DERIVATION OF DESIGN WAVE FROM JOINT BUOY, SATELLITE AND HINDCAST DATA SOURCES

S. VanIseghem

Ifremer, Brest, France

INTRODUCTION

Design waves have been primordial for offshore designer and operators for many years and their characterization has been improved recently with the development of new sensors and new models able to give information about the sea state. To ensure a design study to be reliable, the three following steps are compulsory. The first one is a sensitivity study to determine which sea state parameter is the most relevant for the chosen structure, the second one is the parameter estimation from the available data set and the third one the determination of the estimation reliability. This methodology has been applied as part of the "HOUDIM" (Design Swell) project and some preliminary results have been presented in [8]. This article deals with the last part of the study; how to estimate the reliability and/or uncertainties of the characterization of those design waves when the information is coming from various data sources (in situ, satellite and model). The sea state parameter we have chosen to consider is the significant wave height since it remains a compelling stage to charaterize design waves.
Since buoy data are usually assumed to be the reference, our results have been presented by answering to this crucial question : if sea state characterization is needed on an area where no in situ data is available, is it possible to get reliable characterization from model(s) and /or satellite informations and if yes, how? The test area is the Iroise Sea in the Atlantic Ocean. The first part presents the methodology to make use of the various data sources. Buoy data is available on this area and will be used in this article as a reference only to estimate the reliability of the characterization obtained from the other data sources. This reference climatology is presented in the second part. The third part is relative to the hindcast data and the necessity of a "measured information" like satellite data is explained. The compulsory steps to get reliable estimations from satellite data are presented in part four. The last part is devoted to the estimation of the transfer functions between climatologies from the various data sources.
Conclusions and recommendations are proposed as to the possibility and to the way of using those various data sources to obtain a characterization as reliable as possible.

1 METHODOLOGY TO MAKE USE OF THE DIFFERENT DATA SOURCES

On a given area, reduced sea state parameters used in the sea state characterization can be relative to different sources of data, buoy, satellite or hindcast. It has been observed, see for example [7] and [5], that results differ depending on which source they come from. Potential causes for those differences have been described in [2] : differences in measurement principles, inherent limitations of the measurements principles, systematic off-sets due to incomplete calibration, sampling variability, temporal or spatial offset. Nevertheless, when various data sources are available, the use of a single source for the characterization of design wave would impair the reliability of the result and when only one single source is available it is of great importance to know whether or not a reliable characterization can be obtained from it.
The aim of the methodology is then to quantify the differences between data sources to be able to associate an uncertainty to a design value depending on which data sources have been used to get this estimation. The first step is to estimate climatologies relative to each data source. To get representative climatologies, the following four data conditions have to be satisfied :

1.1 *The four required data conditions*

- **Conditions to get representative climatology**

Condition 1 : homogeneity of the data in space (satellite) and in time (buoy and model)

Condition 2 : sufficient data resolution to avoid under-representation of events

Condition 3 : independence of events to avoid over-representation of events

- **Condition to compare climatologies**

Condition 4 : record durations relative to each data source have to be about the same and numbers of observations along the year have to be about the same.

This last assumption is not theoretically needed but we have been faced to this problem in HOUDIM project and differences have been observed between climatologies because of this difference in the record durations and not because of the sensor in itself. Since the reference period is short (< 1 year), it is compulsory to respect this assumption. The second part of the condition is also important and since the climatology can be assumed to be stationary over 30 days (trend due to season neglectible in 30 days), the number of observations for each data source on any 30 days window will then have also to be about the same.

1.2 *Methodology to compare distributions - Bias function*

A method to estimate bias between two distributions coming from different sources is suggested in this part. The sea state parameter considered is the significant wave height and will be denoted "Hs", the reference data set $(Hs)_{ref}$ and the compared climatology $(Hs)_{other}$.
The aim is to find a formula of the type :

$$Hs_{ref} - Hs_{other} = f(Hs_{other})$$

so that when the reference can be trusted but only the "other" (model or satellite) data is available, the correction $f(Hs_{other})$ can be added to the Hs_{other} distribution to make its distribution equivalent to the Hs_{ref} distribution.
The method, is based on the following statistical result:
Let a and b be two samples of sizes n_a and n_b coming from two sources of data, F_a and F_b their corresponding cumulative distributions. Let qp_a and qp_b be quantiles of order p so that $F_a(qp_a) = p$ and $F_b(qp_b) = p$. Empirical quantiles of F_a and F_b are denoted by respectively Qp_a and Qp_b
A statistical result is that

$$\sqrt{n_a} Qp_a \sim N(\nu_a, \sigma_a) \;\; with \; \nu_a = qp_a \; and \; \sigma_a = \sqrt{(p(1-p))}(\frac{dF_a^{-1}}{dp}(p))$$

If follows that

$$Qp_a - Qp_b \sim N(\nu, \sigma) \; with \; \nu = qp_a - qp_b \; and \; \sigma = \sqrt{\sigma_a^2 + \sigma_b^2}$$

(since estimation errors on Qp_a and Qp_b can be assumed to be independent). For each quantile p, a confidence interval can then be associated to the difference $Qp_a - Qp_b$, see also [4]. An estimation of the bias function between the two distribution (Hs_{ref}) and (Hs_{other}) has then been found and a confidence interval can be associated.
This methodology will be applied in the coming parts and the bias function between the reference data set and the other data sources will be estimated.

2 THE REFERENCE : BEATRICE BUOY DATA

The first data set has been recorded at Beatrice directional buoy located in the Iroise Sea, entrance of the English channel (coordinates 48.5W 5.8N). 2340 sea states of 34 minutes duration have been recorded between 1996 and 1997 with a 2Hz sampling frequency.

2.1 *The four required conditions*

- The buoy data is homogeneous over the 2 years since informations have been recorded with the same acquisition system (MRU central). Significant wave height is estimated every three hours, this temporal resolution is satisfactory and information can be assumed to be representative and event independent
- The equivalent record duration is 0.7 year, see 2.3. Distribution of data over month is presented on figure1

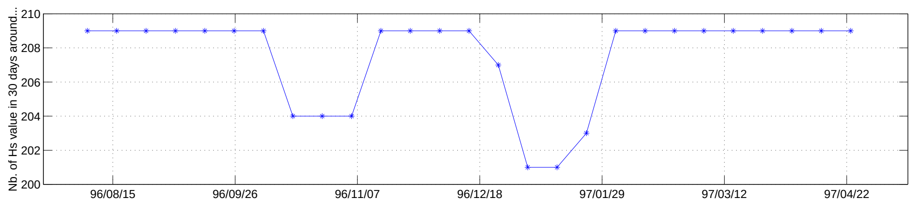

Figure 1: Distribution of the number of observations over months

2.2 *Empirical distribution*

Beatrice Hs distribution is presented on figure 2. The empirical joint distribution of Hs and $T_{1/3}$ is presented in the table below. Hs has been estimated by $H\sigma$. A large proportion of swell can be observed at that location.

Hs (m) / Tz (s)	0-2	2-4	4-6	>6	%
0-6	7	0	0	0	7
6-8	22	11	0	0	33
8-10	13	17	2	0	32
>10	9	13	5	1	28
%	51	41	7	1	100

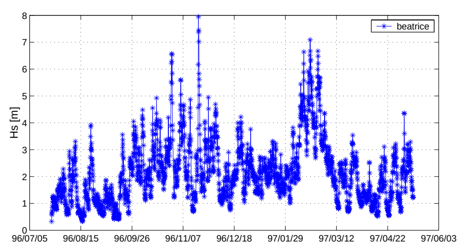

Figure 2: Empirical joint distribution of Hs and Tz at Beatrice location

2.3 *Extreme Hs distribution*

As it has been said in the previous part, 2340 sea states have been recorded. The stationary period of a sea state will be assumed to be 3h. From that we can estimate the equivalent time during which the "Beatrice" climatology has been observed : $T_{eq} \ = \ 3 * 2340 \ = \ 7020\, h \ \ = \ 0.7\, year$
The extreme value distribution of Hs can then be estimated by : $(Hs)_{max} \sim (Hs)^T \ with\ T = \frac{nb.obs.}{T_{eq}}$
Both (Hs) distribution and $(Hs)_{max}$ distributions are presented on figure 3.

3 HINDCAST DATA

Numerical models such as Hindcast models provide sea state information in a very convenient way, reduced parameters Hs are given at any point every six hours. It is also important to keep in mind that this information is dependent on some modeling assumptions. Uncertainties associated to the sea state characterization when this single data source is used are explained in the coming part.

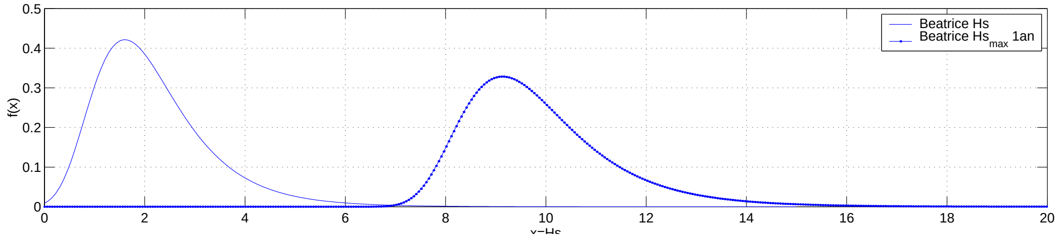

Figure 3: Reference (Hs) distribution and $(Hs)_{max}$ distributions

The Hindcast model used in this article corresponds to the hindcast from the AES40 Wind and Wave Reference Climatology which was funded by Climate Research Branch of Environment Canada, see [6] for details. The period considered correspond exactly to the recorded period of the buoy, from the 22 july 96 to the 13 may 97. The two point model number 5810 (48.1W 5.8N) and 5875 (48.9W 5.8N) are considered. They are distant from about 80km, see figure 5(right).

3.1 *Conditions required*

The model data are homogeneous in time since the algorithm has not evolved between 96 and 97. The temporal resolution is 6 hours which is satisfactory to get a reliable estimation. The equivalent time recorded is 0.35 year. The mean length of stationarity has been found to be 110km (see 4) and a large variability is associated to this mean value. An equivalence between the 3hours duration of stationarity and the 110km length of stationarity can be made and it can then be assumed that the dependence structure between Hs values given any 3hours is the same than the one corresponding to Hs values given any 6hours and from two location distant from about 100km. Information from the 2 model points distant from 80 km have thus been mixed together. The equivalent time record is now approximately equivalent to the one from the buoy (see 1.1). The number of information over each 30 days windows is constant, equal to 30*2*4=240, equivalent to the number of the reference data source. $(Hs)_{hindcast}$ and $(Hs_{max})_{hindcast}$ distributions can then be estimated and compared to the reference distribution.

3.2 *Observed differences between hindcast and reference*

The f function defined in 1.1 by $Hs_{ref} - Hs_{hindcast} = f(Hs_{hincast})$ is estimated in this part :

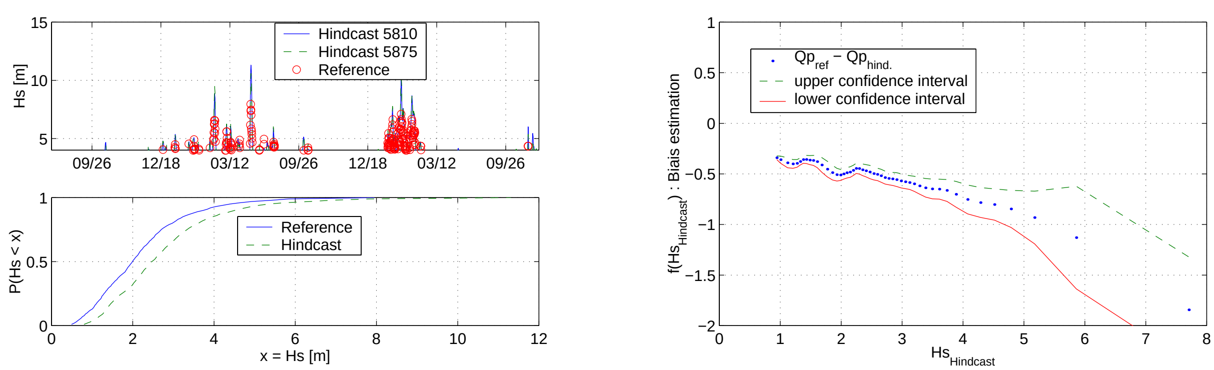

Figure 4: (Hs) distributions - bias between buoy and hindcast

Attention has to be paid on the fact that the higher the quantile is, the less robust the estimation of f is, because of the number of points considered to obtain the estimation. For example, the last point on figure 4 (right) corresponds to $(Q_{0.98})_{ref} - (Q_{0.98})_{hind} = -1.9m$ but its estimation is based on about 40 observations only. Some differences appear, hindcast gives higher Hs estimation than the buoy especially for large Hs . This is illustrated in the next section for a specific given date. A sea state characterization that would rely on Hindcast data only lead then to possible bias in this characterization.

3.3 *On the necessity to take into account satellite data*

Four days of data have been selected, from 97/23/02 to 97/27/02 (written 27/02/97 on figure). Comparison between reference and hindcast data is presented on the top left figure 5. Information provided from the two satellite Ers2 and Topex is presented on the bottom left figure 5 and the satellite location at those dates on the right part of figure 5.

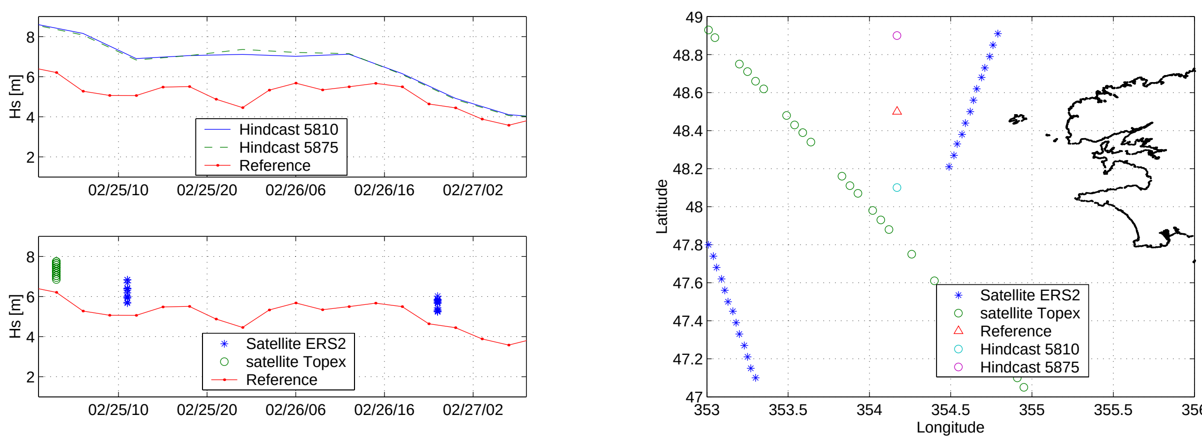

Figure 5: Information available with the satellite

Informations from the satellite agree better with the reference estimation than the hindcast information. Satellite observations are thus of great importance and have to be considered. They can not be used as easily as hindcast data and some intermediate analysis need to be untertaken but this work is worth doing to reduce the characterization uncertainties of design waves.
It has been seen that hindcast data are not sufficient to get reliable estimation. An other question is investigated in the coming part : would it be reasonnable to consider the single satellite altimeters source in order to characterize the climate?

4 SATELLITE DATA

The satellite sensor considered in this article is the altimeter. Between July 96 and May 97, three altimeters are available, ERS1, ERS2 and Topex. Altimeters give Hs estimations along track every 7km with a return period equal to 10 days for Topex and 35 days for ERS1 and ERS2. Successive estimations along a track can not be assumed to be independent. There are thus two ways to proceed to ensure the condition of independent events to be satisfied. The first one is to consider an area where the climate can be assumed to be homogeneous, say about 150x150km, and to consider only one observation on each track, either the median or the closest to the reference buoy. The second way is to consider a larger area and to apply a correction to ensure climate homogeneity. Since longer tracks are obtained, a rupture detection algorithm can be applied to keep information for each detected independent 'event'. Presentation of those two methodologies is made in the following part.

4.1 *Method I : homogeneous area*

Let us consider the 2deg.x2deg. area (47-49N /353-355W), see figure 6 (left), on which climate can be assumed to be homogeneous. To ensure the independent events condition to be satisfied, only one Hs estimation on each satellite track is considered. The number of different tracks on this area is 2 for Topex and 8 for ERS1 and ERS2. So Hs estimation is available on average once every 5 days for Topex and once every 4.5 days for ERS1 and 2.

On this homogeneous area, what effective climate period can be seen from various satellites during this 0.7year (=255days) of reference observed period? This simple calcution gives the result :

$$\text{Number of events observed by satellites} =$$
$$[(Nb._{obs/day})_{Topex} + (Nb._{obs/day})_{ERS1} + (Nb._{obs/day})_{ERS2}] * (Nb_{day})_{of obs.} \sim 165\, events.$$

If we assume that an event corresponds to 3h in time, the climate period observed from the three satellite along the 0.7 year is about 21 days. To get an observed period of the order of 0.7 year we then would need about 10 years of satellite observation with all three altimeters and that is not presently possible.

With the increasing number of altimeters in space, this simple method could be useful in the futur. At the present time and with the data available, it appears that the event selection is too restrictive to get a satellite distribution with an equivalent observed period of the same order as for an in situ distribution.

4.2 *Method II : larger area + bias correction*

We consider now the area presented in figure 6 (right) and the only satellite considered is ERS2 to avoid cross-calibration problems. They are about 14 tracks from ERS2 on the chosen area for each cycle. The mean length of a stationary event is about 110km (result obtained as part as HOUDIM project) so about 4 independent events on each ERS2 track can be assumed to be extracted and ERS2 can then see about two equivalent months of data on this area during 0.7 year. A finer analysis will be performed in 4.2.2, the aim of this crude approximation is only to justify the choice of the area. It will then be possible to compare the satellite climatology with the reference one if we consider about 4 years of ERS2 satellite data. The compulsory step for the climatology to be reliable is to transform the area into an homogeneous one, as studied in the coming part.

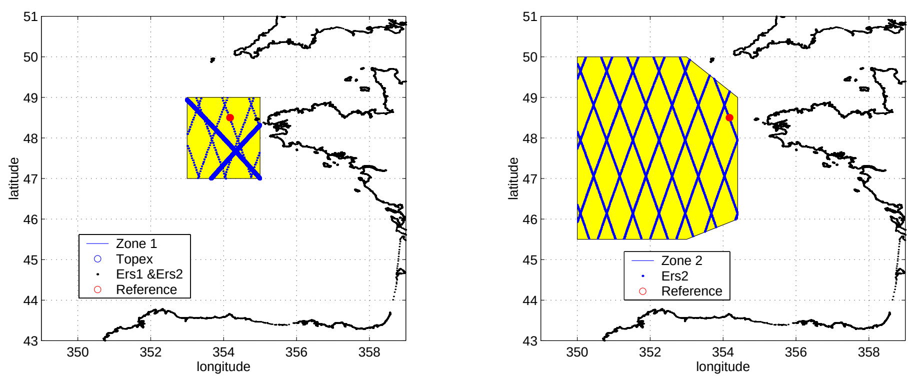

Figure 6: The two areas considered

4.2.1 Bias estimation and application

◇ ***Estimation***

To estimate the geographical bias, 4 years of ERS2 data are considered and the area is split into 38 parts, the centers of which are represented by 'o' on figure 7 (left), at the crossings of the tracks. Hs distributions are estimated on each of these parts. Median values are plotted on figure 7 (right).

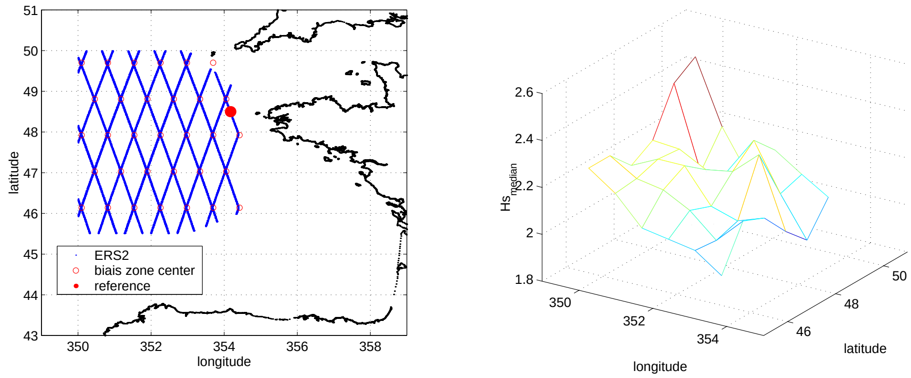

Figure 7: Bias estimation

Two Hs cumulative distributions are presented.
First one - center = 49.8N, 350W
Reference one - including Beatrice buoy.
Since we want the area to be homogeneous, a transfer function will be applied to each non reference climatology so that we obtain 38 comparable climatologies.
This can be written as follows :

$$Hs_{zone\ ref} - Hs_{zone\ x} = f(Hs_{zone\ x})$$

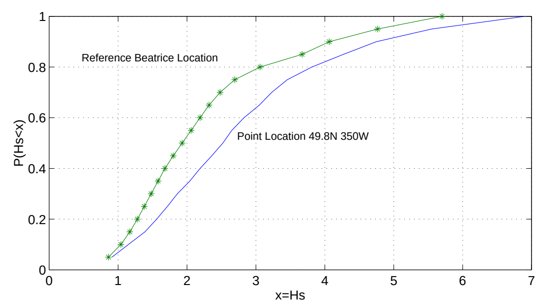

Figure 8: Bias correction

⋄ *Application and results*

Figure 8 presents Hs cumulative distributions for two of the 38 zones. The 37 tranfer functions corresponding to the 37 non reference zones are presented on figure 9 (left). Results are presented on figure 9 (right). The 38 median values have been plotted and it can then be concluded that once those corrections have been applied, the area can be assumed to be homogeneous.

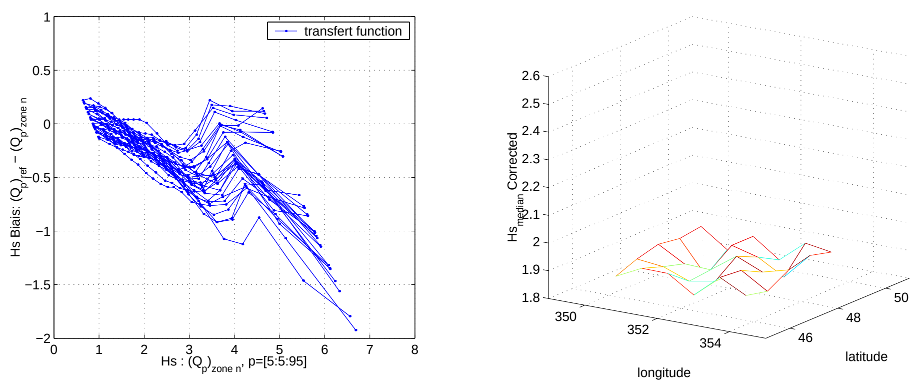

Figure 9: Tranfer functions of the 37 points(left)/ Homogeneity after application(right)

4.2.2 Event decomposition

It has been seen that the choice of a single point for each track was too restrictive to get a reliable climatology, see 4.1, and if no selection is made, the informations are not independent. A rupture detection algorithm, based on the Hinkley statistical test, see [1], has then been implemented along each track.

Hinkley test can be described as follows : A stationary process can be written on the form $H_s(i) = M + \epsilon(i)$. Hypothesis H0 : $\epsilon(i) \sim N(0, \sigma^2)$ is tested against H1 : $\epsilon(i) \sim N(\nu, \sigma^2)\ with\ \nu = aM + b$. A rupture is detected when H0 is rejected, see also [4] for details.

During the four years of ERS2 data, between 1996 to 1999, we have about 40.000 Hs estimations on the chosen area. The rupture detection algorithm detects about 2400 events so the equivalent observed period equals to about 0.75 year. Figure 10 (left) presents the detected lengths of the stationary segments with respect to Hs, figure 10 (right) shows examples of ruptures detection along six tracks.

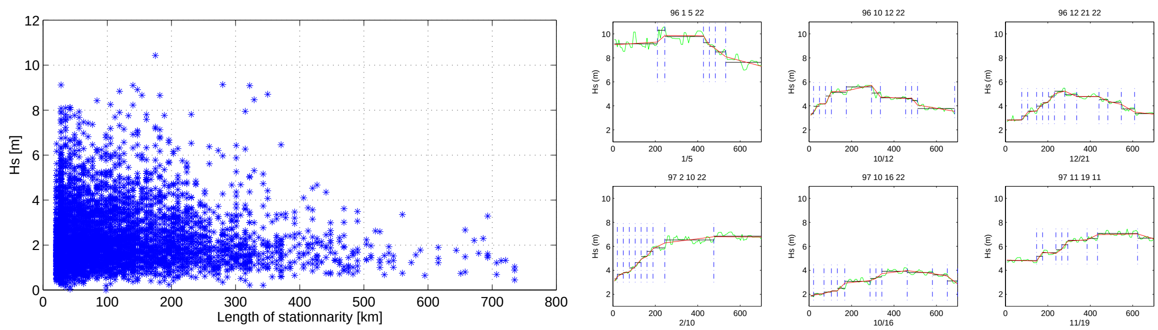

Figure 10: Event decomposition

4.2.3 Conditions required

We do not have considered the data over the months of May June and July to get a number of observations along the year equivalent to the one relative to Buoy data and accounting for the buoy not functioning at that time. Eventually the equivalent total time recording is 0.75 year comparable to the 0.7 year of in situ data. Between 250 and 300 events have been detected in any given month. This is satisfactory even if this variability should be reduced to improve the quality of the methodology.

Informations are homogeneous in space (only the ERS2 satellite has been considered) and in time since, according to hindcast data, no trend appear in the Hs distribution between 1996 and 2000. Informations are now independent and can be used to estimate the satellite distribution.

4.2.4 Climatology estimation

Empirical Hs distribution is plotted on figure 11 (left). The estimated transfer function between satellite and the reference and its corresponding confidence interval are displayed on figure 11 (right).

5 Climatologies from the three data sources - Discussion

It is now possible to give an answer to the question raised in the introduction : if sea state characterization is needed on an area where no in situ data is available, is it possible to get a reliable characterization from model(s) and/or satellite information and if yes, how?

The simple way to answer is "just" apply the corresponding transfer function, see figure 11 (right) to the given data set, satellite and hindcast. One will then get a distribution equivalent to the in situ one and the design value obtained by extrapolation will be "as reliable" as an "in situ design characterization".

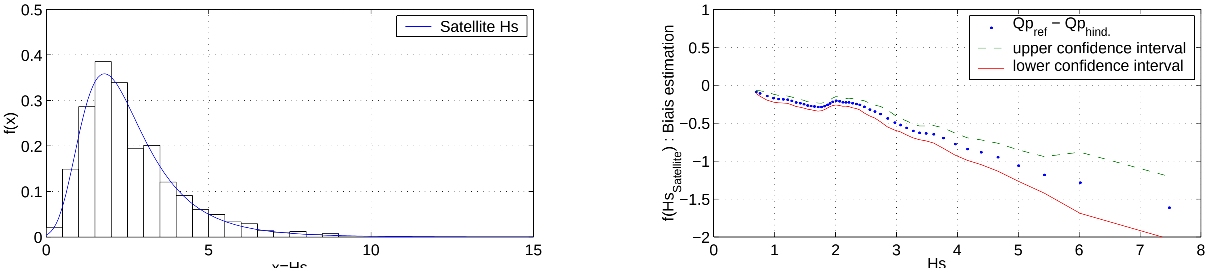

Figure 11: Satellite :(Hs) distribution - Comparison with the reference distribution

It is also interesting to look at the three climatologies simultaneously. They are presented on figure 12 (right). The reference distribution has lower Hs values than hindcast and satellite distributions.
We consider now the satellite and hindcast data sources together. The bias function between the two distributions has been estimated in figure 11 (left). This estimated bias does not exceed 30cm and the simultaneous use of those two data sources would then lead to accurate sea states characterization.

At that stage, it is not possible to decide which one from those two distributions (in situ and hindcast-satellite) is the "true" one. And even, the most probable is that both distributions include some bias.
This leads to an interrogation about the reference choice : the reference is "commonly assumed" to be the buoy but this example shows that this choice is not compulsory. The reference for a design study could then be assumed to be either satellite or hindcast data if the required conditions explained in this article are satisfied.

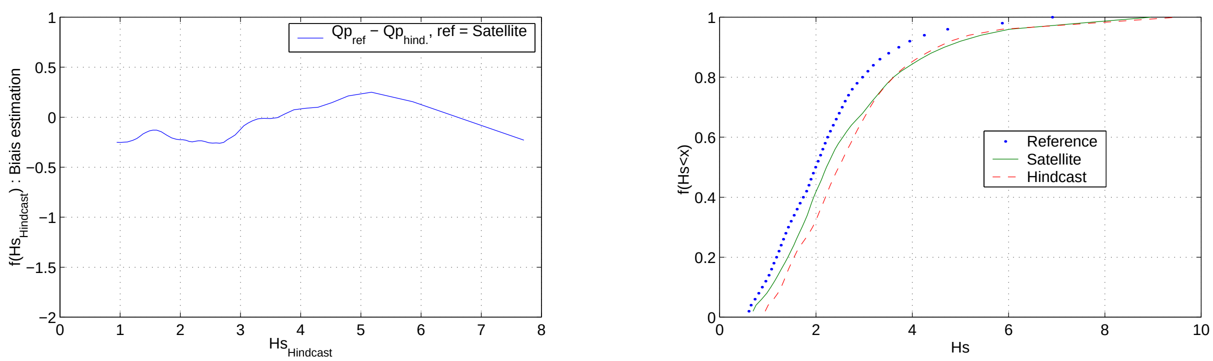

Figure 12: Climatologies from the three data sources

6 Concluding remarks

The question of the reliability of the design wave characterization has been investigated in details in this article. The choice which is commonly made to consider the in situ data set as the reference one has been discussed and it has been concluded that this choice is not compulsory.
The use of satellite and hindcast data with their corresponding advantages and disadvantages have been studied. Hindcast data are dependent on some models assumptions and are not sufficient to predict in an accurate way large Hs values. Satellite information requires many intermediate analyses to satisfy the required conditions. Once the data selection is performed and if the equivalent observed period is sufficient, characterizations of design waves deduced from satellite data exhibit very low uncertainties. It is then possible to get reliable characterization from model(s) and /or satellite information.

In order to decide in our example which one from those two distributions (in situ and hindcast-satellite) is the "true" one, a short additional in situ campaign close to Beatrice buoy appears to be the best solution and would give an answer to an interrogation which is important for all the people involved in maritime activities at this location, entrance of the English Channel.

The methodology based on both model and satellite information will be tested at an other location to analyse the transfer function variability between different locations. Three methodologies to get design wave characterization without in situ data are presented on the figure 13. The one which appears to be the most convenient with low uncertainties uses both satellite and hindcast information.
Further work is on-going in the metocean team at Ifremer to get an operationnal methodology that use hindcast data and satellite data when "available", available being studied in details by taking into account additional parameters like for example hindcast wave directionality.

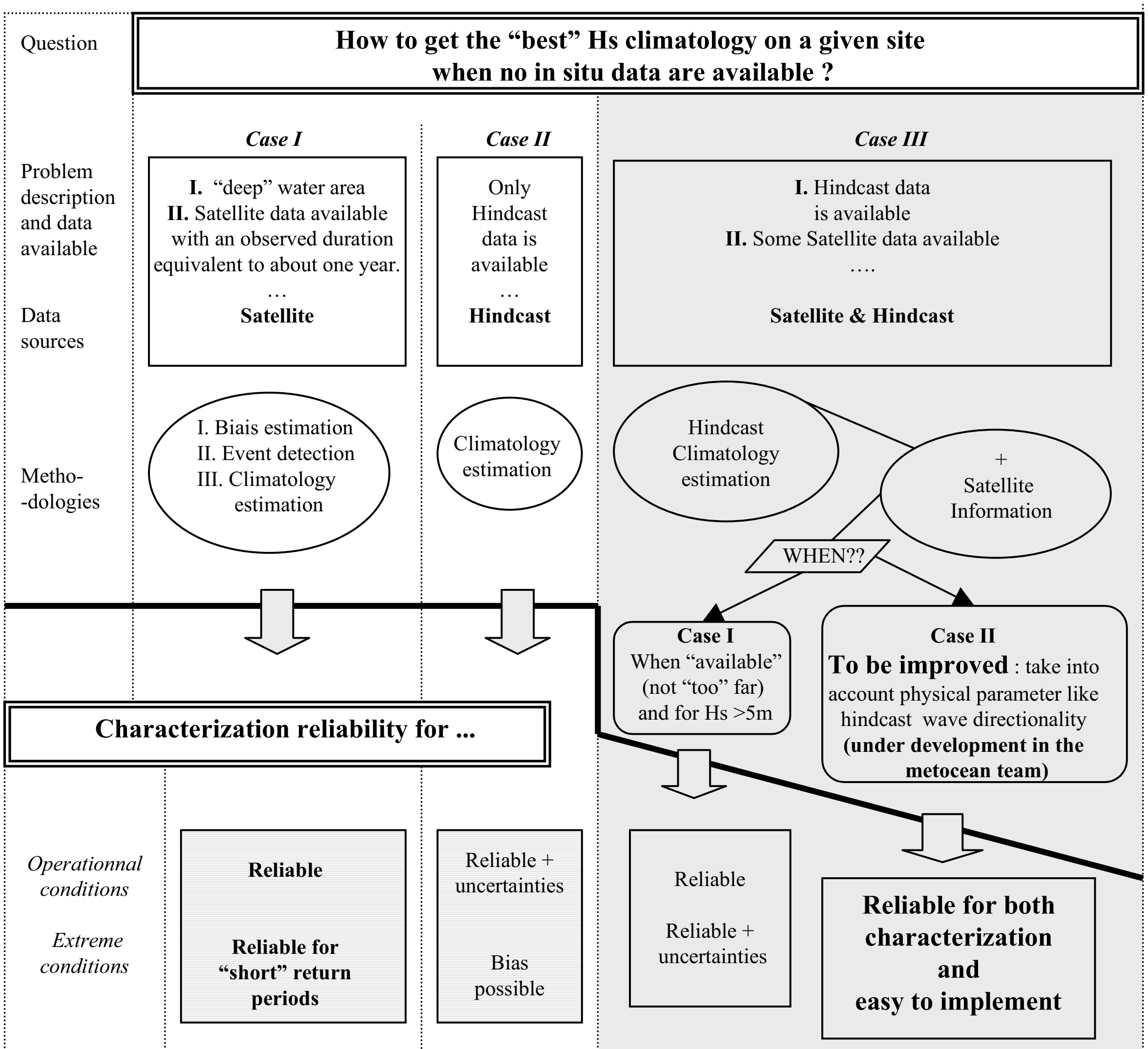

Figure 13: Three different methodologies

Aknowledgements

The author would like to thank Dr. Cardone (OceanWeather) for providing the wave hindcast from the AES40 Wind and Wave Reference Climatology at Béatrice location and G. Goasguen (CETMEF) for Béatrice buoy data set.
Special thanks to J. Ryden (Lund University and post doc. Ifremer) for fruitful discussions on statistical methodologies to merge informations from various data sources.

References

[1] **Hinhley D. V.**, "Inference about the change point in a sequence of random variables" *Biometrika*, 57, p1, 1970.

[2] **Krogstad H.E.**, "Methods for intercomparison of wave measurements" *Coastal Engineering* , 37, p235, 1999.

[3] **Labeyrie J.**, "Stationary and transient states of random seas" *Marines Structures* ,1990.

[4] **Olagnon, M.**, "Use of satellite data for metocean design parameters estimation" *Proc. Oceanology International 96, Brighton (1996)* , Vol. 2, pp. 121-132

[5] **Panchang V., Zhao L.** "Estimation of extreme wave heights using Geosat measurements" *Ocean Engineering*, 1997.

[6] **Swail, V.R. & Cox, A.T.** "On the use of NCEP NCAR reanalysis surface marine wind fields for a long-term north atlantic wave hindcast" *Journal of Atmospheric and Oceanic Technology*, vol 17, 1999, num. 4, pp. 532-545.

[7] **Tournadre J., Ezraty R.** "Local climatology of Wind and Sea State by means of Satellite Radar Altimeter Measurements" *J. Geophysical Research*, 1990.

[8] **VanIseghem, S. & Deleuil, G.** "Improved characterization for design waves" *Proc. I.S.O.P.E, Stavanger*,2000.

Corresponding author:
Sylvie VanIseghem – TMSI/IDM/COM – Ifremer – B.P. 70 – F-29280 PLOUZANÉ (FRANCE)
fax: (+33) 298 22 4650 – email: Sylvie.Van.Iseghem@ifremer.fr