

Improvements in Typhoon Intensity Change Classification by Incorporating an Ocean Coupling Potential Intensity Index into Decision Trees^{*,†}

SI GAO, WEI ZHANG, AND JIA LIU

International Joint Laboratory on Climate and Environment Change and Key Laboratory of Meteorological Disaster, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, China

I.-I. LIN

Department of Atmospheric Sciences, National Taiwan University, Taipei, Taiwan

LONG S. CHIU

Department of Atmospheric, Oceanic, and Earth Sciences, College of Science, George Mason University, Fairfax, Virginia

KAI CAO

Department of Geography, National University of Singapore, Singapore

(Manuscript received 11 May 2015, in final form 9 December 2015)

ABSTRACT

Tropical cyclone (TC) intensity prediction, especially in the warning time frame of 24–48 h and for the prediction of rapid intensification (RI), remains a major operational challenge. Sea surface temperature (SST) based empirical or theoretical maximum potential intensity (MPI) is the most important predictor in statistical intensity prediction schemes and rules derived by data mining techniques. Since the underlying SSTs during TCs usually cannot be observed well by satellites because of rain contamination and cannot be produced on a timely basis for operational statistical prediction, an ocean coupling potential intensity index (OC_PI), which is calculated based on pre-TC averaged ocean temperatures from the surface down to 100 m, is demonstrated to be important in building the decision tree for the classification of 24-h TC intensity change ΔV_{24} , that is, RI ($\Delta V_{24} \geq 25$ kt, where $1 \text{ kt} = 0.51 \text{ m s}^{-1}$) and non-RI ($\Delta V_{24} < 25$ kt). Cross validations using 2000–10 data and independent verification using 2011 data are performed. The decision tree with the OC_PI shows a cross-validation accuracy of 83.5% and an independent verification accuracy of 89.6%, which outperforms the decision tree excluding the OC_PI with corresponding accuracies of 83.2% and 83.9%. Specifically for RI classification in independent verification, the former decision tree shows a much higher probability of detection and a lower false alarm ratio than the latter example. This study is of great significance for operational TC RI prediction as pre-TC OC_PI can skillfully reduce the overestimation of storm potential intensity by traditional SST-based MPI, especially for the non-RI TCs.

1. Introduction

Tropical cyclones (TCs) usually cause huge economic losses and casualties through strong winds, flooding, and storm surge during and after making landfall (e.g., Zhang et al. 2009). Timely and accurate prediction of TC tracks and intensity therefore plays an important role in disaster mitigation. Although both operational track and intensity predictions have been significantly improved in recent decades, the intensity forecast still lags the track forecast

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/WAF-D-15-0062.s1>.

† Earth System Modelling Center Contribution Number 080.

Corresponding author address: Wei Zhang, College of Atmospheric Science, Nanjing University of Information Science and Technology, 219 Ningliu Road, Nanjing 210044, China.
E-mail: wzhang@nuist.edu.cn

and is a big challenge since the improvements of intensity guidance for the warning time frame (24–48 h) are the slowest (DeMaria et al. 2014) and the skill of rapid intensification (RI) prediction is still very poor (Kaplan et al. 2010).

TC intensity change is affected by a combination of complicated physical processes (e.g., Wang and Wu 2004; Elsberry et al. 2013) that can be divided into three main categories: environmental effects such as vertical wind shear and humidity (e.g., Merrill 1988; DeMaria 1996; Hanley et al. 2001; Emanuel et al. 2004; Yu and Kwon 2005; Zeng et al. 2007, 2008, 2010; Wang 2009; Hendricks et al. 2010; Ge et al. 2013); underlying surface forcing such as sea surface temperature, sea spray, and air–sea exchange coefficients (e.g., Emanuel 1986; Shay et al. 2000; Lin et al. 2009; Gao and Chiu 2010; Ito et al. 2015); and internal dynamics such as eyewall replacement and vortex Rossby waves (e.g., Montgomery and Kallenbach 1997; Sitkowski et al. 2011). Physics-based numerical simulation and statistical analysis have been shown to be effective tools for understanding the impacts of the physical processes on TC intensity change. Statistical–dynamical models have been used to predict intensity and rapid intensification probability (e.g., Knaff et al. 2005; Kaplan et al. 2010; Gao and Chiu 2012) and have shown greater skill compared to forecasts of individual physics-based dynamical models; the statistical–dynamical model run with various inputs from dynamical models is still being used to produce consensus forecasts over the western North Pacific (WNP) (DeMaria et al. 2014).

Unlike traditional statistical approaches, data mining is referred to as “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data” (Fayyad et al. 1996; Leung 2010). These approaches adapt data not normally suitable for statistical models with strict assumptions (such as independence, stationarity of the underlying processes, and normality). Data mining methods can be employed to unravel classification, clusters, association rules, decision rules, and other patterns from archived databases (Han and Kamber 2006; Leung 2010). For TC research, association rule mining has been successfully used to discover association rules for the rapid intensification of Atlantic hurricanes (Yang et al. 2007, 2008, 2011). In addition, a decision tree approach [e.g., the C4.5 algorithm; Quinlan (1993)] has also been used for the classification of TC intensity change (Zhang et al. 2013a, hereafter referred to as Z13A), as well as analyses of TC genesis, recurvature, and landfall over the WNP (Zhang et al. 2013b,c; Zhang et al. 2015). The decision tree method makes no prior assumptions for the multicollinearity, multivariate normality, nonautocorrelation, and homoscedasticity of

independent variables while many conventional multiple regression methods need to consider such assumptions. The superiority of decision trees compared to association rules is found in the fact that decision trees can accommodate continuous scale data while thresholds must be used to divide continuous data into categorical or ordinal scales so as to be analyzed by association rules. In contrast, classification algorithms such as neural networks can classify two types of samples. However, it is inconvenient to interpret the results derived from a neural network algorithm. Instead of an input–output black box of a neural network, decision trees can unravel key variables, thresholds for each selected variables, and rules based on the combination of variables and thresholds, and also trace back to the original samples.

In Z13A, the most important variable selected to build a decision tree was the intensification potential obtained by subtracting the current intensity from the maximum potential intensity (MPI; e.g., Emanuel 1988; DeMaria and Kaplan 1994; Holland 1997). MPI is the most important predictor in all of the statistical models (e.g., Knaff et al. 2005; Kaplan et al. 2010; Gao and Chiu 2012) and mining rules for TC intensity prediction (e.g., Yang et al. 2007, 2008, 2011; Z13A), in which MPI was empirically calculated from sea surface temperature (SST). Lin et al. (2013) developed a new Ocean Coupling Potential Intensity index (OC_PI), which effectively reduces overestimation of the archived maximum intensity by SST-based MPI and increased the correlation of maximum intensity estimation, especially for slowing-moving TCs (Lin et al. 2013). It is thus intriguing to examine whether OC_PI can improve the intensity change classification of TCs over the WNP using the C4.5 algorithm. The objectives of this study are to assess (i) whether and to what extent the integration of OC_PI in the classification model can improve its performance and (ii) how the model (i.e., the decision tree) will change after the integration of OC_PI. This study is expected to improve the classification accuracy of TC intensity change and provide valuable references for the prediction, analysis, and mitigation of TC-related hazards in coastal regions.

The data and methodology are described in section 2. Section 3 presents an improved decision tree with OC_PI for TC intensity change classification, followed by a summary and discussion given in section 4.

2. Data and methodology

a. Data

The 6-hourly location and 10-min maximum sustained wind speed (MWS) of WNP TCs are obtained from the final analysis of best-track data produced by the

Regional Specialized Meteorological Center (RSMC) Tokyo-Typhoon Center. The 24-h change in TC intensity (i.e., change in MWS, ΔV_{24} , during 0 and 24 h) can be regarded as a binary classification problem (i.e., RI or non-RI). The following criteria are used to define two classes: RI ($\Delta V_{24} \geq 25$ kt, where 1 kt = 0.51 m s^{-1}) and non-RI ($\Delta V_{24} < 25$ kt), where 25 kt day $^{-1}$ corresponds to the 95th percentile of the 24-h intensity change, which is often used as a threshold to define the RI of TCs (e.g., Kaplan and DeMaria 2003). TC cases are sampled if their intensity changes meet the criteria. The TC cases with $\Delta V_{24} \geq 25$ kt are labeled as class 1 whereas those with $\Delta V_{24} < 25$ kt are labeled as class -1.

Environmental data are derived from the National Centers for Environmental Prediction's (NCEP) Global Forecasting System (GFS) Final (FNL) gridded analysis at $1^\circ \times 1^\circ$ spatial and 6-h temporal resolution. The FNL analysis incorporates the most complete set of observational data and is likely the best option for a long-term operational model archive from NCEP (NOAA/National Centers for Environmental Prediction 2000). The variables used in this study include the wind at 200 and 850 hPa, as well as the air temperature and relative humidity at all levels ranging from 100 to 1000 hPa. The 200-hPa divergence and relative eddy flux convergence, along with the 850-hPa relative vorticity at each grid, are calculated using the wind field and the central difference method. All of the environmental variables are derived by averaging corresponding data within some specific radius from TC centers; that is, the relative humidity, temperature, zonal wind, and vertical wind shear are averaged in the circular ring with an outer radius of 800 km and an inner radius of 200 km, the divergence and relative vorticity are averaged in the circle with a radius of 1000 km, and the relative eddy flux convergence is averaged in the circle with a radius of 600 km.

Daily optimally interpolated SST (OISST) data at 0.25° spatial resolution (Wentz et al. 2000) from the composite of the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) and the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) satellite retrievals are provided by the Remote Sensing Systems (RSS). The other components of upper-ocean temperature profiles, including the depth of the 20°C (D20) and 26°C (D26) isotherms at daily and 0.25° resolutions, are derived using a two-layer reduced-gravity ocean model proposed by Shay et al. (2000) and sea surface height anomaly estimated from multiple satellite altimetry missions, including TOPEX/Poseidon, Jason-1, Jason-2, ERS-1, ERS-2, and the Environmental Satellite (ENVISAT) altimeters (Pun et al. 2007; Lin et al. 2013). Surface-80-m averaged ocean temperature (T80a) and surface-100-m averaged ocean

temperature (T100a) are then calculated by linearly interpolating the upper-ocean temperature profiles. The TC-induced mixing depth that depends on the TC translation speed, size, and intensity, as well as on the upper-ocean thermal structure, is typically 60–100 m (Price 2009). Lin et al. (2013) found that T80a-based OC_PI greatly reduced the overestimation of the TC maximum intensity by SST-based MPI, especially for slow-moving TCs, while T100a-based OC_PI shows the highest correlation with the observed maximum intensity for overall TCs (see their supplementary Fig. 9e); therefore, T80a and T100a are both used for OC_PI calculation to figure out which one is best for the overall TC intensity change classification. Upper-ocean heat content (UOHC) is also calculated from the surface to D26 following Leipper and Volgenau (1972).

MPI can be estimated empirically (DeMaria and Kaplan 1994) and theoretically (Emanuel 1988; Holland 1997). Empirical MPI (MPI_E) is estimated using its statistical relationship with SST:

$$\text{MPI_E} = 29.76 + 108.1e^{0.1141(\text{SST}-30)}. \quad (1)$$

Theoretical MPI (MPI_T) is calculated using temperature sounding data as a function of pressure from NCEP FNL analysis and SST from the RSS and Emanuel's software package (<http://texmex.mit.edu/pub/emanuel/TCMAX/>), and is expressed by

$$\text{MPI_T} = \sqrt{\frac{\text{SST} - T_0}{T_0} \frac{C_k}{C_D} (k^* - k)}, \quad (2)$$

where T_0 is the TC outflow temperature that can be estimated via finding out how high up in the eyewall the air remains warmer than the distant environment by conducting a straightforward assessment of the given vertical profile of temperature in the storm environment; C_k is the enthalpy exchange coefficient; C_D is the drag coefficient; the ratio of C_k to C_D is assumed to be 0.9. Here, k^* is the saturation enthalpy of the sea surface, and k is the surface enthalpy in the TC environment, which can be assumed as the convective available potential energy of the sea surface and the near-surface air at the radius of maximum winds, respectively. The expression of k^* (Emanuel 1994) is shown as

$$k^* = \int_{\text{LNB}}^{\text{LFC}} R_d (\text{SST} - T_{\text{env}}) d \ln p, \quad (3)$$

where LFC and LNB are the level of free convection and level of neutral buoyancy, respectively. In addition, T_{env} is the sounding environmental temperature, R_d is the gas constant, and p is pressure.

TABLE 1. The 19 potential variables used to build the decision tree and their correlation coefficients r with 24-h typhoon intensity change.

Variable	Description	r
MWS0	Initial maximum wind speed (MWS)	-0.46
DMWS	Change in MWS during the past 12 h	0.44
JDAY	Absolute value of Julian day - 248	-0.03
LAT	Latitude of storm center	-0.35
LON	Longitude of storm center	0.18
SPD	Storm translational speed	-0.02
RHLO	Area-averaged (200–800 km) relative humidity at 850–700 hPa	0.16
RHHI	Area-averaged (200–800 km) relative humidity at 500–300 hPa	0.18
U200	Area-averaged (200–800 km) zonal wind at 200 hPa	-0.23
T200	Area-averaged (200–800 km) temperature at 200 hPa	-0.29
$\delta 200$	Area-averaged (0–1000 km) divergence at 200 hPa	0.09
REFC	Relative eddy flux convergence within 600 km at 200 hPa	0.05
SHR	Area-averaged (200–800 km) 200–850-hPa wind shear	-0.31
USHR	Area-averaged (200–800 km) 200–850-hPa zonal wind shear	-0.19
$\zeta 850$	Area-averaged (0–1000 km) 850-hPa relative vorticity	0.02
UOHC	Upper-ocean heat content	0.30
POT_E	Empirical maximum potential intensity (MPI) - MWS0	0.38
POT_T	Theoretical MPI calculated from SST - MWS0	0.43
POT_T100a	Theoretical MPI calculated from preyclone 0–100-m-depth averaged temperature - MWS0	0.45
POT_T80a	Theoretical MPI calculated from preyclone 0–80-m-depth averaged temperature - MWS0	0.42

The value of OC_PI is also computed using Emanuel's software package but by replacing SST with preyclone T80a (this OC_PI is called MPI_T80a),

$$\text{MPI_T80a} = \sqrt{\frac{\text{T80a} - T_0}{T_0} \frac{C_k}{C_D} (k^* - k)}, \quad (4)$$

or by preyclone T100a (this OC_PI is called MPI_T100a),

$$\text{MPI_T100a} = \sqrt{\frac{\text{T100a} - T_0}{T_0} \frac{C_k}{C_D} (k^* - k)}. \quad (5)$$

The reason for using the pre-TC condition is that T80a or T100a calculated from the pre-TC profile can be considered to be a proxy for the "during-TC coupling SST" (Price 2009), which usually cannot be observed. For the stages before and after the peak in TC intensity, MPI_T80a and MPI_T100a are computed based on the oceanic profiles 2 days before the intensity first reaches category 1 (i.e., $\text{MWS} \geq 64$ kt) and 2 days before the peak intensity, respectively.

All of the potential variables are summarized in Table 1. They are collected every 6 h during 2000–11, which represents the maximum period during which all the datasets used in this study are available. Only overwater TC samples are considered here because of the availability of upper-ocean temperature profiles.

b. Decision tree method

The decision tree method is known as a data mining approach for disentangling rules, patterns, and

knowledge for decision-making procedures from archived databases (Quinlan 1993). A root node and a set of splits and leaf nodes are essential components of a decision tree. In classification, a dataset is sequentially divided in line with the decision framework, and a class label (i.e., 1 or -1 in this study) is allocated to each observation according to the leaf node to which this observation belongs. The C4.5 algorithm is a widely used decision tree algorithm whose proper strategy involves the selection of a variable at each node that perfectly partitions samples into several classes via a specified measurement, "information gain," based on information-theoretic "entropy" (Quinlan 1993). The structure of a decision tree is determined by the tests performed at each node. At any given node, the C4.5 algorithm only tests the variable that produces the highest degree of discrimination between local classes.

Formally, we let S be the training set consisting of s data samples, where s_i is the number of observations in S that belong to class i ($i = 1, 2$). Here, S includes RI and non-RI TC cases while $s_1(s_2)$ represents the number of RI (non-RI) samples. The information (entropy) needed to classify S is

$$\text{Info}(S) = - \sum_{i=1}^m \frac{s_i}{s} \log_2 \left(\frac{s_i}{s} \right). \quad (6)$$

Hence, the amount of information needed to partition S into $\{S_1, S_2\}$ by variable X (e.g., relative humidity, temperature, zonal wind, and vertical wind shear) with m distinct values is

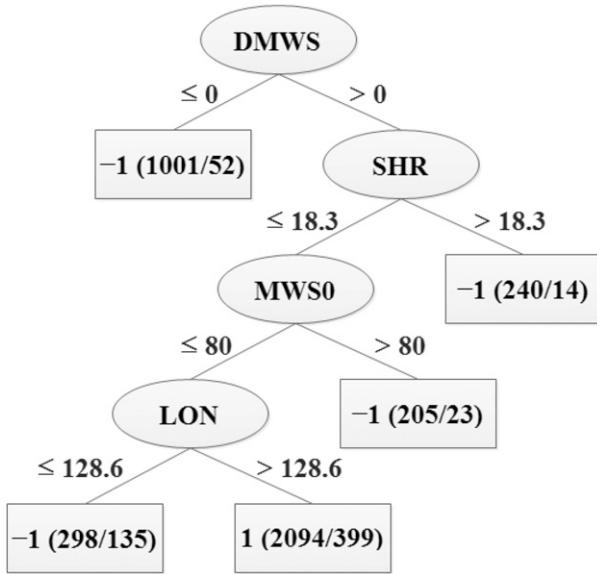


FIG. 1. The decision tree for typhoon intensity change classification constructed from the potential predictors in Table 1 except for UOHC, POT_T100a, and POT_T80a. Rectangles are leaf nodes and ellipses are parent nodes. The numbers in the leaf nodes indicate the classified class label (RI is 1; non-RI is -1), the total number of samples from both classes fulfilling the conditions of each tree path, and the number of misclassified samples.

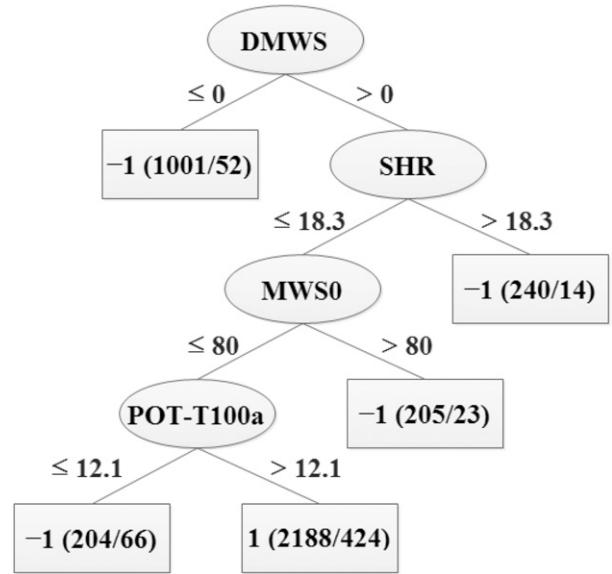


FIG. 2. As in Fig. 1, but for the decision tree that is constructed from all the potential predictors in Table 1.

$$\text{Info}_X(S) = - \sum_{i=1}^m \frac{s_i}{S} \times \text{Info}(S_i). \quad (7)$$

The information gained by partitioning S in accordance with the test X is

$$\text{gain}(X) = \text{Info}(S) - \text{Info}_X(S). \quad (8)$$

The gain criterion performs the selection of a test to maximize this information. The variable with the highest gain is selected for the first splitting in the construction of a decision tree. The succeeding partitions stop when the preset threshold (e.g., minimum leaf size) of the decision tree is reached.

Our database is imbalanced because the sample number of the RI class is much lower than the non-RI class. The classification results produced by the C4.5 algorithm tend to be biased if the two classes are imbalanced (Chawla 2003; Estabrooks et al. 2004; Han et al. 2005). Therefore, resampling is used to avoid biased results. The synthetic minority oversampling technique (SMOTE; Chawla 2003) is employed in this study to oversample the cases in the RI class.

c. Cross validation and independent verification

Cross validation is used for verification of the decision tree trained by samples over the period 2000–10. In

addition, samples during the year 2011 are used for independent verification. The k -fold cross-validation method is summarized as follows. The entire database is first divided into k equal-size parts. The training and validation are then carried out for k iterations. In each

TABLE 2. Means of 19 potential variables for RI class and non-RI class and their differences between two classes (RI and non-RI) based on 2000–11 data. Differences significant at the 95%, 99% and 99.9% level, appear in italic, boldface, and boldface italic respectively.

Variable	Unit	Mean		Difference
		RI	Non-RI	
MWS0	kt	57.7	68.8	<i>-11.1</i>
DMWS	kt	10.1	3.3	6.8
JDAY	day	24.8	28.1	<i>-3.3</i>
LAT	°	17.4	21.3	<i>-3.9</i>
LON	°	139.3	135.2	4.1
SPD	ms ⁻¹	9.6	9.4	0.2
RHLO	%	78.0	76.2	1.8
RHHI	%	61.6	56.9	4.7
U200	ms ⁻¹	-1.7	-0.2	<i>-1.5</i>
T200	°C	-50.7	-49.7	<i>-1.0</i>
δ200	10 ⁻⁶ s ⁻¹	6.2	5.0	1.2
REFC	ms ⁻¹ day ⁻¹	2.0	0.7	1.3
SHR	ms ⁻¹	15.1	17.1	<i>-2.0</i>
USHR	ms ⁻¹	0.1	1.1	-1.0
ζ850	10 ⁻⁶ s ⁻¹	10.8	11.5	-0.7
UOHC	kJ cm ⁻²	86.8	66.7	20.1
POT_E	kt	97.0	63.5	33.5
POT_T	kt	49.2	45.1	4.1
POT_T100a	kt	60.5	15.1	45.4
POT_T80a	kt	65.6	26.3	39.3

TABLE 3. Five unraveled rules governing TC intensity change in the decision tree shown in Fig. 2.

Rule No.	Rule description	Accuracy
1	If DMWS \leq 0 kt, then a TC will not rapidly intensify	949/1001 = 94.8%
2	If DMWS $>$ 0 kt, and SHR $>$ 18.3 m s ⁻¹ , then a TC will not rapidly intensify	226/240 = 94.2%
3	If DMWS $>$ 0 kt, and SHR \leq 18.3 m s ⁻¹ and MWS0 $>$ 80 kt, then a TC will not rapidly intensify	182/205 = 88.8%
4	If DMWS $>$ 0 kt, and SHR \leq 18.3 m s ⁻¹ , MWS0 \leq 80 kt, and LON \leq 128.6°, then a TC will not rapidly intensify	163/298 = 54.7%
5	If DMWS $>$ 0 kt, and SHR \leq 18.3 m s ⁻¹ , MWS0 \leq 80 kt, and LON $>$ 128.6°, then a TC will rapidly intensify	1695/2094 = 80.9%

iteration, a part of the dataset is used for validation whereas the remaining $k - 1$ parts are used for training. This is repeated until each part has been used once for validation. Tenfold cross validation is used in the present study; the decision trees are quite stable during the cross-validation procedure, as indicated in the supplementary material. As such, the numbers of correctly and incorrectly classified samples for 10 iterations are acquired. The prediction accuracy is then obtained by dividing the correctly classified samples by the number of samples in the entire database. Hence, cross validation ensures that each sample of the database can be used for both training and validation, and the training and testing samples are independent in all 10 iterations. This results in the high generalization capability of the decision tree as learned by the C4.5 algorithm.

3. Results

a. Decision trees

Two decision trees are built to show the advancement of OC_PI: the decision tree constructed from the potential predictors in Table 1, except UOHC, POT_T100a (MPI_T100a minus MWS0), and POT_T80a (MPI_T80a minus MWS0), and the decision tree constructed from all potential variables in Table 1. The two decision trees are indicated in Figs. 1 and 2, respectively. To better diagnose the relationship between the unravel rules and the TC intensity change, the composites of all the potential variables for RI and non-RI samples are performed and the results are shown in Table 2. All of the potential variables are significantly different except for storm translational speed, relative eddy flux convergence, 850-hPa relative vorticity, and zonal wind shear.

Figure 1 shows that previous 12-h TC intensity change (DMWS), vertical wind shear (SHR), initial intensity (MWS0), and longitude of TC center (LON) are selected to build this decision tree. These variables all show significant differences between RI and non-RI samples in Table 2 and are commonly used for constructing the RI index (Kaplan et al. 2010; Shu et al. 2012). Shu et al.

(2012) also found the significant difference in LON using data of a longer period and showed a preference for RI to occur eastward of 130°E. The C4.5 algorithm finds a more precise value of 128.6°E. The main physical differences may be a result of the land effect because more energy is dissipated by friction when a TC is approaching the landmasses of the Philippines and eastern China, and there are generally fewer energy supplies around the landmass than are available over the open ocean. The physical explanations of DMWS, SHR, and MWS0 will be discussed later in this section.

Among the five decision rules governing TC intensity change in Table 3, the highest accuracy is 94.8% for rule 1 whereas the lowest accuracy is 54.7% for rule 4. The low accuracy of rule 4 may be due to TCs offshore of East Asia and in the South China Sea (SCS) sometimes experiencing RI when the environmental conditions are favorable (e.g., Gao and Chiu 2010; Yu et al. 2013). Tenfold cross validation of the decision tree shows an accuracy of 83.2% (Table 4), and the accuracy of independent validation is only 83.9% (Table 5). As shown in the confusion matrix from independent validation in Table 5, only 4 out of 11 RI events are correctly classified and 27 out of 31 classified RI events are actually non-RI events; the results also indicate a low probability of detection (POD) of only 36.4% and a high false alarm ratio (FAR) of 87.1%. These statistics indicate the low skill for TC intensity change classification.

DMWS, SHR, MWS0, and POT_T100a are selected to build the decision tree (shown in Fig. 2) from all the potential predictors. We include POT_T100a in the decision tree, indicating its importance to TC intensity change classification. This also suggests that pre-TC

TABLE 4. Confusion matrix from cross validation of the decision tree shown in Fig. 1 (accuracy = 83.2%).

		Classified	
		RI	Non-RI
Observed	RI	1680	239
	Non-RI	406	1513

TABLE 5. Confusion matrix from independent verification of the decision tree shown in Fig. 1 (accuracy = 83.9%).

		Classified	
		RI	Non-RI
Observed	RI	4	7
	Non-RI	27	173

T100a is a better proxy for SST than pre-TC T80a for typhoon RI and non-RI classification.

Table 6 lists five rules governing TC intensity change. Rule 1 (if $DMWS \leq 0$ kt, then a TC will not rapidly intensify) shows that RI generally will not happen if a TC was in a decaying or steady state in the past 12 h, suggesting the importance of persistence for RI occurrence; this finding is consistent with previous statistical studies for TCs in different basins (e.g., Kaplan and DeMaria 2003; Kaplan et al. 2010; Shu et al. 2012). The composites in Table 2 also show that DMWS results for RI and non-RI samples are significantly different at the 99.9% level. This rule shows a high accuracy of 94.8%.

Rule 2 states that if $DMWS > 0$ kt, and $SHR > 18.3 \text{ m s}^{-1}$, then a TC will not rapidly intensify, suggesting that a TC in a high-shear environment will not rapidly intensify even if it was in a developing phase previously. As indicated in Table 2, the vertical wind shears of RI and non-RI samples are significantly different at the 99.9% level, and it has been widely accepted that high vertical wind shear is unfavorable for TC intensification. The most prominent mechanisms are the midlevel warming hypothesis of DeMaria (1996) and the venting hypothesis of Gray (1968) and Frank and Ritchie (2001). The classification accuracy of this rule is as high as 94.2%.

Rule 3 states that if $DMWS > 0$ kt, and $SHR \leq 18.3 \text{ m s}^{-1}$ and $MWS0 > 80$ kt, then a TC will not rapidly intensify. Previous development and relatively low vertical wind shear are both favorable for TC RI, whereas the TC will not rapidly intensify if $MWS0$ is larger than 80 kt. This is reflected in the composite of $MWS0$ in Table 2: the mean initial intensity of RI samples is significantly lower than that of non-RI samples at

the 99.9% level. A recent statistical study by Xu and Wang (2015) also found that RI most frequently occurs when initial TC intensity is in a narrow range of 70–80 kt. The dependence of the intensification rate on the initial TC intensity can be understood based on a simplified dynamical system for TC intensity prediction developed by DeMaria (2009): the intensification rate reaches a maximum when the TC is at its intermediate intensity. This rule has a high accuracy of 88.8%.

Rule 4 states that if $DMWS > 0$ kt, and $SHR \leq 18.3 \text{ m s}^{-1}$, $MWS0 \leq 80$ kt, and $POT_T100a \leq 12.1$ kt, then a TC will not rapidly intensify. If a relatively weak TC was previously intensifying and is in a low-shear environment, although these conditions are all favorable for RI, the TC tends not to rapidly intensify if POT_T100a is relatively low. Again, Table 2 shows the significant difference in POT_T100a between RI and non-RI samples at the 99.9% level. The TC that is close to its MPI hardly undergoes RI. The accuracy of this rule is 67.6%; although it is the lowest in this decision tree, it is much higher than the accuracy of rule 4 in the other decision tree (54.7%).

Rule 5 states that if $DMWS > 0$ kt, and $SHR \leq 18.3 \text{ m s}^{-1}$, $MWS0 \leq 80$ kt, and $POT_T100a > 12.1$ kt, then a TC will rapidly intensify. This rule indicates that intensification in the previous 12 h, low vertical wind shear, relatively weak current intensity, and current intensity far from the MPI_T100a are all favorable for RI, and RI generally take places when all four conditions are satisfied. The importance of POT_T100a for TC intensity change classification is highlighted. The accuracy of this rule is 80.6%.

Tenfold cross validation of this decision tree shows a classification accuracy of 83.5% (Table 7), and the accuracy of the independent verification is 89.6%. The confusion matrix from independent validation in Table 8 indicates that 9 out of 11 RI events are correctly classified and 20 out of 29 classified RI events are actually non-RI events; thus, the POD is 81.8% and the FAR is 70.0%. All of these statistics are better than those of the other decision tree built without OC_PI-related variables. The results demonstrate that the inclusion of

TABLE 6. Five unraveled rules governing TC intensity change in the decision tree shown in Fig. 1.

Rule No.	Rule description	Accuracy
1	If $DMWS \leq 0$ kt, then a TC will not rapidly intensify	949/1001 = 94.8%
2	If $DMWS > 0$ kt, and $SHR > 18.3 \text{ m s}^{-1}$, then a TC will not rapidly intensify	226/240 = 94.2%
3	If $DMWS > 0$ kt, and $SHR \leq 18.3 \text{ m s}^{-1}$ and $MWS0 > 80$ kt, then a TC will not rapidly intensify	182/205 = 88.8%
4	If $DMWS > 0$ kt, and $SHR \leq 18.3 \text{ m s}^{-1}$, $MWS0 \leq 80$ kt, and $POT_T100a \leq 12.1$ kt, then a TC will not rapidly intensify	138/204 = 67.6%
5	If $DMWS > 0$ kt, and $SHR \leq 18.3 \text{ m s}^{-1}$, $MWS0 \leq 80$ kt, and $POT_T100a > 12.1$ kt, then a TC will rapidly intensify	1764/2188 = 80.6%

TABLE 7. Confusion matrix from cross validation of the decision tree shown in Fig. 2 (accuracy = 83.5%).

		Classified	
		RI	Non-RI
Observed	RI	1673	246
	Non-RI	389	1530

POT_T100a improves the performance of the decision tree for TC intensity change classification.

In addition, because China suffers from RI TCs that have genesis in the SCS, the performance of the second decision tree in the SCS region (5° – 20° N, 105° – 120° E) for the whole period 2000–11 is evaluated since there are only 21 samples (all are non-RI events) of independent verification during 2011. The confusion matrix is shown in Table 9. This decision tree shows a total accuracy of 92.0%, one out of two RI events is captured and 103 out of 111 non-RI events are correctly classified.

b. Comparison of different intensification potential variables

It is surprising that the intensification potential POT_E (MPI_E minus MSW0), POT_T (MPI_T minus MSW0), or POT_T80a is not selected in both of the above decision trees, although both of them are significantly correlated with TC intensity change and show significant differences between the RI and non-RI classes. To examine why, the frequency distributions of POT_E, POT_T, POT_T80a, and POT_T100a, along with scatterplots of the other three variables versus POT_T100a for the RI and non-RI classes, are generated using the training data and are shown in Figs. 3 and 4, respectively. Here, POT_T100a is considered to be the reference for comparison since it is an effective OC_PI for classification and T100a-based OC_PI shows the highest correlation with the observed overall TC maximum intensity (Lin et al. 2013).

A notable increase in the frequency of the RI cases is observed for large POT_T100a values (Fig. 3a); 86% of the RI cases take place when POT_T100a exceeds 30 kt compared with only 39% of the non-RI cases, and 61% of the non-RI cases take place when POT_T100a is lower than 30 kt compared with only 14% of the RI cases. Therefore, POT_T100a is a significant variable for the RI and non-RI classification. For POT_T (Fig. 3b), RI and non-RI classes show rather similar distributions, as both events occur more frequently when POT_T is moderate; thus, it is difficult to distinguish between RI and non-RI with moderate POT_T. Figure 4a indicates that POT_T values of the non-RI cases are overestimated much more than those of the RI cases compared to POT_T100a, especially when POT_T100a is low, resulting in

TABLE 8. Confusion matrix from independent verification of the decision tree shown in Fig. 2 (accuracy = 89.6%).

		Classified	
		RI	Non-RI
Observed	RI	9	2
	Non-RI	20	180

high non-RI frequency with moderate POT_T. For POT_E (Fig. 3c), there are two peaks of RI frequency, high fractions of non-RI (47%) also occur around the second RI peak when POT_E is between 45 and 90 kt. Figure 4b shows that the POT_E results of the non-RI cases are also overestimated more than those of the RI cases, especially compared to low POT_T100a, leading to high non-RI frequency with moderate POT_E. Non-RI frequencies appear similar (53% versus 47%) when POT_T80a is lower or larger than 30 kt, although RI occurs much more frequently (89%) with POT_T80a of larger than 30 kt (Fig. 3d). Hence, it has difficulty identifying non-RI with moderate to high POT_T80a since again the POT_T80a results of the non-RI cases are overestimated more than those of the RI cases (Fig. 4c).

The overestimation of SST-based MPI (especially for the non-RI cases) might be because SSTs during TCs cannot be well observed because of rain contamination (Wentz et al. 2000; Gentemann et al. 2010a,b), and the optimally interpolated SSTs during TCs cannot well represent the cold wake produced by TCs (e.g., Mrvaljevic et al. 2013). The overestimation of MPI_T80a (especially for the non-RI cases) might be because T80a is less representative of the upper ocean's cooling effect on storm intensity than T100a. However, the lower but significant correlations of POT_T, POT_E, and POT_T80a (rather than POT_T100a) with TC intensity change seem reasonable because the weak or moderate wind-induced sea surface cooling generally cannot change the sign of the TC intensification rate (Cione and Uhlhorn 2003).

4. Summary and discussion

The OC_PI index proposed by Lin et al. (2013), which can more realistically characterize the ocean's

TABLE 9. Confusion matrix from cross validation and independent verification of the decision tree shown in Fig. 2 for the SCS TCs (accuracy = 92.0%).

		Classified	
		RI	Non-RI
Observed	RI	1	1
	Non-RI	8	103

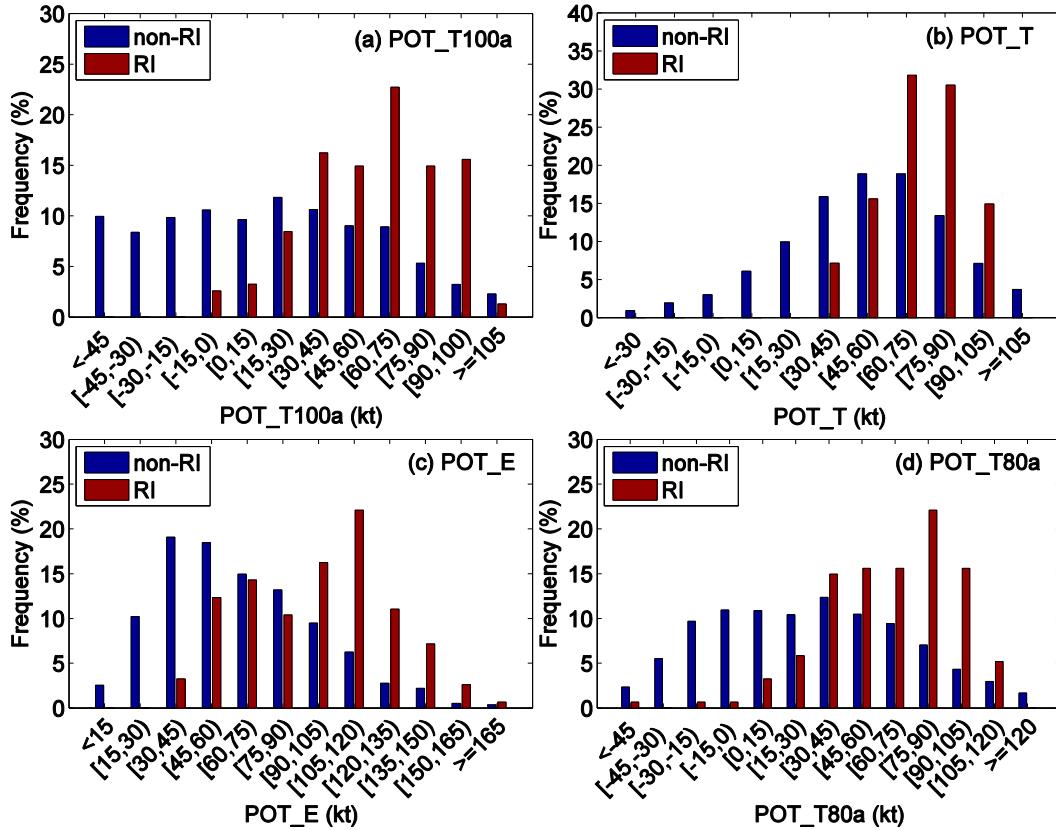


FIG. 3. Frequency distributions of (a) POT_T100a, (b) POT_T, (c) POT_E, and (d) POT_T80a for the RI and non-RI classes.

contribution to TC intensity than traditional SST-based MPI, is used to improve the decision tree for TC intensity change classification. The decision tree built with the database including OC_PI-related variables (i.e., POT_T80a and POT_T100a) outperforms that built with the database including traditional MPI-related variables (i.e., POT_T and POT_E), in terms of total classification accuracy, and POD and FAR for RI events in cross validation and independent verification. DMWS, SHR, MWS0, and LON are selected to build the former decision tree, and DMWS, SHR, MWS0, and POT_T100a are selected to build the latter.

The reasons why the other intensification potential variables are not selected to build the two decision trees are then examined. The results indicate that POT_T, POT_E, and POT_T80a underestimate the upper-ocean cooling effect by overestimating the intensification potential compared to POT_T100a, especially for the non-RI cases, resulting in relatively high non-RI occurrence with moderate values of POT_T, POT_E, and POT_T80a. Therefore, they are not as significant as

POT_T100a for RI and non-RI classifications in the decision trees.

The selection of POT_T100a to build the decision tree is of great significance for operational TC intensity prediction. Simultaneously observing SSTs during TCs is nearly impossible to do correctly by satellites, since infrared or even microwave radiation cannot penetrate through the rainfall around the TC eyewall, but it can penetrate through cloud cover a few days before the TC passage and thus pre-TC upper-ocean temperature profiles can be estimated and used as a good proxy for SSTs during TCs in an effort to calculate MPI, which can ultimately be used for operational TC intensity prediction. The performance of OC_PI in statistical typhoon intensity prediction and RI probability prediction merits being evaluated operationally, which will be carried out in our future studies.

It is noted that a few options for the pre-TC oceanic conditions are tested for OC_PI calculation: (i) at the time of the TC samples, (ii) 2 days before the TC intensity first reaches category 1 for all of the samples of an individual TC, and (iii) 2 days before the TC intensity

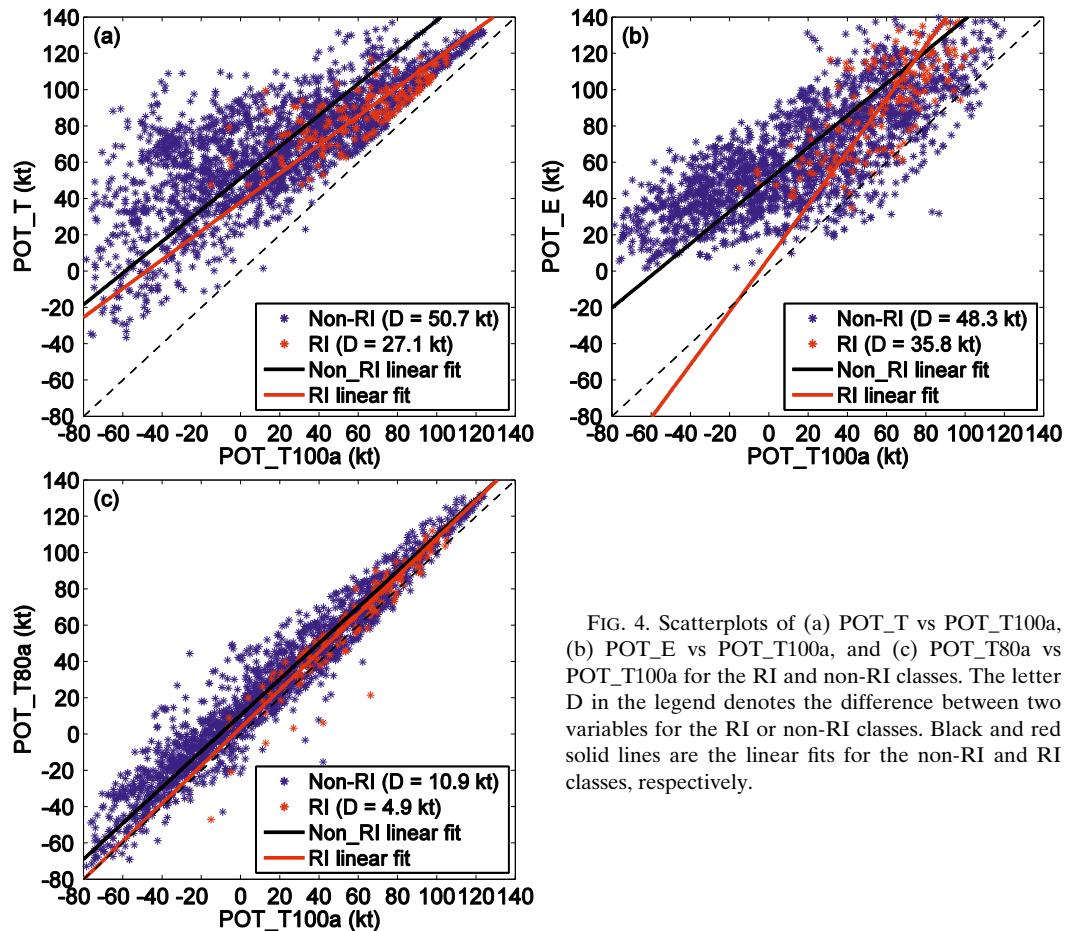


FIG. 4. Scatterplots of (a) POT_T vs POT_T100a, (b) POT_E vs POT_T100a, and (c) POT_T80a vs POT_T100a for the RI and non-RI classes. The letter D in the legend denotes the difference between two variables for the RI or non-RI classes. Black and red solid lines are the linear fits for the non-RI and RI classes, respectively.

first reaches category 1 for the TC samples of an individual TC before the peak and 2 days before the peak intensity for the samples of an individual TC after the peak. The third option shows the best performance and thus is included in our database to build the aforementioned decision trees.

Acknowledgments. The authors are grateful to four anonymous reviewers for their insightful comments and suggestions that strongly improved this paper. This study is supported by the National Natural Science Foundation of China (41201045, 41505035, and 41575078), the Jiangsu Natural Science Funds for Distinguished Young Scholar (BK20140047), the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (14KJB170015), the Jiangsu Shuangchuang Doctor Program, the Science and Technology Foundation for Selected Overseas Chinese Scholars from the Ministry of Human Resources and Social Security of China, the Fund for Selected Overseas Chinese Scholars of Nanjing, and the Startup Foundation for Introducing Talent of NUIST (2014r004 and 2013r116).

REFERENCES

- Chawla, N. V., 2003: C4.5 and imbalanced data sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure. *Proc. Int. Conf. on Machine Learning*, Washington, DC, International Machine Learning Society. [Available online at <https://www3.nd.edu/~dial/papers/ICML03.pdf>.]
- Cione, J. J., and E. W. Uhlhorn, 2003: Sea surface temperature variability in hurricanes: Implications with respect to intensity change. *Mon. Wea. Rev.*, **131**, 1783–1796, doi:10.1175//2562.1.
- DeMaria, M., 1996: The effect of vertical shear on tropical cyclone intensity change. *J. Atmos. Sci.*, **53**, 2076–2088, doi:10.1175/1520-0469(1996)053<2076:TEOVSO>2.0.CO;2.
- , 2009: A simplified dynamical system for tropical cyclone intensity prediction. *Mon. Wea. Rev.*, **137**, 68–82, doi:10.1175/2008MWR2513.1.
- , and J. Kaplan, 1994: Sea surface temperature and the maximum intensity of Atlantic tropical cyclones. *J. Climate*, **7**, 1324–1334, doi:10.1175/1520-0442(1994)007<1324:SSTATM>2.0.CO;2.
- , C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, doi:10.1175/BAMS-D-12-00240.1.
- Elsberry, R. L., L. Chen, J. Davidson, R. Rogers, Y. Wang, and L. Wu, 2013: Advances in understanding and forecasting

- rapidly changing phenomena in tropical cyclones. *Trop. Cyclone Res. Rev.*, **2**, 13–24.
- Emanuel, K. A., 1986: An air–sea interaction theory for tropical cyclones. Part I: Steady-state maintenance. *J. Atmos. Sci.*, **43**, 585–605, doi:10.1175/1520-0469(1986)043<0585:AASITF>2.0.CO;2.
- , 1988: The maximum intensity of hurricanes. *J. Atmos. Sci.*, **45**, 1143–1155, doi:10.1175/1520-0469(1988)045<1143:TMIOH>2.0.CO;2.
- , 1994: *Atmospheric Convection*. Oxford University Press, 580 pp.
- , C. DesAutels, C. Holloway, and R. Korty, 2004: Environmental control of tropical cyclone intensity. *J. Atmos. Sci.*, **61**, 843–858, doi:10.1175/1520-0469(2004)061<0843:ECOTCI>2.0.CO;2.
- Estabrooks, A., T. Jo, and N. Japkowicz, 2004: A multiple resampling method for learning from imbalanced data sets. *Comput. Intell.*, **20**, 18–36, doi:10.1111/j.0824-7935.2004.t01-1-00228.x.
- Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth, 1996: From data mining to knowledge discovery in databases. *AI Mag.*, **17**, 37–54.
- Frank, W. M., and E. A. Ritchie, 2001: Effects of vertical wind shear on the intensity and structure of numerically simulated hurricanes. *Mon. Wea. Rev.*, **129**, 2249–2269, doi:10.1175/1520-0493(2001)129<2249:EOVWSO>2.0.CO;2.
- Gao, S., and L. S. Chiu, 2010: Surface latent heat flux and rainfall associated with rapidly intensifying tropical cyclones over the western North Pacific. *Int. J. Remote Sens.*, **31**, 4699–4710, doi:10.1080/01431161.2010.485149.
- , and —, 2012: Development of statistical typhoon intensity prediction: Application to satellite observed surface evaporation and rain rate (STIPER). *Wea. Forecasting*, **27**, 240–250, doi:10.1175/WAF-D-11-00034.1.
- Ge, X., T. Li, and M. Peng, 2013: Effects of vertical shears and midlevel dry air on tropical cyclone developments. *J. Atmos. Sci.*, **70**, 3859–3875, doi:10.1175/JAS-D-13-066.1.
- Gentemann, C. L., T. Meissner, and F. J. Wentz, 2010a: Accuracy of satellite sea surface temperatures at 7 and 11 GHz. *IEEE Trans. Geosci. Remote Sens.*, **48**, 1009–1018, doi:10.1109/TGRS.2009.2030322.
- , F. J. Wentz, M. Brewer, K. A. Hilburn, and D. K. Smith, 2010b: Passive microwave remote sensing of the ocean: An overview. *Oceanography from Space*, V. Barale et al., Eds., Springer, 13–33.
- Gray, W. M., 1968: Global view of the origin of tropical disturbances and storms. *Mon. Wea. Rev.*, **96**, 669–700, doi:10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2.
- Han, H., W.-Y. Wang, and B.-H. Mao, 2005: Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. *Advances in Intelligent Computing*, D.-S. Huang, X.-P. Zhang, and G.-B. Huang, Eds., Springer, 878–887.
- Han, J., and M. Kamber, 2006: *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 770 pp.
- Hanley, D., J. Molinari, and D. Keyser, 2001: A composite study of the interactions between tropical cyclones and upper-tropospheric troughs. *Mon. Wea. Rev.*, **129**, 2570–2584, doi:10.1175/1520-0493(2001)129<2570:ACSOTI>2.0.CO;2.
- Hendricks, E. A., M. S. Peng, B. Fu, and T. Li, 2010: Quantifying environmental control on tropical cyclone intensity change. *Mon. Wea. Rev.*, **138**, 3243–3271, doi:10.1175/2010MWR3185.1.
- Holland, G. J., 1997: The maximum potential intensity of tropical cyclones. *J. Atmos. Sci.*, **54**, 2519–2541, doi:10.1175/1520-0469(1997)054<2519:TMPIOT>2.0.CO;2.
- Ito, K., T. Kuroda, K. Saito, and A. Wada, 2015: Forecasting a large number of tropical cyclone intensities around Japan using a high-resolution atmosphere–ocean coupled model. *Wea. Forecasting*, **30**, 793–808, doi:10.1175/WAF-D-14-00034.1.
- Kaplan, J., and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea. Forecasting*, **18**, 1093–1108, doi:10.1175/1520-0434(2003)018<1093:LCORIT>2.0.CO;2.
- , —, and J. A. Knaff, 2010: A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **25**, 220–241, doi:10.1175/2009WAF2222280.1.
- Knaff, J. A., C. R. Sampson, and M. DeMaria, 2005: An operational Statistical Typhoon Intensity Prediction Scheme for the western North Pacific. *Wea. Forecasting*, **20**, 688–699, doi:10.1175/WAF863.1.
- Leipper, D. F., and D. Volgenau, 1972: Hurricane heat potential of the Gulf of Mexico. *J. Phys. Oceanogr.*, **2**, 218–224, doi:10.1175/1520-0485(1972)002<0218:HHPOTG>2.0.CO;2.
- Leung, Y., 2010: *Knowledge Discovery in Spatial Data*. Springer-Verlag, 360 pp.
- Lin, I.-I., C.-H. Chen, I.-F. Pun, W. T. Liu, and C.-C. Wu, 2009: Warm ocean anomaly, air sea fluxes, and the rapid intensification of tropical cyclone Nargis (2008). *Geophys. Res. Lett.*, **36**, L03817, doi:10.1029/2008GL035815.
- , and Coauthors, 2013: An ocean coupling potential intensity index for tropical cyclones. *Geophys. Res. Lett.*, **40**, 1878–1882, doi:10.1002/grl.50091.
- Merrill, R. T., 1988: Environmental influences on hurricane intensification. *J. Atmos. Sci.*, **45**, 1678–1687, doi:10.1175/1520-0469(1988)045<1678:EIOHI>2.0.CO;2.
- Montgomery, M. T., and R. J. Kallenbach, 1997: A theory for vortex Rossby waves and its application to spiral bands and intensity changes in hurricanes. *Quart. J. Roy. Meteor. Soc.*, **123**, 435–465, doi:10.1002/qj.49712353810.
- Mrvljajevic, R. K., and Coauthors, 2013: Observations of the cold wake of Typhoon Fanapi (2010). *Geophys. Res. Lett.*, **40**, 316–321, doi:10.1029/2012GL054282.
- NOAA/National Centers for Environmental Prediction, 2000: NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999 (updated daily). National Center for Atmospheric Research Computational and Information Systems Laboratory Research Data Archive, accessed 28 April 2014, doi:10.5065/D6M043C6.
- Price, J. F., 2009: Metrics of hurricane–ocean interaction: Vertically-integrated or vertically-averaged ocean temperature? *Ocean Sci.*, **5**, 351–368, doi:10.5194/os-5-351-2009.
- Pun, I. F., I. I. Lin, C. R. Wu, D. S. Ko, and W. T. Liu, 2007: Validation and application of altimetry-derived upper ocean thermal structure in the western North Pacific Ocean for typhoon intensity forecast. *IEEE Trans. Geosci. Remote Sens.*, **45**, 1616–1630, doi:10.1109/TGRS.2007.895950.
- Quinlan, J., 1993: *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 302 pp.
- Shay, L. K., G. J. Goni, and P. G. Black, 2000: Effects of a warm oceanic feature on Hurricane Opal. *Mon. Wea. Rev.*, **128**, 1366–1383, doi:10.1175/1520-0493(2000)128<1366:EOAWOF>2.0.CO;2.
- Shu, S., J. Ming, and P. Chi, 2012: Large-scale characteristics and probability of rapidly intensifying tropical cyclones in the western North Pacific basin. *Wea. Forecasting*, **27**, 411–423, doi:10.1175/WAF-D-11-00042.1.
- Sitkowski, M., J. P. Kossin, and C. M. Rozoff, 2011: Intensity and structure changes during hurricane eyewall replacement cy-

- cles. *Mon. Wea. Rev.*, **139**, 3829–3847, doi:10.1175/MWR-D-11-00034.1.
- Wang, Y., 2009: How do outer spiral rainbands affect tropical cyclone structure and intensity? *J. Atmos. Sci.*, **66**, 1250–1273, doi:10.1175/2008JAS2737.1.
- , and C.-C. Wu, 2004: Current understanding of tropical cyclone structure and intensity changes—A review. *Meteor. Atmos. Phys.*, **87**, 257–278, doi:10.1007/s00703-003-0055-6.
- Wentz, F. J., C. Gentemann, D. Smith, and D. Chelton, 2000: Satellite measurements of sea surface temperature through clouds. *Science*, **288**, 847–850, doi:10.1126/science.288.5467.847.
- 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:10.1175/WAF-D-14-00141.1.
- Yang, R., T. Jiang, and M. Kafatos, 2007: Improved associated conditions in rapid intensifications of tropical cyclones. *Geophys. Res. Lett.*, **34**, L20807, doi:10.1029/2007GL031241.
- , D. Sun, and T. Jiang, 2008: A sufficient condition combination for rapid intensifications of tropical cyclones. *Geophys. Res. Lett.*, **35**, L20802, doi:10.1029/2008GL035222.
- , T. Jiang, and D. Sun, 2011: Association rule data mining applications for Atlantic tropical cyclone intensity changes. *Wea. Forecasting*, **26**, 337–353, doi:10.1175/WAF-D-10-05029.1.
- Yu, H., and H. J. Kwon, 2005: Effect of TC–trough interaction on the intensity change of two typhoons. *Wea. Forecasting*, **20**, 199–211, doi:10.1175/WAF836.1.
- Yu, Z., D. Wu, and H. Yu, 2013: Rapid intensification near landfall of Typhoon Vicente (2012). *Trop. Cyclone Res. Rev.*, **2**, 1–12.
- Zeng, Z., Y. Wang, and C.-C. Wu, 2007: Environmental dynamical control of tropical cyclone intensity—An observational study. *Mon. Wea. Rev.*, **135**, 38–59, doi:10.1175/MWR3278.1.
- , L. Chen, and Y. Wang, 2008: An observational study of environmental dynamical control of tropical cyclone intensity in the Atlantic. *Mon. Wea. Rev.*, **136**, 3307–3322, doi:10.1175/2008MWR2388.1.
- , Y. Wang, and L.-S. Chen, 2010: A statistical analysis of vertical shear effect on tropical cyclone intensity change in the North Atlantic. *Geophys. Res. Lett.*, **37**, L02802, doi:10.1029/2009GL041788.
- Zhang, Q., Q. Liu, and L. Wu, 2009: Tropical cyclone damages in China: 1983–2006. *Bull. Amer. Meteor. Soc.*, **90**, 489–495, doi:10.1175/2008BAMS2631.1.
- Zhang, W., S. Gao, B. Chen, and K. Cao, 2013a: The application of decision tree to intensity change classification of tropical cyclones in western North Pacific. *Geophys. Res. Lett.*, **40**, 1883–1887, doi:10.1002/grl.50280.
- , Y. Leung, and J. C. L. Chan, 2013b: The analysis of tropical cyclone tracks in the western North Pacific through data mining. Part I: Tropical cyclone recurvature. *J. Appl. Meteor. Climatol.*, **52**, 1394–1416, doi:10.1175/JAMC-D-12-045.1.
- , —, and —, 2013c: The analysis of tropical cyclone tracks in the western North Pacific through data mining. Part II: Tropical cyclone landfall. *J. Appl. Meteor. Climatol.*, **52**, 1417–1432, doi:10.1175/JAMC-D-12-046.1.
- , B. Fu, M. S. Peng, and T. Li, 2015: Discriminating developing versus nondeveloping tropical disturbances in the western North Pacific through decision tree analysis. *Wea. Forecasting*, **30**, 446–454, doi:10.1175/WAF-D-14-00023.1.