



HyTCWaves: A Hybrid model for downscaling Tropical Cyclone induced extreme Waves climate

Sara O. van Vloten^{*}, Laura Cagigal, Ana Rueda, Nicolás Ripoll, Fernando J. Méndez

Geomatics and Ocean Engineering Group, Departamento de Ciencias y Técnicas del Agua y del Medio Ambiente, E.T.S.I.C.C.P., Universidad de Cantabria, Santander, Spain

ARTICLE INFO

Keywords:

Tropical cyclone
Hybrid downscaling
Surrogate model
Vortex-type winds
Extreme value distribution

ABSTRACT

Populated coastlines influenced by tropical cyclone (TC) prone areas call for flood risk hazard assessments, including knowledge on the probability of occurrence of major TC-induced significant wave heights. Due to the scarcity of TC historical records, extreme value analyses often rely on fitting generalized extreme value distribution functions to extrapolate longer return periods. This paper describes a methodology that allows to obtain deterministic estimations of the tail probability distribution using long collections of high-fidelity tracks that reproduce similar historical diversity and frequency trends. Given the large dimensionality of the problem (spatiotemporal variability of track geometry and intensity), we implement a track parameterization to easily identify storms in a parametric space. A hybrid approach significantly reduces computational resources by enabling to narrow the number of non-stationary numerically simulated cases forced with vortex-type wind fields parameterized using the Holland Dynamic Model. The proposed surrogate model, HyTCWaves, is trained with a selected subset of maximum significant wave height (MSWH) spatial fields to which a Principal Component Analysis and interpolation functions are performed. Results show a useful approximation of spatial-based regional extreme value distribution of MSWH induced by TCs. The proposed model is applied to the target location of Majuro atoll.

1. Introduction

Tropical cyclones (TCs), also known as typhoons or hurricanes, are among the world's most destructive natural disasters, bringing strong winds, heavy rainfall, large waves and storm surges that devastate property and cause loss of life (e.g. [Chu and Wang, 1998](#); [Diamond et al., 2012](#); [Stephens and Ramsay, 2014](#)). Substantial research has been directed to assess the wide range of hazards ([Mori and Takemi, 2015](#); [Puotinen et al., 2016](#); [Young, 2017](#); [Cagigal et al., 2022](#)) and damaging impacts ([Emanuel, 2020](#); [Ye et al., 2020](#); [Sajjad and Chan, 2020](#)) over increasingly populated coastlines affected by the passage of TCs, including the potential catalyst effect of compound events, i.e. with king tides ([Leonard et al., 2014](#); [Fakhruddin et al., 2022](#)). Moreover, policy makers responsible for flood risk management and coastal defense infrastructures generally require probabilistic analysis that provide the design wave height for standard return periods, 100-year, or even higher, for instance the 1000-year return values used by the United Kingdom for coastal flood boundary conditions ([Environment Agency, 2018](#)). The extreme value (EV) theory seeks to assess the probability of events that are more extreme than any previously observed. Return values for wind and waves are fundamental to assess the risks associated with human activities, but their computation

is complicated by the paucity of observational records ([Breivik et al., 2014](#)). Particularly when dealing with TCs, only short collections of historical data that span for less than five decades can be gathered at a certain target location. Therefore, long return values can only be extrapolated well beyond the range of available data ([Caires, 2011](#)) with the common use of generalized EV distribution fittings. However, long return period estimates carry a great deal of uncertainty.

Here we aim to address this issue by using high-fidelity storm tracks, namely synthetic storms, which allow to generate large samples of thousands of tracks while preserving historical probability distribution functions of TC parameters and temporal correlation. Nevertheless, dealing with huge amounts of tracks would mean very intensive computational demand to characterize wave climate using numerical simulation of wave propagation processes. Alternatively, past studies have accounted for successful applications of hybrid downscaling methods ([Camus et al., 2011, 2013](#); [Antolínez et al., 2018](#); [Anderson et al., 2021](#); [Ricondo et al., submitted](#)) that combine dynamical simulations with statistical techniques allowing to interpolate the dynamical response of a dataset based on a reduced number of simulated output cases, given that such subset is selected ensuring that it is well distributed and explores the full dimensional space. We incorporate the use of a surrogate predictive model trained with precomputed waves based on

^{*} Corresponding author.

E-mail address: sara.ovanvloten@unican.es (S.O. van Vloten).

high-fidelity simulations offering satisfactory accuracy and enhanced computational efficiency (Smith et al., 2015). Past studies have already established advancements for the application of surrogate modeling techniques for storm surge prediction (Jia et al., 2016) although, to our knowledge, this is the first attempt to predict spatial distribution of maximum TC-induced waves. Moreover, principal component analysis (PCA) is integrated as a dimension reduction technique to enhance computational efficiency.

The focus of the current study is to implement and evaluate the skills of the preliminary proposed methodology to assess the EV distribution of maximum significant wave height (MSWH) due to TCs in a target domain. Both the study site and the data used are presented in Section 2. The subsequent steps of the proposed methodology are explained in Section 3, while the EV distribution and the obtained results are shown in Section 4, followed by summary and discussion in Section 5.

2. Application area and data

2.1. Study site

The methodology presented in this paper can be applied to coastal areas surrounded by open waters and subjected to TC activity. Here we have chosen the area of Majuro, the capital of the Republic of the Marshall Islands, to apply HyTCWaves. These archipelagic islands are located in the north-west equatorial Pacific (Fig. 1a), particularly at the east of the most active concentration of TCs in the western tropical Pacific basin which is located between the Philippines and Guam (Camargo et al., 2007). Majuro is an atoll formed by many little low-lying islands which make them extremely prone to storm surge-induced flooding due to TCs. The area of influence of the study site is moderately exposed to TCs with a mean annual rate of 0.6 events/year for the last 60 years, according to historical records (Knapp et al., 2018). However, devastating events were documented by Spennemann (1996) with a storm surge that washed the entire southern part of Majuro (typhoons in 1905, 1908) as well as severe inundations due to TC Alice (1979). Moreover, Ford et al. (2018) compiled damaging impacts due to more recent TCs (Pamela 1982, Roy 1988, Axel 1992, Gay 1992, Paka 1997).

2.2. Bathymetry

In this study, the bathymetry resulted from a combination between the GEBCO (General Bathymetric Charts of the Ocean, <http://www.gebco.net>) global bathymetric dataset, and a high-resolution nearshore bathymetry dataset available for Majuro atoll (Fig. 1a). The GEBCO's gridded bathymetric dataset (GEBCO_2019 grid) is a global terrain model for ocean and land providing elevation data on a 15 arc-second interval grid (equivalent to ~450 m spatial resolution). However, this global dataset lacks resolution for the low-lying islands of the study site. For that reason, a topobathymetric digital elevation model for the Majuro atoll was developed by the U.S. Geological Survey (USGS, <https://www.usgs.gov/>) in collaboration with the U.S. Department of Interior (DOI) Pacific Islands Climate Science Centre (PI-CSC), in order to support the modeling of storm and tide induced flooding. Its grid spacing is 1 m, and it includes the Majuro atoll and extends offshore to a depth of at least 71 m (<https://doi.org/10.5066/F7416VXX>).

2.3. Tropical cyclone databases

The TC input data is collected from two storm track databases:

- (a) Historical database: the IBTrACS v04r00 database (International Best Track Archive for Climate Stewardship; Knapp et al., 2018) compiles global available records of historical storm tracks from 1851 onwards, and it combines track and intensity estimates from several observational sources (Knapp et al., 2010). IBTrACS includes the track location (longitude and latitude), the sea

level pressure and the maximum wind speed (MWS) at six-hour intervals. In this study we have used the World Meteorological Organization (WMO) dataset, which provides the official pressure and wind speed data reported by the responsible agency at every location, in this case the RSMC Tokyo which is operated by the Japanese Meteorological Agency (JMA).

- (b) Synthetic database: for the purpose of this study, high-fidelity tracks generated by Nakajo et al. (2014) are employed to demonstrate the application of the proposed methodology, however other synthetic databases (e.g. Emanuel et al., 2008; Bloemendaal et al., 2020; Nederhoff et al., 2021) or a combination of several databases could be selected instead. These synthetic tracks from a global stochastic model are sensitive to the joint probability distribution functions of TC parameters such as the track, sea level pressure and forward speed, as well as temporal correlations, thereby reproducing the frequency of TCs and diversity of track geometry. Synthetic tracks include information of the longitude, latitude, time and sea level pressure.

In this study, an area of influence to evaluate TCs-induced waves is established as a 4 degrees radius circle centered in Majuro atoll (~450 km radius). Then a subset of storm tracks that enter and/or cross the circle area is extracted from both historical and synthetic TC databases (Figs. 1b and 1c respectively), with 37 historical storm tracks during the period from 1951 until 2021, and 10,064 synthetic storm tracks. This large subset overcomes the shortage of historical events for this particular region. In order to illustrate the TCs intensity in Figs. 1b and 1c, the track line color indicates the TC category 5, 4, 3, 2, 1 corresponding with minimum central pressures lower than 920, 944, 964, 979, 1000 mbar respectively, while category 0 accounts for tropical storms with central pressures higher than 1000 mbar. While the historical subset intensities range between lower categories (0–2), the synthetic subset intensities include all categories (0–5).

3. Methodology

3.1. Overview

The proposed methodology aims to estimate the wave climate EV distribution induced by TCs in the atoll of Majuro and its immediate surroundings. A standard approach would perform a dynamical downscaling of historical events that crossed or occurred near Majuro, simulated with hindcast wind fields forcing, to conduct an EV analysis fitting to characterize the tail distribution of extreme significant wave heights at particular locations. However, this approach is very limited to the number of occurrences, as TCs are scarce phenomena both in time and space. Moreover, the available and statistically representative period of historical records is too short when the interest is to characterize the EV distribution for return periods of 100 years or more, meaning that estimates must be extrapolated with a high degree of uncertainty.

Alternatively, a hybrid approach is proposed in order to overcome those limitations. On one hand, high-fidelity tracks are employed to significantly populate the sample of historical storm tracks, being accountable for thousands of years so that the tail distribution needs no extrapolation. On the other hand, a surrogate model is proposed to unload the number of dynamical simulations as it is impracticable to numerically simulate thousands of events. Moreover, the hybrid approach includes the use of statistical and clustering techniques that allow to reduce the computational resources demand.

Regarding the generation of high-fidelity tracks, the approach used by Nakajo et al. (2014) basically employs a stochastic downscaling method based on Monte Carlo simulations for a sequential development of TCs calculated statistically from given statistical parameters of TC data. Therefore, artificial tracks are sensitive to the approximations of joint probability distribution functions, by using Principal Component Analysis, of several TC parameters and temporal correlations.

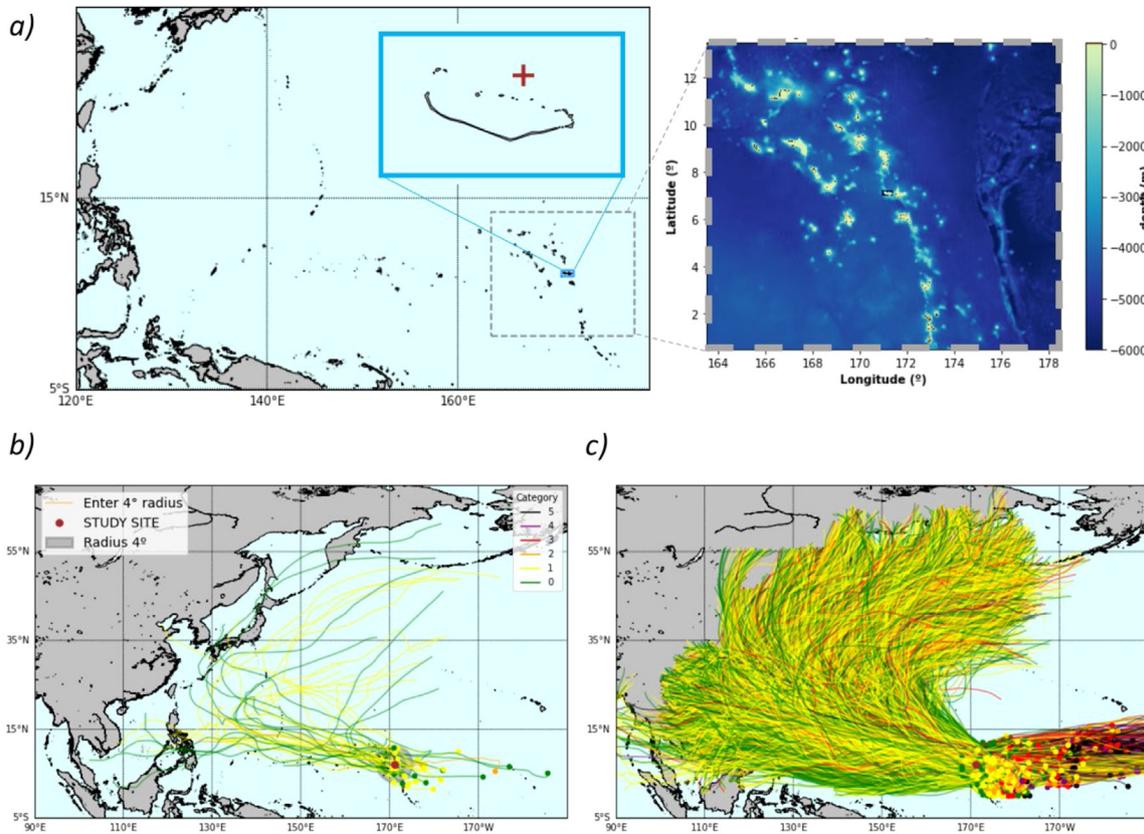


Fig. 1. (a) Study site location of Majuro atoll; gray dashed line encloses the numerical simulation domain and the corresponding GEBCO's bathymetry; blue line zooms the Majuro atoll area; (b) historical, and (c) synthetic storm track subset extraction within a 4 degrees radius circle centered in Majuro.

Fig. 2 illustrates the flow chart of the HyTCWaves methodology, which is composed of the following steps: (a) a stochastic sample of historical and synthetic database events is collected and extracted for a 4 degrees radius circle area of influence, (b) storm tracks are parameterized in terms of a reduced number of variables which characterize its main attributes while reducing the dimensionality of the problem; (c) a selection method is used to generate a subset that preserves TC diversity for a given target location; (d) TC-induced waves are obtained by running dynamic non-stationary simulations forced with parameterized time-varying vortex-type wind fields for each storm track; (e) the spatial-varying conditions for the remaining non-simulated database tracks are reconstructed using PCA coupled with an interpolation technique based on radial basis functions (RBFs); and (f) calculation of the EV distribution. The shaded text boxes (Fig. 2, right column) correspond to the standard dynamical downscaling approach.

The following sections describe in detail the methodology steps, and present the results obtained for the study site as an example of its applicability.

3.2. Track parameterization

The fundamental idea of surrogate-based models is to build a predictive model that is computationally efficient and capable of approximating the value of a function (Goldstein et al., 2019; Anderson et al., 2021). The standard approach includes a learning phase with training examples of past events, linking the design parameters and its corresponding time-varying dynamical response. Therefore, a set of input parameters must be defined to train the surrogate model. The choice of such parameters, in this case, is conditioned by the following criteria: (a) parameters must be available from the storm track database, and (b) parameters must be representative of the track geometry and intensity.

The TC database composed of both historical and synthetic tracks have three common variables: longitude, latitude (location of the storm

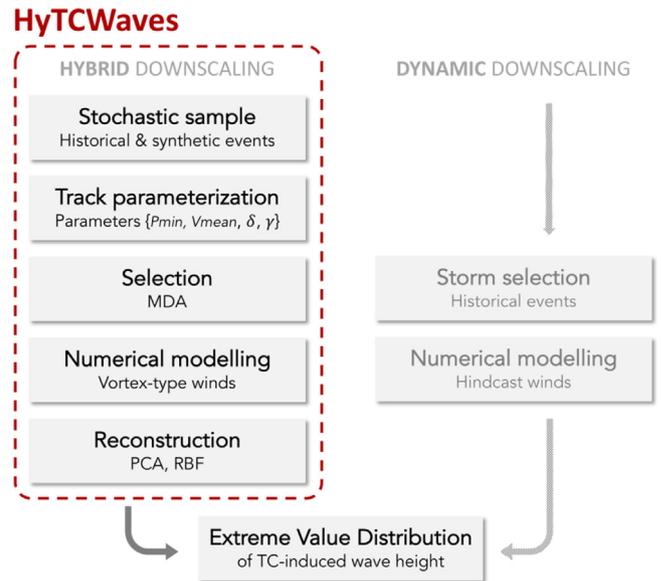


Fig. 2. Flow chart of the HyTCWaves methodology (left column) and the standard dynamical approach (right column).

eye) and central pressure. Besides, the forward speed can be derived from successive storm coordinates. The parametric space aims to explore the diversity of events and to represent each storm by a low-dimensional vector. For that purpose, a simplification of the track geometry within the influence area of 4 degrees radius around Majuro is proposed (Fig. 3) in order to reduce its dimensionality into a small number of variables given by the following geometric, kinematic and

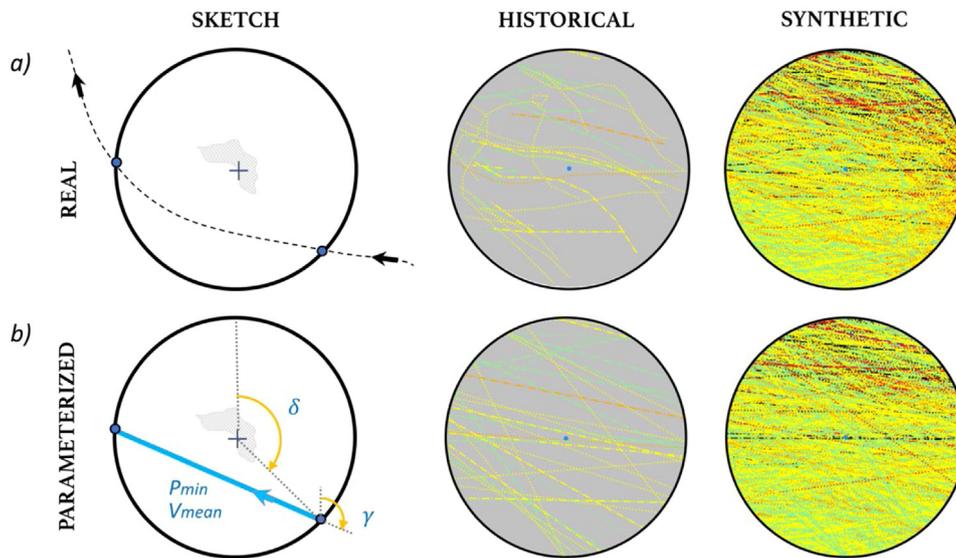


Fig. 3. Storm track parameterization. (a) Real versus (b) parameterized storm tracks. From left to right: sketch, historical and synthetic tracks.

dynamic parameters, which constitute the surrogate model predictor inputs:

- Δ (δ), azimuth of the storm entrance point in the circle influence area;
- Γ (γ), azimuth of the mean direction within the circle using the entrance and exit coordinates;
- P_{min} , minimum central pressure along the storm track coordinates within the circle; and
- V_{mean} , mean forward speed within the circle.

Fig. 3 (left column) illustrates the comparison sketch of a real track geometry versus an approximated track with the ad-hoc parameters. The hypothesis of constraining the track morphology with a constant forward direction (meaning a straight track) simplifies the intrinsic complexity and it allows to represent storm events in a 4-dimensional parametric space. The neglected discrepancies are generally small within the local area of influence, since real and synthetic storms show seldom sinusoidal tracks (Fig. 3, upper row). However, this assumption may limit its applicability in target areas where tracks may be generally more erratic, recurving and convoluted as reported in past review papers (Terry and Feng, 2010; Sharma et al., 2020). A second hypothesis implies that storm intensity is constant along a whole event within the influence area, and the minimum value is set to account for the most adverse possible events. Moreover, the impacts over Majuro will largely depend on the intensity, the RMW, V_{mean} and the minimum distance from land.

All the extracted TCs from historical and synthetic databases for the influence area around Majuro are parameterized, and its results are shown in Fig. 3, where real versus parameterized tracks can be compared for both historical and synthetic tracks. It can be noticed the straightening effect between the circle's entrance and exit coordinates, and the low ratio of quasi-straight and sinuous tracks. Here a 4 degrees radius was established as a balance between accounting for a large enough influence area and ensuring that the simplification of sinuous tracks into straight storms remains a good representation of the TC characteristics. This means that the size of the influence area may be tuned for each location taking into consideration the TC characteristics (i.e. smaller area if convoluted tracks are likely to occur, or larger if tracks are generally quasi-straight). Also, a future line of research could be to perform sensitivity tests on the radius size to the EV analysis. It should be noted that a limitation of this preliminary methodology does not account for potential extreme events produced by TC induced distant swells generated outside the 4 degrees influence area. The

current applicability of the proposed method is restricted to TC-induced waves produced in the study area regional vicinity.

3.3. Selection

The synthetic dataset extracted for the target study site is composed of 10,064 tracks, and the TC track parameters can be visually analyzed with a multi-scatter plot (Fig. 4a), where each dot represents a storm track. When comparing the historical (purple) and synthetic (gray) dots, it can be seen that synthetic tracks explore the parametric space while spanning over the historical data limits, partly because recorded events are very scarce in number. However, some criteria need to be established to remove TC parameters that are not physically plausible or that excessively surmount historical parameters: (1) the minimum central pressure lower limit is set to 860 mbar, corresponding to the minimum historical record plus 1% margin; (2) the mean forward speed upper limit is established by the 99 percentile of historical values plus 5 km/h margin; (3) the azimuth angle (δ) upper limit is 270 degrees since historical tracks are not statistically arriving from the north-west at the target area in the north-west equatorial Pacific; and (4) the entrance angle (γ) lower limit is 40 degrees to remove noticeable synthetic outliers. It should be noted that establishing the filter criteria and the removal of TC parameter outliers is fully dependent on the available historical data within the influence area at the study site. Fig. 4b illustrates the filtered database (gray dots) resulting from applying these criteria, which renders 8773 synthetic tracks.

The updated storm dataset of historical and filtered synthetic tracks for the target study site is composed of 8810 events, a number too large to dynamically downscale as it would be very computationally demanding. Instead, a selection algorithm can be used to determine the subset to eventually train the surrogate model. In this case, the Maximum Dissimilarity Algorithm (MDA) as proposed by Camus et al. (2011) to downscale wave climate to coastal areas has been used. This mathematical tool is used to obtain a reduced number of cases that cover the track variability and guarantee that all possible combinations of track characteristics are represented with a special emphasis on the boundary events of the multivariate parametric space.

The dataset can be defined as $X_i = \{P_{min}, V_{mean}, \delta, \gamma\}_i$ with $i = 1, \dots, N$, the total number of tracks. The scalar and directional variables are previously normalized so that the MDA algorithm is applied over a non-dimensional space, where the target subset of M vectors is initialized by selecting the vector for which its dissimilarity is the largest out of the sample. The remaining M-1 vectors are selected

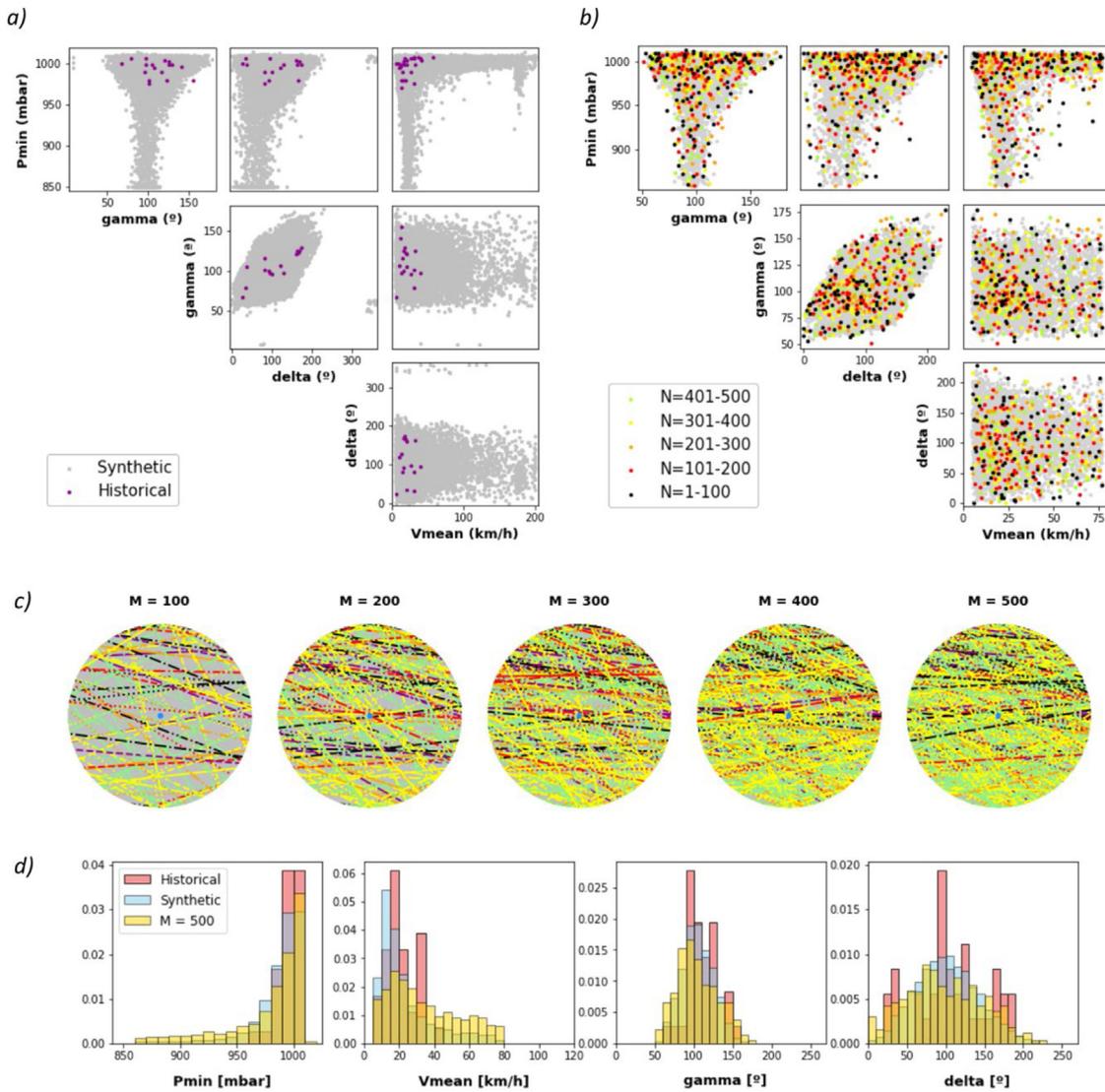


Fig. 4. Storm track parameters: (a) scatter plot of historical (purple dots) and synthetic (gray dots) tracks; (b) scatter plot of filtered dataset (gray dots) and MDA selection (colored dots); (c) parameterized tracks of MDA selection subsets ($M=100, 200, 300, 400, 500$); and (d) normalized histograms of TC parameters ($P_{min}, V_{mean}, \gamma, \delta$).

iteratively by calculating at each iteration the vector which maximizes the dissimilitude in respect to the vectors added to the subset, until a number of M vectors are selected. This selection method has the advantage of allowing an easy iterative analysis of results with incrementing selection sample numbers, as a deterministic tool that further calculates dissimilitude distances with the remaining dataset points.

Fig. 4b shows the multivariate distribution of the MDA subset ($M=500$) in the four-dimensional parametric space, noting that boundary data is included in the selection. The colored dots depict the order in which data points are selected by applying the MDA algorithm to successive badges of 100-intervals (black, red, orange, yellow, green). Fig. 4c illustrates the corresponding parameterized tracks for successive selection samples ($M=100, 200, 300, 400, 500$). It is noted that already in the first 100-badge the selected tracks are well distributed in terms of both angles and intensities. If compared with the historical parameterized tracks (Fig. 3 lower middle), the selection includes more intense events, as expected due to having incorporated synthetic tracks in the dataset sample.

The choice on the selection number is partly subordinated to the simulation computing time. Camus et al. (2011) analyzed the sensitivity effect of the selection number for hybrid model applied to wave propagations and observed that for higher than $M=200$ the decrease of the error was negligible. However, since the current application in this

study involves a higher level of complexity, as we aim to characterize a spatial-based reconstruction and the multivariate problem has more degrees of freedom, we have chosen $M=500$. It seems reasonable in terms of computation, and to explore well enough the parametric space.

Fig. 4d shows the comparative histograms of probability density distribution for the four TC parameters and the historical (red), the synthetic filtered (blue) datasets and the MDA (yellow) subset. It can be noticed that the synthetic filtered dataset exhibits a certain shift of probability distribution to the lower and upper tails, for the minimum central pressure and mean forward speed distributions respectively, consistent with the limiting filtering criteria that was applied over synthetic tracks. Both synthetic angles exhibit a normal probability distribution within the bound limits of historical data. These results account for a synthetic dataset that explores more intense and severe TCs than the historical event records, that will likely account for extreme events. The probability density distribution of the MDA subset shows a general agreement with the dataset despite showing a higher frequency of lower central pressures and fast-moving storms.

3.4. Vortex-type winds parameterization

Once the selected sample of 500 parameterized (straightened) tracks has been defined, the corresponding time-varying wind fields need

to be estimated in order to feed the wave numerical model. Here, a wind parametric model is employed with the advantage of using a small number of input variables. Ruiz-Salcines et al. (2019) evaluated and compared six well-known parametric wind models and concluded that when dealing with a large number of events, the choice of a particular model does not guarantee greater accuracy. The Dynamic Holland Model is based on the vortex-type model developed by Holland (1980, 2008) to provide estimates of maximum winds using an analytic model of the sea level pressure and wind profiles. However, the original model was parameterized to fit the TC instantaneous gradient wind level rather than the winds at surface level of a moving and dynamically developing TC. To overcome this, a modified model was developed by Fleming et al. (2008) to account for the dynamic changes of the TC parameters along the track.

The Dynamic Holland Model estimates the spatial distribution of sea surface wind fields in terms of the following variables: (1) the forward speed, (2) the maximum wind speed at 10 m with a 1-minute sampling interval, (3) the RMW, and (4) the central pressure. However, as it was described the synthetic storm track database does not inform either about the maximum winds or the RMW, therefore these variables must be estimated. On one hand, the IBTrACS storm data provides maximum winds, and the resulting empirical relationship between central pressure and maximum winds depicts a significant correlation (not shown). Therefore, a regression model was performed to obtain 3-order polynomial fittings, for every combination of oceanic basins and Regional Specialized Meteorological Centers (RSMCs), which will be used to estimate maximum winds as a function of the central pressure. Eq. (1) defines the Pmin-Wmax empirical relationship employed for the WMO center and Western Pacific basin, with central pressures in mbar and maximum wind speeds in knots, 1-minute average winds, which are converted to 10-minutes average winds by a factor of 0.93 according to Harper et al. (2010).

$$W_{\max} = -2,85 \cdot 10^{-7} P_{\min}^3 - 8,56 \cdot 10^{-5} P_{\min}^2 + 0,1089 P_{\min} + 297,97 \quad (1)$$

On the other hand, the RMW is a determining factor, along with the decaying rate of wind speed over the distance from the storm eye, to calculate the TC vortex profile that defines the extent or domain subjected to more intense winds. Moreover, Irish et al. (2008) showed that storm size must be considered when estimating surge generation in coastal areas to predict flood risk, even though historically, the famous Saffir–Simpson scale has only provided a classification of damaging TCs intensity in terms of the central pressure and/or the maximum sustained winds. These two variables are generally available whereas there are limited observational measurements of the RMW by radar, flights or satellites (Knaff et al., 2014), which has only been provided by a few RSMCs since 2001 onwards, thus only spanning for the last two decades.

Due to the importance of this relevant variable there have been several efforts aimed at obtaining empirical regression functions, either global (Knaff et al., 2015) or by oceanic basins (Knaff et al., 2007; Tan and Fang, 2018). Despite the uncertainty and complexity involved in its determination, most estimations infer the RMW in terms of the central pressure, the latitude, and the maximum winds, although there are no robust estimates as shown in previous studies. In this study the estimation of Knaff et al. (2015) obtained from observed flight-level TC winds will be employed to calculate the RMW along the storm track, both for historical and synthetic tracks. It is a function of intensity and latitude, in line with several studies that documented the tendency for the RMW to become smaller with increasing intensity (Weatherford and Gray, 1988; Kimball and Mulekar, 2004), and to become larger as the TC moves poleward (Mueller et al., 2006). Evidently, the RMW estimate will have an effect and add on the uncertainty of spatial wind fields and consequently of wave simulations.

3.5. Numerical modeling

The Simulating WAVes Nearshore (SWAN, Booij et al., 1996) third-generation model (version 41.31) is used to predict the wave action evolution in time and space with non-stationary runs forced with large-scale time-varying parametric vortex-type storm wind fields over the numerical domain, which covers the circle influence area and extends to the outer boundaries [163.5°, 178.5°]E, [0.5°, 13.5°]N (Fig. 1a, gray dashed line) to capture the storm regional scale and time evolution over periods of a few days. Simulations are performed for the 500 selected parameterized tracks, which are elongated from the parameterization circle area until reaching the domain boundaries. The extended frame area provides enough computational time to warm up the model's cold start. SWAN was run on a 15 km regular mesh and results were stored at spatially gridded nodes and at a control point located at the north of Majuro.

The SWAN time step is 15 min and frequency space ranges from 0.03 to 1 Hz discretized into 34 bins on a logarithmic scale. The direction space is discretized into 72 sectors (5° for each sector). Simulations activate default white-capping (Komen model), quadruplets and triads. The wind drag coefficient is capped ($C_d < 0.0025$) since SWAN's default bulk wind input formulation (Wu, 1982) may overestimate wind drag coefficients in very intense wind conditions due to TCs (Zijlema et al., 2012), and the wind source term implementation is set to ST6 (Rogers et al., 2012). Compared to Climate Forecast System Reanalysis (CFSR) wind fields, synthetic vortex wind fields may underestimate far-field winds and wave generation by neglecting to consider large-scale meteorological conditions as shown in (Hoeke et al., 2015). A qualitative validation of TCs Ofa (1990) and Paka (1997) vortex-type winds found an agreement with maximum wind warnings issued by the Joint Typhoon Warning Center (JTWC) and the Japan Meteorological Agency (JMA). However, the scope of this study does not include a sensitivity analysis on the wave model which remains an area for future research. Instead, the focus of this study is the development of a hybrid methodology to replicate the outputs of a numerical model. Therefore, the present study does not include a calibration and/or validation of the wave numerical model against measurements of historical TCs which remains an area for future research.

Fig. 5 shows the model's input and output corresponding to the first 100 simulated cases; the TC-induced maximum synthetic vortex wind fields (Fig. 5a) and the MSWH fields (Fig. 5b), namely the swath maps, along the track simulation period. These swath maps provide a representative measure of the track wind/wave footprint. Maximum winds reach up to 60 m/s near the storm eye for most intense storms, and maximum waves are generated up to 14.8 m. The combination of most intense and slow-motion storms usually produces larger waves. Also, the TC forward speed is considered to be the primary factor that contributes to the TC structural asymmetry (Sun et al., 2019) which translates to wave asymmetry due to strong winds upon the right side of the heading direction (in the north hemisphere) during some time dependent on the forward speed, which assimilates to a cumulative energy fetch. The Holland asymmetric model that was transcribed and implemented, takes into account wind asymmetry while obtaining parameterized vortex-type wind fields. In Fig. 5a asymmetry can be observed in some cases where it is more noticeable, for instance in the third-to-last case in the last row (pink rectangle). The TC is moving westwards, and the extent of the maximum wind footprint is larger on the right side from the heading direction, which is translated to a similar pattern in the wave simulations shown in Fig. 5b.

The analysis of the 500 output swath maps illustrates a consistent relationship between irregular spatial distribution of footprint with corresponding TC parameters where the forward speed exceeds the maximum sustained winds. The vortex parameterization calculates the maximum storm wind speed at 10 m by subtracting the storm forward speed from the maximum wind speed, therefore the formulation does not support storms moving faster than the maximum winds. For this reason, the MDA selected cases are filtered to remove the anomalous and non-representative output results.

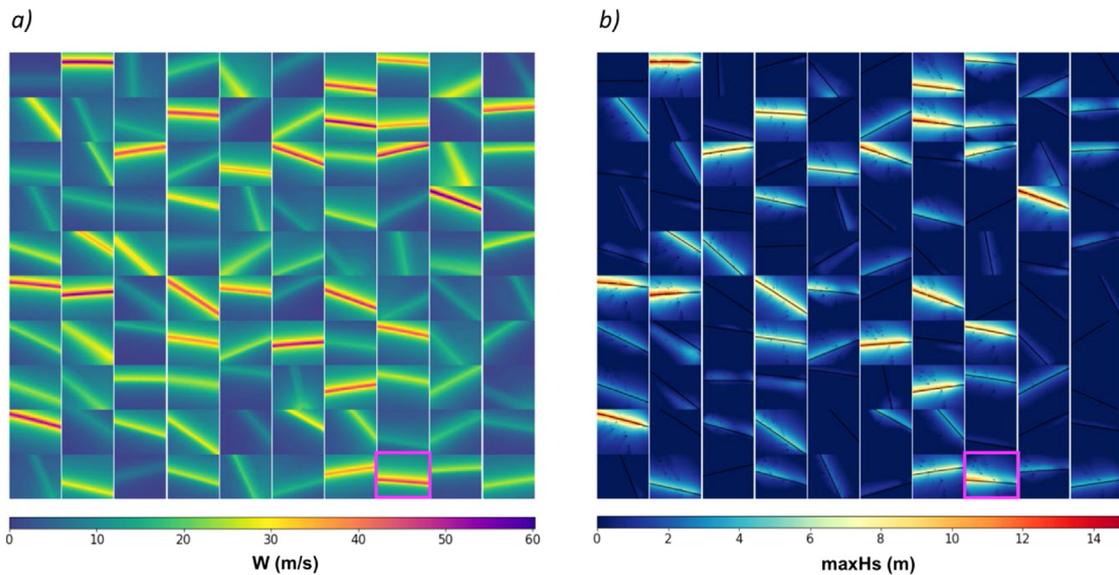


Fig. 5. Grid of (a) maximum wind speed and (b) MSWH swath maps over the simulation period for the first 100 simulated cases.

3.6. Reconstruction

The simulated swath maps of MSWH represent the aggregated variable response, which will be used to train the surrogate model. First, the PCA technique is applied to the spatial swath maps, projecting the original data on a new coordinate system space while aiming to preserve the maximum variance of the data, in such way that new coordinates are sorted in a decreasing order. This allows to reduce the dimensionality of the spatial fields by taking the first n components that explain the most variance, and obtaining the most dominant spatial variability patterns (empirical orthogonal functions, EOFs) with its corresponding temporal coefficients (principal components, PCs). Therefore, the original data $X(x, c)$ can be explained as a linear combination of EOFs (spatial variability) and PCs (case): $X(x, c) = EOF_1(x) \cdot PC_1(c) + EOF_2(x) \cdot PC_2(c) + \dots + EOF_N(x) \cdot PC_N(c)$, with N the dimension of the data grid points. To neglect the initial model warm up period results, only data within the subarea [166°–176°E [2°–12°N] is kept.

The PCA was applied to the standardized sub-area swath maps with the original data's mean and standard deviation in order to remove the effect of dimension scales. The first 4 modes from the PCA analysis explain 92.18% of the variance (65.63%, 19.3%, 4.7%, 2.55% respectively). Fig. 6 shows the results of the first 4 EOFs at the target location multiplied by the standard deviation, and its corresponding standardized PCs. After several tests, it was found that by selecting the first 50 EOFs multiplied by the corresponding PCs, we are able to predict the simulated swath maps as shown in Fig. 6 (dashed box). The spatial correlation of the track footprint is preserved, and the maximum values are correctly estimated.

The PCA technique allows to decompose the wave fields simulated with the MDA sample TC parameters in terms of the most significant modes of oscillation, or in this case, wave features. The linear composition of EOFs and PCs are able to reconstruct the original data employed with PCA. The composition example shown in Fig. 6 shows the MSWH response due to a storm track coming from the east heading WNW, with an extended fetch on the right side since in the northern hemisphere TCs winds move counter-clockwise. The result is consistent with the corresponding TC parameters (960 mbar, 34.3 km/h, gamma 104.5°, delta 87.9°), also regarding the SWH footprint intensity which describes higher maximum values for cases with lower central pressures. The effect of the mean forward velocity is translated in more intense swath maps when the ratio between the maximum winds and the mean forward velocity is close to one, while very fast-moving storms only reach moderate values of maximum SWHs. In those

cases, the storm moves faster than the group velocity and there is not enough fetch to reach fully developed sea conditions. The concept of the extended fetch and how wave trains develop and travel through the TC varying wind field was recently implemented in Kudryavtsev et al. (2021).

Next step is to train the surrogate model with the subset of filtered simulated events, and to estimate the prediction of any other parameterized storm track using RBFs. This is a flexible non-parametric technique used for constructing an approximated function based on the M simulated responses, as it was successfully applied in previous studies (Camus et al., 2011; Gouldby et al., 2014; Rueda et al., 2019). The general form of the RBFs comes as $z(X) \approx RBF(X) = p(X) + \sum_{i=1}^N a_i \cdot \Phi(\|X - X_i\|)$, where $z(X)$ is the surrogate model output (swath maps of the non-stationary aggregated output), X is the vector of input parameters ($p_{min}, v_{mean}, \delta, \gamma$) that determine the track intensity and geometry, and $p(X)$ is formed by a number of monomial equal to the data dimension and a monomial of degree zero, being b the vector of coefficients of these monomials. Therefore, b_i together with a_i , which are the coefficients of the RBFs, are obtained based on the M simulated cases.

The RBFs are trained with the MSWH for the first 50 PCs independently, as this number yielded a good agreement between real and reconstructed swath maps. Moreover, a K -fold cross-validation was performed to assess the statistical model prediction ability. A 4-fold validation using the filtered cases from MDA is shown in Fig. 7. In this case 400 cases are used to train the model, and the plots show a comparison of the remaining reconstructed versus simulated MSWH. Once the RBFs are fitted they can be used to transfer TC parameters with the corresponding swath map, given that the parameters are within the limits of the trained parametric space.

The RBF interpolation technique was previously applied for point-based waves propagation (Camus et al., 2011). From a suitable sample of sea state parameters at a buoy and the corresponding propagated sea state parameters for instance in front of a coastal infrastructure, it is possible to reconstruct the sea surface of any other given sea state by interpolating in the resulting surface from interpolating at the “known” response values. Similarly, the RBF application for the spatial swath maps response due to a sample of TC parameters has proven to work reasonably well with $M=500$. This sample feeds the RBF with information of MSWH consistent with each set of TC parameters which include a wide range of values of central pressure, mean forward speed, angle of entrance and angle of direction. The accuracy of results highly depends on the quality of input SWAN simulations, which have shown

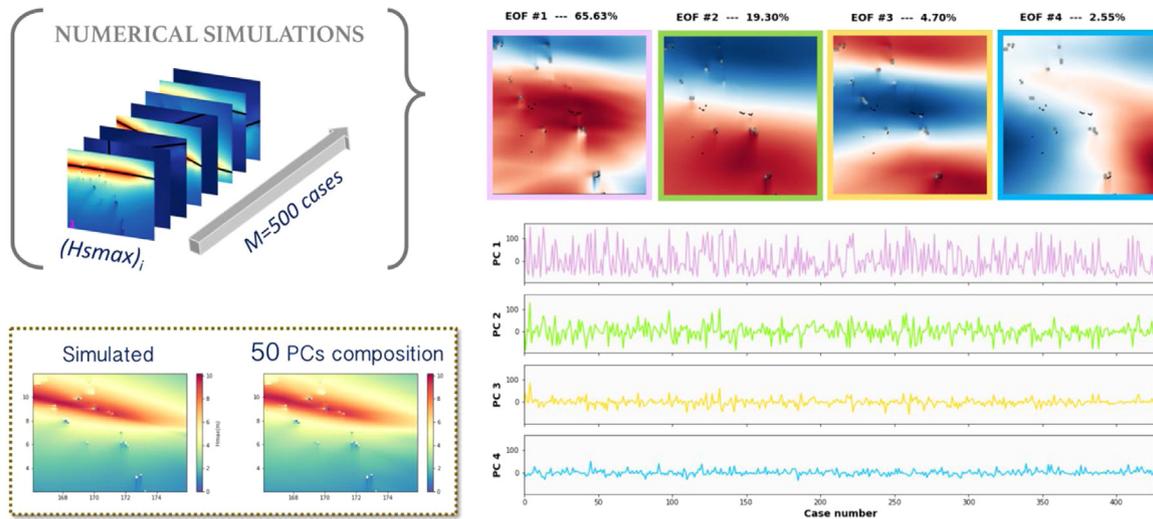


Fig. 6. PCA over simulated swath maps. (Top left) sketch of input swath maps from M simulated cases to compute the PCA; (bottom left) swath map comparison between the numerical simulated case and the prediction using the first 50 EOFs and PCs; (top right) first four EOFs multiplied by the standard deviation; and (bottom right) the corresponding standardized PCs.

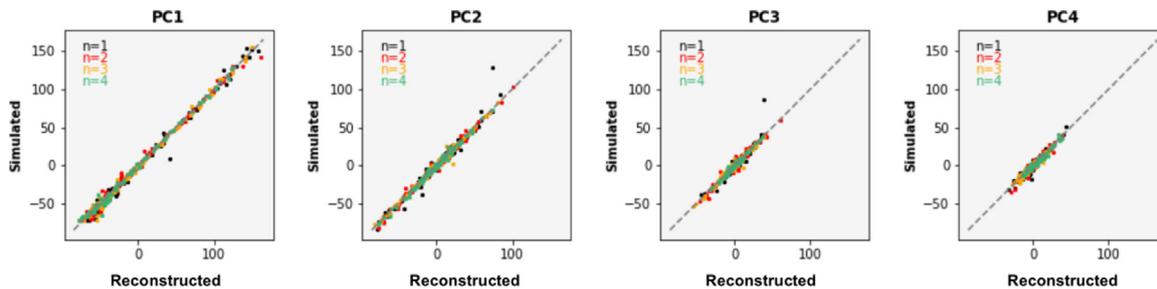


Fig. 7. K-fold cross-validation using MDA simulated cases. Scatter plots of the first 4 PCs comparing simulated and reconstructed values using the interpolated model prediction.

in the previous section to be able to capture features like the extended fetch and wave asymmetry.

4. Extreme value distribution

Extreme values of wave heights govern the design of safety regulations concerning flood risk management, offshore platforms, coastal structures, among other fields of engineering. Thus, it is important to understand and estimate the frequency, intensity and severity of natural hazards (Rueda et al., 2016), as well as the risk of extreme events that threat exposed coastal communities in low-lying coastal areas to support long-term policy decision making. At long term scale, the EV theory aims to quantify the behavior at unusually large values, and to estimate the probability of events larger than any on record (Coles, 2001). Therefore, it mainly deals with the first exceedance of a high-level value expected to occur within the next n years, namely the n -year return value that is exceeded once every n years.

From the multiple methods derived from the EV analysis, one classical method is the annual maxima based on the approximation of the distribution of independent and identically distributed random variables by a member of the generalized extreme value (GEV) family of distribution functions (Haigh et al., 2010). However, the level of uncertainty stems from the quality and quantity of data used. Statistically, longer records imply smaller errors, and additionally, the record should be long enough to encompass the range of variability in extremes (Serafin and Ruggiero, 2014). Longer time series of available historical data usually cover no more than 50 years, meaning that one generally needs to extrapolate well beyond the range of the available data and thus resort to EV analysis to obtain the required return value estimates (Caires, 2011).

In the present study, for the determination of long-return period estimates of MSWH produced by TCs at a particular site, we explore the use of large amounts of virtually independent synthetic events. This approach enables to obtain estimates without the application of GEV functions, where longer return periods than available observational records imply extrapolation estimates, and consequently the tail of the distribution carries larger inherent uncertainty. Instead, the empirical distribution of annual maxima is calculated for modeling the extremes of both the historical and synthetic databases of reconstructed parameterized tracks.

For this purpose, the series of MSWH corresponding to each historical storm track at any given control point from the simulation domain, are independent and identically distributed numerical observations $X = \{X_1, X_2, \dots, X_n\}$, with n observations. These discrete events are grouped by years in order to calculate the spatial annual maxima $Z = \{Z_1, Z_2, \dots, Z_k\}$, with k years. The historical period from season 1979/1980 onwards until 2021, spanning for 42 years, is selected for statistical purposes, since it is considered to be the modern era when geostationary satellite coverage is more widely available than in previous years. Therefore, a total 22 storms that entered or crossed the circle area of influence around Majuro, thereby computing a mean occurrence rate of $\lambda = 0.52$ storm tracks per year. The empirical distribution function evaluated in the k th ordered annual maxima (Z_k) is $\hat{F}(Z_k) = k / (m + 1)$, and the corresponding return period $TR(Z_k) = 1 / (1 - \hat{F}(Z_k))$ is obtained in terms of the mean number of TCs per year and the probability of exceedance of sorted annual maxima.

The same procedure is applied to evaluate the empirical distribution function of the synthetic database. However, the 8773 stochastic events lack chronological time reference, and according to the mean historical rate of occurrence ($\lambda = 0.52$), they account for 15,671 years

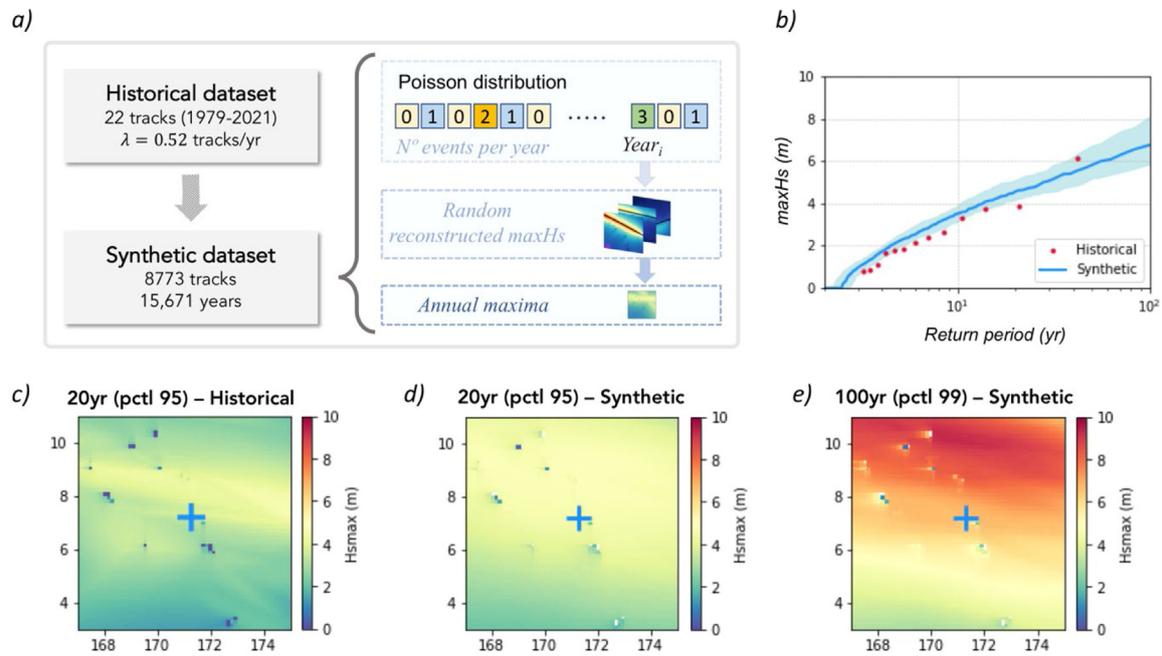


Fig. 8. Extreme value distribution. (a) Flow chart procedure to generate annual maxima swath maps; (b) return probability plot of historical (red dots) and synthetic (blue line) reconstructed swath maps at the control point in Majuro; (c) 20-year return plot for historical data; (d) 20-year and (e) 100-year return plots for synthetic data.

of high-fidelity simulated tracks. Thus, in order to generate plausible events preserving the historical occurrence probabilities, a Poisson distribution with $\lambda = 0.52$ is used to obtain the number of events per year ($N_1, N_2, \dots, N_{15,500}$) for 15,500 years (Fig. 8a). Then, the series of N_i events is populated by randomly extracting a reconstructed swath map from the synthetic database, so that the annual maxima can then be calculated, meaning at each grid node the maximum value of all available event per year is taken to determine the annual maxima (whether multiple, one or none events). In the context of using Monte Carlo simulations of thousands of years, the peak over threshold (POT) method can be used with the Poisson-GPD (Generalized Pareto distribution) model estimating frequencies and intensities, which on the other hand is a method analog for threshold exceedances of the GEV distribution for annual maxima (Coles, 2001), meaning both methods are entirely consistent with one another above the GPD threshold. In order to analyze the confidence interval of the synthetic reconstructed TCs, the series of years have been split in 31 portions of 500 years each. Fig. 8b shows the probability plot comparing the empirical distribution function resulting from the historical (red dots) and synthetic (blue line) reconstructed swath maps, at the control point located in the north of Majuro (Fig. 1a). It can be observed that the mean synthetic distribution and the 90% confidence intervals between the 5th and 95th percentiles mostly match the historical results, and also are capable to extend to longer return periods.

Since the MSWH was reconstructed over a spatial domain, it is possible to obtain the 31 series of 500-year annual maxima at all grid nodes. Moreover, return plots are obtained by computing the q -th percentile of the data along the case axis corresponding to the m -year return values. Fig. 8c and 8d show the spatial variability of the 20-year return plots (95-percentile) for historical and synthetic reconstructed events. The most noticeable difference is the effect of the scarce number of historical data which obtains less track footprints due to less populated calibration years while the synthetic output shows less spatial variability due to the large amount of storm data (in the previous section it was demonstrated that the use of PCA preserves spatial correlation).

The presented methodology is capable of obtaining spatially return period estimates, without the need of fitting analytical distributions to the empirical distribution function (at each point) and thereby

extrapolating extreme values for high return periods. As an example, Fig. 8e illustrates the 100-year return plot (99-percentile), where MSWH reaches up to 10 m in the northern area of the domain and 6.9 m at the control point. Moreover, the proposed methodology can be further applied for regional downscaling simulations with increasing resolution at smaller areas closer to the target location.

5. Summary and discussion

In TC prone areas it is important to acquire knowledge on the probability of occurrence of major TC-induced storm surge and MSWH as key tools for policy management. Here, we aim to characterize low-probability extreme events of regional MSWH produced by TCs. The proposed methodology is based on (a) the use of stochastic sample of high-fidelity storm tracks, (b) parameterization of TC tracks in terms of geometric, kinematic and dynamic parameters to reduce dimensionality, (c) MDA selection of a reduced and representative subset of cases, (d) use of vortex-type parameterized winds for each storm track, (e) non-stationary runs of SWAN forced with time-varying wind fields, (f) the application of PCA over the selected output cases' swath maps and RBF interpolation of the first PCs to reconstruct any TC-induced swath map, and (g) the EV distribution based on synthetic reconstructed TCs.

The developed surrogate model provides a preliminary methodology that allows to obtain estimations of a computationally intensive physical process model using a hybrid approach that only simulates a reduced number of cases that are used to train the model which can reconstruct any given storm within the original parametric space. The surrogate model is applied to one location in Majuro atoll in the equatorial North-west Pacific, exposed to an average of one TC every two years for the last four decades. The storm parameterization allows to reduce the dimensionality of complex track geometries, an assumption which is acceptable in TC prone areas that exhibit seldom convoluted or sinuous tracks. Results are also dependent on the use of the constant minimum central pressure along the track which ensures to yield the most adverse MSWH. The MDA selection algorithm allows for an efficient low number of cases to be simulated to train the model and, the K-fold cross validation demonstrates the good skill of the surrogate model. However, to feed the PCA, good quality trained

outputs are required. In that sense, the surrogate model's ability to reproduce simulated cases from the reconstructed PCs showed an evident improvement when the velocity criteria were applied to remove storms moving at absolute speeds higher than the maximum sustained winds. Overall, this methodology reduces significant computational resources since only a limited number of cases need to be numerically simulated and the reconstruction calculations are not demanding. At last, regional estimates are provided, however the methodology allows to further downscale waves by simulating high-resolution nested meshes at nearshore domains that correctly solve the physical processes and consider the local effect of bathymetry at intermediate water levels. The numerical skill of the methodology has been assessed based on K-fold validation and the EV distribution comparisons. Although it is out of scope of this paper, the SWAN model validations could be addressed in the future with available altimeter data.

HyTCWaves provides new means to gain insight on the upper tail of extreme distribution functions by using a large sample of synthetic storm tracks. High-fidelity tracks developed by Nakajo et al. (2014) have been used in this study, however other synthetic databases can be used in the future, or a combination of them, to evaluate global differences. Moreover, it is possible to include a disturbance factor on the storm intensity and/or frequency to account for climate change scenarios (Mori and Takemi, 2015). Additionally, this study has focused on simplified storm track events, however there is room to evaluate more complex parameterizations in the future that allow for close-to-real representation of tracks (i.e. temporal variability of TC intensity, distance-weighted storm selection). Also, it is recommended to further examine sensitivity analysis on several steps of the methodology: (a), optimization of the minimum number of selection cases needed, (b) effect on introducing a weighted distance in the MDA selection algorithm to obtain more tracks near the target location (less tangent tracks far away), (c) effect on different selection algorithms, (d) size of the influence circle area, and (e) number of PCs trained for interpolation. As described in previous sections, the methodology implements certain assumptions which derive in some limitations the user should take into consideration. The storm track parameterization provides a close-to-real representation of quasi-straight tracks, while convoluted tracks would be poorly represented as well as the hazards produced by such events. Moreover, the applicability is restricted to TC-induced waves produced in the regional vicinity of the study area, defined by a 4 degrees radius in the present study, although a sensitivity analysis of the radius size is recommended in future research. Additionally, the present study does not include a calibration and/or validation of the wave numerical model against measurements of historical TCs which still remains an area for future research. Instead the focus of this study was addressed towards the development of a hybrid methodology that replicates the outputs of a numerical model.

In summary, the proposed methodology has the ability of coupling numerical modeling, data-mining and statistical tools to obtain a cost-effective approach for producing a large database of maximum swath maps induced by TCs. We expect this preliminary methodology will be useful to end-users as a first approximation of regional EV distribution of MSWH induced by TCs.

CRedit authorship contribution statement

Sara O. van Vloten: Conceptualization, Methodology, Software, Validation, Writing – original draft, Visualization. **Laura Cagigal:** Supervision, Writing – review & editing. **Ana Rueda:** Supervision, Writing – review & editing. **Nicolás Ripoll:** Software, Writing – review & editing. **Fernando J. Méndez:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been partially funded by the Beach4Cast PID2019-107053RB-I00 project, granted by the Spanish Ministry of Science and Innovation. AR acknowledge the funding from Juan de la Cierva-Incorporación IJC2020-043907-I/ MCIN/AEI/10.13039/501100011033 and “NextGenerationEU”/PRTR.

References

- Anderson, D., Ruggiero, P., Mendez, F.J., Barnard, P., Erikson, L., O'Neill, A., Merrifield, M., Rueda, A., Cagigal, L., Marra, J., 2021. Projecting climate dependent coastal flood risk with a hybrid statistical dynamical model. *Earth's Future* 9, <http://dx.doi.org/10.1029/2021EF002285>.
- Antolínez, J.A.A., Murray, A.B., Mendez, F.J., Moore, L.J., Farley, G., Wood, J., 2018. Downscaling changing coastlines in a changing climate: the hybrid approach. *J. Geophys. Res. Earth Surf.* 123 (2), 229–251. <http://dx.doi.org/10.1002/2017JF004367>.
- Bloemendaal, N., Haigh, I.D., de Moel, H., Muis, S., Haarsma, R.J., Aerts, J.C.J.H., 2020. Generation of a global synthetic tropical cyclone hazard dataset using STORM. *Sci. Data* 7, 40. <http://dx.doi.org/10.1038/s41597-020-0381-2>.
- Booij, N., Holthuijsen, L.H., Ris, R.C., 1996. The SWAN wave model for shallow water. In: *Proc. 25th Int. Conf. Coastal Engng. Orlando, USA. Vol. 1*. pp. 668–676. <http://dx.doi.org/10.1061/9780784402429.053>.
- Breivik, O., Aarnes, O., Abdalla, S., Bidlot, J., 2014. Wind and wave extremes over the world oceans from very large ensembles. *Geophys. Res. Lett.* 41 (14), 5122–5131. <http://dx.doi.org/10.1002/2014GL060997>.
- Cagigal, L., Mendez, F.J., van Vloten, S.O., Rueda, A., Coco, G., 2022. Wind wave footprint of tropical cyclones from satellite data. *Int. J. Climatol.* 1–10. <http://dx.doi.org/10.1002/joc.7764>.
- Caires, S., 2011. Extreme Value Analysis: Wave Data. World Meteorological Organization/JCOMM, Geneva, Switzerland, p. 33. <http://dx.doi.org/10.25607/OBP-1499>, (JCOMM Technical Report No. 57).
- Camargo, S., Robertson, A., Gaffney, S., Smyth, P., Ghil, M., 2007. Cluster analysis of typhoon tracks, Part I: general properties. *J. Clim.* 20 (14), 3635–3653. <http://dx.doi.org/10.1175/JCLI4188.1>.
- Camus, P., Mendez, F.J., Medina, R., 2011. A hybrid efficient method to downscale wave climate to coastal areas. *Coast. Eng.* 58 (9), 851–862. <http://dx.doi.org/10.1016/j.coastaleng.2011.05.007>.
- Camus, P., Mendez, F.J., Medina, R., Tomas, A., Izaguirre, C., 2013. High resolution downscaled ocean waves (DOW) reanalysis in coastal areas. *Coast. Eng.* 72, 56–68. <http://dx.doi.org/10.1016/j.coastaleng.2012.09.002>.
- Chu, P.S., Wang, J.X., 1998. Modeling return periods of tropical cyclone intensities in the vicinity of hawaii. *J. Appl. Meteorol.* 37 (9), 951–960. [http://dx.doi.org/10.1175/1520-0450\(1998\)037<1\textless>0951:MRPOTC<1\textless>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1998)037<1\textless>0951:MRPOTC<1\textless>2.0.CO;2).
- Coles, S., 2001. An introduction to statistical modeling of extreme values. In: *Springer Series in Statistics*, London, U. K. <http://dx.doi.org/10.1007/978-1-4471-3675-0>.
- Diamond, H.J., Lorrey, A.M., Knapp, K.R., Levinson, D.H., 2012. Development of an enhanced tropical cyclone tracks database for the southwest Pacific from 1840 to 2010. *Int. J. Climatol.* 32 (14), 2240–2250. <http://dx.doi.org/10.1002/joc.2412>.
- Emanuel, K., 2020. Evidence that hurricanes are getting stronger. *Proc. Natl. Acad. Sci. USA* 117, 13194–13195. <http://dx.doi.org/10.1073/pnas.2007742117>.
- Emanuel, K., Sundararajan, R., Williams, J., 2008. Hurricanes and global warming: results from downscaling IPCC AR4 simulations. *Bull. Am. Meteorol. Soc.* 89, 347–367. <http://dx.doi.org/10.1175/BAMS-89-3-347>.
- Environment Agency, 2018. Coastal flood boundary conditions for the UK: update 2018. https://assets.publishing.service.gov.uk/media/60365303e90e0740ac3ea1b5/Coastal_flood_boundary_conditions_for_the_UK_2018_update_-_user_guide.pdf. (last Accessed: 2022).
- Fakhrudin, B., Kintada, K., Hassan, Q., 2022. Understanding hazards: Probabilistic cyclone modelling for disaster risk to the Eastern Coast in Bangladesh. *Prog. Disaster Sci.* 13, <http://dx.doi.org/10.1016/j.pdisas.2022.100216Get>.
- Fleming, J., Fulcher, C., Luettich, R., Estrade, B., Allen, G., Winer, H., 2008. A real time storm surge forecasting system using ADCIRC. In: *Estuarine and Coastal Modeling 10th International Conference*. [http://dx.doi.org/10.1061/40990\(324\)48](http://dx.doi.org/10.1061/40990(324)48).
- Ford, M., Merrifield, M., Becker, J., 2018. Inundation of a low-lying urban atoll island: Majuro, Marshall Islands. *Nat. Hazards* 91 (3), 1273–1297. <http://dx.doi.org/10.1007/s11069-018-3183-5>.
- Goldstein, E.B., Coco, G., Plant, N.G., 2019. A review of machine learning applications to coastal sediment transport and morphodynamics. *Earth-Sci. Rev.* 194, 97–108. <http://dx.doi.org/10.1016/j.earscirev.2019.04.022>.
- Gouldby, B., Mendez, F.J., Guanache, Y., Rueda, A., Mínguez, R., 2014. A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis. *Coast. Eng.* 88, 15–26.
- Haigh, I., Nicholls, R., Wells, N., 2010. A comparison of the main methods for estimating probabilities of extreme still water levels. *Coast. Eng.* 57 (9), 838–849. <http://dx.doi.org/10.1016/j.coastaleng.2010.04.002>.

- Harper, B.A., Kepert, J.D., Ginger, J.D., 2010. Guidelines for Converting Between Various Wind Averaging Periods in Tropical Cyclone Conditions. TCP Sub-Project Report, WMO/TD-No. 1555, World Meteorological Organization.
- Hoeke, R.K., McInnes, K.L., O'Grady, J.G., 2015. Wind and wave setup contributions to extreme sea levels at a tropical high island: a stochastic cyclone simulation study for Apia, Samoa. *J. Mar. Sci. Eng.* 3, 1117–1135. <http://dx.doi.org/10.3390/jmse3031117>.
- Holland, G.J., 1980. An analytic model of the wind and pressure profiles in hurricanes. *Mon. Wea. Rev.* 108 (8), 1212–1218. [http://dx.doi.org/10.1175/1520-0493\(1980\)108<T1\textless>1212:AAMOTW\(T1\textgreater>\)2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1980)108<T1\textless>1212:AAMOTW(T1\textgreater>)2.0.CO;2).
- Holland, G., 2008. A revised hurricane pressure-wind model. *Mon. Wea. Rev.* 136 (9), 3432–3445. <http://dx.doi.org/10.1175/2008MWR2395.1>.
- Irish, J., Resio, D., Ratcliff, J., 2008. The influence of storm size on hurricane surge. *J. Phys. Oceanogr.* 38 (9), 2003–2013. <http://dx.doi.org/10.1175/2008JPO3727.1>.
- Jia, G., Taflanidis, A.A., Nadal-Caraballo, N.C., Melby, J., Kennedy, A., Smith, J., 2016. Surrogate modeling for peak or time-dependent storm surge prediction over an extended coastal region using an existing database of synthetic storms. *Nat. Hazards* 81, 909–938. <http://dx.doi.org/10.1007/s11069-015-2111-1>.
- Kimball, S.K., Mulekar, M.S., 2004. A 15-year climatology of north atlantic tropical cyclones, Part I: size parameters. *J. Clim.* 17, 3555–3575. [10.1175/1520-0442\(2004\)017_3555:AYCONA.2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(2004)017_3555:AYCONA.2.0.CO;2).
- Knaff, J., Longmore, S., Demaria, R., Molenaar, D., 2015. Improved tropical-cyclone flight-level wind estimates using routine infrared satellite reconnaissance. *J. Appl. Meteor. Climatol.* 54 (2), 463–478. <http://dx.doi.org/10.1175/JAMC-D-14-0112.1>.
- Knaff, J., Longmore, S., Molenaar, D., 2014. An objective satellite-based tropical cyclone size climatology. *J. Clim.* 27, 455–476. <http://dx.doi.org/10.1175/JCLI-D-13-00096.1>.
- Knaff, J., Sampson, C., Marchok, T., Gross, J., McArdie, C., 2007. Statistical tropical cyclone wind radii prediction using climatology and persistence. *Weather Forecast* 22, 781–791. <http://dx.doi.org/10.1175/WAF1026.1>.
- Knapp, K.R., Diamond, H.J., Kossin, J.P., Kruk, M.C., Schreck, C.J., 2018. International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4. NOAA National Centers for Environmental Information, <http://dx.doi.org/10.25921/82ty-9e16>.
- Knapp, K., Kruk, M., Levinson, D., Diamond, H., Neumann, C., 2010. The international best track archive for climate stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bull. Am. Meteorol. Soc.* 91, 363–376. [10.1175/2009BAMS2755.1](http://dx.doi.org/10.1175/2009BAMS2755.1).
- Kudryavtsev, V., Yurovskaya, M., Chapron, B., 2021. Self-similarity of surface wave developments under tropical cyclones. *J. Geophys. Res. Oceans* 126 (4), <http://dx.doi.org/10.1029/2020JC016916>.
- Leonard, M., Westra, S., Phatak, A., Lambert, M., Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., Stafford-Smith, M., 2014. A compound event framework for understanding extreme impacts. *WIREs Clim. Change* 5, 113–128. <http://dx.doi.org/10.1002/wcc.252>.
- Mori, N., Takemi, T., 2015. Impact assessment of coastal hazards due to future changes of tropical cyclones in the North Pacific ocean. *Weather Clim. Extrem.* 11, 53–69. <http://dx.doi.org/10.1016/j.wace.2015.09.002>.
- Mueller, K.J., DeMaria, J.A., Knaff, J.P., Kossin, M., Vonder Haar, T.H., 2006. Objective estimation of tropical cyclone wind structure from infrared satellite data. *Weather Forecasting* 21, 990–1005. <http://dx.doi.org/10.1175/WAF955.1>.
- Nakajo, S., Mori, N., Yasuda, T., Mase, H., 2014. Global stochastic tropical cyclone model based on principal component analysis and cluster analysis. *J. Appl. Meteor. Climatol.* 53 (6), 1547–1577. [10.1175/JAMC-D-13-08.1](http://dx.doi.org/10.1175/JAMC-D-13-08.1).
- Nederhoff, K., Hoek, J., Leijnse, T., van Ormondt, M., Caires, S., Giardino, A., 2021. Simulating synthetic tropical cyclone tracks for statistically reliable wind and pressure estimations. *Nat. Hazards Earth Syst. Sci.* 21, 861–878. <http://dx.doi.org/10.5194/nhess-21-861-2021>.
- Puotinen, M., Maynard, J.A., Beeden, R., Radford, B., Williams, G.J., 2016. A robust operational model for predicting where tropical cyclone waves damage coral reefs. *Sci. Rep.* 6 (26009), <http://dx.doi.org/10.1038/srep26009>.
- Ricondo, A., Cagigal, L., Rueda, A., Ripoll, N., Mendez, F.J., 2022. Hywaves: hybrid downscaling of multimodal wave-climate in small pacific. *Islands. Ocean Model.*
- Rogers, W.E., Babanin, A.V., Wang, D.W., 2012. Observation-consistent input and whitecapping dissipation in a model for wind-generated surface waves: Description and simple calculations. *J. Atmos. Oceanic Technol.* 29 (9), 1329–1346.
- Rueda, A.C., Cagigal, L., Pearson, S., Antolinez, J.A.A., Storlazzi, C., Van Dongeren, A., Camus, P., Mendez, F.J., 2019. HyCReWW: A hybrid coral reef wave and water level metamodel. *Comput. Geosci.* 127, 85–90.
- Rueda, A., Camus, P., Mendez, F.J., Tomás, A., Luceño, A., 2016. An extreme value model for maximum wave heights based on weather types. *J. Geophys. Res. Oceans* (121), <http://dx.doi.org/10.1002/2015JC010952>.
- Ruiz-Salcines, P., Salles, P., Robles-Díaz, L., Díaz-Hernández, G., Torres-Freyermuth, A., Appendini, C., 2019. On the use of parametric wind models for wind wave modeling under tropical cyclones. *Water (Switzerland)* 11 (10), <http://dx.doi.org/10.3390/w11102044>.
- Sajjad, M., Chan, J.C.L., 2020. Tropical cyclone impacts on cities: a case of Hong Kong. *Front. Built Environ.* 6, <http://dx.doi.org/10.3389/fbuil.2020.575534>.
- Serafin, K.A., Ruggiero, P., 2014. Simulating extreme total water levels using a time-dependent, extreme value approach. *J. Geophys. Res. Oceans* 119, 6305–6329. <http://dx.doi.org/10.1002/2014JC010093>.
- Sharma, K., Magee, A., Verdon-Kidd, D., 2020. Variability of southwest Pacific tropical cyclone track geometry over the last 70 years. *Int. J. Clim.* 41 (1), 529–546. <http://dx.doi.org/10.1002/joc.6636>.
- Smith, J.M., Taflanidis, A.A., Kennedy, A.B., 2015. Surrogate Modeling for Hurricane Wave and Inundation Prediction. http://dx.doi.org/10.1142/9789814611015_0007.
- Spennemann, D., 1996. Nontraditional settlement patterns and typhoon hazard on contemporary Majuro Atoll, Republic of the Marshall Islands. *Environ. Manage.* 20 (3), 337–348. <http://dx.doi.org/10.1007/BF01203842>.
- Stephens, S.A., Ramsay, Doug, 2014. Extreme cyclone wave climate in the southwest Pacific ocean: Influence of the El Niño southern oscillation and projected climate change. *Glob. Planet. Change* 123, 13–26. <http://dx.doi.org/10.1016/j.gloplacha.2014.10.002>.
- Sun, Z., Zhang, B., Zhang, J.A., Perrie, W., 2019. Examination of surface wind asymmetry in tropical cyclones over the northwest Pacific ocean using SMAP observations. *Remote Sens.* 11 (22), 2604. <http://dx.doi.org/10.3390/rs11222604>.
- Tan, C., Fang, W., 2018. Mapping the wind hazard of global tropical cyclones with parametric wind field models by considering the effects of local factors. *Int. J. Disaster Risk Sci.* 9, 86–99. <http://dx.doi.org/10.1007/s13753-018-0161-1>.
- Terry, J., Feng, C., 2010. On quantifying the sinuosity of typhoon tracks in the western north Pacific basin. *Appl. Geogr.* 30 (4), 678–686. <http://dx.doi.org/10.1016/j.apgeog.2010.01.007>.
- Weatherford, C.L., Gray, W.M., 1988. Typhoon structure as revealed by aircraft reconnaissance, Part I: data analysis and climatology. *Mon. Wea. Rev.* 116, 1032–1043. [http://dx.doi.org/10.1175/1520-0493\(1988\)116_1032:TSARBA.2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1988)116_1032:TSARBA.2.0.CO;2).
- Wu, J., 1982. Wind-stress coefficients over sea surface from breeze to hurricane. *J. Geophys. Res.* 87 (9704).
- Ye, M., Wu, J., Liu, W., He, X., Wang, C., 2020. Dependence of tropical cyclone damage on maximum wind speed and socioeconomic factors. *Environ. Res. Lett.* 15 (9).
- Young, I., 2017. A review of parametric descriptions of tropical cyclones wind-wave generation. *Atmosphere* 8 (10), 194. <http://dx.doi.org/10.3390/atmos8100194>.
- Zijlema, M., van Vledder, G.P., Holthuijsen, L.H., 2012. Bottom friction and wind drag for wave models. *Coast. Eng.* 65, 19–26.