388

Assessment of the ECMWF ensemble prediction system for waves and marine winds

Øyvind Saetra and
Jean-Raymond Bidlot

Research Department

October 2002

For additional copies please contact

The Library
ECMWF
Shinfield Park
Reading, Berks RG2 9AX

library@ecmwf.int

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications.html

Abstract

The ECMWF Ensemble Prediction System (EPS) for waves and marine surface winds has been evaluated using buoy and platform data from the Global Telecommunication System (GTS), covering a period between September 1st 1999 and March 31st 2002, including three full winters. The significant wave height forecasts have in addition been compared to a global data set of altimeter data covering the same period. For forecasting purposes, the ensemble spread may be regarded as a measure of the uncertainties in the deterministic predictions. In order to demonstrate this, the ensemble spread was divided into different classes. An upper bound for the model errors was established by calculating the corresponding percentiles of the errors for each separate class. The results indicate that there is a strong correlation between the ensemble spread and the deterministic forecast skill. This finding is valid for both wave height and wind speed. The reliability of the probability forecasts for wind and waves are good. However, the reliability diagrams indicate a small tendency for over-confidence in the wave probability forecasts for waves above 6 and 8 metres. This is most pronounced in the southern hemisphere, whereas the reliability for the northern hemisphere is relatively good.

Use of the EPS in decision making is studied by a cost-loss model for the relative economic value (Richardson 2000). For comparison, poor-man's ensembles (PME) have also been created by adding normally distributed noise to the control forecasts. This study reveals that the real EPS performs better than both the PME and the control forecasts in terms of relative economic value. When more complex forecasting parameters are considered, such as the joint probability of wave height and period, benefits of the EPS become more pronounced.

1. Introduction

In June 1998, the Ensemble Prediction System (EPS) at the European Centre for Medium Range Weather Forecasts (ECMWF) was coupled to the ocean wave model. From then on, daily ensemble wave forecasts have been available. Although the positive impact on both the atmospheric and the wave forecasts was the main reason for the introduction of the coupling (Janssen et al. 2002), probabilistic forecasts of ocean waves are also potentially very valuable products. For the offshore and shipping industry, such a forecasting tool could have numerous applications, such as ship routing and the planning of high-risk operations. In many activities out at sea, the most critical environmental parameters are ocean waves. Oil rigs in the ocean are designed to withstand almost any possible wind condition, but extreme waves may in some cases result in serious damage to a platform. A common oil rig design criterion requires that the 100-year maximum wave must not touch the platform deck. It is not necessarily feared that the rig itself might topple, but rather that many of the light-weight installations on the platform deck, such as walking bridges and fences, are not designed to withstand the forces from waves. When hazardous or delicate operations are to be performed, ensemble forecasts could be used to estimate the probabilities of weather events that are considered dangerous. Particularly if such activities need to be planned days ahead, probabilistic forecasts of dangerous weather and sea-states can provide valuable information. One example may be the towing and installation of oil rigs. During such operations, it is vital that certain weather and sea-states are avoided. If not, both the risk to human life and the potential economic loss would be enormous. Since this type of probabilistic information is not available from traditional deterministic models, forecasting systems that are able to reliably predict even small probabilities of such hazardous events would be very useful.

In marine forecasting, some sort of floating object is often involved. This may be anything from small barges and ramps to huge vessels. A common feature for all floating objects is that their response to ocean waves is strongly sensitive to the wave frequency, with a maximum response near the resonance frequency of the structure in question. If this situation occurs, the structure might be subject to violent oscillations even for wave heights that generally would be regarded as relatively small. During the construction of a floating bridge in Salhusfjorden in Norway, the bridge modules were transported from The Netherlands to Norway across the North Sea on barges. During one of these transports one module was lost when it fell off the barge due to strong oscillations of the barge. Very much to the surprise of the skipper on the barge, the wave height

was rather small when this happened. It is likely that the resonant periodic motion of the barge due to ocean waves was the cause of this (Johannes Guddal - personal communication). The wave EPS makes it possible to provide forecasts of the response probability. This type of service will perhaps need to be tailored for the users, say for example a particular container vessel. Using information on the response properties of the vessel and the local sea-state, any motion of freedom, such as the pitch and heave, can in principle be estimated. If certain threshold values for these motions are to be avoided, the EPS could be used to issue maps on which areas with a significant probability of these thresholds occurring are highlighted. Indeed, for many applications, the important forecast parameter is not necessarily the wave height alone, but rather the joint probability of wave height and period, or perhaps some other parameter characterizing the wave energy distribution.

One of the objectives of the EU-funded research project SEAROUTES is to investigate the possible usefulness of ensemble predictions for ship routing. As a first step towards such a goal, the forecast system itself should be tested against observations. Validation of the system is of course necessary to take full advantage of probability forecasts. A decision-maker who has to decide whether or not to take action when the forecast threshold probability of a given event is exceeded, or decide which path to follow during an Atlantic crossing, must be confident that the forecast probabilities reflects the true risk that a given event will take place.

The purpose of this study is to assess the 51 member ensemble forecasts currently running operationally at ECMWF. In a recent study, Vogelezang and Kok (1999) have tested the EPS waves by using two buoys in the North Sea for the period from October 1998 to February 1999. In our study, the forecasts are compared to buoy and platform observations for wind speed, wind direction, significant wave height and peak period obtained via the Global Telecommunication System (GTS). The data coverage can be seen in Figure 1, which shows the positions of all the observations. Except for one platform located off the South African

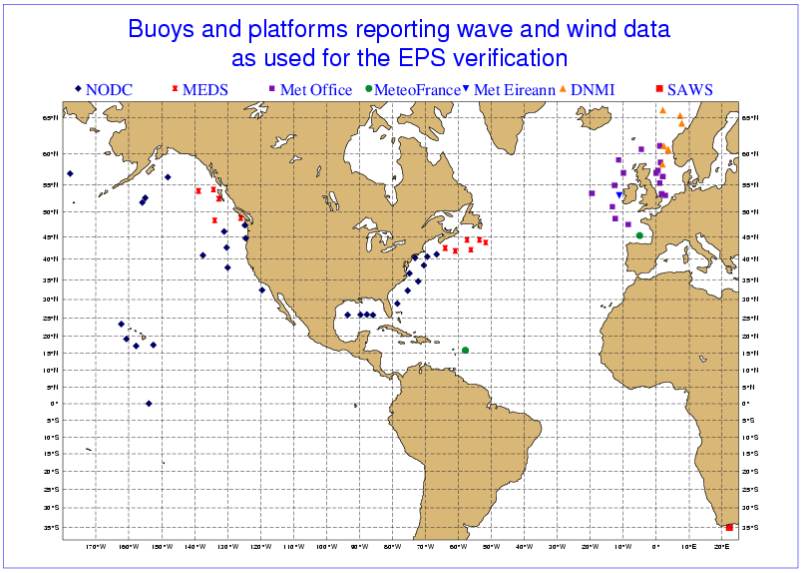

Figure 1 Map of locations for observations used in this study. The data providers are shown above the map. For details, see Table 1.

coast and one buoy on the equator near Christmas Island in the Pacific, all measurements are taken in the Northern Hemisphere. Since the majority of the buoys and platforms are located close to the continents, relatively few observations are obtained over the open oceans. To account for these shortcomings, the forecasts of significant wave height are additionally assessed against satellite altimeter observations, and the results compared with those obtained from the buoy and platform observations. The coverage of the altimeter data is shown in Figure 2. Since the present data-assimilation system for the wave model only uses satellite data, the buoy and platform measurements used here serve as an independent set of observations. This is not the case for the wind forecasts, because most of the observations have entered the data-assimilation system for the atmospheric model. The study covers the period from September 1st 1999 to March 31st 2002, thus including three full Northern Hemisphere winters.

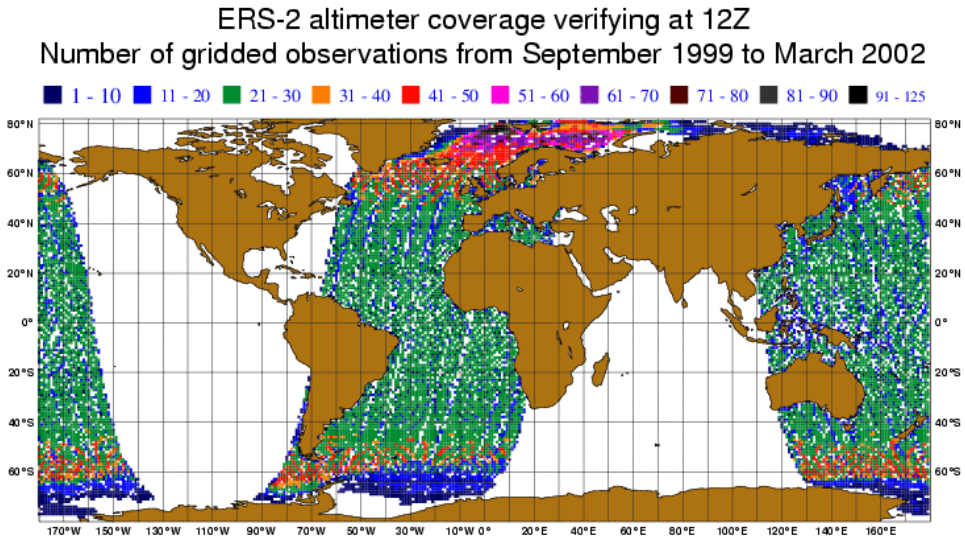

Figure 2 Map showing the distribution of gridded altimeter wave height observations. The colour coding indicates the number of observations.

The structure of this paper is as follows: Section 2 gives a brief description of the ensemble prediction system. Section 3 describes the buoy, platform and altimeter observations used in this study. The deterministic skill of the EPS is discussed in section 4. In Section 5, the ensemble spread is discussed and section 6 covers the relationship between the ensemble spread and the skill of the deterministic forecasts. In section 7, the reliability of the probability forecasts is considered. Section 8 investigates the economic value of the EPS. Section 9 compares the EPS with the altimeter observations, and finally, the conclusions are drawn in section 10.

2. Ensemble Prediction System

In 1998 the coupling between wind and waves was introduced operationally. The resolution of the atmospheric component of the EPS was $T_L159L31$ (spectral triangular truncation with 31 levels in the vertical), which is about 120 km horizontal resolution at mid-latitudes. On 21st November 2000, the new high-resolution EPS was introduced with T_L255, corresponding to approximately 80 km resolution in the horizontal. Note that this change in the EPS was carried out within the period spanned by this investigation. The most important changes introduced to the system during this period are reported by Buizza et. al (2002).

For a detailed description of the atmospheric component of the ECMWF EPS, the reader is also referred to Buizza et al. (2000).

The wave model component in the EPS is the ECMWF version of WAM cycle 4. WAM was developed during the 1980's by an international group of scientists (WAMDI-group 1987, Komen et al. 1994), and marked the introduction of a new generation of ocean wave models. This new model was the first model to solve the full energy balance equation including the non-linear transfer term for wave-wave interaction. Earlier generations of wave models solved the energy balance equation by introducing ad-hoc assumptions on the spectral shape. In some cases, particularly if the wind is changing rapidly, this assumption does not provide a proper description of the sea-state (Janssen et al. 1997). Since its implementation as operational wave model at ECMWF in November 1991, the model has undergone numerous changes and improvements Janssen 2000; Bidlot et al. 2002). A major improvement of the ECMWF forecasting system was made in June 1998, when the coupled atmospheric circulation and wave model system became operational. Traditionally in atmospheric models, the stress at the air-sea interface is calculated by the Charnock relation, where the surface stress is assumed to be only dependent on the local 10-meter wind. In reality however, the surface stress is to a large extent not dependent on local wind conditions alone, but also on the so-called wave age. That is, a young growing wind sea is more dominated by waves in the high-frequency part of the spectrum, whereas an old wind sea has most of its energy shifted to the low-frequent part of the spectrum. This fact has some important implications for the sea-surface roughness as experienced by the atmosphere. For a young wind sea, the surface will be rougher than an old wind sea, since momentum is extracted from the atmosphere more effectively. Therefore, some knowledge of the history of the wave field is needed. To take this into account, Janssen (1991) suggested coupling the atmosphere and wave model using a modified version of the Charnock relation by introducing a wave-induced stress. A significant positive impact of this, both on the atmospheric circulation and the wave field, has been documented (Janssen et al. 2001). As a direct consequence of this implementation in all components of the ECMWF forecasting system, ensemble wave forecasts became available on a daily basis.

The present version of the EPS wave model runs on a 110 km grid resolution with shallow water physics, 12 directional and 25 frequency bins. This resolution was put into operation on November 21st 2000, at the same time as the implementation of the new high-resolution EPS for the atmosphere. The previous version implemented in 1998, used a grid resolution of 1.5° with deep-water physics only. Thus, both the atmospheric and wave component of the EPS had a significant increase in model resolution during the period spanned by this investigation. It is likely that this will have some implication for this investigation, and will therefore be discussed later.

For the atmospheric component, 50 ensemble members are all initiated from the ECMWF analysis where perturbations have been introduced (Buizza et al. 2000), and are therefore labelled perturbed forecasts. A last member, denoted the control run, uses the unperturbed analysis as initial field interpolated to the EPS resolution. For the waves, all ensemble members use the unperturbed analysis as the initial condition. The divergence between the wave ensemble members is therefore due only to different wind forcing when the coupled atmospheric ensemble members are subject to different evolutions (Farina 2002). In this comparison, we have used all EPS forecasts initiated from 12Z. ECMWF has also been producing EPS forecasts from 0Z since 2001. The forecasts from 0Z will not be used in this study.

3. Observations

3.1 In situ

Sea state and ocean surface meteorological observations are routinely collected by several national organisations via networks of moored buoys and fixed platforms deployed in their near- and offshore areas of interest. The geographical coverage of the data is still very limited, and at the present EPS wave model resolution, only a small number of all these stations are within the wave model grid. Nevertheless, about 66 stations that report both wind and wave data can be selected. They are well within the grid of the wave model, in relatively deep water since the EPS wave model was originally set up as a deep water model, and have a high rate of data availability and reliability. Note that there are much more stations that only report wind but they will not be used in this study.

The wave and wind data are transferred continually via the GTS to national meteorological centres and are usually archived with all other synop ship observations even though they are not really ships. In the remainder of the paper, the word buoy will be used to refer to the selected moored buoys or platforms since most of the reliable observations come from moored buoys. Note however that the observation principle for waves is quite different for buoys than platforms. Buoys usually rely on time series analysis of the buoy motion to derive wave spectra whereas radar imaging of the sea surface is employed by platforms to derive the wave spectra. Collocations between these observations and the corresponding model values interpolated to the buoy locations can easily be obtained. A direct comparison between model values and buoy and platform observations is however undesirable as some measurements may still be erroneous. Furthermore, model and observed quantities represent different time and spatial scales.

From the buoy records, time series are reconstructed and used to perform a basic quality check on the data (Bidlot et al. 2002). This quality check procedure will only keep values that are within an acceptable physical range. It will also try to detect faulty instruments by removing all constant records of one day long or more. Finally, it will remove outliers by looking at the deviation from the mean of each data record and from the deviation from one hourly value to the next. Spatial and temporal time scales are made comparable by averaging the hourly observations in time windows of 4 hours centred on the synoptic times. An extra quality check was necessary for wave data from the Norwegian platforms. These wave data are obtained from a Miros wave radar mounted on the platforms. It appears that they do not retrieve realistic wave information under low wind speed situation and in a few other instances. For these stations a minimum wind speed is prescribed (around 5 m/s) and all wave observations obtained when the observed wind speed falls below this minimum are discarded. Furthermore, all wave heights above a maximum value Hs max are also removed from the monthly records. This maximum wave height is based on the saturation wave height due to wind:

$$H_{s\max} = 2 + 2 \cdot \left(0.22 \cdot \frac{U_{10}{}^{2}}{g} \right)$$

where U_{10} is the 10m wind speed and g is the acceleration due to gravity and the term in brackets is the saturation wave height for a given wind speed.

GTS data are unfortunately provided with some truncation. Wave heights are rounded to the closest 0.1 metre, wave periods to the closest second and wind speed to the closest meter per second. Averaging will diminish the effect of these truncations. The resulting errors for wave data are well within what can be

expected from buoy measurements (Monaldo 1988). It is however unfortunate that wind speed observations are encoded with such a large truncation error (up to 0.5 m s-1) since most of them still need to be adjusted to the standard height of 10m.

Buoy anemometers are not usually at an average height of 10 metres. However, the wind observations used here are supposed to represent the wind 10 meters above sea level. Therefore the height of the anemometers has been obtained from the data providers (Table 1) and the wind speed statistics were produced by adjusting the buoy winds to 10m. The wind speed is corrected assuming that on average the wind profile in the planetary boundary layer is neutral:

$$U_{10} = \frac{u_*}{\kappa} \ln(z / z_o)$$

where z is the height above the sea level, κ is the von-Karman's constant and u_* is the friction velocity. The surface roughness is given by Charnock's relation

$$z_0 = \alpha u_*^2 / g$$

where $\alpha = 0.018$, and g is the acceleration due to gravity. Combining these two equations, and using the height where the observation is taken for z, the surface stress is determined. Once the surface stress or the friction velocity is known, the corresponding corrected 10 meter wind speed can then be calculated. Winds from platforms are usually adjusted to 10m by the data providers. A reduction factor is used even though the height of the anemometer could be in the several tens of metres. Winds from platforms are therefore less reliable than buoy observations. The wind directions on the other hand, are assumed to be constant within the turbulent boundary layer, so no correction is done for this parameter. This may be true for the vast majority of the observations here. However, there are a few observations, taken from oil platforms, which are measured at more than 100 metres above sea level. It is likely that directional observations taken at such height, do not properly represent the wind direction at 10 metres.

Finally, statistics were compiled with the quality-controlled data supplemented with a blacklisting (omission) of a few data segments. The blacklisting of certain stations is done each month by collecting information from the data providers (web pages, e-mails...) and by analysing the monthly time series for any suspicious behaviour which has eluded the quality control. Roughly 46,000 wave height data were used. Besides wave height, buoys also report wave period measurements. There is however, no consensus on what type of period should be reported. As shown in table 1, Canadian and US buoys report the period corresponding to the peak in the one-dimensional wave spectrum, the peak period (Tp), whereas the other data providers use a mean period, usually the zero mean crossing period (Tz) which can roughly be equated to the normalised second moment of the frequency spectrum. The peak period has always been a standard output of the operational model, however Tz has only been archived in the operational EPS since the 27th October 2001. We will therefore only show results based on Tp.

ID	locations	latitude	longitude	Data provider	type	Anemo. height	period
41001	US East Coast, E Hatteras	34.68	-72.23	NDBC	6N	5 m	Tp
41002	US South-East Coast, S Hatteras	32.27	-75.42	NDBC	6N	5 m	Tp
41010	US East Florida , Cape Canaveral East	28.91	-78.55	NDBC	6N	5 m	Tp
41100	French West Indies (Antilles 1)	15.90	-57.90	Meteo-France	ODAS	4 m *	
42001	Mid Gulf of Mexico	25.92	-89.68	NDBC	10D	10 m	Tp
42002	Western Gulf of Mexico	25.90	-93.59	NDBC	10D	10 m	Tp
42003	East Gulf of Mexico	25.88	-85.95	NDBC	10D	10 m	Tp
42054	Mid central Gulf of Mexico	26.00	-87.73	NDBC	12D	10 m	Tp
44004	US North East Coast, Hotel	38.50	-70.47	NDBC	6N	5 m	Tp
44008	US North-East Coast, Nantucket	40.50	-69.43	NDBC	3D	5 m	Tp
44011	US North-East Coast, Georges Bank	41.08	-66.58	NDBC	6N	5 m	Tp
44014	US East Coast, Virginia Beach	36.58	-74.84	NDBC	3D	5 m	Tp
44025	US North East Coast, Long Island	40.25	-73.17	NDBC	3D	5 m	Tp
44137	Nova Scotia, East Scotia slope	41.83	-60.94	MEDS	6N	5 m	Tp
44138	Newfoundland, SW Grand Bank	44.26	-53.62	MEDS	6N	5 m	Tp
44139	Newfoundland, Banquerau	44.26	-57.39	MEDS	6N	5 m	Tp
44140	Newfoundland, Tail Of The Bank	43.70	-51.70	MEDS	6N	5 m	Tp
44141	Nova Scotia, Laurentian Fan	42.10	-56.22	MEDS	6N	5 m	Tp
44142	Nova Scotia, La Have Bank	42.50	-64.02	MEDS	6N	5 m	Tp
46001	Gulf of Alaska	56.30	-148.17	NDBC	6N	5 m	Tp
46002	US West Coast,Oregon	42.57	-130.32	NDBC	6N	5 m	Tp
46003	Aleutian Peninsula	51.83	-155.85	NDBC	6N	5 m	Tp
46005	US North-West Coast, W Astoria	46.05	-131.02	NDBC	6N	5 m	Tp
46006	US West Coast, SW Astoria	40.84	-137.49	NDBC	6N	5 m	Tp
46035	Bering Sea	56.91	-177.81	NDBC	12D/6N	5 m	Tp
46036	Canada West Coast, South Nomad	48.35	-133.94	MEDS	6N	5 m	Tp
46041	US North-West Coast, Cape Elisabeth	47.34	-124.75	NDBC	3D	5 m	Tp
46047	US South-West Coast, Tanner Banks	32.43	-119.53	NDBC	3D	5 m	Tp
46050	US West Coast, Yaquina Bay	44.62	-124.53	NDBC	3D	5 m	Tp
46059	US West Coast, California	37.98	-130.00	NDBC	6N	5 m	Tp
46066	Gulf of Alaska, South Aleutians	52.65	-155.00	NDBC	6N	5 m	Tp
46184	Canada West Coast, North Nomad	53.91	-138.85	MEDS	6N	5 m	Tp
46205	Canada West Coast, W. Dixon Entrance	54.16	-134.28	MEDS	3D	5 m	Tp
46206	Canada West Coast, La Perouse Bank	48.84	-126.00	MEDS	3D	5 m	Tp
46208	Canada West Coast, West Moresby	52.52	-132.68	MEDS	3D	5 m	Tp
51001	Hawaii North West	23.40	-162.27	NDBC	6N	5 m	Tp
51002	Hawaii South West	17.15	-157.79	NDBC	6N	5 m	Tp
51003	Hawaii West	19.16	-160.74	NDBC	6N	5 m	Tp
51004	Hawaii South East	17.44	-152.52	NDBC	6N	5 m	Tp
51028	Christmas Island DWA	0.00	-153.88	NDBC	3D	5 m	Tp
62001	Gulf of Biscay, Gascogne	45.20	-5.00	Meteo-France	ODAS	4 m *	Tz
62026	North Sea (K17)	55.30	1.10	Met Office	ODAS	4 m	Tz
62029	UK Celtic Sea shelf break (K1)	48.70	-12.40	Met Office	ODAS	4 m	Tz
62081	UK East Atlantic (K2)	51.00	-13.30	Met Office	ODAS	4 m	Tz
62090	West Ireland (M1)	53.10	-11.20	Met Eireann	ODAS	4 m	Tz
62105	UK East Atlantic (K4)	54.90	-12.60	Met Office	ODAS	4 m	Tz
62106	UK North-East Atlantic (RARH)	57.00	-9.90	Met Office	ODAS	4 m	Tz
62108	UK East Atlantic (K3)	53.50	-19.50	Met Office	ODAS	4 m	Tz
62109	North Sea (K16)	57.00	0.00	Met Office	ODAS	4 m	Tz
62112	North Sea (Brae A)	58.70	1.30	Met Office	platform	120 m *	
62132	North Sea (Auk A)	56.40	2.00	Met Office	platform	103 m *	Tz
62144	North Sea (Clipper)	53.40	1.70	Met Office	platform	70 m *	Tz
62145	North Sea (Sean P)	53.10	2.80	Met Office	platform	70 m *	Tz
62162	North Sea (Kittiwake)	57.40	0.50	Met Office	platform	97 m *	Tz
62163	UK Celtic Sea shelf break (Brittany)	47.50	-8.50	Met Office	ODAS	4 m	Tz
63103	South Norwegian Sea (North Cormorant)	61.20	1.10	Met Office	platform	102 m *	
64045	UK North-East Atlantic (K5)	59.10	-11.40	Met Office	ODAS	4 m	Tz
64046	UK North-East Atlantic (K7)	60.70	-4.50	Met Office	ODAS	4 m	Tz
LF4B	Southern Norwegian Sea (Troll A)	60.60	3.70	DNMI	platform	*** m *	
LF3F	Northern Norwegian Sea (Draugen)	64.30	7.80	DNMI	platform	78 m *	
LF3J	Southern Norwegian Sea (Gullfaks)	61.20	2.30	DNMI	platform	142 m *	
LF3N	Northern Norwegian Sea (Heidrun)	65.30	7.30	DNMI	platform	131 m *	
LF4C	North Sea (Sleipner)	58.40	1.90	DNMI	platform	136 m *	
3FYT	Southern Norwegian Sea (Troll)	60.70	3.60	DNMI	platform	91 m *	
LDWR	Northern Norwegian Sea (Mike)	66.00	2.00	DNMI	platform	15.5 m *	
ZSWAV	South Africa Agulhas Bank	-35.00	22.20	SAWS	platform	113 m *	

* The winds are adjusted to 10m before distribution to the GTS.

Table 1 Buoys and platforms that provide wave and wind data onto the GTS. The data are managed by the data providers: the US National Data Buoy Center (NDBC), the Canadian Marine Environmental Data Service (MEDS), the Met Office, the French meteorological service (Meteo-France), the Irish meteorological service (Met Eireann), the Norwegian meteorological institute (DNMI) and the South African Weather Service (SAWS). Data are collected from 6- meter NOMAD buoys (Navy Oceanographic Meteorological Automatic Device), from 12-, 10- and 3-metre discuss buoys, from UK ODAS buoys (Ocean Data Acquisition System) and from platforms. The height of the anemometers is given and the star (*) indicates that the wind speed data were adjusted to 10m before dissemination onto the GTS. The GTS report can only contain one measure for the wave period. It is usually the peak period (Tp), or the mean zero-crossing period (Tz) or some other mean period.

3.2 Wave data from satellite measurements

Wind and wave data are also available from the radar altimeter on board of the European satellite ERS-2. ECMWF receives the fast delivery ERS data in near real time and archives them. The wave height observations from ERS-2 are of relativly good quality and have been used by the wave model data assimilation since May 1996. Nevertheless, Janssen (2000) showed that the altimeter wave height might actually be slightly too low, especially in situations where waves are steep. A scheme for the correction of altimeter wave heights due to the non-gaussianity of the sea surface elevation and slopes distribution was introduced in the ECMWF operational wave model in July 1999. Deviations from a Gaussian distribution are measured by the skewness factor and the elevation-slope correlation, which depend in a complicated way on the wave spectrum. The wave model spectra are used to estimate these two quantities to derive a correction to the ERS-2 wave height data. Even though altimeter data are processed observations and thus not as such wave model results, their processing will require some information from the model. Before correcting the altimeter wave heights, the data are pre-processed by running them through a quality control procedure which is very similar to the one used for the buoy data except that the processing runs along 30 consecutive data points following the satellite track. A few quality indicators provided with the data are also used to discard suspicious data points. The valid individual altimeter wave height data, which are available in a +- 3 hours time window centred around the main synoptic times, are collocated to the closest model grid point. The average value is computed for all grid boxes with at least two individual observations. The mean position is assumed to coincide with the model grid point and the time at the centre of the time window is taken as the verifying time. These mean values are then corrected using wave model spectra available just before assimilation of the altimeter data. Both averaged data sets are then archived on the same grid as all operational analysis wave model fields (roughly with a 55 km resolution) with the time stamp of the analysis.

Because these corrected wave data are averaged on the wave model grid of the analysis, it is trivial to find the gridded values that are the closest to the EPS grid points and to retrieve the corresponding EPS wave heights for all output steps considered in this study (0,24,48… hours). The data coverage is presented in Figure 2. Note that the altimeter data were averaged over the operational analysis grid of roughly 55 km and not on the EPS grid. Nevertheless, 310,000 gridded observations were obtained. Fast delivery wave heights for low wave heights (below ~ 1.5m) are known to be overestimated (Challenor and Cotton 1997), however, no corrective fit is used here to remove this inconsistency because the comparison between altimeter wave heights and the EPS is intended to focus on high waves.

The altimeter wind speeds are not yet processed onto the wave model grid and are not therefore used in the verification. Furthermore, the ECMWF verification of the ERS-2 fast delivery wind speed product has identified a few periods during which the quality of the retrieved winds had degraded. It was therefore decided not to use the data.

4. Deterministic skill of the EPS

Before the probability forecasts are assessed, we will investigate the deterministic skill of the EPS. This will serve as a reference when judging the performance of the probabilistic forecasts. The deterministic performance of the EPS members, such as the root-mean-square errors and model bias, has important implications on both spread and reliability of the ensemble forecasts (Hamill 2001).

The ensemble forecast can be converted into a deterministic forecast by calculating the ensemble mean, and treating this as the predicted value. Although this quantity has some severe shortcomings when used as a forecast value, its bias and root-mean-square scores may still give valuable insight to the overall behaviour of the EPS system. Figure 3 shows the ensemble root-mean-square error (RMSE) of the control forecast and the ensemble mean for forecast times from 0 to 240 hours, calculated over the whole period and for all buoy data. Here, values for the significant wave height and the peak period are given. The RMSE has been calculated according to

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x_n^f - x_n^o\right)^2}$$

where x_n^f and x_n^o denotes the forecast and the observed value respectively, and N is the total number of data pairs (Wilks 1995). Since all wave ensemble-members are started from the same initial condition, the RMSE for the ensemble mean and the control forecast must be equal for this time step. As the forecast time increases however, the ensemble mean is, as expected, performing better than the control forecast for both parameters shown in this plot.

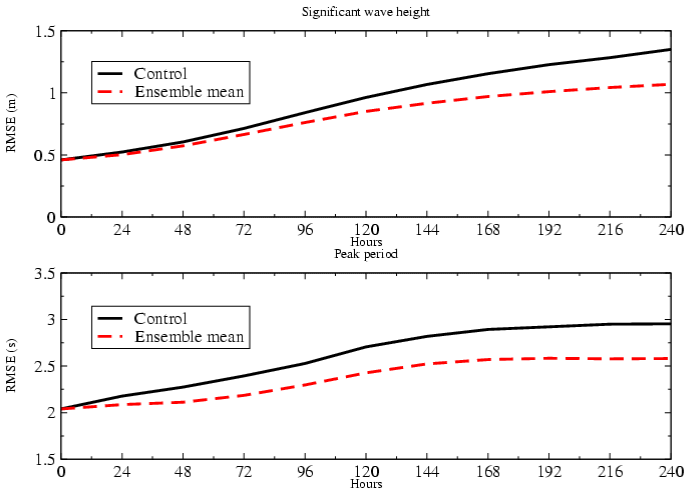

Figure 3 Root-mean-square errors for significant wave height and peak period for all buoy data. The peak-period statistics are for the US and Canadian buoys only.

Figure 4 shows the performance of the 10-meter wind-speed and direction for the control run and the ensemble mean. The ensemble mean direction is defined as the angle of the mean direction-vector, where the mean is found by averaging over all ensemble members. Disregarding the magnitude of the wind vector and taking α as the wind angle relative to the north, the mean direction-vector over N members is then defined as $\bar{V} = \bar{u}\,\mathrm{i} + \bar{u}\,\mathrm{j}$, where

$$\bar{u} = \frac{1}{N}\sum \sin\alpha$$

$$\bar{v} = \frac{1}{N}\sum \cos\alpha\,.$$

Since the wind-speed and direction are output parameters from the atmospheric model, they have a spread also at the initial time step. However, the differences between the control and ensemble mean are too small to be noticeable in these graphs. The ensemble mean performs better than the control forecast in terms of RMSE for all four parameters when compared to the buoy observations used here.

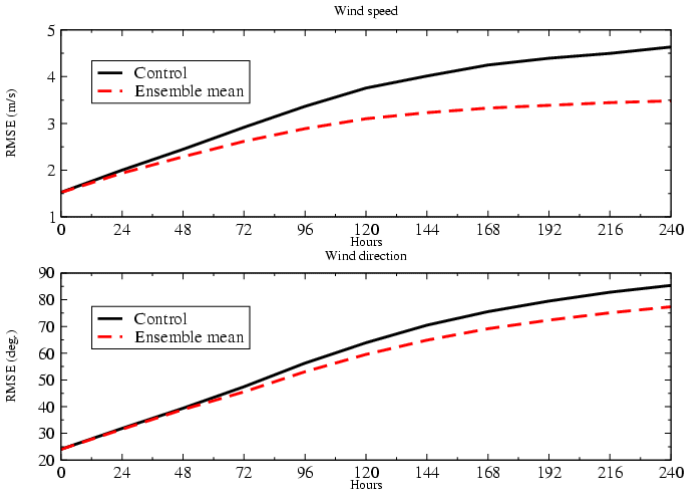

Figure 4 Root-mean-square errors for wind-speed and direction for all buoy data.

For the model bias, the mean errors have been calculated as the mean difference between the model and the observations, such that a positive value would indicate a tendency for the model to be above the observations and vice versa. Figure 5 shows the results for the wave height and the peak period. The bias of the wind-speed and directions are depicted in Figure 6. The bias in the wind direction is calculated as the angular differences between the forecast and the observations, a positive bias indicates that the model has a tendency to be to the right of the observations, and a negative value is to the left. On average, the analysed wind-direction is found to be 3 degrees to the left of the observations. Interestingly, the bias in wave height and wind speed forecasts from the ensemble mean almost vanishes beyond day two. This is not the case for the control forecasts.

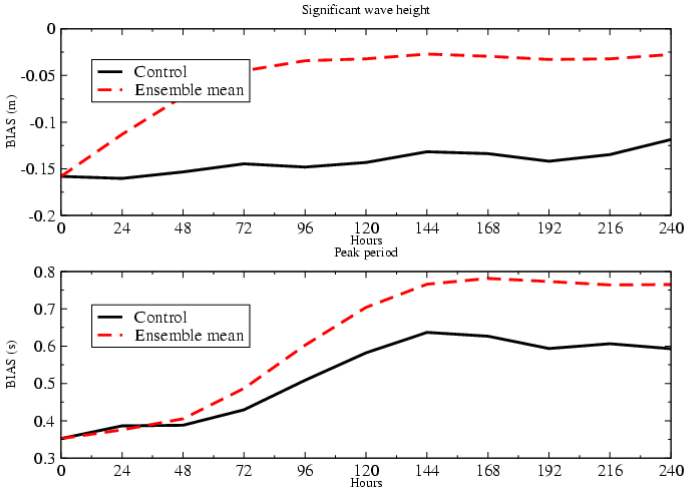

Figure 5 Forecast bias in significant wave height and peak period for the control and ensemble mean forecasts for all buoy data. The peak-period statistics are for the US and Canadian buoys only.

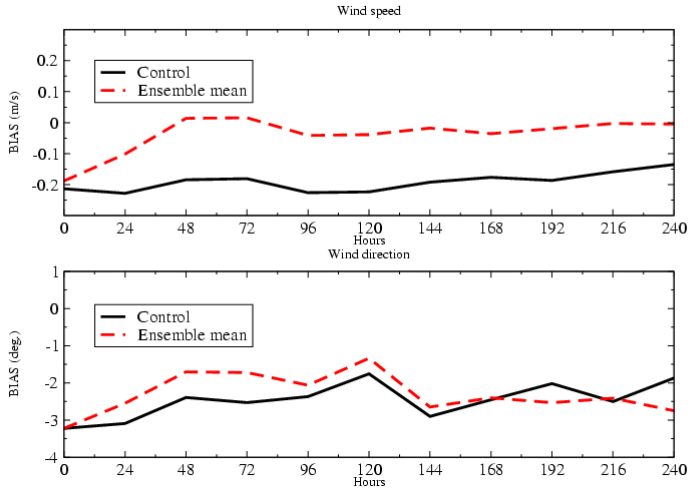

Figure 6 Forecast bias in wind-speed and direction for the control and ensemble mean forecasts for all buoy data.

An example of the seasonal variations in the RMSE is shown in Figure 7, where the monthly mean values for day 5 forecasts of significant wave height and wind-speed are depicted. The scores for two randomly picked ensemble members, number 12 and 34 in this case, are also shown together with the similar scores for the control run and the ensemble mean. The RMSE are lowest during summer and highest in the period from December to April for both parameters.

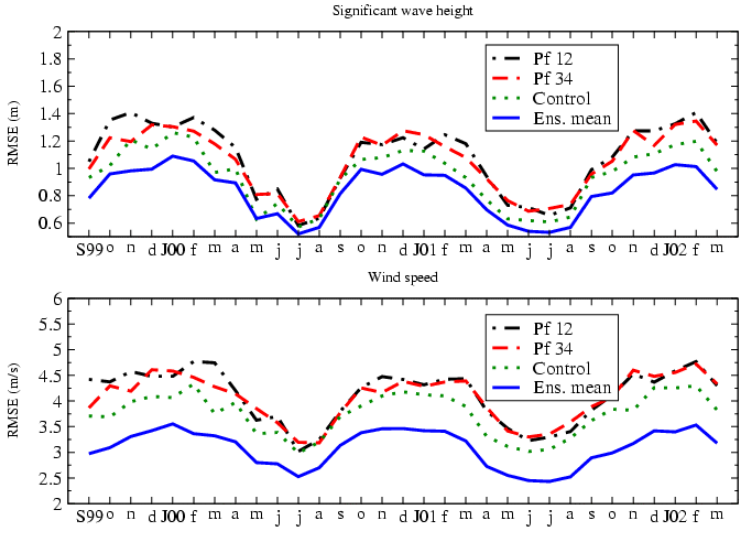

Figure 7 Monthly RMSE for wave height and wind speed day 5 forecasts. The plot depicts the scores of two of the ensemble members (number 12 and 34, chosen at random) together with the results for the control forecast and the ensemble mean. The x-axis starts at September 1999 (S99), and ends in March 2002. All buoy data were used.

It is also interesting to see that the control forecasts perform better than the two random members. This may be a bit surprising, since one might tend to believe that the unperturbed forecast could be regarded as just another ensemble member, and hence expect it to perform as such. As these results indicate, this is probably not the case. The control run is the model forecast based on the unperturbed atmospheric analysis, which is

the most probable state of the atmosphere at the analysis time. All the other ensemble members are initiated from an atmospheric state where perturbations have been introduced. Although the errors are within the uncertainties of the observations used for the analysis, the resulting perturbed initial conditions are possible, but less likely states of the initial atmosphere. Hence, the control forecast should have a better chance of forecasting correctly than a randomly chosen EPS-member. In contrast to the control forecast, they are also run with stochastic physics in the atmospheric component (Buizza et al. 2000).

The average monthly model bias at day 5 for wave height and wind speed is shown in Figure 8. Again, the ensemble mean performs better than the control forecast in terms of bias. Maybe more surprising is that the two individual ensemble members are also doing better than the control. As well as having different initial conditions, the atmospheric component of the ensemble members are run with stochastic physic. This fact is probably the cause for the reduced bias in both ensemble mean and individual ensemble members.

Neither of the figures containing monthly mean values (Fig. 7 and 8) reveals any dramatic changes in connection with the system upgrade in November 2000. One may perhaps argue that the curves for the monthly mean bias for wave height are somehow smoother in the period after the system change, but the difference are not very large, and longer series will probably be needed to determine any trend caused by this upgrade.

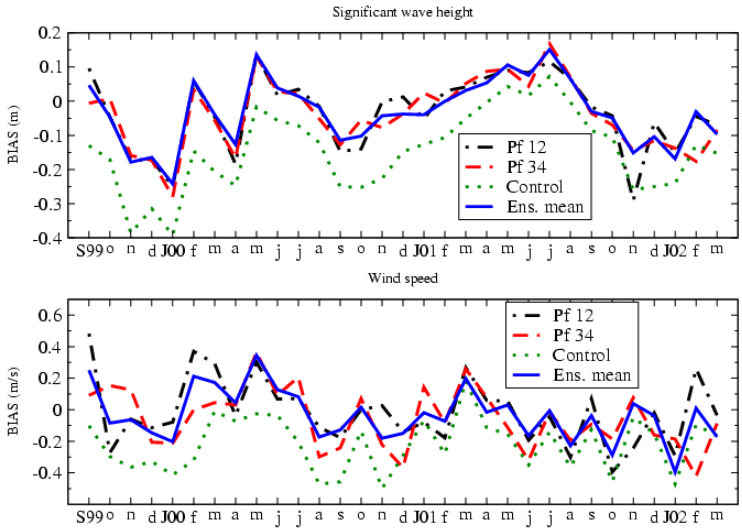

Figure 8 Monthly bias for wave height and wind speed day 5 forecasts. The plot depicts the scores of two of the ensemble members (number 12 and 34) together with the results for the control forecast and the ensemble mean. The x-axis starts on September 1999 (S99), and ends in March 2002. All buoy data were used.

In the results presented above, the ensemble mean was performing best in most cases. It is still important to stress that producing a deterministic prediction by averaging over all members, is not necessarily a recommended way of issuing a forecast from an ensemble. Even though such a forecast performs best when the average is taken over a sufficient number of cases, it is likely that the performance of this mean would be rather poor when trying to forecast rare or extreme events. Being an average, the ensemble mean tends to be much smoother than a single deterministic forecast. Hence, events that deviate much from the mean are less likely to be represented.

5. Ensemble spread

In July 2000, some changes were made to the data assimilation at ECMWF, and unintentionally the scaling factor used to initialise the perturbed ensemble members became too small. The problem was resolved in January 2001, but during this six month period the ensemble spread was too low. This is most apparent in the day 3 forecasts. In Figure 9, the monthly frequencies of observations outside the ensemble range for wave height and wind speed at day 3 are shown. The time of introduction and removal of the error in the initialisation of the system is marked by two vertical dashed lines. For the wind speed, a sudden increase in the number of observations outside the ensemble forecasts are observed in July 2000, when the error was introduced. After January 2001, when the bug was removed, the spread returned approximately to the previous level. Changes in the ensemble spread for the wave height can also be detected in this period, but the signal is weaker for this parameter. In Figure 9, the monthly mean Brier Scores are also plotted. This is essentially the mean-squared error for a probability forecast with 0 for a perfect forecasting system and 1 as upper bound (Wilks 1995). For waves, the graph corresponds to the probability forecasts of wave height above 2 meters. For the wind speed, the threshold value is 10 m/s. No particular differences in the Brier Score are detected during this period. Neither is it possible to detect any changes associated with the system upgrade in November 2000. As for the monthly mean values in Figure 7 and 8, we believe that longer series are needed to determine possible trends in connection with changes to the system.

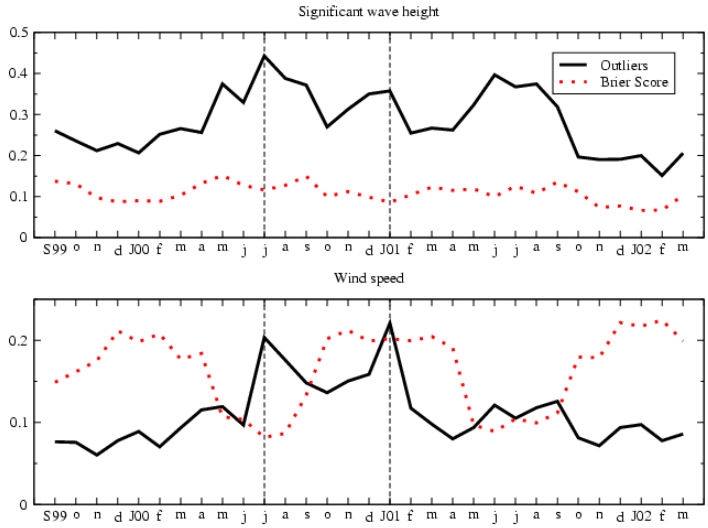

Figure 9 Monthly mean values of Brier score (red dotted line) and frequency of observation outside the ensemble range (black solid line) for the day 3 forecasts. The Brier scores are for probabilities of waves above 2 metre, and wind speed above 10 m/s respectively. The period with a bug in the EPS initialisation is between the two vertical dashed lines.

One common method to asses the ensemble spread is based on the so-called rank histograms, where the frequency distribution for the rank of the observation is visualised in a bar plot (Anderson 1996; Hamill 2001). As was demonstrated by Saetra et al. (2002), the effect of uncertainties in the observations can cause an impression of too low ensemble spread when tested using rank histograms. This can however be compensated by adding noise to the ensemble members, the random noise should have the same distribution as the observation errors. For wave observations, the relative error is estimated to be approximately 12%, as determined by a triple collocation exercise between buoy, altimeter and model (Saleh Abdalla - Personal

communication), and a somewhat lower value for the wind observations. Figure 10 and 11 show the rank histograms for waves and wind when normally distributed noise has been added to the ensemble members. For waves, the standard deviation of the noise was taken to be 12% of the forecasted value for each ensemble member. For the winds, the standard deviation was taken to be 10% of the forecast values.

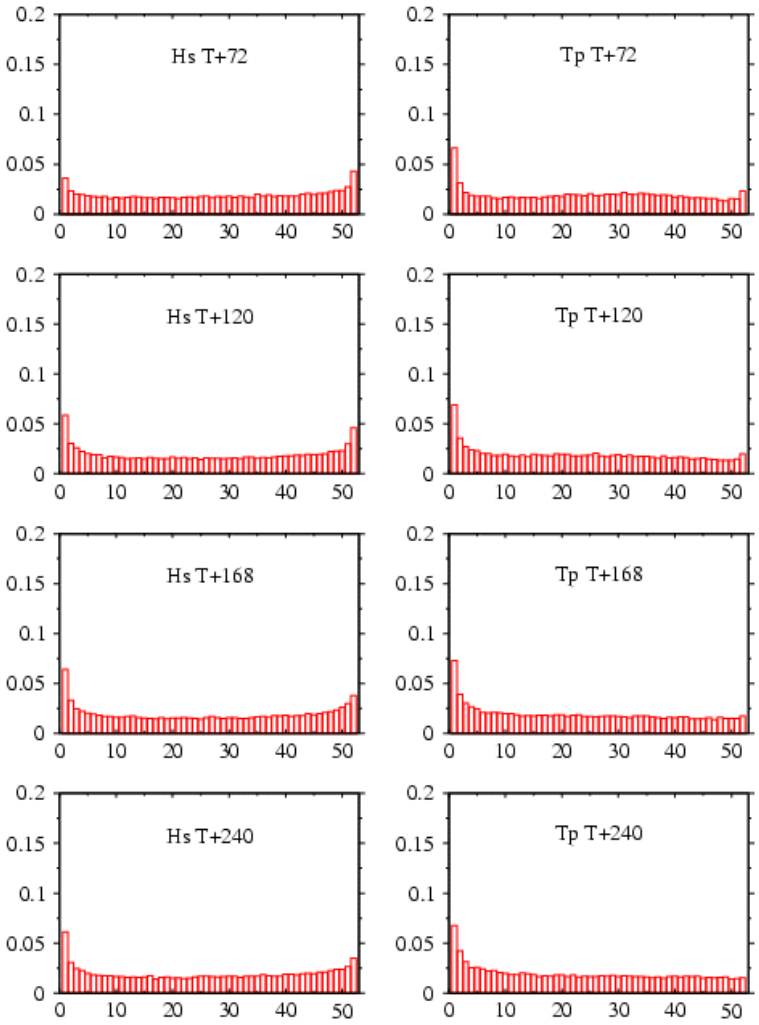

Figure 10 Rank histogram for wave height (left) and peak period (right) when compared with buoy data. Note that Tp is only reported by US and Canadian buoys (see table 1). Forecast steps are 72, 120, 168 and 240 hours. The period covered spans from September 1999 to March 2002.

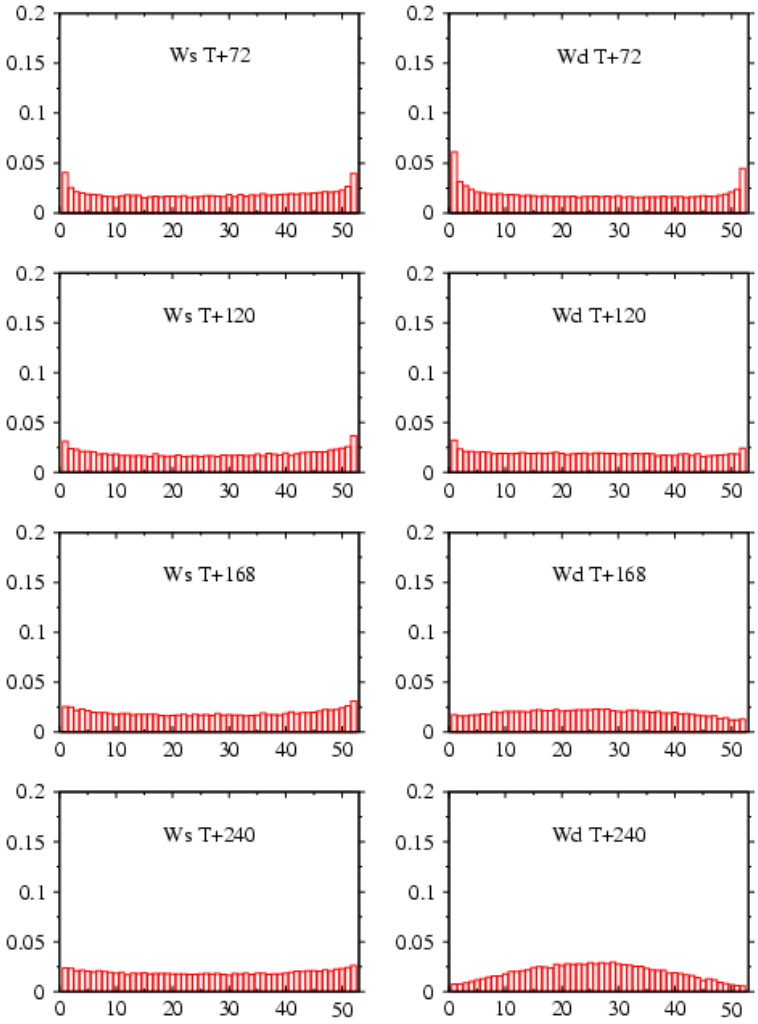

Figure 11 Rank histogram for wind-speed (left) and direction (right) when compared to buoy data. Forecast steps are as in Figure 10.

6. Relation between ensemble spread and skill

For practical purposes, the ensemble spread should somehow be judged in relation to the forecast skill. We would like to be able to interpret the ensemble spread as a measure of the uncertainty of the corresponding deterministic forecast in such a way that we are more confident in the forecast when the ensemble spread is small than when it is large. Ideally, the smaller the spread, the more we should be able to trust that the deterministic forecast is good. However, it is not obvious how this relationship between spread and skill can be tested. We want the forecast errors to be small when the ensemble spread is low, but we may accept small errors even if the ensemble spread is large. A reasonable expectation is that there is an increasing probability of large errors as the spread increases. This can be tested by using the 90-percentile of the absolute errors as a measure of a statistical error bound. For a given spread, we are seeking the value that separates the 10% largest errors from the rest of the data. As a result, if a situation is picked randomly from this data set, the probability is 90% that the corresponding absolute error is smaller than the value given by the 90-percentile. Since the spread is also a stochastic parameter, it must be treated in an equal way by using percentiles. Here, the ensemble spread will be defined as the difference between the upper and lower quartiles of the ensemble.

To relate the percentiles of the errors to the ensemble spread, it is necessary to divide the spread into different classes or bins, and then ranking the observed errors within each class to find the value that constitutes the boundary between the 10% largest errors and the rest.

In Figure 12, the 90-percentile of the absolute error for significant wave height is given as a function of the ensemble spread for the day 5 forecast range. The absolute error is defined as the distance between the observed value and the control forecast. The equivalent results for wind speed are given in Figure 13. For both wave height and wind speed, the 90-percentile shows a clear dependency on the ensemble spread. In fact, the correlation coefficients for these two cases are 0.966 and 0.989 for wave and wind respectively. The data have also been divided in different areas (not shown here) and seasons with different characteristics, and hence different variability. Nonetheless, the spread-skill relationship calculated here does not show any regional or seasonal dependencies.

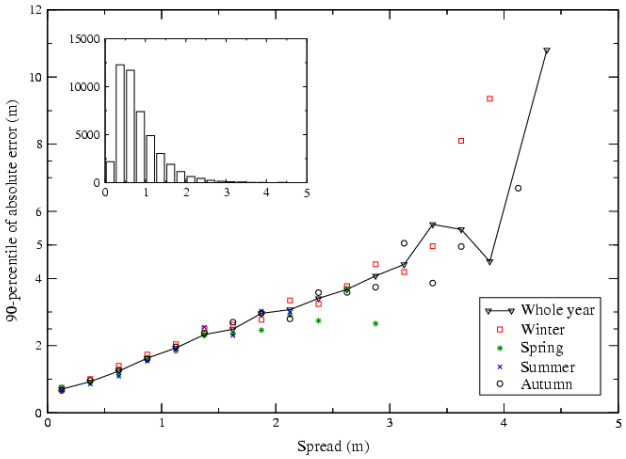

Figure 12 Day 5 forecast spread-skill for wave height, using the 90-percentile of forecast errors. The histogram in the upper left corner shows the frequency distribution of the spread bins. All buoy data are used. The black solid line are the results when all available data are taken into account; the triangles mark the centre point of each bin for spread. The number of cases that have been used for each bin is indicated by the histogram in the upper left corner. The red squares, green stars, blue crosses and black circles denote the results for winter, spring, summer and autumn respectively. Winter is defined as December, January and February, spring as March April and May, summer as June, July and August and autumn as September, October and November.

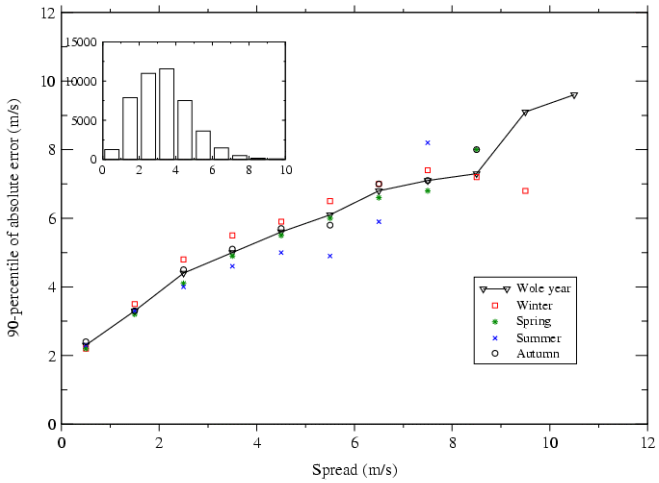

Figure 13 Day 5 forecast spread-skill for wind speed, using the 90-percentile of observation errors. The histogram in the upper left corner shows the frequency distribution. All buoy data are used. See Figure 12 for details.

Of course, the choice of percentile for the observed errors in this case is more or less arbitrary; any other percentile is expected to give qualitatively similar results. That is, except maybe for a percentile close to 100, for which the percentile may be very sensitive to a few outliers. In Figure 14, the relationship between the spread defined as the inter quartile range, and the 75, 80, 85 and 90-percentile for significant wave height is given. Again, the forecast range in the example is day 5. Figure 15 shows the similar results for the wind speed. Note that the gradient is steeper for the waves than for the wind speed. For the wave height, the 75-percentile fits roughly with the diagonal line. A very approximate rule of thumb may then be that the error in the wave forecasts is expected, with 75% probability, to be less than or equal to the inter-quartile range of the wave ensemble.

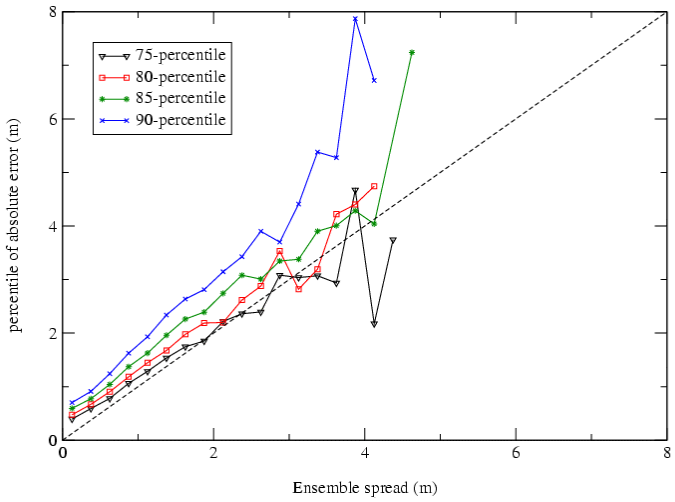

Figure 14 Day 5 forecast spread-skill for wave height, using different percentile as upper bound to errors. All buoy data are used.

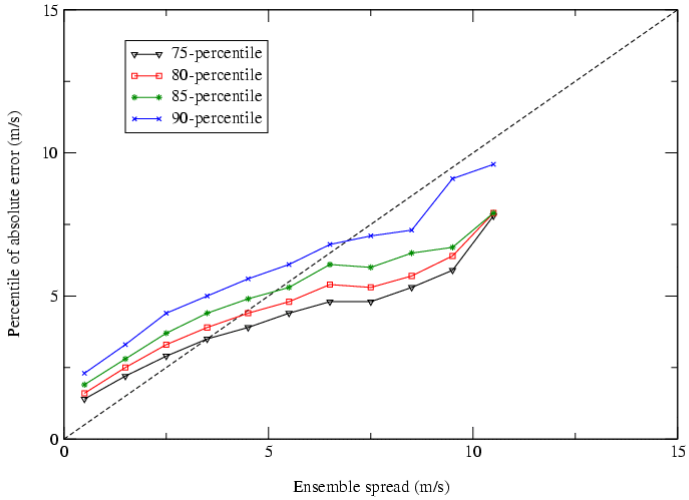

Figure 15 Day 5 forecast spread-skill for wind speed, using different percentile as upper bound to errors. All buoy data are used

7. Reliability of the probability forecasts

In Figure 16, the reliability diagrams (Wilks 1995) for the day 5 forecast probabilities of wave above 2, 4, 6 and 8 m are plotted. For a given event, the forecast probabilities are split into discrete bins ranging from zero to one. For each probability class, the fraction of times the event is observed (with respect to the total number of ensemble forecasts in that class), defined as the observed frequency, is plotted against the corresponding probability. For a perfectly reliable forecasting system, these points lie on the diagonal line. The plots also display the overall Brier score for each event. The graphs shown here are based on the results obtained without adding any noise to the ensemble members to correct for observational errors, as it had to be done for the rank histograms. The results were obtained for both cases, and only minor differences could be detected. Generally, the results indicate good reliability, particularly for the 4 m threshold. For threshold values of 2 and 6 m, the reliability is also quite good, but there is a small tendency for the points to lie below the diagonal line, which indicates that high probabilities are forecasted slightly too often. For the 8 m threshold, the reliability curve shows the behaviour typical for situations with insufficient sample size. Reliability diagrams for wind speed are given in Figure 17. Here, the threshold wind speeds are 10, 14, 17 and 20 m/s. As for the waves, the model apparently has a small tendency to over-forecast, in particular for the highest wind speed thresholds.

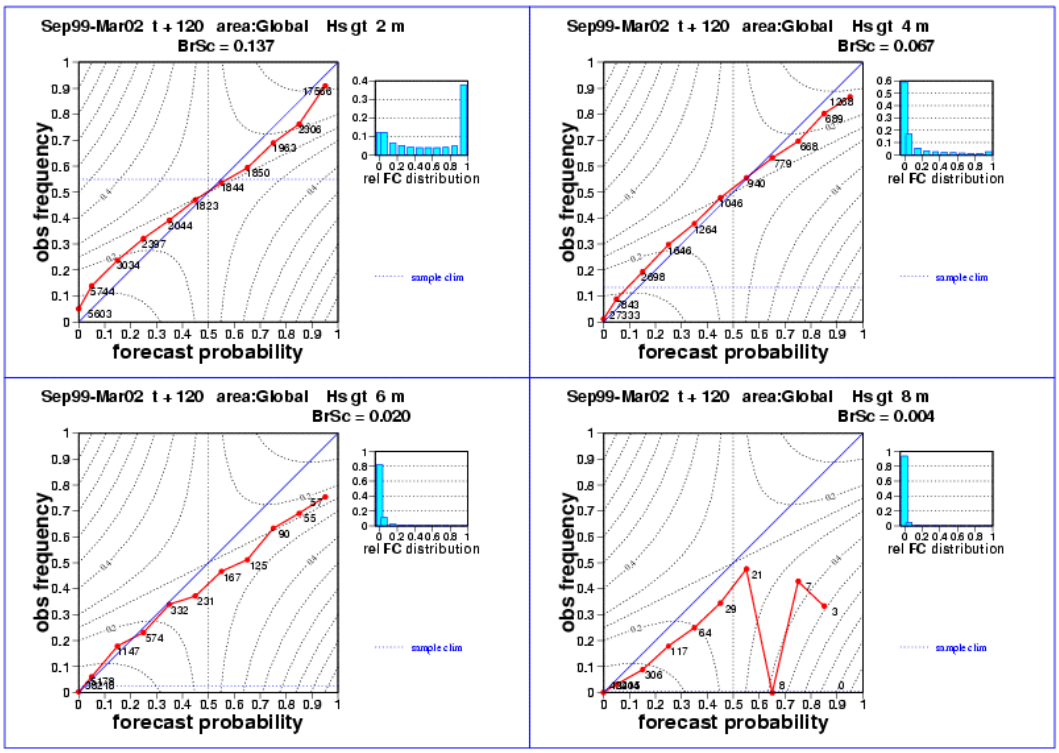

Figure 16 Day 5 reliability diagram for wave height. BrSc stands for Brier Score (see text). All buoy data were used.

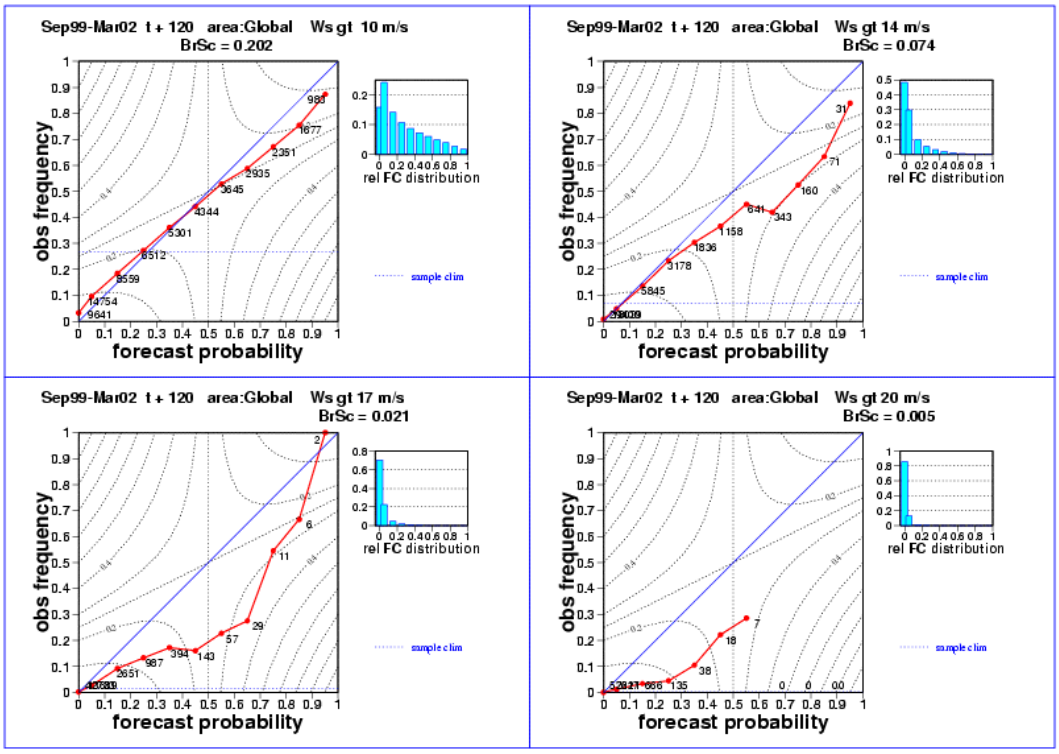

Figure 17 Day 5 reliability diagram for wind speed. BrSc stands for Brier Score (see text). All buoy data were used.

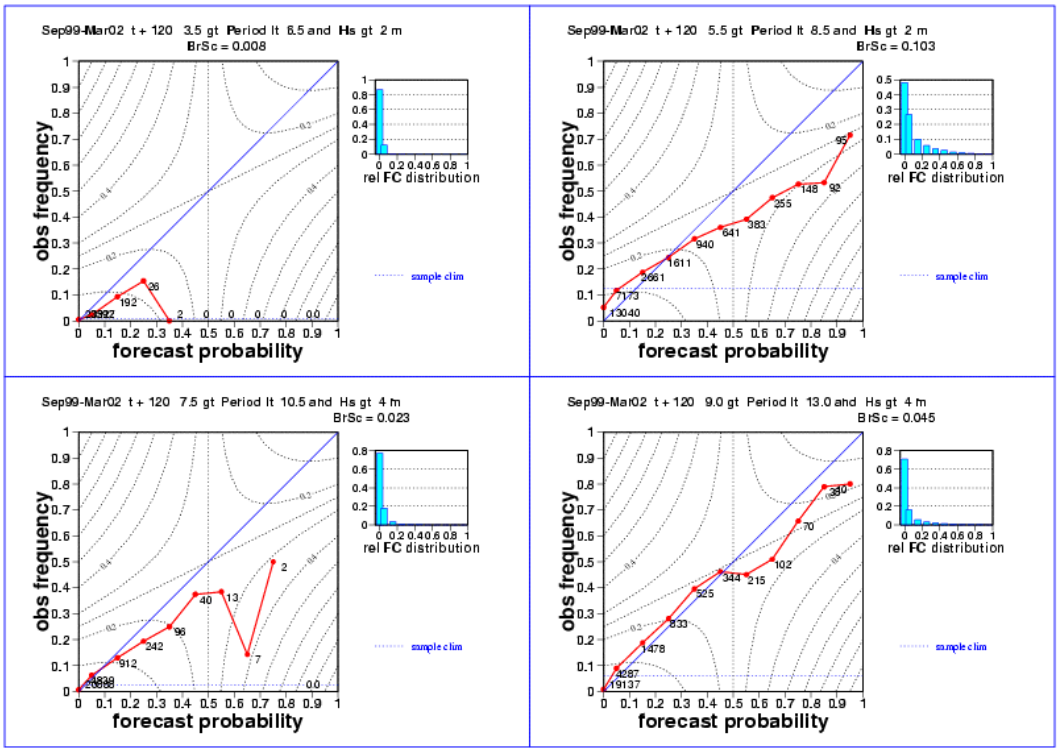

Figure 18 Day 5 reliability for the joint probability of significant wave height and peak period. BrSc stands for Brier Score (see text). All buoy data were used.

In Figure 18, the day 5 forecast reliabilities of the joint probability of wave height and period are given. Here, two relatively low threshold values of 2 and 4 m for wave height have been used. For 2 m wave height, the intervals are for periods between 3.5 s and 6.5 s and between 5.5 s and 8.5 s. The reliability diagram for the first of these is typical for rare events, but with relatively good reliability. The second case is much more common as the reliability curve also reveals. For 4 m threshold wave height, the periods are between 7.5 s and 10.5 s for the first case, and between 9.0 s and 13.0 for the second case. For this threshold value, the first case shows the behaviour typical for rare events with relatively good reliability. The second case also shows quite good reliability, but this is a much more common situation.

8. Economic value of the EPS

It is important to assess the economic value of the ensemble forecasts. In many operations involving weather related risks, the decision on whether to carry out the operation or not must at some point be taken, while the potentially dangerous part of the operation may lie several days into the future. For instance when an oil rig is to be towed, a perilous part of the operation is the installation of the platform on the operation site, in some cases many days after the onset of the operation. In such cases, ensemble forecast should provide valuable information.

Richardson (2000) suggests a method for estimating the relative economic value of weather forecasts, including ensemble forecasts. This method is also well suited for comparing the relative value of the ensemble forecasts with that of traditional deterministic forecasts. The method assumes a situation where a person has to decide on whether to take action to avoid a weather-related risk or not. In this case, we can imagine the situation of an oil rig that is to be towed to the drilling site. If the probability of waves above the dangerous threshold is considered too high, action can be taken to postpone the operation in order to prevent a potential loss L. In this example, taking action involves costs C, associated with the delayed operation. If L_o is the part of the potential loss that is saved by taking action, the cost-loss ratio is defined as $\alpha = C/L_o$. Now, for the forecast, four different outcomes are possible, and may be expressed by a contigency table as shown in Table 2. In this table, a, b, c and d are the frequencies of these outcomes when the forecasts are compared with observations. The hit rate H and the false alarm rate F are then defined as

$$H = a/(a+c)$$

$$F = d/(d+b)$$

	Event observed	**Event not observed**
Event forecasted	*a*	*b*
Event not forecasted	*c*	*d*

Table 2 Contigency table showing the relative frequencies of the four possible outcomes of forecasting a specific event.

The hit rate is the fraction of occurrences correctly forecast and the false alarm rate is the fraction of non-occurrences that was incorrectly forecast. In the simple cost-loss model, the total cost of using the forecast

can now be computed and compared with the hypothetical costs if the decisions were based on either the sample climate or perfect forecasts. The relative economic value is defined as

$$V = \frac{E_{clim} - E_{forecast}}{E_{clim} - E_{perf}}$$

where $E_{forecast}$, E_{perf} and E_{clim} are the expected expenses when using the actual forecast, a perfect forecast or the sample climate to make the decision respectively. According to this, a perfect forecast will score 1 and a forecast that does not perform better then the sample climate will score 0.

For a probabilistic forecast, a probability level for whether to take action or not must be chosen. Richardson (2000) shows that if the forecasting system is reliable, maximum benefit will be obtained if the probability p=a is used. In Figure 19, the relative economic value of the ensemble wave forecast and a deterministic forecast, represented here by the control forecast, is given as a function of the cost-loss ratio. In addition, the EPS has been compared to the poor-man's ensemble (PME), which was constructed by adding normally distributed noise to the control forecasts. The standard deviation used for this is 0.96 m, which is the root-mean square error for the day 5 forecasts for waves. These later figures correspond to the reliability diagrams in Figure 16, and show the results for the day 5 forecast for waves above 2, 4, 6 and 8 m. In the curves for the EPS in this case, the appropriate probability level has been found by calculating the expression for a discrete set of probabilities ranging from 0 to 1 for each cost-loss value, and choosing the one that maximises the economic value.

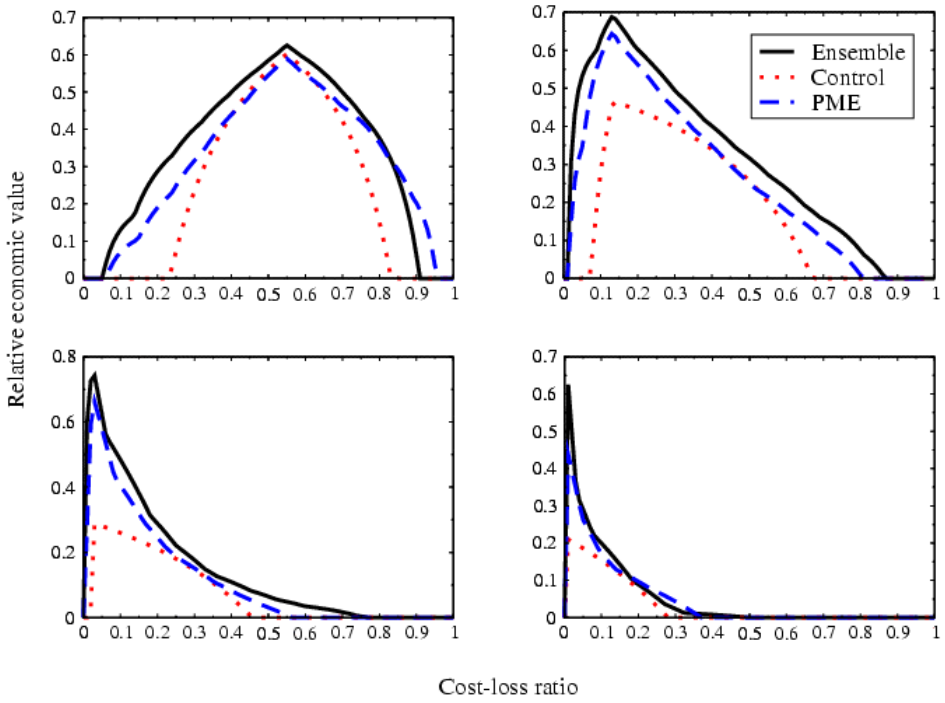

Figure 19 Relative economic value for the day 5 wave height forecasts as function of the cost-loss ratio. The threshold values are 2,4, 6, and 8 meter and corresponds to the values used for the reliabilities in figure 16.

The relative economic value of the forecasts for the joint probability of wave height and period is given in Figure 20. The threshold levels here correspond to the reliability diagrams shown in Figure 18. The standard deviation used to create the PME for the peak period is 2.71 s, corresponding to the root-mean square error at day 5. Encouragingly, for all cases shown, the relative economic value of the EPS is larger than the value of both the control forecast and the PME. The two cases on the left hand side of the plot represent rare combinations of wave height and period. For these cases, relative economic values above climatology are obtained only for very low cost-loss ratios (note the logarithmic cost-loss axis). The results indicate that the relative difference between the PME and the EPS is larger for rare, or complex situations. However, it is very important to remember that this is strictly dependent on the correct choice of probability level for deciding whether or not to take action.

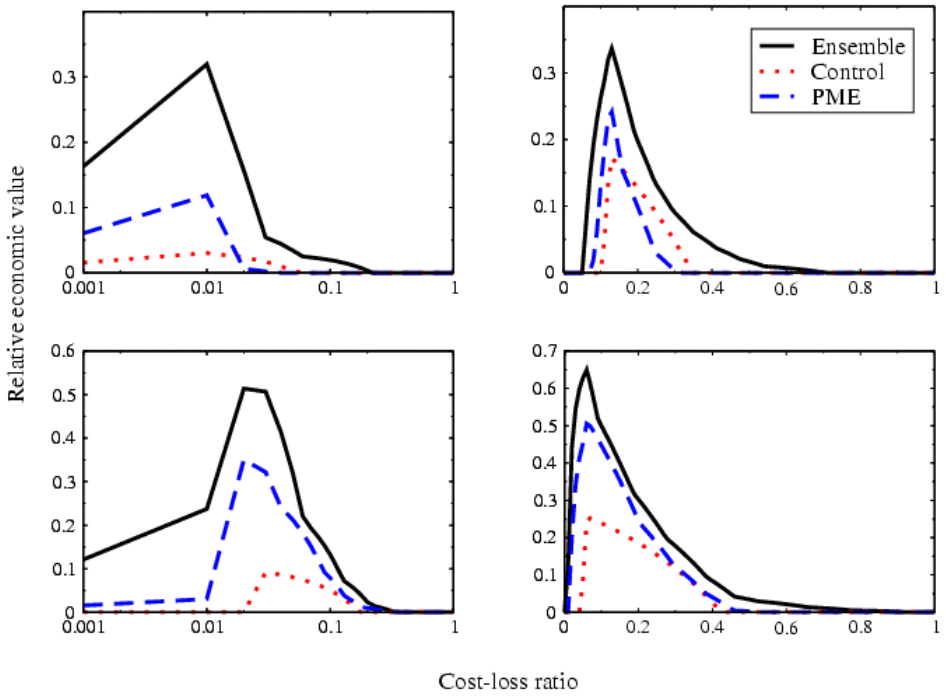

Figure 20 Relative economic value for the day 5 forecasts of the joint probability of wave height and peak period (see Figure 18 for threshold values). Note that the two plots to the left-hand side represent rare combinations of wave height and period and exhibit economic value above climatology only for cost-loss ratios below 0.1. To see the differences more clearly, these two graphs are plotted with logarithmic scale along the cost-loss axis.

Further insight into the performance of different forecasting systems can be obtained by comparing their respective hit rates and false alarm rates. The ROC curves can be obtained by plotting the hit rate against the false alarm rate for probabilities from 0 to 1. For a totally random forecasting system, the hit rate will be equal to the false alarm rate, and results in points along the diagonal line. A perfect forecasting system, with H=1 and F=0, would give one point at the upper left corner of the graph. Here, instead of showing plots of the ROC, we will concentrate on the areas under curves. This can be used as an index of accuracy, with A=1 for a perfect system and A=0.5 for totally random forecasts. The results for wave height and joint probability of waves and periods are given in table 3 and 4. In these tables, the scores of the PME are compared with those of the real ensemble. For wave height, the threshold values are the same as for Figure 19. For the joint probabilities of waves and periods, the four cases correspond to the thresholds used in Figure 20. As

expected form the economic value, the areas under the ROC curves are larger for the real ensemble in all cases.

	Hs > 2m	*Hs > 4m*	*Hs > 6m*	*Hs > 8m*
Ensemble	0.880	0.912	0.914	0.858
PME	0.870	0.885	0.846	0.728

Table 3 Area under ROC-curve for significant wave height and four different threshold values for wave height.

	Case 1	*Case 2*	*Case 3*	*Case 4*
Ensemble	0.682	0.725	0.805	0.876
PME	0.623	0.653	0.748	0.827

Table 4 Area under ROC-curve for joint event of peak period and significant wave height as shown in figure 18.

9. Comparison with altimeter data

To compensate for the shortage of buoy and platform data for the open oceans in general, and the southern hemisphere in particular, the wave EPS has in addition been compared to satellite altimeter data covering the whole globe. The data cover the same three year period as the buoy and platform data. The data shown in the plots focus on the results for the day 5 forecasts. All the other forecast ranges, from day one to day 10, have been examined and the main impression is that the results below are in general representative of the other forecast steps. Starting with the spread-skill relationship, given in Figure 21, the absolute errors show a clear relationship with the ensemble spread. As before, the spread is taken to be the inter-quartile range of the ensemble forecasts. In addition to the global results, the figure also shows the spread-skill relation for different regions; Northern Hemisphere, Southern Hemisphere, Tropics, North Atlantic and Pacific. Here, the North Atlantic and North Pacific are defined as the areas of these oceans that are more than 20° north respectively.

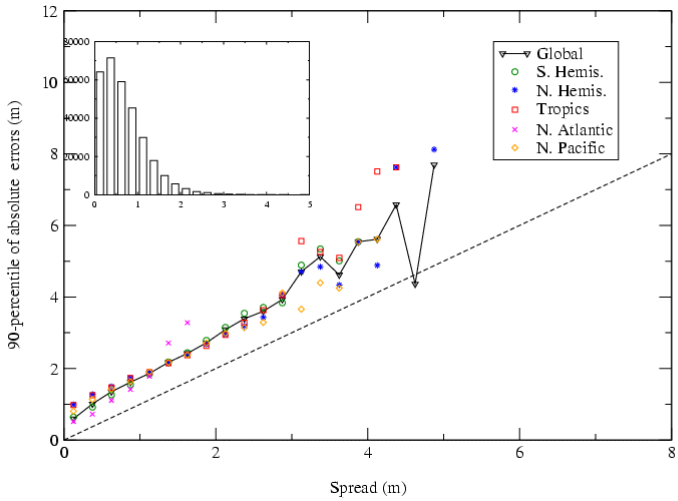

Figure 21 Day 5 forecast spread-skill based on the altimeter wave height. The frequency in the various bins for ensemble spread are depicted in the bar diagram. The Tropics are defined as the area between 20 deg north and 20 deg south.

The reliability of the EPS for day 5 when verified against altimeter data is given in Figure 22. The plot shows the result for the global data set for the threshold values of 2, 4, 6 and 8 meters. These are the same thresholds that were used for the buoy observations. Compared to the results obtained with the buoy data, the performance is slightly poorer. Figure 23 and 24 show the reliability for the Northern and Southern Hemisphere respectively. The corresponding results for the Tropics are given in Figure 25. For the Northern

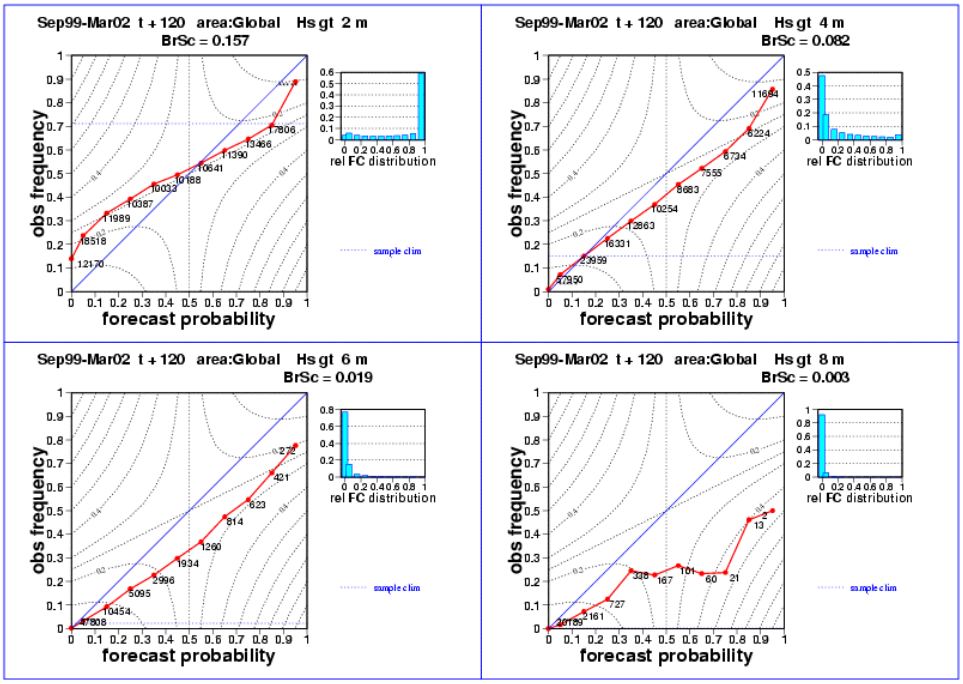

Figure 22 Reliability diagram for significant wave height at day 5 when global altimeter data are considered. The threshold values are shown above each plot, and are the same as those used for the buoy data.

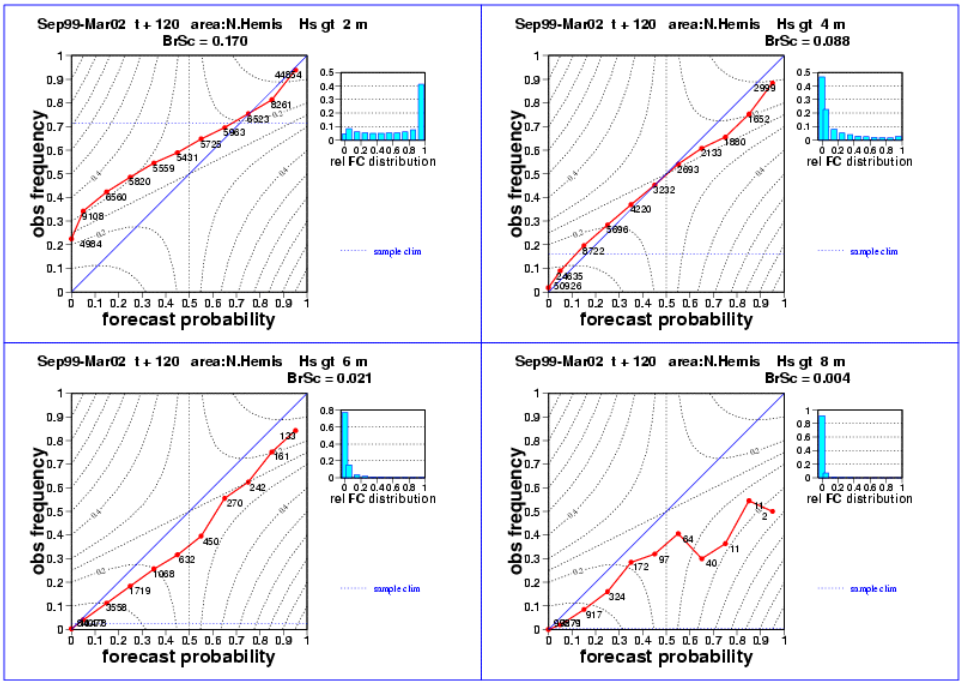

Figure 23 Reliability diagram for significant wave height at day 5 for altimeter data in the northern hemisphere.

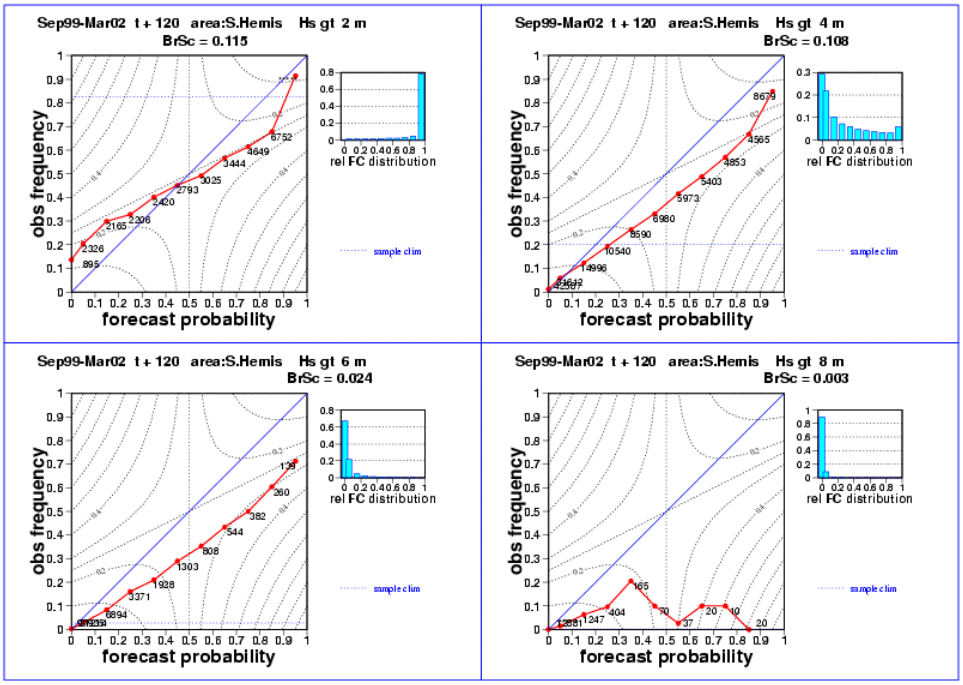

Figure 24 Reliability diagram for significant wave height at day 5 for altimeter data in the southern hemisphere.

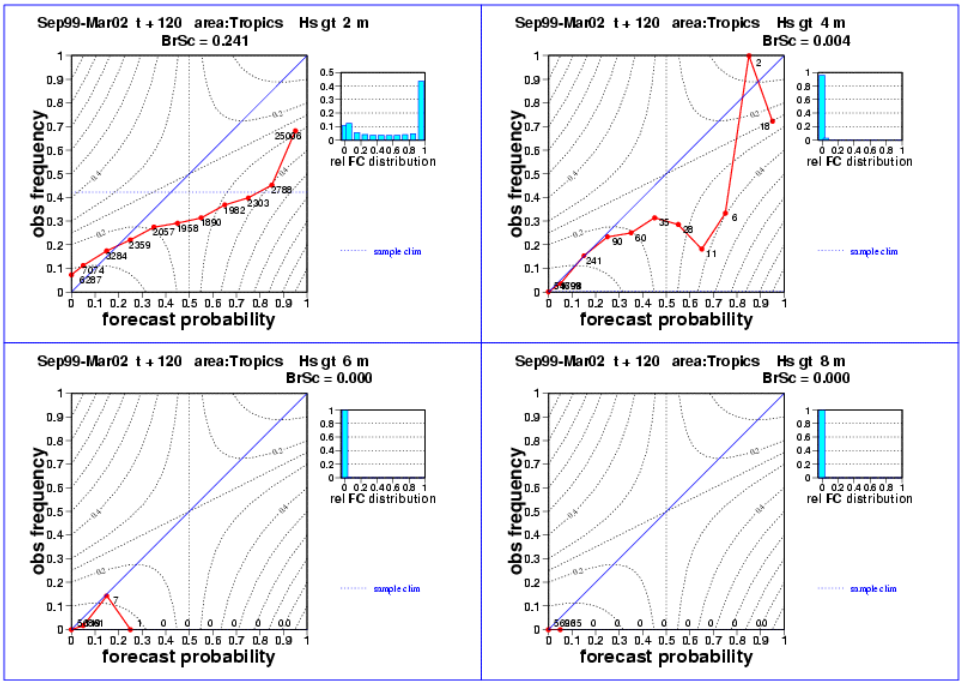

Figure 25 Reliability diagram for significant wave height at day 5 for altimeter data in the tropics. For this area, there are almost no wave height observations above 6 metre.

Hemisphere, the results are comparable to those obtained with buoy observations. Note however that for the 2m threshold value, the tendency to under-forecast low probabilities is further enhanced. We believe this is

caused by the fact that the altimeter has problems measuring the low wave heights (Janssen 2000). For the Southern Hemisphere, the EPS seems to have a small tendency to over-forecast the probability of wave height of more than 4 meters over most probability levels. In the Tropics, there are almost no observations of waves exceeding 6 meters, and even for wave heights above 4 meters, there are very few observations, making it difficult to draw any conclusions. For the lowest threshold value, the forecast system are over-forecasting probabilities above approximately level 0.4.

The analyses of the EPS performance in terms of the relative economic value are shown for waves above 2, 4, 6 and 8 meters in Figure 26. The black solid line represents the relative values of the EPS, and the red dotted line represents the deterministic forecasts, which is the control forecast in this case. At first glance, the results look very similar to those obtained from the buoy and platform data (see Figure 19). However, a closer examination reveals some important differences. The maximum relative economic values are slightly reduced by roughly 0.1. For the lowest threshold value, the economic value does not exceed the climatological value for cost-loss rations below 0.2 when the altimeter data are used. Using the buoy and platform observations, this was achieved already at cost-loss ratios above 0.05. On the other hand, for cost-loss ratios above 0.5 and wave heights above 6 and 8 meters, the economic values have improved compared to the results acquired from the GTS data.

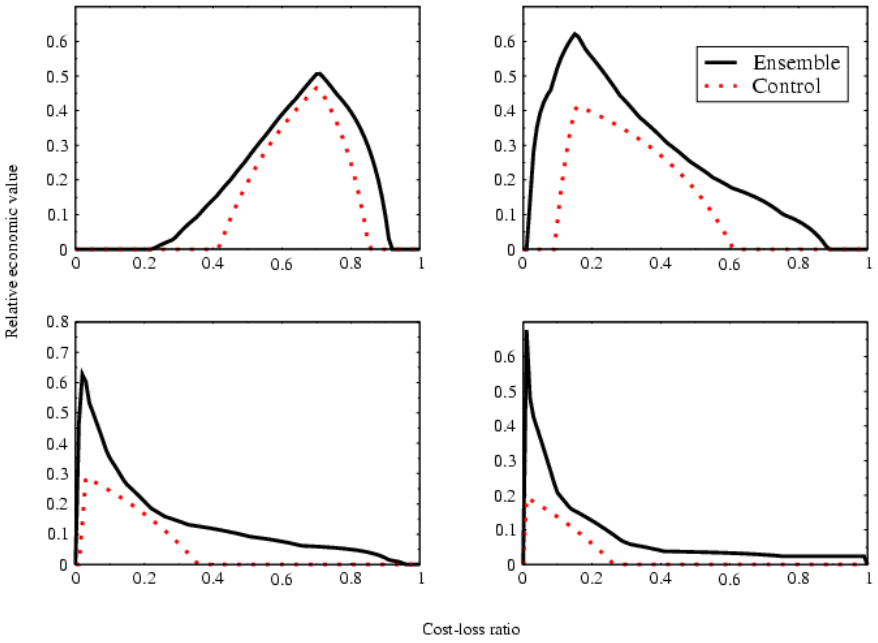

Figure 26 Relative economic value for the day 5 wave height forecasts as function of the cost-loss ratio Results are based on altimeter data. The threshold values are 2,4, 6, and 8 meter and corresponds to the results in figure 19.

10. Conclusions

The ECMWF ensemble prediction system for waves and wind over marine areas has been compared with observed data for the period between September 1999 to March 2002. Two different data sets have been used. First, the model has been compared with quality controlled data from platforms and buoys, available via the Global Telecommunication System (GTS). This amounts to approximately 46,000 observations for wave heights, and about 60,000 observations for the wind speeds. Due to the limited geographical coverage

of the GTS data, which are largely confined to the Northern Hemisphere and mostly located over the continental shelves, the EPS wave heights have in addition been evaluated against global satellite altimeter data. This data set, covering the same time span as the GTS data, consists of 310,000 grid box mean observations. The probabilistic forecasts have been tested for spread, spread-skill relation, reliability and relative economic value.

When viewing the spread in relation to the skill of the deterministic forecast, a relatively strong correlation is demonstrated. From a forecaster's point of view, the ensemble spread should be a measure of how much confidence he or she can have in a particular weather prediction. Small spread equals strong confidence, and vice versa. By sorting the ensemble spread into different bins, and calculating the percentiles of the absolute errors for each bin, an upper bound to the expected forecast error is found. For waves, the slope of the curve is more or less parallel to the diagonal line, indicating that the forecast error of the deterministic model could be expected to bounded by the inter-quartile range of the ensemble spread. For the wind speed, the slope is less steep. Nonetheless, an apparent correlation between spread and skill is also demonstrated.

The reliability of the probability forecasts has been tested by analysing reliability diagrams for four different threshold levels for both wave height and wind speed. The reliability seems to be very good indeed, although the buoy and platform observations indicate a small tendency towards over-confidence in forecasting wave heights above 6 and 8 meters. The reason for this is not clear to us at the moment. Generally, the wave model is known to underestimate high waves (Bidlot et al. 2002), but when looking at individual time series for cases with very high waves, we can see that in most cases a number of the ensemble members have predicted wave heights that are well above the observed values. The reason seems to be that these members have been forced by sufficiently strong wind speeds, resulting in too large probabilities being forecasted for the larger waves classes. For the two lowest threshold levels, the plotted points of observed frequency versus forecasted probability are very close to the diagonal line. To a certain degree, the altimeter observations confirms the above result, at least for the Northern Hemisphere. However, there is a more pronounced tendency for overconfidence in the probability forecast when tested against altimeter data.

In this study, the reliability of four different combinations of wave height and wave period has been calculated. Two of these cases are considered to be rare, although indeed possible combinations of height and period. The reliability as it turns out is rather good. Even for the single most atypical combination, the points in the reliability diagram are located relatively close to the diagonal. However, for the joint probabilities, there is also a general tendency for slightly too confident probability forecasts.

To test the value of the EPS forecasting system for decision making, the method suggested by Richardson (2000) for calculating the relative economic value as a function of the cost-loss ratio has been applied. The value of the forecast is measured relative to that of climatology and perfect knowledge of the future weather. The method also enables comparison with other forecasting methods. The value of the forecasts has been compared with those of traditional deterministic forecasts. A so-called poor-man's ensemble can be created by simply adding normally distributed noise to the deterministic forecast, using information on the error statistics to determine the spread. The spread from such a forecast will be constant for a given forecast range, and consequently can not be used to decide the expected confidence in the deterministic forecast. It still performs relatively well as we have demonstrated in this investigation, even though it is in almost all situations outperformed by the real ensemble. For more complex forecasting parameters, the benefit of using the real ensemble becomes even more apparent. This encouraging result should hopefully serve as an

inspiration for the development of more interesting products based on the EPS. The potential of the wave ensembles as a marine forecasting tool could then be exploited to its full extent.

Acknowledgements

This research is partly funded by the European Research Council through the SEAROUTES project. We thank Peter Janssen, Hans Hersbach, David Richardson, Roberto Buizza and Tim Palmer for support and valuable discussions. We would also like to thank the different organisations that have over the years deployed and maintained the buoy networks, and the platforms operators who disseminated the wind and wave data. In particular, we would like to thank Ian Hendry of the Met Office for providing information on the North Sea platforms, Magnar Reistad from DNMI for his investigative work on the Norwegian platforms and Ian Hunter from the South African Weather Service for providing information and missing data from ZSWAV. Finally, we acknowledge the very useful source of information available from the National Data Buoy Center and the Canadian Marine Environment Data Service web pages. We are also grateful to ESA for the provision of altimeter data.

References

Anderson, J. L., 1996: A Method for Producing and Evaluating Probabilistic Forecasts from Ensemble Model Integrations. *J. Climate*, **9**, 1518-1530.

Bidlot, J., B. Hansen and P. Janssen, 1997a: Modifications to the ECMWF WAM Code, *ECMWF Research Department Tech. Memo*. No.232. 22 pp.

Bidlot, J-R., P. Janssen, B. Hansen and H. Günther, 1997b: A modified set up of the advection scheme in the ECMWF wave model, *ECMWF Research Department Tech. Memo*. No.237, ECMWF, Reading, United-Kingdom.. 31 pp.

Bidlot, J-R., D. J. Holmes, P. A. Wittmann, R. L. Lalbeharry and H. S. Chen, 2002: Intercomparison of the Performance of Operational Ocean Wave Forecasting Systems with Buoy Data, *Wea. Forecasting*, **17**, 287-310.

Buizza, R., D. S. Richardson and T. N. Palmer, 2002: Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles. *Q. J. R. Meteorl. Soc*. in press.

Buizza, R., J. Barkmeijer, T. M. Palmer and D. S. Richardson, 2000: Current status and future developments of the ECMWF Ensemble Prediction System. *Meteor. Appl*., **7**, 163-175.

Challenor,P.G, and P.D. Cotton, 1997: The SOC contribution to the ESA working group calibration and validation of ERS-2 FD measurements of significant wave height and wind speed. pp-95-100, *CEOS wind and wave validation workshop, 3-5 June 1997*, ESTEC, Noordwijk, The Netherlands. ESA WPP-147.

Evans, R. E., Harrison, M. S. J., Graham, R. J. and K. R. Mylne, 2000: Joint Medium-Range Ensembles from The Met. Office and ECMWF Systems. *Mon. Wea. Rev*., **128**, 3104-3127.

Farina, L., 2002: On ensemble prediction of ocean waves. *Tellus*, **54A**, 148-158.

Hamill, T. M., 2001: Interpretation of rank histogram for verifying ensemble forecasts, *Mon. Wea. Rev*., **129**, 550-660.

Janssen, P. A. E. M., B. Hansen and J-R. Bidlot, 1997: Verification of the ECMWF Forecasting System against Buoy and Altimeter Data, *Wea. Forecasting*, **12**, 763-784.

Janssen, P. A. E. M., 1991: Quasi-linear theory of wind wave generation applied to wave forecasting, *J. Phys. Oceanogr*., **21**, 1631-1642.

Janssen, P., 2000: Wave modeling and altimeter wave height data. in '*Satellite, Oceanography and Society'*, edited by D. Halpern, Elsevier Science. 35-56

Janssen, P. A. E. M., J-R. Bidlot and B. Hansen, 2000: Diagnosis of the ECMWF ocean-wave forecasting system, *ECMWF Research Department Tech. Memo*. No.318. 23 pp.

Janssen, P. A. E. M., J. D. Doyle, J. Bidlot, B. Hansen, L. Isaksen and P. Viterbo, 2002: Impact and feedback of ocean waves on the atmosphere. *Advances in Fluid Mechanics, Atmosphere-Ocean Interactions*, Vol. I, WITpress, Ed. W.Perrie. 155-197.

Komen, G. J., L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann, and P. A. E. M. Janssen, Eds., 1994: *Dynamics and Modelling of Ocean Waves*. Cambridge University Press, 533 pp.

Monaldo, F., 1988: Expected difference between buoy and radar altimeter estimates of wind speed and significant wave height and their implications on buoy-altimeter comparisons, *J. Geophys. Res*., **93C**, 2285-2302.

Richardson, D. S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Q. J. R. Meteorl. Soc.*, **126**. 649-667.

Strauss, B., and A. Lanzinger, 1996: Verification of the Ensemble Prediction System (EPS). *ECMWF Newsletter*, 72, 9-15.

Saetra, Ø., J. R. Bidlot, H. Hersbach and D. S. Richardson, 2002: Effects of observation errors on the ensemble statistics. In preparation.

Vogelezang, D. H. P., and C. J. Kok, 1999, Golfhoogteverwachtingen voor de Zuidelijke Noordzee, *KNMI Technisch Rapport*; TR-223, 24 pp.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic press, 467 pp.

WAMDI Group, 1988: A third generation ocean wave prediction model, *J. Phys. Oceanogr.*, **28**, 1775-1810.