AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2015JC011309

On the specification of background errors for wave data assimilation systems

Jesús Portilla-Yandún[1] and Luigi Cavaleri[2]

[1]Escuela Politecnica Nacional, Quito, Ecuador, [2]Institute of Marine Sciences, ISMAR-CNR, Venice, Italy

Key Points:

- A methodology for computing Background Errors in spectral wave data assimilation is presented
- A stochastic solution is offered for the cross assignment of model and observed spectra
- A data assimilation algorithm compatible with these new developments is proposed

Correspondence to:
J. Portilla-Yandún,
jportilla@ymail.com

Citation:
Portilla-Yandún, J., and L. Cavaleri (2016), On the specification of background errors for wave data assimilation systems, *J. Geophys. Res. Oceans*, *121*, 209–223, doi:10.1002/2015JC011309.

Received 9 SEP 2015
Accepted 6 DEC 2015
Accepted article online 13 DEC 2015
Published online 10 JAN 2016

Abstract In this paper, a new methodology is proposed for the computation of Background Errors in wave data assimilation systems. Background errors define the spatial influence of an observation in the model domain. Since at present the directional wave spectrum is the fundamental variable of both state-of-the-art numerical models and most modern instrumentation, this is at the core of the proposed methodology. The advantage of the spectral approach is that the wave spectrum contains detailed information of the different wave systems and physical processes at work (e.g., wind-sea or swells). These systems have different origins and may be driven by different mechanisms, having therefore different spatial structures, length scales, and sensitivity to local wind conditions. The presented method enables making consistent and specific corrections to each component of the spectrum, in time and space. The innovations presented here require an integral look at the data assimilation algorithm for which a suitable scheme is also proposed. Examples of computed background errors are presented for shelf and oceanic basins showing the spatial structures of the different wave systems active in these areas.

1. Introduction

A strong limitation of present wave data assimilation systems is that quite often they consider observed integral parameters to correct the modeled wave spectrum [e.g., *Lionello et al.*, 1992; *Chen et al.*, 2004; *Wittmann and Cummings*, 2005; *Greenslade and Young*, 2005; *Sannasiraj and Goldstein*, 2009; *Dee et al.*, 2011]. This practice has a major drawback because wave conditions are typically complex, involving locally generated waves (wind-sea) and one or more swell systems. In such conditions, integral parameters provide a poor representation of the actual sea truth. For instance, a specific wave height may correspond to many totally different combinations of wave conditions. Therefore, trying to correct a complex sea state with a limited set of observed parameters is obviously risky, certainly too ambitious. Moreover at present several, if not most, wave measuring devices aim at obtaining the directional wave spectrum, and there is a steadily growing amount of such information worldwide [e.g., *National Data Buoy Center*, 2015; *Data Buoy Cooperation Panel*, 2015; *Rijkswaterstaat*, 2015; *Marine Automatic Weather Station Network*, 2015; *The Coastal Data Information Program*, 2015]. In this context, considering the modern modeling and observing technologies, it is a most natural step forward for wave data assimilation systems to put our attention on the whole spectrum, i.e., to move to a fully spectral approach. In this paper, we address some of the key challenges involved in the development of such a scheme, focusing primarily on the computation of a crucial element, the background error covariances.

Most of the present operational wave data assimilation systems (e.g., European Centre for Medium-Range Weather Forecasts, henceforth ECMWF, Meteo France, Australian Bureau of Meteorology) are based on the scheme of *Lionello et al*. [1992], which was designed to assimilate altimeter data (i.e., significant wave height only, *Hs*). The just mentioned limitations were clear since those initial developments, and Lionello et al. thoroughly dealt with some of the related issues. A crucial one is that *Hs* is not the model variable, but just one bulk parameter obtained from integration of the spectrum. Therefore an ad hoc function is needed to transfer the *Hs* corrections to the wave spectrum. However, from information of *Hs* only, the most straightforward and only valid solution for correcting the spectrum is an uniformly, or proportionally distributed, update of energy over the whole frequency-direction domain. The limitation of such a solution follows directly, because in the wave growth mechanism there is a strong dependence between wave height and peak period (*Tp*), in such a way that for an inconsistent balance between wind speed, wave height, and peak period the model quickly overrides the corrections and resumes its former behavior. To overcome this issue, if we are dealing with wind-sea, any update of *Hs* must involve a corresponding update of *Tp*. A similar, but more relaxed, relationship exists for each swell system where the limit is given by the wave steepness. Therefore the update of *Tp* must in turn consider the stage of wave development, i.e., the wave age. Correspondingly, *Lionello et al*. [1992] consider two *Hs-Tp* correction strategies, for wind-sea and swell, respectively. However, they did not split the spectrum, because from the altimeter measurements there is not enough information to properly distribute the correction among the single spectral components. In their scheme, they check the growth stage of the predominant wave system and apply the corresponding wind-sea or swell correction to the whole spectrum. In reality in the ocean there is a whole bunch of intermediate situations between wind-sea and swell.

With the advent of more sophisticated and remote sensing instruments, namely directional buoys and the *Synthetic Aperture Radar* (SAR), it was apparent that many limitations related to the altimeter would be overcome, because these instruments provide estimates of directional wave spectra [*Hasselmann et al*., 1996, 1997]. However, other shortcomings affect the 2-D measured spectral data, some of them addressed in more detail in the discussion. What is relevant for the present approach is that one cannot expect two sources of spectral data to match bin by bin with acceptable accuracy. This is due to the statistical conditions involved in the concept of measuring the spectrum at given time and location [e.g., *Donelan et al*., 2015; *Donelan and Pierson*, 1983]. To deal with this issue, *Gerling* [1992] and *Hasselmann et al*. [1996] devised the partitioning method, which consists of identifying spectral features that can be associated to physical wave systems. Based on this principle, the wave systems found in the observed spectrum would be "cross assigned" to those of the model spectrum and each of them would be updated accordingly. However, this cross-assignment step is not that straightforward for reasons that are not only practical, but mainly conceptual. Even if robust spectral partitions are obtained from the two sources, a perfect match between them is unlikely. In general there will be more partitions in the observed data, some of them corresponding well with model wave systems, others less so or not at all. The point is that there is substantial uncertainty in the observed 2-D spectral estimates [e.g., *Donelan et al*., 2015, discuss well the related limitations], which appears, or it is interpreted, as natural variability. This variability cannot be handled following a deterministic approach as currently done. Based on this argument, the method proposed here suggests a stochastic view of the cross-assignment routine. For this purpose, we define the wave systems in a statistical sense, i.e., independently of time. This new perspective affects every single step of the assimilation routine, starting from the need to assess the wave climate from a spectral perspective. In this paper, we are specifically concerned with the specification of background errors, because we consider this element central to the whole data assimilation problem. Nevertheless, because they are all interconnected, we address several related issues like the spectral correction step and the optimization scheme.

The paper is organized as follows: in section 2, we give an overview of background errors in numerical wave modeling, focusing on their importance in data assimilation. In section 3, the conceptual framework of the statistically defined spectral wave systems is outlined. Section 4 is dedicated to the issue of correcting the wave spectrum. Section 5 explains the methodology proposed here to derive the background errors. In section 6, we present the related results for two different scenarios (ocean and shelf basins). In section 7, several issues related to the data assimilation algorithm, including the optimization scheme, are addressed. Section 8 summarizes our main conclusions.

2. Background Errors in Wave Modeling

One of the inherent characteristics of environmental modeling in regard to data assimilation is the fact that measurements are point specific, in contrast with the spatial model representation. Independently of the robustness and accuracy of numerical models data, assimilation is expected to improve the model predictions. This is not only because of the general perception that observations are more accurate than model results, but mainly because in the Bayesian framework the combination of information is expected to decrease the overall resulting uncertainties [e.g., *Kalnay*, 2003]. The basic question then is how the point measurement information should be distributed into the model spatial domain. Since it is clear that in data assimilation we are confronted in general with an extrapolation problem, the key issue is that a wrong spatial distribution of observed data will directly translate into the introduction of errors into the model, rather than the intended corrections, no matter how accurate observations are [e.g., *Sannasiraj et al*., 2006].

The mathematical formalism to carry out this spatial distribution is well defined, based on the covariance of background errors. The error covariance, between a reference point (i) of the spatial grid (the measurement, hence assimilation, point) and another arbitrary remote grid point (j), is defined as $\langle w_i\, w_j \rangle$, where w is the error signal we want to analyze, in our case specific to a wave system. Given the respective error standard deviations $\sigma(w_i)$, $\sigma(w_j)$, we define their ρ correlation coefficient as:

$$\rho_{i,j} = \frac{\langle w_i w_j \rangle}{\sigma(w_i)\sigma(w_j)} \tag{1}$$

ρ is a very useful parameter, in practice a metrics of the uncertainty of the reference point to represent the remote point.

Thus, the error covariance is fully specified if the other two terms are known. A sound method for estimating the error standard deviation is given for instance by the triple collocation method [e.g., *Janssen et al*., 2007], but the covariance and the correlation coefficient are tightly linked to each other. A common practice (and common mistake as well) is to consider that the correlation decays with distance, generally raised to some negative power [e.g., *Lionello et al*. [1992]; *Voorrips et al*., 1997; *Hasselmann et al*., 1997; *Aouf et al*., 2006]. This assumption carries with it several physical and statistical implications, the most important one being the homogeneity and isotropy of the wavefield. However, when dealing with waves, homogeneity is hardly fulfilled even in oceanic conditions. At the generation zone, waves depend on fetch, so even for a homogeneous wind field, the longer the fetch, the larger the wave height and the lower the peak frequency. So from their origin there is a marked heterogeneity of the main wave parameters. After the generation, waves are subject to dispersion, so heterogeneity is enhanced (although on a larger scale). Similarly, although there is often a preferential direction for waves at every specific site, more in general there are different wave trains with different genesis, traveling in different directions and with different characteristics. So the isotropic condition is hardly met in reality.

Greenslade and Young [2005] evaluated the isotropy of the significant wave height field. For an example, they considered a 3 month period in the Indian Ocean dominated by monsoon regime. In their approach, they used the anomaly between the time-dependent variable and its expected value (climate) as a metrics for background errors. They found that even for the overall wave height there is a marked anisotropy which was then quantified. In nearshore and shelf conditions *Portilla* [2009] and *Waters et al*. [2013] found the homogeneity condition based only on distance as inappropriate, since at that scale there are many geographical features that influence the wave conditions and therefore also the region of homogeneity. In order to estimate that region, *Portilla* [2009] evaluated the spatial variation of several integral parameters, finding that each of them exhibits a different spatial structure and length scale. Therefore, it was obvious that a more robust and consistent approach was required, the natural step being to consider the wave spectrum as such rather than its integral parameters.

The advantage of the wave spectrum is that it contains a lot of specific information about the local wave climate including some characteristics of the state of wind forcing (wind-sea and swell), as we can derive from the steepness and directional distribution of the various wave systems. Its analysis allows us to make specific and dedicated corrections to each component. Therefore, it is important for data assimilation to have a good knowledge of the local wave climate, not merely in terms of integral parameters, but mainly on a spectral basis. This is the subject of the following section.

3. Long-Term 2-D Spectral Statistics

Wave conditions at any specific site are in general complex with wave trains of different characteristics that in turn are the response to different meteorological events. A common and crude classification is that of wind-sea and swell, but even within these two categories there might be other wave conditions possible. In fact, meteorological conditions are highly variable, so wave characteristics will also be different in many aspects such as genesis, fetch, direction, typical heights and periods, among others. The present representation of waves based on the directional wave spectrum, both from observations and models, is a good, and

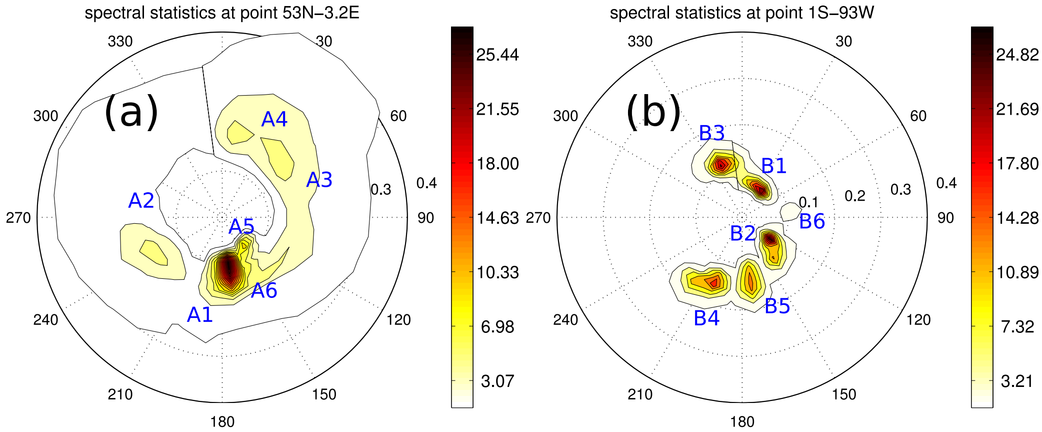

Figure 1. Spectral statistics (occurrence probability of spectral partitions, in percentage) at (a) the North Sea and (b) Equatorial Pacific, obtained from spectral time series from 1979 to 2013.

at present the best, source to derive many of these characteristics. A methodology to develop spectral wave climate variables has been presented by *Portilla et al*. [2015]. The method is based on the statistics of spectral partitions, defined as independent wave systems within the spectrum. The partitioning technique is explained in detail in *Portilla et al*. [2009]; here a brief summary is given for the sake of completeness. The concept of partitioning is based on the assumption that waves originated by particular meteorological events appear as different entities in the wave spectrum. These entities can be identified as spectral patterns, using, for instance, image processing tools. An issue of consideration in this processing is the level of noise inherent to spectral estimates, which depends on the data source, model, or observations. For robust partitioning results this noise needs to be consistently disregarded. In general, a level of about 2% of the total energy is sufficient to characterize noise in observations. In modeled data this level is even lower.

Once consistent and independent wave systems are obtained from a long time series of spectral data, the partitioning technique can be further exploited to characterize the local wave climate on a spectral basis. A proposed methodology for this is given in *Portilla et al*. [2015]. For spectral characterization, several possibilities exist (we can consider different parameters like the mean, median spectrum, or others; some of these were tested in the cited reference). However, from these many possibilities it was found that the "empirical distribution of the peaks of partitions in frequency and direction" is particularly skillful for the purpose of defining wave systems statistically. This bivariate distribution contains information about all possible wave regimes that can occur at a single location, but two particularly useful indicators are their frequency-direction domains, and their occurrence probabilities. This information is crucial to understand the local wave climate in regard to the different possible wave systems. Moreover, it turns out that at every site this distribution is unique and constitutes a sort of wave climate footprint. Therefore, it can also be used to establish the spatial distribution of the individual wave systems. This is a key concept in the present methodology because the spatial distribution of single wave systems derived from model data allows us to understand their structure, extent of homogeneity, isotropy, and therefore their associated correlations, which is the basic information required to estimate the background error covariances. Clearly the statistics can be evaluated on a, e.g., yearly, seasonal or other basis. For the present paper, yearly ones have been considered.

To work out the method, we consider two locations with rather different characteristics, one in the North Sea and the other in the Equatorial Pacific. Figure 1 shows the spectral statistics at the chosen locations. These statistics have been obtained using 35 years of ERA-Interim output from the WAM model operated by ECMWF [see *Dee et al*., 2011, for details about the data set]. These probability density distributions contain two important pieces of information related to the local wave climate. The first is the spectral domain of each wave system, the second is their occurrence probability. From these two, the statistics of all the other spectral parameters (significant wave height, mean wave period, mean direction, among others) for each wave system can be derived. A relatively detailed analysis of the spectral wave conditions of the two considered areas can be found in *Portilla et al*. [2013, 2015]. For the sake of clarity of the overall procedure,

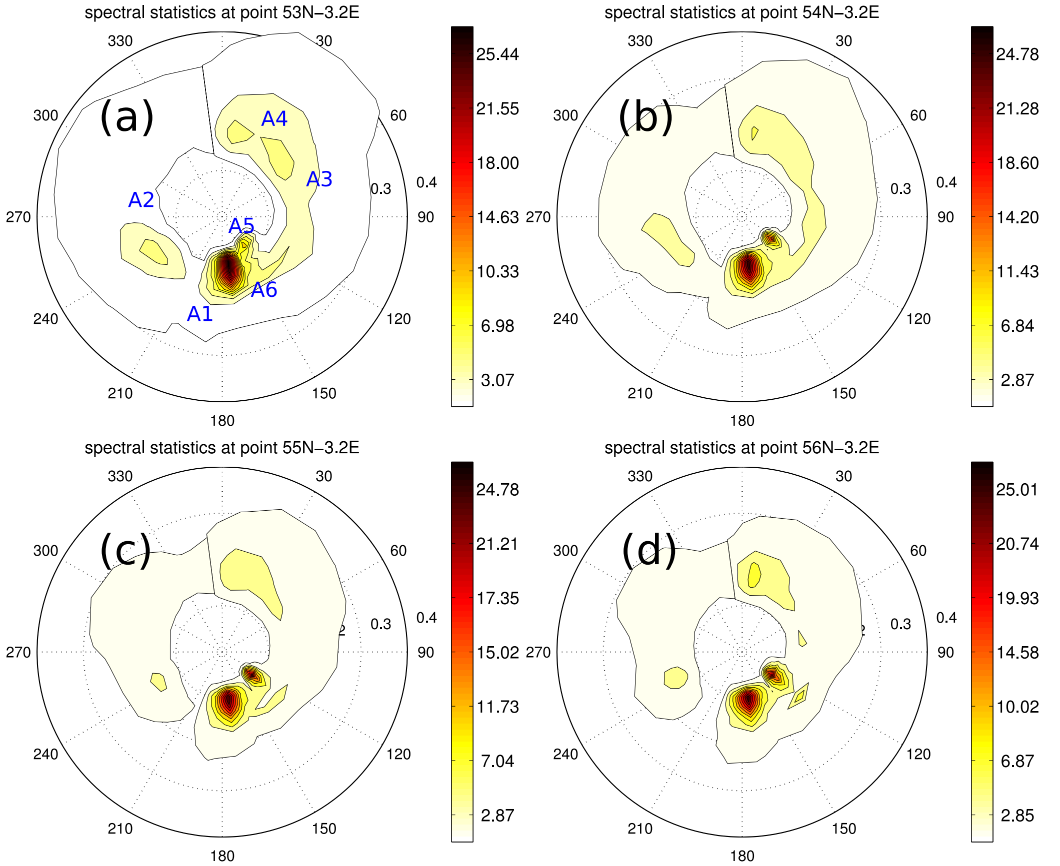

Figure 2. Spectral statistics (occurrence probability of spectral partitions in percentage) of neighboring points in the North Sea. The coordinates are 3.2° in longitude for the four points and from 53° to 56° in latitude [after *Portilla et al.*, 2015].

it is convenient to provide here a brief summary of the main spectral wave systems for which the error covariance is to be estimated. This is done for the North Sea and the Equatorial Pacific.

3.1. Spectral Wave Climate

3.1.1. North Sea

The North Sea is located in the extratropical zone. Therefore, it presents a marked seasonality with strong storms during the boreal winter months (December–January–February), and relatively calm meteorological and wave conditions in the summer (June–July–August). A map of its southern part can be seen in Figure 4, the dot showing the position of the oil tower later used as a reference point. In addition, the whole basin is nearly closed at its west boundary to incoming swells from the Atlantic Ocean, but local waves can be generated by winds blowing over the English channel. In turn, the North Sea is open to the north, so that swell waves generated in the North Atlantic and the Norwegian Sea (to the North of the shown area) penetrate constantly into the domain. Locally generated waves from those directions are also relatively frequent and sometimes associated to extremes due to the long fetch. Other wave systems are related to northeasterly and westerly winds. This information can be derived from spectral statistics as the one shown in Figure 1a. In the southern part of the North Sea, at the location of K13 (53°N–3.2°E), up to seven wave systems can be identified. For the present purposes, from those seven systems we consider only two main groups, the first encompassing all wave systems propagating to south (i.e., A1, A5, and A6), and the second encompassing all the wave systems propagating to north-northeast (A3, and A4). Waves in the first group are the most recurrent, and although their magnitudes in the mean are moderate, as already mentioned, they are associated to the largest extremes in the series. In turn, waves of the second group are less recurrent, but they constitute the most energetic condition on average.

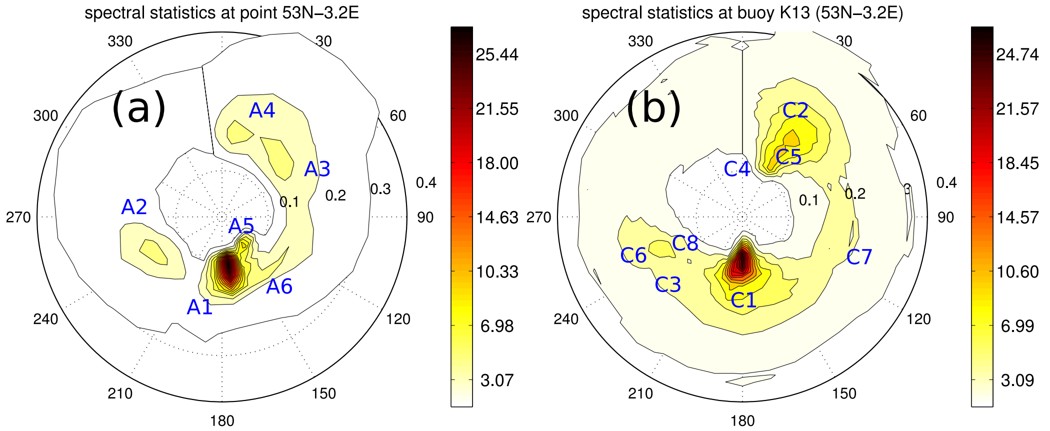

Figure 3. Spectral statistics (occurrence probability of spectral partitions in percentage) of model and observations at the point 53°N–3.2°E in the North Sea corresponding to the Dutch platform K13. (a) Model data (1979–2013), and (b) buoy data (2003–2007) [after *Portilla et al.*, 2015].

3.1.2. Equatorial Pacific

As its name suggests, this area is dominated by tropical conditions, and given the large dimensions of the Pacific Ocean and the geographical position of the reference site at the eastern border (1°S, 93°W), off the coast of Ecuador, it is exposed to swell waves arriving from many different directions (Figure 1b; see also Figure 5 for the geography of the area). Wind conditions vary from moderate to low, preferentially from southwest, but northeasterly winds are also common during the boreal winter months including the wind jets from Central America. Therefore, wave conditions in this zone are dominated by swells. The most recurrent and energetic are generated in the southern hemisphere (mostly active during the austral winter). Following the analysis by *Portilla et al*. [2015], these split into two groups, the first with marked swell characteristics (B1) and the other with characteristics of young swell (B3). In turn, during the boreal winter, swells generated in the northern hemisphere are dominant (B2). These three wave systems are the ones we consider here for developing the background error structures. A similar approach could be followed for the other wave systems seen in Figure 1b (i.e., B4, B5, and B6).

3.2. Cross Assignment

3.2.1. Model to Model Cross Assignment

For the spatial assessment of wave systems, spectral statistics have to be derived for every model grid point. This overall information allows us to understand the spatial structure, dimensions, and other characteristics of each individual wave system. For that it is necessary to find at every model remote location (j), the corresponding wave system related to that of the model observation location (i). The key point is that in these (statistically defined) wavefields we do not have "discontinuity fronts." Rather, the change from point to point is progressive, in the sense that a local feature, e.g., a wave system or partition, is expected to appear, albeit with some little modifications, in the neighboring points, the difference obviously increasing with distance because of the general meteorological pattern, of local coastal features or others. This can be conveniently summarized saying that the field is characterized by a differential variability of the spectra [see, e.g., *Van der Westhuysen*, 2013]. All these properties allow the procedure to be carried out automatically.

We show all this considering the spectral statistics at several neighboring points in the North Sea (Figure 2). The wave system 1 (A1) is probably the easiest to identify in all grid points. Once this wave system has been chosen (as target for data assimilation at the corresponding observation location), given its (f, θ) peak position, the algorithm is set to search in the neighboring points the corresponding (f, θ) coordinates, within some range of variation. Note that the spectral domains of wave systems A5 and A6 are relatively close to that of A1. Besides, they share other characteristics like the propagation direction. For this reason, these three wave systems are considered here as a single group.

Similarly, all the other model wave systems can be easily identified and cross assigned in the model spatial domain. In general, there is a one to one correspondence of wave systems among neighboring points. The possible exceptions are found in the spatial boundaries of a particular wave system. For instance A3 can be

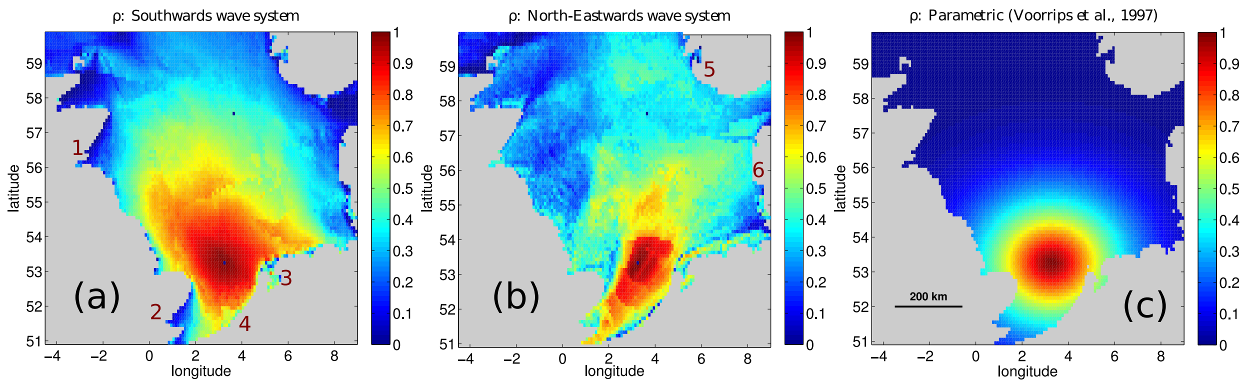

Figure 4. Background errors correlation coefficients in the southern North Sea, at the K13 location (53°N–3.2°E). (a) Southward wave system (buoy C1, model A1–A5–A6), (b) northeastward wave system (buoy C2–C4–C5, model A3–A4), and (c) parametric correlation as used by *Voorrips et al*. [1997]. See Figure 3 for wave system definition.

clearly identified at point 53°N–3.2°E, but it is less and less evident in the other points as one moves northward. This indicates that the boundary has been reached and this condition can be easily specified in the algorithm.

3.2.2. Model to Observations Cross Assignment

One of the most challenging aspects in spectral data assimilation is the ability (or inability) of the algorithm to cross-assign information between model and observations. It is clear that such a task cannot be done directly between the two corresponding matrices. Observed and model spectra typically have different resolutions, frequency ranges, noise levels, among others. At first sight the partitioning technique seems to be the right tool to assist this operation because of the data reduction involved. However, there are other difficulties involved. The spectrum itself is a stochastic variable, and while observations are estimates of that variable obtained via sampling, models are essentially deterministic computations of that stochastic variable. Of course this does not mean model spectra are a faithful representation of the truth, and for serious reasons. Reality can be very complex, with, e.g., wind fields varying over spatial or time scales that are not even represented by meteorological models. As a consequence, apart from the noise, one cannot expect observations partitions to be perfectly consistent in time and this complicates enormously the cross assignment of partitions, to the point that it is nearly impossible to consign this task to a computer algorithm given the level of subjectivity involved. A feasible solution to this problem is presented here, again by means of long-term 2-D spectral statistics. The occurrence probability of spectral partitions defines all possible wave systems at the location (eight wave systems for the example of Figure 3b). All time-dependent partitions and their variability will fall within these eight clusters. Consequently, this cross assignment can be done using statistical stencils from model and observations, so that the time dependence issue is overcome.

To this end, similar spectral statistics need to be derived from observed data, which contrarily to the model ones are point specific. In this case, it is reasonable to expect differences (possibly large) between the observed statistics and those of the model grid point at the same location, which is one of the reasons why data assimilation is required. In any case, observed spectral wave systems need to be matched to those of the model at the same location. However, in spite of the fact that using spectral statistics of both sources reduces significantly the level of subjectivity making this task a lot easier, it is hard to conceive that a mere computer algorithm can do this model to observations cross-assignment task correctly. As a matter of fact only a subjective and expert assessment can guarantee consistent results. Therefore, although it may be possible to develop an algorithm to carry out this task, in the present solution we still leave this step to a human. The great advantage in the present approach is that such a task only needs to be carried out once for every observation point that is introduced in the assimilation system.

For illustration, consider the buoy data from the Dutch platform K13 and model data at the same location (53°N–3.2°E) shown in Figure 3. As a whole, there is a relatively good agreement between both statistics. However, there is not a one to one correspondence. For instance wave systems A1, A5, and A6 from the model cannot be clearly distinguished in the buoy data, but the general trend is consistent, there are swell systems propagating southward (C1) and these make the most recurrent condition at the reference point. Similarly, by visual assessment we can cross-assign A2 from the model to the buoy wave systems C3, C6, and C8. A slightly more challenging case is that of wave systems A3 and A4 from the model. In the buoy

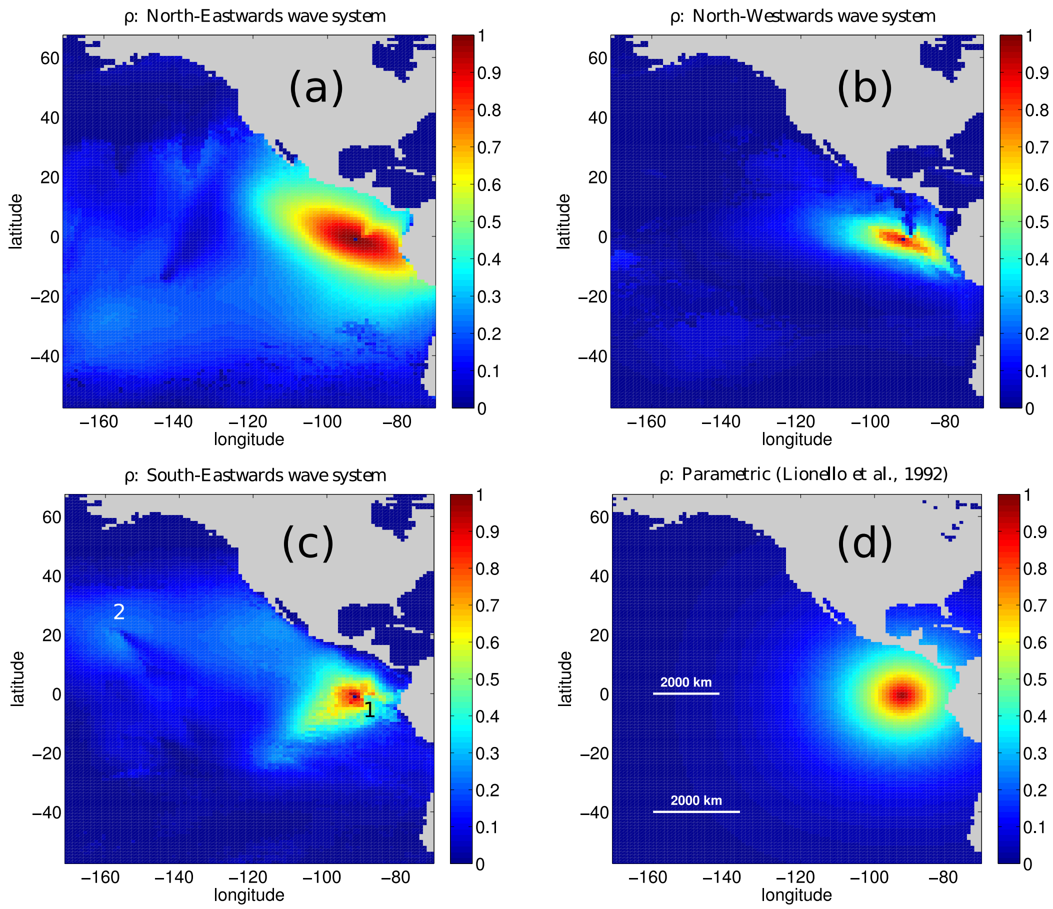

Figure 5. Background errors correlation coefficients in the Equatorial Pacific Ocean, location (1°S–93°W), (a) Northeastward wave system, (b) Northwestward wave system, and (c) Southeastward wave system, (d) parametric correlation as used by *Lionello et al*. [1992].

data there are three wave systems (C2, C4, and C5) with directions similar to A4, but A3 is not clearly represented. The question is then whether we can cross-assign model wave system A3 and A4 to buoy wave systems C2, C4, and C5. For the present application and illustration we have proceeded calling this group *northeastward* wave system. Note however that this decision remains subjective and that other possibilities exist. Note also that consigning such an assignment to a computer automatic procedure can be very difficult. However, while this is a repetitive task, to be repeated (and judged) at each new application, the advantage in the present approach is that these two cross-assignment tasks are carried out on a spectral statistical indicator, hence not on a time-dependent basis, but done only once for every assimilation point.

Note that in the present approach the wave systems are predefined and statistically cross assigned. Therefore, time-dependent spectra have to conform to the statistical stencils. This approach implies a couple of interesting aspects. The first is that every observation point and every considered wave system must have an individually associated background error covariance matrix. The second is that time-dependent spectra are compared (in real-time data assimilation). If a certain wave system is (is not) present in the observation data, but it is not (is) in the model, then the system needs to be introduced in (erased from) the model. It should be noted as well that the wind-sea/swell attribute is considered here implicitly, because each wave system having a specific spectral domain has also well defined characteristics as for instance the height/length, i.e., H/T, ratio, regarding the wind forcing [see *Portilla et al*., 2015].

4. The Spectral Correction Scheme

In principle, the information of the individual wave systems allows to correct as many parameters as observed by the instrument. For instance, wave measuring buoys based on 3-D accelerometers or corresponding systems are capable of providing the 1-D spectrum plus, for each frequency component, the first four Fourier coefficients of its directional distribution, information from which variance density, mean direction, directional spreading, skewness, and kurtosis can be derived [e.g., *Kuik et al.*, 1988]. However, model spectral errors cannot be evaluated and corrected further than what is measured. Therefore, in the present approach the estimation of background errors is tightly linked to the spectral estimation scheme. For the development of the methodology, the following straightforward model is considered:

$$S(f,\theta)^m_{j,estimate} = \alpha_o{}^* S(\beta_o{}^* f, \theta+\delta_o)^m_{i,true} \tag{2}$$

Equation (2) states that the directional spectrum of wave system m at any remote grid point j is estimated using the information of the "true" corresponding wave system at the observation location i. This estimation is carried out with the three parameters α_o, β_o, and δ_o which target the system main wave parameters, i.e., variance density (S), frequency (f), and direction (θ) respectively.

Note that equation (2) implies that a new field of wave conditions, over a spatial domain around the observation/assimilation point, is being "produced" purely on the basis of the observed single point information, obtained by means of a parametric model α, β, and δ in equation (2)), hence with the derived partition characteristics, varying from place to place (grid point to grid point). This new information will be combined with the wave model field, which in turn is built from the physical representation of the dominant processes. In the Bayesian framework, each of these fields is given a weight according to their error statistics. It is in this idea that lies the importance of background errors.

In addition, both wind-sea and swell tend to maintain the respective characteristic relationship between wave height and period. For wind-sea, this dependency is rather strong and related to the growth process and dynamics of a wind-sea. In the case of swell, it is associated to the propagation history of the system and to the low spatial gradient of its characteristics. For the present evaluation of background errors, we expect these relations to be fulfilled to the extent the model is able to represent them. Therefore, for a practical approach we use the inherent relations developed by the model for each system, but attention has to be given to the fact that such a relation could be tested and addressed by data assimilation in the present context.

5. Computation of Background Errors in Spectral Wave Modeling

Equation (1) provides a relationship between the error covariance $\langle w_i\, w_j \rangle$, the error standard deviations $\sigma(w_i)$, $\sigma(w_j)$, and the error correlation coefficient ρ. For the practical use of data assimilation, we need to know $\langle w_i\, w_j \rangle$. This can be derived knowing σ and ρ. Assuming (soon to be discussed) the availability of the true spectra at all the grid points, the ρ between any two points is obtained by evaluating the spectral correction model given in equation (2), quantified through the coefficient of determination (R^2, $\rho \approx sqrt(R^2)$) defined as:

$$\left(R^m_{i,j}\right)^2 = 1 - \frac{\int_\theta \int_f \left[S^m_{j,estimate} - S^m_{j,true}\right]^2 df\, d\theta}{\int_\theta \int_f \left[S^m_{j,true} - \overline{S^m_{j,true}}\right]^2 df\, d\theta} \tag{3}$$

For a given wave system m, $(R^m_{i,j})^2$ provides a relationship between the model grid point (i) (conceived as the observation location) and the generic remote point (j). $S^m_{j,estimate}$ is the spectrum (of m) at the remote point estimated with the scheme given in equation (2) from the true spectrum at the observation location (i.e., $S^m_{i,true}$). With $S^m_{j,true}$ the true spectrum at (j), the numerator evaluates the differences between the spectrum obtained from equation (2) and the true one. The denominator is a normalizing term function of the true spectrum and its expected value, where $\overline{S^m_{j,true}}$ can be defined, for instance, as the average. Note that the differences are evaluated over the whole spectral domain (f, θ), so that the so defined R^2 aims to penalize (i.e., to be highly sensitive to) any difference in the spectrum and not only (to) an integrated quantity. However, given its definition, R^2 is itself an integrated parameter. Therefore, it allows us to summarize the spectral information into a single (spatial grid) parameter, but with a much deeper meaning that what derived, e.g., from the simple wave height.

The set of equations from (1) to (3) gives us a complete description of the background errors specific to every observation/assimilation point and to every wave system. For its full evaluation, we only need a reference for the true spectra at all the grid points over the domain. Since we are interested in background (model) errors, an acceptable valid truth can be emulated using model results. There are two main arguments for this choice. The first one is that we cannot expect to have observations at every model grid point, and, even if we had them, they would not be the true values, but just other estimates. The truth is always unknown. The second is that the current state-of-the-art wave models represent wave conditions based on the description of the main physical processes, which are known to a relatively high degree. Indeed using a good wave model and an up-to-the-level scheme to transform local into remote spectra (equation (2)), better estimates of background errors can be obtained.

Ideally, the spectral time series should be as long as possible. Current standards suggest time series of at least a decade (depending on the application). However, a practical related limitation is the volume of data for storage and processing, because spectra at every grid point need to be saved at regular intervals. A possibility is to use archived data from recognized databases, just like the ERA-Interim results presented in Figures 1 and 2. However, during the course of this study spectral data from ERA-Interim were not openly available, except at a limited amount of points, and the grid resolution in the North Sea ($\sim$0.5°) is lower than what desirable for our purpose. In addition, our main goal here is to illustrate the methodology, rather than obtaining estimates with the highest precision. Therefore we limit the simulation to one, chosen arbitrarily, year period, 2012. In this context, we have emulated a set of true spectra using the WAVEWATCHIII model v3.14 with default physical packages and parameters (see Tolman et al., 2013 for details). Forcing winds and ice coverage concentration in the poles were obtained from the GFS system (Global Forecast System) [e.g., *Yang*, 2013], and bathymetry data corresponds to the ETOPO database [*Smith and Sandwell*, 1997]. Two different environmental scenarios where considered for evaluation, one in the Equatorial Pacific and the other in the North Sea, representing respectively the wide open and deep ocean and the typical conditions in an enclosed sea, with possibly also limited depth effects. The spatial grid resolution is one geographical degree in the Pacific, and one tenth of a degree in the North Sea. The spectrum in both cases was discretized into 29 frequencies from 0.0350 to 0.5047 Hz in geometric sequence, and 24 directions equally spaced. Spectra were saved at 6 h interval.

The scheme parameters α_o, β_o, and δ_o are derived from the transformation, in time-dependent basis, of the true observation location spectra (i) into the true remote spectra (j), both known. Naturally this transformation is based on the same scheme of equation (2), using the time-dependent parameters α, β, and δ obtained via optimization [MATLAB optimization routine *fmincon*, a constrained minimization function in which we target the differences between the two spectra, e.g., *Byrd et al*., 2000]. Note that in fact α, β, and δ are variables themselves, and they carry information about the relation between variance density and frequency for the wave system under consideration. Therefore, the relationship between α_o, β_o, δ_o, and α, β, and δ is not unique. Rather, we find a distribution of the possible values. Here we consider the most recurrent ones, i.e., the peak value of their empirical probability density function. In turn, this choice implies that the background covariance will also be representative of the most recurrent cases. However, note that some data assimilation applications may need to target other conditions, extreme values for instance. In that case, among other conditions, the background errors need to be designed according to the specification of the scheme parameters of equation (2).

Finally, a new time series of estimated spectra around the target point is obtained using equation (2). Since the true spectra are known, we can evaluate the related errors using equation (3). These results are presented in the next section.

6. Results

6.1. North Sea

For the North Sea case, we consider two main wave groups. The first encompasses the wave systems propagating southward (A1, A5, and A6), and the second encompasses the wave systems propagating northeastward (A3, and A4) (see Figure 1a). The southward group is more related to swell conditions, while the northeastward group is more related to wind-sea conditions. The most direct result shown by Figure 4 is the completely different correlation structures of these two wave groups. The correlation domain of the southward waves (Figure 4a) is rather large, with values of 0.8 and higher covering a significant fraction of the whole basin. Due to the southward propagation direction, geographical features like the Firth of Forth and Thames delta (points 1 and 2) generate shadow zones that are decorrelated from the reference (assimilation) point. Something similar occurs at the Dutch and Belgian coasts (points 3 and 4, respectively) in this case not caused by a shadowing effect, but due to the shallow bathymetry inherent to that region. Note that these and other similar effects do not need to be explicitly accounted for in equation (3). Indeed they are generated implicitly due to the changes in the wave spectra over the spatial domain. This is important as it speaks for the robustness of the method. Another characteristic of the southward correlation is the relatively isotropic shape, especially in the zone of high correlation (above 0.8).

For the northeastward waves (Figure 4b), it can be seen that the shape of the high correlation area (above 0.6) exhibits highly anisotropic conditions and it is clearly oriented along the English channel with a fanning out shape toward the north. The associated correlation length scales are also significantly smaller than those of the southward wave systems. Another remarkable characteristic of this correlation is that for the medium values (around 0.5) it covers a great area of the North Sea basin (going up to the Norwegian and Danish coasts, points 5 and 6), indicating the dominance of wind regimes all over the area whose wave generation patterns are affected by orographic effects.

The classical parametric approach, as used by *Voorrips et al*. [1997], Figure 4c, leads to a correlation closer to that of the southward wave systems, but it is clearly ill to fit the complex geography of the area. Its isotropic distribution looks at once more apt to the open space of the oceans rather than enclosed seas, areas where the orography, and in some cases the bathymetry, may play a fundamental role. Its most problematic issue is the unlikely high correlation values between the reference (assimilation) point, and remote points located near the coasts or behind shadowing areas. The fact is that these high correlations force the assimilation scheme to impose wave conditions in the remote points that might be unphysical and may certainly lead to wrong results, but eventually also to model instability, as reported by *Portilla* [2009].

6.2. Equatorial Pacific

Similar results are obtained at oceanic scales (see Figure 5) with the major difference being the length scales of the correlations. Of the three wave systems considered, there is a highly regular correlation for the wave system propagating northeastward (Figure 5a), although with a clearly anisotropic shape. These waves have marked swell characteristics and are originated in the extratropical storm zone of the southern hemisphere. It can be seen that relatively high-correlation values (above 0.4) include the whole Central American coasts. Another interesting feature, not only in this correlation, but in the two other ones as well, is the shadow zone produced by the Galapagos archipelago (where the reference point is located).

The length scales of the northwestward wave system correlation (Figure 5b) are significantly smaller and the anisotropic condition more pronounced. The reason is that this wave system is more associated to the local wind conditions and more specifically to the trade winds that are subject to geostrophic directional changes in the area. Here it is clear that, even though these wave systems have some similarities in the spectral domain (mainly regarding propagation direction), they cannot be assimilated using a single scheme because, although both correspond to oceanic conditions, their spatial structure is significantly different.

In Figure 5c, we see the correlation of the wave systems propagating southeastward, generated in the Northern hemisphere. In this case, the correlation length scales for the higher values (above 0.6) are significantly lower than those of the wave system propagating northeastward, although these last present a more clearly defined isotropic behavior. Similarly as in the case of the North Sea, here a shadow zone is visible along the Central American coasts, caused by the geographical features upstream this area. Again the disturbance of the Galapagos archipelago is noticeable, as also visible is the one due to the Hawaii islands (points 1 and 2 in Figure 5c). The parametric correlation used by *Lionello et al*. [1992] seems to fit better this last wave system (shadow zone difference between two panels). However, a major inconsistency of this structure is the inability to adapt itself to the geographical and physical conditions. For instance, the Caribbean wave conditions seem to have little or no correlation with the wave conditions in the Pacific according to the results of Figures 5a–5c. However, this cannot be specified with the parametric correlation (Figure 5d), and eventually during data assimilation these two basins will be (wrongly) exchanging information. A similar situation occurs for the coastal zone. For the northwestward and southeastward wave systems

(Figures 5b and 5c), the conditions nearshore (Central America) seem to be different than in the open ocean, but this is not taken into account by the parametric correlation.

7. Discussion

We present here a new approach to specify background errors following the spectral framework that is nowadays standard for numerical wave models and most in situ and remotely sensed observations. This new perspective goes far beyond the mere specification of errors and requires a thorough redesign of the data assimilation scheme which itself has to be embedded in the spectral framework. This includes the need of taking into account from the onset the *spectral wave climate* of the considered area. Although this consideration adds an extra rather laborious step to the scheme, it solves two major drawbacks of current data assimilation systems, namely the transfer of information from the wavefields to the wave spectra (for the case of integral parameters), and the cross assignment of spectral information (for the case of spectral data). If wave systems are handled with a stochastic perspective these two steps become actually a lot simpler and the robustness of whole scheme increases significantly.

A data assimilation scheme based on this approach has the structure presented in Figure 6. In this scheme, spectral partitioning is applied to both sources (model and observations), their spectral statistics are derived and wave systems are cross assigned. The data assimilation scheme runs for every wave system with a corresponding background error covariance matrix (BECM), resulting directly into the analyzed spectra. Note that in this scheme the spectral wave climate and background errors play a central role together with the data assimilation (DA), or optimization step. This is not the case in current systems where the major task is performed by the DA scheme, and background errors are only part of the parameterization.

In this regard, it is worth noticing that a common point of discussion in the data assimilation literature spins often around the *optimization scheme*. Currently several approaches exist for this purpose (e.g., Optimal Interpolation, 3DVAR, 4DVAR, Kalman filtering, among others). From the just presented spectral approach and from Figure 6, it is clear that the role of the optimization scheme is to properly combine the two available data sources, i.e., (a) the first guess from the model with (b) the observations based field produced in the DA process (generated, for instance, through equation (1)). In this context, it is not evident that improving the optimization scheme "only" would necessary improve DA results. In turn, it is clear that the observations based field plays a central role.

Consequently, the data assimilation practice in general exerts pressure on observed data. As we have illustrated, integral parameters are too limited for consistent wave data assimilation. Moreover, the spectral wave climate comparison and subsequent cross assignment demand a decent match between model and observations, and this comparison helps identifying common model shortcomings from the onset. Here we have used the case of buoy data to work out the examples, but the method should be valid for any spectral source. There can be specific problems related to a given instrument and its characteristics. We have mentioned the SAR spectra, a high potential data source which, apart from the cross-assignment issues referred to here, have another shortcoming related to the directional ambiguity. If such ambiguity is not properly resolved the cross-assignment task even on a stochastic basis becomes impossible and this renders the whole data set unsuited for data assimilation. This was at least the case for the EnviSat data set, although the more modern instruments carried on board the Sentinel-1 mission seem to be better suited to recognize propagation direction [e.g., *Husson et al.*, 2012].

The example of observed data presented here (Figure 3b) refers to buoy data based on accelerometers, which nowadays encompass a rich network worldwide. In addition, other modern techniques based on electric, acoustic, or electromagnetic sampling provide also estimates of the 2-D spectra. However, we should keep in mind that directional wave estimates are only approximations of the truth. As a matter of fact, at present such spectral directional distribution is unknown, and assumptions have to be necessarily done to derive the 2-D spectrum [e.g., *Kuik et al.*, 1988; *Donelan et al.*, 2015]. This issue is not discouraging for data assimilation because the Bayesian combination of data allows us to specify a weight to every source according to their related uncertainty. Therefore, data assimilation should not be viewed as a numerical artifact to make up model results. On the contrary, the practice of data assimilation compels us to be more demanding in regard to observations and this is a positive driver for further developments. In our example of Figure 3, the match between model and observations is not perfect, there are components in the two sources that do not have a

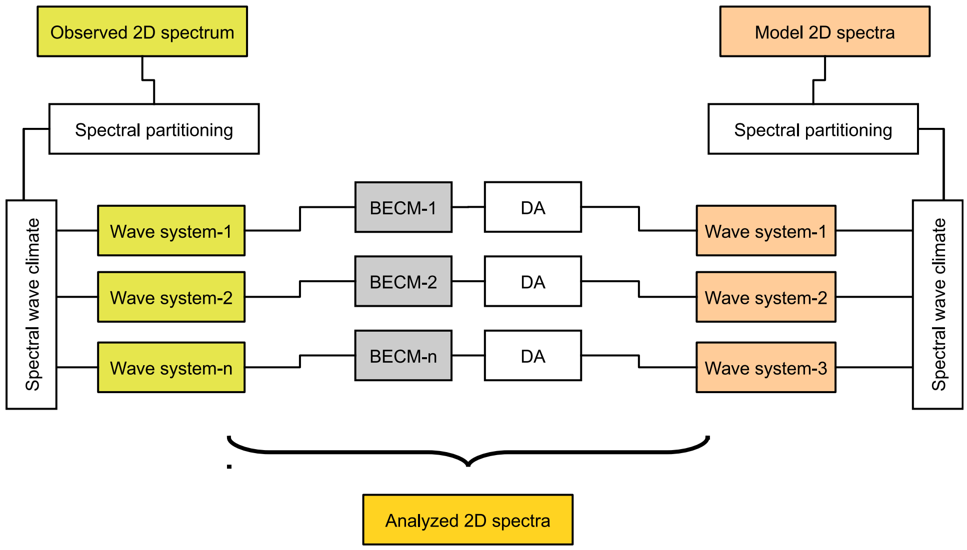

Figure 6. Data assimilation scheme structure in the framework of spectral wave data assimilation. BECM stands for background error covariance matrix.

corresponding pair. This is why the cross assignment between model and observations requires expert human assessment. Nevertheless, the match at the K13 location is very encouraging, with the main spectral features properly represented by the model, probably because wave conditions in the North Sea are not too complex. Such a good level of match cannot be expected at every location, less so in the open ocean. In the context of spectral wave climate and the data assimilation approach presented here, a dedicated evaluation of model performance at spectral level is a necessary step.

Every data assimilation system must be fit to, and exploit the possibilities of, the system it is trying to improve. In this paper, we are proposing a new system whose aim is to take full advantage of the information provided by the wave spectrum. Indeed the latter is the best and most complete description of the sea surface we have at present on a large scale, now universally used and somehow taken for granted. Starting from the historical paper by *Pierson and Marks* [1952], the spectral approach has been the backbone of most wind wave studies and of today highly successful wave forecast systems. If this is to be the case in the future is a tricky question. Except for the deterministic approach presently suitable only for very limited areas (starting from the Euler equation see, e.g., the Boussinesq and mild slope approximations, *Mei* [1983], 510–512 and 86–89, respectively), no other solution seems to be lying in the faint light of an early dawn. However, ideas, supported by the continuously rising computer power, may, and often do, arise unexpected. *Cavaleri* [2006] tried to guess the future, the basic point being how much ground a deterministic approach will gain on the more economical, but somehow simplifying, spectral one. Thinking to the nonlinearity embedded into the evolution of a wavefield (and of its interaction with the atmosphere and the ocean circulation) a possible intermediate solution could be to work with spectra as basic information, but estimating their evolution via the deterministic approach on a physical realization of the wavefield derived at each grid point with random phases from the local known spectrum. Another suggested solution is to work with wave groups rather than single waves, both these physical quantities still waiting for a rigorous, if possible, definition in a full two-dimensional field. An expected extension of any data assimilation system is associated to a more comprehensive view of the interaction between ocean and atmosphere. Here, in the critical layer that is the sea surface where we model waves, many processes act and interact, something that somehow will have to be taken into consideration if we want to improve our modeling, and understanding, of the overall process. Whichever the new solution, and the likely corresponding increase of available information, they will require new approaches also for the assimilation technique. Which and how will depend also on the amount of data available, something that, being basically dependent on technology, is more likely to see a continuous and rapid development and increase. What we know is that the assimilation techniques will, and will have to, exploit all the available information for a better forecast, but also, and probably more important, for a better understanding of the world we live in.

8. Summary and Conclusions

A new methodology is proposed for the computation of background errors in wave data assimilation systems. This methodology aims to conciliate the physical and statistical constraints involved in the definition of background errors. In order to do that the present approach exploits as much as possible the information provided by wave spectra. To this end it is essential to consider from the onset the spectral wave climate of the study area, from which the different wave systems can be recognized and targeted.

Background errors are at the core of any assimilation scheme because data assimilation consists in deriving a new wavefield from the observations at a single point. The derived new wavefield is combined in a Bayesian framework with the background field delivered by the model. Therefore we can have a very sophisticated assimilation scheme and also accurate observations. However, if, because of, e.g., a wrong correlation assumption, the wavefield derived from observations is not correct, this translates into errors rather than corrections in the analysis, no matter how sophisticated the assimilation scheme or how accurate the point observation are.

Under the present view of statistically defined (time independent) wave systems, a solution is offered for a critical step of spectral wave data assimilation, which is the cross assignment of wave systems from different sources. The execution of this task in time-dependent bases is close to impossible, especially considering the model-observations cross assignment, because the level of subjectivity required is high. Following the statistical approach presented here the subjective judgment (still required) is confined to a single and smooth spectral indicator, making this task now feasible.

Given their different origins, the several wave systems composing the local wave climate are governed by completely different correlation structures. The present approach produces correlations that are consistent with the physical processes at work in the area. This is not explicitly introduced in the scheme, but it is obtained from the memory of the system and retrieved from its wave spectra.

Acknowledgments
Spectral wave data from the ERA-Interim archive have been kindly provided by ECMWF (http://www.ecmwf.int/). We acknowledge the help of Luciana Bertotti for data acquisition. Wind fields and ice coverage concentration data used for the WAVEWATCHIII model were obtained from the Global Forecast System of the NCEP-NOAA center (http://www.emc.ncep.noaa.gov/index.php?branch=GFS). We are very grateful to Rijkswaterstaat from the Dutch Ministry of Infrastructure for providing the buoy data at K13 (https://www.rijkswaterstaat.nl/), and to the Research Center on Mathematical Modelling MODEMAT-EPN for their valuable input. J. Portilla acknowledges funding from Escuela Politécnica Nacional (Project PIMI-1402). Luigi Cavaleri took part to this research as a contribution to the EU-funded MyWave project, Theme (SPA.2011.5.'03).

References

Aouf, L., J. M. Lefèvre, and D. Hauser (2006), Assimilation of directional wave spectra in the wave model WAM: An impact study from synthetic observations in preparation for the SWIMSAT satellite mission, *J. Atmos. Oceanic Technol.*, *23*, 448–463, doi:10.1175/JTECH1861.1.
Byrd, R. H., J. C. Gilbert, and J. Nocedal (2000), A trust region method based on interior point techniques for nonlinear programming, *Math. Program.*, *89*(1), 149–185.
Cavaleri, L. (2006), Wave modeling: Where to go in the future, *Bull. Am. Meteorol. Soc.*, *86*, 207–214, doi:10.1175/BMS-87-2-207.
The Coastal Data Information Program (2015) [online]. [Available at https://cdip.ucsd.edu/(online), last accessed 27 Aug. 2015.]
Chen, H. S., D. Behringer, L. D. Burroughs, and H. L. Tolman (2004), A variational wave height data assimilation system for NCEP operational wave models, *Tech. Procedures Bull. Ser. MMAB-2004-04*, 3 pp., NCEP, Camp Springs, Md.
Data Buoy Cooperation Panel (2015) [online] [Available at http://www.jcommops.org/dbcp/, last accessed 27 Aug. 2015.]
Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*, 553–597, doi:10.1002/qj.828.
Donelan, M., and W. Pierson (1983), The sampling variability of estimates of spectra of wind-generated gravity waves, *J. Geophys. Res.*, *88*, 4381–4392, doi:10.1029/JC088iC07p04381.
Donelan, M., A. Babanin, E. Sanina, and D. Chalikov (2015), A comparison of methods for estimating directional spectra of surface waves, *J. Geophys. Res. Oceans*, *120*, 5040–5053, doi:10.1002/2015JC010808.
Gerling, T. W. (1992), Partitioning sequences and arrays of directional ocean wave spectra into component wave systems, *J. Atmos. Oceanic Technol.*, *9*, 444–458, doi:10.1175/1520-0426(1992)009 < 0444:PSAAOD>2.0.CO;2.
Greenslade, D., and I. Young (2005), The impact of altimeter sampling patterns on estimates of background errors in a global wave model, *J. Atmos. Oceanic Technol.*, *22*, 1895–1917, doi:10.1175/JTECH1811.1.
Hasselmann, S., C. Bruning, K. Hasselmann, and P. Heimbach (1996), An improved algorithm for the retrieval of ocean wave spectra from synthetic aperture radar image spectra, *J. Geophys. Res.*, *101*, 16,615–16,629, doi:10.1029/96JC00798.
Hasselmann, S., P. Lionello, and K. Hasselmann (1997), An optimal interpolation scheme for the assimilation of spectral wave data, *J. Geophys. Res.*, *102*, 15,823–15,836, doi:10.1029/96JC03453.
Husson, R., F. Ardhuin, F. Collard, B. Chapron, and A. Balanche (2012), Revealing forerunners on Envisat's wave mode ASAR using the Global Seismic Network, *Geophys. Res. Lett.*, *39*, L15609, doi:10.1029/2012GL052334.
Janssen P. A. E. M., S. Abdalla, H. Hersbach, and J. R. Bidlot (2007), Error estimation of buoy, satellite, and model wave height data, *J. Atmos. Oceanic Technol.*, *24*, 1665–1677, doi:10.1175/JTECH2069.1.
Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation and Predictability*, 364 pp., Cambridge Univ. Press, Cambridge, U. K., ISBN: 9780521796293.
Kuik, A. J., G. Ph. van Vledder, and L. H. Holthuijsen (1988), A method for the routine analysis of pitch-and-roll buoy wave data, *J. Phys. Oceanogr.*, *18*, 1020–1034, doi:10.1175/1520-0485(1988)018 < 3C1020:AMFTRA>3E2.0.CO;2.
Lionello, P., H. Günther, and P. A. E. M. Janssen (1992), Assimilation of altimeter data in a global third-generation wave model, *J. Geophys. Res.*, *97*, 453–474, doi:10.1029/92JC01055.
Marine Automatic Weather Station Network (2015) [online]. [Available at http://www.ncof.co.uk/Marine-Automatic-Weather-Station-Network.html#measurements, last accessed 27 Aug. 2015.]

Mei, C. C. (1983), *Applied Dynamics of Ocean Surface Waves, Adv. Ser. Ocean Eng*., vol. 1, 740 pp., John Wiley, N. Y.

National Data Buoy Center (2015) [online]. [Available at http://www.ndbc.noaa.gov/, last accessed 27 Aug. 2015.]

Pierson, W. J., Jr., and W. Marks (1952), The power spectrum analysis of ocean-wave records, *Earth Space Sci. News*, vol. *33*(6), 834–844.

Portilla, J. (2009), Buoy data assimilation in nearshore wave modelling, PhD dissertation, 193 pp., Katholieke Univ., Leuven.

Portilla, J., F. Ocampo-Torres, and J. Monbaliu (2009), Spectral partitioning and identification of wind sea and swell, *J. Atmos. Oceanic Technol.*, *26*, 107–122, doi:10.1175/2008JTECHO609.1.

Portilla, J., J. Sosa, and L. Cavaleri (2013), Wave energy resources: Wave climate and exploitation, *J. Renew. Energy*, *57*, 594–605, doi:10.1016/j.renene.2013.02.032.

Portilla, J., L. Cavaleri, and G. Ph. Van Vledder (2015), Wave spectra partitioning and long term statistical distribution, *Ocean Modell.*, vol. 96, Part 1, 148–160, doi:10.1016/j.ocemod.2015.06.008.

Rijkswaterstaat, The Netherlands (2015) [online]. [Available at http://www.rijkswaterstaat.nl/kaarten/golven.aspx, last accessed 27 Aug. 2015.]

Sannasiraj, S. A., and M. G. Goldstein (2009), Optimal interpolation of buoy data into a deterministic wind-wave model, *Nat. Hazards*, *49*, 261–274.

Sannasiraj, S. A., V. Babovic, and E. S. Chan (2006), Wave data assimilation using ensemble error covariances for operational wave forecast, *Ocean Modell.*, *14*, 102–121.

Smith, W., and D. Sandwell (1997), Global seafloor topography from satellite altimetry and ship depth soundings, *Science*, *277*, 1956–1962. [Available at http://www.sciencemag.org/content/vol277/issue5334/.]

Tolman, H. L., M. L. Banner, and J. M. Kaihatu (2013), The NOPP operational wave model improvement project, *Ocean Modell.*, *70*, 2–10, doi:10.1016/j.ocemod.2012.11.011.

Van der Westhuysen, A. (2013), *Spatial and Temporal Tracking of Ocean Wave Systems*, Mar. Model. and Anal. Branch, NOAA/NWS/NCEP/EMC, WAVEWATCHIII Winter School presentation. [Available at http://polar.ncep.noaa.gov/waves/workshop/pdfs/wwws_2013_wave_tracking.pdf.]

Voorrips, A. C., V. K. Makin, and S. Hasselmann (1997), Assimilation of wave spectra from pitch-and-roll buoys in a North Sea wave model, *J. Geophys. Res.*, *102*, 5829–5849, doi:10.1029/96JC03242.

Waters, J., L. Wyatt, J. Wolf, and H. Adrian (2013), Data assimilation of partitioned HF radar wave data into Wavewatch III, *Ocean Modell.*, *72*. 17–31, doi:10.1016/j.ocemod.2013.07.003.

Wittmann, P. A., and J. A. Cummings (2005), Assimilation of altimeter wave measurements into Wavewatch III. Oceanography division, *Rep. NRL/PP/7320-04-0003*, Naval Res. Lab., Stennis Space Cent., Hancock, Miss.

Yang, F. (2013), Review of GFS forecast skills in 2013, *NCEP Tech. Rep*., NCEP, Greenbelt, Md. [Available at www.emc.ncep.noaa.gov/GFS/docs/GFS.performance.review.2013.pdf.]