

## Estimating tropical cyclone-induced wind, waves, and surge A general methodology based on representative tracks

Bakker, Tije M. ; Antolínez, Jose A.A. ; Leijnse, T.W.B.; Pearson, Stuart .G.; Giardino, Alessio

### DOI

[10.1016/j.coastaleng.2022.104154](https://doi.org/10.1016/j.coastaleng.2022.104154)

### Publication date

2022

### Document Version

Final published version

### Published in

Coastal Engineering

### Citation (APA)

Bakker, T. M., Antolínez, J. A. A., Leijnse, T. W. B., Pearson, S. . G., & Giardino, A. (2022). Estimating tropical cyclone-induced wind, waves, and surge: A general methodology based on representative tracks. *Coastal Engineering*, 176, Article 104154. <https://doi.org/10.1016/j.coastaleng.2022.104154>

### Important note

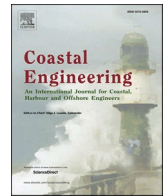
To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

### Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

### Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.



# Estimating tropical cyclone-induced wind, waves, and surge: A general methodology based on representative tracks

Tije M. Bakker<sup>a,b</sup>, José A.A. Antolínez<sup>a,\*</sup>, Tim W.B. Leijnse<sup>b</sup>, Stuart G. Pearson<sup>a,b</sup>, Alessio Giardino<sup>c</sup>

<sup>a</sup> Department of Hydraulic Engineering, Delft University of Technology, 2600 GA, Delft, the Netherlands

<sup>b</sup> Unit Marine and Coastal Systems, Deltares, 2600 MH, Delft, the Netherlands

<sup>c</sup> Sustainable Development and Climate Change Department (SDCC), Asian Development Bank, 6 ADB Avenue, Mandaluyong City 1550, Metro Manila, Philippines

## ARTICLE INFO

### Keywords:

Tropical cyclones  
Flood hazard  
Hybrid downscaling  
Dissimilarity selection  
Synthetic tropical cyclones  
TCWiSE

## ABSTRACT

Tropical Cyclones (TCs) are singular storms causing intense wind, large waves, extreme water levels, and heavy rainfall. TCs prove every year to be one of the most destructive natural phenomena worldwide. The quantitative assessment of the hazards resulting from TCs (i.e., flooding and extreme winds) is challenging since satellite data are only available for recent decades, whereas older historical observations are incomplete and less accurate. In addition, long-term prediction through numerical weather forecasting is still limited. This often results in large uncertainties in the definition of TC hazards associated with events with longer return periods or in areas infrequently impacted by TCs. Even when this information is available, for example through statistical sampling of synthetic TC tracks, the numerical modelling of the associated hazards for all the different TC conditions can lead to computational costs which are often infeasible. Several methodologies that overcome the issues of accuracy and computational efficiency currently exist, but these are not generically applicable, and they tend to focus on specific areas only, for example where TCs typically make landfall. The main contribution of this paper is a novel methodology for the estimation and analysis of TC hydro-meteorological conditions and induced hazards. The method is generically applicable and maximizes accuracy while accounting for computational efficiency. Our approach identifies a smaller but representative set of TC tracks (RTCs) that preserves the information about extremes and the frequency of events of the larger population. The method is successfully applied and validated in a case study in the Bay of Bengal, using a set of synthetic TC tracks representing 1000 years of TC climate. For the best-performing configuration, the required number of scenarios and associated computational costs were reduced by 90% while maintaining accuracy in the simulated offshore storm surges, significant wave height, and windspeeds typically within 10% of the prediction based on the original full set of scenarios. This method is globally applicable and greatly improves the efficiency of TC-related hazard estimation, making it particularly valuable for areas with limited historical data.

## 1. Introduction

Tropical Cyclones (TCs), also called cyclones, hurricanes or typhoons depending on the region, are rotating systems of clouds and thunderstorms, driven by heat transfer from the ocean, and triggered by pre-existing weather disturbances above warm tropical or subtropical waters (Emanuel, 2003). TCs often result in extreme flooding due to a combination of extreme storm surges, waves, and precipitation. According to Peduzzi et al. (2012), on average 87 TCs develop each year, of which 35 make landfall and can affect an estimated 1.53 billion people

and a total asset value of US\$ 16,218 billion (2010 values). However, TC induced hazards and associated risks vary significantly dependent on geographic area. Furthermore, future projections show that a combination of climate change impacts, including sea-level rise and a possible increase in TC intensity in some regions (Knutson et al., 2015), will further enhance the vulnerability of low-lying coastal areas to TCs. Therefore, regions that were rarely affected by TCs may also become prone to more frequent TC events (e.g., Dekker et al., 2018).

The importance of understanding TC hazards and risks can be illustrated by the tragic event of Cyclone Nargis (2008), which made landfall

\* Corresponding author.

E-mail addresses: [j.a.a.antolinez@tudelft.nl](mailto:j.a.a.antolinez@tudelft.nl) (J.A.A. Antolínez), [tim.leijnse@deltares.nl](mailto:tim.leijnse@deltares.nl) (T.W.B. Leijnse).

<https://doi.org/10.1016/j.coastaleng.2022.104154>

Received 8 August 2021; Received in revised form 14 April 2022; Accepted 29 May 2022

Available online 30 May 2022

0378-3839/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

in the Ayeyarwady delta in Myanmar, a location along the Bay of Bengal. According to the International Best Track Archive for Climate Stewardship database (IBTrACS; Knapp et al., 2018), which has been recording data since 1972, no TC had made landfall before at this location (Fritz et al., 2009). Local populations were also not expecting TCs in this area and were therefore unprepared. Areas up to 50 km inland flooded as a result of sustained wind speeds exceeding 210 km/h, in combination with storm surges of more than 5 m and wave heights of 2 m (Fritz et al., 2009). An estimated 138,000 people lost their lives, mainly due to flooding. The occurrence and severity of Nargis could have only been predicted on short notice through operational forecasting, as TCs are singular and short-lived events (Chen et al., 2019; Magnusson et al., 2019). Nevertheless, TCs similar to Nargis could have been anticipated based on historical observations and experience in the Bay of Bengal through probabilistic assessments of TC activity (Lin and Emanuel, 2015) as well as the hydro-meteorological conditions and hazards resulting from TCs. Note that in this paper we refer to hydro-meteorological conditions as the direct effect of TCs (e.g., waves, storm surges, wind and precipitation), while hazards are the result of these hydro-meteorological conditions (e.g., flooding) impacting assets and people in the coastal area.

Several variables influence the hazards resulting from a TC (Lin and Emanuel, 2015), so it is not trivial to assess them. Some of these variables are strictly related to the TC themselves (e.g., track (or landfall) location, wind speed intensity, pressure drop, storm size, forward speed, and precipitation rates). Other factors depend on the geographical area where the TC develops, propagates, and finally makes landfall (e.g., topography and bathymetry, presence of rivers, vegetation or reefs, and presence of possible flood protection measures). Hence, a wide range of TC scenarios as well as detailed models are required to account for the combination of all these factors. This makes the estimation of the hazards challenging because (a) the estimation may be very location specific, (b) historical data (i.e., Historical Tropical Cyclone (HTC) tracks, hydro-meteorological conditions) are generally scarce and cover a relatively short historical period, which does not permit a comprehensive statistical description of the forcing conditions, (c) even when this information is available, computational costs to represent all possible scenarios resulting from the combination of all different variables becomes extremely demanding.

Several methods are available that overcome some of these limitations of general applicability, accuracy, and efficiency. Sebastian et al. (2017) developed a modelling framework to estimate storm surge and precipitation based on synthetic TC landfall locations coupled to a simple one-dimensional empirical wind setup model for Galveston Bay. The approach shows accurate results but is specific to the case for which it was developed and implemented. Zheng et al. (2013) and Xu et al. (2014) applied data-driven techniques to estimate the (joint) probability of occurrence of storms and TC induced water levels and precipitation. As the methods are data-driven, they are accurate within the conditions for which they have been derived, however they are less suitable e.g., to represent the joint probability of occurrence of water levels and precipitation for low frequency events (e.g., the Nargis event) or at areas infrequently impacted by TCs. Similarly, Vousedoukas et al. (2018) and Torres et al. (2015) presented global and local flood hazard estimates, respectively, based on the use of actual and synthetically shifted HTCs. Consequently, results by Vousedoukas et al. (2018) are more accurate in areas frequently impacted by TCs, while those in the case of Torres et al. (2015) are with a specific local focus. For assessment of the impact of climate change on tropical cyclones and associated coastal hazards, several methods exist, see for example Mendelsohn et al. (2012), the pseudo global warming approach by Jyoteeshkumar Reddy et al. (2021) and the dynamical downscaling method by Mori and Takemi (2016).

To overcome the issue of accuracy and limited historical records in areas seldom impacted by (extreme) TCs, several methods have been developed that use Synthetic Tropical Cyclones (STC). STC emulators generate long datasets of realistic TC tracks by sampling relevant

physical parameters within ranges as observed in the HTCs (e.g., wind speeds, forward speed, heading), and have been widely used (see e.g., Vickery et al., 2000; James and Mason, 2005; Emanuel et al., 2006; Nakajo et al., 2014; Haigh et al., 2014; Neumann et al., 2015; Bloemendaal et al., 2020; Nederhoff et al., 2021; Arthur, 2021; Hojjat Ansari et al., 2021). An alternative is the method described by Niedoroda et al. (2010) and used for flood hazard assessments in the U.S. This method makes use of idealized synthetic TC track scenarios with likelihood based on statistics of observed cyclone events, therefore making it accurate but location-specific. Recently, Dullaart et al. (2021) have simulated global storm surge induced by tropical cyclones using a dataset of STCs (Bloemendaal et al., 2020). Nevertheless, these estimates lack the contribution of waves and are based on a global model set-up.

Simulating long synthetic time series of TCs in process-based models (e.g., for the estimation of design waves and water levels or flooding simulations) comes with high computational costs. Computational efficiency is poor as similar tracks can occur several times and are simulated individually. Therefore, a logical solution is to limit the number of STCs that must be simulated. A number of studies have attempted to reduce the number of realistic TC tracks, including manually selecting relevant tracks (e.g., Ou et al., 2002) and by clustering of historical tracks (e.g., Camargo et al., 2007; Choi et al., 2009). However, these studies did not focus on including extremes nor a higher variability of TC events than observed in the HTC record.

Hybrid methods focus on reducing the number of tracks simulated (e.g., Resio et al. (2009) and Toro et al. (2010), and more recently Jia and Taflanidis (2013), Kim et al. (2015), Jia et al. (2016) and Bass and Bedient (2018)). Such methods proved capable of accurately representing a larger set of scenarios. However, they mainly focussed on TCs making landfall and their characteristics at that location. Nevertheless, for a correct representation of TC-induced hydro-meteorological conditions and resulting hazards, TCs that do not make landfall are often also relevant (e.g., a TC passing far offshore may generate waves that propagate to the coastline).

In this study, we present a novel hybrid method for estimating TC-induced hydro-meteorological conditions, which can be used directly as a basis for TC hazard assessments. The method is generally applicable to any location in the world. We use a large dataset of realistic STC track scenarios (generated based on HTC track data) to include a wide range of plausible TC events in the analysis, also covering areas with infrequent TCs or incomplete historical records. From the full set of STCs, we select a comprehensive subset of Representative Tropical Cyclones (RTCs) that describe the hydro-meteorological conditions corresponding to the full set. RTCs are selected based on the physical characteristics of the full STC track, whereby hydro-meteorological conditions resulting from all STC are represented, including the extremes, TCs that pass at large distances, and TCs that do not make landfall. Subsequently, the RTCs are simulated in hydrodynamic models, whereby the limited number of RTCs allows for relatively high-resolution modelling of their complex dynamics. By reconstructing the hydro-meteorological conditions for the original set of STC scenarios, essential information on the probability of occurrence is preserved.

The method was applied to a case study at the Bay of Bengal (BoB). Different configurations of the method were tested, aimed at both downscaling tracks for the full Bay of Bengal area or for more local applications (i.e., O(100 km) coastline) and providing insight into the suitability of the method for different applications at different scales. For the best performing configuration, we also provide some insights in the performance for different number of RTCs, which can be used to balance efficiency and accuracy in future applications. In our case study we make use of a STC dataset generated by a specific emulator and specific hydrodynamic models. However, the method is conceptually relatively simple and flexible, so in principle it can be applied in combination with any TC track dataset and hydrodynamic model.

An overview of the proposed method is described in Section 2. Our application to the BoB is described in Section 3. Results are presented in

Section 4. Assumptions and limitations of the study are discussed in Section 5, while the conclusions of the study are summarized in Section 6.

## 2. Overview of the methodology

A general methodology to estimate tropical cyclone-induced wind, waves, and surge using RTCs is introduced (schematic shown in Fig. 1). In this manuscript we apply it into the offshore region (~30 m depth). The method is split into 6 steps:

1. **Collection of HTC track data**, used as a basis for the analysis (Section 3.2), and pre-processing.
2. **Generation of STC tracks** based on the HTC track data using a TC emulator (Section 3.2). With this step, longer time series of TCs are obtained, covering a wider area in greater density than the HTCs. In our application, a set of STCs including spatio-temporal wind fields derived following Leijnse et al. (2022) was used, generated with the Tropical Cyclone Wind Statistical Estimation Tool (TCWiSE, Nederhoff et al., 2021), although alternative TC emulators could be used. Our work builds on theirs by providing a more efficient and effective way of using TC emulators.
3. **Sampling of a representative subset of RTC tracks** (Section 3.3). In our application, we sample from the STCs, but alternatively the HTC tracks could be used directly for areas where sufficient data is available. In our application, we aim to approximate offshore hydro-meteorological conditions (waves, storm surge and wind) for the full set of TCs. For the design of the RTC subset, physical TC characteristics relevant for quantifying their resulting offshore hydro-meteorological conditions are used.
4. **Simulation of the RTCs** to obtain the hydro-meteorological conditions (Section 3.4). In this study, a dataset following Leijnse et al. (2022) of simulated storm surge and waves for the full sets of HTCs and STCs was used. For validation purposes of this new

methodology, computed hydro-meteorological conditions for the full sets of STCs and HTCs were used. However, this will normally not be required when applying the methodology as proposed here, as a small subset of RTC can be simulated instead of the full STC dataset, therefore leading to large savings in computational time.

5. **Reconstruction of hydro-meteorological conditions** for each of the TCs in the full set (TCrecs) (Section 3.5). In our application STCs are used, giving the reconstructed STCs (STCrecs). By using the full STC set and RTCs, hydro-meteorological conditions for each of the STCrecs are reconstructed stochastically based on similarity of the same physical TC characteristics as used for the selection of RTCs. Because of this step, information about the probability of occurrence of the hydro-meteorological conditions is preserved.
6. **Extreme Value Analysis (EVA)** is applied to the offshore hydro-meteorological conditions of the STCrecs to approximate the quantiles for different Return Periods (RP) (Section 3.6). In our application, EVA is also applied to the sets of HTCs and STCs to assess performance of the method.

The optimal configuration and sensitivity of the methodology to its application in regional and local study cases are additionally discussed by comparing output from subsets of RTCs (ranging from 2 to 1744 TCs in size) with those obtained from a full set of STCs (1745 TCs), and with those modelled from the original HTCs (81 TCs).

## 3. Application

We apply the method described in Section 2 to the case study for the BoB. First the study area is described in Section 3.1. Subsequently, the application of each of the 6 steps of the method is described in Sections 3.2–3.6.

We make use of a dataset derived following Leijnse et al. (2022) that consists of simulated storm surge and wave conditions for a large set of HTC and STC tracks in the BoB (details can be found in the specific sections). This unique dataset makes it possible to assess the performance of the method as derived in this paper, as hydro-meteorological conditions based on RTCs can be compared with those for the full set of STC, as well as with those for HTCs. In our application, we quantify hydro-meteorological conditions in terms of the maximum significant wave height ( $H_{s,L}$ ), maximum storm surge ( $SS_L$ ), and maximum sustained wind speed ( $v_{max,L}$ ) per TC.

### 3.1. Study area

The BoB is located in the north-eastern part of the Indian Ocean (Fig. 2). Countries bordering the BoB are, in clockwise direction: Sri Lanka, India, Bangladesh and Myanmar. Within the bay lie the Andaman and Nicobar Islands in the east (territories of India). The methodology is presented as a regional setting case study, at the scale of the BoB, and at nine (9) specific locations: Batticaloa (Sri Lanka), Madras, Visakhapatnam and Puri (India), Charchenga and Chittagong (Bangladesh), Sittwe and Mawlamyine (Myanmar) and Port Blair at the Andaman and Nicobar Islands (India) (Fig. 2).

On average, there are about 2.3 TC events per year in the Indian Ocean basin of which 1.7 occur in the BoB. The wave conditions, storm surge, and wind speeds resulting from these events vary significantly by TC, and location within the BoB. TCs are relatively infrequent in the Northern Indian Ocean basin compared to other TC basins (e.g., compared to an average of 7.1 TC events per year in the North Atlantic basin (Nederhoff et al., 2021)). However, in terms of casualties, the most extreme TCs of the past decades – including Cyclone Nargis (2008), described in Section 1 – all occurred in this basin.

To illustrate the details of our method, we chose two locations in the Bay of Bengal on which to focus our analysis. Charchenga (Bangladesh) and Batticaloa (Sri Lanka) were selected because of their relative different location and TC climate. Charchenga is fronted by an extensive

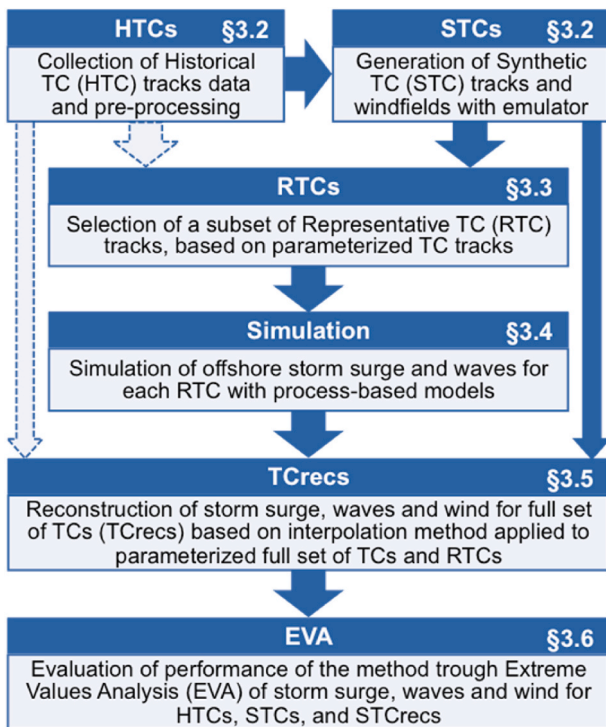
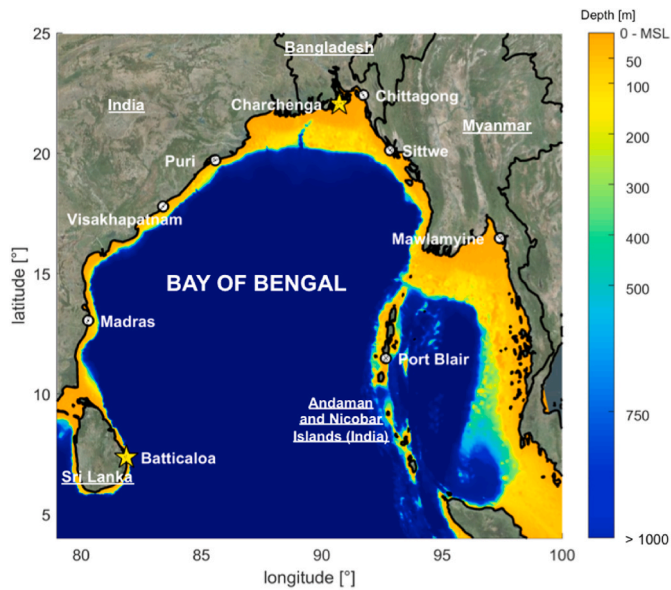


Fig. 1. Schematic overview of the methodology proposed in this paper. In our application, we generate STCs, select the RTCs as a subset from the STCs and reconstruct the Synthetic TCrecs (STCrecs), but alternatively the RTCs could also be selected from the HTCs directly (grey arrows).





**Fig. 2.** Map of the study area: Bay of Bengal (BoB). Surrounding countries (white, underlined) with borders (black) and studied locations (white) are indicated. In our analysis, we focus on Batticaloa and Charchenga locations (yellow star). Depth relative to Mean Sea Level (MSL) is indicated (color scale); shallower continental shelf areas are clearly visible (warmer colors).

(shallow) continental shelf area and TCs make relatively frequent landfall here, including some of the TCs with the highest windspeeds (Fig. 3). At the other side of the study area, Batticaloa is in an area without an extensive continental shelf, where TCs infrequently make landfall and typically have lower windspeeds. Additional results for locations other than Charchenga and Batticaloa can be found in Supplementary material A.

### 3.2. HTC and STC - historical and synthetic TC tracks

HTC tracks were derived from the IBTrACS database version v04r00 (Knapp et al., 2018). From the database the subset of the Joint Typhoon Warning Centre was used, which contains 110 HTC tracks for the Northern Indian Ocean basin for the period 1972–2020. Eighty-one of these originated from the Bay of Bengal, including recent Cyclone

Amphan (2020), see Fig. 3a. HTC tracks are described by the three hourly coordinates of the location of the TC eye (latitude, longitude) and the maximum sustained wind speed ( $v_{max}$ ).

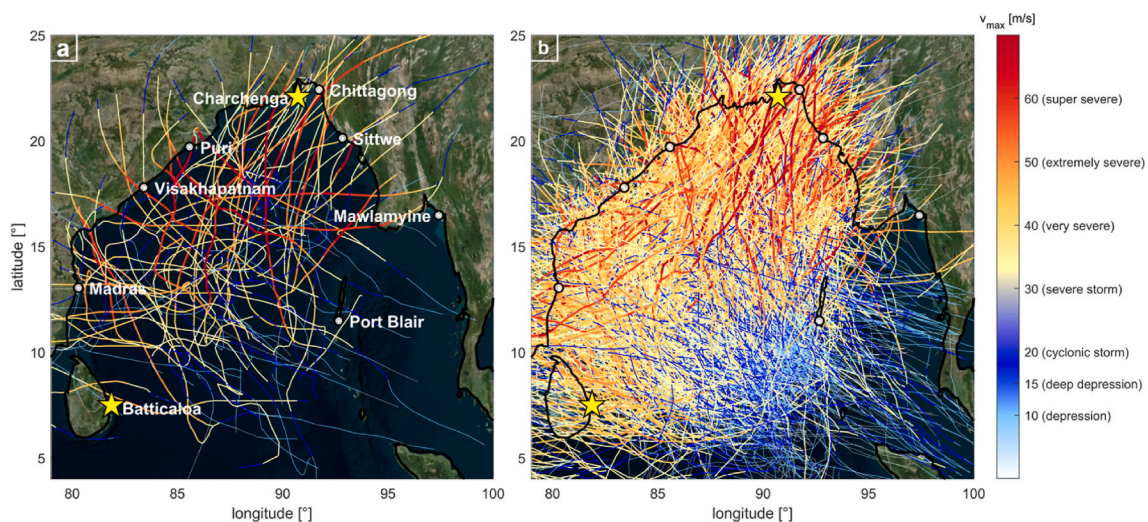
Based on the HTC track data, 1000 years of STC tracks were generated following Leijnse et al. (2022) using a synthetic emulator of TCs, here the TCWiSE tool (Nederhoff et al., 2021). TCWiSE uses an Empirical Track Model based on Markov-chains to generate synthetic TC tracks. It can be used in any oceanic basin and on any (historical) data source. In this tool, tracks are parameterized by the coordinates of the location of the TC eye (latitude, longitude) and the maximum sustained wind speed ( $v_{max}$ ) on a three-hourly basis. This resulted in 1745 STC tracks for the BoB (Fig. 3b), which cover a much wider area and in much greater density than the HTCs. Both qualitative and quantitative comparisons of HTC and STC track genesis, passing, and termination locations show good agreement (Leijnse et al., 2022). Subsequently, TCWiSE is used to create the spatio-temporal wind and pressure fields used as input for model simulations (see Leijnse et al., 2022).

### 3.3. RTC - representative TC track selection

The STC tracks from the dataset (Section 3.2) were screened to select a limited number of tracks representative of the full dataset. RTC tracks were selected using the Maximum Dissimilarity Algorithm (MDA), first described by Kennard and Stone (1969). An important advantage of MDA is that the selected set of scenarios spans the full range of the parameter space, including the extremes or outliers. The hydro-meteorological conditions resulting from the TC depend on many factors spanning a broad parameter space, from the translation speed at the eye of the TC to windspeeds at hundreds of km away of the eye of the TC. However, the number of parameters used to describe the TC scenarios in MDA needs to be limited for optimal performance. Hence, a balanced set of parameters that summarizes all relevant characteristics of the scenarios is required for MDA to be effective.

#### 3.3.1. TC descriptors and parameters

To describe the different physical characteristics of the TC, multiple descriptors can be defined, which in turn are represented by one or more parameters. An overview of the different descriptors and parameters is provided in Table 1. In this study, track location, TC duration, maximum sustained wind speed ( $v_{max}$ ), and translation speed ( $c$ ) of each TC track were used as descriptors. In case the TC hazard for only a specific location is of interest (hereafter called Selected Location, SL), the track



**Fig. 3.** Overview of TC tracks in the Bay of Bengal and studied locations. Charchenga and Batticaloa locations are highlighted (yellow star). (a) 81 HTC tracks for the 1972–2020 period. (b) 1745 STC tracks generated with TCWiSE (corresponding to 1000 years of the current TC climate). Track color is a measure of the maximum sustained wind speed ( $V_{max}$ ) of the TC at that location. Classification according to the North Indian Ocean scale of the India Meteorological Department.

**Table 1**

Overview of descriptors with parameters and applied factors to modify the normalized parameter space for the tested configurations. Three different sets of parameter space modifications were tested: FB, FBMPs, and SLMPs configurations. The number of parameters per descriptor ( $N_{par}/des$ ) is used as correction factor in all configurations. Weight factors ( $F_w$ ) and scaling factors ( $F_s$ ) are applied to each of the normalized parameters in the descriptor according to eq. (1). DTSL (Distance Track to Selected Location) and ATSL (Azimuth of line of DTSL) parameters are only used in configuration SLMPs.

Descriptor	Parameters	Full Basin (FB)		Full Basin Modified Parameter Space (FBMPs)		Selected Location Modified Parameter Space (SLMPs)		
		Names	$N_{par/des}$	$F_w$	$F_s$	$F_w$	$F_s$	
Track location	$lat_{gen}, lon_{gen},$ $lat_{1/3}, lon_{1/3},$ $lat_{2/3}, lon_{2/3},$ $lat_{term}, lon_{term}$	8	1	1	2	1	2	1
Duration of the TC	$Duration$	1	1	1	1	1	1	1
Translation speed (c)	$c_{min}, c_{median},$ $c_{max}$	3	1	1	1	1	1	1
Maximum wind speed ( $v_{max}$ )	$v_{max,min}, v_{max,median},$ $v_{max,max}$	3	1	1	3	3	3	3
Relative location of the TC track	$DTSL$	1	–	–	–	–	3	1/3
	$ATSL$	1	–	–	–	–	1	1

location relative to SL is additionally used. As an alternative to the wind speed, other studies have also used the air pressure (e.g., Toro et al., 2010), but those two variables are related (e.g., Holland, 2008). In our application, wind speeds were preferred over pressure drop as they are the principal drivers for coastal hazards resulting from wind, waves and storm surge (e.g., coastal flooding). Besides, several studies have used the angle of approach and landfall location to characterize the TC track (e.g., Toro et al., 2010; Sebastian et al., 2017). In our application, track location and relative track location descriptors were preferred over angle of approach and landfall location, as in this schematization the location of the full TC track is better described. Furthermore, TCs that do not make landfall can be included in the analysis as they can also contribute to coastal hazards.

The track location descriptor was parameterized by the latitude and longitude coordinates of 4 points: the genesis location, a point at 1/3 along the track, a point at 2/3 along the track, and the termination location of the TC, see Fig. 4a. The total duration of the TC was used directly. For both descriptors  $v_{max}$  and  $c$ , the maximum, median and minimum values along the track were used (parameters  $v_{max,min}$ ,  $v_{max,median}$ ,  $v_{max,max}$  and  $c_{min}$ ,  $c_{median}$ ,  $c_{max}$ , respectively). The track location relative to the SL was parameterized by use of the minimum Distance of the Track to SL (DTSL) and the Azimuth of the line of the minimum distance of the Track to SL (ATSL), whereby effectively the polar coordinates of the point on the track at minimum distance were used with SL as origin (Fig. 4a). To focus the analysis, a domain of interest for the selection was specified, and only the parts of tracks that fell within this region were included (Fig. 4a).

### 3.3.2. TC parameter space modifications

In total, 17 parameters were used for 5 descriptors (when the relative location of the TC track is included) (Table 1). If the Euclidean distances in this 17-parameter space were to be calculated directly, relatively more weight would go to the location of the track (described by 8 parameters) compared to, for example, the duration (1 parameter), because it is described by more parameters. The same applies to  $v_{max}$  (3 parameters),  $c$  (3 parameters) and the relative track location to SL (2

parameters). Consequently, physical TC characteristics described by only one or a few parameters (e.g., the duration) would not be well considered in the selection of RTCs. Furthermore, not all parameters are equally important for estimating the hydro-meteorological conditions resulting from the TC. For example, the distance of a TC track to a location (represented by DTSL) seems more strongly related to the hydro-meteorological conditions at that location than the total duration of the TC (see Fig. 5). In addition, the hydro-meteorological conditions are not necessarily linearly related to the value of a parameter (e.g.,  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  are not linear with DTSL parameter).

To compensate for this kind of unrepresentative weighting, the parameter space can be modified in several ways: (1) the parameter space can be constrained, in our case by including only tracks within a certain area (i.e., within the domain of interest); (2) parameters can be scaled, whereby more weight can be given to TCs with high  $v_{max}$  compared to TCs with low  $v_{max}$ ; and (3) weight factors can be applied to the parameters, whereby the relative importance of (groups of) parameters can be adjusted. The effect of such modifications is illustrated in Fig. 4b–e, where different sets of 9 representative samples were selected out of 200 samples (Fig. 4b), described by 2 parameters (samples generated for illustrational purposes only). In Fig. 4c, no modifications to the parameter space were applied. The samples were normalized, where after representative samples were selected by maximizing the Euclidean distance in normalized parameter space; numbering corresponds to the sequence of selection. Finally, the selected representative samples are shown in original parameter space. Fig. 4d illustrates how weight factors affect the normalized parameter space and the selection of representative samples. Similarly, Fig. 4e illustrates the effect of scaling.

In this study, normalized parameters were corrected for the number of parameters per descriptor ( $N_{par}/des$ ), weighted by multiplication with a factor ( $F_w$ ) and subsequently scaled by raising to the power with a factor ( $F_s$ ), according to eq. (1). Applying a higher (lower) value of  $F_w$  to a parameter makes it more (less) important relative to others. Applying a  $F_s > 1$  ( $F_s < 1$ ) makes the higher (lower) values of that parameter relatively more important than its lower (higher) values.

$$\left[ \frac{F_w}{N_{par}/des} * (norm(parameter)) \right]^{F_s} \quad (1)$$

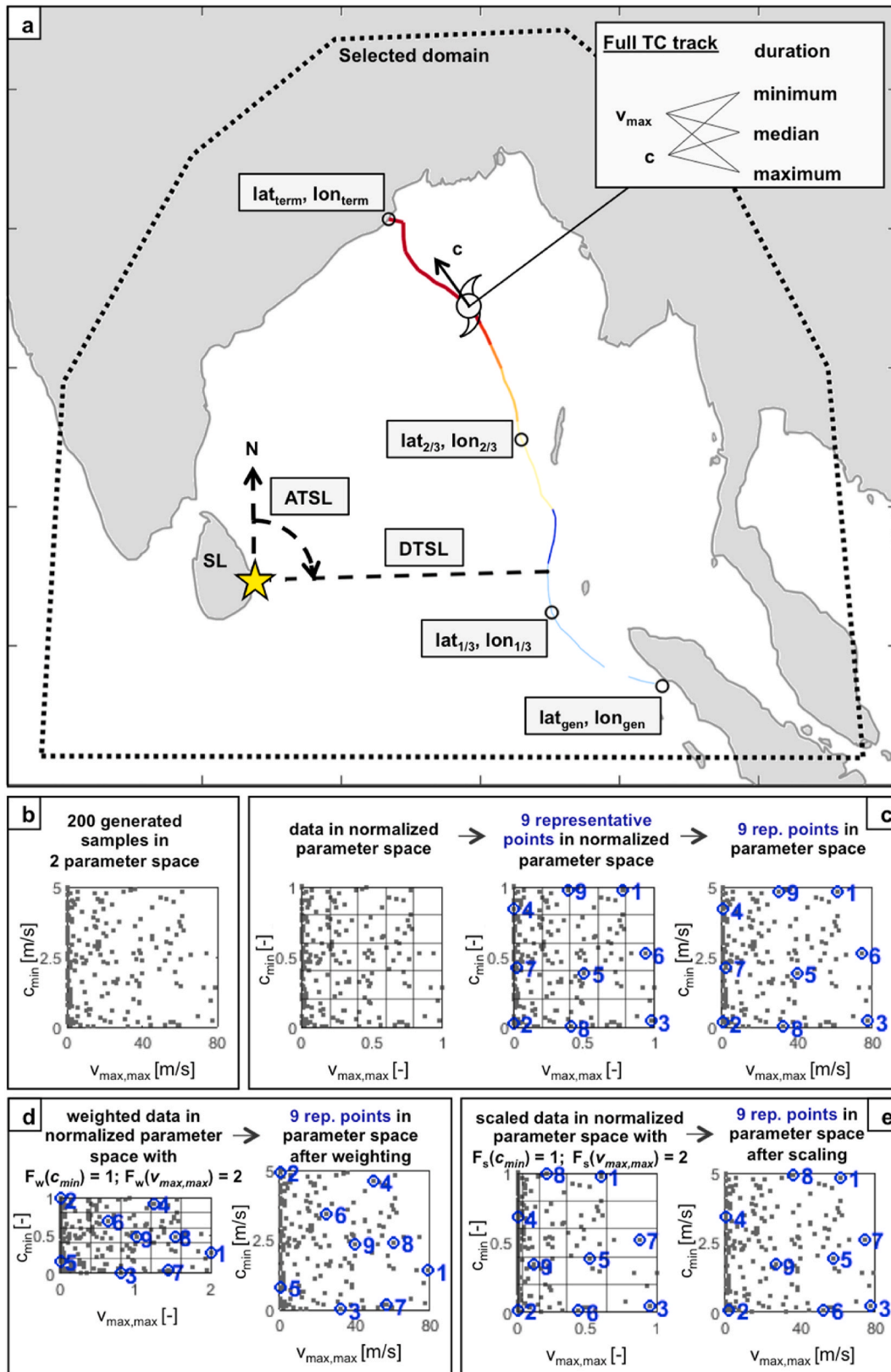
### 3.3.3. Parameter space configurations

To provide some insights into the effect of the definition of the parameter space on track selection and performance of the method, three configurations of the method were tested:

1. Full Basin without modified parameter space (except for correction for number of parameters per descriptor) (FB),
2. Full Basin with Modified Parameter Space (FBMPs), and
3. Selected Location focused with Modified Parameter Space (SLMPs).

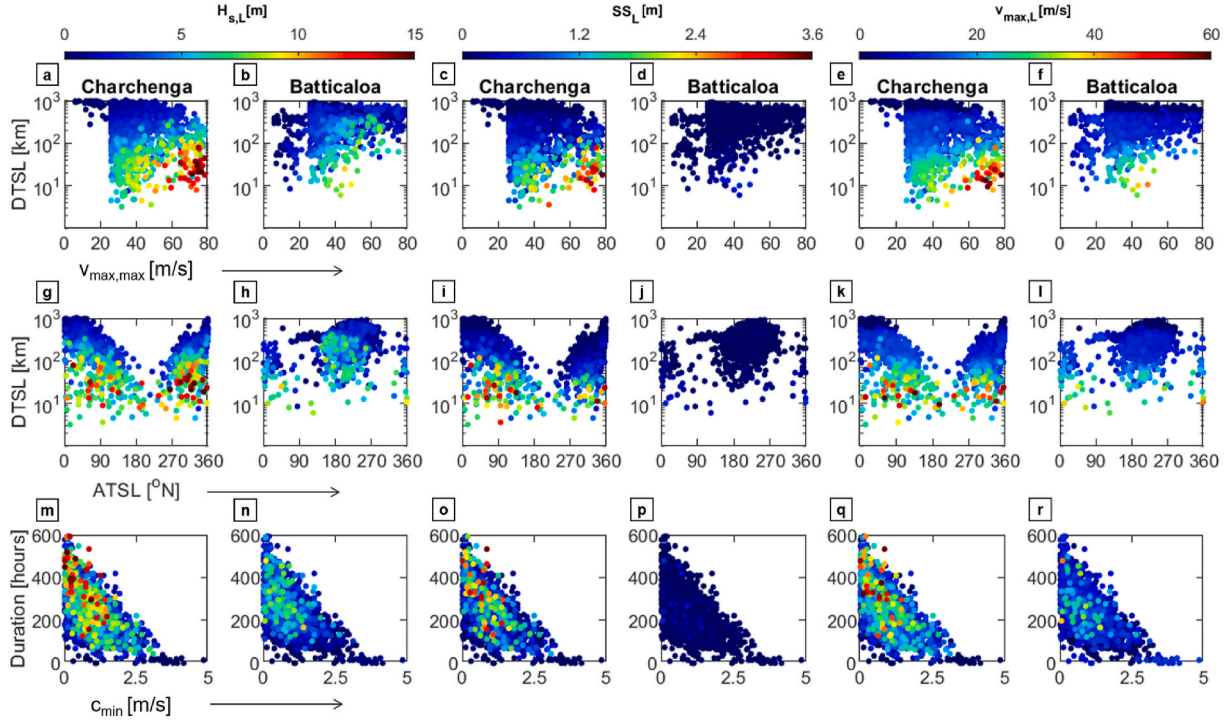
The FB and FBMPs configurations are for the full basin and the relative location of the TC track is not used, whereas in SLMPs the selection is specific to one location (i.e., SL). Full basin configurations are more suitable for large-scale studies ( $\approx 1000$  km) in which there is no specific accuracy requirement for a particular location. Configurations with focus at one location are best suited to more local studies ( $\approx 10$ – $100$  km), where higher accuracy requirements are generally needed for a specific area. An overview of the applied weight and scaling factors applied to each of the configurations is provided in Table 1.

To optimally modify the parameter space (FBMPs and SLMPs configurations), the relation between (combinations of) parameters used to describe the TC and the hydro-meteorological conditions resulting from that TC was analysed qualitatively. For each STC, values of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  at Charchenga and Batticaloa locations are plotted for different combinations of parameters (Fig. 5). Only parameter combinations  $v_{max,max}$  and DTSL (Fig. 5a–f), ATSL and DTSL (Fig. 5g–i), and  $c_{min}$  and



**Fig. 4.** Schematic overview of the parameterization of a TC track and modifications of the parameter space. (a) Example parameterization of a TC track over the Bay of Bengal. TC track color represents the maximum windspeed (warmer colors for higher windspeed). The yellow star indicates the Selected Location (SL). (b–e) Schematic overview of the selection of a subset of 9 representative samples (blue) out of (b) 200 samples (black, generated for illustrational purposes only) in 2 parameter space ( $C_{min}$ ,  $V_{max,max}$ ). (c) Selection of 9 representative samples; numbering corresponds to the sequence of selection. (d) Illustrates the effect of weight factors on the normalized parameter space and selection of representative samples; locations of the grid lines correspond with the grid lines as in (c) to illustrate the effect of the modification. (e) Illustrates the effect of scaling, similarly to (d).





**Fig. 5.** Simulated hydro-meteorological conditions versus different combinations of parameters for the STC. Hydro-meteorological conditions (colorscale) are quantified by the local maximum significant wave height ( $H_{s,L}$ ), storm surge ( $SS_L$ ), and wind speed ( $V_{max,L}$ ) for each TC offshore at Charchenga and Batticaloa locations, and plotted versus different combinations of parameters, used to parameterize the TC. This information is used to guide decisions about how to optimize the weighting and scaling of the parameter space. Parameters are explained in Fig. 4a. Note the logarithmic scale for the DTSL (Distance Track to Selected Location) parameter (a–l).

duration (Fig. 5m–r) are shown for simplicity. The parameters used for the descriptors maximum windspeed (only  $v_{max,max}$  shown), and relative location of the TC track (only  $DTSL$  shown) were found to be good predictors of the hydro-meteorological conditions (Fig. 5a–f). As expected, most extreme hydro-meteorological conditions are caused by TCs with high windspeeds and tracks with smaller distance to the studied location. For the translation speed (only  $c_{min}$  shown), duration and  $ATSL$  descriptors, the relation with the hydro-meteorological conditions is less clear.

The maximum windspeed (only  $v_{max,max}$  shown) of the TC and minimum distance ( $DTSL$ ) are thus better predictors of the  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  compared to the duration, translation speed (only  $c_{min}$  shown) and  $ATSL$  descriptors (Fig. 5). For the track location descriptor, many parameters (8) are required, and even qualitative assessment of the parameters is challenging. However, a proper description of the TC track is essential to quantifying the variability in TC tracks – especially when tracks for the full basin are selected and  $ATSL$  and  $DTSL$  parameters are not included. Hence, the maximum windspeed and minimum distance descriptors were weighted most strongly ( $F_w$  of 3 applied), followed by track location ( $F_w$  of 2 applied), while the duration, translation speed and  $ATSL$  have the least influence ( $F_w$  of 1 applied). Furthermore, following from Fig. 5a–f  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  were highest for TCs with high windspeed (represented by  $v_{max,max}$ ) passing at small distance ( $DTSL$ ). To ensure that the most extreme TCs were well represented by the RTCs, scaling was thus applied in such way that the Euclidean distance between extreme TCs was enlarged. Once the parameter space configuration was chosen, the set of RTCs could be selected.

### 3.4. Simulated TC hydro-meteorological conditions

TC hydro-meteorological conditions for each of the HTC and STC are taken from Leijnse et al. (2022). Offshore wave conditions and nearshore storm surge associated with each TC were dynamically

simulated by coupling a hydrodynamic model (Delft3D-FM, Kernkamp et al., 2011) with a wave model (SWAN, Booij et al., 1999). Spatio-temporal wind and pressure fields, generated with the TCWiSE tool for each of the TCs, were used as input for the model. The water level was set to mean sea level, with no tidal variation applied. For the wave model a constant resolution of  $\sim 2$  km was used, while for the hydrodynamic model a varying resolution of up to  $\sim 3$  km nearshore was used. Wave contributions to storm surge and the storm surge at the coast cannot be resolved at these scales. Bathymetric data for the Bay of Bengal were derived from the GEBCO 2008 global bathymetric data set (Becker et al., 2009). Despite the relatively coarse model resolution, the model runtime, which depends on the TC track, was estimated at about 1 h per TC. Details of the model setup can be found in Appendix A of Leijnse et al. (2022). Appendix A

The model produced time series of storm surge and significant wave height at output points along the coastline including the 9 locations used in this study. Storm surge conditions were obtained at output points along the shoreline of the BoB, while significant wave heights were obtained at the nearest point with a water depth of at least 30 m, so as not to be influenced by the local bathymetric conditions. From the time series of significant wave height and storm surge, maximum values of the significant wave height ( $H_{s,L}$ ) and storm surge ( $SS_L$ ) per TC were obtained. Maximum sustained windspeeds ( $v_{max,L}$ ) per TC for these locations were obtained directly from the wind fields as generated with TCWiSE.

### 3.5. STCrecs - reconstruction of STC hydro-meteorological conditions

Unlike the STC events, the RTC events and associated hydro-meteorological conditions do not have an equal probability of occurrence (i.e., they are selected from the STCs with emphasis on the extremes). Hence, to allow estimates of return periods, we reconstructed the hydro-meteorological conditions for each STC (STCrecs) by inter-



polation from the RTCs. For a given STC, we assigned a weight to each RTC based on the match between their track characteristics. These weights were determined by use of the Softmax function: a relatively simple, one parameter, non-linear interpolation function (Goodfellow et al., 2016), see eq. (2).

$$S(x)_i = \frac{\exp(-\beta * d_i)}{\sum_{j=1}^{N_{RTC}} \exp(-\beta * d_j)} \quad (2)$$

In our application,  $S(x)$  is the weight given to RTC  $i$  to approximate the STC, with  $d_i$  as the (Euclidean) distance between the STC and RTC  $i$ , and  $d_j$  as the (Euclidean) distance between the RTC  $j$  and approximated STC.  $\beta$  is a stiffness parameter and  $N_{RTC}$  is the number of representative TC scenarios. The implementation of eq. (2) by Scott et al. (2020) was used. For  $d_i$  and  $d_j$ , the Euclidean distances calculated in the parameter space between the scenarios were used – the same as were used in MDA to select the set of representative scenarios (Section 3.3).

For each STC, Softmax returned a weight ranging between 0 and 1 for each of the RTCs, resulting in a matrix with weights of size  $N_{STC}$  by  $N_{RTC}$  (the numbers of STCs and RTCs respectively). Hydro-meteorological conditions for each STC were thus approximated independently. The sum of the weights of all RTC to approximate one STC is 1 (by definition). Weights below 0.001 were redistributed to limit the number of RTCs used per STC. By using these weight factors, the hydro-meteorological conditions for each of the STCs were approximated based on the RTCs, resulting in the STCrecs.

### 3.6. Evaluation of performance of the method by EVA

To evaluate performance of the method, we compared the hydro-meteorological conditions resulting from model simulations for each of the STCs with their reconstructed counterpart in the STCrecs. In addition, it was possible to estimate the probability of occurrence for each condition at a specific location (e.g., there is an estimated 1/100-year probability of 12.85 m waves offshore at Charchenga (Table 2)). To estimate this, an EVA was carried out over the values of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$ . A peak over threshold approach was used in combination with fitting of the exponential distribution. For HTC (n = 81), the 80th percentile was used, while for the STCs and STCrecs (n = 1745) the 98th percentile values were used. The method of probability weighted moments was used for estimating the distribution parameters and 95% confidence bands were determined by bootstrapping (1000 samples) (Caires, 2016). The output from this step provides the information necessary to compute the hazards associated with TCs, like flooding.

## 4. Results

### 4.1. Comparison of RTC selection configurations

#### 4.1.1. Spatial comparison of selected RTCs

To assess the spatial distribution of the RTC tracks, we compare the first 81 RTCs selected according to each of the configurations (FB, FBMPs and SLMPs, described in Section 3.3.3) (Fig. 6). In Fig. 6a the selected RTC tracks are presented for the FB configuration, which considers the full basin without modifications to the parameter space. When compared with the same number (81) of HTCs (Fig. 3a), the selected RTC tracks for the FB configuration are slightly more uniformly distributed over the BoB area (Fig. 6a). More importantly, the selected RTCs generally have higher windspeeds compared to the HTCs, whereby a wider range of extreme events is included on the scale of the full basin.

Next, we compare the tracks of the FB configuration with those of the FBMPs configuration, which is modified to prioritize more extreme TCs (Fig. 6b). In the FBMPs configuration, the (extremes of) maximum windspeed ( $v_{max}$ ) descriptor were given a higher weight. Comparison with Fig. 6a indeed shows that these modifications result in the selection of more RTCs with a very high windspeed. This seems advantageous for assessment of hazards (e.g., coastal flooding), as it is typically the extremes that are of interest. However, the downside of this is that the STC tracks at lower latitudes (e.g., near Batticaloa) are less well represented, as TCs in these regions typically have lower maximum windspeeds.

We can further refine our selection of RTCs by focusing on TCs that affect a selected location (SL). When RTC are chosen according to the SLMPs configuration, the relative location of the track to the SL is included and the parameter space modified to prioritize extreme events. The effect of this can be seen in Fig. 6c and d, where Charchenga and Batticaloa were used as SL respectively. As expected, the selected RTC tracks are located closer to their SL compared to the RTCs selected for the full basin (Fig. 6b). Furthermore, the RTC windspeeds around a given SL vary more compared to those derived with the FBMPs configuration (Fig. 6b). This effect is especially visible in Fig. 6c, where more RTCs with lower windspeeds were selected around Charchenga compared to Fig. 6b. In the selected RTCs for Batticaloa (Fig. 6d) this effect is also visible, but due to the relatively high weight given to the (extremes of) maximum windspeed descriptor, some extreme tracks at larger distance from Batticaloa were also selected. While the RTC thus better cover the area of the SL, the selection is not optimal for locations other than the SL. For example, selected RTCs of Fig. 6c with Charchenga as SL are not representative for STCs around Batticaloa, as they cover the area very poorly.

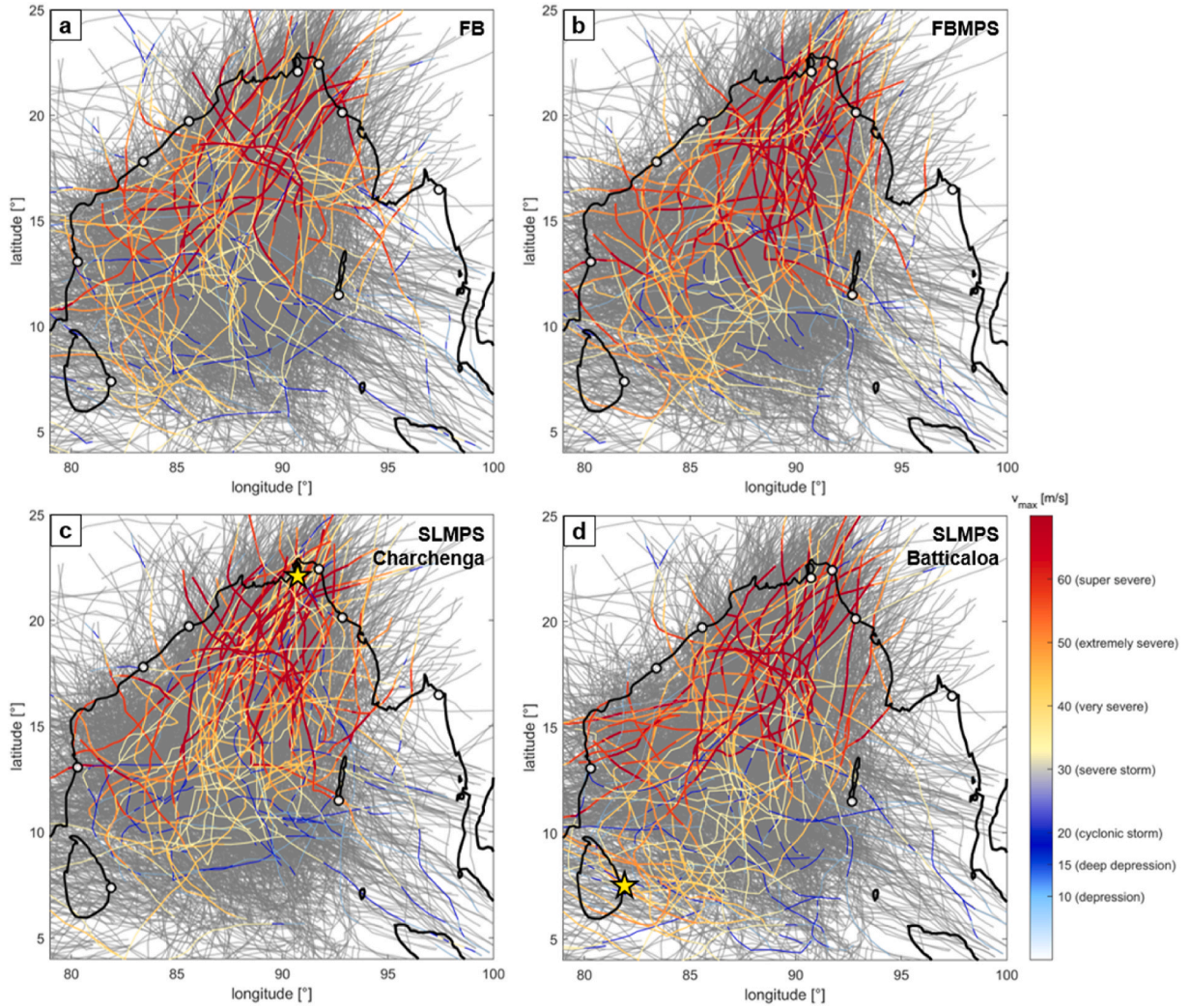
#### 4.1.2. Comparison of RTC hydro-meteorological conditions

Since the overall aim is to identify an optimal set of TCs, which can

**Table 2**

Performance of configuration SLMPs for all studied locations. Values of STCrec based on 175 RTCs ( $\pm 10\%$  of the STC dataset) are compared with values based on the full set of 1745 STCs. Presented are approximated 100-year RP values based on STCs and based on STCrecs (estimated by empirical cumulative distribution functions), Spearman rank correlation coefficient ( $\rho$ ), Relative Root Mean Square Error (RRMSE), and Relative Bias (RB), for significant wave height ( $H_{s,L}$ ), maximum storm surge (SSL) and maximum wind speed ( $V_{max,L}$ ), at each of the studied locations.

Location	$H_{s,L}$					$SS_L$					$V_{max,L}$				
	STC [m]	STCrec [m]	$\rho$ [–]	RRMSE [%]	RB [%]	STC [m]	STCrec [m]	$\rho$ [–]	RRMSE [%]	RB [%]	STC [m/s]	STCrec [m/s]	$\rho$ [–]	RRMSE [%]	RB [%]
Charchenga	13.74	12.85	0.94	7.57	0.91	2.79	2.68	0.94	6.40	0.44	48.5	44.2	0.94	5.84	0.26
Batticaloa	8.26	7.49	0.87	9.43	0.58	0.26	0.26	0.71	5.26	–0.14	27.8	26.9	0.91	6.12	–0.63
Madras	10.44	9.55	0.91	7.90	1.09	0.51	0.39	0.89	7.04	0.83	38.1	35.3	0.94	6.49	0.42
Visakhapatnam	9.19	8.09	0.92	6.47	–0.15	0.36	0.35	0.91	5.59	–0.16	30.3	29.0	0.91	6.56	–0.31
Puri	10.10	9.55	0.89	7.05	0.17	0.49	0.50	0.86	6.97	–0.27	36.9	35.3	0.90	6.33	–0.32
Chittagong	11.93	10.96	0.95	7.99	–0.37	3.46	3.12	0.95	7.15	–0.23	47.0	47.6	0.94	6.44	–0.04
Sittwe	12.58	11.17	0.95	5.55	0.11	1.11	0.93	0.95	5.67	–0.20	42.8	39.3	0.95	5.18	–0.75
Mawlamyine	4.33	3.86	0.72	7.13	–0.98	0.70	0.61	0.86	7.92	0.11	14.5	13.4	0.88	13.58	–1.61
Port Blair	6.40	5.04	0.82	8.94	0.28	0.13	0.11	0.85	6.89	–0.34	26.3	21.7	0.86	7.62	1.23



**Fig. 6.** Distribution of the selected RTCs (in colorscale) over the BoB for different configurations of the selection method as presented in Section 3.3.3. The same number of RTCs (81) as the number of HTCs in the data set (Fig. 3a) are presented. RTCs are selected as a subset of the STCs (grey). (a) Full basin without modified parameter space (configuration FB), (b) Full basin with modified parameter space (configuration FBMPS), (c) Focused on one location (SL) with modified parameter space (configuration SLMPS) for Charchenga, and (d) SLMPS for Batticaloa. SL is indicated by yellow star; the colorscale of the RTC tracks is a measure for the maximum wind speed ( $V_{max}$ ).

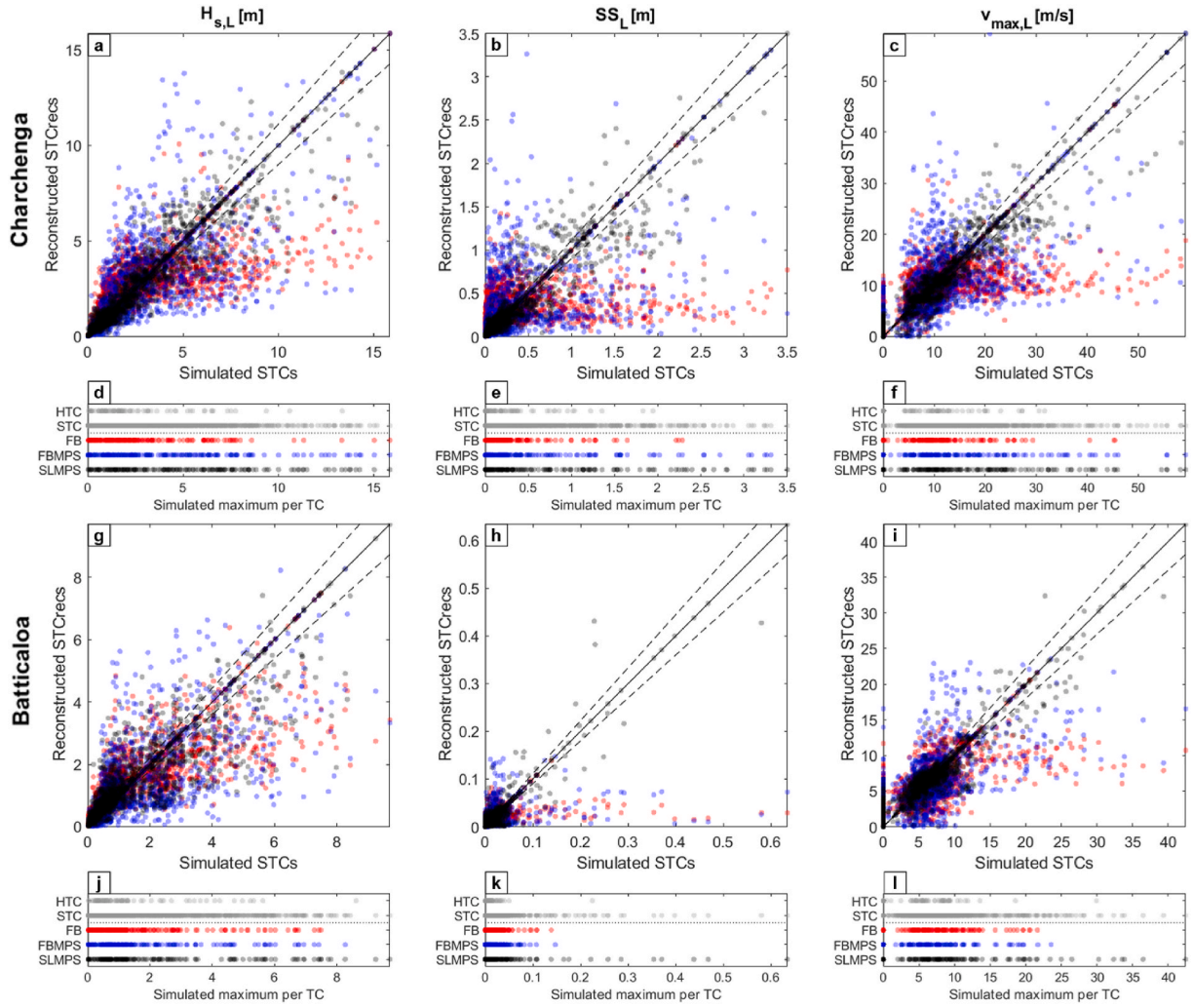
best represent the hydro-meteorological conditions driving coastal hazards at a certain region, here we compare how well each RTC selection configuration can estimate the hydro-meteorological conditions of the full set of STCs. To test this, the values  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  for each STC (based on model simulations) were compared with their STCrec counterpart (based on model simulations for a smaller subset of RTC and interpolation). Only results for Charchenga and Batticaloa locations are presented in full (Fig. 7). For this, 175 RTCs were selected according to each of the configurations. This roughly corresponds to simulating 10% of the initial STCs or, in other words, corresponds to a 90% reduction in computational costs for simulating all STCs. The Softmax function was used for the interpolation, as described in Section 3.5. A stiffness parameter  $\beta$  of 175 ( $= N_{RTCs}$ ) was used for the configurations focused on the full basin (FB and FBMPS), while  $\beta$  of 35 ( $= N_{RTCs}/5$ ) was used in the site-specific SLMPS configuration. The SLMPS configuration selects more extreme RTCs compared to the configurations for the full basin. Therefore, the Euclidean distances between the extreme STCs and the RTCs are smaller for the SLMPS configuration (compared to the full basin configurations), and a lower  $\beta$  value is required for the SLMPS configuration. This principle is further elaborated on in Section 4.2.2, where the effect of the number of RTCs on the optimal  $\beta$  value is discussed (which has a similar effect on the Euclidean distances between

STCs and RTCs).

The aim is to obtain the best possible match between the simulated  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  for each of the STCs and their approximate STCrec counterpart. When a perfect match is obtained, the marker is located on the diagonal in Fig. 7a–c, g–i; 10% error margins of the reconstructed hydro-meteorological conditions are shown as well (dashed lines). Furthermore, the distribution of the selected RTCs over  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  is presented (Fig. 7d–f, j–l) for each of the configurations. To approximate  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  for each of the STCs, the selected RTCs should preferably cover the full range of those parameters as well (at least when the STCs approximately cover the full range as well, which is the case here). Furthermore, the STC(s) causing the most extreme values of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  (or at least one close to this one) should preferably be selected as RTC(s) (one or multiple as the STC associated with the most extreme values of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  respectively could be different). Selection of the extremes is important, as hydro-meteorological conditions for STCs that are not selected as RTC can only be derived by interpolation and not through extrapolation.

First, we compare the FB and FBMPS configurations, which both aim to select RTCs for the full basin (red and blue markers respectively in Fig. 7). Parameter space modifications of the FBMPS configuration, aimed at selecting more extreme TCs, indeed leads to the selection of





**Fig. 7.** Performance of the STCrecs in estimating the maximum STC-induced hydro-meteorological conditions (significant wave height ( $H_{s,L}$ ), storm surge ( $SSL$ ) and wind speeds ( $V_{max,L}$ )) for three different configurations of the selection method, both for Charchenga (a–f) and Batticaloa (g–l) locations. Configurations in order of increasing performance: Full Basin (FB, red) without modifications to parameter space, Full Basin with Modified Parameter Space (FBMPs, blue), and Selected Location focused with Modified Parameter Space (SLMPs, black). (a–c, g–i) hydro-meteorological conditions based on simulation of all 1745 STC (x-axis), versus reconstructed hydro-meteorological conditions of the STCrecs (based on simulation of 175 RTCs and interpolation) (y-axis). When located on the diagonal (black line) they form a perfect match; dashed black lines indicate 10% error margins of the reconstructed hydro-meteorological conditions. (d–f, j–l) Distributions of the different sets of 175 selected RTCs (selected according to configurations FB, FBMPs and SLMPs) over values of  $H_{s,L}$ ,  $SSL$  and  $V_{max,L}$ , corresponding to the panel above. Distributions of 81 HTCs and the full set of 1745 STCs over  $H_{s,L}$ ,  $SSL$  and  $V_{max,L}$  are shown as well.

more relatively extreme RTCs compared to the FB configuration without modifications. This is the case for both Charchenga and Batticaloa locations. Furthermore, this is in agreement with Fig. 6a and b. However, the set of RTCs improved more for Charchenga than Batticaloa; in particular, the effect on the approximation of  $SSL$  (Fig. 7k) at Batticaloa is negligible. Neither configuration FB nor FBMPs reconstructed all STCrecs well with 175 RTCs. Especially TCs causing the most extreme  $SSL$  at Batticaloa are not well captured, but extreme  $H_{s,L}$  and  $V_{max,L}$  are not well represented, either. Extreme values of  $SSL$  at Batticaloa are significantly lower than at Charchenga and the distribution of the  $SSL$ , both for HTCs and STCs, is different (Fig. 7e–k). This is because of the typically direct hit by fully developed TCs at Charchenga while at Batticaloa TCs typically pass at larger distance, and the wider continental shelf at Charchenga (Fig. 2) which allows for higher surge levels.

RTCs derived using the site-specific SLMPs configuration (black markers) are typically more uniformly distributed over  $H_{s,L}$ ,  $SSL$ , and  $V_{max,L}$ , and their extremes are better covered than RTCs generated using the larger-scale FB and FBMPs configurations (Fig. 7d–f, j–l). In addition, the reconstructed maximum values of  $H_{s,L}$ ,  $SSL$ , and  $V_{max,L}$  for each of the

STCs are in better agreement with their simulated values (i.e., in Fig. 7a–c and g–i the black markers are typically closer to the diagonal compared to the red and blue markers of configurations FB and FBMPs respectively). The improvement is especially considerable for Batticaloa location.

In Fig. 7c and i some STCs have zero maximum windspeeds ( $V_{max,L} = 0$  m/s), while their approximated values (based on RTC) are higher. This is caused by simplifications in the windfields as part of the dataset used for this study and derived from TCWise (Nederhoff et al., 2021), and the interpolation method. In particular, wind fields were available for a circular area with a radius of 900 km around the TC eye, while outside this area they were set to 0 m/s. Hence, for STC passing at more than 900 km, simulated  $V_{max,L}$  was 0 m/s, while interpolated values from RTCs may be larger than 0 m/s due to the interpolation procedure. Windspeeds for TC passing at such distances are low and cannot be attributed to the TC (e.g., see Nederhoff et al., 2019).

In view of the better spatial coverage (Section 4.1.1) and representativeness of the extreme cases at specific locations of interest, we can conclude that the site-specific SLMPs configuration performs better than

the full basin-based FB and FBMPs configurations, at least for specific locations. We thus adopt the SLMPs configuration for further analysis in Section 4.2.

#### 4.2. Application of SL configuration

##### 4.2.1. Performance evaluation in representing extreme event conditions

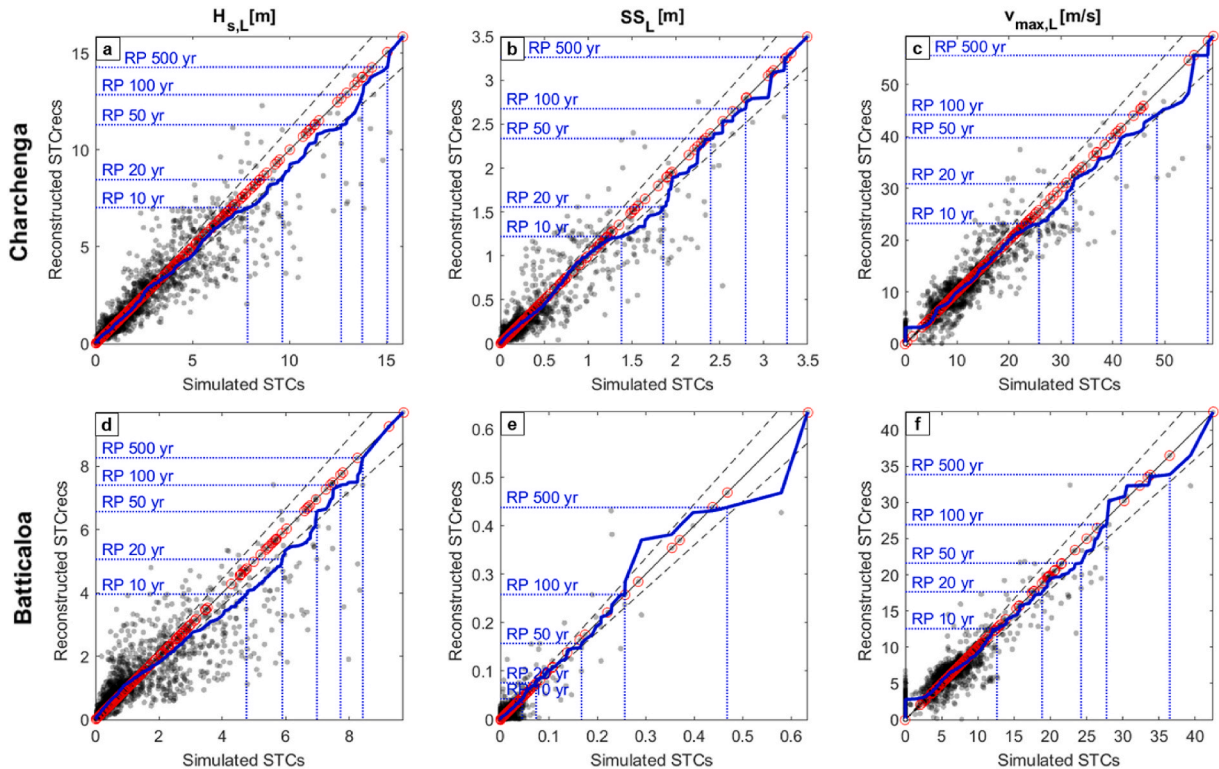
Hydro-meteorological conditions during extreme events are important for defining coastal hazards and risks as well as for the design of coastal protection interventions. In this section, we evaluate how well the RTCs selected with configuration SLMPs represent extreme event conditions for the following parameters:  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$ .

For Charchenga and Batticaloa locations,  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  for selected RTCs are indicated by red markers in Fig. 8; these are located on the diagonal as derived by model simulations. For both Charchenga and Batticaloa locations, the full range of values is well covered by the RTCs, and the STC track(s) that caused the most extreme hydro-meteorological conditions was selected as RTC. In fact, this is also the case for all other studied locations around the BoB (see Supplementary material A) – except for the Sittwe location (see Fig. 2 for this location) where the STC track that caused the second most extreme hydro-meteorological conditions was selected as RTC. Mawlamyine, the studied location closest to the area where Cyclone Nargis made landfall (discussed in Section 1) is relatively sheltered from the waves caused by TCs passing at large distance. Many of the simulated STCs do not affect this location at all, and offshore values for 100-year RP are among the lowest of the studied locations (Table 2). However, the most extreme simulated hydro-meteorological conditions (corresponding to an estimated 1000-year RP) are  $H_{s,L} = 7.98$  m,  $SS_L = 2.14$  m, and  $v_{max,L} = 31.9$  m/s, based either on all STC or 175 RTC. The selected RTCs thus span the full range

of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  and extreme and infrequent STCs were also captured.

Subsequently, we evaluate the performance of our method for all studied locations around the BoB using error metrics. For both the sets of STCs and STCrecs, and for each of the studied locations, estimated 100-year RP values of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  are presented (Table 2). These were estimated based on the Empirical Cumulative Distribution Function (ECDF). Values of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  for the STCs are simulated values, values of the STCrecs were reconstructed based on 175 simulated RTCs and interpolation ( $\beta = 35$ ). In addition, Spearman rank correlation coefficient ( $\rho$ ), Relative Root Mean Square Error (RRMSE), and Relative Bias (RB) are presented. First, the Root Mean Square Error (RMSE) and bias values were calculated based on the values of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  for STCs and STCrecs; these values can also be found in Supplementary material A. Subsequently, RRMSE and RB values were obtained by dividing the RMSE and bias values (of the full set of STCs and STCrecs) by their 100-year RP value based on the ECDF of the STCs (Table 2). The presented RRMSE and RB are thus a percentage of the estimated 100-year RP values of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  at that location, and as such allow easy comparison for the different locations.

For values for  $\rho$  of 1, RRMSE of 0, and RB of 0, estimated (STC) and reconstructed (STCrecs) values of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  form a perfect match. This is found by definition when all 1745 STCs in the dataset are also selected as RTC. For values of RB larger (smaller) than 0, the estimated  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  of STCrecs is larger (smaller) than the simulated  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  respectively. Values of  $\rho$  (mostly  $>0.9$ ) show good agreement for  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  at most locations, but with some deviations (e.g., for  $SS_L$  at Batticaloa). RRMSE is below 10% for all studied locations, except for  $v_{max,L}$  at Mawlamyine (see Supplementary material A) where the windspeeds for part of the TCs passing at large



**Fig. 8.** Performance of the RTCs selected with configuration SLMPs (Selected Location with Modified Parameter Space) and reconstructed STCrecs in estimating the STC-induced hydro-meteorological conditions (significant wave height ( $H_{s,L}$ ), storm surge ( $SS_L$ ) and wind speeds ( $V_{max,L}$ )). Presented are results for STCs (x-axis, based on simulation of 1745 STC) versus results for STCrecs (based on simulation of 175 RTC and interpolation) (black markers, with opacity), for Charchenga (a–c) and Batticaloa (d–f) locations. When located on the diagonal (solid black line) they form a perfect match; dashed black lines indicate 10% error margins of the reconstructed hydro-meteorological conditions. Selected RTCs (red) are indicated; being a simulated subset of the STCs they always form a perfect match. Empirical cumulative distribution function of the full set of STCs is plotted against the one based on the STCrecs (blue). Estimated values for different Return Periods (RP) are indicated.



distance were overestimated (see Section 4.1.2). The Absolute (value of the) Relative Bias (ARB) of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  is below 2% for all locations.

For some individual STCrecs, the approximated values of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  deviate considerably from their STC counterpart. To assess the consequences of these deviations and performance after EVA, ECDFs for the simulated STC were plotted against the ECDF for the reconstructed STCrecs (Fig. 8, in blue). Quantiles for different RPs ranging 10–500 years are indicated as well, to mark the range of quantiles that is typically of interest in TC hazard assessments. Comparison of the ECDF based on RTCs with the ECDF based on STCs shows that the quantiles of interest are typically estimated within 10% error margins (dashed lines). Even though  $\rho$  for  $SS_L$  at Batticaloa was found to be the lowest value of all studied locations (0.71; Table 2), the quantiles corresponding to 10–500-year RPs could be well approximated.

When looking at values of  $\rho$  in Table 2, typically the SLMPS configuration performs better for locations where TCs that directly make landfall lead to most extreme hydro-meteorological conditions (i.e., Charchenga, Visakhapatnam, Sittwe, Chittagong), compared to locations where TCs passing at larger distance can also lead to relatively extreme hydro-meteorological conditions (i.e., Batticaloa, Mawlamyine, Port Blair). This can be expected given the focus in the SLMPS configuration on one location.

#### 4.2.2. Optimization of RTC selection and interpolation

To further improve the efficiency and accuracy of our method, we explored the number of RTCs and interpolation parameter values necessary to make good predictions. Performance of configuration SLMPS for different numbers of RTCs and stiffness parameter  $\beta$  in the Softmax interpolation method was tested for Charchenga and Batticaloa locations (Fig. 9). For combinations of the number of RTCs and  $\beta$ , the hydro-meteorological conditions (i.e.,  $H_{s,L}$ ,  $SS_L$  and  $v_{max,L}$ ) of the STCs were reconstructed and compared with their simulated values. Values of  $\rho$ , the RRMSE and ARB are presented (colorscale) in Fig. 9. The presented RRMSE and ARB are a percentage of their estimated 100-year RP values at that location based on STC (Table 2).

As expected, the STCs can be better approximated when more RTCs are used. The required number of RTCs depends on the required level of accuracy. For each value of  $N_{RTC}$ , the optimal  $\beta$  value is indicated by the black line in Fig. 9 (i.e.,  $\beta$  value the corresponding to the highest value of  $\rho$ , lowest values of RRMSE, and lowest value of the ARB). When more RTCs are used for a site, a larger value of stiffness parameter ( $\beta$ ) is required, since on average for each STC, closer RTCs are available (i.e., with smaller Euclidean distance in parameter space). Hence, with more RTCs available it is more favourable to give the closest RTCs more weight instead of taking values closer to the average of all RTCs. This is especially the case when approximating the hydro-meteorological conditions for the most extreme STCs; even though the focus of the SLMPS configuration was on selecting the extreme STC, there were many more mild rather than extreme RTCs selected. Hence with constant  $\beta$ , using more RTCs thus generally leads to giving more weight to milder RTCs, and thereby underestimating the extreme STCs.

For the SLMPS configuration, a value of  $N_{RTC}/\beta$  of about 10 was estimated as optimal based on  $\rho$ , RRMSE, and ARB values (Fig. 9). However, the calculated values of  $\rho$ , RRMSE, and RB in Table 2 and Fig. 9 are based on hydro-meteorological conditions for all STCs and STCrecs in the BoB. This means that for the calculation of  $\rho$ , RRMSE, and RB, no distinction is made between small deviations in the many STCrecs with relatively mild conditions on the one hand, and large deviations in the few extreme STCrecs on the other hand. For hazard assessment, large deviations in the approximation of the few extreme STCrecs are generally more important. Due to the mechanisms described above, a lower value of  $N_{RTC}/\beta$  performed better for approximating the conditions for extreme TCs and these events are of the most interest. In practice, a slightly lower value thus seems advisable and in this study a value of  $N_{RTC}/\beta$  of 5 was used. Fig. 9 can thus be useful to estimate the required

values of  $N_{RTC}$  and  $\beta$ , but in applications the RTC selection and  $\beta$  should be checked by comparison as in Figs. 7 and 8 (e.g., by comparing  $v_{max,L}$ , see Section 5.3).

#### 4.2.3. Evaluation by EVA of hydro-meteorological conditions

Having optimized our parameters for the RTC selection and interpolation, we can now compute the extreme values of hydro-meteorological variables as a function of return period. An EVA as described in Section 3.6 was performed on the maximum values of  $H_{s,L}$ ,  $SS_L$  and  $v_{max,L}$  for HTC, simulated STCs, and STCrecs (as interpolated from RTCs). Only results for Charchenga and Batticaloa locations are presented (Fig. 10). As expected, comparison of STCs and RTCs shows stronger agreement for larger numbers of RTCs (81, 175, and 350 RTCs shown). With the STCs as benchmark, fits based on 81 RTCs typically perform better than fits based on 81 HTCs;  $H_{s,L}$  for Charchenga is the only exception (Fig. 10).

RTCs are advantageous, especially for approximating higher return period (>50 years) quantiles. Such estimates based on HTCs and STCs deviate considerably at some locations, especially with infrequent TCs (e.g., for Batticaloa). For example, the maximum simulated  $SS_L$  of only 1 out of 81 HTC tracks for the BoB is > 0.1 m at Batticaloa (Fig. 10e), while this is the case for 7 out of the 81 selected RTC tracks. Reconstruction based on 81 RTCs results in 38 out of the 1745 STCrecs with  $SS_L > 0.1$  m. When all 1745 STCs are simulated, also 38 out of the 1745 STCs are found to have  $SS_L > 0.1$  m. Consequences for an EVA can be clearly observed by comparing the fitted distributions for 81 HTCs, for STCrecs based on 81 RTCs, and for all STCs: with the STCs as benchmark, estimated 100-year RP values for  $SS_L$  based on HTCs are underestimated by more than 250%, while those based on RTCs are only underestimated by ~10% at similar computational costs. This approach thus balances accuracy with efficiency in estimating extreme values for hydro-meteorological conditions, particularly for more extreme events not well-captured by historical tracks.

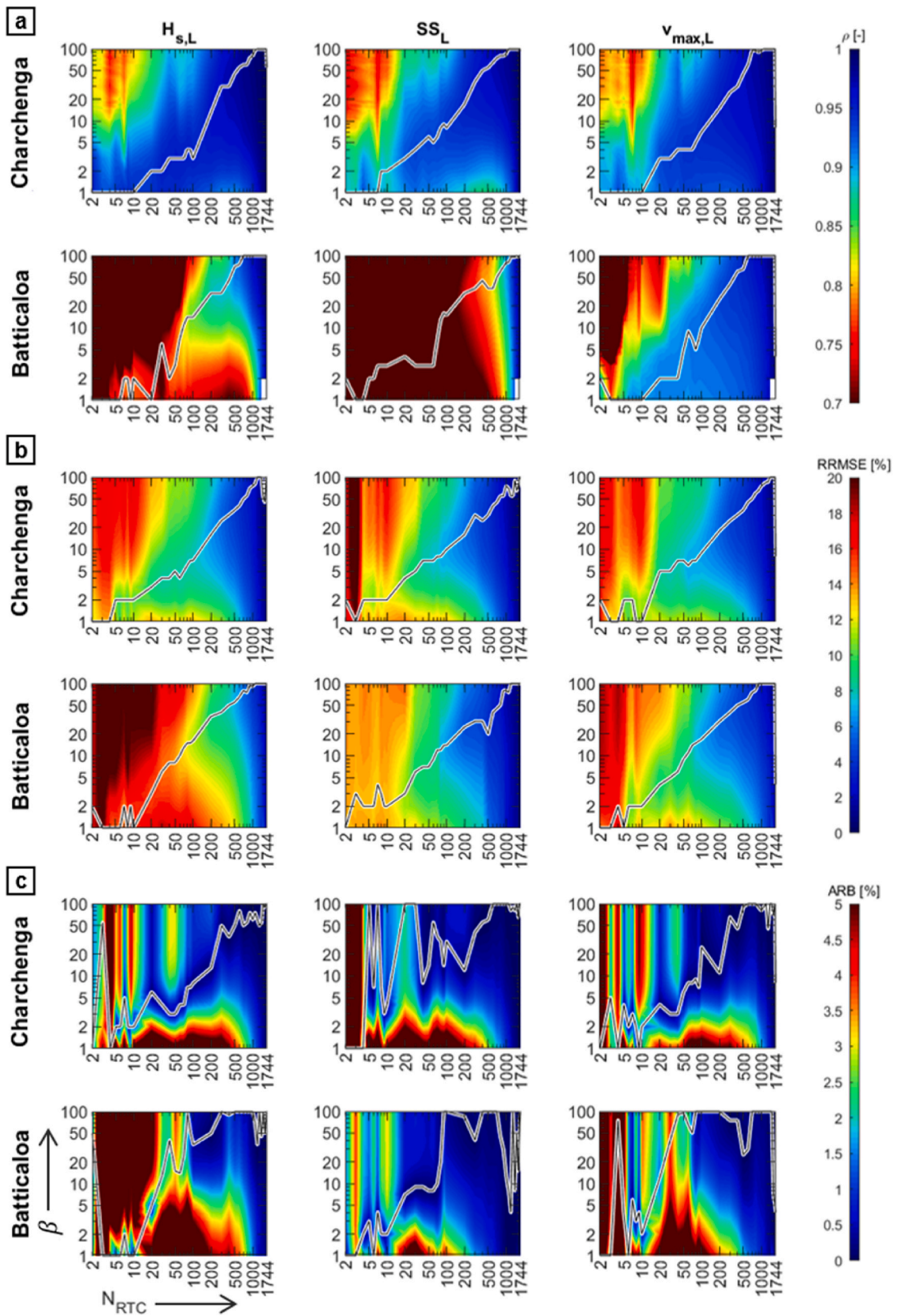
## 5. Discussion

In this section we discuss the main assumptions and limitations of the presented method and the underlying data in our case study. We also provide some suggestions for future applications and research.

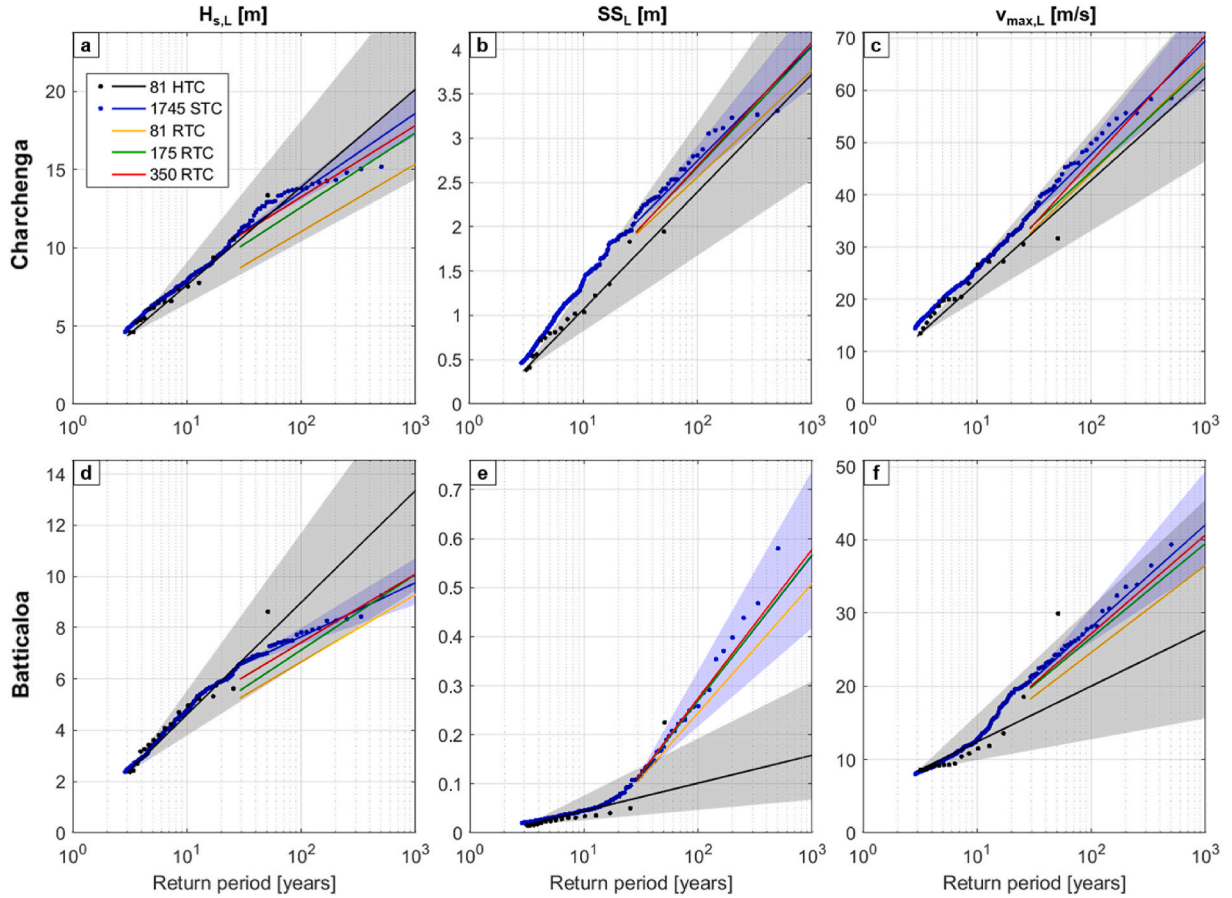
### 5.1. Optimal configuration

For making estimates of TC-induced hydro-meteorological conditions, our method enables a better balance between computational costs per simulation and the number of scenarios simulated, compared to using the HTCs or STCs directly. This approach can be used to increase efficiency by simulating fewer tracks, accuracy by permitting the use of higher resolution models, or a combination of both. However, for selecting the configuration and number of RTCs to be simulated, a trade-off between accuracy and efficiency must be made, in which the spatial scale of the analysis should also be considered.

Three configurations were tested with increasing performance. The first configuration (FB) has no noteworthy advantages and was outperformed by both the second (FBMPS) and third (SLMPS) configurations. For large-scale assessments (e.g., for hazard assessment of coastal areas around the full Bay of Bengal), configuration FBMPS has the clear advantage that one set of RTCs represents the full basin, while with configuration SLMPS a new set of RTCs is selected for each location. For example, to represent STCs at both Charchenga and Batticaloa with configuration SLMPS, 291 RTCs (two times 175 RTCs with 59 overlapping) are needed. Nevertheless, performance at specific sites increased significantly after inclusion of the location of interest in configuration SLMPS compared to configurations FB and FBMPS for the full basin. With 175 RTCs, the SLMPS configuration was capable of selecting TC scenarios that represent the offshore hydro-meteorological conditions for the full set of scenarios and the only configuration



**Fig. 9.** Performance of the SLMPS configuration, focused on one selected location with modifications to the parameter space. Error metrics for different combinations of the number of RTCs (NRTC) and values of the stiffness parameter ( $\beta$ ) in the Softmax interpolation method (logarithmic scales) are presented, for Charchenga and Batticaloa locations. Presented in colorscale are (a) Spearman rank correlation coefficient ( $\rho$ ), (b) Relative Root Mean Square Error (RRMSE), and (c) Absolute value of Relative Bias (ARB) for maximum significant wave height ( $H_{s,L}$ ), storm surge (SSL), and wind speed ( $V_{max,L}$ ). Cooler colors indicate better performance. For each given value of NRTC, the optimum value of  $\beta$  is indicated (black line).



**Figure 10.** Extreme value analysis of hydro-meteorological conditions based on simulation of 81 HTC, 1745 STC and STCres based on different numbers of RTC (NRTC = 81, 175, 350). Hydro-meteorological conditions are quantified in terms of significant wave height ( $H_{s,L}$ ), storm surge (SSL), and maximum wind speed ( $V_{max,L}$ ) for Charchenga (a–c) and Batticaloa (d–f) locations. Exponential distributions (solid lines) with 95% confidence intervals (shaded, only for HTC (black) and STC (blue)) are fitted to the data (markers, only shown for HTCs and STCs). Fitted distributions for SSL (b and e) based on STCs and based on STCres (NRTC = 175, 350) are very similar and partially overlap in the figure.

including extreme events. Hence, for local applications in TC hazard assessment, configuration SLMPs is recommended.

Individual STCres estimates of maximum  $H_{s,L}$ ,  $SSL$ , and  $v_{max,L}$  can deviate considerably from their STC counterpart value (Fig. 8). Hence, reconstructed hydro-meteorological conditions for the individual STCres should not be used. Due to the overall low RB of these estimates (Table 2), this is partially averaged out, and quantiles for ECDFs (Fig. 8) or fitted exponential functions (Fig. 10) typically show good agreement.

As an alternative to MDA selection, other sampling methods such as Latin hypercube or orthogonal sampling could also be considered to select the RTCs. However, MDA allows for selecting the extremes of the parameter space, which is advantageous for extreme value analyses. Note that the Softmax function used as interpolation function is often used as the activation function in the output layer of neural network models that predict a multinomial probability distribution. That is, Softmax is used as the activation function for multi-class classification problems (e.g., Antolínez et al., 2016; Chiri et al., 2019) where class membership is required on more than two class labels. We used the Softmax function similarly and we additionally investigated the sensitivity of its stiffness parameter on the reconstructed values, that in this application is shown to be sensitive to the number of selected RTCs (Section 4.2.2), showing a clear positive correlation between stiffness and a larger RTCs sample. We think this finding is intuitive, once there is a larger number of RTCs around and close to the unknown solution this parameter can be used to constrain the interpolation weights to its nearest neighbours. More complex alternatives to the simple Softmax interpolation based on statistical or machine learning methods that are

mesh-free, such as radial basis functions, are promising for further development.

For some locations, TCs making landfall nearby are most important (e.g., Charchenga), while for other locations extreme tracks passing at larger distance are equally important (e.g., Batticaloa). For example, TC generated swell waves have caused flooding at small islands in the Pacific over 1000 km away from the TC-track (Hoeke et al., 2020). In these environments, even for the SLMPs configuration, it is recommended to use a large selection domain that covers a wide range of tracks. In this way, the most extreme and the closest tracks are emphasized by modifying the parameter space, rather than by limiting the selection domain and excluding part of the tracks from the analysis.

## 5.2. TC tracks and hydro-meteorological conditions

In this study we took STCs and accompanying hydro-meteorological data following Leijnse et al. (2022) as benchmark. It should be noted that the synthetic tracks – and the sampled representative subset – are based on historical track records, thus they are dependent on the quality and length of those records (e.g., Mori et al., 2021). The underlying spatio-temporal statistical distributions (used in STC emulators) suffer from using relatively small datasets (observed or reanalysis data) in cyclogenesis and storm motion prediction. Furthermore, the decay rate of TCs making landfall or with motion on cold water is often parameterized only as a function of time (Kaplan and DeMaria, 1995; Vickery, 2005).

Astronomical tidal variations, being driven independent of the TCs,



were not included in the dataset. This allows for direct comparison of the storm surge and wave conditions for the different TCs. However, for realistic flood hazard assessment, the tide should certainly be considered in the model simulations to account for interactions between tide, storm surge, and waves. Similarly, precipitation was not included but can contribute to the TC flood hazard. Tide and precipitation could potentially be added to the method (e.g., by making combined synthetic TC track, tide and precipitation scenarios, and additional descriptors in the representative scenario selection method). In the RTC selection method, the air pressure deficit was not included for reasons explained in Section 3.3.1. Other STC emulators (e.g., Vickery et al., 2000; James and Mason, 2005; Emanuel et al., 2006; Nakajo et al., 2014; Bloemendaal et al., 2020; Arthur, 2021) do include air pressure and it could also be added as an additional descriptor similar to the windspeed. Adding such variables would help to better describe the TC activity and therefore would potentially help in the selection of key TCs to characterize the TC activity.

### 5.3. General applicability

In this case study, the method was tested at a specific location, and dataset of STC tracks generated with a specific STC emulator (i.e., TCWiSE), which in turn was based on a specific dataset of HTC tracks. However, the method uses key parameters for the characterization of TC activity and models that have shown the potential to be set-up in any specific regional setting. Other datasets of TC tracks, such as STCs generated with other emulators (see Section 5.2) or potentially also data from dynamical downscaling (e.g., Mori and Takemi, 2016; Jyoteeshkumar Reddy et al., 2021) could also be used. For areas where good-quality track records are readily available, HTC tracks could also be used directly (or a combination of HTC and STC tracks). Nevertheless, using synthetic tracks has the clear advantage of accounting for a wider range of events and reducing the uncertainty in extreme value predictions (Leijnse et al., 2022) at relatively low computational costs for generation of the STCs. Hybrid methods by Kim et al. (2015) and Jia et al. (2016) use idealized synthetic TC tracks and can accurately reconstruct time series of hydro-meteorological conditions for the range of TC tracks they have been derived for. In contrast, our method only reconstructs maximum values of hydro-meteorological conditions, but is more flexible and general, as we base our method on TC track parameters that are independent of the TC dataset or emulator, geographical location, or the TC making landfall.

In our case study an existing dataset of STCs with simulated waves and surge levels was used, which allowed extensive validation of the STCrecs estimates of maximum  $H_{s,L}$ ,  $SS_L$  by one-to-one comparison with their STC counterpart. However, in most applications such information will not be available, being a major reason to use this method. For a given desired accuracy level, estimates of the required number of RTCs and the value of stiffness parameter  $\beta$  for the SLMPS configuration, presented in Section 4.2.2 and Fig. 9 can be used as a basis. However, it is highly recommended to perform additional validation of the representativeness of the selected RTCs to the original dataset. For example, by comparison of values of  $v_{max,L}$  for the STCrecs with their STCs counterpart (as in Fig. 8c and g), as  $v_{max,L}$  can be obtained directly from the windfields without performing computationally expensive model simulations (e.g., TCWiSE provides windfields directly). For example, such strategy was applied in a case study for Majuro (Republic of the Marshall Islands, Western North Pacific basin) (Bakker, 2020).

### 5.4. Implications for TC hazard assessment

Quantiles of hydro-meteorological conditions obtained after EVA can be used directly if one is interested in the conditions offshore (as in our application), or propagated further shoreward (e.g., to obtain design values or for flood hazard assessment). Performance was typically better for locations with more frequent TCs (as more data was available), or

where more extreme conditions occur. However, RTC scenarios add value for hazard assessment especially in areas with infrequent TCs, for which sufficient historical TC data is currently unavailable. They are also well-suited to estimating conditions associated with longer return periods when simulation of large sets of synthetic tracks is computationally unfeasible.

In this study, offshore hydro-meteorological conditions were reconstructed. Alternatively, the RTC scenarios may potentially also first be propagated further landwards before reconstructing STCrecs scenarios. For example, such an approach was applied by Bakker (2020) for estimating flooding extents, but validation remains challenging, as propagation of a full set of STC scenarios for a reference data set was not feasible. Other potential future applications include early warning systems, where a good balance between rapid assessment and accurate results is essential.

## 6. Conclusions

A new general method for efficient and accurate modelling of Tropical Cyclone (TC) induced hazards of flooding and extreme winds is presented. In the method, a comprehensive subset of Representative Tropical Cyclone (RTC) tracks is selected from a larger set of tracks by maximizing storm dissimilarity based on TC track characteristics. Extremes of the original set of TC tracks are included, and the information about frequency of TC and forcing conditions is preserved by statistical reconstruction of the original set based on the simulated RTC tracks. The method was successfully applied to a pre-existing dataset of 1745 Synthetic Tropical Cyclone (STC) tracks (equivalent to 1000 years of TC conditions with the frequency of occurrence of the historical activity) with accompanying simulated maximum significant wave height ( $H_{s,L}$ ), storm surge ( $SS_L$ ) and wind speeds ( $v_{max,L}$ ) for nine selected locations in the Bay of Bengal. Performance was tested by comparing simulated values of  $H_{s,L}$ ,  $SS_L$ , and  $v_{max,L}$  of the full simulated set of STCs with their reconstructed value STCrecs, as interpolated from a smaller set of RTCs.

We tested configurations Full Basin (FB), Full Basin Modified Parameter Space (FBMPS) and Selected Location Modified Parameter Space (SLMPS), which offered progressively better performance with each level of additional customization based on parameters and locations of interest. Configurations FB and FBMPS were designed to represent TC tracks for the full Bay of Bengal (regional setting), while SLMPS was designed to represent TC tracks per (project) location. In FBMPS and SLMPS the parameter space was modified whereby tracks with extreme windspeeds were made more important in the selection method. In addition, in SLMPS tracks passing close to the location of interest were given greater weight in interpolation. The approach was capable of making more accurate and comprehensive predictions of TC-induced hydro-meteorological conditions for longer return periods compared to simulating only Historical Tropical Cyclone (HTC) tracks, and with significantly lower computational costs compared to simulating all possible STC tracks (175 vs 1745).

Overall, the SLMPS configuration showed the best performance (typically within 10% deviation) in approximating the quantiles for return periods ranging 10–500 years with 90% reduction in simulated TC scenarios. By using more RTCs, the original set can be better approximated, but the rate of improvement decreases with an increasing number of representative tracks. The optimal value of the stiffness parameter ( $\beta$ ) in the interpolation method increases with increasing number of RTCs. The presented configurations can also guide future applications, especially the parameterization of TC track scenarios.

The method is general so it can be applied to any set of TC tracks worldwide in combination with any hazard model. Our approach enables a more efficient and accurate TC (flood) hazard assessment; especially for infrequently impacted areas for which currently insufficient HTC data is available, when simulating large sets of synthetic tracks is infeasible, numerical models for hydrodynamics and waves need to be run on a higher resolution, or when only a limited set of tracks



can be modelled. Therefore, our method can improve or make feasible detailed flood hazard and risk assessments for areas where this was not feasible before.

## Data and code availability

The IBTrACS dataset for HTC tracks is publicly available from: <https://www.ncdc.noaa.gov/ibtracs/>

The TCWiSE tool for generation of STCs is available through: <https://www.deltares.nl/en/software/tcwise/>

We aim to include the Matlab code for selection of RTCs in a future release of the TCWiSE tool. The Delft3D Flexible Mesh Suite is available through: <https://www.deltares.nl/en/software/delft3d-flexible-mesh-suite>.

## Author contribution

All authors have read and agree to the published version of the manuscript. Conceptualization, methodology and writing – review and editing by all authors; software and resources by TB, JA and TL; formal analysis, investigation, data curation and visualisation by TB, JA and TL; writing – original draft preparation by TB; funding acquisition by AG.

## Disclaimer

The views expressed in the article do not reflect the views of ADB.

## 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.

## Acknowledgements

We acknowledge the Deltares research programs “Seas and Coastal Zones” and “Planning for Disaster Risk Management and Resilience” which have provided funding to this project and to write this paper. We want to thank Stefan Aarninkhof, Jeremy Bricker, and Luisa Torres Dueñas for their feedback during the MSc thesis leading up to this paper. Also, we want to thank the two anonymous reviewers for their critical and useful feedback.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.coastaleng.2022.104154>.

## References

- Antolínez, J.A.A., Méndez, F.J., Camus, P., Vitousek, S., González, E.M., Ruggiero, P., Barnard, P., 2016. A multiscale climate emulator for long-term morphodynamics (MUSCLEmorpho). *J. Geophys. Res.: Oceans* 121, 1–16. <https://doi.org/10.1002/2015JC011107>.
- Arthur, W.C., 2021. A statistical-parametric model of tropical cyclones for hazard assessment. *Nat. Hazards Earth Syst. Sci.* 21, 893–916. <https://doi.org/10.5194/nhess-21-893-2021>.
- Bakker, T.M., 2020. Compound Flood Hazard Assessment of Atoll Islands Based on Representative Scenarios for Typhoons and Non-typhoon Conditions: A Majuro Case Study. Master Thesis, TU Delft. <http://resolver.tudelft.nl/uuid:5fa05ec9-d901-4b47-9cd4-dcc36f5ccaaf>.
- Bass, B., Bedient, P., 2018. Surrogate modeling of joint flood risk across coastal watersheds. *J. Hydrol.* 558, 159–173. <https://doi.org/10.1016/j.jhydrol.2018.01.014>.
- Becker, J.J., Sandwell, D.T., Smith, W.H.F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S.-H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Trimmer, R., Von Rosenberg, J., Wallace, G., Weatherall, P., 2009. Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30 PLUS. *Mar. Geodes.* 32 (4), 355–371. <https://doi.org/10.1080/01490410903297766>.
- Bloemendaal, N., Haigh, I.D., de Moel, H., Muis, S., Haarsma, R.J., Aerts, J.C., 2020. Generation of a global synthetic tropical cyclone hazard dataset using storm. *Sci. Data* 7 (1), 1–12. <https://doi.org/10.1038/s41597-020-0381-2>.
- Booij, N., Ris, R.C., Holthuijsen, L.H., 1999. A third-generation wave model for coastal regions: 1. Model description and validation. *J. Geophys. Res. C: Oceans* 104 (C4), 7649–7666. <https://doi.org/10.1029/98jc02622>.
- Caires, S., 2016. A comparative simulation study of the annual maxima and the peaks-over-threshold methods. *J. Offshore Mech. Arctic Eng.* 138 (5) <https://doi.org/10.1115/1.4033563>.
- Camargo, S.J., Robertson, A.W., Gaffney, S.J., Smyth, P., Ghil, M., 2007. Cluster analysis of typhoon tracks. Part I: general properties. *J. Clim.* 20 (14), 3635–3653. <https://doi.org/10.1175/JCLI4188.1>.
- Chen, J.-H., Lin, S.-J., Magnusson, L., Bender, M., Chen, X., Zhou, L., et al., 2019. Advancements in hurricane prediction with NOAA's next-generation forecast system. *Geophys. Res. Lett.* 46, 4495–4501. <https://doi.org/10.1029/2019GL082410>.
- Chiri, H., Abascal, A.J., Castaneda, S., et al., 2019. Statistical simulation of ocean current patterns using autoregressive logistic regression models: a case study in the Gulf of Mexico. *Ocean Model.* 136 <https://doi.org/10.1016/j.ocemod.2019.02.010>.
- Choi, K.-S., Kim, B.-J., Choi, C.-Y., Nam, J.-C., 2009. Cluster Analysis of tropical cyclones making landfall on the Korean peninsula. *Adv. Atmos. Sci.* 26 (2), 202–210. <https://doi.org/10.1007/s00376-009-0202-1>.
- Dekker, M.M., Haarsma, R.J., de Vries, H., Baatsen, M., van Delden, A.J., 2018. Characteristics and development of European cyclones with tropical origin in reanalysis data. *Clim. Dynam.* 50 (1–2), 445–455. <https://doi.org/10.1007/s00382-017-3619-8>.
- Dullaart, J.C.M., Muis, S., Bloemendaal, N., et al., 2021. Accounting for tropical cyclones more than doubles the global population exposed to low-probability coastal flooding. *Coomun. Earth Environ.* 2, 135. <https://doi.org/10.1038/s43247-021-00204-9>.
- Emanuel, K., 2003. Tropical cyclones. *Annu. Rev. Earth Planet Sci.* 31 (1), 75–104. <https://doi.org/10.1146/annurev.earth.31.100901.141259>.
- Emanuel, K., Ravela, S., Vivant, E., Risi, C., 2006. A statistical deterministic approach to hurricane risk assessment. *Bull. Am. Meteorol. Soc.* 87 (3), 299–314. <https://doi.org/10.1175/BAMS-87-3-299>.
- Fritz, H.M., Blount, C.D., Thwin, S., Thu, M.K., Chan, N., 2009. Cyclone Nargis storm surge in Myanmar. *Nat. Geosci.* 2 (7), 448–449. <https://doi.org/10.1038/ngeo558>.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Cambridge, MA. <http://www.deeplearningbook.org/>.
- Haigh, I.D., MacPherson, L.R., Mason, M.S., Wijeratne, E., Pattiaratchi, C.B., Crompton, R.P., George, S., 2014. Estimating present day extreme water level exceedance probabilities around the coastline of Australia: tropical cyclone-induced storm surges. *Clim. Dynam.* 42 (1–2), 139–157. <https://doi.org/10.1007/s00382-012-1653-0>.
- Hoeke, R.K., Damlamian, H., Aucan, J., Wandres, M., 2020. Severe flooding in the atoll nations of Tuvalu and Kiribati triggered by a distant Tropical Cyclone Pam. *Front. Mar. Sci.* 7, 991. <https://doi.org/10.3389/fmars.2020.539646>.
- Hojjat Ansari, A., Olyaei, M.A., Heydari, Z., 2021. Ensemble generation for hurricane hazard assessment along the United States' Atlantic coast. *Coast. Eng.* 169, 103956. <https://doi.org/10.1016/j.coastaleng.2021.103956>.
- Holland, G., 2008. A revised hurricane pressure–wind model. *Mon. Weather Rev.* 136 (9), 3432–3445. <https://doi.org/10.1175/2008mwr2395.1>.
- James, M.K., Mason, L.B., 2005. Synthetic tropical cyclone database. *J. Waterw. Port, Coast. Ocean Eng.* 131 (4), 181–192. [https://doi.org/10.1061/\(asce\)0733-950x\(2005\)131:4\(181\)](https://doi.org/10.1061/(asce)0733-950x(2005)131:4(181)).
- Jia, G., Taflanidis, A.A., 2013. Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment. *Comput. Methods Appl. Mech. Eng.* 261, 24–38. <https://doi.org/10.1016/j.cma.2013.03.012>.
- Jia, G., Taflanidis, A.A., Nadal-Caraballo, N.C., Melby, J.A., Kennedy, A.B., Smith, J.M., 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 (2), 909–938. <https://doi.org/10.1007/s11069-015-2111-1>.
- Jyoteeshkumar Reddy, P., Sriram, D., Gunthe, S.S., Balaji, C., 2021. Impact of climate change on intense Bay of Bengal tropical cyclones of the post-monsoon season: a pseudo global warming approach. *Clim. Dynam.* 56, 2855–2879. <https://doi.org/10.1007/s00382-020-05618-3>.
- Kaplan, J., DeMaria, M., 1995. A simple empirical model for predicting the decay of tropical cyclone winds after landfall. *J. Appl. Meteorol. Climatol.* 34 (11), 2499–2512. [https://doi.org/10.1175/1520-0450\(1995\)034<2499:ASEMFP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1995)034<2499:ASEMFP>2.0.CO;2).
- Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. *Technometrics* 11 (1), 137–148. <https://doi.org/10.1080/00401706.1969.10490666>.
- Kernkamp, H.W.J., Van Dam, A., Stelling, G.S., de Goede, E.D., 2011. Efficient scheme for the shallow water equations on unstructured grids with application to the Continental Shelf. *Ocean Dynam.* 61 (8), 1175–1188. <https://doi.org/10.1007/s10236-011-0423-6>.
- Kim, S.-W., Melby, J.A., Nadal-Caraballo, N.C., Ratcliff, J., 2015. A time-dependent surrogate model for storm surge prediction based on an artificial neural network using high-fidelity synthetic hurricane modeling. *Nat. Hazards* 76 (1), 565–585. <https://doi.org/10.1007/s11069-014-1508-6>.
- 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 (Accessed: 26-05-2020). <https://doi.org/10.25921/82ty-9e16>.

- Knutson, T.R., Sirutis, J.J., Zhao, M., Tuleya, R.E., Bender, M., Vecchi, G.A., Villarini, G., Chavas, D., 2015. Global projections of intense tropical cyclone activity for the late twenty-first century from dynamical downscaling of CMIP5/RCP4.5 scenarios. *J. Clim.* 28 (18), 7203–7224. <https://doi.org/10.1175/jcli-d-15-0129.1>.
- Leijnse, T.W.B., Giardino, A., Nederhoff, K., Caires, S., 2022. Generating reliable estimates of tropical cyclone induced coastal hazards along the Bay of Bengal for current and future climates using synthetic tracks. *Nat. Hazards Earth Syst. Sci.* 22 (6), 1863–1891. <https://doi.org/10.5194/nhess-22-1863-2022>.
- Lin, N., Emanuel, K., 2015. Grey swan tropical cyclones. *Nat. Clim. Change* 6 (1), 106–111. <https://doi.org/10.1038/nclimate2777>.
- Magnusson, L., Bidlot, J.-R., Bonavita, M., Brown, A., Browne, P., De Chiara, G., Dahoui, M., Lang, S.T.K., McNally, T., Mogensen, K.S., Pappenberger, F., Prates, F., Rabier, F., Richardson, D.S., Vitart, F., 2019. ECMWF activities for improved hurricane forecasts. *Bull. Am. Meteorol. Soc.* 100, 445–458. <https://doi.org/10.1175/BAMS-D-18-0044.1>.
- Mendelsohn, R., Emanuel, K., Chonabayashi, S., Bakkensen, L., 2012. The impact of climate change on global tropical cyclone damage. *Nat. Clim. Change* 2 (3), 205–209. <https://doi.org/10.1038/nclimate1357>.
- Mori, N., Takemi, T., 2016. Impact assessment of coastal hazards due to future changes of tropical cyclones in the North Pacific Ocean. *Weather Clim. Extrem.* 11, 53–69. <https://doi.org/10.1016/j.wace.2015.09.002>.
- Mori, N., Takemi, T., Tachikawa, Y., Tatano, H., Shimura, T., Tanaka, T., Fujimi, T., Osakada, Y., Webb, A., Nakakita, E., 2021. Recent nationwide climate change impact assessments of natural hazards in Japan and East Asia. *Weather Clim. Extrem.* 32, 100309. <https://doi.org/10.1016/j.wace.2021.100309>.
- Nakajo, S., Mori, N., Yasuda, T., Mase, H., 2014. Global stochastic tropical cyclone model based on principal component analysis and cluster Analysis. *J. Appl. Meteorol. Climatol.* 53 (6), 1547–1577. <https://doi.org/10.1175/jamc-d-13-08.1>.
- Nederhoff, K., Giardino, A., van Ormondt, M., Vattvani, D., 2019. Estimates of tropical cyclone geometry parameters based on best-track data. *Nat. Hazards Earth Syst. Sci.* 19 (11), 2359–2370. <https://doi.org/10.5194/nhess-19-2359-2019>.
- 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. <https://doi.org/10.5194/nhess-2020-250>.
- Neumann, J.E., Emanuel, K.A., Ravela, S., Ludwig, L.C., Verly, C., 2015. Risks of coastal storm surge and the effect of sea level rise in the Red River Delta, Vietnam. *Sustainability* 7 (6), 6553–6572. <https://doi.org/10.3390/su7066553>.
- Niederoda, A.W., Resio, D.T., Toro, G.R., Divoky, D., Das, H.S., Reed, C.W., 2010. Analysis of the coastal Mississippi storm surge hazard. *Ocean Eng.* 37 (1), 82–90. <https://doi.org/10.1016/j.oceaneng.2009.08.019>.
- Ou, S.-H., Liao, J.-M., Hsu, T.-W., Tzang, S.-Y., 2002. Simulating typhoon waves by SWAN wave model in coastal waters of Taiwan. *Ocean Eng.* 29 (8), 947–971. [https://doi.org/10.1016/s0029-8018\(01\)00049-x](https://doi.org/10.1016/s0029-8018(01)00049-x).
- Peduzzi, P., Chatenoux, B., Dao, H., De Bono, A., Herold, C., Kossin, J., Mouton, F., Nordbeck, O., 2012. Global trends in tropical cyclone risk. *Nat. Clim. Change* 2 (4), 289–294. <https://doi.org/10.1038/nclimate1410>.
- Resio, D.T., Irish, J., Cialone, M., 2009. A surge response function approach to coastal hazard assessment – part 1: basic concepts. *Nat. Hazards* 51 (1), 163–182. <https://doi.org/10.1007/s11069-009-9379-y>.
- Scott, F., Antolinez, J.A.A., McCall, R., Storlazzi, C., Reniers, A., Pearson, S., 2020. Hydro-morphological characterization of coral reefs for wave runoff prediction. *Front. Mar. Sci.* 7. <https://doi.org/10.3389/fmars.2020.00361>. Article 361.
- Sebastian, A., Dupuits, E., Morales-Nápoles, O., 2017. Applying a Bayesian network based on Gaussian copulas to model the hydraulic boundary conditions for hurricane flood risk analysis in a coastal watershed. *Coast. Eng.* 125, 42–50. <https://doi.org/10.1016/j.coastaleng.2017.03.008>.
- Toro, G.R., Niederoda, A.W., Reed, C.W., Divoky, D., 2010. Quadrature-based approach for the efficient evaluation of surge hazard. *Ocean Eng.* 37 (1), 114–124. <https://doi.org/10.1016/j.oceaneng.2009.09.005>.
- Torres, J.M., Bass, B., Irza, N., Fang, Z., Proft, J., Dawson, C., Kiani, M., Bedient, P., 2015. Characterizing the hydraulic interactions of hurricane storm surge and rainfall-runoff for the Houston–Galveston region. *Coast. Eng.* 106, 7–19. <https://doi.org/10.1016/j.coastaleng.2015.09.004>.
- Vickery, P.J., Skerlj, P.F., Twisdale, L.A., 2000. Simulation of hurricane risk in the US using empirical track model. *J. Struct. Eng.* 126 (10), 1222–1237. [https://doi.org/10.1061/\(asce\)0733-9445\(2000\)126:10\(1222](https://doi.org/10.1061/(asce)0733-9445(2000)126:10(1222).
- Vickery, P.J., 2005. Simple empirical models for estimating the increase in the central pressure of tropical cyclones after landfall along the coastline of the United States. *J. Appl. Meteorol.* 44 (12), 1807–1826. <https://doi.org/10.1175/JAM2310.1>.
- Vousdoukas, M.I., Mentaschi, L., Voukoulas, E., Verlaan, M., Jevrejeva, S., Jackson, L. P., Feyen, L., 2018. Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard. *Nat. Commun.* 9 (1) <https://doi.org/10.1038/s41467-018-04692-w>.
- Xu, K., Ma, C., Lian, J., Bin, L., 2014. Joint probability analysis of extreme precipitation and storm tide in a coastal city under changing environment. *PLoS One* 9 (10), e109341. <https://doi.org/10.1371/journal.pone.0109341>.
- Zheng, F., Westra, S., Sisson, S.A., 2013. Quantifying the dependence between extreme rainfall and storm surge in the coastal zone. *J. Hydrol.* 505, 172–187. <https://doi.org/10.1016/j.jhydrol.2013.09.054>.

## Glossary

### General

- HTC*: Historical Tropical Cyclone (IBTrACS data in our case study)
- STC*: Synthetic Tropical Cyclone (generated with TCWiSE in our case study)
- RTC*: Representative Tropical Cyclone (selected subset of STC)
- TCrec*: reconstructed Tropical Cyclone (based on interpolation)
- STCrecs*: reconstructed Synthetic Tropical Cyclone (based on interpolation, as we reconstruct the STCs we have STCrecs in our case study)
- IBTrACS*: International Best Track Archive for Climate Stewardship (dataset)
- TCWiSE*: Tropical Cyclone Wind Statistical Estimation (tool for STC generation)
- BoB*: Bay of Bengal
- $H_{s,i}$ : maximum significant wave height during a TC at a specific location
- $SS_i$ : maximum storm surge during a TC at a specific location
- $v_{max,i}$ : maximum sustained windspeed during a TC at a specific location

### To describe a TC (track)

- SL*: Selected Location
- descriptor*: describes a characteristic of the TC, quantified by a number of parameters
- parameter*: used to parameterize the TC track, see  $v_{max,min}$  to *ATSL* below
- $v_{max}$ : maximum sustained windspeed of a TC, value specified for moment in time and location (lat,lon), obtained from IBTrACS or TCWiSE
- $c$ : translation speed / forward velocity of a TC (eye), value specified for moment in time and location (lat,lon), approximated by calculating the distance between two consecutive TC locations (lat,lon), divided over the time past; lat, lon and time obtained from IBTrACS or TCWiSE
- $v_{max,min}$ : minimum value of  $v_{max}$  of the full TC track
- $v_{max,median}$ : median value of  $v_{max}$  of the full TC track
- $v_{max,max}$ : maximum value of  $v_{max}$  of the full TC track
- $c_{min}$ : minimum value of  $c$  of the full TC track
- $c_{median}$ : median value of  $c$  of the full TC track
- $c_{max}$ : maximum value of  $c$  of the full TC track
- duration*: total duration / lifetime of the TC
- $lat_{gen}$ : latitude coordinate of cyclogenesis location
- $lon_{gen}$ : longitude coordinate of cyclogenesis location
- $lat_{1/3}$ : latitude coordinate of the TC at 1/3 of its total lifetime
- $lon_{1/3}$ : longitude coordinate of the TC at 1/3 of its total lifetime
- $lat_{2/3}$ : latitude coordinate of the TC at 2/3 of its total lifetime
- $lon_{2/3}$ : longitude coordinate of the TC at 2/3 of its total lifetime
- $lat_{term}$ : latitude coordinate of TC termination location
- $lon_{term}$ : longitude coordinate of TC termination location
- DTSL*: minimum Distance of the (TC) Track to SL
- ATSL*: Azimuth of *DTSL* (in degrees north)

### To select RTCs

- FB*: Full Basin (configuration of the method)
- FBMPS*: Full Basin Modified Parameter Space (configuration of the method)
- SLMPS*: Selected Location Modified Parameter Space (configuration of the method)
- MDA*: Maximum Dissimilarity Algorithm (used to select the RTCs)
- $N_{par/des}$ : Number of parameters in a descriptor
- $F_w$ : weight factor (applied to each of the normalized parameters in the descriptor)
- $F_s$ : scaling factor (applied to each of the normalized parameters in the descriptor)
- $N_{RTC}$ : the number of RTCs
- $N_{STC}$ : the number of STCs

### To reconstruct the STCrecs

- Softmax*: Interpolation function used to reconstruct the STCrecs
- $S(x)_i$ : probability of the STC matching to RTC  $i$
- $\beta$ : stiffness parameter
- $d_i$ : the (Euclidean) distance between the STC and RTC  $i$
- $d_j$ : the (Euclidean) distance between the RTC and approximated STC  $j$

### Errors and extreme value analysis

- EVA*: Extreme Value Analysis
- RP*: Return Period
- ECDF*: Empirical Cumulative Distribution Function
- $\rho$ : Spearman rank correlation coefficient
- RMSE*: Root Mean Square Error
- RRMSE*: Relative Root Mean Square Error, obtained by dividing the RMSE values (of the full set of STCs) by their 100-year RP value based on the ECDF of the STCs
- RB*: Relative Bias, obtained by dividing the bias values (of the full set of STCs) by their 100-year RP value based on the ECDF of the STCs
- ARB*: Absolute value of the Relative Bias (RB)