VERFICATION OF OCEAN WAVE ENSEMBLE FORECAST AT NCEP[1]

Degui Cao, H.S. Chen and Hendrik Tolman

NOAA /National Centers for Environmental Prediction
Environmental Modeling Center
Marine Modeling and Analysis Branch
Camp Springs, Maryland, USA

1. INTRODUCTION

Ensemble forecast is a probabilistic forecast technique which deals with forecast with uncertainty and improves the forecast skills. The National Centers for Environmental Prediction (NCEP) has developed and implemented the ensemble global ocean wave forecast system (EGOWaFS) operationally (Chen, 2006). The EGOWaFS consists of one control member and ten ensemble members using the NWW3 (Tolman at al, 2002 and Tolman, 2003) as the wave model to generate the wave data. The model wind input is obtained from NOAA/NCEP Global Forecast System (GFS) 10m wind field and is updated every three hours. The control forecast is the current operational wave forecast. The ten members are generated using the GFS ensemble wind fields made by the breeding method (Toth and Kalnay, 1997). The initial wave fields of the ten members of wave forecast are the same as that of the control forecast. These ensemble wave forecasts are running four hours per day out to 126 hours. The main outputs of the system include the ensemble mean, spread, spaghetti diagram and probability at different thresholds every six hours, and are posted on NOAA/NCEP website[2] .

The ensemble forecast theory is resolving two main uncertainties existing in the nonlinear numerical forecast objectives: prediction and forecast uncertainties (Toth and Kalnay, 1997).). The prediction uncertainty is from the deterministic nonlinear system itself. The chaotic system is very sensitive to the initial conditions. Very small perturbation of the initial conditions could generate considerable different outputs and lead to big divergence in the model forecast over a finite time. These initial errors could easily come from the data observation and analysis. The forecast uncertainty is from the predictability of the numerical model. Because of our imperfect understanding and knowledge of the natural nonlinear systems, the model can not reflect completely all physical processes of the systems. In addition there are some errors coming from the numerical methods, dynamical formulation and physical parameterizations. Therefore verification of any developed ensemble forecast

---

[1] Marine Modeling and Analysis Branch Contribution Number 261
Corresponding author: Degui Cao. Degui.Cao@noaa.gov

[2] http://polar.ncep.noaa.gov/

system is performed to answer to which extent the ensemble system resolves these uncertainties (Jolliffe and Stephenson, 2003). After the EGOWaFS was set up, Chen (2006) used the significant wave height (H $_S$) and the wind speed at 10m height (U $_{10}$) from nearly 30 deep water buoys to compare with the ensemble forecast. The spread increases with the forecast hour. The ensemble forecast of U $_{10}$ and H $_S$ hit all the observation data while

operational forecasts miss many. Chen also evaluated the system for the storms in May through July, 2004. The ensemble system is more realistic than the deterministic system. Chen (2006) made the conclusions only studying several cases within three months. More ensemble outputs have been generated since then. This study covers the period from June 1, 2006 to March 31, 2007. H $_S$ and U $_{10}$ from the ten ensemble members and the control run are used for verification.

There are five sections in this paper. Section 1 is the introduction. Section 2 gives a brief description of the buoy data used in the study. Section 3 describes the conventional verification methods used: Brier skill and Brier skill score, cost-lost analysis and relative operating characteristics. Section 4 analyzes the results of each verification skill or score. The conclusions are drawn in section 5.

2. BUOY DATA USED FOR VERIFICATION

The wind and wave data from 67 NOAA/NDBC buoys at the same period from June 1 2006 to March 31 2007 are treated as the ground truth for verification. Fig.1 shows the global distribution of the buoys. The hourly and quality controlled buoy data are compared with the ensemble output each hour. The wind and wave climate data are generated using the NDBC datasets from January 1 1997 to December 31 2006. The climate data are hourly averaged at each buoy location without smoothing at larger time scale.

3. METHODOLOGY OF VERIFICATION

3.1 Brier score (BS) and Brier skill score (BSS)

BS measures the mean squared probability error. The quadratic scoring measure for a probabilistic binary forecast is defined as (Jolliffe and Stephenson, 2003)

$$\text{BS}=\frac{1}{N}\sum_{i=1}^{N}(p_i-o_i)^2 \qquad (1)$$

where N is the total event number. $p_i$is the forecast probability, o_i is the observed data (1 for the event happening, 0 for the event not happening). For a probabilistic forecast, p_i is between 0.0 to 1.0 since ensemble forecasts usually give an uncertain forecast unlike the deterministic forecast where the probability is 0 or 1. The best value for a BS is 0 for perfect forecast system.

Murphy (1973) decomposed the above formula into three terms:

$$\text{BS}=\underbrace{\frac{1}{N}\sum_{m=1}^{M}n_m(p_m-\overline{o_m})^2}_{\text{Reliability}} - \underbrace{\frac{1}{N}\sum_{m=1}^{M}n_m(\overline{o_m}-\bar{o})^2}_{\text{Resolution}} + \underbrace{\bar{o}(1-\bar{o})}_{\text{Uncertainty}} \qquad (2)$$

where M is the probability classes. $\bar{o}$ is the climate data.

Reliability measures the difference between the forecast and observed probability distribution. The best reliability value is 0. Resolution is the ability to distinguish forecast from averaged observed data or climate data. The best resolution value is 0. Uncertainty measures the error or variability in the observed data used in either the initial conditions or in the comparisons. Uncertainty is always greater than 0. BS is zero when the forecast is a perfect deterministic forecast.

The BSS is defines as

$$\mathrm{BSS}=\frac{BS-BS_{ref}}{BS_P-BS_{ref}}=1-\frac{BS}{BS_{ref}} \qquad (3)$$

where the reference BS_{ref} could be the climate data, operational run or different forecast system. A positive BSS shows skill in comparison to the reference ($\mathrm{BS}>\mathrm{BS}_{ref}$), otherwise there is no skill.

3.2 Reliability diagram

The reliability diagram plots the observed frequency against the forecast probability for different probability range. There are 11 ranges in this study (0., 0.1, 0.2, …, 1.0). The reliability diagram answers how well the prediction corresponding to its observation. The diagonal line indicates perfect reliability. No resolution line indicates the ensemble system can not catch the events occurred below that probability (see Fig. 4 and 5). The no skill line means no forecast skill if the event occurs below it.

3.3 Cost-lost analysis

3.3.1 Economical value

The so-called decision-analytic model (Jolliffe and Stephenson, 2003) (Table 1) provides the users to alter their actions based on forecast information and different economical purposes.

		Forecast /Action	
		Yes	No
Event	Yes	Hit(h) Mitigated loss ($C+L_u$)	Miss(m) Loss ($L=L_p+L_u$)
	No	False alarm(f) Cost(C)	Correct rejection (c) No cost (N)

Table 1. The costs and losses accrued by the use of the wave prediction, depending on prediction and observed events

If the event does not occur, and the user takes no action, there is no cost, N=0. If the event does not occur, and user takes action, there is a cost C. If the event occurs, and user takes action, there is a cost, C, plus unprotected loss, L_u. If the event occurs and user takes no action, then there is a loss of L_u and protected loss, L_p.

The economic value is defined as:

$$V=\frac{E_{forecast}-E_{c\lim ate}}{E_{perfect}-E_{c\lim ate}} \quad (4)$$

Where $E_{c\lim ate}$ is the expected expenses associated with using the climatological data. $E_{c\lim ate}$= oL_u + min(oL_p, C). o is the climatological frequency of the event through calculating the buoy climate data. $E_{forecast}$ is the expected user expense of a forecast system. $E_{forecast}$ = h(C+L_u) + fC+ m(L_p + L_u). $E_{perfect}$ is the minimum expense of a user, given a perfect forecast system that provides accurate predictions for the occurrence and nonoccurrence of a particular event. $E_{perfect}$ =o (C+ L_u).

3.3.2 Relative operating characteristics (ROC)

ROC measures the ability of the forecast to discriminate between the events and non-events. ROC curve is plotted using hit rate with false alarm rate against a set of varying probability thresholds.

Hit rate = hit/(hit +miss)
False alarm rate=false / (false + correct rejection). (5)

The area under the curve defined as the ROC area is a useful summary measure of a forecast skill. ROC area is 1 which means the perfect forecast. Only the ideal deterministic forecast can reach it. ROC is 0.5 which means no forecast skill. A good ROC indicates by the curve which is close to the upper left corner (low false alarm rate and high probability detection) (see Fig. 7). The ROC can be considered as a measure of potential usefulness and is a good companion of the reliability diagram.

4. RESULTS

Below results for the various validation parameters are presented. BBS is calculated using five days forecasts, the others are obtained for the day 5 forecast.

4.1 Capacity of the ensemble forecast system and reliability of its forecast

In Fig. 2, BSS is plotted for the wind and wave field, the reference is the climate field. The threshold values are 2m, 4m, 6m and 8m for H_S and 10m/s, 14m/s, 17m/s and 20 m/s for U_{10}. The skill scores are rather good for H_S > 2m, 4m and U_{10}>10m/s and 14m/s. But the system lost predictability after Day 4 for 6m and 8m and with all forecast range for U_{10}> 17m/s and 20m/s.These results are mainly caused by a lack of sufficient observation data. We therefore do not discuss Hs>6m, 8m and U_{10}>17m/s, 20m/s in cost-lost analysis and reliability diagram. Fig. 3 are the BSS plots for wave and wind fields, but the reference is the operational NCEP run. All BSS increase over the forecast days. This means the ensemble forecast has higher forecast skill than the deterministic operational forecast.

Fig. 4 and 5 are the reliability diagrams for day 5 of wave height (H_S> 2m, 4m) and wind field (U_{10}> 10m/s, 14m/s). The forecast probabilities are divided into 11 ranges from 0 to 1.0 increasing by 0.1. The observed relative frequency is defined the fraction of the observed events against the total number of the forecast events in the same probability class. The striking results from the diagrams are the “no resolution” lines which are very low, less than 0.1 for wave height and 0.02 for wind fields. These show that this ensemble wave forecast system has very high forecast capacity. It could catch nearly all events occurred. In the low probability, the forecast system is under-forecast. In the high probability, the forecast system is over-forecast. The probability 0.9 and 1 for U_{10}>14m/s are obvious abnormal caused by less observation data.

4.2 Cost-loss analysis of the ensemble forecast system

Fig. 6 is the relative economic value as a function of the cost-loss ratio of the ensemble wave forecast and the operational NWW3 forecast. The benefit of the ensemble forecast system is apparent in these four cases. The economic value of the ensemble system is larger than that of operational run at each ratio. The economic value 0 means the value calculated using the climatology data and 1 means the value using the perfect forecast system. The economic value could be obtained from a larger range of cost-loss ratio using the ensemble forecast than using the operational forecast.

Analyzing the relative operating characteristics (ROC) for different thresholds could be used to perceive the performance of ensemble forecast system (Fig. 7). ROC areas are over 0.85 for H_S>2m and U_{10}>10m/s and 0.98 for H_S>4m and U10>14m/s.

5. DISCUSSION AND CONCLUSION

The ensemble forecast system has better forecast skill than the deterministic operational forecast after the probabilistic improvement. The EGOWaFS system has good forecast capacity which could catch most forecasted events. The ensemble system is under forecasting in low probability and over forecasting in high probability (Fig. 4). The sharpness diagram (histogram) in this figure clearly identifies that the spread generated by the ensemble is too narrow. Furthermore higher wind speeds are biased high, and, not surprisingly, so are higher wave heights. This behavior is well known from deterministic validation of the control run; see for instance, Bidlot et al. (2007), where wind speed and wave height biases are known to drift upward with forecast time. At the five day forecast, all such biases are generally positive. Note that the underestimation of the spread in wave heights in EGOWaFS is not unique to this ensemble system. It is also observed in the ECMWF ensemble forecast system (see Saetra and Bidlot, 2004). A possible explanation for the narrowness of the spread can be found in the design of the ensemble. By perturbing the forcing only, only wind sea perturbations are generated. Swells older than five days are essentially identical in all members of the ensemble. This is caused by the use of identical initial conditions in all ensemble members. Swells therefore do not contribute to the spread in the ensemble, whereas there clearly is uncertainty in the swells at the time of their origination. Swell contributions to the ensemble spread are therefore by definition underestimated. To assess if this can explain the spurious narrowness of the ensemble wave height distribution, we intend to isolate wind seas in the ensemble and in the data, and perform the same validation based on wind seas only.

An additional problem with the validation study presented here is the general lack of data. First, the 10 month period considered here is too short to provide a sufficient sample size of extreme wave conditions. Second, buoys used in this verification are located in the Northern Hemisphere, they can not represent the Southern part. Alternatively, altimeter data for several years should be used in the validation. We intend to use the altimeter data for such a validation study in the near future.

6. REFERENCES

Bidlot J.R., Li J.G., Fauchon M., Chen H.S., Lefevre J.M., Bruns T., Greenslade D., Ardhuin F., Kohno N., Park S., Gomez M., 2007: Inter-comparison of operational wave forecasting systems. In 10th international workshop on wave hindcasting and forecasting.

Chen, H. S., 2006, An ensemble global ocean wave forecast system, NOAA THORPEX PI Workshop, Camp Springs, MD.

Jolliffe I. T. and Stephenson D. B., 2003, Forecast Verification, Wiley.

3). Murphy, A.H., 1973, Hedging and skill scores for probability forecasts. J. App. Met., 12, 215-223.

Saetra O. and Bidlot J.R., 2004, Potential benefits of using probabilistic forecasts for waves and marine winds based on the ECMWF ensemble prediction system. Weather. and Forecasting. 673-689.

Tolman H . L., Balasubramani B., Burroughs L.D., Chaliko D.V., Chao Y.Y., Chen H. S. and Gerald V.M., 2002, Development and implementation of wind-generated ocean surface wave models at NCEP. Weather and Forecasting, 17, 311-333.

Tolman, 2003, Treatment of unresolved islands and ice in wind wave models. Ocean Modeling, 5(2003), 219-231.

Toth, Z. and Kalnay, E., 1997. Ensemble forecasting at NCEP and the breeding method. Mon. Wea. Rev., 125, 3297-3319.

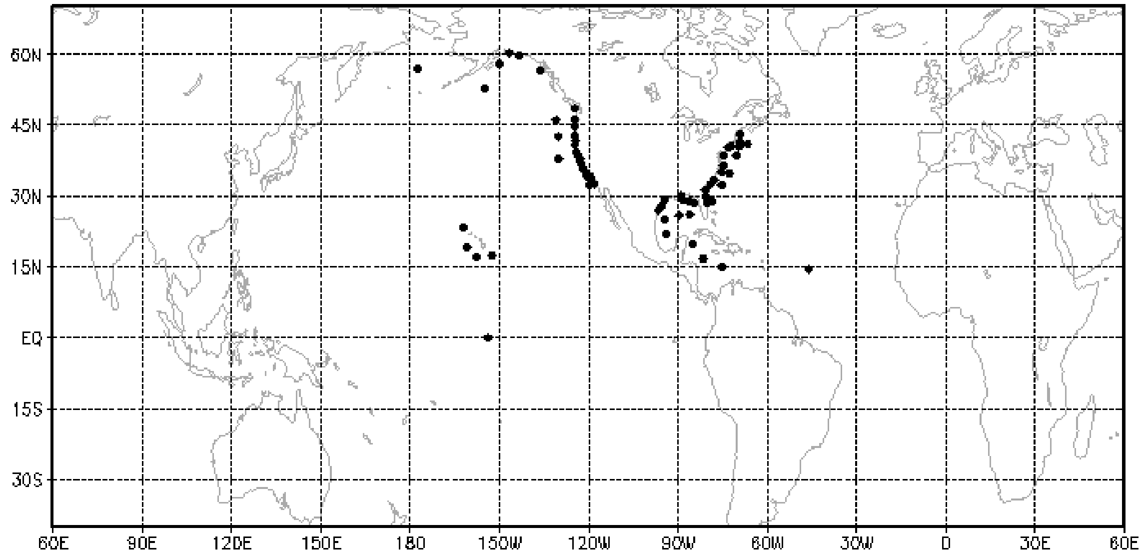


FIG. 1. Global buoy locations used in the ensemble wave forecast verification

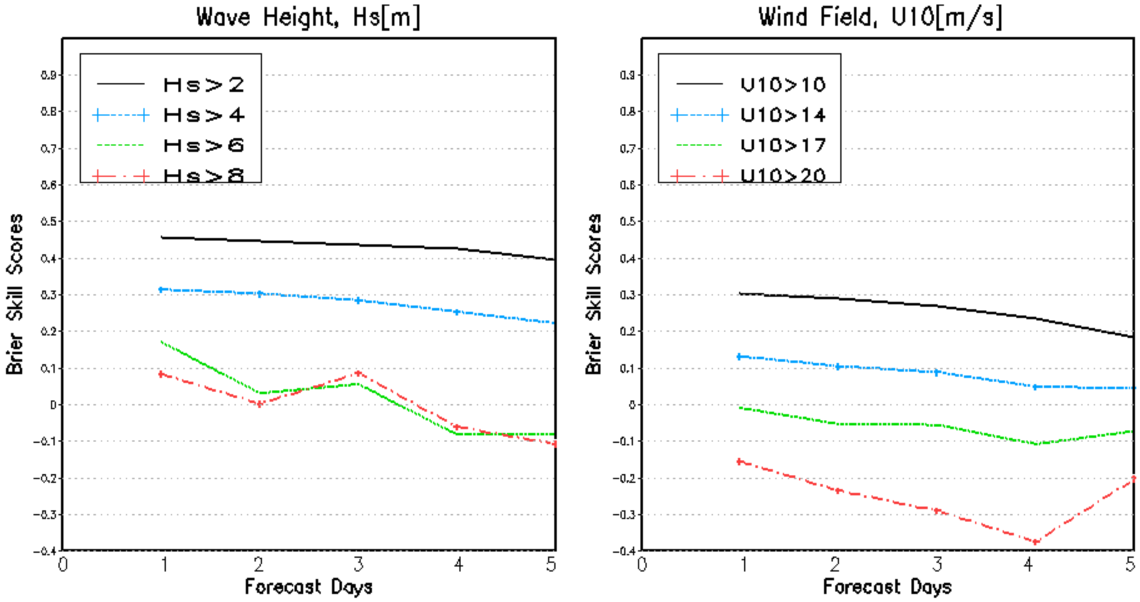


FIG. 2. Brier Skill Scores using the wave and wind climate data as the reference

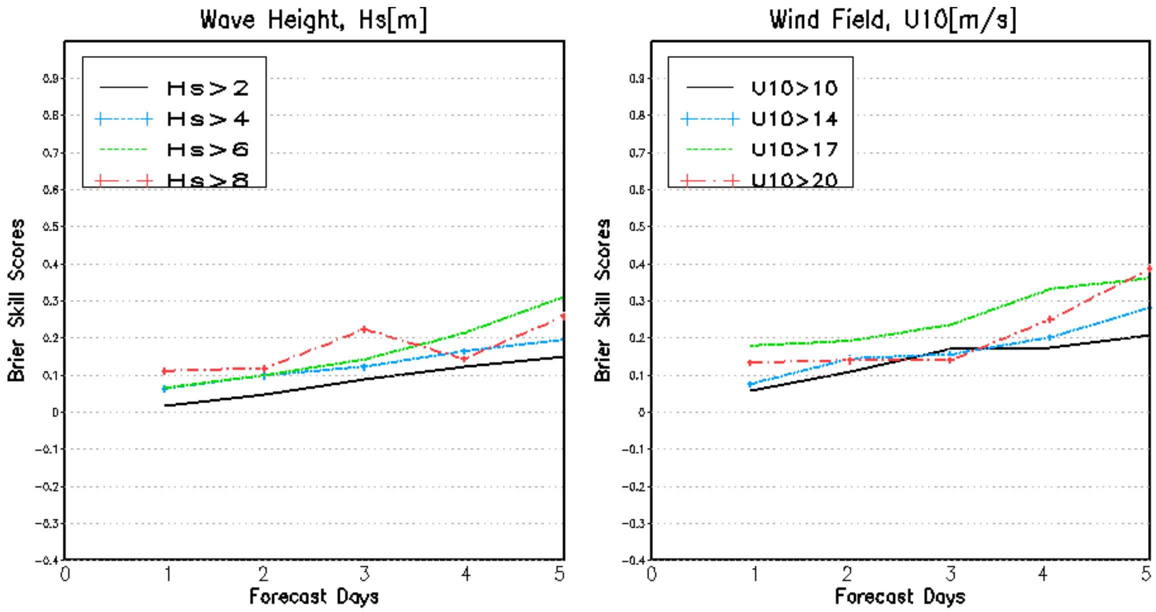


FIG. 3. Brier Skill Scores using the NWW3 operational run as the reference

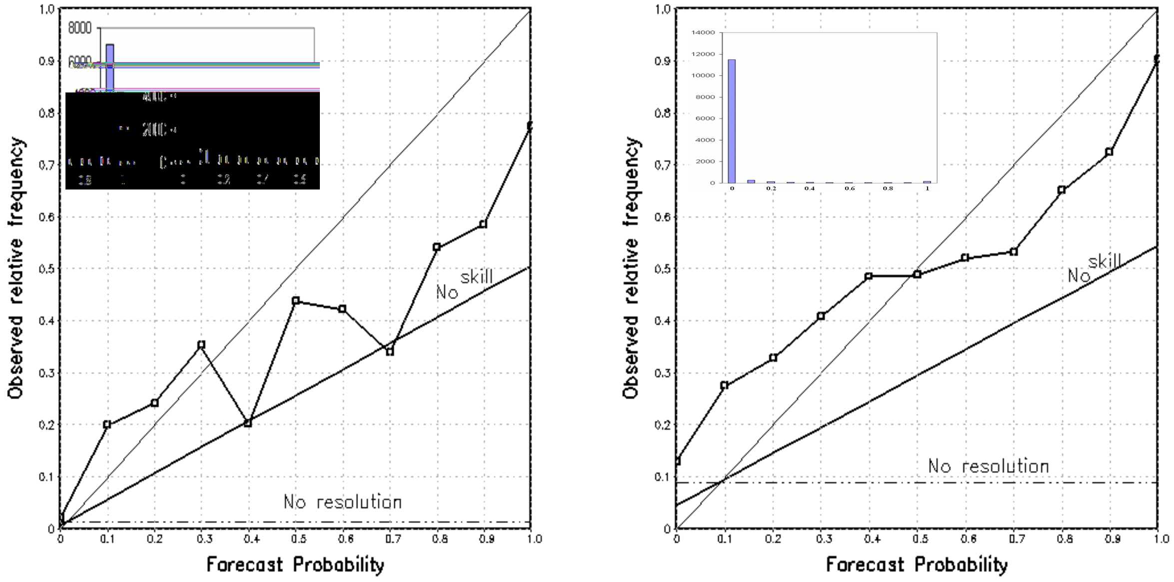


FIG. 4. Day 5 reliability diagram for wave height. Left: $H_S > 2$m, Right: $H_S > 4$m.

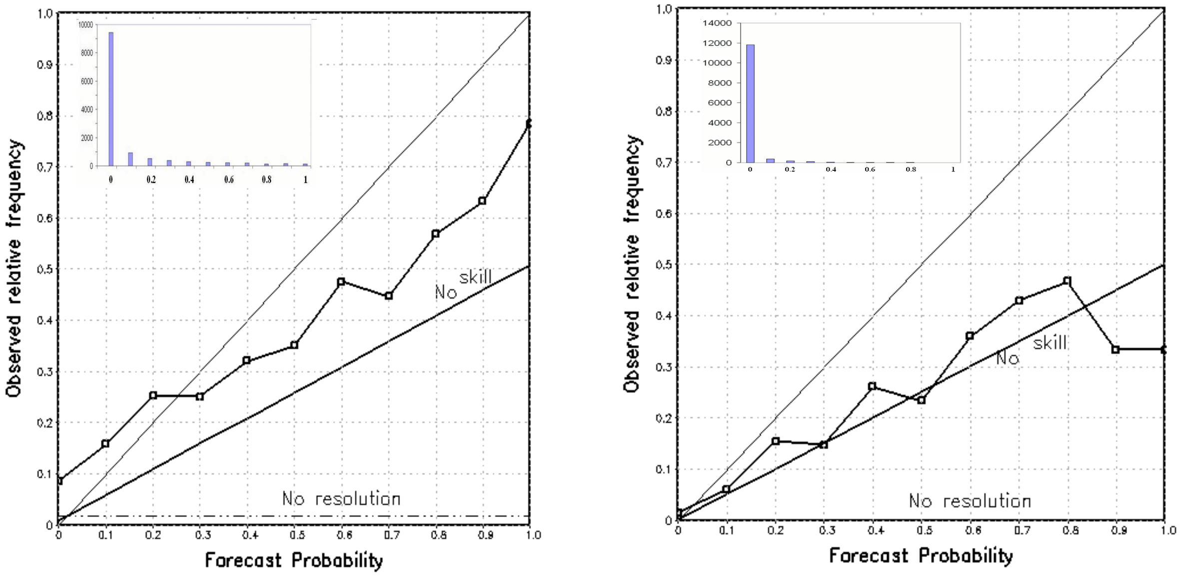
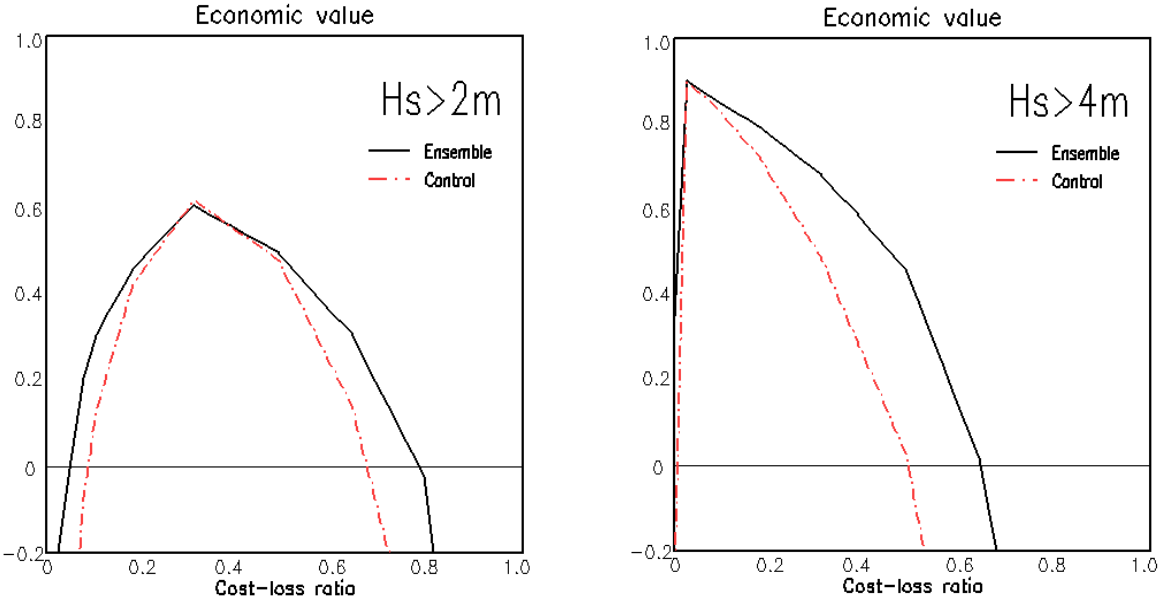


FIG. 5. Day 5 reliability diagram for wind field. Left: U_{10}>10m/s, Right: U_{10}:14m/s.



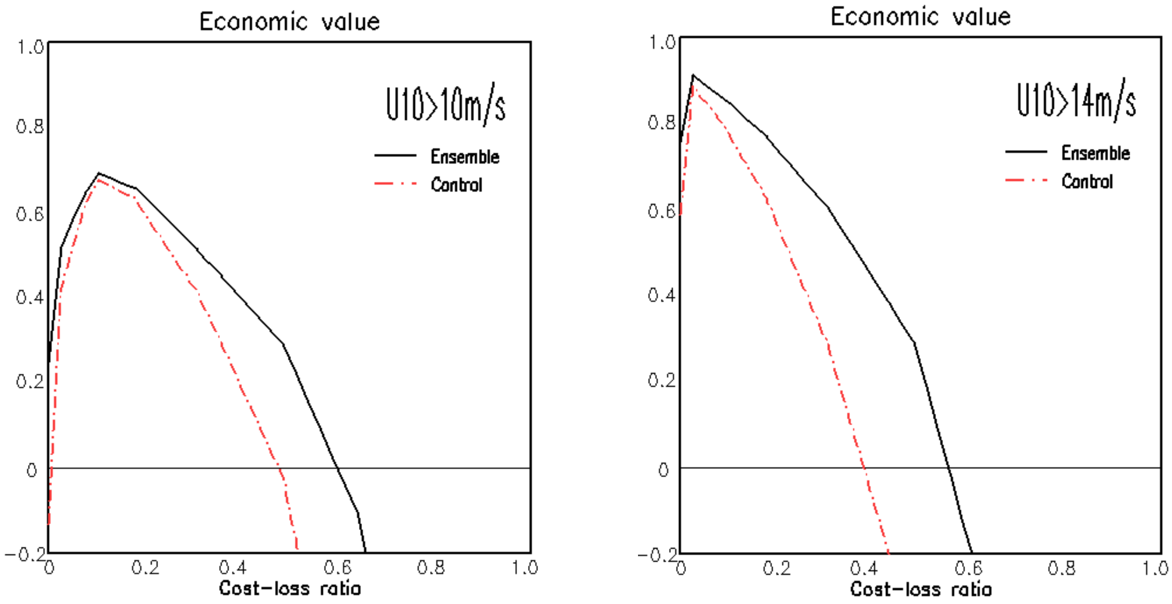


FIG. 6 Day 5 economic value for wave (H $_S$ >2m and H $_S$ >4m) and wind (U10 >10m/s and U10 >14m/s)

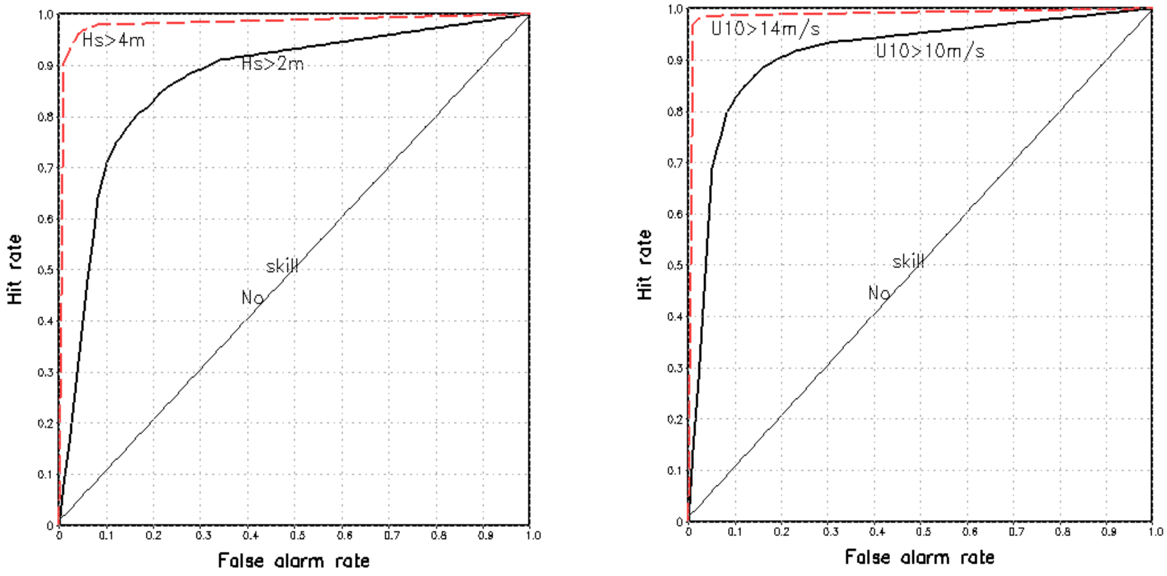


FIG. 7 Day 5 relative operating characteristics (ROC) curve for wave (H $_S$ >2m and H $_S$ >4m) and wind (U10> m/s and U10 > 14m/s)