**CHAPTER 4 - Specific events during the year 2020**

### **Supplementary material**

**4.1 Western Mediterranean record-breaking storm Gloria: an integrated assessment based on models and observations**

**Authors:** Alvarez-Fanjul, E., Pérez Gómez, B., de Alfonso Alonso-Muñoyerro, M., Lorente Jiménez, P., García Sotillo, M., Lin-Ye, J., Aznar Lecocq, R., Ruíz Gil de la Serna, M., Pérez Rubio, S., Clementi, E., Coppini, G., García-León, M., Fernandes, M., García Valdecasas, J., García Valdecasas, J.M., Santos Muñoz, D., Luna Rico, M.Y., Mestres, M., Molina, R., Tintore, J.,Mourre, B., Masina, S., Mosso, C., Reyes, E., Santana, A.

### **Assessment of model forecasting services performance**

### **Wave forecast validation**

In figure C.4.1.1., the evolution of the storm is represented through comparative time series of the significant wave height (Hm0) and the mean direction, from the buoys and the Portus forecasting model. The reanalysis of the Hm0, during Gloria differed in less than 1m from the corresponding buoy measurements. The only exception was Tarragona. The model also presented a smoother evolution of Hm0 than the observations. The predictions in calm periods were smaller than 0.4m. The wave direction was more erratic during the calm period preceding and continuing Storm Gloria, but it turned predominantly Eastern in Valencia, Tarragona and Begur, the three peninsular sites. It was mainly North-Eastern, in Dragonera. The prediction of the wave direction was practically coincident with the measurements.

Interfaz de usuario gráfica

Descripción generada automáticamente

*Figure C.4.1.1. Evolution of the spectral significant wave height (Hm0) during storm Gloria, as measured by the buoy in Dragonera (a), Valencia (b), Tarragona (c) and Begur (d). The re-analyses are also presented. The dates are shown in “day-month-year hour” (UTC). The solid dots are the reanalyses, the lighter dots are the buoy measurements, the darker arrows are the wave directions from the reanalysis, the lighter arrows are the observed wave directions (the positive y-axis is 0º and the direction increases clockwise). CMEMS data product ref-4.1.5 for observations and data from ref-4.1.7 for modeled time series*

### **Sea level forecast validation**

Each ENSURF ensemble member provides a storm-surge signal, mainly differing in the spatial domain of each model and their intrinsic physics. None of these models addresses high frequency processes. Also, even though the spatial and temporal resolutions of the atmospheric forcing differ, all of them use the same model (ECMWF-IFS). This last issue eases the intercomparison among different sea-level forecasts because the atmospheric surface fields (winds and mean seal level pressure) share similarities.

Figure C.4.1.2. show how all models correctly reproduce the basic features of the storm in terms of sea level residuals. For the sake of clarity, only the surge component is presented, although ENSURF forecasts total sea level. The best statistical behavior corresponds, depending on the station, to ENSURF or MedMFC, being generally Nivmar the less accurate system for this particular storm.

It is important to note that analysis of other storms can show different behavior of the forecasting systems. For other historical events, Nivmar has been shown as the most accurate system (Pérez et al., 2021). This is reflected in the behavior of the BMA system, which does not rely on a single model and is the result of the previous statistical training, demonstrating that all considered models are of interest for studying the surges in the region.

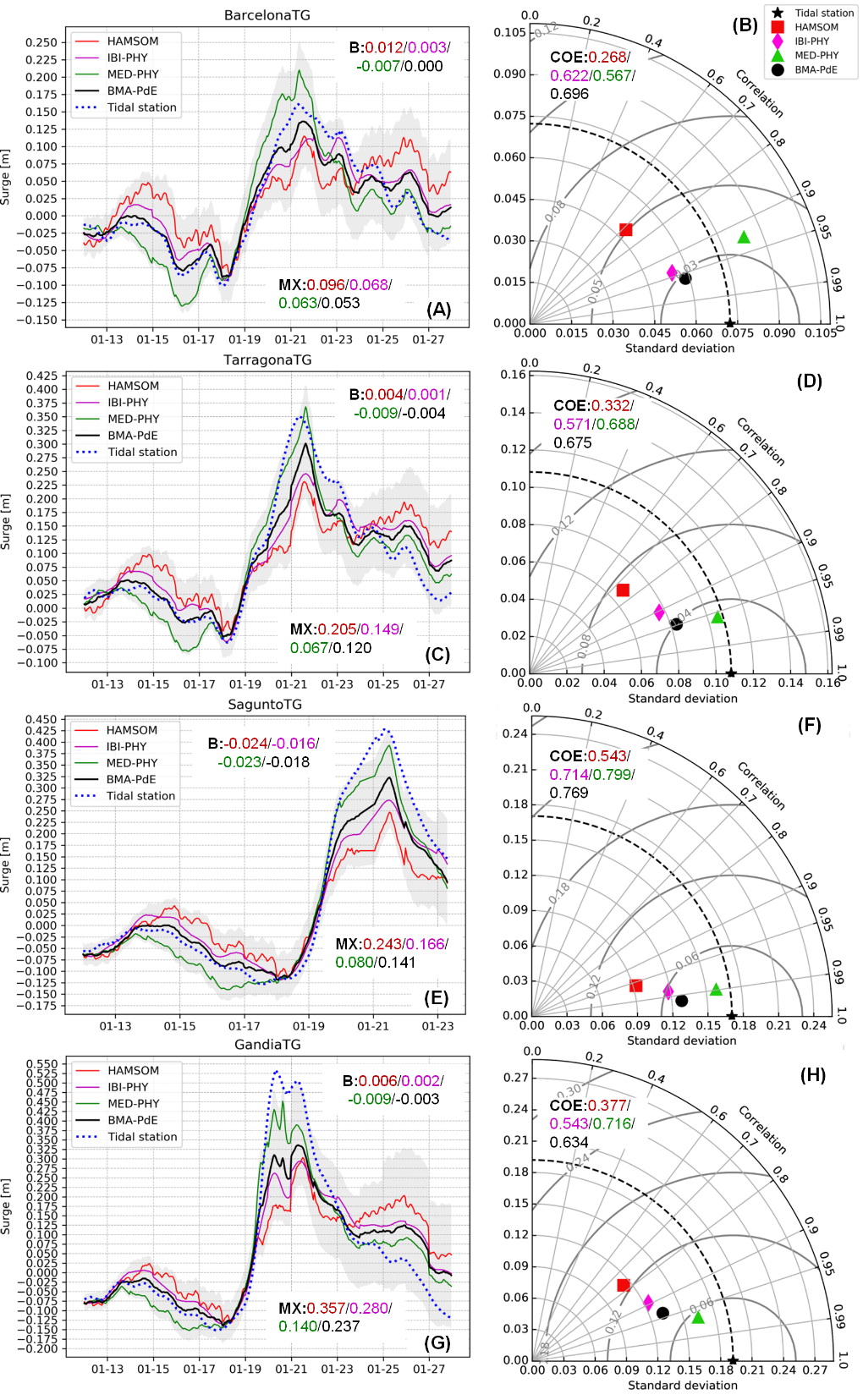
ENSURF produced accurate results during the Gloria event, although underestimating the main peak in the most affected areas (specially Sagunto, Gandía and Almería). In regions where the peak was not so extreme, like Melilla or Málaga, ENSURF outperforms the rest of the models. For example, at Barcelona the BMA shows better statistical values than both IBIMFC and MedMFC. The Correlation of Efficiency - COE (Legates and McCabe, 1999) is 0.696 for the former, vs. 0.622 and 0.567 respectively. It is also important to remark that the surge lies, for all stations, inside the confidence band predicted by ENSURF. The reliability of the uncertainty estimation demonstrates the importance and benefit of the multi-model ensemble method.

For all the stations, the MedMFC is producing the largest surge of all the models. At points with a very large surge, like Almería, Gandía and Sagunto, this produces the more accurate reproduction of the peak of the storm. At Gandía, where the highest surge was recorded, around 54 cm, the underestimation is 14 cm. Nevertheless, in other points with lower values on the peak, like Ibiza, the MedMFC forecast severely overshoots the measurements.

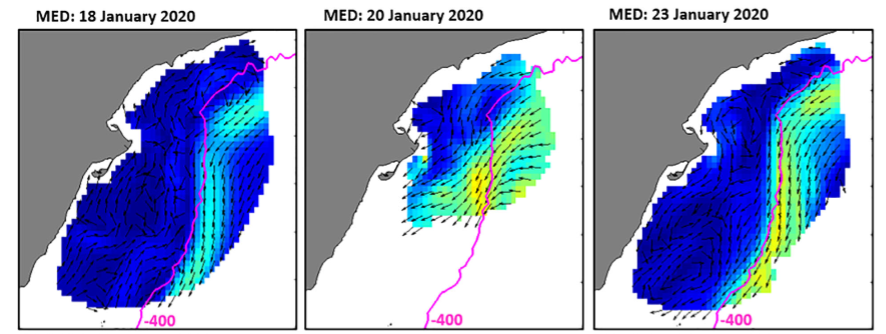
IBIMFC is, for all stations, producing worse statistical results for Gloria than MedMFC and ENSURF. At the peak of the storm, the IBIMFC is producing a much smaller surge than MedMFC. Similarly, Nivmar severely underestimated the main peak of the storm, and the analysis of the Taylor diagram shows that in general, it had a relatively poor performance during this particular event. A detailed analysis of the origin of the differences between the models goes beyond the scope of this section and can be found at (Perez-Gómez et al., 2021).

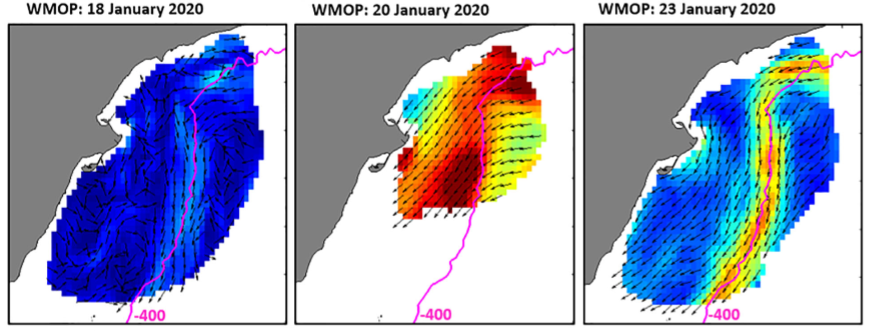
### **Circulation**

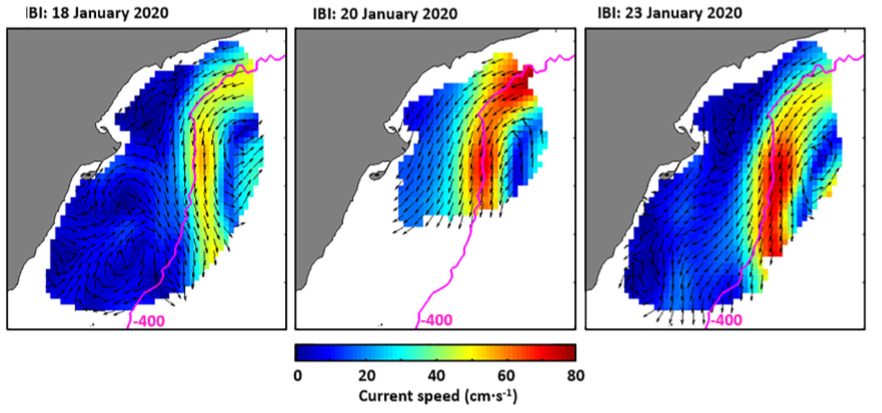
Figure 4.1.10. shows the circulation fields (daily means) as modeled by WMOP, IBIMFC and MedMFC on the domain of the HF radar, that can be directly compared with the observations (figure 4.1.7.). All the models react to the event with an important intensification of the currents, but the differences between then and with observations are significant.



*Figure C.4.1.2. Results of the ENSURF system at several TGs during the storm Gloria: Nivmar (red), IBIMFC (magenta), MedMFC (green), BMA-PdE (black) and observations (blue). Left: surge time series at Barcelona (A), Tarragona (C), Sagunto (E) and Gandía (G). B is the mean bias for each model, in meters (same chromatic index). MX is the maximum error, in meters. The grey areas enclose the 5-95th percentiles of the ENSURF system. Right panels: Taylor diagrams at Barcelona (B), Tarragona (D), Sagunto (F) and Gandía (H). Legend: Nivmar (red square), IBIMFC (magenta rhombe), MedMFC (green triangle), BMA (black dot). COE is the Coefficient of Efficiency for each model (adimensional). CMEMS data product ref-4.1.5 (observations), ref-4.1.1. (MED model), ref-4.1.4. (IBI model) and ref-4.1.7. for ENSURF time series*



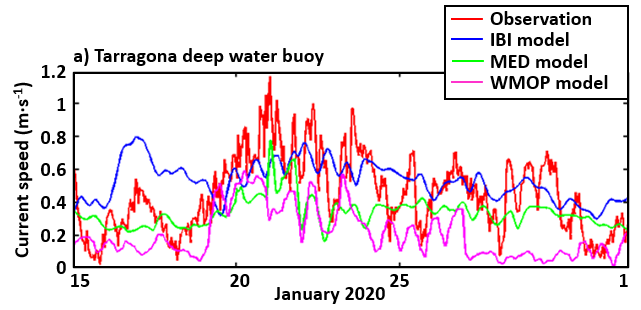




*Figure C.4.1.3. Model output (daily means) over the domain of the HF radars observations. Currents simulated by MedMFC (1stdrow), WMOP (2nd row) and IBIMFC (3rd row), on the 19th (1st column), 20th (2nd column), and 21st (3rd column) of January. CMEMS data product ref-4.1.1. (MedMFC), ref-4.1.4. (IBIMFC) and ref-4.1.6. (WMOP)*

This qualitative comparison against HF radar maps is complemented with skill metrics (RMSE and temporal correlation), spatially averaged over the common domain. The statistical results revealed that the models’ accuracy was higher for the meridional current component, and significantly better during the storm than in the previous same length period (Sotillo et al., 2021). All the models were more capable of dealing with a situation where extreme wind, and not baroclinic mesoscale structures, was the dominant forcing. Correlation index for the meridional component computed in the period before the event range from 0.15 (IBIMFC) to 0.36 (WMOP), improving to reach the interval from 0,39 (IBIMFC) to 0.61 (WMOP). The best agreement with observations during Gloria corresponds to WMOP, outperforming its parent MedMFC system (correlation of 0,61 vs 0,40 and RMSE of 8.43 cm s-1 vs 9.36 cm s-1 for the meridional component). IBIMFC clearly overpredicted the SW slope jet and underestimated the current speed in nearshore areas, the comparison against the HF radar led to slightly lower meridional correlation (0.39) and higher RMSE (13.83 cm s-1).

These relatively poor behavior of all the models in quantitative terms is confirmed by the comparisons of time series at the position of the buoys. Figure 4.1.11. shows the comparison for the position of the B3 buoy. All solutions differ from each other considerably and no one is accurately matching the observed values, even during the maximum storm. All the circulation models show an increase of the current magnitude associated to the arrival of Gloria, but this is masked in IBIMFC due to the existence of large currents associated to the mentioned overprediction of the SW shelf break current. In all cases, all the circulation models failed to properly capture the peak during Gloria and, in general, to reproduce accurately the currents during the whole month at the position of the buoy. As mentioned, this is consistent with the poor correlation indexes described in the validation with the HF radar.



*Figure C.4.1.4. Validation of operational ocean models within situ observations buoys: Timeseries of surface current [speed (in m s−1) and direction (deg) at Tarragona mooring. CMEMS data product ref-4.1.5 (observations), ref-4.1.1. (MED model) and ref-4.1.4. (IBI model).*

### **Conclusions of forecasting services performance**

In general, the forecasting systems performed quite correctly. Nevertheless, the information generated was more accurate for some variables than for others. The extreme values waves (figure C.4.1.1.) and sea level (figure C.4.1.2.) were correctly forecasted, but the circulation models were only able to provide a general description of the response to the storm (see figure C.4.1.3.), failing as usual in forecasting specific values of current for a given point at a specific time (figure C.4.1.4.). This is a well-known problem, mainly related with the difficulties properly collocate mesoscale patterns into the initial 3D-fields of the simulation (Röhrs et al, 2021) and, in this aspect CMEMS, as well as other state of the art circulation forecasting systems, have plenty of room for improvement in the future.

The advantages of storm surge multi-model ensemble forecast system were demonstrated. ENSURF produced accurate results during the Gloria event, although underestimating the main peak in the most affected areas (specially Sagunto, Gandía and Almería). It is also important to remark that the surge lies, for all stations, inside the confidence band predicted by ENSURF. The reliability of the uncertainty estimation demonstrates the importance and benefit of the multi-model ensemble method.