Ocean Modelling 72 (2013) 17–31

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

Data assimilation of partitioned HF radar wave data into Wavewatch III

Jennifer Waters [a,c,*], Lucy R. Wyatt [a,d], Judith Wolf [b], Adrian Hines [c]

[a] School of mathematics and statistics, The University of Sheffield, Sheffield S10 2TN, UK
[b] National Oceanography Centre, Joseph Proudman Building, 6 Brownlow Street, Liverpool L3 5DA, UK
[c] The Met Office, Fitzroy Road, Exeter EX1 3PB, UK
[d] ACORN, school of earth and environmental sciences, James Cook University, Townsville, QLD 4811, Australia

ARTICLE INFO

Article history:
Available online 27 July 2013

Keywords:
Data assimilation
High frequency radar
Wave spectrum
Wave modelling

ABSTRACT

In this study the assimilation of HF radar data into a high resolution, coastal Wavewatch III model is investigated. An optimal interpolation scheme is used to assimilate the data and the design of a background error covariance matrix which reflects the local conditions and difficulties associated with a coastal domain is discussed. Two assimilation schemes are trialled; a scheme which assimilates mean parameters from the HF radar data and a scheme which assimilates partitioned spectral HF radar data. This study demonstrates the feasibility of assimilating partitioned wave data into a coastal domain. The results show that the assimilation schemes provide satisfactory improvements to significant wave heights but more mixed results for mean periods. The best improvements are seen during a stormy period with turning winds. During this period the model is deficient at capturing the change in wave directions and the peak in the waveheights, while the high sea state ensures good quality HF radar data for assimilation. The study also suggests that there are both physical and practical advantages to assimilating partitioned wave data compared to assimilating mean parameters for the whole spectrum.

1. Introduction

Observing and modelling waves in coastal regions is important for applications such as shipping, offshore engineering and development of sea defences. Data assimilation is a process by which models and observations are combined to give the best estimate of the true state, known as the analysis. It is a useful tool for initialising forecasts and optimising hindcasts which in turn can improve our understanding of the ocean and coastal conditions.

Data assimilation into wave models is a relatively new subject. While data assimilation into atmospheric models began in the 1950s and 1960s, data assimilation into wave models was not addressed until the 1980s. In the 1980s the higher quality of the wind fields being produced by atmospheric models and the increase in wave observations through the introduction of ocean satellite radars such as the SAR (Synthetic Aperture Radar) were instrumental for the extension of assimilation to wave models. The majority of wave models in operational use are third generation spectral wave models. In general it is not practical to assimilate the whole two dimensional wave spectrum into these wave models. The main reason is due to the difficulty of calculating an analysis for all the spectral components and the high computational cost involved with doing this. The limited availability of full frequency-direction spectra from observations has also restricted the possibility of assimilating the whole spectrum. Furthermore, correlation exists between spectral components and ideally this would need to be specified if assimilating the spectrum directly. This correlation would be difficult to define and would further increase the cost of the scheme. So as an alternative most authors have chosen to assimilate a selection of mean parameters and wind parameters such as significant waveheight (*Hs*), mean period (*Tz*), wind speed and wind direction. But since the wave models are spectral models, an analysis spectrum needs to be generated by adjusting the original model spectrum using the assimilated parameters.

Various schemes have been proposed, some are as simple as scaling the whole model spectrum to the analysis wave height, for example Esteva (1988) and Bauer et al. (1992), while others (Thomas, 1988; Foreman et al., 1994; Francis and Stratton, 1990) consider the windsea and swell parts of the spectrum separately. Lionello et al. (1992) classified a spectrum as either windsea, swell or mixed windsea and applied different techniques for scaling the spectrum dependent on the classification.

By the mid 1990s the idea of assimilating partitioned spectral wave data had been proposed (Voorrips et al., 1997; Hasselmann et al., 1997). The idea was that rather than splitting the wave spectrum into a windsea and swell component using methods for characterising a windsea from the local wind, a more elegant partitioning method based on the topography of the spectrum could be used to identify all the component wave trains present (allowing for more than one swell wave train). It is then assumed

* Corresponding author at: The Met Office, Fitzroy Road, Exeter EX1 3PB, UK. Tel.: +44 1392 886155; fax: +44 1392 885681.
E-mail address: jennifer.waters@metoffice.gov.uk (J. Waters).

1463-5003/$ - see front matter Crown Copyright © 2013 Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ocemod.2013.07.003

that each partition within the spectrum represents a different wave train with a unique meteorological origin and thus, that each partition is uncorrelated. The analysis mean wave integrated parameters for each partition can then be calculated and each partition can be adjusted separately.

Voorrips et al. (1997) assimilated partitioned pitch and roll buoy spectra into a North Sea implementation of WAM. The framework for the assimilation was an optimal interpolation method and each partition was scaled in energy and shifted in frequency and direction to obtain the partitioned analysis energy, mean frequency and direction. Hasselmann et al. (1997) also used an optimal interpolation method and assimilated partitioned Atlantic ERS-1 satellite radar data into the WAM model.

To date, most assimilation into wave models has been concerned with global or ocean scale models; there has been little consideration of data assimilation into coastal regions until recently (Siddons, 2007; Siddons et al., 2009; Portilla, 2009; Sannasiraj and Goldstein, 2009). Data assimilation into a coastal model poses specific problems compared to a global model. Wave conditions in regional models vary on much shorter temporal and spatial scales and are sensitive to changes in bathymetry and sheltering from coastlines. The error covariances therefore need to be designed to reflect the complex coastlines and bathymetry of a region. It is important that deep water points are not strongly correlated with shallow water locations to avoid instabilities in the data assimilation results. Much like global models, an important source of error comes from the wind forcing but this will occur at different scales in a regional model and regional models may also expect errors from their boundary conditions.

Siddons et al. (2009) assimilated *Hs* and *Tz* data from an OSCR (Ocean Surface Current Radar) HF (High-Frequency) radar located off the East Coast of England into the SWAN model. He tested three different assimilation techniques; 3D-VAR, an ensemble optimal interpolation (ensemble-OI) and an ensemble Kalman Filter (ensemble-KF). The results showed some overall improvements for the 3D-VAR and ensemble-OI methods, however, results from the ensemble-KF method were inconsistent. Siddons et al. (2009) suggested that incorporating spatially correlated errors and removing biases could improve the performance of the data assimilation schemes. He also stressed the need to apply strict quality control to the HF radar data. Portilla (2009) assimilated data from a single buoy off the Belgium Continental Shelf into a near shore configuration of the WAM model. He assimilated mean parameters (*Hs* and *Tz*) using an optimal interpolation scheme and investigated some different methods for parameterising the gain matrix. The parameterisation of the gain matrix allows for information to be spread in way which is consistent with the wave conditions in the region, but makes it difficult to extend the method to multiple observations. Portilla (2009) found improvements to the scatter index and RMSE and showed that in moderate wind conditions the benefit of assimilation could last for several days. Portilla also discussed assimilation of partitioned data and highlighted that the main task for this application would be the specification of an effective partition cross-assignment scheme.

Sannasiraj and Goldstein (2009) also used the optimal interpolation method to assimilate buoy data into WAM. They considered the Arabian sea region and assimilated significant waveheights from 3 different buoys into their model. They found their method to be computationally efficient and noted that the root mean squared error in the analysis waveheights was reduced by 30–50% in their study.

This study considers the assimilation of both mean parameters and partitioned wave data from an HF radar into a Celtic Sea wave model. An optimal interpolation (OI) method is used to assimilate multiple HF radar observations and a technique based on the Quick Canadian (QC) covariance method (Polavarapu et al., 2005) is used to estimate the wave model background error correlations for the region. The background error covariances are parameterised using these correlation lengthscales and parameterisations based on the bathymetry and climatological conditions of the region. Unlike the studies of Portilla (2009) and Sannasiraj and Goldstein (2009), this study implements data assimilation of wave partitions in a coastal region and compares the results to a twin study which assimilates mean parameters of the whole spectrum.

2. Wave model and observations

2.1. Wavewatch III

The model used in this study is Wavewatch III version 2.22 (hereafter WW3). It is a third generation spectral wind-wave model which solves the action balance equation. The full details of the numerical expressions used in WW3 are provided in Tolman (2002), Booij et al. (1999) and Ris et al. (1999). WW3 uses an explicit numerical scheme and in this study the Tolman and Chalikov (1996) combined input and dissipation source term is applied, the Hasselmann et al. (1973) empirical JONSWAP model is used for the bottom friction and the discrete interaction approximation (DIA) of Hasselmann and Hasselmann (1985) is used to model the quadratic non-linear wave-wave interactions.

For this study the WW3 model was run for the Celtic Sea region using NGDC GEODAS bathymetry data from 9 W to 4 W, 50 N to 55 N, with a resolution of $\frac{1^o}{30}$ (see Fig. 1). WW3 was forced with hourly 12 km resolution analysis winds from the UK Met Office atmospheric model and 12 km current and water level fields from the POLCOMS shelf sea model. The Celtic Sea model was nested within a $\frac{1^o}{6}$ North East Atlantic model, which in turn was nested in a lower resolution North Atlantic model. The North Atlantic model was run from 15/12/2004 and provided hourly boundary conditions for the North East Atlantic model which was run from 11/01/2005. The Celtic Sea model was initialised from zero at 11/01/2005 (the first 2 days were considered as spin up) and was forced with hourly boundary conditions from the North East Atlantic model. The WW3 model used spectra with 25 frequencies

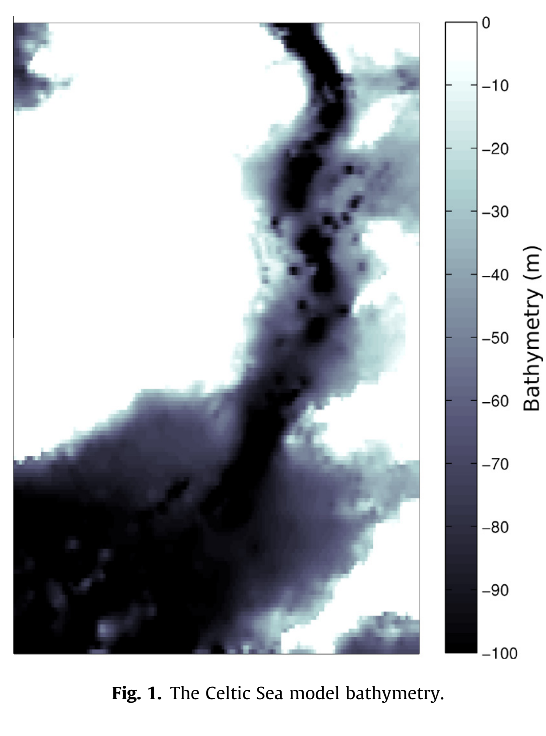

Fig. 1. The Celtic Sea model bathymetry.

covering a range from 0.041–0.404 Hz, and 24 equally spaced directions. For the WW3 Celtic Sea Model a global time step of 1 h was used along with a Courant Friedrichs Lewy (CFL) time step of 60 s. The model produced spectral and field outputs at hourly intervals.

2.2. Observation network

HF radars are remote sensing tools which are capable of measuring waves, currents and winds from electromagnetic backscatter from the ocean surface. Two radars are required to derive current vectors and perform a full directional wave inversion (Wyatt, 1987). They produce good spatial and temporal coverage in coastal regions and are land based so are easy to access and maintain. A dual Pisces radar system was trialled in the Celtic sea between South Wales and North Devon between December 2003 and June 2005. The two radars were located at Nabor Point in North Devon and Castlemartin in South Wales and each had 3 beam directions giving a total of 9 intersection points where directional data is available. Fig. 2 shows the radar sites and cell locations. The radar operated for around 19 min at each beam location. This provided 3 bin locations where data was measured simultaneously, data for the other 6 points were combined using an assumption of wave field stationarity over the timescale of an hour to give hourly coverage for the whole region. The range resolution of the radar was 15 km and the radar was operational at various frequencies between 5–11 MHz. A report on the Celtic Sea Pisces radar by Wyatt et al. (2006) found that the radar data was available for 96% of Nabor Point observations and 97% of Castlemartin observations and that dual radar wave data was consistently available 50–60% of the time. This does not include observations at radar bin 3 where the signal to noise ratio was persistently low with only 16% dual radar wave data availability.

There are also various buoy locations marked on Fig. 2 which are independent observations used for validation. All of these are non-directional buoys with the exception of the Lundy Waverider buoy which is collocated with radar bin 4 and the St Ives buoy. The Lundy buoy produces the frequency spectrum at 64 frequencies between 0.024 Hz and 0.58 Hz.

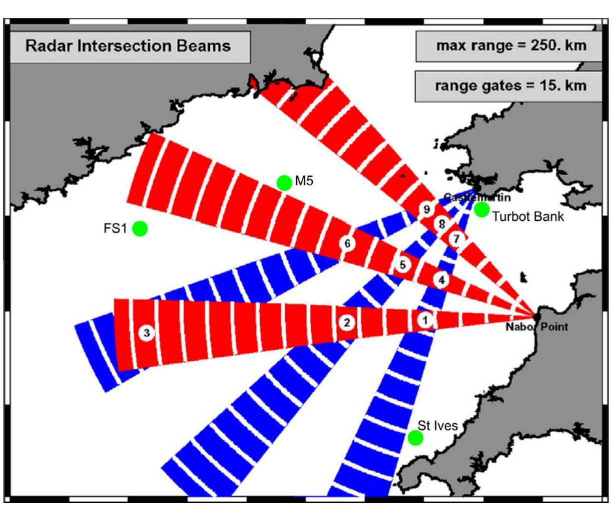

Fig. 2. The Pisces Celtic Sea HF radar. The two radar locations are marked along with the 3 beam directions and 9 dual radar bins. The St Ives, Turbot Bank, M5 and FS1 buoys are marked by circles on the map and the Lundy directional waverider buoy was collocated at radar bin 4.

2.3. Experiment set up

The WW3 Celtic Sea model was run for a 36 day period (plus 2 days of spin up) from the 13/01/2005. Fig. 3 shows the model wind direction and speed during this period. The winds are presented in oceanographic convention (waves travelling towards the direction). The winds travel in a variety of directions during the period, although the predominant wind conditions are winds travelling towards the North-Northeast with a mean wind speed of 9 m/s. There are 2 stormy periods identified in the winds, a period between 16/01/2005 and 20/01/2005 which coincides with a period of no Lundy buoy data and a period between 10/02/2005 and 15/02/2005. This study trials the assimilation of partitioned and mean parameter HF radar data from the Pisces HF radar into the Wavewatch III model for a 7 day period between 22:00 08/02/05 and 22:00 15/02/05. This period was chosen because of the stormy conditions which caused high *Hs* and ensured higher quality HF radar data (see Wyatt et al., 2006 and Section 2.4). There was also good continuous availability of buoy data for validation during this period. Data was assimilated from 8 radar bin locations (radar bin 3 was excluded for the reasons given above). Fig. 3 also shows a close up plot of winds for the trial period. During the period the mean wind speed is 11.5 m/s and the winds are turning from a North–East direction to a South direction. The plot of bathymetry in Fig. 1 shows the shape of the coastline in the region of interest. The region is sheltered to the North–West and East, and therefore typical swell systems in the region will be travelling East to North–East.

2.4. Comparison of observations and model during the experiment period

Fig. 4 shows time series of the HF radar bin 4, Lundy buoy and collocated WW3 data over the 36 day WW3 period. For *Hs*, the buoy, radar and model data compare well. Peaks in the *Hs* associated with the two stormy periods shown in Fig. 3 are identifiable in the data. WW3 has a tendency to underestimate the *Hs*, particularly during peak events. This is in agreement with results from Cavaleri (2009) who showed that wave models miss extreme events. The results for *Tz* are more variable. During periods of low *Hs* the radar *Tz* is very scattered. However, better correlations are observed in the radar data during the stormier periods. A detailed analysis of the HF radar data was provided by Wyatt et al. (2006), and this showed that the HF radar produces higher quality results in high sea states. The mean directions from the buoy, radar and WW3 are quite well correlated. Again, results are improved in the HF radar during high sea states.

3. Spectral partitioning

Spectral partitioning separates a wave spectrum into it's component wave trains. It is a useful technique which allows a spectrum to be represented with a reduced set of statistics without being averaged over the entire spectrum. It has practical applications for windsea and swell identification, swell tracking, noise removal, system validations and comparisons. It can also be used in data assimilation to develop more sophisticated schemes. It is common to assimilate mean parameters such as *Hs* and *Tz* into wave models but assimilating integral parameters from partitions allows for more information on directional characteristics and spectral shape to be included in the analysis spectrum in a physically meaningful way.

The spectral partitioning technique used in this study is a steepest ascent method applied in 8 directions, see Hasselmann et al. (1996), Voorrips et al. (1997) and Hanson and Phillips (2000). It is necessary to apply some post processing of the partitioned

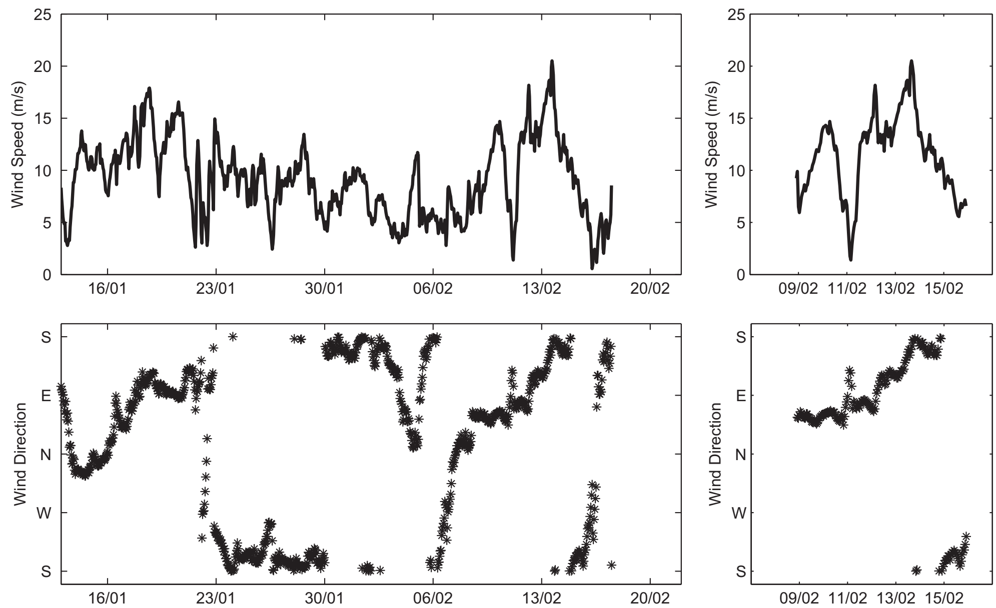

Fig. 3. Model wind speed and direction at the radar bin for location.

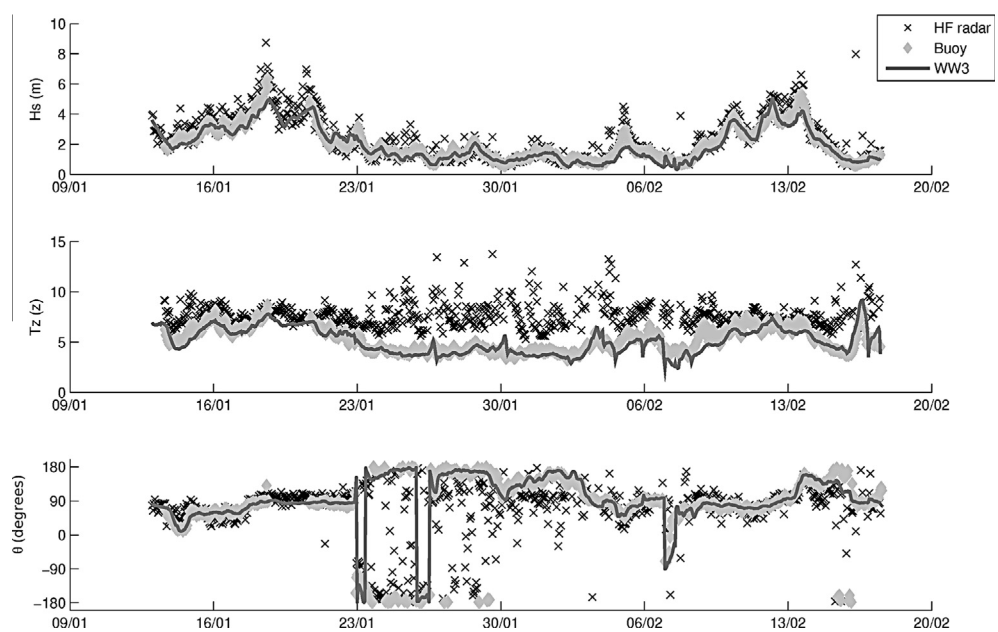

Fig. 4. Time series of data from the HF radar, Lundy buoy and WW3. The top plot is *Hs*, the middle plot is *Tz* and the bottom plot is mean direction.

spectrum to remove and/or combine spurious partitions. In this study the following criteria for combining and discarding partitions are used (based on those of Hasselmann et al., 1997; Hanson and Phillips, 2000):

1. *The peaks are too close* – the partitions should be combined if their peaks are only a grid point apart
2. *The trough separating the peak is not low enough* – the partitions should be combined if the lowest point between the two peaks is higher than A% of the lower of the two peaks.
3. *The spread is larger than the square distance between the peaks* – the partitions should be combined if $B\overline{\delta f^2} > \Delta f^2$ for either of the partitions. See the appendix for definitions of $\overline{\delta f^2}$ and Δf^2.
4. *The partition's energy is too low* – discard partitions with energy below 1% of the total energy. This eliminates any peaks in the noise floor.

For the HF radar spectra, $A = 70$ and $B = 0.5$, for the WW3 spectra $A = 75$ and condition 3 is not applied. These parameters were chosen based on the characteristics of the spectra and by consideration of the results produced and number of partitions determined from hourly spectral data over a 36 day period. Further details of how these criterion have been chosen see Waters (2010). The key findings were that the HF radar produces noisy spectra with a high noise floor and the parameters are tuned to deal with these characteristics. Meanwhile WW3 spectra have a tendency to be smoother than observed spectra with less spurious information and thus condition 3 is not applied and A is set at a larger value to prevent over-combing of wave trains.

4. Data assimilation method

OI was first introduced in the 1960s by Gandin (1965). The analysis x_a is calculated with the following equation

$$\mathbf{x}^a = \mathbf{x}^b + BH^T(HBH^T + R)^{-1}[\mathbf{y}^o - H\mathbf{x}^b] \quad (1)$$

where x^b is the background state and y^o is the observed state. The matrix B is the background error covariances, R is the observation error covariances, H is the linearised observation operator which maps the model onto observation space.

4.1. Error covariances

One of the fundamental challenges of the OI scheme is the specification of the error covariances. In general the observation errors are assumed to be spatially uncorrelated which reduces R to a diagonal matrix of variances. In order to maximise the benefit of a data assimilation scheme, it is necessary to define a spatially correlated B matrix. Defining the spatial correlations of the errors is a non-trivial task. In the majority of studies on wave data assimilation standard statistical correlation functions such as Gaussian functions have been used to parameterise the background error covariances. Some authors have simply estimated an appropriate correlation lengthscale, for example Bender and Golwacki (1996) chose an arbitrary lengthscale of 350 km which corresponded to 6 grid points in their model. Others have applied techniques such as the Hollingsworth and Lonnberg method (Wittmann and Cummings, 2005) or forecast difference ensemble methods (Greenslade and Young, 2005) to estimate the lengthscales.

The above mentioned cases are examples where observations are being assimilated into regional or global models. In this study the model configuration is a high resolution coastal run where background error correlations are more difficult to accurately model. Assimilating into nearshore or coastal regions is problematic because it is difficult to define a background error covariance structure which can well represent a complex coastline and bathymetry. There are also particular issues in how deep and shallow water model points should be correlated. Initial test runs in this study found that if a simple isotropic function is used to generate the background error covariance matrix in a coastal region, problems can arise when assimilating into shallow water locations. For example, the analysis Hs correction can be larger than the model's waveheight at a particular location. Care thus needs to be taken to ensure that the analysis waveheights do not become negative and that changes to Tz are not unrealistic.

Portilla (2009) assimilated data from one buoy into a nearshore model run of WAM off the coast of Belgium. In order to deal with the complexities of assimilating into a coastal region he tested two techniques for parameterising the gain matrix. The first method used long term model estimates. For various mean parameters, correlations between the different model locations were calculated. The gain matrix was then generated as the product of the Hs, Tz, mean first moment period ($T1$), wind speed and peak period correlations and a bathymetry and distance factor. The bathymetry factor prevented assimilation in shallow waters and the distance factor was a Gaussian function dependent on distance of separation. The second method was a dynamic technique where the gain matrix was calculated at each time step using a similar structure but short-term model data. These techniques provide a correlation structure which better suits the shape of the coastline, however, the correlations are based on the spatial variation of the mean parameters, not on their errors. It should also be noted that although this technique uses the framework of an optimal interpolation method, the parameterisation of the gain matrix (the gain matrix is calculated directly rather than from the model and observation errors) reduces the technique to a successive correction method.

Sannasiraj et al. (2006) tested a data assimilation scheme where the background error covariance structure was calculated from an ensemble of model runs generated by an ensemble of windfields. In that study it was assumed that the model was perfect and that errors are due to errors in the input winds.

The methods for calculating the error covariance used in the experiments presented are now described.

4.1.1. Error variances

The variances for the HF radar and WW3 are estimated by taking the mean squared difference (MSD) with the Lundy buoy data. It should be noted that the variance is only equal to the MSD when there is no bias. The HF radar MSD is calculated using the radar bin 4 data and the same variance is assumed for all radar bin locations. The errors in the radar may vary depending on distance from the radar stations and location in the region. However, at this time there is only quantitative data for radar location bin 4 since this is the only location collocated with buoy data. In WW3 the MSD is also calculated from data at the radar bin 4 location and the background error covariance is then scaled to ensure that the variance is equal to that calculated from the data. In this study the MSD calculated over the data assimilation period is used rather than the MSD from a larger data set. This is because the period for the data assimilation has been chosen as a period where the radar performs well; if errors from a longer period were used they are likely to be much larger. The errors are presented in Table 1. The HF radar errors are larger than those from the model, but this does not mean that assimilating the HF radar can not be beneficial. The model and HF radar both have skill in different conditions.

4.1.2. Observation error and the observation operator matrix

In this study the HF radar errors are assumed to be spatially uncorrelated, this reduces the observation error covariance to a diagonal matrix with the variance along the diagonal. Since the same variance is assumed for all observation locations the observation error covariance matrix can be written as:

$$R = VI \quad (2)$$

where V is the HF radar variance (see Table 1), I is an m by m identity matrix and m is the number of observations (which is at most 8).

The observation operator matrix, H, is an n by m matrix, where n is the number of model points and m is the number of observations. In this case it is specified as a simple matrix consisting of ones and zeros. The ones correspond to model points where observations exist. In general, the observation locations are not exactly collocated with the model points, but to maintain the simplicity of H the nearest model point is matched to the observation location. The form of H means that BH^T reduces to an n by m matrix where each column consists of the error covariances of a model point collocated with a radar bin with all other model points in

Table 1
MSD for the HF Radar and WW3.

Parameter	Radar MSD	WW3 MSD
Hs (m)	0.198	0.148
Tz (s)	2.272	0.390
f_m (1/s)	0.001373	0.000293
θ_m (rads)	0.305	0.070

the domain (e.g. such as the data shown in Fig. 5). This reduces the cost of Eq. (1) since the largest matrix to compute and store is an n by m matrix.

4.1.3. Background error correlations

The method used for calculating the background error correlations in this study are based on the Quick Canadian covariance method. The Quick Canadian covariance method (QC) was implemented by Polavarapu et al. (2005) in the Canadian Middle Atmosphere Model data assimilation scheme. In their application, background error covariances were calculated from 6 h difference fields. Jackson et al. (2008) tested the QC method in a troposphere/stratosphere configuration of the Met Office assimilation system. The authors compared the covariances with those produced using a technique referred to as the NMC method (Parrish and Derber, 1992). They found that as well as being a much quicker method, the QC also produced covariances which were of similar quality (and in some cases better quality) than those from the NMC.

The version of the QC method in Eq. (3) was used to calculate the error covariances using *Hs* field data from WW3 outputted at 1 h intervals. The model data was from the 36 day WW3 run described in Section 2.1 and there were a total of 853 output times. Fig. 3 shows that the 36 day period covers a range of different wind conditions. In Eq. (3), $B_{i,j}$ is the background error covariance between model grid locations i and j, the subscripts t and $t+1$ refer to time and time plus 1 h, respectively and $\langle Hs_{t+1} - Hs_t \rangle$ denotes the mean of the 1 h *Hs* difference over the 36 day period:

$$B_{i,j} = \left\langle [(Hs_{t+1} - Hs_t) - \langle Hs_{t+1} - Hs_t \rangle]_i [(Hs_{t+1} - Hs_t) - \langle Hs_{t+1} - Hs_t \rangle]_j \right\rangle \quad (3)$$

Using the 1 h difference fields is consistent with the temporal resolution of the forcing fields and the high temporal variability of a coastal region.

The spatial background error covariance structure for each of the 8 HF radar locations calculated using Eq. (3) are shown in Fig. 5. The covariances are fairly Gaussian in their structures but small close to the coastline and in shallow waters. The correlations are strongest in the North–East direction and there are weaker correlations to the North of the region. This is related to the prevailing wind conditions and direction of swell in this region. The QC method does generate some negative covariances and these were set to zero. Although it is possible to have negative covariances, it is assumed that these regions will generally be far from the observation locations and thus these locations are decorrelated to avoid spurious results. Fig. 5 shows that regions with zero covariance are either close to the coast or far from the radar bins.

4.1.4. A SOAR model for the background error correlations

In this study, the covariances in Fig. 5 are used to determine correlation lengthscales and a parameterisation for the background covariance. Fig. 6 shows a plot of the covariances against distance of separation for all 8 locations. Functions can be fitted to the data to attempt to produce a model for the covariance. Gaussian and second-order autoregressive (SOAR) functions with both one correlation length scale and two length scales are shown on the plot. The SOAR function with two length scales appears to give the best fit to the shape of the data (by eye), this has the form:

$$P(r) = V_1\left(1 + \frac{r_{ij}}{L_1}\right)\exp\left(-\frac{r_{ij}}{L_1}\right) + V_2\left(1 + \frac{r_{ij}}{L_2}\right)\exp\left(-\frac{r_{ij}}{L_2}\right) \quad (4)$$

where $P(i,j)$ is the covariance of model points i and j, V denotes variance, L denotes correlation lengthscales and r_{ij} is the distance between model points i and j. A chi-squared method was used to find the best fit for the lengthscales. These were found to be $L_1 = 11.35$ km and $L_2 = 99.53$ km and this function is shown as the thick black line in Fig. 6. It is generally assumed that the errors within wave model outputs are highly dependent on errors in the windfields. It is therefore interesting that the lower correlation length scale 11.35 km is very close to the 12 km resolution of the Celtic sea winds. It is also possible that the higher length scale could be related to the resolution of the North Atlantic winds which is 70 km. The length scale is possibly larger than this value due to the fact that waves are dispersive and these lower resolution winds will be responsible for swell seen in the Celtic sea region. However, the results in Fig. 6 do appear quite scattered, this is partly due to the assumption of the function fitting that the spatial covariances

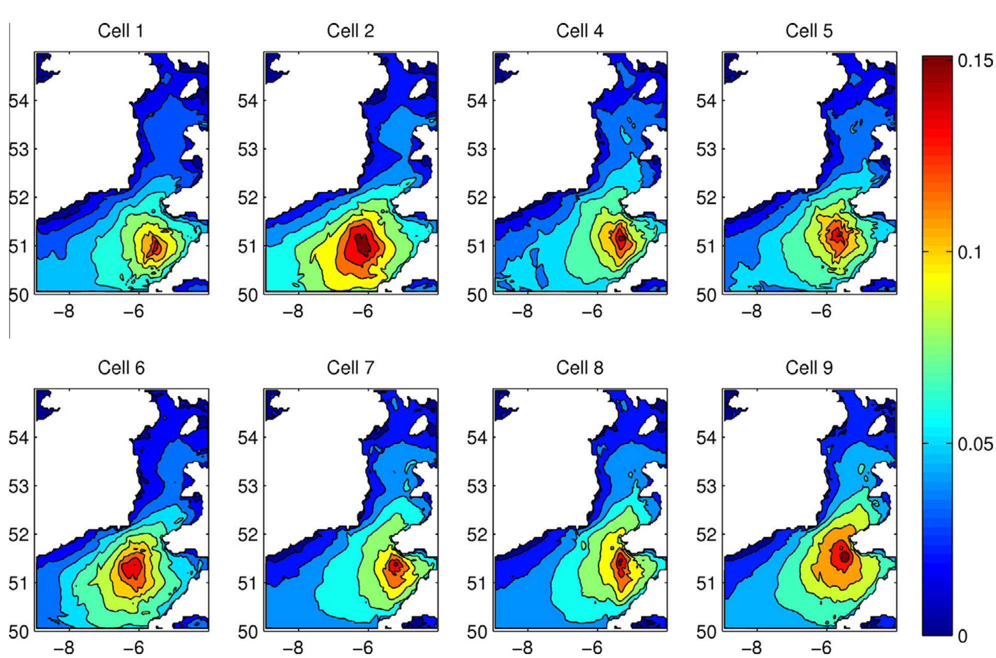

Fig. 5. The background error covariances calculated from the QC method. The covariances are valid at the 8 HF radar locations.

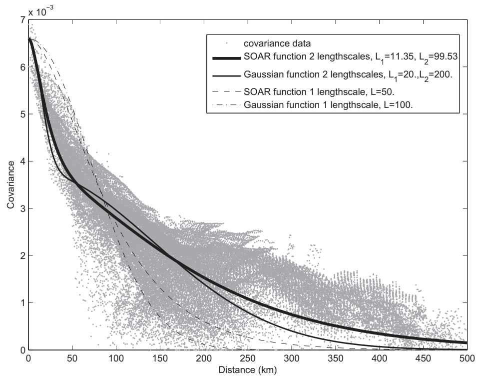

Fig. 6. The QC covariances plotted against distance of separation for all 8 radar bins. L refers to the correlation lengthscales.

are isotropic and homogenous throughout the region. This is unlikely to be true due to the shape of the coastline, but the assumption does provide an indication of the likely correlation length scales.

The correlation function in Eq. (4) is scaled to have a total variance equal to the MSD in Table 1. Thus, for *Hs* the chosen background error covariance is:

$$P_1(i,j) = 0.063\left(1+\frac{r_{ij}}{11.35}\right)\exp\left(-\frac{r_{ij}}{11.35}\right) + 0.085\left(1+\frac{r_{ij}}{99.53}\right)\exp\left(-\frac{r_{ij}}{99.53}\right) \quad (5)$$

Where $r_{i,j}$ is in km and $P_1(i,j)$ is in m^2. Eq. (5) models the spatial covariance structure, however, it does not take into account the variability in depth. Shallow water points are poorly correlated with the observation locations (which have depths between 54 and 104 m) and thus it is necessary to decorrelate observations in shallower water. In order to deal with this an exponential function of the following form is used:

$$P_2(i,j) = \begin{cases} 1 & d_j \geqslant 50\text{ m} \\ \exp\left[-([50-d_j]/23)^4\right] & d_j < 50\text{ m} \end{cases} \quad (6)$$

where $P_2(i,j)$ is a covariance between the model point *i* (assumed to be collocated with a radar bin), *j* is any other model point and d_j is the depth at the model point *j*. The depth threshold of 50 m was chosen as a value just below the depth of the shallowest observation (it is assumed that the depth does not have an impact at the observation locations). The power of four used in the function was selected to ensure a slow fall off in the correlations for depths between 30 and 50 m but correlations converging to zero at depths below 15 m. For model locations with depth below 15 m, assimilating information was found to cause unstable results. Waves in these shallow water locations will be governed by very different processes to those at the observation locations.

A test study assimilating mean parameters from the Lundy buoy using the QC background error covariance found that assimilating into the northern region of the Celtic Sea had a tendency to cause unrealistic results in the *Tz*. This is because this region is much more sheltered from swell and thus does not correlate well with observation locations. In fact, when the QC method is applied with *Tz* rather than *Hs*, shorter spatial correlation length scales are found, particularly to the north. However, the QC method applied with periods is much more sensitive to noise and gave a more spiky covariance structure and thus has not been used. Analysis of the Tz covariances and results from the buoy assimilation test suggest that the covariances need to be reduced North of 51.9 N. The following exponential function is used to gradually decorrelate the Northern part of the domain:

$$f(i,j) = \begin{cases} 1 & y_{(j)} \leqslant 51.9\text{N} \\ \exp[-3.6(51.9-y_{(j)})^2] & y_{(j)} > 51.9\text{N} \end{cases} \quad (7)$$

where $f(i,j)$ is a covariance between the model point *i* (assumed to be collocated with a radar bin), *j* is any other model point and y_j is the latitude at the model point *j*. Eqs. (5)–(7) are combined to provide a description of the error covariances at the model points collocated with radar bins and all other model points in the domain. This can be written as follows:

$$P(i,j) = P1(i,j)P2(i,j)f(i,j) \quad (8)$$

Furthermore, any covariances between the radar bin locations and the North West of Ireland and South of Cornwall are set to zero since these regions are separated from the observation locations by land. Finally, covariances values below 0.09% of the maximum covariance are set to zero to prevent inefficient data assimilation into areas with very low correlation. This value is chosen to have a small impact on the results but to improve the efficiency of the data assimilation. It is a rather arbitrary choice and could be subject to further tuning.

4.2. Data assimilation method

Two data assimilation schemes are tested: a mean parameter assimilation scheme and a partitioned integral parameter assimilation scheme. The methods are described in this section.

4.2.1. Mean parameter (MP) data assimilation

In the MP assimilation the parameters assimilated are *Hs* and *Tz* for the whole spectrum. The analysis mean parameters are calculated from the OI equation (see Eq. (1)). Since WW3 is a spectral wave model it is necessary to calculate an analysis wave spectrum from the analysis *Hs* and *Tz*. This is done by stretching the spectrum in the frequency domain and scaling its energy. The analysis spectrum is calculated from the background model spectrum using the following equation:

$$F^a(f,\theta) = \left(\frac{Hs^a}{Hs^b}\right)^2 \frac{Tz^a}{Tz^b} F^b\left(f\left\{\frac{Tz^a}{Tz^b}\right\},\theta\right) \quad (9)$$

Since the model wave spectra are discretised it is necessary to interpolate the spectrum in order to calculate the analysis. In this study a simple linear interpolation technique is used; it is not possible to use polynomial or spline interpolation techniques as they may allow negative spectral values in low energy regions of the spectrum. In order to deal with cases when the adjusted spectral frequency is outside the model's frequency range an exponential function of the following form is used:

$$F^a(f,\theta) = \left\{ \begin{array}{ll} F^b(f_{min},\theta)\exp(100[f-f_{min}]) & f < f_{min} \\ F^b(f_{max},\theta)\exp(100[f_{max}-f]) & f > f_{max} \end{array} \right\} \quad (10)$$

The exponential function creates a high and low frequency tail for the spectrum.

The radar data is assimilated into the nearest hour and when observations are at half past the hour they are assimilated at the hour preceding them (for example an observation at 01:30 is assimilated at 01:00). Since each radar has three beam locations throughout an hour, the time for a radar bin observation is calculated as the average of the two beams' start times. There is therefore enough temporal uncertainty in the observation time to assume that the observations are valid on the hour.

In order to prevent spurious results from the HF radar from being assimilated into WW3, various quality control thresholds are used. For the radar data to be assimilated it needs to satisfy the following criteria:

- The Radar *Hs* is greater than 1 m.
- The difference between the Radar and WW3 *Hs* is less than 50% of the WW3 *Hs*.
- The difference between the Radar and WW3 *Tz* is less than 50% of the WW3 *Tz*.

The top criterion is based on the findings of Wyatt et al. (2006), who showed that waveheights of 1 m are required to ensure reasonable quality HF radar data. The other criteria are chosen to prevent spurious radar data from being assimilated but to allow for reasonable differences in the data. These thresholds are quite restrictive but in general it is better to assimilate no data rather than assimilate very poor data. It may be necessary to further tune these thresholds in future applications.

4.2.2. Partitioned integral parameter (PIP) data assimilation

In the PIP data assimilation scheme integral parameters from the partitioned data are assimilated. The scheme is applied as follows:

1. Partition the WW3 and observation spectra using the partitioning scheme specified in Section 3. Any partitions in the WW3 spectra which fall below the energy thresholds are superimposed onto the analysis spectra.
2. Calculate and store the integral parameters for each partition. These are energy ($E = [4Hs]^2$), mean frequency $\left(\frac{1}{T_z}\right)$ and mean direction (θ).
3. Cross-assign partitions, this means finding which partitions correspond to the same wave train. Two cross-assignments are necessary: a cross-assignment between the partitions in the observation spectrum and WW3 spectrum at the observation location, and a cross-assignment between the cross-assigned partitions in the WW3 spectrum at the observation location and the partitions in all other WW3 spectra. At the model locations away from the observations, the partitions are cross-assigned with WW3 partitions rather than observation partitions because it is assumed that there will be more consistency between WW3 spectra. If a partition exists in a WW3 spectrum which can not be cross-assigned with a partition in the WW3 spectrum at the observation location, no alterations to that partition will occur.
4. Analysis integral parameters for each cross-assigned partition are calculated using the OI equation.
5. The cross-assigned WW3 partitions are adjusted to have the analysis integral parameter values.
6. WW3 partitions which can not be cross-assigned are simply superimposed onto the analysis wave spectrum. For observation partitions, the first step is to attempt to combine the partition with another observation partition which is cross-assigned. This is done by looking for partitions which satisfy $\frac{|\theta_i-\theta_j|^\circ}{90} + \frac{f_i-f_j}{0.5f_i} < 1$. If observation partitions are still non-assigned then they are superimposed with a scaled energy. The scaled energy is calculated from the OI equation assuming that the background state is zero so that Eq. (1) reduces to $E^a = BH^T(HBH^T + R)^{-1}E^\circ$.

The scheme is similar to the formulation used in Voorrips et al. (1997), however, the overall scheme used here differs in its treatment of non-assigned observation partitions and the cross-assignment of wave trains. Voorrips et al. (1997) required that two wave systems must be of the same type (both windsea or both swell) in order to be cross-assigned. However, an initial data assimilation test using a similar framework was performed and it was found that this condition is unsuitable. In this case all the swell was combined into one partition for simplicity. In many cases a wave system which is classified as windsea at one location, becomes swell at a distant location. Although the overall characteristics of the two wave systems are very similar, and to an observer the partitions would be classified as the same system, small differences in the mean frequency and direction or in the winds can cause a discrepancy. The result of this is that the analysis information is spread poorly across the region and large ridges in the *Hs* fields can arise over a boundary where the wave system's classification suddenly changes.

4.2.3. Cross-assignment 1

The criteria chosen for cross-assigning observation and WW3 partitions for co-located spectra are based on the criteria defined by Voorrips et al. (1997). Partitions are cross-assigned if

$$\frac{|\theta_i - \theta_j|}{50^\circ} + \frac{f_i - f_j}{0.4f_i} < 1 \tag{11}$$

and

$$\frac{E_{min}}{E_{max}} > 0.2 \tag{12}$$

Voorrips et al. (1997) used an energy threshold of 0.05 in his study. This has been increased to 0.2 as it was found that allowing less variability between the radar and model partitions is necessary to prevent noise from the HF radar being cross-assigned with WW3 partitions.

If for a particular observation (model) partition, more than one model (observation) partition satisfies this condition, the partitions with the smallest value of D_1 are cross-assigned, where

$$D_1(i,j) = \frac{|\theta_i - \theta_j|}{50^\circ} + \frac{f_i - f_j}{0.4f_i} + 0.2\frac{E_{max}}{E_{min}} \tag{13}$$

This is similar to the cross-assignment criterion but uses frequency, direction and energy together to determine how similar two partitions are. It is not always the case that the 2 closest partitions in spectral space correspond to the same wave system (for example, some partitions may be noise).

4.2.4. Cross-assignment 2

For cross-assigning WW3 partitions with WW3 partitions, the criteria are:

$$\frac{|\theta_i - \theta_j|}{50^\circ} + \frac{f_i - f_j}{0.4f_i} < 1 \tag{14}$$

and

$$\frac{E_{min}}{E_{max}} > 0.05 \tag{15}$$

If several partitions can be cross-assigned with one partition, the partitions with the smallest value of D_2 are cross-assigned, where

$$D_2(i,j) = \frac{|\theta_i - \theta_j|}{50^\circ} + \frac{f_i - f_j}{0.4f_i} + 0.05\frac{E_{max}}{E_{min}} \tag{16}$$

The prevalence of the energy weighting is reduced in this metric as at distances far from the observation location the energy of a corresponding wave train may vary quite significantly.

4.2.5. Adjusting the partitions

Once the analysis integral parameters for a partition have been calculated the model partition needs to be adjusted to become the analysis partition. A slightly different technique is used than in the MP assimilation scheme. In the MP assimilation method the *Tz* analysis was calculated and the frequencies were scaled to obtain

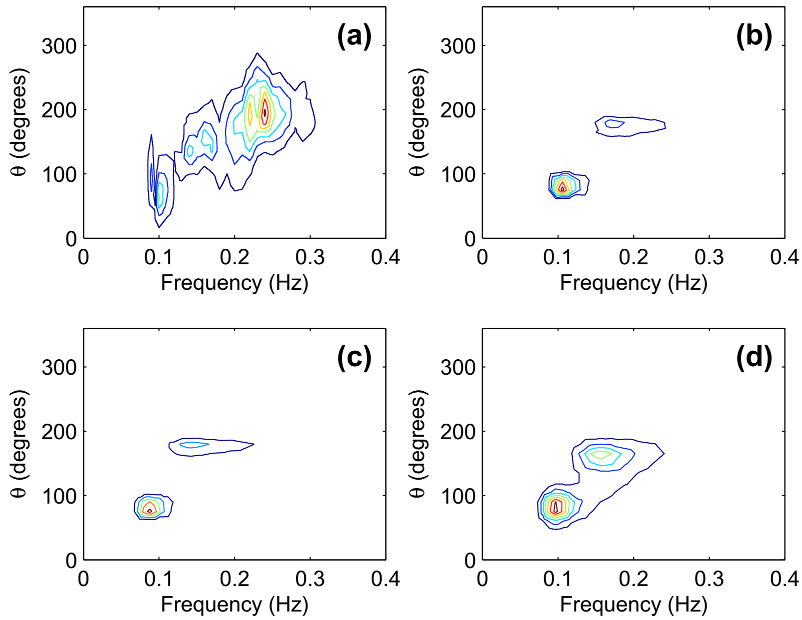

Fig. 7. Wave spectra in $m^2/Hz/degrees$ at 05:00 15/02/05. Plot (a) is the buoy spectrum, (b) is the WW3 spectrum without assimilation, (c) is the WW3 spectrum with MP assimilation, (d) is the WW3 spectrum with PIP assimilation. Maximum spectral density for (a) =0.73, (b) =3.79, (c) =2.67, (d) =1.49. The contours plotted are [0 0.005 0.01 0.05 0.1 0.2 0.4 0.6 0.8 1] times the maximum spectral density of (b).

the analysis period. In the PIP assimilation the mean frequency, $f_m = (1/Tz)$ is used instead, the analysis partition, $F_i^a(f,\theta)$ is calculated as

$$F_i^a(f,\theta) = \left(\frac{E_i^a}{E_i^b}\right) F_i^b(f + \Delta f, \theta + \Delta\theta) \qquad (17)$$

where $\Delta f = f_{mi}^b - f_{mi}^a$, $\Delta\theta = \theta_{mi}^b - \theta_{mi}^a$ and θ_m is the mean direction. Tests assimilating the Lundy buoy data found that using the scheme in Eq. (9) produces the best results for the MP assimilation while the scheme in Eq. (17) produces the best results for PIP assimilation. This is probably due to the fact that in MP assimilation, the whole spectrum is simply being adjusted to produce the analysis Tz and thus scaling in terms of the Tz produces the best results. However, in the PIP each partition is being moved in spectral space and it is thus more consistent to make adjustments in terms of a mean frequency. Note that in Eq. (17) the frequencies are now being shifted rather than scaled, this is so that adjustments in frequency and direction space are consistent. In the case where a negative analysis energy or mean frequency occur for a partition, the partitions analysis energy is assumed to be zero.

All 8 radar bins are assimilated in a PIP assimilation scheme. The same quality control thresholds used for the MP assimilation are applied and additionally if no partitions can be cross-assigned at the observation location then that observation is not assimilated at that time. The treatment of non-assigned partitions has proved to be more problematic when dealing with all 8 radar bins. Simply superimposing the data leads to significant over estimation of the waveheights and poor Tz results. This is probably due to cases where numerous erroneous partitions are superimposed. In order to deal with this problem a new condition was introduced: partitions can only be superimposed if they are considered persistent. When a non-assigned partition arises in a radar spectrum all other available radar locations are considered and if a similar partition occurs in at least one of the other locations the partition is considered persistent. Otherwise the partition is simply discarded. Two partitions are considered similar if they satisfy $\frac{|\theta_i-\theta_j|}{50^\circ} + \frac{f_i-f_j}{0.4 f_i} < 1$, this is based on the cross-assignment criterion used for cross-assigning WW3 and radar partitions.

4.3. Spectral results

In this section some spectral results from the assimilation will be presented. The plots will be analysed to ascertain if physically realistic wave spectra are constructed in the data assimilation scheme and whether the new spectra better correspond to the buoy results. All the spectral plots shown in this section are from the Lundy buoy location. The direction convention for the spectral plots is oceanographic and 0° is North, 90° is East, 180° is South and 270° is West.

In Fig. 7 there are three wavetrains visible in the buoy spectrum but only two in the original WW3 spectrum. The maximum spectral density in the original WW3 spectrum is also considerably larger than that in the buoy spectrum. The MP run produces limited change to the spectral shape from that produced in the original WW3 run, although the size of the spectral peak is reduced. In the PIP run there are more changes to the spectrum. The spectrum appears to contain the two lower frequency wavetrains seen in the buoy spectrum, but is missing the higher frequency wavetrain. This high frequency wavetrain is predominantly outside the frequency range of the HF radar, which observes a maximum frequency of 0.23 Hz at this time.

Fig. 8 is an example where there are two dominant distinct partitions in the original WW3 spectrum, travelling towards the East and South. In the buoy spectrum the energy is less disjointed with the higher frequency wave train having a lower frequency than that in the WW3 spectrum, and travelling in a South–East direction rather than a Southward direction. In the MP spectrum the spectral shape is surprisingly good, with the energy distribution between the two wave trains better than that in (b). The frequency of the higher frequency wavetrain is improved but the frequency of the lower frequency wavetrain is now too low and there is no improvement in the direction of the higher frequency partition. The PIP case gives a good spectral shape which now quite closely matches that of the buoy.

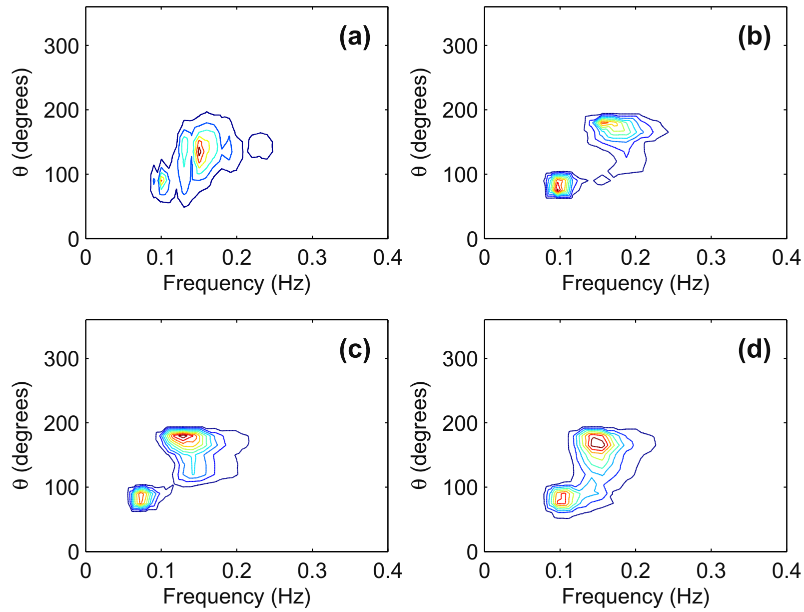

Fig. 8. Wave spectra in m^2/Hz/degrees at 15:00 14/02/05. Plot labels are the same as in Fig. 7. Maximum spectral density for (a) =3.27, (b) =5.17, (c) =5.38, (d) =4.39.

The spectral results for the data assimilation are quite good. It might be supposed that making adjustments to the spectral shape and superimposing observation partitions could cause irregularities in the spectra such as spurious energy and unrealistic results. However, this has been found not to be the case. The changes in the spectral shape in the MP case are generally quite small. This is to be expected as the only alterations really being made to the spectra are a scaling of energy and a shifting of all energy in the frequency domain. Where larger changes are present it is likely that these arise because the small changes in the wave spectra impact on how the spectrum has evolved over time. So, for example in Fig. 8(c) a change at a previous time may have improved the distribution of the energy of the two wavesystems at this time. The PIP assimilation allows more dramatic changes in the spectral shape. The results for PIP show some encouraging results, with the spectra often more closely representing the buoy spectrum. There are likely to be cases where results are not as good and this will be due to poor HF radar data, but in general the quality control thresholds are quite rigorous and ensure good overall results. Overall, the results for the PIP assimilation tend to be subtle and the spectral shapes produced are smooth and realistic.

5. Results

The results for the data assimilation runs are now presented. The *Hs* and *Tz* time series results will first be considered at five buoy locations: the Lundy buoy, St Ives buoy, M5 buoy, Turbot Bank and FS1 buoy. No spectral data were received for the M5, Turbot Bank and FS1 buoy and hence the buoy data in these cases can not be restricted to the WW3 frequency range. Thus at all five buoy locations a high frequency tail is added to the model data in order to extend the outputs to the buoys' frequency range and to maintain consistency throughout comparisons. Spectral and field results are also presented in order to give a thorough overview of the impact of the data assimilation and to highlight possible weaknesses.

The time series plots (Figs. 9 and 10) and statistics (Tables 2 and 3) are now presented for each of the five buoy locations. The root mean squared difference (RMSD) and correlation statistics with respect to the buoys' data are presented. The RMSD is considered the most important statistic within this study. The optimal interpolation technique seeks to minimise the analysis error variance and thus it is assumed that the RMSD gives the best indication of the success of the data assimilation scheme.

From Fig. 9 and Table 2 there is generally a good improvement in the *Hs*, for example the *Hs* RMSD is reduced by 35% in the PIP run compared to free run at the M5 location and the *Hs* correlations are improved at all locations by assimilation. The only location where the *Hs* RMSD is not improved is the St Ives buoy. The St Ives buoy output data at 25 min past the hour, every 2 h. This time lag and lower temporal resolution could be one reason for the mixed results seen at this location. Also, the St Ives buoy is one of the farthest from the radar bins and is in a sheltered region close to the coast which may make it poorly correlated with the radar locations.

The *Tz* results in Fig. 10 and Table 3 are more mixed. The *Tz* is consistently over-predicted at the Lundy buoy, St Ives and FS1 buoy locations in the assimilation runs. However, there are improvements in *Tz* for the Turbot bank and M5 buoy locations. At the Turbot bank location there is a 39% reduction in the MP assimilation run compared to the free run. The improvement at this location may be due to the close proximity of the Turbot bank to 3 radar bin locations. The Celtic Sea HF radar has a tendency to over-predict *Tz* while the WW3 free model run at the M5 and Turbot bank locations tends to under-predict the *Tz* during this period. These relative biases appear to have cancelling affect which results in a good improvement to the model *Tz* in the assimilation experiments. The correlations in Table 3 show an improvement in correlation in the assimilation experiments at all locations except the Lundy buoy. This implies that the HF radar assimilation is improving the modelling of variability in the *Tz* even if it is over-predicting the magnitudes. Overall it is not surprising that the *Tz* results are worse than the *Hs* results as it known that the quality of *Tz* data is poorer from the HF radar (see Table 1).

Comparing the *Tz* results from the MP and PIP assimilation cases shows that when the *Tz* results are poor in the MP assimilation case, they are comparatively much better in the PIP case. For example, in Fig. 10 the *Tz* is significantly over-predicted during the end of the period in the MP results for most of the locations.

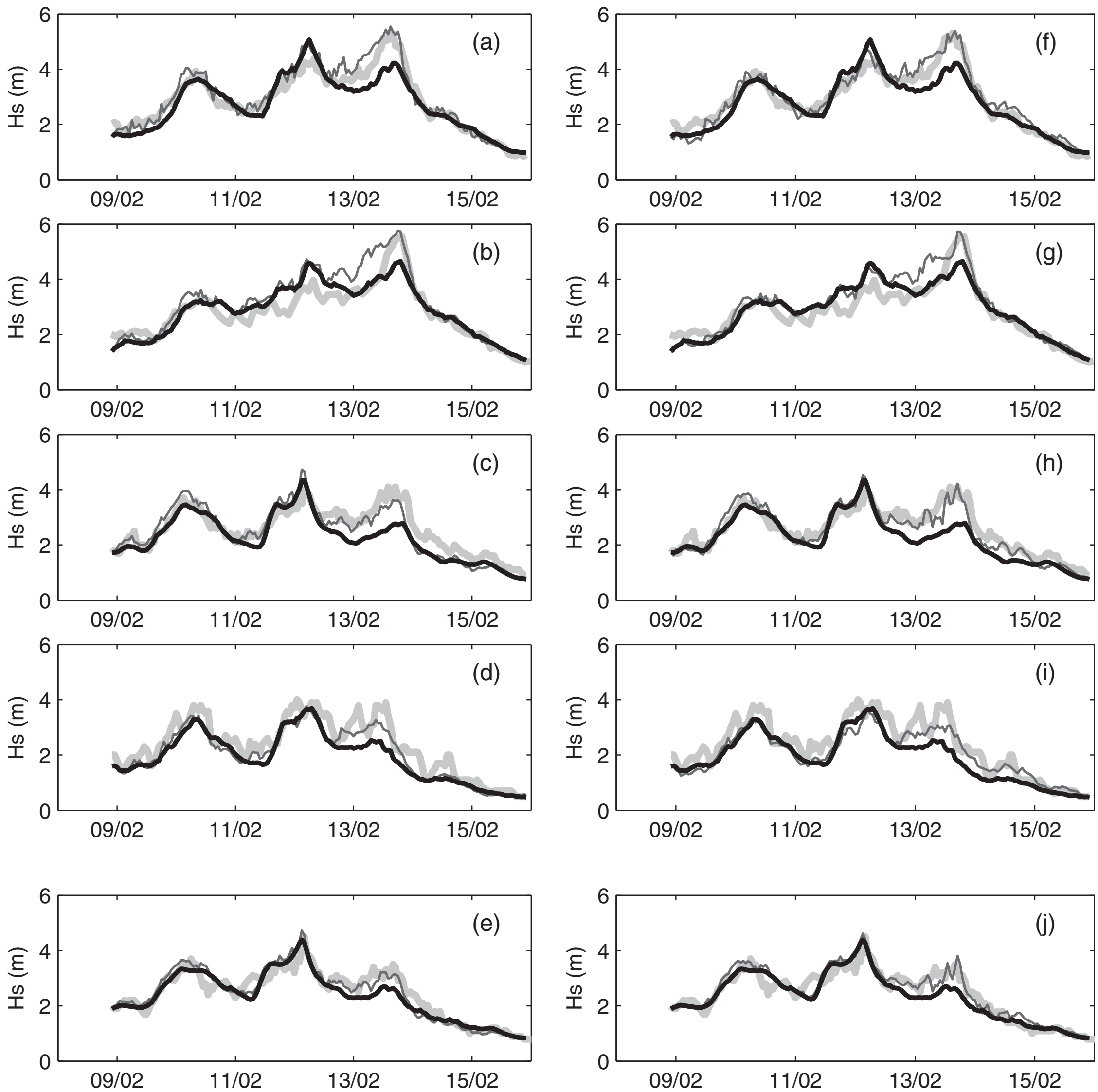

Fig. 9. The thick light grey line is the buoy *Hs*, the black line is the WW3 *Hs* without data assimilation and the thin dark grey line is WW3 *Hs* with assimilation. The plots are for different buoy locations, (a) and (f) are the Lundy buoy, (b) and (g) are the St Ives buoy, (c) and (h) are the M5 buoy, (d) and (i) are the Turbot Bank buoy, (e) and (j) are the FS1 buoy. The left hand plots are using the MP assimilation scheme while the right plots are using the PIP assimilation scheme.

However, the results during this period are much better in the PIP assimilation case. This is also seen in the statistics in Table 3 where the RMSD in significantly smaller for the PIP case compared to the MP case at the Lundy, St Ives and FS1 locations. This is because assimilating partitioned data introduces a new threshold level. Data is only assimilated when at least one partition can be matched. If no partitions are matched the spectral shape is considered too different and the data is not used. Also, one of the main causes of bias in the radar *Tz* is the radar's low frequency cut off, but assimilating partitions removes much of the impact of this. When assimilating the partitioned data, only the parts of the WW3 spectrum which correspond to partitions in the radar spectrum are adjusted. Since the high frequency cut off in the radar spectrum is generally lower than that in the WW3 spectrum, when assimilating partitions, only the region below the radar's maximum frequency will be adjusted. Furthermore, the condition within the PIP scheme which only allows non-assigned radar partitions which are considered spatially persistent to be superimposed into the analysis spectrum is also thought to have a positive impact. The technique allows for spurious radar partitions to be removed, hence reducing the error in the assimilation results. Finally, poorer results in the MP run compared to the PIP run may be related to the correlation lengthscales. When assimilating partitioned data, changes to partitions will only be applied for as long as the partition persists. Thus in some sense, assimilating partitions allows a natural length scale based on the persistence of partitions to exist. This can prevent unrealistic assimilation at locations far from the observations.

In general, the largest improvements in the assimilation cases for both the MP and PIP assimilation are seen during the peak storm period between 13/02 and 14/02 (see Fig. 3). There are potentially several reasons for a more significant improvement at this time. Since the HF radar performs best at high sea states, the quality of the HF radar data peaks during this event (see Fig. 4). Fig. 4 also shows that WW3 under-predicts the *Hs* during this extreme event. As well as the storm event there may be other reasons for the model to perform poorly during this period. The wind directions in Fig. 3 show that winds are turning throughout the storm

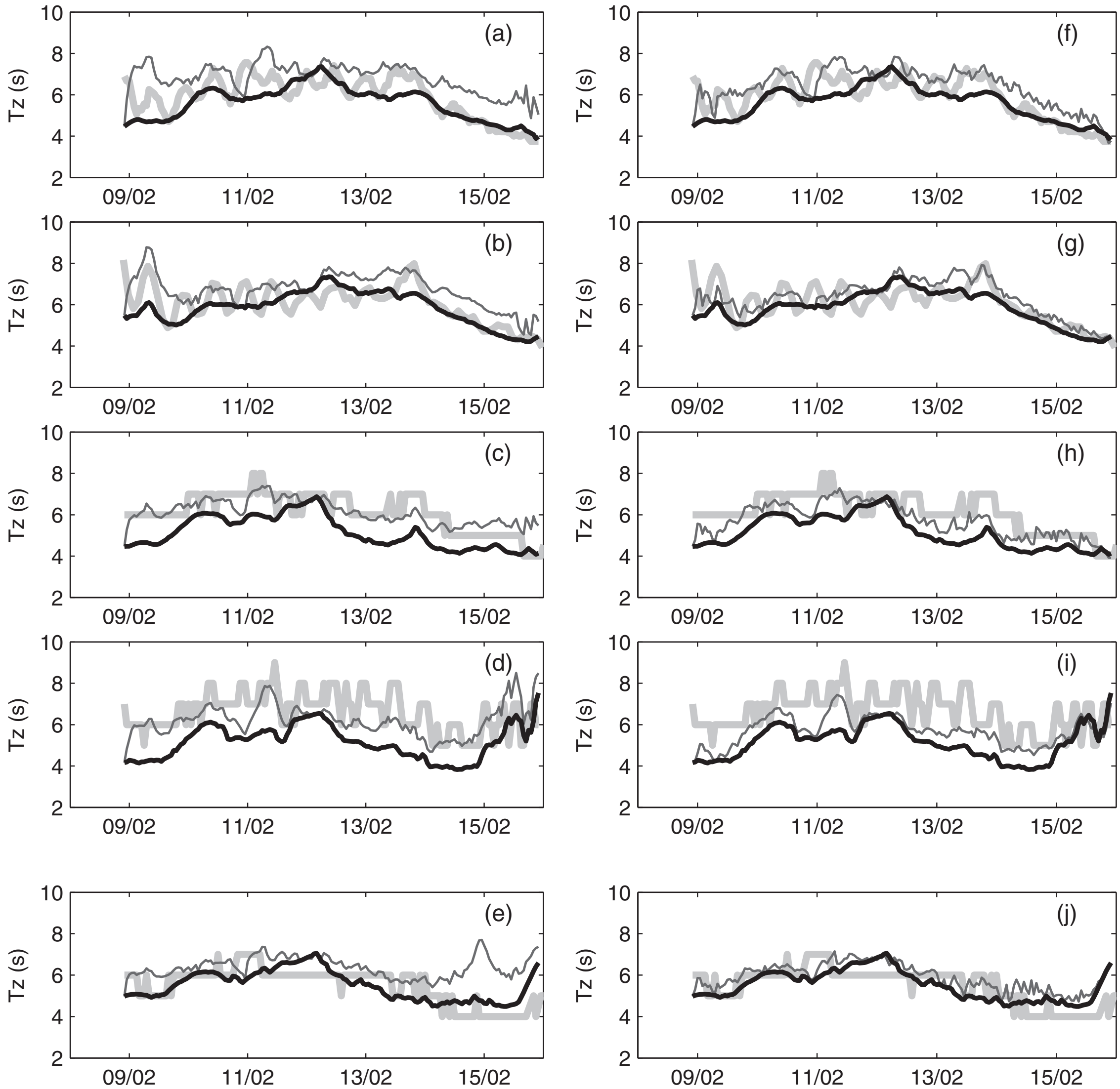

Fig. 10. Same as Fig. 9 but for *Tz*.

Table 2
Hs RMSD (correlation). The RMSD is in m.

Buoy location	No assim	MP	PIP
Lundy	0.39 (0.94)	0.31 (0.98)	0.29 (0.97)
St Ives	0.40 (0.91)	0.55 (0.94)	0.46 (0.94)
M5	0.51 (0.87)	0.38 (0.92)	0.33 (0.94)
Turbot Bank	0.55 (0.93)	0.40 (0.96)	0.43 (0.95)
FS1	0.36 (0.92)	0.30 (0.95)	0.30 (0.94)

Table 3
Tz RMSD (correlation). The RMSD is in s.

Buoy Location	No assim	MP	PIP
Lundy	0.71 (0.77)	1.16 (0.68)	0.76 (0.69)
St Ives	0.71 (0.69)	0.91 (0.70)	0.71 (0.70)
M5	1.17 (0.76)	0.63 (0.71)	0.70 (0.80)
Turbot Bank	1.70 (0.53)	1.03 (0.38)	1.19 (0.61)
FS1	0.68 (0.65)	1.30 (0.16)	0.76 (0.69)

period. This may indicate that the model does not correctly capture the wave conditions in the case of strong turning winds.

Considering the Lundy buoy results in Fig. 9, the PIP assimilation better captures the wave growth during the storm period than the MP assimilation. As the winds grow and turn, the PIP assimilation will allow for the different parts of the spectrum to be adjusted separately and for energy to be redistributed throughout the spectrum. In the case of a growing windsea this allows for the correction of spectral growth.

Fig. 11 shows spectra from the Lundy buoy and WW3 runs during the peak of the storm (12:00 13/02/05). The impact of the turning winds is visible in all spectra. There is old windsea travelling towards the East and new windsea travelling towards the South. In the buoy spectra there is a smooth transition between the new and old windsea. In the original WW3 spectrum the waves appear to have turned too fast and there is too much energy travelling towards the south. There is an improvement in the MP assimilation case with less energy travelling towards the South, but the best results are seen in the PIP assimilation case where the distribution of the energy in frequency and direction of the turning windsea best replicates that of the buoy spectra. This supports the suggestion that the PIP assimilation can provide better results in cases of turning winds.

In the majority of studies on data assimilation into wave models, in particular in those assimilating partitioned data, field plots

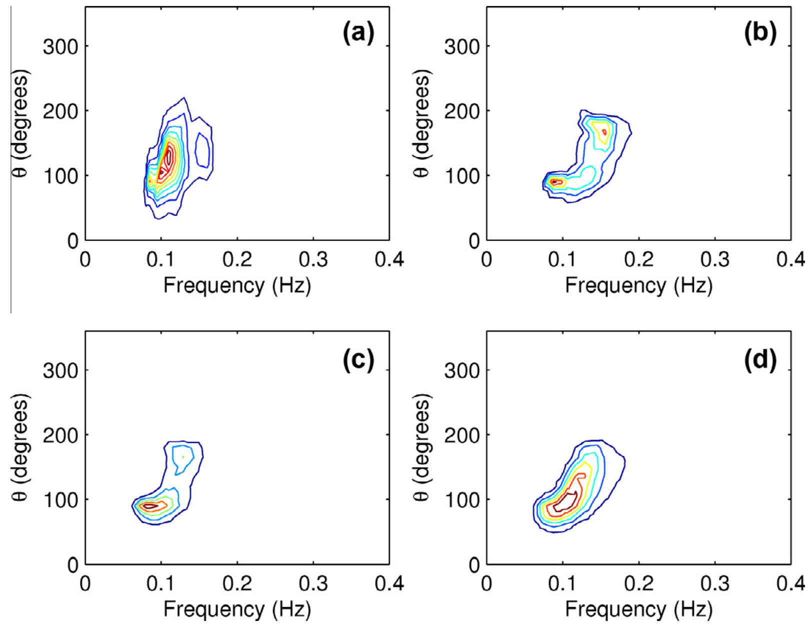

Fig. 11. Wave spectra in m^2/Hz/degrees at 12:00 13/02/05. Plot labels are the same as in Fig. 7. Maximum spectral density for (a) =19.23, (b) =12.80, (c) =28.00, (d) =14.00.

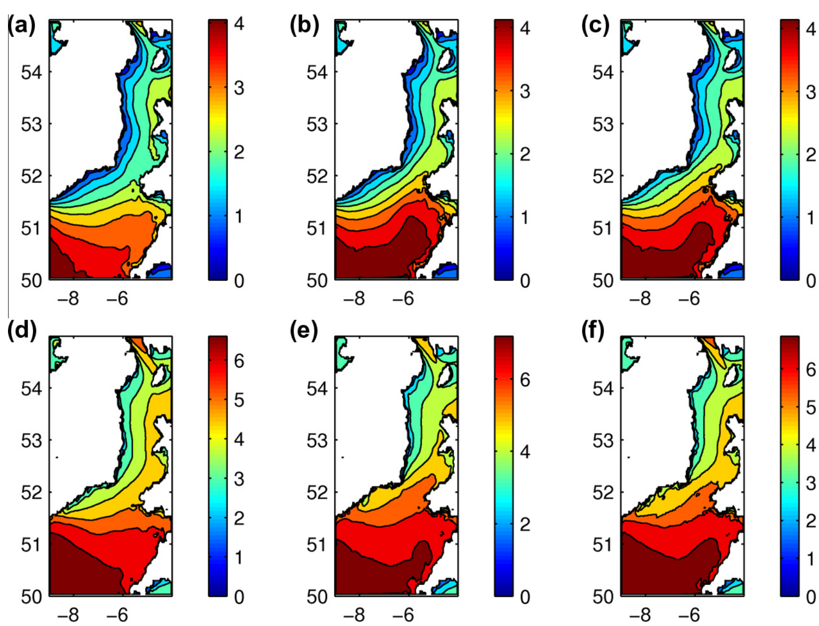

Fig. 12. Field plots at 00:00 13/02/05. The top plots are *Hs* fields (m), the bottom plots are *Tz* fields (s). Plot (a) and (d) are WW3 fields without assimilation, (b) and (e) are WW3 fields with MP assimilation, (c) and (f) are WW3 fields with PIP assimilation.

of mean parameters have not been published. One of the main difficulties in the data assimilation is preserving a realistic spatial structure and avoiding the introduction of noise or false variability. This is a particular problem with assimilating the partitioned data as the changes made to the spectrum depend on the partitions identified at each location and the cross-assignment. Earlier the difficulties of assimilating windsea and swell were discussed. It was found that using a cross-assignment based on the wavetrain's type causes discontinuities in the mean parameter field data. If the field data is not considered, such weaknesses may be overlooked. Considering the field data can also be useful for testing the background error covariance structure and ensuring that the changes are reasonable. However, since there are no observations which can give the spatial coverage and resolution of the model it is difficult to validate the field structures, but the field data is useful for identifying problems.

Fig. 12 shows plots of *Hs* and *Tz* from WW3 for the free run and both assimilation runs. At this time the results from the assimilation experiments show an increase in the *Hs* within the region of the observations. In the *Tz* plots there is an increase in the *Tz* over the South of the region. The fields from the assimilation runs are smooth and there is no indication of noise or discontinuities in the results.

6. Discussion

In this study HF radar wave data has been assimilated into the WW3 model. Two different data assimilation schemes have been described and tested, a mean parameter assimilation scheme and partitioned integral parameter assimilation scheme. A SOAR model for the background error covariances has also been described. This has been specifically designed to deal with complex bathymetry in this coastal region and has been localized for the Celtic Sea area. The results from the assimilation schemes have been validated against various independent buoys in the Celtic Sea region.

In general, both assimilation schemes have produced satisfactory improvements to the *Hs*. The overall improvement in *Hs* at the independent observation locations shows that the data assimilation is having a positive impact on *Hs* away from the assimilated observations locations. The results for the *Tz* are more mixed with improvements to the RMSD only seen at the M5 and Turbot Bank locations. There is a bias in the HF radar *Tz* data and in particular in the MP case this can cause the assimilated *Tz* to be over-predicted. This impact was found to be lessened by assimilating partitions rather than mean parameters. Assimilating partitions allows that only the region of the WW3 spectrum corresponding to the radar spectrum is adjusted. This removes some of the *Tz* bias caused by the HF radar's low cut off frequency. Also, assimilating partitions allows for an extra stage within the data quality control based on spectral shape. Observations are discarded when no partitions can be cross-assigned and thus less poor quality data is assimilated in the PIP run. Furthermore, the PIP run enables a natural correlation length scale based on the persistence of partitions and this prevents unrealistic data assimilation at locations far from the observations. This is particularly beneficial for a coastal domain where correlations are more difficult to accurately model. Finally, in the PIP run, only radar partitions considered spatially persistent are superimposed into the analysis spectrum. This process removes spurious partitions and hence noise from the spectrum.

The largest improvements in the data assimilation schemes are seen during a stormy period with turning winds. The assimilation runs produce the largest improvements to the Hs at this time and comparisons of the PIP and MP assimilation schemes suggest that the PIP assimilation performs better in these conditions. Overall, this study suggests that there are both practical and physical advantages to assimilating partitioned data rather than mean parameters when considering HF radar data. Many of these benefits are likely to extend to other observation types.

Both the assimilation of HF radar wave data and assimilation within coastal regions are relatively new and hence this study provides some preliminary results on the feasibility of such work. Overall there are some encouraging results from the data assimilation which suggest that assimilating partitioned data in a coastal region is feasible and beneficial for *Hs*. This study could therefore be extended to look at assimilating different coastal observation networks. The data assimilation scheme should be applicable in different coastal domains, although the design of the background error covariances is rather specific, so the geography of the region would need to be considered in its construction.

As this is a preliminary study, there are several areas which should be considered for future work. At present radar data is only assimilated when the *Hs* is greater than 1 m. From Wyatt et al. (2006) it was found that a 1 m threshold ensures good quality *Hs* data but a 2 m threshold is needed to produce good quality *Tz* and mean direction data, this might be a more appropriate threshold for future applications. Note that this is specific to this radar configuration, for a radar operating in a different frequency range alternative quality control thresholds would be required. Future work could focus on improving the specification of the background and observation errors by increasing the ensemble size used in the QC calculation or by considering more sophisticated techniques for calculating the background and observation error variances. One of the key advantages of assimilating partitioned data is the potential to also correct the wind forcing using the windsea data. This allows the impact of the data assimilation to be retained over a longer time period. This should be investigated further. This extension to the assimilation scheme would be particularly useful within the framework of a coupled atmosphere-wave model. The assimilation experiments should also be extended to consider different periods so that the impact of the data assimilation in different wave conditions can be analysed. The impact of the assimilation on forecasts should also be addressed in future work to assess how long the impact of data assimilation persists, and in what regions it has the most significant impact.

Acknowledgments

This work was undertaken as part of a Natural Environment Research Council [NE/E529041/1] funded PhD at the University of Sheffield. There was also funding from the Met Office as a CASE partner. The Pisces data was provided by Neptune Radar and collected during a project funded by DEFRA and the UK Met Office. The buoy data were provided by CEFAS and the Met Office.

Appendix A. Definition of spectral spread

The spectral spread $\overline{\delta f^2}$ of a partition is defined as,

$$\overline{\delta f^2} = \overline{(f_x - \overline{f_x})^2} + \overline{(f_y - \overline{f_y})^2} = \overline{f_x^2} + \overline{f_x}^2 + \overline{f_y^2} + \overline{f_y}^2$$

where

$$\overline{f_x} = \frac{1}{e}\iint S(f,\theta) f \cos\theta d\theta df$$
$$\overline{f_y} = \frac{1}{e}\iint S(f,\theta) f \sin\theta d\theta df$$
$$\overline{f_x^2} = \frac{1}{e}\iint S(f,\theta) f^2 \cos^2\theta d\theta df$$
$$\overline{f_y^2} = \frac{1}{e}\iint S(f,\theta) f^2 \sin^2\theta d\theta df$$

where f, θ and $S(f,\theta)$ are frequency, direction and the frequency-direction spectrum respectively and e is total spectral energy. If two partitions have peaks located at (f_1,θ_1) and (f_1,θ_1), then the distance between the peaks Δf^2 is determined by,

$$\Delta f^2 = (f_1 \cos\theta_1 - f_2 \cos\theta_2)^2 + (f_1 \sin\theta_1 - f_2 \sin\theta_2)^2.$$

References

Bauer, E., Hasselmann, S., Hasselmann, K., Graber, H., 1992. Validation and assimilation of SEASAT altimeter wave heights using the WAM wave model. J. Geophys. Res. 97, 12671–12682.

Bender, L., Golwacki, T., 1996. The assimilation of altimeter data into the Australian wave model. Aust. Meteorol. Mag. 45, 41–48.

Booij, N., Ris, R.C., Holthuijsen, L.H., 1999. A third generation wave model for coastal regions. 1: Model description and validation. J. Geophys. Res. 104, 7649–7666.

Cavaleri, L., 2009. Wave modeling – missing the peaks. Am. Meteorol. Soc. 39, 2757–2778.

Esteva, D.C., 1988. Evaluation of preliminary experiments assimilating SEASAT significant wave heights into a spectral wave model. J. Geophys. Res. 15, 14099–14105.

Foreman, S.J., Holt, M.W., Kelsall, S., 1994. Preliminary assessment and use of ERS-1 altimeter wave data. J. Atmos. Ocean. Technol. 11, 1370–1380.

Francis, P., Stratton, R.A., 1990. Some experiments to investigate the assimilation of SEASAT altimeter wave height data into a global wave model. Q. J. R. Meteorol. Soc 116, 1225–1251.

Gandin, L., 1965. Objective analysis of meteorological fields. Israel Program for Scientific, Translation, p. 242.

Greenslade, D.J.M., Young, I.R., 2005. Forecast divergence of a global wave model. Mon. Weather Rev. 133, 2148–2162.
Hanson, J.L., Phillips, O.M., 2000. Automated analysis of ocean surface directional wave spectra. J. Atmos. Ocean. Technol. 18, 277–293.
Hasselmann, K., Barnett, T.P., Bouws, E., Carlson, H., Cartwright, D.E., K. Enke, J.E., Gienapp, H., Hasselmann, D., Kruseman, P., Meerburg, A., Mller, P., Olbers, D.J., Richter, K., Sell, W., Walden, H., 1973. Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). Technical Report 8(12), Deutsche Hydrographic Institute.
Hasselmann, S., Bruning, C., Hasselmann, K., Heimbach, P., 1996. An improved algorithm for the retrieval of ocean wave spectra from synthetic aperture radar image spectra. J. Geophys. Res. 101, 16615–16629.
Hasselmann, S., Hasselmann, K., 1985. Computations and parametrizations of the nonlinear energy transfer in a gravity-wave spectrum. Part 1: A new method for efficient computations of the exact nonlinear transfer integral. J. Phys. Oceanogr. 15, 1369–1377.
Hasselmann, S., Lionello, P., Hasselmann, K., 1997. An optimal interpolation scheme for the assimilation of spectral wave data. J. Geophys. Res. 102, 15823–15836.
Jackson, D.R., Keil, M., Devenish, B., 2008. Use of canadian quick covariances in the Met Office data assimilation system. Q. J. R. Meteorol. Soc. 134, 1567–1582.
Lionello, P., Gunther, H., Janssen, P.A.E.M., 1992. Assimilation of altimeter data in a global third-generation wave model. J. Geophys. Res. 97, 14453–14474.
Parrish, D.F., Derber, J.C., 1992. The National Meteorological Center's spectral statistical-interpolation analysis system. Am. Meteorol. Soc. 120, 1747–1763.
Polavarapu, S., Ren, S., Rochon, Y., Sankey, D., EK, N., Koshyk, J., Tarasick, D., 2005. Data assimilation with the Canadian middle atmosphere model. Atmos–Ocean 43, 77–100.
Portilla, J., 2009. Buoy data assimilation in nearshore wave modelling. PhD Dissertation, Katholieke Universiteit Leuven.
Ris, R.C., Holthuijsen, L.H., Booij, N., 1999. A third generation wave model for coastal regions. 1: Verification. J. Geophys. Res. 104, 7667–7681.
Sannasiraj, S.A., Babovic, V., Chan, E.S., 2006. wave data assimilation using ensemble error covariances for operational wave forecast. Ocean Modell. 14, 102–121.
Sannasiraj, S.A., Goldstein, M.G., 2009. Optimal interpolation of buoy data into a deterministic wind-wave model. Nat. Hazards. 49, 261–274.
Siddons, L.A., 2007. Data assimilation of HF radar data into coastal wave models. PhD Dissertation, The University of Sheffield.
Siddons, L.A., Wyatt, L.R., Wolf, J., 2009. Assimilation of HF radar data into the SWAN wave model. J. Mar. Syst. 77, 312–324.
Thomas, J., 1988. Retrieval of energy spectra from measured data for assimilation into a wave model. Q J. R. Meteorol. Soc. 114, 781–800.
Tolman, H.L., 2002. User manual and system documentation of Wavewatch-III. NOAA/NWS/NCEP/MMAB Technical Note, 2.22 edition.
Tolman, H.L., Chalikov, D.V., 1996. Source terms in a third-generation wind-wave model. J. Phys. Oceanogr. 26, 2497–2518.
Voorrips, A.C., Makin, V.K., Hasselmann, S., 1997. Assimilation of wave spectra from pitch and roll buoys in a North Sea wave model. J. Geophys. Res. 102, 5829–5849.
Waters, J., 2010. Data assimilation of partitioned HF radar wave data into wave models. Ph.D. Thesis, The University of Sheffield.
Wittmann, P.A., Cummings, J.A., 2005. Assimilation of altimeter wave measurements into Wavewatch III. Oceanography division, Naval Research Laboratory, Stennis Space Centre.
Wyatt, L., 1987. Ocean wave parameter measurement using a dual-radar system. a simulation study. Int. J. Remote Sens. 8, 881–891.
Wyatt, L., Green, J.J., Middleditch, A., Moorhead, M.D., Howarth, J., Holt, M., Keogh, S., 2006. Operational wave, current and wind measurements with the Pisces HF radar. EEE J. Ocean. Eng. 31, 819–834.