Applied Ocean Research 165 (2025) 104818

Contents lists available at ScienceDirect

Applied Ocean Research

journal homepage: www.elsevier.com/locate/apor

Resource assessment uncertainty reduction via bias correction: On the temporal and spatial sensitivity analysis

Francesco Callea [a,*], Ander Martinez-Perurena [b,*], Giuseppe Giorgi [a], Markel Penalba [b,c], Gregorio Iglesias [d,e]

[a] Marine Offshore Renewable Energy Lab, Politecnico di Torino, 10129 Torino, Italy
[b] Fluid Mechanics Department, Mondragon University, Loramendi 4, 20500 Arrasate, Spain
[c] Ikerbasque, Basque Foundation for Science, Euskadi Plaza 5, Bilbao, Spain
[d] MaREI, Environmental Research Institute & School of Engineering, UCC, Cork, Ireland
[e] University of Plymouth, School of Engineering, Computing & Mathematics, Plymouth, UK

ARTICLE INFO

Dataset link: https://portus.puertos.es, https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels

Keywords:
Offshore Renewable Energy
Resource assessment
Wave data uncertainties
Bias correction
Wave direction
Sensitivity analysis

ABSTRACT

The present study addresses the necessity of a wide and reliable knowledge of metocean conditions for the design and operation of Offshore Renewable Energy (ORE) installations. Statistical bias correction techniques, including a directionally-sensitive technique, are implemented and tested in several locations around the Spanish coast, using information coming from *in-situ* measurements, with the objective to reduce the uncertainty on wave statistics of re-analysis datasets. In addition, the most effective techniques (*i.e.* EGQM and DAGQM) are taken as a reference to carry out a temporal and spatial sensitivity analysis, aimed at extending the validity of the aforementioned calibration techniques both in time and space. Encouraging results are obtained concerning temporal sensitivity, showing that a period between 4 and 5 years is the ideal for effectively calibrating several years of re-analysis data, and that a window of just 3 months could be enough to achieve a good bias correction. On the other hand, spatial sensitivity analysis does not highlight a trend as clear as the previous one. Uncertainty reduction after bias correction is limited to the buoys surroundings, and no critical radius could be obtained. Due to the multitude of relevant local geographical factors, such as bathymetry and coast morphology, it can be concluded that correction factors obtained in a specific site cannot always be effectively used to calibrate datasets beyond the observation point.

1. Introduction

Offshore Renewable Energy (ORE) technologies represent one of the leading solutions in the transition from fossil fuels to a more sustainable energy supply. Despite their lower level of maturity, they are expected to be among the main carriers towards the goals highlighted by the Paris Agreement (UN, 2015) and the latest report from the International Panel for Climate Change (IPCC) (IPCC, 2023). In fact, the International Energy Agency estimates that around 45% of the CO_2 emissions cut by 2050 will be due to technologies that are currently in the development stage (Stéphanie Bouckaert et al., 2021).

The effective design and operation of ORE facilities is strongly influenced by the quality of the metocean knowledge available. The lack of such data or the uncertainty that affects them can result in significant misestimation in the design stage of the installations. In fact, the bias affecting metocean data has been proven to lead to excessive conservativism in the design of both power production (*PP*) and survivability (*surv*) operating mode of ORE devices (Penalba et al., 2021). Metocean information is also fundamental for the proper estimation of the capital and operational (Konuk et al., 2023) expenditures associated to an ORE facility: an estimation error of the climate conditions at an offshore site can result in a significant increase in their costs and dimensions, preventing them from being competitive in the energy market (Haselsteiner et al., 2021).

To date, *in-situ* measurements are the most reliable approach for collecting metocean data. Despite their accuracy, field measurements tend to be expensive, prone to failures and limited both in temporal and spatial coverage. They typically operate over relatively short periods of time and provide information from one specific point in the ocean. This last issue can be addressed by re-analysis datasets and climate models, hereafter referred to as *assimilated*. They are able to cover long periods of time (some even go back to 1900 (Carreno-Madinabeitia et al., 2021)) with a very high spatial resolution. The problem with *assimilated* datasets is their uncertainty (Cervelli et al., 2022), which often requires

* Corresponding authors.
E-mail addresses: francesco.callea@polito.it (F. Callea), amartinezpe@mondragon.edu (A. Martinez-Perurena).

https://doi.org/10.1016/j.apor.2025.104818
Received 16 May 2025; Received in revised form 30 September 2025; Accepted 13 October 2025
Available online 22 October 2025

the development and application of dedicated bias correction (BC) techniques to reduce the error associated to them.

To the best of authors' knowledge, a BC techniques benchmarking was first carried out by Lemos et al. (2020). Their study focused on the significant wave height (H_s) in offshore applications, using the ERA5 database as the reference datasets. Later on, Penalba et al. (2023) presented an extension of that work in the ORE sector, calibrating the BC techniques using *in-situ* observations as a reference, and comparing the outcomes in four different locations around the Iberian peninsula. The latter introduced new statistical metrics in the BC techniques performance assessment (for instance, the *PDF-score* and the *Distribution Added Value*) and concluded that, among all those analysed in the two papers, only the methods using quantile mapping (QM) were reliable enough and, therefore, worth considering.

However, for bias correction techniques to be effectively applied, it is essential to have a reference dataset (usually derived from observation) that serves as *ground truth*. Several years of data are required for offshore resource assessment purposes. Since long and expensive measuring campaigns may not always be feasible or convenient, it is necessary to rely on data from multi-year numerical simulations, with scarce certainty about their accuracy and no observation ground-truth available for calibration. Furthermore, even when available, on-site measuring points only provide information referred to a very specific point in space. The scarce coverage and availability of observation buoys leaves wide portions of the ocean unmonitored.

The lack of reference data presents a great obstacle for the proper application of BC techniques. Nonetheless, they represent a valuable resource in reducing the uncertainty of metocean data, that benefits decision making throughout the lifecycle of ORE installations. Therefore, acknowledging their importance, the present study aims to explore the capabilities of bias correction techniques by performing a sensitivity analysis, assessing whether their validity can be extended to address two key requirements of the resource assessment process for ORE applications:

1. *Temporal sensitivity* – Using *in-situ* observations as a reference data to correct longer *assimilated* datasets from different time periods, and assessing both the optimal length of reference records and the effect of shorter or longer observational periods on the performance of the bias correction methods, in the attempt to reduce impactful measurement on-site campaigns as much as possible;
2. *Spatial sensitivity* – Relying on information from a single point in space to correct data over a wider area surrounding the observation point, analysing whether performances improve or worsen with the distance from the reference point, as well as whether any correlation emerges.

Both objectives fall within the broader topic of choosing the best set of input data for the BC techniques. The present study integrates an additional directional calibration method and carries out a temporal and spatial sensitivity analysis of bias correction techniques with respect to the reference metocean data employed. The paper is structured as follows: Section 2 introduces the bias correction techniques used in the sensitivity analysis; Section 3 describes the methodology that has been followed, with the statistical metrics for results evaluation and the procedural scheme for sensitivity analysis; Section 4 presents the case study. Section 5 shows the most important results obtained in the process; and Section 6 finally draws the most relevant conclusions out of the work.

2. Bias correction techniques

Bias correction techniques do not require detailed knowledge of the underlying phenomenon. Instead, their purpose is to adjust the values of the *assimilated* dataset so that its distribution aligns more closely with the one of the observed dataset.

In the literature, numerous studies investigate the reduction of metocean data uncertainty through various methodological approaches. Studies focusing on statistical BC techniques (Lemos et al., 2020 and Penalba et al., 2023) provide a comprehensive review of the available statistical techniques, assessing the improvement in some design parameters following application of the calibration methods. Callea et al. (2024) perform a comparison among the methods available, reporting methods based on quantile mapping as the only methods sufficiently reliable for extensive application in resource assessment. The study also dedicates particular attention to the novel direction-based method introduced by Lemos et al. (2020) accounting for wave directions and based on the solution of a non-linear regression problem. It is also the first to introduce a sensitivity analysis of the BC techniques to the time windows used for calibration.

Hence, the analysis carried out in this paper is based on the Gumbel-based Quantile Mapping (referred to as EGQM) and the directionally adjusted Gumbel Quantile Mapping (from here on DAGQM) bias correction techniques, which are introduced in the following. It shall be specified that none of the two methods correct data on a one-to-one basis; instead, they aim to adjust the dataset distribution, tackling systematic bias through a statistical approach.

2.1. Gumbel-based Quantile Mapping

The Quantile Mapping method consists of partitioning the two datasets into quantile ranges and, for each of them, computing a correction factor $X^{QM}(q_j)$ based on the Cumulative Density Function (CDF). The quantile placement is determined according to a Gumbel distribution (GDF) (Gumbel, 1935):

$$F(x;\mu,\beta) = e^{-e^{-(x-\mu)/\beta}} \tag{1}$$

where μ and β are, respectively, the location and scale parameters. In this formulation, the GDF assigns a denser placement of quantiles in the upper tail of the distribution (with more than 50% of quantiles located above the 99th percentile), thereby enhancing the representation of extreme values and seizing the nature of the most severe climate phenomena.

In line with the framework outlined and verified in Penalba et al. (2023), 50 quantile levels are defined from 0.01 to 0.99 (both included), and the correction factor is calculated as:

$$X^{GQM}(q_j) = CDF_{as}^{-1}(q_j) - CDF_{obs}^{-1}(q_j) \tag{2}$$

where the subscript "*as*" refers to the *assimilated* dataset, while "*obs*" indicates the observation data. The difference between the inverse CDFs is then fitted with an *n-th* order polynomial function, and subsequently used to correct the dataset as follows:

$$y^{BC}(q_j) = y^{as}(q_j) + f(X^{GQM}(q_j), n) \tag{3}$$

where y is the metocean variable, and y^{BC} and y^{as} refer to the bias corrected and the *assimilated* datasets, respectively.

2.2. Directionally adjusted Gumbel Quantile Mapping

Following the suggestion that wind/wave characteristics can show strong variations depending on the provenience direction (Semedo et al., 2011, 2018), the BC technique was adjusted to account for the information on wind/wave direction (Mínguez et al., 2011). This is expected to be particularly beneficial, as it allows to separate wave phenomena basing on their provenience, better targeting each wave group's peculiar features. The wave rose is divided into equal sectors with a width α and analyse them one at a time, each sector presenting a shift β from the previous one. Every sector is divided into 50 quantile ranges: for each of them, a different couple of correction factor $a^{DAGQM}(\theta_i, q_j)$ and $b^{DAGQM}(\theta_i, q_j)$ is identified and used to correct the correspondent data. Depending on the sectors' width and the shift, sectors may partially overlap, causing metocean data to be included

in more than one sector. In that case, the data is separately calibrated for every sector it is included in: the result showing the best agreement with the observed reference is chosen, ensuring the best calibration is used.

The DAGQM method, introduced by Lemos et al. (2020), maintains the principle of dividing the data available into quantiles, but instead of a polynomial function, it is based on the solution of a non-linear regression problem. As a result of it, the correction parameters vary along the different wind/wave directions following a cubic spline.

$$y^{BC} = a^{DAGQM}(\theta_i, q_j)(y^{as})^{b^{DAGQM}(\theta_i, q_j)} \tag{4}$$

Where $a^{DAGQM}(\theta_i, q_j)$ and $b^{DAGQM}(\theta_i, q_j)$ are the correction spline parameters, expressed as a function of the wind/wave direction θ_i, and of the quantile range q_j the analysed metocean value belongs to. They are obtained so that the solution of the optimization problem (here, the sum of the square residuals) returns the calibrated values of the variable y:

$$min_{a,b}[\sum(y_i^{obs} - y_i^{BC})^2] =$$
$$min_{a,b} \sum[y_i^{obs} - a^{DAGQM}(\theta_i, q_j)(y_i^{as})^{b^{DAGQM}(\theta_i, q_j)}]^2,$$
$$i = 1, \dots, S \tag{5}$$

where S is the number of sectors.

3. Methodology

The present study is part of a broader research, whose preliminary results were presented in Callea et al. (2024). The effort stems from the necessity to correct long metocean datasets, derived by re-analysis models, using the information available from *in-situ* measurements. Given the time and cost required by such campaigns, the potential to derive correction factors from relatively short *observation* dataset to reduce the bias on a target *assimilated* dataset is explored. The goal is to optimize the identification period, seeking an ideal duration that minimizes measurement campaigns' time and cost. To address the issue related to the limited spatial coverage of *observational* metocean data, the study is further extended in the present. A sensitivity analysis is carried out, using the correction factors obtained in correspondence of the measurement buoy to calibrate re-analysis data at grid-points in the buoy surroundings. In particular, Section 3.1 introduces the statistical metrics used to rate performances in each case study; Sections 3.2 and 3.3 clarify the methods followed to carry out the temporal and spatial sensitivity analysis, respectively. The analysis focuses exclusively on significant wave height (H_s), making use of the information on wave direction as well; wind speed was not considered.

3.1. Statistical metrics

In order to allow a clear assessment and comparison among case studies, some statistical indicators, hereby presented, have been used.

The first and most important is the *bias*, defined as the difference between the reference ground truth (y^{obs}) and the analysed data, either y^{as} or y^{BC}:

$$bias(y^{obs}, y^{as} \| y^{BC}) = y^{obs} - y^{as} \| y^{BC} \tag{6}$$

The dispersion of the bias distribution is also computed. The chosen metric is the Interquantile Range (IQR), defined in (7) as the difference between the 75th and the 25th quantiles of the bias distribution:

$$IQR = q_{75}(bias(y^{obs}, y^{as} \parallel y^{BC})) - q_{25}(bias(y^{obs}, y^{as} \parallel y^{BC})) \tag{7}$$

A higher dispersion may suggest that the BC technique corrects some sea-states better than others. Finally, to facilitate the comparison and provide a unique value for each case, the mean absolute bias (MAB from here on) is computed:

$$MAB(y^{obs}, y^{as} \| y^{BC}) = mean(|y^{obs} - y^{as} \| y^{BC}|) \tag{8}$$

The use of absolute value prevents the comparison from being affected by systematic error generated from averaging positive and negative bias values, which could falsely produce a null bias.

Although these are simple to obtain and interpret, Penalba et al. (2023) introduces two ORE-specific and statistics-oriented metrics. The first is the PDF_{score}, which measures the similarity between two PDFs, for a complete comparison across the whole distribution of the bias (Perkins et al., 2007). It is computed as the sum, over all bins of the distribution, of the minimum value between the two normalized PDFs evaluated at each bin:

$$PDF_{score} = \sum_{m=1}^{M} min(PDF(y_m^{obs}), PDF(y_m^{as})), \tag{9}$$

Where $PDF(y_m^{obs})$ and $PDF(y_m^{as})$ are the two normalized PDFs at the mth bin, and M is the number of bins used to represent the PDF.

The second metric, the *Distribution Added Value* (DAV) performs a direct and normalized comparison between two PDF_{score} (Soares and Cardoso, 2018):

$$DAV = \frac{PDF_{score,BC} - PDF_{score,as}}{PDF_{score,as}} \times 100, \tag{10}$$

where $PDF_{score,BC}$ and $PDF_{score,as}$ represent the PDF_{score} of the corrected and raw *assimilated* datasets, respectively. It is particularly useful as it quantifies, in relative terms, the improvement achieved by each bias correction technique.

3.2. Temporal sensitivity framework

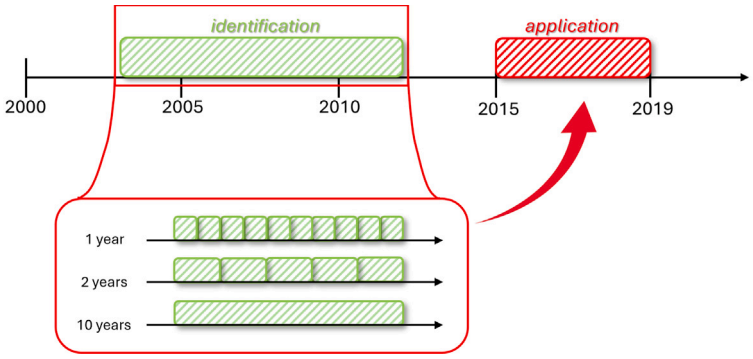
The potential of using bias correction methods by deriving correction factors from an *identification period* and subsequently applying them to reduce the uncertainty of an *application* period has already been successfully evaluated in Callea et al. (2024).

This article investigates the sensitivity of BC techniques to the time window used for the identification of the correction factors. As represented in Fig. 1, an *identification window* spanning between 2003 and 2012 is taken (Fig. 1(a)): this is the "oldest" time window where 10 years of data are available for all sites considered. The period is divided in several ways, by varying the length of the intervals and shifting them 1 year at a time, generating multiple identification windows as shown in Fig. 1(b). Correction factors are separately obtained from each of them and then applied to correct *assimilated* data from an *application* window, which includes data covering from 2015 to 2019. Each correction is carried out individually and the results are later compared. The goal is to assess how the number of *Identification Years* (I.Y.) affects the performance of the BC techniques, evaluating if, within each number of I.Y. used, the selection of a specific identification window over another one of the same length can lead to different results. The analysis is carried out using both techniques described in Section 2. For the DAGQM method, a sector width α of 22.5° is used, with a shift β of 1° at a time. The outcomes from both techniques are compared, with the aim of identifying an established method for metocean data calibration.

Furthermore, in the hope of reducing the burden of a long and expensive *in-situ* measuring campaign, the possibility to reduce the length of the identification period below a year is evaluated. Comprehensively, a cross-comparison among seasons was conducted in order to investigate potential differences in the performance of BC techniques. For this second purpose, four periods within the year 2012 are employed for calibration, while the years 2013–2023 served as validation window.

3.3. Spatial sensitivity framework

The second part of the analysis focuses on the spatial extent of the validity of bias correction techniques, taking an observation point as



(a) Timeline illustrating calibration and validation periods.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1-year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
2-years		2003-04	2004-05	2005-05	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12
3-years			2003-05	2004-06	2005-07	2006-08	2007-09	2008-10	2009-11	2010-12
4-years										
5-years										
6-years										
7-years										
8-years										
9-years									2003-11	2004-12
10-years										2003-12

(b) Visual representation of the sensitivity analysis framework (from (Callea et al., 2024)).

Fig. 1. Schematic representation of the temporal sensitivity analysis methodology, showing timeline and windows definition.

a reference. In this case, the BC techniques are applied to different grid-points located within a 100 km radius around the reference buoy. To reduce the computational cost while still ensuring calibration accuracy, the analysed period has been limited to the two-year window: 2018–2019 are selected as they are the most "recent" years within the same time frame used for the temporal sensitivity analysis. Here, data from measurement buoys are again used as *ground-truth* to calibrate all ERA5 grid-points by the means of the DAGQM method. Ideally, a proper result validation would require wave recordings from real *in-situ* measurement stations at each ERA5 node; since no additional buoys are available in the study areas, data from the downscaled SIMAR oceanographic model dataset (better introduced later in Section 4.2) are taken as a reference. In fact, despite being the result of numerical simulations, SIMAR dataset counts on a way higher spatial resolution than ERA5, and is able to reproduce real wave conditions with very good accuracy, including extreme H_s values (Juan et al., 2024). SIMAR data's great accuracy and nodes distribution make them a valuable resource, enabling their use as a reference for BC techniques spatial validation, despite not being an effective *in-situ* measurement (Section 4.2.1 provides further insights on SIMAR's value as a reference).

In this study, the closest SIMAR grid-point to each analysed ERA5 node is selected, as shown in Fig. 2. The bias distribution of each ERA5 point with respect to SIMAR is computed before and after calibration. It is important to stress that, although SIMAR is the assumed reference for validation, the calibration process is performed using buoy-derived correction factors. The PDF_{score} is used to compare pre a and post-calibration accuracy, assessing the spatial validity of bias correction techniques and investigating a critical radius of applicability.

4. Case study

In this section, the object of the work is presented. Section 4.1 introduces the locations analysed, while 4.2 presents the datasets used.

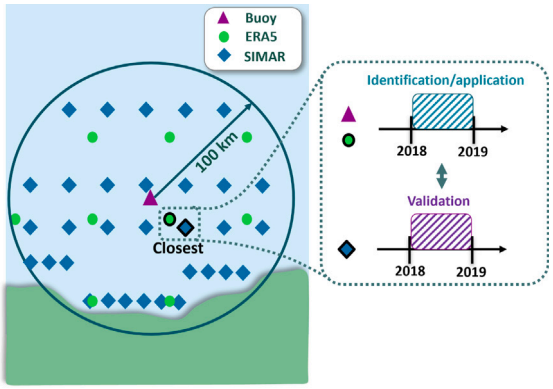
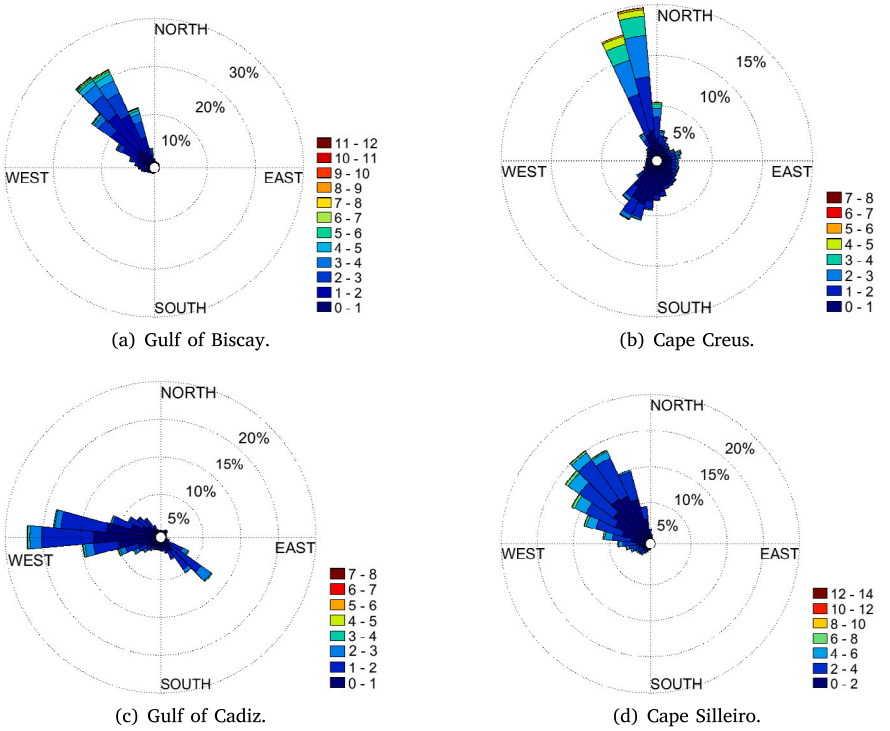


Fig. 2. Methodology for ERA5 spatial correction.

4.1. Geographical locations

As a continuation of past studies on the topic by the same authors, the analysis has been carried out referring to four different buoys located around the Iberian peninsula: namely, Gulf of Biscay, Cape Creus, Gulf of Cadiz and Cape Silleiro. Their location is shown in the map in Fig. 4, along with the associated bathymetry for each of them. All sites are located between 24 and 52 km from the coast, in order to avoid near-shore phenomena such as those related to wave reflection/refraction, tidal effects and seabed interaction.

A cross-comparison was deemed necessary to validate the contribution of wave direction as a key component for bias correction. Fig. 3 illustrates the differences among the four selected sites with respect to the wave rose, where various wave characteristics are represented. Gulf of Biscay shows a clear and narrow predominance of north-western waves, while Cape Silleiro shares the same main wave direction, but on a broader directional spectrum (both governed by Atlantic swell



(a) Gulf of Biscay. (b) Cape Creus.

(c) Gulf of Cadiz. (d) Cape Silleiro.

Fig. 3. Wave roses for the locations analysed, using all data available from the respective buoy measurements (ranging between 1995 and 2023).

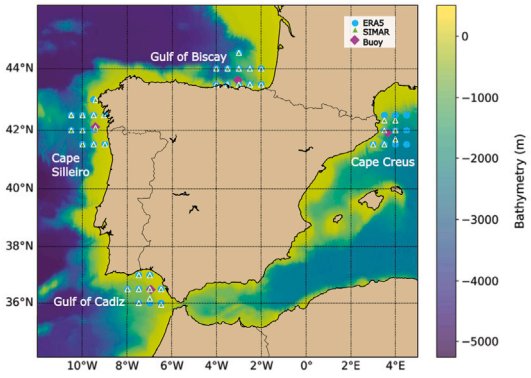


Fig. 4. Geographical distribution of buoys and the corresponding SIMAR and ERA5 grid-points.

waves). The two others, instead, display a more distributed rose, with more than one prevalent sector, as a result of the mixed contribution of swell regimes (slightly predominant in Gulf of Cadiz) and wind sea (the main component in Cape Creus). The more widely distributed the rose, the more the implementation of a directional bias correction technique is expected to have an impact, allowing to address each direction's calibration separately and more adequately.

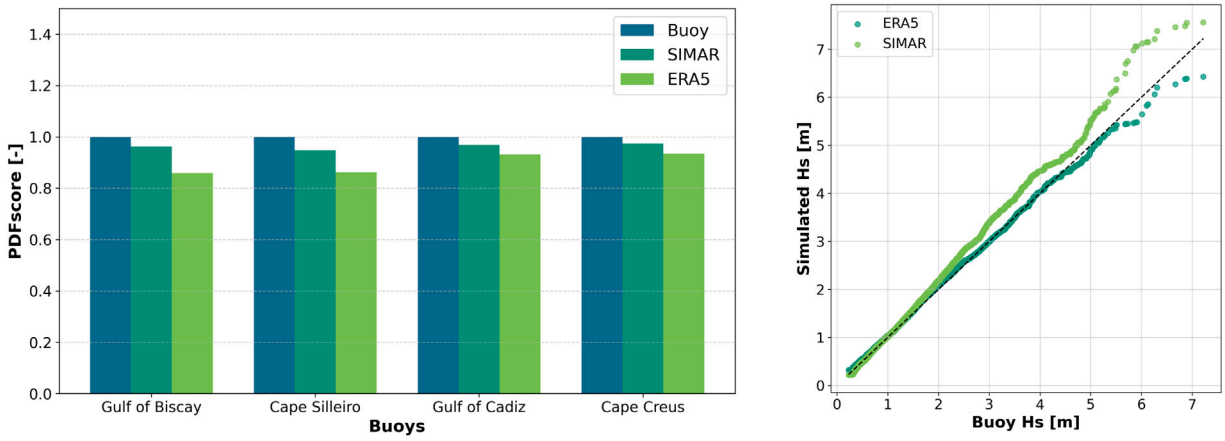
4.2. Wave data

The present analysis mainly focuses on wave-related data and, in particular, on the significant wave height (H_s). *Observation* data related to each buoy are provided by buoys operated by Spanish coastal authorities "*Puertos del Estado*" (del Estado). On the other hand, *assimilated* data for calibration are obtained from the ERA5 re-analysis dataset developed by Copernicus, providing hourly data with a spatial resolution of 0.25° (ECMWF, 2023).

For the spatial sensitivity analysis, *observation* data from specific points are not sufficient. Therefore, the information is completed with data provided by the SIMAR-WANA database. SIMAR data cover the 1958–2005 time range, and consist of simulations run by Puertos del Estado, based on regional models calibrated for the Spanish coastline. WANA data, instead, go from 2006 to present, and are obtained from the sea-state prediction system co-developed by *Puertos del Estado* and AEMET (del Estado, 2024b). The simulations are run over nested grids with increasing resolution, covering the Mediterranean and Atlantic areas around the Iberian Peninsula. The output provides hourly values with a spatial resolution ranging from 0.25° to 0.05° (del Estado, 2024b). The mesh resolution varies based on the geographic characteristics of each area and the wave model used (Puertos del Estado employs WAM, WaveWatch III and SAM models). This configuration ensures an accurate representation of the transformations experienced by waves as they approach the coastline (del Estado, 2024a), offering a better wave characterization than global scale models such as ERA5 (Fig. 2 highlights the SIMAR-WANA grid nearshore refinement).

4.2.1. SIMAR data validation

The scarce distribution of offshore measurement stations calls for the necessity of an additional data reference to properly assess BC techniques' spatial validity. In fact, the whole framework presented in Section 3.3 was primarily motivated by the low availability of measurement buoys far offshore. Indeed, the adoption of two different data references for calibration and validation leads to unavoidable inconsistency that needs to be addressed appropriately. For this reason, this section presents a prior SIMAR data quality validation, comparing them with *in-situ* observations and raw ERA5 values to assess its worth as a baseline for the calibrated ERA5 dataset. The PDF_{score} has been computed for SIMAR and ERA5 nodes corresponding to each measurement buoy shown in Fig. 4 (observations' values, equal to 1, are



(a) PDF_{score} for all datasets at the four locations.

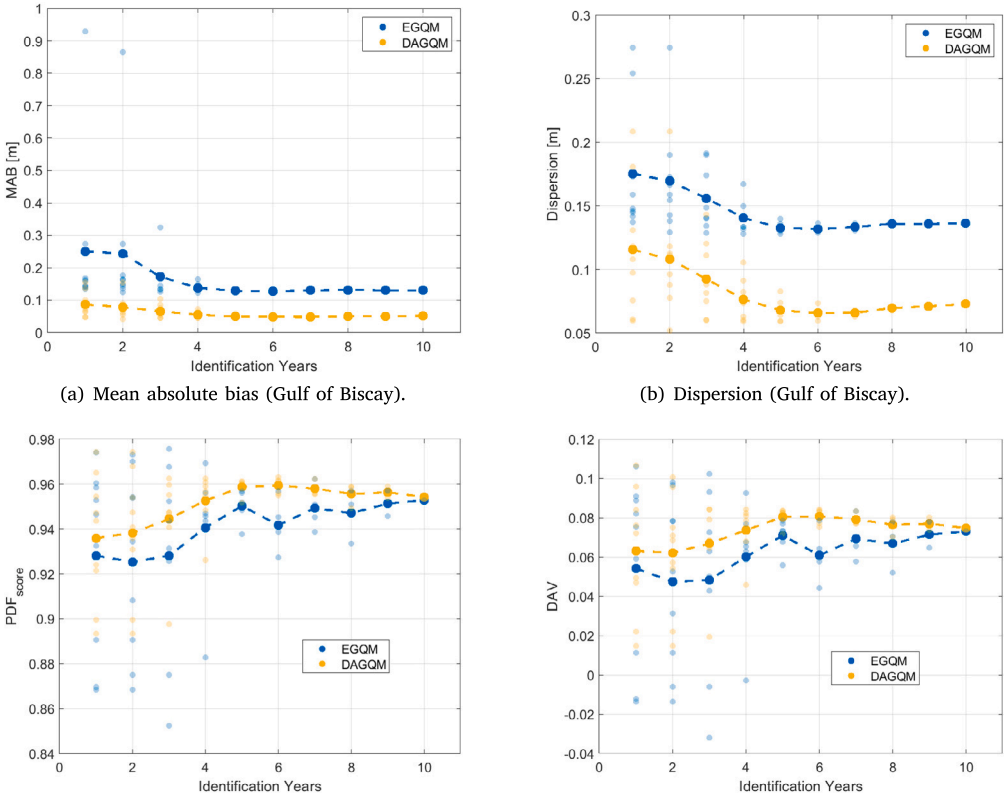
(b) Sorted scatter plot (Gulf of Cadiz).

Fig. 5. Comparison between the ERA5, the SIMAR and the calibrated datasets in correspondence of every observation buoy.

reported in the plots for reference). As reported in Fig. 5(a), SIMAR consistently matches buoy data distribution properties, with significantly higher values than ERA5 at all locations; Gulf of Cadiz is the site case where a smaller difference in PDF_{score} is observed. Fig. 5(b) focuses on the Gulf of Cadiz case, displaying the cumulative scatter plot of SIMAR and ERA5 data in correspondence of the buoy. Apparently, the location is characterized by unusually good ERA5 data quality, with H_s values adjusting well to the buoy observations, explaining the above-average PDF_{score}. Overall, the values obtained at all sites allow to reaffirm the quality of SIMAR dataset. Therefore, even though it is not an effective *in-situ* measurement, it was employed as a valuable reference for BC techniques spatial result validation.

5. Results

This section reports the main results of the study, obtained through the application of both the Gumbel Quantile Mapping (EGQM) and the directionally adjusted Gumbel Quantile Mapping (DAGQM). First, the temporal sensitivity analysis results are shown in Section 5.1. Then, Section 5.2 addresses the spatial sensitivity framework, assessing the



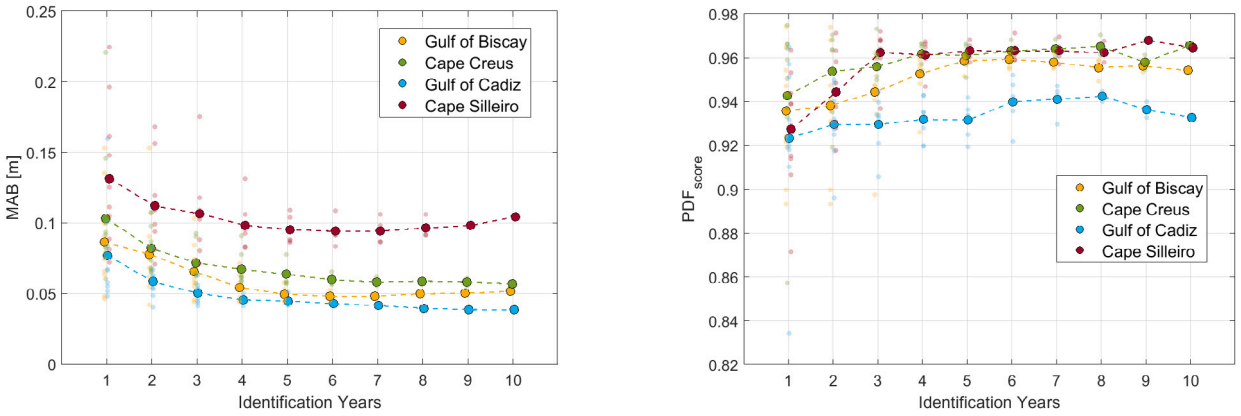
(a) Mean absolute bias (Gulf of Biscay).

(b) Dispersion (Gulf of Biscay).

(c) PDF_{score} (Gulf of Biscay).

(d) DAV (Gulf of Biscay).

Fig. 6. Sensitivity of the BC techniques to the years used for identification.



(a) Mean absolute bias comparison. (b) PDF_{score} comparison.

Fig. 7. Mean absolute bias and PDF_{score} comparison among the four locations analysed (DAGQM technique).

validity of the BC techniques as a function of the distance from the reference observation point.

5.1. Temporal sensitivity

On a first instance, results concerning the Gulf of Biscay case study are reported. Fig. 6 shows a comparison of the EGQM and DAGQM techniques by the means of the metrics stated in Section 3.1: lighter clouds of points represent all the different results obtained for each scenario. Darker and bigger points express the mean among all scenarios, computed for each number of Identification Years. All plots suggest that the directional technique generally outperforms the simple Gumbel Quantile Mapping, showing a mean absolute bias closer to zero and lower dispersion. Additionally, both PDF_{score} and DAV certify a bigger improvement in the quality of the corrected dataset, with respect to the one obtained through the EGQM method. This holds for all the scenarios, independently of the number of years used for calibration. Moreover, point clouds representing the considered scenarios seem to converge more quickly for the DAGQM technique, suggesting it to be less sensitive to the choice of identification period. Both MAB and dispersion plots display very sparse clouds of points when only 1 or 2 years of data are used. Performance improves significantly as more years are included, even though little to no improvement is observed when more than 5 years are used for identification. Both PDF_{score} and DAV seem to confirm this for DAGQM, whereas EGQM displays higher inconsistency, though it generally improves when more years are considered for calibration . Considering all metrics, the DAGQM technique consistently outperform the EGQM one, independently on the number of years considered for calibration. Given its higher accuracy, following the same path underlined in Lemos et al. (2020) and Callea et al. (2024), the DAGQM is assumed as the reference method from here on. Hence, all results from following analyses will refer exclusively to the directional bias correction method.

Assuming it as a reference, a cross-comparison between the four selected locations is shown in Fig. 7. A recurring pattern can be spotted in Fig. 7(a), with MAB progressively decreasing as the number of Identification Years is increased, reaching a plateau when 5 or more years are considered. Similarly, the points representing each scenario tend to converge after more than 3 I.Y. are used. The PDF_{score} plot (Fig. 7(b)) confirms this trend, as all locations converge between 3 and 5 years of input data, and the curves reach their optimum at around 6 years. Interestingly, Cape Silleiro stands out with a higher

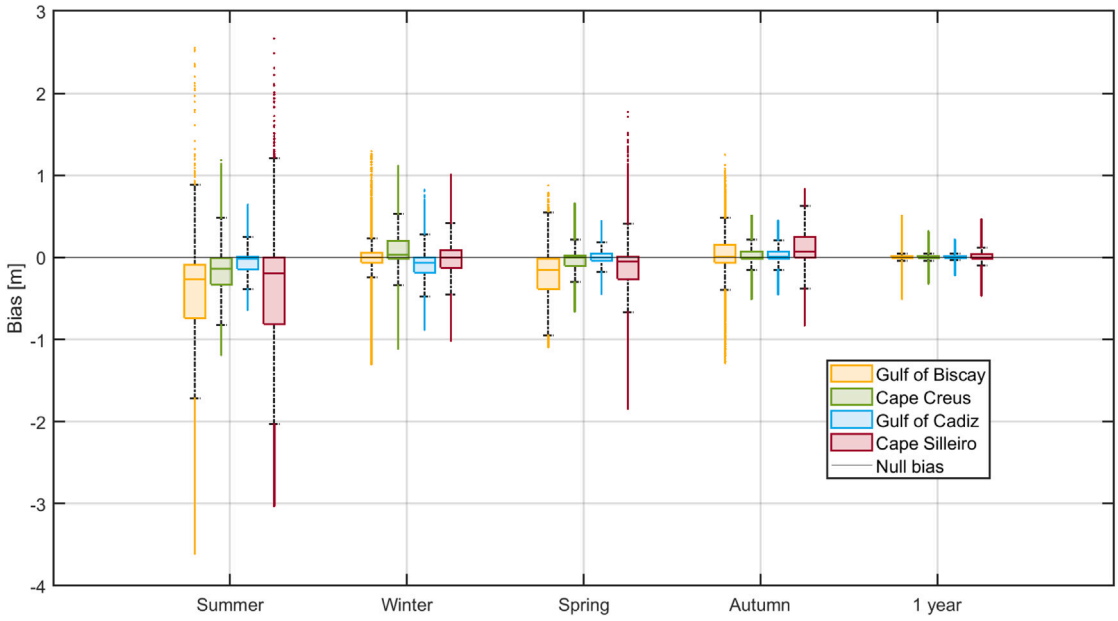


Fig. 8. BC techniques sensitivity to the calibration season (boxplots, for all locations).

mean absolute bias, also resulting more sensitive to the specific time window chosen for identification. Nevertheless, looking at the PDF_{score} it stands as one of the most improved datasets (Fig. 7(b)), suggesting that a poorer quality of the ERA5 data at the site might lead to a higher mean absolute bias despite still reflecting substantial improvement. On the other hand, at the Gulf of Cadiz, the lowest absolute bias does not correspond to the highest improvement measured by PDF_{score}, probably due to a very good quality of the ERA5 data at the site.

In the attempt to obtain a reliable reference database from a shorter and cheaper *in-situ* measurement campaign, a comparison was led between the four seasons of the year, assessing whether one could serve as a better reference than the others. 3-months long periods extracted from the year 2012 have been used for the identification of correction factors, while data belonging to the 2013–2023 window have been corrected. Results are reported in Fig. 8 in the form of boxplots for the four locations. Despite being operationally easier to operate in, thanks to its milder conditions, summer does not provide enough information to properly calibrate the bias correction techniques. Additionally, enclosed basins (such as Gulf of Cadiz and Cape Creus) seem to be better corrected using autumn and spring data, while the two Atlantic sites (Gulf of Biscay and Cape Silleiro) benefit more from winter data. Overall, autumn seems to be best trade-off, providing reliable results across all locations.

Cape Silleiro consistently displays the highest bias among all sites, as discussed for Fig. 7(b), likely due to the lower quality of its raw assimilated data. Nevertheless, the results are encouraging, as backed up by the PDF_{score}, suggesting that trustworthy information can be obtained from 3-month datasets.

5.1.1. Extreme values representation

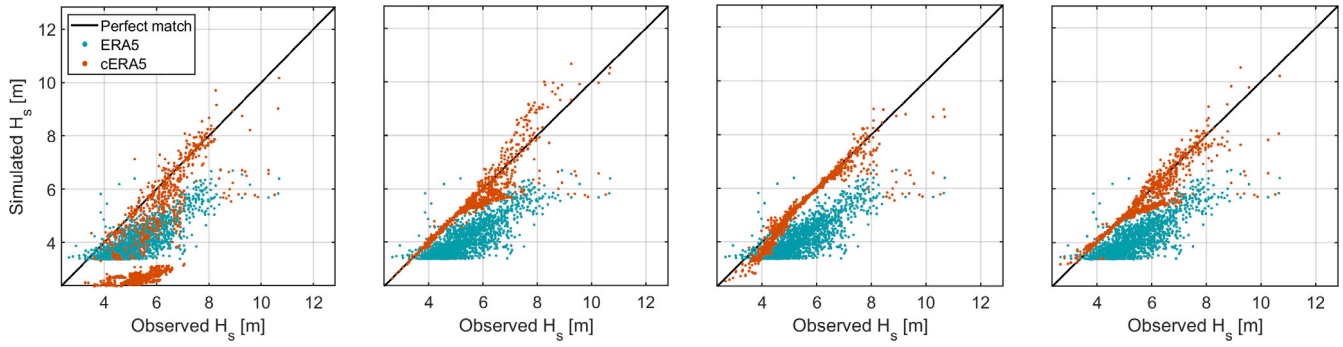
Using shorter time windows considerably increases the risk of missing important information about rare and extreme weather phenomena. Fig. 8 already showed and compared the potential of each season to serve as identification period. As previously discussed, summer calibration was already proven not to be effective, due to excessively mild conditions, while winter, spring and especially autumn provided more encouraging results.

For a deeper understanding, this section presents an insight on the number of high H_s events in each dataset. When focusing on the upper tail of a distribution, quantiles between the 90th and the 99th are generally considered (De Leo et al., 2020; Campos et al., 2019). Following the approach described by Weisse and Günther (2007), the 95th percentile of the ERA5 dataset is used here as a threshold for the classification of extreme events: such value was computed, at each location, for the 2012–2023 correction window. Table 1 reports the occurrence of events classified as extreme per each season of the year 2012 (used for identification), in both the observation and the ERA5 datasets. As expected, summer fails to provide information about extreme wave phenomena, except at Cape Creus, where it still represents the milder season). Overall, autumn exhibits the highest number of high-H_s events; Gulf of Biscay is the only exception in the observed dataset. It is also interesting to note that in all locations (except Gulf of Cadiz) ERA5 models more extreme phenomena than buoys, although these are considered more reliable.

Fig. 9 compares ERA5 distributions (before and after correction) against buoy observations, showing only values above the threshold reported in Table 1. Each box refers to a different season used for calibration; for a matter of space, results are only shown for the Gulf of Biscay case. Coherently with Table 1, the absence of extreme events in summer results in a strong underestimation of extreme H_s values 9(a). On the other hand, winter, spring and autumn all lead to a strong improvement in the representation of higher waves, proving the method's ability to capture extreme values with shorter identification windows. This is due to a key strength of the Gumbel distribution: since more than half quantiles are placed over the 99th percentile (Gumbel, 1935), an adequate representation of the highest tails of the distribution is ensured even when short periods of time are used for the calculation of correction factors.

5.2. Spatial sensitivity

In this section, the results obtained from the spatial sensitivity analysis are summarized. Following the conclusions obtained in Section 5.1, the DAGQM method will be used for the calibration of the ERA5 data, and the will be extracted as the performance evaluation metric. The

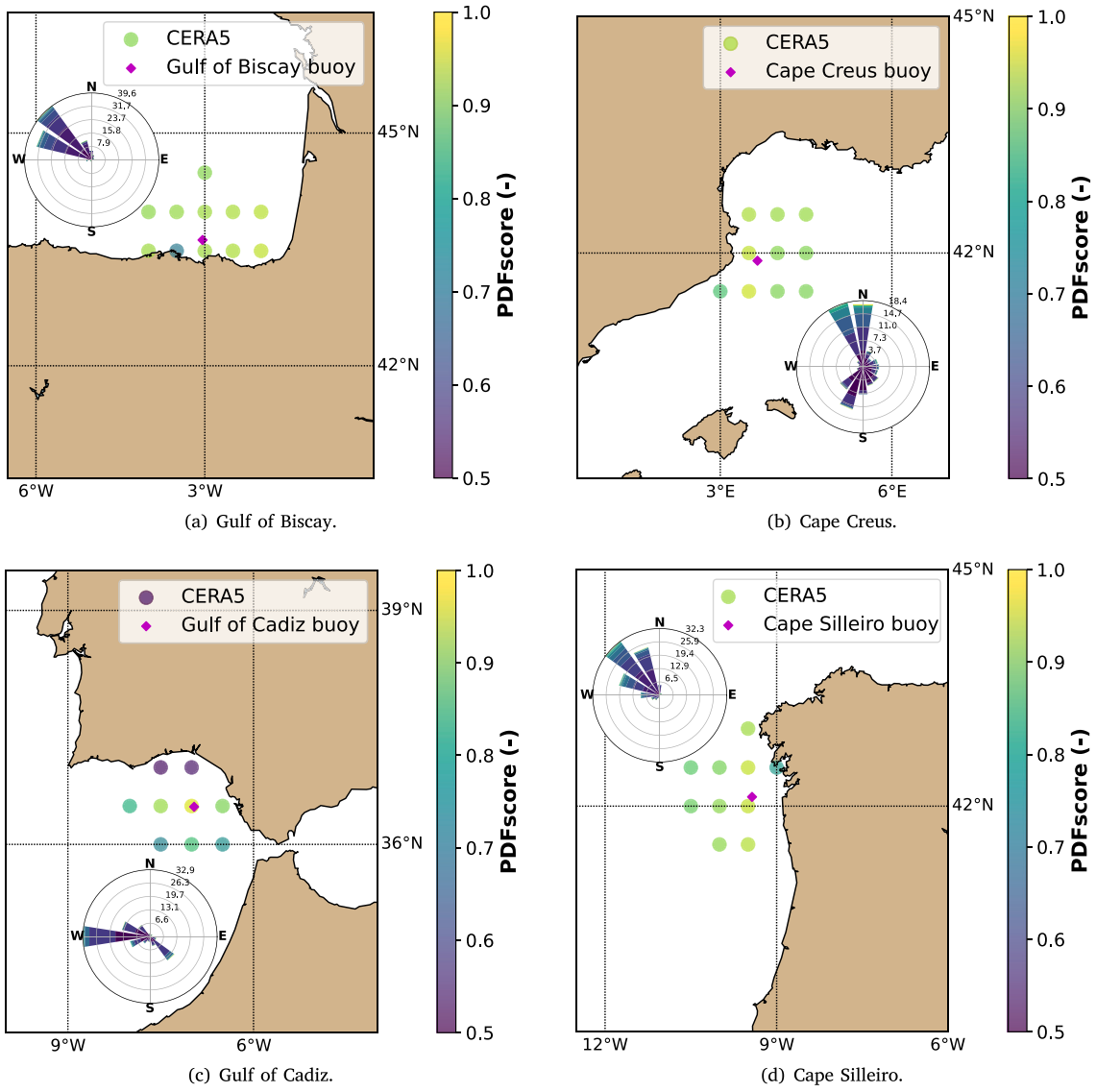


(a) Summer. (b) Winter. (c) Spring. (d) Autumn.

Fig. 9. Scatter diagrams of wave phenomena above the 95th percentile, per each identification window (Gulf of Biscay).

Table 1
Number of extreme events per all Identification windows used (at all locations).

Events > 95th quantile (ERA5)	Threshold [m]	Summer		Winter		Spring		Autumn	
		Buoy	ERA5	Buoy	ERA5	Buoy	ERA5	Buoy	ERA5
Gulf of Biscay	3.37	0	0	122	215	61	121	95	446
Cape Silleiro	4.35	0	0	30	102	33	66	86	221
Gulf of Cadiz	2.37	0	0	19	12	97	82	143	112
Cape Creus	2.67	46	96	375	544	44	72	648	1141



(a) Gulf of Biscay. (b) Cape Creus.

(c) Gulf of Cadiz. (d) Cape Silleiro.

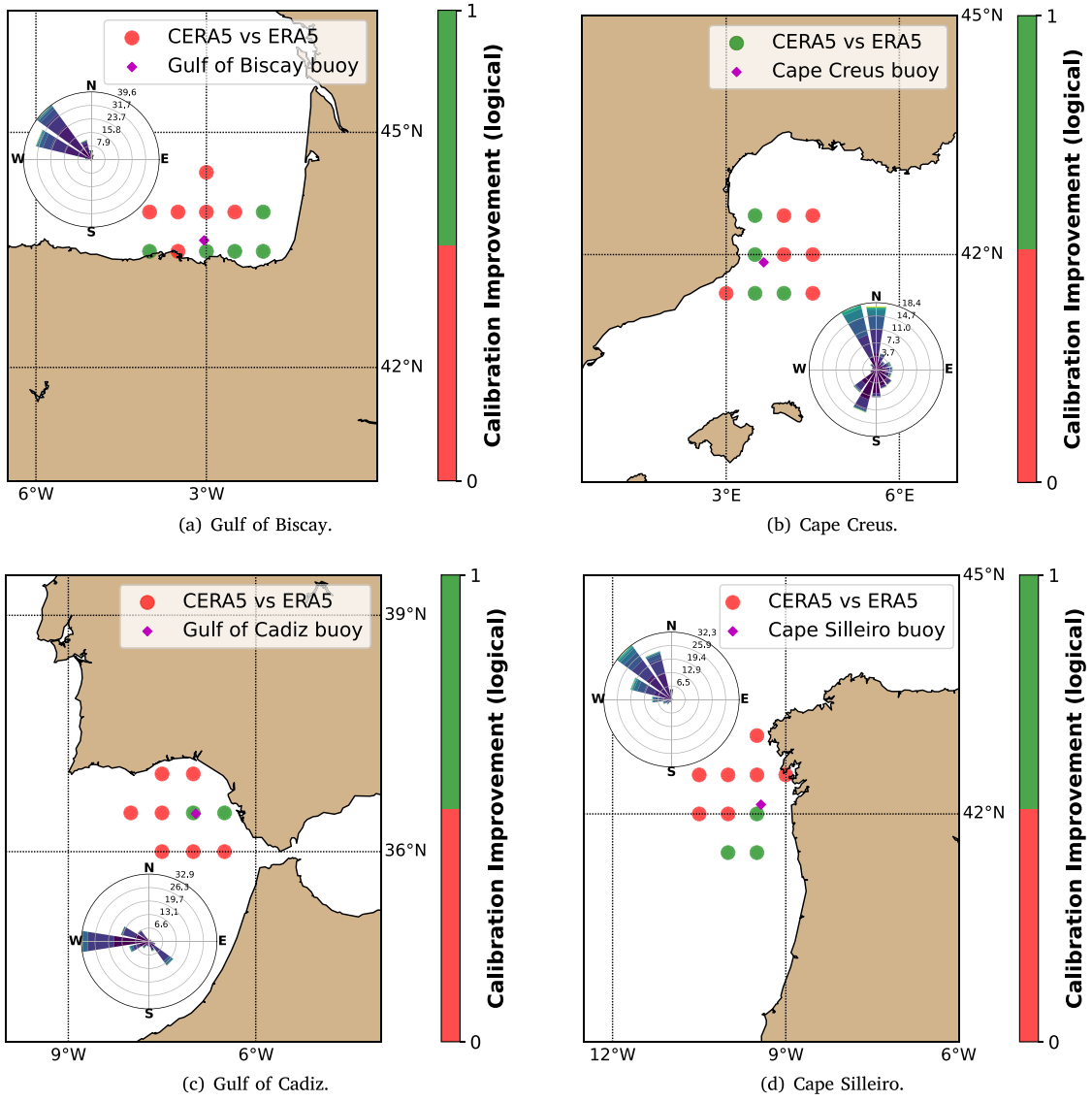
Fig. 10. Calibration performance in the grid-points around the reference buoy, in all locations analysed.

results presented are a combination of spatial graphs and tables that represent the behaviour of the PDF_{score} around the reference buoy.

Regarding the spatial graphs, Fig. 10 shows the PDF_{score} after calibration for each ERA5 coordinate (here referred to as "cERA5") and each study case. In general, it is observed that after bias correction, most grid points show PDF_{score} values below 0.9 (with much lower values in the case of some nodes, especially in Gulf of Cadiz and Cape Silleiro). High agreement (PDF_{score} closer to one) is generally only observed in the buoy proximities. By comparing the PDF_{score} of the raw ERA5 and the calibrated dataset, it is possible to identify those locations where the application of the technique results in an improvement or a deterioration of the accuracy. For easier graphical distinction, the results are represented with logical values in Fig. 11, with 1 (green) indicating the grid-points where PDF_{score} experienced an improvement and 0 (red) corresponding to a worsened PDF_{score}. It can be deduced that data only improve in a small fraction of the total grid points considered for each location, as summarized in Table 2.

Gulf of Biscay emerges as the location with highest number of improved grid-points (5/11) and higher mean PDF_{score} (0.91). Intermediate performances are observed in Cape Silleiro and Cape Creus, while Gulf of Cadiz records both the lowest number of points improved (2/9) and the lowest mean PDF_{score} (only 0.77). This aligns well with what discussed for Fig. 5, which highlighted the already high ERA5 data quality in that area. Analysing the results spatial distribution from Fig. 11, a higher concentration of improved points can be observed in the buoy surroundings for all location. Although no clear common pattern can be identified, results seem to suggest a decreasing validity of bias correction techniques as distance from the reference identification point increases.

Hypotheses can be made on the correlation between the calibration performance and the distance from the buoy. In fact, the bias correction techniques here presented are statistical methods, and do not account for the physical features that might influence the wave phenomena, such as bathymetry, predominant sea type (wind sea or swell) or wind speed and direction. All these may vary significantly over the range of some tens of kilometres, generating very different conditions from one grid-point to another. For example, closer to the shore the local morphology and the seabed characteristics tend to have a stronger influence on wave dynamics, either dissipating or concentrating their energy by the means of phenomena such as refraction, diffraction and reflection which have a strong impact on wave height. In the case of Gulf of Biscay, for example, the bathymetry drastically changes



(a) Gulf of Biscay. (b) Cape Creus.

(c) Gulf of Cadiz. (d) Cape Silleiro.

Fig. 11. Calibration performance represented with logical values.

from −3000 m to −20 m over a relatively short distance (around 40 km, as Fig. 4 shows), calling for geophysical factors which result in very distinct wave characteristics. Therefore, extending calibration to locations far from the buoys presents complex challenges that cannot be adequately addressed using purely statistical bias-correction methods.

6. Conclusions

This article studies the impact of integrating wave direction into bias correction techniques, as well as their applicability beyond the original calibration time frame (*temporal sensitivity*) and location (*spatial sensitivity*). Two variants of the Gumbel Quantile Mapping method – a non-directional (EGQM) and a directional one (DAGQM) – have been tested and compared within the temporal sensitivity framework. Results show that the more specifically-targeted directional approach consistently outperform its non-directional counterpart in all locations. The analysis of the identification time period highlights that 4 to 5 years represent the optimum duration for minimizing mean bias in all case studies. This choice also provides the best trade-off in terms of PDF_{score} across locations, while making the calibration results almost independent of the exact time window considered.

Additionally, the possibility to use periods shorter than 1 year for calibration has been investigated. Each season of the year has been tested as an identification period: summer and spring do not provide sufficient information to represent the full range of possible sea states, while encouraging results were achieved using winter and especially autumn. This finding is promising, as it suggests the possibility of calibrating long datasets through relatively short on-site measuring

Table 2
Spatial calibration results summary (the average PDF_{score} is computed among all grid-points considered at each location).

Metric	Gulf of Biscay	Cape Silleiro	Gulf of Cadiz	Cape Creus
Improved grid-points	5/11	3/10	2/9	4/10
Average PDF_{score} [–]	0.91	0.90	0.77	0.92

campaigns, enabling time and cost savings in the early stages of site assessment for ORE installation projects.

The spatial sensitivity analysis was carried out comparing calibrated ERA5 data against the SIMAR dataset within a 100 km radius of the buoy-defined reference points at all selected locations. However, the outcomes were less satisfactory. While only a few grid-points experienced actual improvement, the correction factors seem to show a restricted validity. These results demonstrate that the statistical bias correction techniques used in this analysis lack versatility for extrapolating to new coordinates, as they do not account for specific local effects. Given the impact of these specific physical properties at each site, it is estimated that in the buoy surroundings, where features do not change significantly, the spatial application of these calibration techniques could be interesting. However, this remains a hypothesis, and supporting this assumption would require a study designed to meet the specified conditions. Furthermore, the lack of consistency and the peculiar configuration of the locations analysed did not allow the identification of a critical radius of applicability. Due to the aforementioned factors, the use of more robust wave models and observational data is recommended to accurately capture wave dynamics at each coordinate. Also, artificial intelligence algorithms are expected to optimally address this issue, as they can effectively handle vast amounts of data and extract specific patterns.

In conclusion, while this analysis highlights the limits of bias correction techniques for spatial extrapolation, it also reaffirms that they significantly improve the quality of *assimilated* data at the buoy location, as previously shown in Callea et al. (2024). This makes them a valuable tool for calibrating data from re-analysis models. Bias correction techniques play a leading role in improving resource assessment for ORE applications, since decisions at any stage of an installation's lifecycle benefit from a more accurate understanding of the conditions at the site. Future work should explore The impact of BC techniques – and extend their applicability – on productivity estimates and other key design parameters.

CRediT authorship contribution statement

Francesco Callea: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Ander Martinez-Perurena:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Giuseppe Giorgi:** Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Markel Penalba:** Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Gregorio Iglesias:** Supervision, Project administration, Funding acquisition.

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.

Acknowledgments

This publication is part of the project PNRR-NGEU which has received funding from the MUR-DM630/2024; funding is also related to the European Union - NextGenerationEU Award Number: Project code PE0000021, Concession Decree No. 1561 of 11.10.2022 adopted by Ministero dell'Università e della Ricerca (MUR), Italy, CUP, Italy E13C22001890001, Project title "Network 4 Energy Sustainable Transition - NEST". It is also part of the research project PID2021-124245OA-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF. A way of making Europe, and the research project funded by the Basque Government's ELKARTEK 2024 program under the grant No. KK-2024/00086. Finally, the authors from the Fluid Mechanics research group at Mondragon University are also supported by the Basque Government's Research Group Program under the grant No. IT1505-22. All authors have read and approved the final version of the manuscript.

Data availability

All of the data used in this study are publicly available for free. Buoy observations and data at SIMAR grid-points are provided by the Spanish authority "Puertos del Estado", through their platform Portus (https://portus.puertos.es), and are free for academic and research purposes. ERA5 data are developed by the European Center for Medium-Range Weather Forecast, and are available at the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels).

References

Callea, F., Martinez-Perurena, A., Zarketa-Astigarraga, A., Penalba, M., Cervelli, G., Giorgi, G., Robertson, B., Iglesias, G., 2024. On the temporal sensitivity of the reference data in bias correction techniques for improving metocean datasets. In: Innovations in Renewable Energies Offshore: Proceedings of the 6th International Conference on Renewable Energies Offshore. (RENEW 2024, 19-21 November 2024, Lisbon, Portugal), CRC Press, pp. 3–10.

Campos, R., Guedes Soares, C., Alves, J., Parente, C., Guimaraes, L., 2019. Regional long-term extreme wave analysis using hindcast data from the south atlantic ocean. Ocean Eng. 179, 202–212. http://dx.doi.org/10.1016/j.oceaneng.2019.03.023.

Carreno-Madinabeitia, S., Ibarra-Berastegi, G., Sáenz, J., Ulazia, A., 2021. Long-term changes in offshore wind power density and wind turbine capacity factor in the Iberian Peninsula (1900–2010). Energy 226, http://dx.doi.org/10.1016/j.energy.2021.120364.

Cervelli, G., Parrinello, L., Moscoloni, C., Giorgi, G., 2022. Comparison of the ERA5 wave forecasting dataset against buoy record. Instrum. Mes. Métrologies 21 (3).

De Leo, F., Solari, S., G., B., 2020. Extreme wave analysis based on atmospheric pattern classification: an application along the Italian coast. Nat. Hazards Earth Syst. Sci. 20 (5), 1233–1246. http://dx.doi.org/10.5194/nhess-20-1233-2020.

ECMWF, 2023. Browse reanalysis datasets. Online, URL https://www.ecmwf.int/en/forecasts/datasets/browse-reanalysis-datasets, (Last Accessed 29 August 2023).

del Estado, P., Prediccion de oleaje, nivel del mar ; Boyas y mareografos, URL http://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx.

del Estado, P., 2024a. Informe sobre la validaciòn mensual del modelo de oleaje de Puertos del Estado. Technical Report, Puertos del Estado, Madrid, Spain, URL https://portus.puertos.es/Portus/pdf/informeValidacion.pdf.

del Estado, P., 2024b. Conjunto de datos SIMAR. URL https://bancodatos.puertos.es/BD/informes/INT_8.pdf.

Gumbel, E., 1935. Les valeurs extrêmes des distributions statistiques. Ann. L'Institut Henri Poincaré 5, 115–158.

Haselsteiner, A.F., Mackay, E., Thoben, K.-d., 2021. Reducing conservatism in highest density environmental contours. Appl. Ocean Res. 117 (November), 102936. http://dx.doi.org/10.1016/j.apor.2021.102936.

IPCC, 2023. Synthesis report of the IPCC sixth assessment report (AR6). Tech. Rep., Intergovernmental Panel on Climate Change (IPCC), ISBN 978-92-9169-151-7 URL https://report.ipcc.ch/ar6syr/pdf/IPCC_AR6_SYR_SPM.pdf.

Juan, N.P., Rodríguez, J.O., Valdecantos, V.N., 2024. Comparison of the SIMAR-WANA, ERA-5, and waverys databases for maritime climate estimations and the implications of coastal protection structures. J. Waterw. Port Coast. Ocean. Eng. 150 (1), 04023021. http://dx.doi.org/10.1061/JWPED5.WWENG-2017.

Konuk, E.-B., Centeno-Telleria, M., Zarketa-Astigarraga, A., Aizpurua, J.-I., Giorgi, G., Bracco, G., Penalba, M., 2023. On the definition of a comprehensive technology-informed accessibility metric for offshore renewable energy site selection. J. Mar. Sci. Eng. 11 (9), http://dx.doi.org/10.3390/jmse11091702, URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85172769954&doi=10.3390%2fjmse11091702&partnerID=40&md5=1122e7c8842ee9b23b9620652f92ed96.

Lemos, G., Menendez, M., Semedo, A., Camus, P., Hemer, M., Dobrynin, M., Miranda, P.M., 2020. On the need of bias correction methods for wave climate projections. Glob. Planet. Change 186 (December 2019), 103109. http://dx.doi.org/10.1016/j.gloplacha.2019.103109.

Mínguez, R., Espejo, A., Tomás, A., Méndez, F., Losada, I., 2011. Directional calibration of wave reanalysis databases using instrumental data. J. Atmos. Ocean. Technol. 28 (11), 1466–1485.

Penalba, M., Aizpurua, J.I., Martinez-Perurena, A., 2021. On the definition of a risk index based on long-term metocean data to assist in the design of Marine Renewable Energy systems. Ocean Eng. 242 (March), 110080. http://dx.doi.org/10.1016/j.oceaneng.2021.110080.

Penalba, M., Guo, C., Zarketa-Astigarraga, A., Cervelli, G., Giorgi, G., Robertson, B., 2023. Bias correction techniques for uncertainty reduction of long-term metocean data for ocean renewable energy systems. Renew. Energy 219, 119404. http://dx.doi.org/10.1016/j.renene.2023.119404.

Perkins, S.E., Pitman, A.J., Holbrook, N.J., McAneney, J., 2007. Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J. Clim. 20 (17), 4356–4376. http://dx.doi.org/10.1175/JCLI4253.1.

Semedo, A., Dobrynin, M., Lemos, G., Behrens, A., Staneva, J., De Vries, H., Sterl, A., Bidlot, J.-R., Miranda, P.M., Murawski, J., 2018. CMIP5-derived single-forcing, single-model, and single-scenario wind-wave climate ensemble: Configuration and performance evaluation. J. Mar. Sci. Eng. 6 (3), 90.

Semedo, A., Sušelj, K., Rutgersson, A., Sterl, A., 2011. A global view on the wind sea and swell climate and variability from ERA-40. J. Clim. 24 (5), 1461–1479.

Soares, P.M., Cardoso, R.M., 2018. A simple method to assess the added value using high-resolution climate distributions: Application to the EURO-CORDEX daily precipitation. Int. J. Climatol. 38 (3), 1484–1498. http://dx.doi.org/10.1002/joc.5261.

Stéphanie Bouckaert, Pales, A.F., McGlade, C., Remme, U., Wanner, B., 2021. Net Zero by 2050: A Roadmap for the Global Energy Sector. Tech. rep., International Energy Agency, Paris, p. 224, URL https://www.iea.org/reports/net-zero-by-2050.

UN, 2015. Adoption of the paris agreement. https://www.iea.org/publications/freepublications/publication/ElectricityInformaion2017Overview.pdf.

Weisse, R., Günther, H., 2007. Wave climate and long-term changes for the Southern North Sea obtained from a high-resolution hindcast 1958–2002. Ocean. Dyn. 57, 161–172. http://dx.doi.org/10.1007/s10236-006-0094-x.