SAMPLING BIAS IN THE ESTIMATION OF SIGNIFICANT WAVE HEIGHT EXTREME VALUES

Fabio Dentale[1,2], Ferdinando Reale[2], Felice D'Alessandro[3], Leonardo Damiani[4], Angela Di Leo[2], Eugenio Pugliese Carratelli[1,2] and Giuseppe Roberto Tomasiccchio[3]

It has been shown before, and it is intuitively evident, that in a Significant Wave Height (SWH) time series, the longer the sampling interval, the lower is the number of events which are above a given threshold value. As a consequence, the use of data with a low time resolution (such as a 3 h sampling, for instance) causes a considerable undervaluation of the extreme SWH values for a given return time RT. In this paper an example of such a bias is provided, and a method is suggested to estimate it on a regional basis. Results may help to improve the use of historical wave meters data which were often collected with a low time resolution, and may also provide a tool to improve the application of Numerical Meteo-Wave models to the evaluation of extremes.

Keywords: extreme waves; sampling bias; wave measurement

INTRODUCTION

The evaluation of the extreme values of Significant Wave Height (SWH) and of their dependence on return period is an all-important step in the design of ships, coastal and offshore structures, as well as in the risk assessment of ship routing.

The basic data for this kind of studies are generally provided by in-situ wave meters, satellite altimeters, meteo-wave models, or by a combination of the three (Chen et al. 2013; Feng et al. 2014a; Mentaschi et al. 2013; Xu et al. 2015). All these sources are affected by errors in various ways and to different extents: the limitations of models and of satellite altimeters as a source of historical data are obvious and are discussed elsewhere (Ganguly et al. 2015; Passaro et al. 2015; Sartini et al. 2015a; Kudryavtseva and Soomere 2016). In situ wave-meters, when available, are normally the best choice but ï as it will be shown in the following ï in certain circumstances they are also affected by a strong bias when used to determine extreme values. A bias in the determination of extreme values is indeed present in all sources of data whenever the sampling of the relevant parameter (SWH in our case) is carried out with too long a time interval as compared with the inherent time constant of the phenomenon (in our case the storm evolution). Fig. 1 exemplifies the problem by comparing SWH values as measured at 30ô and 3 hours intervals.

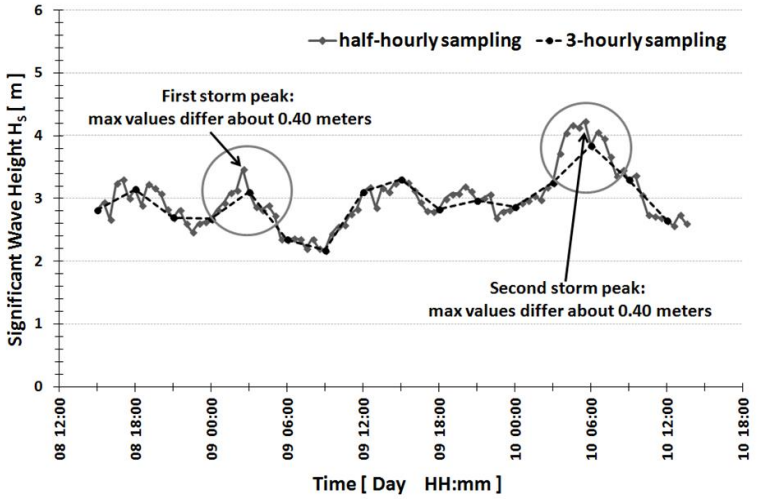

Figure 1. Wave storm in the Tyrrhenian Sea recorded by Ponza (Italy) buoy on 8th-10th November 2010. SWH (H_s) time evolution is sampled with 30ôand 3h time intervals. The maximum values are shown to differ by about 0.40 meters.

[1] CoNISMa - National Inter-university Consortium for Marine Science, Piazzale Flaminio, 9 ï Rome, 00196, Italy
[2] MEDUS - Department of Civil Engineering, University of Salerno, Via Giovanni Paolo II, 132, Fisciano (SA), 84084, Italy
[3] Department of Engineering, University of Salento, Via per Monteroni, Lecce, 73100, Italy
[4] DICATECh, Technical University of Bari, Via Edoardo Orabona, 4, Bari, 70125, Italy

While the general trend of the storm evolution is the same with both sampling intervals, it is also clear that short term oscillations of the storm intensity (Reale et al. 2014) are not revealed by the data taken with a coarser time interval. The wind field is indeed subject to random oscillations which can sometime be considered as generated by wind ñgustynessò (Abdalla and Cavaleri 2002).

The 30ô sampling shows two peaks, the first at 02:30 on 9th November (H_S = 3.47 m) and the second at 05:30 on 10th November with H_S = 4.23 m. If the sampling interval is reduced to 3 hours, the maximum values at the two peaks are respectively 3.10 m (03:00 on 9th November), with a difference of 0.37 m, and 3.84 m (06:00 on 10th November), with a difference of 0.39 m.

It is interesting to note, even if it is only indirectly relevant to the present work, that a similar problem also affects satellite SWH data. Spatial SSSV (Small Scale Storm Variations) are often revealed by altimeter tracks (Reale et al. 2014). See for instance Fig. 2 which shows a typical example of spatial SSSV between CryoSat altimeter data and ECMWF analysis data.

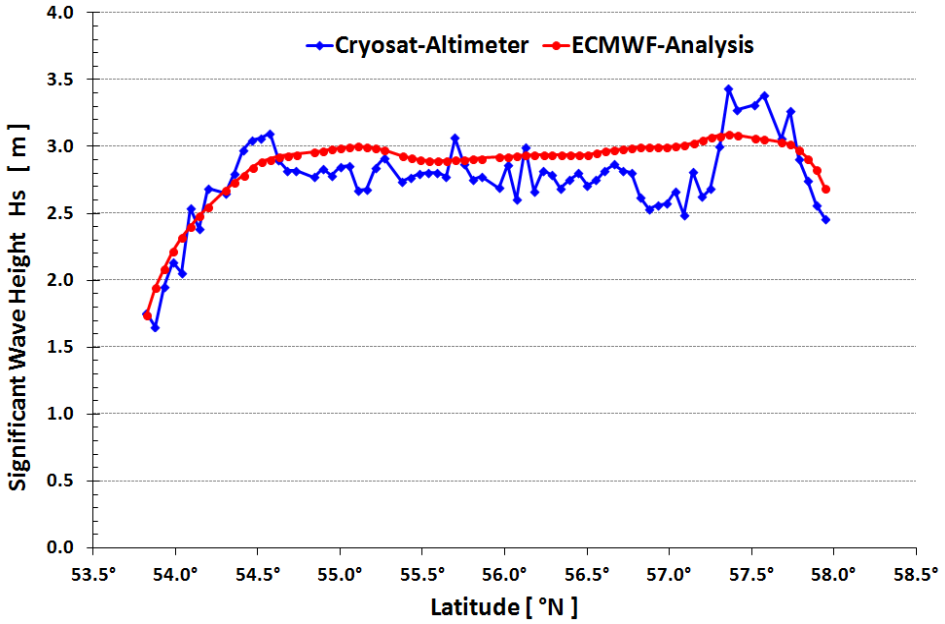

Figure 2. SWH (H_S) data from CryoSat radar altimeter (Cycle 36, Passage 811) and ECMWF analysis on 22th January 2013 at 02:48 in the North Sea.

In any case, generally speaking, it is reasonable to assume that a coarse sampling interval leads to an undervaluation of the maximum value (Cavaleri and Bertotti 2006). This paper reports on an empirical investigation on wave buoy data, aimed at clarifying such bias and at providing some information on its extent and relevance. The work is restricted to wavemeter data, since other sources are also affected by others sources of errors, which have to be dealt with separately.

METHODOLOGY

The analysis was carried out on 6 wave buoy data provided by the Italian National Wavemeter Network (RON): Alghero, Catania, Cetraro, Crotone, Mazara, and Ponza.

These wave buoys have been operating for many years, and between 1989 and 2008 have provided SWH values (H_s) at 30ô intervals. Table 1 shows the data availability for RON buoys considered. Some data are however missing due to various reasons.

Table 1. Italian National Wavemeter Buoys (RON - ISPRA) data series as used for this paper.

Buoy	Start	End
Alghero	01 - July - 1989	05 - April - 2008
Catania	01 - July - 1989	05 - October - 2006
Cetraro	01 - January - 1999	05 - April - 2008
Crotone	01 - July - 1989	15 - July - 2007
Mazara	01 - July - 1989	04 - April - 2008
Ponza	01 - July - 1989	31 - March - 2008

All the original data considered were collected at 30ô time intervals; they were then degraded to 1h, 3h and 6h interval by taking respectively a single SWH every two, six and twelve recorded values. It is worth noting that since SWH is itself an averaged parameter which has to be estimated over a certain duration of time, conceptually there is no proper ñtrueò value; however, the 30ô sample in this context will be considered to be the ñtruthò, since it would not make sense to consider a shorter time interval.

The following example - built upon the data of the RON buoy of Ponza - graphically shows how the sampling interval influences the duration curve (Fig. 3).

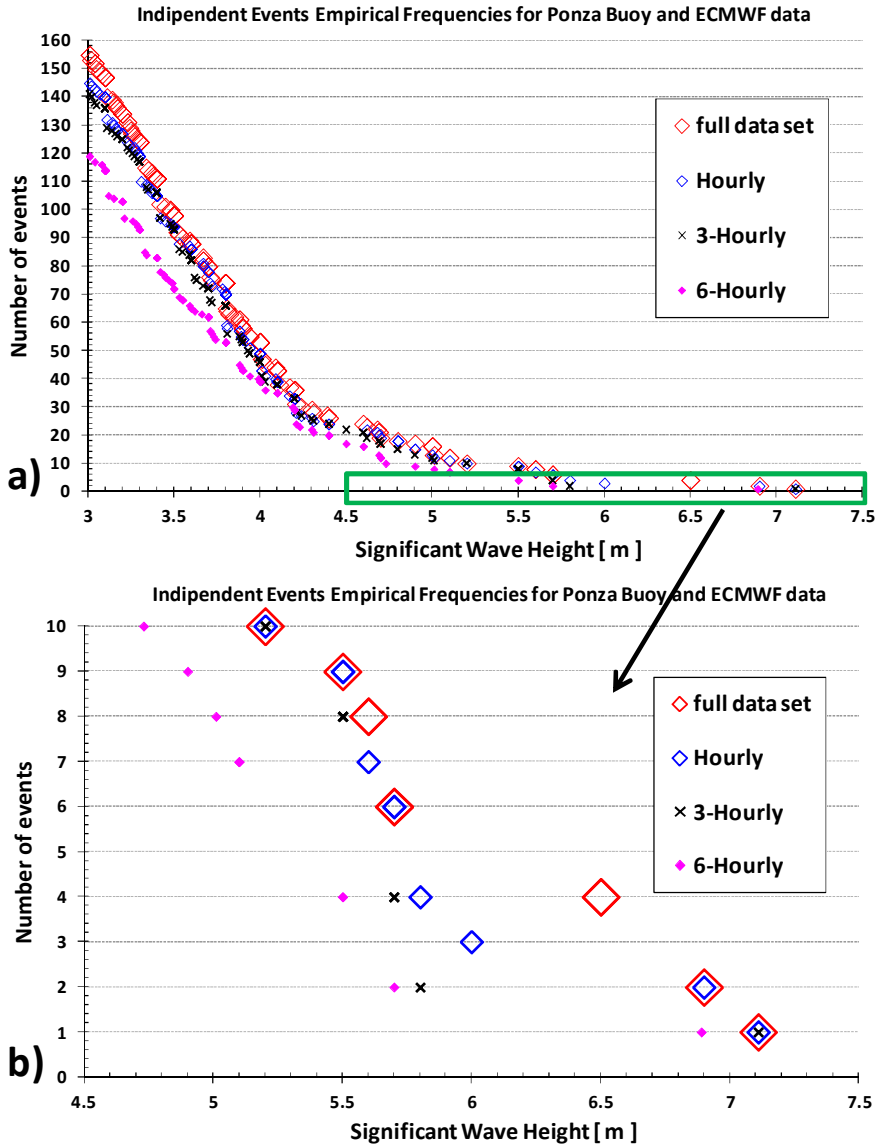

Figure 3. Number of events above a given SWH threshold for Ponza RON buoy Wavemeter: a) higher than 3 meters; b) higher than 4.5 meters.

It is visibly evident that the longer the sampling interval, the lower the number of events which are above a given value, and therefore the duration over such a value.

It is thus to be expected that the choice of the sampling interval will also influence the result of extreme value calculations. An investigation was therefore carried out on all the data available on the above mentioned buoys.

The Peak Over Threshold (POT) method was applied: a standard procedure, as described for instance by Sartini et al. (2015b), was followed to produce the Weibull distribution (Eq. 1).

$$F(H) = 1 - \exp\left[\left(\frac{H-B}{A}\right)^{k}\right] \quad (1)$$

where *A*, *B* and k are known respectively as scale, position and shape parameters. Parameters are estimated by choosing statistically independent storm peaks over a given threshold: after ordering the N_T independent storm peak SWH, (indicated in the following as H_i), in decreasing order, with $i = 1, 2, ..., N_T$, the empirical frequency F_i of each is taken to be (Eq.2):

$$F_i = 1 - \frac{i-\alpha}{N_T+\beta} \quad (2)$$

α and β are computed following Sartini et al. (2015b) as:

$$\alpha = 0.20 + \frac{0.27}{\sqrt{k}} \quad (3)$$

$$\beta = 0.20 + \frac{0.23}{\sqrt{k}} \quad (4)$$

By introducing the reduced variable *y* given by Eq. 5

$$y = \frac{H-B}{A} \quad (5)$$

then its value y_i for the *i*th empirical frequency F_i is calculated by the following Eq. 6

$$y_i = \left[-\ln\left(1-F_i\right)\right]^{\frac{1}{k}} \quad (6)$$

and correspondent estimated values $\overline{H}_i$ by Eq. 7:

$$\overline{H}_i = Ay_i + B \quad (7)$$

Parameters A and B are estimated by minimizing the function $S(A,B)$ given by Eq. 8:

$$S(A,B) = \sum_{i=1}^{N_T}\left(H_i - \overline{H}_i\right)^2 \quad (8)$$

which represents the squared difference between the independent storm peaks H_i and the correspondent empirical frequency F_i estimated values $\overline{H}_i$. The value for the shape parameter distribution was taken as $k = 1.40$ i.e. the best fit value for our data.

Once the distribution parameter are known, the SWH value $H(T_r)$ for a given return period T_r (in years) is evaluated through Eq. 9:

$$H\left(T_r\right) = B + A\left[\ln\left(\lambda T_r\right)\right]^{\frac{1}{k}} \quad (9)$$

where λ is the average yearly number of peaks over the threshold, given by the ratio between the total number of events N_T and the observation length (in years).

An example is shown in Fig. 4 which provides the plots of Eq. 9 for various data set in one of the stations (Alghero).

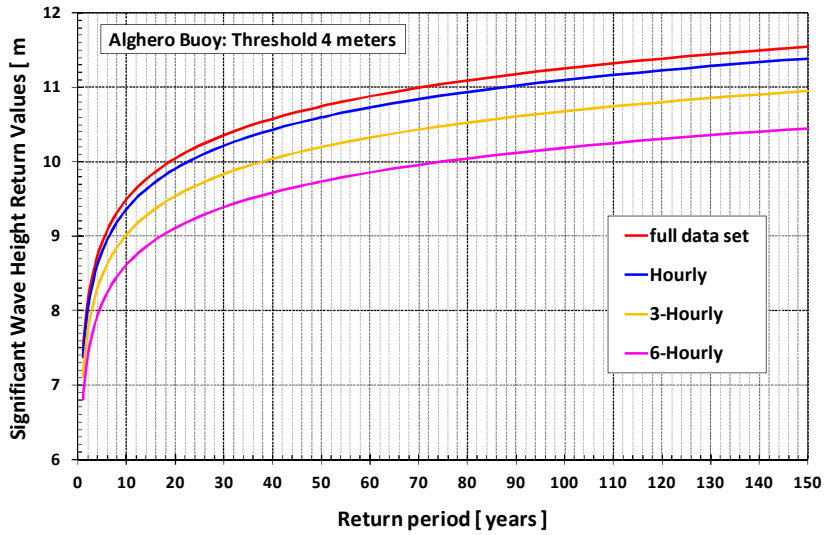

Figure 4. Weibull interpolated H(T$_r$) values vs. return periods T$_r$ for all the data and for various sampling intervals for Alghero RON buoy wavemeter.

There is an obvious trend in the results: as it was to be expected, the higher the sampling interval, the greater the distance from the full data set, i.e. from the 30ô samples. The difference is relevant and can lead in most circumstances to an important under-evaluaution of design waves. The 100 years return wave, for instance, as estimated from the 6 hours data would be one meter lower than the value estimated with the ñtrueò 30ô full data set.

Since, as stated above, in many actual design problems only data with a low sampling rate are available (such is for instance the case of model-generated sea states), it is extremely important to be able to estimate the error deriving by such undersampling. To this end, we have made use again of Eq. 9 to evaluate the SWH as a function of return period T_r for all the available data series.

Indicating by H_r^s the Weibull value of the extreme SWH for a given station r and a given T_r computed from data with sampling interval s (e.g. 1h, 3h, 6h), and by H_r^a the ñtrueò value computed with the whole data sets, i.e. with a 30ô sampling interval, the error is:

$$E_r^s = H_r^s - H_r^a \tag{10}$$

i.e. the difference between the extreme value derived from the undersampled series and the ñtrueò value normalising with H_r^s yields the ñ Estimated Relative Errorò e_r^s (Eq. 11)

$$e_r^s = \frac{H_r^s - H_r^a}{H_r^s} \tag{11}$$

e_r^s is in turn treated as a random variable whose statistical distribution is common to all the data sets from the various buoys in a given geographical area. The statistical parameters of such a distribution, i.e. its mean ε_e^s and its standard deviation $\mathring{u}_e^s$, are estimated as:

$$\varepsilon_e^s = \frac{1}{N}\sum_{r=1}^{N} e_r^s \tag{12}$$

$$\mathring{u}_e^s = \sqrt{\frac{1}{N}\sum_{r=1}^{N}\left(e_r^s - \varepsilon_r^s\right)^2} \tag{13}$$

where N the number of wavemeter buoys considered (6 in this example).

So if H_r^s is known, the expected value of H_r^a, H_r^{est} can be evaluated as

$$H_r^{est} = H_r^s + H_r^s \cdot \mu_e^s \qquad (14)$$

It is therefore an easy task to estimate H_r^{est} from the six hours data H_r^6 by making use of Eq. 14.

Results are shown in Fig. 5 where the Extreme Significant Wave/Return Time functions for H_r^{est} H_r^6 and H_r^a are compared for all the available stations.

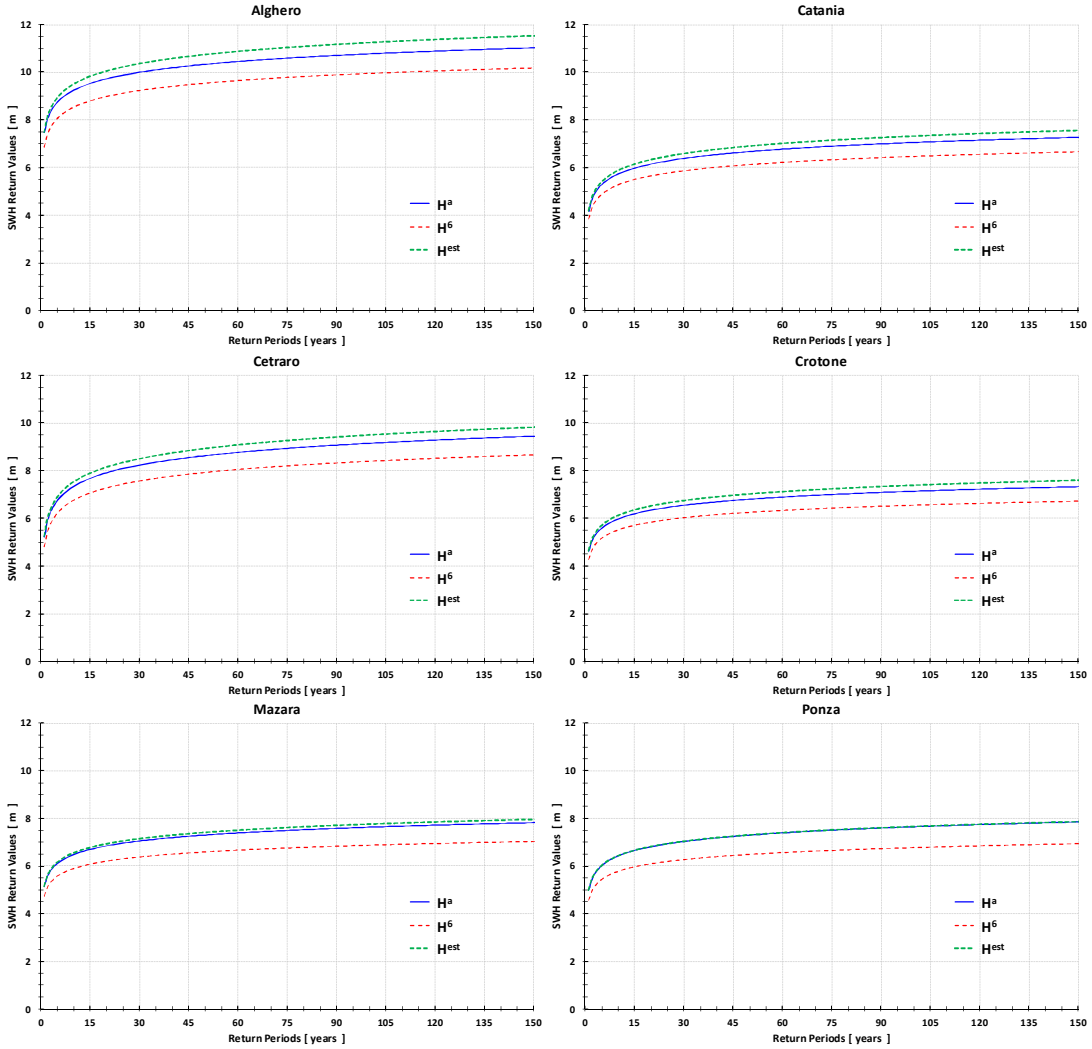

Figure 5. SWH as a function of return period T_r computed from 30ô data H^a (blu line), from 6 hours data H^6 (dashed red line) and estimated data H^{est} from Eq. 14 (dashed green line).

The results are quite impressive, as they seem to show that very good estimates of extreme Significant Wave Heights can be obtained by combining low time resolution data from a single location with error estimates for the whole area; this would imply that the statistical distribution of the Relative Errors is quite consistent. This conclusion, however, might be only hold for the particular West Mediterranean area considered, so the procedure will have to be tested in various sea locations.

CONCLUSION

As the sampling interval of the Significant Wave Height measurements increases, the probability that extreme values may be missed increases as well. Since past historical data records often provide just data sampled at 3 or 6 hour intervals, the estimation of high return time wave heights can be seriously biased, as compared with high density data ï half an hour or less: an application which is of growing importance as wave climate changes are being investigated (Feng et al. 2014b; Passaro et al. 2015; Liang et al. 2016). A method has been supplied to estimate such a bias, and to compute Significant Wave Heights as a function of the return period.

The approach we presented may also provide a tool to improve the application of model generated wave data to the evaluation of extremes, most of the historical data of such models are only available at coarse time intervals.

An obvious future development lies on the necessity of introducing an estimation of the probability of the Relative Error itself by considering not only its expected value but also it variance; it should therefore be possible to estimate a given probability of its exceedance value, thus giving a fuller picture of the incertitude of extreme Significant Wave Heights.

ACKNOWLEDGMENTS

Most of the work described in the paper has been carried out within CUGRI (University Joint Research Centre on Major Hazards). The Authors are grateful to R. Inghilesi, G. Nardone (APAT-ISPRA, Italian Environmental Agency) and L. Torrisi (CNMCA-Italian Air Force Meteorological Office), for useful data and helpful discussions.

REFERENCES

Abdalla, S., and L. Cavaleri. 2002. Effect of wind variability and variable air density on wave modelling, *Journal of Geophysical Research*, 107(C7), 17/1-17/7.

Cavaleri, L., and L. Bertotti. 2006. The improvement of modelled wind and wave fields with increasing resolution, *Ocean Engineering*, 33(5-6), 553-565.

Chen, C., J. Zhu, M. Lin, Y. Zhao, X. Huang, H. Wang, Y. Zhang, and H. Peng. 2013. The validation of the significant wave height product of HY-2 altimeter-primary results, *Acta Oceanologica Sinica*, 32 (11), 82-86.

Feng, X., M.N. Tsimplis, G.D. Quartly, and M.J. Yelland. 2014a. Wave height analysis from 10 years of observations in the Norwegian Sea, *Continental Shelf Research*, 72(1), 47-56.

Feng, X., M.N. Tsimplis, M.J. Yelland, and G.D. Quartly. 2014b. Changes in significant and maximum wave heights in the Norwegian Sea, *Global and Planetary Change*, 113, 68-76.

Ganguly, D., M.K. Mishra, and P. Chauhan. 2015. Deriving sea state parameters using RISAT-1 SAR data, *Advances in Space Research*, 55 (1), 83-89.

Kudryavtseva, N.A., and T. Soomere. 2016. Validation of the multi-mission altimeter wave height data for the Baltic Sea region. *Estonian Journal of Earth Sciences*, 65 (3), pp. 161-175.

Liang, B., X. Liu, H. Li, Y. Wu, and D. Lee. 2016. Wave Climate Hindcasts for the Bohai Sea, Yellow Sea, and East China Sea, *Journal of Coastal Research*, 32(1), 172-180.

Mentaschi, L., G. Besio, F. Cassola, and A. Mazzino. 2013. Developing and validating a forecast/hindcast system for the Mediterranean Sea, *Proceedings of 12th International Coastal Symposium (ICS) 2013. Journal of Coastal Research*, Special Issue 65, 1551-1556.

Passaro, M., L. Fenoglio-Marc, and P. Cipollini. 2015. Validation of Significant Wave Height From Improved Satellite Altimetry in the German Bight, *IEEE Transactions on Geoscience and Remote Sensing*, 53 (4), 2146-2156.

Reale, F., F. Dentale, E. Pugliese Carratelli, and L. Torrisi. 2014. Remote Sensing of Small-Scale Storm Variations in Coastal Seas, *Journal of Coastal Research*, 30(1), 130-141.

Sartini, L., F. Cassola, and G. Besio. 2015a. Extreme waves seasonality analysis: An application in the Mediterranean Sea, *Journal of Geophysical Research: Oceans*, 120(9), 6266-6288.

Sartini, L., L. Mentaschi, and G. Besio. 2015b. Comparing different extreme wave analysis models for wave climate assessment along the Italian coast, *Coastal Engineering*, 100, 37-47.

Xu, Z., W. Zhou, Z. Sun, Y. Yang, J. Lin, G. Wang, W. Cao, and Q. Yang. 2015. Estimating the Augmented Reflectance Ratio of the Ocean Surface When Whitecaps Appear. *Remote Sensing*, 7(10), 13606-13625.