*Manuscript submitted to the Offshore Technology Conference 2016, Houston.*

**OTC-27039-MS**

# Efficient Environmental and Structural Response Analysis by Clustering of Directional Wave Spectra

M. Vogel, Shell Global Solutions US Inc
J. Hanson, WaveForce Technologies LLC
S. Fan, Shell Global Solutions US Inc
G.Z. Forristall, Forristall Ocean Engineering Inc
Y. Li, Shell International Exploration and Production Inc
R. Fratantonio, RPS ASA
P. Jonathan, Shell Projects & Technology

## Abstract

Estimation of environmental and complex structural responses, such as fatigue for risers on deepwater floating production systems, is a critical and generally computationally intensive process. Long term damage estimates require the determination of host vessel motions used for riser stress calculations. In principle, riser stress could be calculated for each of a large number of directional sea states, a considerable computational burden. However, it might be possible to identify a representative subset of directional sea states for vessel motion and subsequent riser stress analysis, such that estimated fatigue characteristics (from the full set of sea states and the subset thereof) were equivalent. This would be advantageous as it would require considerably less computational effort.

In this work we use non hierarchical K-MEANS cluster analysis to partition a large set of directional wave spectra for contiguous sea states at a location offshore Brazil, corresponding to a period of approximately 2 years into a number of clusters. We adopt the set comprised of cluster centroids only as representative sea states for efficient characterization of the environment and structural response.

We demonstrate that the representative sea states provide an efficient basis for estimation of overall sea state bulk, wind sea and swell characteristics. We evaluate the effect of cluster size on the performance of the representative sea states using custom built visualization tools utilizing the Kolmogorov-Smirnov test statistics. The representative sea states are further used as input for a VLCC-class FPSO vessel motion analysis. For heave at the turret, roll motions, and relative vessel heading, distributions of vessel motions from analysis of representative sea states are in excellent agreement with those from analysis of all sea states. Guidelines for the application of the methodology are provided.

# 1. Introduction

Estimating the response of ocean structures subject to environmental loading is in general a complex but important task. For example, estimating riser fatigue on deepwater floating production systems is a critical and generally computationally intensive process. When interest lies in estimating long term damage, riser stress might be calculated for each of a large number of directional sea states. This is a considerable computational burden.

Cluster analysis provides a way of grouping a (typically large) number of individuals (such as directional spectra for consecutive sea states) into a (typically small) number of clusters, such that all members of the same cluster have similar characteristics. Each cluster has a cluster centre (strictly the centroid or centre of mass). All members of a given cluster are more similar to the cluster centroid for that cluster than to any other cluster centroid. It is reasonable therefore to assume all the members of the cluster might be represented by just one directional spectrum, namely that of the cluster centroid (or any directional spectrum near to the cluster centroid). Further, the information contained in the (small) set of cluster centroids might be a reasonable approximation for the information in the original (large) set of directional spectra.

We might expect that the distribution of values of any environmental variable of interest such as significant wave height ($H_S$, or peak period $T_P$, or swell $H_S$, or $H_S$ for a particular directional sector, or mean wave direction $D_M$) for the original (large) set of directional spectra could be approximated by the distribution of values of the same variables calculated from the (small) set of directional spectral centroids, appropriately weighted to take account of the relative size of each of the clusters. For example, a bulk sea state variable like $H_S$ is easily calculated once the directional spectrum is known, so that approximating the full distribution of $H_S$ using the weighted distribution from cluster centroids is not so interesting. However, there are many important environmental variables (such as swell $H_S$, estimated from spectral partitioning) and engineering design variables (such as structural responses) which are not easily calculated from the directional spectrum. Approximating full distributions for these variables using weighted distributions from cluster centroids might represent a huge reduction in the computational complexity of estimating design values.

Specifically, we expect that the statistical distribution of any complex response of a marine structure, subject to ocean environments characterised by each of the original (large) set of directional spectra, is well-approximated by the appropriately-weighted distribution of the same response for the cluster centroids. Therefore, given response analysis for the (small number of) cluster centroids only, we can estimate the full distribution of structural response. An approach similar to this, applied to characterisation of current profiles, has been reported previously in the ocean engineering literature (Prevosto et al. 2012, Jeans et al. 2015).

The objective of this paper is to demonstrate that full distributions of environmental and structural response variables can indeed be well-approximated using weighted distributions of the same variables for cluster centroids only. We outline the approach in application to the estimation of design values for a FSPO offshore Brazil. The layout of the paper is as follows. In Section 2, we describe the observed wave climate at the location, for which a large set of directional spectra are available. Section 3 outilnes the vessel response analysis, and Section 4 provides a summary of the approach to cluster analysis used. In Section 5, we estimate cumulative distribution functions for environmental and vessel response characteristics, by brute-force response analysis for all sea states, and by cluster analysis of all sea states followed by response analysis of cluster centroids only. Section 6 provides a discussion of results, and conclusions of the study.

# 2. Observed Wave Climate

*2.1 Buoy Observations*

Data for this study include two years of wave records from a Wavescan metocean buoy deployed offshore of Brazil at measurement station BC-10 depicted in Figure 1. Station BC-10 is located in Campos Basin at approximately 21 15' S, 039 45' W with a water depth of 1750 m.

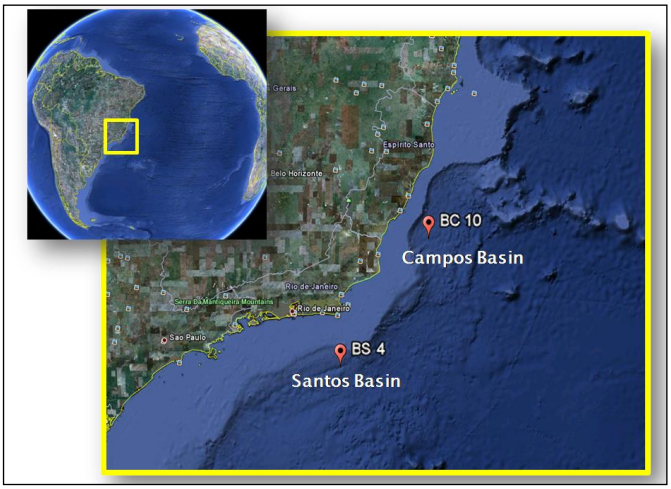


**Figure 1.** Brazil Campos Basin offshore study area with BC-10 station location

The metocean buoy was outfitted with a heave/slope Motion Reference Unit (MRU) for wave direction measurements and a meteorological station. A total of 17,086 hourly wave records were obtained spanning an approximate two-year time period from 1 May 2006 to 17 June 2008. As is common with data collected at sea, there are occasional gaps in the data records. The most significant gaps in the wave data are listed in Table 1.

**Table 1. Significant Observational Data Gaps**

| Station | Start Date | End Date |
|---|---|---|
| BC-10 | 4 NOV 2006 | 1 DEC 2006 |
| BC-10 | 16 MAY 2007 | 17 JUN 2007 |

The Wavescan data were prepared for analysis by converting raw spectral coefficients to directional wave spectra using the maximum likelihood method (Oltman-Shay and Guza, 1984). Resulting spectra were interpolated to 0.01Hz frequency resolution and 10 degree directional resolution. A weighted-average smoothing of the directional wave spectra over a three hour window reduced sampling noise; this generates wave partitions evolving more consistently over time. Sample spectra appear in Figure 2.

Our structural response study required careful preparation of the resulting wave spectra through spectral partitioning (Hanson and Phillips, 2001; Hanson and Fratantonio, 2015). The process of spectral partitioning is used to separate the sea and swell wave components from the directional wave spectra. We employed a watershed algorithm with directional wave age criterion to identify and combine wind sea peaks in the partitioned spectra. An iterative smoothing/combination approach (Portilla et al., 2009)

reduced the number of partitions down to a single sea and swell wave component. The final partitioning step was application of a 7 hour forwards-and-back trend analysis to further reduce discontinuities in evolving sea and swell components resulting from fluctuating winds.

Standard wave parameters computed from the original (bulk) and partitioned spectra include significant wave height $H_S$ (approximated by $H_{m0}$), peak wave period $T_P$ (estimated using a 3-point parabolic fit), and mean wave direction $D_m$ (see O'Reilly et al. 1996). Unless noted, all wind and wave directions reported are directions from which winds and waves propogate respectively, in units of degrees clockwise from North. Note that only original directional spectra are used as input to cluster analysis. Bulk sea-state characteristics (e.g. $H_S$, $T_P$) and partitioned characteristics (e.g. swell $H_S$, $T_P$) are regarded as responses. Interest lies in estimating the statistical distribution of responses as efficiently as possible.

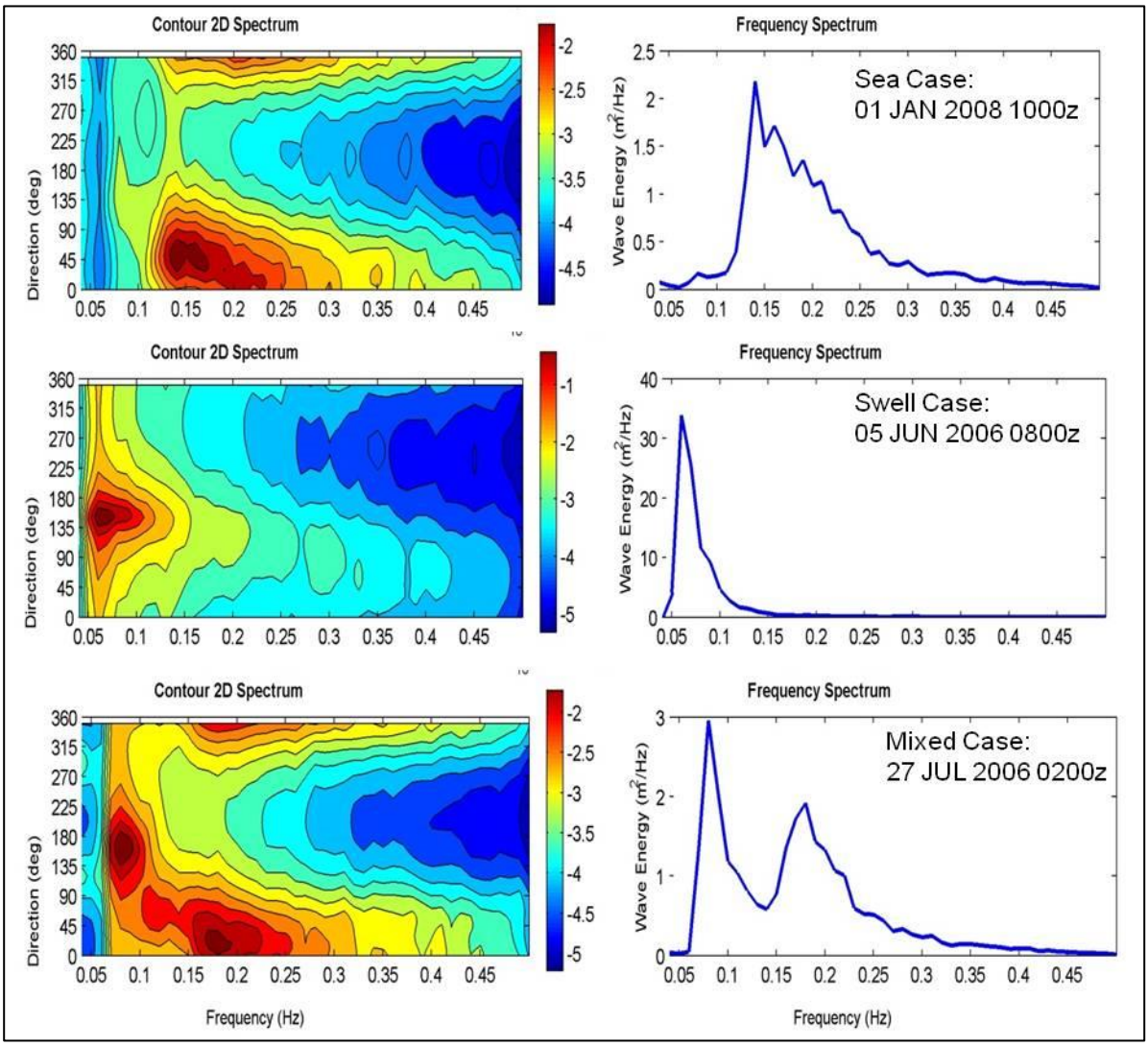


**Figure 2.** Sample directional (left panels) and non-directional (right panels) wave spectral records including a wind sea case (upper panels), swell case (middle panels), and a mixed sea and swell case (lower panels).

## *2.2 Wavefield Description*

The Brazil offshore wavefield is composed of a dynamic mix of local wind sea and swell from regional and remote sources. Basic statistics from the two-year wind and wave record appears in Table 2. Significant wave heights average is 1.8m and peak period is 9.4s, with an extreme wave height of 5.7m

at 15.8s period. Wind speed and significant wave height roses appear in Figure 3. Mean winds are approximately from the north-northeast. It is noteworthy that the principal wave directions appear to be bi-modal, with some wavefield components approximately aligned with the wind, and additional energy propagating from the south.

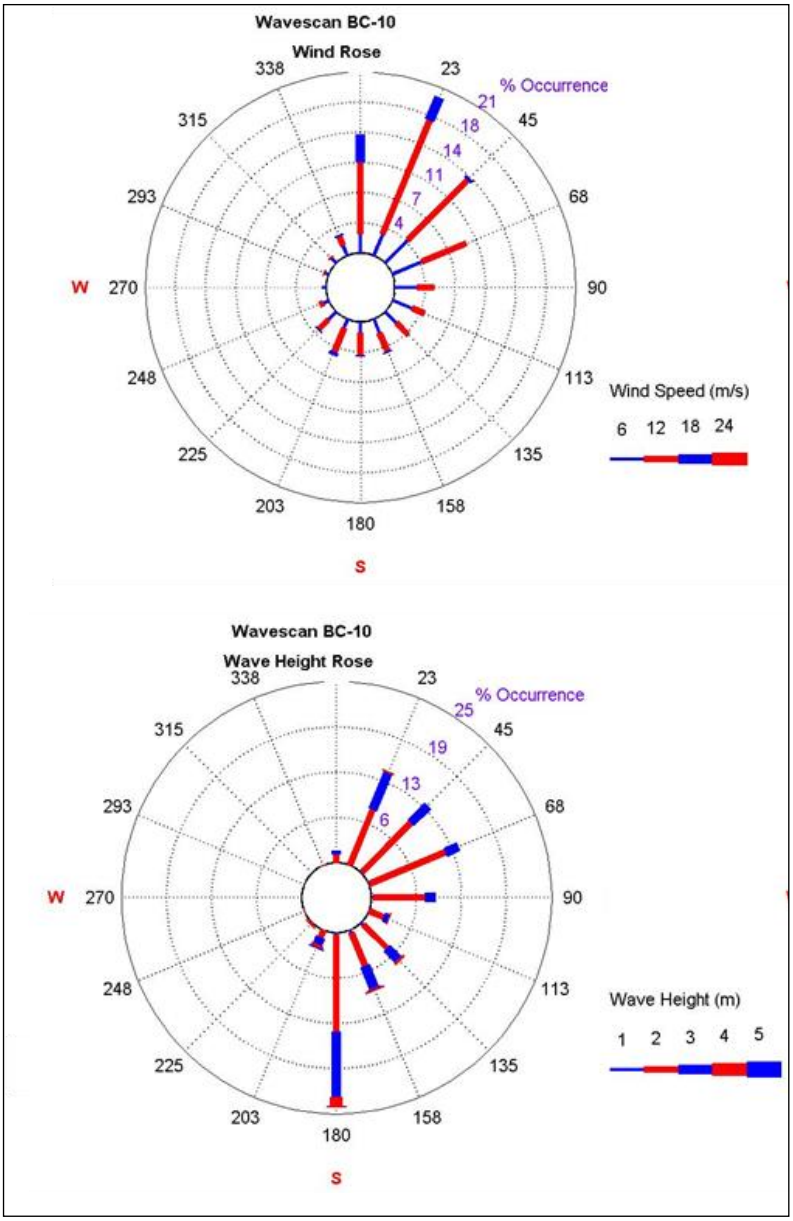


**Figure 3.** Summary wind (upper) and wave (lower) roses at BC-10.

To provide further insight into the wavefield climatology, a statistical summary of the wave partition analysis results is necessary. Separate histograms for wind sea and swell height, period and direction at each station are provided in Figure 4.

The site is dominated by swell, with significantly greater numbers of swell events than wind sea events. The wave period histograms suggest a broad distribution of wave periods, with the most frequent being 11-12s period swell. The wave direction histograms depict a bi-modal distribution, as was indicated by the bulk wave height roses of Figure 3.

**Table 2. Bulk Wind and Wave Statistics**

| Descriptor | Mean | Extreme | Standard Deviation |
|---|---|---|---|
| **Wind Speed (m/s)** | 4.0 | 17.6 | 3.1 |
| **Wind Direction (deg)** | 38.2 | 234.1 | -- |
| **Sig. Wave Height (m)** | 1.8 | 5.7 | 0.5 |
| **Peak Wave Period (s)** | 9.4 | 15.8 | -- |
| **Mean Wave Direction (deg )** | 103.3 | 160.0 | -- |

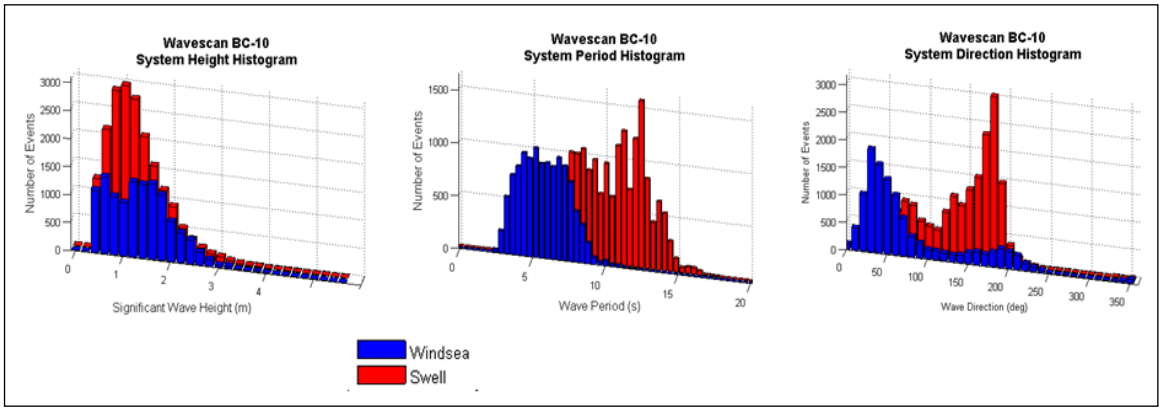


**Figure 4.** Wave system (wind sea and swell) histograms for wave height (left), peak period (center), and wave direction (right).

Here we have confirmation that the northeast waves are primarily wind sea, and the southerly waves are predominantly swell. The southerly swells originate from intense storms at higher latitudes (farther south), primarily during the southern hemisphere winter.

Further detail on the mix of wind sea and swell in the Brazil offshore environment is revealed by a joint-occurrence analysis. The joint occurrences (%) of wind sea and swell are given in Table 3. In this case the partitioning analysis was set up to find a wind sea and one or two swells in each wave record. Only in 17% of records does wind seas occur in the absence of swell, so that at least one swell component is present in 83% of records. Swell in combination with wind sea occurs in 52% of records. Two swell systems occur 29% of the time. There are no instances of calm seas (Significant wave height<0.3 m) in the records from BC-10.

**Table 3. Joint Occurrence Table (%) for BC-10 Wind sea and Swell**

| Conditions | Primary Swell | Primary & Secondary Swell | No Swell | Total |
|---|---|---|---|---|
| **Wind Sea** | 42 | 10 | 17 | 69 |
| **No Wind Sea** | 12 | 19 | 0 | 31 |
| **Total** | 54 | 29 | 17 | 100 |

A significant point from Table 3 is that secondary swell is present nearly 30% of the time at BC-10. This implies that for a standard two-component analysis, nearly 30% of the swell events will actually be a combination of two swell events. When present, secondary swells at BC-10 typically account for 6-25% of the total wave energy in the wavefield.

## 3. Response Analysis

In order to demonstrate the effectiveness of the spectral clustering approach, statistical distributions for vessel motions (surge, sway, heave, roll, pitch and yaw) of the FPSO Espirito Santo at the BC-10 field offshore Brazil were estimated from directional wave spectra for all individual sea states in the original measured two-year sample. For each sea state, vessel motion spectra were calculated using Response Amplitude Operators (RAOs), which are transfer functions of wave period and direction. To justify this approach we make the first order assumption that vessel motion is linearly related to surface waves. Then vessel motion in irregular waves may be calculated by summing motions from regular waves of different amplitudes, frequencies and potentially directions of propagation corresponding to the sea state of interest. Figure 5 illustrates the transfer function principle.

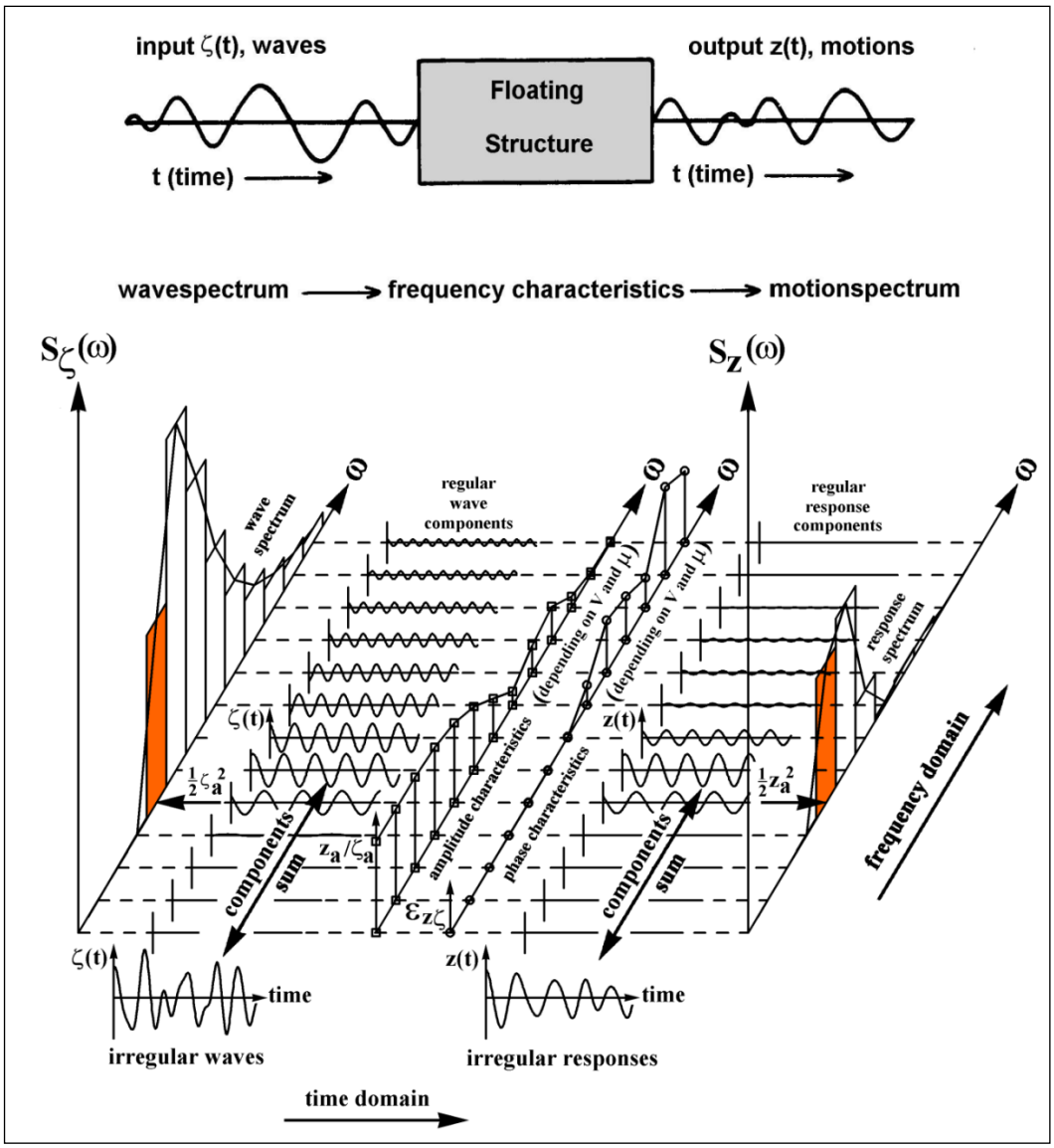


**Figure 5.** Principle of transfer of waves into vessel responses (from Journee and Massie 2001)

Figure 6 shows heave, pitch, roll, surge, sway and yaw RAOs for the FPSO Espirito Santo.

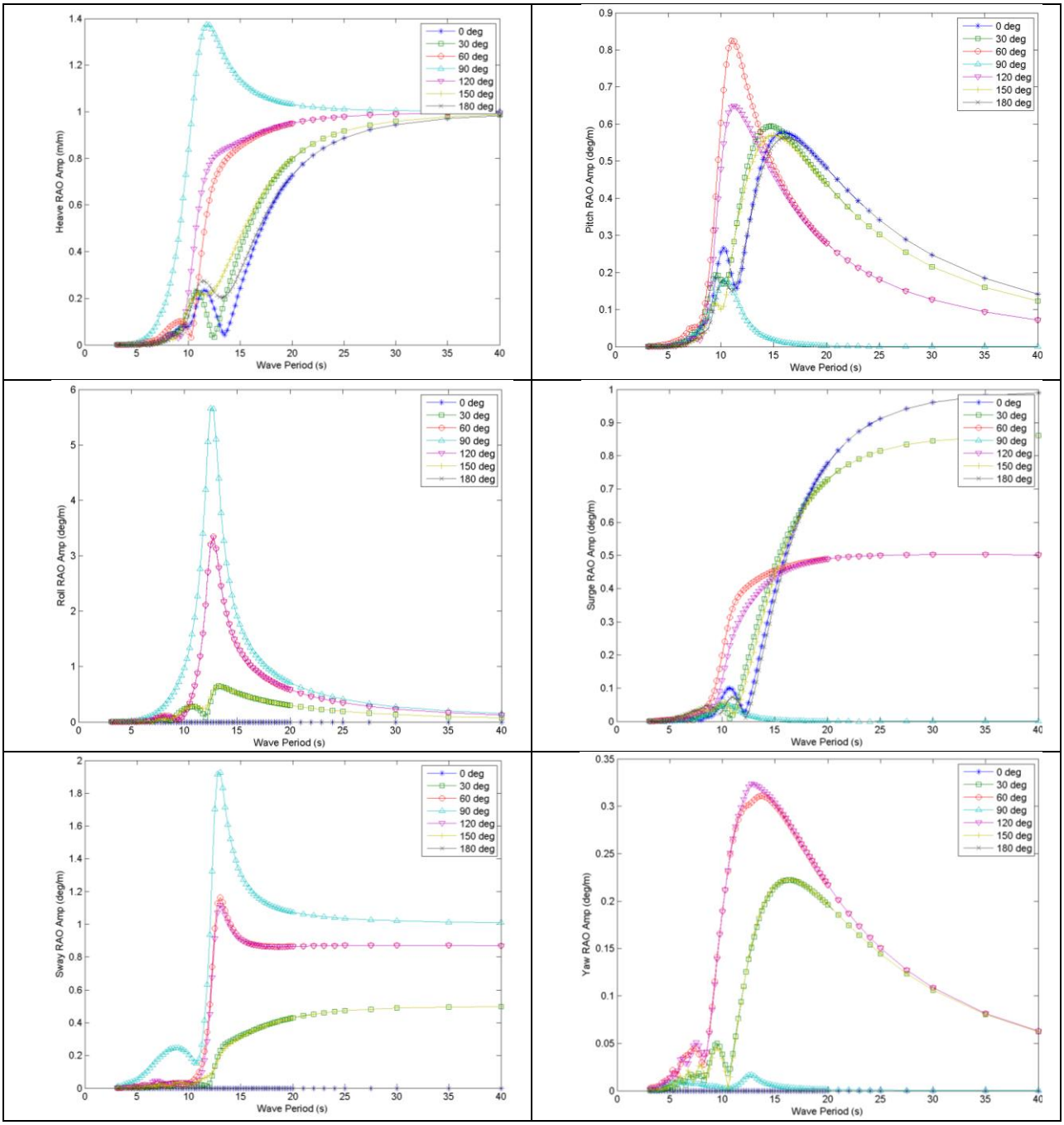


**Figure 6.** RAOs for the FPSO Espirito Santo

The resultant vessel motion spectra were integrated and significant responses estimated using

$$R_S = 4\, m_0^{1/2}$$

where $m_0$ is the zero-order moment of the vessel motion frequency spectrum, equal to the integrated response. The maximum vessel response $R_{max}$ in a three hour sea state associated with each significant response $R_S$ in that sea state was derived in an analogous way to maximum wave heights, using

$$R_{max} = R_S\{(1/2)ln(10800/T_Z)\}^{1/2}$$

where $T_Z = \{m_0/m_2\}^{1/2}$ and $m_2$ is the second response spectral moment.

## 4. Cluster Analysis

The input to the cluster analysis is a set of $n$ directional spectra $X_i$, $i = 1,2,\dots,n$, where each $X_i$ is a $n_f \times n_d$ matrix of wave energy for a sea state resolved into $n_f$ frequencies by $n_d$ directions. Cluster analysis summarises the information in the $n$ spectra in terms of $m$ cluster centroids $\mu_j$, $j = 1,2,\dots,m$, where $m \ll n$. In the cluster analysis, each of the $n$ observations is allocated to exactly one of the $m$ clusters $C_j$, $j = 1,2,\dots,m$, the centroid of which is $\mu_j$, and the number of members of which is $|C_j|$. Allocation of spectra to clusters, and estimation of cluster centroids $\mu_j$ is done by minimising some criterion quantifying the dissimilarity between two spectra. In the current work, we find cluster allocations and centroid estimates such that the within-cluster sum of squares $I^2$

$$I^2 = \sum_{j=1}^{m} \sum_{X_i \in C_j} \|X_i - \mu_j\|^2$$

is minimised with respect to $C_j$, $\mu_j$, $j = 1,2,\dots,m$ for a given value of $m$, where $\|X\|^2$ indicates the sum of the squares of the elements of matrix $X$. There are many possible choices of clustering criterion $I^2$. Since the directional spectra quantify the distribution of wave energy by frequency and direction, in the present application we consider within-cluster sum of squares to be a natural measure of the difference between directional spectra, and therefore a suitable basis for clustering of spectra.

The fundamental motivation for this work is to show that examining the characteristics of sea states corresponding to the $m$ centroids only, and the structural responses of vessels for the centroid sea states, is a good approximation to examining the same characteristics and structural responses for the full set of $n$ spectra. In particular, we study whether the distributions of environmental variables (such as $H_S$, $T_P$) of interest, and of structural responses (such has vessel heave, pitch and roll) estimated using the $m$ centroid sea states are good approximations to the corresponding distributions estimated using spectra for all $n$ sea states. Note that cluster analysis here uses unpartitioned directional spectra only as input, and that bulk sea-state characteristics (e.g. $H_S$, $T_P$), partitioned characteristics (e.g. swell $H_S$, $T_P$) and vessel motion characteristics (e.g. heave and pitch) are all regarded as responses.

Cluster analysis is computationally NP-hard, but the K-MEANS algorithm is a useful pragmatic approach. The algorithm starts with a set of candidate clusters $C_j^0$, $j = 1,2,\dots,m$, then improves the cluster solution by moving spectra between clusters so as to reduce the value of within-cluster sum of squares $I^2$ until further exchange of spectra provides only negligible reduction in $I^2$. Since it is possible that different candidate clusters might provide different cluster solutions, it is important to explore the stability of K-MEANS cluster solutions with respect to candidate clusters.

The number of clusters $m$ must be pre-specified in K-MEANS clustering. It is essential therefore to explore the behaviour of $I^2$ as a function of $m$. As $m$ increases, we expect that $I^2$ decreases since there are more clusters available to explain the same set of spectra. However, it is often the case that a particularly suitable value $m_{Opt}$ of $m$ can be found: e.g. such that the rate of change of $I^2$ with $m$ is large for $m < m_{Opt}$ but that the rate of change of $I^2$ with $m$ is smaller for $m \geq m_{Opt}$, and that the value of $I^2$ at $m_{Opt}$ is relatively near to the value of $I^2$ for much larger values of $m$. In this situation, we claim that $m_{Opt}$ provides an optimal size for the cluster solution.

## 5. Application

We now illustrate the cluster analysis approach to estimate the cumulative distribution functions (cdfs) of environmental and vessel response variables for sea states from the location described in Section 2.

*5.1 Cluster analysis*

Cluster analysis of the original set of $n = 17{,}086$ directional spectra into $m \in [10{,}1000]$ clusters was performed using the sum of squares metric given in Section 4. As discussed below, the $m = 100$ cluster solution provides a good choice for this application. For this reason, we choose to illustrate the cluster solutions using $m = 100$ as a reference cluster solution size.

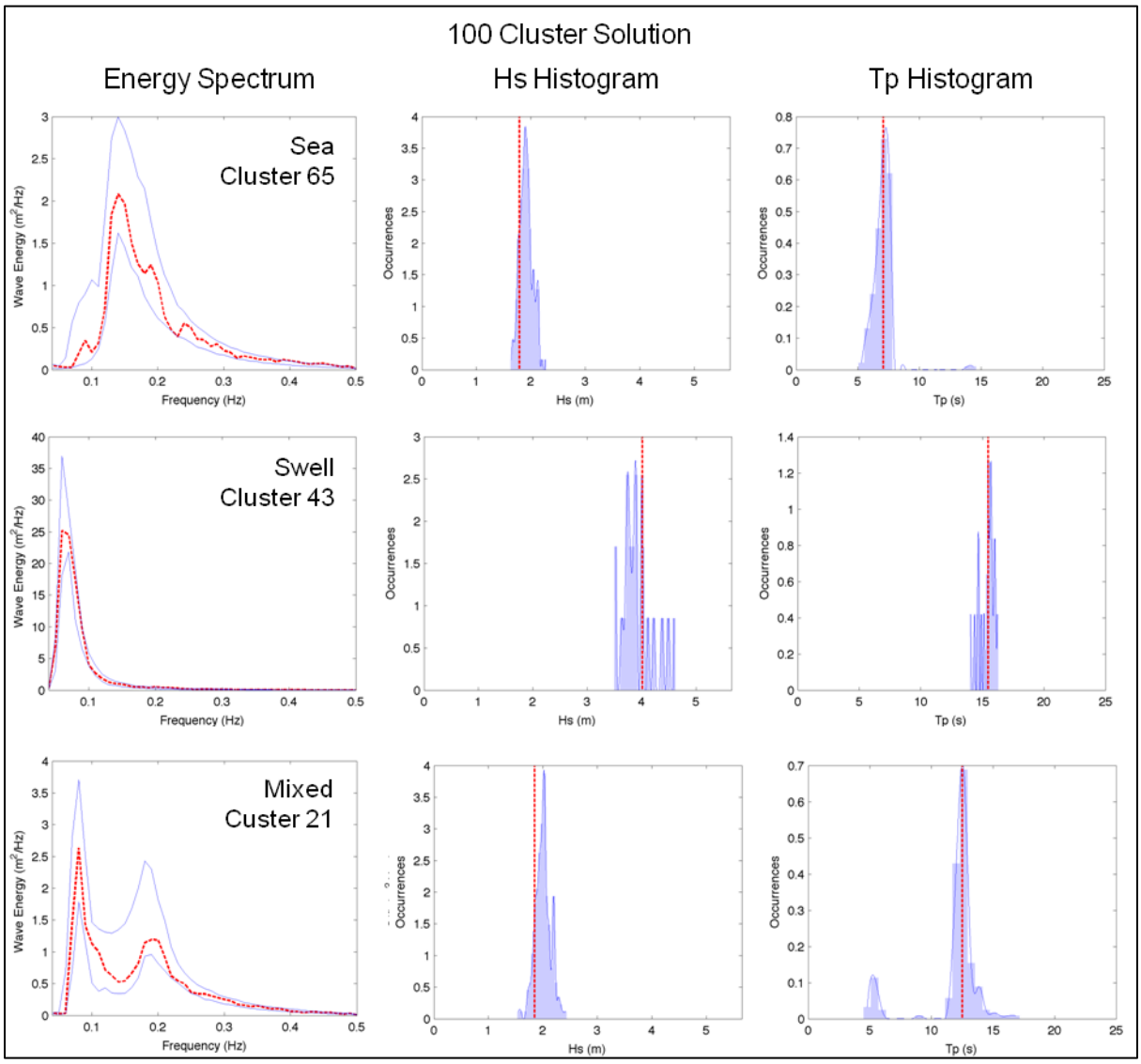


**Figure 7.** Summary statistics for three clusters from the 100 cluster solution, showing the frequency spectrum (left panels), significant wave height histograms (center panels), and peak wave period histograms (right panels) for wind sea (upper panels), swell (middle panels) and mixed wavefield conditions (lower panels). Cluster groups shown include the spectral records depicted in Figure 2. Left hand panels shows a 90% band (pointwise per frequency) for the cluster members (in gray) and the centroid spectrum (in red). The red lines in the center and right-hand panels correspond to cluster centroid values.

Figure 2 identifies three characteristic wind sea, swell and mixed sea states. It is interesting to examine the characteristics of the clusters into which these three sea states are allocated (clusters 65, 43 and 21 respectively, in this case). Figure 7 summarises the characteristics of the three clusters in terms of the

frequency spectrum, and histograms for $H_S$ and $T_P$. The red curves in Figure 7 give the value of the frequency spectrum, $H_S$ and $T_P$ for the cluster centroid. We observe that, by construction, all the sea states in cluster 65 are similar: they exhibit typical wind sea spectra, with $H_S$ near 2m and $T_P$ between 5s and 8s. Similarly cluster 43 contains typical swell spectra with $H_S$ near 4m and $T_P$ around 15s. Cluster 21 exhibits mixed wind sea and swell, with $H_S$ near 2m. The distribution of $T_P$ for cluster 21 is bi-modal, since the peak period can correspond to either the wind sea or swell component of the spectrum. It is clear that the cluster centroid $T_P$ in this case corresponds to the swell peak period near 12s.

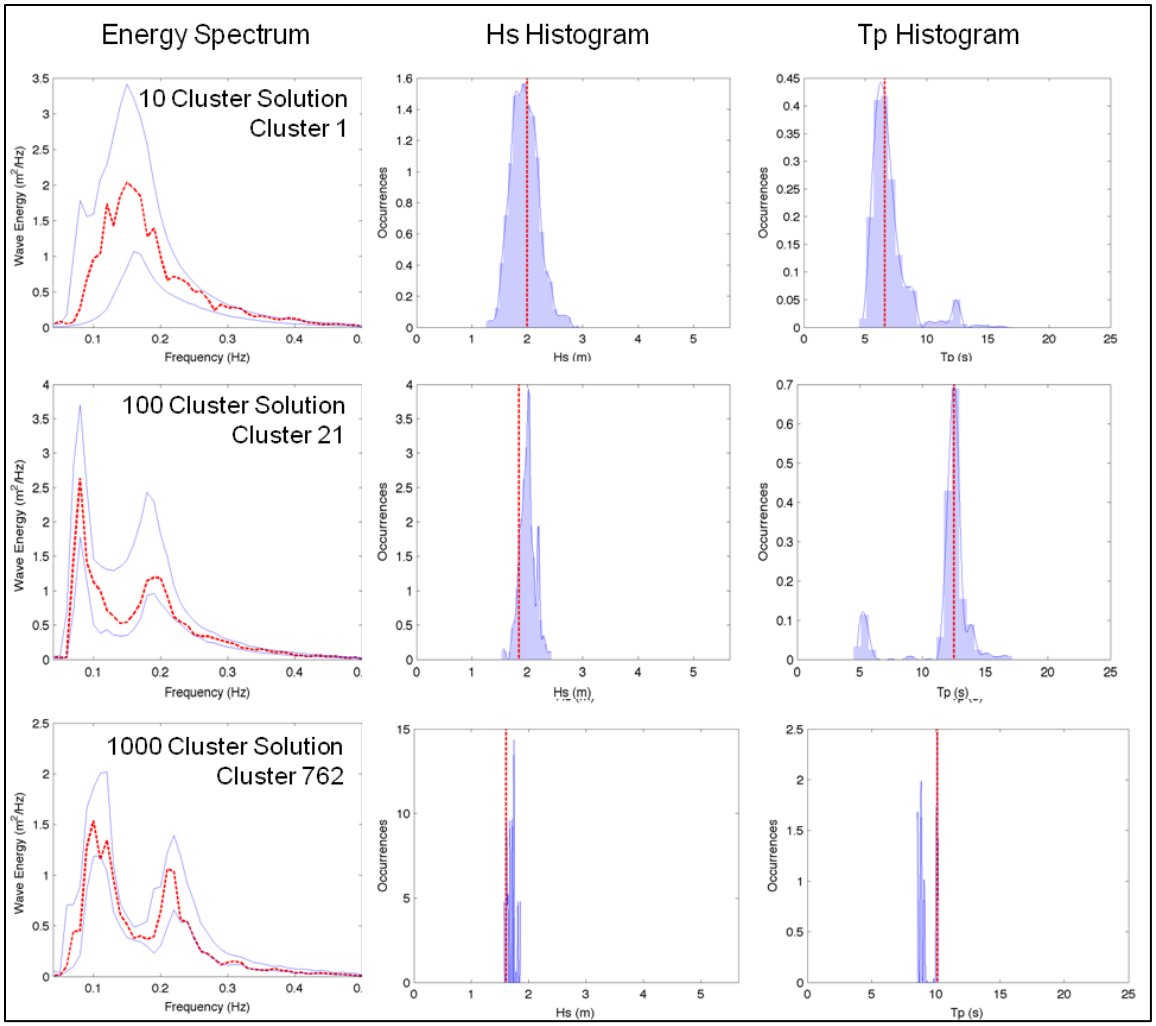


**Figure 8.** Characteristics of the cluster containing the mixed sea state from Figure 2, for 10, 100 and 1000 cluster solutions. Includes the frequency spectrum (left panels), significant wave height histograms (center panels), and peak wave period histograms (right panels). Left hand panels shows a 90% band (pointwise per frequency) for the cluster members (in gray) and the centroid spectrum (in red).

Figure 8 explores how the characteristics of the cluster containing the mixed sea state from Figure 2 change as we vary the size of the cluster solution $m$. When $m$ is small, e.g. $m = 10$, all spectra must be allocated into one of 10 clusters. The number of sea states $m_j$ in cluster $C_j$, $j = 1,2,\ldots,10$ must be relatively large therefore, compared (say) with the $m = 100$ solution. Further, we would expect the variation of frequency spectrum, $H_S$ and $T_P$ within the cluster to be relatively large for $m = 10$. Figure 8 confirms this. The centroid frequency spectrum does not correspond to a mixed sea state at all, but the cluster nevertheless contains mixed sea states. If we increase the size of the cluster solution to $m = 1000$, then the expected size of a cluster must be relatively small, and we would expect the variability in within-cluster characteristics to be relatively small. This is indeed the case from the figure.

Cluster analysis proceeds by minimising the clustering criterion $I^2$ (from Section 4) for given cluster solution size $m$. By design in general, we would expect that $I^2$ reduces with increasing $m$, since there are more clusters to explain the variability of directional spectra, and therefore less unexplained within-cluster variability. We also observe that the within-cluster variability of a function $G$ of the directional spectra (corresponding to $H_S$, $T_P$ or $D_M$, or any of the 6 maximum vessel responses) also reduces with increasing $m$. Figure 9 gives the value of within-cluster total root mean squared (RMS) error $I_G$ for typical environmental and structural responses, where

$$I_G^2 = \sum_{j=1}^{m} \sum_{X_i \in C_j} \left\| G(X_i) - G(\mu_j) \right\|^2$$

and for the case of $G$ representing the maximum heave response, $G(X_i)$ is the maximum heave for directional spectrum $X_i$, $i = 1,2,\dots,n$, and $G(\mu_j)$ is the maximum heave response corresponding to the cluster centroid spectrum $\mu_j$ of cluster $j$, $j = 1,2,\dots,m$. We observe that for all choices of $G$ considered the cluster analysis based on directional spectra $X_i$, $i = 1,2,\dots,n$ provides a reasonable basis for characterisation of environmental and structural response variables.

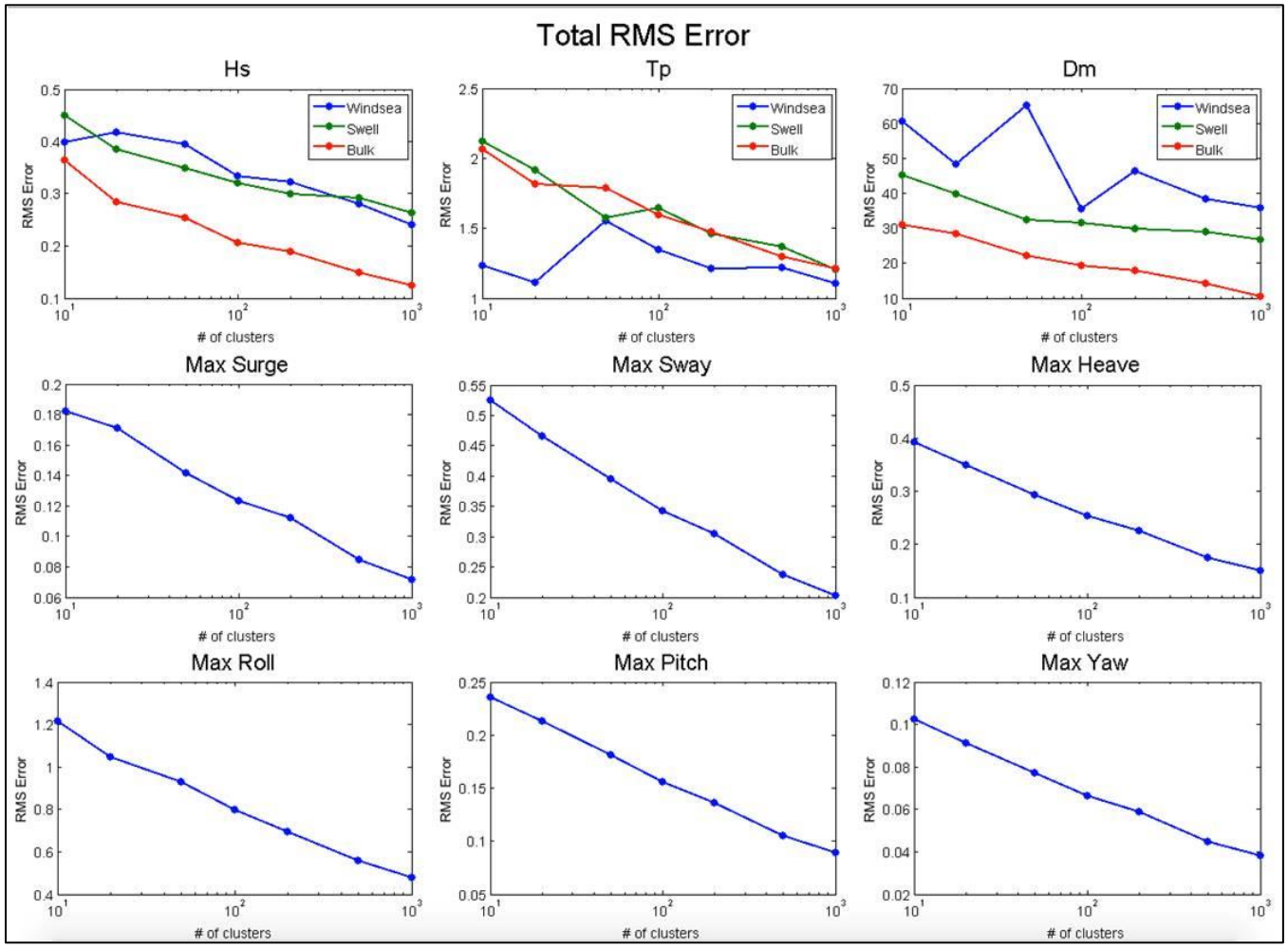


**Figure 9.** Within-cluster root mean square error for wave and vessel response variables, as a function of size of cluster solution.

*5.2 Estimation of statistical distributions for environmental and structural response variables*

Section 5.1 demonstrated that environmental and vessel response characteristics for cluster centroids (of directional wave spectra) provide reasonable approximations for the same characteristics as the full set of directional spectra. In this section, we explore whether cluster centroid characteristics can also be used to estimate the cumulative distribution functions (cdfs) of environmental and structural response variables. We define the empirical cdf $F_G$ of any environmental or structural response variable $G$ using the full set of $n$ directional spectra $X_i$, $i = 1,2,\dots,n$ as

$$F_G(g) = \mathbb{P}(G \le g) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}(G(X_i) \le g)$$

where $\mathcal{I}(x)$ is an indicator function, with value 1 if $x$ is true and 0 otherwise. To calculate $F_G$, we need to evaluate $G(X_i)$ for all $n$ directional spectra, which can be computationally demanding. We want to generate good approximations of $F_G(g)$ for all possible values of $g$ using the corresponding estimate $\tilde{F}_G$ based on the cluster centroids only, where

$$\tilde{F}_G(g) = \frac{1}{n}\sum_{j=1}^{m} \left|C_j\right| \mathcal{I}\left(G\left(\mu_j\right) \le g\right)$$

and $\left|C_j\right|$ is the number of directional spectra in cluster $C_j$ with centroid $\mu_j$, $j = 1,2,\dots,m$, and $\sum_{j=1}^{m}\left|C_j\right| = n$. To calculate $\tilde{F}_G$, we need to evaluate $G\left(\mu_j\right)$ just for the $m$ cluster centroids, which is computationally attractive especially when $m \ll n$.

As a baseline comparison, it is also interesting to examine estimates $\tilde{F}_G^R$ for the cdf based on random selections of $m$ spectra from the full set. The computational complexity of generating this estimate, in terms of the number of response analyses required, is the same as that of generating the cluster centroid estimate $\tilde{F}_G$. However, the $m$ spectra used for $\tilde{F}_G$ are special in that they are known to provide an optimal representation, in terms of $m$ spectra, of the full set of $n$ spectra. Spectra selected at random to estimate $\tilde{F}_G^R$ do not have this property. It is reasonable therefore to expect that $\tilde{F}_G$will be a better estimate for $F_G$ than $\tilde{F}_G^R$.

These comparisons are visualised in Figure 10, which shows estimates for the cdf of significant heave (left-hand side) and of maximum heave (right-hand side) based on the full set of $n$ spectra (in blue, $F_G$) and the set of $m$ cluster centroids (in green, $\tilde{F}_G$). Also shown is the range of values (pointwise with heave value) for 100 different estimates of $\tilde{F}_G^R$, each based on a random subset size $m$ of the full set of spectra. The three rows of the figure correspond to $m = 10$, $m = 100$ and $m = 1000$ cluster solutions. From the figure we observe that the quality of approximation of $F_G$ improves as $m$ increases for both $\tilde{F}_G$and $\tilde{F}_G^R$. However, we also observe that the cluster centroid estimate $\tilde{F}_G$ is always within the bounds corresponding to the random subsample estimate $\tilde{F}_G^R$. The figure also gives good guidance as to the relative lack of fit for $m = 10$, $m = 100$ and $m = 1000$. For the current application, with $n \approx 17000$, were computational constraints to limit the maximum number of response analyses to about 100, it is clear that the cluster centroid-based estimate is preferrable. However, if 1000 response analyses were feasible, for the current application the relative difference between centroid-based and random subsample-based estimates is small, and potentially negligibly so.

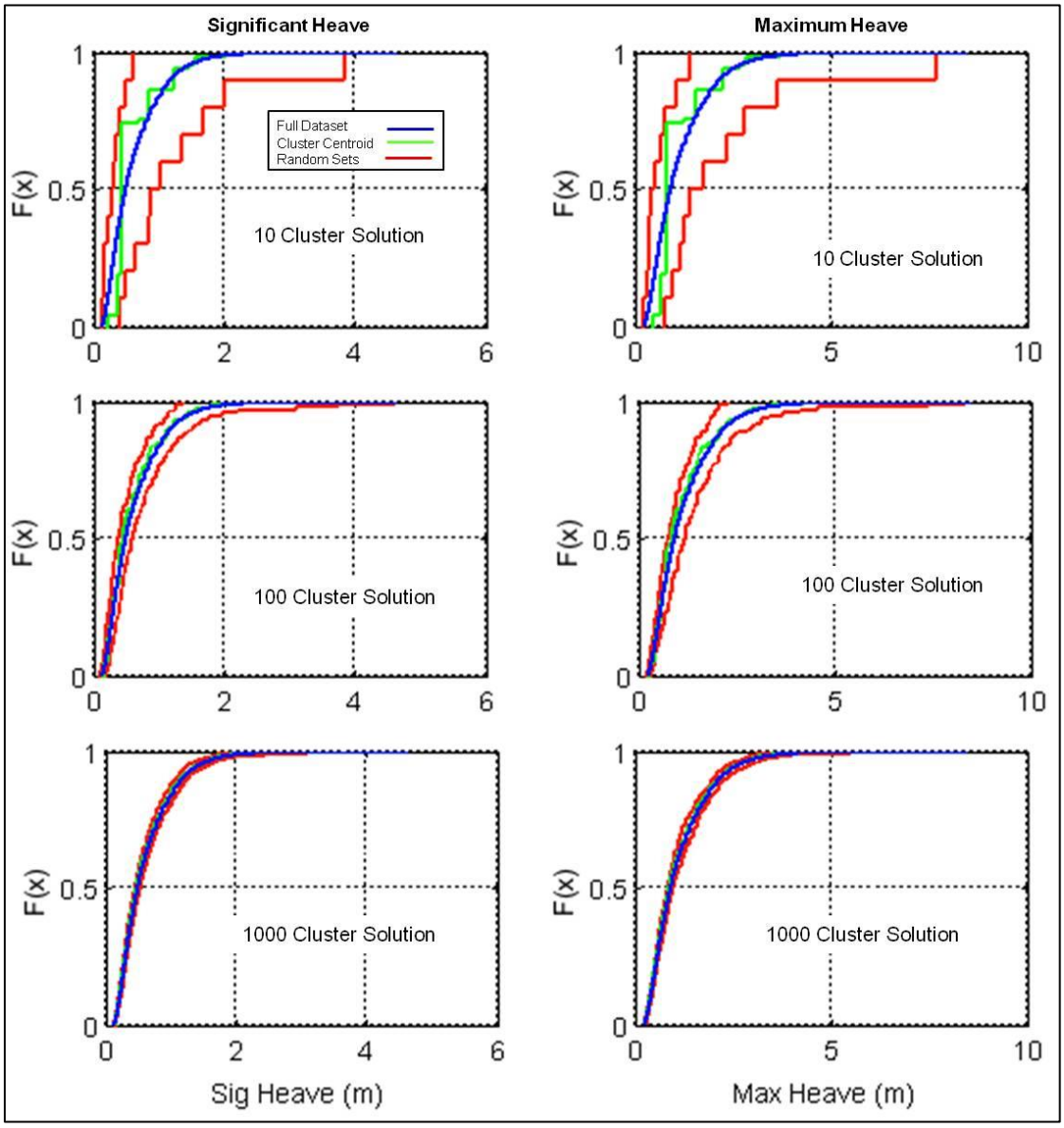


**Figure 10.** Empirical cumulative distribution functions for significant heave (left) and maximum heave (right) for 10 (upper), 100 (middle) and 1000 (lower) cluster solutions. Estimates based on full set of directional spectra (blue) and cluster centroids only (green). The red band is a pointwise range for the cdf estimated from 100 different random subsamples of the full sample.

Figure 11 shows the $m = 100$ cdf estimate for $H_S$, $T_P$ and $D_M$, and the 6 maximum vessel responses. We observe that the cluster centroid-based estimate $\tilde{F}_G$ provides a better approximation to full sample cdf $F_G$ than the random subsample-based estimate $\tilde{F}_G^R$ in each case.

The Kolmogorov-Smirnov (KS) statistic provides a straight-forward means for quantifying the similarity of two cdfs. The KS statistic is the maximum absolute difference between two the values of two cdfs $F_1$ and $F_2$ for any single value $g$ in the common domain of the cdfs

$$K = \max_g |F_1(g) - F_1(g)|$$

We can use the KS statistic $K$ to summarise the lack of fit of the cluster centroid-based estimate $\tilde{F}_G$ to the full sample cdf $F_G$for various choices of environmental and structural response variable $G$, and examine the behaviour of $K$ as a function of the size $m$ of cluster solution.

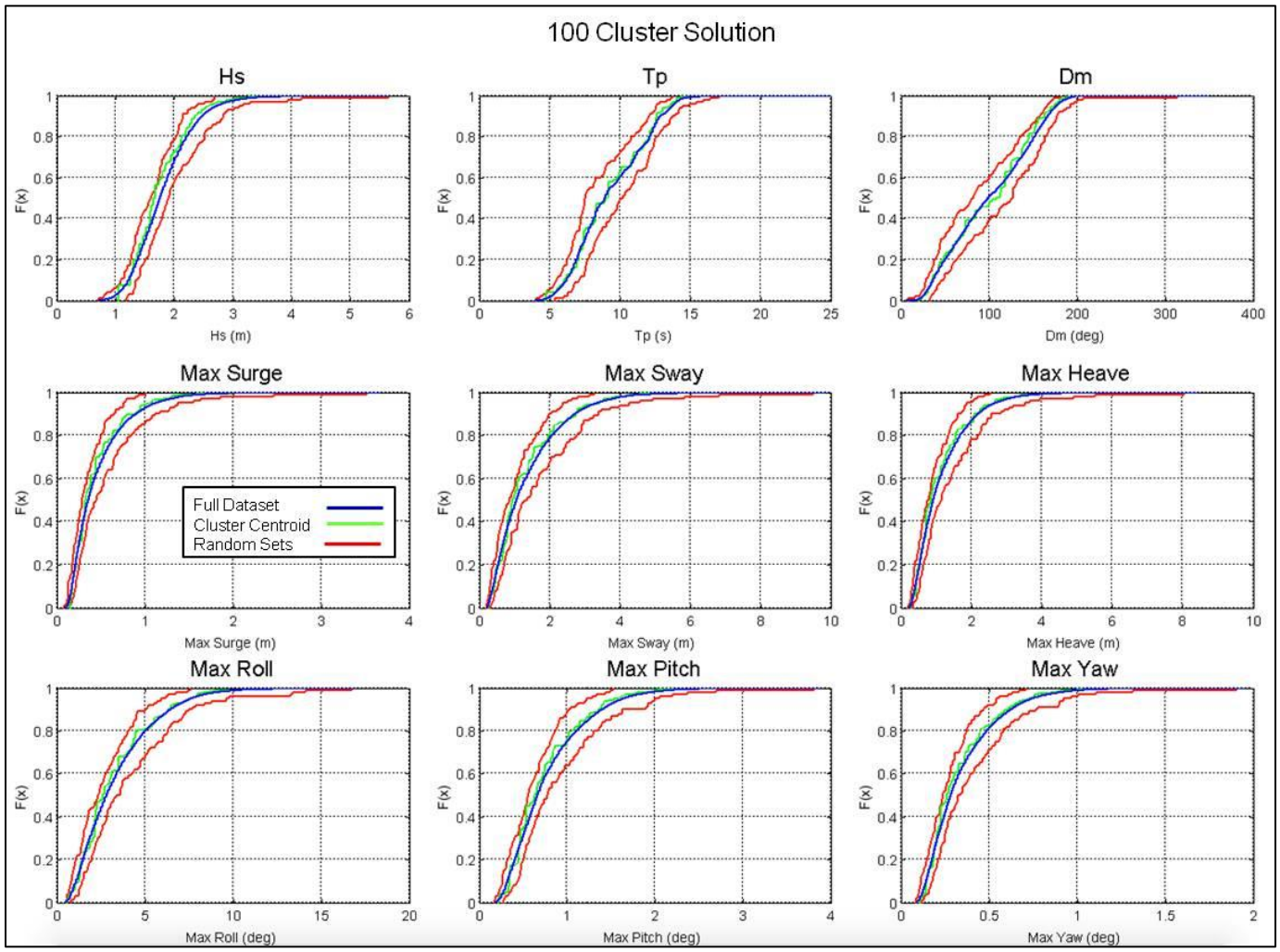


**Figure 11.** Empirical cdfs for significant wave height, peak wave period and mean wave direction (top), maximum surge, sway and heave (middle) and maximum roll, pitch and yaw (bottom) estimated using the full set of directional spectra (blue) and cluster centroids from the 100 cluster solution. Also shown is the range of estimates from 100 different randomly-sampled sets of 100 directional spectra (red).

Figure 12 illustrates that the value of $K$ decreases as a rule with increasing $m$ for $H_S$, $T_P$ and $D_M$ (wind sea, swell and total for each) and the 6 maximum vessel responses. This confirms that the quality of fit of $\tilde{F}_G$ to $F_G$ improves with increasing $m$ in general.

As discussed in Section 4, it is sometimes (but not always) possible to identify a particularly favourable value $m_{Opt}$ for the size $m$ of the cluster solution. From Figure 9, we observe that $I^2$ reduces monotonically with increasing $m$ for all maximum responses. However, the same is not true for the KS statistic $K$ in Figure 12; the absolute gradient of $K$ with $m$ is larger for $m < 100$, but smaller for $m > 100$ for a number of the variables examined, suggesting that perhaps $m = 100$ is a suitable cluster solution size for the current application, especially if estimation of $F_G$ for the variables illustrated is a priority.

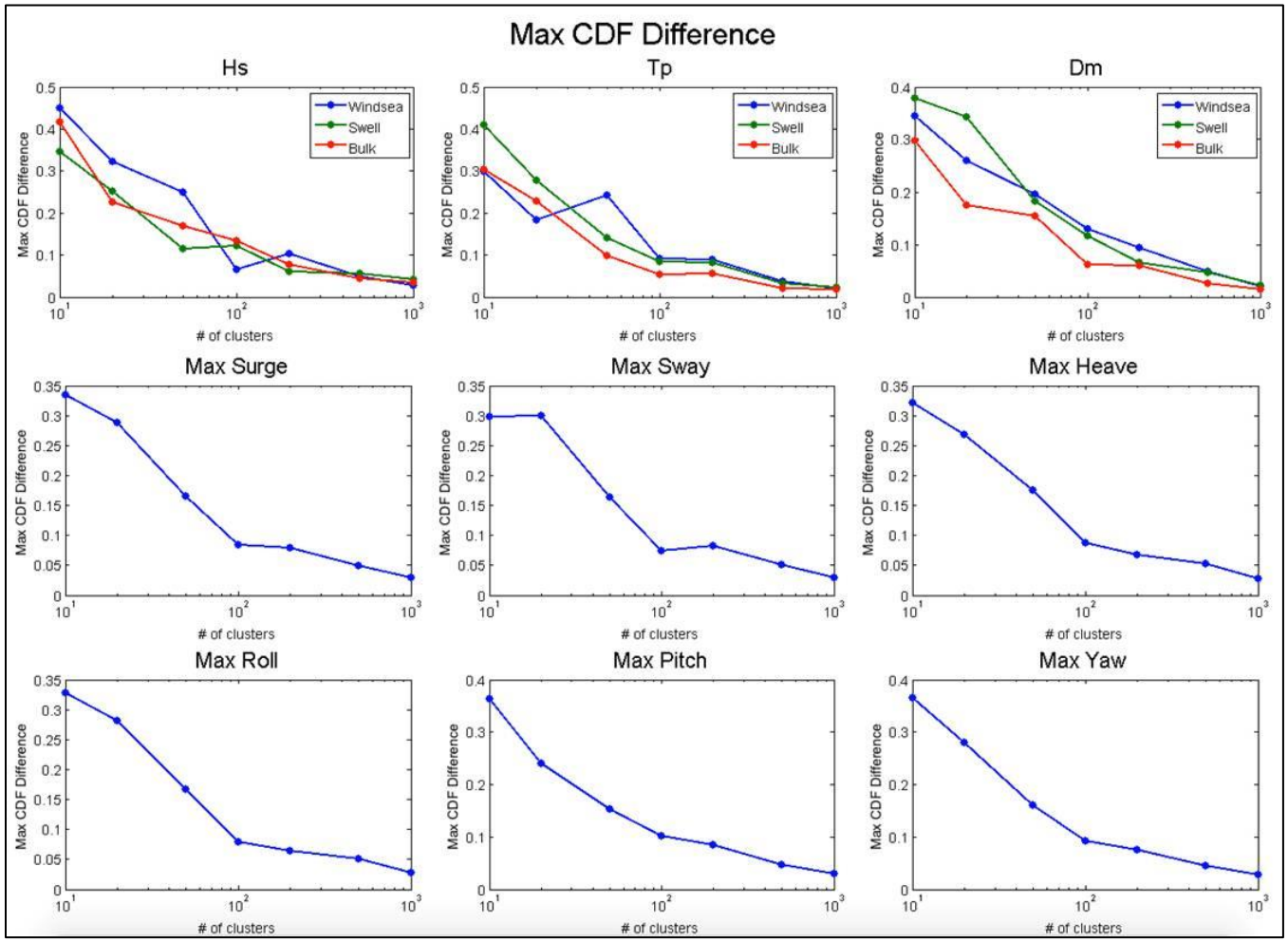


**Figure 12.** Values of Kolmogorov-Smirnov test statistic, comparing cluster centroid-based estimates for cdfs of environmental and vessel response variables, with those based on the full set of directional spectra.

## 6. Discussion and Conclusions

Some ocean basins, such as the Gulf of Mexico, are dominated by wind-driven sea states. In these basins, extreme environments are dominated by wind-wave properties with relatively common directional wave spectral characteristics. This is not the case at other locations, such as offshore Brazil, where seas are dynamic and multi-modal, with competing wind sea and swell effects at play. To characterise such environments well, more sophisticated techniques are needed to identify a small yet relatively representative set of directional wave spectra. In this work we demonstrate that statistical distributions of environmental and structural response variables corresponding to a large set of sea-state directional wave spectra can be well-approximated using weighted distributions of those variables for a small set of representative spectra only. We used representative spectra corresponding to the centroids from a K-MEANS cluster analysis of the full set of spectra. When the number of clusters in the cluster solution is appreciably smaller than the number of original sea states, this results in a computationally-efficient approach to estimation of distributions of complex structural responses in particular.

There are many possible approaches to estimation of small representative subsets of a large set of individuals, such as sea-state directional spectra; cluster analysis is one possible approach, and K-MEANS is one amongst many clustering algorithms, attractive because of its simplicity. When the

dimensionality of the description (i.e. the total number of frequency × directional increments of the directional spectrum) is large, it can be advantageous computationally to first reduce dimensionality by performing principal components (or empirical orthogonal function) analysis of the spectra, and then perform cluster analysis on the reduced data. Jeans et al. (2015) reports the application of K-MEANS to clustering of current profiles. Prevosto et al. (2012) propose using the use of self-organising maps (Liu and Weisberg 2011, Barbariol et al. 2015) as an alternative to cluster analysis, again in application to current profiles.

We outline the approach in application to the estimation of design values for a FSPO offshore Brazil. Here, the wave climate can be characterized as bi-modal with wind seas from the northeast and swell from the south. The wavefield is swell-dominated in winter, and wind sea dominated in summer. Although a wind sea and multiple swell systems occasionally occur simultaneously, a two-component wavefield is generally adequate to describe the offshore environment. For this application, using within-cluster sum of squares as the K-MEANS clustering criterion, we find that cluster centroids from a 100-cluster solution provide a relatively good approximation to the cumulative distribution functions of significant wave height, peak wave period and mean wave direction, and for the 6 maximum vessel responses. In applying the approach, the user is required to specify the clustering minimisation criterion (within-cluster sum of squares here) and an appropriate size of the cluster solution. Since the elements of are directional spectrum are on the energy scale, a within-cluster sum of squares criterion seems appropriate. Choice of the size of the cluster solution is in general more challenging, and should take account of all of the following considerations: (a) the uncertainty in the estimated cumulative distribution function based on cluster centroids as a function of the size of the cluster solution, and (b) the computational resource required for a response analysis for one sea state.

The current work shows that spectral clustering is effective at estimating the statistical distributions of environmental variables derived from full and partitioned directional spectra. If our interest lies in estimation of the statistical distribution of (e.g.) swell $H_S$, spectral partitioning of directional spectra for all sea states is not necessary therefore, and only necessary for directional spectra of sea states corresponding to cluster centroids. The vessel motion analysis reported here assumes a linear relationship between wave input and vessel response; this is adequate for typical sea states and vessel motions. The current work demonstrates that spectral clustering is effective at estimating the statistical distributions of such motions. The transfer function approach may not be as appropriate for extreme environmental conditions and highly non-linear systems, for which full time-domain hydrodynamic response analysis might be required; it would be interesting to demonstrate the effectiveness of the current approach in such cases.

## References

Barbariol, F., F.M. Falcieri, C. Scotton, A. Benetazzo, S. Carniel, and M. Sclavo, Self-Organizing Maps approaches to analyze extremes of multivariate wave climate, Submitted to Ocean Sci. (2015).

Hanson, J.L. and O.M. Phillips, 2001. Automated analysis of ocean surface directional wave spectra, J. Atmos. Oceanic Technol. 18 277-293.

Hanson, J.L., and R. Fratantonio, 2015. XWaves Users Guide, Wave Force Technologies, *http://waveforcetechnologies.com/help/xwaves/getting_started.html.*

Jeans, G., R. Gibson, and O. Jones, 2015. A New Quantitative Assessment of Current Profile Clustering Methods for Riser Engineering, OMAE2015-41429, St. John's, Newfoundland, Canada.
Journee J. M. J. and Massie W. W., 2001. Introduction in offshore hydromechanics, First edition, Delft University of Technology.

Liu, Y., and R.H. Weisberg, 2011. A review of self-organizing map applications in meteorology and oceanography. In: Self-Organizing Maps-Applications and Novel Algorithm Design, 253-272.

Oltman-Shay, J. M., and R. T. Guza, 1984. A data adaptive ocean wave directional-spectrum estimator for pitch and roll type measurements, J. Phys. Oceanogr. 14 1800–1810.

O'Reilly, W. C., T. H. C. Herbers, R. J. Seymour, and R. T. Guza, 1996. A comparison of directional buoy and fixed platform measurements of Pacific swell. J. Atmos. Ocean. Tech. 13 231-238.

Portilla, J., F. O. Torres, and J. Monbaliu, 2009. Spectral partitioning and identification of wind sea and swell, J. Atmos. Oceanic Technol. 26 107-122.

Prevosto, M., G. Z. Forristall, G. Jeans, C. Herry, G. Harte, L. Harrington-Missin, and P. Dooley, 2008. Worldwide approximataions of current profiles for steel riser design – the WACUP project, OMAE2012-83348, Rio de Janeiro, Brazil.