

# The NARVAL software toolbox in support of ocean models skill assessment at regional and coastal scales

Pablo Lorente<sup>1\*</sup>, Marcos G. Sotillo<sup>1</sup>, Arancha Amo-Baladrón<sup>1</sup>, Roland Aznar<sup>1</sup>, Bruno Levier<sup>2</sup>, Lotfi Aouf<sup>3</sup>, Tomasz Dabrowski<sup>4</sup>, Álvaro De Pascual<sup>1</sup>, Guillaume Reffray<sup>2</sup>, Alice Dalphiné<sup>3</sup>, Cristina Toledano<sup>1</sup>, Romain Rainaud<sup>3</sup> and Enrique Álvarez-Fanjul<sup>1</sup>

<sup>1</sup> Puertos del Estado, Madrid, Spain.

<sup>2</sup> Mercator Ocean, Toulouse, France.

<sup>3</sup> MeteoFrance, Toulouse, France.

<sup>4</sup> Marine Institute, Galway, Ireland.

**Abstract.** The significant advances in high-performance computational resources have boosted the seamless evolution in ocean modelling techniques and numerical efficiency, giving rise to an inventory of operational ocean forecasting systems with ever-increasing complexity. The skill of the Iberia-Biscay-Ireland (IBI) regional ocean forecasting system, implemented within the frame of the Copernicus Marine Environment Monitoring Service (CMEMS), is routinely evaluated by means of the NARVAL (North Atlantic Regional VALidation) web-based toolbox. Multi-parameter comparisons against observational sources (encompassing both in situ and remote-sensing platforms) are regularly conducted along with model intercomparisons in the overlapping areas. Product quality indicators and skill metrics are automatically computed not only averaged over the entire IBI domain but also over specific sub-regions of particular interest in order to identify strengths and weaknesses of each model. The primary goal of this work is three-fold. Firstly, to provide a flavor of the basic functionalities of NARVAL software package in order to elucidate the accuracy of IBI near real time forecast components (physical, biogeochemical and waves); secondly, to showcase a number of the practical applications of NARVAL; finally, to present the future roadmap to build a new upgraded version of this software package, which will include the quality assessment of multi-year and interim products, the computation of long-term skill metrics or the evaluation of event-oriented multi-model intercomparison exercises. This synergistic approach, based on the integration of numerical models and multi-platform observational networks, should be useful to comprehensively characterize the highly dynamic sea states and the dominant modes of spatio-temporal variability.

**Keywords:** model skill assessment, ocean forecasting, model intercomparison.

## 1 Introduction

Over recent decades, substantial advances have been achieved in the discipline of operational oceanography thanks to the significant increase in high-performance computational resources, among others: multicore processor-based cluster architectures, massive storage capabilities, optimized parallelization and efficient scalability strategies. Such progresses have boosted the seamless evolution in ocean modelling techniques and numerical efficiency, thereby giving rise to an inventory of operational ocean forecasting systems (OOFs) with ever-increasing complexity. Nowadays, more sophisticated and memory-demanding simulations can be effectively run at shorter time intervals and finer spatio-temporal resolutions for coupled forecast systems that might include entangled air-sea, wave-current and/or biophysical interactions. In this context, the implementation of operational data assimilation schemes has constituted a quantum leap in terms of realistic forecast predictions since they maximize the interconnection of multi-platform ocean observing systems and OOFs. This integrated approach provides critical oceanographic information to support wise decision-making in the marine environment with subsequent societal benefits.

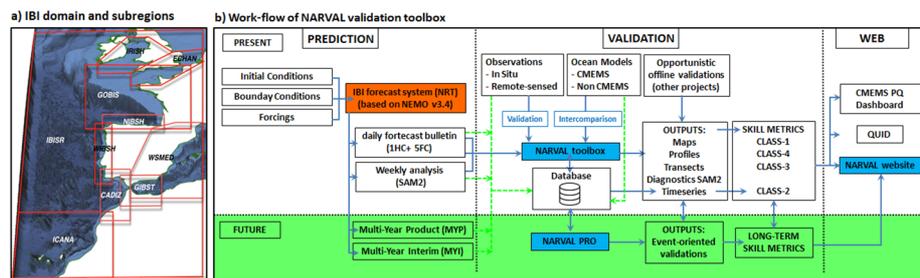
In the framework of the Copernicus Marine Environment Monitoring Service (CMEMS), a global ocean model together with a wealth of nested regional OOFs are currently running in different areas of the European seas and providing paramount oceanographic forecast products. Since the comparison of OOFs against independent observations constitutes a core activity in CMEMS, the development of skill assessment software packages and dedicated web applications is an active theme. In particular, the accuracy of Iberia-Biscay-Ireland (IBI) regional OOFs is routinely evaluated by means of the NARVAL (North Atlantic Regional VALidation) system [1, 2], a web-based toolbox that provides a number of skill metrics automatically computed. NARVAL tool has been implemented to routinely monitor IBI performance and to objectively evaluate model's veracity and prognostic capabilities. Both real-time comparisons ('online mode') and regular-scheduled 'delayed-mode' comparisons (for longer time periods) are performed using a wealth of independent observational sources as benchmark, among others: in situ observations (from moorings, tide-gauges, drifters, gliders and ARGO floats networks) and remote-sensed estimations (provided by satellites and High-Frequency radars -HFR hereinafter-).

NARVAL is modular and flexible enough to assess the quality of a variety of near real-time (NRT) forecast products, encompassing the physical (IBI-NRT-PHY), biogeochemical (IBI-NRT-BIO) and waves (IBI-NRT-WAV) components. Furthermore, model intercomparisons are regularly conducted in the overlapping areas to elucidate pros and cons of each model performance. Product quality indicators and skill metrics are automatically computed not only averaged over the entire IBI domain but also over specific sub-regions of particular interest from a user perspective (i.e. coastal or shelf areas) in order to infer IBI accuracy and the spatiotemporal uncertainty levels (Fig. 1, a).

The main goal of this contribution is to showcase: i) the current practical applications of NARVAL software toolbox to evaluate the performance of IBI-NRT forecasting system and ii) the future roadmap to build a new upgraded version here named

NARVAL-PRO, which will include a number of novelties such as the accuracy assessment of multi-year (MY) and interim (INT) products or the computation of long-term skill metrics (Fig. 1, b).

The paper is organized as follows: Section 2 provides further details about the model quality assessment framework and the workflow of NARVAL software toolbox. Section 3 describes the diverse NARVAL components and outlines a wealth of illustrative examples. Finally, main conclusions, ongoing work and future prospects are summarized in Section 4.



**Fig. 1.** a) IBI regional service domain (IBISR) and defined sub-regions: Irish Sea (IRISH), English Channel (ECHAN), Gulf of Biscay (GOBIS), North Iberian Shelf (NIBSH), Western Iberian Shelf (WIBSH), Gulf of Cadiz (CADIZ), Strait of Gibraltar (GIBST), Western Mediterranean (WSMED) and Canary Islands (ICANA); b) Workflow of NARVAL toolbox and future prospects.

## 2 Quality assessment framework and workflow of NARVAL

The comparison of OOFSS against independent quality-controlled measurements constitutes a core activity in oceanographic operational centers (Fig. 1, b) since it aids: i) to infer the relative strengths and weaknesses in the modelling of several key physical processes; ii) to compare different versions of the same OOFSS and evaluate potential improvements and degradations before a new version is transitioned into operational status; iii) to compare coarse resolution ‘father’ and nested high-resolution ‘son’ systems to quantify the added value of the downscaling approach adopted. With regards to the third aspect, IBI forecast products are regularly inter-compared not only against other CMEMS regional model solutions (e.g. its parent system, the GLOBAL) in the overlapping areas but also against other non CMEMS models by means of NARVAL tool. Complementarily, opportunistic intercomparisons are conducted in the frame of diverse EU-funded projects such as MEDESS-4MS [3].

The agreement between both in situ and remote-sensing instruments and the ocean forecasting system is evaluated by means of computation of a set of statistical metrics traditionally employed in this framework: histograms, bias, root mean squared differences (RMSE), scalar and complex correlation coefficients, current roses, histograms, quantile-quantile (QQ) plots and the best linear fit of scatterplots. Skill metrics have

been defined in four different types, including gridded model output (CLASS-1), time-series at specified locations and sections (CLASS-2), transports through sections and other integrated quantities (CLASS-3), and metrics of forecast capability (CLASS-4).

The statistical metrics regularly generated by NARVAL are online delivered to inform end-users and stakeholders about the quality and reliability of the marine forecast products routinely delivered, fostering downstream services and user uptake. This is achieved thanks to the Quality Information Document -QUID- [4], which is periodically updated and freely available in CMEMS website (<http://marine.copernicus.eu/>). Equally, the skill metrics can be also displayed through the CMEMS Product Quality Dashboard (<http://marine.copernicus.eu/services-portfolio/scientific-quality/>).

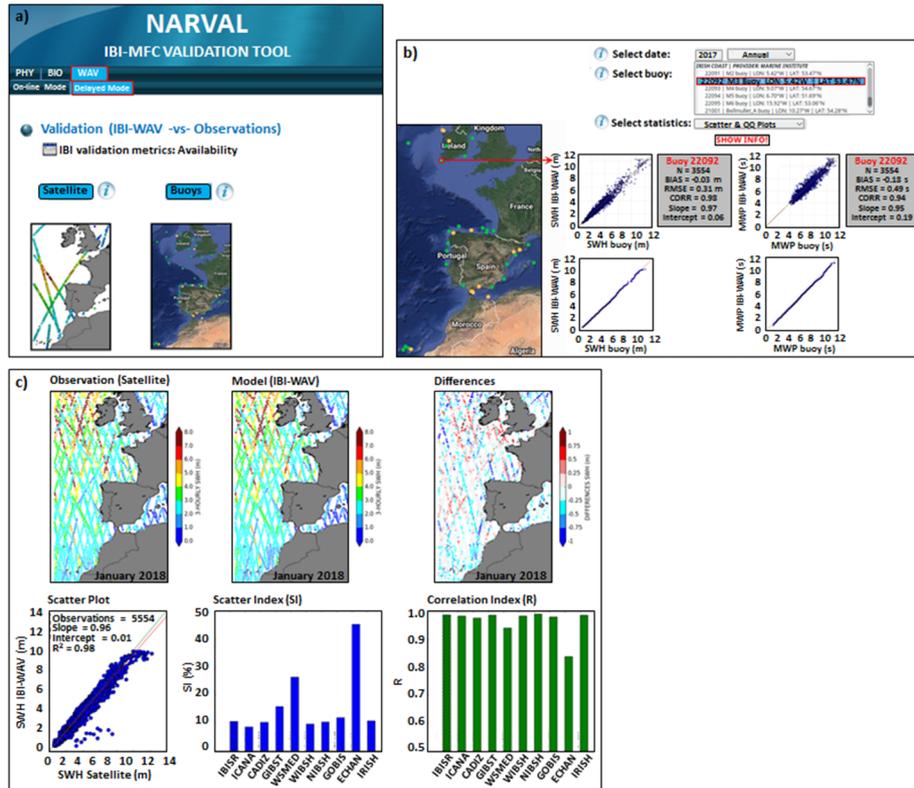
### 3 NARVAL components

In this section, basic features of NARVAL toolbox to evaluate the quality of the three components (physical, biogeochemical and waves) of the IBI near real time operational suite are described.

#### 3.1 IBI-NRT-WAV

The operational IBI near real time wave forecast system (IBI-NRT-WAV), based on MF-WAM, provides a 5-day regional wave forecast, which is updated twice a day (cycles at 00 z and 12 z). The model performs a partitioning technique on wave spectra that allows the separation between sea wind and primary and secondary wave swell systems [5]. The model was implemented on the IBI domain with a grid size of 10 km and with a spectral resolution of 24 directions and 30 frequencies, starting from 0.035 Hz. The IBI-WAV runs are driven by 3-hourly analyzed winds provided by the European Center for Medium-Range Weather Forecasts (ECMWF). The boundary conditions (wave spectra) are provided by the global CMEMS wave system, which uses the assimilation of altimeter wave data.

A specific module has been implemented in NARVAL toolbox to compare this forecast system against all the available observations on both online and delayed mode (Fig. 2, a). The coverage area of IBI regional domain includes an array of deep and coastal buoys which provide hourly-averaged quality-controlled in situ measurements of significant wave height (SWH), mean wave period (MWP), the wave period at spectral peak (PKP), and the mean wave direction (WD). Such buoys are used in concert as a robust benchmark to conduct a multi-parameter skill assessment of the IBI-NRT-WAV. An example of annual (2017) comparison focused on the Irish Sea is here provided: scatter and Quantile-Quantile (QQ) plots for SWH and MWP, along with the associated skill metrics, clearly proved that the model properly captured basic features of the wave regime in a region with one of the most energetic wave climate in Europe, with a number of wave height events clearly above 11 m (Fig. 2, b).



**Fig. 2.** a) Snapshot of the NARVAL website, focused on the skill assessment of IBI-NRT-WAV on “delayed-mode”; b) NARVAL section devoted to compare the model against coastal (orange dots) and deep-water (green dots) buoys: Annual comparison (2017) of SWH (left) and MWP (right) at the model grid point closest to buoy 22092 (red circle): scatter and QQ plots. N represents the number of hourly observations; c) Monthly comparison against satellite-derived observations (January 2018): map of SWH differences, scatter plot and analysis for different sub-regions are presented.

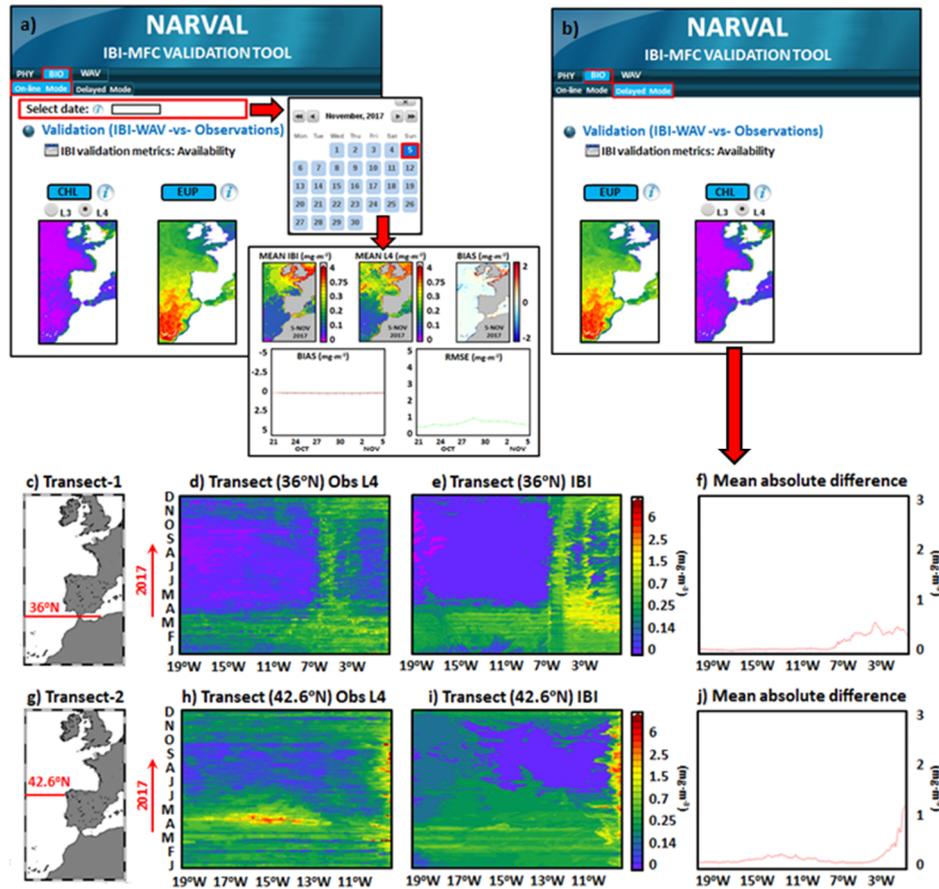
Complementarily, the wave altimetry product used to quantify the skill of IBI-NRT-WAV model comprises a pool of three different satellite missions (Jason-2, Saral/Altika, and Cryosat-2) that is subsequently merged and prepared by Meteo-France. The satellite-sensed SWH estimations have been spatially averaged on a  $0.1^\circ$  grid and chosen for specific three-hourly time steps (00–21 h) in order to associate the measured and simulated significant wave heights and objectively assess the quality of IBI-NRT-WAV on a daily and monthly basis. Maps of differences and the best linear fit of scatterplots are routinely computed through NARVAL. According to the maps shown in Fig. 2-c, the resemblance between observed and simulated significant wave height is significant for January 2018. The statistical results derived from the best linear fit confirm it, with the slope (intercept) close to 1 (0) and the correlation coefficient fairly above 0.95. The monthly analysis for specific sub-regions (defined in Fig.

1-a) reveals that IBI-NRT-WAV accuracy is lower in the Western Mediterranean (WSMED) and in the English Channel (ECHAN) since higher (lower) scatter index (correlation) is observed for those specific sub-regions. This feature has been persistently observed along the entire 2018 (not shown), highlighting the ability of NARVAL to monitor IBI-NRT-WAV performance and detect its strengths and weaknesses [6].

### 3.2 IBI-NRT-BIO

The operational IBI near real time biogeochemical forecast system (IBI-NRT-BIO) is based on NEMO v3.6 circulation model and PISCES v2 biogeochemical model. The latter provides 24 prognostic variables, running simultaneously with the ocean physics, with the same  $1/36^\circ$  horizontal resolution ( $\sim 2$  km). The operational system provides a near-real-time short-term (7-days) forecast of the main biogeochemical variables: chlorophyll, oxygen, iron, nitrate, ammonium, phosphate, silicate, net primary production, euphotic zone depth and phytoplankton carbon.

The skill of IBI-NRT-BIO forecast products is regularly assessed through a devoted module of NARVAL. As this process is seriously handicapped by the scarcity of in situ bio observations, only satellite-derived observations of chlorophyll and euphotic zone are currently being used on both “online mode” (Fig. 3, a) and “delayed mode” (Fig. 3, b). Regarding the former, by selecting a parameter (CHL-L4) and a specific date (5<sup>th</sup> of November 2017) from the calendar, a panel is exposed with daily-averaged maps along with a map of differences and the daily evolution of a variety of skill metrics for the last 15 days (Fig. 3, a). In order to provide a deeper insight into IBI model performance during 2017, a qualitative model-observation comparison was performed: Hovmöller diagrams were computed at two selected transects of constant latitude with the main aim of analyzing the temporal evolution of the daily chlorophyll concentration in key regions like the Strait of Gibraltar or the Galician upwelling system. In this example, a relevant model-observation resemblance is observed in the Strait of Gibraltar and the Alboran Sea ( $6^\circ\text{W}$ - $1^\circ\text{W}$ ), where quasi-permanent peaks of chlorophyll were satisfactorily reproduced by IBI along the entire year 2017 (Fig. 3, c-e). The mean absolute difference (MAD) remained moderate along the transect, with a relative peak detected over the Alboran Sea, indicating a model overestimation (Fig. 3, f). On the other hand, IBI-NRT-BIO appeared to capture basic characteristics of the NW Iberian upwelling system such as the intensification of the chlorophyll concentration during specific summer coastal upwelling events when northerly winds are predominant (Fig. 3, g-i), although the MAD was also higher in this region (Fig. 3, j).



**Fig. 3.** a-b) Snapshot of NARVAL web, focused on the “online mode” and “delayed mode” quality assessment of IBI-NRT-BIO products, respectively; c-j) Transects of constant latitude and the associated Hovmöller diagrams and mean absolute differences.

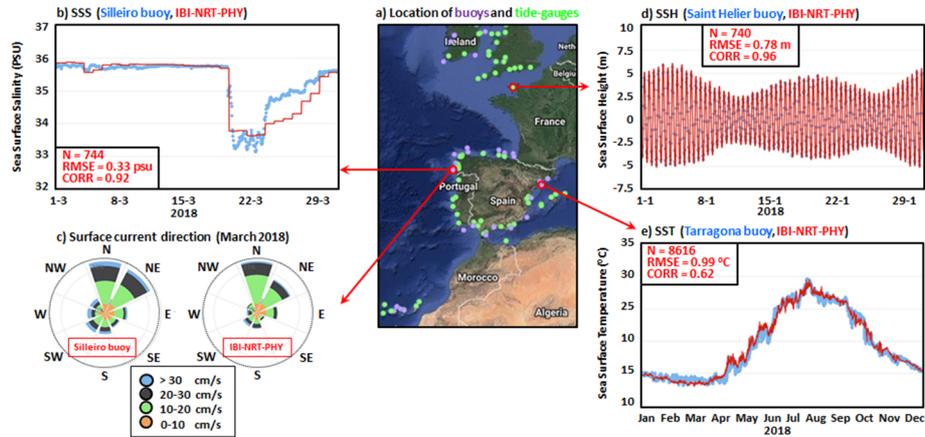
### 3.3 IBI-NRT-PHY

The operational IBI near real time physical forecast system (IBI-NRT-PHY) provides a short-term 5-day hydrodynamic 3D forecast of a range of physical parameters (currents, temperature, salinity and sea level) since 2011 [1]. The system is based on an eddy-resolving NEMO v3.6 model application, run at  $1/36^\circ$  horizontal resolution on an Arakawa-C grid and 50 geopotential vertical levels, assuming hydrostatic equilibrium and Boussinesq approximation. Final products are routinely delivered in a service domain extending between  $19^\circ\text{W}$ - $5^\circ\text{E}$  and  $26^\circ\text{N}$ - $56^\circ\text{N}$ . The IBI run is forced every 3 hours with up-to-date high-frequency meteorological forecasts provided by ECMWF. Lateral open boundary data are interpolated from the daily outputs of the CMEMS GLOBAL system. These are complemented by 11 tidal harmonics built from FES2004 and TPXO7.1 tidal models solutions. A SAM2-based data assimilation

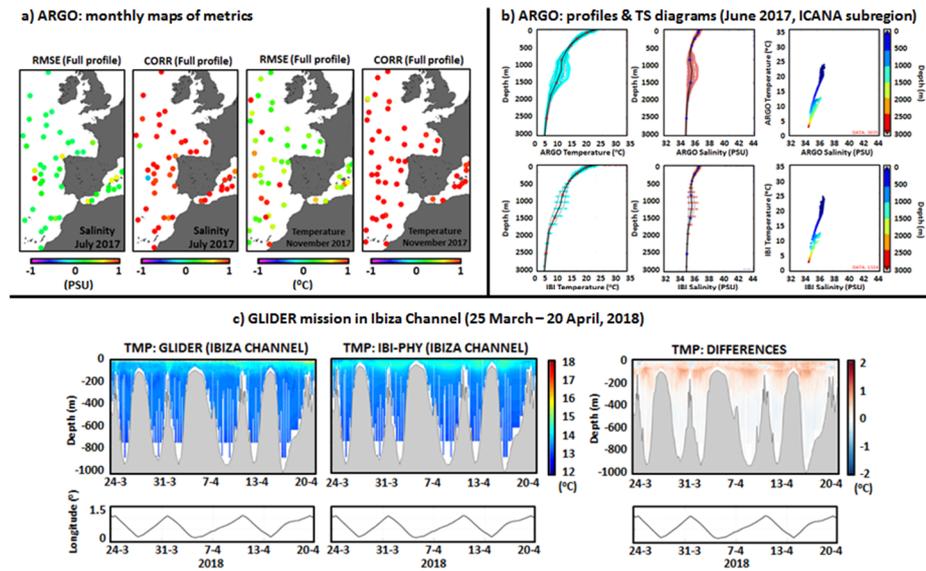
scheme was recently introduced (April 2018) in order to enhance IBI predictive skills but will not be further described here.

NARVAL has been implemented to carry out direct comparison of model outputs with quality-controlled hourly time series of in situ observations. To this aim, the skill assessment software ingests daily model forecasts and extracts the time series on those grid points closest to the available in situ sensors within IBI regional domain (Fig. 4, a). Both modelled and observed datasets are inserted into a relational database (Fig. 1, b) for a long-term storage and subsequently retrieved and visualized through an intuitive georeferenced web interface (not shown). This interactive approach allows computing a variety of CLASS-2 metrics to evaluate the performance of IBI-NRT-PHY system in coastal areas. A number of examples of the multi-parameter skill assessment of IBI-NRT-PHY against in situ hourly observations are presented below (Fig. 4, b-e). Hourly in situ SSS data collected by Silleiro buoy during March 2018 experienced an abrupt decrease from 36 PSU down to 33 PSU during the 20<sup>th</sup> of March due to freshwater river discharges (Fig. 4, b). IBI-NRT-PHY outputs at the closest grid point appeared to properly capture both the sharp drop in SSS values and the persistent low salinity values for the next 4-day period, as reflected by a monthly correlation coefficient of 0.92 and a RMSE of 0.33 PSU. There is also a noticeable resemblance between the monthly current roses derived from in situ observations and model predictions in terms of speed and mean direction (Fig. 4, c), showing the predominance of the so-called Iberian Poleward Current, flowing northwards. The monthly comparison of the sea level in an energetic tidal area such as the English Channel reveals the consistent performance of the forecast system, according to the skill metrics obtained (Fig. 4, d). Finally, IBI-NRT-PHY correctly reproduced the annual cycle of the sea surface temperature in the Western Mediterranean during the entire 2018, as confirmed by a significantly high correlation of 0.99 (Fig. 4, e).

For the sake of completeness, supplementary works with in situ observations in the entire three-dimensional (3D) water column have been undertaken to achieve a comprehensive model skill assessment (Fig. 5). A comparison of IBI temperature and salinity profiles against ARGO floats within IBI regional domain are routinely conducted on a monthly basis thanks to NARVAL web tool. CLASS-4 metrics are computed for the whole water column and for different layers, being the specific levels considered: 0-5 m, 5-200 m, 200-600 m, 600-1500 m, 1500-2000 m (not shown). According to the skill metrics derived from the monthly comparison against ARGO floats, IBI-NRT-PHY seems to properly capture the vertical distribution of temperature and salinity (Fig. 5, a). Furthermore, the resemblance between both datasets is significantly high and the Temperature-Salinity (TS) diagrams look rather alike (Fig. 5, b). As a result of strong alliances with local partners, IBI-NRT-PHY model solution is verified against specific glider missions. In particular, the exercises performed in the Ibiza Channel by SOCIB are used to assess the model 3D consistency. As reflected in Fig. 5-c, the model performance seems to be rather consistent (especially in lower depth levels), whereas moderate discrepancies are mainly found in the first 200 m.

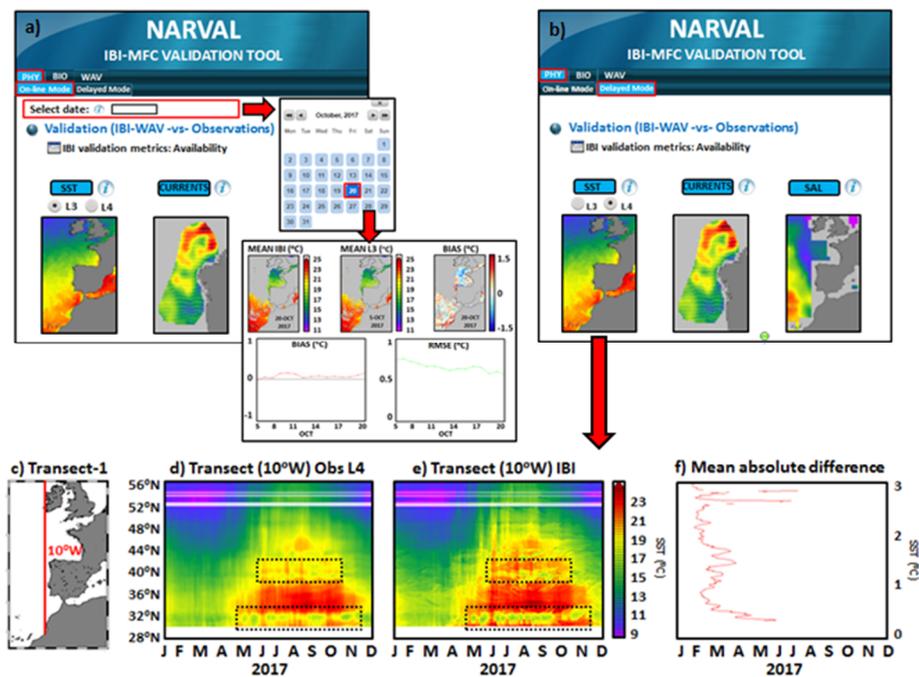


**Fig. 4.** a) Location of buoys (purple dots) and tide-gauges (green dots) used to evaluate the quality of IBI-NRT-PHY forecast system; b-e) CLASS-2 multi-parameter comparison at selected model grid points closest to moorings location: timeseries comparison, surface current roses and skill metrics are automatically computed on a different time basis, ranging from monthly to seasonal and annual frequencies.



**Fig. 5.** a) Maps of metrics (RMSE and correlation) derived from the monthly comparison (July 2017) of IBI-NRT-PHY outputs against in situ full-profiles of salinity and temperature provided by ARGO floats; b) Monthly qualitative comparison between daily model outputs and in situ observations in ICANA sub-region (June 2017): profiles of temperature and salinity along with TS diagrams; c) Glider mission in Ibiza Channel: temperature profiles observed, modelled and differences (courtesy of SOCIB) during March-April 2018.

Complementarily, other skill metrics (CLASS-1 and CLASS-4) are regularly computed by means of NARVAL on both “online mode” (Fig. 6, a) and “delayed mode” (Fig. 6, b). Regarding the former, by selecting a parameter (SST-L3) and a specific date (20<sup>th</sup> of October 2017) from the calendar, a panel is exposed with daily-averaged maps along with a map of differences and the daily evolution of a variety of skill metrics for the last 15 days (Fig. 6, a). Hovmöller diagrams at selected transects of constant longitude (Fig. 6, c) are also calculated along with the mean absolute difference in order to properly monitor the temporal evolution of the daily sea surface temperature in key regions. As it can be observed in Fig. 6 (d-f), diagrams look rather alike and IBI appears to properly capture basic features like the annual cycle or the African and Galician upwelling systems where a sudden cooling (represented by black dotted boxes) take place when northerly winds blow during summertime and move surface waters away from the coast, which are replaced by cooler water that wells up from below.



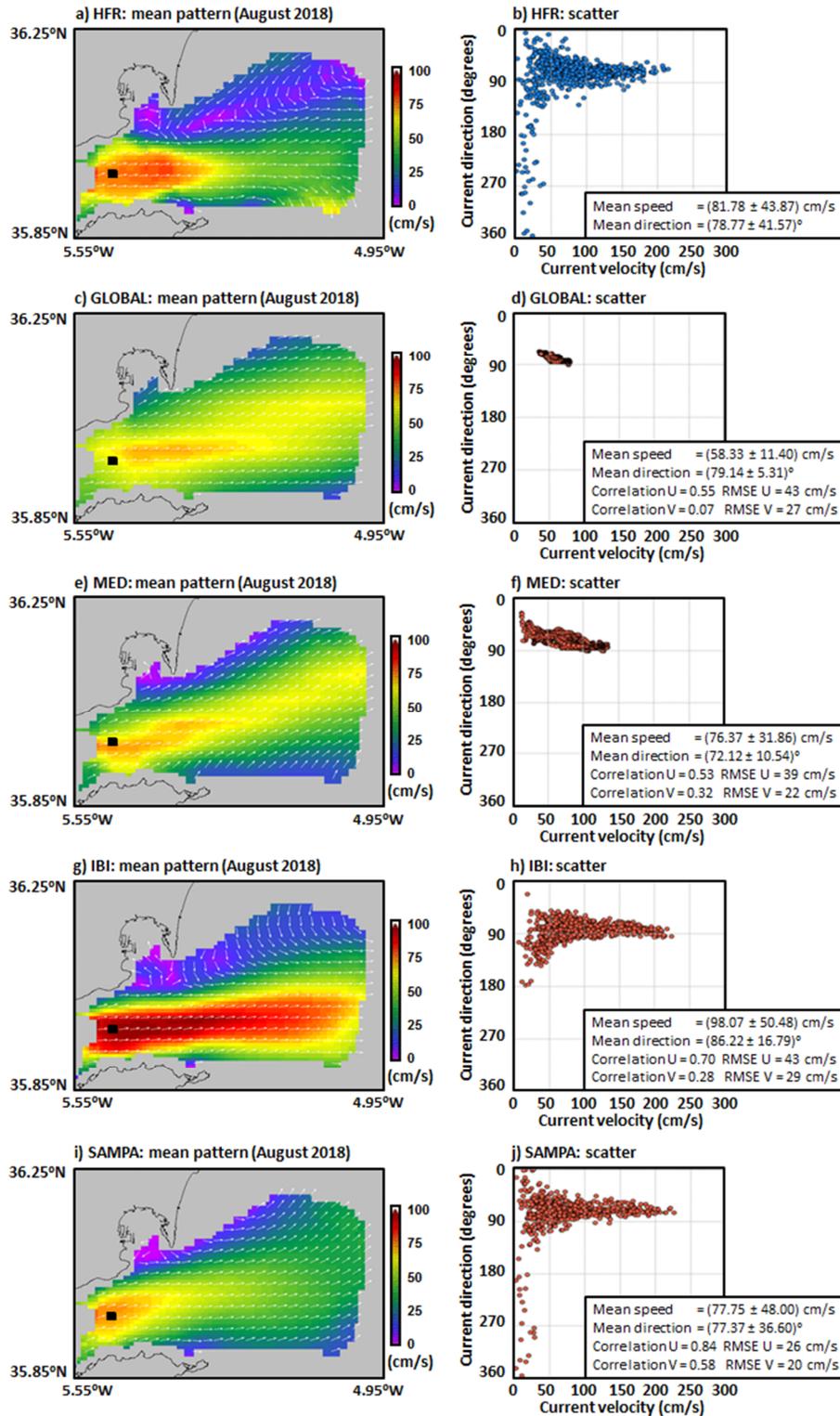
**Fig. 6.** a-b) Snapshot of NARVAL web, focused on the “online mode” and “delayed mode” skill assessment of IBI-NRT-PHY products, respectively; c-f) Transect of constant longitude and the associated Hovmöller diagram and mean absolute differences. Black dotted boxes represent coastal upwelling of cold waters.

An additional aspect addressed in NARVAL toolbox is the multi-parameter inter-comparison of diverse ocean forecast models in the overlapping regions, conducted at the sea surface, ranging from global to regional and local scales [7]. Here we present a

multi-model intercomparison exercise for August 2018 in the Strait of Gibraltar among three CMEMS forecast systems (GLOBAL, IBI and MED – being the last two systems nested to GLOBAL) and the SAMPA high-resolution coastal forecast system (embedded in IBI) in order to elucidate the accuracy of each system to characterize the Atlantic Jet (AJ) inflow dynamic (Fig. 7). To this end, a HFR system has been used as benchmark since it regularly provides quality-controlled hourly maps of the surface currents of the Strait [8]. The qualitative inspection of monthly-averaged circulation maps reveals that each forecast system reproduces reasonably well the eastward AJ inflow into the Mediterranean as previously observed in HFR estimations (Fig. 7, a), but differ in the intensity and direction of the mean surface inflow. Whilst GLOBAL (Fig. 7, c) and MED (Fig. 7, e) appear to underestimate the speed in the Strait of Gibraltar, IBI seem to overestimate it, exhibiting besides a more zonal surface flow (Fig. 7, g).

However, SAMPA outperforms the parent systems (IBI and GLOBAL) and MED by better replicating the orientation and strength of the inflow (Fig. 7, i). A quantitative CLASS-2 comparison at the selected grid point (indicated by a black square in Fig. 7-a) was assessed (Fig. 7, right panels). The scatter plot of HFR-derived hourly current speed versus direction (taking as reference the North and positive angles clockwise) revealed that the AJ flowed predominantly eastwards, forming an angle of  $79^\circ$  (Fig. 7, b). The current velocity, on average, was  $82 \text{ cm}\cdot\text{s}^{-1}$  and reached peaks above  $200 \text{ cm}\cdot\text{s}^{-1}$ . Speeds below  $50 \text{ cm}\cdot\text{s}^{-1}$  were registered along the entire range of directions. Westwards currents, albeit minority, were also observed. In the case of the scatter plot derived from GLOBAL estimations, substantial discrepancies were detected as the variability of both the AJ direction and speed were clearly limited (Fig. 7, d). No flow reversals were detected and peak velocities of the eastward flow were underestimated. In the monthly scatter plot of regional MED (Fig. 7, f) and IBI estimations (Fig. 7, h), surface current velocities below  $20 \text{ cm}\cdot\text{s}^{-1}$  were barely replicated and the AJ inversion was not observed. Despite IBI appeared to properly portray the mean characteristics of the eastwards flow, the model tended to privilege zonal flow directions. By contrast, the scatter plot of SAMPA estimations presented a significant resemblance in terms of prevailing current velocity and direction (Fig. 7, j). The main features of the AJ were qualitatively reproduced and surface flow reversals to the west were properly captured. Accordingly, the skill metrics obtained for SAMPA coastal system are better than those derived for regional and global model solutions.

In summary, this exercise reflects the added value of the dynamical downscaling performed through the SAMPA coastal system with respect to IBI regional solution (in which SAMPA is nested). Overall, a steady improvement in the AJ characterization is evidenced in model performance when zooming from global to coastal configurations, highlighting the benefits of the downscaling approach adopted and also the potential relevance of a variety of factors at local scale, among others: a more refined horizontal resolution, a tailored bathymetry and/or a higher spatio-temporal resolution of the atmospheric forcing.



**Fig. 7.** (left) Monthly-averaged surface circulation patterns for August 2018. Black squares indicate the grid point selected to conduct the comparison; (right) Scatter of hourly current direction (taking as reference the North and positive angles clockwise) versus current speed for August 2018 at the selected grid point.

## 4 Conclusions and future work

In this work, a general overview of the main features of NARVAL toolbox has been presented, with special emphasis on the rigorous skill assessment of a wealth of CMEMS IBI near real time forecasting products (WAV, BIO and PHY). Some of the current practical applications of NARVAL have been showcased, highlighting the benefits of a synergistic approach based on the integration of numerical models (CMEMS IBI) and observational networks, used in tandem to comprehensively characterize the highly dynamic sea states and the dominant modes of spatio-temporal variability.

With the advent of new technologies (coastal altimetry, autonomous underwater vehicles, BIO ARGO floats, etc.), a combined use of multi-platform, multi-scale observing systems encompassing both in situ (buoys, tide gauges, etc.) and remote (HFR, satellite, etc.) sensors will provide further insight into the comprehensive characterization of the shelf's surface circulation and will also contribute positively to a more exhaustive model accuracy assessment.

The future roadmap to build a new upgraded version denominated NARVAL-PRO includes the extension of its capabilities to evaluate the quality of multi-year and interim products or the computation of long-term skill metrics in order to gain insight into the evolution of IBI model performance (Fig. 1, b).

NARVAL is a live software package in terms of seamless evolution and continuous upgrades. Successive versions will be developed with a focus on the inclusion of: i) novel skill metrics; ii) new observational platforms; and iii) new ocean forecasting systems, implemented at both regional and coastal scales. In the framework of My-Coast and OCASO projects, several regional systems are currently being intercompared by means of NARVAL. Likewise, this toolbox will also play a relevant role in order to evaluate the benefits of downscaling approaches in coastal areas since a variety of operational port-scale forecasts products have been recently developed under the umbrella of SAMOA project, aimed at implementing a fully integrated monitoring service to increase safety and efficiency of marine operations in the Spanish harbors.

Ancillary verification exercises should focus on the evaluation of ocean models ability to accurately reproduce singular oceanographic processes. Since the NARVAL tool is devoted to intercompare model solutions on a monthly, seasonal or annual basis, part of the picture is missing due to traditional time averaging. An event-oriented multi-model intercomparison methodology would allow better quantifying the skill of each system to capture small-scale coastal processes. Those oceanographic events subject of further insight might encompass, among others: i) coastal upwelling, downwelling and relaxation episodes; ii) fronts and submesoscale eddies; iii) extreme events.

## References

1. Sotillo, M.G., Cailleau, S., Lorente, P., Levier, B., Aznar, R., Reffray, G., Amo-Baladrón, A. and Álvarez-Fanjul, E.: The MyOcean IBI Ocean Forecast and Reanalysis Systems: Operational products and roadmap to the future Copernicus Service, *Journal of Operational Oceanography*, 8 (1), 1-18 (2015).
2. Lorente P., Piedracoba S., Sotillo M.G., Aznar R., Amo-Baladrón, A., Pascual, A., Soto-Navarro J., Álvarez-Fanjul, E.: Ocean model skill assessment in the NW Mediterranean using multi-sensor data. *Journal of Operational Oceanography*, doi: 10.1080/1755876X.2016.1215224 (2016).
3. Sotillo M.G., Amo-Baladrón, A., Padorno, E., García-Ladona, E., Orfila, A., Rodríguez-Rubio, P., Conte, D., Jiménez J.A., De los Santos, F.J. and Álvarez-Fanjul, E.: How is the surface Atlantic water inflow through the Strait of Gibraltar forecasted? A lagrangian validation of operational oceanographic services in the Alboran Sea and Western Mediterranean, *Deep Sea Research*, 133, 100-117 (2016).
4. Sotillo, M.G., Levier, B., and Lorente, P.: Quality Information Document for the Atlantic - Iberian Biscay Irish- Ocean Physics Analysis and Forecasting Product: IBI\_ANALYSIS\_FORECAST\_PHYS\_005\_001. CMEMS Technical Report, 1-94 (2018).
5. Rainaud, R, Aouf, L, Dalphinet, A., Sotillo, M.G., Alvarez-Fanjul, E., Reffray, G., Levier, B., Lawchune, S., Lorente, P. and Toledano, C. The impact of wave physics in the CMEMS-IBI ocean system Part A: wave forcing validation. *Ocean Sci Discuss*, <https://doi.org/10.5194/os-2018-165> (2019).
6. Lorente, P., M.G. Sotillo, L. Aouf, A. Amo-Baladrón, E. Barrera, A. Dalphinet, C. Toledano et al.: Extreme Wave Height Events in NW Spain: A Combined Multi-Sensor and Model Approach. *Remote Sensing* 10 (1). doi:10.3390/rs10010001 (2018).
7. Lorente, P., Sotillo, M.G., Amo-Baladrón, A., Aznar, R., Levier, B., Sánchez-Garrido, J.C., Sammartino, S., Pascual, A., Reffray, G., Toledano, C. and Álvarez-Fanjul, E.: Skill assessment of global, regional and coastal circulation forecast models: evaluating the benefits of dynamical downscaling in IBI surface waters. *Ocean Sci Discuss*, <https://doi.org/10.5194/os-2018-168> (2019).
8. Lorente P., Piedracoba S., Sotillo M.G., Álvarez-Fanjul, E.: Long-Term Monitoring of the Atlantic Jet through the Strait of Gibraltar with HF Radar Observations. *J. Mar. Sci. Eng.*, 7, 3; doi:10.3390/jmse7010003 (2019).