博士學位論文

Statistical-dynamical typhoon intensity predictions in the western North Pacific based on track pattern clustering, decision tree and ocean coupling predictors

濟州大學校 大學院

海洋氣象學 協同科程

金 成 勳

2017年 8月

태풍 진로 패턴 군집분류 및 의사결정나무, 해양접합 예측인자 기반 북서태평양 통계-역학적 태풍강도예측

지도교수 문 일 주

김 성 훈

이 논문을 이학 박사학위 논문으로 제출함

2017年 6月

김성훈의 이학 박사학위 논문을 인준함

심사위원장 _____

위 원 _____

위 원 _____

위 원 _____

위 원 _____

제주대학교 대학원

2017年 6月

Statistical-dynamical typhoon intensity predictions in the western North Pacific based on track pattern clustering, decision tree and ocean coupling predictors

Sung-Hun Kim
(Supervised by professor Il-Ju Moon)

A thesis submitted in partial fulfillment of the requirement for the degree of Doctor of science.

2017. 6.

This thesis has been examined and approved.

Interdisciplinary Postgraduate Program
in Marine Meteorology

GRADUATE SCHOOL
JEJU NATIONAL UNIVERSITY

Contents

List of tables

List of figures

Abstract

A statistical-dynamical model for predicting tropical cyclone (TC) intensity has been developed using a track-pattern clustering (TPC) method and ocean-coupled potential predictors. Based on the fuzzy c-means clustering method, TC tracks during 2004-2012 in the western North Pacific (WNP) were categorized into five clusters and their unique characteristics were investigated. The predictive model uses multiple linear regressions, where the predictand or the dependent variable is the change in maximum wind speed relative to initial time. To consider TC-ocean coupling effect due to TC induced vertical mixing and resultant surface cooling, we also developed new potential predictors for maximum potential intensity (MPI) and intensification potential (POT) using depth-averaged temperature (DAT) instead of sea surface temperature (SST). All together, we used six static, 11 synoptic, and three DAT-based potential predictors.

Results from a series of experiments for the training period of 2004 - 2012 showed that the use of TPC and the DAT-based predictors improved TC intensity predictions remarkably. The model was tested on predictions of TC intensity for 2013 and 2014 which are not used in the training samples. Relative to the non-clustering approach, the TPC and DAT-based predictors reduced the prediction errors up to 10~28% at most lead time. The present model is also compared with four operational dynamical forecast models. At short leads (up to 24 hours) the present model has the smallest mean absolute errors. After 24-hour lead times, the present model still shows a good skill comparable to the best operational model.

The developed model which uses TPC and DAT-based predictors, CSTIPS-DAT, led to a significant improvement in intensity prediction of TC, but it still showed relatively large errors in a specific cluster, in which strongly-developing TCs and non-developing TCs coexist, particularly in the central WNP where environmental conditions are the most favorable for TC intensification. In order

to improve the prediction skill of CSTIPS-DAT, here we employed a decision tree algorithm which is most popular classification method, to classify the type of TCs in early stage, based on whether they will develop to a strong intensity (maximum wind speed of 70 kt) during their lifetime or not. For the binary classification, a decision tree with four leaf nodes was built using the Classification And Regression Tree (CART) algorithm. According to the discovered rules from the trained tree, DAT, latitude, and DAT-based MPI were major factors in determining the two types. The decision tree has classification accuracy of 92.5% for training period, and of 80.5% for test period. The fact that DAT and DAT-based MPI are selected in the decision tree implies that TC-induced vertical mixing process and pre-existing ocean thermal structures along the track play a major role in determining type of TCs. The present results finally suggest that the TC intensity prediction skill of CSTIPS-DAT can be further improved by establishing independent statistical models for the classified groups.

Keywords: statistical-dynamical model, tropical cyclone, intensity prediction, fuzzy c-means clustering, depth-averaged temperature, multiple linear regressions, decision tree.

1. Introduction

A tropical cyclone (TC) accompanied by strong winds, storm surges, high waves, and flooding is among the most dangerous natural hazards, and poses a great threat to a global population of almost a billion people (Peduzzi et al., 2012). Accurate predictions of TC tracks and intensities are required in order to reduce TC damage. Over the last quarter of a century, TC track forecasts have been steadily improving, while storm intensity prediction remains a big challenge (Knaff et al., 2005) because TC intensification involves highly complex interactions between the atmosphere, ocean, and structure of the storm (Rappaport et al., 2012), which make accurate predictions difficult. With substantial improvement in computational power, high-resolution dynamical modeling has become a useful tool for predicting TC tracks and intensities. For intensity predictions, however, traditional statistical approaches (Jarvinen and Neumann, 1979; Knaff et al., 2003) are still widely used in practice, which provide consistent and basic information.

DeMaria and Kaplan (1994) developed a Statistical Hurricane Intensity Prediction Scheme (SHIPS) combining statistical models and dynamical models. SHIPS, based on a multi-regression technique, uses predictors estimated from a dynamical forecast model as well as climatological and persistence predictors. The SHIPS has been used for hurricane intensity guidance in the North Atlantic (NA) and eastern North Pacific (ENP) at the National Hurricane Center (NHC) (DeMaria and Kaplan, 1994a; DeMaria and Kaplan, 1999; DeMaria et al., 2005). For the western North Pacific (WNP), the SHIPS-based Statistical Typhoon Intensity Prediction Scheme has been used at the Joint Typhoon Warning Center (JTWC) (Knaff et al., 2005). Until recently, most statistical-dynamical models use regression equations, where same predictors are used at each lead time based on the assumption that TC intensity is controlled by the same environmental predictors without taking TC tracks into

account. Recent studies, however, indicate that environmental predictors related to TC activity are strongly dependent on TC track patterns.

Camargo et al. (2007a, 2007b) developed a TC track clustering technique based on a regression mixture model and showed that each cluster has a unique correlation with sea surface temperature (SST) and large scale atmospheric circulation patterns. Chu and Zhao (2011) also developed a mixture Gaussian model to cluster TC tracks over the Weste rn North Pacific into several types and investigated long-term changes of TC attribute s (e.g., frequency, lifespan) for each type. Kim et al. (2011) classified the WNP TCs using a fuzzy c-means clustering method (FCM) and showed that large-scale environment predictors influence TCs in different ways depending on each cluster. Chu et al. (2010) also used FCM-based classification of TCs in the vicinity of Taiwan to investigate relations between the TC passage frequency and the environment of individual TC tracks and developed a statistical model for predicting the seasonal TC frequency for each cluster using a Bayesian regression scheme. Also based on track patterns, Kim et al. (2012) developed a statistical-dynamical model for predictions of seasonal TC activity, which predicts the spatial distribution of TC track density using the correlation between the seasonal TC frequency and the environmental predictors calculated from the dynamical prediction model. These studies strongly suggested that the relationship between TC and environmental factors depends on the TC track patterns and showed that clustering methods can improve the statistical TC predictions. Note that most of the clustering-based studies have focused on predictions of TC frequency and track density, and less on TC intensity.

It is well known that the cooling of the upper ocean by TC-induced mixing is one of the factors influencing TC intensity (Mainelli-Huber, 2000; Shay et al., 2000; Lin et al., 2008 & 2009; Moon and Kwon, 2012). Price (2009) suggested that depth-averaged temperature (DAT) is a better indicator for TC intensity change than the widely-used Ocean Heat Content

(OHC) index. This is because that DAT considers the TC-induced cooling effect and realistically characterizes the ocean contribution to TC intensity. Lin et al. (2013) also argued that DAT is a better index than sea surface temperatures (SST) in estimating the Maximum Potential Intensity (MPI), which is an upper bound of TC intensity widely used as a key predictor. Therefore, it would be interesting to examine the effect of the use of DAT on statistical predictions of TC intensity since no such attempt has been made in the literature. Particularly, it is worthwhile to explore the role of DAT in classified TC groups since the strength of the cooling effect depends on the ocean subsurface thermal structure, as well as various factors such as the storm translation speed, size, and intensity, which are connected to TC tracks.

The first aim of the present study, therefore, is to develop a Statistical Typhoon Intensity Prediction Scheme for the western North Pacific using a track pattern clustering method and DAT-based potential predictors. This new scheme is then used to investigate how much the clustering and the ocean coupling effect may contribute to the improvement of TC intensity predictions. To test the performance of our new scheme, a series of experiment were conducted to predict the intensity of WNP TCs for the training period (2004-2012) and the real-time prediction period (2013 and 2014).

A statistical-dynamical model that predicts the intensity of TC has shown a great performance comparable to dynamical models (e.g., Knaff et al., 2005; Kaplan et al., 2010; Gao and Chiu, 2012). We have also demonstrated that the present new statistical-dynamical model, CSTIPS-DAT, using clustering technique and predictors based on DAT for the WNP has led to a significant improvement in the prediction of TC intensity in the Korea Meteorological Administration (KMA) since 2015. However, the CSTIPS-DAT has a relatively large error for specific clusters, particularly for C2 cluster, which contains most TCs in the Western North Central Pacific (WNCP) that spend their lifetime in the tropic

where the environmental factors are favorable for TC development (Fig. 1). The C2 cluster is characterized by the strongest mean intensity, noticeable intensification and the wide spectrum of intensity which lead large variance of predictand. Therefore, the intensity prediction in this cluster shows relatively good performance for TCs with strong intensity which similar to the mean intensity of the cluster, but poor performance for non-developing TCs deviating from the mean intensity. For example, Typhoon Phanfone (2014), which was developed strongly to the maximum wind speed of 100 kt, was well predicted by CSTIPS-DAT at most prediction time, while Typhoon Faxai (2014), which was not developed to strong intensity, showed a large discrepancy between prediction and observation due to the overestimation (see Fig. 1 and Fig 2). These results suggest that if we can classify the type of TCs in the early stage, based on whether they will develop high intensity during their lifetime or not, and establish independent statistical models for the classified groups, the TC intensity prediction skill will be improved. Particularly for the clusters that have a bimodal distribution in TC intensity, in which many TCs are deviated from mean distribution of TC intensity.

Recently, several studies have noted that the global distribution of lifetime maximum intensity (LMI) of TCs is bimodal (Maganello et al., 2012; Kossin et al., 2013; Zhao et al., 2009; Lee et al., 2016). Knowing what causes the bimodal distribution in LMI is critical to develop the classification skill determining the TC type in early TC stage. However it has no consensus on the causes and mechanisms related to the occurrence of the bimodal distribution. Torn and Snyder (2012) argued that an artificially low number of category 3 hurricanes in the Atlantic may be caused by low resolution of Dvorak technique at category 3 intensities. Soloviev et al (2014) explained the bimodal distribution of LMI using the ratio of surface exchange coefficients Ck/Cd as a function of wind speed with a local maximum around 115 kt, in which the local maxima would be favorable for rapid intensification (RI),

leading to increase number of TCs in the second high-intensity peak. Lee et al. (2016, hereafter; LEE16) reported that the bimodality of the LMI distribution is related to the RI, reflecting two types of TCs: those that undergo RI during their lifetime (RI TCs) and those that do not (non-RI TCs). These types of TCs have an own unimodal distribution of LMI, the peak is around 120 kt for RI TCs and 45 kt for non-RI TCs, respectively. This bimodal characteristic in LMI distribution hampers accurate TC intensity prediction of statistical models due to an increase of variance of the predictand. Therefore, a successful classification of bimodal components from the LMI distribution may contribute to improving the performance of statistical intensity model.

In recent years, a decision tree approach, which is one of classification algorithms, has been widely employed to investigate the mechanism of TC development (Li et al., 2009; Zhang et al., 2013, 2015; Gao et al., 2016). Li et al (2009) employed the decision tree to discover the collective contributions to Atlantic hurricanes using sea surface temperature, water vapor, vertical wind shear, and zonal stretching deformation. Zhang et al. (2013) applied the decision tree to split the binary classification which becomes intensifying or weakening in 24 hours. The developed decision tree uses only three variables but shows a remarkable prediction accuracy of 90.2%. Zhang et al. (2015) investigated the classification of developing and non-developing tropical disturbances to tropical storm in the WNP using the decision tree approach. The developed classification model conducted with six rules, and it has a classification accuracy of 81.7 % for training and hindcast accuracy of 84.6%. Gao et al. (2016) used the decision tree to develop the RI prediction model which classifies RI event and non-RI event. They showed that the pre-storm ocean coupling potential intensity index, which uses DAT instead of sea surface temperature for the calculation of maximum potential intensity (MPI), led an improvement of RI classification accuracy about 6% in test period. All of these previous studies suggest that a decision tree approach can be applied to classify

the bimodal distribution of LMI even though this has never been attempted.

The second aim of the present study is to build a decision tree for classifying the type of TCs in the early stage depending on whether or not TCs will reach a certain high intensity during their lifetime. This classification will contribute to improvement of statistical intensity prediction model as well as a better understanding of the bimodal characteristic in the LMI distribution of TCs in the current climate, ultimately proving clues to the causes of the bimodality.

Section 2 provides a description of the dataset and the clustering and classification method. Section 3 describes the overall results of developed statistical-dynamical intensity prediction model. Section 4 describes the LMI type classification model. Summary and conclusion is in Section 5.

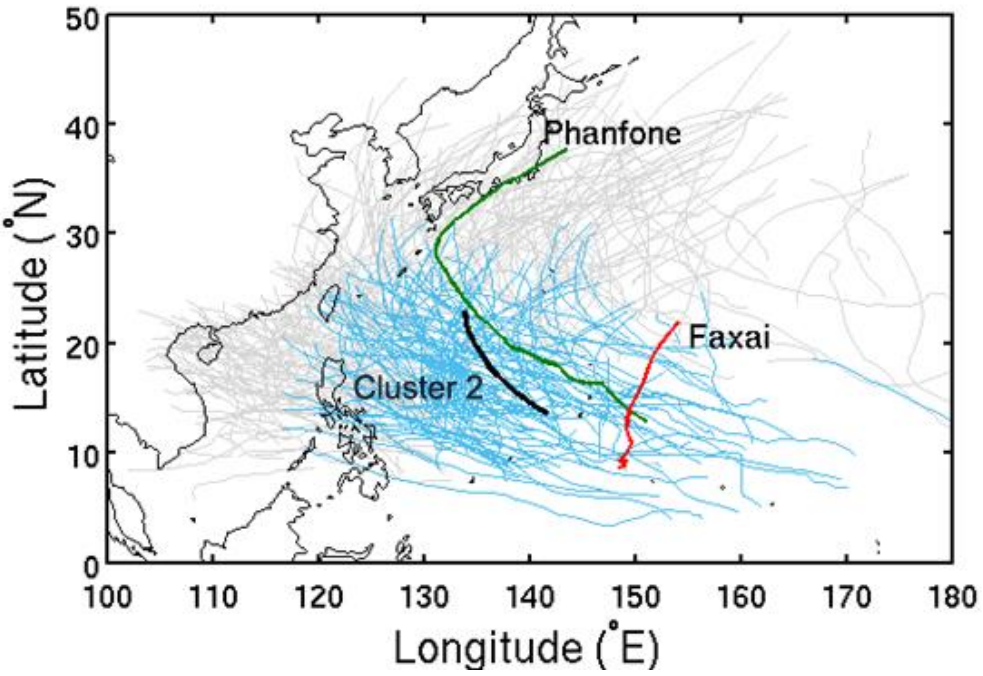

Fig. 1. (a) Tracks of typhoon Faxai (2014; red line) and Phanfone (2014; green line) along with the mean track for cluster 2 (thick black line), individual track for cluster 2 (blue lines) and individual track for WNP (gray lines).

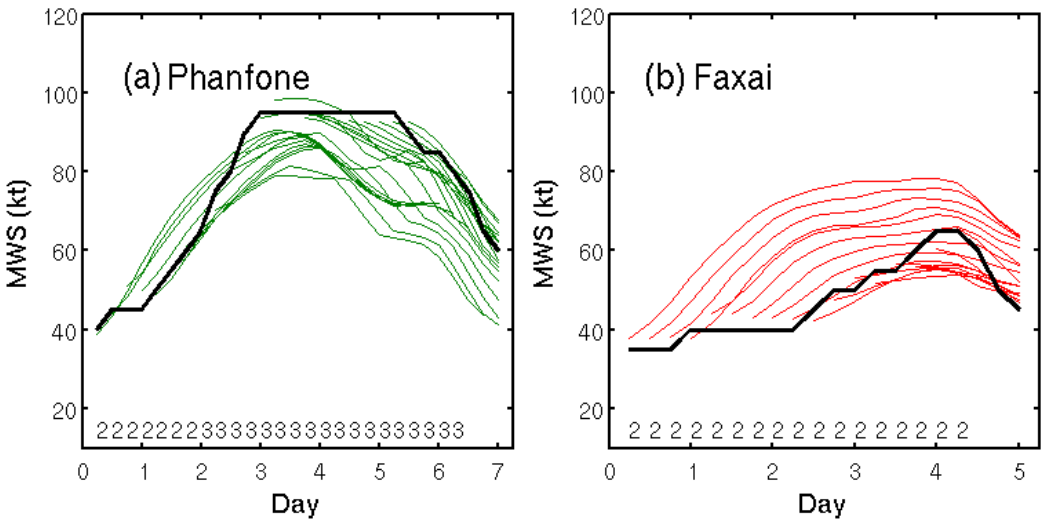

Fig. 2. Results of individual intensity predictions from CSTIPS-DAT for typhoon (a) Phanfone (2014) and (b) Faxai (2014). Thick black line is observation (RSMC best track data) and colored lines are individual CSTIPS-DAT predictions. The numbers above x axis denote the assigned cluster number.

2. Data and method

2.1. Data

The present Statistical Typhoon Intensity Prediction Scheme (hereafter, STIPS) is trained with TC data at 6-hr interval for the period of 2004 – 2012 (203 TCs) and tested independently with TCs in 2013 and 2014 for real-time prediction.

The TC data for the training and test periods were obtained from the best track data provided by the Regional Specialized Meteorological Center (RSMC) in Tokyo and the operational 5-day track prediction results from the Korea Meteorological Administration (KMA). The decision tree is trained with WNCP TCs for the period of 2004–2013 and validated independently with TCs for 2014–2016. The 6-hourly location and maximum wind speed of the TCs are obtained from the best track data produced by RSMC Tokyo.

Oceanic potential predictors were calculated using 3-dimensional ocean data derived from the Hybrid Coordinate Ocean Model (HyCOM) + Navy Coupled Ocean Data Assimilation (NCODA) Global Analysis (GLBa0.08) provided by the U. S. Naval Research Laboratory (NRL). HyCOM is a global ocean circulation model, which uses a Mercator projection and a hybrid coordinate system combining isopycnic, a terrain-following (sigma layers), and z-level coordinates (Bleck, 2002). The horizontal resolution of the model is 1/12°, which covers the global range between 78°S and 47°N. Atmospheric potential predictors were calculated using Global Forecast System (GFS) analysis data with 1 × 1 degree horizontal resolution and at 6-hour intervals provided by the National Centers for Environmental Prediction (NCEP).

In the present STIPS, the TC intensity was defined as maximum wind speed, the intensity change is predicted from the initial forecast time. For predictor selection, we used a

regressions approach that considers all possible subsets of independent variables and selects the variable subsets with a significant correlation at the 95% confidence level. We used the forward selection which is the most common screening procedure in selecting a good set of predictors from a pool of potential predictors. To avoid multicollinearity, potential predictors are carefully selected so that they are not significantly correlated with each other (Aczel, 1989; Fitzpatrick, 1997). During the combined regression process, if the sign of each individual regression coefficient for independent predictors is changed, it is removed from the candidates. If the regression fitting using a combination of predictors is worse compared with an individual predictor, the single best predictor is selected for the parsimonious reason.

2.2. Clustering method

The cluster classification technique is one of multivariate statistical techniques to identify similar object groups in multidimensional space. In other words, based on the statistical similarity between the objects constituting the entire, the objects at a close distance are classified into the same group, and the other objects are included in another group. Bezdek (1981) proposed a fuzzy c-mean clustering (FCM) algorithm to improve Hard c-Means clustering (HCM). FCM is a data classification algorithm that classifies each data point according to degree of belonging to each cluster. The number of sample N, is divided into C fuzzy groups, and the center of the cluster is searched in each group where the objective function of non-similarity measurement is minimum.

An important difference between FCM and HCM is that the FCM uses a fuzzy partition which a given data point can belong to several groups with a degree of membership between 0 and 1. The sum of degrees of membership is always 1. The objective function for FCM is defined by Dunn (1973) as follows.

$$J = \sum_{i=1}^{C} \sum_{k=1}^{N} (u_{ik})^m \|X_k - V_i\|^2 \qquad (1)$$

$$D_{ik} = \|X_k - V_i\|^2 \qquad (2)$$

Here, u_{ik} indicates how close the k-th data (X_k) is from the center of the i-th cluster. The u_{ik} has a value of 0-1 according to D_{ik}. Thus, the higher membership coefficient indicates a strong association between the i-th cluster and the k-th data. This is the main difference that distinguishes FCM from K-means clustering method. And m is a weighting index indicating the influence of the degree of fuzzy membership function. That is, if the value of m is set to a small value, more weight is given to data located closer to the center of the cluster. The cluster center vector (V_i) of the FCM is defined as the weighted average of all data and the membership coefficient. The FCM has a big advantage when it classifies data with unclear boundaries such as track data.

To cluster and classify TC tracks a fuzzy c-means clustering method (FCM, Bezdek, 1981) was used, a useful tool for the analysis of TC tracks (Kim et al., 2011). The fuzzy clustering is an extension of the k-mean clustering method. This algorithm allows objects to belong to several clusters with different degrees of membership. For clustering, 5-day TC tracks are used rather than the entire track of TC lifetime since the present STIPS is set to predict a total of 5 days. The dissimilarity between two tracks is defined as the Euclidean norm of the difference of two vectors which contain the latitudes and longitudes for each TC track.

Determining the optimal number of clusters is the most important process in the cluster analysis. The optimal number of clusters is difficult to determine due to the interpolation problem depending on the shape and size of the data distribution. Previous studies have determined it by calculating and comparing several indices. Partition Coefficient (PC) is an index representing the degree of redundancy between the clusters, and it is difficult to

directly relate to some properties of the data itself. It represents the optimal number of clusters when PC shows the maximum value. Next, Partition Index (SC) is the ratio of the closeness of the individuals in the cluster to the degree of separation. SC is useful when comparing different partitions with the same number of clusters, and a lower SC shows better partitions. The Xie and Beni`s index (XB) represents the ratio of the total and the variation of each cluster, and the Alternative Dunn's Index (ADI) represents the ratio of the minimum distance to the other clustering points to the maximum distance in the cluster. As two indices is lower, it means that a good classification.

Figure 3 shows the optimal cluster decision indices according to the number of clusters. PCs tend to decrease as the number of clusters increases. Considering PC only, the optimal number of clusters is two, but other indexes also tend to decrease with increasing clusters. Therefore, we need to find a suitable trade-off point to reflect all of these characteristics. ADI was similar to PC, but SC and XB were different locally. SC showed a very small reduction rate between five to six and XB showed a local minimum at five. In this study, the number of clusters is determined to be five according to the above results.

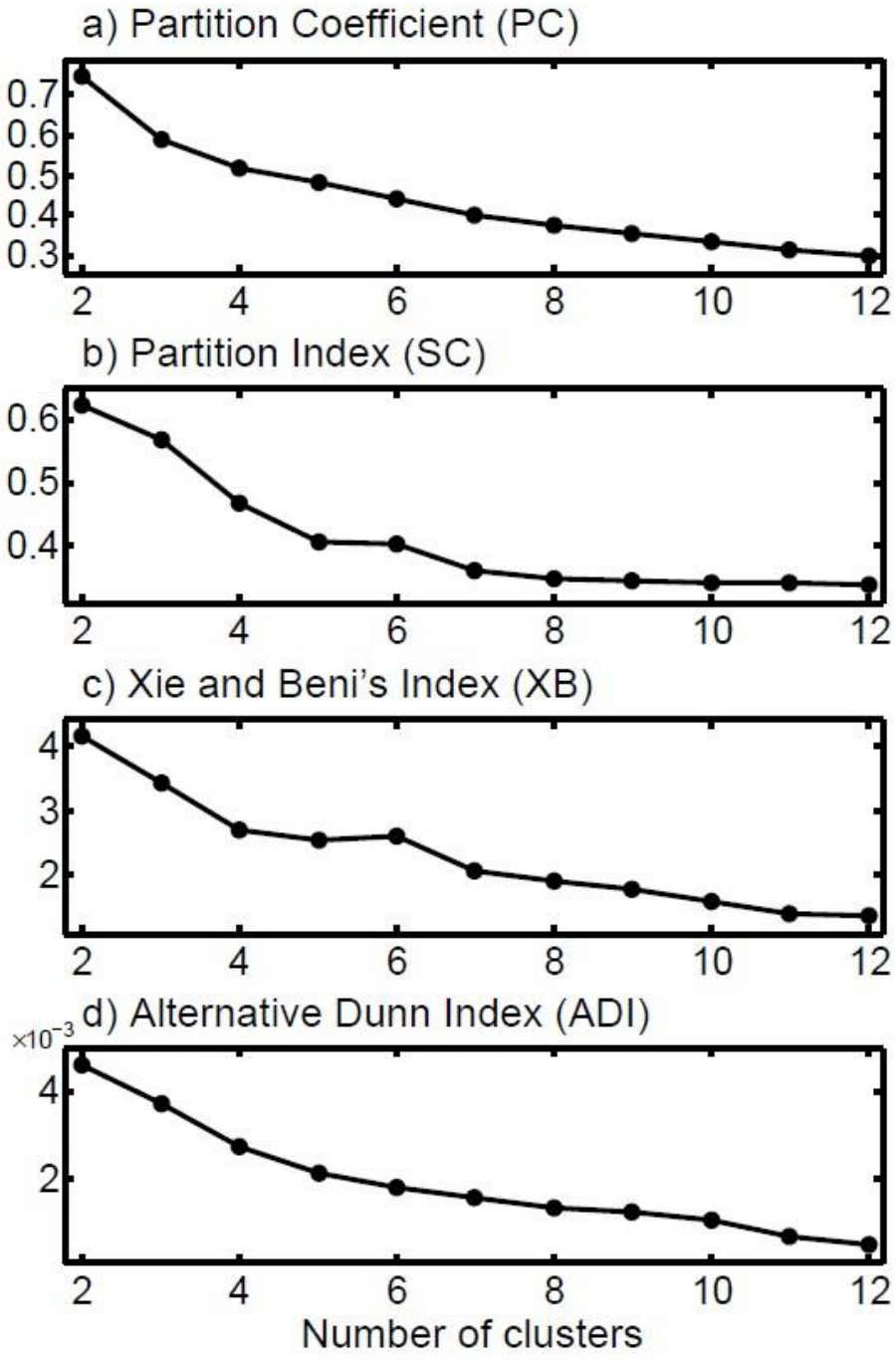

Fig. 3. Value of four scalar validity measures according to number of clusters. (a) Partition coefficient, (b) partition index, (c) Xie and Beni's index and (d) alternative Dunn index.

2.3. Characteristics of classified clusters

In this subsection, we examine the characteristics of the five clusters, C1 to C5, classified by the FCM. The largest cluster among the five clusters is C3 (23.6%). The TCs in C3 make landfall over Japan or the Korean Peninsula (Fig. 4, solid line). C1 is the second largest cluster (23.2%) and includes many TCs heading westward toward Hainan Island in China after passing through the Philippines. C4 (21.3%) includes TCs that land over Taiwan and Eastern China coasts. TCs in C2 (18.5%) mostly head northwestward from the eastern Philippines Sea. Type C5 has the lowest frequency of occurrence (13.4%) and its TCs pass mainly over the Kuroshio extension, farther away from the East Asian coast.

Each cluster exhibits distinctive features in mean maximum intensity (Table 1) and the mean intensity change (Table 2). TCs in C1 reside mainly in the South China Sea, where the upper ocean heat content is relatively lower than other tropical regions, and TCs are mostly likely to make landfall over Hainan Island and Guangdong. C1, along with C5, shows the lowest mean intensity (Table 1) and a distinct weakening tendency (Table 2) over the prediction lead time. TCs in C2 spend their lifetime in the tropic where vertical wind shear (VWS) is weak (Fig. 4a) and SST is high (Fig. 4b), which are favorable conditions for TC development, resulting in the strongest mean intensity and noticeable TC intensification compared with the other clusters. Many TCs in C2 subsequently evolve into TCs in C3 (26%) and C4 (37%), respectively. Most TCs in C3 experience the largest weakening stage as forecast leads increase mainly due to low SST and strong VWS along the track (Fig. 4a and 4b) and they finally make landfall over the Korea peninsula and Japan. Most TCs in C4 initially have relatively strong intensity (table 1), but rapidly weaken after landfall in Taiwan and eastern China. C5 passes over the eastern ocean of Japan where VWS is strong and SST is low, leading to significant weakening in the intensity.

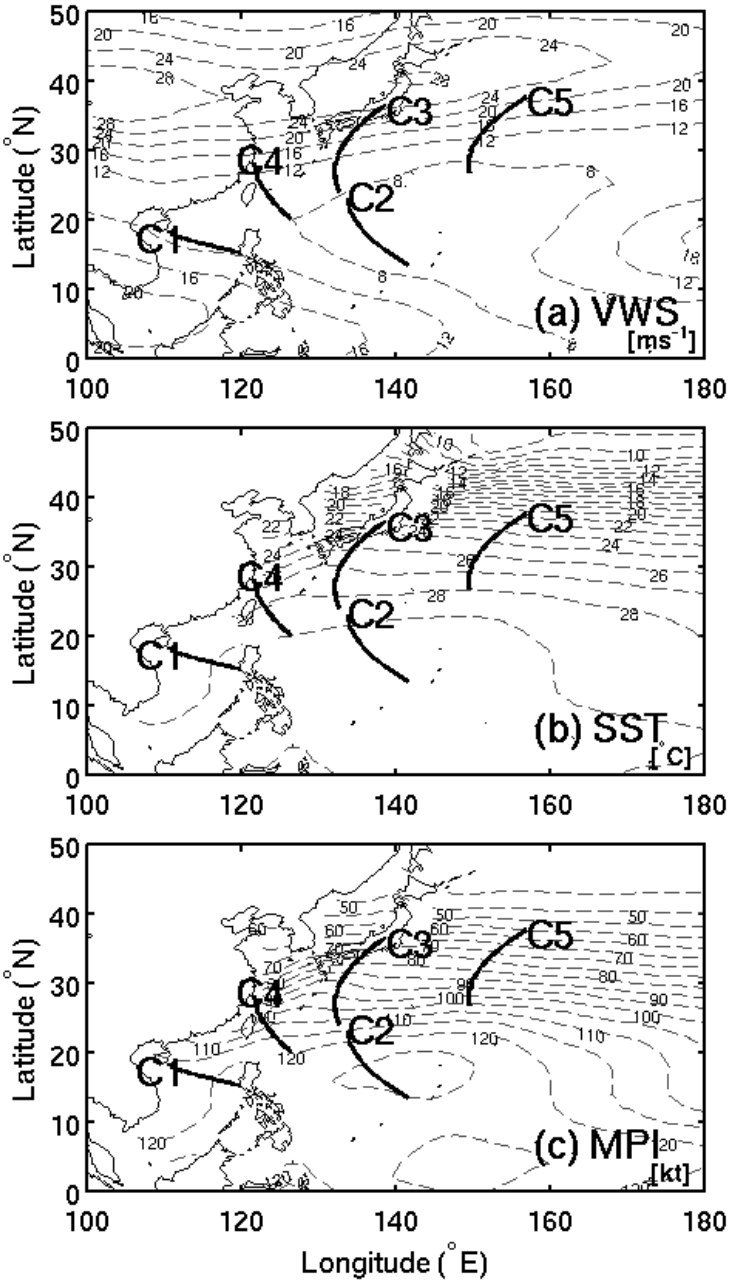

Fig. 4. Spatial distributions of climatological vertical wind shear in ms^{-1} (dashed lines) (a), sea surface temperatures in °C (dashed lines) (b) and MPI in kt (c) for typhoon season (JJASON) in the western North Pacific. Black curves represent the mean track of five clusters.

Also note that in Table 1, the difference in mean values between each cluster type and the non-clustering cases is overwhelmingly significant at the 5% test level (two-tailed t-test). Considering all five lead times and five clusters, 22 out of 25 type-lead show statistical significance. These differences become even more pronounced when the change in maximum wind speed is considered (Table 2). Except for one event (C4 at lead 24 hours), all other 24 type-lead exhibit statistical significance. Based on this simple comparison, it is beneficial to cluster TC tracks in a natural way to reduce variability in TC intensity change.

Table 1. Mean maximum wind speed (knot) of TCs for five clusters at each lead time. Boldface and boldface with asterisk represent that the differences in mean value between each cluster and all TCs are statistically significant at the 95% and 99% confidence level (two-tailed t-test), respectively.

Forecast time (h)	Cluster					
	C1	C2	C3	C4	C5	All TCs
24	**58.1***	**69.9***	**68.5***	**67.0***	**57.5***	64.4
48	**59.4***	**79.1***	67.2	68.0	**57.9***	67.0
72	**58.6***	**83.4***	**63.9***	66.3	**57.4***	67.7
96	**57.5***	**81.6***	**59.8***	**62.2***	**57.5***	66.9
120	**54.9***	**77.0***	**53.5***	**57.3***	**59.7**	64.9

Table 2. Same as in Table 1, but for the change in mean maximum wind speed (knot).

Forecast time (h)	Cluster					
	C1	C2	C3	C4	C5	All TCs
24	**0.1***	**12.2***	**-2.2***	0.6	**-1.6***	1.5
48	**-0.1***	**21.1***	**-4.8***	**0.7**	**-2.4***	3.1
72	**-1.8***	**25.1***	**-7.5***	**-0.9***	**-4.6***	3.8
96	**-4.5***	**23.3***	**-10.3***	**-4.3***	**-7.3***	3.3
120	**-6.7***	**18.3***	**-13.4***	**-10.2***	**-9.7***	2.1

2.4. Benefits of Cluster Analysis

As seen in Tables 1 and 2, the classified clusters show distinctive characteristics in terms of the mean intensity and tendency of intensity change. To quantitatively evaluate the effect of the use of clustering, we compared the standard deviations (SD) of each cluster before and after clustering at each lead time (Table 3), i.e. each cluster's SD from the entire sample mean (σ_{WNP}) and from the respective cluster's mean (σ_c).

$$\sigma_{WNP} = \left(\frac{\sum(X_c - \bar{X}_{WNP})^2}{n_c}\right)^{\frac{1}{2}} \qquad (3)$$

$$\sigma_c = \left(\frac{\sum(X_c - \bar{X}_c)^2}{n_c}\right)^{\frac{1}{2}} \qquad (4)$$

Here, X_c refers to the intensity change (*i.e.*, predictand) of an individual storm; $\bar{X}_c$ is the mean intensity change for each cluster; $\bar{X}_{WNP}$ is the mean intensity change for the entire sample; and n_C the number of samples for each cluster. The comparison of σ_c and σ_{WNP} reveals that clustering reduces the variability of the predictand (i.e., intensity change) with reduction rates ranging from 0.1% to as high as 21.0% (brackets in Table 3). The largest reduction (21%) was observed at 24-h and 72-h lead times in C2. The significant reduction of SD suggests that the clustering-based STIPS can reduce uncertainty in intensity change prediction and thus improve the TC intensity prediction skill. The clustering effects on real TC predictions will be discussed in Section 3.

Table 3. Comparisons of standard deviations (σ_{WNP}) of predictand (intensity change) using the mean value for the entire sample (i.e., non-clustering cases) with those (σ_c) using the mean of each cluster at each lead time. The reduction rates of σ_C relative to σ_{WNP} are indicated in parentheses.

Forecast time (h)		Cluster					
		C1	C2	C3	C4	C5	All TCs
24	σ_{WNP}	15.9	17.4	14.5	16.8	12.0	15.6
	σ_C	15.8	13.7	14.0	16.8	11.6	14.8
		(0.4%)	(21.0%)	(3.3%)	(0.1%)	(3.3%)	(5.3%)
48	σ_{WNP}	22.6	29.3	22.5	25.1	17.5	24.2
	σ_C	22.4	23.2	21.1	25.0	16.7	22.3
		(1.0%)	(20.8%)	(6.3%)	(0.4%)	(5.0%)	(7.9%)
72	σ_{WNP}	25.5	34.7	25.9	28.6	21.6	28.5
	σ_C	24.9	27.4	23.4	28.3	19.9	25.6
		(2.3%)	(21.1%)	(9.8%)	(1.2%)	(7.6%)	(10.5%)
96	σ_{WNP}	29.0	34.6	28.3	28.0	23.7	30.2
	σ_C	27.9	28.3	24.8	27.0	21.3	26.8
		(3.5%)	(18.2%)	(12.3%)	(3.6%)	(10.2%)	(11.5%)
120	σ_{WNP}	31.4	32.7	30.7	26.8	23.9	30.6
	σ_C	30.2	28.3	26.6	23.8	20.9	27.1
		(3.8%)	(13.2%)	(13.4%)	(11.0%)	(12.3%)	(11.3%)

2.5. Decision tree algorithm

The analysis of climate information on TCs has socio-economic implications and scientific significance because it leads to a better understanding of the typhoon activities and related mechanisms (Bengtsson et al., 1996; Emanuel, 2005; Webster et al., 2005). However the volume of the data related to the TCs is huge and varied, and it has increased significantly in recent years, which seems to outstrip the amount of data that can be handled by traditional analytical methods (Zhang et al., 2013). Data mining is the process of finding useful information form the huge and varied database, by automatically or semi-automatically analyzing data from various perspectives (Berry and Linoff, 1997).

Decision tree which one of the most frequently used algorithms, is known as data mining methods for finding rules, patterns, and knowledge from archived databases for decision-making (Quinlan, 1993). Decision tree is easier to understand and explain the results and can be used directly in decision making.

The Classification And Regression Tree (CART) is one of the analysis algorithms of the decision tree. It is easy to interpret the generated rules and has the advantage that both continuous and categorical variables can be used, and it provides a post-pruning strategy to prevent overfitting problem after the decision tree is developed. CART which measures impurity using Gini Index, is an algorithm based on binary split that has only two child nodes from the parent node. Gini index is described as below.

$$G = \sum_{j=1}^{c} \sum_{i \neq j} P(i)P(j) \qquad (5)$$

where *P(i)* is the probability that one entity in each node belongs to the i-th category of

the target variable, c is the number of categories of target variables, respectively. P (i) P (j) is the probability of misclassification that an entity extracted from the i-th category of the target variable belongs to the j-th category of the target variable. The Gini index, which is the sum of the probability of misclassification, is one of the measures of impurity or diversity in each node and can be expressed as follows.

$$\mathrm{G} = \sum_{j=1}^{c} P(j)(1 - P(j)) = 1 - \sum_{j=1}^{c} P(j)^2 = 1 - \sum_{j=1}^{c} (\frac{n_j}{n})^2 \qquad (6)$$

where n is the number of observations included in the node and n_i is the number of observations belonging to the i-th category of the target variable. The Gini index is a probability that two elements randomly extracted from n-elements belong to different groups, and it with two categories of target variables can be expressed as follows.

$$\mathrm{G} = 2\mathrm{P}(1)\mathrm{P}(2) = 2(\frac{n_1}{n})(\frac{n_2}{n}) \qquad (7)$$

The CART algorithm selects the best predictor for reducing the Gini index and the optimal separation of the variable as a child node. The reduction of the Gini index is calculated as follows.

$$\Delta\mathrm{G} = \mathrm{G} - \left(\frac{n_L}{n}\right) G_L - \left(\frac{n_R}{n}\right) G_R \qquad (8)$$

where n is the number of observations in the parent node, and n_R and n_L are the number of observations in the child node, respectively. The CART algorithm divides the node in order that the child node has the smallest degree of impurity, and repeats this process to construct a decision tree.

2.6. Resampling and oversampling technique

The *k*-fold cross validation is one of the most popular resampling techniques used to increase the statistical reliability of model performance measurements. The *k*-fold cross validation method is defined as follows. First, the entire sample is divided into k equal-sized subsamples. One sub-sample is used as the validation data for the test of the model, and the remaining *k*-1 sub-samples are used as the training data. This process is repeated *k*-times until all subsamples are used exactly once for the validation data. The results from each step of the process are averaged to form an evaluation index, which can be used to perform verification. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once. In this study, we use the 10-fold cross validation method.

A classification model was developed using an imbalanced database that is likely to be a biased model that yields results with a higher classification in the training data. In this study, the synthetic minority oversampling technique (SMOTE; Chawla, 2003), the most widely used resampling technique, was used to avoid this problem.

3. Statistical-dynamical typhoon intensity prediction scheme

3.1. Static and synoptic potential predictors

The present STIPS uses various potential static and synoptic predictors, which have been widely used in many statistical TC intensity prediction models (Elsberry et al., 1975; Jarvinen and Neumann, 1979). In this study, six potential static predictors are used for the model development (Table 4): initial maximum wind speed (iWIND), intensity changes within 12 hours (DVMX), longitude (LON) and latitude (LAT) of TC center, storm translation speed (MOV), and the ratio of land within the TC core (LAND). They are mostly obtained from the TC best track information. The potential synoptic predictors, or, environmental predictors, are estimated from the prediction results of the dynamical model. The present model uses 11 potential synoptic predictors (Table 5): sea surface temperatures (SST), maximum potential intensity (MPI), intensification potential (POT), ocean heat contents (OHC), relative humidity of upper and lower layer (RHHI, RHLO), vertical wind shears of upper and lower layers (SH200, SH500), 200-hPa air temperature (T200), 200-hPa zonal wind (U200), and 850-hPa relative vorticity (RV850). The value averaged within 200-km radius from the storm center is used to estimate the oceanic predictors such as OHC, SST, MPI, and POT. For atmospheric predictors such as RHHI, RHLO, SH200, SH500, T200, and U200, they are estimated between 200 and 800 km from the storm center. For R V850, it is averaged within 1000-km radius from the storm center.

iWIND is the most important predictor in the static predictor pool and has a strong negative correlation with the intensity change for all leads and all five clusters (Fig. 5a). DeMaria and Kaplan (1994a) noted that strong storms have less potential for further intensification because their intensities are already closer to MPI. DVMX represents the persistence of intensity change (Fig. 5b), which exhibits the highest correlation with intensity

change in 6-h forecast time for most clusters (i.e., C3 to C5), consistent with the results of Knaff et al. (2005). The relationship between TC locations (longitude and latitude) and intensity change is dependent on the track types (Figs. 5c and 5d). TCs that move eastward after recurving (e.g., C5 in Fig. 5c) tend to weaken due to low SST and high westerly wind shear encountered (Fig. 4). Specifically, the intensification rate of C1 and C4 which are affected by land as TCs move to the west shows a strong positive correlation with LON (Fig. 5c). For LAT, an overall negative correlation with intensity change is dominant (Fig. 5d) since TCs, on average, become weaker when they move northwards due to unfavorable environmental conditions to TC development at higher latitudes.

Table 4. List of six static potential predictors used in the present model.

Predictor	Description
iWIND	initial maximum wind speed (MWS) (kt)
DVMX	12-h change in intensity
LON	Longitude
LAT	Latitude
MOV	storm translational speed (m/s)
LAND	Ratio of landmass within 200 km from center

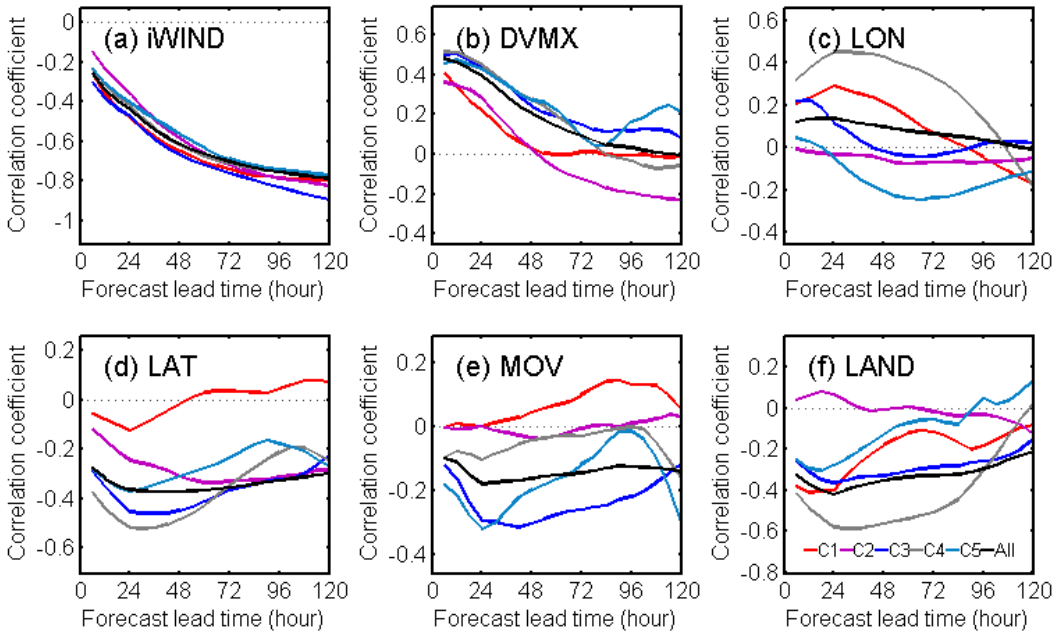

Fig. 5. The correlation coefficients between 6 static predictors (a, iWIND; b, DVMX; c, LON; d, LAT; e, MOV; f, LAND) and the change in TC intensity for five clusters and non-clustering case using all TCs at each forecast time.

MOV is known to be an important factor for TC-ocean interaction (Lin et al., 2008, 2009). Slow moving TCs tend to cause a large cold wake which could limit TC intensification. This relationship is only found in C1, where most TCs move slower and experience large cooling effect due to shallow-ocean mixed layer of the South China Sea. For the other clusters (particularly for C3 and C5), a negative correlation between MOV and the intensity change is shown clearly. As TCs encountered mid-latitude jet stream (C3 and C5), MOV tend to increases but the TC intensity weakens due to strong wind shear and land effects (Fig. 5e).

LAND is defined as the ratio of landmass within radius of 200 km from the TC center. When TCs approach land, relatively dry and cold air mass flows into the outer circulation of

the TCs and the friction over land increases, leading to TC weakening. Therefore, LAND and intensity changes are negatively correlated in most of the clusters except the non-landing clusters (C2 and C5 in Fig. 1.2f). The highest negative correlation is found in C4 in which most TCs with relatively strong intensity rapidly weaken during the landfall period over the mountainous Taiwan and the huge landmass of the eastern China (Fig. 5f).

Table 5. List of eleven synoptic potential predictors used in the present model.

Predictor	Description
SST	Area-averaged (0 km to 200 km) sea surface temperature
MPI	Area-averaged (0 km to 200 km) maximum potential intensity based on empirical equation
POT	MPI minus iWIND
OHC	Area-averaged (0 km to 200 km) ocean heat contents
RHHI	Area-averaged (200 km to 800 km) relative humidity 500 – 300 hPa
RHLO	Area-averaged (200 km to 800 km) relative humidity 850 – 700 hPa
SH200	Area-averaged (200 km to 800km) 200 hPa to 850 hPa vertical wind shear
SH500	Area-averaged (200 km to 800 km) 500 hPa to 850 hPa vertical wind shear
T200	Area-averaged (200 km to 800 km) air temperature at 200 hPa
U200	Area-averaged (200 km to 800 km) zonal wind at 200 hPa
RV850	Area-averaged (0 km to 1000 km) relative vorticity at 850 hPa

Now let's turn our attention to the potential synoptic predictors (Table 5). MPI is the upper bound of TC intensity given the atmospheric vertical profile and pre-cyclone SST (Emanuel, 1988). The MPI is empirically estimated according to DeMaria and Kaplan (1994a): An exponential function is derived by fitting the scatter plot between SST and TC maximum intensity. Here, the SST is extracted from the HyCOM-NCODA data for the period of 2004 - 2014 and averaged within a 200-km radius from the center of each TC. The final regression coefficients (Equation 9 below) are determined using only the upper 95th percentile values of SST at each 0.5 ℃ temperature interval.

$$\mathrm{MPI} = \mathrm{A} + \mathrm{B}e^{\mathrm{C}(\mathrm{SST}-\mathrm{T}_0)}, \qquad (9)$$

where A = 39.91 knots; B= 96.0 knots; C=0.1837 $℃^{-1}$; and T_0 = 30.0 °C. In Fig. 6b, MPI is highly correlated with TC intensity change for C3 as the MPI gradient is large along the TC tracks (Fig. 4c), but weakly correlated for C1 and C4 the influence of land (Fig. 6b). This pattern in MPI is very similar to that of SST (Fig. 6a).

POT is the TC potential future intensity changes (DeMaria and Kaplan, 1994b), defined as the difference between MPI and maximum wind at the initial time.

$$\mathrm{POT} = \mathrm{MPI} - \mathrm{iWIND} \qquad (10)$$

It is reported that POT is the most important predictor in the SHIPS (DeMaria and Kaplan, 1994a and DeMaria and Kaplan, 1999). Indeed, our analysis also shows that POT has the highest correlations with predictand for lead times after 24 hours and this high correlation is maintained for all five clusters (Fig. 6c).

OHC is an oceanic predictor representing the upper ocean thermal structure, expressed as an integral of calories of seawater with temperature above 26 ℃ (Leipper and Volgenau,

1972)

$$\mathrm{OHC} = C_p \rho \int_0^{d26} (T - 26) dz \quad , \quad (11)$$

where C_p is the sea water specific heat capacity; ρ is the water density; *d26* is the 26 °C isothermal depth; and T is the ocean temperature at a specific depth. Lin et al. (2009) suggested that warm eddies and ocean currents could make a critical contribution to rapid intensification of TCs. These warm features are characterized by high OHC and can effectively mitigate TC-induced negative feedback due to sea surface cooling. Analysis reveals that OHC is highly correlated with predictand in most clusters except C1 where the mixed layer is very shallow (Fig. 6d).

Relative humidity is known to affect convective buoyancy, which is a direct source of TC energy (Wu et al., 2012; Bogner et al., 2000; Knaff et al., 2005). In this study, the relative humidity was divided into RHLO and RHHI based on Knaff et al. (2005). RHLO (RHHI) is calculated in atmospheric layer from 850 hPa to 700 hPa (500 hPa to 300 hPa), within an annulus of 200-800 km radius (donut shape) from the center of the TC. Knaff et al (2005) reported that RHHI is statistically important at all forecast lead times, but RHLO is not. However, our results for C2 reveal that both RHHI and RHLO have a moderate positive correlation with intensity change (Figs. 6e and f). This result is an example of how a previously unrecognized predictor could be potentially useful for a particular cluster.

Many studies have shown that VWS is an important predictor for TC intensity change (Jone, 2000; DeMaria, 1996; Wang and Holland, 1996; Frank and Ritchie, 2001; Corbosiero and Molinari, 2002). The VWS is defined by the magnitude of the vector difference between the two different layers, 200–850 hPa (SH200) and 500–850 hPa (SH500):

$$\mathrm{SH200} = \sqrt{(U_{200} - U_{850})^2 + (V_{200} - V_{850})^2} \quad (12)$$

$$\mathrm{SH500} = \sqrt{(U_{500} - U_{850})^2 + (V_{500} - V_{850})^2}, \qquad (13)$$

where $U_{200}(V_{200})$ is 200-hPa zonal (meridional) wind; $U_{500}(V_{500})$ is the 500-hPa zonal (meridional) wind; and $U_{850}(V_{850})$ is the 850-hPa zonal (meridional) wind. Strong VWS disrupts the organized deep convection, which inhibits TC intensification. The wind-related predictors such as VWS (Figs. 6g and h), RV850 (Fig. 6k), and U200 (Fig. 6j) are highly correlated with the intensity change for C3 in which many TCs pass through the subtropical region where the VWS is strong (Fig. 4a). In contrast, a low correlation is found in C1 because the mean motion of TCs moves straight towards the northwest, where there is almost no vertical shear gradient.

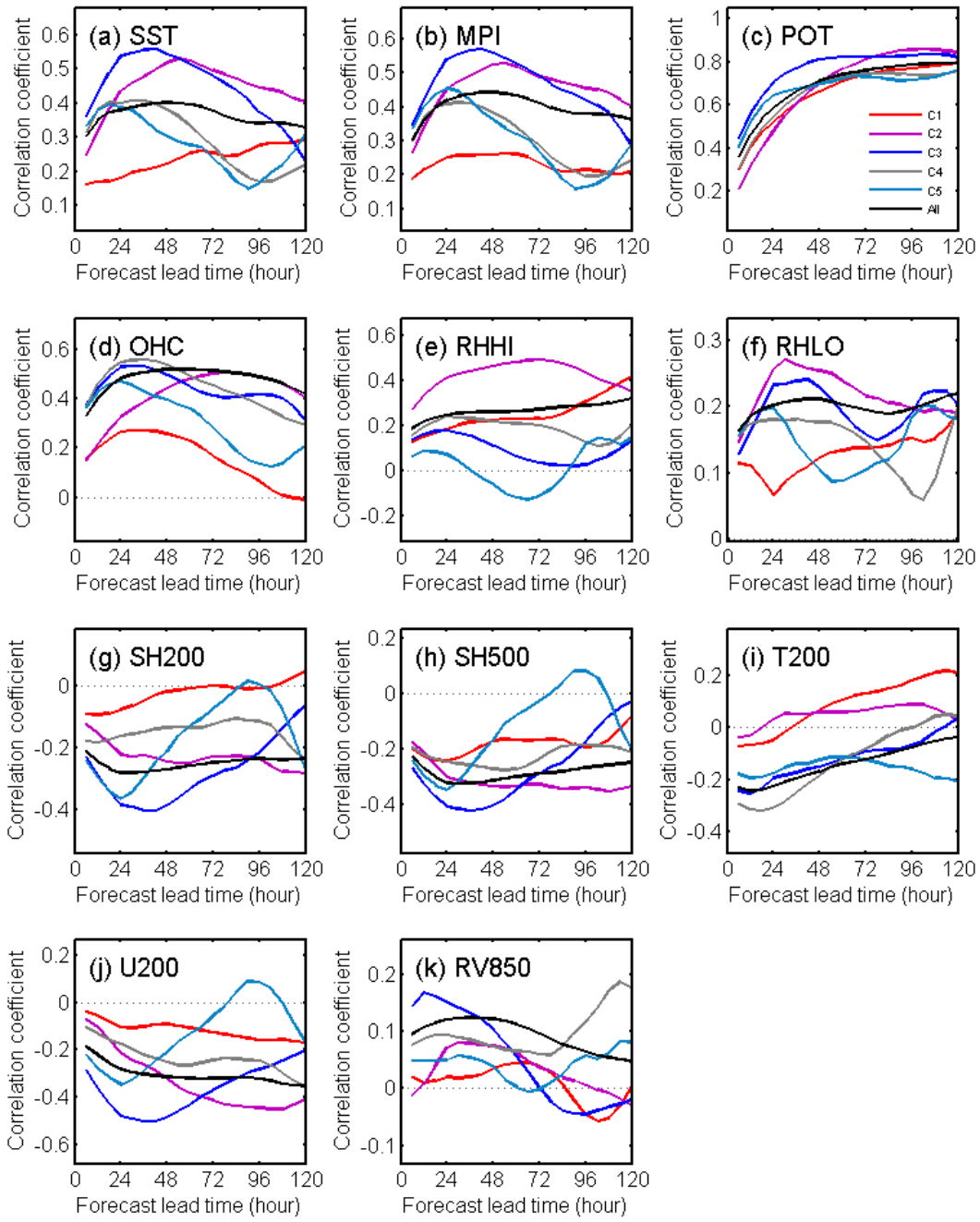

Fig. 6. Same as in Fig 5, but for 11 synoptic predictors (a, SST; b, MPI; c, POT; d, OHC; e, RHHI; f, RHLO; g, SH200; h, SH500; i, T200; j, U200; k, RV850).

3.2. DAT-based potential predictors

TC-induced vertical mixing is affected by various factors such as TC intensity, storm translation speed, Coriolis parameter, and upper ocean structure. The depth of the vertical mixing is known to be typically 60 to 100 m in major tropical cyclones (Price, 2009). In this study, when considering the effect of the TC-induced vertical mixing on the intensity predictions, new MPIs and POTs are additionally calculated using DATs based on various mixing depths from surface to 120 m (10-m interval) instead of using the SST (Table 6).

As the storm intensity increases, DAT using deeper mixing depths is more correlated with TC intensity changes (Lin et al., 2008). To investigate the relation between TC intensity and mixing depth, we classified TCs into three groups according to iWIND (less than 50 knots, 50–80 knots, and over 80 knots) and examined the variations of correlation coefficients between thermodynamic predictors and 24-hours intensity changes as a function of different mixing depths (Fig. 7). It is shown that initial TCs that are stronger generally tend to have higher correlations with the three thermodynamic predictors for deeper mixing depths.

Table 6. List of three DAT-based potential predictors used in the model.

Predictor	Description
DAT10–DAT120	Ocean temperatures averaged from the surface down to various depth (10-120 m, 10-m interval)
MPI10–MPI120	Maximum potential intensity using DAT10–DAT120
POT10–POT120	POT using MPI10–MPI120

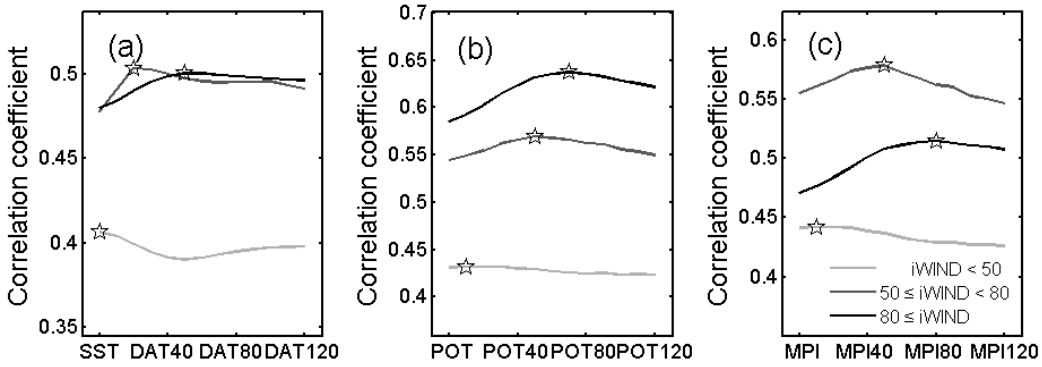

Fig. 7. The comparison of the correlation coefficients between three thermodynamical predictors (a, DAT; b, POT; c, MPI) and the 24-h changes in TC intensity for three groups classified by initial maximum wind speed (iWIND). Open pentagrams represent the location of maximum value for the three groups.

It is interesting to ask whether this feature is also shown when TC tracks are clustered. Figure 8 present the correlation coefficients between the DAT-based predictors and TC intensity change for five track types and non-clustering cases at various lead times. C1 and C5, with relatively weak intensity (see Table 1), have the highest correlation coefficients up to 72 hours leads when mixing depths are shallow. C4 with strong intensity has the highest correlations when mixing depths are deep (Fig. 8). For C2, this relationship is not clear since the differences of the DATs for varying mixing depths are small in the tropics. The result suggests that the inclusion of the DAT-based predictors combined with cluster analysis might further improve the prediction skill of STIPS because the predictors consider realistic interactions of the TC with the ocean that vary depending on clusters.

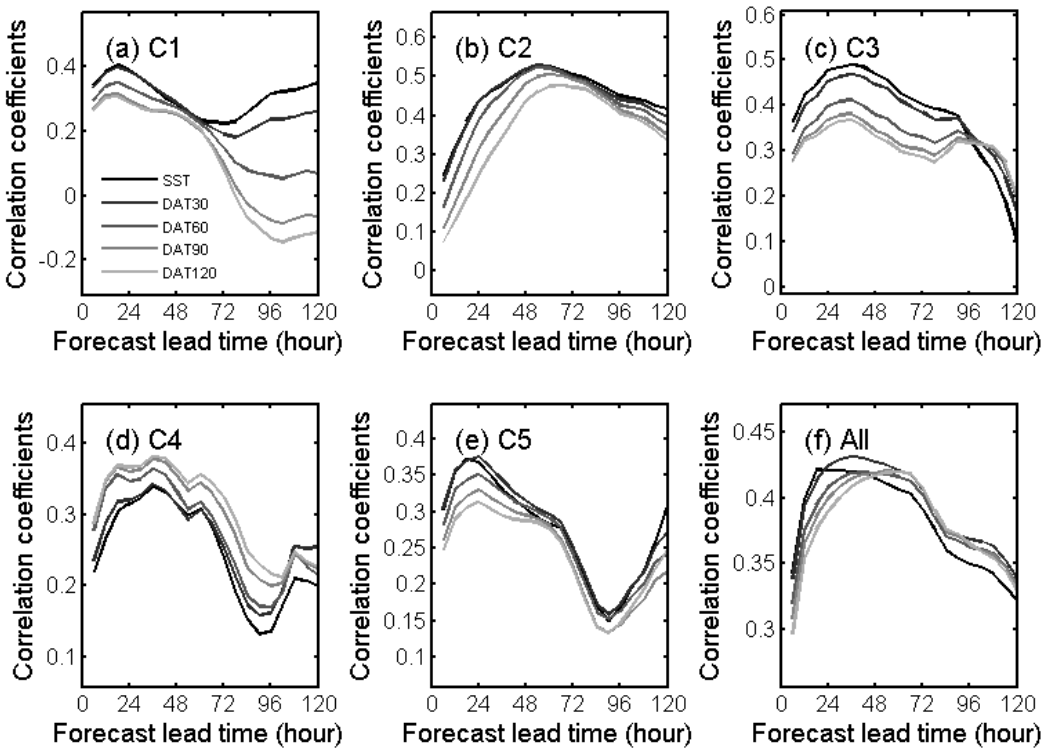

Fig. 8. Correlation coefficients between the various depth-averaged temperatures (SST, DAT30, DAT60, DAT90, and DAT120) and TC intensity change for five clusters (a, C1; b, C2; c, C3; d, C4; e, C5) , non-clustering case using all TCs (f) at each lead time.

3.3 Effects of using clustering and DAT-based predictors

To investigate further the effect of using clustering and DAT-based predictors on statistical-dynamical TC intensity predictions, four sets of experiments were conducted for the training period (Table 7): a run without the use of both clustering and DAT-based predictors (hereafter referred to as the STIPS-SST), a run using clustering only (CSTIPS-SST), a run using DAT-based predictors only (STIPS-DAT) and a run using both clustering and DAT-based predictors (CSTIPS-DAT). We first examined the predictors selected for each experiment (Fig. 9). For non-clustered models, multiple regression equations are developed for each prediction time, yielding a total of twenty equations (120-h prediction, 6-h inter val). However, clustered models have one-hundred equations because five clusters are also considered. Given the large number of equations, we present the ratio of selected final predictors for each model. Here, it should be noted that the present multiple regression model allows the combination of up to five predictors, but only one or two predictors were finally selected due to multi-collinearity issues. Figure 9 shows the ratio of the final selected predictors for each set of experiments. The first predictor selected from the screening pr ocedure for each prediction time is termed "predictor 1", and the second one is termed "predictor 2".

Table 7. Experimental designs for investigating the effect of using clustering and DAT-based predictors. The open circle and the cross indicate the method applied and not applied, respectively.

Model	Clustering	DAT-based predictor
STIPS-SST	×	×
STIPS-DAT	×	○
CSTIPS-SST	○	×
CSTIPS-DAT	○	○

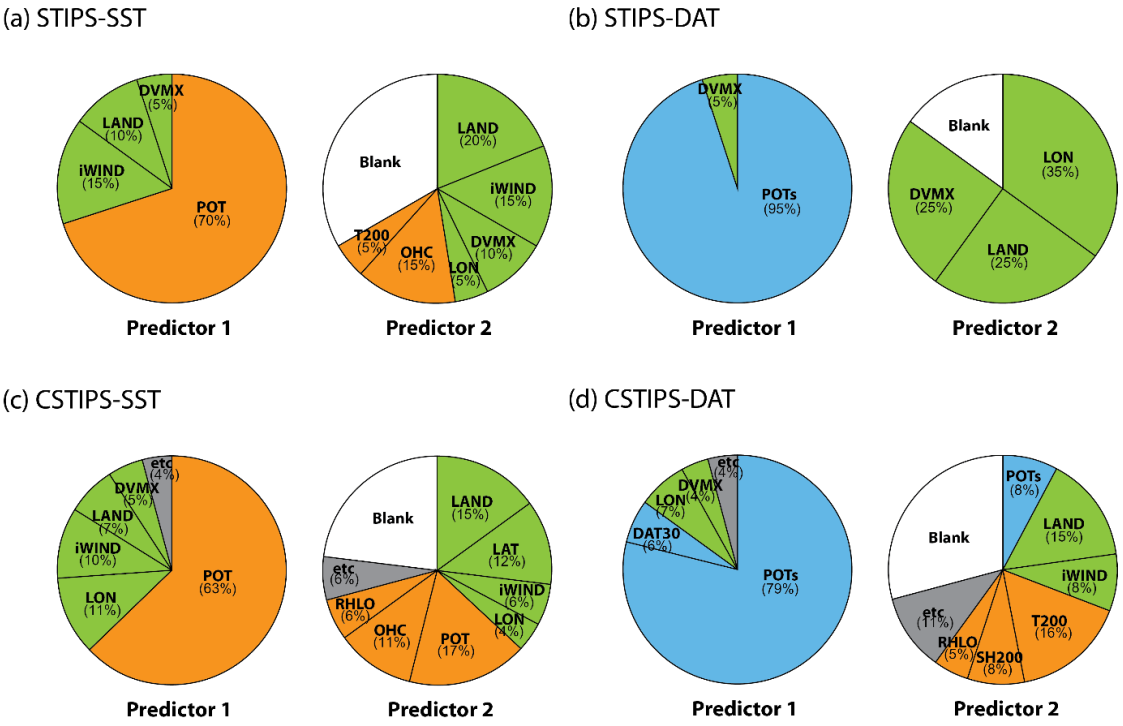

Fig. 9. Pie charts representing the ratio of the final selected predictors for (a) STIPS-SST, (b) STIPS-DAT, (c) CSTIPS-SST, and (d) CSTIPS-DAT. Predictor 1 represents the first selected predictor with the highest correlation coefficients with predictand (intensity change). Predictor 2 represents the second selected predictor in the regression equation. Colors indicate the type of predictor (i.e., static is green, synoptic is orange, DAT-based is blue, and others are gray). If only one predictor is selected, it is marked as blank in Predictor 2.

The most frequently selected predictor was POT (including DAT-based POT) for all experiments. Particularly, for STIPS-DAT, POT was selected 95% of the time as the first predictor 1. iWIND, SST, MPI, and OHC, which had high correlations with intensity change (Fig. 5a), were much less selected compared to POT for most experiments. This is because, when POT was selected, the aforementioned four other predictors were excluded due to their high correlations with POT. DVMX had a particularly good performance in the short-term prediction (Fig. 5b) and this intensity-change predictor is commonly selected as the fir st and second predictors for two non-clustering experiments (Fig. 9a and 9b). LAND

was also one of the main predictors selected in all four experiments. In addition, LON, OHC, RHLO, and T200 are important second predictors. The number of selected predictors for STIPS-SST and CSTIPS-DAT was similar to each other (Fig. 9a and 9d).

The types of selected predictors for each cluster reflect unique characteristics of the clusters such as the tendency of mean TC intensity (or the changes), landfalls, upper-ocean thermal structure, and TC-induced mixing depth. For example, for C1 POTs using relatively shallow DAT (POT, POT10, POT20, POT30, and POT50 with depth ranging from 0 to 50 m) and LAND were chosen as major predictors (Fig. 10a). This is due to the fact that many cases of C1 experience a rapid weakening due to landfall and inadequate ocean thermal conditions with a thin mixed layer, leading to a shallow vertical mixing caused by TCs. In C2, a large percentage of the selected predictors for predictor 1 were the single POT-type predictor (95%). This arises as most TCs in C2 spend their lifetime in the tropical open ocean where oceanic thermal conditions are the most important factors. In this case, land effect and static predictors are minor. C3, characterized by strong weakening tendencies with leads (Table 2), showed that T200 was the most dominant in predictor 2. This is possible because many TCs in C3 travel to the north and are influenced by a large gradient of T200 along the track. Many TCs in C4 make landfall over the East Asian region (The eastern coast of China and Taiwan), and LON and LAND were selected as major predictors. Lastly, for C5, in which most TCs typically decayed over the cold open ocean, POTs with DATs of relatively shallow to medium depth (20 to 80 m) were important predictors. Since TCs in C5 pass through the open ocean where a large gradient of wind shear along the mean track of TC exists (Fig. 4a), vertical wind shear (SH200 & SH500) was also selected. The performance of four models will be presented in the following section.

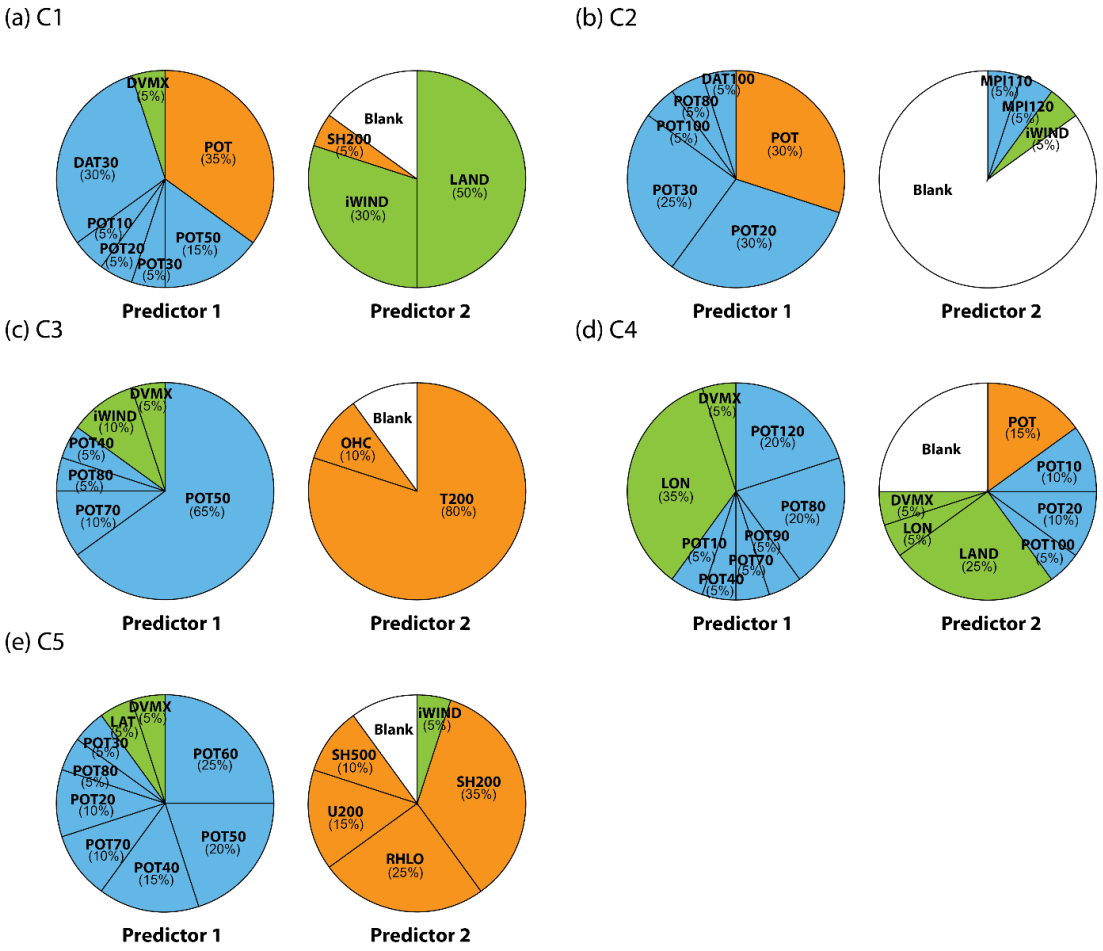

Fig. 10. Same as in Fig. 9, but for five clusters of CSTIPS-DAT (a, C1; b, C2; c, C3; d, C4; e, C5).

3. 4 Comparisons of model performance

In statistical models, the relationship between the predictand and predictors is first established during the training period. This relationship is then applied to an independent dataset. This relationship will then be used to determine the performance of real-time predictions, which is independent of the testing sample. Figures 11 and 12 compare the prediction skills among CSTIPS-DAT, CSTIPS-SST, STIPS-DAT, and STIPS-SST for the training period. The forecast errors between the CSTIPS-SST and CSTIPS-DAT at vari ous leads are small for all clusters (Figs. 11a-c). The comparisons reveal that CSTIPS-DAT, which used both clustering and DAT-based predictors, generally outperformed the other three models. A close examination shows that the most significant improvement was made by applying clustering approach (CSTIPS-SST vs. STIPS-SST, and CSTIPS-DAT vs. STIPS-DAT in Fig. 11f). At 72-h lead time, the improvement is most remarkable; the mean absolute errors (MAEs) are reduced by 3.6 knot (25% improvement) and 1.9 knot (15%) for CSTIPS-SST and CSTIPS-DAT, respectively, compared to non-clustering settings. A similar tendency is also found in the results of R-square (Fig. 12f). The overall improvement in the CSTIP-DAT model is due to the enhanced improvement in some individual clusters. In particular, C5 showed the best prediction skills among the five clusters, most significantly contributing to the model improvement (Fig. 12e). We found that the performance of each cluster is related to the magnitude of σ_C (Table 3). That is, smaller σ_C leads to more skillful predictions. This reaffirms that the clustering approach reduces the variability of predictand, which ultimately results in the improvement of TC intensity prediction.

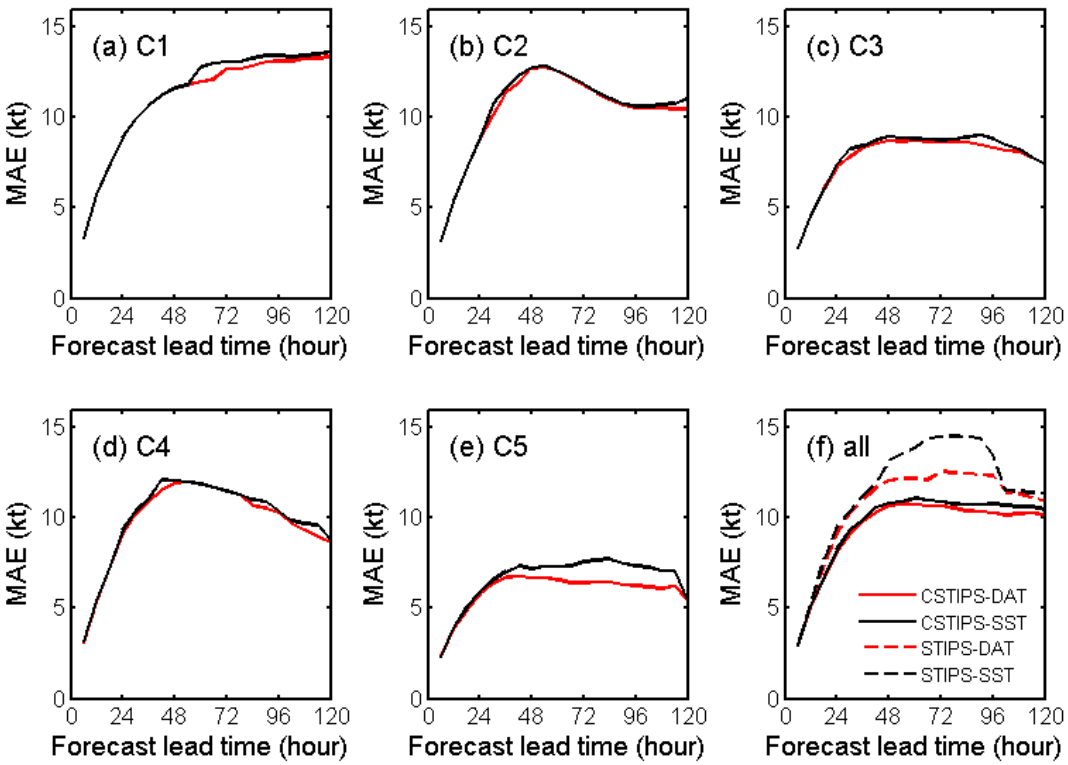

Fig. 11. Comparisons of mean absolute errors (MAEs) of the maximum intensity (kt) for five clusters (a, C1; b, C2; c, C3; d, C4; e, C5) and non-clustering cases using all TCs (f) at each lead time during the training period (2004-2012). For five clusters (a–e), two results from CSTIPS-DAT and CSTIPS-SST are compared. For non-clustering case (f), the results from all experiments in Table 1.6 are compared.

The impacts of DAT-based predictors on TC intensity prediction can be realized by comparing STIPS-DAT with STIPS-SST, and CSTIPS-DAT with CSTIPS-SST (Fig. 11f and 12f). The effect of DAT-based predictors is not large as that of clustering, but they still further improved the performance. The largest improvement (14%) in MAE was found in the comparison with non-clustering experiments (STIPS-DAT and STIPS-SST in Fig. 11f) at lead times between 48 h and 96 h.

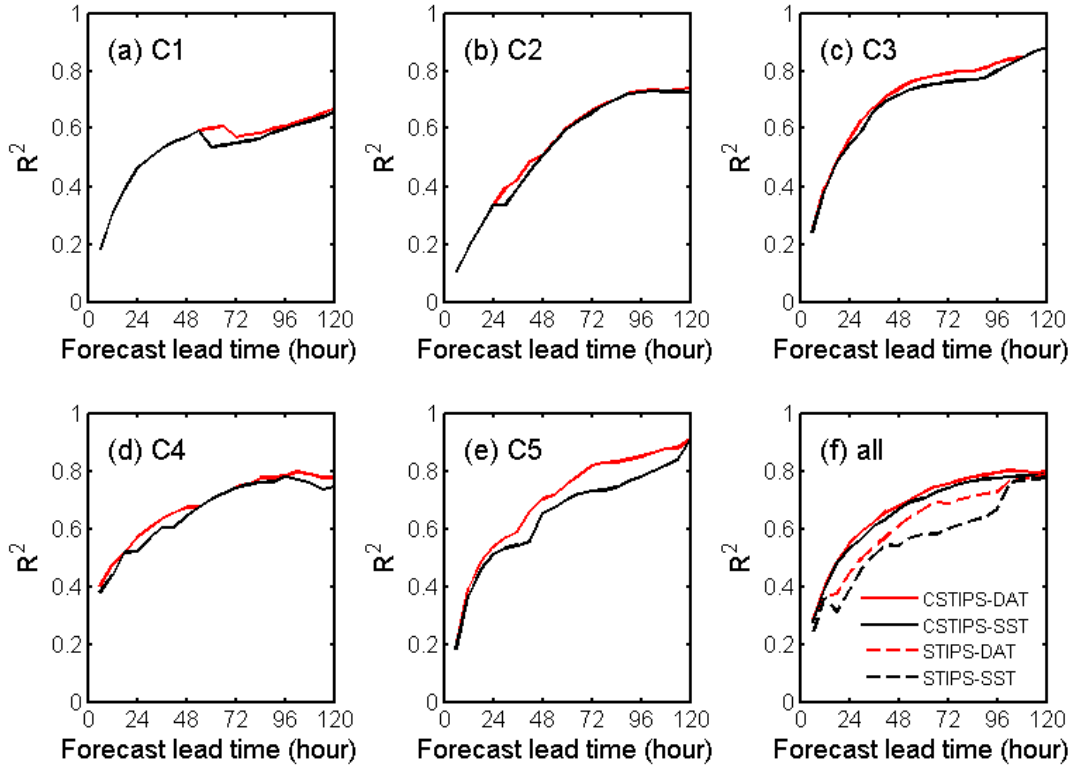

Fig. 12. Same as in Fig 11, but for the R-squared (variance explained).

The model developed during the training period is tested with real-time predictions for two years of 2013 and 2014. Figure 13 illustrates the schematics of the real-time prediction systems. The systems were implemented through the following procedure: first, obtaining the 5-day track forecast information and atmospheric and oceanic conditions at each forecast lead time from KMA; second, calculating various static, synoptic and DAT-based predictors and membership coefficients for each TC; third, determining belong to one of the five clusters with the membership degrees; and fourth, using the statistical model to produce the 5-day intensity prediction.

The next question is how well the models developed in this study perform for an

independent data set? A comparison of MAE, bias, and R-square among the four models shows that the CSTIPS-DAT model outperformed the others at most lead times (Fig. 14) and the contribution of clustering to the improved performance was the largest, similar to the results of the model training. However, the overall mean MAEs for all real-time predictions (Fig. 14a and 14b) were a little larger (about 1.5 knots) than those for the training periods, possibly due to the inaccuracy in real-time track predictions (Knaff et al., 2005) and the fact that the model parameters were fit for the training period, not during the independent period.

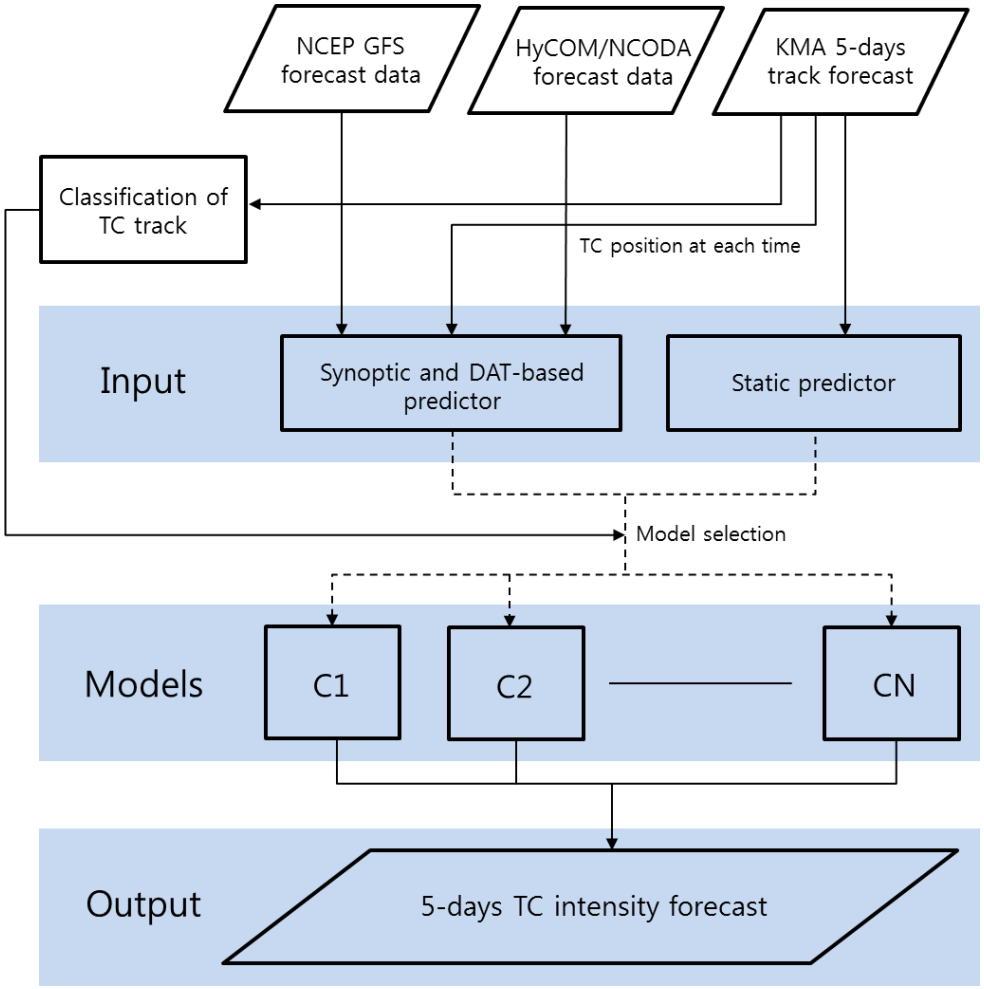

Fig. 13. Schematic of real-time prediction procedure using the CSTIPS-DAT.

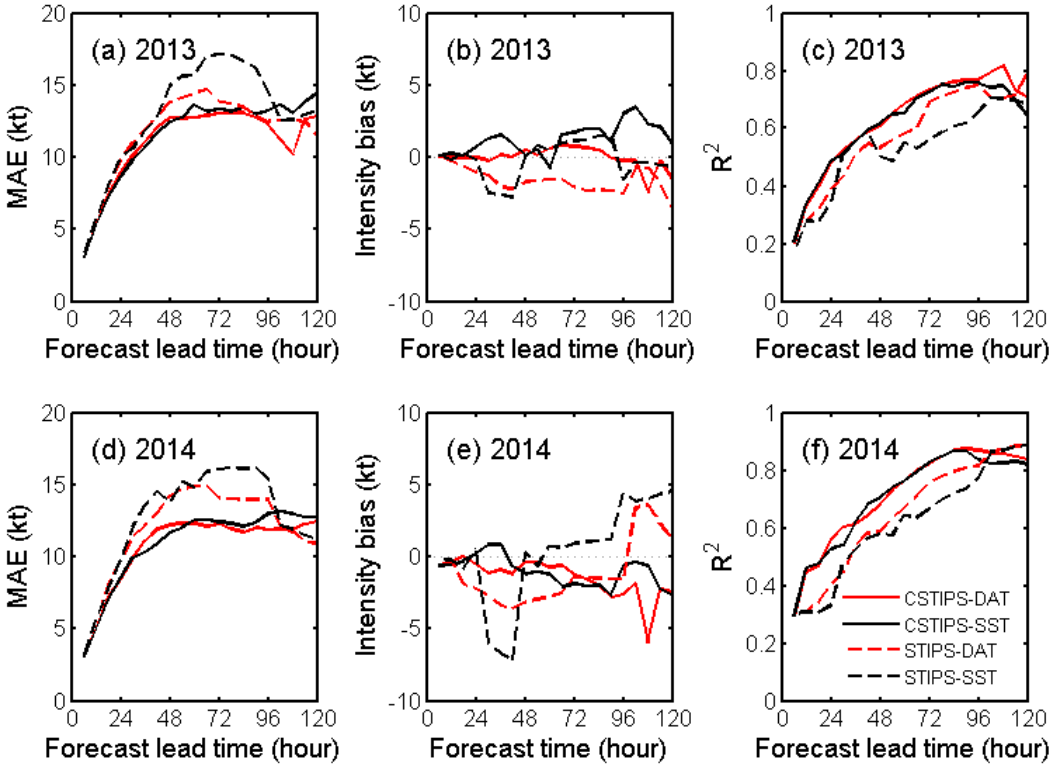

Fig. 14. Comparisons of MAEs (a, b), Bias (b, e), and R-square (c, f) for the real-time TC intensity prediction in 2013 (a–c) and 2014 (d–e). Here, the results from all experiments in Table 6 are compared.

To quantify the effects of the use of clustering and DAT-based predictors, the reduction rates in MAE (i.e., RE) were computed for STIPS-DAT, CSTIPS-SST, and CSTIPS-DAT relative to the control experiment (STIPS-SST):

$$\mathrm{RE} = \frac{100\,(E - E_{STIPS-sst})}{E_{STIPS-sst}}, \qquad (14)$$

where E is the MAE for each model, and $E_{STIPS-sst}$ is the MAE of STIPS-SST. Results indicate that the use of clustering was the main contributor to the reduction in MAE (red and black solid lines in Fig. 15). For clustered models, large reductions about 20-25% were noted

between 48-h and 96-h lead time for both years. However, the improvement disappeared after 96-h lead time, particularly for 2014. This seems to be related to the lack of statistical confidence due to insufficient sample size after 96 h.

To examine the results of individual prediction in each cluster, we selected six typhoons of Pewa (2013), Fitow (2013), Pabuk (2013), Faxai (2014), Kalmaegi (2014) and Phanfone (2014) and analyzed their tracks (Fig. 16a), membership coefficients (Fig. 16b), and intensity prediction results of the CSTIPS-DAT model (Fig. 17). Here, the membership coefficient is an index indicating the similarity of the individual TC track (colored lines in Fig. 16a) to the mean track of the membership cluster (black lines in Fig. 16a), i.e. a larger membership coefficient implies a higher similarity of a particular TC track to its respective mean tr ack. The cluster number assigned at each prediction time (6-hr interval) is given at the bottom of Fig. 17. For example, Typhoon Phanfone started as C2 in the beginning, but its membership gradually changed to C3 with its northward movement (Fig. 17f). The comparison of predicted intensities (thin lines in Fig. 17) with the RSMC best-track data (thick lines) presents an overall good performance for Fitow, Pabuk, Kalmaegi and Phanfone. For Typhoons Pewa and Faxai, however, the predicted intensities were largely overestimated. We found that the inaccurate predictions of these two typhoons are associated with their low membership coefficients (see Fig. 16b). In fact, Pewa and Faxai belonged to C5 and C2, respectively, but their tracks were far from the mean track of its respective cluster (Fig. 16a). The result suggests that an increase of number of clusters might allow more accurate predictions if enough samples were available. It also reconfirms the potential of cluster-analysis of track patterns to improve statistical TC intensity prediction skills.

It is also interesting to compare the CSTIPS-DAT model with the latest operational dynamical models such as Hurricane Weather Research and Forecast (HWRF), Regional

Unified Model (UMR) of the Korean Meteorological Agency, Japan Meteorological Agency Global Spectral Model (JGSM), and the U.S. Global Forecast System (GFS). As shown in Fig. 18, for lead times up to 24 hours, the CSTIPS-DAT model shows smallest MAEs with prediction errors less than 12 knots relative to operational dynamical models for both 2013 and 2014. After 24-h leads, the CSTIPS-DAT model shows a remarkably good skill and is comparable with GFS, which has the lowest MAE among the other three operational models.

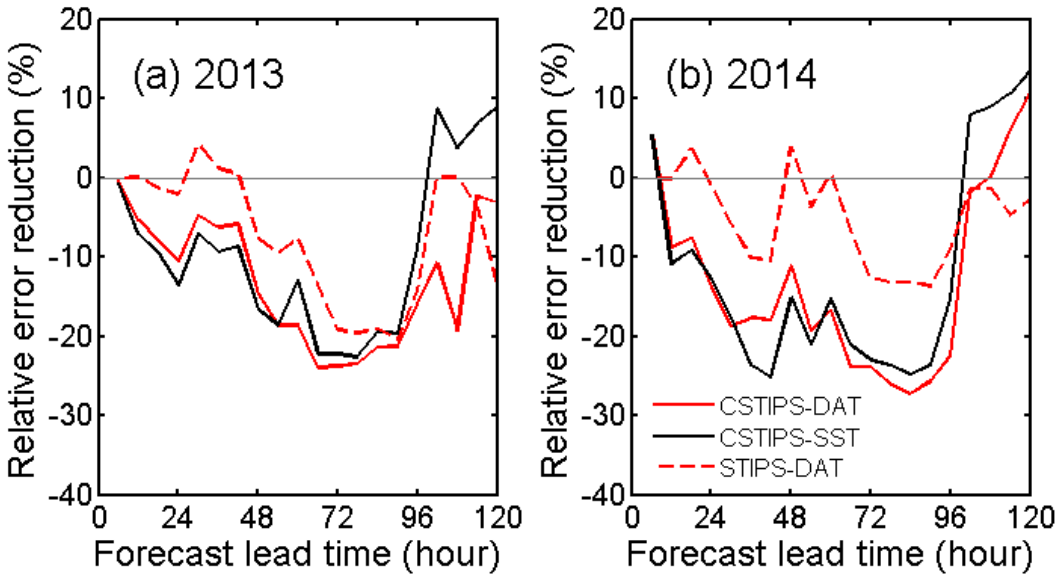

Fig. 15. The reduction rates in MAEs of STIPS-DAT, CSTIPS-SST, and CSTIPS-DAT relative to the control experiment (STIPS-SST) for 2013 (a) and 2014 (b).

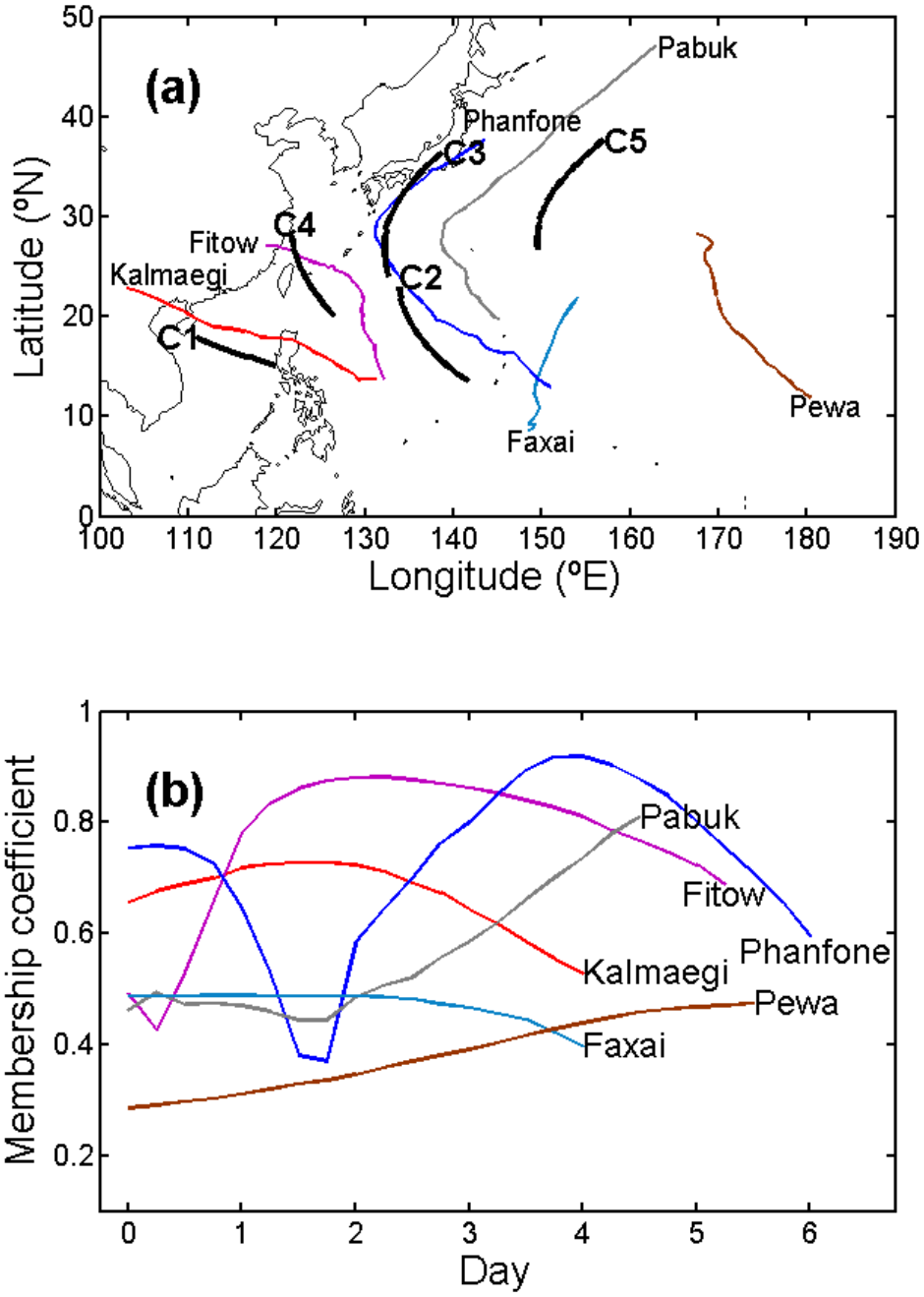

Fig. 16. (a) Tracks (colored lines) of Typhoon Pewa (2013), Fitow (2013), Pabuk (2013), Faxai (2014), Kalmaegi (2014) and Phanfone (2014) along with the mean track for five clusters (thick black lines). (b) Temporal evolution of membership coefficient for six TCs.

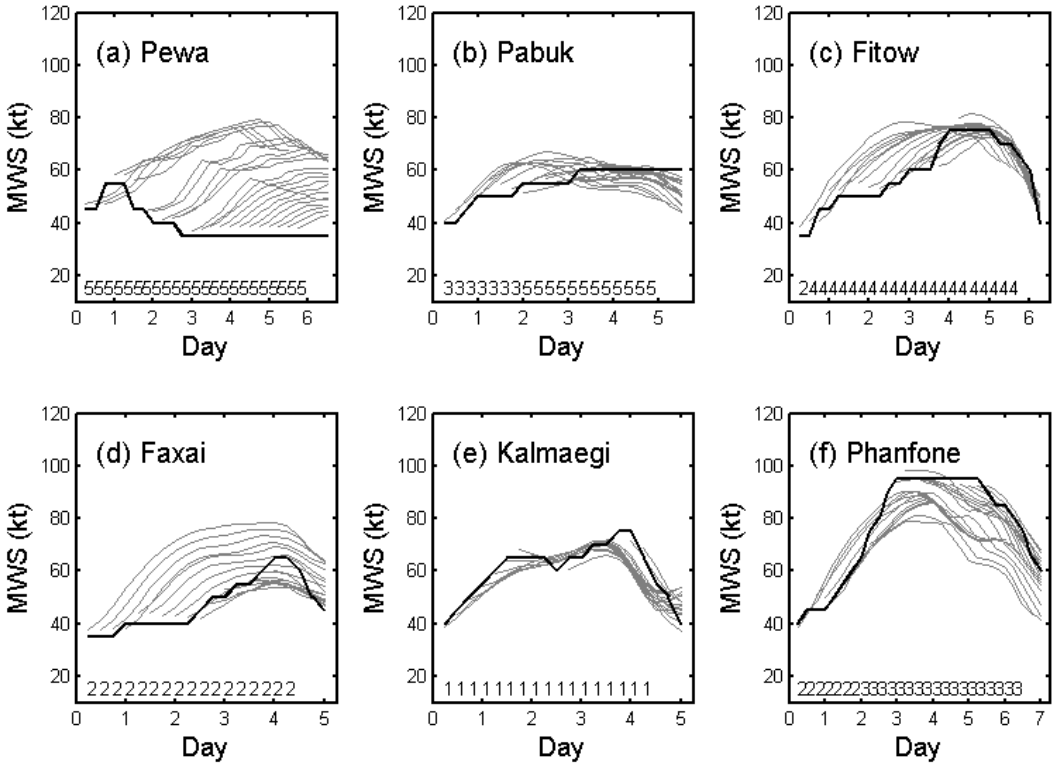

Fig. 17. Results of individual intensity predictions from CSTIPS-DAT for Typhoon (a) Pewa (2013), (b) Fitow (2013), (c) Pabuk (2013), (d) Faxai (2014), (e) Kalmaegi (2014) and (f) Phanfone (2014). Thick line is observation (RSMC best track data) and gray lines are individual CSTIPS-DAT predictions. The numbers above x axis denote the assigned cluster number.

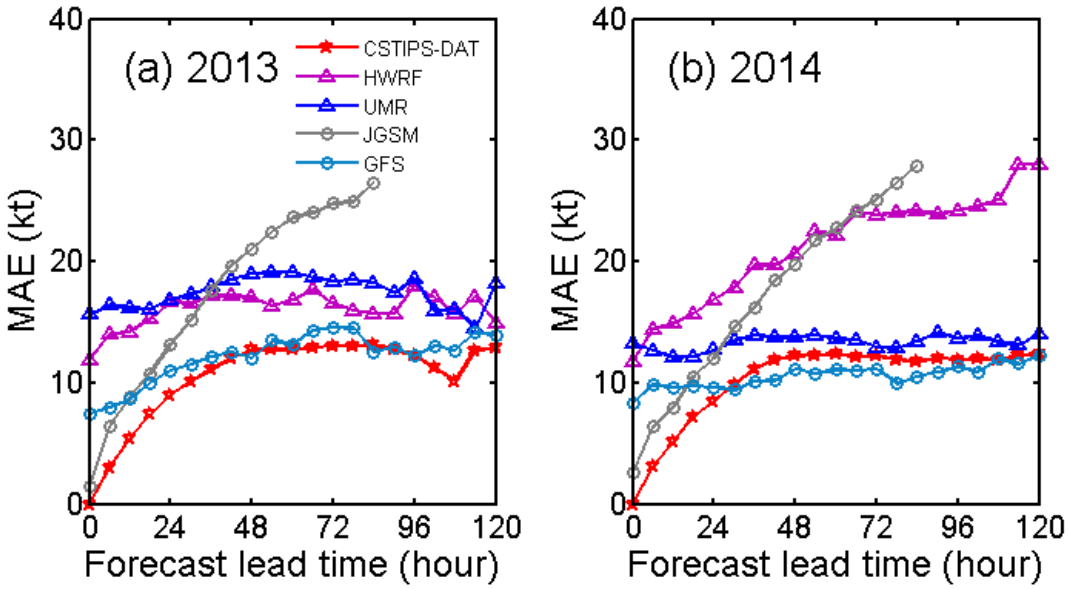

Fig. 18. Comparisons of MAE between results of various operational dynamical models (HWRF, UMR, JGSM and GFS) and the present model (CSTIPS-DAT) for (a) 2013 and (b) 2014.

4. Classification model for determining type of LMI

4.1 Prediction error of WNCP

LMI is an integrated statistic of TC intensification and represents the basic properties of TC climatology among the various criteria of measuring TC intensity (Emanuel, 2000; Kossin et al., 2014;Park et al., 2014; Moon et al., 2015). The distribution of LMI for the WNP in the period of 2004 - 2016 has two local maxima around 50 kt and 100 kt, and a local minimum at 70kt (Fig. 19a). This bimodal distribution in the probability density function (PDF) for LMI also found in the WNCP (Fig. 19b), but the second peak in PDF is much larger than the first peak compared to the WNP. In this study, TCs in the WNCP (*i.e.*, cluster C2) were classified into two groups for training of model based on the local minimum of LMI distribution: TCs with the LMI of above 70 kt (A70) and TCs below 70kt (B70).

A comparison of the MAE and bias of intensity prediction between A70 and B70 reveals that B70 has a larger MAE than A70 after 72-h lead times (Fig. 20a), in which the pattern of B70 is very similar to that of bias (Fig. 20b, blue). This implies that the bias correction of B70, *i.e.*, the correction of overestimation, can lead to the reduction of MAE. It is also important that the tendency of underestimation (*i.e.*, negative bias) for A70 can be corrected by separating TCs into two groups and establishing independent statistical model for each group since the bias of A70 is caused by the shift of the mean values to weak LMI due to the inclusion of overestimating B70.

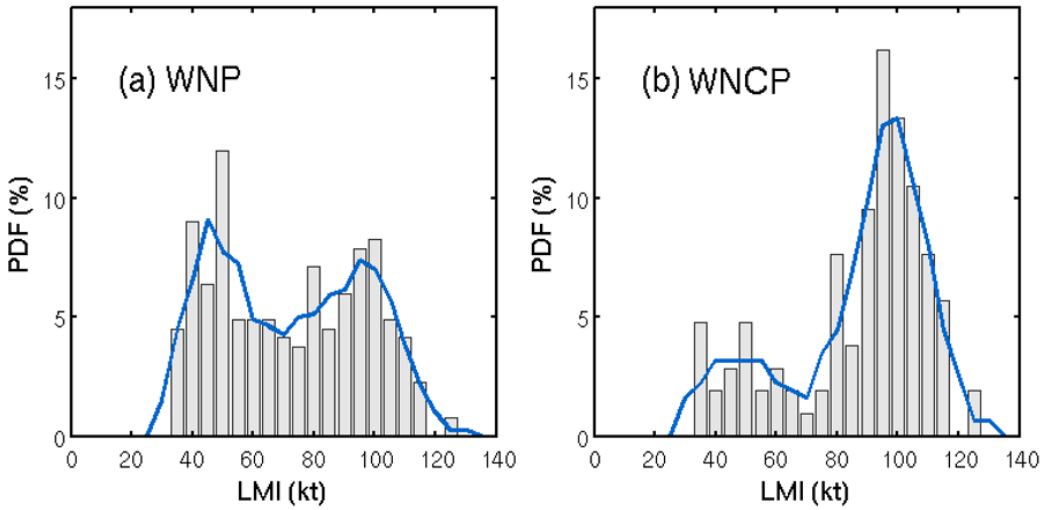

Fig. 19. Distributions of life time maximum intensity. PDFs are calculated using 2004-2016 TCs which belong to (a) Western North Pacific and (b) Western North Central Pacific. The blue line indicates the smoothed PDF.

The unique characteristics of the two groups were also found in the distribution of intensity change (Fig 21). The mean change in intensity within 48-hours was 20.7±24.9 kt for A70 (representing strongly-developing TCs) and 0.7±13 kt for B70 (non-developing TCs), in which the difference between the two was statistically significant. Therefore, if TCs in the WNCP are classified into two groups according to their LMI and the prediction model is separately developed for the classified group, it is expected that the variance of the intensity change can be reduced, leading to improving the intensity prediction skill of TCs.

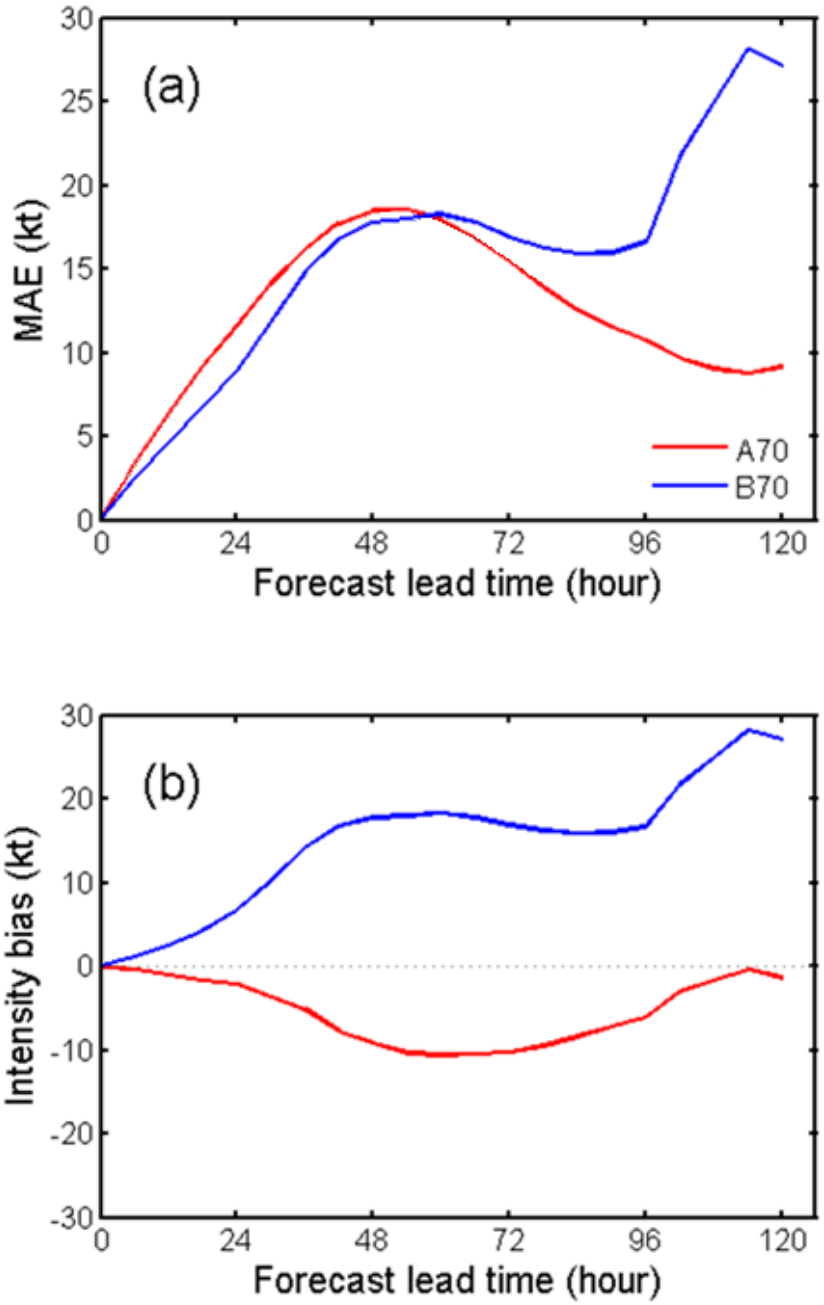

Fig. 20. Comparison of (a) mean absolute errors (MAEs) of maximum intensity and (b) intensity bias for classified two groups (red: A70, blue: B70) and all TCs (black) in Western North Central Pacific at each lead time during the 2013-2014.

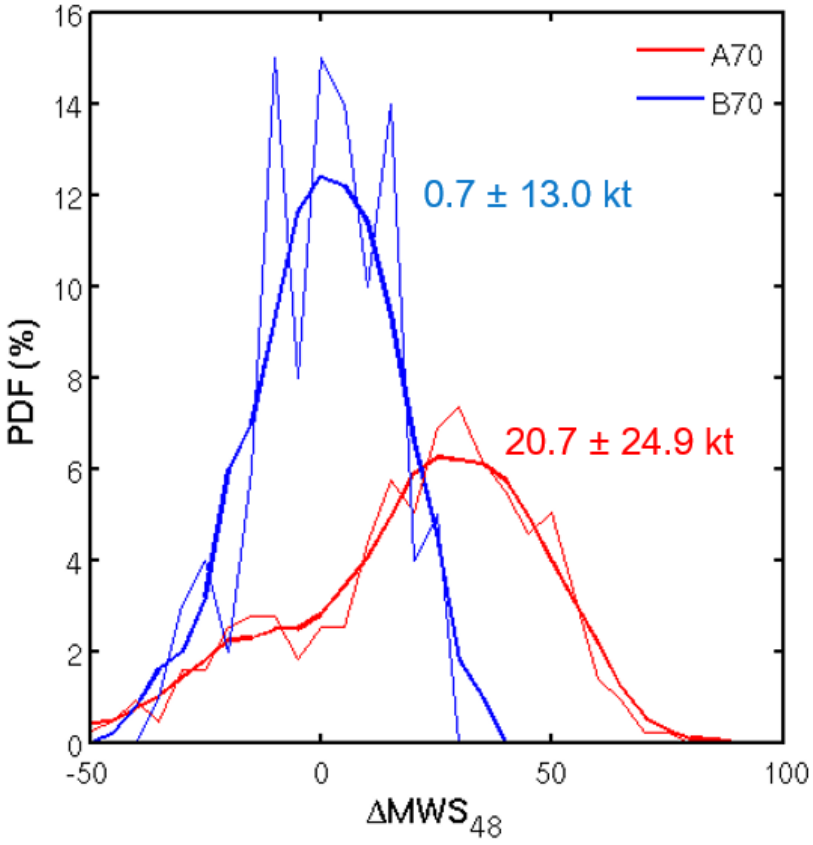

Fig. 21. Comparison of the intensity change in 48-h PDFs for classified two groups (red: A70, blue: B70). The thick lines indicate smoothed PDF.

As shown in Table 8, there are a total of 86 A70 cases and 27 B70 for the period 2004-2016. There are 26 A70 cases and 10 B70 cases in the period 2014-2016. The decision tree is trained by data in 2004-2013, and data in 2014-2016 are used for test. We used SMOTE to increase the number of samples in B70, to solve the problem of imbalance in the training data set. After the A70 and B70 are identified, we calculate the values for each variable.

Table 9 describes the 16 types potential variables included to build the decision tree for classification of two types of LMI. The variables include 4 static variables and 12 synoptic

variables. Such variables are similar to those used to develop CSTIPS-DAT in NTC, but exclude intensification potential (POT; MPI minus initial intensity) -based variables. The POT-based variables, which were most important factors in CSTIPS-DAT, are similar to MPI-based variables at the beginning of TC because TCs initial intensity is close to constant as 30-40kt. Therefore, only MPI was included on behalf of the above two.

The variables were obtained by averaging the values up to 3.25 days after the genesis, which is the sum of LMI-reaching time (1.7 days) and its one standard deviation (1.55 days). The size of averaging area is following CSTIPS-DAT.

DATs and DAT-based MPIs have the highest correlation with LMI among all variables (Table 9). Ocean heat contents (OHC), which is widely used as a representative value of the upper ocean thermal condition, also showed a high correlation coefficient but it was lower than some DATs. This result is consistent with Price (2009), which claims that OHC and DATs contain similar information in sufficiently deep-ocean but DATs have more realistic information than OHC.

Table 8. Number of A70 and B70 during 2004-2013, 2014-2016 and 2004-2016.

Period	A70	B70	Total
2004-2013	60	17	77
2014-2016	26	10	36
2004-2016	86	27	113

Table 9. List of potential variables included in the present model and its correlation coefficient with life time maximum intensity.

Variable	Description	*r*
JDAY	Absolute value of Julian day – 248	-0.27
LAT	Latitude of typhoon location	-0.33
LON	Longitude of typhoon location	0.07
SPD	Storm moving speed	-0.23
D200	Area-averaged (0 km to 1000 km) divergence at 200hPa	0.05
RV500	Area-averaged (0 km to 1000 km) relative vorticity at 500 hPa	0.16
RV850	Area-averaged (0 km to 1000 km) relative vorticity at 850 hPa	0.04
U200	Area-averaged (200 km to 800 km) zonal wind at 200 hPa	-0.28
T200	Area-averaged (200 km to 800 km) air temperature at 200 hPa	-0.39
RHHI	Area-averaged (200 km to 800 km) relative humidity 500 – 300 hPa	0.32
RHLO	Area-averaged (200 km to 800 km) relative humidity 850 – 700 hPa	0.29
SH200	Area-averaged (200 km to 800km) 200 hPa to 850 hPa vertical wind shear	-0.17
SH500	Area-averaged (200 km to 800 km) 500 hPa to 850 hPa vertical wind shear	-0.32
OHC	Area-averaged (0 km to 200 km) ocean heat contents	0.52
DAT - DAT120	Ocean temperatures averaged from the surface down to the various depth (10-120 m, 10-m interval)	0.48 - 0.54
MPI - MPI120	Maximum potential intensity using DAT - DAT120	0.47 - 0.56

4.2 Optimal min-leaf size

The Confusion Matrix can be used to calculate verification measures such as Probability of Detection (POD), False Alarm Rate (FAR) and accuracy (ACC). POD is the ratio of the correct forecasts of event to the actual number of event while FAR is the ratio of incorrect forecasts to the total number of event forecasts. The POD, the FAR and the ACC values are calculated as follow.

$$\mathrm{POD} = \frac{TP}{TP+TN} \qquad (15)$$

$$\mathrm{FAR} = \frac{FP}{FP+TP} \qquad (16)$$

$$\mathrm{ACC} = \frac{TP+FN}{TP+FP+TN+FN} \qquad (17)$$

where TP is the true positives, TN is the true negatives, FP is the false positives, and FN is the false negatives.

Figure 22 shows the classification performance of the decision-tree in training period as a function of the min-leaf-size parameter, which stops splitting when the number of samples in a leaf gets below the specified size. The decision trees showed the highest ACC and POD in min-leaf size in 1, and these values decreased with increasing min-leaf size. FAR showed a distribution of 0~12% according to min-leaf size, and there was no significant trend with min-leaf size.

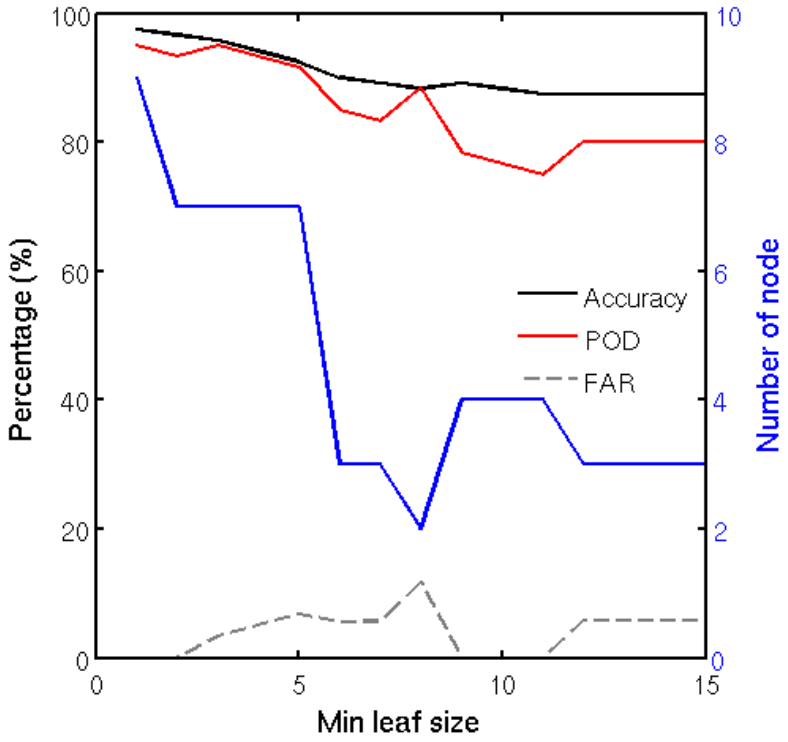

Fig. 22. Comparison of the skill scores (black: Accuracy, red: POD, gray: FAR) and number of node (blue line) at each min-leaf size.

The smaller the min-leaf size, the better the performance, however increasing nodes due to min-leaf size reduction makes complicated decision tree, which may lead the over-fitting. Statistical predictions should train the model to make reliable predictions of the test data. In over-fitting, trained model describes noise instead of the underlying relationship. Excessively over-fitted model has poor performance in the real-time prediction, as it overreacts to minor fluctuations in training data. Therefore, to avoid this prediction instability problem, the optimal min-leaf size should be determined through cross-validation loss comparison. The k-fold cross-validation loss is obtained by averaging k-times of verification performance such as misclassification, it can be expressed follow.

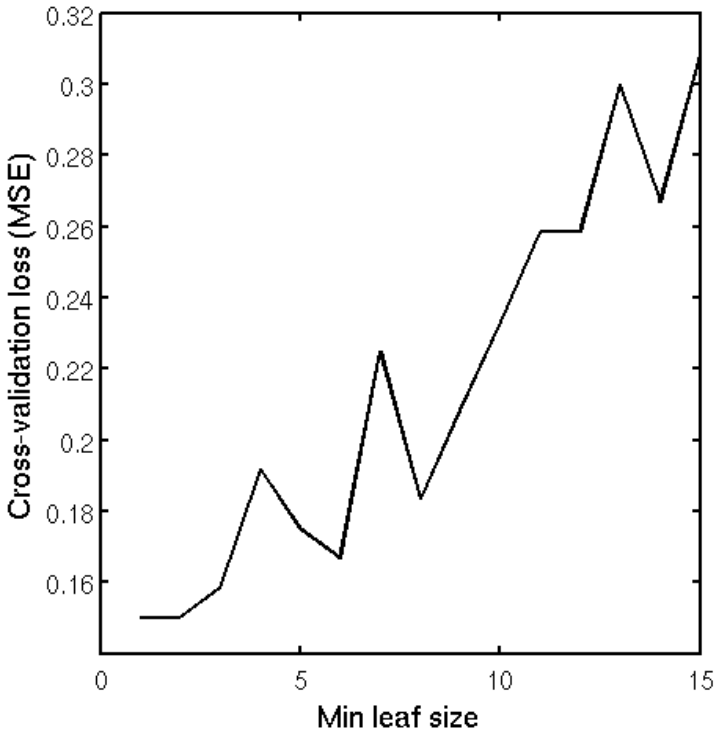

Fig. 23. Distributions of cross-validation loss according to min-leaf size. The k-fold cross-validation method is used.

$$\mathrm{CV} = \frac{1}{k}\sum_{i=1}^{k} MSE_i \qquad (18)$$

where MSE is mean squared error. Figure 23 indicates the comparison of *k*-fold cross-validation loss according to min-leaf size. Overall, the cross-validation loss in our analysis, tended to increase as the min leaf size increased. The loss was the smallest at minleaf size 1-2, followed by local minima at 6 and 8. However, decision trees with minleaf size of 1-2 show nine and seven nodes (fairly complex structure), respectively, whereas with minleaf size of 6 and 8, show three and two nodes, respectively (blue line in Fig 22). In this study, we set the min-leaf size to 6, in which the structure of the decision tree is relatively simple and the cross-validation loss is small.

4.3 Governing rules of classification model

The trained decision tree includes three nodes with four rules (Fig. 24). Table 10 lists four rules governing decision tree. Rule 1 shows that a TC in a low MPI20 environment is difficult to develop with strong intensity. MPI is the most important predictor in previous developed statistical intensity prediction skills (Knaff et al., 2005; Kaplan et al., 2010; Gao and Chiu, 2012). In the rule 1, shallow DAT-based MPI is selected as classification factor, and many weak TCs were classified based on this rule. The weak TCs cannot interact with deep-ocean, thus the shallow DAT-based MPI should be a good criteria for classifying the weak TCs.

Rule 2 states that if MPI20 ≥ 114 kt, and LAT ≥ 22.1ºN, then the LMI will below 70 kt, suggesting that a TC in subtropics will not develop to strong intensity even if it is in high MPI20 environment. TCs with high LAT travel mainly to the north, or occurrence at relatively high latitude. Because the meridional gradient of VWS is large in this region, TC can be suppressed with northward moving.

Rule 3 states that If MPI20 ≥ 114 kt, LAT < 22.1ºN, and DAT100 < 26.3℃, then the LMI will below 70 kt. High MPI20 and low LAT are both favorable for TC intensity, whereas the TC is difficult to develop with strong intensity if DAT100 is lower than 26. 3℃. Previous studies suggested that 100 m is typical depth of vertical mixing by strong TC, thus DAT100 is the realistic value to represent the sea surface thermal condition under the TC (Price, 2009; Balaguru et al., 2015). If DAT100 is lower than 26.3℃, which is close value known as the dew point air-temperature of the tropic, the TC is difficult to develop for strong intensity because the heat flux has a negative sign.

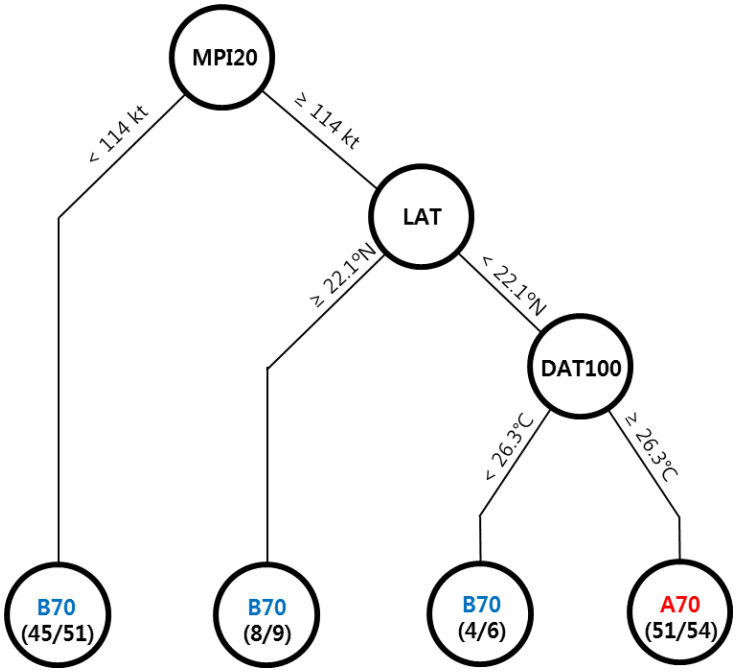

Fig. 24. The decision tree for life time maximum intensity classification constructed from C2. The numbers in parenthesis indicate the total data points and the number of hit samples

Rule 4 states that If MPI20 ≥ 114 kt, LAT < 22.1ºN, and DAT100 ≥ 26.3℃, then the LMI will exceed 70 kt. This rule is the only rule for A70, and it has an accuracy of 94.4%. This rule indicates that TC generally develops with strong intensity when all three conditions are satisfied.

To evaluate the accuracy of this decision tree in classifying A70 and B70, both training and test period are conducted. The decision tree shows a classification accuracy of 92.5% in the training (Table 11), and the accuracy of the test is 80.5%. The confusion matrix of the test period in Table 12 indicates that 24 out of 26 A70 TCs are correctly classified and 5 out of 29 classified A70 TCs are actually B70 TCs; thus, the POD is 92.3% and the FAR is only 17.2%.

Table 10. Descriptions and accuracies of the rules derived from developed decision tree.

Rule NO.	Description	Accuracy
1	If MPI20 < 114 kt, then the LMI will below 70 kt.	45/51 = 88.2%
2	If MPI20 ≥ 114 kt, and LAT ≥ 22.1ºN, then the LMI will below 70 kt.	8/9 = 88.9%
3	If MPI20 ≥ 114 kt, LAT < 22.1ºN, and DAT100 < 26.3℃, then the LMI will below 70 kt.	4/6 = 66.7%
4	If MPI20 ≥ 114 kt, LAT < 22.1ºN, and DAT100 ≥ 26.3℃, then the LMI will exceed 70 kt.	51/54 = 94.4%

Table 11. Confusion matrix from the developed decision tree in training period.

		Model	
		A70	B70
Observed	A70	51	9
	B70	3	57

Table 12. Same as table 11, but test period.

		Model	
		A70	B70
Observed	A70	24	2
	B70	5	5

5. Summary and conclusions

A statistical-dynamical typhoon intensity prediction model in the WNP was developed using track-pattern clustering and DAT-based potential predictors. In this model system (CSTIPS-DAT), TCs were classified into five clusters based on the fuzzy *c*-means clustering method and the distinctive characteristics of each cluster were examined. The present model uses six static, 11 synoptic potential predictors as well as three new DAT-based ocean-coupling predictors that consider TC-induced vertical mixing effect.

To investigate the effect of using clustering and DAT-based predictors in TC intensity prediction, four experiments were conducted for the training period of 2004-2012 and real-time predictions in 2013 and 2014. Our results suggest the use of clustering remarkably reduced the MAE relative to the non-clustered experiments in both training and real-time experiments, about 25% (72-h lead time; CSTIPS-SST) and 20-25% (24-96-h; clustered models), respectively. We found that the classified clusters had unique characteristics in terms of the tendency of intensity change, which reduced the variability (σ_C) of predictand and ultimately led to the improvement of TC intensity prediction. The effect of DAT-based predictors was not significant as much as that of clustering, but still further improved the performance. By using both clustering and DAT-based predictors, various sets of predictors could be skillfully selected depending on each cluster and its unique characteristics related to mean TC intensity, landfalls, upper-ocean thermal structure, and TC-induced mixing depth along the TC tracks. The test performance of the present model for 2013 and 2014 was more skillful than operational dynamical models up to 24-h leads. After 24 hours, the CSTIPS-DAT model still shows skill comparable with or more superior to the operational dynamical models. The present model showed its limitations in the improvement, particularly after 96-h lead time. This seems to be related to over-fitting due to insufficient sample size. Also, the

performance of the present model was not good enough for TCs with low membership coefficients, indicating the low similarity of the individual TC tracks to the mean tracks of their respective membership clusters. This implies that an increase of the number of clusters might be needed to allow higher membership coefficients and consequently more accurate prediction results. If the sample size increases in the future as more TC data are collected, the problem of over-fitting and low membership coefficients is expected to be improved.

The WNCP distribution of LMI for the period 2004-2016 has two local maxima around 50 kt (small peak) and 100 kt (large peak), and a local minima at 70kt. Based on this bimodal distribution of LMI, we have classified TCs in the WNCP into two groups TCs with the LMI of above 70 kt (A70) and TCs below 70kt (B70). Analysis reveals that B70 has a larger MAE than A70 after 72-h lead times mostly due to a significant overestimation in B70. These results suggest that if we can correct the bias of B70 through the binary classification of TCs in the WNCP according to their LMI and develop the independent prediction model for the classified groups, it can contribute to reducing the variance of the intensity change, ultimately leading to improvement of the TC intensity prediction.

In this study, in order to improve the prediction skill of CSTIPS-DAT and better understanding of LMI bimodality, we employed the decision tree algorithm that classifies the type of TCs in the early stage depending on whether or not TCs will intensify to the LMI of above 70 kt during their lifetime. The present decision tree is developed using the CART algorithm and consists of four leaf nodes. MPI20, LAT and DAT100 are selected to classify A70 and B70, in which the splitting values of the three variables are 114 kt for MPI20, 22.1ºN for LAT and 26.3℃ for DAT100. The decision tree has classification accuracy of 92.5% for training period, and of 80.5% for test period.

Lee16 reported that lifetime of TCs is important in determining their LMI. The lifetime of TCs is largely influenced by the track that determines whether or not it make landfall. However, TCs in the WNCP are mostly independent from the lifetime, because they stay mainly in the tropical open ocean. MPI20, LAT, and DAT100, which were selected as classification criteria in the present decision tree, are independent of lifetime and they are factors related to the development of TC. This supports Lee16's claim that the bimodality in LMI distribution can be better explained by the magnitude of the intensity change than the lifetime of the TCs.

It is also important fact that DAT100 has been selected as an important classification criterion of LMI. Correlation coefficients of DATs before classification for LMI showed the maximum value at DAT50 (Fig 25). However, TCs classified by rule 4 (*i.e.* strongly-developing TCs of A70) had the highest correlation at DAT100. Price (2009) proposed the DAT100 as the ocean factor which reflected the TC-induced surface cooling by Saffir-Simpson category 3 TCs (96-113 kts). Interestingly, this category belongs to the second peak of the LMI distribution (Fig. 19b) and about 40% of the TCs in the WNCP belong to this category. On the other hand, the TCs classified by the other rules (*i.e.*, non-developing TCs of B70) has a very low correlation at most DATs (Fig 25 blue line). These results suggest that pre-existing ocean thermal structures along the track is an important factor to determine the LMI for strongly-developing TCs whereas not for non-developing TCs. In this regard, it is necessary to develop an individual model according to the two classified groups since it will allow taking into account the environmental factors varying depending on the intensity of TCs.

Understanding the distribution of LMI is crucial to improve the prediction skill of TC intensity. In this study, the rules discovered in the tree suggest that the upper-ocean

subsurface structure and the vertical mixing process play an important role to determine the LMI in TCs, and provide a clue to explain the causes of the bimodal distribution of LMI. However, the present findings are far from the fully understanding of the causes of the bimodal distribution. Further research using more advanced approaches such as deep-learning technique are guaranteed to solve the problem in the future.

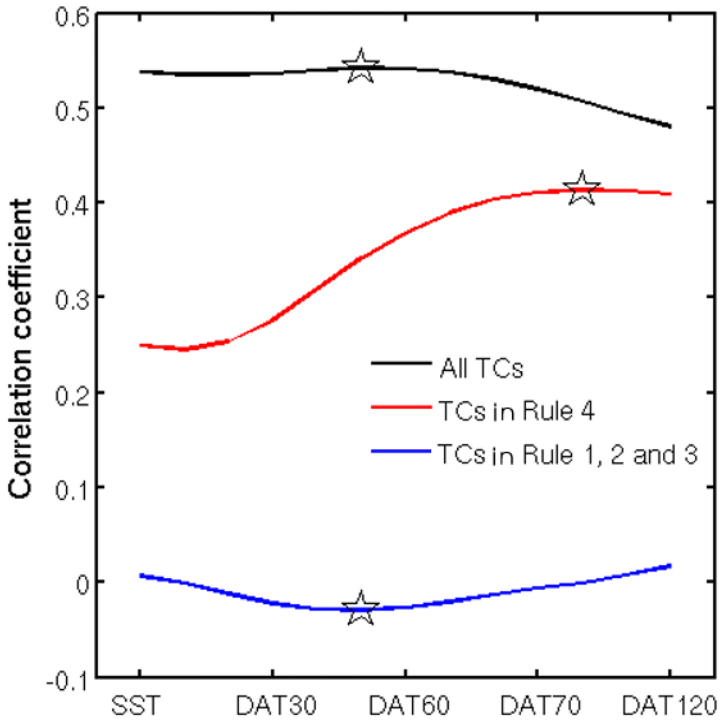

Fig. 25. The comparison of the correlation coefficients between DAT variables and LMI. The black line is non-classified TCs, the red is TCs in rule 4 and the blue line is TCs in rule 1, 2 and 3. Open pentagrams represent the location of maximum value for the three groups.

6. References

Aczel, A. D., 1989: Complete Byahusiness Statistics. Richard D. Irwin, 1056 pp.

Balaguru, K., G. R. Foltz, L. R. Leung, E. D'Asaro, K. A. Emanuel, H. Liu, and S. E. Zedler, 2015: Dynamic Potential Intensity: An improved representation of the ocean's impact on tropical cyclones, *Geophys. Res. Lett.*, **42**, 6739-6746, doi:10.1002/2015GL064822.

Bengtsson, L., M. Botzet, and M. Esch, 1996: Will greenhouse-induced warming over the next 50 years lead to higher frequency and greater intensity of hurricanes? *Tellus*, **48A**, 57–73.

Berry, M. and G. Linoff, 1997: Data Mining Techniques: for Marketing, Sales, and Customer Support. *John Wiley & Sons, New York.*

Bezdek, J. C., 1981: Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic, 256 pp.

Bleck, R., 2002. An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. *Ocean Modelling*, **4**, 55-88.

Bogner, P. B., G. M. Barnes, and J. L. Franklin, 2000: Conditional instability and shear for six hurricanes over the Atlantic Ocean. *Wea. Forecasting*, **15**, 192-207.

Camargo, S. J., A. W. Robertson, S. J. Gaffney, P. Smyth, and M. Ghil, 2007a: Cluster analysis of typhoon tracks. Part I: General properties. *J. Climate*, **20**, 3635–3653.

Camargo, S. J., A. W. Robertson, S. J. Gaffney, P. Smyth, and M. Ghil, 2007b: Cluster analysis of typhoon tracks. Part II: Large-scale circulation and ENSO. *J. Climate*, **20**, 3654–3676.Chu, P.-S., and X. Zhao, 2011: Bayesian analysis for extreme climatic events: A review. *Atm. Res.*, **102**, 243-262.

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*.

Chu, P.-S., X. Zhao, C.-H. Ho, H.-S. Kim, M.-M. Lu, and J.-H. Kim, 2010: Bayesian forecasting of seasonal typhoon activity: A track-pattern-oriented categorization approach. *J. Climate*, **23**, 6654-6668.

Corbosiero, K.L., and J. Molinari, 2002: The effects of vertical wind shear on the distribution of convection in tropical cyclones. *Mon. Wea. Rev*., **130**, 2110-2123.

DeMaria, M., and J. Kaplan, 1994a: A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. *Wea. Forecasting*, **9**, 209–220.

DeMaria, M., and J. Kaplan, 1994b: Sea surface temperature and the maximum intensity of Atlantic tropical cyclones. *J. Climate*, **7**, 1324–1334.

DeMaria, M., 1996: The effect of vertical shear on tropical cyclone intensity change, *J. Atmos. Sci*., **53**, 2076-2087.

DeMaria, M., and J. Kaplan, 1999: An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **14**, 326–337.

DeMaria, M., M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, 2005: Further improvements to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting*, **20**, 531–543.

Emanuel, K. A., 1988: Toward a general theory of hurricanes. *Am. Sci*., **76**, 370-379.

Emanuel, K. A., 2000: A statistical analysis of tropical cyclone intensity. *Mon. Weather Rev*. **128**, 1139–1152.

Emanuel, K. A., 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, **436**, 686–688.

Fitzpatrick, P. J. 1997: Understanding and Forecasting Tropical Cyclone Intensity Change with the Typhoon Intensity Prediction Scheme (TIPS). *Wea. Forecasting*., **12**, 826-846.

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.

Gao, S., and L. S. Chiu, 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.

Gao, S., W. Zhang, J. Liu, I.-I. Lin, L. S. Chiu, and K. Cao, 2016: Improvements in typhoon intensity change classification by incorporating an ocean coupling potential intensity index into decision trees. *Weather Forecasting*, **31**, 95-106, doi: 10.1175/WAF-D-15-0062.1.

Jarvinen, B. R., and C. J. Neumann, 1979: Statistical forecasts of tropical cyclone intensity for North Atlantic Basin. *NOAA Tech. Memo.,* BWS NHC 10, 22 pp. [National Hurricane Center, Coral Gables, FL 22146.]

Kaplan, J., M. DeMaria, 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.

Kim, H.-S., J.-H. Kim, C.-H. Ho, and P.-S. Chu, 2011: Pattern classification of typhoon tracks using the fuzzy c-means clustering method. *J. Climate*, **24**, 488-508.

Kim, H.-S., C.-H. Ho, J.-H. Kim, and P.-S. Chu, 2012: Track-pattern bases model for seasonal prediction of tropical cyclone activity in the western North Pacific. *J. Climate*, **25**, 4660-4678.

Knaff, J. A., M. DeMaria, C. R. Sampson, and J. M. Gross, 2003: Statistical 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting*, **18**, 80–92.

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.

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

Kossin, J. P., K. A. Emanuel, and G. A. Vecchi, 2014: The poleward migration of the location of tropical cyclone maximum intensity. *Nature*, **509**, 349–352.

Lee, C-Y., M. K. Tippett, A. H. Sobel, and S. J. Camargo, 2016: Rapid intensification and the bimodal distribution of tropical cyclone intensity. *Nat. Commun.* **7**:10625 doi: 10.1038/ncomms10625.

Leipper, D., and D. Volgenau, 1972: Hurricane heat potential of the Gulf of Mexico. *J. Phys. Oceanogr.*, **2**, 218–224.

Li, W., C. Yang, and D. Sun, 2009: Mining geophysical parameters through decision-tree analysis to determine correlation with tropical cy-clone development, *Comput. Geosci*., **35**, 309-316, doi:10.1016/j.cageo.2008.02.032.

Lin. I-I., C.-C. Wu, I.-F. Pun, and D.-S. Ko, 2008: Upper-Ocean Thermal Structure and the Western North Pacific Category 5 Typhoons. Part I: Ocean Features and the Category 5 Typhoons Intensification, *Mon. Wea. Rev*, **136**, 3288-3306.

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.

Lin. I-I., I.-F. Pun, and C.-C. Wu, 2009: Upper Ocean Thermal Structure and the Western North Pacific Category-5 Typhoons Part II: Dependence on Translation Speed, *Mon. Wea. Rev*. **137**, 3744-3757.

Lin, I.-I., P. Black, J. F. Price, C. Y. Yang, S. S. Chen, C. C. Lien, P. Harr, N. H. Chi, C. C. Wu, and E. A. D'Asaro, 2013: An ocean coupling potential intensity index for tropical cyclones, *Geophys. Res. Lett*., **40**, 1878–1882, doi:10.1002/grl.50091.

Mainelli-Huber, M., 2000: On the role of the upper ocean in tropical cyclone intensity change. M.S. thesis, Dept. of Meteorology and Physical Oceanography, University of Miami, 73 pp.

Manganello, J. V., K. I. Hodges, J. L. Kinter III, B. A. Cash, L. Marx, T. Jung, D. Achuthavarier, J.D. Adams, E. L. Altshuler, B. Huang, E. K. Jin, C. Stan, P. Towers, and N. Wedi, 2012: Tropical cyclone climatology in a 10-km global atmospheric GCM: Toward weather-resolving climate modeling. *J. Climate*, **25**, 3867-3893.

Moon, I.-J., and S. J. Kwon, 2012: Impact of upper-ocean thermal structure on the intensity of Korean peninsula landfall typhoons, *Progress in Oceanography*, **105**, 61-66.

Moon, I.-J., S.-H. Kim, P. Klotzbach, and J. C. L. Chan, 2015: Roles of interbasin frequency changes in the poleward shifts of the maximum intensity location of tropical cyclones, *Environ. Res. Lett*, **10**, 104004.

Park, D.-S. R., C.-H. Ho, and J.-H. Kim, 2014: Growing threat of intense tropical cyclones to East Asia over the period 1977-2010. *Environ. Res. Lett*. **9**, 014008.

Price, J. F., 2009: Metrics of hurricane-ocean interaction: Vertically-integrated or vertically-averaged ocean temperature? *Ocean Sci.,* **5**, 351–368.

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.

Soloviev, A. V., R. Lukas, M. A. Donelan, B. K. Haus, and I. Ginis, 2014: The air-sea interface and surface stress under tropical cyclones. *Sci. Rep*. **4**, 5306.

Torn, R. D., and C. Snyder, 2012: Uncertainty of tropical cyclone best-track information. *Weather Forecast.* **27**, 715–729.

Wang, Y., and G.J. Holland, 1996: Tropical cyclone motion and evolution in vertical shear. *J. Atmos. Sci.*, **53**, 3313-3332.

Webster, P. J., G. Holland, J. A. Curry, and H.-R. Chang, 2005: Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science*, **309**, 1844–1846.

Wu, L., H. Su, R. Fovell, B. Wang, J. T. Shen, B. Kahn, S. M. Hristova-Veleva, B. Lambrigtsen, E. J. Fetzer, and J. Jiang, 2012: Relationship of environmental relative humidity with North Atlantic tropical cyclone intensity and intensification rate, *Geophys. Res. Lett.*, **39**, L20809, doi: 10.1029/2012GL053546.

Zhang, W., S. Gao, B. Chen, and K. Cao, 2013: 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.

Zhang, W., 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.

Zhao, M., I. M. Held, S.-J. Lin, and G. A. Vecchi, 2009: Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Climate*, **22**, 6653–6678.