A Deep-Learning Reconstruction of Tropical Cyclone Size Metrics 1981–2017: Examining Trends

JING-YI ZHUO[a,b] AND ZHE-MIN TAN[a,b]

[a] *Key Laboratory of Mesoscale Severe Weather, Ministry of Education, Nanjing University, Nanjing, China*
[b] *School of Atmospheric Sciences, Nanjing University, Nanjing, China*

ABSTRACT: Observed climate records of length, homogeneity, and reliability are the basis of climatological studies on tropical cyclones (TCs). However, such data are scarce for TC size in terms of wind field, particularly over the western North Pacific (WNP). This study demonstrates that deep learning can practically bridge this data gap when applied to satellite data. Using transfer learning, deep learning algorithms were developed to estimate reliable TC sizes from infrared imagery for the WNP TCs. The algorithms were then applied to a homogeneous satellite database to reconstruct a new historical dataset of TC sizes, named DeepTCSize, which covers 37 years (1981–2017) over the WNP. DeepTCSize includes multiple TC size quantities, such as wind radii of 17, 26, and 33 m s^{-1} and maximum winds (i.e., R17, R26, R33, and RMW), which have high correlations (R = 0.85, 0.84, 0.79, and 0.76, respectively) with postseason quality-controlled best track data. Comparisons with ocean wind observations were made and this further revealed that DeepTCSize has good quality and is free from spurious error trends, providing an advantage over the historical "best estimates" of TC sizes currently available in the best track archives for the WNP. The new reconstructed TC sizes dataset for the WNP TCs shows significant expanding trends in the annual-mean outer circulations (at a rate of 2% decade^{-1} for R17 and a rate of 2% decade^{-1} for R26), which are mainly associated with weaker storms, as well as a weak contracting trend in the annual-mean inner-core size (RMW).

SIGNIFICANCE STATEMENT: Tropical cyclone (TC) size largely controls the TC-induced hazard and risk. If the size of TC can be determined more efficiently in observations spanning a long-enough period, the climatology and changes in TC can be better modeled and understood. This study applies deep learning methods to reconstruct a new dataset of multiple inner- to outer-core TC size metrics from infrared imagery of satellites for the western North Pacific TCs. The dataset spans 37 years. It is homogenous and has comparable accuracy with the existing "best estimates." Using the dataset, a significant expanding trend was identified in the outer-core size, while the inner-core size exhibits a weak contracting trend. The dataset can be employed in several applications.

KEYWORDS: Tropical cyclones; Climate records; Deep learning

---

1. Introduction

The question of whether climate change influences tropical cyclone (TC) activity has attracted much attention due to its significant potential societal impacts. TC intensity, especially the proportion of the strongest TCs, has been observed via its trends (Holland and Bruyère 2014; Kossin et al. 2013, 2020). These observations show that the average latitude of the lifetime-maximum intensity of TCs has migrated poleward globally since the 1980s (Kossin et al. 2014). These changes can add to TC hazard exposure and mortality risk (Kossin et al. 2016). Even though TC sizes, defined as the radii of winds, often determine the TC destructive potential (Powell and Reinhold 2007) and storm surge risk (Irish et al. 2008), few studies have examined their changes, and no observed anthropogenic influences on TC size have been reported to date (Knutson et al. 2020).

Multiple definitions have been attributed to the term "TC size." According to various physical conditions, the complete radial structure of a TC is frequently thought to be composed of inner-core regions and outer circulations in a leading order (e.g., Frank 1977; Weatherford and Gray 1988). Most studies define the TC size as the areal extension of a TC's outer circulations. However, the radius of maximum wind (RMW) located in the turbulent inner-core region is another critical size metric that is widely adopted (e.g., Irish and Resio 2010). Due to its destructive potential, forecasters operationally pay special attention to the radial extent of 17, 26, and 33 m s^{-1} surface winds (R17, R26, and R33), as well as the RMW. Several studies also refer to TC size in terms of the rainfall area of TCs (Lin et al. 2015; Guzman and Jiang 2021).

Theoretical models that adequately explain TC in nature also aid in the prediction and comprehension of emerging changes in TC activity. For example, the potential intensity theory (Emanuel 1986, 2000), which formulates an upper bound of a TC's peak wind, suggests an increasing TC intensity with global warming (e.g., Sobel et al. 2016). The TC size, on the other hand, only has a limited physical basis. Chavas et al. (2015) developed a physics-based model for the radial structure of the TC tangential winds. Although the model is useful, it cannot be determined solely by environmental parameters, and its time-dependent wind-field variability is prompted by internal TC parameters such as the TC intensity. There could also

be a theoretical upper bound on TC size, but the existing theories (e.g., Khairoutdinov and Emanuel 2013; Chavas and Emanuel 2014; Lu and Chavas 2022) have not yet offered a comprehensive explanation for observational results (Chavas et al. 2016).

Regarding the theoretical limitations, it is important to investigate the physics of TC sizes based on observations. In previous studies, TC size climatology has been developed mostly based on QuikSCAT ocean winds (e.g., Chavas and Emanuel 2010; Lee et al. 2010; Chan and Chan 2012, 2015). In recent times, Chavas et al. (2016) comprehensively revisited TC size climatology using an updated version of the QuikSCAT database. However, the length of QuikSCAT records (1999–2009) was short and scarce [with less than 100 samples per year over each ocean basin, according to Chan and Chan (2015)] to support investigations of variations of TC sizes on the decadal time scale. To circumvent this issue, Knaff et al. (2014) created an infrared (IR)-based TC size climate record over 1978–2011, using 5-kt wind radii (R5; 1 kt $\approx$ 0.51 m s^{-1}) as the size metric. With the 34-yr database, interbasin trends were found (including a decrease in the eastern North Pacific and an increase in the western North Pacific), but none of these trends were statistically significant. The R5-based TC size climatology, on the other hand, differs considerably from the observed one. This is likely due to R5 being an indirect quantity derived empirically from the tangential wind at 500 km (V500), which is diagnosed from numerical models; and it only explains 30% variance of the observed R17 (Knaff et al. 2014). Another issue with the IR-based R5 dataset could be the use of a simple regression model with limitations. Several other studies used reanalysis datasets to generate a large number of long-term TC size records (e.g., Chan and Chan 2018; Schenkel et al. 2018). However, the TC wind profiles, including the TC intensities, were still found to significantly differ from those recorded in best track data (Schenkel and Hart 2012); hence, the findings based on reanalysis data need to be interpreted meticulously. Additionally, the inner-core sizes, such as RMW, were also poorly resolved in the reanalysis data (Schenkel et al. 2017; Bian et al. 2021). As such, improving observed records of TC sizes to supplement our understanding of TC size climatology becomes imperative.

Observational capabilities are one of the root causes for the lack of datasets for TC size that are compliant with the length and homogeneity standards of climate data. When aircraft data are not available (in most tropical regions outside the North Atlantic and the Caribbean Sea), satellite data have been the major source of TC information. The Dvorak technique has been the primary satellite method for estimating TC intensity for more than four decades, based on cloud pattern recognition from visible (VIS) and IR satellite images (Velden et al. 2006). However, there is no such technique for estimating equally reliable TC sizes from satellite imagery. Although several algorithms were developed to derive TC sizes from IR imagery (e.g., Mueller et al. 2006; Kossin et al. 2007; Knaff et al. 2014, 2016), the estimations tend to have large uncertainties (Landsea and Franklin 2013; Knaff et al. 2021). So, the operational estimation of TC sizes has primarily relied on scatterometers (Brennan et al. 2009). For example,

the much-cited NASA's Ku-band scatterometer QuikSCAT (Lungu and Callahan 2006) provided reliable ocean surface wind data for TC outer sizes around 1999–2009. Throughout the 2010s and into the early 2020s, developments in scatterometry (e.g., ASCAT; Figa-Saldaña et al. 2002), L-band radiometers (Reul et al. 2017), and synthetic aperture radar (Horstmann et al. 2015) have continued to complement TC size estimations (Knaff et al. 2021). More recent advances in wind retrieval algorithms have also enabled all-weather TC wind speeds such as those produced from multichannel radiometers WindSat, AMSR-E, and AMSR2 (Meissner et al. 2021). However, many of these satellite sensors still have difficulty in providing accurate observations of RMW for TCs, due to their limited resolutions and deteriorating capability to handle more complex TC inner-core conditions. When aircraft observations are unavailable, the SAR may be the most reliable source of RMW data (e.g., Mouche et al. 2017; Combot et al. 2020). A recent extensive review of TC size observations is presented by Knaff et al. (2021). The review suggested that the quality of the observed TC size records was found to be dependent on the method used and to vary over time and space. Therefore, it is difficult to conduct a direct reanalysis of these records to create a climate-quality dataset.

Global TC forecasting centers have collected most of the historical data of TC sizes over time, constituting what is known as the "best track" (e.g., Knapp et al. 2010; Landsea and Franklin 2013). Partly due to the difficulty in reanalyzing the TC sizes data from diverse sources, only small parts of the TC sizes data in the best tracks to date have undergone rigorous postseason reanalysis (Sampson et al. 2017; Knaff et al. 2021). In the western North Pacific (WNP), both the Joint Typhoon Warning Center (JTWC) and the WMO Regional Specialized Meteorological Centre in Tokyo (referred to as JMA in the remainder of this paper) maintain the best track record of TC sizes. However, significant interagency data discrepancies have been identified (Song and Klotzbach 2016; Kim et al. 2022), indicating that the best track TC sizes are also flawed. In addition to this, how the uncertainty in the best track datasets may result in large, distorted variations and false trends in TC sizes, also remains uninvestigated.

Given the above context, a typical research question will be, is it possible to reproduce TC sizes that meet the quality required for climate research? To achieve this, novel paradigms such as deep learning could offer a promising approach. The rapid advances in deep learning have substantially impacted many scientific fields, including climate science. Deep learning can be used to reconstruct missing climate information (e.g., Kadow et al. 2020), recognize driving climate patterns or extreme events (e.g., Racah et al. 2017; Barnes et al. 2020), climate attribution (e.g., Callaghan et al. 2021) and seasonal-to-decadal climate predictions (e.g., Ham et al. 2019). Recently, major progress has also been made in applying deep learning for TC-related studies. Several studies have demonstrated that deep learning outperforms the Dvorak techniques for estimating TC intensity from satellite imagery (e.g., Chen et al. 2019; Wimmers et al. 2019; Zhuo and Tan 2021). In particular, Zhuo and Tan (2021) introduced a physics-augmented deep-learning method named DeepTCNet

to estimate TC intensity and size from IR imagery. Deep-TCNet's IR-based TC sizes were found to compare well with recon-aided best track data for the North Atlantic TCs. To the best of our knowledge, DeepTCNet is the first model to use deep learning to estimate TC sizes. One of the main goals of this study is to establish an observed TC size dataset for WNP TCs from satellite data using DeepTCNet. The new dataset is then utilized to investigate the climatology and trends of TCs over WNP. The basic idea of this study aligns with the effort to develop consistent reanalysis of TC intensity from satellite data using objective algorithms (e.g., Kossin et al. 2007, 2013); however, in this paper we will concentrate on TC size metrics. Specifically, we focus on the WNP basin because it is the most active ocean basin in terms of TC activity, but there is an apparent lack of a reliable climate dataset of TC wind structure in this region. In addition, to better represent the TC wind field, the R17, R26, R33, and RMW spanning the outer- to the inner-core region of a TC will be considered.

In section 2, transfer learning is applied to DeepTCNet to adapt it for the WNP TCs. The model is then applied to reconstruct a long-term dataset of TC sizes, named DeepTCSize. Section 3 outlines the data used in this study. Section 4 provides an overview and a relatively detailed evaluation of the new dataset. Section 5 presents the applications of DeepTCSize to address the data issue of historical WNP TC size records and for the trend analysis of TC sizes, but no rigorous scientific attempt is made to attribute the observed changes to a specific cause. The concluding remarks are provided in section 6.

2. Deep-learning methods for the TC size estimation

a. DeepTCNet and its adaption for the WNP via transfer learning

DeepTCNet, developed by Zhuo and Tan (2021), is a deep learning–based method that learns to formalize the relationships between the IR imagery (input), TC intensity, and size (output) by iteratively adjusting its parameters until it can optimally predict the training cases. The backbone of Deep-TCNet consists of two-dimensional (2D) convolutional neural networks (CNNs; LeCun et al. 2015). A 2D CNN is a specialized type of artificial neural network designed to process spatial hierarchies of features in data. As a result, it is appropriate for exploiting relevant information of TCs from IR imagery. Deep learning algorithms generally require a large amount of training data to perform well, but the amount of TC data is frequently comparatively small due to the rarity of TCs and the difficulties in observing them. As such, physical knowledge of TCs has been introduced to augment DeepTCNet through data that embody underlying physics. For the TC size estimation, DeepTCNet was trained in a multitask learning framework to simultaneously estimate four wind radii metrics (i.e., R17, R26, R33, and RMW) and two intensity quantities (Vmax and Pc) serving as auxiliary outputs. When a deep learning model is trained on such data, it can learn functions that reflect the physical structure of the data (Karniadakis et al. 2021). In this case, it is the complex physical relationship between the wind structure of TCs and the IR features, possibly including the

cloud shield and distributions of convection. Taking advantage of the new technological capability of DeepTCNet, IR-based TC size estimations were found to compare well to the best track data for North Atlantic TCs with Pearson correlation coefficients as around 0.8 (Zhuo and Tan 2021). These size estimations were also found to be of better quality than those produced by multiple observation platforms (MTCSWA; Knaff et al. 2011). However, the previous version of DeepTCNet only applied to the North Atlantic TCs, considering this region best observes and documents the intensity and size data of TCs by the National Hurricane Center (NHC). Even though the general physical relationship between IR and TC size can hold between different basins, both IR and TC sizes themselves could have significant interbasin variations (e.g., Knaff et al. 2014; Chan and Chan 2015; Chavas et al. 2016). Therefore, the updating of the weights of DeepTCNet for the WNP TCs is required. As a result, this paper first develops reliable satellite observation-based size estimation algorithms for WNP TCs (Fig. 1a). The algorithms are then applied to reconstruct a new long-term TC sizes dataset dating back to the early 1980s (Fig. 1b; section 2b).

First, for the establishment of a model for estimating size metrics for the WNP TCs, transfer learning is carried out to adapt the network weights of DeepTCNet (Fig. 1a). Transfer learning (e.g., Pan and Yang 2010) is a machine learning method where a model developed for a task is reused as the starting point for another model on another task. Transfer learning has unique advantages in handling small-data scenarios; therefore, it is a good choice for training DeepTCNet with little reliable best-tracked TC size data available over the WNP. In addition, the feasibility of our methods is based on the fact that the fundamental physical relationship between the IR and TC sizes is transferable between the North Atlantic and the WNP basins. Specifically, the WNP version of DeepTCNet was trained using transfer learning as follows (Fig. 1a):

1) Pretrain DeepTCNet with NHC data for Atlantic TCs during 2005–20. The 2005–18 ($N = 6597$) data are used for training models and 2019–20 ($N = 1290$) validation samples for the early stop.
2) Fine-tune DeepTCNet parameters with JTWC data for WNP TCs during 2018–20. The 2019–20 data ($N = 1067$) for training and 2018 ($N = 1057$) validation samples for the early stop.

b. DeepTCNet refinement

For the best possible reconstruction of the final historical dataset of TC size metrics, named DeepTCSize (Fig. 1b), here, a two-step procedure is applied to refine DeepTCNet besides using transfer learning for the WNP adaption as shown in the last subsection. In the first step, only one IR imagery is taken as model input to produce the corresponding TC sizes, i.e., input = [IR]. At this step, we found that the IR-based size estimations at the time of TC lifetime-maximum intensity (LMI) were more accurate (e.g., $R = 0.88$; MAE = 36 km for R17; $R = 0.77$, MAE = 13 km for RMW) than all cases (e.g., $R = 0.83$, MAE = 43 km for R17; $R = 0.73$, MAE = 14 km for
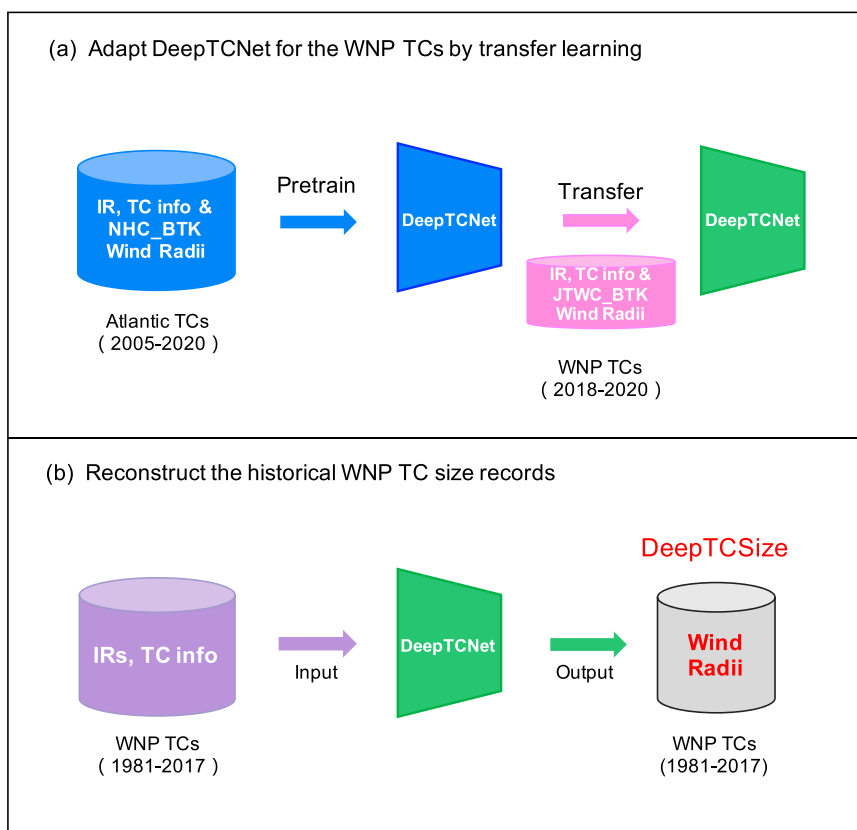
FIG. 1. The framework of the reconstruction of historical TC size records: (a) develop a reliable deep learning model (DeepTCNet) for the estimation of TC sizes/wind radii from IR imagery over the WNP basin and then (b) apply the WNP version of DeepTCNet to reanalyze a long-term satellite imagery record to generate a new dataset (DeepTCSize) of wind radii for WNP TCs over a 37-yr (1981–2017) period.

RMW), based on the 2016–17 test data of WNP TCs; R means the Pearson correlation; MAE means the mean absolute error. Better statistics also hold for R26 and R33 of the LMI cases than all cases. This could be due to the clouds being more symmetric and organized at the LMI stage of a TC, therefore providing higher "predictability" of TC sizes.

Although the accuracy of these TC sizes estimations produced solely from IR imagery by DeepTCNet has been promising in the first step, a second step is employed for further improvements by introducing some TC information besides the IR imagery as the input of DeepTCNet. As the intrastorm variabilities in TC sizes are observed to be smaller than the interstorm differences (e.g., Lee et al. 2010; Schenkel et al. 2018), providing the DeepTCNet some "size bound" besides the IR imagery could help constrain the size estimations to first order. Considering that the DeepTCNet's estimations of the TC sizes are more accurate at the time of LMI found in the first-step results, these data may be used to provide some "size bound." Specifically, we have used the R17 and RMW at the time of LMI (denoted as R17_LMI and RMW_LMI) as auxiliary TC information to augment the estimation of TC sizes from satellite data in the second step; i.e., the input is [IR, R17_LMI, RMW_LMI, DT2LMI], where DT2LMI is a time parameter normalized by the time of LMI. Overall, the second step yields an improvement over the TC sizes estimations obtained in the first step with higher correlation and lower estimation errors (e.g., $R = 0.85$, MAE = 39 km for R17; $R = 0.76$, MAE = 14 km for RMW for all the 2016–17 WNP TC cases).

The rationale behind the two-step procedure is based on the fact that introducing TC information besides the IR imagery can improve the performance of DeepTCNet (Zhuo and Tan 2021) and there is also a demand for a more homogeneous reconstruction of data to aid in the detection of trends in activity (Emanuel et al. 2018). In terms of the homogeneity demand, one can regard the two-step procedure as an interactive reanalysis of IR imagery because the IR imagery was first analyzed to obtain the R17_LMI and RMW_LMI in the first step, then the R17_LMI and RMW_LMI were combined with IR imagery again to produce the final TC size estimations in the second step. As such, the final reconstructed database DeepTCSize is strictly based on one source of satellite imagery data. Although the TC location, which is one important TC size-related factor, can also be used to augment the DeepTCNet for better TC size estimations, TC location has not been employed as the algorithm input because it requires using other data sources such as the best track database.

In this study, the network and training configurations of DeepTCNet remain unchanged as in Zhuo and Tan (2021), except that we 1) refined the input of DeepTCNet intending to achieve the optimal reconstruction of homogeneous TC sizes records, 2) enlarged the input image size to be larger (256×256 grids) by considering that WNP TCs tend to have wider cloud circulations, and 3) used learning rate as 1×10^{-4} to have a slight improvement in the overall performance of DeepTCNet. Also note that in both the first and the second steps, pretraining DeepTCNet with North Atlantic TC data and transfer learning with WNP TC data, as introduced in Second 2a (Fig. 1a), are required. Particularly, in the second step, for the model pretraining, the input R17 and RMW at the time of LMI are directly from the NHC best track to ensure the DeepTCNet learns a correct model starting point; while for the transfer learning to adapt for the WNP TCs and finally the establishment of DeepTCSize, the auxiliary TC information of R17_LMI and RMW_LMI are obtained from the DeepTCNet using IR as the only input in the first step.

As shown in Fig. 1b, the TC size estimation algorithm— DeepTCNet of WNP version refined with the two-step procedure—is applied to a spatiotemporally homogenized record of satellite data (section 3a) to form the record of TC size estimations, named DeepTCSize. DeepTCSize includes four TC size metrics R17, R26, R33, and RMW, provided at every 3 h. Fuller descriptions of DeepTCSize data, including which TC cases they correspond to, are listed in section 4a.

3. Data

In this study, three kinds of data were adopted to develop the algorithms and evaluate the resulting TC size estimations. The first kind of data is the IR imagery from a spatiotemporal homogenized record used as the data source of the TC size estimations. The second kind of data is the best track database used as training data able to fit the DeepTCNet algorithms and as a reference for accessing the TC size estimations. The third kind of data is the remotely sensed TC ocean winds data, also used as a reference for the evaluation of the TC size estimations.

a. IR satellite records

To create a homogeneous record of TC size metrics, the IR imagery from the Gridded Satellite (GridSat-B1) dataset version v02r01 (Knapp et al. 2011) is used. It is good to note that the IR channel (nearly 11 μm) of GridSat-B1 meets the climate data record standards in terms of length, consistency, and continuity. This is because these data have been reprocessed and recalibrated, and the later data have been subsampled spatially and temporally to be homogeneous, with the earlier data (0.07° latitude spatial and 3-h temporal resolution) (Kossin et al. 2013). Therefore, we have attempted to reconstruct a homogeneous record of WNP TC sizes based on the deep-learning reanalysis of GridSat. Since there is a lack of available geostationary satellite data in the Eastern Hemisphere in 1980 (Knapp et al. 2010), our period of analysis starts in 1981 when the satellite data are consistently available. We have cropped the global-scale GridSat-B1 to be a

TC-centered data of 256×256 grids using the best track storm center positions provided by NHC and JTWC. One can also directly use the GridSat data via HURSAT which has been cropped to be TC-centered data (Knapp and Kossin 2007), but HURSAT so far has only been updated through 2015 for the public access. For data preprocessing, IR images with more than 30% invalid values (brightnessnc1 temperature < 140 K or >375 K) were removed. Sample-specific normalization was then applied to each IR image by deducting the average and dividing by the standard deviation.

b. Best track data

The best track data used in this study are taken from the International Best Track Archive for Climate Stewardship (IBTrACS) version 4.0 database (Knapp et al. 2010), which are mainly used to 1) provide TC center locations to crop the IR imagery, 2) feed labels of training data for DeepTCNet algorithms, and 3) provide a reference for the assessment of DeepTCSize data. IBTrACS has collected TC data produced by different regional specialized meteorological centers and tropical cyclone warning centers around the world and provides interpolated 3-hourly estimates of the location, intensity, wind radii, and other parameters covering each TC's life cycle. In line with the focus of our work, we have used the data provided by the NHC, JTWC, and JMA.

NHC maintains the record of R17, R26, and R33 in the geographic quadrants (northeast, southeast, southwest, and northwest) surrounding the storm, which extends back into the late 1980s for the North Atlantic and eastern North Pacific basins (Demuth et al. 2006). However, NHC's poststorm analysis of these data only began in 2004 (Landsea and Franklin 2013). JTWC provides wind radii data starting in the early 2000s, but the poststorm reanalysis only began in 2016 (Sampson et al. 2017). RMW data are both accessible from NHC and JTWC best tracks; however, it is worth noting that the NHC is currently the only agency that conducts poststorm reanalysis of RMW in the world (starting from 2021; Knaff et al. 2021). JMA documents the longest best track records of TC sizes over the WNP, which include the longest and shortest axis of R15 and R26 in geographic octants from 1979 onward (Knaff et al. 2021), but JMA neither provides data for R33 nor RMW. In this paper, we have directly employed JMA R15 as R17. Also, note that the gale region in TCs could be relatively widespread. As mentioned, these agencies have relied on diverse observations, techniques, and standards for analysis. Specifically, both NHC and JTWC have used numerous scatterometers such as ASCAT to measure TC sizes (Sampson et al. 2017), while JMA relied on available observations, such as *Himawari-8* satellite images, radar, surface synoptic observations, ship, buoy, and ASCAT (Muroi 2018). In the absence of the necessary observations, JMA estimates R15 and R26 between central pressure and wind radius using regression equations (Muroi 2018).

In this study, the R17, R26, and R33 refer to nonzero azimuthally averaged 17, 26, and 33 m s^{-1} wind radii. For NHC and JTWC data, we have calculated the wind radii as nonzero average of the existing quadrants. For JMA, the azimuthal-mean wind radii were calculated as the average of the longest
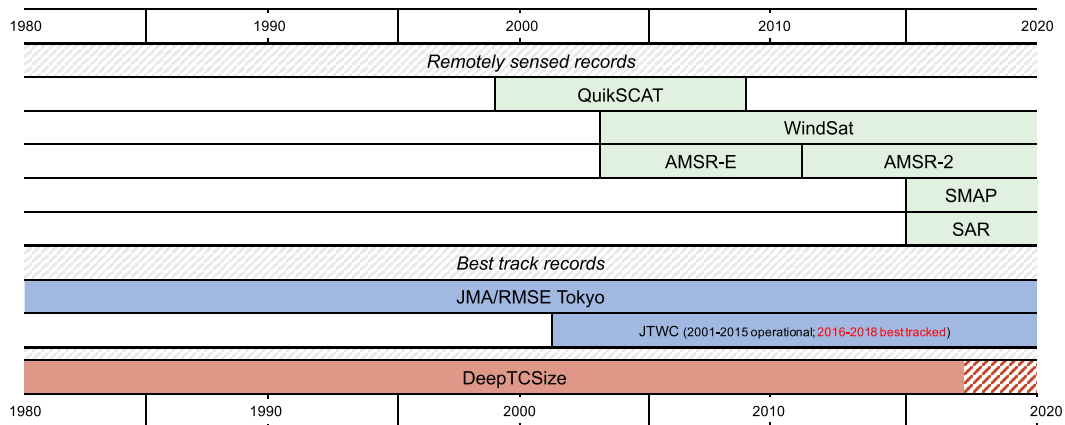
FIG. 2. TC sizes records since 1980 in the western North Pacific. Shaded horizontal areas indicate the periods during which the TC size data were and are available from remotely sensed observations (green), best track datasets (blue), and the DeepTCDat (brown).

and shortest geographic axes. Following Zhuo and Tan (2021), only named storms of the JTWC best track intensity (Vmax) above 17 m s^{-1} are considered in this study. Note that records that did not satisfy the physical relationship R17 > R26 > R33 and R17 ≥ RMW, R26 ≥ RMW, R33 ≥ RMW were removed.

As stated in section 2, we have only used R17, R26, R33, and RMW both from NHC over 2005–20 and JTWC over 2018–20 for our algorithm development, considering these wind radii (except RMW) data have been applied with post-season quality controls or have been observed with improved capabilities (Knaff et al. 2021), i.e., in high quality. Also, we have used the out-of-sample best track wind radii data which is quality controlled at JTWC (i.e., 2016–17 data) as the test set to assess our model and its TC size estimations. The JTWC data in 2001–15 (not postseason reanalyzed) and the JMA data were only introduced for data comparisons and discussions.

c. TC sizes from ocean wind observations

To assess the quality of our TC sizes estimations constructed from IR imagery, several observations of surface winds have also been considered. We used remotely sensed ocean winds from three data archives. The first data archive is R17, R26, and R33 documented in the optimized version of the QuikSCAT database of science quality, which spans the period 1999–2009 (Chavas and Vigh 2014). Following the recommendation of Chavas et al. (2016), we have kept only the QuikSCAT wind radii for which the uncertainty parameter is defined as $\zeta \leq 0.5$. The second data archive is R17, R26, and R33 from the Remote Sensing Systems (RSS) TC-winds products (Meissner et al. 2021), which consists of SMAP, WindSat, and AMSR observations. The L-band SMAP observations provide realistic wind retrievals in TCs and are not affected by rain (Meissner et al. 2017), and the WindSat and AMSR TC-winds are reprocessed from their C-band and X-band channels to be of comparable quality with the SMAP (Meissner et al. 2021). The SMAP, WindSat, and AMSR data span 2015–20, 2003–20, and 2003–20, respectively. The third data archive is the 3-km

averaged co- and cross-polarized SAR winds data, which includes measurements of R17, R26, R33, and RMW. Recent studies have revealed that the maximum winds observed by SAR are well correlated with small biases to those estimated from the airborne SFMR, and the SAR could currently be the most reliable data source for the inner-core size for the TCs in the WNP (e.g., Combot et al. 2020). However, it is good to note that the SAR data are sparse and only a few SAR samples coincident with our reconstructed data are available. These remotely sensed TC size observations are applied to assess the reliability of our reconstructed data of TC sizes.

The remotely sensed ocean wind data were also processed with nonzero azimuthal average, and these data were only used to assess the quality of TC size estimations, and not to train the DeepTCNet.

4. Dataset overview and evaluation

a. Overview of DeepTCSize

As described in section 2b, DeepTCSize is obtained by applying DeepTCNet to a homogenized record of satellite IR imagery. The DeepTCSize data corresponds to the following: 1) named storms of the best track intensity (Vmax) above 17 m s^{-1} over the WNP basin (100°E–180° and 0°–60°N); 2) R17 and RMW data that are valid only for TCs of Vmax ≥ 17 m s^{-1}, R26 data that are valid only for TCs of Vmax ≥ 26 m s^{-1} and R33 data that are valid only for TCs of Vmax ≥ 33 m s^{-1}; 3) cases having TC fullness: TCF = 1 − RMW/R17 ≥ 1 − 17/Vmax = TCF$_0$ to remove unphysical values (Guo and Tan 2017, 2022); 4) exclusion of none tropical systems such as the post-tropical cyclones and miscellaneous disturbances.

Figure 2 displays the temporal availability of DeepTCSize (brown bar). DeepTCSize spans the past 37 years, from 1981 to 2017, to be precise. Due to the use of 3-hourly all-weather IR-channel satellite data (section 3a), DeepTCSize also possesses a continuous and high temporal coverage. The year 1981 is the year in which geostationary satellite data started to become consistently available in the eastern hemisphere (Knapp et al. 2010).
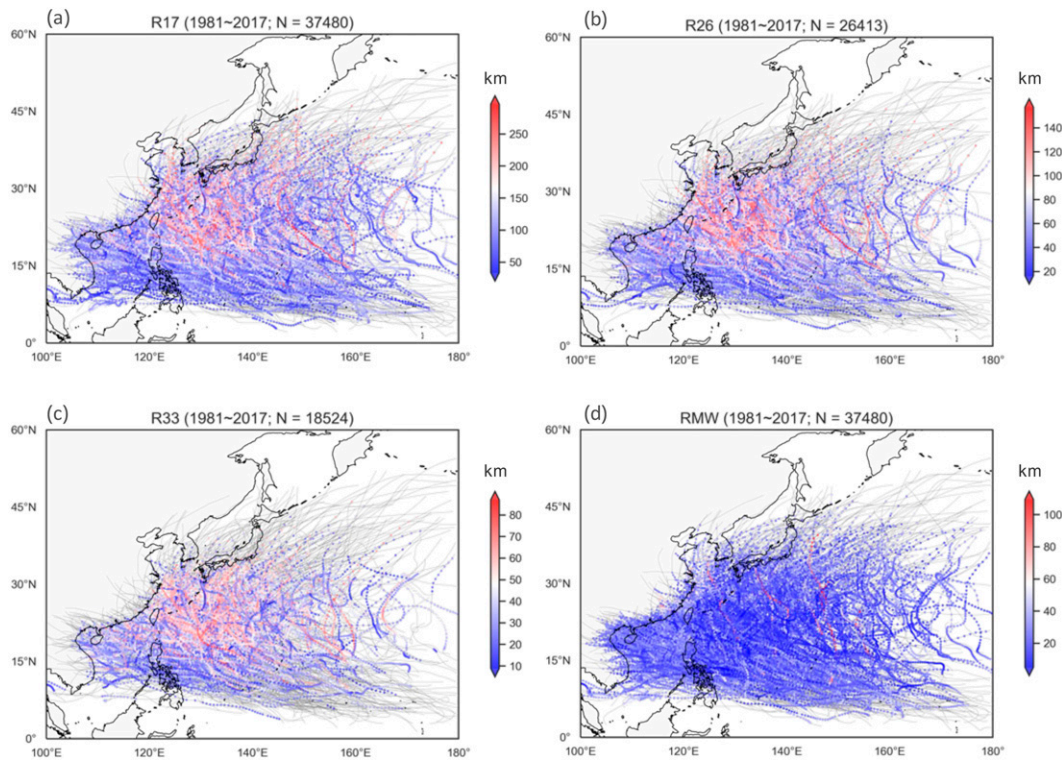
FIG. 3. Maps of the spatial distributions of all samples in DeepTCSize for (a) R17, (b) R26, (c) R33, and (d) RMW.

Also, the 2018–20 data used for DeepTCNet development are not included in DeepTCSize.

In Fig. 2, comparisons between the historical datasets of TC sizes available over the WNP are also shown. Among these datasets, DeepTCSize and JMA maintain the longest records, roughly from 1980, and these records are nearly 2 times longer than those documented by JTWC. For remotely sensed surface wind data of TCs, both WindSat and AMSR (AMSR-E + AMSE-2) cover 18 years (2003–20), QuikSCAT spans 11 years (1999–2009), while the more recent capacities of SMAP and SAR became available in 2015. Although Fig. 2 is not a complete representation of observational availability, it raises some concerns about the quality of the data of historical TC size records before 2000. This is because the observations and techniques available to inform wind radii estimates are limited or nonexistent (Knaff et al. 2021). Figure 2 also reveals the significant changes and advances in observational capabilities over time. This also appears to be one of the major issues of best track database, and it will be discussed further in section 5. In conclusion, of all the datasets listed in Fig. 2, only DeepTCSize provides homogeneous TC size data spanning nearly 40 years. This capability of DeepTCSize could support the reliable detection of long-term trends in WNP TC size, as long as the dataset's accuracy (to be examined in section 4b) is also proven.

The map of the DeepTCSize for the WNP basin (100°E–180° and 0°–60°N) is illustrated in Fig. 3. DeepTCSize comprises four TC size parameters viz, R17, R26, R33, and RMW. The sample size is 37 480 for R17, 26 413 for R26, 18 524 for

R33, and 37 480 for RMW. All the data are in MKS units. As shown in Figs. 3a–c, the spatial distributions of the outer-core sizes R17, R26, and R33 are relatively similar, and they increase with recurving tracks and higher latitudes, as noted in previous studies (e.g., Kimball and Mulekar 2004; Knaff et al. 2014; Chan and Chan 2015). The patterns of the spatial distributions of R17, R26, and R33 also suggest that there is a consistent expansion or contraction of the outer circulations of TC wind fields. RMW is shown in Fig. 3d, and it can be seen that no clear pattern is observed. Therefore, we regrouped the RMW in DeepTCSize by latitudes at a 5° interval and found that the RMW slightly decreases from 35 km south of 10°N to a minimum of 32 km at 15°N, then increases to over 45 km north of 35°N. This observation that RMWs increase with latitude, has been reported in previous studies (e.g., Knaff et al. 2015), and this could be due to fact that R34 also increases with latitude as suggested by Chavas and Knaff (2022). Table 1 shows the data statistics of TC sizes in

TABLE 1. Statistics of TC sizes in DeepTCSize, including data number, median, first and third quartiles, mean, standard deviation (Std), and coefficient of variation (CV).

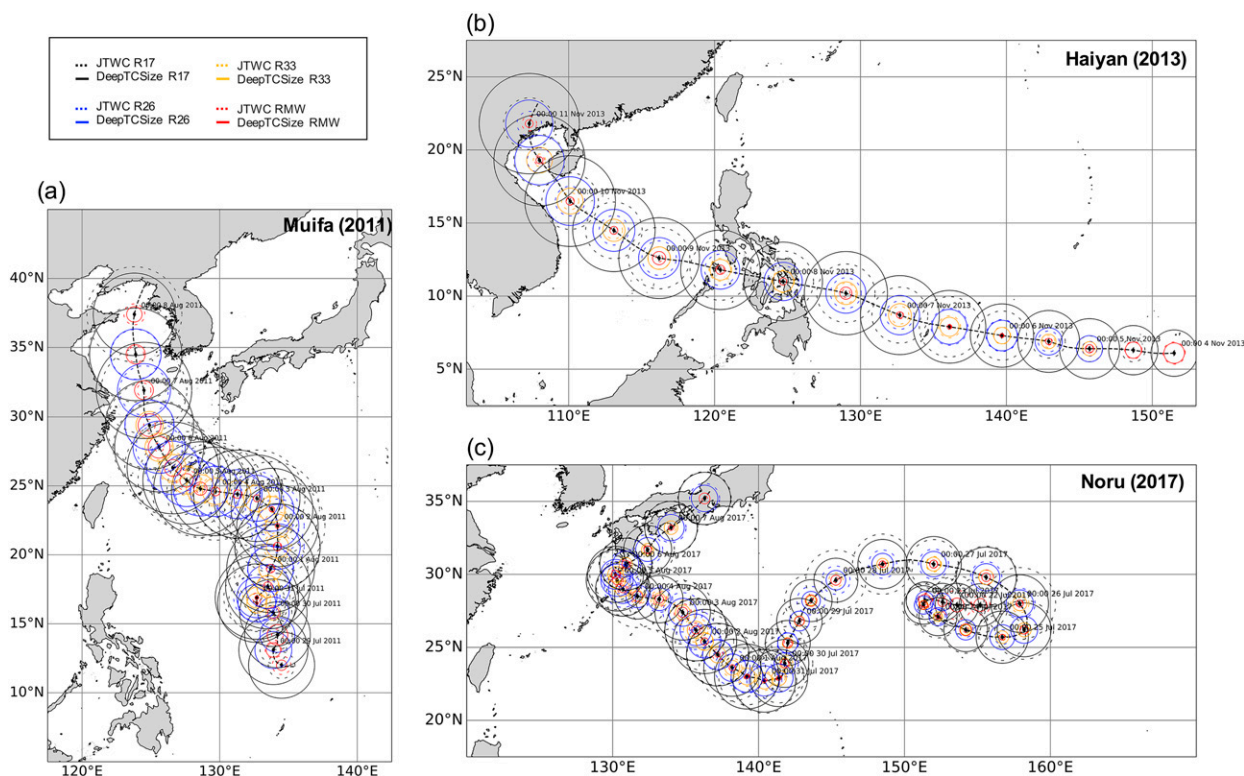| TC size (km)	N	25%	Median	75%	Mean	Std	CV (%)
R17	37 480	171	228	296	240	88	37
R26	26 413	91	123	164	130	49	38
R33	18 524	53	74	96	75	26	35
RMW	37 480	32	47	63	50	24	49

FIG. 4. Wind radii estimations from DeepTCSize (solid-line circle) and JTWC best track (dashed-line circle) for (a) Typhoon Mifua (2011), (b) Typhoon Haiyan (2013), and (c) Typhoon Noru (2017). The estimates are shown every 12 h with R17, R22, R33, and RMW being displayed in black, blue, orange, and red, respectively. Dates/times are shown once a day to the right.

DeepTCSize. The mean value of R17 is 240 km with an interquartile range of [171, 296] km. This descriptive statistic is comparable to the mean R17 values (235 km) for WNP TCs, calculated using QuikSCAT observations (Chan and Chan 2015). Our estimated mean R17 for WNP also is larger than the observed mean R17 in other ocean basins (Chan and Chan 2015). In addition, the mean values for R26, R33, and RMW are 130, 75, and 50 km, respectively.

To demonstrate how the DeepTCSize data would appear to users, Fig. 4 illustrates three long-lived TCs of the WNP that all caused widespread damages—Typhoon Muifa (2011), Typhoon Haiyan (2013), and Typhoon Noru (2017). The DeepTCSize R17, R26, R33, and RMW are in black, blue, orange, and red with solid lines, respectively, showing inter- and intrastorm variations of TC size metrics for these three cases. As a simple comparison, Fig. 4 also shows the size estimations from JTWC in dashed-line circles. The comparisons of the TC size estimations from different agencies will be discussed in section 5a.

b. Evaluation of DeepTCSize

To evaluate the validity of DeepTCSize, we first compared our TC sizes products to coincident best track estimations that have been postseason verified by the JTWC (2016/17 TCs; $N = 1545$). Figure 5 shows the comparison results for R17, R26, R33, and RMW. From the comparisons, it was observed that there is a high correlation between the JTWC best tracks and DeepTCSize for the four wind radii. Specifically, R17 has the highest correlation ($R = 0.85$), followed by R26 ($R = 0.84$), R33 ($R = 0.79$), and RMW ($R = 0.76$). The normalized bias values are 13%, 8%, 8%, and 10% for R17, R26, R33, and RMW, respectively. Among the four wind radii, the outer-most TC size, R17, has the highest correlation coefficient ($R = 0.85$) and the smallest scatter index (SI = 24%) (i.e., the best accuracy). It is good to mention that the accuracy decreases with wind speed values associated with the radii, yielding $R = 0.76$ and SI = 36% for RMW. This is consistent with the inherent uncertainty of the best track wind radii (Landsea and Franklin 2013; Combot et al. 2020). Since the best track wind radii data were mostly derived from indirect methods or low-orbit satellite missions with medium to low resolution (Knaff et al. 2021), a poor resolution during the observations could have limited the precision of the wind radii and RMW data for higher wind speeds, especially the most intense inner-core wind radii RMW. Figure 5 also shows that the data points at the time of LMI, have TC size estimations with higher correlations ($R = 0.90, 0.90, 0.89,$ and 0.80 for R17, R26, R33, and RMW, respectively) and lower errors (MAE = 35, 21, 12 and 12 km for R17, R26, R33, and RMW, respectively).

Using the same test samples, R17, R26, R33, and RMW of DeepTCSize for the Saffir–Simpson storm intensity category were further analyzed. Figure 6 shows that the variations of
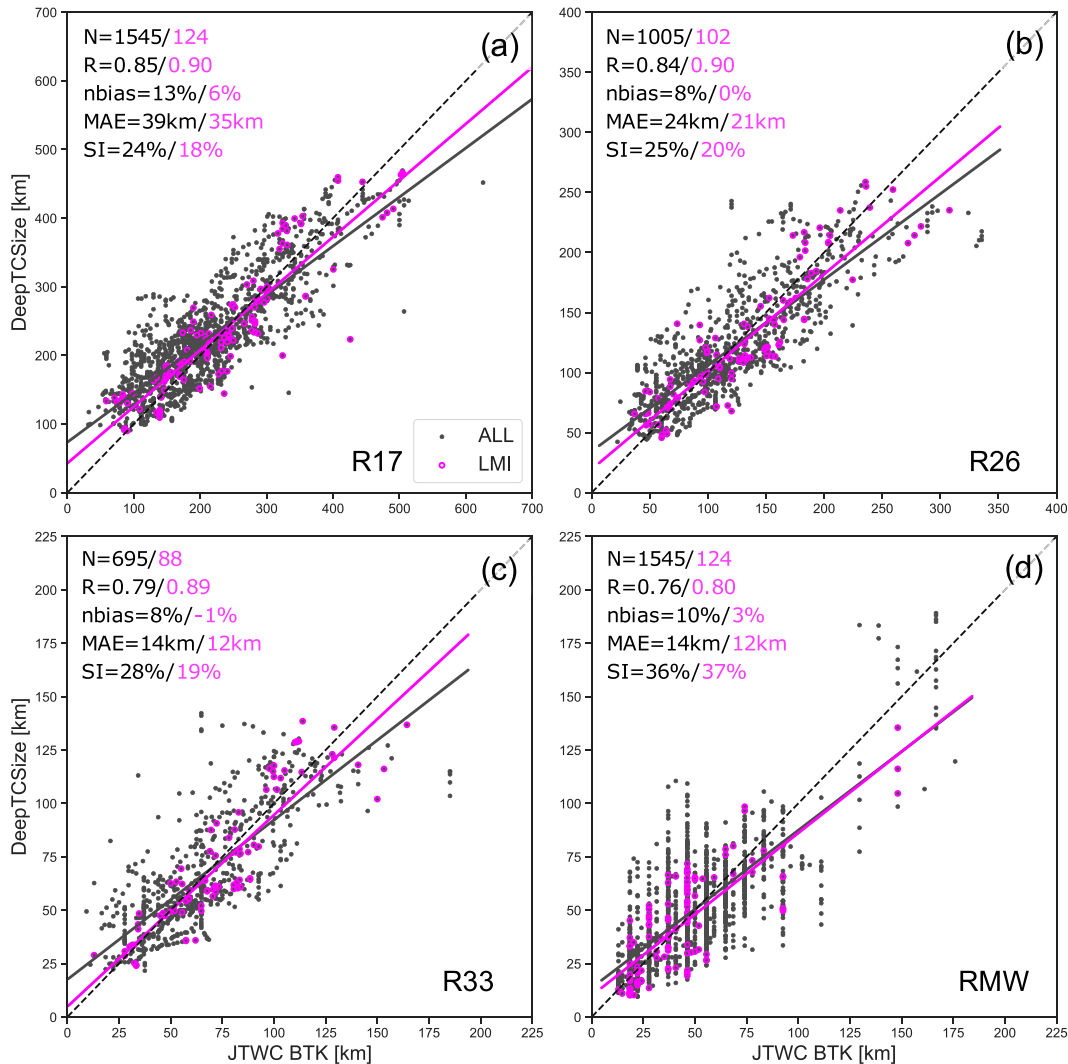
FIG. 5. Comparison of DeepTCSize and best-track data verified postseason in 2016–17 for (a) R17, (b) R26, (c) R33, and (d) RMW for all coincident samples (black dots) and the cases at the time of lifetime maximum intensity (magenta dots). The black dashed line is the "1-to-1" line; black and magenta lines are the linear regression model fits for the data.

the averaged DeepTCSize wind radii (solid lines) as a function of the storm category are identical to those of the best-tracked mean wind radii (dashed lines) for each size metric. One of the most basic TC intensity and size relationships widely observed and studied is that the larger TC outer-core sizes (R17, R26, and R33) and the smaller TC inner-core size RMW when the intensity increases. This is well captured by DeepTCSize. Figure 6 also presents the mean absolute difference (i.e., MAE) between DeepTCSize and the best track data (shaded area) for each storm category. The results show that the MAEs of R17, R26, and R33 are almost constant relative to the TC intensity. However, the MAEs of RMW in DeepTCSize are larger for weak storms and smaller for intense storms. This could be indicative of the higher uncertainty in estimating the RMW of weak storms from IR imagery.

To make the dataset evaluation more robust, we have also compared DeepTCSize with the remotely sensed ocean wind observations described in section 3c. As DeepTCSize has a much higher temporal resolution (3 h) over the sparse and intermittent observations, we have interpolated the Deep-TCSize data to be the time of observations before the calculations. Figure 7 shows the comparisons made between DeepTCSize and QuikSCAT, WindSat, AMSR, and SMAP for R17. High correlations are observed for DeepTCSize. The results for R26, R33, and RMW have also been derived using the same method. Assuming the SMAP as reference values for R26 and R33, the correlation is 0.72 ($N = 101$) and 0.73 ($N = 96$), respectively. As for RMW, the SAR could currently be the only satellite observation platform of proven capability over the WNP (Combot et al. 2020). However, there are only 15 coincident RMW samples for DeepTCSize and SAR and their
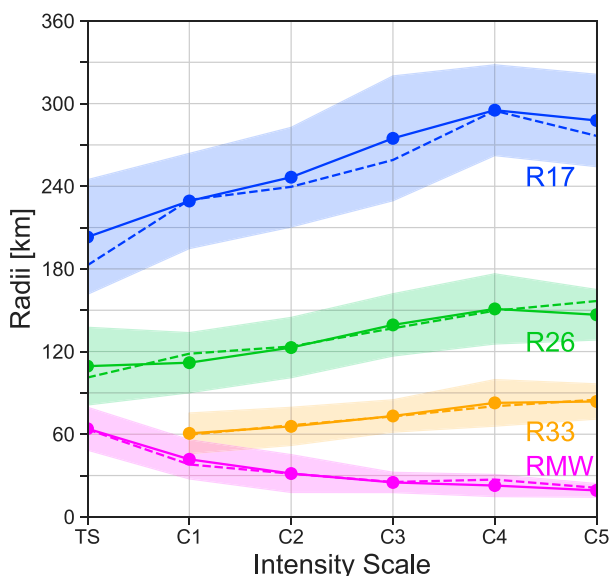
FIG. 6. TC sizes (R17: blue; R26: green; R33: orange; RMW: magenta) from DeepTCSize and JTWC best tracks as a function of (Saffir–Simpson) TC intensity scale for 2016–17 WNP TCs. The mean category values for each TC size parameter are plotted for DeepTCSize (solid lines) and JTWC best tracks (dashed lines), and the shaded areas denote the mean absolute difference between DeepTCSize JTWC best tracks.

correlation is 0.69. In addition, we also verified DeepTCNet via a simple case study.

In all, the MAEs of the four wind radii for WNP TCs are consistent with those predicted for North Atlantic TCs using DeepTCNet (Fig. 4 of Zhuo and Tan 2021). This confirms that the adoption of DeepTCNet is a good choice. The results calculated here using DeepTCSize are also consistent with previous studies that state that the best track wind radii have uncertainties (in terms of SI) in a range from 10% to ~40% (Knaff and Sampson 2015). Although we acknowledge the uncertainty of the TC size data is debatable, it is important to note that our IR-based TC size estimations produced by deep learning have much lower errors (MAEs < 40 km) in total, in comparison to the existing indirect observation-based measurements (MAEs > 46 km) (Demuth et al. 2006; Kossin et al. 2007; Knaff et al. 2011, 2016). Our IR-based TC size estimations are also found to be comparable to the emerging observational capabilities (Reul et al. 2017; Combot et al. 2020). Since the more recent best track data have higher quality due to technology advancements, the good agreement between DeepTCSize and the 2016–17 quality-controlled best track wind radii (Fig. 5) suggests that DeepTCSize is comparable to the "best estimates" currently available for WNP TCs. The relatively detailed assessment of DeepTCSize performed here has increased our confidence in DeepTCSize. In the following section, we utilized DeepTCSize to explore the data discrepancy issue of TC sizes over the WNP, and then analyzed the long-term climatology of WNP TC size.

5. Applications of DeepTCSize

a. Comparison of historical TC sizes data over the WNP

The lack of consistent and reliable historical data has limited the research and understanding of the nature of TC sizes. As noted previously, there are a few historical datasets of TC size now available over the WNP; however, their quality remains either unknown or unverified. In particular, the homogeneity of the datasets remains a serious issue, for they can potentially yield spurious signals. To have a better understanding of existing issues in TC size data and their possible influences on the resultant findings, this section compares DeepTCSize with two widely used best track datasets. Note that the TC wind radii directly observed from remote sensing platforms (see section 3c) are assumed to be the reference values.

Figure 7 shows comparisons of the R17 estimations from DeepTCSize, JTWC, and JMA. In Fig. 7, it can be observed that there are significant biases in the JMA best track database. That is, JMA systematically produces larger R17 estimations, hence the largest absolute errors. In contrast to this, the R17 estimates from both JTWC and DeepTCSize are in strong consonance with the remotely sensed TC size data. In Fig. 7, it can also be observed that the three datasets have relatively good correlations with the remotely sensed data (0.63–0.86). These good correlations may be a result of JMA and JTWC operationally adopting the remotely sensed data to produce TC size estimations (Sampson et al. 2017; Knaff et al. 2021). However, the high correlations between the observations are more robust for the objective IR-based estimations of DeepTCSize. Similar comparisons have been made for R26 (Fig. 8). As expected, the correlations for R26 slightly decrease as the wind speed increases. This is probably due to the already discussed resolution issue (see section 4b). However, for R26, the JMA overestimation issue seems to have been reduced. In all, it can be said that DeepTCSize has the most accurate historical TC size estimations (R17 and R26), even though only IR imagery satellite data were used to generate the dataset. This implies that some operational technology issues may exist in the JMA and JTWC TC size estimations.

Data discrepancy is further explored as the time series of R17 calculated with JMA, JTWC, and DeepTCSize (Fig. 9). It can be observed that JMA's annual-mean R17 is consistently larger than those obtained by DeepTCSize and JTWC, as illustrated in Fig. 7. Note that such data discrepancy among agencies has also been exemplified by previous studies (e.g., Song and Klotzbach 2016; Kim et al. 2022). Much of the disparity can be attributed to agency-specific techniques and policies: JMA estimates TC winds in conjunction with other midlatitude synoptic systems to make more comprehensive TC warnings, whereas JTWC estimates wind radii mainly based on a TC's winds (Kim et al. 2022). Using DeepTCSize as a baseline, we have also found that JTWC's best track database had a significant change in quality around 2016. This is the year JTWC initialized the postseason reanalysis of wind radii (Sampson et al. 2017), and our results indicate that the verification of TC size data at JTWC seems to have a
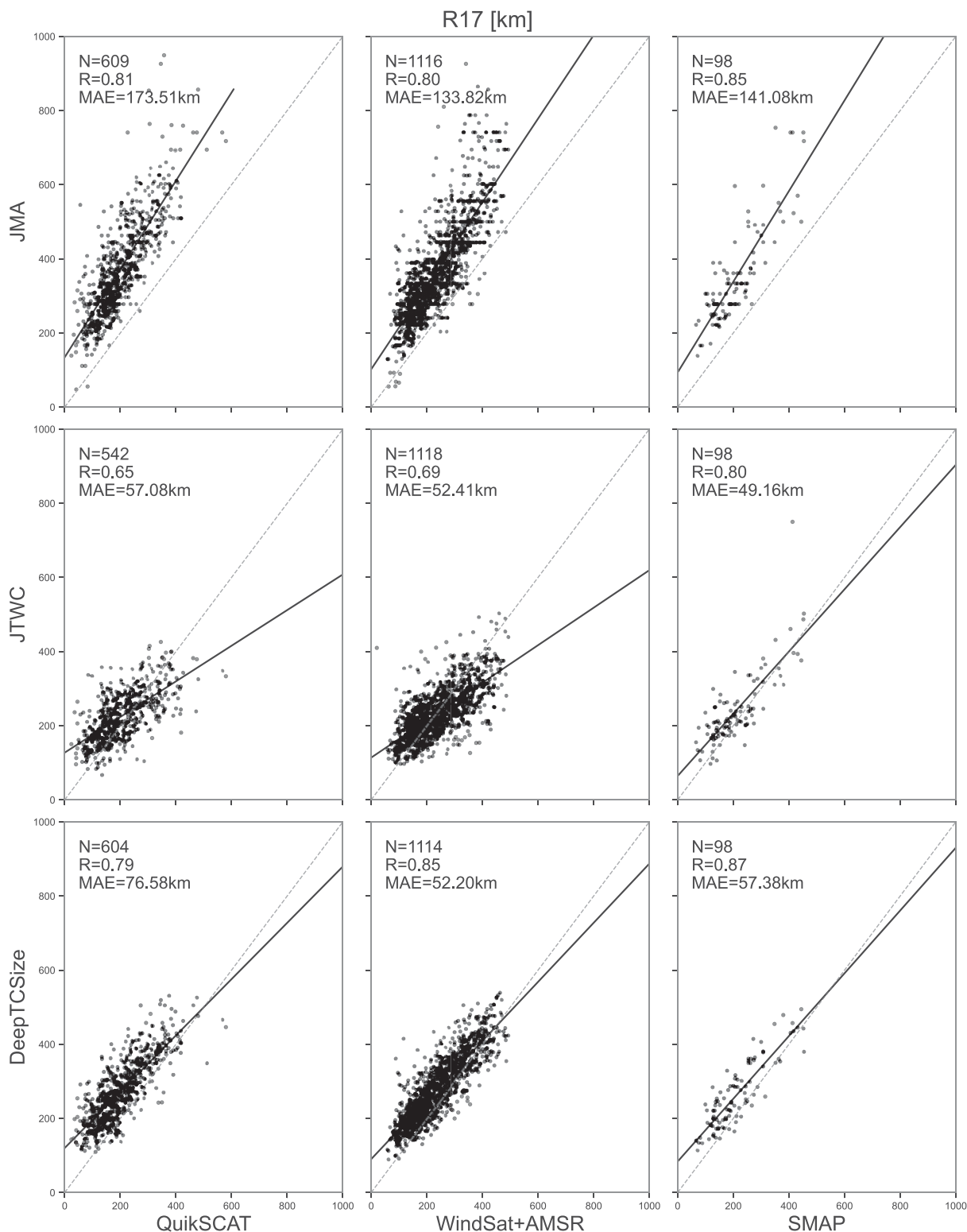
R17 [km]

FIG. 7. As in Fig. 5, but for all coincident samples from historical remotely sensed ocean wind observations [(left) QuikSCAT, (center) WindSat + AMSR, and (right) SMAP] and indirect estimations [(top) JMA, (middle) JTWC, and (bottom) DeepTCSize] for R17. Note that WindSat and AMSR are combined as one source for they have similar data quality.
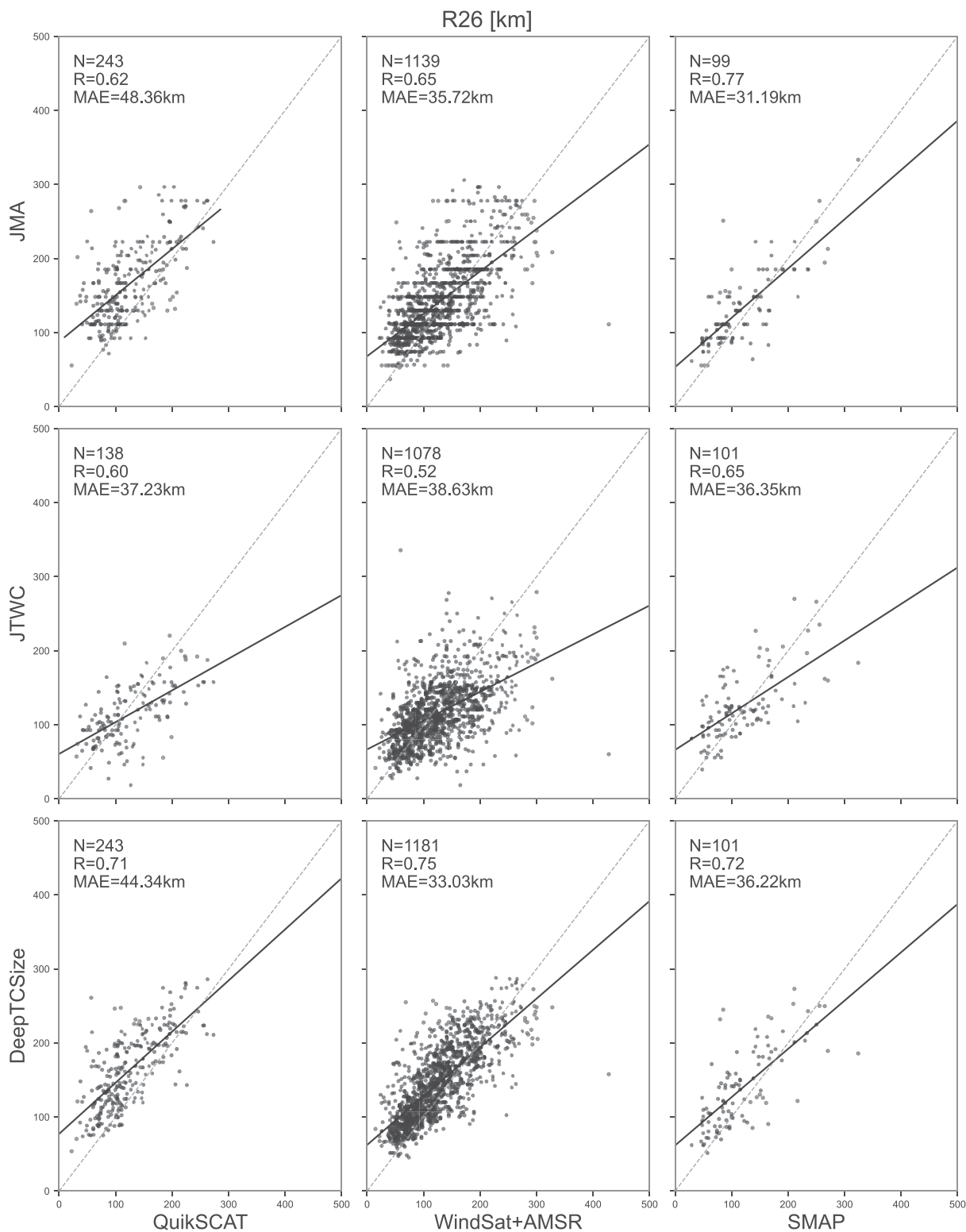
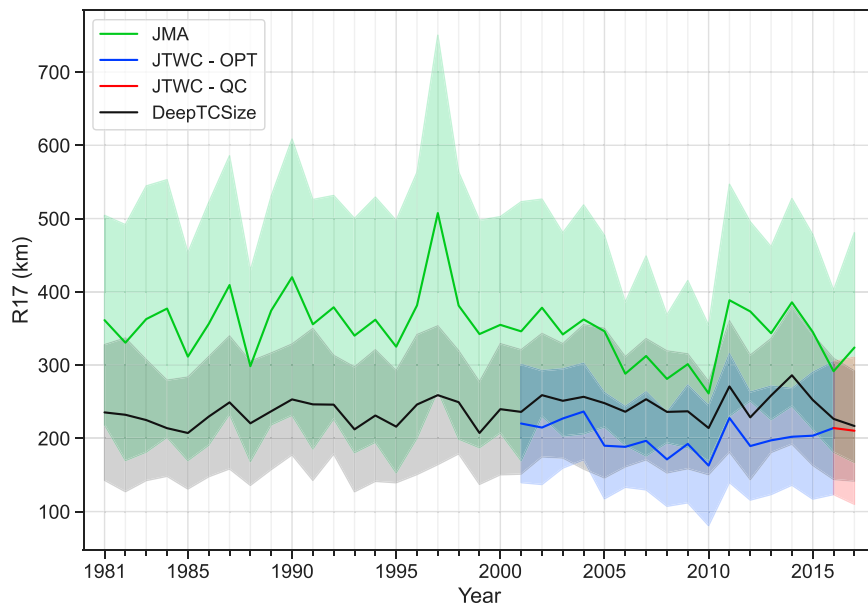## R26 [km]

FIG. 8. As in Fig. 7, but for R26.

FIG. 9. Time series of R17 from JMA best tracks (green; 1981–2017), JTWC best tracks with (red; 2016–17) and without rigorous postseason reanalysis (blue; 2001–15), and DeepTCSize (black; 1981–2017). The lines and shades denote the mean and standard deviation, respectively.

substantial impact. The application of modern techniques at JTWC could be an additional factor (JTWC 2016, 2017). Despite the considerable discrepancy in the mean values, the time series of R17 from JMA, JTWC, and DeepTCSize exhibit comparable interannual variability, suggesting all three datasets have captured the large-scale controls on the TC outer size in terms of R17.

There is another important issue—would the errors in these datasets cause spurious signals of trends in TC size, and how large are the amplitudes of the noise? To answer this question, we have calculated the errors of R17 estimates interpolated from JMA, JTWC, and DeepTCSize against the remotely sensed observations, as a function of time. Since the remotely sensed data from each satellite source remain consistent during their serving periods, if the errors of R17 estimates have statistical trends, then the quality of R17 estimates during one specific period could have changed. Figure 10 depicts error points and the ordinary least squares fit with seasons. During the first decade of the 2000s, JMA exhibited a statistically significant ($p = 0.018$) error trend of R17. In the 2000s, JTWC also shows a weak increasing trend of errors, but it is not statistically significant. In contrast, we have observed no trends in the errors of DeepTCSize compared with the ocean–wind observations. Taking DeepTCSize against WindSat + AMSR (longest coverage over 2003–17) as an example, the trend of error is around 0.32 km decade^{-1} ($p = 0.942$), confirming the good temporal consistency of DeepTCSize. Comparisons of DeepTCSize R26 and R33 with WindSat + AMSR (not shown) also yield near-zero trends in errors. This result increases our confidence in applying DeepTCSize for examining the changes in TC size metrics in the following subsection.

b. Trend analysis

Given DeepTCSize's long-term coverage and confirmed quality and homogeneity, we have used the DeepTCSize data to investigate the trends of TC sizes over the WNP in the 37-yr period (1981–2017) (Fig. 11a). Note that only the TCs lasting ≥2 days have been considered for the trend analysis, and the corresponding maximum wind speed thresholds used for the samples of R17, R26, R33, and RMW are 17, 26, 33, and 20 m s^{-1}, respectively.

The annual mean time series and corresponding ordinary least squares fit of these 37 single values for each TC size parameter are shown in Figs. 11a–d, and the statistics of the slopes are detailed in Table 2. Both R17 and R26 exhibit significant increasing trends: +5.8 km decade^{-1} ($p = 0.04$) for R17 and +3.1 km decade^{-1} ($p = 0.07$) for R26. In contrast to this, the wind radii for high wind speed R33 show a near-zero increasing trend (+0.6 km decade^{-1}), and the TC inner-core size RMW shows a near-zero decreasing trend (-0.2 km decade^{-1}). It is also worth noting that, even though the amplitude of the mean trend of R17 seems to have a small value (+5.8 km decade^{-1}, roughly equal to 2%) for the WNP TCs over 1981–2017, such a trend for a TC with an averaged R17 of 240 km (Table 1) corresponds to an area expansion of +2818 km^2—roughly half of the Shanghai area and 4 times of New York City—per decade, taking the TC wind field as a first-order symmetric circle.

Figure 12 shows the trends of annual mean TC sizes over the 37-yr period when the storms reach their LMI. In Fig. 12, it can be observed that there are increasing trends for outer-wind radii (R17, R26, and R33) and decreasing trends for inner-core size (RMW). However, the amplitudes for these LMI samples are greater than those of all WNP storms (Table 2).
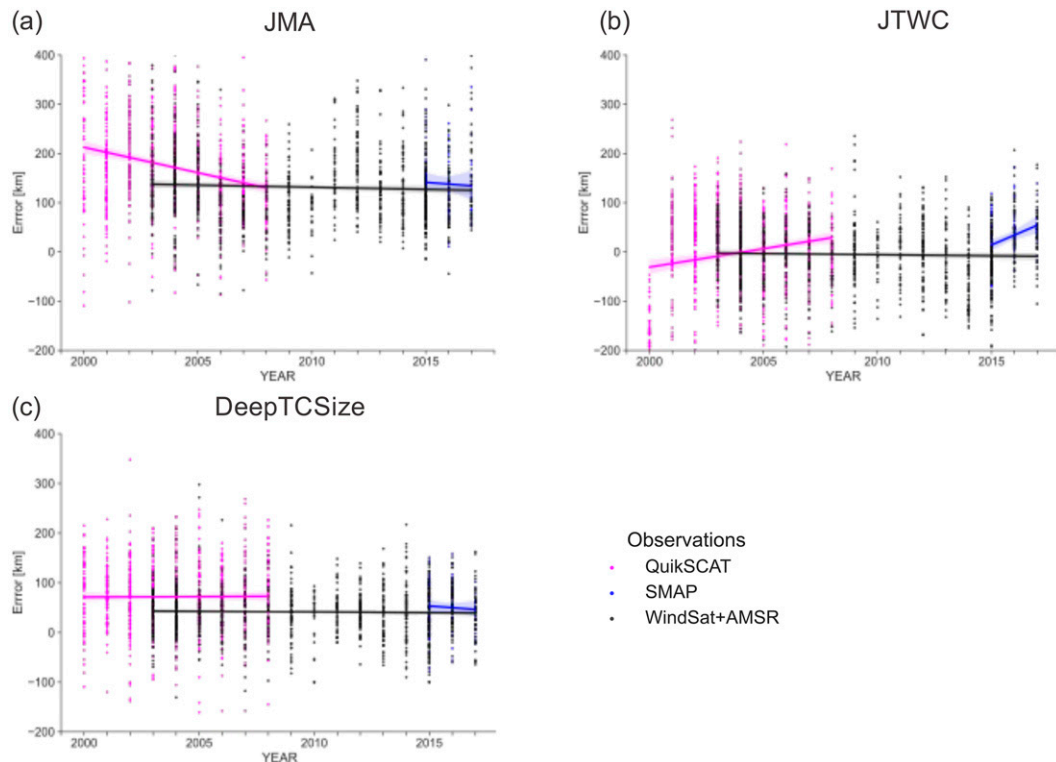
FIG. 10. Time series of errors of R17 estimates in (a) JMA best track database, (b) JTWC best track database, and (c) DeepTCSize over 2010–17, assuming directly observed R17 from remotely sensing (magenta: QuikSCAT; black: WindSat + AMSR; blue: SMAP) as reference values. Dots denote the R17 estimates minus the observed R17 values; the line displays the linear regression model fit. Note that the R17 estimates have been interpolated to be at the same time of observations.

The trends for R17, R26, R33 and RMW indicate metrics of $+7.4$ km decade^{-1} ($p = 0.03$), $+4.1$ km decade^{-1} ($p = 0.02$), $+1.7$ km decade^{-1} ($p = 0.11$), and -0.4 km decade^{-1} ($p = 0.61$), respectively. If the observed trends in the expansion of TC outer circulations persist, the destructive potential and damage hazards of TCs may substantially increase.

As already explained and noted so far, the simple time series is based on the reduction of each year's data to annual-mean single values, such that the trends are obtained with the ordinary least squares fit to these 37 single values. To perform a more rigorous analysis of the trends and their levels of confidence, the quantile regression is applied to the WNP TC size in DeepTCSize (as used for Fig. 11). The quantile regression is used similarly to the trend analysis of TC intensity described in Kossin et al. (2013). Figure 13 shows the quantile regression for each 2.5% quantile from 5% to 95%. In Fig. 13, it can be seen that R17 exhibits significant trends in all quantiles, except for those above the 90th percentile, indicating an overall large shift of TC outer-core size (Fig. 13a). The peak trend, approximately $+8$ km decade^{-1}, is also observed close to the 80% quantile ($\sim$314 km) of the distribution in Fig. 13. The vanished trend at the 0.9 quantiles in Fig. 13 can be partly attributed to the limited sample amount. R26 shows a similar pattern to that of R17 for the slopes of the trends (Fig. 13b). Also in Fig. 13b, the median to large quantiles between 50%

and 80% ($\sim$123–174 km) of R26 exhibit the largest trend of about $+5$ km decade^{-1}. A similar pattern—median to larger quantiles increase the most—is also found for R33, but the trends are weaker ($\sim$0.5%–1.5%; Fig. 13c). In contrast to the abovementioned result, the TC inner-core size RMW exhibits significant negative trends in all quantiles, except for the 90th percentile. In all, the results imply that for the WNP basin over the period 1981–2017, there are expanding trends in the TC outer-core wind radii, while the inner-core size RMW shows a contracting trend, confirming the results calculated with annual-mean wind radii (Fig. 11).

Figure 14 further illustrates the trends for outer- to inner-core wind radii for 1981–2017 WNP TCs cases categorized by different intensity scales. It is interesting to see that TCs with different intensity scales have different changes in their wind structure. The category 0 cases (i.e., tropical storm; 17–32 m s^{-1}) were identified to have expanding trends in the whole wind field from RMW to R26 and R17. The outer circulations of the tropical storms have significantly increased by $+4\%$ decade^{-1} (=8.2 km decade^{-1}; $p = 0.002$) for R17 and $+5\%$ decade^{-1} (=5.3 km decade^{-1}; $p = 0.003$) for R26. In comparisons, the category 1–2 cases (32–48 m s^{-1}) have a significant contracting trend of -3% decade^{-1} ($p = 0.08$) for RMW, while insignificant trends for outer-core wind radii (R17, R26, and R33) were observed. This result also indicates that the upward trends in R17
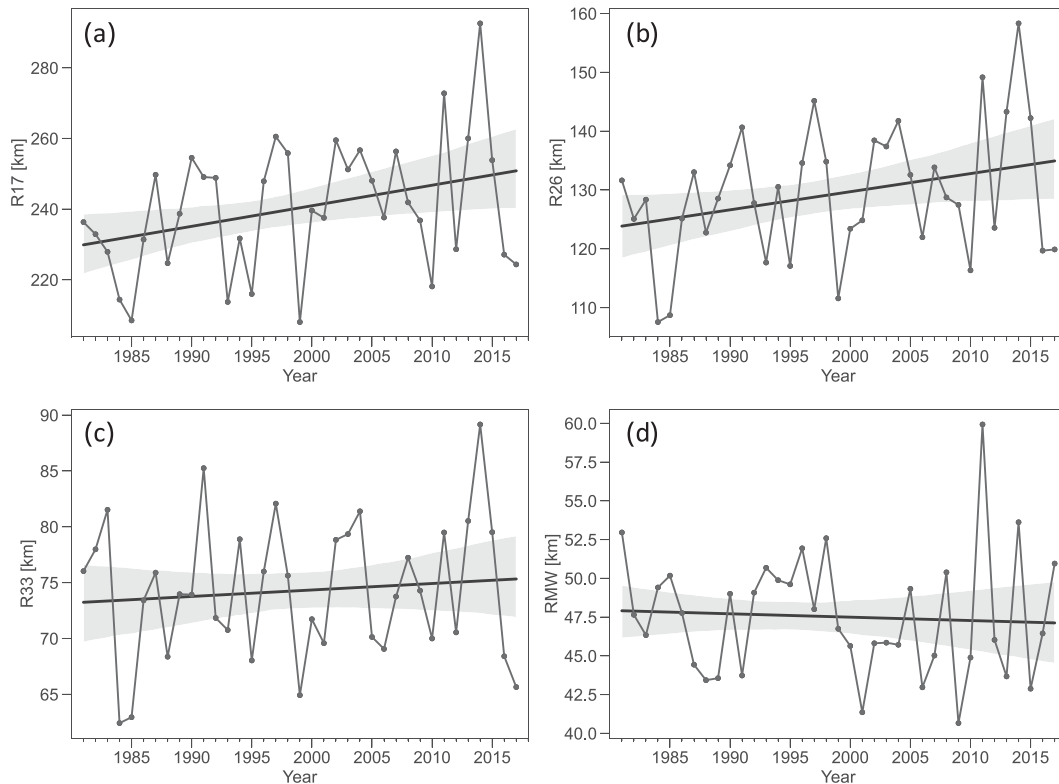
FIG. 11. Time series of annual-mean (a) R17, (b) R26, (c) R33, and (d) RMW over the 37-yr period (1981–2017) for the WNP TCs. Linear trend lines (black line) are shown with their 90% confidence intervals (shaded area).

and R26 (Fig. 12) are primarily associated with weaker storms. As for the very intense cases (category 3–5; ≥48 m s^{-1}), they tend to have no detectable changes in their wind structure. In addition, we have further performed partial correlations to test the influence of latitude since poleward shifts were found for the maximum TC intensity in the WNP (e.g., Kossin et al. 2016), showing that controlling the latitude (calculated from JTWC best tracks) does not change the statistical significances listed above.

In all, our results confirm that the model projections of outer TC size over the WNP are generally on the order of 10% or less (e.g., Kim et al. 2014; Knutson et al. 2015; Yamada et al. 2017). Given that the intensities of TCs are usually accompanied by the expansion of the outer circulations and/or contraction of RMW, as suggested by the physical model (Chavas et al. 2015), the changes in the TC wind field structure based on DeepTCSize are likely to be consistent with the expectations of increasing TC intensity with global warming, as predicted by potential intensity theory (e.g., Holland and Bruyère 2014;

Sobel et al. 2016). However, since this is hypothetical and there is no supporting evidence, it can be said that the observed changes in TC sizes over the WNP using DeepTCSize cannot be fully understood from the trends. Consequently, additional work such as climate model simulations is required.

6. Summary and discussion

This study has utilized deep learning methods to improve the construction of historical TC size datasets over the WNP. The TC size estimation algorithms were developed based on DeepTCNet (a physics-augmented deep learning model designed to infer critical intensity and size information of TCs from IR imagery) (Zhuo and Tan 2021). Transfer learning was initially carried out to establish reliable TC size estimation algorithms suited to the WNP TCs. The algorithms were then applied to a homogenized (spatiotemporal) record of satellite imagery to generate a long-term historical TC size dataset named DeepTCSize. DeepTCSize includes multiple TC size quantities such as R17, R26, R33, and RMW that depict the complete radial wind structure of a TC, and provides a large amount of data (3-hourly; $N = 37\,480$) for TCs over the WNP basin (100°E–180° and 0°–60°N) from 1981 to 2017. The mean values of R17, R26, R33, and RMW are about 240, 130, 75, and 50 km, respectively.

Using two years of coincident samples (2016–17; $N = 1545$), DeepTCSize's data quality was evaluated and high correlations

TABLE 2. Linear trends of annual-mean TC sizes over 1981–2017, calculated from DeepTCSize. Statistics are shown as trend ± standard error, and statistical significance is indicated in bold at the 90% confidence interval.

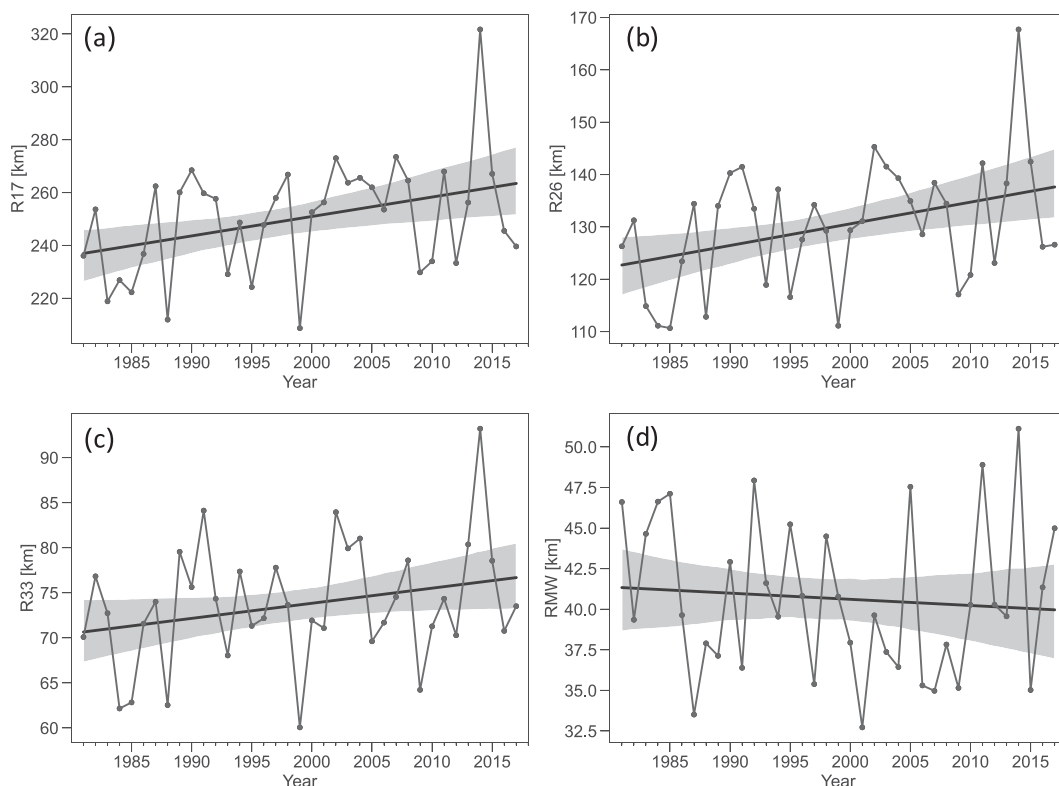|	R17	R26	R33	RMW
All cases	**5.8 ± 2.7**	**3.1 ± 1.7**	0.6 ± 1.0	−0.2 ± 0.6
LMI cases	**7.4 ± 3.1**	**4.1 ± 1.7**	1.7 ± 1.0	−0.4 ± 0.7

Fig. 12. As in Fig. 11, but for the TCs that achieved their lifetime maximum intensity.

(0.76 ∼ 0.85) were found for each of the wind radii parameter in comparison to the quality-controlled best track data. Among all TC size quantities in DeepTCSize, R17 yields the best accuracy in terms of the highest correlations and the lowest uncertainty ($R = 0.85$; SI = 24%). Comparisons with the remotely sensed TC ocean winds from multiple sources further increased the confidence in the quality of DeepTCSize. More importantly, the homogeneity of DeepTCSize is also verified, as near-zero trends were found when compared with the historical ocean wind observations of TC. Summarily, given the good agreement between DeepTCSize and the "best estimates" so far available from JTWC in recent years (the current state of observational capability), the method proposed in this paper can be simply regarded as applying deep learning to extend the current state-of-the-art quality of TC size data to the whole satellite era with consistency.

In this study, we have demonstrated that DeepTCSize fills a large part of the significant data gap reported in the WNP where there has been the absence of reliable long-term observed TC size data. Hence, a variety of potential applications are enabled as a result of this data infilling. For example, we have used DeepTCSize as a baseline to investigate the defective data quality and buried data issues in the best track records maintained by the JMA and JTWC. In our investigations, a significant data discrepancy was observed among these datasets. Particularly, the best track records were found to include spurious trends that were proven to be large enough to even reverse the signs of size changes over the

WNP. This clearly suggests great caution must be exercised when the best track data for TC size is being used for climatological studies. In particular, the past findings based on the best tracks for trend analysis ought to be reconsidered.

Using DeepTCSize, the trends of WNP TC size have been investigated in this study. Our investigations reveal significant increasing trends in the annual-mean TC outer-core size R17 of about +5.8 km (2%) per decade and +3.1 km (2%) for R26 for the WNP TCs over 1981–2017, and no significant trends were identified for R33 and RMW. The TC cases that reach their lifetime maximum intensity (LMI) were found to exhibit larger amplitudes of trends for each size parameter. This is an indication of a potentially higher risk of damage that could be caused by TCs over the WNP, regardless of whether their LMI remain unchanged or even decrease. In addition, we have also applied a more rigorous method (quantile regression) for the trend analysis of WNP TC size. Our findings reveal that there is a significant expanding trend for the TC outer-core sizes (R17, R26, and R33), while the TC inner-core size (RMW) exhibits a contracting trend. And results also show that the WNP TCs with different intensity scales exhibit different trends in their wind structure. The weak tropical-storm-scale storms were identified to have significant expanding trends in the whole wind field. Category 1–2 cases exhibit a significant decreasing trend in the RMW but remain unchanged in their outer circulations. In comparison, the very intense storms tend to have no detectable changes in their wind structure. Our result also suggests that upward trends in R17 and R26 are
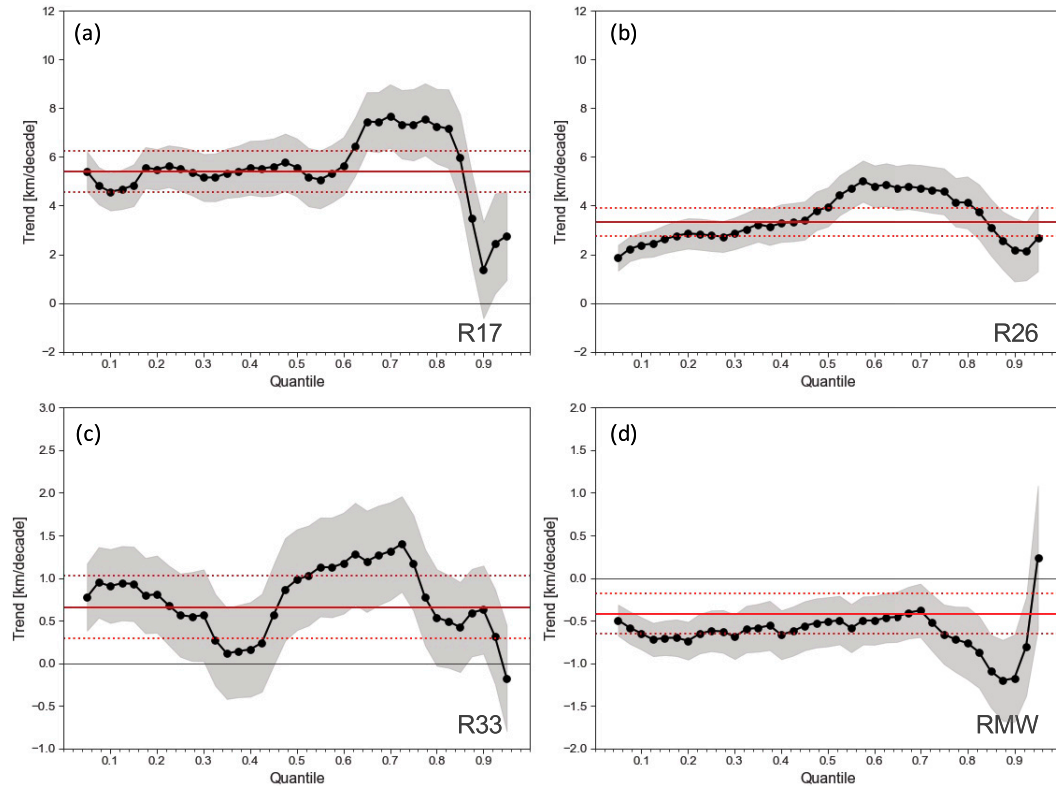
FIG. 13. Trends in the quantiles of (a) R17, (b) R26, (c) R33, and (d) RMW over the 37-yr period (1981–2017) calculated with the TCs as Fig. 11. The black dots represent the trends in the quantiles of each TC size parameter from 0.05 to 0.95 in steps of 0.025. Shading represents pointwise 95% confidence (two-tailed). The red solid line shows the (constant value) trend in the mean as measured by ordinary least squares regression, and the red dashed lines show the confidence interval.

primarily shown to be associated with weaker storms. Given that it may turn out that global warming creates more opportunities for tropical storm formation and higher peak intensities in the WNP, it is essential to investigate the relationship between climate change and TC size, which is a good proxy for rainfall potential.

Since this study is limited to the WNP basin, there is still a need to investigate the global observed TC size climatology. This is because data availability and policies vary from agency to agency over time and the prospect of having a reliable global homogenized reanalysis of TC size is challenging. Presently, the reconstruction of historical TC size records for other basins like the Southern Hemisphere and the north Indian Ocean still requires a careful design of algorithms, including the application of deep learning methodologies with transfer learning as introduced and carried out in this study. In other words, the application of deep learning could help hasten the realization of reliable global homogenized TC size data. This will be more realistic if a less strict data precision standard is considered such as sacrificing accuracy for consistency, as past studies did (e.g., Kossin et al. 2007). In this study, we have applied transfer learning to adapt the DeepTCNet from the North Atlantic to the western North Pacific to obtain relative MAE reduction of 10%, 18%, 14%, and

36% for R17, R26, R33, and RMW, respectively. It is good to note that there is still room for improvement in the performance of the DeepTCNet-based TC size estimation and reanalysis. In other words, it can be expected that as more reliable observations become available for the TC wind field in the future, DeepTCNet could be retrained with these data, and then the improved capability of observations can be deployed to cover more areas via applying DeepTCNet to a reanalysis of historical satellite data.

With the availability of long-term climate datasets for TC size, many studies can be conducted in the future. For example, a more comprehensive study of TC size climatology, such as those previously performed using scarce QuikSCAT data (e.g., Chan and Chan 2015; Chavas et al. 2016), can be explored with DeepTCSize. Additionally, a new dataset of the complete radial wind structure of TCs can be generated by combining DeepTCSize with TC wind structure models (e.g., Chavas et al. 2015). Such a dataset will allow for a holistic analysis of radial wind structure and ultimately the enhancement of the general understanding of the nature of the TC wind structure. For the readers who are interested in studying TC asymmetries, one can use the reconstructed azimuthal-mean wind radii data together with storm motion and location to further fit for four-quadrant wind radii by using the vortex
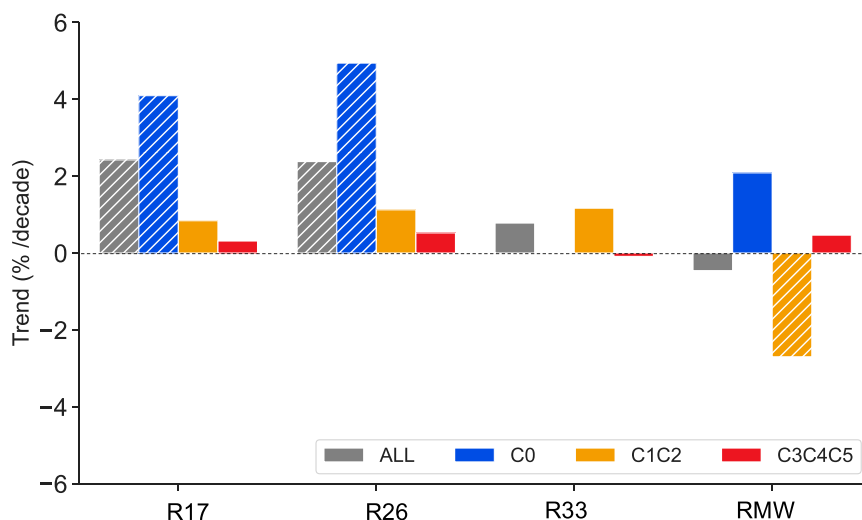
FIG. 14. Per-decade linear trends of annual-mean outer- to inner-core wind radii over 1971–2017 for WNP named storms (gray), Saffir–Simpson category 0 cases (i.e., tropical storm; blue), category 1–2 cases (orange), and category 3–5 cases (red). The 90% significant trends are displayed with "///" hatches.

methodology (Knaff et al. 2016) or some simple regression model. In addition, DeepTCSize can also be used for theoretical comparisons aimed at the development of forecasting algorithms. It can also be used for case studies of specific systems, especially systems for which traditional observations of TC size may have been missing or uncertain in the past.

REFERENCES

Barnes, E. A., B. Toms, J. W. Hurrell, I. Ebert-Uphoff, C. Anderson, and D. Anderson, 2020: Indicator patterns of forced change learned by an artificial neural network. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002195, https://doi.org/10.1029/2020MS002195.

Bian, G.-F., G.-Z. Nie, and X. Qiu, 2021: How well is outer tropical cyclone size represented in the ERA5 reanalysis dataset? *Atmos. Res.*, **249**, 105339, https://doi.org/10.1016/j.atmosres.2020.105339.

Brennan, M. J., C. C. Hennon, and R. D. Knabb, 2009: The operational use of QuikSCAT ocean surface vector winds at the National Hurricane Center. *Wea. Forecasting*, **24**, 621–645, https://doi.org/10.1175/2008WAF2222188.1.

Callaghan, M., and Coauthors, 2021: Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nat. Climate Change*, **11**, 966–972, https://doi.org/10.1038/s41558-021-01168-6.

Chan, K. T. F., and J. C. L. Chan, 2012: Size and strength of tropical cyclones as inferred from QuikSCAT data. *Mon. Wea. Rev.*, **140**, 811–824, https://doi.org/10.1175/MWR-D-10-05062.1.

——, and ——, 2015: Global climatology of tropical cyclone size as inferred from QuikSCAT data. *Int. J. Climatol.*, **35**, 4843–4848, https://doi.org/10.1002/joc.4307.

——, and ——, 2018: The outer-core wind structure of tropical cyclones. *J. Meteor. Soc. Japan*, **96**, 297–315, https://doi.org/10.2151/jmsj.2018-042.

Chavas, D. R., and K. A. Emanuel, 2010: A QuikSCAT climatology of tropical cyclone size. *Geophys. Res. Lett.*, **37**, L18816, https://doi.org/10.1029/2010GL044558.

——, and J. Vigh, 2014: QSCAT-R: The QuikSCAT tropical cyclone radial structure dataset. NCAR Tech. Note NCAR/TN5131+STR, 25 pp., https://doi.org/10.5065/D6J67DZ4.

——, and K. Emanuel, 2014: Equilibrium tropical cyclone size in an idealized state of axisymmetric radiative–convective equilibrium. *J. Atmos. Sci.*, **71**, 1663–1680, https://doi.org/10.1175/JAS-D-13-0155.1.

——, and J. A. Knaff, 2022: A simple model for predicting the tropical cyclone radius of maximum wind from outer size. *Wea. Forecasting*, **37**, 563–579, https://doi.org/10.1175/WAF-D-21-0103.1.

——, N. Lin, and K. Emanuel, 2015: A model for the complete radial structure of the tropical cyclone wind field. Part I:

Comparison with observed structure. *J. Atmos. Sci.*, **72**, 3647–3662, https://doi.org/10.1175/JAS-D-15-0014.1.

——, ——, W. Dong, and Y. Lin, 2016: Observed tropical cyclone size revisited. *J. Climate*, **29**, 2923–2939, https://doi.org/10.1175/JCLI-D-15-0731.1.

Chen, B.-F., B. Chen, H.-T. Lin, and R. L. Elsberry, 2019: Estimating tropical cyclone intensity by satellite imagery utilizing convolutional neural networks. *Wea. Forecasting*, **34**, 447–465, https://doi.org/10.1175/WAF-D-18-0136.1.

Combot, C., A. Mouche, J. Knaff, Y. Zhao, Y. Zhao, L. Vinour, Y. Quilfen, and B. Chapron, 2020: Extensive high-resolution synthetic aperture radar (SAR) data analysis of tropical cyclones: Comparisons with SFMR flights and best track. *Mon. Wea. Rev.*, **148**, 4545–4563, https://doi.org/10.1175/MWR-D-20-0005.1.

Demuth, J. L., M. DeMaria, and J. A. Knaff, 2006: Improvement of Advanced Microwave Sounding Unit tropical cyclone intensity and size estimation algorithms. *J. Appl. Meteor. Climatol.*, **45**, 1573–1581, https://doi.org/10.1175/JAM2429.1.

Emanuel, K. A., 1986: An air–sea interaction theory for tropical cyclones. Part I: Steady-state maintenance. *J. Atmos. Sci.*, **43**, 585–605, https://doi.org/10.1175/1520-0469(1986)043<0585:AASITF>2.0.CO;2.

——, 2000: A statistical analysis of tropical cyclone intensity. *Mon. Wea. Rev.*, **128**, 1139–1152, https://doi.org/10.1175/1520-0493(2000)128<1139:ASAOTC>2.0.CO;2.

Emanuel, K., and Coauthors, 2018: On the desirability and feasibility of a global reanalysis of tropical cyclones. *Bull. Amer. Meteor. Soc.*, **99**, 427–429, https://doi.org/10.1175/BAMS-D-17-0226.1.

Figa-Saldaña, J., J. J. W. Wilson, E. Attema, R. Gelsthorpe, M. R. Drinkwater, and A. Stoffelen, 2002: The Advanced Scatterometer (ASCAT) on the meteorological operational (MetOp) platform: A follow on for European wind scatterometers. *Can. J. Remote Sens.*, **28**, 404–412, https://doi.org/10.5589/m02-035.

Frank, W. M., 1977: The structure and energetics of the tropical cyclone. I: Storm structure. *Mon. Wea. Rev.*, **105**, 1119–1135, https://doi.org/10.1175/1520-0493(1977)105<1119:TSAEOT>2.0.CO;2.

Guo, X., and Z.-M. Tan, 2017: Tropical cyclone fullness: A new concept for interpreting storm intensity. *Geophys. Res. Lett.*, **44**, 4324–4331, https://doi.org/10.1002/2017GL073680.

——, and ——, 2022: Tropical cyclone intensification and fullness: The role of storm size configuration. *Geophys. Res. Lett.*, **49**, e2022GL098449, https://doi.org/10.1029/2022GL098449.

Guzman, O., and H. Jiang, 2021: Global increase in tropical cyclone rain rate. *Nat. Commun.*, **12**, 5344, https://doi.org/10.1038/s41467-021-25685-2.

Ham, Y.-G., J.-H. Kim, and J.-J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572, https://doi.org/10.1038/s41586-019-1559-7.

Holland, G., and C. L. Bruyère, 2014: Recent intense hurricane response to global climate change. *Climate Dyn.*, **42**, 617–627, https://doi.org/10.1007/s00382-013-1713-0.

Horstmann, J., S. Falchetti, C. Wackerman, S. Maresca, M. J. Caruso, and H. C. Graber, 2015: Tropical cyclone winds retrieved from C-band cross-polarized synthetic aperture radar. *IEEE Trans. Geosci. Remote Sens.*, **53**, 2887–2898, https://doi.org/10.1109/TGRS.2014.2366433.

Irish, J. L., and D. T. Resio, 2010: A hydrodynamics-based surge scale for hurricanes. *Ocean Eng.*, **37**, 69–81, https://doi.org/10.1016/j.oceaneng.2009.07.012.

——, ——, and J. J. Ratcliff, 2008: The influence of storm size on hurricane surge. *J. Phys. Oceanogr.*, **38**, 2003–2013, https://doi.org/10.1175/2008JPO3727.1.

JTWC, 2016: Annual tropical cyclone report 2016. Joint Typhoon Warning Center Tech. Rep., 122 pp., https://www.metoc.navy.mil/jtwc/products/atcr/2016atcr.pdf.

——, 2017: Annual tropical cyclone report 2017. Joint Typhoon Warning Center Tech. Rep., 133 pp., https://www.metoc.navy.mil/jtwc/products/atcr/2017atcr.pdf.

Kadow, C., D. M. Hall, and U. Ulbrich, 2020: Artificial intelligence reconstructs missing climate information. *Nat. Geosci.*, **13**, 408–413, https://doi.org/10.1038/s41561-020-0582-5.

Karniadakis, G. E., I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, 2021: Physics-informed machine learning. *Nat. Rev. Phys.*, **3**, 422–440, https://doi.org/10.1038/s42254-021-00314-5.

Khairoutdinov, M., and K. Emanuel, 2013: Rotating radiative-convective equilibrium simulated by a cloud-resolving model. *J. Adv. Model. Earth Syst.*, **5**, 816–825, https://doi.org/10.1002/2013MS000253.

Kim, H.-J., I.-J. Moon, and I. Oh, 2022: Comparison of tropical cyclone wind radius estimates between the KMA, RSMC Tokyo, and JTWC. *Asia-Pacific. J. Atmos. Sci.*, **58**, 563–576, https://doi.org/10.1007/s13143-022-00274-5.

Kim, H.-S., G. A. Vecchi, T. R. Knutson, W. G. Anderson, T. L. Delworth, A. Rosati, F. Zeng, and M. Zhao, 2014: Tropical cyclone simulation and response to CO_2 doubling in the GFDL CM2.5 high-resolution coupled climate model. *J. Climate*, **27**, 8034–8054, https://doi.org/10.1175/JCLI-D-13-00475.1.

Kimball, S. K., and M. S. Mulekar, 2004: A 15-year climatology of North Atlantic tropical cyclones. Part I: Size parameters. *J. Climate*, **17**, 3555–3575, https://doi.org/10.1175/1520-0442(2004)017<3555:AYCONA>2.0.CO;2.

Knaff, J. A., and C. R. Sampson, 2015: After a decade are Atlantic tropical cyclone gale force wind radii forecasts now skillful? *Wea. Forecasting*, **30**, 702–709, https://doi.org/10.1175/WAF-D-14-00149.1.

——, M. DeMaria, D. A. Molenar, C. R. Sampson, and M. G. Seybold, 2011: An automated, objective, multiple-satellite-platform tropical cyclone surface wind analysis. *J. Appl. Meteor. Climatol.*, **50**, 2149–2166, https://doi.org/10.1175/2011JAMC2673.1.

——, S. P. Longmore, and D. A. Molenar, 2014: An objective satellite-based tropical cyclone size climatology. *J. Climate*, **27**, 455–476, https://doi.org/10.1175/JCLI-D-13-00096.1.

——, ——, R. T. DeMaria, and D. A. Molenar, 2015: Improved tropical-cyclone flight-level wind estimates using routine infrared satellite reconnaissance. *J. Appl. Meteor. Climatol.*, **54**, 463–478, https://doi.org/10.1175/JAMC-D-14-0112.1.

——, C. J. Slocum, K. D. Musgrave, C. R. Sampson, and B. R. Strahl, 2016: Using routinely available information to estimate tropical cyclone wind structure. *Mon. Wea. Rev.*, **144**, 1233–1247, https://doi.org/10.1175/MWR-D-15-0267.1.

——, and Coauthors, 2021: Estimating tropical cyclone surface winds: Current status, emerging technologies, historical evolution, and a look to the future. *Trop. Cyclone Res. Rev.*, **10**, 125–150, https://doi.org/10.1016/j.tcrr.2021.09.002.

Knapp, K. R., and J. Kossin, 2007: New global tropical cyclone data set from ISCCP B1 geostationary satellite observations. *J. Appl. Remote Sens.*, **1**, 013505, https://doi.org/10.1117/1.2712816.

——, M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Amer. Meteor. Soc.*, **91**, 363–376, https://doi.org/10.1175/2009BAMS2755.1.

——, and Coauthors, 2011: Globally gridded satellite observations for climate studies. *Bull. Amer. Meteor. Soc.*, **92**, 893–907, https://doi.org/10.1175/2011BAMS3039.1.

Knutson, T. R., J. J. Sirutis, M. Zhao, R. E. Tuleya, M. Bender, G. A. Vecchi, G. Villarini, and D. Chavas, 2015: Global projections of intense tropical cyclone activity for the late twenty-first century from dynamical downscaling of CMIP5/RCP4.5 scenarios. *J. Climate*, **28**, 7203–7224, https://doi.org/10.1175/JCLI-D-15-0129.1.

——, and Coauthors, 2020: Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bull. Amer. Meteor. Soc.*, **101**, E303–E322, https://doi.org/10.1175/BAMS-D-18-0194.1.

Kossin, J. P., J. A. Knaff, H. I. Berger, D. C. Herndon, T. A. Cram, C. S. Velden, R. J. Murnane, and J. D. Hawkins, 2007: Estimating hurricane wind structure in the absence of aircraft reconnaissance. *Wea. Forecasting*, **22**, 89–101, https://doi.org/10.1175/WAF985.1.

——, T. L. Olander, and K. R. Knapp, 2013: Trend analysis with a new global record of tropical cyclone intensity. *J. Climate*, **26**, 9960–9976.

——, K. A. Emanuel, and G. A. Vecchi, 2014: The poleward migration of the location of tropical cyclone maximum intensity. *Nature*, **509**, 349–352, https://doi.org/10.1038/nature13278.

——, ——, and S. J. Camargo, 2016: Past and projected changes in western North Pacific tropical cyclone exposure. *J. Climate*, **29**, 5725–5739, https://doi.org/10.1175/JCLI-D-16-0076.1.

——, K. R. Knapp, T. L. Olander, and C. S. Velden, 2020: Global increase in major tropical cyclone exceedance probability over the past four decades. *Proc. Natl. Acad. Sci.*, **117**, 11975, https://doi.org/10.1073/pnas.1920849117.

Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592, https://doi.org/10.1175/MWR-D-12-00254.1.

LeCun, Y., Y. Bengio, and G. Hinton, 2015: Deep learning. *Nature*, **521**, 436–444, https://doi.org/10.1038/nature14539.

Lee, C.-S., K. K. W. Cheung, W.-T. Fang, and R. L. Elsberry, 2010: Initial maintenance of tropical cyclone size in the western North Pacific. *Mon. Wea. Rev.*, **138**, 3207–3223, https://doi.org/10.1175/2010MWR3023.1.

Lin, Y., M. Zhao, and M. Zhang, 2015: Tropical cyclone rainfall area controlled by relative sea surface temperature. *Nat. Commun.*, **6**, 6591, https://doi.org/10.1038/ncomms7591.

Lu, K.-Y., and D. R. Chavas, 2022: Tropical cyclone size is strongly limited by the Rhines scale: Experiments with a barotropic model. *J. Atmos. Sci.*, **79**, 2109–2124, https://doi.org/10.1175/JAS-D-21-0224.1.

Lungu, T., and P. S. Callahan, 2006: QuikSCAT science data product user's manual: overview and geophysical data products, version 3.0. JPL Tech. Rep. D-18053, 91 pp., https://podaac-tools.jpl.nasa.gov/drive/files/allData/quikscat/L2A/v2/docs/QSUG_v3.pdf.

Meissner, T., L. Ricciardulli, and F. J. Wentz, 2017: Capability of the SMAP mission to measure ocean surface winds in storms. *Bull. Amer. Meteor. Soc.*, **98**, 1660–1677, https://doi.org/10.1175/BAMS-D-16-0052.1.

——, ——, and A. Manaster, 2021: Tropical cyclone winds from WindSat, AMSR2, and SMAP: Algorithm development and testing. *Remote Sens.*, **13**, 1641, https://doi.org/10.3390/rs13091641.

Mouche, A. A., B. Chapron, B. Zhang, and R. Husson, 2017: Combined co-and cross-polarized SAR measurements under extreme wind conditions. *IEEE Trans. Geosci. Remote Sens.*, **55**, 6746–6755, https://doi.org/10.1109/TGRS.2017.2732508.

Mueller, K. J., M. DeMaria, J. Knaff, J. P. Kossin, and T. H. Vonder Haar, 2006: Objective estimation of tropical cyclone wind structure from infrared satellite data. *Wea. Forecasting*, **21**, 990–1005, https://doi.org/10.1175/WAF955.1.

Muroi, C., 2018: Brief history and recent activities of RSMC Tokyo Typhoon Centre. *Trop. Cyclone Res. Rev.*, **7**, 57–64, https://doi.org/10.6057/2018TCRR01.09.

Pan, S. J., and Q. Yang, 2010: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.*, **22**, 1345–1359, https://doi.org/10.1109/TKDE.2009.191.

Powell, M. D., and T. A. Reinhold, 2007: Tropical cyclone destructive potential by integrated kinetic energy. *Bull. Amer. Meteor. Soc.*, **88**, 513–526, https://doi.org/10.1175/BAMS-88-4-513.

Racah, E., C. Beckham, T. Maharaj, S. E. Kahou, Prabhat, and C. Pa, 2017: Extreme weather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *31st Conf. on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, Association for Computing Machinery, 3405–3416, https://dl.acm.org/doi/10.5555/3294996.3295099.

Reul, N., and Coauthors, 2017: A new generation of tropical cyclone size measurements from space. *Bull. Amer. Meteor. Soc.*, **98**, 2367–2385, https://doi.org/10.1175/BAMS-D-15-00291.1.

Sampson, C. R., E. M. Fukada, J. A. Knaff, B. R. Strahl, M. J. Brennan, and T. Marchok, 2017: Tropical cyclone gale wind radii estimates for the western North Pacific. *Wea. Forecasting*, **32**, 1029–1040, https://doi.org/10.1175/WAF-D-16-0196.1.

Schenkel, B. A., and R. E. Hart, 2012: An examination of tropical cyclone position, intensity, and intensity life cycle within atmospheric reanalysis datasets. *J. Climate*, **25**, 3453–3475, https://doi.org/10.1175/2011JCLI4208.1.

——, N. Lin, D. Chavas, M. Oppenheimer, and A. Brammer, 2017: Evaluating outer tropical cyclone size in reanalysis datasets using QuikSCAT data. *J. Climate*, **30**, 8745–8762, https://doi.org/10.1175/JCLI-D-17-0122.1.

——, ——, ——, G. A. Vecchi, M. Oppenheimer, and A. Brammer, 2018: Lifetime evolution of outer tropical cyclone size and structure as diagnosed from reanalysis and climate model data. *J. Climate*, **31**, 7985–8004, https://doi.org/10.1175/JCLI-D-17-0630.1.

Sobel, A. H., S. J. Camargo, T. M. Hall, C.-Y. Lee, M. K. Tippett, and A. A. Wing, 2016: Human influence on tropical cyclone intensity. *Science*, **353**, 242–246, https://doi.org/10.1126/science.aaf6574.

Song, J., and P. J. Klotzbach, 2016: Wind structure discrepancies between two best track datasets for western North Pacific tropical cyclones. *Mon. Wea. Rev.*, **144**, 4533–4551, https://doi.org/10.1175/MWR-D-16-0163.1.

Velden, C., and Coauthors, 2006: The Dvorak tropical cyclone intensity estimation technique: A satellitebased method that has endured for over 30 years. *Bull. Amer. Meteor. Soc.*, **87**, 1195–1210, https://doi.org/10.1175/BAMS-87-9-1195.

Weatherford, C. L., and W. M. Gray, 1988: Typhoon structure as revealed by aircraft reconnaissance. Part I: Data analysis and climatology. *Mon. Wea. Rev.*, **116**, 1032–1043, https://doi.org/10.1175/1520-0493(1988)116<1032:TSARBA>2.0.CO;2.

Wimmers, A., C. Velden, and J. H. Cossuth, 2019: Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery. *Mon. Wea. Rev.*, **147**, 2261–2282, https://doi.org/10.1175/MWR-D-18-0391.1.

Yamada, Y., M. Satoh, M. Sugi, C. Kodama, A. T. Noda, M. Nakano, and T. Nasuno, 2017: Response of tropical cyclone activity and structure to global warming in a high-resolution global nonhydrostatic model. *J. Climate*, **30**, 9703–9724, https://doi.org/10.1175/JCLI-D-17-0068.1.

Zhuo, J.-Y., and Z.-M. Tan, 2021: Physics-augmented deep learning to improve tropical cyclone intensity and size estimation from satellite imagery. *Mon. Wea. Rev.*, **149**, 2097–2113, https://doi.org/10.1175/MWR-D-20-0333.1.