Wave Ensemble Predictions for Safe Offshore Operations

Ana Carrasco, Øyvind Sætra and Jean-Raymond Bidlot

1 Introduction

Many offshore operations require windows of relatively calm weather. One example is the transportation of an oil rig between two locations over a period of, say, 24 or 48 hours. In order to make the necessary preparations the crew needs to know at least three days in advance when a calm weather window is to be expected. The optimal scenario is to have absolutely calm weather, e.g. waves with significant wave heights lower than, say, 1 meter over a 24 hour window. If an unpredicted worsening of the weather arises, let's say the significant wave height rising to 2 meters for some hours, the operation can continue with some difficulties. But with significant wave height larger than 3 metres, even for only a few hours, the operation will fail and big economic losses can be expected. Other examples of operations that require calm weather windows include the maneuvering of ships and platform maintenance.

In this article we will study the utility of wave forecasts of relatively calm weather windows with the requirements of offshore operations in mind. Given a deterministic (high resolution) and an ensemble forecast (probabilistic forecast), a user will have to make the decision of when to start the operation or if it is necessary to cancel it based on the predictions available. Is the probabilistic forecast better than the deterministic one, or are there cases when the deterministic forecast performs just as well? For the probabilistic forecast, are decisions best based on the ensemble mean or on some probability threshold?

In order to answer these questions ten years of wave forecast verification combined with financial information are used as input to the cost lost model traditionally used in the analysis of weather forecasts (Richardson, 2000; Wilks, 2006; Thornes and Stephenson, 2001; Roulston and Ellepola J, 2005). The cost-loss model has been used in the analysis of extreme weather validation, but here we will apply it to forecasts of relative calm weather windows. We perform this exercise by simulating that decisions must be taken concerning operations in an area around the Conocco-Phillips Ekofisk oil field located in the North Sea (Figure 1).

The forecast verification is framed in a *contingency table* which is constructed by using in-situ observations of significant wave height (H_s) and outputs from both the ensemble prediction system (EPS) and the deterministic run from the European Centre for Medium-Range Weather Forecasts (ECMWF).

The forecast skill is evaluated in a statistical sense, and the larger the sample size the more trustworthy the verification results become. In this case the total number of collocated data, i.e. when observations and model predictions both exist, were $n = 30,332$. With such a large sample size one may hope that the analysis of calm weather windows may be statistically more robust in comparison with the verification of extreme events.

2 Data

Ten years of significant wave height data from observations and model, from the January 1999 to December 2009, are used. The observations come from locations near the Ekofisk oil platform (Figure 1). The idea was to select the maximum number of observations that describe the same wave climate. Time series from the ECMWF wave prediction system (deterministic and ensemble) were interpolated onto the observation positions.

Observations

Wave observations from 27 locations in the North Sea were used. The data are part of the information broadcast to meteorological offices via the Global Telecommunication System (GTS). Most of them are from fixed platforms operating in the North Sea, however very little information on the type of wave sensor is available in the GTS data record. Nevertheless, the wave height data generally compare well with model analysis. From the data records, time series are reconstructed to perform a basic quality check on the data, Bidlot et al., 2002. Temporal scales are made uniform by averaging the hourly observations in a window of 4 hrs centered around the verification time. In-situ observations exhibit a high-frequency variability on a time scale of 1 hr.

Not averaging the data will result in a scatter between the models and observations, which can be linked to high frequency variability, not present in the model, Janssen et al., 1997. For a more detailed description of the data treatment, see Bidlot et al., 2002 and Saetra and Bidlot, 2004.

Model predictions

An overview of the operational ensemble prediction system (EPS) at ECMWF is given in Leutbecher and Palmer, 2008 and Buizza, 1997. For these ten years, the ensemble system has been running with 51 members, the *control run* which starts from an unperturbed initial conditions (the analysis after the assimilation step) and 50 from perturbed initial conditions. The EPS is thus based on the notion that forecast uncertainty is dominated by errors in the initial conditions. Stochastic perturbations of the atmospheric model tendencies are also applied throughout the perturbed forecast runs to account for uncertainties in the model physics parametrisation. The *deterministic run* is like the control run but with higher horizontal resolution. In all model integration, the atmospheric model is coupled to the wave model WAM (Janssen et al., 1997, G. Komen et al., 1994). The WAM model marked the introduction of so-called third generation wave models which explicitly accounts for the non-linear interaction between the wave components. For simplicity all members start from the same initial wave conditions as prescribed by the latest operational analysis. The divergence between the wave ensemble members is therefore due only to different wind forcing.

During 1999 to 2009 there have been several changes that are described in details by Leutbecher and Palmer [2008]. For example, the horizontal resolution of the wave ensemble changed from 1.5^o to 1.0^o on 20 November 2000. While the horizontal resolution of the deterministic run changed from 0.5^o to 0.36^o on 1 July 2000. The forecast were only available every 12 hours before 15 March 2005 and every 6 hours since then. The wave model changed from being run just with deep water integration to both deep and shallow water after November 2000. In this study, we will ignore the possible non-uniformity of the forecast quality due to these changes.

3 Methods

The cost-loss model proposed by Richardson [2000] assess the economic value of a forecast if both the monetary loss L (in, say, dollars) due to adverse weather situations and the cost C of preventing weather damage are known in monetary terms. For a given event the cost C is assumed to be less than the loss L. The idea is to consider a hypothetical decision maker who must choose to take action or do nothing. The decision is based on the forecast available.

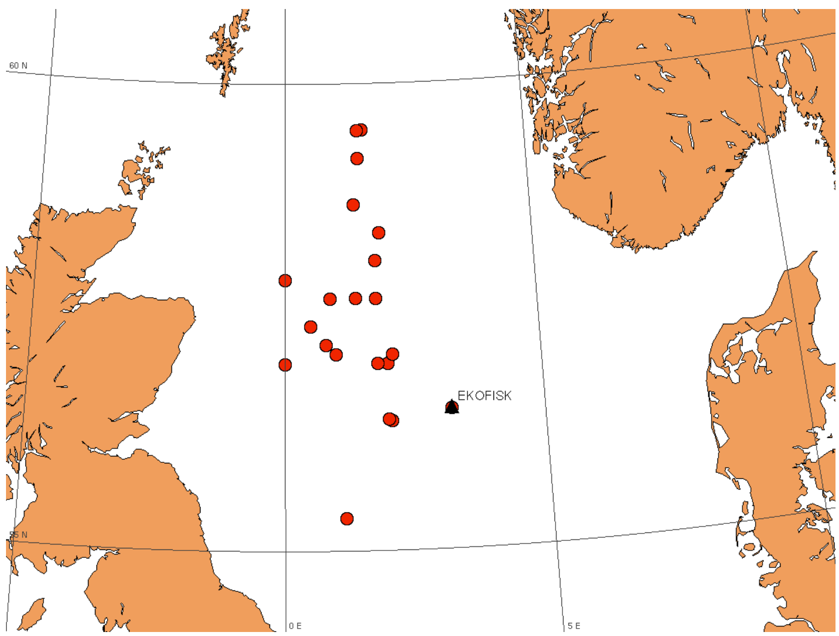

Figure 1: The positions of buoys (red circles) and the Ekofisk oil field (black trangle).

First we take a benign weather event X_0 defined by low significant wave height in a span of time. The obvious X_0 is $H_s \leq 1m$ during an entire 24-hr forecast window of [72 78 84 90 96] hours (the ECMWF fields come at six-hour intervals). However, the benign weather X_0 we will present here is $H_s \leq 3m$ for at least four of the five forecasts [72 78 84 90 96] hours (or $H_s \leq 3m$ for at least two of the three forecasts [72 84 96] when only three forecast are available).

When the benign weather event X_0 is forecast as *yes*, the decision maker choose to do nothing and carries on with the offshore operation. If the event is forecasted to occur (i.e. the waves stay below 3m) and it happens the expenses will be zero dollars. If the event is forecasted to occur but does not happen a loss L is incurred. Such loss L would be the expenses due to damage of equipment or prolonging the operation in time. On the other hand, if bad weather is forecasted, i.e. the benign weather is forecasted as *no*, the decision maker will spend an amount of money C protecting, whether the event occurs or not. Let's say that the protection action is postponing the operation. C is then the expenses associated with the delay. This situation is summarized as:

$$\begin{array}{lcccc} & & \multicolumn{2}{c}{benign\ \ weather\ \ event} & \\ & & \multicolumn{2}{c}{X_o\ \ observed} & \\ Forecast & & yes & no & \\ & yes & 0 & L & no\ \ take\ \ action \\ & no & C & C & yes\ \ take\ \ action \end{array} \tag{1}$$

In order to use the same formulation as Richardson in Jolliffe and Stephenson [2003], we considerer instead the complementary bad weather event X_1, which is: $H_s > 3m$ for at least two of the forecast within the window [72 78 84 90

96] . Then (1) instead becomes:

$$
\begin{array}{llllll}
 & & & bad \quad weather \quad event & & \\
 & & & X_1 \quad observed & & \\
 & & yes & no & & \\
Forecast & yes & C & C & yes & take \quad action \\
 & & & & & \\
 & no & L & 0 & no & take \quad action.
\end{array}
\tag{2}
$$

The *relative economic value* of a given forecasting system is defined as

$$
V = \frac{E_c - E_f}{E_c - E_p} \tag{3}
$$

Here, E_c is the expenses expected when no forecast is available, E_f is the expenses of the forecasting system considered and E_p is the expenses of a hypothetical perfect forecasting system. There are two possible strategies for determining E_c, based on the user knowledge of the area climatology, either to protect all the time therefore no extra cost will be incurred or to never protect and suffer the extra loss when the event happens.

The definition of V is a skill score of expected expenses with climatology as reference. When $V > 0$ the decision maker will gain some economic benefit by using the forecast information. When $V = 0$ the system is as bad as the climatology. The maximum value of $V = 1$ is reached when the system perfectly predicts the future. To evaluate these expenses a *contingency table* is used. This table gives the counts for each of the four possible combinations of forecast and observed event:

$$\begin{array}{cccc} & & X_1 \quad observed & \\ & & yes & no \\ Forecast & yes & a & b \\ & & & \\ & no & c & d \end{array} \tag{4}$$

Here a indicates the number of times the event X_1 was forecasted to occur and did occur (hits); b is the number of times the event was forecast to occur, but did not occur (false alarms); c is the number of times the event was forecast not to occur but did occur (misses); and d is the number of times the event was forecast not to occur and did not occur (correct rejections).

The expression for the expenses thus become:

$$\begin{array}{rcl} E_c & = & min \quad (C, L \quad \frac{(a+c)}{n}) \\ E_p & = & C \quad \frac{(a+c)}{n} \\ E_f & = & \frac{a}{n} \quad C + \frac{b}{n} \quad C + \frac{c}{n} \quad L \end{array} \tag{5}$$

where $n = a + b + c + d$ is the total amount of collocated data. In this case $n = 30,332$. The saving $S =E_c - E_f$ (Thornes and Stephenson [2001]) gives a direct measure of the amount of finacial benefit. These quantities are calculated for the deterministic run, the mean of the ensemble and for the entire ensemble at each probability threshold.

4 Results

When analysing the complementary event to $H_s \leq 1m$ during an entire 24-hr forecast window, which is: $H_s > 1m$ for at least one of the forecast [72 78 84 90 96] the reliability diagram presents a big underforecasting bias and a post-processing calibration was needed it, see Roulston and Ellepola J [2005].

Here we analyse the event $X_1 : H_s > 3m$ for at least two of the forecast within the window [72 78 84 90 96] . The economic value formulated in the equation 3 is shown in Figure 2. For each probability a curve of economic value is calculated. The ensemble economic value is obtained by taking the envelope of all these curves at each C/L. The envelope curve, shown with the label *ENS* in Figure 2a), is then the optimum maximum V of the ensemble system. For the entire range of C/L (the cost-loss ratio) the economic value V of the ensemble is larger than both the deterministic and the ensemble mean. This study will focus in small values of C/L since they can be related to the offshore operations mentioned. It is important to notice that for a range of $0.3 < C/L < 0.45$ the V for ensemble, the ensemble mean and the deterministic are very close. For C/L smaller than 0.05 only the ensemble gives some economic benefit.

A user with C/L of 0.1, 0.2, 0.3 and 0.4, will benefit by postponing the operation when the probabilities are equal or bigger than 0.12, 0.23, 0.32 and 0.39 respectively, see Figure 2b). At these probabilities V reaches its maximum. In a perfectly reliable system the threshold probabilities are exactly equal to C/L (Jolliffe and Stephenson [2003]). This is almost true in this case due to the fact that the system is quite reliable at these probabilities. A measure of the reliability is shown in the values of the bias, $b(p)$, in the Table 1. The closer to one the better. The positive economic value$V$ obtained from the ensemble mean, for C/L=0.1 and C/L=0.2, have a lower value than the peaks of the C/L curves. For a user with C/L=0.3 and $C/L = 0.4$ the deterministic and ensemble mean score almost the same at the probability at which V reaches its maximum. This means that for these user the operation can be postpone when the mean of the ensemble or the deterministic run forecasts the event X_1 since the probabilistic forecast does not seem to be any better.

Assuming that the losses are of the order of two million dollars, the expenses in thousands of dollars obtained according to Eq. 5, are shown in

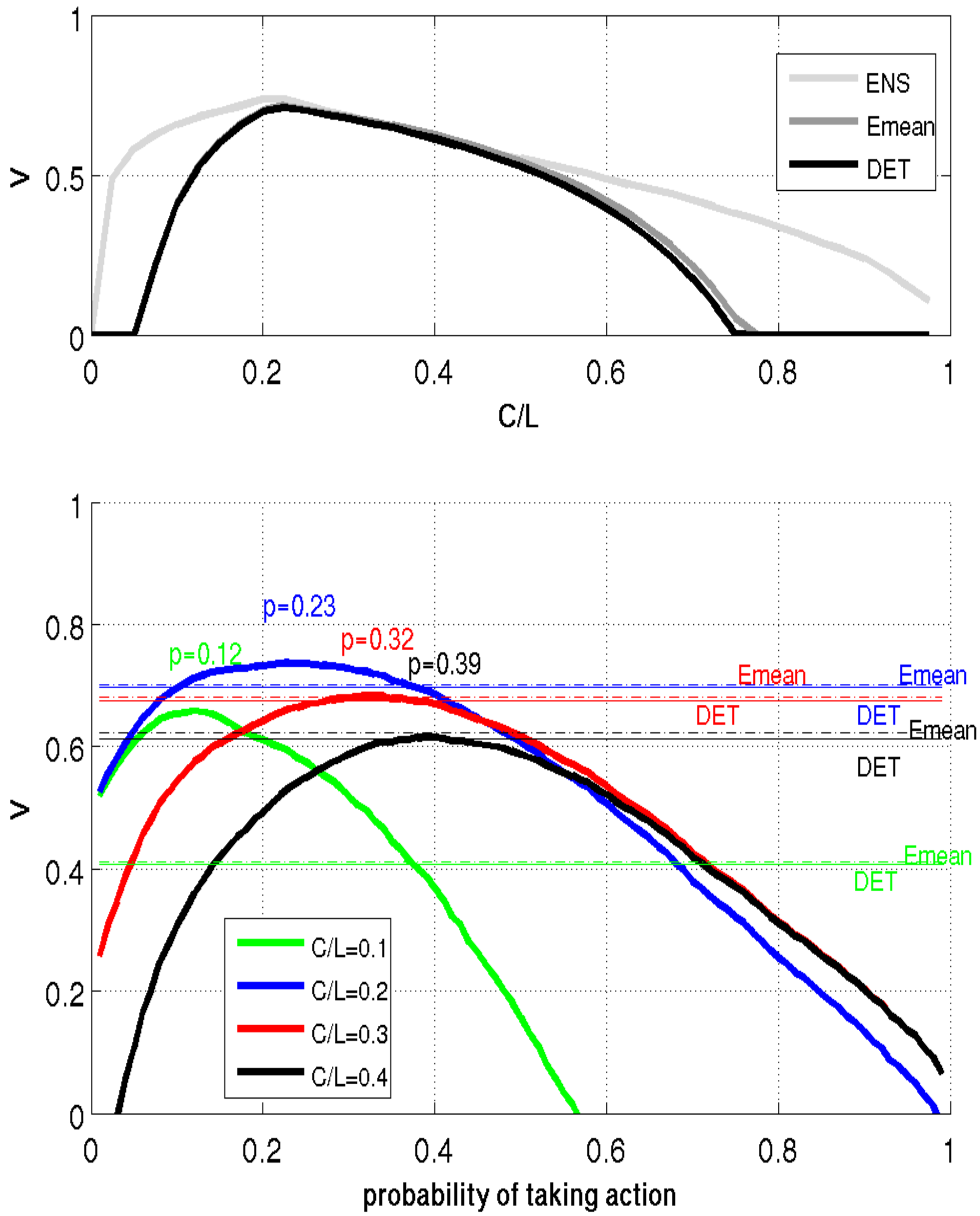

Figure 2: Economic value V for the event X_1. a) V versus the ratio C/L obtained by using the ensemble, *ENS*, the ensemble mean, *Emean*, and the deterministic forecast, *DET*. b) V versus probability of taking action at fixed values of C/L. The probability at which V reaches its maximum are also indicated. The ensemble mean and deterministc forecast are also shown in b).

Table 1. For the probabilistic forecast the expenses and three verification measures are calculated at the probabilities at which the operation should be postponed. The user wants to chose the strategy that will use the least amount of money. The minimum expenses are those obtained for the the perfect forecast, E_p. The expenses closest to the perfect are the expenses obtained from the ensemble forecast $E_f(p)$ in the first two cases. In the cases where $p = 0.32$ and $p = 0.39$, the expenses calculated from the ensemble mean E_f^m are closest to E_p. The next are $E_f(p)$ and finally the expenses calculated with the deterministic forecast E_f^d with a small difference between them. The saving of using $E_f(p)$ is given by E_c-$E_f(p)$. The user with $C/L = 0.2$ will benefit the most.

In order to evaluate the probabilistic system the last three columns of table 1 show the bias, $b(p)$, the false alarm rate, $fa(p)$, and the hit rate, $h(p)$. The probabilistic system presents a light overforcasting since the bias is bigger than one. The closer the hit rate is from one, indicates that the event X_1 happened and was predicted most of the occasions. The closer to zero the false alarm rate is, the better the system was at predicting the benign weather event X_0. The best false alarm was for the user with $C/L = 0.4$, *i.e.* $p = 0.39$, while the best hit rate was for the user with $C/L = 0.1$, *i.e.* $p = 0.12$. Which is more valuable is for the user to decide. The user with $C/L = 0.2$, present the smallest distance between hit rate and false alarm.

In a reliable system, the maximum distance between hit rates and false alarm rates is reached at $V_{max}(C/L = \frac{a+c}{n}) = max(H - F)$ as Richardson explained in the book Jolliffe and Stephenson [2003]. The maximum value of V is reached when the cost/loss ratio equals the base rate: $\frac{a+c}{n}$. This is at $C/L = 0.23$, see Figures 2 which also is the percentage of times the event X_1 happens. The values of cost-loss ratio chosen in this example are very close to V_{max} where the system performs better.

		X_1						
prob.	$E_f(p)$	E_f^d	E_p	E_f^m	E_c-$E_f(p)$	$b(p)$	$fa(p)$	$h(p)$
$p \approx \frac{C}{L}$						$\frac{a+b}{a+c}$	$\frac{b}{b+d}$	$\frac{a}{a+c}$
0.12	112.5	144.8	51.3	144.2	127.5	1.861	0.245	0.96
0.23	187	194	98	193	241	1.54	0.147	0.894
0.32	233.5	234.5	136.9	232.4	194	1.184	0.095	0.833
0.39	269.7	270.4	171	268	158	1.018	0.066	0.776

Table 1: For a loss L= 2000000 dollars displayed are the expenses in 10^3 dollars and three verification measures at the probability at which the operation should be postponed.

5 Conclusions

With the event studied here, it was easy to demonstrate that the probability at which to postpone the operation was close to the cost/loss ratio. Once the probability was chosen, the expenses were calculated and the best expenses were identified. Evaluating how good the forecast system is, and with it how good this selected probabilities can be usefull, is not obvious. Instead of starting by fixing the cost/loss ratio we could analyze the range of cost/loss for which the distance between Hits rate and False Alarm are large enough. Although this range might not be of interest for the industry, we could offer it as the range for which we have a credible answer. This same analysis could be done for other events even if they present a bias. The probability at which to postpone the operation would not be close to the cost/loss ratio. Analyzing how useful these probabilities are is a challenge.

We could use the strategy suggested by Thornes and Stephenson [2001] where they classify the error by types. A forecast error that involves safety has more weight than an error that involves waste of money. Still there is a long way to go with respect of communicating with the users of EPS, for instance, Casati et al. [2008], Demeritt et al. [2010]. Although their decision making process is more complicated we could provide some recommendations for specific cases.

Acknowledgments

This is part of the the Wave Ensemble Prediction for Offshore Operations project (WEPO) which has been funded by the Research Council of Norway in collaboration with Conocco-Phillips.

References

J.-R. Bidlot, D.J. Holmes, P.A. Wittmann, R.L. Lalbeharry, and H.S. Chen. Intercomparison of the performance of operational ocean wave forecasting systems with buoy data. *Wea. Forecasting*, 17:287–310, 2002.

R. Buizza. Potential forecast skill of ensemble prediction and spread and skill distribution of the ecmwf ensemble prediction system. *Mon. Wea. Rev*, 125:99–119, 1997.

B. Casati, L. J. Wilson, D. B. Stephenson, A. Ghelliand M. Pocernich, U. Damrath, E. E. Ebert, B. G. Browne, and S. Masonh. Forecast verification: current status and future directions. *Meteorol. Appl.*, 15:3–18, 2008.

D. Demeritt, S. Nobert, H. Cloke, and F. Pappenberger. Challenges in communicating and using ensembles in operational flood forecasting. *Meteorological Applications*, 17:209–222, 2010.

G. G. Komen, J.L. Cavaleri, M. Donelan, K. Hasselman, and S. Hasselman. *Dynamics and Modelling of Ocean Waves.* Cambridge University Press, 1994.

P. Janssen, B. Hansen, and J.-R. Bidlot. Verification of the ecmwf forecasting system against buoy and altimeter data. *Wea. Forecasting*, 12:763–784, 1997.

Ian T. Jolliffe and D.B. Stephenson. *Forecast Verification. A Practitioner's Guide in Atmospheric Science.* John Wiley and Sons, 2003.

M. Leutbecher and T.M. Palmer. Ensemble forecasting. *J. Comp. Phys.*, 227:3515–3539, 2008.

D. Richardson. Skill and economic value of the ecmwf ensemble prediction system. *Q. J. R. Meteorol. Soc.*, 126:649–668, 2000.

MS. Roulston and Smith Ellepola J. Forecasting wave height probabilities with numerical weather prediction models. *Ocean Engineering*, 32(14 –15): 1841–1863, 2005.

Ø. Saetra and J.-R. Bidlot. Potential benefit of sig probabilistic forecasts for waves and marine winds based on the ecmwf ensemble prediction system. *Wea. Forecasting*, 19:673–689, 2004.

J.E. Thornes and D.B. Stephenson. How to judge the quality and value of weather forecast products. *Meteorol. Appl*, 8:307 – 314, 2001.

DS. Wilks. *Statistical Methods in Atmospheric.* Academic Press, Burlington, MA; 627, 2nd edition, 2006.