



Performance of different forecast systems in an exceptional storm in the Western Mediterranean Sea

L. Bertotti,^a J.-R. Bidlot,^b C. Bunney,^{c†} L. Cavaleri,^{a*} L. Delli Passeri,^d M. Gomez,^e J.-M. Lefèvre,^f T. Paccagnella,^g L. Torrisi,^h A. Valentini^g and A. Vocino^h

^aInstitute of Marine Sciences, CNR, Venice, Italy

^bEuropean Centre for Medium-Range Weather Forecasts, Reading, UK

^cMet Office, Exeter, UK

^dDepartment for National Civil Protection, Rome, Italy

^ePuertos del Estado, Madrid, Spain

^fMétéo-France, Toulouse, France

^gARPA-SIMC, Hydro-Meteo-Climate Service of Emilia-Romagna, Bologna, Italy

^hCNMCA, National Meteorological Service, Rome, Italy

*Correspondence to: L. Cavaleri, ISMAR, Castello 1364/A, Venice 30122, Italy. E-mail: luigi.cavaleri@ismar.cnr.it

†The contribution of these authors was written in the course of their employment at the Met Office, UK, and is published with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

We consider an exceptional storm – ‘Klaus’ (January 2009) – its evolution on the Western Mediterranean Sea, and how the associated wind and wave conditions were modelled by seven of the major systems presently operational in this area. We intercompare the model results and then verify them and the related model ensemble *versus* the available measured data.

Working with short-term forecasts (24 h) only, as expected, each model correctly anticipates the incoming of an exceptional storm. However, even at such limited range, we have found substantial differences among the results of the different models. The differences concern the time the storm should have entered the Western Mediterranean Sea, the peak values of wind speed and significant wave height, the general distribution of the fields, and the locations where the maxima were achieved.

We have compared the model results *versus* the available measured data, wind from scatterometer, waves from altimeter, plus a few buoy data. We have found some inconsistencies in the results, model wind data being on average larger than the measured one, while the opposite was true for wave heights. However, the limited amount of data available and its different times and positions, at and off the centre of the storm, impede the drawing of any definite conclusion in this respect.

On the whole we feel that our results, although related to a single storm, cast doubts on the reliability of a single forecast system to provide sufficiently reliable and accurate forecasts in case of an incoming exceptional storm. The results, both for wind and waves, have improved using an ensemble of the seven considered models. This suggests that there is no relevant systematic error in the used models except, as possibly suggested by our results, in the case of wave generation under very strong wind and very young sea conditions. Copyright © 2011 Royal Meteorological Society and British Crown Copyright, the Met Office

Key Words: forecast; wind; wind waves; extreme conditions; meteorological models; wave models

Received 1 October 2010; Revised 27 June 2011; Accepted 30 June 2011; Published online in Wiley Online Library 12 August 2011

Citation: Bertotti L, Bidlot J-R, Bunney C, Cavaleri L, Delli Passeri L, Gomez M, Lefèvre J-M, Paccagnella T, Torrisi L, Valentini A, Vocino A. 2011. Performance of different forecast systems in an exceptional storm in the Western Mediterranean Sea. *Q. J. R. Meteorol. Soc.* **138**: 34–55. DOI:10.1002/qj.892

1. Introduction

In January 2009 an exceptional storm hit the northern part of Spain and the southern part of France. Moving eastwards, the storm entered the Western Mediterranean Sea, leading to what was estimated as the most violent storm in these areas of the last 10–20 years (Cavaleri and Sclavo, 2006). The purpose of this paper is to check how different forecast systems performed in modelling the storm, with a focus on the surface wind and waves.

An uncounted number of limited area meteorological models are operational in the Mediterranean Sea, typically coupled to, or followed by, a corresponding wave model. For some of them performance statistics are available via the JCOMM project (Bidlot *et al.*, 2002), pioneered and led by the European Centre for Medium-Range Weather Forecasts (Reading, UK, henceforth ECMWF). These statistics provide an indication of the performance of the different (participating) wave systems, both as analysis and forecast.

Our purpose is different. As a case study, we want to verify the performance of the different models when dealing with an exceptional storm. Experience shows that, while all the models are expected to perform reasonably well in the more common situations, the extreme cases often test the models to their limits, simply because some of the physical assumptions they are built on or some empirical relationships may no longer be strictly valid (WISE Group, 2007).

Being based on the analysis of the model results of a single storm, it is not our aim to evaluate the long-term performance of the systems. Rather, this can be considered as an evaluation of these forecast systems under real and extreme conditions – a sort of ‘acidity test’. Our main interest has been to assess how reliable limited area modelling can be when using a single forecast for subsequent preventive actions in the case of a violent storm.

The article is structured as follows. After describing in section 2 the area of interest, we give in section 3 a description of the storm. The areas covered by the different models are briefly described in section 4, while a compact description of the various model configurations is provided in Appendix A. A short description and analysis of the outputs from the different models are discussed in section 5, while in section 6 we compare the model results *versus* the available measured data. In section 7 we explore the performance of the multi-model ensemble. Finally, an overall discussion and conclusions are presented in section 8.

2. The area of interest

The Mediterranean Sea is the longest (almost) enclosed basin in the world. It spans nearly 3600 km in the longitudinal direction (from 6°W to 36°E) and more than 1600 km in latitude (from 30° to 45.5°N). However, these large dimensions are fractioned by the very complicated geometry of the coastline, with elongated peninsulas, namely Italy and Greece, protruding from its northern coast and sectioning the Mediterranean into a number of sub-basins. Large islands (Sicily, Sardinia, Corsica, the Balearic Islands, Crete

and Cyprus – not to mention the clusters in the Aegean Sea) add to fractioning. All this is made evident by the large number of names identifying specific areas, each one with its logical reason and meaning. The meteorological modelling is further complicated by the continuous mountain ranges that characterize the northern border, starting from the Sierra Nevada in Spain and ending with the Taurus mountains in Turkey.

The focus of this paper is on the area shown in Figure 2, extending from 3° to 21°E and from 35° to slightly more than 45°N. On the upper left is the Gulf of Lion where the northwest wind blowing down the Carcassonne pass (between the Pyrenees and the Massif Central) reaches its highest speed. We refer to this wind as the mistral (in France it is called ‘tramontane’, but we stick here to the more common English name ‘tramontana’). The area with typically the most intense wave activity (see, for example, Cavaleri and Sclavo, 2006) is enclosed between the Balearic Islands on the left and Corsica and Sardinia on the right. Once the storm waves pass below Sicily and turn to the left, towards the east, they carry on as swell towards the eastern part of the basin (not in the figure), reaching the coasts of Egypt and Israel almost 3 days later.

3. The storm of 24–25 January 2009

The focus of this paper is the Mediterranean marine part of the storm. ‘Klaus’, as the Free University of Berlin named the storm, developed between 23 and 25 January 2009. It was characterized by a very strong depression that moved in from the Atlantic Ocean into southern France and northern Spain, causing fatalities and havoc in both these countries before entering the Western Mediterranean, where it led to one of the most violent storms of the last two decades. A thorough description of the genesis of the storm and its evolution on the Atlantic Ocean, until its landing on the French–Spanish coasts, is provided by Liberato *et al.* (2011), which we have summarised in the following description.

‘Klaus’ started as a small wave perturbation on 21 January, then moved swiftly eastwards during the following days. The interaction with the uncommonly low-latitude upper-level jet led to an explosive cyclogenesis that reached its maturing stage deepening rate on 23 January in the Bay of Biscay (37 hPa/24 h). Northern Spain and southern France were hit with wind gusts up to 198 km/h measured at low-elevation stations. The strongest winds were concentrated around the Pyrenees. Record wave heights were registered within Spanish seas, with significant wave heights up to 14.88 m and single wave heights up to 26.13 m. Heavy damage was widespread. The toll in human lives was 26.

Given the purpose of this work, our attention is focused on the synoptic situation at the surface. A sequence of maps at 12 h intervals is shown in Figure 1. Starting from the genesis on the Atlantic, the maps show the evolution of the storm in the Western Mediterranean Sea. In terms of mean sea-level pressure (MSLP), at 0600 UTC on 24 January, when storm ‘Klaus’ landed over the western coast of France, a secondary low developed over the sea northeast of Perpignan and southwest of Marseille. This system was associated with strong southwesterly winds over the Balearic Sea. During the

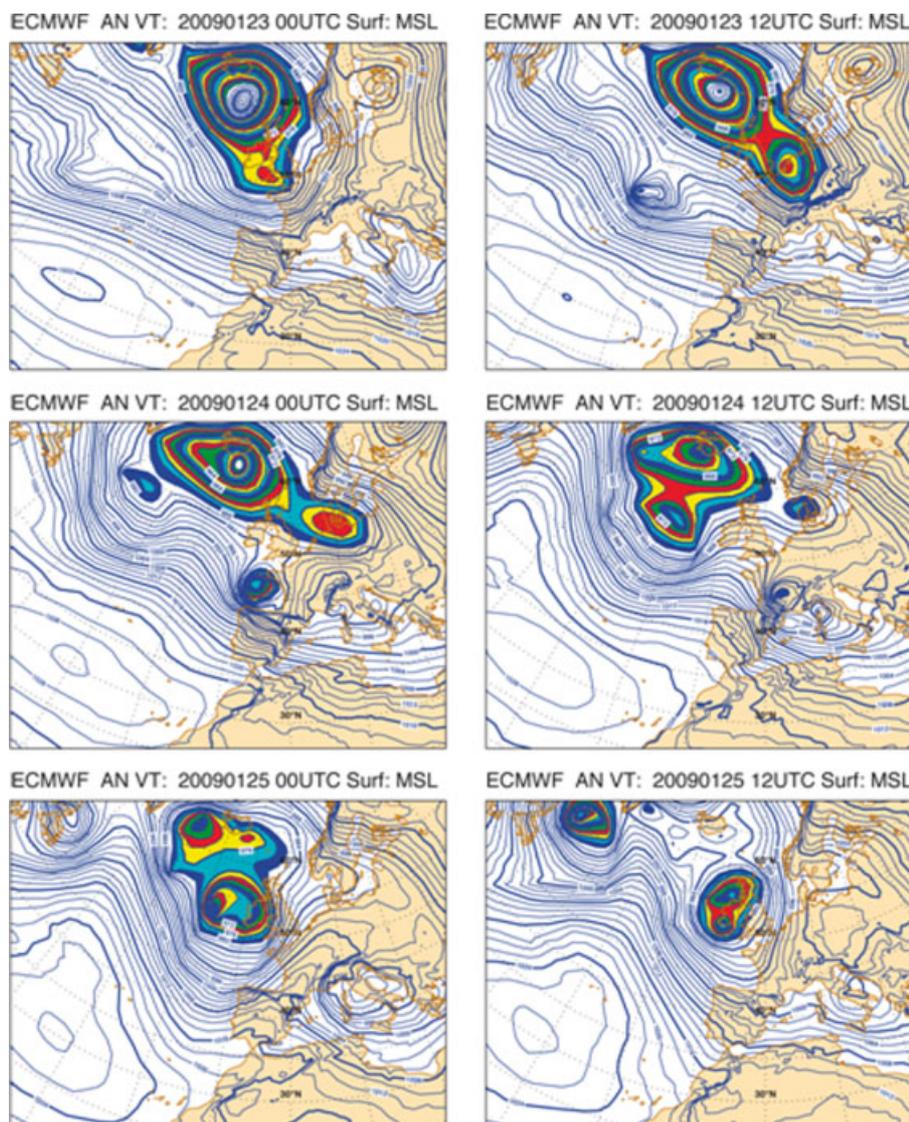


Figure 1. Sequence of surface meteorological maps, at 12-hourly intervals, showing the development of the 'Klaus' storm. Isobars at 2 hPa intervals. ECMWF analysis. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

following 6 h, while 'Klaus' was crossing France north of the Pyrenees, this secondary low further deepened and travelled towards the Ligurian Sea. At 1200 UTC on 24 January it was located between Corsica and mainland Europe. The surface wind veered clockwise, and at 1200 UTC a strong northwesterly flow started affecting the northern part of the Mediterranean Sea, northwest of Corsica. The secondary low continued to travel across the Ligurian Sea and reached the western coast of Tuscany, central Italy, at 1800 UTC. During this time the surface wind further intensified, with a northwesterly flow affecting the northern part of the Mediterranean Sea, west of Corsica, and a northerly flow affecting the Ligurian Sea, north and east of Corsica. This extreme, very strong wind continued to affect the northern part of the Mediterranean Seas during the night of 24–25 January until 0600 UTC on the 25th, when the low-pressure system moved over the central part of the Adriatic Sea, causing strong easterly wind conditions, known as the 'Bora', over its northern part.

From the marine point of view the relevant meteorological information is the surface (10 m height) wind speed U and the consequent waves, represented by the significant wave height H_s . Following the respective international

conventions, we identify winds and waves with their incoming (nautical convention) and flowing (oceanographic convention) directions respectively. Their development during the 2 days across the peak is shown in Figures 2 and 3, representing the ECMWF analysis (25 km resolution for surface wind, 10 km for waves). The panels, from (a) to (f), are at 6 h intervals, starting at 0600 UTC on 24 January. At this time the west/southwest wind has already been acting for a few hours (Figure 2(a) and (b)) leading to substantial waves towards the east (Figure 3(a) and (b)), particularly in the more southerly part of the area. Between 1200 and 1800 UTC the mistral starts blowing in the Gulf of Lion (Figure 2(c)) and, with lower wind speeds, in the area between Sardinia and the Balearic Islands (henceforth Balearic). This leads to cross-sea conditions (Figure 3(c)) with a mean easterly direction.

The peak of the storm is between 1800 UTC on 24 January and 0000 UTC on the 25th. The mistral is now dominating the area (Figure 2(d)), and in so doing is forcing the waves in the same direction (Figure 3(d)). More to the south, towards the African coast, following the general meteorological pattern and the pre-existing situation, the waves have a more longitudinal component. While at this stage the sea

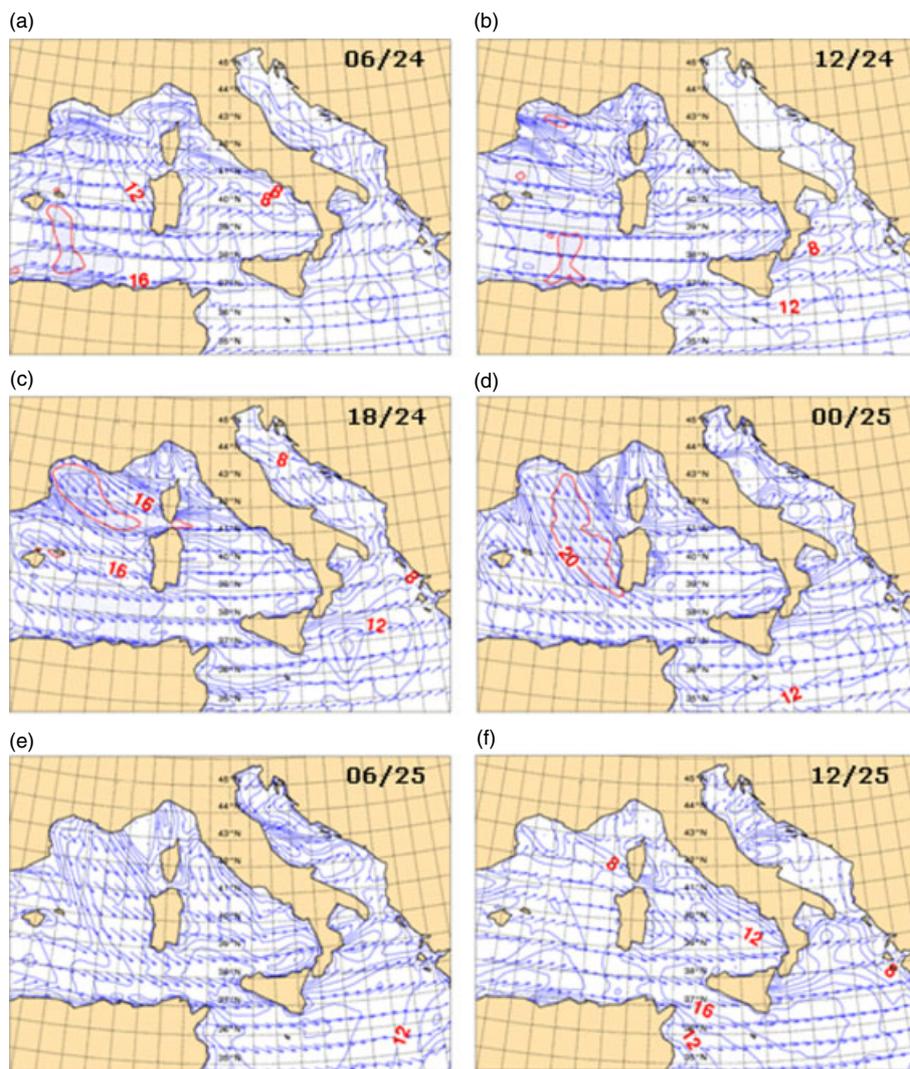


Figure 2. Surface wind conditions in the Western Mediterranean Sea. Panels (a)–(f) are at 6-hourly intervals, starting at 0600 UTC on 24 January 2009. Arrows show wind speed and direction. The panels show ECMWF analysis fields. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

is obviously characterized by cross-sea conditions, the rapid change of direction between Balearic and Sardinia was due to the similar wave periods along the two directions, to east and southeast, leading to strong wave–wave interactions and to a consequently rapid shift to the local wave direction. At 0600 UTC on 25 January the worst of the storm is over, although 7+ m significant height waves are still pounding the west coast of Sardinia (Figure 3(e)). The highest wind zone is approaching the area between this island and Tunisia. In the following hours, while progressively decaying, the storm hits Sicily (Figures 2(e) and 3(e)). During this phase note the shadow zones behind Tunisia and Sicily. Finally, the storm propagates into the Ionian Sea (Figures 2(f) and 3(f); 1200 UTC on 25 January), then carrying on as swell towards the east.

This is the typical development of a mistral storm, albeit of greater intensity than usual, and is a rather common event in the western part of the Mediterranean Sea. Therefore, at least in principle, it would be natural to expect that all the operational high-resolution meteorological models acting on the area would be able to provide a fair description of the storm. Partly this may be affected by the specific area covered by their grids and set-ups. Therefore, before

discussing their performance in the present case of interest, it is convenient to show the area they cover and to give a compact description of their set-ups. This is the subject of the next section.

4. The considered models

Figure 4 provides a general view of the Mediterranean Sea and of the borders of the high-resolution grids used by the seven considered models. The corresponding institutes are ECMWF (UK), Météo-France (France), CNMCA-ISMAR and ARPA-SIMC/DPCN (Italy), Puertos del Estado (Spain), SHOM (France) and UK Meteorological Office (UK). See Appendix A for a more detailed description. Throughout our discussion it will be practical to refer to the different models by a letter A–G (see Table I and Figure 4). This more anonymous addressing will also make it easier to comment in a more aseptic way ‘for good or bad’ about their relative performance.

The high-resolution meteorological and wave grids of A, C, F and G cover the whole Mediterranean basin. E is limited to its western part. The same holds for B, with only a slightly more eastern border. The more limited one is D (dotted

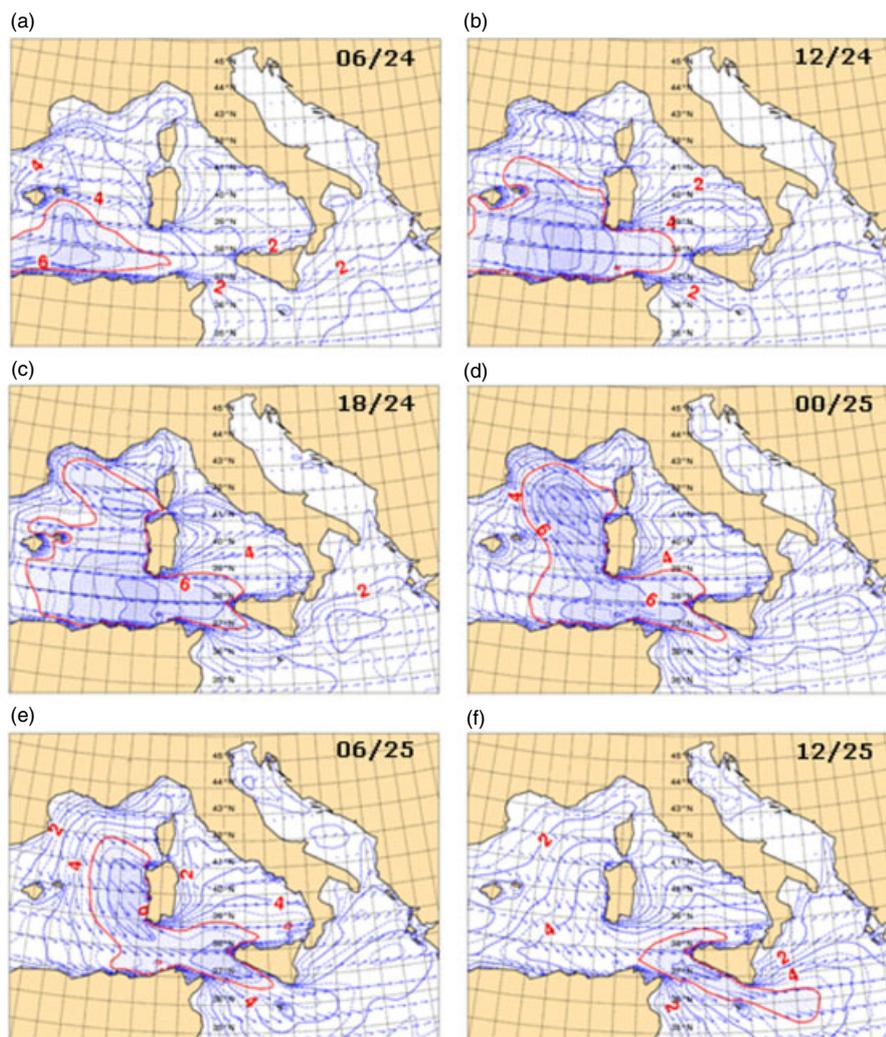


Figure 3. Wave conditions in the Western Mediterranean Sea. Panels (a)–(f) are at 6-hourly intervals, starting at 0600 UTC on 24 January 2009. Arrows show significant wave height and direction. The panels show ECMWF analysis fields. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

Table I. Letters used to indicate the various institutions.

	Institution
A	ECMWF
B	Météo-France
C	CNMCA-ISMAR
D	ARPA-SIMC/DPCN
E	Puertos del Estado
F	SHOM
G	UK Meteorological Office
ENS	Multi-model ensemble

ENS refers to model ensemble.

border), which is, however, as detailed shortly, nested in a larger but coarser grid.

Aiming at analysing the performance of seven models during and in the more intense part of the storm, we have focused our attention and analysis on the area between Balearic and Sardinia, marked with a grey grid in Figure 4. As this exceeds the area covered by D, for this model we have complemented its high-resolution results with those from its coarser grid (up to the continuous line in Figure 4). A

direct comparison in the overlapping area has shown for this storm no practical difference for both wind and wave model results, particularly in the area where the highest wind and wave conditions were found.

A short description of the set-up of each one of the seven considered models is given in Appendix A.

Before proceeding further it is worthwhile stressing that it is not our purpose to reach a judgement on the average performance of the considered models. Clearly the results for a single storm, however exceptional, cannot provide this information. This is particularly the case for the SHOM results that were obtained with a similarly qualified, but not standard, set-up. Rather, we are interested in the consistency and accuracy among/of the various models in case of an exceptional storm. The ‘Klaus’ storm offered such an opportunity. In this respect it is convenient to provide, where available, some information on the long-term performance of the considered models. This is given in Appendix B.

5. The storm according to the different models

In this paper we do not analyse the performance of the models at different forecast ranges. Rather, we focus on their

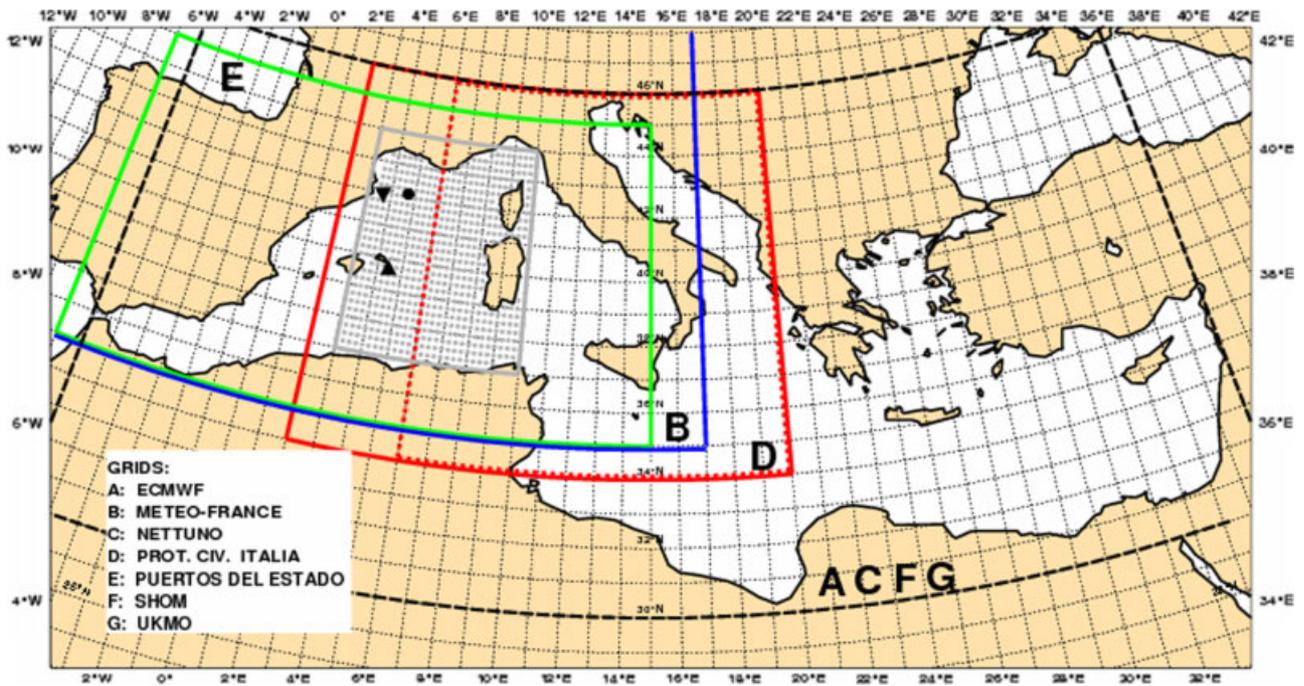


Figure 4. High-resolution grids of the seven considered forecast systems. See the related institutions in the lower-left corner and in Table I. The grey-dotted area shows where most of the wind and wave analysis has been focused. The marks show the position of the wave-measuring buoys 61002 in the Gulf of Lion (○), Begur (▽) and Mahon (△). Begur and Mahon also provide information about wind. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

assumed best performance, i.e. their first 24 h forecast. While this may not be the best short-term forecast available from some of the systems (a number of them issue new forecasts more than once a day), it allows a more fair intercomparison of their performance. The sequence of wind and wave fields we have considered correspond to the +3 to +24 h forecast (or +1 to +24 h if the fields are available at 1 h intervals) issued at 0000 UTC of 23, 24 and 25 January 2009, from which an uninterrupted 72 h time series has been derived for each model. The ECMWF winds and waves considered for the present analysis correspond to their 24 h forecast and are therefore different from the ones shown in Figures 2 and 3 (analysis).

We begin our analysis considering the wind situation at 1800 UTC of 24 January as reproduced by the different model systems. The seven different maps are shown in Figure 5. Clearly, all the models show the mistral storm, but substantial differences are present. B, C, D and G have similar geographical distributions, well extended beyond the Gulf of Lion, with large areas with wind speeds greater than 24 m s^{-1} . Also the ranges of wind speeds are similar. Model A shows a smaller area of intensive speeds. The contrary is true for E and F, particularly the former, which exhibits wind speeds close to 32 m s^{-1} .

The corresponding wave conditions 6 h later, at 0000 UTC on 25 January, are shown in Figure 6. Given the sensitivity of waves to also small differences in the driving wind fields, the differences, as expected, are larger than in Figure 5. Similar fields are suggested by D and F, while A, E and G indicate lighter conditions. While all these models direct the storm towards the more southern part of Sardinia, B puts the peak more to the north, between Corsica and Sardinia. The highest wave conditions are suggested by C, with a peak significant wave height above 11 m. We will discuss panel E_{21} in Figure 6 in a later section.

Apart from these differences, what is surprising is that the above groupings, i.e. similar wave fields, are not necessarily associated with the similar wind fields listed from Figure 5. This brings us to the matter of how the fields evolve in time.

Given the geometry of the basin and the wind direction for the considered area west of Sardinia where the storm reached its peak, the reaction time of the basin is of the order of 12–14 h. Therefore the wave conditions shown in Figure 6 derive from the winds present in the basin in the hours preceding the fields in Figure 5. Besides, given the scale, it is not easy to derive from the two figures the actual maxima displayed by the single models. This limitation is partly overcome by Figure 7, which provides, both for wind speed and wave height, the time series of the field maxima in the area of interest (see Figure 4) according to the different models. Starting from 0000 UTC of 24 January, i.e. 18–24 h before the peak conditions, Figure 7(a) displays in the first 24 h substantial differences among the single models, the range of values being of the order of 40% of their average. Note in particular the wind peak in C at 0900 UTC on 24 January. The models' wind speed range is smaller at the peak of the storm, 1800 UTC, but still about 15% of the overall model mean at this time. The symbols at the lowest central part of the figure indicate the time the highest wind speed was reached for each model. Most models peak at 1800 UTC on 24 January, including C when neglecting, as we have done in Figure 7, its isolated peak 9 h earlier. The exceptions are D and B, which place their peak, quite consistent with their time series, respectively 2 and 6 h later.

The corresponding analysis of the progressive wave height peak values in Figure 7(b) is very enlightening in showing the effects of the integrated wind field input to waves. We notice at once the double peak of C, as just seen for the wind. However, here the second peak is higher, pointing to a different fetch length and hence position. A similar

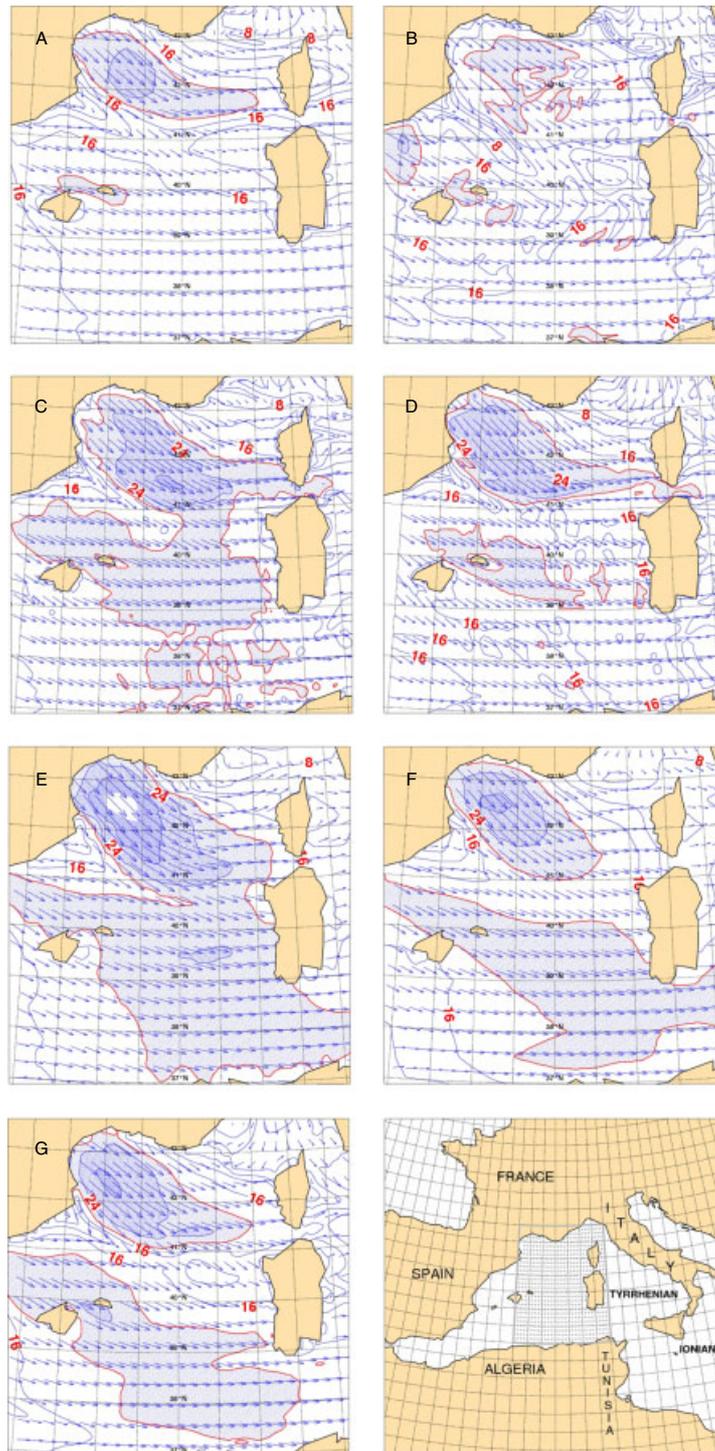


Figure 5. Wind fields at 1800 UTC on 24 January 2009 according to the different sources. Isotachs are at 4 m s^{-1} intervals. The grey-dotted area in the lowest right panel shows where most of the wind and wave analysis has been focused. Arrows and shading show wind speed and direction. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

oscillation in time of the H_s value is shown by D. The times of the maxima are all concentrated between 2100 UTC on 24 January and 0300 UTC on the 25th. However, on the whole, apart from the C peak in Figure 7(a), the wave maxima are more scattered in time with respect to the wind maxima. Considering the 12–14 h reaction time of the basin, such a scatter suggests some substantial differences in the driving wind fields, not only in strength but also in time. Of course, we note the outstanding peak of E (see also Figure 6), almost 2 m higher than any other model, while

no such difference is present in Figure 7(a). The different geometry of the driving wind fields, in space and time, are also reflected in the wide range of H_s model values throughout the storm. Even neglecting the isolated peak by E, the range varies in time between 20% and 40%, implying a 40–100% range in wave energy.

In this respect it is instructive to look at the position where each model reaches its peak conditions. For wind and waves this is shown in Figure 8(a) and (b) respectively. In Figure 8(a) we see the wide distribution of the peak

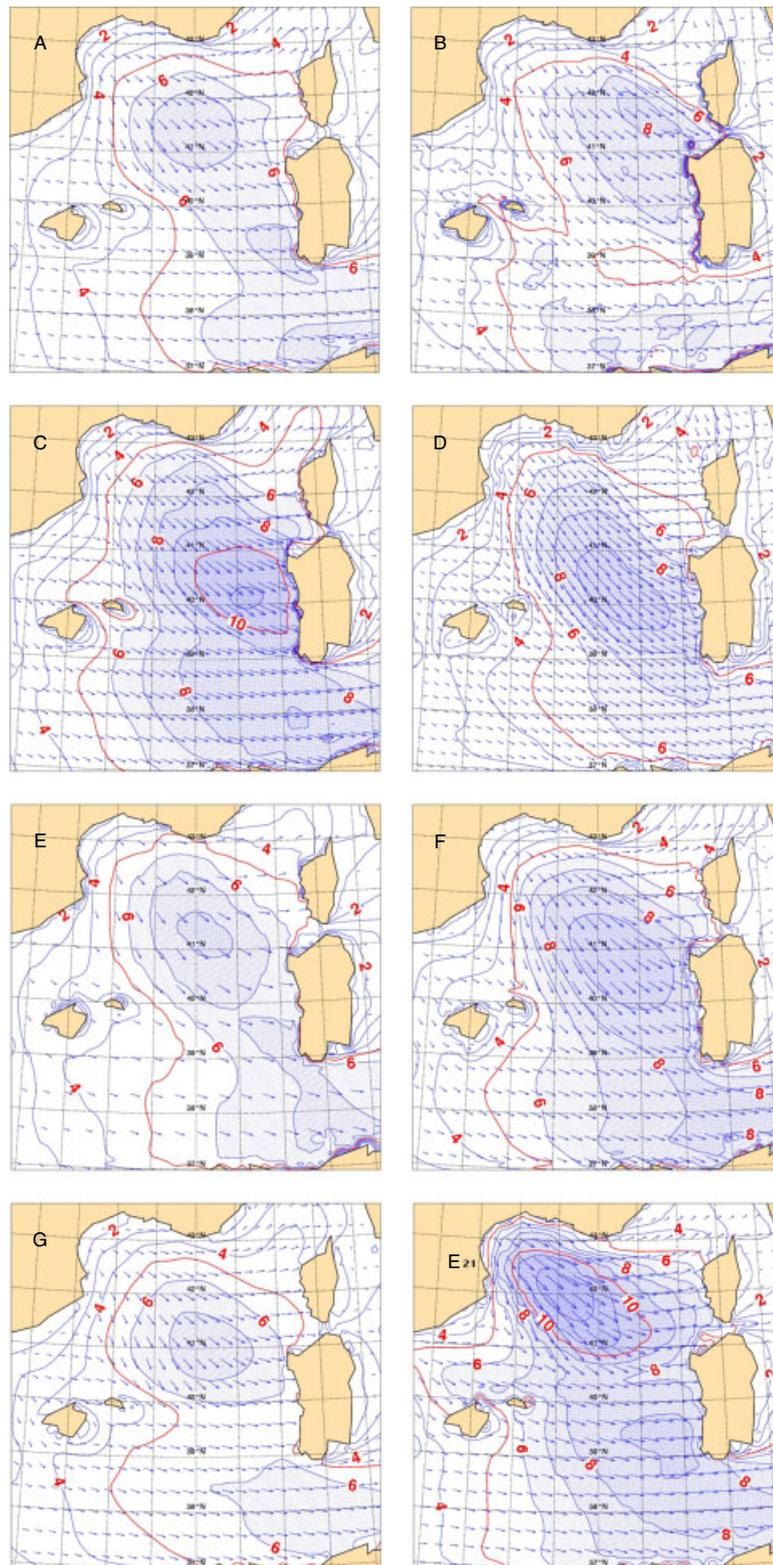


Figure 6. Significant wave height fields at 0000 UTC on 25 January 2009 according to the different sources. Isolines are at 1 m intervals. The lowest right panel shows the E wave fields 3 h earlier. Arrows and shading show wave height and direction. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

positions, spanning about 200 km along the main axis of the Gulf of Lion. The exception is B, which places its peak much further to the east. This is consistent with its more northerly direction noted in Figure 6(b).

The positions of the wave peak values are similarly widely distributed, spanning almost 300 km. Note that the fetch length where the different models achieve their maximum

varies from 200 to almost 550 km. Also, moving from the coast towards offshore, the different sequences, for wind and waves, with which we encounter the peak values of the single models point to substantially different distributions of the wind speed also in the previous hours.

Having commented on the different developments of the considered models, it is now time to explore, within the range

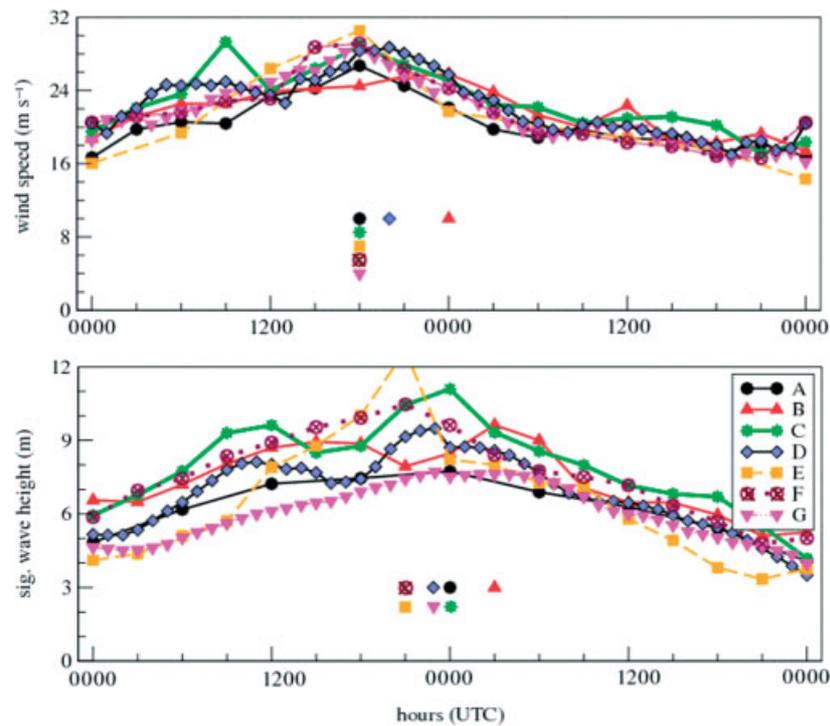


Figure 7. Starting at 0000 UTC on 24 January 2009, plots of the maximum wind speed (upper panel) and significant wave height (lower panel) at different times along the development of the storm according to the different forecast systems. Ticks at 3 h intervals. The symbols in the lower part of each panel show when each system achieves its overall maximum value.

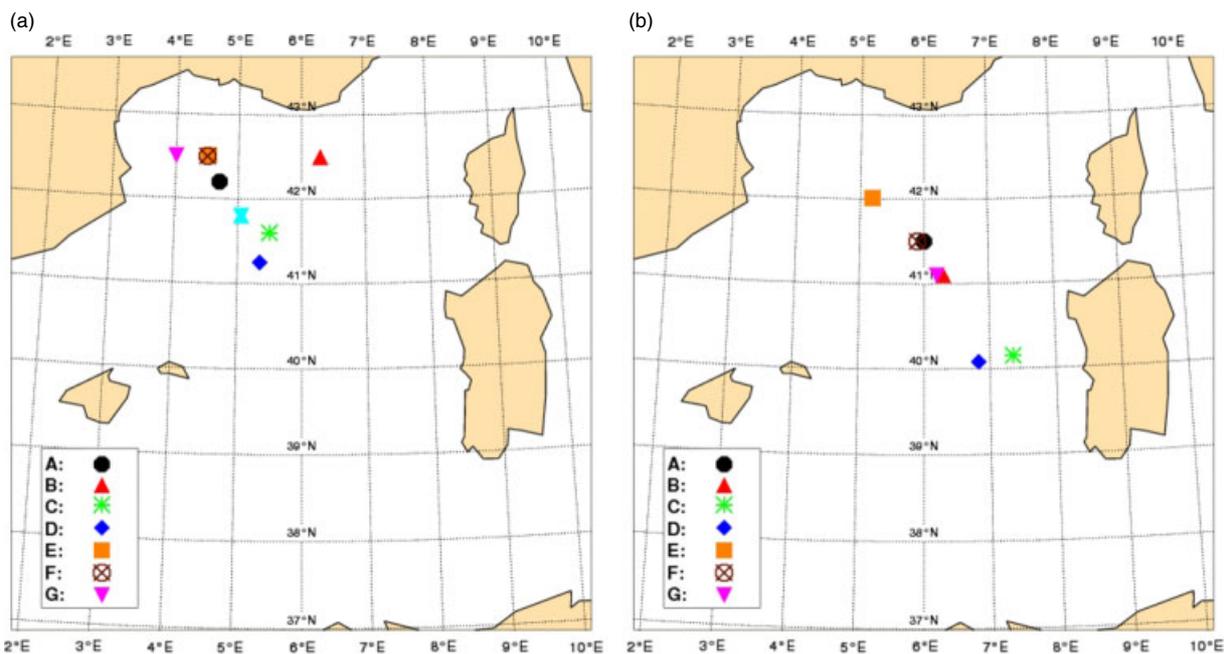


Figure 8. Locations where each model system achieves its maximum overall value (see Figure 7): (a) wind speed; (b) significant wave height. The extra symbol in panel (a) refers to the maximum of the scatterometer data (see Figure 9). This figure is available in colour online at wileyonlinelibrary.com/journal/qj

of the available data, how models fit the measurements. This is the subject of the next section.

6. Comparison with the available measured data

Wind and wave observation data were provided by a combination of satellite and buoy measurements. For wind we have two valid passes by the ASCAT and ERS-2 scatterometers over the area and within the period of

interest, respectively at 2100 and 2200 UTC on 24 January. Within the available altimeters, only Jason made a mildly interesting pass, shown in Figure 11, almost 24 h after the peak. Unfortunately, the most conveniently placed wave buoy, off the northwest corner of Sardinia, was out of service. We have interesting data from the French buoy 61 002 moored in the Gulf of Lion (42°06.18'N, 4°42.18'E), and from two Spanish buoys: Begur (41°54.9'N, 3°38.7'E) and Mahon (39°40.8'N, 4°25.2'E), moored respectively close

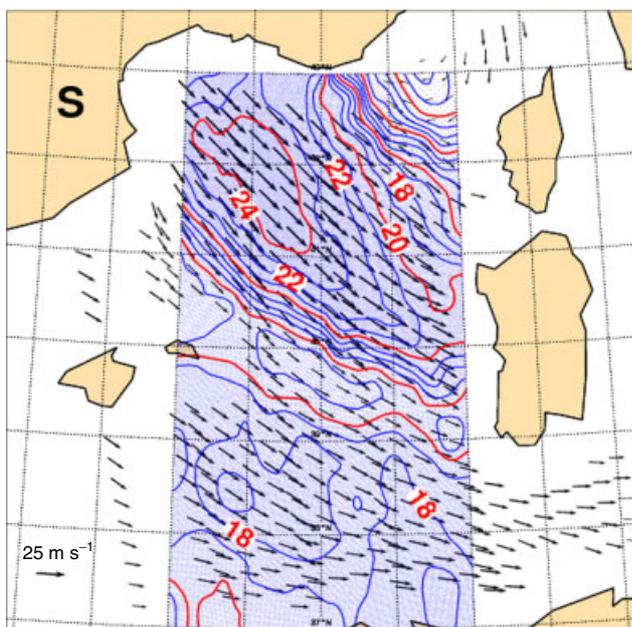


Figure 9. Surface wind field obtained by the ASCAT scatterometer during its pass of 2100 UTC on 24 January 2009. Arrows show wind speed and direction. Isotachs at 1 m s^{-1} intervals have been traced in the shadowed area. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

to the Gulf of Lion and 16 km southeast of Menorca, both of which also provide wind data (see their position in Figure 4).

In Figure 9 we have plotted the isotachs built on the ASCAT wind field. Although the coastal area in the Gulf of Lion is not covered by the data, the structure of the field (at 2100 UTC on 24 January) is well defined. The area of highest wind speeds, more than 24 m s^{-1} , is between 100 and 200 km off the coast of the gulf, with strong transversal gradients on the borders and protruding, with progressively decreasing values, towards the lower end of Sardinia. In the

more southern part the gradients are lower, with the field slowly turning left to enter the Tyrrhenian Sea.

Although there is a 3 h difference, a first comparison can be done with the fields in Figure 5. Here we see that C, D, E and F have, with minor approximations, the correct direction, while A, B and G show a more easterly component. Considering the maxima in Figure 7(a), the more consistent model is A, while all the other models are, at different levels, in excess. Granted that most models place the maximum U at 1800 UTC on 24 January – 3 h earlier than the pass – all the models, with the exception of B, place the maximum consistently with the scatterometer data.

Apart from this direct, but qualitative, comparison between the fields in Figure 5 and the ASCAT one in Figure 9, it is worthwhile quantifying the fit of the models *versus* the scatterometer data with scatter diagrams. These are shown in Figure 10, with the numerical results given in Table II. The comparison has been done, where necessary, with a space and time linear interpolation of the model data at the ASCAT and ERS-2 times and positions, with the results from the two passes summarized for each model in a single diagram. Consistent with the qualitative comparison, most models – five out of seven – show an overestimate of the wind speeds, this varying between 6% and 17%. The best results are provided by A, with a best-fit slope differing from unity by only 3%. Note the logarithmic scale for the number of cases in each shaded pixel. The large clouds of data around the best-fit lines are mostly single cases, so that, with the exception of E, the scatter indices SI (defined as the RMS error divided by the mean measured value) are all lower than, or equal to, 0.16 and as low as 0.11 for C.

To have a more ample view of the performance of the meteorological models, not strictly focused on the peak area and time of the storm, we have considered also all the ASCAT data available in the Mediterranean Sea for 24 and 25 January. Note that early on the 24th (see Figure 2(a)) the storm was already intense with a strong westerly wind,

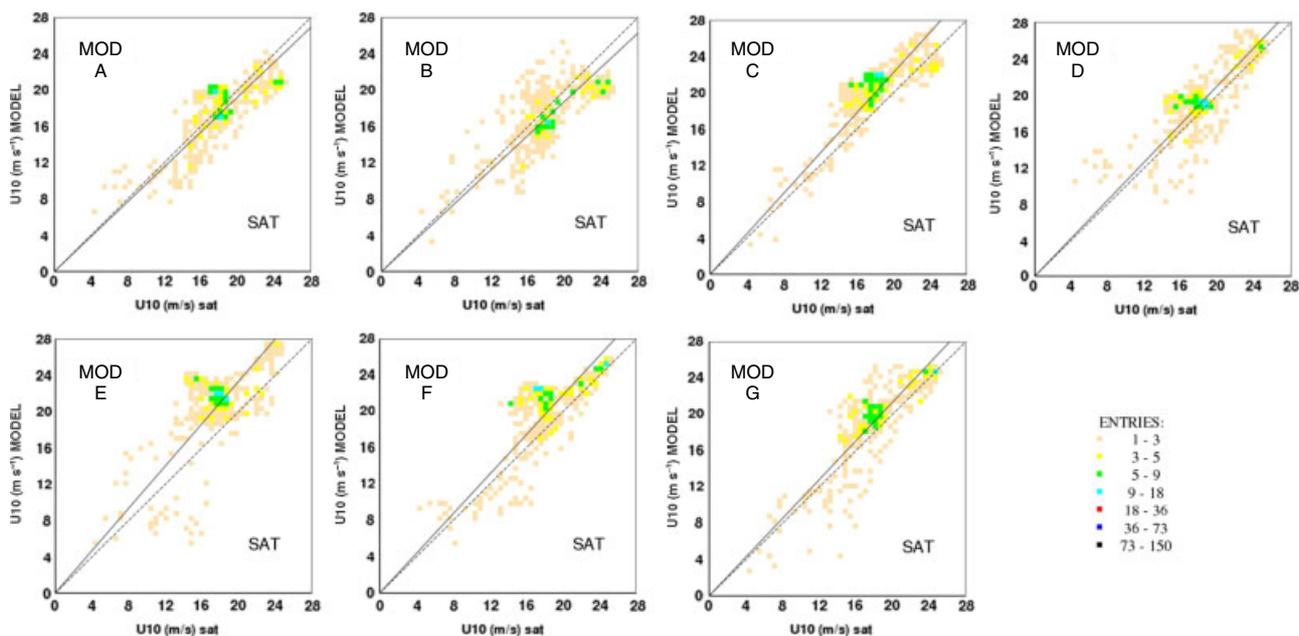


Figure 10. Scatter diagrams for the fits between the seven considered model wind speed fields *versus* the scatterometer data in Figure 9. Shading (see scale on the right) indicates the number of cases in each pixel. In all the diagrams the values go from 0 to 28 m s^{-1} at 4 m s^{-1} intervals. See Table II for the corresponding bias, scatter index and best-fit slope values. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

Table II. Comparison between models wind speeds and scatterometer measured data.

Wind	ASCAT + ERS2 2100–2200 UTC 24 January 2009				ASCAT 24–25 January 2009	
	Mean (m s^{-1})	Bias (m s^{-1})	SI	Slope	SI	Slope
A	18.02	−0.53	0.12	0.96	0.13	0.98
B	18.02	−0.92	0.10	0.94	0.14	0.99
C	18.02	2.23	0.11	1.12	0.18	1.06
D	18.02	1.14	0.13	1.06	0.19	1.04
E	18.02	3.04	0.16	1.16	0.24	1.13
F	18.02	1.38	0.15	1.07	0.15	1.06
G	18.02	0.43	0.07	1.06	0.10	1.02
ENS	18.02	1.07	0.09	1.06	0.10	1.03

Left part (ASCAT + ERS2) refers to the passes at the peak of the storm (see Figure 9). ‘Mean’ is the mean scatterometer value. The right part (ASCAT) refers to all the passes over 2 days. SI is scatter index (RMS error divided by the mean measured value). Slope is the best-fit slope of model *versus* measured data.

while late on the 25th (Figure 2(f)) the still large waves were acting on the Ionian Sea between Italy and Libya. For each model the comparison has been done within the area where the specific model data are available (see Figure 4). The results are on the right side of Table II, again as neutral best-fit slopes and SIs. There are some differences among the performances and hence among the relative scores in the two comparisons, although ‘with a low-pass filter’ the general sequence of values is confirmed. We find it interesting that in the most extensive comparison the SI values are in general appreciably larger than when focused on the peak conditions. One possible reason for this deterioration, at least for some of the models, is that, considering the whole area covered by each model, we implicitly move in most of the cases to its borders, where it merges with the coarser parent model (clearly this is not the case for ECMWF and for the models with a sufficiently large nested grid). If, as expected, especially in the Mediterranean Sea, the parent model is less accurate than the nested model we considered, this may affect a non-negligible band of grid points along the border of the nested model. Another reason may be that the meteorological models, although with some errors, are more up to the point in well-defined conditions – in our case the mistral area. Considering the wider area and a larger time window implies consideration of also much milder conditions, where the models may be less accurate.

We move now to the performance of the wave models. As anticipated, the only significant pass (by Jason) is shown in the upper panel of Figure 11. The pass, at 1638 UTC on 25 January, occurred during the decaying phase of the storm (see Figure 3(f)), when the highest waves were passing, between Sicily and Tunisia, from the Tyrrhenian to the Ionian Sea. Although not at peak time, the waves, at least in the more southern part of the pass, still retain the memory of the previous severe conditions and still provide useful information on the performance of the different models.

As done for the scatterometer, here too the model data have been linearly interpolated in space and time to the various altimeter data times and positions. For an unspecified model the resulting comparison is shown in the larger panel of Figure 11. The horizontal scale (km) starts at the position the satellite first enters the Mediterranean Sea (ascending path). The data are at about 7 km (1 s) intervals. The upper plot concerns wind speed, the lower one wave height, left and right scales respectively. The sharp interruptions correspond to missing data, mostly due to the

presence of land in the surface area covered by the altimeter pulses. The larger ‘black-out’ corresponds to the pass above Italy, before the short crossing of the Adriatic Sea. In the intercomparison with the model data we have neglected this last section and also the short initial one before crossing the small protruding peninsula off Tunisia.

The model wave height distributions along the satellite ground track are shown in Figure 12. For a proper evaluation of these results note that the vertical scale of the plots starts at 3 m. The most evident feature is that all the models display a rather smooth profile of the significant wave height along the pass, as compared to the rather ‘turbulent’ sequence of the altimeter data. No doubt part of this ‘turbulence’ is associated with the random errors in the altimeter data. Abdalla *et al.* (2010) estimate the RMS error of Jason as being on average between 20 and 26 cm in the range of H_s seen in Figure 12. However, after smoothing the altimeter data along the pass, there are still evident differences between the measured and the model significant wave heights. When these differences persist for an extended spatial difference, they are expected to represent a physical truth. Conversely, note that, given the spatial resolution of the models, their smoothness is not necessarily a consequence of the interpolation to the altimeter positions. Models A, B and D fit reasonably well the first part of the pass, say up to 250 km, which is underestimated by the other models. C and F fit well the second part, which is overestimated by the other models. All this is summarized in the statistics detailed in Table III (left part). The best fit is provided by G (best-fit slope = 1.00). Only F has an error larger than 5% (see Figure 12(f)), which is due to its overestimate in the first part of the pass, only partly compensated, as far as the overall fit is concerned, by the good data in the second part (after 300 km). This half-positive result is reflected in its lowest scatter index, 0.09. In any case all the SI values are relatively low, only reaching 0.17.

Using a similar method to that applied to the scatterometer data, we have also analysed how the model results fit the measured data on a more extended time–space scale. For this we have considered all the Jason passes over the Mediterranean during 24 and 25 January. The resulting fits are shown in Figure 13 and summarized on the right-hand side of Table III. We have elected to show these plots rather than those for the pass in Figure 12 because of the limited number of points in the latter case. Note that, owing to the different extension of the grids, a different number

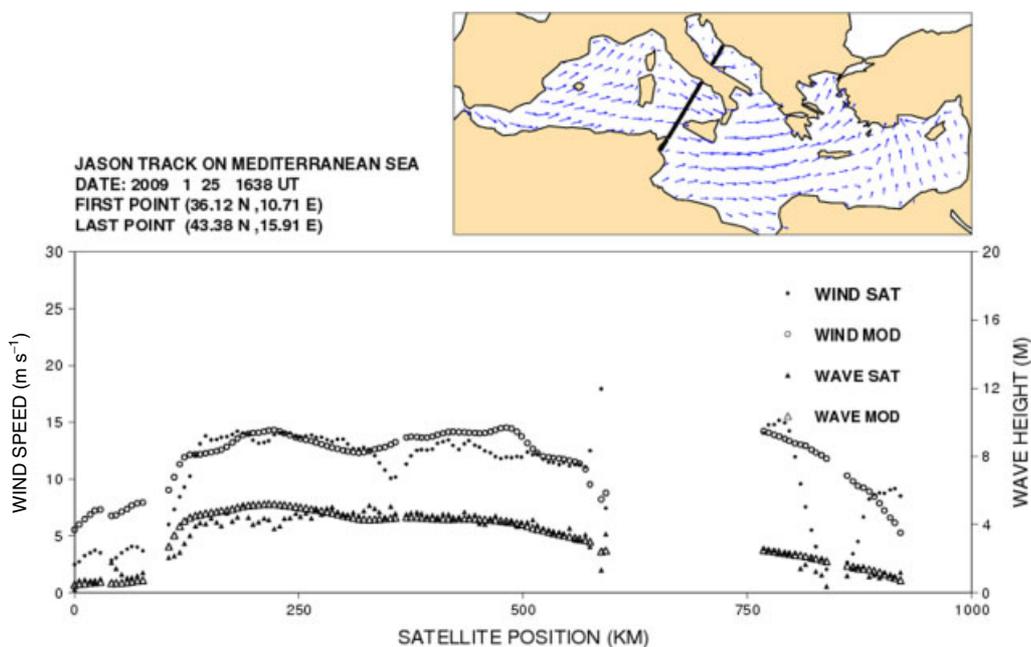


Figure 11. Jason pass at 2009 UTC on 25 January 2009. Upper panel shows ground track. Lower panel shows the comparison of wind and wave model profiles *versus* the altimeter measured data. Horizontal scale in km. Ascending orbit. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

Table III. Comparison between models' significant wave heights and altimeter measured data.

Wave	Jason 1638 UTC 25 January 2009				Jason 24–25 January 2009	
	Mean (m)	Bias (m)	SI	Slope	SI	Slope
A	4.16	−0.20	0.11	0.95	0.19	0.92
B	4.16	−0.20	0.10	0.95	0.25	0.85
C	4.16	0.17	0.13	1.04	0.22	1.00
D	4.16	−0.28	0.10	0.93	0.24	0.90
E	4.16	0.24	0.17	1.05	0.36	0.85
F	4.16	0.37	0.09	1.08	0.24	1.06
G	4.16	0.02	0.12	1.00	0.21	0.96
ENS	4.16	−0.04	0.13	1.99	0.21	0.94

Left part refers to the pass shown in Figure 11. 'Mean' is the mean altimeter value. The right part refers to all the passes over 2 days. SI is scatter index (RMS error divided by the mean measured value). Slope is the best-fit slope of model *versus* measured data.

of data have been used in the comparison of each model, especially in the low wave height range. From Figure 13 and Table III (right) we see that the best fit is provided by C. With the exception of F, all the other models underestimate the significant wave heights throughout the basin. This varies from the 4% of G to the 15% of B and E. Note the large SI of E, apparently connected with its more limited number of points.

We conclude the validation of the model results by considering the conditions at the positions of the three considered buoys, respectively 61 002, Begur and Mahon, whose positions are shown in Figure 4. We begin with 61 002, the buoy in the Gulf of Lion. For our purposes the buoy only provides the significant wave height. U and H_s time histories for the locally relevant 36 h are plotted in Figure 14.

The upper panel depicts the local U time series according to the different models. Before the mistral begins to blow (see Figure 2(a)) all the models evolve coherently, reaching a temporary local peak at 0600 UTC on 24 January. The models start to diverge with the onset of the mistral, their values at 1200 UTC ranging from 6 to 18 m s^{-1} . They

converge towards the peak conditions at 1800 UTC, then progressively slow down in a coherent way. This is consistent with what has been derived from Figures 5, 7 and 8 and together this suggests that, beside a time difference, there are also rather different spatial distributions within the various model wind fields. This is confirmed by the time series of the model wave heights (lower panel in Figure 14), whose values at the buoy position remain rather different in the local decaying phase of the storm, notwithstanding the local similarity of the wind speeds. Note the time series of the recorded H_s – lower than most of the models – during the period H_s is locally growing. Note also the outstanding value of the E model, which is consistent with Figure 7. Together with the scatter diagrams in Figure 10, all this suggests that, at least in the early hours, most models provide a wind field higher than the truth. The results of the comparisons with the 61 002 buoy data are summarized in Table IV. Note the different buoy mean values as, for each model, they refer to the buoy data corresponding to the data (time interval) available for each model. For all the buoys the ENS results will be discussed in the next section.

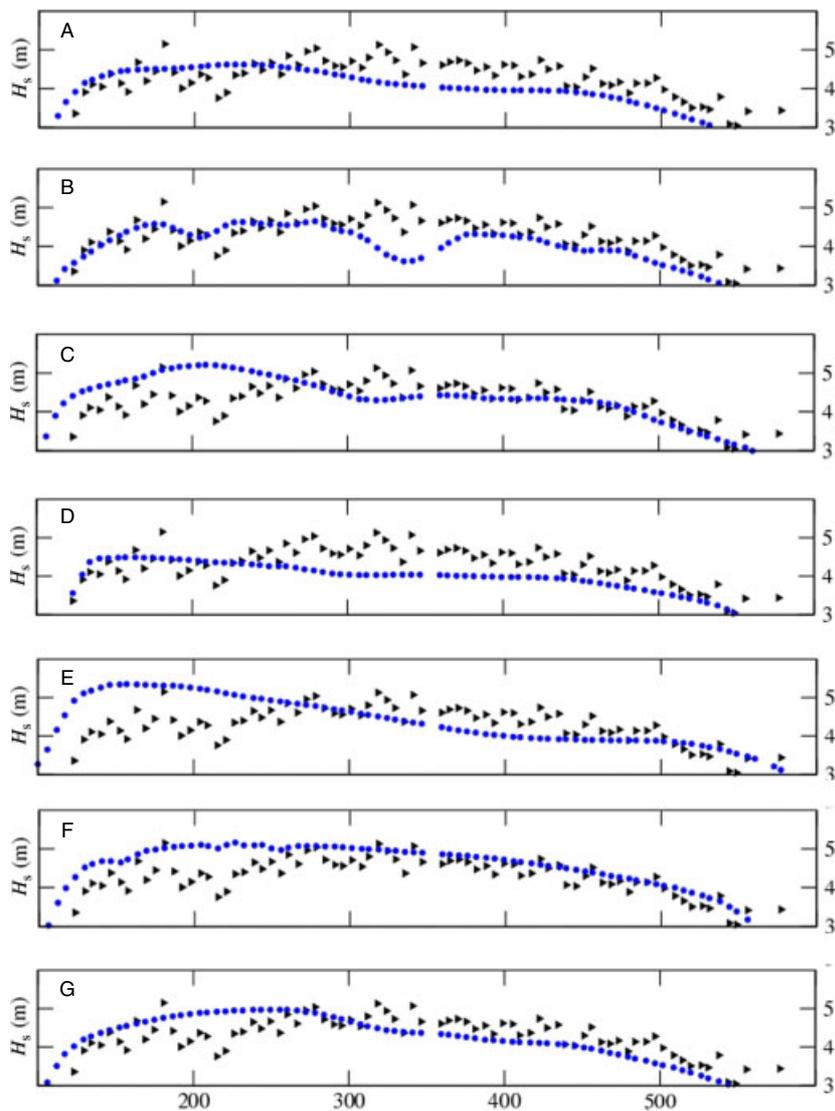


Figure 12. For each one of the seven considered models is shown the comparison between the model (dots) and the altimeter (triangles) significant wave heights along the central part of the ground track in Figure 11. Horizontal scale in km along the ground track of the satellite. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

The results for the Begur buoy, for which also wind data are available, are reported in Table V. The wind speed recorded at this position (hourly data) showed extended and irregular oscillations, down to 1 m s^{-1} , that we suspect were associated with the position of the buoy: aside the main jet and on the lee side of the side mountains. Whatever the reason, these oscillations are barely present in the model results, and whenever present are mostly out of phase. This explains the very large scatter shown in the upper (wind) panel of Table V. Note that, except perhaps for D, on average the bias and slope results are not correspondingly poor. This suggests that the field structures were on average correct, but they missed the relatively short-term variability (with respect to the storm time-scale). This is supported by the corresponding wave results (lower panel). As an integrated effect, in space and time, of the driving wind fields, the wave fields have a strong tendency to dampen, in their distribution, the irregularities of the wind; hence the much lower scatter indices. What is surprising is that in general the biases are negative (underestimated by the models) as compared to the mostly positive wind biases. With reference to the ideal unitary slope, the same is true for the best-fit slopes.

The same tendency, i.e. higher model wind speeds and lower H_s , is present, although at a lower level, at the Mahon buoy (Figure 4), east of Menorca. The results are given in Table VI. Here the scatter indices have more standard values, again the waves displaying their ability to filter out a large part of the wind speed variability. Acknowledging the high and low model tendencies for U and H_s respectively, there is in general a corresponding scaling between the two sets of fields, in that a model having, with respect to the other models, higher wind speeds also has higher wave heights (the exception is G, most likely due to the structure of the fields). However, once associated with the wind–wave dichotomy, this implies that the models performing well for U at Mahon are not the ones scoring better for H_s . Thus the relative good and poor wind performance of D and F are inverted when considering the waves.

7. Multi-model ensemble

In meteorological modelling and its coupled and driven model uses, e.g. wind wave and fluvial models, the deterministic forecast is often complemented with ensemble

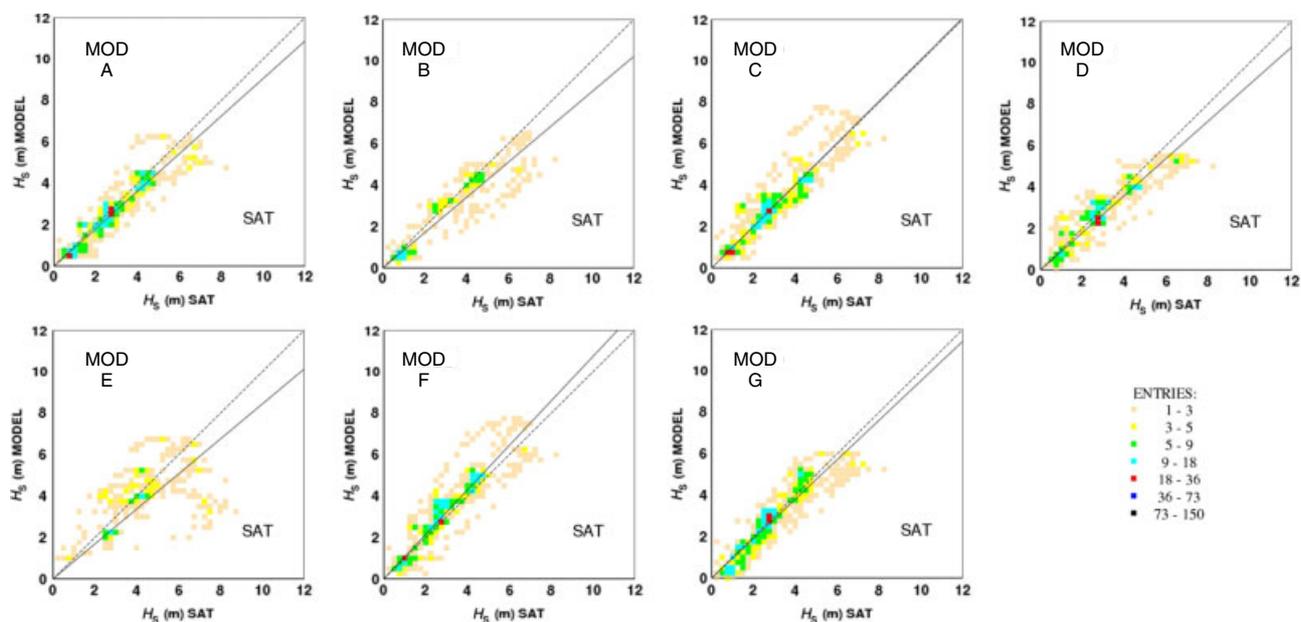


Figure 13. Scatter diagrams for the fits between the seven considered model significant wave heights *versus* the Jason measured values. All the Jason passes over the Mediterranean during 24 and 25 January 2009 are considered. Shading (see scale on the right) indicates the number of cases in each pixel. In all the diagrams the scales go from 0 to 12 m, at 2 m intervals. See Table III for the corresponding bias, scatter index and best-fit slope values. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

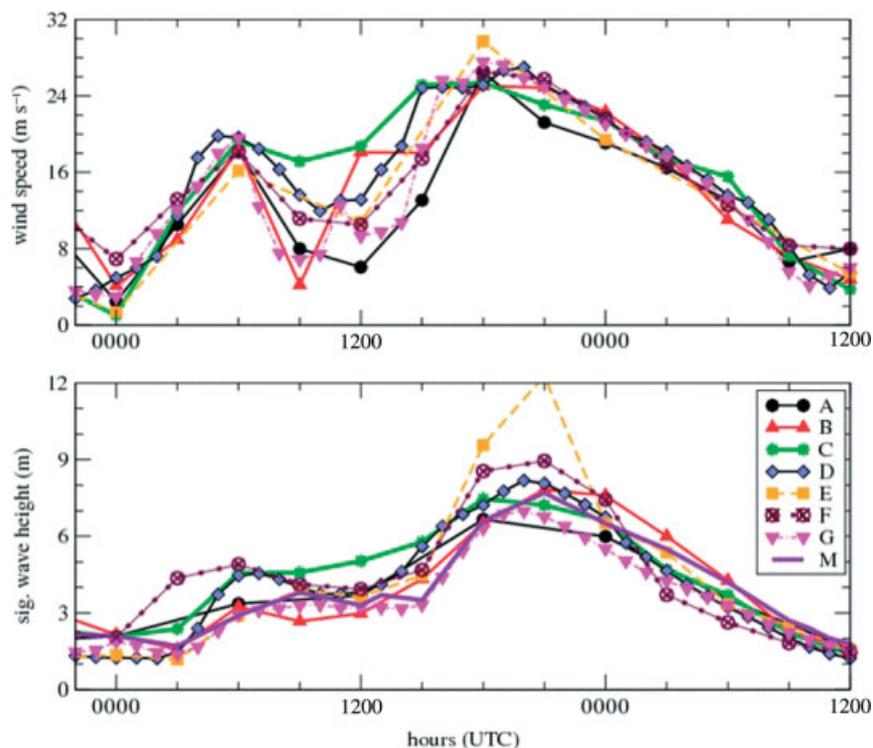


Figure 14. Time series of the wind speed (upper panel) and significant wave height (lower panel) at the position of buoy 61002 moored in the Gulf of Lion, according to the seven considered models. See Figure 4 for the buoy position. Hours start at 0000 UTC on 24 January 2009. Ticks at 3-hourly interval. The lower panel shows also (M) the locally recorded H_s data.

forecasts. The basic idea is that the approximations intrinsic both in the models and in the analyses the forecast starts from can be simulated by slight, controlled modifications both in the model and the analysis – in so doing obtaining a set of alternative forecasts. Besides providing, via their spread, information on the reliability of the deterministic forecast, the ensemble provides a probabilistic approach to meteorological and associated models forecasting. Extensive descriptions of the theoretical background and quality

of the results can be found, for example, in Buizza and Hollingsworth (2002), Buizza (2008) and Zsoter *et al.* (2009).

A similar, but different, approach is to consider not the variations of a single forecast system, but the outputs of N different, independent forecast systems. If the systems have comparable general reliability, the statistical combination of their outputs is shown to improve the prediction of the events. Applied preferably on time-scales longer than the

Table IV. Comparison of the model wave heights *versus* the measured data at buoy 61002 in the Gulf of Lion (see Figure 3 for its position).

Wave	Buoy 61002 in the Gulf of Lion			
	Mean (m)	Bias (m)	SI	Slope
A	3.91	−0.07	0.11	0.98
B	4.02	0.06	0.14	1.04
C	4.02	0.41	0.25	1.09
D	3.77	0.03	0.24	1.05
E	4.02	0.48	0.39	1.24
F	4.02	0.48	0.34	1.14
G	3.77	−0.47	0.10	0.88
ENS	4.02	−0.09	0.19	1.00

The different buoy mean values depend on the time interval with which the data are available from each model. SI is scatter index (RMS error divided by the mean measured value). Slope is the best-fit slope of model *versus* measured data.

Table V. Comparison of the model wind speeds and wave heights *versus* the measured data at the Begur buoy close to the Gulf of Lion (see Figure 3 for its position).

Wind	Buoy at Begur			
	Mean (m s ^{−1})	Bias (m s ^{−1})	SI	Slope
A	8.98	0.19	0.35	0.98
B	9.83	0.63	0.35	1.10
C	9.53	0.84	0.39	1.09
D	8.20	2.50	0.56	1.25
E	9.53	0.04	0.42	0.99
F	9.53	−0.40	0.43	0.90
G	9.11	0.99	0.32	1.13
ENS	9.53	0.68	0.40	1.06

Wave	Buoy at Begur			
	Mean (m)	Bias (m)	SI	Slope
A	2.41	−0.33	0.17	0.86
B	2.53	−0.36	0.24	0.85
C	2.48	0.01	0.17	0.99
D	2.47	−0.39	0.20	0.88
E	2.48	−0.29	0.23	0.96
F	2.48	−0.28	0.19	0.89
G	2.50	−0.40	0.11	0.85
ENS	2.48	−0.29	0.14	0.90

The different buoy mean values depend on the time interval with which the data are available from each model. SI is scatter index (RMS error divided by the mean measured value). Slope is the best-fit slope of model *versus* measured data.

medium-range forecasts, the approach has been shown to provide useful information on seasonal (Hagedorn *et al.*, 2005) and spring flows forecasting (Regonda *et al.*, 2006), among others. Hagedorn *et al.* (2011, personal communication) have further extended the concept, making a multi-model ensemble of ensemble forecasts.

In the present paper, having at our disposal the outputs of seven different models, we have explored how their multi-model ensemble fares with respect to the single models. Seven models and a single storm provide too limited information to derive any extended meaningful statistics. Rather, we have used their combined information to derive alternative statistics that are reported as ENS in the last row of Tables II–VI. Owing to the different output time intervals used by the individual models, hourly ensemble

statistics have been computed using data from only those models that have an output at that time.

It is clear that a multi-model ensemble performs on average better than the single models when the ensemble members are distributed around the truth. When the model distribution is substantially shifted with respect to observations, then the same will be true for the ensemble. This is clearly visible in the ENS results. An example of the first kind is given by the comparison with altimeter data during the 25 January pass (Table III). As we have seen in Figure 12, all the models follow the general trend of the altimeter data, but with limited positive or negative differences with respect to it. This is reflected in the good performance of the ensemble, with a 2 cm bias and a unitary best-fit slope. On the contrary, the strongly biased

Table VI. Comparison of the model wind speeds and wave heights *versus* the measured data at the Mahon buoy close to Minorca of the Balearic Islands (see Figure 3 for its position).

Wind	Buoy at Mahon			
	Mean (m s ⁻¹)	Bias (m s ⁻¹)	SI	Slope
A	10.28	1.45	0.18	1.14
B	10.70	1.41	0.25	1.13
C	10.63	1.84	0.25	1.18
D	10.33	0.73	0.25	1.10
E	10.63	1.53	0.27	1.15
F	10.63	3.06	0.26	1.27
G	10.59	2.34	0.21	1.22
ENS	10.63	1.76	0.20	1.17

Wave	Buoy at Mahon			
	Mean (m)	Bias (m)	SI	Slope
A	3.45	-0.41	0.17	0.90
B	3.47	-0.53	0.12	0.85
C	3.44	0.08	0.17	1.04
D	3.61	-0.60	0.10	0.85
E	3.44	-0.42	0.14	0.91
F	3.44	-0.02	0.16	1.00
G	3.44	-0.44	0.09	0.88
ENS	3.44	-0.34	0.08	0.92

The different buoy mean values depend on the time interval with which the data are available from each model. SI is scatter index (RMS error divided by the mean measured value). Slope is the best-fit slope of model *versus* measured data.

comparison with buoy data, which are generally different for each location and parameter, leads to similarly biased results in the ensemble.

With the aim of providing an overall judgement on the usefulness of the multi-model ensemble, we have tried to draw an overall classification of the combined wind/wave performance of the different models, including the ensemble. The procedure is briefly described in Appendix C. The multi-model ensemble is ranked third in wind modelling, first by far in wave modelling and a clear first in the overall wind/wave score.

8. Discussion

In the present paper we have tried to assess how high-resolution models performed in a very severe storm in the Mediterranean Sea. The storm we have selected, 24–25 January 2009, was one of the most severe in south-western Europe during the last 10–20 years, resulting in substantial damage and fatalities in the countries it crossed along its path. See Section 3 for a description of the storm's development. For the purpose of this study, we have selected seven major meteorological and wave models operational on the Mediterranean area or part of it. All the models cover the area of interest where we have focused our attention.

Our purpose was not to derive from the analysis of the data an estimate of the general performance of the single-model systems. This independent information, where available, is provided in Appendix B. Clearly a single storm, particularly if exceptional, is not sufficient to judge the long-term quality of the system results. In fact, owing to the extreme intensity of the storm, there is the possibility that some of the assumptions or parametrizations at the

base of each model, especially the wave models, may not be completely correct in such conditions; hence the interest for exploring their performance in the 'Klaus' storm. Similar extreme conditions occur associated with hurricanes and typhoons in other parts of the world. For such events, a reliable forecast is vital for civil protection purposes. The intercomparison among the different model results and their cross-check *versus* measured data provide a good, although simple, picture of their performance in such a situation. This may be less the case for SHOM because in the considered period this system was run with a different from the usual wind set-up.

8.1. Main findings

8.1.1. General

Obviously, all the models show the presence of the exceptional storm. However, we have found substantial differences among their results. Considering the time evolution, there is a clear difficulty in properly modelling the entrance of the storm in the Mediterranean Sea. While all the models agree (Figure 14(a)) about the time of the first wind peak, there is an enormous spread in the transient phase to the second, higher peak 12 h later. The models converge again at the peak, but the differences in the previous hours have obvious consequences on the corresponding wave fields.

The difficulty of modelling wind, and hence waves, in the sea just out of a mountain range is clearly exemplified by the wind record at the 61002 and Begur buoys (Figure 4) just off the coast of the Gulf of Lion. The models show no trace of the large and long oscillations recorded in the area that are characteristic of the winds in the lee of the mountains.

8.1.2. Wind

While all the models show maximum wind speeds above 25 m s^{-1} , there are substantial differences in the specific peak values, ranging from 26 to 30 m s^{-1} . Most of the models agree on the time of the peak: 1800 UTC on 24 January. However, the distribution of the respective spatial locations is quite spread, with the distance from the coast ranging from 100 to 230 km. Again, this is connected to the difficulties of modelling the atmospheric situation over and in the lee of a sharp mountain range.

Compared to the data from two scatterometer passes close to the time and area of the peak conditions, most of the models show an overestimate of the wind speeds.

8.1.3. Waves

This has obvious consequences on the corresponding wave fields. Even neglecting the very large E peak (for which an explanation will soon be given), the range of the maximum H_s is from $<8 \text{ m}$ to $>11 \text{ m}$: a difference of almost 100% in the energy of the system. The fetch where these values are reached varies from 200 to 500 km (Figure 8).

Owing to the lack of altimeter passes during the most interesting phase of the storm, wave heights could be checked on an extended path only in the aftermath of the storm, i.e. the day after the peak, when still vigorous, but lower, waves were progressively moving towards the more eastern part of the basin. On the other hand, because observed and modelled wave conditions were very strongly related to the respective values in the previous hours, they are still good indicators of the general performance. Here we have found that all the models clearly show the storm and largely follow the wave profile along the track. However, differences are present, not only in the finer details, as the short-term variability of the measured data, but also in the general trend.

A matter of concern is the fact that, when compared to the measured values, modelled wind speeds appear higher than the truth, but the opposite is true for wave heights. In principle this could be due to the overall wind structures. However, the fact that this happens for all the model systems while also using three different wave models (WAM, WAVEWATCH III, SWAN) hints to something where more attention is probably needed. Indeed the local conditions, characterized, at least at short fetches, by very strong wind and very short waves, with intense breaking of practically all the waves in the field, remind us of the extreme conditions in hurricanes where the wind wave generation has characteristics different from what we expect to observe in regular storms.

8.2. General

Some of the differences found among the different systems may be due to the different time steps with which the wind information is passed to the wave models. There is a clear difference between, for example, the fully coupled system of ECMWF, where the meteorological and wave models are integrated with a reciprocal continuous exchange of information, and Puertos del Estado, where the wind information was passed to the wave model, with no coupling of wave data back to the wind model, once every 6 h. Obviously, the frequency of the wind input to the wave model is critical in the case of a rapidly varying storm – the more so if it is exceptional, as ‘Klaus’ was. Indeed the first

results we obtained stimulated Puertos del Estado to explore the specific situation. It was found that a short transient period of very high wind speeds in the meteorological model was applied as the 6-hourly input to the wave model, leading to the very large wave heights seen in Figure 6 (E_{21}). A positive result of this study was the decision of Puertos del Estado to move to a 1 h wind forcing interval to drive the wave model.

8.3. Ensemble

We have explored whether the model results could be improved by considering their multi-model ensemble. The results, reported in the last row of Tables II–VI, show that this is indeed the case. In most cases the ensemble results rank among the best performers of the set, being the only ones always in this position. We have also attempted to extract from the various tables, although in a rather simple way, a summary score capable of providing a hint on the overall performance of each system. In this score the multi-model ensemble has a clear first position.

This is a positive result. The fact that the ensemble – better, its mean – is on average closer to the measured values suggests that there is no, or at most a limited, systematic error in the models. If this is the case, the spreading around the truth is associated with the differences present in the initial analyses, numerics, parametrizations, etc. In this respect a further development, well beyond the present paper, would be to analyse, where available, the ensemble of the single models or, even more, following Hagedorn *et al.* (2005) and Park *et al.* (2008), the ensemble of the ensembles. Hagedorn *et al.* (2011, personal communication) point out that the model ensemble can be further improved if applied to only a subset of the available models, selected on the base of the statistics of their previous performance. Clearly this is not possible when acting on a single storm.

8.4. Model performance

The above scoring was done only to judge the potential usefulness of a model ensemble. It was not our purpose to reach a judgement on the performance of each model, neither to address the problem of explaining in detail the reasons for the single errors; obviously this is not possible within the time-scale we have considered. Clearly the above results trigger a number of questions, the first one being why the models perform in a certain way. For instance, as already mentioned, we are puzzled by the fact that the results show an overestimate for wind speed and an underestimate of the wave heights. If this is not simply justified by the limited amount of measured data, also shifted in time and space, then we have to look for a more physical explanation. The mentioned generation (and dissipation) in an extremely young sea is a real possibility. Also, following Abdalla and Cavaleri (2002), the strong oscillations in the mistral in the lee of the mountains and propagated into the sea – oscillations not reproduced by the meteorological models – may have contributed to the $U-H_s$ apparently inconsistent results.

Many more aspects would – and probably will – be worth a deeper analysis; for instance, the fact that some models show similar results from one point of view, e.g. the peak position, only to differ substantially when we consider their overall distributions. Clearly, as each model system is a specific and

self-standing machine, the further deepening of these many aspects is far beyond the scope of this study. Our aim was to verify the reliability of these high-resolution local models when used to forecast a severe storm. Purposely, and not only for practical reasons, during our discussion we have referred to the different models in an anonymous way. However, it would be of interest to compare the performance of the considered storm with the long-term performance of the various systems. Appendix B provides some statistics for five of the seven sources, derived from one or two years of data in the Western Mediterranean Sea. Four of the sources have a long-term participation in the JCOMM intercomparison (see Appendix B). The results in Figure BI and Tables BI and BII concern the Western Mediterranean, and more specifically seven buoys that include the three we have used in this study (see Figure 4 for their position); the other are distributed along the coast of Spain as far as Gibraltar. Given the buoy positions, we feel that the statistics reported in Appendix B cannot be compared with the present open sea results, say those in Figures 9–13. We limit ourselves, and still in very general terms, to the corresponding results, although for only three of the buoys, shown in Tables IV–VI.

For the bias, the results for the present experiment seem consistent with the long-term ones. As expected, given the rather different field structures we have found in the seven systems (Figures 5 and 6), there is a wide spread in our results. Nevertheless, by and large, with the exception of the wind at Mahon on the main flow of the storm, there is some coherence between the single storm and the long-term values. The same is true for the best-fit slope, again with the exception of the wind at Mahon. We find the main differences in the scatter. Acknowledging the large SI wind values at Begur, likely because of the close mountain effect, the SI average values derived from the experiment are substantially lower than the long-term ones. This is likely related to the more coherent structure of the fields during a storm and the larger value of the normalizing mean measured value. Indeed, in the calmer (but not always) conditions that typify the Mediterranean Sea, the weather patterns may be less defined and the SI correspondingly larger. This is consistent with the differences in Tables II and III when considering first the peak and area of the storm, and then the whole Mediterranean Sea along a slightly more extended period. From this perspective, our comments should be positive. However, it is clear that the level of approximation acceptable in 'normal' conditions may not be sufficient in a very severe storm, when time, location and values of the maxima are obviously relevant. Each one of us, as each reader, can derive his/her own judgement, but we strongly suggest that the main message of this study is neither in the performance of the single models nor in their average with respect to their long-term performance. Rather, as by now repetitively said, we point our attention on the differences found during a very severe storm among the results of seven of the best models operational in the Mediterranean Sea. In a time when meteorological and wave modellers often claim optimal results, for practical purposes, but also for more fundamental reasons, this may be a matter of concern.

Appendix A: the Seven Considered Models

We provide here a short description of the seven considered forecast systems operational in, or part of, the Mediterranean Sea.

A: European Centre for Medium-Range Weather Forecasts (ECMWF) (UK)

The wave model is a modified version of WAM, cycle 4 (Bidlot *et al.*, 2007; Janssen, 2008). In its global configuration, it is fully coupled to the global atmospheric model (Janssen, 2004). The wave model provides a feedback to the atmosphere in the form of information on the change of surface roughness in relation to the varying momentum and energy flux from the atmosphere of the growing waves. ECMWF also runs operationally an uncoupled version of the same code for the seas around Europe, in particular the Mediterranean Sea. Data from this configuration were used in the comparison. For the storm of January 2009, the horizontal resolution of the stand-alone model was of the order of 28 km, with a spectral resolution of 24 directions and 30 frequencies, starting from ~ 0.035 Hz. A 10 m wind speed and direction from the coupled global atmospheric model with resolution of the order of 25 km (T799) were used to force the model at 6-hourly intervals. Forecasts are produced twice daily from the 0000 and 1200 UTC analyses. Altimeter wave height and ASAR data were used in the wave model assimilation.

Information on the atmosphere and wave models are available at

<http://www.ecmwf.int/products/forecasts/guide/index.html>

Detailed documentation can be found at

<http://www.ecmwf.int/research/ifsdocs/>

B: Météo-France (France)

VAG is the model currently in operation at Météo-France for the Mediterranean Sea, and is based on a 'second-generation' parametrization of wave–wave interactions together with wind generation and dissipation formulations equivalent to the WAM cycle 4 model (Fradon, 1997; Fradon *et al.*, 2000; Lefèvre *et al.*, 2003). It is driven by ALADIN winds (limited area model operational at Météo-France; ALADIN International Team, 1997), wind forcing every 3 h, four times a day. The horizontal resolution of VAG and ALADIN is approximately 10 km.

It should be noted that ALADIN is nested into ARPEGE/IFS (Courtier *et al.*, 1991), and that ARPEGE/IFS is also used at ECMWF but with slightly different settings, and in a different operational context. The models are covering only the western part of the Mediterranean Sea (above 30°N and west to 17°E). VAG is coupled to a model covering the whole Mediterranean basin and driven by ARPEGE winds, allowing the generation of waves in the Eastern Mediterranean.

C: CNMCA (Italian Meteorological Service) and ISMAR-CNR (Italy)

Nettuno (Bertotti *et al.*, 2010) is the wave forecast system running operationally at CNMCA (Centro Nazionale di Meteorologia e Climatologia Aeronautica of the Italian Meteorological Service) developed in cooperation with the ISMAR-CNR Institute. This system is based on the ECMWF version of the WAM model (<http://www.ecmwf.int/research/ifsdocs/>) forced by the atmospheric wind of COSMO-ME (<http://www.cosmo-model.org/con>

Table B I. Long-term performance of the meteorological and wave models for four of the considered systems.

(a) Day 1 forecast statistics for 10 m wind speed from July 2008 to June 2010

wind ($n = 6547$)	buoy mean (m s^{-1})	bias (m s^{-1})	SI	Slope
ECMWF	6.05	-0.31	0.29	0.94
Puertos del Estado	6.05	-0.15	0.40	0.97
SHOM	6.05	-0.38	0.30	0.93
UK Met. Office	6.05	-0.43	0.32	0.94

(b) Day 1 forecast statistics for significant wave height from July 2008 to June 2010.

H_s ($n = 7128$)	buoy mean (m)	bias (m)	SI	Slope
ECMWF	1.10	-0.14	0.24	0.89
Puertos del Estado	1.10	-0.08	0.34	0.93
SHOM	1.10	-0.16	0.28	0.89
UK Met. Office	1.10	-0.13	0.33	0.92

Table B II. Long-term performance of the CNMCA-ISMAR system. Day 1 forecast statistics from July 2008 to June 2010.

CNMCA – ISMAR	Bias	SI	Slope
Significant wave height (m)	-0.08	0.21	0.98
Wind speed (m s^{-1})	-0.76	0.32	0.97

tent/tasks/operational/usam/default.htm), which is the 7 km CNMCA operational set-up of the non-hydrostatic regional model developed by the Consortium for Small-Scale Modelling (COSMO; Steppeler *et al.*, 2003). COSMO-ME is initialized by the CNMCA 3D-VAR data assimilation system (Bonavita and Torrisi, 2005) and driven by the IFS boundary conditions.

The WAM model is integrated twice per day (0000 and 1200 UTC) up to 72 h over the entire Mediterranean basin with a grid spacing equal to 3'. The model is run with 36 directions and 30 frequencies starting from ~ 0.05 Hz and it is forced by the hourly COSMO-ME wind. Three-hourly forecast plots of significant wave, mean period and mean direction are available at <http://www.meteoam.it/modules.php?name=Nettuno> and http://ricerca.ismar.cnr.it/MODELLI/ONDE_MED_ITALIA/ONDE_MED_ITALIA.php.

A 1' implementation of WAM over the Italian seas forced by 2.8 km COSMO-IT model is currently pre-operational at CNMCA.

D: ARPA-SIMC and Italian Civil Protection (Italy)

MEDITARE (Valentini *et al.*, 2007) is the sea-state forecasting system implemented in 2006 by ARPA-SIMC (Hydro-Meteo-Climate Service of Emilia-Romagna, <http://www.arpa.emr.it/sim>) as the centre of competence for meteorological and sea modelling of the National Civil Protection Department in Rome (DPCN), which is still financing its maintenance and improvement. It is based on the SWAN model (Booij *et al.*, 1999), version 40.51AB, and consists of a sequence of nested runs that start from a coarse run over the Mediterranean Sea at a resolution of ~ 25 km. Such a run produces the necessary boundary conditions for the following run over the whole Italian domain at a resolution of 8 km. The following step is designed to achieve quite high resolutions (about 800 m)

in small coastal domains, by means of the same nesting technique.

The SWAN model is run in third-generation mode, with the wind growth formulation by Komen *et al.* (1994) for exponential growth and by Cavaleri and Malanotte-Rizzoli (1981) for linear growth. The computation of nonlinear triad and quadruplet wave interactions is activated; bottom friction is taken into account by means of the Madsen *et al.* (1988) formulation. The spectral resolution is 36 directions and a frequency range of 0.05–1 Hz. Operationally the system provides one run per day at 0000 UTC with a forecast range of 72 h, with hourly outputs.

The model is driven by the 10 m hourly wind computed by the meteorological model COSMO-I7 (COSMO Newsletter, 2004) and by the wind computed by IFS-ECMWF outside the COSMO-I7 integration domain. COSMO-I7 is one of the Italian 7 km resolution implementations of the non-hydrostatic numerical weather prediction model COSMO (Steppeler *et al.*, 2003) that is operational at ARPA-SIMC.

The wave forecasts are available at <http://www.arpa.emr.it/sim/?mare>

E: Puertos del Estado (Spain)

The Puertos del Estado wave forecast system is based on the WAM cycle 4 model. The system is run in a twice a day cycle with a horizon of 72 h.

The system consists of two applications: one for the Western Mediterranean Sea, with a resolution of 5' in its western part, and another for the North Atlantic Ocean. To complete the coverage, four nested applications have been developed for the Cantabric Sea, with a resolution of 2.5'; the Strait of Gibraltar, with a resolution of 1'; and the Gulf of Cadiz and the Canary Islands, both with a resolution of 5' (Gomez and Carretero, 2005).

The WAM model is run with 24 directions and 25 frequencies in the Atlantic Ocean and 30 frequencies in

the Mediterranean Sea. A two-way nesting procedure for the WAM model, to increase the resolution in one single run, has been developed at Puertos del Estado.

The results of the 15 previous days of each forecast cycle are validated in real time *versus* 20 buoys of the Puertos del Estado buoy network, and the corresponding time series are updated twice a day together with the forecast maps and numerical tables at the Puertos del Estado web site: www.puertos.es.

The 10 m wind fields used to drive the system come from the HIRLAM model application operated at the Spanish Meteorological Agency. The spatial resolution is 16 km and the time resolution is 6 h.

8.5. F: SHOM – Service Hydrographique et Oceanographique de la Marine (France)

The SHOM system of the French Navy provides daily forecasts of the wave conditions on the globe and, at higher resolution, in the Mediterranean Sea. The forecast is based on version 3.14-SHOM of the WAVEWATCH III wave model (Tolman, 2009). It is amply described in Magne *et al.* (2010), using the parametrization TEST405, further discussed in Ardhuin *et al.* (2010) With respect to the standard WAVEWATCH III formulation, new parametrizations for the spectral dissipation of wind-generated waves have been introduced. The rates of dissipation have no predetermined spectral shapes and are a function of the wave spectrum and wind speed and direction. Different formulations are used for the dissipation of swell and for dissipation due to wind–wave breaking. An additional source of short-wave dissipation is considered due to the influence of swell on the short waves.

The model is run twice a day with 0.1° resolution, with a 6-day forecast range. As a rule, the system uses the wind forecast of ECMWF, available at 6 h intervals with 0.5° nominal resolution (increased to 3 h and 0.25° respectively starting 20 February 2009; hence after the considered storm). However, during January 2009, due to lack of connection with ECMWF, the wave model was forced with NCEP GPS winds, with the wind speeds enhanced by 10% to correct for biases in the Mediterranean Sea.

G: UK Meteorological Office (UK)

The Met Office's North Atlantic and European (NAE) wave forecasting system is based around the WAVEWATCH-III (WW3) spectral wave model (Tolman, 2007) using the Tolman and Chalikov (1996) source terms and modified to include the second-order advection scheme of Li (2007). It runs four times a day out to $T + 60$ h with a 1 h temporal resolution and a spectral resolution of 25 frequencies and 24 directions.

The NAE configuration covers the whole of the Mediterranean Sea and northeast Atlantic with a spatial resolution of ~ 12 km ($0.11^\circ \times 0.11^\circ$) on a rotated pole grid (pole located at 37.5° N, 177.5° E). Boundary conditions are provided by the Met Office's Global WW3 configuration. At this time, surface currents are not included as an input to the model.

Hourly 10 m wind forcing is provided by the NAE configuration of the Met Office's Unified Model (UM) (Davies *et al.*, 2005). The UM is a suite of coupled atmospheric and ocean models that produces NWP forecasts and climate predictions on both global and regional scales. The NAE

configuration of the UM runs on the same horizontal grid as the NAE wave model with 38 levels in the vertical.

Appendix B: Long-Term Performance of the Considered Systems

In this appendix we provide, where available, some information on the long-term performance, for both wind and waves, of the considered systems.

Starting in 1996, ECMWF pioneered, and has been maintaining since then, a monthly intercomparison of the performance of the participating institutions. Started as a friendly exercise among a very few partners, the project, now endorsed by JCOMM (Joint Technical Commission for Oceanography and Marine Meteorology), gained momentum and it includes 13 members at the time of writing. More detailed information is provided by Bidlot *et al.* (2002, 2007). The validation, *versus* wind and wave buoy data, is done on a global basis and in particular also on the Mediterranean Sea. For our present purposes we show in Figure BI two plots showing the bias (significant wave height and wind speed) for the Western Mediterranean Sea for four (ECMWF, Puertos del Estado, SHOM, UK Meteorological Office) of the seven models considered in the present paper. Seven buoys have been considered: the three marked in Figure 4, plus

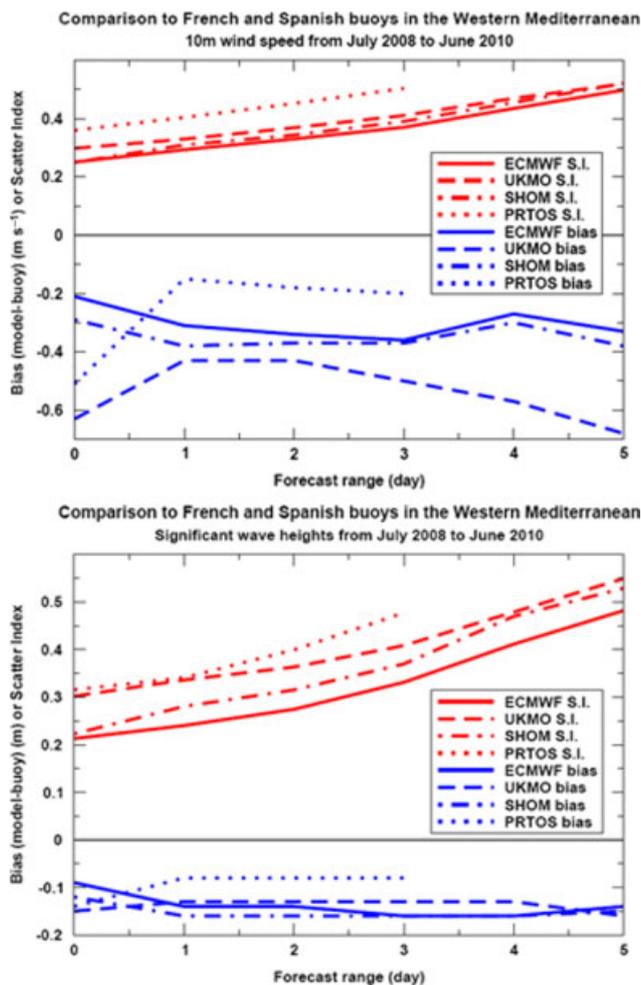


Figure BI. Wind speed and significant wave height scatter index and bias at different forecast ranges (days) for four of the considered systems. The period considered is July 2008 to June 2010. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

four others, one of which is close to Majorca (the largest Balearic Island). Of this last group three are distributed along the Spanish coast up to the Alboran Sea, close to Gibraltar. The results, shown for an extended forecast range where available, are given at 24 h intervals (24, 48, . . .). The overall results for 24 h – the range of interest in the present paper – for bias, scatter index and best-fit slope, are reported in Table BI. They refer to the period July 2008 to June 2010.

The results for Météo-France are not available separately for the Mediterranean. Some information about their performance can be found in Arduin *et al.* (2007).

The results for CNMCA-ISMAR are reported in Table BII. They refer to the period July 2008 to June 2010 and concern the Begur and Mahon buoys (see Figure 4) and altimeter data (for wave height).

Appendix C: Overall Scoring of the Seven Models + Model Ensemble

Given the bias, scatter index and slope parameter values in Tables II–VI, we look for a method, simple but reasonably representative, to classify the overall performance of the single models. The only aim was to see how the multi-model ensemble ENS scored with respect to the seven considered wind and wave models.

We have three output parameters, eight models, three tables for wind speed (II, V, VI) and four tables for significant wave heights (III–VI). Tables II and Table III provide two values each for scatter index and best-fit slope.

For each model and parameter, e.g. wind speed, we have evaluated the average parameter values using the available table values (four for U , five for H_s). These average values are used to score the single model: 1 for the best one, 2 for the second, . . . , 8 for the poorest one. Note that for bias we have considered absolute values.

This provides three scores for each model (one parameter) that we can add, reaching the overall classification (for wind speed). A similar procedure has been followed for significant wave height. As a further step the two (wind and wave) summary scores have been added, providing an overall single score for each model, and hence a classification of the relative performance of a single model within the group.

The method is obviously crude. For instance, the 1 to 8 scoring does not consider the quantified differences of, for example, the scatter index, but simply their sequential ordering, from the best to the poorest one. However, there was no point in devising a more sophisticated system, having only one storm at our disposal. Our only interest was to determine, also with some approximation, how the multi-model ensemble fared with respect to the seven models. The numerical results, not reported for the reasons repetitively explained in the paper, supported our opinion.

Acknowledgements

The description of the genesis and evolution of the ‘Klaus’ storm on the Atlantic Ocean has been derived from Liberato *et al.* (2011). Its surface description and evolution in the Mediterranean Sea have been provided by Roberto Buizza. Rob Hines put his master hand to most of the figures.

References

Abdalla S, Cavaleri L. 2002. Effect of wind variability and variable air density on wave modelling. *J. Geophys. Res.* **107**(C7): 17–11717.

- Abdalla S, Janssen PAEM, Bidlot J-R. 2010. Error estimation of Jason-2 wind and wave products. *Mar. Geod.* **33**(S1): 239–255.
- ALADIN International Team. 1997. The ALADIN project: mesoscale modelling seen as a basic tool for weather forecasting and atmospheric research. *WMO Bull.* **46**(4): 317–324.
- Arduin F, Bertotti L, Bidlot J-R, Cavaleri L, Filipetto V, Lefèvre J-M, Wittman P. 2007. Comparison of wind and wave measurements and models in the Western Mediterranean Sea. *Ocean Eng.* **34**(3–4): 526–541.
- Arduin F, Rogers E, Babanin A, Filipot J-F, Magne R, Roland A, van der Westhuysen A, Queffelec P, Lefèvre J-M, Aouf L, Collard F. 2010. Semi-empirical dissipation source functions for wind-wave models. Part I: definition, calibration and validation. *J. Phys. Oceanogr.* **40**: 1917–1941.
- Bertotti L, Cavaleri L, De Simone C, Torrisi L, Vocino A. 2010. Il sistema di previsione del mare ‘NETTUNO’. *Riv. Meteorol. Aeronaut.* **70**(1): 25–36.
- Bidlot J-R, Holmes DJ, Wittmann PA, Lalbeharry R, Chen HS. 2002. Intercomparison of the performance of the operational wave forecasting systems with buoy data. *Weather Forecast.* **17**: 287–310.
- Bidlot J-R, Janssen PAEM, Abdalla S, Hersbach H. 2007. A revised formulation of ocean wave dissipation and its model impact. ECMWF Tech. Memo. 509. ECMWF, Reading, UK. <http://www.ecmwf.int/publications/>
- Bonavita M, Torrisi L. 2005. Impact of a variational objective analysis scheme on a regional area numerical model: the Italian Air Force Weather Service experience. *Meteorol. Atmos. Phys.* **88**(1–2): 39–52.
- Booij N, Ris RC, Holthuijsen LH. 1999. A third-generation wave model for coastal regions. Part I: Model description and validation. *J. Geophys. Res.* **104**(C4): 7649–7666.
- Buizza R. 2008. Comparison of a 51-member low-resolution (TL399L62) ensemble with a 6-member high-resolution (TL799L91) lagged-forecast ensemble. *Mon. Weather Rev.* **136**: 3343–3362.
- Buizza R, Hollingsworth A. 2002. Storm prediction over Europe using the ECMWF Ensemble Prediction System. *Meteorol. Appl.* **9**: 1–17.
- Cavaleri L, Malanotte Rizzoli P. 1981. Wind wave prediction in the Adriatic Sea. Theory and numerical experiments. *J. Geophys. Res.* **86**(C11): 10961–10973.
- Cavaleri L, Sclavo M. 2006. The calibration of wind and wave model data in the Mediterranean Sea. *Coastal Eng.* **53**: 613–627.
- COSMO Newsletter. 2004. *Operational applications: ARPA-SIM (BOLOGNA)*, Vol. 6. Deutscher Wetterdienst (DWD); 25–26.
- Courtier P, Freydier C, Geleyn J-F, Rabier F, Rochas M. 1991. The ARPEGE project at Météo France. In *ECMWF 1991 Seminar Proceedings: Numerical Methods in Atmospheric Models*, Vol. 2, ECMWF, 9–13 September 1991; 193–231.
- Davies T, Cullen MJP, Malcolm AJ, Mawson MH, Staniforth A, White AA, Wood N. 2005. A new dynamical core for the Met Office’s global and regional modelling of the atmosphere. *Q. J. R. Meteorol. Soc.* **131**: 1759–1782.
- Fradon B. 1997. *Modélisation numérique de l’état de la mer: comparaison des performances et de limites d’un modèle de deuxième génération et d’un modèle de troisième génération dans le cadre de l’expérience SEMAPHORE*. Thèse de doctorat de l’Université Paris VII.
- Fradon B, Hauser D, Lefèvre J-M. 2000. Comparison study of a second-generation and of a third-generation wave prediction model in the context of the SEMAPHORE experiment. *J. Atmos. Ocean Technol.* **17**: 197–214.
- Gómez Lahoz M, Carretero Albiach JC. 2005. Wave forecasting at the Spanish coasts. *J. Atmos. Ocean Sci.* **10**(4): 389–405.
- Hagedorn R, Doblas-Reyes FJ, Palmer TM. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting. 1: Basic concept. *Tellus* **57**(3): 219–233.
- Janssen PAEM. 2004. *The Interaction of Ocean Waves and Wind*. Cambridge University Press: Cambridge, UK.
- Janssen PAEM. 2008. Progress in ocean wave forecasting. *J. Comput. Phys.* **227**: 3572–3594.
- Komen GJ, Cavaleri L, Donelan M, Hasselmann K, Hasselmann S, Janssen PAEM. 1994. *Dynamics and Modelling of Ocean Waves*. Cambridge University Press: Cambridge, UK.
- Lefèvre J-M, Kortcheva A, Stephanescu S. 2003. Performance of several ocean wave forecasting systems for high swell conditions. In *Proceedings of ISOPE 2003 Conference*, Hawaii, HI, May 2003.
- Li JG. 2007. Upstream non-oscillatory advection schemes suitable for ocean wave models. In *Proceedings of the 10th International Workshop on Wave Hindcasting and Forecasting and Coastal Hazard Symposium*, Hawaii, HI.
- Liberato MLR, Pinto JG, Trigo IF, Trigo RM. 2011. Klaus: an exceptional winter storm over northern Iberia and southern France. *Weather DOI*: 10.1002/wea.755.

- Madsen OS, Poon YK, Graber HC. 1988. Spectral wave attenuation by bottom friction: theory. In *Proceedings of the 21st International Conference on Coastal Engineering*, ASCE; 492–504.
- Magne R, Ardhuin F, Roland A. 2010. Waves forecast and hindcast from globalocean to the beach (in French). *Eur. J. Environ. Civil Eng.* **14**: 149–162.
- Park YY, Buizza R, Leuchtecher R. 2008. TIGGE: preliminary results on comparing and combining ensembles. *Q. J. R. Meteorol. Soc.* **134**: 2029–2050.
- Regonda SK, Rajagopalan B, Clark M, Zagana E. 2006. A multimodel ensemble forecast framework: application to spring seasonal flows in the Gunnison River Basin. *Water Resour. Res.* **42**: W09404, DOI:10.1029/2005WR004653.
- Steppeler J, Doms G, Shatter U, Bitzer HW, Gassmann A, Damrath U, Gregoric G. 2003. Meso-gamma scale forecasts using the nonhydrostatic model LM, *Meteorol. Atmos. Phys.* **82**: 75–96.
- Tolman HL. 2007. *User Manual and System Documentation of WAVEWATCH-III*, version 3.12. MMAB/NCEP/NOAA, July 2007.
- Tolman HL. 2009. *User Manual and System Documentation of WAVEWATCH-III*, version 3.14. Tech. Rep. 276, NOAA/NWS/NCEP/MMAB.
- Tolman HL, Chalikov D. 1996. Source terms in third generation wind-wave model. *J. Phys. Oceanogr.* **26**: 2497–2518.
- Valentini A, Delli Passeri L, Paccagnella T, Patrono P, Marsigli C, Cesari D, Deserti M, Chiggiato J, Ribaldi S. 2007. The sea state forecast system of ARPA-SIM. *Boll. Geofis. Teor. Appl.* **48**: 333–349.
- WISE Group – Cavaleri L, Alves J-HGM, Ardhuin F, Babanin A, Banner M, Belibassakis K, Benoit M, Donelan M, Groeweg J, Herbers THC, Hwang P, Janssen PAEM, Janssen T, Lavrenov IV, Magne R, Monbaliu J, Onorato M, Polnikov V, Resio D, Rogers WE, Sheremet A, McKee Smith J, Tolman HL, van Vledder G, Wolf J, Young I. 2007. Wave modelling: the state of the art. *Prog. Oceanogr.* **75**(4): 603–674.
- Zsoter E, Buizza R, Richardson D. 2009. ‘Jumpiness’ of the ECMWF and UK Met Office EPS control and ensemble-mean forecasts. *Mon. Weather Rev.* **137**: 3823–3836.