For REProcessed IN SITU product (WAVES):
INSITU_GLO_WAV_REP_OBSERVATIONS_013_045

http://doi.org/10.13155/53380

Issue: 2.0

Contributors: Marta de Alfonso, Fernando Manzano Muñoz

Approval Date by Quality Assurance Review Group:
## CHANGE RECORD

<table>
<thead>
<tr>
<th>Issue</th>
<th>Date</th>
<th>§</th>
<th>Description of Change</th>
<th>Author</th>
<th>Checked By</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>10/12/2017</td>
<td>all</td>
<td>First version of document</td>
<td>Marta de Alfonso</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fernando Manzano Muñoz</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>10/01/2018</td>
<td>all</td>
<td>Modifications with the final configuration of the INSTAC WAVES REP product</td>
<td>Marta de Alfonso</td>
<td></td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

## I Executive summary

1.1 Products covered by this document

1.2 Summary of the results

1.3 Estimated Accuracy Numbers

## II Production Subsystem description

## III Validation framework

## IV Validation results

IV.1 Coverage in time of the WAVES product

IV.2 Coverage in space of the WAVES product

IV.3 Information on the quality of the data

## V References
I EXECUTIVE SUMMARY

I.1 Products covered by this document

The document describes the Quality of the Delayed Mode (REP) WAVES product delivered by the CMEMS INS TAC.

The following document applies to the following list of products described in CMEMS Catalogue

<table>
<thead>
<tr>
<th>Short Description</th>
<th>Product code</th>
<th>Area</th>
<th>Delivery Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLOBAL REP</td>
<td>INSITU_GLO_WAV_REP_OBSERVATIONS_013_045</td>
<td>GLOBAL</td>
<td>Yearly</td>
</tr>
</tbody>
</table>

Table 1: List of INS TAC products for which this document applies.

These products integrate observation 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 Centers (NODCs) and JCOMM global systems (Argo, GOSUD, OceanSITES, GTSP, DBCP) and the Global telecommunication system (GTS) used by the Met Offices.

Data are distributed on full level (no interpolation). They are available in a dedicated directory to waves (INSITU_GLO_WAV_REP_OBSERVATIONS_013_045) of GLOBAL Distribution Unit in one file per platform. This directory is updated once a year.

Four of the INSTAC regions (BAL, NWS, IBI and MED) are performing their own wave validation over the data contained in their respective files (“BO_”, “NO_”, “IR_” and “MO_”). These REP files are distributed also in the corresponding REP regional directories. The GLOBAL Distribution Unit gets these files together with the validated “GL_” files and distributes the final WAVES product in the GLOBAL dedicated directory mentioned above.

I.2 Summary of the results

The wave observations are aggregated by the In Situ Thematic Assembly Center and the data set is provided to users together with metadata information on the platforms that were used to perform the observations. The quality of the observation is tested using 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 (see Table 4).

The In Situ TAC relies on observing systems maintained by institutes that are not part of the In Situ TAC and CMEMS service is not contributing to the maintenance and setting up of the observing systems it uses.

- The platforms mostly used to measure waves are moorings (buoys and lightvessels). There are deep water platforms but also coastal stations that are affected by local bathymetry and coastal processes, so it should be considered.
In some regions the number of available platforms is on a critical low level to provide an adequate representative overall view of the state of the ocean. Some of the areas are clearly undersampled and in some others data is not available. The In Situ TAC is dedicating a great effort to gather all the wave observations that will continue in the following months for both operational stations and historical data sets.

The percentage of data flagged as ‘good data’ is quite high (over 95%) and varies slightly from region to region.

The temporal coverage of wave measurements starts with a low and stable number of platforms in the early 90’s. At the beginning of the second decade (about 2002) it starts to grow and increase a lot during the last two years (2016-2017).

### I.3 Estimated Accuracy Numbers

The following table summarizes the accuracy of the measurements that can be expected depending on the sensors. This is the best accuracy then a user can expect for the in situ data to which a quality flag “Good data” (see Table 4) has been applied after validation process.

The definition of the reference values is obtained from different sources. The specific reference is given in the tables below and the values are given for the different parameters.

<table>
<thead>
<tr>
<th>Wave sensor</th>
<th>Measured time series</th>
<th>Estimated parameters (due to the statistical variability)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vertical displacement (heave)</td>
<td>Period</td>
</tr>
<tr>
<td>Waverider</td>
<td>0.5% of the measured value</td>
<td>0.5% of the measured value</td>
</tr>
<tr>
<td>(Datawell)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavesense</td>
<td>0.1 m</td>
<td>0.15 s</td>
</tr>
<tr>
<td>(Oceanor)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triaxys (Axys)</td>
<td>1% of the measured value</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All wave sensors</td>
<td>&lt; 5% of the estimated value</td>
<td>&lt; 5% of the estimated value</td>
</tr>
</tbody>
</table>

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

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.
II PRODUCTION SUBSYSTEM DESCRIPTION

The INS-TAC is a distributed centre organized around 7 oceanographic regions: the global ocean and the 6 EUROGOOS regional alliances (see Figure 1). It involves 15 partners from 10 countries in Europe. It doesn’t deploy any observing system and relies on data that are obtained exclusively funded by other sources than CMEMS.

![Figure 1: The INS TAC components](image)

© EU Copernicus Marine Service – Public
The in-situ TAC architecture is decentralized. However, quality of the products delivered to users must be equivalent wherever the data are processed. The different functions implemented by the global and regional components of the In Situ TAC are summarized in Figure 2.

![Figure 2: Functions implemented by an in situ TAC component](image)

Each region implements the 4 core functions:

- **Data Acquisition**: Gather data available on international networks or through collaboration with regional partners.
- **Data Quality control**: Apply automatic quality controls that have been agreed at the In Situ TAC level. These procedures are defined by parameter, elaborated in coherence with international agreement, in particular SeaDataNet, and documented in CMEMS Catalogue.
- **Product validation**: Assess the consistency of the data over a period of time and an area to detect data that are not coherent with their neighbours but could not be detected by automatic QC.
- **Product distribution**: Make the data available within CMEMS and to the external users.

Each region has organized the activities according to the expertise and background in data management for operational oceanography.

1. The 4 functions are implemented in one institute per region.
2. Acquisition and QC is done by platforms and one institute takes care of the validation and distribution is centralized.

In any case, the Global component of the In Situ TAC collects the data from the regional components and integrates them into the global product acting as a backup of the regional centres. The main distribution channel for the INS TAC is authenticated FTP. OGC viewing service (WMS) and SUBS Subsetter access are developed within CMEMS.
III VALIDATION FRAMEWORK

The In Situ TAC is in contrast to the MFCs dedicated to assure the accuracy of in situ observations through mainly two validation channels. These two channels consist of the real time quality control (RTQC) of the in situ observations and the validation in delayed mode and assessment of the product out of the quality controlled data sets.

The assessment performed by 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.

For the first channel, a set of metrics were developed, and it is illustrated in Table 3. These metrics are described in detail in the document for the Real Time Quality Control for Waves: Copernicus Marine In Situ Tac Data Management Team (2016).

By performing the QC tests, the QC flags proposed by the method are assigned to the data. The QC flag scale is presented in Table 4.

<table>
<thead>
<tr>
<th>Short description</th>
<th>Applicability of metrics for Time Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impossible date</td>
<td>X</td>
</tr>
<tr>
<td>Impossible location</td>
<td>X</td>
</tr>
<tr>
<td>Position on land</td>
<td></td>
</tr>
<tr>
<td>Global range</td>
<td>X</td>
</tr>
<tr>
<td>Regional range</td>
<td>X</td>
</tr>
<tr>
<td>Pressure increase</td>
<td></td>
</tr>
<tr>
<td>Spike</td>
<td>X</td>
</tr>
<tr>
<td>Stuck value</td>
<td>X</td>
</tr>
<tr>
<td>Grey list</td>
<td></td>
</tr>
<tr>
<td>Sensor Drift</td>
<td></td>
</tr>
<tr>
<td>Rate of change</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 3: Metrics used for the quality control of WAVE data
For REProcessed IN SITU product (WAVES):
INSITU_GLO_WAV_REP_OBSERVATIONS_013_045

Ref : CMEMS-INS-QUID-013-045
Date : 10/01/2018
Issue : 2.0

<table>
<thead>
<tr>
<th>Code</th>
<th>Meaning</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No QC was performed</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>Good data</td>
<td>All real-time QC tests passed.</td>
</tr>
<tr>
<td>2</td>
<td>Probably good data</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Bad data that are potentially correctable</td>
<td>These data are not to be used without scientific correction.</td>
</tr>
<tr>
<td>4</td>
<td>Bad data</td>
<td>Data have failed one or more of the tests.</td>
</tr>
<tr>
<td>5</td>
<td>Value changed</td>
<td>Data may be recovered after transmission error.</td>
</tr>
<tr>
<td>6</td>
<td>Not used</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>Not used</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>Interpolated value</td>
<td>Missing data may be interpolated from neighbouring data in space or time.</td>
</tr>
<tr>
<td>9</td>
<td>Missing value</td>
<td>-</td>
</tr>
</tbody>
</table>

*Table 4: Quality control flag scale*

The second channel consists of a series of additional procedures that have been performed on the REP product and are listed below:

• **Automatic delayed mode quality checks.**

Some data is recovered in delayed mode (not transmitted in real time). The checks performed are the same as in real time (Table 3). It allows a better detection of spikes due to it’s not performed over the last value received, but we have the previous and the next value.

• **Visual inspection of wave parameter series and scatter plots of wave height-period.**

For this visual inspection a new tool to download and draw the whole series and the scatter of wave height-period has been developed. It allows to check in one graph possible malfunctions and outliers.

Figure 3 shows an example of wrong data detection with this tool graphs in a platform in the MED Sea (platform: 61281) (upper figure) and the same data after the validation process (lower figure).
Figure 3: Example of visual inspection graph with detection of wrong data due to a malfunction in buoy 61281 in the MED region (upper figure) and the same data after data validation (lower figure).

- **Comparison with nearby stations.**

To assess the consistency of the data, we have performed comparisons between nearby stations. This process should be done carefully, taking into account only those platforms affected by the same wave conditions and filtering the data to obtain a match up data set for comparison.

We are going to show two examples of the comparisons performed. Figure 4 shows the position of the platforms used for both examples.
In the first comparison (platforms 6201030-62024), both stations are very close, located in deep water and are affected by the same wave conditions, so we expect to obtain very good results in the comparison. In the second case (platforms: 62105-62093) stations are a bit far away and the outer one is in more open water conditions, so the results are not expected to be so good.

We show in Figure 5 time series, scatter plot and quantile-quantile plot (qq plot) of significant wave height (Hm0 estimator) for the examples of nearby buoys in the IBI region. Both the scatter and the qq plots show very good agreement, mainly in the first case with very low dispersion.

Table 5 shows the results of metrics from the comparison of nearby stations. In both cases, buoys are owned by different providers and the wave analysis could be different and wave sensors are different. Even though, the results in the first case are excellent with correlation of 0.97, very low
bias (close to zero), and low root mean square difference (RMSD) and scatter index (SI). In the second case, even when we expect not so good results, the correlation is over 0.90, so we can check the consistency of the data.

<table>
<thead>
<tr>
<th>MOORING Buoys</th>
<th>N</th>
<th>BIAS</th>
<th>RMSD</th>
<th>SI</th>
<th>LR SLOPE</th>
<th>LR OFFSET</th>
<th>CORR</th>
</tr>
</thead>
<tbody>
<tr>
<td>6201030-62024</td>
<td>4172</td>
<td>-0.071</td>
<td>0.3423</td>
<td>0.1578</td>
<td>0.9529</td>
<td>0.0311</td>
<td>0.9710</td>
</tr>
<tr>
<td>62093-62105</td>
<td>2542</td>
<td>0.301</td>
<td>0.7419</td>
<td>0.2429</td>
<td>0.9428</td>
<td>0.4762</td>
<td>0.9338</td>
</tr>
</tbody>
</table>

*Table 5: Metrics from comparison of neighbour stations*

- **Comparison with model reanalysis**

Whenever possible, it is advisable to participate in the model reanalysis validation process with in situ data. It allows to detect other kind of malfunctions impossible to see looking only at the data and also permits to acquire perspective of the consistency of the data in the region.

In the IBI region the validation carried out by the IBI-MFC has been made in coordination of the IBI INSTAC and it has been very fruitful for both CMEMS components. The validation has been very exhaustive and can be fully consulted in the IBI-MFC QUID for waves (Sotillo et al, 2017: [http://marine.copernicus.eu/documents/QUID/CMEMS-IBI-QUID-005-005.pdf](http://marine.copernicus.eu/documents/QUID/CMEMS-IBI-QUID-005-005.pdf)).

In this document we are showing only some general results to give and idea of the validation process carried out. Figure 6 shows the location of the INSTAC platforms used for the IBI-MFC WAV model validation and all the subareas in which the general domain has been divided.

![Figure 6: Location of the coastal (orange dots) and deep waters (green dots) buoys employed to conduct the IBI qualification during a two-year period (2014-2015). The IBI service domain (IBISR, red rectangle) has been split into different sub-regions of particular interest, which are denoted with rectangles of different colours.](image)

© EU Copernicus Marine Service – Public
Figure 7 reveals the spatial distribution of the statistical metrics obtained for each buoy employed. Every dot represents the value of bias (left) and root mean squared difference (RMSD, right) obtained for the period comprised between the first of January 2014 and the 31st of December 2015. As it can be observed, such metrics mostly emerged in the ranges [-0.2 m, 0.2 m] and [0m, 0.3 m], respectively.

Likewise, Figure 8 shows the distribution of the correlation index (left) and the scatter index (right) for the period, lying predominantly in the ranges [0.88-1] and [0-0.35], respectively.

Figure 7: Maps of skill metrics derived from the comparison of the significant wave height provided by the IBI-WAV product and the observations from the mooring buoys in INSTAC: bias (left) and root mean squared difference or error (right) obtained for an entire two-year period (2014-2015).
Figure 8: Maps of skill metrics derived from the comparison of the significant wave height provided by IBI-WAV product and the observations from the mooring buoys: correlation index (left) and scatter index (right) obtained for an entire two-year period (2014-2015).

A summary of the regional statistical results of the significant wave height (SWH) derived from the validation exercise is provided in Table 6. Metrics are gathered for the whole IBI service domain (IBISR). Likewise, metrics are computed using only coastal (CO) and Deep-water (DW) mooring buoys.

<table>
<thead>
<tr>
<th>REGION: IBISR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOORING BUOYS</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>COSTAL</td>
</tr>
<tr>
<td>DEEP (35)</td>
</tr>
<tr>
<td>TOTAL</td>
</tr>
</tbody>
</table>

Table 6: Summary of skill metrics obtained from the comparison of SWH: IBI-WAV products against the coastal and deep-water mooring buoys in IBI INSTAC for the period 2014-2015. Metrics averaged for the entire IBI Service Area (IBISR). N, RMSD, SI, CORR, LR_SLOPE and LR_OFFSET represent, respectively, the number of data analysed, the root mean squared error, the scatter index, the correlation index and the results of the best linear fit.
Finally, we show in Figure 9 the Taylor diagram for deep water stations (circles) and coastal stations (stars). This diagram resumes the general results found in the validation process. It can be appreciated the good results of deep water stations with correlation over 0.95 and standard deviation below 0.4 for most of them and good results but no so much for coastal stations. In the case of coastal stations, waves are affected by the bathymetry and the coastline and suffer several transformations. The model output is chosen in the nearest gridpoint to the mooring and the position can differ several kilometers. In open waters this difference is not important due to the wave field is very uniform, but close to the coast, it can affect the results.

![Taylor Diagram](image)

**Figure 9.** Taylor diagram for of IBI-WAV model product against observational data from buoys in IBI INSTAC for the 2014-2015 period.
IV VALIDATION RESULTS

IV.1 Coverage in time of the WAVES product

The following figures give an overview of the number of WAVE measurements since 1990.

Figure 10 shows the measurements in the whole GLOBAL ocean. The number is low and stable for the first decade (90’s), at the beginning of the second decade (about 2002) it starts to grow and increase a lot during the last two years (2016-2017). This increase is due mainly to the incorporation of the wave networks in USA and Canada.

Figure 10: Bar diagram providing the number of WAVE measurements per month since 1990 for the whole GLOBAL INSTAC.

In the following figure (Figure 11) appears similar bar diagrams for the different regions. It’s clear that for ARCTIC and Black Sea, the number of measurements is low and sparse, with very few platforms. For the BALTIC Sea, the number is stable during the first half and grows a bit in the last fifteen years. The rest of the regions (NWS, IBI and MED) show similar trends with low rate of observations during the 90’s and a high and progressive increase from 2000.
For REProcessed IN SITU product (WAVES):

INSITU_GLO_WAV_REP_OBSERVATIONS_013_045

Ref : CMEMS-INS-QUID-013-045
Date : 10/01/2018
Issue : 2.0
Figure 11: Bar diagram providing the number of WAVE measurements per month since 1990 for the different regions (ARCTIC, BALTIC, NWS, IBI, MED and Black Sea).
IV.2 Coverage in space of the WAVES product

The following figures (Figure 12 and Figure 13) provide a synoptic view of the coverage in space of the WAVES product within INSTAC.

![Map of INSTAC platforms providing wave information in European Seas.](image)

*Figure 12. Map of INSTAC platforms providing wave information in European Seas.*

The European Seas are covered in different levels depending on the region: IBI, BAL and NWS are uniformly covered whereas ARC and MED have an important lack of data in some areas like the Norwegian coast, the North of Africa and the Italian coast. Only one station is currently working in the Black Sea.
Figure 13. Map of INSTAC platforms providing wave information at global level.

At global level out of European Seas, North America is well covered thanks to the buoy national networks deployed in USA and Canada. The rest of the world have spare moored stations in the Pacific Ocean and some others in South America.

Finally, in Figure 14 we show the number of platforms in the GLOBAL region compared with the rest of parameters. Waves are in blue colour with a maximum number of 402 platforms.

Figure 14: Number of platforms available daily in the GLOBAL per parameter at the end of 2017.
IV.3 Information on the quality of the data

These diagrams give a synoptic view on the quality of the data available in the Historical product per region.

The output data quality information about waves provided by the GLOBAL INSTAC is displayed in the Figure 15, providing the percentage of observations flagged with a good data quality flag.

Similar graphs appear in Figure 16 for each region.

![GLOBAL INSTAC](image)

*Figure 15. Percentage of observations flagged with the ‘good’ data flag for GLOBAL INSTAC.*
Figure 16. Percentage of observations flagged with the ‘good’ data flag for every region. Information is provided for significant wave height parameter.
V REFERENCES

Copernicus Marine In Situ Tac Data Management Team (2016). Copernicus In Situ TAC, Real Time Quality Control for WAVES. http://doi.org/10.13155/46607.