# Error Characterization of Significant Wave Heights in Multidecadal Satellite Altimeter Product, Model Hindcast, and In Situ Measurements Using the Triple Collocation Technique

GUILLAUME DODET,[a] SALEH ABDALLA,[b] MATIAS ALDAY,[a] MICKAËL ACCENSI,[a] JEAN BIDLOT,[b] AND FABRICE ARDHUIN[a]

[a] *Univ. Brest, CNRS, Ifremer, IRD, Laboratoire d'Océanographie Physique et Spatiale, IUEM, Plouzané, France*
[b] *European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom*

(Manuscript received 6 January 2022, in final form 18 March 2022)

ABSTRACT: Ocean wave measurements are of major importance for a number of applications including climate studies, ship routing, marine engineering, safety at sea, and coastal risk management. Depending on the scales and regions of interest, a variety of data sources may be considered (e.g., in situ data, Voluntary Observing Ship observations, altimeter records, numerical wave models), each one with its own characteristics in terms of sampling frequency, spatial coverage, accuracy, and cost. To combine multiple source of wave information (e.g., for data assimilation scheme in numerical weather prediction models), the error characteristics of each measurement system need to be defined. In this study, we use the triple collocation technique to estimate the random error variance of significant wave heights from a comprehensive collection of collocated in situ, altimeter, and model data. The in situ dataset is a selection of 122 platforms provided by the Copernicus Marine Service In Situ Thematic Center. The altimeter dataset is the ESA Sea State CCI version1 L2P product. The model dataset is the WW3-LOPS hindcast forced with bias-corrected ERA5 winds and an adjusted T475 parameterization of wave generation and dissipation. Compared to previous similar analyses, the extensive (~250 000 entries) triple collocation dataset generated for this study provides some new insights on the error variability associated to differences in in situ platforms, satellite missions, sea state conditions, and seasonal variability.

KEYWORDS: Sea state; Buoy observations; Satellite observations; Uncertainty; Numerical analysis/modeling

## 1. Introduction

Historical wave data are important for many scientific and engineering applications (climate science, marine safety, coastal and offshore structures, coastal risk management). The sea state is observed by a variety of instrument networks, including Voluntary Observing Ships, seismometers, moored and drifting buoys, satellite altimeters, and satellite-borne synthetic aperture radars (Ardhuin et al. 2019). In addition, wave model hindcasts and reanalysis provide accurate gridded wave information at global scale and over several decades, information that gets better with every new reanalysis thanks to constantly improving physics parameterization, resolution, forcing accuracy, and numerical methods (e.g., Hersbach et al. 2020; Alday et al. 2021; Law-Chune et al. 2021). These different sources complement each other but there is a need to improve the standardization of data format, metadata, and uncertainty characterization among these sources (Merchant et al. 2017). In particular, an understanding of errors and uncertainties is crucial in order to make the most effective and appropriate use of a given dataset whether for analysis, calibration, validation, or assimilation purposes (Parker 2016). For instance, numerical weather prediction models benefit from wave data assimilation using both significant wave height (Hs) from satellite altimeters (Abdalla and Janssen 2017) and wave spectra partitions from other radar systems (Aouf et al. 2021) the assimilation process is made most efficient when the properties of the measurement uncertainties are fully understood.

Triple collocation (TC) is a robust technique used to assess random error variance in three independent measurement systems. Initially introduced to oceanic and meteorological communities by Stoffelen (1998) to estimate error characteristics in surface wind speed from in situ anemometer, satellite scatterometer, and model forecast, the technique has since been applied to a wide range of geophysical parameters, such as wind stress (Portabella and Stoffelen 2009), sea surface temperature (O'Carroll et al. 2008), or soil moisture (Gruber et al. 2016). Regarding ocean wave measurements, Caires and Sterl (2003) compared ERA-40 model reanalysis, in situ measurements and *European Remote-Sensing Satellite 1* (*ERS-1*) and TOPEX altimeter measurements and found the largest errors in the model results and the lowest errors in the in situ measurements. They also identified a strong negative bias in *ERS-1* measurements. Janssen et al. (2007) applied the TC technique to several combinations of in situ buoys, *ERS-1*, *ERS-2*, and *Envisat* altimeter data and ECMWF wave model forecasting data (with and without altimeter data assimilation), and demonstrated the importance of making use of an independent dataset to obtain consistent error statistics. Contrary to Caires and Sterl (2003), they obtained the largest errors in the in situ buoy data and the lowest errors in the

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JTECH-D-21-0179.s1.

*Corresponding author*: Guillaume Dodet, guillaume.dodet@ifremer.fr

DOI: 10.1175/JTECH-D-21-0179.1

model analysis. Later, Abdalla et al. (2011) applied the TC technique to *Jason-1*, *Jason-2*, and *Envisat* altimeter data, in situ buoy data, and a stand-alone version of the ECMWF wave model forecast (with no altimeter data assimilation), in order to ensure the independence criterion between model and altimeter data. They obtained the lowest error for *Jason-2* (5.4%), followed by *Envisat* (6.2%), *Jason-1* (7.8%), buoys (8.6%), and the stand-alone model hindcast (9.7%). They also demonstrated that these error results were significantly dependent on the selected distance criteria between altimeter and in situ matchups, which controls the total number of collocated measurements. Later, Abdalla and De Chiara (2017) showed that TC wind speed error results were robust if at least a few thousand matchups were considered. With the available wave buoy network, this number can be achieved on an annual basis using a 50-km distance criterion, although relaxing the distance criterion to 200 km is advisable to quadruple the number of available matchups (Abdalla et al. 2011). More recently, Dodet et al. (2020) computed the error statistics for 10 calibrated altimeter missions based on dual comparisons with in situ measurements and found that *ERS-1* was biased low (~5%) while *Envisat*, *Jason-2*, *CryoSat-2*, *SARAL*, and *Jason-3* were biased high (5%–10%). Comparisons with an independent wave hindcast (no assimilation) revealed normalized root-mean-squared error (NRMSE) between 9% and 13% with largest errors for *ERS-1* and *ERS-2* and lowest errors for *SARAL* and *CryoSat-2*. All these results tend to indicate that the error estimates of in situ, altimeter and model Hs data and the relative performance between these systems are not unequivocally defined yet and requires further investigation. Moreover, the payloads of in situ platforms and altimeter missions have been regularly evolving over the last decades, together with their measuring capabilities, which requires dedicated analysis of the temporal, interplatform, and intermission error measurement variability. A major difficulty in performing error analysis with the TC technique is the need to have a sufficiently high number of collocated measurements between independent systems to produce robust error estimates. In this study, we have developed an extensive triple collocation dataset (~250 000 entries) over the period 1993–2018, based on three recently delivered datasets, namely,

- the multimission ESA Sea State CCI version 1 altimeter dataset (Dodet et al. 2020);
- a selection of in situ wave platforms from the Copernicus Marine Service In Situ Thematic Assembly Center (CMEMS INSTAC) database (http://www.marineinsitu.eu/);
- and a WW3 wave model hindcast forced by ERA5 winds implemented at Ifremer/LOPS (Alday et al. 2021).

The main objective of this study is to investigate the Hs error characteristics in these three multidecadal datasets to better constrain error variability associated with differences in satellite missions, in situ platforms, sea state conditions, and seasonal variability. A secondary objective is to perform an assessment of systematic and random errors in the ESA Sea State CCI version 1 dataset. The primary error assessment method employed to accomplish the objectives is the triple collocation technique.

The manuscript is organized as follows: the three datasets are described in section 2, section 3 describes the methods used to generate the triple collocation dataset and compute error estimates using the TC technique, section 4 presents the results of the triple collocation analysis, and section 5 discusses and summarizes the results.

## 2. Wave height data sources

In this section, we describe the three data sources we used to derive the triple collocation dataset and estimate the measurement errors.

### *a. Altimeter data from the Sea State CCI v1 dataset*

The ESA Sea State CCI v1 dataset (Piollé et al. 2020) is a multimission altimeter product that spans over the period 1992–2018, and covers the following missions (ranked by date of launch): *ERS-1*, TOPEX, *ERS-2*, *Geosat Follow-On* (*GFO*), *Jason-1*, *Envisat*, *Jason-2*, *CryoSat-2* (low-resolution mode acquisition only), *SARAL*, and *Jason-3*. The L2P product contains three flavors of 1-Hz along-track significant wave height records: (i) the raw significant wave height as provided in the Geophysical Data Records (GDR), (ii) the calibrated significant wave height that was calibrated using buoy matchups and mission crossovers, and (iii) the denoised significant wave height after application of the EMD-based denoising technique developed by Quilfen and Chapron (2021) on the calibrated significant wave height. Further information on this dataset can be found in Dodet et al. (2020). For this study, we have considered the significant wave heights calibrated with (ii), and, for specific analysis, denoised with (iii).

### *b. In situ wave platforms from CMEMS INSTAC*

Several thousands of in situ platforms have been deployed in oceans, seas, and lakes over the last 50 years in order to collect information on wave and other met/ocean parameters. These platforms are often operated by state agencies, following different procedures, making the collection, analysis and intercomparison of the recorded measurements particularly challenging. The INSTAC is a component of the European CMEMS and its role is to ensure consistent and reliable access to a range of in situ data for service production and validation. For this purpose, CMEMS INSTAC collect multisource/multiplatform data, and perform consistent quality control before distributing the data in a common format to the CMEMS Marine Forecasting Centres (MFC). The data can be found at http://www.marineinsitu.eu/.

Two types of products are available: (i) near-real-time (NRT) products are automatically quality controlled within 24 h from acquisition to feed MFC for their forecasting activities and (ii) reprocessed (REP) products are built for reanalysis activities or climate research and provide integrated products over the past 25–50 years. These latter are assessed by scientific team and suspicious measurement are checked to discriminate sensor anomaly from real ocean signal. For this study, we have used the reprocessed CMEMS INSTAC

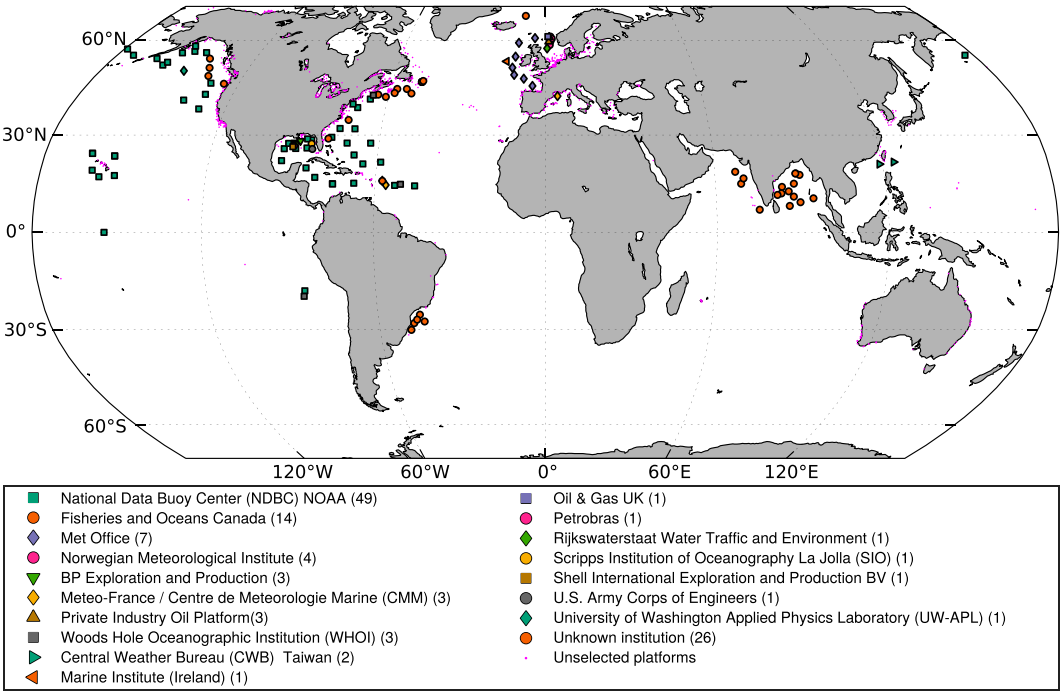


FIG. 1. Location of all available CMEMS INSTAC reprocessed in situ wave platforms (pink dots) and selected platforms (colored symbols) located >100 km from the coast, in >50-m water depth, and deployed over at least one year during the 1992–2018 period. The name of the main providers and the number of associated platforms are given in the legend.

products. Among the 1180 available in situ wave platforms, only a subset of 122 platforms was selected, based on the following criteria: the platforms being located more than 100 km offshore, in more than 50-m water depth, and deployed for at least one year over the period 1992–2018. The distance to the coast was estimated from the Open Street Map coastline polygons (https://osmdata.openstreetmap.de/data/coastlines.html) and the depth was computed from the IOC/IHO General Bathymetric Chart of the Oceans (GEBCO) 2020 gridded data (15 arc s) interpolated at the buoy location (https://www.gebco.net/data_and_products/gridded_bathymetry_data/). For most long-term operating platforms, the nominal location changes during the lifetime of the platform, and a new position is generally set after maintenance operations. While the distance between subsequent positions is often small enough (a few kilometers) to consider that the platform is measuring similar sea state conditions, it may be much larger in some cases (e.g., >500 km). Unfortunately, the mode of acquisition, the accuracy, and the resolution of the position information, when they are provided, strongly differ from one platform/provider to the other. Using the latitude–longitude information in the CMEMS INSTAC dataset, we found that 50% of the 1180 wave platforms are given a single position for the whole deployment period, including the 75 platforms deployed over more than 10 years. Among the 591 platforms for which time varying positions are provided, more than 15% show position shifts exceeding 100 km. In some cases, the position time series may show singularities that cannot correspond to any realistic displacement. To reject outlier position and make use of the time-varying position in the in situ platform–altimeter matchup analysis, we computed for each platform stepwise segments of constant position separated by a minimum of 10 km and we rejected the wave measurements for which the position was suspicious (singularity) or indicated some drifting. Median position of the CMEMS INSTAC wave platforms are shown in Fig. 1 and maximum distance between real and median position for the 122 selected buoys are provided in Table 1 of the online supplemental material.

### c. *Model outputs from WW3-LOPS hindcast*

The implemented WW3 model employs a global grid with spatial resolution of 0.5° extending from latitudes −78° to 83°. Spectral discretization considers 24 directions, equivalent to a directional resolution of 15°, and 36 frequencies exponentially spaced with a 1.1 increment factor (ranging from 0.034 to 0.95 Hz).

The model was forced using hourly surface wind fields from the ECMWF fifth-generation reanalyses ERA5 (Hersbach et al. 2020). Surface currents are included to take into account wave–current interactions at global scale. The current fields, which are the sum of geostrophic and Ekman components, were taken from the CMEMS-Globcurrent product (MULTIOBS_GLO_PHY_REP_015_004). These 3-hourly fields were generated based on the method by Rio et al. (2014) with the updated mean dynamic topography from CNES-CLS (Mulet et al. 2021). Finally, ice daily concentration considering a 1 m thickness, and icebergs distribution are taken, respectively, from the

TABLE 1. Absolute and relative errors of each measurement system for different data processing: (single obs) single point in situ and altimeter observations are used; (superobs) altimeter and in situ superobservations are used; (no EMD) EMD denoising is not applied to altimeter observations; (EMD) EMD denoising is applied to altimeter observations.

| | Absolute error (m) | | | | Normalized error (%) | | | |
|---|---|---|---|---|---|---|---|---|
| System | Single obs, no EMD | Superobs, no EMD | Single obs, EMD | Superobs, EMD | Single obs, no EMD | Superobs, no EMD | Single obs, EMD | Superobs, EMD |
| In situ | **0.17** | **0.14** | 0.17 | **0.15** | **7.5** | **6.5** | 7.3 | **6.5** |
| Altimeter | 0.22 | 0.16 | **0.16** | **0.15** | 9.8 | 7.3 | **7.0** | 6.6 |
| Model | 0.23 | 0.23 | 0.23 | 0.23 | 10.2 | 10.2 | 10.1 | 10.0 |

Ifremer SSMI-derived product (Girard-Ardhuin and Ezraty 2012) and the Ifremer-Altiberg icebergs distribution database (Tournadre et al. 2016).

Wave generation and swell dissipation parameterizations correspond to the wind input and wave dissipation source terms described in Ardhuin et al. (2010) and Leckler (2013), including the parameters' adjustments and wind input corrections from test T475 detailed in Alday et al. (2021). Partial blocking of wave energy by icebergs is considered with the parameterizations from Ardhuin et al. (2011). Wave evolution and nonlinear (wave to wave) interactions are described with the discrete interaction approximation (DIA) (Hasselmann and Hasselmann 1985; Hasselmann et al. 1985). The generated hindcast covers a total of 28 years, from 1993 to 2020 with data output each 3 h. No data assimilation was used in the generation of this dataset. This new model hindcast appears to be generally more accurate than the previous versions of the WW3-LOPS hindcast (e.g., Rascle and Ardhuin 2013), and can be more accurate than the ERA5 Hs estimates, in particular in strong current regions and for Hs $>$ 7 m. More details on the parameterizations, numerical implementation, and validation can be found in Alday et al. (2021) and in the WW3 user manual (WAVEWATCH III Development Group 2019).

## 3. Methods

### a. Altimeter–in situ platform matchups

Matchups between altimeter and in situ platform measurements were computed within 100-km distance and 1-h time window. The along-track altimeter and in situ Hs observations were smoothed with 50-km and 2-h running average, respectively, in order to reduce high-frequency variability and match the 0.5° model resolution. Indeed, assuming an average wave group velocity of 7 m $s^{-1}$ (corresponding to ~9-s wave period in deep water), a 2-h in situ record will "see" wave energy that has traveled over ~50 km. Finally, the 3-hourly model outputs were linearly interpolated in space at the buoy location and at the time of the satellite–in situ matchup.

### b. Error estimation from triple collocation

Following Vogelzang and Stoffelen (2012), we consider three data sources $X$, $Y$, and $Z$ giving collocated estimates ($x$, $y$, $z$) of the true (unknown) quantity $t$ with associated random errors $\epsilon_i (i = x, y, z)$. We assume that $X$ is a bias-free calibrated reference system while $Y$ and $Z$ are subjected to calibration errors, which are linearly related to $t$ for the whole range of values under consideration. The measurements can, therefore, be written:

$$x = t + \epsilon_x, \tag{1}$$

$$y = a_y t + b_y + \epsilon_y, \tag{2}$$

$$z = a_z t + b_z + \epsilon_z, \tag{3}$$

where $a_i$ and $b_i$ $(i = y, z)$ are the least squares slopes and intercepts. We assume that the random errors $\epsilon_i$ are unbiased ($\langle \epsilon_i \rangle = 0$, with the angle brackets $\langle \cdot \rangle$ representing the average over all data), uncorrelated with the true signal ($\langle \epsilon_i t \rangle = 0$), and uncorrelated with each other ($\langle \epsilon_i \epsilon_j \rangle = 0$ for $i \neq j$). Under these assumptions, it is possible to estimate the variances of the random error $\langle \epsilon_i^2 \rangle$ as a function of the variances $C_{ii}$ and covariances $C_{ij}$ of the signals, as follows:

$$\langle \epsilon_x^2 \rangle = C_{xx} - \frac{C_{xy} C_{xz}}{C_{yz}}, \tag{4}$$

$$\langle \epsilon_y^2 \rangle = C_{yy} - \frac{C_{xy} C_{yz}}{C_{xz}}, \tag{5}$$

$$\langle \epsilon_z^2 \rangle = C_{zz} - \frac{C_{xz} C_{yz}}{C_{xy}}. \tag{6}$$

Moreover, the calibration coefficients can be computed as

$$a_y = \frac{Cyz}{C_{xz}}, \tag{7}$$

$$a_z = \frac{Cyz}{C_{xy}}, \tag{8}$$

$$b_y = \langle y \rangle - a_y \langle x \rangle, \tag{9}$$

$$b_z = \langle z \rangle - a_z \langle x \rangle. \tag{10}$$

Note that the uncorrelated errors ($\langle \epsilon_i \epsilon_j \rangle = 0$) assumption does not always hold when the measurement systems provide observations at different scales, for instance when one of the measurement systems has a coarser resolution than the other two systems (Vogelzang and Stoffelen 2012). In such a case, the high-frequency variability measured by the two fine-scale

measurements systems will not be detectable for the coarse-resolution system and therefore be regarded as error. If one wants to compare error variance at the scale of the coarse-resolution system, a priori knowledge on the so-called representation (or "representativeness") error is required in order to subtract it from (add to) the coarse (fine)-resolution system error estimates. For instance, Vogelzang and Stoffelen (2012) used the wind variance density spectra from numerical models and scatterometer observations to estimate the representation errors between satellite observations and model results. Another widely used method is to filter high-resolution observations in order to match the spatial scales of the coarser-resolution measurement system (e.g., Janssen et al. 2007; Abdalla et al. 2011; Caires and Sterl 2003). In our study, the selected model implementation uses a 50-km grid spacing, and the buoy and along-track altimeter data have been smoothed with 2-h and 50-km moving windows, respectively, in order to filter out high-resolution variability and match the model resolution. We therefore consider that the representativeness error is negligible. However, in order to estimate the impact of the in situ and altimeter data filtering, the triple collocation analysis has also been applied to the single point measurements and compared to the 50-km averaged "superobservations."

Since error estimates are highly sensitive to outliers, an outlier detection method has been implemented based on the iteratively reweighted least squares robust regression method, which assign a weight between 0 and 1 to each data pair during the regression iterative process. In our case, we apply the robust regression method successively to the altimeter–in situ, altimeter–model, and in situ–model data pairs, and reject the data triplet if the weight is below 0.1 in at least one of the three dual comparisons.

In the remainder of the study, two error estimates are given: the standard deviation of the random error $\sqrt{\langle \epsilon_i^2 \rangle}$ (named absolute error) and the normalized standard deviation of the random error $\sqrt{\langle \epsilon_i^2 \rangle}/\langle \mathrm{Hs}_i \rangle$ (named the relative error), where $\mathrm{Hs}_i$ refers to the mean Hs of the considered measuring system.

## 4. Results

Considering the entire triple collocation dataset and assuming that in situ platform and altimeter missions share identical error characteristics, we obtain the error statistics given in Table 1. Although these results are presented for different types of data processing (single or superobservations, with and without EMD denoising applied on altimeter data), we will first focus on the results obtained for the buoy and altimeter superobservations and no EMD denoising applied on altimeter data (second and sixth columns). We see that the Hs error is the lowest for in situ data (6.5%), followed by altimeter data (7.3%) and model data (10.2%). As far as altimeter data are concerned, these results are in line with the ones of Janssen et al. (2007) and Abdalla et al. (2011), who obtained normalized errors comprised between 5% and 10% for *ERS-2*, *Envisat*, *Jason-1*, and *Jason-2*. However, they significantly differ for what concerns the in situ measurement errors. Indeed, Abdalla et al. (2011) obtained larger errors for in situ measurements (8.6%) than for altimeter data and Janssen et al. (2007) found the largest error in in situ measurements compared to altimeter and model data, with values comprised between 8% and 13%. The potential reasons of these discrepancies will be discussed in section 5.

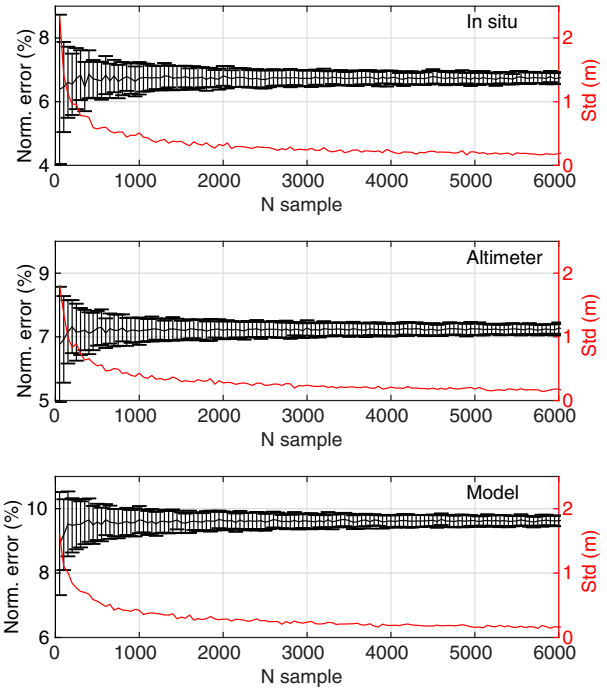


FIG. 2. Normalized error as a function of sample size. Error bars and red line represents the standard deviation estimated from 100 random realizations of each sample size.

To investigate error characteristic of the three measurements systems, the triple collocation analysis is then performed over different samples, corresponding to distinct wave platforms, satellite missions or Hs range. Since the results of triple collocation analysis depends on the sample size, we first perform a sensitivity analysis by randomly sampling the triple collocation dataset for sample sizes varying from 50 to 6000. For each sample size, 100 realizations are performed and the mean standard deviation of the errors are computed. Figure 2 shows the mean and standard deviation of the normalized error for each sample size and each measurement system. We can see that the standard deviation of the normalized error computed for 100 random sampling realizations is rapidly increasing for sample size lower than 500. For a sample size of 1000, the 95% prediction intervals of the mean normalized error are 6.5% ± 1.0% for in situ platform measurements, 7.3% ± 0.9% for altimeter measurements and 10.2% ± 0.8% for model outputs. These prediction intervals can be reduced by ~60% with 6000 samples. However, in order to analyze buoy errors individually and altimeter error at a yearly resolution, we have used a minimum sample size of 1000. The

TABLE 2. Number of years of observation, number of in situ–altimeter matchups, and average absolute and normalized errors and standard deviation of yearly errors per altimeter mission.

| Mission | No. of years | No. of matchups | Absolute error (m) | Normalized error (%) | Std dev(%) |
|---|---|---|---|---|---|
| *ERS-1* | 4 | 4628 | 0.20 | 9.29 | 0.42 |
| TOPEX | 13 | 25 311 | 0.16 | 6.90 | 0.32 |
| *ERS-2* | 17 | 30 975 | 0.20 | 8.90 | 0.65 |
| *GFO* | 9 | 21 206 | 0.16 | 6.97 | 0.30 |
| *Envisat* | 11 | 28 889 | 0.16 | 6.69 | 0.26 |
| *Jason-1* | 12 | 36 637 | 0.16 | 6.70 | 0.26 |
| *Jason-2* | 11 | 40 025 | 0.15 | 6.84 | 0.30 |
| *CryoSat-2* | 10 | 25 577 | 0.15 | 7.12 | 0.39 |
| *SARAL* | 6 | 24 138 | 0.12 | 5.72 | 0.23 |
| *Jason-3* | 3 | 12 528 | 0.16 | 7.18 | 0.43 |
| All missions | 27 | 249 914 | 0.16 | 7.23 | 0.36 |

number of matchups obtained for each altimeter mission is provided in Table 2.

### a. *In situ platform Hs errors*

In situ measurements are often considered as the gold standards for the calibration and validation of remote sensing instruments (e.g., radar altimeter) and numerical models. In that case, in situ measurements are assumed to represent the true quantity, and the errors or calibration coefficient of the other measurement systems are estimated with respect to these true measurements. However, no measuring device is perfect and ignoring measurement error characteristics during calibration or validation may severely impact the results (Stoffelen 1998). Moreover, in situ platforms involve a great number of measurement settings, such as offshore platforms, lightships, and moored buoys, and while buoys are the most common they can have a great diversity of hull shape and payload, which affect the accuracy of the measurement. In addition, long-term monitoring platforms are regularly replaced by more advanced (or cheaper) devices which may have different measurement error characteristic (Ardhuin et al. 2019).

Here we attempt to quantify and compare the measurement errors of in situ platforms regularly used for altimeter calibration and model validation. The top panel of Fig. 3 shows the normalized error of 72 platforms. Compared to the 122 platforms presented in Fig. 1, 50 platforms were removed from the analysis because (i) the number of matchups was lower than 1000 and/or (ii) the resolution of the measurements was too coarse (0.5 m). This latter criteria rejected all buoys located in the Bay of Bengal and Arabian Sea (see Table 1 in the supplementary material). For the remaining platforms, the main resolution was either 0.1 or 0.01 m, and the time periods for which the Hs records were given with a resolution $\geq 0.5$ were removed from the analysis.

The Hs random errors of in situ platforms vary between 0.06 and 0.62 m, which corresponds to 3%–10% of the mean Hs at each platform location (Fig. 3, top panel). The lowest normalized error (3.4%), which correspond to a 0.1 m absolute error for a mean Hs of 2.9 m, is obtained for platform 46246 (Ocean Weather Station Papa) located at 50°N, 145°W (in 4250-m depth), and recording ocean weather parameters since 2010. The highest normalized error (9.7%), which correspond to a 0.2-m absolute error for a mean Hs of 1.9 m, is obtained for platform 44024 located at 42.325°N, 65.909°W (in 225-m depth), and recording ocean weather parameters from 2004 until 2017. Note that the resolution of platform 46246 is 0.01 m while the resolution of platform 44024 is 0.1 m, which could partly explain these differences. To evaluate the impact of the instrument resolution on the error values presented in Fig. 3, the same analysis is performed after rounding all buoy measurements to the first decimal, hence mimicking a common 0.1-m resolution for all platforms. On average the error of the platform with an original resolution lower than 0.1 m shows a limited increase of 4.5%, and the overall ranking remains mostly unchanged. Moreover, no significant correlation was obtained between original platform resolutions and Hs errors.

Looking at the random errors characteristics of the platforms operated by the NDBC, MEDS, and Met Office, there is no systematic random error differences between these datasets. However, the bias between the in situ platforms and the altimeter data on one hand (Fig. 3, middle panel) and between the in situ platforms and the model on the other hand (Fig. 3, bottom panel) indicate that almost all MEDS platforms underestimate Hs measurements in comparison to altimeter records and model results. Considering all MEDS platforms, the average bias is $-0.08$ m with respect to altimeter data and $-0.14$ m with respect to the model results. Such underestimation was already evidenced by Durrant et al. (2009), who found bias difference around 0.25 m when comparing *Envisat* and *Jason-1* altimeter data with NDBC and MEDS buoys separately.

### b. *Altimeter errors*

Satellite altimeter records represent the largest database of Hs measurements, and allow for the reconstruction of continuous time series from 1992 onward with a quasi-global coverage. However, radar altimeter systems, including sensor technology, onboard and ground processing techniques, have drastically evolved over the past 30 years. It is therefore expected to obtain different error characteristics for each considered mission. Moreover, over its lifetime, a satellite mission encounter a number of issues that may impair its measuring capabilities [as an example, see Abdalla (2021) for a list of *Jason-2* issues], which also require particular attention. Here, we use the triple collocation technique in order to estimate

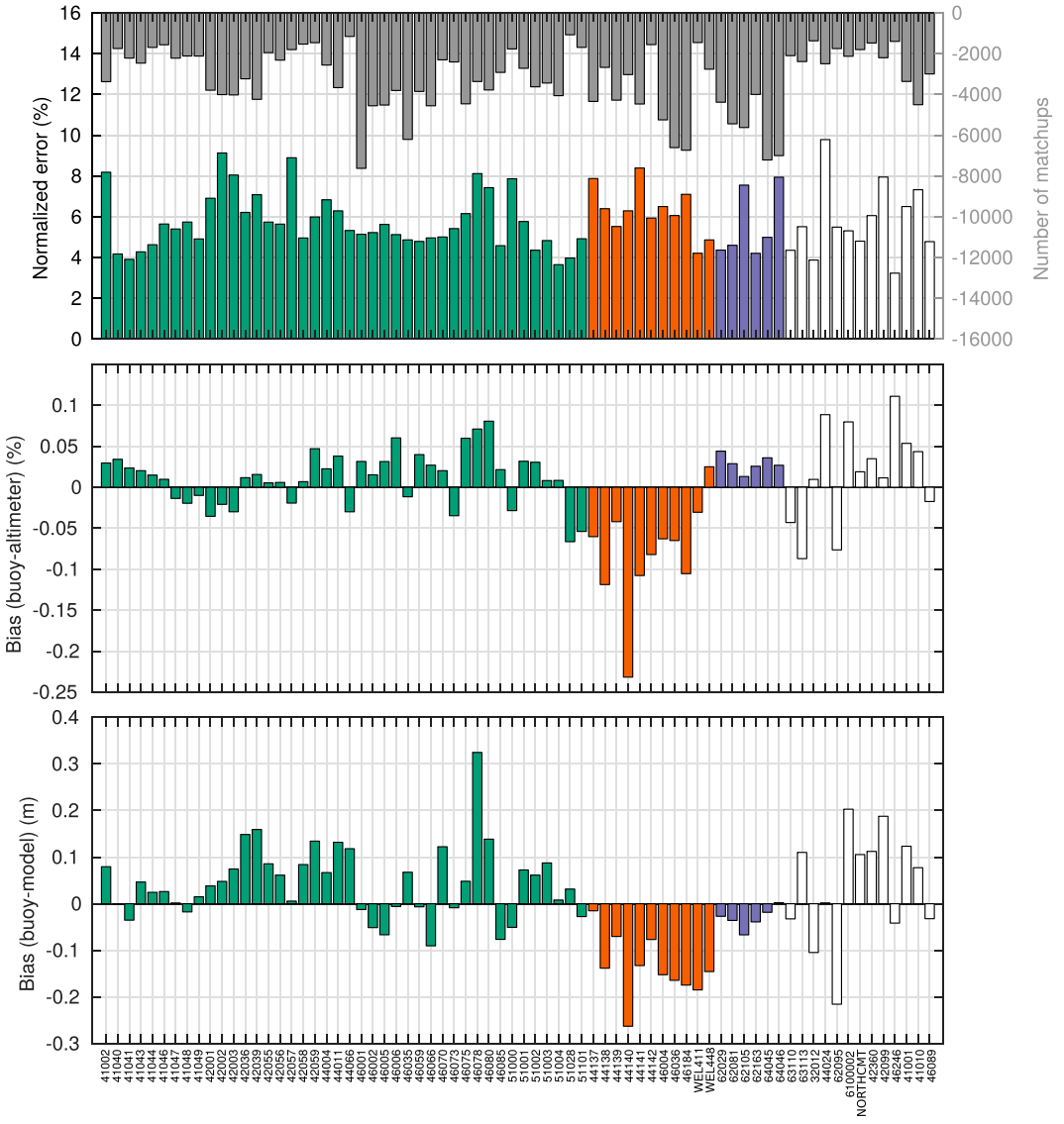


FIG. 3. (top) Normalized error and number of matchups for 72 in situ wave platforms. (middle) Bias between in situ and altimeter data. (bottom) Bias between in situ and model data. The colors of the bar indicate the platforms operated by NDBC (green), MEDS (orange), Met Office (purple), and other institutions (white).

the error in altimeter, in situ, and model data, for each mission on a yearly time scale. Figure 4 shows the time series of the normalized errors for the three measurement systems over the acquisition period of each altimeter mission of the ESA Sea State CCI dataset. The gray bars indicate the number of in situ–altimeter matchups found every year. We can see that this number varies between 716 (*ERS-1*, 1996) and 4845 (*Jason-2*, 2016), with an average of 2600. The temporal trends in the number of matchups can be attributed to the increasing number of in situ platform measurements. Other differences may be related to the different satellite orbits, and in particular their inclination, ranging from 66° for the TOPEX/Jason orbits to above 90° for the ERS/*Envisat*/*SARAL* and *CryoSat-2* orbits, which directly affect the duration a satellite is passing over dense in situ network area, mostly located within 10°–60°N.

A number of error characteristics can be observed. First, we note some interannual variability in the altimeter measurement errors, with variations up to 2.4% for *ERS-2* between 1995 and 1999. This variability may be attributed to several factors: the varying number of available matchups from year to year (see also Fig. 2), temporary issues with the radar altimeter system affecting the quality of the measurements, or the interannual variability of the wave climate. Second, if we compare the mean altimeter measurement errors between each mission (Table 2), we see that *ERS-1* and *ERS-2* show the largest normalized errors (9.29% and 8.90%, respectively), while *SARAL* show the lowest normalized error (5.75%) and also the lowest interannual variability, with a standard deviation of the yearly error equal to 0.23%. The second lowest mean normalized error (6.69%) and standard deviation

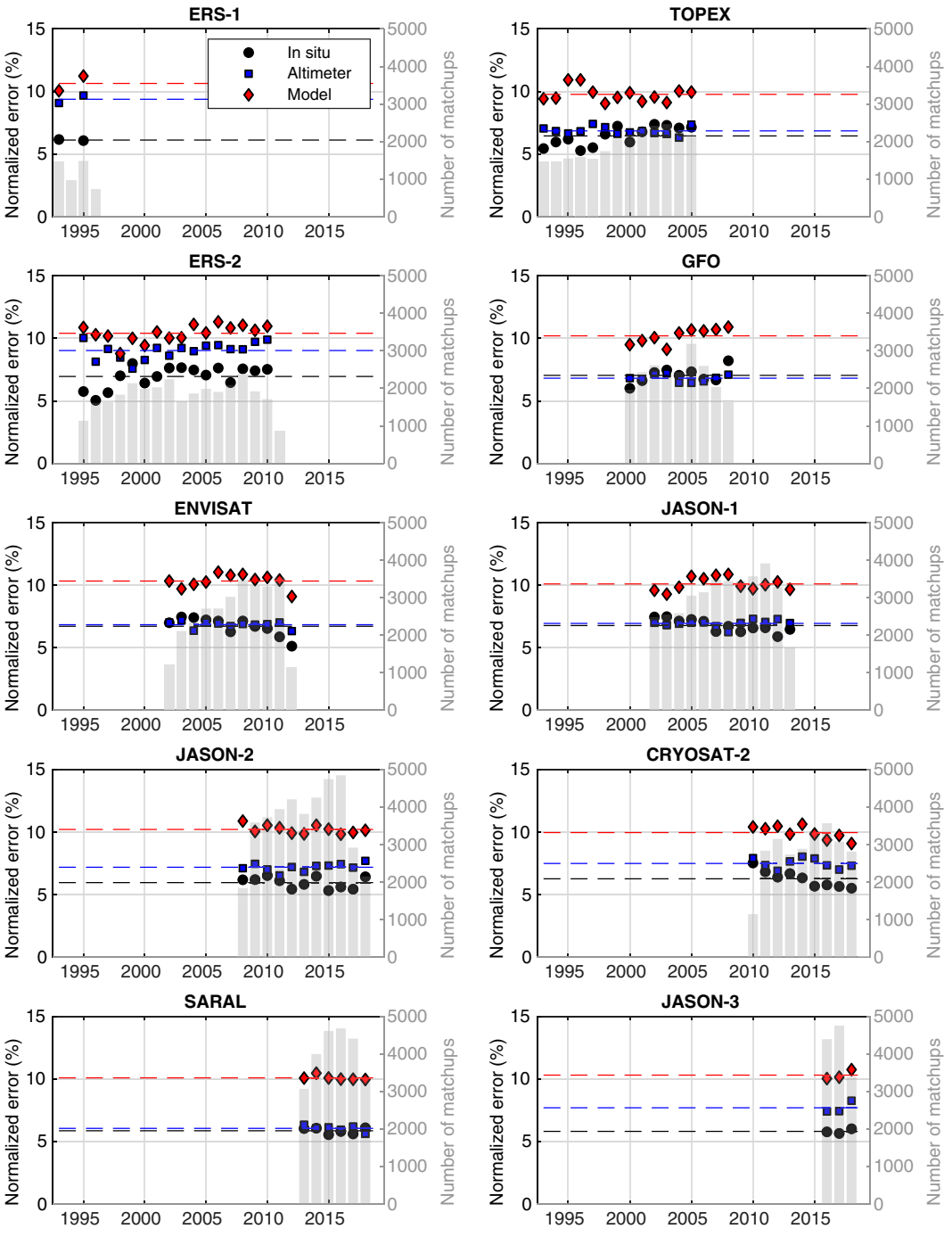


FIG. 4. Normalized error for in situ (black circles), altimeter (blue squares), and model (red diamonds) data, computed at yearly interval for each altimeter mission of the Sea CCI dataset. Gray bars indicate the number of matchups used to compute the error from the triple collocation technique. Dashed color lines represent the mean error values of each measurement system.

(0.26%) is obtained for the *Envisat* mission, launched in 2002 to service continuity of the ERS and operated until 2012. Comparing the error between the different measurement systems, we see that Hs measurements from *Envisat*, *Jason-1*, and *SARAL* present similar error levels than in situ observations, illustrating the excellent behavior of the radar systems on board these missions. Finally, looking at the averaged in situ and model errors for each altimeter mission (black and red dashed lines in Fig. 4), we note a very good consistency across all missions, which confirms the robustness of the triple collocation approach.

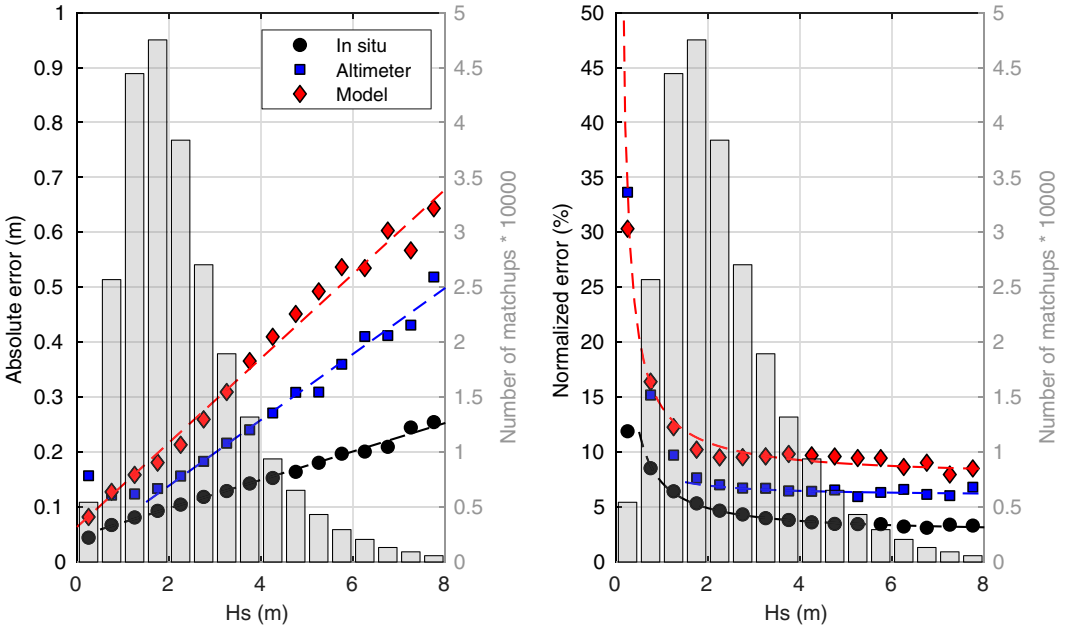


FIG. 5. (left) Absolute and (right) normalized errors for in situ (black circles), altimeter (blue squares), and model (red diamonds) data as a function of Hs. Gray bars indicate the number of matchups used to compute the error from the triple collocation technique.

### c. *Hs dependence of measurement errors*

Errors in Hs measurement systems are known to depend on the sea state conditions, and error models for wave measurement systems are often assumed to be linear over the whole range of measurements. However, all measurement systems are affected by technological constraints, in particular for very low and high sea states, and linear error models may not be valid in such conditions. Here, we use the results of the triple collocation analysis to investigate the relationships between Hs random errors and sea state conditions for in situ, altimeter, and model data. For this purpose, we look at the errors of the three measurement system as a function of Hs by computing triple collocation results for bins of 0.5-m width, from 0 to 8 m. To gather a sufficient number of matchups (>1000) for each bin, we assume that all altimeter and all in situ platform share similar error characteristics and we merge them all together. This way, we obtained robust statistics up to ~8 m. Figure 5 illustrates the relationship between measurements errors and Hs value for the three considered measurement systems.

The most striking feature is the quasi-linear relationship between the absolute error and Hs (Fig. 5, left panel). This relationship holds for Hs between 0.5 and 8 m for in situ measurements, 1.5–8 m for altimeter measurements, and 0–8 m for model data. Below 1.5 m, we note that altimeter error strongly deviates from the linear function. Moreover, the relative performance of each measurement system is given by the slopes of the linear error function, which is maximum for model data (0.07) and twice as high for altimeter (0.06) as for in situ measurements (0.03). Because the intercept of the linear function is different from 0 for each measurement system, the error function of the normalized errors is modeled by the slope of the linear function plus an inverse function. At low Hs (<1 m), altimeter errors are the highest (0.15 m, 28.2%) and in situ errors are the lowest (0.03 m, 9.2%). For Hs comprised between 0.5 and 8 m, measurements errors are systematically lower for in situ measurements, followed by altimeters and model data.

### d. *Seasonal variability of Hs errors*

Using 4 years of data, Janssen et al. (2007) put in evidence a strong seasonal variability of buoy and altimeter Hs errors. Here, we take advantage of our multidecadal triple collocation dataset to further investigate this topic. Figure 6 shows the climatology of the absolute and normalized errors for in situ, altimeter, and model data, estimated from the 27-yr TC dataset. The absolute error of each of the three systems shows a clear decrease during the summer period, associated with a decrease of the mean Hs. The Hs seasonality obviously results from the distribution of the in situ platforms, mostly localized in the Northern Hemisphere, where the sea state seasonality is well pronounced. In terms of normalized errors, the opposite behavior is observed, with significantly larger errors during the summer period, peaking in July/August, and corroborating Janssen et al.'s (2007) results. It is also clear from this figure that the error seasonality is significantly stronger (3% error difference between summer and winter) for altimeter data, than for model (1.4% difference) and buoy data (0.8% difference). From the normalized error/Hs relationship presented in Fig. 5, it is straightforward to relate these differences to the relative errors at low sea state obtained for each measurement system.

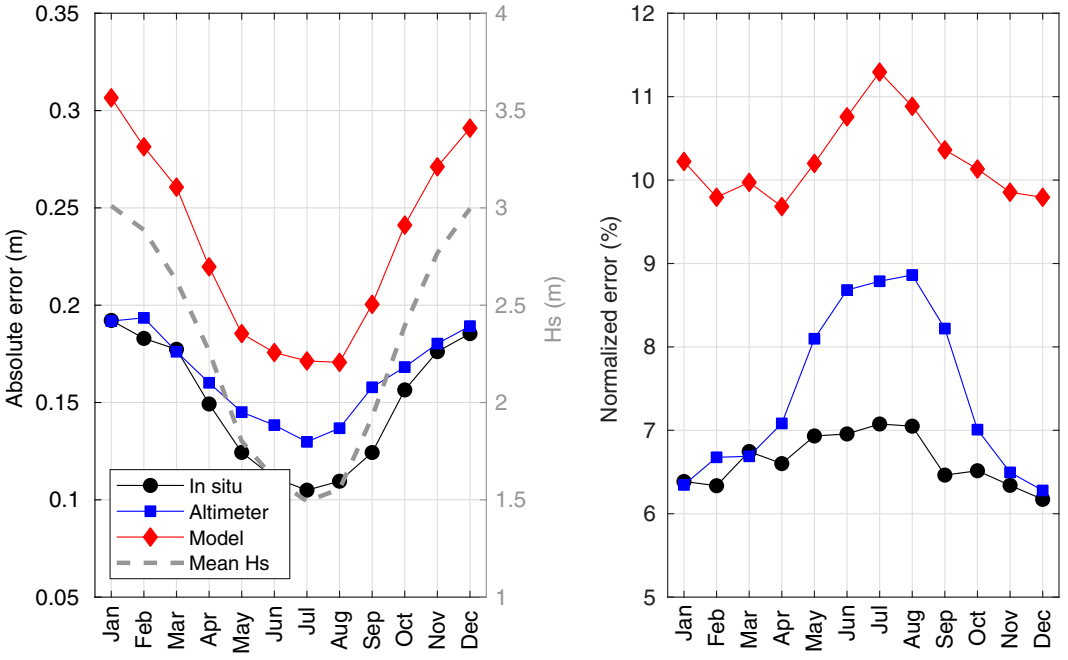


FIG. 6. Monthly climatology of the (left) absolute and (right) normalized error for in situ (black circles), altimeter (blue squares), and model (red diamonds). Mean Hs climatology is shown in the left panel with a dashed gray line.

### *e. Impact of data filtering on Hs errors*

The three measurement systems that we consider in this study provide Hs measurements at different scales. In particular, the in situ and altimeter observations contain some high-frequency signal that cannot be detected by the coarse-scale model. If part of the buoy and altimeter high-frequency signals were correlated, the triple collocation analysis would translate it into an increased model error making the error comparisons more subtle to interpret. To overcome this issue and estimate errors in the three measurement systems at the same spatial scale (the one of the model), altimeter and in situ observations are averaged to match the 50-km model resolution. We should therefore expect a reduced model error if the high-frequency signal, common to the fine-scale systems but invisible to the model, is filtered out. To test this assumption, we carried out the triple collocation analysis on single point in situ and altimeter measurements (first and fifth columns in Table 1), and compared the results to the ones obtained with the averaged superobservations (second and sixth columns in Table 1). We see that both in situ and altimeter errors are significantly reduced, while model error remains unchanged. As a consequence, we can assume that the high-resolution signal contained in the altimeter and in situ measurements are (mostly) uncorrelated, and is considered as noise within the triple collocation analysis. Therefore, a meaningful comparison of altimeter, in situ and model errors can only be done after filtering the high-frequency signals, and hold with respect to the resolution of the model. For the Sea State CCI dataset, a particular effort has been devoted to filter the high-frequency noise that interfere with the altimeter signal at scales lower than 100 km. The novel approach based on empirical mode decomposition (EMD) and wavelet thresholding, and described by Quilfen and Chapron (2021) is applied to 1-Hz along track Hs records in order to conserve nonlinear features, such as strong gradients, and extreme values, without significant smoothing. To estimate the impact of the EMD-denoising method on the altimeter Hs errors, the triple collocation analysis was carried out considering EMD-denoised single and superobservations (columns 3, 4, 7, and 8 in Table 1). We see a clear reduction of the altimeter error with EMD denoising when the single observations are considered. In this case the altimeter error is even lower than the in situ measurement error. When superobservations are used, we see that altimeter and in situ error levels are comparable, which confirms the improved accuracy of the denoised Hs altimeter data provided in the Sea State CCI product.

### *f. Altimeter calibration*

While this work focuses on the estimation of the random errors in each of the measurement systems, the results of the altimeter calibration performed during the triple collocation analysis are worth mentioning. Indeed the triple collocation is a powerful technique to calibrate a measurement system against another by taking into account random measurement errors in both systems. Moreover the calibration of altimeter wave measurements is critical for merging historical missions with recent ones and producing consistent long-term multimission records, necessary for climate trends studies (Young and Ribal 2019) or extreme wave analysis (Izaguirre et al. 2011). In Fig. 7, the altimeter calibration coefficients $a_y$ and $b_y$ of Eqs. (7)–(9) are plotted for each mission at yearly time scale (blue symbols) and averaged over the whole period of

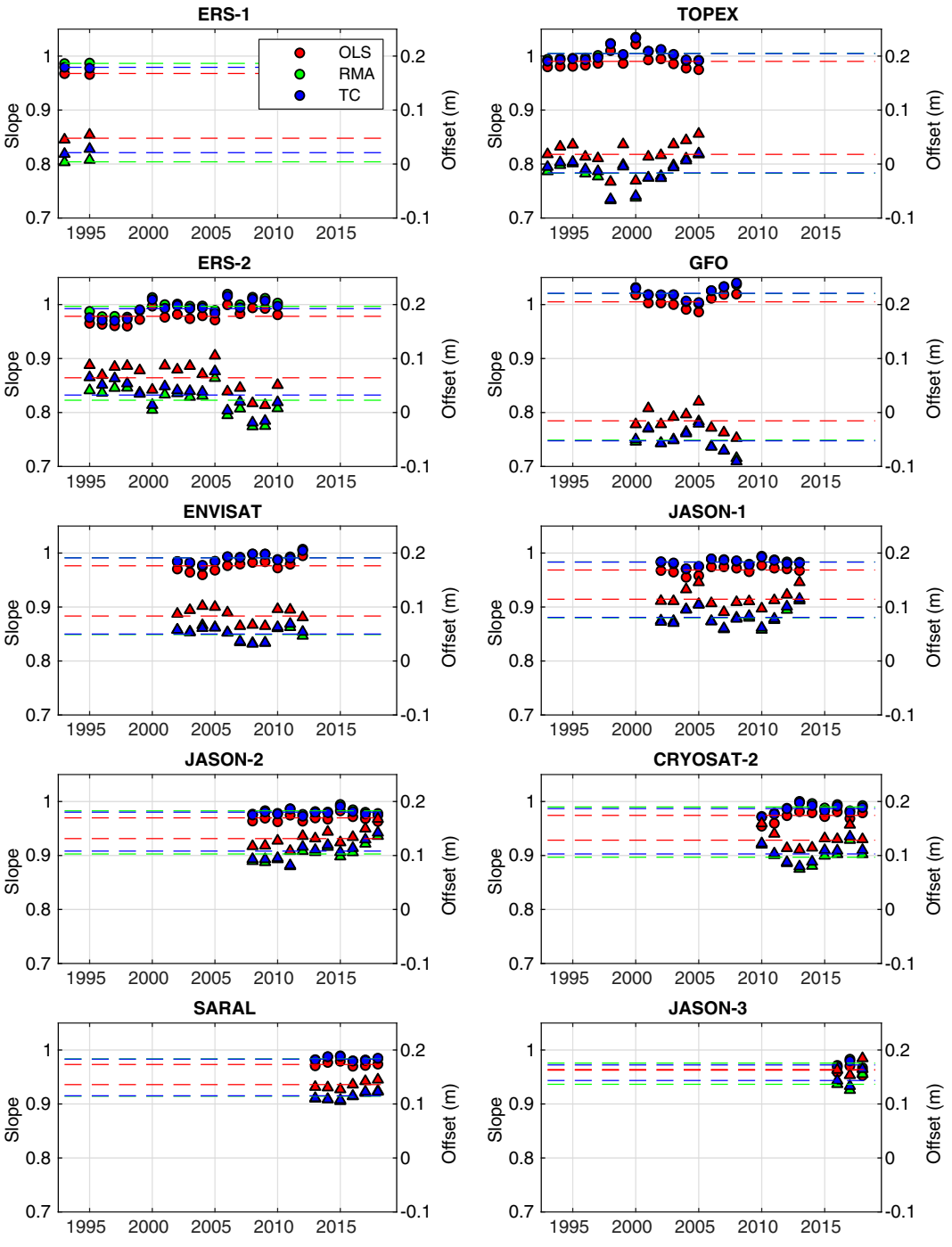


FIG. 7. Slopes (circles) and offsets (triangles) of the linear corrections estimated from ordinary least squares regression (OLS; red), reduced major axis regression (RMA; green), or triple collocation technique (TC; blue), computed at yearly interval for each altimeter mission of the Sea CCI dataset. Dashed colored lines represent the mean value computed over the time period covered by each altimeter mission. For OLS and RMA methods, dual comparisons between altimeter and in situ platform records were used.

measurements (dashed lines). For comparisons, the slopes and offsets of the linear corrections estimated from ordinary least squares (OLS) and reduced major axis (RMA) regressions are also shown (gray and red symbols and lines).

First of all, we can observe for all missions a significant interannual variability in the slope and offset values (see Table 3). This variability is maximum for *ERS-2* and minimum for *SARAL*, and may result from nonnominal functioning

TABLE 3. Mean, minimum, maximum, and standard deviation of the slopes and offsets of the linear corrections for each mission based on annual triple collocation analysis.

| | Slope | | | | Offset (m) | | | |
|---|---|---|---|---|---|---|---|---|
| Mission | Mean | Min | Max | Std dev | Mean | Min | Max | Std dev |
| *ERS-1* | 0.979 | 0.978 | 0.979 | 0.001 | 0.023 | 0.019 | 0.028 | 0.007 |
| TOPEX | 1.003 | 0.990 | 1.034 | 0.013 | −0.014 | −0.065 | 0.019 | 0.025 |
| *ERS-2* | 0.992 | 0.970 | 1.015 | 0.014 | 0.033 | −0.019 | 0.076 | 0.028 |
| GFO | 1.022 | 1.004 | 1.040 | 0.012 | −0.052 | −0.091 | −0.021 | 0.021 |
| *Envisat* | 0.990 | 0.977 | 1.005 | 0.008 | 0.052 | 0.032 | 0.069 | 0.013 |
| *Jason-1* | 0.983 | 0.971 | 0.993 | 0.006 | 0.083 | 0.060 | 0.115 | 0.017 |
| *Jason-2* | 0.979 | 0.973 | 0.991 | 0.005 | 0.109 | 0.082 | 0.142 | 0.018 |
| *CryoSat-2* | 0.985 | 0.971 | 0.998 | 0.008 | 0.105 | 0.080 | 0.135 | 0.018 |
| *SARAL* | 0.984 | 0.980 | 0.988 | 0.004 | 0.114 | 0.107 | 0.123 | 0.007 |
| *Jason-3* | 0.970 | 0.963 | 0.980 | 0.009 | 0.147 | 0.133 | 0.165 | 0.016 |

of the radar instrument, changes in the in situ platform networks or interannual wave climate variability. Moreover, significant differences in the slopes and offsets are observed from one mission to another. Considering the TC analysis, mean slopes vary from 0.97 (*Jason-3*) to 1.02 (*GFO*) and mean offsets vary from -0.05 m (*GFO*) to 0.15 m (*Jason-3*). In particular, we note that for *Envisat*, *Jason-1*, *Jason-2*, *CryoSat-2*, *SARAL*, and *Jason-3*, the offset is significantly larger (0.05–0.15 m) than for the prior missions (−0.05 to 0.03 m). This result is in close agreement with the validation results presented by Dodet et al. (2020, their Fig. 10) who obtained positive biases of similar magnitudes for these missions. The slopes and offsets presented in Dodet et al. (2020, their Table 3) are between 0.99 and 1.13 and −0.05 and 0.18 m, respectively. These values were used to calibrate the Geophysical Data Record products of each altimeter mission included in the Sea State CCI dataset. However, comparisons of the calibrated dataset against buoy data showed that some of the missions (*Jason-2*, *Jason-3*, *CryoSat-2*, and *SARAL*) were high biased, indicating potential issues in the calibration methodology. In the present manuscript, the slopes and offsets estimated from the TC analysis of the Sea State CCI calibrated altimeter missions are between 0.97 and 1.02 and −0.05 and 0.15 m. These values are on the same order as for the original calibration formula, which confirms the need for an additional calibration step to reduce the bias in some of the altimeter missions. After applying the TC calibration formula to the Sea State CCI dataset, the ranges of slopes and offsets are reduced by an order of magnitude (0.998–1.001 and −0.001 to 0.004 m, respectively). It is therefore particularly relevant to apply these new corrections to the Sea State CCI dataset v1 before investigating long-term statistics. Finally, we see that the TC and RMA are in very close agreement, indicating that in the absence of a triple collocation dataset, the reduced major axis regression is a robust alternative to TC analysis for accounting errors in the reference measurements during the calibration process.

## 5. Summary and discussion

In this study, we have used collocated data from three independent wave measurement systems, namely, the altimeter data from the ESA Sea State CCI v1 dataset, a selection of 122 mooring platforms within the CMEMS INSTAC dataset, and the WW3-LOPS model hindcast, in order to derive error characteristics for each of these systems. A first analysis based on the entire triple collocation dataset (250 000 entries) revealed that in situ measurements presented the lowest normalized errors (6.5%), followed by altimeter (7.3%) and model (10.2%) data, when altimeter and in situ observations were filtered to match the 50-km resolution of the model. These results contrast with the ones from Janssen et al. (2007), who obtained systematically larger errors (~8%) for in situ measurements than for altimeter data (~6%). As explained by these authors, a potential reason for these large buoy errors is that buoy data of questionable quality might have infiltrated their dataset. Regarding the relatively low errors they obtained for *ERS-2* (6.4%) in comparison to our results (8.9%), we have not found any clear explanation. It is particularly puzzling that these authors obtained a similar level of errors between *ERS-2* and *Envisat*, while we obtained systematic higher errors for *ERS-1* and *ERS-2* than for *Envisat* and the other missions. This intermission variability in measurement errors will therefore require further research.

For the first time, the Hs random errors of in situ measurements estimated from triple collocation analysis are intercompared between a set of platforms operated by different state agencies. These results show that the normalized errors are relatively low (3%–10%) for all the platforms considered herein, provided that the number of matchups is sufficient to derive robust statistics. In our case, 1000 matchups were found to give sufficiently accurate results (see Fig. 2) while preserving the possibility to carry out detailed analysis (such as interplatform comparisons). Moreover, the consistent error levels for model and in situ data across all altimeter missions confirm the robustness of our approach. The random errors present significant interplatform variability but they do not present systematic trends between the three major data providers: NDBC, MEDS, and Met Office. However, the biases between in situ data and altimeter data, and between in situ data and model data, confirmed a systematic error (negative bias of ~0.10 m) of the MEDS in situ measurements in comparison to the other platform networks (see Fig. 3), as

previously found by Durrant et al. (2009). To further examine the impact of the different hulls and payloads of the in situ platforms on the random errors, it would be very valuable to have access to more detailed information on the instrument and platform characteristics in the CMEMS INSTAC database.

Another key results of this triple collocation analysis is the intercomparison of the Hs random errors at yearly time scale for the 10 altimeter missions included in the ESA Sea State CCI dataset v1 (Fig. 4 and Table 2). We note a significant intermission variability of the mean and standard deviation of the annual errors, with *SARAL* showing the lowest error, as evidenced by previous authors (e.g., Sepulveda et al. 2015), and *ERS-1* and *ERS-2* showing the largest errors. The good performance of the *SARAL* mission can be attributed to the larger bandwidth and higher pulse repetition frequency (PRF) of the Ka-band AltiKa instrument on board *SARAL*, which result in a smaller effective footprint and a higher resolution than conventional Ku-band instrument (Verron et al. 2015). Regarding the lower performance of *ERS-1* and *ERS-2*, it could be attributed to the reduced number of range gates (64) used to sample the radar waveforms in comparison to the more recent missions (128). Indeed, although the extension to 128 bins was principally used for the estimation of antenna mispointing from the waveform data and is not directly used in the calculation of Hs (Quartly et al. 2001), these extended waveforms allows for a better characterization of rain cells and improved rain flagging (Tournadre 1998, 2004). Noteworthy are the *Envisat* and *Jason-1* missions showing the second and third lowest normalized errors, respectively, with very stable error characteristics over their more-than-10-yr-long measurement time period.

Since we used calibrated altimeter data for the triple collocation analysis and since the calibration involved in situ platform data that have also been used in this TC analysis, the assumption of independence between these measurement systems may be called into question. To investigate this point, we compared the random errors for the three measurements systems, using either the raw or the calibrated Hs altimeter data. We found a 17% increase of the altimeter error, a 5% increase of the in situ error, and a 1% decrease of the model error when using the raw Hs altimeter data in comparison to the calibrated altimeter data. These changes in in situ and model errors indicate that the altimeter calibration introduces some (artificial) correlation between altimeter and in situ errors. However, the impact of this correlation is relatively small.

Assuming that altimeter missions share similar error characteristics, it was also possible to investigate the Hs dependence of the random errors and compare this relationship between altimeter, in situ and model data (Fig. 5). It is shown that the absolute random error in each measurement system can be modeled by a linear relationship over the ranges 0.5–8, 1.5–8, and 0–8 m, for the in situ, altimeter, and model systems, respectively, with the largest slope (7%) obtained for the model data and the lowest slope (3%) obtained for the in situ data. For Hs below 1 m the altimeter errors are found to be the highest and in situ errors to be the lowest. This increased error of altimeter measurement at low sea state can be directly linked to the limited range resolution of the instruments (~0.5 m for Ku-band radar), which is insufficient to adequately sample the steep leading edge of radar waveforms backscattered from calm seas, resulting in a noisy estimate of Hs below 2 m (Smith and Scharroo 2015).

Finally, linear calibration coefficients were deduced from TC analysis for each altimeter mission at annual time scale, using the in situ data as reference. First, the errors show significant interannual variability, which is likely to increase the uncertainty in the estimate of calibration coefficients and may partly explain some differences in the altimeter corrections computed by different authors for different time periods (e.g., Queffeulou 2004; Zieger et al. 2009; Ribal and Young 2019). Regarding the altimeter measurements considered in this study, we note that the calibration slopes are close to 1 for all missions but the offsets show significant deviation from 0. Taking these new corrections into account is likely to improve intermission consistency. Also, since the most recent missions present larger calibration offsets than the early ones, it is possible that the long-term trends are impacted (Timmermans et al. 2020).

*Acknowledgments.* This research has been funded by the European Space Agency as part of the Sea State CCI project of the Climate Change Initiative (CCI) (ESA ESRIN, Contract 4000123651/18/I-NB). We are thankful to Stéphane Tarot and Ludovic Drouineau from Ifremer Systèmes d'informations scientifiques pour la Mer (SISMER) for providing detailed information on the CMEMS INSTAC wave in situ dataset.

*Data availability statement.* All data used in this study are publicly available. The L2P product from the ESA Sea State CCI v1 dataset can be downloaded from https://data.ceda.ac.uk/neodc/esacci/sea_state/data/v1.1_release/l2p/. The Copernicus Marine Service in situ dataset can be downloaded from http://www.marineinsitu.eu/. Ifremer wave hindcast can be downloaded at fttp://ftp.ifremer.fr/ifremer/dataref/ww3/GLOBMULTI_ERA5_GLOBCUR_01/GLOB-30M.

## REFERENCES

Abdalla, S., 2021: Are Jason-2 significant wave height measurements still useful. *Adv. Space Res.*, **68**, 802–807, https://doi.org/10.1016/j.asr.2019.08.032.

——, and G. De Chiara, 2017: Estimating random errors of scatterometer, altimeter, and model wind speed data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **10**, 2406–2414, https://doi.org/10.1109/JSTARS.2017.2659220.

——, and P. Janssen, 2017: Monitoring waves and surface winds by satellite altimetry: Applications. *Satellite Altimetry over Oceans and Land Surfaces*, CRC Press, 46 pp.

——, ——, and J.-R. Bidlot, 2011: Altimeter near real time wind and wave products: Random error estimation. *Mar. Geod.*, **34**, 393–406, https://doi.org/10.1080/01490419.2011.585113.

Alday, M., M. Accensi, F. Ardhuin, and G. Dodet, 2021: A global wave parameter database for geophysical applications. Part 3:

Improved forcing and spectral resolution. *Ocean Modell.*, **166**, 101848, https://doi.org/10.1016/j.ocemod.2021.101848.

Aouf, L., D. Hauser, B. Chapron, A. Toffoli, C. Tourain, and C. Peureux, 2021: New directional wave satellite observations: Towards improved wave forecasts and climate description in Southern Ocean. *Geophys. Res. Lett.*, **48**, e2020GL091187, https://doi.org/10.1029/2020GL091187.

Ardhuin, F., and Coauthors, 2010: Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation. *J. Phys. Oceanogr.*, **40**, 1917–1941, https://doi.org/10.1175/2010JPO4324.1.

——, J. Tournadre, P. Queffelou, and F. Girard-Ardhuin, 2011: Observation and parameterization of small icebergs: Drifting breakwaters in the Southern Ocean. *Ocean Modell.*, **39**, 405–410, https://doi.org/10.1016/j.ocemod.2011.03.004.

——, and Coauthors, 2019: Observing sea states. *Front. Mar. Sci.*, **6**, 124, https://doi.org/10.3389/fmars.2019.00124.

Caires, S., and A. Sterl, 2003: Validation of ocean wind and wave data using triple collocation. *J. Geophys. Res.*, **108**, 3098, https://doi.org/10.1029/2002JC001491.

Dodet, G., and Coauthors, 2020: The Sea State CCI dataset v1: Towards a sea state climate data record based on satellite observations. *Earth Syst. Sci. Data*, **12**, 1929–1951, https://doi.org/10.5194/essd-12-1929-2020.

Durrant, T. H., D. J. M. Greenslade, and I. Simmonds, 2009: Validation of *Jason-1* and *Envisat* remotely sensed wave heights. *J. Atmos. Oceanic Technol.*, **26**, 123–134, https://doi.org/10.1175/2008JTECHO598.1.

Girard-Ardhuin, F., and R. Ezraty, 2012: Enhanced Arctic sea ice drift estimation merging radiometer and scatterometer data. *IEEE Trans. Geosci. Remote Sens.*, **50**, 2639–2648, https://doi.org/10.1109/TGRS.2012.2184124.

Gruber, A., C. H. Su, S. Zwieback, W. Crow, W. Dorigo, and W. Wagner, 2016: Recent advances in (soil moisture) triple collocation analysis. *Int. J. Appl. Earth Obs. Geoinf.*, **45**, 200–211, https://doi.org/10.1016/j.jag.2015.09.002.

Hasselmann, S., and K. Hasselmann, 1985: Computation and parameterizations of the nonlinear energy transfer in a gravity-wave spectrum. Part I: A new method for efficient computations of the exact nonlinear transfer integral. *J. Phys. Oceanogr.*, **15**, 1369–1377, https://doi.org/10.1175/1520-0485(1985)015<1369:CAPOTN>2.0.CO;2.

——, ——, J. Allender, and T. Barnett, 1985: Computation and parameterizations of the nonlinear energy transfer in a gravity-wave spectrum. Part II: Parameterizations of the nonlinear energy transfer for application in wave models. *J. Phys. Oceanogr.*, **15**, 1378–1391, https://doi.org/10.1175/1520-0485(1985)015<1378:CAPOTN>2.0.CO;2.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Izaguirre, C., F. J. Méndez, M. Menéndez, and I. J. Losada, 2011: Global extreme wave height variability based on satellite data. *Geophys. Res. Lett.*, **38**, L10607, https://doi.org/10.1029/2011GL047302.

Janssen, P. A. E. M., S. Abdalla, H. Hersbach, and J.-R. Bidlot, 2007: Error estimation of buoy, satellite, and model wave height data. *J. Atmos. Oceanic Technol.*, **24**, 1665–1677, https://doi.org/10.1175/JTECH2069.1.

Law-Chune, S., L. Aouf, A. Dalphinet, B. Levier, Y. Drillet, and M. Drevillon, 2021: WAVERYS: A CMEMS global wave reanalysis during the altimetry period. *Ocean Dyn.*, **71**, 357–378, https://doi.org/10.1007/s10236-020-01433-w.

Leckler, F., 2013: Observation et modélisation du déferlement des vagues. Ph.D. thesis, Université Européenne de Bretagne, 240 pp., http://tinyurl.com/leckler-thesis.

Merchant, C. J., and Coauthors, 2017: Uncertainty information in climate data records from Earth observation. *Earth Syst. Sci. Data*, **9**, 511–527, https://doi.org/10.5194/essd-9-511-2017.

Mulet, S., and Coauthors, 2021: The new CNES-CLS18 global mean dynamic topography. *Ocean Sci.*, **17**, 789–808, https://doi.org/10.5194/os-17-789-2021.

O'Carroll, A. G., J. R. Eyre, and R. W. Saunders, 2008: Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *J. Atmos. Oceanic Technol.*, **25**, 1197–1207, https://doi.org/10.1175/2007JTECHO542.1.

Parker, W. S., 2016: Reanalyses and observations: What's the difference? *Bull. Amer. Meteor. Soc.*, **97**, 1565–1572, https://doi.org/10.1175/BAMS-D-14-00226.1.

Piollé, J.-F., G. Dodet, and Y. Quilfen, 2020: ESA Sea State Climate Change Initiative (Sea_State_cci): Global remote sensing multi-mission along-track significant wave height, L2P product, version 1.1. Centre for Environmental Data Analysis, accessed 1 March 2022, https://doi.org/10.5285/f91cd3ee7b6243d5b7d41b9beaf397e1.

Portabella, M., and A. Stoffelen, 2009: On scatterometer ocean stress. *J. Atmos. Oceanic Technol.*, **26**, 368–382, https://doi.org/10.1175/2008JTECHO578.1.

Quartly, G. D., M. A. Srokosz, and A. C. McMillan, 2001: Analyzing altimeter artifacts: Statistical properties of ocean waveforms. *J. Atmos. Oceanic Technol.*, **18**, 2074–2091, https://doi.org/10.1175/1520-0426(2001)018<2074:AAASPO>2.0.CO;2.

Queffeulou, P., 2004: Long-term validation of wave height measurements from altimeters. *Mar. Geod.*, **27**, 495–510, https://doi.org/10.1080/01490410490883478.

Quilfen, Y., and B. Chapron, 2021: On denoising satellite altimeter measurements for high-resolution geophysical signal analysis. *Adv. Space Res.*, **68**, 875–891, https://doi.org/10.1016/j.asr.2020.01.005.

Rascle, N., and F. Ardhuin, 2013: A global wave parameter database for geophysical applications. Part 2: Model validation with improved source term parameterization. *Ocean Modell.*, **70**, 174–188, https://doi.org/10.1016/j.ocemod.2012.12.001.

Ribal, A., and I. R. Young, 2019: 33 years of globally calibrated wave height and wind speed data based on altimeter observations. *Sci. Data*, **6**, 77, https://doi.org/10.1038/s41597-019-0083-9.

Rio, M.-H., S. Mulet, and N. Picot, 2014: Beyond GOCE for the ocean circulation estimate: Synergetic use of altimetry, gravimetry, and in situ data provides new insight into geostrophic and Ekman currents. *Geophys. Res. Lett.*, **41**, 8918–8925, https://doi.org/10.1002/2014GL061773.

Sepulveda, H., P. Queffeulou, and F. Ardhuin, 2015: Assessment of SARAL/AltiKa wave height measurements relative to buoy, Jason-2, and CryoSat-2 data. *Mar. Geod.*, **38** (Suppl. 1), 449–465, https://doi.org/10.1080/01490419.2014.1000470.

Smith, W. H. F., and R. Scharroo, 2015: Waveform aliasing in satellite radar altimetry. *IEEE Trans. Geosci. Remote Sens.*, **53**, 1671–1682, https://doi.org/10.1109/TGRS.2014.2331193.

Stoffelen, A., 1998: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *J. Geophys. Res.*, **103**, 7755–7766, https://doi.org/10.1029/97JC03180.

Timmermans, B. W., C. P. Gommenginger, G. Dodet, and J.-R. Bidlot, 2020: Global wave height trends and variability from new multimission satellite altimeter products, reanalyses, and

wave buoys. *Geophys. Res. Lett.*, **47**, e2019GL086 880, https://doi.org/10.1029/2019GL086880.

Tournadre, J., 1998: Determination of rain cell characteristics from the analysis of TOPEX altimeter echo waveforms. *J. Atmos. Oceanic Technol.*, **15**, 387–406, https://doi.org/10.1175/1520-0426(1998)015<0387:DORCCF>2.0.CO;2.

——, 2004: Validation of Jason and Envisat altimeter dual frequency rain flags. *Mar. Geod.*, **27**, 153–169, https://doi.org/10.1080/01490410490465616.

——, N. Bouhier, F. Girard-Ardhuin, and F. Rémy, 2016: Antarctic iceberg distributions 2002–2010. *J. Geophys. Res. Oceans*, **121**, 327–349, https://doi.org/10.1002/2015JC011178.

Verron, J., and Coauthors, 2015: The SARAL/AltiKa altimetry satellite mission. *Mar. Geod.*, **38**, 2–21, https://doi.org/10.1080/01490419.2014.1000471.

Vogelzang, J., and A. Stoffelen, 2012: Triple collocation. NWPSAF Tech. Rep. NWPSAF-KN-TR-021, 24 pp., https://knmi-scatterometer-website-prd.s3-eu-west-1.amazonaws.com/publications/triplecollocation_nwpsaf_tr_kn_021_1.0.pdf.

WAVEWATCH III Development Group, 2019: User manual and system documentation of WAVEWATCH III version 6.07. NOAA/NWS/NCEP/MMAB Tech. Note 333, 465 pp.

Young, I. R., and A. Ribal, 2019: Multiplatform evaluation of global trends in wind speed and wave height. *Science*, **364**, 548–552, https://doi.org/10.1126/science.aav9527.

Zieger, S., J. Vinoth, and I. R. Young, 2009: Joint calibration of multiplatform altimeter measurements of wind speed and wave height over the past 20 years. *J. Atmos. Oceanic Technol.*, **26**, 2549–2564, https://doi.org/10.1175/2009JTECHA1303.1.