ARTICLE

Check for updates

Improved estimates of extreme wave conditions in coastal areas from calibrated global reanalyses

Valeria Fanti [1✉], Óscar Ferreira [1], Vincent Kümmerer [1] & Carlos Loureiro [1,2,3]

The analysis of extreme wave conditions is crucial for understanding and mitigating coastal hazards. As global wave reanalyses allow to extend the evaluation of wave conditions to periods and locations not covered by in-situ measurements, their direct use is common. However, in coastal areas, the accuracy of global reanalyses is lower, particularly for extreme waves. Here we compare two leading global wave reanalyses against 326 coastal buoys, demonstrating that both reanalyses consistently underestimate significant wave height, 50-year return period and mean wave period in most coastal locations around the world. Different calibration methods applied to improve the modelled extreme waves, resulting in a 53% reduction in the underestimation of extreme wave heights. Importantly, the 50-year return period for significant wave height is improved on average by 55%. Extreme wave statistics determined for coastal areas directly from global wave reanalyses require careful consideration, with calibration largely reducing uncertainty and improving confidence.

[1] Centre for Marine and Environmental Research (CIMA/ARNET), Universidade do Algarve, Faro 8005-139, Portugal. [2] Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA Scotland, UK. [3] Geological Sciences, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Durban 4001, South Africa. ✉email: vfanti@ualg.pt

xtreme wave conditions cause large disruption in coastal areas[1] and their impacts are likely to increase due to climate change[2,3], including sea level rise[4], while also being enhanced by expanding coastal populations[5,6]. Therefore, characterising the nearshore wave climate is fundamental for present and future coastal management, as extreme wave events can potentially affect a wide range of socio-economic activities. Storms or extreme wave events enhance sediment transport[7] and modify coastal geomorphology[8,9], cause erosion and flooding[10] and must be considered when designing and implementing coastal structures or marine energy devices[11,12]. Knowledge of the extreme wave climate is also necessary to characterize coastal hazards, quantify the risks associated with different return periods, and propose management actions as a function of the expected consequences[13]. If the extreme wave parameters used in such analyses are biased (i.e. underestimated or overestimated), the error is propagated to the estimation of the related coastal hazards and risks. This can lead to incorrect coastal risk assessments, affecting and compromising the planning, the mitigation efforts, and ultimately the adoption of risk reduction measures.

At a global level, reliable in-situ measurements of waves in coastal areas (e.g. buoys) are still scarce, comprise records with relatively short duration (often only a few years or decades), have limited geographical distribution[14] and are subject to errors[15,16] that require appropriate quality control. Comprehensive characterisation of extreme wave conditions derived from satellite altimetry data is also irregular in time and space, scarcely available in coastal regions and high latitudes, and requires complex post-processing techniques for retrieving accurate information in coastal areas[17]. Wave hindcast models are a powerful alternative for wave climate analysis, allowing to minimise the lack of data by providing consistent global datasets of past wave conditions, including coastal areas, and temporally and spatially augmenting wave observations.

Several global and regional wave reanalyses, derived from the assimilation of atmospheric and oceanographic data from models, satellites and in-situ observations, have been developed and improved in the last decades, and their reliability and use for coastal hazard analysis increased substantially[18–21]. Until recently, the most widely used wave reanalyses on a global scale were the 1.5° spatial resolution ERA40[22] and 1° ERA-Interim[23] produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), the CFSR/NCAR (0.5°/0.07°) produced by NCEP/NOAA[24] and GOW2[25] (0.5°/0.25°) from IH Cantabria. In the past few years, new global wave reanalyses have been produced, including the 0.5° ERA5[26] by ECMWF and 0.2° WAVERYS[27] by the Copernicus Marine Environment Monitoring Service (CMEMS). These last generation and up-to-date wave reanalyses are becoming the most commonly used reanalyses globally, and several studies have demonstrated the accuracy and improvements they have achieved[28–32]. Compared to wave buoys, reanalyses have the advantage of being a physically consistent reconstruction of past wave climate, with no gaps in space and time. However, they have coarse resolutions[33] that cannot fully resolve the complexity of coastlines, accurately incorporate regional coastal winds, simulate depth-limited wave propagation processes across shallow parts of the continental shelf, or entirely incorporate local processes such as bottom friction, wave breaking and sub-grid island shadowing[34,35]. Despite being quite reliable in the open ocean, wave reanalyses can still exhibit systematic errors and thus consistently over or under predict wave parameters in coastal areas with respect to in-situ observations[36–39]. Consequently, modelled outputs from wave reanalysis might not be directly applicable to coastal areas without previous validation and calibration[40–42].

To address this limitation, several bias correction methodologies[36,43–45] and improved calibration methods[12,46–50]

have been developed to produce wave time series that better fit observations. Some authors focused on finding linear relationships between buoy and reanalysis data for the n-year return period[12,47,51], others proposed spatial calibration that considers multiple observations[49], while more recently a calibration that depends on mean wave direction was proposed[48,50]. At regional and local scales, downscaling has also been used to better resolve the physical processes and improve reanalysis accuracy[52,53]. However, this method does not correct systematic errors and instead propagates the errors of the global reanalyses[54], which are normally used as boundary conditions to force the downscaled local wave propagation models. Despite these limitations, wave data from global reanalyses have been routinely and increasingly applied in coastal studies without validation and/or calibration[20,55–62] due to their easiness of use and because measured data are often not available. The systematic errors of global reanalyses are often amplified for extreme values[39,63], resulting in greater under or over prediction[38,64] and potentially causing large errors and uncertainties in the determination of the return periods of extreme events[12,65], incorrectly estimating consequences and potentially misrepresenting coastal risks and influencing the subsequent coastal management decisions.

Aiming to improve the characterisation of extreme wave conditions and associated return periods for coastal areas based on validated and calibrated global wave reanalysis data, this work evaluates ERA5 and WAVERYS reanalysis against a database of 326 coastal buoys. Results indicate a globally consistent underestimation of extreme wave conditions and associated return periods, which can be considerably reduced using a range of calibration approaches based on in-situ observations. To overcome data limitations in many coastal areas worldwide, a global calibration is proposed and shown to improve accuracy and reduce uncertainty in estimates of extreme wave conditions in coastal areas. The findings of this work provide an improved understanding of the limitations of global wave reanalyses and enhance their applicability around the world's coasts.

Results

Overview. The evaluation of the ERA5 and WAVERYS reanalysis time series for significant wave height (H_s), mean wave period (T_m), peak wave period (T_p) and mean wave direction (Dir_m), was performed against data from 326 globally distributed coastal buoys using standard errors metrics (see Materials and Methods). Independent calibration of the wave reanalysis data for each individual buoy location was performed using four transfer functions, improving estimates of wave parameters in areas where in-situ data is available. Considering all the locations where a negative relative bias in extreme wave conditions was identified in the reanalyses as a single time series, the same transfer functions were applied to determine and validate a global calibration for extreme wave conditions (see Materials and Methods).

Evaluation of global wave reanalysis. When compared against the 326 coastal buoys, both reanalyses show a very strong correlation for H_s, with a mean Pearson correlation coefficient (R, Eq. (5)) of 0.91 for WAVERYS and 0.90 for ERA5 (all R values reported are statistically significant for p-value <0.01). Overall, WAVERYS performed better than ERA5, presenting higher or identical R coefficients for H_s, T_m, T_p and Dir_m (Supplementary Fig. 1a, b). On average, the lower scatter index (SI, Eq. (6)) for H_s, 22% for WAVERYS and 25% for ERA5, and smaller root mean square error (RMSE, Eq. (7)), 0.33 m and 0.37 m respectively (Supplementary Figs. 2a, b and 3a, b), confirm the improved performance of WAVERYS, reinforcing the findings of Law-Chune et al.[27]. Although the reanalyses display both negative and
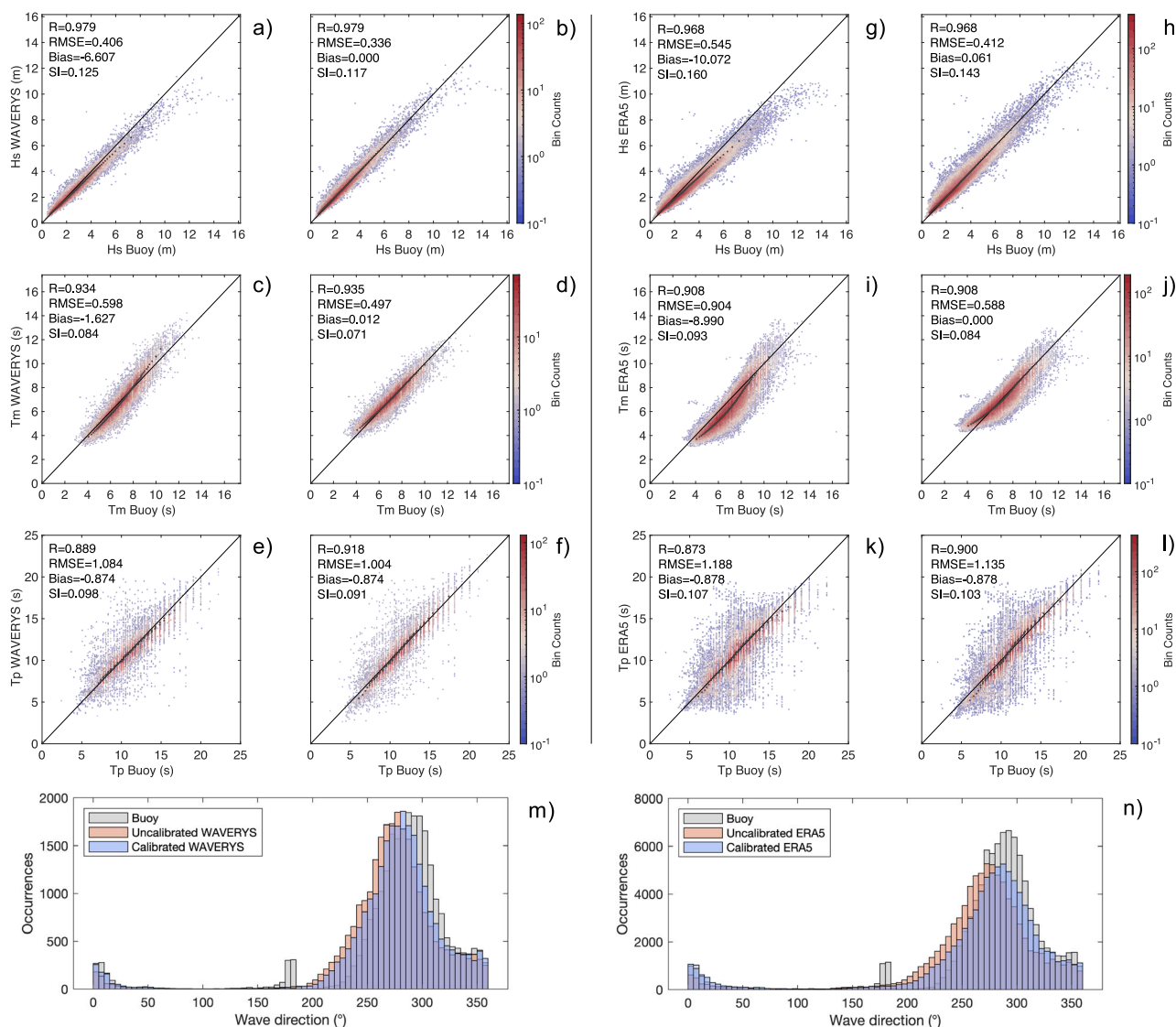
Fig. 1 Calibration of wave reanalysis parameters for Belmullet. Example of original (**a, c, e, g, i, k**) wave parameters (H_s, T_m and T_p) and calibrated for individual buoy locations (**b, d, f, h, j, l**). Results shown for the Belmullet buoy location, in western Ireland, for WAVERYS (**a–f**) and ERA5 (**g–l**). Results for mean wave direction (in °) are shown for original (light red) and calibrated for individual buoy locations (light blue) for WAVERYS (**m**) and ERA5 (**n**) and compared to the buoy data (grey). The black dots indicate the percentiles (from 1st to 99th).

positive relative biases (Eq. (8)) for H_s (Supplementary Fig. 4a, b), there is a consistent underestimation of H_s in the upper quantiles, above the 95th percentile for over 95% of the buoy locations (Fig. 1a, g for Belmullet in Ireland and Supplementary Fig. 5a, b) and above the 99th percentile for 98% of the 326 locations considered (Supplementary Fig. 6a, b). Positive biases were found around islands (e.g., in the Azores, Canary Islands, Caribbean Islands and Taiwan) and channels (e.g., Strait of Gibraltar and English Channel), which evidences the limitations of global reanalyses to resolve the sheltering effects of small islands and complex coastal configurations. The relative bias computed using only data from the higher percentiles (>95th percentile) is negative for 94% (78%) of the buoy locations in WAVERYS (ERA5), with results indicating an average underprediction of −16% (−9%) in the two reanalyses (Fig. 2a, b). Given the underestimation of the extreme wave heights, the resulting H_s 50-year return periods are underestimated on average by 1.11 m for WAVERYS and 0.80 m for ERA5 (Supplementary Fig. 7a, b). Return periods were only determined for locations with more than 12 years of observations, which corresponds to 30% of the

326 buoys used for calculation of the error metrics. Underestimation is highest in enclosed basins or sheltered locations: specifically the Sea of Japan with deviations of 3.08 m for the H_s 50-year return period determined using WAVERYS; and coastal areas in South Korea and the Balearic Sea with 2.95 m deviation for ERA5 (Supplementary Fig. 7a, b). Overall, the H_s value for the 50-year return period is underestimated in 89% (80%) of the locations considered for extreme value analysis in WAVERYS (ERA5).

Regarding T_m and T_p, results indicate strong correlation for both reanalyses, with an average R of 0.80 and 0.61 for WAVERYS and 0.79 and 0.59 for ERA5, respectively (Supplementary Fig. 1c–f). Underestimation was generally observed for T_m, with higher values identified in ERA5 around islands and channels (Supplementary Fig. 4c, d). The average relative bias for T_m is −9% (−13%) for WAVERYS (ERA5), increasing to −12% (−18%) when calculated for the data above the 95th percentile (Supplementary Fig. 8a, b). The underestimation of T_p in both reanalyses is observed mostly in the higher percentiles (Supplementary Fig. 8c, d). In fact, while the average relative bias for T_p is
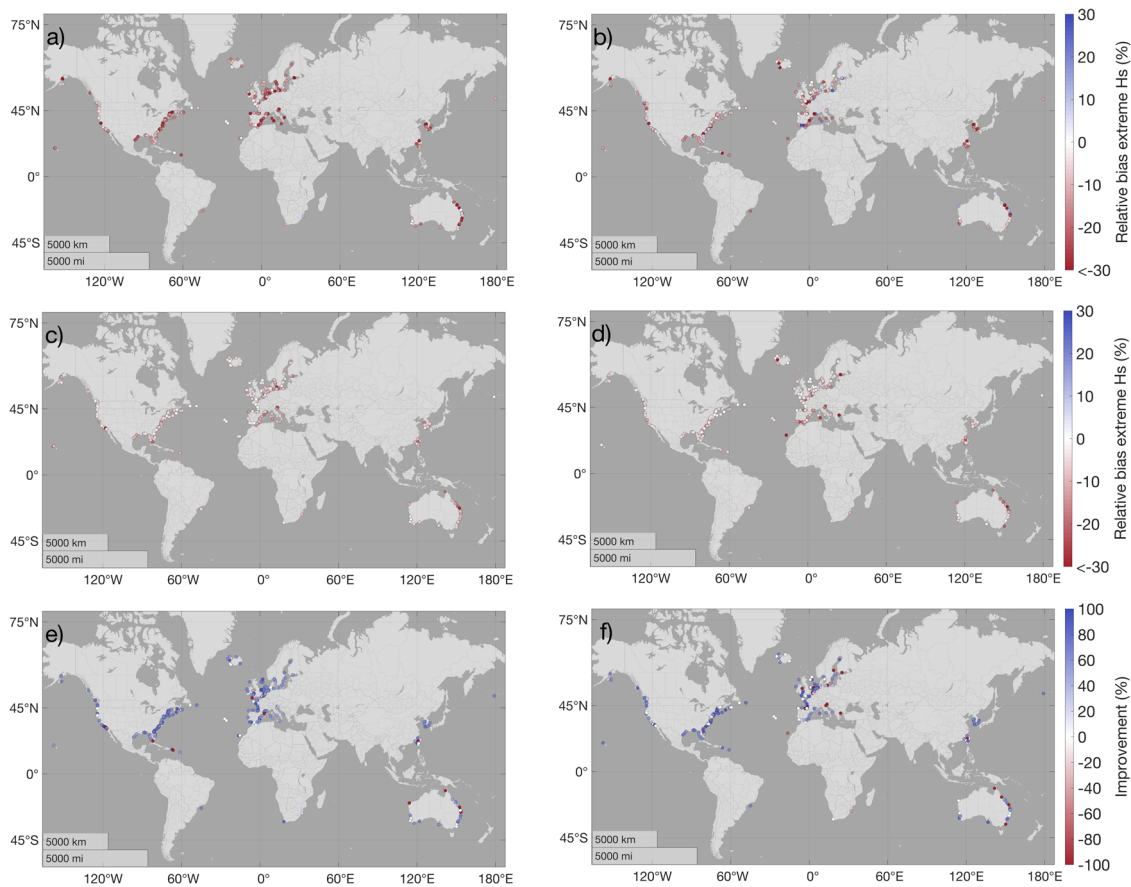
Fig. 2 Relative bias for extreme H_s values. Global distribution of bias (in %) for extreme (above the 95th percentile) values of H_s between wave buoys and original reanalysis data (**a**, **b**), individual calibration (**c**, **d**) and percentual improvement obtained with the calibration (**e**, **f**) for WAVERYS (**a**, **c**, **e**) and ERA5 (**b**, **d**, **f**). Red (blue) values represent an underestimation (overestimation) from the reanalysis (in **a–d**).

1% (−1%) in WAVERYS (ERA5), it increases to −15% (−20%) when considering only the values above the 95th percentile, with both reanalyses showing underestimation for 95% of the buoy locations (Supplementary Fig. 8c, d). This is reflected in the mean difference between buoy and reanalysis data for the extreme T_m and T_p values (Supplementary Figs. 5c–f and 6c–f). A total of 183 buoys incorporated in this analysis included directional information, allowing a comparison of Dir_m between buoy and reanalysis data. From the Dir_m histograms, the dominant wave regimes are well reproduced by both reanalyses, although with some deviation in directional peaks (Fig. 1m, n). The average absolute difference between the circular mean direction of the buoys and the reanalyses is 17° (20°) for WAVERYS (ERA5), with the largest differences identified in the Azores islands, Strait of Gibraltar and areas of the Baltic sea (Supplementary Fig. 9a, b). Overall, both reanalyses evidence a strong correlation of Dir_m with buoy data, with an average R of 0.64 (0.62) for WAVERYS (ERA5) (Supplementary Fig. 1g, h). The highest R values (>0.90) were obtained along the West coast of the USA and the North coast of the Netherlands, while the lowest were in sheltered locations in the Caribbean, Azores and Strait of Gibraltar ($R < 0.2$ for both reanalyses). The average RMSE of Dir_m was 40° (44°) for WAVERYS (ERA5), with the lowest performance in the same sheltered locations, where RMSE exceeds 80° (Supplementary Fig. 2g, h). The relative bias for Dir_m is positive for 63% (61%) of buoy locations for WAVERYS (ERA5), indicating that reanalysis waves are dominantly biased in a counter-clockwise direction, with an average circular bias of less than 2° in both reanalyses (Supplementary Fig. 4g, f). The deviations in Dir_m between buoy

and reanalysis data are consistent in both reanalyses for 85% of the cases. These results highlight a similar performance between both reanalyses in terms of Dir_m with differences in buoy and reanalysis likely due to the complex interaction between wind-sea and swell components[39] and coastal configuration.

Calibration of reanalyses for individual buoy locations. Depending on the method applied to calibrate the reanalysis data, different improvements between R and relative bias are achieved. When the first-degree polynomial (Eqs. (9) and (10)) is used, R is unmodified, and the bias is corrected. When rotation around the mean is applied instead (Eq. (12)), the time series rotates and the bias remains the same, but R is improved. When a power function is used (Eq. (11)), R remains identical while the bias is improved. Additionally, in a limited number of cases (1–15% depending on the model/variable considered), the original reanalysis data were already accurate, and no calibration was applied. In 71% of the cases, either a power function (Eq. (11)) or rotation (Eq. (12)) was applied to correct H_s in the reanalyses, with R improving by 34% (26%) for WAVERYS (ERA5) (Supplementary Fig. 10a, b). For the locations that required bias correction (i.e., the reanalyses were corrected with a first-degree polynomial (Eqs. (9) and (10)) or a power function (Eq. (11)), changes in the relative bias resulted in an average improvement of 99% (97%) for WAVERYS (ERA5). SI and RMSE improved with all the calibration methods applied, with SI reduced on average by 18% for both reanalyses, while the RMSE for H_s improved by 18% (20%) for WAVERYS (ERA5) (Supplementary Fig. 10a, b). The reanalysis time series for T_m were calibrated with a first-degree polynomial (Eq. (10)) in 49% of the

cases. The calibration of T_m resulted in a relative bias improvement of 86% (69%) for WAVERYS (ERA5). Calibration of T_p was performed using the rotation around the mean (Eq. (12)) for 59% of the cases, with 15% not requiring calibration. Overall, T_p calibration contributed to an improvement of 36% (24%) in relative bias for WAVERYS (ERA5). It is important to note that in some cases the bias value in T_p is related to the presence of secondary swells with longer wave period, which are not correctly captured in either reanalysis. In terms of Dir_m, the difference in the circular mean computed between the buoys and the reanalyses was reduced from an average of 17° to 10° for WAVERYS and from 20° to 13° for ERA5. The relative bias decreased by 21% (16%) on average for WAVERYS (ERA5), but no significant improvements were obtained for R, SI and RMSE. While no calibration of the directions was deemed necessary for 50% of the buoy locations, results obtained for the calibrated sites show considerable improvements (e.g., Belmullet, Fig. 1m, n).

For extreme wave conditions, i.e. H_s above the 95th percentile, the calibration of H_s reduced the average underestimation by 0.30 m (0.15 m) for WAVERYS (ERA5) (Supplementary Fig. 10b). When considering H_s above the 99th percentile, the calibration improves results further, with a reduction in the average underestimation of 0.45 m (0.22 m) for WAVERYS (ERA5) (Supplementary Fig. 10). The relative bias for H_s values above the 95th percentile (Fig. 2c, d) was reduced to −7% (−8%), which corresponds to an average improvement of 56% (36%) for WAVERYS (ERA5) (Fig. 2e, f and Supplementary Fig. 11a, b). The calibration for individual buoy locations allowed for relevant improvements in the H_s return period values, with differences in the 50-year return period estimated with buoy observations and with reanalysis data reduced by ~0.68 m (0.37 m) for WAVERYS (ERA5), corresponding to a 61% (46%) average improvement (Fig. 3b and Fig. 4a, b). This included areas like the southeast coast of Spain, where the underestimation of the 50-year return period in WAVERYS was corrected by 2.5 m with the calibration for individual buoy locations. The underestimation of the relative bias for the extreme (above the 95th percentile) T_m (T_p) was reduced to −11 % (−12%) for both reanalyses, while the average underestimation of the extreme T_m (T_p) values was reduced to 0.8 s (1.5 s).

Global reanalysis calibration. Based on the common underestimation of H_s and T_m in both reanalyses for most coastal areas,

a global calibration for each reanalysis dataset is proposed. The equations for the global calibration of H_s and T_m were determined by identifying the best-fit parameters considering the entire dataset as a unique time series, but excluding the data from buoys located in areas where extreme wave heights are overestimated by the reanalyses. These are areas where the extreme relative bias is positive (Fig. 2a, b), and can be found in sheltered coasts and areas with complex coastal configurations, which represent 6% (22%) of buoy locations for WAVERYS (ERA5). For both reanalyses, a first-degree polynomial performed better for the global calibration of H_s, indicated by Eq. (1) for WAVERYS and Eq. (2) for ERA5, while a first-degree polynomial with intercept provided the best correction for T_m, shown in Eq. (3) for WAVERYS and Eq. (4) for ERA5. No calibration method provided a significant improvement for T_p, in the reanalysis datasets.

Based on the data from a group of randomly selected buoys (30% of the dataset) excluded from the global reanalysis calibration, a validation of the global calibration equations using the error metrics (detailed in Materials and Methods) was performed (Table 1). The mean difference between buoy observations and the globally calibrated reanalysis data for H_s values above the 95th percentile improved on average by 0.14 m (0.10 m) for WAVERYS (ERA5), while the relative bias for H_s above the 95th percentile, was reduced to −12% (−11%) (Table 1). The averaged relative bias for extreme T_m was reduced by 4% (12%) for WAVERYS (ERA5), indicating that the underestimation was partially corrected. After applying the global calibration to all the buoy locations where a negative relative bias for extreme wave conditions was identified, the 50-year return period H_s values were corrected on average by 0.30 m (0.16 m) for WAVERYS (ERA5) (Fig. 3b) and improvements in the overall return period estimates were observed in 90% (70%) of the locations for WAVERYS (ERA5) (Fig. 4c, d).

$$H_{sCWYS} = 1.077 * H_{sWYS} \qquad (1)$$

$$H_{sCERA5} = 1.045 * H_{sERA5} \qquad (2)$$

$$T_{mCWYS} = 0.870 * T_{mWYS} + 1.124 \qquad (3)$$

$$T_{mCERA5} = 0.928 * T_{mERA5} + 1.156 \qquad (4)$$
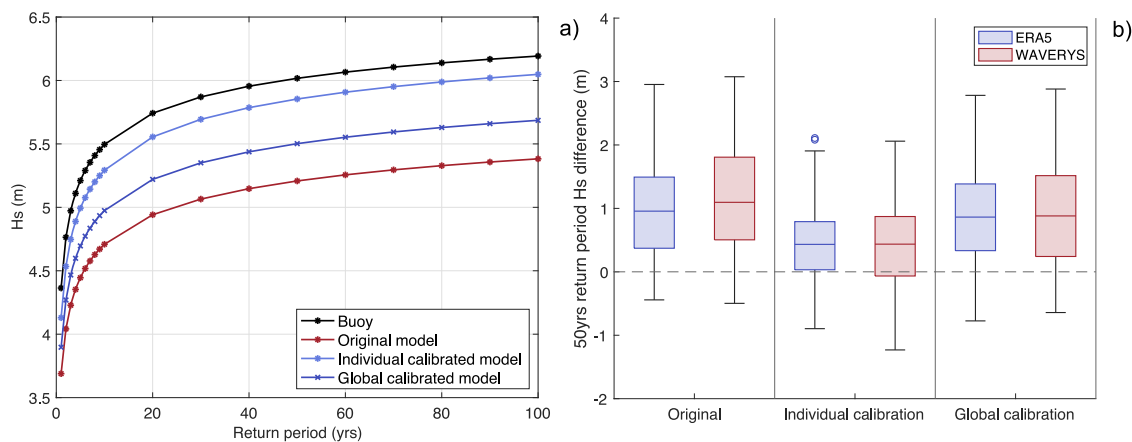
Fig. 3 Effect of individual and global calibration on H$_s$ 50-year return periods. a Return period values for original WAVERYS reanalysis (red) and buoy data (black), and with the individual (light blue) and global (dark blue) calibration for the location in the coast of South Carolina, USA. **b** Box plot of differences in H$_s$ 50-year return period between buoy data (only buoys with negative relative bias for extreme H$_s$) and reanalysis data, estimated with the original reanalysis data, individual calibration, and global calibration for WAVERYS and ERA5. The bottom and top of each box plot are the 25th and 75th percentile, the line in the middle the median, the whiskers correspond to 99.3% of the data and the outliers are reported as circles.
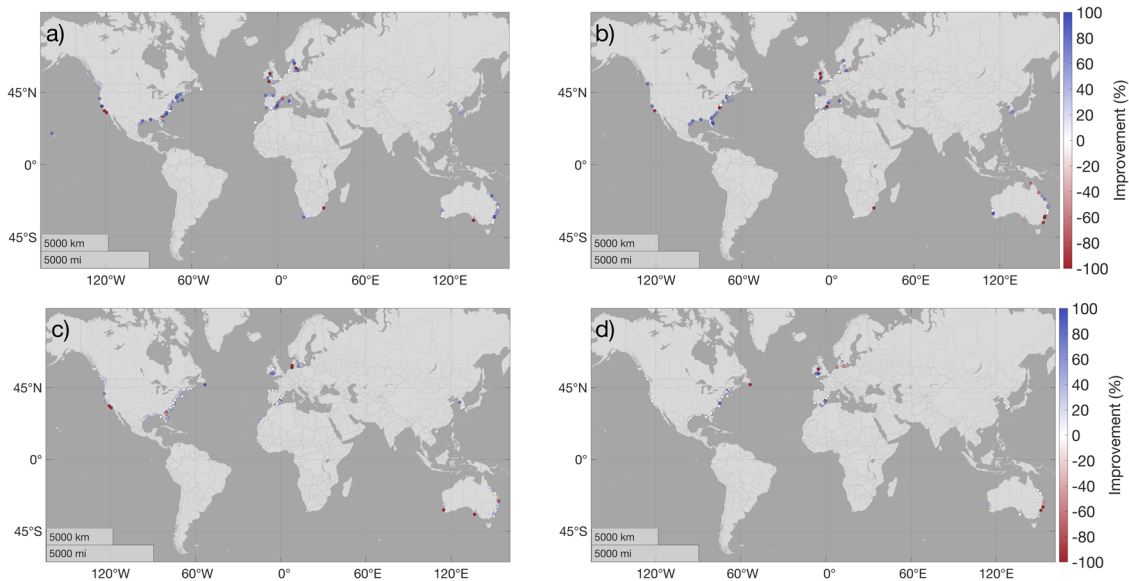
Fig. 4 Relative improvement of the H$_s$ 50-year return period. Global distribution of the improvement (in %) for the 50-year return period H$_s$ calculated for the wave buoys and individual calibration (**a**, **b**) and global calibration (**c**, **d**) for WAVERYS (**a**, **c**) and ERA5 (**b**, **d**).

Table 1 Error metrics for original and calibrated WAVERYS and ERA5 global wave reanalyses.

Significant wave height (m)

Average of all buoys with extreme bias <0	R		Relative bias for H$_s$ (T$_m$) >95th percentile (%)		Mean difference for values >95th percentile		Mean difference for values >99th percentile	
	WAVERYS	ERA5	WAVERYS	ERA5	WAVERYS	ERA5	WAVERYS	ERA5
Original reanalysis	0.91	0.91	−17	−14	0.55	0.54	0.86	0.78
Individual calibration	0.94	0.93	−7	−6	0.21	0.19	0.36	0.32
Global calibration (30%)	0.91	0.90	−12	−11	0.41	0.44	0.65	0.72
Global calibration (70%)	0.91	0.91	−12	−12	0.39	0.38	0.66	0.66
Mean wave period (s)								
Original reanalysis	0.80	0.80	−13	−20	0.92	1.61	1.25	2.13
Individual calibration	0.83	0.87	−11	−11	0.83	0.87	1.24	1.28
Global calibration (30%)	0.80	0.81	−9	−8	0.86	0.69	1.11	1.21
Global calibration (70%)	0.80	0.80	−8	−9	0.68	0.76	1.10	1.18

Calibration results including the correlation coefficient R, relative bias of the values above the 95th percentile, mean differences for values above the 95th and 99th percentiles determined from the buoys and the original, individual and global calibrated reanalysis data, considering the randomly selected 70% of buoys used to find the global equations and the remaining 30% buoys used to validate the equations.

Discussion

In the past decade, the performance of global reanalyses to estimate (hindcast) past wave conditions has improved significantly, and reanalysis data are now widely applied in different fields. They are highly reliable in the open ocean[27,29,66], while their inaccuracies have been increasingly reported in coastal areas[35,40,42]. Here, the evaluation of global reanalysis against data from 326 buoys distributed around the world's coasts allowed to determine that the higher-resolution WAVERYS reanalysis performed better than the more widely used ERA5 reanalysis. These results are consistent across most error metrics (R, RMSE, and SI) and for the most used wave parameters (H$_s$, T$_m$, T$_p$ and Dir$_m$) in coastal areas. Law-Chune et al.[27] also compared these reanalyses with buoy observations from the Copernicus Marine Service observations database, and although a different set of coastal buoys was considered (up to 200 m water depth), WAVERYS performed better than ERA5, regardless of the coastal setting and geographical location of the buoys[27]. Spatially, lower SI values (Supplementary Fig. 3a, b) are found on the north-eastern Pacific coast and larger SI, up to 30%, is observed in semi-enclosed

regions such as the Mediterranean Sea[27]. In terms of wave direction, the mean differences for Dir$_m$ are consistent with previous research[27], with higher values found in enclosed basins and around small islands. So far, comparisons of the global performance of WAVERYS and ERA5 against coastal buoy observations are limited, but a recent regional assessment by Crespo et al.[67] for the south-western Atlantic Ocean, an area underrepresented in terms of wave buoy data, also identified improved performance by WAVERYS for mean and extreme H$_s$.

The main contribution of this work is the detailed analysis of the errors in the estimation of extreme wave conditions in global wave reanalyses and the impact this has on the determination of return period values, which are fundamental for coastal risk analysis. A main finding is the overall underestimation of coastal H$_s$ in global reanalyses, especially for extreme wave heights (H$_s$ > 95th percentile), as highlighted by average negative biases of −16% (−9%) in WAVERYS (ERA5). Underestimation of extreme H$_s$ in ERA5 based on comparison with buoy data was also found in Chinese waters[40], around Japan[42], and in the Mediterranean sea[53,68]. Cases of extreme H$_s$ overestimation were

identified for 6% (22%) of the buoy locations when compared to WAVERYS (ERA5). The overestimation in WAVERYS occurred mainly in coastal areas where the reanalysis resolution is insufficient to resolve nearshore wave changes across complex shelf bathymetries (in the presence of reefs, headlands, sheltering land), while ERA5 was found to overestimate extreme H_s for buoy locations that are closer to the coastline due to its coarser resolution. Previous work evidenced the overestimation of wave heights in ERA5 for locations in the Arabian Sea[69], as well as its precursor ERA-Interim in several coastal locations around India[41,70,71]. However, no regional patterns were identified in the present study for overestimated values of H_s. Instead, such cases corresponded to very specific locations, often surrounded by cases of underestimation in extreme waves and most likely related to the limitations of coarse resolution global models to resolve local wave propagation and transformation.

Underestimation of T_m was also found in most coastal areas analysed, with an average relative bias of −9% (−13%) for WAVERYS (ERA5). T_m was overestimated in less than 10% of the buoy locations considered in both reanalyses, mostly coincident with the sites where H_s was also overestimated, and for very long wave periods (>95th percentile). Overestimation of T_m in ERA5 is not uncommon and has been reported along the Chinese coast[40], Arabian Sea, Bay of Bengal[72] and in the Mediterranean Sea[53]. T_p was also consistently underestimated by global reanalyses for 90% of the buoy locations and for conditions above the 95th percentile, with average relative biases of −15% (−20%) for WAVERYS (ERA5), confirming previous results for the North Indian Ocean[72] and in the Greek seas[68].

Assessments of the performance of global wave reanalysis rarely consider wave direction, mainly because in coastal areas the wave direction is highly dependent on the model resolution and the ability to simulate small scale processes[53]. However, accurately determining wave direction is extremely important for coastal studies related to sediment transport[73] and coastal erosion[74]. Although the calibration of wave direction cannot be applied at a global scale and instead requires local buoy data, application of a shift in the distribution of wave directions as demonstrated in this work, allows to improve the representation of the peaks in the directional distribution and achieve a better average direction in the reanalysis datasets (see Fig. 1n, m as an example). Furthermore, WAVERYS and ERA5 present similar wave direction distributions, and this is likely due to the use of the same 2-minute ETOPO2 bathymetric grid and the fact that WAVERYS is forced with the ERA5 10-m wind conditions.

The systematic underestimation of wave parameters in the ERA reanalysis products developed by ECMWF is not new. It was reported for the previous reanalysis, ERA-Interim, when compared to buoy data from the National Data Buoy Centre in the USA[38], in the Atlantic coast of Spain and the Italian coast[75]. It is widely known that global wave models, such as WAM used by both WAVERYS and ERA5 in different versions, tend to underestimate extremes in H_s, and hence T_p[63]. This is mostly due to uncertainties in wind forcing[39], but also because of limitations in the computation of various wave processes (bottom friction, wave breaking, island shadowing, and fetch length)[34,63]. However, more than the numeric errors in reanalysis datasets, the potential risks associated with using model outputs for coastal applications that systematically under or overpredict wave conditions lay in the likely under or overestimation of the outcome. This includes assessments of wave energy and impact, erosional potential, flooding severity, amongst others. For example, deviations in the modelled H_s are propagated quadratically in terms of wave energy calculation (or linearly in the case of T_p). This is particularly relevant for extreme H_s conditions, as the higher percentiles are poorly estimated by wave models[63].

Coastal hazard and risk assessments are often based on the definition of extreme values for different return periods[13]. The widespread global underestimation of the wave parameters associated with a given return period based on original reanalysis data, highlighted in the results presented here for most coastal locations, will result in a potential minimisation of the hazard or risk level. Despite this limitation, global wave reanalyses have and continue to be widely used for coastal studies without calibration or without quantifying their uncertainties. The applications of uncalibrated reanalyses include analysis of wave climate[71,76] and extreme events[57,58], coastal wave energy assessments[70,77,78], estimation of overtopping[20], shoreline retreat[62] and coastal flooding[56]. As the recognition of the limitations of global reanalyses increases, it becomes evident that it is fundamental to improve return period estimates from global reanalyses[65] to ensure the accurate definition of coastal hazards and risk analysis at local to global scales.

To minimise the underestimation that is common in global wave models, particularly for extreme conditions, different calibration techniques have been developed to correct wave reanalysis[36,48,50] or explicitly the extreme wave heights[79] and their return periods[12,47,51]. Although directional calibration techniques exist[48,50] for H_s, the calibration of the direction itself, and the periods, have not been addressed. Underestimation of H_s extremes occurs randomly in wave models and not all extreme H_s values need correction[46]. Thus, the application of simple parametric corrections can also lead to errors and uncertainties, since many complex factors influence the prediction of extreme events[63]. However, the results obtained by applying four different transfer functions to calibrate time series of wave parameters, demonstrated that it is possible to improve the estimation of extreme wave conditions by correcting the overall underestimation of H_s, T_m and T_p in reanalysis datasets. Substantial improvements were observed (Fig. 3a and Fig. 4a, b) in the estimation of the H_s values associated with long return periods (50 years), which are often used for hazard and risk assessments in coastal areas[13], demonstrating the importance of calibrating wave reanalyses before their use in coastal applications.

In ideal conditions, calibration should be performed against wave buoy data from a location close to the reanalysis grid node. This not only improves the reanalysis but allows extending, with higher accuracy, the often limited temporal coverage of wave buoy time series. Once a reanalysis is calibrated for a specific location, the results can be extended to adjacent areas with similar morphology and wave climate. However, when in-situ observations are not available, this work proposes a global calibration equation for H_s and T_m, to be applied in coastal areas where ERA5 and WAVERYS reanalyses are known to underestimate wave heights and mean periods. This excludes locations with complex coastal configurations, sheltered by islands, headlands or reefs, as well as nearshore areas where shallow water processes cannot be properly captured by global coarse resolution reanalyses. In those cases, calibrated reanalysis could be used after downscaling with a regional wave model that better resolves the impacts of complex coastal configurations on wave propagation. Based on the validation performed with the buoys that were not included in the global calibration equations, the estimation of H_s for different return periods using the global calibration achieves improvements for 90% (70%) of the buoys compared to WAVERYS (ERA5), although the calibration for individual buoy locations allows a greater improvement. The main potential of the global equations proposed is their broader applicability when compared to the individual calibration, which require in-situ data to be applied. The underestimation of the widely used H_s 50-year return period was reduced on average by 25% (15%) for WAVERYS (ERA5). This allows more reliable results in any

coastal engineering application that requires the use of return periods, whether for design purposes, coastal hazard and protection analysis or evaluation of wave energy devices. Although global calibration equations have the merit of being a simple method to correct wave reanalysis data, their use should consider the specific characteristics of the wave climate in the study areas. In areas where the overall H_s time series is commonly overestimated by global reanalyses, such as in extremely shallow or sheltered areas, the global equation would further increase the error (i.e. the overestimation). Future work can potentially explore the development of basin scale calibrations to be applied for areas where reanalyses have clear and distinct regional patterns. This would allow a regionally consistent calibration for those areas where a global calibration may not be suitable.

In conclusion, by comparing two of the most recent and widely used global wave reanalyses, WAVERYS and ERA5, with observations from 326 coastal buoys in different areas of the world, this work demonstrates that global reanalyses have important limitations in the determination of extreme wave parameters in coastal areas, including widely used return period estimates. This highlights the need to perform site specific validation and calibration before using wave reanalyses in coastal areas in order to reduce their uncertainties. Such limitations can be problematic, particularly when assessing coastal hazards and associated risks. To address this, this work proposes efficient and reproducible calibration approaches to improve wave reanalyses in coastal areas, either using site-specific observations where available, or globally derived approximations.

Materials and methods

Two of the most recent and advanced wave reanalyses were selected for evaluation: ERA5[29], the latest reanalysis released by the European Centre for Medium-Range Weather Forecasts (ECMWF) with 0.5° resolution, and WAVERYS[27], a 0.2° resolution reanalysis by the Copernicus Marine Environment Monitoring Service (CMEMS). The reanalyses have different resolutions, dissipation terms, and white capping terms. WAVERYS includes wave-current interactions but is not coupled with an atmospheric model like ERA5. Both reanalyses have been extensively validated against buoy and satellite data (e.g. refs. [27,40,42,66],). In this work, the calibration of H_s, T_m, T_p and Dir_m was performed for 326 buoys located in distinct coastal areas around the world (Supplementary Table 1).

Wave reanalysis data. ERA5[29] is produced by the ECMWF, covers the period from 1950 to the present at hourly intervals, and it has improved spatial and temporal resolution and performance in comparison to its predecessors (ERA-40 and ERA-Interim)[28,29]. ERA5 provides records of the atmosphere, land surface, and ocean waves. The atmospheric model is coupled to WAM, a third-generation spectral WAve Model developed by the WAMDI Group[80]. ERA5 also assimilates H_s measured from satellites (ERS1, ERS2, SARAL, CryoSat2, Jason1, Envisat) into the predicted wave spectra[29]. The ERA5 wave data are available from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/). For the present study, synthetic parameters, derived from the wave spectra, were used. The parameters considered are the H_s of combined wind waves and swell, T_p, T_m, and Dir_m.

WAVERYS is the first global wave reanalysis produced by CMEMS[27] and released in December 2019. It has a spatial resolution of 0.2° and covers the period from 1993 to 2019 with a temporal resolution of 3 h. WAVERYS is not coupled with an atmospheric model but includes the 3 h surface currents from GLORYS12V1[81] global physical reanalysis. The altimeter wave data from ERS1 until Sentinel-3A missions[27] were assimilated into the model. WAVERYS is forced with the ERA5 10-m wind fields and the wave model used is MF-WAM version 4 implemented by Météo-France, which differs from the version used in ERA5 for the input and dissipation terms[39].

For both reanalyses, the data from the cell closer to the mooring location of each buoy were considered, unless the location was beyond a radius of 0.099° (0.249°) for WAVERYS (ERA5). In those cases, the data from adjacent cells were interpolated using a weighted approach based on the distance between the buoy location and adjacent grid cells centre. Regardless of the use of weighted averages or the values for individual grid cells, it must be recognised that reanalysis data are estimates of the average conditions for entire grid cells[82]. This has implications for the consistency between buoy and reanalysis data, not only because of the spatial scale but also because the grid cell depth will not represent exactly the buoy depth. This may contribute to differences in wave parameters between wave reanalysis and coastal buoys, particularly as depth influences wave transformation. However, the improved resolution of both reanalysis datasets, the developments in model source

terms and quality of bathymetric grids, as well as improved assimilation of a larger range of observations[27,29], provides increased confidence that buoy datasets are spatially consistent with the reanalysis datasets. In addition, the buoy data were linearly interpolated to a 1-hour temporal resolution (as described below), and such averaging over time is considered to bring the temporal and spatial scales of buoy and reanalysis data closer together[83].

Wave buoy data. Buoys are available for limited locations around the world, with the majority in the Northern Hemisphere, particularly in the North Pacific and North Atlantic, and large data gaps in the Southern Hemisphere[14]. In addition, buoy data are often discontinuous or cover only short periods, which makes the analysis of the return periods unfeasible in many locations[12]. This work used buoy measurements to evaluate and calibrate the two wave reanalyses. The in-situ wave observations were extracted from the global network of buoys provided by the CMEMS, with product name INSITU_GLO_WAV_REP_OBSERVATIONS_013_045[84], supplemented by additional buoys sourced from hydrographic agencies for underrepresented areas. The CMEMS database includes 1489 buoys with information about latitude, longitude and time series of H_s and in some cases T_m, T_p and Dir_m, together with quality flags (from 0 to 9, more details in the product user manual[85]) for each parameter. In cases where only the 1D wave spectra were available, the synthetic parameters were computed from the spectral moments. The criteria for selecting coastal buoys from the database included: i) water depth, only buoys in water depths of less than 100 m were considered (water depth was estimated using the global GEBCO2022 bathymetry); and ii) record length, only buoys with more than one complete year of data and including the winter months were considered. Buoys in lagoonal areas, small enclosed basins, protected by headlands or around small islands (less than ~22 km in length, which is the spatial resolution of WAVERYS) were removed, as their setting is not properly resolved by the relatively coarse resolution of global reanalyses. The buoys situated in coastal locations where both reanalyses had empty cells were also automatically excluded. All buoy records were inspected to remove spikes, quality control flags, flat lines, wrongly assigned variables and values outside the acceptance range[86]. The criteria and filters used reduced the set of buoys considered in this study to 326 (Supplementary Table 1). The time series from the buoys were interpolated at 1-hour intervals to ensure consistency with model outputs and homogenise variable temporal resolutions (e.g., some buoys increase the sampling frequency to 30 min during storm events) or fill in missing records for data gaps of less than 6 h. Gaps longer than 6 h were excluded from the analysis. The records from the buoys ranged from 1 year to 42 years, with 83 buoys including more than 12 years of data. These 83 buoys were used for the estimation of return periods, as described below. The number of buoys used for comparison with WAVERYS and ERA5 differs depending on the presence of a reanalysis cell at the buoy location and on the temporal range of the reanalysis, as WAVERYS only covers the period from 1993 to 2019. Moreover, while 326 buoys have records for H_s, only 210 also recorded T_m, 270 T_p and 183 Dir_m.

Evaluation. The performance of wave reanalysis was evaluated against wave data (for H_s, T_m, T_p and Dir_m) from the in-situ buoy records for each of the selected coastal locations (Supplementary Table 2). Standard error metrics were used to evaluate the hourly time series, namely the Pearson's correlation coefficient (R), scatter index (SI), root mean square error (RMSE) and relative bias (bias):

$$R = \frac{\sum_{i=1}^{N}(M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{N}(M_i - \bar{M})^2 \sum_{i=1}^{N}(O_i - \bar{O})^2}} \qquad (5)$$

$$SI = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}[(O_i - \bar{O}) - (M_i - \bar{M})]^2}}{\frac{1}{N}\sum_{i=1}^{N}O_i} \qquad (6)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2} \qquad (7)$$

$$bias = \frac{\sum_{i=1}^{N}(M_i - O_i)}{\sum_{i=1}^{N}O_i} * 100 \qquad (8)$$

Where O_i represents the observations, M_i the modelled data from the wave reanalysis and N is the number of measurements. The overbar refers to the mean. A negative (positive) bias represents an underestimation (overestimation) from the wave reanalysis. For the wave directions circular R, SI, RMSE and bias were estimated taking into consideration the fact that when comparing two angular variables in the sector from −90° to 90° the values ranging from 270° to 360° need to be corrected to −90° and 0°.

Calibration

Calibration for individual buoy locations. To calibrate the wave reanalysis (y) against the buoy observations (x) four transfer functions were tested to better represent the behaviour of wave data from buoys with different locations around the world. The functions were applied to H_s, T_m and T_p of each buoy. The selection of the best fit function was performed automatically by finding the fit that optimizes both the standard error metrics for the entire time series and for the

extremes (i.e. the values above the 95th percentile). This was done by estimating the improvement obtained with each fit (in terms of R, RMSE, SI, mean absolute difference, and bias) and finding the fit that would optimise the most parameters (Supplementary Table 3). The transfer functions used were a first-degree polynomial without and with intercept (Eqs. (9) and (10)), a power function (Eq. (11)), and a rotation around the mean (Eq. (12)), from which Eq. (13) is derived:

$$y' = ay, \tag{9}$$

$$y' = ay + b, \tag{10}$$

$$y' = ay^b, \tag{11}$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(1-a) & -\sin(1-a) \\ \sin(1-a) & \cos(1-a) \end{bmatrix} \begin{bmatrix} x - x_c \\ y - y_c \end{bmatrix}, \tag{12}$$

$$y' = \sin\beta(x - x_c) + \cos\beta(y - y_c), \text{ with } \beta = (1-a) \tag{13}$$

Linear regression methods, such as Eqs. (9) and (10), have been widely used to correct modelled wave data against observations, and have been applied to the full H_s time series[36], to the extreme H_s[79] or to correct the H_s associated with specific return periods[12,47,51]. Here, the equations are applied to the entire original reanalysis wave time series from WAVERYS and ERA5 to obtain a calibrated time series (y') and as a result, the bias is corrected. The power function has also been used in previous studies[36,48,50], with the correction parameters a and b varying with the direction of wave propagation. In this study, the parameters were instead constant variables for the entire time series. The rotation around the mean (x_c, y_c) was, to the authors' knowledge, applied here for the first time in wave data calibration. For this purpose, a rotation matrix was used (Eq. (12)), which depends on the angular coefficient found through the first-degree polynomial fit with intercept. This method was deemed appropriate to correct both the underestimation of the higher H_s values from the reanalysis and the overestimation of the lower H_s values. The bias is not modified. The rotation matrix in Eq. (12) refers to a counterclockwise rotation, which corresponds to cases where the model underestimates the H_s compared to the observations. In case of overestimation, a clockwise rotation matrix is required, which can be obtained by changing the sign of both sines in Eq. (12).

For calibrating the wave directions, a different method was implemented. First, the wave roses from the buoys and from the reanalysis were compared visually. If differences between the observed and modelled wave direction distributions were observed, a calibration was performed. This was done by dividing wave direction data into 2, 3 or 4 sectors depending on the number of peaks present in the distribution (e.g., for the Belmullet buoy in Fig. 1m, n two sectors were identified: from 0° to 180° and from 180° to 360°). For each sector, the difference between the model and buoy averaged directions was determined and subtracted/added to the reanalysis data, according to the need to rotate each sector in a counterclockwise/clockwise direction.

Global calibration. A global calibration was implemented to improve the wave reanalyses time series where these are known to underestimate H_s and T_m, particularly above the 95th percentile of the wave height and period distributions. Underestimation of the extreme bias in the reanalysis datasets (Fig. 2a, b) occurs in 94% (78%) of buoy locations for WAVERYS (ERA5), while overestimation was observed in 6% (22%) of buoy locations for WAVERYS (ERA5). This smaller subset of buoys where reanalyses overestimate the extreme bias was excluded from the global calibration, as these locations correspond to areas of complex coastal configuration, poorly resolved in global wave reanalyses models. Such locations are found mainly in areas sheltered by islands (e.g., Hawaii, Azores, Sardinia), headlands (e.g., Cape Cod) or reefs (e.g., Great Barrier Reef), as well as in sections of enclosed basins (e.g., parts of the Baltic and North Sea), as global reanalysis models have known limitations in resolving sheltering effects[35].

From the buoy datasets identified for the global calibration, 70% were randomly selected and used to derive the equations for the global calibration while the remaining 30% were used to independently validate the approach. A sensitivity analysis (Supplementary Figs. 12 and 13) was performed by selecting different percentages of buoys for the derivation of the equations and for the validation, as well as different random groups, which did not result in observable differences in the error statistics. After combining the data from the randomly selected buoys and calibrating ERA5 and WAVERYS reanalysis for those locations as a unique time series, Eqs. (9)–(11) were applied to find the best-fit calibration equations to be used globally to correct the underestimation of the H_s and T_m values above the 95th percentile. The equations found were then validated against the data from the remaining 30% of buoys.

Calculation of return periods. The calculation of the return periods of H_s was based on the Peak-Over Threshold method[87] and a de-clustering algorithm to characterize extreme H_s events[88], following the approach proposed by Oikonomou et al. (2020)[89]. Independently and identically distributed extreme value datasets were obtained by applying a threshold based on the 95th percentile value[90] of the H_s distribution and an independence criterion for separating consecutive events based on the extremal index calculation. Each extreme value dataset was then fitted

to the Generalized Pareto Distribution to determine the return periods of an extreme occurrence in H_s. This was performed for each location using the wave buoy data (for buoys with more than 12 years of complete data), the original, individually calibrated and globally calibrated reanalysis data.

References

1. Coco, G. & Ciavola, P. *Coastal storms: processes and impacts.* (John Wiley & Sons, 2017).
2. Meucci, A., Young, I. R., Hemer, M., Kirezci, E. & Ranasinghe, R. Projected 21st century changes in extreme wind-wave events. *Sci. Adv.* **6**, eaaz7295 (2020).
3. Lobeto, H., Menendez, M. & Losada, I. J. Future behavior of wind wave extremes due to climate change. *Sci. Rep.* **11**, 1–12 (2021).
4. Vitousek, S. et al. Doubling of coastal flooding frequency within decades due to sea-level rise. *Sci. Rep.* **7**, 1–9 (2017).
5. Neumann, B., Vafeidis, A. T., Zimmermann, J. & Nicholls, R. J. Future coastal population growth and exposure to sea-level rise and coastal flooding - a global assessment. *PLoS One* **10**, e0118571 (2015).
6. van Dongeren, A. et al. Introduction to RISC-KIT: Resilience-increasing strategies for coasts. *Coast. Eng.* **134**, 2–9 (2018).
7. Wesselman, D. et al. The effect of tides and storms on the sediment transport across a Dutch barrier island. *Earth Surf. Process. Landforms* **43**, 579–592 (2018).
8. Morton, R. A. & Sallenger, A. H. Morphological impacts of extreme storms on sandy beaches and barriers. *J. Coast. Res.* **19**, 560–573 (2003).
9. Houser, C., Hapke, C. & Hamilton, S. Controls on coastal dune morphology, shoreline erosion and barrier island response to extreme storms. *Geomorphology* **100**, 223–240 (2008).
10. Leaman, C. K. et al. A Storm Hazard Matrix combining coastal flooding and beach erosion. *Coast. Eng.* **170**, 104001 (2021).
11. Guillou, N., Lavidas, G. & Chapalain, G. Wave energy resource assessment for exploitation-A review. *J. Mar. Sci. Eng.* **8**, 705 (2020).
12. Neary, V. S. et al. Characterization of extreme wave conditions for wave energy converter design and project risk assessment. *J. Mar. Sci. Eng.* **8**, 1–19 (2020).
13. Ferreira, O. et al. Storm-induced risk assessment: evaluation of two tools at the regional and hotspot scale. *Coast. Eng.* **134**, 241–253 (2018).
14. Data Buoy Cooperation Panel. *Global data buoy observations, a DBCP implementation strategy* (Data Buoy Cooperation Panel, 2018).
15. Gemmrich, J., Thomas, B. & Bouchard, R. Observational changes and trends in northeast Pacific wave records. *Geophys. Res. Lett.* **38**, 1–5 (2011).

16. Janssen, P. A. E. M., Abdalla, S., Hersbach, H. & Bidlot, J. R. Error estimation of buoy, satellite, and model wave height data. *J. Atmos. Ocean. Technol* **24**, 1665–1677 (2007).

17. Vignudelli, S. et al. Satellite altimetry measurements of sea level in the coastal zone. *Surv. Geophys.* **40**, 1319–1349 (2019).

18. Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C. J. H. & Ward, P. J. A global reanalysis of storm surges and extreme sea levels. *Nat. Commun.* **7**, 11969 (2016).

19. Muis, S. et al. A high-resolution global dataset of extreme sea levels, tides, and storm surges, including future projections. *Front. Mar. Sci.* **7**, 1–15 (2020).

20. Almar, R. et al. A global analysis of extreme coastal water levels with implications for potential coastal overtopping. *Nat. Commun.* **12**, 1–9 (2021).

21. Vousdoukas, M. I. et al. Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard. *Nat. Commun.* **9**, 1–12 (2018).

22. Uppala, S. M. et al. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**, 2961–3012 (2005).

23. Dee, D. P. et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).

24. Chawla, A., Spindler, D. M. & Tolman, H. L. Validation of a thirty year wave hindcast using the Climate Forecast System Reanalysis winds. *Ocean Model* **70**, 189–206 (2013).

25. Perez, J., Menendez, M. & Losada, I. J. GOW2: A global wave hindcast for coastal applications. *Coast. Eng.* **124**, 1–11 (2017).

26. Hersbach, H. et al. ERA5 hourly data on single levels from 1979 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, https://doi.org/10.24381/cds.adbb2d47 (2018).

27. Law-Chune, S. et al. WAVERYS: a CMEMS global wave reanalysis during the altimetry period. *Ocean Dyn.* **71**, 357–378 (2021).

28. Hoffmann, L. et al. From ERA-Interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmos. Chem. Phys.* **19**, 3097–3214 (2019).

29. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).

30. Sharmar, V. & Markina, M. Validation of global wind wave hindcasts using ERA5, MERRA2, ERA-Interim and CFSRv2 reanalyzes. *IOP Conf. Ser. Earth Environ. Sci.* **606**, 012056 (2020).

31. Hersbach, H. et al. Global reanalysis: goodbye ERA-Interim, hello ERA5. *ECMWF Newsletter.* 17–24, https://doi.org/10.21957/vf291hehd7 (2019).

32. Wang, J. & Wang, Y. Evaluation of the ERA5 Significant Wave Height against NDBC Buoy Data from 1979 to 2019. *Mar. Geod.* 1–15, https://doi.org/10.1080/01490419.2021.2011502 (2021).

33. Feng, H, Vandemark, D., Quilfen, Y., Chapron, B. & Beckley, B. Assessment of wind-forcing impact on a global wind-wave model using the TOPEX altimeter. *Ocean Eng.* **33**, 1431–1461 (2006).

34. Cavaleri, L. et al. Wave modelling - The state of the art. *Prog. Oceanogr.* **75**, 603–674 (2007).

35. Cavaleri, L. et al. Wave modelling in coastal and inner seas. *Prog. Oceanogr.* **167**, 164–233 (2018).

36. Lemos, G. et al. On the need of bias correction methods for wave climate projections. *Glob. Planet. Change* **186**, 103109 (2020).

37. Bidlot, J. R., Holmes, D. J., Wittmann, P. A., Lalbeharry, R. & Chen, H. S. Intercomparison of the performance of operational ocean wave forecasting systems with buoy data. *Weather Forecast* **17**, 287–310 (2002).

38. Stopa, J. E. & Cheung, K. F. Intercomparison of wind and wave data from the ECMWF Reanalysis Interim and the NCEP Climate Forecast System Reanalysis. *Ocean Model* **75**, 65–83 (2014).

39. Baordo, F., Clementi, E., Iovino, D. & Masina, S. *Intercomparison and assessement of wave models at global scale.* CMCC Research Paper (CMCC, 2020).

40. Shi, H. et al. Evaluating the accuracy of ERA5 wave reanalysis in the water around China. *J. Ocean Univ. China* **20**, 1–9 (2021).

41. Kumar, V. S. & Naseef, T. M. Performance of ERA-interim wave data in the nearshore waters around India. *J. Atmos. Ocean. Technol.* **32**, 1257–1269 (2015).

42. Hisaki, Y. Intercomparison of assimilated coastal wave data in the Northwestern Pacific Area. *J. Mar. Sci. Eng.* **8**, 579 (2020).

43. Aguiar, D. F., McInnes, K., Filho, E. M., Romeu, M. A. R. & Fontoura, J. A. S. Extreme wave analysis based on 31 years data from WW3 model: study off Southern Brazilian Coast. *An. Acad. Bras. Cienc.* **93**, 1–17 (2021).

44. Haerter, J. O., Hagemann, S., Moseley, C. & Piani, C. Climate model bias correction and the role of timescales. *Hydrol. Earth Syst. Sci.* **15**, 1065–1079 (2011).

45. Maraun, D. Bias correcting climate change simulations - a critical review. *Curr. Clim. Chang. Rep.* **2**, 211–220 (2016).

46. Caires, S. & Sterl, A. A new nonparametric method to correct model data: application to significant wave height from the ERA-40 re-analysis. *J. Atmos. Ocean. Technol.* **22**, 443–459 (2005).

47. Caires, S. & Sterl, A. 100-year return value estimates for ocean wind speed and significant wave height from the ERA40 data. *J. Clim.* **18**, 1032–1048 (2005).

48. Mínguez, R., Espejo, A., Tomás, A., Méndez, F. J. & Losada, I. J. Directional calibration of wave reanalysis databases using instrumental data. *J. Atmos. Ocean. Technol* **28**, 1466–1485 (2011).

49. Tomas, A., Mendez, F. J. & Losada, I. J. A method for spatial calibration of wave hindcast data bases. *Cont. Shelf Res.* **28**, 391–398 (2008).

50. Espejo, A. et al. Directional calibrated wind and wave reanalysis databases using instrumental data for optimal design of off-shore wind farms. In *OCEANS 2011 IEEE - Spain* 1–9, https://doi.org/10.1109/Oceans-Spain.2011.6003592 (2011).

51. Sterl, A. & Caires, S. Climatology, variability and extrema of ocean waves: the web-based KNMI/ERA-40 wave atlas. *Int. J. Climatol.* **25**, 963–977 (2005).

52. Bellotti, G., Franco, L. & Cecioni, C. Regional downscaling of copernicus ERA5 wave data for coastal engineering activities and operational coastal services. *Water* **13**, 859 (2021).

53. Vannucchi, V., Taddei, S., Capecchi, V., Bendoni, M. & Brandini, C. Dynamical downscaling of era5 data on the north-western mediterranean sea: From atmosphere to high-resolution coastal wave climate. *J. Mar. Sci. Eng.* **9**, 1–29 (2021).

54. Bertotti, L., Cavaleri, L., Soret, A. & Tolosana-Delgado, R. Performance of global and regional nested meteorological models. *Cont. Shelf Res.* **87**, 17–27 (2014).

55. Fernández-Montblanc, T., Vousdoukas, M. I., Mentaschi, L. & Ciavola, P. A Pan-European high resolution storm surge hindcast. *Environ. Int.* **135**, 105367 (2020).

56. Kirezci, E. et al. Projections of global-scale extreme sea levels and resulting episodic coastal flooding over the 21st Century. *Sci. Rep.* **10**, 1–12 (2020).

57. Marcos, M. et al. Increased extreme coastal water levels due to the combined action of storm surges and wind waves. *Geophys. Res. Lett.* **46**, 4356–4364 (2019).

58. Mentaschi, L., Vousdoukas, M. I., Voukouvalas, E., Dosio, A. & Feyen, L. Global changes of extreme coastal wave energy fluxes triggered by intensified teleconnection patterns. *Geophys. Res. Lett.* **44**, 2416–2426 (2017).

59. Nienhuis, J. H. et al. Global-scale human impact on delta morphology has led to net land area gain. *Nature* **577**, 514–518 (2020).

60. Vousdoukas, M. et al. Developments in large-scale coastal flood hazard mapping. *Nat. Hazards Earth Syst. Sci. Discuss.* **0**, 1–24 (2016).

61. Vousdoukas, M. I., Voukouvalas, E., Annunziato, A., Giardino, A. & Feyen, L. Projections of extreme storm surge levels along Europe. *Clim. Dyn.* **47**, 3171–3190 (2016).

62. Vousdoukas, M. I. et al. Sandy coastlines under threat of erosion. *Nat. Clim. Chang.* **10**, 260–263 (2020).

63. Cavaleri, L. Wave modeling-missing the peaks. *J. Phys. Oceanogr.* **39**, 2757–2778 (2009).

64. Campos, R. M. & Soares, C. G. Comparison and assessment of three wave hindcasts in the North Atlantic Ocean. *J. Oper. Oceanogr.* **9**, 26–44 (2016).

65. Morim, J. et al. Understanding uncertainties in contemporary and future extreme wave events for broad-scale impact and adaptation planning. *Sci. Adv.* **9**, 1–13 (2023).

66. Stefanakos, C. Intercomparison of wave reanalysis based on ERA5 and WW3 databases. *Proc. Int. Offshore Polar Eng. Conf.* **3**, 2506–2512 (2019).

67. Crespo, N. M. et al. Western South Atlantic Climate Experiment (WeSACEx): extreme winds and waves over the Southeastern Brazilian sedimentary basins. *Clim. Dyn.* **60**, 571–588 (2022).

68. Kardakaris, K., Boufidi, I. & Soukissian, T. Offshore wind and wave energy complementarity in the greek seas based on ERA5 data. *Atmosphere* **12**, 1–17 (2021).

69. Bruno, M. F., Molfetta, M. G., Totaro, V. & Mossa, M. Performance assessment of ERA5 wave data in a swell dominated region. *J. Mar. Sci. Eng.* **8**, 214 (2020).

70. Haripriya, R., Misra, A., Jeyaraj, S. & Ramakrishnan, B. Wave energy assessment for 39 years along the Exclusive Economic Zone (EEZ) of India. *Reg. Stud. Mar. Sci.* **45**, 101809 (2021).

71. Kumar, V. S., George, J. & Joseph, D. Hourly maximum individual wave height in the Indian shelf seas—its spatial and temporal variations in the recent 40 years. *Ocean Dyn.* **70**, 1283–1302 (2020).

72. Muhammed Naseef, T. & Sanil Kumar, V. Climatology and trends of the Indian Ocean surface waves based on 39-year long ERA5 reanalysis data. *Int. J. Climatol.* **40**, 979–1006 (2020).

73. Harley, M. D. et al. Extreme coastal erosion enhanced by anomalous extratropical storm wave direction. *Sci. Rep.* **7**, 1–9 (2017).

74. Mortlock, T. R., Goodwin, I. D., McAneney, J. K. & Roche, K. The June 2016 Australian East Coast Low: Importance of wave direction for coastal erosion assessment. *Water* **9**, 1–22 (2017).

75. Dentale, F. et al. Extreme wave analysis by integrating model and wave buoy data. *Water* **10**, 1–15 (2018).

76. Bergsma, E. W. J., Almar, R., Anthony, E. J., Garlan, T. & Kestenare, E. Waves variability along the world's continental shelves and coasts: monitoring opportunities from satellite Earth observation. *Adv. Sp. Res.* **69**, 3236–3244 (2022).

77. Sanil Kumar, V. & Anoop, T. R. Wave energy resource assessment for the Indian shelf seas. *Renew. Energy* **76**, 212–219 (2015).
78. Hemer, M. A. et al. A revised assessment of Australia's national wave energy resource. *Renew. Energy* **114**, 85–107 (2017).
79. Stephens, S. A. & Gorman, R. M. Extreme wave predictions around New Zealand from hindcast data. *New Zeal. J. Mar. Freshw. Res.* **40**, 399–411 (2006).
80. Group, T. W. The WAM Model - a third generation ocean wave prediction model. *J. Phys. Oceanogr.* **18**, 1775–1810 (1988).
81. Lellouche, J. M. et al. Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time 1/12° high-resolution system. *Ocean Sci.* **14**, 1093–1126 (2018).
82. Caires, S., Sterl, A., Bidlot, J. R., Graham, N. & Swail, V. Intercomparison of different wind-wave reanalyses. *J. Clim.* **17**, 1893–1913 (2004).
83. Caires, S. & Sterl, A. Validation of ocean wind and wave data using triple collocation. *J. Geophys. Res. Ocean.* **108**, 1–16 (2003).
84. Copernicus Marine in situ TAC. *Copernicus Marine In Situ - Global Ocean Wave Observations Reanalysis* (SEANOE, 2022) https://doi.org/10.17882/70345.
85. Loic, P. D. L. V., Marta de A., Fernando, M.-M. & Sylvie, P. *Product user manual. For REProcessed IN SITU product (WAVES).* CMEMS-INS-PUM-013-045. https://doi.org/10.13155/54853 (2021).
86. Bushnell. *Manual for real-time quality control of in-situ surface wave data: a guide to quality control and quality assurance of in-situ surface wave observations version 2.1.* https://repository.library.noaa.gov/view/noaa/19817 (2019) https://doi.org/10.25923/7yc5-vs69.
87. Ferreira, J. A. & Guedes-Soares, C. An application of the peaks over threshold method to predict extremes of significant wave height. *J. Offshore Mech. Arct. Eng* **120**, 165–176 (1998).
88. Ferro, C. A. T. & Segers, J. Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65**, 545–556 (2003).
89. Oikonomou, C. L. G., Gradowski, M., Kalogeri, C. & Sarmento, A. J. N. A. On defining storm intervals: extreme wave analysis using extremal index inferencing of the run length parameter. *Ocean Eng.* **217**, 107988 (2020).
90. Harley, M. Coastal storm definition. In *Coastal storms: processes and impacts* 1–21 (John Wiley & Sons, 2017).

Acknowledgements

Author contributions

V.F. retrieved data, performed the statistical analysis, and wrote the manuscript with support from O.F. and C.L. V.K. contributed to developing Matlab scripts. All authors have contributed to the final manuscript and agreed to the published version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-023-00819-0.

Correspondence and requests for materials should be addressed to Valeria Fanti.

Peer review information *Communications Earth & Environment* thanks Jichao Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Joe Aslin. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.