



sea state
cci

Product Validation and Algorithm Selection Report (PVASR)


version 3.0, 14 September 2021

Contents

List of Acronyms	4
1. Introduction	5
1.1 Overview of document	6
1.2 General Principles of the Round Robin for Altimetry	6
1.2.1 Definitions	7
2. Description of data	8
2.1 Altimeter (LRM and DD)	8
2.2.1 Input Data	8
2.2.2 Output Data	11
2.2 Synthetic Aperture Radar	12
2.2.1 Input Data	12
2.2.2 Output Data	16
2.3 Round Robin datasets	16
2.3.1 Altimeter (LRM and DD)	16
2.3.2 Synthetic Aperture Radar	16
3. Round Robin procedures	20
3.1 Altimeter (LRM and DD)	20
3.1.1 Outlier Analysis	20
3.1.2 High Frequency Noise Analysis	21
3.1.3 Comparison with in-situ Data	21
3.1.4 Comparison with Model Data	21
3.1.5 Representation of Scales of Variability	22
3.2 Synthetic Aperture Radar	22
3.2.1 Outlier Analysis	23
3.2.3 Comparison with In-situ Data	23
3.2.4 Comparison with Model Data	24
4. Round Robin results	24
4.1 Outlier Analysis	25
4.2 Noise Analysis	27
4.3 Wave Spectral Variability Analysis	30
4.3.1 Jason-3	31
4.3.2 Sentinel-3A	32
4.3.3 Overview of Spectral Analysis	32
4.4 Comparison against Wave Models	32
4.4.1 Jason-3	33
4.4.2 Sentinel-3	34
4.5 Comparison against In-situ Data	36
4.5.1 Jason-3	37
4.5.2 Sentinel-3	38

4.5.3 Overview of Buoy Comparison Results	40
4.5.4 Perspectives	40
4.6 Round Robin SAR	41
5. Weighting matrix for results	49
5.1 Altimeter	49
5.2 Synthetic Aperture Radar	49
6. Summary	51
7. References	51

Author	Approved	Signature	Date
Florian Schlembach, Marcello Passaro, Björn Tings, Graham D. Quartly, Jean Bidlot, Francesco Nencioli, Fabrice Collard	Ellis Ash		14/09/2021
ESA Acceptance			

Issue	Date	Comments
1.0	11/02/2019	First Version
1.1	25/11/2019	Updates following ESA review
2.0	20/05/2020	Version two with updates following Round Robin conclusion
3.0	14/09/2021	Version three with corrections and updates on SAR Round Robin

List of Acronyms

ADP	Algorithm Development Plan
ATBD	Algorithm Theoretical Basis Document
cci	Climate Change Initiative
DD	Delay-Doppler
DtC	Distance to Coast
E3UB	End-to-End ECV Uncertainty Budget
ECV	Essential Climate Variable
FFT	Fast Fourier Transform
GDR	Geophysical Data Record
GTS	Global Telecommunication System
L4	Level 4
LRM	Low Rate Measurement
LUT	Look-Up Table
MAD	Median Absolute Deviation
MLE	Maximum Likelihood Estimator
NRCS	Normalized Radar Cross-Section
OSTST	Ocean Surface Topography Science Team
PHCP	Percentage of High Correlation
PLRM	Pseudo Low Rate Measurement
PTR	Point Target Response
PVASR	Product Validation and Algorithm Selection Report
RA	Radar Altimeters
RR	Round Robin
R.m.s.	Root mean square
RMSE	Root mean square error
S3A	Sentinel-3A
S3B	Sentinel-3B
SAR	Synthetic Aperture Radar
SI	Scatter Index
SSH	Sea Surface Height
S.D.	Standard Deviation
SWH	Significant Wave Height
WHALES	Wave Height Adaptive Leading Edge Subwaveform (retracker)
w.r.t	with respect to
WV	Wave (mode for SAR)

1. Introduction

This document presents the Product Validation and Algorithm Selection Report (PVASR) for the Sea State Climate Change Initiative (SS_cci), deliverable 2.1 of the project.

The SS_cci project is part of the ESA Climate Change Initiative, and aims to identify, produce and validate a sea state essential climate variable (ECV). Requirements for sea state have been expressed formally by GCOS only for a single variable: the significant wave height (SWH). SWH will therefore be the key parameter on which the evaluation of the algorithms will be based.

In order to identify the best performing algorithm or combination of algorithms, the SS_cci project held an open Round Robin (RR): an algorithm intercomparison exercise following the protocol reported in this document, which summarises the agreement found during the first months of the project within the Consortium. By maximising the number of users participating in the Round Robin exercise, ESA expected to identify the best algorithms for a future operational system. The chosen algorithm(s) were then to be implemented in an end-to-end system to generate the SS_cci data records.

The SS_cci project aims at providing SWH estimations from two different classes of sensors: Satellite Radar Altimeters (RA) and Synthetic Aperture Radars (SAR). Given the very different nature of these measurements, two different Round Robin exercises were planned. Routines and techniques to evaluate algorithms for RA and SAR were programmed by different teams and are described completely separated from each other in this document.

Concerning RA, the participation to the Round Robin was publicly advertised during the Ocean Surface Topography Science Team (OSTST) Meeting and the “25 Years of Progress in Satellite Altimetry” Conference in September 2018. Personal invitations were sent by email to the following scientists, authors and co-authors of recent publications focused on the retrieval of SWH from Satellite Altimetry: F. Birol (LEGOS), C. Buchhaupt (TU Darmstadt), L. Fenoglio-Marc (University of Bonn), S. Dinardo (EUMETSAT), W.H. Smith (NOAA), F. Peng and X. Deng (University of Newcastle), R. Roscher (University of Bonn), Xi-Feng Wang (ESST, Kyushu University), N. Kurtz (NASA), D. Sandwell (NOAA). Preliminary expressions of interest in participating were obtained from S. Dinardo (participation to Altimetry DD Round Robin), D. Sandwell (participation to Altimetry LRM Round Robin) and F. Peng (University of Newcastle).

Internal participation to the Altimetry Round Robin was guaranteed by TUM (participating to both LRM and DD Round Robin), CLS (participating to both) and IsardSAT (participating to DD Round Robin). The final list of RA algorithms evaluated is found in Section 4 of this report.

Internal participation to the SAR Round Robin is guaranteed by ODL, Ifremer and DLR. A personal invitation was sent by email to Xiaoming Li (Chinese Academy of Sciences).

1.1 Overview of document

This document is organised into the following sections:

Section 1: this section introduces the main purpose of the document and sets the general rules of the Round Robin experiment

Section 2: this section describes the input and output dataset for Satellite Altimetry and Synthetic Aperture Radar and the external dataset required to produce the statistics

Section 3: this section describes the procedures on which the Round Robin is based

Section 4: this section presents the results of the Round Robin for each participant. It includes the results of the Satellite Altimetry Round Robin in this version in this first version

Section 5: this section describes the decision process followed to select the winning algorithms

Section 6: this section summarises the main findings

1.2 General Principles of the Round Robin for Altimetry

In order to make the Round Robin as transparent as possible, a set of preliminary rules and requirements for participation are defined.

- 1) The Round Robin exercise is a transparent process. TUM and PML will, under request, share the code used for the Round Robin.
- 2) The criteria of the Round Robin must be quantitative.
- 3) The rules of the Round Robin have been approved collegially by the Consortium.
- 4) TUM, as leader of the Altimetry-Algorithm Development group, and the Science Leader of the SS CCI Project have the final word in clarifying disputes on methodology.
- 5) Proposed changes to the Round Robin methodology after the start of the exercise (KO+9) will not be considered.
- 6) The Round Robin assesses the quality of the Ku-band significant wave height at 20-Hz. The providers are invited to avoid bad practice such as: forcing to absolute zero the SWH, using external data to force retrievals at NaN. The participants shall provide a 1-0 flag to assess bad retrievals and shall describe the criteria used for it. If the algorithm allows, the authors shall provide the estimations of σ_0 . Finally, the participants must suggest the best strategy to average from 20-Hz to 1-Hz data, since the final product will be distributed as 1-Hz measurements. The Round Robin involves both internal consistency checks (outliers, along-track variability...) and validation with external data (buoys and models), as

described in this document.

1.2.1 Definitions

Here we report the definition of terminology used in the following sections of the document.

Closest point: “closest point” is defined as the median SWH of the 51 high-frequency (“51 20-Hz”) closest points to the buoy, including NaNs. Unrealistic estimations, i.e. outside the interval [-0.25 m, 25 m], are excluded.

Distance to Coast (DtC): “Distance to coast” is the distance of each 20-Hz point from the nearest coast, computed using the “Distance to Nearest Coastline: 0.01-Degree Grid: Ocean” available from <http://pacioos.org>. In this dataset, “Distances were computed with GMT using its intermediate-resolution coastline and then gridded globally at a spatial resolution of 0.04 degrees. Bilinear interpolation was then applied to increase the spatial resolution to 0.01 degrees.”

Outliers: outliers are considered points for which SWH=NaN, which lie outside [-0.25m, 25m] and/or which are more than 3*MAD away from the median of the closest 20 points.

Noise: noise is defined as the standard deviation of the 20-Hz SWH within a 1-Hz distance.

2. Description of data

2.1 Altimeter (LRM and DD)

2.2.1 Input Data

The Round Robin evaluation focuses on a subset of Jason-3 and Sentinel-3A mission data.

The following table provides the list of input products collected and made available to round-robin participants, for both missions:

Mission	Product	Version
Jason-3	SGDR	version D
Sentinel-3A	SR_1_SRA_BS_	Operational NTC (TBC: unless 2018 reprocessing can be provided by Eumetsat)
	SR_1_SRA_A_	Operational NTC (TBC: unless 2018 reprocessing can be provided by Eumetsat)

For the comparisons with models and in-situ data, participants needed to process 2 full years of data along a small number of selected tracks passing close to the in-situ data. The analysis covers the period July 2016-June 2018 to cover a balanced mix of winter/summer conditions.

The input data from Jason-3 and Sentinel-3A were collected by Ifremer repository only, to avoid later discrepancies if different sources are used.

The following tracks for Jason-3 were selected:

Pass	Area	Potentially collocatable in-situ observations
026	Gulf of Mexico	42039, 42003, 42097, 42057
039	South Pacific - Canadian East Coast - Iceland	32012 44141, 44139, 41046, (42058), (42065), (TFGRV), (TFKGR)

045	New Zealand, Hawaii, Canadian West Coast	(NZban), 51209, 51002, 51100, 51000, 46004, 46208, 46183, (45150), (45141)
063	North Atlantic to Norwegian Sea	41040, (41066), FAWV3, LF3N, LF5I, FL7I, LF3N
065	US East Coast Offshore - Caribbean Sea	44174, (44024), 44008, 44004, 41001, 41002, 42057
070	English Channel - Celtic Sea	62105, 62107, 62023, 62094 (the Porthleven buoy could be added)
094	North Sea to Iceland	62129, 62130, 62134, LF4C, LF5C, 62102, 62161, 64046, FAWV3, FSWV2, TFBLK
113	West Ireland, Norwegian Sea	62095, 62105, 62106, 62048, (64046), 62302, 63118, 64041, 63115, 62169, 63117 63101
120	North Sea	EURO, 62142, 63058, 62144, 62289, 62142, 62148, 62127, 62131, 62165, 62293, 64045
127	Korea	21229, 22105, 22106, 22190, (22104), (22188)
137	Canary Islands to Finland (Portuguese and Finnish data will need to be acquired)	(13131), FARO, 62192, 62025, 62001, (62074), 62288, 62044, 62286, 25077, FIMR3
163	North Sea - Irish Sea - Brazil	62146, 62122, 62119, 62094, 31229, 31231, 31375
172	Western Europe	62095, 62081, 62029, 62163, (62001), 62025, 61417
206	US West coast offshore	46076, 46085, 46184, 46004, (46036), (46006), 46022, 46212, 46255, 46219, (46047)
243	US East Coast - Iceland	41010, 41025, 41062, 44087, 44066, 44097, 44090, 44020, 44018, 44005, (TFBLK), TFKGR
244	Norwegian Sea - Arabian Sea	LF3N, LF5I, LF7I, 23456, 23451, 23494, 23453

The following tracks for Sentinel-3A have been selected:

Pass	Area	Potentially collocatable in-situ observations
011	South Pacific to Gulf of Alaska	46006, 46085
074	Norwegian Sea - Spain	LFB2, (62046), 6201, 62X55, 62X54, 62163, 62083, (62084)
084	Antarctica to Western Canada	46006, 46036, (46208), 46183
099	Hawaii to Bering Sea	51001, 51003, 51101, 46073, (46035)
115	Portugal to Iceland	62085, 62191, 62084, 62082, 62029, 62081, 62095, TFSRT, TFSTD
142	Gulf of Alaska to Hawaii	51206, 51100, 51000, 56246, 46085
181	Southern California offshore	46047, 46219, 46255, 46249, (46053), (46087)
207	Caribbean Sea - Gulf of Mexico	42056, 42003, (42039), 41012
226	West of Canada to Antarctica	46206, (46002), 46059
250	US East coast offshore to Caribbean Sea	44017, 44066, (44087), (41001), 41002
285	Mediterranean to Norwegian Sea	61002, (61431), 64046, FAWV1, FAWV3, FAWV4
319	Caribbean to Canada	41300, 41044, (44150), 44024
355	Hawaii -Bering Sea	51000, 51100, 51004
371	West Scotland to the Mediterranean Sea	62048, 62047, 62091, (62301), Pembrokeshire buoy, 62078, 61281
393	India	23091, 23093, 23455
463	Caribbean to Florida	42058, 41010, 41004
474	Norwegian Sea - South Atlantic	(FAWV2), (FAWV3), 64045, 62105, 62095

478	US East Coast Offshore - South Pacific	44032, 44005, 44018, (44020), 41001, (41002), 32012
491	US East Coast	44032, 44005, 44018, (44020), 41001, (41002), 32012
493	South Pacific	32012
508	Gulf of Mexico - Lake Superior	42055, 42002, (45004)
513	Mediterranean Sea -North Sea	61021, 62142, 63058, 62144, 62148, 63059, 63111, 62302, 63118, 64041
514	New Zealand to Alaska	46075, NzBan
530	Portugal to Norwegian Sea	62200, FARO, 62025, 62001, 62293, 62162,62118, 62155, 62168, 62161, 62130, 62128, LF8F, 63057, LF4H, LF6T, LF5B, LF5E, LF5A, LF7I, LF5I, LF3N, LF5T
592	Labrador Sea to South eastern Pacific	45138, (44005), (44037), 44008, 41001
646	Iceland - Brazil	TFSRT, TFGRS, 31375, 31053
656	Gulf of Alaska to Hawaii	51209, (51003), 51001, 51101, (46066), 46078, 46264, 46080, 46076
695	US West Coast Offshore	46083, 46004, 46036, (46005), 46002, 46059
703	Korea	22189, 22106, 22190, 22105
741	North Sea - Mediterranean Sea	52134, 62116, 62104, 62122, 62119, 62120, 62154, 62123, 61001

2.2.2 Output Data

The output data will have to be given in NetCDF format, with individual files for each cycle and for each track within it. The following fields are necessary to participate in the Round Robin: time, latitude, longitude, Significant Wave Height, Quality Flag. All fields should be provided at a 20-Hz posting rate. Estimations of Backscatter coefficient and Range are encouraged, but not mandatory. The time record is to be exactly as provided on the original waveform product, as this may be used in collocating the coverage of different algorithms.

The algorithms that take part in the Round Robin are described in the Algorithm Theoretical Basis Document (ATBD). External participants were also requested to provide an analogous documentation.

2.2 Synthetic Aperture Radar

2.2.1 Input Data

The Round Robin evaluation focuses on a subset of Sentinel-1 Wave Mode data. For the comparisons with models and in-situ data, participants needed to process 1 full year of data. The analysis covers the period of January 2019 to December 2019 to provide a balanced mix of winter/summer conditions.

The following list contains 60 relative orbits numbers with regular coverage of NDBC buoys within ~50km collocation distance:

Orbits Pacific:

	relative orbit number	type	Sentinel-1 A/B		collocated NDBC buoy, S1A/S1B	wv1 wv2	collocation, km (A/B)
1	001	descending	-	B	46070 B	2	57
2	005	ascending	A	B	46005 A/B	1	52/9
3	006	ascending	-	B	46066 B	2	50
4	008	ascending	-	B	51209 B	2	29
5	014	descending	A	B	51004 A 46085 B	1 1	34 47
6	020	ascending	A	B	46085 A 46006 B	2 2	20 28
7	028	descending	A	B	46005 A/B	1	21/23
8	029	descending	-	B	46078 B 46066 B	1 2	22 49
9	030	descending	-	B	46071 B	2	4
10	035	ascending	A	B	46078 A/B	1	26/41
11	043	descending	A	B	46184 A/B	1	36/29
12	049	ascending	A	B	46004 A	2	6/47
13	050	ascending	-	B	46075 B	2	36

14	051	ascending	A	B	46075 A/B	2	3/3
15	058	descending	-	B	46001 B	1	20
16	059	descending	A	B	46073 B 51209 A	1 1	47 52
17	064	ascending	-	B	46001 B 46080 B	2 2	27 57
18	066	ascending	A	B	52201 A/B	1	54/54
19	067	descending	A	-	52211 A	2	38
20	078	ascending	A	B	46005 A/B 46036 A/B 46208 B 43010 B 46002 A/B	2 1 2 1 2	13/14 0/52 30 52 35/39
21	087	descending	A	-	46246 A	1	45
22	088	descending		B	46072 B	2	44
23	093	ascending	A	B	46001 A 46246 B	1 1	28 30
24	101	descending	A	B	46004 A/B	2	39/39
25	102	descending	-	B	46006 B	1	20
26	103	descending	A	B	46071 A/B	1	37/37
27	110	ascending	A	B	51209 A/B	1	5/46
28	112	descending	-	B	41040 B	2	57
29	116	descending	-	B	46085 B	2	9
30	122	ascending	A	B	46184 A/B 46006 B 46085 B	2 1 1	38/23 56 19
31	131	descending	-	B	51208 B 46078 B	1 2	35 25
32	132	descending	-	B	52201 B	2	16
33	137	ascending	-	B	46078 B	1	57

34	138	ascending	-	B	46035 B	1	40
35	139	ascending	A	B	52201 A/B	2	17/24
	140	ascending	-	B	52211 B	2	57
36	147	descending	-	B	46070 B	1	46
37	151	ascending	-	B	46036 B 46205 B	2 2	40 51
38	153	ascending	-	B	46070 B	2	18
39	160	descending	-	B	46001 B	1	33
40	161	descending	A	B	46072 B 51209 A/B	1 2	37 50/25
41	174	descending	A	-	46036 A	1	21

Orbits Atlantic:

	relative orbit number	type	Sentinel-1 A/B		collocated NDBC buoy, S1A/S1B	wv1 wv2	collocation, km (A/B)
42	002	ascending	-	B	44139 B	1	40
43	018	ascending	-	B	44011 B 44150 B	1 2	35 43
44	025	descending	A	B	41049 A 41052 B 41049 B 41056 B	2 2 2 2	27 35 57 20
45	031	ascending	A	B	41041 A/B	1	39/39
46	033	ascending	-	B	44095 B 41001 B 41047 B 44099 B 41025 B	2 1 1 2 2	32 39 32 33 54
47	040	descending	-	B	41047 B	1	36
48	047	ascending	-	B	44137 B	1	27

49	054	descending	-	B	41044 B	1	30
50	062	ascending	-	B	44008 B 41048 B 42085 B	2 1 2	49 29 23 (near land)
51	068	descending	-	B	41041 B	2	44
52	076	ascending	-	B	44139 B	2	40
53	091	ascending	A	B	44011 A/B 41049 A/B	2 2	41/- 42/21
54	104	ascending	A	B	41041 A/B	2	47/39
55	113	descending	-	B	41002 B 41001 B 44066 B 44005 B	2 1 2 1	50 28 35 45
56	120	ascending	-	B	44137 B	2	22
57	135	ascending	-	B	41046 B 41047 B 41001 B 41115 B	2 1 1 2	0 46 50 10 (near land)
58	148	ascending	A	-	3100053 A	2	40 *EMONNET
59	156	descending	A	B	41044 A/B	2	30/30
60	164	ascending	-	B	41052 B	1	45

As reference data two separate datasets are used:

1. In-situ measurements (ground truth) from buoys
2. Model measurements (ground truth) from MFWAM distributed by CMEMS

For some orbits buoy collocation from both Sentinel-1 platforms A and B are possible. This means from the list of 60 relative orbit numbers 87 WV products are collocatable with buoys. For buoy validation about 1.100 vignettes for each wv1 and wv2 were expected, resulting in ~2.200 data samples in the buoy validation dataset. For model validation only vignettes within $-60^\circ < \text{latitude} < 60^\circ$ are considered to avoid ice coverage. For all 87 WV products about 100.000 vignettes for each wv1 and wv2 were expected, resulting in ~200.000 data samples in the model validation dataset. Each reference data sample must be temporally interpolated

to the image acquisition time between the two neighboring measurements not further apart than six hours.

2.2.2 Output Data

The output data are given in NetCDF format, with individual files for each orbit. The following fields were necessary to participate in the Round Robin: time, latitude, longitude, Significant Wave Height, Quality Flag. Estimations of Significant Wave Length and Significant Wave Direction was encouraged, but not mandatory. The time record was to be exactly as provided on the original waveform product, to co-locate the coverage of different algorithms. The algorithms that take part in the Round Robin are fully described in the Algorithm Theoretical Basis Document (ATBD). External participants were requested to provide an analogous documentation.

2.3 Round Robin datasets

2.3.1 Altimeter (LRM and DD)

In-situ dataset

As part of its wave forecast verification activities within the Joint Commission for Oceanography and Marine Meteorology, ECMWF gathers wave observations. Most of the data are obtained from the data received via the Global Telecommunication System (GTS), where atmospheric and oceanographic data are disseminated to weather forecasting institutions. A few data sets are also supplied directly to ECMWF. Most data are from a wide range of moored buoys, except for data from the North Sea, the Norwegian Sea and the Gulf of Mexico. A basic quality control procedure is performed following Bidlot et al. (2002) and the resulting hourly time series have been made available to the project.

Wave Model dataset

A long global wave model hindcast has been produced using the latest version of ECMWF wave model (CY46R1, ECMWF 2019). The model spatial resolution is 0.125x0.125 degrees and the output is hourly, making it ideal to be collocated with altimeter data passes. The necessary hourly wind forcing and sea ice cover information come from ECMWF latest reanalysis (ERA5). The hindcast covers the period from 1979 to present. These data are preferred to using directly ERA5 output instead, because, ERA5 wave model data are on a coarser resolution (0.36 degree), are based on an older version of the wave model code and most of all, ERA5 has used altimeter data from (ERS-1, ERS-2, ENVISAT, Jason-2, Jason-2), rendering it not independent. The latest hindcast has been found to be even better than ERA5 (Jean Bidlot personal communication). Over the course of 2019, it will become a product made available alongside ERA5 data.

2.3.2 Synthetic Aperture Radar

In-situ measurements (ground truth) from buoys

The 50 km collocation distance defines the distance from the nearest border or corner of a 20km×20km S1-WV vignette to the buoy location, this means ~60 km distance from vignette centre.

For in-situ measurements 54 buoys providing wave heights were selected:

- 49 buoys from NOAA buoy stations network NDBC (National Data Buoy Center)
- 8 buoys from ECCC (Environment and Climate Change Canada)
- 1 buoy from EMODnet

The stationary NOAA buoys typically provide hourly measurements and the ECCC buoys store the data more unsystematically. For validation the data of each buoy are temporally interpolated. The data with a measurement time gap over 6h were excluded.

The following table lists all buoys with their corresponding geographic coordinate:

Name	Lat	Lon	Data source	Group
43010	10.051	-125.032	NDBC	Group-1 Alaska, Canada
46001	56.304	-147.920	NDBC	
46002	42.612	-130.537	NDBC	
46004	50.930	-136.100	ECCC	
46005	46.140	-131.070	NDBC	
46006	40.782	-137.397	NDBC	
46035	57.026	-177.738	NDBC	
46036	48.350	-133.940	ECCC	
46066	52.785	-155.047	NDBC	
46070	55.082	175.153	NDBC	
46071	51.125	179.012	NDBC	
46072	51.672	-172.088	NDBC	
46073	55.031	-172.001	NDBC	
46075	53.983	-162.041	NDBC	
46078	55.556	-152.582	NDBC	
46080	57.947	-150.042	NDBC	
46085	55.868	-142.492	NDBC	
46184	53.910	-138.850	ECCC	
46205	54.190	-134.320	ECCC	
46208	52.520	-132.690	ECCC	
46246	50.033	-145.200	NDBC	

51209	-14.264	170.493	NDBC	Group-2 South-West Pacific
52201	7.083	171.392	NDBC	
52211	15.268	145.662	NDBC	
51001	24.453	-162.000	NDBC	Group-3 Pacific Hawaii
51002	17.037	-157.696	NDBC	
51004	17.604	-152.364	NDBC	
51101	24.361	-162.075	NDBC	
51205	21.018	-156.425	NDBC	
51202	17.037	-157.696	NDBC	
51208	22.285	-159.574	NDBC	
41001	34.502	-72.522	NDBC	
41002	31.892	-74.930	NDBC	
41025	35.006	-75.402	NDBC	
41040	14.554	-53.045	NDBC	
41041	14.441	-46.033	NDBC	
41043	21.124	-64.830	NDBC	
41044	21.582	-58.630	NDBC	
41046	23.822	-68.384	NDBC	
41047	27.514	-71.494	NDBC	
41048	31.838	-69.585	NDBC	
41049	27.490	-62.938	NDBC	
44005	43.201	-69.128	NDBC	
44008	40.504	-69.248	NDBC	
44011	41.070	-66.588	NDBC	
44014	36.606	-74.840	NDBC	
44066	39.618	-72.644	NDBC	
44095	35.750	-75.330	NDBC	

44099	36.914	-75.720	NDBC	
44137	42.260	-62.000	ECCC	
44139	44.240	-57.100	ECCC	
44150	42.500	-64.020	ECCC	
41052	18.249	-64.763	NDBC	Group-4 North Atlantic, Caribbean Sea
41056	18.261	-65.464	NDBC	
41115	18.376	-67.280	NDBC	
42085	17.869	-66.532	NDBC	
3100053	-23.478	-43.984	EMONNET	p

Model measurements (ground truth) from MFWAM distributed by CMEMS

For model “ground truth” MFWAM model data from CMEMS (COPERNICUS) with spatial resolution of 1/12 degree will be used. The results are stored in 3h time intervals and are therefore temporally interpolated.

http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=GLOBAL_ANALYSIS_FORECAST_WAV_001_027

3. Round Robin procedures

3.1 Altimeter (LRM and DD)

Here we report the statistics chosen to evaluate the Round Robin. The evaluation of the Round Robin is based on tables reporting the key statistics for each of the participants.

The statistics (unless differently specified) are reported separately for each of the following categories:

GEOGRAPHICAL CATEGORIES

- Full Dataset: all data without any distinction
- Coastal Data: all data in which DtC < 20 Km, < 10 Km, < 5 Km
- Open Ocean Data: all data in which DtC > 20 Km

SEA STATE CATEGORIES

- Low Sea States: all data in which $0\text{m} < \text{SWH} < 2\text{m}$
- Average Sea States: all data in which $2\text{m} < \text{SWH} < 5\text{m}$
- High Sea States: all data in which $\text{SWH} > 5\text{m}$
- Very High Sea States: all data in which $\text{SWH} > 10\text{m}$

3.1.1 Outlier Analysis

For detecting outliers, the following three criteria are defined:

- **invalid**: Data missing (already set to NaN) or quality flag set to 'bad' (1).
- **out_of_range**: If a value is out of the expected range of [0.25, 25] m. (Note noisy estimations may sometimes return sub-zero values.)
- **mad_factor**: This criterion compares the value with its 20 closest neighbours (10 before and 10 after). It is implemented using median and median absolute deviation (MAD), which are statistically robust measures. Data are discarded if they exceed median plus $3 * 1.4826 * \text{MAD}$, with median and MAD calculated on 20-point sliding windows, and the factor 1.4826 converts the MAD to SD equivalent for a normal distribution.

If one of the criteria is applicable, the individual measurement is considered to be an outlier. The evaluated metrics are the total number of outliers and the number of outliers in the coastal zone as a function of the dist2coast, considering bands of less than 20, 10, and 5 km from the coast. This analysis is performed on the 20 Hz-data, and the total number of outliers is given by the sum of the individual outlier types excluding potential overlapping indices (one measurement might be marked by multiple types, e.g., mad_factor and out_of_range). Furthermore, as an exception of the outlier analysis, there is no sea-ice flag considered, when reading the netCDF files. The sea-ice flag would be considered as another kind of quality flag, marking a measurement as invalid. The outlier analysis should be on estimations of SWH in the ocean and do not take into account points on surfaces, for which such estimation does not exist, such as land and sea-ice covered areas.

3.1.2 High Frequency Noise Analysis

The noise is computed as the standard deviation of the 20-Hz SWH estimations within a 1-Hz block. For each category, the statistics is obtained by computing the median of all the noise estimations.

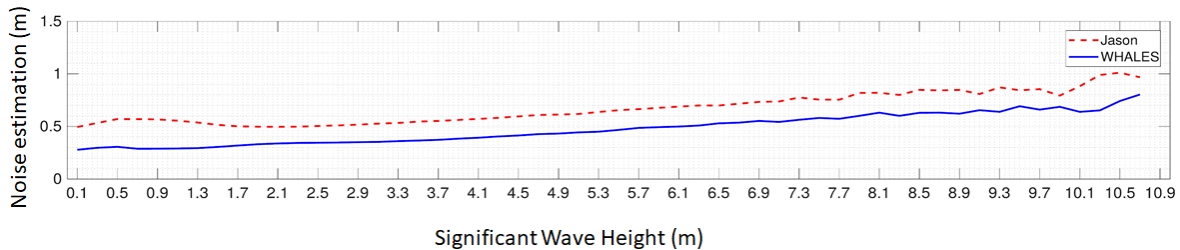


Figure 1: Visual example of noise analysis for varying SWH performed on a test area with Jason-3 data and an algorithm under development for the SS_cci project.

3.1.3 Comparison with in-situ Data

Data from wave buoys at the time of each satellite pass are compared with the satellite observations at the closest point, as defined in section 1.3. For a meaningful altimeter average we compute the mean of the observation at the buoy location and of 25 points either side of this. This implies the mean is calculated over 51 points (~17 km along track) and although some of these points will be over land they should have no impact because all land or 'bad' data are excluded. Statistics are separated according to the distance from coast and sea state (< 2m ; 2-5m ; > 5m). The following statistics are calculated, using all buoys and all passes:

- Number of points used in comparison
- Slope and Intercept of best fit line
- Median bias
- Standard Deviation of Differences
- Correlation

Of these the latter three were the ones used in intercomparing the performance of the various candidate algorithms (see also Schlembach et al., 2020).

3.1.4 Comparison with Model Data

Model grid points and altimetry will be coupled by considering the median of the SWH 20-Hz measurements from altimetry within the grid point. The following statistics are provided:

- Correlation
- Standard Deviation of the difference between SWH from altimetry and SWH from model
- Slope of the linear fit (regression line of SWH from altimetry -vs- SWH from model scatter

plot)

- Median bias between SWH from altimetry and SWH from model.

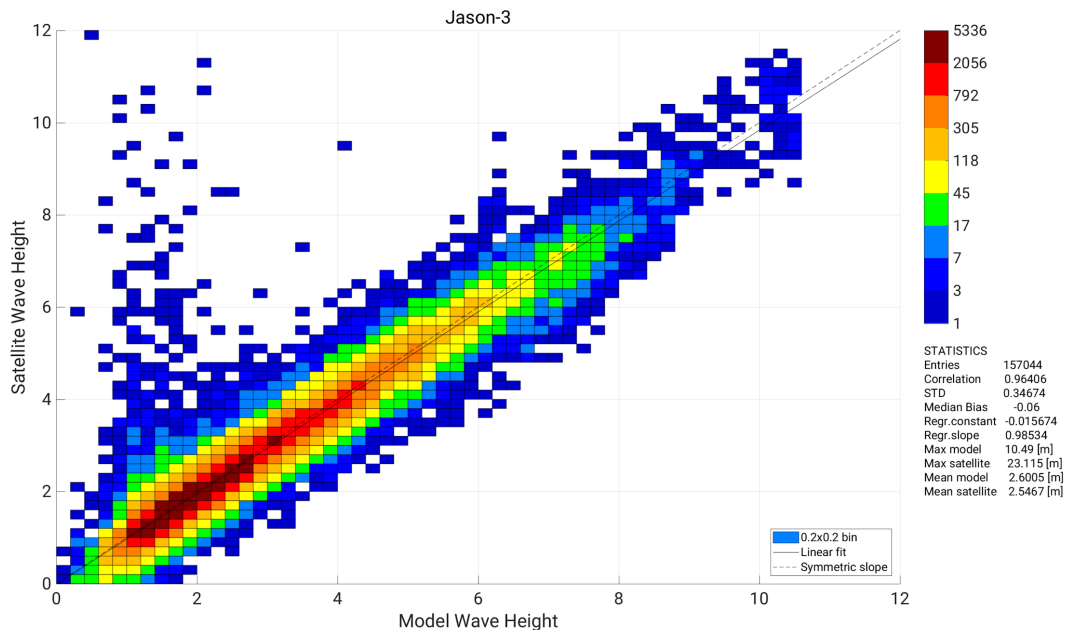


Figure 2: A density plot in a test area illustrating key statistics in the comparison between Jason-3 data and the Météo-France global wave model (MFWAM) available from the Copernicus Marine Service.

3.1.5 Representation of Scales of Variability

Spectra were calculated for SWH in the same way they are used for Sea Surface Height (SSH) to compare the performance of different retracers. We used FFT applied to the 20 Hz SWH data, using segments of at least 1024 points (~330 km), with Hann weighting applied. Provided the retracker algorithms treat each waveform independently, and there is no along-track smoothing applied, the spectra at very short scales should be white noise, with a plateau level consistent with the value for "high-frequency noise" (section 3.1.2). At large scales, the spectra should be dominated by the real geographical variations in SWH, and providing all estimators are unbiased their spectra will converge.

The main assessment considered in this section is the nature of the spectra in the ranges 25-50 km and 50-100km, as these are the scales requested by GCOS, and are also the scales of models that might assimilate altimeter SWH data. Spectra were assessed at these scales for all retracers, with the results partitioned according to median wave height, in order to ascertain whether there are problems associated with differently shaped waveforms (i.e. leading edge slope).

3.2 Synthetic Aperture Radar

The statistics chosen to evaluate the SAR Round Robin are similar to the Altimetry Round Robin. As the amount of data samples with sea state above 12m is expected low, in contrast to the Altimeter Round Robin high and very high sea state is comprised into one category.

Further, a category for rough sea state has been added. This means the statistics for the SAR Round Robin are categorized as follows:

GEOGRAPHICAL CATEGORIES

- Full Dataset: all data without any distinction, as the amount of ground-truth data samples compared to altimetry is less due to non continuous acquisitions.
- $-60^{\circ} < \text{LAT} < 60^{\circ}$ to avoid ice coverage

SEA STATE CATEGORIES

- Low Sea States: all data in which $0\text{m} < \text{SWH} < 1.5\text{m}$
- Moderate Sea States: all data in which $1.5\text{m} < \text{SWH} < 3\text{m}$
- Rough Sea State: all data in which $3\text{m} < \text{SWH} < 6\text{m}$
- High and Very High Sea States: all data in which $\text{SWH} > 6\text{m}$

NB: In the first Round Robin for Synthetic aperture radar, only significant wave height is evaluated since only one algorithm is providing wave period.

3.2.1 Outlier Analysis

For the definition of Outliers please see the previous sections 1.3. The percentage of Outliers is defined for each of the listed categories. For each sea state bin, the outliers are the values with

$$\text{ABS}(\text{SWH}_{S1} - \text{SWH}_{\text{ground-truth}}) > 3 * \text{RMSE}_{\text{bin}}$$

As the different RMSE_{bin} leads to different number of outliers, in order to get the comparable results, RMSE_{bin} for outliers' definition was taken from Stopa results (see Tab.X2).

3.2.3 Comparison with In-situ Data

The collocation of in-situ buoy data with SAR measurements is only valid, when the distance between the respective buoys and the SAR subscene are at most 50 km apart from each other. The in-situ buoy data and SAR are coupled by comparing the median of the SWH and parameters measurements from SAR with the buoy's SWH. The Statistics are divided according to the categories previously defined. The following statistics are provided:

- Root Mean Square Error (RMSE)
- No-sea-state percentage, i.e. proportion of bad values of SAR vignettes for which no sea state estimation is possible
- Outliers percentage: proportion of SAR derived parameters that differ from in-situ data by more than 3 times the local RMSE for each sea state domain.
- Median bias between SWH from SAR and from buoy data.

3.2.4 Comparison with Model Data

The same MFWAM model distributed by CMEMS used for altimetry dataset, is used for SAR comparison. Model grid points and SAR are coupled by considering the median of the SWH parameter measurements from SAR within the grid point. The Statistics are divided according to the categories previously defined. The following statistics are provided:

- Root Mean Square Error (RMSE)
- No-sea-state percentage, i.e. proportion of bad values of SAR vignettes for which no sea state estimation is possible
- Outliers percentage: proportion of SAR derived parameters that differ from model data by more than 3 times the local RMSE for each sea state domain.
- Median bias between SWH from SAR and from model.

4. Round Robin results

Running the validation with retrackval produces a large set of figures that can not all be presented and discussed in this report. Instead, a small subset of the figures will be taken into account in the following sections, in order to provide a comprehensive analysis of the results.

The following table provides the list of all the algorithms evaluated in the exercise. For a clarification of the “denoising”, see section 4.2 of this report. Note that the algorithms from the following authors are internal to the project and described in the ATDB document: TUM, PML/TUM, CLS/CNES and isardSAT.

	Retracking Algorithms	Altimeter Mode	Author	Denoised
J3	MLE-3 (reference)	LRM	-	No
	MLE-4 (reference)	LRM	-	No
	WHALES	LRM	TUM	No
	WHALES_adj	LRM	PML/TUM	Yes
	WHALES_realPTR	LRM	PML/TUM	No
	WHALES_realPTR_adj	LRM	PML/TUM	Yes
	Brown-Peaky	LRM	UON	No
	TALES	LRM	UniBonn	No
	Adaptive	LRM	CLS/CNES	No
	Adaptive_HFA	LRM	CLS/CNES	Yes
	STARv2	LRM	UniBonn	Yes (inherently)
Total Number				11
S3A	SAMOSA (reference)	DDA	SAMOSA project [6]	No
	WHALES-SAR	DDA	TUM	No
	DeDop-Waver	DDA	isardSAT	No
	LR-RMC	DDA	CLS/CNES	No
	LR-RMC_HFA	DDA	CLS/CNES	Yes
	MLE-4-PLRM (reference)	PLRM	-	No
	TALES-PLRM	PLRM	UniBonn	No
	STARv2-PLRM	PLRM	UniBonn	Yes (inherently)

4.1 Outlier Analysis

Figure 3 shows the number of total outliers of J3 vs. S3A as a function of dist2coast. For open-ocean scenarios, Brown-Peaky (J3) and WHALES-SAR (S3A) have the least amount of total outliers, accounting for 8.1% and 15.8%. The outliers of all retrackerers stay below 20%, with the exception of STARv2-PLRM, which amounts to 27%. The J3 retrackerers tend to be less prone to outliers, but since different datasets (due to its nature to have different orbits) were used, this hypothesis needs to be considered with care. Among the new retrackerers, J3 datasets contain less outliers. This is likely to be a consequence of a more conservative use of the qual-flag, since in the standard products (MLE-3 for J3 and SAMOSA for S3A) the opposite is observed. When approaching coast, the number of outliers is significantly increased for both LRM and DDA retrackerers. The number of outliers range from 27.9% to around 50% (LRM) and from 42.5% to around 60% (DDA) for the best performing retrackerers (approaching coast in the intervals 20 km, 10 km, and 5 km), which are Brown-Peaky and TALES for J3 and SAMOSA, WHALES-SAR, and TALES-PLRM for S3A. It appears that the number of outliers quasi-linearly increases with a decreasing dist2coast. The differences of the amount of total outliers can be quite large, e.g. when considering areas that are very close to the coast (less than 5 km), SAMOSA is able to produce estimates for almost 50% of the measurements, whereas Adaptive retrieves only 16.5% of valid SWH samples, which is very few.

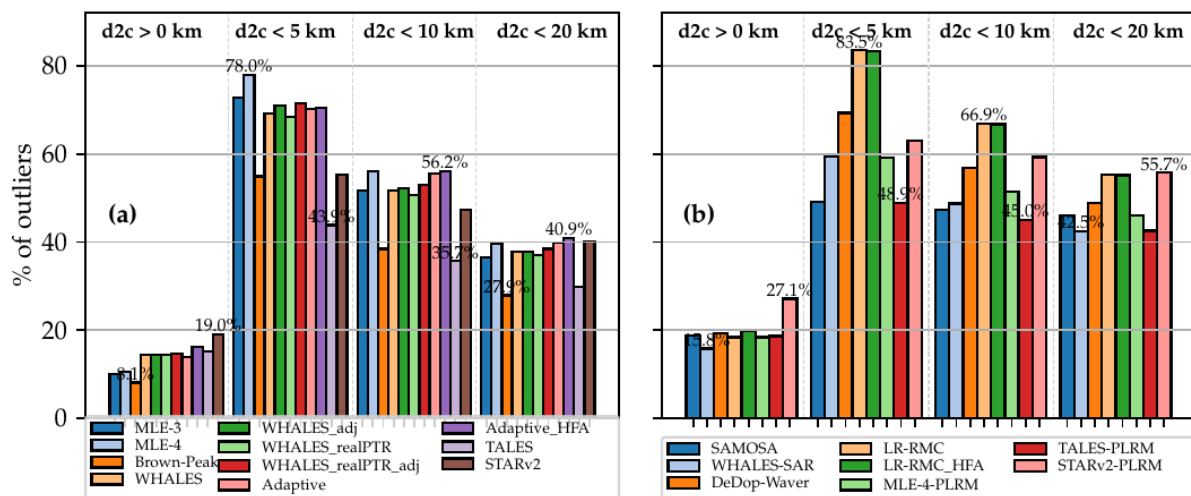


Figure 3: The total number of outliers as a function of dist2coast for (a) J3 and (b) S3A.

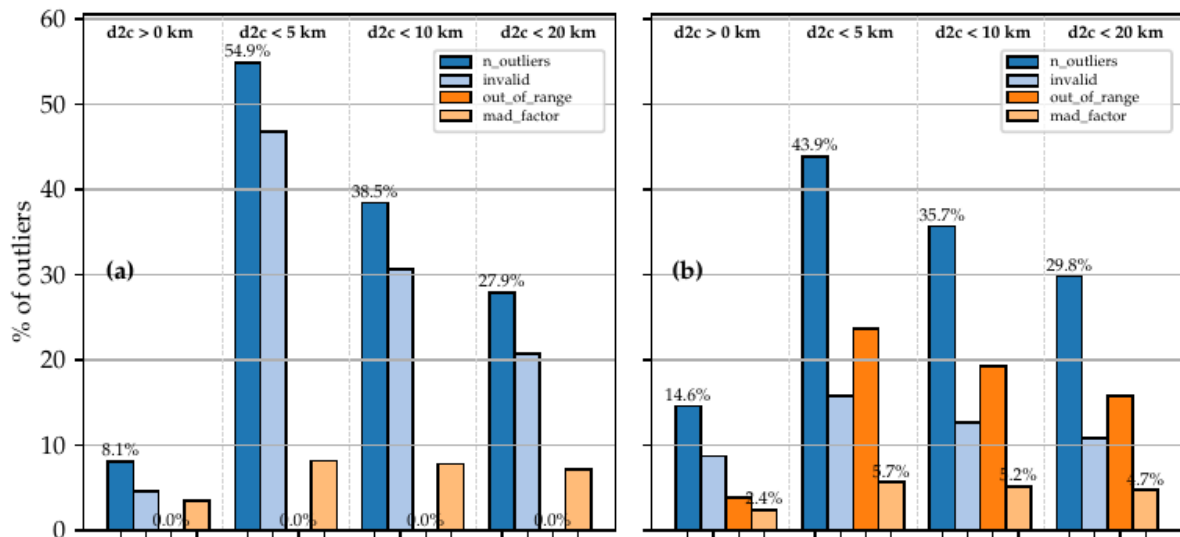


Figure 4: A comparison of outlier types as a function of dist2coast for the retrackerers (a) Brown-Peaky and (b) TALES.

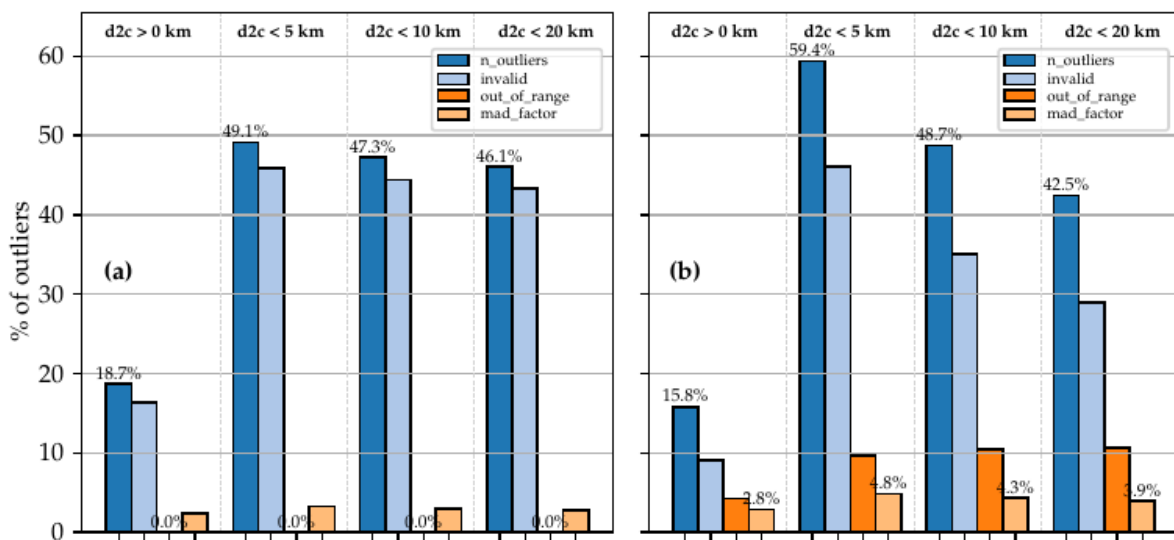


Figure 5: A comparison of outlier types as a function of dist2coast for the retrackerers (a) SAMOSA and (b) WHALES-SAR.

Figure 4 depicts the distribution of the outliers types invalid, out_of_range, and mad_factor as a function of dist2coast for the retrackerers Brown-Peaky and TALES. They both follow a subwaveform approach that discards the trailing edge of the waveform, which is mostly contaminated by spurious signals in the coastal zone. One might thus expect a similar outlier behaviour, but their number of outliers differ. TALES exhibits a significant amount of about 15-23 % of out_of_range outliers, whereas Brown-Peaky shows none but instead an increased amount of invalid estimates (21 % vs. 11 % for dist2coast < 20 km). This underlines the role of qual-flag: It is up to the strategy of the individual retracker, whether to decide if an estimate is set to be bad or remained as a potential outlier (out_of_range or mad_factor). Interestingly, in general, the fraction of mad_factor outliers is increased only slightly by a factor of about two when approaching coast, whereas the total amount of outliers increases significantly, meaning that the mad_factor does only weakly depend upon

the dist2coast. In contrast, the estimate is either good or very bad or missing in the coastal zone, yielding the measurement to be an outlier of type invalid or out_of_range. In Figure 5, the outlier types of the two DDA retrackerers SAMOSA and WHALES-SAR are shown. Both exhibit a low amount of total outliers, as shown in Figure 3. SAMOSA's major fraction of outliers mostly accounts for invalid estimates, no out_of_range values, and only few mad_factor-type outliers. This signifies that it is capable of identify correctly, which values might be reliable estimates. The total amount remains relatively constant in the coastal zone with about 48 %, even when further decreasing the dist2coast. In comparison, as shown in Figure 5-(b), WHALES-SAR (which still exhibits one of the best outlier characteristic of the investigated retrackerers) sets some SWH estimates to out_of_range values. This might arise from the fact that only a subwaveform is considered. Likewise as with SAMOSA, the number of invalid points make up the majority of outliers and the number of mad_factor samples remain relatively constant between about 3-5 %.

To conclude the outlier analysis, one can state the following points:

- The number of outliers is significantly increased in the coastal zone and increases further when approaching coast.
- In open-ocean, the number of total outliers amounts to less than 20 %.
- Most of the retrackerers' outlier types are invalid samples, which originate from measurements, whose qual-flag is poorly defined.

with the following retracker showing the least amount of outliers:

- J3: Brown-Peaky, TALES
- S3A: SAMOSA, WHALES-SAR
- S3A-PLRM: TALES-PLRM

4.2 Noise Analysis

In this section the intrinsic noise of the retracked datasets is evaluated. As described in RR definition document, the noise is defined as the SD of a 20-Hz measurement of the along-track series.

An important fact to consider is that some of the investigated retracked datasets already have a denoising technique applied:

- J3: WHALES_adj, WHALES_realPTR_adj, Adaptive_HFA, and STARv2
- S3A: LR-RMC_HFA, STARv2-PLRM

These datasets exhibit a reduced noise performance and need to be evaluated separately. It also has to be noted that some of the denoising techniques can be applied independently a-posteriori after the SWH estimates (L2 data) were retrieved so they can be applied to arbitrary retrackerers. Other retrackerers such as the STARv2/STARv2-PLRM algorithm have an inherent denoising implied.

Figure 6 depicts the median noise values of (a) J3 and (b) S3A classified by the area of

interest: overall, coast, and open-ocean. For the J3/LRM-retrackers without denoising, Adaptive and WHALES have the least and second least median noise values of about 0.23 m. While also including denoised datasets into consideration, Adaptive_HFA has the lowest median noise values with about 0.12 m, followed by STARv2 with about 0.18 m. For the S3A algorithms without denoising, DeDop-Waver shows the lowest median noise values with about 0.32 m. Among the denoised algorithms, STARv2-PLRM has the least median noise values with a slight increase for the coastal scenario from 0.17 m to 0.25 m. This demonstrates the effectiveness of the used denoising techniques.

When analysing the dependence of the median noise values for open-ocean and coastal scenarios, one can notice a slight increase of noise, which is more or less pronounced on the individual algorithms. For instance, STARv2's value is increased from about 0.18 m to about 0.24 m, whereas Adaptive_HFA's value is increased by just 0.01 m. In conclusion, it can be stated that there is only a minor dependence on the dist2coast.

Figure 7 depicts the noise as a function of SWH and the different sea states. The plots demonstrate the strong dependence of the sea state. The results are in accordance with the ones shown in Figure 6.

For LRM, Adaptive exhibits the best noise performance for all sea states (no denoising applied). With denoising applied, Adaptive_HFA wins for low and average sea states, whereas STARv2 outperforms Adaptive_HFA for high and very high conditions.

With respect to S3A and low sea states, the noise levels of WHALES-SAR, DeDop-Waver, LR-RMC_HFA (denoised), and STARv2 (denoised) are at a similar level. For average, high and very high sea states, STARv2-PLRM shows significantly low noise values. This might be explained by the nature of the STARv2-PLRM algorithm, for which neighbouring SWH estimates are taken into account for reducing abrupt changes in the estimates and thus reducing the SD of the 20-Hz measurements (the same applies to the LRM version of STARv2). Thereafter comes the LR-RMC algorithm as the second best of the S3A retrackers at average, high, and very high sea states.

For very low SWHs of less than 1 m, one can observe an increased noise level. This is due to the fact that sea states with very low wave heights induce a waveform with a very steep slope of the leading edge, which thus is undersampled. Smith and Scharroo (2015) has investigated this issue and suggested a simple zero-padding, with which this effect can be mitigated. Comparing the noise level of the standard retrackers MLE-4 and SAMOSA (LRM vs. DDA) for low and average sea states, one can conclude that the performance is improved significantly, which is in accordance to Gommenginger et al. (2013).

Furthermore, it can be stated that most of the novel retrackers show significant improvements across all sea states, as compared with the standard retrackers MLE-4 and SAMOSA. This is particularly pronounced for high and very high sea states, for which both retrackers show at least twice as much noise level as compared to the novel approaches.

When comparing the absolute noise levels evaluated here with those mentioned in the literature, one can observe that they differ from each other. For instance, Fenoglio-Marc et al. (2015) has conducted a study in the German Bight and estimated the noise values of 6.7 cm and 13 cm (for SWH values around 2 m) for DDA and PLRM (considering open-ocean measurements with dist2coast \geq 10 km), respectively. In ESA CP40 project report (2015),

noise values were retrieved to be 8.5 cm and 11.09 cm for DDA and LRM for open-ocean scenarios across all sea states. Arduin et al. (2019) has compared the noise for a full cycle of the missions J3, S3B-LRM, and S3B-SARM, and estimated them to be 0.50 cm, 0.47 cm and 0.38 m for SWH values of around 2 m, respectively. From this, it can be concluded, that the intrinsic noise performance strongly depends upon the region of interest, the sea state, and whether the coastal zone is included or not in the considerations. With the values evaluated in this RR exercise, ranging from 0.12 m to 0.70 m, the findings are in accordance to the literature.

To sum up the noise analysis, the following can be stated:

- The intrinsic noise shows only a minor of the dist2coast and strong dependence on the sea state.
- The noise is reduced for most of novel retracking algorithms.
- DDA retrackers show a better noise performance than their adapted PLRM counterpart.
- The following retrackers exhibit the best intrinsic noise characteristic:
 - J3: First: Adaptive, Second: STARv2 (although inherently denoised).
 - S3A: First: STARv2-PLRM (although inherently denoised), Second: LR-RMC.

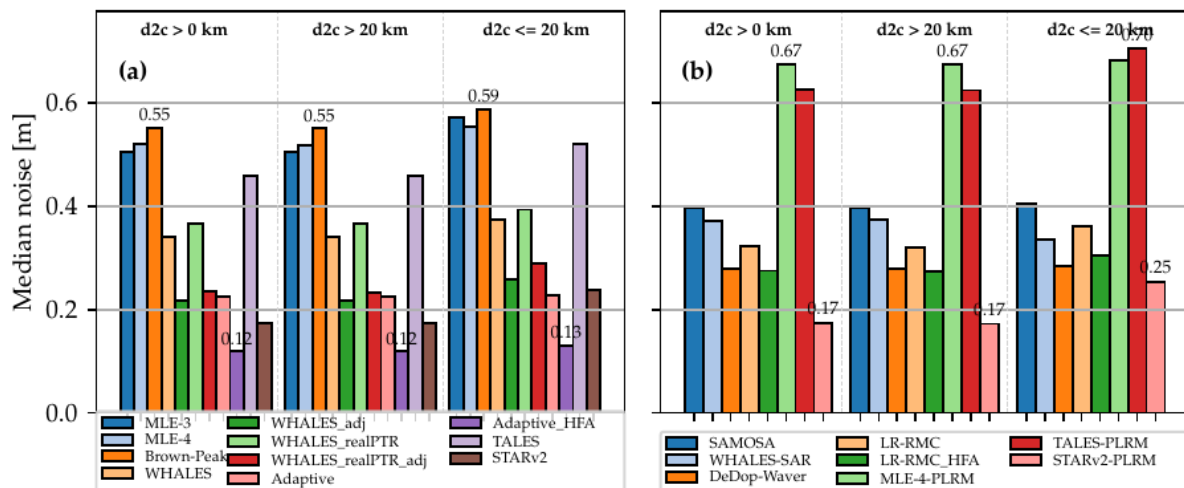


Figure 6: Median noise as a function of dist2coast for (a) J3- and (b) S3A-retracking algorithms.

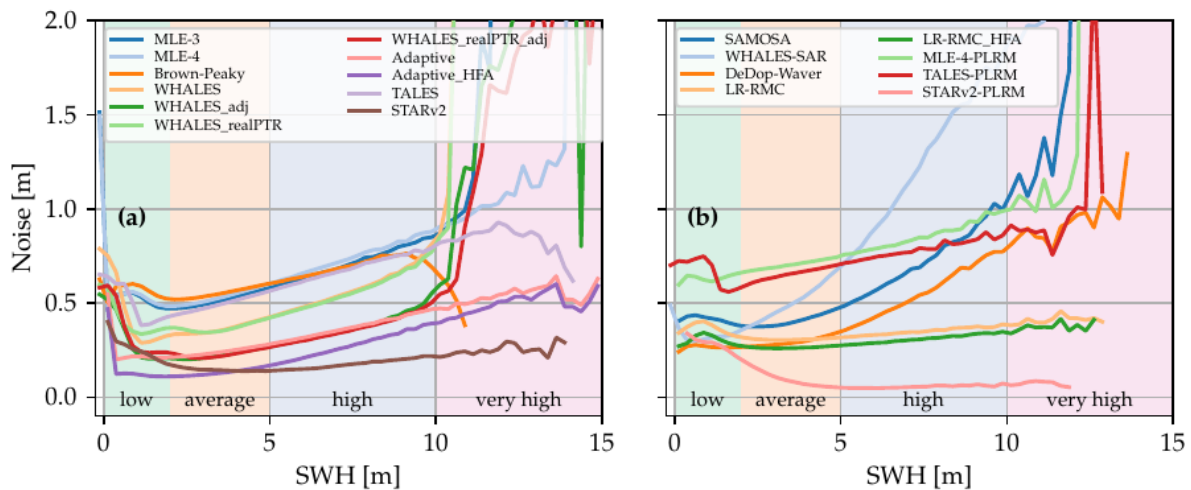


Figure 7: Noise level of the individual retrackerers as a function of SWH for (a) J3- and (b) S3A-retracking algorithms with the sea state noted at the bottom.

4.3 Wave Spectral Variability Analysis

Spectra were calculated according to the Welch method described earlier (1024-pt FFT, with Hann weighting and overlapping of data spans by 50% to maximise the number of effectively independent spectra), with order 50000 spectra being calculated for each retracker (across all wave height conditions). The results are shown in Figures 9 and 10.

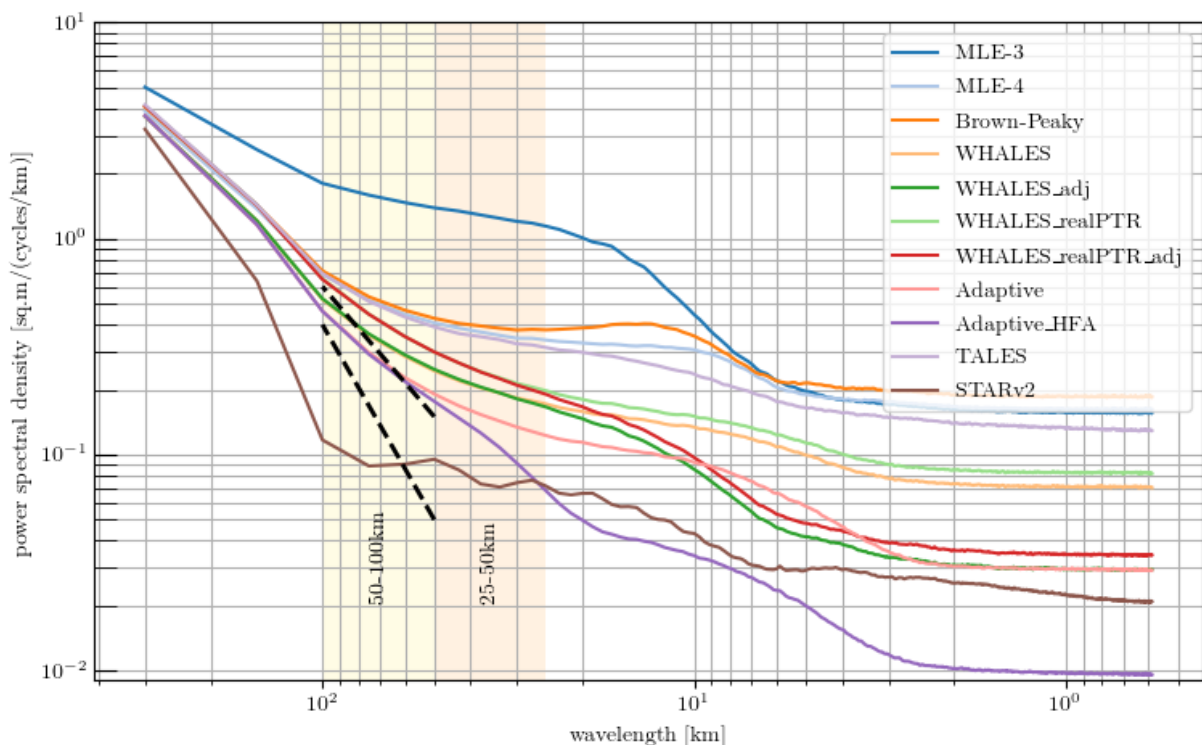


Figure 8: Spectra of SWH derived from all files provided for J3

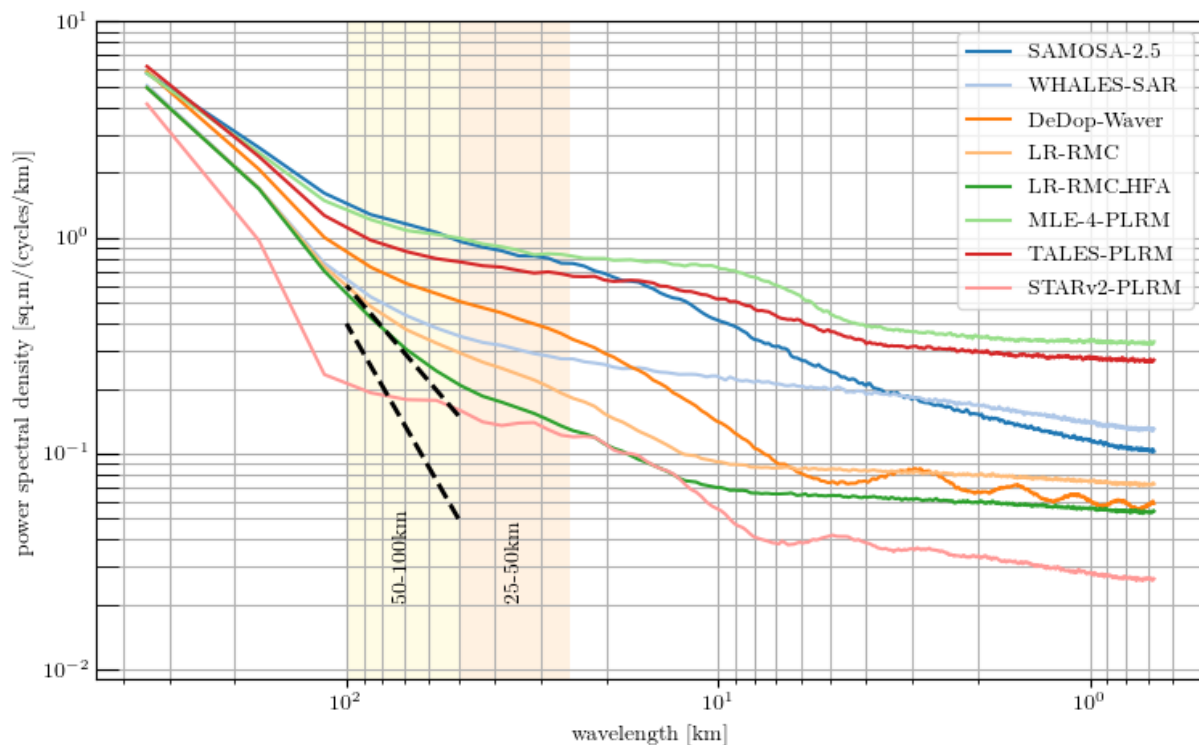


Figure 9: Spectra of SWH derived from all files provided for S3A.

The MLE-4 algorithm has been used for all LRM altimeter missions in the last decade, being provided within the standard data products from the agencies, and thus it serves as a useful reference with which to compare the others. At the high wavenumber (short wavelength) end it is dominated by white noise, with a constant level being shown for scales under ~ 3 km. At the long wavelength range it should represent the real scales of variability of the geophysical signal. The spectral slope in this region indicates the cascade of energy from larger scales to smaller ones and may vary with the environment. Inbetween (in the range 7-30 km) there is a "spectral hump" (Dibarboure et al., 2014) that was first noticed in spectra of Sea Level Anomaly but could also be noted for σ_0 (Quarty, 2009). This relates to the size of the instrument footprint, as small features directly on the satellite track will affect all returns in a similar manner whilst the satellite traverses that distance. Dibarboure et al. (2014) were able to show that such effects (for SLA) were minor for Cryosat-2 as it has a much reduced footprint.

4.3.1 Jason-3

For Jason-3 (Figure 8a), the other default algorithm (MLE3) has even higher spectral levels than MLE-4 due to the existence of other waveform characteristics (e.g. skewness or mispointing) that are leaking into the SWH estimates. In the absence of anomalous waveforms, the Brown-Peaky algorithm conforms closely to the MLE-4 and thus its spectra is similar. A number of algorithms focus primarily on the information within the leading edge, neglecting the returns from the wider irradiated disk on the sea surface, and thus in effect they utilise a narrower footprint. This is borne out by the spectra of their products having a much weaker or non-existent spectral hump.

Secondly, a number of algorithms differ in the level of the plateau of white noise at the smallest scales. This relates to the precision i.e. the measure of short-scale variability detailed earlier, with the algorithms that employ an adjustment to overcome the effects of correlated noise i.e. those labelled 'adj' or 'HFA' showing a clear improvement over their parent product. The effect for 'Adaptive_HFA' is over a wider range of scales because of how it was implemented. It is noted that the spectrum for STARv2 does not flatten off completely, probably due to the intrinsic method it uses for determining a SWH profile for each track.

4.3.2 Sentinel-3A

For the spectra from Sentinel-3A (Figure 8b) many similar features are noted. The shapes of the spectra for MLE-4, TALES and STARv2 applied to the PLRM data are similar to their LRM equivalents, but with the plateau level at the short wavelength end differing on account of the different statistics for the measurement noise due to fading. In the long wavelength regime the spectra are on a par with those for Jason-3. Somewhat surprisingly the DDA spectra show some energy in the region of the “spectral hump”, and their curves do not flatten off completely at very short wavelengths.

4.3.3 Overview of Spectral Analysis

The main purpose of the spectral evaluation was to ascertain the variability at the scales of 25-50 km and 50-100 km (highlighted in Figure 8) that were of interest to the GCOS community. For this measure the best LRM/PLRM performance is achieved by STARv2 and then Adaptive_HFA, with LR-RMC_HFA having the lowest power levels of the spectra of DDA retracers. However, there is some concern that these products may have applied excessive smoothing removing some of the power levels associated with the real geophysical signals, but there is no recognised agreement on what the true spectra of SWH should be like.

4.4 Comparison against Wave Models

The statistics of the comparison of the 1-Hz retracked data against the ERA5-based hindcast (ERA5-h) wave model, which does not assimilate any satellite altimetry data, are shown in Figure 10 and Figure 11 for J3 and S3A as a function of dist2coast and the sea state on the left and right columns, respectively. The three metrics correlation, median bias, and SDD are presented and discussed in this section.

It needs to be emphasised at this point of the analysis that the comparison against the wave model is limited to the resolution of the ERA5-h wave model (which is 18 km). Since the posting rate of the SWH series and the model are reduced to 1-Hz, potential high-frequency variations of the SWH series might thus be masked or some retracers that inherently smooth the SWH series might benefit from this type of analysis (e.g. STARv2 or the _HFA variants). In consequence, this means that if a retracking algorithm such as STARv2 is strongly filtering an SWH series, it might show an excellent correlation against the wave

model and a low SDD, which is shown in the following subsections. However, at the same time, the smoothed SWH series lacks a significant amount of energy.

Moreover, a wave model has limitations in the coastal zone, where wave interactions with bathymetry and land-shading effects require often regional nested very high resolution models to improve the simulations. Therefore this assessment is complementary to the use of a ground-truth such as a large buoy dataset, but can still be useful to derive further noise characteristics of the retracking w.r.t. an independent source and erroneous estimates of SWH (although realistic and therefore not detectable by the outlier analysis) near the coast.

4.4.1 Jason-3

Figure 10 depicts the comparison statistics against the ERA5-h model for the retracked J3 datasets. In the first row, the correlation coefficient is presented as a function of dist2coast and sea state. Apart from MLE-3, all retrackerers show a very good correlation with a coefficient of around 0.97 for the overall and open-ocean scenario. However, in the coastal zone, MLE-3, MLE-4, and TALES show a deteriorated performance with a correlation of 0.8-0.85 (0.49 for MLE-3). The rest of the retrackerers show a very high correlation of 0.96 in the coastal zone. For average and high sea state the differences are less pronounced with most of the retrackerers showing a high correlation of around 0.92. For low sea states, Adaptive, and STARv2 prove the best correlation with values of around 0.87, whereas MLE-3, MLE-4, and TALES show degraded correlations. For very high sea states, all retrackerers (apart from Brown-Peaky) show similar degraded correlations of around 0.72.

The median bias in the second row of Figure 10 depicts how much the SWH is different from the model, with lower values indicating a more accurate dataset. In this case, the median bias depends on both the area of interest and the sea state. For coastal scenarios, it can be said that the estimates tend to be overestimated (meaning the retracked value is higher than the model, whereas for open-ocean the retrieval are rather underestimated. With an increasing sea state, the median bias also tends to increase. STARv2 strongly underestimates large wave heights, showing a median bias of 0.51 m. Abdalla et al. 2018 (Figure 6-top) have plotted the monthly bias for MLE-4 (J2 mission), which ranges at around 0.1 m (sign was aligned, since the definition of the bias is vice versa), which is in accordance of the median bias value of around 0.08 m in Figure 10 (centre row, left for dist2coast > 0 km).

The last row of Figure 10 depicts the SDD. All retrackerers show comparable SDD values of around 0.20-0.40 m for most of the conditions with the exception of increased values for very high sea states. MLE-3, MLE-4, and (partially) TALES show increased values for low and average sea states and particularly in the coastal zone. In Abdalla et al. 2015, an SDD value of 0.20 m was reported for MLE-4 for the SARAL/AltiKa mission, which is in good agreement with the value of about 0.25 m, when considering an average sea state as shown in Figure 10 (last row, right).

In conclusion, most of the retrackerers show similar performances in terms of correlation, median bias, and SDD for coastal, open-ocean scenarios and most of the sea states, when comparing the 1-Hz retracked data with the ERA5-h wave model. MLE-3, MLE-4, and TALES exhibit a deteriorated performance, especially in the coastal zone.

4.4.2 Sentinel-3

The assessed S3A retrackerers show for the overall and open-ocean scenarios a very high correlation against the ERA5-h wave model of around 0.96 (with the exception of SAMOSA: 0.89). This also applies to the coastal zone, where all retrackerers exhibit a high correlation of around 0.95, with the exception of MLE-4-PLRM (~0.75), TALES-PLRM (~0.85) and the standard L2 product SAMOSA (0.43). All retrackerers (apart of SAMOSA) show a very good correlation of ~0.9 of for average and high sea states. With respect to low sea states, LR-RMC/LR-RMC_HFA have the highest correlation of 0.88, followed by WHALES-SAR, STARv2-PLRM, and DeDop-Waver, which amount to around 0.83. SAMOSA shows the worst correlation across all sea states. None of the algorithms are able to correctly retrieve very high sea states, with an average correlation of about 0.2 (SAMOSA: around 0.0). The inaccurate estimates for very high sea states might be explained by the very few samples that are available in all datasets: Out of the 512 netCDF files (pole-to-pole tracks), there are only around 260 1-Hz SWH estimates apparent (in comparison for the J3 analysis: around 2100). STARv2-PLRM shows an inverse correlation of -0.20 for very high sea states, which again can be attributed to the scales of denoising that are likely to be too wide to correctly observe areas with very high waves. The authors of Abdalla et al. 2018 have reported correlation values of 0.98 and 0.94 for the CS2 NE Atlantic and Pacific Box, with the latter one being placed in the open-ocean. These are in a rough accordance to the evaluated value of 0.90, as shown in Figure 11 (top row, left, dist2coast > 0 km).

The median bias has only a minor dependence of the dist2coast. LR-RMC, LR-RMC_HFA, MLE-4-PLRM, TALES-PLRM, and STARv2-PLRM show a very small median bias of less than 0.10 m in the coastal zone. SAMOSA exhibits a very large bias in the coastal zone with a value of -0.47 m, which is in accordance to the degraded correlation, as discussed before. The LR-RMC variants incorporate almost no median bias in open-ocean and in all sea states with less than 0.05 m. For very high sea states, most of the retrackerers have a high median bias. This is as expected from the correlation analysis and might be due to the very few samples that are available for such high sea states. The bias that was retrieved for SAMOSA in the NE Atlantic or Pacific Box of CS2 is very low with values of about -0.08 m and -0.03 m (sign swapped by authors, since the definition of the bias is swapped), as shown in (Figure 4 in Abdalla et al. 2018), which is very well in accordance with the bias value of -0.06 m for the average sea state (Figure 11, centre row, right).

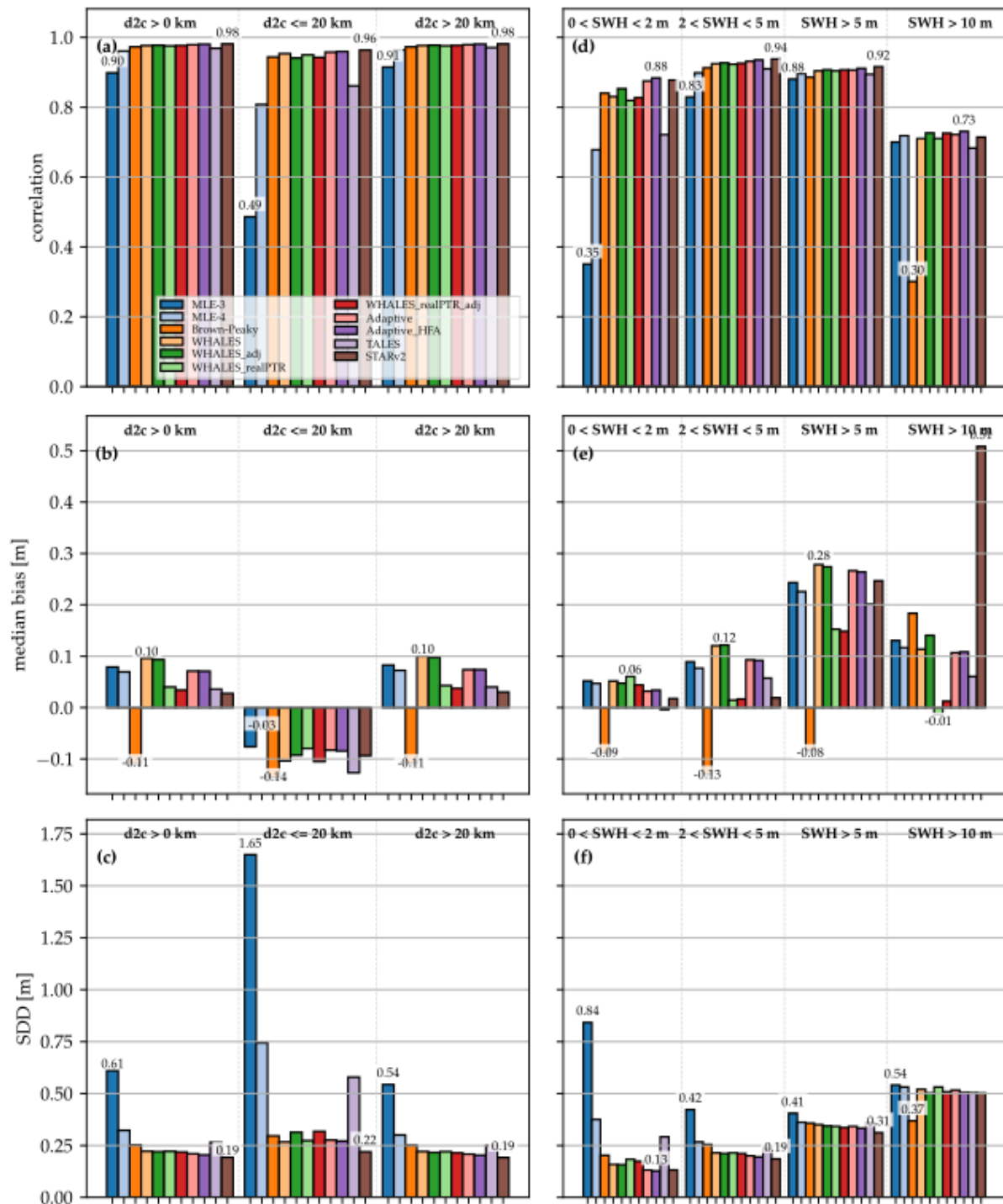


Figure 10. Comparison of the correlation coefficient, median bias, SDD against ERA5-h model of the individual J3 retracers as a function of dist2coast on the left-hand side and as a function of SWH on the right-hand side.

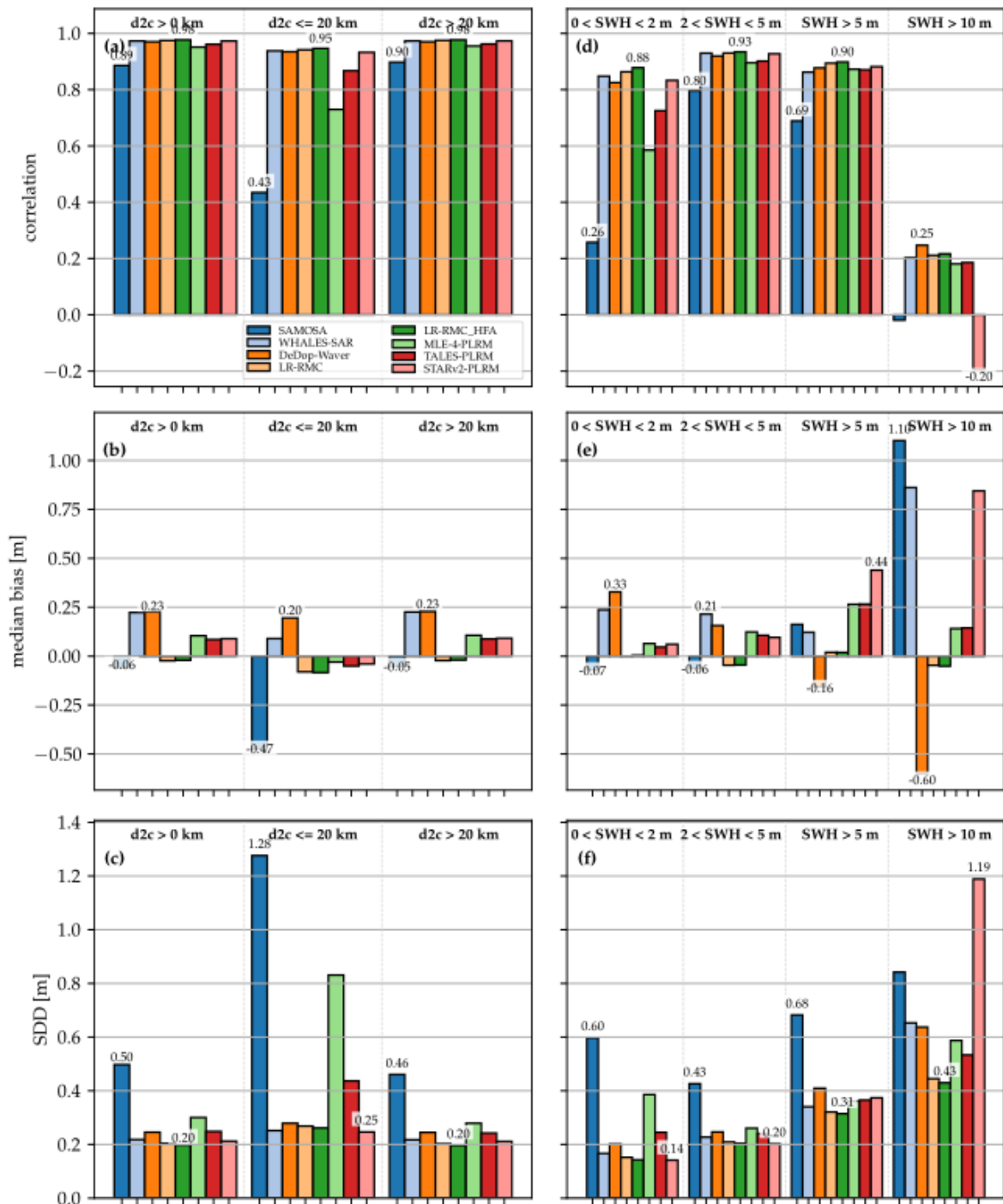


Figure 11. Comparison of the correlation coefficient, median bias, SDD against ERA5-h model of the individual S3A retracers as a function of dist2coast on the left-hand side and as a function of SWH on the right-hand side.

4.5 Comparison against In-situ Data

The statistics for the comparison of the median of 51 nearest 20 Hz observations with buoys are shown in Figure 12 and Figure 13 for J3 and S3A as a function of dist2coast and the sea state on the left and right columns, respectively. The three metrics (correlation, median bias, and SDD) are presented and discussed in this section.

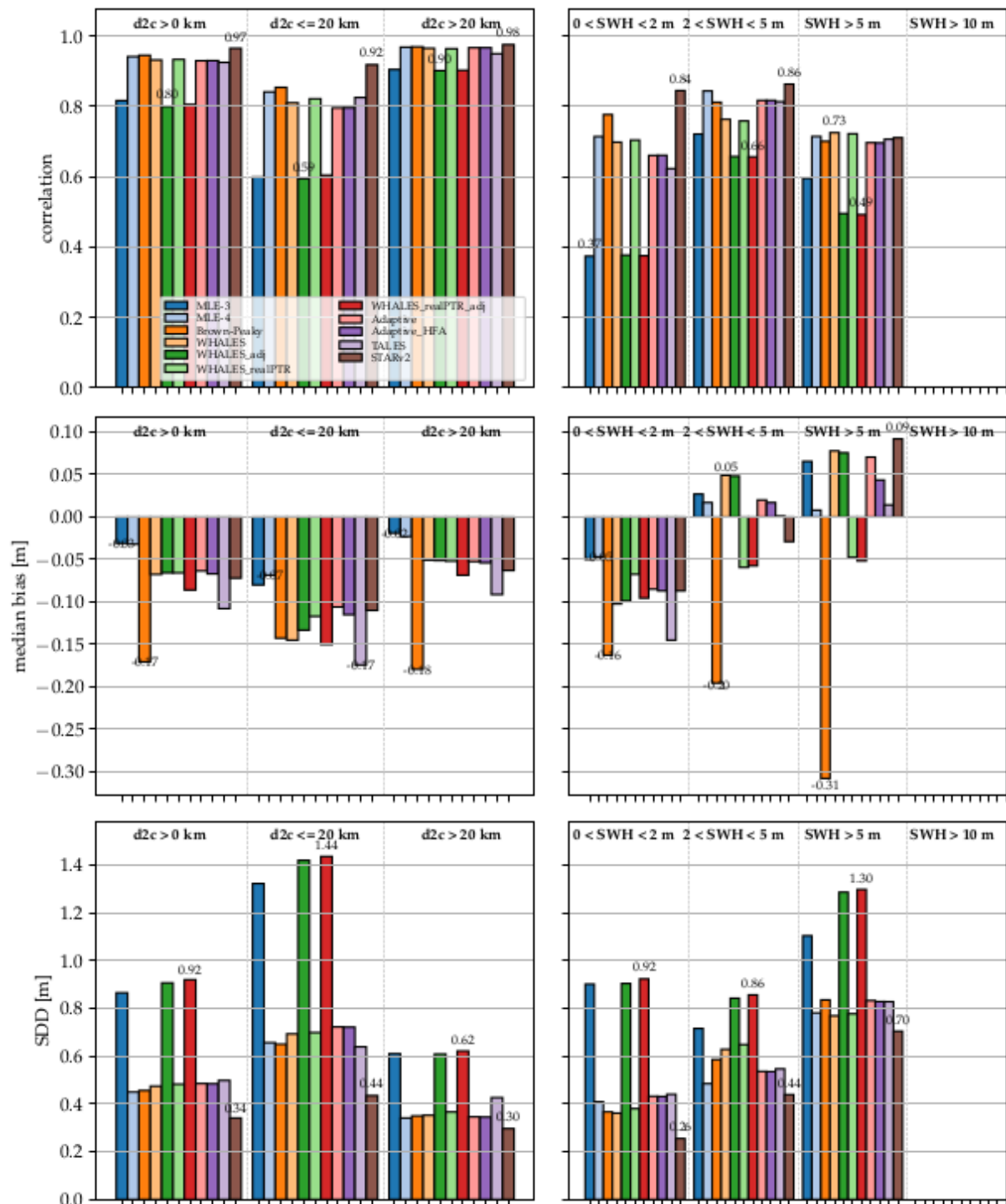


Figure 12. Comparison of the correlation coefficient, median bias, SDD against in situ buoys for the individual J3 retrackers as a function of dist2coast on the left-hand side and as a function of SWH on the right-hand side.

4.5.1 Jason-3

Figure 12 depicts the comparison statistics against buoy data for the retracked J3 datasets. In the first row, the correlation coefficient is presented as a function of dist2coast and sea

state. All the retrackers produce SWH values that are highly correlated with the buoy data for locations far from the coast. This is consistent with many previous studies, which have restricted validation data to those buoys in deep water and show correlations of >0.9 (Ref1, ref2). In this scenario, the MLE-3, WHALES_adj and WHALES_realPTR_adj perform worse than the default (MLE-4) algorithm, with STARv2 being slightly better than all the others. It is a little surprising that the algorithms reducing the small-scale noise on the estimates also diminish the correlation with independent standards. In the coastal zone (here defined as $\text{dist2Coast} < 20$ km) performance of all algorithms is reduced significantly, but the same pattern in relative performance remains i.e. MLE-3, WHALES_adj and WHALES_realPTR_adj are the worst, but in this case STARv2 is more clearly the best. As well as the chance of potential land contamination of the waveform data, a switch in comparisons from open ocean to coastal will also result in a different set of typical SWH conditions, with lower SWH values nearer the coast. The plots on the right-hand side show the correlation properties as a function of SWH, mixing together both the coastal and open ocean data. (All correlation values are much less than for the general open ocean case in the left-hand column, because in each case data are only being compared over a narrow SWH span, which affects the correlation estimate more than the bias and SDD.) As expected all algorithms struggle to perform as well at $\text{SWH} < 2$ m, because of the poor waveform sampling of the leading edge. However, yet again the same three algorithms perform much worse than MLE-4 and STARv2 is clearly the best. There is also some differentiation between the other algorithms with Brown-Peaky's performance exceeding the others for these low wave heights. STARv2 is also the best performer in the 2-5m category, being the only one to perform better than the default product (MLE-4), with all the good algorithms giving similar correlations for the high SWH conditions.

In terms of bias, all algorithms return a negative value (i.e. buoy value less than altimeter estimate) for both open ocean and coastal conditions. In most cases, the discrepancy is of order 5 cm (except for Brown-Peaky), and could of course be easily corrected for within the algorithms. However, the error is not a simple offset but a function of the wave height conditions encountered. Indeed, for this measure, MLE-4 provides the most stable bias of the algorithms assessed.

The third row shows the standard deviation of the differences, which will include errors in the altimeter algorithms, errors in the buoy data and errors due to the mismatch in locations. However, the lowest errors will still indicate the best performing algorithms, and will tend to mirror the results shown in the plots of correlation coefficient. Thus in terms of SDD, MLE-3, WHALES_adj and WHALES_realPTR_adj show worse performance than MLE-4, with STARv2 being the best. For this measure, STARv2 shows its superiority in all dist2Coast and SWH categories.

4.5.2 Sentinel-3

The data analysis for Sentinel-3A (Figure 13) shows many similar patterns, but with more extreme results. In particular there are very good correlations for all algorithms in the open ocean, apart from the default algorithm (SAMOSA). The effect is even more pronounced in the coastal zone, which is very surprising given that DDA algorithms are expected to do much better in this regime than LRM-based ones. For example in a study of Sentinel-3 performance compared with coastal buoys around the UK, Nencioli and Quartly (2019) found

that correlations remained high and SDD low as SAMOSA was used near the coast, whereas the PLRM default (MLE-4) got worse in those situations. The two new algorithms applied to the PLRM data (TALES and STARv2) generally perform better than the default (MLE-4).

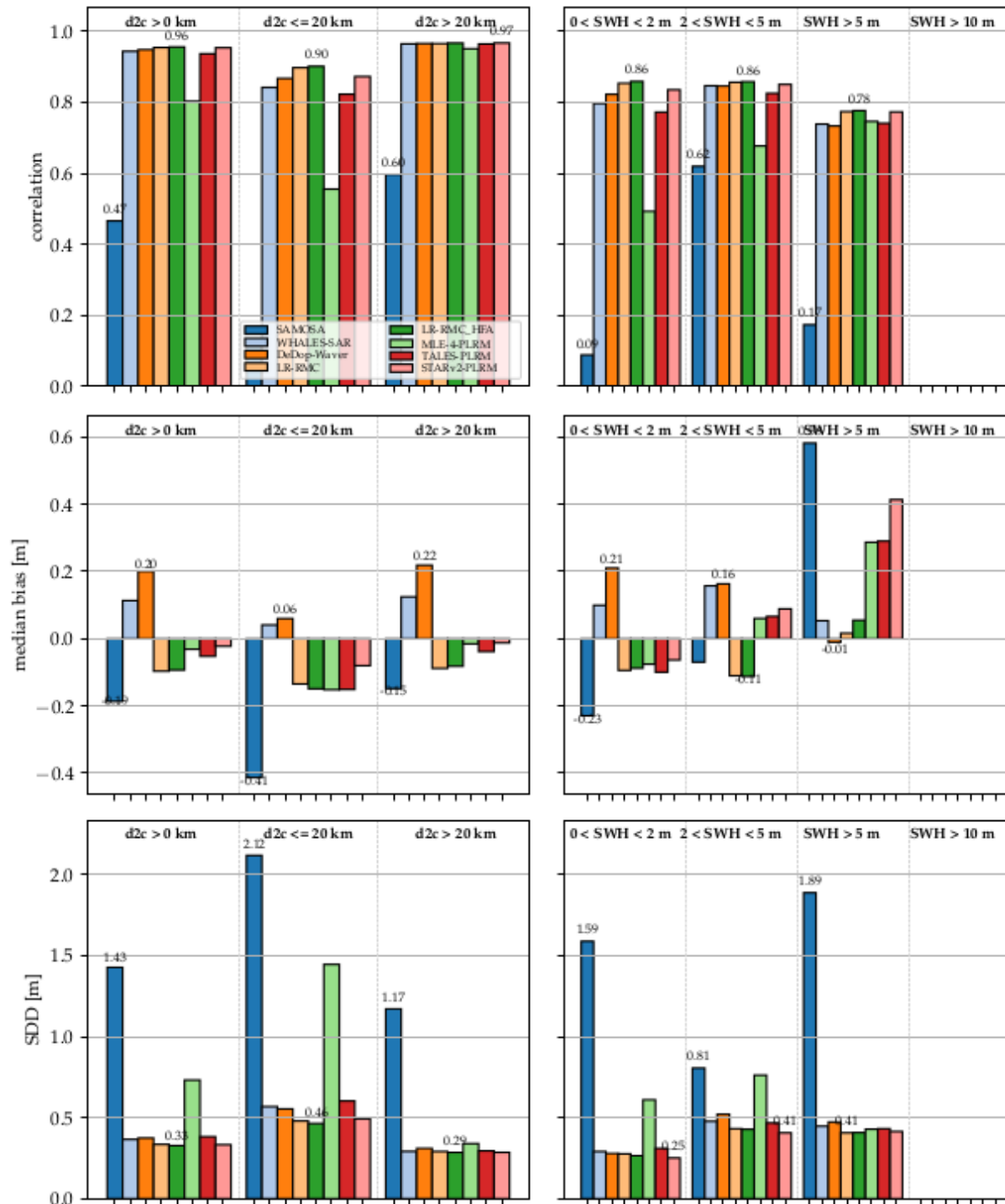


Figure 13. Comparison of the correlation coefficient, median bias, SDD against in situ buoys for the individual S3A retrackerers as a function of dist2coast on the left-hand side and as a function of SWH on the right-hand side.

In terms of bias, SAMOSA shows the most extreme values: most negative in the coastal zone, most negative in the 0-2 m range and most positive in the SWH>5 m category. The variability in bias also seems larger for several algorithms than was the case for J3 data, possibly reflecting less prior effort in tuning of algorithms to buoys. The plots of SDD show a fairly consistent picture that all the new algorithms perform better than the default ones (SAMOSA and MLE-4), giving errors of <0.3m, which is slightly better than that shown for J3 (bottom row of Figure 12). However, it should be borne in mind that in each case buoys are compared with the nearest track, and that due to their different spatial sampling strategies the typical separation of buoys from S3A passes is much less than the separation from J3 passes

4.5.3 Overview of Buoy Comparison Results

STARv2 provides the best overall performance for both J3 (LRM) and S3A (PLRM); of the DDA algorithms, the best is LR-RMC_HFA. One caveat on this set of results is that the methodology for STARv2 fits an evolving SWH profile through large sections of data, and thus the median of 51 20-Hz observations that we use here actually incorporates information from a larger spatial extent. This makes its estimates less sensitive to random noise; however the magnitude of this effect is hard to quantify and is believed to vary with SWH regime. Secondly, we note that the SDD values coming out of this Round Robin analysis are a little higher than those from some other researchers (REF, REF). This is because the buoy selection was done without sight of the altimeter sampling pattern, such that in practice some buoys are not well positioned for validation of the chosen tracks. The methodology laid down in the original Round Robin document was retained, although other researchers choose to be more selective about which buoy-altimeter match-ups are included e.g. see the methodology of Nencioli and Quartly (2019). The relatively poor performance of SAMOSA is a surprise. It is possible that the algorithm may be good and this may be a flagging issue i.e. that the supplied quality flags with the SAMOSA product let too many poor values through. However, the Round Robin exercise was designed as an assessment of the products (estimated SWH values and associated quality-flagging), so there has been no effort devoted to separately assessing the efficacy of the flagging. We note that a new version of SAMOSA has been implemented since we performed the Round Robin analysis, but this version has not been assessed.

4.5.4 Perspectives

An important fact that was not considered in this evaluation is the numbers of valid 1-Hz measurements that are mainly attributed to the supplied quality flag. As described in Section 4.1, the retrackerers show significant differences in terms of number of outliers. Two of the best performing algorithms LR-RMC and STARv2 also show the highest number of outliers within 20 km of the coast. It becomes obvious here that it is a trade-off between quality and quantity of the measurements. The new retrackerers analysed in this study are provided with an effective quality flag that allows reliability of the estimates in the coastal zone, but the amount of good quality data differs significantly among the datasets.

In an extension of the work by Quartly & Kurekin (2020), some of the buoy assessments were redone using 20-point averages (effectively 1 Hz means), but with different requirements (1, 5, 10 or 20) on the number of valid measurements used in calculating the average. Not surprisingly all algorithms showed an improvement when averages based on only a few points were discarded, but the improvement was much greater for some than others. This emphasises how important correct flagging is in the evaluation and utilisation of an algorithm.

For future assessments, we recommend to take also into account the number of valid measurements.

4.6 Round Robin SAR

Generally, both algorithms show quite high precision with RMSE of around 26 cm for wv1 and 28 cm for wv2 by comparison with CMEMS and around 41 cm by comparisons with NDBC. The resulting RMSE is in the order of ground truth noise, as the scattering between CMEMS and WW3 hindcast and also NDBC results in RMSE of 28 cm by comparison for individual NDBC buoy locations. This high precision RMSE of in average 27 cm has not yet been published in literature and is an achievement performed within the framework of this project.

The Round Robin comparisons show both algorithms having identical SWH results with a total score difference of ~ 0.001 . DLR has a little better RMSE achieved and Ifremer has slightly less filtered data (No-Sea-State percentage). By CMEMS validation, the total RMSE averaged for wv1 and wv2 is 0.263 m (DLR) and 0.273 m (Ifremer), the no-sea-state percentage is 1.08 % (DLR) and 0.24 % (Ifremer). For the NDBC validation the similar number can be obtained: RMSE of 0.414 m (DLR) and 0.445 m (Ifremer) and no-sea-state percentage of 1.45 % (DLR) and 0.83 % (Ifremer).

The resulting total scores for both algorithms calculated according to the formulas given in section 5.2. results in ~ 0.14 .

For total comparisons for SWH (all conditions) estimated by DLR and Ifremer the following Tables show the detailed comparison of both algorithms. The comparisons are also presented in [Figure 14](#) (for RMSE), [Figure 15](#) (for BIAS), [Figure 16](#) (for outliers) and [Figure 17](#) (for no-sea-state percentage).

SAR Round Robin total comparison					
		wv1		wv2	
		DLR	IFREMER	DLR	IFREMER
RMSE, m	CMEMS	0.24635	0.26013	0.28019	0.28548
	NDBC	0.41577	0.42903	0.41231	0.46226
BIAS, m	CMEMS	-0.03593	-0.00757	-0.02535	0.01016

	NDBC	-0.10321	0.11152	-0.10045	0.12034
No-sea-state percentage, %	CMEMS	0.65810	0.11721	1.50771	0.12251
	NDBC	1.50771	0.12251	2.33298	1.18153
Outliers percentage, %	CMEMS	0.87650	0.57940	0.84942	0.47539
	NDBC	0.97120	0.86455	1.12433	0.65217

SAR Round Robin CMEMS S1 wv1									
Domain SWH, m	SWH %	RMSE, m		BIAS, m		No-sea-state percentage, %		Outliers percentage, %	
		DLR	IFREMER	DLR	IFREMER	DLR	IFREMER	DLR	IFREMER
0.0 - 1.5	11.764	0.28008	0.31014	-0.12171	0.17958	2.79656	0.10586	0.43639	0.60466
1.5 - 3.0	61.959	0.19675	0.20573	-0.03136	-0.04886	0.50908	0.10996	0.68691	0.42965
3.0 - 6.0	24.491	0.30428	0.31337	-0.01216	0.01922	0.05302	0.13759	0.97499	0.50920
6.0 <	1.786	0.51955	0.58345	0.02849	-0.17505	0.08078	0.16407	1.09047	0.12325
TOTAL	100.00	0.24635	0.26013	-0.03593	-0.00757	0.65810	0.11721	0.87650	0.57940

SAR Round Robin CMEMS S1 wv2									
Domain SWH, m	SWH %	RMSE, m		BIAS, m		No-sea-state percentage, %		Outliers percentage, %	
		DLR	IFREMER	DLR	IFREMER	DLR	IFREMER	DLR	IFREMER
0.0 - 1.5	11.893	0.34212	0.35084	-0.15318	0.19698	5.24198	0.13409	0.43233	0.50656
1.5 - 3.0	61.740	0.22708	0.22315	-0.02274	0.04539	1.35065	0.11037	0.89351	0.42080
3.0 - 6.0	24.580	0.33291	0.34659	0.02356	0.01367	0.18875	0.14450	0.89729	0.40756
6.0 <	1.786	0.55897	0.60751	0.02022	0.12333	0.24029	0.16234	1.04125	0.20325
TOTAL	100.00	0.28019	0.28548	-0.02535	0.01016	1.50771	0.12251	0.84942	0.47539

SAR Round Robin NDBC S1 wv1									
Domain SWH, m	SWH %	RMSE, m		BIAS, m		No-sea-state percentage, %		Outliers percentage, %	
		DLR	IFREMER	DLR	IFREMER	DLR	IFREMER	DLR	IFREMER

0.0 - 1.5	28.737	0.37024	0.43051	-0.24418	0.34686	2.01342	0.65359	0.67114	0.65789
1.5 - 3.0	51.302	0.37357	0.36582	-0.10264	0.04284	0.00000	0.56180	1.01730	1.12994
3.0 - 6.0	18.322	0.50478	0.52764	0.04289	0.01574	0.00000	0.00000	1.05263	1.05263
6.0 <	1.639	0.95620	0.84081	0.66728	0.56993	0.00000	0.00000	0.00000	0.00000
TOTAL	100.00	0.41577	0.42903	-0.10321	0.11152	0.57859	0.47801	0.97120	0.86455

SAR Round Robin NDBC S1 wv2									
Domain SWH, m	SWH %	RMSE, m		BIAS, m		No-sea-state percentage, %		Outliers percentage, %	
		DLR	IFREMER	DLR	IFREMER	DLR	IFREMER	DLR	IFREMER
0.0 - 1.5		0.34758	0.44386	-0.23071	0.35301	4.97238	1.10803	0.27624	0.28011
1.5 - 3.0		0.37292	0.36480	-0.07603	-0.01096	1.00251	1.27877	1.75439	1.03627
3.0 - 6.0		0.50125	0.59251	0.04294	-0.01076	0.00000	1.20482	1.18343	0.60976
6.0 <		1.16420	1.15387	0.74060	-0.71695	0.00000	0.00000	0.00000	0.00000
TOTAL		0.41231	0.46226	-0.10045	0.12034	2.33298	1.18153	1.12433	0.65217

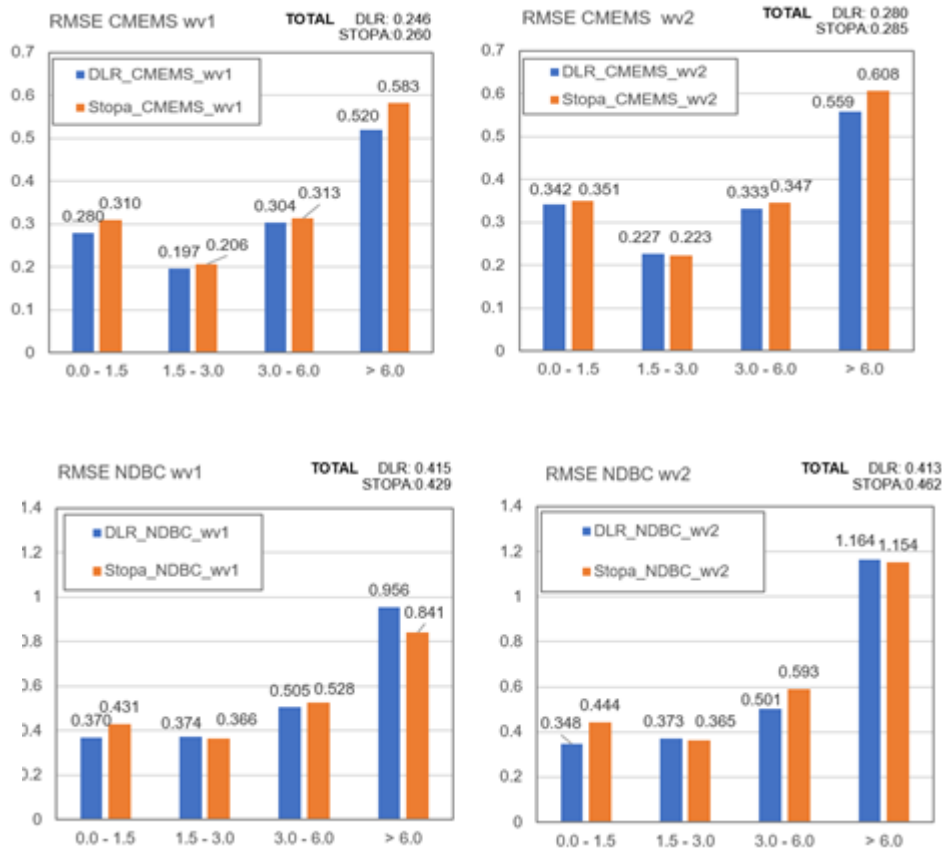


Figure 14: SAR Round Robin detailed comparison RMSE

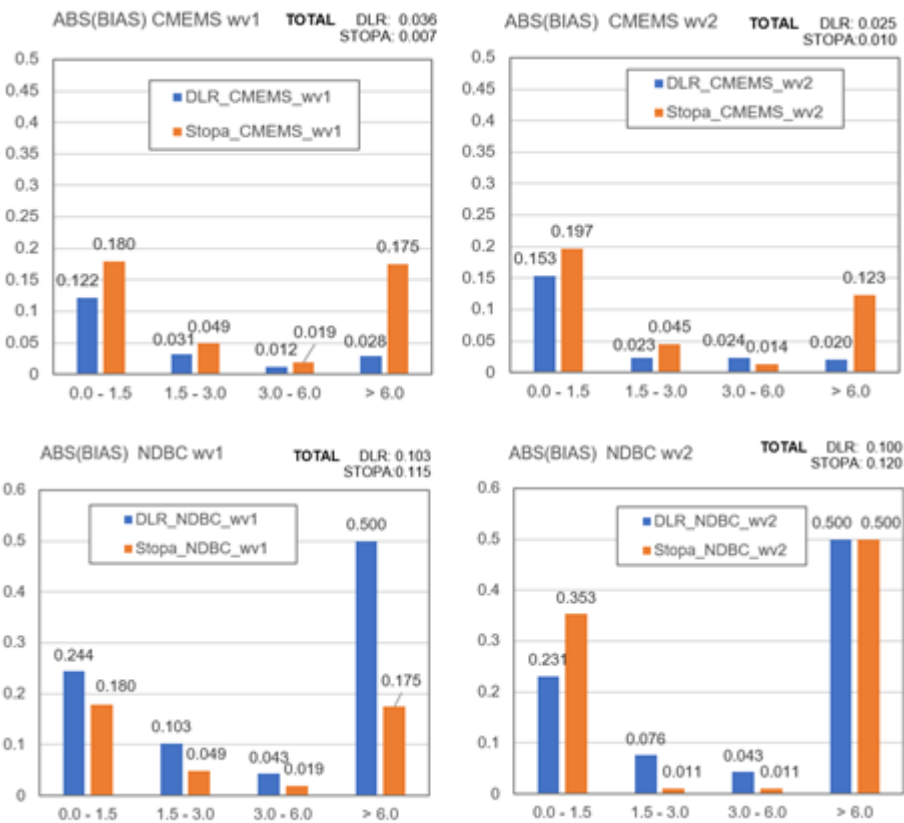


Figure 15: SAR Round Robin detailed comparison ABS(BIAS).

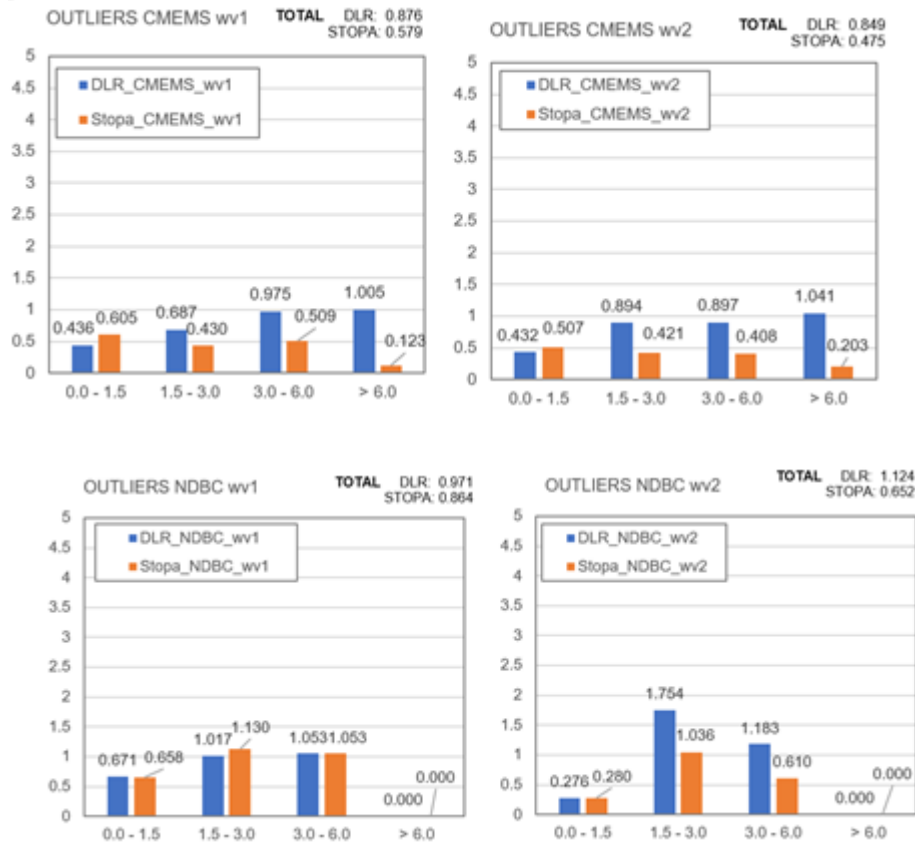
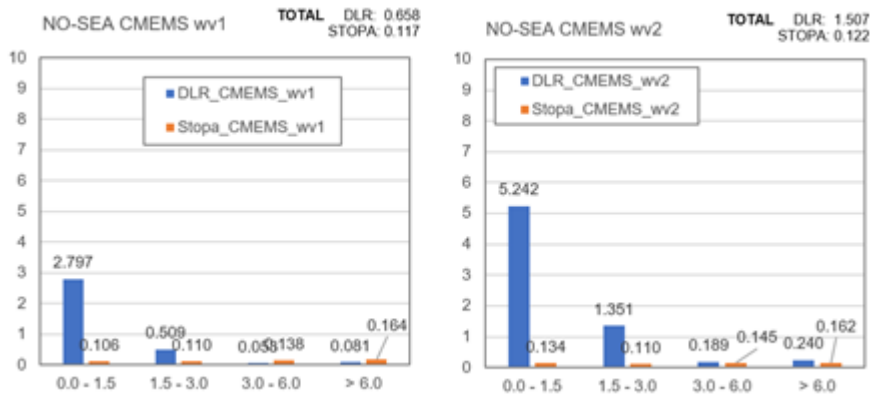


Figure 16: SAR Round Robin detailed comparison. Outliers with local SWH >3(local RMSE) for each bin.



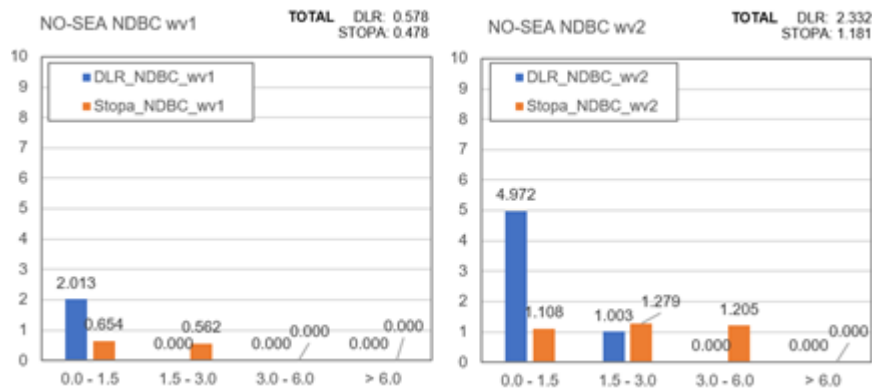


Figure 17: SAR Round Robin detailed comparison. Percentage of no-sea-state percentage (non-valid).

Figure 18 shows the RMSE distribution for four sea state domains and mean RMSE noise estimated.

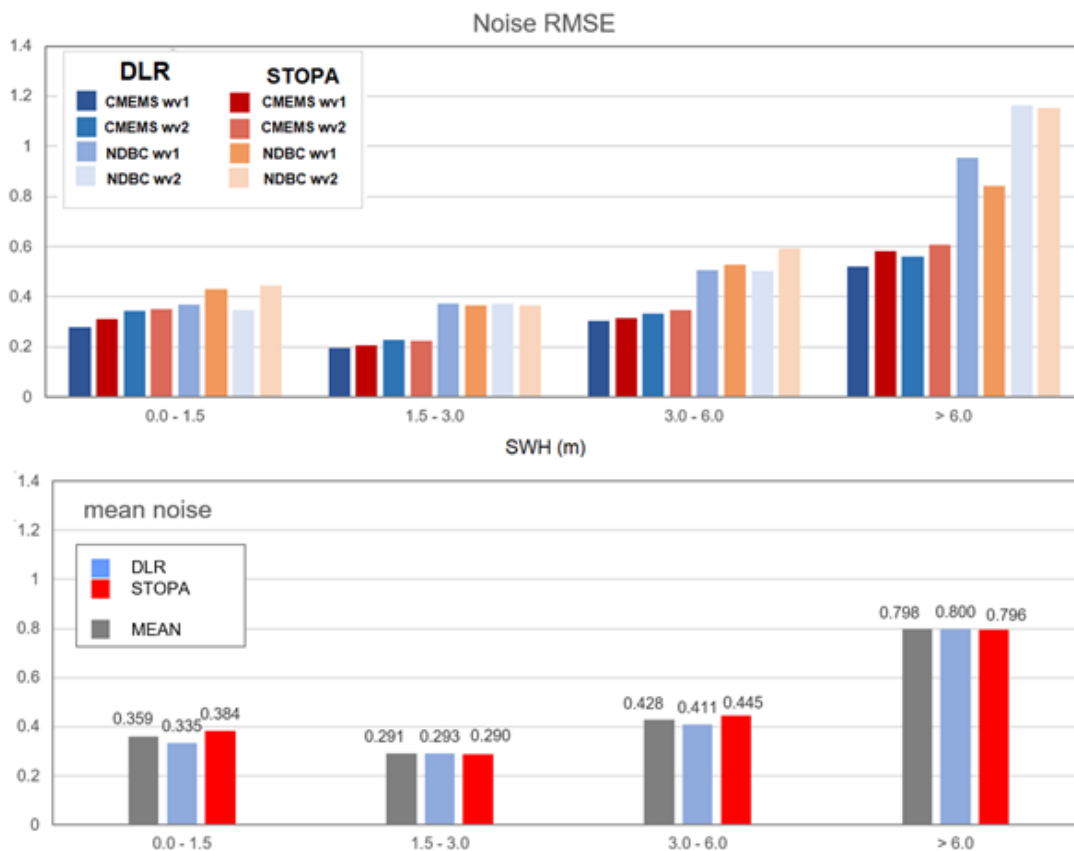


Figure 18 RMSE distribution for four sea state domains and mean RMSE noise estimated.

The noise analysis shows similar trends for both algorithms:

- the accuracy averaged for forecast model and buoy measurements is in the order of ~30 cm for the sea state category of moderate sea state, where ~60% of all data points are located

- this averaged accuracy declines with increasing SWH values to ~42 cm for rough sea state with SWH about 4-5m and to ~80 cm for very high sea state over 6m SWH. This is connected to both accuracy of the SAR methods and to an increased error in the ground truth, which also grows proportional to the larger sea state values. However, if one connects the local RMSE with the mean value for this sea state domain, the same scatter index SI can be obtained.
- for low sea state (SWH<1.5) m some difficulties can be seen in comparison to moderate sea state, the lower accuracy does not match the tendency of higher error with higher sea state described before. This effect is connected to the specifics of SAR imaging of sea surface, where the short and small waves cannot be imaged individually, but are only visible as noise. Although it is possible to derive their characteristics from the noise, the accuracy is slightly lower in contrast to visible wave patterns.

Note that the validation was carried out for $-60^{\circ} < \text{LAT} < 60^{\circ}$ in order to avoid ice coverage. However, as analyses showed, sea ice can be encountered until $-60^{\circ} < \text{LAT} < 55^{\circ}$ so that both SWH, from the model and estimated from S1 WV, may be affected. For the DLR algorithm a number of outliers and no-sea-state flags are especially high in this area. In case the validation area is reduced to $55^{\circ} < \text{LAT} < 60^{\circ}$, the resulting total RMSE improves by ~1.5 cm.

Figure 19 shows the distribution of squared difference between SWH estimated form S1 and ground truth (CMEMS). On the graph all 270.000 RR data points are shown, the horizontal axis means latitude of acquired and compared data. As can be seen, in both algorithms, an increase of these differences appears at $\text{LAT} < -58^{\circ}$. This inconsistency is due to sea ice.

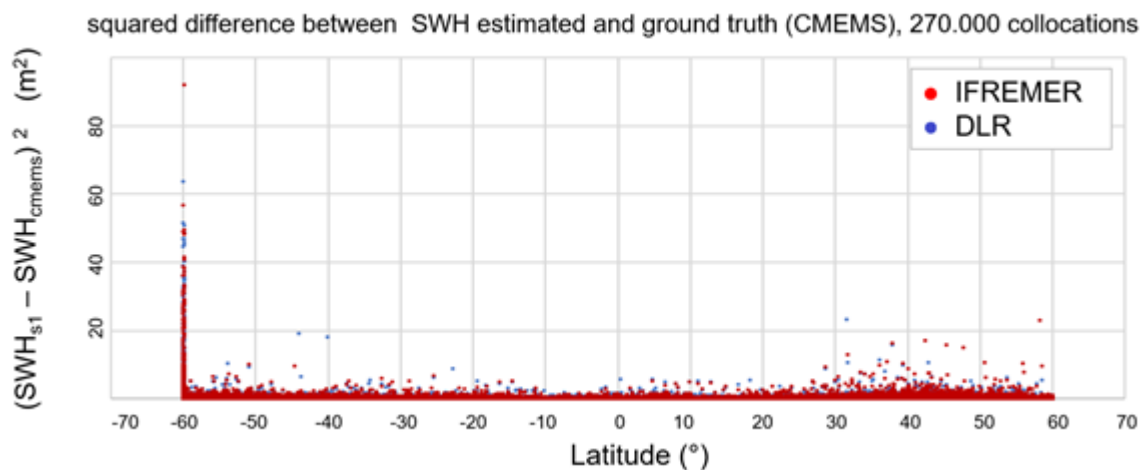


Figure 19: Distribution of squared difference between SWH estimated form S1 and ground truth (CMEMS). For both algorithms, the ice coverage impact is quite visible for $\text{LAT} < -58^{\circ}$. The complete comparison was done for $-60^{\circ} < \text{LAT} < 60^{\circ}$ in order to avoid ice coverage. As can be seen this masking is not enough to completely eliminate the ice uncertainty produced by both: CMEMS and S1.

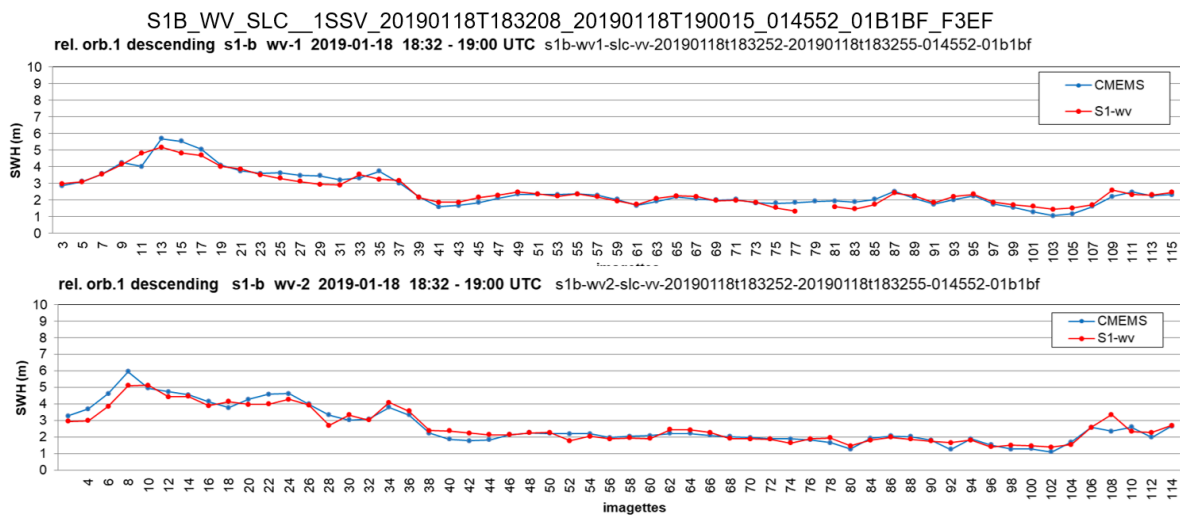


Figure 20: An example of an along-orbit comparison for one long overflight of around 12.000 km (product ID S1B_WV_SLC_1SSV_20190118T183208_20190118T190015_014552_01B1BF_F3EF) plotted separately for wv1 (even numbers on the first graph, 200 km between each points) and wv2 (odd numbers on the second graph) vignettes. Red points mean DLR SWH and blue are CMEMS SWH temporally interpolated to 3h outputs.

In order to prove that the method does not include a smoothing, along-track comparisons and comparisons of PDFs were carried out. A typical example of an along-orbit comparison for a long overflight of around 12.000 km is presented in Figure 20. The wv1 (even numbers on the first graph, 200 km between each points) and wv2 (odd numbers on the second graph) imagettes are plotted separately. The red points stands for DLR SWH and blue points stands for CMEMS SWH temporally interpolated to 3h outputs.

Figure 21 shows the PDF for SWH distribution in ground truth (CMEMS, blue curves in both graphs) and in SWH estimated from S1 WV vignettes (red curves) built for both algorithms. The Ifremer algorithm comparison is on the left panel, the DLR results are on the right panel, the differences between PDF-S1 and PDF-CMEMS for both algorithms are green. As the DLR algorithm was tuned with CMEMS data, the comparison results in a slightly better difference. The Ifremer algorithm was tuned using altimeter data, this probably results in a light smoothing effect for low wave heights.

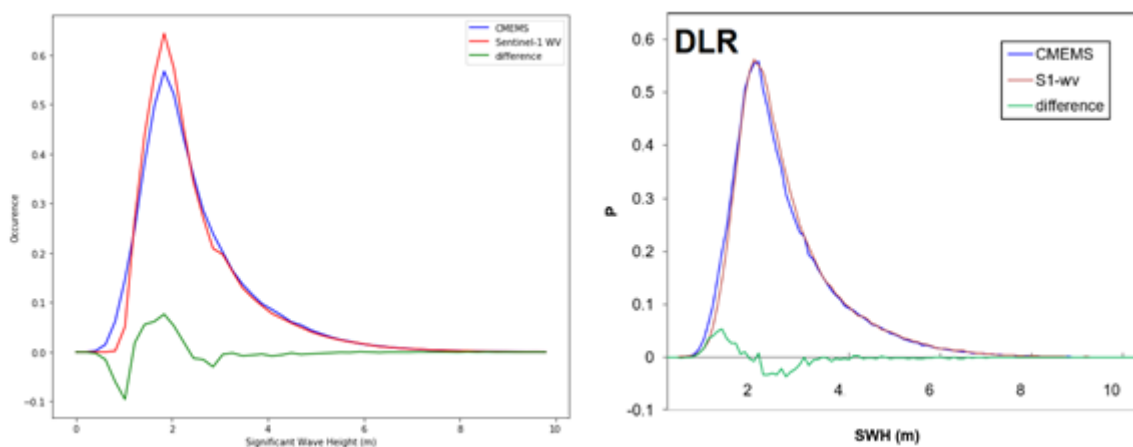


Figure 21 PDF for SWH distribution for CMEMS and estimated from S1 WV imagettes.

5. Weighting matrix for results

5.1 Altimeter

The Altimeter Algorithm Development Team and in particular PML and TUM, which are responsible for the Round Robin exercise, brought the results of the Round Robin without applying any weighting matrix. Considering these, the Consortium has taken a decision in collegiality during a progress meeting.

The results and criteria of this decision process have been summarised and approved by ESA in the document “Round Robin: Final selection and ranking of algorithms” [Sea_State_cci_RR_Final_Selection_v1.1-signed.pdf].

5.2 Synthetic Aperture Radar

The four performance parameters defined in section 3.2 were calculated for each algorithm, with a weighting according to the significant wave height ranges. The performance parameters are calculated for Sentinel1 WV1 and WV2 wave mode and considered with equal weighting in the total score:

$$\text{total score} = 50\% \cdot \text{WV1}_{\text{score}} + 50\% \cdot \text{WV2}_{\text{score}}$$

The separate evaluations using either collocations with CMEMS model hindcast data or with NDBC buoy data are weighted stronger towards the model data according to:

$$\text{WV1}_{\text{score}} = 70\% \cdot \text{CMEMS}_{\text{score}} + 30\% \cdot \text{NDBC}_{\text{score}}$$

$$\text{WV2}_{\text{score}} = 70\% \cdot \text{CMEMS}_{\text{score}} + 30\% \cdot \text{NDBC}_{\text{score}}$$

All four performance parameters RMSE, BIAS, Outliers percentage, No-sea-state percentage are normalized for each of the four sea state domains:

$$\text{PARAMETER_NORMALIZED}_{\text{bin}} = \frac{\text{PARAMETER} - \text{PARAMETER}_{\text{min}}}{\text{PARAMETER}_{\text{max}} - \text{PARAMETER}_{\text{min}}}$$

The following Table lists the Min and max boundaries for normalization of RMSE, BIAS, Outliers percentage and No-sea-state percentage

bin	SWH range, m	RMSE, m		No-sea-state, %		Outliers, %		ABS(BIAS), m	
		min	max	min	max	min	max	min	max
1	0m-1.5m	0.2	0.6	0.0	30	0.0	5	0.0	0.5

2	1.5m-3m	0.2	0.6	0.0	30	0.0	5	0.0	0.5
3	3m-6m	0.2	1.0	0.0	50	0.0	10	0.0	0.5
4	>6m	0.2	1.0	0.0	60	0.0	10	0.0	0.5

The score for each wv1 and wv2 for CMEMS and NDBC is a sum of products:

$$WV1_{score} / WV2_{score} = \sum_{N=1}^4 K_N \text{PARAMETER_NORMALIZED}_N$$

The following Table shows the weightings for each bin of SWH for each of the performance parameters: RMSE, BIAS, Outliers percentage, No-sea-state percentage

bin	Weighting factor K(N) for different parameters									
	N	SWH range, m	RMSE		No-sea-state percentage		Outliers percentage		BIAS	
			CMEMS	NDBC	CMEMS	NDBC	CMEMS	NDBC	CMEMS	NDBC
1		0m-1.5m	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
2		1.5m-3m	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62
3		3m-6m	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24
4		>6m	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

The final decision process for all algorithms selected for production was similarly held on a collegiality basis, as defined in the previous subsection for the Altimeter Round Robin.

6. Summary

The PVASR aimed at presenting the plan for the Round Robin exercise that determined the algorithm(s) to be used in the generation of the official SS_cci product. The Round Robin exercise has been separated according to the sensors used to produce the SWH estimations: Synthetic Aperture Radars and Radar Altimeters.

Key statistics that were computed on the test datasets have been described in the document and focus on internal evaluation and comparison with buoys and model data.

The RR for SAR shows for both algorithms quite similar results with difference of total score of ~0.001%. DLR has a little better RMSE and Ifremer has slightly less filtered data (No-Sea-State percentage). On a collegially basis, it was decided that both algorithms should be used for processing, as the Ifremer algorithm is independently trained from model data and the DLR algorithm can provide additional sea state parameters like wave periods or partially integrated parameters.

7. References

Bidlot J.-R., D. J. Holmes, P. A. Wittmann, R. Lalbeharry, H. S. Chen, 2002: Intercomparison of the performance of operational ocean wave forecasting systems with buoy data. *Wea. Forecasting*. 17. 287-310.

Smith, Walter H.F., and Remko Scharroo. 'Waveform Aliasing in Satellite Radar Altimetry'. *IEEE Transactions on Geoscience and Remote Sensing* 53, no. 4 (2015): 1671–1682. <https://doi.org/10.1109/TGRS.2014.2331193>.

Dinardo, Salvatore, Luciana Fenoglio-Marc, Christopher Buchhaupt, Matthias Becker, Remko Scharroo, M. Joana Fernandes, and Jérôme Benveniste. 'Coastal SAR and PLRM Altimetry in German Bight and West Baltic Sea'. *Advances in Space Research* 62 (1 September 2018): 1371–1404. <https://doi.org/10.1016/j.asr.2017.12.018>.

Gommenginger, Christine, Cristina Martin-Puig, Laiba Amarouche, and R. Keith Raney. 'Review of State of Knowledge for SAR Altimetry over Ocean. Report of the EUMETSAT JASON-CS SAR Mode Error Budget Study'. Monograph, 21 November 2013. <https://eprints.soton.ac.uk/366765/>.

Fenoglio-Marc, L., S. Dinardo, R. Scharroo, A. Roland, M. Dutour Sikiric, B. Lucas, M. Becker, J. Benveniste, and R. Weiss. 'The German Bight: A Validation of CryoSat-2 Altimeter Data in SAR Mode'. *Advances in Space Research* 55, no. 11 (June 2015): 2641–2656. <https://doi.org/10.1016/J.ASR.2015.02.014>.

Ardhuin, Fabrice, Justin E. Stopa, Bertrand Chapron, Fabrice Collard, Romain Husson, Robert E. Jensen, Johnny Johannessen, et al. 'Observing Sea States'. *Frontiers in Marine Science* 6 (2019). <https://doi.org/10.3389/fmars.2019.00124>.

Abdalla, Saleh. 'SARAL/AltiKa Wind and Wave Products: Monitoring, Validation and Assimilation'. *Marine Geodesy* 38, no. sup1 (10 September 2015): 365–80. <https://doi.org/10.1080/01490419.2014.1001049>.

Abdalla, Saleh, Salvatore Dinardo, Jérôme Benveniste, and Peter A.E.M. Janssen. 'Assessment of CryoSat-2 SAR Mode Wind and Wave Data.' *Advances in Space Research* 62, no. 6 (2018): 1421–1433. <https://doi.org/10.1016/j.asr.2018.01.044>.

ESA. 'Cryosat Plus for Oceans: CP4O', 23 February 2015, 65.

Dibarboure, G., F. Boy, J.D. Desjonqueres, S. Labroue, Y. Lasne, N. Picot, J.C. Poisson, and P. Thibaut, 2014: Investigating Short-Wavelength Correlated Errors on Low-Resolution Mode Altimetry. *J. Atmos. Oceanic Technol.*, 31, 1337–1362, [htxps://doi.org/10.1175/JTECH-D-13-00081.1](https://doi.org/10.1175/JTECH-D-13-00081.1)

Quartly, G.D., 2009, 'Optimizing σ^0 information from the Jason-2 altimeter.' *IEEE Geosci. Rem. Sensing Lett.*, 6 (3), 398-402. doi: 10.1109/LGRS.2009.2013973

Nencioli, F. & G.D. Quartly, 2019. 'Evaluation of Sentinel-3A wave height observations near the coast of Southwest England', *Remote Sensing*. 2019, 11(24), 2998 (20pp.) doi: 10.3390/rs11242998

Quartly, G.D. and A.A. Kurekin, 2020. 'Sensitivity of altimeter wave height assessment to data selection.' *Remote Sens.* 2020, 12, 2608. doi: 10.3390/rs12162608

Schlembach, F., M. Passaro, G.D. Quartly, A.A. Kurekin F. Nencioli, G. Dodet, J.-F. Piollé, F. Arduin, J. Bidlot, C. Schwatke, F. Seitz, P. Cipollini, and C. Donlon, 2020. 'Round Robin assessment of radar altimeter Low Resolution Mode and Delay-Doppler retracking algorithms for significant wave height.' *Remote Sens.* 2020, 12(8), 1254; doi: 10.3390/rs12081254 & doi: 10.3390/rs13061182