



sea state
cci

Product Validation Plan (PVP)

version 1.1, 25 November 2019

| | | |
|---|---|---|
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Contents

| | |
|---|-----------|
| List of Acronyms | 4 |
| 1. Introduction | 5 |
| 2. Validation methodology / diagnoses | 6 |
| 2.1 Triple Collocation Technique (TCT) | 6 |
| 2.1.1 Introduction to TCT | 6 |
| 2.1.2 Derivation of TCT | 8 |
| 2.1.3 Suggested Procedure | 10 |
| 2.1.4 Practical Considerations | 11 |
| 2.2 Satellite to validation datasets comparison metrics | 17 |
| 3. Validation data | 17 |
| 2.1 In situ data | 17 |
| 2.2 Wave enabled drifters | 19 |
| 2.3 Microseism data | 19 |
| 2.4 Wave model data | 21 |
| 2.5 Independent satellite measurements | 22 |
| 4. Validation in the Round Robin process | 22 |
| 5. ECV product validation and User assessment | 24 |
| 6. References | 25 |

| Author | Approved | Signature | Date |
|-----------------------------------|-------------------------------|---|------------|
| Fabrice Collard, Saleh Abdalla | Fabrice Ardhuin, Ellis Ash |  | 25/11/2019 |
| ESA Acceptance | | | |

| Issue | Date | Comments |
|-------|------------------|--|
| 1.0 | 17 August 2019 | First version for ESA approval |
| 1.1 | 25 November 2019 | Update to address comments from ESA review |

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 |
| ECMWF | European Centre for Medium-range Weather Forecasts |
| ECV | Essential Climate Variable |
| FFT | Fast Fourier Transform |
| GDP | Global Drifter Programme |
| GDR | Geophysical Data Record |
| GPS | Global Positioning System |
| GTS | Global Telecommunication System |
| IFS | Integrated Forecasting System |
| L4 | Level 4 |
| LRM | Low Rate Measurement |
| LUT | Look-Up Table |
| MAD | Median Absolute Deviation |
| MERIS | Medium Resolution Imaging Spectrometer |
| MLE | Maximum Likelihood Estimator |
| NWP | Numerical Weather Prediction |
| OSTST | Ocean Surface Topography Science Team |
| PHCP | Percentage of High Correlation |
| PLRM | Pseudo Low Rate Measurement |
| PTR | Point Target Response |
| PVP | Product Validation Plan |
| 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 |
| TCT | Triple Collocation Technique |
| w.r.t | with respect to |
| WV | Wave (mode for SAR) |

1. Introduction

This document presents the Product Validation Plan (PVP) for **Sea_State_cci**, deliverable 2.5 of the project. It describes the validation approach, and lists reference data sets to be used in the validation of the Sea State Climate Research Data Product.

Further sections of the document are structured as follows:

Section 2 describes validation methodology focusing on the Triple Collocation Technique.

Section 3 presents the sources of validation data.

Section 4 describes the validation approach for the altimetry Round Robin process.

Section 5 refers to the activity of validation and User Assessment and its reporting.

A further report, the In Situ Database Report, will be prepared in Autumn 2019 and will elaborate on section 3 and present details of available data.

2. Validation methodology / diagnoses

As no Reference wave measurements can be considered error free, we propose, in the frame of the Sea State ECV to use the triple collocation technique that allows to evaluate the ECV uncertainty, even in the presence of reference measurement error. Additionally, the validation will be performed by independent intercomparison of the satellite measurements to all individual collocated validation datasets.

2.1 Triple Collocation Technique (TCT)

2.1.1 Introduction to TCT

Geophysical measurements can be verified by comparing against a reference such as in-situ (also termed here as buoy) observations and the output of Numerical Weather Prediction (NWP) models like the Integrated Forecasting System (IFS) of European Centre for Medium-Range Weather Forecasts (ECMWF). However, this type of verification does not provide proper error estimates for use in various applications like data assimilation which is very important for the success of weather forecasting. In-situ sea-state measurements are often considered as the proper reference for verification. Irrespective of the accuracy of this statement, their limited availability at few locations mainly around the European and American coasts makes any verification to be of limited applicability. The global NWP models, on the other hand, provide an attractive alternative for verification due to their global coverage. It is important to stress that both in-situ observations and model prediction contain errors.

A “measurement error” can be defined as the deviation of the measurement from the truth. Unless under controlled conditions, the geophysical truth is usually unknown and hence the error. There are several causes of errors some are related to the instrument used for the measurement (hereafter the term “instrument” is used to refer to a measuring device, a prediction model or any measurement tool) like inaccuracy, noise and deficiencies in the principle of measurements and some due to ambient conditions which are not accounted for. Other sources of errors like human errors or communication errors are difficult to deal with and are ignored here. Estimating errors in individual measurements is usually very close to impossible due to the randomness factor they involve. However, statistical description of errors is possible.

If a given truth was to be measured using the same instrument a large number of times we end up with a probability distribution typically similar to that given in *Figure 1* (the shape of the distribution is irrelevant to the discussion hereafter). The errors can be classified into two categories:

1. bias error which is a systematic error that defines the accuracy of the instrument, and
2. random error which defines the precision of the instrument.

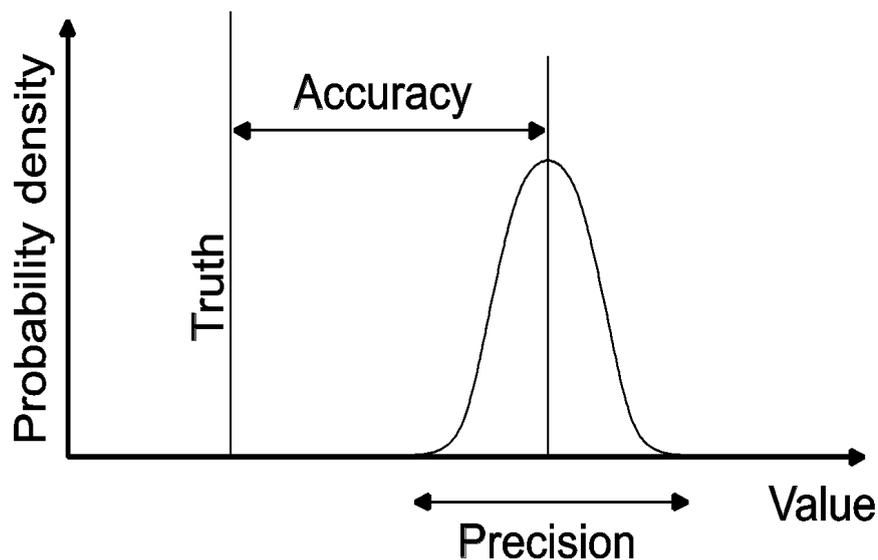


Figure 1: Typical error distribution (the shape of the distribution is irrelevant).

Quantifying the “absolute bias” of a measurement is not possible due to the absence of a standard reference. Buoys or other in-situ instruments may serve as the standard reference since they are usually subject to proper calibration. Although this may be correct under controlled environmental conditions like laboratories, it is usually not the case in the open ocean with harsh atmospheric and marine conditions. For example, it was found that significant wave height (SWH) measurements from different *in-situ* instruments (buoys and platforms) do not necessarily provide consistent results. In particular, a systematic 10% bias was identified between the US and the Canadian buoy networks (Bidlot et al., 2008). Due to the absence of any other options, however, *in-situ* measurements are generally accepted as the standard reference as far as the bias is concerned.

The variance of the “random error” can be estimated even in the absence of an absolute truth. Triple collocation technique (TCT), which was proposed by Stoffelen (1998) and was improved later by several researchers like Zwieback et al. (2012), has been used for this purpose during the last two decades. This technique can be summarised as follows: given three independent estimates of the truth, T , with unknown random errors it is possible to show that the error variance in each estimate can be found using the total (co-)variances of the three data sets in addition to the “unknown” covariances of the errors. Further assumptions are needed to estimate the error covariances. The assumption of uncorrelated errors, for example, nullifies the error covariance terms. If this assumption is not correct, the error estimates will not be correct. It is also important to note that although the errors in two data sets may not be correlated directly, it may be possible to have a pseudo-correlation due to the nonlinear nature of both errors (e.g. Janssen et al., 2007).

Stoffelen (1998) was the first to apply TCT to estimate the errors of wind measurements from ERS-1 scatterometer, buoys and model analysis. Caires and Sterl (2003) used the same technique to estimate the errors in the 40-year ECMWF Re-Analysis (ERA-40) surface wind speed and significant wave height. Tokmakian and Challenor (1999) implemented the technique to estimate the mean sea level anomalies from the model, ERS-2 and TOPEX/Poseidon altimeters. Freilich and Vanhoff (1999) and Quilfen et al. (2001) used a similar approach with an assumption that the true wind speed is Weibull distributed. Janssen

et al. (2007) estimated significant wave height errors in ERS-2, buoy and ECMWF model analysis, background and hindcast. They extended the method to include two extra sources of information, i.e. quintuple collocation, to estimate the covariances due to existence of correlations between the errors in ERS-2 and Envisat observations. This was the way out to estimate the errors in Envisat RA-2 wave heights in a collocation data set that involves ERS-2 data and model background data. Triple collocation technique was also used by Abdalla and Janssen (2007) to estimate the error of the total column water vapour from the microwave radiometers on board Envisat and Jason-1 as well as the Medium-Resolution Imaging Spectrometer (MERIS) on board Envisat and the ECMWF model analysis. Abdalla et al. (2011a and 2011b) carried out a systematic evaluation of the surface wind speed and SWH from three radar altimeters, buoys and model predictions. The technique was also used in other fields like soil moisture (Scipal et al., 2008; Gruber et al., 2016; and Yilmaz and Crow, 2014); sea surface temperature (O'Carroll et al., 2008) and near-surface humidity (Kinzal et al., 2016). McColl et al. (2014) extended the technique to provide an estimation of the correlation coefficient of the measurement systems with respect to the unknown truth. Furthermore, Zwieback et al. (2016) extended the approach to account for quadratic relationships.

2.1.2 Derivation of TCT

The following derivation is based on Abdalla and De Chiara (2017).

Consider three measuring systems X_j , $j=1, 2, 3$ (e.g. altimeter, in-situ/buoy and model) measuring the same truth at several locations and instants of time. Each raw measurement X_{ji}^r can be related to the truth T_i , with i runs from 1 to N , using an approximate linear error model in the following form:

$$X_{ji}^r = \alpha_j + \beta_j T_i + e_{ji} \quad (1)$$

where X_{ji}^r is the i^{th} measurement taken by measuring system j , α_j is the fixed bias in system j , β_j is the (mis-)calibration factor of system j , T_i is the i^{th} (unknown) truth being measured by the three systems, e_{ji} is the random error in the i^{th} measurement done by system j .

Contrary to expectations, estimating the absolute fixed bias, α_j , in systems measuring quantities like significant wave height is usually non-trivial (except under very controlled laboratory environments). Therefore, it is usually estimated with respect to an arbitrary reference level. There are three possible choices for the reference: using a standard reference, selecting one of the systems as the reference, or adjusting the three systems so that they have the same mean (leading to a zero relative-bias in each system). In order to avoid the complications that can be caused by the biases, the last option can be selected. This choice does not affect the random error estimate, which is the focus of this part.

The calibration factor, β_j , is estimated, in relative sense, using the iterative procedure suggested by Janssen et al. (2007). By removing the bias and using the calibration factor at each iteration, (1) can be written as:

$$X_{ji} = T_i + e_{ji} \quad (2)$$

where X_{ji} is the unbiased and calibrated version of the raw measurement X_{ji}^r . Expressing the measurements of the three systems using (2) and writing the mean squares of the differences between each pair of measurements, one can easily write three equations in the form:

$$N^{-1} \sum (X_{ji} - X_{ki})^2 = N^{-1} \sum (e_{ji} - e_{ki})^2 \quad (3)$$

Here, N is the number of collocation triplets and the summation is done over all collocations. Eq. (3) is repeated three times for pairs of j and k (1 and 2; 1 and 3 and finally 2 and 3). Using the notation: $\langle X \rangle \equiv N^{-1} \sum (X_i)$; Eq. (3) can be written for systems j and k as:

$$\langle X_j^2 \rangle - 2 \langle X_j X_k \rangle + \langle X_k^2 \rangle = \langle e_j^2 \rangle - 2 \langle e_j e_k \rangle + \langle e_k^2 \rangle \quad (4)$$

Again (4) represents three equations by repeating j and k for the pairs: (1,2), (1,3) and (2,3). $\langle X_j^2 \rangle$, $j=1, 2, 3$ are the measurement variances of the three systems, and $\langle X_j X_k \rangle$ are the measurement covariances (three of them). Similarly, $\langle e_j^2 \rangle$ and $\langle e_j e_k \rangle$ are the error variances and covariances, respectively. The former group are known from the statistics of the collocated triplets while the latter are unknowns.

Solving for the error variances and rearranging:

$$\langle e_1^2 \rangle = \langle X_1^2 \rangle - \langle X_1 X_2 \rangle - \langle X_1 X_3 \rangle + \langle X_2 X_3 \rangle + \langle e_1 e_2 \rangle + \langle e_1 e_3 \rangle - \langle e_2 e_3 \rangle \quad (5)$$

$$\langle e_2^2 \rangle = \langle X_2^2 \rangle - \langle X_1 X_2 \rangle - \langle X_2 X_3 \rangle + \langle X_1 X_3 \rangle + \langle e_1 e_2 \rangle + \langle e_2 e_3 \rangle - \langle e_1 e_3 \rangle \quad (6)$$

$$\langle e_3^2 \rangle = \langle X_3^2 \rangle - \langle X_2 X_3 \rangle - \langle X_1 X_3 \rangle + \langle X_1 X_2 \rangle + \langle e_2 e_3 \rangle + \langle e_1 e_3 \rangle - \langle e_1 e_2 \rangle \quad (7)$$

Note that the first 4 terms (in capital letters) on the right-hand side of (5)-(7) are known from the statistics of the collocated triplets while the last three (the error covariances which have the letter e in them) are unknowns. So there are three equations (5)-(7) with six unknowns: the error variances on the left hand side and the error covariances on the right hand side.

Considering two data sets x_i and y_i with $i = 1, 2, \dots, N$ that have zero means, the Pearson correlation coefficient between the two data sets is defined as:

$$r^2 = \langle x y \rangle^2 / (\langle x^2 \rangle \langle y^2 \rangle) \quad (8)$$

Therefore, the covariance $\langle x y \rangle$ can be interpreted as a measure for the correlation between the two data sets.

Assuming that the random errors in the three measuring systems are uncorrelated (i.e. all the error covariances are zeros), the number of unknowns in the system of equations defined in (5)-(7) reduces to three which are the error variances that appear on the left-hand side of the three equations. The solution is straightforward.

2.1.3 Suggested Procedure

A flowchart of the iterative procedure suggested originally by Janssen et al. (2007) and modified later by Abdalla et al. (2013) is shown in Figure 1. The procedure assumes that one of the systems (say the first one, X_1) is calibrated (i.e. $\beta_1 = 1$). The other two systems need

to be calibrated (i.e. To find β_2 and β_3) accordingly. Any of the three systems can be assumed to be calibrated but the final calibration factors of the other two systems will be relative to the selected system. Therefore, one needs to select the one which is believed to be well calibrated. This selection does not impact the estimated random errors.

The iteration starts by assuming that the calibration factors of the other two systems equal to 1. The error variances $\langle e_1^2 \rangle$, $\langle e_2^2 \rangle$ and $\langle e_3^2 \rangle$ are computed using (5)-(7).

The neutral regression (e.g. Marsden, 1999) is used to compute the calibration factors. Conventional regression is not suitable as it assumes that one of the measuring systems is error-free. This is almost impossible in real work measurements. The calibration factor of the second system β_2 is computed using:

$$\beta_y = [-B + (B^2 - 4 A C)^{1/2}] / 2 A \quad (9)$$

where:

$$A = \gamma \langle X_1 X_2 \rangle ; \quad (10)$$

$$B = \langle X_1^2 \rangle - \gamma \langle X_2^2 \rangle ; \quad (11)$$

$$C = - \langle X_1 X_2 \rangle ; \text{ and} \quad (12)$$

$$\gamma = \langle e_1^2 \rangle / \langle e_2^2 \rangle \quad (13)$$

Similarly, β_3 can be found by replacing X_2 above with X_3 . The changes in the values of the calibration factors β_2 and β_3 are tested against a given tolerance. If the change is large, the procedure is repeated with the new values of the calibration factors. Otherwise, the procedure is terminated by adjusting the estimated error variances to account for the collocation distance (see Paragraph e of Subsection 2.1.4).

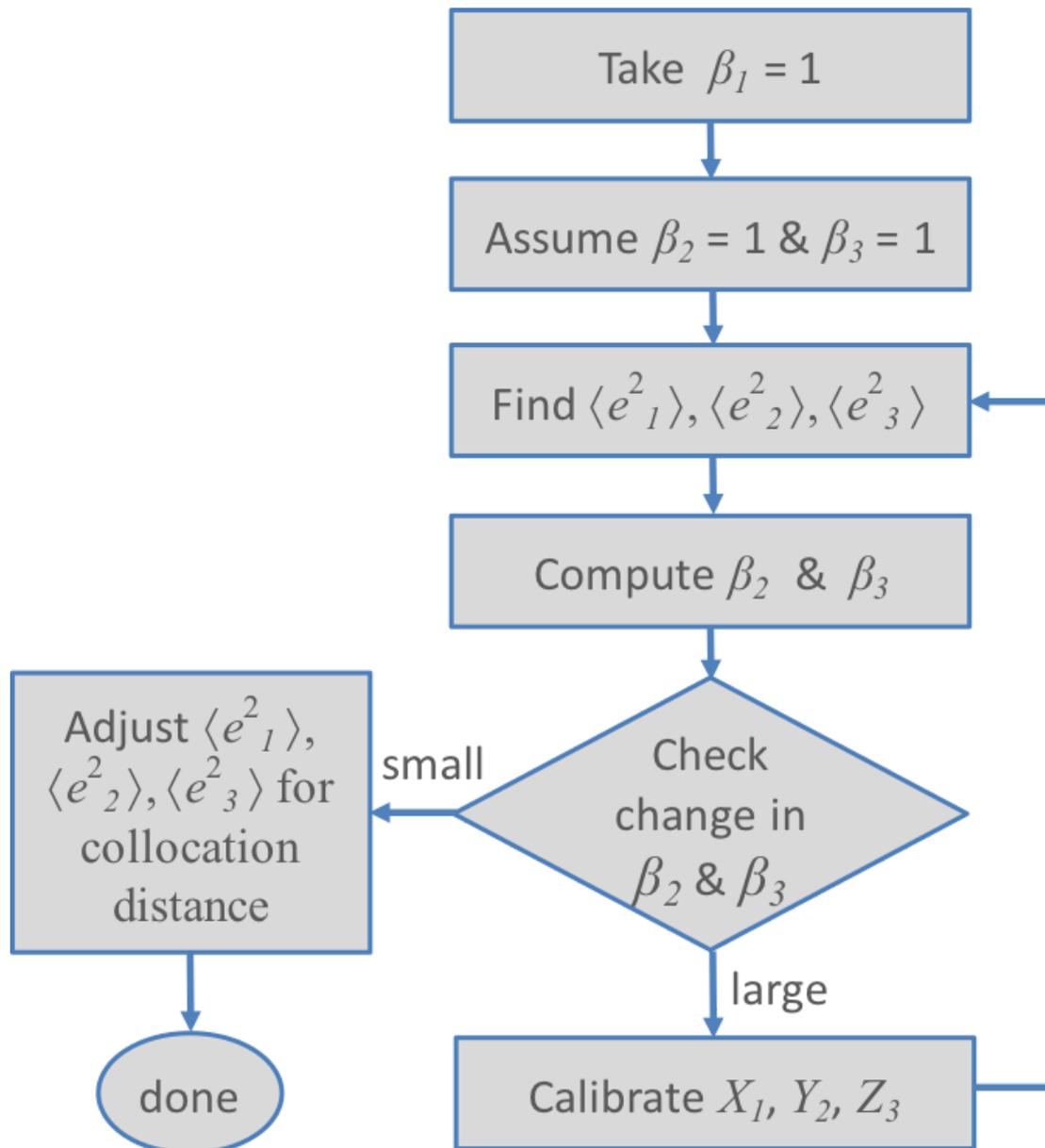


Figure 2: Flow chart of the procedure to implement the TCT.

2.1.4 Practical Considerations

a) Handling error correlations

The error correlations which are represented by the error covariances in the three equations (5)-(7) are usually not known. The best strategy is to carefully select the three data sets with errors that are independent from errors of the other two. Sometimes it is possible to select or design three measuring systems with uncorrelated random errors (Zwieback et al., 2012). Janssen et al. (2007) used quadruple and quintuple collocations to handle the error correlations between ERS-2, ENVISAT and the wave model. Gruber et al. (2016) and Zwieback et al. (2012) developed methods to account for error correlations that are specific to the soil moisture. Abdalla and De Chiara (2017) avoided error correlations due to assimilating scatterometer wind data into ECMWF NWP model by considering model forecasts at medium ranges. However, this is not always possible especially for historical data sets and for expensive-to-conduct data sets.

It is important to stress that the correlation of concern is the “error correlation” not the “measurement correlation”. The measurements of any two systems are highly correlated but this does not mean that their errors are correlated. Examples for possible reasons to have error correlations between two measuring systems can be:

- If the two measuring systems share the same principles of measurement, the assumptions used and the processes that are unaccounted for while making the measurements cause considerable amount of error correlation. For example, Jassen et al. (2007) showed that ERS-2 and ENVISAT radar altimeter SWH measurements have significant error correlations.
- Use of data from one system in calibrating the retrieval algorithm of the second system also introduces a considerable amount of error correlation.
- If one of the systems assimilates the measurement of the second system (e.g. as in the case of NWP system), this is another source of error correlation.

If one of the system triplets is a model, especially in the case of NWP models, error correlation is a serious issue. NWP models assimilate a large number of measurements coming from a wide range of measuring systems. Furthermore, NWP model fields are considered an attractive source of data for developing retrieval algorithms or for instrument calibration. In order to eliminate the impact of error correlation in this context, it is possible to use the following treatments:

1. use other methods or prior knowledge to evaluate or estimate the error correlations;
2. use of the model forecasts at medium-ranges rather than the model analysis or model forecasts at short ranges. For assimilated observations the correlation between the errors of the model and the other system decreases with the increase in the forecast lead time;
3. conduct numerical experiments where the model is run without assimilating the data from the systems used in the triple collocation. This blacklisting should extend to other systems which can be correlated with the systems under concern. However, the results related to model error will then not represent the errors in the full NWP system;
4. use of additional measurement systems and perform quadruple, quintuple, etc. collocation techniques.

b) Representativeness Error

The representativeness error which emerges due to scale differences among various measuring systems is usually an issue for triple collocations as well as for direct comparisons (see for example Stoffelen, 1998 and Vogelzang et al., 2011). If one of the systems is able to resolve finer scales compared to the others, triple collocation penalises it and accounts the sensed actual small-scale variability as an additional error. There are several ways to deal with the representativeness error either by using assumptions based on the understanding of the systems as in Stoffelen (1998) or by performing spectral analysis as in Vogelzang et al. (2011). In order to avoid this issue, the scales of the three measuring systems are brought to same level. Measurement system with the coarsest scale dictates the selected scale for all systems. Averaging is used to “coarsen” the scale of finer systems. The term “scale” here is used to refer to the effective resolution or the ability of the

measuring system to resolve true variations. The scale is usually higher than the resolution, discretization or sampling interval/spacing due to internal correlations. In case of ECMWF IFS numerical model, for example, the TL1279 model grid spacing is 16 km but its intrinsic effective resolution (scale) can be between 4 to 8 grid spacings (which is equivalent to 60-125 km) depending on the accepted level of “resolution” (Abdalla et al., 2013c).

Radar altimeters sample the sea surface at a typical rate of several hundreds of pulses per second (e.g. more than 1700 pulses per second for Jasons). An on-board averaging reduces this rate to about 20 Hz. Further processing produces 1-Hz products sampled every 6-7 km which are traditionally made available to the users. Details can be found in Chelton et al. (2001). Janssen et al. (2007), Abdalla et al. (2011a and b) and Abdalla and De Chiara (2017) average the 1-Hz products, which has a typical scale of 6-8 km (slightly larger than the sampling interval), along the track to give a scale of about 75-100 km to be comparable to the IFS effective resolution.

For in-situ data, measurements are carried out temporally at individual locations. The spatial scale as such does not have any physical meaning. Using Taylor’s hypothesis, it is possible to relate spatial scales to the temporal scales using the mean velocity of the flow. In the case of wind waves, the typical wave period is 8 seconds (frequency of 0.125 Hz). The group velocity, which represents the mean flow velocity of the wave energy, is slightly higher than 6 m/s which is about 22 km/hr. The scale of 100 km is equivalent to about 4.5 hours. Therefore, most of verification studies that involve NWP and in-situ data adopt 5-hour averaging of in-situ data.

If data sets involved in the TCT are all roughly of the same scale, the representativeness error becomes less of an issue. However, the error estimates are valid only for that specific scale.

c) Biases

As discussed earlier, biases cannot be found in absolute sense. However, large bias differences among the three measuring systems confuse the TCT and the resulting error estimates will be wrong. In the absence of the truth or an unbiased measuring system, all data systems should be brought to the same level. This can be done by adopting one of the three data sets as the reference. The other two data sets are bias corrected with respect to the reference data set.

Another option is to adjust each data set by subtracting its mean from each measurement in that set. The mean of each adjusted data set becomes zero.

d) Calibration Factors

After assuming that one of the measurement systems is well calibrated, the calibration factors of the other two systems are estimated using an iterative procedure as proposed by Janssen et al. (2007) as already mentioned above in Sub-section 2.1.3 and depicted in Figure 2. On the other hand, Su et al. (2014) derived several formulae to evaluate the calibration factors (they call them scaling coefficients). In particular, they derived a formula for triple collocation (their Eq. 13) which can be written as:

$$\beta_2 = \text{cov}(X_2, X_3) / \text{cov}(X_1, X_3) \quad (14)$$

where $\text{cov}(x,y)$ is the covariance operator. The results of (14) were found to coincide (within 5 significant digits) with the calibration factors computed iteratively as suggested by Janssen et al. (2007).

e) Number of Collocations and the Impact of the Collocation Distance

Usually in-situ data are used as one of the triplets in TCT. For sea-state measurements there is a limited number of measuring stations scattered mainly around Europe and North America. The collocated altimeter and buoy measurements should be very close to each other to represent the same truth and, at the same time this restriction should be relaxed to have enough triplets to yield statistically representative error estimates. If restrictive criteria for the collocation are applied without allowing any tolerance in space and time, the number of the triplets would be too small to provide representative error statistics. Abdalla et al. (2013b) showed that over a whole year, restricting the collocation distance between altimeter measurements (ENVISAT, Jason-1, or Jason-2) and in-situ measurements to below 50 km reduces the number of collocation significantly to a dozen of hundreds as can be seen in Figure 3.

Janssen et al. (2007) suggested the use of a relaxed collocation distance of 200 km and time difference of 2 hours. To ensure that the triplet measure the same truth restrictions on the model predictions are imposed. Their collocation procedure is very simple in a sense that any altimeter measurement within 200 km from a buoy location is collocated with the buoy measurement (within 2 hours). If there is a large gradient or discontinuity of the SWH between the locations of the two measurements, the “natural” difference between them will be interpreted as an additional error that cannot be separated from the random error. For example, if the buoy is located at one side of an island or a peninsula while the altimeter track is at the other side, for sure both measurements will not be representing the same truth. To eliminate, as much as possible, this kind of added error, any triplet is rejected when the model estimates at the altimeter location and at the buoy location differ by more than 5%. The assumption here, which is a fair one, is that the model is able to reproduce the true geophysical variability. Therefore, too different “model” SWH values is a strong indication that the altimeter and the buoy measurements do not represent the same geophysical truth. Furthermore, any triplet with the model mean wave direction at the altimeter and at the buoy locations are different by more than 45 degrees is rejected. This is another measure intended to filter out the field inhomogeneities due to, for example, atmospheric fronts. It should be noted that in the criteria mentioned above, the 5% and the 45 degrees are empirical values based on experience.

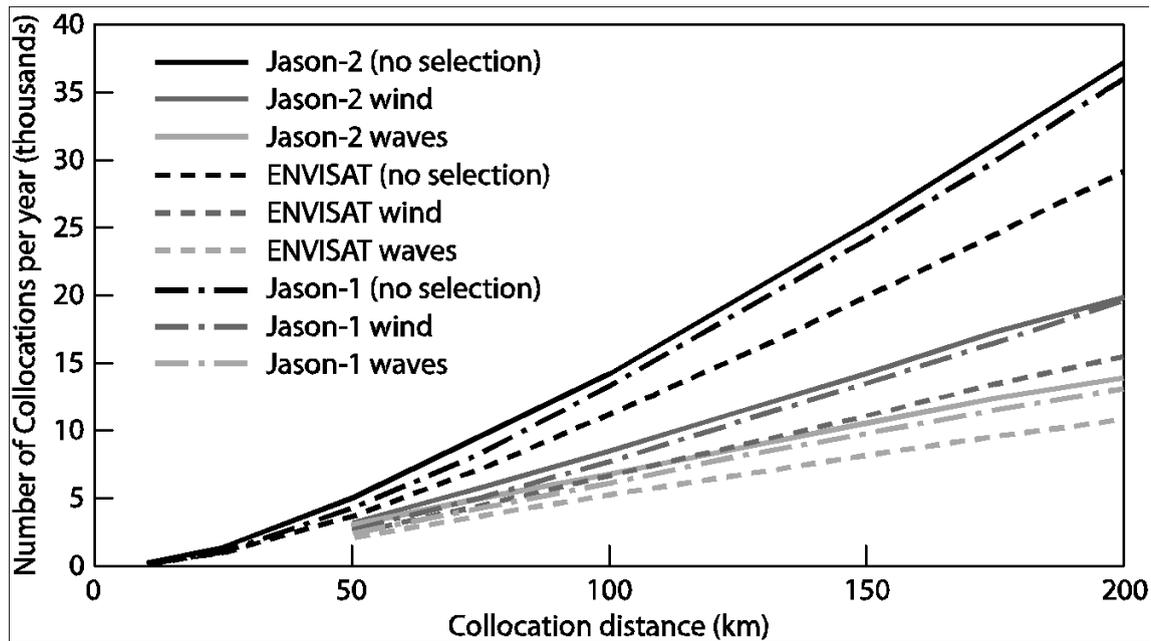


Figure 3: Number of collocations for various restriction conditions for the period from 1 August 2009 to 31 July 2010 (from Abdalla et al., 2013a)

To account for the change in error estimates due to the collocation distance, Abdalla et al. (2013b) repeated the error estimates for different collocation distances as shown in Figure 4 for triplets that involve a model hindcast (model run without data assimilation), buoy and an altimeter. The exercise was repeated for three altimeters: ENVISAT, Jason-1 and Jason-2. Abdalla et al. (2013a) estimated the impact of the collocation distance between the altimeter and the buoy as given in the legend at the right-hand side top of Figure 4. It is clear that the change of error with respect to the collocation distance is linear. Altimeter (at different levels but have more or less same slope) and buoy SWH errors increase by increasing the collocation distance while the model error is decreasing. The rates of the change per 100 km are 0.021 m, 0.0004 m, and -0.19 m for the altimeter, the buoy and the model hindcast, respectively. More or less same rates with respect to the collocation distance (line slopes in Figure 4) were found for each triplet whether it is ENVISAT, Jason-2 or Jason-1 in addition to the model and the buoys.

The result above can be utilized to increase the number of collocations by adopting the 200-km restriction of the collocation distance. The error estimates are then adjusted by using the results in Figure 4. An argument can be raised if one needs to adjust for a zero-collocation distance or for a collocation distance depending on the scale of the model and super-observations (which is typically between 75 and 100 km). Using the distance equivalent to the scale (e.g. 75 km) instead of the 0-km distance, would add about 0.02 m to the altimeter errors (model error by about 0.017 m while buoy errors will not change). Abdalla et al. (2013b) decided that a collocation distance equals to the scale of the data would be the proper selection.

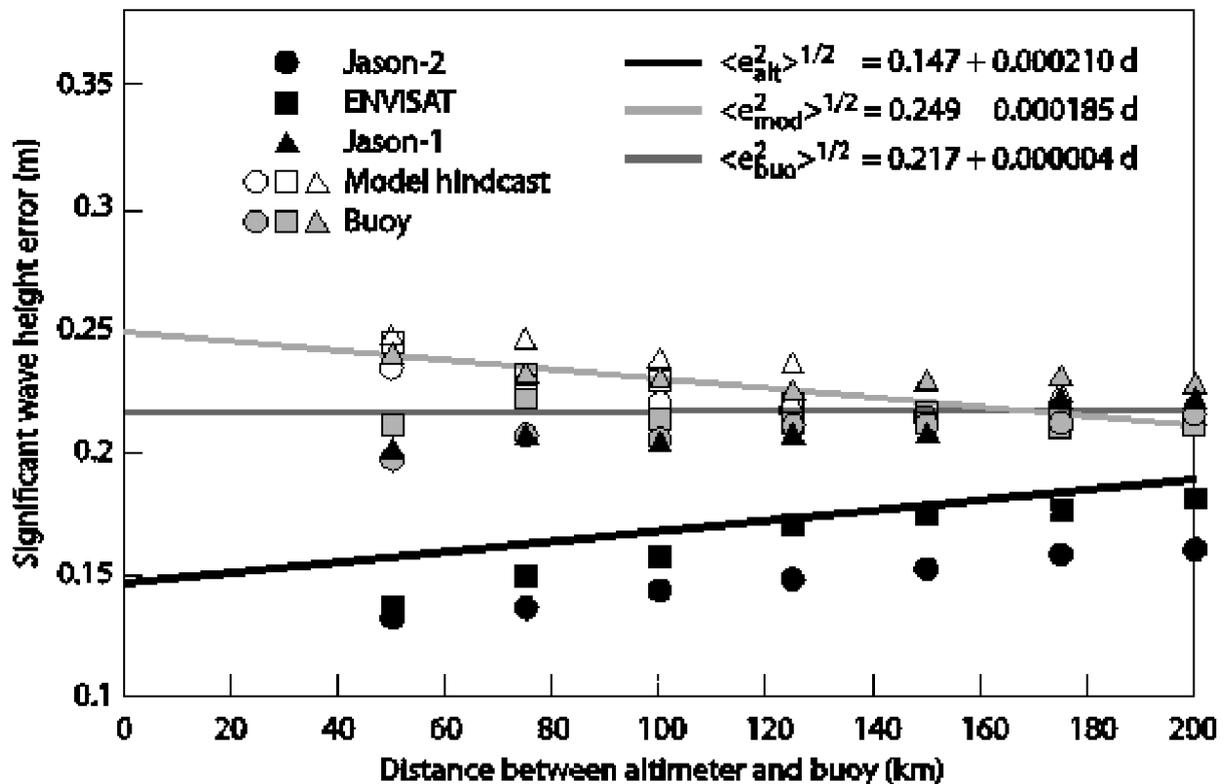


Figure 4: Change of SWH errors as functions of the maximum allowed collocation distance. The linear regression fits are given. The errors $\langle e_{alt}^2 \rangle^{1/2}$, $\langle e_{mod}^2 \rangle^{1/2}$, $\langle e_{buo}^2 \rangle^{1/2}$ of the altimeters, the model and the buoys are in m while the collocation distance, d , is in km. (from Abdalla et al., 2013a)

f) Robustness of the Results

Abdalla and De Chiara (2017), who used the TCT to estimate the errors in the altimeter, scatterometer and model winds, applied the standard boot-strapping procedure with replacement on the original data samples. They created 200 new samples, with half the size, by random selection from the original sample. The error variances were estimated for each new sample. The mean and the standard deviation (SD) of the 200 estimates of error variances were computed and the 95% confidence interval was established. This is similar to the procedure followed by Caires and Sterl (2003) where the 95% confidence intervals were computed simply as 1.96 SD of errors emerging from the bootstrap samples.

Since their sample sizes were rather large (35,000 collocations), Abdalla and De Chiara (2017) found that halving the sample does not impact the results. Intuitively, one expects the sample size to play an important role in the success of the TCT. The question is whether there is a critical sample size that is required for the success of this type of analysis. The answer may not be straightforward, but the bootstrapping may provide a guideline. They subsampled the original sample randomly 200 times for each given size and computed the errors, the ensemble means of the errors and the 95% confidence intervals. This was repeated for different sizes. They concluded that, as a crude estimate, sample sizes of few thousands of collocations are needed to get acceptable error estimates. Note that this exercise was carried out for wind speed not SWH.

2.2 Satellite to validation datasets comparison metrics

Significant wave height (H_s) will be evaluated using RMS difference with respect to the colocated significant wave height of validation dataset. The collocation distance in time and space will be as short as possible but determined based on the minimum of collocation point number that needs to be representative (of the order of 100 minimum)

Wave spectra L2 products will be evaluated using RMS on integral parameters per spectral partitions. The following parameters will be computed in the L2 product and compared with the validation datasets using a RMS difference: H_s , dir peak and spread, wl peak and spread.

Additional metrics such as scatter index and correlation coefficient will be estimated.

QQplots will also be provided to enable comparison of the different Sea State ECV products statistics with the validation datasets statistics.

3. Validation data

We describe here all reference validation datasets that will be used to validate the Sea State ECV.

2.1 In situ data

Moored buoys with more than 10 years of data that are sufficiently far from coastlines in deep water are important reference time series for validation and calibration of the satellite observations. Buoys from the NDBC, CDIP, MEDS, and OCEANSITES networks with these criteria are shown in the following Figure 5. More importantly, these data will help in validating the consistency of the satellite records for climate applications.

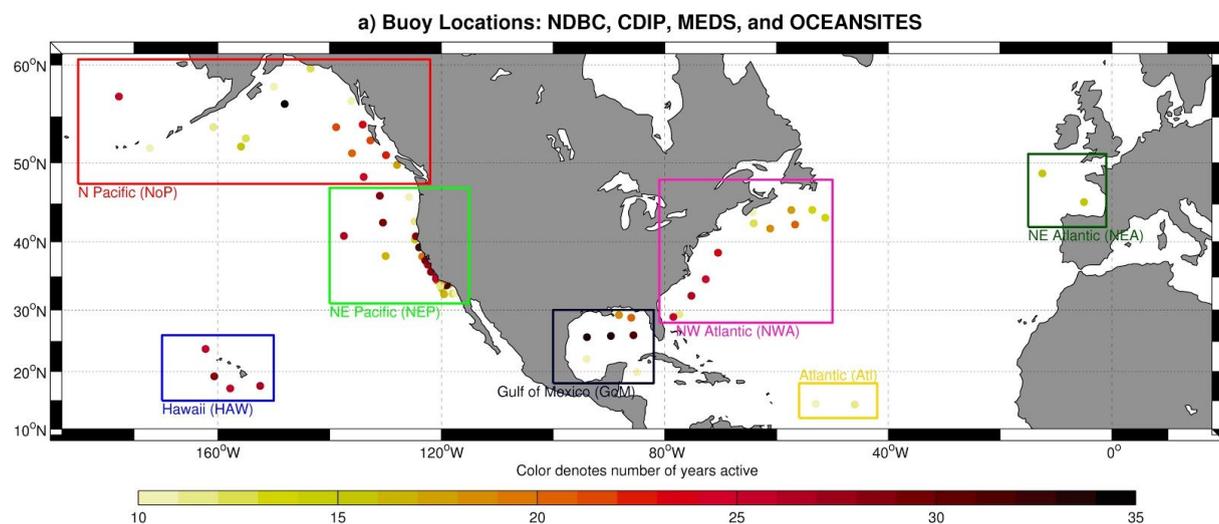


Figure 5: Buoys with at least 10 years of observations, >50 km from the coast, and in deep water. The color code represents the length of the buoy time series in years.

A major concern of the time series from the moored buoys is the consistency. Changes in sensor payloads, geographical displacements, sensor configurations (e.g. placement of the sensors on the buoy), and hull type all affect the homogeneity of the historical records. Changes due to these issues were observed on buoys from the MEDS and NDBC networks in the NE Pacific (Gemrich et al., 2011). These calibration issues are more widespread and exist on buoys near Hawaii, Gulf of Mexico, and NW Atlantic (Livermont, 2017). It is proposed that a subset of the buoys, possibly those with the longest time series, shown in the above figure are analyzed in more detail assessing the consistency utilizing the metadata describing the payload and hull changes. For this endeavour we propose to use existing tools such as RHtestV4 developed to assess homogeneity in time series developed by Wang and Feng (2013). This task will be performed in collaboration with Bob Jensen of the USACE and Ian Young of University of Melbourne.

For the **in situ wind and wave** measurements that will be used for the validation of CCI products, the project will rely on the **CMEMS in situ TAC** that collects all buoy data from the following sources:

- NDBC (including CDIP)
- MEDS
- CEREMA
- Meteo-France
- various buoys over European seas over Baltic, Azores, Spain, Portugal, Greece

This service takes over the task previously performed through **ESA/GlobWave** project and both teams work together to ensure a smooth and seamless transition. While not yet fully operational today, it will provide the full backlog for all these networks, as well as the wave spectra when available, and will then be a major asset for the CCI Sea State project.

The data are fully homogenized (variable names and format), quality controlled and delivered in NetCDF format. The illustration below shows the coverage over September 2017.

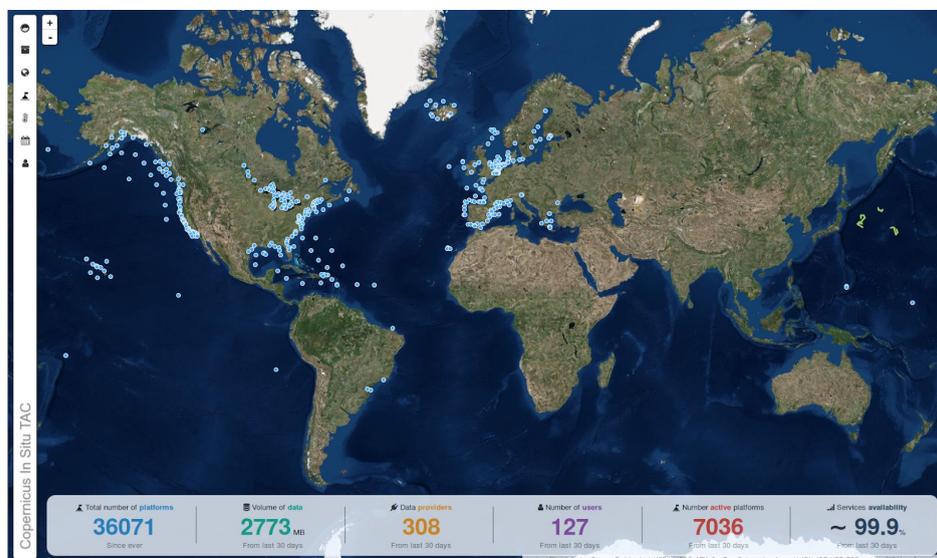


Figure 6: Coverage of in situ data over September 2017

2.2 Wave enabled drifters

The Global Drifter Program (GDP) (Niiler 2001) has been created to manage the deployment of surface drifters to follow surface currents.

Since 2016 the GDP drifters use exclusively the Iridium satellite system because of the shorter data latency compared to the Argos satellite system (1 minute vs 90—120). Since the geolocation computed from the Doppler shift of the transmitter carrier frequency for Iridium is less accurate (~2 km) than the one obtained from Argos satellites (~200 m) [Lopez et al. [2014]] all of the Iridium drifters are equipped with a Global Positioning System (GPS) engine (accuracies of ~2m—50 m rms).

Since the GPS engine can also be used to obtain directional wave properties, an exciting opportunity that is actively pursued by the GDP is the addition of directional wave spectra estimates from drifters. The advantages of GPS-derived wave properties are both practical [Herbers et al., 2012] and financial. Undrogued drifters can be turned into directional wave riders and can become the first in situ global network of wave sensors. It is anticipated that wave forecasting model will greatly benefit from this application.

Luca Centurioni from Scrips in San Diego agreed to share with the project data from these wave enabled drifters deployed in the Central pacific and in the southern ocean where little other wave measurements are possible.

2.3 Microseism data

Seismic records of ground displacement anywhere on Earth contain the signature of ocean waves in a broad frequency band that typically ranges from 0.003 Hz to 1 Hz, this signature is known as microseisms. The quantitative link between waves and microseism is well understood for the vertical displacements in the “secondary microseism” peak that is around 0.2 Hz (e.g. Bernard 1941, Longuet-Higgins 1950, Hasselmann 1963, Ardhuin & Herbers 2013). This is dominated by Rayleigh waves generated where ocean waves of opposing direction and same frequency are found in the ocean. Such seismic sources can be put in three broad classes of events (Ardhuin et al. 2011).

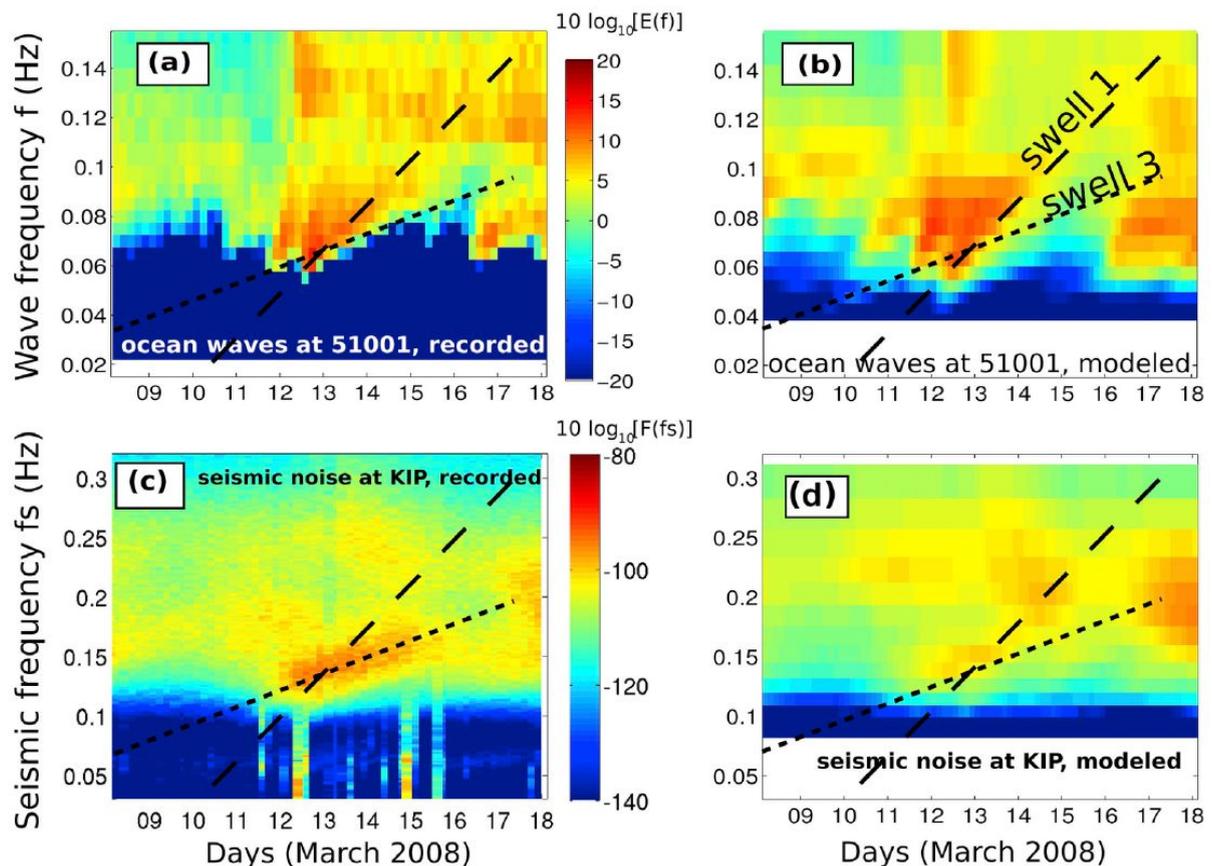


Figure 7: Example of seismic spectrogram (bottom) observed and modeled, compared to wave spectra (top). Taken from Ardhuin et al. (2011).

Due to the importance of opposing directions, there is no simple relation between wave height and microseism amplitude. However, for a same ocean wave spectrum shape, an increase in microseism correspond to an increase in wave height. As a result, microseism amplitudes can be use to track the magnitude of ocean waves over time (Bernard 1990, Grevemeyer et al. 2000). Any given station gives some sea state index in a region that varies in size from a few hundreds of kilometers to a few thousand of kilometers with a clear influence of regional sea ice for example (Stutzmann et al., 2009, Stutzmann et al. 2012, Sergeant et al. 2013). Furthermore, secondary microseism can provide independent constraints on the amount of ocean wave reflection coefficient. Secondary microseisms are generated by the interaction of ocean waves of similar frequency and coming from opposite directions. The resulting pressure fluctuations have the double of the frequency of the ocean waves. From these pressure sources, seismic waves propagate everywhere in the ocean and the Earth and also have the double of the frequency of the ocean waves. Primary microseisms are generated by the direct coupling of the ocean waves with the sloping ocean-continent boundary and the resulting seismic waves have the same frequency as the ocean waves.

The processed data will include all Geoscope stations (see map below, <http://geoscope.ipgp.fr/index.php/en>). Data are converted into seismic acceleration and spectrogram are computed using windows of 3 hours. Three-hourly spectra data for the vertical components with 512 frequencies takes about 7 Mb for a full year. The total volume for 40 stations over 30 years is only 8 Gb, including quality flags indicating the likely occurrence of earthquakes. Synthetic spectrograms will be computed and the comparison

between observed and modelled spectrogram will enable to constrain the ocean wave coastal reflection coefficient.

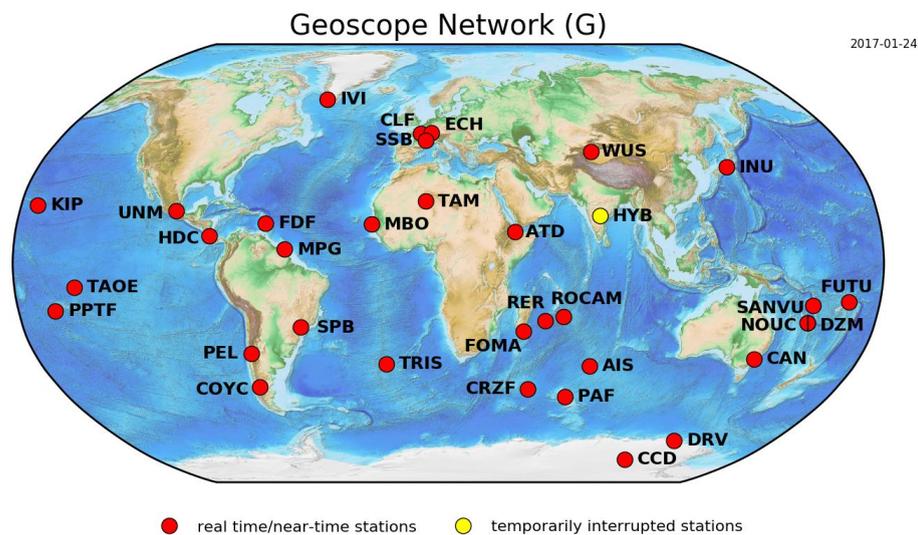


Figure 8: Extension of the Geoscope network of seismic stations

Ongoing developments on seismic array processing also shows that a single array can be used to get a map of microseism sources (Farra et al., 2015, Meschede et al., 2017). We will further explore the capability of using long-term arrays such as the Grafenberg array for mapping sources and their evolution over the past 40 years.

2.4 Wave model data

While wave model data cannot be considered as ground truth, they provide valuable large number of collocation with satellite data allowing to validate the sensitivity of the retrieved wave measurements with respect to the wide variety of metocean conditions.

Model estimates from MFWAM distributed by CMEMS will be used.

This global wave system of Météo-France is based on the wave model MFWAM which is a third generation wave model. MFWAM uses the computing code ECWAM-IFS-38R2 with a dissipation terms developed by Ardhuin et al. (2010). The model MFWAM was upgraded on november 2014 thanks to improvements obtained from the european research project « my wave » (Janssen et al. 2014). The model mean bathymetry is generated by using 2-minute gridded global topography data ETOPO2/NOAA. Native model grid is irregular with decreasing distance in the latitudinal direction close to the poles. At the equator the distance

in the latitudinal direction is more or less fixed with grid size $1/10^\circ$. The operational model MFWAM is driven by 6-hourly analysis and 3-hourly forecasted winds from the IFS-ECMWF atmospheric system and by the Mercator $1/12$ global ocean surface current distributed by CMEMS. The wave spectrum is discretized in 24 directions and 30 frequencies ranging from 0.035 Hz to 0.58 Hz. The model MFWAM uses the assimilation of altimeters with a time step of 6 hours. The global wave system provides analysis 4 times a day, and a forecast of 5 days at 0:00 UTC. The wave model MFWAM uses the partitioning to split the swell spectrum in primary and secondary swells.

2.5 Independent satellite measurements

We note that CFOSAT SWIM instrument provides a novel wave scatterometer that can provide wavelength and direction information well collocated in time with Sentinel-1 A/B data as both missions are on a dust down orbit (equator crossing time near 6am/6pm local time). These SWIM derived information can be compared to SAR derived wavelength and directions. Note that SWIM instrument is rather recent and careful validation of these measurements is still ongoing.

4. Validation in the Round Robin process

Validation in the Round Robin process will use its own methods suitable for fair intercomparison of algorithms over a representative and common subset of the reference validation datasets. The agreed metrics for this are defined in five parts as follows:

Part 1: outliers analysis

Definition: outliers are considered points for which lie outside $[-0.25\text{m}, 25\text{m}]$, which SWH=NaN, and/or which are more than three times the median absolute deviation away from the median of the closest 20 points.

1.1 Total Number of Outliers

1.2 Number of Outliers in the Coastal Zone (distance to coast < 20 Km, < 10km and < 5km). "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."

Part 2: noise analysis as a function of distance to coast

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

2.1 Median of all valid noise values as a function of distance to coast, with open

ocean represented by distance to coast > 20km

2.2 Median of noise at $0\text{m} < \text{SWH} < 1\text{m}$ (low sea states) in open ocean and coastal zone

2.3 Median of noise at $1.5\text{m} < \text{SWH} < 2.5\text{m}$ (average sea states) in open ocean and coastal zone

2.4 Median of noise at $\text{SWH} > 6\text{m}$ (high sea states) in open ocean and coastal zone

2.5 Median of noise at $\text{SWH} > 12\text{m}$ (very high sea states) in open ocean and coastal zone

Part 3: comparison with in-situ data

Definition: buoys shall be grouped into “open ocean buoys” and “coastal (but exposed) buoys” (see presentation by Quartly et al. at CAW 2018). Statistics will be separated accordingly.

Definition: “closest point” is defined as the median SWH of the 51 high-frequency (“51 20-Hz”) closest points to the buoy, including NaNs and after unrealistic estimations are excluded. Unrealistic estimations are excluded in the following way: first points outside the interval $[-0.25\text{ m}, 25\text{ m}]$ are excluded, secondly the 3- sigma criterion is used to eliminate remaining unrealistic estimations.

For each of the following statistics, the mean values over all the “open ocean buoys” on one side and the “coastal buoys” on the other side will be considered for the assessment.

3.1 PCHC at closest point

The Percentage of Cycles for High Correlation (PCHC, Passaro et al. 2015) is a statistic designed to take into account both the correlation (Pearson correlation coefficient) between time series and the number of observations available. The test is performed in an iterative way: First of all, for the selected location, the correlation of the buoy time series with the entire set of altimetry retrievals is checked; if the correlation coefficient is lower than a certain threshold, then the cycle with the maximum discrepancy (quantified as the maximum of the absolute value of the difference) between the buoy value and altimeter retrieval is excluded. This exclusion is iterated until the correlation rises above that threshold, at which point the percentage of cycles left provides the measure of the general quality of the retracked altimetry values.

3.2 Standard Deviation between SWH from altimetry at closest point and buoy.

3.3 Slope of the linear fit (regression line of $\text{SWH}_{\text{altimetry}}$ -vs- SWH_{buoy} scatter plot)

3.4 Median bias between SWH from altimetry at closest point and buoy

Part 4: comparison with model output

Model grid points and altimetry will be coupled by considering the median of the SWH 20-Hz measurements from altimetry within the grid point. Coastal and Open Ocean statistics will be divided: Coastal statistics will include measurements and model data closer than 20 km to the coast. The following statistics will be provided:

4.1 Correlation

4.2 Standard Deviation of the difference between SWH from altimetry and SWH from model.

4.3 Slope of the linear fit (regression line of $\text{SWH}_{\text{altimetry}}$ -vs- $\text{SWH}_{\text{model}}$ scatter plot)

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

Part 5: representation of scales of variability

Along-track spectra of SWH will be calculated for open ocean segments of track of at least 1024 points (~330 km length) using Welch's method. A useful measure to be extracted is the "signal variance at the scales of interest", which will be the mean spectral level averaged over wavelengths of 100-50km and 50-25 km.

5. ECV product validation and User assessment

Work Package 4500 is dedicated to the final validation and intercomparison of ECV products. All the validation diagnoses defined previously and applied in the round-robin phase could be applied in this task. The results of validation are reported in the Product Validation and Intercomparison Report, produced in three versions during the second two years of the project.

Work Package 5000 covers assessment of products from a climate perspective and includes Case Studies for the specific assessment of ECV products for particular applications:

- Extremes at the coast
- Tropical swell and storms
- Links with Copernicus Climate Change Service
- Wave near ice

Results will be reported in the Climate Assessment Report and peer-reviewed journal publications.

6. References

- Abdalla, S. and Janssen, P. A. E. M. (2007). "Monitoring and Validation of Water Vapour Products from ENVISAT MWR and MERIS", in Proc. Envisat Symposium 2007, Montreux, Switzerland, April 2007
- Abdalla, S., Janssen, P. A. E. M. and Bidlot J.-R. (2011a). "Altimeter near real time wind and wave products: Random error estimation", *Mar. Geod.*, vol. 34, pp. 393-406.
- Abdalla, S., Janssen, P. A. E. M. and Bidlot J.-R. (2011b). "On the Accuracy of Surface Wind and Wind-Wave Data," in Proc. 10th Int. Conf. on the Mediterranean Coastal Environment, MEDCOAST 11, E. Ozhan (Ed.), Mediterranean Coastal Foundation, Dalyan, Turkey, pp. 799-808.
- Abdalla, S. and De Chiara, G. (2017) "Estimating Random Errors of Scatterometer, Altimeter and Model wind speed Data", *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10:5, 2406-2414. doi: 10.1109/JSTARS.2017.2659220.
- Abdalla, S., Isaksen, L., Janssen, P. A. E. M. and Wedi, N. (2013c) "Effective Spectral Resolution of ECMWF Atmospheric Forecast Models", *ECMWF Newsletter*, n. 137, pp. 19-22.
- Ardhuin, F., Stutzmann, E., Schimmel, M., and Mangeney, A., "Ocean wave sources of seismic noise," *J. Geophys. Res.*, 116, p. C09004, 2011. doi:10.1029/2011JC006952.
- Ardhuin, F. and Herbers, T. H. C., "Noise generation in the solid earth, oceans and atmosphere, from nonlinear interacting surface gravity waves in finite depth," *J. Fluid Mech.*, 716, 316–348, 2013. doi:10.1017/jfm.2012.548.
- Ardhuin, F., R. Magne, J-F. Filipot, A. Van der Westhuyzen, A. Roland, P. Queffelec, J. M. Lefèvre, L. Aouf, A. Babanin and F. Collard : Semi empirical dissipation source functions for wind-wave models : Part I, definition and calibration and validation at global scales. *Journal of Physical Oceanography*, March 2010.
- Bernard, P., "Sur certaines propriétés de la boule étudiées à l'aide des enregistrements sismographiques," *Bull. Inst. Oceanogr. Monaco*, 800, 1–19, 1941.
- Bernard, P., "Historical sketch of microseisms from past to future," *Phys. Earth Planetary Interiors*, 63, 145–150, 1990. doi:0031-9201(90)90013-N.
- Bidlot, J.-R., Durrant, T. and Queffelec, P. (2008). "Assessment of the systematic differences in wave observations from moorings," Presentation at JCOMM Technical Workshop on Wave Measurements from Buoys, New York, U.S, October 2008.
- Caires, S., and A. Sterl, Validation of ocean wind and wave data using triple collocation, *J. Geophys. Res.*, 108(C3), 3098, doi:10.1029/2002JC001491, 2003

-
- Chelton, D. E., Ries, J. C., Haines, B. J., Fu, L.-L. and Callahan, P.S. (2001). "Satellite altimetry," in: L.-L. Fu, A. Cazenave (ed), *Satellite Altimetry and Earth Sciences*, Academic Press.
- Farra V., E. Stutzmann, L. Gualtieri, M. Schimmel and F. Ardhuin. Ray-theoretical modeling of secondary microseism P-waves. 2016, *Geoph. J. Int.*, 206, 1730-1739, doi: 10.1093/gji/ggw242
- Freilich, M. H. and Vanhoff, B. A. (1999). "QuikScat vector wind accuracy: Initial estimates," In *Proceed. QuikScat Cal/Val Early Science Meeting*, Pasadena, CA, Jet Propulsion Laboratory.
- Gemmrich, J., B. R. Thomas, and R. Bouchard, **2011**: Observational changes and trends in northeast Pacific wave records. *Geophys. Res. Lett.*, 38, L22601, doi:10.1029/2011GL049518.
- Grevemeyer, I., Herber, R., and Essen, H.-H., "Microseismological evidence for a changing wave climate in the northeast Atlantic Ocean," *Nature*, 408, 349–351, 2000.
- Grob, M., Maggi, A., and Stutzmann, E., "Observations of the seasonality of the antarctic microseismic signal, and its association to sea ice variability," *Geophys. Res. Lett.*, 38, p. L11302, 2011. doi:10.1029/2011GL047525.
- Gruber, A., Su, C. H., Zwieback, S., Crow, W., Dorigo, W. and Wagner, W. (2016). "Recent advances in (soil moisture) triple collocation analysis," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 45, pp. 200-211.
- Gruber, A., Su, C. H., Crow, W. T., Zwieback, S., Dorigo, W. A. and Wagner, W. (2016b). "Estimating error cross-correlations in soil moisture data sets using extended collocation analysis," *J. Geophys. Res. Atmos.*, vol. 121(3), pp. 1208–1219. doi:10.1002/2015JD024027.
- Hasselmann, K., "A statistical analysis of the generation of microseisms," *Rev. of Geophys.*, 1, 2, 177–210, 1963.
- Herbers, T. H. C., P. F. Jessen, T. T. Janssen, D. B. Colbert, and J. H. MacMahan (2012), *Observing Ocean Surface Waves with GPS-Tracked Buoys*, *J Atmos Ocean Tech*, 29(7), 944-959.
- Janssen, P. A. E. M., Abdalla, S., Hersbach, H. and Bidlot, J.-R. (2007). "Error estimation of buoy, satellite, and model wave height data," *J. Atmos. Oceanic Technol.*, vol. 24, pp. 1665–1677.
- Janssen, P., L. Aouf, A. Behrens, G. Korres, L. Cavalieri, K. Christiensen, O. Breivik : Final report of work-package I in my wave project. December 2014.

- Kinzel, J., Fennig, K., Schröder, M., Andersson, A., Bumke, K. and Hollmann, R. (2016). "Decomposition of Random Errors Inherent to HOAPS-3.2 Near-Surface Humidity Estimates Using Multiple Triple Collocation Analysis," *J. Atmos. Oceanic Technol.*, vol. 33, pp. 1455–1471. doi: 10.1175/JTECH-D-15-0122.1.
- Livermont, E., 2017. Correcting changes in the NDBC wave records of the United States. 15th International Workshop on Wave Hindcasting and Forecasting, Liverpool, UK, September 10-15.
- Longuet-Higgins, M. S., "A theory of the origin of microseisms," *Phil. Trans. Roy. Soc. London A*, 243, 1–35, 1950.
- McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M. and Stoffelen, A. (2014). "Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target", *Geophys. Res. Lett.*, vol. 41, pp. 6229-6236.
- Meschede M., E. Stutzmann, V. Farra, M. Schimmel, F. Arduin. The effect of water column resonance on the spectra of secondary microseism P waves 2017 *J. Geophys. Res.*, 122, 8121-8142., doi:10.1002/2017JB014014
- Niiler, P. P. (2001), *The world ocean surface circulation*, in *Ocean Circulation and Climate*, edited by G. Siedler, J. Church and J. Gould, pp. 193-204, Academic Press.
- O'Carroll, A. G., Eyre, J. R. and Saunders, R. W. (2008). "Three-Way Error Analysis between AATSR, AMSR-E, and In Situ Sea Surface Temperature Observations," *J. Atmos. Oceanic Technol.*, vol 25, pp. 1197-1207. doi: 10.1175/2007JTECHO542.1.
- Quilfen, Y., Chapron, B. and Vandemark, D. (2001). "The ERS scatterometer wind measurement accuracy: evidence of seasonal and regional biases," *J. Atmos. Oceanic Technol.*, 18, 1684-1697.
- Scipal, K., W. Dorigo and R. deJeu, "Triple collocation — A new tool to determine the error structure of global soil moisture products," 2010 IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, 2010, pp. 4426-4429. doi: 10.1109/IGARSS.2010.5652128
- Scipal, K., Holmes, T., De Jeu, R., Naeimi, V. and Wagner, W. (2008). "A possible solution for the problem of estimating the error structure of global soil moisture data sets," *Geophys. Res. Lett.*, 35, L24403.
- Stoffelen, A., 1998: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *J. Geophys. Res.*, 103, 7755–7766, doi:10.1029/97JC03180.
- Stutzmann, E., M. Schimmel, G. Patau, A. Maggi. Global climate imprint on seismic noise. 2009, *Geochemistry, Geochem. Geophys. Geosyst.*, 10, Q11004, doi:10.1029/2009GC002619

-
- Stutzmann E. , F. Ardhuin, M. Schimmel, A. Mangeney, G. Patau, Modelling long-term seismic noise in various environments 2012, Geoph. J. Int, doi:10.1111/j.1365-246X.2012.05638.x
- Su, C. H., Ryu, D., Crow, W. T. and Western, A. W. (2014). "Beyond Triple Collocation: Applications to Soil Moisture Monitoring," J. Geophys. Res. Atmos., vol. 119, pp. 6419-6439. doi: 10.1002/2013JD021043.
- Tokmakian, R. and Challenor, P.G. (1999). "On the Joint Estimation of Model and Satellite Sea Surface Height Anomaly Errors", Ocean Modell., vol. 1, pp. 39-52.
- Vogelzang, J., Stoffelen, A., Verhoef, A. and Figa-Saldaña, J. (2011). "On the quality of high-resolution scatterometer winds," J. Geophys. Res., vol. 116, C10033, 14 p. doi:10.1029/2010JC006640.
- Wang, X. L. and Y. Feng, published online July 2013: RHtestsV4 User Manual Climate Research Division, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada. 28 pp. [Available online at <http://etccdi.pacificclimate.org/software.shtml>]
- Yilmaz, M. T. and Crow, W. T. (2014). "Evaluation of Assumptions in Soil Moisture Triple Collocation Analysis," J. Hydrometeorol., vol. 15, pp. 1293-1302. doi:10.1175/JHM-D-13-0158.1.
- Zwieback, S., Scipal, K., Dorigo, W. and Wagner, W. (2012). "Structural and statistical properties of the collocation technique for error characterization", Nonlin. Proc. Geophys., vol. 19, pp. 69-80.
- Zwieback, S., Su, C. H., Gruber, A., Dorigo, W. A. and Wagner, W. (2016). "The Impact of Quadratic Nonlinear Relations between Soil Moisture Products on Uncertainty Estimates from Triple Collocation Analysis and Two Quadratic Extensions," J. Hydrometeorol., vol. 17, pp. 1725–1743. doi: 10.1175/JHM-D-15-0213.1.