Tellus (2005), 57A, 204–216
Printed in UK. All rights reserved

TELLUS

Approximation of ensemble members in ocean wave prediction

By LEANDRO FARINA[1]*, ANTÔNIO M. MENDONÇA[2] and JOSÉ P. BONATTI[2], [1]*Instituto de Matemática, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 91509-900, Brazil;* [2]*Centro de Previsão de Tempo e Estudos Climáticos, Instituto Nacional de Pesquisas Espaciais, Cachoeira Paulista, SP 12630-000, Brazil*

(Manuscript received 6 November 2003; in final form 2 November 2004)

ABSTRACT

An efficient method for generating members in a ocean wave ensemble prediction system is proposed. A linearization of the wave model WAM is used to obtain approximations of the ensemble members. This procedure was originally introduced in a dynamical assimilation scheme where Green's functions play a central role. The evaluation of member approximations can be carried out in a fraction of the time required by the full model integration. This aspect of the method suggests a way of increasing the ensemble size as well as refining the model resolution without increasing computational costs.

1. Introduction

Ensemble prediction is a technique in which several forecasts are produced based on an ensemble of different possible objects, such as initial conditions, forcings and/or model parameters. The advantages of an ensemble prediction system (EPS) are well known. Amongst its benefits are greater reliability for the solution, the generation of several possible predictions and the probabilities associated with them as well as the capability of predicting better extreme events. In December 1992, operational EPSs were put into activity by the US National Meteorological Center (NCEP) and by the European Centre of Medium-Range Weather Forecasts (ECMWF). These systems are being continuously upgraded and are the subject of current research worldwide.

The ensemble method employed in atmospheric modelling has an analogue for ocean wave prediction. From the present point of view, a wave ensemble prediction system (WEPS) can essentially go in two directions to produce ensemble solutions or members: (a) to create perturbations of the forcing wind fields or/and (b) to generate perturbations of the initial wave spectrum. These approaches are described and analysed by Farina (2002) and some potential benefits of wave ensemble prediction are presented in Janssen (2000) and Hoffschildt et al. (1999) where the ECMWF wave ensemble forecasts, operational since June 1998, are employed.

* Corresponding author.
e-mail: farina@mat.ufrgs.br

Ensemble prediction generates a very large amount of data. This gives rise to two problems: how to interpret these new data, giving meaningful products, and the high computational demands for generating an ensemble of forecasts. This work will deal with the latter problem, in the context of ocean wave ensemble prediction. Given a computational setting, the cost of a model integration in ensemble mode is dictated by factors such as the number of members in an ensemble and the resolutions of the model runs. Usually a compromise must be made, reducing what is at first sight the ideal number of members and the resolutions in order to accommodate an operational EPS in centres where the usage computational resources is optimized. Ideally, one would like to increase the number of members, as this procedure would generally improve probability forecasts as well as capture possible extreme events, which are unpredictable otherwise. The resolutions used are also a very important issue. An EPS normally employs a coarse resolution for members, keeping only the control run fine. However, high-resolution ensemble prediction systems (HEPS) are appearing; for example Buizza et al. (2003) recently showed qualitative performance improvement and benefits of a HEPS.

Evidently, a mechanism able to reduce the cost in the generation of the members favours the increase in the number of members as well as the resolution refinement. In this paper we essentially compare two wave EPSs; the usual one where the members are obtained by full non-linear integration of the wave model and another where the members are calculated by the approximated linearized model. These wave EPSs are based on forcing provided by an atmospheric EPS. A fast method for generating approximated members in a ocean wave ensemble is

proposed. This method uses a linearization of the third-generation wave model WAM, introduced by Bauer et al. (1996), where Green's functions play a central role and a means of evaluating member approximations in a fraction of the time required by the full model integration. We noticed that the evolution of these member approximations gives information on severe sea states often not predicted by the usual ensemble members. Similar linearization approaches are described for ocean general circulation models in Stammer and Wunsch (1996) and Menemenlis and Wunsch (1997).

The outline of the paper is as follows. Section 2 briefly presents the wave model and in Section 3, the wave ensemble prediction system is described with the exposition of the fast method and the linearization used in its construction. An extension to this method is also discussed in this section, where a sub-ensemble is defined by using a local dimension concept. Section 4 introduces the perturbation method used to get the atmospheric ensemble forcings. This method is employed by the operational EPS of the Centro de Previsão de Tempo e Estudos Climáticos (CPTEC). Section 5 introduces and analyses the numerical results. In Section 6, we make some concluding remarks and perspectives are discussed.

2. Wave model

To model the waves, we adopt the wave model WAM (Komen et al., 1994), which is a tested and powerful tool for operational wave prediction. The essential structure of this model and the governing equation is presented below. We assume that in deep water the evolution of the ocean wave spectrum can be described by the wave action balance equation:

$$\frac{\mathrm{D}}{\mathrm{D}t}F(\boldsymbol{k},\boldsymbol{x},t) = S_{\mathrm{nl}}(\boldsymbol{k},\boldsymbol{x},t) + S_{\mathrm{in}}(\boldsymbol{k},\boldsymbol{x},t;\boldsymbol{U}) + S_{\mathrm{ds}}(\boldsymbol{k},\boldsymbol{x},t), \quad (1)$$

where $\boldsymbol{k}$ is the wavevector, $\boldsymbol{x}$ is a point on the mean free surface, t is time and $F(\boldsymbol{k}, \boldsymbol{x}, t)$ is the wave energy density, with the wave action defined by $F(\boldsymbol{k}, \boldsymbol{x}, t)/\omega(\boldsymbol{k})$, where ω is the intrinsic angular frequency. The term S_{nl} is the rate of variation of wave energy at wavevector $\boldsymbol{k}$ and at the position $\boldsymbol{x}$ due to the non-linear wave–wave interactions. This process of interaction is weakly non-linear in the sense that resonance is contributing only for groups of four waves in deep water, a case where the dispersion relation is given by $|\boldsymbol{k}| = \omega^2/g$. A representation of this process is given in terms of the Boltzmann integral:

$$\begin{aligned}\frac{\partial}{\partial t}N = 4\pi \int & T_{0123}[N_1N_2(N_3+N) - N_3N(N_1+N_2)] \\ & \times \delta(\mathbf{k}_1+\boldsymbol{k}_2-\boldsymbol{k}_3-\boldsymbol{k})\delta(\omega_1+\omega_2-\omega_3-\omega) \\ & \times \mathrm{d}\boldsymbol{k}_1\,\mathrm{d}\boldsymbol{k}_2\,\mathrm{d}\boldsymbol{k}_3, \end{aligned} \quad (2)$$

where $N_i = N(\boldsymbol{k}_i, t)$, $N = N(\boldsymbol{k}, t)$ is the wave action spectrum and T_{0123} denotes the kernel of this integral equation.

The expression for T_{0123} may be seen in (Hasselmann, 1962). Equation (2) incorporates conservation of action, moment and energy. The resonance condition $\boldsymbol{k}_1 + \boldsymbol{k}_2 = \boldsymbol{k}_3 + \boldsymbol{k}_4; \omega_1 + \omega_2 = \omega_1 + \omega_2$ present in the delta distributions selects those wave groups that contribute to S_{nl}. In the model, eq. (2) is not actually used; an approximation of this, called the DIA (discrete interaction approximation) is instead employed and details of this can be found in Komen et al. (1994).

The term S_{in} gives the energy rate transferred from the wind $\boldsymbol{U}(\boldsymbol{x}, t)$ to the wave surface and S_{ds} represents the wave energy dissipation rate.

In order to solve eq. (1), the knowledge of the spectrum F at a time t_0 and the wind field forcing $\boldsymbol{U}(\boldsymbol{x}, t)$ must be prescribed. One of the most widely used parameters obtained from the solution of the problem modelled above is the significant wave height H_{s}, defined as the average height of the 1/3 highest waves. It can be shown that (Ochi, 1998)

$$H_{\mathrm{s}} = 4\sqrt{E},$$

where E is the total wave energy at position $\boldsymbol{x}$ and at time t, given by

$$E = \int_0^\infty F(\boldsymbol{k},\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{k}.$$

3. Methodology

In the present work, we will consider an ocean wave EPS on which only the forcing of the system, i.e. the surface wind fields $\boldsymbol{U}_j(\boldsymbol{x}, t)$, $1 \le j \le N$ is perturbed. The respective model integrations produces N solutions, or members. The initial wind fields can be generated by the breeding method (Toth and Kalnay, 1997), perturbations based on empirical orthogonal functions (EOF) (Zhang and Krishnamurti, 1999) or by singular vectors (Molteni et al., 1996). In Section 4 we describe how we got the wind field perturbations based on EOFs. These perturbations were used for the numerical experiments in Section 5.

The number of members in an ensemble prediction system can be large and its computation extremely costly. This fact is seen by recalling that each member, or solution, requires the integration of the balance eq. (1). Typically, we will be interested in using the maximum possible number of members and in practice the effective computation of the members is only limited by available computational resources. Thus, if N is the number of members, the cost in an EPS, adding the control, is proportional to

$$(N+1)s, \quad (3)$$

where s is the computational cost for calculating the solution of eq. (1).

The idea of the method to be described in this section is to obtain approximations of each member and with low computational cost. Before showing how these approximations are obtained, let us show how the balance equation can be linearized, which is an important issue in its own right.

3.1. Linearization

We adopt the linearization procedure proposed by Bauer et al. (1996) and introduce it in the context of assimilation of wave data into the WAM model. This scheme uses impulse response or Green's functions and its steps will be followed now.

Equation (1) can be written as

$$S(F, \boldsymbol{U}) = \frac{\mathrm{D}F}{\mathrm{D}t},$$

where $S = S_{\mathrm{nl}} + S_{\mathrm{in}} + S_{\mathrm{ds}}$ denotes the total source term. Thus, a perturbation $\boldsymbol{u}$ of the wind field and the corresponding wave spectrum e are related through the Taylor expansion

$$S(e, \boldsymbol{u}) = S(0, \boldsymbol{0}) + \boldsymbol{u} \cdot \frac{\partial S}{\partial \boldsymbol{U}} + e\frac{\delta S}{\delta F} + \frac{1}{2}\left(|\boldsymbol{u}|\boldsymbol{u} \cdot \frac{\partial^2 S}{\partial \boldsymbol{U}^2} + e^2\frac{\delta^2 S}{\delta F^2}\right) + \cdots,$$

where $\delta S/\delta F$ denotes the functional derivative of S with respect to F. Let us now invoke the assumption that the differences between the spectra originating from two perturbed wind fields in the ensemble obey a linear dynamics. This hypothesis is partially corroborated by the numerical experiments in Farina (2002) and was employed in the same of related physical situations by Bauer et al. (1996), Stammer and Wunsch (1996) and Menemenlis and Wunsch (1997), for instance. Thus, neglecting the $\mathcal{O}(|\boldsymbol{u}|^2)$, $\mathcal{O}(e^2)$ and $\mathcal{O}(|\boldsymbol{u}|e)$ terms, we have

$$S(e, \boldsymbol{u}) = \boldsymbol{u} \cdot \frac{\partial S}{\partial \boldsymbol{U}} + e\frac{\delta S}{\delta F}, \tag{4}$$

or

$$Le = \frac{\partial S}{\partial \boldsymbol{U}} \cdot \boldsymbol{u},$$

where $L = (\mathrm{D}/\mathrm{D}t) - \Lambda$ with $\Lambda = \delta S/\delta F$. Formally, we can write

$$e = L^{-1}\left(\frac{\partial S}{\partial \boldsymbol{U}} \cdot \boldsymbol{u}\right).$$

Using Green's functions G, we can express e explicitly by

$$e(\boldsymbol{k}, \boldsymbol{x}, t) = \iiint G(\boldsymbol{k}, \boldsymbol{x}, t; \boldsymbol{\xi}, \tau)\frac{\partial S}{\partial \boldsymbol{U}} \cdot \boldsymbol{u}(\boldsymbol{x}, t; \boldsymbol{\xi}, \tau)\, \mathrm{d}\boldsymbol{\xi}\, \mathrm{d}\tau. \tag{5}$$

Thus, the error $r_2(e, \boldsymbol{u})$ in approximating $S(e, \boldsymbol{u})$ with expression (4) is for sufficiently smooth S, such that

$$\lim_{\substack{e\to 0 \\ \boldsymbol{u}\to 0}} \sum_{\substack{\alpha_1,\alpha_2\le 2 \\ \alpha_1+\alpha_2=2}} \frac{r_2(e, \boldsymbol{u})}{|\boldsymbol{u}|^{\alpha_1} e^{\alpha_2}} = 0$$

and the error in e is therefore of the order $L^{-1}(r_2(e, \boldsymbol{u}))$.

In practice, however, eq. (5) must be simplified if we wish to tackle the problem computationally. With this goal, we use the behaviour observed in the dynamics of wind seas, namely that for a small perturbation $e(\boldsymbol{k}, \boldsymbol{x}, t)$, there exists a highly localized domain in space and time on which a perturbation is more influential. We then assume that this characteristic can be modelled by delta distributions acting on *influence points* $\boldsymbol{\xi}_0(\boldsymbol{k}, \boldsymbol{x}, t)$ and $\tau_0(\boldsymbol{k}, \boldsymbol{x}, t)$ that represent these domains. Thus, we write

$$G(\boldsymbol{k}, \boldsymbol{x}, t; \boldsymbol{\xi}, \tau)\frac{\partial S}{\partial \boldsymbol{U}} = \delta(\boldsymbol{\xi} - \boldsymbol{\xi}_0)\delta(\tau - \tau_0)\boldsymbol{W}(\boldsymbol{k}, \boldsymbol{x}, t). \tag{6}$$

From eqs (5) and (6), we have

$$e(\boldsymbol{k}, \boldsymbol{x}, t) = \boldsymbol{W}(\boldsymbol{k}, \boldsymbol{x}, t) \cdot \boldsymbol{u}_0(\boldsymbol{x}, t), \tag{7}$$

where $\boldsymbol{u}_0 = \boldsymbol{u}(\boldsymbol{\xi}_0, \tau_0)$. The *impact function* $\boldsymbol{W} = (W_1, W_2)$ must be determined in such a way to represent the past evolution of the sea state and $\boldsymbol{\xi}_0$ and $\tau_0(\boldsymbol{k}, \boldsymbol{x}, t)$ are found using the wave age parameter, a ratio between the wave group velocity and the wind velocity. We then see that $\boldsymbol{u}_0$ is indirectly a function also of $\boldsymbol{k}$. The parameters

$$\boldsymbol{W}, \boldsymbol{\xi_0} \quad \text{and} \quad \tau_0(\boldsymbol{k}, \boldsymbol{x}, t) \tag{8}$$

can be calculated using values of the source functions already in use and required by the control integration of eq. (1). Such an integration, where the functions in eq. (8) are produced, we refer to as an *enhanced integration* of the model. See (Bauer et al., 1996) for further details of this procedure, although there this terminology is not used. As no additional source function is necessary, the computation of $e(\boldsymbol{k}, \boldsymbol{x}, t)$ can be carried out during the control and non-linear integration of the model. As the main cost in the WAM model is due to the evaluation of its source functions, the added computational cost is then comparatively very small. However, the memory storage requirement of this enhanced integration increases.

3.2. Fast method

Let us now describe the fast wave ensemble method. Suppose the pair of data $(F_0(\boldsymbol{k}, \boldsymbol{x}, t_0), \boldsymbol{U}_0(\boldsymbol{x}, t))$ are prescribed. From these we get the control solution, the spectrum $F_0(\boldsymbol{k}, \boldsymbol{x}, t)$, for all t in the time interval considered. N other solutions, or members of the ensemble $\mathcal{E}_N = \{F_j(\boldsymbol{k}, \boldsymbol{x}, t); 1 \le j \le N\}$ are obtained from the data of the ensemble $\mathcal{D}_N = \{(F_j(\boldsymbol{k}, \boldsymbol{x}, t_0), \boldsymbol{U}_j(\boldsymbol{x}, t)); 1 \le j \le N\}$. We now choose a sub-ensemble, $\mathcal{B}_M$ formed by the M members of $\mathcal{D}_N$ that have greatest degree of linear independency or variance. These concepts will be made precise in the next section.

The ensemble solution is then obtained in the following way. F_0 and the members of $\mathcal{B}_M$ are calculated by full, non-linear integration of the model and the other members by

$$F_j = F_{jb} + e_j, \quad j \ne jb,$$

where $F_{jb} \in \mathcal{B}_M$ and e_j are computed using

$$e_j(\boldsymbol{k}, \boldsymbol{x}, t) = \boldsymbol{W}(\boldsymbol{k}, \boldsymbol{x}, t) \cdot (\boldsymbol{U}_j - \boldsymbol{U}_{jb})(\boldsymbol{k}, \boldsymbol{x}, t). \tag{9}$$

The spectrum F_{jb} is chosen in $\mathcal{B}_M$ by minimizing $|\boldsymbol{U}_j - \boldsymbol{U}_{jb}\|$.

The complexity of this algorithm can be assessed in the following way. The cost of an ensemble prediction system, as shown by eq. (3), is proportional to $(N + 1)s$. In the fast method, the

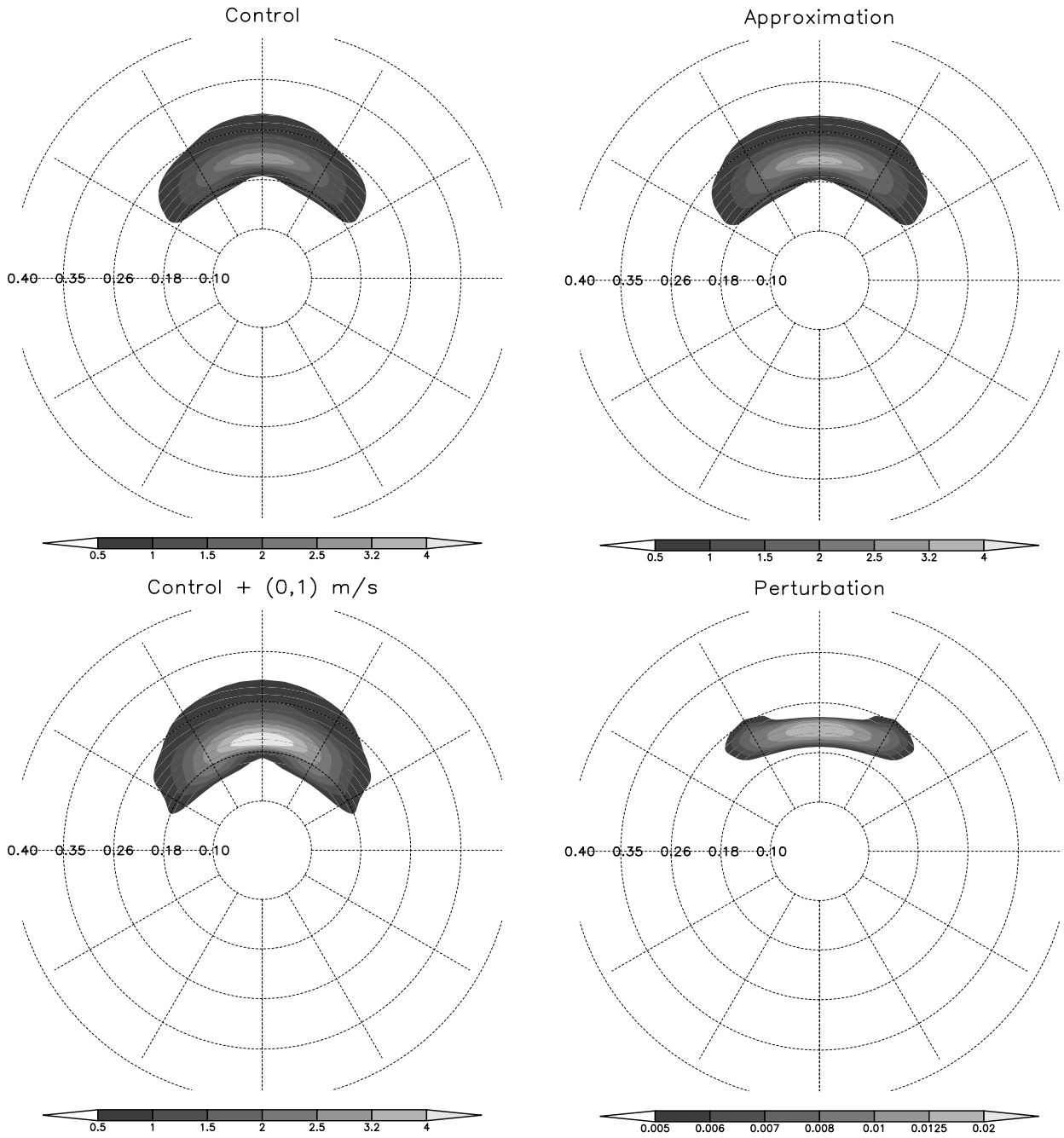

Fig 1. Spectra at a point in an open ocean 6° away from a uniform wind field blowing at a height of 10 m from south to north with a speed of 10 m s^{-1} (the control case) and of 11 m s^{-1}. The spectra at the top are obtained from the non-linear model integration. At the bottom of the figure are the approximation $F_a(f, \theta)$ of the 11 m s^{-1} spectrum and the perturbation $e(f, \theta)$, as functions of frequency and direction and for $t = 6$ days.

Fig 1. (cont'd).

number of full integrations has a cost of $\mathcal{O}(M + 1)$, so that its total cost is $\mathcal{O}(M + 1) + q$, where q/M is the overhead cost in an enhanced integration. Neglecting the overhead cost, the acceleration of the fast method will be proportional to $(N + 1)/(M + 1)$.

Computing only the control solution F_0 by the full, non-linear model, i.e. taking $M = 0$, the scheme proposed is then roughly $N + 1$ times faster than the conventional EPS. This option is adopted in the numerical experiments reported in Section 5.

3.3. *Determining* $\mathcal{B}_M$

We will consider determining the sub-ensemble $\mathcal{B}_M$. This procedure is inspired by Patil et al. (2001), where the local

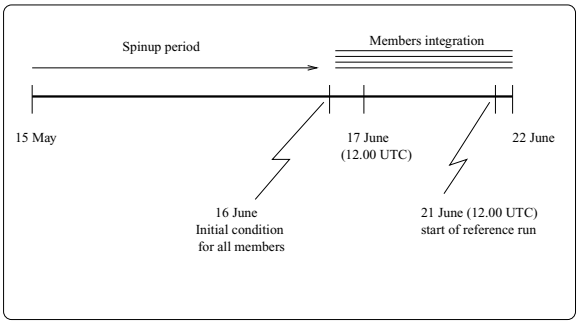

Fig 2. Scheme showing the procedure adopted for the realistic simulation of June 2000. The control wave model run is started cold on 15 May 2000 and integrated until 16 June 2000. All the ensemble members are integrated from 16 June to 22 June. The reference state is obtained from a 12 h run with the best winds available.

dimensionality of the atmosphere is studied (see also Francisco and Muruganandam, 2003).

Divide the spatial domain, that usually consists of the oceanic portion of the globe, into subdomains where local dimensions will be determined. Fixing $t = t_m$, consider k points on each of these subdomains where the fields $\boldsymbol{U}_j(\boldsymbol{x}, t_m)$, $1 \leq j \leq N$ are evaluated. Considering the two components of the vector $\boldsymbol{U}_j$, this discretization can be arranged in a matrix A, with $2k$ rows and N columns. These columns are called *local bred vectors* by Patil et al. (2001) as they can be obtained by an ensemble breeding method (Toth and Kalnay, 1997). We look into determining the dimension of the space spanned by the bred vectors. Empirical orthogonal functions are employed. The covariance matrix of A is $C_{N\times N} = A^{\mathrm{T}}A$, where A^{T} is the transpose of A. Since the covariance matrix is non-negative definite and symmetric, its N eigenvalues λ_i are non-negative and have eigenvectors v_i such that Av_i form an orthonormal basis for the space spanned by the columns of A. Thus, the eigenvalues λ_i measure how much the column vectors of A point in the direction of v_i and $\sigma_i^2 := \lambda_i$ represents the quantity of variance with respect to v_i. In order to indentify a local dimension generated by the N local bred vectors in each subdomain, we define the following statistic over the values of σ_i.

$$\psi(\sigma_1, \sigma_2, \ldots, \sigma_N) = \left(\sum_{i=1}^{N} \sigma_i\right)^2 \left(\sum_{i=1}^{N} \sigma_i^2\right)^{-1}. \tag{10}$$

Thus, ψ assumes values in $[0, N]$. These values represent a local dimension of the field $\boldsymbol{U}$. The value of ψ also supplies a means of determining the members of $\mathcal{B}_M$; in particular one can take $M = \mathrm{int}(\psi)$ where int denotes the closest integer. These members are the ones with greater total variance $\mu(\sigma_i^2)$, where μ is an average over all subdomains.

4. Atmospheric perturbations

The procedure employed to generate the atmospheric perturbed initial conditions is based on the method introduced by Zhang and Krishnamurti (1999) originally proposed for hurricane forecasting using the Florida State University global model. This method, called *EOF-based perturbation*, was developed observing the fact that during the initial integration period of atmospheric models, perturbations to a reference state grow linearly. By this hypothesis, one can construct an ensemble of *optimal perturbations* using empirical orthogonal functions. This approach is outlined by the following steps:

(1) n random small factors, with the same order of magnitude of the forecasting errors are added to the control analysis.

(2) The resulting n fields are integrated for 36 h (optimal interval) storing the solutions every 3 h.

(3) The control forecast is subtracted from each of the n solutions at each time increment of 3 h. This generates n temporal series.

(4) An EOF analysis of the temporal series is carried out on a domain of interest. This analysis allows us to find eigenvectors (modes) associated with the largest eigenvalues. These modes are the optimal perturbations.

(5) The optimal perturbations are rescaled in order to make its standard deviation of the same order as the initial perturbations.

(6) Adding and subtracting these optimal perturbations from the control analysis produces an ensemble of $2n$ initial perturbed states.

Thus, two groups of members in the ensemble can be identified: the negative and the positive. This classification will be used in Section 5.

The optimal interval of 36 h was introduced by Zhang and Krishnamurti (1999) who observed that the perturbations present an approximately linear growth up to 36 h. This supports the use of EOFs to determine the optimal perturbations.

Aiming at hurricane prediction, Zhang and Krishnamurti (1999) proposed perturbations with respect to the hurricane initial condition and computation of the empirical orthogonal functions in the neighbourhood of the hurricane. This approach is effective in studying the evolution of a localized extreme event. However, for the atmosphere general circulation, the application of perturbations at some specific event does not seem reasonable. Therefore, two main modifications are introduced to the EOF-based perturbation method. The first deals with the perturbed region that was originally confined to the neighbourhood of the hurricane. For global prediction, Coutinho (1999) noticed that restricting the perturbations to subdomains limited in latitude and longitude, as for instance a rectangular region over South America, does not produce good results. The domain isolation would affect the perturbation growth in regions relevant to synoptic system evolution. the results of Coutinho (1999) also show thatconsidering more extensive regions, such as 45°S to 30°N, 0°E to 360°E, improves the performance of the method. In the present study, we performed perturbations in the domain (65°S,10°N)

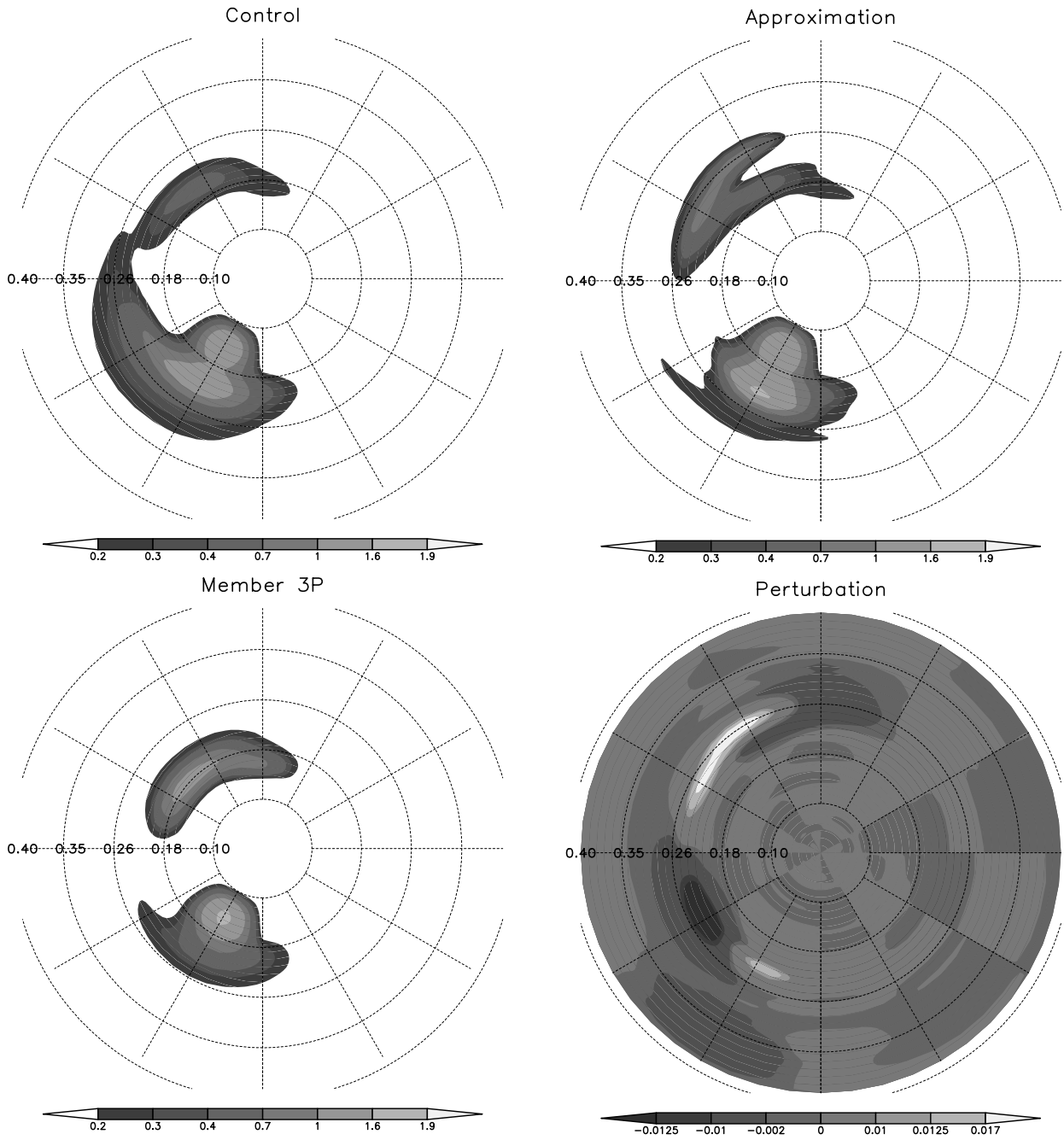

Fig 3. Spectra as functions of frequency (Hz) and direction for simulation of June 2000 described in the text. For this situation, the mean wave direction is 250° from the east and measured clockwise, the peak period is 10.5 s and $H_s = 2.24$ m.

Fig 3. (cont'd).

as we aim to study a case focusing on the rough Southern Hemisphere oceans. The second modification refers to the intensity of the initial perturbations and to the optimal perturbation rescaling. In Zhang and Krishnamurti (1999) it is suggested that the initial perturbations should be of the order of 3 h forecasting errors: 3 m s^{-1} for the wind and 0.6 K for temperature. Originally it was also recommended that the standard deviation of the optimal perturbations in relation to the total average should be 1.5 m s^{-1} for the wind and 0.7 K for the temperature. Since the perturbation region was altered, more appropriate values for the initial perturbations and for the rescaling have been tested. Thus, increased values of 5.0 m s^{-1} and 1.5 K, suggested by Daley and Mayer (1986) showed better results than the

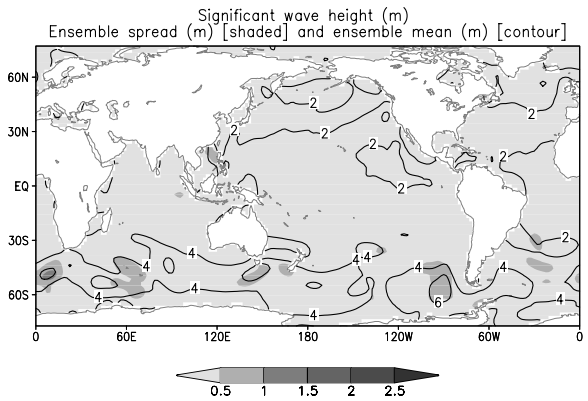

Fig 4. The ensemble spread (shaded) and ensemble mean (contour) of significant wave heights for 12:00 UTC of June 2000 predicted 36 h before by the non-linear WEPS.

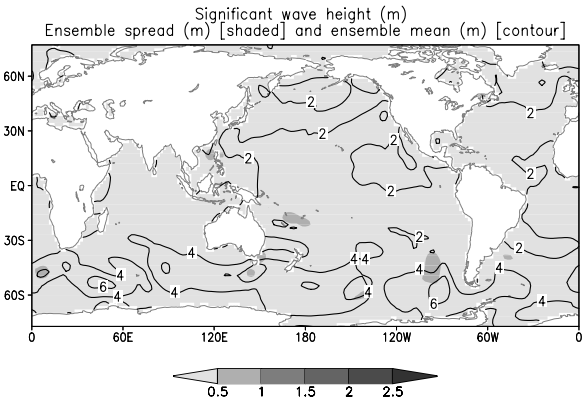

Fig 5. The ensemble spread (shaded) and ensemble mean (contour) of significant wave heights for 12:00 UTC of June 2000 predicted 36 h before by the linearized WEPS.

original ones with respect to the performance of the ensemble mean, measured by anomaly correlations.

The atmospheric model used for the generation of the WAM model forcing, the wind stress, is the CPTEC/COLA global model (Cavalcanti et al., 2002). This spectral model was executed with horizontal resolution of 1.875°, 28 levels of σ and with subgrade physical processes through parametrizations. The control initial conditions were obtained from NCEP. The wind stress used by the wave model spinup is from the above initial conditions, while the forecasted wind field ensemble was generated by the integration of the CPTEC/COLA global model in the ensemble mode for a period of 144 h. For the initial condition of 16 June 2000, adopted in an experiment in Section 5, a total of 20 perturbed fields were produced.

5. Numerical results

The WAM configuration used in the experiments has a global domain, with wind stress updated every 3 h. The spatial and wind grids have resolution of 1.875°. Further, the WAM model was implemented with additional steps to incorporate the capability to produce the impact function $\boldsymbol{W}$ and the influence points $\boldsymbol{\xi_\theta}(\boldsymbol{k}, \boldsymbol{x}, t)$ and $\tau_0(\boldsymbol{k}, \boldsymbol{x}, t)$ during its integration. This procedure allows the construction of linearized solutions such as $F_a = F_0 + re$, where F_0 is the control solution and r is a fitting factor. Note that we take $M = 0$ in the numerical experiments we are about to report.

Firstly, consider a simple and idealized situation where a uniform wind field is blowing from south to north with speed of 10 m s^{-1} at a height of 10 m. Let this situation be the control one. In order to make a comparison, consider a similar setting with the difference that the wind speed is 11 m s^{-1}. Consider now a point in an open, deep ocean, subject to the waves and 6° away from the location where the wind starts to blow. In this case, from eq. (9), the perturbation is given by the second component of the impact vector function, i.e. $e = W_2(f, \theta)$. The two situations are depicted in Fig. 1 which shows the two spectra obtained from the non-linear model integration, the approximation $F_a(f, \theta)$ and the perturbation $e(f, \theta)$, as functions of frequency and direction and for $t = 6$ days. This figure shows a good and promising agreement between the approximation, the linearized solution and the spectrum generated by non-linear integration with the 11 m s^{-1} wind speed forcing. This numerical experiment closely reproduces a similar simulation presented in Bauer et al. (1996) and the results support each other.

For the next case study which we offer, realistic data are taken. Twenty wind field perturbations are generated by the EOF-based method explained in Section 4. These perturbations are separated in two groups of 10, denoted by 1P, 2P, . . . ,10P and 1N, 2N, . . . , 10N, where P and N here stand for positive and negative, respectively.

The time of the simulation is within a Southern Hemisphere winter. This is the season when the Southern Oceans, the roughest overall (based on satellite altimeter data over 9 yr, Young (1999) presented global statistics of significant wave heights and concluded that the Southern Ocean clearly has the most extreme, year round, wave conditions), are more agitated. So, the events we analyse are located on this half of the globe. To allow spinup of the wave model and to incorporate swell in the wave field, the control wave model run is started cold on 15 May 2000 and integrated until 00:00 UTC 16 June 2000, when the initial spectrum condition is used by all members of the EPS. Parallel to this procedure, all approximating members obtained by linearization are generated by the fast method (see diagram in Fig. 2).

Figure 3 shows the control spectra, the member 3P and its linearized spectra along with the perturbation e used to get the approximation. The point where this bimodal sea occurs is $\boldsymbol{x} =$ (6.5°N,210°E) at time $t = 6$ days. These data represent a 144 h prediction for 00:00 UTC of 22 June 2000. The wind speed is 5.5 m s^{-1} blowing at an angle of 138°, from the east, measured

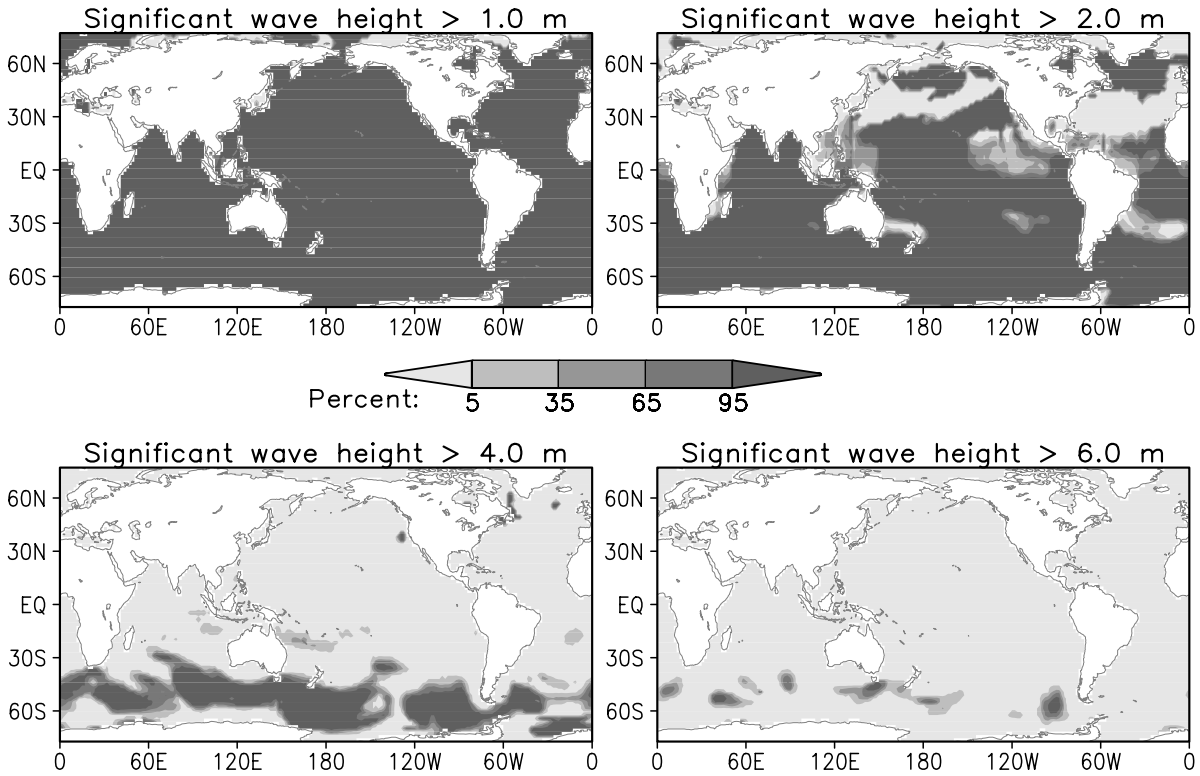

Fig 6. Probability distributions of the 36 h significant wave heights prediction for 12:00 UTC of 17 June 2000 above 1, 2, 4 and 6 m are shown for the non-linear WEPS.

clockwise. The mean wave direction is 250°, the mean period is 10.5 s and $H_s = 2.24$ m. Thus, the wave spectrum at the point studied reflects a wind sea with swell. Even after this extended period of time, we can see that linearized solution clearly reproduces some of the characteristics of the 3P member. Perhaps the main difference between the control and the 3P member is their respective connection and disconnection shown in the graphic of the two spectral modes; the one due to the wind sea blowing from the northeast and the other due to the swell from the southwest. In the perturbation spectrum there is a (f, θ) region with negative values forcing the separation of the two modes in the approximation. Similarly there is strongly positive region in the perturbation inducing an enhancement of the northernmost mode, in agreement with the original ensemble member. However, there is also a peak, at roughly 23 Hz, 240° that makes the approximation inherit characteristics of the control spectrum in the southern mode. This *heritage* aspect is also noted in other examples to come. The extra features present in the approximated member can be explained by the fact that the results are obtained from a 144 h prediction when the linearization hypothesis is no longer fully valid.

Figures 4 and 5 show the ensemble spread and ensemble mean for 12:00 UTC of 17 June 2000, predicted 36 h before by the non-linear WEPS and by the linearized WEPS, respectively. The ocean wave and atmospheric conditions for the events that we study were characterized by a strong swell front off the coast of Argentina, a wind sea in the northeast of Australia and a predominantely wind sea state in the South Atlantic, due to strong winds on that area.

The ensemble means are found to be very similar. We also notice that the ensemble spread is larger in the non-linear WEPS almost everywhere, with the exception of a region of wind sea in the northeast of Australia. In Figs. 6 and 7 probability distributions of significant wave heights above 1, 2, 4 and 6 m are shown for the two WEPS. Although the results are similar in the different EPSs, we note the stronger peaks at the probabilities above 4 m near northeast Australia and in some areas of the South Atlantic, in the linearized WEPS. Next we analyse individual cases.

Figures 8 and 9 show the significant wave heights for 12:00 UTC of 17 June 2000, predicted 36 h before by the member 1P, by its linearized approximation and by the control run. We notice the approximation is a combination of the control and the member 1P. Focusing our analysis at latitudes higher than 20°S where the perturbation produces relevant differences, it is seen that some of the strong wave events predicted by member 1P and absent in the control are present in the 1P approximation. For instance, the peak regions near (40°W, 55°S) and in the south of New Zealand. This desirable characteristic also occurred in the other member comparisons that we performed.

For predictions with a longer range the linearization hypothesis starts to break down, and although the approximations are

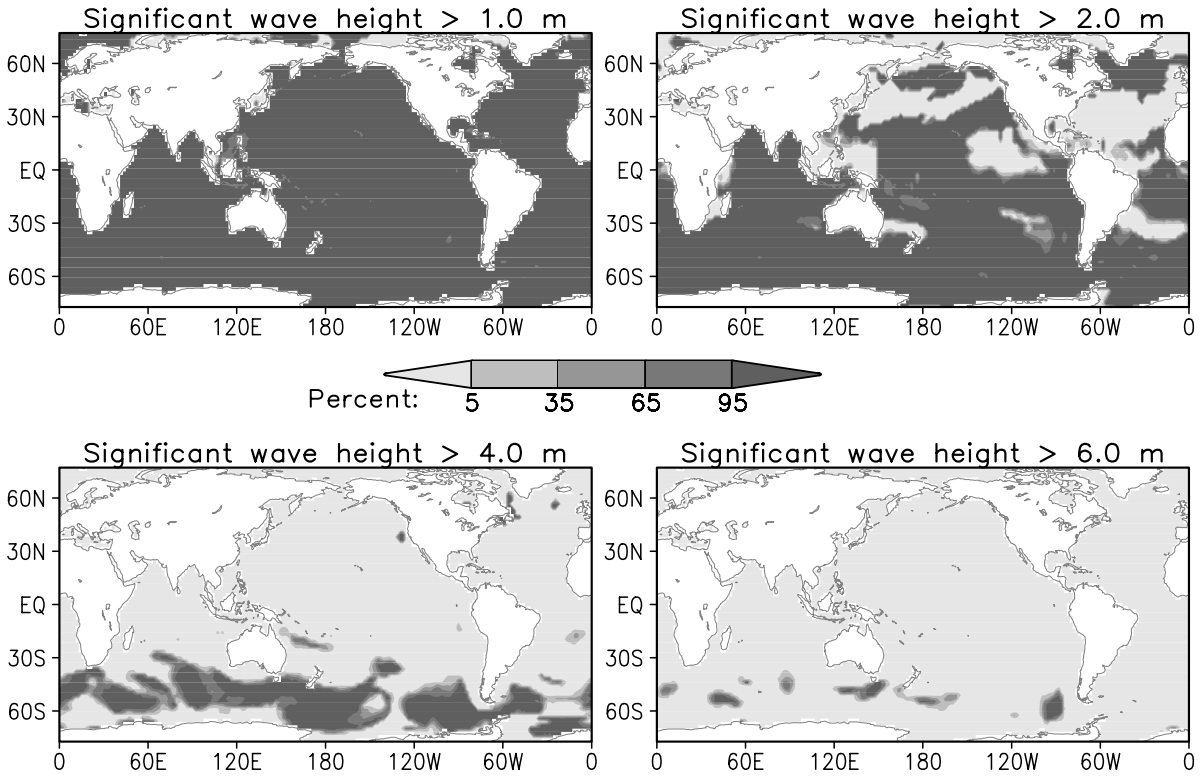

Fig 7. Probability distributions of the 36 h significant wave heights prediction for 12:00 UTC of 17 June 2000 above 1, 2, 4 and 6 m are shown for the linearized WEPS.

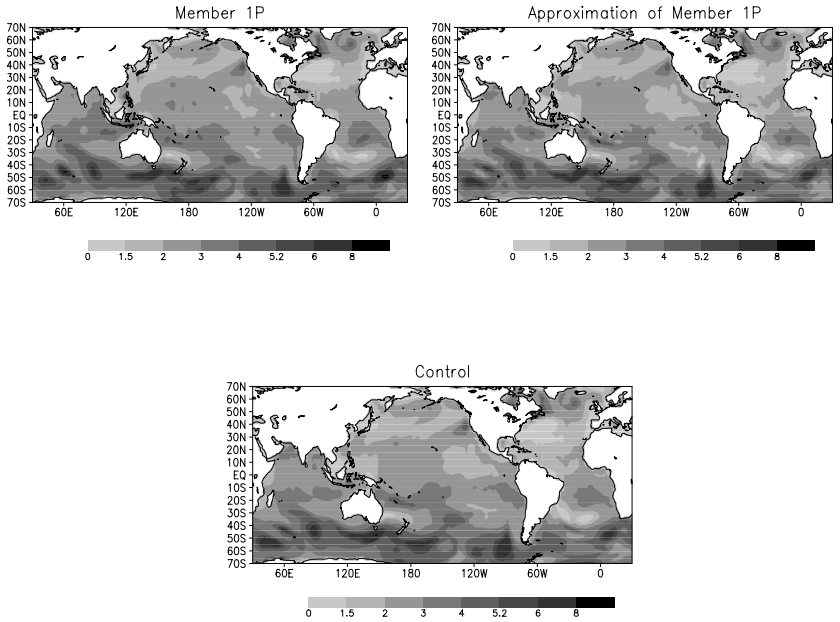

Fig 8. Global significant wave heights for 12:00 UTC of 17 June 2000, predicted 36 h before by the member 1P, by its linearized approximation and by the control run.

smooth excessive spatial variation is observed. Nevertheless very interesting aspects have been observed in our comparisons. For this reason we now show some of the results for 144 h forecasts. Consider the ensemble spreads and means illustrated in Figs. 10 and 11. One important feature observed is that, unlike the results for the 36 h prediction, the spreads for the 144 h prediction show overall larger values in the linearized WEPS than in the non-linear WEPS. The ensemble means have similar values on

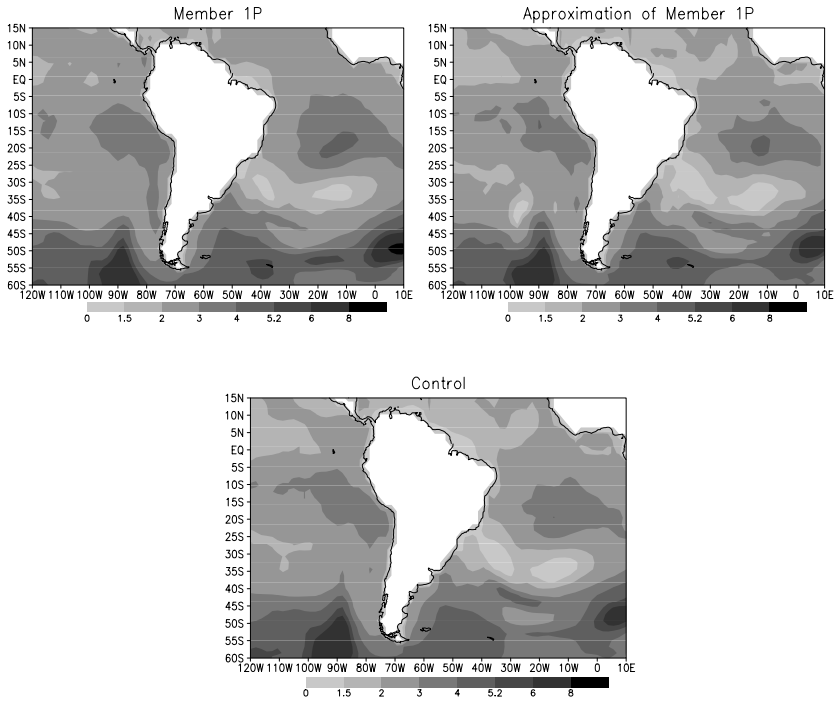

Fig 9. Significant wave heights about South America, for 12:00 UTC of 17 June 2000, predicted 36 h before by the member 1P, by its linearized approximation and by the control run.

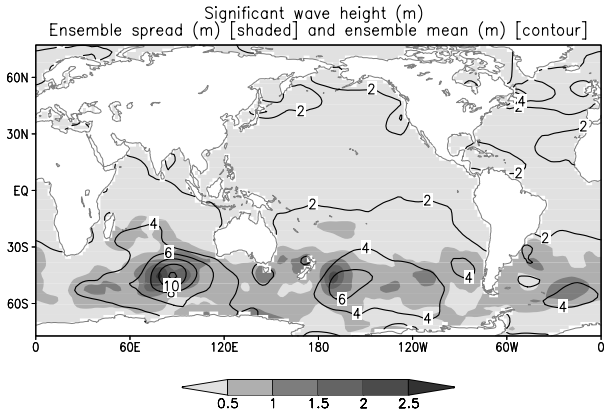

Fig 10. The ensemble spread (shaded) and ensemble mean (contour) of significant wave heights for 00:00 UTC of 22 June 2000, predicted 144 h before by the non-linear WEPS.

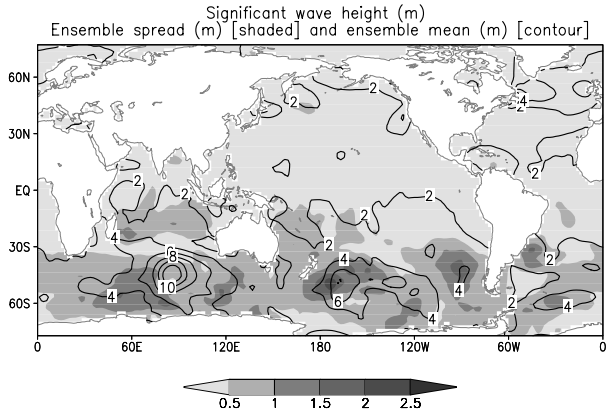

Fig 11. The ensemble spread (shaded) and ensemble mean (contour) of significant wave heights for 00:00 UTC of 22 June 2000 predicted 144 h before by the non-linear WEPS.

translated contour lines. The approximated WEPS shows higher spreads in South Australia, the northern Indian Ocean and to the southwest and east of South America. We remark that the non-linear WEPS indicates a larger spread at roughly 80–90°E and 50°S. Figures 12 and 13 represent the probability distributions for the 144 h forecast. These distributions for waves higher than 2 m are more uniform in the non-linear WEPS while the linearized WEPS shows smaller probability values. The regions with a high probability of wave heights above 4 m are roughly equivalent for both systems with the exception of the 35–65% probability shown in the area near the northwest of Australia and southwest of Sumatra. The probability distributions for waves higher than 6 m are similar in both systems with some extra areas suggesting strong events by the approximated WEPS, mainly in the South Pacific.

Concluding this section, Figs. 14 and 15 show the root mean square errors (RMSE) between the ensemble members and the reference solution (solid lines) and between the approximation and the reference solution (dashed lines). The results are obtained from the simulation of the realistic case described above (see Fig. 2). In fact, the RMSE in the 36 h approximation are overall smaller than the non-linear members themselves. This gives an added value for the linearized solutions. On the other hand, the RMSE for the 144 h approximation are always larger then the RMSE of the original members, as expected. It is interesting, however, to note that both curves in Fig. 15 have their maxima

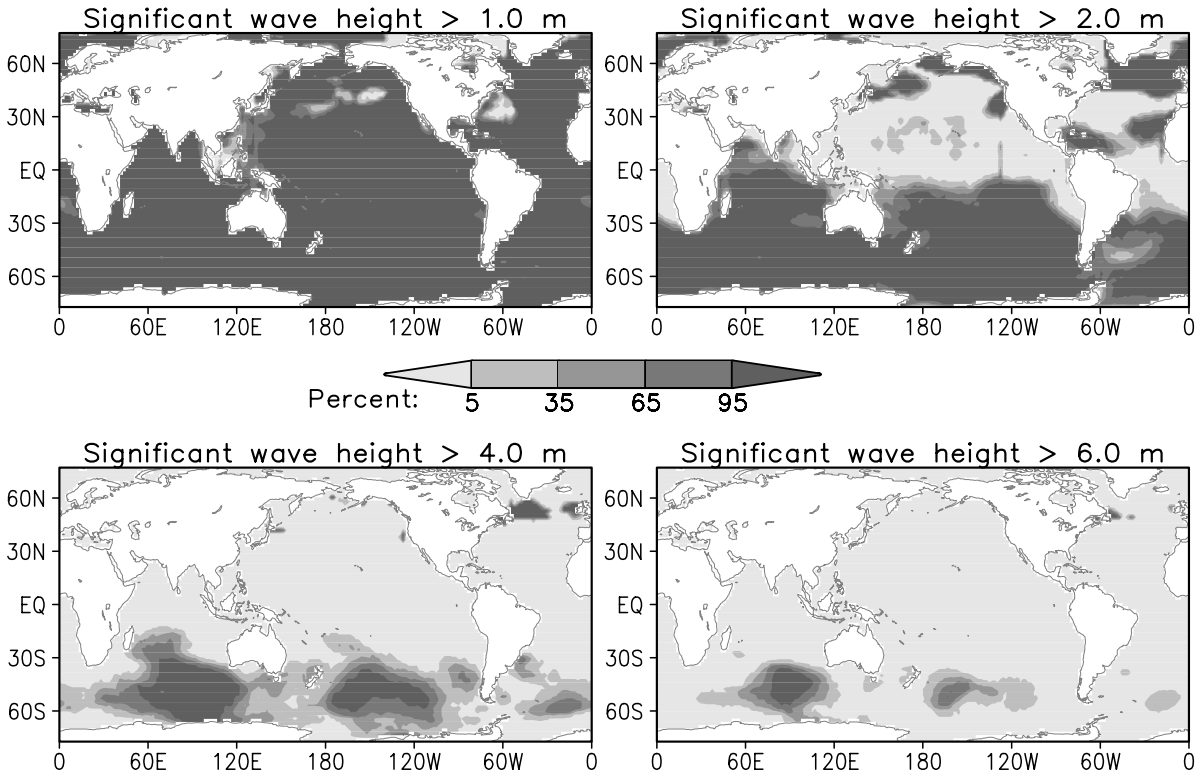

Fig 12. Probability distributions of the 144 h significant wave heights predictions for 00:00 UTC of 22 June 2000 above 1, 2, 4 and 6 metres are shown for the non-linear WEPS.

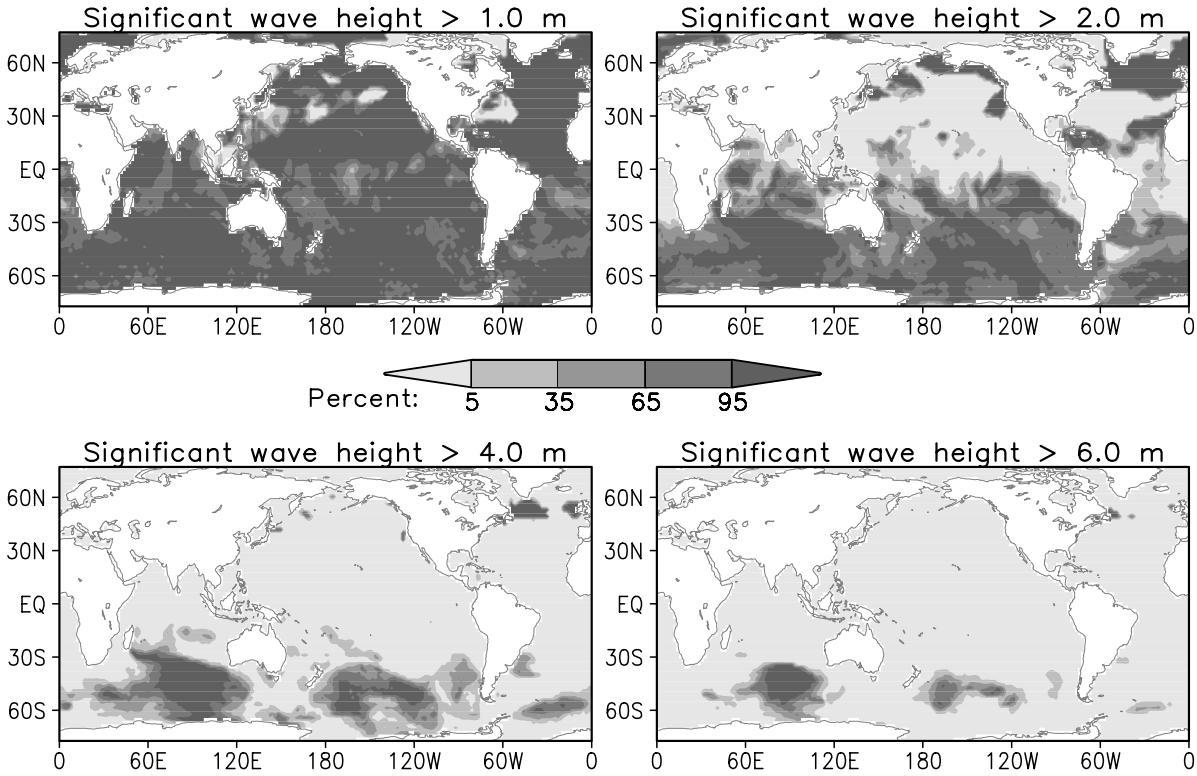

Fig 13. Probability distributions of the 144 h significant wave heights predictions for 00:00 UTC of 22 June 2000 above 1, 2, 4 and 6 m are shown for the linearized WEPS.

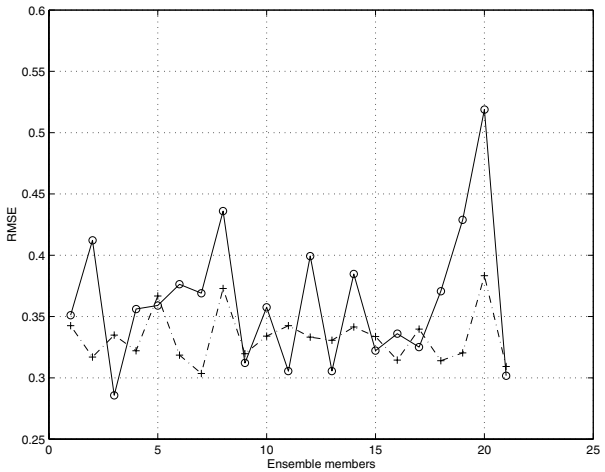

Fig 14. Root mean square 36 h prediction error between the ensemble members and the reference solution (solid line) and between the approximation and the reference solution (dashed line). The members are denoted by the numbers 1 to 20, in the order 1N, 1P, 2N, 2P, . . . , 10N, 10P. The control is represented by abscissa 21.

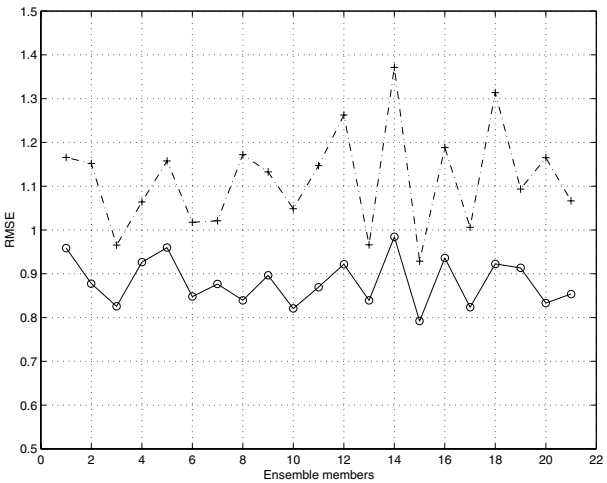

Fig 15. Root mean square 144 h prediction error between the ensemble members and the reference solution (solid line) and between the approximation and the reference solution (dashed line). The members are denoted by the numbers 1 to 20, in the order 1N, 1P, 2N, 2P, . . . , 10N, 10P. The control is represented by abscissa 21.

and minima alternating in the same order, making the delineation of the curves similar. We observe that no particular geographical dependence has been shown by the linearized WEPS.

6. Conclusions

A method for evaluating approximations of members in a ocean wave ensemble prediction system is developed. These approximations are obtained by a single run of the wave model using a linear relation between the difference of two ensemble forcings and the respective difference in the wave spectra. Since all the member approximations can be obtained in a single run, the reduction in the computational cost of an EPS is proportional to the number of members if one substitutes the approximations for the original members.

Numerical results show that typically a member approximation is a combination of the member to be approximated and a control solution obtained by non-linear integration of the model. In the results with forecasts of 36 h, the following behaviour has been observed. Smooth fields of significant wave heights for the approximations are obtained, supporting the validity of the linearization hypothesis. Even though the ensemble spread is larger in the non-linear WEPS, the probability distributions suggest areas of strong wave activity better predicted by the linearized WEPS. Some severe sea states are shown to be predicted only by approximating members. The overall global root mean square errors of the approximations outperform the corresponding non-linear member errors.

For forecasts of 144 h, the linearization hypothesis breaks down, even though interesting results are observed. The selected cases studied indicate and confirm the high sensitivity of the wave model to wind fields.

It is important to observe that for forecasts of 36 h the ensemble spread in the non-linear system is slightly larger than the linearized one. However, this situation is inverted as the prediction time increases: the linearized WEPS presents areas of larger spreads compared with the non-linear WEPS.

The question of how the linearized members would perform with better quality wind predictions and also how well they agree with observational data are points of interest that deserve to be investigated further.

The results suggests that in forecasts of the order of 144 h, the approximations should be calculated *in addition* to the conventional ensemble members. This would double the size of the ensemble with small computational cost and could provide forecast information not present in the original WEPS. The two classes of members in the new ensemble should be clearly identified and separately treated when forecasts are interpreted. We note that due to the hypothesis used in Section 3.1, the method may not be applicable in areas dominated by ocean swell.

The results are still preliminary, and one direction of further research is to attempt to improve the accuracy of approximations in short-range forecasts by using a higher-order approximation.

7. Acknowledgments

We thank Dr Carlos A. Nobre for allowing access to the CPTEC computer resources for the first author in his present institution, Dr Eva Bauer for providing the code for assimilation of wave data using impulse response functions and Professor Mark Thompson for the revision of the text.

Part of this work was carried out while the first author was affiliated with the Centro de Previsão de Tempo e Estudos Climáticos (CPTEC). The first author acknowledges financial support by CNPq (process no 307517/2003-9).

References

Bauer, E., Hasselmann, K., Young, I. R. and Hasselmann, S. 1996. Assimilation of wave data into the wave model WAM using an impulse response function method. *J. Geophys. Res.* **101**(C2), 3801–3816.

Buizza, R., Richardson, D. S. and Palmer, T. N. 2003. Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles. *Q. J. R. Meteorol. Soc.* **129**(C), 1269–1288.

Cavalcanti, I. A. F., Marengo, J., Satyamurty, P., Nobre, C. A., Trosnikov, I. et al. 2002. Global climatological features in a simulation using the CPTEC/COLA AGCM. *J. Climate* **21**, 2965–2988.

Coutinho, M. M. 1999. Previsão por conjuntos utilizando perturbações baseadas em componentes principais. *Masters Dissertation*, INPE, São José dos Campos, Brazil.

Daley, R. and Mayer, T. 1986. Estimates of global analysis error from the global weather experiment observational network. *Mon. Weather Rev.* **114**, 1641–1653.

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

Francisco, G. and Muruganandam, P. 2003. Local dimension and finite time prediction in spatiotemporal chaotic systems. *Phys. Rev. E* **67**, 066204.

Hasselmann, K. 1962. On the non-linear energy transfer in a gravity-wave spectrum, part 1: general theory. *J. Fluid Mech.* **12**, 481–500.

Hoffschildt, M., Bidlot, J.-R., Hansen, B. and Janssen, P. A. E. M. 1999. *Potential Benefit of Ensemble Forecasts for Ship Routing*. ECMWF Technical Memorandum no 287, Reading, UK.

Janssen, P. A. E. M. 2000. Potential benefits of ensemble prediction of waves. *ECMWF Newsletter* **86**, 3–6.

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

Menemenlis, D. and Wunsch, C. 1997. Linearization of an oceanic general circulation model for data assimilation and climate studies. *J. Atmos. Ocean. Technol.* **14**, 1420–1443.

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996. The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.* **122**, 73–119.

Ochi, M. K. 1998. *Ocean Waves, the Stochastic Approach*. Cambridge University Press, Cambridge.

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A. and Ott, E. 2001. Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.* **86**, 5878–5881.

Stammer, D. and Wunsch, C. 1996. The determination of the large-scale circulation of the Pacific Ocean from satellite altimetry using model Green's functions. *J. Geophys. Res.* **101**, 18 409–18 432.

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

Young, I. R. 1999. *Wind Generated Ocean Waves*. Elsevier, Amsterdam.

Zhang, Z. and Krishnamurti, T. N. 1999. A perturbation method for hurricane ensemble predictions. *Mon. Weather Rev.* **127**, 447–469.