Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Ocean Engineering 32 (2005) 1841–1863

OCEAN ENGINEERING

www.elsevier.com/locate/oceaneng

Forecasting wave height probabilities with numerical weather prediction models

Mark S. Roulston[a,b,*], Jerome Ellepola[c], Jost von Hardenberg[b,d], Leonard A. Smith[a,b]

[a]*Pembroke College, Oxford University, OX1 1DW, UK*
[b]*Centre for the Analysis of Time Series, London School of Economics, WC2A 2AE, UK*
[c]*Shell Global Solutions International B.V., 2596 HR The Hague, The Netherlands*
[d]*CIMA, University of Genoa, 1-17100 Savona, Italy*

Received 20 August 2004; accepted 19 November 2004
Available online 7 April 2005

Abstract

Operational weather forecasts now allow two week probabilistic forecasts of wave height. This paper discusses methods for generating such forecasts from numerical model output from the European Centre for Medium Range Weather Forecasting Ensemble Prediction System. The ECMWF system produces Monte Carlo style forecasts of wave statistics out to a lead time of 10 days. This ensemble of forecasts is designed to reflect uncertainty in current global weather conditions. In this paper a method for post-processing the ensemble forecasts of wave statistics is described and demonstrated using the significant wave height forecasts for four locations of interest to the offshore industry. The post-processing results in ensemble forecasts which have improved reliability, and which are better predictors of the expected error in the forecast.

Keywords: Wave forecast; Probability; Ensemble; Numerical weather prediction

1. Introduction

Wave forecasting is now an integral part of operational weather forecasting at several weather forecasting centres (Bidlot and Holt, 1999). The incorporation of wave models

* Corresponding author. Address: Department of Meteorology, The Pennsylvania State University, URL: www.lsecats.org, PA 16802, USA. Fax: +1 814 865 3663.
E-mail address: roulston@met.psu.edu (M.S. Roulston).

0029-8018/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.oceaneng.2004.11.012

into numerical weather prediction models can improve atmospheric forecasts by allowing the transfer of momentum between the ocean's surface and the atmosphere to be better modelled (Saetra and Bidlot, 2002). In addition, the forecasts of the wave fields that are generated are valuable forecast products in their own right. These wave models model the evolution of the two-dimensional wave spectra at each gridpoint. The spectra can be used to provide forecasts of sea state statistics such as significant wave height, swell height and windsea height, as well as wave direction and predominant wave period. For example, short to medium range (1–10 days) forecasts of wave conditions can assist ship routing and the scheduling of offshore operations. The European Centre for Medium Range Weather Forecasting has further enhanced the potential value of wave forecasting by incorporating a wave model into their 'ensemble prediction system'. The ECMWF ensemble prediction system (EPS) was introduced in the 1990s in an attempt to quantify state-dependent uncertainty in medium range (3–10 days) forecasts. State-dependent uncertainty refers to the component of uncertainty that depends upon the current stability conditions in the atmosphere (Palmer et al., 1994). At present, the ECMWF EPS produces an ensemble of 51 global 10-day forecasts, which are produced daily using a version of the ECMWF numerical weather prediction model. One forecast in this ensemble is initialized using the best estimate of the current state of the atmosphere. The remaining 50 are initialized using initial conditions constructed by perturbing the best estimate, to reflect the uncertainty in this best estimate, and model its impact on the subsequent forecast. The idea is to provide a crude Monte Carlo style sample of the likely behaviours of the model. More detailed discussions of ensemble forecasting at ECMWF can be found in the meteorology literature (Molteni et al., 1996; Palmer, 2000).

Since the ECMWF wave model is coupled to the atmospheric model, there is a forecast of the global wave field associated with each member of the ensemble. The wave model predicts the directional wave spectra at each grid point but only about two dozen summary statistics are archived (Bidlot, 2001). These statistics include quantities such as significant wave height, significant swell height and significant height of locally generated wind waves.

It must be stressed that ensemble forecasting is largely an attempt to model one source of forecast uncertainty—initial condition error. In recent years, ECMWF have included 'stochastic physics' in an attempt to improve the models estimation of other sources of uncertainty (Buizza et al., 1999) and other researchers have introduced ensembles consisting of forecasts from different models (Evans et al., 2000). But, regardless of how the ensembles are constructed, ensemble forecasts generated by numerical modelling alone can only ever provide a partial description of all the sources of uncertainty that affect a numerical weather prediction. To provide an assessment of the total uncertainty associated with a forecast, and indeed to provide a probabilistic forecast suited to decision making under uncertainty, some type of statistical post-processing of the ensemble forecast is required. This post-processing will incorporate information about past verifications and the accuracy of past forecasts.

In this paper, one post-processing technique consisting of linear regression and 'best member dressing' is applied to ECMWF ensemble forecasts of significant wave height at four locations shown in Fig. 1. The sites selected were Bonga, off the coast of Nigeria, Maui off New Zealand and the Campos and Santos Basins off the coast of Brazil.

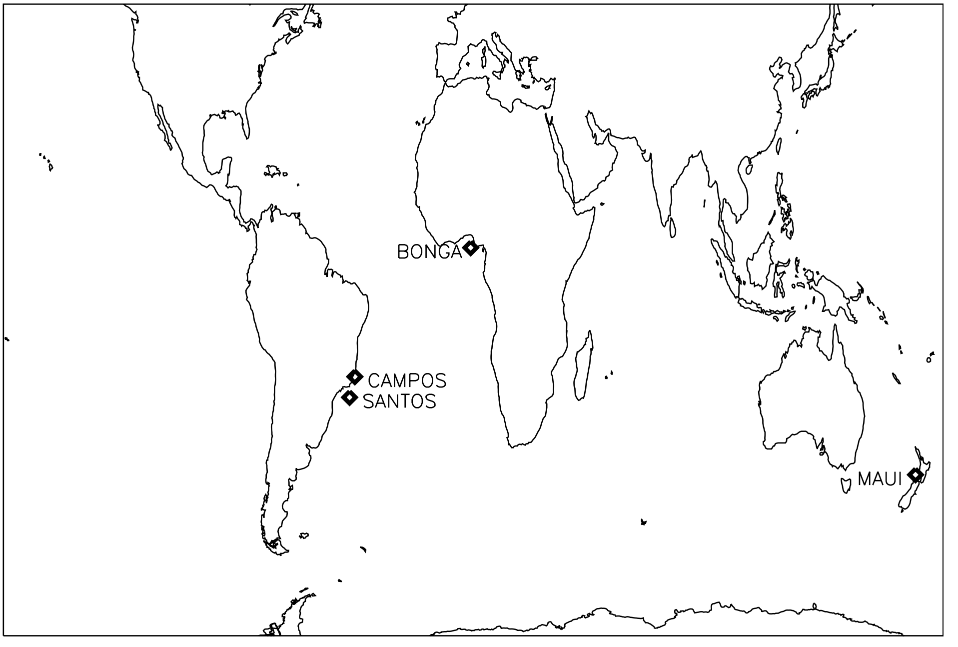

Fig. 1. The locations of the four sites studied in this paper. Offshore oil extraction is underway at these four sites.

Offshore oil and gas exploration and production is underway in all of these locations and the skill of wave forecasts for these places is thus of particular interest to the offshore energy industry. The aim of this paper is to both illustrate the post-processing technique through an application, and also to demonstrate that, with appropriate post-processing, the ECMWF ensemble prediction system can provide skillful and reliable probabilistic forecasts of wave fields at lead times of more than one week.

2. Method

2.1. Forecast regression

Wave forecasts deviate from the subsequent observations due to model imperfections and uncertainty in the initial conditions. Certain types of model imperfections—such as systematic biases—can be corrected for by fitting past forecasts to observations using linear regression. Furthermore, it has been shown that applying a linear regression to a finite size ensemble leads to better estimates of parameters of the forecast uncertainty distribution, even in the perfect model case (Leith, 1974). The first step in the post-processing of ensemble forecasts is thus to apply a linear transformation of the form

$$e'_{t,l,q,i} = m_{t,l,q} e_{t,l,q,i} + c_{t,l,q} \quad (1)$$

where $e_{t,l,q,i}$ is the ith ensemble member of the ensemble forecast made on day t for lead time l and for quantity q, $e'_{t,l,q,i}$ is the transformed ensemble member and $m_{t,l,q}$ and $c_{t,l,q}$ are the coefficients obtained by performing a least squares fit between past ensemble means and past verifications. The ensemble mean of the forecast on day t for quantity q at lead time l is given by

$$\langle e\rangle_{t,l,q} = \frac{1}{N}\sum_{i=1}^{N} e_{t,l,q,i} \tag{2}$$

where N is the size of the ensemble. In this paper the ensemble means and verifications from the two month period immediately preceding day t (making an allowance for the lead time) were used in the linear regression. The subscripts t and l on the coefficients m and c emphasize the fact that these coefficients were updated for each forecast, always using the previous two months, and that each lead time, l, had its own coefficients.

2.2. Dressing

The technique of dressing is an attempt to account for sources of error in the forecast not accounted for by the way that an ensemble forecast is constructed. *Best member* dressing is one way of estimating the distribution that should be used to 'dress' the ensemble. That is, the distribution that should be convolved with the distribution of raw ensemble members. An outline of the method is given here although a more complete description can be found in Roulston and Smith (2003).

Let v denote a forecast verification. This could mean the subsequent direct observation of a predicted quantity, or it could, as in this paper, refer to a subsequent *analysis*—an estimation of the current state of the ocean made by assimilating observations into a model. Given an ensemble of forecasts x_i ($i=1,\ldots,N$), and having defined a measure of distance such as root mean square distance, we can define the *best member* as the ensemble member which is closest to v. An expression for the probability that v will take on a particular value can be written

$$\mathrm{Prob}(v = x) = \sum_{i=1}^{N} \mathrm{Prob}(v = x|x_i \text{ is best}) \times \mathrm{Prob}(x_i \text{ is best}) \tag{3}$$

where the expression Prob(x_i is best) indicates the probability that the ith member is the best ensemble member. The expression in Eq. (3) is exact, and its exactness does not actually depend on the precise details of how the best member is defined. If the forecast contains no information, however, $\mathrm{Prob}(v=x|x_i \text{ is best})=\mathrm{Prob}(v=x)$ and so while Eq. (3) will be true, it would not be useful. If all the ensemble members are *exchangeable*, that is there is no a priori reason to believe that one ensemble member is more likely to be the best member than any other then it is possible to write

$$\mathrm{Prob}(x_i \text{ is best}) = \frac{1}{N} \tag{4}$$

This assumption of exchangeability would not be the case for an ensemble consisting of forecasts from different models (and is not strictly the case for the so called *control* member of the ECMWF ensemble, since it alone has not been perturbed from the best estimate of the initial condition). Nevertheless, in the current study the assumption was made. It is possible to modify the best member approach to exclude this assumption but this was not done in this study. The second assumption made regarding Eq. (3) is that

$$\mathrm{Prob}(v = x | x_i \text{ is best}) = \rho(x - x_i) \tag{5}$$

That is, the probability that $v=x$ given that a particular ensemble member is best is a probability distribution the shape of which is independent of the value of any x_i. This assumption is harder to justify on purely theoretical grounds. Making the assumptions given in Eqs. (4) and (5) allows Eq. (3) to be rewritten as

$$\mathrm{Prob}(v = x) \approx \frac{1}{N} \sum_{i=1}^{N} \rho(x - x_i) \tag{6}$$

The 'dressing' distribution, $\rho(\varepsilon)$, can be determined by selecting a parametric distribution, such as a Gaussian, and optimizing the parameters of the distribution using historical ensemble-verification pairs. This approach was adopted for univariate temperature forecasts by Smith et al. (2001). An alternative parameter free approach is to identify $\rho(\varepsilon)$ with the 'best member' error distribution. This is the distribution of the error (distance from the verification) on the best member of the ensemble. In this approach the best member must be identified as such in a high enough dimensional space to avoid projection effects leading to a misidentification of the best member and subsequent poor estimation of the best member error distribution (Roulston and Smith, 2003). In practice, this means that the best member of the ensemble should be identified using multiple forecast quantities, either in time and/or in space. Roulston and Smith used a method of 'false best members' to decide whether a given set of variables is too small, but there is, as yet, no theory for guiding this variable selection. For example, in this paper the best member of each ensemble was identified as the one with the lowest total squared error with respect to the subsequent analysis, where each forecast variable was normalized by its standard deviation in the ensemble. That is, the best member was defined as the one with the lowest value of χ^2 given by

$$\chi^2_{t,i} = \sum_q \sum_l \frac{(v_{t,l,q} - e'_{t,l,q,i})^2}{\sigma^2_{t,l,q}} \tag{7}$$

where

$$\sigma^2_{t,l,q} = \frac{1}{N} \sum_{i=1}^{N} (e'_{t,l,q,i} - \langle e' \rangle_{t,l,q})^2 \tag{8}$$

Note that the linearly transformed variables defined in Eq. (1) are used in Eqs. (7) and (8). The three subscripts signify that there is a separate value of σ for each forecast, lead time and forecast variable. The index l represents the lead times and $l=1,\ldots,10$. The index q represents the forecast variables and $q=1,\ldots,8$, where the 8 variables that contributed to the squared error are shown in Table 1. Only the single location for which forecasts were being produced was used in the process of identifying the best members. When the

Table 1
The wave model variables used to identify the best member of each ensemble forecast.

SWH	Significant wave height
PP1D	Peak period of 1d spectra
MWP	Mean wave period
SHWW	Significant height of wind waves
MPWW	Mean period of wind waves
SHPS	Significant height of swell
MPPS	Mean period of swell
WIND	10 m wind speed modified by wave model

duration of the forecasts is much longer than the forecast lead time of interest, one would select best members based on a shorter time scale than the full forecast lead time.

Once the best member of each of the past ensembles for which the subsequent analyses were available had been determined, the errors associated with these members were calculated by subtracting the value of the best member from the verification (in this case the analysis). New 51-member ensemble forecasts were then produced using the algorithm outlined below.

1. regress forecast using Eq. (1) with coefficients fitted using forecasts from the past two months
2. select a random member of the regressed ensemble
3. select a random best member error from the previous two months
4. add the selected error to the selected ensemble member to create a member of the ‘dressed’ ensemble
5. repeat steps 2–4 50 more times to create a new 51-member ensemble

Note that in steps 2 and 3 the members of the regressed ensemble and the best member error are selected *with replacement*. It should also be noted that for a forecast issued on a particular day, both the regression and the dressing procedures were performed using only forecasts from the past two months for which complete verifications were available. That is, forecasts issued in the period from 60 days before the particular day to 10 days before the particular day. This point is important because the number of past forecasts required to implement a post-processing or recalibration method affects its practical value. The existence of a substantial archive of past forecasts can enhance the value of a forecast model by allowing better estimates of its error statistics (Smith, 2003). The processed ensemble was constructed to have 51 members to allow a fair comparison with the raw ensemble. The processed ensemble could have a larger number of members. With 51 members and a relatively short archive of 50 error samples a processed ensemble with 2550 different members could in principle be constructed.

Fig. 2 shows a comparison of raw 51-member ECMWF ensembles with ‘regressed and dressed’ 51-member ensembles for significant wave height forecasts made on two different days for Maui off the coast of New Zealand. Notice that the post-processing tends to broaden the ensemble. Ideally, of course, the ensemble should be as narrow as possible, but not so narrow that it is not an accurate indicator of the uncertainty in the forecast.

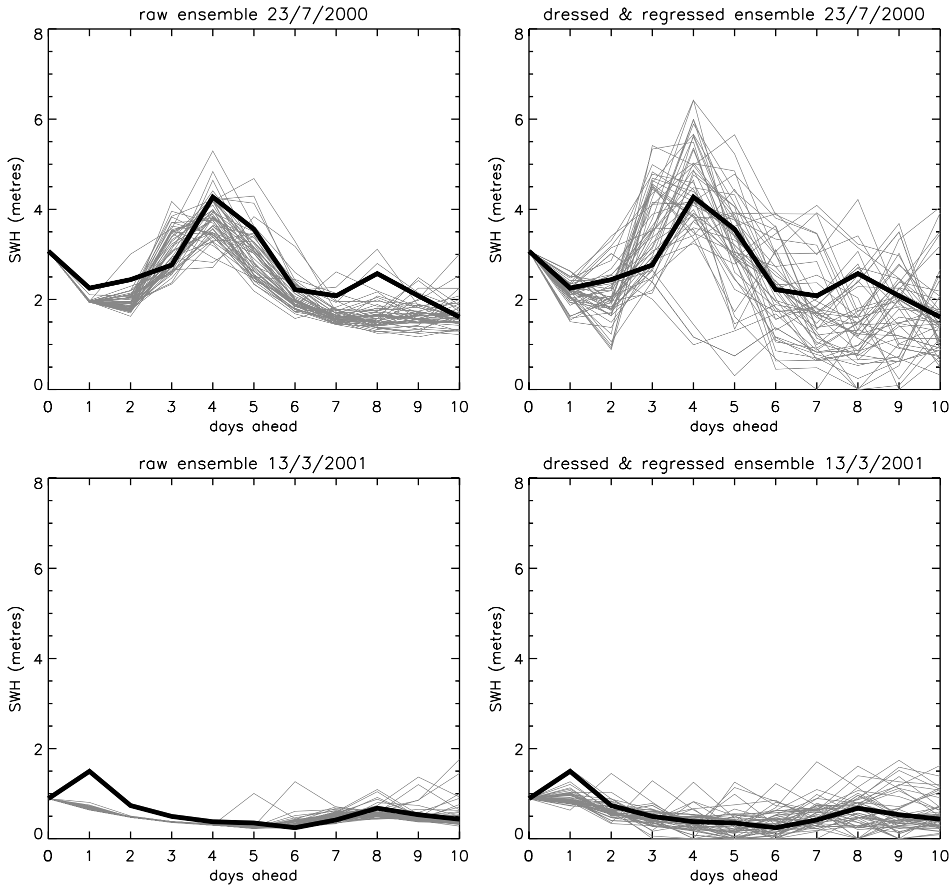

Fig. 2. An example of the impact of post processing on the raw 51-member ECMWF ensembles of significant wave height. The panels on the left show unprocessed ensemble forecasts for SWH at Maui for two different days. The panels on the right show the corresponding regressed and dressed forecasts. The thick black curve is the corresponding analyses. Notice that post-processing widens the distribution of ensemble members, and in doing so makes it a better representation of the actual forecast error, although this cannot be properly demonstrated using individual forecasts.

2.3. Evaluating ensemble forecasts

When discussing the accuracy of single 'deterministic' forecasts[1] it is common to use linear correlation between the forecast and the verification as an indicator of forecast skill. The skill of the ensemble mean can be evaluated in the same way. If $\langle e \rangle_t$ is the ensemble mean of forecast t and v_t is the corresponding verification, then the coefficient of linear

[1] The word 'deterministic' refers to what are also known as 'point forecasts'. A single prediction, which may come from a model with 'stochastic physics' that might be interpreted as the mean, mode or median of the uncertainty distribution depending on how it was constructed.

correlation is given by

$$r = \frac{\sum_t (\langle e \rangle_t - \langle \bar{e} \rangle)(v_t - \bar{v})}{\sqrt{\sum_t (\langle e \rangle_t - \langle \bar{e} \rangle)^2} \sqrt{\sum_t (v_t - \bar{v})^2}} \tag{9}$$

where the overbar indicates averaging over forecasts.

The value of ensembles, however, extends far beyond the skill of the ensemble mean as a deterministic forecast. Ensembles are designed to provide information about the likely error in the forecast. One way of assessing this capability is the *spread-skill* relationship. There are several variations of how spread and skill are defined in this context but in the present work the following definitions were used.

$$\mathrm{SPREAD}_t = \frac{1}{N} \sum_{i=1}^{N} |e_{t,i} - \langle e \rangle_t| \tag{10}$$

$$\mathrm{SKILL}_t = |v_t - \langle e \rangle_t| \tag{11}$$

Examination of Eqs. (10) and (11) shows that the spread is the mean absolute difference between the ensemble members and the ensemble mean, and the skill is just the absolute error of the ensemble mean with respect to the subsequent verification. Spread-skill relationships are sometimes plotted for individual forecasts. This however leads to a large amount of scatter in the relationship, even in the case of perfect ensembles, because a single realization of the actual error of a forecast is a poor estimate of the expected error associated with the distribution. It can be shown that, even in the case of perfect normally distributed ensembles with a lognormal distribution of ensemble spreads, the spread-skill correlation cannot exceed 0.80 (Houtekamer, 1993). In the current study the forecasts were split into groups of 70 forecasts with similar ensemble spreads. The average absolute error associated with each group was then plotted against the average ensemble spread associated with the group.

The skill of ensemble forecasts is often decomposed into two components; *sharpness* and *reliability* (Murphy, 1997). The sharpness describes how narrow the ensemble forecast is, while the reliability describes whether the spread is a fair reflection of the forecast uncertainty. If an ensemble forecasting system is reliable then it is equally probable that the verification will fall into each of the intervals between the ranked ensemble members. This property can be evaluated using *rank histograms* (also called '*Talagrand Diagrams*'). Rank histograms are constructed by sorting the members of each ensemble into order (based on a single forecast quantity at a single lead time) and then building a histogram of the frequency with which the verification falls into each of the intervals between the ensemble members. In the case of a 51-member ensemble there are 52 intervals; 50 intervals between members and two open intervals at the ensemble extremes. A flat rank histogram is a necessary, although not a sufficient, condition of reliable forecasts (Hamill, 2001).

When estimating skill measures of forecasts, as when estimating the value of any statistic from data, the uncertainty of the estimate should itself be estimated. In the current study uncertainty estimates were made using a bootstrap resampling technique (Efron and Tibshirani, 1986). The original time series of forecast-verification pairs was broken into blocks, each of 20 days in length. The block length was chosen so that the forecasts in successive blocks would be uncorrelated. These blocks were then sampled, with replacement, to create a new time series of forecast-verification pairs of the same length as the original. The skill measure in question was then calculated for this time series of forecast-verification pairs. This process was repeated a number of times to produce a number of estimates of the skill measure. From this collection of estimates the mean estimate and its standard error were estimated.

3. Results and discussion

3.1. Impact of post-processing: skill in the mean

The linear correlations between ensemble means and the verification for significant wave height forecasts are shown in Figs. 3–6. From these figures it can be seen that the mean of the ECMWF ensembles has significantly higher skill than the persistence forecasts shown for comparison. Persistence forecasts simply assume that current conditions will persist into the future. It should also be noted that the post-processed (regressed and dressed) forecasts in the right-hand panels retain this skill. The post-processing method is not designed to improve the skill in the mean of the ensemble it is designed to give a better prediction of forecast *uncertainty*. The only case where post-processing introduces a significant change in the skill of the ensemble mean is in the Campos Basin (Fig. 5) where the means of the post-processed ensembles have a higher correlation skill score than the raw ensembles.

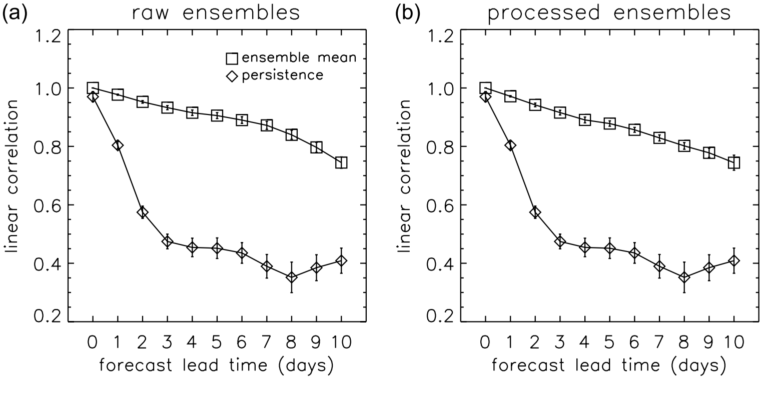

Fig. 3. The linear correlation between the ensemble means and the analyses for the Bonga site for the period March 1999 to December 2001 for significant wave height. The left panel is the correlation between the mean of the unprocessed ECMWF ensemble and the analyses, while the right panel is the correlation between the mean of the regressed and dressed ensemble and the analyses. The skill of a persistence forecast (assume current conditions persist) is shown for comparison. Note that post-processing has little impact on this measure of the skill of the ensemble mean, this is not surprising as the post-processing is designed to improve predictions of forecast uncertainty, not of the ensemble mean. The error bars are standard error estimates obtained using bootstrap resampling of 20 day blocks.

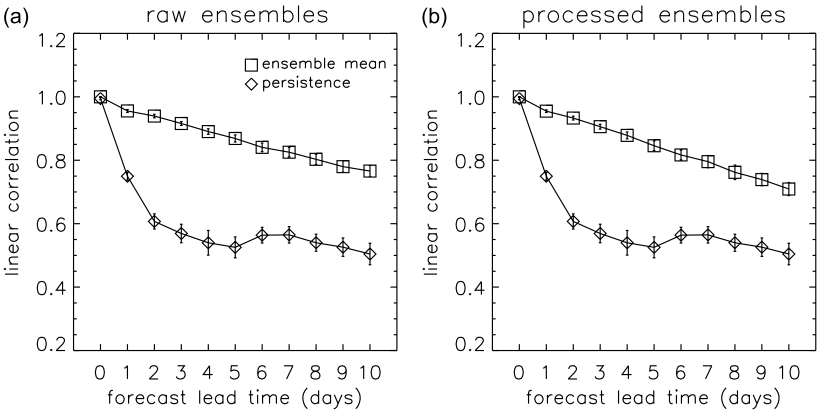

Fig. 4. As Fig. 3 but for Maui.

3.2. Impact of post-processing: skill in the distribution

As mentioned in the previous section, however, the value of ensemble forecasts lies in their ability to quantify forecast uncertainty rather than their ability to improve forecast skill in the traditional deterministic sense. Figs. 7–10 show the spread-skill relationships for the ensembles at the four locations at lead times of 1, 4 and 10 days. The top panels in each figure show the skill (averaged over 70 forecasts) plotted against the spread for the raw ensembles of significant wave height. The bottom panels show the same thing for the post-processed (regressed and dressed) ensembles. At all four locations the plotted points in the upper panels tend to lie above the dashed diagonal line implying that the spread of the raw ensembles underestimates the actual expected error. In all four locations the spread-skill relationship of the post-processed ensembles tends to lie close to the diagonal. In some cases, for larger spreads, the points appear to systematically fall below the diagonal implying that in these cases the post-processed ensemble spread slightly

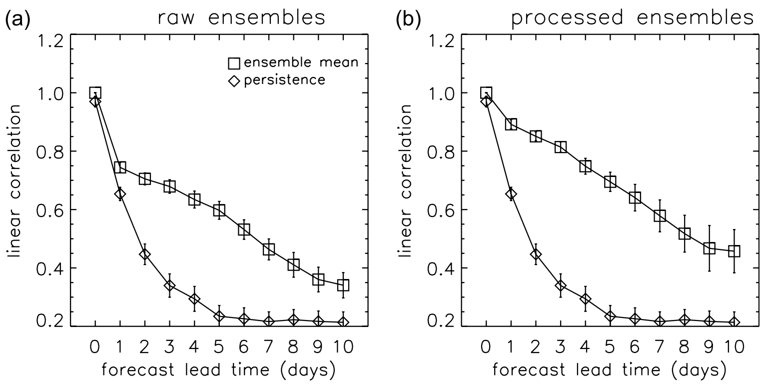

Fig. 5. As Fig. 3 but for the Campos Basin. In this case post-processing improves the skill of the ensemble mean at most lead times.

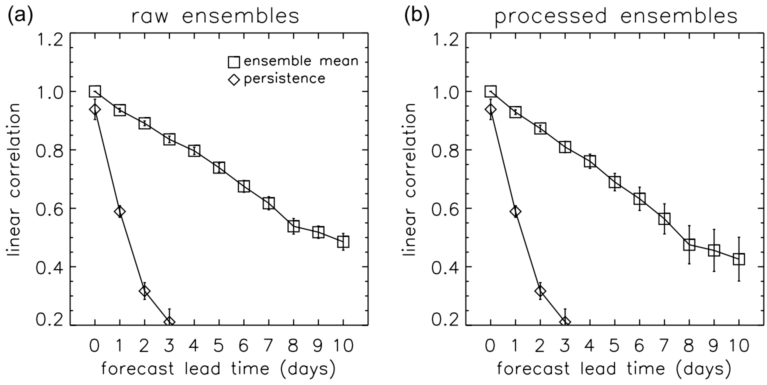

Fig. 6. As Fig. 3 but for the Santos Basin.

overestimates the actual expected error. This happens primarily at a lead time of 10 days. The overall effect of post-processing, however, is to increase the spread of the ensembles such that the mean absolute deviation of ensemble members from the ensemble mean is a good quantitative predictor of the expected absolute error between the verification and the ensemble mean. This is consistent with the ensemble being a good estimate of the distribution from which the verification is drawn, but for a further assessment of the reliability of the ensembles rank histograms were used.

Figs. 11–14 show rank histograms for the ensembles at the four locations at lead times of 1, 4 and 10 days. The top panels show the histograms for the unprocessed 51-member ECMWF ensembles while the bottom panels show the histograms for the post-processed 51-member ensembles. It is clear from the top panels that the verification falls outside the unprocessed ensemble far more often than would be expected for reliable ensembles. In a reliable 51-member ensemble one would expect the verification to fall outside the ensemble $2/52 \approx 3.8\%$ of the time. At lead times of 1 and 4 days the verification falls outside more than 30% of the time at Santos Basin and Maui and over 60% of the time at Campos Basin and Maui. At a lead time of 10 days the verification falls outside the ensemble more often than it should at Bonga, Maui and the Campos Basin, although the Santos Basin forecasts are consistent with reliable forecasts at 10 days. At the shorter lead times the verification falls both above and below the unprocessed ensemble too often. There is some asymmetry, however. In Fig. 12 for example it can be seen that at 4 days the verification tends to fall above the unprocessed ensemble far more often than it falls below. This is indicative of a forecast bias, which regressing the ensembles should mitigate. The bottom panels of Figs. 11–14 show that post-processing results in rank histograms that are close to uniform. Such rank histograms are a necessary condition of reliable ensembles.

3.3. Relative impact of regression and dressing

The post-processing method consists of two steps—regression and dressing. The relative impact of these two steps is examined in Figs. 15 and 16. From these figures it can be seen that regressing the forecasts, without dressing leads to ensembles, which do not

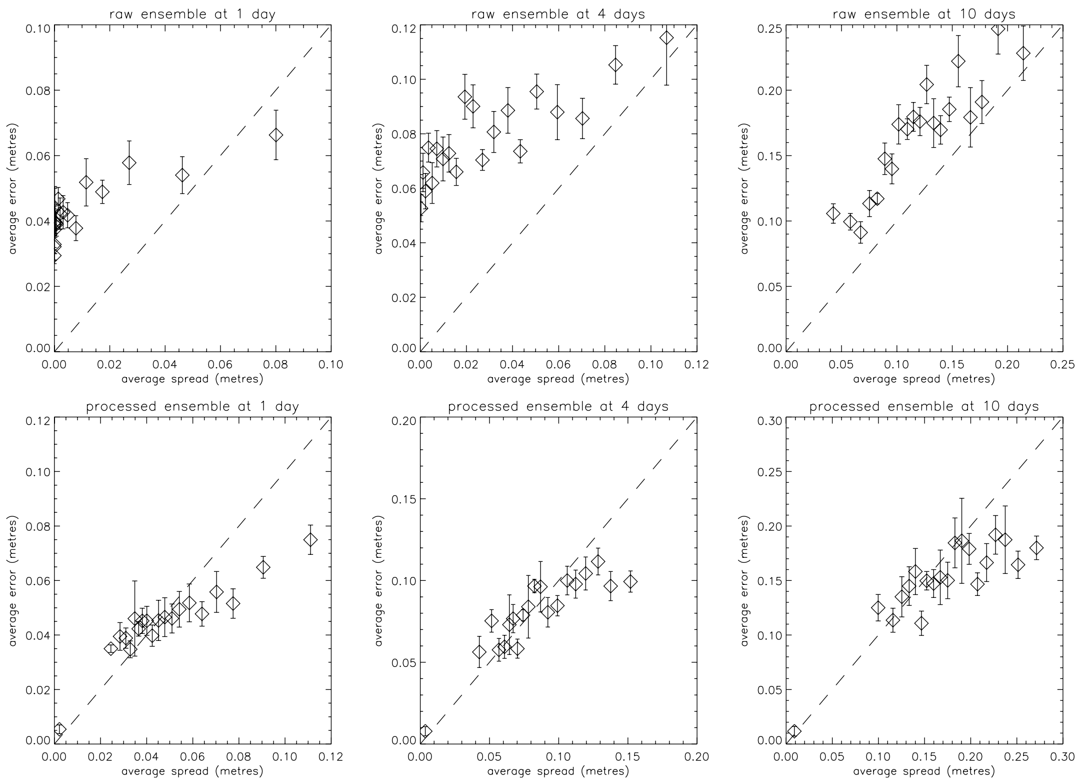

Fig. 7. A comparison of the spread-skill relationship of significant wave height forecasts at Bonga. Each point was obtained by calculating the average error of the ensemble mean of 70 ensembles between March 1999 and December 2001 with similar ensemble spreads. The dashed lines are the lines along which the points would be expected to fall if the ensembles were perfect. Points above this line correspond to cases when the ensemble spread underestimates the true error, points below this line mean that the ensemble spread is larger than the true error. The top panels show the spread-skill relationship for the raw 51-member ECMWF ensembles, while the bottom panels show the relationship for a 51-member ensemble obtained by regressing and dressing the raw ensemble. The error bars represent the standard error estimates obtained by bootstrap resampling. Note that the raw ensembles generally underestimate the error. The processed ensembles provide a better indication of the expected forecast error, although there is a tendency for them to overestimate the larger errors.

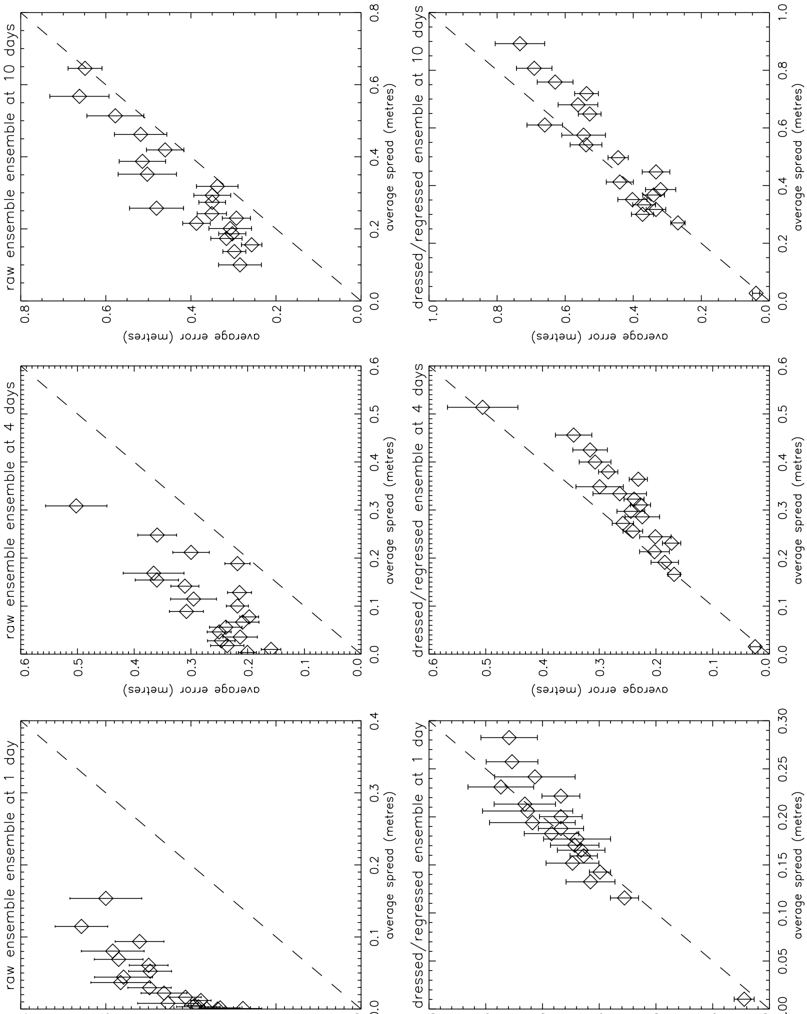

Fig. 8. As Fig. 7 but for Maui.

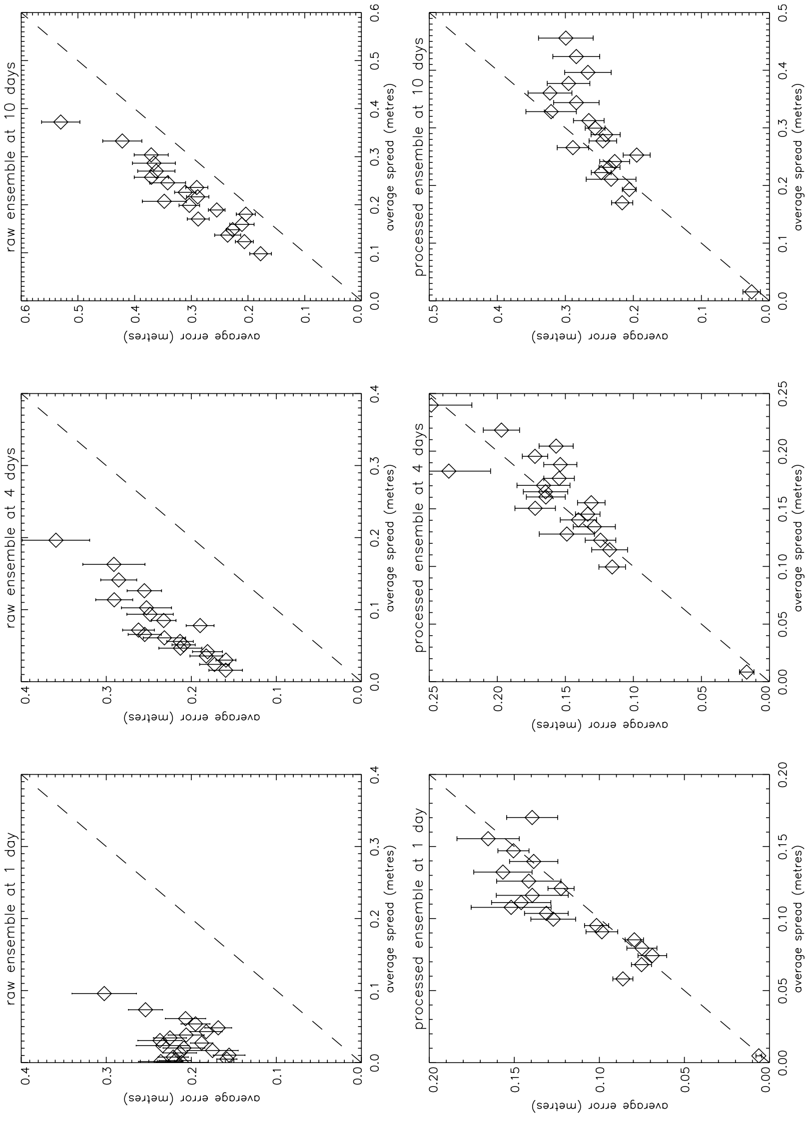

Fig. 9. As Fig. 7 but for the Campos Basin.

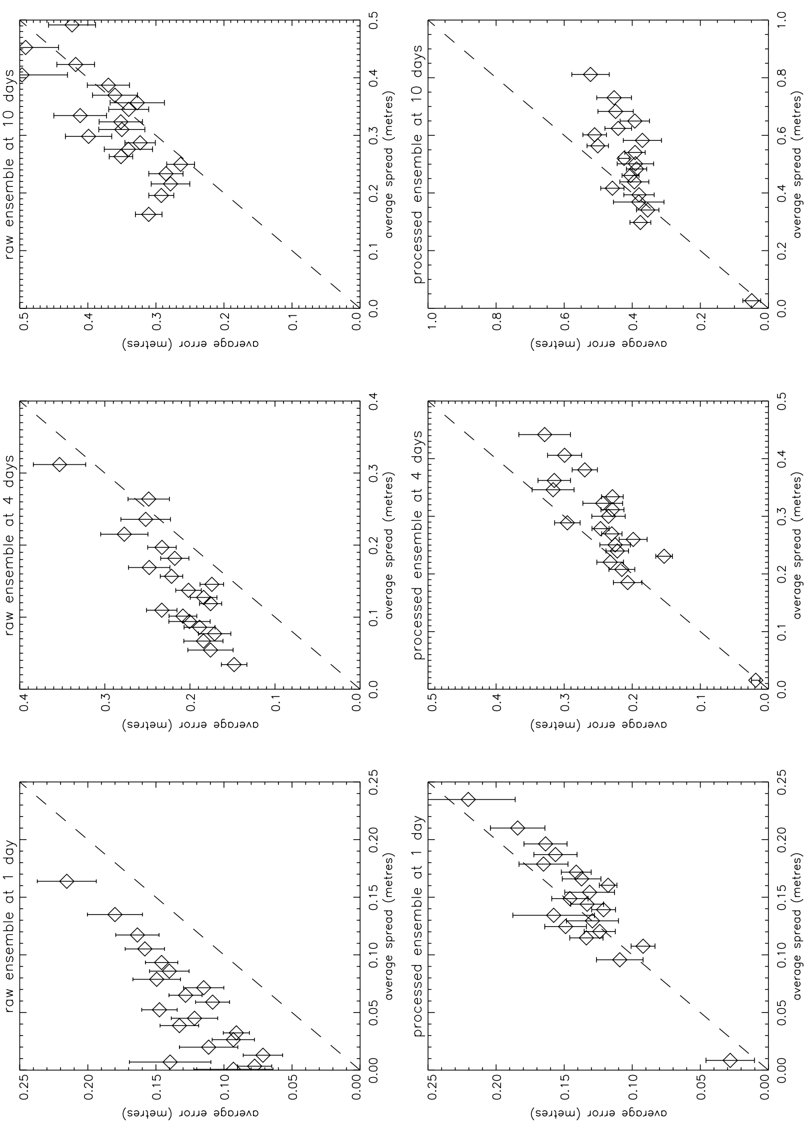

Fig. 10. As Fig. 7 but for the Santos Basin.

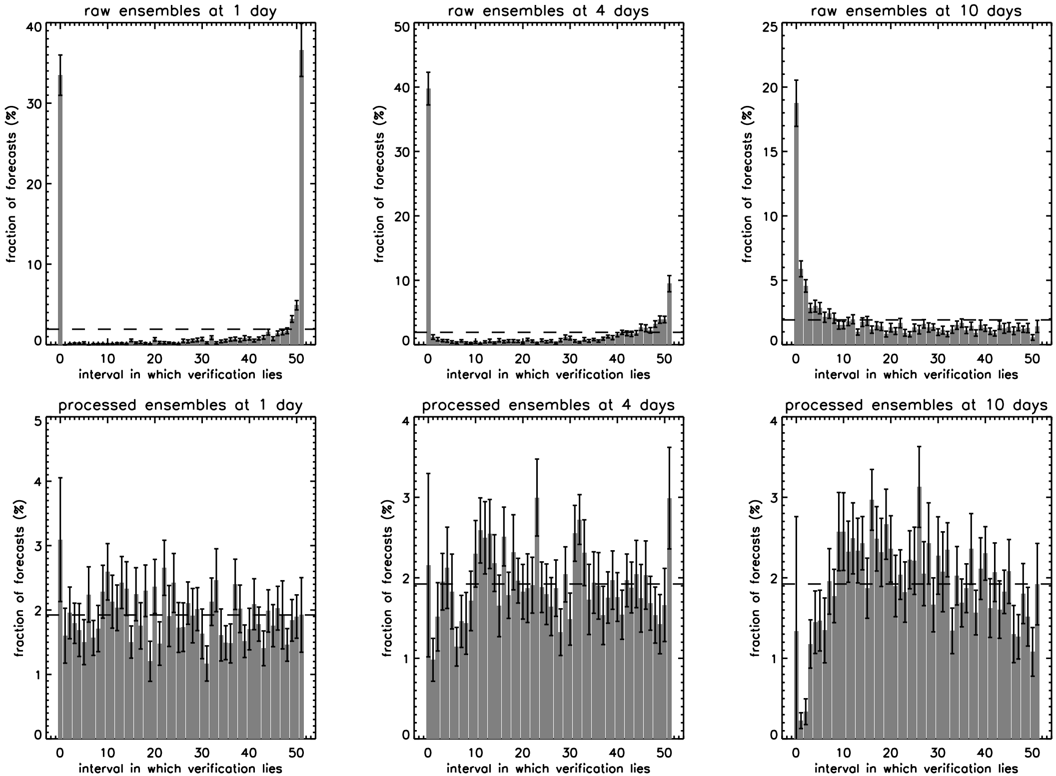

Fig. 11. A comparison of *rank histograms* of the 51-member unprocessed ECMWF ensemble and a 51-member ensemble constructed by regressing and dressing the ECMWF ensemble. The top panels show the rank histograms of the unprocessed ensembles at lead times of 1, 4 and 10 days. The bottom panels show the histograms of the processed ensemble for the same lead times. The histograms were constructed using forecasts for the period March 1999 to December 2001. The error bars are the standard error estimates obtained by bootstrap resampling the forecasts in blocks of 20 days. The dashed lines represent the fraction of forecasts that would lie in each interval in a flat rank histogram ($1/52 \approx 1.9\%$). It can be seen that the raw ensembles have rank histograms which are overpopulated at the extremes whereas the processed ensembles have histograms which are nearly flat—a necessary condition of reliable forecasts. Note that the verification lies outside the raw ensembles far more often that would be expected for reliable forecasts—70% of the time at 1 day compared to the expected $2/52 \approx 3.8\%$ of the time.

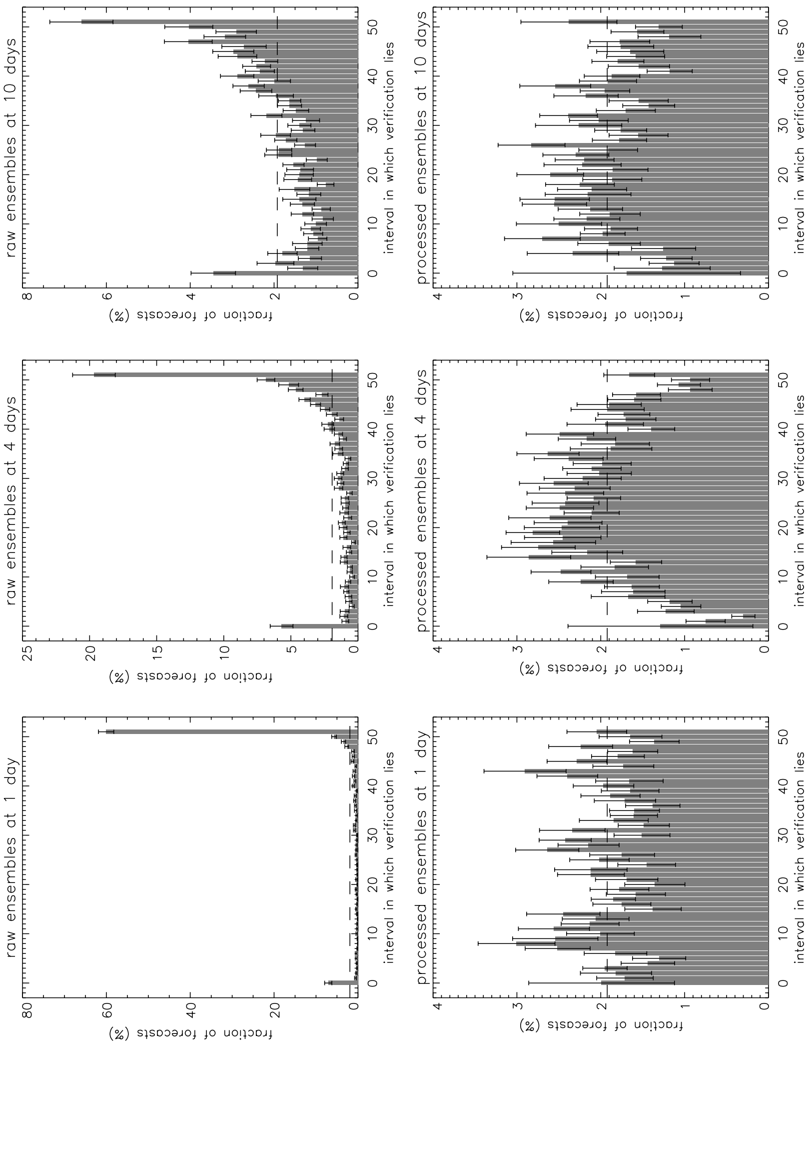

Fig. 12. As Fig. 11 but for Maui.

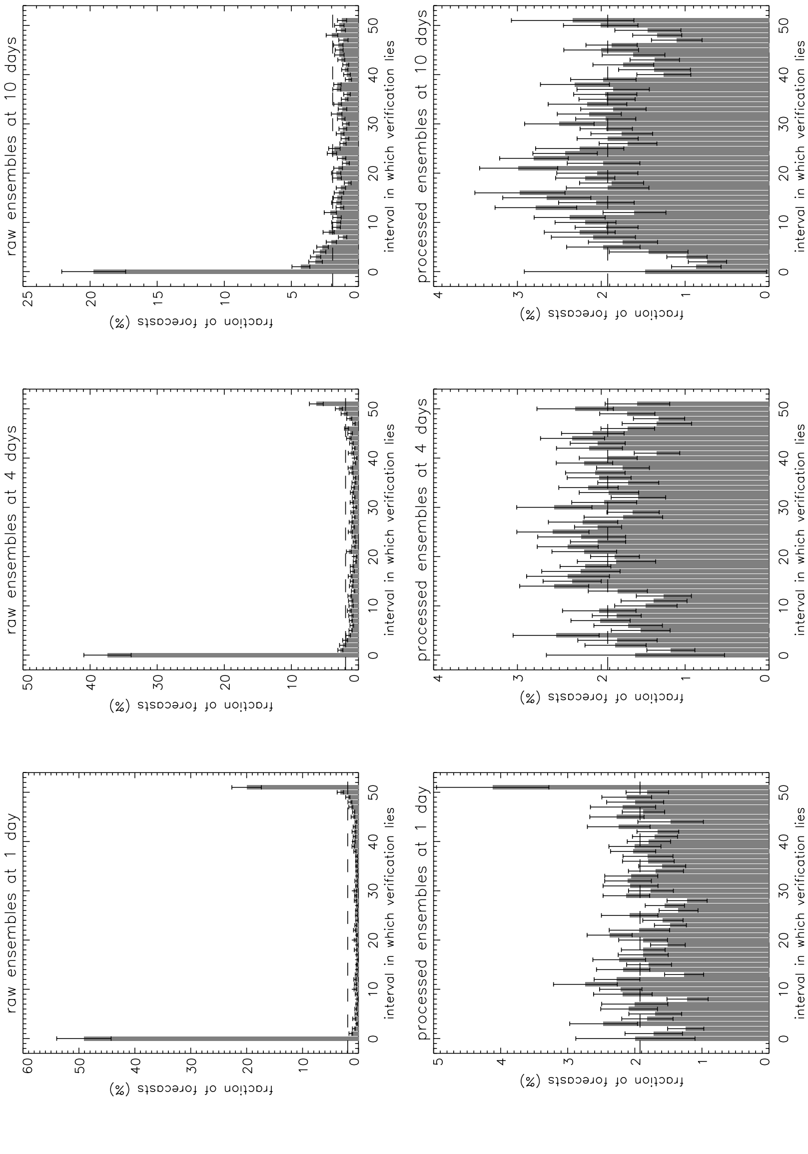

Fig. 13. As Fig. 11 but for the Campos Basin.

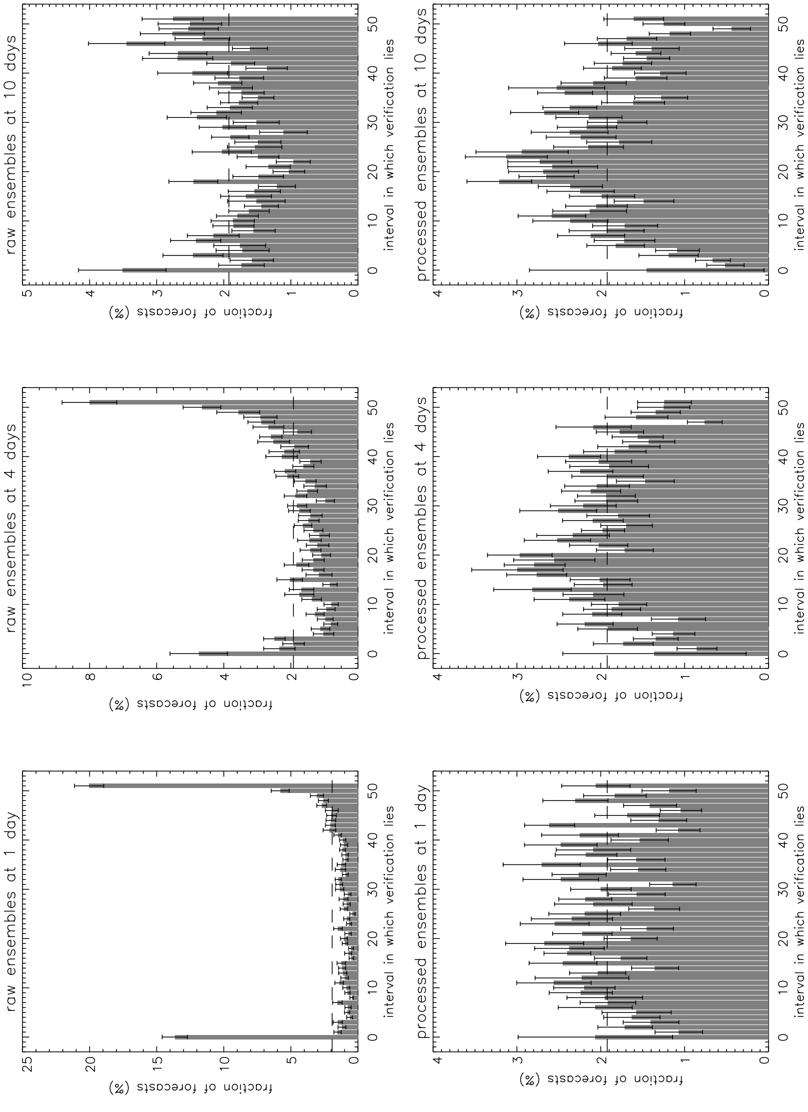

Fig. 14. As Fig. 11 but for the Santos Basin.

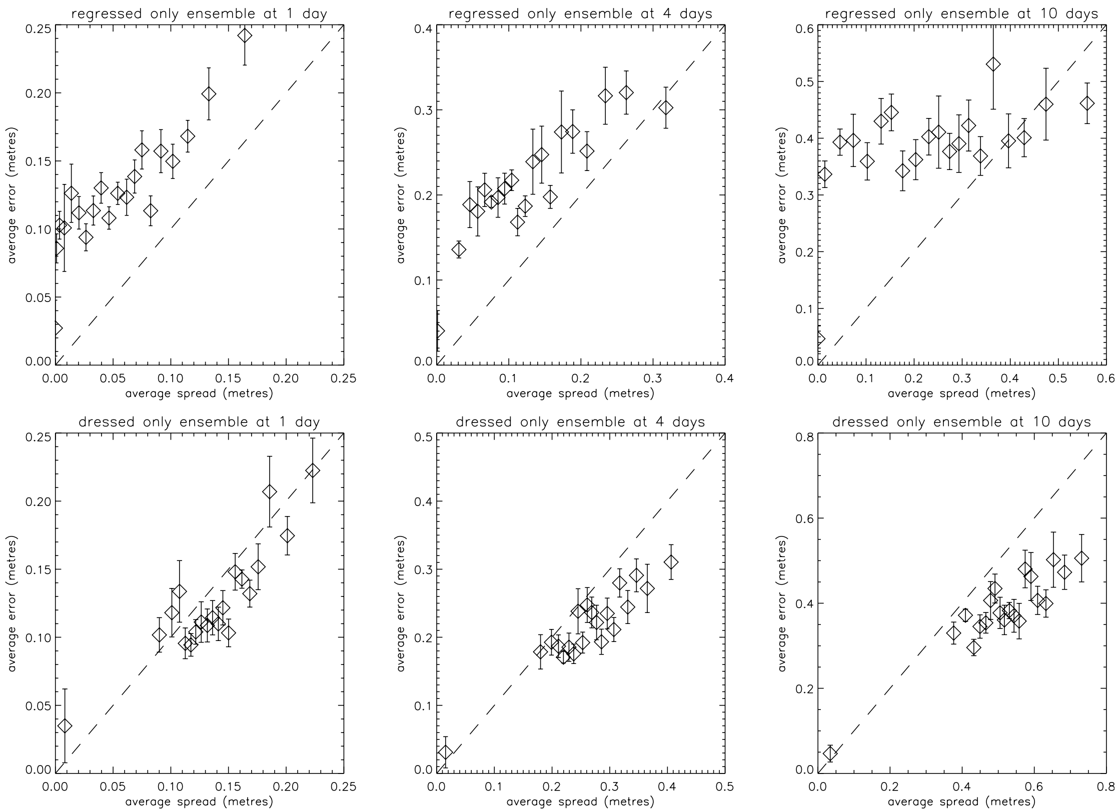

Fig. 15. A comparison of the spread-skill relationship of ensemble forecasts which have only been regressed but not dressed (top panels) with those which have been dressed but not regressed (bottom panels). The forecasts are for significant wave height in the Santos Basin. Notice that if the forecasts are regressed but not dressed the errors are greatly underestimated, even more than for the raw ensemble at 10 days (compare with Fig. 10). If the forecasts are only dressed, without first regressing them, the results are very similar to the full processing step.

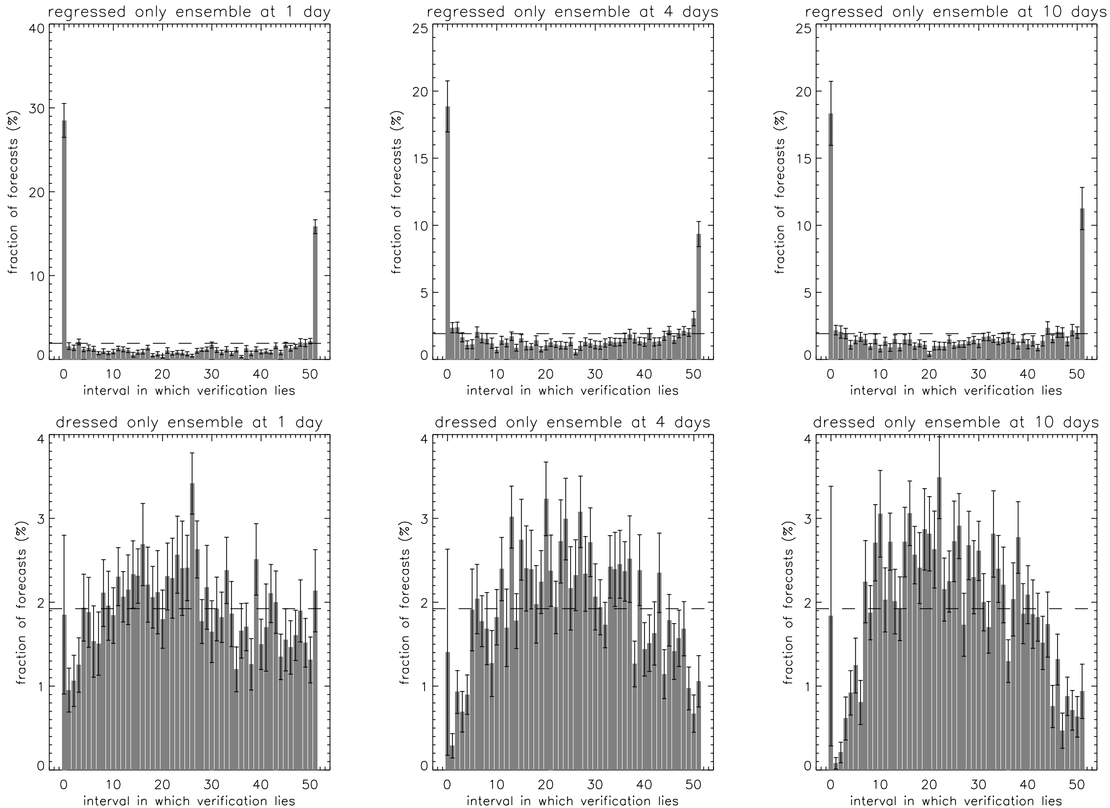

Fig. 16. A comparison of the reliabilities of ensemble forecasts which have only been regressed but not dressed (top panels) with those which have been dressed but not regressed (bottom panels). The forecasts are for significant wave height in the Santos Basin.

have enough spread to accurately reflect forecast uncertainty. In contrast, dressing the forecasts without regressing them improves the reliability the forecasts. At a lead time of 10 days the rank histogram of the dressed but unregressed ensembles appears to have too few forecasts populating the outer intervals suggesting that the dressed ensembles are too wide, although the statistical significance of this effect is small. The dressing stage of the post-processing technique has the biggest impact on the forecasts but it should be used in conjunction with the regression stage to correct for additive and multiplicative forecast bias and to prevent the post-processed ensembles from having too much spread.

4. Conclusions

A post-processing technique that involves regression and 'best member dressing' of ensemble forecasts of significant wave height has been applied to ECMWF ensemble forecasts of significant wave height at four locations of interest to the offshore oil industry. For any given day the post-processing technique used only the previous two months of forecasts. The evaluation period was from March 1999 to December 2002. The ensemble mean of the post-processed ensembles retain the same skill as the unprocessed ensembles but the post-processed ensembles provide much better predictions of the likely forecast error. This information is of considerable value to forecast users who have a nonlinear or asymmetric exposure to wave variability.

The relatively short database required for the technique to provide improvements is of practical significance, as it means that the method can be implemented in the absence of a large historical database of past forecasts and verifications. It has been found that a similarly small database of past forecasts is needed to recalibrate mesoscale weather forecasts using a Bayesian technique similar to the dressing method presented in this paper (Raftery et al., 2003).

Acknowledgements

The authors wish to thank the European Centre for Medium Range Weather Forecasting for providing their forecasts and Royal Dutch Shell for funding the research. They would also like to thank Chris Graham and George Forristall for discussions and advice. This research was funded in part by NOAA grant 1-RAT-S592-04001 and ONR DRI N00014-99-1-0056.

References

Bidlot, J.-R., 2001. ECMWF wave-model products. ECMWF Newsletter 91, 9–15.

Bidlot, J.-R., Holt, M.W., 1999. Numerical wave modelling at operational weather centres. Coastal Eng. 37, 409–429.

Buizza, R., Miller, M., Palmer, T.N., 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Quart. J. Roy. Met. Soc. 125, 2887–2908.

Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. 1, 54–77.
Evans, R.E., Harrison, M.S.J., Graham, R.J., Mylne, K.R., 2000. Joint medium-range ensembles from the Met. Office and ECMWF systems. Mon. Wea. Rev. 128, 3104–3127.
Hamill, T.M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. Mon. Wea. Rev. 129, 550–560.
Houtekamer, P.L., 1993. Global and local skill forecasts. Mon. Wea. Rev. 121, 1834–1846.
Leith, C.E., 1974. Theoretical skill of Monte-Carlo forecasts. Mon. Wea. Rev. 102, 409–418.
Molteni, F., Buizza, R., Palmer, T.N., Petroliagis, T., 1996. The ECMWF ensemble prediction system: methodology and validation. Quart. J. Roy. Met. Soc. 122, 73–119.
Murphy, A.H., 1997. Forecast verification. In: Katz, R.W., Murphy, A.H. (Eds.), Economic Value of Weather and Climate Forecasts. Cambridge University Press, Cambridge, pp. 19–74.
Palmer, T.N., 2000. Predicting uncertainty in forecasts of weather and climate. Rep. Prog. Phys. 63, 71–116.
Palmer, T.N., Buizza, R., Molteni, F., Chen, Y.Q., Corti, S., 1994. Singular vectors and the predictability of weather and climate. Phil. Trans. Roy. Soc. 348, 459–475.
Raftery, A.E., Balabdaoui, F., Gneiting, T., Polakowski, M., 2003. Using Bayesian model averaging to calibrate forecast ensembles. Tech. Report no. 440, Dept. of Statistics, Univ. of Washington.
Roulston, M.S., Smith, L.A., 2003. Combining dynamical and statistical ensembles. Tellus A 55, 16–30.
Saetra, O., Bidlot, J.-R., 2002. Assessment of the ECMWF ensemble prediction system for waves and marine winds. ECMWF Tech. Memo. 2002;, 388.
Smith, L.A., 2003. Predictability Past Predictability Present, ECMWF Seminar on Predictability 2003 pp. 291–292.
Smith, L.A., Roulston, M.S., von Hardenberg, J., 2001. End to end ensemble forecasting: towards evaluating the economic value of the ensemble prediction system, ECMWF Technical Memo. 336