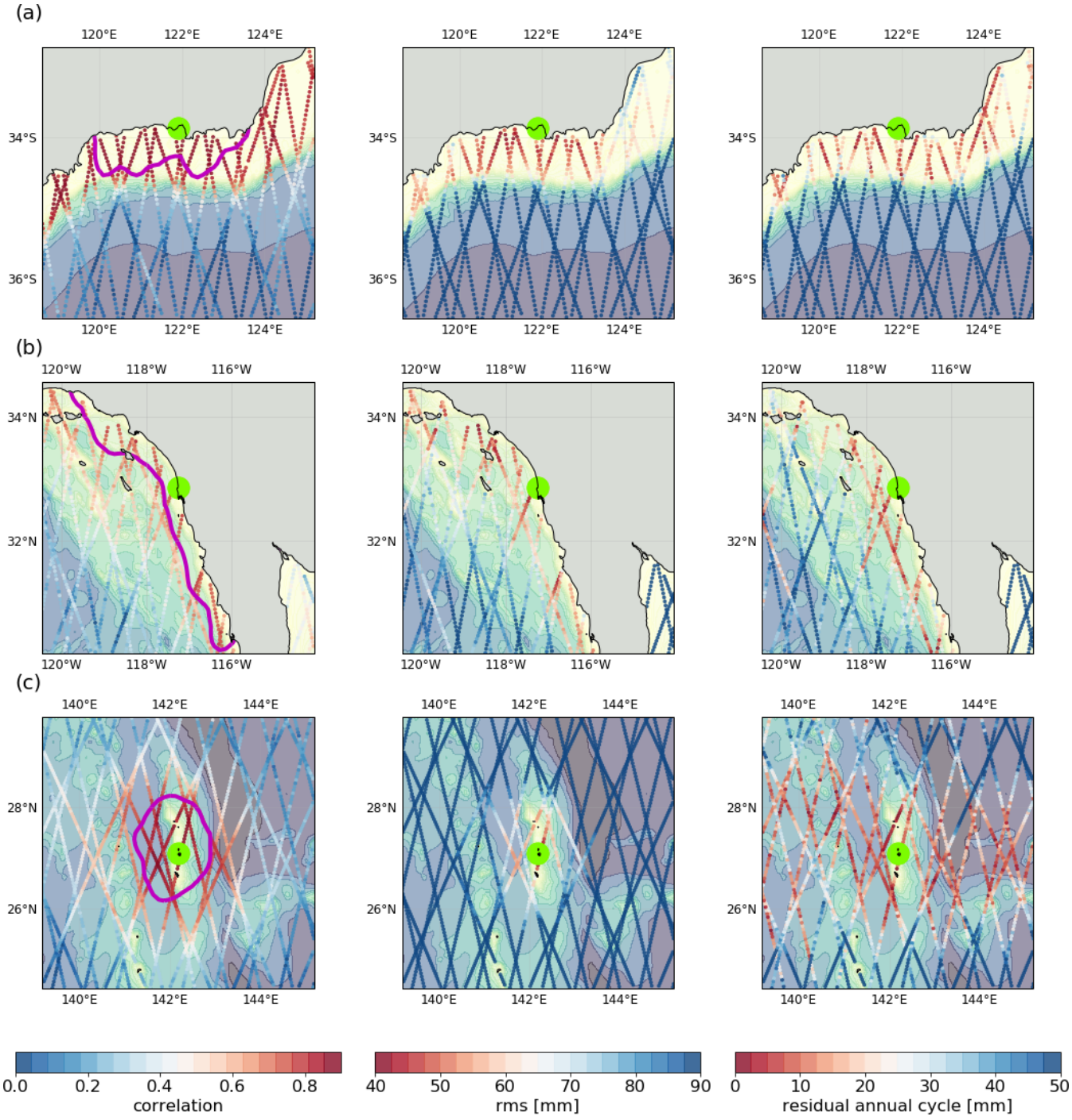


Figure 1. Zone of Influence: Different coherent zones of sea level variability are identified by different statistical criteria S. The columns show correlations, $\mathrm{RMS}_{\mathrm{SAT\text{-}TG}}$ and the residual annual cycle from left to right. The metrics are computed on every point of the 1 Hz along-track product, comparing the performance of altimetry measurements with the tide gauges, highlighted in green (center). (a) shows the South-Coast of Western Australia, (b) the western coast of North America (tide gauge in San Diego) and (c) Chichijima island (Japan). The 'color' contour in the first column indicates a Zone of Influence built from 20% of the best-correlated SLAs within a 300 km radius. The underlying contours denote the underlying bathymetry.

Comparing these three examples, we also observe that absolute values of the statistics differ from site to site. Correlations of along-track data near the Australian coastline, for instance, outperform the ones in example Figure 1b. The same holds for $RMS_{SAT\text{-}TG}$ values. These differences not only indicate different degrees of coherency, but can also stem from regional deviations in the quality of data, i.e. quality of tide gauge records or error sources in the altimetric product, such as tidal adjustments or coastal corrections. Differences can also be caused by coastal properties, e.g. when the tide gauges are located in sheltered areas, which separates the in situ variability from the one measured at distant altimeter tracks. We analyse therefore the use of relative thresholds, to select the SLAs, since setting absolute thresholds as in Kleinherenbrink et al. (2018) might not be applicable in all cases. Figure 1c) also shows that different statistics can determine different extents of the ZOI, considering that rather poorly correlated areas are partially characterised by low residual annual cycle amplitudes.

A correct choice of the ZOI based on a sub-set of high performance ~~performant~~ SLAs can significantly reduce the SAT-TG residuals as are exemplified in Figure 2. Here, we show three time series of SAT-TG differences for the Australian site (see Figure 1a). The first series (Figure 2a) indicates much lower residual noise, when the time series is constructed from the 20% best SLAs (according to the $RMS_{SAT\text{-}TG}$). Here, the ALES-GESLA-ZOI residuals outperform those of the other combinations ALES-PSMSL-250km and AVISO-PSMSL-250km, which are still affected by a pronounced annual cycle not related to VLM.

While using relative thresholds can reduce the noise of $VLM_{SAT\text{-}TG}$ time series for individual stations, we seek to identify a globally optimal ZOI definition and associated criteria and thresholds, which lead to largest improvements of uncertainty and accuracy of $VLM_{SAT\text{-}TG}$. Therefore, we vary the relative thresholds X between 0.0 and 0.975 (with a stepsize of 0.025), which refers to using 100% and 2.5% of the best performing SLAs according to each criteria. For each threshold and criterion we derive an individual global $VLM_{SAT\text{-}TG}$ trend and uncertainty dataset. We validate the performance of the trend estimates for a specific ZOI definition in accordance with section 3.4. Optimal parameter X, S are then suggested for the global application (section 4.2).

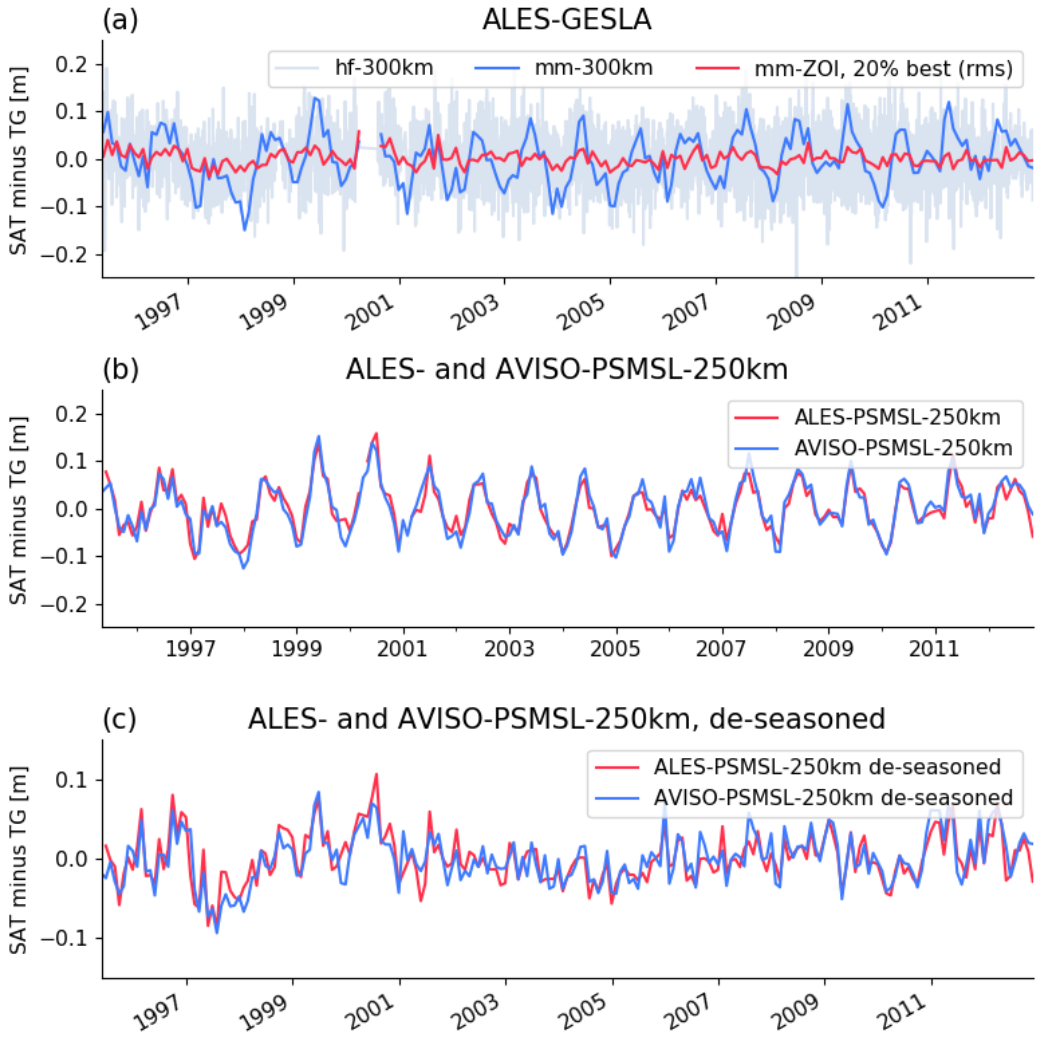


Figure 2. Shown are 'SAT minus TG' time series for different datasets and configurations for the TG in Figure 1a). (a) Monthly mean (mm) time series for ALES-GESLA, when all SLAs are averaged in a 300 km radius (blue) and when SLAs are comprised of the 20% most representative anomalies based on the $RMS_{SAT\text{-}TG}$ between altimetry and TG (red). The grey line denotes the underlying high-frequency time series. (b) Monthly mean differenced time series for ALES-PSMSL and AVISO-PSMSL, which are based on a 250km-radius selection of SLAs. (c) Same as (b) but with the annual and semi-annual signals removed.

3.3 Statistical analysis: Trend and uncertainty estimation

We fit the differenced time series to a combination of a deterministic model and stochastic noise models with the Maximum Likelihood Estimation (MLE) method. Parameters of the deterministic model are comprised of a constant offset A and a linear trend B. The annual and semi-annual signals are expressed by harmonic functions with the annual and semi-annual frequencies $\omega_{1,2}$ and amplitudes $C_{1,2}$ and $D_{1,2}$.

$$y(t) = A + Bt + \sum_{i=1}^{2} C_i cos\, 2\pi t\omega_i + D_i sin\, 2\pi t\omega_i \quad (1)$$

When combining altimetry and tide gauges for VLM estimation, several sources can contaminate the differenced time series and inflate the actual 'red'-noise (low-frequency) content in the residuals, which generates auto-correlated signals in the data. ~~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, to avoid underestimation of the uncertainty of the parameters, we take into account auto-correlation in the residuals of the detrended and deseasoned time series. We describe the power spectral density of the noise with a combination of a power-law and a white noise model (using the Hector software (Bos and Fernandes, 2019)). The power-law process assumes that time-correlated noise power is proportional to f^{κ}, which for negative spectral indices κ describes increasing power at lower frequencies f and a white-noise process when $\kappa = 0$ (Agnew, 1992). Santamaría-Gómez et al. (2011) showed that this combination (of power-law and white noise model) represents the best approximation of the noise content for 275 GNSS station position time series. This combination was also implemented in studies concerned with $VLM_{SAT\text{-}TG}$ estimation (WM16, Kleinherenbrink et al. (2018); Ballu et al. (2019)). In particular, the spectral index κ can contribute to detect the intrusion of low-frequency signals in the differenced time series. Next to the spectral index κ we estimate the individual fractions of the power-law and white noise models, as well as the total variance σ^2 which scales the amplitude of the noise. 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.

3.4 Validation of $VLM_{SAT\text{-}TG}$ with VLM_{GNSS} trends

To validate SAT-TG-based trend estimates, we use the ULR6a GPS solution provided by the GNSS data assembly centre SONEL (Systeme d'Observation du Niveau des Eaux Littorales, http://www.sonel.org). The reanalysis covers 19 years of GNSS data from 1995 to 2014, which are processed within the ITRF2008, consistent with the reference frame of altimetry orbits. The primary coordinates provided by GNSS are geocentric Cartesian coordinates (X, Y, Z, Vx, Vy, Vz). For the comparison with vertical trends inferred from other techniques, they are converted to ellipsoidal coordinates (latitude ϕ, longitude

λ and ellipsoidal height h, and V_ϕ. V_λ, V_h). Thus, we compare GNSS ellipsoidal height trends (V_h) with SAT-TG trends. It should be mentioned that, while the altimetry trends refer to the so-called TOPEX/Poseidon ellipsoid, the GNSS vertical trends refer to the GRS80 (Geodetic Reference System, 1980; Moritz (2000)) ellipsoid. Although there is difference of 70 cm between the semi-major axes of both ellipsoids, the GNSS and SAT vertical trends can be compared without degradation of precision, as both ellipsoids are geocentric and have the same orientation with respect to the Earth's body (e.g., the ellipsoid minor axes coincide with the mean Earth's rotation axis, and the major axes are on the Earth's equatorial plane). We take into account GNSS stations which are closer than 1 km to a tide gauge. With this constraint we aim to avoid potential differential vertical motions between the tide gauge and the GNSS-antenna (WM16).

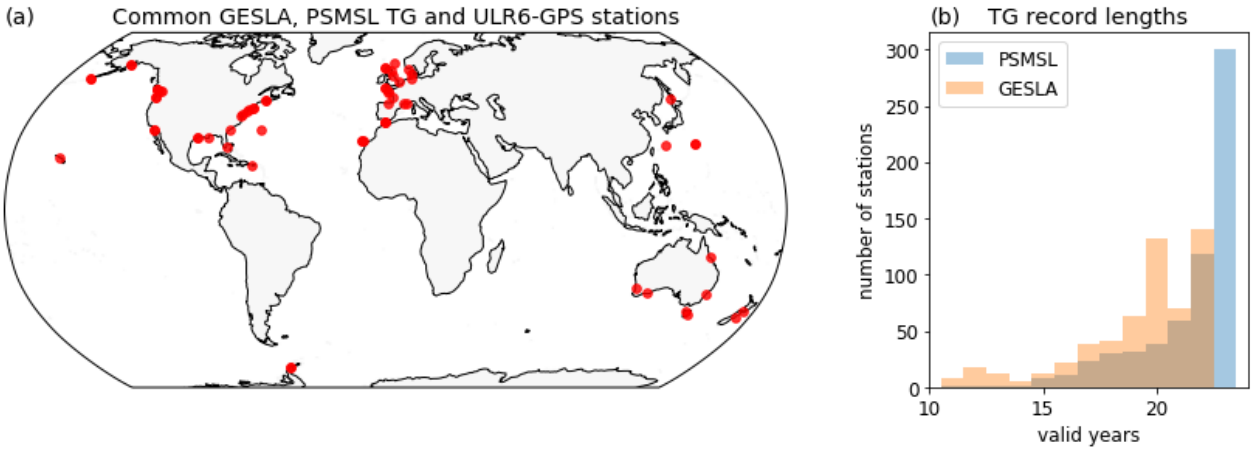


Figure 3. a) Global Distribution of 52 common GESLA, PSMSL and ULR6-GNSS stations, which meet the described requirements. b) Number of all TG stations sorted by the amount of months which contain valid data (here shown in valid years) in the period 1993-2015.

The tide gauge locations and record lengths differ among the presented experimental datasets (section 3.1). Therefore, we define several requirements for the validation of those experimental-datasets, to obtain a consistent set of TG and GNSS validation pairs. In contrast to PSMSL records GESLA-TG observations only last until 2015. Even when PSMSL TG records are limited to before 2015, they still contain more months of valid data than GESLA (see Figure 3b). Hence, we align the time period covered by the PSMSL-TGs to the corresponding GESLA-TGs for all following experimental datasets. Generally, we only take into account SAT-TG time series, when they cover at least 120 months of valid data. Note, that the outlier analysis (section 2.4) or coupling of high frequency TG data in the ZOI can reduce the length of the SAT-TG time series for GESLA TGs. Taking into account all these requirements, we obtain 52 common GESLA and PSMSL TGs, which provide a neighbouring GNSS station within 1 km distance. These pairs are validated for ALES-PSMSL-250km, ALES-GESLA-250km and AVISO-PSMSL-250km. The ALES-GESLA-ZOI combination includes six more stations. The resulting validation pairs are shown in Figure 3a), where a higher coverage in northerly and midlatitude regions is evident.

We compute the $RMS_{\Delta VLM}$ and the median of the differences (ΔVLM) of $VLM_{SAT\text{-}TG}$ and VLM_{GNSS} for a given dataset combination. To take into account the derived formal errors (U) of the estimate we compute the weighted RMS_w as follows:

$$RMS_w = \sqrt{\sum_{i=1}^{n}(w_i(VLM_{GNSS_i} - VLM_{SATTG_i})^2)} \tag{2}$$

with weights

$$w_i = \frac{\sqrt{(U^2_{GNSS_i} + U^2_{SATTG_i})}^{-1}}{\sum_{i=1}^{n}(\sqrt{(U^2_{GNSS_i} + U^2_{SATTG_i})}^{-1})} \tag{3}$$

We analyse also the median of the absolute value of differences (|ΔVLM|). This metric is less prone to extreme deviations and can thus consolidate the evaluation of the dataset performances. We generally assume that GNSS provides a more accurate estimation of the linear component of the VLM with a smaller error than $VLM_{SAT\text{-}TG}$, ~~also~~ despite shorter time span of measurements. Hence, for the purposes of this paper and as done in all studies concerning $VLM_{SAT\text{-}TG}$ estimation, we define as measures of accuracy the $RMS_{\Delta VLM}$ and additionally the median of |ΔVLM|. We include the spectral index κ (see section 3.3) as it helps to understand the level of auto-correlation of the time series. All statistics other than the $RMS_{\Delta VLM}$ denote median values (of all $VLM_{SAT\text{-}TG}$ estimates) for a specific dataset configuration.

4 Results

4.1 Comparison of different datasets configurations based on a 250 km average selection

We compare performances of the three datasets which are constructed from 250 km radius SLA averages in Table 3. Validation against GNSS vertical velocities reveals that the gridded combination AVISO-PSMSL-250km slightly outperforms ALES-PSMSL-250km in terms of accuracy. Both the $RMS_{\Delta VLM}$ and the median of absolute trend differences are ~~8~~9% lower for AVISO-PSMSL-250km. This confirms that, if all the available altimetry data within a wide region are compared against monthly values of TGs, the use of a gridded product outperforms the along-track performances (WM16). Kleinherenbrink et al. (2018) similarly compared an along-track combination of 250 km-SLA averages (from RADS) and PSMSL tide gauge data with the AVISO-PSMSL combination from WM16. They found a small $RMS_{\Delta VLM}$ reduction of 0.1 mm/year when using the along-track product without any correlation thresholds applied. WM16's trends were however based on 1° radius-averages of SLAs (in contrast to the 250 km selection), and record lengths were not equalized as in this study.

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 et al., 2015). Still, the observed $RMS_{\Delta VLM}$ of AVISO-PSMSL-250km (1.50 mm/year) is comparable to WM16 result (1.47 mm/year). In contrast to trend accuracies, the uncertainties are 5% lower for ALES-PSMSL-250km than for AVISO-PSMSL-250km. As in WM16, the spectral index κ of the interpolated gridded product is lower than for the along-track data. Both κ indices (-0.56 and ~~-0.42~~ -0.39) also match well those found by WM16 for AVISO (-0.5) and the along-track product (-0.4, GSFC). The larger spectral index (~~-0.42~~ -0.39) is associated with reduced power of the noise at low frequencies and thus indicates reduced contamination of the SLA signal by sea-level variations that do not represent those measured at the tide gauge. This enhanced comparability is also reflected in the lower trend uncertainties of ALES-PSMSL-250km (0.69 mm/year) compared to AVISO-PSMSL-250km (0.73 mm/year). The differences between the characteristics of the residuals of the datasets can partially be explained by the resolution of the data: Due to the spatial filtering of the data, the gridded solution AVISO incorporates information of SLAs beyond the 250 km radius and thus contains time-correlated SL-signals with stronger deviations ~~which stronger deviate~~ from the TG records.

Table 2. ~~Statistics of different SAT-TG combinations. ΔVLM refers to the differences of and trends. *X* denotes the relative level of comparability above which data is included.~~

X	$RMS_{\Delta VLM}$ mm/year	weighted $RMS_{\Delta VLM}$ mm/year	med. |ΔVLM| mm/year	med. ΔVLM mm/year	med. uncertainty	spectral index κ
			ALES-PSMSL-250km (52 stations)			
	1.63	1.54	1.22	-0.51	0.69	-0.42
			AVISO-PSMSL-250km (52 stations)			
	1.50	1.48	1.12	0.56	0.73	-0.56
			ALES-GESLA-250km (52 stations)			
	1.46	1.45	1.13	-0.22	0.79	-0.41
			ALES-GESLA-ZOI (best RMS_{SAT-TG}, 58 stations)			
0	1.52	1.41	1.00	-0.46	0.85	-0.47
0.1	1.38	1.34	0.88	-0.27	0.87	-0.44
0.2	1.34	1.32	0.91	-0.38	0.82	-0.45
0.3	1.31	1.32	0.83	-0.41	0.77	-0.47
0.4	1.31	1.31	0.84	-0.40	0.77	-0.45
0.5	1.30	1.31	0.84	-0.23	0.73	-0.46
0.6	1.30	1.29	0.83	-0.30	0.70	-0.49
0.7	1.28	1.27	0.83	-0.43	0.67	-0.47
0.8	1.28	1.28	0.87	-0.50	0.59	-0.44
0.9	1.51	1.47	0.90	-0.44	0.57	-0.43

Table 3. Statistics of different SAT-TG combinations. ΔVLM refers to the differences of $VLM_{SAT\text{-}TG}$ and VLM_{GNSS} trends. X denotes the relative level of comparability above which data is included.

X	$RMS_{\Delta VLM}$ mm/year	weighted $RMS_{\Delta VLM}$ mm/year	med. |ΔVLM| mm/year	med. ΔVLM mm/year	med. uncertainty	spectral index κ
			ALES-PSMSL-250km (52 stations)			
	1.68	1.57	1.28	-0.87	0.69	-0.39
			AVISO-PSMSL-250km (52 stations)			
	1.50	1.48	1.12	0.56	0.73	-0.56
			ALES-GESLA-250km (52 stations)			
	1.51	1.47	1.14	-0.39	0.79	-0.39
			ALES-GESLA-ZOI (best $RMS_{SAT\text{-}TG}$, 58 stations)			
0	1.54	1.45	0.98	-0.46	0.86	-0.45
0.1	1.39	1.36	0.9	-0.27	0.86	-0.44
0.2	1.34	1.33	0.88	-0.36	0.83	-0.47
0.3	1.32	1.36	0.83	-0.44	0.78	-0.46
0.4	1.3	1.38	0.87	-0.37	0.76	-0.45
0.5	1.29	1.4	0.86	-0.26	0.73	-0.47
0.6	1.3	1.43	0.87	-0.31	0.71	-0.47
0.7	1.28	1.39	0.82	-0.41	0.66	-0.48
0.8	1.28	1.37	0.86	-0.41	0.58	-0.43
0.9	1.53	1.58	0.97	-0.43	0.61	-0.46

In comparison with the low-frequency datasets (ALES-PSMSL-250km and AVISO-PSMSL-250km), the high-rate set-up ALES-GESLA-250km improves the $RMS_{\Delta VLM}$. The absolute bias of trend differences decreases more substantially to 0.39 mm/year (compared to -0.87 mm/year) ~~and is with -0.22 mm/year slightly lower than WM16's estimate of 0.25 mm/year~~. 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. 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. Therefore ~~Given the strong improvement in the bias,~~ the ALES-GESLA coupling is further developed based on a better definition of the ZOI in the next section.

4.2 The Zone of influence improves VLM estimates

We investigate how the ZOI selection of SLAs fosters quality of SAT-TG VLM estimates. As addressed in section 3.2, we build the ZOI upon different criteria of comparability: $RMS_{SAT\text{-}TG}$, correlation and the residual annual cycle. First, we focus

on the results of using the $RMS_{SAT\text{-}TG}$ of the detrended differenced time series (Table 3 and Figure 4, ALES-GESLA-ZOI). We observe that the $RMS_{\Delta VLM}$, the median of absolute and total differences, as well as trend uncertainties decrease towards higher relative thresholds. The statistics converge to a minimum when the ZOI is restricted to the 30-20% best data. To compare ALES-GESLA-ZOI with the other dataset combinations, we compute the statistics for the same 52 tide gauges used in these configurations (because the shown statistic in Table 3 refer to a larger set of 58 stations). At the 20% thresholds, we obtain similar performances with a $RMS_{\Delta VLM}$ of 1.29 mm/year, median uncertainty of ~~0.57~~0.51 mm/year and a median of absolute differences ($|\Delta VLM|$) of 0.86 mm/year. Thus, the improvements of $RMS_{\Delta VLM}$ compared to the plain 250 km-radius selection (ALES-GESLA-250km) is ~~12~~15% and ~~28~~35% for uncertainties. Hence, we find more substantial, nearly linear reductions of trend uncertainty with increasing relative thresholds compared to trend accuracy ($RMS_{\Delta VLM}$, Table 3, ALES-GESLA-ZOI). As demonstrated for different time series in Figure 2, selecting e.g. highly correlated SLAs efficiently reduces the noise of the residuals. Correspondingly, at higher levels of comparability, the variance, which scales the amplitudes of the considered noise models, decreases (not shown).

Because the spectral index (for ALES-GESLA-ZOI) is slightly lower (~~-0.44~~-0.43 at 20% level) than for ALES-GESLA-250km (~~-0.41~~-0.39) it cannot account for the uncertainty improvements. Here, the lower κ index reveals an relative increase of power at low frequency (i.e. time scales longer than months). Thus the bulk of improvements we see in uncertainty (comparing ALES-GESLA-ZOI and ALES-GESLA-250km) stems from the reduction of the power law and white noise amplitudes in the residuals. This is in turn caused by improvements of the comparability of tide gauge and altimetry measurements at high-frequency (i.e. days). We argue that extending the maximal radius selection from 250 km to 300 km to construct the ZOI (as done for ALES-GESLA-ZOI) increases the low-frequency noise (indicated by κ). However, with this selection we capture more altimetry tracks with similar highly correlated high-frequency signals (see Figure 1), which again contribute to sampling density and reduced white noise. This further substantiates our choice to select SLA within a larger 300 km radius, which is also supported by observed larger-scale coherency of coastal sea level trends (see section 3.2).

$RMS_{\Delta VLM}$ and trend uncertainty 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 ~~closely~~ 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 accuracy ($RMS_{\Delta VLM}$). ~~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 VLM SAT-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.

When setting this optimal threshold to 20%, the ALES-GESLA-ZOI set-up outperforms the other investigated configurations. Figure 4 compares the scatter of estimated $VLM_{SAT\text{-}TG}$ against GNSS trends of all datasets. For ALES-GESLA-ZOI, we find lower $VLM_{SAT\text{-}TG}$ trend uncertainties and reduced spread of the estimates with respect to the 1:1 line ~~bisector~~ (Figure 4). ~~This result proves the importance of using such a refined selection procedure (ZOI), as this approach outstrips the improvements (in terms of trend accuracy and uncertainty) induced by the different altimeter or tide gauge data combinations.~~ 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.

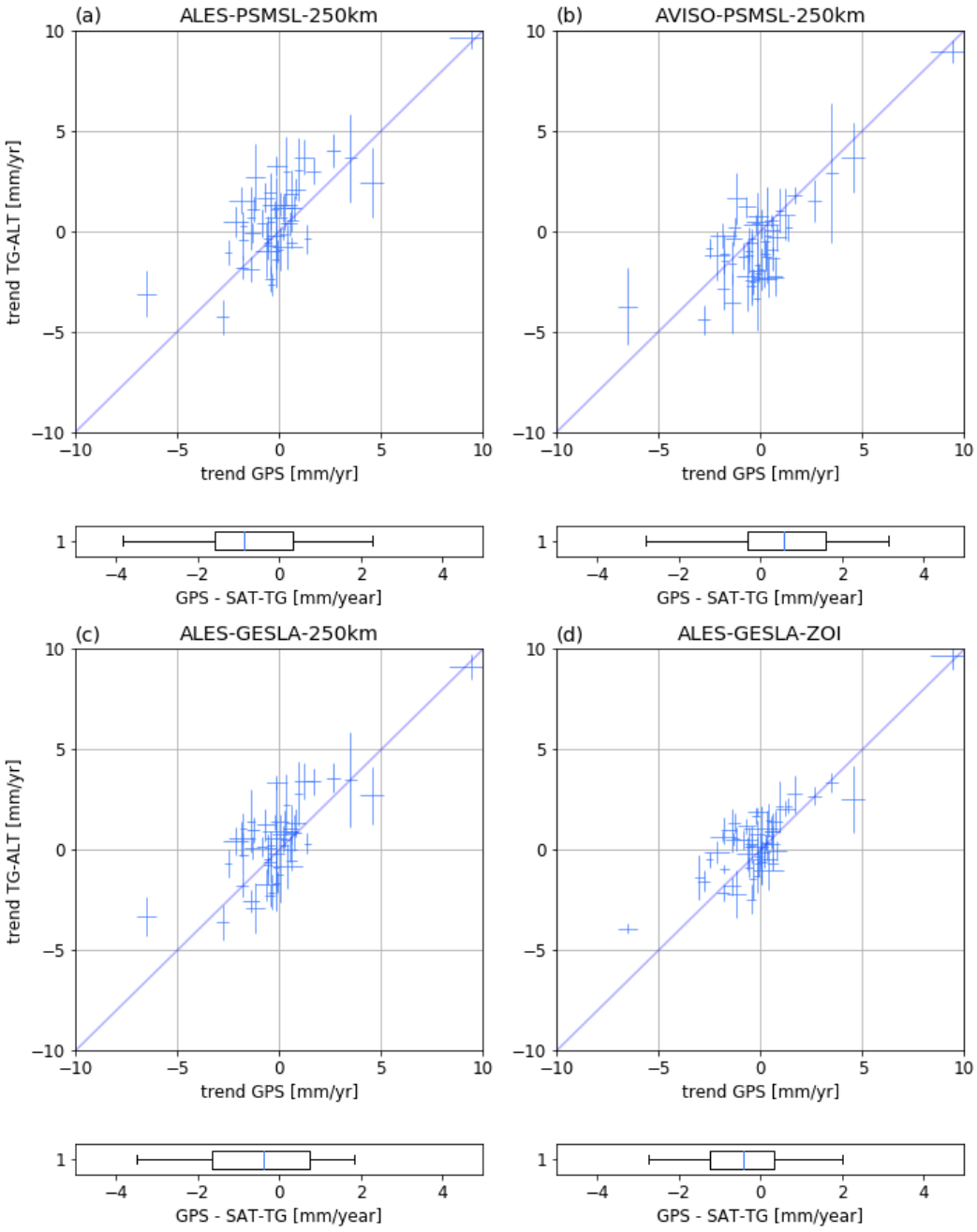


Figure 4. Scatter and boxplots compare estimated SAT-TG trends and GNSS trends, as in WM16 Figure 14. a) ALES-PSMSL-250km b) AVISO-PSMSL-250km c) ALES-GESLA-250km d) ALES-GESLA-ZOI (at 20% threshold based on RMS-criterion). Error bars denote the 1 sigma trend uncertainties of the individual estimates.

Figures 5a and 5c illustrate the influence of applying different criteria on the performance of estimated trends. Generally, increasing relative $RMS_{SAT\text{-}TG}$ or correlation thresholds yields similar optimal ranges ($\sim 20\%$) for both $RMS_{\Delta VLM}$ or uncertainty of $VLM_{SAT\text{-}TG}$ trends and can thus be interchangeably used. At lower relative threshold levels (20-60%), however, application of the RMS-criterion yields slightly reduced $RMS_{\Delta VLM}$ values compared to correlations. Hence, for this set of tide gauges a SLA-selection based on the minimum $RMS_{SAT\text{-}TG}$ generally provides more accurate trend estimates (in terms of $RMS_{\Delta VLM}$). The ~~criterion~~ residual annual cycle criterion only weakly reproduces the improvements provided by the other criteria and is less suited to confine the ZOI. This observation emphasizes the need of matching the data according to the high-frequency comparability (RMS, correlation), because selecting the data based on the residual annual cycle (i.e. low frequency comparability), limits the performance of the estimates (Figure 5c). 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.

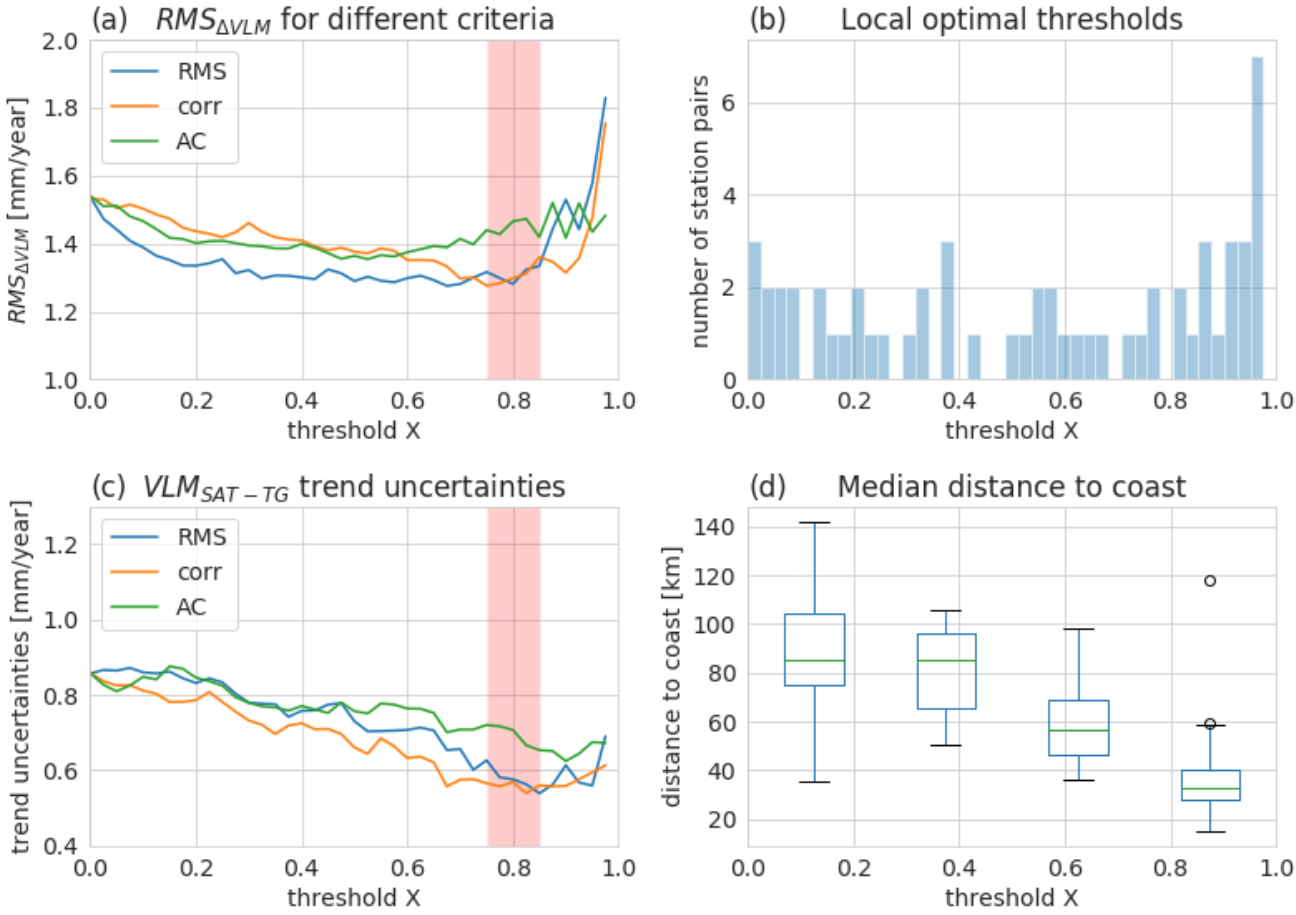


Figure 5. Performance of $VLM_{SAT\text{-}TG}$ trend estimates for ALES-GESLA-ZOI. a) $RMS_{\Delta VLM}$ for different relative thresholds (step size 2.5%) and different selection criteria: $RMS_{SAT\text{-}TG}$ (blue), correlation (red) and residual annual cycle (AC, green); c) same as (a) but for median uncertainty. b) Distribution of best performing relative thresholds for individual stations. The local optimal threshold is defined at the minimum of the absolute difference of $VLM_{SAT\text{-}TG}$ 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.

5 Discussion

The integration of the ZOI primarily reduces the uncertainty of $VLM_{SAT\text{-}TG}$ trend estimates. Over a considerable range of thresholds (80 - 20% of best performing data) trend accuracies do not improve as strongly as the uncertainties decrease. This is in line with Kleinherenbrink et al. (2018), who showed that for a highly correlated sub-set of TGs, increasing absolute correlation thresholds would not significantly reduce the $RMS_{\Delta VLM}$. We thus strive to understand better why trend estimates do not always improve when selecting highly comparable (w.r.t. tide gauge) or closely located absolute SLA measurements. This question ultimately leads to the discussion of the importance of identifying the small-scale dynamical components of local sea-level variability, given that long-term absolute sea level trends are large-scale signals.

5.1 Space and time dependencies of coastal sea level trends

The results presented in Figure 5 and Table 3 denote metrics and performances derived from the global TG-GNSS dataset for ALES-GESLA-ZOI ~~average 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 $VLM_{SAT\text{-}TG}$ 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 $RMS_{\Delta VLM}$. 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. Overall, the optimal levels X are broadly distributed from 0 to 0.975. We find highest concentrations between 0.8-0.975, which slightly exceeds the range of the global optimum. 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 ~~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).

In contrast to these examples, we find very low local optima for some stations (Figure 5b). Here, local $VLM_{SAT\text{-}TG}$ and GNSS trend differences do not converge to a minimum when increasing the comparability of SAT and TG observations. Accordingly, in these cases, vertical land motion estimates do not necessarily benefit from high coastal resolution of the data, because a low relative threshold is simultaneously linked to a larger-scale selection of SLAs (Figure 5d). At the lower level ranges, for instance at 0-0.2, SLAs have an average distance of 95 km to the coast. Supposing that the sources of these larger scales of coherency

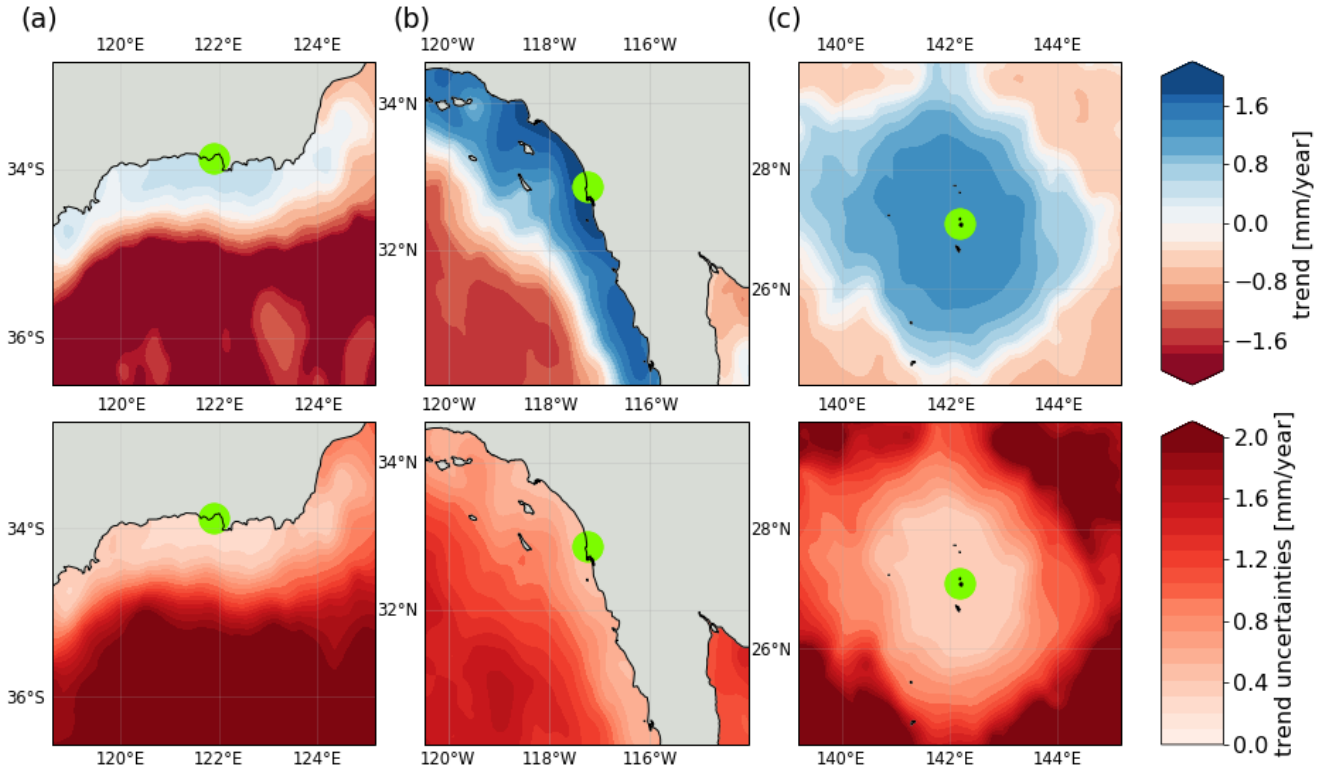


Figure 6. $VLM_{SAT\text{-}TG}$ trends (first row) and uncertainties (second row) mapped onto relative correlation levels. The mapping and interpolation method is further elucidated in the Appendix A. We show the same stations (a,b,c) as in Figure 1.

of coastal SL trends would be known, a more advanced adaption to these additional factors would further increase accuracy of VLM estimates. An associated ideal selection of trends, based on optimal individual levels shown in Figure 5d would largely reduce to $RMS_{\Delta VLM}$ 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.

To further shed light on the relationships between dynamical-sea-level-based SLA selection and spacial coherence of trends and uncertainties, we show trend and uncertainty maps (Figure 6) in accordance with those in Figure 1 (displaying maps of statistical criteria). Here, we map linear trends and uncertainties onto observed levels of comparability defined by the correlation-criterion. We thus compute these $VLM_{SAT\text{-}TG}$ trends over different coastal regions (see further details in the Appendix A). As a result, we observe sharp trend gradients consistent with the degree of comparability: Figure 6a for instance, shows high small-scale variability of trends, because trends off the slope-current-region are detached from the trends in the along-shore continental shelf region. Trends in Figure 6a and 6b project onto the far-reaching along-shore correlations as ob-

served in Figure 1 showing consistent signals over several hundreds of kilometers along the coast. Uncertainty maps further pronounce the importance of the application of highly resolved coastal altimetry data (Figure 6 lower row). These examples show, that at individual locations the ~~The~~ use of less comparable SLAs can increase the uncertainty by a factor of three to four. Therefore, for the majority of cases, these results promote using high relative levels of comparability to define the ZOI for trend estimation. However, we also observe that the coherency of trends (highlighted by the strength of absolute trend gradients) can be differently expressed at different coastal regions.

Bathymetric and coastal properties can cause large discrepancies in responses of coastal sea level variability as they modify the character of the impact of large-scale atmospheric forcing or remote variability from the deeper ocean (Woodworth et al., 2019). Hence, an advanced analysis of SL coherency and the role of bathymetry might facilitate further enhancements of trend accuracy based on SAT and TG. We note~~,~~ that physical origins might, however, not necessarily cause the spread of individual optimal thresholds (Figure 5b). If our assumption, that GNSS-trend estimates perfectly represent the linear trend over the time span of the altimetry/tide gauge records was not met, the shown individual thresholds would erroneously reflect local optima. Ruling out these sources of error is thus a prerequisite to further study physical explanations for different extents of the ZOI.

Next to site-dependent physical factors, the spatial-scales of trend coherency might also depend on the time span of the observations themselves. Global maps of sea level trends, for example, even when derived from two decades of observations, still show distinct pattern of natural/forced variability and thus shade signals of ocean mass or steric contributions (e.g. Stammer et al. (2013)). Similarly, coastal sea level trends that are computed in the ZOI are affected by local interannual sea level variability on top of the secular trend. Therefore, the importance to adopt the concept of the ZOI for improving trend accuracy might also be influenced by the actual time span covered by the record.

To investigate this time-scale-dependency, we truncate the $\mathrm{VLM}_{\mathrm{SAT\text{-}TG}}$ time series such that we obtain different experimental ALES-GESLA-ZOI sets with maximum record lengths from 10 to 18 years. We repeat the same validation analysis against GNSS trends as in section 4. Figure 7a) encompasses anomalies of the $\mathrm{RMS}_{\Delta\mathrm{VLM}}$ with respect to the mean $\mathrm{RMS}_{\Delta\mathrm{VLM}}$ for a dataset of a specific time scale which is given in Figure 7b (red). The same evolution is shown for trend uncertainties in Figure 7c).

Mean $\mathrm{RMS}_{\Delta\mathrm{VLM}}$ as well as mean uncertainties (which are averaged over all relative thresholds for a specific maximum record length) substantially decrease with increasing record length (Figures 7a and 7c). Both statistics approximately follow the theoretical proportionality of uncertainty and sample size n of $1/\sqrt{n}$ (assuming no serial correlation). The evolution of the $\mathrm{RMS}_{\Delta\mathrm{VLM}}$ anomaly shows that selecting SLAs in a ZOI at high relative thresholds more substantially reduces the $\mathrm{RMS}_{\Delta\mathrm{VLM}}$ on shorter time scales (e.g. 10 years) than on longer time scales (Figure 7a, e.g. at 18 years). At long time scales, the $\mathrm{RMS}_{\Delta\mathrm{VLM}}$ anomalies do not significantly improve between the 80% and 20% thresholds, which we also observe in the previous analysis in Figure 5a. We argue that the transition time scale where the improvements of $\mathrm{RMS}_{\Delta\mathrm{VLM}}$ flatten (14-16 years), marks when the high frequency coastal sea level dynamical variability is superseded by dynamics producing large-scale sea level trends. In other words, this is the time scale in which coastal sea level trends start to merge with the offshore trends. The tendency of increasing spatial scales with time is also reflected by the increasing distances to coast of the measurements for an optimal ZOI at a specific time scale (Figure 7b). The time-scale-dependency could explain the mismatch of trend accuracy and uncertainty

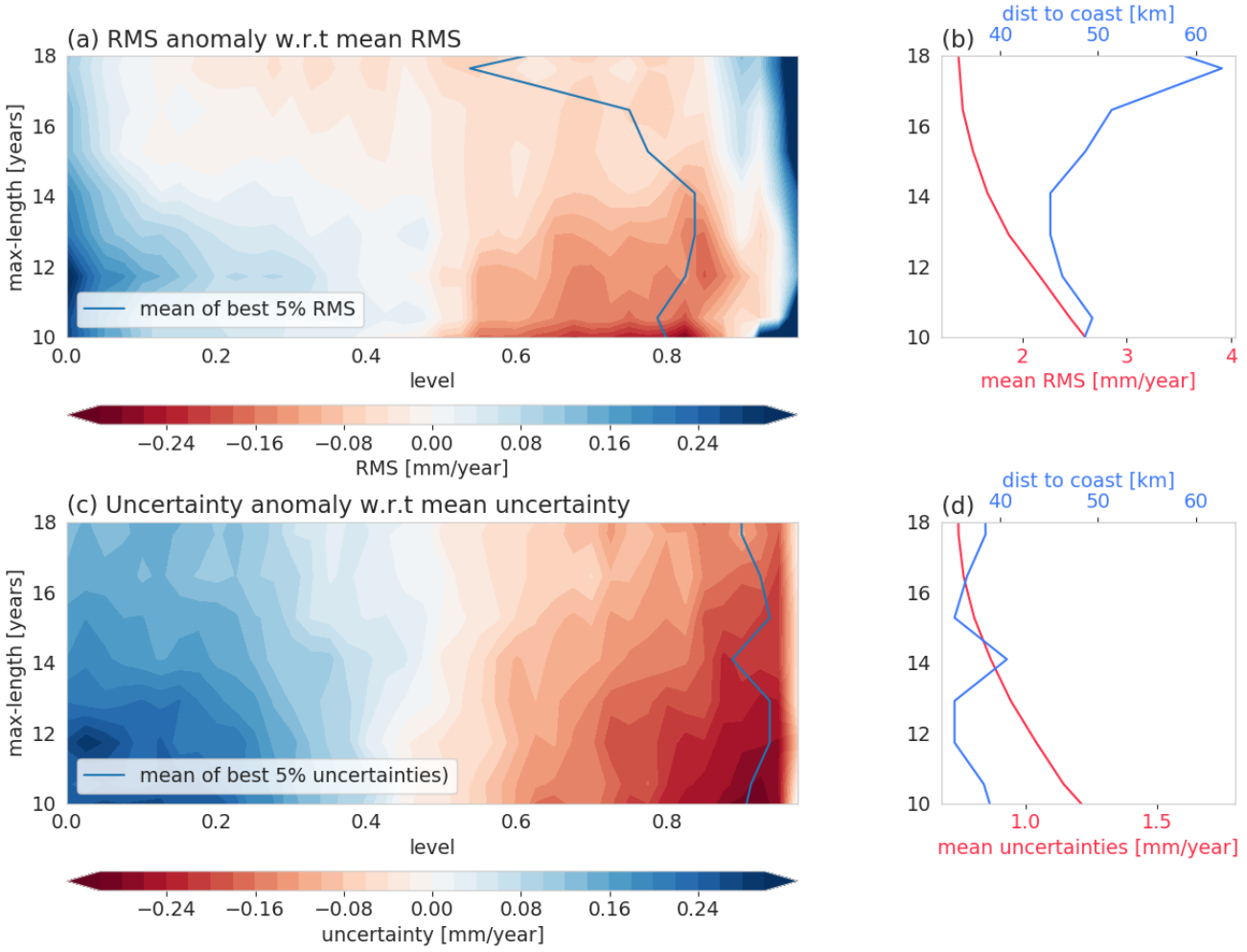


Figure 7. Time and space dependencies of trend uncertainty and accuracys: a) Evolution of the $RMS_{\Delta VLM}$ anomaly (SAT-TG vs. GNSS trend) for subsets of ALES-GESLA-ZOI, depending on a relative threshold X (x-axis) and a maximum record length (y-axis). $RMS_{\Delta VLM}$ anomaly is defined as the departure from the mean $RMS_{\Delta VLM}$ (shown in b) averaged over all thresholds X for a specific maximum record length. In b) we also show in blue the mean distance to coast of the measurements, associated with the average of the best 5% ZOI-levels per time scale, shown in a). c,d) Same as a,b) but for uncertainties.

improvements when using higher levels of comparability. This is also supported by Kleinherenbrink et al. (2018), who showed little sensitivity for SAT-TG combinations which had minimum lengths of 15 years.

The same evaluation for the dependency of uncertainty on time and level of comparability X demonstrates that using the ZOI nearly constantly improves trend uncertainties at any time scale. Hence, even though spatial scales of trend coherency might increase with time, an ideal match of altimetry and tide gauges should be based on a ZOI.

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., 2019; 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 to altimeter bias drift, nonlinear VLM from contemporary mass redistribution (CMR) changes were shown to cause differences between $VLM_{SAT\text{-}TG}$ 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 trend on 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 $VLM_{SAT\text{-}TG}$ and VLM_{GPS}. They expose that $VLM_{SAT\text{-}TG}$ 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.

5.3 Comparison with previous results

Based on optimal relative thresholds, we estimated an $RMS_{\Delta VLM}$ between SAT-TG and GNSS trends of 1.28 mm/year and a median uncertainty of ~~0.59~~0.58 mm/year at 58 sites. Our approach of combining along-track altimetry, high frequency tide gauge data and a refined SLA selection-scheme improves the performance of VLM estimation compared to using gridded

altimetry products and constant spatial SLA averages (WM16: $RMS_{\Delta VLM}$: 1.47 mm/year; uncertainty 0.80 mm/year). Other studies further emphasized the importance of spatial resolution in coastal zones, considering the decreasing temporal and spatial scales of sea level variability in such areas. With the focus on coastal sea level trends, Cipollini et al. (2017) demonstrated that the along-track X-TRACK product contained much more valid data close to the coast than AVISO, not only due to the spatial down-sampling in AVISO, but also to less adapted coastal processing. Here, we tackle both issues by implementing an advanced coastal along-track altimetry product. Because we find that much of the observed high-performing altimetry data has a close vicinity to the coast, our results underpin that along-track data is the best choice for coastal sea level trend estimation, and substantiate the results of Kleinherenbrink et al. (2018).

Accuracies of estimated $VLM_{SAT\text{-}TG}$ expressed by $RMS_{\Delta VLM}$ are in the order of Kleinherenbrink et al. (2018)'s result of 1.20 mm/year. These results can, however, not unequivocally compared due to different validation settings. We extend their analysis by investigating a variety of other criteria of comparability and find that the $RMS_{SAT\text{-}TG}$ of the differenced $VLM_{SAT\text{-}TG}$ time series provides the most robust estimates compared to correlations or residual annual cycle. Our results also propose that increasing the radius of selection denotes another improvement for VLM estimates. Practically, the approach of using absolute thresholds, which was put forward by Kleinherenbrink et al. (2018) almost halved the number of considered stations from 294 to 155, when setting an absolute correlation threshold to 0.7. Applying relative thresholds, facilitates the estimation of trends at lower correlated stations, which would be rejected otherwise. This is crucial, because it was frequently shown, that correlations between altimetry and tide gauges are highly variable across the globe (WM16). Hence, we maintain the main advantage of using tide gauges for VLM estimation: The large global distribution compared to continuous GNSS-measurements.

6 Conclusions

We investigate potential improvements of combining altimetry and tide gauges for coastal vertical land motion estimation. The innovations of our approach are twofold: (1) For the first time, we exploit a global network of high frequency tide gauge data (GESLA) and dedicated coastal altimetry (ALES) to determine VLM at a variety of co-located GNSS stations. Secondly (2) we define a Zone of influence, to identify coherent zones of coastal SL variability which optimizes the combination of altimetry and tide gauges. We rate improvements of both innovations against various SAT-TG datasets, which are comprised of along-track and gridded altimetry, as well as high (daily) and low-frequency (monthly) tide gauge combinations.

Combining high frequency tide gauge with coastal altimetry data (ALES-GESLA-250km) yields modest improvements of trend accuracies, compared to a monthly gridded or monthly along-track combination, when averaging SLAs in a radius of 250 km. The high spatio-temporal resolution of the data, however, provides the foundation to identify coherent zones of sea level variability. We confine a Zones of Influence by using relative thresholds of comparability based on $RMS_{SAT\text{-}TG}$, correlation and residual annual cycle of the altimetry and tide gauge timeseries. We identify a global optimal threshold, when selecting 20% of the data with the lowest $RMS_{SAT\text{-}TG}$. At this threshold, validation against GNSS velocity estimates (at 58 stations) yields a $RMS_{\Delta VLM}$ of $VLM_{SAT\text{-}TG}$ and VLM_{GNSS} differences of 1.28 mm/year with a median formal uncertainty of $VLM_{SAT\text{-}TG}$ trends of ~~0.59~~0.58 mm/year. This refined selection method improves trend accuracy by ~~12~~15% and uncertainty

by ~~28~~35% compared to the 250 km-average selection. The smaller degree of improvements of trend accuracy compared to uncertainty is explained by the increasing space-scales of sea-level trend components with progressing time scales. We show that in many cases, capturing small scale features of coastal sea level variability within few tens of kilometers from the coast is vital for VLM_{SAT-TG} estimation and constantly reduces trend uncertainty of the estimates. We thus promote using relative levels of comparability and dedicated coastal ~~dedicated~~ altimetry matched with high-frequency tide gauges to confine ZOIs~~,~~ and increase the number of VLM estimations along the global coastline with improved ~~and decrease their~~ uncertainty.

Data availability. ULR6a GNSS trend estimates are obtained from the data assembly centre SONEL (Systeme d'Observation du Niveau des Eaux Littorales, https://www.sonel.org/-Vertical-land-movement-estimate-.html?lang=en, Santamaría-Gómez et al. (2016)). GESLA tide gauge data are available at http://www.gesla.org (Woodworth et al., 2016) and PSMSL data at https://www.psmsl.org/data/obtaining/ (Holgate et al., 2013). ALES along-track data are processed at DGFI-TUM (https://www.dgfi.tum.de/en/) with OpenADB (https://openadb.dgfi.tum.de). Averaged DT-MSLA AVISO gridded altimetry data are obtained from https://www.aviso.altimetry.fr.

Appendix A: Methods

In Figure 6 we map linear VLM_{SAT-TG} trends and uncertainties onto observed levels of comparability set by the correlation-criterion. First, we group observed SLA time series in 0-20th, 20-40th, 40-60th etc. percentile-ranges, sorted by their correlations with the TG time series. Then, we merge the altimetry time series for each group and calculate their associated VLM_{SAT-TG} trends. The resulting VLM_{SAT-TG} trends are hereinafter defined on the altimetry tracks, categorized by the aforementioned groups of comparability. To better illustrate the different zones of coherency the trends are interpolated onto a regular grid (100x100 nodes i.e. 6 km resolution) and thus smoothed as seen in Figure 6. We use linear radial basis functions to interpolate the data.

Appendix B: Significance of median biases

To gain a better understanding of when the VLM_{SAT-TG} and VLM_{GNSS} difference distributions are significantly biased we create a Monte-Carlo experiment to check the H_0 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 20000 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 B1 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.

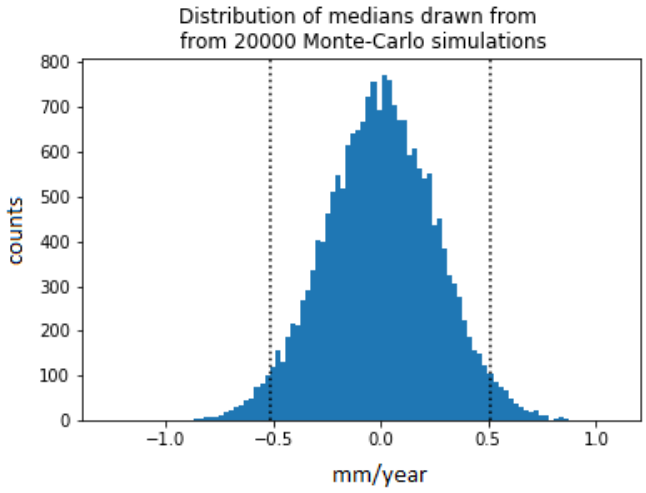


Figure B1. Histogram of median values of randomly sampled sub-sets. The sub-sets consists of 52 samples (according to the number of TGs in AVISO-PSMSL-250km) and are randomly drawn from normally distributed values with zero mean and a standard-deviation of 1.5 mm/year (according to the RMS of AVISO-PSMSL-250km).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

Author contributions. J.O. and M.P. conceptualized and designed the study. J.O. wrote the manuscript and is the author of the full software code used in this study. M.P. is the author of the ALES retracking algorithm and mentored the work of J.O.; C.S. and D.D. are responsible for the altimetry database organisation and the data structure. L.S. provided assistance in the use of GNSS data. F.S. provided the basic resources making the study possible and coordinates the activities of the institute. All authors read and commented on the final manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was funded by the Deutsche Forschungsgemeinschaft (DFG) (grand agreement 411072120) and the Technical University of Munich (TUM) in the framework of the Open Access Publishing Program. We thank the data-providers GESLA, PSMSL, SONEL and AVISO for the opportunity to use their products. We thank Sergiy Rudenko, Ashwita Chouksey and Michael Hart-Davis for their help and comments. We are very grateful for the comments of the reviewers Alvaro Santamaría Gómez and Christopher Watson, which strongly improved the manuscript.

References

Ablain, M., Cazenave, A., Valladeau, G., and Guinehut, S.: A new assessment of the error budget of global mean sea level rate estimated by satellite altimetry over 1993–2008, Ocean Science, 5, 193–201, https://doi.org/10.5194/os-5-193-2009, https://os.copernicus.org/articles/5/193/2009/, 2009.

Agnew, D. C.: The time-domain behavior of power-law noises, Geophysical Research Letters, 19, 333–336, https://doi.org/10.1029/91GL02832, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/91GL02832, 1992.

Andersen, O. B., Nielsen, K., Knudsen, P., Hughes, C. W., Fenoglio-marc, L., Gravelle, M., Kern, M., Fenoglio-marc, L., Gravelle, M., Kern, M., and Polo, S. P.: Improving the Coastal Mean Dynamic Topography by Geodetic Combination of Tide Gauge and Satellite Altimetry Improving the Coastal Mean Dynamic Topography by Geodetic Combination of Tide Gauge and Satellite Altimetry, Marine Geodesy, 0, 1–29, https://doi.org/10.1080/01490419.2018.1530320, https://doi.org/10.1080/01490419.2018.1530320, 2018.

Ballu, V., Gravelle, M., Wöppelmann, G., de Viron, O., Rebischung, P., Becker, M., and Sakic, P.: Vertical land motion in the Southwest and Central Pacific from available GNSS solutions and implications for relative sea levels, Geophysical Journal International, 218, 1537–1551, https://doi.org/10.1093/gji/ggz247, https://doi.org/10.1093/gji/ggz247, 2019.

Bloßfeld, M., Angermann, D., and Seitz, M.: DGFI-TUM Analysis and Scale Investigations of the Latest Terrestrial Reference Frame Realizations, in: International Symposium on Advancing Geodesy in a Changing World, edited by Freymueller, J. T. and Sánchez, L., pp. 3–9, Springer International Publishing, Cham, 2019.

Bos, M. and Fernandes, R.: Hector user manual, pp. 1–43, 2019.

Bosch, W. and Savcenko, R.: Satellite altimetry: Multi-mission cross calibration, in: International Association of Geodesy Symposia, https://doi.org/10.1007/978-3-540-49350-1_8, 2007.

Bosch, W., Dettmering, D., and Schwatke, C.: Multi-mission cross-calibration of satellite altimeters: Constructing a long-term data record for global and regional sea level change studies, Remote Sensing, https://doi.org/10.3390/rs6032255, 2014.

Bouin, M. N. and Wöppelmann, G.: Land motion estimates from GPS at tide gauges: a geophysical evaluation, Geophysical Journal International, 180, 193–209, https://doi.org/10.1111/j.1365-246X.2009.04411.x, https://doi.org/10.1111/j.1365-246X.2009.04411.x, 2010.

Brooks, B. A., Merrifield, M. A., Foster, J., Werner, C. L., Gomez, F., Bevis, M., and Gill, S.: Space geodetic determination of spatial variability in relative sea level change, Los Angeles basin, Geophysical Research Letters, 34, https://doi.org/10.1029/2006GL028171, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL028171, 2007.

Calafat, F. M., Wahl, T., Lindsten, F., Williams, J., and Frajka-Williams, E.: Coherent modulation of the sea-level annual cycle in the United States by Atlantic Rossby waves, Nature Communications, 9, https://doi.org/10.1038/s41467-018-04898-y, 2018.

Carrère, L. and Lyard, F.: Modeling the barotropic response of the global ocean to atmospheric wind and pressure forcing - comparisons with observations, Geophysical Research Letters, 30, https://doi.org/10.1029/2002GL016473, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002GL016473, 2003.

Carrère, L., Lyard, F., Cancet, M., and Guillot, A.: FES 2014, a new tidal model on the global ocean with enhanced accuracy in shallow seas and in the Arctic region, in: EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts, p. 5481, 2015.

Carrère, L., Faugère, Y., and Ablain, M.: Major improvement of altimetry sea level estimations using pressure-derived corrections based on ERA-Interim atmospheric reanalysis, Ocean Science, 12, 825–842, https://doi.org/10.5194/os-12-825-2016, https://os.copernicus.org/articles/12/825/2016/, 2016.

Carson, M., Köhl, A., Stammer, D., Slangen, A., Katsman, C., Wal, R., Church, J., and White, N.: Coastal Sea Level Changes, Observed and Projected during the 20th and 21st Century, Climatic Change, 134, 269–281, https://doi.org/10.1007/s10584-015-1520-1, 2016.

Cazenave, A., Dominh, K., Ponchaut, F., Soudarin, L., Cretaux, J. F., and Le Provost, C.: Sea level changes from Topex-Poseidon altimetry and tide gauges, and vertical crustal motions from DORIS, Geophysical Research Letters, https://doi.org/10.1029/1999GL900472, 1999.

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.

Church, J. and White, N.: Sea-Level Rise from the Late 19th to the Early 21st Century, Surveys in Geophysics, 32, 585–602, https://doi.org/10.1007/s10712-011-9119-1, 2011.

Church, J. A., Clark, P., Cazenave, A., Gregory, J., Jevrejeva, S., Levermann, A., Merrifield, M., Milne, G., Nerem, R., Nunn, P., Payne, A., Pfeffer, W., Stammer, D., and Unnikrishnan, A.: 2013: Sea level change, Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 1137–1216, https://doi.org/10.1017/CB09781107415315.026, 2013.

Cipollini, P., Calafat, F. M., Jevrejeva, S., Melet, A., and Prandi, P.: Monitoring Sea Level in the Coastal Zone with Satellite Altimetry and Tide Gauges, https://doi.org/10.1007/s10712-016-9392-0, 2017.

Cleveland, W. S. and Devlin, S. J.: Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting, Journal of the American Statistical Association, 83, 596–610, http://www.jstor.org/stable/2289282, 1988.

Collilieux, X. and Woppelmann, G.: Global sea-level rise and its relation to the terrestrial reference frame, Journal of Geodesy, 85, 9–22, https://doi.org/10.1007/s00190-010-0412-4, 2009.

Couhert, A., Cerri, L., Legeais, J.-F., Michaël, A., Zelensky, N., Haines, B., Lemoine, F., Bertiger, W., Desai, S., and Otten, M.: Towards the 1 mm/y stability of the radial orbit error at regional scales, Advances in Space Research, 55, https://doi.org/10.1016/j.asr.2014.06.041, 2015.

Dangendorf, S., Marcos, M., Wöppelmann, G., Conrad, C. P., Frederikse, T., and Riva, R.: Reassessment of 20th century global mean sea level rise, Proceedings of the National Academy of Sciences, https://doi.org/10.1073/pnas.1616007114, https://www.pnas.org/content/early/2017/05/16/1616007114, 2017.

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

Ducet, N., Le Traon, P. Y., and Reverdin, G.: Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2, Journal of Geophysical Research: Oceans, 105, 19 477–19 498, https://doi.org/10.1029/2000JC900063, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JC900063, 2000.

Fenoglio, L., Schöne, T., Illigner, J., Becker, M., Manurung, P., and Khafid: Sea Level Change and Vertical Motion from Satellite Altimetry, Tide Gauges and GPS in the Indonesian Region, Marine Geodesy, 35, https://doi.org/10.1080/01490419.2012.718682, 2012.

Fernandes, M. J., Lázaro, C., Ablain, M., and Pires, N.: Remote Sensing of Environment Improved wet path delays for all ESA and reference altimetric missions, Remote Sensing of Environment, 169, 50–74, https://doi.org/10.1016/j.rse.2015.07.023, http://dx.doi.org/10.1016/j.rse.2015.07.023, 2015.

Frederikse, T., Landerer, F. W., and Caron, L.: The imprints of contemporary mass redistribution on local sea level and vertical land motion observations, Solid Earth, 10, 1971–1987, https://doi.org/10.5194/se-10-1971-2019, https://se.copernicus.org/articles/10/1971/2019/, 2019.

Hamlington, B. D., Thompson, P., Hammond, W. C., Blewitt, G., and Ray, R. D.: Assessing the impact of vertical land motion on twentieth century global mean sea level estimates, Journal of Geophysical Research: Oceans, https://doi.org/10.1002/2016JC011747, 2016.

Hawkins, R., Husson, L., Choblet, G., Bodin, T., and Pfeffer, J.: Virtual Tide Gauges for Predicting Relative Sea Level Rise, Journal of Geophysical Research: Solid Earth, 124, 13 367–13 391, https://doi.org/10.1029/2019JB017943, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JB017943, 2019.

Hay, C. C., Morrow, E., Kopp, R. E., and Mitrovica, J. X.: Probabilistic reanalysis of twentieth-century sea-level rise, Nature, https://doi.org/10.1038/nature14093, 1990.

Holgate, S. J., Matthews, A., Woodworth, P. L., Rickards, L. J., Tamisiea, M. E., Bradshaw, E., Foden, P. R., Gordon, K. M., Jevrejeva, S., and Pugh, J.: New Data Systems and Products at the Permanent Service for Mean Sea Level, Journal of Coastal Research, pp. 493–504, https://doi.org/10.2112/JCOASTRES-D-12-00175.1, https://doi.org/10.2112/JCOASTRES-D-12-00175.1, 2013.

Hughes, C. and Meredith, M.: Coherent sea-level fluctuations along the global continental slope, Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 364, 885–901, https://doi.org/10.1098/rsta.2006.1744, 2006.

Hughes, C. W., Fukumori, I., Griffies, S. M., and Huthnance, J. M.: Sea Level and the Role of Coastal Trapped Waves in Mediating the Influence of the Open Ocean on the Coast, Surveys in Geophysics, 40, 1467–1492, https://doi.org/10.1007/s10712-019-09535-x, https://doi.org/10.1007/s10712-019-09535-x, 2019.

Idžanović, M., Gerlach, C., Breili, K., and Andersen, O.: An Attempt to Observe Vertical Land Motion along the Norwegian Coast by CryoSat-2 and Tide Gauges, Remote Sensing, 11, 744, https://doi.org/10.3390/rs11070744, 2019.

Jevrejeva, S., Moore, J. C., Grinsted, A., Matthews, A. P., and Spada, G.: Trends and acceleration in global and regional sea levels since 1807, Global and Planetary Change, https://doi.org/10.1016/j.gloplacha.2013.12.004, 2014.

King, M. A., Keshin, M., Whitehouse, P. L., Thomas, I. D., Milne, G., and Riva, R. E. M.: Regional biases in absolute sea-level estimates from tide gauge data due to residual unmodeled vertical land movement, Geophysical Research Letters, 39, https://doi.org/10.1029/2012GL052348, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GL052348, 2012.

Kleinherenbrink, M., Riva, R., and Frederikse, T.: A comparison of methods to estimate vertical land motion trends from GNSS and altimetry at tide gauge stations, Ocean Science, 14, 187–204, https://doi.org/10.5194/os-14-187-2018, https://os.copernicus.org/articles/14/187/2018/, 2018.

Kolker, A. S., Allison, M. A., and Hameed, S.: An evaluation of subsidence rates and sea-level variability in the northern Gulf of Mexico, Geophysical Research Letters, 38, https://doi.org/10.1029/2011GL049458, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011GL049458, 2011.

Kuo, C. Y., Shum, C. K., Braun, A., and Mitrovica, J. X.: Vertical crustal motion determined by satellite altimetry and tide gauge data in Fennoscandia, Geophysical Research Letters, 31, https://doi.org/10.1029/2003GL019106, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003GL019106, 2004.

Kurapov, A., Erofeeva, S., and Myers, E.: Coastal sea level variability in the US West Coast Ocean Forecast System (WCOFS), Ocean Dynamics, 67, https://doi.org/10.1007/s10236-016-1013-4, 2016.

Landskron, D. and Böhm, J.: Refined discrete and empirical horizontal gradients in VLBI analysis, Journal of Geodesy, 92, 1387–1399, https://doi.org/10.1007/s00190-018-1127-1, https://doi.org/10.1007/s00190-018-1127-1, 2018.

Mazzotti, S., Jones, C., and Thomson, R. E.: Relative and absolute sea level rise in western Canada and northwestern United States from a combined tide gauge-GPS analysis, Journal of Geophysical Research: Oceans, 113, https://doi.org/10.1029/2008JC004835, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JC004835, 2008.

Moritz, H.: Geodetic Reference System 1980, Journal of Geodesy, 74, 128–133, https://doi.org/10.1007/s001900050278, 2000.

Nerem, R. S. and Mitchum, G. T.: Estimates of vertical crustal motion derived from differences of TOPEX/POSEIDON and tide gauge sea level measurements, Geophysical Research Letters, https://doi.org/10.1029/2002gl015037, 2003.

Passaro, M., Cipollini, P., Vignudelli, S., Quartly, G. D., and Snaith, H. M.: ALES: A multi-mission adaptive subwaveform retracker for coastal and open ocean altimetry, Remote Sensing of Environment, 145, 173–189, https://doi.org/10.1016/j.rse.2014.02.008, http://dx.doi.org/10.1016/j.rse.2014.02.008, 2014.

Passaro, M., Cipollini, P., and Benveniste, J.: Annual sea level variability of the coastal ocean: The Baltic Sea-North Sea transition zone, Journal of Geophysical Research: Oceans, 120, 3061–3078, https://doi.org/10.1002/2014JC010510, 2015.

Passaro, M., Müller, F. L., and Dettmering, D.: Lead detection using Cryosat-2 delay-doppler processing and Sentinel-1 SAR images, Advances in Space Research, 62, 1610–1625, https://doi.org/10.1016/j.asr.2017.07.011, 2018.

Peltier, W.: GLOBAL GLACIAL ISOSTASY AND THE SURFACE OF THE ICE-AGE EARTH: The ICE-5G (VM2) Model and GRACE, Annual Review of Earth and Planetary Sciences, 32, 111–149, https://doi.org/10.1146/annurev.earth.32.082503.144359, https://doi.org/10.1146/annurev.earth.32.082503.144359, 2004.

Petit, G. and Luzum, B.: IERS Conventions, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt, Germany, 2010.

Pfeffer, J. and Allemand, P.: The key role of vertical land motions in coastal sea level variations: A global synthesis of multisatellite altimetry, tide gauge data and GPS measurements, Earth and Planetary Science Letters, 439, 39–47, https://doi.org/10.1016/j.epsl.2016.01.027, http://dx.doi.org/10.1016/j.epsl.2016.01.027, 2016.

Pfeffer, J., Spada, G., Mémin, A., Boy, J.-P., and Allemand, P.: Decoding the origins of vertical land motions observed today at coasts, Geophysical Journal International, 210, 148–165, https://doi.org/10.1093/gji/ggx142, https://doi.org/10.1093/gji/ggx142, 2017.

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.

Piccioni, G., Dettmering, D., Schwatke, C., Passaro, M., and Seitz, F.: Design and regional assessment of an empirical tidal model based on FES2014 and coastal altimetry, Advances in Space Research, https://doi.org/https://doi.org/10.1016/j.asr.2019.08.030, http://www.sciencedirect.com/science/article/pii/S0273117719306131, 2019.

Poitevin, C., Wöppelmann, G., Raucoules, D., Cozannet, G. L., Marcos, M., and Testut, L.: Vertical land motion and relative sea level changes along the coastline of Brest (France) from combined space-borne geodetic methods, Remote Sensing of Environment, 222, 275–285, https://doi.org/https://doi.org/10.1016/j.rse.2018.12.035, http://www.sciencedirect.com/science/article/pii/S0034425718305960, 2019.

Ponte, R. M.: Low-Frequency Sea Level Variability and the Inverted Barometer Effect, Journal of Atmospheric and Oceanic Technology, 23, 619–629, https://doi.org/10.1175/JTECH1864.1, https://doi.org/10.1175/JTECH1864.1, 2006.

Pugh, D. and Woodworth, P.: Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes, Cambridge University Press, https://doi.org/10.1017/CBO9781139235778, 2014.

Riva, R., Frederikse, T., King, M., Marzeion, B., and Van den Broeke, M.: Brief communication: The global signature of post-1900 land ice wastage on vertical land motion, The Cryosphere, 11, 1327–1332, https://doi.org/10.5194/tc-11-1327-2017, 2017.

Sanli, D. and Blewitt, G.: Geocentric sea level trend using GPS and >100-year tide gauge record on a postglacial rebound nodal line, Journal of Geophysical Research, 106, 713–720, https://doi.org/10.1029/2000JB900348, 2001.

Santamaría-Gómez, A., Gravelle, M., and Wöppelmann, G.: Long-term vertical land motion from double-differenced tide gauge and satellite altimetry data, Journal of Geodesy, 88, 207–222, https://doi.org/10.1007/s00190-013-0677-5, 2014.

Santamaría-Gómez, A., Gravelle, M., Dangendorf, S., Marcos, M., Spada, G., and Wöppelmann, G.: Uncertainty of the 20th century sea-level rise due to vertical land motion errors, Earth and Planetary Science Letters, 473, 24–32, https://doi.org/10.1016/j.epsl.2017.05.038, http://dx.doi.org/10.1016/j.epsl.2017.05.038, 2017.

Santamaría-Gómez, A., Bouin, M.-N., Collilieux, X., and Wöppelmann, G.: Correlated errors in GPS position time series: Implications for velocity estimates, Journal of Geophysical Research: Solid Earth, 116, https://doi.org/10.1029/2010JB007701, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JB007701, 2011.

Santamaría-Gómez, A., Gravelle, M., Collilieux, X., Guichard, M., Míguez, B. M., Tiphaneau, P., and Wöppelmann, G.: Mitigating the effects of vertical land motion in tide gauge records using a state-of-the-art GPS velocity field, Global and Planetary Change, 98-99, 6 – 17, https://doi.org/https://doi.org/10.1016/j.gloplacha.2012.07.007, http://www.sciencedirect.com/science/article/pii/S0921818112001476, 2012.

Santamaría-Gómez, A., Gravelle, M., and Wöppelmann, G.: GPS Solution ULR6, SONEL Data Center, https://doi.org/10.26166/sonel_ulr6a, https://www.sonel.org/-Vertical-land-movement-estimate-.html?lang=en, 2016.

Saraceno, M., Strub, P. T., and Kosro, P. M.: Estimates of sea surface height and near-surface alongshore coastal currents from combinations of altimeters and tide gauges, Journal of Geophysical Research: Oceans, https://doi.org/10.1029/2008JC004756, 2008.

Scharroo, R. and Smith, W. H. F.: A global positioning system–based climatology for the total electron content in the ionosphere, Journal of Geophysical Research: Space Physics, 115, https://doi.org/10.1029/2009JA014719, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JA014719, 2010.

Scharroo, R., Leuliette, E., Lillibridge, J., Byrne, D., Naeije, M., and Mitchum, G.: RADS: Consistent multi-mission products, in: Proceedings of Symposium on 20 Years of Progress in Radar Altimetry, vol. 20, p. 59–60, 2012.

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

Slangen, A. B. A., Carson, M., Katsman, C. A., van de Wal, R. S. W., Köhl, A., Vermeersen, L. L. A., and Stammer, D.: Projecting twenty-first century regional sea-level changes, Climatic Change, 124, 317–332, https://doi.org/10.1007/s10584-014-1080-9, https://doi.org/10.1007/s10584-014-1080-9, 2014.

Snay, R., Cline, M., Dillinger, W., Foote, R., Hilla, S., Kass, W., Ray, J., Rohde, J., Sella, G., and Soler, T.: Using global positioning system-derived crustal velocities to estimate rates of absolute sea level change from North American tide gauge records, Journal of Geophysical Research: Solid Earth, 112, https://doi.org/10.1029/2006JB004606, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006JB004606, 2007.

Stammer, D. and Böning, C. W.: Mesoscale Variability in the Atlantic Ocean from Geosat Altimetry and WOCE High-Resolution Numerical Modeling, Journal of Physical Oceanography, 22, 732–752, https://doi.org/10.1175/1520-0485(1992)022<0732:MVITAO>2.0.CO;2, https://doi.org/10.1175/1520-0485(1992)022{%}3C0732:MVITAO{%}3E2.0.COhttp://0.0.0.2, 1992.

Stammer, D., Cazenave, A., Ponte, R. M., and Tamisiea, M. E.: Causes for Contemporary Regional Sea Level Changes, Annual Review of Marine Science, 5, 21–46, https://doi.org/10.1146/annurev-marine-121211-172406, https://doi.org/10.1146/annurev-marine-121211-172406, pMID: 22809188, 2013.

Sánchez L., B. W.: The role of the TIGA project in the unification of classical height systems, In: Drewes H. (Ed.): Geodetic Reference Frames, IAG Symposia 134: 285-290, Springer, https://doi.org/10.1007/978-3-642-00860-3_44, 2009.

Veit, E. and Conrad, C. P.: The impact of groundwater depletion on spatial variations in sea level change during the past century, Geophysical Research Letters, 43, 3351–3359, https://doi.org/10.1002/2016GL068118, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL068118, 2016.

Wada, Y., van Beek, L. P. H., Sperna Weiland, F. C., Chao, B. F., Wu, Y.-H., and Bierkens, M. F. P.: Past and future contribution of global groundwater depletion to sea-level rise, Geophysical Research Letters, 39, https://doi.org/10.1029/2012GL051230, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GL051230, 2012.

Watson, C., White, N., Church, J., King, M., Burgette, R., and Legresy, B.: Unabated global mean sea-level rise over the satellite altimter era, Nature Climate Change, 5, https://doi.org/10.1038/nclimate2635, 2015.

Woodworth, P. L., Hunter, J. R., Marcos, M., Caldwell, P., Menéndez, M., and Haigh, I.: Towards a global higher-frequency sea level dataset, Geoscience Data Journal, https://doi.org/10.1002/gdj3.42, 2016.

Woodworth, P. L., Melet, A., Marcos, M., Ray, R. D., Wöppelmann, G., Sasaki, Y. N., Cirano, M., Hibbert, A., Huthnance, J. M., Monserrat, S., and Merrifield, M. A.: Forcing factors affecting sea level changes at the coast, Surveys in Geophysics, 40, 1351–1397, http://nora.nerc.ac.uk/id/eprint/523283/, 2019.

Wöppelmann, G., Martin Miguez, B., Bouin, M.-N., and Altamimi, Z.: Geocentric sea-level trend estimates from GPS analyses at relevant tide gauges world-wide, Global and Planetary Change, 57, 396–406, https://doi.org/10.1016/j.gloplacha.2007.02.002, 2007.

Wöppelmann, G. and Marcos, M.: Vertical land motion as a key to understanding sea level change and variability, Reviews of Geophysics, 54, 64–92, https://doi.org/10.1002/2015RG000502, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015RG000502, 2016.

Wöppelmann, G., Gravelle, M., Guichard, M., and Prouteau, E.: Fourth Progress Report on the GNSS at Tide Gauge Activities: SONEL Data Holdings & Tools to access the data, presented at the XVIth GLOSS group of experts meeting, hosted by the Korean Hydrographic and Oceanographic Agency, Busan (Republic of Korea), 11-13 April 2019. Online version available at: https://www.sonel.org/IMG/pdf/ge16_gnssattg_activities_sonel_report_v2.pdf, 2019.

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:

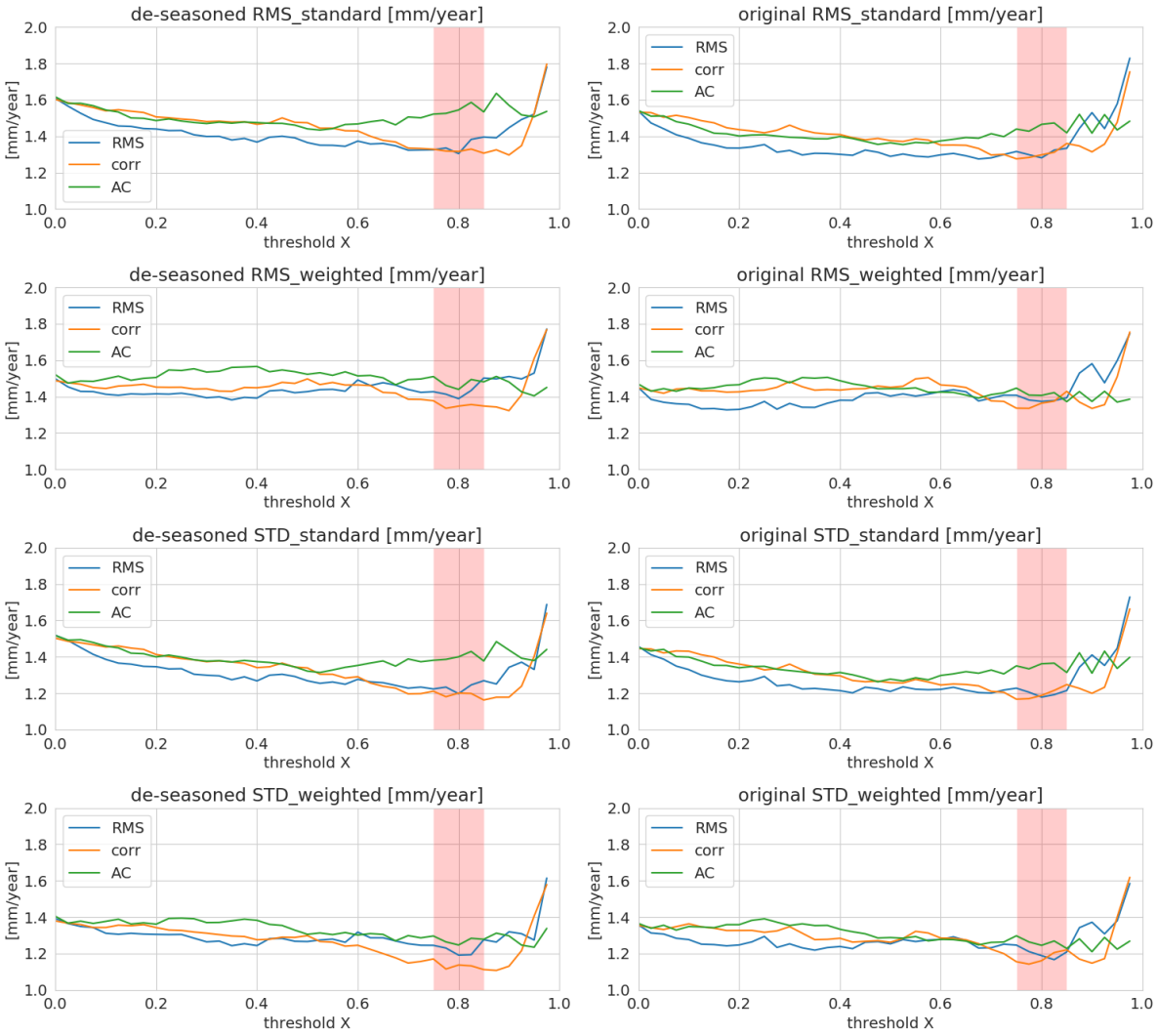
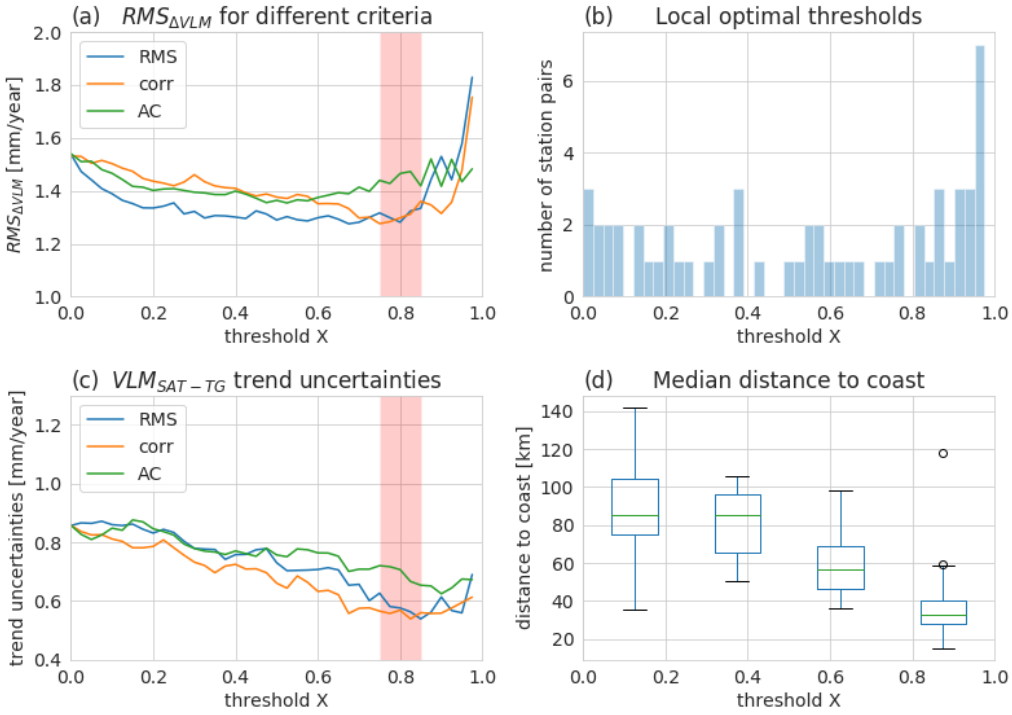


Figure R1: Comparison of statistics when the data was de-seasoned (left) or not (right) before matching. First row shows the unweighted RMS, the second row shows the weighted RMS of trend differences, the third row shows the un-weighted standard deviation and the last column shows the weighted standard 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

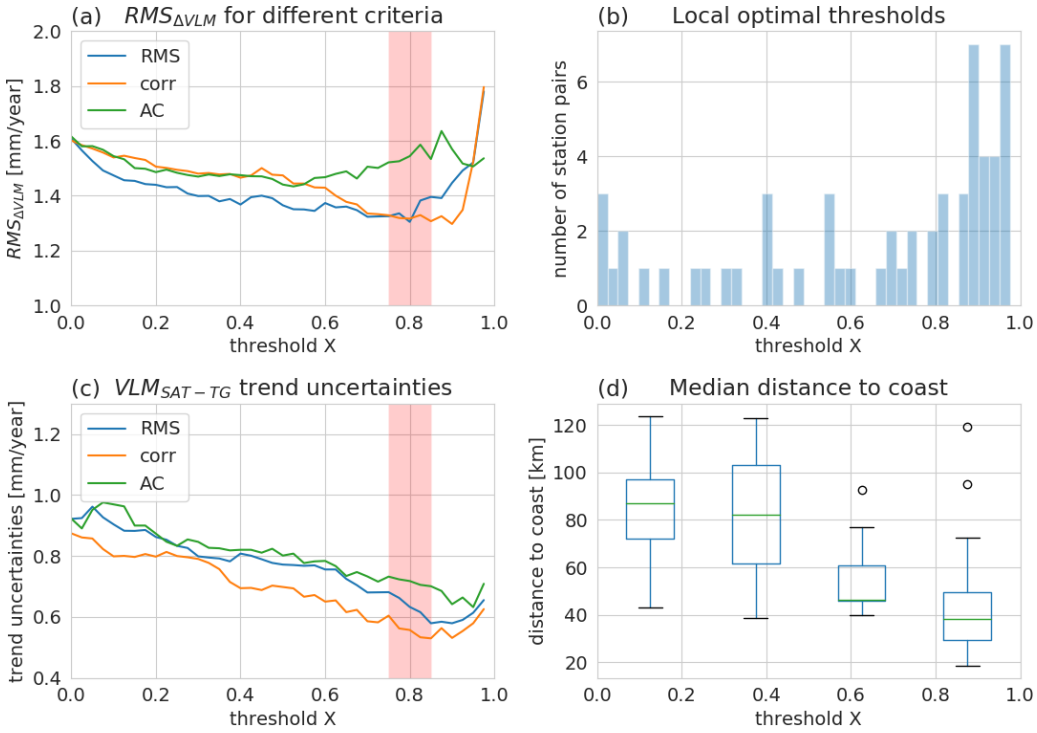


Figure R2: Shown are A) statistics when data was only detrended but not de-seasoned; B) statistics based on detrended and de-seasoned data. Figure captions are as in Figure 5: Performance of VLMSAT-TG trend estimates for ALES-GESLA-ZOI. a) RMSΔVLM for different relative thresholds (step size 2.5%) and 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):

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

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

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

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.'

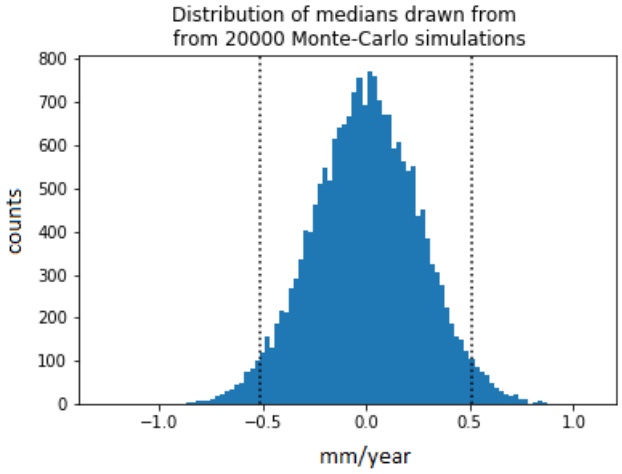
In the results section, we add following lines to the paragraph from L385:

'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 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.'

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.

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 paper. 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):

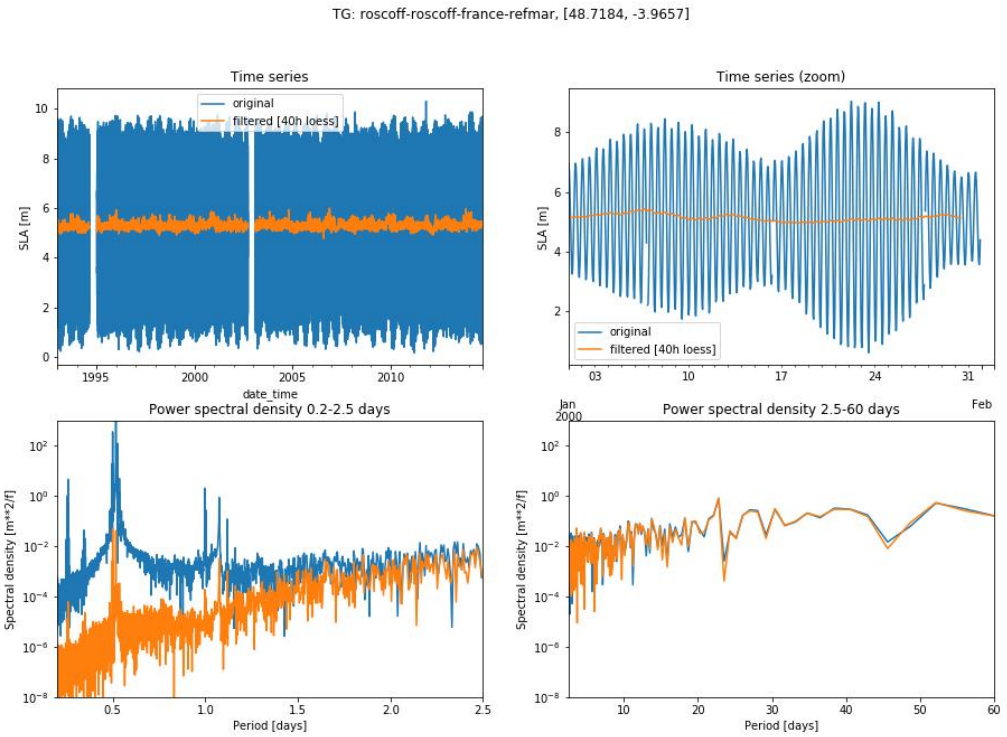
'For all missions satellite orbits in ITRF2008 are used, mostly processed by CNES (GDR-E). For ERS-2, GFZ VER11 orbits are applied.'

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)



B)

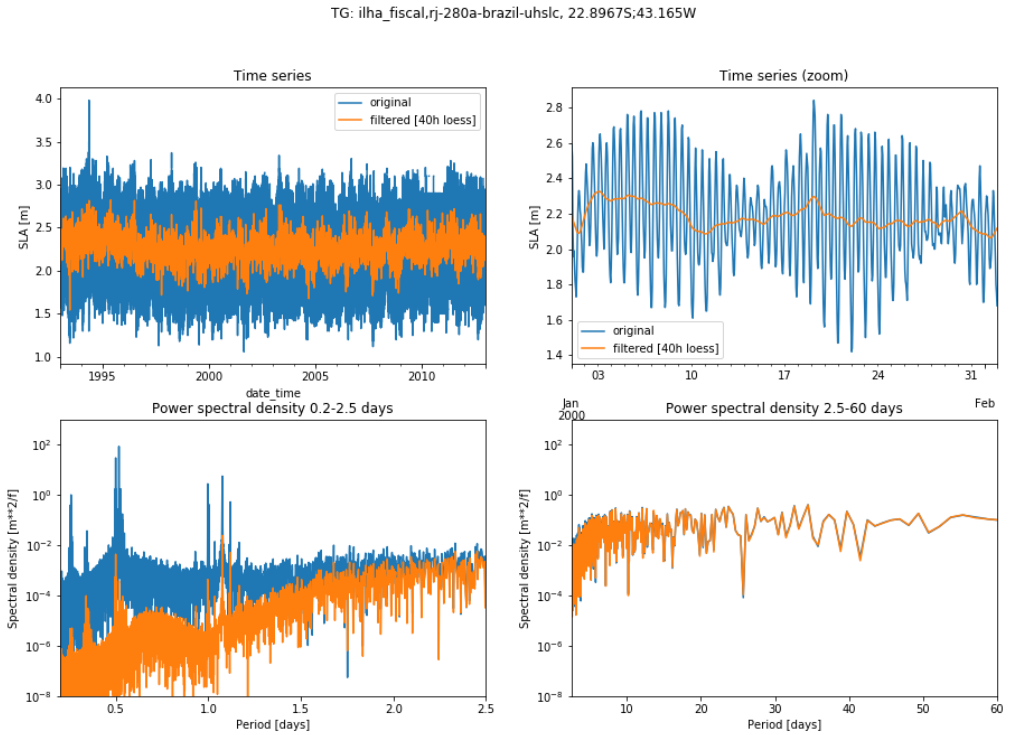


Figure R1: Time series (upper rows) and spectral densities (lower rows) for two different TG stations (A) France and B) Brazil). The original time series is shown in blue and the filtered one in orange. Spectral density estimates are obtained from using Welch's method and are based on time series with 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.

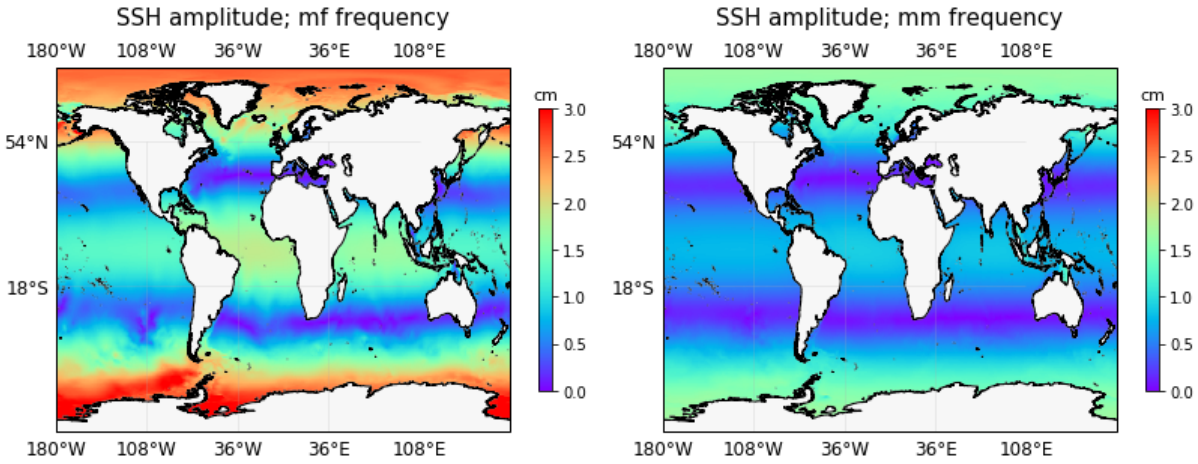


Figure R2: Modelled (FES2014) SSH amplitude of two fortnightly and monthly tidal constituents: Mf (left, period 13.66 days) and Mm (right, period 27.55 days)

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.

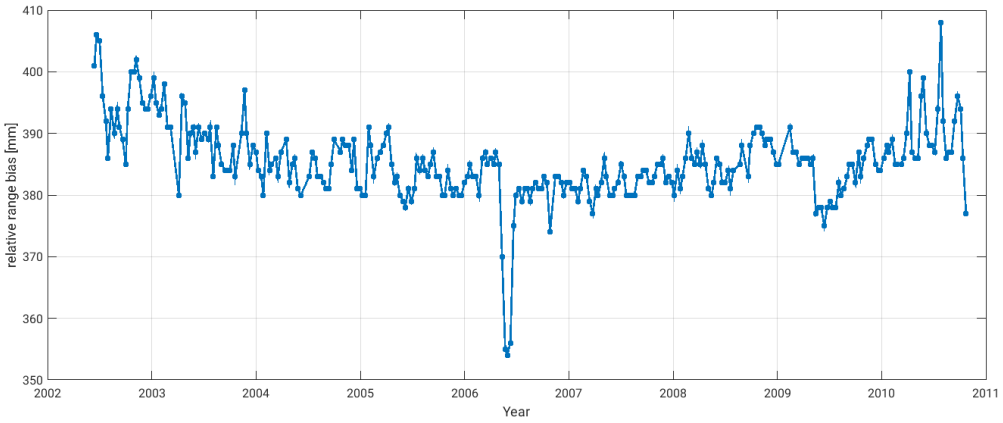


Figure R3: Envisat relative range bias per 10-day cycle (w.r.t. Jason-1) [mm]

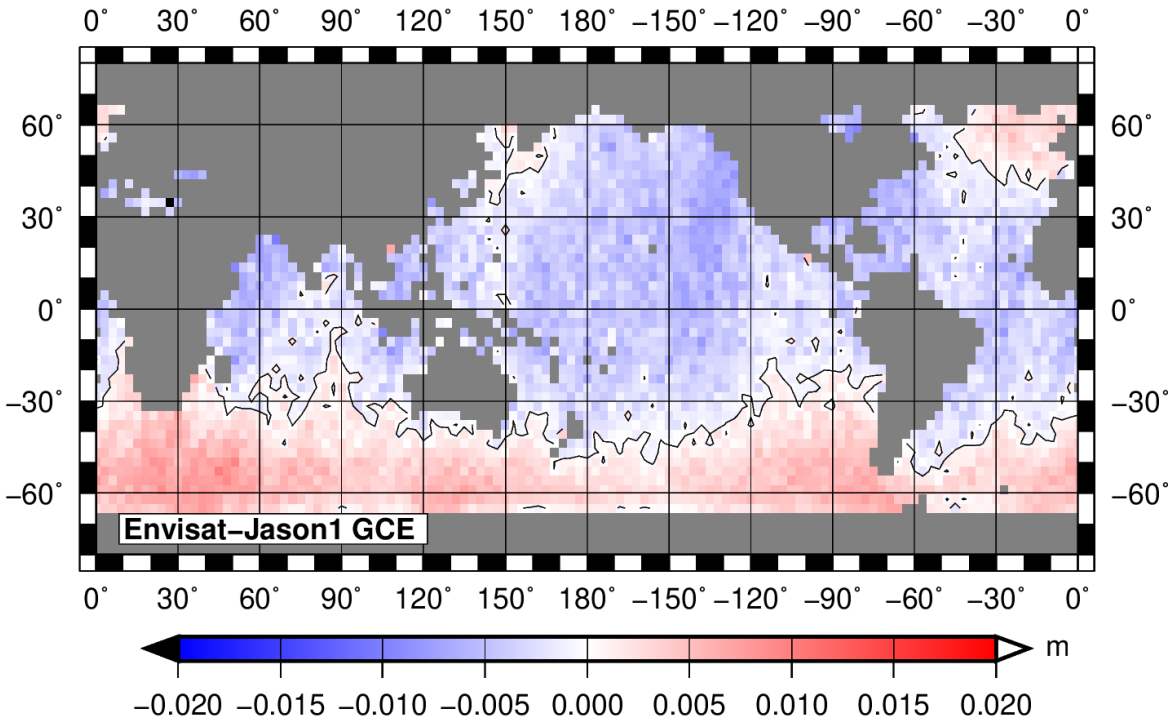


Figure R4: Geographically correlated mean errors of Envisat w.r.t. Jason-1 [mm]

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.

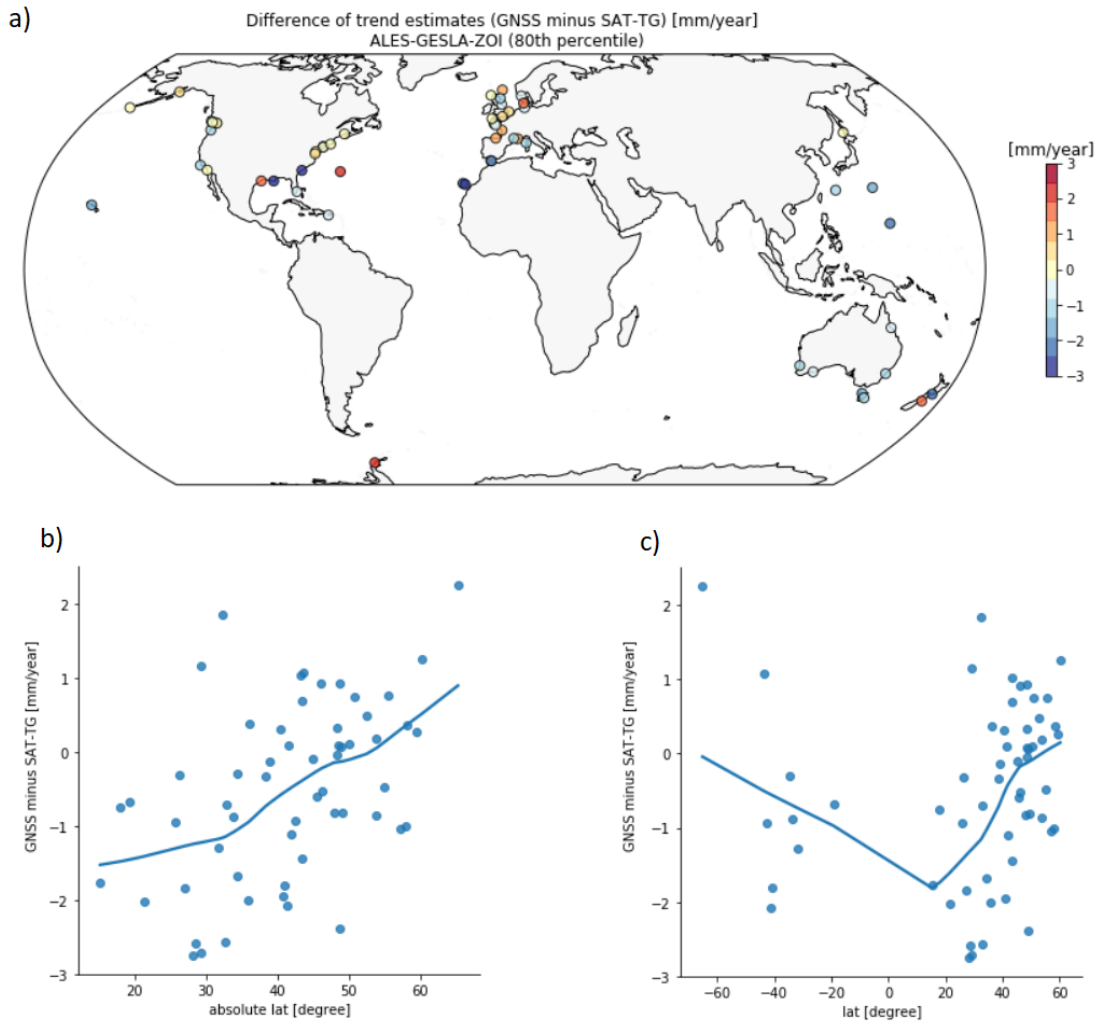


Figure R5: Trend differences of GNSS minus SAT-TG VLM [mm/year] for ALES-GESLA-ZOI (80th percentile) (a) Map of trend differences; (b and c) Trend differences as a function of absolute latitude and latitude. Solid lines represent locally weighted regressions (Cleveland et al., 1988).

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.

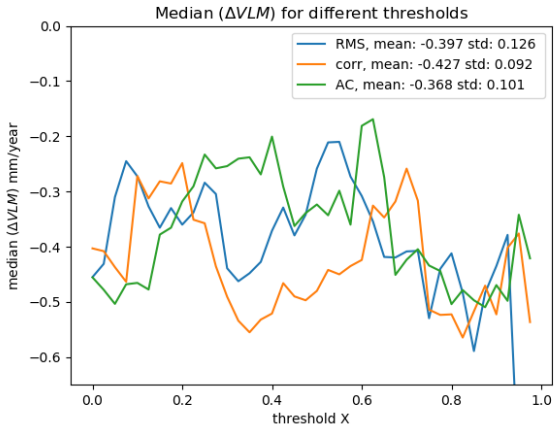


Figure R6: Shown is the evolution of the bias (median of trend differences) depending on different criteria and thresholds for ALES-GESLA-ZOI (as in Figure 5)

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')