



The Controlled Four-Parameter Method for cross-assignment of directional wave systems

Andre Luiz Cordeiro dos Santos ^a, Felipe Marques dos Santos ^b , Nelson Violante-Carvalho ^c , Luiz Mariano Carvalho ^d , Helder Manoel Venceslau ^a

^a Mathematics Institute, Federal Center for Technological Education of Rio de Janeiro (CEFET-RJ), Brazil

^b National Oceanography Centre, Southampton, UK

^c Ocean Engineering Program, Rio de Janeiro Federal University (COPPE-UFRJ), Brazil

^d Applied Mathematics Department, Institute of Mathematics and Statistics, Rio de Janeiro State University (UERJ), Brazil

ARTICLE INFO

Keywords:

Cross-assignment of wave spectra partitions

Wind-generated waves

Wave spectra assessment

Wave spectra assimilation

ABSTRACT

Cross-assignment of directional wave spectra is a critical task in wave data assimilation. Traditionally, most methods rely on two-parameter spectral distances or energy ranking approaches, which often fail to account for the complexities of the wave field, leading to inaccuracies. To address these limitations, we propose the Controlled Four-Parameter Method (C4PM), which independently considers four integrated wave parameters. This method enhances the accuracy and robustness of cross-assignment by offering flexibility in assigning weights and controls to each wave parameter. We compare C4PM with a two-parameter spectral distance method using data from two buoys moored 13 km apart in deep water. Although both methods produce negligible bias and high correlation, C4PM demonstrates superior performance by preventing the occurrence of outliers and achieving a lower root mean square error across all parameters. The negligible computational cost and customization position C4PM as a potential alternative for wave data assimilation, possibly improving the reliability of forecasts and model validations.

1. Introduction

The wave directional spectrum is the fundamental representation of a sea state, providing a detailed description of the energy distribution as a function of both frequency and direction — hence crucial for understanding the complexities of wave dynamics and interactions. Accurate analysis of the spectrum is vital for various applications, including climate studies, coastal management and maritime safety (Cavaleri et al., 2007; Ardhuin et al., 2019). One of the key challenges is the cross-assignment of spectral partitions, which involves identifying and matching collocated wave systems from different datasets or models — see a discussion about spectral partitioning in Gerling (1992), Violante-Carvalho et al. (2005), Portilla-Yandún et al. (2015) and Portilla-Yandún et al. (2019). Cross-assignment is essential for data assimilation, where observational data are integrated into numerical models (Aouf et al., 2006; Hauser et al., 2021), and for assessing measurements, enabling comparison of results from various sources.

Currently, most (if not all) methods for cross-assignment are based either on two-parameter spectral distance or on ranking the energy content of each system — the most energetic partitions in one spectrum are paired to the correspondingly ranked partitions in the other spectrum

(Jiang et al. 2022, Hauser et al. 2021). Energy ranking methods are more prone to inaccuracies, mainly when the number of partitions of each spectra differ. Cross-assigning only the most energetic partition might overcome this limitation but significantly reduce the possible number of matches. Conversely, methods based on a two-parameter spectral distance rely exclusively on frequency and direction (see, among many others, Hasselmann et al., 1996; Hanson and Phillips, 2001; Li and Saulter, 2012; Wang et al., 2020; Smit et al., 2021; Aouf et al., 2021; Santos et al., 2021; Jiang et al., 2022; Ricondo et al., 2023; Wu et al., 2024). However, this approach has drawbacks. The main limitation is that it can result in errors, mainly because partitions close in frequency but significantly apart in direction (or vice versa) are mismatched — leading to potential discrepancies caused by outliers. These inaccuracies can propagate through data assimilation processes, resulting in suboptimal model performance and potentially misleading data interpretations.

To overcome these limitations, we propose a novel methodology for cross-assigning partitions, termed the Controlled Four-Parameter Method (C4PM). This approach incorporates four bulk wave spectral parameters: significant wave height, peak wave period, peak wave

* Corresponding author.

E-mail address: felipe.santos@noc.ac.uk (F.M. dos Santos).

<https://doi.org/10.1016/j.ocomod.2025.102584>

Received 8 December 2024; Received in revised form 11 June 2025; Accepted 11 June 2025

Available online 26 June 2025

1463-5003/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

direction, and peak wave spreading. In the comparative analysis conducted, in which more than 30,000 partition match-ups were independently obtained by the C4PM and 2PM algorithms, our technique consistently resulted in lower *RMSE* values across all cross-assigned wave parameters. These findings demonstrate the robustness and superior accuracy of the proposed method, highlighting its potential to enhance the assimilation of observational data into numerical models and, consequently, to improve forecast accuracy. C4PM has proven to be an effective and versatile tool for cross-assignment, offering control over each primary wave parameter and enabling their prioritization through a customizable weighting vector used in the distance calculation between partitions.

The structure of the current analysis is organized as follows: Section 2 presents the theoretical framework of C4PM, which includes the definition of the semimetric on which the method is based, as well as a detailed presentation of the cross-assignment scheme used. Section 3 outlines the buoy data and the processing methodologies employed. Section 4 discusses the results, while the conclusions are summarized in Section 5. Finally, the error metrics and wave parameter formulations are described in Appendix.

2. The cross-assignment problem

This section addresses the cross-assignment problem, beginning with a basic concept, followed by a brief description of a classic technique and its limitations, and concluding with a detailed explanation of the C4PM method.

2.1. Matching: a fundamental concept

Assuming that $A(\mathbf{k})$ and $B(\mathbf{k})$ are two measurements of the wavenumber spectrum associated with a certain sea state. After a partitioning process of $A(\mathbf{k})$ and $B(\mathbf{k})$, consider the additive decomposition:

$$\begin{cases} A(\mathbf{k}) = A_1(\mathbf{k}) + A_2(\mathbf{k}) + \dots + A_p(\mathbf{k}) \\ B(\mathbf{k}) = B_1(\mathbf{k}) + B_2(\mathbf{k}) + \dots + B_m(\mathbf{k}) \end{cases} \quad (1)$$

where $A_1(\mathbf{k}), A_2(\mathbf{k}), \dots, A_p(\mathbf{k})$ and $B_1(\mathbf{k}), B_2(\mathbf{k}), \dots, B_m(\mathbf{k})$ are, respectively, the partitions¹ of $A(\mathbf{k})$ and $B(\mathbf{k})$. It is supposed from now on that $p \leq m$ and that the explicit dependence on the wavenumber vector $\mathbf{k}$ will be omitted.

We call *n-matching* between A and B a set

$$\mathcal{M} = \left\{ \left\{ A_{i_1}, B_{j_1} \right\}; \left\{ A_{i_2}, B_{j_2} \right\}; \dots; \left\{ A_{i_n}, B_{j_n} \right\} \right\}, \quad (2)$$

formed by $n \leq p$ of pairs² of partitions of A and B with the property that no partition of A is associated with more than one partition of B and vice versa, where $1 \leq i_k \leq p$, $1 \leq j_k \leq m$, and $k \leq n$. Ideally, a cross-assignment is a matching between A and B such that coupled partitions exhibit the highest possible concordance in their oceanic characteristics. Therefore, it is crucial to employ a technique capable of accurately identifying compatible wave systems.

2.2. A two-parameter method (2PM)

Typically, the cross-assignment of spectral partitions heavily relies on evaluating a two-parameter spectral distance. In this context, the fundamental concept behind the technique proposed in Hasselmann et al. (1996) is the association of a partition with its so-called

(two-parameter) characteristic wavenumber vector. In that study, the expression

$$\Delta(A_i, B_j) = \frac{\|\mathbf{k}_c(A_i) - \mathbf{k}_c(B_j)\|}{\sqrt{\|\mathbf{k}_c(A_i)\|^2 + \|\mathbf{k}_c(B_j)\|^2}}, \quad (3)$$

where $\mathbf{k}_c(A_i)$ and $\mathbf{k}_c(B_j)$ are the characteristic wavenumber vectors of A_i and B_j , respectively, was proposed for the distance³ between partitions of A and B . For the definition of characteristic wavenumber vector (CWV), see Appendix, Table A.2. Thus, the *n*-matching given by (2) is the cross-assignment between A and B , in the Hasselmann sense, if the following conditions are met:

- (i) $n \leq p$ is the largest possible
- (ii) $\Delta(A_{i_1}, B_{j_1}) \leq \Delta(A_{i_2}, B_{j_2}) \leq \dots \Delta(A_{i_n}, B_{j_n}) \leq R$
- (iii) $\Delta(A_{i_k}, B_{j_k}) = \min\{\Delta(A_{i_k}, B_j) : j = 1, 2, \dots, m\}$ for each $k = 1, 2, \dots, n$.

Here, R is a specified critical value, which acts as a cutoff line defining which pairs will be in the cross-assignment. Hereafter, we will refer to this technique as 2PM, and its algorithm is illustrated in Fig. 1. The method calculates all possible distances between the partitions of A and B , eliminates pairs with distances exceeding the threshold value R , and finally establishes partition matchups by prioritizing the smallest distances.

One of the main shortcomings of this approach is that even if the spectral distance between two partitions is considered small — below a given threshold — it does not guarantee a strong agreement between the wave systems they represent. Figs. 2 and 3 show pairs of corresponding partitions obtained using the 2PM algorithm, initialized with $R = 0.75$, a reference value suggested by Hasselmann et al. (1996), known to produce fairly reasonable results. Although the partition pairs exhibit distances significantly below this threshold, they display substantial discrepancies in certain wave parameters. Indeed, in Fig. 2, the distance between the partitions is equal to 0.59, but the partitioned significant wave height is approximately nine times larger than its assigned counterpart, meaning that unrelated wave systems were associated in the cross-assignment process. A more subtle case of mismatch with distance equal to 0.32 is depicted in Fig. 3. Despite similar values of partitioned peak wave periods, partitioned significant wave heights and partitioned peak directional spreadings, the wave systems propagating towards the southern quadrant are separated by 120°.

2.3. The controlled four-parameter method — C4PM

As shown in 2.2, 2PM methods for cross-assignment have significant limitations, thereby leaving margin for mismatches in the cross-assignment process. In this context, the *Controlled Four-Parameter Method* (C4PM) is introduced as a more robust and efficient tool, providing comprehensive control over wave parameter deviation and their hierarchical importance.

2.3.1. The weighted semimetric

A semimetric defined on the partitions of A and B is a function d such that for any $1 \leq i \leq p$ and $1 \leq j \leq m$, the following conditions hold: $d(A_i, B_j) \geq 0$, $d(A_i, B_j) = d(B_j, A_i)$, and $d(A_i, B_j) = 0$ if, and only if, $A_i = B_j$. To define our spectral distance, each partition is associated with a four-dimensional vector whose components represent the partitioned integrated wave parameter values that characterize it: partitioned significant wave height (*PSWH*), partitioned peak wave period (*PPWP*), partitioned peak wave direction (*PPWD*), and partitioned peak wave spreading (*PPWS*), precisely in that order. This

¹ We use the term “partition” to refer to an independent wave system that is a constituent of the wave spectrum

² unordered

³ the double-bar symbol designates the Euclidean norm

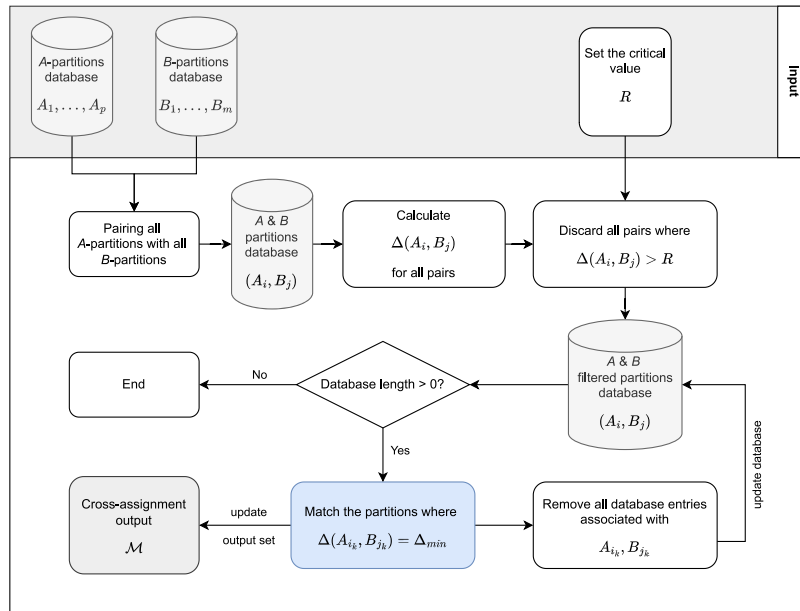


Fig. 1. 2PM algorithm diagram.

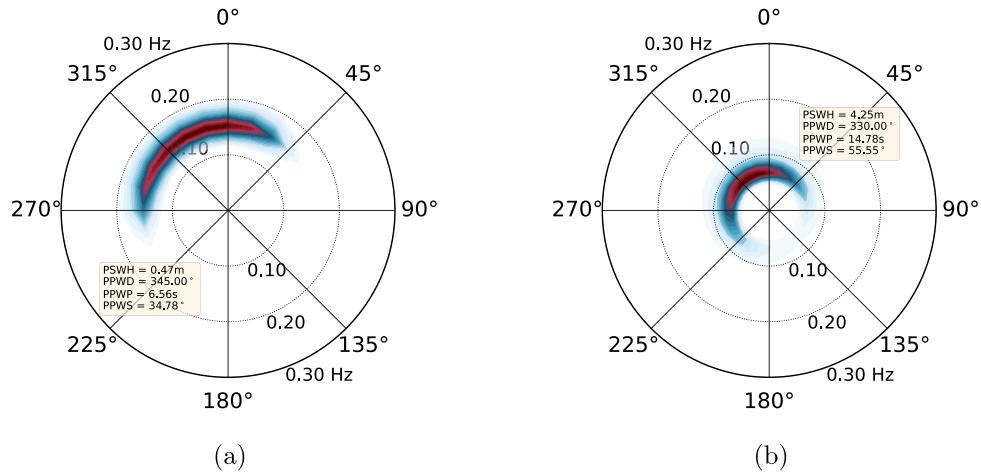


Fig. 2. Directional wave spectra from NDBC buoys: (a) 51001 and (b) 51101 on 12 Jan 2023 at 18:40:00. The boxes indicate the values of the partitioned integrated wave parameters — partitioned significant wave height ($PSWH$), partitioned peak wave period ($PPWP$), partitioned peak wave direction ($PPWD$) and partitioned peak directional spreading ($PPWS$). The Δ -distance between partitions is 0.59.

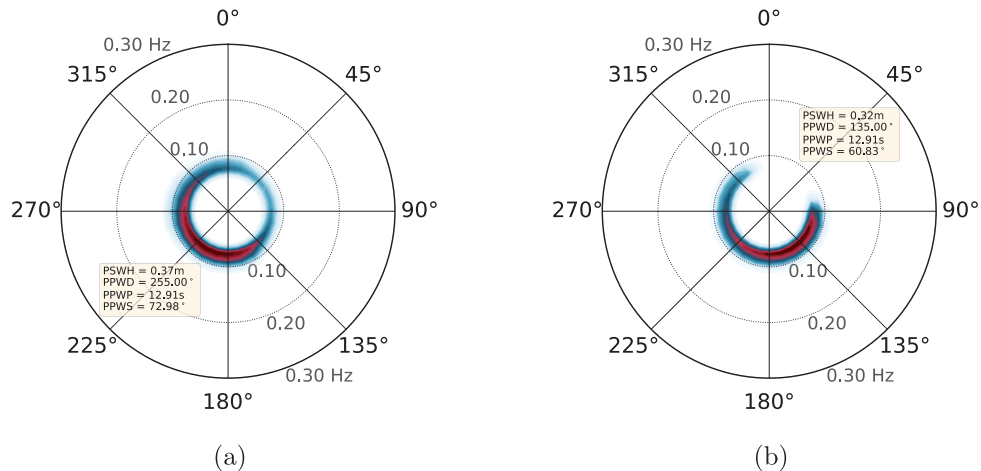


Fig. 3. Directional wave spectra from NDBC buoys: (a) 51001 and (b) 51101 on 02 Aug 2023 at 01:10:00. The boxes indicate the values of the partitioned integrated wave parameters. The Δ -distance between partitions is 0.32.

approach incorporates more information about each partition compared to the distance defined in Eq. (3). Thus, for each, $1 \leq i \leq p$ and $1 \leq j \leq m$, we consider the associations: $A_i \leftrightarrow (h(A_i), t(A_i), \theta(A_i), \sigma(A_i))$ and $B_j \leftrightarrow (h(B_j), t(B_j), \theta(B_j), \sigma(B_j))$, where the symbols h, t, θ and σ represent, respectively, the significant wave height, peak period, peak direction, and directional spread of the partitions they accompany. To set our spectral distance (which is a weighted semimetric) d between the partitions A_i and B_j as

$$d(A_i, B_j) = \mathbf{v}(A_i, B_j) \cdot \mathbf{w}, \quad (4)$$

a dot product, where $\mathbf{v}(A_i, B_j)$ represents the variation vector, encapsulating the deviations of the partitioned integrated wave parameters between A_i and B_j , and $\mathbf{w}$ is the weighting vector, whose components assign weights to each partitioned integrated wave parameter. These vectors are defined as:

$$\mathbf{v}(A_i, B_j) = (v_h(A_i, B_j); v_t(A_i, B_j); v_\theta(A_i, B_j); v_\sigma(A_i, B_j)), \quad (5)$$

where

$$v_z[A_i, B_j] = \frac{|z(A_i) - z(B_j)|}{\max\{z(A_i), z(B_j)\}} \quad \text{for } z = h, t, \sigma \quad (6)$$

and

$$v_\theta[A_i, B_j] = \frac{1}{180^\circ} \min\{|\theta(A_i) - \theta(B_j)|, 360^\circ - |\theta(A_i) - \theta(B_j)|\} \quad (7)$$

are the variation functions, both bounded above by 1 (by construction). The weighting vector $\mathbf{w} = (w_h, w_t, w_\theta, w_\sigma)$ consists of positive scalars w_h, w_t, w_θ and w_σ , which satisfy the normalization condition $w_h + w_t + w_\theta + w_\sigma = 1$ ⁴ and reflect the relative significance of each wave parameter in the computation of the distance between two partitions. It is important to emphasize that the weighting vector offers endless possibilities; for instance, the reliability of the analyzed data serves as a rational criterion for assigning weights to each partitioned integrated wave parameter. For illustrative purposes, the choice of weights $w_h = 0.04, w_t = 0.32, w_\theta = 0.32$ and $w_\sigma = 0.32$ indicates that significant height has reduced importance in the distance calculation and, consequently, will not be decisive in the formation of match-ups.

Eq. (4) explicitly represents the normalized weighted arithmetic mean of the partitioned wave parameter deviation between A_i and B_j , and is rewritten as:

$$d(A_i, B_j) = v_h(A_i, B_j) w_h + v_t(A_i, B_j) w_t + v_\theta(A_i, B_j) w_\theta + v_\sigma(A_i, B_j) w_\sigma. \quad (8)$$

If the weighting vector $\mathbf{w}$ is balanced, meaning $w_h = w_t = w_\theta = w_\sigma$, the distance in Eq. (8) simplifies to the arithmetic mean of the deviation of the partitioned wave parameters. Notably, in contrast to the distance defined by Eq. (3), if $d(A_i, B_j) = 0$, then A_i and B_j fully agree in all their corresponding wave parameters, as d is a semimetric. Conversely, if $d(A_i, B_j) \approx 1$, it indicates that at least one corresponding wave parameter between A_i and B_j exhibits a high degree of discrepancy.

2.3.2. A new cross-assignment formulation

This section introduces a new cross-assignment framework and demonstrates how the algebraic structure of the proposed semimetric, defined as the normalized weighted arithmetic mean of the wave parameter variations between a pair of counterpart partitions, can effectively improve cross-assignment accuracy. To achieve this, consider a control vector $\mathbf{c} = (c_h, c_t, c_\theta, c_\sigma)$, where each coordinate satisfies $0 \leq c_h, c_t, c_\theta, c_\sigma \leq 1$. The n -matching defined by (2) is referred to as the *controlled four-parameter cross-assignment* between A and B , relative to the weighted semimetric d and governed by the control vector $\mathbf{c}$, if the following conditions are satisfied:

- (i') $n \leq p$ is the largest possible

- (ii') $\mathbf{v}(A_{i_k}, B_{j_k}) \leq \mathbf{c}$ for each $k = 1, 2, \dots, n$

$$(iii') \sum_{k=1}^n d(A_{i_k}, B_{j_k}) = \min \left\{ \sum_{k=1}^n d(A_{\rho(k)}, B_{\tau(\rho(k))}) : \rho \in I_{n,p}, \tau \in I_{p,m} \right\}$$

are valid.⁶

Condition (ii') provides control over the deviation of the cross-assigned partitioned wave parameters of the matchups; the previously chosen constraining scalars c_h, c_t, c_θ , and c_σ ultimately control the distance between the matched partitions. If the partitions A_{i_k} and B_{j_k} are $\mathbf{c}$ -controlled — i.e., condition (ii') is valid — then the distance between these partitions is

$$d(A_{i_k}, B_{j_k}) \leq w_h c_h + w_t c_t + w_\theta c_\theta + w_\sigma c_\sigma \quad (9)$$

The optimality condition (iii') characterizes controlled cross-assignment as the matching between A and B with the shortest possible length. The quantity calculated in (iii') represents the length of the cross-assignment; among all the n -matchings of $\mathbf{c}$ -controlled partitions of A and B , this is the one with the smallest length.

If $c_h = c_t = c_\theta = c_\sigma = 1$, the cross-assignment problem is unconstrained, and in this case, there is always a solution. Conversely, for instance, if $c_h = c_\theta = c_\sigma = 1$ and $c_t = 0.3$, there are no constraints on the values of significant wave heights, peak wave directions, or peak directional spreading for the matched partitions. However, a significant constraint is imposed on the values of peak wave periods for the matched partitions. In each matchup, the smallest period is at least $(1 - 0.3) = 0.7$ of its cross-assigned counterpart. The largest period of that matchup — or in other words, the smallest period is at least 70% of the largest period. In this case, the existence of a constrained cross-assignment depends on the dataset being analyzed, as it is possible that no pair of wave spectra partitions considered satisfies the specific constraint of the control vector on the variation vector.

In summary, this section introduced a framework for cross-assignment, which consists of the following: given two partitioned wave spectra (measured) of a specific sea state, the objective is to find the *best* matching between their spectral partitions, ensuring that the discrepancies in all four partitioned integrated wave parameters are controlled *a priori*. To address this problem, we developed a computational routine called the *Controlled Four-Parameter Method* (C4PM), whose diagram is presented in Fig. 4. Specifically, we denote the method as u-C4PM when it operates in uniform mode, meaning the control vector is uniform, with $c_h = c_t = c_\theta = c_\sigma = r$. In this sense, the conditions

- (i'') for $z = h, t$ and σ , $z(B_j) \geq (1 - r)z(A_i)$ if $z(A_i) \geq z(B_j)$ (otherwise, swap A_i with B_j) and

- (ii'') for $z = \theta$, the (lower) angle between the peak directions $\theta(A_i)$ and $\theta(B_j)$ does not exceed $180^\circ \cdot r$

are necessary and sufficient for the partitions A_i and B_j to be r -uniformly controlled, that is, $\mathbf{v}(A_i, B_j) \leq \mathbf{c}$ with $\mathbf{c} = r(1, 1, 1, 1)$.

3. Data and methods

3.1. NDBC buoy data

We selected two buoys, operated by the National Data Buoy Center (NDBC), with IDs 51001 and 51101, for the period from January 2023 to December 2023 — details available at <https://www.ndbc.noaa.gov/>. These buoys are situated approximately 13 km apart in deep water off the coast of Hawaii (Fig. 5). Their proximity, location in

⁵ the symbol $\leq$ indicates that the components of the variation vector do not exceed the corresponding components of the control vector.

⁶ the symbols $I_{n,p}$ and $I_{p,m}$ represent, respectively, the sets of all injective functions from $\{1, 2, \dots, n\}$ into $\{1, 2, \dots, p\}$ and from $\{1, 2, \dots, p\}$ into $\{1, 2, \dots, m\}$.

⁴ purely algebraic fact

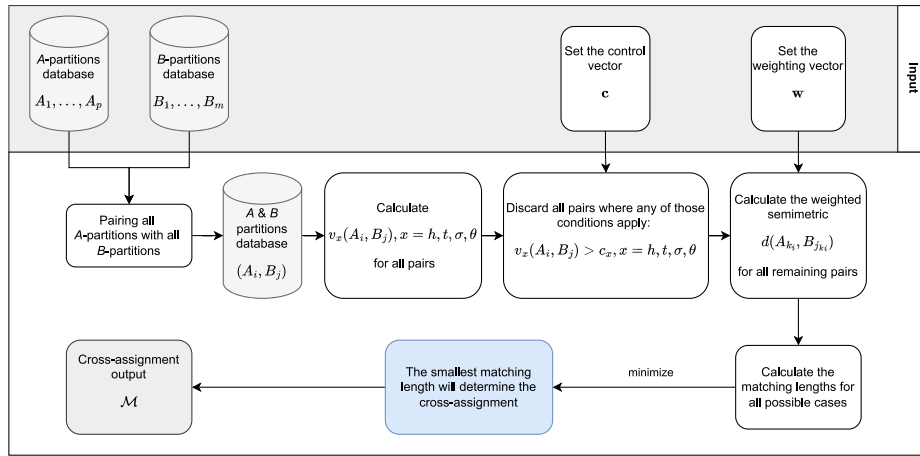


Fig. 4. C4PM algorithm diagram.

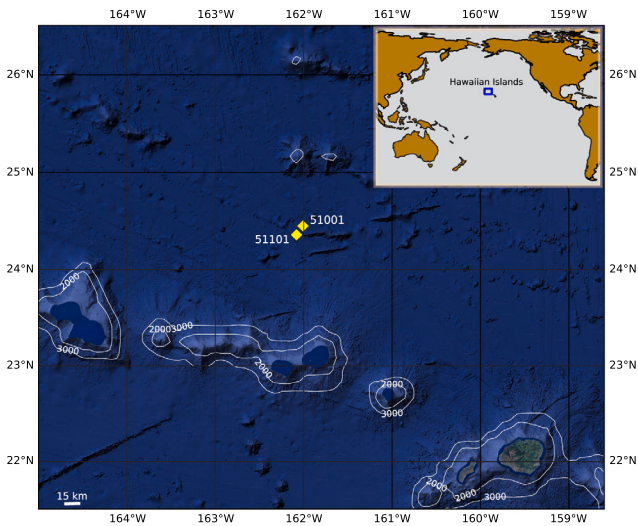


Fig. 5. Location of the NDBC buoys 51001 (24.451 N, 162.008 W, and 4906 m depth) and 51101 (24.359 N, 162.081 W, and 4860 m depth), which are 13 km apart, in deep water off Hawaii.

similar deep-water conditions and payload specifications (both carry the NDBC Self-Contained Ocean Observing Payload — SCOOP) indicate that both buoys provide relatively comparable measurements. This implies a reasonable number of high-quality pairs can be identified between them, while also reducing challenges associated with applying cross-assignment methodologies, as will be discussed in the following sections. Both buoys collect meteorological and oceanographic data, including the directional wave spectra necessary for cross-assignment techniques. Wave data from these buoys are available at 30 minute intervals, yielding a total of 17,100 pairs of wave spectra.

Fig. 6 shows, over the selected time period of 2023, the scatter plots of wave parameters: significant wave height (SWH), peak wave period (PWP), peak wave direction (PWD) and peak wave spreading (PWS). The wave parameters were downloaded directly from the NDBC page — with exception of PWS which was computed as outlined in Table A.2 — therefore the spectra were not partitioned. The statistical parameters are listed in Table A.1.

As expected, the biases across the four parameters are minimal. However, low-frequency occurrences representing significant differences among the spectral peak pairs are evident, as indicated by the light blue shades, particularly in relation to direction and period (Fig. 6b and c). For SWH , the $RMSE$ is 0.20 m with a correlation coefficient equal to 0.98, indicating a strong level of similarity. Although

the agreement for PWP is generally good, numerous outliers in Fig. 6b contribute to an increased $RMSE$ and lower correlation coefficient. A similar pattern is seen in Fig. 6c, where the outliers introduce substantial discrepancies. The lower correlation coefficient among the four wave parameters is associated with PWS , a parameter that is challenging to estimate accurately from single point measurements such as directional buoys (Kuik et al., 1988).

3.2. Data processing

The NDBC provides wave spectral data in the form of five parameters concerning the Fourier series expansion of the buoy's directional wave spectrum: the non-directional spectral density, the first normalized directional Fourier coefficient, the mean wave direction, the second normalized directional Fourier coefficient and the main wave direction. The time series for each buoy were downloaded for the year 2023. The data are then time-synchronized and consolidated into a database, removing any incomplete entries (i.e. missing data) to ensure that each record has complete information for all five wave parameters from both buoys, at the same date and time. Each row of this database is then processed according to NDBC's recommendations in order to generate the directional wave spectra for both buoys, which includes the use of the weighted directional spreading function, presented in Earle et al. (1999), to handle eventual negative values. More than being physically meaningless, negative values in a directional wave spectrum would raise issues during the partitioning and parameter calculations phases.

Once the directional wave spectra are obtained, they are partitioned according to the methodology proposed by Hasselmann et al. (1996), and their partitioned integrated wave parameters are calculated as outlined in Table A.2. Finally, all noisy partitions are removed according to the following criteria:

- (a) $PSWH \leq 0.25$ m;
- (b) $PPWP \leq 5$ s and $PSWH \leq 10\%$ SWH .

These values are consistent with prior studies employing partition-based noise filtering (Violante-Carvalho et al., 2004; Hanson et al., 2009; Devaliere et al., 2009). While slight variations exist across studies, the physical rationale remains comparable. The selected thresholds effectively separate noise from physically meaningful wave systems in our dataset. A full sensitivity analysis, while useful, would exceed our scope.

At this stage, the partition database is consolidated, with the greatest possible amount of matches being 30,956. Both cross-assignment techniques, 2PM and C4PM, were applied to the same dataset, each generating its respective output databases, as illustrated in the processing workflow shown in Fig. 7.

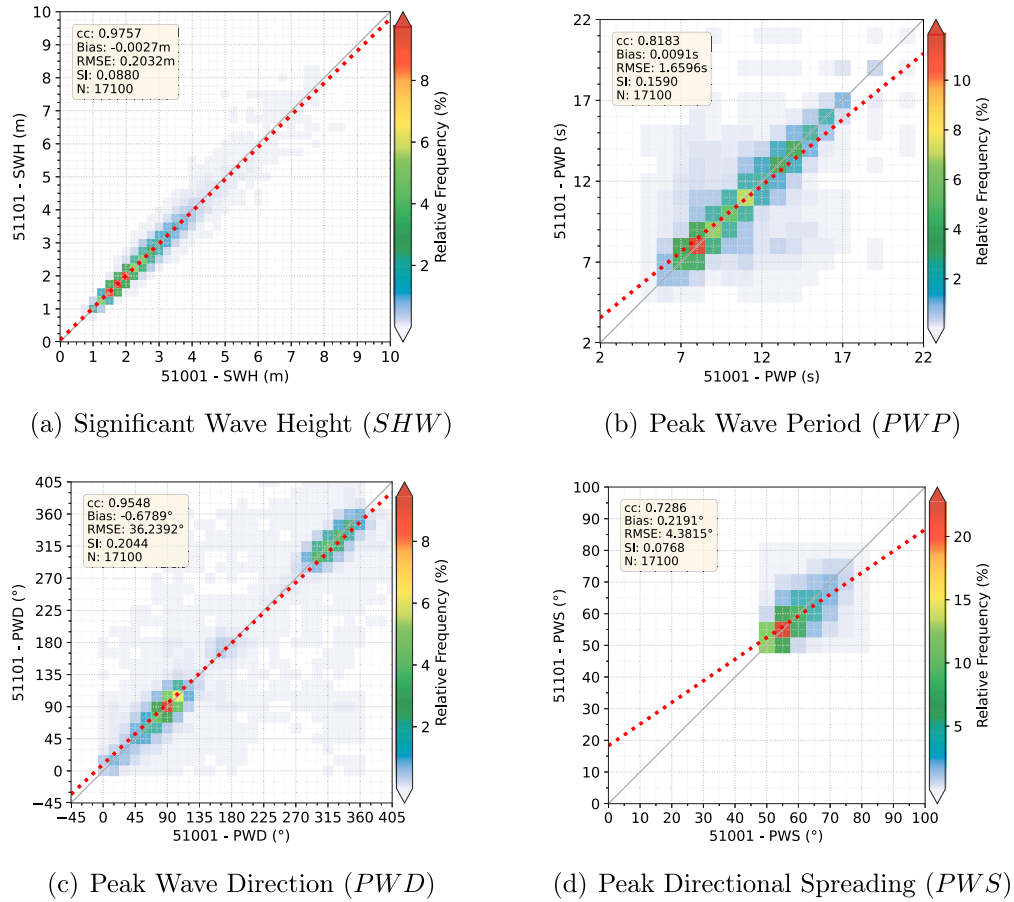


Fig. 6. Scatter plots between buoys 51001 and 51101 over the year 2023. (a) *SWH*, (b) *PWP*, (c) *PWD* and (d) *PWS*. The boxes indicate Pearson correlation coefficient (*cc*), Bias, Root Mean Square Error (*RMSE*), Scatter Index (*SI*) and total number of points (*N*). The red dotted line indicates the best fit from a linear regression using the least-squares method.

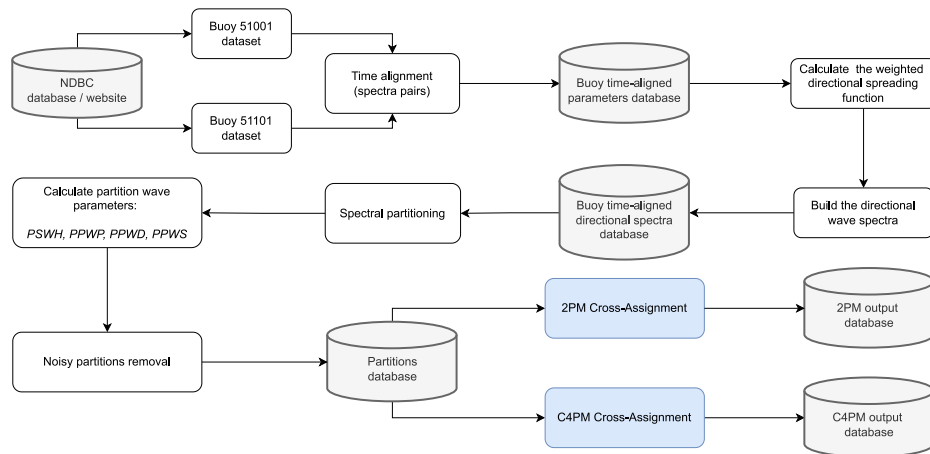


Fig. 7. Data processing diagram.

4. Results

In general, cross-assignment tasks employ spectra from different sources, such as numerical models, *in situ* measurements, or remote sensing, which naturally have some degree of discrepancy between them. In many cases, the number of partitions in each paired spectrum differs, with some partitions missing and others being spurious. In these

common cases, cross-assignment has to rely heavily on the ability of the employed method to distinguish good matchups. In this sense, some experiments are proposed and analyzed to demonstrate the abilities of C4PM, taking 2PM as a reference.

It is important to note that all C4PM experiments described in this section rely on a balanced weight vector to establish a general approach for comparison with the reference method.

Table 1
2PM and u-C4PM progression of the number of matchups against equidistant control values.

l	2PM		u-C4PM	
	R_l	Q_l	r_l	q_l
1	0.2	26213	0.1	7338
2	0.4	28321	0.2	18430
3	0.6	29192	0.3	24112
4	0.8	29654	0.4	27236
5	1.0	30103	0.5	29177
6	1.2	30454	0.6	29915
7	1.4	30649	0.7	30463
8	1.6	30744	0.8	30799
9	1.8	30850	0.9	30919
10	2.0	30956	1.0	30956

4.1. Sensitivity test

Validating C4PM and assessing its performance is a central issue. To this end, we highlight the parallelism between the roles of the critical value (R) in 2PM and the control vector (c) in C4PM. Ultimately, both act as cutoff parameters and, albeit through fundamentally different mechanisms, regulate the distance between partitions matched in each cross-assignment process.

We define the progressive arithmetic sequence of cutoff values as follows:

$$\begin{cases} R_l = \frac{l}{10} R_{\max} & \text{(critical value)} \\ r_l = \frac{l}{10} r_{\max} & \text{(control value)} \end{cases}, \quad (10)$$

for $l = 1, 2, \dots, 10$. The l th experiment in this test involves determining the number of matchups Q_l and q_l generated by the 2PM and u-C4PM⁷ runs, respectively, when initialized with input data R_l and r_l . It can be verified that $R_{\max} = 2$ and $r_{\max} = 1$ (for the selected dataset) correspond, in this order, to the maximum possible number of matchups $Q_{\max}$ and $q_{\max}$, both of which are equal to $N_{\max} = 30,956$ (reached for $l = 10$).

The results are presented in Table 1, which highlight a key distinction in the spectral distance behaviors of C4PM and 2PM. Specifically, the distribution of matchups performed by C4PM exhibits a progressive and scaled increase from the 1st to the 5th metric deciles, followed by a stabilization trend up to the 10th metric decile. In contrast, the distribution of matchups performed by 2PM begins with a very high value in the 1st metric decile, with only a relatively small increase observed up to the 10th metric decile. However, the most significant observation is that Q_1 , the number of match-ups in the first metric decile of 2PM, is remarkably close to q_4 , the number of match-ups in the fourth metric decile of C4PM. This demonstrates that, while the 2PM algorithm fails to distinguish matches based on discrepancies in their wave parameter partitions, the C4PM algorithm, with a nearly equivalent number of matches, is capable of distinguishing four metrically distinct groups of match-ups, namely: the match-ups whose d -distances are at least equal to r_i , but less than r_{i+1} , where $i = 0, 1, 2$ and 3 .⁸ This will become even more apparent in the subsequent sections.

4.2. Accuracy test

As a step toward evaluating the performance of C4PM, this test analyzes the results obtained by running the 2PM and u-C4PM algorithms, each initialized with input values that yield the top 0.3% (equivalent to 99 pairs) of the best matchups generated by each technique (Table 2). The input cutoff values were determined and, as expected, are very small: $R = 4.7 \times 10^{-5}$ (the critical value for 2PM) and $r = 1.0 \times 10^{-2}$

Table 2
Contrasting u-C4PM and 2PM in refined settings.

<i>RMSE</i> : a strict range				
Method	<i>PSWH</i> (m)	<i>PPWP</i> (s)	<i>PPWD</i> (°)	<i>PPWS</i> (°)
2PM	0.18	0.69	13.57	4.91
u-C4PM	0.01	0.00	0.00	0.31

(the control value for u-C4PM). Because the matchups generated by u-C4PM are r -uniformly controlled, it follows that, for all cross-assigned partitioned integrated wave parameters (except for partitioned wave peak directions), the smaller parameter is at least 99% of the larger one. Additionally, the partitioned and cross-assigned wave peak directions differ by no more than 1.8° . In contrast, among the 99 matchups produced by 2PM, 51 pairs exhibit at least one cross-assigned partitioned integrated wave parameter (excluding partitioned wave peak directions) where the smaller parameter is less than 90% of the larger one. For partitioned wave peak directions, the difference in these cases is no less than 18° . These discrepancies are evidently reflected in the RMSEs of the partitioned integrated wave parameters, as summarized in Table 2, and serve to justify the results presented therein. Although the contrasted matchup groups lie within a particularly narrow range, this result highlights a key distinction in the performance of the methods: C4PM successfully prevented the formation of poorly matched partitions, whereas 2PM, even when operating at an extremely low critical value, did not.

4.3. A broader comparison

In this test, the largest set of all possible matchups obtained from the u-C4PM and 2PM runs will be divided into quintiles, and their corresponding *RMSEs* will be presented and analyzed.

To formalize this, let Q_l^* and q_l^* represent the number of matchups produced by the 2PM and C4PM algorithms, respectively, when initialized with the highest possible corresponding input data R_l^* and r_l^* . These cutoff values are uniquely defined so that the equations

$$Q_l^* = q_l^* = \frac{l}{5} N_{\max}, \quad (11)$$

for $l = 1, \dots, 5$ are valid. This scheme enables a fair and equitable comparison between the techniques, as the cutoff values are established to compare groups with an equal number of match-ups and increasing flexibility in their formation.

Table 3 summarizes the performances of 2PM and C4PM across all quintiles. For both methods, the *RMSE* values of all integrated wave parameters do not consistently decrease with each advancing quintile as the cutoff values progress. Notably, these tables partially reflect the results discussed in 4.2. Consider the first quintile of each method. The 6191 matchups generated by C4PM are r_1^* -uniformly controlled with $r_1^* = 0.0851$. Therefore, for all pairs of cross-assigned integrated wave parameters, the smaller value is at least 91% of the

⁷ In uniform mode, the control vector is of the form $c = r(1, 1, 1, 1)$.

⁸ Adopt $r_0 = 0$ for convenience.

Table 3
2PM and u-C4PM performance for quintilian critical values.

<i>l</i>	$Q_l^* = q_l^*$	R_l^*	2PM				r_l^*	u-C4PM			
			RMSE per quintile					RMSE per quintile			
			<i>PSWH</i> (m)	<i>PPWP</i> (s)	<i>PPWD</i> (°)	<i>PPWS</i> (°)		<i>PSWH</i> (m)	<i>PPWP</i> (s)	<i>PPWD</i> (°)	<i>PPWS</i> (°)
1	20%	0.0044	0.22	0.72	13.17	4.59	0.0851	0.11	0.00	9.48	2.23
2	40%	0.0142	0.23	0.76	14.81	5.00	0.1266	0.15	0.72	9.70	2.92
3	60%	0.0415	0.26	0.81	17.11	5.63	0.2050	0.20	0.74	13.28	4.05
4	80%	0.1420	0.29	0.92	20.24	6.81	0.3258	0.26	0.90	17.08	5.78
5	100%	2.0000	0.32	1.80	34.74	9.22	1.0000	0.31	1.65	34.71	8.50

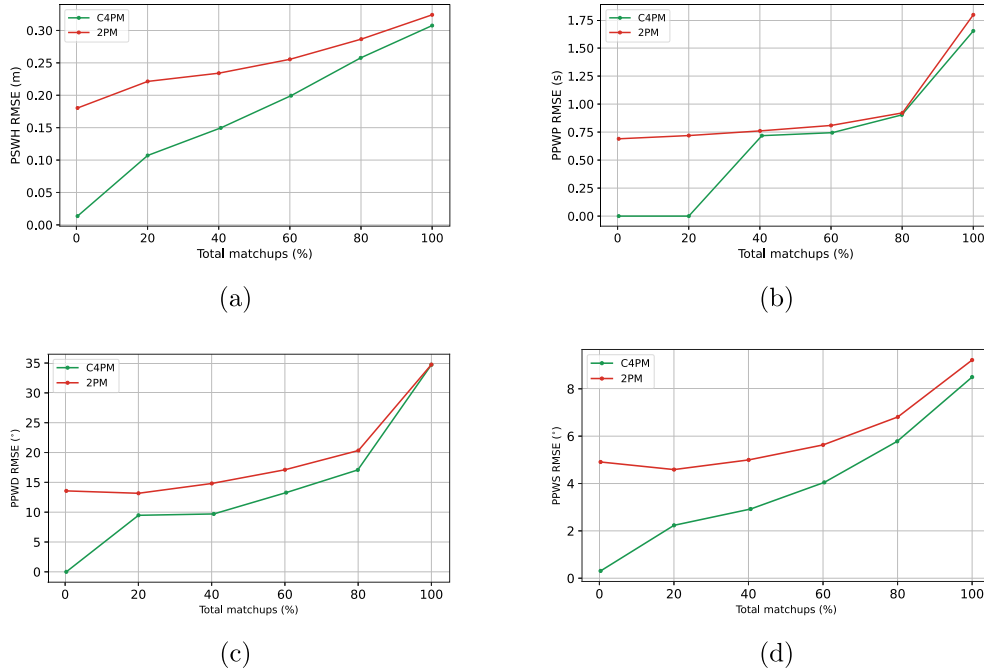


Fig. 8. *RMSEs*: (a) *PSWH*, (b) *PPWP*, (c) *PPWD* and (d) *PPWS*.

larger one, except for cross-assigned partitioned peak wave directions, whose differences do not exceed 15.4° . In contrast, these conditions do not hold for the 6191 matchups corresponding to the first quintile under 2PM, as evidenced by their higher *RMSE* values. Specifically, 3647 matches include at least one cross-assigned partitioned integrated wave parameter where the smaller value is less than 91% of the larger one, excluding wave peak directions, which differ by at least 15.4° . This discrepancy accounts for the higher *RMSE* values observed for 2PM compared to C4PM within this quintile.

An analogous analysis can be performed for subsequent quintiles, yielding similar conclusions, while noting that the *RMSE* values become increasingly similar as the quintiles progress. The additional, more visual representation of the evolution of *RMSE* values against the percentage of matchups for each experiment — shown in Fig. 8 — further corroborates this observation. In particular, the figure demonstrates that C4PM outperformed 2PM in forming matches, regardless of the percentage of total matches considered. In other words, C4PM consistently produced more accurate matchups.

Fig. 9 illustrates the *PPWP* evolution for C4PM and 2PM across the first and second quintiles. In the first quintile, C4PM exhibits a concentrated distribution along the diagonal, indicative of accurate matches and the absence of outliers. Conversely, 2PM demonstrates a more dispersed distribution, with significant outliers reflecting lower match quality. In the second quintile, this trend persists. C4PM continues to maintain a narrow distribution along the diagonal, highlighting its ability to generate reliable matches. In contrast, 2PM displays a broader distribution accompanied by many outliers, further emphasizing its inferior match quality. These observations align with the *RMSE* values

reported in Table 3. A lower global *RMSE* generally corresponds to improved match quality and fewer outliers, particularly under stricter matching conditions. The consistent superior performance of C4PM in both quintiles underscores its robustness and reliability in producing accurate matches, even under challenging scenarios.

4.4. Two settings confronted

The results obtained from running the 2PM and C4PM algorithms on relatively broad input data are compared. This test does not aim to identify equivalent configurations between the two methods — if such equivalence is even possible — but rather to analyze the behavior of C4PM and the results it produces. Notably, this comparison underscores one of C4PM's key capabilities: its ability to independently control discrepancies between cross-assigned partitioned integrated wave parameters of different natures, depending on the application, type of measurement, or desired level of accuracy.

Following Hasselmann et al. (1996), $R = 0.75$ is used as the critical input value for the 2PM algorithm, while the control vector $\mathbf{c} = (0.2, 0.3, 0.2, 0.6)$ is applied as input to the C4PM algorithm. This control vector imposes the following limits on parameter discrepancies within a matchup: the smallest significant wave height must be at least $(1 - 0.2) = 0.8$ of the largest significant wave height in the same matchup; similar limitations apply to discrepancies in peak wave periods and peak directional spreads. Additionally, cross-assigned peak wave directions cannot differ by more than $0.2 \times 180^\circ = 36^\circ$. Fig. 10 illustrates the results of this comparison. Both methods exhibit negligible bias, high correlation coefficients, and low *RMSE* for all partitioned

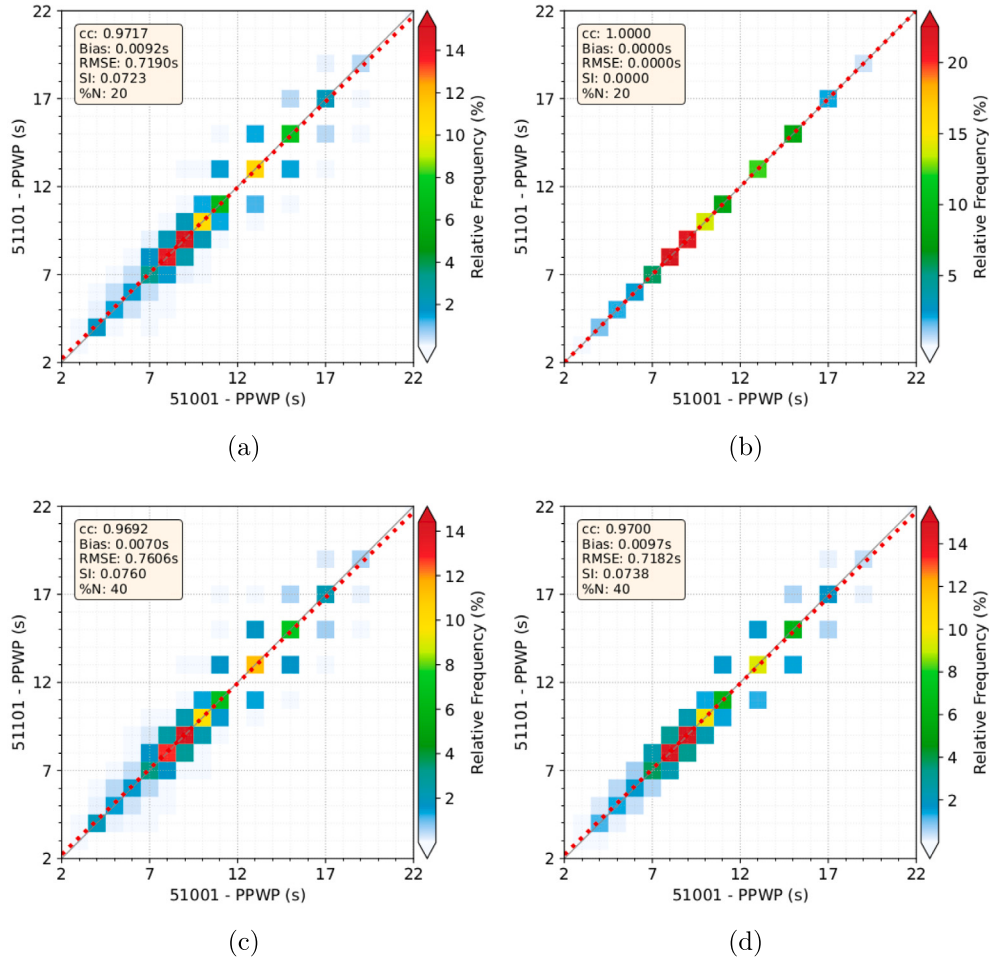


Fig. 9. PPWP scatter plot for the first two quintiles. Left column, 2PM [ac]; right column C4PM [bd]. First line is the first quintile [ab]; second line is the second quintile [cd]. The red dotted line indicates the best fit from a linear regression using the least-squares method.

integrated wave parameters. However, closer examination of Fig. 10[a, c, e, g], which displays the 2PM results, reveals the presence of outliers, indicating the occurrence of unlikely matchups. Such outliers, while rare, may pose challenges in applications such as data assimilation. Notably, these outliers persist even when the critical value for 2PM is significantly reduced (plots not shown). In contrast, Fig. 10[b, d, f, h], presenting the C4PM results, shows matchups tightly clustered around the 1:1 line, indicating the absence of outliers. This result highlights the effectiveness of C4PM in controlled matchup formation, further demonstrating its superiority in avoiding spurious matches.

5. Summary and conclusions

Cross-assignment of wave spectra represents a combinatorial optimization challenge in physical oceanography. The core question is how to determine the best matching between partitions of wave spectra from different sources, specifically those whose oceanic characteristics exhibit the highest level of concordance. Addressing this problem requires a reliable method for measuring the similarity between partitions and a robust criterion to determine when pairs of partitions should form the cross-assignment.

This work proposes a novel cross-assignment methodology based on a spectral distance that incorporates four integrated wave parameters, termed the Controlled Four-Parameter Method (C4PM), which offers a high degree of customizability, enabling the assignment of distinct weights to the integrated wave parameters during the distance calculation while also allowing for *a priori* control of discrepancies among the corresponding parameters throughout the cross-assignment process.

To evaluate the proposed method, we compared C4PM with an existing approach based on a two-parameter spectral distance (denoted as 2PM). Thousands of data points from two buoys located 13 km apart were analyzed. The experiments explored the progression of cutoff values for each technique and their qualitative impact on the resulting groups of matchups. The results consistently demonstrated the superiority of C4PM over 2PM in nearly all aspects.

The weighted semimetric underlying C4PM proved to be well-scaled, providing superior capability in identifying matches with highly concordant oceanic features. In contrast, the spectral distance employed in 2PM exhibited limitations, allowing for the formation of mismatched pairs in numerous cases. When operating with very strict cutoff values, C4PM identified only perfect matches, i.e., measurements from nearly identical pairs of partitions, while 2PM produced a significant proportion of relatively poor matches even within the same cutoff range.

Further analysis, which divided the total set of possible matches into quintiles based on cutoff values, revealed that C4PM consistently outperformed 2PM across all groups, particularly in the first three quintiles. The *RMSE* values achieved by C4PM for each partitioned integrated wave parameter were consistently lower than those of 2PM, highlighting the accuracy of its matchups. This distinction was further underscored by the presence of outliers in the 2PM results, which persisted even when stricter cutoff values were applied. These outliers, indicative of poorer match quality, were absent in the C4PM results, demonstrating the robustness of its controlled matchup formation.

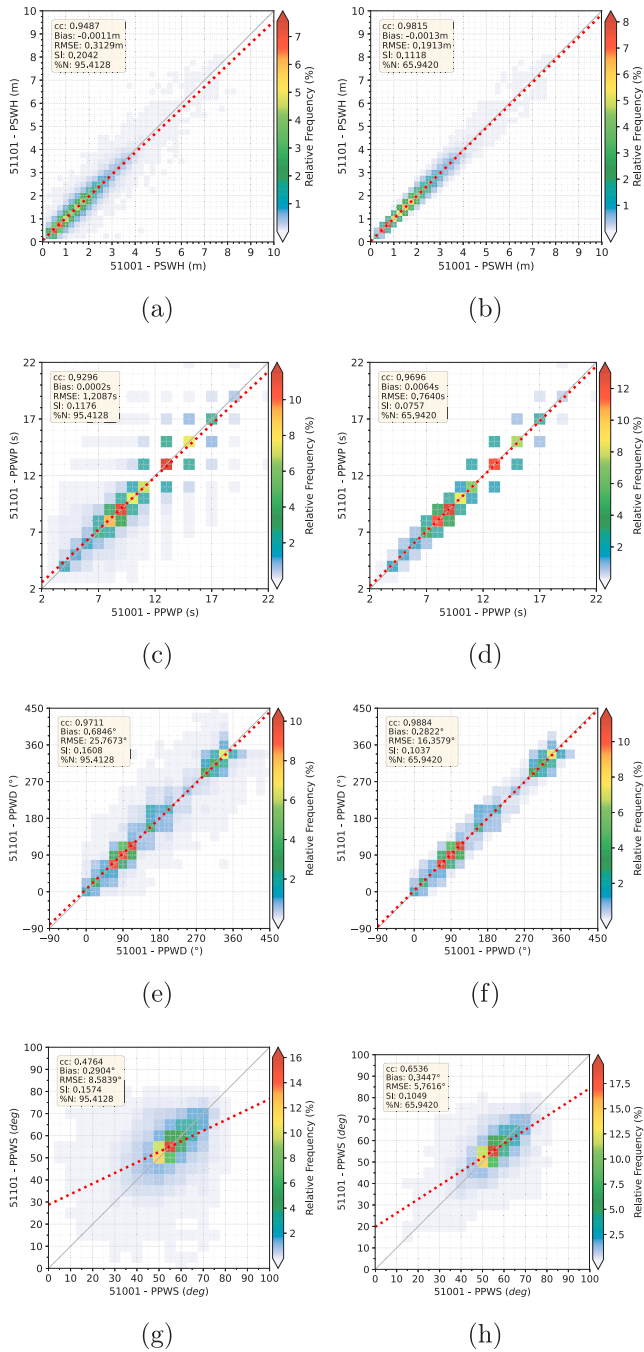


Fig. 10. Scatter plots between buoys 51001 and 51101 for *PSWH*, *PPWP*, *PPWD* and *PPWS*, respectively from top to bottom. Left column: 2PM; right column: C4PM. The red dotted line indicates the best fit from a linear regression using the least-squares method.

C4PM's ability to assign specific weights and deviation limits to each parameter provides a level of flexibility that enables tailored cross-assignments to suit the requirements of different datasets and applications. Furthermore, its negligible computational overhead and planned release as an open-source Python package enhance its accessibility and practicality for broader use. In summary, the results presented here demonstrate significant improvements over traditional spectral distance methods for the cross-assignment of directional wave spectra. C4PM is a highly effective tool for wave data assimilation, where precise cross-assignments are critical for improving the reliability and accuracy of forecast models.

Table A.1

Statistical parameters.

Parameter	Expression
Mean	$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i$
Bias	$\frac{1}{n} \sum_{i=1}^n (y_i - x_i)$
Root mean square error (<i>RMSE</i>)	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$
Scatter index (<i>SI</i>)	$\frac{(RMSE)}{\bar{X}}$
Pearson correlation coefficient (<i>cc</i>)	$\frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{Y})^2}}$

Table A.2

Wave parameters. $S(f, \theta)$ represents the directional wave spectrum.

Parameter	Expression
<i>SWH</i>	$4\sqrt{\int_0^{2\pi} \int_0^\infty S(f, \theta) df d\theta}$
<i>PWP</i>	$\frac{1}{\text{argmax}(S(f))}$
<i>PWD</i>	$\text{atan2}\left(\int_0^{2\pi} S(f_p, \theta) \sin(\theta) d\theta, \int_0^{2\pi} S(f_p, \theta) \cos(\theta) d\theta\right)$
<i>PWS</i>	$\sqrt{2 \left(1 - \sqrt{\left(\frac{\int_0^{2\pi} S(f_p, \theta) \cos(\theta) d\theta}{\int_0^{2\pi} S(f_p, \theta) d\theta} \right)^2 + \left(\frac{\int_0^{2\pi} S(f_p, \theta) \sin(\theta) d\theta}{\int_0^{2\pi} S(f_p, \theta) d\theta} \right)^2} \right)}$
<i>CWV</i>	$\frac{4\pi^2}{g} \left(\int_0^\infty \int_0^{2\pi} f S(f, \theta) \cos \theta d\theta df, \int_0^\infty \int_0^{2\pi} f S(f, \theta) \sin \theta d\theta df \right)$

CRedit authorship contribution statement

Andre Luiz Cordeiro dos Santos: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Felipe Marques dos Santos:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nelson Violante-Carvalho:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Luiz Mariano Carvalho:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Helder Manoel Venceslau:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis, Conceptualization.

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.

Appendix A. Statistics and wave parameters

Tables A.1 and A.2 provide the formulas for the statistical parameters and wave parameters, respectively.

Data availability

Data will be made available on request.

References

- Aouf, L., Hauser, D., Chapron, B., Toffoli, A., Tourain, C., Peureux, C., 2021. New directional wave satellite observations: Towards improved wave forecasts and climate description in Southern Ocean. *Geophys. Res. Lett.* 48, e2020GL091187. <http://dx.doi.org/10.1029/2020GL091187>.
- Aouf, L., Lefevre, J.-M., Hauser, D., 2006. Assimilation of directional wave spectra in the wave model WAM: An impact study from synthetic observations in preparation for the swimsat satellite mission. *J. Atmos. Ocean. Technol.* 23, 448–463. <http://dx.doi.org/10.1175/JTECH1861.1>.
- Ardhuin, F., Stopa, J.E., Chapron, B., Collard, F., Husson, R., Jensen, R.E., Johannessen, J., Mouche, A., Passaro, M., Quartly, G.D., Swail, V., Young, I., 2019. Observing sea states. *Front. Mar. Sci.* 6, <http://dx.doi.org/10.3389/fmars.2019.00124>.
- Cavaleri, L., Alves, J.-H., Ardhuin, F., Babanin, A., Banner, M., Belibassakis, K., Benoit, M., Donelan, M., Groeneweg, J., Herbers, T., Hwang, P., Janssen, P., Janssen, T., Lavrenov, I., Magne, R., Monbaliu, J., Onorato, M., Polnikov, V., Resio, D., Rogers, W., Sheremet, A., McKee Smith, J., Tolman, H., van Vledder, G., Wolf, J., Young, I., 2007. Wave modelling — the state of the art. *Prog. Oceanogr.* 75, 603–674. <http://dx.doi.org/10.1016/j.poccean.2007.05.005>.
- Devaliere, E.-M., Hanson, J.L., Luettich, R., 2009. Spatial tracking of numerical wave model output using a spiral search algorithm. In: 2009 WRI World Congress on Computer Science and Information Engineering. <http://dx.doi.org/10.1109/CSIE.2009.1021>.
- Earle, M.D., Steele, K., Wang, D.W., 1999. Use of advanced directional wave spectra analysis methods. *Ocean Eng.* 26, 1421–1434.
- Gerling, T.W., 1992. Partitioning sequences and arrays of directional ocean wave spectra into component wave systems. *J. Atmos. Ocean. Technol.* 9, 444–458.
- Hanson, J.L., Phillips, O.M., 2001. Automated analysis of ocean surface directional wave spectra. *J. Atmos. Ocean. Technol.* 18.
- Hanson, J.L., Tracy, B.A., Tolman, H.L., Scott, R.D., 2009. Pacific hindcast performance of three numerical wave models. *J. Atmos. Ocean. Technol.* 26, 1614–1633. <http://dx.doi.org/10.1175/2009JTECHO650.1>.
- Hasselmann, S., Brüning, C., Hasselmann, K., Heimbach, P., 1996. An improved algorithm for the retrieval of ocean wave spectra from synthetic aperture radar image spectra. *J. Geophys. Res. Ocean.* 101, 16615–16629. <http://dx.doi.org/10.1029/96JC00798>.
- Hauser, D., Tourain, C., Hermozo, L., Alraddawi, D., Aouf, L., Chapron, B., Dalphin, A., Delaye, L., Dalila, M., Dormy, E., Gouillon, F., Gressani, V., Grouazel, A., Guitton, G., Husson, R., Mironov, A., Mouche, A., Ollivier, A., Oruba, L., Piras, F., Rodriguez Suquet, R., Schippers, P., Tison, C., Tran, N., 2021. New observations from the SWIM radar on-board CFOSAT: Instrument validation and ocean wave measurement assessment. *IEEE Trans. Geosci. Remote Sens.* 59, 5–26. <http://dx.doi.org/10.1109/TGRS.2020.2994372>.
- Jiang, H., Mironov, A., Ren, L., Babanin, A.V., Wang, J., Mu, L., 2022. Validation of wave spectral partitions from SWIM instrument on-board CFOSAT against in situ data. *IEEE Trans. Geosci. Remote Sens.* 60, 1–13. <http://dx.doi.org/10.1109/TGRS.2021.3110952>.
- Kuik, A.J., van Vledder, G.P., Holthuijsen, L.H., 1988. A method for the routine analysis of pitch-and-roll buoy wave data. *J. Phys. Oceanogr.* 18, 1020–1034. [http://dx.doi.org/10.1175/1520-0485\(1988\)018<1020:AMFTRA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0485(1988)018<1020:AMFTRA>2.0.CO;2).
- Li, J.-G., Saulter, A., 2012. Assessment of the updated envisat ASAR ocean surface wave spectra with buoy and altimeter data. *Remote Sens. Environ.* 126, 72–83. <http://dx.doi.org/10.1016/j.rse.2012.08.018>.
- Portilla-Yandún, J., Cavaleri, L., Van Vledder, G.P., 2015. Wave spectra partitioning and long term statistical distribution. *Ocean. Model.* 96, 148–160. <http://dx.doi.org/10.1016/j.ocemod.2015.06.008>.
- Portilla-Yandún, J., Valladares, C., Violante-Carvalho, N., 2019. A hybrid physical-statistical algorithm for SAR wave spectra quality assessment. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 12, 3943–3948. <http://dx.doi.org/10.1109/JSTARS.2019.2944115>.
- Ricondo, A., Cagigal, L., Rueda, A., Hoeke, R., Storlazzi, C.D., Méndez, F.J., 2023. Hywaves: Hybrid downscaling of multimodal wave spectra to nearshore areas. *Ocean. Model.* 184, 102210.
- Santos, F.M., Santos, A.L., Violante-Carvalho, N., Carvalho, L.M., Brasil-Correa, Y.O., Portilla-Yandun, J., Romeiser, R., 2021. A simulator of synthetic aperture radar (SAR) image spectra: The applications on oceanswell waves. *Int. J. Remote Sens.* 42, 2981–3001. <http://dx.doi.org/10.1080/01431161.2020.1847352>.
- Smit, P., Houghton, I., Jordanova, K., Portwood, T., Shapiro, E., Clark, D., Sosa, M., Janssen, T., 2021. Assimilation of significant wave height from distributed ocean wave sensors. *Ocean. Model.* 159, 101738.
- Violante-Carvalho, N., Ocampo-Torres, F., Robinson, I., 2004. Buoy observations of the influence of swell on wind waves in the open ocean. *Appl. Ocean Res.* 26, 49–60.
- Violante-Carvalho, N., Robinson, I.S., Schulz-Stellenfleth, J., 2005. Assessment of ERS synthetic aperture radar wave spectra retrieved from the Max-Planck-Institut (MPI) scheme through intercomparisons of 1 year of directional buoy measurements. *J. Geophys. Res. Ocean.* 110, <http://dx.doi.org/10.1029/2004JC002382>.
- Wang, X., Husson, R., Jiang, H., Chen, G., Gao, G., 2020. Evaluation on the capability of revealing ocean swells from Sentinel-1A wave spectra measurements. *J. Atmos. Ocean. Technol.* 37, 1289–1304. <http://dx.doi.org/10.1175/JTECH-D-19-0060.1>.
- Wu, L., Sahlée, E., Nilsson, E., Rutgersson, A., 2024. A review of surface swell waves and their role in air–sea interactions. *Ocean. Model.* 190, 102397.