Development of a wave ensemble system at the Met Office

François-Xavier Bocquet *

Met Office, Exeter, United Kingdom

1. Introduction

Increasingly model users are moving towards a risk-based approach when making operational decisions. This creates the need for a wave ensemble prediction system in order to quantify the uncertainty in the model output, by means of the ensemble spread. For example, heavy-lift operations in the offshore industry typically need a long window of calm seas to be completed safely and the confidence in the length of the window needs to be high before a decision to proceed with the operation is made.

Wave ensemble systems are currently run operationally at several operational centres, such as the European Centre for Medium Range Weather Forecasting (Saetra and Bidlot 2004) and NOAA/NCEP (Cao et al. 2007). The replacement of the Met Office NEC SX-8 supercomputer by a new IBM cluster has now made the operational implementation of a short-range wave ensemble at the Met Office possible and this paper presents the proposed system design. Results from the preliminary verification of the system are also detailed, based on a three-month period spanning from February to April 2009.

This paper is structured as follows. Section 2 describes the models used in this trial, Section 3 outlines the development of the system used for the trial, Section 4 presents the preliminary verification statistics for the trial period. Section 5 shows some potential applications derived from the wave ensemble results. Finally, conclusions and further work are discussed in Section 6.

2. Description of the models

This section briefly describes the operational atmospheric ensemble system used to provide the forcing of the wave ensemble as well as the operational wave model run at the Met Office.

a. *Atmospheric forcing*

The driving force behind the wave ensemble is the atmospheric forcing and this was provided by the Met Office Global and Regional Ensemble Prediction System (MOGREPS), which became operational in August 2005 . MOGREPS uses an Ensemble Transform Kalman (ETKF, Wang and Bishop (2003)) filter to generate a set of 23 perturbed atmospheric conditions as described in (Bowler et al. 2008) plus an unperturbed control member. The perturbed analysis fields are obtained from a linear combination of the forecast perturbations from the previous cycle. This ensures a good approximation of the analysis error covariance matrix. This method of constructing perturbations works well for short-range prediction and ensures that the physically most significant error modes in the system are selected.

* *Corresponding author address:* François-Xavier Bocquet, Met Office, Fitzroy Road, Exeter, EX1 3PB, UK.
E-mail: francois.bocquet@metoffice.gov.uk

b. *Wave model*

The wave model used in this ensemble trial is the 3.1.4 version of WAVEWATCH III (Tolman 1997), customised for operational running at the Met Office. The model uses the Tolman-Chalikov source term scheme(Tolman and Chalikov 1996), the Discrete Interaction Approximation for evaluating the non-linear interactions and a second order advection scheme (Li 2008) for increased runtime performance. The spectral resolution is 25 frequencies by 24 directions.

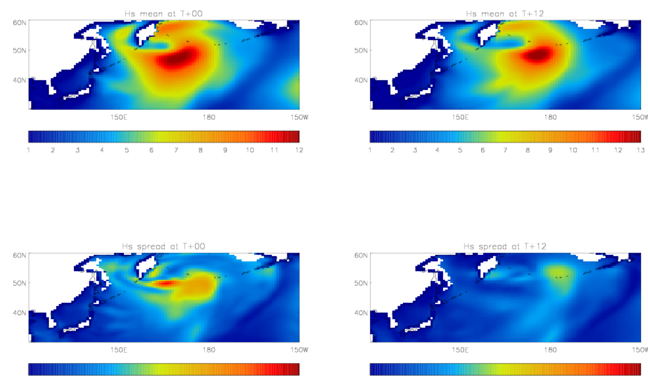
3. Development of the wave ensemble system

The aim of a wave ensemble system is to sample from a range of possible forecasts in a way which is consistent with the error structure of the observations and of the model itself. This can be achieved by an appropriate choice of atmospheric forcing, Initial Conditions (IC), Boundary Conditions (BC) and choice of a scheme which perturbs the physics of the model itself. The atmospheric forcing has already been discussed in the previous section and this section mainly focuses on IC's and BC's.

a. *Initial conditions*

There is evidence in literature (Chen et al. 2004) that the IC's do not significantly affect the performance of a wave ensemble prediction system. This is due to the fact that the differential equations used in the wave model are weakly non-linear and highly dissipative, so that the effect of the initial wave field on the wave forecasts dies away monotonically in a few days. All memory of the IC's of the wind-sea is then lost but a memory of the swell remains. This would tend to indicate that starting from a single field for all the ensemble members is suitable in cases where swell is absent.

This assumption was tested by running the wave in two steps. The first step consisted of a spin-up phase, with initial conditions provided by the operational wave model. During this phase the wave model was run for a five-day period, using successive 12 hours forecasts driven by a set of 24 forcing winds provided by MOGREPS to build up spread in the system. The final seastates of this run were then used as IC's for the second part of the test. This second part consisted of running the model for another 5 days from these IC's, but now under a single wind forcing. Fig. 1 shows how fast the ensemble member's wave fields would converge to a common seastate, or a close approximation thereof. From the start of the single wind forcing,



(a) T+0 (b) T+12

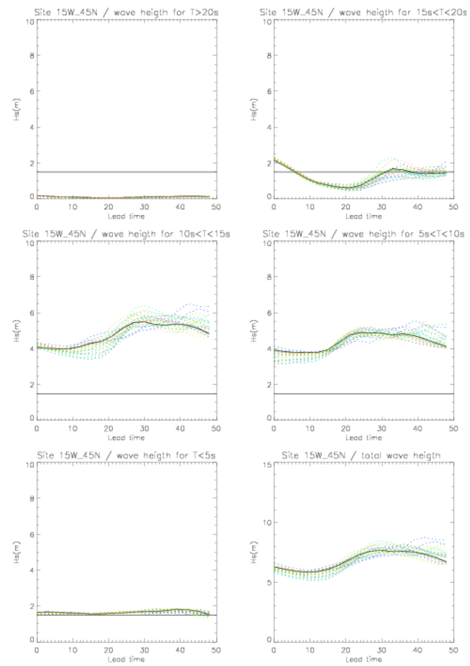
FIG. 1. Hs mean (top) and spread (bottom) for a 24 member ensemble forced with identical winds.

the spread reduces rapidly and about half of the spread vanishes within one day. While this implies that the choice of initial conditions is not crucial for medium-range forecasts, it still shows that for short lead times it will have a significant influence on the performance of the system. Therefore a decision was made to use the T+24 forecast from each ensemble member of the previous cycle as the initial condition for the equivalent ensemble member in the current cycle. Starting from an identical IC would caused the system to exhibit a lack of spread in the early stages of the forecast and was deemed unsuitable.

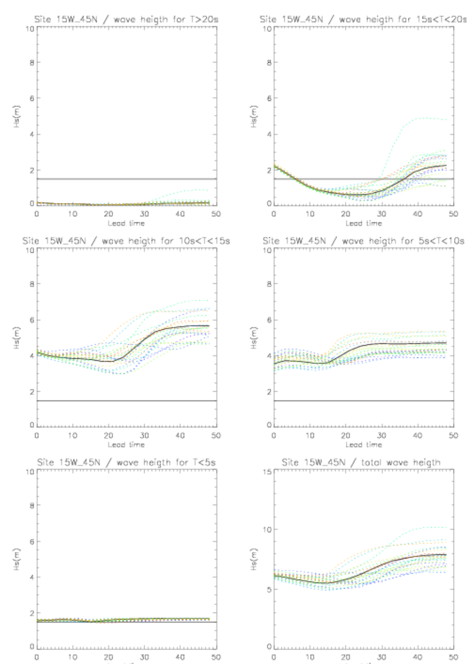
Perturbing the initial state of the system through other methods such as ETKF or singular vectors was not considered in the present study as this would require the development of a wave data-assimilation system which was outside the scope of this project.

b. *Boundary conditions*

Two options were available for the BC's in the regional model: using a set of identical BC's issued from the operational high-resolution $60km$ global model, or the more computationally expensive option of using individual BC's for each ensemble member obtained from a global ensemble run at a $90km$ resolution. A two week test run was performed to investigate the impact of both choices of BC's. The average difference in significant wave height between the two runs was found to be at most $0.2m$ at T+12. But in some locations, the spectral shape varied substantially with differences in the low frequency end of the spectrum. The most convenient way to investigate the effect of the BC is to look at the five-bin outputs of the model, where the spectrum is divided in five frequency bins and integrated to provide parameters such as the significant wave height. This gives a more detailed understanding than just looking at the overall significant wave height and period. The comparison of the five-bin output between the NAE with ensemble BC and the NAE with deterministic BC shows little difference when looking at the total significant wave height and total mean period. But a more detailed investigation, using five-bin outputs shows some difference in which the energy is divided up between the bins, as can be seen for example in Figs. 2. In this example, variability is lost in the case with the deterministic BC b, especially so in the lower frequency bins corresponding to swell.



(a) Deterministic boundary conditions



(b) Ensemble boundary conditions

FIG. 2. Comparison of five-bin outputs between a regional ensemble with deterministic and ensemble BC's. There is marked difference in energy between the 2 cases in the 10-15s and 15s-20s bands.

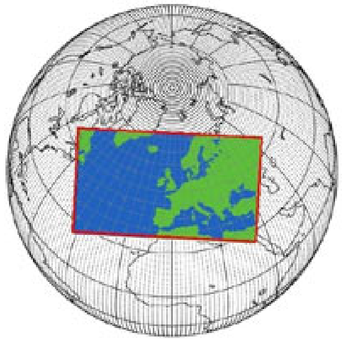


FIG. 3. View of the Regional North-Atlantic and European (NAE) domain.

The loss of spread in the lower frequencies is likely to be an issue with end users interested in this part of the spectrum. Therefore, considering the results of this test, the choice was made to use a global ensemble to introduce uncertainty in the swell entering the boundaries of the regional model.

c. Trial set-up

Given the previous considerations, the decision was made to run both a global $90km$ and regional $24km$ ensemble covering the North-Atlantic and European domain (Fig. 3). These resolutions are lower than the operational resolutions of $60km$ and $12km$ in order to reduce computational costs. Each ensemble member gets its initial conditions from the previous cycles' member and the lateral BC's of the regional ensemble are obtained from the global ensemble. This system is optimised for short-range predictions, up to 54 hours after the issue date for the regional model and up to 72 hours after issue date for the global model. The system was spun-up during a 10 day period using a single deterministic run that provided the initial conditions for each ensemble member. It was then run from February 2009 until April 2009.

4. Verification results

Although both a global and a regional ensemble have been run, this section will focus on the results from the North-Atlantic and European domain. The observational data used in this work was provided by a network of 44 buoys located in the region of interest as seen in Fig. 4.

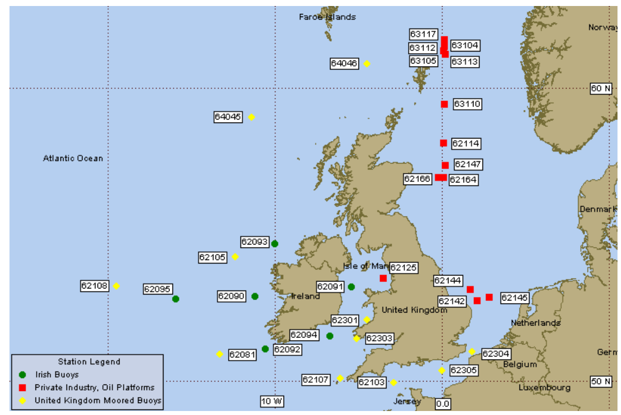


FIG. 4. Location of the wave buoys used in the verification scheme (from NDBC website http://www.ndbc.noaa.gov).

a. Spread and error statistics

A first test of the model was to compare performance of the ensemble mean with that of the unperturbed control member to provide a deterministic forecast of wave height. The ensemble mean is expected to outperform the unperturbed control member as the averaging process filters out the less predictable elements in the forecast. To verify this hypothesis, a series of plots showing the average RMSE as a function of lead time for the control member, the ensemble mean and the $12km$ high resolution run have been produced in Figure 5. There is a tendency for the ensemble mean RMSE to be slightly lower than the RMSE of the unperturbed control member. In most cases the wave ensemble spread, defined as the standard deviation of the wave height over the ensemble spread, is very low and therefore difference between the control and ensemble mean plots are expected to be small. Only in some circumstances does the spread become significant,such as near active generation areas. But these are not necessarily well sampled by the observations network, explaining the close proximity of both.

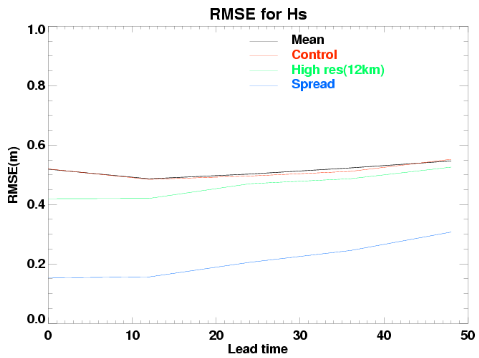


FIG. 5. Evolution of the RMS error for the control, ensemble mean and high-resolution forecasts as a function of lead time. The evolution of the ensemble spread is also shown.

In conclusion, the deterministic performance of the ensemble mean is not better than that of the control member although this could be due to some sampling problems. The ensemble system still has some value as it provides information about the forecast uncertainty through means of the ensemble spread.

b. Probabilistic skill of the ensemble

There are two main questions we want to answer when assessing the probabilistic skill of an ensemble. First, whether the ensemble spread accurately represents the forecast error and secondly whether the probabilities of an event occurring can reliably be derived from the ensemble.

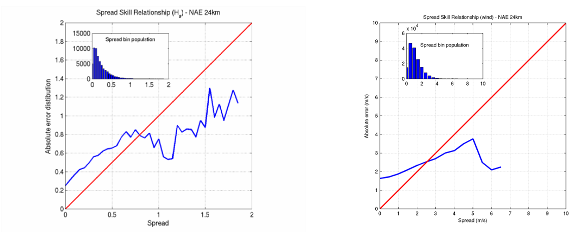
1) SPREAD-SKILL RELATIONSHIP

The spread or uncertainty in the ensemble forecast should reflect the skill of the ensemble mean. Assuming a model which perfectly represents the waves physics and assuming no observational errors, the spread should equal the error in the ensemble mean. An important caveat comes from the fact that the error RMSE is built from two parts: the observational error and the forecast error. The spread of the ensemble is only supposed to represent the forecast error. The observational error for significant wave height is typically of the order of 10% of the observation.

How close we are to this idealised situation can be determined by examining spread vs. skill diagrams, as described in Saetra and Bidlot (2004). The spread-skill diagrams in Figure 6 show the 80^{th} percentile of the absolute error of the ensemble mean against the ensemble spread. The first diagram covers the significant wave height H_s and the second the wind speed v_w. The ensemble spread is defined as the standard deviation of the variable over the sample of 24 members. The diagram was obtained by subdividing the spread into bins and calculating the absolute error percentiles per spread bin. A histogram of the population of the spread bins ha been added to clarify how representative the points on the spread-skill curve are.

Fig. 6 shows clearly that for low spread values ($< 0.8m$ for H_s and $< 3ms^{-1}$ for v_w), when the forecast uncertainty is expected to be low, it is indeed the case although the spread does not capture the full magnitude of the forecast error. For higher values of the spread, the system tends to overpredict the forecast error. This type of diagram should allow forecasters to judge quantitatively how much forecast error to expect. For low spreads, the deterministic forecast should be adequate, but for high spread situations, one would typically expect a large forecast error and examination of the individual ensemble members to judge the situation will be required.

Finally, a look at the histograms reveals that the population in the high spread bins is quite low. This indicates that an extension of this trial will be required to accurately sample from the less predictable events, which are typically associated with high seastates.

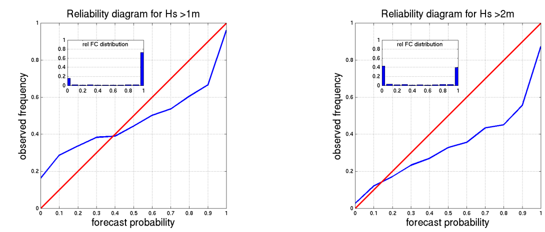


(a) H_s (b) v_w

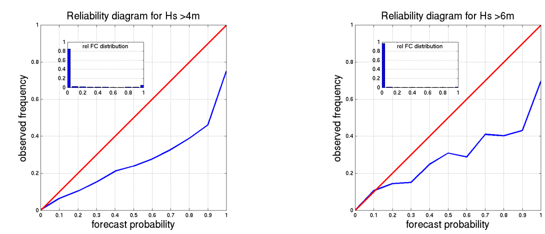
Fig. 6. Spread-skill diagram for significant wave height H_s and wind-speed v_w. The forecast spread is subdivided in bins of $0.05m$ width for H_s and $0.5ms^{-1}$ width for v_w. For each spread bin, the 80th percentile of the absolute error of the mean ensemble forecast is represented. The histogram also shows the total population per spread bin.

2) Reliability diagrams

A key step in validating any ensemble prediction system is to verify the probabilities derived from it with the frequency of occurrence of these events. A set of reliability diagrams (Wilks 1995) has been produced for the trial period for a range of events (e.g. $H_s > 4m$). The forecast gives an estimate of which forecast bin this probability lies in. For each probability bin, the occurrence rate of the event is calculated. If the probabilistic forecast were perfectly reliable, the forecast probability and occurrence rate would be identical resulting in a diagonal plot. However, the system tends to underpredict to underpredict the occurrence of events at the lower end of the probabilities and to overpredict for high forecast probabilities (Figs. 7 and 8). For higher thresholds, there is a similar behaviour in the wind and wave reliability diagrams, altough the overprediction of high probability events is more marked in the wave model. Those could be in part due to a positive bias in the wave model, and in part due to model uncertainties which have not been taken into account in this trial.



(a) $H_s > 1m$ (b) $H_s > 2m$



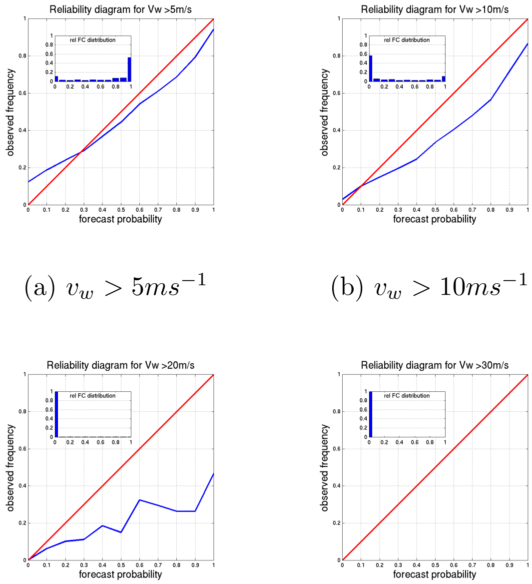
(c) $H_s > 4m$ (d) $H_s > 6m$

Fig. 7. Reliability diagrams for significant wave height exceeding $1m$, $2m$, $4m$ and $6m$. The histogram represents the probability of use of the forecast. The probabilities have been binned in bins with a width of 0.1 .

When examining the statistics for the $4m$ and $6m$ thresholds, it was found that the observed significant wave heights and wind speeds were somewhat under the typical climatological values and as a result thereof, the statistics for extreme events are less representative.

c. Other sources of uncertainty

While atmospheric forcing is the main contributor to the forecast uncertainty in a wave model, model uncertainty will also affect the forecast. The effect of adding model uncertainty will be investigated in the future through the use of a perturbed parameter scheme.



(a) $v_w > 5ms^{-1}$ (b) $v_w > 10ms^{-1}$

(c) $v_w > 20ms^{-1}$ (d) $v_w > 30ms^{-1}$

FIG. 8. Reliability diagrams for wind speeds exceeding $5ms^{-1}$, $10ms^{-1}$, $20ms^{-1}$ and $30ms^{-1}$.The histogram represents the probability of use of the forecast. The probabilities have been binned in bins with a width of 0.1 .

5. Applications

A key issue with probabilistic forecasting is conveying complex information in a way which is understandable for non-specialists. Two different categories of derived products are possible and have been tested: area based products and site-specific products.

Area based products provide probabilistic information for the whole domain. The mean and spread plots (Fig. 1) allows to see the mean forecast which can be seen as a deterministic forecast and the spread which allows to identify areas with high spread and therefore high forecast uncertainty. In addition to this, postage stamp plots allow to visualise each ensemble member and forecast exceedence plots allow to see the probability associated with certain events.

Site-specific products complement the previous examples and can convey more information in a compact format. Meteograms are now widely used in ensemble forecasting and provide a useful insight into the probabilistic forecast. Here we also show operational window statistics which were developed for use in marine operations that are highly dependent on sea-state and more specifically to the energy in certain frequency bands. An operational window is defined as an interval during which one or several operational thresholds are not exceeded, e.g. $H_s(15s < Tm < 20s) < 2m$ The wave spectrum is subdivided in 5 bands: 0-5s,5-10s,10-15s,15-20s and > 20s and is integrated to provide an equivalent wave height for those subdivisions of the spectrum. For each of these bands, it is possible to define an energy threshold which should not be exceeded during the operation. It is then possible to derive statistics for the duration of this operation window, as shown in Fig. 9. This could be used in operational risk analysis.

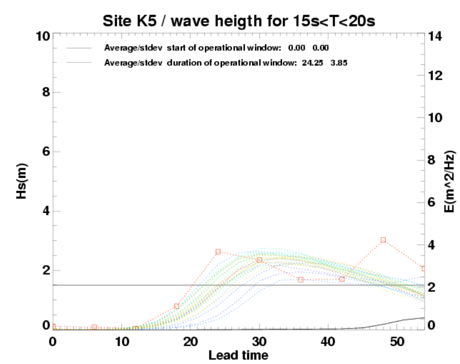


FIG. 9. Operational window for the 15-20s frequency band at the K5 buoy for the forecast issued on 03/02/2009. The orange line corresponds to the observed spectrum, retrieved from the spectral buoy. The window has an average duration of 24.25 hours with a standard deviation on the duration of 3.85 hours.

This type of product highlights the importance of selecting appropriate boundary conditions as it was found the use of a single boundary condition for all the regional ensemble members reduced the spread in the low frequency part of the spectrum and affected the quality of these forecast. A rigorous validation of this method will be presented in a later paper.

6. Conclusions and future work

The present paper identified a suitable set-up for a regional wave ensemble covering the North-Atlantic and European waters. The ideal configuration consisted of a global and regional ensemble, where the main function of the global ensemble was to provide suitable boundary conditions to the regional model so that the uncertainty in the swell was well represented. The atmospheric forcing was provided by the operational runs from the MOGREPS system. The initial conditions for each ensemble member were provided by the previous cycle's forecast for that member, in order to avoid an underspread at low lead times. Perturbation of the model physics was not considered at this stage.

This system was run for a three month period and validated against in-situ observations. The validation results showed a good reliability of the forecast probabilities, although the statistics for extreme events were not reliable due to the small sample size of these events during the period considered and a longer trial period will be required to make the statistics more significant. The performance of the ensemble mean showed no significant improvement on that of the unperturbed control member and the proposed explanation was a lack of observations in the high spread areas. Most observations were in low spread areas where the ensemble mean and ensemble control member are expected to be close.

The present work only investigated the effects of the forecast error, due to uncertainty in the winds. In practice model errors also contributed to the uncertainty and to include this effect it will be necessary to investigate the use of perturbed physics where a range of physical parameters in the model are sampled from their expected distribution.

In the future, the verification will be improved in two ways. Satellite observations will be used in the verification as well to provide a better insight into model performance away from coastal waters where most buoys are situated. The verification work will also be extended to include tools such as economic value analysis, in order to judge the economic value of using the wave ensemble in risk-based applications.

7. Acknowledgements

Several people within the Met Office have provided valuable input for this work and I would like to extend my thanks to Chris Bunney, Andrew Saulter, Jonathan Flowerdew and Ken Mylne for the many useful discussions.

REFERENCES

Bowler, N., A. Arribas, K. Mylne, K. Robertson, and S. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **134**, 703–722.

Cao, D., H. Chen, and H. Tolman, 2007: Verification of ocean wave ensemble forecast at NCEP, NOAA/NWS/NCEP/EMC/MMAB technical note nr. 261. Tech. Rep. 261, NCEP, NOAA/NWS/NCEP/EMC/MMAB.

Chen, H. S., B. D., L. D. Burroughs, and H. L. Tolman, 2004: A variation wave height data assimilation system for NCEP operational wave models. Tech. Rep. MMAB/2004-04, NCEP, 1-16 pp.

Li, J. G., 2008: Upstream nonoscillatory advection schemes. *Mon. Weather Rev.*, **136**, 4709–4729.

Saetra, O. and J.-R. Bidlot, 2004: Skill and relative economic value of the ECMWF ensemble prediction system. *WF*, **19**, 673–689.

Tolman, H. and D. Chalikov, 1996: Source terms in a 3rd generation wind-wave model. *J. Phys. Oceanogr.*, **26**, 2497–2518.

Tolman, H. L., 1997: User manual and system documentation of WAVEWATCH III. Tech. Rep. 151, NCEP.

Wang, X. and C. H. Bishop, 2003: A comparison of breeding and Ensemble Transform Kalman Filter ensemble forecast systems. *JAS*, **60**, 1140–1158.

Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences*. Academic Press.