Forecast Divergences of a Global Wave Model

DIANA J. M. GREENSLADE

Bureau of Meteorology Research Centre, Melbourne, Australia

IAN R. YOUNG

Swinburne University of Technology, Melbourne, Australia

(Manuscript received 2 July 2004, in final form 31 January 2005)

ABSTRACT

One of the main limitations to current wave data assimilation systems is the lack of an accurate representation of the structure of the background errors. One method that may be used to determine background errors is the “NMC method.” This method examines the forecast divergence component of the background error growth by considering differences between forecasts of different ranges valid at the same time. In this paper, the NMC method is applied to global forecasts of significant wave height (SWH) and surface wind speed (U10).

It is found that the isotropic correlation length scale of the SWH forecast divergence ($L_{SWH}$) has considerable geographical variability, with the longest scales just to the south of the equator in the eastern Pacific Ocean, and the shortest scales at high latitudes. The isotropic correlation length scale of the U10 forecast divergence ($L_{U10}$) has a similar distribution with a stronger latitudinal dependence. It is found that both $L_{SWH}$ and $L_{U10}$ increase as the forecast period increases. The increase in $L_{SWH}$ is partly due to $L_{U10}$ also increasing. Another explanation is that errors in the analysis or the short-range SWH forecast propagate forward in time and disperse and their scale becomes larger. It is shown that the forecast divergence component of the background error is strongly anisotropic with the longest scales perpendicular to the likely direction of propagation of swell. In addition, in regions where the swell propagation is seasonal, the forecast divergence component of the background error shows a similar strong seasonal signal. It is suggested that the results of this study provide a lower bound to the description of the total background error in global wave models.

1. Introduction

Several operational numerical weather prediction (NWP) centers around the world run wave forecasting models that routinely assimilate satellite significant wave height (SWH) data (e.g., Breivik and Reistad 1994; National Meteorological Operations Centre 1999) or directional wave spectra (Bidlot 2002). There has been a considerable amount of research into the development of wave data assimilation systems in recent years (e.g., Lionello et al. 1992; Greenslade 2001; Mastenbroek et al. 1994). A major limitation to current assimilation systems is the specification of the model (or background) errors. This has not previously been explored to any great extent for wave models.

One method that may be used to investigate the structure of the background errors is the “NMC method” (Parrish and Derber 1992; Rabier et al. 1998). It was first given the name in Rabier et al. (1998) in reference to a publication by Parrish and Derber (1992) who were then based at the U.S. National Meteorological Center [NMC; now the National Centers for Environmental Prediction (NCEP)]. This method considers the forecast divergence component of the background error growth. Correlations and variances of the differences between forecasts of different ranges valid at the same time are determined and can be used to construct the background error covariances. In this work, forecasts of SWH from a global wave model are investigated. A general assumption is that most of the errors occurring in wave forecasts are due to errors in the surface winds (Cardone et al. 1995). Therefore, the
surface wind fields that are used to force the wave model are also examined.

A brief review of previous wave data assimilation research is given in section 2 with an emphasis on specification of the wave model background errors. Previous work using the NMC method is also presented here. The wave model used in this work is a version of the third-generation ocean wave prediction (WAM) model (WAMDI Group 1988; Komen et al. 1994). This is described in section 3. The details of the method used to calculate the correlations and variances are presented in section 4. Results for isotropic and anisotropic correlations are presented and discussed in section 5. Further issues are addressed in section 6 and finally section 7 provides a summary.

2. Background

Current operational wave data assimilation systems at, for example, the European Centre for Medium-Range Weather Forecasts (ECMWF; Lionello et al. 1992) and the Australian Bureau of Meteorology (Greenslade 2001) use the sequential method of statistical interpolation (SI) to combine first-guess wave model fields with the observations to obtain analyzed wave fields. Details of the SI algorithm can be found in Lionello et al. (1992) or Greenslade (2001). One of the requirements in the application of SI techniques is the specification of the background error correlation matrix, $P$. This is a symmetric $N_{\text{obs}}$ by $N_{\text{obs}}$ matrix ($N_{\text{obs}}$ is the number of observations) whose element $(k, j)$ is given by

$$P_{kj} = \frac{\langle (H_p^k - T^k)(H_p^j - T^j) \rangle}{\sigma_P^k \sigma_P^j},$$

where $H_p$ is the model first guess (prediction or background), $T$ is the true field, $(\ldots)$ is the expected value, and $\sigma_P$ is the background rms error, that is,

$$\sigma_P^k = \sqrt{\langle (H_p^k - T^k)^2 \rangle}.$$  

In other words, the value of element $(k, j)$ of matrix $P$ is the correlation between the background error at observation location $k$ and the background error at observation location $j$. If the background error (and likewise the observational error, $O$) is known precisely, then the rms error of the analysis can be minimized. In that case, the technique is known as optimal (or optimum) interpolation. However, in practice, $P$ and $O$ are not known exactly as they represent the difference between the background (or observation) and the unknown truth. For this reason, the term SI is used here.

A wide range of structures has been used in the literature to describe $P$. A review of these can be found in Greenslade (2004). They generally have the form

$$P_{kj} = \left( 1 + \frac{|x_k - x_j|}{L} \right)^a \exp\left[ -c \left( \frac{|x_k - x_j|}{L} \right)^b \right],$$

where $L$ is the decorrelation length scale and $|x_k - x_j|$ is the distance between the points $k$ and $j$. The values of $a, b, c$, and $L$ used by various authors are listed in Table 1 and some of the resulting $P_{kj}$ curves are plotted in Fig. 1. The $P_{kj}$ curves are plotted as function of the distance $d = |x_k - x_j|$.

Some of these structures for $P_{kj}$ have been ad hoc estimates, while others have provided some justification for the choice of background error structure. For example, they may be based on knowledge of the error structures of atmospheric fields, which are generally better known. Most of the studies have assumed isotropy and homogeneity in the background error structure, with the exception of the studies in which $L$ is defined in degrees of latitude and longitude. (This means that the background errors are defined to have a larger zonal spatial scale at low latitudes than at high latitudes, resulting in both anisotropy and inhomogeneity.) To date, there has been no extensive effort made

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lionello et al. (1992)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>15°</td>
</tr>
<tr>
<td>Mastenbroek et al. (1994)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-135 km</td>
</tr>
<tr>
<td>Breivik and Reistad (1994)</td>
<td>0</td>
<td>2</td>
<td>1/2</td>
<td>200–40 km</td>
</tr>
<tr>
<td>Young and Glowacki (1996)</td>
<td>0</td>
<td>2</td>
<td>1/2</td>
<td>350 km</td>
</tr>
<tr>
<td>Greenslade (2001)</td>
<td>0</td>
<td>2</td>
<td>1/2</td>
<td>300 km</td>
</tr>
<tr>
<td>Voorrips et al. (1997)</td>
<td>0</td>
<td>3/2</td>
<td>1</td>
<td>200 km</td>
</tr>
</tbody>
</table>

![Fig. 1. Some example functions used for the background error correlations at midlatitudes.](image-url)
to determine the spatial scale of the background errors in wave models on a global basis. A major limitation to current implementations of SI systems and indeed potential future systems is the lack of an accurate representation of the background errors.

There are several methods that have commonly been used to estimate the background error correlations in meteorology and oceanography. One of these is the observational method of Hollingsworth and Lönnberg (1986). This method uses observations from a long-term, dense, homogenous observational network and examines the difference between the observations and the background field. Greenslade and Young (2004, hereafter GY) used altimeter data to determine background errors in SWH. Their results showed that the length scales of the background errors varied significantly over the globe, with the largest scales at low latitudes and shortest scales at high latitudes. Very little seasonal or year-to-year variability in the correlation length scales was detected. Conversely, the magnitude of the background error variance was found to have considerable seasonal and year-to-year variability. These results will be discussed further in section 6.

Ensemble techniques can also be used to estimate background error correlations. For wave models, which are strongly forced systems, ensembles could be produced by perturbing the wind forcing fields or by using forcing fields from different forecasting centers. The differences between wind fields from different centers may provide an indication of where the greatest uncertainties lie. Another way to obtain different forcing fields could be to force the wave model with only 24- and 48-h forecast wind fields and compare the resulting wave fields.

In data assimilation, the background field is typically a model forecast from the previous assimilation step, that is, the previous analysis. Therefore, the background error consists of two components: the model error and the analysis error. The model error describes the amount by which the error due to the model has grown since the previous analysis time. This could be due to either errors in the wind forcing, or errors internal to the wave model. The analysis (i.e., the wave field after the observations have been incorporated via data assimilation) will also include some error. This is because the length scales and error variances defined in the assimilation system represent expected values and are unlikely to be correct at all places and at all times. In addition, most areas of the ocean are not updated during every assimilation period. Observational error also precludes a perfect analysis. The NMC method examines only the forecast divergence component of the total background error. While this does not give a complete picture of the background error, an examination of the forecast divergence alone can provide a guide to the structure of the background errors, and can also provide some insight into the error growth in wave models.

The forecast divergence is estimated by examining the difference between pairs of model forecasts of different ranges (e.g., a 48- and a 24-h forecast) valid at the same time, using the shorter-range forecasts as the validation. This method has been used to examine background errors in NWP systems at several forecasting centers. One of the first studies to apply this technique to an atmospheric model was Baker et al. (1987). They considered the differences between analyzed and 6-h forecast geopotential height over a 60-day period. Larger correlation scales were found in the Tropics than in the midlatitudes. It was suggested that the geographical variation could be due to deficiencies in the analyses due to the lack of data in some areas.

In Parrish and Derber (1992), 24-h forecasts were compared with model analyses over a period of 30 days. The forecast error variance over 24 h could therefore be determined. This was then empirically scaled to convert the variances to values representative of the 6-hourly interval at which data assimilation was performed. No attempt was made to examine the spatial structure of the forecast errors directly using this method.

Rabier et al. (1998) adapted and extended the method by considering the differences between 48- and 24-h forecasts of model variables such as temperature and specific humidity. It was suggested that although the choice of forecast fields used is fairly arbitrary, there are certain advantages in using the 24- and 48-h fields. First, the use of the 24-h field as the verifying field avoids issues involved with using the analysis. For example, there may be problems in the analysis associated with the irregular spacing of observations. In addition, the 24-h difference between the forecast periods is long enough so that the forecasts are not too similar and short enough so that the results will still be applicable to the desired 6-hourly time period, that is, it will still be within the linear error growth regime. Both error variances and correlation scales were considered in Rabier et al. (1998). It was shown that results using this method compared reasonably well to results using other methods, for example, the observational method (Hollingsworth and Lönnberg 1986).

Other authors that have used this method to determine background error statistics for atmospheric systems include Steinle et al. (1995), Desroziers et al. (1995), and Ingleby et al. (1996).

The method has been applied to SWH fields in Voor-
rips (1998). The spatial correlation structure of the background error for total wave energy was estimated by considering differences between the model forecast at 24 h and the model analysis over a 2-month period within the North Sea. The resulting structure for the background error correlation matrix for SWH is listed in Table 1 and shown in Fig. 1.

There are some advantages to the NMC method over the observational method. These are mainly related to not having to rely on the availability of a high-quality observational dataset. The method uses only model output that is generally on a regular grid. The method is relatively straightforward and easy to apply and information on the background errors can be obtained in all directions, on all spatial scales (larger than the resolution of the model grid), and at all locations. In addition, for NWP systems, it provides the opportunity to examine the error structure of fields that are difficult to observe, for example, humidity.

However, there are limitations to its application. The inherent assumption that the structure and magnitude of the forecast divergence is similar to that of the forecast errors is difficult to justify. If the structure of the forecast errors varies little over the time range being examined, then this is a reasonable assumption, but in the case of SWH, the long-term swell component of the wave field can complicate the situation substantially. This is related to the issue of needing to remain within the linear error growth regime. The SWH field may be better represented by the sum of two different linear error growth regimes: the wind–sea regime and the swell regime.

3. Wave model

The modeled wave fields used in this work are those produced operationally and archived at the Bureau. The wave model run at the Bureau is the Australian version of the WAM model (AUSWAM; National Meteorological Operations Centre 1999). The WAM model is a third-generation wave model that solves the wave transport equation explicitly without assuming a form for the evolving spectrum. The wave transport equation is

$$\frac{\partial F}{\partial t} + \nabla \cdot (c_g F) = S_{in} + S_{nl} + S_d,$$

where $F(f, \theta)$ is the wave spectrum (i.e., the spectral density of wave energy) as a function of frequency ($f$) and direction ($\theta$), $c_g$ is the group velocity, and the terms on the right-hand side represent the source terms: $S_{in}$ is the energy input due to wind forcing, $S_{nl}$ the nonlinear energy transfer between groups of resonant waves, and $S_d$ the dissipation of energy due to white capping.

In the most recent version of the WAM model (cycle 4, released in December 1991) the physics included a dynamic coupling between wind and waves (Janssen 1989, 1991). Bender (1996) compared the original cycle-3 (Snyder et al. 1981; Komen et al. 1984) and cycle-4 physics to independent waverider buoy data over a 1-month period (July 1992) and concluded that a wave forecasting model for the Australian region should be based on the cycle-3 physics. Although this led to underprediction of wave heights, it was shown that by upgrading the first-order propagation numerics to third order (which led to overprediction of wave height) and increasing the magnitude of the dissipation term by approximately 40%, a good match with the buoy observations could be achieved. Therefore, the present implementation of the WAM model at the Bureau uses the cycle-3 physics with increased dissipation and third-order upwinding numerics.

Wave spectra are discretized into 12 directional bins, centered at 15°, 45°, 75°, etc. This “staggering” of the directional bins is to avoid having spectral energy propagating directly along the axes of the north–south coordinate system (Bidlot et al. 1997). There are 25 frequency bins ranging from 0.0418 to 0.4114 Hz. Deepwater physics only is used. The propagation and source term time steps are 20 and 10 min, respectively. For the global version of the model, the north–south extent of the domain is 78°N–78°S. Surface parameters such as SWH are archived every 12 h.

Forcing fields for the global wave model are wind velocities at 10 m above sea level. These are obtained from the Bureau’s global atmospheric model (GASP; Seaman et al. 1995). Surface winds are obtained from the lowest level of GASP via Monin–Obukhov theory with empirical stability functions (Garratt 1992). The time period considered in this work is a 2-yr period from April 1999 to March 2001. This time period was chosen because there were no major upgrades or changes in resolution to either GASP or AUSWAM during this time. The 10-m wind fields are instantaneous snapshots of the surface fields and are provided to the wave model at 12-hourly intervals and 2.5° spatial resolution. These are linearly interpolated in time to 3-hourly intervals, and bilinearly interpolated in space to the resolution of the wave model grid. The spatial resolution of the wave model is 3° and wave forecasts are made out to 96 h. The only relevant change made to the operational system during this time period is that the assimilation of the European Remote Sensing satellite (ERS-2) SWH data commenced in August 1999 (Greenslade 2001). Thus, data assimilation
was being performed over most of the time period considered here.

4. Method

In this work, 48-h forecasts of SWH and 10-m wind speed (U10) are compared to 24-h forecasts valid at the same time over the 2-yr period to examine the magnitude, spatial scale, and temporal and geographical variability of forecast divergence. In addition, over a period of 1 yr, forecasts valid at all time periods greater than 24 h (up to 96 h) are compared to the 24-h forecast in order to examine the variability of the forecast divergence with forecast period.

The background error correlation matrix in Eq. (1) can be expressed as the spatial error correlation between two points, $j$ and $k$ (Daley 1991):

$$R_{jk} = \frac{(O_j - B_j)(O_k - B_k)}{\sqrt{(O_j - B_j)^2 (O_k - B_k)^2}} = \rho(r, \theta).$$

(5)

For the purposes of this work, $B$ represents a forecast field (of either SWH or U10) and $O$ is the verifying field valid at the same time (i.e., the 24-h forecast field) so Eq. (5) actually describes a forecast divergence correlation matrix. Correlations of the forecast divergence are represented as $\rho(r, \theta)$, where $r$ is the great circle distance between points $j$ and $k$ and $\theta$ is the angle (from north) of the line joining them. The temporal mean SWH or U10 is removed at each grid point to ensure that the difference fields are unbiased. Correlations are computed for 3-monthly time periods and boxes of side length 20° in latitude and longitude globally. The average represented by the overbar is the time average over all matching pairs of model fields at 12-hourly intervals within the 3-month time period. The correlations are then bin averaged into bins of 10-km width and 5°.

An example of a 48–24-h (t48–t24) forecast divergence correlation for SWH is shown in Fig. 2a. This example is for a 20° box in the southeastern Pacific Ocean centered at (40°S, 110°W) for the 3-month period April–June 2000. Figure 2b shows the same correlation as a function of distance alone, $\rho(r)$. As expected, the correlations are highest at short distances and decay toward longer distances. The decay in the two-dimensional correlation function is slower in the southeast direction than the northeast direction. In other words, over the 24-h forecast period the spatial scale of the forecast divergence is larger in the southeast direction than in the northeast direction. This behavior can be interpreted as follows. Consider a model grid point within the domain. Assume that at this grid point, at a particular time within this 3-month period, the difference in SWH between the 48-h forecast and the verifying 24-h forecast is large. Then according to Fig. 2a the SWH at a point $x$ km to the southeast of this grid point is more likely to also differ between the 24- and 48-h forecast than the SWH at a point $x$ km to the northeast of this grid point. Note that the correlation functions in Fig. 2 do not provide any information about the magnitude of the forecast divergence, but only the spatial scales and the directions in which the forecasts are most likely to diverge. To estimate the magnitude of the expected difference between 48- and 24-h forecasts, the variance needs to be examined (see below).

Analytic descriptions of the correlations are obtained by applying curve-fitting procedures to the 3-monthly correlations. In addition, average correlations over longer time periods are obtained by averaging the 3-monthly correlations together binwise and then applying the curve-fitting procedure. The following forms were selected after testing many variations of Eq. (3) (and its two-dimensional equivalent) and selecting the curve that resulted in the lowest mean-square error (mse) on average (Greenslade 2004). The analytic form used for the two-dimensional, or anisotropic case is

$$\rho(r, \theta) = \exp\left(-\frac{d}{a_3}\right),$$

(6)

with

$$d^2 = r^2\left[\frac{1}{a_1^2}\cos^2(\theta - a_2) + a_2^2 \sin^2(\theta - a_2)\right],$$

(7)

where $a_1$, $a_2$, and $a_3$ are the parameters to be fitted. The best-fit surface for the example shown in Fig. 2a is represented in the figure by the solid ellipses. In addition, the isotropic, or one-dimensional case is considered. The analytic function that resulted in the lowest mse for the isotropic correlations is the second-order autoregressive function, that is,

$$\rho(r) = \left(1 + \frac{r}{a_1}\right) \exp\left(-\frac{r}{a_1}\right),$$

(8)

and the best-fit curve is again shown in Fig. 2b. For this example, the value of $a_1$ for the isotropic case is 276 km. This is the isotropic length scale of the spatial correlation of the t48–t24 forecast divergence. For the remainder of this paper, this will be referred to as $L$, the isotropic correlation length scale.

The variance of the forecast divergence is also calculated as follows:

$$E_{d}^2 = \frac{1}{K} \sum_{k=1}^{K} (B_k - O_k)^2,$$

(9)
where $K$ is the number of model grid points within the $20^\circ$ box.

5. Results

First, the mean results over 1 yr are considered for t48–t24 during the year April 2000–March 2001. Note that wave data assimilation was included in the wave forecasting system throughout this time period. The operational Bureau data assimilation system has $P_{hj}$ defined with a length scale of approximately 200 km. This means that variability with this length scale is being inserted into the analysis. It is difficult to know exactly what the impact of this would be on estimates of the spatial structure of forecast divergence.

Maps of the isotropic correlation length scale over the globe are shown in Fig. 3 for SWH ($L_{SWH}$) and U10 ($L_{U10}$). It can be seen that the geographical distribution of the length scales is similar for both parameters. The longest length scales occur just south of the equator in the eastern Pacific and Atlantic Oceans. It can be seen that there is in general a strong latitudinal dependence,
with the smallest $L$ at high latitudes. The regions in which $L$ is large are those in which the two forecasts differ on a large spatial scale. The physical basis for these geographical distributions will be discussed further in section 5a(1).

To represent the anisotropic case, the 0.5-level contour line of the two-dimensional analytical correlation function $\rho(r, \theta)$ (see Fig. 2a) is drawn at the center of each 20° box. For the remainder of this work, these will be referred to as “error ellipses.” These are shown in Fig. 4. The relative size of the ellipses generally follows the pattern of the isotropic length scale. For SWH, the ellipses are generally anisotropic, with the longest scales in the east–west direction or close to it. There are some areas where there is less anisotropy, for example, at high latitudes. The error ellipses for U10 follow a similar pattern, with a tendency to be less anisotropic at lower latitudes. The implications of the anisotropy will be discussed further in section 5a(2).

a. Forecast period

The impact of the length of the forecast period is now considered. In particular, forecasts of different ranges are compared to the 24-h forecast. Again, forecasts during the year April 2000–March 2001 are examined.

1) ISOTROPIC CASE

Figure 5 shows how $L_{SWH}$ varies with forecast period. A noticeable feature of Fig. 5 is that the correlation length scales become larger everywhere as the forecast period increases. For example, in the central Indian Ocean, the length scale increases from approximately 300 km for t36–t24 to more than 450 km for t96–t24. Note also the longest scales in the eastern Pacific Ocean lengthening from 400 to 650 km as the forecast period increases. This has been shown to be true for atmospheric systems also. For example, Bengtsson and Gustavsson (1971) show how the spatial scale of forecast error autocorrelation of 500-mb geopotential height increases as the forecast period increases. In other words, for short-range forecasts, most of the error is in the short scales while longer-range forecasts have increasing errors in the larger scales. The trend toward larger spatial scales for longer forecast periods seen in Fig. 5 may be due to this trend occurring in the wind fields and being transferred into the wave fields. This can be investigated by examining the length scale of the forecast divergence of the winds used to force the wave model. This is shown in Fig. 6.

It can be seen that $L_{U10}$ does lengthen somewhat as the forecast period increases, but not to the same extent as in the SWH forecast divergence. For example, in the central Indian Ocean $L_{U10}$ lengthens from approximately 300 km to more than 350 km and in the eastern Pacific Ocean, it lengthens from 450 to 550 km. This suggests that the lengthening of scales in the surface wind forecast divergence does not fully explain the lengthening of scales seen in the SWH forecast divergences.
For SWH, there is another simple physical explanation. The swell component of a wave field is that component of the wave spectrum that exists independently of the local surface winds. Errors in SWH that exist in the short-range forecasts (either due to errors in the analysis, errors in the wind forcing, or errors internal to the wave model) can propagate forward in time, independent of the wind forcing in the longer-term forecasts, similar to the way in which swell propagates. As these “swell errors” propagate, they disperse and their spatial scale becomes larger. Thus, the differences between short-range and long-range forecasts will occur on larger scales. In other words, the errors existing in the SWH fields at longer forecast periods are more affected by large-scale swell than the errors at shorter forecast periods. The lengthening of $L_{\text{SWH}}$ seen in Fig. 5 is likely to be due to a combination of this error propagation and the wind error pattern.

The latitudinal dependence of $L_{U10}$ is more pronounced than that of $L_{\text{SWH}}$. This supports the idea that the forecast divergence of the SWH is made up of two components: one that is strongly coupled to the wind field and therefore strongly latitudinally dependent, and a component consisting of the propagation of swell errors, which superimposes a nonlatitudinally dependent pattern.

2) ANISOTROPIC CASE

The variation of the anisotropy of the forecast divergence with forecast period is now considered. Figure 7 shows plots of the error ellipses for increasing forecast periods for three different locations. Consider the box centered in the Indian Ocean at (30°S, 80°E; Figs. 7a,b). For U10, the ellipses become slightly more anisotropic as the forecast period increases, but not to the extent that the SWH ellipses do. So, in this area, the increase of $L_{\text{SWH}}$ with the forecast period seen in Fig. 5 is due only to the lengthening of the scales in one particular...
direction. Since the U10 ellipses do not vary much over the forecast period, this reinforces the hypotheses that 1) the increase in scales in the SWH forecast divergence is due to swell errors propagating and dispersing, and 2) the SWH error regime consists of two separate components—the swell error and the wind–sea error.

The SWH error ellipses in Fig. 7b lengthen predominantly in one direction, that is, the spatial scale of the differences between long- and short-term forecasts is larger in the southeast direction. In this region there are frequent swell systems propagating toward the northeast throughout the Indian Ocean. These are forced by the storm systems that move from west to east along 40° to 50°S. The error ellipses are aligned with their short axes in the direction of propagation of the swell. It is likely then, that the structure of the correlations in this area is related to errors in the generation of swell within the wave model. The swell errors propagate and disperse perpendicular to their direction of propagation.

In the eastern Pacific (Figs. 7c,d), the opposite occurs: the U10 ellipses vary very little over the forecast period, but the SWH ellipses become larger in all directions. This suggests that the errors in the SWH field at this location are arriving from or are propagating in many directions. This is reasonable, as over a period of a year, this region of the ocean is exposed to swell (and therefore swell errors) arriving from both the Southern Ocean and the northern Pacific Ocean.

The North Atlantic location (Figs. 7e,f) is close to the European coast. A similar trend to that seen in the Indian Ocean location is seen here, but with the ellipses aligned toward the northeast rather than the southeast. This suggests that the errors in the SWH field at this location are mainly due to errors in swell systems propagating from the northwest.

3) Variance

The magnitude of the variance of the forecast divergence [see Eq. (9)] for two different forecast ranges versus the 24-h forecast is shown in Fig. 8. These plots...
represent the average variance over the year April 2000–March 2001. For each 20° box there are the same number of model grid points in any time period. In addition, there is approximately the same number of time levels for each 3-month time period. So it is possible to obtain variance statistics for time periods longer than 3 months by simply calculating the arithmetic mean of the variance from individual 3-month time periods (bearing in mind that this procedure assumes stationarity). The distribution of the variance shown in Fig. 8 is similar for t36–t24 and for t96–t24. Maximum variances occur in the North Pacific and at around 50°S.
in the Indian Ocean sector of the Southern Ocean. These are the areas where SWH is generally highest (Young 1999).

As can be seen in Fig. 8 the magnitude of the variance for \( t_{96} - t_{24} \) is considerably larger than that of \( t_{36} - t_{24} \). This can also be seen in Fig. 9 where the average variance over the globe of the difference between each forecast (at 12-hourly intervals) and the 24-h forecast is shown. (Again, note that calculating the arithmetic mean of the variance over the globe assumes homogeneity, and Fig. 8 demonstrates that this is clearly not the case). The background error variance should increase with time until it reaches the level of climatological error variance, where it should asymptote (Daley 1991).

It can be seen that while the variance of the forecast divergence does indeed increase with time, there are no signs of it asymptoting within this time period.

A data assimilation system requires a value of the background rms error, which is appropriate for the time interval at which data is inserted into the model. The choice of this time interval requires a trade-off to be made between having a large time window (which ensures more observations available) and a smaller time window (in which the assumption of simultaneous observations in the SI system is more reasonable). The operational Bureau wave data assimilation performs the assimilation every 3 h. The issue of sparsity of the altimeter data is likely to be alleviated by the use of observations from multiple altimeters (Skandrani et al. 2004). From Fig. 9 the appropriate spatially averaged value of \( E_B^2 \) [see Eq. (9)] at 3 h (i.e., \( t_{27} - t_{24} \)) is approximately 0.02 m\(^2\). However, as discussed in section 2, the values shown here represent forecast error divergence—the appropriate value would be 0.02 m\(^2\) only if the analysis were perfect.

b. Seasonal variability

In their study of atmospheric variables, Rabier et al. (1998) found that the isotropic and homogenous component of the forecast error was very stable with season. The results of GY also showed that there was little seasonal variability in \( L_{\text{swth}} \). However, there are grounds for expecting to see seasonal variability in the forecast divergence of SWH. For example, during Southern Hemisphere (SH) winter, swell arriving in the central Pacific would mostly be from the SH, while during Northern Hemisphere (NH) winter, one would expect the central Pacific to be dominated by swell propagating from the NH. This is therefore best investigated by examining the seasonal variability of the anisotropic forecast divergence.

To obtain a time series longer than one season, correlations for specific 3-month seasons over 2 yr are averaged together, as described in section 4. An even longer time series would be desirable, but the time period used here is limited to 2 yr in order to avoid any major changes in the operational system. It has been shown (in GY) that system changes such as increases in the frequency of the forcing wind fields can have considerable impact on the SWH error correlation structure.

In this section, the seasonal variability of the error ellipses for the locations shown in Fig. 7 is considered. First consider the box centered at (30°S, 80°E), in the central Indian Ocean. Seasonal variability is shown here in Fig. 10. Figures 10a,b show the error ellipses for

![Fig. 8. Mean variance (m\(^2\)) during the year Apr 2000–Mar 2001 of (a) \( t_{36} - t_{24} \) and (b) \( t_{96} - t_{24} \) forecast.](image)
U10 and SWH for January–March 2000 and 2001 and Figs. 10c,d are for July–September 1999 and 2000. It can be seen that the spatial scales of the correlations for both parameters are generally larger in SH winter and slightly more tilted, but generally, there are no large differences in angular distribution between the two seasons. This region has swell generated in the Southern Ocean and propagating toward the Australian coast all year. The shape and size of the Indian Ocean (in particular its very small NH area) means that there is very little seasonal variability in the generation of swell in this region.

On the other hand, one would expect the Pacific Ocean to show a seasonal signal. As mentioned earlier, there should be more swell generated in the hemisphere experiencing winter and this would dominate the error ellipses in forecasts in the central Pacific Ocean. Figure 11 is the same as Fig. 10 except it shows results for the box centered in the eastern equatorial Pacific Ocean. It can indeed be seen that the SWH fields (Figs. 10b,d) show error ellipses tilted approximately perpendicular to the direction from which one expects swell to be propagating. The U10 ellipses for SH summer are quite different to those for SH winter, and these are not strongly related to the patterns in the SWH error ellipses. This demonstrates that the signal in the SWH fields must be related to swell propagation, and is not related to the local wind fields.

6. Discussion

These results can be compared to the results of GY, in which the same problem was addressed, but with a different method. In GY, the observational method of Hollingsworth and Lönnberg (1986) was used. Correlations of the differences between satellite altimeter observations of SWH and modeled SWH were examined globally. The modeled wave fields used were from AUSWAM at 0.5° spatial resolution over the 4-yr time period April 1998–March 2002. The winds used to force AUSWAM were from the data assimilation cycle of the atmospheric model GASP, described in section 3. Assimilation of wave observations was not included in the 4-yr wave model run.

The major results of GY are reproduced here in Fig. 12. This figure can be compared to the various panels in
Fig. 5. It can be seen that the global distribution of isotropic correlation length scales obtained from the forecast divergence method is in general quite similar to that obtained from the observational method. Note however, that the magnitude of the length scales in Fig. 12 is larger than that of even the t96–t24 forecast divergence in Fig. 5. The correlations considered in GY were calculated from the difference between “analyses” and observations, where the analyses are wave fields forced by winds from the data assimilation cycle of the atmospheric model. However, any component of the wave fields that was incorrectly generated would propagate forward in time and disperse and become large-scale swell error. Thus, the modeled wave fields examined in GY represent long-term forecasts, considerably longer than the t96–t24 forecast range considered here. Therefore the length scales calculated from those modeled wave fields and shown in Fig. 12 are likely to be overestimated.

As discussed earlier, the error structure that should be defined within a data assimilation system is the structure of the background error at the frequency of data assimilation. If the present results using the NMC method are used as a proxy for the background error (i.e., ignoring analysis error), then this means that both the error variance and the length scale must be scaled down to 3 h. Figure 5 suggests that if the length scale grows linearly, then at a forecast range of 3 h, the appropriate length scale is quite short. If a perfect analysis were produced then the appropriate error structure to use is indeed the forecast divergence structure found in this work. However, as discussed earlier, the background error describes the difference between the background and the truth, and this must include analysis error, error in the forcing and forecast (or model) error. No analysis is perfect: the length scale and error variances defined in the assimilation system represent

![Fig. 11. Same as Fig. 10 but for a box in the eastern Pacific centered at 0°, 90°W.](image)

![Fig. 12. Isotropic SWH model error correlation length scale (km) over the globe using the observation method (reproduced from GY).](image)
expected values and are unlikely to be correct at all times. In addition, most areas of the ocean are not updated every assimilation period. This means that there will always be some long-term or remotely forced errors in the analysis and thus also in the forecast, and so the length scales and variances based on the forecast divergences alone are likely to be underestimated. One way to address this issue could be to use a combination of results from GY (in which there is likely to be too much remotely forced error) and results from this work (in which there is too little remotely forced error). Thus the results of GY provide an upper bound to the error structure of the background errors and error variance and the results of this work provide a lower bound. Note that there is quite a large difference between the length scales shown in Fig. 12 and those in Fig. 5. This suggests that there still remains significant uncertainty in our knowledge of the structure of the background errors in wave models.

In section 5b it was shown that the swell component of the error in forecast wave fields is likely to be anisotropic and to have a seasonal signal. The question of how to incorporate this information in a data assimilation system arises. To take advantage of this, one would need to know whether the error in the modeled wave field is in the wind–sea or swell component of the wave spectrum. Most current operational wave data assimilation systems assimilate satellite altimeter observations, which provide estimates only of SWH. This does not provide any information on whether the error is in the swell or wind–sea portion of the spectrum, so it is not easy to use this information. However data assimilation systems that use synthetic aperture radar observations of the wave spectrum (e.g., Hasselmann et al. 1997; Breivik et al. 1998; Aouf et al. 2005, manuscript submitted to *J. Atmos. Oceanic Technol.*), or in situ buoy observations may be able to apply these results in updating the wave spectrum.

7. Summary

In this work, the “NMC method” has been used to examine the structure and magnitude of the forecast divergences of an operational wind/wave forecasting system. The forecast error divergence is used as a proxy for the short-range forecast error. It is found that the isotropic correlation length scale of the SWH forecast divergence has considerable geographical variability, with the longest scales (approximately 450 km for the t48–t24 forecast divergence) just to the south of the equator in the eastern Pacific Ocean, and shortest scales (approximately 250 km) at high latitudes.

It was shown that $l_{SWH}$ increases as the forecast period increases. This is partly due to the length scales of the surface wind forecast divergence increasing. Another explanation is that errors in the short-range SWH forecast propagate forward in time and disperse and their scale becomes larger. Examination of the anisotropy in the forecast divergences confirmed that longer-range forecasts are likely to be dominated by errors in swell systems. This error has the largest spatial scales perpendicular to the expected direction of propagation of swell.

The results compared well, qualitatively, to those found using an alternate method in GY, but differ in magnitude. Limitations to both methods were discussed and it was suggested that the results in this work should be used as a lower bound and the results of GY as an upper bound to estimates of the structure of the background error in global wave models. Typical operational wave data assimilation systems define the spatial structure of the background error to be homogenous over the globe. This work and that of GY has shown that the spatial scale in fact varies significantly in space. It is generally larger than the typical values used in operational systems, particularly at low latitudes. These results can easily be incorporated into a background error correlation matrix, and could be expected to provide more accurate wave analyses and therefore wave forecasts. This will be tested in a wave data assimilation system in future work.

**Acknowledgments.** The authors are grateful for the comments and suggestions provided by Oscar Alves, Peter Steinle, and three anonymous reviewers.

**REFERENCES**


——, ———, H. Schyberg, and J. Sunde, 1998: Assimilation of ERS


