Ocean Sci. Discuss., 8, 761–800, 2011
www.ocean-sci-discuss.net/8/761/2011/
doi:10.5194/osd-8-761-2011

Ocean Science
Discussions

ENSURF: multi-model sea level forecast – implementation and validation results for the IBIROOS and Western Mediterranean regions

B. Pérez[1], **R. Brower**[2], **J. Beckers**[2], **D. Paradis**[3], **C. Balseiro**[4], **K. Lyons**[5], **M. Cure**[5], **M. G. Sotillo**[1], **B. Hacket**[6], **M. Verlaan**[2], and **E. Alvarez Fanjul**[1]

[1]Organismo Público Puertos del Estado (OPPE), Madrid, Spain
[2]Deltares, Delft, The Netherlands
[3]Meteo-France, Toulouse Cedex, France
[4]MeteoGalicia, Santiago de Compostela, Spain
[5]Irish Marine Institute, Co. Galway, Ireland
[6]Norwegian Meteorological Institute, MET-NO, Norway

Received: 1 March 2011 – Accepted: 15 March 2011 – Published: 8 April 2011

Correspondence to: B. Pérez (bego@puertos.es)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

ENSURF (Ensemble SURge Forecast) is a multi-model application for sea level forecast that makes use of existing storm surge or circulation models today operational in Europe, as well as near-real time tide gauge data in the region, with the following main goals:

- providing an easy access to existing forecasts, as well as to its performance and model validation, by means of an adequate visualization tool
- generation of better forecasts of sea level, including confidence intervals, by means of the Bayesian Model Average Technique (BMA)

The system was developed and implemented within ECOOP (C.No. 036355) European Project for the NOOS and the IBIROOS regions, based on MATROOS visualization tool developed by Deltares. Both systems are today operational at Deltares and Puertos del Estado respectively. The Bayesian Modelling Average technique generates an overall forecast probability density function (PDF) by making a weighted average of the individual forecasts PDF's; the weights represent the probability that a model will give the correct forecast PDF and are determined and updated operationally based on the performance of the models during a recent training period. This implies the technique needs the availability of sea level data from tide gauges in near-real time. Results of validation of the different models and BMA implementation for the main harbours will be presented for the IBIROOS and Western Mediterranean regions, where this kind of activity is performed for the first time. The work has proved to be useful to detect problems in some of the circulation models not previously well calibrated with sea level data, to identify the differences on baroclinic and barotropic models for sea level applications and to confirm the general improvement of the BMA forecasts.

1 Introduction

During last decades the availability of sea level forecasts, including not just the astronomical tide but also the meteorological component, has become critical for some countries where the magnitude of storm surges can reach several meters and cause considerable damage and inundation along the coast. This is the case of North Sea surrounding states, for example, where the land is both low-lying and densely populated, and historic extreme storm surges have caused even thousands of dead. More recently, even regions less prompt to these dramatic events start to make use of these forecasts, such as the Mediterranean coast, where the meteorological component is of the same order of magnitude than the tide, and its knowledge becomes useful for large vessel manoeuvering inside the harbours, or for dredging operations.

Numerical modelling of surges is able to provide a good description of how a real sea responds to a real imposed weather field, allowing taking into consideration irregular boundaries and depths of the oceans, that affect surge propagation and magnitude, and including non-linear surface and bottom stresses. Barotropic 2-D models have proved to be adequate for this application during the last 30 years (Flather, 1981, 1987; Alvarez-Fanjul et al., 2001) and have been the basis of the existing operational sea level forecasts up to now. Nevertheless, the recent improvement on computer skills and performance of 3-D baroclinic models for current forecasts, and their operational implementation, has lead to the availability of sea level forecasts coming also from these more sophisticated and general models in some regions; although not always considered with the importance it has, validation of sea level output of general circulation models is critical for a correct characterization of the sea surface elevation and consequently for an adequate description of the circulation patterns.

It is important to stress, however, that numerical models are not enough by themselves to provide a precise description of real total sea level, as they always present a bias with respect to observations that needs to be corrected by making use of real time tide gauge data. So a complete operational sea level forecast system should be always based on the use of numerical models and tide gauge data.

So storm surge forecasting remains one of the central tasks of operational forecasting centres throughout Europe, at the same time that the new more general ocean circulation systems. All these systems provide deterministic and independent forecasts of sea level for their specific regions, sometimes geographically coincident in part, and had not been compared and even less combined in order to improve their skills at the common domains or points. Recent studies have demonstrated the advantages of the multi-model and the ensemble approach for validation and improvement of predictive capabilities. Both facts generated the idea of creating the ENSURF system (Ensembles SUrge Forecast), within ECOOP European project (European COastal shelf sea OPerational observing and forecasting system), Contract No. 036355, whose overall goal is to consolidate, integrate and further develop existing European coastal and regional seas operational systems. ENSURF constitutes one of the main products of this project (http://www.ecoop.eu/summary.php), as it represents a perfect example of this integration, not only because it involves different forecasting systems but also because it makes use of observations to improve the final forecasts.

2 ENSURF system: objectives and general description

ENSURF is a multi-model application for sea level forecast that makes use of existing storm surge/circulation models today operational in Europe, as well as near-real time tide gauge data in the region. So it is an integration of existing operational sea level forecasts, with potential for relocation in new coastal areas and the following main objectives:

1. providing an easy access to existing forecasts as well as to the performance and validation of the different models, through an adequate visualization tool
2. generation of better forecasts of sea level, including confidence intervals, by means of statistical techniques such as the Bayesian Model Average (BMA)
3. becoming a joint European service in the framework of ECOOP project

The system was implemented for the NOOS and IBIROOS regions (Fig. 1), and it is running operationally at Deltares (http://noos.deltares.nl) and Puertos del Estado (http://ensurfibi.puertos.es), respectively, based on Matroos visualization tool, developed by Deltares. The reason for the implementation of these two components was the different status and experience on sea level data exchange policy, both from models and observations, at the two regions. Initially it was not possible to develop a component for the MOON region, due to the lack of enough operational models with output of sea level in the Mediterranean Sea. Nevertheless, in this work we have included the Western Mediterranean, where Spanish and French forecasts and data were available. The system has shown its usefulness as a user-friendly operational validation tool, and its ability to improve existing forecasts by means of the Bayesian Model Average Technique. It is the first time such a kind of tool is implemented for the South of Europe and, although only implemented for sea level applications within this project, it would be immediate and easy to do it for any other parameter such as waves or currents.

We present in this paper the implementation performed at Puertos del Estado for the IBIROOS region, for which it was necessary to select the locations of the storm surge forecasts within available tide gauges and to establish the data fluxes (real time measurements and forecasts) between partners. The system is nowadays operational and ready to incorporate more stations and sources in the future (Fig. 2). We will also show at the end the first validation results of the different models and the performance of the Bayesian Model Average Technique, for a specific period.

ENSURF is based on MATROOS (Multifunctional Access Tool for Operational Ocean Data Services) visualization tool developed by Deltares. It is installed in a computer where automatic scripts are in charge of the acquisition of data from both models and tide gauges through the Internet, by means of ftp sites. So the first step is establishing the adequate data flux and formats for an operational integration in our system. Through this scheme, both time series of data (forecasts and observations) and forecasted fields can be included in an internal Data Base, allowing an easy access and visualization by remote users (Fig. 3).

The models output can be simply the surge component (when they are forced just with meteorological forecasts) or the total sea level (including the tide). In the first case the tide needs to be added later in order to provide a total sea level forecast. On the other hand, sometimes the models are run in a barotropic mode, as it has been said normally sufficient for storm surge applications, and others the sources of the forecasts are general circulation or baroclinic models, which in principle include all the different sea level signals (e.g. density changes). For the first time all these different applications can be validated in near-real time thanks to ENSURF system.

2.1 ENSURF-IBIROOS sources and data

The sources currently contributing with operational sea level forecasts to the IBIROOS component of ENSURF are shown in Table 1. As already mentioned, the characteristics of the models differ, some being barotropic and others baroclinic, and also accounting for different kinds of forcings, leading to different outputs of sea level (depending on just having meteorological forcing or including also the tide).

Of course, implemented by different institutions in different countries, the domains of the models are also diverse (Fig. 4), although sharing part of the coastline in some cases; these will be the coastlines and harbours where the advantage of multi-model approach to improve the forecasts will be explored. A brief description of each source, without entering into too many details, is given below.

2.1.1 Nivmar system

In operation since 1998 at Puertos del Estado, it is based on the Hamsom circulation model and the use of near-real time tide gauge data from the REDMAR network (Alvarez Fanjul et al., 2001). The model is run vertically integrated in barotropic mode, forced only with meteorological data (atmospheric pressure and wind) from the HIRLAM meteorological model (Unden et al., 2002; HIRLAM project: http://hirlam.org). The forecast is run twice a day (00:00 h and 12:00 h UTC cycles), with a 72 h forecast horizon, and the domain covers the Spanish Atlantic coast and Canary Islands as well as the whole Mediterranean Sea.

Hamsom (Backhaus, 1983; Rodríguez et al., 1991; Alvarez et al., 1997) uses a finite difference semi-implicit scheme on a variable size grid, being the resolution of the central area of the domain $10' \times 15'$ for latitude and longitude, respectively. For the bottom friction it makes use of a quadratic function in terms of the current velocity, and for the wind stress it uses the Charnock parameterization (Charnock, 1955), which consists on the use of a constant non-dimensional surface roughness or Charnock coefficient ($\alpha = z_0 g/W^2$, where z_0 is the rougness length, W the friction velocity and g the gravitational acceleration). The open boundary conditions consist of the inverted barometer effect (Atlantic). The HIRLAM meteorological model is a limited area model with 0.16° and 6 h spatial and temporal resolution, being run twice daily by the AEMET (Spanish Meteorological Agency).

The bathymetry employed is the DTM5 data set (GETECH, 1995). Output data are hourly values of meteorological residual at all the points of the domain (no tide) and total seal level at special points (harbours) where a tide gauge is available, what allows to add the tidal component derived from observations to the model result.

Nivmar includes a simple data assimilation scheme for the forecast at the harbours, improving the results of the predictions by correcting the mean value of the simulated residuals. The correction is done by adding a constant value which is the difference of the means of the predicted and the observed time series during a recent time window. This is in fact the same technique used by the BMA to deal with the bias problem that will be explained later.

2.1.2 ESEOAT system

ESEOAT is an ocean forecasting system operational at Puertos del Estado since 2006 (Sotillo et al., 2007, 2008). It is based on the 3-D baroclinic model POLCOMS (Proudman Oceanographic Laboratory Coastal Ocean Modelling System), which uses a finite differences scheme, and covers the Iberian Atlantic waters with an 1/20° horizontal resolution and 34 vertical S-levels. A flux/radiation boundary condition scheme is used for elevations and water column mean velocities; relaxation of temperature and salinity and inverse barometer conditions are also included at the open boundaries. Near bed velocities are computed by means of friction coefficients. The system is forced with the same meteorological fields than Nivmar (HIRLAM system). The wind stress makes use of the Charnok parameterization. Bathymetry used is derived from GTOPO30 data base and a tidal forcing, based on 15 harmonic constituents imposed at the open boundaries, is also considered. Hourly outputs of total sea level (including tides) and surface fields are provided, as well as daily averaged 3-D fields (temperature, salinity and currents). ESEOAT does not include tide gauge data assimilation.

2.1.3 Meteo-France sources

Meteo-France provides three different forecasts to the system, which make use of the same circulation model (the Météo-France Storm Surge model), with different meteorological forcing. The storm surge model is a 2-D barotropic model which uses finite differences on an uniform grid, with a resolution of $5'$ (around 9 km), the Chezy bottom roughness condition (Chézy, 1776), which implies the dependence of the bottom friction coefficient on a constant Chezy value for the whole domain, i.e. no depth dependency, and the Wu formulation for the wind stress (Wu, 1982), which considers it varies linearly with U_{10}, wind velocity measured at 10 m above the mean sea surface. As boundary condition it uses the output of the global ocean storm surge circulation model, and as bathymetry the GEBCO$1' \times 1'$ plus local and regional fixes. There is no tide forcing. The three forecasts correspond to the following meteorological forcings:

- Metfr_ecmwf: IFS: ECMWF global model with 4DVar, 25 km, 0.5° every 6 h.
- Metfr_arpege: Arpege: Météo-France global model with 4DVar, 23 km, 0.25°, every 3 h.
- Metfr_aladin: Aladin: Météo-France, LAM+3DVar coupled by Arpege, 9 km, 0.1°, every 3 h.

The output consists of 10 min surge or meteorological sea levels at all the grid points. No data assimilation from tide gauge data.

2.1.4 IMI system

Operational circulation forecasting system developed by the Marine Institute for the Northeast Atlantic, based on the 3-D baroclinic ROMS model, which uses a finite differences scheme on a c-grid, rotated by 45°. The resolution is 2.5 km. As bottom friction condition it makes use of a logarithmic B.L. with roughness coefficient. As open boundary conditions it uses MERCATOR psy2v2 with radiation, TPXO and 7 tides with Flather+Chapman. For the wind stress it uses the C. Fairall COARE method. The bathymetry is based on the Irish SWATH and the GEBCO UK Admiralty data sets. The meteorological forcing is provided by the NCEP GFS 1° 3 h fields and tide is included with 10 harmonic constituents. The output consists of 10 min total sea level at tide gauge locations and 3 h fields of salinity, temperature, currents and sea surface height. No data assimilation of tide gauges is performed.

2.1.5 Metga system

MeteoGalicia operational storm surge forecast is based on a 2-D-barotropic version of Mohid circulation model that uses a finite volumes numerical scheme and the Large and Pond, 1981, parameterization of the wind stress. Although several spatial scales have been defined with the aim of defining the storm surge processes in Galicia Coast and inside the Rias, for ENSURF we use by now just the coarse resolution grid covering the Iberian Peninsula with 0.06° resolution. The bathymetries used were obtained without any type of filtering, using as source of bathymetries GEBCO with 30 s of arc of resolution and data from local nautical charts to correct near coast zones. Meteorological forcing is provided by the local atmospheric model WRF with boundary conditions provided by the GFS global model. WRF model is running daily in 3 nested grids with 36, 12 and 4 km resolution forcing the different MOHID scales with 1 h temporal resolution. Also an inverted barometer effect is imposed at the open boundary. The system output produces daily three days forecast, with hourly values of meteorological sea level and current velocity fields, as long as surface elevation maps.

2.1.6 Tide gauge data

A common set of tide gauge stations were selected for reception of sea level data in near-real time. All the models must provide their output at all these special points included in its geographical domain; the purpose of this is not just the validation of the different models with observations at the harbours, but also the implementation of the Bayesian Model Average Technique (BMA) for statistical forecast at these specific locations, as it will be explained later. The important role of tide gauge data for improving sea level forecasts at the coast is recognized since the implementation for example of the Nivmar system (Alvarez-Fanjul et al., 2001) and it is also mentioned by Mourre et al. (2006), who found how the use of tide gauges lead to better global statistical performance of high-frequency barotropic models.

Data sampling can vary from 10 to 60 min (multiples of 10), and latency required can be of several hours. Automatic quality control of data in near-real time was implemented for this ENSURF-IBIROOS component, to avoid wrong values to enter MATROOS and affect models calibration and BMA results. Time needs to be Universal Time. As it will be explained in next section, for each individual tide gauge entering ENSURF, at least one year of data is required for previous computation of the tide. This will be needed to compute the total sea level provided by the system at a particular harbour and also for the implementation of near-real time quality control of the observations. Sea level data from tide gauges have been kindly provided by the following institutions: SHOM (France), POL (UK), SLEAC (Denmark), Marine Institute (Ireland), Geographic Institute (Portugal) and Puertos del Estado (Spain).

2.2 Tide, bias and datum correction

Several facts difficult the immediate comparison between different sea level forecasts and observations, which imply the need of making some decisions and pre-processing before sea level data enter MATROOS tool:

- Some models provide total sea level (tide + meteorological + density effects) and others just surge component (meteorological variations).
- Reference or datum of sea levels differ between models and data: models refer their output to "mean sea level", which in this case it is a spatial average that depends on the model domain and the boundary conditions. "Mean sea level" from a tide gauge station is a temporal and local (one point) average, so it depends on the period of data and the station position. To further complicate things, observations of sea level from tide gauges are normally referred to the "harbour" or "chart" datum, i.e., close to the Lowest Astronomical Tide, not to mean sea level.
- When the model includes the tide, this differs from the one obtained from observations as models use just a few set of harmonics, and not all the models use the same set. The most precise tide at a particular harbour comes from harmonic analysis of tide gauge observations.

One of the consequences of this is that all the models present significant bias with respect to sea level observations, both in surge and total sea level (Fig. 5), as well as differences in the tide and reference. In order to minimize the bias problem during the period of ECOOP project, and facilitate the visualization and comparison within MATROOS tool, this bias was computed for all the sources and stations based on two months of data previous to the Target Operational Period (TOP) of ECOOP project, which started on January 2009, and then it was applied operationally to the sources before integration into the system. This was not needed obviously for Nivmar, as the bias is already corrected in this case by the use of tide gauge data in near-real time.

Differences in the tide, on the other hand, have been solved within MATROOS in the following way:

- For models providing just total sea level: harmonic analysis is performed to one year of model output (for all the grid points (Fig. 6) and at the tide gauge points (Fig. 7)). From the obtained harmonic constants, the tide (Model Tide) can be computed and the surge component (total – tide) of the forecast extracted from:

  Forecasted surge = Forecasted Total Sea Level – Model Tide

- A harmonic analysis is also performed to one year of tide gauge observations with the same software (to avoid any differences due to the number and set of constituents used), in order to compute, in the same way:

  Observed surge = Observed Total Sea Level – Observations Tide

- Finally, the total sea level forecasted by the ENSURF system for a particular source will be the result of the Tide obtained from the observations and the Forecasted surge:

  Total Sea Level ENSURF = Forecasted Surge + Observations Tide

One of the advantages of this need of pre-computing and extracting the tide from the models that provide total sea level is that it has allowed the detection of problems in the introduction of the tide in some of the sources, which after harmonic analysis and tide extraction showed large and irregular oscillations on the residuals. On the contrary, a normal appearance of the model surge component implies the good quality of the tide of the model (Fig. 7). For this task we have used the Foreman harmonic analysis and prediction software (Foreman, 1977).

3 Bayesian Model Average (BMA) technique

One of the advantages of multi-model systems is that provide the opportunity of improving individual forecasts by means of statistical techniques, such as the Bayesian Model Average (BMA), which was for the first time employed in social and health sciences, and later applied to dynamical weather forecasting models by Raftery in 2005 (Raftery et al., 2005). In 2008 the technique was implemented for forecasting sea level at stations along the Dutch coast line, making use of six different forecasts from the NOOS region (Beckers et al., 2008).

As one of the main objectives of ENSURF, we will present later the validation results of several implementations of the BMA for the IBIROOS and Western Mediterranean regions, as compared with the validation of the existing independent sources. The BMA will also provide a statistical forecast including confidence intervals.

3.1 Description of the technique

When selecting a particular model for prediction there is always a source of uncertainty that is normally ignored and then underestimated. The Bayesian Model Average (Leamer, 1978; Kass and Raftery, 1995; Hoeting et al., 1999) solves this problem by conditioning, not on a single "best" model, but on an ensemble of models, becoming a standard method for combining predictive distributions from different sources. Our uncertainty about the best of these sources is quantified by the BMA.

It is important to stress that the dominant approach to probabilistic weather forecasting has been the use of ensembles in which a model is run several times with different initial conditions or model physics (Leith, 1974; Toth and Kalnay, 1993; Molteni et al., 1996; Hamill et al., 2000). In our case, the approach is slightly different as we make use of existing operational systems based on different models and even physics, and of course more limited in the number of members.

The basic idea is to generate an overall forecast probability density function (PDF) by means of a weighted average of PDF's centered on the individual bias-corrected forecasts; the mean of this total PDF is expected to have a smaller root mean square (RMS) error than those of the different models, i.e. there should be an improvement of the performance with respect to those of the individual forecasts (Fig. 8). The weights used on this average represent the probability that a particular model will give the correct forecast PDF, and this is determined and updated operationally based on the performance of the models during a recent training period. This implies the technique needs the availability of sea level data from tide gauges in near-real time, as it has been mentioned before. Moreover, the overall PDF, being reasonably well-calibrated, allows providing a forecast confidence interval, which is important for many practical applications. Of course, the BMA weights can also be used to assess the usefulness of ensemble members and for their selection.

The variance of the total PDF is the result of two components: the first one associated to the spread of the ensemble members, the second one to the variance of the individual model forecast PDF's. This latter component should also be determined over a training period, which can be different from the training period mentioned earlier, although in ENSURF the same training period is used to determine the BMA weight and the variance of the individual models.

The computation of the optimal BMA forecast PDF is done by means of the EM algorithm, an iterative algorithm that alternates between two steps, the E (or expectation) step and the M (or maximization) step:

1. The E-step starts from an initial guess for the weights $w(k)$ of each individual model and estimation of the matrix $z(k,s,t)$, which represents the probability that model k gives the best forecast for station s at time t:

$$z^j(k,s,t) = \frac{w(k)g(k,s,t)^{j-1}}{\sum_i w(i)g(i,s,t)} \tag{1}$$

Where j refers to the j-th iteration of the algorithm, and $g()$ represents the probability that the observed value obs*(s,t)* was predicted correctly by model k, i.e., the forecast PDF of each model which is assumed to be a normal distribution with variance $\sigma(k)$:

$$g(k,s,t)=\frac{\sigma(k)}{\sqrt{\pi}}\exp\left(\frac{(\text{obs}(s,t)-\text{forecast}(k,s,t))}{2\sigma(k)^2}\right) \tag{2}$$

2. The M-step consists then on the determination of weights $w(k)$ and variances $\sigma(k)$ of each of the models (k), based on the values of *z(k,s,t)*:

$$w^j(k)=\frac{1}{n}\sum_{s,t} z^j(k,s,t) \tag{3}$$

$$\sigma^j(k)^2=\frac{1}{n}\sum_{s,t}\sum_{k} z^j(k,s,t)(\text{obs}(s,t)-\text{forecast}(k,s,t))^2 \tag{4}$$

   Where n is the number of observations in the training period.

These two steps are repeated until convergence by using a convergence criterium or by fixing the number of iteration cycles that should guarantee convergence. Beckers found that 10 iterations are normally sufficient. In ENSURF implementation, being an operational application, weights and variances from the previous time step are used as a starting point for the new iteration. Once the convergence is reached, the overall forecast mean for each of the stations can be computed from:

$$\text{forecast}(\text{overall},s,t\_fc)=\sum_{k} w(k)\text{forecast}(k,s,t\_fc) \tag{5}$$

And the overall forecast confidence intervals can then be obtained by integrating the weighted sum of the individual forecast PDFs.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀I | ▶I
◀ | ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

3.2 BMA experiments for ENSURF-IBIROOS

We have implemented several BMA's making use of the existing capability of Matroos visualization tool. The BMA is applied to the surge component of sea level forecast and this is done basically because this component can be approximated by a normal distribution to a reasonable degree of accuracy, which is not the case for total sea level including tides, especially for strong semidiurnal tidal regimes. This is the reason why all the validation results at the end of this section refer to the surge or meteorological component. The total sea level forecast is nevertheless available operationally in ENSURF as we add the tide computed from tide gauge data to the different forecasts, including the BMA, as we explained in previous section (Fig. 2).

As it has been mentioned, the implementation of ENSURF for IBIROOS and the Western Mediterranean represents the first activity of operational multi-model forecast in the region, so several experiments were performed during the development of the system. In many cases, some of the sources available had not been particularly validated with respect to sea level, as their initial objective was general ocean forecast including parameters such as currents, temperature or salinity. In these cases, ENSURF has allowed the detection of important problems sometimes related to the tide of the model, others to the boundary conditions or to the re-initialization scheme; all these problems leave its clue in the sea level time series. As a consequence of this situation, those sources which presented lower Correlation Index with observations were not included in the BMA implementations of ENSURF. Several institutions are still working on the improvement of some aspects of the models that hopefully will provide better forecasts of sea level in the near future. Such is the case of MeteoGalicia (Spain) or Marine Institute (Ireland).

So the following initial BMA's were implemented in the region, taking into account the reliable sources available and their common domains (we will distinguish between Atlantic and Western Mediterranean coast):

Atlantic: available sources: *nivmar, eseoat, imi, metfr* (3 sources) and *metga*. In this case, we have output from two baroclinic sources, *eseoat* and *imi*. Four BMA's were implemented avoiding metga and imi sources, due to their low Correlation Index. A maximum number of 5 sources is available (TP being the Training Period):

- BMA0: *eseoat* and *nivmar*, TP = 15 days
- BMA_ibi1: *eseoat, nivmar, metf*r (3 sources), TP = 7 days
- BMA_ibi2: *eseoat, nivmar, metfr* (3 sources), TP = 4 days
- BMA_ibi3: *eseoat, nivmar, metfr* (3 sources), TP = 15 days

Mediterranean: available sources: *nivmar* and *metfr* (3 sources). In the Mediterranean the four sources are barotropic. We used all the sources available, 4 in total, for the BMA implementation, changing also the TP, as in the Atlantic coast:

- BMA_med1: *nivmar, metfr* (3 sources), TP = 7 days
- BMA_med2: *nivmar, metfr* (3 sources), TP = 4 days
- BMA_med2: *nivmar, metfr* (3 sources), TP = 15 days

The BMA was implemented at particular stations or harbours, that were selected based on the availability of enough sea level forecasts or sources, and automatic near-real time quality control of tide gauge data (Fig. 9). At the present stage of ENSURF implementation, there are still several harbours where only one forecast exists. On the other hand, the quality control procedure was at the moment of starting this work only applied for Puertos del Estado tide gauges or REDMAR network (Pérez et al., 2008). This software is currently being extended to all IBIROOS stations within MyOcean project, and will be applied by Puertos del Estado also for other Mediterranean stations included in ENSURF in the future. All the BMA's were in operation during ECOOP Training Operational Period (TOP). Results of the validation of all the sources and the BMA's will be shown in next section.

It should be kept in mind that both observations and forecast data may not be complete, so the BMA allows dealing with missing data, and a weight $w(k)$ and a *sigma(k)* will be determined as long as there is at least one forecast-observation combination in the training period. This may lead to problems with the BMA, difficult to avoid. On the other hand, taking into account the problem of the bias between models and observations previously mentioned, the technique is implemented in ENSURF with a bias correction, which in fact improves greatly the results and makes the BMA reasonable, independently of this difference.

4 Validation results

Basic statistic parameters (Root Mean Square Error: RMSE, Correlation Index: CI, Maximum Error: RMAX and Mean difference: Bias) were computed from the comparison between the different models and the tide gauge observations, including the different BMAs, which have been treated here as additional sources, for the period September 2008 to December 2009. In order to synthesize all the data, we have averaged and plotted the CI and RMSE parameters of all the stations and sources on the Iberian Atlantic coast and on the Mediterranean coast (Figs. 10 and 11, respectively). In the following sections we will point to different aspects of the conclusions from these first validation results.

4.1 Barotropic vs. baroclinic sources

One of the first objectives of the validation, before checking the performance of the BMA, was comparing the output of baroclinic general circulation models operational in the region, with the standard storm surge applications based on barotropic, vertically-integrated models. This is a test that at the moment of writing this paper was only possible on the Atlantic facade.

Figure 12 shows the RMSE and CI of several sources for Bilbao station (North of Spain), with respect to tide gauge observations. Best performance is found for *nivmar*, what is not strange as it is the only source that automatically and dynamically corrects the bias based on the observations. However, the interesting thing is that Meteo-France forecasts (*mf-aladin*, *mf-arpege*, and *mf-ecmwf*), also barotropic but without tide gauge data assimilation, get better statistical results than the baroclinic forecasts from *eseoat* and *imi*. This is something that happens often at the stations of the Atlantic coast, and which is an important and interesting conclusion about the capability of baroclinic models to reproduce correctly sea level variations; averaged statistical parameters for all the Atlantic stations (Fig. 10) confirm this point. Also from Fig. 10, some problems become evident with sources *imi* and specially *metga_sm*, which show poorer statistics, the reason why they were not used for the BMA implementation; institutions responsible for these systems are investigating the causes. On the other hand, *eseoat* shows a relatively good performance, taking into account that it does not make use of tide gauge data, although not enough to improve the results of the barotropic sources. This is also an even more interesting result if we take into account that meteorological forcing resolution for *eseoat* is comparable to the ones of the barotropic sources (two of the *metfr* sources in fact have coarser resolution) and that horizontal resolution of *eseoat* ocean model is higher ($3'$ facing the $10' \times 15'$ resolution of *nivmar* and $5'$ of *metfr*).

4.2 General performance of the BMA's

If we pay attention now to the capability of the BMA's implemented to outperform the individual forecasts, it can be seen from Fig. 10 that this is in general the case for the Atlantic coast, with higher CI and lower RMSE for practically all of the BMA's, but more clearly for *bma_ibi2*, the one with 4 days of training period; the best source is *nivmar*. Nevertheless, the differences between these values of CI and RMSE are sometimes very small, and probably not statistically significant.

Results for the Mediterranean are presented in Fig. 11. In this case there were no baroclinic sources available in ENSURF at the moment of writing this paper.

Nevertheless it can be said that the BMA's do not improve the results of *nivmar* so clearly now, being only bma_med2 with 4 days of training period the one showing a slight improvement; on the other hand, Meteo-France results are poorer than in the Atlantic. One possible reason for the latter is the presence of a boundary on MeteoFrance model domain around Sardinia and Corsica, which disturbs the results at these stations; taking into account the experience with *nivmar* system, we think the Mediterranean sea should be completely covered by the model domain.

It can be said that generally all the BMA's present very good results in the Atlantic, improving the performance of the best of the sources, and that this is not always the case in the Mediterranean. It is important to clarify that these are mean statistic parameters, and that results can differ slightly depending on the station and the period of data.

4.3 Influence of the data period on validation results

In order to determine the influence of the data period used for the validation, and taking into account the existence of months with very low storm activity in the initial period September 2008 to December 2009, we performed the computation of statistical parameters for other periods. We selected for the Atlantic stations the stormy season of January to February 2009, where most of the largest surge events since the implementation of ENSURF were present (see example of results for the two periods at Gijón station, Table 2).

The first evident result is that all the sources show an improvement of their statistical validation parameters, especially significant for those performing worse for the previous period, such as *imi* or *metga_sm*. One possible explanation for this could be the proximity in time to the period used for adjustment of the bias; in fact most of the sources present a drift along the months with respect to observations. Some institutions are investigating the reason for this drift; an interesting point is that, although easy to understand that barotropic models do not include all low frequency variations of sea level, especially those related to steric effects, these should be present in baroclinic sources such as *eseoat* and *imi* that should include any kind of forcing for sea level variability. This problem may indicate that it is necessary to implement an operational bias correction based on observations, as *nivmar* does, for the rest of sea level forecasts.

Another important point is that the BMA's, that improved the statistic parameters in practically all the stations in the Atlantic for the initial period September 2008 to December 2009, only get this improvement for the 50% of the stations (Gijón, Bonanza, Huelva and Vigo), when limiting the period of study to January–February 2009. In the rest of stations (Bilbao, Santander, Coruña and Vilagarcía) *nivmar* gets the best results. This is in some sense a strange result, as it seems the BMA it is working now with in principle "better" models, at least during these two months, if we take into account the explanation above.

However, when repeating the validation analysis for a different stormy period, from mid-November 2009 to January 2010, further away from the bias correction months, the BMA performs better than for the whole period, especially in the Mediterranean, where the BMA shows a larger improvement with respect to "*nivmar*", something contradictory with the first result.

So as a general conclusion, results are better for the individual sources when selecting a shorter period for validation, and BMA gets better performance when using the whole period of data of one year or more, or the last months of this period, farther away from the bias correction. From this we can conclude that statistical results depend not only on the period of data being stormy or not, but probably also on other problems such as gaps in the sources or observations, which will vary for different periods, as well as on the presence of malfunction on the tide gauges. The latter is something that is minimized in this study, as a delayed quality control was applied in some way to data stored in Matroos data base, but that it is difficult to avoid in the operational mode, even with near-real time quality control procedures, as these are never perfect and able to deal with all kinds of malfunction.

4.4 Performance during the peak of a storm

The results presented up to now represent average statistic parameters, obtained from a relative long period of data, and so they reflect the general behaviour of the models and the BMA for all meteorological conditions. However, the objective of a good sea level forecast should be an adequate simulation of the peak of a storm. Figure 12 shows the forecasts for the largest storm of the period of study, at Bilbao station. As we pretend to check the improvement of the BMA, for the sake of simplicity we show just the output from *nivmar* and *eseoat* sources, as well as the output of the three different BMA's with different training periods.

It can be seen that the peak of the storm is better reproduced by *nivmar* source, and that in this case, the BMA's available do not improve the forecast, although they do better than *eseoat* in any case. This is an important result that should be explored in detail. We also see at this particular harbour that the BMA which gets closer to the peak is bma_ibi1, with 15 days of training period, although this can be different at another harbour. In fact, if we have a look to the statistical parameters for the whole period for Bilbao, we see that these are better for a training period of 4 days (*bma_ibi2*). This is contradictory with what it is expected and with results for the North Sea (Beckers et al., 2008), where peaks were reproduced better with shorter training periods.

It is also remarkable the reasonably good forecast of *nivmar* of the peak of the storm. One possible explanation for this is the adequate representation of the continental platform, very narrow here, for the *nivmar* system: it was manually and carefully corrected before final implementation. This is something that was not done in the rest of the sources, not so focused on sea level, such as *eseoat*.

It seems, in any case, that a better determination of the BMA weights and parameters may be needed for an adequate forecast of extreme events. Beckers et al. (2008) suggest already this idea and propose to determine these weights based on the performance of the models during extreme meteorological conditions instead of during a recent training period.

5 Conclusions and future work

ENSURF has proved its utility as a validation and multi-model forecast tool, and has become the first experience of exchange of operational forecasts for the IBIROOS and Western Mediterranean regions. It has allowed the detection of problems on existing operational models that had not been previously well calibrated with respect to sea level observations, and for the first time a better statistical forecast is feasible based on existing operational systems and is already implemented and operational for sea level in the NOOS and IBIROOS regions. Extension to other parameters and regions is possible and easy.

First validation results of the surge component, based on the comparison between tide gauge data and the forecasts at the harbours show that, at least for the IBIROOS region, baroclinic models do not reach the good performance of barotropic models for storm surge applications and that there is a general improvement of performance of the BMA forecast, more clear in the Atlantic stations. This improvement becomes less evident if we change the data period and concentrate on the stormy season and there is some difficulty of the BMA to reproduce the peak of the storm, as compared to *nivmar* source, for example.

On the other hand, availability of near-real time data from tide gauges, with automatic quality control to avoid wrong data to enter the system, becomes one of the main conditions to provide accurate forecasts of sea level at the harbours and of course for the implementation of the BMA technique.

Future work will focus on the addition of new sources and extension to the whole Mediterranean, the more relevant and urgent ones being the operational forecasting systems established within MyOcean project for IBIROOS and MOON. Within the same project, the automatic quality control of tide gauge data will be applied to the rest of sea level stations in Europe, what will allow completing the BMA implementation and validation for other countries contributing to ENSURF. Finally a detailed study of the influence of the training period in the BMA performance or the extension to 2-D fields should be the goal in the near future.

Acknowledgements. This work has been supported through the EU ECOOP project, Contract No. 036355. We would like to thank the personnel working at the different institutions who have collaborated in providing the sources to ENSURF. We thank also Ronan Creach (from SHOM, France), Vibeke Hess (from DMI, Denmark), Elizabeth Bradshaw (from POL, UK) and Gonzalo Crisostomo (from IGN, Portugal), for their support and cooperation for making the tide gauge data available. Finally, we thank José Antonio Zaballos for the development of the ENSURF-IBIROOS front page.

References

Alvarez Fanjul, E., Pérez Gómez, B., and Rodríguez, I.: A description of tides in the Eastern North Atlantic, Prog. Oceanogr., 40, 217–244, 1997.

Alvarez Fanjul, E., Pérez, B., and Sánchez-Arévalo, I. R.: Nivmar: A storm surge forecasting system for the Spanish Waters. Scientia Marina, 65, 145–154, 2001.

Backhaus, J. O.: A semi-implicit scheme for the shallow water equations for application to shelf sea modeling, Cont. Shelf Res., 2(4), 243–254, 1983.

Beckers, J. V. L., Sprokkereef, E., and Roscoe, K. L.: Use of Bayesian model averaging to determine uncertainties in river discharge and water level forecasts. Proc.4th International Symposium on Flood Defence: Managing Flood Risk, Reliability and Vulnerability, Toronto, Ontario, Canada, 6–8 May 2008.

Charnock, H.: Wind stress on a water surface, Q. J. Roy. Meteorol. Soc., 81, 639–640, 1955.

Chézy, A.: Formule pour trouver la vitesse de l'eau conduit dan une rigole donnée, Dossier 847 (MS 1915) of the manuscript collection of the École National des Ponts et Chaussées, Paris, 1776, reproduced in: G. Mouret, Antoine Chézy: histoire d'une formule d'hydraulique, Annales des Ponts et Chaussées 61, Paris, 1921, 1776.

Flather, R. A.: Practical surge prediction using numerical models, pp. 21–43, in: Floods Due to High Winds and Tides, edited by: Peregrine, D. H., London: Academic Press, 109 pp., 1981.

Flather, R. A.: Estimates of extreme conditions of tide and surge using a numerical model of the north-west European continental shelf. Estuarine, Coast. Shelf Sci., 24, 69–93, 1987.

Foreman, F. G. G.: Manual for tidal heights analysis and predictions, Institute of Ocean Sciences, Patricia Bay, Pacific Marine Report no. 77(10), 101 pp., 1977.

GETECH: Geophysical Exploration TECHnology internal report, Dept. Earth Sciences, University of Leeds, UK, 1995.

Hamill, T. M., Snyder, C., and Morss, R. E.: A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles, Mon. Weather Rev., 128, 1835–1851, 2000.

Hoeting, J. A., Madigan, D. M., Raftery, A. E., and Volinsky, C. T.: Bayesian model averaging: A tutorial (with discussion), Stat. Sci., 14, 382–401, A corrected version is available on line at www.stat.washington.edu/www/research/online/hoeting1999.pdf, 1999.

Kass, R. E. and Raftery, A. E.: Bayes factors, J. Amer. Stat. Assoc., 90, 773–795, 1995.

Large, W. G. and Pond, S.: Open ocean momentum flux measurements in moderate to strong winds, J. Phys. Oceanogr., 11, 324–336, 1981.

Leamer, E. E.: Specification Searches, Wiley, 370 pp., 1978.

Leith, C. E.: Theoretical skill of Monte-Carlo forecasts, Mon. Weather Rev., 102, 409–418, 1974.

Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T.: The ECMWF ensemble system: Methodology and validation, Q. J. Roy. Meteorol. Soc., 122, 73–119, 1996.

Mourre, B., De Mey, P., Mènard, Y., Lyard, F., and Le Provost, C.: Relative performance of future altimeter systems and tide gauges in constraining a model of North-Sea high-frequency barotropic dynamics, Ocean Dynamics, 56, 473–486, 2006.

Pérez, B., Vela, J., and Alvarez-Fanjul, E.: A new concept of multi-purpose sea level station: example of implementation in the REDMAR network, in: Proceedings of the Fifth International Conference on EuroGOOS, May 2008: Coastal to global operational oceanography: achievements and challenges, Exeter, UK, 2008.

Poole, D. and Raftery, A. E.: Inference for Deterministic Simulation Models: The Bayesian Melding Approach, J. Am. Statist. Association; ABI/INFORM Global, p. 1244, 2000.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowsly, M.: Using Bayesian Model Averaging to Calibrate Forecast Ensembles, Am. Meteorological Soc., 133, 1155–1174, 2005.

Rodríguez, I., Alvarez, E., Krohn, J., and Backhaus, J.: A mid-scale tidal analysis of waters around the north-western corner of the Iberian Peninsula, Proc. Computer Modelling Ocean Eng., 91, Balkema, 1991.

Sotillo, M. G., Jordi, A., Ferrer, M. I., Conde, J., Tintoré, J., and Alvarez-Fanjul, E.: The ESEOO regional ocean forecasting system. In: Proceedings of the ISOPE-2007, The 17th International Offshore Ocean and Polar Engineering Conference, Lisbon, Portugal, 2007.

Sotillo, M. G., Alvarez-Fanjul, E., Castanedo, S., Abascal, A. J., Menéndez, J., Emelianov, M., Olivella, R., García-Ladona, E., Ruiz-Villarreal, M., Conde, J., Gómez, M., Conde, P., Gutiérrez, A. D., and Medina, R.: Towards an operational system for oil-spill forecast over Spanish waters: Initial developments and implementation test, Marine Pollution Bulletin, 56, 686–703, 2008.

Toth, Z. and Kalnay, E.: Ensemble forecasting at the NMC: The generation of perturbations, B. Am. Meteorol. Soc., 74, 2317–2330, 1993.

Unden, P., Rontu, L., Jarvinen, H., Lynch, P., Calvo, J., Cats, G., Cuxart, J., Eerola, J., Fortelius, C., Garcia-Moya, J. A.,Jones, C., Lenderlink, G., McDonald, A., McGrath, R., Navascues, B., Nielsen, N. W., Odegaard, V., Rodriguez, E., Rummukainen, M., Room, R., Sattler, K., Sass, B. H., Savijarvi, H., Schreur, B. W., Sigg, R., The, H., and Tijm, A.: HIRLAM-5 Scientific Documentation. HIRLAM-5 Project, c/o Per Unden SMHI, SE-601 76, Norrkopping, Sweden, 2002.

Wu, J.: Wind-stress coefficients over sea surface from breeze to hurricane, J. Geophys. Res., 87, 9704–9706, 1982.

Table 1. Sources or models contributing to ENSURF-IBIROOS. Name of the sources are the ones used within Matroos visualization tool.

Institution	Country	Source/model	Output
OPPE	Spain	Nivmar: Hamsom model (barotropic) including tide gauge data assimilation	Surge and total sea level
		Eseoat: Polcoms model (baroclinic)	Total sea level
Meteo-France	France	Metfr_arpege: barotropic, met forcing 0.25°	Surge
		Metfr_aladin: barotropic, met forcing 0.1°	Surge
		Metfr_ecmwf: barotropic, met forcing 0.5°	Surge
Marine Institute	Ireland	Imi: ROMS model (baroclinic)	Total sea level
MeteoGalicia	Spain	Metga_sm: Mohid model (barotropic)	Surge

Table 2. Statistical parameters of the validation for the different sources at Gijón station (North Spain), for the whole period and for just the stormy months of January and February 2009. Data of RMSE in meters.

Gijón station	Sep 2008–Dec 2009		Jan 2009–Feb 2009	
Source	RMSE (m)	C.I.	RMSE (m)	C.I.
Nivmar	0.042	0.94	0.042	0.96
Eseoat	0.055	0.90	0.047	0.95
Mf-aladin	0.072	0.83	0.047	0.96
Mf-arpege	0.072	0.83	0.046	0.96
Mf-ecmwf	0.071	0.83	0.045	0.96
Metga_sm	0.080	0.74	0.045	0.96
Imi	0.102	0.58	0.052	0.97
BMA	0.036	0.96	0.038	0.97

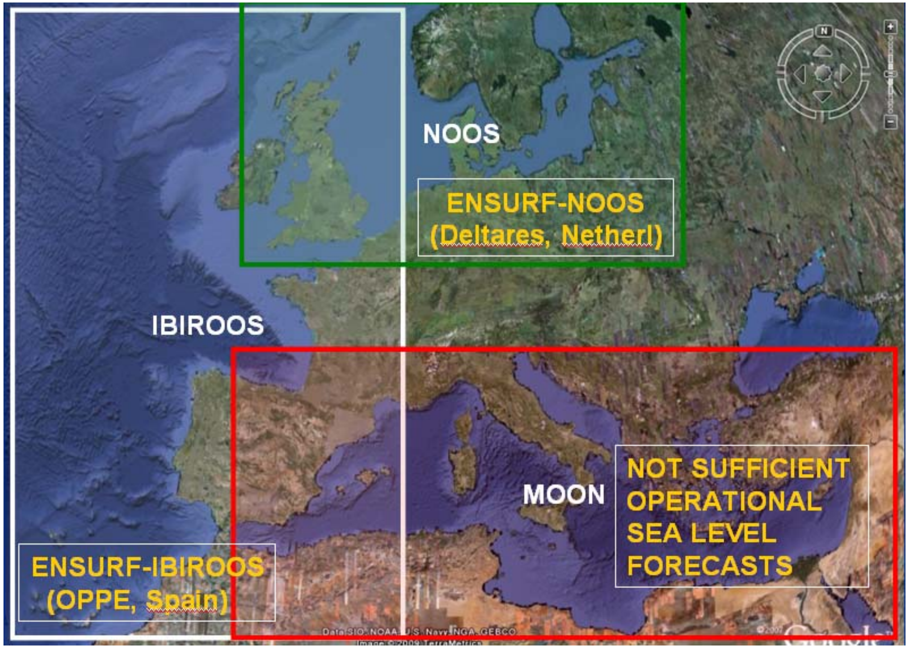
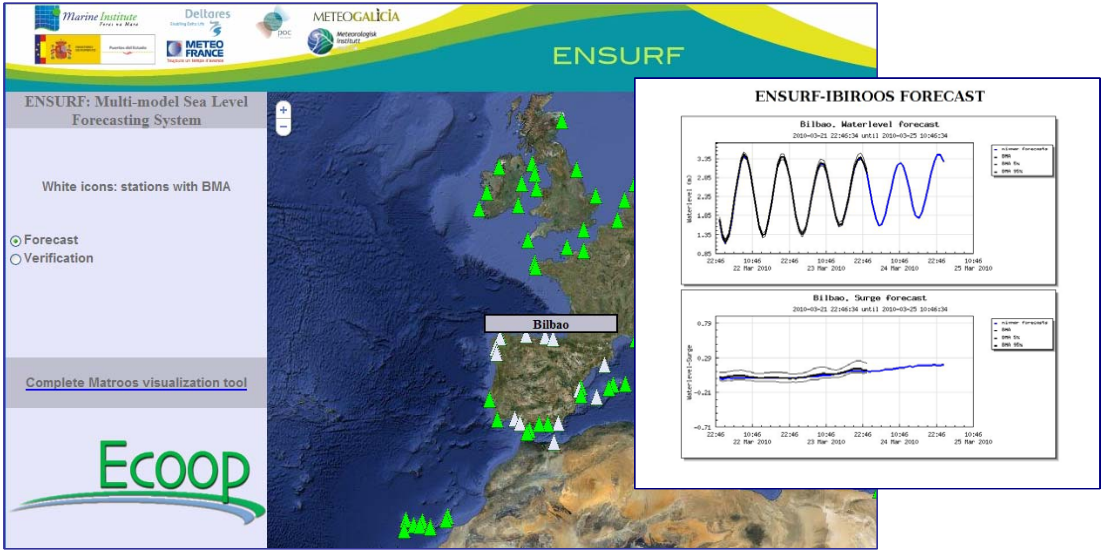


Fig. 1. ENSURF components for the three main operational oceanographic regions.



ENSURF: Multi-model Sea Level Forecasting System

White icons: stations with BMA

● Forecast
○ Verification

Fig. 2. ENSURF IBIROOS portal: http://ensurfibi.puertos.es. It allows the display of the forecast including the BMA confidence interval or the verification with tide gauge data. The figure shows an example of forecast for Bilbao station, both for total and meteorological sea level.

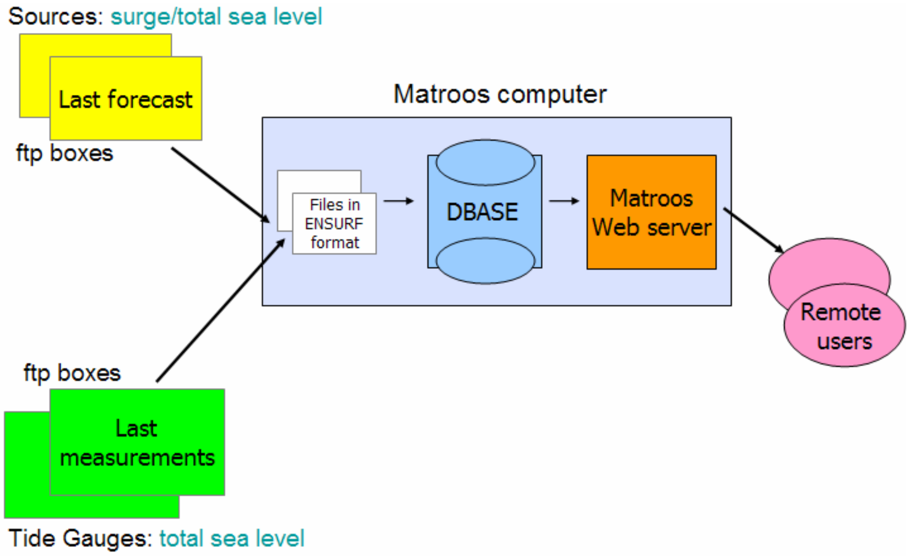


Fig. 3. ENSURF system architecture, showing the data flow and Matroos structure.

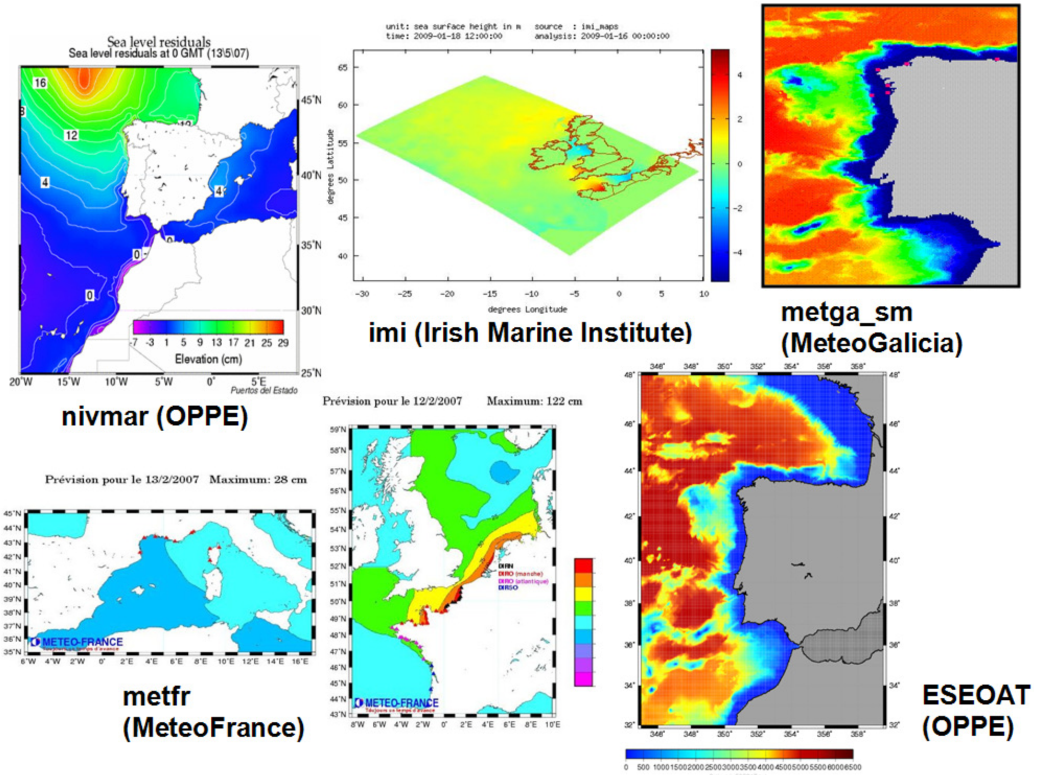
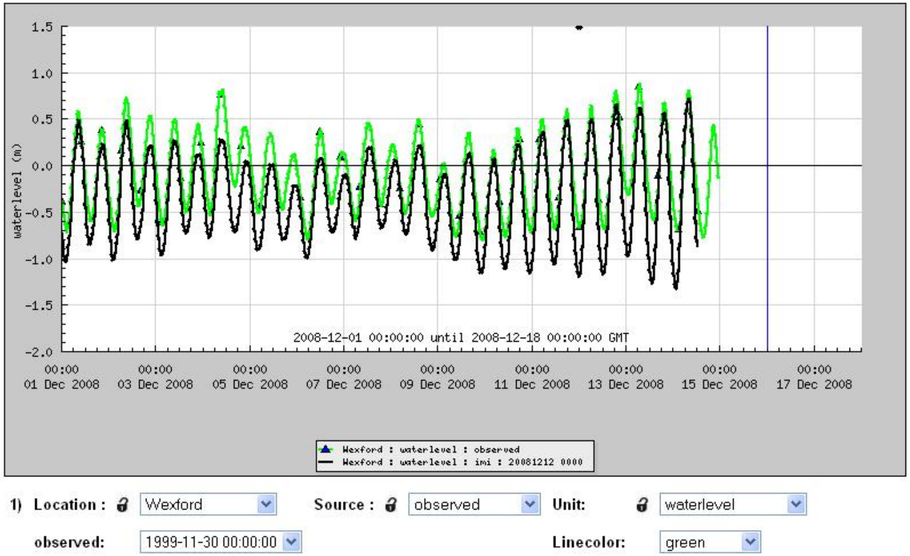


Fig. 4. Domains of the sources available for the ENSURF-IBIROOS component. Nivmar covers the whole Mediterranean sea, but only results at Western Mediterranean are presented.



1) Location : Wexford Source : observed Unit: waterlevel

observed: 1999-11-30 00:00:00 Linecolor: green

Fig. 5. Example of bias between Marine Institute of Ireland (imi source) forecast and tide gauge observations at Wexford harbour (ENSURF-IBIROOS).

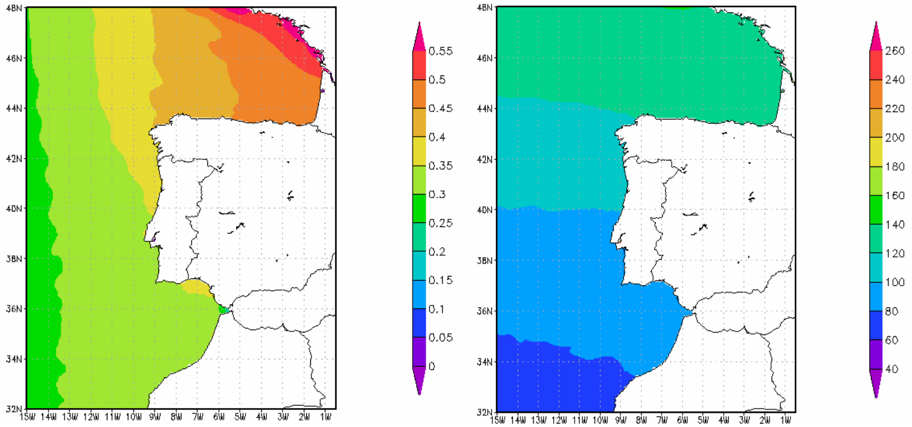


Fig. 6. Output of S2 harmonic constituent (amplitude, left, and phase, right) result of the harmonic analysis of one year of data at all the grid points of "eseoat" source (ENSURF-IBIROOS).

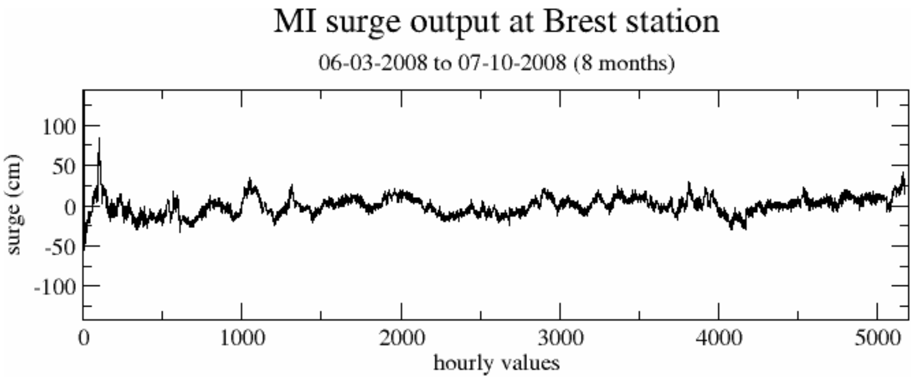
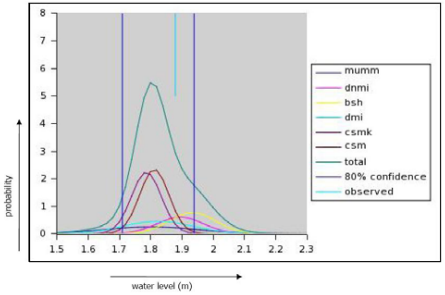
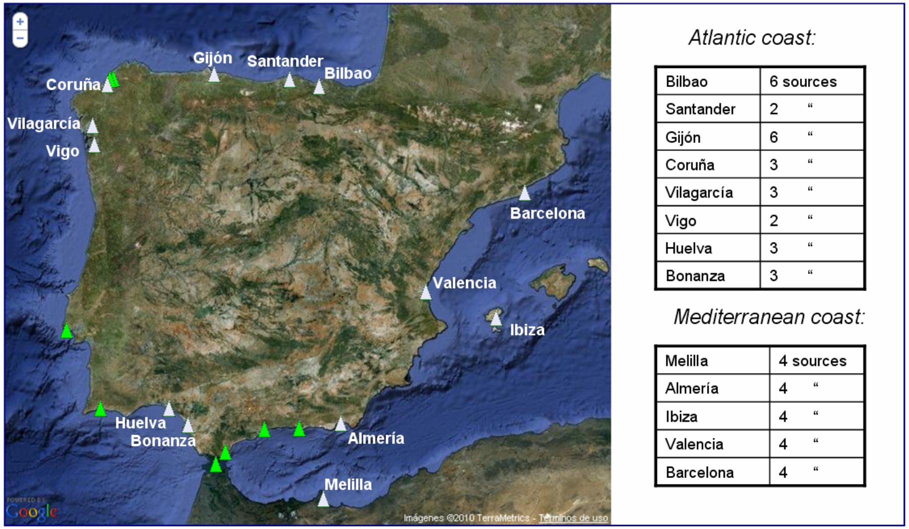


Fig. 7. Surge component of “imi” source (Marine Institute) at Brest tide gauge, after tide extraction. The normal appearance of the surge from the model (no spikes or reference changes or tidal oscillations), allows to confirm the correct introduction of the tide in the model. This was not always the case during ENSURF development.



Model	Average weight	RMS error (m)
MUMM	0.15	0.16
DNMI	0.15	0.19
BSH	0.19	0.15
DMI	0.16	0.18
DCSMK	0.23	0.09
DCSM	0.23	0.09
BMA	-	0.08

Fig. 8. Left: individual and overall pdf's for a single 24 h forecast (27 October 2006 13:00, Delfzjil), based on 6 models for the North Sea component of ENSURF. The 80% confidence interval is marked on dark blue, the actual observed values was 1.87 m (blue), within the confidence interval. Right: results for the BMA and the individual forecasts for the period 2003–2006 (6 stations) (extracted from Beckers et al., 2008).



Atlantic coast:

Bilbao	6 sources
Santander	2 “
Gijón	6 “
Coruña	3 “
Vilagarcía	3 “
Vigo	2 “
Huelva	3 “
Bonanza	3 “

Mediterranean coast:

Melilla	4 sources
Almería	4 “
Ibiza	4 “
Valencia	4 “
Barcelona	4 “

Fig. 9. Stations for which the BMA was implemented (white icons) for the first ENSURF-IBIROOS implementation, based on the availability of quality control of tide gauge data in near-real time, and more than two sources of forecast.

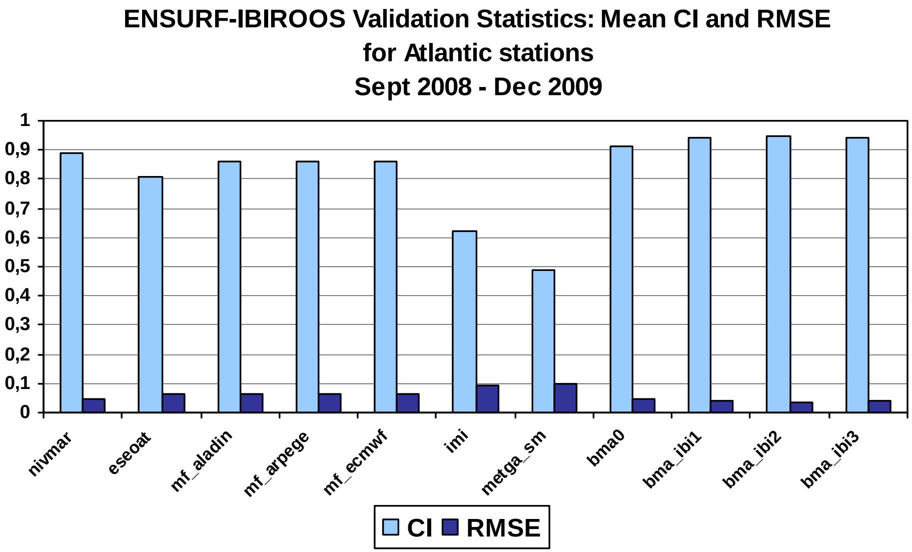


Fig. 10. Mean CI (Correlation Index) and RMSE (Root Mean Square Error) for the sources and stations of the Iberian Atlantic coast (September 2008 to December 2009).

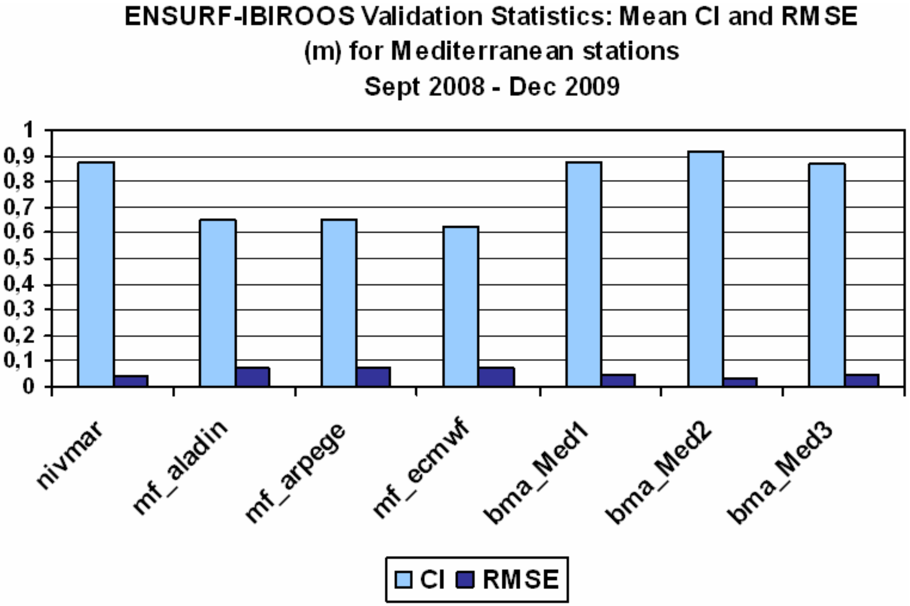
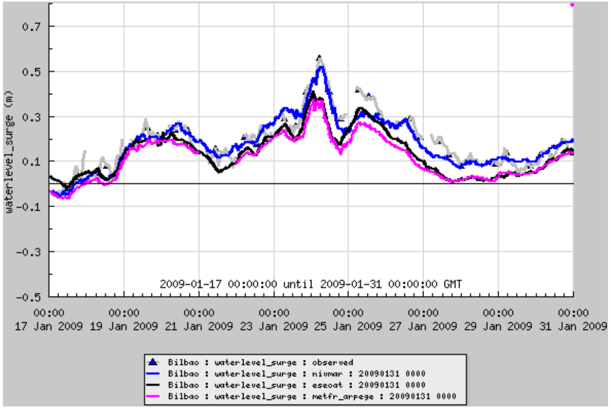


Fig. 11. Mean CI (Correlation Index) and RMSE (Root Mean Square Error) for the sources and stations of the Mediterranean coast (September 2008 to December 2009).



Model	RMSE	CI
Nivmar	0.041	0.96
Eseoat	0.062	0.91
Mf-aladin	0.049	0.95
Mf-arpege	0.049	0.95
Mf-ecmwf	0.050	0.95
Imi	0.054	0.95

Fig. 12. Comparison of baroclinic models (*eseoat* and *imi*) and barotropic ones (*nivmar* and *metfr* models) for the stormy period January–February 2009 at Bilbao station. Blue colour in the table (right) used for the sources with better statistical parameters. (Data in meters, RMSE: root mean square error, CI: Correlation Index).