THE EFFECT OF ALTIMETER SAMPLING PATTERNS ON ESTIMATES OF WAVE MODEL ERROR CORRELATIONS

D. J. M. Greenslade

Bureau of Meteorology Research Centre,
Melbourne, Australia

I. R. Young

Faculty of Engineering, Computer and Mathematical Sciences, University of Adelaide
Adelaide, Australia

1. INTRODUCTION

Many operational forecasting centres run systems to assimilate satellite altimeter data into wave models. A major limitation to the current operational systems is the lack of knowledge on the structure of the model errors. One method used to determine the error correlations is to use data from a long-term observational network and examine the differences between model forecasts and the observations. This work describes an initial step in the calculation of wave model error correlations from altimeter data, namely, an examination of the effect of the sparse irregular altimeter sampling pattern.

The model error correlation matrix is introduced in Section 2 of this paper. In section 3, a set of "true error correlations" is calculated from modelled wave fields, using the model climatology as the background field. In Section 4, the modelled wave fields are sampled along simulated altimeter ground tracks and error correlations re-calculated from this simulated altimeter data. The results are discussed in Section 5 and compared to the "true error correlations". A brief summary is presented in Section 6.

2. THE MODEL ERROR CORRELATION MATRIX

Current operational wave data assimilation systems commonly use the technique of Statistical Interpolation (SI) to combine the satellite altimeter data with the first-guess model fields to obtain analysed fields of Significant Wave Height (SWH). A detailed description of this method can be found in Lionello *et al*. (1995) or Greenslade (2001). An important component of the algorithm is the specification of the error correlation matrix of the model prediction, P_{kj}. This is a symmetric N_{obs} by N_{obs} matrix (N_{obs} = the number of observations) with element *(k, j)* given by:

$$P_{kj} = \frac{\left\langle \left(H_p^k - T^k\right)\left(H_p^j - T^j\right)\right\rangle}{s_p^k s_p^j} \tag{1}$$

where H_p are the first-guess predictions from the model, s_p is the model prediction *rms* error, T is the true SWH and <…> is the expected value. 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 major limitation to the current implementation of SI systems is the lack of a good representation of P_{kj}. It is usually assumed to be isotropic and to have a form similar to:

$$P_{kj} = \exp\left\{-\frac{1}{2}\frac{\left|x_k - x_j\right|^2}{L^2}\right\} \tag{2}$$

where L is the decorrelation length scale and $|x_k - x_j|$ is the distance between the observation locations k and j. An example of P_{kj} with L = 300 km is shown in Figure 1.

In meteorology, a method commonly used to determine the model error correlations is to consider observations from a long-term observational network and examine the difference between model forecasts and the observations (eg. Seaman, 1982, Hollingsworth and Lönnberg, 1986). In the past it has been difficult to apply this method within the field of oceanography. This is mainly due to the lack of a long-term observational network with reasonable space-time sampling characteristics. Thus other methods have been sought.

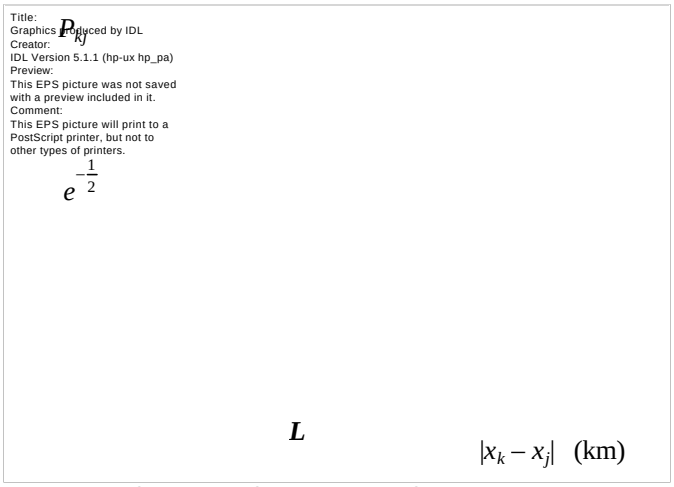

Figure 1. An example of a typical function used for background error correlations.

An alternative method to estimate P_{kj} is to examine the difference between model analyses and model forecasts, ie, assume that the model analysis is the truth. Voorrips et *al*. (1997) apply this method to an ocean wave model and estimate the spatial correlation structure of the forecast error by considering differences between the model forecast at 24 hours and the model analysis. In this case, the analysis was a model run using analysed wind fields, ie, no wave data are assimilated.

The advantage of this method is that there is "data" at all model gridpoints and so it is possible to obtain information on correlations in all directions. However, a major disadvantage of this method is that there is a heavy reliance on the ability of the wave model to provide high quality model analyses to be used as the "truth". This could cause problems in several ways. Even if no wave data assimilation is used in the determination of the model analysis, the analysed SWH may be corrupted by the use of imposed correlation length scales in the assimilation process within the atmospheric model which is used to provide wind fields to the wave model. In addition, it is quite possible that the spatial structure of the errors in the model forecast is similar to the spatial structure of errors in the model analysis, which would result in inaccurate error correlations.

The ultimate aim of this work is to determine error correlations via the first method described above, ie, by considering the differences between model forecasts and observations. Highest quality wave observations generally come from wave-rider buoys or platforms which are fixed in space. However, the spatial distribution of buoys is extremely poor - for practical reasons they are generally located in coastal regions - and so it is difficult to obtain any information on the spatial error correlation structure from buoy data. Satellite altimeters, however, now provide a comprehensive global long-term network of wave observations, which can be used to determine P_{kj}.

In this paper, the ability of a typical altimeter sampling pattern to obtain estimates of model error correlations is examined. This is achieved by first calculating "error correlations" from modelled wave fields. These are considered to be the "true" error correlations. The modelled wave fields are then sampled along simulated altimeter ground tracks, error correlations are re-calculated from this simulated altimeter data and the results are compared to the "true" error correlations.

Equation (1) can be expressed as the spatial error correlation between two points, j and k, ie, (Daley, 1991)

$$R_{jk} = \frac{\overline{(O_j - B_j)\,(O_k - B_k)}}{\sqrt{\overline{(O_j - B_j)^2}\;\;\overline{(O_k - B_k)^2}}} = r(r,q) \qquad (3)$$

where r is the error correlation as a function of great circle distance, r, and angle, q, the overbar represents a time-average, O_i are "observations" and B_i represent the background values. The actual fields used for these variables depend on the method that is being used to determine the error correlations.

3. ERROR CORRELATIONS FROM WAVE MODEL

The wave model used here is AUSWAM, a version of WAM cycle 4 (WAMDI Group, 1988, Komen *et al*., 1994) with cycle 2 physics (Snyder, 1981) and 3rd-order upwinding numerics. No wave data assimilation was used in the construction of the modelled wave fields used in this work. The domain spans the globe from 78° N to 78° S at 0.5° spatial resolution. Fields of SWH are output every 6 hours. Other details of the wave model (eg, wind fields) are not relevant for this work and so are not listed here.

3.1 Correlation computations

A range of time periods and spatial domains were considered for the calculation of the wave model correlations. The size of the domain can play an important role because fields of SWH are typically not homogenous over the ocean. Generally, the larger the area considered, the larger the length scale of the correlation. On the other hand, on time scales of weeks to months, SWH can be assumed to be stationary, so the time period chosen has less of an impact on the correlation functions than the spatial domain. The selected time periods and domains were partly dictated by the altimeter sampling pattern.

The time period chosen needed to be long enough so that at any location, there were several repeat observations from the altimeter. It was also desirable to have it short enough to enable detection of any seasonality in the correlations. Thus a time period of three months was chosen. The spatial dimensions of the domain needed to be large enough so that prior and subsequent altimeter ground tracks could be used (this is discussed in detail later). However, the motivation for this work is data assimilation, and in data assimilation processes, one is most interested in what happens at small scales. A box size of 20° in latitude and longitude was found to be the best compromise between these two requirements.

For each pair of gridpoints j, k within the domain, R_{jk} was calculated according to Equation (3). Here, O_i are the 6-hourly model SWH values and B_i is the model climatology, ie.

$$B_i = \overline{O_i} = \frac{1}{N_t}\sum_{t=1}^{N_t} O_i(t) \qquad (4)$$

where N_t is the number of 6-hourly model fields within the 3-month time period. In addition, the great circle distance, r (km) and q, the great circle bearing (angle in degrees clockwise from North) between points j and k are calculated. Thus the "errors" considered here are in fact anomalies from a 3-month climatological average.

Correlations were averaged into 1° by 1 km bins. Figure 2(a) shows an example of the model error correlations for a box centred at (60W, 0N), ie, in the north-western Indian Ocean for the time period July - September 1998. Clearly, $R_{jk} = R_{kj}$ and so the function is invariant under a 180° rotation.

This correlation function can be interpreted as follows. Consider a model gridpoint within this 20° box. Assume that at this gridpoint, at a particular time, the deviation in SWH from the model climatology is large. Then firstly, and most obviously, it can be seen that the SWH at gridpoints close to this gridpoint is also likely to deviate strongly from the climatology, whereas SWH at gridpoints further away is less likely to deviate strongly from the climatology.

For this particular region, the decay in the correlation function is slower in the zonal direction than the meridional direction, ie, correlations are larger in the east-west direction than the north-south direction. From this elongated shape of the correlation function, it can therefore be seen that a gridpoint x km to the east or west is more likely to have a large deviation from the climatology than a point x km to the north or south. Similarly, if SWH at this gridpoint is close to the climatology, then points to the east or west are more likely to be close to the climatological values than points to the north or south. The equivalent model error correlation using only every fifth model gridpoint is shown in Figure 2(b). It can be seen that although the resolution is coarser and there is a "banded" effect in the correlation function, the major features of this plot are similar to those in Figure 2(a).

3.2 Fitting to analytic functions

The calculated correlations $r(r,q)$ were then fitted to analytic functions. A commonly used function in studies of surface pressure and temperature error correlations (Julian and Thiebaux (1975), Seaman (1982)) is:

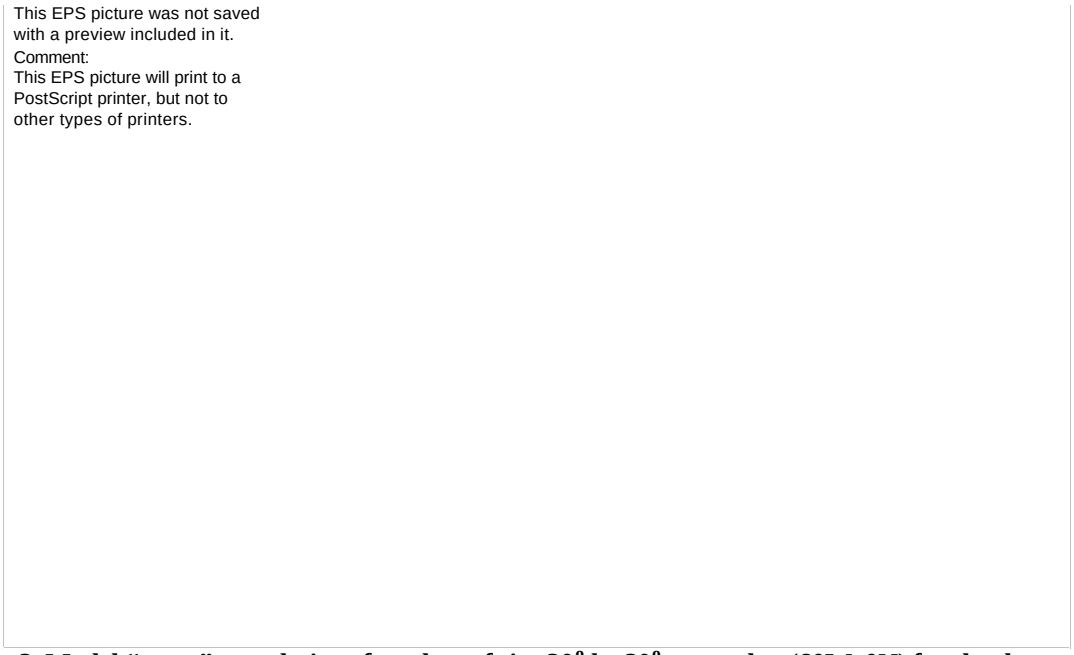

Figure 2. Model "error" correlations for a box of size 20° by 20° centred at (60W, 0N) for the three-month time period July - September 1998 (a) using all model grid points and (b) using every fifth model gridpoint.

$$r(r,q) = (a_4 \cos(a_5 d) + a_6 - a_4)\exp\{a_3 d\} \tag{5}$$

where

$$d^2 = r^2\left(\frac{1}{a_1^2}\cos^2(q - a_2) + a_1^2 \sin^2(q - a_2)\right) \tag{6}$$

The justification for including a cosine dependency is that often the temperature and pressure error correlation functions decay to negative values before returning to zero at longer distances. For SWH on the scales of interest here, this is not generally the case, so a simpler form for the correlation function can be considered by removing the cosine term. In addition, for the time being, the constraint that $r_{(r=0)} = 1.0$ is included. These simplifications are not vitally important because for this work it is not so much the actual *form* of the correlation function that is important, as the impact that the altimeter sampling pattern has on estimates of that form. The simplified function is:

$$r(r,q) = \exp\{a_3 d\} \tag{7}$$

with d given by Equation (6). This results in elliptical contours (which seems reasonable considering Figure 2) with the parameter a_1 related to the eccentricity (e) of the ellipse, a_2 gives the tilt of the ellipse (angle of the major axis in degrees from North) and a_3 defines a length scale (or a rate of decay). The procedure used to find the best estimate for a_i was a NAG Fortran library subroutine for finding an unconstrained minimum of a sum of squares. In particular,

$$F(a_1, a_2, a_3) = \sum_{i=1}^{m}\left(r(r_i, q_i) - \exp\{a_3 d_i\}\right)^2 \tag{8}$$

is minimised.

Figure 3 shows contours of the analytic functions that were found to be the best fit to the model error correlations shown in Figure 2. The best fit values for the parameters in this case are shown in Table1.

	a_1	a_2	$1/a_3$
All points	1.50	86.0	2168.9
Every fifth point	1.50	85.0	2052.6

Table 1. Parameter values for the functions contoured in Figure 3.

Figures 2, 3 and Table 1 demonstrate that the thinning of the model data by using only every fifth gridpoint has only a minor effect on the results. So to reduce computational time, all further correlations were calculated using only every fifth model gridpoint.

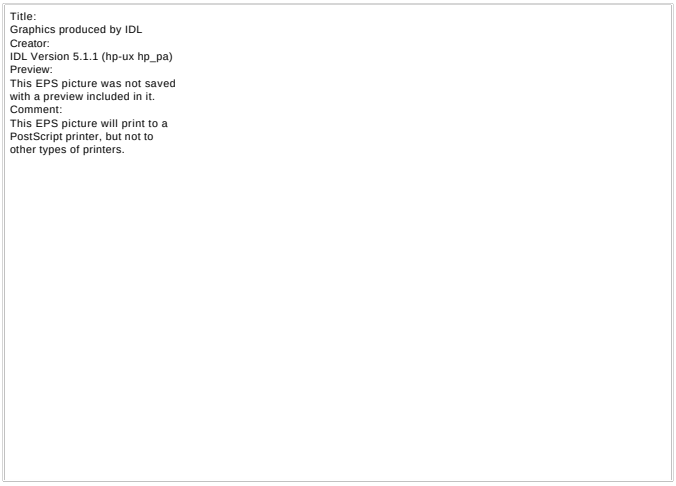

Figure 3. Best-fit analytic surfaces for the correlation functions shown in Figure 2. The solid line is for the case of all model gridpoints and the dashed line for the case using every fifth point.

In addition to the anisotropic (2-D) fitting, the correlations were fit to isotropic (1-D) functions. Candidate functions were a Gaussian, equivalent to Equation (2):

$$r(r) = \exp\left\{-\frac{1}{2}\frac{r^2}{a_1^2}\right\} \tag{9}$$

and a 2nd-order auto-regressive function:

$$r(r) = (1 + a_1 r)\exp\{-a_1 r\} \tag{10}$$

Figure 4 shows the same correlations as in Figure 2 (b) but as a function of distance alone, along with the best fit 1-D functions. Overall, it was found that the 2nd-order auto-regressive function was a better fit to the correlations than the Gaussian. For the remainder of this paper, only the auto-regressive function is considered. The fact that Equation (10) is a better fit for the isotropic case than Equation (9) may indicate that the best fit anisotropic function is an auto-regressive function with angle-varying distance, ie, Equation (10) with r given by d in Equation (6). However, this is not considered here.

The best fit value for the parameter a_1 in the 1-D case (ie, the value of a_1 in Equation (10) represented by the long dashes in Figure 4) was found to be $a_1 = 1.035 \times 10^{-3}$. The error correlation length scale is defined as $1/a_1$ and in this case is approximately 966 km.

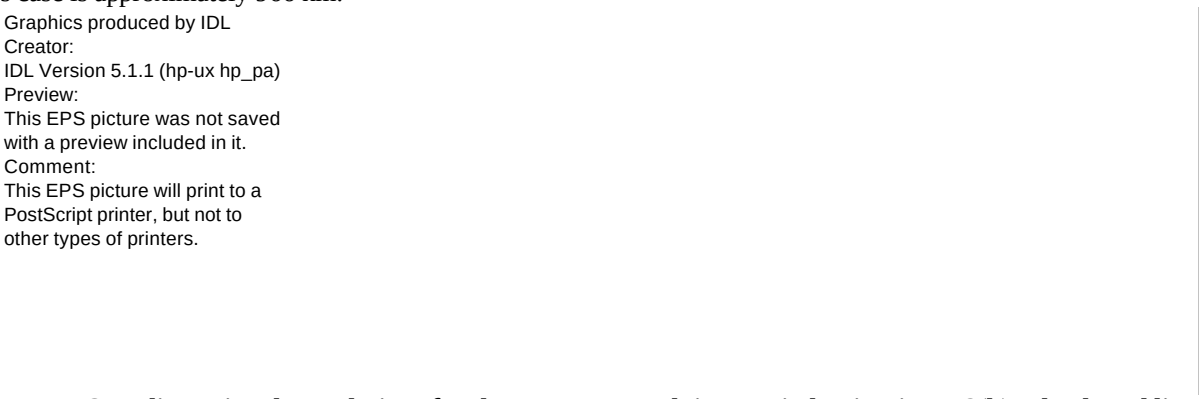

Figure 4. One-dimensional correlations for the same area and time period as in Figure 2(b). The dotted line is the best fit Gaussian curve and the dashed line is the best fit auto-regressive curve.

3.3 Regional effects

Model error correlations were calculated as described above for 20o boxes at 10^o intervals over the globe for the three-month time period July - September 1998. Best fit curves for both the isotropic and anisotropic cases were determined for each box. Correlations for an additional time period (Jan – Mar 1999) were also calculated, however, the details are not presented here, for brevity.

3.3.1 Isotropic case

Figure 5 shows how the 1-D error correlation length scale ($1/a_1$ in Equation (10)) varies over the globe for this time period. It can immediately be seen that the length scale varies quite significantly in space.

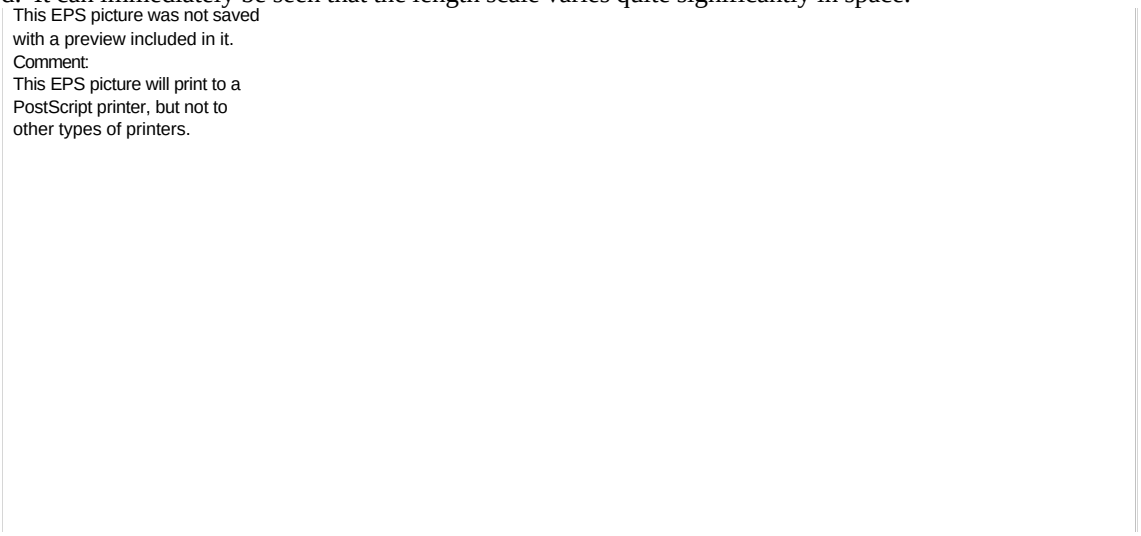

Figure 5. One-dimensional error correlation length scale (km) from modelled SWH fields over the globe for 20^o boxes at 10^o intervals for the time period July to September 1998.

The longest scales are to be found in the northern Indian Ocean (Arabian Sea) and in the eastern equatorial Pacific. Longer scales imply that the model deviates from the climatology on a large spatial scale. This means that areas of anomalously high or low SWH are large during this time period. The area of long length scales in the Arabian Sea is associated with the Indian Monsoon. From May to September, winds in this region blow persistently from the south-west, and so during this time period, the fetch for this area is relatively long, stretching along the East African coast. Thus an area of persistently high SWH develops in the Arabian Sea (Young, 1999). The size of the correlation length scales in Figure 5 reflects this.

The shortest length scales (< 500 km) can be seen on the western boundaries of the ocean basins and also in the Southern Ocean in a band of short scales around 40^oS. The short scales in this area reflect the size of the storm areas that propagate from west to east along this latitude band.

3.3.2 Anisotropic case

The correlation functions shown in Figures 2 and 3 indicate that the model error correlation is clearly not isotropic. An indication of the anisotropy of a particular correlation function can be obtained from inspecting just one contour level, since each contour holds the same information on the eccentricity and tilt of the ellipse. Figure 6 shows the 0.5-level contour of the anisotropic model error correlations for each 20^o box.

The relative size of each ellipse corresponds well to the 1-D length scales shown in Figure 5. In particular, note that the largest ellipses occur in the Eastern Pacific and the Arabian Sea. Overall, it can be seen that there is generally a large amount of anisotropy in the error correlations. Areas that are particularly anisotropic are the Indian Ocean to the west of Australia and the Pacific Ocean to the west of South America. If the actual model errors have the same structure as the anomalies shown here, this may have implications for current data assimilation schemes, which use simple isotropic correlation functions.

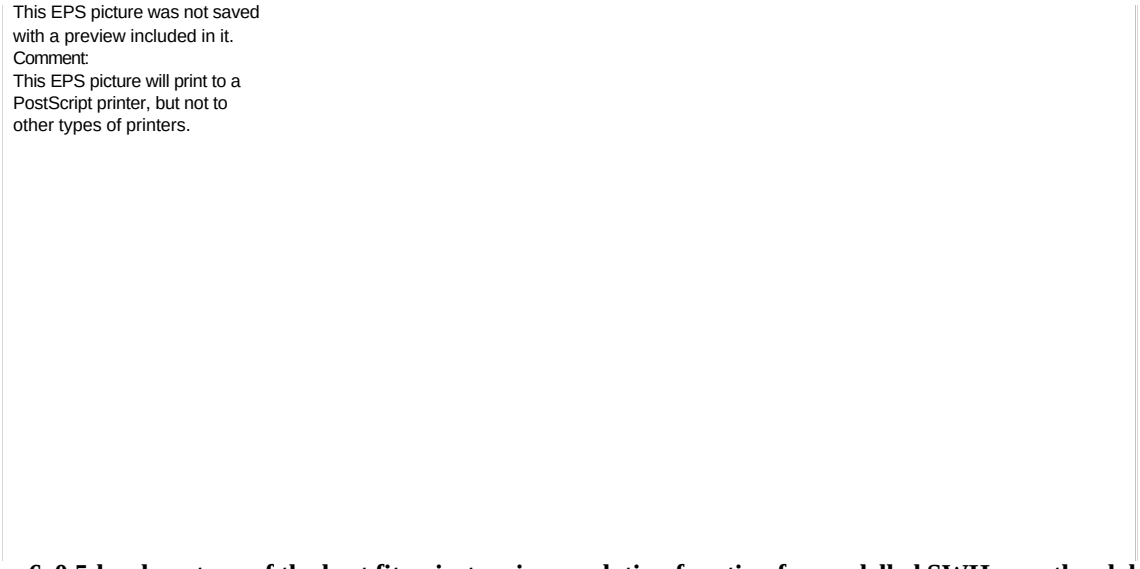

Figure 6. 0.5-level contour of the best fit anisotropic correlation function for modelled SWH over the globe for 20^o boxes at 10^o intervals for the time period July to September 1998.

4. ERROR CORRELATIONS FROM SIMULATED ALTIMETER DATA

Altimeter ground tracks (ie, latitude, longitude and time) were created for a satellite with an altitude of 788 km and an orbital inclination of 108^o. These are the parameters of the GEOSAT mission. The repeat period of the GEOSAT altimeter is 17.05 days. This means that within a three-month period, each ground track will be sampled 5 or 6 times. The orbital period of the satellite, ie, the time between subsequent ascending (or descending) equator crossings is approximately 100 minutes. The along-track spacing of observations was set to be 20 km.

A set of ground tracks was generated for every 20^o box for the three-month period described above. For each box, the background field (ie, the model climatology) and the 6-hourly model fields were interpolated to the simulated altimeter observation locations. Bilinear interpolation was required for the background fields (two space dimensions) and trilinear interpolation for the 6-hourly model fields (two space dimensions plus time). Figure 7 shows an example of a model background field and a set of altimeter ground tracks.

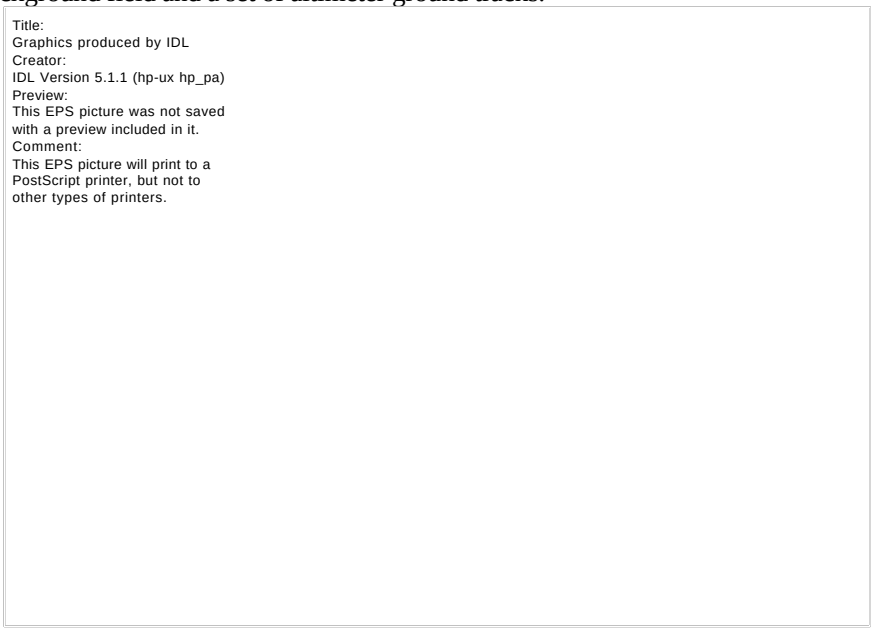

Figure 7. A model background field, ie, climatological SWH (m) with simulated altimeter observation locations.

The calculation of error correlations from the simulated altimeter data is complicated by the fact that an altimeter dataset represents a time series and few of the observations can be considered to be simultaneous. To calculate a correlation between two observation locations j and k according to Equation (3), there are two criteria that must be satisfied:

1. There must be simultaneous observations at each location. This is easily satisfied for the model fields, but for altimeter data, there are *no* observations that are simultaneous. However, simultaneous can be defined as: occurring within a short enough time period (say, t_{max}) so that the field of interest does not vary significantly. This is discussed further below

2. Not only must criterion 1 be satisfied, but it must also be satisfied more than once within the time period of interest. If there is only one observation at each location then R_{jk} reduces to the trivial case of $R_{jk} = 1$. This is one of the reasons a time period of three months was chosen, as opposed to a shorter time period.

The temporal distribution of the altimeter observations within a 20^o box on the earth's surface is complex. For a 20^o box near the equator, one overpass takes approximately 6 minutes. If this is, say, an ascending pass, then the time to the next observation within the box depends on whether the next ascending pass falls within the box or not. If it does, then the time to the next overpass is approximately 100 minutes (ie, the orbital period). If the next ascending pass falls outside the box, then the next observation falling within the box will come from a descending pass. This will occur after the earth has gone through half a rotation period (ie, 12 hours) and so the time to the next observation within the box may be either approximately 10.8 hours or 12.5 hours.

The combination of the two criteria outlined above and the complexities of the altimeter sampling pattern means that if t_{max} = 15 minutes, then the set of data-pairs between which correlations can be calculated consists of only along-track combinations. This severely limits the amount of directional information on the correlations. However, if t_{max} is extended to 2 hours then this allows the inclusion of subsequent (or prior) same-direction ground-tracks. This provides additional information in the zonal direction.

The disadvantage of using t_{max} = 2 hours is that there is the possibility that the SWH field will have altered within the 2 hour period, and this may then corrupt the correlation calculations. Situations in which the SWH field would be expected to alter the most rapidly are those in which the wind speed and/or wind direction changes abruptly. The time-scale of the response of the mean wave direction to a change in wind direction has been shown to be greater than 5 hours (Günther *et al*., 1981, Komen *et al*., 1994). This indicates that 2 hours is a reasonable time period during which it can be assumed that the SWH field will not change significantly.

Following Equation (3), error correlations were calculated from the simulated altimeter data for all 20^o boxes over the globe at 10^o intervals. For these calculations, B_i is the model climatology interpolated to the altimeter observation locations, and the O_i are the simulated altimeter observations. The time average is the sum over the number of time-levels for which observations occur at both locations j and k. These correlations were then fitted to the same analytic functions used for the model error correlations (Equation (10) for the 1-D case and Equation (7) for the 2-D case).

Figure 8 (a) shows an example of the error correlations from the model, ie, the "true" error correlations and 8(c) and (e) show the equivalent error correlations for the altimeter sampled case for the two different values of $t_{max.}$ The best fit analytic surfaces are shown in the bottom panels. This case is for a box centred at (70W, 20S), in the central Indian Ocean. Table 2 gives the values of a_i for each case.

	a_1	a_2	$1/a_3$
Model	1.92	123.4	1737.1
Alt. t_{max} = 15 min.	3.46	169.4	1679.6
Alt. t_{max} = 2 hrs	2.32	132.2	1908.3

Table 2. Parameter values for the surfaces shown in Figure 8(b), (d) and (f).

Consider first the altimeter case with t_{max} = 15 minutes (Figure 8(c) and (d)). The first point to note is that as opposed to the true error correlations, the correlations using the simulated altimeter data do not look much like an ellipse. Correlations are calculated only between along-track data pairs, so there is little information available on the value of the correlations in the east-west direction. Some basic features are still evident, however. The correlations are higher at shorter scales and the correlations in the north-west corner are lower than in the south-west corner. This

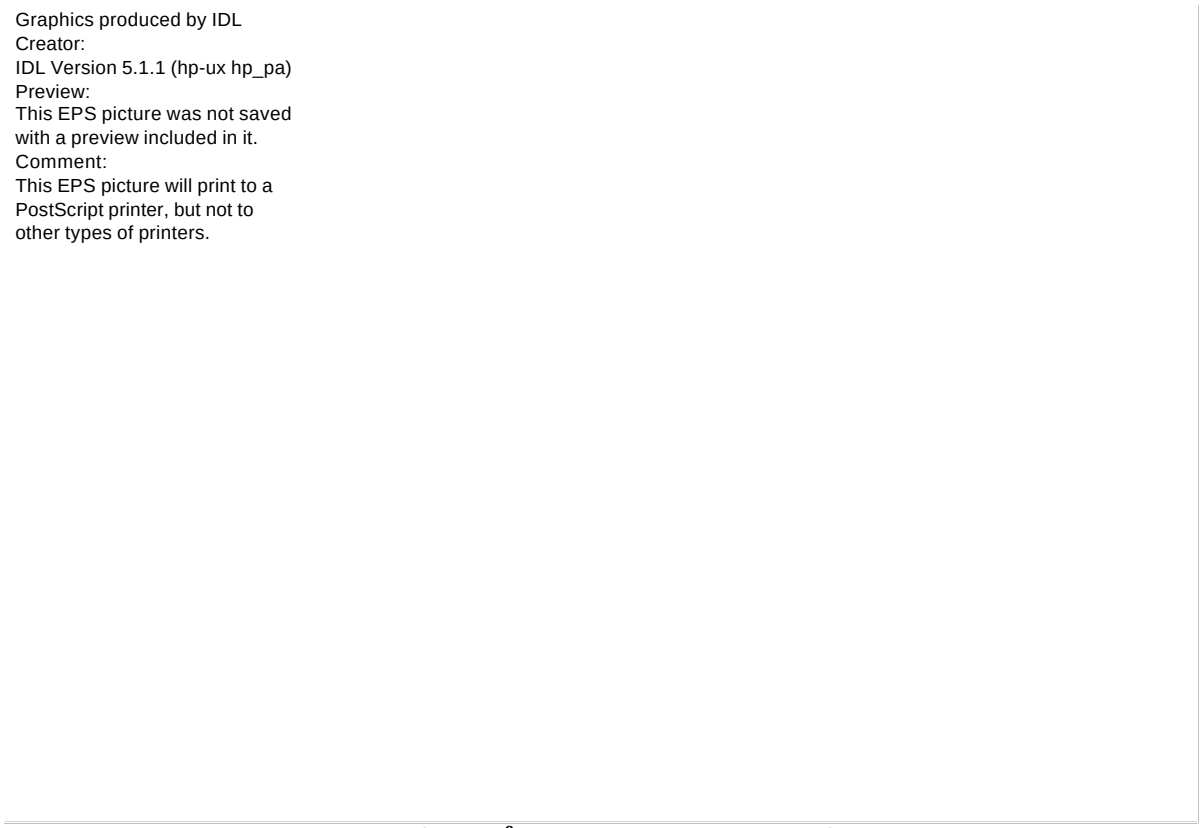

Figure 8. (a) Model error correlations for a 20^o box centred at (70W, 20S) for the time period July to September 1998. (b) Best fit to the correlations shown in (a). (c) Same as (a) but using data from simulated altimeter ground tracks and calculating correlations with t_{max} = 15 minutes. (d) Best fit to the correlations shown in (c). (e) Same as (a) but using data from simulated altimeter ground tracks and calculating the correlations with t_{max} = 2 hours. (f) Best fit to the correlations shown in (e).

leads to an analytic function for which the main features have some similarity to those of the model correlations. In particular, the major axis of the ellipse is in the correct quadrant and the decay rate/length scale is similar.

Figure 8(e) and (f) show how the correlations change with the inclusion of subsequent/prior ground tracks, ie. t_{max} = 2 hours. The best fit analytic function is now considerably closer to that of the model error correlation and this is reflected in the a_i values. Although a_3 is now slightly further from the "true" value, a_1 and a_2 are quite close to those of the model error correlations.

In Figure 9 the equivalent plots for the 1-D correlations along with the best fit auto-regressive curves are shown. The value of $1/a_1$ in each case is shown in Table 3. Note the two distinct tails in the model error correlations, representing the higher values in the south-eastern quadrant and lower values in the north-eastern quadrant of the 2-D correlations. With the t_{max} = 2 hour case, the extra correlation values at large values of r (ie. correlations between observations from different ground tracks) are not in general close to those of the model, but they do force the fitted curve to be closer to the "true" correlation function.

	$1/a_1$ (L)
Model	692.9
Alt t_{max} = 15 min.	566.5
Alt t_{max} = 2 hrs	610.4

Table 3. Parameter values for the curves shown in Figure 9.

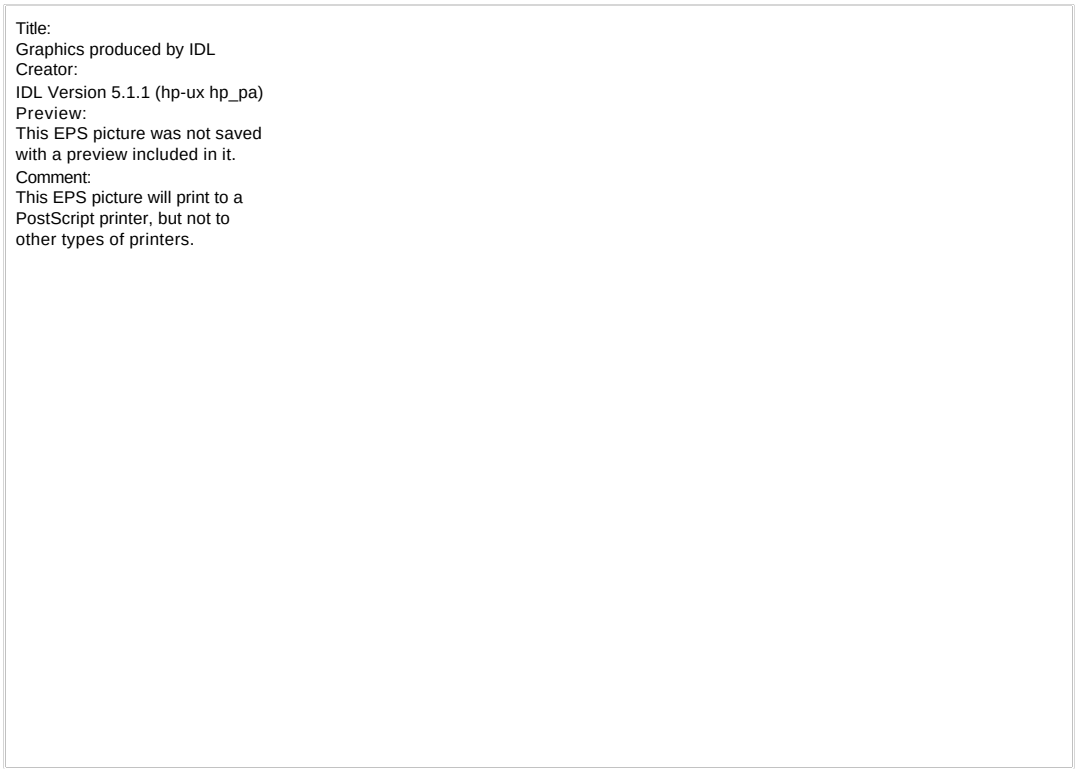

Figure 9. Same as Figure 8 but for isotropic fitting. The dashed line in each panel represents the best fit auto-regressive function. (a) Correlations from model output. (b) Correlations from simulated altimeter data with t_{max} = 15 minutes. (c) Correlations from simulated altimeter data with t_{max} = 2 hours

The model error correlation length scale is defined to be the true length scale, L, and in this case, L = 692.9. The length scale obtained from simulated altimeter data with t_{max} = 2 hours (L_{alt_2hr}) underpredicts the true length scale by approximately 12%. Similarly, the length scale obtained from simulated altimeter data with t_{max} = 15 minutes (L_{alt_15min}) underpredicts the true length scale by approximately 18%. These values are fairly typical and will be discussed further in Section 5.

4.1 Regional effects

As in the case of the modelled wave fields, correlations were calculated for all 20^o boxes over the globe for the same 3-month time period from simulated altimeter data. In this section, only the case with t_{max} = 2 hours is presented.

4.1.1 Isotropic case

Figure 10 shows how L_{alt_2hr} varies over the globe for this 3-month time period. Comparing this to Figure 5, it can be seen that most of the main features are evident. In particular, the longest scales occur in the Eastern Pacific and the Arabian Sea, and a band of short scales appears in the Southern Ocean, although in this case it occurs at around 30^{o}S. In addition, the pattern in the Atlantic Ocean (N and S) is very similar in the two plots. There are a few obvious anomalies occurring south of Japan and west of Tasmania. In these areas, L_{alt_2hr} is considerably larger than L. These anomalies are basically a result of the full model error correlation having a different shape at short scales and at long scales. Detailed discussion of these cases is not presented here, due to space constraints.

4.1.2 Anisotropic case

The fitting of anisotropic functions to the correlations from the simulated altimeter data is now considered. Figure 11 shows the 0.5 level contour of the error correlations from the simulated altimeter data with t_{max} = 2 hours for each 20^o box. Comparing this to Figure 6 it can be seen that there are many locations where the ellipses are very different to the model ellipses. In many regions, the contours are not closed - this indicates that a_1 is too big. On the other hand, there are some encouraging similarities between the plots.

The most eccentric ellipses are generally found in the eastern equatorial Pacific. South of the equator, these are generally tilted in the right direction. In the northernmost and southernmost regions the comparison is relatively good ie,

Figure 10. Same as Figure 5 but for correlations calculated from simulated altimeter data.

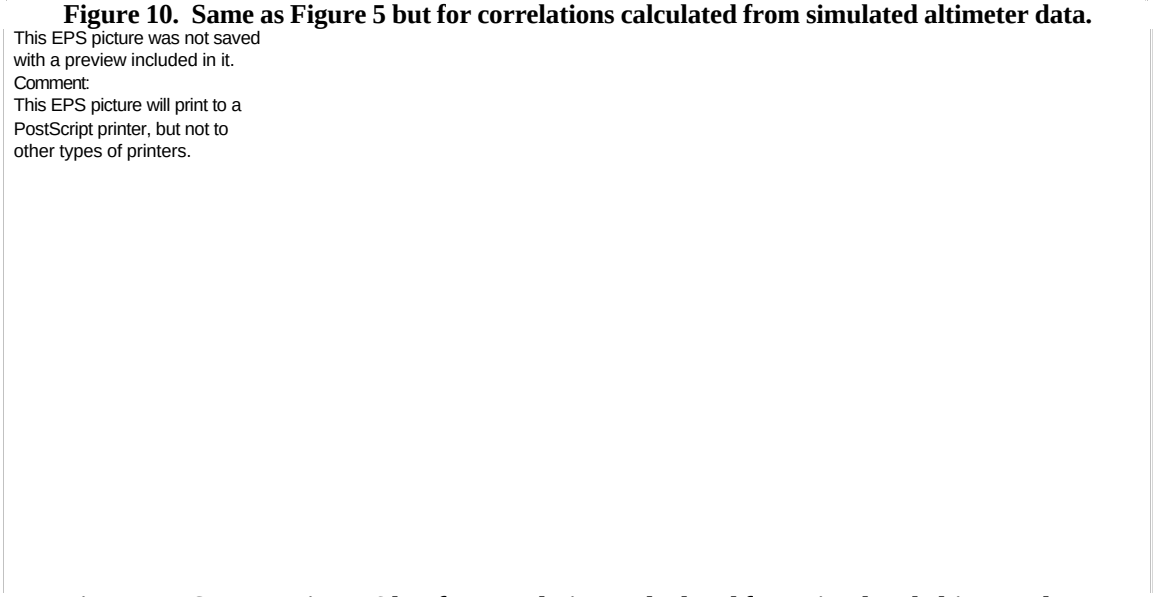

Figure 11. Same as Figure 6 but for correlations calculated from simulated altimeter data.

the size of the ellipses is about right and they are generally isotropic or slightly north-south elongated. The improved performance of the altimeter sampling pattern at higher latitudes is partly due to the higher density of observations in these regions, and also due to their distribution. The convergence of meridians means that the altimeter ground tracks become more east-west aligned at higher latitudes (see Figure 7). This means that at higher latitudes, the data-pairs for the correlation calculations occur over a much larger range of angles and so there is more information available on the correlations in the east-west direction.

5. DISCUSSION

In this section, the 1-D correlation length scales from the simulated altimeter data are examined further. Figure 12 shows the difference between L and L_{alt_2hr} as a percentage of L. Over most of the ocean, the altimeter sampling pattern causes the length scale to be underpredicted (lighter shading). The areas where L is underpredicted by the largest amount are generally near the centres of ocean basins, while areas where the altimeter overpredicts the length scale are generally near the coast, or along the southernmost latitude band.

The solid line in Figure 13 shows a histogram of the differences shown in Figure 12. Results from the second time period (Jan – March 1999) are also included here (the results are very similar for both time periods). Boxes with

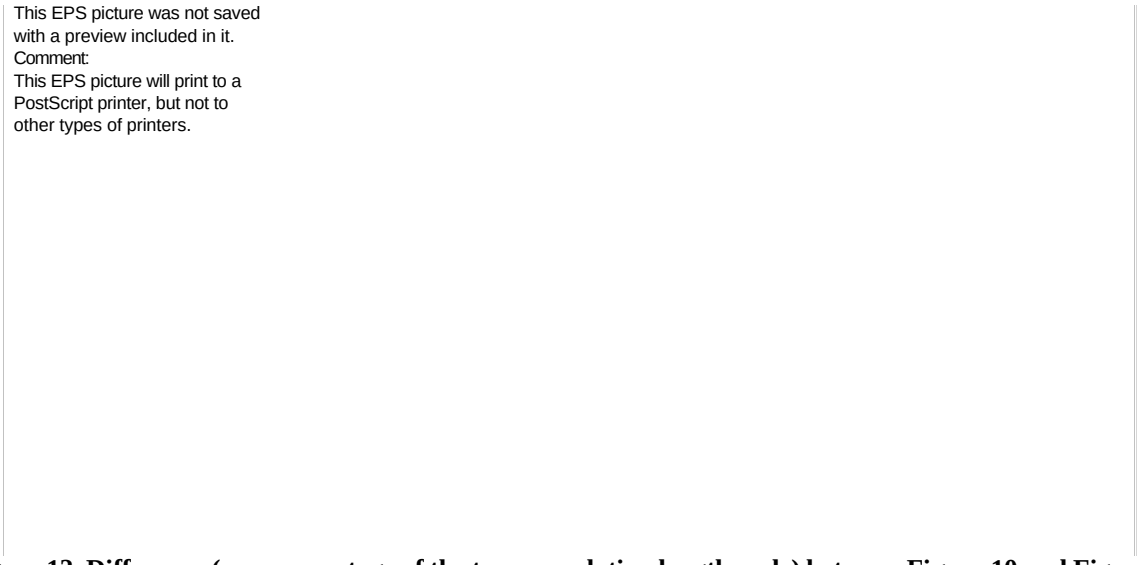

Figure 12. Difference (as a percentage of the true correlation length scale) between Figure 10 and Figure 5.

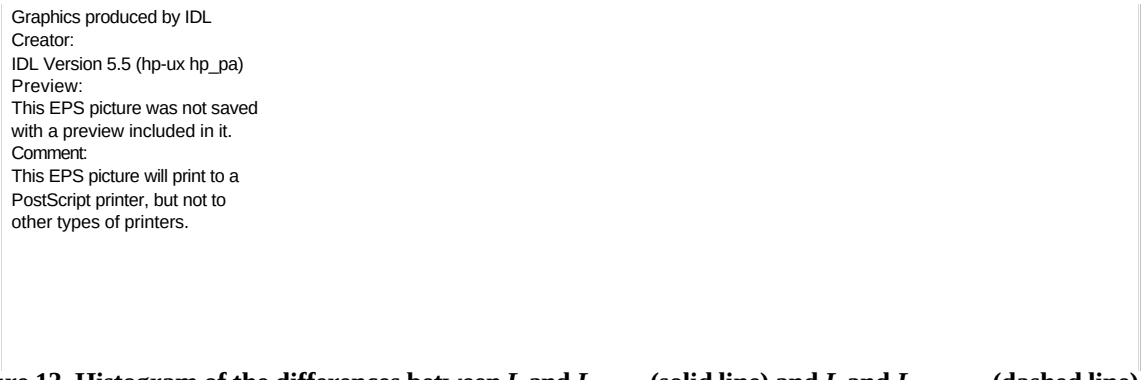

Figure 13. Histogram of the differences between L and L_{alt_2hr} (solid line) and L and L_{alt_15min} (dashed line) as a percentage of L for two 3-month time periods.

significant amounts of land are not considered. The mean difference is about 10.5%, suggesting that a possible course of action might be to calculate L_{alt_2hr} and then increase it by 10.5% to obtain the true correlation length scale. (Strictly speaking this should be 11.7% since if $L_{alt\_2hr} = 89.5\% L$ then $L = 111.7\% L_{alt\_2hr}$). However, the standard deviation here is relatively large (20.4%) indicating that there is not much confidence in this method.

The two anomalies referred to in the previous section showed that the inclusion of prior and subsequent altimeter ground tracks may, in certain situations, degrade the estimation of L. Since a method is sought which provides a consistent correction to L_{alt}, and not necessarily the method that produces values closest to the true L, it is worthwhile to examine here the estimates of L_{alt_15min}.

The global distribution of L_{alt_15min} and the global distribution of differences (not shown here) are very similar qualitatively to Figures 10 and 12. The corresponding histogram of the differences is shown in Figure 13 (dashed line). The mean difference between L and L_{alt_15min} is 14.0%, ie, it is larger on average than the difference between L and L_{alt_2hr}. However the standard deviation for the 15-minute case is 18.5%. Thus a more confident estimate of the true correlation length scale can be obtained by calculating L_{alt_15min} and increasing it by 16.3%.

The amount by which the altimeter underpredicts (or overpredicts) the 1-D correlation scale should be a function of the shape of the true 2-D correlation function. If the 2-D correlation is isotropic, then the correlation length scale is the same in all directions, and the limited sampling of the surface by the altimeter should not have any impact on the determination of the length scale. As the 2-D correlation function becomes more anisotropic, and the ellipses become more eccentric, then the underprediction of L by the altimeter becomes dependent upon the angle between the altimeter ground tracks and the major axis of the ellipse.

This will be investigated in future work. One approach could be to examine length scales obtained from ascending and descending altimeter ground tracks separately. Other future work involves examining differences between real (not simulated) altimeter data and modelled wave fields to determine the actual model error correlations. In addition, the spatial and temporal variability of the correlation length scales will be investigated and the impact of this variability on data assimilation schemes will be considered.

6. SUMMARY

The ultimate aim of this work is to use satellite altimeter SWH data to determine the structure of wave model errors. In the work presented in this paper, the effect of the sparse, irregular altimeter sampling pattern has been considered. This was achieved by first calculating "true error correlations" from modelled wave fields, using the model climatology as the background field. The modelled wave fields were then sampled along simulated altimeter ground tracks, error correlations re-calculated from this simulated altimeter data and the results compared to the "true error correlations". It was found that regional variability in the 1-D error correlation length scale is captured quite well by the simulated altimeter data. The correlation length scale is generally underpredicted as a result of the irregular altimeter sampling pattern. It was shown that the best estimates of the correlation length scale are achieved by using only along-track altimeter data. Determination of any anisotropy in the error correlations from altimeter data appears less successful, although there are promising results at high latitudes.

6. REFERENCES

Daley, R. 1991. Atmospheric Data Analysis, Cambridge University Press, Cambridge, U.K., 457pp.

Greenslade, D. J. M., 2001. The assimilation of ERS-2 significant wave height data in the Australian region, *J. Mar. Sys*, **28**, 141-160.

Günther, H., Rosenthal, W. and Dunckel, M. 1981. The response of surface gravity waves to changing wind directions. *J. Phys. Oceanogr*., **10**, 718 - 728.

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**, 111 - 136.

Julian, P., and Thiebaux, H. J. 1975. On some properties of correlation functions used in optimum interpolation schemes, *Mon. Wea. Rev*., **103**, 605 - 616.

Komen, G. J., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S. and Janssen, P.A.E.M. 1994. Dynamics and modelling of ocean waves. Cambridge University Press, Cambridge, UK, 532 pp.

Lionello, P., Günther, H., and Hansen, B. 1995. A sequential assimilation scheme applied to global wave analysis and prediction. *J. Mar. Sys*., **6,** 87 - 107.

Seaman, R. S. 1982. A systematic description of the spatial variability of geopotential and temperature in the Australian region. *Aust. Met. Mag*., **30**, 133 - 141.

Snyder, R. L., Dobson, F.W., Elliott, J.A. and Long, R.B. 1981. Array measurements of atmospheric pressure fluctuations above surface gravity waves, *J. Fluid Mech*., **102**, 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, *J. Geophys. Res*, **102**, 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. *J. Phys. Oceanogr*., **18**, 1775 - 1810.

Young, I.R. 1999. Wind Generated Ocean Waves. Elsevier Science, Oxford, U.K. 288 pp.