International Journal of Offshore and Polar Engineering (ISSN 1053-5381)

Vol. 35, No. 3, September 2025, pp. 244–251; https://doi.org/10.17736/ijope.2025.ak72

http://www.isope.org/publications

Extending Wave in situ Measurements Through Gaussian Process Regression: An Experimental Campaign

Leonardo Gambarelli and Edoardo Pasta
Dipartimento di Ingegneria Meccanica e Aereospaziale, Politecnico di Torino
Turin, Italy

Francesco Ferri
Department of the Built Environment, Aalborg University
Aalborg, Denmark

Giuseppe Giorgi
Dipartimento di Ingegneria Meccanica e Aereospaziale, Politecnico di Torino
Turin, Italy

Wave energy is recognized as one of the most promising sources of clean and abundant energy. Nevertheless, as of today this technology is still not commercially viable due to a number of reasons, such as the harshness of the sea environment, the expenses needed for the deployment and maintenance of devices in open ocean, and the lack of information regarding wave parameters worldwide. Indeed, a proper characterization of the resource in a site is of quintessential importance for assessing the productivity of the site and dimensioning the supporting system of a device. This work aims to address the problem of the lack of data by resorting to spatial prediction techniques, using data gathered through an experimental campaign conducted at the wave basin facility at the Ocean and Coastal Engineering Laboratory in Aalborg University. During this campaign, two months of data from a real in situ measuring device were replicated in the basin. In the middle of the basin, some concrete blocks were deployed underwater to replicate a sudden shift in the bathymetry, which should act as a disturbance to the wave propagation and arise nonlinear phenomena. Nineteen wave gauges recorded the wave elevation for the whole time. A scenario where only some of the measuring devices were working was replicated by considering only the data from a subsample of wave gauges and inferring the parameters in the locations of the other devices from them, through a Gaussian Process Regression (GPR) algorithm. The proposed algorithm was able to interpolate the parameters at the other locations, at the expense of a relatively low error, indicating that this set up could be used to increase the spatial coverage of the wave-measuring buoys deployed worldwide or to provide an estimate of the parameters at a buoy that is not working, e.g., for maintenance operations.

INTRODUCTION

Wave energy is emerging as a highly promising energy source (Terrero Gonzalez et al., 2021), expected to be instrumental in future energy production. It is anticipated to be a significant factor in the transition away from fossil fuels (von Schuckmann et al., 2020). The appeal of wave energy lies in its advantages, such as its high predictability and availability (Mackay et al., 2010), distinguishing it from other renewable resources like solar or wind. This technology should play a key role in helping to move away from fossil fuels, per the Communication from the European Commission of November 2020 (EU, 2020), setting the ambitious goals of installed capacity of at least 1 GW of ocean (wave and tidal) energy by 2030 and 40 GW by 2050. Despite its potential, as of now wave energy technologies have not achieved commercial viability, primarily due to their prohibitively high Levelized Cost of Energy (LCoE) (Astariz and Iglesias, 2015; Giglio et al., 2023). This elevated LCoE stems from several factors. First, there is no universally optimal device design suitable for all operational scenarios, creating a significant challenge (Trueworthy and DuPont, 2020; Giorcelli et al., 2022). Additionally, there is no effective control systems for all these devices in all of their working conditions (Ringwood et al., 2023). There are also the substantial maintenance and operation costs (Centeno-Telleria et al., 2022), as these activities must be carried out in oceanic environments, which significantly escalates expenses and rapidly damages the devices deployed. Moreover, there is a scarcity of comprehensive global data on wave parameters (Yaakob et al., 2016). This hinders the development and optimization of these technologies, especially regarding the spatial distribution of the wave parameters, which are essential for a proper location assessment (Konuk et al., 2023), not only for wave energy production, but also for offshore wind energy production (Faraggiana et al., 2022). Also, using computer-aided simulation is not perfectly feasible for simulating the ocean environment; these tools can be employed for studying the spatial variability of the waves and can give accurate results (Ashton et al., 2014; Zhou et al., 2015), but since the dynamics regulating the ocean are given by the Navier-Stokes equations, which are highly nonlinear and notoriously difficult to solve, simulating even brief periods accurately typically demands considerable computational resources (Griffies, 2000), hardly making it a viable option. In addition, post-processing operations, like bias correction techniques, are often needed *ad*

Received August 19, 2024; updated and further revised manuscript received by the editors December 30, 2024. The original version (prior to the final updated and revised manuscript) was presented at the Thirty-fourth International Ocean and Polar Engineering Conference (ISOPE-2024), Rhodes, Greece, June 16–21, 2024.

KEY WORDS: Wave energy, Gaussian Process Regression, in situ measurements, resource assessment, spatial gap-filling.

hoc for the results of those numerical simulation to be accurate (Penalba et al., 2023). To address these issues, further research and innovation are needed to make wave energy a more economically feasible and reliable energy source. This study aims to tackle one of the key challenges in wave energy development: the scarcity of essential wave parameters data, which are crucial for selecting the most optimal locations for deploying wave energy converters (WECs) (Yavor et al., 2021). Typically, wave energy parameters are measured using buoys equipped with measuring devices (McLeod and Ringwood, 2022). However, these devices, much like the deployed WECs themselves, require regular maintenance. They are prone to becoming non-operative for extended periods due to the harsh marine environment's toll on the equipment. Addressing this issue is vital for improving the feasibility and efficiency of wave energy projects, as reliable data on wave parameters are fundamental for strategic deployment and optimization of wave energy technologies.

This paper proposes a solution to the challenge of maintaining continuous wave parameter data collection in situations where some measuring devices become inoperative. We simulate a real-world scenario in a specific area where several measuring devices are deployed but only a subset remains operational. In this context, we employ Gaussian Process Regression (GPR) to reconstruct wave parameters at the non-operative buoys using data from the correctly working devices. This approach was tested in a carefully designed experimental campaign to gather the necessary data.

The structure of this paper is organized as follows. First, a comprehensive description of the experimental setup, including the methods for data collection and processing, is provided. Then, an explanation of the adaptation and implementation of GPR for this particular application is given. After that, the presentation and discussion of the results obtained from the experiment is provided. Finally, a compilation of conclusions drawn from these results, providing insights into the effectiveness and potential applications of this approach in wave energy parameter measurement, is presented.

EXPERIMENT SETUP

The experimental phase of this study was carried out in the wave basin at the Ocean and Coastal Engineering Laboratory of Aalborg University. This facility boasts a wave basin measuring 19.3 meters in length, 14.6 meters in width, and 1.5 meters in depth, with an active testing area of 13 meters by 8 meters. During the experiments, the water depth was maintained at 0.9 m. The basin is equipped with an advanced segmented wavemaker system. This system features 30 independent wave paddles, each capable of precise control, and includes active absorption technology.

For the purpose of the experiments, the basin's floor was modified by adding concrete blocks to mimic sudden changes in the bathymetry, as it is well known that a changing bathymetry is a source of nonlinearites for the wave propagation (Madsen et al., 2006). Indeed, the literature indicates that sudden changes in bathymetry can heighten the risk of initiating nonlinear phenomena such as wave slamming (Paulsen et al., 2019). Additionally, a set of 19 wave gauges was deployed to record wave heights, replicating the function of in situ measurement instruments. The generated waves in the tank were manipulated to slowly vary their spectrum, mirroring actual wave data and spectral parameters obtained from measurements of the DanWEC site (Brodersen et al., 2013). This setup provided a controlled environment to accurately simulate real-world ocean conditions for this study.

In this research, we replicated two months' worth of oceanic conditions observed at the DanWEC site in the wave basin, using a scale factor of 1:200. This scale factor was selected to align with the dimensions of the wave basin. Two hundred four different sea states were launched one after the other in this way, each one lasting for around 450 s. The 19 wave gauges were installed around the strategically-placed concrete blocks within the basin and played a crucial role in providing both training and validation datasets for the models being tested in this study. The wave gauges recorded the time series of the wave elevation in the points where they were deployed. From this time series it was possible to derive the frequency spectrum to calculate the wave parameter of interest H_s and T_e by using the spectral moments.

$$H_s = \sqrt{4m_0} \tag{1}$$

$$T_e = \frac{m_{-1}}{m_0} \tag{2}$$

An image illustrating the wave gauges and the concrete blocks setup is presented in Fig. 1, while a plant showing the spatial distribution of the wave gauges is presented in Fig. 2. Figure 3 shows the exact positioning of the concrete blocks.

From Fig. 2, it is possible to notice the presence of a wave gauge particularly far away from the others, in position (10.335, 4.2) m, that acts as an undisturbed measurement recording the wave parameters away from the influence of the nonlinearities generated by the concrete blocks. The waves are generated and propagated along the positive y-axis, as shown in Fig. 3. Table 1 provides the precise positions of the wave gauges. For a more detailed description of the instrumentation used and of the experimental environment, the reader can refer to Faedo et al. (2023).

Fig. 1 Disposition of the wave gauges and of the concrete blocks in the basin

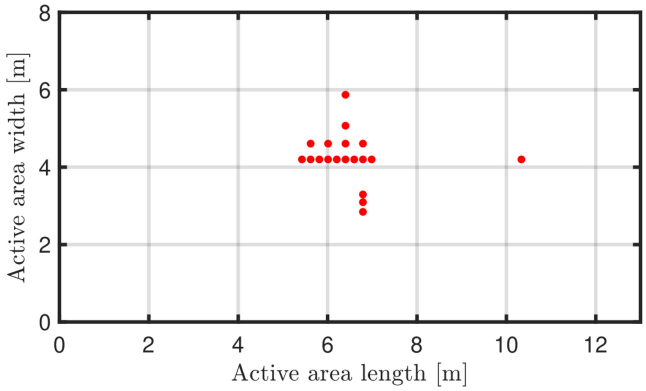

Fig. 2 Spatial distribution of the wave gauges in the active area of the basin

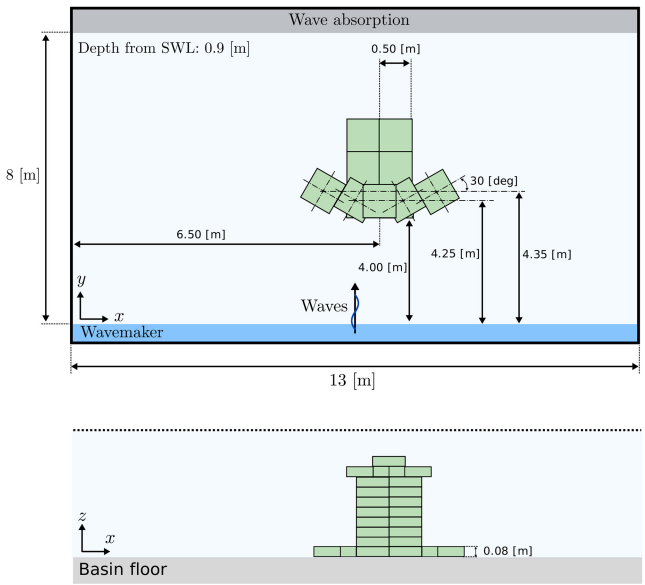

Fig. 3 Scheme of the disposition of the concrete blocks

Wave gauge	x coordinate [m]	y coordinate [m]
1	6.7900	2.8450
2	6.7900	3.0950
3	6.7900	3.2950
4	6.9850	4.2000
5	6.5950	4.2000
6	6.2050	4.2000
7	5.8150	4.2000
8	5.4250	4.2000
9	6.7900	4.6080
10	6.4000	4.6080
11	6.0100	4.6080
12	5.6200	4.6080
13	6.4000	5.8700
14	10.3350	4.2000
15	6.7900	4.2000
16	6.4000	4.2000
17	6.0100	4.2000
18	5.6200	4.2000
19	6.4000	5.0700

Table 1 Positions of the wave gauges

In Figs 4 and 5, it is possible to see an example of the measured parameters H_s and T_e, respectively, along with their location in the basin.

Figures 6 and 7 illustrate the time series of the wave parameters H_s and T_e across the 204 sea states, as measured by the undisturbed wave gauge located at position $(10.335, 4.2)$ m.

SPATIAL REGRESSION ALGORITHM

In this work, a Gaussian Process Regression (GPR) has been used for performing the spatial regression of the parameters measured by the wave gauges.

GPRs have been chosen for this work since they seem suited for this kind of interpolation for a number of reasons, like their ability to provide an estimate of the uncertainty of the prediction, which can be useful when moving towards real applications (Neal, 2000). Additionally, GPRs are suited to work with a low quantity of data that can even be sparse in space (Rasmussen, 2004), as

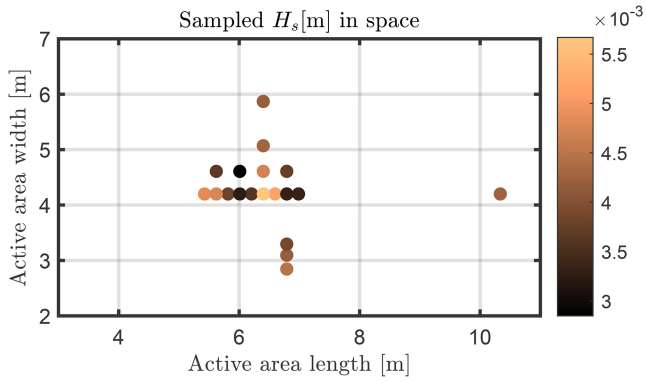

Fig. 4 H_s measured values

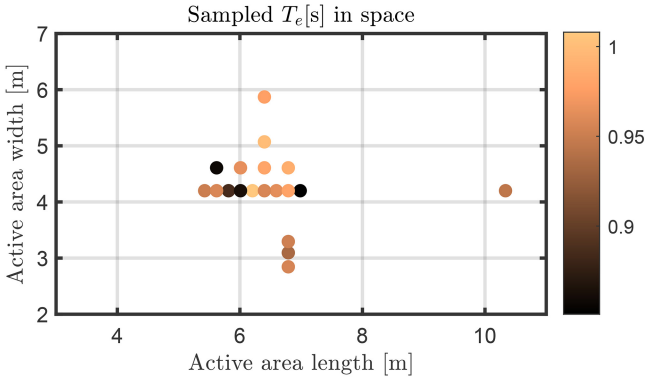

Fig. 5 T_e measured values

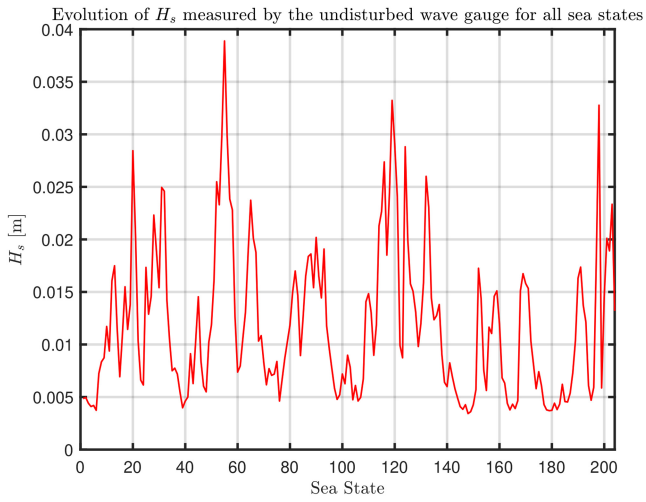

Fig. 6 H_s time series on the undisturbed wave gauge

in our experimental setup. Furthermore, GPRs are abstract and general enough to be able to describe almost all phenomena, but they also have tunable hyperparameters that can be tailored for the specific application (Seeger, 2004).

GPRs expand upon and generalize traditional Kriging techniques, extending their application beyond geostatistics (Ebden, 2015; Christianson et al., 2022). A Gaussian Process is theoretically defined as a collection of random variables, where any finite subset of those variables follows a multivariate normal distribution (Rasmussen and Williams, 2006). In this context, the parameter of interest (i.e., T_e and H_s in this application) is viewed as outcomes of a stochastic process, which is characterized by two key functions: a mean function denoted as $m(x)$ and a covariance or kernel function, represented as $k(x, x')$. Calling the generic input point $x \in R^d$ (i.e., in this study, length and width positions in the

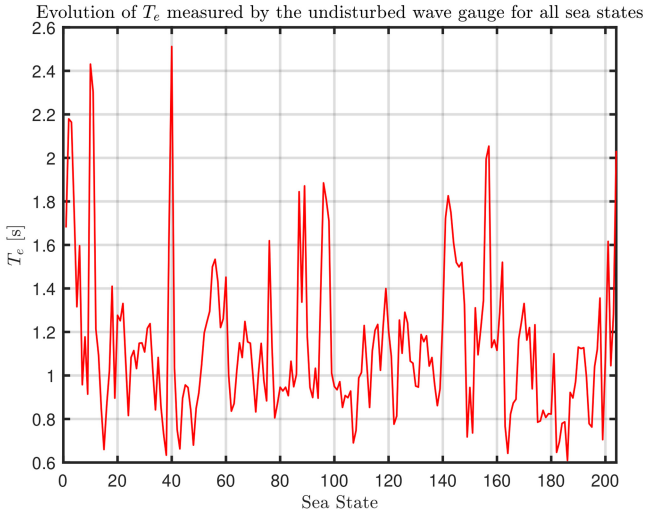

Fig. 7 T_e time series on the undisturbed wave gauge

tank), where d is the dimension of our input space (in our case $d = 2$), N as the number of training points, and T as the number of testing points, let us define several key elements:

1. $X \in R^{d\times N}$: This is the matrix that contains all the training input points. In our spatial regression context, these would be the coordinates of the locations where the measurements of the wave parameters were taken.
2. $f \in R^N$: These are the training output values, which represent the actual measurements observed at the training points.
3. $X^* \in R^{d\times T}$: This refers to the testing input points, essentially the other points in the input space where we aim to predict or estimate the output (in our case these points would be the positions of all the other wave gauges not used in the training set).
4. $f^* \in R^T$: This is the estimate of the output at the testing points.

Given these definitions, considering a constant mean function equal to 0, the estimate f^* at the testing points is represented as a normal distribution. The mean of this distribution is calculated as $K(X^*, X)K(X, X)^{-1}f$, while its variance is given by $K(X^*, X^*) - K(X^*, X)K(X, X)^{-1}K(X, X^*)$.

$$f^*|f, X, X^* \sim \mathcal{N}\big(K(X^*, X)K(X, X)^{-1}f, \\ K(X^*, X^*) - K(X^*, X)K(X, X)^{-1}K(X, X^*)\big) \tag{3}$$

In the above formulation:

- $K(X, X) \in R^{N\times N}$ is the training-training covariance matrix computed evaluating the kernel function between all possible training-training pairs of points.
- $K(X^*, X) \in R^{T\times N}$ is the testing-training covariance matrix computed evaluating the kernel function between all possible testing-training pairs of points. Since kernel function must be symmetric, $K^T(X^*, X) = K(X, X^*)$.
- $K(X^*, X^*) \in R^{T\times T}$ is the testing-testing covariance matrix computed evaluating the kernel function between all possible testing-testing pairs of points.

The mean of the prediction $K(X^*, X)K(X, X)^{-1}f$ contains two important terms:

1. $K(X^*, X)$ is the testing-training covariance matrix, also known as cross-variance matrix. This matrix represents how much the points where predictions are made (test points) are related to the training points. This term captures the influence that the observed data (training points) have on the predictions at new, unobserved locations (test points).
2. $K(X, X)^{-1}$ is the inverse of the training-training covariance matrix or simply the covariance matrix of the training points. It is used to "normalize" or adjust the influence of the training data on the prediction. By taking the inverse of the covariance matrix of the training points, it effectively removes redundant information and adjusts for the interdependence of the training data. This process ensures that the prediction is not overly influenced by the correlations in the training data, so that two points collocated (or at very small distance) are actually seen as just one point and not two distinct ones.

A common practical choice during the implementation is setting $T = 1$ and just estimating one testing point at a time, since considering the testing-testing covariance would only increase the computational effort required without improving the accuracy.

As can be seen from the equations, here the main part of the model design is performed by the choice of the kernel function $k(x, x')$. Indeed, the kernel function is the key component of the algorithm that models the spatial auto-correlation of the considered stochastic process (Chen, 2021). Different kernel functions exist, each one describing a different spatial auto-correlation in a stochastic process. Some of the most used kernel functions are:

- The Squared Exponential (SE) kernel:

$$k_{\text{SE}}(x, x') = \sigma^2 \exp\left(-\frac{||x - x'||^2}{2l^2}\right) \tag{4}$$

- The Linear Exponential (LE) kernel:

$$k_{\text{LE}}(x, x') = \sigma^2 \exp\left(-\frac{||x - x'||}{l}\right) \tag{5}$$

- The Matérn kernel:

$$k_{\text{Mat}}(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}||x - x'||}{l}\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}||x - x'||}{l}\right) \tag{6}$$

Typical choices for the ν parameter are 1.5 and 2.5 (Rasmussen and Williams, 2006). $\Gamma(\nu)$ is the Gamma Function of ν, and K_{ν} is the modified Bessel function of the second kind.

- The Rational Quadratic (RQ) kernel:

$$k_{\text{RQ}}(x, x') = \sigma^2\left(1 + \frac{||x - x'||^2}{2\alpha l^2}\right)^{-\alpha} \tag{7}$$

Here, typical choices for the α parameter are 0.5 and 2 (Rasmussen and Williams, 2006).

The kernels presented above, as most common kernels, are stationary kernels, in the sense that they depend only on the distance $||x - x'||$ and not on the specific values of both x and x'. All those kernel function have an amplitude set by σ, which quantifies the amount of variance of the stochastic process that can be inferred by its spatial auto-correlation described by the kernel. These functions have also a characteristic length, l, which sets the range inside of which it is possible to make spatial inference. For comparison, Fig. 8 reports together the above described kernels using an $l = 1$ and a $\sigma = 1$.

To account for measurement noise and to increase the numerical stability of the algorithm, another term is added to the kernel function: $\sigma^2_{err}\delta_{xx'}$, where σ^2_{err} is the error measurement variance (that can be seen as the average difference of two collocated measurements), while $\delta_{xx'}$ is the Kronecker Delta of x and x'. Due to the presence of the Kronecker Delta, this error variance term is added to the kernel only when $x = x'$ (only on the diagonal of the kernel matrix). In our setup, σ^2_{err} was set to 0.001 and 0.0001 for the T_e and H_s, respectively, and it was added mainly for numerical stability since the measurements were pretty accurate.

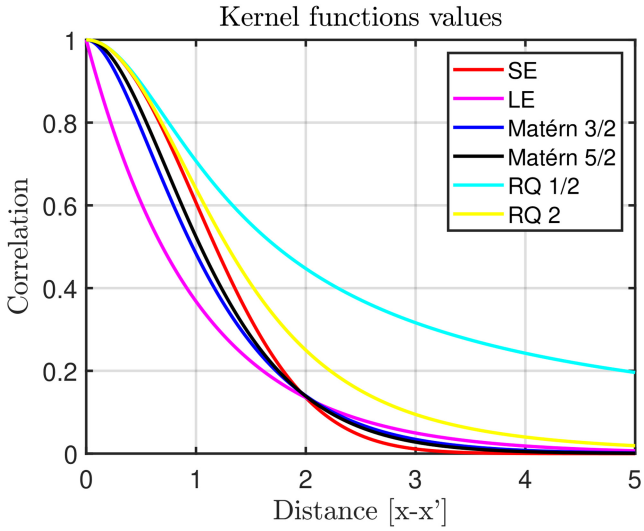

Fig. 8 Most commonly used kernel functions, with unitary amplitude and characteristic length

In the considered experimental setup, there were a total of 204 time snapshots in which each wave gauge recorded H_s and T_e.

Since the behavior of the parameters changed significantly between those temporal snapshots, each time snapshot has been considered alone (no temporal correlation is considered in this work). Also, the variance of the parameters changed, and for this reason the σ parameter was set as the empirical standard deviation of the parameter (seen from the selected training points). The number of training points changed between 3 and 14 to test scenarios with different levels of sparsity of the training points. The selection of the training points was performed randomly (excluding the undisturbed wave gauge) 50 times for each setup. In the end, it was possible to obtain $204 \times 11 \times 50$ setups. The characteristic length l for the reconstruction still needs to be specified. Different values for this parameter have been tested, starting from the average distance between the wave gauges $l = d_{avg} = \binom{N}{2}^{-1} \sum_x \sum_{x', x' \neq x} \|x - x'\| = 1.2755$ m, where N is the number of considered points, and then trying $d_{avg}/2 = 0.6377$ m, $d_{avg}/4 = 0.3189$ m, $d_{avg}/8 = 0.1596$ m, and $d_{avg}/16 = 0.0797$ m. The average distance between the wave gauges was chosen as the reference for the l parameter, as it represents a characteristic length scale of the process under study. In this context, the wave gauges play a central role in describing the spatial variability of the process. The parameter measured by the undisturbed wave gauge has been used as an estimate of the mean of the process, which was then subtracted from the measurements at the training points, and only the errors with respect to this mean were actually interpolated at the testing points.

For each of these reconstructions, the Mean Absolute Relative Error (MARE) was computed.

$$\text{MARE} = \frac{1}{T} \sum_{i=1}^{T} \frac{|y_i - \hat{y}_i|}{y_i} \tag{8}$$

where T is the number of considered testing points (equal to 18 minus the number of training points), y_i is the real value of the estimated parameter at the ith testing point, and $\hat{y}_i$ is the estimate of the parameter at the ith testing point. All of the reconstructions were repeated considering 6 different kernels:

1. SE kernel
2. LE kernel
3. Matérn 3/2 kernel
4. Matérn 5/2 kernel
5. RQ 0.5 kernel
6. RQ 2 kernel

RESULTS

The reconstruction results are summarized in Fig. 9, which shows the average MARE across all time instants and random permutations of the training points, for all kernels, characteristic lengths, and varying numbers of training points for both T_e and H_s.

These results are further detailed in Table 2 and Table 3, presenting the average MARE with respect to different numbers of training points, characteristic lengths, and kernels used in the reconstruction for both H_s and T_e.

For both T_e and H_s, using the full d_{avg} as the characteristic length in the reconstruction yielded poor performance. In this setup, the LE kernel, which provides the least spatial correlation with fixed parameters, performed the best (as shown in Fig. 8, where the LE kernel has the lowest area under the curve). This suggests that better results could be achieved by using a shorter characteristic length.

Indeed, reducing the characteristic length gave better results for both parameters and for all kernels. As can be seen from Fig. 9 and from Table 2 and Table 3, the best performances are obtained for both T_e and H_s when using an RQ 0.5 kernel with a characteristic length of $d_{avg}/8$, with a MARE of 0.10936 and 0.024781 for H_s and T_e, respectively.

Reducing the characteristic length even further starts to worsen the results, as can be seen from the higher MARE for the $d_{avg}/16$ setup, when compared to the $d_{avg}/4$ and $d_{avg}/8$ setups. Another thing that can be seen from the results is that the errors for H_s are almost one order of magnitude higher compared with the errors of T_e. This was already found in Gambarelli et al. (2023) and can be attributed to the fact that the Coefficient of Variance, equal to the mean of a variable divided by its standard deviation, is higher for H_s and lower for T_e, indicating that T_e has small perturbations of its values with respect to its mean, which makes it easier for T_e to achieve better metric performances.

$$\text{CoV}_x = \frac{\sigma_x}{\mu_x} \tag{9}$$

Using the optimal kernel setup (RQ 0.5) with a characteristic length of $d_{avg}/8$, the variation in errors across different buoys was

H_s	SE	LE	Matérn 3/2	Matérn 5/2	RQ 0.5	RQ 2
d_{avg}	0.15313	0.11552	0.13197	0.14248	0.14131	0.14893
$d_{avg}/2$	0.146	0.11255	0.11875	0.1257	0.12572	0.13603
$d_{avg}/4$	0.11717	0.11236	0.11094	0.1116	0.11164	0.11388
$d_{avg}/8$	0.1134	0.11676	0.11499	0.1144	0.10936	0.11157
$d_{avg}/16$	0.12405	0.12291	0.12329	0.12349	0.11259	0.12062

Table 2 Average MARE for H_s, for different kernels and characteristic lengths

T_e	SE	LE	Matérn 3/2	Matérn 5/2	RQ 0.5	RQ 2
d_{avg}	0.038696	0.02598	0.029925	0.033582	0.034155	0.036625
$d_{avg}/2$	0.035172	0.025256	0.026795	0.028807	0.029481	0.032678
$d_{avg}/4$	0.028051	0.02503	0.025011	0.02535	0.025541	0.026235
$d_{avg}/8$	0.025646	0.026088	0.025816	0.025738	0.024781	0.025121
$d_{avg}/16$	0.027634	0.027404	0.027487	0.027526	0.025331	0.026954

Table 3 Average MARE for T_e, for different kernels and characteristic lengths

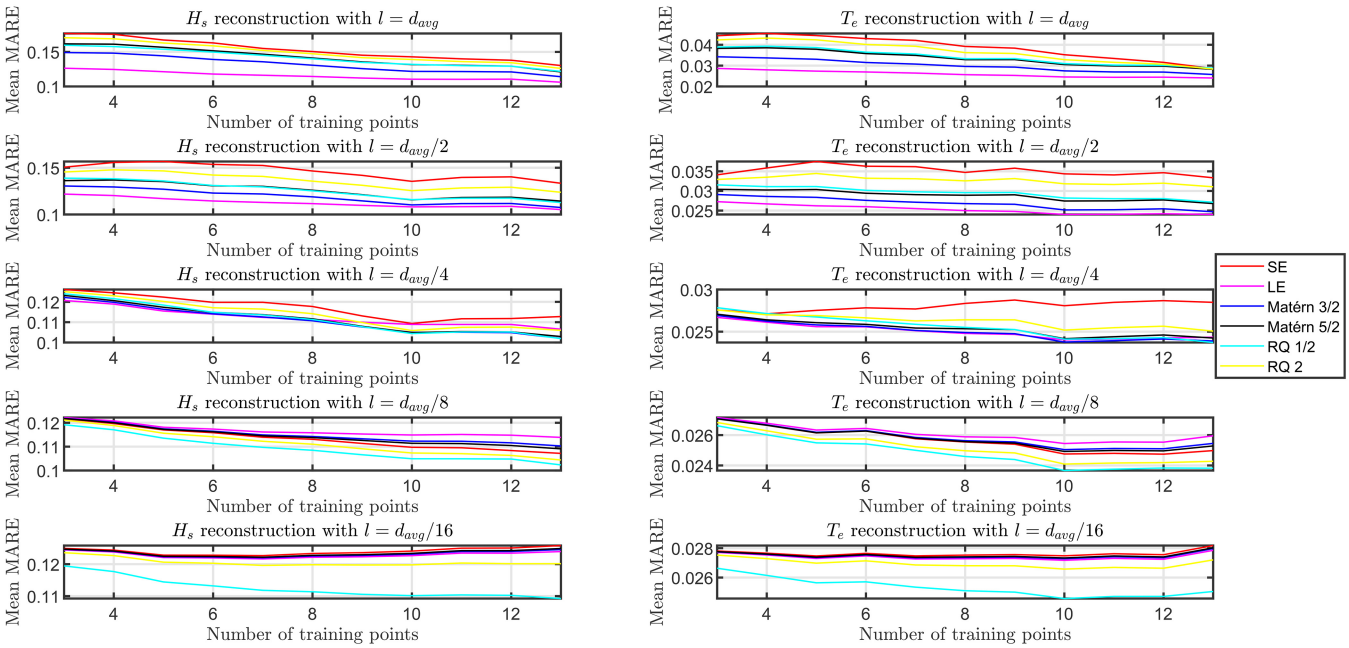

Fig. 9 Obtained average MARE for different kernels, characteristic lengths and number of training points

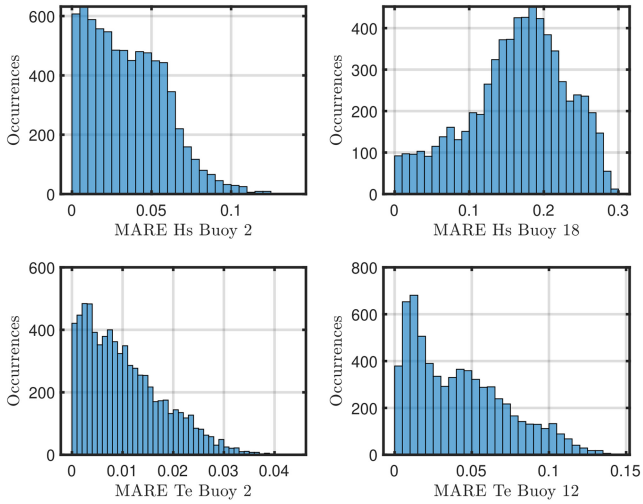

Fig. 10 Obtained MARE values for the considered buoys, with RQ 0.5 kernel and 7 training points setup

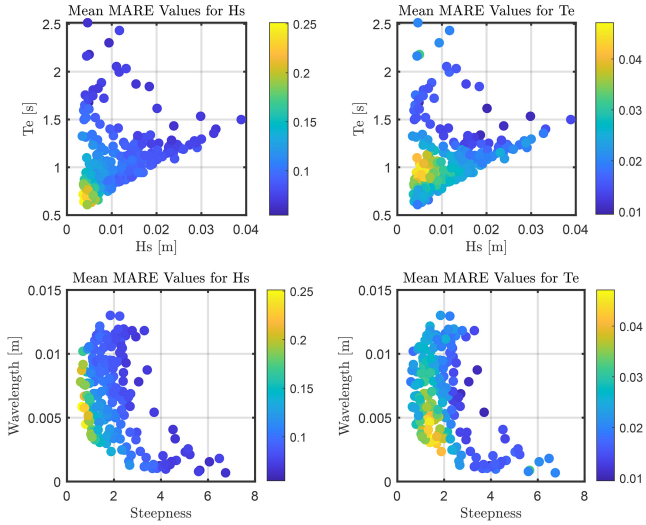

Fig. 11 Averaged MARE for H_s and T_e, for different sea state with RQ 0.5 kernel and 7 training points setup, as a function of the sea state's H_s and T_e or wavelength and steepness

further analyzed. Buoy 2, located at $(6.79, 3.095)$ m, exhibited the lowest reconstruction errors for both H_s and T_e. By contrast, the highest errors were observed for H_s at Buoy 18, positioned at $(5.62, 4.2)$ m, and for T_e at Buoy 12, located at $(5.62, 4.608)$ m. Figure 10 illustrates the distributions of the obtained MAREs for these variables under the setup with 7 training points.

A final analysis was conducted to account for the wave condition. Using the measurements from the undisturbed buoy as a representation of the specific wave condition, the average MARE for each sea state was calculated and plotted in Fig. 11, with the RQ 0.5 kernel and 7 training points setup. This is presented as a function of the undisturbed H_s and T_e, or equivalently as a function of the wavelength and steepness (derived from the undisturbed H_s and T_e).

As shown in Fig. 11, certain sea states are more challenging to reconstruct, resulting in higher average MARE. Specifically, the reconstruction of H_s performs worse for sea states characterized by low values of both H_s and T_e, or equivalently, by low values of wavelength and steepness. By contrast, the reconstruction of T_e exhibits lower performance for sea states with low H_s values and T_e around 1 s, or equivalently, with low wavelength and steepness values near 2.

CONCLUSIONS

This preliminary work represents an initial attempt to estimate wave parameters at a specific location using data from nearby buoys, a common scenario in practical applications (e.g., when a measuring device fails and requires maintenance). GPRs were explored for this task, proving effective in providing accurate wave parameter estimates. These findings demonstrate the potential of GPR to address practical engineering challenges, such as ensuring continuous wave parameter monitoring for the strategic deployment and maintenance of wave energy converters in real-world scenarios.

Several conclusions can be drawn from the results. First, because the variance of the wave-generating process varies over time, the kernel's amplitude, which addresses spatial auto-correlation, can be adjusted instant-by-instant using the empirically-derived standard deviation from the training points in each snapshot. Second, the characteristic length that best captures the process variability appears to be $d_{avg}/8$ for both T_e and H_s, indicating that this scale captures the dominant features in the data. However, this result may be influenced by the spatial arrangement of our wave gauges. Changing the positions of the measuring devices (either closer together or farther apart) could shift the optimal characteristic length toward different scales. Additionally, the optimal kernel for both T_e and H_s is the RQ 0.5 kernel, which seems to best describe the spatial auto-correlation of the process. The fact that the RQ 0.5 kernel yields the highest correlation over greater distances suggests that the process exhibits a significant degree of nonlocality, meaning the influence of any given point extends over a broader range rather than being predominantly local.

In conclusion, while significant improvements were observed with the l parameter set at $d_{avg}/8$, further refinement of this parameter near this value could potentially yield even better results. A more detailed examination of the l parameter around $d_{avg}/8$ might reveal subtle nuances in the data, enhancing the model's accuracy. Further analysis of the results reveals that the performance metric varies across the considered buoys and sea states. Certain buoys are more significantly impacted by nonlinear phenomena, making their reconstruction more challenging. Last, a note on σ^2_{err}: this parameter was not critical in this setup due to the high accuracy achievable in a laboratory setting. However, in real-world applications, this parameter becomes crucial, and proper tuning could significantly improve accuracy.

ACKNOWLEDGEMENTS

This work is part of the project (second and third authors) AIMS: Artificial Intelligence to Monitor our Seas – funded by the European Union - NextGeneration EU - Ministry of University and Research - under the PRIN 2022 PNRR Call for Proposals (D.D.1409 of September 14, 2022). The first author is part of the project NODES, which has received funding from the MUR – M4C2 1.5 of PNRR with grant agreement no. ECS00000036. Support was also received from the framework COST Action 17105 - WECANet (second and fourth authors).

REFERENCES

Ashton, I, Van-Nieuwkoop-McCall, J, Smith, H, and Johanning, L (2014). "Spatial Variability of Waves Within a Marine Energy Site Using In-Situ Measurements and a High Resolution Spectral Wave Model," *Energy*, 66, 699–710. https://doi.org/10.1016/j.energy.2013.12.065.

Astariz, S, and Iglesias, G (2015). "The Economics of Wave Energy: A Review," *Renewable Sustainable Energy Rev*, 45, 397–408. https://doi.org/10.1016/j.rser.2015.01.061.

Brodersen, H, Nielsen, K, and Kofoed, J (2013). "Development of the Danish Test Site DanWEC," *Proc Eur Wave Tidal Energy Conf*, Aalborg, Denmark, EWTEC.

Centeno-Telleria, M, Aizpurua, J, and Penalba, M (2022). "An Analytical Model for a Holistic and Efficient O&M Assessment of Offshore Renewable Energy Systems," *SSRN Electron J*. https://doi.org/10.2139/ssrn.4273477.

Chen, Y (2021). "An Analytical Process of Spatial Autocorrelation Functions Based on Moran's Index," *PLOS ONE*, 16(4), 1–27. https://doi.org/10.1371/journal.pone.0249589.

Christianson, R, Pollyea, R, and Gramacy, R (2023). "Traditional Kriging Versus Modern Gaussian Processes for Large-scale Mining Data," *Stat Anal Data Min*, 16(5), 488–506. https://doi.org/10.1002/sam.11635.

Ebden, M (2015). "Gaussian Processes: A Quick Introduction," arXiV preprint. https://doi.org/10.48550/arXiv.1505.02965.

EU (2020). *An EU Strategy to Harness the Potential of Offshore Renewable Energy for a Climate Neutral Future*, COM(2020) 741, European Commission.

Faedo, N, et al. (2023). "SWELL: An Open-access Experimental Dataset for Arrays of Wave Energy Conversion Systems," *Renewable Energy*, 212, 699–716. https://doi.org/10.1016/j.renene.2023.05.069.

Faraggiana, E, et al. (2022). "A Review of Numerical Modelling and Optimisation of the Floating Support Structure for Offshore Wind Turbines," *J Ocean Eng Mar Energy*, 8, 433–456. https://doi.org/10.1007/s40722-022-00241-2.

Gambarelli, L, Pasta, E, and Giorgi, G (2023). "On Spatial Interpolation of Ocean Energy Source Variables: A Comparative Analysis," In *Proc Eur Wave Tidal Energy Conf*, Bilbao, Spain. EWTEC, Vol 15.

Giglio, E, et al. (2023). "Estimating the Cost of Wave Energy Converters at an Early Design Stage: A Bottom-Up Approach," *Sustainability*, 15(8), 6756. https://doi.org/10.3390/su15086756.

Giorcelli, F, et al. (2022). "Wave Energy Converter Optimal Design Under Parameter Uncertainty," *Proc 41st Int Conf Ocean Offshore Arct Eng*, Hamburg, Germany, ASME, 8, V008T09A085. https://doi.org/10.1115/OMAE2022-81464.

Griffies, SM (2000). *Fundamentals of Ocean Climate Models*, Princeton University Press, 528 pp.

Konuk, E-B, et al. (2023). "On the Definition of a Comprehensive Technology-informed Accessibility Metric for Offshore Renewable Energy Site Selection," *J Mar Sci Eng*, 11(9), 1702. https://doi.org/10.3390/jmse11091702.

Mackay, EB, Bahaj, AS, and Challenor, PG (2010). "Uncertainty in Wave Energy Resource Assessment. Part 2: Variability and Predictability," *Renewable Energy*, 35(8), 1809–1819. https://doi.org/10.1016/j.renene.2009.10.027.

Madsen, PA, Fuhrman, DR, and Wang, B (2006). "A Boussinesq-type Method for Fully Nonlinear Waves Interacting With a Rapidly Varying Bathymetry," *Coastal Eng*, 53(5–6), 487–504. https://doi.org/10.1016/j.coastaleng.2005.11.002.

McLeod, I, and Ringwood, J (2022). "Powering Data Buoys Using Wave Energy: A Review of Possibilities," *J Ocean Eng Mar Energy*, 8, 417–432. https://doi.org/10.1007/s40722-022-00240-3.

Neal, R (2000). *Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification,* Technical Report 9702, University of Toronto, Canada. https://doi.org/10.48550/arXiv.physics/9701026.

Paulsen, BT, de Sonneville, B, van der Meulen, M, and Jacobsen, NG (2019). "Probability of Wave Slamming and the Magnitude of Slamming Loads on Offshore Wind Turbine Foundations," *Coastal Eng*, 143, 76–95. https://doi.org/10.1016/j.coastaleng.2018.10.002.

Penalba, M, et al. (2023). "Bias Correction Techniques for Uncertainty Reduction of Long-Term Metocean Data for Ocean Renewable Energy Systems," *Renewable Energy*, 219, 119404. https://doi.org/10.1016/j.renene.2023.119404.

Rasmussen, CE (2004). *Gaussian Processes in Machine Learning*, Springer Berlin Heidelberg, Berlin, Heidelberg, 63–71. https://doi.org/10.1007/978-3-540-28650-9_4.

Rasmussen, CE, and Williams, CKI (2006). *Gaussian Processes for Machine Learning*, MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001.

Ringwood, JV, Zhan, S, and Faedo, N (2023). "Empowering Wave Energy with Control Technology: Possibilities and Pitfalls," *Annu Rev Control*, 55, 18–44. https://doi.org/10.1016/j.arcontrol.2023.04.004.

Seeger, M (2004). "Gaussian Processes for Machine Learning," *Int J Neural Syst*, 14(02), 69–106. https://doi.org/10.1142/S0129065704001899.

Terrero González, A, Dunning, P, Howard, I, McKee, K, and Wiercigroch, M (2021). "Is Wave Energy Untapped Potential?" *Int J Mech Sci*, 205, 106544. https://doi.org/10.1016/j.ijmecsci.2021.106544.

Trueworthy, A, and DuPont, B (2020). "The Wave Energy Converter Design Process: Methods Applied in Industry and Shortcomings of Current Practices," *J Mar Sci Eng*, 8(11), 932. https://doi.org/10.3390/jmse8110932.

von Schuckmann, K, Holland, E, Haugan, P, and Thomson, P (2020). "Ocean Science, Data, and Services for the UN 2030 Sustainable Development Goals," *Mar Policy*, 121, 104154. https://doi.org/10.1016/j.marpol.2020.104154.

Yaakob, O, Hashim, FE, Mohd Omar, K, Md Din, AH, and Koh, KK (2016). "Satellite-based Wave Data and Wave Energy Resource Assessment for South China Sea," *Renewable Energy*, 88, 359–371. https://doi.org/10.1016/j.renene.2015.11.039.

Yavor, KM, Bach, V, and Finkbeiner, M (2021). "Resource Assessment of Renewable Energy Systems—A Review," *Sustainability*, 13(11), 6107. https://doi.org/10.3390/su13116107.

Zhou, G, Huang, J, Yue, T, Luo, Q, and Zhang, G (2015). "Temporal-spatial Distribution of Wave Energy: A Case Study of Beibu Gulf, China," *Renewable Energy*, 74, 344–356. https://doi.org/10.1016/j.renene.2014.08.014.