The Analysis of Tropical Cyclone Tracks in the Western North Pacific through Data Mining. Part II: Tropical Cyclone Landfall

WEI ZHANG

Department of Geography and Resource Management, The Chinese University of Hong Kong, Sha Tin, and Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, Hong Kong, China

YEE LEUNG

Department of Geography and Resource Management, and Institute of Environment, Energy and Sustainability, The Chinese University of Hong Kong, Sha Tin, and Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, Hong Kong, China

JOHNNY C. L. CHAN

Guy Carpenter Asia-Pacific Climate Impact Center, School of Energy and Environment, City University of Hong Kong, Hong Kong, China

(Manuscript received 13 January 2012, in final form 7 November 2012)

ABSTRACT

This is the second paper of a two-part series of papers on the analysis of tropical cyclone (TC) tracks in the western North Pacific Ocean. In this paper, TC landfalls in the South China Sea and western North Pacific basins are investigated through the data-mining approach. On the basis of historical TC archives, the C4.5 algorithm, a classic tree algorithm for classification, has been employed to quantitatively discover rules governing TC landfall. A classification tree, with 14 leaf nodes, has been built. The path from the root node to each leaf node forms a rule. Fourteen rules governing TC landfall across the Chinese coast have been unraveled with respect to the selected attributes having potential influence on TC landfall. The rules are derived by the attributes and splitting values. From the classification tree, split values, such as 27°N latitude, 130°E longitude, 141°E in the west extension index, and 0.289 in the monsoon index have been shown to be useful for TC forecasting. The rules have been justified from the perspective of meteorology and knowledge of TC movement and recurvature (e.g., deep-layer mean winds and large-scale circulation). The research findings are also consistent with existing results concerning TC movement and landfall. Both the unraveled rules and the associated splitting values can provide useful references for the prediction of TC landfall over China.

1. Introduction

As shown by Zhang et al. (2013), which is Part I of this two-part paper, tropical cyclone (TC) recurvature and landfall are two TC-related scientific problems of significant economic and social implications. Part I presents the analysis of TC recurvature based on the C4.5 algorithm. In this paper, particular attention will be paid to TC landfall problems.

The greatest damage caused by TCs tends to occur during or after they make landfall. In addition, the structure, intensity, and motion of the TC will change dramatically during the landfall process. Improved forecasting of TC landfall provides many potential benefits for coastal communities. Landfall has, therefore, attracted great attention from scientists in recent years, after it was highlighted by the U.S. Weather Research Program (Marks and Shay 1998).

Considerable research has investigated the seasonal or annual forecasts of TC landfall frequency in the Atlantic and western North Pacific (WNP) Ocean basins. Particular attention has also been focused on landfall over regional areas, for example, southern China and the Korean Peninsula. The variability of seasonal TC landfall

Corresponding author address: Yee Leung, Dept. of Geography and Resource Management, Institute of Environment, Energy and Sustainability, The Chinese University of Hong Kong, Sha Tin, Hong Kong, China.
E-mail: yeeleung@cuhk.edu.hk

DOI: 10.1175/JAMC-D-12-046.1

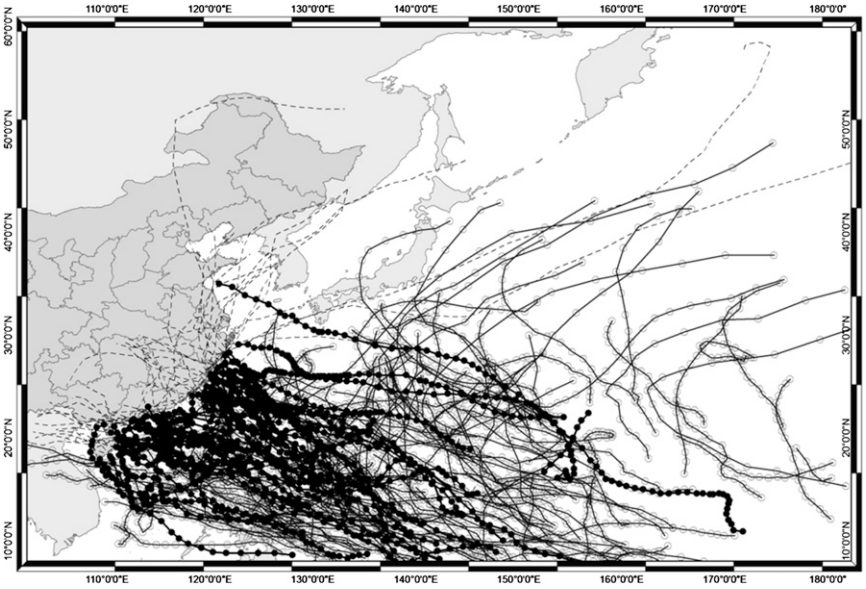

FIG. 1. The landfalling and nonlandfalling TCs for the period 2000–09. The landfalling TC tracks consist of two parts, i.e., prior to (thick solid curve) and after (dashed curve) landfall. The solid circles are the prior-to-landfall observations of landfalling TCs. The thin curves and hollow circles denote nonlandfalling TC tracks and observations, respectively.

frequency was attributed to the modulation of El Niño–Southern Oscillation (ENSO), the Madden–Julian oscillation (MJO), the quasi-biennial oscillation (QBO), the Atlantic multidecadal oscillation, and the Pacific decadal oscillation (PDO). The variability of large-scale circulation patterns (the subtropical high, monsoon troughs, and the midlatitude westerlies) associated with ENSO or other oscillations are likely to be responsible for most of the seasonal or annual variations in TC landfall in the WNP and Atlantic basins.

TC–land interaction during landfall has been studied through observational analysis and dynamic modeling over the decades (Tuleya and Kurihara 1978; Tuleya et al. 1984; Bender et al. 1985; Chan and Liang 2003; Wong and Chan 2006; Wong 2007; Wong et al. 2008). The impact of topography on landfalling TCs is more significant than roughness variation, especially in mountainous areas (Chang 1982). Observations (Brunt 1968; Hamuro et al. 1969) have confirmed that high correlations exist between the areas of maximum rainfall during TC landfall and the mountainous terrain of the region. Brand and Blelloch (1974) found that topographical effects, if not properly accounted for, can cause significant errors in the forecasting of a TC's motion. Chang (1982) also observed that the interaction between the terrain and the TC caused strong easterlies to develop to the north of Taiwan, accelerating the TC motion in his case.

To recapitulate, previous research on TC landfall prediction was concerned with seasonal or annual TC landfall frequency according to the variability of large-scale circulation modulated by ENSO, QBO, and MJO; with the precipitation distribution associated with TC landfall; with changes in their tracks and structures when TCs passed through an island by studying land–TC interaction; and with the TC decaying mechanisms post-landfall. However, few studies have investigated the mechanisms controlling TC landfalls over the Chinese coast based on the historical TC database. In addition, little research has been performed on TC forecasting solely with regard to TC landfall. Up until recently, few objective forecasting aids in operational use have been specifically designed to identify landfall situations. Given our existing understanding of TC landfall mechanisms based on observational and statistical analysis and dynamic modeling (Elsner and Liu 2003; Liu and Chan 2003; Bhowmik et al. 2005; Brettschneider 2008; Goh and Chan 2010), large-scale circulation exerts a dramatic influence on TC landfall in the South China Sea (SCS) and WNP basins. As stated in Part I, TCs forming in the SCS and WNP move in a westward or northwestward direction under a westward steering current when the subtropical high is relatively strong and shifts westward. These TCs tend to make landfall along the Chinese coast when the steering current is persistently strong and westward. By contrast, westward-moving TCs will turn to the north and then to the northeast if the steering flow is weakened or even changes direction from westward to eastward as a result of the

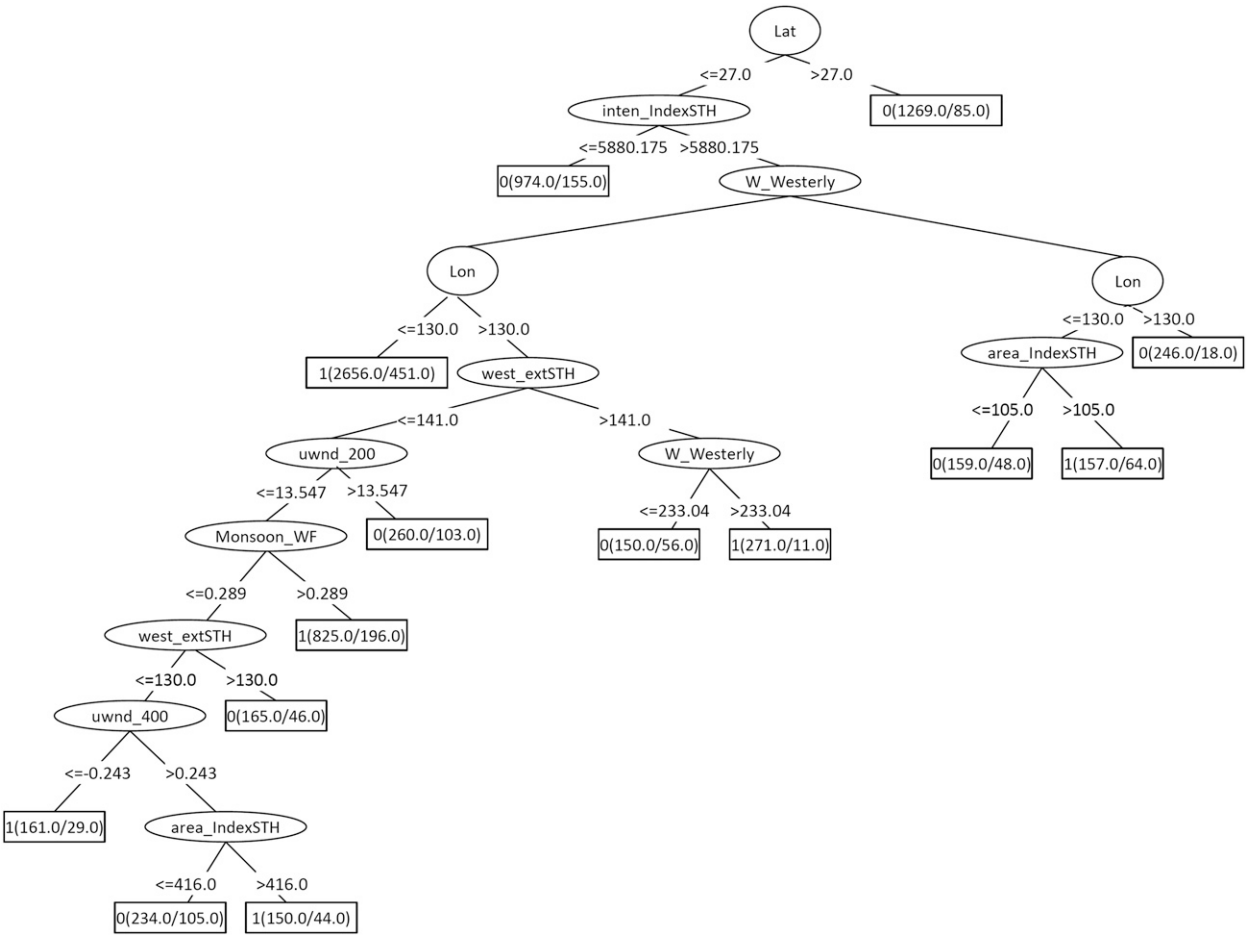

FIG. 2. The classification tree governing TC landfall, as constructed by the C4.5 algorithm.

interaction of large-scale circulation. In general, TCs will move far away from the Chinese coast without making landfall there because of the reversal of the steering flow from westward to eastward.

Data mining is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Fayyad et al. 1996; Fayyad and Stolorz 1997; Fayyad 1997; Miller and Han 2001; Han and Kamber 2006; Leung 2010). Given the rapid rate of growth of the TC-related datasets in term of storage volume and dimensions (i.e., attributes), as well as the increasing amount of useful knowledge (e.g., structures, processes, relationships, regularities, and patterns) hidden in TC-related datasets, this study seeks to unravel the rules and regularities governing TC landfall over the Chinese coast from the historical TC archives through data-mining methods (here, the C4.5 algorithm). The discovered rules can provide useful references for forecasting aids specifically for TCs that make landfall over the Chinese coast.

This study is organized as follows. Section 2 introduces the study area and data source. The methodology is introduced in section 3. Section 4 explains and analyzes the results generated by the C4.5 algorithm. Concluding remarks are discussed in section 5.

2. Study area, data, and methodology

The study area covers the SCS and WNP. All of the TCs that occurred in this study area from 2000 to 2009 are employed as learning samples for rule discovery. The data related to TC landfall consists of two major classes: TC best-track data and meteorological data.

It is noted that landfalling TCs in this study are TCs that make landfall over the Chinese coast. The TC best-track data are made available from the Japan Meteorological Agency's (JMA) Regional Specialized Meteorological Center Tokyo (RSMC Tokyo). This postanalysis best-track dataset consists of 6-h estimates of TC position (latitude and longitude), minimum central pressure (MCP), and the 10-min maximum sustained wind speed of all of the TCs in the WNP basin, including the SCS, starting from 1951. There are a total of 6920 TC sample

TABLE 1. The 14 unraveled rules governing TC landfall.

Rule	Attributes	Accuracy
If lat > 27°, then the TC will *not* make landfall along the Chinese coast	Lat	1184/1269 = 0.933
If lat ≤ 27° and inten_IndexSTH ≤ 5880.175, then the TC will *not* make landfall along the Chinese coast	Lat and intensity index of the subtropical high	819/974 = 0.841
If lat ≤ 27°, inten_IndexSTH > 5880.175, W_Westerly ≤ 477.463, and lon ≤ 130°, then the TC will make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, and lon	2205/2656 = 0.830
If lat ≤ 27°, inten_IndexSTH > 5880.175, W_Westerly ≤ 477.463, lon > 130°, west_extSTH > 141, and W_Westerly > 233.04, then the TC will *not* make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, lon, and zonal wind in the 500-hPa layer	260/271 = 0.959
If lat ≤ 27°, inten_IndexSTH > 5880.175, W_Westerly ≤ 477.463, lon > 130°, west_extSTH > 141, and W_Westerly ≤ 233.04, then the TC will make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, lon, and westward extension index of the subtropical high	94/150 = 0.627
If lat ≤ 27°, inten_IndexSTH > 5880.175, W_Westerly > 477.463, and lon > 130°, then the TC will *not* make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, and lon	228/246 = 0.927
If lat ≤ 27, inten_IndexSTH > 5880.175, W_Westerly > 477.463, lon ≤ 130°, and area_IndexSTH ≤ 105.0, then the TC will *not* make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, lon, and area index of the subtropical high	111/159 = 0.698
If lat ≤ 27°, inten_IndexSTH > 5880.175, W_Westerly > 477.463, lon ≤ 130°, and area_IndexSTH > 105.0, then the TC will make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, lon, and area index of the subtropical high	93/157 = 0.592
If lat ≤ 27°, inten_IndexSTH > 5880.175, W_Westerly ≤ 477.463, lon > 130°, west_extSTH ≤ 141, and uwnd_200 > 13.547, then the TC will *not* make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, lon, westward extension index of the subtropical high, and zonal wind in the 200-hPa layer	157/260 = 0.604
If lat ≤ 27°, inten_IndexSTH > 5880.175, W_Westerly ≤ 477.463, lon > 130°, west_extSTH ≤ 141, uwnd_200 ≤ 13.547, and Monsoon_WF > 0.289, then the TC will make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, lon, westward extension index of the subtropical high, zonal wind in the 200-hPa layer, and monsoon index	629/825 = 0.762
If lat ≤ 27°, inten_IndexSTH > 5880.175, W_Westerly ≤ 477.463, lon > 130°, west_extSTH ≤ 141, uwnd_200 ≤ 13.547, Monsoon_WF≤0.289, and west_extSTH > 130, then the TC will *not* make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, lon, westward extension index of the subtropical high, zonal wind in the 200-hPa layer, and monsoon index	119/165 = 0.721
If lat ≤ 27°, inten_IndexSTH > 5880.175, W_Westerly ≤ 477.463, lon > 130°, west_extSTH ≤ 141, uwnd_200 ≤ 13.547, Monsoon_WF≤0.289, west_extSTH ≤ 130, and uwnd_400 ≤ −0.243, then the TC will make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, lon, westward extension index of the subtropical high, zonal wind in the 200- and 400-hPa layers, and monsoon index	132/161 = 0.820
If lat ≤ 27°, inten_IndexSTH > 5880.175, W_Westerly ≤ 477.463, lon > 130°, west_extSTH ≤ 141, uwnd_200 ≤ 13.547, Monsoon_WF ≤ 0.289, west_extSTH ≤ 130, uwnd_400 > −0.243, and area_IndexSTH > 416, then the TC will make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, lon, zonal wind in the 500-hPa layer, westward extension index of the subtropical high, and area index of the subtropical high	106/150 = 0.710

TABLE 1. (*Continued*)

Rule	Attributes	Accuracy
If lat ≤ 27°, inten_IndexSTH > 5880.175, W_Westerly ≤ 477.463, lon > 130°, west_extSTH ≤ 141, uwnd_200 ≤ 13.547, Monsoon_WF ≤ 0.289, west_extSTH ≤ 130, uwnd_400 > −0.243, and area_IndexSTH ≤ 416, then the TC will make landfall along the Chinese coast	Lat, intensity index of the subtropical high, westerly index, lon, zonal wind in the 500-hPa layer, westward extension index of the subtropical high, and area index of the subtropical high	129/234 = 0.551

points derived from 222 TC tracks, 53 of which are landfalling TC tracks and the rest are nonlandfalling (Fig. 1). As shown in Fig. 1, the landfalling TC tracks consist of two parts: prior to (thick solid curve) and after (dashed curve) landfall. The solid circles are prior-to-landfall observations of landfalling TCs. It should be noted that only the prior-to-landfall observations of landfall TCs are employed in the following analyses because the observations after landfall exert little influence on TC landfall. Therefore, the tracks after landfall are depicted by dashed curves and the observations after landfall are omitted in Fig. 1. The thin curves and hollow circles denote nonlandfalling TC tracks and observations, respectively. The landfalling TC tracks contain 1150 observations and nonlandfalling TC tracks contain 5770 observations. It should be noted that observations of landfalling TC tracks prior to landfall are labeled class 1 and nonlandfalling TC tracks are labeled class 0. Because of the uneven number of TC samples in the two classes, resampling methods are used to ensure a more balanced class distribution.

The meteorological variables (e.g., wind fields and geopotential heights in different atmospheric layers) are the same as those described in Part I. It should be noted that factors that affect TC recurvature also play an important role in affecting TC landfalling. The C4.5 algorithm is again employed to discover rules governing TC landfalling. Details about the study area, data, and methodology can be found in Part I and are not repeated here.

3. Results analysis and interpretation

a. Results

The total classification accuracy of the decision tree constructed by C4.5 (see Fig. 2) is 81.3287% by 10-fold cross validation. Among all the variables listed in Table 1, selected by the C4.5 algorithm to build the decision tree, are the latitude and longitude of the TC center (lat and lon), the intensity index and area index of the subtropical high (inten_IndexSTH and area_IndexSTH), the average zonal wind within a 6°–8° radius in the 400-hPa layer (uwnd_400), the westerly index (W_Westerly), the westward extension index of the subtropical high (west_extSTH), and the monsoon index proposed by Wang and Fan (1999) (Monsoon_WF). In Fig. 2, a 1 means landfall and a 0 means nonlandfall. The rectangles in Fig. 2 are leaf nodes whereas ellipses or circles show the parent nodes. A path from the root node to the leaf node represents a rule, which can provide references for TC-landfall prediction. The tree contains 14 rules governing TC landfall in the SCS and WNP basins (see Fig. 1). Taking the leaf node 0(1269.0/85.0) as an example, the 0 before the parenthesis means the TC will not make landfall along the Chinese coast and 1269.0 and 85.0 indicate that, among the total of 1269 samples of the leaf node, there are 1184 (i.e., 1269 − 85) nonlandfalling samples and 85 landfalling samples, respectively. Each rule can be justified by meteorological and TC theories. In Table 1, the first column states the unraveled rules. The second column depicts the attributes involved in each rule. The third column in Table 1 shows the classification accuracy of each rule. This accuracy is calculated by dividing the number of correctly classified samples by the total number of samples in the leaf node.

b. Verification and interpretation

The results unraveled by the data-mining algorithm should be verified and interpreted according to our knowledge of TC movement and atmospheric circulation. Verification and interpretation are essential

TABLE 2. TC tracks used for verification.

TC	Landfall	Time of genesis
Omais	No	1200 UTC 22 Mar 2010
Conson	Yes	1200 UTC 11 Jul 2010
Chanthu	Yes	0600 UTC 17 Jul 2010
Dianmu	No	0000 UTC 7 Aug 2010
Mindulle	No	0000 UTC 22 Aug 2010
Lionrock	Yes	1800 UTC 27 Aug 2010
Kompasu	No	1200 UTC 28 Aug 2010
Namtheun	No	0600 UTC 29 Aug 2010
Malou	No	1200 UTC 11 Sep 2010
Meranti	Yes	0000 UTC 7 Sep 2010
Fanapi	Yes	1800 UTC 14 Sep 2010
Malakas	No	0600 UTC 20 Sep 2010
Megi	Yes	0000 UTC 13 Oct 2010
Chaba	No	1800 UTC 23 Oct 2010

TABLE 3. Confusion matrix of the verification.

	Landfall (prediction)	Nonlandfall (prediction)	Accuracy (%)
Landfall (obs)	109	14	88.618
Nonlandfall (obs)	45	158	77.833
Overall accuracy			81.902

components of the data-mining process (Han and Kamber 2006). In addition to cross validation in the training and classification step, the decision tree in Fig. 2 is further verified by the JMA–RSMC TC best-track data in 2010 for its validity. In what follows, we give the interpretation of the classification tree and the associated rules derived from the tree based on the theories of TC movement and landfall.

Of the 14 TCs that occurred in 2010, 6 made landfall over the Chinese coast. The 14 TC tracks with 326 TC observations are used to test the decision tree. Relevant information about the 14 TCs in 2010 is given in Table 2. The confusion matrix is shown in Table 3. Of the 123 landfalling and 213 nonlandfalling TC observations, 109 and 158 are correctly classified, respectively. The classification accuracy is 81.9018% (Table 3), very similar to the 81.3287% obtained in the training set. Therefore, the rules derived from the classification tree rather accurately characterize TC landfall over the Chinese coast and can be employed to predict TC landfall in the region.

The interpretation of the classification tree governing TC landfall resembles that of the classification tree governing TC recurvature, discussed in Part I. The key is to consider the composite steering flow (e.g., deep-layer mean winds) (Holland 1983, 1993; Velden and Leslie 1991), the 500-hPa geopotential height (depicting the subtropical high) of the samples in each leaf node, and the distance between the TC center and the Chinese coast.

The rule derived from node 0(1269.0/85.0) is simple, but the classification accuracy is the second highest among all of the rules. This rule is stated as follows: If a TC moves to the north of 27°N, then the TC will not make landfall over the Chinese coast. A possible explanation for this is that latitudes north of 27°N are in the midlatitude area with a strong westerly wind. When a TC moves into this region, it will be steered by the westerly wind to move away from the Chinese coast. The composite deep-layer mean winds (i.e., averaging from the 850–300-hPa layer) as well as the 500-hPa geopotential height of all the samples in this leaf node, overlap,

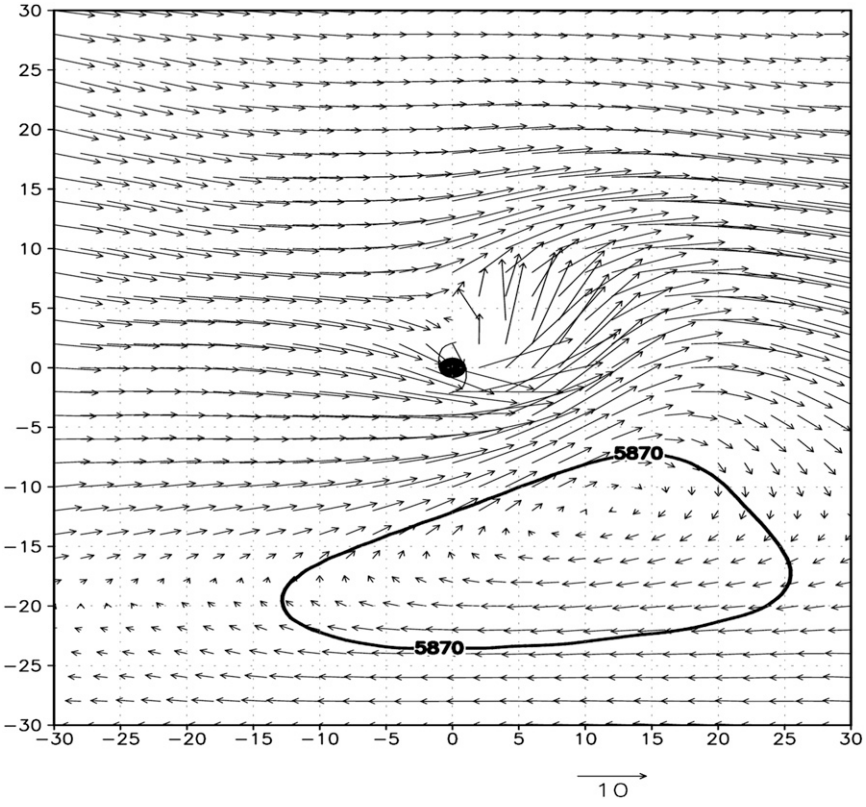

FIG. 3. Composite deep-layer mean winds and geopotential height in the 500-hPa layer of samples in leaf node 0(1269.0/85.0). The thick contours indicate the center of the subtropical high.

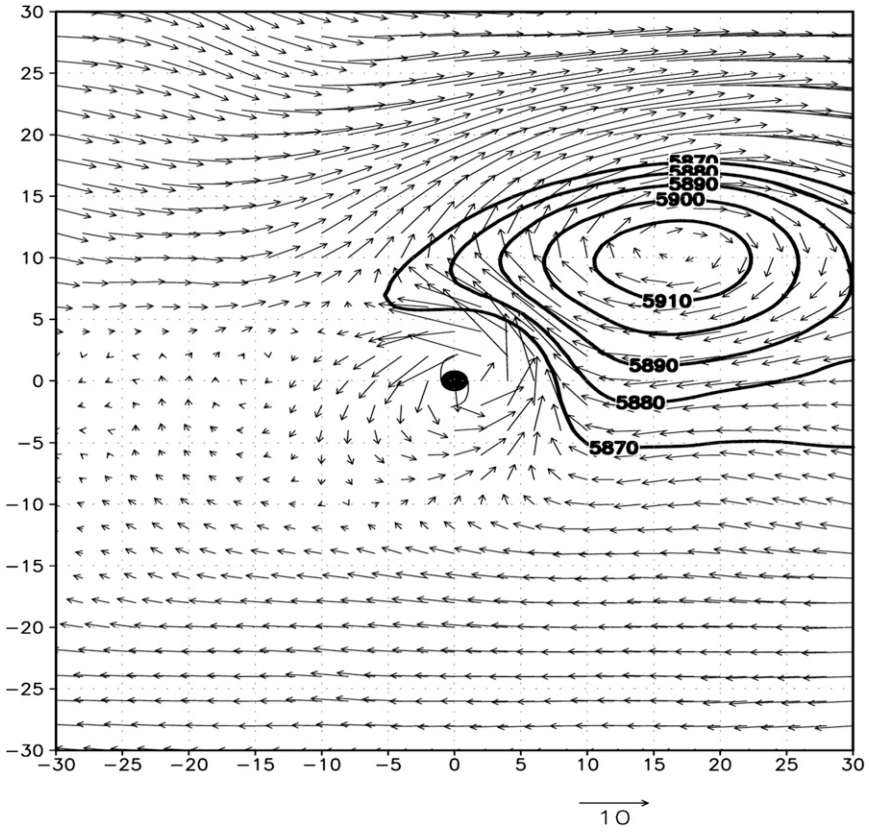

FIG. 4. As in Fig. 3, but for leaf node 0(234.0/105.0).

according to Fig. 3. The center of the subtropical high is depicted by the contour of Z_{500} = 5870 gpm (Z_{500} represents the 500-hPa geopotential height). Figure 3 illustrates that the wind fields surrounding the TC are eastward and that the TC is situated to the north of the subtropical high (depicted by the 5870-gpm contour). Therefore, the TC tends to be steered by the eastward flow to move far away from the Chinese coast. Latitude 27°N is a key element for determining TC landfall along the Chinese coast, since it is the first attribute selected by the C4.5 algorithm to build the classification tree. This latitude can thus be used as a crucial factor for forecasting TC landfall along the Chinese coast.

The rule constructed from leaf node 0(234.0/105.0), which has the lowest classification accuracy, can be stated as follows: If a TC moves to the south of 27°N, the intensity index of the subtropical high is >5880.175 gpm, the westerly index is <477.463, the longitude of the TC center is not to the west of 130°E, the western ridge of the subtropical high shifts to the west of 141°E, the 200-hPa zonal wind is <13.547 m s^{-1}, the monsoon index is <0.289 (strong monsoon), the western ridge of the subtropical high is to the west of 130°E, the average 400-hPa zonal wind is >-0.243 m s^{-1}, and the area_Index of the subtropical high is <416, then the TC will not make landfall across the Chinese coast. The western part of the subtropical high moves to the north of the TC center by about 5°–10°. The steering current caused by the subtropical high leads the TC to move westward for a certain distance, followed by turning or recurvature after the TC moves beyond the subtropical high (see Fig. 4). The tendency for TCs to make landfall is not remarkable. This is in accordance with the classification accuracy of 0.551. It is noted that this node has the lowest classification accuracy because 105 landfall TC samples are classified as nonlandfalling. All of the TC samples prior to landfall are taken into consideration. However, whether a TC will make landfall or not is not significantly influenced by the large-scale circulation and other local circulation patterns during its early life span. TC samples, which are incorrectly classified, are in the early part of their life span. Therefore, the large-scale circulation and circulation patterns surrounding TCs suggest nonlandfalling TCs. However, as the TCs traverse the ocean they tend to make landfall over the Chinese coast.

The rule derived from leaf node 0(974.0/155.0) can be stated as follows: If a TC is situated to the south of 27°N and the intensity index of the subtropical high is

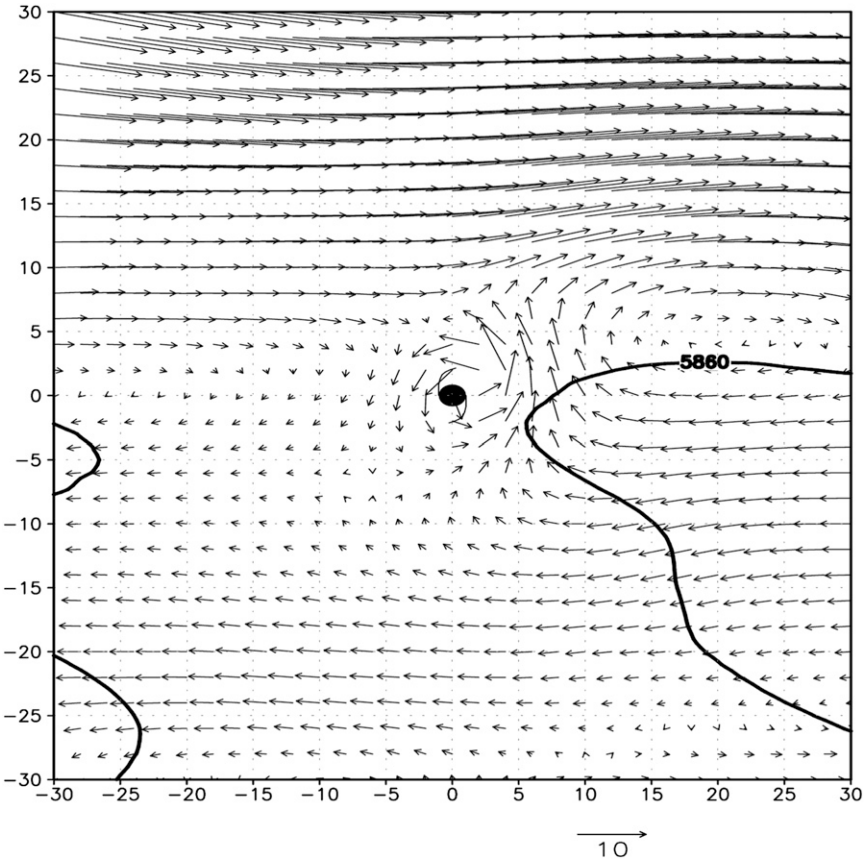

FIG. 5. As in Fig. 3, but for leaf node 0 (974.0/155.0).

<5880.175 gpm, then the TC will not make landfall along the Chinese coast. This rule has an accuracy of 0.841. Based on this rule, the intensity of the subtropical high is relatively weak (inten_indexSTH ≤ 5880.175 gpm). The subtropical high corresponding to this rule is too weak to draw the 5870-gpm contour. Therefore, only the 5860-gpm contour of the geopotential height in the 500-hPa layer is plotted in Fig. 5. Furthermore, the average deep-layer mean wind fields associated with this rule are northward and northeastward. The TC will be steered by the northward and northeastward flow to recurve (see Fig. 5). Therefore, TCs under this rule will not make landfall along the Chinese coast.

The rule derived from leaf node 0(246.0/18.0) says the following: If a TC is located to the south of 27°N, the intensity index of the subtropical high is >5880.175 gpm, the westerly index is >477.463 (strong westerly), and the TC center is located to the east of 130°E, then the TC will not make landfall along the Chinese coast. From this statement, this rule involves the subtropical high and westerlies. The intensity index of the subtropical high indicates the average 500-hPa geopotential height of grids with a geopotential height larger than 5870 gpm in the region (10°–60°N and 100°E–180°). This rule also reveals that the interactions between the relatively strong subtropical high and strong westerlies eventually cause TCs to recurve without making landfall along the Chinese coast when they are located relatively far away from it (>130°E). It is shown in Fig. 6 that the subtropical high lies to the east of the TC center. Although the steering flow may drive the TC westward for some distance, it has an appreciable tendency to recurve. Since the steering flow and TC's position are far from the Chinese coast, the TC will not make landfall across this coast subsequently.

The rule obtained from leaf node 1(157.0/64.0) can be described as follows: If a TC is located to the south of 27°N, the intensity index of the subtropical high is >5880.175 gpm, the westerly index is >477.463 (strong westerly), the longitude of the TC center is to the west of 130°E, and the area index of the subtropical high is >105, then the TC will make landfall along the Chinese coast. This rule suggests that when the TC moves closer to the Chinese coast, the strong and large-area subtropical high depicted by the intensity index and area index of the subtropical high drives the TC to make landfall along

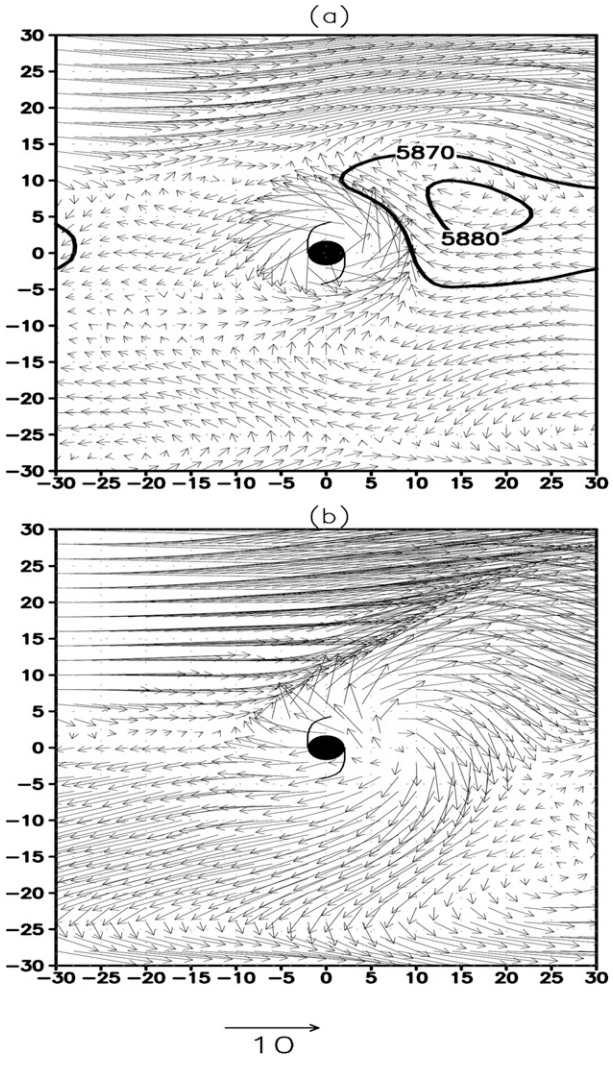

FIG. 6. (a) Composite deep-layer mean winds and 500-hPa geopotential height and (b) composite 200-hPa wind fields of the samples in leaf node 0(246.0/18.0). The thick contours indicate the center of the subtropical high.

the Chinese coast. Figure 7 shows that the center of the subtropical high lies to the north of the TC. Therefore, the prevailing easterly along the southern part of the subtropical high will eventually lead the TC to make landfall over the Chinese coast. The composites in Fig. 7 are consistent with what the rule indicates. It should be noted that the classification accuracy of this rule is only 0.592, and 64 nonlandfalling samples are classified as landfall samples. A portion of the nonlandfalling TC samples are located equatorward of the Chinese coast. These TC samples will not make landfall over the Chinese coast because the strong easterly steering flow leads the TC to move westward. In addition, some of the nonlandfalling TC samples experience recurvature close to the Chinese coast. Therefore, despite large-scale circulation suggesting TC recurvature in this rule, the TC samples located equatorward of the Chinese coast and the recurving ones are responsible for the error rate in classification.

The rule obtained from leaf node 1(2656.0/451.0) is stated as follows: If a TC moves to the south of 27°N, the intensity index of the subtropical high is >5880.175 gpm, the westerly index is <477.463, and the longitude of the TC center is to the west of 130°E, then the TC will make landfall across the Chinese coast. This rule means that the subtropical high is strong in terms of its intensity index, and the westerly is weak near the TC. It is noted that the longitude in this rule is to the west of 130°E. The wind fields surrounding the TC are plotted in Fig. 8. The deep-layer mean wind is easterly, which is consistent with the 200-hPa composite wind fields (Fig. 8). This indicates that the easterly prevails from lower to higher levels. Therefore, the easterly along the southern part of the subtropical high will eventually lead the TC to make landfall along the Chinese coast.

The rule formed by the path from the root node to leaf node 1(825.0/196.0) is as follows: If a TC is located to the south of 27°N, the intensity index of the subtropical high is >5880.175 gpm, the westerly index is <477.463, the longitude of the TC center is to the east of 130°E, the western ridge of the subtropical high shifts to the west of 141°E, the 200-hPa zonal wind is <13.547 m s^{-1}, and the monsoon index is >0.289 (weak monsoon), then the TC will make landfall along the Chinese coast. This rule shows that the mean intensity of the subtropical high is strong, the westerly is weak near the TC, and a monsoon with a southern and southwestern component is weak. Figure 9 illustrates that the easterly along the southern part of the subtropical high will eventually lead the TC to make landfall along the Chinese coast. This rule integrates the subtropical high, the monsoon systems, the midlatitude westerlies, and the outer radius environmental wind fields of TCs in the 200-hPa layer. It is fully consistent with existing findings on TC movement (Harr and Elsberry 1991, 1995a,b; Chen et al. 2009).

The rule derived from leaf node 0(271.0/11.0) says the following: If a TC moves to the south of 27°N, the intensity index of the subtropical high is >5880.175 gpm, the westerly index is <477.463, the longitude of the TC center is to the east of 130°E, the western ridge of the subtropical high is to the east of 141°E, and the westerly index is >233.04, then the TC will not make landfall across the Chinese coast. This rule has the highest classification accuracy among all 14 of the rules. This rule indicates that the subtropical high ridge shifts to the east of 141°E and the westerly index varies from 233.04 to 477.463. In addition, the intensity index (>5880.175 gpm) represents a strengthened subtropical high. As shown in

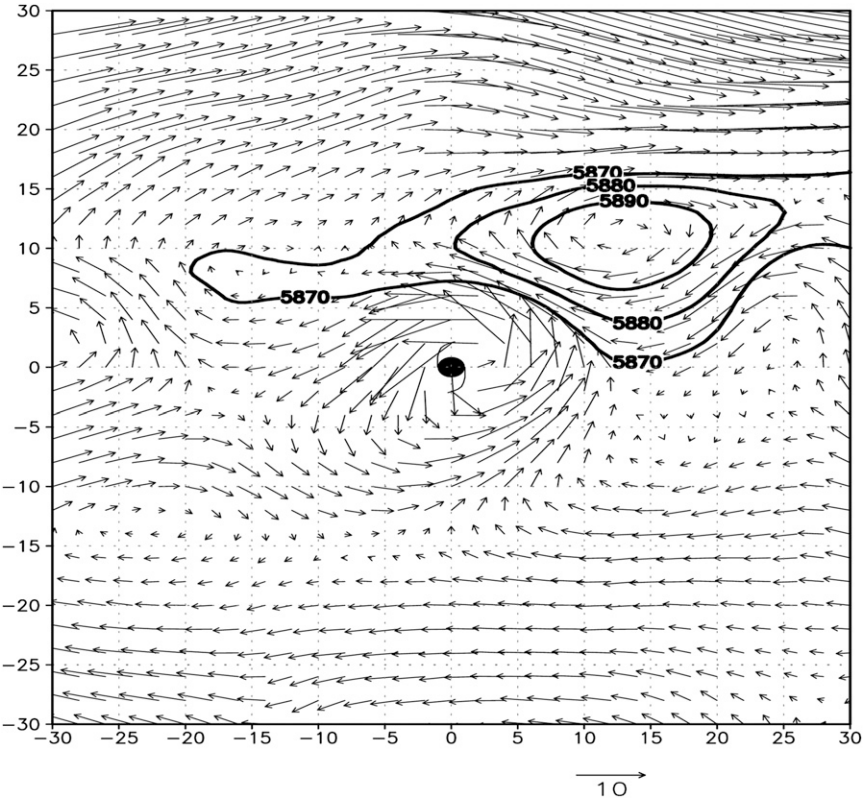

FIG. 7. As in Fig. 6, but for leaf node 1(157.0/64.0). The plot is relative to the center of the composite TC at coordinates (0, 0) with the typhoon symbol.

Fig. 10, the subtropical high retreats to the east of the TC center in spite of the relatively high intensity index; and the westerly begins to exert its influence on the TC center. The southern component of the western part of the subtropical high steers the TC northward and northeastward, and it subsequently recurves without making landfall over the Chinese coast. Therefore, the composite deep-layer mean wind and the subtropical high are fully in line with what is indicated by the rule.

The rule obtained from leaf node 1(150.0/56.0) can be described as follows: If a TC moves to the south of 27°N, the intensity index of the subtropical high is >5880.175 gpm, the westerly index is <477.463, and the longitude of the TC center is to the east of 130°E, the western ridge of the subtropical high is to the east of 141°E, and the westerly index is <233.04, then the TC will make landfall along the Chinese coast. This rule indicates that the subtropical high ridge retreats to the east of 141°E, and the westerly index is as low as 233.04. As described in Fig. 11, the subtropical high combined with another high pressure system is located to the north of the TC. The prevailing easterly surrounding the TC steers the TC westward, and it eventually makes landfall along the Chinese coast. It can be observed that this rule differs from the rule induced by leaf node 0(271.0/11.0) on the westerly index. The westerly index is smaller than that of leaf node 0(271.0/11.0). It can be proven by the prevailing eastern component (see Fig. 11).

In summary, the longitude at 130°E is frequently chosen as the splitting longitude by the C4.5 algorithm when building a classification tree for TC landfall and recurvature in both parts of this series. It indicates that this longitude (130°E) is crucial for TC recurvature and landfall. In operations, particular attention should thus be paid to this longitude, because it plays an important role in differentiating whether or not the TC will recurve and make landfall along the Chinese coast based on results presented in this two-part series. The latitude at 27°N is the first chosen attribute of the classification tree (see Fig. 2). This latitude is also a key indicator of TC landfall. Based on our findings, if a TC moves to the north of 27°N, it probably will not make landfall along the Chinese coast. The 200- and 400-hPa zonal winds are also selected by the C4.5 algorithm. Nevertheless, none of the meridional winds is selected by the C4.5 algorithm in this study. As demonstrated by various investigations, the zonal winds surrounding TC centers have more significant impacts on TC movement than the meridional

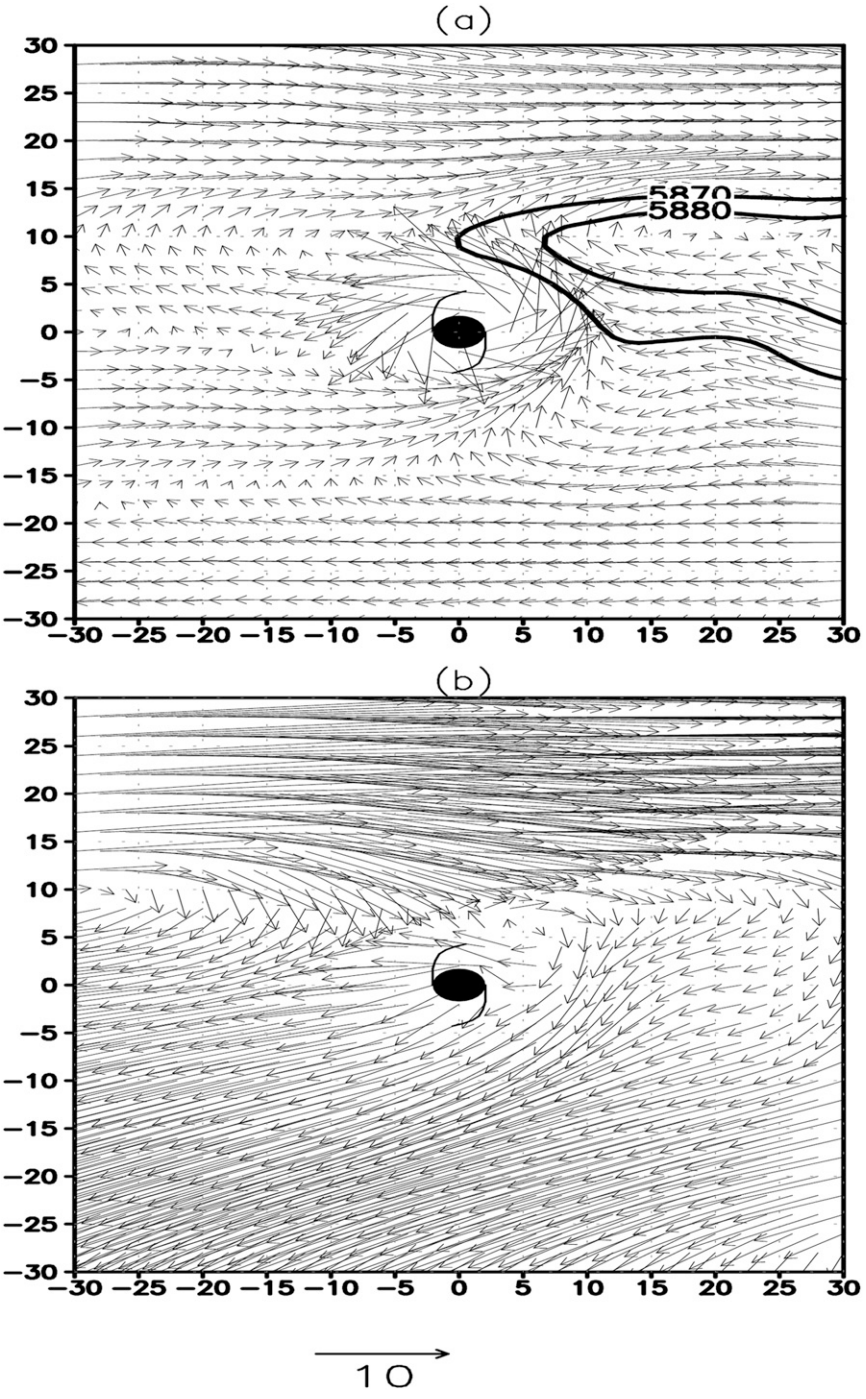

FIG. 8. As in Fig. 6, but for leaf node 1(2656.0/451.0).

winds (e.g., Chen et al. 2009; Harr and Elsberry 1991, 1995a,b). In addition to the attributes discussed above, other attributes selected by the C4.5 algorithm to build the classification tree are the indices for measuring large-scale circulation (e.g., the subtropical high, the monsoon systems, the midlatitude westerlies), which include the westerly index, indices of the subtropical high (the intensity index, area index, and west extension index), and the monsoon index. These indices are used here to quantify the status of the large-scale circulation. The results are in good agreement with previous research efforts and their findings about the modulation of large-scale

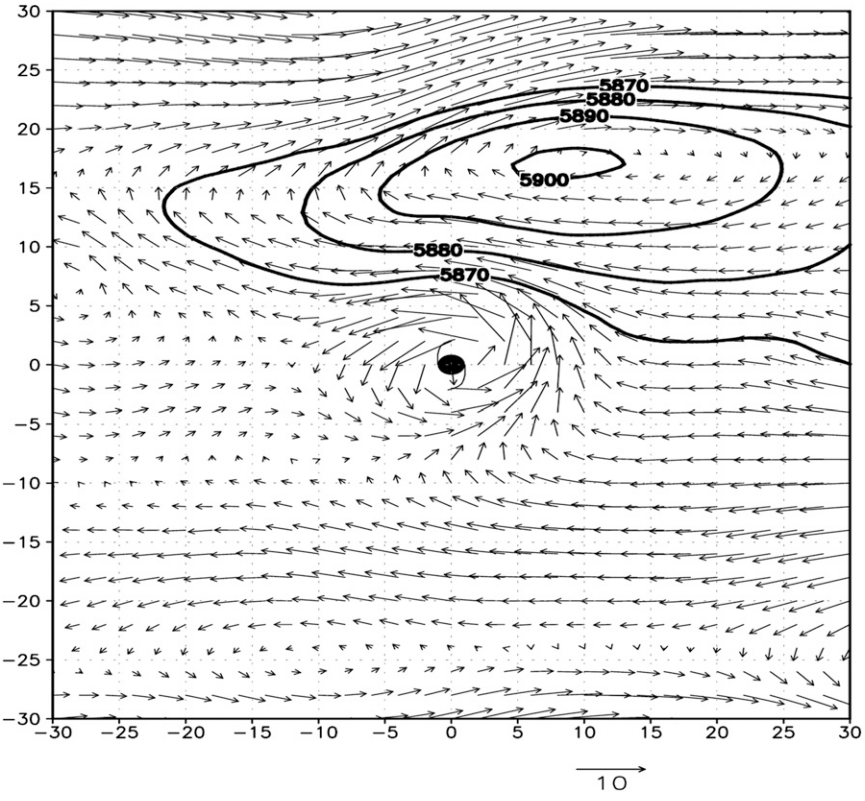

FIG. 9. Composite deep-layer mean winds and 500-hPa geopotential height of the samples in leaf node 1(825.0/196.0). The plot is relative to the center of the composite TC at coordinates (0, 0) with the typhoon symbol. The thick contours indicate the center of the subtropical high.

circulation on TC movement (e.g., Harr and Elsberry 1991, 1995a,b; Elsberry 2004; Chen et al. 2009).

c. *Case study*

To demonstrate the potential of this decision tree for forecasting TC landfall, we use Typhoon Conson as a case study. After developing out of a tropical disturbance east of the Philippines at 1200 UTC 11 July 2010, Conson quickly developed as it moved nearly due west. It intensified into a severe tropical storm by 12 July. By 16 July, it attained typhoon status as it was close to the southern island of Hainan, China. After making landfall over China at its peak intensity with sustained winds of 36 m s^{-1}, it weakened in the Gulf of Tonkin as a result of less favorable conditions. It eventually made landfall over Vietnam on 17 July and dissipated during the following day.

We use the decision tree (Fig. 2) to forecast whether this TC will make landfall over China or not, using the variables quantifying large-scale circulation, wind fields surrounding TCs, and variables characterizing TCs. We start from the genesis (13.9°N, 132.6°E) of Conson at 1200 UTC 11 July 2010. Based on the decision tree (Fig. 2), the first splitting variable is lat (with a splitting value of 27°N), which is the latitude at the TC center. The latitude of the genesis of Conson is 13.9°N, which is smaller than 27°N. The genesis follows the left branch to inten_IndexSTH, which is the intensity index of the subtropical high. The value of inten_IndexSTH for TC genesis is 5915.131 gpm, which is larger than the splitting value of 5880.175 gpm. Therefore, the genesis follows the right branch to W_Westerly, which is the index of the midlatitude westerlies. This index (223.248 gpm) is smaller than the splitting value (477.463). The genesis follows the leaf branch to the longitude of the TC center (lon). Since the longitude of the TC genesis is 132.6°E, the genesis then follows the right branch to west_extSTH. Since the value for this variable (121°E) is smaller than 141°E (the splitting value), the path goes from the left branch to uwnd_200. The genesis follows the left branch to Monsoon_WF with a splitting value of 0.289 because of the fact that the value of uwnd_200 (10.09 m s^{-1}) is smaller than the splitting value of 13.547 m s^{-1}. The value of Monsoon_WF at TC genesis is 11.698. Subsequently, the path of the genesis terminates at the leaf node [1(825/196)]. This leaf node suggests that this TC

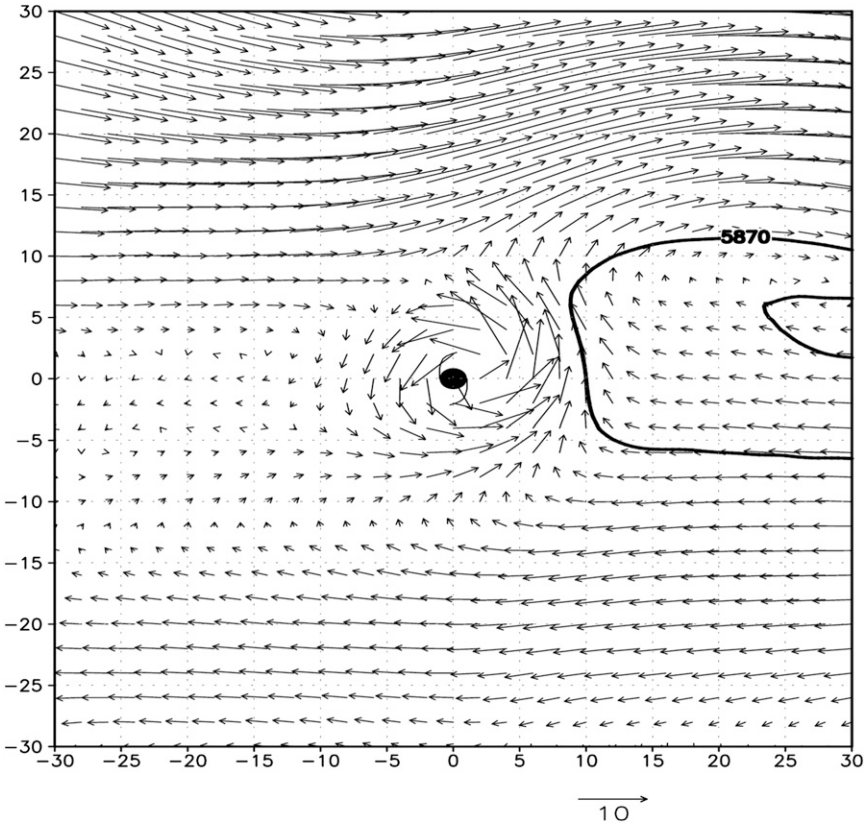

FIG. 10. As in Fig. 6, but for leaf node 0(271.0/11.0).

will make landfall over the Chinese coast. At 1200 UTC 11 July 2010, Conson moved to 14.2°N, 131.8°E. We infer again whether it will make landfall over China at this observation. This observation first follows the left branch to inten_IndexSTH because the latitude is smaller than the splitting value of lat (27°N). The inten_IndexSTH value (5910.133) means that this observation follows the left branch to W_Westerly. After we have compared the values of other variables with their splitting values, this observation follows the same path as the genesis to the left node [1(825/196)]. Therefore, we can infer at 1200 UTC 11 July 2010 that Conson will make landfall over China. For the subsequent TC observations after genesis, we follow the same process to infer whether Conson will make landfall or not. We can infer at every TC stage that Conson will make landfall over the Chinese coast. The decision tree can therefore provide useful information and sound logic for forecasting TC landfall.

4. Conclusions and discussion

In part of our series of papers, we have investigated TC landfall in the SCS and WNP basins through the data-mining approach. Based on the historical TC archives, the C4.5 algorithm, a classic tree algorithm for classification, has been employed to quantitatively evaluate the rules governing TC landfall.

A classification tree (Fig. 2), with 14 leaf nodes, has been built. The path from the root node to each leaf node forms a rule. Fourteen rules governing TC landfall across the Chinese coast have been unraveled with respect to the selected attributes having potential influence on TC landfall. The rules are derived by the attributes and splitting values. Based on the classification tree, split values, such as 27°N latitude, 130°E longitude, 141°E in the west extension index, and 0.289 in the monsoon index (Wang and Fan 1999), have been shown to be useful for TC forecasting. The rules have been justified from the perspective of meteorology and our knowledge of TC movement and recurvature (e.g., deep-layer mean winds and large-scale circulation). The research findings are based upon existing theories on TC movement and landfall (Harr and Elsberry 1991, 1995a,b; Chen et al. 2009). Both the unraveled rules and the associated splitting values can provide useful references for TC landfall prediction.

New local conditions (e.g., latitude and longitude) have been derived from the decision tree. In addition,

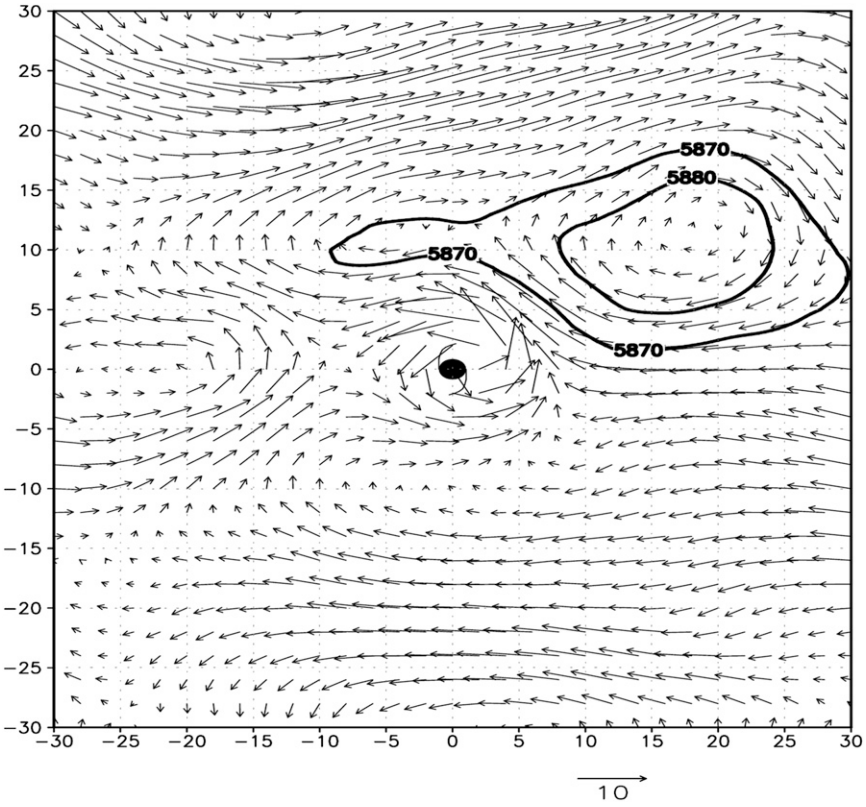

FIG. 11. As in Fig. 6, but for leaf node 1(150.0/56.0).

some rules indicate that the stage (developing, mature, and weakening) of a TC (Gray 1997, 1998; Lee et al. 1989) should be taken into account when forecasting TC landfall. The analyses of the effects of local conditions and TC stages on TC landfalls carried out in this paper are parallel to those in Part I.

Analysis results indicate that the classification accuracy is 81.3287% when latitude and longitude are taken into consideration. However, when these two parameters are excluded from the analysis, the accuracy decreases to 74.2446%. Such results show that local conditions play a central role in the forecasting of TC landfall. As discussed in Part I, TCs go through the developing, mature, and weakening stages. For assessment, we divide the TC samples into three basic groups: developing stage, mature stage, and weakening stage. The classification accuracies with respect to these stages are 81.3287%, 92.6216%, and 94.3314%, respectively. This suggests that forecasting TC landfall becomes more accurate when TCs move from the time they first form to the time they weaken.

This study is limited by several factors. First, the basic framework for the analysis of TC landfalling is based on the steering flow concept. In addition to the steering flow, thermodynamic processes such as diabatic heating affect TC motion, including TC recurvature (Li and Chan 1999; Wu and Wang 2000; Chan et al. 2002). Such processes may bias the influence of steering flow on the direction of a TC track, and this will further affect TC landfall. Second, our study focuses on the classification of landfalling and nonlandfalling TCs. Although the decision tree built in this study can provide useful references for inferring whether a TC will make landfall over the Chinese coast, our research findings cannot be used directly to forecast, for example, exactly when and where a TC will make landfall over the Chinese coast.

In future research, data mining and dynamic modeling may complement each other in the study of TC landfall. The research findings of this study, for example, can be used to improve the dynamic models by injecting new mechanisms, fine-tuning the parameters, and verifying their outputs. The variability of seasonal TC landfall frequency is attributed to the modulation of ENSO, MJO, QBO, AMO, and PDO (Liu and Chan 2003; Goh and Chan 2010). The indices of these oscillations may exert a certain degree of influence on TC landfall. These indices can also be taken as potential attributes of TC landfall in future studies.

To recapitulate, we have examined in this two-part series of papers TC recurvature and landfall using the C4.5 algorithm. Two classification trees stipulating

various situations under which a TC will recurve or make landfall have been unraveled from 2000–09 data and verified using 2010 TC data. This study is, however, limited by data availability (e.g., only 6-h intervals are available), the lack of data related to internal TC dynamics, and the limitations of the spatial resolution of TC-related meteorological data. In future studies, data with shorter time intervals (e.g., 3 h, 1 h, 0.5 h, and 15 min) and 3D radar data should be employed to unravel rules relating to TC recurvature and landfall. The stability of the rule sets against varying time intervals can then be examined to further improve the consistency and reliability of this approach.

Acknowledgments. This research was jointly supported by the Geographical Modeling and Geocomputation Program under the Focused Investment Scheme of the Chinese University of Hong Kong, the National Natural Science Foundation of China (Grant 41201045), and the 973 project (2012CB955800) of the Ministry of Science and Technology of China. Work of the third author (JCLC) was supported by the General Research Fund of the Research Grants Council of the HKSAR government with Grant CityU 100210.

REFERENCES

Bender, M., R. Tuleya, and Y. Kurihara, 1985: A numerical study of the effect of a mountain range on a landfalling tropical cyclone. *Mon. Wea. Rev.,* **113,** 567–583.

Bhowmik, S. K. R., S. D. Kotal, and S. R. Kalsi, 2005: An empirical model for predicting the decay of tropical cyclone wind speed after landfall over the Indian region. *J. Appl. Meteor.,* **44,** 179–185.

Brand, S., and J. Blelloch, 1974: Changes in the characteristics of typhoons crossing the island of Taiwan. *Mon. Wea. Rev.,* **102,** 708–713.

Brettschneider, B., 2008: Climatological hurricane landfall probability for the United States. *J. Appl. Meteor. Climatol.,* **47,** 704–716.

Brunt, A., 1968: Space–time relations of cyclone rainfall in the northeast Australian region. *Civil Eng. Trans.,* **10,** 40–46.

Chan, J. C. L., and X. D. Liang, 2003: Convective asymmetries associated with tropical cyclone landfall. Part I: *f*-plane simulations. *J. Atmos. Sci.,* **60,** 1560–1576.

——, F. M. F. Ko, and Y. M. Lei, 2002: Relationship between potential vorticity tendency and tropical cyclone motion. *J. Atmos. Sci.,* **59,** 1317–1336.

Chang, S. W. J., 1982: The orographic effects induced by an island mountain range on propagating tropical cyclones. *Mon. Wea. Rev.,* **110,** 1255–1270.

Chen, T. C., S. Y. Wang, M. C. Yen, and A. J. Clark, 2009: Impact of the intraseasonal variability of the western North Pacific large-scale circulation on tropical cyclone tracks. *Wea. Forecasting,* **24,** 646–666.

Elsberry, R. L., 2004: Monsoon-related tropical cyclones in East Asia. *East Asian Monsoon,* C.-P. Chang, Ed., Series on Meteorology of East Asia, Vol. 2, World Scientific, 463–498.

Elsner, J. B., and K. B. Liu, 2003: Examining the ENSO–typhoon hypothesis. *Climate Res.,* **25,** 43–54.

Fayyad, U., 1997: Knowledge discovery in databases: An overview. *Inductive Logic Programming,* N. Lavrač and S. Džeroski, Eds., Springer, 1–16.

——, and P. Stolorz, 1997: Data mining and KDD: Promise and challenges. *Future Gener. Comput. Syst.,* **13,** 99–115.

——, G. Piatetsky-Shapiro, and P. Smyth, 1996: From data mining to knowledge discovery in databases. *AI Mag.,* **17,** 37–54.

Goh, A. Z.-C., and J. C. L. Chan, 2010: An improved statistical scheme for the prediction of tropical cyclones making landfall in south China. *Wea. Forecasting,* **25,** 587–593.

Gray, W. M., 1979: Hurricanes: Their formation, structure and likely role in the tropical circulation. *Meteorology over the Tropical Oceans,* D. B. Shaw, Ed., Royal Meteorological Society, 155–218.

——, 1998: The formation of tropical cyclones *Meteor. Atmos. Phys.,* **67,** 37–69.

Hamuro, M., and Coauthors, 1969: Precipitation bands of Typhoon Vera in 1959 (Part 1). *J. Meteor. Soc. Japan,* **47,** 298–308.

Han, J., and M. Kamber, 2006: *Data Mining: Concepts and Techniques.* 2nd ed. Morgan Kaufmann, 770 pp.

Harr, P. A., and R. L. Elsberry, 1991: Tropical cyclone track characteristics as a function of large-scale circulation anomalies. *Mon. Wea. Rev.,* **119,** 1448–1468.

——, and ——, 1995a: Large-scale circulation variability over the tropical western North Pacific. Part I: Spatial patterns and tropical cyclone characteristics. *Mon. Wea. Rev.,* **123,** 1225–1246.

——, and ——, 1995b: Large-scale circulation variability over the tropical western North Pacific. Part II: Persistence and transition characteristics. *Mon. Wea. Rev.,* **123,** 1247–1268.

Holland, G. J., 1983: Tropical cyclone motion—Environmental interaction plus a beta-effect. *J. Atmos. Sci.,* **40,** 328–342.

——, 1993: Tropical cyclone motion. *Global Guide to Tropical Cyclone Forecasting,* G. Holland, Ed., WMO, 1–46.

Lee, C., R. Edson, and W. M. Gray, 1989: Some large-scale characteristics associated with tropical cyclone development in the north Indian Ocean during FGGE. *Mon. Wea. Rev.,* **117,** 407–426.

Leung, Y., 2010: *Knowledge Discovery in Spatial Data*. Springer-Verlag, 360 pp.

Li, Y. S., and J. C. L. Chan, 1999: Momentum transports associated with tropical cyclone recurvature. *Mon. Wea. Rev.,* **127,** 1021–1037.

Liu, K. S., and J. C. L. Chan, 2003: Climatological characteristics and seasonal forecasting of tropical cyclones making landfall along the south China coast. *Mon. Wea. Rev.,* **131,** 1650–1662.

Marks, F., and L. Shay, 1998: Landfalling tropical cyclones: Forecast problems and associated research opportunities. *Bull. Amer. Meteor. Soc.,* **79,** 305–323.

Miller, H. J., and J. Han, 2001: Geographic data mining and knowledge discovery: An overview. *Geographic Data Mining and Knowledge Discovery,* H. J. Miller and J. Han, Eds., Taylor and Francis, 3–32.

Tuleya, R. E., and Y. Kurihara, 1978: A numerical simulation of the landfall of tropical cyclones. *J. Atmos. Sci.,* **35,** 242–257.

——, M. Bender, and Y. Kurihara, 1984: A simulation study of the landfall of tropical cyclones. *Mon. Wea. Rev.,* **112,** 124–136.

Velden, C. S., and L. M. Leslie, 1991: The basic relationship between tropical cyclone intensity and the depth of the environmental steering layer in the Australian region. *Wea. Forecasting,* **6,** 244–253.

Wang, B., and Z. Fan, 1999: Choice of South Asian summer monsoon indices. *Bull. Amer. Meteor. Soc.,* **80,** 629–638.

Wong, M. L. M., 2007: Modeling the effects of land–sea roughness contrast on tropical cyclone winds. *J. Atmos. Sci.,* **64,** 3249–3264.

——, and J. C. L. Chan, 2006: Tropical cyclone motion in response to land surface friction. *J. Atmos. Sci.,* **63,** 1324–1337.

——, ——, and W. Zhou, 2008: A simple empirical model for estimating the intensity change of tropical cyclones after landfall along the south China coast. *J. Appl. Meteor. Climatol.,* **47,** 326–338.

Wu, L., and B. Wang, 2000: A potential vorticity tendency diagnostic approach for tropical cyclone motion. *Mon. Wea. Rev.,* **128,** 1899–1911.

Zhang, W., Y. Leung, and J. C. L. Chan, 2013: The analysis of tropical cyclone tracks in the western North Pacific through data mining. Part I: Tropical cyclone recurvature. *J. Appl. Meteor. Climatol.,* **52,** 1394–1416.