ArticleReach Paging Slip / Cover Sheet

10/13/2011

Citation Information

Journal: Journal of Atmospheric & Ocean Science

Article: The impact of inhomogenous background errors on a global wave data assimilation

Author: , Diana J. M. Greenslade & Ian R. Young

ISSN: 1741-7538

EISSN:

Volume: 10

Issue: 2

Quarter:

Season:

Number:

Month:

Day:

Year: 2005

Pages: 61 - 93

Citation Source: null

Patron DOI: 10.1080/17417530500089666

Note:

Staff Note:

Paged Locations: za006, >>9geor<<, 6wash

Holdings Information

Location: INTERNET

Call Number:

HOLDINGS: (03 01, 2005)- (12 31, 2005)

Request ID: 594697

Central Ariel IP 205.227.91.137

Central Ariel arlir, delivery

Patron:

Ariel IP Address

>>

<<

Odyssey IP Address

Patron's Library Information

Address: Australian National University, Menzies Library
(Bldg 2), CANBERRA ACT 0200,
AUSTRALIA

Email: docdel@anu.edu.au

Fax: 011-61-2-6125-4063

Patron Information

Patron Number: .p1082211@9anu0

Patron Type: ARLIR staff

Patron Email: margaret.richardson@anu.edu.au

Preference: Electronic

Courtesy of Library Services

For further assistance, please contact your library. NOTICE: This material may be protected by copyright law
(Title 17 U.S. Code).

The impact of inhomogenous background errors on a global wave data assimilation system

DIANA J. M. GREENSLADE†* and IAN R. YOUNG‡

†Bureau of Meteorology Research Centre, GPO Box 1289K, Bureau of Meteorology,
Melbourne, Victoria, 3001, Australia

‡Swinburne University of Technology, PO Box 218 Hawthorn, Melbourne,
Victoria, 3122, Australia

(Received in final form 16 February 2005)

One of the main limitations in current wave data assimilation systems is the lack of an accurate representation of the structure of the background errors. In this work, models for the observational error variance, background error variance and background error correlations are developed based on the results of previous studies. These are tested in a global wave data assimilation system and the resulting wave forecasts are verified against independent observations from buoys. Forecasts of significant wave height show substantial improvement over the Australian Bureau of Meteorology's current operational wave forecasting system. However, forecasts of peak period are not similarly improved. The regional impacts of the new assimilation scheme are found to vary on a seasonal basis. Overall, it is shown that the inclusion of a latitudinally dependent background error, and improved specification of the background and observational error variances can reduce the root-mean-square error of 24-hour forecast Significant Wave Height by almost 10%.

Keywords: Data assimilation; Significant wave height; Background errors

1. Introduction

Several operational Numerical Weather Prediction (NWP) centres run wave forecasting systems that routinely assimilate satellite observations. One of the main limitations in current wave data assimilation systems is the lack of an accurate representation of the structure of the background errors. Typically, the background errors are set to be isotropic and homogenous and the spatial scale is defined on an *ad hoc* basis. Recent work (Greenslade and Young (2004a), Greenslade and Young (2005) (henceforth referred to as GY1 and GY2, respectively)) has attempted to quantify more accurately the structure of the background errors for global wave data assimilation systems. GY1 and GY2 found that the structure of the background error in Significant Wave Height (H_s) varies significantly over the globe. It was also demonstrated that the background errors may be significantly anisotropic in places.

*Corresponding author. Email: dag@bom.gov.au

In this work, new models for the structure of the background errors in a global wave data assimilation system are described, based on the results of GY1 and GY2. These new models are tested in a data assimilation system to determine whether they add any skill to the system. The wave forecasts are compared to forecasts from the data assimilation system that is currently implemented operationally at the Australian Bureau of Meteorology (the Bureau) and evaluated against observations of H_s and Peak Period (T_p) from wave buoys.

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. Summaries of the methods used in and the results of GY1 and GY2 are given in this section. The wave model used in this work is a version of the WAM model (WAMDI Group 1988, Komen *et al.* 1994). This is described in section 3. The details of the data assimilation method are presented in section 4 and the specific experiments run are described in section 5. Results are presented and discussed in section 6 and finally section 7 contains a summary and outlook.

2. Background

Current operational wave data assimilation systems at, for example, the European Centre for Medium-Range Weather Forecasts (ECMWF), the Bureau and Meteo-France use the sequential method of Statistical Interpolation (SI) to combine first-guess wave model fields with the observations to obtain analysed wave fields. Lionello *et al.* (1992) were the first authors to document SI as a method to construct analysed H_s fields, although it had been in use in NWP previously (Lorenc 1981). The SI algorithm is described here.

At each model gridpoint, i , the analysed (corrected) H_s , H_{an}^i , is expressed as a linear combination of the first-guess predictions from the model, H_p^i , and the observations, H_{obs}^i ,

$$H_{an}^i = H_p^i + \sigma_p^i \sum_{j=1}^{N_{obs}} W_{ij} \left[\frac{H_{obs}^j - H_p^j}{\sigma_p^j} \right] \quad (1)$$

N_{obs} is the number of observations and σ_p is the standard deviation of the model prediction error, i.e.,

$$\sigma_p^i = \sqrt{\langle (H_p^i - T^i)^2 \rangle} \quad (2)$$

where T is the true H_s and $\langle \dots \rangle$ is the expected value. Note that H_p and H_{obs} are assumed to be unbiased, even though this is often not the case in practice (see section 6.1). The weights (W_{ij}) are chosen to minimise the root-mean-square (rms) error of the analysis. Assuming that the errors in the model are not correlated with the measurement errors, the weights are therefore given by

$$W_{ij} = \sum_{k=1}^{N_{obs}} P_{ik} M_{kj}^{-1} \quad (3)$$

where the matrix M is

$$M_{kj} = P_{kj} + O_{kj} \quad (4)$$

P and O are called the error correlation matrices of the model prediction (or background) and observations respectively and must be specified.

One of the major limitations in the application of SI techniques is the specification of the model error correlation matrix, P . This is a symmetric N_{obs} by N_{obs} matrix whose element (k, j) is given by

$$P_{kj} = \frac{\langle (H_p^k - T^k)(H_p^j - T^j) \rangle}{\sigma_p^k \sigma_p^j} \quad (5)$$

In other words, the value of element (k, j) of matrix P is the correlation between the model error at observation location k and the model error at observation location j .

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:

$$P_{kj} = \left(1 + \frac{|x_k - x_j|}{L}\right)^a \exp\left[-c\left(\frac{|x_k - x_j|}{L}\right)^b\right] \quad (6)$$

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 figure 1.

Table 1. Values of the parameters in equation (6) used by various authors.

Author	a	b	c	L
Lionello <i>et al.</i> (1992)	0	1	1	15°
Mastenbroek <i>et al.</i> (1994)	1	1	1	~135 km
Breivik and Reistad (1994)	0	2	1/2	200–40 km
Young and Glowacki (1996)	0	2	1/2	350 km
Greenslade (2001)	0	2	1/2	300 km
Voorrips <i>et al.</i> (1997)	0	3/2	1	200 km

Figure 1. Some example functions used for the background error correlations at mid-latitudes.

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. Most of the studies have assumed isotropy and homogeneity in the background error structure. (Note however, that Lionello *et al.* (1992) defined the length scale of the background error correlations as a function of degrees of latitude and longitude, rather than kilometres. This means that the model errors are defined to have a larger zonal spatial scale at low latitudes than at high latitudes, resulting in both anisotropy and inhomogeneity.) An important limitation to current implementations of SI systems and indeed potential future systems is the lack of an accurate representation of the model or background errors.

The recent studies of GY1 and GY2 have examined the structure of P_{kj} more closely. GY1 applied the observational method of Hollingsworth and Lönnberg (1986) and considered the difference between bias-corrected modelled H_s fields and bias-corrected ERS-2 altimeter observations of H_s over a four-year period. It was found that the spatial scale of the background error varied considerably over the globe, with the largest scales at low latitudes and the shortest scales at high latitudes. This distribution of background error spatial scales is similar to the expected distribution of spatial scales of the signal. Specifically, at low latitudes, the sea-state is typically dominated by large-scale swell systems, while at high latitudes the sea-state is dominated by smaller scale wind-sea systems. GY1 detected very little seasonal or year-to-year variability in the spatial scales. The following second-order autoregressive function was proposed for the background error correlation matrix.

$$P_{kj} = \left(1 + \frac{|x_k - x_j|}{L}\right) \exp\left[-\frac{|x_k - x_j|}{L}\right] \quad (7)$$

with L varying as shown figure 2.

In GY2, the ‘NMC method’ (Parrish and Derber 1992, Rabier *et al.* 1998) was used to examine the structure and magnitude of background errors in an operational wave forecasting system. The ‘NMC method’ uses the forecast divergence as a proxy for the short-range forecast error. The forecast divergence can be estimated by examining the differences between pairs of model forecasts of different ranges

Figure 2. Proposed values for L (km) in equation (7) (reproduced from GY1).

valid at the same time. For example $t48-t24$ represents the difference between 48-hour and 24-hour forecasts of H_s . Using the structure given in equation (7), the average length scale (L) of the $t48-t24$ forecast divergence of H_s over one year found in GY2 is shown in figure 3. It can be seen that the isotropic correlation length scale of the $t48-t24$ forecast divergence has a similar structure to that found in GY2 but the scales are generally shorter. In GY2, limitations to both methods were discussed and it was suggested that the results from GY2 could be used as a lower bound and the results of GY1 as an upper bound to estimates of the spatial scale of the background error in global wave models.

In GY1, it was not possible to use altimeter data to determine whether there is any anisotropy in the background errors because of the limited directional sampling of the satellite altimeter. However, the ‘NMC method’ uses model data from a regular grid and this permits the examination of correlations in all directions. The anisotropic results from GY2 are shown in figure 4. Each ellipse in this figure

Figure 3. Average isotropic correlation length scale of the $t48-t24$ forecast divergence of H_s for the year April 2000 to March 2001 (reproduced from GY2).

Figure 4. 0.5-level contour of the correlation of the $t48-t24$ forecast divergence of H_s for the year April 2000 to March 2001 (reproduced from GY2).

represents the 0.5-level contour of the anisotropic correlation of the t48–t24 forecast divergence of H_s for the year April 2000 to March 2001. The relative size of the ellipses generally follows the pattern of the isotropic length scale shown in figure 3. The ellipses are mostly 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.

3. Wave model

The wave model used is AUSWAM, a version of the WAM model (WAMDI Group 1988, Komen *et al.* 1994). The WAM model is a third generation wave model which solves the wave transport equation explicitly without assuming a form for the evolving spectrum. The wave transport equation is:

$$\frac{\partial F}{\partial t} + \nabla \cdot (c_g F) = S_{in} + S_{nl} + S_d \quad (8)$$

where $F(f, \theta)$ is the wave spectrum as a function of frequency and direction, 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 non-linear energy transfer between groups of resonant waves and S_d the dissipation of energy due to whitecapping.

In the most recent edition 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 wave buoy data over a one-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, and compensating for the reduced numerical diffusion by 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 discretised into 12 directional bins, centred 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 co-ordinate system (Bidlot *et al.* 1997). There are 25 frequency bins ranging from 0.0418 to 0.4114 Hz. Deep water physics only are used. The propagation and source term time steps are 20 min and 10 min respectively. For the global version of the model, the north–south extent of the domain is 78°N to 78°S . The ice-edge is not explicitly included in the wave model. 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 (Hess *et al.* 1995).

For this work, AUSWAM is implemented on a global grid at 1° resolution and 48-hour forecasts (forced by forecast winds from GASP) are made every 12 h over a period of a month. Each model run begins with a 12-hour assimilation period (forced by analysed winds from GASP) starting from the previous wave analysis,

and output fields of integrated wave parameters (i.e., H_s, T_p) etc. are archived at 3-hourly intervals.

Restart files were created at 00Z on May 1 2001 and the model was spun-up without data assimilation for one month. The data assimilation experiments were then run for the month of June 2001. This month was chosen because the altimeter data (used for assimilation) and buoy data (used for verification) available during this month were of a high quality.

4. Data assimilation

During each 12-hour assimilation period, H_s observations from the ERS-2 altimeter are assimilated every 3 h. Ideally, there would be observations available globally within each 3-hour period. However, the altimeter observes only a small portion of the ocean surface within 3 h. So expectations for the impact of any data assimilation scheme must be limited, particularly when only one altimeter is considered. Fast-delivery ERS-2 data (H_{FD}) are bias-corrected as follows (Greenslade and Young 2004b):

$$\begin{aligned} H_{adj} &= 0.1 && \text{for } H_{FD} \leq 0.72 \\ H_{adj} &= 2.187 H_{FD} - 1.582 && \text{for } 0.72 < H_{FD} \leq 1.375 \\ H_{adj} &= 1.153 H_{FD} - 0.160 && \text{for } H_{FD} > 1.375 \end{aligned} \quad (9)$$

before assimilation. In addition, the altimeter data undergo extensive quality control. This involves two stages. Firstly, there is an initial check for error against the background field, followed by a cross-validation check for consistency with nearby data (Young and Glowacki 1996).

Observational errors from the altimeter are generally assumed to be spatially uncorrelated and the definition of the matrix O in equation (4) usually reflects this by being set to be diagonal. However, there is some evidence (Janssen *et al.* 2003) to suggest that the processing of the individual ERS-2 altimeter observations introduces some error correlation on scales of approximately 70 km. This should not affect the experiments in this work because of the way in which the altimeter data is thinned. Specifically, for each assimilation step, the analysis is run N times, with $1/N$ of the altimeter observations used each time (every N th observation) and the resulting N analyses are averaged together. N is set to 100 for these experiments and the altimeter observations are approximately 7 km apart, so the impact of any spatial correlations will be eliminated. The main reason for this thinning of the observations is to reduce the computational requirements for the data assimilation, but a useful side-effect is that it eliminates the need to consider the potential correlations in the observations. Note that a consequence of such thinning is that the analysis increments will have a smaller amplitude than if all the data were used. An alternative to this data-thinning is the creation of ‘super-observations’ with similar spatial scale as the model grid by averaging together a number of the individual altimeter observations.

During each data assimilation step, the first-guess, or background field from the wave model is combined with the ERS-2 observations to produce the analysed H_s field. This is done via SI and has been described in section 2. The parameters that are adjusted between each experiment in this work are: the magnitude of the

observational error variance (σ_{obs}^2), the magnitude of the background error variance (σ_{p}^2) and the background error correlation matrix (P_{kj}). The specific values used for these parameters for each experiment are described in section 5.

Once the analysed H_{s} has been determined, the next step is to adjust the wave spectrum. H_{s} is proportional to $\sqrt{E}$ where E is the integral of the wave spectral energy over all frequencies and directions; thus, there is no unique method of adjusting the wave spectrum to obtain the analysed H_{s} . AUSWAM uses the ‘swell strategy’, which involves the constraint that the integral wave steepness remains constant. This is achieved by shifting energy to different frequencies as the total energy is adjusted. Specifically, at every model gridpoint where the analysed H_{s} (H_{an}) is different to the first-guess H_{s} (H_{p}), the analysed spectrum, $F_{\text{an}}(f, \theta)$, is obtained from the predicted spectrum, $F_{\text{p}}(f, \theta)$ and the mean frequency of the predicted spectrum, $\bar{f}$ as

$$F_{\text{an}}(f, \theta) = a F_{\text{p}}(f + \Delta f, \theta) \quad (10)$$

where

$$\Delta f = \bar{f} - \left(\frac{\bar{f}}{a^{1/4}} \right) \quad \text{and} \quad a = \left(\frac{H_{\text{an}}}{H_{\text{p}}} \right)^2 \quad (11)$$

Note that since this involves a frequency shift, the frequencies at which energy is defined in the analysed spectrum are likely to be different to the frequencies at which the spectra are defined within the model. So a final step in the assimilation involves interpolating the analysed wave spectra back to the defined frequencies of the wave model. The major advantage of the ‘swell strategy’ is that there is little impact on the dissipation source term in the wave model, which in AUSWAM, is proportional to the square of the wave steepness.

5. Experiments

Models for σ_{obs}^2 , σ_{p}^2 and P_{kj} are described in this section. Specific model runs are listed at the end of the section and in table 2.

5.1. Observational error variance

The observational error describes that part of the observation that is not ‘signal’. In this context, the signal is the true H_{s} . Thus the observational error includes

Table 2. Details of experiments. ϕ is latitude in degrees.

Model run	σ_{obs} (m)	σ_{p} (m)	P_{kj}	L (km)
No_assim	—	—	—	—
Assim_op	0.5	0.5	$\exp \left[-\frac{1}{2} \frac{ \mathbf{x}_k - \mathbf{x}_j ^2}{L^2} \right]$	300
Assim_op_700	0.5	0.5	$\exp \left[-\frac{1}{2} \frac{ \mathbf{x}_k - \mathbf{x}_j ^2}{L^2} \right]$	700
Evar	Equation (12)	Equation (13)	$\exp \left[-\frac{1}{2} \frac{ \mathbf{x}_k - \mathbf{x}_j ^2}{L^2} \right]$	300
L_Lat_Evar	Equation (12)	Equation (13)	Equation (7)	$650 - 5.5 \phi $
L_Exact_Evar	Equation (12)	Equation (13)	Equation (7)	Figure 2
Aniso	Equation (12)	Equation (13)	Equation (17)	Equation (17)

instrument error and the ‘error of representativeness’. The ‘error of representativeness’ is a measure of the error caused by the misrepresentation of the signal at all scales smaller than the sampling interval (Daley 1991). (Note that the ‘error of representativeness’ is often also used in the modelling context to refer to variability that appears in the observations but cannot be resolved by the model grid). Typically for satellite estimates of H_s , σ_{obs}^2 is set to be constant and for the operational Bureau system, $\sigma_{\text{obs}}^2 = 0.25 \text{ m}^2$ ($\sigma_{\text{obs}} = 0.5 \text{ m}$).

Estimates of the observational error variance can be obtained indirectly via the method of Hollingsworth and Lönnberg (1986). In GY1, estimates of the observational error variance were obtained from the value of the correlation between ERS-2 H_s and modelled H_s at a spatial lag of zero (Daley 1991), under the assumption that the observational errors are not spatially correlated. These were average values expected within boxes of side length 20° in latitude and longitude over periods of 3 months. σ_{obs}^2 was found to vary seasonally, so the relevant values of the observational error variance for the data assimilation experiments in this work are those for the Apr–Jun season. This is reproduced from GY1 (as standard deviation) in figure 5(a). It can be seen that σ_{obs} is quite variable over the globe and less

Figure 5. (a) Standard deviation (m) of the error in ERS-2 H_s observations for Apr–Jun (b) Mean observed H_s (m) for Apr–Jun.

Figure 6. Standard deviation of the error in ERS-2 H_s observations for Apr–Jun as a function of mean observed H_s for Apr–Jun. Dashed line represents the current Bureau operational value.

than 0.5 m everywhere. Note that in GY1, the assumption was made that the observational errors are not spatially correlated and as discussed in section 4, this may not be justified. The potential impact of this is that the values shown in figure 5(a) may be underestimated.

Figure 5(a) suggests the possibility of setting σ_{obs} to be dependent on latitude. However, it is more likely to be a function of observed H_s , which in turn is generally strongly correlated with latitude. The distribution of σ_{obs} in figure 5(a) can be compared to the mean observed H_s during Apr–Jun, which is shown in figure 5(b). A similar pattern appears with the lowest values of H_s occurring at low latitudes and highest values at high latitudes in the Southern Hemisphere. This information is shown as a scatter plot in figure 6. Each cross on this figure represents the values within one 20° box. As one might expect, the standard deviation of error in observed H_s increases for higher H_s . However, there is also an increase in the error at lower H_s , with the minimum error occurring at approximately 2 m. This is a reflection of the problems that altimeters have observing low H_s . A simple two-branched linear model was applied to these data and the best-fit was found to be:

$$\begin{aligned} \sigma_{\text{obs}} &= 0.54 - 0.14 H_s & \text{for } H_s \leq 2.0 \text{ m} \\ \sigma_{\text{obs}} &= 0.15 + 0.06 H_s & \text{for } H_s > 2.0 \text{ m} \end{aligned} \quad (12)$$

These fits are shown by the solid lines in figure 6. The slight discontinuity that this introduces for $H_s = 2$ m should not result in a significant impact. This analysis was also performed for the Jul–Sep time period and the resulting best-fit lines were very similar. This suggests that the relationship in equation (12) is time-independent. The current Bureau operational value is also shown on figure 6. It can be seen that the current operational value is generally too high.

5.2. Background error variance

In GY1, estimates for the background error variance were determined concurrently with the estimates for observational error variance. The results for σ_p are shown

Figure 7. (a) Standard deviation (m) of the error in modelled H_s for Apr–Jun from GY1. (b) Mean modelled H_s for Apr–Jun.

in figure 7(a). The value of σ_p according to this method ranges from 0.4 m to more than 1 m. As in the previous section, these represent average values over a 20° box and a 3-month time period. Note that the high values of σ_p along the southernmost latitude are likely to be associated with the lack of an explicit ice-edge in the wave model. The mean modelled H_s for the Apr–Jun season is shown in figure 7(b). Similarly to σ_{obs} , a relationship between σ_p and H_s is suggested. The values from figure 7 are presented as a scatter plot in figure 8. The scatter plot this time suggests a constant value of σ_p for low modelled H_s , and a linearly increasing function for higher values of modelled H_s . Setting the cut-off value to be 2 m, the best-fit line was found to be:

$$\begin{aligned} \sigma_p &= 0.5 & \text{for } H_s \leq 2.0 \text{ m} \\ \sigma_p &= 0.03 + 0.23H_s & \text{for } H_s > 2.0 \text{ m} \end{aligned} \quad (13)$$

Again, an analysis of σ_p and H_s for a separate season showed that this relationship is time-independent.

The current Bureau operational value ($\sigma_p = 0.5$ m) is indicated with a dashed line in figure 8. It can be seen that this value is appropriate for low values

Figure 8. Standard deviation of the error in background H_s for Apr–Jun as a function of mean modelled H_s for Apr–Jun. Dashed line represents the current operational value.

of modelled H_s , but is clearly too low for high H_s . Daley (1991) shows that if σ_p is considerably underestimated, then the analysis would be more accurate if the background were completely ignored. This suggests that by simply defining more accurate error variances within the assimilation system, a positive impact on the analysed H_s and therefore the forecast H_s could be obtained. This possibility is examined in section 6.

5.3. Background error correlations

5.3.1. Isotropic. As described earlier, the background error correlations were examined in both GY1 and GY2. Isotropic correlation length scales from these two papers were shown in figures 2 and 3. It was suggested in GY2 that the results shown in figure 2 represent an upper bound to the length scales, and those of figure 3 represent a lower bound. Seaman (1983) shows that, in theory, the expected analysis error is greater when L is underestimated than when it is overestimated by the same amount. In other words, it is preferable to define the spatial scale of the background error correlations to be too large rather than too small (within reasonable bounds). Therefore, the ‘upper bound’ results of GY1 are used here. Figure 2 suggests that the length scales are strongly dependent on latitude.

These values are therefore plotted as a function of latitude in figure 9. No distinction is made between northern and southern hemispheres here. A simple linear model was applied to this data and the best fit was found to be:

$$L(\phi) = 647.8 - 5.62|\phi| \quad (14)$$

where ϕ is latitude in degrees and L is in km. This can be approximated as

$$L(\phi) = 650 - 5.5|\phi| \quad (15)$$

and this approximate value is shown in figure 10. Comparing this with figure 2, it can be seen that this is a reasonable approximation. The most obvious differences occur around the central Pacific where the approximate length scales are somewhat shorter than the actual length scales.

The use of an inhomogeneous background error correlation function introduces some complications into the SI procedure. The prediction error correlation matrix,

Figure 9. Background error correlation length scale as a function of latitude. Each cross represents the value of L for one 20° box.

Figure 10. Length scale of the background error correlations as found in GY1 approximated as a function of latitude.

P_{kj} , represents the correlation between the model error at location k and the model error at location j , where j and k represent either a model gridpoint or an observation location. However, the defined length scale of the model error correlation may have different values at the two locations, so the question arises: which length scale should be used? This contradiction occurs because some basic statistical assumptions are being contravened. In order to be able to calculate the autocorrelation function of a process, the process must be stationary up to order 2 (Priestley 1981). In the case of spatial autocorrelations, this means that the mean and variance of the process must be independent of location, and the autocovariance (and therefore also the autocorrelation) should depend only on the distance between two points, and not

on their location. For the case of global H_s , these assumptions clearly do not hold. Indeed, the very concept of a correlation function that varies in space is ill-defined. However, stationarity (in the spatial sense, i.e., homogeneity) is unlikely to exist in the global situation, so for practical purposes these statistical limitations must simply be kept in mind. Standard practice in NWP and ocean data assimilation systems is to use the arithmetic or geometric mean of the two length scales. Here, the geometric mean is used, i.e.,

$$L_{kj} = \sqrt{L_k L_j} \quad (16)$$

where L_k is the length scale at point k and L_j is the length scale at point j .

5.3.2. Anisotropic. It is reasonable to expect that the background errors in a global wave model would be anisotropic. This is because the H_s signal itself is anisotropic (due to, for example, swell systems dispersing parallel to their direction of propagation) and the background errors may reflect this. If an anisotropic P_{kj} were incorporated into the data assimilation system, it could provide a more physically based adjustment to the first-guess modelled H_s field.

In GY1, it was found that the altimeter data did not provide information in enough directions to be able to use them to determine the anisotropy in the background errors. In GY2, however, anisotropy in the forecast divergences could be examined. The assumption is made here that the structure of the background errors is the same as the structure of the t48–t24 forecast divergence and the anisotropic results from GY2 are used. Although one would expect to see a seasonal signal in the anisotropy (e.g., more swell from the winter hemisphere, as discussed in GY2) only the annually averaged values shown in figure 4 are used here. The analytic correlations represented by the ellipses in this figure are implemented as:

$$P_{kj} = \exp \left[-\frac{d}{A_{kj}} \right] \quad (17)$$

with

$$d^2 = |x_k - x_j|^2 \left(\frac{1}{B_{kj}^2} \cos^2(\theta_{kj} - C_{kj}) + B_{kj}^2 \sin^2(\theta_{kj} - C_{kj}) \right) \quad (18)$$

where $|x_k - x_j|$ is the great circle distance between points j and k and θ_{kj} is the angle from North of the line joining them. A_{kj} , B_{kj} , and C_{kj} are the three parameters which define the ellipses in figure 4 (i.e., length scale, eccentricity and tilt, respectively). These vary from 20° box to 20° box. As for the isotropic background error correlations in the previous section, complications arise when using a correlation function that varies geographically. In this case, each parameter (A_{kj} , B_{kj} , and C_{kj}) was treated in the same way as in the previous section, i.e., by using the geometric mean of the parameters at the two locations, k and j .

5.4. Model runs

The following seven experiments were run. The details of each experiment are summarised in table 2.

No_assim Assimilation of ERS-2 data is not performed.

Assim_op Assimilation is performed as in the current operational AUSWAM. In this case, $\sigma_{\text{obs}} = 0.5 \text{ m}$, $\sigma_p = 0.5 \text{ m}$ and P_{kj} is given by:

$$P_{kj} = \exp \left[-\frac{1}{2} \frac{|x_k - x_j|^2}{L^2} \right] \quad (19)$$

with $L = 300 \text{ km}$.

Evar It is useful to check whether any improvement arises from simply defining the error variances more accurately, so for this run the background error structure is as for **Assim_op**, and the assumed values for the observational and background error variances are given by equations (12) and (13).

L_Lat_Evar For this run, the values for the background and observational error variances are set to be the same as in **Evar**, i.e., as a function of modelled and observed H_s respectively. In addition, the background error correlation matrix is defined as proposed by GY1, with the length scales as described by equation (15).

L_Exact_Evar This run examines the sensitivity of the latitudinal approximation to figure 2. σ_{obs} , σ_p and P_{kj} are exactly as in the previous run, but instead of L given by equation (15), the actual values within each 20° box shown in figure 2 are used.

Assim_op_700 This run is exactly the same as **Assim_op** except $L = 700 \text{ km}$ is used. This is done in order to ensure that any improvement seen in the model skill is not simply due to larger background error correlation length scales, but is due to the geographic variability of the length scales.

Aniso For this run, the error variances are set as in **Evar**, i.e., as a function of observed and modelled H_s . The background error correlations are set to be anisotropic and are given by equations (17) and (18).

6. Results

Figure 11 shows an example of the first-guess fields in the Pacific Ocean at the same time level (09Z on June 21st 2001) for each of the assimilation experiments described above. The location of the (thinned) ERS-2 observations for this assimilation time step (i.e., observations occurring between 7:30Z and 10:30Z) are also shown in this figure. Before considering the performance of each assimilation experiment, there are a few points of interest in these first-guess fields. The assimilation schemes have been running for approximately 20 days at this point, so even though the same wind forcing fields have been used, there are some differences evident in the first-guess fields. Note, in particular, the area of high ($> 5 \text{ m}$) H_s to the south of Australia. The maximum H_s in this area varies widely between each run – it just reaches 8 m for the operational system (**Assim_op**) while it is more than 10 m for **L_Lat_Evar** and **L_Exact_Evar**. On the other hand, the other region of high H_s (in the southeastern corner of the domain) appears similar (in shape and magnitude) for all the experiments. This implies that either the model has forecast this wave system correctly, or there have been no recent altimeter observations of it.

It is also interesting to consider the region of high H_s in the northern Pacific as the ERS-2 ground tracks pass close to this system. Figure 12 shows the

Figure 11. First-guess fields of H_s (m) over the Pacific Ocean for each assimilation experiment valid at 09Z on June 21st, 2001. Contours are at intervals of 1 m. The crosses mark every 1/100th ERS-2 observation during this time period.

Figure 12. H_s observations from the ERS-2 altimeter along the westernmost ground track indicated in figure 11.

bias-corrected H_s observations from the satellite along the westernmost ground track indicated in figure 11. It can be seen that the observed H_s is > 4 m at 50°N . None of the first-guess fields have forecast $H_s > 4$ m at the location of the ground track, so it appears that either the model has forecast the location of the peak H_s incorrectly, or it has underestimated the spatial extent of the area of high H_s . The structure of the first-guess H_s field is similar in all the experiments for this region, so this is a useful example that can be examined more closely to provide an indication of how the definition of the background errors has an impact on the analysed H_s field.

Figure 13 shows the H_s increment field for this assimilation step for each experiment. The increment field is the value of H_s at each gridpoint that is added to or subtracted from the first-guess H_s to give the analysed wave field and is given by the second term on the right-hand side of equation (1).

There are some very clear differences evident between the various assimilation experiments. Comparing panels (a) and (e), it can be seen that the effect of increasing L , the length scale of the background error correlations, is to spread the information from the altimeter over a larger area. The magnitude of the increment, however, does not change substantially. A comparison of panels (a) and (b) demonstrates the effect of changing the values of σ_p and σ_{obs} . For the modelled wave system in the northern Pacific, σ_p will be larger for the **Evar** case than for **Assim_op** (see figure 8). σ_{obs} , on the other hand, will be smaller for the **Evar** case than for **Assim_op** (see figure 6). It can be seen that this results in a larger magnitude of the increment. Also, note that for **Evar**, the position of the maximum increment is not at the observation location, but at the location of the maximum modelled H_s . This is a reflection of the higher uncertainty associated with the modelled H_s . In areas of high H_s (and therefore also high σ_p), the analysis is permitted to deviate more widely from the first-guess field. A similar pattern can be seen for all the cases where σ_p and σ_{obs} are defined as a function of H_s (b, c, d and f) as opposed to the cases where they are defined to be a constant value (a and e).

The increment for **L_Lat_Evar** in panel (c) demonstrates the effect of implementing background error correlations that vary with latitude. Note that near the equator, the spatial extent of the increment is larger than **Assim_700**. On the other hand, in the high latitudes, the spatial scale of the increment is generally smaller for **L_Lat_Evar** than for **Assim_op_700**. Little difference can be seen between (c) and (d) except the edges of the increment appear slightly less smooth for (d). This arises because the length scale of the background error correlations in **L_Exact_Evar**

Figure 13. Increment fields (m) over the Pacific Ocean for each assimilation experiment valid at 09Z on June 21st, 2001. The crosses mark every 1/100th ERS-2 observation during this time period. Contours are at intervals of 0.25 m.

can vary abruptly from 20° box to 20° box. The same applies to **Aniso**. This effect is undesirable because it will introduce noise into the model.

6.1. Verifications

The wave forecasts from each experiment are verified against observations of H_s and T_p from a series of wave buoys. Most of these are National Data Buoy Center (NDBC) buoys with some additional buoys situated around the Australian coast. Details of the buoys used for verification are listed in table 3 and their locations are shown in figure 14. Figure 15 shows time series of observed H_s and T_p at buoy 46002 in the northeastern Pacific along with 24-hour forecasts from three of the experiments: **No_assim**, **Assim_op** and **L_Lat_Evar**. Verification statistics for these 24-hour forecasts at this location are shown in table 4. The bias, rms, Scatter Index (si) and linear correlation coefficient (R) are defined as:

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N H_i - O_i \quad (20)$$

Table 3. Details of buoys used in this work.

Region	WMO Code	Location	Owner	Lat.	Lon.	Jun '01	Jan '03
North-West Atlantic	41001	Hatteras	NDBC	34.8N	72.8W	X	
	41002	S. Hatteras	NDBC	32.2N	75.4W		X
	41004	Edisto	NDBC	32.5N	79.1W		X
	41010	Canaveral E.	NDBC	28.9N	78.5W	X	X
	44004	Hotel	NDBC	38.4N	70.6W	X	X
	44005	Gulf of Maine	NDBC	43.2N	69.2W	X	
	44008	Nantucket	NDBC	40.5N	69.4W	X	X
	44011	Georges Bank	NDBC	41.1N	66.6W	X	X
Gulf of Mexico	44014	Virginia Beach	NDBC	36.6N	74.8W	X	X
	42001	Mid Gulf	NDBC	25.9N	89.7W	X	X
	42002	W. Gulf	NDBC	25.9N	93.6W		X
	42003	E. Gulf	NDBC	25.9N	85.9W	X	
	42036	W. Tampa	NDBC	28.5N	84.5W	X	X
	42039	Pensacola	NDBC	28.8N	86.0W	X	X
	46001	Gulf of Alaska	NDBC	56.3N	148.2W	X	X
	46002	Oregon	NDBC	42.5N	130.3W	X	X
East Pacific	46005	Washington	NDBC	46.1N	131.0W	X	X
	46006	S.E. Papa	NDBC	40.9N	137.5W	X	X
	46035	Bering Sea	NDBC	56.9N	177.8W		X
	46047	Tanner Banks	NDBC	32.4N	119.5W	X	X
	46059	California	NDBC	38.0N	130.0W	X	X
	46066	S.Aleutians	NDBC	52.6N	155.0W	X	X
	51001	N.W. Hawaii	NDBC	23.4N	162.3W	X	
	51002	S.W. Hawaii	NDBC	17.2N	157.8W	X	X
Central Pacific	51003	W. Hawaii	NDBC	19.2N	160.8W	X	
	51004	S.E. Hawaii	NDBC	17.4N	152.5W	X	X
	51028	Christmas Is.	NDBC	0.0	153.9W	X	
Australia	55020	Eden	MHL	37.3S	150.2E	X	
	55026	Strahan	Bureau	42.1S	145.0E	X	X
	56002	N. Rankin	Woodside CSIR	19.6S	116.1E	X	X

Figure 14. Location of buoys used in this study (a) Northern hemisphere and (b) Southern hemisphere.

$$\text{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N (H_i - O_i)^2} \quad (21)$$

$$\text{si} = \frac{\sqrt{(1/N) \sum_{i=1}^N ((H_i - \bar{H}) - (O_i - \bar{O}))^2}}{\bar{O}} \quad (22)$$

$$R = \frac{\sum_{i=1}^N (H_i - \bar{H})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (H_i - \bar{H})^2 (O_i - \bar{O})^2}} \quad (23)$$

where H_i is the modelled H_s , O_i is observed H_s , N is the number of observations and an overbar represents the mean value.

Consider first the time series of H_s . For the run with no data assimilation, it can be seen that there is a significant amount of negative bias in the modelled H_s . This bias is likely to be due to bias in the wind forcing fields (Kepert *et al.* 2005). The two assimilation runs also start off with this negative bias since data assimilation starts only on June 1. After a few days, the impact of the altimeter data can be seen as the forecast H_s starts to match the observed H_s a little better. The aim of a data

Figure 15. Observed H_s and T_p (thick solid line) and 24-hour forecasts from **No_assim** (thin solid line), **Assim_op** (dashed) and **L_Lat_Evar** (dotted) during June 2001 at the buoy: 46002.

Table 4. Verification statistics for the time series shown in figure 15.

	No_assim	Assim_op	L_Lat_Evar
(a) H_s			
Bias	-0.61	-0.27	-0.15
R	0.79	0.79	0.87
si	0.26	0.27	0.23
rms	0.74	0.50	0.40
(b) T_p			
Bias	-2.6	-0.42	1.11
R	0.16	0.0	0.15
si	0.29	0.43	0.56
rms	3.7	3.9	5.2

assimilation scheme is the correction of variable errors, rather than the elimination of any systematic bias, although this can be a useful side-effect. If the bias is known, it should be reduced by some other means, for example by tuning the dissipation source term (see equation (8)) or by adjusting the wind forcing. It can be seen both in figure 15 and in the verification statistics in table 4 that **L_Lat_Evar** corrects the bias in H_s to a greater extent than the **Assim_op** case. Thus it appears that one effect of including more accurate error variances and inhomogenous background error correlations is simply to increase the H_s . However, note that the si is also reduced in **L_Lat_Evar**. This is encouraging, as it indicates that these results do contribute to adding skill to the wave model.

Now consider the results for T_p . The observed T_p ranges from 3.9 s to 16.7 s with some rapid changes between high and low values. These large abrupt changes can occur when both swell and wind-sea are present and the peak energy in each system is approximately equal. At around June 19–21, there is a prolonged period of high T_p corresponding to a period of low observed H_s . It is likely that there were low winds during this time period, and therefore a lack of wind-sea energy, so that the swell was dominant. This causes the low peak frequency, or high T_p seen around this time.

Due to the sensitivity of T_p to the presence of multiple wave systems, it is difficult for numerical wave models to capture its variability. However, some general features can be noted, and the overall verification statistics can give a general indication of the impact of the data assimilation. Note that at this location, the **No_assim** T_p is almost always too low – this is reflected in the verification statistics, where there is a bias of -2.6 s in T_p . This indicates that there is either too much wind-sea energy or not enough swell in the model. Since the H_s is also generally too low, lack of low-frequency swell is the likely candidate. This could be caused by excessive dissipation for low frequency wave components in AUSWAM (Alves *et al.* 2002) or lack of growth in the wind-sea spectrum.

At some times, the assimilation schemes improve the prediction of T_p (e.g., days 7–9 and days 22–25). However, on numerous occasions, the assimilation causes many rapid switches in T_p which do not appear in the observations and this degrades the verification statistics, particularly around days 10 to 15. This may be due to the sensitivity of T_p to multiple wave systems, however, the fact that the observed T_p is fairly stable at this time indicates that this is probably not the case. A more likely explanation is that the ‘swell strategy’ used in all these experiments has shifted energy to the low frequencies unnecessarily. This could be remedied by altering the method by which the assimilation adjusts the wave spectrum. This possibility is not examined here, but is an area for which further investigation is required.

In summary, it can be seen that at this location the rms error in H_s is improved substantially by **L_Lat_Evar** over the Bureau’s current operational assimilation scheme. While this is mainly due to a reduction in the systematic bias, there is also reduction in the variable errors. The verifications of T_p do not, however, show similar improvement.

Verification statistics for H_s averaged over all the buoys listed in table 3 are shown graphically in figure 16. Results for all of the experiments described in section 5.4 are shown here, with the exception of **L_Exact_Evar** because it was found to be almost identical to the results of **L_Lat_Evar**. The results in figure 16 generally reflect the features seen in the time series at buoy 46002. The thick black lines represent the **No_assim** run. It can be seen that without data assimilation, there is a negative bias of approximately 0.4 m in H_s . This bias actually decreases slightly as the forecast period increases. This may simply be due to chance, although a possible explanation is that the surface winds become less negatively biased as the forecast period becomes longer. Overall, the rms error of the model run without data assimilation is approximately 0.6 m.

The operational assimilation scheme accounts for a large portion of this bias – the bias in this case is less than 0.2 m for all forecast periods. This bias reduction is reflected in the significant reduction in the rms error over the **No_assim** run. However, note again that the si is also reduced in the operational assimilation scheme indicating that the assimilation does more than just correct for bias. As the forecast period increases, the bias tends to become more negative, indicating that

Figure 16. H_s verification statistics averaged over 26 buoys during June 2001. The mean observed H_s during this time was 1.4m (a) Bias (b) si and (c) rms error.

the model will relax back to the **No_assim** case without the inclusion of altimeter data. The **Evar** case is a slight improvement over the operational assimilation – the bias and si are further reduced, resulting in an rms error that is smaller than that of the operational case.

Now consider the other three cases shown in figure 16. The bias for **L_Lat_Evar** has a smaller magnitude than that of **Assim_op_700** and **Aniso**, for all forecast periods and all three runs have similar si values, with **Aniso** and **Assim_op_700** being slightly better at short forecast periods. Overall the rms values for these three runs are very close with **Aniso** slightly better at short forecast periods and **L_Lat_Evar** slightly better at longer forecast periods. These differences however, may not be statistically significant.

It was seen in figure 13 that for **L_Lat_Evar**, the information from the altimeter is spread over a much greater area than for **Assim_op**. The possibility exists that the improvement seen in figure 16 could be obtained by simply increasing the spatial scale of the background errors everywhere. It is for this reason that the **Assim_op_700** case was run. The results from these model runs have not conclusively shown that incorporating inhomogeneity into the scale of the background errors contributes skill to the model. However, this would be very difficult to demonstrate with the limited locations of the verifying observations, i.e., the buoys tend to be in coastal areas. Overall, these three cases have reduced the rms error in the 24-hour forecast H_s by almost 10% compared to the Bureau's operational data assimilation system.

Verification statistics for T_p for all model runs (except **L_Exact_Evar**) averaged over all buoys are shown in figure 17. It can be seen that the T_p for the case of no data assimilation is biased low. The operational system corrects this bias, but it is still slightly low and all the other assimilation runs have resulted in a positive bias in T_p . This suggests that these assimilation schemes are overcorrecting towards the low wave frequencies. As the forecast period increases, the assimilation cases gradually tend back towards the **No_assim** case. So as the model propagates forward without the constraint of the altimeter data and the assimilation scheme, it relaxes back to a state with not enough low-frequency swell. Note that **Aniso**, **L_Lat_Evar** and **Assim_op_700** are relaxing back quite slowly and at 48 h there is still a slight positive bias. The si indicates that the **No_assim** case actually performs best, with **Assim_op** performing better than the other assimilation schemes. As was seen with the examination of the time series at buoy 46002, the assimilation schemes can induce rapid changes in the T_p . So even though the overall bias in T_p is reduced, suggesting that the amount of swell in the model has been corrected, it is not necessarily included in the model at exactly the right times.

The rms errors show that the operational assimilation scheme seems to perform best in terms of predicting T_p . **L_Lat_Evar** and **Aniso** are in fact the worst schemes overall. However, this is mainly due to the variable errors, so little weight should be given to these results. Note that the quality of the impact of the data assimilation scheme in terms of the T_p bias is not consistent with the quality of the impact in terms of the SWH bias. On average, the assimilation schemes introduce a shift in the analysed peak periods, and this shift increases as the decorrelation length scale increases (from **Assim_op** to **Assim_op_700** or **L_Lat_Evar**, for example). This correction seems too high though, when compared to the impact on the SWH. For **L_Lat_Evar**, for example, the SWH bias is almost zero (figure 16) while for the analysed T_p , the bias is the largest. This apparent inconsistency may be due to some problems in the assimilation scheme (for example, the assumptions made in distributing the energy within the spectra). This negative impact is unfortunately not likely to be alleviated with the use of additional altimeters (Skandrani *et al.* 2005).

Figure 17. Same as figure 16 but for T_p . The mean observed T_p during this time period is 8.3 s. (a) Bias (b) si and (c) rms error.

In terms of forecasting H_s , the **Aniso** scheme works well and is an improvement over the current operational assimilation scheme. However, it should be borne in mind that many approximations were made for this **Aniso** case. For example, the background error was assumed to have the same structure as the t48–t24 forecast divergence and this is unlikely to be the case. This was shown in GY2, where it was

demonstrated that the isotropic correlation scale obtained via the ‘NMC method’ is likely to be too short. It is quite likely then, that the spatial scales of the anisotropic structures used in this work are also too short. In addition, the annual averaged value of the anisotropic correlations has been used even though there is strong evidence to suggest that there is a seasonal signal in the anisotropic background error structure (GY2). Despite these limitations, the **Aniso** case performs very well. This suggests that there is significant potential for improvement in the use of anisotropic background error correlations.

6.2. Regional impacts

It is interesting to examine the impact of the various data assimilation schemes in different regions of the globe. The buoys were divided into the five regions shown in table 3. Figure 18 shows the rms error of the 24-hour forecast H_s in each region. The information in the ‘All’ column replicates the 24-hour values shown in panel (c) of figure 16.

In the Australian region (and note that this includes only 3 locations), the **Aniso** and **L_Lat_Evar** cases perform slightly better than the other assimilation cases. For the Central Pacific Ocean, the **Aniso** case is marginally better than the other cases. Also note that **Assim_op_700** performs slightly better here than **L_Lat_Evar**. It is at this location that the change in the background error length scales has made the biggest impact. This is to be expected because it is in this region that the biggest difference between the operational length scales and the new length scales exists. The average rms error at buoys in this region is reduced by almost 20%

Figure 18. rms error of 24-hour forecast H_s in different regions of the globe.

by **L_Lat_Evar** compared to the operational assimilation scheme. This is a significant impact, considering that the rms error is already relatively low in this region. The northeast Pacific also shows significant improvement with the rms error for the **L_Lat_Evar** case 16% lower than that of the operational assimilation case.

Very little impact can be seen in the Gulf of Mexico. All of the assimilation schemes perform similarly here and show only minor improvement over the case with no data assimilation. The Gulf of Mexico is almost an enclosed basin, and very little long-term swell penetrates into this area. Therefore most of the wave systems observed by the buoys would be locally generated and there is little opportunity for the altimeter to observe and correct any of these systems. It may also be that the ‘swell strategy’ is not the most appropriate assimilation strategy for this region.

Finally, in the northwest Atlantic a similar pattern to the Gulf of Mexico is seen, with all of the assimilation schemes performing only slightly better than the case with no data assimilation. The case with no data assimilation performs relatively well here. This is likely to be a seasonal effect – this is in the northern hemisphere summer (prior to the hurricane season) so winds and wave heights are generally lower, resulting in lower rms errors.

6.3. January 2003

It is relevant to note that the time period for which these experiments have been run occurs entirely within the four-year period that was examined closely in GY1. This means that the positive impact seen may result from tuning the assimilation system to a particular description of the model errors. It is therefore useful to perform a test from a time period that is independent of the four-year period examined in GY1.

Some of the experiments were therefore re-run for the month of January 2003. As well as being outside the four-year time period, this is during northern hemisphere winter, so it provides the opportunity to examine seasonality. The main aim here, however, is to ensure that **L_Lat_Evar** provides a similar improvement to that which was seen in the previous section, so only **No_assim**, **Assim_op**, **L_Lat_Evar** and **Aniso** were repeated for this second time period.

Note that the details of the data assimilation scheme will not change for this time period. Specifically, the definition of P_{kj} for **L_Lat_Evar** will be identical because it was found that the distribution of the isotropic L does not vary from season to season. It was also shown that the relationships between σ_{obs} and observed H_s , and σ_p and modelled H_s were time independent. And finally, P_{kj} for **Aniso** will not change because an annually averaged value is being used.

Verification statistics for H_s and T_p for this time period are shown in figures 19 and 20, respectively. Note that a slightly different set of buoys is used here (see table 3) because of missing data. It can be seen that the overall rms error for H_s is higher for all runs compared to June 2001. This is because most of the buoys are in the northern hemisphere, and during the northern hemisphere winter, the H_s and therefore the rms error is generally higher. The results tend to follow the same pattern as the results for June 2001, although **L_Lat_Evar** provides a clearer improvement over **Aniso** this time. Note that the rms error of 24-hour forecast H_s is again reduced by 10% with the inclusion of inhomogenous background errors and more accurate error variances.

Regional impacts for this time period are shown in figure 21. There are a few differences between the regional impacts seen here and those for June 2001.

Figure 19. H_s verification statistics averaged over 23 buoys during January 2003. The mean observed H_s during this time period is 2.6m. (a) Bias (b) si and (c) rms error.

The biggest impact from **L_Lat_Evar** is now seen in the Australian region, where the rms error is reduced by 19%. The impacts seen in the Central Pacific, the north-east Pacific and the Gulf of Mexico are similar to those seen during June 2001. In the Northwest Atlantic the overall rms errors are much larger than those

Figure 20. Same as figure 19 but for T_p . The mean observed T_p during this time periods is 9.9s. (a) Bias (b) si and (c) rms error.

for June 2001 – an expected seasonal difference. Also, **L_Lat_Evar** now performs better than the operational assimilation scheme in the northwest Atlantic.

In summary, it can be concluded that the models for P_{kj} , σ_{obs} and σ_p that have been developed in this work are applicable to time periods outside the four-year time period that was examined in GY1 and GY2.

Figure 21. rms error of 24-hour forecast H_s in different regions of the globe for January 2003. The colour of each square corresponds to the legend in figures 19 and 20.

7. Summary

In this work, models for the observational error variance, background error variance and background error correlation for a global wave model have been developed based on the results of previous studies. These models were implemented in a data assimilation system and tested over two month-long time periods. The results were verified against independent observations of H_s and T_p from buoys. It was found that more accurate specification of the observational and background error variances alone could provide some improvement to the forecast skill of the wave model. Also, simply increasing the spatial scale of the background errors resulted in a substantial improvement, but the biggest improvement came from incorporating improved error variances and geographically varying background errors. Using background error correlation length scales that varied from 20° box to 20° box did not show any improvement over using background error correlation length scales that were dependent only on latitude.

Overall, the recommended configurations to include in an SI system such as that described by equations (1–5) are as follows.

Observational error:

$$\begin{aligned}
 \sigma_{\text{obs}} &= 0.54 - 0.14 H_s & \text{for } H_s \leq 2.0 \text{ m} \\
 \sigma_{\text{obs}} &= 0.15 + 0.06 H_s & \text{for } H_s > 2.0 \text{ m}
 \end{aligned} \tag{24}$$

Background error:

$$\begin{aligned}\sigma_p &= 0.5 && \text{for } H_s \leq 2.0 \text{ m} \\ \sigma_p &= 0.03 + 0.23 H_s && \text{for } H_s > 2.0 \text{ m}\end{aligned}\quad (25)$$

Background error correlation matrix:

$$P_{kj} \left(1 + \frac{|x_k - x_j|}{L_{kj}} \right) \exp \left[- \frac{|x_k - x_j|}{L_{kj}} \right] \quad (26)$$

where

$$L_{kj} = \sqrt{L_k L_j} \quad \text{and} \quad L_i = 650 - 5.5\phi_i \quad (27)$$

$|x_k - x_j|$ is the great circle distance between points k and j and ϕ is the absolute value of the latitude.

The regional impacts of the new assimilation scheme were found to vary on a seasonal basis. During June 2001, the biggest impact was seen in the central and northeast Pacific while during January 2003, the biggest impacts were seen in the Australian region.

The rms errors in T_p did not show the same positive impact as the verifications of H_s . This may be due to the sensitivity of T_p to the presence of multiple wave systems or incorrect frequency shifting during the data assimilation. However, the systematic errors in T_p did show marked improvement, and due to the noisiness of T_p , less weight should be given to the variable errors of this parameter.

In summary, the inclusion of a latitudinally dependent background error, and the improved specification of the background and observational error variances has improved the skill of the wave model by almost 10% compared to the Bureau's current wave data assimilation system, when skill is defined as rms error of 24-hour forecast H_s . These improvements were found despite the sparseness of the satellite altimeter data used in the assimilation. There is potential for even more improvement if observations from multiple satellite altimeters are used (Skandrani *et al.* 2005).

The results of this work have shown that the Bureau's operational wave forecasting system could immediately benefit by improving the structure of the background error correlations and the background and observational error variances. This applies specifically to AUSWAM, and the errors resulting from its specific implementation and wind forcing fields. However, it is likely that these results are, to a certain extent, applicable to other operational wave forecasting systems. For example, it is highly likely that the background errors for any wave forecasting system are significantly broader in the tropics than at high latitudes and thus other wave forecasting systems could also benefit from the use of a latitudinally dependent background error correlation structure.

Acknowledgements

The authors would like to thank Jeff Kepert and the anonymous reviewers for comments on the manuscript. Financial assistance in the form of a PhD scholarship for D.J.M. Greenslade was provided by the Australian Bureau of Meteorology.

References

- ALVES, J.H.G.M., GREENSLADE, D.J.M. and BANNER, M.L., 2002, Impact of a saturation dependent dissipation source function on operational hindcasts of wind-waves in the Australian region. *Global Atmosphere and Ocean System*, **8**, pp. 239–267.
- BENDER, L.C., 1996, Modification of the physics and numerics in a third-generation ocean wave model. *Journal of Atmospheric and Oceanic Technology*, **13**, pp. 726–750.
- BIDLOT, J.R., JANSSEN, P.A.E.M., HANSEN, B. and GÜNTHER, H., 1997, A modified set up of the advection scheme in the ECMWF wave model. *ECMWF Technical Memorandum*, Vol. 237 (Reading: ECMWF).
- BREIVIK, L.A. and REISTAD, M., 1994, Assimilation of ERS-1 altimeter wave heights in an operational numerical wave model. *Weather and Forecasting*, **9**, pp. 440–451.
- DALEY, R., 1991, *Atmospheric Data Analysis*, p. 457 (Cambridge, U.K.: Cambridge University Press).
- GREENSLADE, D.J.M., 2001, The assimilation of ERS-2 Significant Wave Height data in the Australian region. *Journal of Marine Systems*, **28**, pp. 141–160.
- GREENSLADE, D.J.M., 2004, The structure of the background errors in a global wave model. PhD thesis, University of Adelaide, Australia.
- GREENSLADE, D.J.M. and YOUNG, I.R., 2004a, Background errors in a global wave model determined from altimeter data. *Journal of Geophysical Research*, **109**, C09S04, doi:10.1029/2003JC002215.
- GREENSLADE, D.J.M. and YOUNG, I.R., 2004b, A validation of ERS-2 fast delivery significant wave height, BMRC Research Report No. 97. *Bur. Met.*, Australia.
- GREENSLADE, D.J.M. and Young, I.R., 2005, Forecast divergences of a global wave model, submitted to *Monthly Weather Review* (In press).
- HESS, G.D., COLMAN, R.A. and MCAVANEY, B.J., 1995, On computing screen temperatures, humidities and anemometer-height winds in large-scale models. *Australian Meteorological Magazine*, **44**, pp. 139–145.
- HOLLINGSWORTH, A. and LÖNNBERG, P., 1986, The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, pp. 111–136.
- JANSSEN, P.A.E.M., 1989, Wave-induced stress and the drag of air flow over sea waves. *Journal of Physical Oceanography*, **19**, pp. 645–754.
- JANSSEN, P.A.E.M., 1991, Quasi-linear theory of wind-wave generation applied to wave forecasting. *Journal of Physical Oceanography*, **21**, pp. 1631–1642.
- JANSSEN, P.A.E.M., ABDALLA, S. and HERBACH, H., 2003, Error estimation of buoy, satellite and model wave height data, *ECMWF Technical Memorandum*, Vol. 402 (Reading: ECMWF).
- KEPERT, J.D., GREENSLADE, D.J.M. and HESS, D., 2005, Assessing and improving the marine surface winds in the Bureau of Meteorology Numerical Weather Prediction Systems, BMRC Research Report No. 105, Rev. Met., Australia.
- KOMEN, G.J., HASSELMANN, S. and HASSELMANN, K., 1984, On the existence of a fully developed wind-sea spectrum. *Journal of Physical Oceanography*, **14**, pp. 1271–1285.
- KOMEN, G.J., CAVALERI, L., DONELAN, M., HASSELMANN, K., HASSELMANN, S. and JANSSEN, P.A.E.M., 1994, *Dynamics and Modelling of Ocean Waves*, p. 532 (Cambridge, UK: Cambridge University Press).
- LIONELLO, P., GÜNTHER, H. and JANSSEN, P.A.E.M., 1992, Assimilation of altimeter data in a global third-generation wave model. *Journal of Geophysical Research*, **97**, pp. 14453–14474.
- LORENC, A.C., 1981, A global three-dimensional multivariate statistical interpolation scheme. *Monthly Weather Review*, **109**, pp. 701–721.
- MASTENBROEK, C., MAKIN, V.K., VOORRIPS, A.C. and KOMEN, G.J., 1994, Validation of ERS-1 altimeter wave height measurements and assimilation in a North Sea wave model. *Global Atmosphere and Ocean System*, **2**, pp. 143–161.
- PARRISH, D.F. and DERBER, J.C., 1992, The national meteorological center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, pp. 1747–1763.
- PRIESTLEY, M.B., 1981, *Spectral Analysis and Time Series*, p. 890 (London: Academic Press).

- RABIER, F., McNALLY, A., ANDERSSON, E., COURTIER, P., UNDÉN, P., EYRE, J., HOLLINGSWORTH, A. and BOUTTIER, F., 1998, The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: structure functions. *Quarterly Journal of the Royal Meteorological Society*, **124**, pp. 1809–1829.
- SEAMAN, R.S., 1983, Objective analysis accuracies of statistical interpolation and successive correction schemes. *Australian Meteorological Magazine*, **31**, pp. 225–240.
- SEAMAN, R., BOURKE, W., STEINLE, P., HART, T., EMBERY, G., NAUGHTON, M. and RIKUS, L., 1995, Evolution of the Bureau of Meteorology's global assimilation and prediction system, Part 1: analyses and initialization. *Australian Meteorological Magazine*, **44**, pp. 1–18.
- SKANDRANI, C., LEFEVRE, J.-M. and QUEFFEULOU, P., 2005, Impact of multi-satellite altimeter data assimilation on wave analysis and forecast. *Marine Geodesy*, **27**, pp. 511–533.
- SNYDER, R.L., DOBSON, F.W., ELLIOT, J.A. and LONG, R.B., 1981, Array measurements of atmospheric pressure fluctuations above surface gravity waves. *Journal of Fluid Mechanics*, **102**, pp. 1–59.
- VOORRIPS, A.C., MAKIN, K. and HASSELMANN, S., 1997, Assimilation of wave spectra from pitch-and-roll buoys in a North Sea wave model. *Journal of Geophysical Research*, **102**, pp. 5829–5849.
- WAMDI GROUP (S. HASSELMANN, K. HASSELMANN, E. BAUER, P.A.E.M. JANSSEN, G. KOMEN, L. BERTOTTI, P. LIONELLO, A. GUILLAUME, V.C. CARDONE, J.A. GREENWOOD, M. REISTAD, L. ZAMBRESKY and J.A. Ewing), 1988, The WAM model – A third generation wave prediction model. *Journal of Physical Oceanography*, **18**, pp. 1775–1810.
- YOUNG, I.R. and GLOWACKI, T.J., 1996, Assimilation of altimeter wave height data into a spectral wave model using statistical interpolation. *Ocean Engineering*, **23**(8), pp. 667–689.