IN SITU TAC

INSITU_GLO_WAV_DISCRETE_MY_013_045

Issue: 5.1

Contributors: Marta de Alfonso, Fernando Manzano, Alejandro Gallardo

Approval date by the Copernicus Marine Service product quality coordination team: 24/11/2023

CHANGE RECORD

When the quality of the products changes, the QuID is updated and a row is added to this table. The third column specifies which sections or sub-sections have been updated. The fourth column should mention the version of the product to which the change applies.

Issue	Date	§	Description of Change	Author	Validated By
1.0	10/12/2017	all	First version of document	Marta de Alfonso Fernando Manzano	
2.0	10/01/2018	all	Modifications with the final configuration of the In Situ TAC wave REP product	Marta de Alfonso	
3.0	14/01/2019	all	Upgrade for CMEMS Phase2 and after In Situ TAC wave REP product update	Marta de Alfonso Alejandro Gallardo Fernando Manzano	Loïc Petit de la Villéon
4.0	04/09/2020	all	Wave REP product update and inclusion of wave spectra	Marta de Alfonso Alejandro Gallardo Fernando Manzano	Jerome Gourrion
4.0	30/11/2021	all	Correction of external links	Stéphane Tarot	Stéphane Tarot
4.1	21/02/2022	all	Update for Copernicus Marine 2 & added update frequency	Marta de Alfonso, Fernando Manzano, Alex Gallardo	Stéphane Tarot
5.0	06/06/2022	all	New product naming, new Production Centre and new template	Marta de Alfonso, Fernando Manzano, Alex Gallardo	Stéphane Tarot
5.1	02/08/2023	I.1, I.3.1	New dataset with hourly data	Marta de Alfonso, Fernando Manzano, Alex Gallardo	Stéphane Tarot

TABLE OF CONTENTS

I EXECUTIVE SUMMARY

I.1 Products covered by this document

The document describes the quality of the Delayed Mode Multi-Year (MY) WAVE product delivered by the Copernicus Marine Service In Situ Thematic Assembly Centre (In Situ TAC).

This document applies to the following list of products described in Copernicus Marine Service Catalogue (**Table 1**):

Short Description	Product code	Area	Delivery Time
GLOBAL MY	INSITU_GLO_WAV_DISCRETE_MY_013_045	GLOBAL	Twice a year

***Table 1**: List of In Situ TAC products for which this document applies.*

This product integrates observations aggregated and validated from the Regional EuroGOOS consortium (Arctic-ROOS, BOOS, NOOS, IBI-ROOS, MONGOOS) and Black Sea GOOS as well as from National Data Centres (NODCs), JCOMM global systems (OceanSITES, DBCP) and the Global telecommunication system (GTS) used by the Met. Offices. The In Situ TAC relies on observing systems maintained by institutes that are not part of the In Situ TAC, and Copernicus Marine Service is not contributing to the maintenance and setting up of the observing systems it uses.

The Multi-Year WAVE product is a global product and provides two kinds of files: one with integrated parameters, with the string “TS_MO” in the filename (TS: Time Series, MO: Mooring), computed from the wave spectrum (e.g. significant wave height, peak period, mean direction) or zero crossing parameters (e.g. maximum wave height, mean height) and another one with the string “WS_MO” in the filename (WS: Wave Spectra, MO: Mooring) with the spectral information (scalar spectrum and directional functions like mean direction and angular spreading depending on the frequency). The files with integrated parameters contain also other physical and meteorological variables measured by the same platform. The complete list of variables distributed by the In Situ TAC can be found in the Copernicus Marine In Situ TAC physical parameters list (https://doi.org/10.13155/53381). Each file contains only the wave parameters provided by the platform which is a subset of the complete list of wave parameters.

The Production Unit performs the wave validation in a centralized way twice a year, then the product is assessed through the metrics described in Section IV (Validation Results) and the final wave product is distributed in the dedicated product mentioned above.

Since December 2020, files with wave spectra are included in the Multi-Year WAVE product. Since 2023, hourly data is delivered in a dedicated dataset: cmems_obs-ins_glo_wav_my_na_PT1H.

I.2 Summary of the results

For this product, it is important to differentiate between the data validation process and the metrics employed for assessment. The data validation is performed through automatic procedures, visual inspection and comparison to other sources. Quality flags are positioned to inform the users of the level of confidence attached to the observations. The quality of the product is assessed through benefit assessment metrics described in Section IV (Validation Results) based on quality percentage and temporal and spatial coverage for scalar waves (wave height and period), directional waves (wave direction) and wave spectra.

The main results of the validation process are the following:

- Moorings (buoys and light vessels) are the platforms mostly used to measure waves. It should be considered there are deep water platforms but also coastal stations that are affected by local bathymetry and coastal processes.
- The temporal coverage of wave measurements starts with a low and stable increase in number of platforms from the 1970's to the early 1990's. At the beginning of the 2000 the number of platforms starts to grow, and, during the last two decades, the increase is higher due to the effort in integration of new providers, stations and historical datasets.
- Regarding the spatial coverage, most of the platforms are located in the Northern Hemisphere (Europe, North America, Japan and India), with some stations in South America and Australia. In some regions the number of available platforms is at a critical low level to provide an adequate representative overall view of the state of the ocean (Arctic, Southern and Eastern Mediterranean, Southern Black Sea and most of the Southern Hemisphere; see Figure 2 to locate the areas). Some of the areas are clearly under sampled and in some other areas data is not available. To gather all the wave observations, the In Situ TAC is dedicating a great effort that will continue in the future for both operational stations and historical data sets.
- The percentage of data flagged as 'good data' is over 95%.

I.3 Estimated Accuracy Numbers

Table 2 summarizes the accuracy of the measurements that can be expected depending on the sensors. This is the best accuracy that a user can expect for the in situ data to which a 'good data' quality flag (see **Figure 4Error! Reference source not found.**) has been applied after the validation process.

The definition of the reference values is obtained from the user manual of the sensors. The specific reference is given in Table 2 and the values are given for the different parameters.

Wave sensor	Measured time series		
	Vertical displacement (heave)	**Period**	**Direction**
Waverider (Datawell, https://datawell.nl/)	0.5% of the measured value	0.5% of the measured value	0.4 – 2 deg (dep. On latitude)
Wavesense (Fugro, https://www.fugro.com/)	0.1 m	0.15 s	1 deg
Triaxys (Axys, https://axys.com/)	1% of the measured value	1%	3 deg
	Estimated parameters (due to the statistical variability)		
	Wave heights[1]	**Wave periods[1]**	**Wave directions[1]**
All wave sensors	< 5% of the estimated value	< 5% of the estimated value	< 10 deg

Table 2. *Accuracy numbers for measured time series and wave estimated parameters for different wave sensors.*

I.3.1 Hourly dataset

Since 2023 a new dataset with hourly data and rounded timestamps is provided. This subsection shows the results of a reliability study carried out for different sampling intervals.

In early 2023, a total of 1732 platforms measuring waves were distributed in the WAVE MY product. Studying the median of the sampling interval, most of these platforms have a time sampling of one hour (693 platforms) or less (559 platforms). About other platforms, 25 platforms measure every 2 hours, 408 every 3 hours, 29 stations every 6 hours and only 18 provide data every 12 or 24 hours.

To verify if a linear interpolation of discrete sampling is possible, in order to fill gaps, we have performed a reliability study. This study assesses how the linear interpolation of wave data could affect the results depending on the different time samplings. Eight platforms (moored buoys) with long time series (more than 10 years) of hourly data exposed to different wave conditions and moored in different geographical

[1] Based on numerical time series simulation and intercomparing tests. The uncertainty in estimated parameters is inherent to the stochastic process and it's due to the statistical variability.

areas (coastal, open and deep waters, Atlantic and Pacific Ocean, enclosed seas) have been selected to get a representative sample. An image of the selected platforms can be seen in Figure 1. The information about the platforms is as follows: 55019 (Pacific Ocean, coastal), 51003 (Pacific Ocean, open and deep water), 42020 (Caribbean Sea, enclosed sea, coastal), 41010 (West Altantic Ocean, open waters), 6200082 (Bay of Biscay, open and deep water), 6400046 (East Atlantic Ocean, open and deep water), 6100417 (Mediterranean Sea, enclosed sea, open waters) and VaderoarnaWR (North Sea, coastal). The label of each platform is the code used in Copernicus to identify it and corresponds to the number assigned by the World Meteorological Organization (WMO) to moored buoys or the name of the platform in case it does not have WMO number assigned.

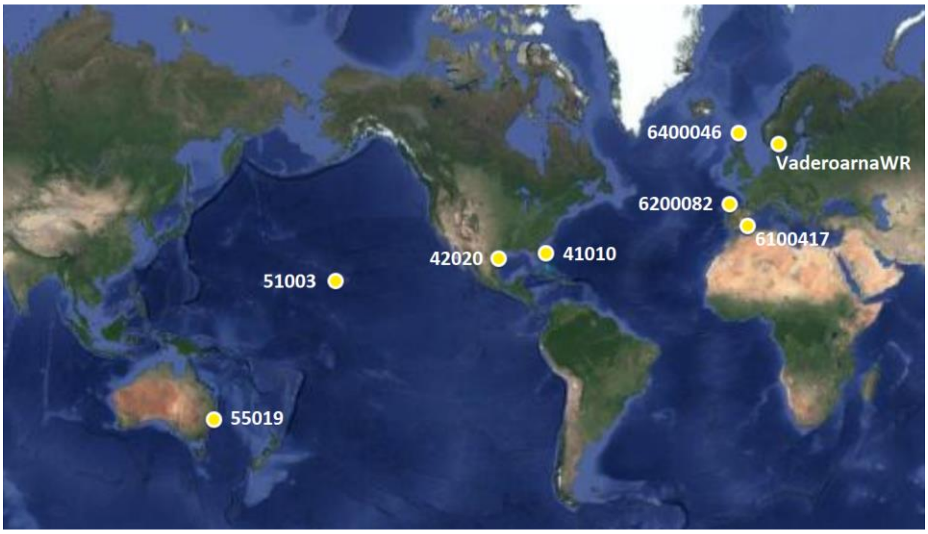


Figure 1. *Location of the platforms selected for the reliability study.*

The study has been conducted as follows: for each platform, the hourly data time series is subsampled to 2, 3, 6, 12 and 24 hours. We proceed to refill every subsampled series with linear interpolation and then, the interpolated series are compared with the original series and usual metrics are computed: Bias, Root Mean Square Error (RMSErr), Absolute Mean Error (AbsMErr), Scatter Index (ScatIndex), Correlation Coefficient (CorrCoef), and Slope and intercept of the linear regression.

The results are very similar for all the stations and the mean values of the metrics are shown Table 3:

Sampling	Bias	RMSErr	AbsMErr	ScatIndex	CorrCoef	Slope	Intercept
2H	0.000	0.09	0.04	0.05	1.00	0.99	0.02
3H	0.000	0.11	0.06	0.06	0.99	0.99	0.02
6H	0.000	0.15	0.09	0.08	0.99	0.97	0.05
12H	-0.002	0.21	0.13	0.13	0.97	0.94	0.10
24H	-0.002	0.33	0.21	0.20	0.93	0.87	0.22

Table 3. *Mean values of metrics from comparison between original series and series subsampled and interpolated linearly for different sampling.*

The results obtained showed very good agreement for samplings less or equal to 6 hours, with a correlation coefficient over 0.99 and a scatter index under 0.1. For samplings over 6 hours (12 and 24 hours) results get worse. Thus, a threshold of 6 hours was chosen as limit for the time sampling for applying the interpolation.

The data from 1252 platforms (72.3% of the total number of platforms in the product at the time of this study) with sampling equal or less than 1 hour are not interpolated except in occasional gaps less or equal than 6 hours. 462 stations (26.7%) with sampling between 1 and 6 hours are filled with linear interpolation of the data, and only 18 (1%) platforms are delivered with not filled time series. All the interpolated data are clearly identified with a quality flag equal to 8 (interpolated value) following **Figure 4**.

II PRODUCTION SYSTEM DESCRIPTION

Production centre : In Situ TAC**.**

Production Units (PU) name: Puertos del Estado, Spain; NOLOGIN CONSULTING, Spain.

Production system name: Global Ocean In Situ Reprocessed Waves (Copernicus Marine Service names: INSITU_GLO_WAV_DISCRETE_MY_013_045)

Description:

The In Situ TAC is a distributed centre organized around 7 oceanographic regions: the global ocean and the 6 EuroGOOS regional alliances (see **Figure 2**). It involves 17 partners from 11 countries in Europe. It doesn't deploy any observing system and relies on data that are obtained exclusively funded by other sources than Copernicus Marine Service.

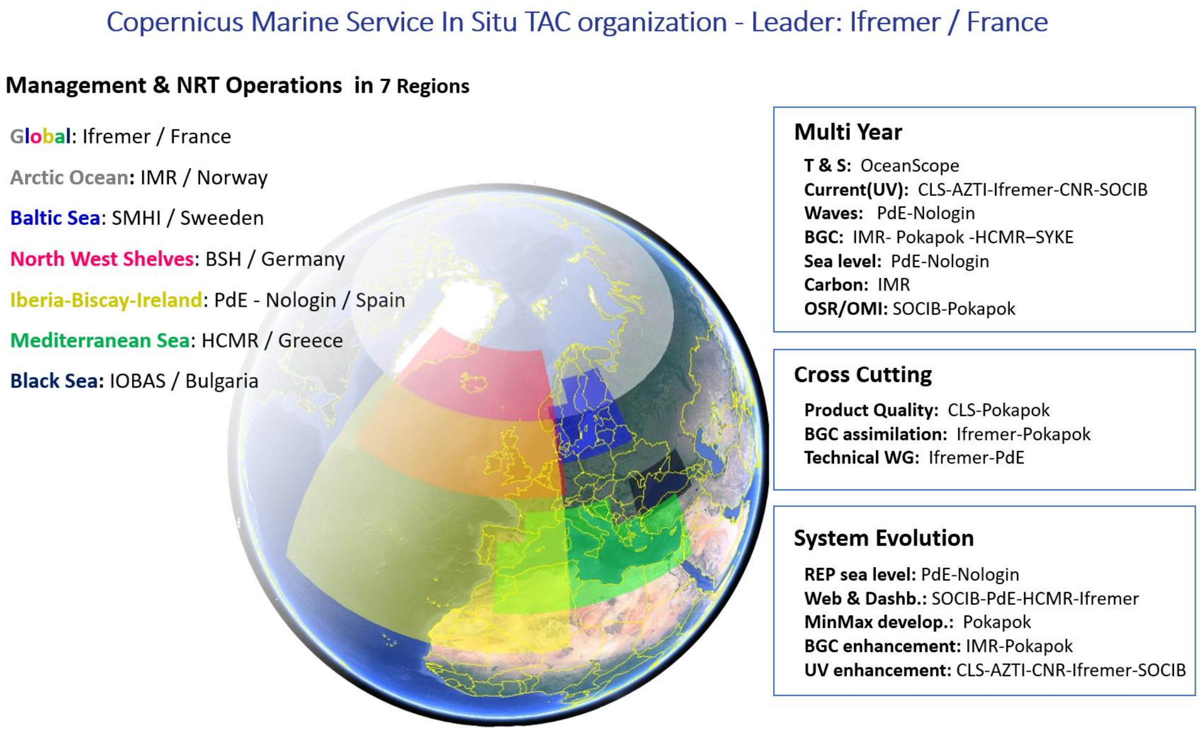


***Figure 2**: The In Situ TAC components.*

The In Situ TAC architecture is decentralized for the Near Real Time (NRT) production, but the production of the reprocessed multiyear products is centralized in a Production Unit composed by one or several institutions in charge of it.

The Global Ocean In Situ Multi-Year WAVE product was created for the first version with the wave observations measured in Near Real Time (frozen copy of Global multiparametric NRT product INSITU_GLO_PHYBGCWAV_DISCRETE_MYNRT_013_030). From this multiparameter NRT product, wave data is extracted, and a thorough validation process is carried out including automatic tests, visualization and comparison between stations and with other sources. For the successive versions, with data extension in time, is produced using the previous version of the product and extended in time with the wave observations measured in Near Real Time product. After the validation process, the NetCDF files are generated and checked, the product is assessed through the benefit assessment metrics describing the coverage and then distributed to the users. **Figure 3** shows the main processing elements for the product.

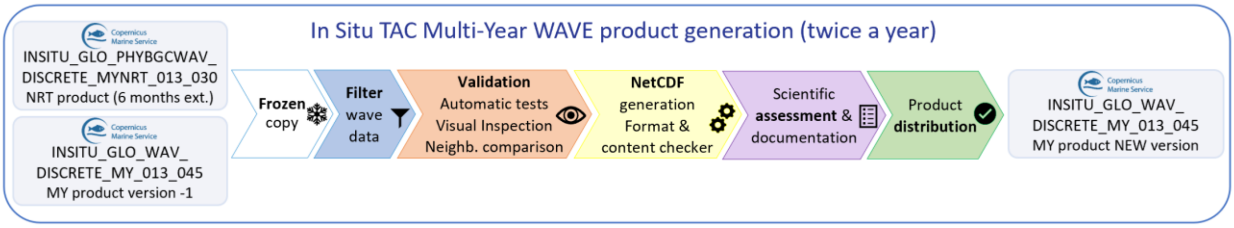


Figure 3: *Scheme of the processing elements of the Multi-Year WAVE product.*

The whole validation and update process for this product is performed twice a year: around June for temporal extension of six months and at the end of the year (November-December) for temporal extension of six months and also several possible modifications or improvements of the product including full reprocessing.

III VALIDATION FRAMEWORK

The In Situ TAC is dedicated to assure the accuracy of in situ observations through two main validation steps: automatic tests and the validation by visual inspection. Then, the assessment of the product is performed. These two validation steps are described in the following sections (III.1, III.2 and III.3).

The assessment performed by data providers is not described in this document because it is different for each platform and variable in the time. Most of the times they are not even documented in the metadata attached to the provided data.

Quality Control (QC) flags are assigned to the data after performing QC tests. The QC flag scale is presented in **Table 4**.

Code	Meaning	Comment
0	No QC was performed	-
1	Good data	All real-time QC tests passed.
2	Probably good data	These data should be used with caution.
3	Bad data that are potentially correctable	These data are not to be used without scientific correction.
4	Bad data	Data have failed one or more of the tests.
5	Value changed	Data may be recovered after transmission error.
6	Value below detection	The level of the measured phenomenon was too small to be quantified/detected by the technique employed to measure it. The accompanying value is the quantification/detection limit for the technique or zero if that value is unknown.
7	Nominal value	Data were not observed but reported. Example: an instrument target depth.
8	Interpolated value	Missing data may be interpolated from neighbouring data in space or time.
9	Missing value	The value is missing, is not reported, is not applicable.

***Table 4**: Quality control flag scale*

III.1 Automatic delayed mode quality checks

In the first validation step, a set of metrics were developed as automatic tests to be applied to the wave parameters. They are illustrated in **Table 5**.

Short description	Applicability of metrics for Time Series
Impossible date	X
Impossible location	X
Position on land	
Global range	X
Regional range	X
Pressure increase	
Spike	X
Stuck value	X
Grey list	
Sensor Drift	
Rate of change	

Table 5: *Metrics used for the quality control of WAVE data.*

The limits of some of these tests are region dependent, like the range limits for significant wave height that are from 0 to 25 meters in the Atlantic Ocean, but it is restricted to 0 to 10 meters for the Mediterranean Sea, an enclosed sea, where the waves cannot develop to get so high values.

These metrics and region dependent limits are described in detail in the document Copernicus In Situ TAC, Real Time Quality Control for WAVES (http://dx.doi.org/10.13155/46607).

The wave parameters are single values for each time stamp, but we have also the wave spectral density, a vector depending on the frequency. For the wave spectral density we cannot apply the tests over the different values of the vector, so the range tests are applied over the total energy of the spectrum (zero order moment, m0), that is, the integral of the spectrum over the whole frequency interval. The significant wave height is obtained as four times the squared root of the zero order moment (WMO, 1998):

$$SWH = Hm0 = 4 * \sqrt{m0} \quad (1)$$

so, operating the formula, we can determine the corresponding limits by the following relationship:

$$m0 = (\frac{Hm0}{4})^2 \quad (2)$$

If we are using 25 meters as the highest limit in range, for the energy m0 the limit will be:

$$(\frac{25}{4})^2 = 39.0625 \text{ m}^2\text{s} \quad (3)$$

III.2 Visual inspection of wave parameter series and scatter plots of wave height-period

For this visual inspection a tool to download and draw the whole time series of wave height and period and the scatter of wave height-period has been developed. This allows to check on one graph possible malfunctions of the measurement system and outliers not detected by the automatic tests.

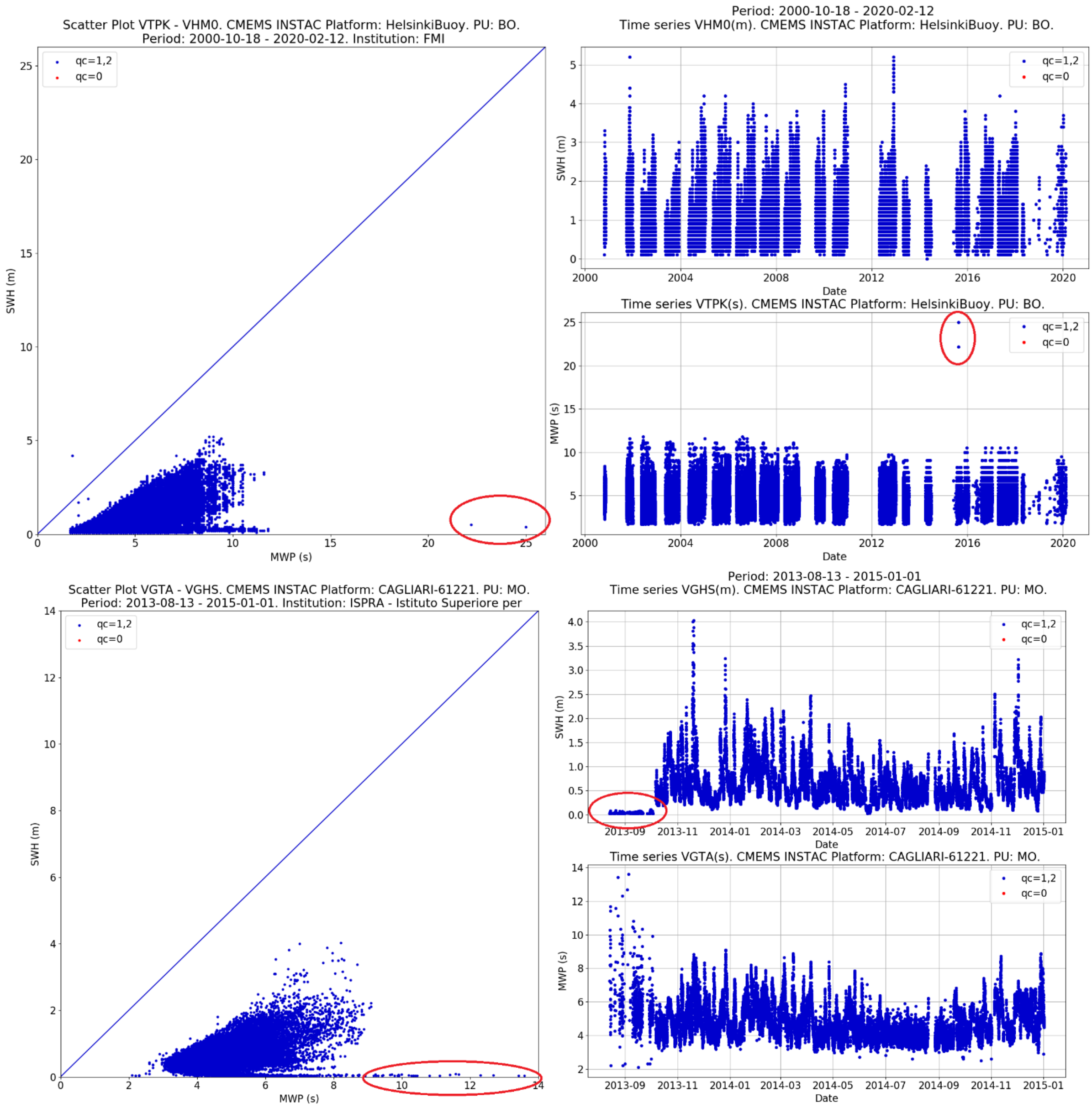


Figure 4: *Example of visual inspection graph with detection of spikes in Helsinki buoy in the Baltic Sea (upper panel) and anomalous data due to a malfunction in the CAGLIARI buoy in the Mediterranean Sea (lower panel). Detections are circled in red.*

Figure 4 shows two examples of outliers detection. In the upper panel two spikes in the peak period are shown detected in Helsinki Buoy measurements in the Baltic Sea and the lower panel shows an anomalous period detected in the platform CAGLIARI measurements in the Mediterranean Sea. In Figure 4, the detections are circled in red. On the left side, scatter plots of wave height-period and on the right side, time series of wave height and wave period where the x axis represents the date. The legend indicates the quality flag of the data.

III.3 Visual inspection of wave spectra

The wave spectra are visualized in a graph showing its evolution in time, in order to check if the spectra are well formed, and to detect wrong spectra.

In **Figure 5** the spectral data are computed in discrete frequencies. In the upper panel the frequencies are variable and in the lower panel are fixed and with low resolution for frequencies higher than 1 Hz. The apparent gaps in the lower panel are due to these fixed frequencies.

Figure 5 (upper figure) shows an example of a timeline for wave spectra collected by platform 6200085 moored in the Gulf of Cadiz (Southwest of Europe) for the period 2018-2021, with no wrong spectrum detected. Note that this platform measures two kinds of wave sea states, swell with energy in frequencies below 0.1 Hz (periods over 10 seconds) and wind sea with energy in frequencies around 0.15 Hz (6-7 seconds).

In **Figure 5** (bottom) the evolution of wave spectra during January-March 2022 is shown for platform 46184 in the West Coast of Canada. A problem in the spectrum (bad data) was detected at the end of March 2022 (circled in red) with high energy for frequencies lower than 0.025 Hz (>40 seconds) and qualified accordingly during the validation process.

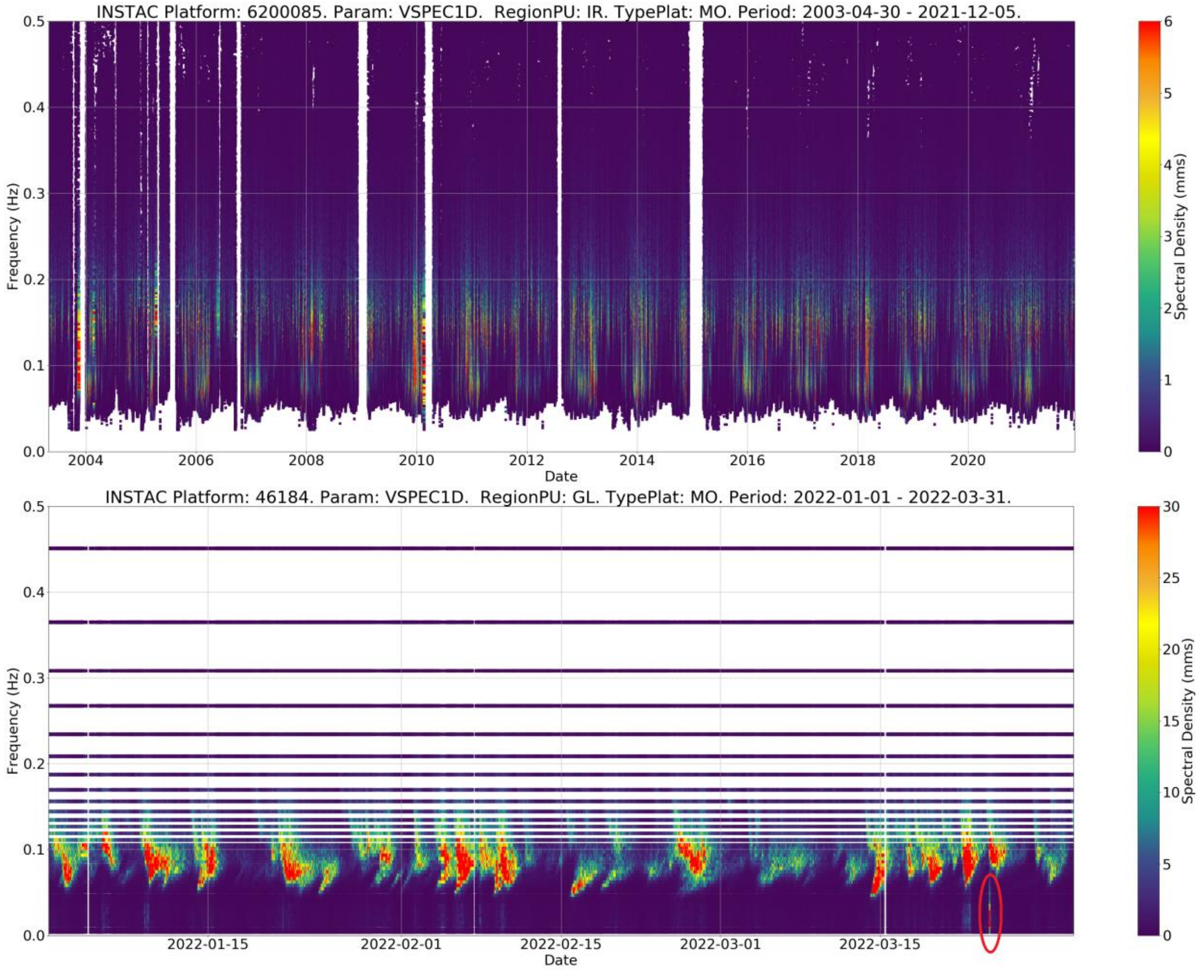


***Figure 5:** Visual inspection of wave spectra for platforms 6200085 (Southwest of Europe) for period: 2018-2021 with no detection (upper panel) and 46184 (West Coast of Canada) for period: January-March 2022. Anomalous spectral data are detected (circled in red) at the end of March 2022 (lower panel).*

IV VALIDATION RESULTS

This section shows the results of the metrics used for the assessment of the product.

IV.1 Temporal coverage of the wave product

This section gives an overview of the wave measurements temporal coverage since 1970.

Figure 6 shows the number of platforms providing wave measurements by years in the whole global ocean and in every Copernicus Marine Service region for wave heights, period, direction and spectra.

Table 6shows the estimated number of platforms making measurements of (i) scalar waves (height and period), (ii) wave direction and (iii) wave spectra during the period 1970-2022 at global scale. The temporal coverage has been divided into five different decades from 1970 up to now (2022). The metric includes the mean number of platforms and the standard deviation.

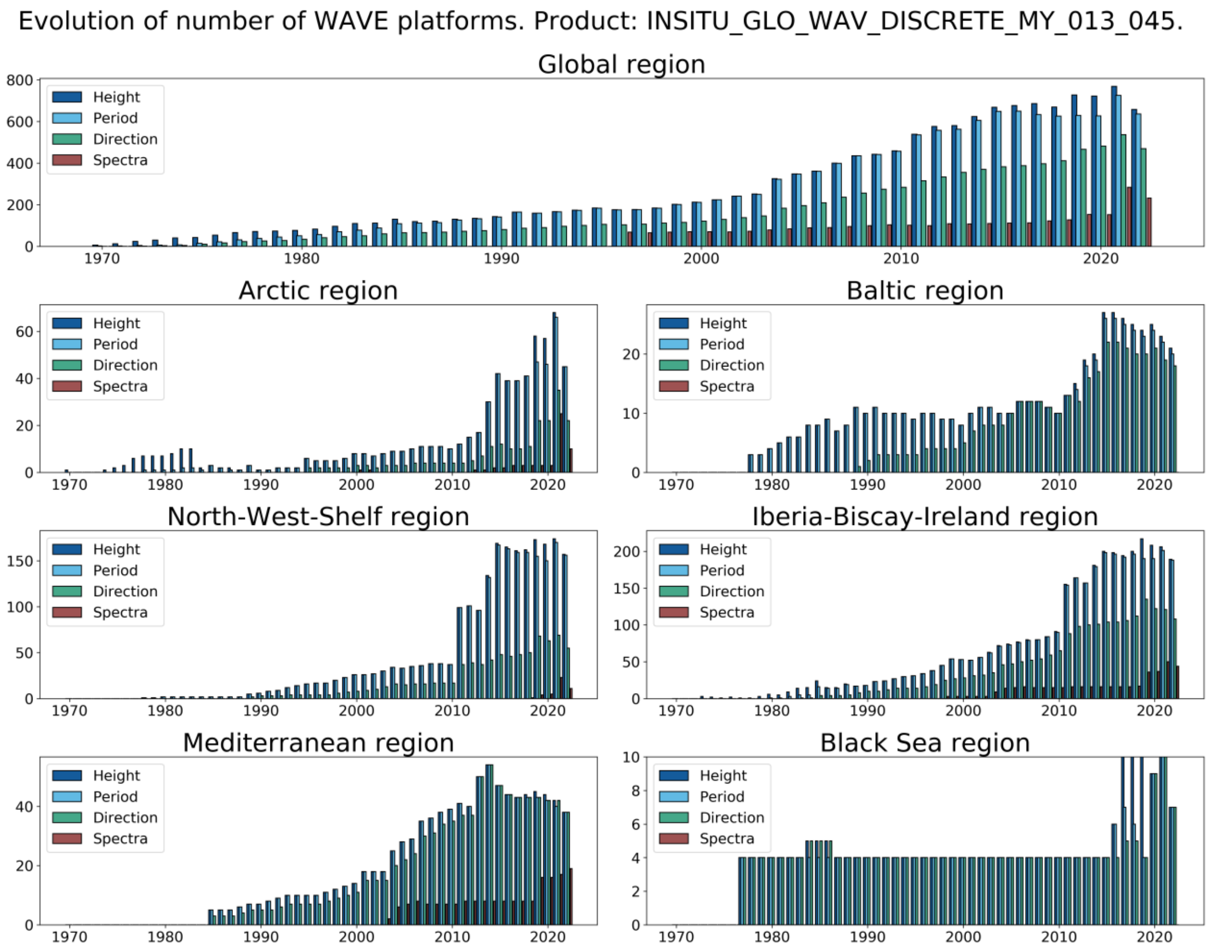


***Figure 6**: Bar chart with the number of platforms providing wave measurements (heights, periods, directions and spectra) per year since 1970 for the whole GLOBAL In Situ TAC and the different regions.*

The number of platforms is low and stable with a slight increasing trend for the first decades (until the 1990's). At the beginning of the 2000 decade, they experience a clear increase mainly due to the incorporation of European historical datasets and the wave networks of US and Canada.

For wave spectra there is no coverage in the first two decades, and the data collection started in the mid 1990's and grew over recent years due to the technical improvements in the measurement and processing methods that have allowed the provision of spectral information.

WAVES (Global)	Coverage (number of platforms)									
	1970-1980		1980-1990		1990-2000		2000-2010		2010-2022	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Height	41.7	22.9	111.0	19.1	173.0	15.2	324.0	82.8	642.8	82.0
Period	16.9	14.9	93.4	27.6	172.1	15.3	323.0	82.8	607.3	63.5
Direction	10.9	10.4	57.7	13.3	99.4	10.6	188.8	52.0	399.1	69.8
Spectra	0.0	0.0	0.0	0.0	27.1	33.2	84.9	11.7	139.9	53.9

Table 6: *Estimated coverage (number of platforms) providing scalar waves (height and period), wave direction and wave spectra during the period 1970-2022 at global scale by decades.*

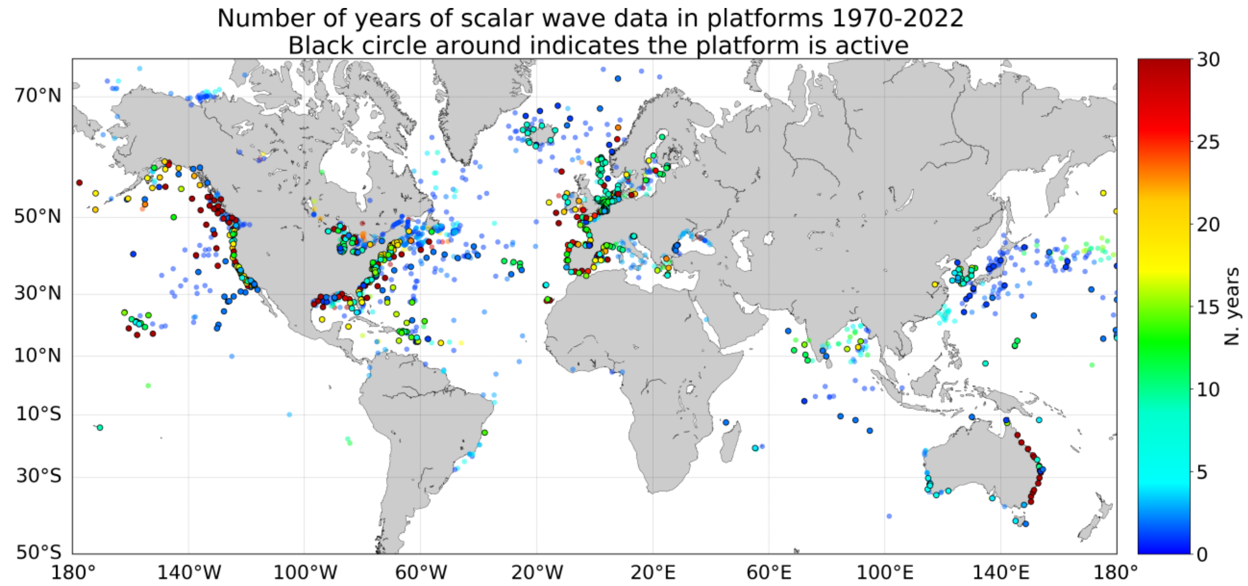
IV.2 Spatial coverage of the wave product

The following figures (Figures 7 to 9) provide a synoptic view of the spatial coverage of the wave product within In Situ TAC, where the colour signifies the number of years covered by the station. A coloured point with a black outline indicates that the platform is active (data provided in the last two months of the validation period; in this case, November and December 2022).

Figure 7 shows the distribution of platforms providing scalar waves (heights and periods) in the global ocean (upper figure) and for European Seas (lower figure). Most of the platforms are concentrated in the Northern Hemisphere and especially along the coasts of Europe, North America, Japan, Korea and India. In the Southern Hemisphere there are some stations in South America and Australia. In European Seas the coverage is high except than in the Arctic, Southern and Eastern Mediterranean and Southern Black Sea areas.

Figure 8Figure 8 shows the distribution of platforms providing directional waves. Most of the platforms are in the Northern Hemisphere and especially in Europe and North America. In the Southern Hemisphere there are some platforms in Australia. In European Seas the coverage is high, except in the Arctic, Southern and Eastern Mediterranean and Black Sea areas. The patterns are similar to the ones found **Figure 7**, but the coverage is noticeable less in some areas (West Pacific, North Indian and Northeast Atlantic Oceans). These differences are explained by the absence of directional wave sensor in those areas. A similar situation is observed with the British network and some platforms in the North Sea and in the US coast.

Figure 9 represents the distribution of platforms providing wave spectra. Only Spanish, French, US and Canadian networks provide this information, as the French and the Canadian ones are very recent (less than 10 years). The recent merge of the spectral information from drifters and saildrones has allowed the availability of spectra in open waters.



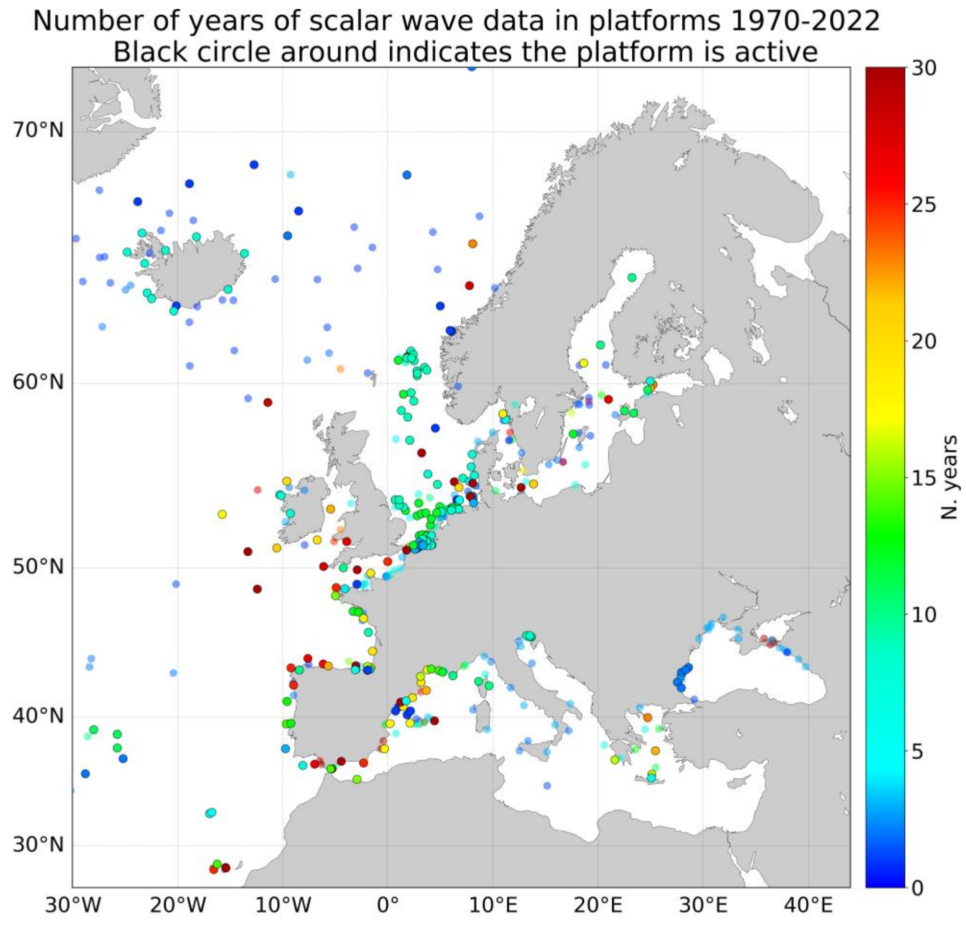


***Figure 7**: Map with the spatial distribution of platforms providing scalar waves in the global ocean (upper panel) and European Seas (lower panel). The colour signifies number of years covered by the station. A coloured point with a black outline indicates that the platform is active (data provided in the last two months of the validation period).*

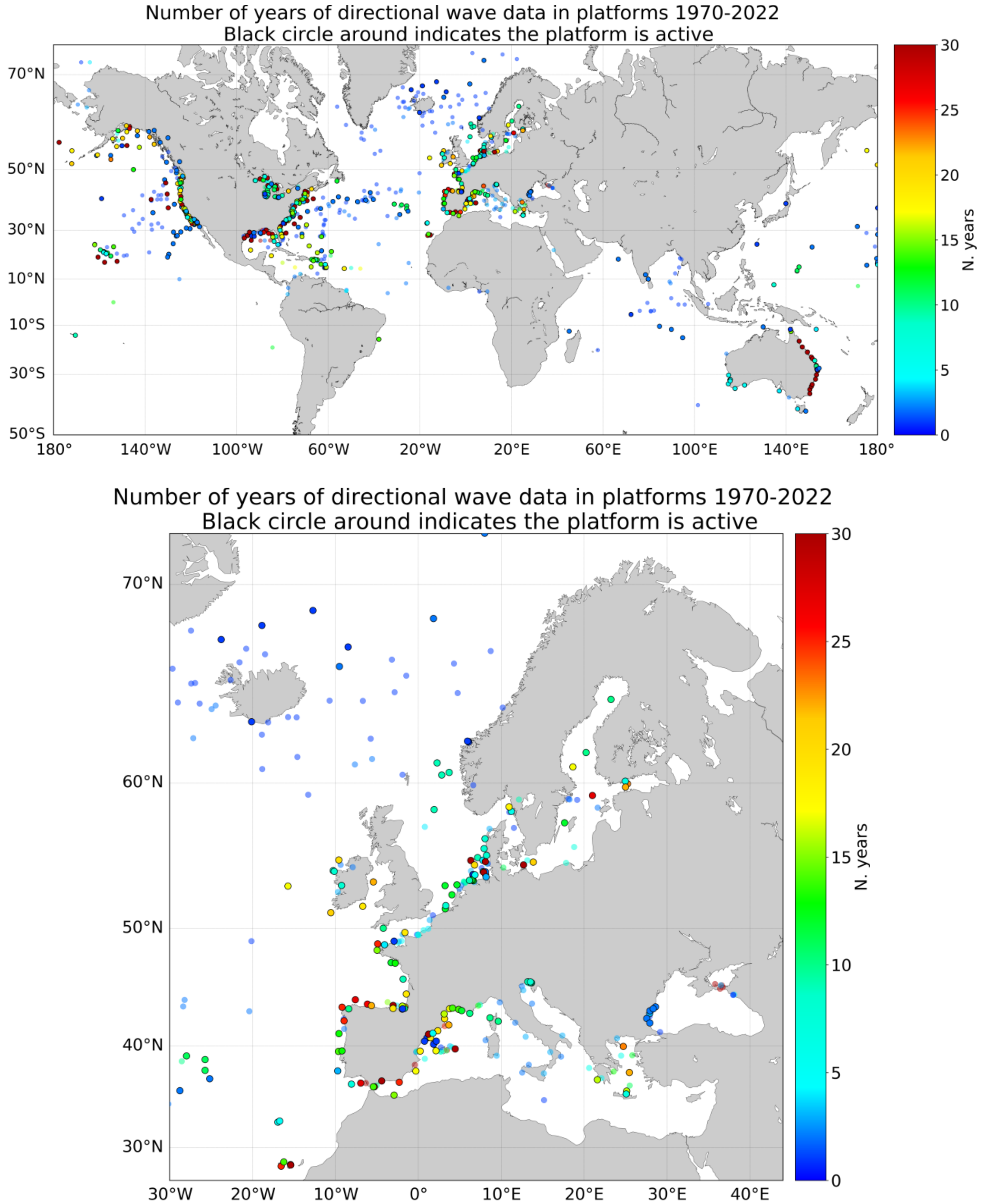
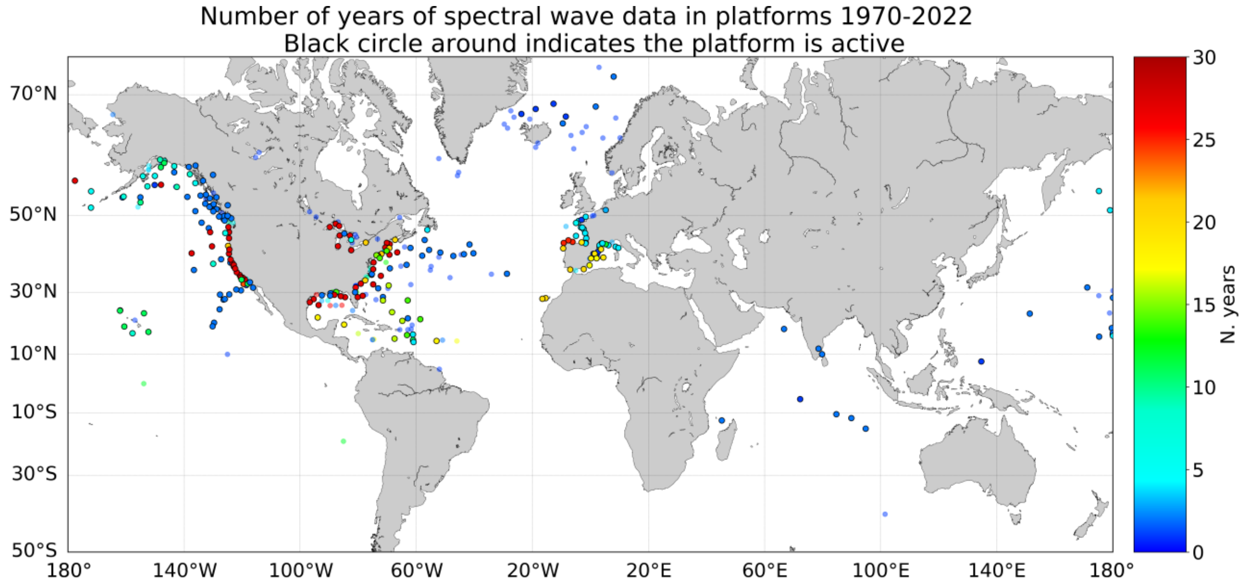


Figure 8: *Map with the spatial distribution of platforms providing directional waves in the global ocean (upper panel) and European Seas (lower panel). The colour signifies number of years covered by the station. A coloured point with a black outline indicates that the platform is active (data provided in the last two months of the validation period).*



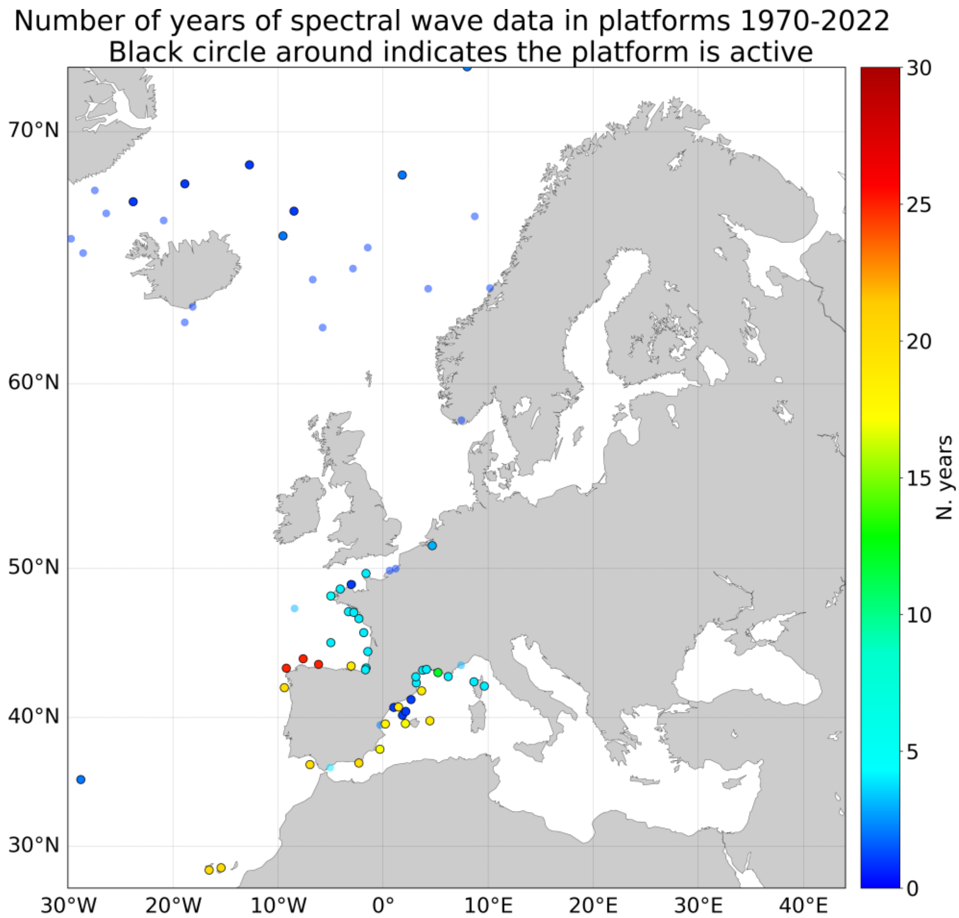


Figure 9: *Map with the spatial distribution of platforms providing wave spectra in the global ocean (upper panel) and European Seas (lower panel). The colour signifies number of years covered by the station. A coloured point with a black outline indicates that the platform is active (data provided in the last two months of the validation period).*

IV.3 Information on the quality of the data

An analysis of observations flagged as 'good data' after the validation process has been performed. The results shows that the percentage of observations that have been flagged as 'good data' is very high: 98.61%.

V SYSTEM'S NOTICEABLE EVENTS, OUTAGES OR CHANGES

Nothing to mention

VI QUALITY CHANGES SINCE PREVIOUS VERSION

The validation and assessment of this product have been performed in a distributed way among regions since the product launch. Since mid-2022, the process is centralized in a Production Unit, so, even when the validation procedures and the assessment metrics remain the same, all the processes have been centralized.

VII REFERENCES

Copernicus Marine In Situ Team. Copernicus In Situ TAC, Real Time Quality Control for waves. CMEMS-INS-WAVES-RTQC. https://dx.doi.org/10.13155/46607

Copernicus Marine In Situ Tac Data Management Team. Copernicus Marine In Situ TAC - physical parameters list. https://doi.org/10.13155/53381

WMO, 1998. GUIDE TO WAVE ANALYSIS AND FORECASTING. Secretariat of the World Meteorological Organization, Geneva, Switzerland. 1998. WMO-No. 702. https://www.wmo.int/pages/prog/amp/mmop/documents/WMO%20No%20702/WMO702.pdf.