

Supporting Information

Wave climate and atmospheric forcing along the Brazilian Equatorial Margin

Theoretical and Applied Climatology

Matheus Bonjour Laviola da Silva; Danilo Couto de Souza;
Fernando Tulio Camilo Barreto; Leonardo Carvalho de Jesus;
Amauri Costa Bevilacqua; Marcia Carolina de Oliveira Costa;
Victor Azevedo Godoi; Carlos Leandro da Silva Junior

Corresponding author: Matheus Bonjour Laviola da Silva
OceanPact, Rua da Glória 122, Rio de Janeiro, RJ 20241-180, Brazil
`matheus.laviola@oceanpact.com`

Text S1. Selection of the number of wave patterns

Two families of criterion can be applied to the choice of K , and here they disagree. The first measures compactness: how tightly the days of a regime cluster around their centroid, and how far apart the centroids lie. The second measures reproducibility: how far the same partition is recovered when the classification is refitted on resampled data or from different initialisations. We report both, and select on the second.

Compactness declines monotonically with K over the range examined (Table S1, Figure S1a). The silhouette coefficient is highest at $K = 4$ (0.2245) and falls to 0.2192 at $K = 6$, a difference of 2.4%, before dropping more sharply from $K = 7$ onward. The minimum standardised separation between centroids behaves in the same way, holding near 0.82 up to $K = 6$ and falling 13.3% at $K = 7$. This monotonic decline is expected rather than informative: silhouette-type indices reward compact groups separated by gaps in the feature space, and sea state is a continuum in which such gaps do not exist. Applied alone to data of this kind, the criterion selects the smallest partition by construction, and would do so whatever the physical structure of the regimes.

Reproducibility behaves differently, and is not monotonic (Figure S1b). $K = 6$ maximises the median bootstrap adjusted Rand index (0.939), its 5th percentile (0.871), and the minimum adjusted Rand index across seeds (0.995). The contrast with $K = 4$ is instructive: the partition favoured by the silhouette has a 5th-percentile bootstrap ARI of 0.763, lower than that of $K = 7$ and second lowest in the range examined. The most compact partition of this dataset is therefore among the least stable under resampling, which is the behaviour expected when a genuine regime is being merged into a neighbour. Beyond $K = 6$, reproducibility and separation fall together, and the smallest regime occupies 6.7% of days at $K = 7$ and $K = 8$, too sparse to support the composite analysis of the main text.

We therefore adopt $K = 6$. The choice rests on stability, not on optimality: no index examined here has an interior optimum at $K = 6$ in the compactness family, and we do not present one. The six patterns are additionally supported by physical review of their matched centroids under the 1993–2022 refit, in which no pattern changes compass sector, the largest directional rotation is 1.71° , and the largest shift in mean H_s is 3.9 cm.

Table S1: Selection of the number of wave patterns over the 1993–2022 analysis period. For each candidate K : silhouette coefficient; minimum standardised separation between centroids; fraction of days in the smallest regime; median and 5th percentile of the adjusted Rand index (ARI) across 100 calendar-year block bootstrap replicates; and the minimum ARI across 20 random seeds. The selected partition is shown in bold.

K	Silhouette	Min. separation	Smallest regime (%)	Boot. ARI median	Boot. ARI q_{05}	Seed ARI min.
4	0.2245	0.8276	13.5	0.930	0.763	0.983
5	0.2192	0.8271	13.2	0.929	0.864	0.988
6	0.2192	0.8176	9.2	0.939	0.871	0.995
7	0.2042	0.7089	6.7	0.906	0.859	0.980
8	0.1956	0.6679	6.8	0.852	0.635	0.940

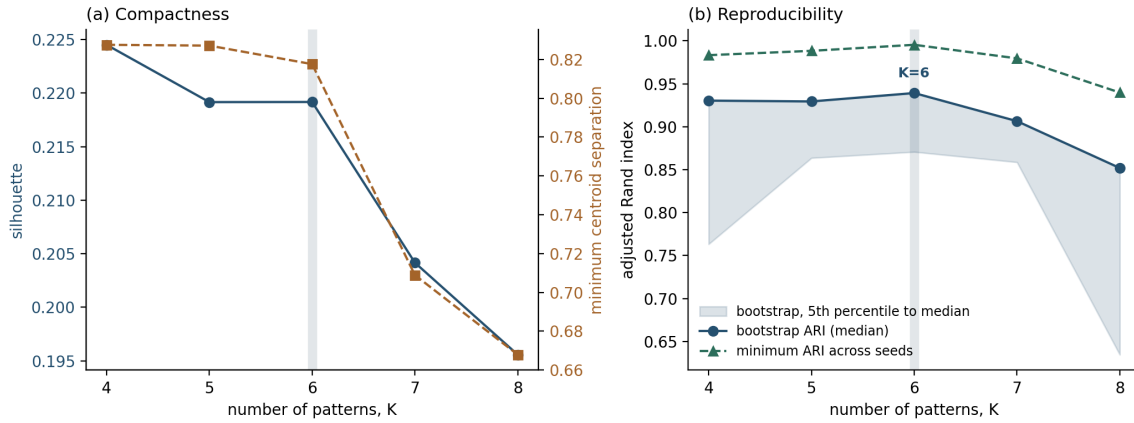


Figure S1: Criteria for the number of wave patterns, 1993–2022. (a) Compactness: silhouette coefficient (left axis) and minimum standardised separation between centroids (right axis), both declining with K . (b) Reproducibility: median adjusted Rand index across 100 calendar-year block bootstrap replicates, with the band spanning its 5th percentile to the median, and the minimum adjusted Rand index across 20 seeds. The selected partition $K = 6$ is highlighted in both panels.

Table S2: Observations assimilated by the WAVERYS reanalysis, by mission, variable and period of use, after Law-Chune et al. (2021) and Table 3 of the product quality documentation (*E.U. Copernicus Marine Service Information (CMEMS)*, 2024). Altimeter significant wave height is reprocessed by CERSAT-IFREMER before assimilation. In situ observations are not assimilated, which is what allows the Fortaleza buoy to serve as an independent reference; the producers themselves validate the reanalysis against moored buoys and against the non-assimilated HY-2A altimeter. Two missions present in the CCI record used here fall outside this table: Sentinel-3B, which it does not list, and SARAL before 1 March 2016. Together they supply 8.1% of the CCI observations collocated over the study domain in 1993–2022. The periods for the two spectral instruments follow the text of the quality document rather than its Table 3, which lists a start of March 2016 for CFOSAT, a mission launched in October 2018.

Mission	Assimilated variable	Period of use
ERS-1	Significant wave height	1 Jan 1993 – 2 Jun 1996
TOPEX/Poseidon	Significant wave height	1 Jan 1993 – 8 Oct 2005
ERS-2	Significant wave height	15 May 1995 – 4 Jul 2011
GFO	Significant wave height	7 Jan 2000 – 1 Aug 2006
Jason-1	Significant wave height	15 Jan 2002 – 21 Jun 2013
Envisat	Significant wave height	14 May 2002 – 8 Apr 2012
Jason-2 and Jason-3	Significant wave height	From 4 Jul 2008
CryoSat-2	Significant wave height	From 14 Jul 2010
SARAL	Significant wave height	From 1 Mar 2016
Sentinel-3A	Significant wave height	From 1 Mar 2016
Sentinel-1 SAR	Directional wave spectra	From Mar 2016
CFOSAT SWIM	Directional wave spectra	From Feb 2020
<i>In situ</i> buoys	<i>none</i>	—

Table S3: Directional statistics by wave pattern at the three analysis points over 1993–2022. Directions follow the coming-from convention, so a mean direction in the N–NE quadrant denotes waves arriving from N–NE and propagating towards S–SW. Sectors are centred on the cardinal points (N spans 348.75° – 11.25°) and the three most frequent are listed with their share of days. Quadrants are disjoint 90° intervals, so the four quadrant columns sum to 100% (the S–SW and W–NW quadrants are omitted, together accounting for less than 7% at every point and pattern). Because the E sector straddles the 90° boundary, a pattern may have E as its dominant sector while most of its days fall in the N–NE quadrant. Circular dispersion and the resultant length R are given as concentration measures.

WP	$\overline{H_s}$ (m)	$H_s > 2$ m (%)	Three leading sectors	Mean dir. ($^\circ$)	Disp. ($^\circ$)	R	N–NE (%)	E–SE (%)
<i>Foz do Amazonas</i>								
WP1	2.38	88.6	NE 87%, ENE 11%, NNE 2%	49.3	6.4	0.994	100.0	0.0
WP2	1.50	6.9	NE 48%, ENE 42%, E 8%	58.5	13.1	0.974	97.6	2.4
WP3	1.95	44.4	NE 75%, NNE 21%, ENE 3%	41.2	10.0	0.985	100.0	0.0
WP4	2.43	85.1	NE 48%, NNE 46%, N 6%	31.8	11.9	0.979	98.7	0.0
WP5	1.52	9.9	NE 56%, ENE 25%, NNE 13%	49.6	15.6	0.964	98.8	1.0
WP6	1.21	0.6	ENE 41%, E 39%, NE 14%	75.1	17.1	0.957	80.4	19.6
<i>Barreirinhas</i>								
WP1	1.81	20.3	NNE 46%, NE 44%, ENE 9%	36.9	12.8	0.975	100.0	0.0
WP2	1.56	10.9	E 46%, ENE 29%, NE 17%	71.4	21.0	0.935	80.2	19.8
WP3	1.66	9.9	NNE 49%, NE 34%, ENE 13%	35.7	16.9	0.958	99.8	0.0
WP4	1.94	38.9	NNE 59%, N 25%, NE 12%	20.6	14.8	0.967	95.6	0.0
WP5	1.54	11.4	E 35%, ENE 31%, NE 17%	64.5	24.8	0.911	87.1	12.8
WP6	1.74	22.1	E 86%, ESE 8%, ENE 4%	93.1	10.9	0.982	20.7	79.3
<i>Potiguar</i>								
WP1	1.73	13.3	NE 34%, ENE 26%, NNE 23%	52.9	23.0	0.923	94.1	5.9
WP2	1.66	9.8	ESE 52%, E 33%, ENE 10%	98.8	17.9	0.952	25.7	74.3
WP3	1.64	8.5	ENE 30%, NE 29%, NNE 21%	53.7	23.5	0.919	94.8	5.1
WP4	1.83	25.8	NNE 38%, N 22%, NE 21%	31.3	25.6	0.905	90.6	2.5
WP5	1.59	9.3	ESE 45%, E 32%, ENE 12%	95.2	23.2	0.921	31.7	68.3
WP6	1.97	45.4	ESE 75%, SE 23%, E 1%	118.8	7.8	0.991	0.6	99.4

Text S2. Analysis-period boundary

The annual domain-mean WAVERYS series were screened with the Pettitt test (Pettitt, 1979), which indicated a shift around the transition into the altimeter-assimilation era (Figure S2). The vertical line denotes the 1993 product boundary adopted in the main text. The wave-pattern classification, persistence and transition analyses use only 1993–2022; no trend is inferred from the complete 1980–2022 record.

Text S3. Construction of the daily feature vector

The classification described in Section 3.3 of the main text operates on a vector of eight numbers per day rather than on the gridded field itself. This section records how that vector is built and how missing data are handled, since the second point separates a wave-field classification from the weather-pattern classifications on which the procedure is based.

From three-hourly fields to eight features

The three-hourly WAVERYS fields of significant wave height H_s , mean wave direction and peak period T_p are first reduced to daily means. Each day is then summarised by the spatial mean and the spatial standard deviation of H_s , of the two directional components, and of T_p , giving eight features.

WAVERYS mean wave direction follows the coming-from convention and is a circular variable, so it cannot be averaged directly: the mean of 355° and 5° is 180° rather than 0° . Direction θ is

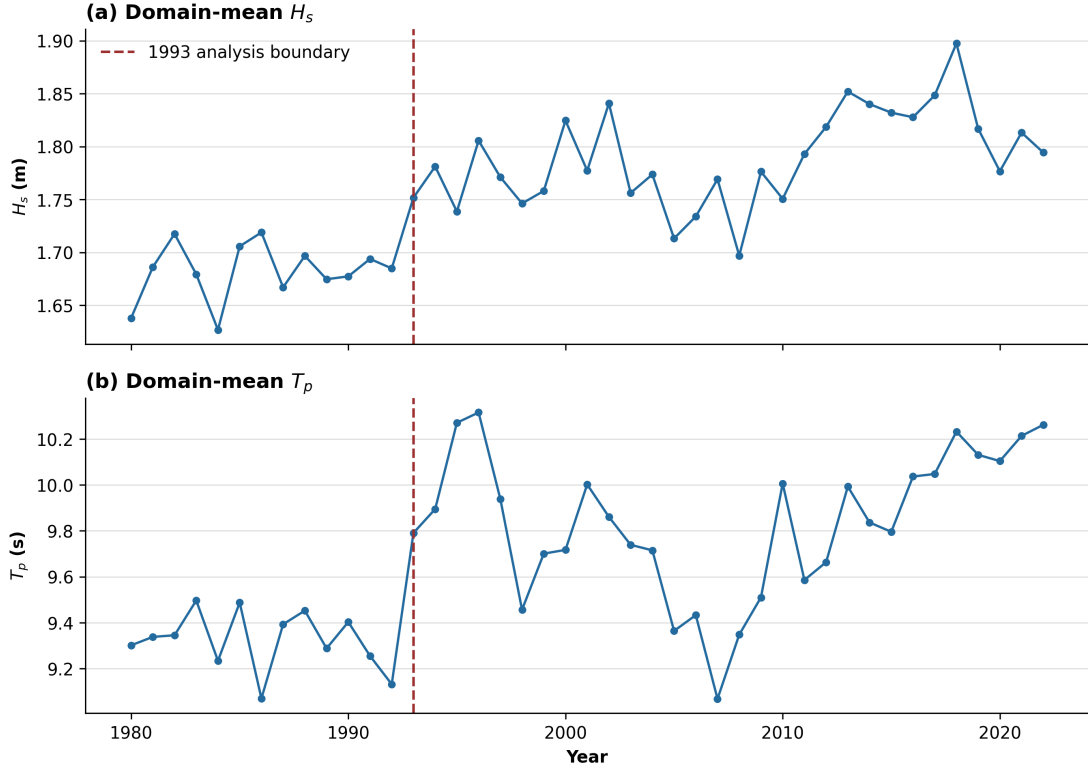


Figure S2: Annual domain-mean WAVERYS (a) significant wave height and (b) peak period over 1980–2022. The dashed vertical line marks the 1993 analysis boundary associated with the onset of altimeter assimilation.

therefore decomposed at each grid cell into the propagation-vector components

$$u = -\sin \theta, \quad v = -\cos \theta, \quad (1)$$

before the spatial statistics are taken, so that periodicity is preserved. The daily mean direction is itself obtained by averaging $\sin \theta$ and $\cos \theta$ over the eight three-hourly steps and recovering the angle with $\arctan(\cdot, \cdot)$.

All eight features are standardised to zero mean and unit variance over the study period. Without that step the distance metric would be dominated by T_p , whose numerical range in seconds is an order of magnitude wider than that of H_s in metres and of the directional components, which are bounded by ± 1 . This is a departure from da Silva et al. (2025) and Leal et al. (2026), who report that standardisation did not improve the partition of their pressure and wind fields; there the variables entering the distance share a common physical range, and here they do not.

Land, missing cells, and why the mask is fixed

Part of the model domain is land, and the ocean cells are not all valid at every step. Three rules govern this.

First, a single valid-ocean mask is taken from the first daily H_s field of the record and applied unchanged to every subsequent day. This matters more than it appears. If the mask were recomputed each day, the spatial mean of one day and the spatial mean of another would summarise different areas, and the Euclidean distance between the two feature vectors – which is the quantity K-means minimises – would confound a change in the sea state with a change in the region being averaged. Coastal cells that fall in and out of validity would then generate spurious structure in the partition. Fixing the mask removes that mechanism at the cost of discarding the few cells that are valid on some days only.

Table S4: Consistency of WEVERYS with the ESA Sea State CCI altimeter record over 1993–2022 at a 25 km distance-to-coast threshold, computed over all collocated observations and over the subset that WEVERYS does not assimilate. That subset is Sentinel-3B, absent from Table S2, over its whole record from June 2018, and SARAL before 1 March 2016; together they are 1 097 525 of the 13 641 077 collocated pairs, or 8.0%. The bias is unchanged to within 0.01 m in every stratum. The higher root mean square error and lower correlation of the subset follow from a sample twelve times smaller and drawn from a single mission at a time, without averaging across missions.

Stratum	All missions				Not assimilated			
	n	Bias (m)	RMSE (m)	r	n	Bias (m)	RMSE (m)	r
Foz do Amazonas	369 572	+0.082	0.188	0.955	28 276	+0.083	0.203	0.946
Barreirinhas	305 475	+0.073	0.182	0.916	27 590	+0.083	0.202	0.890
Potiguar	293 525	+0.080	0.183	0.898	23 495	+0.079	0.195	0.881
Open domain	13 641 077	+0.084	0.185	0.937	1 097 525	+0.087	0.200	0.926

Second, within the mask the spatial summaries ignore missing cells, uniformly and without latitude weighting. The domain spans 10°S–10°N, over which the cosine of latitude varies by less than 2%, so area weighting would not change the ordering of days.

Third, any day left with a non-finite feature after those two steps is rejected. Over 1993–2022 no day met that condition, so the classification covers the complete record of 10 957 days.

What is not done here

The clustering uses no dimensionality reduction: the eight features are used directly, without a prior principal-component step. It also uses no wave data beyond the BEqM domain; the Atlantic-wide fields enter only afterwards, in the lag composites, and never inform the partition.

References

- E.U. Copernicus Marine Service Information (CMEMS)* (2024) Quality information document for the global ocean waves reanalysis GLOBAL_MULTIYEAR_WAV_001_032. Reference CMEMS-GLO-QUID-001-032, issue 1.5. URL: <https://documentation.marine.copernicus.eu/QUID/CMEMS-GLO-QUID-001-032.pdf>. Table 3 lists the assimilated satellite missions and their periods of use. Accessed 2026-09-04.
- Law-Chune, S., Aouf, L., Dalphinnet, A., Levier, B., Drillet, Y. and Drevillon, M. (2021) WEVERYS: a CMEMS global wave reanalysis during the altimetry period. *Ocean Dynamics*, **71**, 357–378.
- Leal, K. B., de Souza, D. C., Laviola-da-Silva, M. B., da Pereira, S. S., Körting, T. S. and de Souza Robaina, L. E. (2026) Weather patterns associated with coastal disasters in Santa Catarina, Southern Brazil. *International Journal of Climatology*, **46**, e70384.
- Pettitt, A. N. (1979) A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C*, **28**, 126–135.
- da Silva, M. B. L., de Souza, D. C., Barreto, F. T. C., Tecchio, R., de Freitas Pimentel dos Anjos, R., Gramscianinov, C. B. and De Camargo, R. (2025) Classification of synoptic weather patterns associated with extreme wave events in different regions of western south atlantic. *Natural Hazards*. URL: <https://doi.org/10.1007/s11069-025-07166-7>. Accessed: 2025-04-30.