

HYDROCOASTAL

SAR/SARin Radar Altimetry for Coastal Zone and Inland Water Level

Technical Note - Impact Assessment Report on Data Assimilation Deliverable D2.7

Sentinel-3 and CryoSat-2 SAR/SARin Radar Altimetry for Coastal Zone and Inland Water
ESA Contract 4000129872/20/I-DT

Project reference: HYDROCOASTAL_ESA_TN_D2.7
Issue: 1.0

This page has been intentionally left blank

Change Record

Date	Issue	Section	Page	Comment
August 29, 2023				Addressed comments raised by ESA

Control Document

Process	Name	Date
Written by:	Henrique Guarneri, Cornelis Slobbe, and Martin Verlaan	08/2023
Checked by	David Cotton	August 29, 2023
Approved by:		

Subject	Radar Altimetry for Coastal Zone and Inland Water Level	Project	HYDROCOASTAL
Author	Organisation	Internal references	
Henrique Guarneri	TU Delft		
Cornelis Slobbe	TU Delft		
Martin Verlaan	TU Delft / Deltares		

	Signature	Date
For HYDROCOASTAL team		11/09/23
For ESA		

Table of Contents

Table of Contents	4
1. Introduction	5
1.1. Background, Motivation, and Objectives	5
1.2. Scope of this Document	6
1.3. Document Organisation	6
2. Data and Model	6
2.1. TOPEX/Poseidon and Jason-derived Sea Level Anomalies	6
2.2. Along-track Tidal Product	7
2.3. Sentinel-3 Data	7
2.4. Tide Gauge Data	8
2.5. The Dutch Continental Shelf Model version 7	9
3. Methods	10
3.1. The Data Assimilation Procedure	10
3.2. Validation	11
4. Results and Discussion	11
4.1. Performance Assessments based on the model-only simulation	11
4.2. The Impact of Assimilating TPJ-derived Low-frequency Water Levels	16
5. Summary and Conclusion	19
6. References	21

1. Introduction

1.1. Background, Motivation, and Objectives

Exploiting the wealth of satellite radar altimeter data to calibrate regional, high-resolution hydrodynamic models used to forecast, among others, instantaneous water levels is one of the main objectives of the research cooperation between TU Delft and Deltares. In a straightforward implementation, we would assimilate the *total* water level relative to a well-defined vertical reference surface. However, based on twin experiments with a conceptual 1D coastal ocean model, we discovered that a direct assimilation of *total* water levels into the model is not feasible. The difficulties are due to i) the low temporal resolution of altimeter observations, ii) the high degree of freedom present in the data assimilation procedure, and iii) the shallow waters' nonlinear dynamics. The first implies that the altimeter time series and the simulation period need to cover decades to accumulate enough information. The second refers to the fact that differences between model and observation could have many causes. Since the system is 'under-observed' it becomes impossible to narrow down the correct update. Hence, we get an 'innovation state' with the wrong variables being updated. This, in turn, affects the dynamics in the wrong way. As an example, in the experiments we observed shock waves that were traveling through the model domain.

Subsequent research focused on an approach in which we exploit altimeter data to improve separate components of the water levels *in shallow water* (i.e., in areas where nonlinear interactions play a major role). Recently, we developed and implemented a procedure that allows to estimate tides from TOPEX/Poseidon and Jason (TPJ) satellite radar altimeter data in shallow waters showing significant improvements compared to tides obtained with a regional, high-resolution 2D tide-surge model (Guarneri et al., 2022). These altimeter-derived tides (more precisely, *residual* tides since we used the regional, high-resolution 2D tide-surge model as background model to reduce the signal variance) can be used to improve the model's tidal or low-frequency (periods > one month) water level representation. In addition to year-to-year variations in the seasonal water level, these low-frequency water level variations also include variations over longer time periods. For the most part, they are caused by physical processes not included in the model's governing equations or forcing terms, such as far field atmospheric forcing, baroclinic processes, major ocean currents, steric variations, or precipitation (Asher et al., 2019). For the current generation of storm surge models, they are a major source of errors in the operational forecasting of water levels. For this reason, a procedure was recently developed by Asher et al. (2019) to correct for this on the basis of low-frequency water levels from tide gauge records. Such a procedure is currently also being implemented in the Netherlands.

The main objective of this impact study is to explore whether satellite altimeter data can be used for this purpose. Doing so would largely increase the number of locations for which data are available. Specifically, *we aim to improve the low-frequency water level variability of the regional, high-resolution, 2D tide-surge Dutch Continental Shelf Model version 7 (DCSM) in Dutch coastal waters by the assimilation of TPJ-derived low-frequency water levels and to validate the approach by using Sentinel-3 (S3) water levels*. The added value of using S3 water levels as control data is that these data allow us to get much closer to the coast. The S3 data being used in this study are produced in the context of the HYDROCOASTAL project. A secondary objective is to assess the performance of i) the DCSM-derived tide-surge water levels compared to the FES tidal and Dynamic Atmospheric Correction (DAC) surge water levels, ii) the various retracker products included in the HYDROCOASTAL data product, and iii) an in-house developed unfocussed SAR product produced at a posting rate of 80 Hz (Ehlers, 2023a/b). The choice for 80 Hz is based on the results of Egido et al. (2021) and Ehlers et al. (2023b). The latter have shown that the required posting rate may reach up to 100 Hz for unfocussed SAR, depending on processing settings and sea state.

1.2. Scope of this Document

This document is the Impact Assessment Report (IAR) for HYDROCOASTAL and it corresponds to the deliverable D2.7 of the project. The scope of this report is to compile and assess the main findings of the impact activities.

1.3. Document Organisation

Sect. 1: A short introduction defining the scope of this report.

Sect. 2: Brief introduction to the data and model being used in this study.

Sect. 3: Brief introduction to the methods applied in this study.

Sect. 4: Results and discussion.

Sect. 5: Summary and Recommendations.

2. Data and Model

This section describes all datasets being used in the study, their processing, as well as the 2D Dutch Continental Shelf Model version 7 (DCSM).

2.1. TOPEX/Poseidon and Jason-derived Sea Level Anomalies

The TPJ-derived low-frequency water levels were obtained from the X-TRACK along-track sea level anomaly (SLA) product. This product is generated with the X-TRACK processing system (Birol et al., 2017) developed by the Center of Topography of the Ocean and Hydrosphere (CTOH) in Toulouse. It uses as input the measurements and parameters provided in the Geophysical Data Records plus additional corrections and auxiliary data specifically distributed by the CTOH. The product was created to improve the completeness and quality of sea surface height information received from satellite altimetry in coastal waters. The SLAs have been projected onto reference tracks with a point separation of about 6–7 km. In this study, we used the X-TRACK data of version 2017 (DOI is 10.6096/CTOH_X-TRACK_2017_02). The data cover the time span February 1993–May 2020. The data from 1993 through 2018 have been used to estimate the residual tides (see Sect. 2.2), the data from January 2017 onwards were assimilated into the model. Note that more recent TPJ data are included in the XTRACK SLA product version 2022. Switching to this newer version was not possible in this project, as the projection of the data to the reference tracks is different from the one in version 2017. This means that the model-derived time series over the entire data time span (1993 onwards) needed to compute the along-track tidal product have to be re-generated, which is very time-consuming.

To compute the water levels, we restored the applied tidal (FES) and surge (DAC) corrections. These corrections are provided with the data. The low-frequency water levels being assimilated into the model are computed as:

$$\text{low-frequency water level} = \text{observed water level} - \text{DCSM-derived tide-surge water level} - \text{residual tide correction}$$

Here, the residual tide correction is obtained from the in-house developed along-track tidal product (see Sect. 2.2). In our area of interest, tides and surge are by far the most important contributors to high-frequency water level variations. Hence, after removal, mainly the low-frequency water level variations

remain in addition to errors in DCSM's representation of the surge water levels. One possibility is to suppress these latter errors using filtering. However, the sampling frequency is very low. It was therefore decided to include these errors in the noise model of the low-frequency water levels. Note that in the assimilation, we only use the time series at crossover locations.

2.2. Along-track Tidal Product

The in-house developed along-track tidal product was estimated from the X-TRACK TPJ water levels (i.e., the SLAs + FES tide + DAC surge) covering the time span from 1993 to 2018. Here, the tide-surge DCSM model was used as a prior model. That is, *residual tides* were estimated after removing the DCSM-derived tide-surge water levels from the X-TRACK TPJ water levels. The DCSM-derived water levels include the nonlinear interaction between tides and surge, which was found to be a main contributor to the lower accuracy of altimeter-derived tides in shelf and coastal waters (Guarneri et al., 2023). A key asset of our approach is *to treat the prior model as stochastic*. Doing so avoids the location-dependent selection of the set of tidal constituents that can be estimated from the altimeter data. It does require knowledge of the error variance-covariance matrix of the model-derived tides. This matrix was empirically determined from the vector differences between observation- and model-derived tides at tide gauge locations. A detailed description of the method is the topic of a publication currently in preparation. For a validation of the method, we refer to Guarneri et al. (2022).

2.3. Sentinel-3 Data

The Sentinel-3 (S3) data used to assess the impact of the data assimilation were produced in the context of the HYDROCOASTAL project. They cover the area outlined in Fig. 2b. The data product includes the output of three retracker; the Spatio-Temporal Altimeter Retracker for SAR altimetry (STARS) of the University of Bonn, the Multiple Waveform Persistent Peak (MWaPP) retracker developed by DTU, and ESA's L2 SAMOSA 2.5 retracker (referred to as 'ESA retracker'). To assess the performance of the pure DCSM model (i.e., without applying data assimilation), we used the output of all three retracker. In the impact assessment, we only used the output of the STARS retracker. Quality flags included in the data product (i.e., 'flags_ubo_range', 'flags_dtu', and 'flags_ESA') have been applied. In this study, we used all data from 2017 to 2021. The water levels were computed using the following equation:

$$\text{water level} = \text{altitude} - \text{retracked range} - \text{quasi-geoid} - \text{corrections}$$

The quasi-geoid model is the official Dutch height reference surface computed by TU Delft, i.e., the NLGEO2018 model (Slobbe et al., 2019). The corrections include:

- Ionosphere correction (Model Ionospheric Correction)
- Dry tropospheric correction (university of Porto)
- Wet tropospheric correction (university of Porto)
- Sea state bias correction
- Solid earth tide correction
- Pole tide correction

To exclude water levels over tidal flats in the Wadden Sea, we used the thresholding waveform classification scheme developed by Bij de Vaate (2022). Some small changes were applied, as the scheme developed by Bij de Vaate (2022) was tailored to identify waveforms contaminated by sea ice. The classification scheme only requires the observed waveform as input. All data points classified as

'land' or 'unknown' were removed. We also removed all data points i) outside polygons outlining the North Sea and Wadden Sea basins, and ii) outside the DCSM model domain.

Part of the data preprocessing involves a stacking of all water levels acquired during different cycles. That is, we binned all data to form time series covering the time span from 2017 to 2021.

To analyze performances as a function of distance to the coast, we interpolated the distance to the nearest coastline from the global product provided by the NASA Ocean Biology Processing Group (OBPG) and R.P. Stumpf. (2012).

For S3A, we processed all data with relative orbit number 370 acquired between January 2017 and January 2022 using our in-house developed software package (Ehlers, 2023a/b). This package provides, among others, the data at a higher posting rate than what is available from the HYDROCOASTAL products. In this study, we use the unfocussed SAR product at a posting rate of 80 Hz (referred to as the 80Hz UF-SAR product). The ranges are obtained by retracking the sub-waveform with the SAMOSA2 retracker. Here, the sub-waveform contains the part of the waveform between the first range gate and the range gate that is 20 range gates beyond the one with the maximum value. All geophysical corrections listed above are obtained from the HYDROCOASTAL product.

2.4. Tide Gauge Data

Tide gauge data, which we used to validate the performance of the data assimilation, were obtained from an in-house developed database. The data were acquired by different national authorities in the countries Belgium, Denmark, France, Germany, Ireland, the Netherlands, Norway, Sweden, and Great Britain. All time series have been visually inspected for outliers. In total, we obtained 149 time series covering the period January 2017 to May 2020. No tide gauges are used that are outside DCSM's model domain or located on tidal flats. Fig. 1 shows a map of the tide gauges.

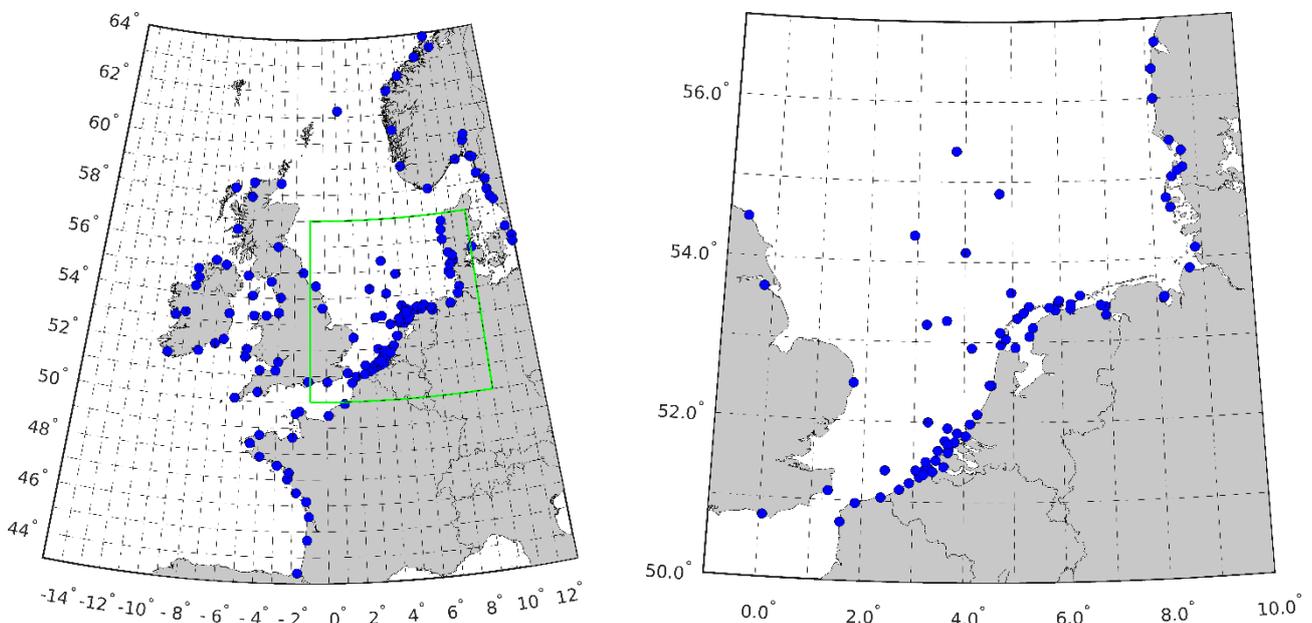


Fig. 1: Locations of tide gauges used to validate the performance of the data assimilation. The right panel shows the area outlined in green.

2.5. The Dutch Continental Shelf Model version 7

The 2D Dutch Continental Shelf Model - Flexible Mesh (DCSM), (Zijl and Groenenboom, 2019) is the successor of the version in Zijl et al. (2013, 2015). The model describes the tide-surge water level variability for the northwest European continental shelf between 15° W to 13° E and 43° N to 64° N (see Fig. 2) by solving the depth-integrated shallow-water equations for hydrodynamic modeling of free-surface flows (Leendertse, 1967; Stelling, 1984). In doing so, the nonlinear tide-surge interaction is accounted for. Contrary to the previous version, it uses the Delft3D Flexible Mesh Suite (or D-HYDRO Suite) that allows for the use of unstructured grids. For the model used in this study, the minimum grid size is approximately 840 × 930 m in Dutch waters (see Fig. 2). To reduce the uncertainty of the bottom roughness, an automated calibration using the 'Doesn't Use Derivative' algorithm (Ralston and Jennrich, 1978) has been performed. In doing so, all 2017 data from 195 tide gauges were used. Here, extra weight in the cost function has been given to the Dutch coastal tide gauges, since the model is primarily intended to obtain an accurate water level representation in Dutch coastal areas.

At the northern, western, and southern open boundaries, water level boundary conditions were applied. When modeling the tide-surge water levels, they are composed of the sum of the astronomical water levels and the surge. The tides were obtained from a harmonic expansion of 32 tidal constituents retrieved from the global ocean tide model FES2012 (Carrère et al., 2013) supplemented with the solar annual Sa constituent obtained from an earlier version of the model. The surge at the open boundaries was approximated by the time- and space-dependent inverse barometer correction. Also, in the model domain, the tidal potential was simulated and accounts for a smaller part of the tides, compared to the tides forced from the open boundaries. The time- and space-varying atmospheric wind and pressure forcings were obtained from the ECMWF's ERA5 reanalysis dataset (Hersbach et al., 2020).

In this project, a small routine was developed that linearly interpolates the model-derived water level time series at surrounding grid points to the altimeter data locations. This routine was used to generate the time series at the TPJ crossover locations. For the S3 data being used in the validation, a routine was developed to extract the water levels at the altimeter locations from time series of maps.

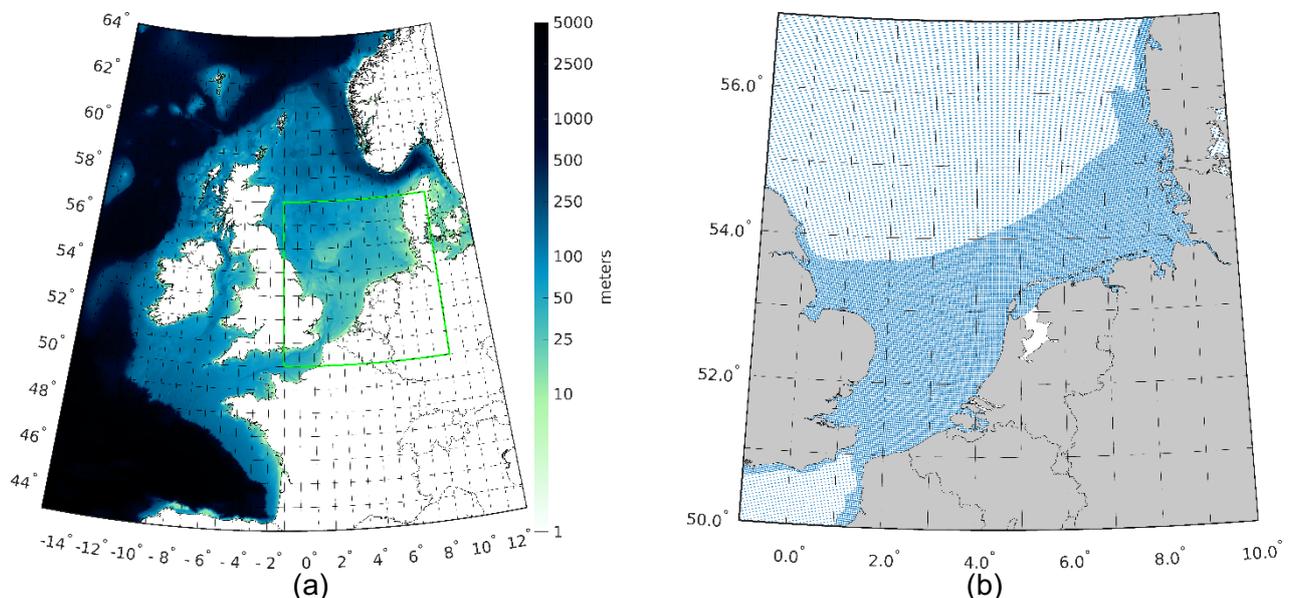


Fig. 2: Map of the bathymetry over the domain of the DCSM model used in this study (a). The part of the North Sea and the Wadden Sea included in the green box covers our main area of interest. For this area S3 data are available. In (b), we show the mesh of the DCSM model in the main area of interest.

3. Methods

3.1. The Data Assimilation Procedure

The data assimilation scheme developed and applied in this study is inspired by the scheme proposed by Asher et al., (2019). It comprises two steps:

Step 1: Replace the observation-derived mean water level over the entire simulation period by the model-derived mean water level – This study aims to improve the model representation of the low frequency water level variability. This does *not* comprise the contribution to the zero frequency, i.e., the mean water level (MWL). This contribution, among others, includes the datum shift between the vertical reference of the observed and modeled water levels. The contribution is removed by replacing for each time series the observation-derived MWL with the model-derived one. The latter are computed by running a simulation over the entire period from January 2017 to May 2020.

Step 2: Perform the data assimilation – The assimilation of the bias-corrected TPJ-derived water levels obtained in step 1 is based on the Kalman filter (Kalman 1960), which provides optimal state estimates by sequentially combining model and observations considering their uncertainties. More specifically, we used the asynchronous ensemble Kalman filter (Sakov, et al. 2010), which is essentially equivalent to the four-dimensional variational data assimilation (4D-Var). It allows to assimilate observations not acquired at the time of the update, referred to as asynchronous observations. As such, there is no need to stop model integration for analysis every time a new altimeter observation is available. This significantly reduces the computational load and thus simulation times without compromising accuracy. The number of ensemble members equals 40. This number has been found in experiments involving the assimilation of low-frequency water levels from tide gauge records. Note that localization has been applied (Hamill tapering function); the localization radius equals 1.200 km.

The differences between the observation-derived water levels and the model-derived tide-surge water levels include (apart from noise) the low-frequency water levels and the residual tidal signal (see Sect. 2.1). Since we only want to improve the model's low-frequency water level representation, we added the residual tidal water level to the modeled tide-surge water level before computing the innovation. The low-frequency water level signal missing in the model has been included in the model noise. In doing so, an artificial time-varying pressure field was introduced that was converted to water levels by means of the inverted barometer hypothesis. The spatiotemporal characteristics of these pressure fields mimic those of the missing low-frequency signal. The fields were described by an autocorrelated stochastic process whose parameters were empirically determined from experiments (not part of this study) involving the assimilation of tide gauge derived low-frequency water levels (T. Zijlker, personal communication, June, 2023). The noise in the TPJ-derived low-frequency water levels was assumed to be on the same level as the model noise, i.e., both datasets get equal weight.

The data assimilation was conducted using the open-source data assimilation toolbox OpenDA; a generic framework for parameter calibration and data assimilation applications (El Serafy et al., 2007, Van Velzen and Segers, 2010). To run the simulations in parallel, we split the model domain into 12 sub-regions. The runs itself were performed on the Dutch National Supercomputer Snellius using a total of 256 cores.

3.2. Validation

To assess changes in the model representation of the low-frequency water level variability, we compared the standard deviations (SDs) of the residual water level time series obtained with and without application of data assimilation. Here, the SDs were computed as $1.4826 \times$ the median absolute deviation (Cook and Weisberg, 1982; Rousseeuw and Croux, 1993)). An improved representation implies lower standard deviations. Note that the SD is only computed in case the time series includes at least 30 points (in case we use data from 2017 to 2021) or 15 points (in case we use data from 2017 to May 2021). At the tide gauge locations, we compute the SDs over the *monthly-mean* residual water level time series.

4. Results and Discussion

Two model simulations were conducted. The first simulation, referred to as the ‘model-only’ simulation, served as the reference. That is, the impact of the assimilation of the TPJ-derived low-frequency water levels is quantified with respect to the results obtained with the model-only simulation. In the model-only simulation, no data assimilation has been applied (remember, however, that the DCSM model has been calibrated using coastal tide gauges, see Sect. 2.5). The second model simulation assimilates the TPJ-derived low-frequency water levels. It is referred to as the ‘RA-assimilation’ simulation. The RA-assimilation simulation spans the period from January 2017 to May 2020 (see Sect. 2.1). Note that in all simulations we used a spin-up period of 10-days (i.e., the simulations started on December 22, 2016).

Besides being used as reference, the model-only simulation will also be used to assess the performance of the DCSM-derived tide-surge water levels compared to the FES tidal and DAC surge water levels. Moreover, we will use the outputs of this simulation to assess the performance of the various retracker products included in the HYDROCOASTAL data product as well as that of the in-house developed 80Hz UF-SAR product. In doing so, we will use a simulation time span covering the period from January 2017 to January 2022.

Sect. 4.1 presents and discusses the results of the performance assessments of the DCSM-derived tide-surge water levels compared to the FES tidal and DAC surge water levels as well as that of the HYDROCOASTAL retracker products and the 80Hz UF-SAR product. In Sect. 4.2, we present and discuss the results of the assimilation of the TPJ-derived low-frequency water levels.

4.1. Performance Assessments based on the model-only simulation

The results are summarized in Figs. 3–5. Fig. 3 shows per retracker product the SDs of the residual water level time series obtained after removing the DCSM tide-surge water levels (3a, 3c, and 3d) or the FES tidal and DAC surge water levels (3b), Fig. 4 shows the corresponding histograms, and Fig. 5 shows the SDs as a function of the distance to the coast. In Fig. 6, we show a map of the number of points included in the time series (6a), as well as the behavior of the number of points as a function of distance to the coast for the different retrackers (6b). The results show the following:

- The tide-surge water levels obtained with the DCSM model outperform the tidal and surge water levels obtained with the FES and DAC models included in the HYDROCOASTAL data product (cf. Figs. 3a and 3b). The median SD reduces from 16 to 6 cm (cf. Figs. 4a and 4b). In Fig. 5, we observe for larger distances to the coast a stronger increase of the median SD in case we use the FES tidal and DAC surge water levels to compute the residual water levels.

- The STARS retracker shows the best overall performance (cf. Figs. 3a, 3c, and 3d, as well as Figs. 4a, 4c, and 4d). We understand that the strong increase in SDs obtained for the ESA retracker in the north-west part of the S3 data domain (see Fig. 2b) is due to interpolation errors in generating the HYDROCOASTAL product; in some cases, the residual water levels gradually increase to tens of meters. Unfortunately, the data product does not include a proper quality flag that allows us to exclude these water levels.
- Up to 1.5 km off the coast, the median SD is lowest for the STARS retracker (differences in median SD reach approximately 7.5 cm). Between 1.5 and 10.5 km off the coast, the ESA retracker provides the lowest median SDs (differences reach approximately 6.5 cm) – it is not affected by the interpolation error at these locations. For any distance ≥ 10.5 km, the STARS retracker again provides the lowest median SDs (absolute differences compared to the ESA retracker are on the order of 1–2 cm).
- The number of water levels near the coast obtained with the STARS retracker is generally lower compared to the numbers obtained with the MWaPP and ESA retrackers (Fig. 6b).
- The number of time series including at least 30 points obtained in the Wadden Sea is low: 262 out of 1513 (STARS), 355 out of 1529 (MWaPP), and 366 out of 1536 (ESA).

The improved performance of the DCSM model in computing the *residual* water levels is explained by the fact that the model includes the nonlinear tide-surge interactions. Guarneri et al. (2023) found a comparable performance with respect to the usage of FES and DAC in case the residual water levels were computed by removing DCSM tidal water levels and DCSM surge water levels separately. Our results show that the nonlinear tide-surge interactions dominate the error budget in case we remove the tidal and surge water levels obtained with separate models. We want to emphasize that the DCSM model is not perfect; the results presented in (Guarneri et al., 2022) show that improvements in the model representation of the tidal water levels are possible. The residual water levels based on which the SDs are calculated contain, in addition to uncertainty in the altimeter data and unmodelled (baroclinic) water level variations, also these model errors.

The difference in overall performance of the STARS and MWaPP retrackers is expected; the latter is not designed for open sea waters where waves are present. The results obtained with the ESA retracker are flawed by the interpolation error in open waters away from the coast and towards the edge of the domain. We want to emphasize that the area covered by this case study is limited, as such we cannot provide conclusive findings. In addition, despite the fact that DCSM is calibrated using tide gauges, many of which are located in coastal waters, part of the increase in SDs in coastal waters is due to larger model errors. Some of the complex dynamics in these shallow waters cannot be resolved with a 2D model either. The results obtained in this study do not allow to determine the separate contributions of the uncertainty in the altimeter data, the DCSM model, and the unmodeled water level variations.

The differences in the number of obtained water levels near the coast are substantial up to 2.5 km from the coast. The STARS retracker generally provides the lowest numbers. At the same time, it shows the best performance near the coast. This suggests that it is successful in identifying bad (i.e., contaminated) data.

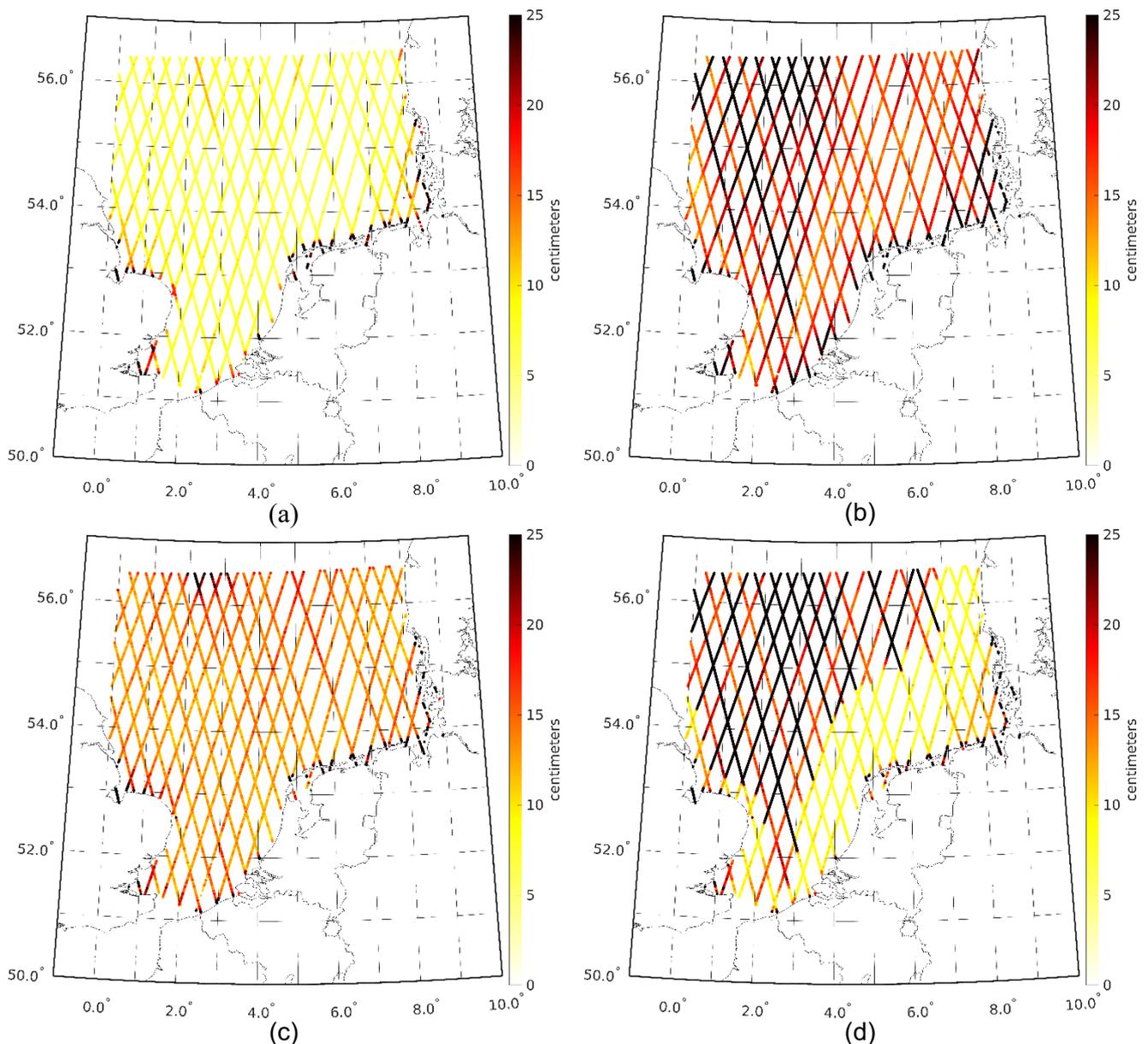


Fig. 3: SDs of the S3 residual water level time series covering the period 2017 to 2021. The S3 sea surface heights were obtained with the STARS retracker ((a) and (b)), the MWaPP retracker (c), and the ESA retracker (d). In (a), (c), and (d) the residual water levels were computed by removing the DCSM tide-surge water levels obtained with the 'model-only' simulation. In (b), we removed the FES tidal and DAC surge water levels. Note that we only included the time series for which the number of points ≥ 30 .

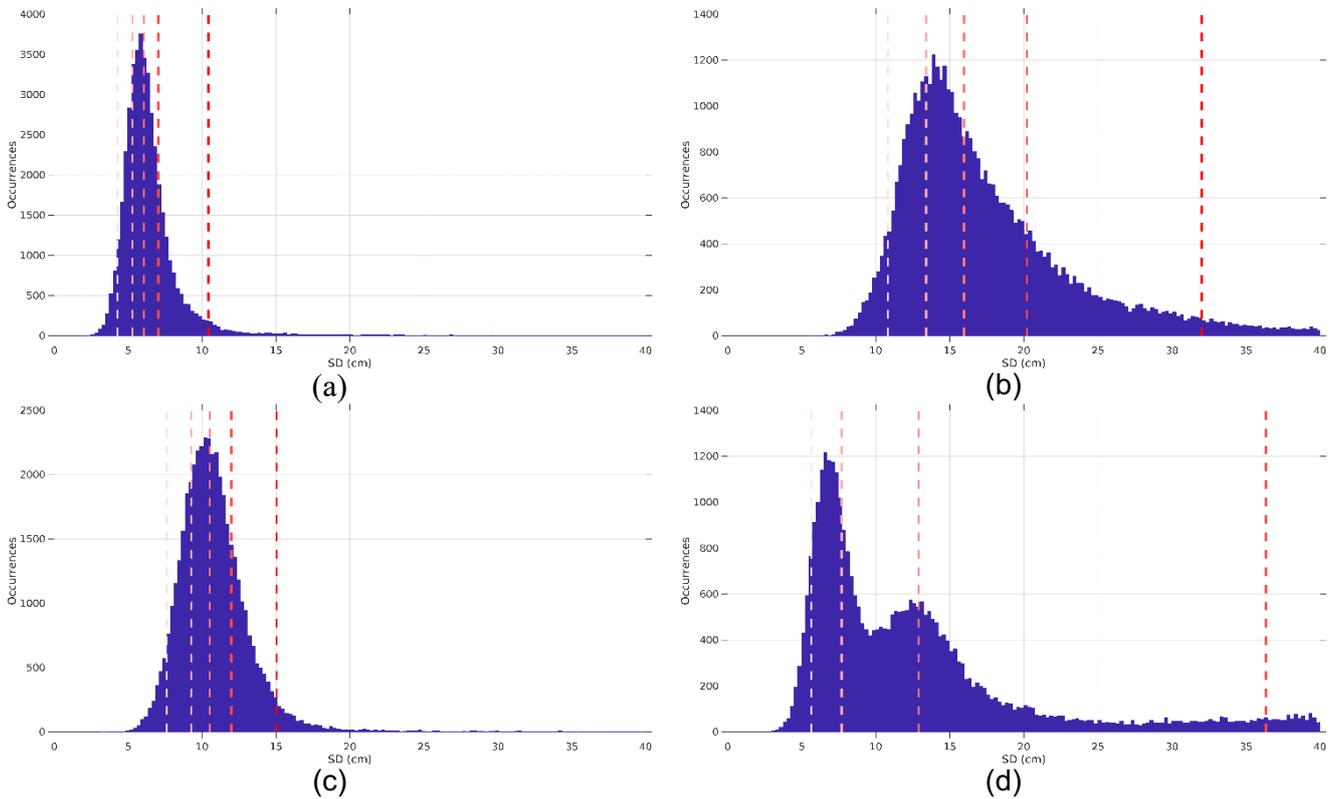


Fig. 4: Histograms of the SDs of the S3 residual water level time series covering the period 2017 to 2021 (i.e., the ones shown in Fig. 3). The dashed vertical lines show the 5, 25, 50, 75, and 95 percentiles. The S3 sea surface heights were obtained with the STARS retracker ((a) and (b)), the MWaPP retracker (c), and the ESA retracker (d). In (a), (c), and (d) the residual water levels were computed by removing the DCSM tide-surge water levels obtained with the 'model-only' simulation. In (b), we removed the FES tidal and DAC surge water levels. Note that we only included the time series for which the number of points ≥ 30 .

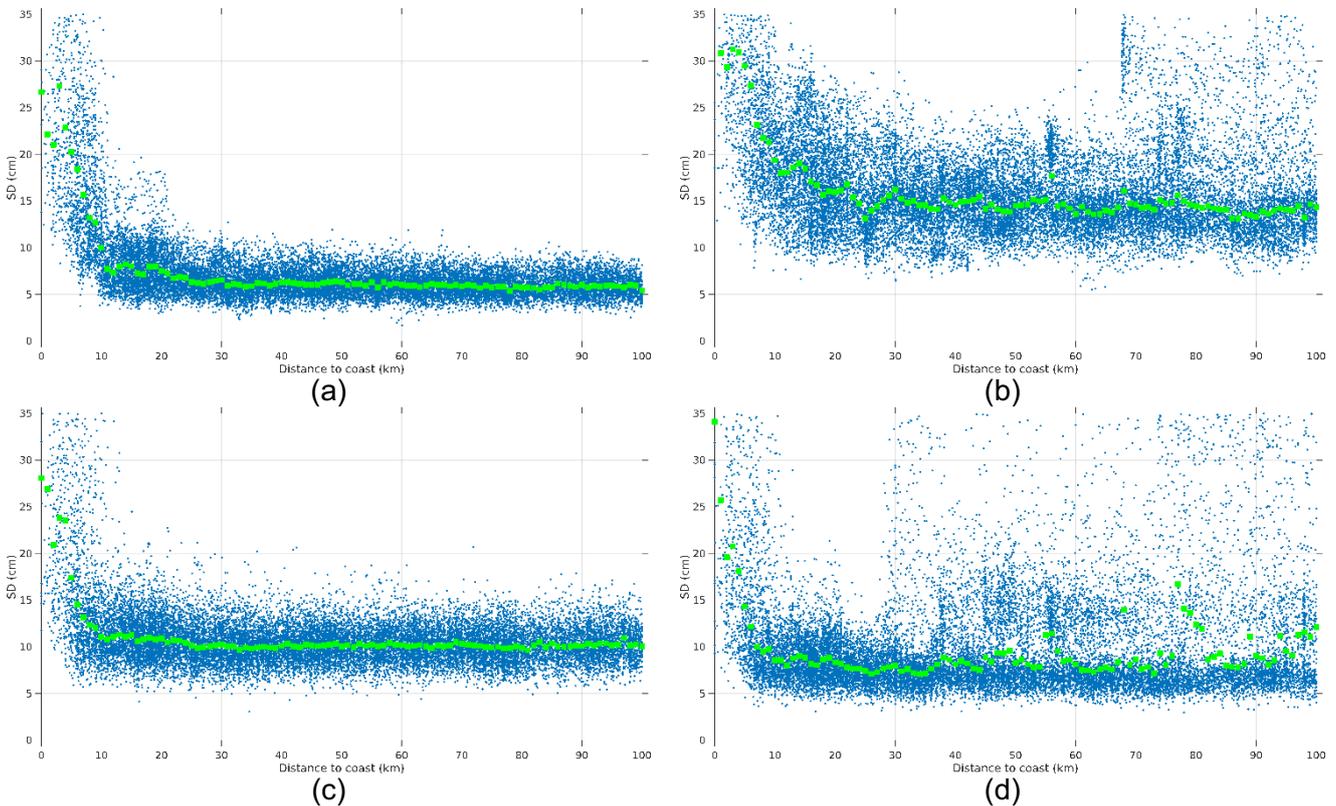


Fig. 5: SDs of the S3 residual water level time series covering the period 2017 to 2021 (i.e., the ones shown in Fig. 3) as a function of the distance to the coast. The green squares indicate the median of all values in the bin. The S3 sea surface heights were obtained with the STARS retracker ((a) and (b)), the MWaPP retracker (c), and the ESA retracker (d). In (a), (c), and (d) the residual water levels were computed by removing the DCSM tide-surge water levels obtained with the ‘model-only’ simulation. In (b), we removed the FES tidal and DAC surge water levels. Note that we only included the time series for which the number of points ≥ 30 .

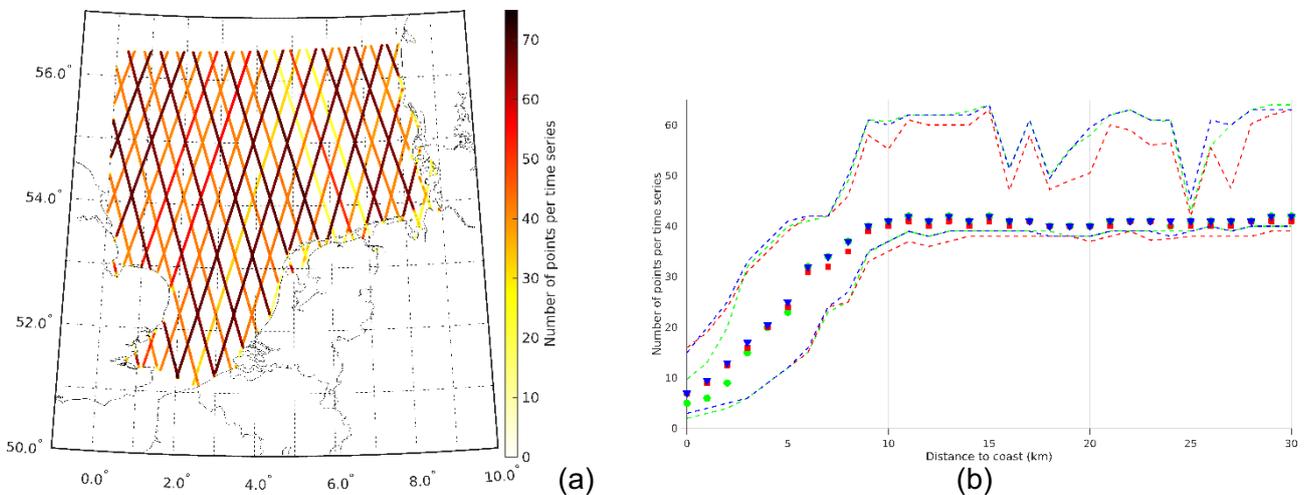


Fig. 6: Number of points per residual water level time series obtained with the STARS retracker and the DCSM-derived tide-surge water levels (a). In (b), we show the median number of points (markers) and the 25 and 75 percentiles (dashed lines) as a function of distance to the coast for the STARS (green), MWaPP retracker (red), and the ESA retracker (blue).

To assess the potential of SAR altimeter data at high posting rates (i.e., the 80Hz UF-SAR product), we repeat the analysis for the data with relative orbit number 370 (see Fig. 7a). The results are summarized in Fig. 7. The performance of the TU Delft derived 80Hz UF-SAR product is comparable to the performance of the product obtained with the ESA retracker. This was expected because we used a comparable retracker (i.e., we applied SAMOSA2 to a part of the waveform). The most striking difference is that we get closer to the coast; i.e., we retrieve more water levels near the coast (see Fig. 7c). The main reason for this is that we apply our retracker to part of the waveform only. Indeed, the median SD associated with distances up to 0.75 km from the coast seems flawed but fits the overall increase of the median SD while approaching the coast. More analysis is needed to assess whether this behavior is general or particular to this track.

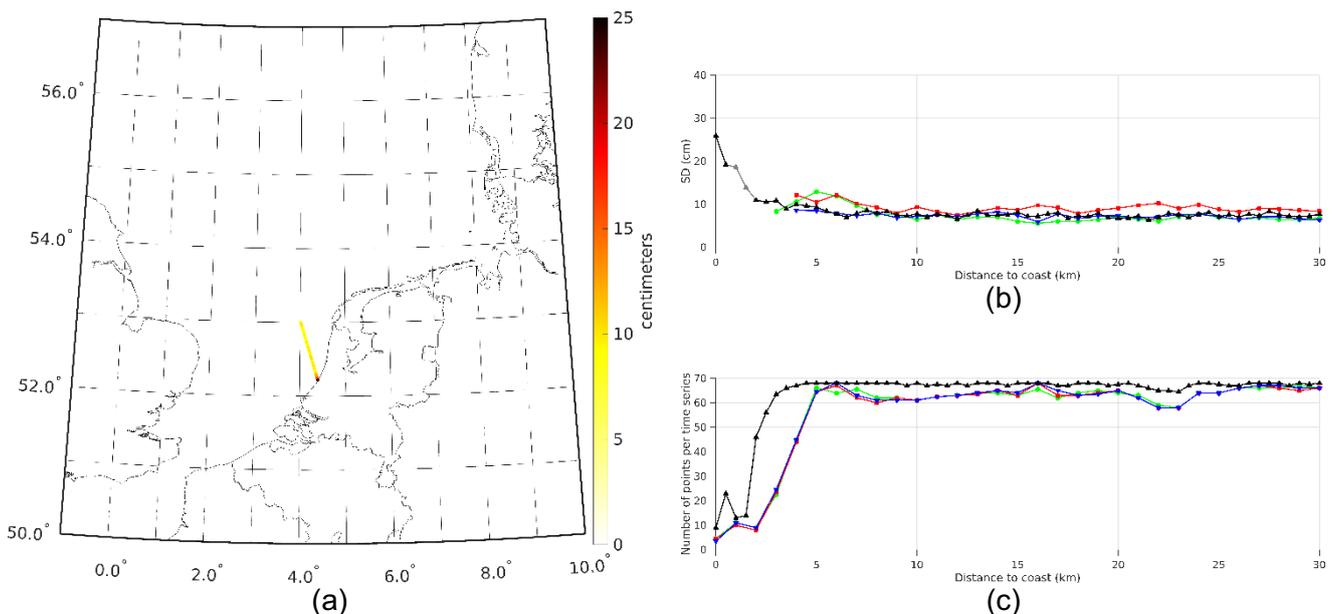


Fig. 7: Map showing the SDs of the S3 residual water level time series covering the period 2017 to 2021 obtained using the TU Delft derived 80Hz UF-SAR product (a). The residual water levels were computed by removing the DCSM tide-surge water levels obtained with the ‘model-only’ simulation. In (b), we compare the median SD as a function of distance to the coast obtained for this product (black/gray line) with the ones obtained using the STARS (green), MWaPP retracker (red), and the ESA retracker (blue). Note that we only included the time series for which the number of points ≥ 30 . **The gray points indicate time series for which the number of points ≥ 10 .** In (c), we compare the median number of points per time series as a function of distance to the coast for the various products/retrackers.

4.2. The Impact of Assimilating TPJ-derived Low-frequency Water Levels

The results of the impact of the assimilation of TPJ-derived low-frequency water levels are summarized in Figs. 8-9. Fig. 8 shows the SDs of the residual water level time series obtained without (Fig. 8a) and with (Fig. 8b) assimilating TPJ-derived low-frequency water levels. Fig. 9 shows for both scenarios the SDs of the residual monthly-mean water level time series at the tide gauge locations. The results show the following:

- There are some small changes in the SDs of the S3 residual water level time series in case data assimilation is applied (cf. Figs. 8a and 8b). The medians of all SDs for both simulations are comparable: 5.9 cm (model-only simulation) versus 6.1 cm (RA-assimilation simulation).
- The SD of the monthly-mean residual water level time series decreases at 148 out of 149 tide gauges in case data assimilation is applied. The median SD drops from 6.2 cm to 2.8 cm. The impact is observed throughout the entire model domain, but most prominent in the German Bight.
- In Fig. 9b we observe that at a few tide gauge locations the SD of the monthly-mean residual water level time series is significantly larger compared to the values at nearby tide gauges. In most cases, the model lacks performance to represent the monthly-mean water level. At tide gauge A12, which is located in the center of the North Sea (55.3833° N and 3.8° E), the large SD is caused by a spurious trend in the observation-derived time series. This likely points to errors in the tide gauge record.

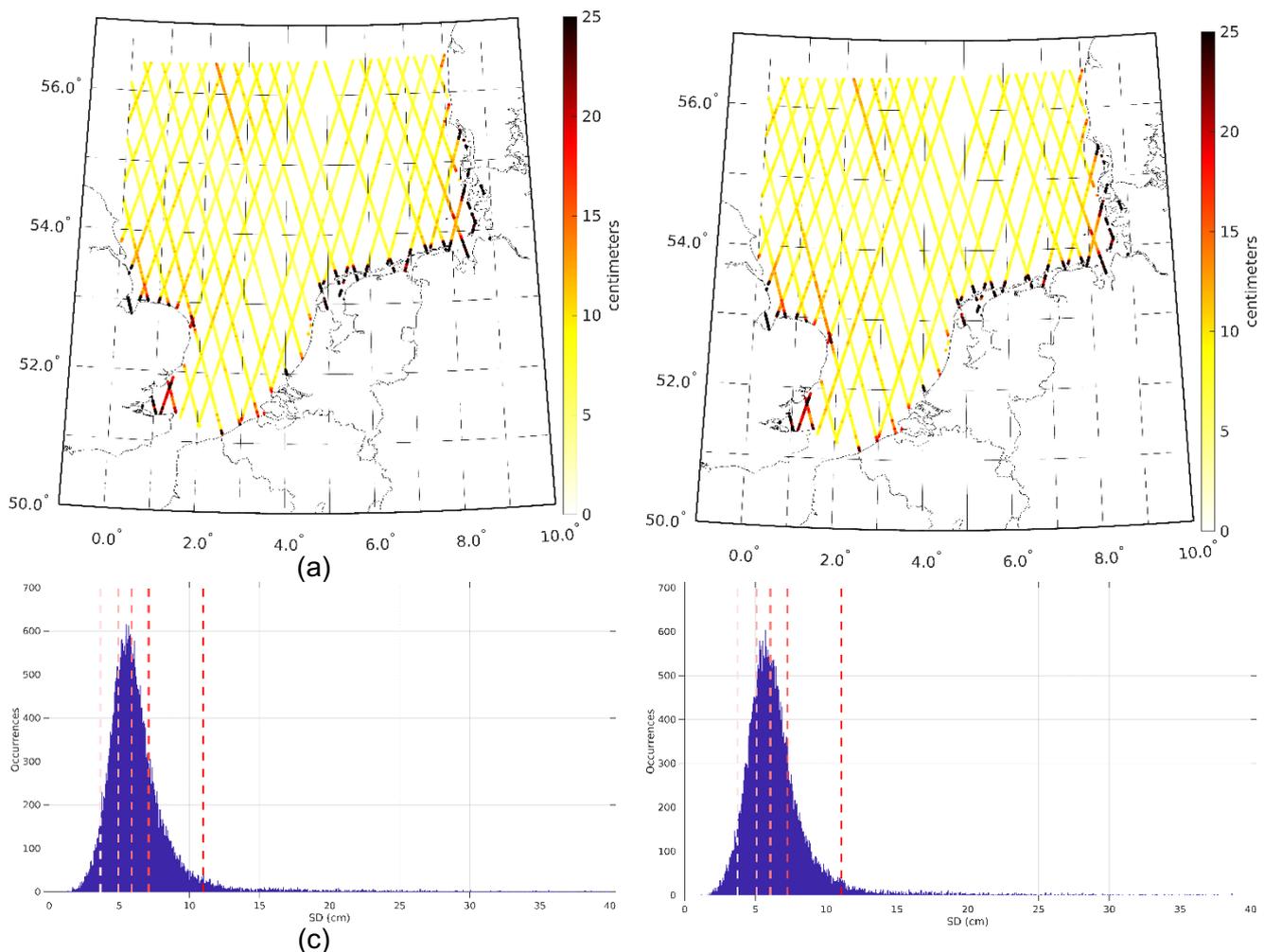


Fig. 8: SDs of the S3 residual water level time series obtained without (a) and with (b) assimilating TPJ-derived low-frequency water levels and the corresponding histograms ((c) and (d)). The S3-derived water levels are obtained with the STARS retracker. **The time series cover the period January 2017 until May 2020.** Note that we only included the time series for which the number of points ≥ 15 .

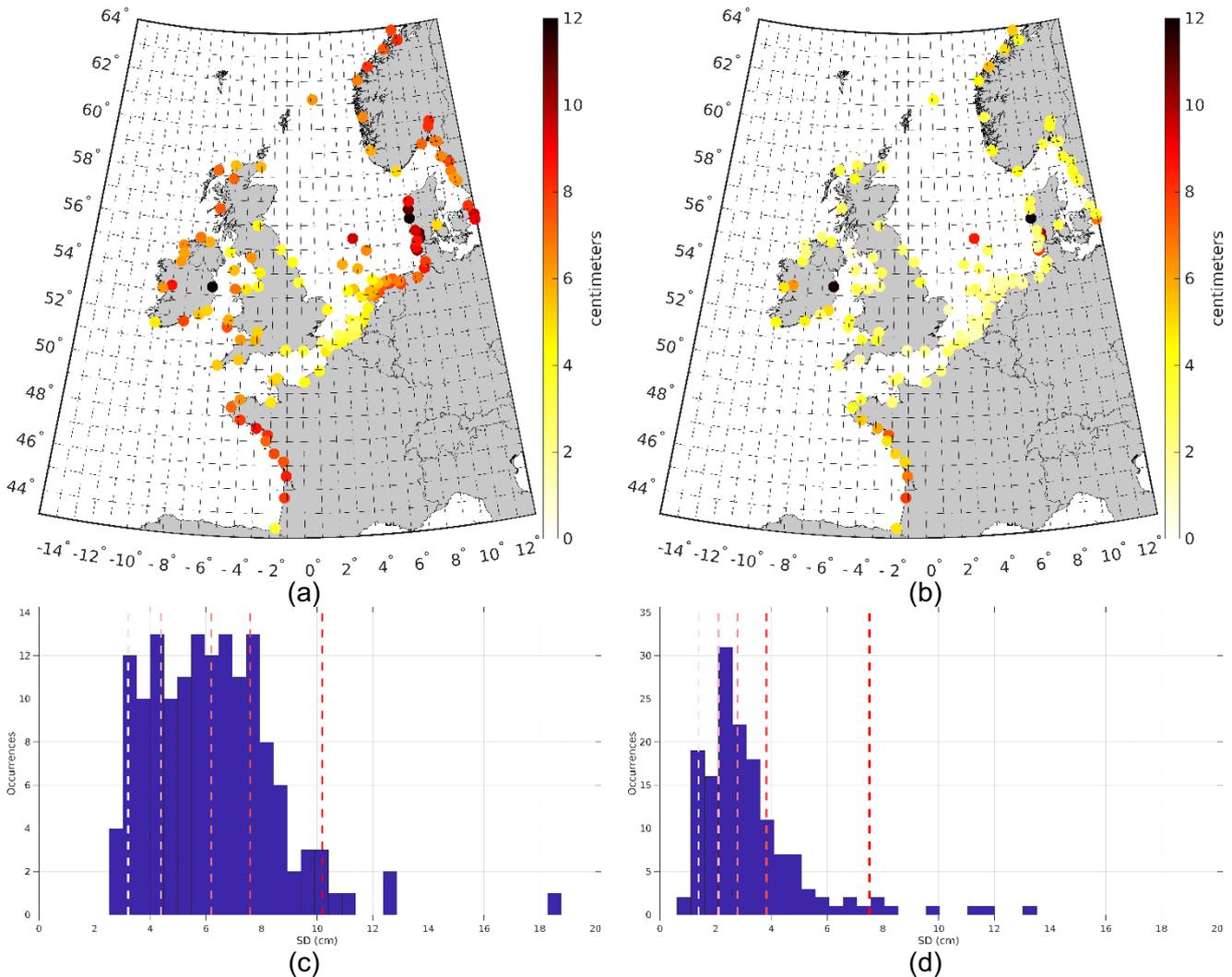


Fig. 9: SDs of the monthly-mean residual water level time series obtained without (a) and with (b) assimilating TPJ-derived low-frequency water levels and the corresponding histograms ((c) and (d)). The dashed vertical lines in the histograms indicate the 5, 25, 50, 75, and 95 percentiles. **The time series cover the period January 2017 until May 2020.** Note that we only included the time series for which the number of points > 15.

The most probable explanation for the observed differences in impact between the S3-derived results and the tide gauge derived results is the fact that the S3 residual water levels still contain the tidal errors in the DCSM-derived tide-surge water levels. At the tide gauges, these average out when computing the monthly-mean water levels. Due to the poor temporal sampling we cannot do this with the S3 time series. Note that the length of the S3-derived time series is anyway low, in particular for the S3b satellite. The impact of the data assimilation itself is substantial. Whether or not there is any advantage of using the altimeter-derived low-frequency water levels in the assimilation over tide gauge derived is the topic of future studies.

5. Summary and Conclusion

To exploit the wealth of satellite radar altimeter data in calibrating the regional, high-resolution 2D tide-surge DCSM model covering the northeast Atlantic including the North Sea and Wadden Sea requires an approach that can be applied to the separate water level variability contributors. In this study, we aim to improve DCSM's ability in representing the low-frequency water level variability by assimilating data acquired by the TOPEX/Poseidon and Jason (TPJ) satellites. This variability, caused by physical processes not included in the model's governing equations or forcing terms, is a major source of errors in the operational forecasting of water levels. To validate the impact of the data assimilation, we used i) S3-derived water levels acquired over the southern North Sea and Wadden Sea that were produced in the context of the HYDROCOASTAL project, and ii) tide gauge records required at 149 locations throughout the DCSM model domain. A secondary objective is to assess the performance of i) the DCSM-derived tide-surge water levels compared to the FES tidal and DAC surge water levels, ii) the various retracker products included in the HYDROCOASTAL data product, and iii) an in-house developed unfocussed SAR product produced at a posting rate of 80 Hz.

From the results and subsequent analysis, we conclude:

- 1) Computing sea level anomalies (i.e., water levels corrected for tides and surge) requires consideration of the nonlinear tide-surge interactions. Over the North Sea, well known for the nonlinear interactions between tides and surge, we observed a 10 cm (i.e., more than 50%) reduction of the median SD when using the DCSM-derived tide-surge water levels compared to the use of the FES tidal and DAC surge water levels included in the HYDROCOASTAL product.
- 2) In the coastal waters of the southern North Sea, the STARS retracker provides the lowest median SDs of the residual water level time series up to 1.5 km off the coast (differences in median SD reach approximately 7.5 cm) and beyond 10.5 km off the coast (differences in median SD compared to the ESA retracker are on the order of 1–2 cm). Between 1.5 and 10.5 km it is the ESA retracker (differences in median SD reach approximately 6.5 cm). In the coastal waters, the STARS retracker provides the lowest number of water levels.
- 3) Outside of coastal waters, the results of the ESA retracker in the HYDROCOASTAL product are in many cases unreliable, due to the previously noted interpolation error. Water levels gradually rise to tens of meters. The quality flag included in the product does not label these data as unreliable.
- 4) Processing SAR data at higher posting rates and using a sub-waveform retracker may allow more data to be retrieved in coastal waters. In this study, we analyzed the potential of an in-house developed unfocussed SAR product generated at a posting rate of 80 Hz. In doing so, we used all data acquired along the track with relative orbit number 370. The median SDs as a function of distance to the coast are comparable to the ones obtained with the ESA retracker. We do, however, obtain significantly more data. The key reason for this is the fact that we applied our retracker to part of the waveform.
- 5) The impact of the assimilation is substantial. At the tide gauge locations, the median SD of the residual monthly-mean water levels reduced from 6.2 cm to 2.8. The impact cannot be assessed from the HYDROCOASTAL data. The most likely explanation is the fact that these data are still impacted by the tidal errors in the DCSM-derived tide-surge water levels.

Based on our findings, we recommend to:

1. Provide tide-surge water level corrections in future (coastal) altimeter products obtained from a model considering the nonlinear interactions.

2. Assess the potential of SAR altimeter data processed at higher posting rates. In this study, we only considered one track. More research is needed to determine whether usage of higher posting rates generally results in more data.
3. Assess the performance of assimilating TPJ-derived low-frequency water levels compared to the performance of the approach that relies on coastal tide gauges.
4. Note the problem with the ESA data in open water towards the edge of the domain, due to interpolation errors, advise other users of the product and correct for future products.

6. References

Asher, T.G., Luettich, R.A., Jr., Fleming, J.G., and Blanton, B.O.: Low frequency water level correction in storm surge models using data assimilation. *Ocean Model.* 144, 101483. <https://doi.org/10.1016/j.ocemod.2019.101483>, 2019.

Bij de Vaate, I., Martin, E., Slobbe, D.C., Naeije, M., and Verlaan, M.: Lead Detection in the Arctic Ocean from Sentinel-3 Satellite Data: A Comprehensive Assessment of Thresholding and Machine Learning Classification Methods, *Marine Geodesy*, 45:5, 462-495, DOI: [10.1080/01490419.2022.2089412](https://doi.org/10.1080/01490419.2022.2089412), 2022.

Birol, F., Fuller, N., Lyard, F., Cancet, M., Niño, F., Delebecque, C., Fleury, S., Toubanc, F., Melet, A., Saraceno, M., and Léger, F.: Coastal applications from nadir altimetry: Example of the X-TRACK regional products, *Advances in Space Research*, 59, 936–953, <https://doi.org/https://doi.org/10.1016/j.asr.2016.11.005>, 2017.

Carrère, L., Lyard, F., Cancet, M., Guillot, A., and Roblou, L.: FES2012: A New Global Tidal Model Taking Advantage of Nearly 20 Years of Altimetry, in: *20 Years of Progress in Radar Altimetry*, edited by Ouwehand, L., vol. 710 of ESA Special Publication, p. 13, 2013.

Cook RD, Weisberg S (1982) *Residuals and influence in regression*. Chapman and Hall, New York

Egido, A., Dinardo, S., Ray, C., 2021. The case for increasing the posting rate in delay/Doppler altimeters. *Adv. Space Res.* 68 (2), 930–936. <https://doi.org/10.1016/j.asr.2020.03.014>.

Ehlers, F., Schlembach, F., Kleinherenbrink, M., & Slobbe, D.C. Validity assessment of SAMOSA retracking for fully-focused SAR altimeter waveforms. *Advances in Space Research*, 71(3), 1377-1396. <https://doi.org/10.1016/j.asr.2022.11.034>, 2023a.

Ehlers, F., Slobbe, C., Verlaan, M., & Kleinherenbrink, M. A noise autocovariance model for SAR altimeter measurements with implications for optimal sampling. *Advances in Space Research*, 71(10), 3951-3967. <https://doi.org/10.1016/j.asr.2023.02.043>, 2023b.

El Serafy, G. Y., H. Gerritsen, S. Hummel, A. H. Weerts, A. E. Mynett, and M. Tanaka: Application of data assimilation in portable operational forecasting systems—the DATools assimilation environment. *Ocean Dynamics*, 57 (4): 485–499. doi: 10.1007/s10236-007-0124-3, 2007.

Guarneri, H., Verlaan, M., Slobbe, D. C., Zijl, F., Pietrzak, J. D., Keyzer, L. M., & Klees, R.: Improved shallow waters tidal estimates using satellite radar altimetry data and numerical modeling. Poster session presented at Ocean Surface Topography Science Team meeting 2022, Venice, Italy. <https://doi.org/10.24400/527896/a03-2022.3605>, 2022.

Guarneri, H., Verlaan, M., Slobbe, D. C., Veenstra, J., Zijl, F., Pietrzak, J., M., S., Keyzer, L., Afrasteh, Y., and Klees, R.: The impact of nonlinear tide–surge interaction on satellite radar altimeter-derived tides, *Mar. Geod.*, 46, 251–270, <https://doi.org/10.1080/01490419.2023.2175084>, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.

Kalman, R. E.: A new approach to linear filtering and prediction problems. 1960

Leendertse, J. J.: Aspects of a Computational Model for Long-period Water-wave Propagation, Rand Corporation for the United States Air Force Project Rand, 1967.

NASA Ocean Biology Processing Group (OBPG) and R.P. Stumpf. 2012. Distance to Nearest Coastline: 0.01-Degree Grid: Ocean. Distributed by the Pacific Islands Ocean Observing System (PacIOOS). http://pacioos.org/metadata/dist2coast_1deg_ocean.html. Accessed May 14, 2019.

Ralston, M. L. and Jennrich, R. I.: Dud, A Derivative-Free Algorithm for Nonlinear Least Squares, *Technometrics*, 20, 7–14, <http://www.jstor.org/stable/1268154>, 1978.

Rousseeuw PJ, Croux C (1993) Alternatives to the median absolute deviation. *J Am Stat Assoc* 88(424):1273–1283. <https://doi.org/10.1080/01621459.1993.10476408>.

Sakov, P., Evensen, G., Bertino, L.: Asynchronous data assimilation with the EnKF, *Tellus A: Dynamic Meteorology and Oceanography*, 62:1, 24-29, DOI: 10.1111/j.1600-0870.2009.00417.x, 2010.

Slobbe D.C., Klees R., Farahani H.H., Huisman L., Alberts B., Voet P., De Doncker F.: The Impact of Noise in a GRACE/GOCE Global Gravity Model on a Local Quasi-Geoid. *Journal of Geophysical Research: Solid Earth* 124(3):3219–3237, DOI 10.1029/2018jb016470, 2019.

Stelling, G. S.: On the construction of computational methods for shallow water flow problems, Ph.D. thesis, Delft University of Technology, Delft, Rijkswaterstaat Communications 35, 1984.

Van Velzen, N. and A. J. Segers.: A problem-solving environment for data assimilation in air quality modelling. *Environmental Modelling & Software*, 25 (3): 277–288. doi: 10.1016/j.envsoft.2009.08.008, 2010.

Zijl, F., Verlaan, M., and Gerritsen, H.: Improved water-level forecasting for the Northwest European Shelf and North Sea through direct modelling of tide, surge and non-linear interaction, *Ocean Dynam.*, 63, 823–847, <https://doi.org/10.1007/s10236-013-0624-2>, 2013.

Zijl, F., Sumihar, J., and Verlaan, M.: Application of data assimilation for improved operational water level forecasting on the northwest European shelf and North Sea, *Ocean Dynam.*, 65, 1699–1716, <https://doi.org/10.1007/s10236-015-0898-7>, 2015.

Zijl, F. and Groenenboom, J.: Development of a sixth-generation model for the NW European Shelf (DCSM-FM 0.5nm), Tech. rep., Deltares, available online at: https://publications.deltares.nl/11203715_004.pdf (accessed July 18, 2022), 2019.