Product Validation and Intercomparison Report (PVIR)

Version 3.0, 31 January 2022

Author	Approved	Signature	Date
Ben Timmermans Christine Gommenginger Guillaume Dodet Jean Bidlot	Christine Gommenginger, Ellis Ash		31 Jan 2022
ESA Approval			

Issue	Date	Comments
0.1	October 2019	First draft
1.0	June 2020	First issue
2.0	20/10/2021	Version 2 with updates for v2 data including SAR validation
3.0	31/01/2022	Version 3 with updates for v3 data including Sentinel-3A data

Contents

1. Introduction

This document represents the Product Validation and Intercomparison Report (PVIR) for the CCI+ Sea State project, corresponding to Deliverable 4.1 of WP4000.

Version 3.0 of the PVIR was prepared subsequent to the completion of the third year of the project, concurrent with the release of CCI Sea State V3 products. The aim of the PVIR is to present results of consistency checks and validation of available CCI products. The PVIR is a working document that is updated after each CCI product release, typically towards the end of each year of the CCI+ Sea State project.

The Product Validation and Intercomparison Report is structured to address the requirements outlined in the ESA Statement of Work. Validation results are presented into subsections as follows:

- Summary of improvements in validation (as part of internal product calibration) between V1 and V2 altimetry datasets, due to updated retracking and calibration approach.
- Consistency checks: assessment of CCI climatological sea state from altimetry (Hs) and imaging SAR (Hs & Tm2) against other sea state products.
- Validation of CCI multivariate (Hs & Tm2) and seasonal sea state information from imaging SAR observations with respect to in situ data and other sea state products, at selected long term sites.
- Assessment of impact of annual sampling increase in Hs estimation from the combined Version 3 L2P altimeter and imaging SAR observations [2002 - 2019], with respect to in situ data in the North Pacific.
- Future work: suggested priorities for onward investigation.

Material in this document has been published (in this form or slightly revised) in the following peer-reviewed publications:

- Timmermans, B., C. Gommenginger, G. Dodet and J. R. Bidlot (2020). "Global Wave Height Trends and Variability from New Multimission Satellite Altimeter Products, Reanalyses, and Wave Buoys." Geophysical Research Letters 47(9): e2019GL086880.

- Dodet, G., J. Piolle, Y. Quilfen, S. Abdalla, M. Accensi, F. Ardhuin, E. Ash, J. Bidlot, C. Gommenginger and G. Marechal (2020). "The Sea State CCI dataset v1: Towards a sea state Climate Data Record based on satellite observations." Earth Syst. Sci. Data Discuss 2020: 1-28.

2. Validation summary of CCI altimeter observations: Version 1 -> Version 2

2.1 Introduction

This section summarises validation results during product inter-calibration and covers the analysis of systematic and random errors in the Sea State CCI altimeter L2 products (1Hz along-track records) based on comparisons with in situ observations and model results. A primary objective is to provide quantitative information on the overall level of accuracy of the altimeter product developed during the Sea State CCI project. A second objective is to assess the impact of the different processing methods implemented within the project: in particular the new waveform retracking and the inter-calibration and denoising method implemented in version 2 dataset.

The CCI v2 altimeter products comprise the following missions: Jason-1, Envisat, Jason-2, Cryosat-2, SARAL-Altika, Jason-3. Together these span 2002 to 2019 (complete years), while Cryosat-2 also provides partial data for 2020. Validation results reported in this section are not relevant to data from imaging SAR. See Section 3.2 for more details.

2.2 Methodology

2.2.1 Selection of in situ platforms

For the validation of the CCI altimeter data, 72 quality controlled in situ platforms were selected from the reprocessed product (WAVE-REP) of the CMEMS INSTAC database (http://www.marineinsitu.eu/). This selection corresponds to offshore platforms (>100 km from the coast) with more than 1000 matchups with altimeter data (Figure V1). This selection contains stations mostly from the US National Data Buoy Center (60%), the Canadian Marine Environmental Data Service (15%) and the UK Met Office (8%). The remaining stations (17%) come from other US and European organisms. Note that this set of platforms was also used to calibrate the reference Jason-2 mission (see Product User Guide). Match-ups between altimeter and in situ measurements are identified when the closest approach of the altimeter ground track to the in situ platform location is less than 100 km, within a 30 minute time window. Altimeter and in situ data are filtered with 50 km and 2 hr moving average, respectively, in order to filter out high-frequency variability.
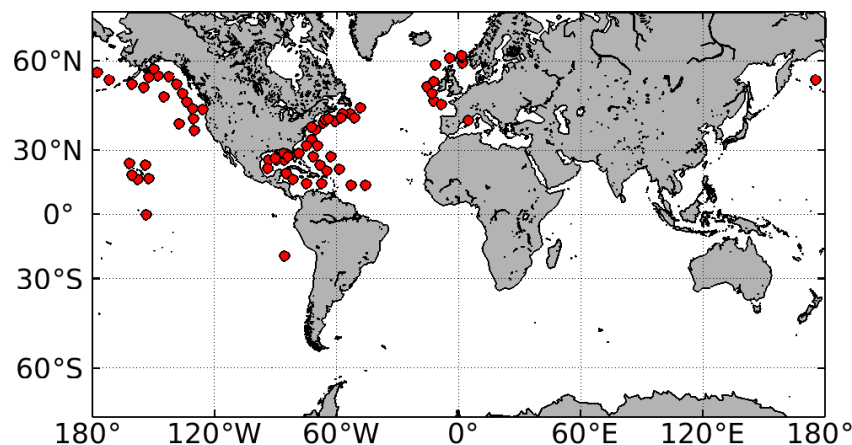
Figure V1: Selection of the in situ platforms from the CMEMS INSTAC database used for the validation of the CCI altimeter observations.

2.2.2 Comparisons with model results

Because in situ wave platforms are sparse, and mostly located along coastal margins in the North Hemisphere, they can only sample a limited amount of the sea state conditions observed by satellite altimeters. Model reanalysis are therefore extremely valuable datasets to investigate altimeter measurement errors in regions where in situ observations are missing. In order to complete the validation of the CCI altimeter data, comparisons against the ERA5 reanalysis have been performed. For this purpose, the model outputs are interpolated at every 1 Hz along-track position using bilinear interpolation. For each mission, we used 30 consecutive days of observations. Altimeter data were filtered with a 50 km moving average in order to filter out high-frequency variability. Since version 1 and version 2 of the Sea State CCI altimeter products share the same along-track posting (except for Cryosat-2, for which a different data source is used between v1 and v2), it is possible to perform the comparisons for both versions using the same sampling of data pairs. For this purpose, we computed error statistics against model results, using only valid data common to version 1 and version 2.

2.3 Validation of CCI altimeter observations

2.3.1 Comparisons of altimeter data (uncalibrated and calibrated) with in situ data

Error statistics between altimeter records and in situ measurements are presented for CCI variables *swh* and *swh_adjusted* in order to assess the impact of the calibration methodology. These statistics were computed after outliers removal, using only data pairs with a weight > 0 in an iteratively reweighted least-squares robust regression. Table V1 shows the results for each mission, before (*swh*) and after calibration (*swh_adjusted*).

Table V1: Statistical results for swh (No corr.) and swh_adjusted (Corr.) in version 2 dataset based on comparisons with in situ platform.

Mission	#values	Bias (m)		NRMSE (%)		SI (%)		R^2	
		No corr.	Corr.	No corr.	Corr.	No corr.	Corr.	No corr.	Corr.
Jason-1	11433	0.05	-0.01	11.26	10.28	9.55	8.87	0.96	0.97
Envisat	14131	0.07	-0.01	10.6	9.8	8.8	8.47	0.97	0.97
Jason-2	17146	0.07	0	11.12	9.7	9.24	8.34	0.97	0.97
CryoSat-2	795	-0.09	0.02	09.07	08.04	7.29	7.18	0.97	0.97
Saral	10909	0.24	0.02	14.29	9.22	8.35	7.94	0.97	0.97
Jason-3	7043	0.07	0.02	10.88	9.45	8.98	8.11	0.97	0.98
Mean	10443	0.1	0.01	9.78	8.32	8.8	8.3	0.97	0.97

Overall, we note a significant reduction of the differences between the altimeter and in situ measurements after the calibration is applied. In particular, the bias decreases from 0.10 m to 0.01 m on average, and the bias of each individual mission after calibration is lower than 0.03

m, which ensures a better long-term consistency of the multi-mission merged altimeter dataset. Note that the null bias obtained for Jason-2 is a consequence of the calibration methodology, in which the reference mission Jason-2 was corrected to match these in situ data. The Scatter Index decreases from 8.8% to 8.3% on average, with the lowest SI obtained for Cryosat-2 (7.18%) and the highest SI obtained for Jason-1 (8.87%).

If we compare the distribution of in situ data and CCI v2 uncalibrated altimeter data (Figure V2, left panels), we see that the major differences concern low sea state values (below 1 m). Indeed, altimeter data contain a much lower number of low sea state data and a higher number of average sea state data (between 1-2 m) than in situ data.
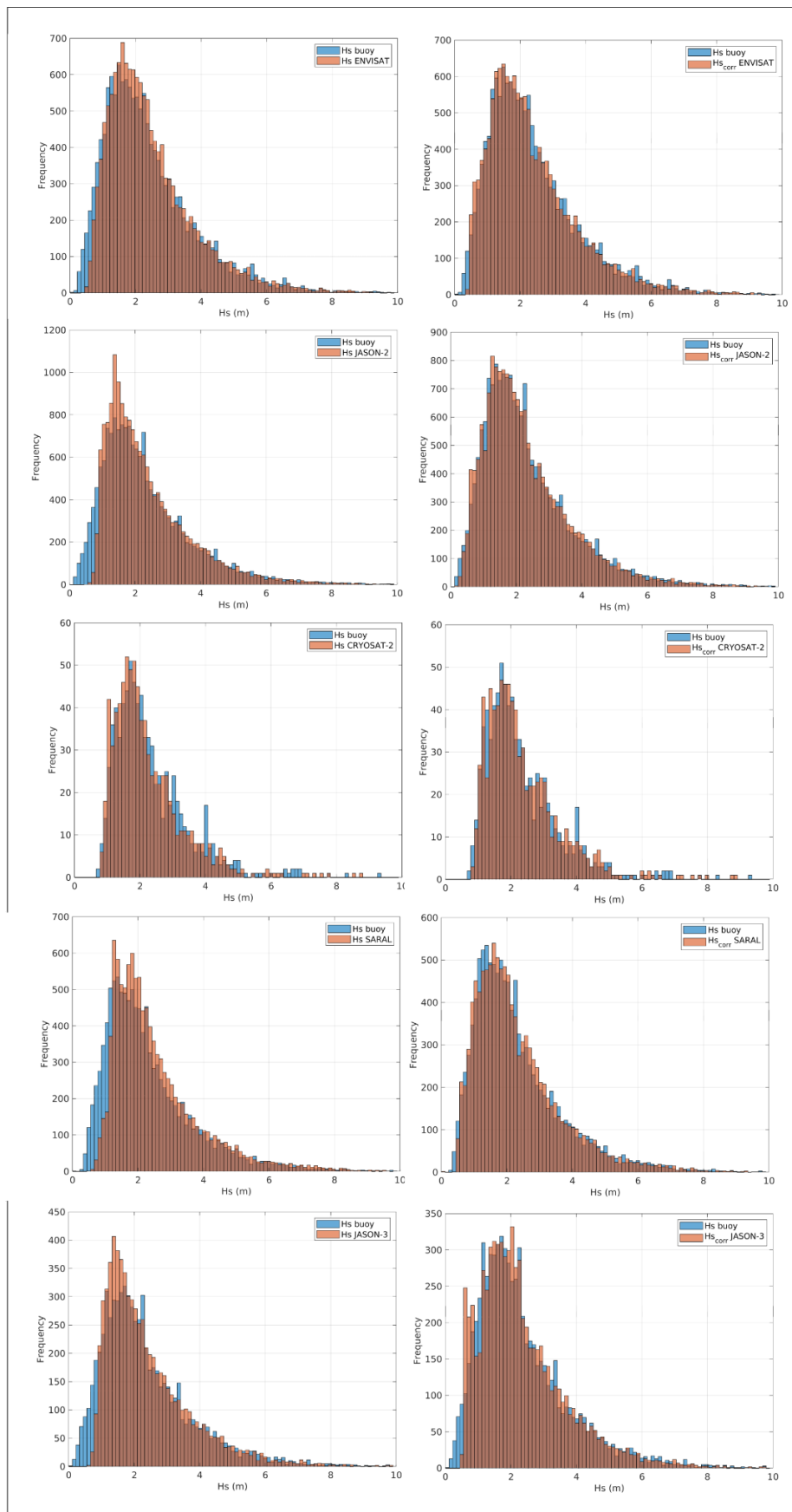
Figure V2: Distribution of significant wave heights from in situ measurements (blue) and CCI altimeter v2 data (red) before (left) and after (right) calibration is applied to the altimeter data.

For Cryosat-2 this difference is less clear because there are almost no in situ matchups below 1 m, as a consequence of the different orbit of this mission, which samples a greater fraction of the ocean's high latitudes compared to other missions. The overestimation of altimeter measurements at low sea state is likely attributed 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 (Smith and Scharroo, 2015). Another possible source of error could be the instrumental corrections derived from the standard MLE-3 and MLE-4 retrackers and applied to the 20 Hz swh records produced by the WHALES retracker (see PUG). In order to reduce this overestimation at low sea states, empirical calibration Look-Up Tables have been computed for each mission based on comparisons at altimeter-in situ matchups (Jason-2) and crossovers with Jason-2 (other missions). Comparisons between the distribution of in situ data and CCI v2 calibrated altimeter data (Figure V2, right panel) clearly show the better match between these observations after the calibration is applied to the altimeter data.

2.3.2 Comparisons of version 1 and version 2 products with ERA5

Error statistics between altimeter records and ERA5 data are presented for the variable *swh_denoised* from both version 1 and version 2 datasets in order to estimate the improvements between these two products (Table V2). Note that given the short period of time used for this analysis (30 consecutive days during the boreal winter period), these results are not necessarily representative of all sea state conditions.

Table V2: Statistical results for swh_denoised in version 1 and version 2 CCI datasets based on comparisons with ERA5

Mission	Bias (m)		RMSE (m)		SI (%)		R^2	
	V1	V2	V1	V2	V1	V2	V1	V2
Jason-1	0.07	-0.01	0.24	0.22	7.77	7.47	0.97	0.97
Envisat	0.09	0.03	0.23	0.2	7.36	6.92	0.98	0.98
Jason-2	0.09	0.02	0.23	0.2	7.02	6.78	0.98	0.98
Saral	0.07	0	0.21	0.2	6.89	6.95	0.98	0.98
Jason-3	0.08	0.01	0.24	0.21	7.51	7.13	0.97	0.97
MEAN	0.08	0.01	0.23	0.21	7.31	7.05	0.98	0.98

We note from these comparisons that the CCI v1 dataset is systematically high-biased (between 0.07-0.09 m) in comparison to the ERA5 reanalysis. While this could be attributed to an under-estimation of *swh* by the model, we believe it is rather the result of the inter-calibration methodology implemented for version 1 dataset. Indeed, for this preliminary release, calibration formulas derived prior to the CCI project were used to calibrate Jason-2 (and Envisat), and were shown a posteriori to be high-biased with respect to in situ data (Dodet et

al., 2020). Since Jason-2 was used as a reference mission to inter-calibrate the most recent missions in the version 1 dataset, a positive bias has also impacted the other missions. In the v2 dataset, the calibration formula of Jason-2 was revised using comparisons with in situ data, and the remaining missions were inter-calibrated against Jason-2. As a result, this new version is now characterized with very low biases (< 0.04m) when compared to in situ data (Table V1) and ERA5 reanalysis (Table V2).

The improvements of version 2 with respect to version 1 is also visible in the RMSE and SI statistics, which present a slight decrease from 0.23 m to 0.21 m, and from 7.31% to 7.05%, respectively, from version 1 to version 2 datasets. This error decrease can be likely attributed to the WHALES retracking algorithm that has been applied to the radar waveforms of the Sensor Geophysical Data Record in version 2, while version 1 used the standard products provided by space agencies. The fact that error statistics were computed using the same sample of valid data common to version 1 and version datasets exclude the hypothesis that improvements may be related to a different data editing procedure.

2.4 Conclusions

The analysis of error metrics between the CCI altimeter data and in situ data and model results confirm the ability of altimeter missions to provide accurate measurements of the significant wave heights at global scale. The version 2 of the CCI altimeter dataset presents biases lower than 0.05m when compared to buoy data and model results. The Scatter Index are lower than 9% when computed against in situ data, and lower than 8% when computed against ERA5 reanalysis. The coefficients of determination (R^2) are systematically higher than 0.96. All these metrics confirm the excellent performance of the altimeter missions included in the ESA Sea State CCI dataset, and the robustness of the processing algorithms developed during the project in order to improve these performance. In particular, the inter-calibration method allows for a better representation of significant wave heights lower than 1 m and a reduction of the bias with in situ and model data. Moreover, the WHALES waveform retracker implemented for the version 2 dataset results in a reduction of the random errors.

3. Consistency checks

3.1 Introduction

This section deals with the evaluation, at the global scale, of the mean seasonal sea state climatology (Hs) as provided in CCI L2 altimetry sea state products. Comparisons are made with the satellite product of Ribal & Young (2019) over the same period. Also included is a summary of characteristics of sea state variables (Hs and Tm2) provided by the newly released imaging SAR product from the Envisat and Sentinel-1 missions, as compared to CCI altimetry products. Note that observations from imaging SAR constitute the first formal release of this data (as opposed to a version 2).

The formal characterisation of uncertainties is a major objective of the CCI Sea State project. In this report, only preliminary assessments of uncertainties are presented. Noting that, for version 2, updates to the altimetry products are limited to six missions (see 3.2.1) and, for version 3, the addition of Sentinel-3A, therefore constrained to between 2002 and 2020 (complete years), a more detailed analysis of long term variability and its uncertainty is postponed until a longer record is available.

The first formal release of observations from scanning aperture radar (SAR) operating in "wave mode", referred to here as "imaging SAR", does however provide the opportunity to examine improved sampling rates and data availability, and extent of spatial coverage. Single mission estimates of monthly summaries tend to be subject to considerable uncertainty on spatial scales less than a few hundred kilometres so these are typically aggregated to grid cells of 4 degrees. Note also that, since imaging SAR data provides, at a minimum, (total) significant wave height and zero-crossing period (Tm2), the opportunity exists to examine the joint distribution of these variables on global scales. Both Hs and Tm2 are examined in the report but analysis of the joint distribution is not presented at this time. In addition to total Hs and Tm2, a number of other sea state parameters are provided for Sentinel-1, A & B (see section 3.2.1). Analysis of these variables is not included at this time. Also, two separate processing algorithms are available for Sentinel-1 acquired data, these being from DLR and QUACH, both selected following the results of the "Round Robin" evaluation exercise. In this report, the data from DLR are presented.

With regards to product inter-comparisons, there is little consistency in the metrics used in different products to estimate uncertainty. Due to these considerations, in this report, we do not explicitly evaluate statistical robustness of differences. This issue is highlighted in Section 6 as a topic for further investigation.

3.2 Data description

3.2.1 CCI L2 version 2

The CCI L2 V2.06 product provides along track 1 Hz observations from altimeters from the following missions: Jason-1, Envisat, Jason-2, Cryosat-2, SARAL-Altika, Jason-3. Together these span 2002 to 2019 (complete years), while Cryosat-2 also provides partial data for 2020. A number of variables are provided in the L2 product including the uncalibrated source GDR product total significant wave height data (*swh*), the calibrated data (*swh_adjusted*) and a

further level of quality control where outliers have been identified and removed from the calibrated data product (*swh_denoised*). In addition, information on quality control is provided in terms of both a data quality flag and information on why the quality was degraded (where applicable). In this report we make use only of *swh_adjusted* and consider "Good" data only (*swh_quality* = 3).

In addition to wave height from altimetry, a number of sea state parameters have been developed from imaging SAR. These data were derived from Envisat (12/2002 - 04/2012) and Sentinel-1 A (12/2014 - 02/2021) & B (04/2016 - 02/2021). For Envisat, these include total significant wave height (*swh*) and the second moment of wave period (*Tm2*), and for Sentinel-1 (DLR processing), these include total significant wave height (*swh*), windsea, significant wave height of the first two swell partitions and the zeroth, first and second moments of wave period (*Tm0, Tm1, Tm2*). In addition, information on both observation uncertainty and quality control is provided in a format similar to altimetry data (*swh_quality, swh_rejection_flags*).

In this report total swh (*swh*) and *Tm2* are considered.

3.2.2 CCI L2 version 3

Sentinel-3A data (07/2016 - 12/202) is provided with the version 3 release, together with additional data for Cryosat-2, SARAL-Altika and Jason-3 complete to 12/2020.

3.2.3 CCI L3 & L4 datasets

Level 3 combined 1 Hz data, and Level 4 gridded products are not evaluated in this report.

3.2.4 Ribal and Young (2019)

The dataset described by Ribal and Young (2019; RY2019 hereafter) is a level 2 altimetry product, providing 1 Hz measurements of Hs and wind speed from all relevant satellite altimeter missions. This dataset has been recently extended in time to 2020 (complete years) using the same calibration and quality control methodology as described in the 2019 publication. All 1 Hz observations are geo- and time-referenced and provided in separate files for 1 degree grid cells. Uncertainty information relating to the 1 Hz observations is not explicitly provided although the sampling distribution can be derived for given geospatial constraints.

3.2.5 ERA5

The European Centre for Medium Range Weather Forecasts (ECMWF) reanalysis dataset, ERA5, is an industry leading reanalysis product. Monthly mean Tm2 (second moment of wave period) at a resolution of 0.5 degrees has been used for comparative analysis. No separate estimate of uncertainty is provided, although be cognisant that these data are not sampled in the same way as for sequential satellite observations. ERA5 hourly Hs can be obtained, from which the Hs probability distribution can be better resolved. For comparison within this report, ERA5 data has been aggregated to 4 degree grid cells.

3.3 Assessment of Sea State ECV

3.3.1 Global seasonal (JFM, JAS) mean climatology (1 deg) and differences.

Climatological comparisons of aggregated 1 Hz data points are made between CCI L2 V2 and RY2019 (updated). Seasonal averages, as opposed to annual averages, are shown to reflect the strong seasonal influences on sea state. Data were aggregated by taking the median value for the segment of the track that passed through the relevant 1 degree square grid cell and then taking the mean of those medians for the full record duration. This is similar to the approach used to derive the CCI L4 product (which could be used for such a comparison, as performed in PVIR V1). For the CCI L2 product, *swh_adjusted* was employed a quality flag value of "3", corresponding to "Good" data, was applied.

Figure 1(A) shows climatological mean Hs for JFM over 2002-2019 at 1 degree resolution from the CCI L2 V2 altimetry products. High values corresponding to northern hemisphere winter storm tracks are evident. Figure 1(B) shows the difference between CCI V2 L2 and RY2019 mean states. Substantial similarity is evident across most of the global oceans although small discrepancies (up to ~25 cm) appear to be predominantly negative. This is most pronounced in parts of the North Atlantic and Southern Ocean.

(A) 2002-2019[JFM] CCI2021: Climatological Hs mean

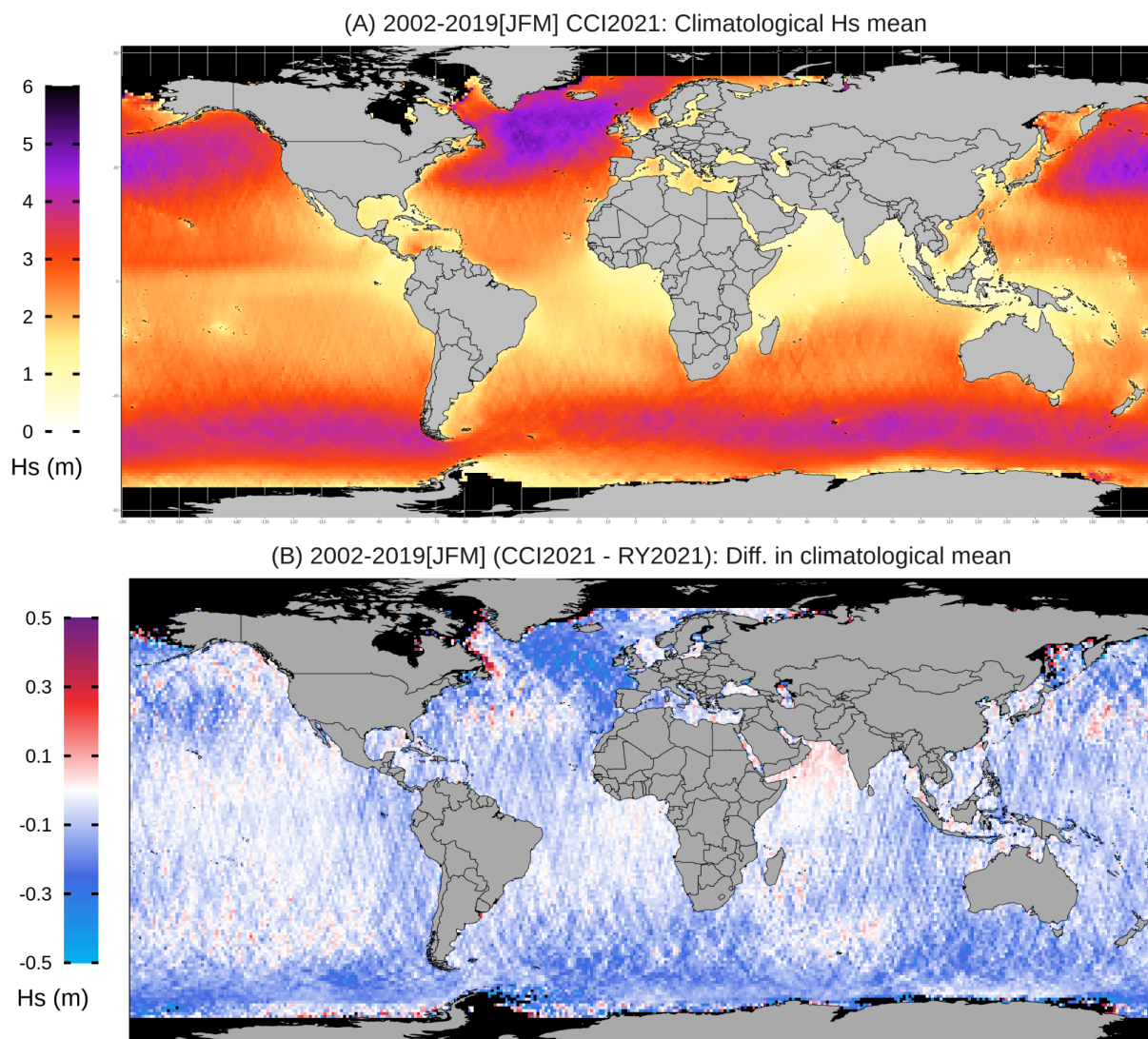(B) 2002-2019[JFM] (CCI2021 - RY2021): Diff. in climatological mean

Figure 1: A) Climatological seasonal mean Hs for the aggregated CCI L2 altimetry products over 2002-2019 at 1 degree resolution; B) Difference in mean with respect to RY2019.

Figure 2(A) shows climatological mean Hs for JAS over 2002-2019 at 1 degree resolution of the CCI L2 V2 altimetry products. High values corresponding to energetic conditions of the Southern Ocean are evident. Figure 1(B) shows the difference between CCI V2 L2 and R&Y2019 mean states. Close agreement between the two products is evident across most of the global oceans. Notable discrepancy is mostly negative and limited predominantly to the Southern Ocean. Some small positive differences are apparent in tropical regions and sheltered seas such as the Gulf of Mexico and Mediterranean Sea, where sea states are typically low.

In both Figures 1 and 2, "trackiness" of the data is apparent, attributable to the high resolution and fairly short record (~17 years).

(A) 2002-2019[JAS] CCI2021: Climatological Hs mean

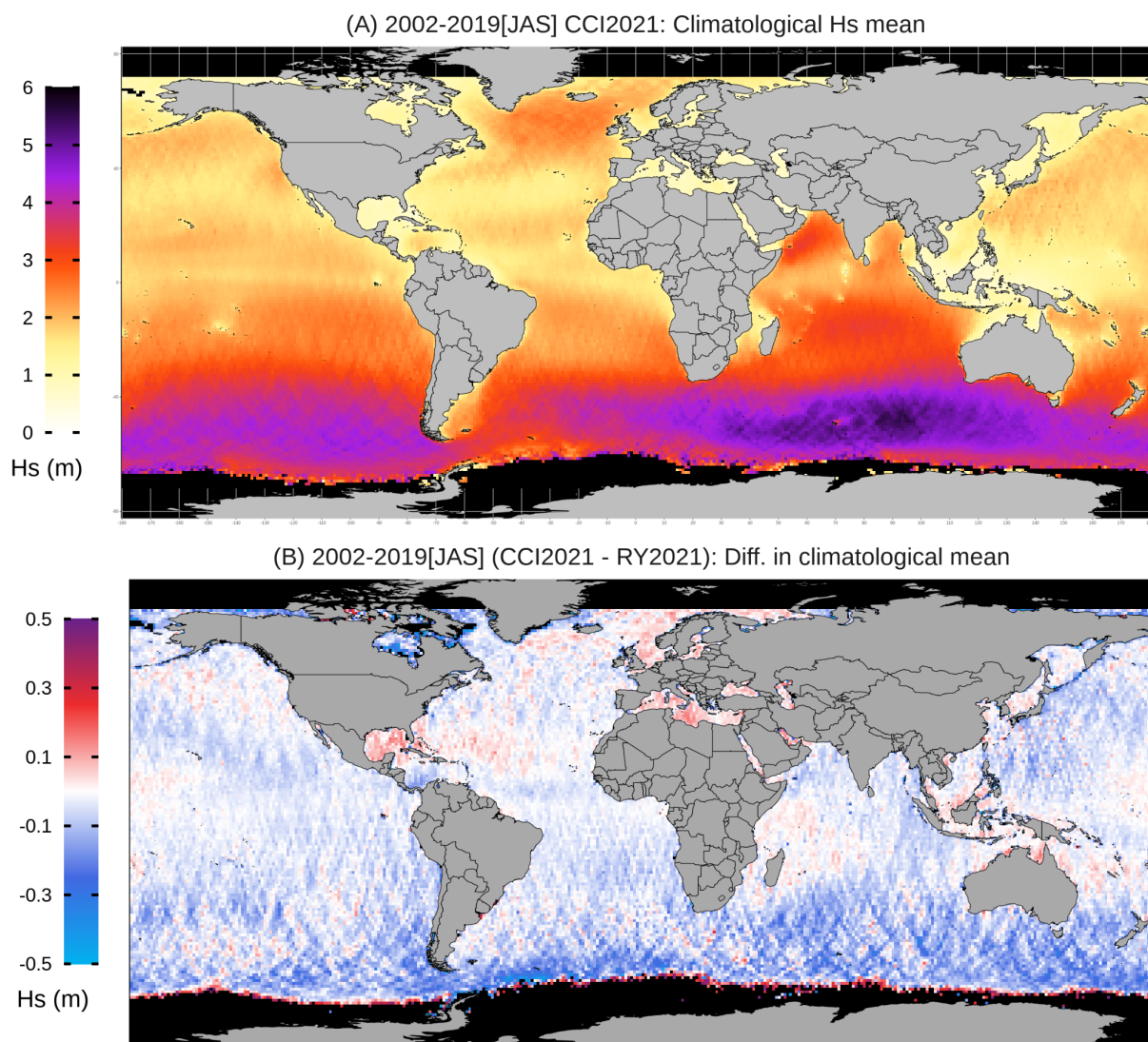(B) 2002-2019[JAS] (CCI2021 - RY2021): Diff. in climatological mean

Figure 2 : A) Climatological seasonal mean Hs for the aggregated CCI L2 altimetry products over 2002-2019 at 1 degree resolution; B) Difference in mean with RY2019.

3.3.2 Global estimates of Hs from imaging SAR (4 deg)

Estimates of sea state variables are now available from imaging SAR. Here we examine global coverage of significant wave height (*swh*) from both Envisat and Sentinel-1 (A & B).

The sparsity of sampling from a single mission is somewhat limiting with respect to the calculation of long term climatological sea state properties. Figure 3 shows the global count statistics for Envisat, where the data are aggregated to 4 degrees. Figure 4 shows similar data for Sentinel-1.
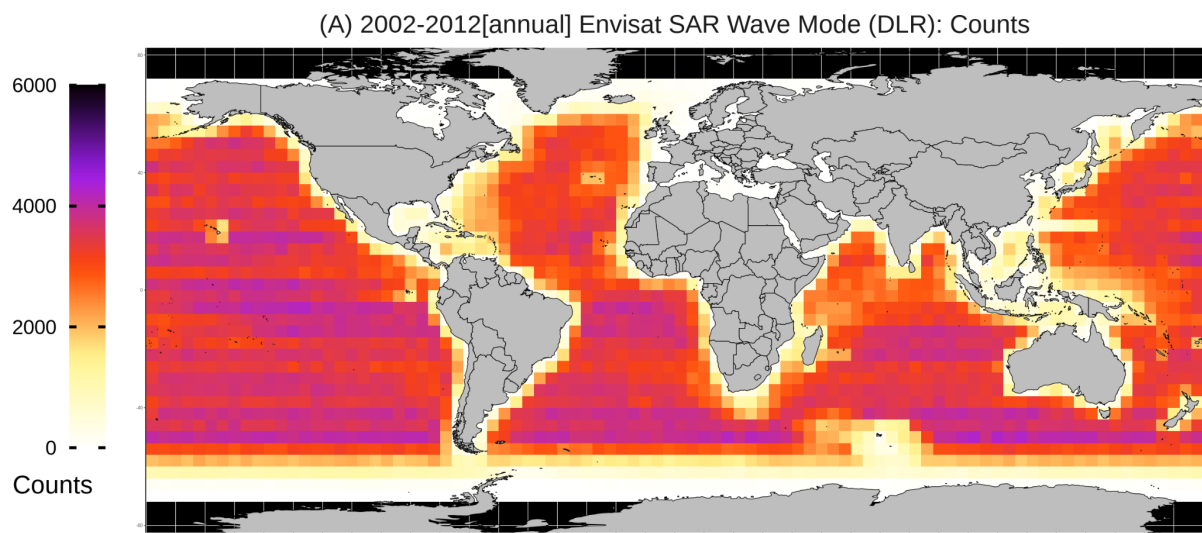

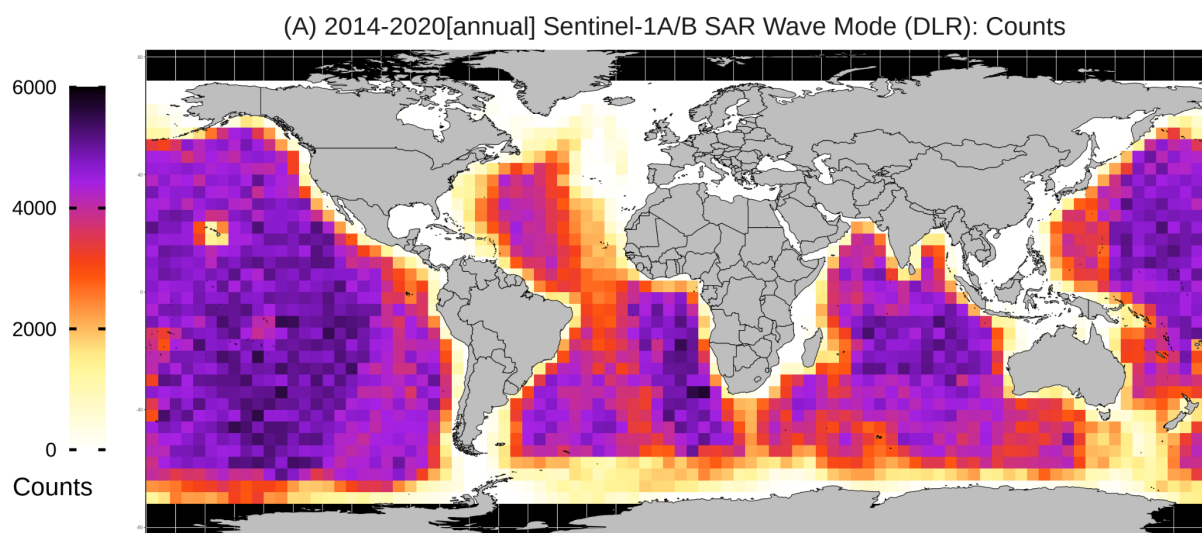Figure 3 : Total counts at 4 degree aggregation for Envisat imaging SAR.

Figure 4 : Total counts at 4 degree aggregation for Sentinel-1 imaging SAR.

Data aggregation at this scale smooths out the trackiness but also reveals the limitations of spatial coverage which, for both missions, is largely absent in coastal zones, sheltered seas and polar regions. This is due to the lack of SAR "wave mode" operation in these regions where other instrument modes are preferable. Additionally, much of the North Atlantic is not

covered due to lack of wave mode operation in this area. To a crude approximation, Envisat samples at approximately half the rate of Sentinel-1. This equates, approximately, to a sampling rate of 1 and 2 per day for Envisat and Sentinel-1 respectively. The summary of counts over the entire mission record is somewhat misleading because, for Sentinel-1 in particular, temporal sampling is heterogeneous. Although a small amount of data is available at the end of 2014, it is not until the launch of Sentinel-1B (04/2016) that observation counts increase substantially. Figure 5 shows annual counts for 2014 to 2020.
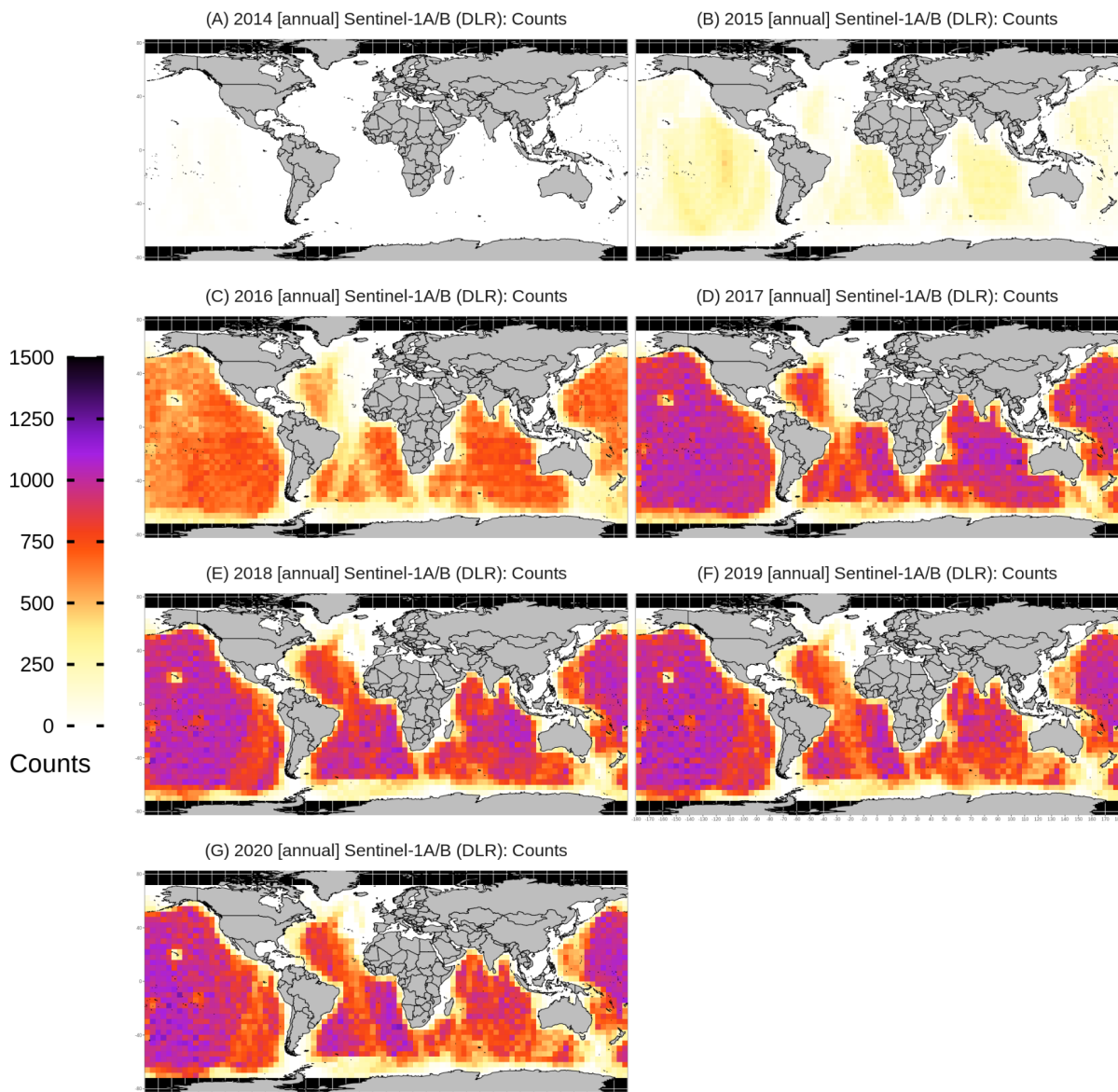

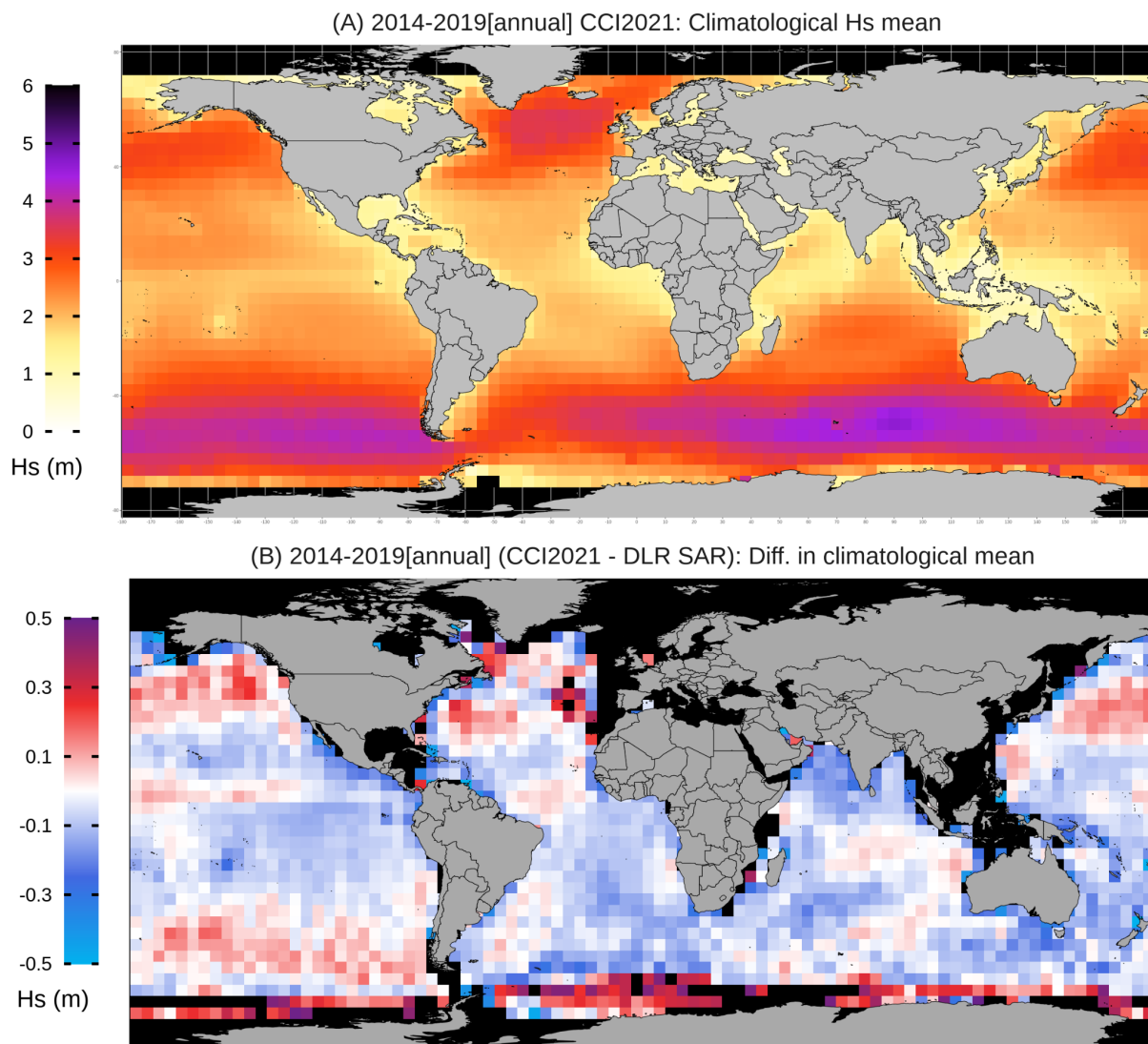*Figure 5 : A)-F) Annual counts at 4 degree aggregation for Sentinel-1 imaging SAR.*

Figure 6 : A) Climatological (annual) mean Hs for the aggregated CCI L2 altimetry products over 12/2014 - 12/2019 at 1 degree resolution; B) Difference in mean Hs with Sentinel-1 over the same period.

Figure 6 (A) shows climatological mean Hs from the L2 altimetry products (12/2014-12/2019). The difference between that and the same quantity from Sentinel-1 is shown in panel (B). Globally, differences are small (~0.1 m) with some larger regional differences. Considerable regional heterogeneity is indicative of sampling variation in the Sentinel-1 data. In addition, differences may also be attributable to considerable variation in temporal sampling rate. For Sentinel-1, sampling rate increases considerably after 2016 (see e.g. Figures 12 & 13).

3.3.3 Global mean climatology for Tm2

Noting potential impacts of low sampling rate, it is now possible to examine observations of wave period on the global domain. Figure 7 (A) shows global mean Tm2 observed by Sentinel-1 between 12/2014 and 12/2020. Limitations on spatial coverage can be seen although coverage is extensive. Regions corresponding to the development of larger sea states, typically in the southern hemisphere, are evident. Since no other direct or indirect observations on the global scale are available, we compare with ERA5 reanalysis. Figure 6(B) shows the difference when compared to ERA5. The products tend to agree within ~0.5 s, with closer

agreement over fairly large areas. Areas of most substantial disagreement are in the North Atlantic, the tropics and some coastal locations. ERA5 data appears to exceed the Sentinel-1 observations in much of the northern hemisphere.
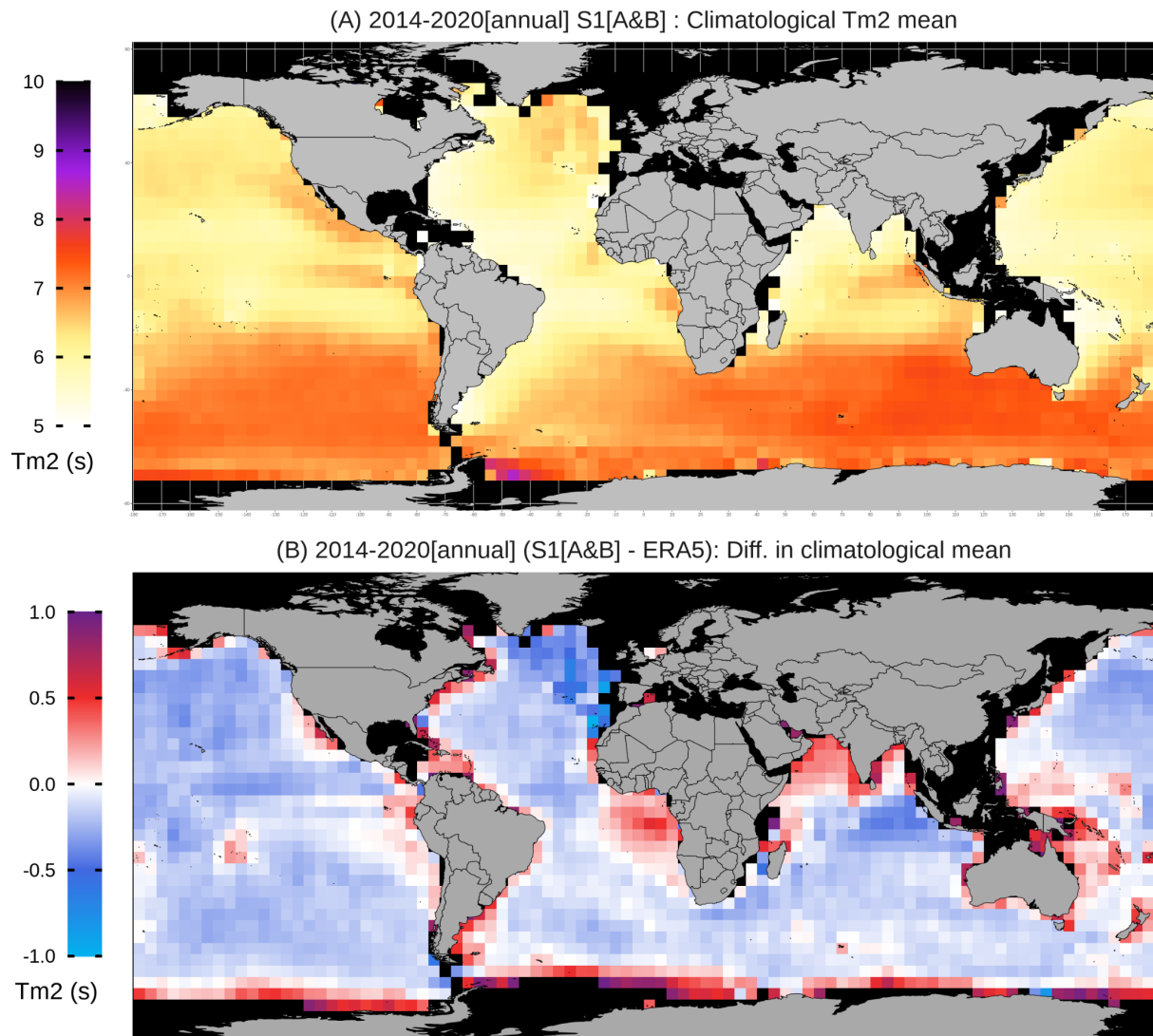

Figure 7 : A) Climatological (annual) mean Tm2 for the aggregated CCI L2 altimetry products over 12/2014-12/2020 at 4 degree resolution. B) Difference in mean Tm2 with ERA5 over the same period.

In contrast, Figure 8 (A) shows global mean Tm2 observed by Envisat between 12/2002 and 04/2012. Spatial and temporal sampling are consistent across years (not shown) and overall the open ocean is consistently observed. Sampling rates are reduced in coastal areas and sheltered seas. However, the remarkable result, shown in Figure 8 (B), is the small but consistent positive difference between Envisat and ERA5 for Tm2. While striking, the difference of ~0.5 s in Tm2 is, in fact, not attributable to errors in either dataset but due to the calculation of Tm2. Since it requires an integration over the wave frequency spectrum, it is sensitive to the high frequency cut-off. The Envisat data set is well calibrated to moored buoys (see e.g. Figure 19), but for Tm2 computed for ERA5, the cut-off is much higher than the buoy. This interesting situation occurs particularly for Tm2 that depends on higher order moments of the wave frequency distribution. It is not so pronounced for other measures of periods, such as the energy period. This issue raises questions relating to the consistency of measurements

and how dataset intercomparisons can be conducted for measures of wave period. A more detailed examination of this issue is provided in Annex A (Section 7).
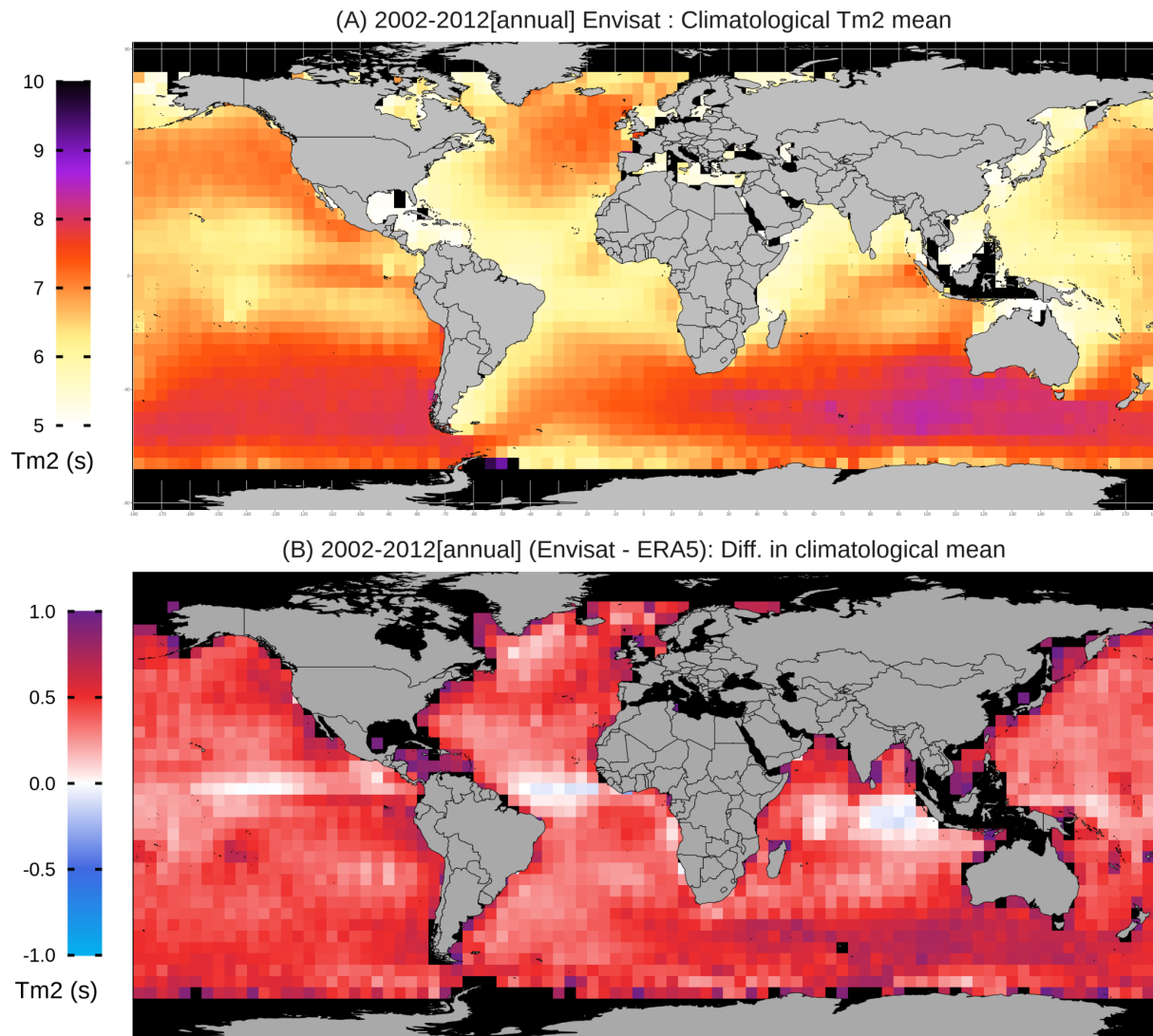

Figure 8 : A) Climatological (annual) mean Tm2 for the aggregated CCI L2 altimetry products over 12/2002-04/2012 at 4 degree resolution. B) Difference in mean Tm2 with ERA5 over the same period.

4. Validation of CCI multivariate sea state information from imaging SAR observations with respect to *in situ* data and other sea state products

4.1 Introduction

Our ability to evaluate the temporal consistency of imaging SAR data is somewhat limited by the short duration of the record, and both the sparsity of observations and limited spatial coverage (see e.g. Figures 3, 4 & 5). For example, limitations on SAR spatial coverage exclude a large number of data buoys are available on the U.S. east coast, Gulf of Mexico and Caribbean. However, in situ observations of both Hs and Tm2 are available at a number of locations concurrent in the Pacific. Imaging SAR observations are therefore initially assessed by evaluating time series of mean Hs at a number of relevant locations with available long-term in situ wave buoy datasets. Buoy data are obtained from the US National Data Buoy Centre (NDBC) which provide some of the longest and most complete wave buoy records.

The geographical locations of the selected buoy sites are shown in Figure 9. Included in these comparisons, for reference, are time series of the annual mean Hs obtained from CCI L2 V2 altimetry product and the satellite altimetry derived by RY2019.
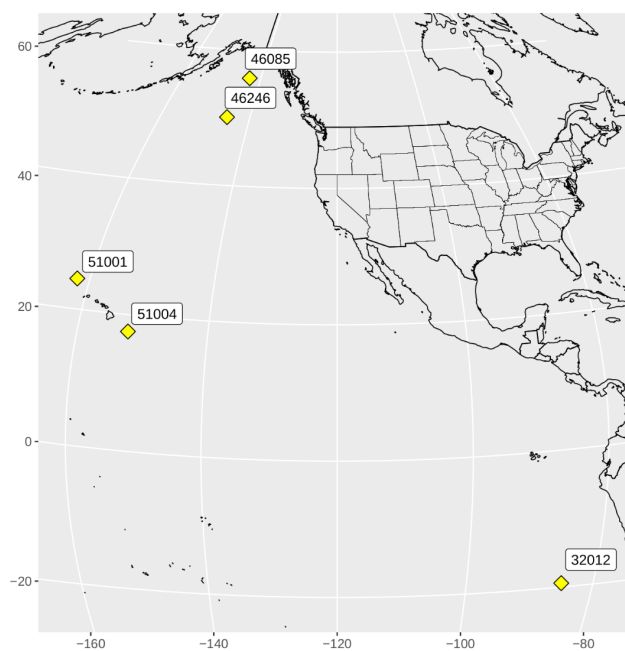

Figure 9 : Geographical locations of the NDBC wave buoys used for validation of the Sea State CCI products in this section.

4.2 Methodology

4.2.1 Comparison

Results are based on comparisons of either annual or seasonal mean Hs and Tm2 from buoys and gridded products aggregated to a common grid (typically 1 or 4 degrees). *In-situ*

observations from data buoys are typically recorded at 30 minutes or hourly intervals. Buoy data have therefore been aggregated as appropriate to enable comparison with CCI and other products over the period for which enough buoy data are available. We note that buoy records frequently feature data gaps that can range from hours to months, assumed to represent intermittent instrument malfunction or seasonal maintenance schedule.

Geospatially, the collocation between the gridded satellite products and the buoys has been achieved by selecting the (4 degree) grid cell that coincides most closely with the buoy.

One notes:
- The choice of buoys is driven by sites that coincide well with satellite coverage. This is more limited for the imaging SAR (see e.g. Figures 3 and 4), so five deep water offshore locations are considered here.
- No correction has been applied to the buoy measurement to account for changes in buoy hull type, maintenance or operation.
- The impact of missing data & quality control choices: we applied a threshold to retain only buoy records where a minimum of 85% of possible observations from the buoys are available, otherwise the full year is discarded. For this reason, missing years are clearly evident in all buoy data time-series in Section 4.3.

4.2.2 Temporal trend analysis

Some short term temporal trends are shown in the following figures and have been computed assuming a linear regression model (year covariates) under Gaussian errors, in a manner similar to that in version 1 of this report. Trend analysis is not a major focus of this report, in part due to the limited duration of time series available in the version 2 of the Sea State CCI altimetry V2 products.

4.3 Time series of seasonal mean Hs at selected buoy locations

As a consistency check for agreement between data sources with respect to long term variability, Figures 10 & 11 show time series of seasonal (JFM) mean Hs from two wave buoys in the North Pacific compared with CCI SAR (Sentinel-1 and Envisat), CCI L2 and RY2019. Note that since the number of observations from the individual SAR missions is low, data have been aggregated over 4x4 degree grid cells in order to provide a more stable estimate of the seasonal mean. Since the two buoys shown here (stations 46246 & 46085) lie in the northern hemisphere, JFM was chosen to isolate higher energy sea states. Regression lines (dashed coloured lines) were fitted to the buoy time series, using the methodology described in Section 4.2.
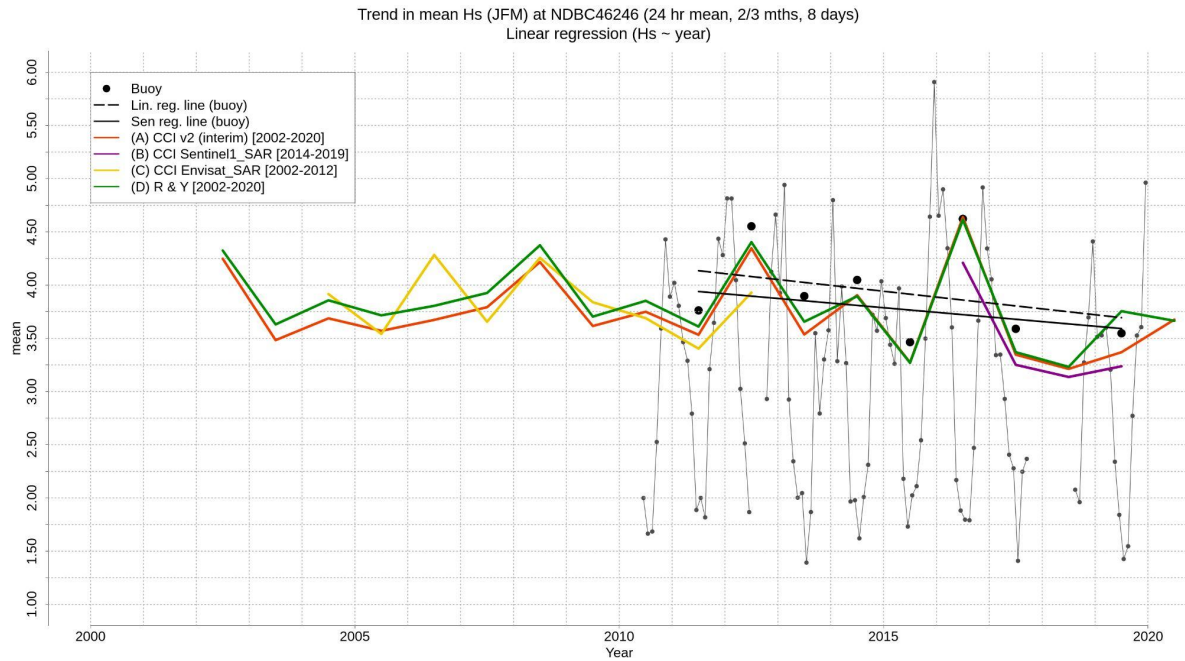
4.3.1 Significant wave height from SAR

Figure 10 : Time series of seasonal mean (JFM) Hs from buoys and various satellite products at NDBC wave buoy station 46246 in the North Pacific.
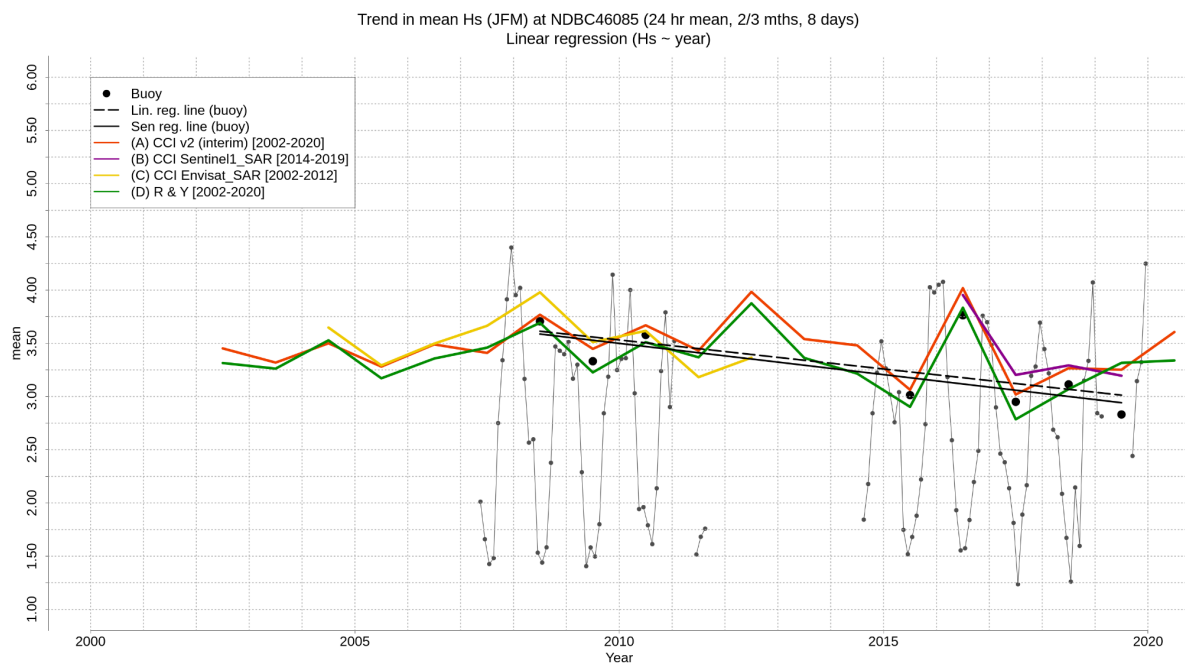

Figure 11 : Time series of seasonal mean (JFM) Hs from buoys and various satellite products at NDBC wave buoy station 46085 in the North Pacific.

4.3.2 Observation counts

In addition to the estimates of seasonal Hs, the relative stability of successive mean estimates is informed by the observation count, which varies considerably. Figures 12 & 13 show that observation counts from altimetry (CCI V2 and RY2019) have increased substantially with

increasing missions. However, intra-mission counts also vary considerably. Imaging SAR observations from Envisat are fairly stable from 2004 onwards but had substantially fewer (~50%) counts in 2003. Note that only a few observations were available in 12/2002. On the other hand, observations from Sentinel-1 were quite limited until 2017, when Sentinel-1B was launched. Interestingly however, the abundance of observations from Sentinel SAR in 2018 and 2019 amounted to approximately 50% of the counts provided by the combined CCI V2 altimetry products in the same period. Additional active missions (HY-2, Sentinel-3A, Sentinel-3B) are included in the RY2019 dataset, increasing the observation count by ~100%. Observations from imaging SAR therefore provide an appreciable increase in observations throughout the record although noting that, when used in isolation, intra-mission data density may vary considerably. Note also that only two locations in the North Pacific are considered here so these conclusions may not apply elsewhere.
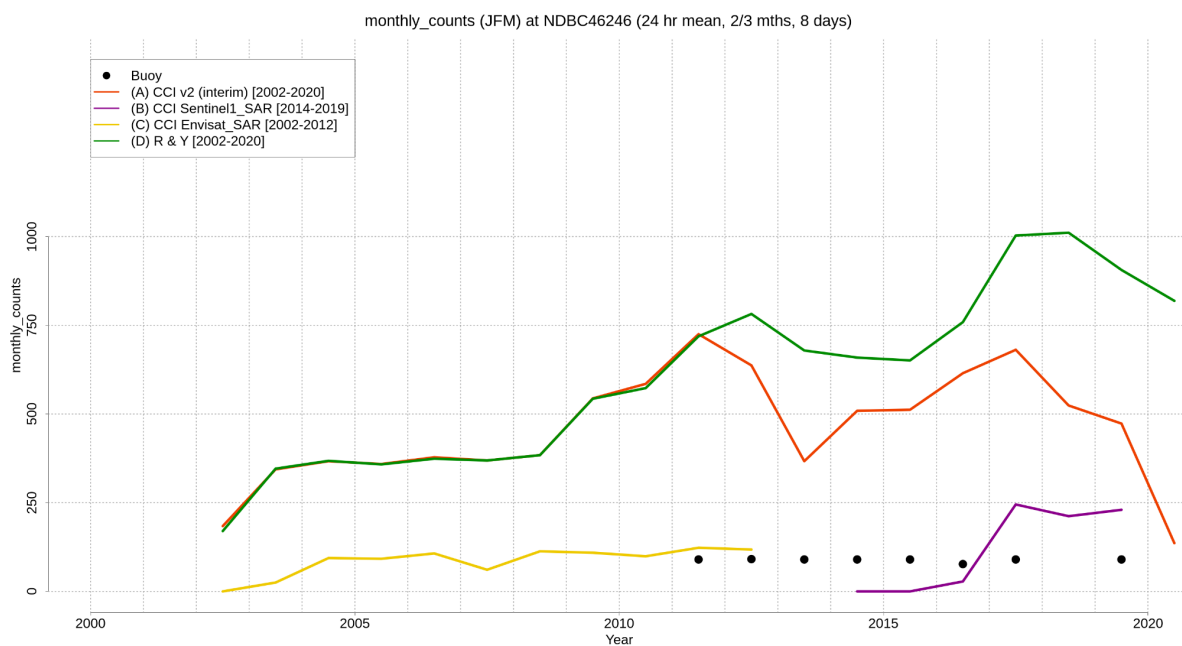

Figure 12 : Time series of seasonal (JFM) observation counts from buoys and various satellite products at NDBC wave buoy station 46246 in the North Pacific.
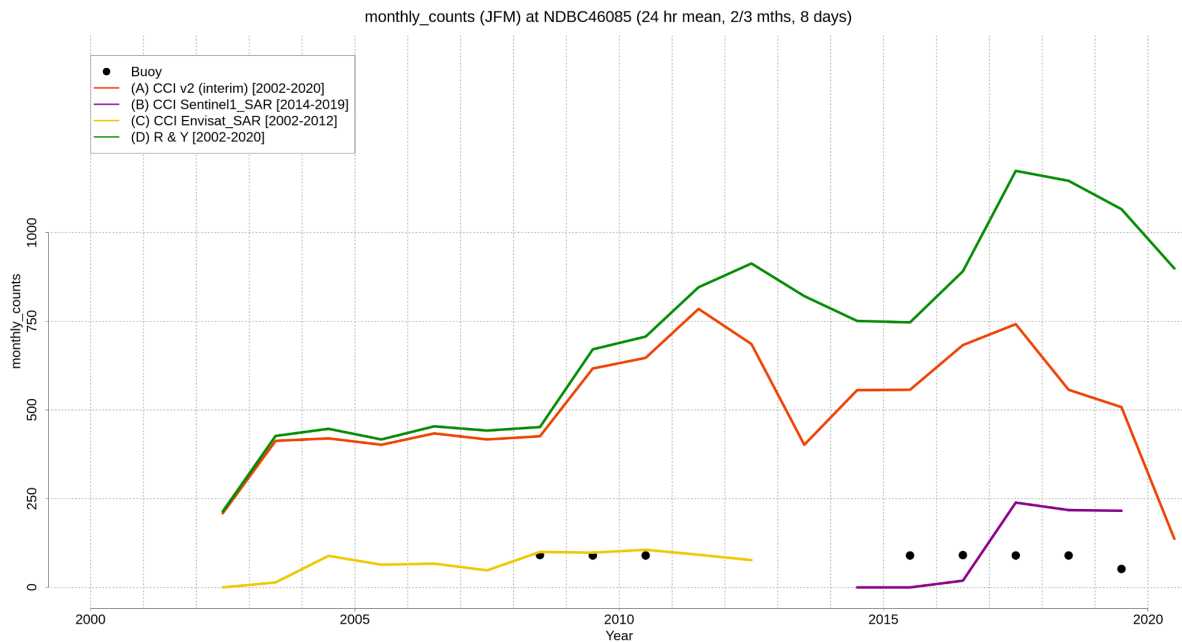
Figure 13 : Time series of seasonal (JFM) observation counts from buoys and various satellite products at NDBC wave buoy station 46085 in the North Pacific.

4.3 Paired analysis for Sentinel-1 imaging SAR derived Hs and Tm2 at selected buoy locations

In order to examine SAR observed Hs and Tm2 more closely, hourly time series from SAR and in situ buoys are compared. Owing to the lack of coverage in nearshore locations, these analyses are limited to sites in the open ocean.

4.3.1 Significant wave height from Sentinel-1 imaging SAR

Observations of Hs approximately co-located in space and time with five moored buoys are shown in Figure 14. NDBC buoy stations 46246, 46085, 51004, 51001 and 32012 have been used. For hourly time series, the observations may be separated by up to one hour temporally and a sampling of radius of 100 km around the buoy was chosen. Noting that the locations considered here are in the open ocean and subject to slowly spatially varying conditions, more stringent criteria were not shown to have a measurable impact on the results.
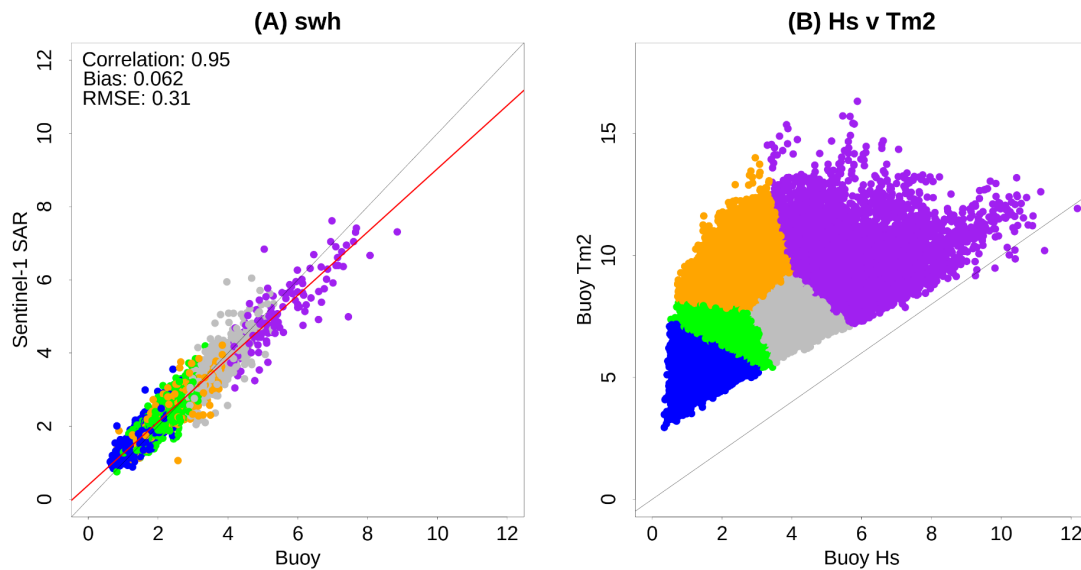
Buoys (100km): 46246,46085,51004,51001,32012

Figure 14 : A) Scatter plot of Sentinel-1 observations of Hs at NDBC wave buoy stations in the North Pacific. B) Joint distribution of Hs and Tm2 at all the buoys. Note that colours are generated through cluster recognition and designate sea state characteristics, as opposed to specific buoys or geographic location.

In Figure 14, observations from all five buoys have been aggregated in order to examine sea state generally. Panel (A) shows high correlation (0.95) and low bias (0.06 m). RMSE is 0.31 m. In order to examine whether there is a dependence on sea state, panel (B) shows the joint distribution of Hs and Tm2 (aggregated from all the buoys). Colours are assigned through cluster analysis, based upon the joint distribution, and not related to specific buoys, thus the joint sea state is identified independently of buoy location. Qualitatively, the colours can therefore be used to identify whether agreement between the satellite observations and in situ measurements are sea state dependent. However, the agreement seen in Figure 14(A) appears to be approximately consistent over all sea states. In general, lower energy sea states appear to be somewhat overestimated and higher energy sea states slightly underestimated.

These results, while encouraging overall, are decomposed by location and season to reveal detail. Figure 15 shows ten panels corresponding to the JFM and JJA seasons comparisons at each buoy location. The regional and seasonal sea state characteristics can be seen but the consistency of agreement between the satellite data and in situ observations is striking. Bias is almost always positive, particularly for lower sea states.
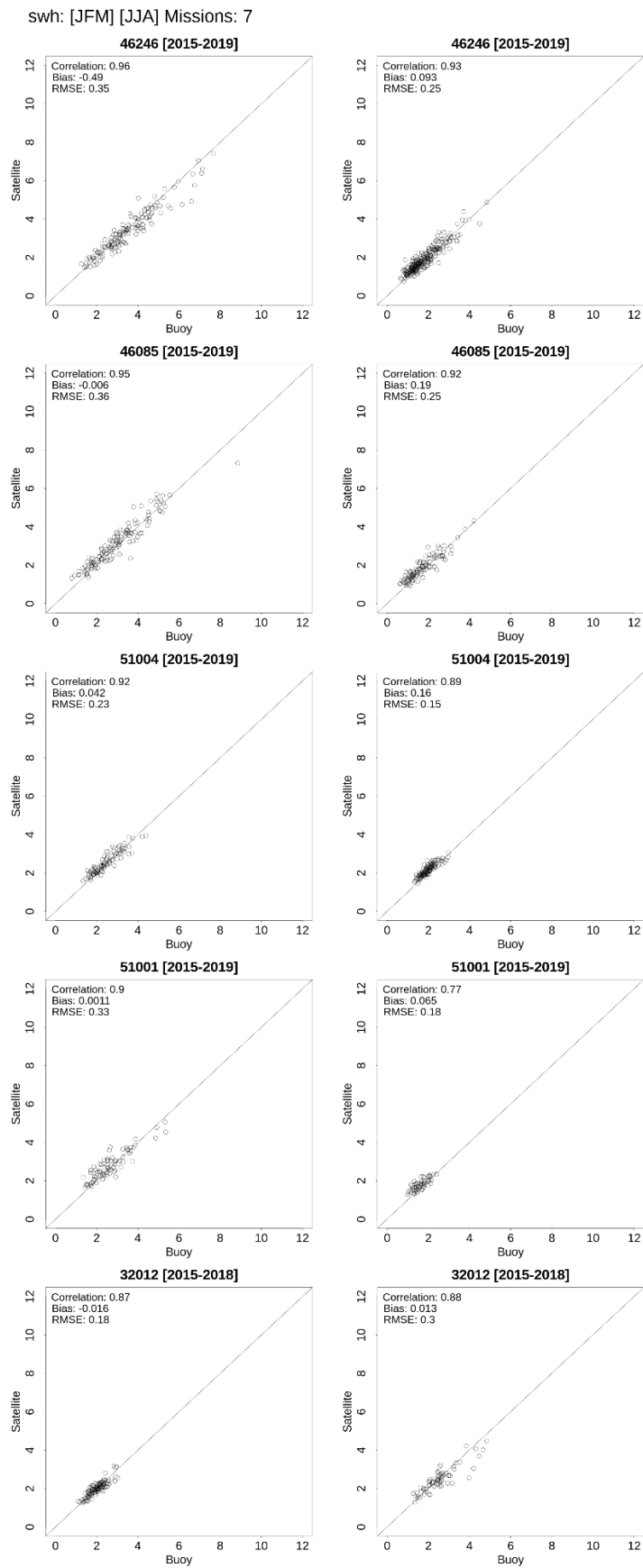
Figure 15 : Scatter plots of Sentinel-1 observations of Hs at NDBC wave buoy stations in the Pacific for JFM (left panels) and JJA (right panels).

4.3.2 Zero-crossing period (Tm2) from Sentinel-1 imaging SAR

Observations of Tm2 are analysed in the same way as section 4.3.1. Tm2 approximately co-located in space and time with five moored buoys are shown in Figure 16. Again, NDBC buoy stations 46246, 46085, 51004, 51001 and 32012 have been used.
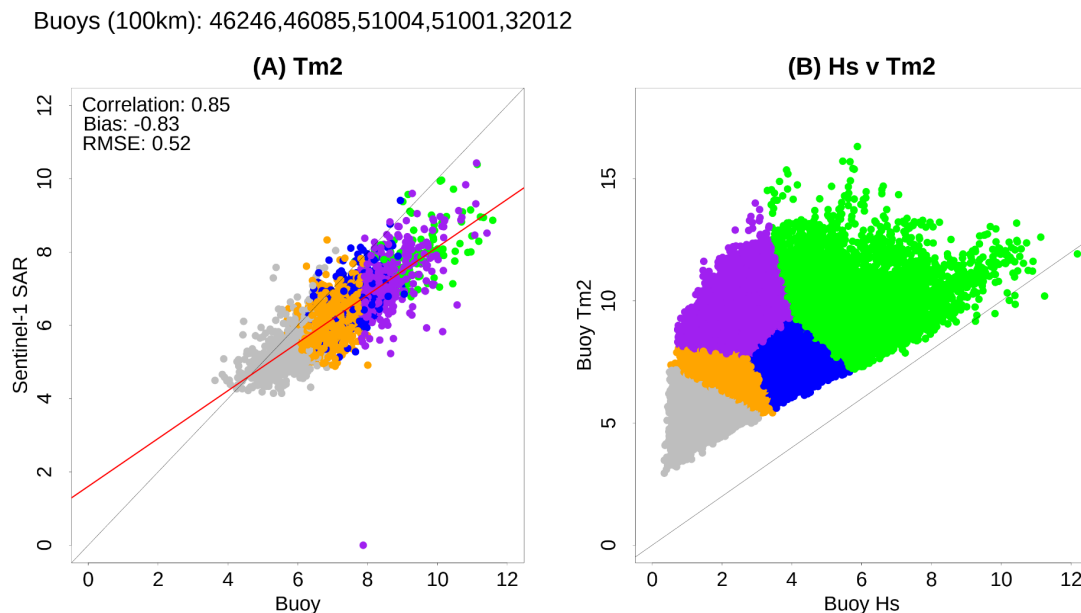

Figure 16 : A) Scatter plot of Sentinel-1 SAR observations of Tm2 at NDBC wave buoy stations in the North Pacific. B) Joint distribution of Hs and Tm2 at all the buoys (same as Fig. 12B). Note that colours are generated through cluster recognition and designate sea state characteristics, as opposed to specific buoys or geographic location.

In Figure 16, observations from all five buoys have been aggregated in order to examine sea state generally. Panel (A) shows good correlation (0.85) and a bias of -0.83 s. RMSE is 0.51 s. In order to examine whether there is a dependence on sea state, panel (B) shows the joint distribution of Hs and Tm2 (aggregated from all the buoys). Colours are assigned through cluster analysis, based upon the joint distribution, and not related to specific buoys, thus the joint sea state is identified independently of buoy location. Qualitatively, the colours can therefore be used to identify whether agreement between the satellite observations and in situ measurements are sea state dependent. However, as for Hs (Figure 14), the agreement seen in Figure 16(A) appears to be approximately consistent over all sea states. In general, Tm2 appears to be somewhat underestimated in all sea states, and increases as sea state grows.

Decomposition by location and season is shown in Figure 17. While the general characteristics seen in Figure 16 are evident, it is apparent that at least one location, buoy station 46085, shows considerably lower bias (-0.13 and -0.29 s) than others across both seasons. Elsewhere, bias is less than -0.5 s (often less than -1 s) although correlation is typically greater than 0.8, sometimes above 0.9. RMSE lies between 0.23 and 0.58 s.
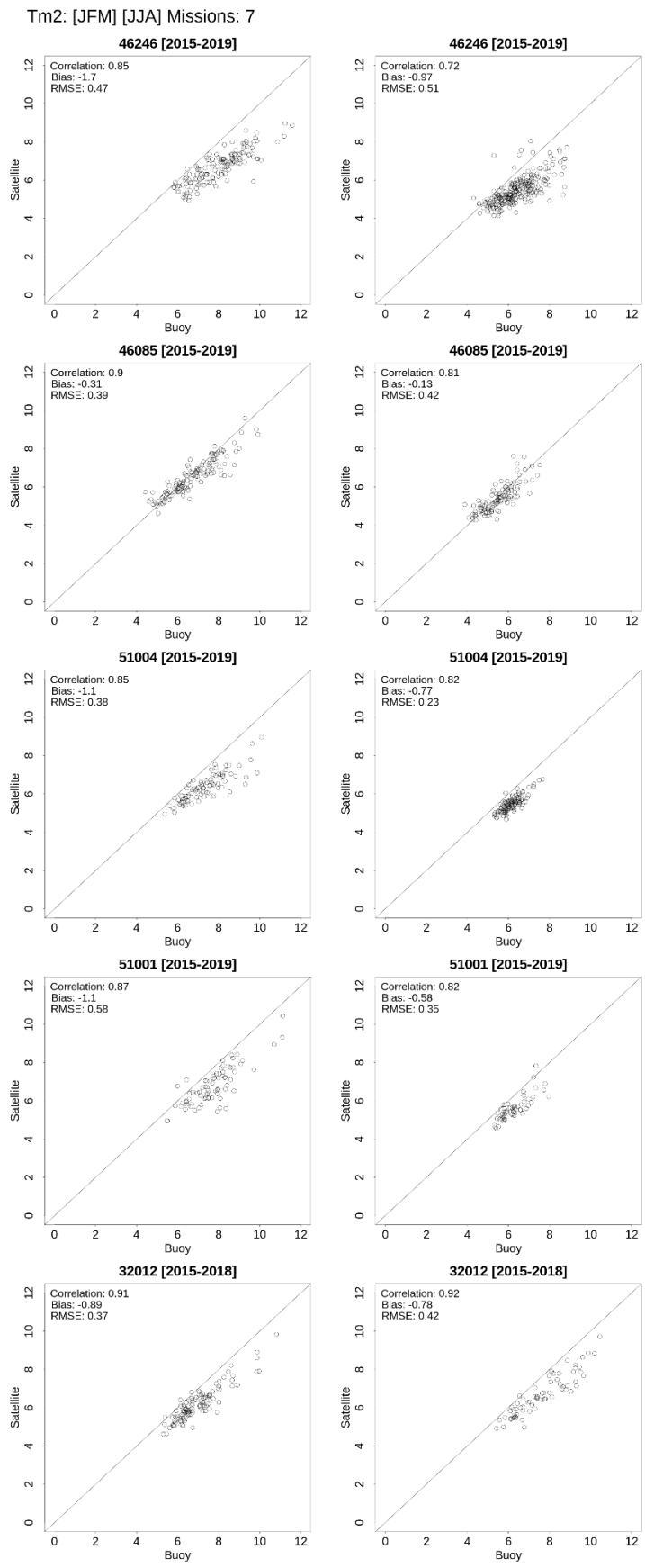
Figure 17 : Scatter plot of Sentinel-1 observations of Tm2 at NDBC wave buoy stations in the Pacific for JJA (left panels) and JFM (right panels).

4.4 Paired analysis for Envisat imaging SAR derived Hs and Tm2 at selected buoy locations

In order to examine SAR observed Hs and Tm2 more closely, hourly time series from SAR and in situ buoys are compared. Owing to the lack of coverage in nearshore locations, these analyses are limited to sites in the open ocean.

4.4.1 Significant wave height and Tm2 from Envisat imaging SAR

For comparison, observations of Hs and Tm2 from Envisat imaging SAR, approximately co-located in space and time with five moored buoys, are shown in Figures 18 & 19. NDBC buoy stations 46246, 46085, 51004, 51001 and 32012 have been used, as for Sentinel-1.

In Figure 18, observations of Hs exhibit generally low bias (typically negative) with RMSE that tends to scale with sea state. Correlation ranges from 0.71 to 0.94 but on average is in the lower part of this range. Both correlation and RMSE are noticeably poorer than for Sentinel-1 where, for example, correlation in both seasons and at nearly all buoys exceeds 0.9.

A comparison of performance of observations of Tm2 (see Figure 19) reveals similar properties to Hs. Specifically, bias in Envisat measurements is generally small and negative, whereas correlation and RMSE appear generally better for Sentinel-1. As for Sentinel-1, no dramatic dependence of performance on sea state is evident from the results although a more detailed analysis would be required to make a more robust assessment.
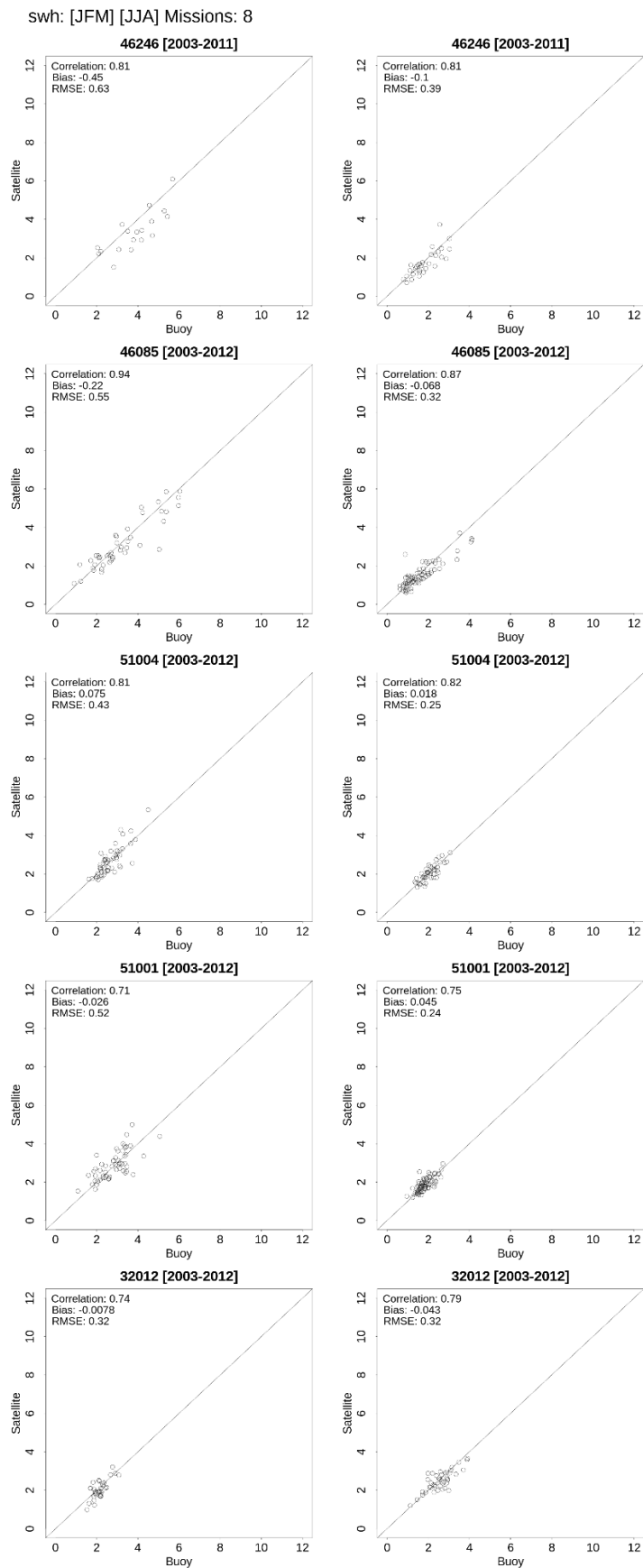
Figure 18 : Scatter plots of Envisat observations of Hs at NDBC wave buoy stations in the Pacific for JFM (left panels) and JJA (right panels).
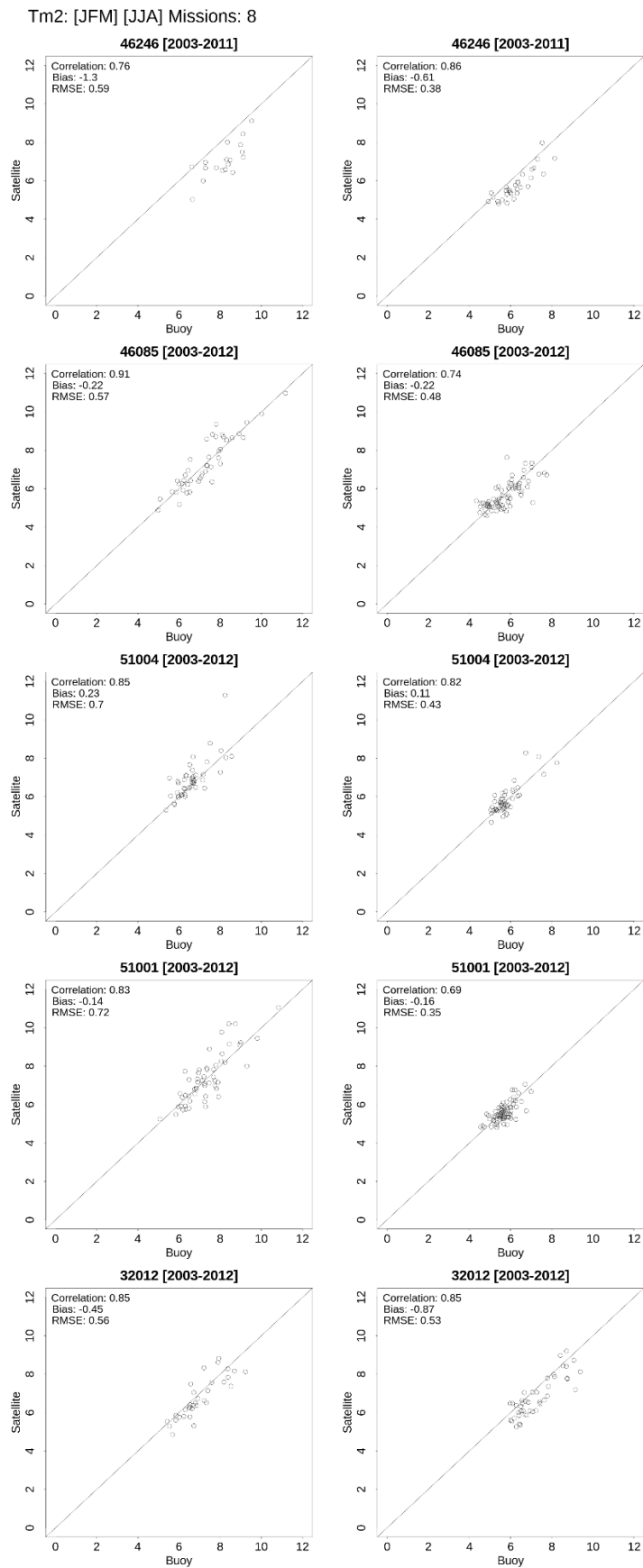
Figure 19 : Scatter plot of Envisat observations of Tm2 at NDBC wave buoy stations in the Pacific for JJA (left panels) and JFM (right panels).

4.5 Discussion & Conclusions

At the global scale, observations of Hs from Sentinel-1 imaging SAR appear to agree well with that from the altimetry products of the Sea State CCI. At a limited number of locations in the Pacific Ocean, temporal variation appears to be well represented and a more detailed comparison with in situ observations reveals good seasonal agreement.

Observations of Tm2 from Sentinel-1 imaging SAR also appear to agree with in situ measurements at a limited number of locations. Furthermore, at the global scale, SAR observations were found to be in good approximate agreement with ERA5. However, regional differences were identified that are as yet not attributable. We note considerable variability in sampling throughout the duration of the Sentinel-1 data which may contribute to differences but there are no other global scale observations of wave period with which to perform additional comparison. Noting also that ERA5 wave period data is based largely upon numerical simulation, there is no strong reason to believe that estimates of Tm2 provided in ERA5 are more accurate.

5. Assessment of annual sampling increase in Hs estimation from the combined Version 3 L2P altimeter and imaging SAR observations [2002 - 2019]

5.1 Introduction

In Section 4 (see e.g. Figures 12 & 13), when combined with data obtained from altimeters, the imaging SAR datasets derived from Envisat and Sentinel-1A/B are shown to result in a substantial overall increase in observation counts in two geographic areas. Given the reasonably spatially homogeneous coverage of those missions over large parts of the global oceans (see e.g. Figures 3 & 4), it is clear that analyses in many regions may benefit from increased sampling provided by these missions. Note, however, that the lack of coastal coverage limits this improvement to investigations on larger scales. In this section we examine the characteristics of annual mean Hs estimates from the combined version 3 data set (L2P) with respect to Hs from moored buoys shown in Figure 20.
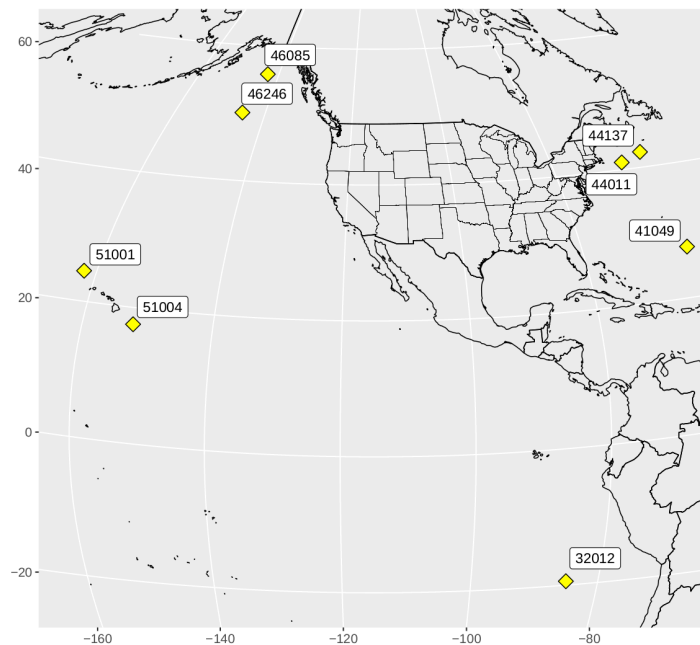

Figure 20 : Geographical locations of wave buoys used for validation of the Sea State CCI products in this section. Note that for reasons of brevity, only limited results are shown.

In particular we highlight the numbers of observation counts and, using a bootstrap resampling method, show that annual mean estimates from the CCI product are almost always statistically consistent with measurements from the buoys. It can be seen that as the sample density increases (in more recent years), discrepancy in the mean is more robustly detected. However, sampling from both satellite and buoys is still insufficient to resolve differences below O(10cm) and is also regionally and sea state dependent. Furthermore, this problem becomes more challenging when sampling is reduced further, e.g. by considering extremes (not covered here).

We start by examining data from 2019 only, since this is the most data abundant year available. Intercomparison between satellite and buoy is performed by using only paired hourly data, thus ensuring the same temporal period is represented by the two sources. Owing to the very large number of comparisons and figures only representative examples are shown. Following this, similar analyses are summarised for the complete record of the

version 3 data [2002 - 2019]: Jason-1 (ALT), Envisat (ALT), Jason-2 (ALT), Cryosat-2 (ALT), SARAL Altika (ALT), Jason-3 (ALT), Sentinel-1 (I-SAR), Envisat (I-SAR), Sentinel-3A (ALT). Note that in the following figures, #1 - #9 correspond to the various missions in the above order.

5.2 Statistical methods

In this section, analyses are based on hourly time series. For comparisons with buoys, satellite data are sampled within a radius of 100 km. The buoys considered here lie offshore and the local sea state tends to be homogeneous on this spatial scale. A smaller radius also leads to lower sampling rate. In addition to standard statistical methods, a bootstrap approach is used to evaluate whether estimates of the annual mean from the buoy and satellite are statistically different. Briefly, under the null hypothesis, samples of the sea state (within a given year) obtained from either a buoy or satellite (within 100 km of the buoy position) are assumed to be drawn from the same distribution. The pooled data (from both sources) is sampled using the number of samples from the satellite(s) (which is typically very much lower than the buoy). The mean is computed for each sampling round. This is carried out 10,000 times leading to a (Gaussian) sampling distribution for the mean that provides a basis for a significance test. A statistically significant difference (e.g. at the 5% level where the satellite observed mean lies outside 2 s.d.) could potentially indicate an important difference.

5.3 Intercomparison during 2019

As an example, Figure 21 shows data from Cryosat-2 (#4) at buoy 46246 located in the North Pacific (See Figure 20). The same analysis is shown for Sentinel 3-A (ALT, #9), Figure 22 and Sentinel-1 (I-SAR, #7) in Figure 23.
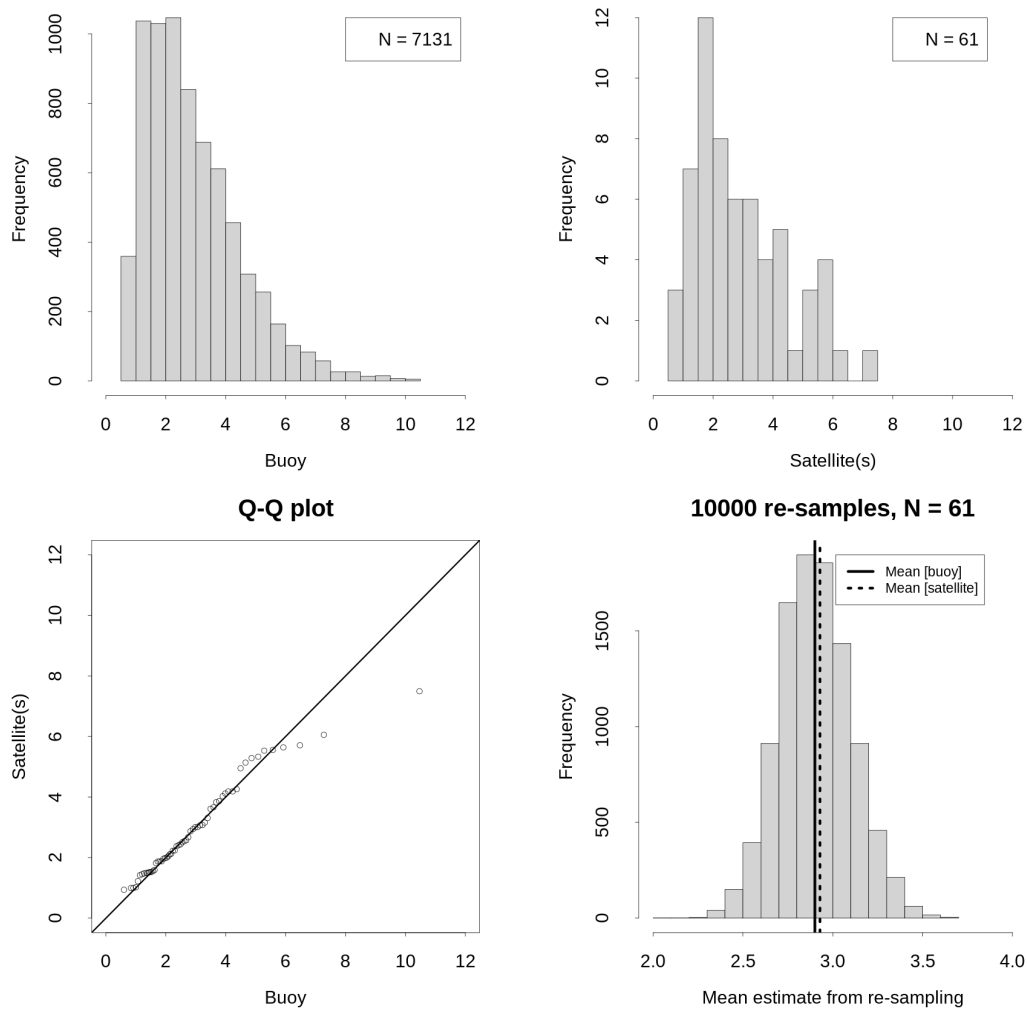
46246 [swh] 2019 [annual] Mission #: 4

Figure 21 : Comparison of annual Hs (m) observations from buoy and satellite (Cryosat-2, ALT #4). Top row: Histograms. Bottom-Left: Q-Q plot. Bottom-Right: Bootstrap estimate of the mean from the buoy data, based upon satellite samples (N=61).

Figure 21 is remarkable in the sense that there is excellent agreement in the mean (bottom-right panel). As a result, the estimate from the satellite falls well within the bootstrap sampling distribution. However, the spread of the distribution is quite wide owing to the low number of samples (N=61). In fact, a difference of approximately +/- 0.5 m could reasonably be argued to be within the sampling uncertainty (say, within 2 s.d.). A much larger number of samples is required to perform a more precise comparison.
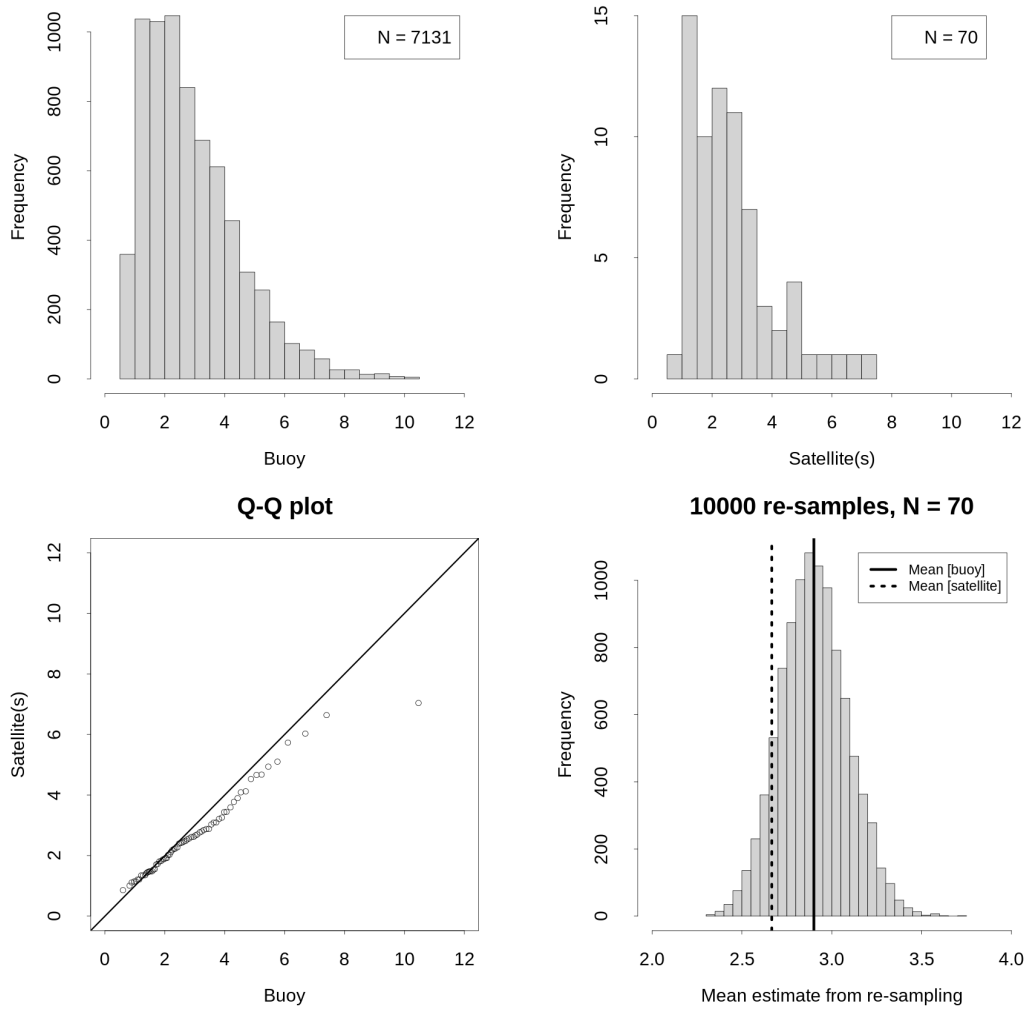
46246 [swh] 2019 [annual] Mission #: 9

Figure 22 : Comparison of annual Hs (m) observations from buoy and satellite (Sentinel-3A, ALT #9). Top row: Histograms. Bottom-Left: Q-Q plot. Bottom-Right: Bootstrap estimate of the mean from the buoy data, based upon satellite samples (N=70).
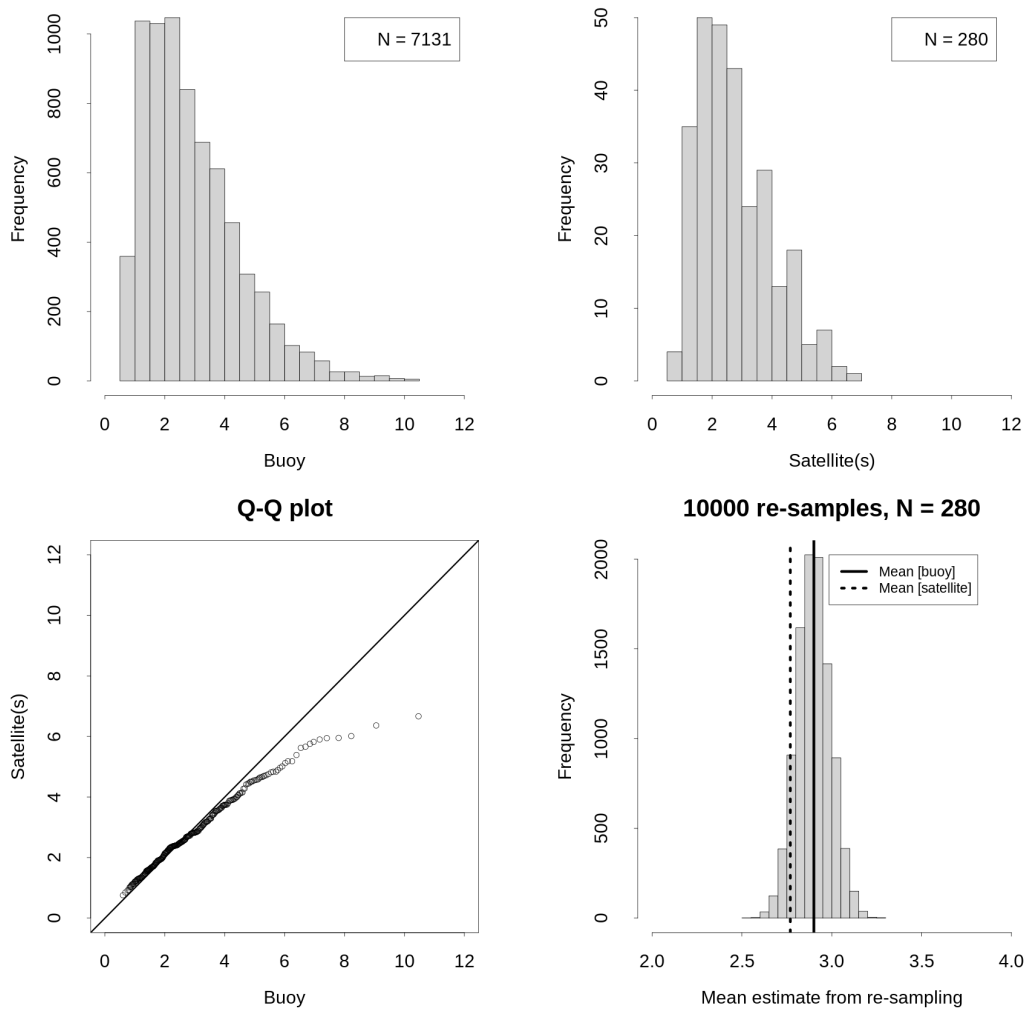
46246 [swh] 2019 [annual] Mission #: 7

Figure 23 : Comparison of annual Hs observations from buoy and satellite (Sentinel-1A/B, I-SAR #7). Top row: Histograms. Bottom-Left: Q-Q plot. Bottom-Right: Bootstrap estimate of the mean from the buoy data, based upon satellite samples (N=280).

A similar situation is seen for Sentinel-3A in Figure 22. The number of samples (N=70) is slightly larger than for Cryosat-2 although the absolute difference between the buoy mean and satellite mean is somewhat larger (0.25 m), but still well within sampling uncertainty. The Q-Q plot suggests a slight underestimate of Hs at higher values. Figure 23 tells a somewhat different story. Sentinel-1 imaging SAR data is considerably more abundant, N=280 compared with N=61 for Cryosat-2 and N=70 for Sentinel-3A, but disagreement in the mean of approximately 0.15 m is apparent (see also Figure 13, top right panel). However, bootstrap sampling based on N=280 suggests that the mean difference is not statistically significant at (say) the 5% level. Observe also that this is in spite of the reduced variance of the sampling distribution, compared with Cryosat-2 (Figure 21), attributable to the much larger sample.

5.4 Analyses of complete records (at selected buoys)

At 46246, the summary of a similar analysis, as shown in Figures 21, 22 & 23, but for the entire in situ observed record, and the entire combined CCI dataset (ALT & I-SAR), is shown in Figure 24.
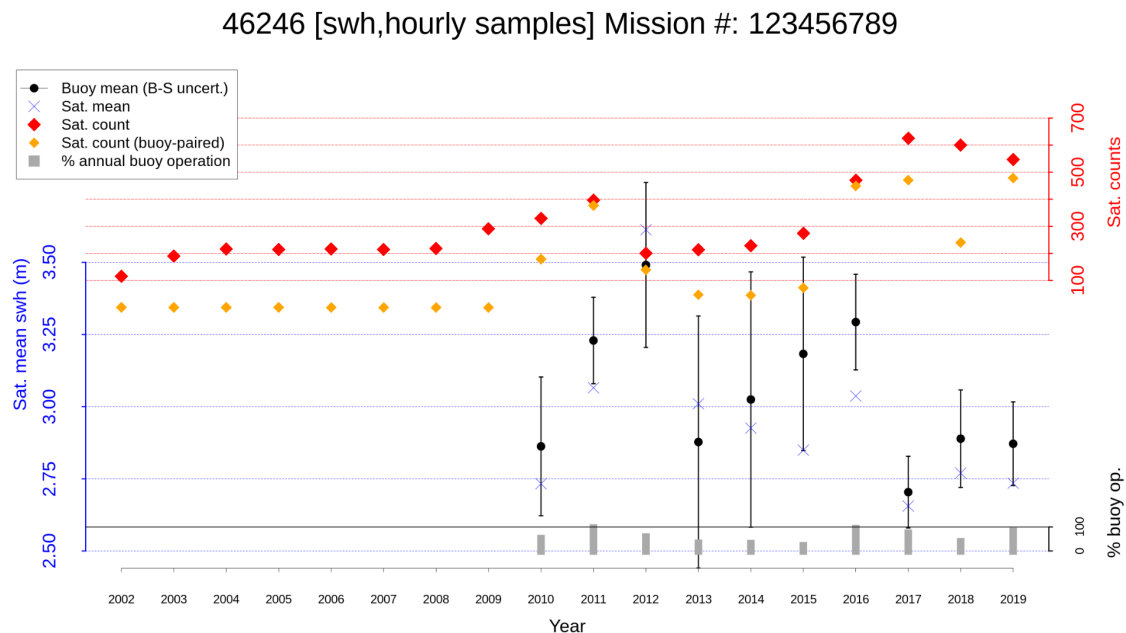

Figure 24 : Time series intercomparison of annual mean Hs at buoy 46246. Data availability from the buoy is shown in the bar chart at the bottom. Total satellite counts are shown in the time series at the top (orange & red diamonds). Black error bars show the 2 s.d. spread due to sampling uncertainty.

With reference to the bottom right panels of Figures 21, 22 & 23, in Figure 24, blue crosses correspond to the satellite estimate (the dashed black line), black circles correspond to the mean estimate from the buoy (the solid black line) and the error bars correspond to +/-2 s.d. of the sampling distribution. Total combined observation counts from the CCI products are shown by red diamonds. In order to maintain a consistent comparison between buoy and satellite, periods affected by data loss are removed from the comparison. Such periods are dominated by typically prolonged periods of operational outage that affect buoys. The satellite record tends to suffer from few disruptions. Data availability from the buoy is shown in the bar chart at the bottom. This also corresponds with the orange diamonds that indicate loss of data for direct comparison. For example, in 2018, ~50% of hourly buoy data is missing. As a result, although a total of ~600 satellite observations are available, the orange diamond indicates that less than 50% (~250) of those could be used for comparison.

We make a few specific comments. Firstly, interannual variability of total satellite data is high following 2010, as missions are introduced or removed from service. While there was a considerable drop in data abundance in 2012, following the retirements of Jason-1 and Envisat, the latest years show the largest observation counts. This corresponds to the steady reduction in uncertainty. Encouragingly, we see that agreement in mean Hs between the satellite record and the buoy is good, and, with the exception of 2016, falls within the expected statistical variability. At this particular location, the satellite product appears to slightly underestimate mean Hs.
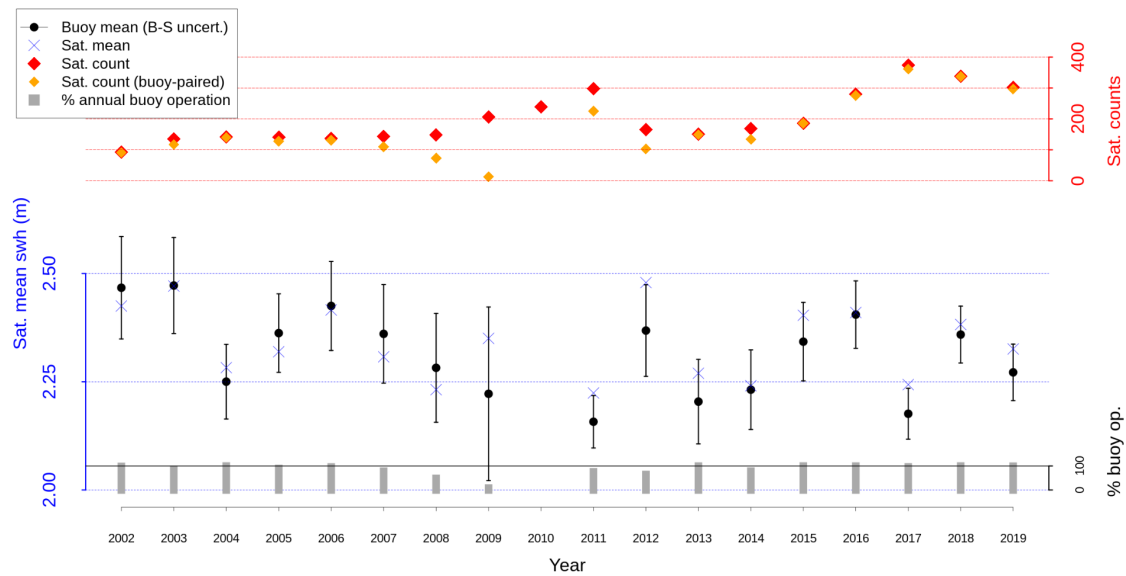
Figure 25 : Time series intercomparison of annual mean Hs at buoy 51004. Data availability from the buoy is shown in the bar chart at the bottom. Total satellite counts are shown in the time series at the top (orange & red diamonds). Black error bars show the 2 s.d. spread due to sampling uncertainty.

For comparison at a different location, a similar analysis is shown at 51004 near Hawaii in Figure 25. Data availability from the buoy is longer term and far more consistent throughout the record. Temporal variability is well captured and good consistency between satellite and buoy mean Hs estimates is apparent in all available years.

Additional locations have been examined in the same way but are not shown here for brevity. In general, consistent with the calibration approach, very few instances of statistically robust disagreement in annual mean Hs have been identified.

6. Future work

This section outlines some of the issues and ideas that have been raised by analyses so far, and would benefit from further study.

6.1 Multivariate analysis of the sea state

While results suggest that multivariate information such as Hs and Tm2 obtained from imaging SAR (Envisat and Sentinel-1A/B, to date) is of high quality, analysis of the joint distribution remains to be performed. This offers the potential to validate and explore sea state, and longer time climate, much more comprehensively. For example, information about the joint distribution can be used to validate models and reanalysis, swell can be examined explicitly (from Sentinel-1A/B) and multivariate quantities such as wave power can, potentially, be assessed globally, without reliance on numerical simulation.

6.2 Estimating uncertainty

In this report, Section 4 has revealed the impact of the increasing data density from satellites. Increasing data abundance allows us to evaluate more robustly where disagreement between products exists, and the rapid rate of increase of data is encouraging - particularly from imaging SAR products. However, it is clear from this, and earlier work, that the heterogeneous distribution of temporal sampling throughout the satellite record gives rise to difficulties in obtaining robust long term analyses. Questions still remain over whether the complete continuous record, from 1992 (version 1 data) can be usefully exploited for the evaluation of trends and extremes.

As such, more detailed examination and quantification of uncertainty is recommended. For example, weighted regression schemes and bootstrap methods may be appropriate. Nonetheless, it is clear that relatively low levels of sampling will continue to hamper analyses of longer term variability from satellite data.

6.3 Impact of buoy platform changes

It is well known that data buoys receive maintenance periodically throughout their operational life times. This can lead to fairly small, but relevant, systematic shifts in observed variables such as Hs that have been shown to affect the evaluation and robustness of long-term trends (Gemmrich et al., 2011). These effects have not been considered in this report, as they raise very difficult questions about the suitability of the existing in situ wave observing network for climate applications (Timmermans et al., 2020).

6.4 Regional trends / link with climate oscillations and indices

Regional modes of variability can strongly affect sea state and are of great importance when estimating trends, particularly when data records are short. On interannual to decadal timescales, sea state may be affected by climate variability indices such as the North Atlantic Oscillation (Dodet et al. 2010) or Pacific Decadal Oscillation. It is important to verify that the CCI products accurately capture such variability, and that these are correctly accounted for in estimates of longer-term climate variability.

7. Annex A - Impact of the high frequency cut-off on the computation of the mean wave periods

A1.1 Introduction

When comparing mean wave periods, it is important to consider the frequency range over which the mean wave periods are determined. For certain applications, the impact of the high frequency cut-off might be important.

Mean wave periods can be computed from the wave spectrum $F(f, \theta)$. Defining the spectral moment of order n as:

$$m_n = \int_{f_0}^{f_c} \int_0^{2\pi} df \, d\theta \, f^n \, F(f, \theta)$$

Where f is frequency, θ the wave direction, f_0 the low frequency cut-off and f_C the high frequency cut-off. Then the following mean wave periods have historically been defined

Energy mean wave period:

$$T_M = T_{-1,0} = \frac{m_{-1}}{m_0}$$

Zero-crossing mean wave period:

$$T_Z = T_{0,2} = \sqrt{\frac{m_0}{m_2}}$$

Reciprocal of the mean frequency:

$$T_1 = T_{0,1} = \frac{m_0}{m_1}$$

In-situ observations such as those made by buoys are limited in their ability to observe high frequency waves above a cut-off frequency $\sim$ 0.4 - 0.5 Hz. However, wave models generally discretize their spectrum in the range, 0.035Hz < f < 0.5 - 1.0 Hz, BUT generally the frequency range is extended to infinity, by assuming a f^{-5} high frequency tail, such as done in WAM and WW3.

Historically, models have output T_M, but T_Z and T_1 might also be available nowadays. T_Z is often available from in-situ observations, but not necessarily the wave spectrum. Therefore, if one is unable to recompute the mean periods, does the choice of f_C matter when comparing mean wave periods between observations and model output?

To investigate this issue, we have run the latest version of standalone ecWAM (CY48R1), with 36 directions, 36 frequencies (0.035Hz < f < 0.97 Hz), with a global resolution of Tco639 (18 km), using ECMWF operational analysis forcing, and no wave data assimilation. We have

output global full resolution 2d-spectra every 15 days from 2018-06-05 0 UTC to 2019-06-15 0 UTC. From these spectra, we have recomputed mean periods for different f_C (0.34 Hz < f_C < 0.97 Hz) as given by the frequency discretization of the model and compared to equivalent full range values.

A1.2 Zero-crossing mean wave period

Figure A1 shows a comparison of the zero-crossing mean wave period evaluated up to a cut-off frequency of about 0.5 Hz with its counterpart computed over the full frequency range for all grid points with water depth larger than 2 m and a significant wave height of at least 0.1 m.
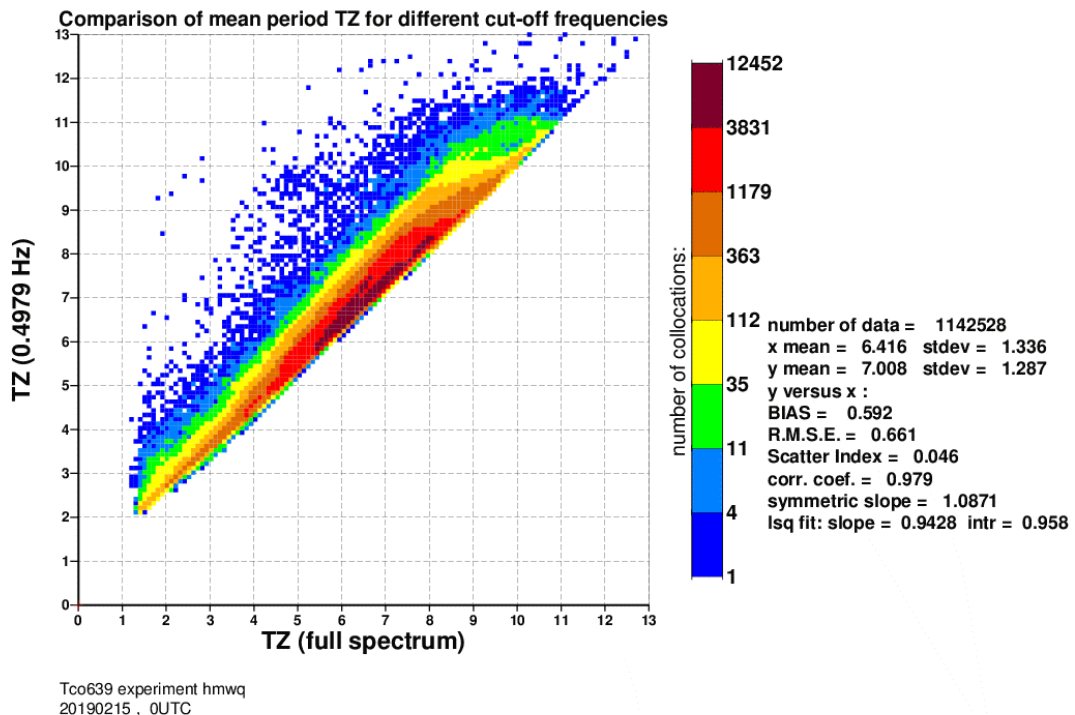

Figure A1: Comparison of zero-crossing mean wave periods determined over the full frequency range (x-axis) and with a frequency cut-off of 0.4979 Hz (y-axis).

We can now combine the different dates and the different sampled cut-off frequencies to show the impact of the cut-off frequency on T_Z. Figure A2 summarises the mean bias for different cut-off frequencies with respect to the full range T_Z. As expected, it is larger for lower cut-off frequency. The blue dots are the different dates, and the red curve is the regression line giving the mean bias. The scatter in the blue points illustrates the uncertainty in the mean bias estimate.

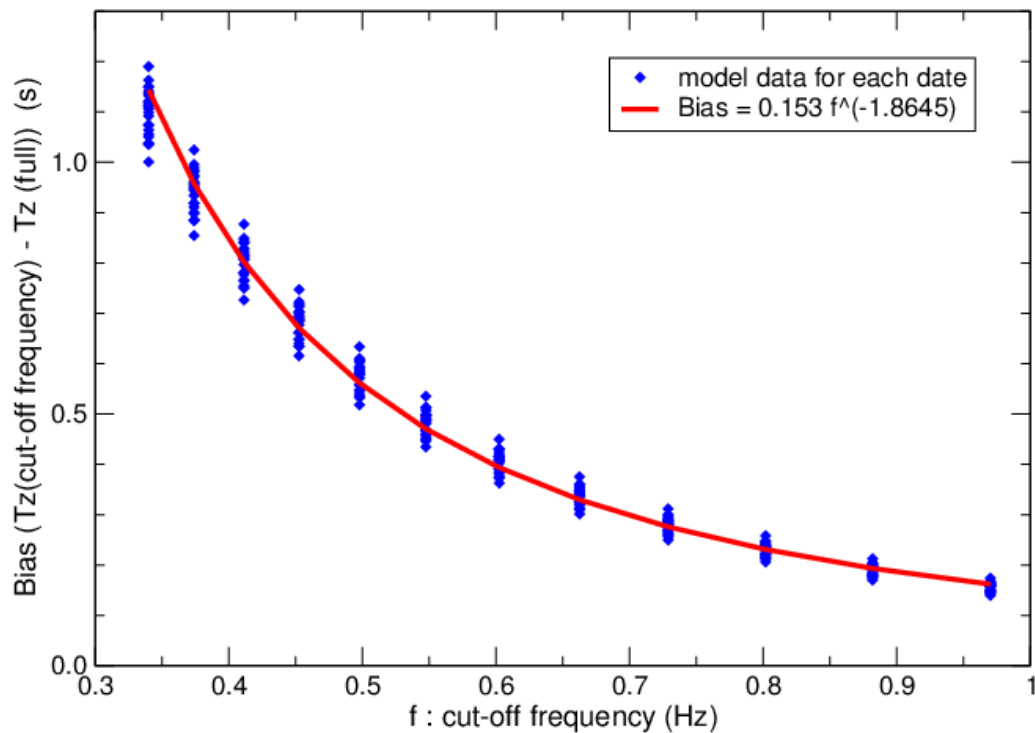## Difference in Tz for different cut-off frequency

Figure A2: Mean difference between the zero-crossing mean wave period computed up to a given cut-off frequency (x-axis) and the corresponding value determined over the full frequency range. Each blue symbol corresponds to one date from 2018-06-05 0 UTC to 2019-06-15 0 UTC every 15 days. The red curve is the fit for all data points.

From Figure A2, it is clear that one needs to be careful when comparing model outputs with other estimates of the zero-crossing mean wave period. For instance, moored buoy observations will tend to systematically return values that are on average about 0.5 second larger than the model equivalent. Similarly, remote senses estimates might have been tuned to a particular data set and hence will have an associated cut-off frequency.

A1.3 Energy mean wave period

A similar analysis can be done for the energy mean wave period T_M. Figure 3 shows a comparison of T_M evaluated up to a cut-off frequency of about 0.5 Hz with its counterpart computed over the full frequency range for all grid points with water depth larger than 2m and a significant wave height of at least 0.1 m. Compared to T_Z (Figure A1), the impact of the choice of the cut-off frequency is much reduced on T_M. This is to be expected since T_M is by definition a lot less sensitive to the high frequency part of the spectrum.
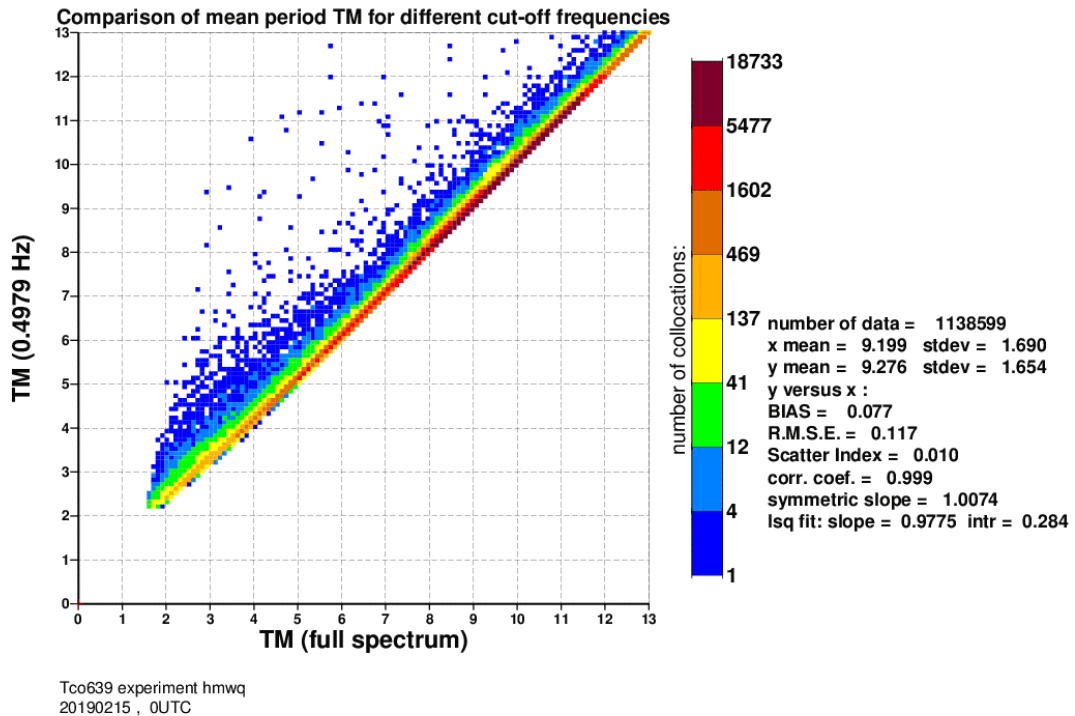
Figure A3: Comparison of energy mean wave periods determined over the full frequency range (x-axis) and with a frequency cut-off of 0.4979 Hz (y-axis).

Figure 4 summarises the mean bias for different cut-off frequencies with respect to the full range T_M. When compared to Figure A1, note that the vertical scale is much reduced. For typical buoy data, the mean bias would be less than 0.1 second when compared to the full range equivalent.

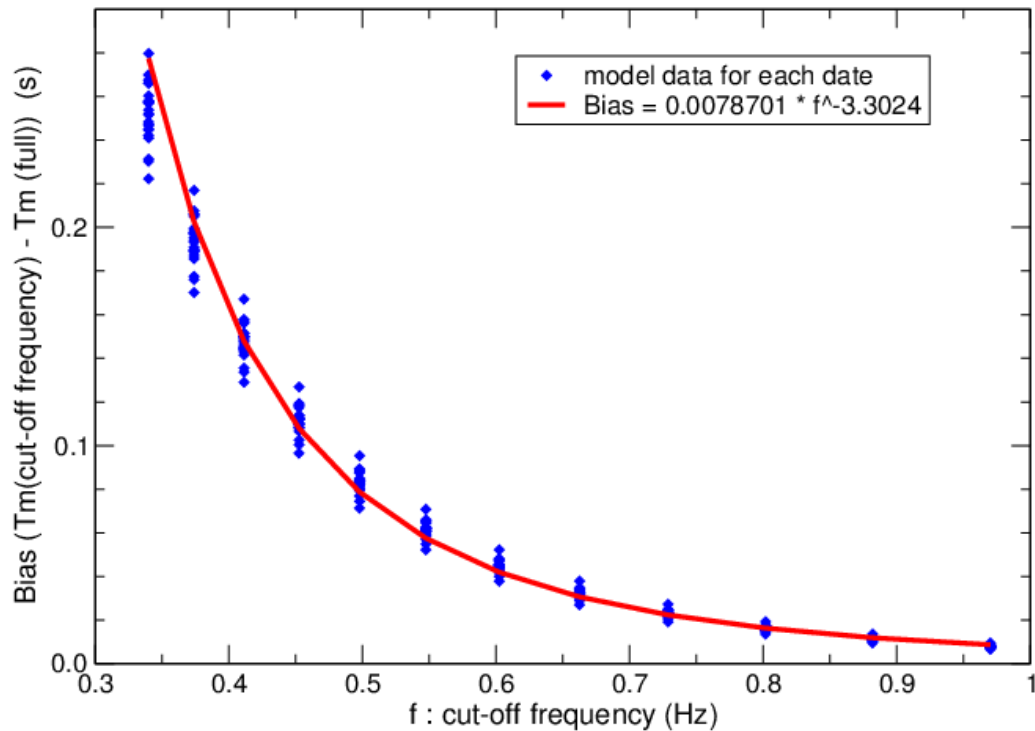## Difference in Tm for different cut-off frequency

Figure A4: Mean difference between the energy mean wave period computed up to a given cut-off frequency (x-axis) and the corresponding value determined over the full frequency range. Each blue symbol corresponds to one date from 2018-06-05 0 UTC to 2019-06-15 0 UTC every 15 days. The red curve is the fit for all data points.

A1.4 Conclusions

This simple analysis demonstrates how mean wave period estimates can be sensitive to the high frequency cut-off. It is important to keep this in mind when comparing model output with other estimates. The cut-off frequency might be imposed by the ability of a given instrument to observe the high frequency part of the wave spectrum (buoys), or implicitly imposed (SAR) by the training of the retrieval algorithm on a specific data set.

8. References

Dodet, G. et al. (2010), Wave climate variability in the North-East Atlantic Ocean over the last six decades, *Ocean Modelling*, https://doi.org/10.1016/j.ocemod.2009.10.010

Dodet, G. et al. (2020), The Sea State CCI dataset v1: Towards a sea state Climate Data Record based on satellite observations. Earth Syst. Sci. Data Discuss 2020: 1-28. https://doi.org/10.5194/essd-2019-253.

Gemmrich et al. (2011), Observational changes and trends in northeast Pacific wave records, *Geophysical Research Letters*, https://doi.org/10.1029/2011GL049518

Morim et al. (2019), Robustness and uncertainties in global multivariate wind-wave climate projections, *Nature Climate Change*, https://doi.org/10.1038/s41558-019-0542-5

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

Sen, P. K. (1968), Estimates of the Regression Coefficient Based on Kendall's Tau, *Journal of the American Statistical Association*, https://doi.org/10.1080/01621459.1968.10480934

Timmermans, B. et al. (2019), An evaluation of the consistency of extremes in gridded precipitation data sets, *Climate Dynamics*, https://doi.org/10.1007/s00382-018-4537-0

Timmermans, B. et al. (2020), Global Wave Height Trends and Variability from New Multimission Satellite Altimeter Products, Reanalyses, and Wave Buoys. Geophysical Research Letters 47(9): https://doi.org/10.1029/2019GL086880

Young et al. (2012), Investigation of trends in extreme value wave height and wind speed, Geophysical Research Letters, https://doi.org/10.1029/2011JC007753

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