On the spatial interpolation of ocean energy source variables: A comparative analysi

On the spatial interpolation of ocean energy source variables: A comparative analysis

Leonardo Gambarelli, Edoardo Pasta, and Giuseppe Giorgi

***Abstract*—In the context of wave energy systems development, the estimation of wave parameters, such as the energy period (T_e), over the entire ocean surface is of paramount importance. These information are crucial for estimating the energy harvesting potential of deployment sites, designing wave energy converters (WECs), and planning optimal maintenance intervention frequency. However, measuring T_e at every point in the ocean is impossible due to the vastness of the ocean and due to the cost and difficulties of installing and maintaining wave instrumentation buoys, since these have to survive in marine environment. As a consequence, the amount of data available is too limited and sparse in space, making it impractical to perform these analyses with precision. To address such data scarcity and sparsity, we analyse in this paper various spatial interpolation techniques employed to fill the spatial gaps in the wave parameter datasets. Three types of interpolators are considered: linear interpolator, spline interpolator, and radial basis functions (RBFs) interpolator. These algorithms are trained and tested on a public dataset of wave parameters from Copernicus Marine Service in an area between the coastlines of South England and North France. To simulate the available data scarcity and sparsity, only limited percentages of the ocean area are considered covered and available in the training stage (from 1% to 5%). The performance of each interpolator is evaluated in terms of Normalized Root Mean Square Error (NRMSE) achieved by the algorithm in reconstructing the parameters at the unsampled locations. The results of this study demonstrate the feasibility of spatial gap-filling of wave parameter, and demonstrates that the RBF algorithm outperforms the other two algorithms in terms of robustness to different training points sampling, when working with low training percentages.**

***Index Terms*—Spatial interpolation, Ocean wave energy, Radial basis functions**

I. Introduction

WAVE energy is one of the most promising source of energy and it is supposed to play a key role for energy production in future years. Wave energy will most likely contribute significantly in moving away from fossil fuels [1]. This source is interesting for a number of reasons like its high predictability and availability [2], compared to the other renewable resources. Another interesting characteristic is its high energy potential, theoretically being able to produce almost double the current world energy demand each year [3]. Moreover, the Communication from the European Commision of November 2020 [4] sets the ambitious targets of installed capacity of at least 1 GW of ocean (wave and tidal) energy by 2030, and 40 GW, by 2050. Up to today, the main obstacle towards the commercialization of wave energy is its way too high Levelised Cost of Energy (LCoE) [5]–[7]. This high LCoE is given by a number of factors, like the lack of a suitable control system [8], the absence of a unique design optimal for all possible scenarios [9], the high cost for maintenance of the devices [10] and the scarcity of data regarding wave parameters worldwide [11], which are mainly gathered through in-situ measurements with buoys [12]. The latter problem in particular makes the assessment of a specific location difficult, increasing the uncertainties related to the amount of energy that a particular device would produce in a specific location [13]. A precise and accurate location assessment is of paramount importance for the economical viability of wave energy because it makes possible to choose the most optimal location for energy harvesting, it helps in indicating which of the WECs design is more suited for that specific location and it gives an idea of the kind of control system suited for that specific setup. For this reasons, the ideal scenario would be that of filling the ocean everywhere with measuring devices, but this is practically impossible due to how vast the ocean actually is and how costly these devices are, not to mention the problems related to their disposal. Indeed these devices have both an high capital investment cost, being pretty sophisticated devices, and also an high operative cost, being devices that have to survive in marine environment, which is a particularly hostile environment. Such type of measuring devices require constant maintenance interventions. All these factors contribute to the fact that sea wave measures are pretty expensive, and from this arises the need of exploiting these measures as much as possible, in order to cover the maximum amount of measured space with the minimum amount of installed devices. However, extending these measurements spatially in the highly nonlinear system of the ocean is a demanding task.This task is typically done through hindcasting with data assimilating (DA) techniques, where real measurements of the parameter of interest are used to correct and update numerical models simulating the system. The drawback is that, in order to be accurate, these numerical models require huge amounts of computational time for running.

This paper would like to address this problem of the sparsity of the data using data-driven techniques, as

This work is part of the project NODES which has received funding from the MUR – M4C2 1.5 of PNRR with grant agreement no. ECS00000036.

L. Gambarelli, E. Pasta, and G. Giorgi are with *Marine Offshore Renewable Energy Lab.*, DIMEAS, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Turin, Italy. (e-mail: leonardo.gambarelli@polito.it).

Digital Object Identifier:
https://doi.org/10.36688/ewtec-2023-619

it is already being done for other complex, non-linear problems related to other energy sources [14]. The rest of this article is structured in the following way: Section 2 presents the various devices employed for recording wave measurements, along with the problematics that regard them, with a focus on the need of interpolating and extrapolating those measurements; Section 3 presents the three interpolation techniques that are being used in this article for interpolating wave parameters; Section 4 explains in greater detail how those algorithm are set up for this specific application; Section 5 presents the results obtained from the algorithms; Section 6 makes the final conclusion that can be derived from the results.

II. Measuring wave energy parameters

A comprehensive review of all the various instruments used for measuring sea waves is proposed in [15] and in [16]. Although recent technological improvements make possible the use of satellite altimeters and other remote sensing instruments, they introduce other problems, like cloud coverage [17] or their inability to measure pressure [18]. Up to today, the majority of the sea wave recordings is made by wave buoys installed and measuring only in-situ. All measuring devices can be grouped into three major classes:

- Devices measuring below the sea surface: hidden and protected from eventual disturbances at the surface, they measure the pressure under the wave (current meters, acoustic profilers).
- Devices measuring at the sea surface: the most commonly used, usually consisting in an measuring instrument attached to a buoy that records wave elevation (wave buoys, wave staffs).
- Devices measuring above the sea surface: the most recent ones, contactless, they are able to cover a larger area with their measurements but those measurements are less accurate with respect to the in-situ ones (satellites, altimeters, LIDAR).

The below sea and at the sea instruments usually convert the mechanical signal (pressure/elevation) into an electric one with a transducer before recording it, while the above the sea ones usually take a photo of the sea and after some post-processing are able to output a number of interest. Even if many of these devices are actually deployed worldwide in the oceans, due to how vast the sea is, the big part of it is still unrecorded. Indeed this measuring devices are sparse in space, and so the measures of the sea parameters are also sparse in space. This implies the need of extrapolating and interpolating those recorded data into places where there is actually no measuring device.

III. Using Spatial interpolation techniques

In order to address the problem of the sparsity in space of wave measurements, spatial interpolation techniques offer an handy tool for artificially extending the available measures into unmeasured locations. Three spatial interpolation techniques are presented here:

- Linear interpolator: the most simple interpolator, which, over not uniformly gridded data, first performs triangulation of the scattered sample points and, after that, interpolates over the points between each triangle using the vertices of that triangle. There are several variants of this method depending on the policy used for interpolating and extrapolating. These can be linear (interpolate/extrapolate the points linearly with respect to the sampled ones) or to the nearest (interpolate/extrapolate the points to the nearest to them in Euclidean distance). Within the specific application of this paper, only the setup with linear interpolation and linear extrapolation is considered. The MATLab™ command `scatteredinterpolant` is used to perform this interpolation.
- Thin plate spline interpolator: a spline-based interpolation technique greatly used in image processing and data mining [19]. It is called thin plate spline since it can be imagined as a thin sheet of metal which is forced to pass to some specific points, the sampled ones. Mathematically speaking, it is an optimally interpolating surface: between all the surfaces that pass through the sampling points, this surface is the one that minimizes the overall bending energy of the surface itself. However, for real applications, the exact passage over a set of sampled points is a strong condition. Better results can be obtained by relaxing this condition and allowing the interpolator to pass nearby these points still minimising the bending energy: a trade-off parameter p is used to regulate this trade-off. For n training points, the overall problem can be casted in terms of an optimization problem where the functional to be minimized has the form of:

$$E(f) = p\sum_{j=1}^{n} ||y_j - f(x_i)||^2 + \\ + (1-p)\iint [\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_1 \partial x_2} + \frac{\partial^2 f}{\partial x_2^2}] dx_1 dx_2. \tag{1}$$

  In the formula above, the sum represents the error with respect to the training points (the sampled ones) while the integral quantifies the overall bending energy of the surface. The MATLab™ command `tpaps` is being used to perform this interpolation, which requires the parameter p as an additional input.
- Radial basis functions (RBF) interpolator: a technique for interpolating, used in cartography and medical imaging, particularly interesting for scattered and multidimensional data [20]. Broadly speaking, radial basis function is a general function approximator. Given n sampled points, an RBF interpolator can be built from them as:

$$s(x) = \sum_{j=1}^{n} \lambda_j \phi(||x - x_j||). \tag{2}$$

  In the equation above, x is the unsampled location

Fig. 1. Selected geographical zone.

where the interpolation is attempted, n is the number of basis functions used, $s(x)$ is the RBF interpolator, λ_j is the j-th expansion coefficient, obtainable imposing the n equations of the passage to the sampled points, and $\phi(r)$ is the function used as the basis. It is important to noticed that the basis function does not take as input the actual position of the point where we want to interpolate, but only the distance r between that point and the j-th training point, which is the center of the j-th basis, this is why it is called 'radial'. Various basis function can be used, but the most common ones and the one adopted in this work is the Gaussian function:

$$\phi(r) = e^{-\epsilon r^2}, \tag{3}$$

which is a basis function with an adjustable parameter ϵ, called spreading, which quantifies the 'amount of smoothness' of the interpolator, a sort of chosen characteristic length for the function we want to interpolate. Another parameter that must be set when dealing with the RBF interpolator is the number of basis functions to be used, (also called number of neurons in the literature). The default choice is to use a number of basis function equal to the number of training points, but better results can be achieved when this number is reduced in order to avoid high computational complexity and to decrease the risk of overfitting in training (the training error goes to 0 when the number of basis function equals the number of training/sampling points, having the same number of degrees of freedom and of constraints). The MATLab™ command `newrb` is being used to perform this interpolation.

IV. Experience setup

The data used for testing those algorithms have been taken from the online public dataset "Atlantic-Iberian Biscay Irish- Ocean Wave Analysis and Forecast" provided by Copernicus Marine Service, available for everyone at [21]. The parameter VTPK, which is the energy period T_e, has been downloaded for all the year 2022 and on the geographic coordinates shown in the Fig. 1, while an example surface of Te is showed in Fig. 2. The missing white points are terrain.

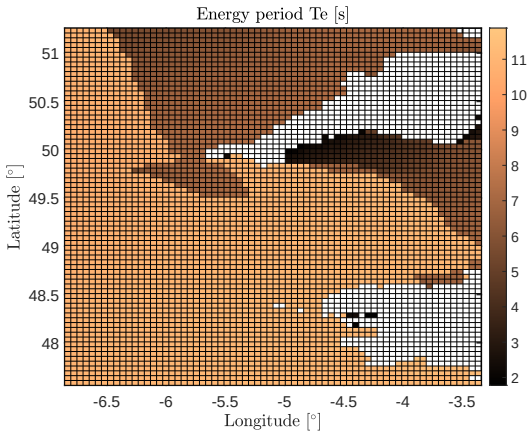

Fig. 2. Example colormap of the energy period Te.

On the original dataset, the maps of the parameter have a resolution of 114 cells on the longitude axis and of 295 on the latitude axis. The spatial resolution is very different between the 2 axis and the number of cells may be too high. This could translate into an extremely high computational time for the algorithms, specially for the RBF one. Indeed, for reducing the learning time of the algorithms and in order to have almost the same spatial resolution for both the axis, the data are resampled to a new resolution of 69x75 cells, from the original one of 114x295. At the end, the map that is being considered contains 5175 points, of which 941 are terrain and are not considered in the analysis, leaving an effective number of 4234 spatial points.

Always for practical reasons, not all the 2913 time instants in the year of the original dataset are considered and only 24 temporally equidistant snapshots are taken into account during the test, in order to have around a snapshot each 2 weeks. After a preliminary analysis, three training percentages are selected to perform the spatial reconstruction, in order to simulate different levels of sparsity of the data:

- 5 %, 211 points
- 2 %, 84 points
- 1 %, 42 points

The above mentioned points are randomly selected between all the points that are not terrain. All the algorithms tested are known for performing poorly at the edges of the considered field. For this reason, all

the 4 borders, here taken with a thickness of 10 cells, are not considered when evaluating the performance of the reconstructions of the algorithms. An example of the selection of the training points is shown in Fig. 3.

For the purpose of having an unbiased estimator

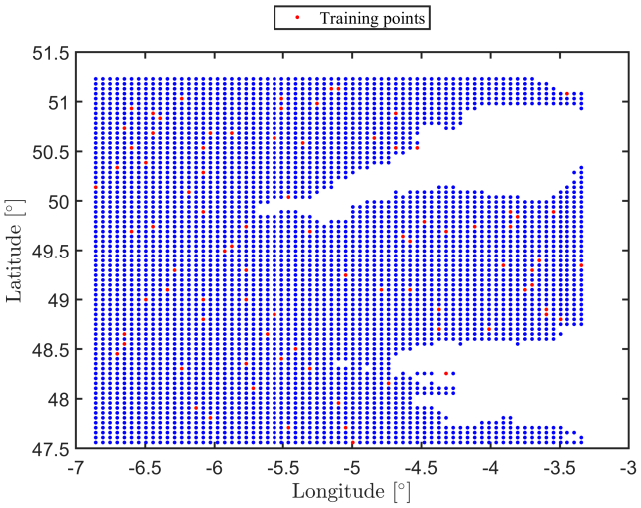

Fig. 3. Example of the distribution of the training points. Blue points represent validation points, while red points represents training points.

of the performance, each configuration is repeated 50 times, every time selecting different training points so that the randomness introduced by selecting only a portion of the points can be reduced. All the points that are not being used for training are instead used for testing (if not excluded for being at the borders). The last thing to be set remains the hyperparameters of the RBF interpolator, spread and number of neurons: in order to account for the dynamics of the sea, an exhaustive exploration of the hyperparameters space is performed for each of the 24 time instants, selecting the best values of the spreading and of the number of neurons (with respect to the achieved validation NRMSE). This hyperparameters space is limited by the dimensions of the map and the number of training points (which should act as an upper limit for the number of neurons, to avoid overfitting). To obtain optimal results, the exploration of the hyperparameters space, that has been done for RBF, should be done also for the p parameter of the thin plat spline, but after a preliminary check it was possible to conclude that changing p does not affect significantly the results of the spline interpolator, and so a fixed value of 0.5 is used for p at all time instants. This is probably due to the fact that the concept of bending energy has no analog physical interpretation when talking about the surfaces of the energy period.

V. Obtained results

Examples of the original surface and of the reconstructed ones are shown in Fig. 4, for all the the interpolators. As it can be seen from the figure, all the interpolating methods tend to have problems when going at the border of the map, i.e. when extrapolating. The difference between interpolation and extrapolation is that the former is performed inside the convex hull of the training points, while the latter is performed outside the convex hull. Moreover extrapolating is notoriously more challenging than interpolating and so it makes sense that the reconstruction performs the worst near the borders and the terrain. For this reason all the points near the borders are excluded from the NRMSE calculation.

For each instant, the error between the interpolated value and the true value is computed, for each point not used in the training and not excluded from testing for being near the borders of the map. After the error for each point have been computed, it is possible to define the overall error in the reconstruction by using the NRMSE defined as:

$$NRMSE = \frac{\sqrt{\frac{\sum_{j=1}^{n} ||\hat{y_j} - y_j||^2}{n}}}{\bar{y}} \tag{4}$$

where n is the total number of points used for validation, $\bar{y}$ is the mean value of the parameter between all points, $\hat{y_j}$ is the j-th interpolated value and y_j is the actual j-th value. This number can be interpreted as a mean percentage error that penalizes more outliers. Since for each specific setup, 50 different selections of the training points are taken, 50 NRMSEs are obtained for each setup. The mean and the stanard deviation of these 50 values are used to evaluate the performance of the specific setup. Fig. 5 depicts the evolution of the mean and of the standard deviation of the NRMSE for all the 24 time instants for the 3 methods and for the 1%, 2% and 5% cases respectively. From Fig. 5 it can be seen how the dynamics of the sea makes the performance of each interpolator change in time, with some time instants being more easily reconstructable than others. Also the robustness of this performance is affected by the time instants and here, unlike with the mean value, the effect of the time can be different between the 3 interpolators.

The values of the mean NRMSE and of its standard deviation averaged over all time instants are presented in the histograms in Fig. 6. Finally in Fig. 7 are presented the optimal hyperparemeters of the RBF for each time instants and each training percentage.

As it can be appreciated from the figures, the three algorithms perform almost the same, in terms of average NRMSE when reconstruction T_e, with the RBF performing slightly worse, and the linear interpolator slightly better, even without significant differences. This can be attributed to the specific form of the function we are interpolating, characterized by zones that are almost totally linear, with eventual discontinuities like the step function one. As expected, the average NRMSE decreases by increasing the percentage of training points.

More interesting is the comparison between the standard deviations of the NRMSE: for the 1% and the 2% cases, the RBF is way more robust than the other two algorithms, probably due to the fact that it is a

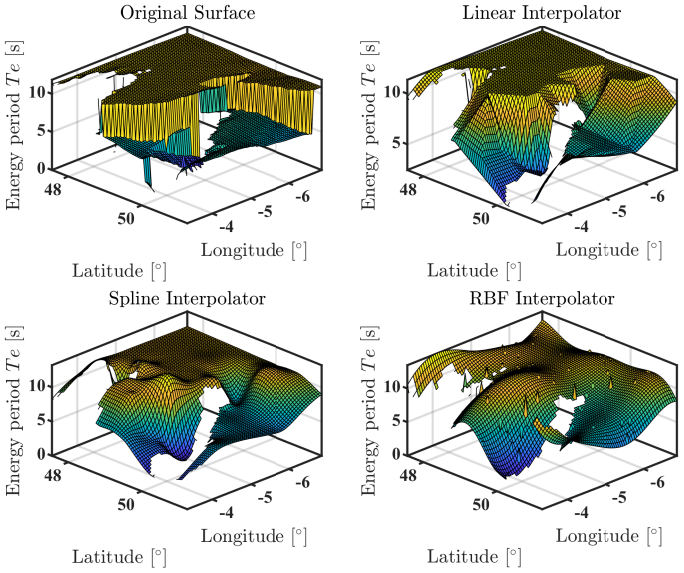

Fig. 4. Example of the real surface for Te and of the interpolated ones for all the 3 interpolators tested, 2% training points

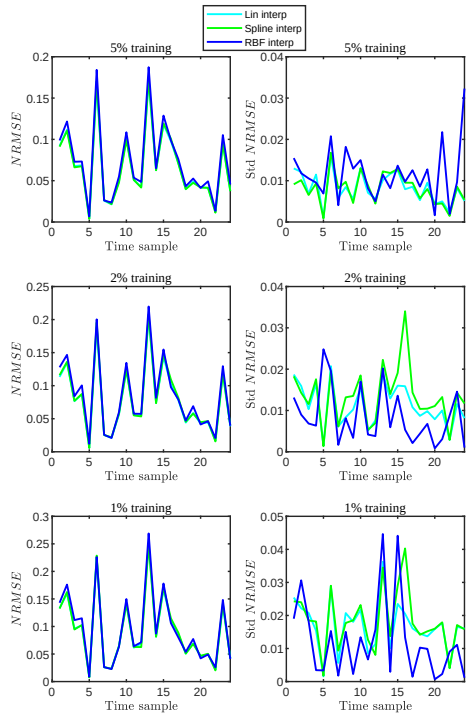

Fig. 5. Obtained average NRMSE and its standard deviation for all scenarios in the 24 time instants

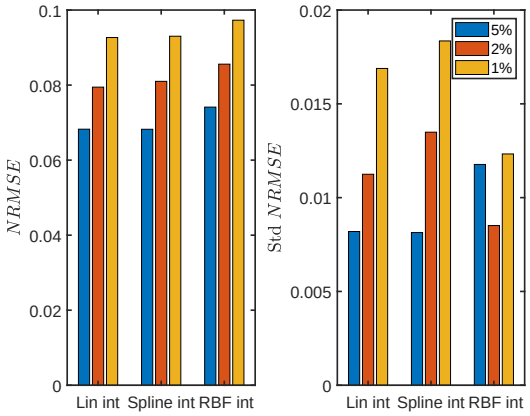

Fig. 6. Obtained average NRMSE and its standard deviation between the 24 time instants, for all training percentages and interpolation methods

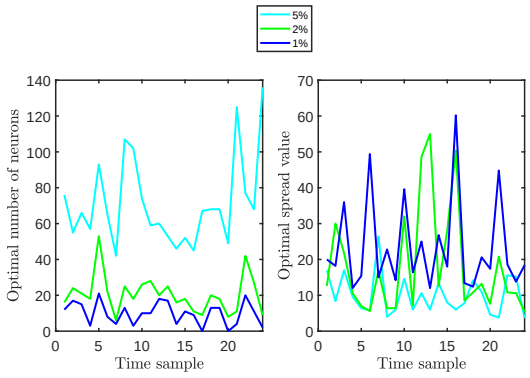

Fig. 7. Evolution of the optimal RBF hyperparameters in the 24 time instants

global interpolator and so can better deal with the stochasticity derived from selecting randomly a very small percentage of points. However, moving to the 5% scenario, the contrary happens, with the linear interpolator and the spline interpolator being more robust than the RBF one, thus favoring local interpolators when the sampling density is high enough. Moreover the standard deviation of the NRMSE of the RBF increases when passing from 2% to 5% probably because having a more dense distribution makes more likely that there is an outlier or a point that resents of the coast effects between these training points. Again this can be attributed to the global nature of the RBF, which resents more of the presence of outliers in the training. Performing a pre-processing of the data for removing the outliers before the train may help in achieving a better and more robust performance and an higher accuracy for all the 3 algorithms, especially for the RBF one.

VI. Conclusions

Three interpolation methods have been tested for reconstructing the energy period T_e from sparse measurements, with different densities of these measurements (5%-2%-1%). Although the average performance for the three methods is almost the same, the RBF method seems more suited when the percentage of training points is low (1% and 2%) since it is less subject to the randomness of selecting few points and considers all the training points when placing a new basis. On the other end, when moving to a finer distribution, the traditional local interpolators seem more suited for the job, particularly the linear interpolator since it performs the best, probably also due to the shape of the functions that are being interpolated. However, a note should be drawn on the fact that the RBF interpolator, being more sophisticated and having tunable hyperparameters, , it may seem unfair to compare it with the other methods considering only the achieved performance. Indeed we are assuming to care more about the achieved performance of the algorithm in unknown locations rather than its required computational time. Anyway, in a real application this requires a third data set in addition to the training and the testing ones, also called validation data set [22]: a sort of second training data set used for evaluating the performance with different hyperparameters and for selecting the most optimal one, before testing it on the unseen test data. This dataset can be artificially generated using ad hoc techniques like cross-validation.

References

[1] J. Cruz, *Ocean Wave Energy*, 2008.

[2] T. K. Brekken, A. von Jouanne, and H. Y. Han, "Ocean wave energy overview and research at Oregon State University," *2009 IEEE Power Electronics and Machines in Wind Applications, PEMWA 2009*, 2009.

[3] A. Terrero González, P. Dunning, I. Howard, K. McKee, and M. Wiercigroch, "Is wave energy untapped potential?" *International Journal of Mechanical Sciences*, vol. 205, no. February, 2021.

[4] E. Commision, "Legislative train 04.2022 1," no. May 2020, pp. 2021–2022, 2022.

[5] E. Giglio, E. Petracca, B. Paduano, C. Moscoloni, G. Giorgi, and S. A. Sirigu, "Estimating the Cost of Wave Energy Converters at an Early Design Stage : A Bottom-Up Approach," pp. 1–39, 2023.

[6] M. Lehmann, F. Karimpour, C. A. Goudey, P. T. Jacobson, and M. R. Alam, "Ocean wave energy in the United States: Current status and future perspectives," *Renewable and Sustainable Energy Reviews*, vol. 74, no. March, pp. 1300–1313, 2017.

[7] C. M. E. G. G. G. Alberto Vargiu, Riccardo Novo and G. Mattiazzo, "An Energy Cost Assessment of Future Energy Scenarios: A Case Study on San Pietro Island ."

[8] N. Faedo, G. Giorgi, J. V. Ringwood, and G. Mattiazzo, "Optimal control of wave energy systems considering nonlinear Froude–Krylov effects: control-oriented modelling and moment-based control," *Nonlinear Dynamics*, vol. 109, no. 3, pp. 1777–1804, 2022. [Online]. Available: https://doi.org/10.1007/s11071-022-07530-3

[9] F. Giorcelli, S. A. Sirigu, E. Pasta, D. G. Gioia, M. Bonfanti, and G. Mattiazzo, "Wave Energy Converter Optimal Design Under Parameter Uncertainty," *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE*, vol. 8, no. June, 2022.

[10] M. P. Manu Centeno-Telleria, Jose Ignacio Aizpurua, "An Analytical Model for a Holistic and Efficient O&M Assessment of Offshore Renewable Energy Systems."

[11] O. Yaakob, F. E. Hashim, K. Mohd Omar, A. H. Md Din, and K. K. Koh, "Satellite-based wave data and wave energy resource assessment for South China Sea," *Renewable Energy*, vol. 88, pp. 359–371, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.renene.2015.11.039

[12] I. McLeod and J. V. Ringwood, "Powering data buoys using wave energy: a review of possibilities," *Journal of Ocean Engineering and Marine Energy*, vol. 8, no. 3, pp. 417–432, 2022.

[13] S. Ramos, H. Díaz, C. Guedes Soares, and G. Lavidas, "Identifying compatible locations for wave energy exploration with different wave energy devices in Madeira Islands," *Developments in Renewable Energies Offshore - Proceedings the 4th International Conference on Renewable Energies Offshore, RENEW 2020*, no. October, pp. 111–122, 2021.

[14] V. T. G. T. A. N. Enrigo Giglio, Gabriele Luzzani and F. Grimaccia, "An Efficient Artificial Intelligence Energy Management System for Urban Building Integrating Photovoltaic and Storage ," vol. IEEE Access, vol 11, pp. 18 673–18 688, 2023.

[15] G. B. Rossi, A. Cannata, A. Iengo, M. Migliaccio, G. Nardone, V. Piscopo, and E. Zambianchi, "Measurement of sea waves," *Sensors*, vol. 22, no. 1, pp. 1–36, 2022.

[16] S. Pennino, A. Angrisano, V. D. Corte, G. Ferraioli, S. Gaglione, A. Innac, E. Martellato, P. Palumbo, V. Piscopo, A. Rotundi, and A. Scamardella, "Sea state monitoring by ship motion measurements onboard a research ship in the antarctic waters," *Journal of Marine Science and Engineering*, vol. 9, no. 1, pp. 1–12, 2021.

[17] X. Liu and M. Wang, "Filling the Gaps of Missing Data in the Merged VIIRS SNPP/NOAA-20 Ocean Color Product Using the DINEOF Method ."

[18] L. C. et al., "Global in situ Observations of Essential Climate and Ocean Variables at the Air–Sea Interface ."

[19] F. L. Bookstein, "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[20] G. Wright, "Radial Basis Function Interpolation: Numerical and Analytical Developments."

[21] Copernicus Marine Service, "Atlantic-Iberian Biscay Irish- Ocean Wave Analysis and Forecast ." [Online]. Available: https://data.marine.copernicus.eu/product/IBI_ANALYSIS_FORECAST_WAV_005_005/download

[22] Domingos P., "A Few Useful Things to Know About Machine Learning," *Communications of the ACM*, vol. 55, no. 10, 2012. [Online]. Available: https://dl.acm.org/citation.cfm?id=2347755