Reviewer #3

General comments

This manuscript presents a new dataset created using Synthetic Aperture Radar (SAR) altimetry data from Sentinel-3A/B and CryoSat-2. Thirteen years of the SAR data are analyzed to retrieve measurements of the significant wave height (Hs), zero crossing wave period (Tz) and wave power density (WPD) around eleven oceanic and coastal regions.

The high resolution data is retrieved using the SAMOSA+ retracker and empirical functions. The new dataset contains crucial information for the assessment of the wave parameters, needed for marine spatial planning, offshore renewable energy development and maritime safety.

Although the manuscript presents a new dataset with products of SAR observation, it lacks details on the methodology of data retrieval and details on limitations and error estimates regarding the two different satellites used.

Below the comments and questions on the manuscript.

Suggested technical corrections

1. *Line 14-17: Suggest rewriting the sentence to make it easier to follow and more clear: " Marine renewable energy is a promissing source for countries with extensive maritime zones. This is due to a growing interest in renewable energy sources for electricity generation to fight climate change, rising global energy consumption, competition for land use (agriculture, urban planning, and protected areas), and high population density."*

Fixed

2. *Line 46-47: Suggest rewriting the sentence to make it easier to follow and more clear: "This article describes the method used to determine the wave period, the validation of the Hs and Tz obtained from satellite altimetry, the structure of the final dataset and finally the applications of the wave power density dataset."*

Fixed

3. Line 98: Suggest rewriting the sentence to make it more coherent and clear: " with a Root Mean Square Error (RMSE) of 0.22 m and a correlation coefficient (CC)".

Fixed

Line 124-129: Suggest rewriting this list of the satellite's data availability or create a table for better visualization of the information.

The table was added:

Variable name and description	Units
Geodetic Latitude at 20 Hz	degree north
Geocentric Longitude at 20 Hz	degree east
Time of Surface Reflection Point at 20 Hz, seconds since 01-Jan-2000 00:00:00 UTC	seconds
Significant Wave Height at 20 Hz	meters
Sigma Nought at 20 Hz (corrected for atmospheric attenuation)	dB
Quality flag: 1 – labeled as an outlier, 0 – labeled as non-outliers	-
Zero-crossing wave period calculated based on SWH and Sigma0	seconds
Wave Power Density per Unit Crest at 20 Hz	W/meter
Distance to Coastline at 20 Hz	meters
Misfit between Power Waveform Model and Power Waveform Data at 20 Hz	count

Table 2. The variables provided in the WAPOSAL dataset.

5. Line 65: Clarify and explain with more detail what is the "misfit".

In the revised version of the paper, we added the explanation below:

In the retracking procedure, the measured reflected radar waveform is fitted to a theoretical model, and several parameters, including the significant wave height and sigma^0, are estimated. The 'misfit' quantifies the discrepancy between the measured waveform and the fitted model waveform. Therefore, a larger 'misfit' indicates poorer agreement between the observations and the model used in the retracking procedure. The dataset contains information on the 'misfit' between the model and the measured altimeter waveform, which was later used for quality control.

6. Line 109: The reference "(Cahill 2012)" is not present in the list of references. Another article is present in the references as (Cahill 2014)

Fixed

Specific comments

1. Line 43: Mentions that "the data is validated against in-situ wave buoys and/or ERA5 data." suggestion to explain what type of validation technique

is applied to the data, and highlight how the data is divided between validation with ERA5 and validation with buoys.

For each buoy and each ERA5 node, the collocated dataset was randomly partitioned into two equal subsets: a training set and a validation set. The training set was used to estimate the regression coefficients a and b, while the validation set was employed to evaluate model performance in terms of bias, scatter index, and correlation coefficient. This procedure was repeated 1000 times to obtain a stable estimate, and the resulting ensemble was used to compute averaged estimates of both the regression coefficients and the statistical metrics. We followed the methodology outlined in Mackay (2008, https://doi.org/10.1029/2007JC004438).

We have added this information to the text.

2. Line 53: Mentions that the "data in high resolution Synthetic Aperture Radar (SAR) mode were considered and reprocessed." Suggestion to describe how much data has been used for each location, if there is a minimum threshold of number of images from SAR that are needed for the location to be able to determine the wave height, wave period and wave energy density.

We are thankful for an important comment. Indeed, for several locations, the number of collocated samples was insufficient. Thus, we used only those locations with more than 100 points for further processing. This information has been added to the text. For several locations, the number of collocated samples is >2500 points. We have performed linear regression for each buoy separately.

3. Line 74-77: The method presented by Gommenginger et al. in 2003 calculates the zero crossing wave period for deep water, suggestion to consider and explain limitation on how that applies to the selected buoys and ERA5 points' data since the area of interest is a coastal area. Or state whether the selected buoys and ERA5 points are located in deep water.

We are grateful for this important comment on the applicability of the deep-water formulation for wave power density in a study focused on coastal regions. We fully agree with the Reviewer's comment. The deep-water formulation is strictly valid only when the water depth exceeds half the wavelength. In the revised manuscript, we have clarified this assumption and explicitly discussed its implications. We acknowledge that for longer-period swell waves, the deep-water assumption may not be strictly valid in parts of the shelf, where intermediate-depth conditions can occur. This may lead to a

moderate overestimation of wave power density when the deep-water formulation is used. We have therefore revised the manuscript to explicitly state the deep-water assumption and discuss the potential bias introduced by neglecting finite-depth effects. Future work will address this limitation by implementing finite-depth formulations of wave power density alongside high-resolution bathymetry and numerical wave models, enabling a more accurate assessment near the coastline.

Moreover, suggestion to describe more thoroughly how the coefficients a and b are calculated.

The coefficients were calculated using linear regression for X and Tz, we have added this phrase to the text:

Therefore, following Gommenginger et al. (2003), we have re-evaluated, for each collocated data set, the coefficients (a and b) of the following equation using linear regression, with collocated Tz in situ wave buoy data and altimeter Hs and σ0 data.

$Tz = aX + b$ (1)

4. Line 87: Mentions "the selected ERA5 grid nodes" without assessing how and how many grid nodes are selected. Did you select the closest node to the location? Did you consider all the nodes in the radius of 40 km from the location? Or did you interpolate the values of the closest nodes? In addition, how many nodes for location did you use? And does this number change depending on the locations?

ERA5 data were used as a substitute for buoy observations. For each area of interest, several ERA5 grid points within the coastal zone were selected for analysis. An example for the Madeira region is shown in Figure 1, where 12 grid points were selected. The number of selected grid points varied by region (6–12 points for the French Guiana, Madeira, and Canary Islands regions), depending on the size and spatial extent of the study area. The grid points were selected to provide representative coverage of the coastal zone. No spatial interpolation was applied; instead, ERA5 data from each selected grid point were used directly. Time series of significant wave height and wave period were extracted from the ERA5 dataset and then processed in the same manner as the wave buoy observations.

The information is added to the text.

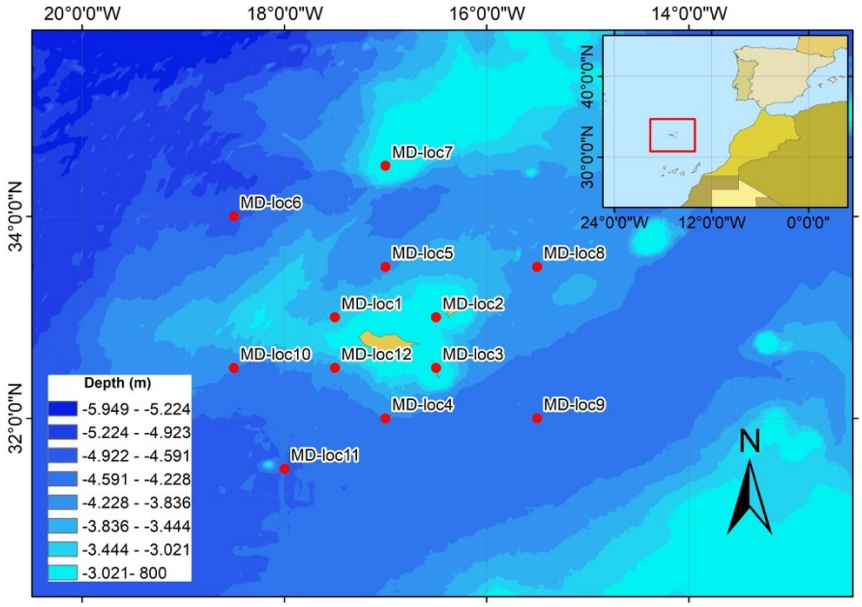

Figure1. The locations of the twelve ERA5 grid points in the region surrounding the Madeira Archipelago, whose data were used to obtain the linear regression coefficients.

5. Line 89/90: Mentions that "the time interval was 45 minutes, and the spatial distance was 40 km", how were these spatial and temporal conditions for the data collocation selected? Did you follow any previous study done or is there any reference in literature? Did you consider the possibility of the results changing if you selected different spatial and temporal conditions?

In Gommenginger (2003, https://doi.org/10.1029/2003GL017743), the collocation criteria are 100 km in space and 1 hour in time. In Mackay (2008, https://doi.org/10.1029/2007JC004438), the spatial collocation criterion was 100 km and the temporal criterion was 30 minutes. However, the authors used NDBC buoy data, and these buoys are located farther from the coast. The buoys from the network around Europe that we used are located mostly in the coastal zone; therefore, we have limited the spatial distance to 40 km to ensure higher homogeneity of the data around the buoy, and the collocation time was set to less than 1 hour, on the one hand, to capture the variability of wave parameters, and on the other hand, to accumulate sufficient statistics.
We acknowledge that different spatial and temporal thresholds could lead to somewhat different validation statistics. However, a sensitivity analysis of the collocation criteria was beyond the scope of the present study. Therefore, while we recognize that the results may vary depending on the selected criteria and local conditions, we have limited our analysis to the

adopted collocation scheme and have clarified this limitation in the revised manuscript.

6. Line 93: Assesses "training and validation" which was never mentioned before therefore it is suggested to discuss this part of the method in more detail, specifically describing how the datasets were divided, if they had the same or different sizes, and if there are any constraints on the selection of the data. It would be extremely useful to have more quantitative data on this part of the methodology.

The training and validation datasets are similar in size: we use one half of the dataset for training and the other for validation. The dataset is randomly split, and this procedure was repeated 1000 times to obtain a stable estimate, and the resulting ensemble was used to compute averaged estimates of both the regression coefficients and the statistical metrics as discussed in (Mackay, 2008, https://doi.org/10.1029/2007JC004438).

The information is added to the revised manuscript.

7. Line 97: States that " for CryoSat-2, similar results are obtained", without giving any type of numerical value of the error metrics like RMSE or CC. It is suggested to add the graph of the CryoSat-2 results, or at least report the numerical qualitative data as done for Sentinel-3A/B data. It is also suggested to integrate information about differences between the Sentinel-3A/B and CryoSat-2, and where such differences may come from, biases or differences in the satellite sensors swath.

We have added the information on the error metrics for Cryosat-2: For Hs, bias=0.02 m, RMSE=0.24 m, CC=0.98; for Tz, bias = -0.0003 s, RMSE=0.57 s, CC = 0.9.

8. Line 116/117: mentions "observation time in seconds since 01-01-2000 for each antenna footprint', without specifying the time zone taken as reference for the observation time data.

The time is indicated in the UTC time zone, and the information is added to the text.

9. Line 169: States " The satellite-derived WPD closely follows the ERA5 observations.", without discussing any numerical value nor quantitative data. What is the bias and RMSE between the satellite derived WPD and the ERA5 observation?

Wave power density derived from CryoSat-2 was compared with ERA5 observations, yielding a bias of 3.5 kW/m and RMSE of RMSE=9.9 kW/m. For Sentinel-3A/B, the corresponding bias was also 3.5 kW/m, while the RMSE was slightly lower at 9.7 RMSE=9.7 kW/m. These results indicate good agreement between the satellite-derived and ERA5 wave power density estimates.

This information is added to the text.