

## Measure-Correlate-Predict techniques to improve long term numerical datasets: application to wave potential estimate

Youen KERVELLA<sup>1</sup>, Meidhy DEMAGNY<sup>2</sup>,  
François BATIFOULIER<sup>1</sup>, Raphaël DUBOIS<sup>2</sup>

1. OPEN OCEAN, 15 rue Kepler, 29200 Brest, France.
2. OPEN OCEAN, 97 rue Oberkampf, 75011 Paris, France.  
[youen.kervella@openocean.fr](mailto:youen.kervella@openocean.fr)

### Abstract:

Many long-term numerical databases (LTND) are distributed by research centers from various countries (like Previmer, NCEP, ..). These LTND usually provide integrated wave parameters which enable the computation of operational and also extreme statistics thanks to their duration (generally between 10 and 30 years).

Although these LTND are validated in several places, it is rarely the case at your location of interest and therefore data extracted from a LTND doesn't necessarily offer the best results at this point. Nevertheless, it is possible to improve the results of these LTND in a "post-processing" step, ie without re-running the numerical simulations with data assimilation.

This paper will present different techniques of Measure-Correlate-Predict (MCP) to improve the results of these LTND using in-situ data. Some of these techniques are then implemented to modify a long-term wave database in northern Bay of Biscay in France. The effect of this change is evaluated in terms of significant wave heights and wave peak period, and also in term of theoretical wave energy potential calculated using the deep waters and JONSWAP approximations, by comparing the obtained results with in-situ measurements.

A discussion is then opened on the way of propagating these local corrections and on the way of making it on larger spatial extents and on different physical types of LTND (wave, current, wind), thanks to the Metocean Analytics software (<http://www.openocean.fr/en/metocean-analytics/>).

**Keywords:** Measure-Correlate-Predict techniques, Wave datasets, In situ measurements, Wave potential.

### 1. Introduction

The cost of the electrical energy production of a marine renewable energy farm, composed of tidal turbines, offshore wind turbines or wave energy converters, depends on its annual power output as well as on the investment, financing and operating and maintenance costs (MANWELL *et al.*, 2009; NELSON, 2009; SPERA, 1994; SHERMAN *et al.*, 1983; HAU, 2005). The annual power output of a device has a strong

influence on this specific cost. In order to get a reliable prediction of this annual power output, knowledge of the inter-annual variability of the resource characteristics is needed, and so on long-term dataset are essential, but often unavailable.

The use of statistical methods, commonly known as Measure-Correlate-Predict (MCP) techniques, is used as a way of getting round this lack of long-term data. MCP algorithms are used to assess the long-term resource at target sites using short-term (one- or two-year) on-site data and concurrent data at nearby reference locations (which also have long-term data). The accuracy of long-term predictions obtained using MCP methods is subject to the availability of a nearby in-situ measurement, the used correction methodology, and the dependence of this correction on physical features such as the bathymetry, the topography, the distance between reference stations, and the type of local climate regime.

As mentioned in Zhang (ZHANG *et al.*, 2014), a wide variety of MCP techniques have been reported in the literature, mainly to estimate wind resource, such as: linear regression (VELASQUEZ *et al.*, 2011; PEREA *et al.*, 2011), variance ratio (PEREA *et al.*, 2011; CARTA & VELASQUEZ, 2011), Weibull scale (CARTA & VELASQUEZ, 2011), artificial neural networks (ANNs) (HAU, 2005; VELASQUEZ *et al.*, 2011; MOHANDÉS *et al.*, 1998) or support vector regression (SVR) (MOHANDÉS *et al.*, 2004; ZHAO *et al.*, 2010).

This paper focuses on simple MCP techniques (linear, exponential and polynomial regressions) as a first step applied to wave datasets: significant wave heights ( $H_s$ ), wave peak periods ( $T_p$ ) and wave potential using a deep water and a JONSWAP approximation (WPJA).

Zhang (ZHANG *et al.*, 2014) develops an hybrid MCP technique for wind resource assessment at target farm site, by using data from multiple reference stations. The weight of each reference station was determined based on the distance and on the elevation differences between the target farm site and each reference station. In our paper, the elevation of the input buoys (Candhis Belle-Ile and Ile d'Yeu) and of the target site (Candhis Four) is the same (sea surface), thus, the weight is determined based on the distance between the target site and each buoy. However, wave systems (defined by a range in direction and frequency) can significantly differ between two relatively close locations, due to modifications of wave propagation driven by local bathymetry changes or the presence of the islands. Moreover, wave systems may not propagate in the same way, resulting in different relationships between wave parameters at the reference location and at the target location. That is why, in a second step, wave parameters ( $H_s$  and  $T_p$ ) are broken down into peak direction classes in order to establish MCP relationships specific to wave directional systems.

## 2. Datasets and methodology

### 2.1 Long Term Numerical Dataset (LTND)

The wave long-term numerical dataset used in the present paper comes from a numerical simulation performed by Ifremer. This database has been performed using the numerical wave model WaveWatchIII® (WW3) version 4.09 (BOUDIERE *et al.*, 2013). WW3 is a third generation spectral wave model based on the conservation equation for the density of wave action. The propagation scheme used in this configuration is an explicit propagation for unstructured grid (ROLAND, 2008; ROLAND *et al.*, 2009). The use of unstructured meshes permits to adapt the grid resolution at different scales in the same computational domain, from the coastal zone (refined mesh of ~ 200m) to offshore (mesh ~ 10km). The time step is about 1-hour. It extends from 43.29°N to 52.90°N and from 8.54°W to 4.72°E and covers a period from 1994 to 2012. The setup used in this configuration for the generation and dissipation of waves (ARDHUIN *et al.*, 2009; ARDHUIN *et al.*, 2010) is the one that was developed during the research project IOWAGA (Integrated Ocean Waves for Geophysical Applications and —other) and tested in preparation mode in the context of operational demonstrator Previmer (LECORNU *et al.*, 2008). The evolution and nonlinear wave interactions are modeled by the DIA method (Discrete Interaction Approximation) (HASSELMANN *et al.*, 1985). The simulated sea-state conditions were performed on a high-resolution bathymetry stretching from southern North Sea to the northern coast of Spain, covering the entire continental shelf of the Bay of Biscay. The bathymetry was obtained using data from the SHOM (Service Hydrographique et Océanographique de la Marine) for the coastline and measurement campaigns conducted by IFREMER and SHOM for the entire field: 100m and 500m DTM (LOUBRIEU *et al.*, 2008).

The wind fields used to force the model are from CFSR reanalysis (Climate Forecast System Reanalysis, (SAHA *et al.*, 2010) conducted in 2010 by the NCEP (National Centers for Environmental Prediction). These wind fields were re-analyzed over the period 1979-2009. Their spatial resolution varies from 0.25° at the equator up to 0.5° at higher latitudes. Currents, water levels and storm surges were calculated using the hydrodynamic code MARS2D (Model for Applications at Regional Scale). MARS2D is a model developed by IFREMER (LAZURE & DUMAS, 2008) and based on shallow water equations. It consists of seven nested models whose resolution differs according to rank (ranks 0, 1 and 2). Data from Météo-France were used as meteorological forcing for the model MARS2D. Ranks 0 and 1 are forced using data from meteorological model ARPEGE 0.5° (COURTIER *et al.*, 1994; COURTIER *et al.*, 1991) with a 6-hours time step. Models of rank 2, with higher resolution, are forced with data from the meteorological model AROME 0.025° (SEITY *et al.*, 2011) with a 1-hour time step. Input current and water level data was derived from an atlas of harmonic components. A replay of tide data was performed over one year (2008) and an analysis of harmonic

components of the tide for each of the seven models too. Tidal currents and levels can thus be evaluated for each year over the entire field. They are updated every 30 minutes and are interpolated onto the wave model mesh. A large set of in situ data from various sources, including ocean surveys like data from the Cetmef Candhis buoys network and Météo-France buoys along the French coast, was used and statistics show a good agreement overall with the observations (BOUDIERE *et al.*, 2013).

## 2.2 Measurements

In-situ measurement datasets used in the present paper come from the Cetmef Candhis buoys network, distributed through the Copernicus (European Earth Observation Program) program. Belle-Ile and Ile d'Yeu buoys are used in this work as input parameters of the different MCP techniques (see section methodology for more details), and Four buoy is used in order to validate the results as it is located between the two other previous buoys (figure 1). The main characteristics of these in-situ datasets are summed up in the table 1.

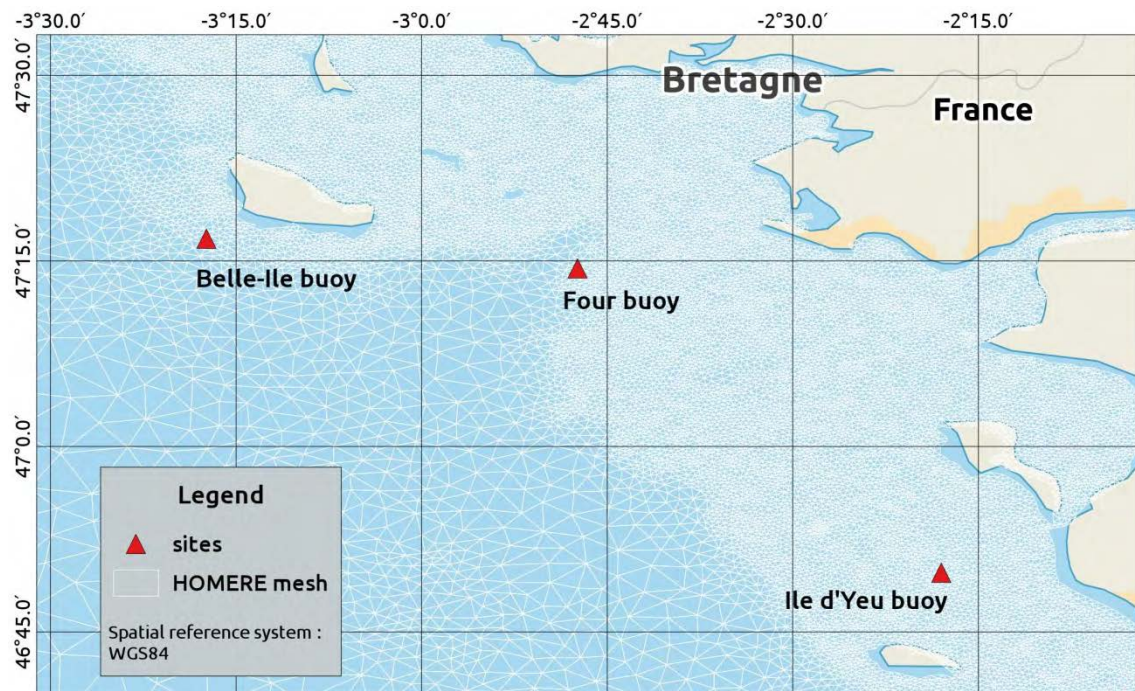


Figure 1. Reference and target sites in the North of Bay of Biscay (France).

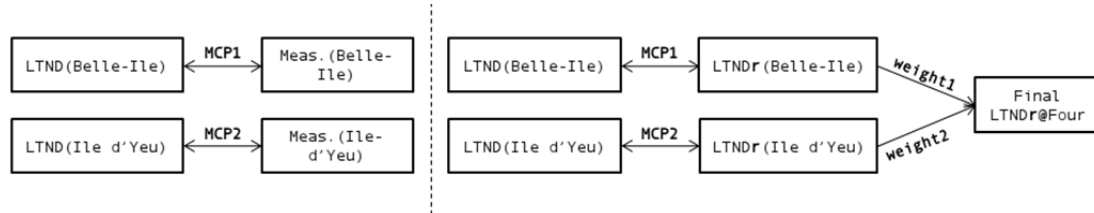
*Table 1. location covered period, available variables and distance to HOMERE grid point for each measurement buoy.*

<i>Buoy</i>	<i>location</i>	<i>period</i>	<i>variables</i>	<i>Distance to grid point</i>
<i>Belle-Ile</i>	<i>3.285°W</i>	<i>01/01/2012</i>	<i>Hs, Tp, Dp</i>	<i>250 m</i>
	<i>47.285°N</i>	<i>21/05/2014</i>		
<i>Ile d'Yeu</i>	<i>2.288°W</i>	<i>04/07/2005</i>	<i>Hs</i>	<i>350 m</i>
	<i>46.833°N</i>	<i>30/11/2015</i>		
<i>Four</i>	<i>2.787°W</i>	<i>18/01/2010</i>	<i>Hs, Tp, Dp</i>	<i>270 m</i>
	<i>47.239°N</i>	<i>14/12/2014</i>		

### 2.3 Methodology

Figure 2 illustrates the overall structure of the proposed methodology, which includes:

- MCP (MCP regression are listed below) between in-situ measurements (Belle-Ile or Belle-Ile and Ile d'Yeu) and LTND at same locations.
- Weights determination based on the distance between the target site (Four) and each reference site (Belle-Ile and Ile d'Yeu).
- Computation of modified wave parameters time series and wave potential time series, and comparison with in-situ measurements at target site (Four).



*Figure 2. MCP methods between measurements and LTND at reference sites (left) and computation of the final LTND at target site (right).*

Several MCP techniques are proposed, classified by step: STEP 1 corresponds to a linear regression between in-situ measurements and LTND at one reference site (Belle-Ile), STEP 2 corresponds to a linear regression between in-situ measurements and LTND at two reference sites (Belle-Ile and Ile-D'Yeu, same technique than STEP 1), STEP 3 corresponds to an exponential regression between in-situ measurements and LTND at these two reference sites and STEP 4 corresponds to a polynomial regression (order of 3).

For example, figure 3 presents the STEP 1 regression for Hs (left) and STEP 4 regression for Tp (right) at Belle-Ile location.

Using the modified Hs and Tp time series, the wave potential is then estimated at the target site. The wave potential can be approximated, in deep waters, by the relation (CORNETT, 2008):

$$P_0 = (\rho g^2 / 64\pi) * H_s^2 * T_e$$

where  $T_e$  is the energy wave period.

Since  $T_e$  is unavailable in a large number of public wave databases or in-situ measurement databases, it must be estimated using other available variables. Thus,  $T_e$  is generally approximated as:  $T_e = \alpha T_p$  with the coefficient  $\alpha$  depending on the shape of the wave spectrum; a conservative value of 0.9 is commonly used and corresponds to a standard JONSWAP spectrum with a peak factor  $\gamma = 3.3$ . In the following, this value is used to compute wave potential (named WPJA: Wave Potential with JONSWAP Approximation) at the target site (Four buoy). The previous assumptions have no impact on the results since the article methodology only deals with comparisons.

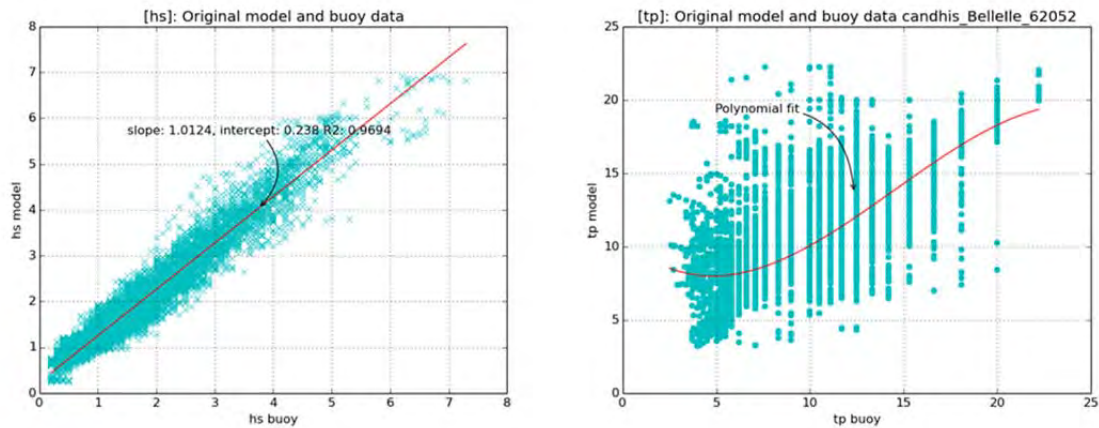


Figure 3. scatter plots of  $H_s$  (left, STEP 1) and  $T_p$  (right, STEP 4) at Belle-Ile location.

## 4. Results

### 4.1 Modified time series

The following table shows the scores (bias and normalized root mean square error) of the original LTND and the 4 MCP-modified LTND at the target site (Four buoy), in terms of wave parameters.

Table 1.  $H_s$  and  $T_p$  scores of original and modified LTNDs at the target site.

Steps	$H_s$		$T_p$	
	BIAS [m]	NRMSE [%]	BIAS [s]	NRMSE [%]
Original	0.21	3.82	1.03	7.86
STEP1	-0.04	2.75	0.75	9.45
STEP2	-0.02	2.74	0.75	9.45
STEP3	0.44	6.21	1.32	7.76
STEP4	0.44	6.16	1.33	7.78

In terms of significant wave height, the best score is obtained with the STEP 1 and STEP 2 MCPs (linear regressions): 2 cm of bias and less than 3% of NRMSE. In fact, the relation between measured and simulated  $H_s$  is often very linear, as illustrated in the figure 3.

Regarding the wave peak period, the results are less biased (0.75 second) using the linear regressions but the NRMS errors are lower using the exponential and polynomial regressions (7.8 % versus 9.4 %) since the relations between measured and simulated wave periods are more complex (figure 3).

#### 4.2 Wave potential estimate

The following table shows the scores (bias and normalized root mean square error) of the original LTND and the 4 MCP-modified LTND at the target site (Four buoy), in terms of wave potential (using deep waters and JONSWAP approximations).

The best score is obtained using the linear regression (MCP 1). In fact, the wave potential highly depends on the value of the significant wave height (square relationship) and as discussed above, the relation between measured and simulated  $H_s$  is very linear. The overall wave potential bias goes from 3661 kW/m (original LTND) to 880 kW/m (STEP 1 MCP), that is to say a bias divided by 4.

<i>Steps</i>	<b>WPJA</b>
	<b><i>BIAS [kW/m]   NRMSE [%]</i></b>
<i>Original</i>	3661   2.45
<i>STEP1</i>	880   2.12
<i>STEP2</i>	1086   2.18
<i>STEP3</i>	6890   3.19
<i>STEP4</i>	6560   3.05

*Table 3: WPJA scores of original and modified LTNDs at the target site.*

#### 4.3 Wave peak direction classes

The wave peak period time series at point Belle-Ile is finally divided in wave peak direction classes, in order to simplify the relationships between measured and simulated peak periods at this point. Then a STEP 1 MCP (linear regression) is applied to each directional class and the final modified LTND at the target site (Four buoy) is reconstituted.

The figure 4 shows the results in terms of peak periods obtained at the target site versus the number of directional classes at point Belle-Ile (on the left) and the results in terms of wave potential obtained at the target site versus the number of directional classes at point Belle-Ile (on the right).

It highlights the fact that the bias on wave peak period can be reduced by this way and both the bias (reduced through 20%) and NRMSE (reduced through 8%) on wave potential too.

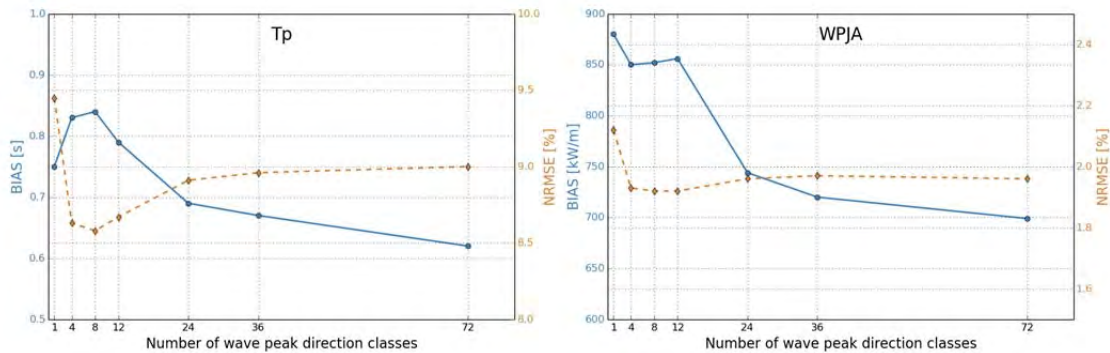


Figure 1: scores of modified LTND at the target site versus number of directional classes,  $T_p$  (left) and WPJA (right).

## 5. Discussion

Several MCP techniques were implemented on long-term wave dataset in reference site and performances were compared at a target site. The results of these modifications are shown in terms of wave parameters (table 2) and in terms of wave potential according to deep-water and JONSWAP approximations (table 3). The gain provided by these techniques is significant with a wave potential bias divided by 4. In a second step, the analysis was broken down into wave peak direction classes, resulting in reduced bias for a higher number of classes. The identification of wave systems in terms of wave peak direction and wave peak period classes could be improved. A limitation of this work is the lack of in-situ data in each reference site.

In future studies, the question of the impact of the number of reference sites and the length of the period covered by measurements should be addressed. The artificial neural network technique will also be tested in an automated way in several areas around the world thanks to the Metocean Analytics software. We will also take an interest in how to implement these techniques on ocean current, by applying filters on the current components.

## 6. References

- MANWELL J.F., MCGOWAN J.G., ROGERS, A.L. (2009). *Wind energy explained*. (2nd ed.) John Wiley & Sons, Great Britain. <http://dx.doi.org/10.1002/9781119994367>
- NELSON V. (2009). *Wind energy*. (1st ed.) CRC Press, FL. <http://dx.doi.org/10.1201/9781420075694>
- SPERA D.A. (1994). *Wind turbine technology*. ASME Press, New York.



- SHERMAN J.M., GRESHAM M.S., FERGASON D.L. (1983). *Wind systems life cycle cost analysis*. (1st ed.) WindBooks, New York.
- HAU E. (2005). *Wind turbines*. (2nd ed.) Springer, New York.
- VELASQUEZ S., CARTA J.A., MATIAS J.M. (2011). *Comparison between anns and linear mcp algorithms in the long-term estimation of the cost per kwh produced by a wind turbine at a candidate site: a case study in the canary islands*. Appl Energy, Vol. 88(11), pp 3689–881. <http://dx.doi.org/10.1016/j.apenergy.2011.05.007>
- PEREA A.R., AMEZCUA J, PROBST O. (2011). *Validation of three new measure-correlate-predict models for the long-term prospection of the wind resource*. J Renew Sustain Energy, Vol. 3(2), 023105. <http://dx.doi.org/10.1063/1.3574447>
- CARTA J.A., VELASQUEZ S. (2011). *A new probabilistic method to estimate the long-term wind speed characteristics at a potential wind energy conversion site*. Energy, Vol. 36(5), pp 2671–85. <http://dx.doi.org/10.1016/j.energy.2011.02.008>
- MOHANDES M.A., REHMAN S., HALAWANI T.O. (1998). *A neural networks approach for wind speed prediction*. Renew Energy, Vol. 13(3), pp 345–54. [http://dx.doi.org/10.1016/S0960-1481\(98\)00001-9](http://dx.doi.org/10.1016/S0960-1481(98)00001-9)
- MOHANDES M.A., HALAWANI T.O., REHMAN S., HUSSAIN A.A. (2004). *Support vector machines for wind speed prediction*. Renew Energy, Vol. 29(6), pp 939-947. <http://dx.doi.org/10.1016/j.renene.2003.11.009>
- ZHAO P., XIA J., DAI Y., HE J. (2010). *Wind speed prediction using support vector regression*. In: The 5th IEEE conference on industrial electronics and applications (ICIEA), IEEE, Taiwan; 2010. <http://dx.doi.org/10.1109/iciea.2010.5515626>
- ZHANG J., CHOWDHURY S., MESSAC A., HODGE B.-M. (2014). *A hybrid measure-correlate-predict method for long-term wind condition assessment*. Energy Conversion and Management, Vol. 87, pp. 697-710. <http://dx.doi.org/10.1016/j.enconman.2014.07.057>
- BOUDIERE E., MAISONDIEU C., ARDHUIN F., ACCENSI M., PINEAU-GUILLOU L. and LEPESQUEUR J. (2013). *A suitable metocean hindcast database for the design of Marine energy converters*. International Journal of Marine Energy, Vol. 3-4, pp e40-e52. <http://dx.doi.org/10.1016/j.ijome.2013.11.010>
- ROLAND A. (2008). *Development of WWM II: Spectral wave modelling on unstructured meshes*. PhD thesis, Technische Universität Darmstadt, Institute of Hydraulic and Water Resources Engineering.
- ROLAND A., CUCCO A., FERRARIN C., HSU T.-W., LIAU J.-M., OU S.-H., UMGIESSER G., ZANKE U. (2009). *On the development and verification of 2-D coupled wave-current model on unstructured meshes*, J. Mar. Syst., Vol. 78, pp S244-S254. <http://dx.doi.org/10.1016/j.jmarsys.2009.01.026>
- ARDHUIN F., CHAPRON B., COLLARD F. (2009). *Observation of swell dissipation across oceans*, Geophys. Res. Lett. Vol. 36, L06607. <http://dx.doi.org/10.1029/2008GL037030>

- ARDHUIN F., ROGERS E., BABANIN A., FILIPOT J.-F., MAGNE R., ROLAND A., VAN DER WESTHUYSEN A., QUEFFEULOU P., LEFEVRE J.-M., AOUF L., COLLARD F. (2010). *Semi-empirical dissipation source function for wind-wave models: part 1, definition, calibration and validation at global scales*, J. Phy. Oceanogr., Vol. 40, pp 1917–1941. <http://dx.doi.org/10.1175/2010JPO4324.1>
- LECORNU F., PAILLET J., RAVANEL H. (2008). *PREVIMER-Coastal observations and forecast – summary of 2 years demonstration and future perspectives*. Proc, Operational Coastal Oceanography conf., Brest, 2008.
- HASSELMANN S., HASSELMANN K., ALLENDER J., BARNETT T. (1985). *Computation and parameterizations of the nonlinear energy transfer in a gravity-wave spectrum. part 2: Parameterizations of the nonlinear energy transfer for application in wave model*”, J. Phys. Oceanogr., Vol. 15 , pp 1378–1391. [http://dx.doi.org/10.1175/1520-0485\(1985\)015<1378:CAPOTN>2.0.CO;2](http://dx.doi.org/10.1175/1520-0485(1985)015<1378:CAPOTN>2.0.CO;2)
- LOUBRIEU B., BOURILLET J.F., MOUSSAT E. (2008). *Bathy-morphologie régional du Golfe de Gascogne et de la Manche*, modèle numérique 2008, Ifremer, Brest, Tech. Rep., DCD/GM/CTDI/08-01, 2008.
- SAHA S. et al (2010). *The NCEP Climate Forecast System Reanalysis*, Bulletin Am. Meteor. Soc., Vol. 91, pp 1015–1057. <http://dx.doi.org/10.1175/2010BAMS3001.1>
- LAZURE P., DUMAS F. (2008). *An external-internal mode coupling for a 3D hydrodynamical model for applications at regional scale (MARS)*, Adv. Water Res., Vol. 31(2), pp 233–250. <http://dx.doi.org/10.1016/j.advwatres.2007.06.010>
- COURTIER P., THEPAUT J., HOLLINGSWORTH A. (1994). *A strategy for operational implementation of 4D-VAR, using an incremental approach*, Q. J. R. MeteorolSoc., Vol. 120, pp 1367–1387. <http://dx.doi.org/10.1002/qj.49712051912>
- COURTIER P., FREYDIER C., GELEYN J.-F., RABIER F., ROCHAS M. (1991). *The ARPEGE project at Météo-France*, ECMWF Seminar Proc.: Numerical Meth. Atmosph. Models II, pp 193–231.
- SEITY Y., BROUSSEAU P., MALARDEL S., HELLO G., BENARD P., BOUTTIER F., LAC C., MASSON V. (2011). *The AROME-France convective scale operational model*, Mon. Wea. Rev., Vol. 139, pp 976–991. <http://dx.doi.org/10.1175/2010MWR3425.1>
- CORNETT A. M. (2008). *A global wave energy resource assessment*. ISOPE 2008.