

Global classification of wave height, period, and direction multivariate distributions using principal component analysis

Ophélie Choupin^{a*}, Joseph Harari^a, László Könözy^b

^a Institute of Oceanography of the University of São Paulo, São Paulo, Brazil.

^b Cranfield University, College Road, Cranfield, Bedfordshire, MK430 AL, United Kingdom.

* Correspondence to: Ophélie Choupin (o.choupin@net.estia.fr)

Abstract. Research on the classification of ocean surface wave patterns uses coarse spatial resolutions and approaches that are less usable for applications such as annual energy estimates for wave energy converters. Consequently, this study investigates the classification of 1-3-dimensional histograms of wave height, period, and direction at a 0.4° spatial resolution. Histograms of global grid-points have been decomposed into their main modes using principal component analysis. Each mode consists of a histogram and a map quantifying its contribution to each local wave pattern. The first mode displayed a meridional gradient of the wave height, a northeast-to-southeast gradient of the wave period, and a mix of both for the wave direction. Dominating wave systems consist of a) 1.9 m, b) 14.5 s in the basins' southeast, c) 12.5 s north and south, and d) 9.5 s north-west. Generated remotely, c) travel and evolve over long distances to reach coasts, as b) pools, while d) is mainly generated locally or northwards. K-means was used to cluster wave patterns into 15 classes using the 10 dominant modes. Some clusters characterise regions in different basins. 1-parameter-based clusters provide more regions and alternations of clusters nearshore, while clustering the parameters together provides larger regions and less nearshore noise. The wave height and period combined classification showed good correlation and difference between the cluster's average histogram and those of the grid-points in that cluster, while the direction was lower in regions affected by strong currents or topographic obstacles. Consequently, combining all parameters greatly decreases these two metrics.

HIGHLIGHTS

- Main wave systems of individual and joint wave height, period, and direction
- Meridional and zonal dominating patterns of the wave height and period climates
- High accuracy of each parameter's 15 clusters, except negative in identified areas
- Joint classifications split swells from seas whilst merging the northern oceans
- All approaches' classifications depict a meridional gradient of the Southern Ocean

KEYWORDS

Principal Component Analysis (PCA); Empirical Orthogonal Function (EOF); Global wave clustering; Wave height; Wave period; Wave direction; Wave Distribution Matrix (WDM).

1 INTRODUCTION

1.1 MOTIVATIONS

Ocean surface wave patterns come from a mix of local and remote wind generation. Many fields would benefit from a classification of the main wave climates since understanding the general climatology of waves can help prepare coastal and offshore structures, city management, and future ship routes, among others, to account for general wave patterns. Therefore, solid climatology analyses of waves are necessary to then help make projections of how such climates would evolve. Such maps can also improve understanding of the origin, propagation, and evolution of wave patterns worldwide. In the context of Marine Renewable Energy (MRE), the most accessible way to compute the energy production of Wave Energy Converters (WECs) is by using histograms of the wave parameters [1]. Hence, classifying the waves based on histograms, currently lacking in the literature, could foster the mapping and the synergies of WECs worldwide [2] by estimating the energy produced for each wave climate and each technology class or device directly [3]. The use of histograms to assess the global climatology of waves is the main contribution of this research.

1.2 WAVE CLIMATE DESCRIPTION AND SCIENTIFIC GAPS

Ocean surface gravity waves are generated by the balance between the friction and pressure the wind exerts on the water surface and gravity that attempts to bring the now unstable surface back to its mean level [4]. However, local wind profiles do not necessarily correspond to local wave climates. Based on the data from the Centre for Australian Weather and Climate Research (CAWCR) dataset of the Commonwealth Scientific and Industrial Research Organisation [5], Figure 11Figure 1 provides wind and wave information for three offshore key points worldwide. The three locations were selected to illustrate contrasting wind–wave relationships: a remote generation-dominated site in the Southern Ocean; a mixed-source point in the Eastern Pacific influenced by both hemispheres; and a region in the Bay of Bengal where local and remote wave characteristics diverge from the prevailing wind field. This selection highlights how wave climates are shaped by both local winds and long-range swell propagation.

The heating of Earth by the Sun and Earth’s rotation generate atmospheric circulation cells that lead to main wind trends: as the Earth spins to the right at the level of the equator, 30°S–30°N, the winds go westwards, with the Coriolis effect, the midlatitudes 30–60°N/S receive opposite winds going eastwards, and higher than 60°N/S, winds go back westwards [6]. These sources of wave generation, along with their temporal evolution in increasing opposition polewards between both hemispheres, lead to various wave-generation wind trends that are also complexified by other variables, such as the world’s topography. Figure 11Figure 1.d,f,i provide examples of 30-year mean 10-m high wind occurrence as a function of its direction, coloured by the dominant wind speeds, for Point 1, in the Bay of Bengal, 2, in the Eastern Pacific low latitudes, and 3, in the Southern Ocean belt, respectively (statistics of wind velocity and direction are provided in Appendix A). The wind roses highlight “westerlies” (eastwards winds, e.g., Points 3) and “easterlies” (westwards winds, e.g., Points 2). Point 1 shows an example of a location diverging from these patterns with its two-system wind profile. As waves come out from the generation area, they spread and evolve worldwide from an initial set of height, period, and direction functions of the wind velocity, coverage area (also called the fetch), and direction.

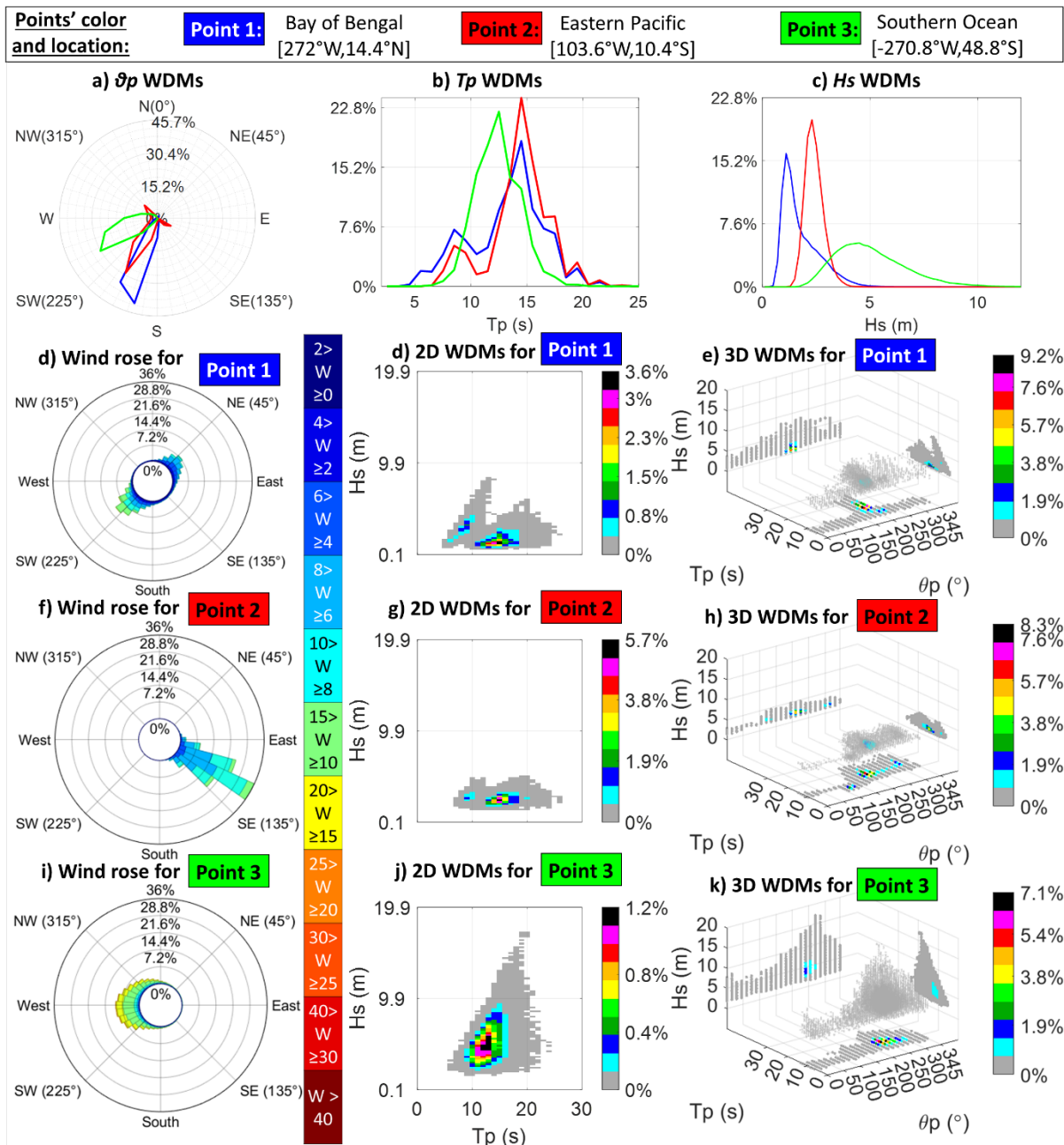


Figure 11. Selected three key points to illustrate the local 30-year occurrence (number of hours translated in percentage of 30 years of hours, hence in %) of wind direction and speed (W , m/s), panels d, f, and i, and for the significant wave height (H_s , m), wave peak period (T_p , s), and wave peak direction (θp , °). Individual θp , T_p , and H_s occurrences in 1-Dimensional Wave Distribution Matrices (WDMs) are provided in panels a-b. Joint H_s and T_p occurrences in 2-Dimensional (2D) WDMs in panels d, g, and j. Joint θp , T_p , and H_s occurrences in the centres of the 3-Dimensional (3D) WDMs are provided in panels e, h, and k. The three walls in the 3D panels provide the sum of the number of hours over the integrated axis, translated into percentages over the 30 years of hours. All directions indicate where the waves or winds come from.

Time series of bulk wave parameters, such as the significant wave height (H_s), peak period (T_p), and peak direction (θp), are widely used to track the temporal evolution of wave climates and highlight the changes or consistencies between spatially spread points. For instance, Dodet et al. [7] found contrasting trends across the 1953-2009 northeast Atlantic region (30-60°N), where southern and northern regions responded differently to the North Atlantic Oscillation due to the varying influences of the Azores High and the Icelandic Low. Generally, boreal winter exhibits higher H_s and longer T_p than summer, with a) stronger and positive variations of H_s and T_p northwards than in the south, and b) more southward wave origins in the Northern Hemisphere, resulting in

northeastward propagation. However, directional variations differ: counterclockwise seasonality is stronger in northern zones, while southern ones tend to shift clockwise interannually.

Global spatial patterns were further segmented by Gulev and Grigorieva [8], distinguishing, in the Northern Hemisphere, seas (generated locally) versus swells (generated remotely), and Chen et al. [9] highlighted the east-west basin split between swells in the east and pools of seas in the west, worldwide. The temporal dynamics of these systems vary: Gulev and Grigorieva [8] noted a swell-and-sea lower correlation (when not anti-correlated) in the North Atlantic than in the North Pacific Ocean, despite a general increase of H_s (from 10-40 cm in 1958-2002). They linked swell trends to the number of cyclones and storm rate of change (with the North Atlantic apparently more affected by such variations), while sea trends were more associated with changes in local wind speed. In the Southern Hemisphere (0° - 60° S), Hemer et al. [10] showed clockwise shifts for decreasing H_s in the Western Pacific and counterclockwise trends for increasing H_s in high latitudes (1985-2009), shaped by Southern Oscillation Index (SOI, with clockwise shifts during SOI-) and Southern Annular Mode (SAM). Notably, Tasman Sea and Western Pacific patterns correlate strongly with SOI, while the Indian sector of the Southern Ocean (and the Eastern Pacific Ocean) better aligns with SAM (increased H_s and counterclockwise switch in SAM+, especially in austral autumn-winter). The rest of the Southern Ocean undergoes little variation due to South America's shadow.

Gulev and Grigorieva [8] employed wind roses that are histograms of the wind speed and direction time series. Similarly, Figure 11Figure 1.d,f,i show the wind roses for the considered three points. Based on the fact that waves are generated by winds, Figure 11Figure 1.a-c show the histograms of Points 1-3 30-year hourly occurrence, here referred to as Wave Distribution Matrices (WDMs), of ∂p , T_p , and H_s . Figure 11Figure 1.d,f,i and Figure 11Figure 1.a-c show that wind and wave patterns often diverge. For example, Point 2 has the narrowest unidirectional wind pattern, but displays the least unidirectional wave spread, with the bulk of the waves diverging by 75° from the main wind direction. Similarly, despite contrasting wind profiles, Points 1 and 2 yield similar two-peak T_p distributions. Conversely, Point 3, with a broad wind spread, produces a more unidirectional wave pattern. These inconsistencies confirm the well-known fact that, locally, wind patterns may not fully explain wave patterns, emphasising the need to analyse wave patterns directly and in combination.

While 1D WDM of H_s is often assessed individually, such as for extreme event analyses [11], T_p is often merged with H_s into 2-Dimensional (2D) WDMs instead of being assessed independently. Figure 11Figure 1.d,g,j illustrate 2D WDM examples for Points 1-3. Using such 2D WDMs for 2000-2014, Rusu and Ocean [12] highlighted the broader ranges of 2D WDM values in the mid-high latitudes. H_s is higher in the south, joined by long wave periods south of South America and Africa, and near Portugal. In low latitudes, 2D WDMs' ranges are narrow with high occurrences for various patterns of wave periods. Further, Rusu and Ocean [12] showed a general increase in the parameters in their hemisphere's respective winter, except for a few points located north of Papua New Guinea, south of India, and west of Mexico. Worldwide, similar locations' seasonal-pattern divergences from the rest of their respective hemisphere and basin have been found in various mid-to-low latitude regions (e.g., east and west of Australia and often around islands) [13]. Rusu and Ocean [12] highlighted a broad spread of wave direction in high latitudes, especially in the north, while the south possesses more waves from the west. Near

the Brazilian coasts and south of Madagascar, waves mainly spread in the south-east quadrant. In the mean latitudes, four points west of the north and south American coasts, the African coastline, and Western Australia show a dominance of waves from $\sim 200^\circ$, east of the Philippines from $30-85^\circ$, and west of Portugal from around 315° . 2D WDMs have been widely used in the wave energy domain to estimate the wave energy production of WECs, similar to Aristodemo and Algieri Ferraro [1], for whom 2D WDMs unlocked the possibility to numerically test 13 machines along the Italian shoreline. As Aristodemo and Algieri Ferraro [1] used a higher spatial-resolution wave model than Rusu and Ocean [12], the 2D WDMs display multi-peak patterns, similar to those of Points 1 and 2 in Figure 11. Figure 1.d-g. Where H_s generally provides 1-peak 1D WDM, these 2D WDMs demonstrate different hearts of high occurrence not only in T_p but also in H_s . In terms of device design, this highlights the need for optimisations for two or more wave systems. Indeed, if only 1D WDMs were considered for these places, perhaps the devices would be tuned for the two dominant T_p peaks, but H_s would remain the same.

The wave direction is generally assessed independently of H_s - T_p 2D WDMs (e.g., [12]), whereas annual energy production may deviate by 5-10% when direction is excluded [14], and studies have provided methods to account for the wave direction dependency of such devices [3]. Figure 11. Figure 1.e,h,k provide the 3-Dimensional (3D) WDMs (combining ϑ_p , T_p , and H_s) of Points 1-3, showing hidden multivariate regimes. For instance, some wave systems exhibit distinct directional spreads or multiple peaks tied to periods (e.g., Point 2) or broad directional spreads of the highest-occurrence regions (e.g., Point 3). This demonstrates the limitations of 1-2D WDMs in fully capturing sea-state complexity. Such complex conditions could force non-weather-vane devices to be installed in specific directions (at the optimal trade-off for energy production) or ships to adjust their routes to use the locations where the bulk of the waves advance in their direction and avoid those where waves could slow down their routes or endanger their ships.

Wave climate studies often rely on Principal Component Analysis (PCA) and k-means clustering. PCA extracts dominant spatial-temporal patterns [15], while k-means groups data with respect to similar characteristics [16]. The aforementioned distinction between seas and swells is often assessed in the wave spectrum, a more detailed level of wave representation. It provides H_s (as the zero moment of the wave spectrum), T_p (the wave period associated with the main peak), and ϑ_p (the wave direction associated with T_p). Using PCA, Echevarria et al. [17] determined for each spatial point of a global coarse 10° spatial-resolution grid, five representative wave-generated systems: two in the low latitudes, northerly- and southerly-generated easterlies; the north and south swell generation areas; and monsoon-generated waves isolating the Arabian Sea and the Bay of Bengal. Using an 8° grid, Lobeto et al. [18] extended this to sixteen clusters, mostly zonal except for the one corresponding to the swell pool in the Pacific Ocean [9]; only the Southern-Ocean climate spans multiple oceans in both studies. While spectral representations offer a detailed picture of wave climates, they often prove too complex for broad engineering applications, such as wave energy assessments similar to Satymov et al. [19]. Spectral outputs are not expressed in units of time or frequency, making it difficult to integrate them directly into engineering metrics such as annual energy production [20]. On the one hand, a distinction between the east and the west coasts is added to the zonal patterns by the coastal study of Fairley et al. [21], considering wave parameters meant to

benefit the wave-energy sector, which biases fundamental wave climate analyses. On the other hand, time series of power and direction have been clustered using k-means and crossed between wind and wave results [22], linking atmospheric systems (e.g., southerlies, easterlies, and westerlies) to wave regimes (e.g., El Niño, SAM, and the Pacific Decadal Oscillation). Finally, the zonal global climates are meridionally split by handmade clustering studies such as Young [23] based on statistics of similar parameters, Alves [24] for its wave spread analysis, and Campana [25] for military purposes. Despite their contribution to wave climate knowledge, these studies do not offer an unbiased foundation for estimating WEC energy production. Notably, a) they lack the 30-year hourly investigation of H_s , T_p , ϑp , 2D, and 3D WDMs; b) the studies clustering the waves have a resolution that is too coarse to capture more wave spatial variations; or c) some studies force the classification to suit the devices' needs rather than reflecting "raw" wave conditions, limiting their applicability to other sectors such as navigation and coastal planning. A WDM-based framework, as proposed here, complements spectral analyses, benefiting both the physical oceanographic and engineering fields. This would also clarify the relevance of basin divisions and reveal similarities between distant wave regimes.

1.3 RESEARCH GOAL AND STRUCTURE

The objective is to provide a comprehensive understanding of the global wave distributions of wave height, period, and directional occurrences independently and joined distributions integrated over 30 years in order to determine their main regional particularities that eventually define the global wave climates. Unlike most previous global wave-climate assessments, which relied on univariate statistics or coarse-resolution summaries of H_s , or even spectral analyses expressed in average spectral energy that is less suitable for various engineering purposes than hourly occurrences, this study applies a fully multivariate classification based on the 3D WDMs encompassing H_s , T_p , and ϑp occurrences. This approach depicts the evolution of each independent parameter, then the widely used 2D WDM of the merged H_s - T_p , and finally the complete 3D WDM, enabling a physically consistent identification of wave regimes at a finer spatial resolution than in earlier climatologies. As such, it bridges statistical pattern recognition with the physical processes governing global wave generation and propagation. Section 2 describes the data used in this research. Section 3 describes the different steps of the combined PCA and k-means method used to consistently extract the most representative modes and obtain a classification for 1D, 2D, and 3D WDMs worldwide, along with the methods to compare the different approaches' classifications and assess their accuracy. Section 4 presents and assesses the results that lead to the final classification, followed by a discussion of the findings. Section 5 provides the conclusions of this research.

2 GLOBAL MODEL HINDCAST AND STUDY DATA

The Centre for Australian Weather and Climate Research (CAWCR) dataset of the Commonwealth Scientific and Industrial Research Organisation provides the wave parameters [5]. The CAWCR wave hindcast is based on the widely used and verified National Oceanic and Atmospheric Administration's WAVEWATCH III [26] forced using surface winds (at 10 m height, for the 0.3° and hourly spatial and temporal resolutions, respectively) from the Climate Forecast System Reanalysis (CFSR, [27]). CFSR's ice concentration was considered following Tolman [28]

with a six-hour temporal resolution. Details about the model are provided in Durrant et al. [5], and Section 4.1 provides validations of the model.

The CAWCR model provides aggregated and spectral wave parameters over a global grid with a spatial resolution of 0.4° (~30-50 km), which falls within the feasibility ranges provided by the International Electrotechnical Commission Technical Specification (IEC TS) for Wave Energy Resource Assessment [29], particularly near the “Design-Type” category. This dataset has been used in previous spectrum-based classifications such as Echevarria et al. [17], although they focused on the temporal evolution of the wave spectrum; they reduced the spatial coverage of the grid to a handful of points spread worldwide at a 10° spatial resolution, whereas the present study aims to map bulk parameter clusters globally. The chosen resolution strikes a necessary balance between spatial detail and computational feasibility, especially considering the processing of high-dimensional histograms across all sea points and time steps on a shared hyper-performance computing infrastructure. Similar studies, such as Colosi et al. [13] using the dataset validated by Rascle and Ardhuin [30], used a slightly coarser resolution (0.5°) and were still able to identify fine-scale features linked to topography, currents, and swell propagation, which are the types of spatial complexity targeted in this research.

CAWCR’s spectral wave model uses a discretisation of 29 frequencies increasing exponentially from 0.038-0.5 Hz. The directional discretisation adopts 24 bins at 15° intervals, which is a common discretisation used in wave energy field-related studies, such as Venugopal et al. [31] for wave farm analysis, as well as in more general studies about the wave resource [32], and global directional spectral classifications [18]. Furthermore, this directional resolution seems to comply with the requirements established for feasibility studies of WECs and other marine renewables [29].

This study uses time series of bulk wave parameters: H_s , T_p , and ϑ_p . It is worth noting that the IEC TS [29] recommended the mean wave period and direction for wave resource assessment. However, these aggregated wave parameters have been selected because WECs’ power is often described using T_p [1] and have been classified based on T_p as well (better helping future mapping of the WEC classes globally) [33], even for the global operation and maintenance feasibility of WECs [34], and because wave-direction-dependent devices’ energy production is characterised more effectively using the prominent peak of the wave spectra, whilst mean spectral parameters may misguide any WEC installation [3]. Furthermore, the power of some devices is still expressed as a function of T_p [14]. Nonetheless, studies have provided research on the use of one wave period parameter in WEC performances using the relation $1.12T_e = 1.29T_z = T_p$, where T_e is the mean wave period, and T_z is the zero-crossing wave period. This relation has been obtained assuming a Joint North Sea Wave Project (JONSWAP) spectrum with a peak enhancement factor of 3.3 [35] (another example of such a relation for another wave spectra standard shape being given in the IEC TS [29]); and further, research has been provided to investigate the effect of the wave spectrum shapes on the WEC performances for a point-absorber [36] and more complex WEC shapes [14] estimating errors from 16.5% to 5-8% respectively. These H_s , T_p , and ϑ_p parameters are obtained for entire oceans and large seas. To consider a representative enough time period for wave climate analyses [37], the recommended 30-year period [29] was defined from 1981-2010. Grid points located south of 65°S have been removed. Ice concentrations used in the model also helped the selection of grid-points, following

the recommendations of Durrant et al. [5]. Grid points located in ice-covered regions were removed from the analysis. The ice mask is based on that ice database and uses an ice concentration exclusion threshold of 0.25. Moreover, a five-round iterative procedure was implemented to discard any isolated active ocean points surrounded by more than five inactive points.

3 WAVE DISTRIBUTION MATRIX CLASSIFICATION METHOD

This research aims to determine, for the five approaches described in Section 3.1, the worldwide regions of similar wave climates described in this study by means of single-to-multivariate histograms that are the WDMs described in Section 3.2. First, the large number of points worldwide is reduced to the main modes that are obtained using PCA, as PCA estimates the Empirical Orthogonal Function (EOF) decomposition of the global WDMs. PCA is described in Section 3.3. The classes are obtained using these modes and the k-means algorithm described in Section 3.4. Once the classes are obtained, they are each representative of one or more zones across the world and therefore a method to assess the accuracy and representativeness of the clusters and their zones is needed, and is thereby described in Section 3.5.

3.1 OVERVIEW OF THE FIVE APPROACHES

This study applies a stepwise classification of the global wave climate based on the three key bulk parameters: significant wave height (H_s), peak period (T_p), and peak direction (ϑ_p). To progressively explore the complexity of wave regimes, three levels of classification are performed: individually (1D), jointly as H_s – T_p (2D), and fully combined as H_s – T_p – ϑ_p (3D). Each step adds a new dimension to the sea-state characterisation.

- 1D classification: Three separate classifications are performed, one for each parameter (H_s , T_p , ϑ_p). These highlight regions where a single parameter dominates, without accounting for interactions with the others. Though simple, 1D classifications help isolate the independent behaviour of each parameter and reveal how much they alone structure the global wave regimes.
- 2D classification: Based on joint histograms of H_s and T_p , this level combines wave energy content and maturity. It reveals regions with similar energy levels but different sea-state types, such as swells versus wind seas. This approach already improves physical interpretability over 1D and is commonly used in wave energy assessments.
- 3D classification: Using joint histograms of H_s – T_p – ϑ_p , this classification is the most complete approach. It captures both the energetic and directional properties of the waves, which are essential for identifying complex patterns such as opposing wave trains or wide directional spreads. Though more demanding, the 3D method best reflects real sea-state variability and applications such as routing or device orientation.

The 1D–2D–3D framework enables a layered understanding of global wave patterns. It also allows evaluating how much information is gained by including more parameters and which combinations better capture the dominant physical drivers behind the observed sea states.

3.2 INITIALISATION OF THE PRINCIPAL COMPONENT ANALYSIS

Each classification approach (H_s , T_p , ϑp , 2D, and 3D) is based on the histograms of the hourly occurrence of the time series from 1981-2010. Following Choupin et al. [14], the histogram discretisation is 0.2 m H_s (from 0.1 m), 1 s for T_p (from 0.5 s), and 15° ϑp (from 0°), forming either univariate (H_s , T_p , or ϑp), bivariate (2D: H_s - T_p) or trivariate (3D: H_s - T_p - ϑp) WDMs.

To reduce the dataset, all the sea-state bins that were never reached in any location are excluded. Hence, the final active bin ranges span 0.1-19.9 m (H_s), 2.5-34.5 s (T_p), and 0-345° (ϑp). The total number of bins for a full WDM is S : H for H_s , T for T_p , Θ for ϑp , HT for 2D WDM, and $HT\Theta$ for 3D WDM. However, as unpopulated bins have been removed from the WDMs, the number of considered bins is set to be s with $1 \leq s \leq S$. Eventually, M being the number of grid points considered, for each WDM's j^{th} grid-point, $1 \leq j \leq M$, and i^{th} binned sea-state, $1 \leq i \leq s$, the number of hours, $O_{i,j}$, of sea-states occurring during 1981-2010 is computed from the H_s , T_p , and ϑp 30-year time series. These reduced histograms are the WDMs depicted in Section 1.2.

Each WDM is reshaped into a column vector and is stacked across M grid points to form an Input Matrix (IM), where each row represents one sea-state bin and each column one location. This process is repeated for each of the five approaches, and Figure 22 provides an illustration of how IM is obtained for the full 3D (H_s , T_p , and ϑp sea-state description) distribution for an example of a model with few grid points along a coast. Each IM is then decomposed into the EOF as per Section 3.3.

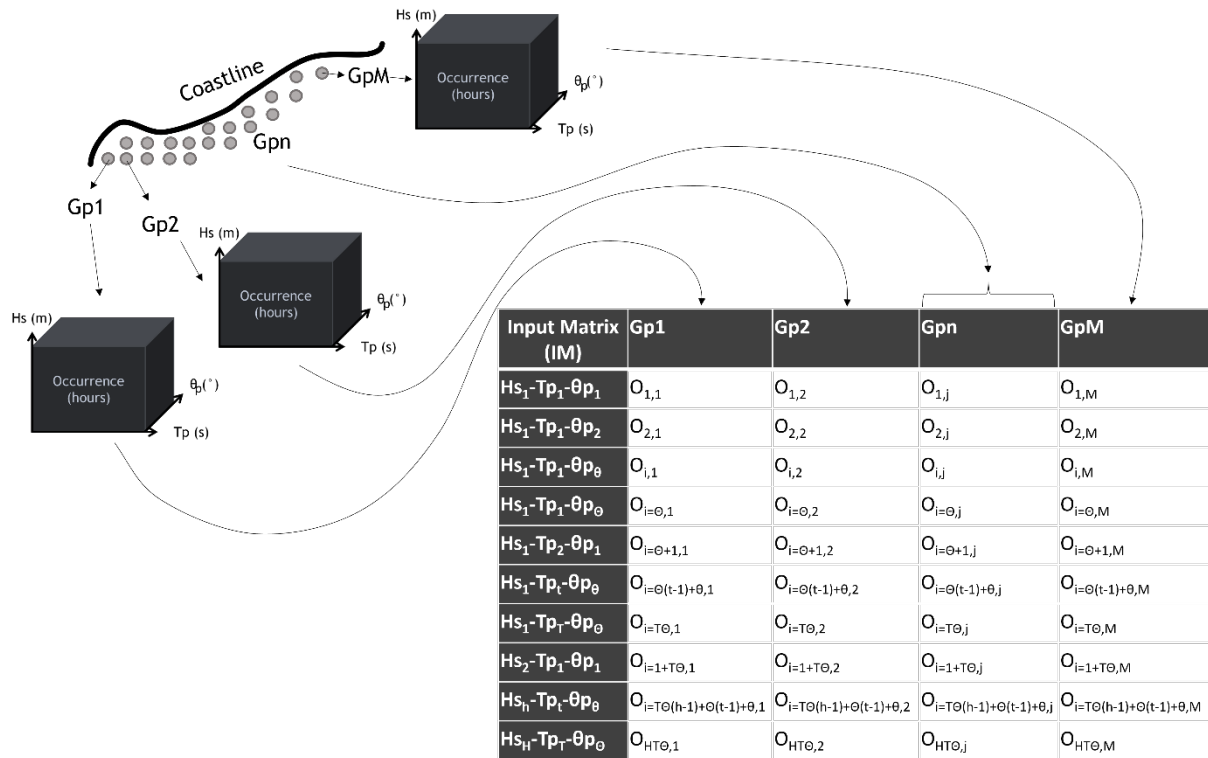


Figure 22. Illustration of the input matrix for a set of Grid-Points (GP) along a coastline. H_s is the significant wave height, T_p the wave peak period, and ϑp the wave peak direction. The occurrence O has been accumulated over 30 years for each bin to obtain the different statistics. H is the number of bins over H_s , T for T_p , Θ over ϑp , and M is the number of grid points considered in the analysis. In this example, IM is considered to be composed of all $S=HT\Theta$ sea-state possibilities.

3.3 PRINCIPAL COMPONENT ANALYSIS: AN EMPIRICAL ORTHOGONAL FUNCTION DECOMPOSITION

The first step of PCA is to centralise $IM(s,M)$ by subtracting each IM's row mean. That column vector, consisting of IM's row means, is the Space-Mass vector $SM(1,s)$. This subtraction yields the Adjusted Matrix (AM, the same size as IM). The Covariance matrix, $C(s,s)$, is computed as $C=AM \times AM^T$. PCA mainly consists of solving the eigenvalue problem: $CX=\lambda X$, where $\lambda(s,s)$ is a diagonal matrix containing the eigenvalues λ_i of C , and each i^{th} row vector of $X(s,s)$ is the eigenvector associated with the eigenvalue λ_i . Each eigenvalue λ_i gives a measure of the fraction of the total variance of C explained by the mode i . Each eigenvector X_i is the Empirical Orthogonal Function (EOF) of the principal pattern of variability of the i^{th} WDM, and the eigenvalues are a measure of the relative importance of those patterns. The EOF patterns obtained represent standing oscillations of the single or multivariate wave distributions, and their spatial variation is depicted by the Principal Component (PC) maps, given for a grid point j by Eq. (1):

$$PC_{i,j} = \sum_{n=1}^s X_{i,n} AM_{n,j} \quad (1)$$

Each mode of variability of the wave distribution is characterised by the pair of EOF and PC of the mode. Here, each EOF describes the mode's representative WDM, and the associated PC provides the spatial distribution of coefficients describing how that mode best fits the local distribution. Eventually, for each j^{th} grid point, its distribution can be entirely obtained by adding to the local SM the sum of all modes' product of EOF and local PC, such that for that j^{th} point, IM's i^{th} bin is obtained via Eq. (2):

$$IM_{i,j} = SM_i + \sum_{n=1}^s X_{i,n} PC_{n,j} \quad (2)$$

When PCA is most accurate, the sum of the local SM and the product of the first mode's (i.e. those with the highest λ values) EOF and local PC should provide a result similar to the local WDM. To better understand the percentage contribution to the modes' representativeness, the eigenvalues have been transformed into percentage values, $\lambda_{\%}$ (%), using Eq. (3):

$$\lambda_{\%i} = \frac{100 \lambda_i}{\sum_{n=1}^s \lambda_n} \quad (3)$$

3.4 CONSISTENT K-MEANS CLUSTERING USING THE PRINCIPAL COMPONENTS

The clustering consists of grouping the projected points (with coordinates provided by the PCs, sorted so that the first mode has the highest λ and so on) based on their distance from one another in the PC-projection space. Each group leads to a cluster characterising a representative class. In this study, computing the Euclidean distance between every pair of points, as in Choupin et al. [33] (weighing M^2 number of calculations, regardless of the number of calculations, to determine the clusters by comparing all distances), is not feasible within a reasonable amount of time due to the very large number of individuals considered. Therefore, the k-means algorithm, used for instance by [38] to classify sea-states of WDMs, must be used instead. Indeed, the k-means algorithm only computes the distance between the k cluster centroids and each individual point, leading to a much smaller number of calculations of about $p k M \ll M^2$ since $k \ll M$, and it takes about $p=10$ rounds to reach convergence,

i.e. when the overall distance between each cluster and surrounding points is minimised. A two-run k-means algorithm is therefore used, following Azzellino et al. [39], such that the centroids of the first round are used as the starting point of the second, delivering, as such, the final clusters. Defining the starting point of the k-means clustering is the first constraint of this method. To ensure that the results are the same for each run, the starting point method of Mayo [40] has been implemented here. This study's final clustering sets the number of clusters to 15 and the number of PCs to 10 (discussed in Appendix B). However, the results also consider an additionally lower number of clusters to assess the main trends in the wave distribution despite being too low in clustering potential.

For any k-means-based classification, there is no reason for each cluster to be indexed in the same way, e.g., a cluster generally located in the Northern Hemisphere would be indexed as cluster 1, and one in the Southern Hemisphere would be cluster 2, while for another approach, it would be the opposite. Therefore, a period-crossing algorithm has been developed to ensure consistent indexing between the different approaches' clusterings (as a helping feature of this research and not a core component of the methodology, this algorithm is presented in Appendix C).

3.5 CLUSTER RECONSTRUCTION: WAVE DISTRIBUTION MATRIX, SPATIAL MAPS, AND REPRESENTATIVENESS ASSESSMENT

Tracing back the clustered grid points to the map reveals the cluster's global coverage. Points near one another may gather into a zone. Each cluster (or sub-cluster) may contain more than one zone, similar to the Köppen-Geiger climate classification maps [41]. Nonetheless, each cluster's representative WDM is obtained by averaging the WDMs of all the grid-points (regardless of the distance between their zones) contained in that cluster, since they belong to the same cluster. Each pair of cluster's WDM and associated zone map is a wave climate.

Finally, correlation, Eq. (4), and the Relative Mean Absolute Difference (RMAD), Eq. (5), based on RMAE (Relative Mean Absolute Error) from Walstra et al. [42] are computed between each cluster's grid point's WDM (no matter its zone) and the cluster's WDM. This provides an estimate of the wave climate's representativeness globally and for each zone. Also, this provides information on which zone is the most representative of its climate and how the transition is structured from one climate zone to another. Considering N , the number of points in one cluster k , for a point i in cluster k , the correlation and RMAE are calculated using Eq. (4) and Eq. (5), respectively:

$$correlation(i|_k) = \frac{\sum_{n=1}^N (o_{i|k_n} - \bar{o}_{i|k_n})(o_{k_n} - \bar{o}_{k_n})}{\sqrt{\frac{\sum_{n=1}^N (o_{i|k_n} - \bar{o}_{i|k_n})^2}{N}} \sqrt{\frac{\sum_{n=1}^N (o_{k_n} - \bar{o}_{k_n})^2}{N}}} \quad (4)$$

$$RMAE(i|_k) = \frac{\sum_{n=1}^N |o_{i|k_n} - o_{k_n}|}{\sum_{n=1}^N |o_{i|k_n}|} \quad (5)$$

4 RESULTS AND DISCUSSIONS

The results section is divided into three parts: Section 4.2 provides an overview of the general statistics of the global wave patterns depicted with H_s , T_p and ϑp WDMs; Section 4.3 applies the PCA to the WDMs and describes their first modes of variability; and Section 4.4 presents and discusses the classification of the wave climate. For precision purposes, most results are provided in terms of the number of hours over the entire 30-year period.

4.1 MODEL VALIDATION

Durrant et al. [43] validated the model both globally, using satellite altimetry, and regionally, using buoy observations around the United States. The global validation for 1982-2010 compared H_s with altimeter data (using GEOSAT, ERS1, TOPEX, ERS2, GFO, JASON1, ENVISAT, and JASON2, approximately in that chronological order, with some overlap between the missions). Durrant et al. [43] examined 2005 in further detail. Results indicated the largest errors in the Antarctic regions, exceeding 1 m, associated with positive biases (discussed in earlier studies [44]). Smaller errors are found in nearshore regions, rarely exceeding 1 m and generally remaining below 0.5 m, with no consistent bias sign (see discussion of coastal modeling and fetch-source term potential issues [45]). The model performs well in open oceans despite a slight negative bias in the western tropical Pacific that is suspected to be linked to a missing shoreline-reflection effect, possibly due to the abundance of islands (discussed in earlier studies [46]). They also offered a speculative remark on the potential influence of ocean currents on the results. Temporal validations highlighted consistency between the model and the altimetric data, with interannual variability associated with phenomena such as El Niño, including a reduction in values around 1993-1994. In addition to the study by Durrant et al. [19], multiple studies have validated the CAWCR model across diverse oceanic regions, using both satellite and buoy observations (Table 11Table 1).

Table 11. Non-exhaustive list of research validating the Centre for Australian Weather and Climate Research (CAWCR). All the symbols and abbreviations are provided in the glossary.

Reference	Location (overall period – some devices may have various periods within the given range)	Type of data	Data validated	Skill metric used*
[17]	Global (1982-2010) United States, including Hawaii regions and the Gulf of Mexico (2004-2014) South Pacific: Fiji, Vanuatu, Western Samoa, Cook Islands, and Rarotonga, Tuvalu-Funafuti, Tonga-Eua, and Tonga-Tongatapu (1987-1994)	Satellites Buoys	H_s H_s , $U10$, T_p	Bias, RMSE (n)Bias, RMSE, SI, R
[47]	United States, including Hawaii regions and the Gulf of Mexico (2004-2014)	Buoys	Directional wave spectra (monthly), H_s (monthly)	R , (n)RMSE
[48].	Southeast Brazil: Itajaí in Santa Catarina state and Rio Grande in Rio Grande do Sul state (2009-2015)	Buoys	Directional wave spectra, ϑm , H_s	MAE, Corr, RMSE, SI
[49]	South Africa (1979-2020)	Buoys	E , $U10$	R , nRMSE
[50]	Australia (2015-2020)	Satellites	Directional wave spectra, H_s , Te , ϑm ,	(n and mean)Bias, (n)RMSE, SI, R , slope
[51]	Australia (1985-2012) Australia (1974-2013)	Satellite (same as [17]) Buoys	H_s , E H_s	Bias, RMSE, SI, R Bias, RMSE, SI, R , slope

*It is worth noting that most of them also provided qualitative validations using maps, time series and percentile comparisons, QQ-plots, probability plots, and scatter density plots, along with comparisons between means and standard deviations. Specifically, Oliver et al. [49] used Taylor diagrams as well.

Due to CAWCR's 0.4° spatial grid resolution, model grid points may not precisely match the water depths at the measurement locations, and multiple buoys may fall within the same CAWCR cell. As previously mentioned, the wave model performance degrades nearshore owing to the limited representation of bathymetric effects, including submerged obstructions (e.g., reefs). The model's weaknesses in high latitudes have also been reported in Australian assessments, with an overestimation of E 's upper percentiles, whereas in lower latitudes these percentiles are underestimated [51]. T_e generally exhibits correlations of approximately 0.1, slightly lower than the 0.9 typically obtained for H_s across all locations (details on the CAWCR skill over H_s from the studies in Table 11Table 1 are summarised in Table 22Table 2). However, although T_e has positive biases below 2 s around Australia, which is similar to the values obtained for T_p in the United States regions, nRMSE and nBias generally indicate better performance for T_e than for H_s . It is worth remembering that T_e and T_p describe different portions of the spectrum and cannot be directly compared. Durrant et al. [43] noted that T_p can exhibit abrupt increases (about +10 s) when incoming swell systems rapidly dominate local sea states. Directional accuracy is more limited, with lower correlations and nRMSE values ranging globally from ~15°, especially in the North Indian Ocean, to over 50° in the Western Pacific and Antarctic regions [50]. In general, CAWCR provides robust H_s characterisation and captures seasonal directional variability with reasonable skill, although the spectral spreads tend to narrow in broader directionality cases. Furthermore, RMSE values tend to be higher in boreal winter and lower in boreal summer, a pattern suspected to come from similar errors of the CFSR winds [17]. Extreme cyclonic events seem generally underestimated by CAWCR, with an underestimation of the sea wave systems, and relatively better performance for the swell wave systems (based on the studies considering the directional wave spectrum in Table 11Table 1). Finally, E shows biases of about a +0.17 kW/m and 0.85 correlation (hence lower than H_s) around Australia [51]. This region's correlations reach up to ~0.85, and are characterised by low RMSD (~0.6–0.8). The standard deviations obtained in this region also closely match the observations along the South African coast, with a few stations exhibiting moderate overestimations of variance. Overall, the performance around Australia is consistent with expectations for wave hindcasts in coastal settings [49].

Table 22. Significant wave height skill metric of CAWCR per location. All the symbols and abbreviations are provided in the glossary.

Location (note on the Bias if available)	RMSE (m)	nRMSE (%) – MAE (m) in the Brazilian case	Correlation (–)	SI (–)
United States	0.35-0.55	<15 ~25 in the Gulf of Mexico during the Atlantic hurricane season	>0.8 >0.9 in some cases	<0.2
Brazil	0.33-0.43	<0.3	0.9	<0.25
Australia (from -0.03 to +0.04 m for [51] and mean biases around ±0.1-0.2 m, lower equatorwards for [50])	0.44 (0.07 higher than the global average [17])	~0.20 equatorwards ~0.30 in the Western Pacific ~0.40 near the Antarctic	~0.9	0.18-0.3

Kumar et al. [52] pointed out that bulk parameters such as H_s , T_p , and ϑp can misrepresent mixed sea states where short-period local seas overlap with long-period swells, particularly affecting Stokes drift, HF radar, and

classification interpretations. This reinforces the importance of carefully validating both sea and swell signals in the bulk outputs when using them for classification or energy assessments. Such concerns have guided the design of the clustering scheme and the interpretation of mixed conditions in regions where both systems coexist. Overall, CAWCR demonstrates high skill across multiple ocean basins and scales, with typical H_s RMSEs between 0.3 and 0.6 m, correlation coefficients above 0.85–0.90, and angular errors typically below 15–20°. These results are consistent across both open-ocean and moderately sheltered locations. The model is especially proficient at reproducing swell-dominated regimes, while also providing sufficient directional and spectral accuracy to support classification tasks using bulk parameters. These validations confirm CAWCR’s suitability for global wave occurrence analysis and clustering, as developed in this study.

4.2 GLOBAL ASSESSMENT OF WAVE PARAMETER DISTRIBUTIONS

Figure 33 provides the averages and standard deviations for all 1D and 2D sea state distributions (exceptionally, H_s - θ_p and T_p - θ_p are also provided for a complete overview, and complementary global statistics on each 1D parameter are provided in Appendix A, along with the full 3D mean provided in Appendix B). It must be noted that the mean matrices are the SM removed from the IM of each corresponding 1-3D WDM provided in Section 4.3. Eventually, the minima are sparsely provided since they are mainly zeros.

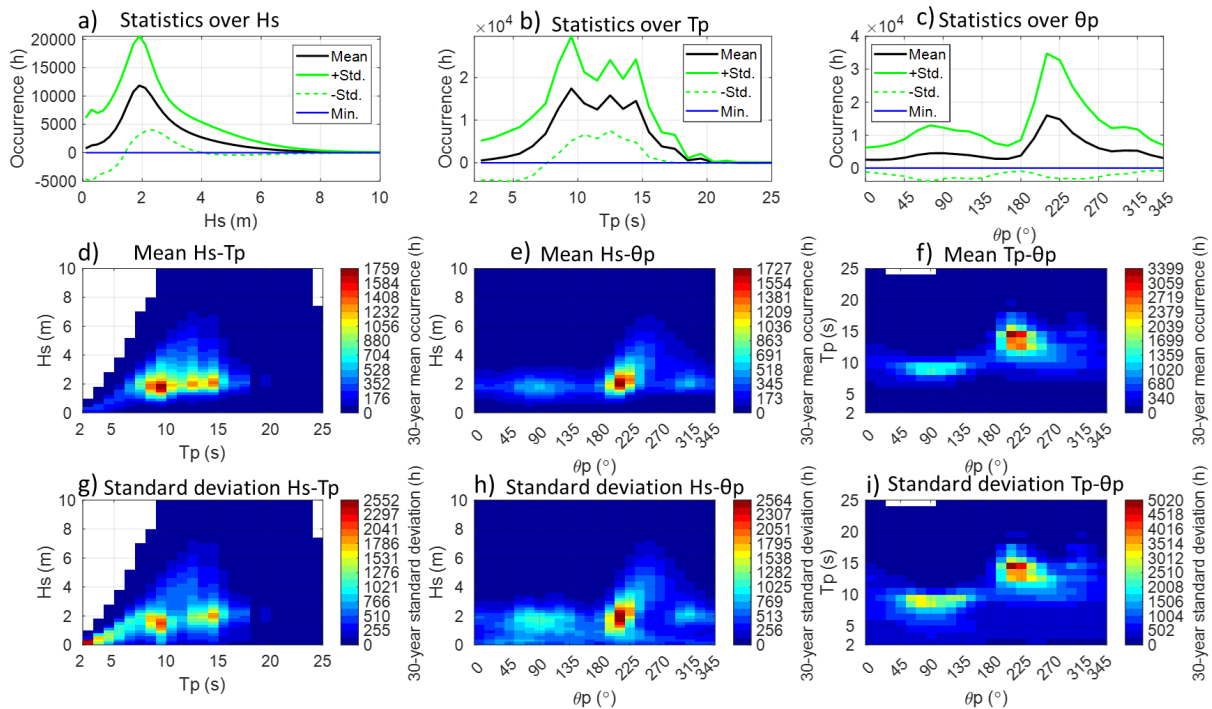


Figure 33. Sea-state parameters 30-year global distribution statistics. H_s is the significant wave height, T_p , the wave peak period, and θ_p , the wave peak direction.

The 1D histograms (top panels of Figure 33) highlight a single H_s peak (around 1.9 m), three main peaks of the average T_p (highest peak at 9.5 s, followed by 12.5 s and 14.5 s, with a slight spike around 19.5 s), and three peaks for θ_p (highest peak around 210°, and two lower peaks around 90°, and 315°). It is worth noting that a 10-s wave period is a good rule of thumb threshold to separate the seas (wind waves, with smaller periods) from the swells (with higher periods) [53]. However, this distinction remains simplistic. Other diagnostics, such as the wave age parameter (e.g., $\alpha = c_p/W$, with α the wave age, c_p the phase velocity, and W the 10-m wind speed,

whose brief statistics are shown in Figure A.5), provide a more physically based criterion to distinguish wind-sea and swell regimes [54], but their computation requires detailed and co-located wind information. As this study focuses solely on wave occurrences, such analyses are beyond its current scope and are left for future developments. As such, two peaks of T_p are more likely to correspond to the swells, while one is for seas near the threshold. The latter occurs slightly more than the two former ones. T_p peaks reappear in the H_s - T_p average diagram, while the values remain around the 1.9 m peak of H_s . However, it is worth noting the high H_s - T_p standard deviation for the first and last T_p peaks. In contrast, the H_s - T_p highest standard deviation is in low H_s and low T_p ranges. The multimodal distribution of ϑp is more prominent with H_s than with T_p . The H_s - ϑp distribution spread around the 1.9 m H_s peak possesses all three ϑp peaks in the same order of intensity. The 210° peak remains the highest for ϑp , H_s - ϑp , and T_p - ϑp . However, where H_s - T_p is aligned with the peaks of 1D H_s and T_p , T_p - ϑp demonstrates two peaks (14.5s-210° and 8.5s-90°) while both 1D T_p and ϑp possess three peaks. Furthermore, for all three approaches, the maxima at 210° are associated with higher standard deviations.

The average 2D panels (Figure 33Figure 3.d-f) are similar to the full 3D accumulated 2D projections on each H_s , T_p , and ϑp axis. There is one global peak that dominates all other bins at 2.1m-14.5s-210° and provides the largest 3D occurrence patch. To better interpret these distributions, Figure 44Figure 4 provides the maps associated with these key bins for the 1D distributions. For instance, Figure 44Figure 4.a provides the accumulation of all hours over the 30-year period falling in the 1.8-2.0 m bin (centred in 1.9 m) of H_s . The maps of the statistics over the time series of H_s , T_p and ϑp are provided in Appendix A. For instance, Figure A.1.a provides the mean of the 30-year temporal evolution of H_s . Furthermore, complementary maps associated with the three key bins of the 2D distributions are provided in Figure B.11Figure B.1. Figure 55Figure 5 provides similar maps for the maximum 3D bin associated with each peak and key bin of 3D's three 2D walls.

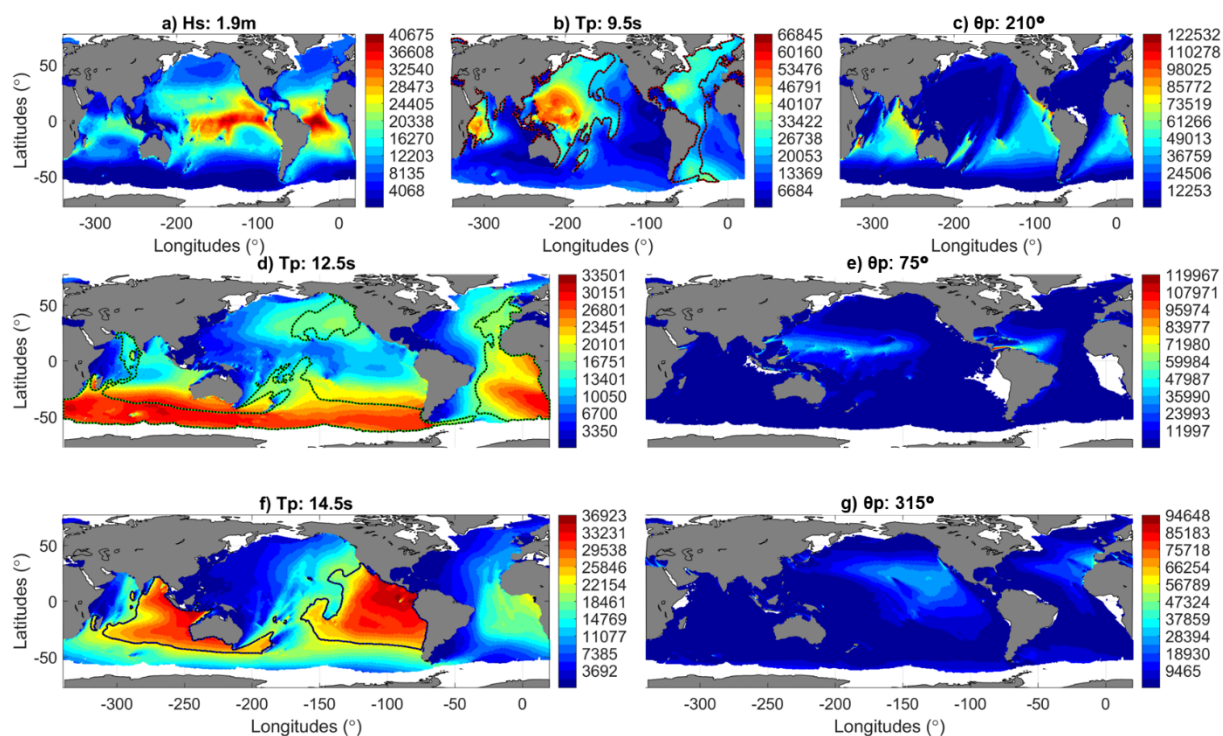


Figure 44. Mapping of the 30-year (1981-2010) occurrence (in hours) of 30-year average sea-state distributions' peaks and key occurrence bins for 1-dimensional distributions of H_s , the significant wave height, T_p , the wave peak period, and ϑp ,

the wave peak direction. The first row provides the highest peak. Dashed lines have been added to Tp 's subfigures to contour the zones of each peak obtained in Figure 1111Figure 11.n, in their respective colour and black.

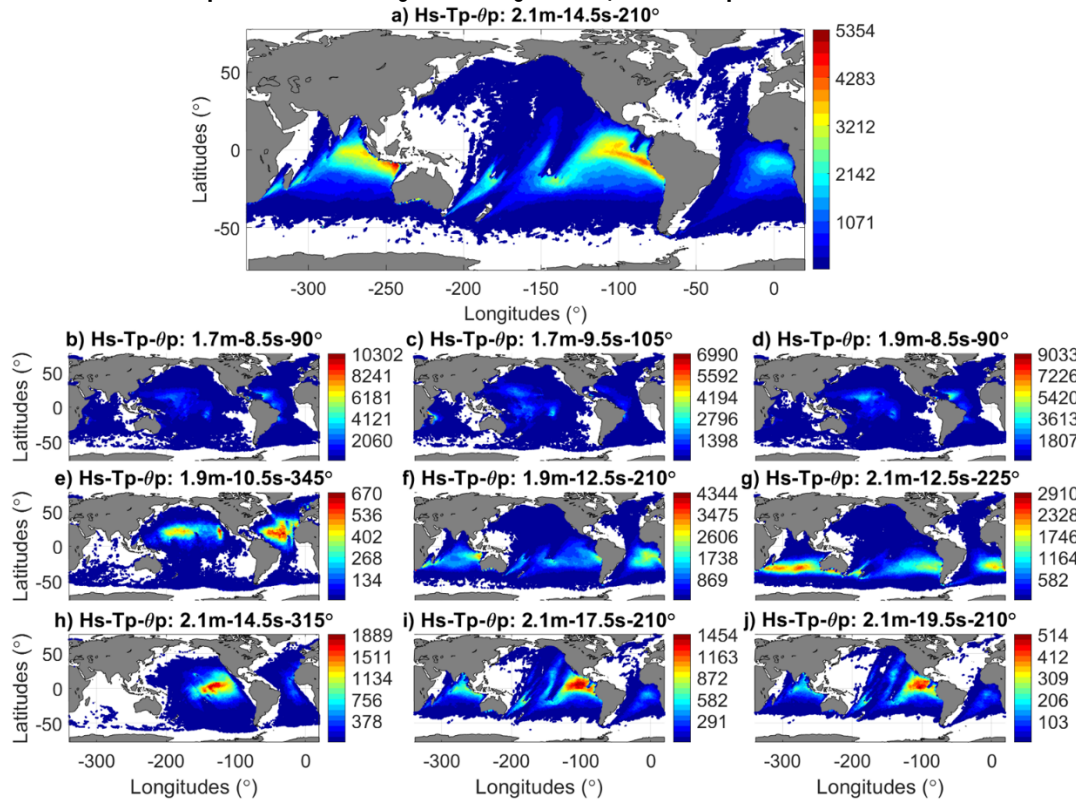


Figure 55. Mapping of the 30-year (1981-2010) occurrence (in hours) of 30-year average sea-state distributions' peaks and key occurrence bins of the 3-dimensional H_s (significant wave height), T_p (wave peak period), and θ_p (wave peak direction).

Not shown here, yet palpable in Figures 4-5 and Figure B.11Figure B.1, as H_s bin range increases, waves expand their coverage from fetch-limited domains, generally closed seas, to eastern coasts, later including northern high latitudes, the Arabian Sea, and the Bay of Bengal. Around 0.6-0.8 m, the entire North Atlantic is also included. Offshore low latitudes and the Pacific-Indian section of the Southern Ocean remain uncovered. At the bin centred on 1.1 m, the entire Atlantic Ocean is covered. Reaching 1.4-1.6 m, only rare locations in the south lack these waves. Then, from 3.4-3.6 m, regions in the Philippines and along the equator below Africa no longer have higher waves. This equatorial region extends westwards as the H_s range keeps increasing until covering all longitudes at 5.4-5.6 m, to then expand polewards. After 7.8-8 m, the fetch-limited domains start losing bins in higher waves. Beyond 11.6 m, waves are rare in mid-low latitudes. The high latitudes host most of the tallest waves, but these waves are increasingly sparse. Figure 44Figure 4.a shows the spatial distribution of occurrences of sea-state with H_s in 1.8-2.0 m. Following the increasing H_s scattering, the 1.8-2.0 m bin displays a clear meridional gradient with the most frequent occurrences within the tropical zone, peaking at ~40,000 hours in 30 years (15%), and less frequent occurrences (as low as 4,000 hours) in the high latitudes and fetch-limited domains.

As T_p range increases, the high-occurrence regions move from fetch-limited domains to offshore in a similar pattern to H_s , although after 22 s (already absent from the fetch-limited domains) the withdrawal pattern mainly expands from the east coasts, later joined by the west coastal-side to leave a northward-decreasing 20°S-60°N region around 150°W from 26 s (not shown here, yet palpable in Figures 4-5 and Figure B.11Figure B.1). In 9-10 s, Figure 33Figure 3.b of the 9.5 s peak, Figure 44Figure 4.b shows four regions of highly concentrated waves

reaching nearly 25% of the whole period, namely west tropical Pacific coasts (especially near the Philippines and Indonesia) and north of Madagascar, covering most of the Arabian Sea. Waves in 12-13 s (Figure 44Figure 4.d) moderately cover the north, i) with a reasonable zonal gradient, ii) with low occurrences in the low latitudes, and iii) higher occurrences in the south (although twice as less present than for the 9.5 s peak) at a certain 315° in the South Atlantic. Finally, 14-15 s, slightly more present than 12-13 s, highlights waves in the east of the main basins, extending slightly into the Southern Ocean, while absent in the Mediterranean Sea. On the one hand, 9-10 s regions are principally in the low occurrence regions of 14-15 s, and seem to be characterised by seas; on the other hand, the regions of 14-15 s high occurrence waves are associated with the swell pools [9]. The 14-15 s bin in the Bay of Bengal provides one side of the bimodal typical profile of the sea-state distribution in this region, as observed in Figure 11Figure 1.

Joining ϑp to this analysis, the bin centred on 210° provides regions with up to nearly 47% of the 30 years (Figure 44Figure 4.c). These regions form wedge-shaped structures extending from the Southern Ocean towards the low-latitude eastern margins of the Indian, South Pacific, and South Atlantic basins. The tip of the Atlantic wedge sits near the coast of Togo, while in the Pacific Ocean and the Bay of Bengal, the wedge stretches northeastwards, reaching up to approximately 25° latitude. These structures are consistent in the 2D distribution with stronger values along the Indian Ocean nearshore, particularly west of the Indonesian coastline. While for $Tp\text{-}\vartheta p$, reasonably high occurrences appear along the Mexican coastline, a high-occurrence trail offshore the Mexican coast, extending up to the Peruvian coastline, stands out. Just as mentioned for increasing Hs and Tp ranges, only the small and large energetic $Hs\text{-}Tp$ waves display large wave-bin-free areas (see example of Figure B.11Figure B.1.a-d). The multidirectional global wave distribution increases the presence of regions not reached by the waves (Figure B.11Figure B.1.e-i), especially in 3D (Figure 55Figure 5). There are great shadowing effects of the continents in the direction of the waves, such as for the westward-based examples (bins centred on 90° and 105°) and the great example of the waves coming from the south propagating east and northeastwards. Southern waves propagating northeastwards reaching the east shores of the basins had sufficient travel distance so that dispersion processes allowed them to increase their period greatly [55]. As such, the 14.5 s regions gather waves mainly originating from large fetches in the high latitudes south, while the bulk of the 9.5 s waves mainly come from smaller fetch areas north of the oceans. South-southeastwards waves (especially Figure 55Figure 5.e,h), much less frequent than the other peaks, are present around the 20°N, and are mostly absent from the Indian Ocean basin. These results make sense of the pools of swells and seas, the first mainly composed of waves coming from the south and the latter of those coming from the north, although waves from the north still contribute relatively to the east coasts. Finally, based on Alves [24] and Figure 11Figure 1, the southeast, westward, and small- Tp waves of the Indian Ocean basin mainly result from local conditions.

4.3 EMPIRICAL ORTHOGONAL FUNCTIONS AND PRINCIPAL COMPONENTS

Hs has been widely investigated in the literature in many specific locations and in statistical maps. Regions of high mean significant wave height are located in the north and south of the basins. Decomposing 30-year histograms of the Hs WDM using PCA provides the principal modes of Hs 's global distribution, which also highlights its local patterns (see Section 3.3 and Appendix D). Figure 66Figure 6 shows the EOFs and PCs of the

first three modes of the global 30-year *Hs* WDM. EOFs can be considered as transfer functions allowing transitions from one distribution pattern to the next. Where the values of the first PC are equal to zero, then the next mode without a zero value in the PC dominates at that location. This is further described in Appendix D. The principal component of the first mode (PC1) of *Hs* shows a clear latitudinal gradient, which can be directly related to the meridional distribution of surface winds at the global scale, with the lower *Hs* values being generated by extratropical storms in the midlatitudes (see Figure A.55Figure A.5).

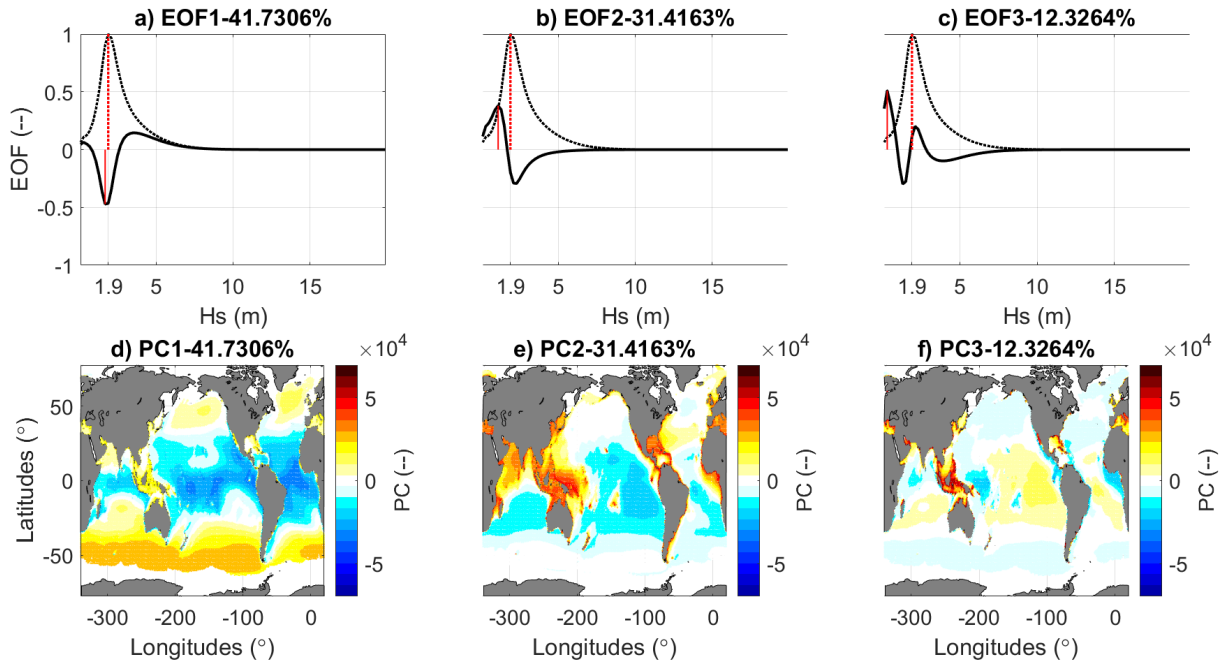


Figure 66. First three Empirical Orthogonal Functions (EOFs, top-panels solid lines) and Principal Components (PCs) maps (bottom-panels) of the significant wave height (*Hs*) for the entire integrated 30-year period. The dashed lines of the EOF figures show the normalised global mean *Hs* distribution (*Hs* normalised space mass vector, see Section 3.3). The red lines underline the maxima to help compare the EOFs, solid for the EOF and dashed for the mean. The title indicates whether the subfigure is an EOF or PC, the mode concerned, and the associated percentage of the total variance values.

The product of each PC-EOF pair contributes to the reconstruction of the original matrix, as shown in Equation (2). Consequently, the sign of individual components is arbitrary: multiplying both the PC and its corresponding EOF by -1 does not alter the physical result. Therefore, apparent sign differences across parameters (e.g., *Hs*-*Tp* versus mean direction) do not imply anti-correlation. As such, although the PC1 of *Hs* seems anti-correlated to the mean of the *Hs* time series (Figure A.11Figure A.1.a), in fact, the sign does not matter as long as the associated EOF is considered properly with its related PC. Hence, the absolute correlation between the time series of *Hs* mean and PC1 is 0.73. The first mode appears strongly correlated with the 1.9 m distribution mapping of Figure 44Figure 4, with a correlation of 0.91, higher than that with the global *Hs* average. The associated EOF1 illustrates how the peak of the *Hs* distribution shifts to the left (i.e. in lower *Hs* values) from the tropical region to the midlatitude regions. This agrees with higher wave heights in the tropical regions than in midlatitudes. The second mode differentiates the open-ocean regions from fetch-limited domains and coastal areas. Indeed, smaller waves are prominent in coastal regions, and the negative PC2 values in offshore areas provide a right shift towards a strong (31%) 2-m wave contribution. The third mode presents a multimodal EOF. Mode 3 encompasses the

contribution of locally generated waves with its first peak, accentuated in fetch-limited domains and nearshore areas, and with greater heights offshore.

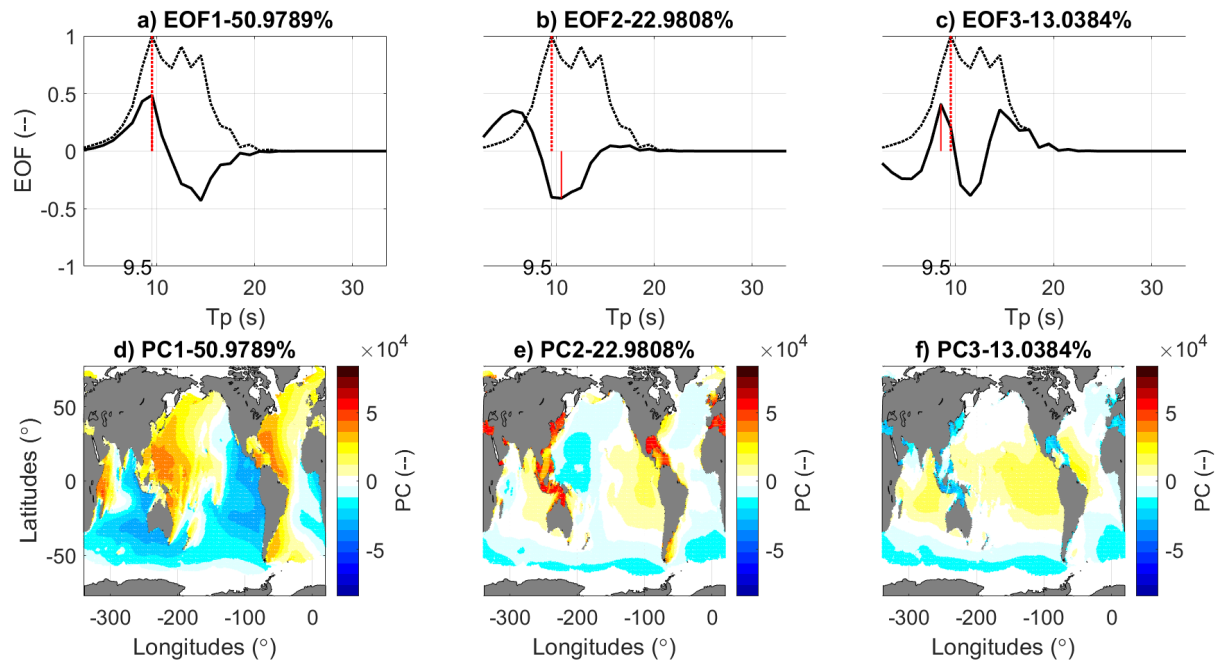


Figure 77. Same as Figure 66, but for the wave peak period (T_p).

Figure 7 provides the first three EOFs and PCs for T_p . The first mode of T_p is largely dominated by seas versus swells WDMs and pools. The continental shadow effect observed in Section 4.2 on the wave patterns is also visible in this mode and further discussed below. Similar to H_s , i) the second mode is more associated with fetch-limited domains and coasts with a weak connection to open seas, and ii) T_p 's third mode provides more multimodality.

In the EOF decomposition of the joint H_s - T_p 2D histogram (Figure 8), the first mode is largely shaped by T_p , as the distribution of PC1 values aligns with the contours of the seas (swell) pools in the northern (southern) oceans (Figure 8.d); while the two hearts of EOF1 are located around 9.5 s and 14.5 s (Figure 8.a), both characteristics of seas and swells, respectively. Mode 2 mainly reflects H_s variations, forming a clear meridional gradient (Figure 8.e). Mode 3 reintroduces the distinction between fetch-limited domains and the open ocean, resembling the separations highlighted by the second modes of H_s and T_p individually.

H_s , T_p , and their joint EOF decomposition now highlight the large coastal occurrences in these ranges (see also Figure B.1), whilst offshore areas are characterised by more energetic waves generally generated by large fetches. This explains the H_s - T_p 's highest standard deviation of Figure 3 in low H_s and low T_p ranges, while the seas/swells pools show the high H_s - T_p standard deviations for the first and last T_p peaks.

It is worth noting that the first two modes contribute about 20% each to the distribution. Despite this equal split, the third mode accounts for 13% of the variation, similar to the third modes of H_s and T_p , whilst their first two modes' $\lambda_{\%}$ accumulate about 70% versus 40% for 2D. Therefore, less than 50% is represented by the first three modes, which demonstrates a limitation in fully capturing the variability of the 2D histograms.

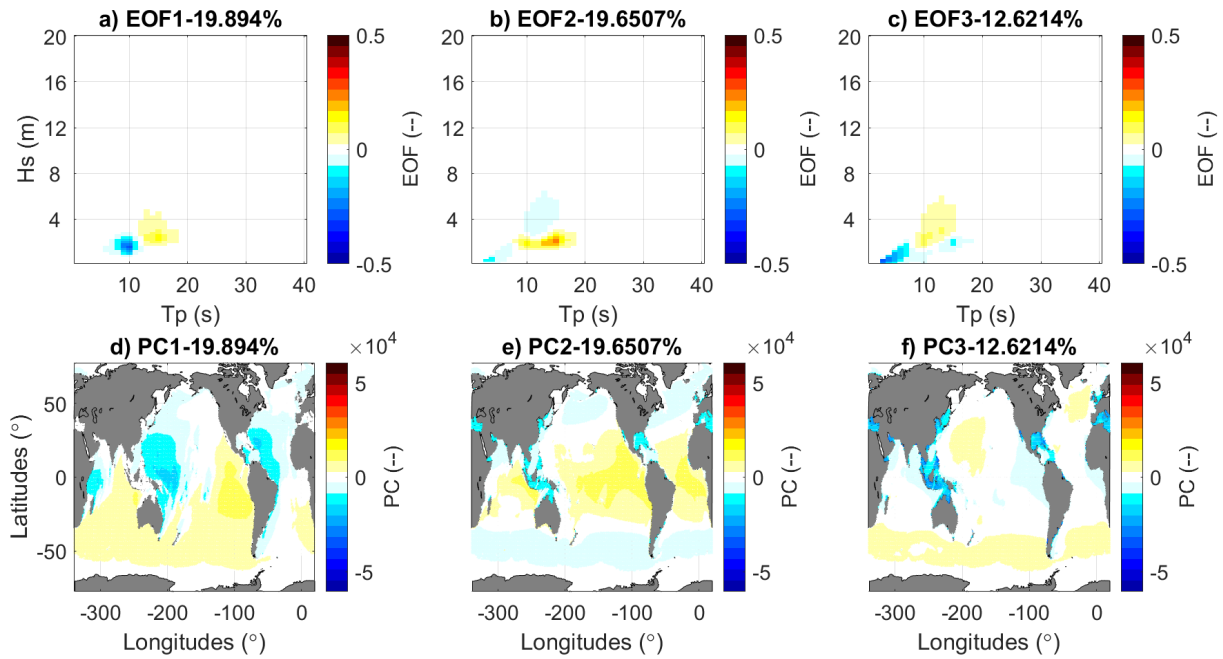


Figure 88. First three Empirical Orthogonal Functions (EOFs) and Principal Components (PCs) of the 2-Dimensional approach (H_s - T_p) for the entire integrated 30-year period. The title indicates whether the subfigure is an EOF or PC, the mode concerned, and the associated percentage of the total variance values.

Figure 99Figure 9 shows that ϑp is initially split between the waves going southwestwards and those going northeastwards, with a fairly zonal gradient of PC1. Mode 2 is characterised by a meridional gradient of waves coming from the south and southwest, with an ongoing minor contribution to the east side of the quadrant. Mode 3 is characterised by the swell pools and land shadows, and seems to favour southwestwards waves as it increases the 45-105° WDM region and cancels the 225° peak of PC1. It is worth noting that ϑp 's accumulated first three modes' $\lambda\%$ reach about 70% of the distribution, which is about 10% above 2D values and 10% below H_s and T_p values. This is due to the multimodal character of the directional distribution.

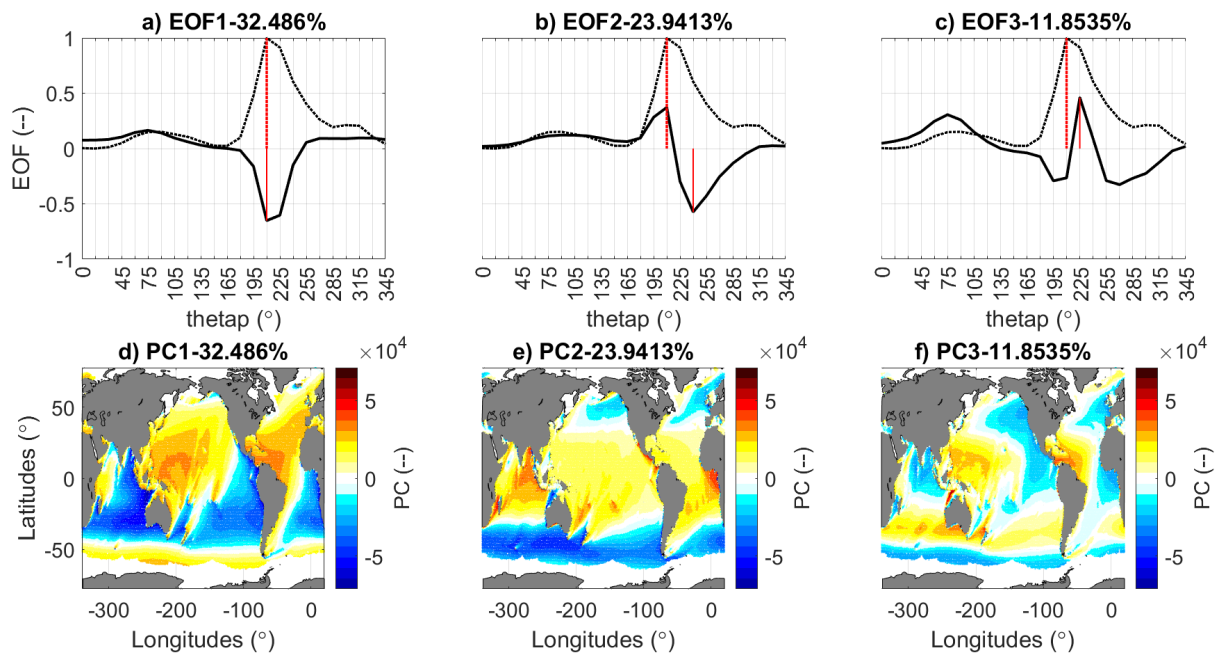


Figure 99. Same as Figure 66Figure 6, but for the wave peak direction (ϑp).

Once the three parameters are merged into 3D space, the EOFs must be displayed in cubes, such that Figure 10.10a-c's walls provide only the extrema to highlight the maximum/minimum variations appearing in each 2x2-plane direction ($Hs-Tp$, right, $Hs-\vartheta p$, left, and $Tp-\vartheta p$, bottom). A first remark is that, as observed for the other approaches, for all 1-3D approaches, the lower $\lambda\%$, the more peaks and noisier the EOF becomes.

3D's mode 1 emphasises mid-to-low-latitude oceanographic patterns, featuring a southwest-to-northeast diagonal dialogue arrangement that carves asymmetrical diamond-like spreads. These shapes emerge from the interplay of Tp and ϑp , with their wider bases principally over the land masses along northwest-southeast directions. This asymmetry forms tilted triangular lobes with mirrored extensions, each underlining (to the west) swell- versus (to the east) sea-dominated zones, and directional gradients. The EOF distinguishes peaks of seas versus swells, echoing the global 3D average (Figure B.22) except for the waves coming from the northwest quadrant. Mode 2 focuses on the patterns of Hs , drawing zonal bands that separate the low-to-mid latitudes from those of the higher latitudes of the Southern Ocean, with a low contribution from the Northern Hemisphere. Mode 3 provides more contrast between zonal and meridional regions. The 3D WDMs have very similar trends (extended directional spread around 90° for seas, with a bidirectional swell branch pattern between $180-270^\circ$, with a greater spread of the waves in high Hs for mode 2) except for the alternating signs across the three main occurrence cores. These cancel or increase the overall WDM hearts, e.g., EOF2's high Hs waves are in the south, and the three modes display a general positive accumulation of the seas from $20^\circ S-30^\circ N$ western pools of the Atlantic and Pacific Oceans.

The extension of South America southwards cuts the continuity of the Southern Ocean, and as the westerlies blow and build up eastwards, the South American continent provides a shadow effect on the west side of the South Atlantic Ocean. This discontinuity has repercussions visible not only in all ϑp 's modes but also in the other parameters. This partly explains the north-to-south positive continuity of the west side of the Atlantic Ocean in ϑp and Tp , PC1, and 3D PC2. Then, the southeast Atlantic Ocean connects to the Indian Ocean, itself fully connected to the Pacific Ocean in terms of negative PC values. Eventually, the southeastern side of the Pacific Ocean, with negative PC values, is practically disconnected from the southwestern side of the Atlantic Ocean, with positive PC values, at the Drake Passage.

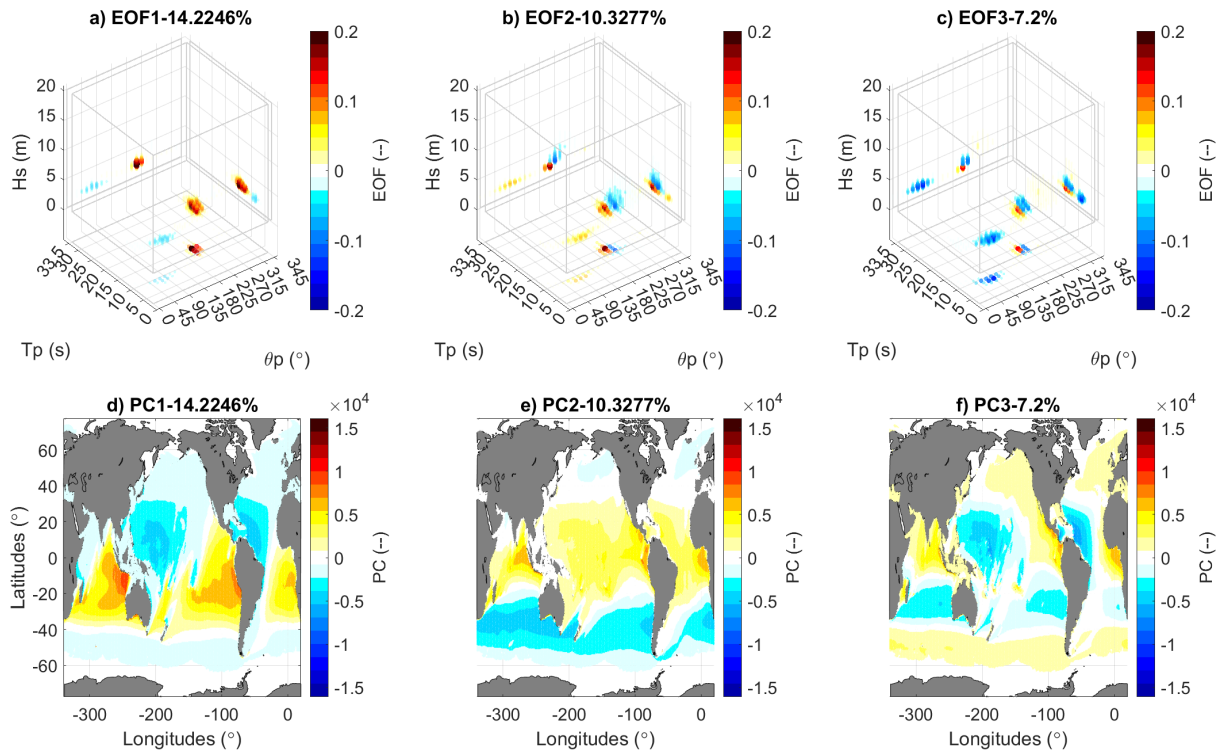


Figure 1010. First three Empirical Orthogonal Functions (EOFs) and Principal Components (PCs) of the 3-Dimensional space (based on the significant wave height, H_s , the wave peak period, T_p , and the wave peak direction θ_p) for the entire integrated 30-year period. The extrema of the projection of the EOF on the 2-Dimensional walls have been provided along with squares to help place the 3D distribution and the projections. The title indicates whether the subfigure is an EOF or a PC, the mode concerned, and the associated percentage of the total variance.

The first modes lack the large amplitudes associated with the fetch-limited domains observed in the previous parameters, especially visible in Figure 66Figure 6 for H_s , as well as for the 2D space (Figure 88Figure 8). This might be due to the diluted effect of H_s by the two other parameters (T_p and θ_p), which introduces similarities and may even overpower H_s . Then, not shown here, 3D's 4th mode emphasises the Indian-based wave variations, probably forcing the rest of the EOF decomposition. The northwestern incoming waves affect the 3D EOF decomposition only in the 5th mode (with a 3% representativeness of the global distribution), and all 7th-to-10th modes. The global wave climate organisation was observed to be dominated by the waves coming from the southwest, notably from the Southern Ocean. H_s - T_p amplitudes also align with the common, almost linear trend of the 3D average, such that an increase in H_s in the WDM generally involves an increase in T_p . However, where most of the data is located in smaller H_s - T_p pairs for the average, the strongest peak of the 3D first mode comes after the low H_s - T_p pairs. This is also the case for 3D's second mode.

To clarify the similarity between θ_p , T_p , 2D, and 3D's first mode, Table 33Table 3 provides the correlation between the PC1 of these approaches. As mentioned for H_s , the sign of the PC maps is not relevant as long as there is consistency between the EOF and PC pairs. To help compare the results with the figures, it has been specified which θ_p and T_p maps have been multiplied by -1. It is worth highlighting how the signs of the peaks of θ_p 's EOF1 indeed align better when θ_p 's mode is multiplied by -1, with high positive values around 210° and reasonably lower ones around 75°. This is also the case for T_p transformed into $-T_p$ and 2D and 3D, with both now possessing a positive peak around 14.5 s and a negative peak around 9.5 s. All correlations are reasonably high and positive, above 0.7. This table emphasises the effect of θ_p on 3D as its first and most prominent mode.

This is also the case for T_p on 2D, but it is worth remembering that 2D is nearly equally described by its first two modes. Surprisingly, ϑp and 2D (ϑp -independent approach) do not have the lowest correlation, and this may be explained by the reasonable correlation between ϑp and T_p . Yet, despite this similarity, the effect of ϑp on 3D dominates that of T_p , leading to $(-T_p) \times 3D$ and $2D \times 3D$ being the lowest correlations among these values. This stresses the lack of studies assessing the wave climates that may be encountered when the wave direction is omitted.

Table 33. Correlation values between the first Principal Components of the wave peak direction, ϑp , the wave peak period, T_p , the 2-Dimensional (2D) approach of H_s-T_p , and the 3-Dimensional (3D) approach of $H_s-T_p-\vartheta p$. A colour legend has been added from red for the higher correlations, through white, to blue for the lowest values, to help better compare the approaches.

Pair of approaches' PC1	$(-\vartheta p) \times (-T_p)$	$(-\vartheta p) \times 2D$	$(-\vartheta p) \times 3D$	$(-T_p) \times 2D$	$(-T_p) \times 3D$	$2D \times 3D$
Correlation value	0.76	0.74	0.87	0.92	0.72	0.73

4.4 DISCUSSION ON CLASSIFICATION

This section is divided into three subsections. Section 4.4.1 investigates the steps from 2-4 clusters in the wave classification for each wave parameter to help to understand a) their most important patterns; and b) the final 15-cluster and more complex classifications shown in Section 4.4.2 for each of the five approaches. Section 4.4.3 discusses and investigates these final classifications.

4.4.1 The first four classifications: the main trends of the wave parameters

Figure 11 displays the first three classifications of H_s (first three rows), T_p (next three rows), and ϑp (last three rows). Each set of three rows contains: 1) a top row projecting the clusters' colours on the first three PCs (those with the highest λ in the 10 PCs used for the classification); 2) a mid-row mapping the clusters following the same colour code as the top row; and 3) a bottom row that provides the WDM of each cluster wave climate. Since the crossing-period algorithm (Appendix C) has been developed only to cross analyses having the same number of clusters, the colouring is consistent per column.

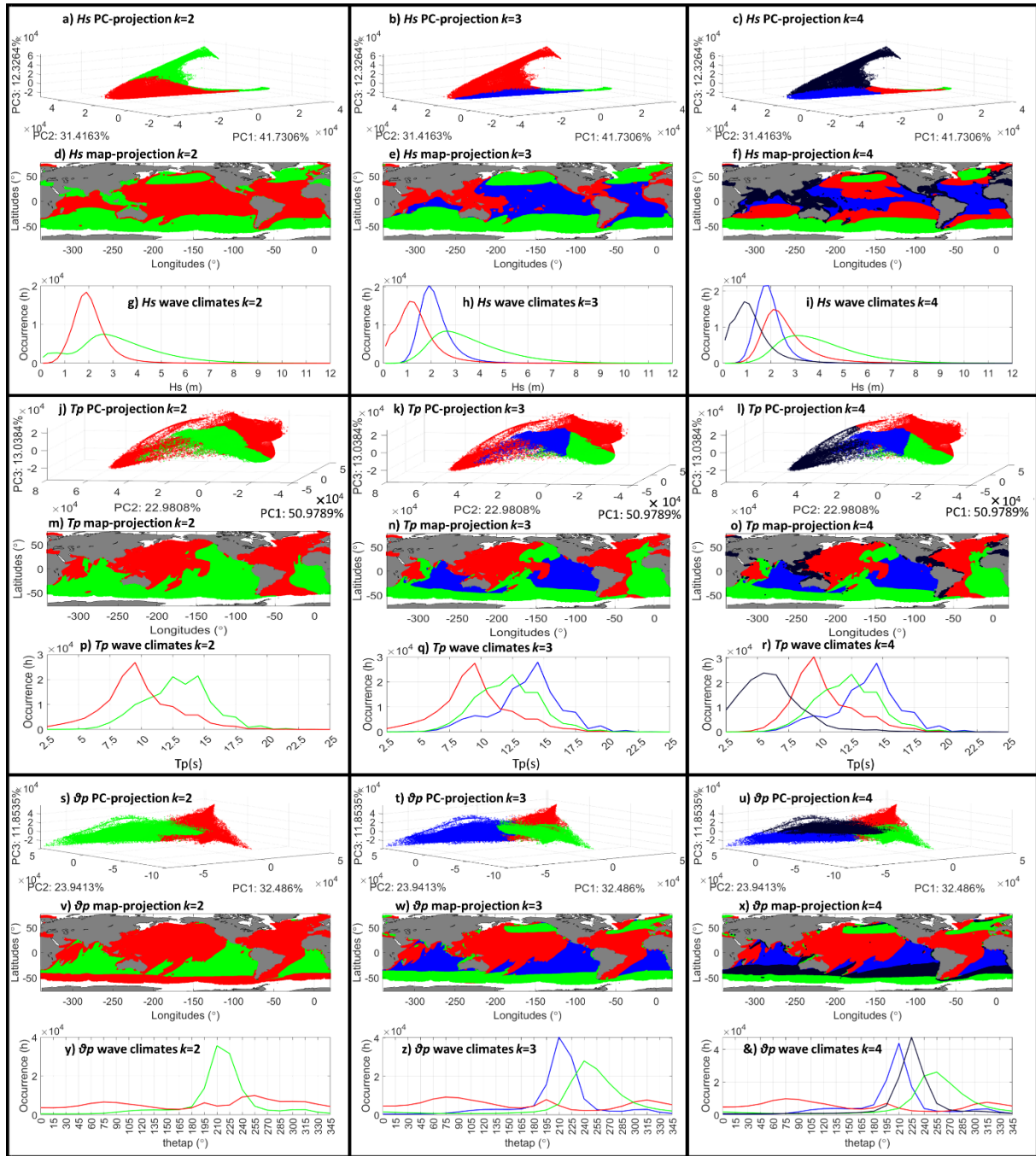


Figure 11. 30-year wave climate classification using 2-4 clusters and 10 modes for the significant wave height (H_s), wave peak period (T_p), and wave peak direction (θ_p). Panels a-b, j-l, and s-u are the clustering of Principal Component (PC1-3) space for the three highest representative modes. Panels d-f, m-o, and v-w project the obtained clustering of the right-above PC-space points on the globe. Panels g-i, p-r, and y-& provide the Wave Distribution Matrices of each right-above cluster.

At $k=2$, H_s is dominated by its zonal trends observed in Figure 66Figure 6.d (PC1), with a symmetry along the equator for the Pacific and Atlantic Oceans, which is impossible to obtain for the Indian Ocean due to its limited extension northwards. T_p is also dominated by southwest-northeast diagonal splits of the oceans observed in Figure 77Figure 7.PC1. Areas in the continental shadow (e.g., South America, Figure 99Figure 9) are removed from the southeastern-based wedge-like cluster. θ_p also draws the positive-versus-negative contours of its Figure 99Figure 9.d (PC1), and somewhat lies between both H_s and T_p . Fetch-limited domains affect the classifications from $k=2$ for H_s , $k=3$ for T_p , and in higher k for θ_p . Although with $k=3$ it could be interpreted based on H_s that

each new k increment involves the next higher mode's principal effect, Tp does not display the effect of the fetch-limited domains in its second mode. Consequently, increasing the number of classes in the classification does not fully evolve "linearly" in the increasingly representative order of the modes.

For $k=2$, the wave climate of ϑp (Figure 1111Figure 11.v) is very similar to that of Tp (Figure 1111Figure 11.j). However, the swell pools in ϑp dominate the distributions such that the Southern Ocean is merged with the northwestward regions in the red cluster. This leads to a spiky green WDM of waves coming from the southwest, and another wave climate much more broadly spread directionally. Next, the red climate of Figure 1111Figure 11.z is divided between the zones dominated by the westerlies in the mid-high latitudes, and the rest of the oceans. As k increases, ϑp 's climates do not really generate new well-defined zones as for Hs and Tp , but instead, the zone wave climates evolve as southwest-northeast diagonally directed transitional zones similar to a gradient contour map in the Southern Hemisphere.

While at every level, ϑp provides new climates that seem to appear "unexpectedly", the first two levels of Hs and Tp display a strong connection between a "mother" zone (green) at $k=2$ and the two "children" zones, red and green in the case of Hs , and green and blue for Tp at $k=3$. This is striking in Tp , where each child cluster contributes to one of the two peaks of the mother-cluster; and in Hs , where (at $k=2$) the low-kurtosis blue wave climate is removed from its low- Hs tail that becomes a whole new climate at $k=3$. However, in general, even though the new clusters often appear as sub-clusters of the previous k -level, each new cluster provides its own very distinctive WDM that may not be determinable in the mother-cluster.

At $k=3$, each of the three Tp clustering and WDM pairs is associated with one of the three peaks at 9.5 s, 12.5 s, and 14.5 s maps observed in Figure 44Figure 4. The northwest cluster at $k=2$ marks the distinction between the 9.5 s waves, while the southeast cluster marks the 12 s (which includes the 14.5 s waves). Then at $k=3$, the blue zones are highly correlated with the high-occurrence zones displayed in Figure 44Figure 4.e for the 14.5 s waves (happening in the Indian and Pacific Oceans but not in the Atlantic Ocean). This observation is consistent with Hs , for which the contours of high 1.9 m occurrence waves (Figure 44Figure 4.a) become increasingly accurate with increasing k values. Eventually, that cluster is the only one that remains throughout the increasing k , and the others keep on being redefined as collections of new clusters. This also seems to be the case for ϑp , but its 2nd and 3rd peaks are so weak compared to the rest of the distributions and the 1st peak that their effects are not visible for such low levels of k .

Finally, the spread of each classification and the contact between the clusters are visible in the associated PC-projection space, where the spread of the data and their repartition are more transparent as the number of clusters increases. For instance, the coastal and fetch-limited domains are clearly separated from the high-latitude regions, forming a structure in the three PCs for Hs (Figure 1111Figure 11.a-c) that resembles an open "crab claw" distribution. The upper-left extension or "fixed finger" (appearing in red and black in Figure 1111Figure 11.b and Figure 1111Figure 11.c) gathers the fetch-limited domains and nearshore regions. The "propodus", initially shown in red in Figure 1111Figure 11.a, becomes partially absorbed into the "fixed finger"

from $k=3$ onwards, separating the mid-to-low-latitude regions influenced by topography. Finally, the “*dactyl*” (extending into low PC3, negative PC2, and positive PC1 quadrant) captures the high-to-mid latitudes, where a latitudinal gradient gradually unfolds with increasing k . Similarly, with ϑp ’s increasing k , the red gets relocated to the back of the PC space, and the other climates form a band at the front of the PC space. This is a similar evolution to that of Tp . It should be recalled that only 3 PCs are displayed over the 10 considered, yet the range of PC values tends to decrease as $\lambda_{\%}$ increases, reducing their effect on the k-means classification. Overall, the spread of the data in the PC space is clearly different from that of the global map, drawing the differences in the wave climates. However, even though some points are clearly separated from others, there is always a transition area that invites considering more than only four clusters.

4.4.2 The final classification: classification using 15 clusters

Figure 12 provides the 15-cluster results for each of the five approaches, based on applying the k-means algorithm to the first 10 modes as introduced in Section 3.4. Each approach is presented by linking the cluster representative WDMs (the method to obtain each cluster’s WDM has been provided in Section 3.5) and their corresponding geographical zones. Namely, Figure 12.a provides the 1D WDMs of the 15 clusters of Hs WDMs classification; while Figure 12.d maps down the zones of each cluster following the same colour code (e.g., the dark green curve in Figure 12.a is the mean of all the points in the cluster of the same colour, mainly located in the southern high latitudes, as in Figure 12.d). Similarly, Figure 12.b provides the 1D WDMs of the 15 clusters of Tp with a colour code corresponding to the cluster map of Tp in Figure 12.e. The last 1D approach is ϑp , whose 15 clusters’ WDMs and zones are provided in Figure 12.c and Figure 12.f, respectively. The map of the classification of the 2D Hs - Tp histograms in 15 clusters is provided in Figure 12.g, and each 2D WDM of each cluster is provided in the subpanels of Figure 12.i-w, following Figure 12.g’s code colour linked to the 2D WDMs through the corresponding colour of the colourbox of the subpanel titles. This colourbox link between the maps and the WDMs has also been provided for the 3D approach (i.e. classification of the Hs - Tp - ϑp histograms), whose 15-cluster classification map is provided in Figure 12.h. For better visibility, the 3D WDMs have been provided separately in Figure 13, Figure 14, and Figure 15.

To help compare the classifications, the colours of the clusters in Figure 12 have been harmonised between the five approaches using the crossing-period algorithm (see Appendix C), but it is worth remembering that this is only a visual help. Although the 1D WDMs have been provided under the same occurrence scale, 2-3D WDMs have been provided with an adjusted colour scale to help better appreciate where each WDM’s heart of high occurrence values lies. Therefore, the reader is invited to keep in mind the colour scales’ order of magnitude. Nonetheless, the ranges of Hs , Tp , and ϑp are consistent between the subfigures. In the following, because of the 75 clusters considered (5 approaches, 15 clusters per approach), each specific cluster is referred to by its approach type and its index as in Figure 12’s legend. For instance, the royal blue cluster number 1 (cluster 1) is referred to as $Hs1$, $Tp1$, $\vartheta p1$, 2D1, and 3D1, for Hs , Tp , ϑp , the 2D cluster (of combined Hs and Tp), and the 3D cluster (combining all three parameters), respectively.

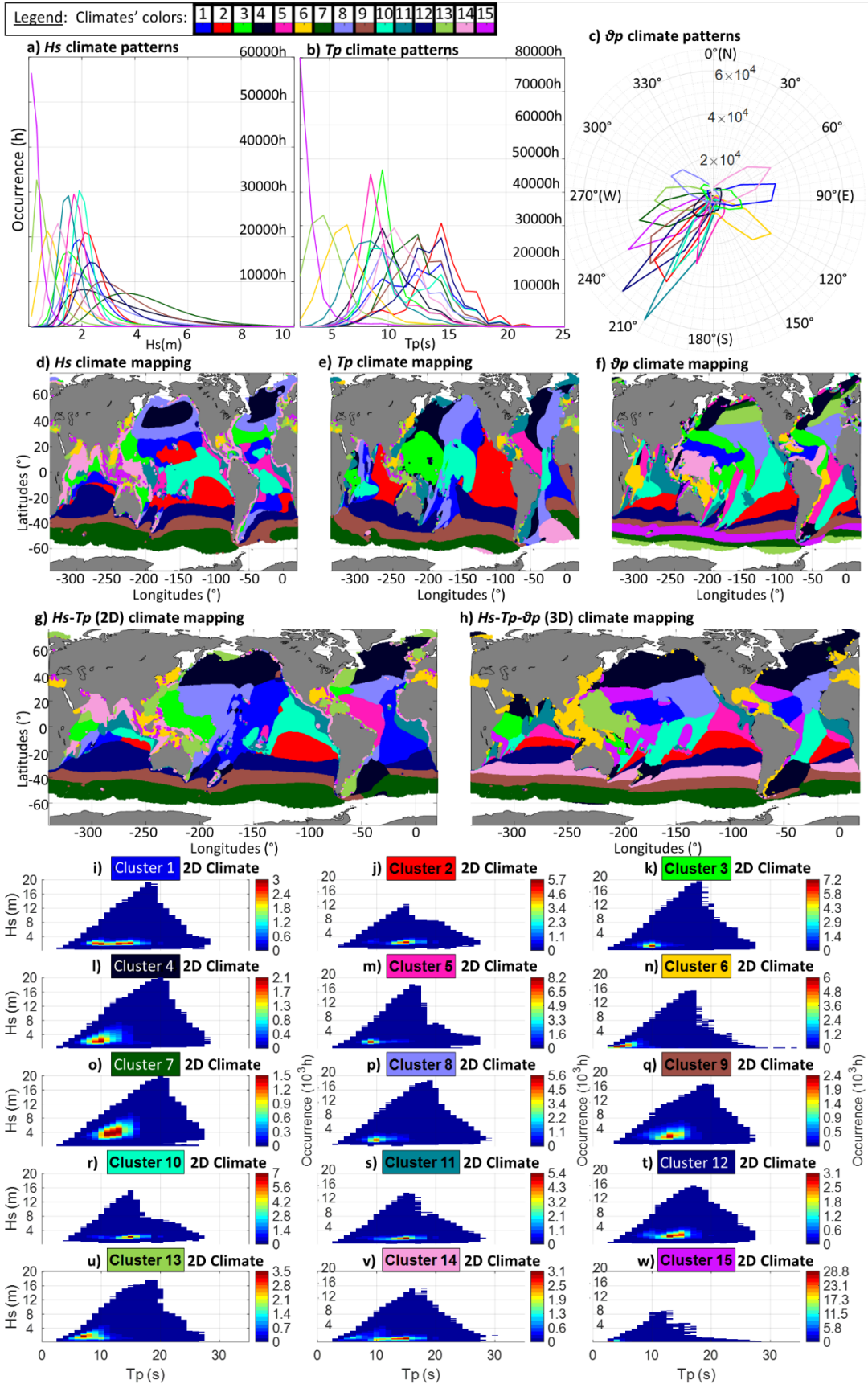


Figure 1212. 30-year wave climate classification using 15 clusters and 10 modes for the significant wave height (H_s), wave peak period (T_p), and wave peak direction (θ_p), along with the 2-Dimensional (2D) space of H_s and T_p , and the 3-Dimensional (3D) space composed of the three parameters. The climates' Wave Distribution Matrices (WDMs) are provided only for the three 1D approaches, and the 2D, Figures 13-15 provide the WDMs for the 3D approach.

1-dimensional 15-cluster consistency with the 2-4 cluster classifications

The general 1D clustering is consistent with that observed in Figure 11, both in the world cut and in the WDMs. Firstly, the equatorial symmetry of H_s is conserved (Figure 12.d); the zonal gradient of T_p is emphasised (Figure 12.e); and the mid-term pattern between H_s and T_p of ∂p (Figure 12.f) is distinguishable, with, instead of the symmetrical pattern of H_s , an asymmetrical pattern of the wave climates. Secondly, H_s does not present multi-peak wave climates, unlike the other two wave parameters. It is also worth noting that none of ∂p 's clusters comes from the north, either on the 0° side or near 360°. PCA is mainly a linear technique, so if many waves had occurred at either side, they would have generated two distinct clusters while, in fact, lying in a similar ∂p range. It is obvious that locations worldwide have waves coming from the north, but these results highlight the weakness of northern bins compared to other ∂p bins and motivate the validation of this classification method (further discussed in Section 4.4.3 and Appendix F).

Observations on the 1-dimensional approaches in the Southern Ocean regions

All three approaches display a general agreement on the Southern Ocean wave climate as a group of transitional zonal bands circling the entire Southern Hemisphere. H_s and T_p provide a meridional gradient of three main layers before the red cluster (no. 2), whilst ∂p provides seven, including the red and turquoise clusters (no. 2 and 10, respectively). This gradient resembles the evolution of the Sea Surface Temperature and its effect on the wind, and thereby the wave climate [56]. The zonal organisation of these clusters reflects the dominance of the Southern Hemisphere westerlies, where persistent storm tracks generate long-period swells that radiate equatorwards [23]. The progressive reduction of H_s and the increase in skewness towards lower latitudes indicate swell attenuation and dispersion as energy propagates northwards into warmer, more stable atmospheric conditions [24].

For H_s (Figure 12.d), this gradient consists of increasing skewness for decreasing H_s from 3.5 m for less than 10k hours (~4% of 30 years) to 2.1 m for more than 20k hours (~8%). For T_p , in the southernmost region (Figure 12.e), Figure 12.b shows that 12.5 s dominates, while going northwards, 14.5 s grows to overtake the 12.5 s peak in T_{p12} (dark blue) and then totally disappears in T_{p2} (red), as T_{p2} represents the swell pools but only in the Pacific and Indian Oceans. These long-period regimes correspond to the classic Southern Ocean swell corridors feeding the eastern basins, while the shorter, more variable periods closer to the westerly frontiers indicate mixed wind-sea conditions driven by transient cyclones [57].

The Atlantic Ocean is characterised by another series of clusters due to South America's eastern shadow, highlighted in Section 4.2. For ∂p , the high southern latitudes (Figure 12.f) are characterised by increasingly high skewness over the seven clusters, with waves coming from 270° shifting southeastward until 210° (Figure 12.c). This systematic directional veering marks the eastward curvature of dominant swell trains under the combined influence of westerly steering and the Coriolis effect [58], emphasising the dynamic continuity between the Antarctic storm belt and the subtropical swell-dissipation zones [57]. While H_s keeps a single-peak format in the red cluster (Figure 12.a), both T_p and ∂p present two additional

peaks at 9.5 s and 19.5 s for Tp , and at 105° and between 300-315° for ϑp , reflecting the coexistence of long-period swell arriving from the Southern Ocean and locally forced trade-wind seas in tropical and eastern-basin regions [54].

Different types of clusters and zones, and the connection between them

The peaks of θp_2 are even more pronounced in θp_{10} , such that θp_{10} appears as a zone that receives contributions from others. This zone and the surrounding key ones display a sort of evolving story from the first to the last. However, these are not necessarily distinct groups of zones that largely diverge from other zones (unlike coastal zones assessed postliminary). These “story zones” are quite prominent when observing Figure 12.12 Figure 12.d-f:

- a. For Hs distribution, from its symmetrical zones to the high latitudes:
 1. in the Pacific Ocean: Hs_{10} , Hs_2 , Hs_{12} in the south versus Hs_1 in the north, and
 2. in the Atlantic Ocean: Hs_5 , Hs_{10} , Hs_1 ;
- b. For Tp from west to east:
 1. in the low latitudes: Tp_3 , Tp_{10} , to Tp_2 , and
 2. in the high-north latitudes: Tp_{11} , Tp_4 , Tp_8 , Tp_1 ; in the Atlantic Ocean, Tp_1 appears only south;
- c. For ϑp in its asymmetrical pattern for both the Atlantic and Pacific Oceans: θp_{14} and θp_1 are surrounded by θp_3 affected by θp_8 in the northeast, and
 1. θp_5 for the Atlantic, and
 2. a mix of θp_5 and θp_{10} for the Pacific.

These are not only visible on the maps but also in WDMs in Figure 12.12 Figure 12.a-c. A striking example is Tp_{10} 's two peaks: one that could appear to come from the contribution from Tp_3 (at 12.5 s) and the other from Tp_2 (at 14.5 s). However, it is worth noting that when considering these peaks, the prevalent wave direction should always be kept in mind, and ϑp 's classification provides a good insight into the direction in which the “story” between these main zones should be read. ϑp highlights the dominance of the Ferrel cells (introduced in Section 1) in the Southern Ocean, while in the north, the circular generation areas between the continents provide a split in the directions. Then, Tp_4 somewhat aligns with θp_4 , sitting in the west of the northern basins. Eastward, though, Tp_8 and Tp_1 are cut horizontally by the θp_5 - θp_4 - θp_7 - θp_{13} - θp_8 gradient of waves nearly coming from southwest to northwest, almost perpendicular to θp_{10} 's main-peak waves. In the north, Tp evolves west to east with a decreasing peak at 9.5 s, while its kurtosis increases. This reflects the influence of the north on the whole Pacific and Atlantic Oceans, while the eastern coasts are strongly influenced by waves coming from the Southern Ocean.

Figure 12.12 Figure 12.d-f display an alternative to these story zones: the contour zones. These zones appear to border and embrace the shape of other zones, and do not make a clear climate block, but rather a meandering region between two or more compact zones, similar to what can be observed in the Köppen-Geiger climate classification [41]. This is the case for Hs_8 and is practically the case for θp_3 . This particularity of some clusters is more or less preserved depending on the ocean. Such is the case for Hs_3 and Tp_1 , which both appear as independent zones in the North Pacific and transitional zones in the Indian and South Pacific Oceans. Both the

story zones and the contour zones are quite reflective of the continuity of the data observed in the PC-projection spaces of Figure 11. Figure 11.

Observations on the 1-dimensional approaches in the northern mid-to-high latitudes

Overall, in the northern high latitudes, the three wave parameters provide very different patterns throughout Figure 12. Figure 12.d-f. H_s is 2-climate-based, although H_{s4} seems to be the central climate, and H_{s8} is a contour climate. H_{s4} is similar to H_{s7} and H_{s9} (low skewness and positive kurtosis, Figure 12.a) except that H_{s4} has a much lower H_s peak at 1.9 m, while H_{s8} is at 1.7 m. Together, they span the core of the midlatitude storm-track band ($\sim 40\text{--}60^\circ$ N) and its subtropical margins ($\sim 30\text{--}40^\circ$ N), with decreasing H_s outside the generation zone of H_{s4} , mostly towards the lower latitudes [23]. This H_s reduction is indicative of the more energetic swells travelling southwards out of the generation zone, crossing regions through which local wave generation may be negligible as the westerlies gradually weaken, accompanied by a reduction of the storm frequency [59].

Figure 12.b and e show that T_p is composed of story zones, which reduce T_p 's maximum occurrence while increasing its number of peaks towards the east, practically changing from neutral kurtosis and skewness to positive and then negative kurtosis and skewness. This evolution suggests a shift from energetic, predominantly wind-sea conditions in the western basins, which are directly forced by recurrent cyclonic activity [59], to mixed or swell-influenced regimes towards the east, where the decayed output of those same storm systems propagates after partial dispersion and refraction have occurred [54].

The split in ϑp in the northernmost regions mentioned above leads to waves moving eastwards initially dominated by the westerlies, and, on the eastern side of the split, they slowly move south-southeastwards. Eastward-directed trains follow the curvature imposed by Coriolis steering and the large-scale wind stress curl of the westerlies [56]. Conversely, the western side of the split, although characterised by a very flat omnidirectional trend, tends to travel from northeastwards to southwestwards, from one zone to the next in the southern direction. This is illustrated by the wave spectra shown by Echevarria et al. [17], and reflects the dual influence of the Aleutian Low (Pacific) and Icelandic Low (Atlantic), which steer the midlatitude westerly storm tracks [60], while lower-latitude deflection links to the tropical convergence zone influence [61]. This results in the waves reaching the Philippine coasts, thereby being more affected by the northern generation than the southern one. This also explains why the west is mainly separated from the east in the Atlantic and Pacific Oceans south of the northern generation areas.

Observations on the 1-dimensional approaches in the mid-to-low latitudes

Figure 12.e and f show that despite the similarities that θp_4 – θp_3 – θp_1 – θp_{14} and Tp_4 have in both the Atlantic and Pacific eastern low-latitude oceans, these are bounded by Tp_3 in the Pacific and Tp_5 in the Atlantic. Tp_3 and Tp_5 have similar climates, except that Tp_5 peaks at 8.5 s, lower than Tp_3 's, due to the smaller fetch of the Atlantic Ocean compared to the Pacific Ocean. When comparing the maps and WDMs of H_s and T_p (Figure 12.d and e for the respective maps, and a and b for the WDMs), in the southern and northern oceans,

they evolve in opposition out of the generation zones, which is especially visible coming out from the polar main generation zones (Southern Ocean, North Atlantic, and North Pacific). The effect of the southern (northern) large (small) fetches on Tp and its travelling and increasing patterns was discussed in Section 4.2. Then, as Tp increases, due to the conservation of energy with very little wave-energy loss during travel across the oceans [35], Hs decreases. Therefore, the waves that hit the eastern coast coming from the south are generally characterised by longer Tp and shorter Hs than the waves hitting the western coasts coming from the north.

These spatial contrasts reflect the fundamental organisation of the global wave climate in the mid- to low-latitude belts. In both hemispheres, the midlatitude westerlies and the mid-to-low-latitude persistent trade-wind systems create strong longitudinal asymmetries: energetic wind-seas generated in the western storm-track regions evolve into decayed long-period swells as they propagate eastwards under minimal local wind forcing [54]. The conservation of wave energy along these long trajectories explains the inverse Tp - Hs relationship, as energy spreads over increasing spatial scales and partly decays and dissipates through nonlinear interactions and air-sea friction [57]. The smaller Atlantic fetch limits the development of longer periods compared with the Pacific, consistent with global observations of weaker swell dominance in the Atlantic sector [24]. However, the two hemispheres remain inherently asymmetric: the Southern Ocean forms a near-continuous swell generator, transmitting energy equatorward from the circumpolar westerlies across an almost unbroken oceanic belt, while in the Northern Hemisphere, ocean-basin geometry constraints fragment the storm tracks of the North Atlantic and North Pacific, restricting fetch and reducing sustained swell development [17]. Consequently, eastern oceanic margins are primarily swell-receiving zones, whereas western margins are active wave-generation zones from the easterly trade winds [62] and western tropical storms [63]. This structure is maintained by the combined action of the westerlies, the subtropical highs, and the equatorial trade-wind belts that steer wave propagation and control energy dissipation across basins [23].

1-dimensional results distinguish offshore and open-sea regions from nearshore and closed-sea areas

Figure 12.12 Figure 12.d-f show a final and clearly distinguishable set of clusters that merges both the fetch-limited domains and nearshore areas. The latter generally display a smooth gradient from the offshore zones to the nearshore zones. Notably, when looking at Figure 12.12 Figure 12.a-c, the perpendicularity of the waves to the coast from refraction [4] is generally maintained in the cluster (e.g., θ_{p11} only affects the eastern coasts); and Hs and Tp tend to increase skewness and kurtosis towards lower values due to nearshore loss of energy (e.g., from bottom friction among other factors [64]). However, as seen regarding the story-zones, Tp remains peakier near the western coasts (Tp_{11}) and in the fetch-limited domains (e.g., Tp_{13} and Tp_6 in the Philippines, Gulf of Mexico, Persian Gulf, Mediterranean Sea). The main clusters orchestrating this nearshore climate are those of index 13, 6 and 15, the latter being the nearest to the coast and the most affected by the lowest values of Hs and Tp . These nearshore gradients are physically consistent with the depth-dependent dissipation and refraction processes [65] represented in the CAWCR wave hindcast, where bottom friction, depth-induced breaking, and spectral refraction parameterisations jointly reduce short-period energy and favour the persistence of longer-period components towards the coast [26]. In semi-enclosed or fetch-limited domains, basin confinement and

weaker wind forcing lead to shorter T_p and higher spectral peaking, as reproduced by the model through the same physical source terms [23].

Comparison between the five approaches' behaviour in the Southern Ocean

In the south, 2D (Figure 1212Figure 12.g) is similar to H_s and T_p (Figure 1212Figure 12.d-e), while 3D (Figure 1212Figure 12.h) introduces an additional layer in the cluster gradient, due to the effect of ϑp (Figure 1212Figure 12.f). In the north, though 2D and 3D concur in a single zone from the North Pole to about 30°N, therefore merging the variety of H_s , T_p , and ϑp clusters. This is also visible in the south under the effect of South America's shadow, where 3D gathers in one zone, T_{p11} , T_{p4} , and T_{p8} , θp_4 , H_{s3} , H_{s1} , and H_{s8} along with 2D13 and 2D4 (other clusters could be mentioned since 3D4 overlaps, such as clusters T_{p14} , θp_5 , and θp_{10} , or 2D8, but they are not fully contained in 3D4). 3D4 also encapsulates 2D13 and 2D4 in the north. 2D13's WDM (Figure 1212Figure 12.u) is mainly located in low ranges of the wave parameters (along with 2D6 and 2D15 of the nearshore and closed sea regions) with a higher peak range (~1.5-3.5k h about 1.2-0.9% of 30 years) than 2D4 (~0.8-0.4%) displayed in Figure 1212Figure 12.l. The result of this mix of zones provides a WDM of 3D4 (Figure 1212Figure 12.u) that is practically unique, with the exception of 3D6 having the same broad range of ϑp (Figure 1313Figure 13), but located at much lower values of H_s and T_p . Therefore, where the Southern Ocean is characterised by a similar cluster gradient, behind the South American continent's east coast and in the high northern latitudes, 2D and 3D (especially) fail to capture the many variations of the two-to-three parameters. This limitation mainly stems from the extreme zonal continuity and energetic dominance of swell in the Southern Ocean [54], which results in a blur of the multidimensional contrasts despite the added angular resolution of the 3D classification.

Comparison between the five approaches' behaviour in mid-to-low latitudes

In the mid-to-low latitudes, beyond the long mid-zones of cluster 1 in the Atlantic and Pacific Oceans, the 2D classification reveals very limited similarity across the three oceans (Figure 1212Figure 12.g). In contrast, the 3D classification preserves several consistent patterns between the Pacific and Atlantic Oceans (Figure 1212Figure 12.h). These similarities include a symmetric arrangement of 3D15 around 3D1, with eastern flanks dominated by 3D8 and 3D10, 3D10 being connected in the south to 3D2. Such similarities slightly extend further eastwards with 3D5 and 3D11. However, this extension remains limited: for instance, 3D5 is isolated in the north between 3D10 and 3D13 in the Atlantic Ocean, while it also connects to 3D8 in the Pacific Ocean. This richer structure of 3D, in comparison to 2D, arises from i) 3D's higher number of zones in the low-to-mid latitudes than 2D, particularly from 3D4 and 3D6, and ii) the influence of ϑp . Consequently, a) all approaches have the same number of clusters; b) the main lines connecting one cluster to the next worldwide are broadly consistent between 2D and 3D (although 3D adds more lines to 2D); and c) the multi-peak character of some T_p (Figure 1212Figure 12.b) and especially ϑp (Figure 1212Figure 12.c) clusters in these regions are gradually conserved from 2D to 3D. Altogether, a)-c) hinders 2D's ability to resolve seasonal or directional wave systems, which is an important limitation for various applications, including ship routing or WEC deployment strategies [66].

Comparison between the Indian, Atlantic, and Pacific Oceans

Although 2D displays some local multi-peak patterns (e.g., Figure 1212Figure 12.g,i,r,v), 3D is the approach most inclined to highlight the contributions of local and remote waves together. This is specifically visible in the Indian Ocean (Figure 1212Figure 12.g for 2D and Figure 1212Figure 12.h for 3D). Firstly, it is worth recalling the difference between the Atlantic and Indian Oceans' high northern latitudes. While the Atlantic and Indian Oceans have a narrower generation area than the Pacific Ocean, surprisingly, the Indian Ocean (Arabian Sea and Bay of Bengal included) tends to more closely follow the Pacific Ocean patterns, especially for Tp . This is the case for the large swell pools visible in Tp 's clustering of the Indian and Pacific Oceans (Tp2). This reflects the greater directional variability and coexistence of multiple wave systems in the mid-to-low latitudes [17], where swell propagation from remote generation zones combines with local wind-sea forcing, a complexity more effectively captured by the 3D approach. These Tp2's zones are subdivided by the 2D and 3D approaches. In the Pacific, the 3D case embraces Tp2's contour better than 2D. 3D splits the swell pools leading to 3D11 along the shoreline, and 3D5 and 3D2, which together embrace the swell pool shape. In contrast, the Atlantic's narrower width and the blocking effect of South America limit such eastward swell continuity, enhancing the dominance of locally generated wind-seas [13]. The link between the Indian and Pacific Oceans is mainly due to the lack of obstacles, such as the South American landmass, in the Atlantic Ocean.

Then the Bay of Bengal and the Arabian Sea have their own climates and wave generation processes, both strongly affected by local (in boreal winter, the Northern Hemisphere, being more energetic than the Southern Hemisphere) and remote (in the boreal summer, the opposite of the boreal winter) waves. These seasonal reversals correspond to the monsoon cycle, where alternating wind regimes generate short-period local seas [9] that coexist with the southward-propagating swells from the Southern Ocean [67], producing multi-peak Tp - Hs - ∂p signatures captured by the 3D approach. The seasonal effect, integrated over 30 years, has less influence in the Pacific and Atlantic Oceans since large generation areas are not only in the south but also in the north, while the Indian Ocean only possesses a southern generation area. As such, where 2D14 mainly joins both seas, ∂p distinguishes them in 3D. Still, 2D14 WDM (Figure 1212Figure 12.v) divides waves generated locally (low heart of wave parameters), and remote waves from the austral winter Southern Ocean (high Tp heart). The 3D15-3D3-3D10 Indian Ocean pattern is also obtained off the eastern Australian shores. 3D10 (Figure 1414Figure 14) and a little 3D3 (Figure 1313Figure 13) display three peaks of different combinations of Tp - Hs - ∂p ; and 3D10 is dominated by the southern waves. Eventually, in the Atlantic and Pacific Oceans, 3D8 (Figure 1414Figure 14) is the most striking multi-peak cluster, and so both 3D10 and 3D8 have lower mean maximum occurrence than 3D3 from such a multi-input wave climate: from the north, south, and locally. These differences emphasise the importance of both basin geometry and atmospheric circulation in shaping the multidirectional wave regimes, which 3D captures more comprehensively than lower-dimensional approaches [57].

Nearshore and closed sea regions for the 2- and 3-dimensional approaches

Fetch-limited domains, fragmented across a mix of clusters 6 and 13 (Figure 1212Figure 12.d-h), with sporadic contributions from 2D14 (Figure 1212Figure 12.g), are unified under cluster 6 in the 3D classification (Figure

Figure 12.h). However, a singular pattern emerges in the Caribbean Sea where the two distinct zones from the Western Pacific and Atlantic Oceans, 3D1 and 3D5, meet and interlock. 3D1 and 3D5, settled in sharp juxtaposition over the basin, evoke the complementary duality reminiscent of a yin-yang arrangement. Overall, the contrast between the multi-cluster character of the 1D classifications and the minimised cluster variety of the 2D and especially 3D classifications suggests that coastal, semi-enclosed, and nearshore regions could greatly benefit from a coastal classification. Such a study would exclude offshore waves to concentrate on the global shelves and enhance the depiction of dynamics tied to shelf and nearshore processes.

On the eastern and western sides of the oceans

Overall, the mid-to-low latitudes' east and west climates are very different. Where cluster 11 is generally present along the eastern coasts, H_s provides H_{s11} also in the Western Pacific (Figure 12.d), and T_p confines it to the west of the basins (Figure 12.e). Cluster 11 is generally single-peak based, although with a distinctly rounded peak for T_p around 8.5 s (Figure 12.b), whilst 2D (Figure 12.s) and 3D (Figure 15) provide it around 14.5 s, corresponding to the eastern high T_p trends analysed earlier here. Vice versa, although the western-Pacific zone of H_s does not exist in 2D and 3D, they both provide waves consistent with its 1.5 m peak. Western Pacific H_{s11} is contained in 3D13 and 2D3. However, where 3D13 is practically a single-zone cluster, 2D3 appears as a nearly symmetrical axis for 2D13 in all three basins. This symmetry probably comes from the influence of H_{s3} (Figure 12.a), which also peaks at 1.5 m, where 2D3 displays its maximum (Figure 12.k). 2D3's main zones, though, are in the Western Pacific and Indian Oceans. Therefore, the uniqueness of 3D13 probably comes from the intersection between 2D3 and the set of θ_{p3} - θ_{p1} - θ_{p14} , which is only present in the Pacific and Atlantic Oceans. Conversely, where 2D5 appears as a mainly single-zone wave climate from the effect of T_{p5} (versus T_{p3}), it mainly becomes the pair of 3D1 and 3D15 that appears west of 3D13 in both the Atlantic and Pacific Oceans. 3D1 (Figure 13) diverges mainly from 3D15 (Figure 15) by 3D1's more pronounced 3rd peak at 195°. 3D13 completely lacks any 2nd or 3rd peak. Further, 3D13 is centred on the main peak of 3D15. Finally, 3D1 revolves around 9.5 s, 1.5-1.3 m, between 60°-105° (3D15 being a tad more southwards) with a tail northwards for longer T_p (that forms the 2nd peak for 3D15 and 3D1). It is worth noting that 3D1 coincides with the Pacific Ocean's main concentration of 1.9m-10.5s-345° waves (Figure 5.e). The waves gather and merge into 3D13, such that the offshore waves (of 3D15) coming from the south lose their effect increasingly through 3D1. While on the western coast, 3D11 (Figure 15) is characterised by dominating very large T_p waves from the south. This, and the trends earlier observed in the Indian Ocean, highlight the different phenomena affecting the different sides of the oceans.

In conclusion, these west-east contrasts encapsulate the physical realism progressively gained from 1D to 3D approaches. By jointly resolving H_s , T_p , and θ_p , the 3D classification distinguishes swell-dominated eastern margins from wind-sea-dominated western ones. Long-period regimes along eastern coasts reflect Southern Ocean-generated swell that propagates into the Indian and Pacific basins [24], and the associated long-period energy signal along eastern margins is evident in the global atlas [68]. The 3D WDMs, such as 3D8 and 3D10, distinguish the northward-propagating swell systems with higher-energy waves (mostly higher T_p) from those

going southwards, originating from the aforementioned geometric confinement of northern storm tracks, opposed to the continuous energy input from the circumpolar Southern Ocean [17]. In contrast, the shorter-period, higher-skewness conditions on western sides arise from trade-wind forcing and storm-track variability [23], further shaped by boundary-current refraction, such as in the Kuroshio and Brazil Currents [13]. In the Indian Ocean, the southwest and northeast monsoons intensify local seas in the Arabian Sea, while the Bay of Bengal remains more exposed to Southern Ocean swells [9]. Together, these mechanisms illustrate how the 3D approach integrates local and remote forcing to reproduce global wave-climate asymmetries.

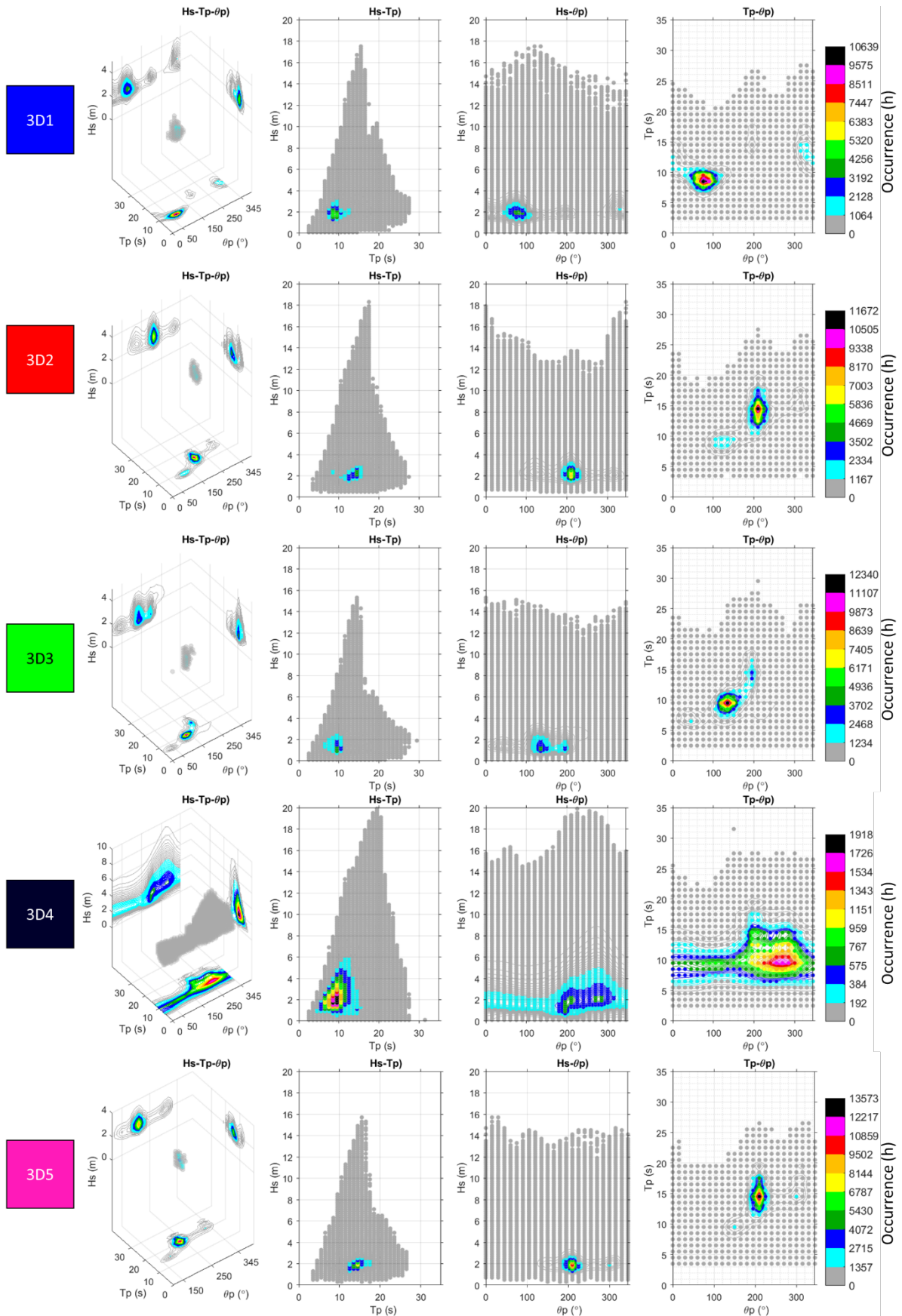


Figure 1313. 30-year wave climate classification using 15 clusters and 10 modes for the full 3-Dimensional (3D) space of the significant wave height, H_s , peak period, T_p , and peak direction, θ_p . Here, the first to the sixth classes are represented following the colouring of Figure 1212Figure 12.h.

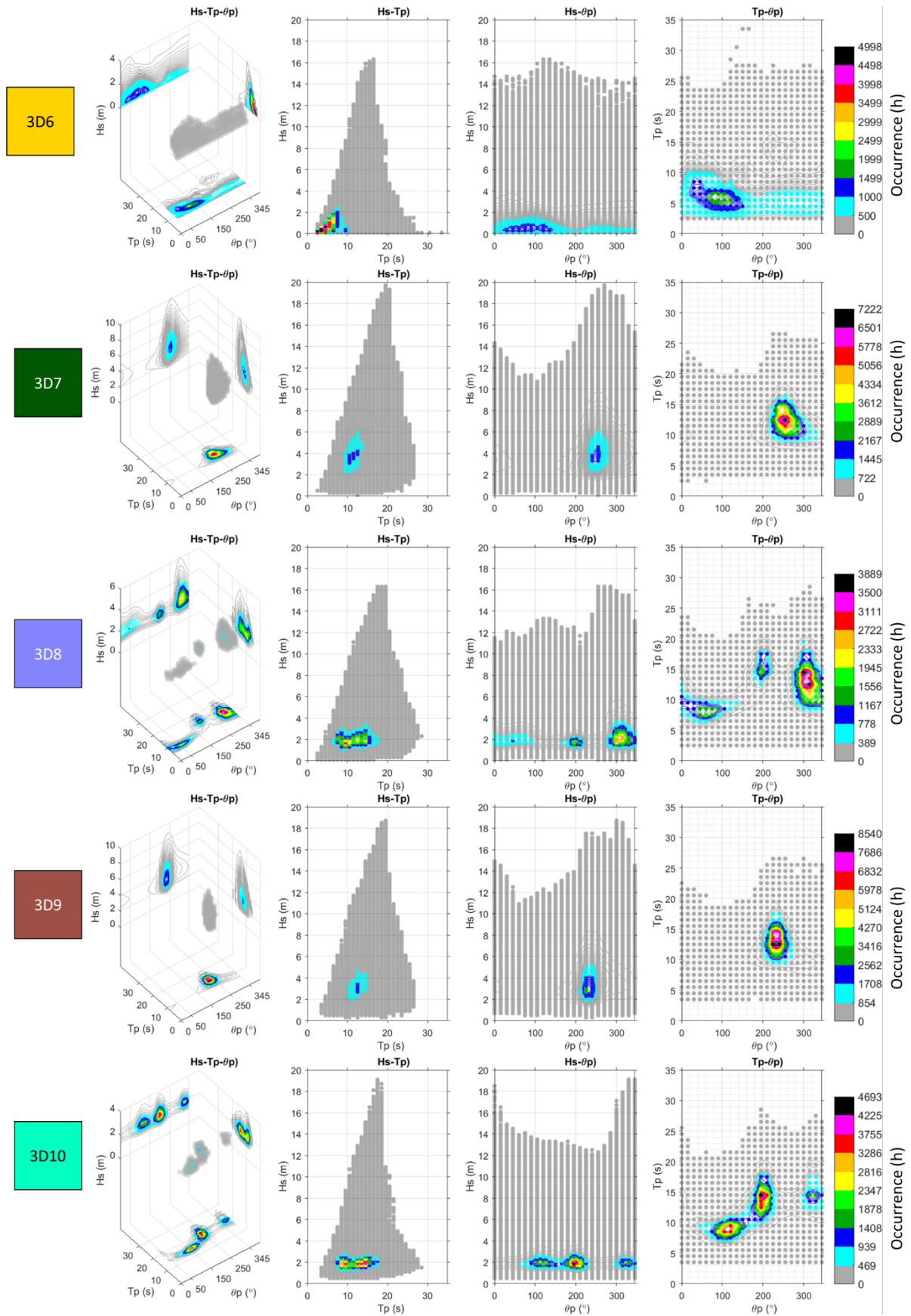


Figure 1414. 30-year wave climate classification using 15 clusters and 10 modes for the full 3-Dimensional (3D) space of the significant wave height, H_s , peak period, T_p , and peak direction, θ_p . Here, clusters 6 to 10 are provided with a numbering of the clusters that follows the colouring of Figure 1212Figure 12.h.

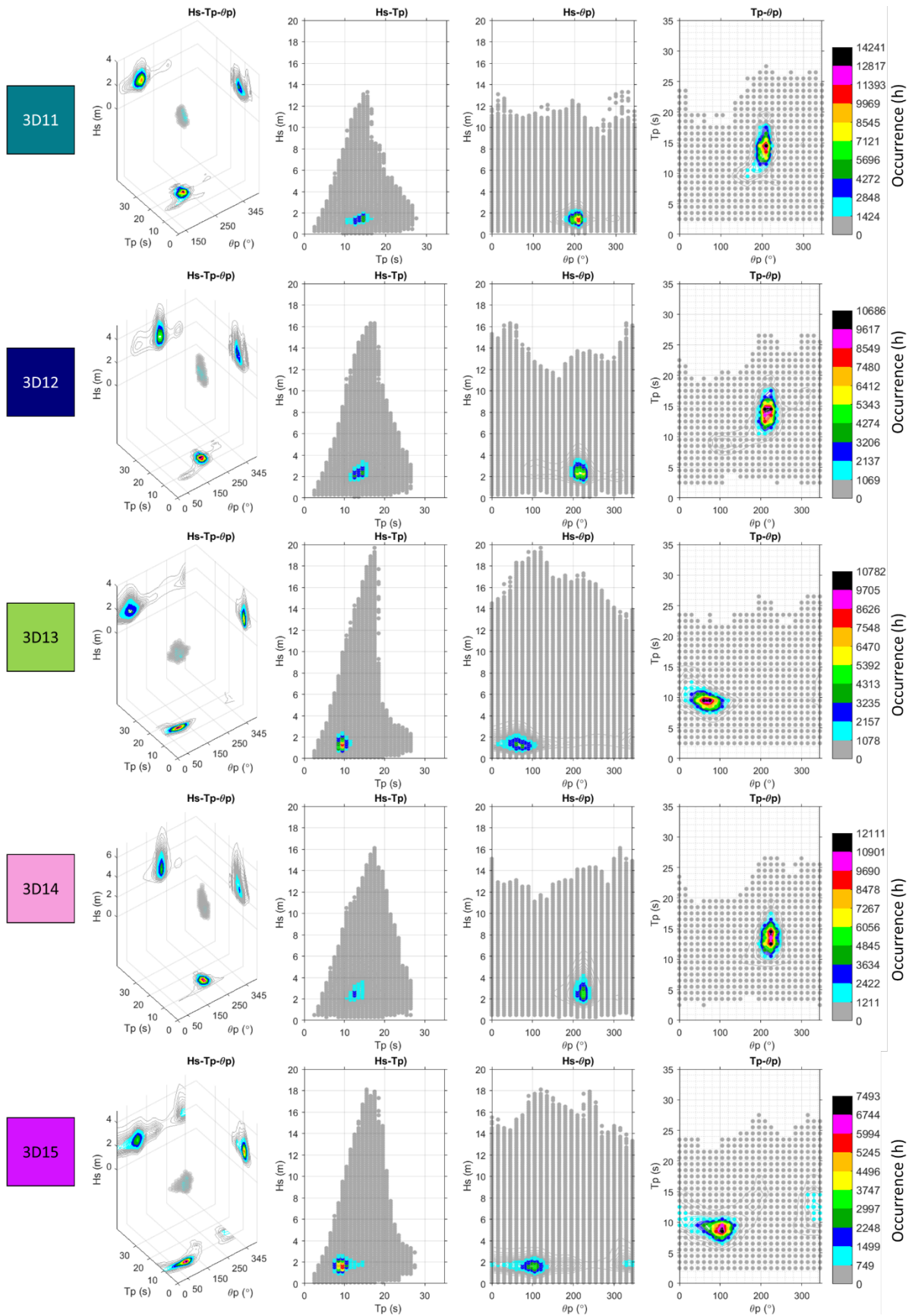


Figure 1515. Last classes of the 30-year wave climate classification using 15 clusters and 10 modes for the full 3-Dimensional (3D) space of the significant wave height H_s , peak period T_p , and peak direction θ_p . The numbering of the clusters follows the colouring of Figure 1212Figure 12.h.

Discussion on the limits of the inter-approach cluster comparison and clusters' spreads across basins

Cluster-indexing between approaches shows some inconsistencies, the main ones being listed here. Cluster 15, generally a coastal area consisting of small waves, appears offshore with a completely different shape for ϑp and 3D. Cluster 13, generally coastal, becomes a directional gradient cluster in the Southern Ocean for ϑp . Cluster 14 is mainly a closed-sea or nearshore cluster, whereas Tp shows it as the anomaly in the southernmost region, and 3D as part of the Southern Ocean cluster gradient. 3D8 is located in the mid-eastern Atlantic and Pacific Oceans, while 2D8 is in the western part of the Pacific Ocean. Yet, 3D8's three peaks around 1.9-1.7 m for the distinct groups of ϑp - Tp (south-14.5 s, northeast-9.5 s, and east-8.5 s), somehow match those of ϑp 8. Both 2D8 and 2D1 peak around 1.7-1.9 m, although Tp ranges around 9.5 s. Hence, inconsistencies lie mainly with nearshore/coastal and east-versus-west regions.

Despite apparent divergences in cluster indexing between approaches, the five classifications consistently outline the principal structures of the global wave climate. These variations mostly reflect differences in the physical sensitivity of each parameter: Hs emphasises energy magnitude, Tp the temporal maturity of the wave field, and ϑp its directional forcing. Yet, together they delineate coherent oceanic regimes that correspond to established atmospheric and wave-generation systems. To illustrate this connection, Table 4 summarises and references the literature of the dominant oceanographic processes governing each of the 15 global clusters across the five approaches. The identified clusters align closely with well-established atmospheric and oceanic circulation features. High-latitude clusters correspond to storm-track and Southern Ocean swell systems driven by westerlies and extratropical cyclones (e.g., clusters 4, 7, and 9), whereas subtropical bands reflect the progressive dispersion and decay of these swells as they enter more quiescent gyre regions (e.g., Hs 8 and cluster 12). Tropical and equatorial clusters are shaped by the persistent trade-wind regimes (e.g., ϑp 1, ϑp 5, Tp 3, 2D3, and 3D13) and, in the Indian Ocean, by the seasonal monsoon reversals (e.g., 3D3, Hs 14, and 2D14). Coastal and semi-enclosed clusters capture the strong attenuation of wave energy due to limited fetch, bathymetric effects, and continental shadowing (e.g., cluster 6 and Hs 15, Tp 15, and 2D15). Together, these spatially consistent physical regimes demonstrate that the multivariate WDM-based classification not only reproduces known global wave-generation mechanisms but also provides an integrated framework linking local and remote forcing (e.g., ϑp 8, 3D8, Tp 10, and 3D10) within a unified statistical–physical description of the world's sea-state variability.

Table 44. Brief physical description of the 15 clusters' main processes, per approach. SO stands for Southern Ocean, SE/E for SouthEast/East, and NE for NorthEast. Clusters dominated by contour zones are underlined (a * is added when there is a mix of zone types), while story clusters defined previously are in italic font, and the gradient zones are underlined with italic font.

Cluster index	Significant wave height, Hs	Wave peak period, Tp	Wave peak direction, ϑp	2-Dimensional space, Hs - Tp	3-Dimensional space, Hs - Tp - ϑp
1	Southwards subtropical transition fed by trade winds [23]	<u>*Merging point of local trades, and swell propagation from northern and southern generation [9]</u>	Northern trade wind-driven seas [23]	<i>Merging point of local trades, and swell propagation from northern and southern generation [9]</i>	Northern trade wind-driven seas and northern swell dispersion [23]

2	Tropical trade-wind seas with embedded swell influence in both hemispheres [54]	Long-period swell pools [9]	<u>Zonal SO bands; gradual veering with latitude</u> [23]	<i>Southern eastern subtropical regimes under subtropical highs and decayed swell</i> [54]	<u>SO northernmost swell decay and dispersion belt</u> [24]
3	<u>*Western regions of the basins and transitional trade generation, storm generation, and swell decay</u> [23]	Trade wind-driven seas pool [9]	<u>Coexistence of northwest, southeast, and east-originated waves</u> [17]	<i>Trade wind-driven seas pool</i> [9]	Multi-input regime combining SO swell, monsoon-modulated seas, and local generation [24]
4	High-latitude storm-track seas of the North Pacific and North Atlantic, around 1.9 m, bridging energetic generation zones and subtropical margins [23]	Shorter-period, wind-sea-dominated storms in western high latitudes with eastward dispersion of longer swells [59]	Westerly-steered storm directions, reflecting Aleutian and Icelandic Low influence [59]	High-latitude wind-sea regime plus secondary zone in the South Atlantic lee of South America's shadow [13]	Composite storm-track regime capturing both North Atlantic/Pacific belts and the SO lee-shadow zones; swell-dominant blurring under westerlies. [23]
5	Equatorial open ocean decayed swells [24]	NE-trade wind-driven Atlantic seas pool [23]	Meridional SSW swell corridors from SO wave refraction [24]	NE-trade wind-driven Atlantic seas pool [23]	SO final northernmost zone gradient towards the swell pools [9]
6	Nearshore/closed seas; fetch-limited [54]	Nearshore/closed seas; fetch-limited [54]	SE trade winds under Mascarene and South Atlantic highs [69]	Nearshore/closed seas; fetch-limited [54]	Nearshore/closed seas; fetch-limited [54]
7	<u>High-latitude southernmost belt of the SO swell belt westerlies dominated</u> [59]	<u>High-latitude southernmost belt of the SO swell belt westerlies dominated</u> [24]	<u>High-latitude southernmost belt of the SO swell belt westerlies dominated</u> [58]	<u>High-latitude polewardmost belt of the swell belt westerlies dominated</u> [17]	<u>High-latitude southernmost belt of the SO swell belt westerlies dominated</u> [22]
8	<u>Subtropical gyre belt transition; Northern swell dispersion and decay</u> [59]	<u>*Northern swell generation from low-pressure systems and nearly meridional transitional zones in the south</u> [69]	Northern weak westerly-easterly transition reduced local forcing bidirectional swell receiver from the northern and Southern Hemispheres [17]	<i>Swell dispersion, storm feeding, and trade wind-driven generation</i> [54]	Northern weak westerly-easterly transition reduced local forcing bidirectional swell receiver from the northern and Southern Hemispheres [17]
9	<u>High-latitude southernmost belt of the SO swell belt dispersion westerlies dominated</u> [59]	<u>High-latitude southernmost belt of the SO swell belt dispersion westerlies dominated</u> [24]	<u>High-latitude southernmost belt of the SO swell belt dispersion westerlies dominated</u> [58]	<u>High-latitude southernmost belt of the SO swell belt dispersion westerlies dominated</u> [17]	<u>High-latitude southernmost belt of the SO swell belt dispersion westerlies dominated</u> [22]
10	Equatorial-tropical superimposition (Pacific) and independent presence (Atlantic) of north and south	Bimodal wave system coexistence of the southern swells and local trade generation [17]	<u>High-latitude northernmost belt of the SO swell belt dispersion westerlies dominated</u> [58]	<i>Swell dispersion, storm feeding, and trade wind-driven generation</i> [54]	Southern swell-dominated mixed sea-swell regime, shaped by basin geometry and southern storm tracks

	originated swells [24]				[54]
11	Trajectory end of the SO swells, arriving from the west from the veering with latitudes in weak equatorial regions [23]	Western coast sheltered weak wave generation regions [23]	Trajectory end of the SO swells, arriving from the west from the veering with latitudes [23]	Trajectory end of the SO swells, arriving from the west from the veering with latitudes [23]	Trajectory end of the SO swells, arriving from the west from the veering with latitudes [23]
12	<u>High-latitude southernmost belt of the SO swell belt dispersion westerlies dominated</u> [59]	<u>High-latitude southernmost belt of the SO swell belt dispersion westerlies dominated</u> [24]	<u>High-latitude southernmost belt of the SO swell belt dispersion westerlies dominated</u> [58]	<u>High-latitude southernmost belt of the SO swell belt dispersion westerlies dominated</u> [17]	<u>High-latitude southernmost belt of the SO swell belt dispersion westerlies dominated</u> [22]
13	Nearshore/closed seas; fetch-limited [54]	Nearshore/closed seas; fetch-limited [54]	<u>High-latitude belt of western swells, westerly-dominated</u> [58]	Fetch-limited and nearshore-attenuated regime west-side class [70]	Midlatitude storm track western-basin regime [23]
14	Monsoon-forced local seas with admixed SO swell [13]	SO storm-track seas/young swell [23]	Easterlies and storm regime trade-dominated [23]	Bimodal monsoon–swell mix: local short-period seas vs remote swell [13]	<u>High-latitude southern belt of the SO swell belt dispersion westerlies dominated</u> [22]
15	Nearshore/closed seas; fetch-limited [54]	Nearshore/closed seas; fetch-limited [54]	<u>High-latitude southern belt of the SO swell belt dispersion westerlies dominated</u> [22]	Nearshore/closed seas; fetch-limited [54]	Equatorially symmetric trade-wind dominated [9]

Overall, clusters with zones situated in various oceans could be observed in Figure 11. This is significantly highlighted here with many more clusters: zones in different oceans still belong to the same cluster, no matter their location and, thereby, the different physical phenomena that lead to their generation. This questions the recurrent approach that consists of separating the oceans in strictly wave climate analyses, such as in Lobeto et al. [18]; doing so may enhance the possibilities of redundancies in the wave classification.

4.4.3 Discussion about the final classification

Figure 16 and Figure 17 provide the correlation and RMAD (see Section 3.5) of the results investigated in Section 4.4.2. The correlation provides an idea of how similar the WDMs' patterns are compared to the WDM of the cluster they belong to. RMAD provides a measure of the gap between the values of the WDMs and their associated cluster's WDM.

All approaches are characterised by positive correlations, except for ϑp , which is the only approach that shows negative values. Although the southern correlations are near 1, they decrease north of 30°S, especially in the western basins. Firstly, the lower correlation values generally follow the contours of the zones. These contours stress the transition of the waves from one zone to the next. There are, however, locations where the transition is more radical. Many of these abrupt transitions happen near or in the shadow of islands (behind the islands in

the direction of the waves). Further, the lowest and most negative correlations occur in areas where currents are strong, especially along the Kuroshio Extension, the Gulf Stream, the Malvinas/Falkland Current and Convergence Zone, and the Somali Current. Even the Agulhas Currents appear to affect the distributions. This dependency on the current is evident in the zones of *Hs*, with, for instance, the branch of *Hs*3 following *Hs*14 for the Gulf Stream, and in *Tp*, with the Agulhas Return Current for *Tp*9 (Figure 1212Figure 12). The Agulhas Return Current leads somehow to the Antarctic Circumpolar Current that seems to be mainly captured by cluster 9 for *Hs*, *Tp*, and 2D, splitting the Southern Ocean into its southern and northern parts. However, while *Tp* follows current main streams far less, and *Hs* is next in line, they both show larger spreads of good correlations and RMAD values than *ϑp*. *ϑp* hence appears much more sensitive to currents than *Hs* and *Tp*. Furthermore, a correlation between the *Hs* zones and the InterTropical Convergence Zone (ITCZ) could be investigated, especially with the trends in the Indian Ocean (e.g., *Hs*3 and *Hs*1) and the symmetry in the other two oceans, which are fairly tilted rather than lying perfectly along the equator [71]. This direct connection with the ITCZ is also perceivable in the contours of the cuts of the zones of the other approaches. Altogether, the current and topography have a direct impact on the winds [56] along with the ITCZ [72], and as these winds generate the waves, currents, topography, and the ITCZ also have an indirect impact on these waves, along with the direct impacts seen such as those from the currents, especially on *ϑp* as per the correlations and RMAD maps.

Hs and *Tp* approaches provide an almost-perfect correlation map, except for the coastal zones, the Arabian and Caribbean Seas, and around the Canary Islands. Even at their highest correlation, RMAD still highlights the transitional zones mentioned above. In particular, cluster 2 and *Tp*1 in the Pacific Ocean, *Hs*5 in the Atlantic Ocean equatorial region, and *Tp*3 and *Tp*10 in the Indian Ocean exhibit a lower RMAD transitional zone than the rest of the offshore clusters. Furthermore, *Hs*7 shows a heart of reduced RMAD near Kerguelen Island (also pointed out in wind anomalies by Chelton et al. [56]), which diverges from the general pattern of higher representativeness in the middle of the zone, getting lower away from this heart. These patterns are reproduced in 2D, as evidenced by their correlation. 2D (and even more 3D) shows that combining parameters decreases the representativeness of the clustering, suggesting the need for more clusters in the classification. Despite being more detailed, 2D (and 3D) manages to capture fewer variations.

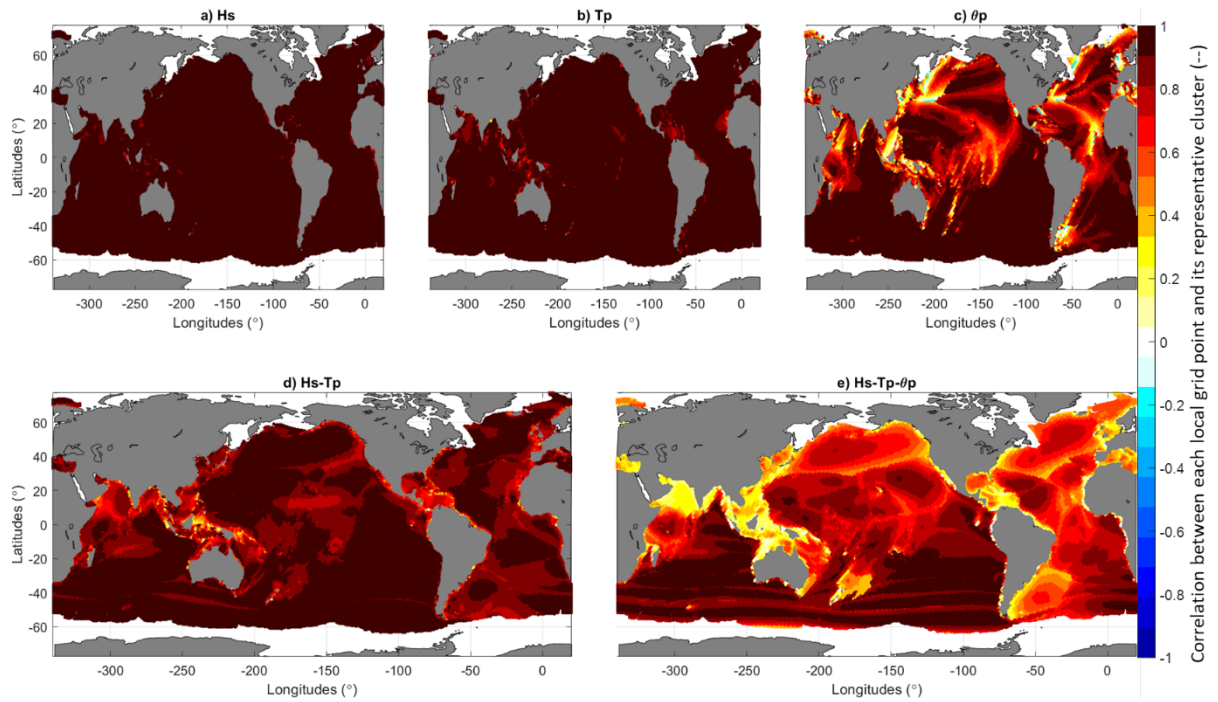


Figure 1616. Correlation surface of the climates' means versus the climates' contributing grid-points displayed in Figure 1212Figure 12.

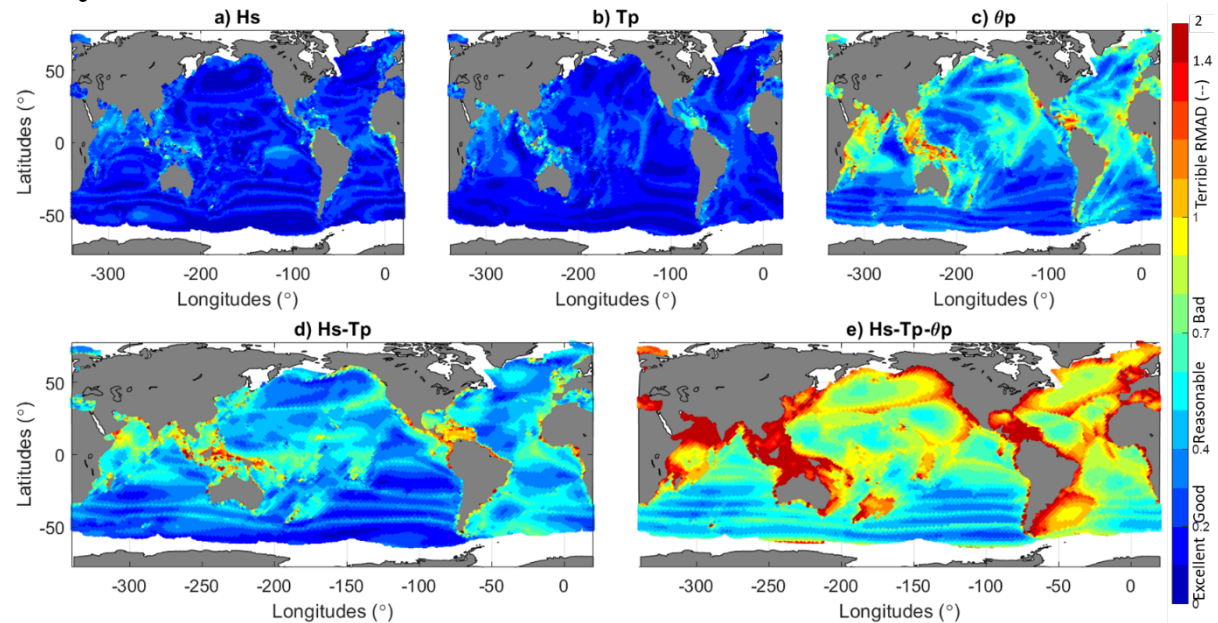


Figure 1717. Relative Mean Absolute Difference (RMAD, unitless) between the clusters, Figure 1212Figure 12 and their respective grid-points. The main ranges of colour interpretation have been provided on the colourbar.

Section 4.4.2 showed that 3D significantly reduces the clustering variation and, thereby, the cluster instability by reducing the noise of cluster alternation through gathering them into a single cluster (see examples of 3D4 and 3D6 investigated in Section 4.4.2). However, this higher stability, with lower cluster-alternation noise, leads to much lower representativeness (correlation/RMAD) of the clusters. This suggests that, despite being the most stable and detailed approach, 3D would profit from a higher number of wave classes. Some tests (not provided here as out of the scope of this research) showed that increasing the number of clusters would lead to the division of 3D4 and 3D6 into zones that more closely follow θp 's split.

The fact that the linearity of PCA may not be a significant problem in ϑp -related classifications is demonstrated in 1D with a map still dominated by good representativeness (most grid points have a near-1 correlation and are below the RMAD-bad threshold). However, ϑp significantly reduces 2D's representativeness in 3D. Yet, in most regions (aside from the transition zones), the representativeness values are below the bad-RMAD threshold and above 0.5 correlation. The critical areas are mainly those nearshore and fetch-limited domains of 3D6. This is also the case for 3D4, which poorly represents the zones near Argentina and in the Arabian Sea, below which 3D3 does not perform so well either. This can lead to future work assessing coastal and fetch-limited domains, on the one hand, and offshore grid points excluding the coast and closed sea grid points, on the other hand. This may also lead to considering other classification methods capable of integrating a linear model or an input form that is circular.

Finally, Figure 18 provides the independent 1D WDMs of the six largest zones of the aforementioned Hs2, Tp1, and θp 4 selected for their distinctive features, such as being a story-type or contour-type cluster, being symmetrical, or having lower correlation or RMAD. It is worth noting that the number of zones for a cluster may increase drastically when containing various zones characterised by a single or a few grid points. One may expect that each zone would contribute to a peak of the multi-peak WDMs, whereas Figure 18 highlights very common trends across the zones, confirming the notion of belonging to the same cluster, grouping similar WDMs. However, some zones do diverge from the cluster's WDM, especially in the Indian Ocean, showing more noise in their WDMs. Hence, what makes them part of the cluster nonetheless is their highest occurrence peaks, which are thereby among the most important features in this wave classification.

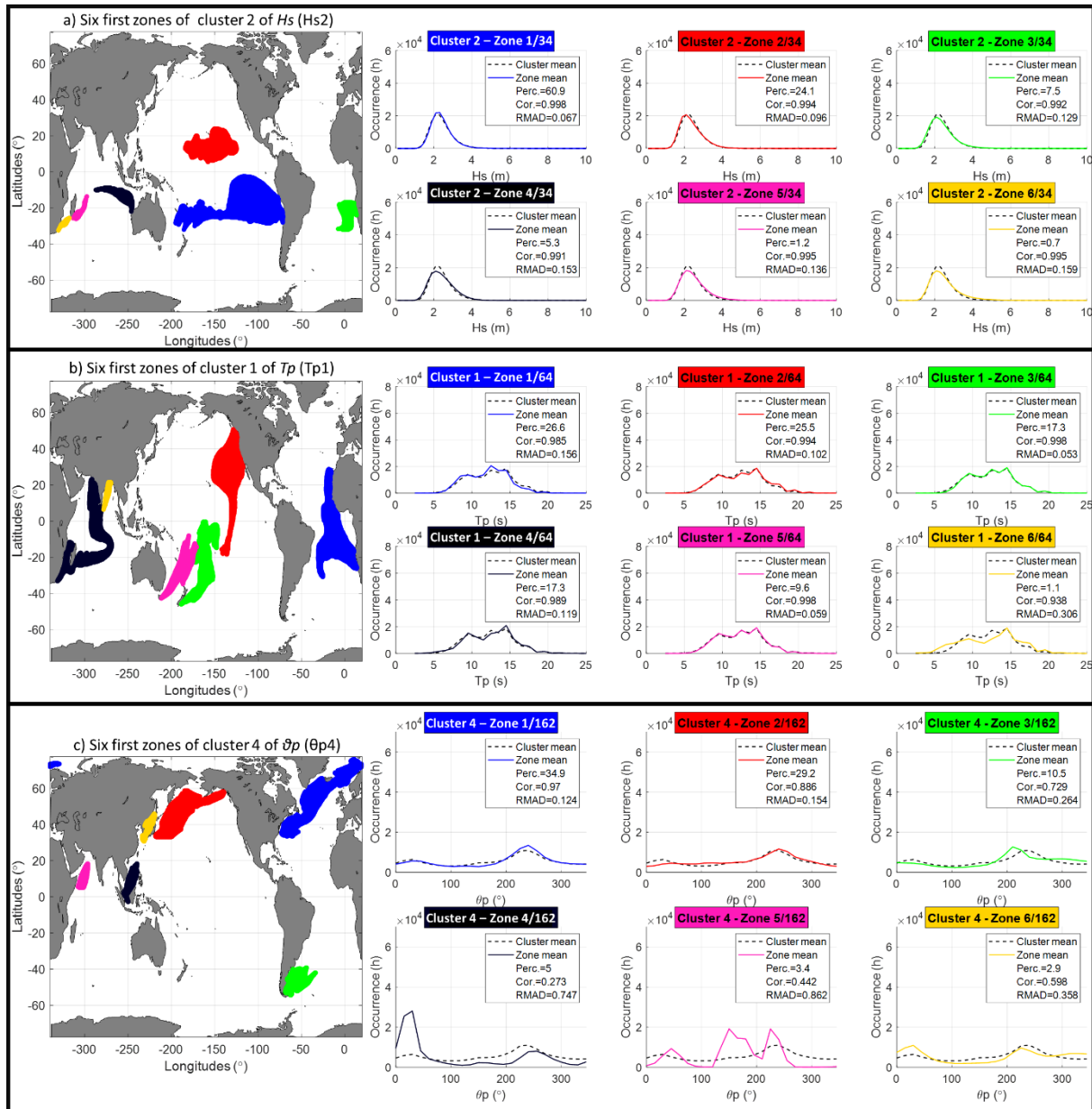


Figure 1818. Significant wave height (H_s), peak period (T_p), and peak direction (θ_p) zones for their cluster 2, 1, and 4 over 15 (from Figure 1212Figure 12), respectively. Each zone's colour and index over the total number of zones of the referred cluster is provided in the title of the subfigure. The Percentage (Perc.) of points in the zone relative to the entire cluster is provided in the legend. Below Perc., the Correlation (Cor.) of that zone wave distribution matrix and the one of the cluster is provided, along with the Relative Mean Absolute Difference (RMAD) between them.

While the present classification assumes a stationary wave climate over 1981-2010 to ensure spatial comparability across approaches, recent studies demonstrate that both natural variability and anthropogenic forcing introduce measurable long-term changes in global wave regimes. Tropical and extratropical wave climates are influenced by warming-driven variability and altered storminess patterns [22]. These are accompanied by increasing global wind speeds and significant wave heights, which are largely governed by swell trends and modulated by major climate modes, such as the North Atlantic Oscillation, Antarctic Oscillation, and El Niño Southern Oscillation [73]. Kamranzad et al. [74] further showed that 30-year averaging windows may still underrepresent multi-decadal fluctuations in swell and wave energy potential. Collectively, these findings emphasise that the apparent spatial stability of the classified clusters likely conceals the evolving magnitudes

and directional balances of the major wave climate modes. Future work should therefore incorporate temporally segmented analyses to evaluate how these global regimes are dominated by one phase or the other of the dipole-type of modes, which include the aforementioned North Atlantic Oscillation and El Niño Southern Oscillation, among others [75], and how they respond to interannual and multi-decadal climate variability along with their long-term changes.

5 CONCLUSION

This research provided a novel approach to wave-climate analysis, grounded in an engineering perspective inspired by wave-energy-converter assessments. The method discretises H_s (significant wave height), T_p (wave peak period), and ϑ_p (wave peak direction associated with T_p) independently, and combined into 2D (2-Dimensional space of H_s - T_p) and 3D (3-Dimensional space of H_s - T_p - ϑ_p) histograms that represent, for each bin, the 30-year occurrence in hours. These WDMs (Wave Distribution Matrices) are computed at every grid point worldwide. PCA (Principal Component Analysis) is then applied to identify the dominant representative modes for each 1D (1-Dimensional space: independent H_s , T_p , and ϑ_p), 2D, and 3D approach. A k-means classification subsequently uses the first 10 dominant modes to cluster the world into the main groups of similar WDMs. As such, it significantly reduces the computational complexity by bypassing the need to consider all individual points (a reduction factor of ~20,000). Finally, these groups are clusters, each containing one or more zones in the same or different ocean basins.

The modes proved capable of representing multiple wave systems simultaneously. For H_s , T_p , and ϑ_p , the first modes generally reflect the mean conditions and resemble the mean and especially the maximum mean-peak occurrence maps. Across all five approaches, the first mode generally depicts a two-peak global system, while the second and third modes refine key regional features of coastal and semi-fetch-limited domains for H_s , T_p , and 2D, and broader-scale atmospheric or current-related structures for ϑ_p and 3D. H_s is characterised by a strong meridional gradient distinguishing high latitudes from lower-latitude regions. T_p displays a clear northeast-to-southwest oceanic division mostly driven by the Southern Ocean, delineating south- and north-originating wave systems, orchestrated between seas and swell pools. ϑ_p nuances this structure further, reflecting the atmospheric circulation cells. Notably, the first modes reveal a few zones that, despite arising in different basins and from different generation processes, consistently group into similar clusters. This suggests that the oceans can be jointly assessed to avoid redundancy in the results, with only a handful of modes already offering a strong overview of the global wave structure.

Although the modes provide a clear view of the dominant global wave patterns, the classification does not evolve in a hierarchical structure. Because the world's wave regimes form a continuous domain, increasing the number of climates mainly rearranges previous zones rather than subdividing them. A higher-k classification zone may overlap two lower-k classification zones, which justifies the use of k-means over hierarchical methods. The wave parameters' main trends remain stable through the k-level classifications: zonal for H_s , diagonal for T_p , and a mix of both for ϑ_p .

All five approaches revealed a zonal wave-climate gradient across the Southern Ocean, extending from high to mid latitudes. From south to north, throughout these transitional bands, tall waves lose amplitude while increasing in period and rotate from eastward to northeastward directions. Despite the recurrent modal linkage between the North Atlantic and Pacific Oceans and the Southern Ocean, the cluster maps of Tp and ϑp often segregate them into distinct zones. Westward versus eastward waves in the east and west, respectively, and lower Tp (and Hs) elongating (decreasing) equatorwards. Western ocean basins are more dominated by the northern-originating waves, while eastern basins tend to have multi-peaked wave systems dominated by the southern swell, typically more energetic due to the Southern Ocean's larger wave-generation area. The Indian Ocean and the adjacent gulf and seas emerge as a region with irregular classification patterns across approaches. South America and nearby current systems produced localised zone(s) similar to those in the north. Coastal and fetch-limited domains consistently form distinct clusters across all approaches, due to their unique wave patterns.

These results revealed not only west-east and south-north trends, but also emphasised the interplay between remote versus local wave influences. Effects from non-wave parameters, such as the currents, topography, and ITCZ (InterTropical Convergence Zone), are also evident. Local effects, especially in the Indian Ocean regions, are particularly pronounced where the northern high-latitude wave-generation areas of the Atlantic and Pacific Oceans are absent. The Indian and northern Pacific Oceans are similarly classified in Tp . 2D maps reflect Tp 's clustering in the mid-low latitudes, while the north mainly consists of one dominating wave climate. 2D tends to reduce the high coastal-cluster variation in Hs and Tp (also present in ϑp). Finally, the complete 3D approach provided the least noisy classification, but with a lower representativeness than the other classifications. Hs and Tp yield the highest representativeness, followed by 2D and ϑp . ϑp 's lower (sometimes negative) correlations in regions affected by currents or topography obstacles essentially contribute to 3D's relatively reduced performance. Notably, 3D collapses coastal and fetch-limited domains into a single dominant wave climate, similar to the high-latitude grouping in the Northern Hemisphere, whereas mid-to-low and southern latitudes remain more fragmented than in 2D.

To preserve comparative consistency, the same number of clusters was used for all approaches. However, results suggest under-clustering in northern, fetch-limited domains and coastal areas that may benefit from higher k values. Future improvements could involve: 1) separating offshore versus coastal and closed sea classifications; 2) increasing the spatial resolution in nearshore zones to assess the wave climate evolution from offshore to coastal areas; 3) evaluating circular methods for the PCA and classification when ϑp is involved. Additionally, this study assumes a stationary wave climate, meaning that potential interdecadal variations or long-term trends related to climate change or oscillatory modes, such as El Niño and La Niña or the North Atlantic Oscillation, are not explicitly captured. Moreover, it is worth noting that although a review of the in-situ validations conducted for the CAWCR hindcast used here has been provided, this work does not perform complementary in situ validation. The interpretation of the results regarding the distinction between swells and seas is principally conducted using a 10-s threshold, which may not hold in all regions. The brief wind overview could be further considered to compute the wave age and help conduct this distinction. Furthermore, this study only considers

the dominant wave spectral peak. Future studies may explore the consideration of more spectral partitions, which would require balancing spatial resolution and the number of partitions with regards to computational capacities, and better refine the separation between swells and seas. Ultimately, the next step in this research is to introduce the temporal evolution effect on these climates, enabling the quantification and projection of long-term trends in wave regimes under changing climatic conditions.

DATA AVAILABILITY

The data used to conduct this research are available at the links provided in the references. Concerning the data resulting from this research, the five approaches' fifteen classes indices are available at https://archive.org/details/classification_thetap_15clusters_1_10_30yearfull.

GLOSSARY

Short name	Units (N/A for Non-Applicable, -- for non-dimensional)	Full Name
1D	N/A	1-Dimensional (either H_s , T_p , or ϑp)
2D	N/A	2-Dimensional (mainly H_s and T_p)
3D	N/A	3-Dimensional (joint H_s , T_p , and ϑp)
C	Context/variable-dependent (e.g., m, s, °, m/s)	Covariance matrix
CAWCR	N/A	Centre for Australian Weather and Climate Research
CFSR	N/A	Climate Forecast System Reanalysis
Corr.	--	Correlation
EOF	--	Empirical Orthogonal Function
H	--	Total number of bins over H_s
H_s	m	Significant wave Height
IEC TS	N/A	International Electrotechnical Commission Technical Specification
IM	Context/variable-dependent (e.g., m, s, °, m/s)	Input Matrix
ITCZ	N/A	InterTropical Convergence Zone
M	--	Number of grid-points
MAE		Mean Absolute Error
Max	Context/variable-dependent (e.g., m, s, °, m/s)	Maximum

Min	Context/variable-dependent (e.g., m, s, °, m/s)	Minimum
PC	--	Principal Component
PCA	N/A	Principal Component Analysis
R	--	Pearson (linear) correlation coefficient
RMSE	Context/variable-dependent (e.g., m, s, °, m/s)	Root-Mean-Square Error
nRMSE	--	normalised Root-Mean-Square Error
RMAE/D	--	Relative Mean Absolute Error/Difference
SAM	N/A	Southern Annular Mode
SOI	N/A	Southern Oscillation Index
S	--	Total number of sea-states
s	--	Number of selected sea-states
SI	--	Scatter Index
Std	Context/variable-dependent (e.g., m, s, °, m/s)	Standard-deviation
T	--	Total number of T_p bins
Te	S	Wave energy or mean period
θ	--	Total number of bins over ϑp
ϑm	° (degrees)	Mean wave direction
ϑp (thetap)	° (degrees)	Wave peak direction (unless mentioned otherwise it indicates where the waves come from)
ϑw (thetaw)	° (degrees)	Wind direction (unless mentioned otherwise it indicates where the winds come from)
T_p	s	Wave peak period
W	m/s	Wind speed
WDM	N/A	Wave Distribution Matrix
WEC	N/A	Wave Energy Converter
X	N/A	Eigenvector matrix

ACKNOWLEDGEMENTS

The first author would like to greatly acknowledge both supercomputers (High-Performance Computers, HPCs), Delta 2 (previously Delta) from Cranfield University and SSH aquia4 from the University of São Paulo. Deepest appreciation to Guillaume Dodet, who contributed greatly to this research. Thanks especially to Karl Jenkins and Irene Moulitsas for the support of this project and support with access to Cranfield's HPC (led by László Könözy) and guidance in using it. A kind word to Rodger Tomlinson and Amir Etemad-Shahidi, with whom this adventure started. A final acknowledgement for the precious advice and support of Jean-Marc Delouis, Jean Tournadre, Tina Odaka, Bertrand Chapron, Nicolas Rasclé, Lisa Maillard, Marine De Carlo, Simon Barbot, Fanny Girard-Ardhuin, and Fabrice Ardhuin, along with Arthur Avenas, Edouard Gauvrit, Clément Pouplin, Aline Zribi, Cédric Prevost, and Mickaël Accensi. Furthermore, the first author conducted this research under Doctor of Science/Philosophy scholarships, as detailed in the Financial Support section below.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTION

Ophélie Choupin: Conceptualisation, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualisation, Writing – original draft preparation, Writing – review & editing. **Joseph Harari:** Funding acquisition, Project administration, Resources, Supervision, Investigation, Validation, Writing – review & editing. **László Könözy:** Data curation, Software, Resources, Validation, Writing – review & editing.

FINANCIAL SUPPORT

The main author started this research in 2018 while receiving the Postgraduate Research Scholarships from Griffith University. This specific research then got financed by a scholarship processo 88887.614992/2021-0 from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) /PROEX (Programa de Excelencia Acadêmica), later replaced by two FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) scholarships, processos n° 2022/06765-8 and n° 2022/13873-1. This work was also supported by ISblue project, Interdisciplinary graduate school for the blue planet (ANR-17-EURE-0015) and co-funded by a grant from the French government under the program "Investissements d'Avenir" embedded in France 2030.

Appendix A : Wave and wind parameters' statistics on the time series

To help better understand the results, statistics have been conducted on the major parameters involved in this study and in relation to this study. Namely, the first three main parameters considered here are H_s (Figure A.11Figure A.1), T_p (Figure A.22Figure A.2), and ϑ_p (Figure A.33Figure A.3). As ϑ_p is a directional parameter, the circular statistics from the package of Berens [76] have been used. As one of the most considered parameters in the wave energy field, and combining both the wave height and period, the wave energy is also provided here in Figure A.44Figure A.4, with the aim of helping understand the H_s - T_p bivariate distribution and EOF decomposition. Finally, the gravity waves considered here are those generated by the wind, in order to better understand the source of the results, so the 10 m wind total amplitude and direction statistics are shown in Figure A.55Figure A.5 and Figure A.66Figure A.6, respectively.

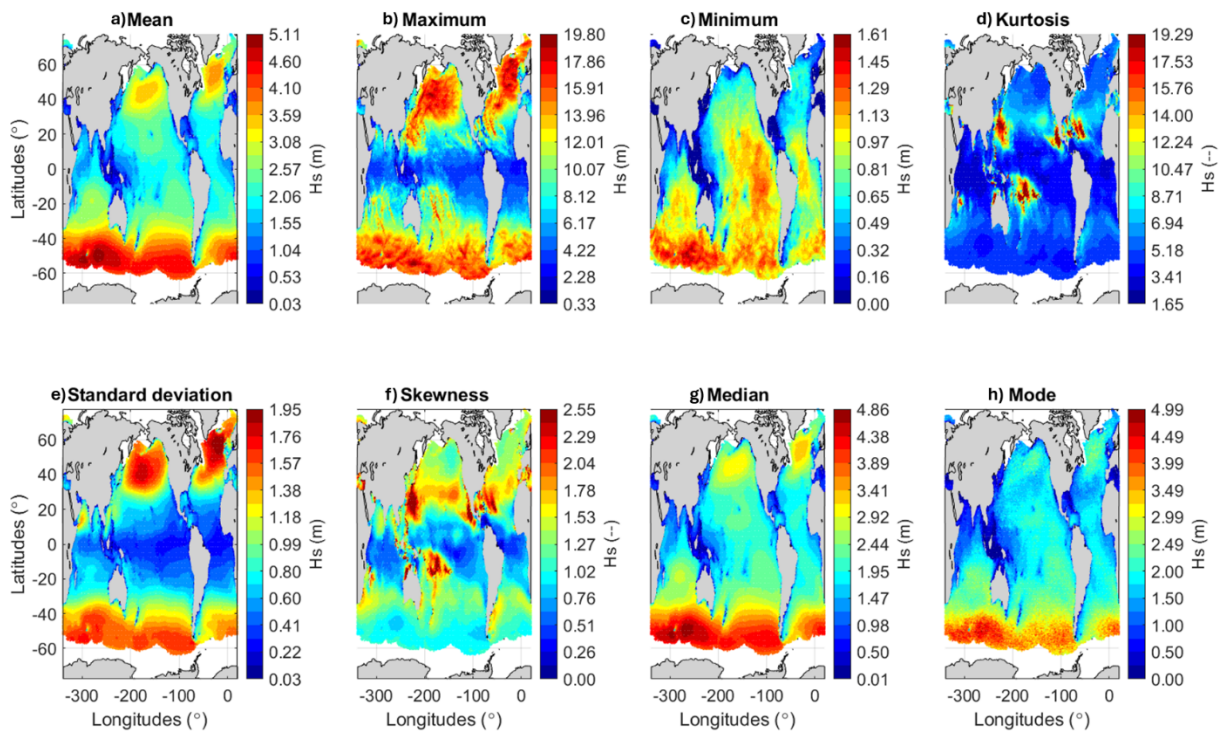


Figure A.11. Statistics over the significant wave height (H_s) global histograms over the 30-year period from 1981 to 2010.

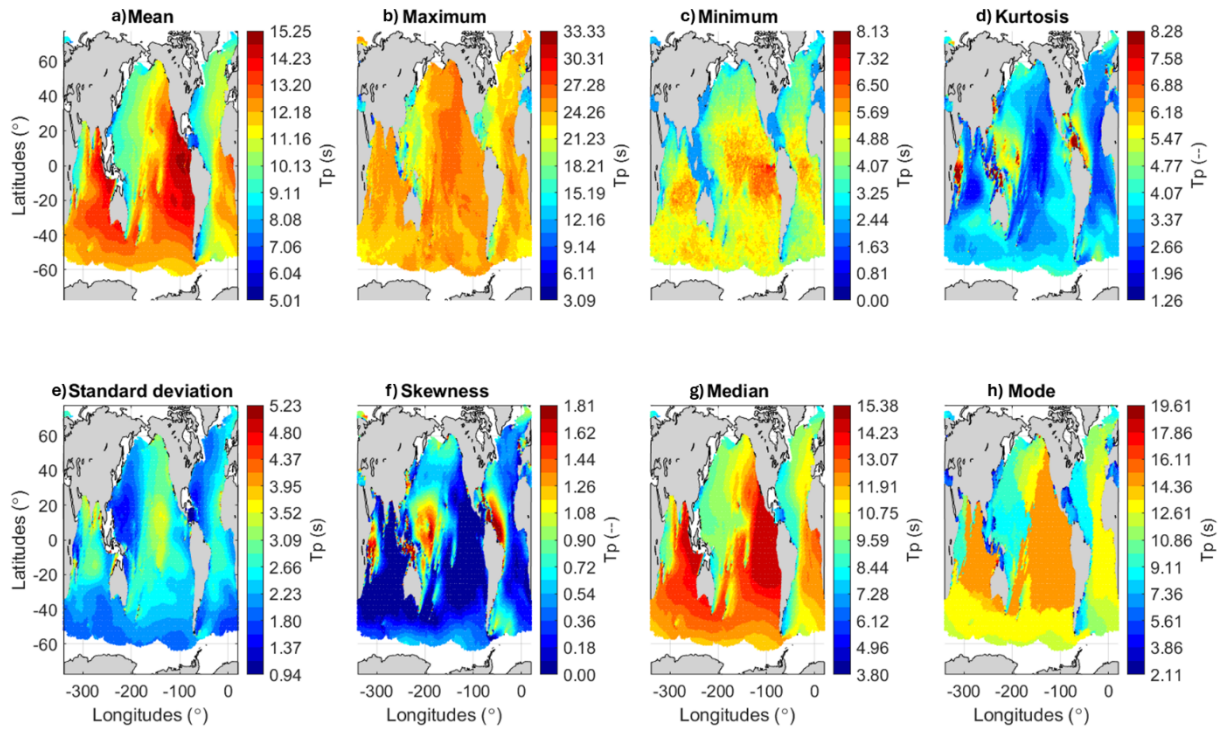


Figure A.22. Statistics over the wave peak period (T_p) global histograms over the 30-year period from 1981 to 2010.

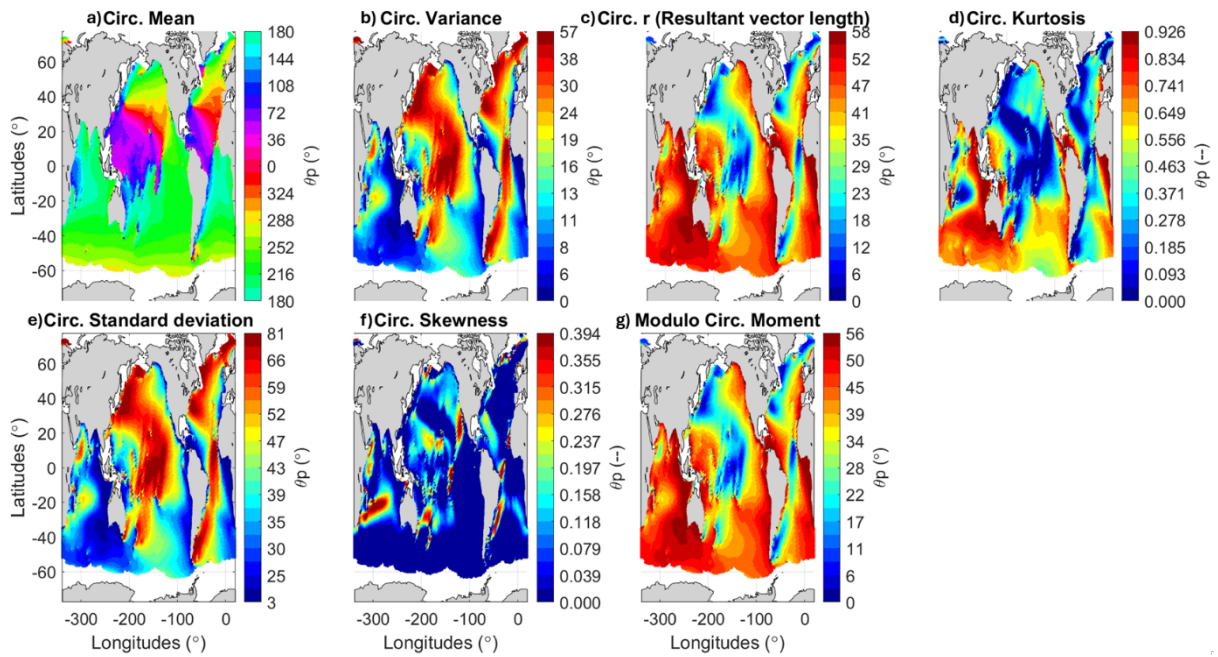


Figure A.33. Circular (Circ.) statistics over the wave peak direction (θ_p) global histograms over the 30-year period from 1981 to 2010.

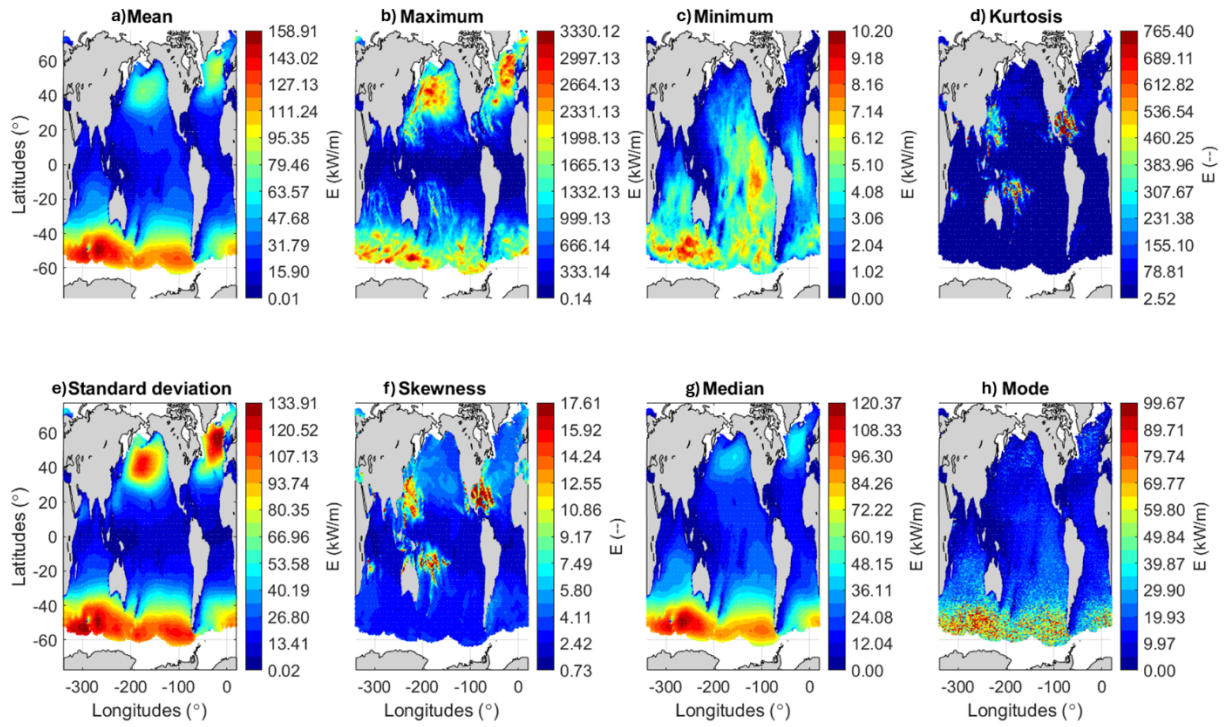


Figure A.44. Statistics over the wave energy (E) global histograms over the 30-year period from 1981 to 2010.

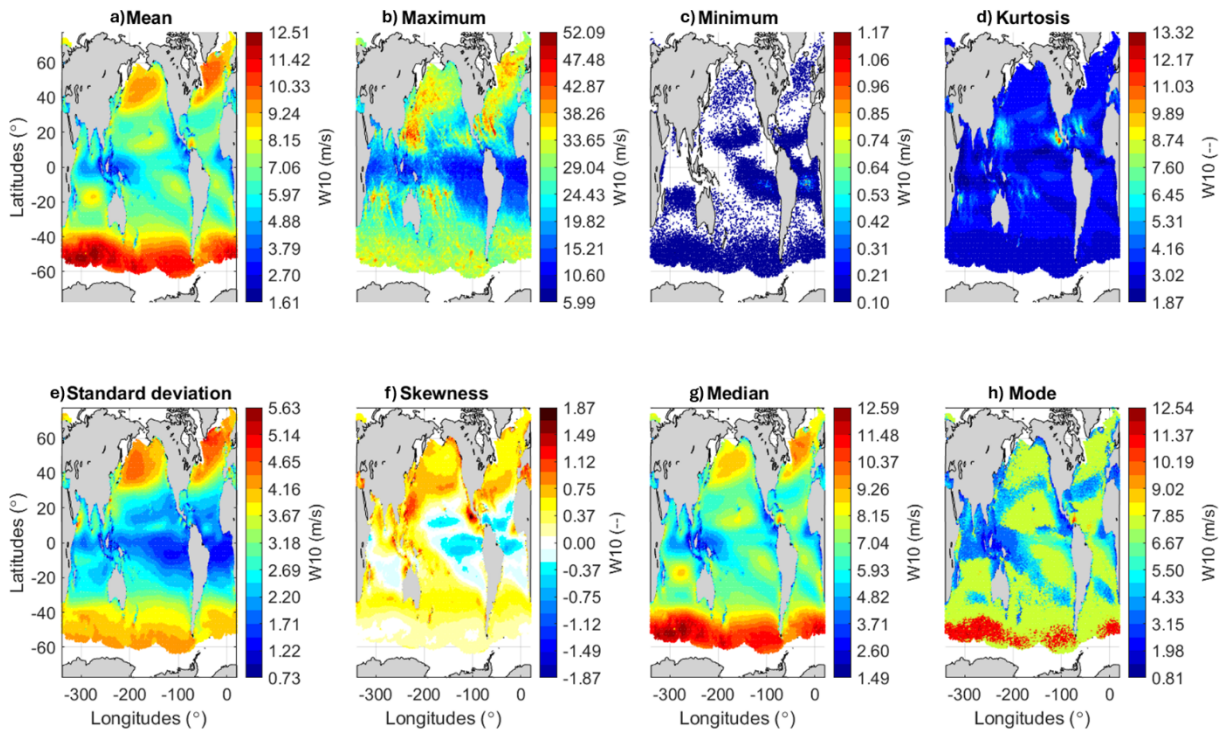


Figure A.55. Statistics over the wind amplitude (W) global histograms over the 30-year period from 1981 to 2010. Zeros are provided in white colour.

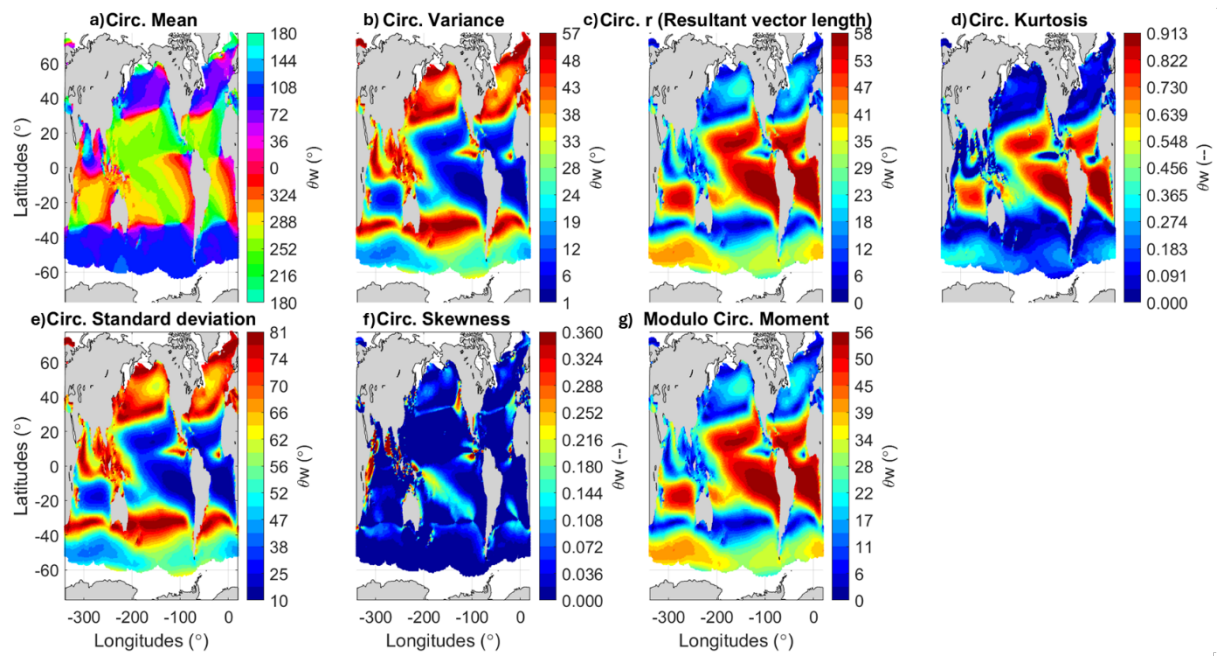


Figure A.66. Circular (Circ.) statistics over the wind direction (θ_w) global histograms over the 30-year period from 1981 to 2010.

Appendix B : Complementary statistics of wave parameters' histograms

Figure B.11 and Figure B.22 provide complementary figures to Section 4.2. Figure B.11 is the occurrence maps of the 2D key bins of Figure 33. Figure B.22 provides the mean WDM (and thereby SM) of the 3D approach.

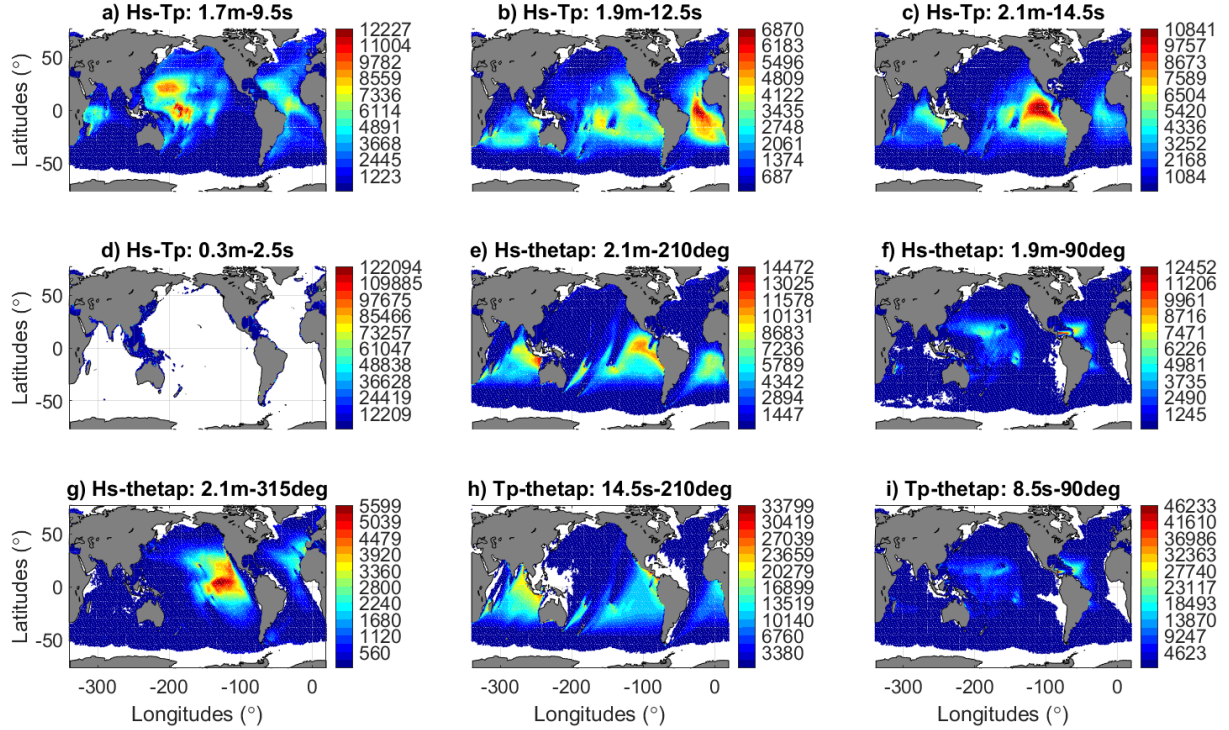


Figure B.11. Mapping of the 30-year (1981-2010) occurrence (in hours) of 30-year average sea-state distributions' peaks and key occurrence bins for 2-dimensional distributions of H_s , the significant wave height, T_p , the wave peak period, and θ_p , the wave peak direction.

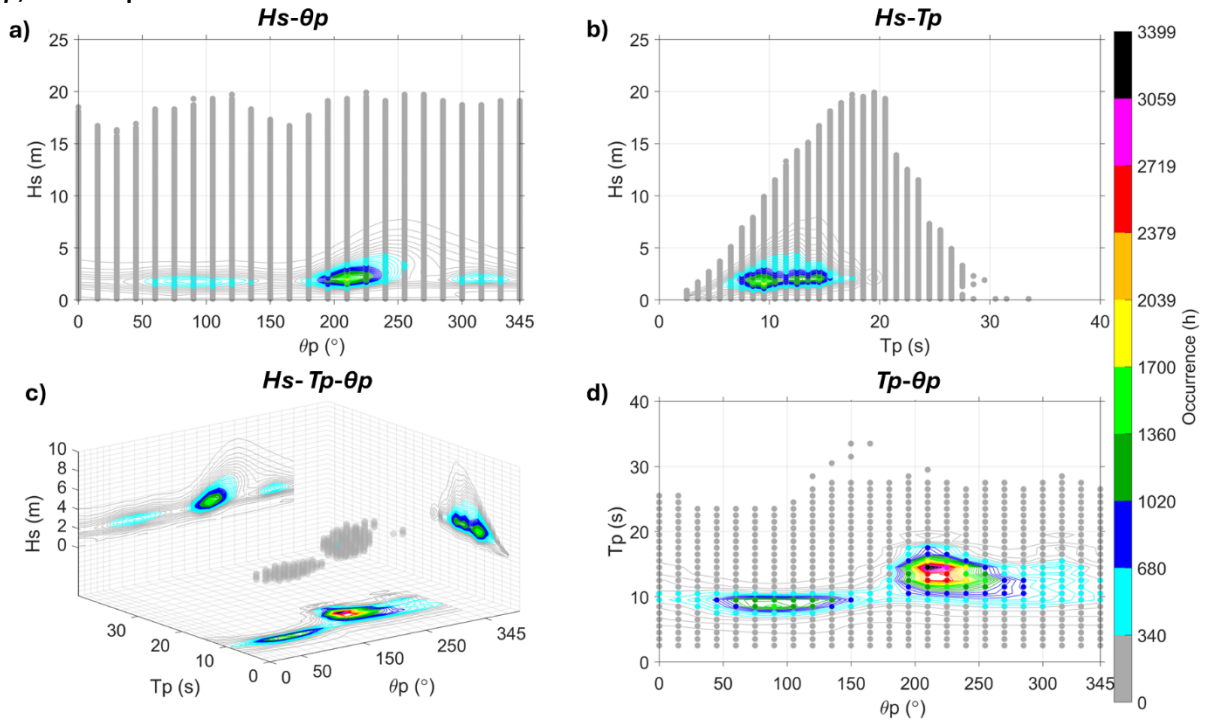


Figure B.22. Sea state full 3-dimensional (3D) mean 30-year global distribution. H_s is the wave significant height, T_p , the wave peak period, and θ_p the wave peak direction. Walls are the accumulated projection of the 3D distribution in the considered axis. Dots are the actual points of the data, and the contourlines have been added to keep track of the 2-

Dimensional wall projections from the 3D distribution. In the 3D panel, *Hs-Tp- θp* , all points below 20% of the top boundary of the minimum range have been removed.

Appendix C : Crossing the cluster indexes

As mentioned in Section 3.4, there is no reason that the k-means indexing match between the approaches (for instance, all clusters along the equator to be referred to under index 1, and those around the poles as index 2). In this section, each approach (1-3D) is considered as a period based on the fact that this algorithm was developed to cross-classify various periods (e.g., each year within a 10 or 30-year period). Each obtained cluster must, therefore, be matched (it will take the index of the cluster that has the most grid-points in common with the maximum representation of all clusters) so the clusters have consistent indexing among all periods. This is done following Figure C.11Figure C.1's k-means index consistency method. For example, using the clustering of *Hs* as a Reference Clustered Vector (RCV), if the second Column Vector (CV) has a very similar spatial clustering but different indexing, this algorithm will change CV's initial indexing to match RCV. Thus, if RCV's 1st cluster matches CV's 3rd cluster, then the final Crossed Period (CP) 2nd column vector will have the points under CV's idx=3 being set to idx=1, and so on. Eventually, the RCP will be replaced by CP to process the next steps of the crossing-period method.

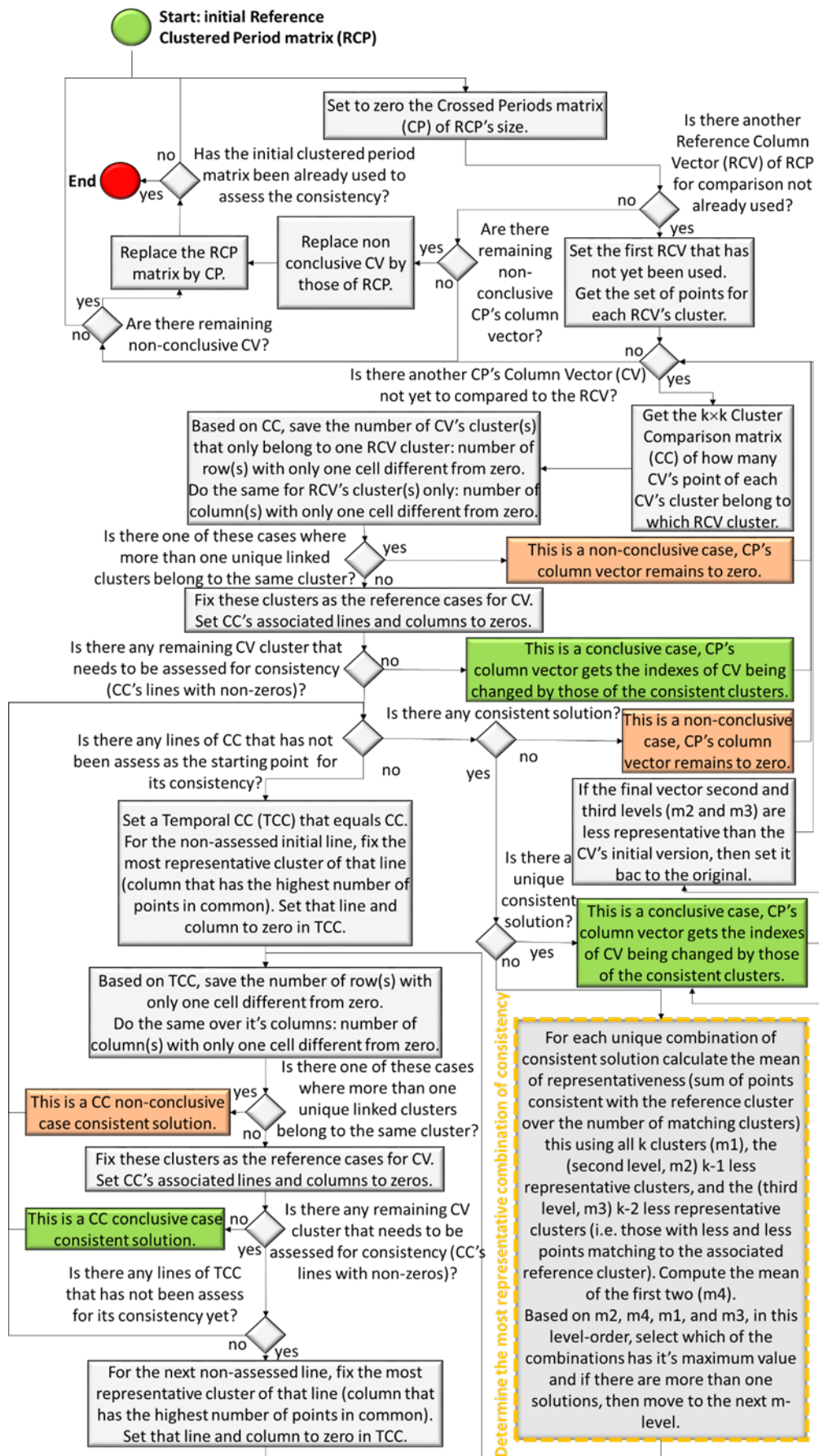


Figure C.11. Flow chart of the k-means index consistency algorithm.

Appendix D : Investigating the decomposition of H_s in the first three modes

Figure D.11 provides an assessment of the distribution and the EOF decomposition of two points of the world. The first point (top row) is located at the minimum value of the first PC of the global 30-year mean H_s , while the second point (bottom row) is associated with the maximum value obtained for that same PC. The left column provides the maps of the points' location and colouring of the first PC. The second column provides the initialisation of the data (IM), i.e. the initial distribution of the points, its centralised version (IM-SM), and SM. The third column provides the multiplication of the first three modes: each EOF by its respective PC value at the considered points. The fourth column provides the sum of the first two and first three modes as an example of how adding modes leads little by little to the initial IM-SM distribution.

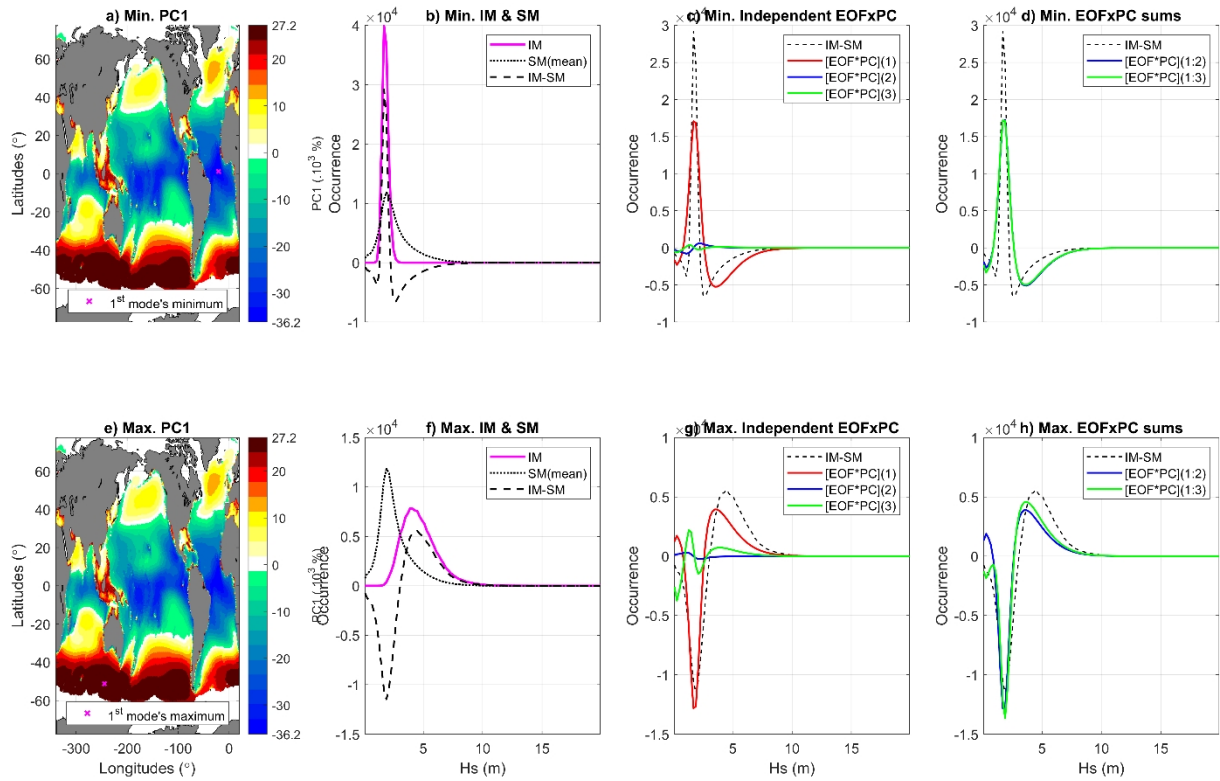


Figure D.11. Reconstruction of the Initial Distribution (IM) minus the global Space Mass (SM) of locations of maximum and minimum values of the first mode of the significant wave height (H_s) Empirical Orthogonal Function (EOF) decomposition, using the three first modes of EOFs and Principal Components (PCs).

Appendix E : Optimal k selection

As mentioned in Section 3.4, for the classifications of the PCs, both the number of modes to use and the number of clusters to consider are necessary to determine. The second main constraint of the k-means clustering is the need to pre-determine the number of clusters. The Elbow method [77] is the common method in wave-related analyses to determine the number of clusters in a distribution; for instance, Fairley et al. [21] used it to classify different sets of wave statistical parameters globally independently, or alongside the Silhouette criterion [78] to classify El Niño coastal hazard worldwide [79] or to group global wind and wave patterns [80]. The inflexion point (at the elbow of the curve) shall provide the optimal number of clusters as in Odériz et al. [22]. However, as the Elbow method has been shown not to be necessarily accurate compared to other clustering methods [81], alternative methods, Caliński and Harabasz [82], Davies-Bouldin [83], and Gap [84] have been considered. A sensitivity test over the number of modes considered was conducted for 1 to 50 cluster possibilities. Figure E.11 provides the best number of cluster(s) obtained from each k-test approach for clustering using from one to many PCs considered. Curves are generally converging in a range of values after considering 10 modes, leading to above 90% representativeness for both H_s and T_p and about 90% for ∂p , about 80% for 2D, and about 70% for 3D. For its elbow shape and as the maximum value of the Gap method generally provided outlier values, the tangent method used in the Elbow method to obtain the elbow-curve optimal point has also been alternatively implemented here and referred to as the "Gap using Elbow" curve in Figure E.11. Furthermore, Figure E.2 and Figure E.3 show the improvement of the correlation and RMSD as the number of clusters increases from 2 to 4. All six tests provided two ranges of optimal cluster values, the first between 2 and 5 clusters and the second between 15 and 20 clusters. To simplify the analysis and because the first range is not descriptive enough of the global sub-processes, the second range of clusters has been selected. Specifically, the lower threshold, i.e. 15 clusters, has been retained to assess the global clustering.

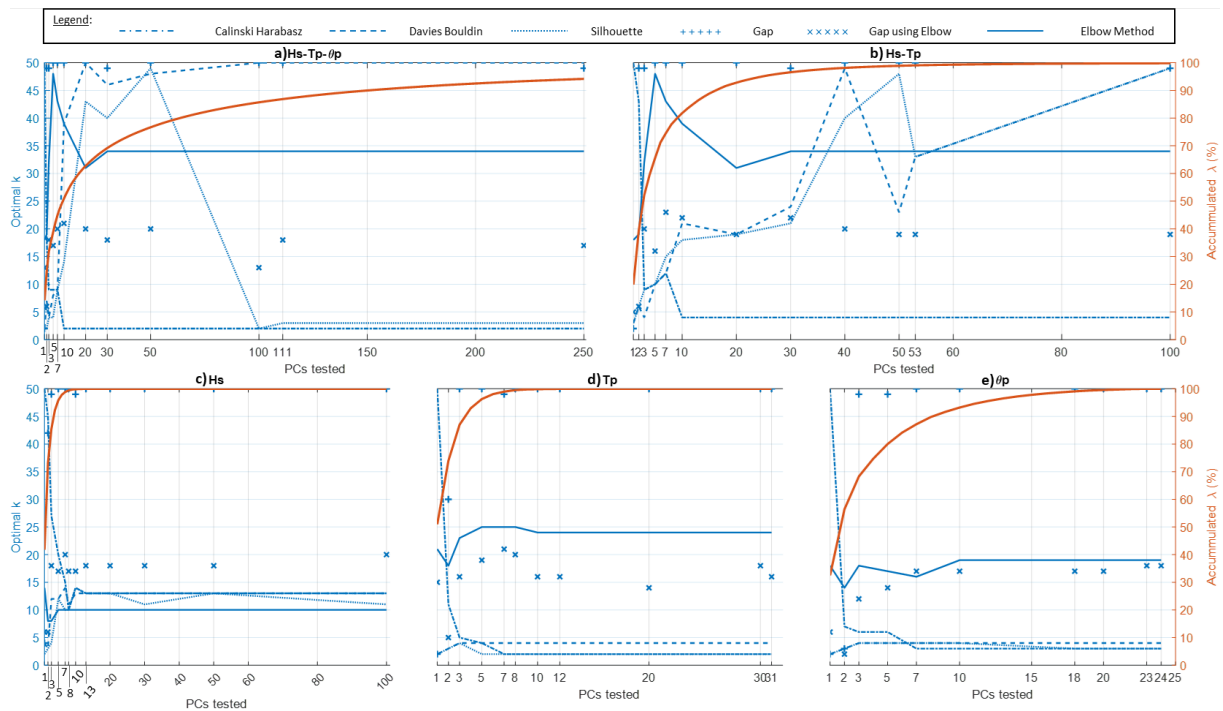


Figure E.11. Selection of the optimal k (number of clusters from the k-means algorithm) as a function of the number of Principal Components (PCs) to compute the distance of the k clusters for various k-selection methods.

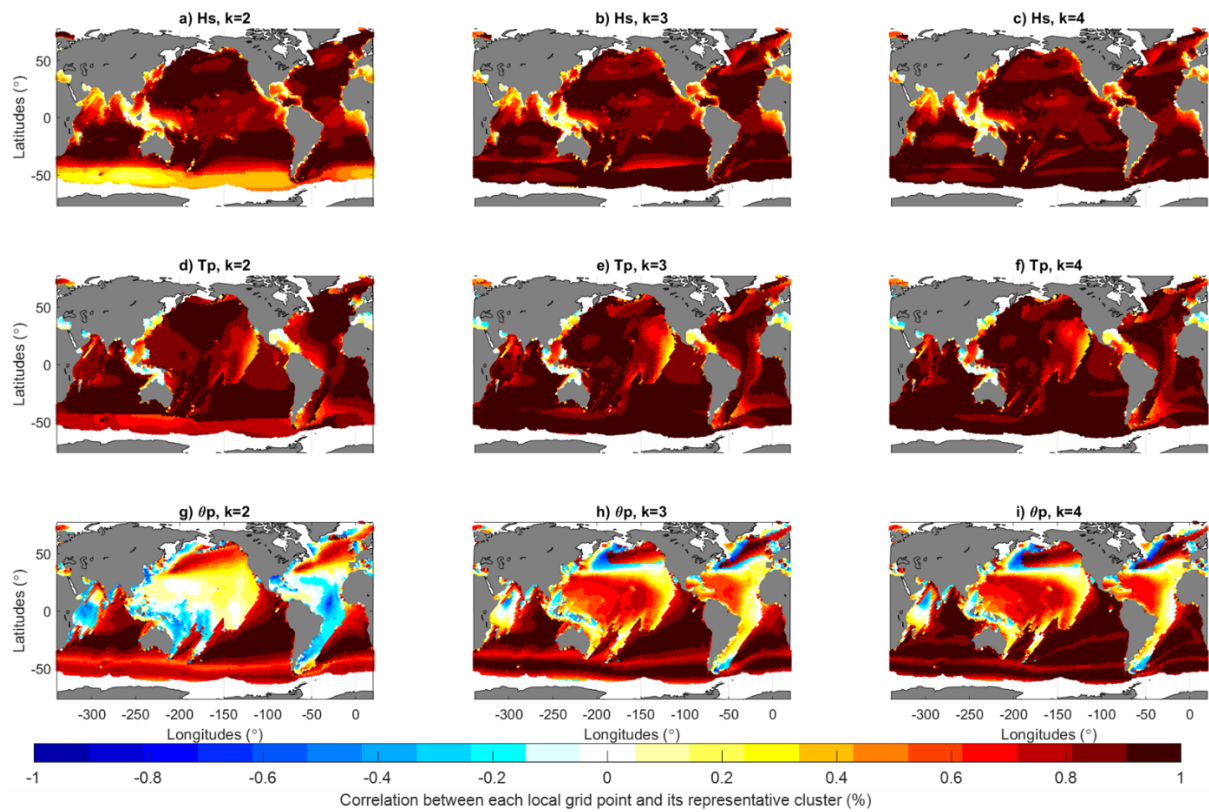


Figure E.22. Correlation (%) between the clusters' representative wave distribution matrices obtained in Figure 1111Figure 11 for the three 1-Dimensional approaches of the significant wave height (H_s), wave peak period (T_p), and wave peak direction (θ_p) and their respective grid-point's wave distribution matrices.

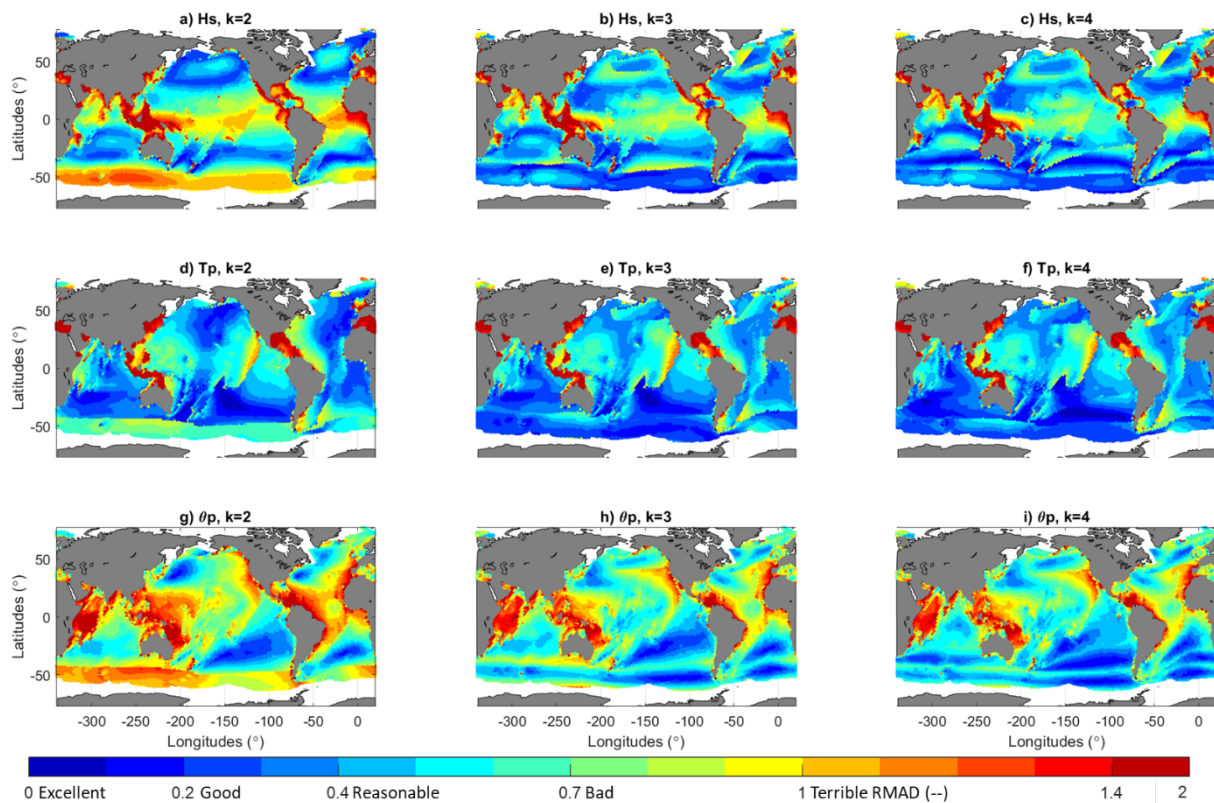


Figure E.23. Relative Mean Absolute Difference (RMAD, unitless) between the clusters' representative wave distribution matrices obtained in Figure 1111Figure 11 for the three 1-Dimensional approaches of the significant wave height (H_s), wave peak period (T_p), and wave peak direction (θ_p) and their respective grid-point's wave distribution matrices.

Appendix F : Considerations on the Directional Convention in Mode-based Classification

This section evaluates whether the circular nature of ϑp affects PCA and clustering results, especially near the $0^\circ/360^\circ$ boundary. To explore this, ϑp is redefined using a counterclockwise convention with the north at 90° and east at 0° , instead of the standard clockwise system with north at 0° . This test allows for the assessment of directional symmetry and any potential bias introduced at cardinal boundaries. Figure F.1 displays the first three Empirical Orthogonal Functions (EOFs) and their associated Principal Component (PC) maps under this modified convention in the same fashion as Figure 9 for direct comparison with the original ϑp analysis. Figure F.2 provides the clustering results for $k=2,4$ (corresponding to Figure 11) and $k=15$ (corresponding to Figure 12). The comparison reveals that, although the EOF structures and low- k clustering results remain generally consistent, noticeable differences appear at higher k values. In particular, one cluster, cluster 1, located above cluster 14 in Figure 12, largely disappears under the new convention, suggesting that the standard ϑp representation preserves more spatial distinction in the directional field. These results confirm that while PCA remains stable for ϑp under both conventions, the original clockwise-from-north system offers slightly more detailed discrimination in classification outcomes. A more comprehensive assessment with even higher cluster counts or region-specific tests may help refine this insight in future studies.

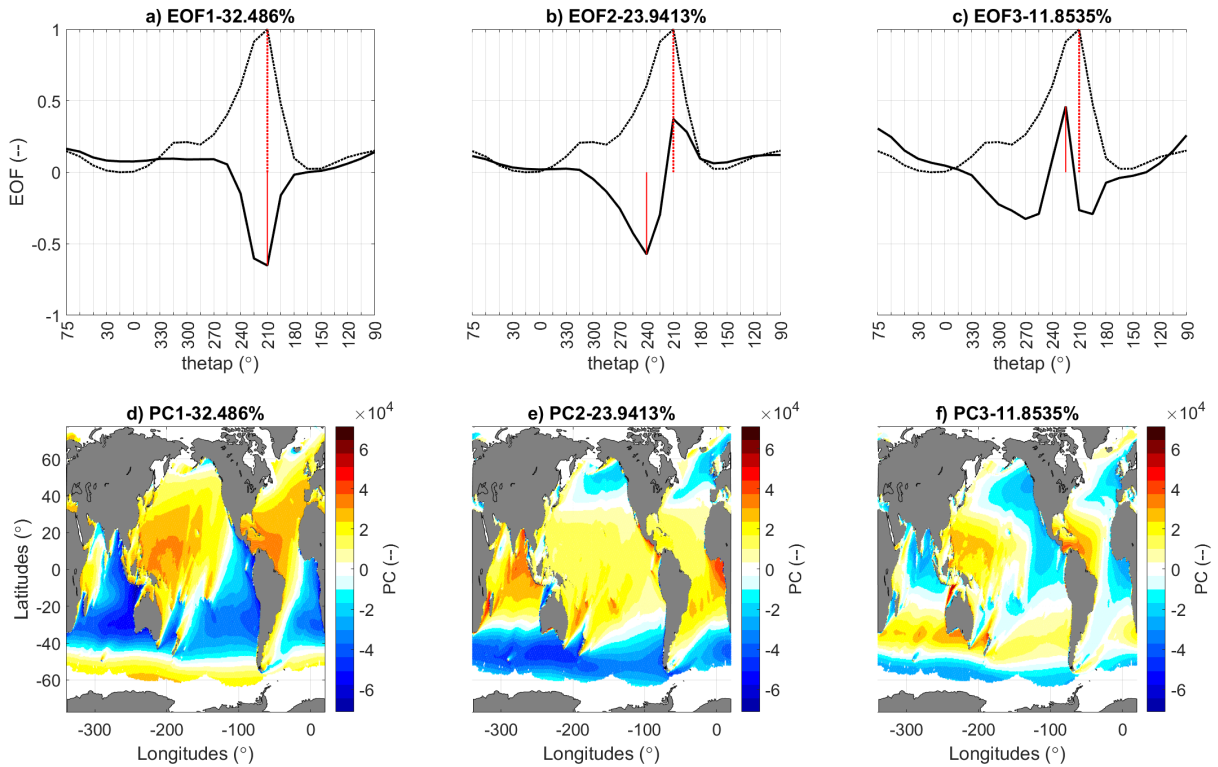


Figure F.11. First three Empirical Orthogonal Functions (EOFs, top-panels solid lines) and Principal Components (PCs) maps (bottom-panels) of the peak wave direction (ϑp) using the East counterclockwise convention for the entire integrated 30-year period. The dashed lines of the EOF figures show the normalised global mean ϑp distribution (ϑp normalised space mass vector, see Section 3.3). The red lines underline the maxima to help compare the EOFs, solid for the EOF and dashed for the mean. The title indicates whether the subfigure is an EOF or PC, the mode concerned, and the associated percentage of the total variance values.

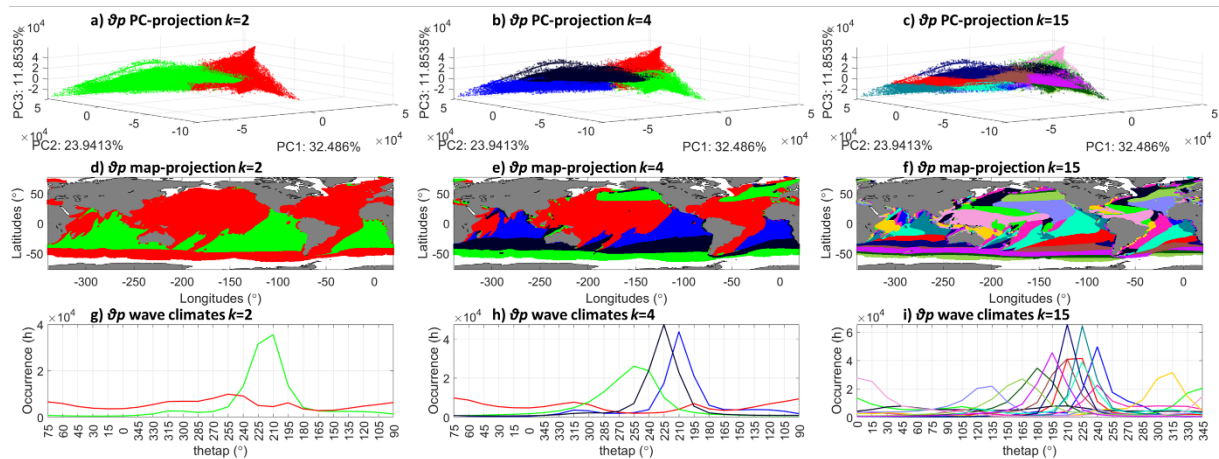


Figure F.22. 30-year wave climate classification using 2, 4, and 15 clusters and 10 modes for the wave peak direction (θp) using the East counterclockwise convention. Panels a–c are the clustering of the highest three Principal Components (PC1–3) space. Panels d–f project the obtained cluster of the right-above PC-space points on the globe. Panels g–i provide the Wave Distribution Matrices of each right-above cluster.

REFERENCES

- [1] F. Aristodemo and D. Algieri Ferraro, 'Feasibility of WEC installations for domestic and public electrical supplies: A case study off the Calabrian coast', *Renewable Energy*, vol. 121, pp. 261–285, June 2018, doi: 10.1016/j.renene.2018.01.012.
- [2] D. Clemente, P. Rosa-Santos, and F. Taveira-Pinto, 'On the potential synergies and applications of wave energy converters: A review', *Renewable and Sustainable Energy Reviews*, vol. 135, p. 110162, Jan. 2021, doi: 10.1016/j.rser.2020.110162.
- [3] O. Choupin, F. Pinheiro Andutta, A. Etemad-Shahidi, and R. Tomlinson, 'A decision-making process for wave energy converter and location pairing', *Renewable and Sustainable Energy Reviews*, vol. 147, p. 111225, May 2021, doi: 10.1016/j.rser.2021.111225.
- [4] F. Ardhuin, *Ocean waves in geosciences*. 2024. doi: 10.13140/RG.2.2.16019.78888/11.
- [5] T. Durrant, M. Hemer, C. Trenham, and D. Greenslade, 'CAWCR Wave Hindcast 1979–2010'. CSIRO, 2013. doi: 10.4225/08/52817E2858340.
- [6] J. Harari, *Noções de oceanografia*. Portal de Livros Abertos da USP, 2023. doi: 10.11606/9786599585401.
- [7] G. Dodet, X. Bertin, and R. Taborda, 'Wave climate variability in the North-East Atlantic Ocean over the last six decades', *Ocean Modelling*, vol. 31, no. 3, pp. 120–131, Jan. 2010, doi: 10.1016/j.ocemod.2009.10.010.
- [8] S. K. Gulev and V. Grigorieva, 'Variability of the Winter Wind Waves and Swell in the North Atlantic and North Pacific as Revealed by the Voluntary Observing Ship Data', *Journal of Climate*, vol. 19, no. 21, pp. 5667–5685, Nov. 2006, doi: 10.1175/JCLI3936.1.
- [9] G. Chen, B. Chapron, R. Ezraty, and D. Vandemark, 'A Global View of Swell and Wind Sea Climate in the Ocean by Satellite Altimeter and Scatterometer', *J. Atmos. Oceanic Technol.*, vol. 19, no. 11, pp. 1849–1859, Nov. 2002, doi: 10.1175/1520-0426(2002)019%3C1849:AGVOSA%3E2.0.CO;2.
- [10] M. A. Hemer, J. A. Church, and J. R. Hunter, 'Variability and trends in the directional wave climate of the Southern Hemisphere', *International Journal of Climatology*, vol. 30, no. 4, pp. 475–491, 2010, doi: 10.1002/joc.1900.
- [11] A. F. Haselsteiner and K.-D. Thoben, 'Predicting wave heights for marine design by prioritizing extreme events in a global model', *Renewable Energy*, vol. 156, pp. 1146–1157, Aug. 2020, doi: 10.1016/j.renene.2020.04.112.
- [12] L. Rusu and F. Onea, 'The performance of some state-of-the-art wave energy converters in locations with the worldwide highest wave power', *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 1348–1362, Aug. 2017, doi: 10.1016/j.rser.2016.11.123.
- [13] L. V. Colosi, A. B. Villas Bôas, and S. T. Gille, 'The Seasonal Cycle of Significant Wave Height in the Ocean: Local Versus Remote Forcing', *Journal of Geophysical Research: Oceans*, vol. 126, no. 8, p. e2021JC017198, 2021, doi: 10.1029/2021JC017198.
- [14] O. Choupin, A. Têtu, B. Del Río-Gamero, F. Ferri, and JP. Kofoed, 'Premises for an annual energy production and capacity factor improvement towards a few optimised wave energy converters configurations and resources pairs', *Applied Energy*, vol. 312, p. 118716, Apr. 2022, doi: 10.1016/j.apenergy.2022.118716.

- [15] A. C. Eckert-Gallup, C. J. Sallaberry, A. R. Dallman, and V. S. Neary, 'Application of principal component analysis (PCA) and improved joint probability distributions to the inverse first-order reliability method (I-FORM) for predicting extreme sea states', *Ocean Engineering*, vol. 112, pp. 307–319, Jan. 2016, doi: 10.1016/j.oceaneng.2015.12.018.
- [16] S. Lloyd, 'Least squares quantization in PCM', *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982, doi: 10.1109/TIT.1982.1056489.
- [17] E. R. Echevarria, M. A. Hemer, and N. J. Holbrook, 'Seasonal Variability of the Global Spectral Wind Wave Climate', *Journal of Geophysical Research: Oceans*, vol. 124, no. 4, pp. 2924–2939, 2019, doi: 10.1029/2018JC014620.
- [18] H. Lobeto, M. Menendez, I. J. Losada, and M. Hemer, 'The effect of climate change on wind-wave directional spectra', *Global and Planetary Change*, vol. 213, p. 103820, June 2022, doi: 10.1016/j.gloplacha.2022.103820.
- [19] R. Satymov, D. Bogdanov, M. Dadashi, G. Lavidas, and C. Breyer, 'Techno-economic assessment of global and regional wave energy resource potentials and profiles in hourly resolution', *Applied Energy*, vol. 364, p. 123119, June 2024, doi: 10.1016/j.apenergy.2024.123119.
- [20] International Electrotechnical Commission, 'Marine energy - Wave, tidal and other water current converters - Part 100: Electricity producing wave energy converters - Power performance assessment; IEC TS 62600-100 Ed. 1', International Electrotechnical Commission: Geneva, Switzerland, Aug. 2012. [Online]. Available: <https://www.en-standard.eu/iec-ts-62600-100-2012-marine-energy-wave-tidal-and-other-water-current-converters-part-100-electricity-producing-wave-energy-converters-power-performance-assessment/>
- [21] I. Fairley *et al.*, 'A classification system for global wave energy resources based on multivariate clustering', *Applied Energy*, vol. 262, p. 114515, Mar. 2020, doi: 10.1016/j.apenergy.2020.114515.
- [22] I. Odériz *et al.*, 'Natural Variability and Warming Signals in Global Ocean Wave Climates', *Geophysical Research Letters*, vol. 48, no. 11, p. e2021GL093622, 2021, doi: 10.1029/2021GL093622.
- [23] Young I.R., 'Seasonal variability of the global ocean wind and wave climate', *International Journal of Climatology*, vol. 19, no. 9, pp. 931–950, June 1999, doi: 10.1002/(SICI)1097-0088(199907)19:9<931::AID-JOC412>3.0.CO;2-O.
- [24] J.-H. G. M. Alves, 'Numerical modeling of ocean swell contributions to the global wind-wave climate', *Ocean Modelling*, vol. 11, no. 1, pp. 98–122, Jan. 2006, doi: 10.1016/j.ocemod.2004.11.007.
- [25] J. D. Campana, 'Multi-Scale Summary of Global Wave Statistics', Naval Surface Warfare Center, Carderock Division, Naval Architecture and Engineering Department, West Bethesda, MD, Technical Report, May 2023. [Online]. Available: <https://apps.dtic.mil/sti/trecms/pdf/AD1140103.pdf>
- [26] H. L. Tolman, 'User manual and system documentation of WAVEWATCH III TM version 3.14', *Technical note, MMAB Contribution*, vol. 276, no. 220, 2009.
- [27] S. Saha *et al.*, 'The NCEP climate forecast system reanalysis', *Bulletin of the American Meteorological Society*, vol. 91, no. 8, pp. 1015–1058, 2010.
- [28] H. L. Tolman, 'Treatment of unresolved islands and ice in wind wave models', *Ocean Modelling*, vol. 5, no. 3, pp. 219–231, Jan. 2003, doi: 10.1016/S1463-5003(02)00040-9.
- [29] International Electrotechnical Commission (IEC), 'Marine energy - Wave, tidal and other water current converters - Part 101: Wave energy resource assessment and characterization; IEC TS 62600-101:2015', International Electrotechnical Commission: Geneva, Switzerland, Technical Specification ISBN 978-2-8322-2724-4, June 2015. Accessed: July 27, 2021. [Online]. Available: <https://webstore.iec.ch/publication/22593&preview=1>
- [30] N. Rasclé and F. Ardhuin, 'A global wave parameter database for geophysical applications. Part 2: Model validation with improved source term parameterization', *Ocean Modelling*, vol. 70, pp. 174–188, Oct. 2013, doi: 10.1016/j.ocemod.2012.12.001.
- [31] V. Venugopal, R. Nimalidinne, and A. Vögler, 'Numerical modelling of wave energy resources and assessment of wave energy extraction by large scale wave farms', *Ocean & Coastal Management*, vol. 147, pp. 37–48, Oct. 2017, doi: 10.1016/j.ocecoaman.2017.03.012.
- [32] M. Gonçalves, P. Martinho, and C. Guedes Soares, 'A 33-year hindcast on wave energy assessment in the western French coast', *Energy*, vol. 165, pp. 790–801, Dec. 2018, doi: 10.1016/j.energy.2018.10.002.
- [33] O. Choupin, A. Têtu, and F. Ferri, 'Wave energy converter power and capture width classification', *Ocean Engineering*, vol. 260, p. 111749, Sept. 2022, doi: 10.1016/j.oceaneng.2022.111749.
- [34] R. Guanche, A. de Andrés, I. J. Losada, and C. Vidal, 'A global analysis of the operation and maintenance role on the placing of wave energy farms', *Energy Conversion and Management*, vol. 106, pp. 440–456, Dec. 2015, doi: 10.1016/j.enconman.2015.09.022.

- [35] A. Pecher and J. P. Kofoed, Eds, *Handbook of Ocean Wave Energy*, 1st ed. 2017 edition. Springer, 2016.
- [36] J. Prendergast, M. Li, and W. Sheng, 'A Study on the Effects of Wave Spectra on Wave Energy Conversions', *IEEE Journal of Oceanic Engineering*, pp. 1–13, 2018, doi: 10.1109/JOE.2018.2869636.
- [37] J. E. Stopa, F. Ardhuin, E. Stutzmann, and T. Lecocq, 'Sea State Trends and Variability: Consistency Between Models, Altimeters, Buoys, and Seismic Data (1979–2016)', *Journal of Geophysical Research: Oceans*, vol. 124, no. 6, pp. 3923–3940, 2019, doi: 10.1029/2018JC014607.
- [38] J. Lavelle and J. P. Kofoed, 'Representative spectra of the wave resource from real sea wave measurements', in *10th European Wave and Tidal Energy Conference, 2013*, Technical Committee of the European Wave and Tidal Energy Conference, 2013.
- [39] A. Azzellino, V. Ferrante, J. P. Kofoed, C. Lanfredi, and D. Vicinanza, 'Optimal siting of offshore wind-power combined with wave energy through a marine spatial planning approach', *International Journal of Marine Energy*, vol. 3–4, pp. e11–e25, Dec. 2013, doi: 10.1016/j.ijome.2013.11.008.
- [40] M. Mayo, 'An Arithmetic-Based Deterministic Centroid Initialization Method for the k-Means Clustering Algorithm', MSc Thesis, TSYS School of Computer Science, 2016. [Online]. Available: https://csuepress.columbusstate.edu/theses_dissertations/241
- [41] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, 'Present and future Köppen-Geiger climate classification maps at 1-km resolution', *Sci Data*, vol. 5, p. 180214, Oct. 2018, doi: 10.1038/sdata.2018.214.
- [42] D. J. R. Walstra, J. A. Roelvink, and J. Groeneweg, 'Calculation of Wave-Driven Currents in a 3D Mean Flow Model', pp. 1050–1063, Apr. 2012, doi: 10.1061/40549(276)81.
- [43] T. Durrant, D. Greenslade, M. Hemer, and C. Trenham, *A global wave hindcast focussed on the Central and South Pacific*, vol. 40. 2014.
- [44] F. Ardhuin, J. Tournadre, P. Queffelec, F. Girard-Ardhuin, and F. Collard, 'Observation and parameterization of small icebergs: Drifting breakwaters in the southern ocean', *Ocean Modelling*, vol. 39, no. 3, pp. 405–410, Jan. 2011, doi: 10.1016/j.ocemod.2011.03.004.
- [45] F. Ardhuin *et al.*, 'Semiempirical Dissipation Source Functions for Ocean Waves. Part I: Definition, Calibration, and Validation', Sept. 2010, doi: 10.1175/2010JPO4324.1.
- [46] F. Ardhuin and A. Roland, 'Coastal wave reflection, directional spread, and seismoacoustic noise sources', *Journal of Geophysical Research: Oceans*, vol. 117, no. C11, 2012, doi: 10.1029/2011JC007832.
- [47] M. D. Earle, K. E. Steele, and D. W. C. Wang, 'Use of advanced directional wave spectra analysis methods', *Ocean Engineering*, vol. 26, no. 12, pp. 1421–1434, Dec. 1999, doi: 10.1016/S0029-8018(99)00010-4.
- [48] C. W. Saidelles, M. S. Ramos, L. Farina, J. L. Nicolodi, J. Albuquerque, and F. J. Méndez, 'Directional correction of ERA5 and CAWCR wave spectrum partitions in the Southeast Brazil', *Ocean Engineering*, vol. 265, p. 112475, Dec. 2022, doi: 10.1016/j.oceaneng.2022.112475.
- [49] B. Oliver, J. Veitch, and C. J. C. Reason, 'Variability in High Wave Energy Events Around the Southern African Coast', *Journal of Geophysical Research: Oceans*, vol. 127, no. 3, p. e2021JC018255, 2022, doi: 10.1029/2021JC018255.
- [50] S. S. Khan, E. R. Echevarria, and M. A. Hemer, 'Ocean Swell Comparisons Between Sentinel-1 and WAVEWATCH III Around Australia', *Journal of Geophysical Research: Oceans*, vol. 126, no. 2, p. e2020JC016265, 2021, doi: 10.1029/2020JC016265.
- [51] M. A. Hemer *et al.*, 'A revised assessment of Australia's national wave energy resource', *Renewable Energy*, vol. 114, pp. 85–107, Dec. 2017, doi: 10.1016/j.renene.2016.08.039.
- [52] N. Kumar, D. L. Cahl, S. C. Crosby, and G. Voulgaris, 'Bulk versus Spectral Wave Parameters: Implications on Stokes Drift Estimates, Regional Wave Modeling, and HF Radars Applications', June 2017, doi: 10.1175/JPO-D-16-0203.1.
- [53] B. Kinsman, *Wind Waves: Their Generation and Propagation on the Ocean Surface*, First Edition. New York: Dover Publications, 1965.
- [54] A. Semedo, K. Sušelj, A. Rutgersson, and A. Sterl, 'A Global View on the Wind Sea and Swell Climate and Variability from ERA-40', *J. Climate*, vol. 24, no. 5, pp. 1461–1479, Mar. 2011, doi: 10.1175/2010JCLI3718.1.
- [55] F. Collard, F. Ardhuin, and B. Chapron, 'Monitoring and analysis of ocean swell fields from space: New methods for routine observations', *Journal of Geophysical Research: Oceans*, vol. 114, no. C7, 2009, doi: 10.1029/2008JC005215.
- [56] D. B. Chelton, M. G. Schlax, M. H. Freilich, and R. F. Milliff, 'Satellite Measurements Reveal Persistent Small-Scale Features in Ocean Winds', *Science*, vol. 303, no. 5660, pp. 978–983, Feb. 2004, doi: 10.1126/science.1091901.
- [57] L. Wu, E. Sahlée, E. Nilsson, and A. Rutgersson, 'A review of surface swell waves and their role in air–sea interactions', *Ocean Modelling*, vol. 190, p. 102397, Aug. 2024, doi: 10.1016/j.ocemod.2024.102397.

- [58] L. G. Bennetts *et al.*, 'Closing the Loops on Southern Ocean Dynamics: From the Circumpolar Current to Ice Shelves and From Bottom Mixing to Surface Waves', *Reviews of Geophysics*, vol. 62, no. 3, p. e2022RG000781, 2024, doi: 10.1029/2022RG000781.
- [59] J. Lodise, S. Merrifield, C. Collins, P. Rogowski, J. Behrens, and E. Terrill, 'Global Climatology of Extratropical Cyclones From a New Tracking Approach and Associated Wave Heights From Satellite Radar Altimeter', *Journal of Geophysical Research: Oceans*, vol. 127, no. 11, p. e2022JC018925, 2022, doi: 10.1029/2022JC018925.
- [60] F. Song, L. R. Leung, J. Lu, and L. Dong, 'Future Changes in Seasonality of the North Pacific and North Atlantic Subtropical Highs', *Geophysical Research Letters*, vol. 45, no. 21, p. 11,959-11,968, 2018, doi: 10.1029/2018GL079940.
- [61] G. Berry and M. J. Reeder, 'Objective Identification of the Intertropical Convergence Zone: Climatology and Trends from the ERA-Interim', Mar. 2014, doi: 10.1175/JCLI-D-13-00339.1.
- [62] P. Lionello, R. D'Agostino, D. Ferreira, H. Nguyen, and M. S. Singh, 'The Hadley circulation in a changing climate', *Annals of the New York Academy of Sciences*, vol. 1534, no. 1, pp. 69–93, 2024, doi: 10.1111/nyas.15114.
- [63] M. C. Kruk, K. R. Knapp, and D. H. Levinson, 'A Technique for Combining Global Tropical Cyclone Best Track Data', *Journal of Atmospheric and Oceanic Technology*, vol. 27, no. 4, pp. 680–692, Apr. 2010, doi: 10.1175/2009JTECHA1267.1.
- [64] M. Folley and T. J. T. Whittaker, 'Analysis of the nearshore wave energy resource', *Renewable Energy*, vol. 34, no. 7, pp. 1709–1715, July 2009, doi: 10.1016/j.renene.2009.01.003.
- [65] J. A. Battjes and J. P. F. M. Janssen, 'Energy Loss and Set-Up Due to Breaking of Random Waves', *Coastal engineering*, pp. 569–587, 1978, doi: 10.1061/9780872621909.034.
- [66] B. Del Río-Gamero, O. Choupin, N. Melián-Martel, and J. Schallenberg-Rodriguez, 'Application of a revised integration of methods for wave energy converter and farm location pair mapping', *Energy Conversion and Management*, vol. 303, p. 118170, Mar. 2024, doi: 10.1016/j.enconman.2024.118170.
- [67] K. R. Wodzicki and A. D. Rapp, 'Long-term characterization of the Pacific ITCZ using TRMM, GPCP, and ERA-Interim', *Journal of Geophysical Research: Atmospheres*, vol. 121, no. 7, pp. 3153–3170, 2016, doi: 10.1002/2015JD024458.
- [68] R. A. Arinaga and K. F. Cheung, 'Atlas of global wave energy from 10 years of reanalysis and hindcast data', *Renewable Energy*, vol. 39, no. 1, pp. 49–64, Mar. 2012, doi: 10.1016/j.renene.2011.06.039.
- [69] I. Simmonds and K. Keay, 'Variability of Southern Hemisphere Extratropical Cyclone Behavior, 1958–97', Feb. 2000, Accessed: Oct. 17, 2025. [Online]. Available: https://journals.ametsoc.org/view/journals/clim/13/3/1520-0442_2000_013_0550_voshec_2.0.co_2.xml
- [70] A. Amores and M. Marcos, 'Ocean Swells along the Global Coastlines and Their Climate Projections for the Twenty-First Century', Jan. 2020, doi: 10.1175/JCLI-D-19-0216.1.
- [71] H. Lashkari, Z. Mohammadi, and G. Keikhosravi, 'Annual Fluctuations and Displacements of Inter Tropical Convergence Zone (ITCZ) within the Range of Atlantic Ocean-India', *Open Journal of Ecology*, vol. 7, no. 1, Art. no. 1, Jan. 2017, doi: 10.4236/oje.2017.71002.
- [72] M. Belmonte Rivas and A. Stoffelen, 'Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT', *Ocean Science*, vol. 15, no. 3, pp. 831–852, June 2019, doi: 10.5194/os-15-831-2019.
- [73] C. Zheng *et al.*, 'Global trends in oceanic wind speed, wind-sea, swell, and mixed wave heights', *Applied Energy*, vol. 321, p. 119327, Sept. 2022, doi: 10.1016/j.apenergy.2022.119327.
- [74] B. Kamranzad, K. Amarouche, and A. Akpinar, 'Linking the long-term variability in global wave energy to swell climate and redefining suitable coasts for energy exploitation', *Sci Rep*, vol. 12, no. 1, Art. no. 1, Aug. 2022, doi: 10.1038/s41598-022-18935-w.
- [75] B. G. Reguero, I. J. Losada, and F. J. Méndez, 'A global wave power resource and its seasonal, interannual and long-term variability', *Applied Energy*, vol. 148, pp. 366–380, June 2015, doi: 10.1016/j.apenergy.2015.03.114.
- [76] P. Berens, 'CircStat: A MATLAB Toolbox for Circular Statistics', *Journal of Statistical Software*, vol. 31, pp. 1–21, Sept. 2009, doi: 10.18637/jss.v031.i10.
- [77] D. J. Ketchen and C. L. Shook, 'The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique', *Strategic Management Journal*, vol. 17, no. 6, pp. 441–458, 1996, doi: 10.1002/(SICI)1097-0266(199606)17:6<3C441::AID-SMJ819%3E3.0.CO;2-G.
- [78] P. J. Rousseeuw, 'Silhouettes: A graphical aid to the interpretation and validation of cluster analysis', *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.

- [79] I. Odériz, R. Silva, T. R. Mortlock, and N. Mori, 'El Niño-Southern Oscillation Impacts on Global Wave Climate and Potential Coastal Hazards', *Journal of Geophysical Research: Oceans*, vol. 125, no. 12, p. e2020JC016464, 2020, doi: 10.1029/2020JC016464.
- [80] B. Robertson, G. Dunkle, J. Gadas, G. Garcia-Medina, and Z. Yang, 'Holistic marine energy resource assessments: A wave and offshore wind perspective of metocean conditions', *Renewable Energy*, vol. 170, pp. 286–301, June 2021, doi: 10.1016/j.renene.2021.01.136.
- [81] E. Schubert, 'Stop using the elbow criterion for k-means and how to choose the number of clusters instead', *ACM SIGKDD Explorations Newsletter*, vol. 25, no. 1, pp. 36–42, June 2023, doi: 10.1145/3606274.3606278.
- [82] T. Caliński and J. Harabasz, 'A dendrite method for cluster analysis', *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, Jan. 1974, doi: 10.1080/03610927408827101.
- [83] D. L. Davies and D. W. Bouldin, 'A Cluster Separation Measure', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.
- [84] R. Tibshirani, G. Walther, and T. Hastie, 'Estimating the Number of Clusters in a Data Set Via the Gap Statistic', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 63, no. 2, pp. 411–423, July 2001, doi: 10.1111/1467-9868.00293.



DSPACE

<https://dspace.org/>

Global classification of wave height, period, and direction multivariate distributions using principal component analysis

Choupin, Ophélie; Harari, Joseph; Könözy, László Z.

2026-01

Elsevier

<https://doi.org/10.1016/j.pocean.2025.103660>

Choupin O, Harari J, Könözy L. (2026) Global classification of wave height, period, and direction multivariate distributions using principal component analysis. Progress in Oceanography, Volume 241, January 2026, Article number 103660