Journal of the Meteorological Society of Japan

EARLY ONLINE RELEASE

The DOI for this manuscript is
DOI:10.2151/jmsj.2022-048
J-STAGE Advance published date: August 30th, 2022
The final manuscript after publication will replace the preliminary version at the above DOI once it is available.

Tropical cyclone size identification over the Western North Pacific using support vector machine and general regression neural network

Xiaoqin LU[1]

Shanghai Typhoon Institute and Key Laboratory of Numerical Modeling for Tropical Cyclone, China Meteorological Administration, Shanghai, China

Wai-kin WONG

Hong Kong Observatory, Hong Kong, China

Hui YU

Shanghai Typhoon Institute and Key Laboratory of Numerical Modeling for Tropical Cyclone, China Meteorological Administration, Shanghai, China

Xiaoming YANG

Shanghai Ocean University, Shanghai, China

November 22, 2021

----

1) Corresponding author: Mrs. Xiaoqin Lu, Key Laboratory of Numerical Modeling for Tropical Cyclones and Shanghai Typhoon Institute, China Meteorological Administration, 166, Puxi Road, Shanghai 200030, China.
E-mail: luxq@typhoon.org.cn
Tel: +86-21-54896094
Fax: +86-21-64391966

Abstract

Knowledge about tropical cyclone (TC) size is essential for disaster prevention and mitigation strategies, but due to the limitations of observations, TC size data from the open ocean are scarce. In this paper, several models are developed to identify TC size parameters, including the radius of maximum wind (RMW) and the radii of 34 (R34), 50 (R50), and 64 (R64) knot winds, using various machine learning algorithms based on infrared channel imagery of geostationary meteorological satellites over the Western North Pacific (WNP). Through evaluation and verification, the trained and optimized support vector machine models are proposed for RMW and R34, while the general regression neural network models are set up for R50 and R64.

According to the independent-sample evaluations against aircraft observations (1981–1987) / Joint Typhoon Warning Center best track data (2017–2019), the mean absolute errors of R34, R50, R64, and RMW are 54 / 58, 34 / 38, N/A / 21, and 25 / 25 km, respectively. The corresponding median errors are 39 / 46, 34 / 31, N/A / 17, and 17 / 19 km, respectively. There is an overall slight underestimation of the parameters, which needs to be analyzed and improved in future study. Despite aircraft observations of TCs in the WNP having ceased in the late 1980s, this new dataset of TC sizes enables a thorough estimation of wind structures covering a period of 40 years.

Key words tropical cyclone size; geostationary satellite; machine learning method; Western North Pacific

1 Introduction

Tropical cyclone (TC) intensity and size are two key factors to determine its destructiveness (Guo and Tan 2017). Cocks and Gray (2002) emphasized that the wind strength and spatial coverage of the TC outer circulation, rather than its central position and intensity, determine the overall risk of disaster due to TC. Therefore, research on estimating and forecasting TC size is undoubtedly essential for disaster prevention and mitigation strategies. Due to limitations in monitoring methods, TC size information is often obtained indirectly. At present, measurements of TC structure are mostly carried out in the Atlantic Ocean due to routine aircraft observations in the western part of this ocean basin (Kossin et al. 2007). Elsewhere, in-situ observations of TCs are mainly from ships, buoys, and meteorological stations on islands in various ocean basins, thus TC size data is very scarce in the open sea. Consequently, TC data generally describe the location and intensity of the TC center, but the description of TC size is rather limited. In the Western North Pacific (WNP; including the South China Sea), only the Regional Specialized Meteorological Center in Tokyo includes the major and minor axis of TC wind ellipses, whilst the Joint Typhoon Warning Center (JTWC) of the US Navy has issued the wind circle radius since 2001, including the wind radii of 34-kt, 50-kt and 64-kt surface winds (R34, R50, and R64) in four quadrants, as well as the radius of maximum surface winds (RMW). However, the above wind radii are generally analyzed subjectively (Knaff et al. 2016) and details of the TC size estimation methodology are unclear.

Various approaches have been employed to investigate TC size, including using synoptic charts (Brand 1972; Merrill 1984), a combination of aircraft and ground observations (Shea and Gray 1973; Weatherford and Gray 1988a and 1988B; Croxford and Barnes 2002; Cocks and Gray 2002), best track data (Lu et al. 2011; Xu and Wang 2015 and 2018; Guo and Tan 2017; Lin and Chou 2018), model reanalysis datasets (McKenzie 2017; Schenkel et al. 2017 and 2018), and satellite observations (Liu and Chan 1999; Lee et al. 2010; Chan and Chan 2012 and 2015; Knaff et al. 2014 and 2016; Wu et al. 2015; Lu et al. 2017), amongst others. The results are different from each other, but they do show a certain degree of consensus in characteristics such as seasonal variations and geographical differences in TC size. However, due to the different analysis data and size definitions (McKenzie 2017), the spatiotemporal characteristics and size variation over long periods remain uncertain.

Satellite data is a primary choice for TC size analysis given the higher coverage in both space and time compared with in-situ measurements from conventional observation platforms. Many studies have used spaceborne scatterometer observations directly to describe TC size and establish TC size datasets (Liu and Chan 1999; Chavas and Emanuel 2010; Chan and Chan 2015). However, the retrieved wind from the scatterometer has a poor temporal resolution, and the accuracy of wind retrieval decreases when the wind speed is more than 30 m s^{-1} (Knaff et al., 2011). Therefore, geostationary satellite observations with high spatiotemporal resolution have become preferable for operational applications. At the same time, geostationary meteorological satellites have the ability to capture a whole TC (Mueller et al. 2006) and can therefore provide better data for analyzing the fine structural features of TCs.

Demuth et al. (2004 and 2006) applied advanced microwave sounding unit (AMSU) retrieved wind and model parameters to estimate TC size. The mean absolute errors (MAEs) of the R34, R50, and R64 were 16.9, 13.3, and 6.8 nautical miles, respectively. Combining the basic TC information (center intensity and location), Mueller et al. (2006) used the infrared (IR) band of geostationary meteorological satellites and aircraft observations to establish a multiple linear regression algorithm that could estimate the RMW of a TC. The MAE was 27.3 km. Using IR observations, a regression model was established for estimating the TC intensity (Maximum Sustained Wind, MSW), R34, R50, R64, and RMW based on the mean radial profile and the principal mode of the empirical orthogonal function (EOF) of the brightness temperature (BT) outside the TC center (Kossin et al. 2007). The estimated MAEs of the R34, R50, R64 and RMW were 44.8, 36.6, 26.9, and 21.1 km, respectively. It was found that including IR observation data can reduce the estimation error in multivariate linear models. Lajoie and Walsh (2008) estimated the TC eye wall structure (radius of the TC eye and RMW) using satellite cloud images, radar, and aircraft observations. Compared with aircraft observations, the MAE of the RMW was 2.8 km, which is better than that of Kossin et al. (2007). The sample size in the above studies was relatively small, and the estimation method involved utilizing multi-platform observations (Kossin et al., 2007), including satellite IR imagery, radar, and aircraft observations. Therefore, the method is not easily applicable in operational use, especially for some agencies that find it difficult to obtain multi-platform observations in real time.

Knaff et al. (2011, 2014, 2016) successively developed a TC surface wind field retrieval algorithm (Multiple satellite platform Tropical Cyclone Surface Wind Analysis, MTCSWA) integrated with multi-satellite observations and an objective TC size retrieval technology using only the IR band BT from geostationary satellite observations. The retrieval accuracy was acceptable in operational applications (Knaff et al. 2010 and 2015), but the model involved complex operations such as a variational data-fitting algorithm that is difficult to be implemented in real time. Furthermore, the grid data product of the MTCSWA has not been publicly released. Lu et al. (2017) used the 1980–2009 geostationary satellite observation dataset (Knapp and Kossin 2007) to establish a linear objective estimation model of TC size (defined as the R34) based on the correlation between the radial distribution features of BT, its gradient in the IR band, and TC size. However, there may be a complex nonlinear relationship between remote sensing information and these key physical elements. Hence, it is necessary to establish a more advanced or robust technique to estimate the detailed wind structure of TCs.

Machine learning (ML) is an approach to establish an approximate model of a given problem, such that it can effectively represent the nonlinear relationship between multiple factors and the target predictand(s) (Kim et al. 2019). At present, ML methods include the multi-layer perceptron (MLP), radial basis function (RBFN), general regression neural network (GRNN), k-nearest neighbour (KNN), support vector machine (SVM), decision tree (DT), and several others (Specht 1991; Ghosh and Krishnamurti 2018; Fuchs et al. 2018; Zhang et al. 2019; Kim et al. 2019 and 2020; Neetu et al. 2020). Zhang et al. (2019) evaluated TC genesis forecasts in the WNP using KNN, SVM, DT, and linear methods. The results showed that the performance of the SVM was better than that of the linear method. Kumler-Bonfanti et al. (2020) used ML to identify tropical and extratropical cyclones and found that ML is more efficient and accurate than conventional methods. However, there is no optimal ML algorithm suitable for all cases, and the performance of a ML algorithm depends not only on the algorithm technique, but also on the application type and input data. ML algorithms have been shown to greatly improve the accuracy of TC intensity estimation (Ghosh et al. 2018; Chen et al. 2019; Wimmers et al. 2019), but the application of ML to TC size recognition is quite limited. Wimmers et al. (2019) noted that ML has great potential in estimating TC parameters such as gale wind radius and other structural characteristics.

This paper establishes the nonlinear models between observations obtained from geostationary meteorological satellites and TC size using ML. We carry out an objective TC size estimation and construct a TC size climate dataset with fine structural characteristics in the WNP. Section 2 introduces the data, whilst the ML methods and TC size estimation tests are discussed in Section 3. The construction and validation of the TC size dataset are illustrated in Section 4. A summary and conclusions will be given in Section 5.

2 Dataset

Lu et al. (2017) showed that there was no significant influence on the estimation of TC size using different series of satellite data. Thus, the IR observation dataset of HURSAT-B1 (1981–2016; Knapp and Kossin, 2007) and FY-2G (2017–2019; Lu and Gu 2016) are used as inputs for the model learning phase, testing, and estimation. The HURSAT-B1 dataset contains seven geostationary meteorological satellite observations combined, including FY-2 from the China Meteorological Administration (CMA), Meteosat-2 to Meteosat-9 from EUMETSAT, GMS-1 to GMS-5, MTSAT-1R to MTSAT-2R and Himawari-8 from the Japan Meteorological Agency (JMA), and GOES-1 to GOES-13 from the United States National Oceanic and Atmospheric Administration (NOAA). All the observations are interpolated onto a regular latitude–longitude grid with a resolution of 0.07 degrees (~8 km) around the TC center. The temporal resolution is 3 hours. The FY-2G dataset is obtained from the National Satellite Meteorological Center of the CMA. The spatial resolution of the FY-2G infrared band is 5 km and the temporal resolution is 0.5 hours. To ensure the consistency of input in model training and estimation, FY-2G data is interpolated onto an 8 km grid. Furthermore, only those satellite observations at 0000, 0600, 1200, and 1800 UTC are selected in the calculation to match the time resolution of the best track record.

In this study, the R34, R50, and R64 in the northeast, southeast, southwest, and northwest quadrants (NE, SE, SW, and NW, respectively), and RMW from the JTWC best track data are taken as the ground truth for training and evaluating the performance of the ML model. The observation times were 0000, 0600, 1200, and 1800 UTC. The TC serial number, name, location (longitude, latitude) and intensity (MSW) are included in this dataset. In addition, aircraft observation reports near the surface of the TC center (1981–1987) and periphery (1985–1987) in the WNP (Bai et al. 2019) are used to validate and assess the performance of the ML model in this study. The aircraft observations of TC centers include the observation time, MSW, and RMW. The TC periphery observations include the gale wind speed, the observed location, and time.

During the final TC size dataset construction, the TC tracks and intensity data from the IBTrACS (v04r00) dataset (Knapp et al. 2010) covering the time period from 1981 to 2019 were used for the position and intensity of the TC center, in order to match the TC center where the HURSAT gridded dataset was centered. It also included the TC serial number, name, center longitude, center latitude, and TC intensity at 0000, 0600, 1200, and 1800 UTC. The intensity grade includes tropical depression (TD; 10.8 ≤ MSW ≤ 17.1 m/s), tropical storm (TS; 17.2 ≤ MSW ≤ 24.4 m/s), severe tropical storm (STS; 24.5 ≤ MSW ≤ 32.6 m/s), typhoon (TY; 32.7 ≤ MSW ≤ 41.4 m/s), severe typhoon

(STY; 41.5 ≤ MSW ≤ 50.9 m/s), and super typhoon (SuperTY; MSW ≥ 51 m/s). Note that this study only considers TCs in the WNP region unless otherwise specified.

3 Methods

3.1 Machine learning algorithms

In this study, five regression-based ML algorithms with various fitting functions are selected to conduct the experiments and evaluation of TC size estimation (Specht 1991; Ghosh and Krishnamurti 2018; Fuchs et al. 2018; Zhang et al. 2019; Kim et al. 2019 and 2020). They are the MLP, GRNN, RBFN, SVM, and DT.

An MLP is a common artificial neural network (ANN) algorithm that consists of an input layer, an output layer with one or more hidden layers that apply weights to the inputs and direct them through an activation function to the output. An MLP is fully connected between different layers and performs well on nonlinear data that each node (neuron) is connected with all other nodes in the preceding layer. An RBFN is a kind of ANN using a radial basis function as the activation function to prescribe how the weighted sum of input is transferred to output from neurons in a layer of the network. The output in RBFN is a linear combination of the radial basis function of inputs and the neuron parameters (i.e. the coefficient in the weight to generate output). A GRNN is a modified RBFN with faster convergence (Specht 1991; Ghosh and Krishnamurti 2018). An SVM, which is a non-parametric statistical learning technique, builds a hyperplane to separate the dataset into a discrete, predefined number of classes. It utilizes a kernel function to transform the dimension of the data into a higher one to identify an optimal hyperplane (Mountrakis et al. 2011; Lee et al. 2016). A DT is a process of data classification through a series of rules. In a DT, the data samples are partitioned into subdivisions repeatedly based on decision rules that resemble branches in a tree (Zhu et al. 2019). The advantage of the DT is to allow intuitive interpretation of and physical insights into the classification rules, as it includes conditions ("if-then-else" rules) based on the relative importance of predictors. In summary, ML can automatically and objectively represent nonlinear relations between key features of satellite observations and the target physical parameters (Kim et al. 2019 and 2020; Zhang et al. 2019).

The five machine learning algorithms are given in Table 1 with empirical and experiential parameters (Specht 1991; Ghosh and Krishnamurti 2018; Fuchs et al. 2018; Zhang et al. 2019; Kim et al. 2019 and 2020; Kumler-Bonfanti et al. 2020). In the following section we determine the best model and input scheme according to an independent sample test performance.

3.2 Determination of input schemes for the ML methods

Previous studies have shown that TC intensity, wind structure, and TC cloud structure are closely related (Dvorak 1975; Velden et al. 1998; Demuth et al. 2006; Mueller et al. 2006; Kossin et al. 2007; Lajoie and Walsh 2008; Elizabeth et al. 2014; Knaff et. al. 2014, 2016; Lu et al. 2017). The radial profile characteristics of infrared cloud-top BT clearly indicate TC intensity, inner and outer core size, and their variation. An analysis of the correlation between TC wind structure parameters (RMW, R64, R50, and R34) and TC intensity (MSW) using 12,529 samples during the period 2001–2017, revealed that the TC inner size (RMW) and R34 are correlated with TC intensity (the correlation coefficients are −0.53 and 0.55, respectively, which are statistically significant at the 99% confidence level). Moreover, there is also a positive correlation between TC intensity and the R64 and R50 (the correlation coefficients are 0.39 and 0.49, respectively, at the 99% confidence level). Lu et al. (2017) also determined from satellite infrared observation that the BT profile distribution, intensity, and location of the TC cloud top are related to the TC size as represented by the R34.

Consequently, in this study, the BT profile in the region from the TC center to a specified radius (R), the TC center position, and TC intensity are used as inputs in the ML algorithm to estimate the TC size. The TC size is expressed in terms of the RMW, R34 (mean value of the four quadrants), R50 (mean value of the four quadrants), R64 (mean value of the four quadrants), and R34-1, R34-2, R34-3, R34-4, R50-1, R50-2, R50-3, R50-4, R64-1, R64-2, R64-3, and R64-4 (where the suffix -1 indicates the NE quadrant, -2 indicates the SE quadrant, -3 the SW quadrant, and -4 the NW quadrant). Here, the BT profile is obtained by calculating the azimuthal average of each grid annulus in each quadrant in the region from the TC center to the radius R. Finally, the estimation accuracy using different input schemes is evaluated.

a. Determination of the best input scheme and ML model for the R34 and RMW

We consider samples with an intensity above TS between 2001 and 2016 (11,060 samples). Taking the R34 from the JTWC best track data as the ground truth, 8742 samples between 2001 and 2013 are used for model training (Zhou, 2020), and 2318 samples between 2014 and 2016 are used for the independent sample test. In the experiments, R is variously set to be 10, 20, 30, 40, 50, 60, 70, or 80 grid points away from the TC center (the spatial resolution of the grid is ~8 km, which is consistent with that of the satellite data). Then, the longitude (Lon) and latitude (Lat) of the TC center, TC intensity (MSW), and BT radial profile (BTP) within the radius R are used as inputs for the ML models in the eight different input scheme experiments. The test results are shown in Fig. 1.

Fig. 1 shows that as the input BTP radius moves from the inner core (10 grid points from TC center) to the outer edge (80 grid points from TC center), the estimation errors of different methods significantly differ from one another. The estimation errors of the MLP (blue line) and SVM (red line) decrease first and then increase with R, with the smallest estimation errors when R is between 40 and 60 grid points. That is, an input radius between 320 and 480 km from the TC center results in the best estimation of the true TC size. The estimation errors of the GRNN (green line) and DT (magenta line) also decrease first, then remain constant when R is larger than 20 grid points in the case of the GRNN and when R is larger than 40 grid points in the case of the DT. The estimation error of the RBFN (cyan line) increases monotonically with R. This performance may be related to the models themselves and their basic parameters, which were set according to experience and test errors. As this test only assesses the basic performance of five algorithms in estimating TC size, the parameters of the model itself are not thoroughly investigated.

The mean estimation error (black line) of the five methods demonstrates that the average error decreases at first and then increases. The minimum error is at 40 grid points away from the TC center, which indicates that an input BTP within 320 km of the TC center results in the best estimation of the TC size. This is consistent with Lu et al. (2017), who showed that the BT distribution and its gradient in the range of 40–50 km (TC inner core region) and 256–288 km (outer region) from the TC center have the highest correlation with TC size. Hence, the BTP information 320 km from the TC center contains the most relevant characteristics of the TC core and periphery, and 40 grid points is thus determined as the input R of the R34 estimation scheme. Similarly, 40, 30, 30, 20 and 40 grid points are determined as the input R for the RMW, R34-1, R34-2, R34-3, and R34-4 estimation schemes, respectively.

The estimation accuracy of the MLP (blue line), DT (magenta line), and SVM (red line) is <50 km using the optimal estimation scheme, which is smaller than that of the other two algorithms and is better than that of the wind radii estimates in operational forecasts (and in the best track records) (Knaff et al.

2010 and 2015). However, the normal distribution and probability density function of the estimation results from these three methods demonstrate that the SVM results have a more reasonable normal distribution and pass the 95% confidence test. The analysis plot is not shown here because of limited space. Hence, the SVM is selected as the final estimation model for the R34 and RMW.

b. Determination of the best input scheme and ML model for the R50 and R64

The R50 and R64 have been available in best track data from the JTWC since 2004. In total there are 4350 samples matched with the HURSAT satellite observations up to 2016. Here, 3519 samples from 2004 to 2014 are used to train the models (Zhou, 2020), and 831 samples from 2015 to 2016 are used as test samples. The test methods of different input schemes (i.e., different input R) are the same as those introduced in Section *a*. However, as the R50 and R64 are also restricted by the value of the R34, the R34 estimation value is also regarded as an additional input to the R50 and R64 estimation models.

The test results are shown in Fig. 2 and Table 2. There is little difference between the estimation errors of different methods as the input BTP radius moves from the inner core (10 grid points from the TC center) to the outer edge (80 grid points from the TC center). The estimation errors decrease and then increase with R for both the R50 (Fig. 2a) and R64 (Fig. 2b). The mean estimation error (black line) of the five methods demonstrates that the average error decreases first and then increases. The minimum error is at 20 grid points from the TC center, meaning that the BTP within 160 km of the TC center results in the best estimation of the R50 and R64. Therefore, 20 grid points is chosen as the optimal model input. Table 2 shows that the GRNN algorithm performs best in the estimation of the R50 and R64. The MAEs of the mean and in each quadrant are all smaller than those of the other four methods, so the GRNN algorithm is selected as the final estimation model for the R50 and R64.

c. Further optimization of the models

Following the determination of the optimal ML models and input schemes, the ML models are retrained with the same samples to fine-tune the parameters further. Finally, the parameters that give the minimum MAE are employed to construct the TC size dataset. In the final regression SVM models, the Automatic Optimization of Hyper-parameters (Mountrakis et al. 2011; Lee et al. 2016) is the most effective for the R34 and RMW estimation.

The advantage of the GRNN is its convenient network parameter setting function. The performance of the GRNN network can be adjusted by setting only one parameter, denoted '*Spread*' (also known as the bandwidth) (May et al. 2010; Ghosh and Krishnamurti 2018). In the experiments, the initial Spread is set to 0.1 and increases to 100 in intervals of 0.1. The results show that the MAE decreases with the increase of Spread, but after reaching a certain value, the MAE levels out and then begins to increase. We find that there is a minimum estimated MAE for both the R50 and R64 when the bandwidth is set to 9.8 and 23.8 in the GRNN models, respectively. We note that all of the above models are convergent.

4 TC size dataset construction and estimation error analysis

4.1 TC size dataset construction in the WNP

Based on the trained ML models and the determined input schemes, the TC size dataset in the WNP during the period between 1981 and 2019 is constructed using the infrared band observations of HURSAT B1 (1981–2016), FY-2G (2017–2019), and the IBTrACS data. The TC size dataset includes 19,995 samples and 940 TCs above TS intensity, with information about the RMW and wind radius (km) of the R34, R50, and R64 in four quadrants. It should be noted that as the sample from 2001 to 2013 are incorporated in the training phase, the interpretation of the constructed TC size dataset during that period may need further attention to avoid possible influence of data over-fitting.

The TC size distribution for various size parameters (R34, R34-1, R34-2, R34-3, R34-4, R50, R50-1, R50-2, R50-3, R50-4, R64, R64-1, R64-2, R64-3, R64-4, and RMW) is shown in Fig. 3. The mean R34, R50, R64, and RMW are 179, 100, 63, and 47 km, respectively, and the median values are 173, 94, 60, and 48 km, respectively. The distribution and probability density function of R34 show that the estimated R34 has a normal distribution centered at about 180 km. In addition, 99.9% of the estimated R34 values are below 400 km, and only about 5% are below 100 km.

4.2 Independent-samples validation and estimation error analysis

Taking the best track data from the JTWC during the period between 2017 and 2019 and the available aircraft reports (Bai et al. 2019) between 1981 and 1987 as the ground truth, we now assess the estimated TC sizes and analyze the errors. .

a. Independent-samples validation based on the JTWC best track data between 2017 and 2019

Taking the JTWC best track data as the ground truth, 1035 independent samples between 2017 and 2019 are used for validation. The results show that the respective MAEs of the mean estimated R34, R50, R64, and RMW are 58, 38, 21, and 25 km; the corresponding median errors are 46, 31, 17, and 19 km; and the standard deviations are 47, 33, 18, and 26 km. There is a clear correlation between the estimated values and the best track data for the R34 (Fig. 4), with a correlation coefficient of 0.39, which is statistically significant at the 95% confidence level (T-test was used for all tests of statistical significance). The blue ellipse in Fig. 4, which is the 95% confidence interval based on a normal distribution, contains most of the samples. There are few outliers (red crosses). The figure shows that the estimated R34 is consistent with that from the JTWC best track data. However, the centroid of the data is slightly lower than the fitting line, indicating that the overall estimated values of R34 are slightly smaller than the best track data; i.e., R34 is slightly underestimated.

The estimated median error is smaller than the MAE for all estimated parameters. This indicates that there are some samples with large bias that caused the larger MAE. Hence, considering R34 as an example, all samples are divided into subgroups by latitudinal zone, size, month, and intensity category to analyze in detail the characters of the estimation errors.

The error box-plot of R34 estimation in different latitudinal zones (Fig. 5) shows the best estimation accuracy for samples between 10° and 30° north (the median error was between approximately −8 and −10 km). The estimation accuracy worsened for samples between 30° and 40° north (median error, about −25 km), equatorward of 10° north (median error, about −57 km), and poleward of 40° north (median error, about −82 km). The estimation method did not perform well for TCs at lower latitudes (<10° north), as the associated cloud clusters of TCs were loosely organized during their early stage of the life cycle. As the TCs moved to higher latitudes (above 40° north), most were recurved and steered by the mid-latitude westerlies so that the superposition with the westerlies may have resulted in larger actual values of R34 than those underestimated by the proposed models.

The sampled TCs are divided into five groups from small to large according to the R34 value in the JTWC best track data: ≤100 km, 100–200 km, 200–300 km, 300–400 km, and ≥400 km. The estimation biases for the different size groups (Fig. 6) clearly increase in magnitude with increasing storm size. The estimated mean bias is between −50 and 50 km when the size is smaller than 300 km, but larger storms have estimated mean bias between −100 and −170km, indicating serious underestimation, i.e., the model's performance is limited for large TCs (defined as those above the 95th percentile of storm size). The estimated MAE for sample values above the 95th percentile is 161 km, which means that the estimated errors of the model for high-value samples are 2.8 times the average (58 km). This shows that the model does not adequately describe abnormal samples or outliers, which is a weakness of the regression method in general, whether linear or nonlinear.

The error bars of R34 estimation for different months (Fig. 7) are variable: the bias in January and December is large, with a mean of about −70 km, while the mean bias for February–April and November is about 0 km, indicating good estimation. The mean bias gradually increases in magnitude from around 0 km in June to −40 km in October, which may be related to the TCs in the WNP being larger from September to October (Guo and Tan, 2017; Lu et al., 2017).

There is no clear regularity of estimation bias of R34 in the different intensity categories. The accuracy is better for TS, TY, and SuperTY categories, whose estimation showed median errors between −4 and −10 km. The estimation of STS and STY showed median errors between −31 and −34 km. The analysis plot is not shown here because of limited space.

The spatial distribution of estimation bias of R34 (Fig. 8) indicates its underestimation near land, such as the coastal areas of the Philippines, East China, and the Korean Peninsula. When a TC is close to land, friction may lead to an inclination of the TC in the vertical direction. Then the BTP across the weak convection away from the center is obtained due to the misalignment of the center of the high-level cloud top and the surface center, which results in an underestimation of R34 in the model. On the other hand, R34 is overestimated in the region where a TC has just formed. It is plausible that dense cloud clusters associated with developing TCs may provide the model with false BTP features suggesting stronger convection, leading to overestimation.

Overall, the above validation shows that the proposed models perform satisfactorily in providing accurate and reliable estimated wind radii, except for at certain latitude regions or for unusually large TCs.

b. Independent-samples validation based on aircraft observations between 1981 and 1987

We now evaluate the estimated mean R34 using data from aircraft observation reports of the TC center and periphery, obtained during the period 1981–1987 in the WNP (Bai et al. 2019). The evaluation neglects R64 as there is no matched observation sample. The TC center observations are used here for RMW evaluation, with a total of 584 matching samples. R34 and R50 are evaluated based on the matching samples of the peripheral observation time and wind speed. Among them, there are 109 matched samples for R34 evaluation, but only 19 matched samples for R50 evaluation.

The validation results show that the MAEs between the mean estimated R34 (109 samples), R50 (19 samples), and RMW (584 samples) and the aircraft observations are 54, 34, and 25 km, respectively; the median errors are 39, 34, and 17 km, respectively; the standard deviations are 38, 22, and 22 km, respectively. This accuracy is slightly better than that of the validation result based on JTWC best track data between 2017 and 2019.

For the matched R34 samples, the mean observation radius of the wind speed between 15 and 21 m s^{-1} is defined as the observed R34. The estimated MAE, median error, and standard deviation are 54, 39, and 38 km respectively. Fig. 9 shows the corresponding scatter plot between the estimated R34 and observations; the correlation coefficient is ~0.45 (significant at the 95% confidence level). The blue ellipse is the 95% confidence interval based on a normal distribution, which contains most of the samples. The magenta ellipse is the range within one standard deviation of all samples. The figure shows that the estimated dataset is also consistent with the R34 values obtained from aircraft observation.

There are 19 matched samples for R50 evaluation. The mean observation radius of the wind speeds between 21.5 and 27.5 m s^{-1} is defined as the observed R50. The estimated MAE, median error, and standard deviation are 34, 34, and 22 km, respectively. The correlation coefficient between the estimated R50 and the observations is ~0.505 (significant at the 95% confidence level).

There are 584 matched samples for RMW evaluation. The estimated MAE, median error, and standard deviation are 25, 17, and 22 km, respectively. To analyze the error distribution, all samples are also divided into subgroups by latitude and intensity (Fig. 10).

The estimation error bars of RMW in different latitudinal zones (upper panel, Fig. 10) show that the range of estimation bias varies between approximately −40 and 20 km for all samples and that the mean bias is between −30 and −10 km. Most samples appear underestimated. The estimation accuracy decreases from lower to higher latitudes. The increasing underestimation with increasing latitude is broadly attributed to superposition with the westerlies, which is consistent with the analysis in Section 4.2***a***.

The estimation error bars for RMW in different intensity categories (lower panel, Fig. 10) show that the mean bias is between −20 and 0 km. The estimation accuracy improves from TS to SuperTY. Stronger TCs favor tighter cloud clusters near their centers, which can be better represented by the model due to the more prominent BTP features.

Overall, the estimated mean R34, R50, and RMW are mostly consistent with the observations. The MAEs for estimation of R34 and R50 (54 and 34 km, respectively) from aircraft observation are smaller than those from the JTWC best track data (58 and 38 km, respectively). However, the median estimation error is smaller than the MAE for all validations, which indicates that the larger MAE was caused by high-value samples. This indicates a slightly larger bias at high values, which may have originated from the combined effect of the estimation methods and the observation samples. For example, the samples at high latitudes have increased R34 and RMW owing to superposition with the westerlies; at the same time, the estimation model does not perform well with the disordered TC cloud structure caused by the westerlies.

Nevertheless, the estimation errors of this study are still smaller than those from operational wind radii estimates, which can be as large as 25%–40% of the radii themselves (Knaff et al. 2010 and 2015).

c. Comparison with previous research

Lu et al. (2017) put forward a linear stepwise regression method to estimate mean TC size (in terms of the R34) using the same satellite data as in this study. The estimated median error was 40 km, which is slightly larger than the value in this study (39 km, compared with aircraft observations). However, in this study, more TC size parameters are estimated and much more detailed information about the TC wind structure is provided, including the R34, R50, and R64 in four quadrants, as well as the RMW. Moreover, the ML algorithm used in this study may be able to reveal the nonlinear relationship between satellite observations and the TC wind field structure, whereas the linear method cannot.

The models and validation conclusions in this study are only suitable for the WNP region. As few studies have estimated the TC wind field structure in the WNP, we here compare the estimation accuracy of this study with comparable studies in the Atlantic. The estimation accuracy of R34, R50, R64, and RMW in this study is equivalent to that in some previous studies (Mueller et al. 2006; Knaff et al. 2011 and 2016). The MAEs for estimation of R34, R50, and R64 by Knaff et al. (2011 and 2016) are about 65, 35, and 23 km, respectively. The validation data for the Atlantic are closer to the ground truth as they are supported by aircraft observations. However, short-term aircraft observations and the best track dataset integrating multiple observations as the verification dataset can also be used to validate TC wind structure estimation in the WNP region, which is a workaround available to relevant studies in this region.

5 Summary and Conclusions

In this paper, identification models of size for TCs in the WNP were proposed based on the infrared channel imagery of geostationary meteorological satellites. Several different machine learning algorithms were tested for different TC size parameters, including RMW, R34, R50, and R64. It is obtained that RMW and R34 can be best estimated by the support vector machine models, while R50 and R64 can be best estimated by the general regression neural network models. These models are used to set up a dataset of TC size for a nearly 40-yr period in the WNP region.

Evaluation of the TC size datasets was conducted using independent samples based on aircraft observations (1981–1987) and JTWC best track data (2017–2019). The results show that the estimated MAEs for R34 are 54 and 58 km, respectively. These MAEs are comparable to the accuracy of wind radius estimates in previous studies. The estimated accuracy for 10°N to 30°N is higher than that for other latitudes, and the errors are larger near coastal areas than open seas. The estimation accuracy of RMW increases with increasing intensity of TC. There are overall slight underestimations of the models, which will require future study.

The models proposed here are constructed and validated based on JTWC best track data and past aircraft observations in the WNP. As there are few aircraft observations in WNP to verify the TC size dataset, further study would be required to implement and validate the performance of the proposed models, such as using datasets for the western Atlantic, where more aircraft reconnaissance observations are available. Moreover, this study has demonstrated a feasible way to carry out relevant research and develop a methodology to estimate TC size or representative parameters for TCs in the WNP. It is anticipated that the proposed algorithms could be improved in future using more observations to enhance the ML models and validate the testing results.

All in all, this study shows that infrared images contain important information about the low-level wind field. By transforming the two-dimensional BT field to the azimuthal mean profile and extracting the distribution features, it can be used as the main predictor in a ML algorithm to estimate the wind radii of the R34, R50, R64, and RMW. However, the performance of the ML algorithm is limited for unusually large or small TCs. This needs to be improved by using or combining multiple algorithms in the future.

All of the algorithms in this study can be implemented in real-time operational applications (Fig. 11) or in post-seasonal analysis as a reference for operational TC forecasting. In addition, the estimation dataset in this study provides important parameters regarding TC evolution in the WNP and may benefit model initialization of TC structure in regions such as the WNP, where aircraft observations and reconnaissance data are relatively limited.

Acknowledgments

This study was jointly supported by the Shanghai Natural Science Foundation (21ZR1477300), the Key Projects of the National Key R&D Program (2018YFC1506300), the FengYun Application Pioneering Project (FY-APP-2021.0106), the Typhoon Scientific and Technological Innovation Group of Shanghai Meteorological Service, the WMO Typhoon Landfall Forecast

Demonstration Project (TLFDP), and the Open Foundation of Fujian Key Laboratory of Severe Weather (2021).

Data Availability Statement

The datasets generated and/or analyzed in this study are available from the corresponding author on reasonable request.

The data supporting the findings of this study are available from National Centers for Environmental Information (NCEI) of US, Joint Typhoon Warning Center of US, and National Meteorological Satellite Center of China. The public access address of HURSAT, IBTrACS, JTWC best track, and FY2G satellite dataset is https://www.ncei.noaa.gov/products/hurricane-satellite-data?name=summary, https://www.ncei.noaa.gov/products/international-best-track-archive?name=bib, http://www.metoc.navy.mil/jtwc/jtwc.html?western-pacific, and http://satellite.nsmc.org.cn/PortalSite/Data/Satellite.aspx, separately.

References

Bai, L.N., H. Yu, P. G. Black, Y. L. Xu, M. Ying, J. Tang, and R. Guo. 2019: Re-examination of the tropical cyclone wind–pressure relationship based on pre-1987 aircraft data in the Western North Pacific. *Wea. Forecasting*, **34**, 1939-1954. DOI: https://doi.org/ 10.1175/WAF-D-18-0002.1

Brand, S. 1972: Very large and very small typhoon of the Western North

Pacific Ocean, *J. Meteor. Soc. Japan*, **50 (4)**, 332-341. DOI: https://doi.org/10.2151/jmsj1965.50.4_332

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. DOI: https://doi.org/ 10.1175/MWR-D-10-05062.1

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

Chavas, D. R., and K. A. Emanuel. 2010: A QuikSCAT climatology of tropical cyclone size. *Geophysical research letter.* **37**, L18816. DOI: https://doi.org/ 10.1029/2010GL044558

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. DOI: https://doi.org/10.1175/WAF-D-18-0136.1

Cocks, S. B., and W. M. Gray. 2002: Variability of the outer wind profiles of western North Pacific typhoons: Classifications and techniques for analysis and forecasting. *Mon. Wea. Rev.* **130**, 1989-2005. DOI: https://doi.org/ 10.1175/1520-0493(2002)1302.0.CO;2

Croxford, M., and G. M. Barnes. 2002: Inner core strength of Atlantic tropical cyclones. *Mon. Wea. Rev*. **130**. 127-139. DOI: https://doi.org/ 10.1175/1520-0493(2002)1302.0.CO;2

Demuth, J. L., M. Demaria, and J. A. Knaff. 2004: Evaluation of Advanced Microwave Sounding Unit tropical-cyclone intensity and size estimation algorithms. *J. Appl. Meteor.*, **43**, 282-296. DOI: https://doi.org/10.1175/1520-0450(2004)043<0282:EOAMSU>2.0.CO;2

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. DOI: https://doi.org/ 10.1175/1520-0450(2004)0432.0.CO;2

Dvorak, V. F. 1975: Tropical cyclone intensity analysis and forecasting from satellite imagery. *Mon. Wea. Rev*. **103**, 420–430. DOI: https://doi.org/10.1175/1520-0493(1975)103<0420:TCIAAF>2.0.CO;2

Elizabeth, R. S., B. S. Barrett, and C. M. Fine. 2014: Relationships between tropical cyclone intensity and eyewall structure as determined by radial profiles of inner-core infrared brightness temperature. Mon. Wea. Rev., 142, 4581-4597

Fuchs, J., J. Cermak, and H. Andersen. 2018: Building a cloud in the southeast Atlantic: Understanding low-cloud controls based on satellite observations with machine learning. *Atmos. Chem. Phys.* **18**, 16537–16552. DOI: https://doi.org/ 10.5194/acp-18-16537-2018

Ghosh, T., and T. N. Krishnamurti. 2018: Improvements in hurricane intensity forecasts from a multimodel superensemble utilizing a generalized neural network technique. *Wea. Forecasting*, **33**, 873–885. DOI:

https://doi.org/10.1175/WAF-D-17-0006.1

Guo, X., and Z. M. Tan. 2017: Tropical cyclone fullness: A new concept for interpreting storm intensity. *Geophys. Res. Lett.*, **44 (9)**, 4324-4331. DOI: https://doi.org/10.1002/2017GL073680

Kim M., J. Cermak, H. Andersen, J. Fuchs, and R. Stirnberg. 2020: A new satellite based retrieval of low-cloud liquid-water path using machine learning and Meteosat SEVIRI Data. *Remote Sens*. **12**, 3475. DOI: https://doi.org/ 10.3390/rs12213475

Kim M., M. S. Park, J. Im, S. Park, and M. Lee. 2019: Machine learning approaches for detecting tropical cyclone formation using satellite data. *Remote Sens*. **11(10)**, 1195. DOI: https://doi.org/ 10.3390/rs11101195

Knaff, J. A., 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. DOI: https://doi.org/ 10.1175/2011JAMC2673.1

Knaff, J. A., and B. A. Harper. 2010: Tropical cyclone surface wind structure and wind-pressure relationships, Proc.WMO Seventh Int. Workshop on Tropical Cyclones, La Reunion, France, WMO, p.KN1.4

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

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, DOI: https://doi.org/10.1175/WAF-D-14-00149.1.

Knaff, J. A., 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. DOI: https://doi.org/10.1175/MWR-D-15-0267.1

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

Knapp, K. R.; M. C. Kruk, D. H. Levinson, H. J. Diamond, C.J. Neumann. 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. Bull. *Am. Meteorol. Soc*. **91**, 363–376.

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. DOI: https://doi.org/10.1175/WAF985.1

Kumler-Bonfanti C., J. Stewart, D. Hall, and M. Govett. 2020: Tropical and extratropical cyclone detection using deep learning. *J. Appl. Meteor. Climatol*, **59**, 1971-1985. DOI: https://doi.org/ 10.1175/JAMC-D-20-0117.1

Lajoie, F., and K. Walsh. 2008: A technique to determine the radius of maximum wind of a tropical cyclone. *Wea. Forecasting*. **23**, 1007-1015. DOI:

https://doi.org/ 10.1175/2008WAF2007077.1

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. DOI: https://doi.org/10.1175/2010MWR3023.1

Lee, S., J. Im, J. Kim, M. Kim, M. Shin, H. C. Kim, and L. J. Quackenbush. 2016: Arctic sea ice thickness estimation from CryoSat-2 satellite data using machine learning-based lead detection. *Remote. Sens*. **8**, 698. DOI: https://doi.org/ 10.3390/rs8090698

Lin, S., and K. Chou. 2018: Characteristics of size change of tropical cyclones traversing the Philippines. *Mon. Wea. Rev*. **146**, 2891–2911. DOI: https://doi.org/10.1175/MWR-D-18-0004.1

Liu, K. S., and J. C. L. Chan. 1999: Size of Tropical cyclone as inferred from ERS-1 and ERS-2 data. *Mon. Wea., Rev*. **127**, 2992-3001. DOI: https://doi.org/10.1175/1520-0493(1999)1272.0.CO;2

Lu, N. M., and S. Y. Gu. 2016: Review and prospect on the development of meteorological satellites. *Journal of Remote. Sens* (in Chinese). **20 (005)**, 832-841. DOI: https://doi.org/ 10.11834/jrs320166194

Lu, X. Q., H. Yu, and X. T. Lei. 2011: Statistics for size and radial wind profile of tropical cyclones in the western North Pacific. *Acta Meteor. Sinica*. **25(1)**, 104-112. DOI: https://doi.org/10.1007/s13351-011-0008-9.

Lu, X. Q., H. Yu, X. M. Yang, and X. F. Li. 2017: Estimating tropical cyclone size in the Northwestern Pacific from geostationary satellite infrared images, *Remote Sens*. **9**, 728. DOI: https://doi.org/10.3390/rs9070728

May, R. J., H. R. Maier, and G. C. Dandy. 2010: Data splitting for artificial neural networks using SOM-based stratified sampling. *Neural Netw*, **23**, 283–294, https://doi.org/10.1016/ j. neunet.2009.11.009.

McKenzie T. 2017: A climatology of tropical cyclone size in the Western North Pacific using an alternative metric [D]. The Florida State University.

Merrill, R. T. 1984: A comparison of large and small tropical cyclones. *Mon. Wea. Rev*. **112**, 1408-1418. DOI: https://doi.org/10.1175/1520-0493(1984)1122.0.CO;2

Mountrakis, G., J. Im, and C. Ogole. 2011: Support vector machines in remote sensing: A review. *Isprs J. Photogramm. Sens*. **66**, 247–259. DOI: https://doi.org/ 10.1016/j.isprsjprs.2010.11.001

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. DOI: https://doi.org/ 10.1175/WAF955.1

Neetu, S., M. Lengaigne, J. Vialard, M. Mangeas, C. E. Menkes, I. Suresh, J. Leloup, and J. A. Knaff. 2020: Quantifying the benefits of nonlinear methods for global statistical hindcasts of tropical cyclones intensity. *Wea. Forecasting*, **35(3)**, 807-820. DOI: https://doi.org/ 10.1175/WAF-D-18-0201.1

Schenkel, B. A., 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. DOI: https://doi.org/10.1175/JCLI-D-17-0122.1

Schenkel, B. A., N. Lin, D. Chavas, 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. DOI: https://doi.org/10.1175/JCLI-D-17-0630.1

Shea D. J., and W. M. Gray. 1973: The hurricane's inner core region. I. Symmetric and asymmetric structure. *J. Atmos. Sci.* **30**, 1544 – 1564. DOI: https://doi.org/10.1175/1520-0469(1973)0302.0.CO;2.

Specht, D. F. 1991: A general regression neural network. IEEE Trans. *Neural Netw*, **2**, 568–576. https://doi.org/10.1109/ 72.97934.

Velden, C. S., T. L. Oleander, and R. M. Zehr. 1998: Development of an objective scheme to estimate tropical cyclone intensity from digital geostationary satellite infrared imagery. *Wea. Forecasting*. **13**, 172–186. DOI: https://doi.org/ 10.1175/1520-0434(1998)0132.0.CO;2

Weatherford, C. L., and W. M. Gray. 1988a: Typhoon structure as revealed by aircraft reconnaissance. Part I: Data analysis and climatology. *Mon. Wea. Rev*. **116**, 1032-1043. DOI: https://doi.org/10.1175/1520-0493 (1988)1162.0.CO; 2

Weatherford, C. L., and W. M. Gray. 1988b: Typhoon structure as revealed by aircraft reconnaissance. Part II: Structural variability. *Mon. Wea. Rev.*, **116**, 1044-1056. DOI: https://doi.org/10.1175/1520-0493(1988)1162.0.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. DOI: https://doi.org/ 10.1175/MWR-D-18-0391.1

Wu, L., W. Tian, Q. Liu, J. Cao, and J. A. Knaff. 2015: Implications of the observed relationship between tropical cyclone size and intensity over the western North Pacific. *J. Climate.* **28**, 9501-9506. DOI: https://doi.org/ 10.1175/JCLI-D-15-0628.1

Xu, J., and Y. Wang. 2015: A statistical analysis on the dependence of tropical cyclone intensification rate on the storm intensity and size in the North Atlantic. *Wea. Forecasting*. **30**, 692–701. DOI: https://doi.org/10.1175/WAF-D-14-00141.1

Xu, J., and Y. Wang. 2018: Dependence of tropical cyclone intensification rate on sea surface temperature, storm intensity, and size in the Western North Pacific. *Wea. Forecasting*, **33**, 523–537. DOI: https://doi.org/10.1155/2019/9456873

Zhang, T., W. Lin, Y. Lin, M. Zhang, H. Yu, K. Cao, and W. Xue. 2019: Prediction of tropical cyclone genesis from mesoscale convective systems using machine learning. *Wea. Forecasting*, **34(4)**, 1035-1049. DOI: https://doi.org/ 10.1175/WAF-D-18-0201.1

Zhou, Z. H. 2020: Machine learning. *Beijing: Tsinghua university press*, **p 1-22**.

Zhu, X., N. Li, and Y. Pan. 2019: Optimization performance comparison of three different group intelligence algorithms on a SVM for hyperspectral imagery classification. *Remote. Sens*. **11**, 734. DOI: https://doi.org/10.3390/rs11060734

List of Figures

Fig. 1 Difference between the R34 from JTWC best track data and that estimated by various methods using different input radii from the TC center (2318 samples between 2014 and 2016). The x-axis is the number of grid points used for the input BTP. The spatial resolution of the grid is ~8 km.

Fig. 2 Errors in (a) R50 and (b) R64 estimated using different algorithms and different input radii from the TC center (831 samples from 2015 to 2016). The figure illustrations are the same as in Fig. 1.

Fig. 3 Box plots of R34, R50, R64, and RMW estimated for TCs in the WNP between 1981 and 2019 (19,995 samples, 940 TCs above TS intensity).

Fig. 4 Scatter plot of IR-predicted R34 vs R34 from the JTWC best track data between 2017 and 2019 (1035 samples). The black line is a linear fit between the two variables. N is the number of samples, and R^2 is the correlation coefficient (statistically significant at the 95% confidence level).

Fig. 5 Estimation bias of R34 in different latitudinal zones compared with the JTWC best track data in the WNP between 2017 and 2019. The sample size is as in Fig. 4. Numbers in parentheses give the sample size.

Fig. 6 Estimation bias of R34 for TCs grouped by size compared with the JTWC best track data in the WNP between 2017 and 2019. The sample size is as in Fig. 4. Numbers in parentheses give the sample size.

Fig. 7 Estimation bias of R34 for TCs in different months compared with the JTWC best track data in the WNP between 2017 and 2019. The sample size is as in Fig. 4. Numbers in parentheses give the sample size.

Fig. 8 Spatial distribution of estimation bias of R34 compared with the JTWC best track data in the WNP between 2017 and 2019. The sample size is as in Fig. 4. The number in each grid is the sample size.

Fig. 9 Scatter plot of the IR-predicted R34 vs R34 from aircraft observations between 1981 and 1987. The black line is a linear fit between the two variables. N is the number of samples, and R^2 is the correlation coefficient (statistically significant at the 95% confidence level). The color represents the density of the scatter points.

Fig. 10 Error bars for estimation of RMW at different latitudes (upper) and for different intensity categories (lower) compared with aircraft observations in the WNP between 1981 and 1987 (584 samples). Numbers in parentheses gives the sample size.

Fig. 11 Flow chart of the algorithms implemented in real-time operational applications.

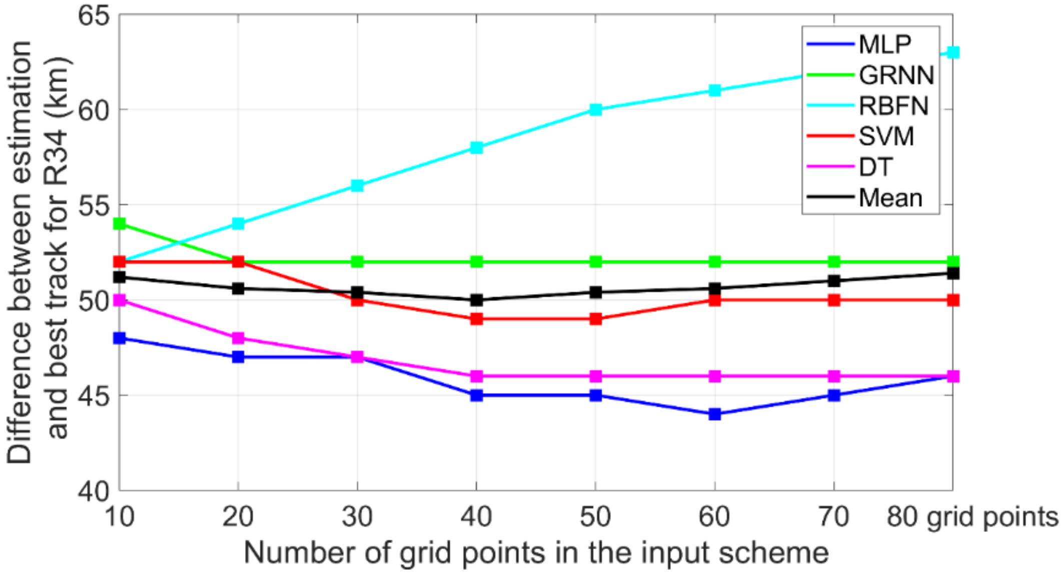


Fig. 1 Difference between the R34 from JTWC best track data and that estimated by various methods using different input radii from the TC center (2318 samples between 2014 and 2016). The x-axis is the number of grid points used for the input BTP. The spatial resolution of the grid is ~8 km.

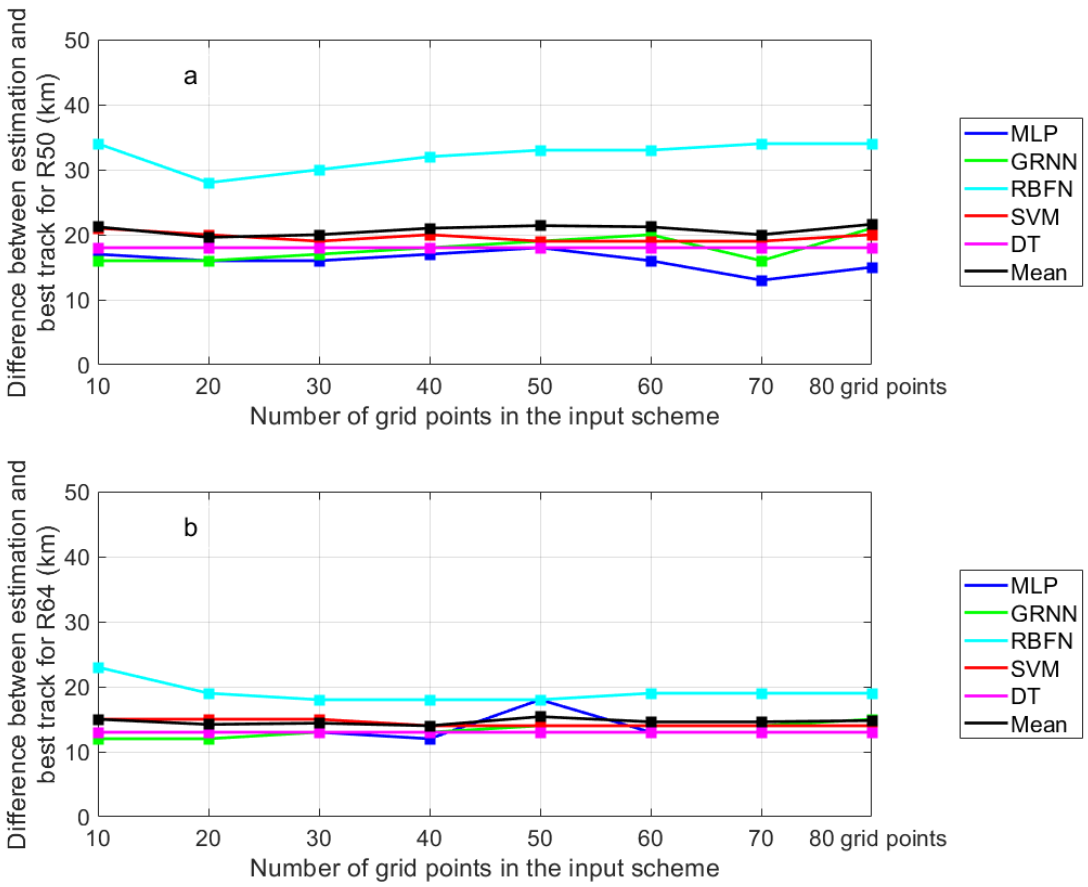


Fig. 2 Errors in (a) R50 and (b) R64 estimated using different algorithms and different input radii from the TC center (831 samples from 2015 to 2016). The figure illustrations are the same as in Fig. 1.

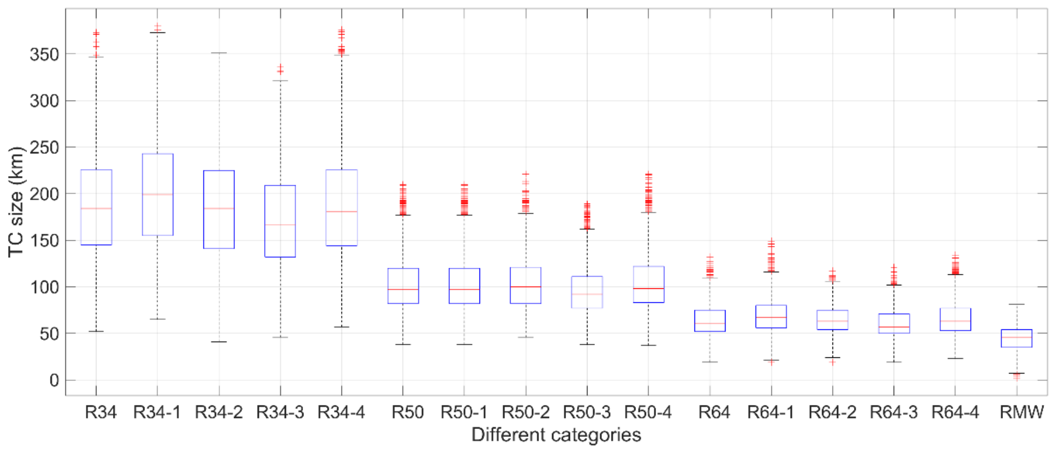


Fig. 3 Box plots of R34, R50, R64, and RMW estimated for TCs in the WNP between 1981 and 2019 (19,995 samples, 940 TCs above TS intensity).

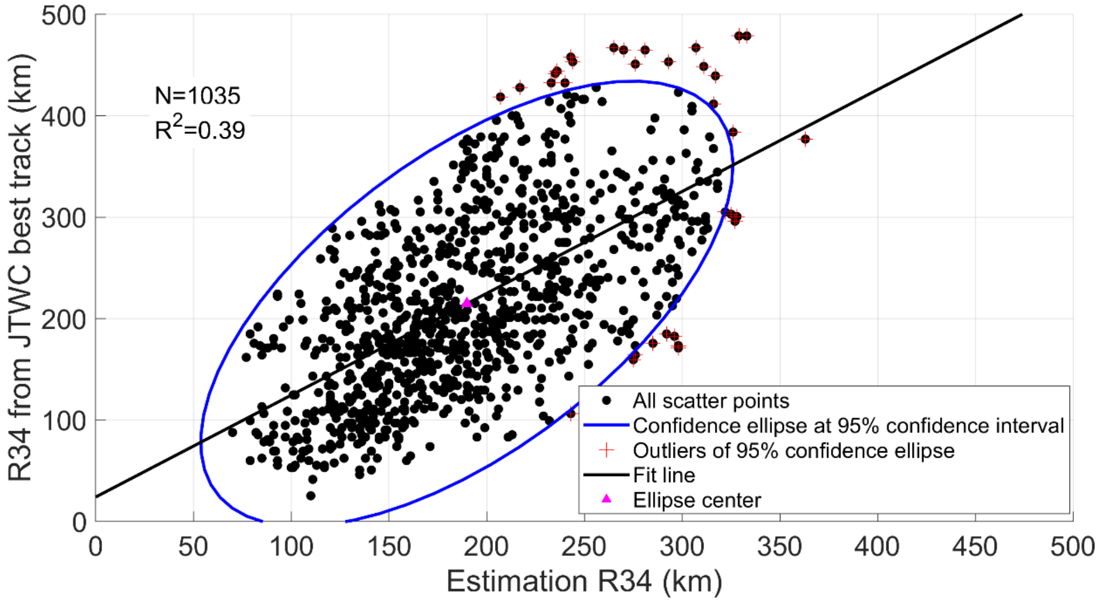


Fig. 4 Scatter plot of IR-predicted R34 vs R34 from the JTWC best track data between 2017 and 2019 (1035 samples). The black line is a linear fit between the two variables. N is the number of samples, and R^2 is the correlation coefficient (statistically significant at the 95% confidence level).

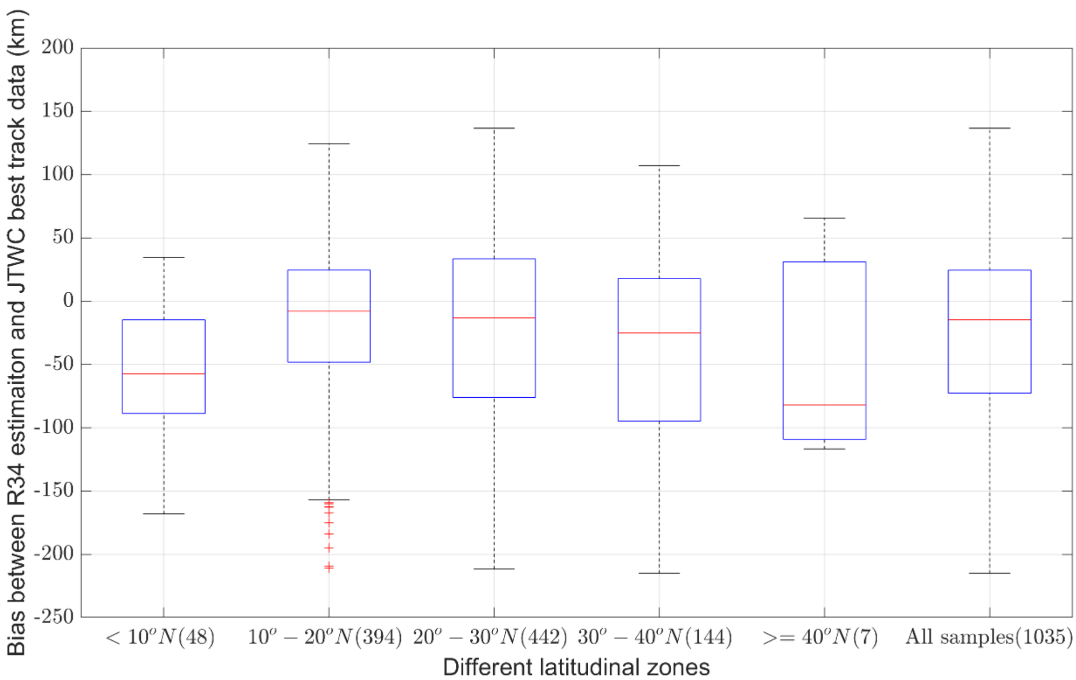


Fig. 5 Estimation bias of R34 in different latitudinal zones compared with the JTWC best track data in the WNP between 2017 and 2019. The sample

size is as in Fig. 4. Numbers in parentheses give the sample size.

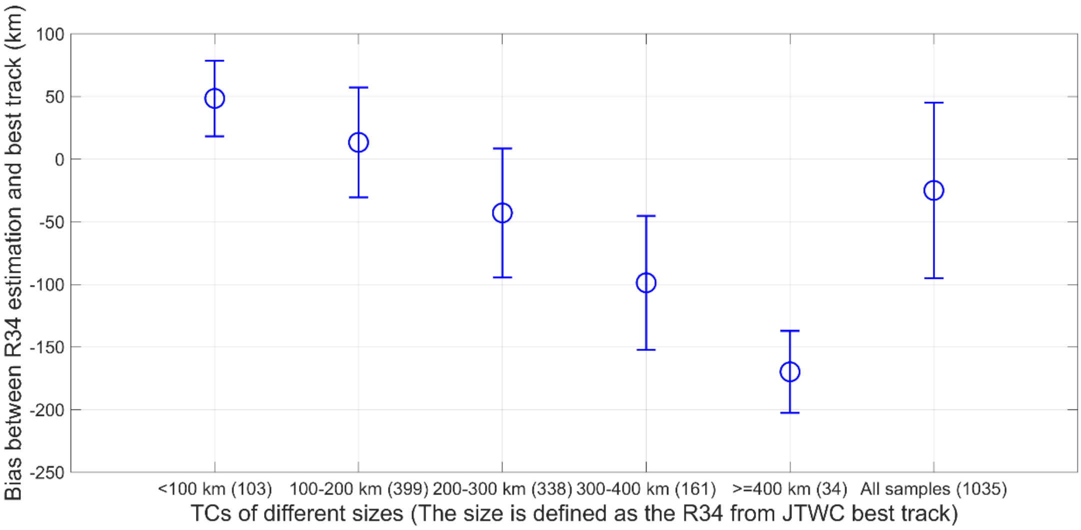


Fig. 6 Estimation bias of R34 for TCs grouped by size compared with the JTWC best track data in the WNP between 2017 and 2019. The sample size is as in Fig. 4. Numbers in parentheses give the sample size.

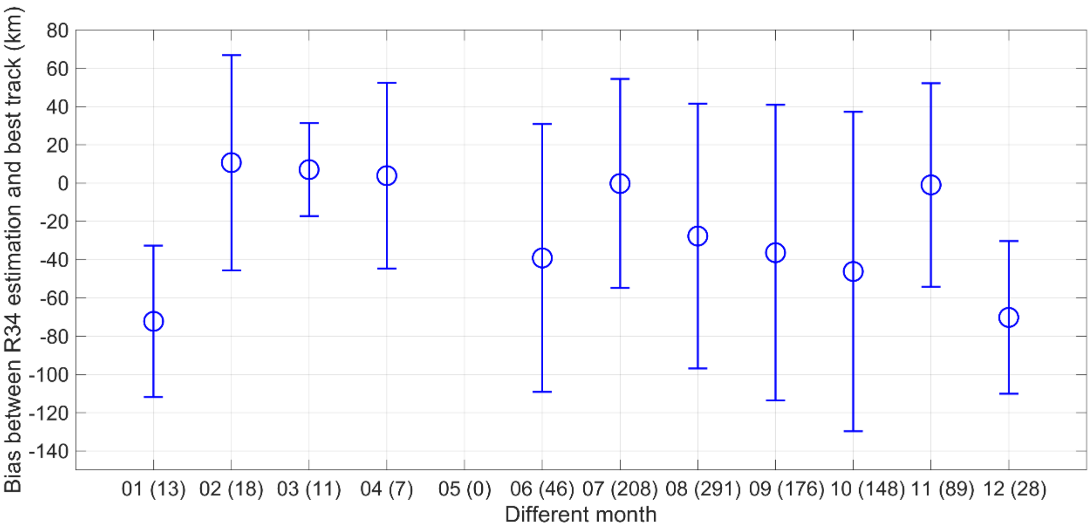


Fig. 7 Estimation bias of R34 for TCs in different months compared with the JTWC best track data in the WNP between 2017 and 2019. The sample size is as in Fig. 4. Numbers in parentheses give the sample size.

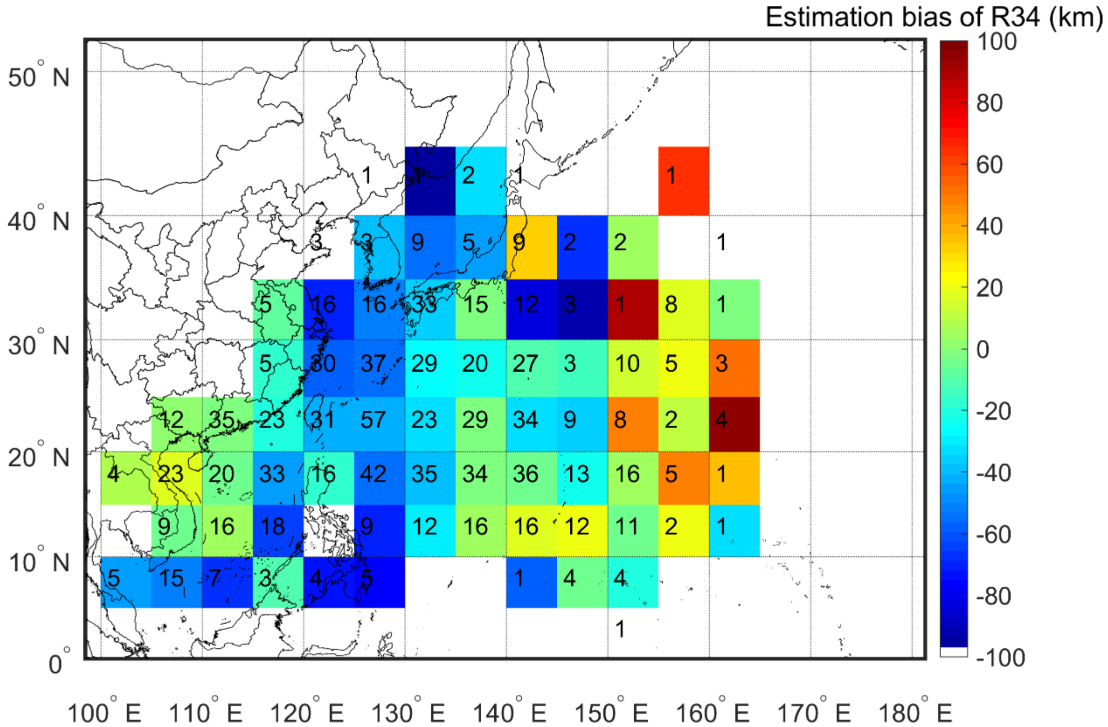


Fig. 8 Spatial distribution of estimation bias of R34 compared with the JTWC best track data in the WNP between 2017 and 2019. The sample size is as in Fig. 4. The number in each grid is the sample size.

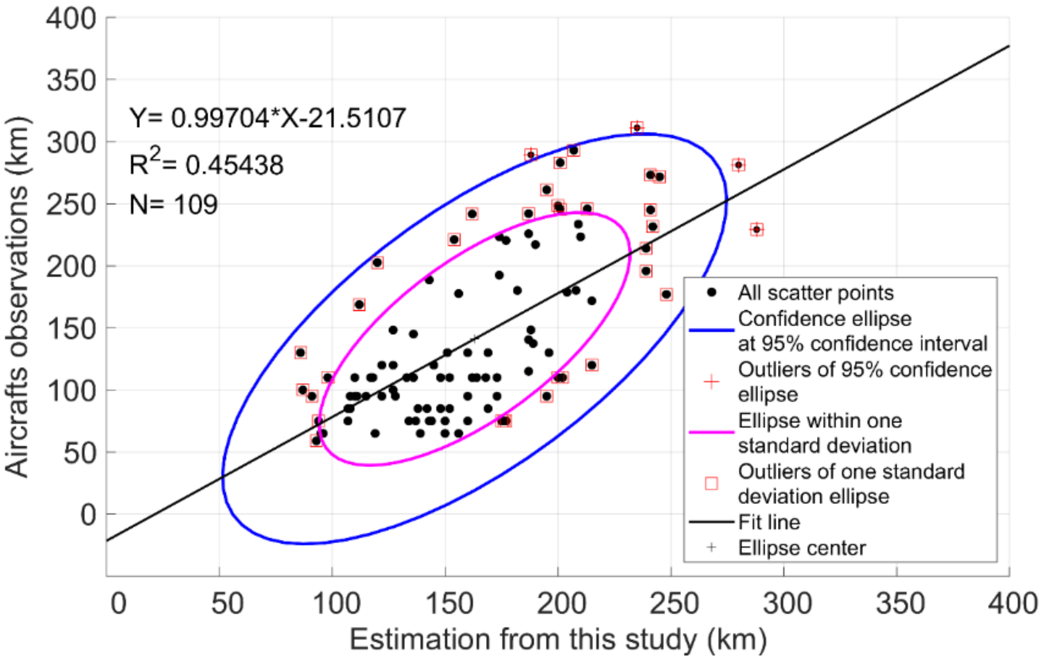
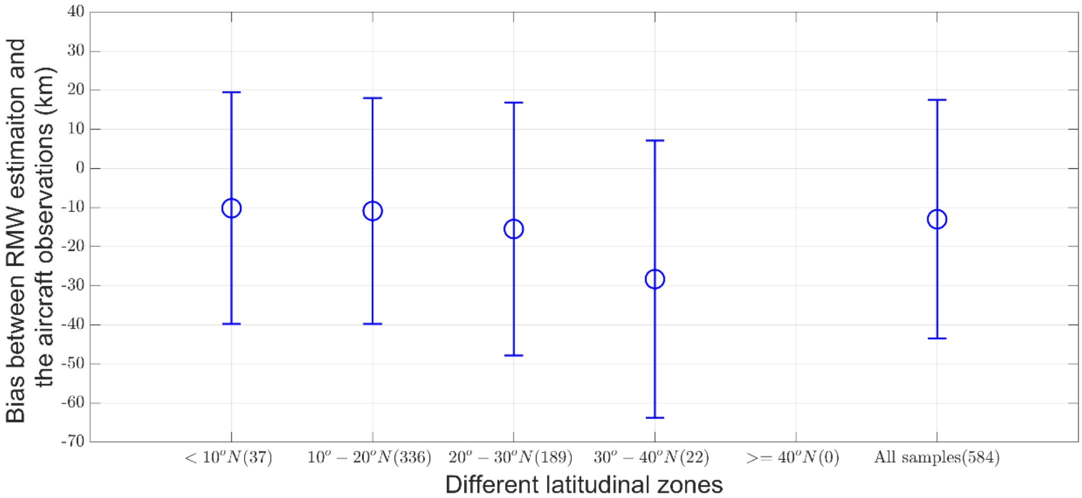


Fig. 9 Scatter plot of the IR-predicted R34 vs R34 from aircraft observations between 1981 and 1987. The black line is a linear fit between the two variables. N is the number of samples, and R^2 is the correlation coefficient (statistically significant at the 95% confidence level). The color represents the density of the scatter points.



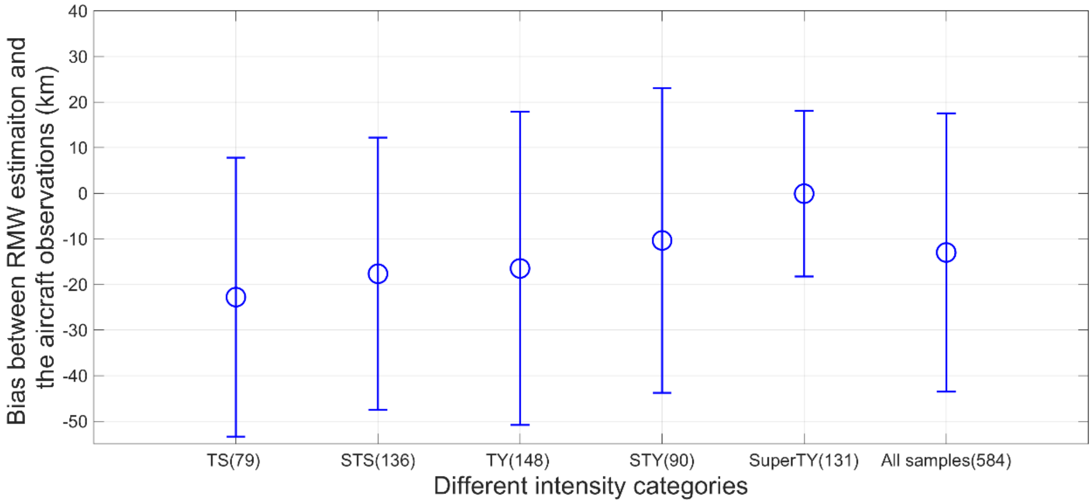


Fig. 10 Error bars for estimation of RMW at different latitudes (upper) and for different intensity categories (lower) compared with aircraft observations in the WNP between 1981 and 1987 (584 samples). Numbers in parentheses gives the sample size.

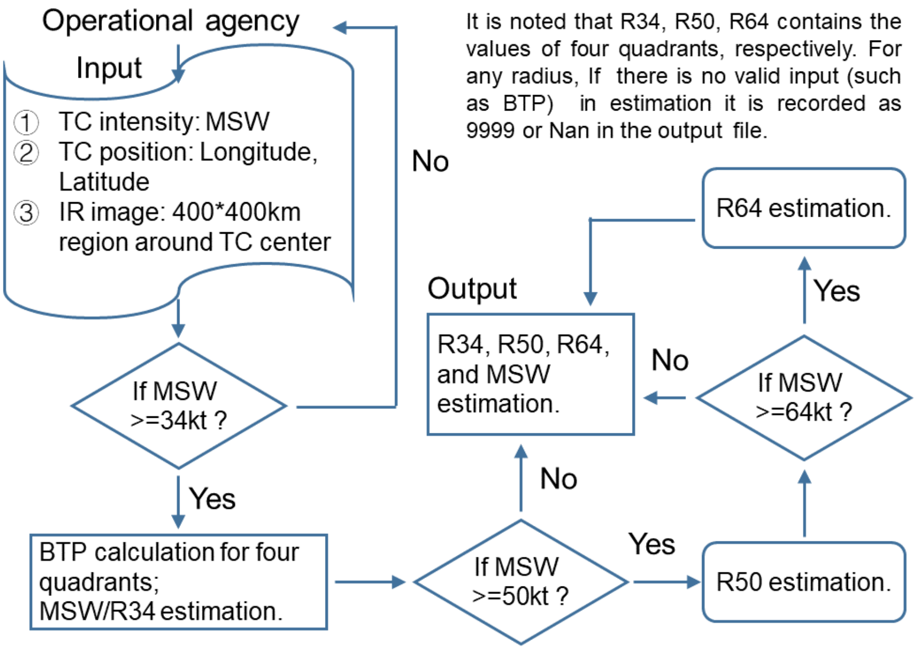


Fig. 11 Flow chart of the algorithms implemented in real-time operational applications.

List of Tables

Table 1 Parameters for the different machine learning methods used in the experiments.

Table 2 Difference between the JTWC best track data and the mean R50 and R64 estimated using different input schemes (MAE, km), 831 samples.

Table 1 Parameters for the different machine learning methods used in the experiments.

Algorithm	Parameter setting	Input
Multi-Layer Perceptron (MLP)	Epochs = 10000; Learning rate = 0.005; Learn function = ’tansig’; Transform function = ’purelin’; Max fail = 10; Goal = 0.01; Perform function = ’mse’; Hidden layer size = log2 N (N is the input size).	The longitude (Lon) and latitude (Lat) of TC center, TC intensity (MSW), and (BT) radial profile (BTP) within the radius R
General Regression Neural Network (GRNN)	Spread = 25.	Ditto
Radial Basis Function Network (RBFN)	Maximum number of neurons = 1000; Number of neurons to add between displays = 10; Spread = 25; Goal = 0.01.	Ditto
Support Vector Machine (SVM)	Kernel function = ’gaussian’; Kernel scale = ’auto’;	Ditto
Decision Tree (DT)	Number of trees = 50; Method = ’regression’.	Ditto

Table 2 Difference between the JTWC best track data and the mean R50 and R64 estimated using different input schemes (MAE, km) (831 samples)

	MLP	GRNN	RBFN	SVM	DT
R50	16	15	38	20	17
R64	12	11	46	14	13