

OPEN  
ANALYSIS

# Estimation of global tropical cyclone wind speed probabilities using the STORM dataset

Nadia Bloemendaal <sup>1</sup>✉, Hans de Moel<sup>1</sup>, Sanne Muis <sup>1,2</sup>, Ivan D. Haigh<sup>3</sup> & Jeroen C. J. H. Aerts<sup>1</sup>

Tropical cyclones (TC) are one of the deadliest and costliest natural disasters. To mitigate the impact of such disasters, it is essential to know extreme exceedance probabilities, also known as return periods, of TC hazards. In this paper, we demonstrate the use of the STORM dataset, containing synthetic TCs equivalent of 10,000 years under present-day climate conditions, for the calculation of TC wind speed return periods. The temporal length of the STORM dataset allows us to empirically calculate return periods up to 10,000 years without fitting an extreme value distribution. We show that fitting a distribution typically results in higher wind speeds compared to their empirically derived counterparts, especially for return periods exceeding 100-yr. By applying a parametric wind model to the TC tracks, we derive return periods at 10 km resolution in TC-prone regions. The return periods are validated against observations and previous studies, and show a good agreement. The accompanying global-scale wind speed return period dataset is publicly available and can be used for high-resolution TC risk assessments.

## Introduction

Tropical cyclones (TCs) are amongst the deadliest and costliest natural disasters, affecting people, economies and the environment in coastal areas around the globe. In 2019, Cyclone Idai caused over 1,000 fatalities and displaced 3 million people upon landfall in Mozambique<sup>1</sup>. In 2017, Hurricanes Harvey, Irma and Maria entered the top-5 costliest Atlantic hurricanes ever, with combined losses estimated at \$220 billion<sup>2</sup>. To minimize future loss of life and property, it is crucial to perform accurate TC risk assessments and identify high-risk locations so that appropriate protection measures can be designed.

Wind is one of the major hazards associated with TCs and can do substantial damage to housing, infrastructure and ecosystems both in coastal regions and far inland<sup>3</sup>. Moreover, wind correlates with the intensity of other TC-induced hazards, such as storm surges, waves and precipitation<sup>4–6</sup>. To enhance our understanding of TC risk at the global scale, it is therefore essential to analyze wind speed probabilities in coastal zones. Risk is commonly calculated as the integrated value of expected damages over all exceedance probabilities –the inverse of these probabilities being return periods (RPs)<sup>7</sup>. As such, accurately calculating risk requires information on a wide range of RPs. Simpson and Lawrence<sup>8</sup> empirically estimated TC RPs along 80 km-long coastal segments of the US coastline based on historical TCs. However, RPs could not be calculated for those coastal segments that were not hit by a TC in the 85 years of observations. This shows that, due to the short length of the observational record, data often needs to be aggregated over larger spatial regions to perform a RP analysis, hereby omitting the spatial heterogeneity. Moreover, estimating RPs comes with large uncertainties, especially for RPs exceeding the length of the observational record. To overcome these limitations, the methodology of synthetic TC track generation has been developed over the past few decades<sup>9–12</sup>. In this approach, TCs, extracted from either historical data<sup>9,13</sup> or climate model simulations<sup>14</sup>, are statistically resampled and modeled to generate synthetic, but realistic, TCs. Using a Monte Carlo approach, this procedure is repeated recurrently to construct a TC dataset having the same statistical characteristics as the input dataset, but spanning hundreds to thousands of years.

<sup>1</sup>Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam, 1081 HV, Amsterdam, the Netherlands.

<sup>2</sup>Deltares, 2600 MH, Delft, The Netherlands. <sup>3</sup>School of Ocean and Earth Science, National Oceanography Centre, University of Southampton, European Way, Southampton, SO14 3ZH, United Kingdom. ✉e-mail: [nadia.bloemendaal@vu.nl](mailto:nadia.bloemendaal@vu.nl)

Using synthetic data enables the analysis of higher RPs and at higher spatial resolution than previously possible. In an accompanying paper we have presented the global synthetic model STORM (Synthetic Tropical cyclOne geneRation Model)<sup>12</sup>. The STORM dataset spans 10,000 years of global TC activity under present-day climate conditions, based on observed TC tracks. Here, we demonstrate usage of the STORM dataset by creating wind speed RPs at three different spatial scales: (i) basin level; (ii) within 100 km for 18 selected coastal cities and 63 islands; (iii) at 10 km resolution in TC-prone regions. This dataset is unique in presenting (high) RPs at a global scale for all TC-prone regions. More importantly, it represents an important step forward to calculating global TC damages and risk.

## Results

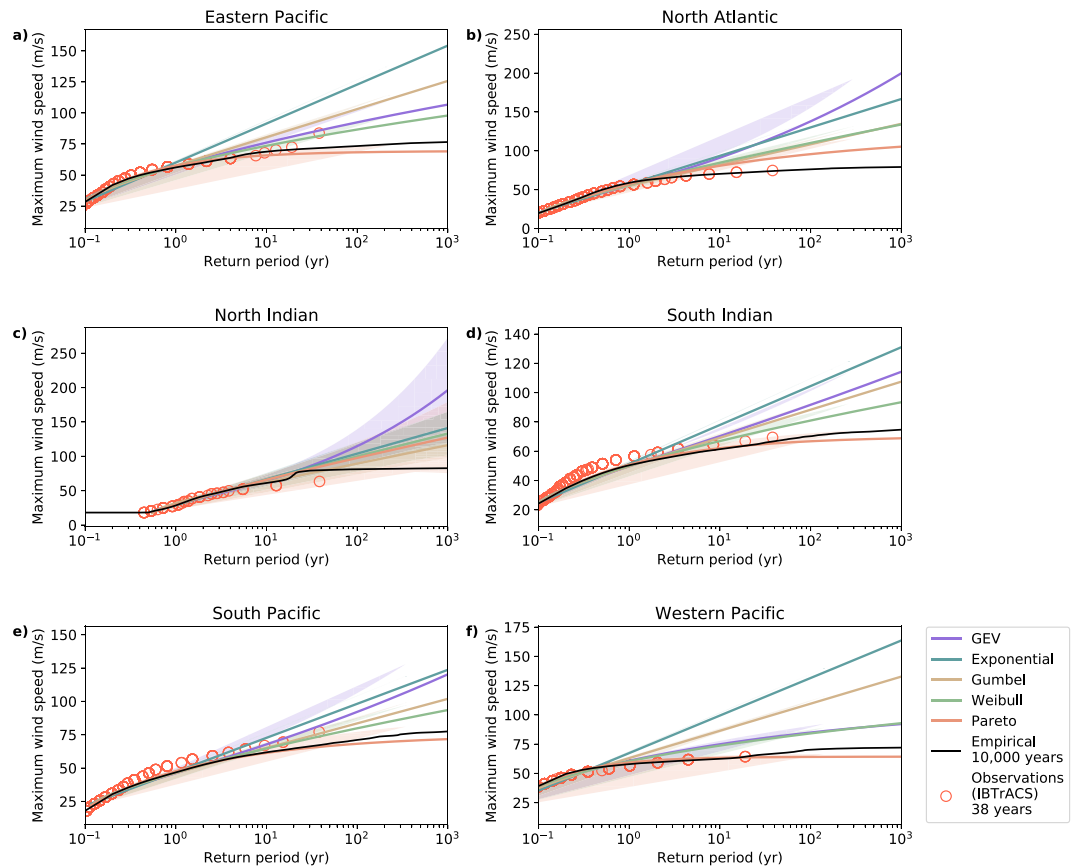
**Deriving return periods from the STORM dataset.** Calculating RPs of (extreme) wind speeds in the STORM dataset can be done either empirically or statistically. When using an empirical approach, RPs are directly calculated from wind speeds ranked in order of magnitude using formulas like Hazen's or Weibull's plotting formula<sup>15,16</sup>. A benefit of this approach is that no specific shape of the RP-curve is assumed; RPs are calculated per given wind speed without interpolation or smoothening of the data. However, the highest RP is limited by the temporal length of the data as this method does not allow for extrapolation beyond this timespan. TC risk assessments typically require information on extreme events that have not been observed yet. Therefore RPs are often determined by fitting extreme value (EV) distributions<sup>17</sup> to historical data. This way any RP can be estimated, even those beyond the range of observations. To ensure there is enough data for a good fit, this approach is generally carried out for ocean basins or relatively large coastal sections. Such fitted RPs are strongly influenced by the selected EV distribution, especially for higher RPs<sup>18</sup>. Furthermore, short records have large uncertainties, and typically multi-decadal records are needed for reliable estimates of the tail (high RPs)<sup>19</sup>.

Here we compare these two approaches to estimate basin-scale RPs using the synthetic TCs from STORM. We apply Weibull's plotting formula to the maximum 10-meter 10-minute average sustained wind speeds (max U10) in the full STORM dataset (10,000 years). We use max U10 because this is globally the most commonly reported value of wind speeds<sup>20</sup>, but other averaging periods can easily be obtained using conversion factors<sup>21</sup>. Next, we fit five EV distributions (the generalized extreme value, exponential, Gumbel, Weibull and Pareto distribution) to 1,000 random realizations of 38 years of data sampled from the STORM dataset. This 38-year length was chosen as the STORM dataset was created using 38 years of historical data (1980–2017) from the International Best Track dataset for Climate Stewardship (IBTrACS<sup>20</sup>).

At basin-level, the empirically derived STORM-RPs agree with the observed RPs (Fig. 1). In the North Atlantic, North Indian, and Western Pacific, STORM-RPs compare well with observations. In the Eastern Pacific and the Southern Hemisphere basins, max U10 in STORM are lower than the highest observed counterparts. Additionally, for four out of five EV distributions the max U10 values are substantially higher than the empirically derived values, particularly at RPs exceeding 100-yr (Fig. 1). Compared to the empirical curve, the worst-performing EV distributions are the exponential, generalized extreme value (GEV) and Gumbel distributions, which deviate from the empirical curve above the 10-yr RP. The EV distributions cannot capture the asymptotic behavior of the TC intensity, caused by environmental constraints such as the Maximum Potential Intensity<sup>22,23</sup>. Consequently, at high RPs, max U10 from EV distributions are higher than their empirically derived counterparts, with a maximum difference of 117 m/s for the GEV in the North Atlantic at the 1,000-yr RP. For the North Indian, we observe a kink in the empirical curve around the 20-yr RP: this is likely caused by an absence of certain max U10 in the dataset, which are then excluded from the RP calculation. This absence is driven by a limited spatial distribution of sea-surface temperatures (SSTs), causing a more frequent occurrence of higher max U10. Figure 1 shows that the Pareto distribution is the best-performing EV distribution compared to the empirical approach. However, for the North Indian, also the Pareto distribution shows substantially higher RPs compared to the empirical curve. Overall, the empirical probabilities of the observations are well in line with the estimates from the STORM dataset, and are considerably lower than using EV fits at RPs exceeding 100-yr.

**Tropical cyclone return periods for coastal cities.** Besides basin-scale RPs, we derive RPs for specific coastal locations using a 100 km radius to capture those TCs that have a substantial impact. We demonstrate this here for 18 coastal cities, but a similar dataset is available for 63 islands (see Data availability Statement). The RP-curves (subplots in Fig. 2; see Table 1) show that probabilities of a TC event occurring within 100 km differ strongly per city. Cities that are not regularly hit by a TC include San Diego (USA), Mumbai (India), and Muscat (Oman) with RPs for a Category-1 exceeding 100-yr. These relatively high RPs are driven by multiple TC characteristics. Firstly, TCs are generally deflected from San Diego and Mumbai<sup>24</sup> and instead move out over the open ocean. This is caused by TCs being embedded within the prevailing easterly (westward) flow at these latitudes. Secondly, TCs dissipate when they make landfall. As Muscat is located in the relatively narrow Gulf of Oman, most TCs in this region will likely have passed land upon approaching Muscat. Lastly, both Mumbai and Muscat lie in the North Indian, where approximately 2 TCs form per year<sup>12</sup>, hereby further decreasing the chances of being hit by a TC in any given year. Of the cities considered here, Taipei (Taiwan) and Tokyo (Japan) experience TCs most often, with RPs for a Category-1 lower than 4-yr. Both cities are located in the Western Pacific, the most active basin with 22.5 TC formations per year<sup>12</sup>. Of the cities considered, San Juan (Puerto Rico) and Chittagong (Bangladesh) are most often affected by strong TCs, with Category-5 TCs having a 131-yr RP for both cities. San Juan's central position in the tropical Atlantic Ocean combined with frequent Category-5 TC formations (approximately 1-in-6-years; see Fig. 1) likely drives these relatively low RPs. Chittagong's relatively low RP is likely due to the high SSTs in the Bay of Bengal, enhancing TC intensification and thus generating strong TCs.

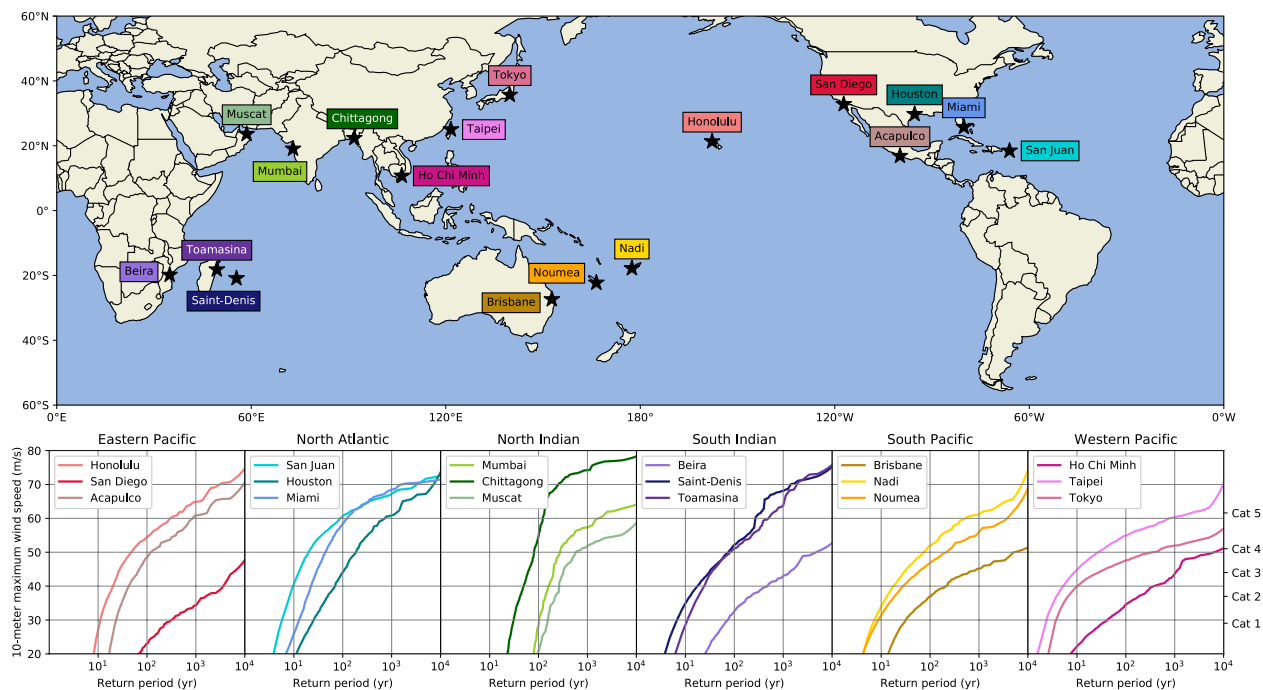
**Spatial distribution of extreme wind speeds.** Using STORM, we can also derive RPs at high (10 km) spatial resolution. While a single extreme event might be captured well in historical datasets (e.g. Hurricane Irma



**Fig. 1** Comparison of the wind speed return periods based on fitting five different extreme value distributions to 1,000 random realizations of 38 years and applying an empirical distribution (the Weibull plotting formula) to the full 10,000 years of data. Data is aggregated at basin level for each of the 6 ocean basins (a–f). The extreme value distributions are the generalized extreme value distribution (purple), the exponential distribution (blue), the Gumbel distribution (also known as the Generalized Extreme Value Distribution Type-I; yellow), the Weibull distribution (also known as the Generalized Extreme Value Distribution Type-II; green), and the Pareto distribution (red). Shaded areas indicate the 95%-confidence interval based on the bootstrap with 1,000 realizations. Empirically derived return periods from observations (IBTrACS) are given as red scatter points. See Methods for a full description of the basin domains.

in IBTrACS), there are not enough events in such datasets to robustly calculate high RPs. To calculate max U10 at 10 km resolution, we fit a 2D-parametric wind model to each synthetic TC (see Methods). Note that RPs inherently depend on the spatial scale they are computed at. At basin-scale, multiple TCs form every year, each one potentially reaching a given max U10. At a high-resolution grid cell, a TC passage can be rare. Hence, for equal max U10, RPs are lower when computed at higher resolution.

Figure 3 shows that max U10 increases between the 100-yr and 1,000-yr RP level, with largest increases in the Bay of Bengal, the North Atlantic, and west of Hawaii in the Eastern Pacific, and less profound in the mid-latitudes and over land. There is distinct spatial variation within basins driven by the strong relationship between SSTs and TC activity. This is for instant evident in the Bay of Bengal, where SSTs<sup>25</sup> of approximately 29 °C drive TC intensification, resulting locally in max U10 exceeding 65 m/s for a 1,000-yr RP. Aside from basin-scale variability in max U10, the two insets in Fig. 3 display the variability in max U10 at smaller scales. Distinct differences in max U10 are visible around the Philippines (inset Fig. 3a), with lower max U10 on the east side caused by the (westward) passage of TCs over the archipelago. Along the US coastline (inset Fig. 3b), stronger TCs make landfall more frequently on the Florida-North Carolina coastline than around New York City (NYC). This is driven by two factors: (i) TCs generally move north-westward near the Florida-North Carolina regions (onshore direction), whereas TCs are deflected north-eastward near NYC (offshore direction); and (ii) SSTs are higher along the southern coastline, supporting intense TCs, whilst the lower SSTs around NYC drive a weakening of TCs. The apparent re-intensification of TCs near 40°S is partly caused by relatively high SSTs of 17–22 °C. Another cause is that these mid-latitude regions are mostly dominated by extratropical cyclones, which follow a different intensification process than TCs. STORM, however, does not model the extratropical transition of TCs and as such may underestimate RPs in these regions. Another feature visible in predominantly the mid-latitudes is the dotted patterns (e.g. North of Hawaii, Fig. 3). These patterns are caused by the passage of few TCs, combined with a higher translational speed at these latitudes. As we use 3-hourly intervals, the max U10 values appear as dots.



**Fig. 2** Return periods of maximum wind speed (10-minute 10-meter average) within a radius of 100 km for a selection of coastal cities. Color tones correspond to the different basins: North Atlantic (blue), Eastern Pacific (red), Western Pacific (pink), North Indian (green), South Indian (purple), and South Pacific (yellow). Graphs show the return period on the x-axis versus the corresponding maximum wind speed (in m/s) on the y-axis. Category-classifications are based on the Saffir-Simpson scale (converted from 1-min to 10-min thresholds, see Methods).

City	Country	Cat 1 (29 m/s)	Cat 2 (37.6 m/s)	Cat 3 (43.4 m/s)	Cat 4 (51.1 m/s)	Cat 5 (61.6 m/s)
Honolulu	Hawaii (USA)	11	17	25	54	442
San Diego	USA	247	1870	5034	>10,000	>10,000
Acapulco	Mexico	22	35	62	156	1684
San Juan	Puerto Rico	5	8	12	22	131
Houston	USA	23	52	92	205	1230
Miami	USA	12	20	28	48	155
Mumbai	India	101	140	199	285	2403
Chittagong	Bangladesh	30	44	59	83	131
Muscat	Oman	151	235	315	768	>10,000
Beira	Mozambique	64	238	1179	6817	>10,000
Saint-Denis	Réunion	6	13	28	87	347
Toamasina	Madagascar	10	19	31	101	616
Brisbane	Australia	28	111	466	9013	>10,000
Nadi	Fiji	7	14	27	87	1093
Noumea	New-Caledonia	8	20	51	263	4755
Ho Chi Minh	Vietnam	33	225	970	9673	>10,000
Taipei	Taiwan	2	4	8	36	2200
Tokyo	Japan	4	7	23	556	>10,000

**Table 1.** Return periods (yr) of Category 1–5 wind speeds occurring within 100 km for 18 coastal cities. Return periods are derived from the STORM dataset. Wind speeds are given as 10-meter 10-minute sustained average values, see Methods.

**Spatial distribution of return periods of tropical cyclones.** Besides calculating max U10 for specific RPs, we can also reverse the procedure and compute the RP for given TC-categories. Figure 4 illustrates the RPs of a Category-1 (max U10 ≥ 29 m/s) and a Category-3 TC (max U10 ≥ 43.4 m/s) on the converted Saffir-Simpson Scale<sup>26</sup>, see Methods. There are large spatial variations, but for all basins RPs are lowest for a Category-1 or Category-3 event in the tropical regions. For large parts of the Western Pacific and the eastern part of the Eastern

Basin name	Basin domain
Eastern Pacific	5°–60°N 180°–coastline of North America on the North Atlantic
North Atlantic	5°–60°N coastline of North America on the Eastern Pacific - 360°
North Indian	5°–60°N 30°–100°E
South Indian	5–60°S 10°–135°E
South Pacific	5–60°S 135°–240°E
Western Pacific	5–60°N 100°–180°E

**Table 2.** Basin definitions, as adapted from Bloemendaal, *et al.*<sup>12</sup>.

Category	Wind speed threshold	
	1-minute average sustained	10-minute average sustained
Category 1	64–82 kt	29.0–37.1 m/s
	32.9–42.2 m/s	
Category 2	83–95 kt	37.6–43.0 m/s
	42.7–48.9 m/s	
Category 3	96–112 kt	43.4–50.7 m/s
	49.3–57.6 m/s	
Category 4	113–136 kt	51.1–61.6 m/s
	58.1–70 m/s	
Category 5	≥137 kt	≥61.6 m/s
	≥70 m/s	

**Table 3.** Tropical cyclone wind speed categorization on the Saffir-Simpson Scale. Conversion between 1-minute average sustained wind speeds in knots and in m/s is done using a factor of 0.5144, conversion between 1-minute and 10-minute average sustained wind speeds (in m/s) is done using a conversion factor of 0.88<sup>21</sup>.

Pacific, Category-1 TCs have an approximate 2-yr RP. In the other basins, these RPs lie between 5 and 20-yr. Category-3 TCs, however, occur less frequent, ranging between 10-yr RP for the Western Pacific to 90-yr RP for the North Indian.

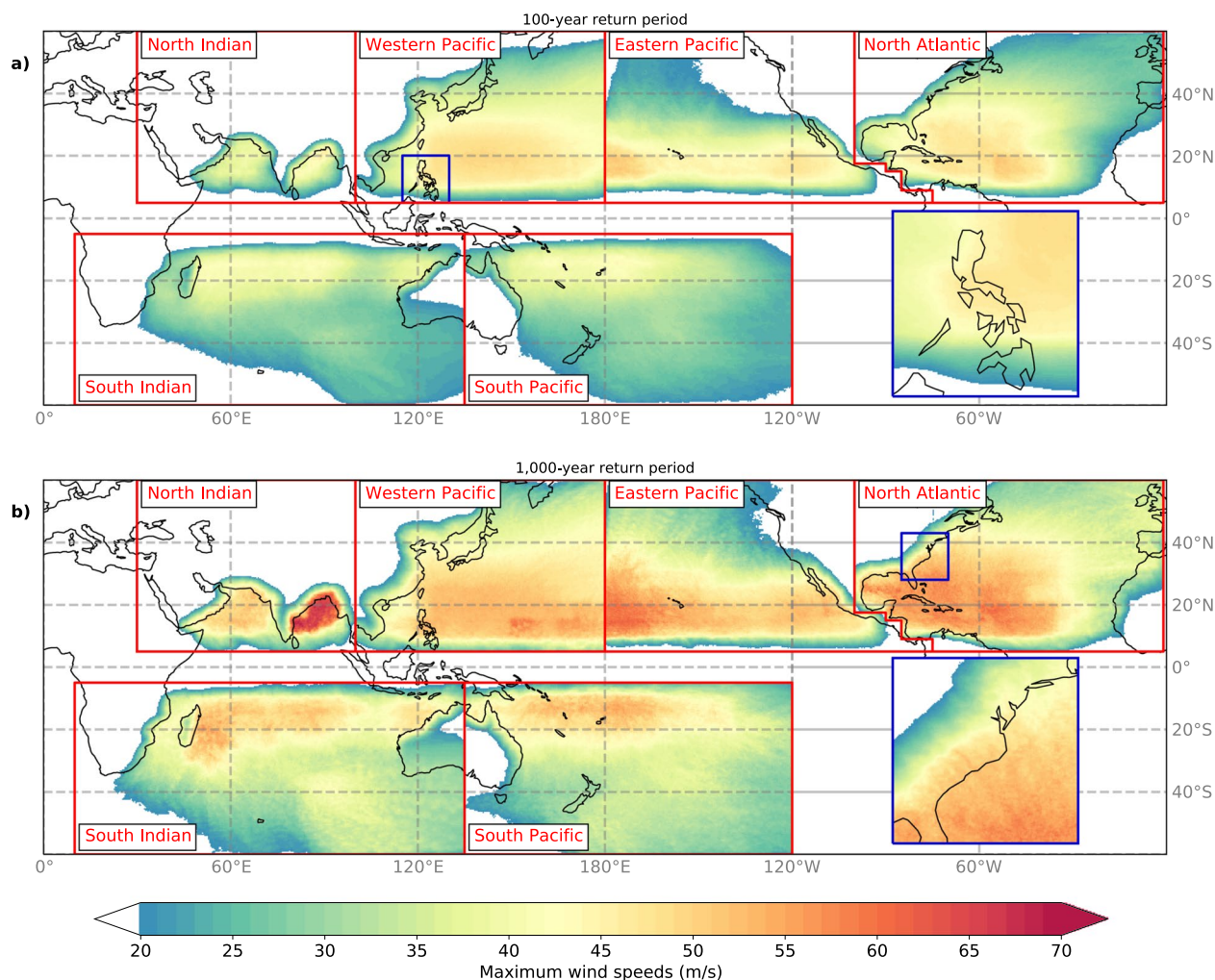
## Discussion

**Comparison to other synthetic datasets.** At basin-level, we have shown that the STORM-RPs compare well with observations (Fig. 1). At local scale, however, data can be scarce. Additionally, the observational dataset used here only spans 38 years, making it unfit for RP analysis past this timespan. Hence, here we compare our results to other studies that derived RPs based on thousands of years of synthetic TC tracks. We first compare model outcomes at the local scale, after which we discuss global-scale patterns in RPs.

For 18 cases we compare max U10 for given RPs with STORM, and for Mumbai we compare the RPs for given categories (see Online-only Table 1). Sobel, *et al.*<sup>24</sup> reported a 49 to 97-yr RP for a Category-1 TC within 150 km from Mumbai, agreeing well with the STORM-RP of 66-yr. Similarly, they estimate Category-3 RPs around 500-yr, while STORM-RPs are around 138-yr. For a Category-1 TC occurring in Mumbai, Sobel, *et al.*<sup>24</sup> list a 224 to 236-yr RP; compared to a STORM-RP of 95-yr. For the city itself, RPs vary between 3,000 to 10,000-yr, compared to 550-yr in STORM. STORM-RPs are predominantly lower than those in Sobel, *et al.*<sup>24</sup>. Sobel, *et al.*<sup>24</sup>, however, speculate that they underestimate the TC hazard. Further differences are likely driven by the use of different track modeling methods, wind field parameterization schemes (including different resolutions) and RP estimation techniques.

STORM performs well for given RPs: for 13 out of 18 cases, absolute differences between max U10 in STORM and other models lie within 5 m/s. The largest difference is 18 m/s for Darwin, Australia<sup>27</sup>. This relatively large difference (–38.3%) is likely caused by the fact that Darwin is located near the South Indian-South Pacific basin border, and is hit by TCs originating in both basins. STORM, however, models TCs per basin, cutting off South Pacific TCs at the basin boundary. Townsville and Port Hedland are located further away from the basin boundary and max U10 are in better agreement. Additionally, we observe relative differences of –35.9% and –25.1% for the Federated States of Micronesia (FSM) and Palau, respectively. In STORM, the lower basin boundary in the Western Pacific is set at 5°N. As these island countries lie at around 7°N, and modelled TCs generally deflect away from the Equator, this means most TCs in STORM pass north of the island countries. Conversely, the AIR Tropical Cyclone Model<sup>28</sup> models the Pacific region as a whole, and thus TCs have a higher probability of affecting FSM and Palau.

At the global scale, Lee, *et al.*<sup>29</sup> calculated RPs for Category-1 and 3 TCs using the CHAZ model. The general spatial patterns in the CHAZ model (Figure 12–13 in Lee, *et al.*<sup>29</sup>) and STORM (Fig. 4) agree well, particularly in the Western and Eastern Pacific, including low RPs (around 1 to 10-yr for Category-1 and in the order of 10-yr for Category 3) in the Western and Eastern Pacific. For the North Atlantic and North Indian, STORM-RPs are lower than CHAZ-RPs. However, Lee, *et al.*<sup>29</sup> illustrate that their RPs are higher in the North Atlantic than observations. In parts of the Southern basins, STORM-RPs are around 90-yr while CHAZ-RPs are approximately 10 to 50-yr for a Category-3.



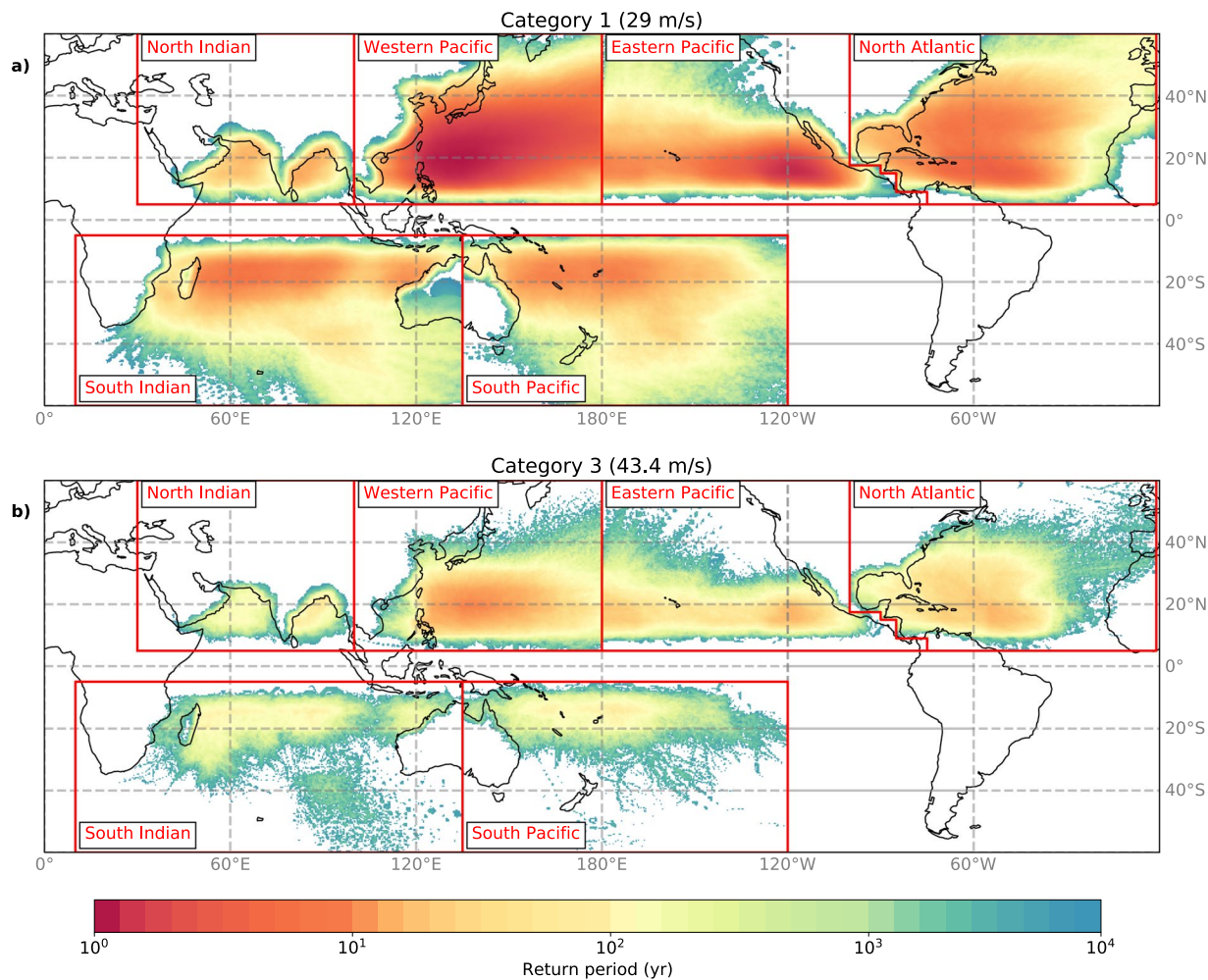
**Fig. 3** Spatial distribution of 10-meter 10-minute sustained maximum wind speeds (m/s) at 10 km resolution, derived from applying a 2D-wind parametrization to the synthetic tropical cyclone tracks in STORM. The wind speeds are the average value of 1,000 random realizations of 10,000 years of data (sampled with replacement) and determining RPs using Weibull's plotting formula to each realization, performed at each coordinate at the 1-in-100-year (a) and the 1-in-1,000-year return period (b), respectively. The return period-analysis is carried out at the basin scale: as such, there is often no smooth transition of wind speed values at the basin boundaries. Inset figures show the distribution of wind speeds around the Philippines (a) and the United States East Coast (b) at the given return period.

In conclusion, the STORM-RPs show good agreement with other studies, with differences in max U10 for a given RP often being less than 5 m/s. At the global scale, we observe similar spatial patterns of RPs for a Category-1 TC, but deviations at smaller scales occur when assessing Category-3 RPs.

**Limitations and future research directions.** In previous sections, we have demonstrated that the STORM-RPs for max U10 perform well. There are, however, some limitations regarding the usage of this dataset, which we will briefly reflect upon here, as well as giving directions for future research.

First, the STORM dataset is based on average present-day climate conditions (1980–2017), and as such does not capture climate variability. The climatology represented by STORM may be biased by the phases of multi-decadal variability contained in the 38-year period of record that was used to generate the dataset, which may not be representative for longer timescales. Moreover, the STORM dataset cannot be used to assess climate trends on decadal timescales or the effects of climate oscillations on TCs (e.g. the El Niño Southern Oscillation or the Madden-Julian Oscillation<sup>30</sup>). Future research could study these aspects by e.g. using ensemble runs, or by generating synthetic TCs per oscillation phase.

Second, we used an easily applicable empirical inland decay function<sup>31</sup> to model the decay of TC wind speeds after landfall, and combine this with a 2D-parametric wind field model<sup>32,33</sup>. This decay function was derived using USA landfalling events, and as such may perform less well elsewhere. Moreover, the function assumes that a TC starts to decay after the TC eye crosses land. In reality, the inland surface winds will decay prior to landfall in response to enhanced surface friction caused by the land mass<sup>34</sup>. The 2D-parametric model does not include the influence of land, and therefore inland wind speeds may be overestimated. The use of a numerical boundary



**Fig. 4** Spatial distribution of return periods (yr) at 10 km resolution, derived from applying a 2D-wind parametrization to the synthetic tropical cyclone tracks in STORM. The return periods are the average value of 1,000 random realizations of 10,000 years of data (sampled with replacement) and determining RPs using Weibull's plotting formula to each realization at each coordinate at Category 1 (wind speeds  $\geq 29$  m/s) (a), and Category 3 tropical cyclone strength (wind speeds  $\geq 43.4$  m/s) (b), respectively.

layer model which includes the effects of terrain on the wind field would result in a better representation of the temporal evolution of the TC wind field over land<sup>35</sup>.

Last, the 2D-parametric wind field model used here assumes the asymmetry in the TC wind field arises from background flow. In extratropical regions, however, enhanced wind shear, caused by large-scale background flows or nearby troughs<sup>36</sup>, may also induce asymmetry. Furthermore, STORM does not model the extratropical transition of TCs, so systems in these regions may be represented incorrectly and end-users should therefore pay attention when using this regional data.

**Concluding remarks.** We have demonstrated the application of the STORM dataset to generate a novel, open-access dataset of wind speed RPs for all TC basins. We empirically derived RPs at three spatial levels: at basin-level, within 100 km of selected coastal locations, and at 10 km resolution. First, we demonstrated the benefit of using such large synthetic dataset, composed of 10,000 years of TC activity for present-day climate conditions, over using a climatological dataset of 38 years for the calculation of RPs. Compared to the empirically derived RPs, fitting a continuous EV distribution to 38 years of data typically leads to higher max U10, especially for RPs exceeding 100-yr. Second, we calculated RPs for TCs within 100 km of 18 coastal cities, and found that RP-curves differ substantially between locations. Finally, we estimated RPs at 10 km resolution by applying a 2D-wind field model to the synthetic TCs. This dataset is applicable for high-resolution TC wind risk assessments, particularly at the local scale such as Pacific island countries or the Caribbean.

To assess our model performance, we compared the STORM-RPs against those derived from other synthetic models across different spatial scales, and found that results generally agree well. Near basin boundaries, however, RPs can be lower compared to other literature. This is because STORM is run at the basin-scale, and TCs are cut off at the basin boundaries. For regional-scale studies, this issue can be solved by applying STORM using adjusted basin boundaries. On the global scale, we observe similar spatial RP patterns for a Category-1 event compared to literature, but larger spatial differences arise when looking at the distribution of Category-3 RPs in the Southern basins<sup>30</sup>.

In conclusion, this study is unique in that it is the first to estimate (high) RPs at 10 km resolution on a global scale. It represents an important step forward in global TC wind risk assessments, particularly for island countries and TC-scarce regions. Furthermore, this research can contribute to an improved quantification of other TC-induced hazards such as storm surge and precipitation<sup>4–6</sup>. To estimate the RPs of TC-induced storm surges, the 2D-wind fields can be used to force a hydrodynamic model<sup>37</sup>. TC precipitation fields are closely related to max U10<sup>4</sup> fields and the distance from the eye<sup>35</sup>. These properties can be used to construct a parametric 2D-precipitation field model similar to the parametric wind field model, to assess TC precipitation risk.

## Methods

Our approach is based on estimating the empirical RPs on basis of the synthetic TCs derived from the STORM dataset<sup>12</sup>. This dataset is created using historical data from the International Best Track Archive for Climate Stewardship (IBTrACS<sup>20</sup>) as input dataset for the Synthetic Tropical cyclOne geneRation Model (STORM). The development of this dataset has been described in detail in Bloemendaal, *et al.*<sup>12</sup>. Here, we provide a brief description of the STORM dataset, but for more details on the methodology and validation we direct readers to Bloemendaal, *et al.*<sup>12</sup>.

**Data.** The STORM dataset, a global synthetic dataset comprised of 10,000 years of synthetic TCs under present-climate conditions, is used for the calculation of the return periods (RPs). The STORM dataset was generated using STORM. This model takes the following IBTrACS data as input: the latitudinal and longitudinal position of the TC, maximum 10-meter 10-minute average sustained wind speeds (max U10), mean sea-level pressure (MSLP), and the size of the TC eye (Radius to maximum winds; Rmax). Averaged environmental conditions are modeled using monthly-mean sea-surface temperatures (SST) and MSLP fields from ERA5.<sup>38</sup> From this, autoregressive formulas model consecutive changes in the longitudinal/latitudinal position of the TC (in °), the minimum pressure (in hPa) and the maximum wind speed (in m/s) at every time step during a TC's lifetime. In addition, STORM also simulates the size of the TC eye, represented via the radius to maximum winds (in km).

STORM is validated in Bloemendaal, *et al.*<sup>12</sup>. Results show that STORM preserves the TC statistics as found in the IBTrACS input dataset, which indicates a good model performance. The average number of both genesis and landfalling events in the STORM dataset, as well as landfall intensity was found to closely correspond (within one standard deviation) to those in the IBTrACS dataset. The largest deviations in max U10 along a TC track were found to be approximately 2 m/s in the STORM dataset compared to the IBTrACS dataset.

**Estimation of return periods and 2D wind field parameterization.** Using the 10,000 years of TC activity in the STORM dataset, we empirically calculate the max U10 for different RPs using Weibull's plotting formula<sup>16</sup>, see Equation 1. The Weibull plotting formula has been demonstrated to be the best performing empirical formula for the estimation of return periods<sup>39</sup>.

$$P_{exc}(\vec{v}) = \frac{i}{n+1} \cdot \frac{n}{m} \quad (1a)$$

$$T(\vec{v}) = 1/P_{exc}(\vec{v}) \quad (1b)$$

Here,  $P_{exc}(\vec{v})$  represents the exceedance probability  $P_{exc}$  for a given maximum wind speed  $\vec{v}$  at rank  $i$ .  $n$  is the total number of events in the set, and  $m$  the total length of the dataset (in years; here,  $m = 10,000$ ). The return period  $T(\vec{v})$  is then given as the inverse of  $P_{exc}$ . Because the STORM dataset represents 10,000 years of TC activity, we empirically calculate RPs up to 10,000 years.

To demonstrate the performance of the empirical distribution compared to extreme value distributions (see Results), we fit five continuous extreme value distributions to the STORM dataset. These distributions include the Generalized Extreme Value, Exponential, Gumbel, Weibull, and Pareto distribution. Estimation of the optimal parameters for each of the distributions was done using Python's *lmoments* package<sup>40</sup> at 0.1-year intervals, up to 1,000 years. For the estimations of RPs at basin scale and within a 100 km radius, we apply Equation 1 directly to max U10 values in the STORM dataset. However, for assessing RPs of wind speeds at 10 km resolution, we need to convert the TC tracks, that includes the longitudinal/latitudinal position of the TC eye, maximum wind speed, and radius to maximum winds, to 2D-wind fields. For this, we follow the parametric approach of Holland<sup>32</sup>. We follow the same approach as was suggested by Lin and Chavas<sup>33</sup> and Muis, *et al.*<sup>41</sup>. The Holland B parameter is calculated following Lin and Chavas<sup>33</sup>. The surface wind is converted to a gradient wind using a wind speed reduction factor of 1/0.85<sup>42</sup>. The asymmetry in the surface wind is accounted for by adding the surface background winds to the wind field<sup>33</sup>, which is approximated by the translational speed of the TC. We compute the 2D-wind field using a mesh with 10 km resolution. To optimize computational costs, this calculation is done in parallel using a separate mesh per basin. For each synthetic TC, we store the max U10 at each grid cell whenever max U10  $\geq 20$  m/s. Lastly, we apply Equation 1 to the max U10 values at every grid cell to estimate the RPs.

**Basin definitions.** The basin definitions used in this paper are adapted from Bloemendaal, *et al.*<sup>12</sup>, see Table 2.

**Saffir-Simpson scale.** We use the Saffir-Simpson Scale<sup>26</sup> as an additional metric to communicate about wind speeds. The categorization on this scale, however, is done using 1-minute average sustained wind speeds, whereas the STORM wind data is given as a 10-minute average value. For this reason, we convert the 1-minute threshold values on the Saffir-Simpson scale to a 10-minute equivalent using a conversion factor of 0.88<sup>21</sup>, see Table 3.



Parameters		
Name	Unit	Notes
Longitude (lon)	Degrees east	Ranges from 0–360°, with prime meridian at Greenwich
Latitude (lat)	Degrees north	
Return period (rp)	Year	Given at 10, 20, 30 ... 100, 200, 300 ... 1,000, 2,000, 3,000 ... 10,000-yr return period
Variables		
Name (dependency)	Unit	Notes
Mean (lat, lon, rp)	m/s	Mean wind speed of 1,000 random realizations, at each return period
Standard deviation (lat, lon, rp)	m/s	Standard deviation in wind speeds of 1,000 random realizations, at each return period
5% confidence interval (lat, lon, rp)	m/s	5% confidence interval value derived from 1,000 random realizations, at each return period
95% confidence interval (lat, lon, rp)	m/s	95% confidence interval value derived from 1,000 random realizations, at each return period

**Table 4.** Overview of entries in the “fixed return periods” dataset. Wind speeds are given as 10-meter 10-minute sustained average values.

Parameters		
Name	Unit	Notes
Longitude (lon)	Degrees east	Ranges from 0–360°, with prime meridian at Greenwich.
Latitude (lat)	Degrees north	
Wind speed (u)	m/s	Given at 20, 25, 30.. 75 m/s, and at 29, 37.6, 43, 50.7, and 61.6 m/s to comply with the category thresholds on the Saffir-Simpson scale <sup>26</sup> (see Methods)
Variables		
Name (dependency)	Unit	Notes
Mean (lat, lon, u)	year	Mean return period of 1,000 random realizations, at each wind speed
Standard deviation (lat, lon, u)	year	Standard deviation in return periods of 1,000 random realizations, at each wind speed
5% confidence interval (lat, lon, rp)	year	5% confidence interval value derived from 1,000 random realizations, at each wind speed
95% confidence interval (lat, lon, rp)	year	95% confidence interval value derived from 1,000 random realizations, at each wind speed

**Table 5.** Overview of entries in the “fixed wind speeds” dataset. Wind speeds are given as 10-meter 10-minute sustained average values.

Entry	Column name	Notes
1	City	Name of city
2	Country	Name of country
3	Latitude	Latitudinal coordinates of capital city as given in Google Maps ( <a href="https://maps.google.com">maps.google.com</a> ).
4	Longitude	Longitudinal coordinates of capital city as given in Google Maps ( <a href="https://maps.google.com">maps.google.com</a> ), ranging from 0–360° with prime meridian at Greenwich.
5–31	Return period	On “Fixed return periods”-tab: maximum wind speed (m/s) at given return period
5–20	Wind speed	On “Fixed wind speeds”-tab: return period (yr) at given wind speed. Wind speeds are given at 5 m/s intervals and at values corresponding to Category thresholds on the Saffir-Simpson Scale, see Methods.

**Table 6.** Overview of entries in the “cities” dataset. Wind speeds are given as 10-meter 10-minute sustained average values.

## Data availability

The STORM dataset was developed in Bloemendaal, *et al.*<sup>12</sup> and is publicly accessible via the 4TU.Centre for Research Data repository<sup>43</sup>.

This paper is accompanied by four different datasets:

- (i) The “fixed return periods” dataset, with wind speed estimates (m/s) for a predefined range of return periods for every longitude/latitude position of the 10 km grid per basin (demonstrated in Fig. 3, see Table 4);
- (ii) The “fixed wind speeds” dataset, with return period estimates (yr) for a predefined range of wind speeds for every longitude/latitude position of the 10 km grid per basin (demonstrated in Fig. 4, see Table 5);
- (iii) The “cities” dataset, with return periods (yr) for a predefined range of wind speeds (m/s) and wind speeds for a predefined range of return periods, occurring with 100 km of the respective city (demonstrated in Fig. 2; see Table 6).

Entry	Column name	Notes
1	City	Name of capital city
2	Country	Name of Island Country
3	Latitude	Latitudinal coordinates of capital city as given in Google Maps ( <a href="https://maps.google.com">maps.google.com</a> ).
4	Longitude	Longitudinal coordinates of capital city as given in Google Maps ( <a href="https://maps.google.com">maps.google.com</a> ), ranging from 0–360° with prime meridian at Greenwich.
5–31	Return period	On “Fixed return periods”-tab: maximum wind speed (m/s) at given return period
5–20	Wind speed	On “Fixed wind speeds”-tab: return period (yr) at given wind speed. Wind speeds are given at 5 m/s intervals and at values corresponding to Category thresholds on the Saffir-Simpson Scale, see Methods.

**Table 7.** Overview of entries in the “islands” dataset. Wind speeds are given as 10-meter 10-minute sustained average values.

(iv) The “island” dataset, with return periods (yr) for a predefined range of wind speeds (m/s) and wind speeds for a predefined range of return periods, occurring within 100 km of the capital city of the respective island. We included the Small Island Developing States and a set of other islands in this dataset (63 islands in total; see Table 7). This dataset has not been demonstrated in the paper, but application is similar to the “cities” dataset.

Datasets (i) and (ii) are netCDF4 (.nc)-files, dataset (iii) and (iv) are excel-files (.xlsx). Each of these datasets can be found at the 4TU.Centre for Research Data repository<sup>44–47</sup>.

### Code availability

All Python algorithms used to generate the datasets described here are publicly available via <https://github.com/NBloemendaal/STORM-return-periods>. The STORM Python script can be found on <https://github.com/NBloemendaal/STORM> and the data-preprocessing is available via <https://github.com/NBloemendaal/STORM-preprocessing>.

Received: 2 June 2020; Accepted: 14 October 2020;

Published online: 10 November 2020

### References

1. The Guardian. *Cyclone Idai caused \$2bn of damage and affected millions, says World Bank*, (2019).
2. NOAA. *Costliest U.S. Tropical Cyclones*, <https://www.ncdc.noaa.gov/billions/dcmi.pdf> (2020).
3. Kruk, M. C., Gibney, E. J., Levinson, D. H. & Squires, M. A. Climatology of Inland Winds from Tropical Cyclones for the Eastern United States. *J Appl Meteorol Climatol* **49**, 1538–1547, <https://doi.org/10.1175/2010jamc2389.1> (2010).
4. Cervený, R. S. & E. Newman, L. Climatological Relationships between Tropical Cyclones and Rainfall. *Mon Weather Rev* **128**, [https://doi.org/10.1175/1520-0493\(2000\)128<3329:CRBTCA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3329:CRBTCA>2.0.CO;2) (2000).
5. Phadke, A. C., Martino, C. D., Cheung, K. F. & Houston, S. H. Modeling of tropical cyclone winds and waves for emergency management. *Ocean Eng* **30**, 553–578, [https://doi.org/10.1016/S0029-8018\(02\)00033-1](https://doi.org/10.1016/S0029-8018(02)00033-1) (2003).
6. Bloemendaal, N. *et al.* Global modeling of tropical cyclone storm surges using high-resolution forecasts. *Clim Dynam*, <https://doi.org/10.1007/s00382-018-4430-x> (2018).
7. Ward, P. J., De Moel, H. & Aerts, J. C. J. H. How are flood risk estimates affected by the choice of return-periods? *Nat. Hazards Earth Syst. Sci.* **11**, 3181–3195, <https://doi.org/10.5194/nhess-11-3181-2011> (2011).
8. Simpson, R. H. & Lawrence, M. B. Atlantic hurricane frequencies along the US coastline. (National Weather Service, 1971).
9. Emanuel, K., Ravela, S., Vivant, E. & Risi, C. A Statistical Deterministic Approach to Hurricane Risk Assessment. *Bull. Am. Meteor. Soc.* **87**, 299–314, <https://doi.org/10.1175/BAMS-87-3-299> (2006).
10. Vickery, P. J., Skerlj, P. F. & Twisdale, L. A. Simulation of Hurricane Risk in the U.S. Using Empirical Track Model. *J. Struct. Eng.* **126**, 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)) (2000).
11. James, M. K. & Mason, L. B. Synthetic Tropical Cyclone. *Database. J Waterw Port C Div* **131**, 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)) (2005).
12. Bloemendaal, N. *et al.* Generation of a global synthetic tropical cyclone hazard dataset using STORM. *Sci. Data* **7**, 40, <https://doi.org/10.1038/s41597-020-0381-2> (2020).
13. Haigh, I. D. *et al.* Estimating present day extreme water level exceedance probabilities around the coastline of Australia: tropical cyclone-induced storm surges. *Clim Dynam* **42**, 139–157, <https://doi.org/10.1007/s00382-012-1653-0> (2014).
14. Lin, N., Emanuel, K., Oppenheimer, M. & Vanmarcke, E. Physically based assessment of hurricane surge threat under climate change. *Nat. Clim. Change* **2**, 462–467, <https://doi.org/10.1038/nclimate1389> (2012).
15. Hazen, A. Flood flows: a study of frequencies and magnitudes in *Flood flows: a study of frequencies and magnitudes* (John Wiley & Sons, 1930).
16. Weibull, W. *A statistical theory of the strength of materials*. (Generalstabens litografiska anstalts förlag, 1939).
17. Kotz, S. & Nadarajah, S. *Extreme Value Distributions: Theory and Applications*. (Imperial College Press, 2000).
18. Esteves, L. S. Consequences to flood management of using different probability distributions to estimate extreme rainfall. *J. Environ. Manage* **115**, 98–105, <https://doi.org/10.1016/j.jenvman.2012.11.013> (2013).
19. Wahl, T. *et al.* Understanding extreme sea levels for broad-scale coastal impact and adaptation analysis. *Nat Comm* **8**, 16075, <https://doi.org/10.1038/ncomms16075> (2017).
20. Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J. & Neumann, C. J. The International Best Track Archive for Climate Stewardship (IBTrACS) Unifying Tropical Cyclone Data. *Bull. Am. Meteor. Soc.* **91**, 363–376, <https://doi.org/10.1175/2009BAMS2755.1> (2010).
21. Harper, B. A., Kepert, J. D. & Ginger, J. D. Guidelines for converting between various wind averaging periods in tropical cyclone conditions. (World Meteorological Organization, 2008).
22. Holland, G. J. The Maximum Potential Intensity of Tropical Cyclones. *J. Atmos. Sci.* **54**, 2519–2541 [https://doi.org/10.1175/1520-0469\(1997\)054<2519:tmplot>2.0.co;2](https://doi.org/10.1175/1520-0469(1997)054<2519:tmplot>2.0.co;2) (1997).
23. Emanuel, K. A. The dependence of hurricane intensity on climate. *Nature* **326**, 483–485, <https://doi.org/10.1038/326483a0> (1987).
24. Sobel, A. H. *et al.* Tropical cyclone hazard to Mumbai in the recent historical climate. *Mon Weather Rev* **147**, 2355–2366, <https://doi.org/10.1175/mwr-d-18-0419.1> (2019).

25. Hersbach, H. *et al.* Global Reanalysis: goodbye ERA-Interim, hello ERA5. 17–24 (2019).
26. Simpson, R. H. & Saffir, H. The hurricane disaster-potential scale. *Weatherwise* **27**, 169–186, <https://doi.org/10.1080/00431672.1974.49931702> (1974).
27. Cook, G. D. & Nicholls, M. J. Estimation of Tropical Cyclone Wind Hazard for Darwin: Comparison with Two Other Locations and the Australian Wind-Loading Code. *J Appl Meteorol Climatol* **48**, 2331–2340, <https://doi.org/10.1175/2009JAMC2013.1> (2009).
28. Commonwealth of Australia. Current and future tropical cyclone risk in the South Pacific: South Pacific regional risk assessment. (Australian government, 2013).
29. Lee, C.-Y., Tippett, M. K., Sobel, A. H. & Camargo, S. J. An Environmentally Forced Tropical Cyclone Hazard Model. *JAMES* **10**, 223–241, <https://doi.org/10.1002/2017MS001186> (2018).
30. Camargo, S. J., Robertson, A. W., Barnston, A. G. & Ghil, M. Clustering of eastern North Pacific tropical cyclone tracks: ENSO and MJO effects. *Geochem Geophys* **9**, <https://doi.org/10.1029/2007gc001861> (2008).
31. Kaplan, J. & DeMaria, M. A Simple Empirical Model for Predicting the Decay of Tropical Cyclone Winds after Landfall. *J. Appl. Meteorol.* **34**, 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) (1995).
32. Holland, G. J. An Analytic Model of the Wind and Pressure Profiles in Hurricanes. *Mon Weather Rev* **108**, 1212–1218 [https://doi.org/10.1175/1520-0493\(1980\)108<1212:aamotw>2.0.co;2](https://doi.org/10.1175/1520-0493(1980)108<1212:aamotw>2.0.co;2) (1980).
33. Lin, N. & Chavas, D. On hurricane parametric wind and applications in storm surge modeling. *J Geophys Res - Atmos* **117**, n/a–n/a, <https://doi.org/10.1029/2011JD017126> (2012).
34. Done, J. M. *et al.* Modelling global tropical cyclone wind footprints. *Nat. Hazards Earth Syst. Sci.* **20**, 567–580, <https://doi.org/10.5194/nhess-20-567-2020> (2020).
35. Yu, Z. & Wang, Y. Rainfall Distribution in Landfalling Tropical Cyclones in *Extreme Weather* (IntechOpen, 2018).
36. Ritchie, E. A. & Elsberry, R. L. Simulations of the Transformation Stage of the Extratropical Transition of Tropical Cyclones. *Mon Weather Rev* **129**, 1462–1480, [https://doi.org/10.1175/1520-0493\(2001\)129<1462:Sottso>2.0.Co;2](https://doi.org/10.1175/1520-0493(2001)129<1462:Sottso>2.0.Co;2) (2001).
37. Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C. J. H. & Ward, P. J. A global reanalysis of storm surges and extreme sea levels. *Nat Comm* **7**, <https://doi.org/10.1038/ncomms11969> (2016).
38. Hersbach, H. & Dee, D. ERA5 reanalysis is in production. *ECMWF Newsletter* (2016).
39. Makkonen, L. Plotting positions in extreme value analysis. *J Appl Meteorol Climatol* **45**, 334–340 (2006).
40. Python Software Foundation. *lmoments* 0.2.3, <https://pypi.org/project/lmoments/> (2020).
41. Muis, S. *et al.* Spatiotemporal patterns of extreme sea levels along the western North-Atlantic coasts. *Sci Rep*, <https://doi.org/10.1038/s41598-019-40157-w> (2019).
42. Powell, M. *et al.* State of Florida hurricane loss projection model: Atmospheric science component. *J Wind Eng Ind Aerod* **93**, 651–674, <https://doi.org/10.1016/j.jweia.2005.05.008> (2005).
43. Bloemendaal, N. *et al.* STORM IBTrACS present climate synthetic tropical cyclone tracks. *4TU.Centre for Research Data*, <https://doi.org/10.4121/uuid:82c1dc0d-5485-43d8-901a-ce7f26cda35d> (2019).
44. Bloemendaal, N., De Moel, H., Muis, S., Haigh, I. D. & Aerts, J. C. J. H. STORM tropical cyclone wind speed return periods. *4TU.Centre for Research Data*, <https://doi.org/10.4121/uuid:779b9dfd-b0ff-4531-8833-aaa9c0cf6b5a> (2020).
45. Malmstadt, J. C., Elsner, J. B. & Jagger, T. H. Risk of Strong Hurricane Winds to Florida Cities. *J Appl Meteorol Climatol* **49**, 2121–2132, <https://doi.org/10.1175/2010JAMC2420.1> (2010).
46. Ellis, K. N., Trepanier, J. C. & Hodges, R. E. Using Synthetic Tropical Cyclones to Characterize Extreme Hurricanes Affecting Charleston, South Carolina. *J Appl Meteorol Climatol* **55**, 883–892, <https://doi.org/10.1175/JAMC-D-15-0215.1> (2016).
47. Garner, A. J. *et al.* Impact of climate change on New York City's coastal flood hazard: Increasing flood heights from the preindustrial to 2300 CE. *PNAS* **114**, 11861, <https://doi.org/10.1073/pnas.1703568114> (2017).

## Acknowledgements

We thank Job Dullaart and Anaïs Couasnon for their help in the development and verification of the methodology used in this study. We also acknowledge SURFsara ([www.surf.nl](http://www.surf.nl)) for the support in using the Lisa Computer Cluster. NB and JCJHA are funded by a VICI grant from the Netherlands Organization for Scientific Research (NWO) (Grant Number 453-13-006) and the ERC Advanced Grant COASTMOVE #884442. IDH was funded by NERC Grant CompFlood (Grant Number NE/S003150/1). SM received funding from the research programme MOSAIC with project number ASDI.2018.036, which is financed by the NWO.

## Author contributions

All authors designed the initial analyses ideas. N.B., H.d.M., and S.M. developed the methodology and analyzed the data model output. N.B. drafted the initial version of the manuscript. All authors were actively involved in the interpretation of the analysis results and the writing process.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to N.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020