Synthetic wave creation

Aina Sánchez Parada, Elena Ruiz de la Cuesta Castaño, and Manel Piera Garrigosa
(Dated: June 4, 2023)

In recent years, the issue of sea level rise and its wide-ranging impacts has raised significant concern. Modeling beach response ti sea level rise is important to design adaptation measures but the wave forcing must be previously established. By analyzing 18 years of sea wave data from a beach in Catalonia, our study aims prepare data analysis for the simulation of potential wave climates based on the collected statistics.

I. INTRODUCTION

Predicting specific beach responses to climate change involves analyzing a variety of factors, including coastal geomorphology, sea level rise, wave climate, sediment dynamics, storm events and human interventions. While it is challenging to make precise predictions due to the complexity and variability of coastal systems, scientists and researchers employ a combination of observational data and computer models to improve our understanding in this area.

This project is based on the analysis of historical wave data with the objective of creating multiple synthetic series for the mathematical modelling of 18 years of a Catalan beach called Cala Castell. In this way, the impact of the stochastic variability and potential future changes in wave patterns on this particular beach could be evaluated. The study works with the measurements of the wave parameters at the Inlet because this data will determine the evolution of the beach morphology. We have faced two challenges. First, available wave data in front of the beach has gaps that must be filled in using other sources. Second, a wave climate emulator has been built that is able to generate synthetic wave conditions that maintain the statistical properties of the original data.

II. METHODS

A. Data Sources

The wave data has been obtained from four different sources. On the one hand, the BUOYCB is a device located in front of Cap Begur that provides 18 years of direct measurements of the wave conditions. Next, the Inlet source is the propagation of BUOYCB data at 14.5 m in front of Cala Castell. This data was already available and the propagation was based on a wave model called SWAN.

On the other hand, there is the SIMAR73 source, which is located 73 meters from the coast in front of Cala Castell and it is made up of time series of wave parameters from numerical modelling. They come, therefore, from simulated data and not from direct measurements of nature. SIMARCB is another source with simulated data from a point in deep water located in front of Cap Begur.

The three relevant parameters of each source to be studied are the period, the height and the direction of the waves.

B. Data Analysis

The first step is to clean the bad data. This step consists on identifying incoherent data and suppressing it from the data series. Then, for each source, we do the following. When at a certain moment the height of the waves does not exist, we eliminate the period and the direction at that instant of time. The direction of incidence of the waves is given with respect to the north. We must subtract the direction from the normal to the bathymetric line where the source is located to obtain directions with respect to the normal. Subsequently, we quantify the distribution and duration of gaps in the four sources individually.

In order to fill in the gaps in the Inlet, we first need to quantify which of the other sources shows a better correlation with it. We generate histograms for the cross sources, analyzing the height, direction, and period of the Inlet in relation to the variables obtained from the other sources. The strongest correlation is observed between BUOYCB and SIMARCB, which it can be proven with the graphs below.

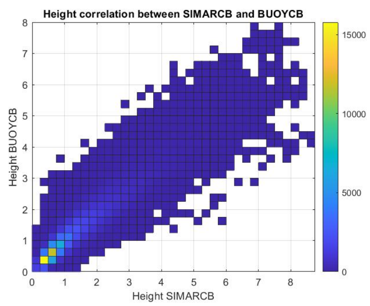

Figure 1. Correlation between the sources BUOYCB and SIMARCB for the height parameter.

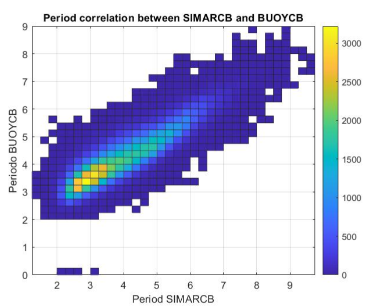

Figure 2. Correlation between the sources BUOYCB and SIMARCB for the period parameter.

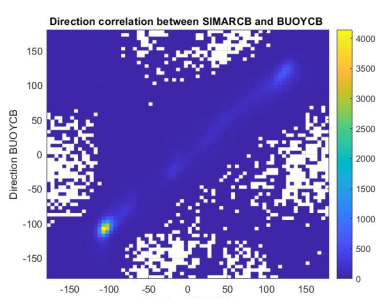

Figure 3. Correlation between the sources BUOYCB and SIMARCB for the direction parameter.

C. Data Imputation

The filling of the Inlet series' gaps was carried out using two methods: interpolation and the Nearest Neighbour Prediction Method (NNPM). The linear interpolation method uses the previous and following values of the gap to find approximate values. This method should work significantly well for gaps of length between 1 and 3 hours. The second method is a procedure designed to fill long gaps and aims to provide series that have the minimum error with respect to the real values (unknown at this point of the analysis).

The NNPM assumes that the correlation between the Inlet and the SIMARCB data is strong enough to ensure that if measurements are similar at SIMARCB for certain times, then measurements at the Inlet are also similar at those times. The procedure is the following: for each gap in the Inlet series, the code looks for the value of the measurements at that moment in the SIMARCB. Next, it searches for other times with similar measurements and saves those positions. Finally, it takes the values of the measurements at the Inlet corresponding to those moments and (once it is ensured that there is available data at those times) the average of all the data collected is taken. This mean value is then the approximation used to substitute the unknown Inlet's value.

The similarity conditions were set according to previous knowledge on the subject and later on checked to verify that the results obtained were satisfactory. The first usage of NNPM method was set to accept as similar values those measurements that differed only in a 5% for the height and the period and in 5°for the direction.

To assess the accuracy of the two methods a test was made by creating gaps in the Inlet series at random positions. The gaps created were of 1 hour length to test the interpolation method and of 15 days length to test the NNPM. The objective was to evaluate the approximation at those instants of time and compare it to the known value.

D. Statistical Analysis

In order to reproduce synthetic time series that could be plausible in reality, a statistical analysis of the only existent real time series (the data from the past) needs to be done. The data from the last 18 years of wave height, period and direction. Each one of these three variables is fitted to a marginal distribution with the Maximum Likelihood method. The variables that describes wave direction, requires a special analysis. The model allows to indicate that a variable is circular by including the key "circular" with the value True. The distribution can be constructed from multiple Probability Models so that it can properly represent the non-exceedance probability of all the plausible values.

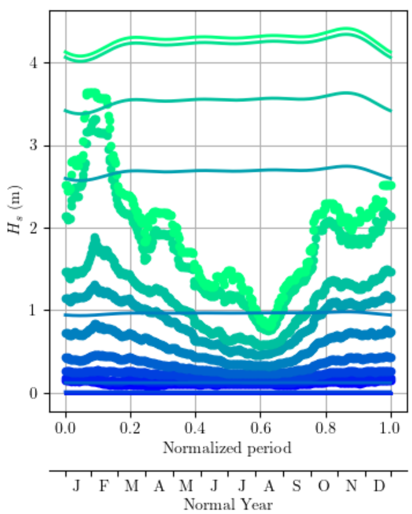

Figure 4. Non-Stationary Cumulative Distribution Function for the wave height where the greener lines represent the higher percentiles.

Once the marginal probability structure of the random variables has been obtained, the temporal multivariate dependency is inferred by using autoregressive methods. After all these procedures the simulation process can be done by retrieving the information from the marginal fit, and the multivariate and temporal dependency analysis to produce a large number of statistically equivalent time series.

III. RESULTS

A. Data Imputation validation

First of all, the interpolation gives pretty good results as it can be seen in figures 5, 6 and 7. The plot of the approximated and real values result in a linear correlation with acceptable width, which shows that the accuracy of this method is an appropriate choice for small gaps.

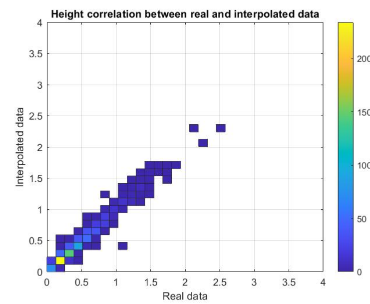

Figure 5. Results of the Interpolation validation test for the height parameter (real data versus interpolated values).

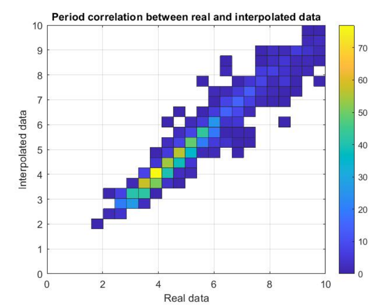

Figure 6. Results of the Interpolation validation test for the period parameter (real data versus interpolated values).

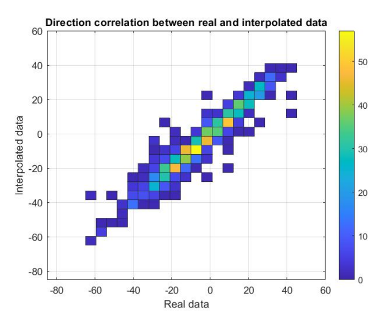

Figure 7. Results of the Interpolation validation test for the direction parameter (real data versus interpolated values).

Secondly, it is vital to check the behaviour of the NNPM. As seen in figures 8, 9 and 10 the accuracy of NNPM is lower than the interpolation method but it is relevant to mention that this procedure is to be applied to long gaps in which the interpolation method would offer poor results. The results presented support the use of these methods for a proper data imputation.

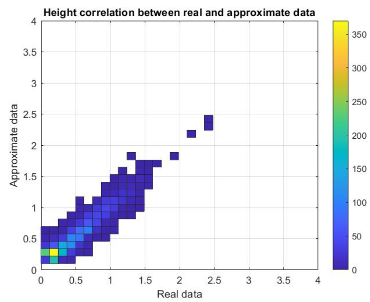

Figure 8. Results of the NNPM validation test for the height parameter (real data versus approximate values).

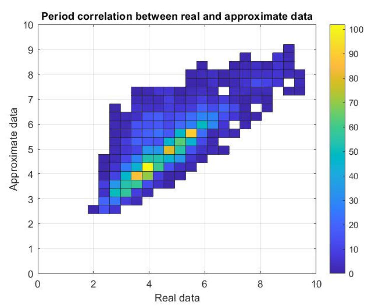

Figure 9. Results of the NNPM validation test for the period parameter (real data versus approximate values).

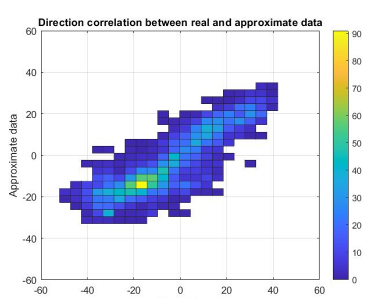

Figure 10. Results of the NNPM validation test for the direction parameter (real data versus approximate values).

B. Data Imputation on Inlet series

In some of the gaps of the Inlet series, the NNPM did not find any similar moments, so for the remaining unknown values the similarity condition was loosen up. Still afterwards, some gaps remain unfilled so for the gaps between 1 and 3 hours that are still left, the interpolation is again applied.

In the end, the number of gaps present in the final Inlet series is 418 gaps in each of the parameters of interest. This represents a reduction of 98.2% with respect to the number of gaps in the original series. Furthermore, the remaining gaps represent a 0.3% of the total data of the Inlet series. In the following figure the data imputation is presented and compared with available data of the original Inlet series.

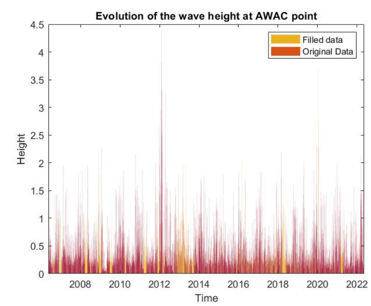

Figure 11. Original Inlet series and Inlet series filled with interpolation and NNPM.

IV. DISCUSSION AND CONCLUSIONS

Although at the end of the project the time series still contains some unknown data, the amount of gaps has been reduced in a 98.2% and the gaps represent a really small part of the total data. Furthermore, an exhaustive analysis of the correlation between sources has been highly successful and has opened new research paths to find even more accurate ways to replenish unknown data.

On this note, it would be enlightening to study how the NNPM method behaves when using the SIMAR73 source as the auxiliary database.

After the completion of the statistical analysis, the next steps would involve creating the synthetic statistically equivalent time series to apply them to the Q2DMorfo and the study of the beach evolution using such series. This would help us understand how the morphology of the beaches change depending on the succession of different climatic events. As well as improving the predictions of present models for the evolution of sea shores.

[1] Francesca Ribas, Laura Portos-Amill, Albert Falqués, Jaime Arriaga, Marta Marcos, Gerben Ruessink (2022) *Impact of mean sea-level rise on the long-term evolution of a mega-nourishment.*

[2] M. Cobos, P. Otiñar, P. Magaña, A. Lira-Loarca, A. Baquerizo (2021) *MarineTools.temporal: A Python package to simulate Earth and environmental time series.*

[3] A. Fernández-Mora, D. Calvete, A. Falqués, H. E. de Swart (2015) *Onshore sandbar migration in the surf zone: New insights into the wave-induced sediment transport mechanisms.*

[4] M. D'Anna, D. Idier, B. Castelle, J. Rohmer, L. Cagigal, F.J. Mendez (2005) *Effects of stochastic wave forcing on probabilistic equilibrium shoreline response across the 21st century including sea-level rise.*

[5] Robert G. Dean, Robert A. Dalrymple (2001) *Coastal Processes with Engineering Applications.*