The Discovery of Tropical Cyclone Dynamics in Western North Pacific through Data Mining

ZHANG, Wei

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
in
Geography and Resource Management

The Chinese University of Hong Kong
September 2011

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Thesis/Assessment Committee

Professor CHEN, David Yongqin (Chair)

Professor LEUNG, Yee (Thesis Supervisor)

Professor FUNG, Tung (Committee Member)

Professor LIU, Kam-Biu (External Examiner)

Professor YAN, Yuk Yee (External Examiner)

ABSTRACT

of thesis entitled:

The Discovery of Tropical Cyclone Dynamics in Western North Pacific through Data Mining

Submitted by ZHANG, Wei
for the degree of Doctor of Philosophy
at The Chinese University of Hong Kong
in September 2011

The booming population and property values of coastal areas, together with changing climatic conditions, mean that tropical cyclones (TCs) are coming to wreak increasing levels of damages on human beings. Data mining methods (e.g., C4.5 algorithm) are used to tease out hidden information on the patterns, regularities or structures of cyclonic behaviors including TC recurvature, TC landfall, post-landfall TC analysis, and intensity related to landfall and recurvature from the archived TC data in western North Pacific, including South China Sea. A web-based system is built for TC analysis and prediction to combine the outputs of data mining and dynamic models. The primary contributions of this thesis may be summarized as follows:

1. A decision tree representing TC recurvature with 18 leaf nodes is built through the C4.5 algorithm based on the potential variables that may influence TC recurvature. Eighteen rules are derived from the tree by following a pathway from the root node to each leaf node. All these rules can be explained from the meteorological perspective. The chosen variables and cutting values in the decision tree (e.g., 123 °E and 130 °E) identify the particular importance in estimating and predicting of controlling the TC recurvature.

2. A decision tree stipulating TC landfall with 14 leaf nodes is built by C4.5 algorithm selecting potential variables and cutting values to build the tree. Therefore, 14 rules are derived from the tree by tracking the path from the root node to each leaf node. The rules comprised by chosen variables and cutting values are interpreted by TC movement

theories. The classification result of the decision tree is verified according to 2010 TC best track data. The chosen variables and cutting values of the tree provide references for predicting TCs' landfall.

3. The analysis of the historical post-landfall tracks of TCs over China unravels three clusters using Finite Mixture Model (FMM)-based clustering. The clusters are interpreted by the concepts of geopotential height, water vapor supply and baroclinic energy from mid-latitude circulations. This form of cluster analysis for the first time identifies the clusters hidden in the TCs that made landfall over the Chinese coast.

4. By revisiting the intensity change associated with recurvature, the study determined that TCs are prone to peak their intensities near their recurvature. We also found that the recurving points of TCs peaking in intensity after recurvature tend to be located to the southeast of those of TCs peaking in intensity prior to recurvature. The recurving TCs sustain their maximum intensity for significantly longer than straight-moving TCs. The TCs tend to weaken significantly from landfall–24h to landfall+24h.

5. The study builds a novel web-based system for TC analysis and prediction. This system is web-based and has great potential for technology transfer. Its user-friendly interface offers a multi-criteria query system for tracking and visualizing TC movements on the basis of mined data and dynamic modeling. It can play a significant role in TC analysis and prediction for professionals and general users.

6. This thesis focuses on how large-scale circulation and meteorological variables surrounding TC centers influence TC motion. TCs actually interact with large-scale circulation. At the same time, the TC structures and internal convections also exert impacts on TC motions. With the development of advanced observing equipment (e.g., radar, satellites and other forms of reconnaissance), the variables related to internal TC dynamics can be collected to complement the analysis of large-scale circulation.

摘要

隨著氣候逐漸變化以及沿海地區人口以及財產的劇增，熱帶氣旋對人類所造成的損失日趨嚴重。本文應用數據挖掘方法（比如 C4.5 算法）從西北太平洋，包括南中国海的熱帶氣旋歷史數據庫中發現氣旋問題中的模式、規律性和結構。這些氣旋問題包括熱帶氣旋轉向、登陸、登陸後熱帶氣旋路徑的分析以及轉向和登陸熱帶氣旋的強度特徵。同時，本文構建了一個用於熱帶氣旋分析和預報的網絡系統。本文的主要貢獻总结如下：

1. 本文利用 C4.5 算法生成了一棵控制颱風轉向的擁有 18 個葉子節點的決策樹。18 條規則，從根節點到葉子節點的路徑，從該決策樹導出。所有這些規則可以通過氣象理論解釋。決策樹所選擇的變量以及分割值（例如東經 123 和 130 度）在控制熱帶氣旋轉向中具有特別意義。

2. C4.5 算法選擇潛在的變量和分割值，構建了一棵具有 14 個葉子節點，控制著熱帶氣旋登陸的決策樹。14 條規則，從根節點到葉子節點的路徑，從該決策樹導出。這些規則可以由熱帶氣旋移動理論解釋，決策樹分類的效果由 2010 年熱帶氣旋資料驗證。決策樹所導出的規則和分割值可以為熱帶氣旋登陸預報提供參考。

3. 基於 FMM 的聚類算法，本文從歷史的登陸中國颱風的登陸後路徑中發現了三類登陆路径。這三個類通過位勢高度，水汽輸送以及中緯度環流的斜壓能量傳輸進行解釋。登陸中國的歷史路徑中隱藏的類第一次通過聚類分析挖掘出來。

4. 本文分析了登陸熱帶氣旋的強度變化。熱帶氣旋趨向于在轉向時達到最大強度。轉向之後達到最大強度的熱帶氣旋的轉向點位於轉向前達到最大強度的熱帶氣旋的轉向點的偏東南方。轉向颱風維持最大強度的時間比直行颱風維持最大強度的時間要顯著地長。登陸熱帶氣旋的強度只有在登陆前一天至登陆后一天顯著降低。

5. 本文建立了一個基於網絡的新穎熱帶氣旋分析和預報系統。該系統既具有學術價值，同时具有技術轉移的巨大潛力。本系統拥有友好的用戶介面和強大的多標準查

詢系統，用於跟蹤和可視化熱帶氣旋的移動。同時，該系統結合了數據挖掘和動力模式的功能。它將為專業用戶以及公眾在熱帶氣旋的分析和預報中發揮重要作用。

6. 本文主要研究大尺度環流以及環繞熱帶氣旋中心的氣象因子對其路徑的影響。事實上，熱帶氣旋與大尺度環流相互作用。熱帶氣旋的結構和內部對流也影響熱帶氣旋的路徑。隨著雷達、衛星以及飛機等觀測手段的進步，熱帶氣旋內部動力結構的因子能夠補充大尺度環流的分析。

ACKNOWLEDGEMENT

I would like to express my deepest appreciation to my thesis supervisor, Professor Yee Leung, for his not only patient, but also inspiring supervisions, encouragements and supports throughout my PhD study. He is the giant on whose shoulder I stood and will stand. I have received efficient and strict training from him on doing research, starting from finding a research problem, coming up with the solutions, interpreting results, and finally writing up the technical papers/reports. What I have learnt from him is not just knowledge, but also the correct attitudes towards doing research, and the wisdom and way to become a scientist. I am always moved by his meticulous approach, curiosity and passion in research. What I have learnt from him is definitely an invaluable wealth of my whole life.

I am indebted to Prof. Yongqin David Chen and Prof. Tung Fung, my committee members for their helpful suggestions and discussions of my term presentations. Special thanks go to Professor Johnny Chan at the City University of Hong Kong for his helpful and inspiring advice and suggestions on my research in tropical cyclones.

I gratefully appreciate Prof. Leung Kwong-Sak and Prof. Wong Man Hon at the Department of Computer Science and Engineering for their valuable discussions and supports in my collaboration on the development of the web-based typhoon system. I am also greatly grateful to Prof. Wang Yuanfei at East China Normal University, Prof. Ngar-Cheung Gabriel Lau at Princeton University, Prof. Russell L. Elsberry at the Naval Postgraduate School of U.S.A., Prof. Long Chiu at George Mason University, Prof. Lam Chiu Ying, Prof. Lin Hui, Prof. Huang Bo, Prof. Chau Kwai Cheong and Prof. Lam Kin Che at The Chinese University of Hong Kong and Prof. Alex Lau at Hong Kong University of Science and Technology for their helpful discussions. I would like to thank Prof. Francis Chi Yung Tam at the City University of Hong Kong for his teaching when I was taking his atmospheric courses. I would like to thank Prof. Hans.-F Graf and Dr. Michael Herzog for their supervisions on atmospheric sciences during my four-month visit to the Atmospheric Center at University of Cambridge. Thanks also go to all faculty

members and staff in the department of geography and resource management at CUHK. I cannot imagine how difficult my study will be without their assistance and help.

I am greatly grateful to Ge Erjia for her encouragement and help throughout my study. I would also like to thank Gao Si and Chen Bin for their discussions, helps and suggestions during my PhD study. I am grateful to the inspiring discussions with Alex Hoffman, Paul Griffths, and Tobias Gerken in the group of atmospheric sciences at the University of Cambridge. The time I spent in The Chinese University of Hong Kong is full of happiness and joys, thanks go to the fellow students Li Rongrong, Zhou Yu, Wong Wai Fung, Li Yun, Xu Yong, Qin Jie, Li Lu, Lin Hui, Cao Kai, Wang Kai, Liu Chunlin, Guo Chunlan, Gao Xuehua, Shi Xiaoxin, Meng Lina, Luo Ming, Liu Lu, Liu Ye and other students in this department. I would also like to express my gratitude to the undergraduates who I have taught at GRM for their supports and appreciations of my teaching.

Last but not least, my heart-felt appreciation goes to my parents and my sister for their supports and understanding all along.

ZHANG Wei

June 2011

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AMO	Atlantic Multi-decadal Oscillation
ANOVA	Analysis of Variance
API	Application Programming Interface
ATCF	Automated Tropical Cyclone Forecasting System
ATCW	Australian Tropical Cyclone Workstation
CART	Classification and Regression Tree
CHAID	Chi-squared Automatic Interaction Detection
CISK	Convective Instability of the Second Kind
CLIPER	Climatology and PERsistence
DEM	Digital Elevation Model
EASM	East Asian Summer Monsoon
EM	Expectation-Maximum
ENSO	El Nino-Southern Oscillation
FMM	Finite Mixture Model
FNL	Final Analyses
GFDI	Geophysical Fluid Dynamics Interpolated
GFNI	Geophysical Fluid Dynamics Model-Navy (interpolated)
GFS	Global Forecasting System
GIS	Geographic Information Systems
GrADS	Grid Analysis and Display System
HKO	Hong Kong Observatory
HTML	Hyper Text Markup Language
JMA	Japan Meteorological Agency
JSP	Java Server Page
JTWC	Joint Typhoon Warning Center
KDD	Knowledge Discovery
LGEM	Logistic Growth Equation Model
LHR	Latent Heat Release

MCP	Minimum Central Pressure
MJO	Madden-Julian Oscillation
MM5	The Fifth-Generation NCAR / Penn State Mesoscale Model
MPI	Maximum Potential Intensity
MSW	Maximum Sustained Wind
NASA	National Aeronautics and Space Administration
NCC	National Climate Center in China
NCAR	National Center for Atmospheric Research
NCEP	National Center for Environmental Protection
NHC	National Hurricane Center
OI	Optimally Interpolated
OpenAMF	Open Action Message Format
PDO	Pacific Decadal Oscillation
QBO	Quasi-Biennial Oscillation
RSMC	Regional Specialized Meteorological Center Tokyo
SCS	South China Sea
SHIFOR	Statistical Hurricane Intensity Forecast model
SHIPS	Statistical Hurricane Intensity Prediction Scheme
SLHF	Surface Latent Heat Flux
SLP	Sea Level Pressure
SST	Sea Surface Temperature
STR	Subtropical Ridge
TC	Tropical Cyclone
TCAPS	Tropical Cyclone Analysis and Prediction System
TIPS	Typhoon Intensity Prediction Scheme
TMI	TRMM Microwave Imager
TMPA	TRMM Multisatellite Precipitation Analysis
TRMM	Tropical Rainfall Measuring Mission
USGS	United States Geological Survey
WISHE	Wind-Induced Surface Heat Exchange
WNP	western North Pacific

WNPSH	Western North Pacific Subtropical High
WWW	World Wide Web

CHAPTER 1: Introduction

1.1 Research Background

Losses caused by weather-related damage have been rising by 10% year on year (Cameron and Allen, 2005). In particular, the economic damage and disruption caused by tropical cyclones (TCs) have risen in scale enormously in recent decades due largely to the rising coastal populations and the increasing value of infrastructure (Knutson et al., 2010; Pielke Jr et al., 2008). The projections derived from the theory and high-resolution dynamic models illustrate that global warming will cause the global average intensity of TCs to increase by 2-11% by 2100, and that the frequency of the strongest TCs as well as the precipitation rate within 100 km of the TC center is projected to rise appreciably in the 21st century (Knutson et al., 2010). Climate change is thus a factor that has the potential to impact on the future variation in the degree of damage incurred by TCs.

TCs are intense geophysical (atmospheric) vortices that form over warm tropical oceans. It has long been recognized that TCs derive their energy largely from the release of latent heat following the condensation of water vapor in the atmospheric boundary layer (Ooyama, 1969; Shay, 2010). Of the TC-triggered disasters, strong winds, heavy and torrential rainstorms and, worst of all, the cumulative effect of storm surges and astronomical tides are the three major elements (Chan, 2005a; Mohanty and Gupta, 2008; Simpson and Riehl, 1981b). TCs cause tremendous damage to coastal regions each year. The western North Pacific (WNP), including South China Sea (SCS), has had the greatest number of TCs among the oceans. The TCs that occur in WNP cause severe damages to the coastal regions of Asia. China, which has a coast that is vastly exposed to TCs, has suffered disastrous losses inflicted by such climatic processes. For example, typhoon 'Nina' caused more than 10,000 deaths in Henan province, China, in 1975. More recently, the tropical storm 'Bilis' led to more than 1000 deaths due to landslides and flooding caused by the torrential rainstorms in Hunan province in 2006 after making landfall over Fujian province when it decayed into a tropical depression. TC forecasting, including TC

tracking and intensity predictions, are, therefore, of great significance to the scientific and public community with regard to disaster management. As described by Chan (2010), among all of the elements of TC forecasting, track prediction has always been the top priority because an incorrect prediction of the future location of a TC will render all other predictions, such as intensity, rainfall and storm surge, meaningless. The prediction of the track of a TC is one of the most difficult and challenging problems in current international TC research (Mohanty and Gupta, 2008). As illustrated in Heming and Goerss (2010), steady reductions in TC track forecast errors have been achieved by the U.S. Navy's Operational Global Atmospheric Prediction System (NOGAPS), the U.S. Navy version of the Geophysical Fluid Dynamics Laboratory model (GFDN), the Met Office (UKMO) model, and the Global Forecast System (GFS) in both WNP and North Atlantic, while the 72-hour consensus forecasts in the North Atlantic reached less than 250 km in 2008. Although TC track forecasts have become increasingly accurate over the last few decades, researches on TC intensity greatly lag behind those on TC movement (Knaff et al., 2005; Wang and Wu, 2004). Besides, TC recurvature (e.g., Dobos and Elsberry, 1993; George and Gray, 1977; Joint Typhoon Warning Center (JTWC), 1988; Riehl and Shafer, 1944) and landfall (e.g., Bender *et al.*, 1985; Chan and Liang, 2003; Tuleya *et al.*, 1984; Tuleya and Kurihara, 1977; Tuleya and Kurihara, 1978; Wong and Chan, 2006a; Wong *et al.*, 2008), which are closely related to TC movement, have been challenging the capability and accuracy of dynamic model-based and statistical TC prediction schemes. As the volume of TC-related databases is increasing rapidly, quantitative analyses (e.g., data mining and statistics-based methods) are promptly required to unravel from the historical multi-source, multi-level, and multi-scale TC data the hidden and useful regularities, patterns and rules related to TC tracking and intensity to complement the dynamic models, which simulate the physical processes through mathematical equations. Therefore, the research problems that are detailed below include TC recurvature, landfall, post-landfall TC tracks, characteristics of intensities related to TC recurvature and landfall, as well as a platform for the analysis and prediction of TCs in WNP, including SCS.

1.2 Research Problems

1.2.1 Tropical Cyclone (TC) Recurvature

Recurvature is a special type of TC tracking, i.e., turning from westward to the north and eventually to the northeast in the Northern Hemisphere (e.g., Dobos and Elsberry, 1993; George and Gray, 1977; Joint Typhoon Warning Center (JTWC), 1988; Riehl and Shafer, 1944). In general, large track forecast errors typically occur when a TC that has been predicted to recurve moves westward. Similarly, a sudden recurvature of a TC predicted to maintain a westward track can engender large forecasting errors (Dobos and Elsberry, 1993). Since nearly half of all WNP TCs eventually recurve, forecasters frequently face these potential recurvature situations (Dobos and Elsberry, 1993). During the last few decades, numerous statistical and observational studies have been conducted on TC recurvature in terms of its physical mechanisms and operational aids (e.g., Chen *et al.*, 2009; Evans *et al.*, 1991; George and Gray, 1977; Harr and Elsberry, 1991, 1995a; Holland and Wang, 1995; Krishnamurti and Subrahmanyam, 1982). Meteorological agencies nowadays take into account the TC track as a whole and do not pay special attention to TC recurvature (Elsberry, 2008, personal communication). Our current understanding has led to a framework of qualitative analysis for recurvature; for example, the influence of large-scale circulation (e.g., the subtropical high, monsoon and mid-latitude westerlies) on recurvature (e.g., Evans *et al.*, 1991; Harr and Elsberry, 1991, 1995a). However, this remains a substantial challenge for the meteorological agencies around the world.

1.2.2 TC Landfall

The greatest damage incurred by TCs usually occurs while or after they make landfall. A landfall is defined as the process of the entire TC system moving over land (Tuleya and Kurihara, 1977; Tuleya and Kurihara, 1978). On average, the coast of China annually undergoes about eight times of landfalling TCs with intensities reaching tropical storm or higher (Li *et al.*, 2004c). TCs forming in the WNP, including SCS, are generally steered

by an easterly flow to head north or northwestward when the subtropical high is strong and shifts westward. These TCs may make landfall along the coast of China if the steering current is persistently strong and westward. However, the westward-moving TCs will "turn to the north and then the northeast" (JTWC, 1988) if the steering current changes direction from westward to eastward due to the interactions of large-scale circulation. Under this change of steering flow, TCs will move far away from the Chinese coast without making landfall over it. TC landfall and recurvature are closely related since they are both controlled by large-scale circulation. Past attention focused on the seasonal or annual forecasts of TC landfall frequency in different basins, on modeling the TC-land interaction during the landfall process, on modulations of variability in atmospheric-ocean oscillations (e.g., the El Nino-Southern Oscillation (ENSO), the Madden-Julian Oscillation (MJO), the Pacific Decadal Oscillation (PDO), the Quasi-Biennial Oscillation (QBO), and the Atlantic Multi-decadal Oscillation (AMO)) on TC landfall, and on historic and prehistoric (paleo-) records of landfalling TCs study. Although the influence of large-scale circulation on TC landfall is vital in predicting landfall accurately, it has been paid little attention in the past. In this thesis, we will discover the rules and regularities that control recurvature and landfall by investigating them together through data mining.

1.2.3 Post-landfall TCs

TCs from WNP always cost tremendous lives and property after making landfalls over the Chinese coast. They begin to change in terms of their structure, intensity, and track during and after landfall. After landfall, the moist transfer is cut off and the TC interacts with the land surface. TCs tend to weaken because latent heat is released rapidly through the increased condensation of water vapor induced by TC-land interactions. The variability of weakening differs with regard to land roughness; for example, plains weaken TCs more slowly than do mountains and dense, high buildings. TC movement is more strongly influenced by steering current when TCs degrade after landfall. Latent heat exchange occurs when energy is used to convert water into a higher state (liquid to gas), or energy released as water is converted into a lower state (gas to liquid). TCs lose their energy due to the latent heat release (LHR). Our understanding of the mechanisms and

patterns associated with post-landfall TC tracks can play a vital role in our preparation for, response to and mitigation of TC-related disasters (e.g., floods, storm surges, landslides and debris flows). However, to date, discussions on post-landfall TCs have been dominated by modeling rainstorms and sustaining (Li *et al.*, 2004b; Yuan *et al.*, 2008; Yuan *et al.*, 2007), but much less attention has been paid to the characteristics of the tracks; for example, track types and lengths after landfall. Cluster analysis has been employed to unravel TC tracks and intensities in the TC database among various basins (e.g., Camargo *et al.*, 2004; Camargo *et al.*, 2008; Camargo *et al.*, 2007a, b; Cheng *et al.*, 2008; Gaffney and Smyth, 1999; Gaffney *et al.*, 2007; Gaffney, 2004; Harr and Elsberry, 1991, 1995a; Lee *et al.*, 2007). The application of cluster analysis to post-landfall TC tracks has not appeared in the literature yet. Therefore, we investigated the post-landfall TC tracks through cluster analysis to find the patterns and regularities in this study.

1.2.4 Intensities related to Recurvature and Landfall

The interaction between TC movement (recurvature and landfall) and intensity plays an important role in TC prediction in WNP. In spite of the paramount significance of this interaction, relatively little attention has been paid to the interactions between TC movements and intensities (Evans and McKinley, 1998; Velden and Leslie, 1991). The maximum intensity corresponding to recurvature has been investigated during the last few decades (Evans and McKinley, 1998; Knaff, 2009; Riehl, 1972). Past researches on the relationship between TC landfall and intensity focus on the decay rate during or after landfall (Bhowmik *et al.*, 2005a; Choi and Kim, 2007; Choi *et al.*, 2010; Wong *et al.*, 2008). However, the maximum intensity near the TC recurvature remains ambiguous to date. We investigate the relationship between maximum intensity and recurvature with the application of Maximum Potential Intensity (MPI) conceptions derived from Sea Surface Temperature (SST) (e.g., Emanuel, 1986, 1988, 1991, 1997; Emanuel *et al.*, 2004; Malkus and Riehl, 1960), the characteristics of TCs peaking at their maximum intensity prior to or after recurvature, and the duration for which the maximum intensity of straight movers and recurvers is sustained. The decay rate of landfalling TCs is also worthy of further study in order to identify the changes prior to and after landfall.

1.2.5 Multi-source Databases

The rapid development of TC-related observation means and technologies has enabled us to collect data from numerous sources, such as satellite sensors, radar devices, aerial photographs, digital scanning devices, mobile devices and global positioning systems. Different collecting devices obtain data through various data formats, spatial and temporal scales, resolutions and measurements. TC monitoring and researches involve a variety of organizations and institutes around the globe. From the 1950s, airplanes began to be used to record metrological variables outside or inside the TC (Chan, 1985a, 2006). These satellite observations are used to derive variables related to TCs from the early 1960s. With the development of radar technology and remote sensing equipment, satellite images, with their larger coverage and higher resolution, have been increasingly applied to TC study and operational forecasting. With respect to TC tracking, a number of the best tracks based on satellite observations have been collected by authoritative agencies (e.g., Japan Meteorological Agency (JMA), Joint Typhoon Warning Center (JTWC), Hong Kong Observatory (HKO), and Shanghai Typhoon Institute (STI)). With regard to the meteorological variables, the re-analysis and final analysis of the meteorological data sets (e.g., geopotential height, pressure, humidity, and zonal and meridional winds) provide conveniently available databases on which to base simulations or real-time predictions. Dealing with multi-source data is an indispensable procedure for discovering knowledge from the historical database. Therefore, data integration is examined in this study.

1.2.6 Platform for TC analysis and prediction

A TC platform is required for public consumption and professional use. This platform facilitates the efficient and effective visualization of historical TC archives, retrieval of data for analysis, as well as TC forecasting in a web-based environment. It contains a user-friendly interface, an efficient query system, facilities for the employment of dynamic models and data mining algorithms for TC analysis and prediction, particularly on TC landfall and recurvature. Existing TC systems can be grouped into several

categories: (1) stand-alone systems for professional use only, with a limited capability for TC visualization and query (Sampson and Schrader, 2000; Woodcock, 1995); (2) web-based systems for visualization, with a limited capability for TC analysis and prediction; (3) geographic information system (GIS)-based systems for simple TC analysis (Kong et al., 2008; Kumar et al., 1998); and (4) web-based systems for information dissemination, with limited data analysis capability. Therefore, a web-based, free-access GIS system should be built to bring GIS technology and TC knowledge to researchers and the general public at little or no cost. It should have functionalities for the visualization of historical TC archives, complex data analysis and information distribution for both research and professional use, and also for the consumption of the general public. A comprehensive TC analysis and prediction system with all of the above features as well as powerful data mining and dynamic modeling capabilities has yet to be developed. This study includes the construction of such a prototype platform on which data mining methods, such as clustering, the classification and regression tree (CART) and case-based algorithm and dynamical models such as The Fifth-Generation NCAR / Penn State Mesoscale Model (MM5), can also be used individually or in synchrony to accomplish more accurate TC prediction.

1.3 Research Objectives and Significances

The overall topic of this thesis is TC analysis and prediction. The overall objective is to be accomplished through the achievement of the following sub-objectives:

1. To unravel useful patterns and rules hidden in data that control TC recurvature from the historical TC database.
2. To uncover valuable regularities and rules hidden in data that govern TC landfall from the archived database related to TC landfall.
3. To discover clusters of tracks hidden in the historical post-landfall TC tracks.
4. To shed light on the relationships between TC tracks and intensity, especially the intensity change of recurving and landfalling TCs.
5. To develop a powerful web-based system with a platform independent interface for the tracking, prediction and visualization of TC movements, landfalls, and

recurvatures via dynamic modeling and data mining based on an assimilated, multi-source, multi-scale and multi-level TC database.

In order to accomplish each sub-objective, corresponding tasks are to be fulfilled accordingly. These tasks are described as follows:

For the first sub-objective, the tasks are:

- to summarize the potential parameters influencing TC recurvature;
- to establish a comprehensive database containing the potential parameters;
- to unravel hidden rules controlling TC recurvature from the database by building a decision tree;
- to justify the findings from the meteorological perspective;
- to employ the findings for TC recurvature prediction.

With regard to the second sub-objective, the tasks are:

- to summarize the potential parameters affecting TC landfall;
- to establish the database archiving the parameters;
- to define rules governing TC landfall from the database by choosing variables and cutting values to build a decision tree;
- to verify rules from the meteorological perspective;
- to utilize the findings to predict landfall.

The third sub-objective is fulfilled by the completion of the following tasks:

- to determine the number of clusters and parameters for cluster analysis;
- to discover clusters hidden in the historical TC tracks;
- to interpret the clusters from the meteorological perspective.

The fourth sub-objective comprises the following tasks:

- to revisit the relationship between TC maximum intensity and recurvature;
- to verify the results from the meteorological perspective (e.g., SST and MPI);
- to establish the spatial characteristics of TCs that attain their maximum intensity prior to or after recurvature;

- to test whether the duration for which recurvers and straight movers sustain their maximum intensity differs significantly;
- to find the intensity change before and after landfall.

The fifth sub-objective is to be realized through the accomplishment of the following interrelated tasks using innovative design concepts:

- to build a web-based platform with a user-friendly interface and effective visualization capability using advanced GIS technology and scientific speed-up techniques:
- to ensure that the system can run on any standard computer. Dynamic models based on physical processes in meteorology are utilized to generate data for tracking TCs, particularly in relation to the dynamic changes in terms of landfall and recurvature;
- to develop novel data mining algorithms to discover the rules and regularities of TC landfall and recurvature. The data mining algorithms and the dynamic model will mutually enrich each other in a complementary and integrated manner;
- to construct a multi-source, multi-scale and multi-level remote sensing database consisting of atmospheric and meteorological data such as wind fields at different levels, humidity, sea-level temperature and pressure, geopotential height, rainfall, wind shear, best track data;
- to facilitate effective and efficient TC track analysis and forecasting;
- to use state-of-the-art software tools such as Adobe Flash and Sun's Java in conjunction with Google Maps Application Programming Interface (API), to develop the system for public consumption on the internet.

The significance and contribution of this thesis are summarized as follows:

1. The research findings, as well as the platform built based on part of the findings can advance our understanding of the recurvature, landfall, post-landfall tracks and intensity change in recurving and straight-moving TCs, and facilitate TC analysis, prediction, and visualization for both professionals and the general public.

2. The rules and regularities discovered in the database are helpful for the mitigation, prediction and management of TC-related hazards.
3. Previous researches focus on employing dynamic modeling and simulation to model the physical processes of TCs. This thesis provides a new angle, i.e., data mining methods, to investigate the TC-related problems. Data mining methods are employed to uncover useful information and knowledge (e.g., patterns and regularities) from the historical database. This can complement the dynamic models to advance our understanding and simulation of TC dynamics.
4. The methods and algorithms employed in this thesis can also be applied beyond the TC-related problem to other meteorological fields; for example, monsoons, winter storms, floods, drought, and rainfall.

1.4 Thesis Structure

There are nine chapters in this thesis. Chapter One constitutes the introduction, where the research background, research problems, research objectives and significance as well as thesis structure are introduced.

Chapter Two is the background of the research. The previous researches related to TC movement, TC recurvature, TC landfall, TC intensity, and data mining are reviewed and summarized. The background of the research provides directions for examining the research problems.

Chapter Three focuses on the research methodology. In this chapter, the data resources, data mining algorithms and statistical methods are elucidated. The reasons why we choose these methods are described in this chapter.

Chapter Four discusses TC recurvature. In this chapter, the tree-based algorithm (i.e., C4.5 algorithm) is used to unravel the rules governing TC recurvature from the historical TC recurvature database that consists of various potential variables and factors that may influence recurvature. The rules are justified from the meteorological perspective; for

example, we examine the composite geopotential height and steering flow in the square anchoring at the TC center.

Chapter Five examines TC landfall. As in Chapter Four, the rules governing TC landfall are unraveled using the C4.5 algorithm from the database that includes the potential parameters influencing TC landfall. The rules are also verified from the meteorological perspective. We also compare the results of this chapter with those from Chapter Four due to the close relationship between landfall and recurvature.

Chapter Six studies post-landfall TCs. The Finite Mixture Model (FMM)-based clustering algorithm is utilized to unravel the patterns hidden in historical post-landfall TC tracks. Three clusters with different "shapes" are uncovered. We justify these clusters by examining the meteorological parameters, such as steering current, geopotential height, and horizontal and vertical water vapor supply, as well as three-dimensional wind fields.

Chapter Seven investigates the relationship between TC tracks and intensity, especially the intensity change in recurving and landfalling TCs. We examine the intensities by the statistical methods (e.g., the Student's t test, Analysis of Variance (ANOVA)) to determine whether there exist significant differences between the various groups. SST and MPI are considered for interpreting the statistical relationships.

Chapter Eight discusses the web-based system for TC analysis and prediction. Based on state-of-the-art software, such as Google Maps API, Java platform, Oracle database, and Adobe Flash, we build a prototype for a full-fledged, web-based system for TC analysis and prediction by complementing data mining and dynamic modeling.

Chapter Nine consists of the summary, suggestions for future work and the conclusion.

CHAPTER 2: Background of Research

A TC is one kind of extreme natural event, with the potential to cause the tremendous loss of human lives and social property due to excessive, torrential rainfall, flash flooding and strong winds. TC track (e.g., Carr and Elsberry, 1990; Chan, 2005a; Holland, 1983a, 1984; Holland and Lander, 1993; Holland *et al.*, 1995) and intensity (e.g., Camargo and Sobel, 2005; Emanuel, 2005; Mandal et al., 2007; Velden and Leslie, 1991; Wang and Wu, 2004; Wong and Chan, 2004) are the two paramount topics in TC-related research in terms of physical mechanisms, observational analyses, numerical simulations, and short- or long-term predictions. TC recurvature and landfall associated with TC motion are of significant social and economic concern. Further post-landfall TC research is required because TCs cause most of the damage both during and after landfall. The intensity variability of recurving and landfalling TCs can provide a useful reference for operational forecasts. These TC-related problems are investigated through data mining in this thesis. Therefore, the following subsections review the literature on these topics; i.e., TC recurvature, TC landfall, post-landfall TCs, TC intensity, a system for TC analysis and prediction, and data mining.

2.1 TC movement

The prediction of TC tracks is one of the most difficult and challenging problems for current international TC research (Chan, 2010). The physical processes of TC movement involve dynamic and thermodynamic processes because TC energy comes from the evaporation of sea water and is released as latent heat in clouds later (Chan, 2005a; Kepert, 2010). The solutions for TC track forecasting via a barotropic and baroclinic framework are both proposed and investigated with paramount concern. The pressure is only a function of temperature and the air density is constant in a barotropic atmosphere while in the real atmosphere the density is essentially a function of both temperature and pressure. This kind of fluid is regarded as baroclinic and horizontal temperature gradients commonly exist in such a fluid (Chan, 2005a).

2.1.1 Barotropic view

The barotropic framework mainly refers to the steering level/flow (Chan, 1984, 1985b; Chan and Gray, 1982; Elsberry, 1987; George and Gray, 1977), Beta effect (β-effect) or Coriolis effect (Adem and Lezama, 1960; Chan and Williams, 1987; Fiorino and Elsberry, 1989; Holland, 1983b; Rossby, 1948), asymmetric circulation (Elsberry, 1990; Fiorino and Elsberry, 1989), environmental structure (Chan, 1995b; Ngan and Chan, 1995; Williams and Chan, 1994), vortex structure (Chan and Williams, 1987; Fiorino and Elsberry, 1989) and energy budget (Chan and Cheung, 1995; Li and Wang, 1994; Wang and Li, 1995). Although the motion of most TCs can be explained by the above mechanisms (Ngan and Chan, 1995), there still exist many circumstances that cannot be properly explained by them. As summarized in Chan (Chan, 2005a), barotropic dynamics appear to dominate in the physics of TC motion in most cases. These barotropic concepts have been applied to the numerical prediction of TC movement with great success (Carr and Elsberry, 1995; Heming et al., 1995).

2.1.2 Baroclinic view

Although the barotropic view has done much to advance our understanding of TC movement, barotropic dynamics are inadequate for describing all of the physical processes associated with TC motion (Chan, 2005a). For example, in a real atmosphere, the internal structure of a TC varies with height. Cyclonic flows prevail in the lower layers and anti-cyclonic flows at the top. In addition, the effect of diabatic heating and the vertical wind shear of the environment will also interact with the TC and affect its motion (Chan et al., 2002; Ko and Chan, 2000). Therefore, from the mid-1990s, much attention has been paid to using the baroclinic approaches to improve our understanding of TC movement dynamics. These processes involved in baroclinic effects include the effect of the vertical structure of the TC vortex (Shapiro, 1992; Wang and Li, 1992; Wang and Holland, 1996b), the effect of the environmental vertical shear (Flatau *et al.*, 1994; Wang and Holland, 1996a, c; Wu and Emanuel, 1993) and the diabatic heating process (Chan *et al.*, 2002; Flatau *et al.*, 1994; Holland and Wang, 1995; Wang and Holland, 1996c).

2.2 TC recurvature

According to JTWC (1988), TC recurvature is "the turning of a TC from an initial path west and poleward to (a subsequent heading) east and poleward". Recurvature is also defined as "westward moving cyclones will sometimes turn to the north and then to the northeast" (George and Gray, 1977). Dobos and Elsberry (1993) stated that, "for the northern hemisphere, recurvature means the change from a northwestward to a northeastward track into the mid-latitude westerlies". Recurvature is referred to simply as "the directional change that many TCs undergo when their motion changes from westward and poleward to eastward and poleward" (O'Shay and Krishnamurti, 2004). In this study, the definition of recurvature is "westward moving cyclones turning to the north and then to the northeast". Numerous researches have investigated recurvature from the perspective of observational study and dynamic modeling from the 1950s.

2.2.1 Observational Study

Synoptically, the approach of a westerly trough and the eastward retreat of the subtropical ridge (STR) are found to be favorable conditions for recurvature (e.g., Riehl and Shafer, 1944). If the base weakens considerably west of a TC in conjunction with an eastward-shifting middle-latitude westerly trough and keeps low, northward recurvature will occur. If the upper westerlies cannot maintain themselves at low latitudes, the TC will circumvent the northern trough and continue in a westward direction. Similar rules are stated for several types of abnormal TC tracks, and some observations are collected on the genesis of tropical storms (Riehl and Shafer, 1944). The upper streamline analysis may be used as an additional parameter for predicting recurvature (Hill and Malkin, 1965).

From the 1950s to the present, numerous studies have been conducted statistically and observationally on TC recurvature in terms of physical mechanisms and operational aids. The subtropical high, monsoon systems, and westerlies are regarded as the most important synoptic systems controlling TC recurvature (Chen *et al.*, 2009; Evans *et al.*, 1991; George and Gray, 1977; Harr and Elsberry, 1991, 1995a; Holland and Wang, 1995;

Krishnamurti *et al.*, 1992; Miller and Han, 2001). Synoptically, a westerly trough and eastward retreat of STR are found to be favorable conditions for TC recurvature (e.g., Elsberry, 1990; Evans et al., 1991; George and Gray, 1977; Holland, 1984; Riehl and Shafer, 1944). On the contrary, a strong subtropical high sitting to the north of the TC, with a major trough in the westerlies located far to the west of the TC, is a typically synoptic pattern for non-recurvature (e.g., George and Gray, 1977; Holland and Wang, 1995).

Observational studies have found the outer circulations of TC (e.g., 5-7° radius from the TC center) to be good indicators of TC recurvature. For instance, the 300 hPa zonal wind (Burroughs and Brand, 1973), the zonal and meridional geostrophic wind component of the 200 hPa layer (George and Gray, 1977; Hodanish and Gray, 1993), the 400 hPa zonal wind (Hodanish and Gray, 1993), the 500 hPa zonal wind (Chan et al., 1980; George and Gray, 1977; Hodanish and Gray, 1993), and the deep-layer mean flow (Fitzpatrick, 1992; Holland, 1984) play an important role in determining TC recurvature within a certain time (for example, 24 hours) preceding recurvature. Average 200 hPa and 500 hPa zonal and meridional winds are typical examples of this. Statistical methods have also been employed to provide prediction schemes for TC recurvature, including discriminant analysis (Ford et al., 1993), regression analysis (Lage, 1982; Leftwich, 1980) and empirical orthogonal functioning (Ford et al., 1993; Lage, 1982). Besides, cluster analysis has been used to examine the recurving TC tracks for seasonal or annual characteristics in a variety of basins (e.g., Choi et al., 2009; Ko and Hsu, 2009)

2.2.2 Dynamic Modeling Study

In addition to the observational analyses, dynamic models, based on the barotropic (e.g., Evans et al., 1991) and baroclinic perspectives (e.g., Holland and Wang, 1995; O'Shay and Krishnamurti, 2004), have been utilized to study TC recurvature for a long time. The vorticity advection (e.g., Elsberry, 1990), the momentum between a TC and its environment (e.g., Krishnamurti et al., 1992; Li and Chan, 1999), diabatic heating (e.g., Chan et al., 2002; Holland and Wang, 1995), baroclinic waves due to divergent circulation (Hodyss and Hendricks, 2010), and the interaction of the TC and the mid-

latitude flow as the TC recurves northward (Archambault et al., 2009; Kim and Jung, 2009; Reynolds et al., 2009; Schumacher et al., 2010) all provide explanations for TC recurvature through dynamic modeling.

2.2.3 Remark

The observational analysis, statistical methods and dynamic modeling based on the barotropic and baroclinic framework have been used to examine and advance our understanding of TC recurvature. However, it remains a great challenge for TC forecasters all around the world. With the developments of technologies related to TCs, it is thus worthwhile to revisit TC recurvature. First, the observational equipment (satellites, aircrafts and radars), methods and facilities have undergone rapid developments in the last few decades. The data acquisition methods and quality have greatly advanced in terms of their accuracy, spatial and temporal resolution, and dimensions. Further, the study is equipped with more recent knowledge about TC motion and recurvature.

2.3 TC Landfall

The greatest damage incurred by TCs occurs largely during or after their making landfalls. Besides, the structure, intensity and motion of TCs will change dramatically during the landfall process. Improved TC landfall forecasting provides many potential benefits for coastal communities. TC landfall has, therefore, attracted much attention from scientists in recent years after being highlighted in the U.S. Weather Research Program (Marks and Shay, 1998). The seasonal and annual variability of TC landfall frequency is of great concern to scientists, governments and insurance companies. Statistical modeling has long been applied to investigate seasonal and annual TC landfall (e.g., Blackwell *et al.*, 2006; Kim *et al.*, 2009a; Medlin *et al.*, 2007) and dynamic models have been widely used to simulate the rainfall distribution and TC-land interaction during and after landfall.

Numerous researches have investigated the seasonal or annual forecasts of TC landfall frequency in the Atlantic basin (e.g., Bove et al., 1998; Larson et al., 2005; Lehmiller et al., 1997; O'Brien et al., 1996) and WNP (Chan, 1995a; Elsner and Liu, 2003; Saunders

et al., 2000; Wu *et al.*, 2004). Particular attention has also focused on landfall over regional areas; for example, Southern China (Goh and Chan, 2010; Liu and Chan, 2003) and the Korean Peninsula (e.g., Fudeyasu et al., 2006). The variability of seasonal TC landfall frequency was attributed to the modulation of ENSO, MJO, QBO, AMO and PDO (e.g., Goh and Chan, 2010; Liu and Chan, 2003). The variability of large-scale circulation (the subtropical high, monsoon trough and westerly anomaly) associated with ENSO events or other oscillations is likely to be responsible for most of the seasonal or annual variations in TC landfall activity in WNP (e.g., Goh and Chan, 2010; Liu and Chan, 2003) and the Atlantic basin (Brettschneider, 2008; Lehmiller et al., 1997). Besides, historic and prehistoric records of landfalling TCs are also reconstructed using geological proxies, such as coastal lagoon sediment as well as the isotopic composition of stalagmites, tree-rings, and coral in both the Atlantic basin (e.g., Donnelly et al., 2001; Liu and Fearn, 2000; Reading, 1990) and WNP (for example, Liu et al., 2001; Louie and Liu, 2003).

With regard to rainfall distribution, its relationship with landfall is widely investigated through dynamic modeling (Blackwell *et al.*, 2006; Kim *et al.*, 2009a; Medlin *et al.*, 2007; Yu and Cheng, 2008). In addition, the surface wind distribution and decaying characteristics are continuously investigated using empirical models or observational analysis due to the direct damage induced by the strong, raging winds of TCs (e.g., Bhowmik *et al.*, 2005b; Fujibe and Kitabatake, 2007; Powell, 1982; Powell and Houston, 1998; Powell *et al.*, 1991; Wong *et al.*, 2008).

TC-land interaction during landfall has been studied through observational analysis and dynamic modeling over the last few decades (refer to Bender *et al.*, 1985; Chan and Liang, 2003; Dong *et al.*, 2010; Wong and Chan, 2006b; Wroe and Barnes, 2003). The impact of topography on landfalling TCs is more significant than roughness variation, especially in mountainous areas (Chang, 1982). Observations have corroborated the significant correlations between regions of peak rainfall during TC landfall and the mountainous land surface (Brunt, 1968; Hamuro and Coauthors, 1969). Brand and Blelloch (1973, 1974) found that the topographical effects, if not properly accounted for,

can cause significant errors in the forecasting of TC motion. Chang (1982) also observed that the interaction between the terrain and the TC caused strong easterlies to develop north of the island, accelerating the TC in his case.

To recapitulate, previous researches on TC landfall were concerned with the seasonal or annual TC landfall frequency according to the variability of large-scale circulation (e.g., the subtropical high, monsoonal systems and westerlies) modulated by ENSO, QBO and MJO, with precipitation and wind distribution associated with TC landfall, and with changes in the tracks and structures when a TC passed through an island by studying land-TC interaction, with TC decaying mechanisms after its making landfall. However, few researchers have investigated the mechanisms controlling TC landfall over the Chinese coast based on the historical TC database. Besides, few researchers have investigated the TC forecast scheme solely for landfall. Up until recently, few objective forecast aids in operational use have been specifically designed to identify landfall situations. Given our existing understanding of the TC landfall mechanisms based on observational and statistical analysis and dynamic modeling (e.g., Brettschneider, 2008; Elsner and Liu, 2003; Goh and Chan, 2010; Liu and Chan, 2003), large-scale circulation exerts a dramatic influence on TC landfall activities in the SCS and WNP basins. Of particular scientific interest and value is unraveling the rules and regularities governing TC landfalls over the Chinese coast from the historical TC archives.

2.4 Post-landfall TCs

Landfall is taken to be the process of the entire TC system moving over land (Tuleya and Kurihara, 1977). When a TC moves over land, the cutting off of the water vapor supply, the increased surface roughness, the interaction with mid-latitude systems, and the decreasing upwards motion in the TC center all make the post-landfall TCs very different from those on the ocean. The TC-land interaction is summarized in the last section. Empirical models and observational analysis were used to predict the decay rate of TC intensity after landfall over South China (Wong et al., 2008), the Indian region (Bhowmik *et al.*, 2005b), and the U.S. coast (Goldman and Ushijima, 1971; Kaplan and Demaria, 1995, 2001). Apart from these studies, the heavy rain mechanism associated with post-

landfall TCs has undergone numerous investigations through dynamic modeling (e.g., Tao et al., 1994; Yu et al., 2010; Zhang et al., 2007). There have been several studies related to the sustaining and movement of typhoons post-landfall (Li et al., 2004a; Li et al., 2004b; Yuan et al., 2008; Yuan et al., 2007).

In summary, these existing investigations on post-landfall TCs focus on the rainstorm mechanism associated with mesoscale circulations and landfall surface, as well as the mechanism and characteristics of the sustaining mechanisms of TC after landfall, through dynamic modeling. From these investigations, the sustaining of TC is largely influenced by the water vapor supply, mid-latitude interaction and upper-level wind flows. However, due to the complexity of TC-land interaction post-landfall, there has been far less research on post-landfall TC characteristics than on TC tracks and intensity. The study of the characteristics of post-landfall TC movements (e.g., types and shapes) has received little attention during the last few decades. We discovered the knowledge from the historical TC tracks through data mining methods.

2.5 TC intensity and Movement

2.5.1 TC Intensity

TC track forecasts are becoming increasingly accurate (Wang and Wu, 2004), thanks largely to a combination of better observations, especially via satellite (Soden et al., 2001) and dropsonde (Aberson and Franklin, 1999; Burpee et al., 1996). Nonetheless, the research on TC intensity (defined as either the maximum surface wind or the minimum central surface pressure) is far less developed than that on TC movement (e.g., Fitzpatrick, 1997; Knaff et al., 2005; McBride and Holland, 1987; Wang and Wu, 2004). In spite of the application of numerical modeling and the availability of advanced remote sensing equipment and observations, little progress regarding the prediction of TC intensity and the intensity change rate has been made in the past (DeMaria et al., 2005; Emanuel et al., 2004; Wang and Wu, 2004). The best TC intensity forecast scheme is still statistically based, not dynamic model based (DeMaria and Kaplan, 1999; Wu and Wang, 2004). The reasons for this include the inadequate observations over the ocean, the spatial and

temporal model resolution, the initial and boundary condition of dynamic models, and our lack of understanding of the physical processes involved in controlling the intensity change, especially the rapid intensification defined by Holliday and Thompson (1979) as a central pressure drop of at least 42 hPa per day (Fitzpatrick, 1997; Wang and Wu, 2004). The operational prediction of rapid intensification has proven to be particularly difficult, due largely to our lack of understanding of the physical mechanisms underlying rare events (Elsberry et al., 2007; Kaplan et al., 2010). Since the intensity change is influenced at any time by sophisticated physical processes and systems controlling the internal TC structure and the interplay between the TC and both the surrounding environment and ocean (Hong *et al.*, 2000; Wang, 2002; Wu and Cheng, 1999), understanding TC intensity changes is far more complex and difficult than understanding TC movement (Fitzpatrick, 1997; Wang and Wu, 2004).

Based on the current understanding, TC intensity and intensity change are modulated by the interaction between upper-tropospheric trough and TC, internal TC dynamics, ambient wind flows, MPI, and internal TC dynamics (Wu and Wang, 2004). MPI is defined as the theoretical limit of the strength of TC and it is a function of SST (Emanuel, 1986, 1988, 1997; Emanuel et al., 2004). The intensity is proven to be remarkably influenced by internal TC dynamics including spiral rainbands, vortex Rossby waves, mesoscale vortices and eyewall processes (e.g., Chen and Yau, 2001; Diercks and Anthes, 1976; Holland and Dietachmayer, 1993; Kossin *et al.*, 2002; Kuo *et al.*, 1999; Kurihara, 1976; Lewis and Hawkins, 1982; Macdonald, 1968; Marks *et al.*, 2008; Montgomery and Lu, 1997; Montgomery *et al.*, 2002; Raymond and Jiang, 1990; Reasor *et al.*, 2000; Simpson *et al.*, 1997; Wang, 2001; Wang, 2002; Willoughby, 1978; Willoughby *et al.*, 1984)

Statistical schemes for TC intensity have been proposed and used in operations; for example, the Statistical Hurricane Intensity Forecast model (SHIFOR) (Jarvinen and Neumann, 1979), the Typhoon Intensity Prediction Scheme (TIPS) (Fitzpatrick, 1997), the Statistical Hurricane Intensity Prediction Scheme (SHIPS) (DeMaria and Kaplan, 1999; DeMaria et al., 2005; Knaff et al., 2005) and the Logistic Growth Equation Model

(LGEM) (DeMaria, 2009). Dynamic models have been used to simulate and predict TC intensity and intensity change. As verified in Gentry and Lackmann (2010), grid-length and domain size suggestions for operational predictions are provided; for operational prediction, a grid length of 3 km or less is recommended. Up until recently, only a limited number of models (e.g., the Geophysical Fluid Dynamics Interpolated model (GFDI; Bender et al., 2007) and the Geophysical Fluid Dynamics Model-Navy (interpolated) model (GFNI; Rennick, 1999)) have attained resolutions whereby TC structure (e.g., intensity) can be successfully identified (Heming and Goerss, 2010).

2.5.2 Relationship between TC Intensity and TC Track

TC landfall and recurvature are of paramount interest in TC research. The relationship between TC track and intensity concentrates on the relationship between TC recurvature and landfall with intensity. The interaction between TC movement and intensity plays an important role in operational TC prediction. In spite of its great significance, relatively little attention has been paid to the interaction between TC movement and intensity. For example, Velden and Leslie (1991) investigated the relationship between TC intensity and direction of movement using the depth of the environmental steering layer, while Evans and McKinley (1998) discussed the relationship between intensity and motion changes. Past researches on the association between TC landfall and intensity focus on the decay rate both during and after making landfall along the South China coast (Wong et al., 2008), Korean Peninsula (Choi and Kim, 2007; Choi et al., 2010), Indian coast (Bhowmik *et al.*, 2005a), U.S. coast (Powell *et al.*, 2004), and Japanese coast (e.g., Fujibe and Kitabatake, 2007).

The maximum intensity corresponding to recurvature has been investigated during the last few decades (Evans and McKinley, 1998; Knaff, 2009; Riehl, 1972, referred to as EM98, K09 and R72 respectively). It has been reported that TC maximum intensity tends to be coincident with TC recurvature (Evans and McKinley, 1998; Riehl, 1972). R72 found that roughly two-thirds of TCs that reach typhoon/hurricane intensity attained their maximum lifetime intensity within 12 h of recurvature in WNP. Instead of 45 and 70% of all WNP TCs peaking within ±12 h and ±24 h of recurvature, as reported in EM98 for the

period 1980-1996, the datasets for 1980-2006 in K09 suggested that these percentages are close to 28.9 and 50.0% respectively. It is inferred in K09 that the long-held conclusion that most TCs attain their peak intensities near or close to recurvature is not supported by the observations.

2.5.3 Remark

TC intensity research lags far behind that of TC movement due largely to our lack of understanding of the complicated physical processes that are responsible for the intensity change. Based on the existing literature, most TCs decay after landfall. There also exists some unusual intensification of TCs post-landfall. However, the specific characteristics of TC intensity change before and after making landfall over the Chinese coast have seldom been examined. The relationship between TC recurvature and maximum intensity remains ambiguous, based on K09, and the TC intensity changes have not yet been explored comprehensively with respect to TC recurvature. The intensity changes in recurving and landfalling TCs are examined comprehensively through the consideration of MPI.

2.6 Data Mining and Knowledge Discovery

Dynamical models based on physical mechanisms and processes are broadly used for TC simulation and real-time TC prediction. However, the dynamical models are always confined by initial and boundary conditions and also affected by chaos (Lorenz, 1965; Lorenz, 1969). Another problem is that dynamic modeling precludes the historical databases that include useful information and knowledge. The data mining and knowledge discovery (KDD) algorithm can be used to discover knowledge and information from the TC database and enrich the dynamic models for TC analysis and prediction.

KDD refers to "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad and Stolorz, 1997; Fayyad *et al.*, 1996). By definition, KDD is "the overall process of discovering useful knowledge from

data whereas the data mining method refers to a particular step, the application of specific algorithms for extracting patterns, structures, processes and relationships from data, in this process" (Fayyad et al., 1996; Lavrac et al., 1997; Leung, 2010). More recently, as defined in Han and Kamber (2006), data mining, "also popularly referred to as knowledge discovery from data, is the automated or convenient extraction of patterns representing knowledge implicitly stored or captured in large databases, data warehouses, the Web, other massive information repositories, or data streams". Kantardzic (2003) referred to data mining as "the entire process of applying a computer-based methodology, including new techniques, for discovering knowledge from data". Based on their views, data mining is treated as a synonym for KDD. We tend to regard data mining as a synonym for KDD throughout this dissertation.

Data mining has become an active research field that is of scientific and commercial concern nowadays (Han and Kamber, 2006; Leung, 2010). It is based on machine learning, pattern recognition, database systems and statistics (Koperski et al., 1996). Data mining methods include statistical approaches; for example, correlation analysis, regression analysis, clusters analysis and principal component analysis. The heuristic algorithms such as genetic algorithm, neural network and support vector machine also lay foundation for data mining. Data mining bypasses the traditional statistics to adapt data that are not normally suitable for statistical models with strict assumptions (such as independence, stationarity of the underlying processes, and normality). The data archived in real-life databases are, however, sometimes nonnumeric, noisy, inconsistent and incomplete. In the meantime, data mining is more strongly inductive than traditional statistical analysis (Han and Kamber, 2006; Miller and Han, 2009).

Spatial data mining involves "the discovery of structures and processes in spatial and temporal data" (Leung, 2010). Spatial data mining, also called "geographic data mining and knowledge discovery", refers to "the extraction of implicit knowledge, spatial relationship or other patterns not explicitly stored in spatial database" (Koperski and Han, 1995; Koperski *et al.*, 1996; Roddick and Spiliopoulou, 1999; Roddick *et al.*, 2001;

Shekhar and Chawla, 2003; Shekhar *et al.*, 2010). This dissertation adheres to the definition of spatial data mining of Leung (2010).

Although spatial data mining or data mining have been widely applied to various fields, its applications in the field of meteorology in general and TC research in particular are relatively few. Data mining algorithms such as clustering algorithms, association rule analysis and classification have been employed to investigate TC tracks, cyclogenesis and intensities in TC databases (e.g., Camargo *et al.*, 2004; Camargo *et al.*, 2008; Camargo *et al.*, 2007a, b; Cheng *et al.*, 2008; Gaffney and Smyth, 1999; Gaffney *et al.*, 2007; Gaffney, 2004; Harr and Elsberry, 1995a; Lee *et al.*, 2007; Li *et al.*, 2009; Yang *et al.*, 2011). In terms of temporal data mining, wavelet analysis and Fourier Transformation have been used to unravel knowledge from historical TC-related variables for a long time (e.g., Chan, 2005b; Liu and Chan, 2008).

A data-rich but information-poor situation is ascribed to the increasing abundance of data, together with the urge requirement for sophisticated techniques and tools to handle data (Han and Kamber, 2006; Miller and Han, 2009). The TC-related database is particularly difficult to visualize and understand because the data grows along two dimensions: the number of fields (also depicted as attributes) and the number of cases. The rates of growth of TC-related data sets exceed much faster than the rates that the traditional "manual" analysis methods can handle. Useful knowledge (e.g., structures, processes, relationships, regularities and patterns) underlying the TC mechanism is often hidden in the historical TC database. Therefore, the data mining approach is suitable for unraveling from the TC database the rules governing TC-related processes throughout their movements.

2.7 Summary

The existing literature related to research problems such as TC recurvature, TC landfall, TC intensity and post-landfall TCs has been reviewed in this chapter. It is summarized as follows:

1. TC recurvature remains a considerable challenge given the advances in TC movement research. Different from simulating the physical processes underlying TC recurvature, data mining methods are used to discover the useful and hidden knowledge (e.g., rules, patterns, regularities and structures) governing the recurvature.

2. Past researches on TC landfall focus on TC-land interaction, rainfall and wind distribution through dynamic modeling or observational analysis. TC landfall is similar to TC recurvature because they are both largely influenced by large-scale circulation. Therefore, TC landfall is also examined through data mining in general to discover the knowledge underlying the landfall mechanisms.

3. Post-landfall TCs have received little attention in the past, since TCs tend to decay rapidly post-landfall. However, extreme disasters are always triggered by post-landfall TCs. Post-landfall TCs have been studied frequently in the past few years. Cluster analysis is employed to discover the types or patterns hidden in the historical post-landfall TC tracks.

4. TC intensity is a very complicated problem. Our prediction and understanding of TC intensity or intensity change lag far behind those of TC movement. The relationship between intensity and movement are far less well understood. Therefore, the intensity of recurving and landfalling TCs are investigated through statistical models to find the characteristics of intensity change.

5. Machine learning, pattern recognition, database systems and statistics have laid the foundation for data mining. Spatial data mining is more complicated than general data mining due to the complications of spatial data and consideration of spatial dependence. The rapid growth of a TC-related database and the useful knowledge hidden in the data sets urgently require powerful tools for data mining. Therefore, we seek to explore TC dynamics through data mining in this dissertation.

CHAPTER 3: Research Methodology

With respect to the research problems, the research background is summarized in Chapter 2. Based on the research background, the tasks that need to be conducted are summarized as: to use data mining methods to examine TC recurvature and landfall; to conduct a cluster analysis of post-landfall TC tracks; to analyze the relationship between TC movement and intensity; and to build a web-based TC platform for TC analysis and prediction. In this chapter, the research framework is proposed for investigating these problems. It comprises data mining methods that are used to analyze a TC-related database. The main research framework is shown in Figure 3-1. The left side of Figure 3-1 describes the problems to be solved: TC recurvature, TC landfall, post-landfall TC analysis, intensities related to TC movement and TC platform. The middle of Figure 3-1 illustrates the datasets to be used to solve these problems. These datasets are listed as follows: JMA TC best track dataset, National Center for Environmental Protection/ National Center for Atmospheric Research (NCEP/NCAR) reanalysis dataset, Tropical Rainfall Measuring Mission (TRMM) precipitation and SST datasets and the United States Geological Survey (USGS) DEM dataset. In order to handle the datasets and discover useful knowledge to solve the problems, C4.5 algorithm, FMM-based clustering algorithm, ANOVA, the Student's *t* test and dynamic modeling are employed and shown in the right side of Figure 3-1.

3.1 Data Source

The database built in this dissertation includes the Regional Specialized Meteorological Center (RSMC) Tokyo at the JMA best track dataset, NCEP-NCAR reanalysis and final analysis dataset, the NCEP Global Forecasting System (GFS) Final analysis (FNL) dataset, the SST and precipitation rate of TRMM data and USGS DEM. The collected raw data are integrated together or transmitted into appropriate formats. The datasets cover a range of atmospheric levels (e.g., 1000 hPa, 850 hPa, and 200 hPa), and different spatial and temporal resolutions, and are collected by different satellites from authorized meteorological organizations. The database is, therefore, multi-source, multi-scale and

multi-level. The datasets are processed and analyzed in the proper data formats that can be accommodated by data mining tools and methods. The datasets and analysis methods are discussed in detail in the following subsections.

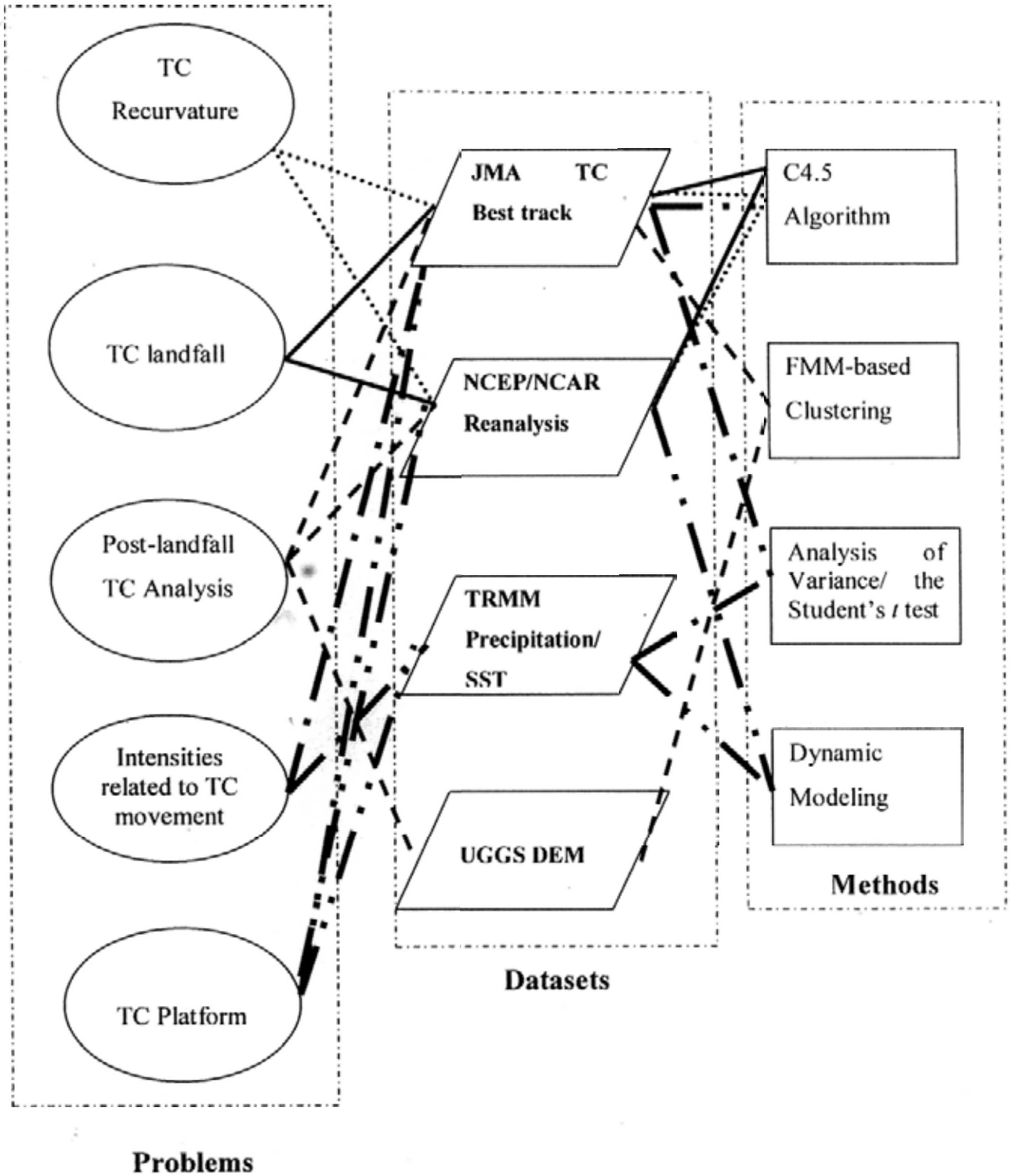


Figure 3-1 Conceptual framework and research design

3.1.1 Japan Meteorological Agency (JMA) Regional Specialized Meteorological Center (RSMC) TC Best Track

The TC best track data are made available from JMA RSMC Tokyo. This post analysis best track data consist of 6-h estimates of position (latitude and longitude), Minimum Central Pressure (MCP), and 10-minute maximum sustained wind speed (MSW) for all named TCs in the WNP basin including the SCS from 1951 to the present. The best track datasets from 2000 to 2010 are selected for data mining because the meteorological variables released by the NCEP GFS FNL with 1° × 1° spatial resolution are available for the time range from 30/07/1999 to the present at 6-hour time intervals. The datasets for examining TC recurvature and landfall are introduced as follows.

Recurving TCs with certain characteristics are retained as follows: (1) the TC sustains for at least 72 hours in best track, (2) the recurvature point has the minimum longitude among the entire track, (3) the latitude of the subsequent point is greater than the latitude of the recurvature point (i.e., no looping tracks or equatorward recurving TCs). In total, 2552 sampling TC points were obtained from 102 TC tracks, 63 of which are recurving tracks and the rest are straight-moving ones, consistent with the definitions of "recurvature". It is noted that the recurvature points of 63 recurving TC tracks are on the ocean and that TCs that recurved over land are excluded, since the respective mechanisms for recurvature on the ocean and on land should be driven by different mechanisms because of the TC-land interaction and the cutting off of water vapor for post-landfall TCs. Recurvers and straight movers are two classes of TC samples. Of the recurving TC tracks, 1560 points previous to TC recurvature are drawn as samples and these are labeled as class "0". On the contrary, the 992 sample points in straight-moving tracks are labeled as class "1".

With regard to TC landfall, there is a total of 6920 sampling TC observations derived from 222 TC tracks, 53 of which are landfalling TC tracks and the rest are non-landfall

TC tracks. The landfalling TC tracks contain 1150 observations and the non-landfall TC tracks contain 5770 observations. It is noteworthy that, with respect to the observations of landfalling TCs, only the observations prior to landfall are sampled and labeled as class "1". All observations of non-landfall TC tracks, on the contrary, are labeled as class "0".

3.1.2 National Center for Environmental Protection/ National Center for Atmospheric Research (NCEP/NCAR) Datasets

The meteorological variables (e.g., SLP, wind fields and geopotential heights in different atmospheric layers) are derived from NCEP GFS FNL analysis which covers the time range from 30/07/1999 to the present at 6-hour time intervals (Yang et al., 2006). The web link to the NCEP "FNL" dataset is http://dss.ucar.edu/datasets/ds083.2/. These NCEP FNL Operational Global Analysis data are on 1.0° x 1.0° degree grids continuously at every six hours. This product is from GFS that is operationally run four times a day in near-real time at NCEP. The analyses are available on the surface, at 26 pressure levels from 1000 hPa to 10 hPa (e.g., 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 and 1000 hPa), in the surface boundary layer and at some sigma levels, the tropopause and a few others. The parameters include SLP, relative vorticity, surface pressure, relative humidity, vertical velocity, geopotential height, SST, soil values, zonal (u-) and meridional (v-) winds and ozone parameters.

Although the NCEP/NCAR reanalysis data (Kalnay et al., 1996) are available for a longer time period, from 1948 to present, the resolution of 2.5° x 2.5° degrees of NCEP/NCAR is coarser than that of the NCEP FNL dataset. The link to the NCEP/NCAR reanalysis data is http://www.esrl.noaa.gov/psd/data/gridded/data.ncep reanalysis.html. Therefore, we derive the meteorological variables from the NCEP FNL dataset and NCEP/NCAR dataset according to the requirements of spatial resolution. The TC samples for investigating TC recurvature and landfall are only extracted for the period from 2000 to 2010 by considering the higher resolution and available time period of the NCEP FNL dataset. The NCEP/NCAR reanalysis dataset from 1985 to 2009 is used to study the relationship between TC intensity and movement (e.g., landfall and recurvature).

3.1.3 Sea Surface Temperature (SST) and Precipitation Rate by Tropical Rainfall Measuring Mission (TRMM)

In order to examine the relationship between TC movement and intensity, MPI plays a key role in quantifying the changes in intensity. MPI is derived from SST and TC intensity (DeMaria and Kaplan, 1994; Knaff et al., 2005). The precipitation rate is taken as an indicator of TC intensity. Precipitation rate and SST are accessed by TRMM datasets. The rainfall dataset used in this study is the TRMM Multi-satellite Precipitation Analysis (TMPA) (Berg et al., 2006; Huffman et al., 2007; Kummerow et al., 1998; Kummerow et al., 2000). TRMM is a joint project between the National Aeronautics and Space Administration (NASA) and Japan's National Space Development Agency, starting from 1997 (Wentz et al., 2002). It is basically designed to monitor and study the characteristics of tropical rainfall and the related release of latent heat energy that maintains global atmospheric circulation, and modulates the global weather and climate. TMPA with fine resolutions (0.25° × 0.25° and 3-hour) integrates precipitation from a number of satellites, as well as land surface rainfall gauge products. The TRMM rainfall covers the area between 50°N and 50°S for the period from 1998 to the present and is used in this study. It is available at ftp://disc2.nascom.nasa.gov/data/s4pa/TRMM_L3/TRMM_3B42/.

The through-cloud capacities of microwave radiometers offer a meaningful profile of global SST (Kummerow et al., 2000; Ricciardulli and Wentz, 2004; Wentz et al., 2000). Therefore, scientists have produced a daily, optimally Interpolated (OI) SST product at a 0.25° × 0.25° degree (~25 kilometers) resolution. This dataset greatly suits for research activities that desire a complete, daily SST profile but not one with missing data because of orbital gaps or environmental conditions excluding SST retrieval. The TRMM Microwave Imager (TMI) on NASA's TRMM satellite was the first microwave radiometer that calibrates well with accurate through-cloud SST retrieval. TMI Optimum Interpolation (OI) daily SST (Gentemann et al., 2004) at a 0.25° × 0.25° spatial resolution together with JMA RSMC Tokyo TC best track data are utilized to estimate MPI, which is one of the most important predictors for the TC intensity prediction models. TMI OI

SST analysis is available from the RSS website at http://www.ssmi.com and covers the latitude band 40°S–40°N for the period from January 1998 to the present.

3.1.4 United States Geological Survey (USGS) Digital Elevation Model (DEM)

The elevation at the TC center is used to build the clustering model for post-landfall TC tracks because the surface roughness and elevation play an important role in TC movement post-landfall (e.g., Chan and Liang, 2003; Li *et al.*, 2004a; Li *et al.*, 2004b; Tuleya *et al.*, 1984; Yuan *et al.*, 2008; Yuan *et al.*, 2007). The elevation data are the GTOPO30, available from USGS website. GTOPO30 is a global digital elevation model (DEM) with a horizontal spatial resolution of 30 arc seconds (approximately 1 kilometer). Of note is that GTOPO30 was based on several vector and raster sources of topographical characteristics (i.e., land surface). The USGS DEM data files represent cartographic information digitally in a raster format (e.g., tiff). The link to the GTOPO30 dataset is *http://edc.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30_info*.

3.1.5 Data Integration

Data integration is defined as "combining data from multiple sources into a coherent data store, as in data warehousing" (Han and Kamber, 2006). It seeks to overcome "the problem of combining data residing at different sources, and providing the user with a unified view of these data" (Lenzerini, 2002; Naumann et al., 1999). The sources usually include multiple databases, data cubes, or flat files. Data integration increases frequently due to the fact that the volume and requirement to share the existing data sources is rising rapidly. Data integration is nowadays a critical problem in our increasingly interconnected but intrinsically heterogeneous environment. Numerous data sources are made available via public information systems such as the World Wide Web. The primary purpose of data integration is to offer programmatic and human users with a comprehensive and combined interface to multiple, heterogeneous data sources (Lenzerini, 2002; Naumann et al., 1999). The users may have the illusion that they are individually visiting a single, homogeneous database for a specified requirement. The data integration process can now be automated in many cases (Genesereth, 2010). The most important issues technically related to data integration include: standards,

interoperability, vertical topology, semantic, reference system, data model, format, and data quality.

The research problems in this dissertation are utilizing comprehensively the multi-source, multi-scale and multi-level spatial databases (e.g., NCEP/NCAR, JMA, USGS, TRMM datasets) for data mining. The data associated with different data sources have different data formats, spatial and temporal scales and atmospheric levels. The integration of multi-sourced spatial data owing to diversity of data providers is thus of great scientific and practical concern. In addition, the technical integration and interoperability of the data sources play an important role in the processes of TC analysis and prediction. Data integration associated with TCs is largely the match of datasets geometrically and topologically, as well as the match of spatial and temporal resolution, vertical levels, referencing system and data formats. Therefore, the processes of utilizing the numerous TC-related data sources fall within the sphere of data integration.

3.2 Methodology

3.2.1 Data Mining-C4.5 Algorithm

Much useful knowledge and information is hidden in the historical TC database. On the other hand, operational and professional people are trapped and stuck in a dilemma caused by the lack of sufficient useful references and knowledge. Data mining is the process of discovering patterns and structures from database (Leung, 2010). It is used to unravel the rules and regularities underlying TC dynamics (e.g., the dynamics associated with TC recurvature, landfall and post-landfall movement). Specifically, tree-based algorithms (e.g., CART and C4.5 algorithm) are employed to find the rules controlling TC recurvature and landfall.

KDD refers to "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad and Stolorz, 1997; Fayyad *et al.*, 1996). By definition, KDD is "the overall process of discovering useful knowledge from data whereas the data mining method refers to a particular step in this process, the

application of specific algorithms for extracting patterns, structures, processes and relationships from data" (Fayyad et al., 1996; Lavrac et al., 1997; Leung, 2010). As pointed out in Chapter 2, we tend to regard data mining as a synonym for KDD throughout this dissertation.

A decision tree is a typical data mining algorithm for unraveling rules and selecting features from databases for a decision making process (Breiman *et al.*, 1984a; Quinlan, 1987). A decision tree is defined as "a classification procedure that recursively partitions a data set into smaller sub-divisions based on a series of tests defined at each branch (or node) in the tree" (Quinlan, 1987). The tree consists of a root node (derived from all of the data), a set of internal nodes (splits), and a set of terminal nodes (leaves) (Friedl and Brodley, 1997). It is noteworthy that each node in the tree has only one parent node, and two descendant nodes or child nodes or more. A dataset is classified by sequentially partitioning it based on the decision framework defined by the tree, and a "class label" is then assigned to each observation according to the leaf node into which the observation belongs to (Friedl and Brodley, 1997).

Decision trees are superior to traditional classification procedures such as maximum likelihood classification in a number of ways. In particular, decision trees are based on strictly non-parametric procedures and any assumption with regard to the distributions of the input data is not required. In addition, they can handle non-linear relationships between variables and classes, accommodating missing values, and are capable of dealing with both numeric and categorical input (Fayyad and Irani, 1992; Hampson and Volper, 1986). Finally, decision trees are of considerable interest because the classification structure is explicit and thus easily interpretable (Friedl and Brodley, 1997).

A classic decision tree is CART model described in Breiman et al. (1984a). CART is a non-parametric statistical method that has been developed for handling classification problems from categorical or continuous dependent variables (Breiman *et al.*, 1984a). CART generates a classification tree based on categorical dependent variables. The continuous dependent variables, however, produce a regression tree. In classification and

regression trees, the essential objective of CART is to generate an accurate set of data classifiers by unraveling the predictive structure of the problem of interest (Breiman *et al.*, 1984a). Corresponding to CART, a tree-structured decision space is achieved by recursively partitioning the data at each node according to a statistical test raising the homogeneity of the training data in the resultant descendant nodes (Breiman *et al.*, 1984a; Fayyad and Stolorz, 1997). Other decision tree algorithms, such as Chi-squared Automatic Interaction Detection (CHAID), ID3 and C4.5 (Quinlan, 1993), are also proposed for classification. CHAID and ID3 have limitations in classifying continuous variables. Under this condition, continuous variable must be converted into categories in order to be digested by CHAID and ID3. However, C4.5 algorithm can be used for continuous variables. Considering the continuous variables employed in this study (e.g., wind fields and the indices of large-scale circulation), the CART and C4.5 algorithms will be employed for the current classification. The construction of a CART classification rule is based on the definitions of three primary elements: (1) the sample-splitting rule; (2) the goodness-of-split criteria, and (3) the criteria for choosing an optimal or final tree for analysis (Breiman *et al.*, 1984a). Classification tree is built in CART by applying predefined splitting rules and goodness of split criteria in the node-splitting process at each step.

In general, the three steps involved in the tree building process (Yohannes and Webb, 1999) are:

(1) growing a large tree (a tree with a large number of nodes);

(2) combining some branches of this large tree to generate a series of sub-trees of varying sizes (different numbers of nodes);

(3) choosing an optimized tree according to "measures of accuracy of the tree".

In CART, the variable splitting criteria are the Gini impurity criterion and the Twoing criterion for nominal variables. Taking the multi-class problem as an example, the two-class problem is simply a special case. Given a node t with the estimation of class probabilities $p(j|t)$, $j = 1,\ldots, J$, a measure of node impurity given t

$$i(t) = \Phi(p(1|t), \ldots, p(J|t)) \qquad (3.1)$$

is defined and a search conducted for the split that results in the greatest reduction in node or, tree impurity. The original function was

$$\phi(p_1,\ldots,p_J) = -\sum_j P_j \log P_j \qquad (3.2)$$

It is noted that the Gini diversity index was adopted and can also be written as

$$i(t) = 1 - \sum_j p^2(j|t)) \qquad (3.3)$$

In the two-class problem, the index is reduced to

$$i(t) = 2p(1|t)p(2|t) \qquad (3.4)$$

The interpretation of the Gini index is interesting. Other than utilizing the plurality rule to classify elements in a node t, it adopts the rule that randomly assigns an element from the node to class I with probability p(i|t). The estimated probability that the element belongs to class j is p(j|t). Therefore, the estimation of the misclassifying probability corresponding to this rule is the Gini index

$$\sum_{i \neq j} p(i|t)p(j|t) \qquad (3.5)$$

It has proven that, for any split s, $\Delta i(s, t) > 0$, actually it is strictly concave (Breiman *et al.*, 1984a), so that $\Delta i(s, t) = 0$ only if $p(j|t_L) = p(j|t_R) = p(j|t)$, $j = 1,\ldots,J$.

As another powerful classification algorithm, the C4.5 algorithm (Quinlan, 1993) is also a supervised learning method based on decision-tree induction. The fundamental strategy is to select an attribute that will best separate samples into individual classes via a measurement; here the measurement is the 'Information Gain Ratio', based on information-theoretic 'entropy'. The primary objective is to find the minimum information required to retain the least "impurity" of the partitions (Han and Kamber, 2006).

Let S be the training set consisting of s data samples, and $s(C_i)$ is the number of records in S that belong to class C_i (for i=1, 2 , ..., m). The information (entropy) required to classify S is

$$Info(S) = -\sum_{i=1}^{m} \frac{s(C_i)}{s} \log_2\left(\frac{s(C_i)}{s}\right) \qquad (3.6)$$

Hence, the amount of information required to subdivide S into $\{S_1, S_2 \ldots S_v\}$ by attribute A (the number of distinct values of attribute A is 'v') is

$$Info(A \mid S) = -\sum_{j=1}^{v} \frac{S_j}{S} Info(S_j) \tag{3.7}$$

The gain is calculated as

$$gain\ ratio(A \mid S) = \frac{gain(A \mid S)}{Info(A \mid S)} \tag{3.8}$$

where

$$gain(A \mid S) = Info(S) - Info(A \mid S) \tag{3.9}$$

3.2.2 Finite Mixture Model-based Clustering

It is assumed that there are several patterns (i.e., clusters) hidden in the historical TC tracks from the meteorological perspective. Some empirical analyses indicate that TCs may display similar moving characteristics post-landfall. FMM has long been applied to cluster analysis. In statistics, samples belonging to different statistical distributions are, at times, mixed. It is somehow difficult to establish the distributions and estimate the parameters for the differentiation of these distributions. TC tracks are assumed to be generated from several unknown distributions (i.e., clusters), that must be identified in order to implement a cluster analysis based on the post-landfall TC samples. In this thesis, FMM is used for the clustering of the post-landfall tracks of TCs that made landfall over the Chinese coast.

As discussed in Everitt and Hand (1981), a mixture density is defined as follows. Let g(x;θ) be a d-dimensional probability density function depending on an m-dimensional parameter vector θ and let $H(\theta)$ be an m-dimensional cumulative distribution function. Then

$$f(x) = \int g(x;\theta)\, dH(\theta) \tag{3.10}$$

where $H(\theta)$ is called the mixing distribution. This definition is perfectly general but most applications are concerned with a subset of this general definition. This subset appears when H is discrete and assigns positive probability to only a finite number of points $\{\theta_i;$

i=1… c}. The finite mixture is then produced by summing up the discrete points to replace the integral in equation 3.10.

$$f(x) = \sum_{i=1}^{c} H_i(\theta_i)\ g(x;\theta_i) \qquad (3.11)$$

The parameters in equation 3.11 consist of three types: c, the number of components in the finite mixture, the mixture proportions $H_i(\theta_i)$ (which is usually denoted by p_i), and the component parameter vectors θ_i. After Pearson (1894) attempted to estimate the five parameters in a mixture of two normal distributions using moments estimation, a variety of estimation methods have been devised and used for the estimation of the parameters of mixture distributions, varying from maximum likelihood approaches, through informal graphical techniques and Bayesian estimation, to the powerful, widely-used Expectation-Maximum (EM) algorithm (McLachlan and Krishnan, 1997).

FMM has been extensively applied to unravel structures and patterns from databases (Biernacki *et al.*, 2002; Everitt, 1988; Leung *et al.*, 2002; Leung *et al.*, 2006; McLachlan *et al.*, 2002). However, the finite mixture framework was not used for TC track clustering until Camargo et al. (Camargo *et al.*, 2007a, b) employed the FMM for clustering TC tracks in WNP.

The model used in Camargo et al. (2007a, b) is

$$y_i = X_i\beta + \varepsilon_i \quad \varepsilon_i \sim \mathcal{N}(0, \Sigma_k) \qquad (3.12)$$

$$P(y_i \mid X_i) = \mathcal{N}(X_i\beta, \Sigma_k) \qquad (3.13)$$

where X_i is the regression matrix with an n×p dimension. β denotes the regression coefficients. y_i is the i^{th} TC track, consisting of n latitude×longitude points (i.e., observations). ε_i is an $n_i \times 2$ matrix of multivariate Gaussian noise, with a zero mean and a 2×2 covariance matrix Σ_k, which contains the diagonal elements σ_{k1}^2 and σ_{k2}^2 (the noise variances for each longitude and latitude observation).

$$X_i = \begin{pmatrix} 1 & \cdots & x_{i1}^{p} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{in}^{p} \end{pmatrix} \qquad (3.14)$$

In order to take into account other variables (e.g., elevation, central pressure, zonal and meridional wind fields, and the indices of the monsoon or the subtropical high), the

values of these variables at each TC point are integrated into X_i. Therefore, the tuned clustering model changes its X_i into

$$T_i = \begin{pmatrix} 1 & \cdots & x_{i1}^p Z_{11} & \cdots & Z_{k1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & \cdots & x_{in}^p Z_{1n} & \cdots & Z_{kn} \end{pmatrix} \quad (3.15)$$

Here, T_i (the modification of X_i) becomes an n× (p+k+1) matrix and β turns out to be a (p+k+1)×2 matrix of regression coefficients. Therefore, the model is modified into

$$y_i = T_i\beta + \varepsilon_i \quad \varepsilon_i \sim \mathcal{N}(0, \sum_k) \quad (3.16)$$

The conditional density for the i^{th} TC track, depending on membership of the k^{th} cluster, is then defined as

$$p(y_i|t_i, \theta_k) = f(y_i|T_i\beta_k, \sum_k)$$

$$= (2\pi)^{-n_i}|\sum_k|^{-n_i/2} exp\{-\frac{1}{2} tr[(y_i - T_i\beta_k)\sum_k^{-1}(y_i - T_i\beta_k)]'\} \quad (3.17)$$

where $\theta_k = \{\beta_k, \sum_k\}$.

This leads to the regression mixture model with K clusters

$$p(y_i|t_i, \phi) = \sum_k^K \alpha_k p_k (y_i|t_i, \theta_k) = \sum_k^K \alpha_k f_k (y_i|T_i\beta_k, \sum_k), \quad (3.18)$$

where α_k is the proportion of a randomly chosen TC track belonging to cluster k (it is noted that $\sum_k \alpha_k = 1$). Here ϕ represents the overall set of mixture parameters $(\alpha_k, \beta_k, \text{and } \sum_k)$.

Providing $Y = \{y_1, \ldots, y_n\}$ is the complete set of n TC tracks and $T = \{t_1, \ldots, t_n\}$ is the set of associated measurement times and related meteorological variables, then the full probability density of Y given T is

$$p(Y|T, \phi) = \prod_i^n \sum_k^K \alpha_k f_k(y_i|T_i\beta_k, \sum_k) \quad (3.19)$$

The parameters are estimated by maximizing this likelihood expression to implement clustering. The EM algorithm finds the local maximum likelihood through iterative estimations. After the model has been learnt, one of the K models with which each TC track is most probably associated can be inferred.

The E-step and M-step form the EM algorithm. In the E-step, we calculate the membership probability, that TC track i belongs to cluster k as follows

$$w_{ik} = \frac{\alpha_k f_k(y_i|T_i\beta_k,\Sigma_k)}{\sum_j^K \alpha_j f_j(y_i|T_i\beta_j,\Sigma_j)} \tag{3.20}$$

It is noteworthy that w_{ik} is equal to the ratio of the likelihood of trajectory i under cluster k, to the total likelihood of TC track i under all clusters. The $w_{ik} = w_{ik}\,\mathbf{I}_{n_i}$, where $\mathbf{I}_{n_i}$ is an n_i vector of ones. Let $W_k = \mathrm{diag}(w'_{1k}, \ldots, w'_{nk})$ be an N×N diagonal matrix. In the M-step, we use W_k to calculate the mixture parameters. The process of calculation is almost equivalent to the weighted least squares in regression analysis. The estimated parameters are

$$\begin{aligned}
\widehat{\beta_k} &= (X'W_kX)^{-1}X'W_kY \\
\widehat{\Sigma_k} &= \frac{(Y-X\widehat{\beta_k})'W_k(Y-X\widehat{\beta_k})}{\sum_i^n w_{ik}} \\
\widehat{\alpha_k} &= \frac{1}{n}\sum_i^n w_{ik}
\end{aligned} \tag{3.21}$$

The EM algorithm iteratively performs E and M steps until convergence. Convergence is identified when the ratio of the incremental improvement in log-likelihood to the initial incremental improvement during the second iteration drops below a certain threshold (e.g., 1×10^{-6}). The algorithm randomly selects a set of membership weights W_k and performs the M step. The solution with the highest likelihood obtained from 20 starts of the EM algorithm was used.

Another problem related to the clustering is the selection of the number of clusters. In this study, the number of clusters is determined using cross-validation, and is decided when the maximum likelihood is reached, given a specified number of clusters. This is discussed in detail in Chapter 6.

3.2.3 Statistical Methods

It is an essential step to test whether there is a significant difference between certain independent samples. For example, is there a significant difference between the lifespans of recurving and straight-moving TCs? Is TC intensity significantly different prior to and post-landfall? Are the lifespans, elevations at the TC center and intensities of several TC clusters significantly different? Do the recurving and straight-moving TCs sustain for the

same time period after attaining their maximum intensity? Which group sustains for longer? Because these problems are of great scientific concern, the Student's *t* test and ANOVA are employed to verify the relationship between them.

a. The Student's *t* test

The Student's *t* test is used to test whether the difference between two groups is significantly positive or negative. Therefore, it plays an essential role in testing, for example, whether the sustaining time of the maximum intensity of recurving TCs is significantly longer than that of straight-moving TCs.

The Student's *t* test is used to test the null hypothesis (H_0) that the two normally distributed populations have equal means (Efron, 1969; Sokal and Rohlf, 1969; Zar, 1974). This test is commonly referred to as an "independent samples" t-test due to the fact that it is mostly implemented when the statistics derived from the two samples under comparison are not overlapped.

If the two populations are assumed to have equal variances, the t statistic is calculated as follows

$$t = \frac{\overline{X_1} - \overline{X_2}}{S_{x_1 x_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n2}}} \tag{3.22}$$

where $\overline{X_1}$ and $\overline{X_2}$ are the sample means of X_1 and X_2 and n_1 and n_2 are the sample sizes of group 1 and group 2 respectively. The numerator of the t statistic is the difference between the two means or averages, while the denominator is taken to be a measure of the variability or dispersion of the value of X.

$$S_{x_1 x_2} = \sqrt{\frac{(n_1 - 1)S_{x_1}^2 + (n_2 - 1)S_{x_2}^2}{n_1 + n_2 - 2}} \tag{3.23}$$

where $S_{x_1}^2$ and $S_{x_2}^2$ are the variances of two samples respectively. It is noteworthy that $S_{x_1 x_2}$ represents the estimation of the standard deviation of the two samples. In this way, its square is the unbiased estimation of the variance. The total number of degrees of freedom is defined as the total sample size minus two (i.e., $n_1 + n_2 - 2$) in the significance testing. Given the two-group samples, we accept or reject the null hypothesis by comparing the *t* statistic with the critical t value at a selected significance level (e.g., 0.05).

b. Analysis of Variance (ANOVA)

ANOVA is used to test whether there is a significant difference between several groups. It is utilized to test, for example, whether or not the average TC intensities prior to and post-landfall differ significantly. It is also employed to test whether the lifespan, elevations at the TC center and intensities of different TC clusters differ significantly. Different time periods prior to or post-TC recurvature, as well as TC landfall, are tested via ANOVA.

ANOVA is a set of general statistical models and their relevant procedures, in which the observed variance in a variable of interest is subdivided into components ascribed to different sources of variation (Neter et al., 1990). ANOVA carries out a statistical test of whether or not the means of several groups are all equal, and thus extends the t test to more than two groups in its simplest form (Neter et al., 1990).

If sample observations are obtained from K groups ($K \geq 2$), and we attempt to compare two or more sample means based on interval or ratio data, the statistical hypotheses are

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$

H_1: At least two of the μ_i are unequal.

The logic of ANOVA depends on the comparison of the variance between groups, as measured by the difference between the sample means, and the variance within groups, as measured by the value of variance between each sample's mean averages across all groups.

For example, in a one-way ANOVA, statistical significance is tested for by comparing the F test statistic, since the total sum of squares S (variance) is partitioned into components related to the effects used in the model (Bray and Maxwell, 1985; Girden, 1992; Harris, 1994; Kim and Kohout, 1975; Silk, 1981).

A one-way ANOVA represents that only one factor or variable of interest is considered. The total variation is thus subdivided into two parts – explained by this factor and unexplained by this factor (Silk, 1981). The total variation in the observations may be represented by the total sum of the squares (SS_T) (it is noted that SS = sum of squares), which may be subdivided into two components, the between-group sum of squares (SS_A) (i.e., the SS attributed to factor A) and the within-group or error sum of squares (SS_E), according to the equations:

$$\sum_i \sum_j (Y_{ij} - \bar{Y})^2 = \sum_i n_i (Y_i - \bar{Y})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_i)^2 \qquad (3.24)$$

$$(SS_T) \quad = \quad (SS_A) \quad + \quad (SS_E)$$

where SS_T represents the sum of the squared deviations of each sampling observation from the sample grand mean (the mean of all the sampling observations), SS_A the sum of the squared deviations between each sample group mean and the sample grand mean multiplying the sample size (n_i) of each group, and SS_E the sum of the squared deviations of individual observations from their sample means (Silk, 1981).

If H_0 is assumed to be true, the sample means of Y_1 and Y_2 and Y_3 indicate the estimations of the population mean μ, and SS_A represents the unbiased estimation of the variance, σ^2, of the population. Therefore, the between-group variance is:

$$S_B^2 = \frac{SS_A}{K-1} \qquad (3.25)$$

where K is the number of groups, so (K-1) is the degrees of freedom associated with this variance estimation. SS_E is another unbiased estimation of σ^2 related to the error variance:

$$S_E^2 = \frac{SS_E}{N-K} \qquad (3.26)$$

where N indicates the total number of sampling observations, and (N-K) is the degree of freedom.

It is shown that the ratio:

$$F_{x_1:x_2} = \frac{S_B^2}{S_E^2} \qquad (3.27)$$

is as an F distribution with x_1=K-1 and x_2=N-K degrees of freedom when the null hypothesis is true. Given the numbers of observations and groups, the degrees of freedom would be known. The remaining task is to test whether the F ratio exceeds the critical F value at a certain significance level (e.g., 0.05).

The characteristics of different TC groups (e.g., recurving and straight-moving TCs, landfall and non-landfall TCs, TCs attaining maximum intensity prior to or post-recurvature) can be found through the Student's *t* test and ANOVA, given a certain significance level.

3.3 Summary

In this chapter, research datasets including JMA RSMC TC best track data, the NCEP/NCAR reanalysis dataset, the NCEP FNL dataset, TRMM data and USGS DEM are introduced. The data mining methods, including C4.5 algorithm, CART, FMM-based clustering, and statistical methods (e.g., The Student's *t* test and ANOVA), are also briefly discussed. These methods are applied in order to mine knowledge from the given database in subsequent chapters.

Chapter 4: TC Recurvature

4.1 Introduction

TC recurvature and landfall pertaining to TC motion are of significant social and economic concern. Until recently, the largest TC track prediction error was still caused by TC recurvature, especially sudden or sharp recurvature. Recurvature signifies a special type of TC track, turning from westward towards the north and eventually to the northeast in the Northern Hemisphere. In general, large track forecast errors typically occur when a TC that has been predicted to recurve continues on a westward track. Similarly, the sudden recurvature of the TC that was predicted to move in a westward direction can engender large forecasting errors. Since nearly half of all WNP TCs eventually recurve, forecasters are frequently challenged by these potential recurvature situations (Dobos and Elsberry, 1993). Therefore, a comprehensive understanding of the mechanisms and rules stipulating TC recurvature is of great theoretic and operational significance. The results will also play an essential role in disaster preparation, management and mitigation. Furthermore, the accurate and real-time prediction for TC recurvature can issue timely forewarnings for people living in coastal regions.

TCs cause most damage during or after landfall that is taken to be the process of the entire TC system moving over land (Tuleya *et al.*, 1984). TCs forming in SCS and WNP are steered northward or northwestward when the subtropical high is strengthened and shifts westward. These TCs may make landfall along the Chinese coast if the steering current is persistently strong and westward. However, westward-moving TCs will turn towards the north and then the northeast if the steering current changes its direction from westward to eastward due to the interactions between large-scale circulation (e.g., the subtropical high, monsoonal systems and mid-latitude westerlies). TCs will move far away from the Chinese coast without making landfall over it due to the reversion of the steering flow. TC recurvature and landfall are both dramatically influenced by large-scale circulation. The two problems are, therefore, investigated together through data mining

methods and shown in this chapter and Chapter 5 given their similarities in internal mechanisms and the intimate relationships. Chapter 4 and Chapter 5 aim to discover from the archived historical TC data the rules governing the TC recurvature and landfall and present the results on TC recurvature and landfall respectively.

Given the aforementioned discussions on TC recurvature in section 2.2, observational analyses and dynamic modeling have been applied to TC recurvature research for several decades. Nonetheless, TC recurvature still engenders the largest errors in TC track prediction. The abundance of data, together with the requirement for powerful data analysis techniques and tools, has been regarded as a data rich but information poor situation. The TC-related database is particularly difficult to visualize and understand because the data can grow along two dimensions: the number of fields (also depicted as attributes) and the number of cases. The rates of growth of TC-related data sets far outstrip the amount of information that traditional "manual" analysis methods can handle. Useful knowledge (e.g., structures, processes, relationships, regularities and patterns) underlying the TC recurvature mechanism is often hidden in the historical TC database. Therefore, the data mining approach is suitable for unraveling from the TC database rules governing the recurvature of TCs throughout their movements.

In recent years, a number of data mining algorithms have been employed to unravel TC tracks and intensities from the TC. However, with regard to these researches, the emphasis is placed on the application of clustering methods to the historical TC tracks and only a few data mining derived results have been successfully applied to analyze TC movement. Tree-based algorithms are widely used for the classification due largely to their simplicity, interpretability, self-explanatory nature, adaptability to processing datasets with errors or missing values and ability to unravel rules and regularities (Breiman *et al.*, 1984b; Friedl and Brodley, 1997; Quinlan, 1987, 1993).However, decision-tree methods have seldom been employed to analyze TC recurvature. TC recurvature study is treated as a binary classification problem (recurvature and non-recurvature) in this chapter. The primary objective of this chapter is, therefore, to unravel the knowledge and rules governing TC recurvature through a decision tree based

algorithm – the C4.5 algorithm – for classification (Quinlan, 1993) from the macroscopic (e.g., the subtropical high, monsoonal systems, and westerlies) and microscopic perspective (i.e., outer radius environmental wind fields around TCs). This objective is attained by using the C4.5 algorithm to analyze all of the potential features that may influence TC recurvature.

This chapter is organized as follows. Section 4.2 introduces the study area and data source. Methodology is discussed in section 4.3. Section 4.4 explains and analyzes the results generated by C4.5 algorithm. The conclusion is drawn in section 4.5.

4.2 Study Area and Data Source

All of the TCs that occurred in the study area from 2000-2010 are employed as samples. The TC-related dataset consists of two major classes: TC best track data and meteorological data.

The TC best track data are made available from JMA RSMC Tokyo. These post-analysis best track data consist of 6-hourly estimates of position (latitude and longitude), MCP, and the 10-min MSW of all of the TCs in the WNP basin, including SCS from 1951 to the present. There are 2552 sampling TC points obtained from 102 TC tracks, 63 of which are of recurved tracks and the rest are of straight-moving tracks. It should be noted that the recurvature points of 63 recurving TC tracks are in the ocean. In this study, TCs that recurved over land are excluded because the mechanisms for recurvature over the ocean and over land differ due to typhoon-land interaction and the cutting off of water vapor for post-landfall TCs. Thus, our TC samples comprise recurvers over the ocean and non-recurvers (straight-movers). The recurvers should retain certain characteristics, as follows: (1) the TC sustains for at least for 72 hours in its best track; (2) the recurvature point has the minimum longitude among the entire track; and (3) the latitude of the subsequent point cannot be less than the latitude of the recurvature point (i.e., no looping tracks or equator-ward recurving TC tracks). Of the recurving TC tracks, 1560

observations prior to TC recurvature are employed as samples and labeled class "0". For the straight-moving TC tracks, all 992 observations are labeled class "1".

The meteorological variables (e.g., wind fields and geo-potential height in different atmospheric layers) are derived from the NCEP GFS FNL at 6-hour time intervals from 30/07/1999 to the present (Yang et al., 2006). These NCEP FNL Operational Global Analysis data are on continuous 1.0° x 1.0° degree grids for every six hours. This GFS is run four times a day in near-real time at NCEP. The data are made available on the surface, at 26 mandatory (and other pressure) levels, varying from 1000 hPa to 10 hPa, in the surface boundary layer and at some sigma levels, the tropopause and a few others. The parameters include SLP, relative vorticity, surface pressure, relative humidity, vertical velocity, geopotential height, SST, soil values, zonal (u-) and meridional (v-) winds and ozone parameters. Although the NCEP/NCAR reanalysis data (Kalnay et al., 1996) are available for a longer time period, the resolution of a 2.5° x 2.5° degree of NCEP/NCAR is coarser than that of the NCEP FNL dataset. Therefore, we derive the meteorological variables from the NCEP FNL dataset other than NCEP/NCAR dataset. Due to the higher resolution and time-period availability of the NCEP FNL dataset, the TC samples are only extracted from 2000-2010.

4.3 Methodology

The C4.5 algorithm is employed to unravel the rules for TC recurvature from the potential factors that may influence this phenomenon. These factors are categorized into three groups: the variables relating to large-scale circulation, the variables measuring the circulation surrounding TCs, and the variables characterizing TCs (see Table 4-1). In Table 4-1, the variables are displayed in abbreviations. In the group "Circulation surrounding TC", "uwnd_200" and "vwnd_200" are respectively the average zonal and meridional wind of 6 - 8° radial belts at the 200hPa level. The other variables in the same group are defined likewise. The five variables chosen to measure the strength and position of the large-scale circulation (i.e., the subtropical high, EASM, westerlies) that largely control TC recurvature (e.g., Harr and Elsberry, 1991, 1995a; Hodanish and Gray, 1993) are: area index (area_IndexSTH), intensity index (inten_IndexSTH), and westward

extension index of the subtropical high (west_extSTH) in WNP (for measuring the strength of the subtropical high and the position of the subtropical high ridge), the EASM index by Wang and Fan (1999) (Monsoon_WF), as well as the westerly index (W_Westerly). In what follows, we first examine these potential factors and then give a brief discussion of C4.5 algorithm.

Table 4-1 The three-group potential attributes influencing TC recurvature

Groups	*Potential Variables*
Circulation surrounding TC	uwnd_200, uwnd_300, uwnd_400, uwnd_500, uwnd_600, uwnd_700, uwnd_800, uwnd_850, uwnd_1000, vwnd_200, vwnd_300, vwnd_400, vwnd_500, vwnd_600, vwnd_700, vwnd_800, vwnd_850, vwnd_1000
Large-scale circulation	area_IndexSTH, inten_IndexSTH west_extSTH, Monsoon_WF W_Westerly
Characteristics of TC	Lon, Lat, Pressure (Central Pressure of TC Center)

4.3.1 Indices for Large-scale Circulation

A number of indices have been proposed to measure the position and intensity of large-scale circulation (e.g., the monsoon systems, subtropical high and mid-latitude westerlies) in the SCS and WNP. The indices depicting the large-scale circulation belong to the potential parameters that influence TC recurvature.

The East Asian Summer Monsoon (EASM) is the most influential and essential component of the Asian climate systems (Chang et al., 2000; Chen and Yoon, 2000; Ding, 1992, 2007; Ding and Chan, 2005; Lau et al., 1988; Wang and Li, 2004; Wang et al., 2008) in the study area due largely to orographical forcing: huge thermal contrasts

between Eurasia and the Pacific, which are the world's largest continent and ocean basin respectively. It is also remarkably affected by the world's highest mountain - the Tibetan Plateau. Meteorologically, there are five categories of EASM indices, namely the "east-west thermal contrast" index and the "north-south thermal contrast" index that are constructed according to the vertical shear of zonal winds, the shear vorticity (often expressed by a north-south gradient of the zonal winds), the "southwest monsoon" indices, which directly measure the strength of the low-level East Asian monsoon winds using the 850-hPa southwesterly flow, and the "South China Sea monsoon" indices (Wang et al., 2008). In Wang et al. (2008), 25 existing EASM indices were examined in terms of two observed major modes of inter-annual variation in the precipitation and circulation anomalies from 1979-2006. They ultimately recommend an index, the reversed Wang and Fan index (Wang and Fan, 1999) (hereafter WF), which is almost "identical to the leading principal component of the EASM and greatly facilitates real-time monitoring". The Wang and Fan index belongs to the "shear vorticity" index and was first proposed to quantify the variability of the WNP summer monsoon. This index was defined by the U850 in the region (5°–15°N, 90°–130°E) minus U850 in the region (22.5°–32.5°N, 110°–140°E), where U850 represents the 850 hPa zonal wind. Physically, the WF shear vorticity index quantifies well the variations in both the subtropical high and WNP monsoon trough, which are the crucial components of the EASM circulation system (Tao and Chen, 1987). Therefore, the WF shear vorticity index is chosen to measure the status of EASM, which places significant influences on TC recurvature (Carr and Elsberry, 1995; Chen, 2009; Harr et al., 1996; Lander, 1994, 1996; Mao and Wu, 2008).

The Western North Pacific Subtropical High (WNPSH) is the large-scale circulation exerting the most dramatic influence on TC movement, especially TC recurvature through the steering flow (e.g., Chen *et al.*, 2009; Evans *et al.*, 1991; George and Gray, 1977; Harr and Elsberry, 1991, 1995a; Holland and Wang, 1995). Kasahara (1959) first noted that the position of the subtropical high plays a crucial role in TC movement. He suggested that a vector mean of the 500 and 700 hPa steering flows should be used for forecasting TC movement. The 500 hPa geopotential heights have been widely used to

measure the WNPSH (e.g., Sui et al., 2007; Sun and Ying, 1999; Zhang and Yu, 1998; Zhou et al., 2009). Besides, the National Climate Center in China (NCC) announced the monthly indices of westward extension as well as the north edge and intensity of the WNPSH, based on monthly mean 500 hPa geopotential heights.

The WNPSH indices announced by NCC have been used in previous studies (e.g., Chen et al., 2001a; Chen et al., 2001b). The NCC defined indices to describe the status of the WNPSH. These indices include the WNPSH intensity index, westward extension index, and area index according to the mean 500 hPa geopotential heights in the weather charts published by the China Meteorological Administration (Chen, 1999). By considering the pre-existing subtropical high indices defined by NCC and other scholars (e.g., Lu, 2001; Lu *et al.*, 2007), we define the WNPSH indices as follows: intensity index is defined as the average geopotential height of the points with their geopotential height larger than 5870 gpm in a rectangular region (100 °E to 180 °E and 10 °N to 60 °N); the WNPSH area index is defined as the number of grids with their geopotential height larger than 5870 gpm in this region; the "west extension index" is defined as the longitude at the western edge of the subtropical high. This definition of the "westward extension index" is in line with that of NCC. The three indices are extracted and calculated from the FNL reanalysis data to measure the WNPSH.

The mid-latitude westerlies also have a significant impact on TC recurvature through upper-tropospheric zonal winds (~200 hPa) (e.g., George and Gray, 1977; Guard, 1977); for instance, it is observed that, when the base lowers considerably west of a TC in connection with an eastward-shift mid-latitude trough, and remains low, northward recurvature will occur (Riehl and Shafer, 1944). Rossby (1939) defined a typical westerly index derived from the 500 hPa geopotential height: IR = H35 – H55. It should be noted that IR is the westerly index defined by Rossby, and H35 and H55 represent the 500 hPa geopotential height at 35°N and 55°N respectively. Here, we calculate the IR from 100 °E to 180 °E in longitude to measure the status of the mid-latitude westerlies, and the larger the IR, the stronger the westerlies.

4.3.2 Outer Radius Flow Surrounding TCs

A variety of indices depicting large-scale circulation are presented in the previous subsection. Apart from the large-scale synoptic systems surrounding TCs, the outer radius (e.g., 5-7°, 6-8°) environment wind fields from the TC center also play a significant role in determining whether or not a TC will recurve. Numerous radial circles have been proposed to investigate wind fields in recent decades; for example, Centry (1983) found that, when the mean upper-tropospheric zonal wind (~200hPa) 1500–2000 km northwest and west of a TC exceeds 20 m/s, the TC will recurve at a probability of 80%. Zonal wind profiles at 8° showed that significant changes in the ambient wind fields concur with no immediate effect on the motion of the TC. Chan (1984) showed that the 5-7° radial circle from the TC center is the best region for the indication of the wind fields surrounding TCs. Based on their rawinsonde database, Hodanish and Gray (1993) found that the 1-3 oscants in the 6-8 ° radial circle from the TC center is the best region for the wind fields. Fitzpatrick (1992) verified that these upper tropospheric zonal winds at a 6° radius to the northwest of a TC are probably a crucial factor in determining whether the TC will recurve or remain on a west-northwest course. The result indicated that the composite deep-layer mean (850-300 hPa) wind fields averaged within the radius 5-7° from the TC center provide a significant component of the steering flow for TCs (Fitzpatrick, 1992). In order to choose the most appropriate outer radius for calculating the ambient wind fields, the radial circles, such as 4-6°, 5-7°, 6-8°, 7-9° and the 1-3 oscants in these radial belts, are tested for classification accuracy.

4.3.3 Variables Characterizing TC

As a result of the atmospheric characteristics of different places, TC movements are largely related to the position in which the TCs are located. The longitude and latitude of a TC center are therefore potential factors influencing TC recurvature. Although TC intensity has not proven theoretically to have an influence on TC recurvature, several studies (e.g., Evans and McKinley, 1998; Knaff, 2009; Riehl, 1972) have found a relationship between recurvature and intensity. The longitude, latitude and intensity (here

central pressure: hPa) of the TC center are thus chosen as potential parameters for mining TC recurvature rules.

4.3.4 Determining Parameters

To choose the most appropriate radial belt and algorithm, the C4.5 algorithm and CART are compared with respect to their classification performances. The first column of Table 4-2 represents the widths of the radial circles within which the environmental flows (e.g., 200 hPa and 500 hPa zonal wind fields) are averaged. The second and third columns indicate the classification accuracy by CART when the minimum leaf sizes are 20 and 50 respectively. The fourth and fifth columns show the classification accuracy by the C4.5 algorithm when the leaf sizes are 20 and 50 respectively. From the table, the radial belt (6-8°) outperforms the others and the C4.5 algorithm performs better than CART when the other parameters are the same. Therefore, the 6-8° radial belt is employed to measure the ambient wind flow and the C4.5 algorithm is used to analyze the potential parameters.

Table 4-2 The classifying accuracy by CART and C4.5

Radial Belt	CART (20 OBJ)	CART (50 OBJ)	C4.5 (20 OBJ)	C4.5 (50 OBJ)
5-7°	82.0061 %	78.1394 %	84.8392 %	80.8959 %
5-7° (123 octants)	81.4701 %	78.2542 %	85.0306 %	80.2833 %
6-8°	**81.9296 %**	**78.8285 %**	**85.7198 %**	**84.364%**
6-8(123 octants)	81.6233 %	78.5988 %	85.3752 %	80.513%
7-9°	82.121 %	78.6753 %	85.0306 %	80.2067 %
7-9° (123 octants)	82.0827 %	78.6753 %	84.6861 %	80.7427 %
8-10°	82.0827 %	78.1011 %	84.3798 %	79.5176 %
8-10° (123 octants)	81.3936 %	78.7136 %	84.2266 %	80.9724 %
4-7°	81.6233 %	77.5268 %	84.1884 %	81.049 %
4-7° (123 octants)	81.317 %	77.8714 %	85.3369 %	80.8193 %
5-8°	81.6998 %	78.2542 %	84.6478 %	81.5084 %
5-8° (123 octants)	81.5467 %	78.5988 %	85.5283 %	80.0153 %
6-9°	82.0061 %	78.4839 %	84.3032 %	80.7427 %
6-9° (123 octants)	81.9296 %	78.5605 %	85.4518 %	80.5896 %
7-10°	82.1593 %	78.2542 %	85.0689 %	80.2067 %
7-10° (123 octants)	81.2404 %	78.7519 %	84.4946 %	81.2021 %

4.4 Results and Interpretations

4.4.1 Results

The C4.5 algorithm is implemented in Weka 3.6.2 (a collection of machine learning algorithms for data mining tasks). The algorithms in Weka can either be applied directly to a dataset or called from some user-defined Java codes. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also suitable for the development of new machine learning schemes (the software is open source and available at http://www.cs.waikato.ac.nz/ml/weka/index.html). To lower the complexity of the classification tree, the minimum leaf size is set at '50' because the smaller the minimum leaf size, the more complicated the classification tree becomes. Ten-fold cross validation is used for verification. Other parameters are used according to their default setting. The settings are: "Binary Split" is true; confidence factor is "0.25"; debug is "false"; numFolds are "3"; reducedErrorPruning is "true"; saveInstanceData is "false"; seed is "1"; subtreeRaising is "true"; unpruned is "false"; and uselaplace is "false".

It is worth noting that the used TC samples consist of the TC observations prior to the TC recurvature and observations of all the straight-moving TC. The observations after recurvature may carry useful information and knowledge underlying the mechanisms of TC recurvature. All observations of recurving and straight-moving TCs are classified using C4.5 algorithm. In addition, with respect to the recurving TCs, we can also extract the equal points from straight-moving TCs by the average number of observations within recurving TCs. Therefore, three pairs of TC datasets are classified using the samples from 2000-2009 and verified by TCs in 2010 (Table 4-3). It can be observed that the accuracies of verifications of three TC groups are quite close to those of training accuracies. The accuracies of verifications are larger than 80% among the three TC groups. The classification trees of these three groups are similar to each other with little difference in splitting values. Therefore, the classification results derived from the partial recurving and full straight-moving TCs are discussed in the following sections.

Table 4-3 The classification accuracies of different TC groups

TC Groups	Training Accuracy	Verifying Accuracy
Partial Recurving and Full Straight-moving TCs	83.3640%	80.5970%
Full Recurving and Full Straight-moving TCs	82.2169%	81.6832%
Partial Recurving and Partial Straight-moving TCs	81.1235%	80.4023%

From the result, we can see that Lon, Lat, Pressure, uwnd_200, uwnd_1000, vwnd_800, vwnd_850, vwnd_1000, area_IndexSTH, west_extSTH and Monsoon_WF are chosen by the C4.5 algorithm to build the decision tree (Figure 4-1), stipulating 18 unraveled rules (see Table 4-4) governing TC recurvature. The average accuracy of TC recurvature prediction by the C4.5 algorithm is 84.364%. In the figure, "1" means recurvature and "0" means non-recurvature, respectively. The rectangles are leaf nodes whereas the ellipses or circles are parent nodes. Every path from the root node to a leaf node represents a rule that can be used as a reference for TC recurvature prediction.

Table 4-4. The 18 unraveled rules governing TC recurvature

Rule Number	Rules	Attributes	Accuracy
1	If Lon≤130, and Pressure>996 hPa, then TC will NOT recurve.	Longitude, Central Pressure	251/263=0.954
2	If Lon≤130, and Pressure≤996 hPa, and west_extSTH≤127, and uwnd_1000>0.898, then TC will NOT recurve.	Longitude, Central Pressure, west edge of the subtropical high ridge, 1000 hPa zonal wind	64/92=0.696
3	If Lon≤130, and Pressure≤996 hPa, and west_extSTH≤127, and uwnd_1000≤0.898, vwnd_1000≤ 3.624, then TC will recurve.	Longitude, Central Pressure, west edge of the subtropical high ridge, 1000 hPa zonal and meridional wind	216/270=0.800
4	If Lon≤130, and Pressure≤996 hPa, and west_extSTH≤127, and uwnd_1000≤0.898, and vwnd_1000> 3.624, then TC will NOT recurve.	Longitude, Central Pressure, west edge of the subtropical high ridge, 1000 hPa zonal and meridional wind	35/50=0.700
5	If Lon>130, and Pressure≤1006 hPa, and area_IndexSTH≤314, and uwnd_200≤-3.57, then TC	Longitude, Central Pressure, area index of the subtropical high, 200 hPa zonal wind	44/72=0.611

	will recurve.		
6	If Lon>130, and Pressure≤1006 hPa, area_IndexSTH≤314, and uwnd_200>-3.57, and west_extSTH<133, and Lon>146, then TC will recurve.	Longitude, Central Pressure, west edge of subtropical high ridge, area index of the subtropical high, 200 hPa zonal wind	126/129=0.977
7	If Lon≤130, and Pressure≤996 hPa, and west_extSTH≤127, and Lon≤123, then TC will NOT recurve.	Longitude, Central Pressure, west edge of the subtropical high ridge	266/286=0.930
8	If Lon≤130, and Pressure≤996 hPa, and west_extSTH≤127, and Lon>123, and Monsoon_WF≤-5.168, then TC will recurve.	Longitude, Central Pressure, west edge of the subtropical high ridge, Monsoon Index	50/74=0.676
9	If Lon>130, and Pressure≤1006 hPa, and area_IndexSTH≤314, and uwnd_200>-3.57, and west_extSTH>133, then TC will recurve.	Longitude, Central Pressure, west edge of subtropical high ridge, area index of the subtropical high, 200 hPa zonal wind	591/617=0.958
10	If Lon>130, and Pressure≤1006 hPa, and area_IndexSTH≤314, and uwnd_200>-3.57, and west_extSTH≤133, and Lon≤146, and Monsoon_WF≤-1.732, then TC will recurve.	Longitude, Central Pressure, west edge of the subtropical high ridge, area index of the subtropical high, 200 hPa zonal wind, Monsoon Index	109/120=0.908
11	If Lon≤130, and Pressure≤996 hPa, and west_extSTH≤127, and Lon>123, and Monsoon_WF>-5.168, then TC will NOT recurve.	Longitude, Central Pressure, west edge of the subtropical high ridge, Monsoon Index	88/110=0.800
12	If Lon>130, and Pressure≤1006 hPa, and area_IndexSTH≤314, and uwnd_200>-3.57, and west_extSTH≤133, and Lon≤146, and Monsoon_WF≤-1.732, and vwnd_800>-0.022 then TC will recurve.	Longitude, Central Pressure, west edge of the subtropical high ridge, area index of the subtropical high, 200 hPa zonal wind , Monsoon Index, 800 hPa meridional wind	47/70=0.671

13	If Lon>130, and Pressure≤1006 hPa, and area_IndexSTH>314, and Monsoon_WF>6.241, then TC will NOT recurve.	Longitude, Central Pressure, area index of the subtropical high, Monsoon Index	56/74=0.757
14	If Lon>130, and Pressure≤1006 hPa, and area IndexSTH>314, and Monsoon_WF≤6.241, and Lat>16 then TC will recurve.	Longitude, Central Pressure, area index of the subtropical high, Monsoon Index, Latitude	278/349=0.801
15	If Lon≤130, and Pressure>1006 hPa, and vwnd_850>-1.695, then TC will recurve.	Longitude, Central Pressure, 850 hPa meridional wind	44/73=0.603
16	If Lon≤130, and Pressure>1006 hPa, and vwnd_850≤-1.695, then TC will NOT recurve.	Longitude, Central Pressure, 850 hPa meridional wind	55/60=0.917
17	If Lon>130, and Pressure≤1006 hPa, and area_IndexSTH>314, and Monsoon_WF≤6.241, and Lat ≤16, then TC will NOT recurve.	Longitude, Central Pressure, area index of the subtropical high, Monsoon Index, Latitude	46/87=0.530
18	If Lon>130, and Pressure≤1006 hPa, and area_IndexSTH≤314, and uwnd_200>-3.57, and west_extSTH≤133, and Lon≤146, and Monsoon_WF≤-1.732, and vwnd_800≤-0.022 then TC will recurve.	Longitude, Central Pressure, west edge of the subtropical high ridge, area index of the subtropical high, 200 hPa zonal wind , Monsoon Index, 800 hPa meridional wind	35/50=0.700

Table 4-4 lists 18 rules that are derived from the classification tree (Figure 4-1). It is noteworthy that "longitude" and "pressure" are two variables characterizing TCs (Table 4-1). These two variables are the first and second factors to be chosen in building the decision tree. Rule 1 is decided merely by "longitude" and "pressure". The remaining 17 rules can be subdivided into three groups:

Group 1: rule 2, rule 3, rule 4, rule 5, rule 6 and rule 7. This group is largely determined by the western edge of the subtropical high.

Group 2: rule 8, rule 9, rule 10, rule 11, rule 12, rule 13, rule 14, rule 15 and rule 16. This group is mainly controlled by the high center of the subtropical high. It should be noted that the rules of Group 2 are also modulated by the western edge of the subtropical high and monsoonal systems. However, the area index of the subtropical high is the first factor that is chosen to split the decision tree after "longitude" and "central pressure".

Group 3: rule 17 and rule 18. This group is determined by the mean 850 hPa meridional wind.

Group 1 and Group 2 take up the 15 rules of the remaining 17 rules. It implies that, besides the variables characterizing TC, the subtropical high plays a crucial role in modulating TC recurvature.

Taking leaf node "0(263.0/12.0)" as an example, the "0" before the bracket means non-recurvature and "263.0" and "12.0" indicate that, among the 263 samples of the leaf node, there are 251 (263 minus 12) non-recurvature samples and 12 recurvature samples, respectively. Each rule can be justified by the meteorological processes and theories on TC dynamics. The first column of Table 4-4 indicates the rule numbers. The second column describes the rules unraveled. The third column depicts the attributes contained in each rule. The fourth column shows the classification accuracy of each rule. The accuracy is calculated by dividing the number of samples correctly classified by the total number of samples in the leaf node. From the fourth column, the highest classification accuracy is 0.977 whereas the lowest is 0.530. It is apparent that the "longitude" of the TC center appears in every rule as it is the first attribute to be selected by the C4.5 algorithm. For substantiation, interpretations of some of these rules are made in the following subsection. Longitude (Lon) is chosen for the splitting of the root node. For the two resultant child nodes of the root node, central pressure (the measurement of TC intensity) is the splitting variable.

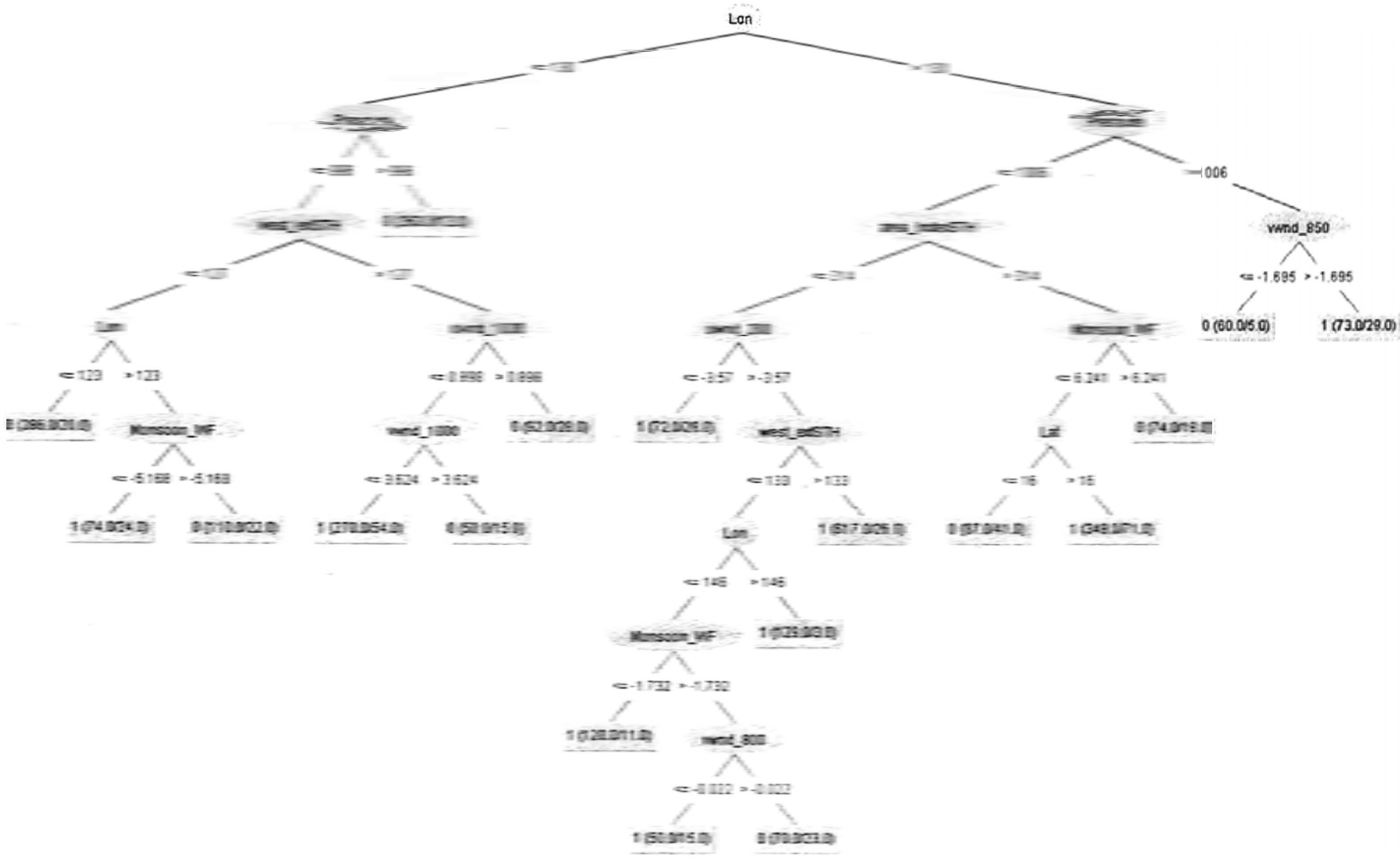


Figure 4-1. Rules governing TC recurvature unravelled by C4.5

Previous investigations have found that some synoptic patterns can identify recurvature; for example, a TC under the influence of a weak subtropical high and southward westerly trough tends to recurve, whereas a strong and westward-shift subtropical high cause the TC to move westward or northwestward. In fact, the tree built by the C4.5 algorithm (Figure 4-1) contains not only the empirical knowledge based on these existing synoptic patterns, but also the rules derived from characteristics such as central pressure, latitude, longitude, and the indices for depicting large-scale circulation. Furthermore, the built classification tree transfers the qualitative empirical rules to quantitative rules characterized by selected variables and splitting values. The rules derived from the classification tree include rules that confirm existing TC theories, and rules that contain new knowledge about TC recurvature. The new rules are: Rule 1, Rule 2, Rule 3, Rule 4, Rule 5, Rule 12, Rule 14 Rule 15, Rule 16, Rule 17 and Rule 18. The rules confirming the existing TC theories are: Rule 6, Rule 7, Rule 8, Rule 9, Rule 10, Rule 11, Rule 13 (Table 4-5).Therefore, this classification tree can provide references for TC recurvature prediction.

4.4.2 Verification and Interpretation

Verification and interpretation are essential components of data mining and always entitled "pattern evaluation" (Han and Kamber, 2006). The rules unraveled by the C4.5 algorithm are interpreted from the meteorological perspective. The decision tree in Figure 4-1 is thus verified using the JMA RSMC TC best track in 2010 to justify whether or not it can correctly classify the newly recurving and non-recurving TCs.

The decision tree built by the C4.5 algorithm is verified using the 2010 RSMC JMA TC best track data, which are not involved in training the classification model. Of the fourteen TCs occurring in 2010, seven TC tracks with 134 observations were selected for verification. The requirements for the testing dataset are consistent with those for selecting training datasets, as discussed in section 4.2. Of the seven selected tracks, four are recurvers and three are straight-movers. The fundamental information about the test

TC tracks are shown in Table 4-5. The total classification accuracy of the decision tree is 80.597%, which is only slightly lower than the 84.364% for training accuracy. Table 4-6 shows the confusion matrix of verification. The non-recurvature TCs are all correctly classified, whereas 26 TC points of recurving TCs are classified as non-recurvature. It can be inferred that the straight-movers are easier to predict than recurvers.

Table 4-5 Selected TC tracks in 2010 for Verification

Names of TCs in 2010	**Recurvature**	**Time of Genesis**
CONSON	No	2010071112
CHANTHU	No	2010071706
MINDULLE	No	2010082200
MALOU	Yes	2010090112
MALAKAS	Yes	2010092006
MEGI	Yes	2010101300
CHABA	Yes	2010102318

Table 4-6 Confusion Matrix of the Verification

	Recurvature	*straight*
Recurvature	42	26
Straight	0	66

TC tracks are largely determined by the steering current, especially the 500 hPa (mid-tropospheric) environmental flow. It has been found that the steering level is greatly influenced by the intensity of the TC. The stronger the intensity of the TC, the deeper (e.g., 300 hPa) the steering level. The deep-layer mean flow is, in general, the pressure-weighted averaged flow from 850 hPa to 300 hPa layer. Therefore, the deep-layer mean flow indicates the steering current that determines the TC movement to a large extent. It has been reckoned that deep-layer means, e.g. 850-300 hPa, are most reliable for forecasting (Holland, 1993b). Therefore, the decision tree with unravelled rules is interpreted from the perspective of the deep-layer mean. Each leaf node has a certain number of samples. The composite deep-layer mean in the leaf node are, therefore, obtained within the 60°×60° longitude and latitude square anchoring at the centers of the

TC samples. The composites of the deep-layer mean wind fields indicate the steering flow around the TC center. The direction of the steering flow can thus affect whether a TC recurves or not.

The rule with the highest classification accuracy of 0.977 is generated by the leaf node '1(129.0/3.0)'. It can be described as follows: "If the longitude of a TC center is to the east of 130°E, and the central pressure of the TC center (TC intensity) is smaller than 1006 hPa, and the area_Index of the subtropical high is smaller than 314, and the average zonal wind within the 6-8° radius in the 200 hPa layer is larger than -3.57 m/s, and the western edge of the subtropical high shifts to the west of 133 °E, and the longitude of the TC center is located to the east of 146°E, then the TC will recurve."

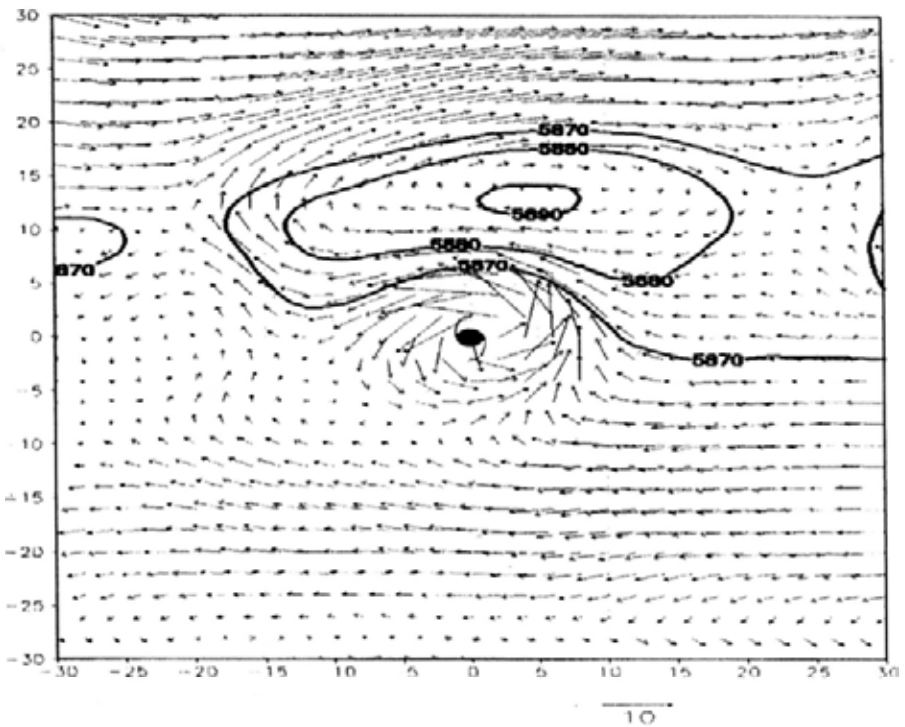


Figure 4-2. The overlay of the composite deep-layer mean wind fields (vector, unit: ms^{-1}) and 500 hPa geopotential heights (contour, unit: gpm) of the TC samples in the leaf node "1(129.0/3.0)". The thick contours represent the high center of the subtropical high. The TC symbol denotes the relative TC center with the coordinate (0, 0). The axes on the bottom and left of the plot represent the x and y ordinates that represent east and north respectively.

As observed in Figure 4-2, the subtropical high is located to the north of the TC center. This is a normal situation, since the TC lies to the east of 146 °E, to its west beyond which the subtropical high always shifts. Supposing the TC moves westward for 20 degrees (east of 126 °E), it will lie to the west of the subtropical high and will be steered by the north-eastward wind in the western part of the subtropical high to recurve. This constitutes the reason why the subtropical high sitting to the north of a TC center still leads to TC recurvature.

The rule with the lowest classification accuracy (0.530) among all the rules is generated by the leaf node '0(87.0/41.0)' that can be described as follows: "If the longitude of the TC center lies to the east of 130°E, and the central pressure of the TC center (TC intensity) is smaller than 1006 hPa, and the area_Index of the subtropical high is larger than 314, and the Monsoon_WF index is smaller than 6.241, and the latitude of the TC center is to the south of 16°N, then the TC will not recurve."

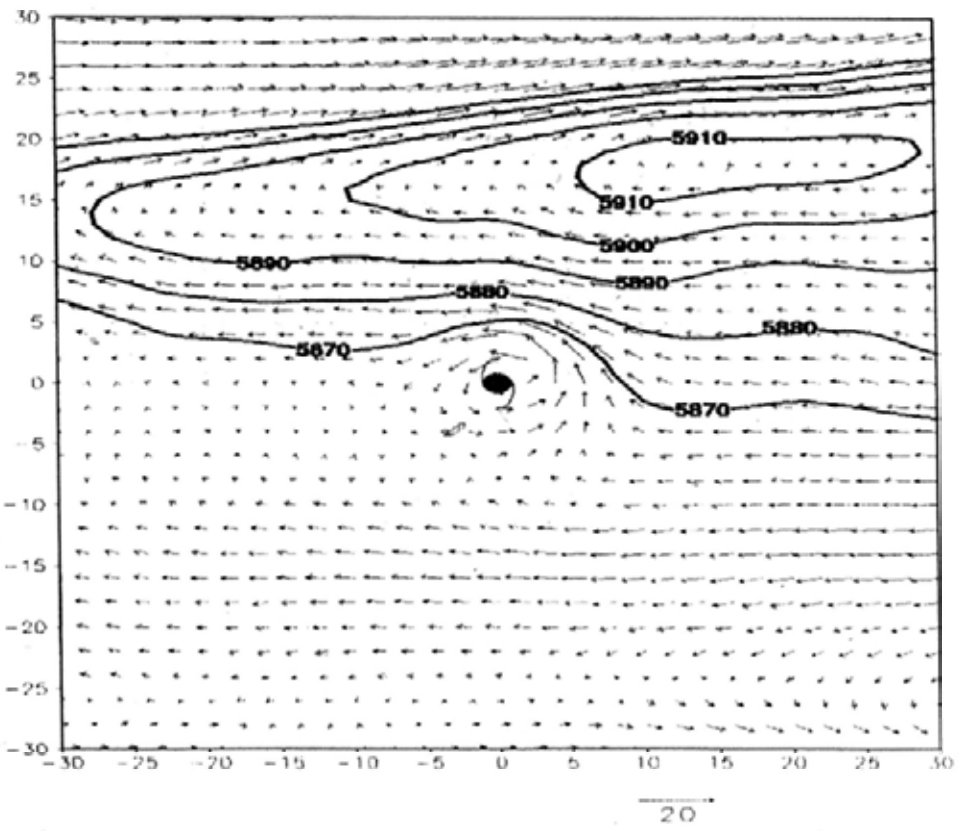


Figure 4-3. As in Figure 4-2 except for the leaf node '0(87.0/41.0)'.

As can be observed in Figure 4-3, the subtropical high is strengthened and displaces to the west of the TC center for more than 30 degrees. The TCs tend to be steered westward under the westward steering current in the southern part of the subtropical high. Therefore, they will move straight under the influence of the easterly flow.

Taking the rule formed by the path from the root node to the leaf node "1(617.0/26.0)" as another example, it can be stated as: "If the longitude of a TC is to the east of 130 °E, and the central pressure is less than 1006 hPa, and the number of grids within the 5870 gpm contour of the 500 hPa geopotential height is less than 314, and the 200 hPa zonal wind averaged within the 6-8° radial belt is mainly eastward, and the western edge of the subtropical high displaces to the east of 133 °E, then the TC will recurve." According to this rule, TCs will recurve under the influence of a weak and retreating subtropical high and moderate westerlies (Figure 4-4). The composite 500 hPa geopotential heights and deep-layer mean wind fields derived from the TC samples belonging to this rule are in agreement with the steering flow (Figure 4-4). The contour of 5870 gpm illustrates the status of the subtropical high. We can observe that the subtropical high retreats to the west of the TC, and the steering flow, illustrated by the deep-layer mean flow from the 850 hPa to 300 hPa layer, tends to turn to the north and northeast. As a result, this steering flow surrounding the TC center leads the TC to recurve.

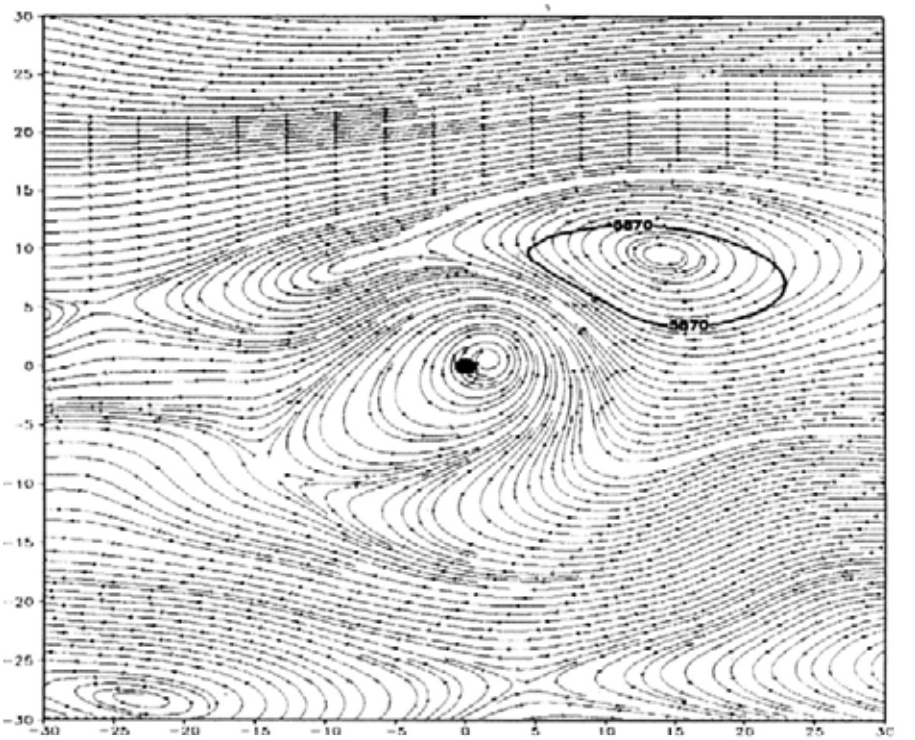
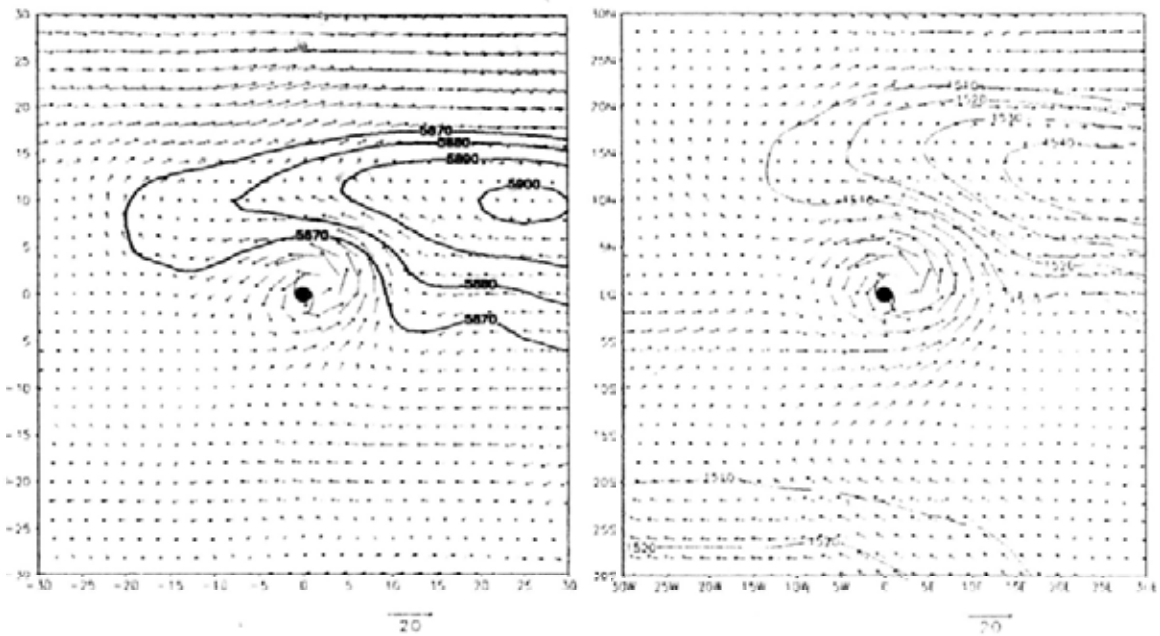


Figure 4-4 As in Figure 4-2 except for the leaf node "0(617.0/26.0)".

The rule derived from the leaf node '1(120/11)' is stated as: "If the longitude of the TC center is to the east of 130°E, and the central pressure of the TC center (TC intensity) is smaller than 1006 hPa, and the area_Index of the subtropical high is smaller than 314, and the average zonal wind within 6-8° radius in 200 hPa layer is larger than -3.57 ms^{-1}, and the western edge of the subtropical high progresses to the west of 133 °E, and the longitude of the TC center is to the west of 146 °E, and the monsoon index (Wang and Fan, 1999) is smaller than -1.732 (representing a strong summer monsoon), then the TC will recurve." The monsoon index used in this study (Wang and Fan, 1999) is negatively correlated with the strength of EASM. Therefore, a low monsoon index (here smaller than -1.732) refers to the relatively high intensity of a summer monsoon. Given the existing findings on monsoon-TC interactions (Carr and Elsberry, 1995; Chen, 2009; Harr and Chan, 2004; Harr et al., 1996; Lander, 1994), the strong summer monsoon in SCS and WNP tends to provide a northward and north-eastward wind component for TC propagation. Therefore, a strong monsoon plays a positive role in TC recurvature here. With regard to this rule, the relatively weak subtropical high, relatively high zonal wind in the 200 hPa layer and strong summer monsoon will cause the TC to recurve. The

composite deep-layer mean is plotted in Figure 4-5. From the figure, although the subtropical high ridge shifts westward beyond the TC center, the TC still tends to recurve due to the strong monsoon (depicted by the wind fields in the 850 hPa layer) and its own longitude (130° E <Lon ≤146 ° E).



(a) (b)

Figure 4-5. The composite wind fields (vectors, unit: ms^{-1}) and geopotential heights (contour, unit: gpm) in the (a) 500 hPa and (b) 850 hPa layers of the samples in the leaf node '1(120.0/11.0)'. The thick contours represent the high center of the subtropical high. The TC symbol denotes the relative TC center with the coordinate (0, 0). The axes on the bottom and left of the plot represent the x and y ordinates that represent east and north respectively.

The rule derived from the leaf node '0(286.0/20.0)' can be stated as: "If a TC moves to the west of 130 °E, and the central pressure of the TC center is smaller than 996 hPa, and the western edge of the subtropical high extends to the west of 127 °E, and the longitude of the TC center is to the west of 123 °E, then the TC will move straight." This rule is mainly associated with the longitude of the TC center and the status of the subtropical high. The longitude at 123° E is quite close to the Chinese coast. The easterly steering flow may cause the TC to move straight. The westward-shift subtropical high provides a

prevailing easterly component for the steering flow in its southern part. Figure 4-6 illustrates the overlay of the composite steering flow and 500 hPa geopotential heights. The steering flow given in Figure 4-6 is westward and the subtropical high shifts to the west of the TC center (depicted by the TC symbol). Under the condition that the TC moves to the west of 123 °E, and the subtropical high displaces westward, the TC will go westward or north-westward under the steering of the easterly flow in the southern part of the subtropical high.

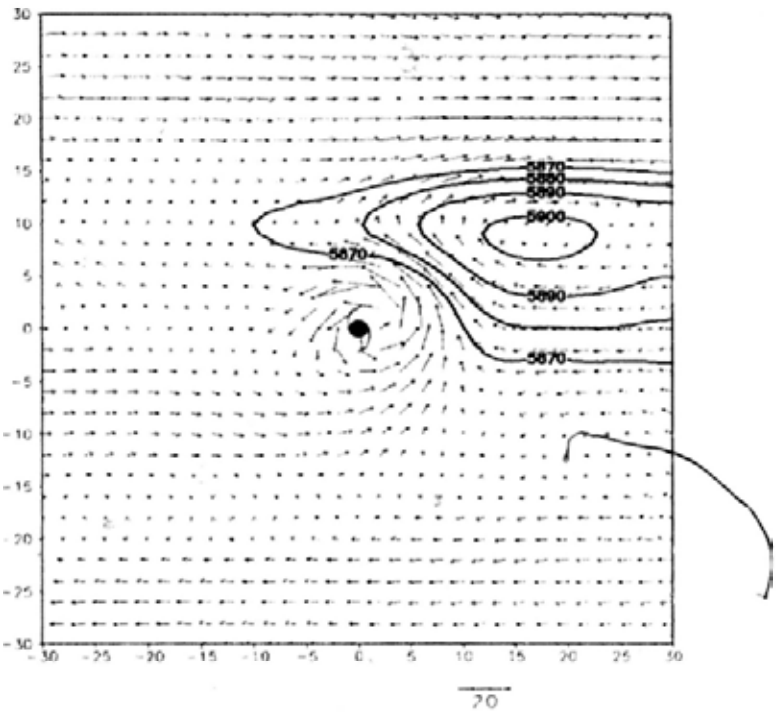


Figure 4-6. As in Figure 4-2 except for the leaf node '0(286.0/20.0)'

The rule obtained from the leaf node "0(74.0/18.0)" is stated as: "If a TC is located to the east of 130 °E, and the pressure of the TC center is smaller than 1006 hPa, and the area index of the subtropical high is larger than 314, and the monsoon index is larger than 6.241 (weak monsoon), then the TC will not recurve." This rule indicates a strong subtropical high, weak monsoon, and moderate TC intensity. The composite deep-layer mean wind and 500 hPa geopotential heights are overlaid in Figure 4-7. In the figure, the subtropical high is significantly strengthened and situated to the north of the TC center. The steering flow surrounding the TC is largely westward. Therefore, the TC tends to move westward under the influence of the strong subtropical high and the weak monsoon (Figure not shown).

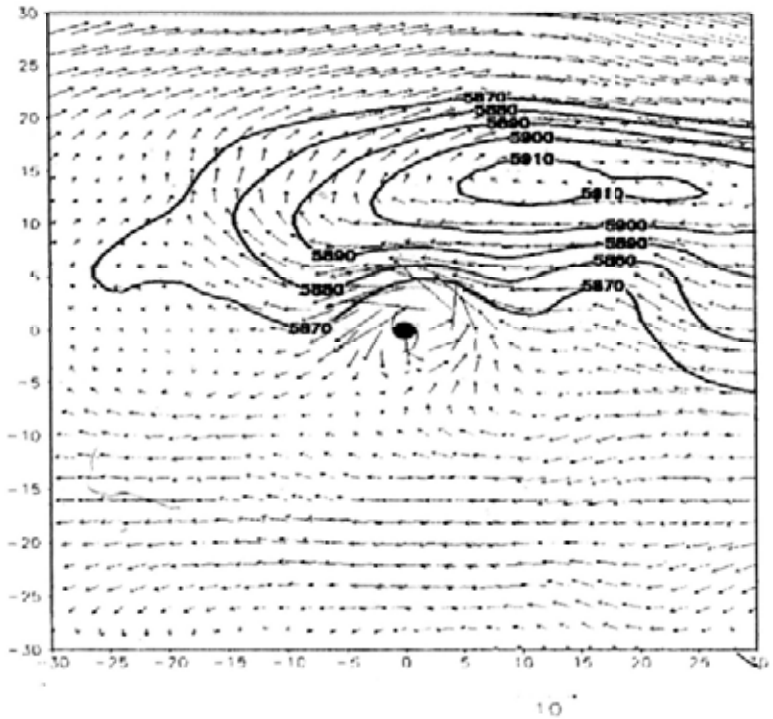


Figure 4-7. As in Figure 4-2 except for the leaf node "1(270.0/54.0)"

The rule derived from the leaf node "1(270.0/54.0)" is stated as: "If a TC moves to the west of 130 °E, and the central pressure of the TC is smaller than 996 hPa, and the western edge of the subtropical high displaces to the east of 127°E, and the average 1000 hPa zonal wind within the 6-8° radius from the TC center is smaller than 0.898 ms^{-1}, and the average meridional wind within the 6-8° radius from the TC center in the 1000 hPa layer is smaller than 3.624 ms^{-1}, then the TC will recurve." This rule includes the low-level wind fields (i.e., 1000 hPa layer). This means that the low-level zonal and meridional winds also influence TC movement to some degree. The rule states that the western edge of the subtropical high retreats and the low-level wind fields are weakened. It is observed that the subtropical high ridge retreats to the east of 127° E and the low-level wind fields are weak (Figure 4-8). The TC moves northward or northeastward along the western part of the subtropical high and eventually recurves.

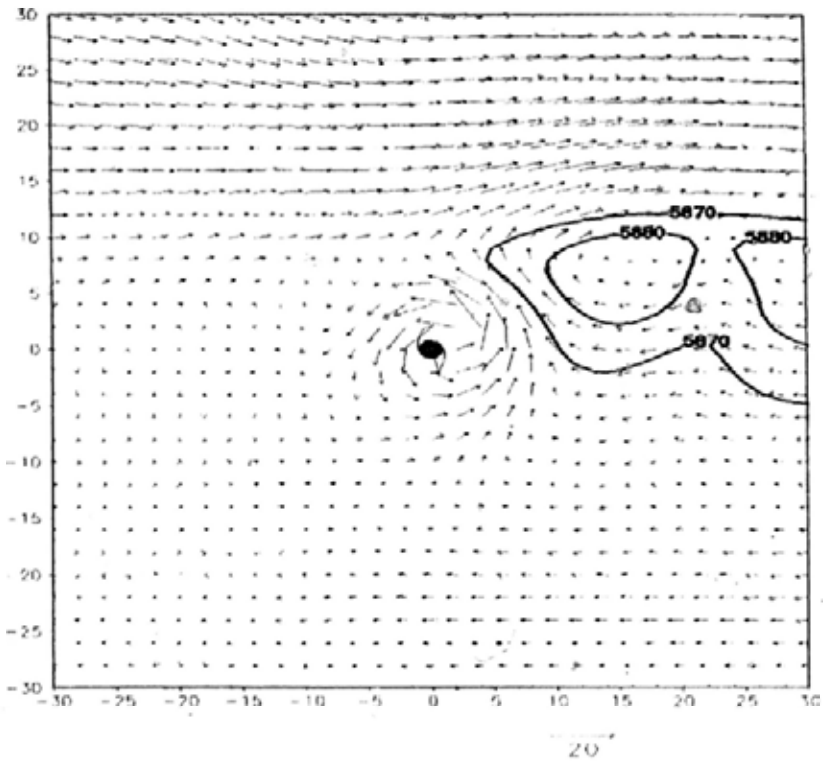


Figure 4-8. As in Figure 4-2 except for the leaf node "1(270.0/54.0)"

The rule obtained from the leaf node "0(110.0/22.0)" can be stated as: "If the longitude of the center of a TC is to the west of 130 °E, and the central pressure of the TC center is smaller than 996 hPa, and the western edge of the subtropical high ridge shifts to the west of 127 °E, and the TC propagates to the east of 123 °E, and the monsoon index is larger than -5.168, then the TC will not recurve." It is indicated by this rule that, if a TC is located to the east of 123 °E, the subtropical high displaces westward, and the monsoon is relatively weak, then the TC will move westward or north-westward due to the strong steering flow (see Figure 4-9). The conditions of the "recurving" leaf node "1(74.0/24.0)" generated by the binary split are similar to those for this rule except that the monsoon index is smaller than -5.168. Due to the strong southerly component of a strong monsoon (refer to Figure 4-10), the rule formed by the leaf node "1(74.0/24.0)" leads to recurvature. This result is in full agreement with the previous findings on TC-monsoon interaction (e.g., Chen *et al.*, 2009; Harr and Elsberry, 1991, 1995a).

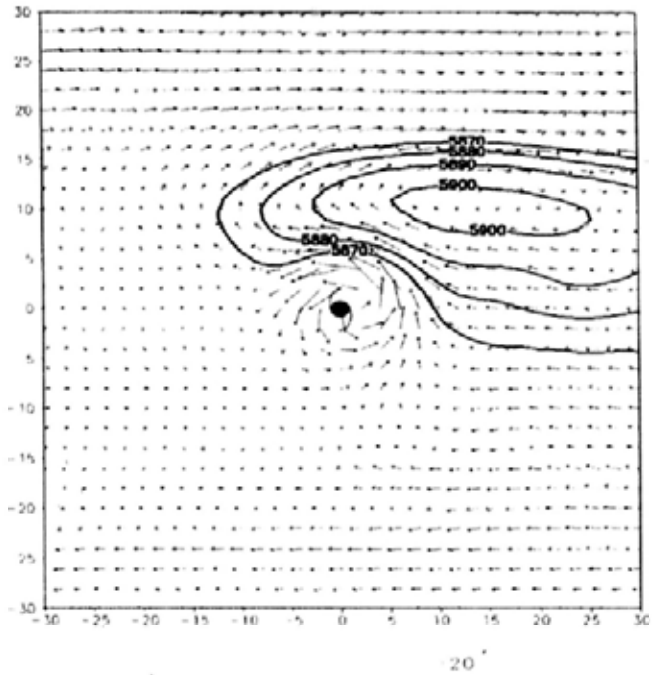


Figure 4-9. As in Figure 4-2 except for the leaf node "0(110.0/22.0)"

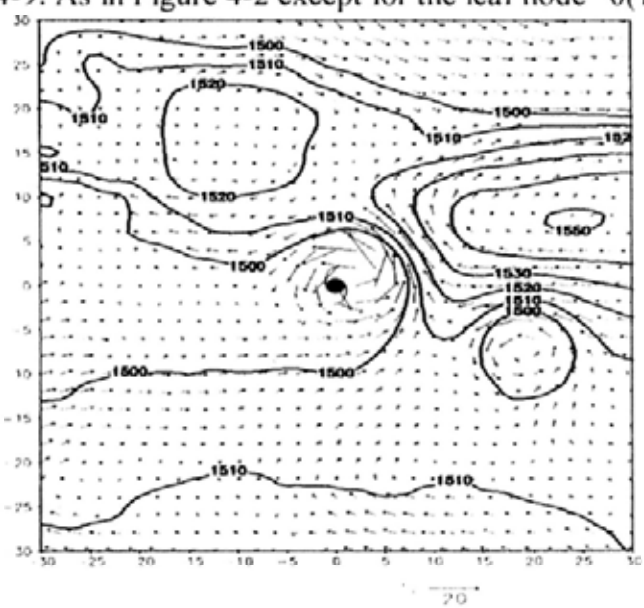


Figure 4-10. The composite deep-layer mean wind fields and geopotential height in the 850 hPa layer of the leaf node "1(74.0/24.0)". The thick contours depict the region with a geopotential height larger than 1500 gpm. The TC symbol denotes the relative TC center with the coordinate (0, 0). The axes on the bottom and left of the plot represent the x and y ordinates that represent east and north respectively.

The rule produced by the leaf node "0(263.0/12.0)" is stated as: "If the TC moves to the west of 130 °E and the intensity of the TC center is larger than 996 hPa, then the TC will not recurve." It is observed in Figure 4-11 that the subtropical high lies to the north of the

TC center. It should be noted that the TC center is adjacent to the Chinese coast (Lon < 130 °E). Therefore, the TC will move westward under the prevailing easterly flow from the southern part of the subtropical high and will not recurve.

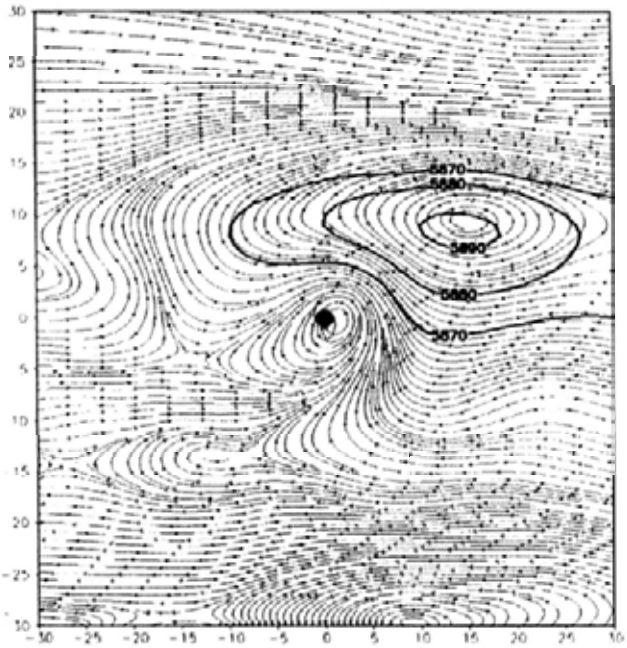


Figure 4-11 As in Figure 4-2 except for the leaf node "1(263.0/12.0)"

Due to the space limitations, not every interpretation of the rules is presented in this study. The eight rules are taken as examples for interpretation. In fact, all of the rules are justified from the perspective of meteorology and are consistent with the existing theories and findings. Apart from the modulation of the subtropical high, monsoon systems, and westerlies on TC recurvature, its intensity, the latitude and longitude of the TC center, and the ambient wind fields also have an impact on TC recurvature. As is consistent with the existing results about the significant modulation of the subtropical high on TC movement (e.g., Elsberry and School, 1987; Evans *et al.*, 1991; George and Gray, 1977; Harr and Elsberry, 1991, 1995a), the west extension index and area index of the subtropical high are selected by the C4.5 algorithm to identify TC recurvature. The monsoon index "Monsoon_WF" (Wang and Fan, 1999) was chosen three times to construct the decision tree (Figure 4-1). This monsoon index should be paid particular attention when measuring the impact of monsoons on TC recurvature. A variety of splitting values in the classification tree can provide references for TC forecasting, since these values can best differentiate the recurving and non-recurving cases based on the

archived TC data. For example, TC recurvature is highly sensitive to longitude "130 °E" because this longitude is first chosen by the C4.5 algorithm to build the classification tree. Besides, longitude "123 °E" should be paid attention in operational aids for TC recurvature. Based on the data mining results, these critical values will provide useful references for TC recurvature forecasting.

4.5 Summary

This chapter presents a detailed study of TC recurvature through data mining over the SCS and WNP. The investigations are particularly focused on unraveling the hidden rules and regularities governing TC recurvature. The historical TC tracks database is comprised of recurving TCs and straight-movers. The TC points of recurving TCs preceding TC recurvature are labeled "1" and those of straight-movers are labeled "0". The investigation becomes a binary classification problem. The potential parameters exerting an influence on TC recurvature are categorized into three groups: circulations surrounding TCs (for example, average 200 hPa zonal and meridional winds within 6-8° radius from the TC center), large-scale circulation (i.e., indices for measuring the subtropical highs, monsoon systems and westerlies) and variables characterizing TCs (i.e., longitude, latitude, and central pressure of the TC center). A classic tree algorithm - C4.5 is applied to the classification. Totally, Lon, Lat, Pressure, uwnd_200, uwnd_1000, vwnd_800, vwnd_850, vwnd_1000, area_IndexSTH, west_extSTH and Monsoon_WF are chosen by the C4.5 algorithm to build the decision tree. The parameters measuring large-scale circulation and those characterizing the TCs play significant roles in building the classification tree. Altogether, 18 rules are unraveled from the processed database.

Most of the 18 rules can be explained by the existing theories and are supported by other empirical findings on TC recurvature (e.g., Chan *et al.*, 1980; George and Gray, 1977; Guard, 1977; Hodanish and Gray, 1993; Holland and Wang, 1995) according to the verification of the results, by calculating and investigating the composite wind fields, geo-potential heights and deep-layer mean wind fields in a square ($60° \times 60°$) centering at the TC centers. These synoptic surrounding wind fields, especially deep-layer mean

winds (from the 850 hPa to the 300 hPa layer) signify the steering flow around the TCs for interpreting the rules. The differences between the rules and the existing theories and findings on TC recurvature are that the rules discovered in this investigation are quantitative. However, the existing findings on TC recurvature, such as "a strong subtropical high and weak monsoon cause TCs to recurve" are mainly qualitative results, or a "rule of thumb". The rules governing TC recurvature from a quantitative perspective can provide useful references for TC track forecasters.

Dynamic models have undergone rapid advancements in the last few decades and are now broadly used in TC track and intensity prediction. Given the rules unraveled by the C4.5 algorithm and the development of dynamic models, the data mining methods and the dynamic model will mutually enrich each other in a complementary and integrated manner. The research findings can be used to improve the dynamic models by tuning the parameters and verifying the modeling results.

CHAPTER 5: TC Landfall

5.1 Introduction

As shown in Chapter 4, TC recurvature and landfall are two TC-related scientific problems of significant economic and social concern. Chapter 4 presents the analysis of TC recurvature based on the C4.5 algorithm. In this chapter, particular attention will be paid to TC landfall.

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

Numerous researches have investigated the seasonal or annual forecasts of TC landfall frequency in the Atlantic basin, WNP. Particular attention has also focused on landfall over regional areas; for example, Southern China and the Korean Peninsula. The variability of seasonal TC landfall frequency was attributed to the modulation of ENSO, MJO, QBO, AMO and PDO. The variability of large-scale circulation (e.g., the subtropical high, monsoon troughs and westerly anomalies) associated with ENSO events or other oscillations are likely to be responsible for most of the seasonal or annual variations in TC landfall in WNP and the Atlantic basin.

TC-land interaction during landfall has been studied through observational analysis and dynamic modeling during the last few decades (refer to Chan and Liang, 2003; Tuleya *et al.*, 1984; Wong and Chan, 2006a, b, 2007). The impact of topography on landfalling TCs is more significant than roughness variation, especially in mountainous areas (Chang, 1982). Observations (Brunt, 1968; Hamuro and Coauthors, 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 (1973, 1974) found that topographical effects, if not considered appropriately, can bring forth significant errors in the forecasting of TC motion. Chang (1982) also observed that the interaction between the terrain and the TC caused strong easterlies to develop north of the island, accelerating the TC in his case.

To recapitulate, previous researches on TC landfall prediction were concerned with seasonal or annual TC landfall frequency according to the variability of large-scale circulation (e.g., the subtropical high, monsoonal systems and mid-latitude westerlies) 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 researches have investigated the mechanisms controlling TC landfalls over the Chinese coast based on the historical TC database. Besides, few researches have investigated TC forecasting solely with regard to 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 (e.g., Brettschneider, 2008; Elsner and Liu, 2003; Goh and Chan, 2010; Liu and Chan, 2003), large-scale circulation exerts a dramatic influence on TC landfall in the SCS and WNP basins. As stated in Chapter 4, TCs forming in the SCS and WNP move in a northward or northwestward direction under an easterly steering current when the subtropical high is relatively strong and displaces 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 due to the interactions of large-scale circulation. In general, TCs will move far away from the Chinese coast without making landfall there due to changes in the steering flow from westward to eastward.

Given the rapid rate of growth of the TC-related data sets 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 the TC-related dataset, this chapter 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 chapter is organized as follows. Section 5.2 introduces the study area and data source. The methodology is introduced in section 5.3. Section 5.4 explains and analyzes the results generated by the C4.5 algorithm. The concluding remarks are discussed in section 5.5.

5.2 Data source and Methodology

All of the typhoons that occurred in the study area from 2000-2009 are used as samples. The data sources related to TC landfall consist of two major classes of data: TC best track data and meteorological data.

The TC best track data are made available from JMA RSMC Tokyo. This post analysis best track data consists of 6-hourly estimates of position (latitude and longitude), MCP, and the 10-min MSW of all named TCs in the WNP basin, including the SCS, from 1951 to the present. There is a total of 6920 sampling TC points derived from 222 TC sample tracks, 53 of which are landfall TC tracks and the rest are non-landfall ones. Landfalling TC tracks contain 1150 observations and non-landfall TC tracks contain 5770 observations, respectively. It is noteworthy that, with respect to the observations of landfalling TCs, only the TC observations prior to landfall are labeled class "1". Non-landfall TC tracks, on the contrary, are labeled class "0". Due to the imbalance in the number of TC samples in the two classes, re-sampling methods (here oversampling) are used to make the class distribution more balanced.

The meteorological variables (e.g., wind fields and geopotential height in different atmospheric layers) are derived from the NCEP GFS FNL at 6-hourly intervals from 07/30/1999 to the present (Yang et al. 2006). These NCEP FNL Operational Global Analysis data are organized on continuous 1.0° x 1.0° degree grids at 6-hourly intervals. This GFS is run four times a day in near-real time at NCEP. The data are made available on the surface, at 26 mandatory (and other pressure) levels, varying from 1000 hPa to 10 hPa (e.g., 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800 850, 900, 950 and 1000 hPa), in the surface boundary layer and in some sigma levels, the tropopause and several others. The parameters include surface pressure, SLP, geopotential height, relative vorticity, temperature, SST, soil values, ice cover, relative humidity, zonal (u-) and meridional (v-) wind fields, vertical motion, and ozone parameters. Although the NCEP/NCAR reanalysis data (Kalnay et al., 1996) are available for a longer time period, the resolution of 2.5° x 2.5° degrees of NCEP/NCAR is coarser than that of the NCEP FNL dataset. Therefore, we derive the meteorological variables from the NCEP FNL dataset other than the NCEP/NCAR dataset. For the higher resolution and time-period availability of the NCEP FNL dataset, the TC samples are only extracted from 2000-2009.

As explained in section 4.1, both TC recurvature and landfall are influenced by large-scale circulation and the wind fields surrounding the TC. The potential factors that determine whether or not the TCs make landfall across the Chinese coast are categorized into three groups: variables relating to large-scale circulation, variables measuring the circulations surrounding TCs, and variables characterizing TCs (see Table 5-1). Table 5-1 presents the variables displayed in an abbreviated form in order to be consistent with the variables in the figure showing the classification tree (see Figure 5-1). In the group "Circulation surrounding TC", "uwnd_200" and "vwnd_200" are respectively the average zonal and meridional wind of the 6-8 degree radial belt at the 200-hPa level. The other variables in the same group are defined likewise. Five variables are chosen to measure the strength and position of large-scale circulation (i.e., the subtropical high, EASM and westerlies) that largely control TC recurvature (e.g., Harr and Elsberry, 1991, 1995a): area index (area_IndexSTH), intensity index (inten_IndexSTH), westward extension index of subtropical highs (west_extSTH) in WNP (for measuring the strength

of the subtropical high and the position of the subtropical high ridge), the EASM index by Wang and Fan (1999) (Monsoon_WF), as well as the westerly index (W_Westerly). These factors are analyzed by the C4.5 algorithm to build the classification tree.

Table 5-1. The potential attributes influencing TC landfall

Groups	Potential Variables
Circulation surrounding TC	uwnd_200, uwnd_300, uwnd_400, uwnd_500, uwnd_600, uwnd_700, uwnd_800, uwnd_850, uwnd_1000, vwnd_200, vwnd_300, vwnd_400, vwnd_500, vwnd_600, vwnd_700, vwnd_800, vwnd_850, vwnd_1000
Large-scale circulation	area_IndexSTH, inten_IndexSTH west_extSTH, Monsoon_WF W_Westerly
Characteristics of TC	Lon, Lat, Pressure (Central Pressure of TC Center)

a. Indices for Large-scale Circulation

A variety of indices have been proposed to measure the position and intensity of large-scale circulation. The indices depicting large-scale circulation form part of the potential parameters that influence TC landfall.

EASM is the most influential and essential component of the Asian climatic systems (Chen and Yoon, 2000; Ding, 2007; Ding and Chan, 2005; Wang et al., 2008) in the study area due largely to orographical forcing. The reversed Wang and Fan index (Wang and Fan, 1999) is generally identical to the leading principal component of the EASM. The Wang and Fan index belongs to the "shear vorticity" index and was initially proposed to measure the variability of the WNP summer monsoon. This index was defined by the U_{850} in (5°–15°N, 90°–130°E) minus U_{850} in (22.5°–32.5°N, 110°–140°E), where U_{850} denotes the zonal wind in the 850 hPa layer. In physical sense, the WF shear vorticity index represents the variations of the subtropical high and WNP monsoon trough, which are the crucial parts of the EASM circulations (Tao and Chen, 1987). The

WF shear vorticity index is thus chosen to measure the status of the EASM, which place significant influences on TC movement, and further on TC landfall (Chen, 2009; Elsberry, 2004; Harr and Chan, 2004; Ko and Hsu, 2006).

The WNPSH indices announced by NCC have been used broadly in previous studies (e.g., Chen et al., 2001a). The NCC defined five indices to describe the location and intensity of the WNPSH, including the WNPSH intensity index, the westward extension index, the mean ridge index, and the area index according to the monthly mean 500 hPa geopotential heights in the weather charts published by the Chinese Meteorological Administration (Chen, 1999). By considering the pre-existing subtropical high indices defined by NCC and other scholars (e.g., Lu and Dong, 2001; Lu *et al.*, 2007), we used these definitions to measure the status of the subtropical high as follows. The WNPSH intensity index is defined as the average geopotential height of points with their geopotential height greater than 5870 gpm in a rectangular region (100 °E to 180 °E and 10 °N to 60 °N). The WNPSH area index is defined as the number of grids with their geopotential height greater than 5870 gpm in the region (100 °E to 180 °E and 10 °N to 60 °N). The WNPSH "westward extension index" is defined as the longitude of the most westerly position of the subtropical high ridge. This definition of the "westward extension index" is in line with that of NCC. The three indices are extracted and calculated from FNL reanalysis data to measure the WNPSH, for simplicity.

The mid-latitude westerlies also exert a significant impact on TC landfall through upper-tropospheric zonal winds (~200 hPa) (e.g., George and Gray, 1977; Guard, 1977); for instance, it is observed that, if the base is weakened sufficiently west of a TC in conjunction with an eastward-shift mid-latitude trough, and remains low, northward recurvature will occur (Riehl and Shafer, 1944). Rossby (1939) defined a typical westerly index by the 500 hPa geopotential heights: $I_R = H_{35} - H_{55}$. Of note is that I_R is the westerly index defined by Rossby (1939), and H_{35} and H_{55} represent the geopotential heights at 35 °N and 55 °N respectively on the 500 hPa layer. Here, we calculate the I_R from 100 °E to 180 °E in longitude to measure the mid-latitude westerlies. The larger the I_R, the stronger the westerly.

b. Wind Fields Surrounding TC

Various indices depicting large-scale circulation are presented in the previous subsection. Apart from large-scale circulation, the outer radius (e.g., 5-7°, 6-8°) environmental wind fields from the TC center also play significant roles in determining whether TCs will make landfall across the Chinese coast. During the last few decades, numerous radial circles have been proposed to investigate wind fields (Chan et al., 1980; Fitzpatrick, 1992; Hodanish and Gray, 1993). In order to choose the most appropriate radius for calculating the wind fields from the TC's center, radial circles such as 4°-6°, 5°-7°, 6°-8°, 7°-9° and the 1-3 oscants in these radial circles are tested by the accuracy of classification. The 6-8° radius from the TC center is chosen to calculate the average wind fields surrounding the TC, in order to conform with the procedure followed in Chapter 4.

TC tracks are associated closely with the position where TCs are located as a result of the atmospheric characteristics of different places. The longitude and latitude of TCs' centers are therefore regarded as potential parameters influencing TC landfall. Although intensity has not been proven to influence TC landfall in terms of physical mechanism, several studies (Evans and McKinley, 1998; Knaff, 2009; Riehl, 1972) have found a relationship between recurvature and intensity. Given the close relationship between TC recurvature and TC landfall, TC intensity as well as the longitude and latitude of TC centers characterizing TCs are chosen as the potential parameters for this study.

5.3 Results analysis and interpretation

5.3.1 Results

The C4.5 algorithm is implemented in Weka 3.6.2, that is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from user-defined Java codes. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is

also well-suited for developing new machine learning schemes (the software is open source and available at http://www.cs.waikato.ac.nz/ml/weka/index.html). The minimum leaf size is set at '150' by considering the complicity of the classification tree. Ten-fold cross validation is used for verification. Other parameters are used according to their default setting. The settings are: "Binary Split" is true; confidence factor is "0.25"; debug is "false"; numFolds are "3"; reducedErrorPruning is "true"; saveInstanceData is "false"; seed is "1"; subtreeRaising is "true"; and unpruned is "false"; uselaplace is "false".

It should be noted that the TC samples used in this study contain the TC observations prior to landfall and observations of all non-landfall TCs. The observations post-landfall may carry useful information and knowledge about landfalling TCs. All of the observations of non-landfalling and landfalling TCs are classified using the C4.5 algorithm to see whether there is a significant improvement in accuracy. In addition, with respect to landfalling TCs, we can also extract the equal points from non-landfalling TCs from the average number of observations within landfalling TCs. Therefore, three pairs of TC datasets are classified used samples from 2000-2009, verified by TCs in 2010. The results are shown in Table 5-2. It can be observed that the verifying classification accuracy of the first group (Partially Landfalling and Full Non-landfalling TCs) is quite close to the training accuracy. The testing accuracy of the other two groups is, however, considerably lower than the training accuracy. The other two TC groups fail to improve the classification accuracy. Therefore, the results derived from the partially non-landfalling and fully landfalling TCs are discussed in the following sections.

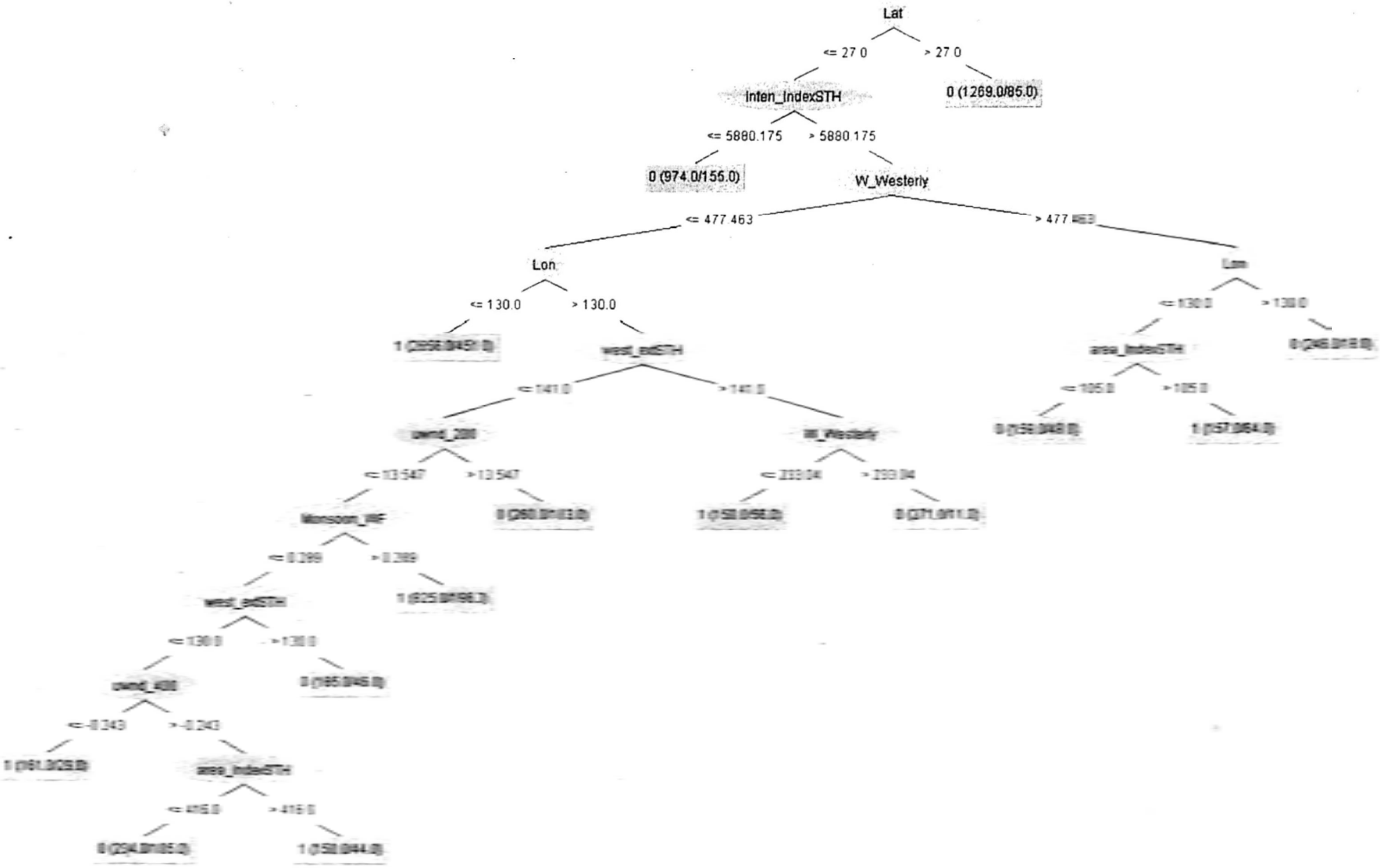


Figure 5-1 The classification tree governing the TC landfall built by C4.5 algorithm

Table 5-2 The classification accuracies of different TC groups

TC Groups	*Training Accuracy*	*Verifying Accuracy*
Partial Landfalling and Full Non-landfall TCs	81.3287%	81.9018%
Full Landfalling and Full Non-landfall TCs	83.4101%	67.5000%
Partial Landfalling and Partial Non-landfall TCs	79.4174%	69.6000%

The total classification accuracy of the built decision tree (see Figure 5-1) is 81.3287% by ten-fold cross-validation. Among all the variables listed in Table 5-1, 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 (Wang and Fan, 1999) (Monsoon_WF). In Figure 5-1, "1" means "landfall" and "0" means "non-landfall", respectively. The rectangles are leaf nodes whereas ellipses or circles are the parent nodes. Every path from the root node to the leaf node represents a rule, which can provide references for TC landfall prediction. Fourteen rules governing TC landfall in the SCS and WNP basins are derived from the classification tree built by the C4.5 algorithm (see Table 5-3). Taking leaf node "0(1269.0/85.0)" as an example, the "0" before the bracket means "the TC will not make landfall along the Chinese coast" and "1269.0" and "85.0" indicate that, among the total 1269 samples of the leaf node, there are 1184 (i.e.,1269 minus 85) non-landfall samples and 85 landfall samples, respectively. Each rule can be justified by meteorological and TC theories.

Table 5-3 lists the 14 rules that are derived from the decision tree (Figure 5-1). Other than rule 1 that is merely determined by "latitude" of TC center, the remaining 13 rules are categorized into three groups:

Group 1 consists of rule 2, rule 4, rule 8, rule 9, rule 12 and rule 14. The rules in group 1 are decided by the condition of the subtropical high in the last split.

Group 2 has rule 6, rule 7, rule 10, rule 11, and rule 13. This group is determined by the factors measuring the westerlies (e.g., westerly index and 200 hPa or 400 hPa zonal winds) and monsoonal systems in the last split.

Group 3 is comprised of rule 3 and rule 5. This group is controlled by the "longitude" of a TC center.

The indices characterizing large-scale circulation (e.g., the subtropical high and mid-latitude westerlies) are chosen to build the tree after the "latitude" of a TC center is selected in the first split. In other words, the large-scale circulation severely affects TC landfall activity based on the decision tree (Figure 5-1).

Table 5-3 The 14 unraveled rules governing TC landfall

Rule Number	Rules	Attributes	Accuracy
1	If Lat>27, then the TC will not make landfall across the Chinese coast.	Latitude	1184/1269=0.933
2	If Lat≤27, and inten_IndexSTH≤5880.175, then the TC will not make landfall across the Chinese coast.	Latitude, intensity index of the subtropical high	819/974=0.841
3	If Lat≤27, and inten_IndexSTH>5880.175, and W_Westerly>477.463, and Lon>130, then the TC will not make landfall across the Chinese coast.	Latitude, intensity index of the subtropical high, Westerly Index, Longitude	228/246=0.927
4	If Lat≤27, and inten_IndexSTH>5880.175, and W_Westerly>477.463, and Lon≤130, and area_IndexSTH≤105.0, then the TC will not make landfall across the Chinese coast.	Latitude, intensity index of the subtropical high, Westerly Index, Longitude, area index of the subtropical high	111/159=0.698
5	If Lat≤27, and inten_IndexSTH>5880.175, and W_Westerly≤477.463, and Lon≤130, then the TC will make landfall across the Chinese coast.	Latitude, intensity index of the subtropical high, Westerly Index, Longitude	2205/2656=0.830
6	If Lat≤27, and inten_IndexSTH>5880.175, and W_Westerly≤477.463, and Lon>130, and west_extSTH>141, and W_Westerly>233.04, then the TC will	Latitude, intensity index of the subtropical high, Westerly Index, Longitude, zonal wind in the 500 hPa layer	260/271=0.959

	not make landfall across the Chinese Coast.		
7	If Lat≤27, and inten_IndexSTH>5880.175, and W_Westerly≤477.463, and Lon>130, and west_extSTH≤141, and uwnd_200≤13.547, and Monsoon_WF>0.289, then the TC will make landfall across the Chinese coast.	Latitude, intensity index of the subtropical high, Westerly Index, Longitude, westward extension index of the subtropical high, zonal wind in the 200 hPa layer, Monsoon Index	629/825=0.762
8	If Lat≤27, and inten_IndexSTH>5880.175, and W_Westerly>477.463, and Lon≤130, and area_IndexSTH>105.0, then the TC will make landfall across the Chinese coast.	Latitude, intensity index of the subtropical high, Westerly Index, Longitude, area index of the subtropical high	93/157=0.592
9	If Lat≤27, and inten_IndexSTH>5880.175, and W_Westerly≤477.463, and Lon>130, and west_extSTH≤141, and uwnd_200≤13.547, and Monsoon_WF≤0.289, and west_extSTH>130, then the TC will NOT make landfall across the Chinese coast.	Latitude, intensity index of the subtropical high, Westerly Index, Longitude, westward extension index of the subtropical high, zonal wind in 200 hPa layer, Monsoon Index	119/165=0.721
10	If Lat≤27, and inten_IndexSTH>5880.175, and W_Westerly≤477.463, and Lon>130, and west_extSTH≤141, and uwnd_200>13.547, then the TC will not make landfall across the Chinese coast.	Latitude, intensity index of the subtropical high, Westerly Index, Longitude, westward extension index of the subtropical high, zonal wind in the 200 hPa layer	157/260=0.604
11	If Lat≤27, and inten_IndexSTH>5880.175, and W_Westerly≤477.463, and Lon>130, and west_extSTH≤141, and	Latitude, intensity index of the subtropical high, Westerly Index, Longitude, westward	132/161=0.820

	uwnd_200≤13.547, and Monsoon_WF≤0.289, and west_extSTH≤130, and uwnd_400≤-0.243, then the TC will make landfall across the Chinese coast.	extension index of the subtropical high, zonal wind in the 200 hPa and 400 hPa layers, Monsoon Index	
12	If Lat<27, and inten_IndexSTH>5880.175, and W_Westerly≤477.463, and Lon>130, and west_extSTH≤141, and uwnd_200≤13.547, and Monsoon_WF<0.289, and west_extSTH≤130, and uwnd_400>-0.243, and area_IndexSTH≤416, then the TC will make landfall across the Chinese coast.	Latitude, intensity index of the subtropical high, Westerly Index, Longitude, zonal wind in 500 hPa layer, westward extension index of the subtropical high, area index of the subtropical high	129/234=0.551
13	If Lat≤27, and inten_IndexSTH>5880.175, and W_Westerly≤477.463, and Lon>130, and west_extSTH>141, and W_Westerly<233.04, then the TC will make landfall across the Chinese Coast.	Latitude, intensity index of the subtropical high, Westerly Index, Longitude, westward extension index of the subtropical high	94/150=0.627
14	If Lat≤27, and inten_IndexSTH>5880.175, and W_Westerly≤477.463, and Lon>130, and west_extSTH≤141, and uwnd_200≤13.547, and Monsoon_WF≤0.289, and west_extSTH≤130, and uwnd_400>-0.243, and area_IndexSTH>416, then the TC will make landfall across the Chinese coast.	Latitude, intensity index of the subtropical high, Westerly Index, Longitude, zonal wind in 500 hPa layer, westward extension index of the subtropical high, area index of the subtropical high	106/150=0.710

In Table 5-3, the first column represents the rule numbers. The second column describes the unraveled rules. The third column depicts the attributes involved in each rule. Each attribute may occur two or three times per rule. However, the attribute occurs only once

in the second column to avoid redundancy and misunderstanding. Besides, the fourth column shows the classification accuracy of each rule. This accuracy is calculated by dividing the number of samples classified correctly by the total number of samples in the leaf node. The rules derived from the classification tree include rules that confirm existing TC theories, and rules that are new in theory. The new rules are: Rule 1, Rule 2, Rule 3, Rule 5, and Rule 11. The rules confirming existing TC theories are: Rule 4, Rule 6, Rule 7, Rule 8, Rule 9, Rule 10, Rule 12, Rule 13 and Rule 14 (Table 5-3).

5.3.2 Verification and Interpretation

The data mining results unraveled by the algorithms should be verified and interpreted according to the background knowledge in terms of TC movement and atmospheric circulation. Verification and interpretation are essential components of the data mining process (Han and Kamber, 2006). The decision tree shown in Figure 5-1 is verified by the RSMC JMA TC best track in 2010 to justify whether it can correctly classify the landfall of newly-occurred TCs. We interpret the classification tree and the rules derived from the tree according to TC theory on movement and landfall.

The decision tree built by the C4.5 algorithm is justified using the 2010 RSMC JMA TC best track data with 14 TC tracks, which are not involved in training the classification model. Of the fourteen TCs occurring in 2010, six made landfall over the Chinese coast while the rest did not. The fourteen TC tracks with 326 TC observations are used to test the decision tree. The TC name, landfall and time of genesis of the 14 TCs in 2010 are elaborated in Table 5-4. A confusion matrix illustrating the verification results is shown in Table 5-5. Of the 123 landfalling and 213 non-landfalling TC observations, 109 and 158 are correctly classified, respectively. The classification accuracy is 81.9018%, compared with 81.3287% for the training model. Therefore, the rules derived from the classification tree characterize TC landfall over the Chinese coast and the classification tree can provide useful references for the prediction of TC landfall there.

Table 5-4 TC tracks in 2010 for verifying the landfall classification Tree

2010 TC Names	*Landfall (Yes or No)*	*Time of Genesis*
OMAIS	No	2010032212
CONSON	Yes	2010071112
CHANTHU	Yes	2010071706
DIANMU	No	2010080700
MINDULLE	No	2010082200
LIONROCK	Yes	2010082718
KOMPASU	No	2010082812
NAMTHEUN	No	2010082906
MALOU	No	2010090112
MERANTI	Yes	2010090700
FANAPI	Yes	2010091418
MALAKAS	No	2010092006
MEGI	Yes	2010101300
CHABA	No	2010102318

Table 5-5 Confusion matrix of the verification

	landfall	*non-landfall*
landfall	109	14
non-landfall	45	158

The classification tree shown in Figure 5-1 is interpreted from the meteorological aspect based on the existing theories and research findings on TC movement. The interpretation of the classification tree governing TC landfall resembles that of the classification tree governing TC recurvature, shown in Chapter 4. The primary idea is to consider the composite steering flow (e.g., deep-layer mean winds), 500 hPa geopotential height (depicting the subtropical high) of the samples in each leaf node and the distance from the location of the TC center to the Chinese coast.

The rule derived from node "0(1269.0/85.0)" is quite simple, but the classification accuracy is the highest of all the rules. This rule is stated as: "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 the latitude in excess of 27°N is in the mid-latitude area with a strong westerly flow. When a TC moves into this region, it will be steered by the westerly flow to move away from the Chinese coast. The composite deep-layer mean winds (i.e., averaging from 850 hPa to 300 hPa layer), as well as the 500 hPa geopotential height of all the samples in this leaf node are overlapping, according to Figure 5-2. The subtropical high is depicted by the contour of Z_{500} = 5870 m (Z_{500} represents 500 hPa geopotential height). Figure 5-2 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 westerly flow to move far away from the Chinese coast. Latitude 27°N is a key value 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. The first variable means the greatest improvement in classification accuracy. This latitude can be used as a crucial value in the operational forecasting of TC landfall across the Chinese coast.

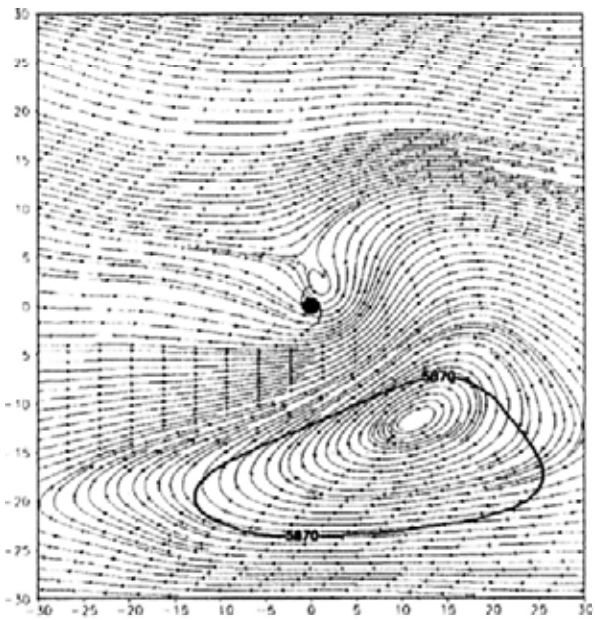


Figure 5-2 The composite deep-layer mean streamline and geopotential heights in 500 hPa layer of the samples in leaf node "0(1269.0/85.0)". The trick contour indicates the high center of the subtropical high. The TC symbol denotes the relative TC center with

the coordinate (0, 0). The axes on the bottom and left of the plot represent the x and y ordinates that represent east and north respectively.

The rule constructed from the leaf node "0(234.0/105.0)", which has the lowest classification accuracy among the rules, is addressed as: "If a TC moves to the south of 27 °N, and the intensity index of the subtropical high is higher than 5880.175, and the westerly index is smaller than 477.463, and the longitude of the TC center is not to the west of 130 °E, and the western ridge of the subtropical high is to the west of 141 °E, and the 200 hPa zonal wind is smaller than 13.547 m/s, and the monsoon index is smaller than 0.289 (strong monsoon), and the western ridge of the subtropical high is to the west of 130 °E, and the 400 hPa zonal wind is larger than -0.243 m/s, and the area_Index of the subtropical high is smaller than 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 degrees. 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 Figure 5-3). The tendency for TCs to make landfall is not remarkable. This is in accordance with the classification accuracy of 0.551.

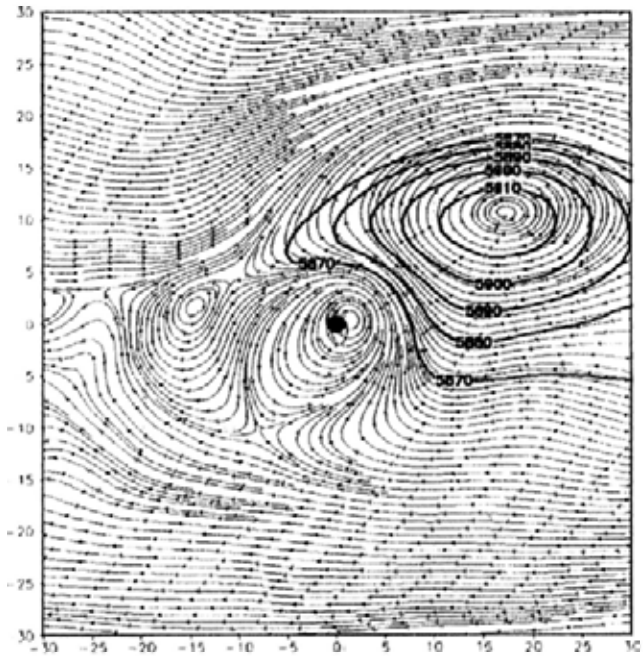


Figure 5-3 As in Figure 5-2 except for the leaf node "0(234.0/105.0)".

The rule derived from the leaf node "0(974.0/155.0)" is stated as: " If a TC is situated to the south of 27 °N and the intensity index of the subtropical high is smaller than 5880.175 gpm, then the TC will not make landfall across the Chinese coast." This rule has an accuracy of 0.942. Based on this rule, the intensity of the subtropical high is relatively weak (inten_indexSTH $\leq$ 5880.175 gpm). A 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 Figure 5-4. Besides, 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 Figure 5-4). Therefore, TCs under this rule will not make landfall across the Chinese coast.

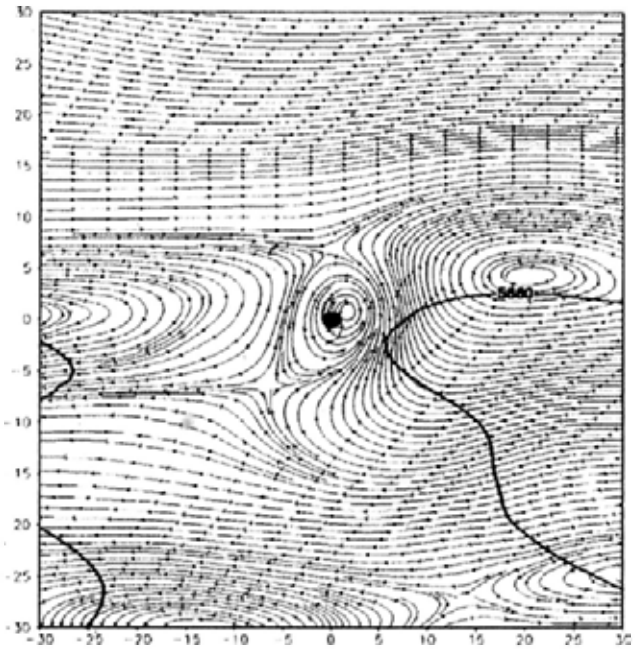


Figure 5-4 As in Figure 5-2 except for the leaf node "0 (974.0/155.0)".

The rule derived from the leaf node "0(246.0/18.0)" is described as: "If a TC is located to the south of 27 °N, and the intensity index of the subtropical high is higher than 5880.175 gpm, and the westerly index is larger than 477.463 (strong westerly) and the longitude of the TC center displaces to the east of 130 °E, then the TC will not make landfall across 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 the grids with their geopotential height larger than 5870 gpm in the region (100 °E to 180 °E and 10 °N to 60 °N). This rule reveals that the interactions

between the relatively strong subtropical high and strong westerlies eventually cause TCs to recurve without making landfall across the Chinese coast when they are located relatively far away from it (>130 °E). It is shown in Figure 5-5 that the subtropical high lies to the east of the TC center. Although the steering flow may drive the TC westward for a 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.

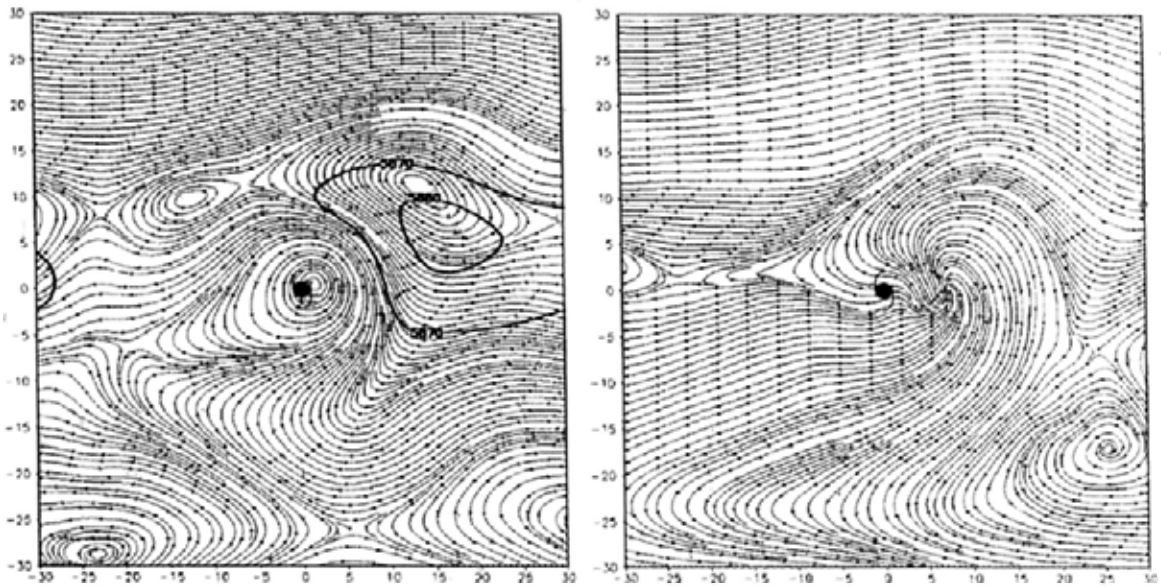


Figure 5-5 The composite deep-layer mean streamline and 500 hPa geopotential heights (left), and the composite 200 hPa wind fields (right) of the samples in the leaf node "0(246.0/18.0)". The TC symbol denotes the relative TC center with the coordinate (0, 0). The axes on the bottom and left of the plot represent the x and y ordinates that represent east and north respectively.

The rule obtained from the leaf node "1(157.0/64.0)" is described as: "If a TC propagates to the south of 27 °N, and the intensity index of the subtropical high is higher than 5880.175 gpm, and the westerly index is larger than 477.463 (strong westerly), and the longitude of the TC center is to the west of 130 °E, and the area index of the subtropical high is larger than 105, then the TC will not make landfall across the Chinese coast." The rule means that, when the TC moves closer to the Chinese coast, the strong, wide subtropical high depicted by the intensity index and area index of the subtropical high drives the TC to make landfall across the Chinese coast. Figure 5-6 shows that 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 across the Chinese coast. The composites in Figure 5-6 are consistent with what the rule infers.

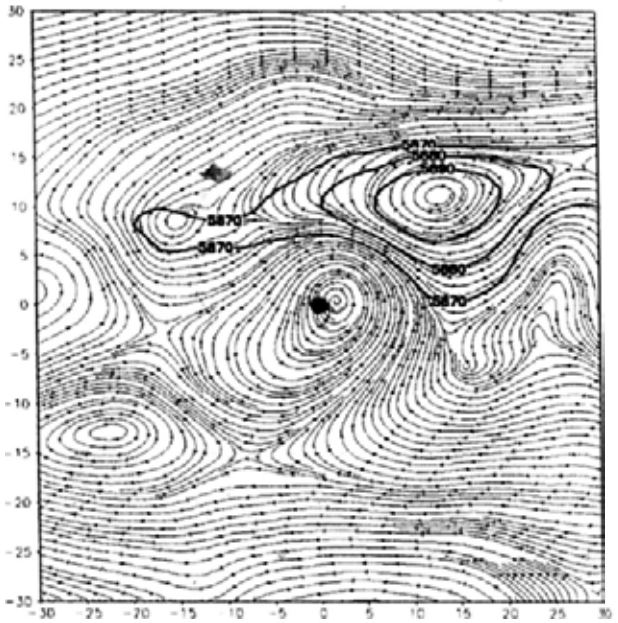


Figure 5-6 As in Figure 5-2 except for the leaf node "1(157.0/64.0)".

The rule obtained from the leaf node "1(2656.0/451.0)" is stated as: "If a TC moves to the south of 27 °N, and the intensity index of the subtropical high is higher than 5880.175 gpm, and the westerly index is smaller than 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 flow is weak near the TC. It is noteworthy that the longitude in this rule is to the west of 130 °E. The environmental wind fields surrounding the TC are plotted in Figure 5-7. The deep-layer mean wind is easterly, which is consistent with the 200 hPa composite wind fields (Figure 5-7). This indicates that the environmental easterly flow 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 across the Chinese coast.

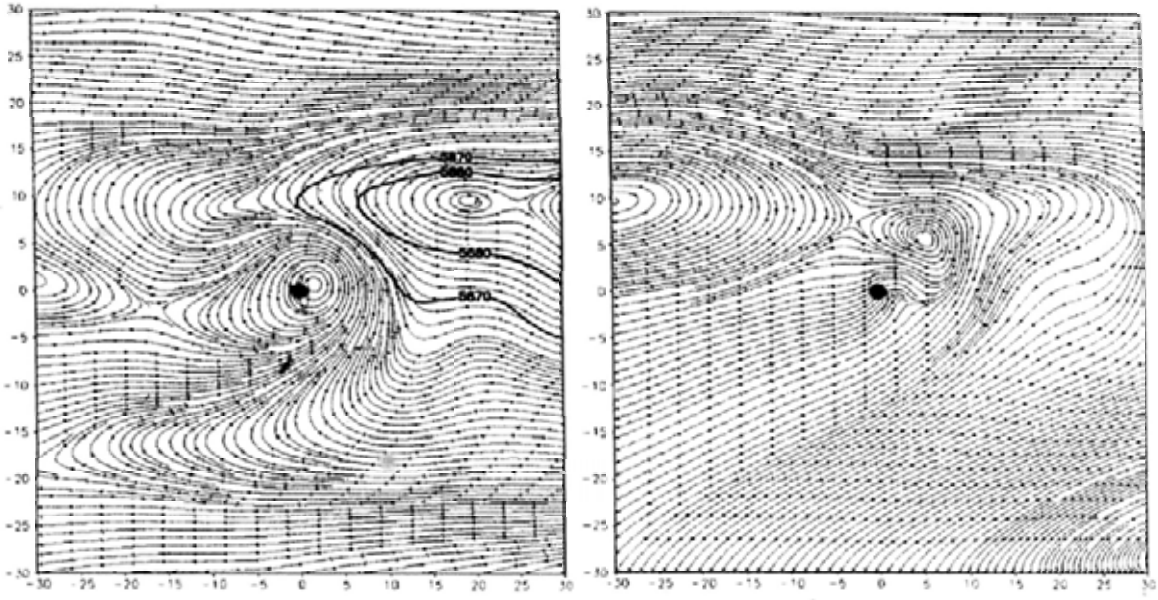


Figure 5-7 As in Figure 5-2 except for the leaf node "1(2656.0/451.0)".

The rule formed by the path from the root node to the leaf node "1(825.0/196.0)" is stated as: "If a TC is located to the south of 27 °N, and the intensity index of the subtropical high is higher than 5880.175 gpm, and the westerly index is smaller than 477.463, and the longitude of the TC center is not to the west of 130 °E, and the western ridge of the subtropical high is to the west of 141 °E, and the 200 hPa zonal wind is smaller than 13.547 m/s, and the monsoon index is larger than 0.289 (weak monsoon), then the TC will make landfall across the Chinese coast." This rule shows that the mean intensity of the subtropical high is strong, the westerly flow is weak near the TC, and the southern and southwestern components of the monsoon system are weak. Figure 5-8 shows that the easterly flow along the southern part of the subtropical high will eventually lead the TC to make landfall across the Chinese coast. This rule integrates the subtropical high, monsoonal system, 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 (Chen *et al.*, 2009; Harr and Elsberry, 1991, 1995a).

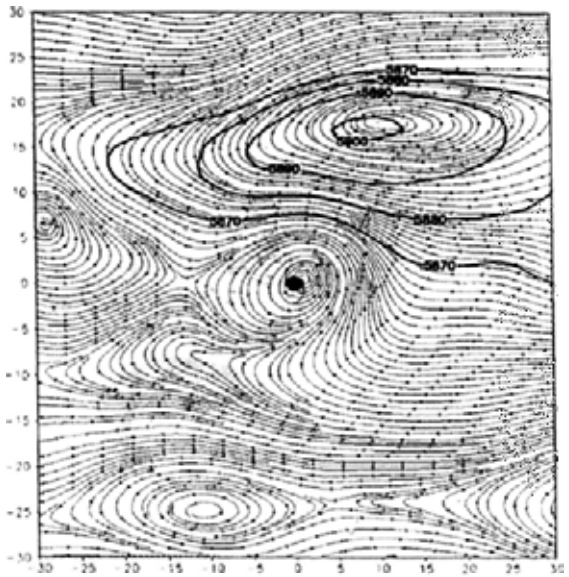


Figure 5-8 As in Figure 5-2 except for the leaf node "1(825.0/196.0)".

The rule derived from the leaf node "0(271.0/11.0)" is described as: "If a TC moves to the south of 27 °N, and the intensity index of the subtropical high is higher than 5880.175 gpm, and the westerly index is smaller than 477.463, and the longitude of the TC center is to the east of 130 °E, and the western ridge of the subtropical high is to the east of 141 °E, and the westerly index is smaller than 233.04, then the TC will not make landfall across the Chinese coast." This rule indicates that the subtropical high ridge retreats to the east of 141 °E, and the westerly index varies from 233.04 to 477.463. Besides, the intensity index (>5880.175 gpm) represents the strong subtropical high. As shown in Figure 5-9, the subtropical high retreats to the east of the TC center in spite of the relatively high intensity index and the westerlies begins to exercise influence on the TC center. The southerly component of the eastern part of the subtropical high steers the TC northward and northeastward, and it subsequently recurves without making landfall across the Chinese coast. Therefore, the composite deep-layer mean wind and subtropical high are fully in line with what is indicated by the rule.

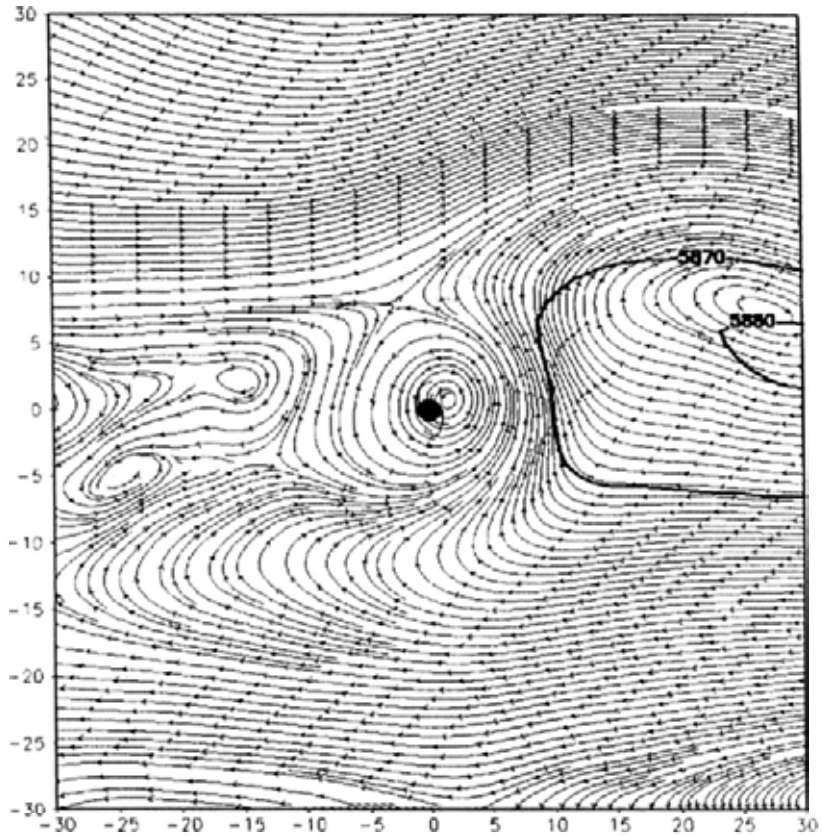


Figure 5-9 As in Figure 5-2 except for the leaf node "0(271.0/11.0)" .

The rule obtained from the leaf node "0(150.0/56.0)" is described as: "If the TC moves to the south of 27 °N, and the intensity index of the subtropical high is higher than 5880.175 gpm, and the westerly index is smaller than 477.463, and the longitude of the TC center is to the east of 130 °E, and the western ridge of the subtropical high is to the east of 141 °E, and the westerly index is smaller than 233.04, then the TC will not make landfall across the Chinese coast." This rule indicates that the subtropical high ridge shifts to the east of 141 °E, and that the westerly index is as low as 233.04. As described in Figure 5-10, the subtropical high integrating with another high pressure system is located to the north of the TC. The prevailing easterly surrounding the TC steers the typhoon vortex westward, and eventually it will make landfall across the Chinese coast. It is seen that this rule differs from the rule induced by leaf node "0(150.0/56.0)" on the westerly index. The westerly index is smaller than that of leaf node "0(150.0/56.0)". It can be proved by the prevalent eastern component induced by the subtropical high (see Figure 5-10).

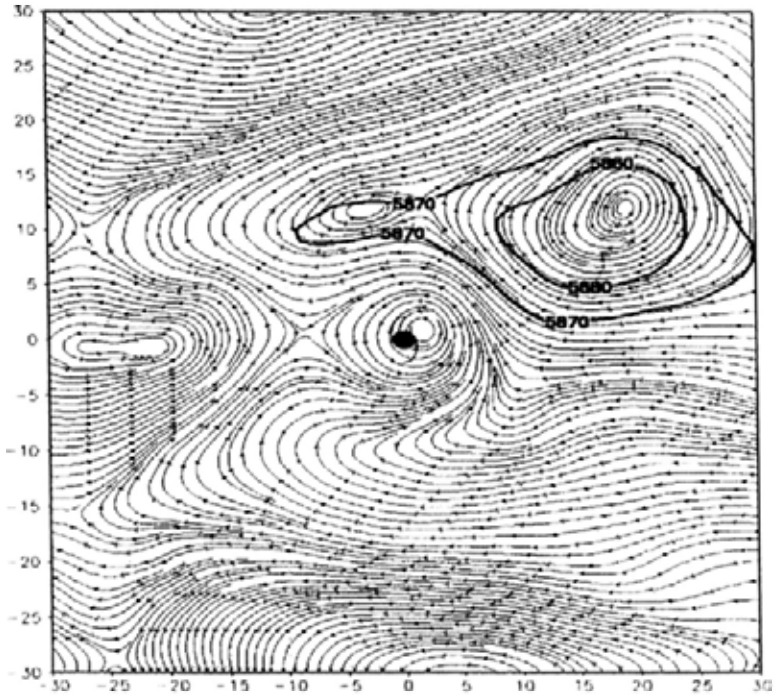


Figure 5-10 As in Figure 5-2 except for the leaf node "0(150.0/56.0)"

In summary, the longitude at 130 °E was frequently chosen as a splitting value by the C4.5 algorithm when building a classification tree for TC landfall and recurvature in both this study and Chapter 4 of the two-part series. It is indicated that this longitude (130 °E) is critical for TC recurvature and landfall. In terms of operational aid, particular attention should be paid to this longitude, since it plays an important role in forecasting whether or not the TC will recurve and make landfall across the Chinese coast based on the series. The latitude at 27 °N is the first chosen attribute of the classification tree (see Figure 5-1). 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 will probably not make landfall across the Chinese coast. The 200 hPa and 400 hPa zonal winds are also selected by the C4.5 algorithm. Nonetheless, none of the meridional winds is selected by the C4.5 algorithm in this study. As proven by various investigations, the zonal winds surrounding TC centers have more significant impacts on TC movement than meridional winds (e.g., Elsberry, 1987; Evans *et al.*, 1991; George and Gray, 1977; Harr and Elsberry, 1991, 1995a; Holland and Wang, 1995). In addition to the attributes discussed above, selected by the C4.5 algorithm to build the classification tree are the indices for measuring large-scale circulation (e.g., the subtropical high, monsoons and westerlies), which include the westerly index, indices of

the subtropical high (the intensity index, area index, and west extension index), and the monsoon index. The indices are used here to quantify the status of the large-scale circulation. The results are in good agreement with previous researches and findings about the modulation of large-scale circulation on TC movement.

5.4 Summary

This research investigated TC landfall in the SCS and WNP basins through data mining methods. Based on the historical TC archives, the C4.5 algorithm, a classic tree algorithm for classification, is employed quantitatively to discover the rules and regularities governing TC landfall.

A classification tree, with 14 leaf nodes, is built by the C4.5 algorithm. The path from the root node to each leaf node forms a rule. Eventually, 14 rules governing TC landfall across the Chinese coast are unraveled from the potential attributes which may influence TC landfall. The rules are derived by the attributes and splitting values. 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), are useful references for operational aids for TC forecasting. The rules are verified from the perspective of meteorology and TC theories on movement and recurvature (e.g., deep-layer mean winds and large-scale circulation). The findings of this study are also consistent with the existing results on TC movement and landfall (e.g., Chen *et al.*, 2009; Harr and Elsberry, 1991, 1995a). Both the unraveled rules and the splitting values of these rules can provide useful references for operational aids for TC landfall prediction. It is noteworthy that TC intensity (central pressure) is chosen to build the classification tree governing TC recurvature. This implies that TC intensity is associated with TC recurvature. This is consistent with the previous researches on the relationship of recurvature with TC intensity (Evans and McKinley, 1998; Knaff, 2009; Riehl, 1972). This relationship is further elaborated in Chapter 7.

With the advancement of dynamic modeling, data mining methods and dynamic modeling can complement each other in studies of, for example, TC recurvature and landfall. The research findings of this study can be used to improve the dynamic model

by fine-tuning the parameters and verifying the modeling results. The variability of seasonal TC landfall frequency is attributed to the modulation of ENSO, MJO, QBO, AMO and PDO (e.g., Goh and Chan, 2010; Liu and Chan, 2003). The indices of these oscillations may exercise a certain influence on TC landfall. These indices can be taken as potential attributes of TC landfall in future study.

The present, two-part study (Chapter 4 and Chapter 5) examined TC recurvature and landfall using the C4.5 algorithm. Two classification trees with derived rules are unraveled and verified using the 2010 TC dataset. In the meantime, this study is limited by data availability (e.g., 6-hourly intervals, a lack of data about internal TC dynamics and the spatial resolution of TC-related meteorological data). In further study, data with shorter time intervals (e.g., 3-hour, 1-hour, ½ hour, and 15-minute) and 3-D radar data are expected to unravel the rules relating to TC recurvature and landfall. In order to find any complementarity and enrich the rules, comparisons can be made between 6-hourly rules and different time-interval rule sets in order to check rule-set consistency and stability.

A real-time system for TC analysis and prediction is of great operational significance, based on the integration of the rules governing TC recurvature and landfall. A prototype has been built and is described in Chapter 8.

Chapter 6: Cluster Analysis of Post-landfall TC Tracks

6.1 Introduction

TCs cause most of their damage in coastal or inland regions during or after landfall. China, with a coast that is severely exposed to TCs, has suffered disastrous losses inflicted by such climatic processes; for example, typhoon 'Nina' caused more than 10,000 deaths in Henan province, China, in 1975 post-landfall. More recently, the tropical storm 'Bilis' led to more than 1000 deaths in Hunan province in 2006 after its making landfall in Fujian province, after which it decayed to a tropical depression. Therefore, our understanding of the characteristics of TCs post-landfall is of great significance both for scientific research and operational forecasting.

During the last few decades, numerous researches have been conducted on the TC landfall process and activities; for example, the relationship between rainfall distribution and landfall is widely investigated by dynamic modeling (Blackwell *et al.*, 2006; Kim *et al.*, 2009a; Medlin *et al.*, 2007; Yu and Cheng, 2008; Yu *et al.*, 2010). In addition, surface wind distribution and variability are paid wide attention by empirical models and observational analysis due to the direct damage induced by the strong, raging winds around TCs (Bhowmik *et al.*, 2005b; Fujibe and Kitabatake, 2007; Powell, 1982; Powell and Houston, 1998; Powell *et al.*, 1991).

TC-land interaction during landfall has been studied through observational analysis and dynamic modeling in the last few decades (Bender *et al.*, 1985; Chan and Liang, 2003; Shen *et al.*, 2002; Tuleya *et al.*, 1984; Wong and Chan, 2006b, 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 have confirmed that a high correlation exists between the areas of maximum rainfall during the TC landfall and the mountainous terrain of the region (Brunt, 1968; Hamuro and Coauthors, 1969). Brand and Blelloch (1973, 1974) found that topographical effects, if not considered properly,

can cause significant errors in the forecasting of a TC's motion. Chang (1982) also observed that the TC-terrain interaction caused strong easterlies to develop to the north of the island, accelerating the TC.

With respect to studies on post-landfall TC, empirical models and observational analysis were used to predict the decay rate of TC intensity after making landfall over the South China coast (Wong et al., 2008), in the Indian region (Bhowmik *et al.*, 2005b), and over the U.S. coast (Goldman and Ushijima, 1971; Kaplan and Demaria, 1995, 2001; Vickery, 2005). In addition to these studies, the heavy rainfall mechanism associated with post-landfall TCs has undergone numerous investigations (e.g., Tao et al., 1994; Yu et al., 2010; Zhang et al., 2007). There have been several studies related to the sustaining and movement of TCs post-landfall (Li et al., 2004a; Li et al., 2004b; Yuan et al., 2008; Yuan et al., 2007). Therefore, these existing investigations on post-landfall TCs focus on the rainstorm mechanism associated with mesoscale circulations and landfall surface, as well as the mechanisms and characteristics of TC sustaining post-landfall through dynamic modeling. Despite the rapid development of observational analysis and dynamic models, the study of the characteristics of post-landfall TCs' movement has received little attention in the last few decades.

Data Mining, in general, is the process of extracting hidden and useful patterns and information from data (Fayyad and Stolorz, 1997; Han and Kamber, 2006; Leung, 2010). In recent years, a number of data mining algorithms have been employed to unravel TC tracks and intensities from a TC database (e.g., Camargo *et al.*, 2004; Camargo *et al.*, 2008; Camargo *et al.*, 2007b; Cheng *et al.*, 2008; Gaffney and Smyth, 1999; Gaffney *et al.*, 2007; Gaffney, 2004; Harr and Elsberry, 1991; Lee *et al.*, 2007). The K-means clustering method (MacQueen, 1967) has been applied to TCs in WNP (Elsner and Liu, 2003) and north Atlantic (Elsner, 2003) areas. It is pointed out in Camargo (2007b) that K-means cannot accommodate tracks of different lengths. The FMM-based clustering algorithm, used to fit the geographical shapes of the trajectories, enables the clustering to be implemented in a strictly probabilistic framework and adapt TC tracks of various lengths (Camargo et al., 2007a, b). FMM-based clustering has been applied to TC tracks

in WNP (Camargo et al., 2007a, b), the North Atlantic (Kossin et al., 2010; Nakamura et al., 2009), and the Eastern North Pacific (Camargo et al., 2008). However, little attention has been focused on the study of post-landfall TC tracks through cluster analysis (e.g., K-means, FMM-based, and fuzzy clustering). Clustering or cluster analysis is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters) (Han and Kamber, 2006; Leung, 2010). These clusters can be natural groups of variables, data-points or objects that are similar to each other according to a defined similarity (e.g., distance and probability) (Leung, 2010). Cluster analysis is taken as an initial exploration of the data and always sets the stage for the further mining of structures and processes (Han and Kamber, 2006; Leung, 2010). It is unsupervised learning because the clusters, structures and patterns to be unraveled are unknown a priori. With the development of TC observation and storage equipment, it is natural that useful patterns and knowledge are hidden and unknown a priori among the historical post-landfall TC tracks. Therefore, cluster analysis is used to discover the hidden clusters, structures and patterns from the collected post-landfall TC tracks. As found in previous researches that land surface or elevation have a significant influence on post-landfall TC tracks (e.g., Li et al., 2004a; Li et al., 2004b; Yuan et al., 2008; Yuan et al., 2007), the elevation at the place where the TC center is located should be taken into account in post-landfall TC studies. Therefore, in the present study, we will apply cluster analysis to the post-landfall tracks of TCs making landfall over the Chinese coast. The primary objective of this research is to unravel the patterns and structures hidden in the historical post-landfall TC tracks, to analyze the characteristics of each cluster and to interpret each cluster from the viewpoint of TC theory, especially the theory on TC movement and sustaining. The FMM-based clustering algorithm is used for KDD.

This chapter is organized as follows. Section 6.2 introduces the methods and data source. The results based on the FMM-based clustering algorithm are elaborated in section 6.3. Section 6.4 explains and analyzes the clustering results. The conclusion is drawn in section 6.5.

6.2 Data and Methods

6.2.1 Data

The cluster analysis in this study aims to group TCs with similar "shape type" and "length" into clusters. The "shape type" of a curve fitted by polynomials is represented by the coefficients of the polynomials. A TC track is depicted by connecting the discrete longitude-latitude observations. The polynomials are based on the time orders of TC observations (refer to section 2.4). It is suggested by previous researches that land surface or elevation has a significant influence on post-landfall TC tracks (e.g., Chan and Liang, 2003; Li *et al.*, 2004a; Tuleya *et al.*, 1984; Wong and Chan, 2006b, 2007; Wong *et al.*, 2008; Yuan *et al.*, 2008; Yuan *et al.*, 2007). Therefore, the longitude, latitude, TC time orders and elevation at the TC center are used to build the clustering model. The longitude, latitude and TC time orders are derived from the TC best track. DEM data are used to derive the elevation at the TC center for the cluster analysis. The procedures for building the models are elaborated in section 2.4.

The TC dataset used in this chapter is the JMA best track (available online at http://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/besttrack.html). The best track data set from 2000 to 2010 is used to determine the TC track types. Only TCs that attain at least tropical storm intensity level are considered in our analysis. The composite analysis utilized different data sources. The atmospheric composites (e.g., geopotential height) were obtained from NCEP FNL Operational Global Analysis data, with a 1°×1° spatial resolution and 6-hour intervals (the web access is http://dss.ucar.edu/datasets/ds083.2/). The cluster analysis is applied to time intervals from 2000-2010 and from 1951-2010 and we found that the clusters of post-landfall tracks for both time periods are essentially the same. This sensitivity analysis lends credence to the results derived from the post-landfall TC tracks for the period from 2000-2010.The elevation data are the GTOPO30, available at the website of USGS. The DEM data (tiff file format) is processed on ArcGIS 9.3 platform to extract the elevations at each TC center.

6.2.2 Clustering Methodology

The curve clustering algorithm used in this study is based on the algorithms employed in Camargo (2007b) and Gaffney (2004). It is based on FMM (e.g., Everitt and Hand, 1981; McLachlan and Basford, 1988; McLachlan and Krishnan, 1997; McLachlan and Peel, 2000; Wong et al., 2008), which is capable of modeling probability densities, especially non-Gaussian density, using a small set of basic probability component densities (Camargo et al., 2007a, b). This algorithm employs mixed polynomial regression models (i.e., curves) to fit the geographical "shape" of the TC tracks and model a TC's longitudinal and latitudinal positions versus time (Gaffney et al., 2007). Through FMM, non-Gaussian density functions can be expressed as a mixture of uni-modal component probability distribution functions. The model is fit to the data by maximizing the likelihood of the parameters, given the historical TC datasets. In order to integrate the elevation at the TC center, the model is tuned by taking the elevation at the TC center as an independent variable. The tuned model integrates the TC's longitudinal and latitudinal observations, time and elevation at the TC center.

The assumption is that each TC track is generated by one of K different regression models, each having its own shape parameters. The objectives of modeling are to estimate and learn the parameters of all K models given the TC tracks, and to infer to which of the K models each TC track is most likely to belong. Each TC track is assigned to the mixture component (i.e., the cluster) which is the most likely to have generated the track, given the FMM. The assigned cluster has the highest posterior probability, given the TC track. Expectation Maximization (EM) algorithm is used to estimate these model parameters. The EM algorithm is straightforward to implement and use, and it has been widely used for parameter estimation in the FMM problem. The FMM can be defined in a manner that analogous to that for mixtures (DeSarbo and Cron, 1988; Gaffney et al., 2007; McLachlan and Basford, 1988; McLachlan and Peel, 2000). The details of the FMM-based clustering algorithm are presented in chapter 3.

6.2.3 Number of clusters

A sophisticated problem associated with cluster analysis is to determine the most appropriate number of clusters. The log-likelihood is defined as the log-probability of the observed data under the model, which can be seen as a goodness-of-fit metric for probabilistic models. Used as an objective measure, one selects the number of clusters for which the log-likelihood is greatest across a candidate set of values. It is shown in Figure 6-1 that the log-likelihood attains its maximum value when the number of clusters is set to be three, given the post-landfall TC tracks from 2000-2010, by ten-fold cross-validation. The log-likelihood fluctuates from 4 to 7 and is smaller than the value when the number of clusters is 3.

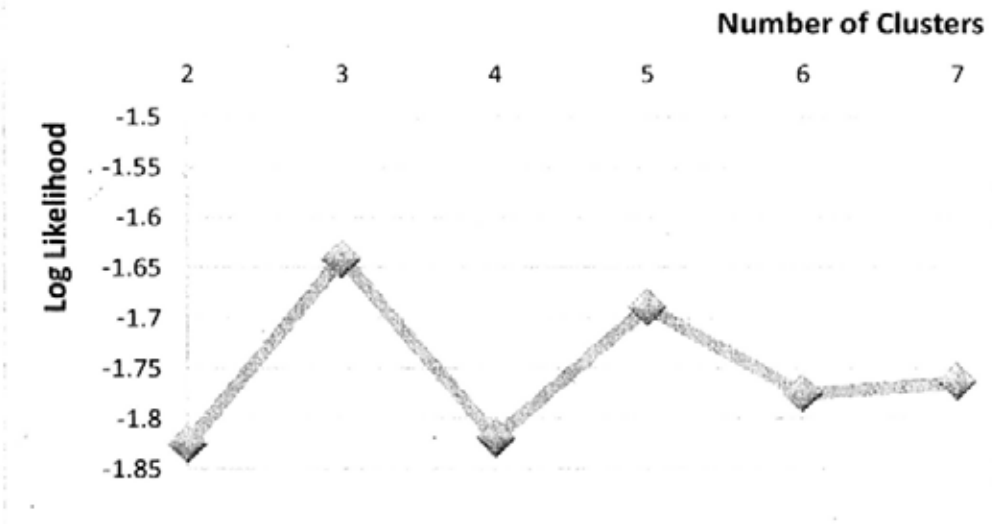


Figure 6-1 Log-likehood values for differenet numbers of TC track clusters (i.e., 2-7).

The clustering results with different number of clusters are shown in Figure 6-2, given the post-landfall TC tracks from 2000-2010. The number of clusters is set at 2, 3, 4, 5, 6 and 7. It can be observed that the difference between clusters 3 and 4 is due to splitting cluster 1 into 2 clusters. However, the two resultant split clusters have quite similar directions and "shapes". The characteristics of cluster analysis with a number of clusters larger than 4 are similar in nature. As the number of clusters increases, the clustering processes seem to split the existing cluster. Therefore, the number of clusters is set at three during the implementation of the algorithm. The cluster analysis results of post-

landfall TC tracks from 1951-2010 is shown in Figure 6-3 to verify whether or not the clustering is sensitive to the amount of datasets. There are 358 post-landfall TC tracks and 3826 TC points as a whole. The TC tracks are clustered, with a number of clusters from 2 to 7 (Figure 6-3). It can be also observed that three appears to be the most appropriate choice for the number of clusters. As the number of clusters increases, they are generated in a similar manner, as shown in Figure 6-2. A greater number of clusters do not carry richer information. In other words, three clusters are appropriate for the post-landfall TC samples from 1951-2009. Meanwhile, the cluster analysis based on FMM is not sensitive to the number of samples or length of time in this case.

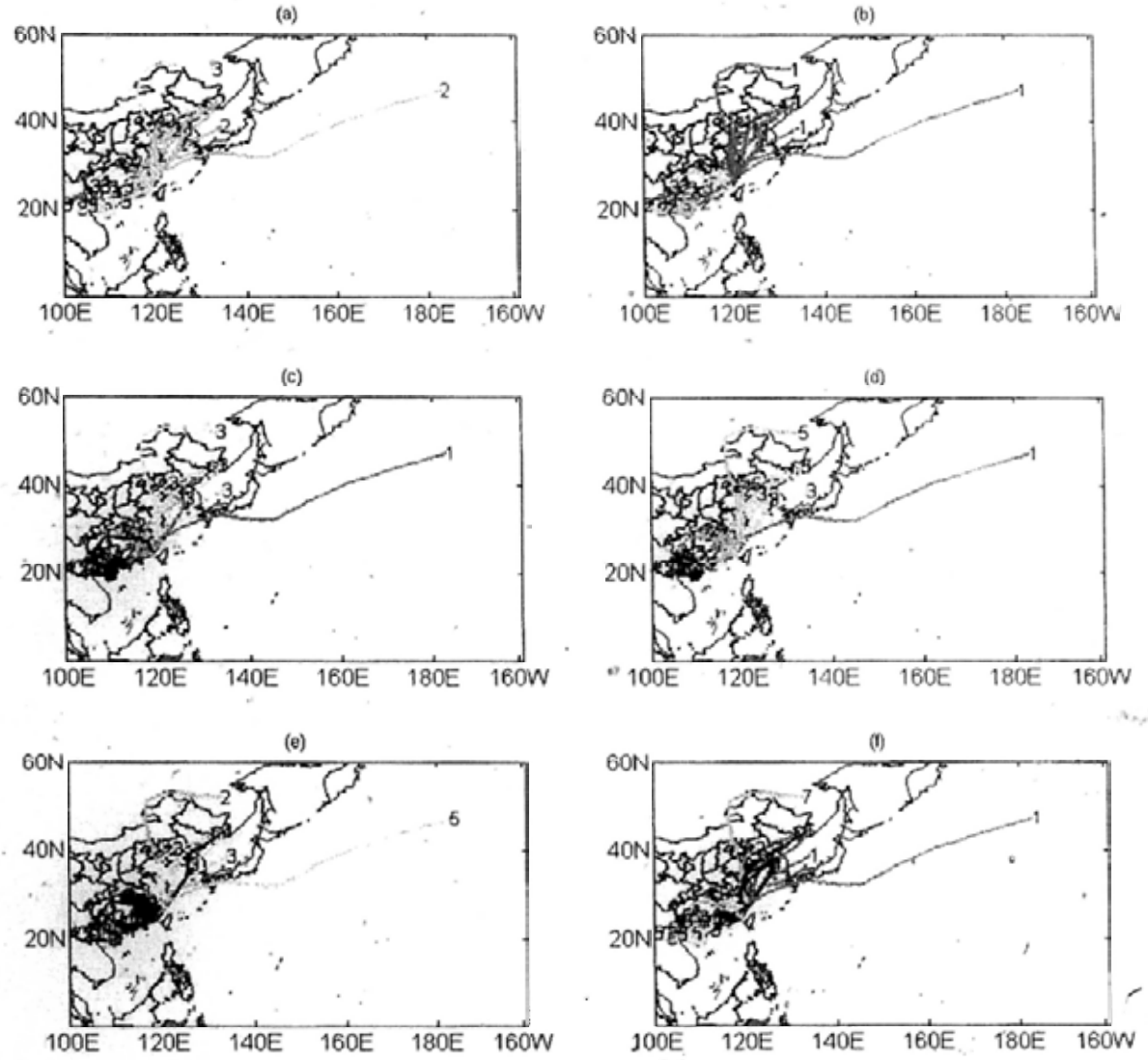


Figure 6-2 The clustering results for post-landfall TC tracks from 2000-2010 by setting the number of clusters at: (a) 2 (b) 3 (c) 4 (d) 5 (e) 6 (f) 7

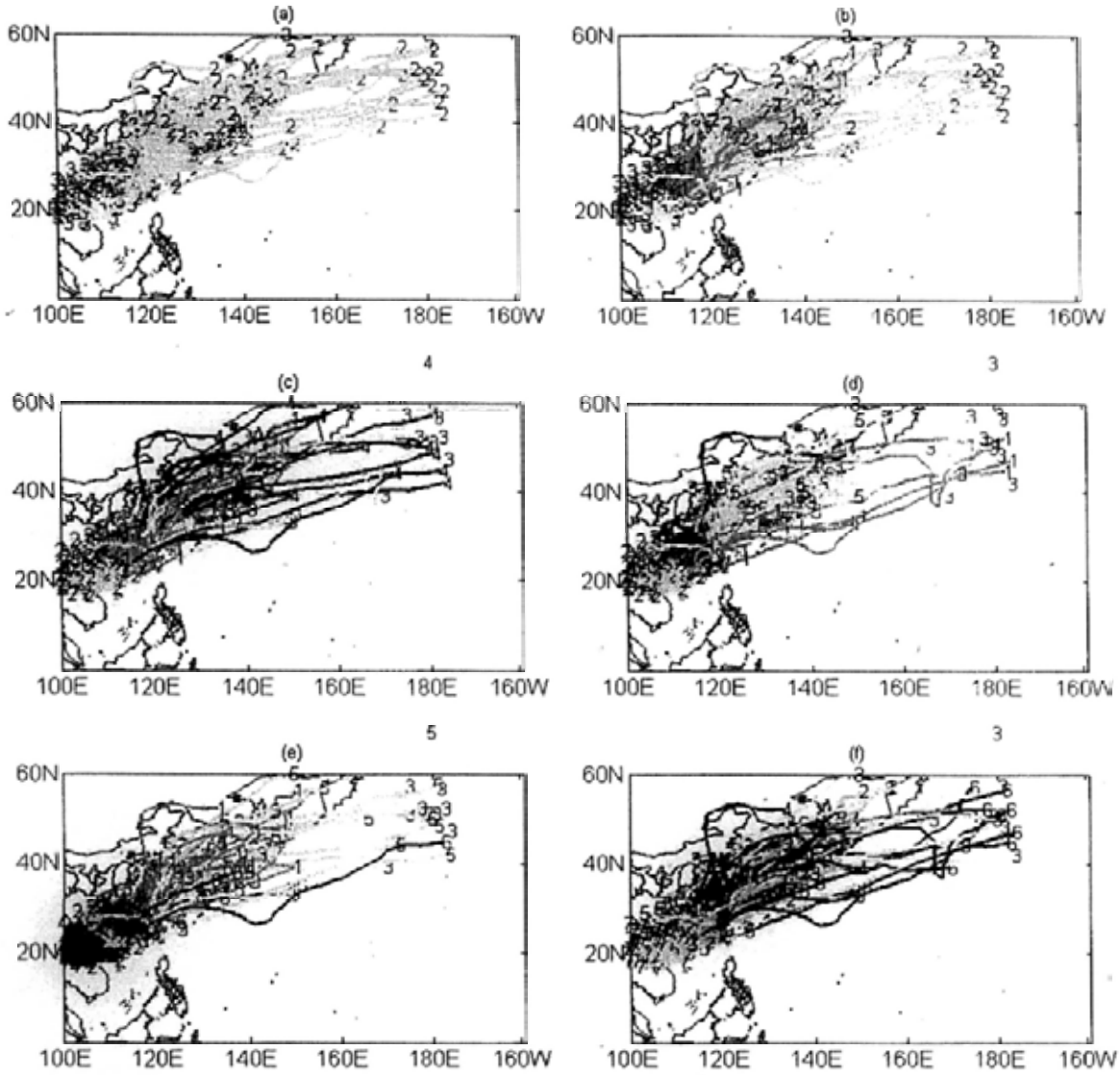


Figure 6-3 The clustering results for post-landfall TC tracks from 1951-2010 by setting the number of clusters at: (a) 2 (b) 3 (c) 4 (d) 5 (e) 6 (f) 7

6.2.4 Model Improvements

The building processes of the clustering model accommodating the variables that may influence TC tracks post-landfall are elaborated in subsection 3.2.2. The clustering model employs a polynomial regression model to simulate TC tracks and integrating the spatial alignment. Apart from this, many other ways of building a clustering model exist, such as the spline regression model or time alignment (Gaffney, 2004). A general, iterative ML

procedure called EM algorithm provides an efficient framework for parameter estimation in the mixture model context (Dempster et al., 1977; McLachlan and Krishnan, 1997). EM is an approximate root-finding procedure that is utilized to search the root of the likelihood equation and it iteratively searches for a set of parameters Θ maximizing the probability of the observed TC data.

Based on the existing studies on TC landfall (Bender *et al.*, 1985; Chan and Liang, 2003; Shen *et al.*, 2002; Wong and Chan, 2007; Wong *et al.*, 2008), the land surface interacts with TCs post-landfall. This TC-land interaction modulates TC movement post-landfall. In addition, environmental flows and large-scale circulation have proven to exert a dramatic influence on TC movement. These parameters should influence TC movement post-landfall. Therefore, the elevation at the TC center and the parameters characterizing large-scale circulation are taken as the potential parameters affecting post-landfall TC clusters. These parameters are, therefore, integrated for cluster analysis. In addition to the time and elevation, other parameters are categorized into three groups: parameters measuring large-scale circulation (the subtropical high, monsoon systems and westerlies), parameters describing the wind fields (the 200 hPa and 850 hPa zonal winds), and parameters characterizing the TCs (e.g., central pressure, unit: hPa). Figure 6-4 illustrates the clustering results for different groups of parameters: (a) time; (b) elevation and time; (c) elevation, central pressure, time and parameters measuring large-scale circulation; (d) elevation, time and wind fields; (e) elevation, time, and parameters measuring large-scale circulation; and (f) elevation, time, parameters measuring large-scale circulation and wind fields. The clustering result based on time is inferior to that generated by the model on time and elevation at each TC center. The results based on time and elevations are more appropriate than those derived from other groups of parameters. Therefore, three clusters are found to be derived from post-landfall TC tracks (Figure 6-4b).

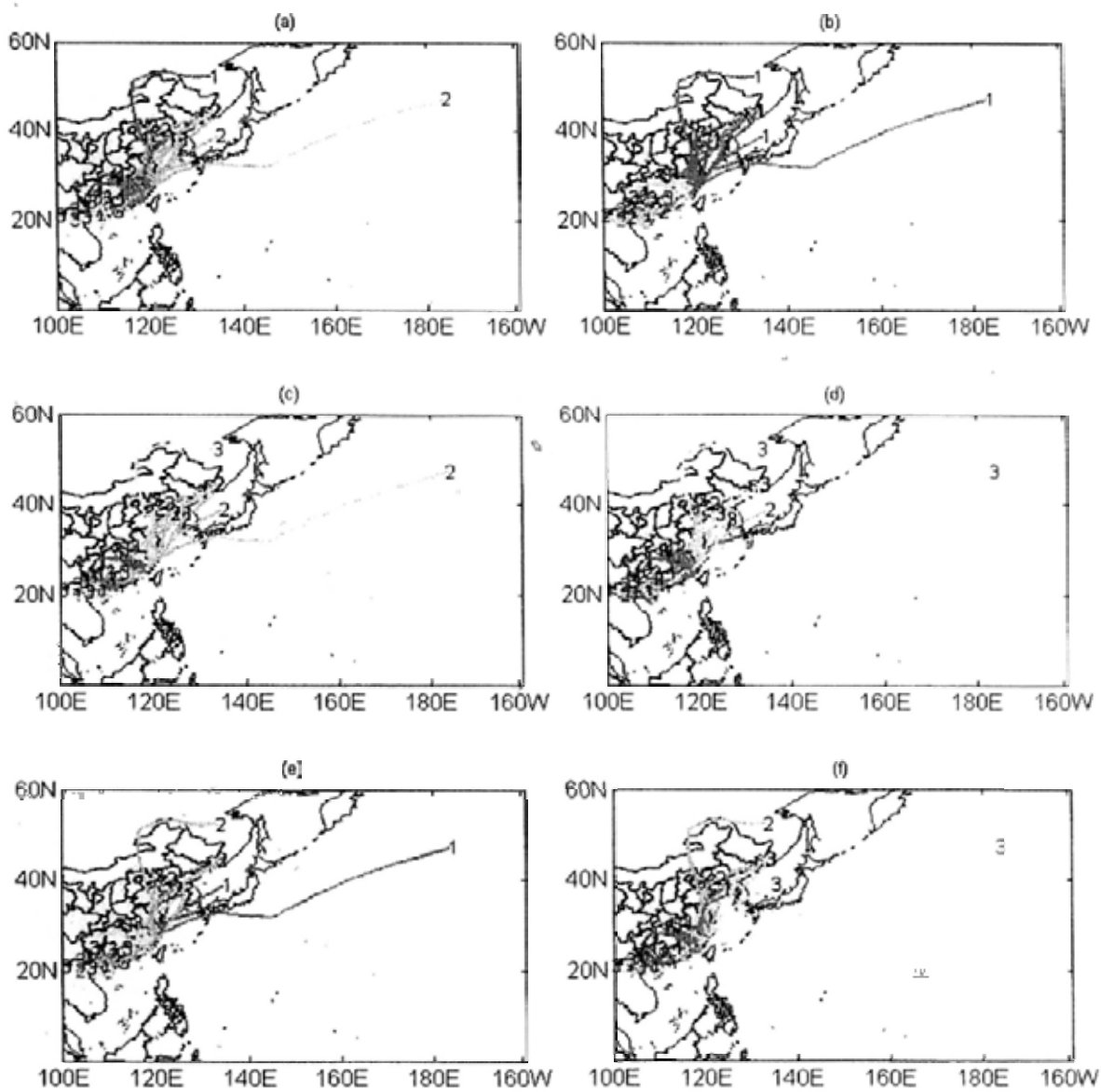


Figure 6-4 Clustering of TC tracks over WNP, during the period 2000-2010, based on the variables: (a) time (b) elevation and time (c) elevation, central pressure, time and large-scale circulation parameters (d) elevation, time and wind fields (e) elevation, time, and large-scale circulation parameters and (f) elevation, time, large-scale circulation parameters and wind fields

6.3 Results

Based on the FMM-based clustering algorithm introduced in the previous section, the historical post-landfall TC tracks are grouped into three clusters (Figure 6-5). Each

cluster has its special characteristics in terms of spatial distribution and lifespan.

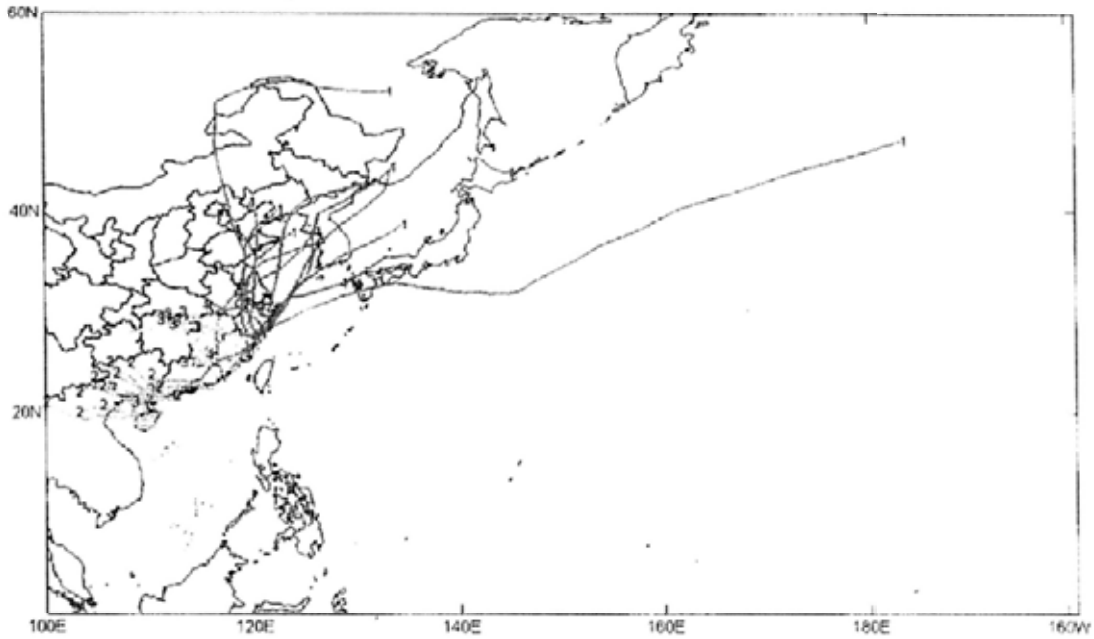


Figure 6-5 The three clusters unraveled from the historical post-landfall TC database for the period 2000-2010 using an FMM-based clustering algorithm. The number at the end of the track signifies the cluster number.

The post-landfall tracks in Cluster 1 make landfall over the southeast coast of China, including Guangdong province, Fujian province, and Zhejiang province. Following landfall, cluster 1 heads northward or northeastward. Most of them move back over the ocean and make secondary landfall over Liaoning province, the Korean Peninsula or Japan. These TCs eventually recurve to the mid-latitude region. Based on past researches on the extra-tropical transition of TCs in WNP, cluster 1 is prone to transit to extra-tropical cyclones during its course. Cluster 3 makes landfall over Guangdong province and Fujian province. Post-landfall, the TCs move over land and eventually dissipate inland. Cluster 2 makes landfall in Hainan province and across the western Guangdong coast. Post-landfall, the TCs move inland to Guangxi province or cross the ocean to make secondary landfall over the eastern Vietnamese coast. The moisture latent heat flux and baroclinic potential energy (Harr and Elsberry, 2000; Harr et al., 2000; Sinclair, 2002; Thorncroft and Jones, 2000) from the mid-latitude synoptic systems are two crucial factors for TC sustaining (Chen et al., 2004; Li et al., 2004a; Li et al., 2004b). The large-scale circulation (e.g., the subtropical high) has a significant influence on TC tracks post-landfall (Chen et al., 2004). The TC motion and sustaining mechanism post-landfall are

also influenced by the elevations and land surface (e.g., Brunt, 1968; Chan and Liang, 2003; Chang, 1982; Hamuro and Coauthors, 1969); for example, TCs may experience track deflections when passing through Taiwan (e.g., Wu, 2001; Wu and Kuo, 1999; Yang *et al.*, 2008). The three clusters discovered from the post-landfall TC tracks are interpreted from the perspective of a moisture supply, large-scale circulation, elevation and baroclinic potential energy.

6.4 Interpretations

TC movements are largely controlled by the steering flow (e.g., Chan, 1985b, 2005a; Holland, 1993a; Holland, 1983b), which is characterized by the streamline and the subtropical high. The composite 500 hPa geopotential height and streamline of each cluster are plotted to show the environmental flow surrounding TCs. Given the characteristics of the post-landfall tracks of three clusters, these tracks are attributed to the environmental flows surrounding TCs post-landfall. Figure 6-6 illustrates that the steering flows surrounding the TC centers in the three clusters differ vastly. The steering flow (i.e., deep-layer mean wind fields from 850 to 300 hPa layer) surrounding the TC in cluster 1 is largely westerly. The subtropical high is located to the southeast of the TC center. However, the steering flows of clusters 2 and 3 are mostly easterly. The steering currents of clusters 2 and 3 drive the TCs to move westward after landfall. Figure 6-7 illustrates the remarkable difference between the 500 hPa geopotential heights of the three clusters in the region 70 °E to 160 °W longitude and 0°N to 70 °N latitude. The subtropical high depicted by the 5870 gpm or more contours is located to the east of 130 °E. However, the subtropical highs of clusters 2 and 3 shift westward to about 110 °E and 90 °E respectively. Cluster 1 moves northward and northeastward under the influence of an eastward-shift subtropical high (Figure 6-7). However, clusters 2 and 3 tend to go westward because the subtropical high shifts westward (see Figure 6-7).

The elevations of land surface where the TC centers are located influence the sustaining and decay rate of TCs post-landfall. Given the characteristics of TC landfall locations and tracks, the average elevations of TCs in cluster 1 are higher than those for cluster 2

according to the Student's *t* test. Besides, the average elevations of the TCs in cluster 3 are higher than those for cluster 2 according to the Student's *t* test. No significant relationship is observed between clusters 1 and 3. The average elevation of cluster 2 is the lowest of the three (see Table 6-1). Cluster 1 makes landfall over more northern areas than does cluster 3 (see Figure 6-5). Therefore, the mean elevations of cluster 1 are lower than those for cluster 3, despite the fact that the relationship is insignificant (see column 'F-value' in Table 6-1). The lower elevation of cluster 1 and the westerly due to the northern latitude cause the TC to recurve.

Table 6-1 Comparison of Average Elevations of Clusters 1, 2 and 3 within 12 hours post-landfall by the Student's *t* test

	F-value	**Sig.**	**t-value**	**d.f.**
Cluster1-Cluster2	11.245	.001	1.606	56
Cluster2-Cluster3	6.271	.015	-1.891	52
Cluster1-Cluster3	.218	.643	-.254	50

TC tracks are usually influenced by the steering flow (e.g., Chan, 1985b, 2005a; Holland, 1993a; Holland, 1984). Besides, some studies have found that the moisture supply and baroclinic potential energy play significant roles in sustaining TCs post-landfall.

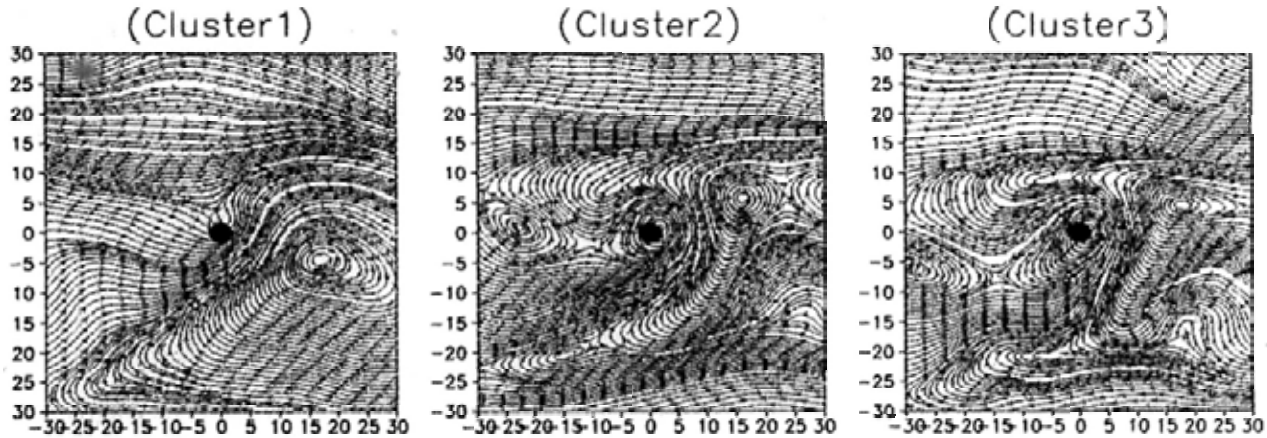


Figure 6-6 Composite streamlines of three clusters in the region 60° × 60°, centering at the TC center. Coordinate (0, 0) marks the typhoon's center.

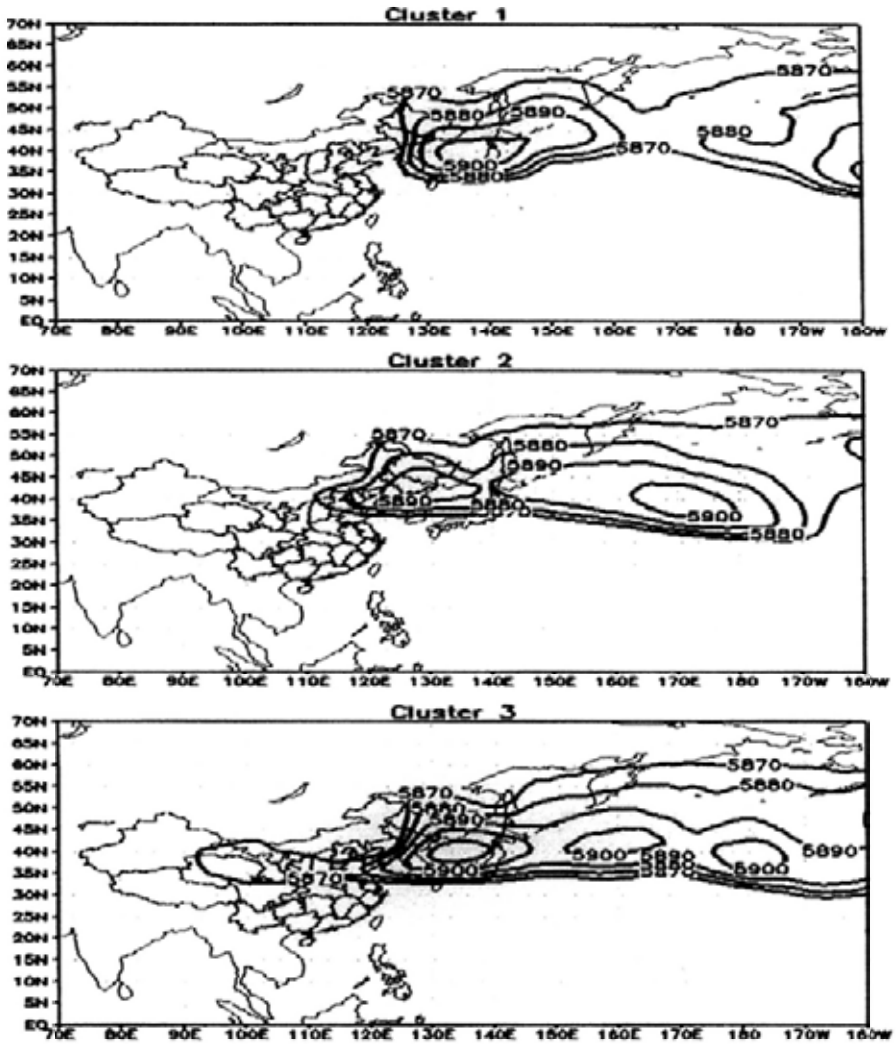


Figure 6-7 The composite 500 hPa geopotential height of three clusters based on the post-landfall TC tracks for the period 2000-2010. The 5870 gpm or higher contours indicate the high center of the subtropical high.

The characteristics of water vapor supply (e.g., moisture transfer) and wind fields on the 850 hPa of clusters 1, 2 and 3 are depicted in Figure 6-8. It is shown that the average moisture transfer of cluster 1 is weak and is cut off during the average time for this cluster. The moisture transfer of cluster 1 at landfall is relatively strong. The moisture transfer is cut off at 24 hours post-landfall and becomes dramatically weaker at 48 hours post-landfall.

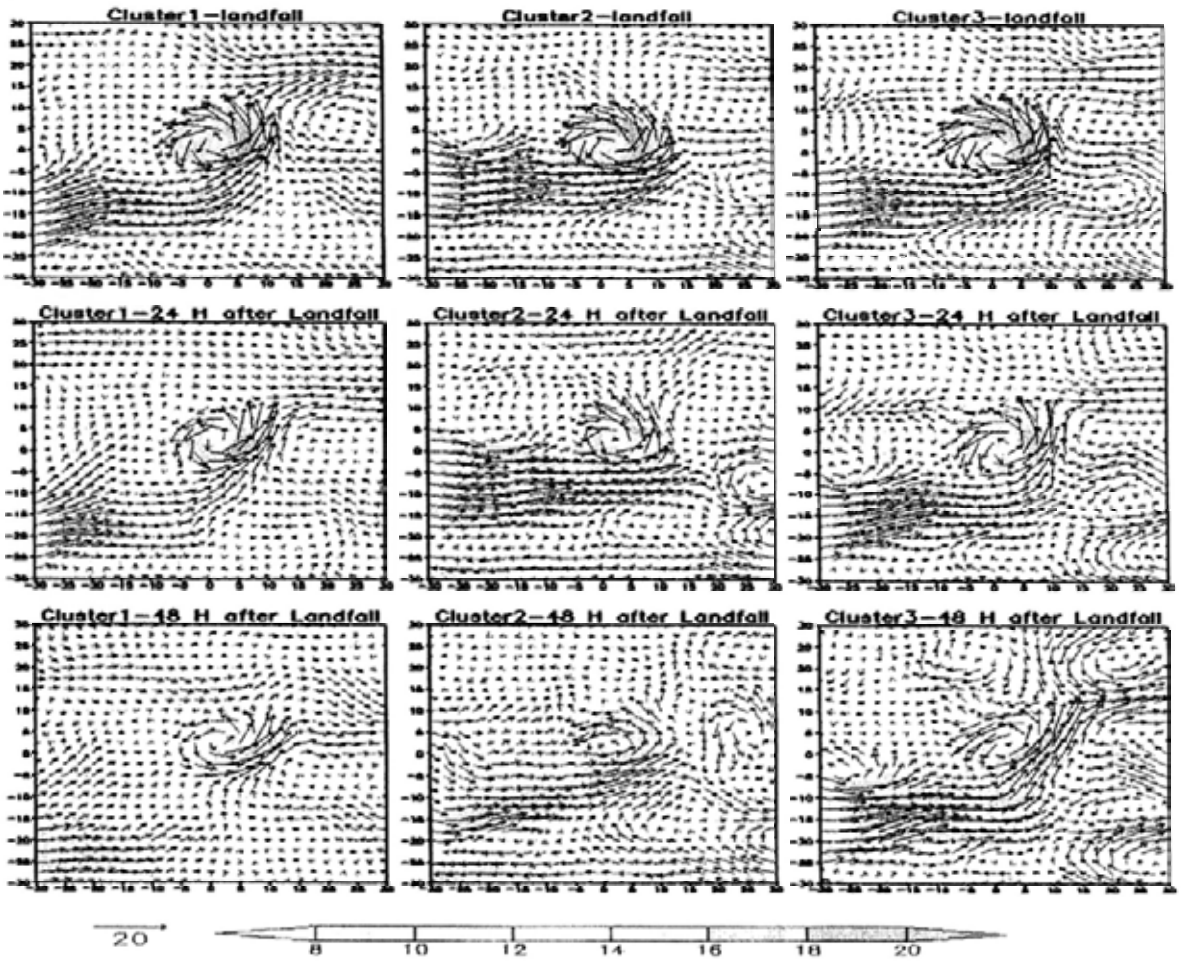


Figure 6-8 The composite 850 hPa water vapor flux (unit: $g \cdot s^{-1} \cdot hPa^{-1} \cdot cm^{-1}$) of clusters 1, 2 and 3 at different times related to landfall. The first row marks the water vapor flux at landfall; the second row the water vapor flux 24 hours post-landfall; and the third row the water vapor flux 48 hours post-landfall.

However, the water vapor fluxes of clusters 2 and 3 are much stronger than that for cluster 1 for all of the selected time periods; for example, at landfall, and 24 or 48 hours post-landfall. The moisture supply at 48 hours post-landfall is still strong for clusters 2 and 3, especially cluster 3 (see Figure 6-8). It is noted that cluster 1 contains the post-landfall TC tracks sustaining for a longer time than those of clusters 2 and 3. The existing findings on TC development and sustaining indicate that the moisture supply sustains the TC to a large extent. However, cluster 1 sustains for a much longer time than clusters 2 and 3, given the weak moisture supply. Therefore, there should be parameters other than moisture supply to sustain the TCs in cluster 1 post-landfall.

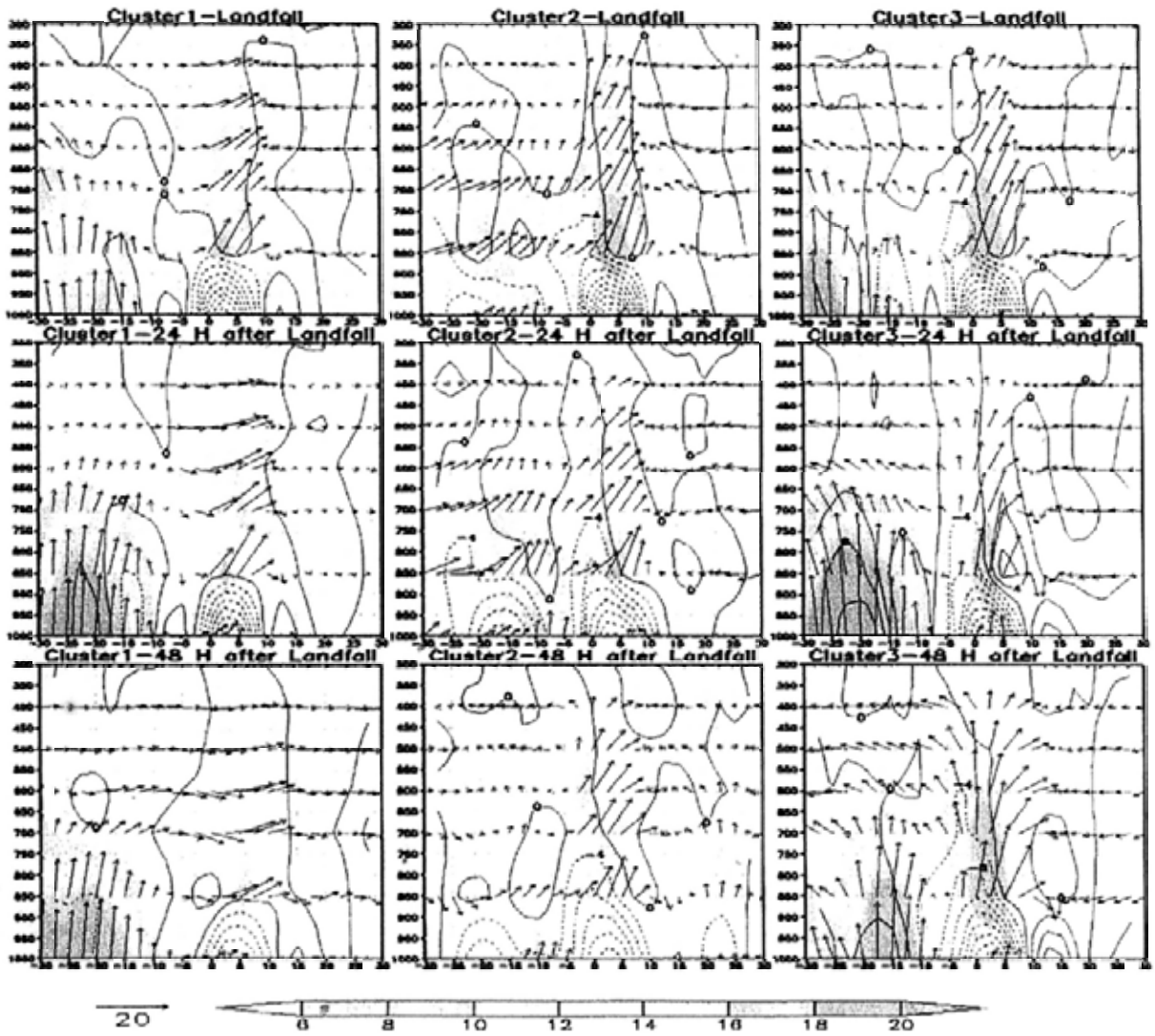


Figure 6-9 The latitudinal profile map of vertical water vapor flux divergence (unit: 10^{-8} g·s^{-1}·hPa^{-1}·cm^{-2}) and the vertical water vapor flux ω*q (10^{-4} g·kg^{-1}·hPa^{-1}·s^{-1}) of clusters 1, 2 and 3 at different times. The first row marks the vertical water vapor flux (from 1000 hPa to 300 hPa layer) at landfall; the second row the vertical water vapor flux 24 hours post-landfall; and the third row the vertical water vapor flux 48 hours post-landfall. It is noted that -30 and 30 along the horizontal axis mean 30° to the south and north of the TC center respectively.

Figure 6-9 shows that the vertical motion of the water vapor of cluster 1 is much weaker than that of clusters 2 and 3. The vertical motion and moisture transfer disappear at 48 hours after landfall. However, those of clusters 2 and 3 continue to rise rapidly to the high-level troposphere at 48 hours after landfall. It is noted that the composites in Figure

6-9 are averaged over the TC samples that still sustain and live in some time period (e.g., 24 hours, 48 hours) post-landfall. Therefore, with respect to the TCs existing post-landfall for 48 hours in clusters 2 and 3, the horizontal and vertical moisture transfer plays an important role in sustaining post-landfall TCs.

The profile of westerly and vertical velocity is shown in Figure 6-10. The vertical profile is from the 1000 to the 200 hPa layer. The solid contours signify the westerly whereas the dashed contours the easterly. In the WNP basin, the prevailing westerly indicates the proximity of westerlies. In the first column of Figure 6-10, the three vertical profiles from top to bottom are from different time periods: at landfall, and 24 and 48 hours post-landfall. It can be observed that the solid contour region moves increasingly closer to the TC center as the TC stays longer over land. In other words, the westerlies are closer to the TCs of cluster 1 post-landfall. At 48 hours post-landfall, westerlies almost occupy the low-, middle- and high-level troposphere (700 hPa and above) of cluster 1 (see first column of Figure 6-7). Westerly prevails to the south of the three clusters because of the cyclonic flow surrounding TCs. The TCs in cluster 1 interact with westerlies and obtain their baroclinic potential energy from them via interaction. However, the interaction between TCs and westerlies cannot be observed in clusters 2 and 3 because the westerlies lie far away from the TCs in these clusters. Therefore, clusters 2 and 3 do not have any opportunity to obtain energy from the mid-latitude synoptic systems. The upward motion disappears 48 hours post-landfall in cluster 1, whereas the upward motion (depicted by the pressure vertical velocity) of clusters 2 and 3 prevails at 48 hours post-landfall. This helps to explain the reason why cluster 1 sustains much longer than clusters 2 and 3 without a sufficient water vapor supply.

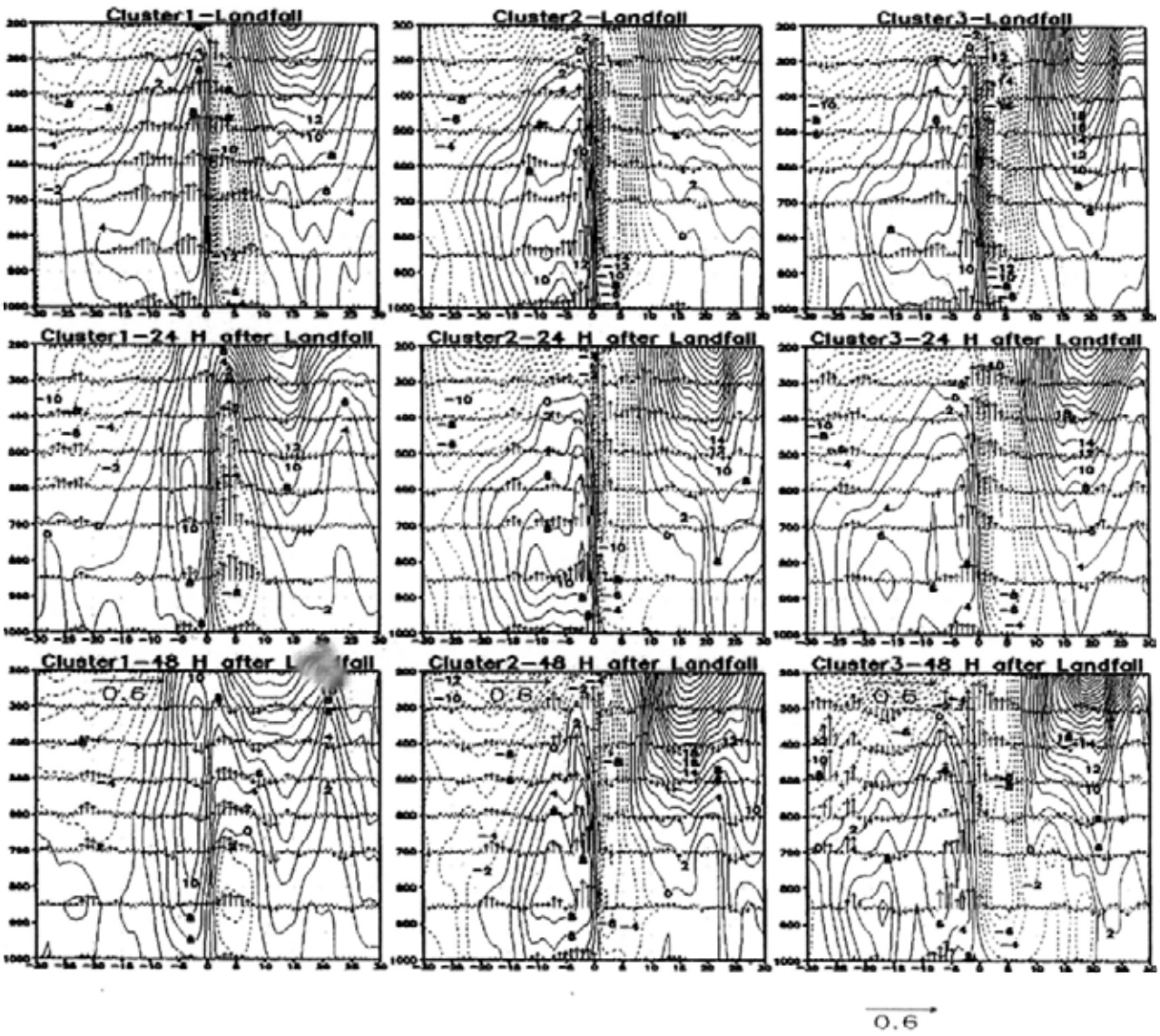


Figure 6-10 Latitudinal profile of pressure vertical velocity (ω means pressure vertical velocity, unit: Pa/s) and zonal wind of clusters 1, 2 and 3(the arrows mark the direction and strength of the vertical velocity and the contours show the zonal wind, unit: m/s). The horizontal axis indicates the latitudes from south (negative) to north (positive).

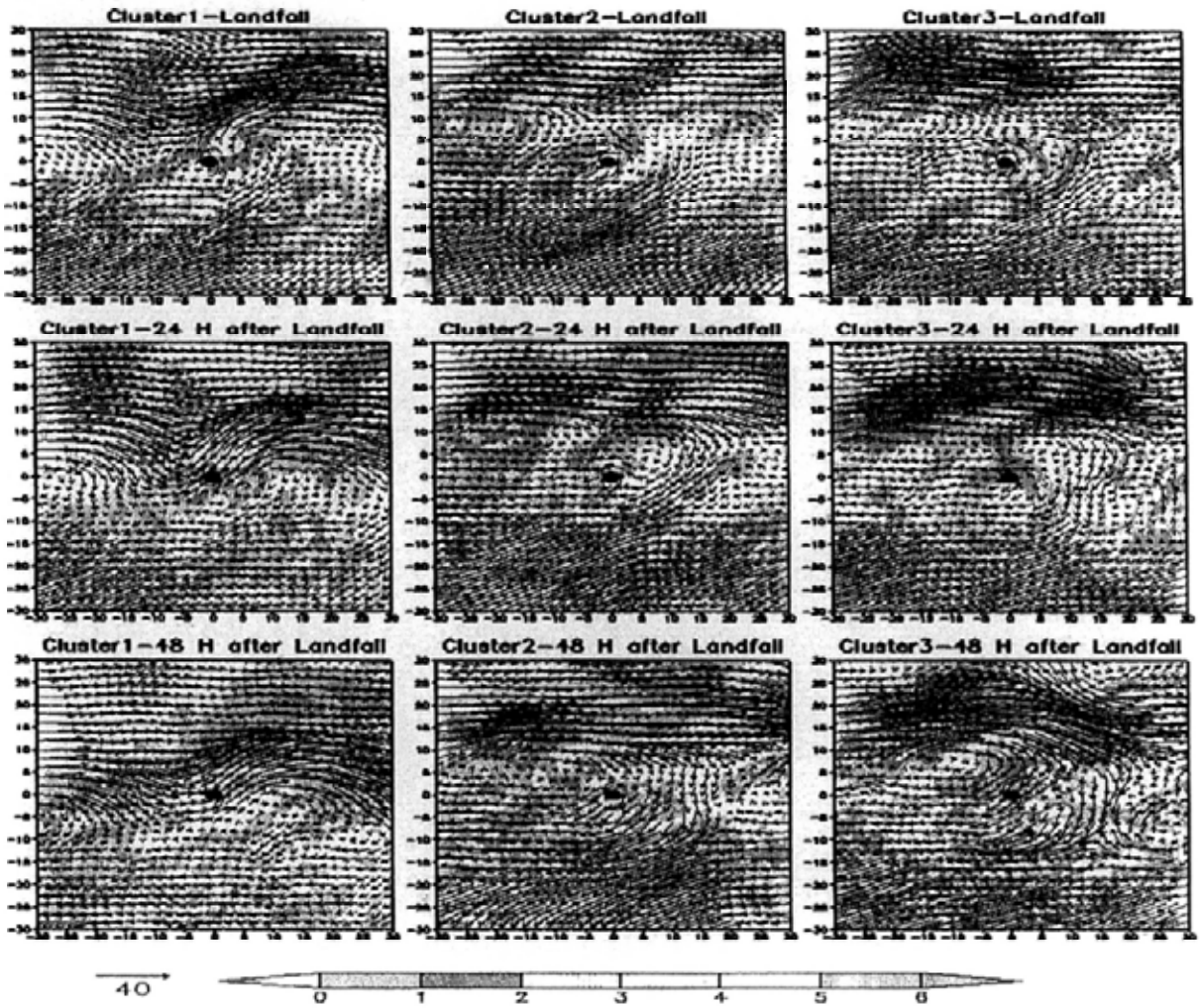


Figure 6-11 The 200 hPa divergence (unit: $\times 10^{-6}$) and zonal wind (unit: m/s) of clusters 1, 2 and 3 at different times post-landfall over the region 60°× 60° centering at TC centers. The symbol at coordinate (0, 0) indicates the typhoon's center.

The TCs have convergence in their low level and divergence in their high level respectively (Bister and Emanuel, 1997; Charney and Eliassen, 1964; Craig and Gray, 1996). Convergence and divergence are good indicators for convections to sustain the TC structure during its lifespan due to the Convective Instability of the Second Kind (CISK) or Wind-Induced Surface Heat Exchange (WISHE) mechanisms (Charney and Eliassen, 1964; Craig and Gray, 1996). Meanwhile, the 200 hPa zonal wind indicates the high level environmental surrounding flow. Figure 6-11 illustrates how the divergence of cluster 1 to the northeast of the TC is larger than that for clusters 2 and 3. There is an appreciable westerly close to cluster 1 at all times post-landfall. However, analogous divergence and

wind field patterns do not exist in clusters 2 and 3. Indeed, the 200 hPa wind fields close to clusters 2 and 3 are mostly easterly by contrast to westerly for cluster 1, which can also be observed in Figure 6-10, which depicts the vertical profile of the zonal wind as well as the upward motion. Therefore, the divergence of cluster 1 indicates that TCs maintain their cyclonic structure post-landfall. The prevailing 200 hPa westerly surrounding cluster 1 implies that cluster 1 interacts actively with the mid-latitude westerlies and obtains baroclinic energy from the mid-latitude synoptic systems in order to maintain the cyclonic structures.

Based on the discussion above, we interpret clusters 1, 2 and 3 from the meteorological perspective. Cluster 1 has a long lifespan and recurves post-landfall, whereas clusters 2 and 3 are short-lived. Cluster 2 mostly moves westward post-landfall. Eventually, most of the TCs in cluster 2 make secondary landfalls across the Vietnamese coast. Cluster 3 moves inland and subsequently disappears. A possible explanation for this is that the subtropical high of cluster 1 propagates more eastward than that of clusters 2 and 3. Besides, the steering flow of cluster 1 is westerly, whereas those of clusters 2 and 3 are easterly. The landfall locations of cluster 1 have higher elevations than those for clusters 2 and 3. Therefore, cluster 1 moves northward and recurves post-landfall, while clusters 2 and 3 move westward or northwestward under the influence of the strong subtropical high and prevailing easterly.

The horizontal and vertical water vapor supply of cluster 1 is weaker than that of clusters 2 and 3. However, it sustains for much longer than clusters 2 and 3. From the vertical profile (from the 1000 to 200 hPa layer) of zonal wind, the westerlies are much closer to cluster 1 than to clusters 2 and 3. The interaction between cluster 1 and the mid-latitude systems (e.g., westerlies) provides the baroclinic potential energy for cluster 1. The strong divergence and southwesterly wind near the TCs of cluster 1 on the high level (200 hPa layer) indicate that cluster 1 maintains its cyclonic structure well. The baroclinic potential energy may play an important role in sustaining the TCs and maintaining the cyclonic structure of TCs.

6.5 Summary

TC landfall has long been considered as one of the most crucial problems in TC study since it triggers tremendous damage to coastal areas. However, the characteristics of TCs post-landfall have been paid insufficient attention by the scientific community.

FMM-based clustering algorithms are applied to post-landfall TC tracks to discover the hidden clusters. Elevations are taken into account in this clustering model because existing studies have found that the landfall surface or elevations influence TC movements post-landfall. Three clusters are uncovered from the historical post-landfall TC tracks. Cluster 1 makes landfall along the Fujian and Zhejiang coast and sustains for a long period, during which it mostly recurves to the mid-latitude region. Cluster 2 makes landfall in Hainan province and across the western coast of Guangdong province. Most of the TC tracks in cluster 2 move over the ocean and make secondary landfalls over the Vietnamese coast and western Yunnan province. Cluster 2 eventually dissipates inland due to the rough landfall surface in western Vietnam and the Himalayas near Yunnan province. Cluster 3 makes landfall over the Guangdong and Fujian provinces. Subsequently, the TCs in cluster 3 move inland and disappear.

The clustering results are interpreted and justified from the meteorological perspective (e.g., the geopotential height, steering flow, vertical and horizontal water vapor supply). We find that cluster 1 moves northward and northeastward and recurves into the mid-latitude region because of a retreating subtropical high and westerly steering flow surrounding the TCs. By contrast, clusters 2 and 3 move westward or northwestward as a result of the westward-shift subtropical high and westward steering flow. Cluster 1 sustains for a longer time than clusters 2 and 3 in spite of its weak horizontal and vertical water vapor supply. It is found that the TCs in cluster 1 interact actively with westerlies during the post-landfall period. On the other hand, we cannot observe any analogous interactions with westerlies in clusters 2 and 3. Besides, the composite 200 hPa divergence of cluster 1 post-landfall is stronger than that for clusters 2 and 3. This

explains why the TCs sustain much better than clusters 2 and 3 post-landfall to some degree.

The focus of this chapter is on the discovery of the patterns in historical post-landfall TC tracks hidden in the TC data. With the development of the dynamic modeling of land-TC interactions, the predictions of post-landfall TC tracks and TC-triggered rainstorms will be advanced in the future. In order to make progress in disaster response, mitigation and management, the data mining algorithms, such as FMM-based algorithms used in this study, will complement dynamic modeling in post-landfall TC study.

CHAPTER 7: Tracks and Intensities

7.1 Introduction

TC tracks and intensities are two basic problems of scientific and operational concerns. TC track forecasts have become increasingly accurate over last several decades (Wang and Wu, 2004), thanks largely to a combination of better observations, especially the satellite (Soden et al., 2001) and dropsonde (Aberson and Franklin, 1999; Burpee et al., 1996). Nonetheless, researches on TC intensities greatly lag that of TC movement (e.g., Knaff et al., 2005; Wang and Wu, 2004). TC landfall and recurvature are of paramount interest in TC researches. The relationship between TC movement and intensities concentrates on the relations of TC recurvature and landfall to intensities. The interaction between TC movement and intensity plays an important role in operational TC prediction. In spite of the great significance in the interaction, relatively little attention has been paid to the interactions between TC movements and intensities. For example, Velden and Leslie (1991) investigated the relationship between TC intensity and moving direction using the depth of the environmental steering layer.

Past researches on the association of TC landfall with intensity focus on the decaying rate during or after making landfall in the South Chinese coast, Korean Peninsula, the Indian coast, the U.S. coast, and the Japanese coast. Based on the existing research findings, most of TCs decay after making landfall. Meanwhile, there exists some unusual intensification of TCs after landfall. However, the specific characteristics of TC intensity change before and after making landfall over the Chinese coast have seldom been examined.

The maximum intensity corresponding to recurvature is under investigation in the last several decades (Evans and McKinley, 1998; Knaff, 2009; Riehl, 1972). It has been reported that TC maximum intensity tends to be coincident with TC recurvature (Evans and McKinley, 1998; Riehl, 1972). Riehl (1972) found that roughly two-thirds of the TCs

that reach typhoon/hurricane intensity attained their maximum lifetime intensity within 12 h of recurvature in the WNP. Instead of 45 and 70% of all WNP TCs peaking within ±12 h and ±24 h of recurvature as reported in EM98 for the period 1980–1996, the datasets from 1980–2006 in Knaff (2009) suggested these percentages are close to 28.9 and 50.0% respectively. It is inferred in Knaff (2009) that the long-held belief that most TCs attain their maximum intensities near recurvature is not supported by the observations. The TC intensity records prior to 1985 should be used with caution (Chu et al., 2002). Nonetheless, the relationship between TC recurvature and maximum intensity is still ambiguous based on Knaff (2009) and the TC intensity changes have not been explored comprehensively with respect to TC recurvature.

MPI is proposed to estimate an upper bound on TC intensity for given atmospheric and oceanic conditions. The MPI is mainly governed by SST (e.g., Emanuel, 1986, 1988, 1997; Emanuel et al., 2004; Malkus and Riehl, 1960). MPI has not been applied to the relationship between intensity and recurvature yet. In this study, MPI will be used to analyze the potential of TCs to intensify related to recurvature.

Instead of counting the frequency of landfalling and recurving TCs, the Student's *t* test and ANOVA are employed to analyze whether the differences between the indicators of several groups are significant. The attempt of this study is to shed light on the relationship between TC tracks (e.g., landfall and recurvature) and intensities through the Student's *t* test and ANOVA.

Therefore, the intensity change with regard to TC recurvature and landfall is investigated by taking into account the MPI and SST to determine the mechanisms controlling the intensification or weakening. The characteristics of two-group TCs (i.e., peaking intensities prior to and after recurvature) are studied by ANOVA and the Student's *t* test to decide whether there is a significant difference. Meanwhile, the Student's *t* test is employed to determine whether the average durations for sustaining maximum intensities between straight movers and recurvers are significantly different.

This chapter is organized as follows. Section 7.2 introduces the data source and methodology. The intensities of landfall TCs are presented in section 7.3. Section 7.4 presents the analysis on intensities of recurving TCs. The conclusion is drawn in section 7.5.

7.2 Data and Methodology

a. Data for Recurving TCs

JTWC TC best track is used in Knaff (2009) to analyze the characteristics of intensity change corresponding to recurvature in WNP and other basins. The TC intensity of JTWC is measured by a 1-min MSW for all named TCs. All of the TCs occurring in these study areas between 1985 and 2009 are drawn as samples. The TC best track data used in this study is available from JMA RSMC Tokyo.

This post analysis best track data consist of 6-h estimates of position (latitude and longitude), MCP, and 10-min MSW for all named TCs in the WNP basin, including the SCS, from 1951 to the present (available online at http://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/besttrack.html). The MSW value of the JTWC best track is therefore larger than that of the JMA best track.

The meteorological parameters, such as latent heat net flux, divergence, relative humidity, water vapor, and upper-troposphere temperature, are extracted from the NCAR/NCEP reanalysis dataset (Kalnay et al., 1996) (available online at http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.pressure.html).The rainfall dataset used in this study is TRMM Multisatellite Precipitation Analysis (TMPA) (Huffman et al., 2007). TMPA at fine resolutions (0.25° × 0.25° and 3 hourly) combines precipitation estimates from multiple satellites, as well as land surface rainfall gauge analyses where possible. The TRMM rainfall covers the area between 50°N and 50°S for the period from 1998 to the present.

TMI Optimum Interpolation (OI) daily SST data (Gentemann et al., 2004) at a 0.25° × 0.25° resolution together with RSMC Tokyo best track data are used to estimate MPI, which is one of the important predictors employed in the TC intensity prediction models. TMI OI SST analysis is available from the RSS website at http://www.ssmi.com and covers the latitude band 40°S–40°N from January 1998 to the present.

In order to investigate the relationship between recurvature and intensity, the MPI is taken into account. However, the SST daily data are only available from January 1998. The TC best track data from 1998 to 2009, together with the NCEP/NCAR dataset over the same period, are applied to the investigation. This study focuses on TCs that form in WNP. Therefore, recurvature is defined as turning from westward to the north and eventually to the northeast in the Northern Hemisphere (Joint Typhoon Warning Center (JTWC), 1988). The point of recurvature is consistent with K09 as the first point where the first derivative of longitude with regard to time is equal to zero and the second derivative of the longitude is positive (i.e., the minimum). However, the theoretical point of recurvature may lack recording due to the 6-hour intervals of the best track data. The point which is nearest to the theoretical point of recurvature is thus regarded as the practical point of recurvature if the theoretical point of recurvature has no records. Not all recurving TCs are drawn as samples. Indeed, the TCs are precluded if they have the following characteristics: (1) the point of recurvature is over land; (2) the maximum intensity of TC is less than 17 m s^{-1}; (3) the looping TCs or equatorward-recurving TCs; and (4) the latitude of any future point cannot be less than the latitude of the recurvature point. This methodology leads to 164 recurving TCs in WNP between 1985 and 2009. Of the 164 recurving TC cases, 137 (0.835) TCs achieve at least typhoon/hurricane intensity (i.e., 65 knot or 33 m s^{-1}), whereas 17 (0.104) TCs reach Category 3 of the Saffir-Simpson Hurricane Scale (Simpson and Riehl, 1981b). It is noteworthy that in order to be consistent with the R72, EM98 and K09, the definition of "time reaching the maximum intensity of the lifespan" is the first time that the TC attains the largest intensity.

7.3 Recurvature

7.3.1 Recurvature and Intensity

It is presented in K09 that are the percentages of TCs that attain the maximum intensity within a specific time period related to recurvature (e.g., recurvature ±12 h). Table 7-1 gives the percentages of TCs attaining the maximum intensity at times related to recurvature. Of the TCs with tropical storms' strength or greater (≥17 m/s), 0.30 and 0.46 reach maximum intensity within ±12 h and ±24 h respectively. With regard to TCs of typhoon strength or above (≥ 33 m/s), 0.27 and 0.42 of them peak in intensity within ±12 h and ±24 h. With respect to severe typhoons (≥52m/s), only 0.16 of these storms attain the maximum intensities within both ±12 h and ±24 h respectively. However, these severe typhoons tend to peak in intensity prior to or at the recurvature point (0.84). Of these severe typhoons, few reach their greatest intensity after recurvature.

Table 7-1 The statistics based on the percentages of TCs peaking in relation to the timing of recurvature. The intensity levels are set in consistence with Table 1 of K09 for comparison. V_{max} is the maximum sustained wind.

Intensity Level	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
≥17m/s	0.30	0.46	0.63	0.75	0.37	0.62	0.17
≥33m/s	0.27	0.42	0.60	0.73	0.42	0.67	0.16
≥52m/s	0.16	0.16	0.37	0.63	0.83	0.84	0.11
52m/s≥V_{max}>33m/s	0.29	0.47	0.64	0.74	0.36	0.65	0.17
33m/s>V_{max}≥17m/s	0.41	0.56	0.72	0.82	0.23	0.46	0.21

(1) The percentage of TCs that attained maximum intensity at ±12 h of recurvature.
(2) The percentage of TCs that attained maximum intensity at ±24 h of recurvature.
(3) The percentage of TCs that attained maximum intensity at ±36 h of recurvature.
(4) The percentage of TCs that attained maximum intensity at ±48 h of recurvature.
(5) The percentage of TCs reaching maximum intensity>24 h prior to recurvature.
(6) The percentage of TCs reaching maximum intensity prior to or at recurvature.
(7) The percentage of TCs reaching maximum intensity>24 h after recurvature.

These results are compared to K09 as follows. It is observed in Table 7-1 that tropical storms (33m/s>vmax≥17m/s) are most likely to peak in intensity when close to recurvature. In comparison to 0.32 and 0.57 for all the WNP TCs peaking within ±12 h of

recurvature as reported in K09, these percentages are 0.41 and 0.56 respectively in this study. We also find that strong TCs tend to peak in intensity prior to recurvature (0.84). Among these strong hurricanes (≥52m/s) attaining maximum intensity before recurvature, most of them (0.83) reach their maximum intensity prior to recurvature-24h (i.e., 24 hours previous to recurvature. Others are defined likewise). This is much larger than the corresponding probability (0.46) in K09. In agreement with K09, a fraction of TCs achieve their maximum strength after recurvature among all the intensity levels, especially at recurvature+24h.

The timing of maximum intensity is defined as the first time the TC achieves maximum intensity. If this timing is extended to be any time(s) that the TC attains its greatest intensity, the results are distinct. As shown in Table 7-2, the proportion of TCs that reach the maximum intensity is much higher than that in Table 7-1, within ±12 h, ±24 h, ±36 h and ±48 h. Another appreciable difference is that the proportion of TCs that peak 24 hours after recurvature is much higher than that shown in Table 7-1. However, the ratio of TCs that peak prior to or at TC recurvature resembles those of TCs in Table 7-1.

Table 7-2 The statistics based on the percentages of TCs peaking in relation to the timing of recurvature (as in table 1 but the timing of recurvature is not confined to the first TC peak intensity). The intensity levels are set in consistence with Table 1 of K09.

Intensity Level	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
≥17m/s	46%	64%	77%	85%	44%	63%	52%
≥33m/s	43%	61%	75%	84%	48%	67%	47%
≥52m/s	32%	53%	74%	84%	74%	84%	52%
52m/s≥V_{max}>33m/s	45%	61%	75%	83%	44%	64%	52%
33m/s>V_{max}≥17m/s	53%	65%	78%	88%	30%	45%	68%

*Columns are defined as in Table 7-1.

The relationship between the maximum intensity and recurvature is ambiguous to some degree based on the previous investigations. It is meaningful further to analyze this relationship. The average TC intensities (both MSW and MCP) at different times related to recurvature are derived from the JMA TC best track data between 1985 and 2009 (see

Figure 7-1). We take into account ±48 hours in relation to recurvature. From the figure, it is observed that the average intensities of recurvers tend to peak at or near recurvature in terms of MSW and MCP. According to the average TC intensities, the TC tends to reach its maximum intensity at or near recurvature.

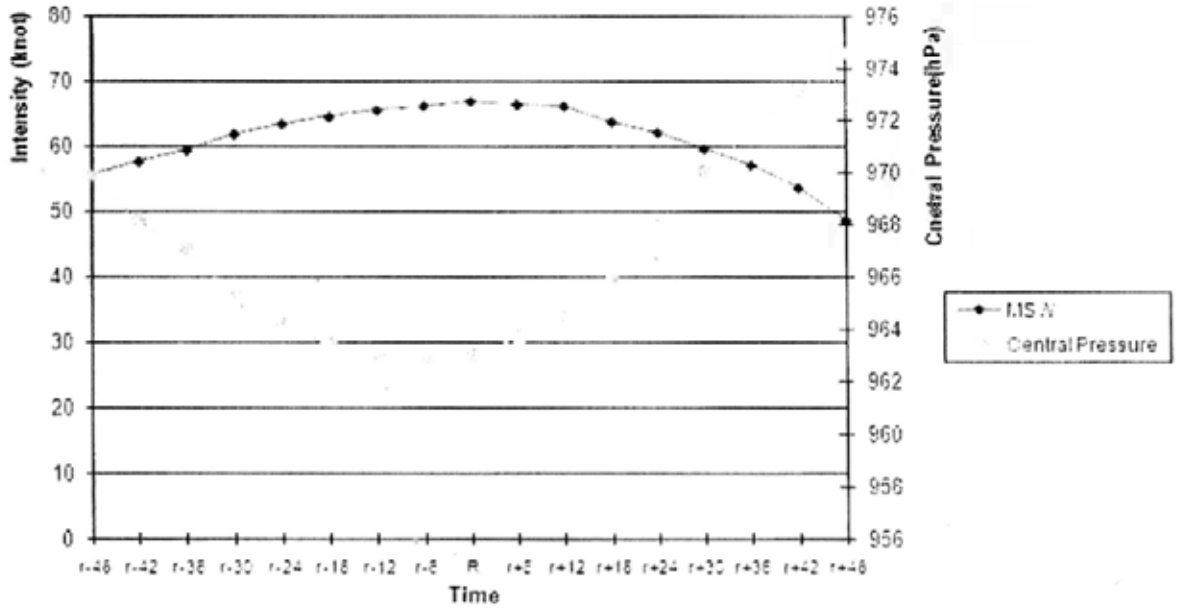


Figure 7-1 The average central pressure and maximum sustained wind of TCs at different times related to recurvature. "R" at the x-axis indicates the time at recurvature whereas "r-6" marks 6 hours prior to recurvature. Other times are defined likewise. The left vertical axis is the maximum sustained wind (unit : Knot) while the right vertical axis indicates the central pressure (unit : hPa).

The Student's *t* test is employed to justify whether the average TC intensity (depicted by MSW (ms^{-1})) changes significantly corresponding to recurvature. The average intensities associated with recurvature are evaluated at seven points: recurvature±12 h, recurvature±24 h, recurvature±36 h, and at recurvature. It is shown in Table 7-3 that the average intensity at recurvature is significantly larger than that at both 24 and 36 hours preceding TC recurvature at a 0.05 significance level. The average intensity at recurvature is also significantly larger than that at recurvature+24h and recurvature+36h. The difference between intensity at recurvature and at recurvature±12h is not remarkable. However, the difference is significantly remarkable between intensity at recurvature and intensity at recurvature-24h, and between intensity at recurvature and recurvature+24h. Therefore, Table 7-3, together with Figure 7-1, indicates that TC tends significantly to

peak in intensity at or near recurvature, at least within ±24 hours in WNP. This result is consistent with R72 and EM98: that a TC tends to attain its maximum intensity close to recurvature.

Table 7-3 The Student's *t* test results for differences in intensities related to TC recurvature from 1985 to 2009. Rec marks the average intensity at recurvature. Pre36H and Aft36H indicate the average intensity at 36 hours previous to and after recurvature respectively. Others are defined likewise.

Intensity change	mean	t-value	d.f.	Sig (2-tailed)
Rec-Pre36H	7.844	4.432	166	0.000
Rec-Pre24H	3.383	2.361	166	0.019
Rec-Pre12H	1.437	1.569	166	0.119
Rec-Aft12H	1.018	1.910	166	0.059
Rec-Aft24H	4.611	4.752	166	0.009
Rec-Aft36H	10.359	7.108	166	0.000
Pre36H-Aft36H	2.515	0.959	166	0.339
Pre24H-Aft24H	1.228	0.608	166	0.544
Pre12H-Aft12H	-0.419	-0.337	166	0.737

Since the TRMM SST dataset is only available from 1998 to 2009, the TC samples during this time period are analyzed for comparison in order to be consistent with the available TRMM SST dataset. The analysis of the TC samples from 1998 to 2009 indicates similar but not identical results (Table 7-4). It shows that the average intensity at recurvature is larger than that at recurvature-36h, recurvature+24h and recurvature+36h. Different from Table 7-3, the difference between at recurvature and recurvature-24h is, however, not statistically significant from the table. MPI, an upper bound on TC intensity for given oceanic and atmospheric conditions (e.g., Emanuel, 1986, 1988, 1991), is used to analyze the potential for intensification. The empirical MPI function (DeMaria and Kaplan, 1994; Knaff et al., 2005) is fit by historical records of SST and TC intensity and shown as:

$$\mathrm{MPI} = A + Be^{C(T-T_0)} \qquad (7.1)$$

where A=38.21 knot, B=170.72 knot, C= 0.1909 °C^{-1}, and T_0=30 °C. T remarks the SST at the TC center. The potential for intensification is evaluated using MPI minus current intensity. As shown in Table 7-5, the intensification potential at recurvature is larger than that at recurvature-36h at a significance level of 0.096. This means that the intensifying trend at recurvature-36h is more evident than that at recurvature. Besides, the intensification potential is significantly larger than those at recurvature+12h, recurvature+24h and recurvature+36h. It can also be observed that the potential TCs intensifications previous to recurvature are much higher than those after recurvature (see Pair 7, 8 and 9 of Table 7-5). The intensification potential based on MPI is in full line with the results in Table 7-3. Figure 7-2 shows that the average SST around the TC center tends to decrease as TC moves from previous to recurvature to after recurvature, since TC tends to move into the mid-latitude region with a lower surface temperature after recurvature.

Table 7-4 The Student's *t* test results for differences in intensities related to TC recurvature from 1998 to 2009. Rec marks the average intensity at recurvature. Pre36H and Aft36H indicate the average intensity at 36 hours previous to and after recurvature respectively. Others are defined likewise.

Intensity change	**mean**	**t-value**	**d.f.**	**Sig (2-tailed)**
Rec-Pre36H	6.667	2.564	62	0.013
Rec-Pre24H	2.143	1.189	62	0.239
Rec-Pre12H	-0.079	-0.070	62	0.945
Rec-Aft12H	0.794	0.904	62	0.369
Rec-Aft24H	3.730	3.218	62	0.002
Rec-Aft36H	8.413	4.569	62	0.000
Pre36H-Aft36H	1.746	0.467	62	0.642
Pre24H-Aft24H	1.587	0.617	62	0.539
Pre12H-Aft12H	0.873	0.489	62	0.626

Table 7-5 The Student's *t* test table for intensifying potential (MPI minus current intensity) at different times related to TC recurvature between 1998 and 2009.

Intensity change	Mean	t-value	d.f.	Sig (2-tailed)
Rec-Pre36H	-9.98	-1.694	58	0.096
Rec-Pre24H	-3.39	-0.673	58	0.504
Rec-Pre12H	-6.29	-0.987	59	0.328
Rec-Aft12H	11.34	2.234	54	0.030
Rec-Aft24H	12.59	2.734	59	0.008
Rec-Aft36H	15.73	2.524	53	0.015
Pre36H-Aft36H	27.67	4.461	53	0.000
Pre24H-Aft24H	16.56	3.008	58	0.004
Pre12H-Aft12H	15.60	2.573	54	0.008

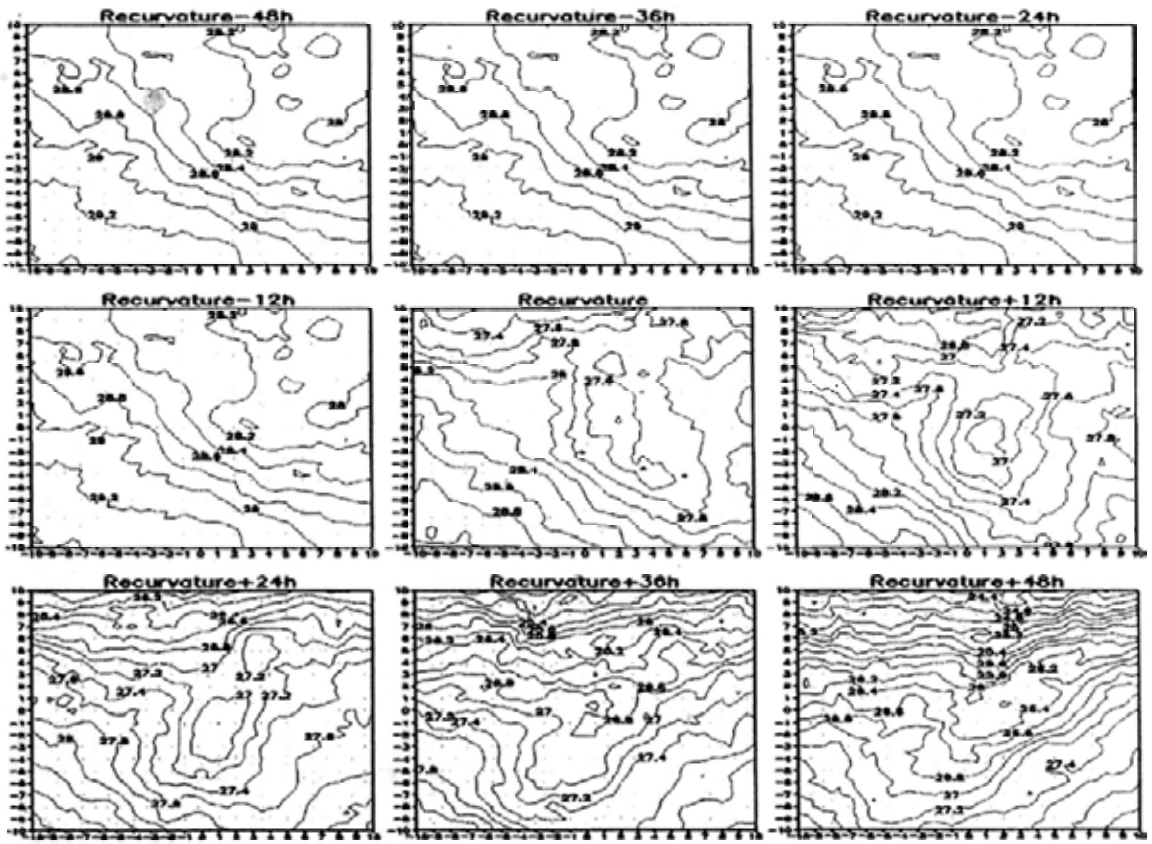


Figure 7-2 The average SST of a rectangular region (20°×20°) centering at TC center

7.3.2 Points of TC Recurvature

Recurving TCs may peak in intensity prior to or after recurvature. The recurving points are shown in Figure 7-3 for TCs between 1984 and 2009. Green rectangular and red circular points indicate that the recurvature points of TCs attaining their maximum intensity after and prior to recurvature respectively. The characteristics of recurving points are derived from the ANOVA table (Table 7-6) for the three attributes: longitude, latitude and MCP. It can be observed that the latitude is significantly different between the two groups (i.e., peaking prior to or after recurvature). The longitude is different between the two groups at a significance level of 0.063. The central pressures of the two groups significantly differ from each other. Therefore, the latitude, longitude and MCP of the two groups of recurving TCs are significantly different. The recurving points' peaking intensity after recurvature tends to be located to the southeast of that of the TCs' peaking intensity prior to recurvature, since the relatively southern recurving points provide more space for intensification due to the higher SST, strong latent hear flux and smaller vertical wind shear (e.g., DeMaria et al., 2005; Emanuel et al., 2004; Knaff et al., 2005). As a TC moves further north, it in general faces a lower SST, oceanic heat content and mid-level cold moisture, as well as a rising vertical wind shear. These conditions are associated with TC weakening.

Table 7-6 ANOVA table for the recurving points of TCs attaining maximum intensity prior to or after recurvature

Variable		Sum of Squares	d.f.	Mean Square	F	Sig.
Latitude	Between Groups	199.399	1	199.399	6.593	0.011
	Within Groups	5111.381	169	30.245		
	Total	5310.780	170			
Longitude	Between Groups	682.060	1	682.060	3.494	0.063
	Within Groups	32990.941	169	195.213		
	Total	33673.001	170			
Central Pressure	Between Groups	4472.492	1	4472.492	7.614	.006
	Within Groups	99272.503	169	587.411		
	Total	103744.994	170			

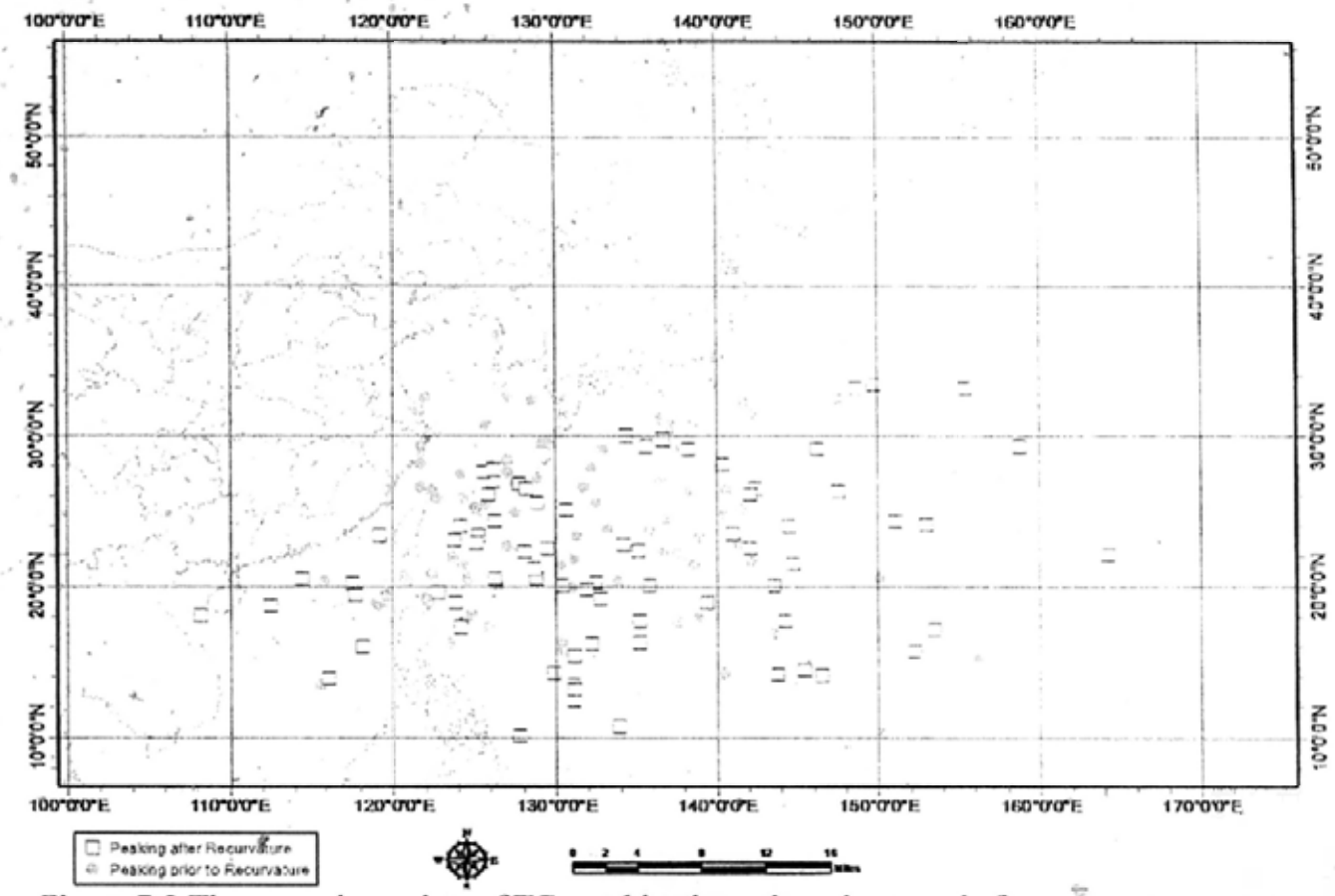


Figure 7-3 The recurving points of TCs peaking intensity prior to and after recurvature. The green rectangular points indicate the recurvature points of those TCs attaining their maximum intensity after recurvature whereas the red circular points are the recurvature points of those TCs attaining their maximum intensity prior to recurvature.

7.3.3 The Sustaining of Maximum Intensity

After TCs attain their maximum intensity, they sustain the same intensity for some time. Little attention has been paid to the sustaining of TCs' maximum intensity with regard to recurvers and straight movers (Figure 7-4). It is noteworthy that the timing of attaining the maximum intensity agrees with the previous definition as being the first time that a TC peaks in intensity. It is of great interest to evaluate the time period for which recurvers or straight movers sustain their maximum intensity after first attaining it. This investigation is accomplished by ANOVA and the Student's *t* test, to see whether the means of the two groups are statistically different.

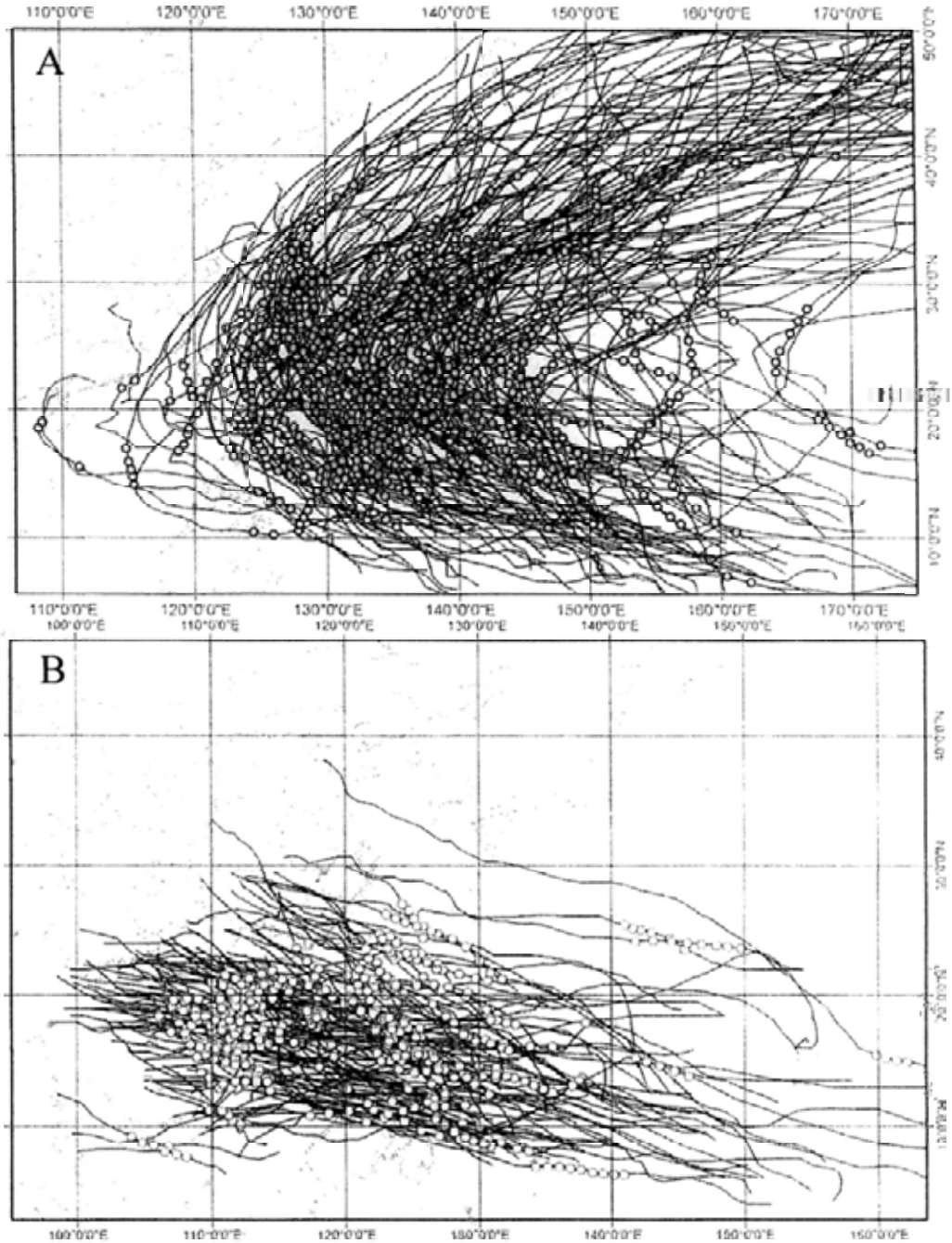


Figure 7-4 The TC tracks and points to their lifetime maximum intensity after first peaking in intensity of (A) recurvers and (B) straight movers from 1985 to 2009

The ANOVA table is shown in Table 7-7. From the table, the maximum intensity and its duration of straight movers and recurvers are different at a significance level of 0.00. Besides, the Student's *t* test indicates that the average duration of the maximum intensity of recurvers is significantly larger than that of straight movers. The average MCP of recurvers during peak time is significantly larger than that of straight movers. This infers

that the intensity of straight movers tends to be larger than that of recurvers during peak time. The difference is largely attributed to the SST. The straight movers peak in intensity and sustain their maximum intensity in relatively low-latitude regions. However, recurvers peak in intensity and sustain their maximum intensity in relatively higher latitude regions, even in the mid-latitude area. The lower SST and higher vertical wind shear associated with the higher latitude ocean lead to the weaker intensity of recurvers. The relatively short duration of peaking intensity of straight moving TCs can be ascribed to the landfall, which causes the water vapor supply and oceanic latent heat to be cut off. Most straight movers experience landfall during their lifespan. Once the TC or its outer rim interacts with the coastal areas about one day prior to the landfall, the TCs begin to weaken. Nonetheless, in spite of the lower SST and sometimes higher vertical wind shear, recurvers can sustain their maximum intensity for longer due to the continuous ocean heat content and water vapor supply.

Table 7-7 The results of (a) the Student's *t* test and (b) ANOVA for maximum intensity and the duration of peak intensity for recurvers and straight movers

(a)

Variables	**t-value**	**d.f.**	**Sig (2-tailed)**
Duration of Maximum Intensity	4.480	287	0.000
Maximum Intensity	4.533	287	0.000

(b)

Variables	**F-value**	**d.f.**	**Sig (2-tailed)**
Duration of Maximum Intensity	20.070	287	0.000
Maximum Intensity	20.500	287	0.000

7.4 Landfall and Intensity

7.4.1 Decay related to landfall

The ±48 hours related to TC landfall are divided into intervals of 12 and 24 hours to form 7 and 8 pairs. TC landfalls coincide with the weakening of intensity due to the cutting off of the water vapor supply, structural change, land-TC interaction and rainstorms induced by TCs, from the previous observational analysis (e.g., Bhowmik *et al.*, 2005a; Fujibe and Kitabatake, 2007; Wong *et al.*, 2008). The time at which the TCs begin to weaken associated with TC recurvature is investigated using the Student's *t* test at intervals of 12 and 24 hours respectively (Table 7-8).

Table 7-8 The Student's *t* test results for the difference between the two groups' average intensity of TCs at intervals of (a) 12 hours and (b) 24 hours related to TC landfall. "Pre48H" indicates the average intensity at 48 hours previous to landfall. The others are defined likewise.

(a)

intensity change	mean	t-value	d.f.	Sig (2-tailed)
Pre48H-Pre36H	-1.607	-1.235	60	0.221
Pre36-Pre24H	0.016	0.017	60	0.996
Pre24H-Pre12H	-2.361	-2.079	60	**0.042**
Pre12H - Landfall	-3.311	-3.366	60	**0.001**
Landfall–Aft12H	-8.545	-3.570	54	**0.001**
Aft12H – Aft24H	-4.902	-4.834	50	**0.000**
Aft24H - Aft36H	-1.182	-0.559	43	0.579
Aft36H – Aft48H	-0.947	-0.349	37	0.729

(b)

intensity change	mean	t-value	d.f.	Sig (2-tailed)
Pre48H-Pre24H	-1.590	-0.842	60	0.403
Pre24H -Landfall	-5.672	-3.036	60	**0.004**
Landfall – Aft24H	-14.117	-5.486	50	**0.000**
Aft24H – Aft48H	-2.421	-0.977	37	0.335

Pre36H –Pre12H	-2.344	-1.305	60	0.197
Pre12H – Aft12H	-11.618	-4.288	54	**0.000**
Aft12H - Aft36H	-7.000	-3.297	43	**0.002**

From the table, with a time interval of 12 hours (Table 7-8a), the average intensity tends to be stable from 48 hours to 24 hours preceding TC landfall and the intensity difference begins to be remarkable at 24 hours prior to landfall at a significance level of 0.042. The decay continues until 24 hours after landfall, at which the intensity difference after 12 hours is not remarkable. With respect to the table with a time interval of 24 hours (Table 7-8b), the average intensity difference between landfall-24h and at landfall is significant. Besides, the average intensity difference between Landfall and LandfallAft24 is also significant. Similar differences cannot be observed between LandfallPre48 and LandfallPre24, between LandfallPre36 and LandfallPre12, and between LandfallAft24 and LandfallAft48. Therefore, the analyses between the intensities of landfalling TCs are not sensitive to the time intervals (e.g., 12 hours, 24 hours) and TCs experience significant decay only during ±24h related to TC landfall. After landfall for 24 hours, the TCs mostly decay to tropical depression.

7.4.2 Interpretation

TC intensity is associated with a wide range of parameters, such as SST, which is related to MPI (e.g., Emanuel, 1986, 1988, 1997; Emanuel et al., 2004; Malkus and Riehl, 1960), the upper bound on TC intensity, vertical wind shear, high-level relative humidity, and surface flux (DeMaria et al., 2005). Landfalling TCs interact with land surface prior to and after landfall. There exists a long time period without SST observation after TCs make landfall. Therefore, the SST is not employed for interpretation. LHR through condensation and precipitation processes, particularly in the TC's inner core (a radius of 110 km from the TC's center), is crucial to TC development and maintenance (Charney and Eliassen, 1964; Kuo, 1965). The rainfall as a proxy of condensational heating is employed to evaluate TC strength by case study (e.g., Heymsfield et al., 2001; Jiang and Halverson, 2008; Rodgers et al., 2000; Simpson et al., 1998). The relationship between

landfall and intensity are interpreted from the perspective of latent heat flux, precipitation rate, vertical wind shear, and relative humidity.

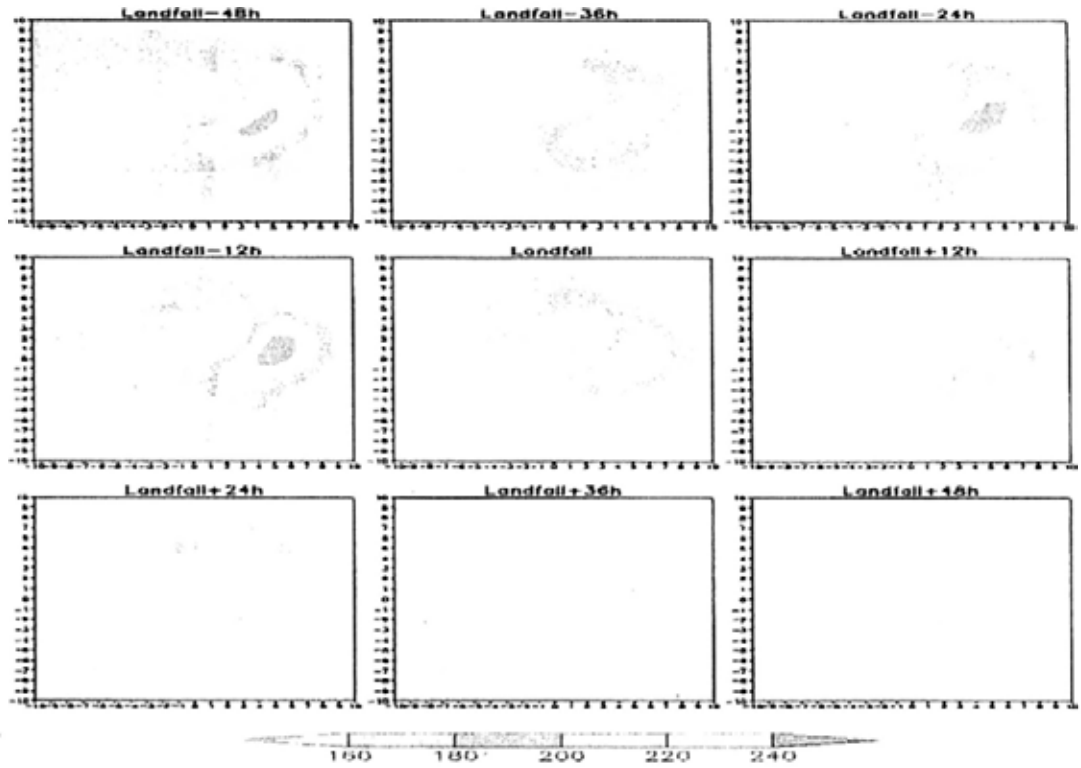


Figure 7-5 The latent heat flux ($w \cdot m^{-2}$) related to TC landfall. The coordinate (0, 0) indicates the TC's center. The rectangular region (20° × 20°) is centering at the relative TC center. Landfall±48h means 48 hours previous to and after TC landfall respectively. Others are defined likewise.

The composite SLHF is shown in Figure 7-5, which includes the composited SLHF at 9 time periods related to the landfall (e.g., landfall-48h, landfall-36h). The composite SLHF tends to decrease from landfall-24h to landfall+24h. However, the SLHF is nearly stable after landfall+24h. The SLHF remains stable prior to landfall-24h. Therefore, the intensity change associated with TC landfall is partly attributed to SLHF. The composite analysis of precipitation rate by TRMM is displayed in Figure 7-6. The precipitation rate in relation to landfall tends to drop after landfall-24h. However, the difference between landfall-24h and landfall-12h is not appreciable. The difference between landfall-12h and the times after that are significant from the observational analysis. The precipitation has been used as a proxy to measure intensity. Therefore, the characteristics of TC intensity change related to landfall can be attributed to the precipitation rate.

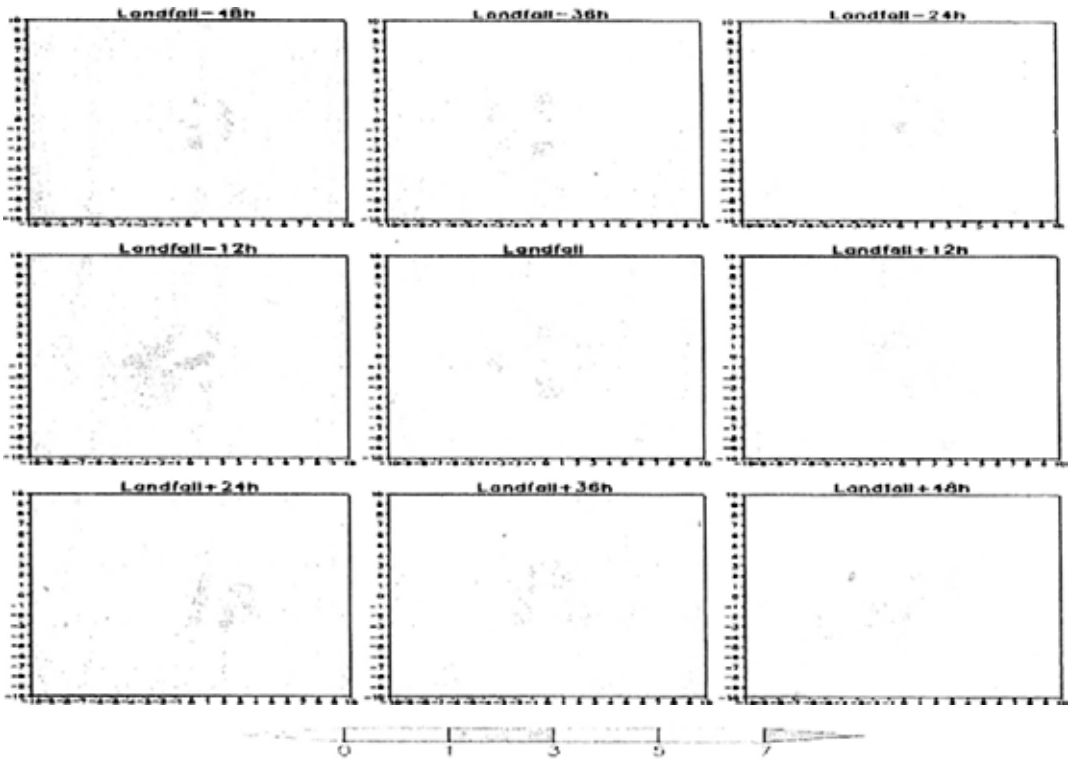


Figure 7-6 The precipitation rate ($mm \cdot hr^{-1}$) related to TC landfall. The coordinate (0, 0) indicates the TC's center. The rectangular region (20° × 20°) is centering at the relative TC center. Landfall±48h means 48 hours previous to and after TC landfall respectively. Others are defined likewise.

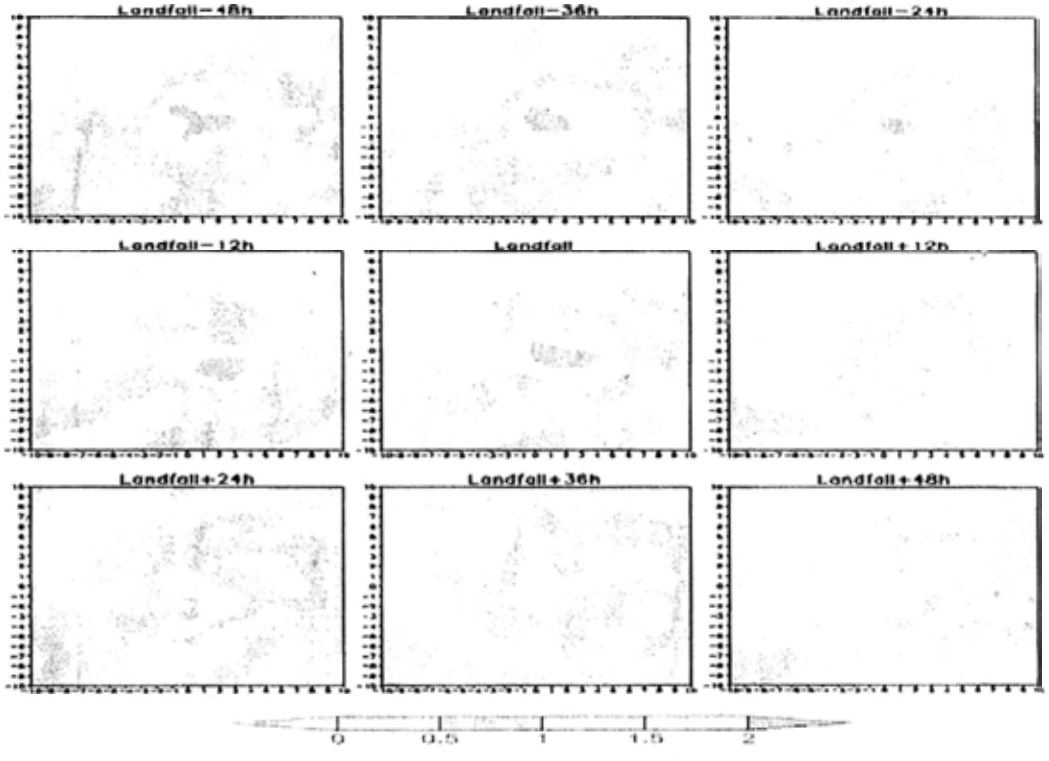


Figure 7-7 The 200 hPa divergence ($\times 10^{-5}$) related to TC landfall. The coordinate (0, 0) indicates the TC's center. The rectangular region (20° × 20°) is centering at the relative TC center. Landfall±48h means 48 hours previous to and after TC landfall respectively. Others are defined likewise.

The high-level divergence (e.g., 200 hPa level) is an important indicator of TC intensity, which is associated with TC structure and development (DeMaria et al., 2005). The composite 200 hPa divergences at time periods prior to landfall differ insignificantly. However, the composite 200 hPa divergence tends to decrease greatly after landfall (Figure 7-7). The TC intensity therefore drops after landfall.

The water vapor supply plays an important role in TC sustainment. The composite 850 hPa water vapor and wind fields are overlaid for investigation. As shown in Figure 7-8, the 850 hPa wind fields around the TC center are mostly northeastward at all times related to landfall. However, the water vapor supply weakens after landfall. After landfall+36 h, the water vapor supply is cut off. The cut-off water vapor supply largely explains the death of post-landfall TCs. The weakening water vapor flux leads to the lower TC intensity. Therefore, the TC weakens after landfall.

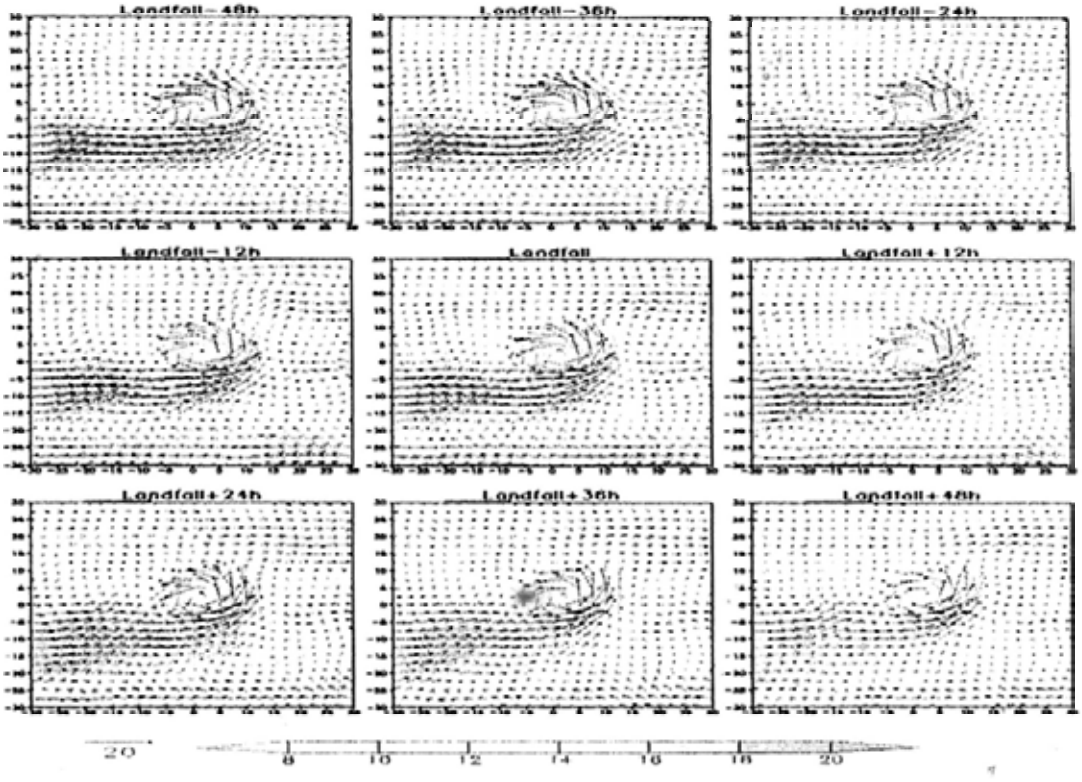


Figure 7-8 The 850 hPa water vapor supply ($g \cdot s^{-1} \cdot hPa^{-1} \cdot cm^{-1}$) and wind fields ($m \cdot s^{-1}$) related to TC landfall. The coordinate (0, 0) indicates the TC's center. The rectangular region (20° × 20°) is centering at the relative TC center. Landfall±48h means 48 hours previous to and after TC landfall respectively. Others are defined likewise.

In summary, the relationship between TC landfall and intensity is interpreted from the perspective of latent heat flux, precipitation rate, 200 hPa divergence, and 850 hPa wind fields and water vapor supply. These composite indicators indicate that the TCs tend to weaken significantly during the period from landfall-24h to landfall+24h.

7.5 Summary

The relationship between TC tracks and intensity is investigated in this chapter. It is found that, at least in WNP (including the SCS), TCs tend to peak in intensity close to the point of recurvature. We also found that the recurving points of TCs peaking their intensity after recurvature are located to the southeast of the recurving points of TCs peaking intensity prior to recurvature. The recurving TCs sustain their maximum intensity for significantly longer than straight moving TCs. The TCs also tend to weaken significantly from landfall–24h to landfall+24h. It is consistent with the previous study that the TCs begin to interact with the coastal areas one day prior to landfall. Meanwhile, the TCs decay to tropical depression or dissipate at landfall+24h. Therefore, the TCs will not weaken significantly after landfall+24h. The results derived from the characteristics of TC intensity change related to TC movement should be of both scientific and operational significance.

CHAPTER 8: A Novel Web-based System for Tropical Cyclone Analysis and Prediction

8.1 Introduction

It is of considerable importance to correctly predict the TC track (Aberson and Sampson, 2003; Chan, 2005a; Harr and Elsberry, 1991) and intensity (Velden and Leslie, 1991; Wang and Wu, 2004; Wong and Chan, 2004) for effective disaster analysis, response, mitigation and management (Simpson and Riehl, 1981a). Landfalls (e.g., Lyons, 2004; Wu *et al.*, 2003) and recurvatures (e.g., Chan et al., 1980; Cheung, 2006; Hodanish and Gray, 1993; Krishnamurti et al., 1992) of TC are particularly critical problems since most damages are caused during TC landfall and recurvature.

Observational research, dynamic modeling and data mining have been key approaches to TC prediction and analysis in terms of intensity, track, landfall and recurvature. A number of systems have also been implemented for TC forecasting. MacAfee (MacAfee, 1997) developed the Canadian Hurricane Center Forecaster's Workstation to forecast and analyze hurricane tracks by the analysis of larger-scale tropical weather patterns and global dynamic hurricane modeling. The Automated Tropical Cyclone Forecasting System (ATCF) developed by NRL Monterey (Sampson and Schrader, 2000) is a computer-based platform designed to automate and optimize the forecasting process at the U.S. Department of Defence and tropical cyclone warning centers of the National Weather Service. It provides capabilities to track and forecast movement, construct messages, and disseminate warnings. Version 3.2 is now widely used in tropical cyclone warning centers and enables analysts to track and forecast storms in the USA. The Australian Tropical Cyclone Workstation (ATCW) (Woodcock, 1995), developed by the Bureau of Meteorology in Australia, is in operations at the Western Australia regional office in Perth and in New Caledonia (Woodcock, 1995). ATCW used CLImatology and PERsistence (CLIPER, Neumann and Hope, 1972) as a forecasting scheme for current TC tracks. In brief, these TC forecasting systems are designed for professional use at various meteorological agencies and are not made available to the public. Furthermore,

system developments are only based on stand-alone edition workstations. These systems, however, do not provide case-by-case visualization and query of historical TCs and do not have facilities for web-based services.

With regard to software development, quite a few TC tracking software have been made available in recent years. There are, for example, 'Eye of the Storm', 'Global Tracks', 'Hurricane Watch 2000', 'HurrTrak' , 'Merlin', 'Personal Hurricane Center',' Storm', 'Tempest Hunter' and 'Tracking the Eye' on the webpage of National Oceanic and Atmospheric Administration (NOAA). However, these systems are mainly developed for the visualization of historical TC archives. Only a few are developed for simple TC tracking, TC analysis, or the display of predictions issued by authoritative meteorological organizations. The systems at JTWC, JMA, National Hurricane Center (NHC) and HKO are typical examples. These TC tracking software are developed just for the visualization of the historical TC database via the web, but they lack the capabilities for TC analysis and prediction.

As for building a GIS for prediction, analysis, tracking and visualization of TC, the fundamental spatial database and API are relatively expensive to access, particularly for the general public. There is a number of popular GIS software, such as ArcGIS server, ArcIMS, MapExtreme, MapX, which provide useful interfaces for system development. For example, a desktop-version of a GIS-based TC system is built for the simple analysis of historical and active TCs (Kong et al., 2008). A GIS-and-remote- sensing-based system was developed to monitor the TC movement in 1998 (Kumar et al., 1998). However, it is considerably more difficult to develop a web-based system for rigorous TC analysis with functionalities for complicated visualization, analysis and data mining. Very few TC forecasting systems are built by web-based GIS software. Existing web-based GIS systems for TCs are, however, mainly built for information dissemination and map distribution via the Internet, with only a few having limited data analysis capability. As to the accessibility of GIS, by virtue of its technical complexity and high cost, GIS has usually been made accessible only to the government and public administration, where it plays a vital role in the vast majority of its everyday operations (Masser et al., 1996).

Data mining is the process of extracting hidden and useful patterns and information from data (Fayyad et al., 1996; Han and Kamber, 2006). In recent years, tremendous data mining algorithms have been employed to unravel TC tracks and intensities in TC database (e.g., Camargo *et al.*, 2004; Camargo *et al.*, 2008; Camargo *et al.*, 2007a, b; Gaffney and Smyth, 1999; Gaffney *et al.*, 2007; Gaffney, 2004; Lee *et al.*, 2007). However, only a few data mining derived results have been successfully applied to predict TC movement. On the other hand, dynamic models have been employed to predict TC movements over the years (Juneng et al., 2007; Mandal et al., 2007).The two approaches can actually complement each other to achieve more reliable and accurate prediction of TC movements. Data mining can unravel useful information for the enrichment of the dynamic models, and the mechanisms underpinning the dynamic models can be used to guide the data mining process.

To recapitulate, existing TC systems can be grouped into several categories: (1) stand-alone systems for professional use only, with limited capability for TC visualization and query; (2) web-based systems for visualization, with limited capability for TC analysis and prediction; (3) GIS-based system for simple TC analysis; (4) web-based system for information dissemination, with limited data analysis capability. Therefore, a web-based and free-access GIS system should be built to bring GIS technology and TC knowledge to the researchers and general public at little or no cost. Thus, it should have functionalities for the visualization of historical TC archives, complex data analysis and information distribution for research and professional use, but also for the consumption of the general public. A comprehensive TC analysis and prediction system with all the above features as well as powerful data mining and dynamic modeling capabilities has yet to be developed. This chapter introduces our (a group effort in collaboration with the Department of Computer Science and Engineering at The Chinese University of Hong Kong) construction of such a platform on which data mining methods such as clustering, CART and case-based algorithm and dynamical models such as MM5 can also be used individually or in synchrony to accomplish more accurate TC prediction (Leung et al., 2011).

The overall objective is to develop a powerful web-based system for the tracking, prediction and visualization of TC movements, landfalls, and recurvatures via dynamic modeling and data mining in assimilated multi-source, multi-scale and multi-level TC database. It is to be realized through the accomplishments of the following interrelated sub-goals with innovative design concepts. Firstly, a web-based platform with user-friendly interface and effective visualization capability is built with advanced GIS technology and scientific speed-up techniques. The present system uses Google map products as a basis for development since it is web-based, and is open and free in terms of geographic data management and API. Secondly, the system can run on any computers and is transportable. Dynamic models based on physical processes in meteorology are utilized to generate data for tracking typhoons, particularly on the dynamic changes of landfalls and recurvatures. Thirdly, novel data mining algorithms are developed to unravel rules and regularities for TC landfalls and recurvatures. The data mining algorithms and the dynamic model will mutually enrich each other in a complementary and integrated manner. Fourthly, a multi-source, multi-scale and multi-level remote sensing database consisting of atmospheric and meteorological data such as wind fields in different levels, humidity, sea-level temperature and pressure, geopotential height, rainfall, wind shear, best track data ,etc., is constructed for effective and efficient typhoon track analysis and forecasting. Lastly, using state-of-the-art software tools such as Adobe Flash and Sun's Java in conjunction with Google Maps API, the system is developed for public consumption in the internet.

The significance of this system lies in its efficient and effective visualization of historical TC archives and the retrieval of data for analysis and forecasting in the web-based environment. Its user-friendly interface and efficient query system greatly facilitate public consumption and professional use. Its facilities for the employment of dynamic models and data mining algorithms make it a powerful system for TC analysis and prediction, particularly on TC landfalls and recurvatures. Furthermore, the TC database is extensible for future applications.

In what follows, the overall architecture of this system is first discussed in section 8.2. System functionality and implementation are scrutinized in section 8.3. Prediction of TCs with data mining algorithms and dynamic model is then elaborated in section 8.4. This chapter is concluded with a summary and outlook for further research.

8.2 Overall architecture of the system

8.2.1 Overall system architecture

The present system uses Google map as a basis for development since it is web-based, flexible, and efficient and is open and free in terms of geographic data management and API. The tropical cyclone analysis and prediction system (TCAPS) is constructed on the basis of the Client/Server architecture. The overall system architecture is depicted in Figure 8-1.

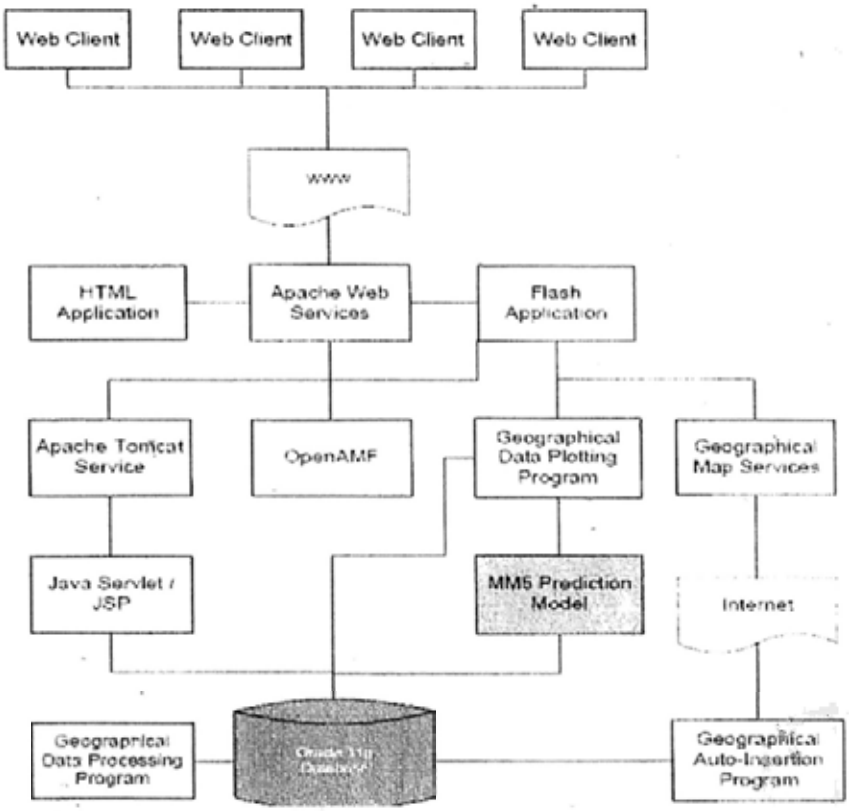


Figure 8-1 Overall architecture of the tropical cyclone analysis and prediction system (TCAPS)

When clients access the system through the World Wide Web (WWW), a request will be submitted to the server for onward submission to the Apache Web Services. The Apache Web Services will then transfer these messages to relevant services or applications. Apache Web Services act as the web gateway to be shown on the WWW.

Specifically, Apache Tomcat Services provide a gateway for different applications to transfer interactions and communications to JSP application or Java Servlet. The Apache Web Services, on the other hand, provide a web gateway for various services or applications. Flash Applications render dynamic graphical user interface for the visualization of geographical and meteorological data and connection to Google Map Services, Open Action Message Format (OpenAMF) and Geographical Data Plotting Program, as well as display of TC best track on Google Map. The Geographical Data Auto-Insertion Program can be employed to extract the downloaded TC-related meteorological data, transform the geographical data to specified format, and insert the meteorological data into database. The Meteorological Data Plotting Program transforms the meteorological data as images in JPEG formats used as images on super-imposed layers of the TC Track Viewer. Google Map Services connect the background world map data from the Internet to other applications. The HTML Applications supply static WebPages through Apache Web Services for user navigation and are used mostly for organization purpose, i.e. linking of different applications and WebPages. Meanwhile, Java Servlet / JSP Applications offer database query services and Java Web Applications through the Apache Tomcat Services. The OpenAMF manages Java-to-Flash Remote Services and provides linkages for Flash Action Script and Java Class aiming at retrieving data from the database. The Oracle Database stores geographical data and TC-related data in a relational database. In order to extract the MM5 modeling results, the Result Extraction Program can extract results obtained by the MM5 model.

8.2.2 General Functionality of the TC track viewer

The TC Track Viewer of TCAPS provides the map view, terrain view and satellite view through Google Map API (Figure 8-2). The link to the system is *http://pc89075.cse.cuhk.edu.hk:8080/myapp2/NewFlashGIS.html*. TC analysis and

prediction are hidden by default in order to provide more space for analysis and decrease the loading time of the TC track viewer.

Figure 8-2 Three basic views of the TC Track Viewer

The TC track viewer is developed by Flash Action Script 3.0, and cannot be directly connected to the Oracle database. Therefore, Java programming is used to connect the TC track viewer to the database. Java Servlet, OpenAMF and Java programs form a linkage between the database and the TC track viewer. The viewer is built on Flash and is embedded into the HTML. It provides not only a high level graphical user interface for the visualization of TC and its related attributes, but also contains different kinds of multi-criteria spatial queries in terms of TC track, recurvature and landfall. Basic functionalities of the TC Track Viewer are summarized as follows:

1. Animation of TC tracks

This tool makes a TC recur in 6-hour increments. The animation starts at a predefined time and from there users can play, pause, go backward or forward to view the TC tracks. The viewer can be set at any year starting from 1951.

2. Layered display of features

To facilitate visualization, atmospheric features can be shown in layers. SLP, SST, cloud formation, humidity, wind field, wind shear on land surface and high altitude can be individually displayed as image layers, or can be superimposed onto one another. It runs

in parallel with the animation of TC tracks to give the atmospheric and oceanic conditions throughout the TC movement.

3. Data analysis

The viewer provides facilities for users to incorporate different attributes into the analysis, and results can be displayed on the line chart which is movable and with changeable opacity. It depicts the correlation between selected attributes.

4. Information display via cursor

TC information of each point along a track can be retrieved and displayed by clicking the mouse cursor at the point. It can also show its atmospheric and oceanic background information. Such information can be hidden by moving the cursor away from it.

5. Data search

For easy manipulation, the viewer can show or hide a wide range of data searching panels such as TC classification. Users can set the visibility of these panels.

6. Result Selector Panel and Result List

Results obtained from the query are shown on the Result Selector Panel and user can select "proceed" or "show details" on the panel. After clicking "show details", the Result List will be shown automatically, giving brief descriptions of the results. User can further select the results on the Result List.

7. TC prediction

Data mining algorithms are put in store for TC track prediction via FMM, spatial similarity and temporal similarity, and dynamic analysis by MM5.

8. TC genesis analysis

Since TCs originated at different locations in the Pacific Ocean and SCS will have different characteristics in terms of track, intensity and landfall characteristics, analysis of the initial condition, called genesis analysis, at the point where TC gains its intensity to become tropical storm is important to the prediction of its subsequent development. This analysis focuses on the number of TCs formed in a specific time interval at a latitude-longitude grid cell. Each latitude-longitude grid cell is colored according to the frequency of TC occurrences at that location over the years.

8.3 System Functionality and Implementation

8.3.1 Building the TC Database

In TCAPS, the TC archives are organized in terms of TC tracks, meteorological variables and satellite cloud images. TC tracks display TC points constituting the polyline in 6-hour interval. The track data are downloaded from JMA. Meteorological variables (e.g., wind field, wind shear, SLP, SST, rainfall and humidity) are important factors affecting TC track, intensity, genesis, landfall and TC triggered disasters (e.g., flood, land slide, storm surge). Therefore, these variables are collected from authoritative meteorological agencies and can be superimposed in TCAPS. Satellite cloud image can display not only the TC structure; it can also indicate TC intensity and motion. In the present system, the study area mainly covers WNP and SCS, that is 0°N – 45°N in latitude and 100°E -180°E in longitude. Therefore, the meteorological variables are extracted within the specified study area. A description of the TC database is given in Table 8-1.

Table 8-1 Description of the TC database

Data	Source	Time Resolution	Spatial Resolution (lat x long)	Image range[1]	Raw data range[2]
TC track	JMA	Every 6 hours	N/A	1951 to 2009	1951 to 2009
Wind at sea level of 10m	NCEP/NCAR	Every 6 hours	2.5° x 2.5°	2003 to 2007	2003 to 2007
Wind shear	NCEP/NCAR	Every 6 hours	2.5° x 2.5°	2000 to 2007	2003 to 2007
SLP	NOAA	Every 6 hours	2° x 2°	2003 to 2007	none
SST	NASA/NOAA	Daily	0.1° x 0.1°	2003 to 2007	2003 to 2007
Rainfall	NASA	Every 3 hours	0.25° x 0.25°	2006 to 2007	2006 to 2007
Satellite cloud images	NASA	Every 3 hours	1/6° x 1/6°	About 10 days	N/A
Humidity	NCEP/NCAR	Every 6 hours	2.5° x 2.5°	2003 to 2008	2003 to 2008

[1] "Image range" stands for the temporal data range for the image layers superimposed in the system

[2] "Raw data range" stands for the temporal data range for the raw data stored in the background database.

The comprehensive TC database in TCAPS integrates TC best track database, and other meteorological variables (such as SST, SLP, humidity, wind fields in different atmospheric levels) acquired from NOAA, NASA, NCEP/NCAR and JMA. A tool kit is constructed to facilitate data retrieval.

The TC track data are derived from the JMA best track with 6-hour interval. It is organized by TC number, intensity, date, central pressure, latitude and longitude. The SST data and images are downloaded from the remote sensing systems (RSS) under the sponsorship of NOAA and NASA. The data have a spatial resolution of 0.25° by 0.25° with a daily interval for the years starting from 1951. SLP data and images have a 6-hour interval. The wind at sea level of 10m, wind shear and humidity are collected from NCEP/NCAR reanalysis (Kalnay et al., 1996). SLP is downloaded from the website of NOAA with a 2.5° x 2.5° rectangular grid resolution in latitude. Satellite cloud images in a 3-hour interval are obtained from NASA. Due to the large storage requirement, about 10 days of satellite cloud images are temporarily collected for preliminary analysis. The SQL loader, Ferret and Grid Analysis and Display System (GrADS) are employed to transform the original raw data into the format suitable for TCAPS operations.

8.3.2 System Requirements

The construction of the TCAPS platform involves Windows XP workstation and Linux Ubuntu workstation. The Windows XP workstation is set up as web server, while the Linux Ubuntu workstation is set up to run the MM5 for TC prediction (Table 8-2).

Table 8-2 The operating systems and functions of the TCAPS platform

Operation System	Windows XP	Linux (Ubuntu)
Installed Application(s)	Apache HTTP Server 2.0.48 Tomcat 6.0 Oracle 11g Release 2	MM5 Prediction Model
Functions	Storing TC data Hosting the system (GIS)	TC Track Prediction

Besides the software listed in Table 8-2, we also use other state-of-the-art software, for example, Adobe Flash, SQL Loader, gfortran, OpenAMF, GrADS, Statistical Product and Service Solutions (SPSS) and Ferret. The system architecture and functionalities are detailed in the discussion to follow. We first discuss the query system in subsection 8.3.3. TC prediction is then examined in section 8.4.

8.3.3 Spatial Query

TC query plays a key role in TC analysis in TCAPS. The purpose of TC query here is to retrieve TC tracks from the TC database with respect to the multiple spatial or non-spatial criteria, to mine and visualize useful information hidden in the historical TC database by powerful algorithms, and to provide useful reference for TC prediction in terms of genesis, track, intensity and landfall. The spatial query system is depicted in Figure 8-3. The TC query system efficiently finds the required TC tracks with respect to multiple criteria (e.g., spatial, temporal, GIS-based) via the effective query outlined in Figure 8-4.

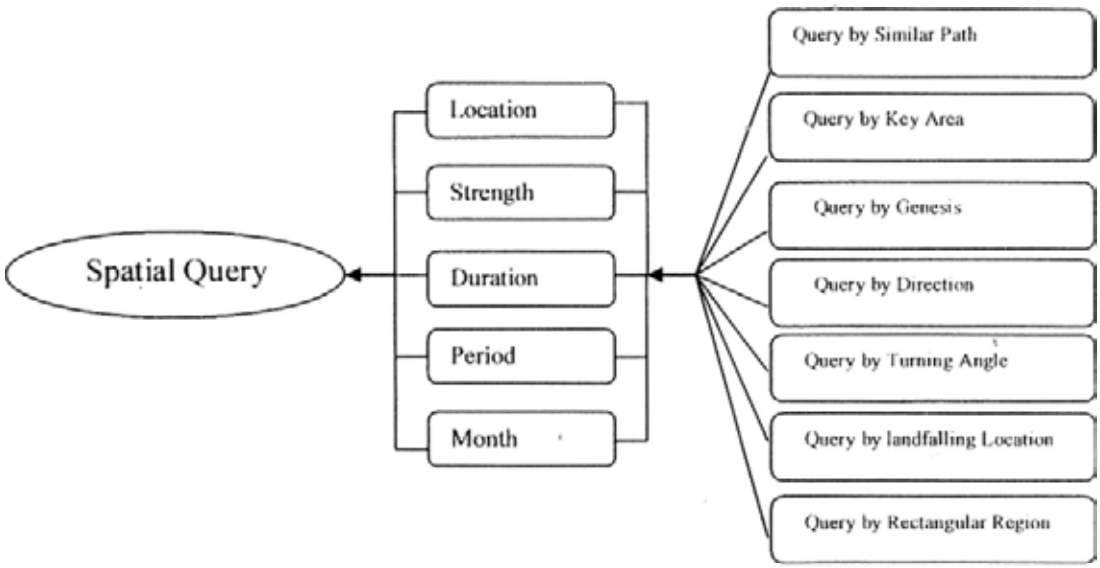


Figure 8-3 The Spatial Query System

Among these query methods, Query by similar path, Query by key area, and Query by landfalling locations find TC tracks passing through a buffer zone with certain radius, or through one or two arbitrarily specified areas (i.e., rectangle, circles and arbitrary polygon) and coastal areas (e.g., Guangdong province, Hong Kong) respectively. Query

by Genesis obtains useful information on the frequency of occurrence of historical TCs in a certain region, and unravels the TC hot spots in WNP and the SCS. Query by direction and Query by turning angle search TCs moving in a certain direction and turning in a certain angle respectively.

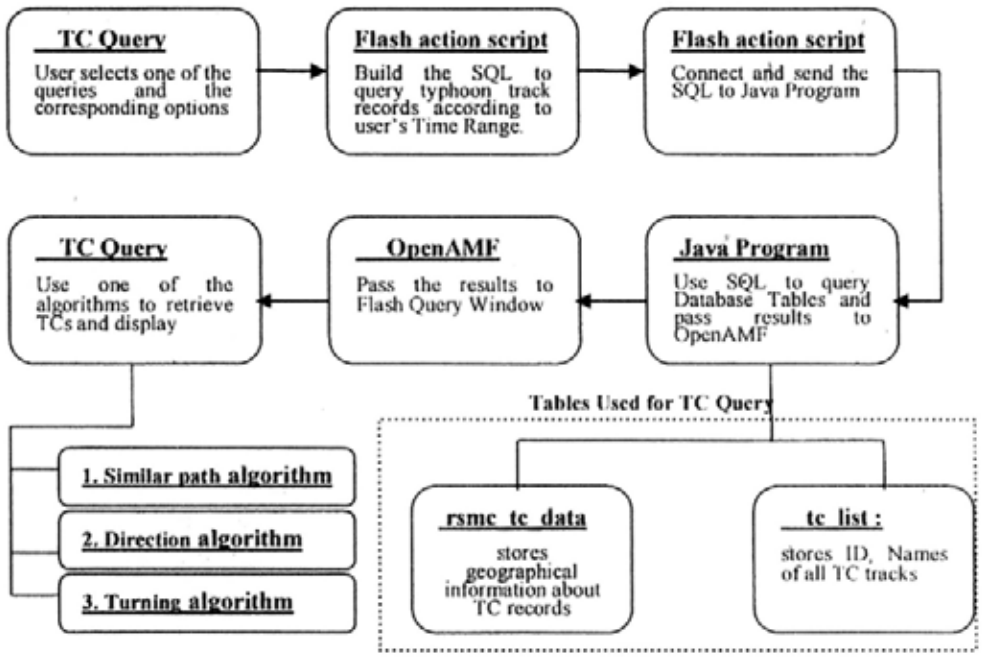


Figure 8-4 The TC query interface and procedure

During each query, users can specify location, intensity (e.g., tropical depression, tropical storm), duration (e.g., 10 days, 12 days), period (e.g., from 1951 to 2007), and month of the TC. Therefore, each query is multicriteria in space and time. Various queries are further explained in the discussion to follow.

1) Similar Path Query

Similar path query is a kind of case-based reasoning (Aamodt and Plaza, 1994; Kolodner, 1993; Kurbalija et al., 2009). It aims at the retrieval of all TC tracks that passed through a specified buffer zone. The buffer zone treats the selected TC track segment as the buffer center. This algorithm is divided into two sub-functions, namely "Locating TC tracks passing through a rectangular region" and "Finding tracks with similar path".

(a) Locating TC tracks passing through a rectangular region

The query aims at the retrieval of all TC tracks that pass through an area specified by the user. As shown in Figure 8-5, the semi-transparent blue rectangle represents the region of interest on the map. By dragging the 2 purple points freely at the corners of the rectangle, the user can move or resize the region on the map. The query algorithm then finds and displays all TC tracks passing through the rectangular region on the viewer.

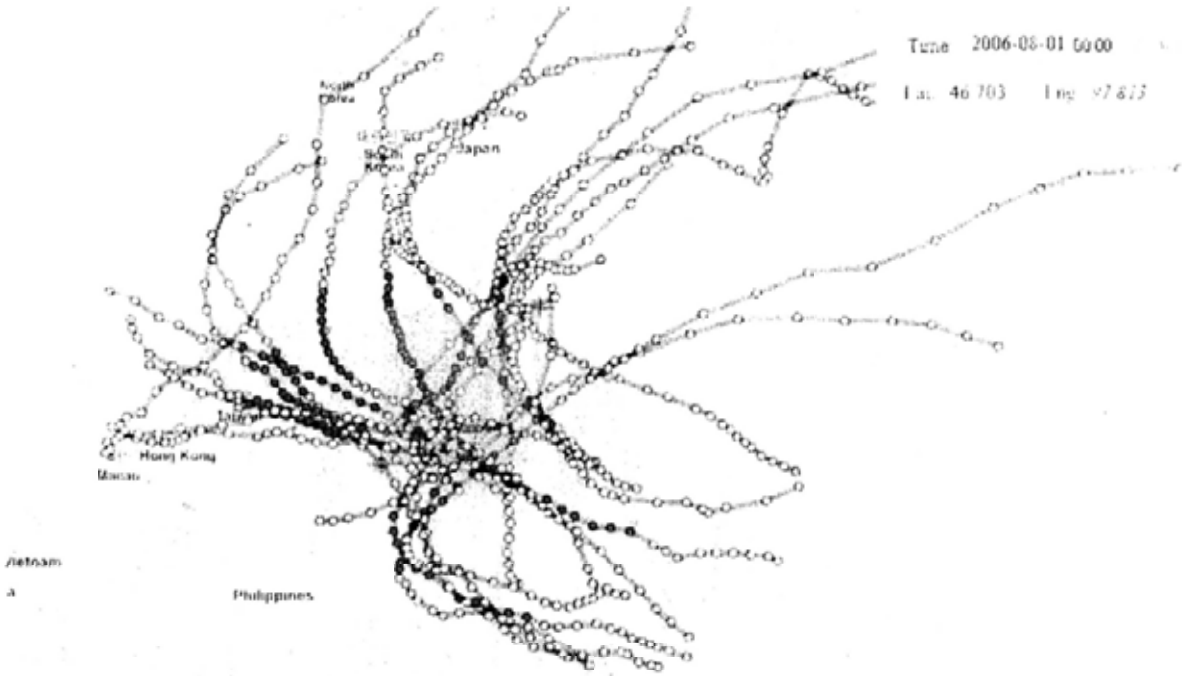


Figure 8-5 Locating TC tracks passing through a rectangular region

(b) Finding tracks with similar path

The purpose of this query is to retrieve all tracks going through a path specified by the user. As depicted in Figure 8-6, users can draw points by left-clicking the mouse on the screen bounded by the semi-transparent blue rectangle, and then draw the subsequent points to form the shape of the path in pink color. A TC track is regarded as having a similar path if it falls within the specified path or if its starting and ending points locate at the two ends of the specified path(e.g., labelled by Area 1 and Area 2 in Figure 8-7).

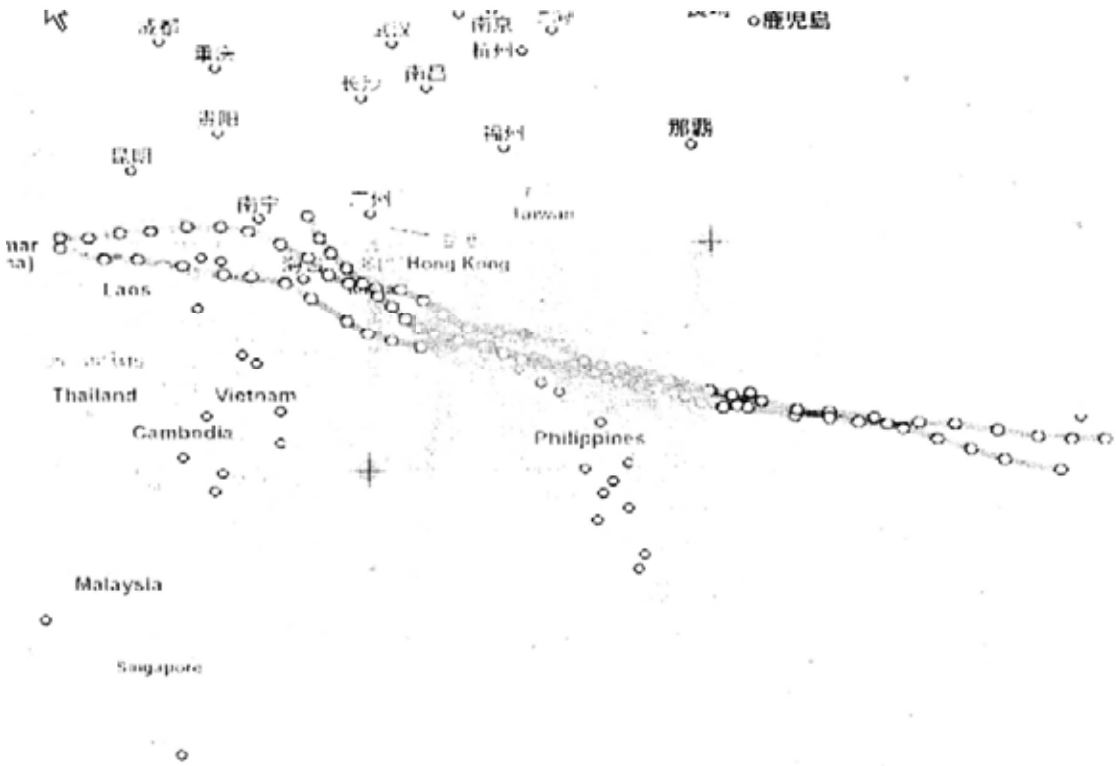


Figure 8-6 Finding tracks with similar path

The Similar path algorithm can be modified to predict the future path of an active TC track by using the subsequent paths of the tracks to which it is similar. The integration of this query with seasonal similarity and clustering results for TC movement prediction is discussed in section 8.4.1.

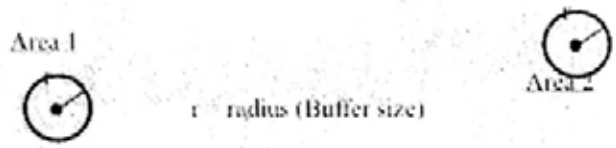


Figure 8-7 The buffer for querying by similar path

2) TC Genesis Query

Genesis of a TC is the point at which the TC first reaches the intensity level of tropical storm. TCs occurring at different locations might possess different characteristics in terms of shape, intensity, landfall and recurvature in their subsequent developments (Harr and Elsberry, 1991; Liu and Chan, 2003). Thus, this query might be instrumental to the prediction of TCs based on their initial conditions.

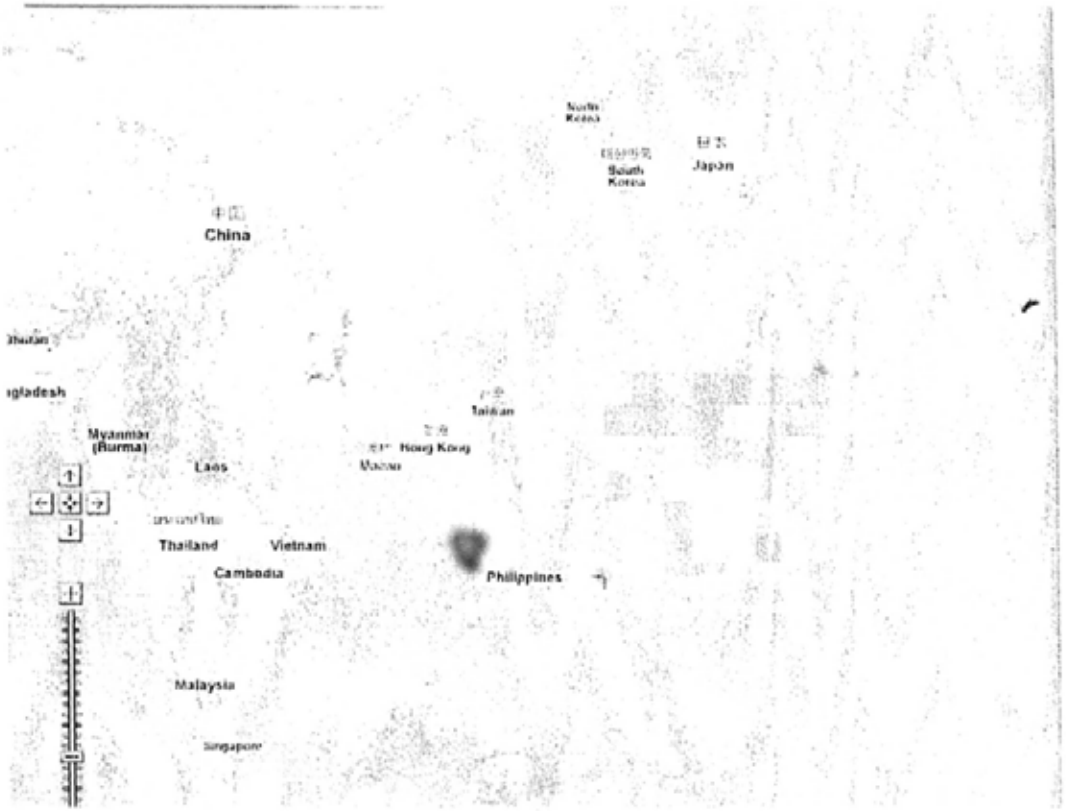


Figure 8-8 Visualization of TC genesis by latitude-longitude grids

TC genesis deals with the counting of TCs that started within a specified latitude-longitude range, duration, and time (month) range. The records of TCs are retrieved from the database. The starting latitude and longitude of a TC is defined as the earliest record of the TC. Since the records can be ordered by latitude and longitude according to the query, no sorting of the records is required. The counting result will be passed to a function for visualization in color. Grids with different number of occurrences are displayed in different colors (Figure 8-8). TC tracks that start from individual grids can also be displayed (Figure 8-9).

Figure 8-9 TC tracks from various latitude-longitude grids

3) Query by Key Areas

This function allows users to retrieve TCs passing through one or two areas of interest. Single area query enables users to analyze the degree to which the area is subjected to TC attacks (Figure 8-10). By specifying another area, the system can reveal all TCs that have passed through the previously specified area and the current area (Figure 8-11). The query can be done with respect to time and other options.

Figure 8-10 Single area query

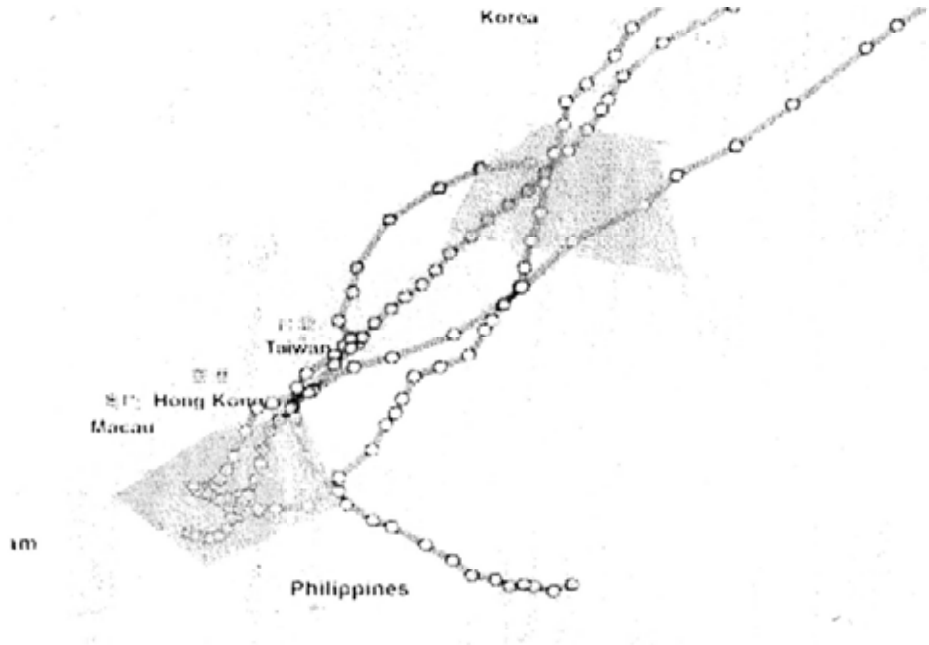


Figure 8-11 Two-area query

4) Query by Direction

Each TC moves along certain direction throughout its life. Directions of TCs in the WNP and the SCS exhibit different characteristics in terms of landfall, recurvature and intensity. This algorithm helps to find the movement directions of TCs.

Users can query the database in eight directions: East, North East, North, North West, West, South West, South and South East (Table 8-3) and can select the threshold, percentage of the TC path that lies in the selected direction, as a basis of query.

Table 8-3 TC directions (Degree)

Direction	Degrees
East	0 ~ 22.5; or 337.5 ~ 360
Northeast	22.5 ~ 67.5
North	67.5 ~ 112.5
Northwest	112.5 ~ 157.5
West	157.5 ~ 202.5
Southwest	202.5 ~ 247.5
South	247.5 ~ 292.5
Southeast	292.5 ~ 337.5

Figure 8-12 is the result of a query with respect to the northeast direction and a 40% threshold.

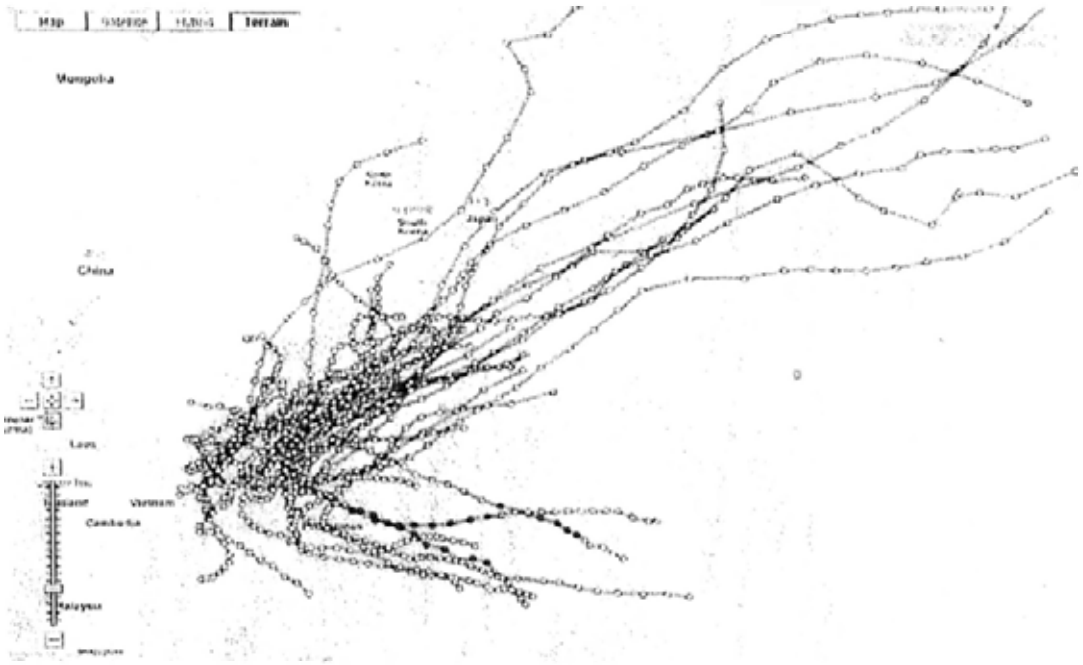
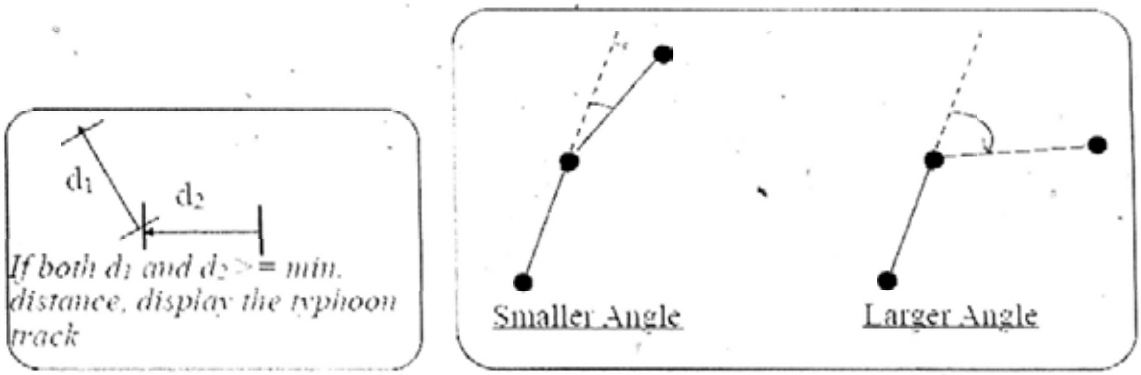


Figure 8-12 TCs from 1st January 1951 to 1st January 2007 with 40% of the paths moving northeast

5) Query by Turning Angle

Affected by various factors, TCs might change course in their movements. Sudden turning of a TC always causes considerable damages to areas lying on its path if they are caught unprepared. The query is to retrieve TC tracks that change courses at certain angles and at certain points in time.

To retrieve relevant TC tracks, users can specify turning angles, number of turns, and minimum distance between two adjacent segments at the turning point. The turning angle is determined by using the vector dot product $\vec{a} \bullet \vec{b} = |\vec{a}||\vec{b}| \cos\theta \Rightarrow \theta = \cos^{-1}(\vec{a} \bullet \vec{b} / |\vec{a}||\vec{b}|)$ to find the angle between 2 sequential segments (Figure 8-13). The user can choose a threshold value to see whether the turning angle of a TC track is larger than the threshold angle.



(a) the algorithm (b) specification of angle

Figure 8-13 Query by turning angle (a) the algorithm (b) specification of angle (It defines the angle between 2 sequential segments d1 and d2.)

Figure 8-14 depicts TC tracks retrieved for the period from January 1st 1951 to January 1st 1965 with turning angles ≥70 degrees, one single turn and minimum distance 0.6 degree in latitude or longitude.

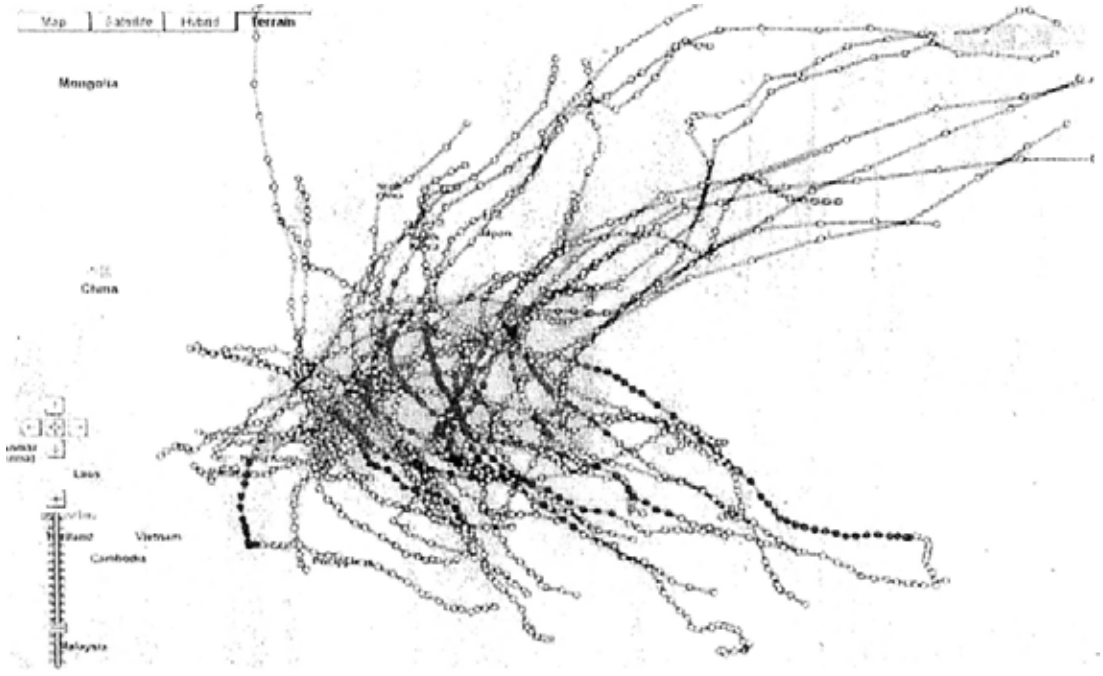


Figure 8-14 TCs retrieved by turning angle (The blue rectangle demonstrates the selected query area relevant TCs passed through.)

6) Query by landfalling location

TC landfall (Ching et al., 2000; Lyons, 2004) always inflicts severe damages to coastal areas. TCs that landfall in different coastal areas may be similar in track shape and trend.

Query by landfalling location can help to analyze all TCs that landed in certain coastal area in selected months and periods. In Figure 8-15, it is clear that most of the TCs that made landfall in Guangdong province and moved in a more or less straight path, with only a few that recurved.

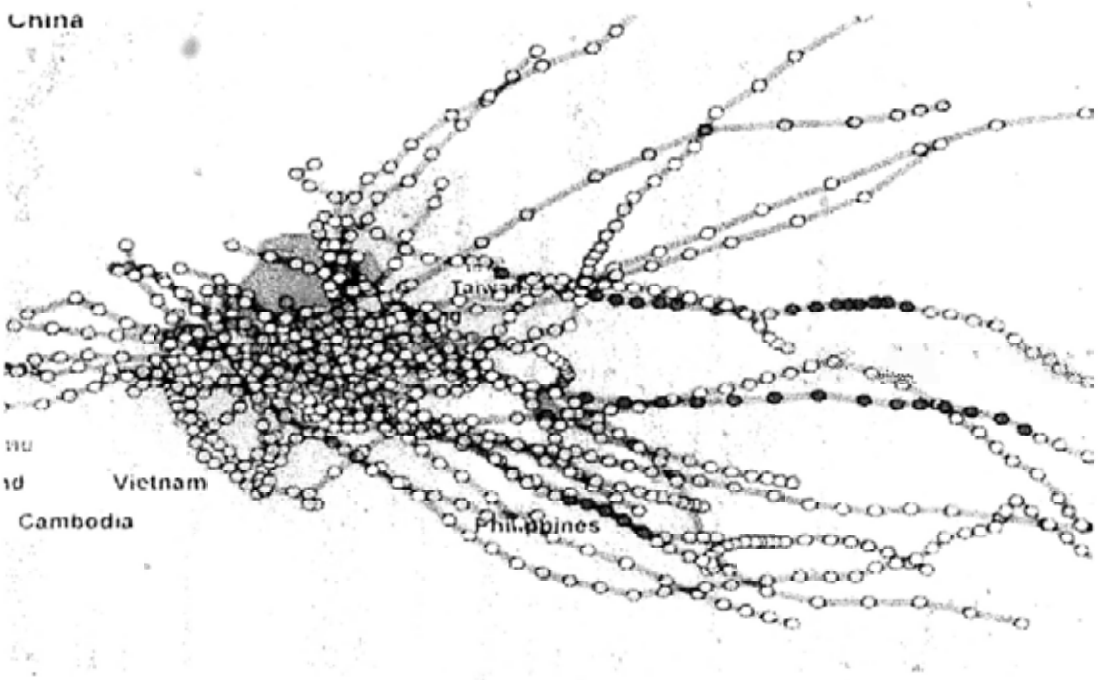


Figure 8-15 TC tracks landed in Guangdong province from 1997 to 2007

7) Additional Query Options

When users are performing any of the above queries, they could set additional options for TC retrieval. The additional options are about period, location, length and duration. Users can specify the query period in terms of year, month and day; and/or location in terms of coastal area, region or a city; and/or strength according to, for example, Table 8-4. Query results are tabulated in the result list showing TC IDs, names and start dates. Selected TCs are then plotted onto the Google map for visualization.

Table 8-4 The intensity level of TC by the Chinese Meteorological Administration

Grades of Tropical Cyclone	Strength on Knots
Super Typhoon	More than or equals to 100
Severe Typhoon	81 ~ 100
Typhoon	64 ~ 81
Severe Tropical Storm	48 ~ 61
Tropical Storm	34 ~ 48
Tropical Depression	Less than or equals to 33

8.4. Prediction of Tropical Cyclones

8.4.1 Mining of Historical Tracks for TC Prediction

Among other topics, prediction of TC track is a main area of research in TC analysis and is of public concern. To facilitate analysts to make prediction, the present web-based system renders the data mining approach and the dynamic-model approach to TC prediction. It also offers the integrated use of both approaches to enhance prediction.

The data mining approach, at present, offers TC-pattern clustering by the FMM and TC recurvature prediction by CART. The dynamic model, MM5, on the other hand, is incorporated into the current system for TC-track prediction.

The movement of a current TC may be predicted by the movements of the past TC tracks exhibiting similar characteristics. Thus, data mining methods can be employed to unravel the characteristics of relevant TC patterns in the past and to shed light on the prediction of the current TC. TC track clustering through FMM, for example, has been performed over the years (Camargo *et al.*, 2004; Camargo *et al.*, 2008; Camargo *et al.*, 2007a, b; Gaffney and Smyth, 1999; Gaffney *et al.*, 2007; Gaffney, 2004). Camargo et al. (Camargo *et al.*, 2007a, b) discovered 7 clusters in historical TC track archive of WNP, while Gaffney (2004) found 10 clusters in similar TC-track database.

Integrating spatial similarity, temporal similarity and the clustering procedure of Camargo et al. (2008), we constructed a TC track prediction scheme, which is similar to the case-based prediction scheme and CLImatology and PERsistenc (CLIPER) (Neumann and Hope, 1972) scheme.

In the present system, the FMM-based clustering model is implemented within the Matlab platform (Camargo *et al.*, 2004; Camargo *et al.*, 2008; Camargo *et al.*, 2007a, b; Gaffney and Smyth, 1999; Gaffney *et al.*, 2007; Gaffney, 2004). The parameters β, σ^2 and v^2 in equation (1) (Gaffney, 2004), characterizing the shape of each cluster can be obtained as follows:

$$\begin{aligned} p_k(y_i) &= \int p_k(y_i, t_i)dt_i \\ &= \int p_k(y_i \mid t_i)p_k(t_i)dt_i \\ &= \int \mathrm{N}(y_i \mid X_i\beta_k + t_i, \sigma_k^2\mathrm{I})\mathrm{N}(t_i \mid 0, v_k{}^2)dt_i \\ &= \mathrm{N}(y_i \mid X_i\beta_k, v_k{}^2\mathrm{1} + \sigma_k^2\mathrm{I}) \end{aligned} \tag{8.1}$$

where y_i is the i-th curve of length n_i (it denotes the i-th TC track), $p_k(y_i)$ is the probability of track y_i belonging to the k-th cluster, t_i is a curve-specific translation scalar for the i-th track, β_k is the $p \times 1$ mean coefficient vector of the k-th cluster, and $\sigma_k{}^2$ and $v_k{}^2$ are variances of y_i and t_i in the Gaussian model for the k-th cluster respectively . Here, according to Gaffney (2004),

$$\begin{aligned} \hat{v}_k{}^2 &= 1/n\sum_i g(\hat{t}_{ik}), \\ \hat{\sigma}_k{}^2 &= 1/N\sum_i \hat{f}(\hat{t}_{ik}), \\ \hat{\beta}_k &= [\sum_i X_i{}' X_i]^{-1}\sum_i X_i{}'(y_i - \hat{t}_{ik}) \end{aligned} \tag{8.2}$$

where $\hat{v}_k{}^2$, $\hat{\sigma}_k{}^2$, $\hat{\beta}_k$ are the estimations of $v_k{}^2$, $\sigma_k{}^2$ and β_k, n and N are the total number of sample TC tracks, $\hat{d}_{ik}$ is the estimation of t_{ik}, and t_{ik} is the curve-specific translation scalar of the i-th track for the k-th cluster, X_i is the Vandermonde matrix evaluated at the i-th TC track and X_i' is the transpose of X_i.

Using the longitude and latitude of each point of a new TC track as input of the clustering model, we can calculate the probability of the current active track belonging to each TC class using equation (8.1). Essentially, the clustering procedure employs the TC position orders (e.g., 1, 2, and 3) as a basis in the polynomial regression. The shapes of the TC tracks (curves) are depicted by the parameters of the polynomial regression model. Different shapes are integrated into the FMM according to the distribution of the error terms. Then, the current TC will be assigned to the class with the largest posterior probability. The characteristics of the TC class and the historical TC tracks belonging to this class can be used as a reference for forecasting. Once an active TC track has been assigned to a cluster, all the historical tracks of that cluster will be retrieved and employed to predict the shape and movement of an active track.

Besides the clustering results, spatial similarity or case-based reasoning via the "Query by Similar Path" can also be employed to find useful historical tracks for prediction. In general, TCs occurring in a certain season have special or similar characteristics according to the position and strength of the subtropical high and monsoon trough (Ho et al., 2004; Lehodey et al., 1997; Lu and Dong, 2001). Thus, historical TC tracks occurred in the same season as the current active track can be used to predict the movement of the latter. This is called "Temporal similarity" or "Seasonal similarity". Thus, TC tracks unravelled satisfy both the clustering results and the spatial and temporal similarity query results. The tracks obtained from these three procedures will then be cross compared and integrated, and the subsequent points of movement of the integrated track can be employed to make more reliable prediction of the current track.

Figure 8-16 shows the flow of the above prediction procedure, and Figure 8-17 depicts a reasonably accurate experimental result of such prediction. The blue buffer is the buffer zone of the selected TC track segment. The yellow line with blue dots is the predicted 72-hour track. The other is the real TC track.

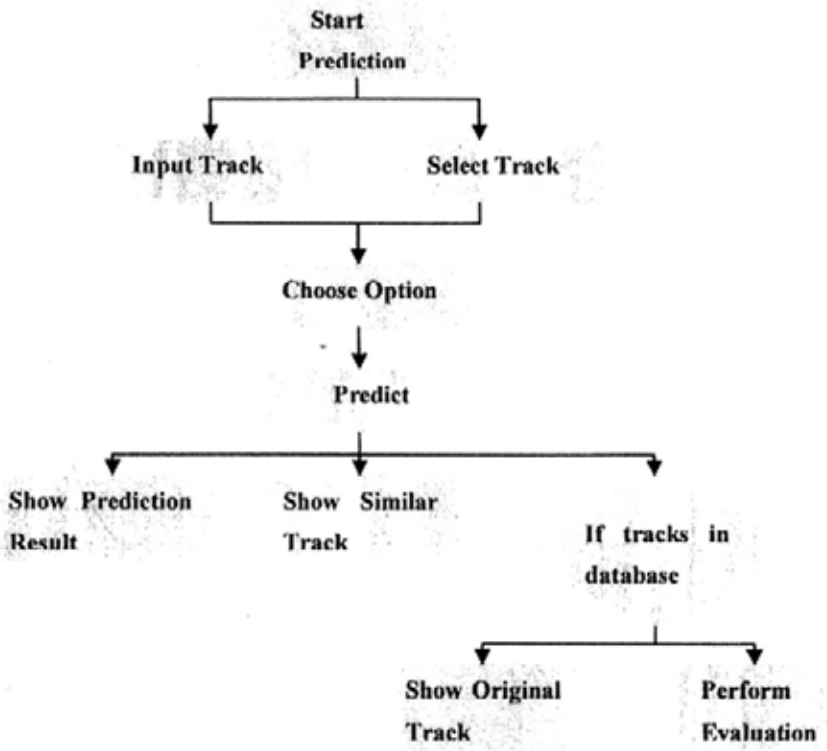


Figure 8-16 The flow of the scheme for TC track prediction

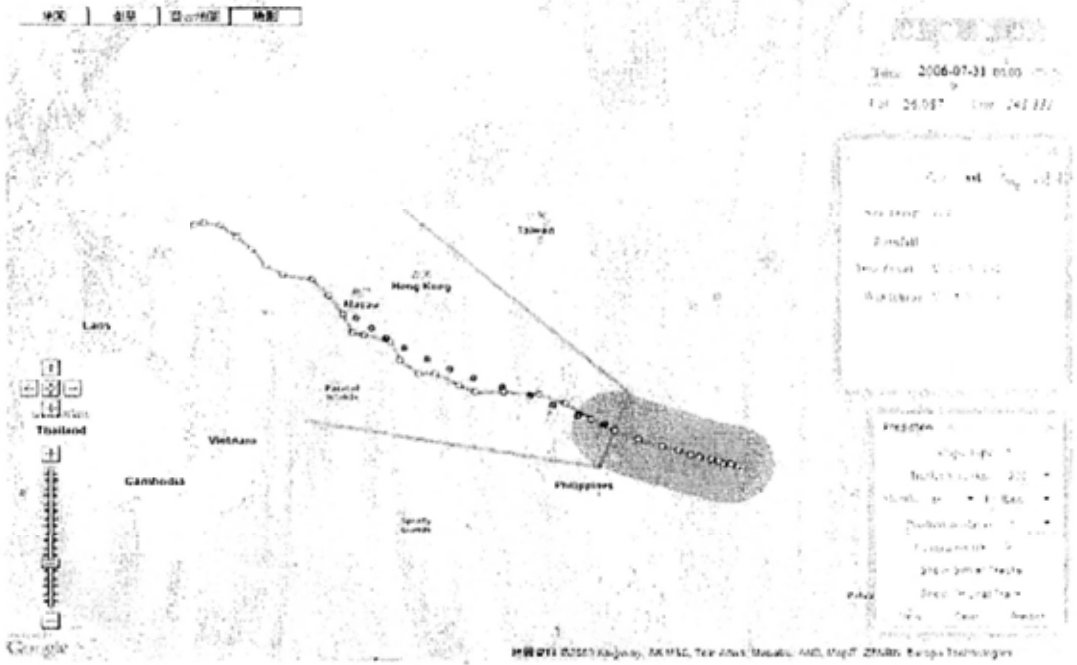


Figure 8-17 TC track predicted by FMM

8.4.2 Prediction Scheme based on MM5

In place of making prediction through the method of clustering, the direct modeling of TC dynamics has been attempted over the years. Among others, MM5 is one of the most widely-used numerical weather prediction systems (e.g., Deng et al., 2004) developed partly for such purpose. Its construction is based on the physical processes and mechanisms plausibly governing TC development and dynamics and it has been employed to predict TC movements (Juneng et al., 2007; Mandal et al., 2007; Ramsay and Leslie, 2008). Due to high computational cost and complexity, we have implemented MM5 into the 32 x 4 parallel computation environments, and it is run using a quadcore.

There are many atmospheric datasets available on the Internet, and these datasets have different characteristics. The two major datasets used by MM5 in the present system are "FNL" and "GFS data", both of which are published by NCAR/NCEP. The web link to "FNL" dataset is http://dss.ucar.edu/datasets/ds083.2/, whereas the link to"GFS" dataset is http://www.nco.ncep.noaa.gov/pmb/products/gfs/. "FNL" is the abbreviation for "Final analyses". This dataset contains historical data in high resolution. By using this dataset, we can use MM5 to perform past TC simulations. "GFS" is the abbreviation for Global Forecasting System. The "GFS" database contains high resolution data for future prediction. The dataset can be used by MM5 to perform TC prediction.

Since "GFS data" is using GRIdded Binary 2 (GRIB2) format, MM5 cannot read it directly. Data preprocessing is needed to convert data from the GRIB2 format to the GRIB1 format. NCEP also provides a tool called "cnvgrib", which can convert a GRIB2 file to GRIB1 file and vice versa.

Figure 8-18 shows the experimental result of a MM5 prediction of the sudden recurving of a TC. The TC in this case is "CHANCHU" whose lifespan is from 2006-5-9 20:00 to 2006-5-19 17:00, and its serial number is '200601'. The 12 red dots in Figure 8-18 are the points in time of the predicted track of the active TC after 20:00 on 14th of May. For comparison, the real TC track (in rainbow colors) is also shown in Figure 8-10. It can be observed that MM5 successfully predicted the sudden recurvature of this TC.

Figure 8-18 TC track forecasting by MM5

8.4.3 TC Recurvature Analysis through CART

Different from dynamic modeling, the data mining approach aims at the discovery of the mechanisms/rules characterizing the conditions under which TC might recurve. Instead of using process models, the data mining approach tries to unravel from TC data rules governing the recurvature of TCs throughout their movements. Data mining, in general, is the process of extracting hidden and useful patterns and information from data (Fayyad and Stolorz, 1997; Han and Kamber, 2006; Leung, 2010). In recent years, a number of data mining algorithms have been employed to unravel TC tracks and intensities in TC database (e.g., Camargo *et al.*, 2004; Camargo *et al.*, 2008; Camargo *et al.*, 2007a, b; Cheng *et al.*, 2008; Gaffney and Smyth, 1999; Gaffney *et al.*, 2007; Gaffney, 2004; Harr and Elsberry, 1991, 1995b; Lee *et al.*, 2007). However, only a few data mining derived results have been successfully applied to predict TC movement. The data mining approach and dynamic modeling can actually complement each other to achieve more reliable and comprehensive prediction of TC movements. Data mining can uncover useful information for the enrichment of the dynamic models, and the mechanisms underpinning the dynamic models can be used to guide the data mining process.

For TCs in the northern hemisphere, recurvature is the change from a north-westward to a north-eastward track moving into the mid-latitude westerlies with the prevailing westerly flow (Dobos and Elsberry, 1993). TC recurvature in the Northern Hemisphere is sensitive to the change speed of the zonal and meridional wind in the 200 hPa and 500 hPa layers within the octant 1, 2, 3 (north and west) of the 6 - 8 degree radial belt (Hodanish and Gray, 1993). Chan et al. (1980) demonstrated that zonal and meridional wind (u and v wind) in the 200 hPa, 300 hPa, 400 hPa, 500 hPa, 700 hPa, 800 hPa, 850 hPa, 900 hPa, 1000 hPa layers (especially the 200 hPa, 500 hPa, 700 hPa layers) in the 5 – 7 degree radius circle play an important role in deciding the recurvarture. Mean values of geopotential height (Lage, 1982; Leftwich, 1980) on the 500 hPa layer for 5 – 7 degree radius circle are important factors in TC recurvature. TC intensity can also affect TC recurvature. However, the physical mechanism leading to TC recurvatures still needs to be exactly determined (Bao and Sadler, 1983; Evans and McKinley, 1998; Knaff, 2009).

The present system provides a data mining procedure to discover conditions under which TCs may change their directions of movement. The idea is to unravel from data mechanisms affecting TC recurvatures. The ultimate purpose is to integrate the dynamic modeling and the data mining approaches to achieve more in depth and accurate prediction. The data processing methods of Hodanish and Gray (1993) is chosen to calculate the attributes relevant to TC recurvatures. They are, for example, zonal and meridional wind in 200 hPa, 500 hPa and 700 hPa layers, the geopotential height in 200 hPa and 500 hPa layers, in the octant 1, 2, 3 (north and west) for 6 - 8 degree radial belt and MSW.

CART is chosen as a method for data mining because of its simplicity and interpretability (Breiman *et al.*, 1984b; Leung, 2010). The primary purpose is to unravel rules for TC recurvature. The sample data of the initial experiment were taken from 2000 to 2009. The potential factors that influence TC recurvature are categorized into three groups: the variables relating to large-scale circulation, the variables measuring the circulation surrounding TC, the variables describing the characteristics of TC (see Table 8-5).

Table 8-5 The potential attributes influencing TC recurvature

Groups	Potential Variables
Circulation surrounding TC	uwnd_200, uwnd_300, uwnd_400, uwnd_500, uwnd_600, uwnd_700, uwnd_800, uwnd_850, uwnd_1000, vwnd_200, vwnd_300, vwnd_400, vwnd_500, vwnd_600, vwnd_700, vwnd_800, vwnd_850, vwnd_1000
Large-scale circulation	area_IndexSTH, inten_IndexSTH west_extSTH, Monsoon_WF W_Westerly
Characteristics of TC	Lon, Lat, Pressure (Central Pressure of TC Center)

The area_IndexSTH, inten_IndexSTH and west_extSTH are the area index, intensity index and westward extension index of the subtropical high respectively. Monsoon_WF is the monsoon index proposed by Wang and Fan (1999), and W_Westerly is the westerly index.

In Table 8-5, the variables are displayed in abbreviations. In the group "Circulation surrounding TC", "uwnd_200" and "vwnd_200" are respectively the average zonal and meridional wind of 6-8 degree radial belt at 200-hPa level. The other variables in the same group are defined likewise. Five variables chosen to measure the strength and position of "large-scale circulation" (i.e., the subtropical high, EASM, westerlies) that largely control TC recurvature (e.g., Hodanish and Gray, 1993) are area index (area_IndexSTH), intensity index (inten_IndexSTH), and westward extension index of the subtropical high (west_extSTH) in WNP (for measuring the strength of the subtropical high and position of STR), EASM index by Wang and Fan (1999) (Monsoon_WF) (demonstrated by to be nearly identical to the leading principal component of the EASM and greatly facilitates real-time monitoring), as well as westerly index (W_Westerly) proposed by Rossby (1939). For the characterization of TCs, we use the Longitude (Lon) and latitude (Lat) of TC center, as well as Central Pressure of TC (Pressure) in the CART mining process.

From the result, we can see that Lon, Lat, Pressure, uwnd_200, uwnd_1000, vwnd_800, vwnd_850, vwnd_1000, area_IndexSTH, west_extSTH and Monsoon_WF are chosen by CART to build the decision tree (Figure 8-19) stipulating 18 unravelled rules governing TC recurvature. We can observe that the average accuracy of TC recurvature prediction by CART is 84.364%. In the figure, "1" means recurvature and "0" means non-recurvature respectively. The rectangles are leaf nodes, while ellipses or circles are parent nodes. Taking leaf node "0(263.0/12.0)" as an example, "0" in front of the bracket means non-recurvature and "263.0" and "12.0" indicate that among the 263 samples of the leaf node, there are 251 non-recurvature samples and 12 recurvature samples respectively. A path from the root node to the leaf node represents a rule, which can provide reference for TC recurvataure prediction. Each rule can be justified by meteorological and TC theories. Taking the rule formed by the path from root node to leaf node "1(617.0/26.0)" as an example it can be stated as:

"If the longitude of a TC center is to the east of 130 °E, the central pressure is less than 1006 hPa, the number of grids within the 5880 gpm contour of 500 hPa geopotential height is less than 314, the average zonal wind at 200hPa is mainly westerly, and the western ridge of the subtropical high shifts to the east of 133 °E, then the TC will recurve". According to this rule, TCs will recurve under the conditions of weak and retreating subtropical high and moderate mid-latitude westerlies. In terms of the geopotential height and deep-layer mean wind vector calculated by the TC samples belonging to this rule are all in line with the analysis (Figure 8-20). The 5880 gpm contours of Figure 8-20 represent the high center of the subtropical high. The deep-layer mean wind is, in general, the pressure-weighted mean wind averaged from 850 hPa to 200 hPa layer. Therefore, the deep-layer mean indicates the steering current that determines the TC movement to a large extent. It has been reckoned that the deep-layer mean wind fields, e.g. 850-200 hPa, are best for the forecasting of TC movement (Holland, 1993b).

The above CART result is only for the purpose of pedagogy. A more comprehensive study will be carried out to discover rules governing TC landfalls and recurvatures within the complex monsoonal system.

Figure 8-19 Rules governing TC recurvature unravelled by CART (The rectangles are leaf nodes, while ellipses and circles are parent nodes.)

To recapitulate, the present system provides facilities for the prediction of TC movement by the data mining and dynamic modeling approaches. The initial implementations of FMM and CART for data mining and MM5 for dynamic modeling have shown the effectiveness of both approaches. The next step is to perform more in-depth investigation of these methods and to take advantages of both approaches to achieve more accurate prediction through their effective integration.

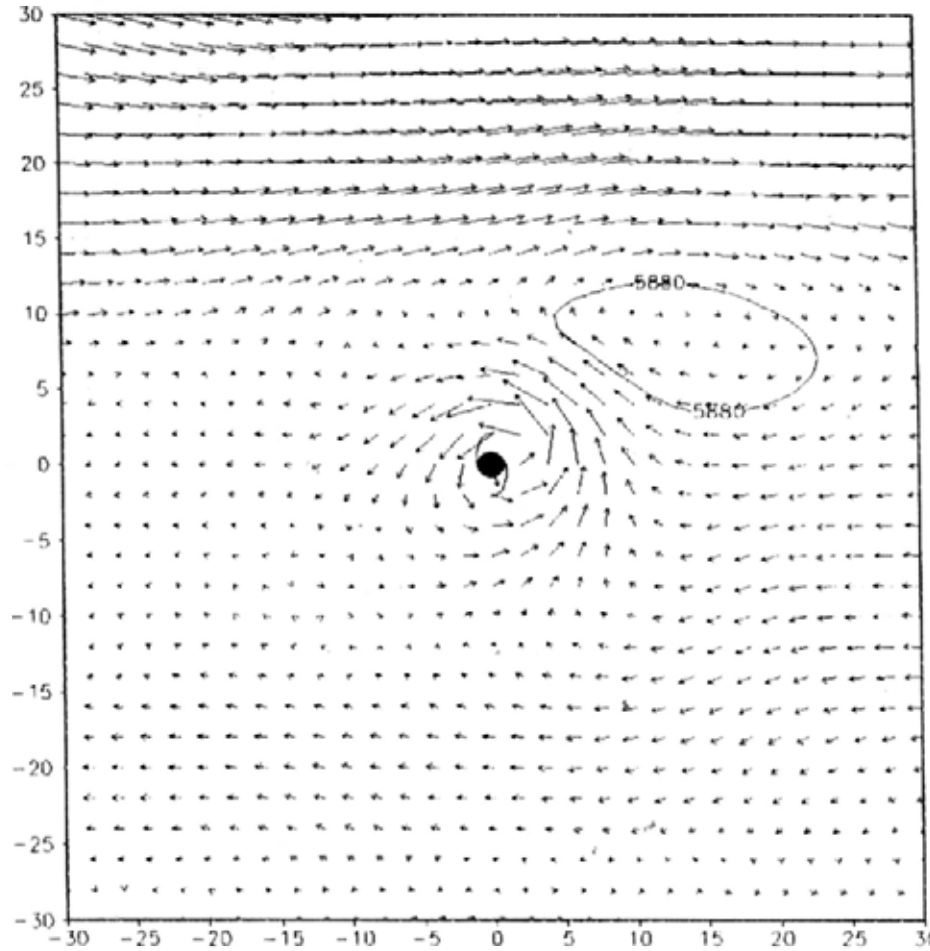


Figure 8-20 The composition of deep-lay mean wind fields and 500 hPa geopotential height of the samples in the CART node (0(263.0/12.0)). The TC symbol denotes the relative TC center with the coordinate (0, 0). The axes on the bottom and left of the plot represent the x and y ordinates that represent east and north respectively.

8.5 Summary

In this chapter, novel methods and advanced information technologies have been employed to build a TC analysis and prediction system suitable for academic research, practical application, and public education. For technical research, the system provides facilities for process models and data mining algorithms. Though they are made transparent to users, a more flexible interface can be built in the future for models management and algorithm implementation to facilitate researchers to perform TC analysis and prediction. For the general public, the system provides user-friendly

interface for TC query and visualization. It is thus a web-based system with high academic and educational value and has great potential for technology transfer.

The significance of this system lies in its efficient and effective visualization of historical TC archives and the retrieval of data for analysis and forecasting in the web-based environment. Its user-friendly interface and efficient query system greatly facilitate public consumption and professional use. Its facilities for the employment of dynamic models and data mining algorithms make it a powerful system for TC analysis and prediction, particularly on TC landfalls and recurvatures. Furthermore, the TC database is extensible for future applications.

The present system is a prototype that can be extended into a fully fledged system in which TC dynamics can be studied within the monsoonal system affecting the development of TCs (Elsberry, 2004; Harr and Chan, 2004). The data mining function of the system, in particular, can be extended to unravel rules stipulating the interplay between TC and MJO (Madden and Julian, 1971, 1972; Madden and Julian, 1994), QBO (Ebdon, 1960; Reed et al., 1961) and ENSO including El Niño and La Niña occurrence (Chan, 2000; Wang *et al.*, 2003). To increase its function and applicability, it can be further developed into a real-time system with monitoring and mitigation capabilities.

Chapter 9: Conclusion and Discussions

9.1 Conclusion

Extreme and more intense TCs tend to occur more frequently under climate change (Knutson et al., 2010). TC intensity and movement are two problems of paramount scientific and operational concern. TC movement research has dramatically advanced in the last few decades. However, TC intensity research lags behind that on TC movement. TC recurvature has been investigated since the 1940s by meteorologists using observational studies and modeling physical processes. It is still a challenging field for operational forecasters and researchers all around the world.

TC landfall, the process of a TC moving over land, is also of significant importance because previous researches on TC landfall were concerned with seasonal or annual TC landfall frequency according to the variability of large-scale circulation (e.g., the subtropical high, monsoonal systems and westerly troughs) modulated by ENSO, QBO and MJO, with the precipitation and wind distribution associated with TC landfall, with the changes in their tracks and structures when TCs passed through an island by studying land-TC interaction, and with TC decaying mechanisms post-landfall. The Chinese coast, as the landfall location, is seldom examined regarding the characteristics of TC landfall in the light of the combination of large-scale circulation and environmental fields surrounding TC centers.

When a TC moves overland, the cutting-off of the water vapor supply, greater surface roughness, interaction with mid-latitude systems, and decreasing upwards motion at the TC center make post-landfall TCs dramatically different from those on the ocean. Due to the complexity and difficulty of studying TC-land interaction post-landfall, there has been far less research conducted on post-landfall TC characteristics than on TC tracks and intensity. The specific characteristics of TC intensity change both before and after landfall over the Chinese coast have seldom been examined. The relationship between

TC recurvature and maximum intensity remains ambiguous based on K09, and TC intensity change has not yet been explored comprehensively with respect to TC recurvature.

Given the shortcomings and limitations of the current TC platforms, a web-based, free-access GIS system should be built to make GIS technology and TC knowledge available to researchers and the general public at little or no cost. It should have functionalities for the visualization of historical TC archives, complex data analysis and information distribution for both research and professional use, and also for the consumption of the general public.

Therefore, our understanding of TC recurvature, landfall, post-landfall characteristics and track-intensity interaction plays a vital role in the preparation for, mitigation and management of TC-related natural hazards. Dynamic modeling has been widely employed to simulate the physical processes of TC dynamics since the 1960s. However, predicting TC remains a tremendous challenge for operational forecasters. The abundance of data, together with the requirement for powerful data analysis techniques and tools, has led to a data rich but information poor situation (Han and Kamber, 2006; Kantardzic, 2003). The TC-related database is particularly difficult to visualize and understand because the data can grow along two dimensions: the number of fields (also depicted as attributes, features or variables) and the number of cases. The rate of growth of TC-related data sets proceeds much faster than the amount of material that traditional "manual" analysis methods can handle. Useful knowledge (e.g., structures, processes, relationships, regularities and patterns) underlying TC dynamics and mechanisms is inherently hidden in the historical TC database. Data mining is defined as "the nontrivial process of indentifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad and Stolorz, 1997; Fayyad *et al.*, 1996). Nonetheless, it is currently seldom used to investigate TC recurvature, landfall, and intensity, as well as track-intensity interactions in this study. Therefore, data mining methods, together with statistical methods, are employed to unravel the rules, patterns and regularities from the historical TC archives. The research findings are summarized as follows:

TC Recurvature

Two decision trees governing TC recurvature and landfall respectively are built via the C4.5 algorithm – a powerful, tree-based classification algorithm. In particular, 18 rules are derived from the decision tree for TC recurvature. Totally, Lon, Lat, Pressure, uwnd_200, uwnd_1000, vwnd_800, vwnd_850, vwnd_1000, area_IndexSTH, west_extSTH and Monsoon_WF are chosen by the C4.5 algorithm to build the decision tree (the explanations of these parameters are shown in section 4.3). The parameters measuring large-scale circulation and characterizing TCs play a significant role in building the classification tree. These 18 rules can be explained by the existing theories and supported by other empirical findings on TC recurvature (e.g., Chan *et al.*, 1980; Dobos and Elsberry, 1993; George and Gray, 1977; Guard, 1977; Harr and Elsberry, 1995a; Li and Chan, 1999; Riehl and Shafer, 1944) according to the verification of the results by calculating and investigating the composite wind fields, geo-potential heights and deep-layer mean winds in a square (60° × 60°) centering on the TC centers. These ambient wind fields, especially the deep-layer mean winds (from the 850 hPa to 300 hPa layer) signify the steering flow around the TCs for interpreting the rules, as well as by the 2010 TC best track data that is not involved in training the model. The differences between the rules and the existing theories and findings on TC recurvature are that the rules discovered in this investigation are strictly quantitative. However, the existing findings on TC recurvature, such as "A strong subtropical high and weak monsoon cause TC to recurve" are mainly qualitative results, or "rule of thumb". The rules governing TC recurvature from a quantitative perspective can provide useful references for TC track forecasters. A variety of splitting values in the classification tree can provide references for TC forecasting, since these values can best differentiate the recurving and non-recurving cases based on the archived TC data; for example, TC recurvature is vastly sensitive to the longitude "130 °E" because this longitude is first chosen by the C4.5 algorithm to build the classification tree. Besides, longitude "123 °E" should be noted in operational aids for TC recurvature. Based on the data mining results, these critical values will also provide useful references for TC recurvature forecasting.

TC landfall

A classification tree, with 14 leaf nodes, is built using the C4.5 algorithm for TC landfall. The path from the root node to each leaf node forms a rule. Eventually, 14 rules governing TC landfall over the Chinese coast are unraveled from the potential attributes (see Table 4-1) that may influence TC landfall. The rules are derived from the attributes and splitting values. Split values such as 27 °N in latitude, 130 °E in longitude, 141 °E in west extension index and 0.289 in monsoon index (Wang and Fan, 1999) are useful references for operational aids for predicting TC landfall. The rules are verified from the perspective of meteorology and the existing TC theories on movement and recurvature (e.g., deep-layer mean winds and large-scale circulation), as well as from the 2010 TC best track data that is not involved in training the model. The findings of this study are also consistent with the existing findings, results and theories on TC movement and landfall (e.g., Chen *et al.*, 2009; Harr and Elsberry, 1991, 1995a; Wang and Fan, 1999). Both the unraveled rules and the splitting values of these rules can provide useful references for operational aids for TC landfall prediction.

Post-landfall TCs

TCs trigger most of their damage during and post-landfall. The mechanisms or patterns characterizing post-landfall TC movement are insufficiently examined. FMM-based clustering algorithms are applied to post-landfall TC tracks to discover the hidden clusters underlying the historical TC tracks. Elevations at TC centers are taken into account in this clustering model because existing studies have found that the landfall surface or elevations influence TC movement post-landfall. Three clusters are uncovered from the historical post-landfall TC tracks. Cluster 1 makes landfall along the Fujian and Zhejiang coast and sustains for a long time, during which the TCs mostly recurve into the mid-latitude region. Cluster 2 makes landfall in Hainan province and along the western coast of Guangdong province. Most of the TC tracks in cluster 2 move over the ocean and makes secondary landfalls over Vietnam and western Yunnan province. Cluster 2 eventually dissipates inland due to the rough landfall surface in western Vietnam and the Himalayas near Yunnan province, China. Cluster 3 makes landfall over Guangdong and

Fujian provinces. Subsequently, the TCs in cluster 3 move inland and dissipate. The clustering results are interpreted and justified from the meteorological perspective (e.g., the geopotential height, steering flow, vertical and horizontal water vapor supply). We find that cluster 1 move northward and northeastward and recurve into the mid-latitude region because of the eastward-retreating subtropical high and westerly steering flow surrounding the TCs. By contrast, clusters 2 and 3 move westward or northwestward as a result of the westward-shift subtropical high and easterly steering flow. TCs of Cluster 1 sustain for a longer time than clusters 2 and 3 in spite of their weak horizontal and vertical water vapor supply. A plausible explanation for this is that the TCs in cluster 1 interact actively with westerlies during the post-landfall period. On the other hand, we cannot observe analogous interactions with westerlies in clusters 2 or 3. Besides, the composite 200 hPa divergence of cluster 1 is stronger than that of clusters 2 and 3 post-landfall. This explains why cluster 1 sustains for much longer than clusters 2 and 3 post-landfall to some degree.

TC Track and Intensity

The relationship between TC track and intensity is worthy of investigation. Statistical methods (e.g., the Student's *t* test and ANOVA) are used to test whether the variables of two groups (e.g., recurving and straight moving, landfall and non-landfall) are significantly different. It is found that, at least in WNP (including SCS), TCs tend to peak in intensity close to recurvature. We also found that the recurving points of TCs peaking in intensity post-recurvature are located to the southeast of those of TCs peaking in intensity prior to recurvature. It is also detected that recurving TCs sustain their maximum intensity for significantly longer than do straight moving TCs. TCs also tend to weaken significantly from landfall–24h to landfall+24h. It is consistent with the previous findings on TC landfall that TCs begin to interact with coastal areas one day prior to landfall. Meanwhile, they decay to a tropical depression or dissipate at landfall+24h. Therefore, TCs will not weaken significantly after landfall+24h. The results on the characteristics of TC intensity related to TC movement are of scientific and operational significance.

TC Platform

Novel methods, state-of-the-art software (e.g., Google Maps API, Oracle database, Adobe Flash, Java) and advanced information technologies have been employed to build a typhoon analysis and prediction system suitable for academic research, practical application, and public education. It is a web-based system with high academic value and great potential for technology transfer. The system has a user-friendly interface and powerful multi-criteria query system for the tracking and visualization of TC movement. It is also equipped with data mining and dynamic-modeling capabilities for the analysis and prediction of TC landfall and recurvature. It is thus a user-friendly and flexible system for research, professional, and the general public's consumption.

9.2 Implications

This dissertation has examined several TC-related problems through data mining. The research findings have a wide range of implications for our understanding and prediction of these problems.

Firstly, this study advances our understanding and paves the way for the more accurate prediction of TC recurvature. The C4.5 algorithm examines TC recurvature using the variables measuring large-scale circulation, meteorological variables surrounding the TC centers and variables characterizing TCs. The rules facilitate a quantitative framework for TC recurvature and, being derived from the classification tree of recurvature, provide new insights into this phenomenon. The key splitting values, such as longitude "130 °E" and "123 °E", are valuable references for the operational prediction of TC recurvature.

Secondly, this study enhances our understanding of TC landfall and paves the way for its accurate prediction over the Chinese coast. The C4.5 algorithm examines TC landfall using large-scale circulation and meteorological variable because the TC landfall over China and recurvature are closely related. The rules provide a quantitative framework for TC landfall and, being derived from the classification tree of landfall, provide new insights into TC landfall. The key splitting values, such as 27 °N in latitude, 130 °E in

longitude, 141 °E in the west extension index and 0.289 in the monsoon index, are valuable references for the operational prediction of TC landfall. The characteristics of TC landfall are associated with the location of TC centers (i.e., longitude and latitude), but few studies investigate the relationship between location and TC landfall. The findings indicate that TC landfall is indeed concerned with TC location. A plausible explanation for this is that the location reflects the status of large-scale circulation.

Thirdly, for the first time, cluster analysis has been applied to post-landfall TC tracks. In particular, three clusters have been unravelled from the historical post-landfall TC tracks in China. This result can be explained by the empirical findings of Chen and Ding (1979). Different from the previous research findings on sustaining TCs post-landfall, it is found that the water vapor supply and baroclinic potential energy are both important factors largely controlling TC sustaining. Previous studies accept the water vapor supply as the most important factor in TC sustaining.

Fourthly, studies of the relationship between TC tracks and intensity double the complexity and difficulties. Previous studies found that TCs tend to attain their maximum intensity close to recurvature. Knaff (2009) challenges this result by revisiting the relationship using the JTWC best track dataset. It is echoed in this study that TCs reach their peak intensity close to recurvature. This result may play a significant role in TC forewarning.

Lastly, the web-based system for TC prediction and analysis can help to facilitate the operational processes and procedures. The current TC platforms around the world focus on professional users. This system can fulfil the requirements of research/professional users, but also those of the general public. Currently, although several professional TC platforms exist, such as the Canadian Hurricane Center Forecaster's Workstation (MacAfee, 1997) , the Automated Tropical Cyclone Forecasting System (Sampson and Schrader, 2000) of the U.S. Department of Defense and the National Weather Service, and the Australian Tropical Cyclone Workstation (Woodcock, 1995), these systems are stand-alone. Web-based systems are mainly developed for the dissemination of

information or visualization of the historical TC database via the web, but they lack the facility for TC analysis and prediction. Therefore, the system in this study may fill the gap between a stand-alone professional TC platform and a web-based system for information dissemination. It provides useful references for researchers and operational personnel regarding everyday operations and forecasts, as well as meteorological information and knowledge that may be of interest to the general public.

9.3 Directions for Further Research

This study attempts to discover knowledge about TC dynamics through data mining. We investigate TC recurvature, landfall, and post-landfall problems, as well as the relationship between TC movement and intensity, and build a novel web-based platform for TC analysis and prediction. The current limitations and future directions are summarized as follows:

1. Since this study is based on existing TC theories and findings, the potential factors may change with the development of these theories. The existing theory on TC movement focuses mainly on a barotropic framework. The baroclinic processes of TC dynamics (e.g., thermodynamic processes) are still under investigation by meteorologists and are far less well understood. It has been pointed out that the study of TC intensity lags far behind that of TC movement. Therefore, this study simply marks the beginning of using data mining methods to discover knowledge from historical TC dynamics, and it will be further advanced by the development of the existing TC theories, data mining algorithms and dynamic models.

2. The study area is focused on WNP, including SCS. The Chinese coast is selected as the landfall location for the current research. As WNP and the Atlantic are both in the Northern Hemisphere and influenced by large-scale circulation, the research methods and findings in this study can also be applied to study the recurvature, landfall, and intensity of TCs in the Atlantic. Meanwhile, the research methods can also be applied to the Southern hemisphere (e.g., cyclones in Australia) by redefining the recurvature, and the indices of the subtropical high and monsoonal systems.

3. As the spatial resolution of NCEP/NCAR meteorological variables is 1°×1° only from 1999 to the present, and the time interval of the variable is 6 hours, the quality of the data variables may limit the results to some extent. The NCEP/NCAR reanalysis dataset from 1951 to the present has a spatial resolution of 2.5°×2.5°, which seems too coarse for studying TC problems. As the data quality is increasingly improving according to the advanced observation equipment and technology, this problem may be solved in the future.

4. The present system is a prototype that can be extended into a full-fledged system through which TC dynamics can be studied within the monsoonal system affecting the development of TCs (Elsberry, 2004; Harr and Chan, 2004; Liu and Chan, 2003). The data mining function of the system, in particular, can be extended to unravel the rules stipulating the interplay between TC and MJO (Madden and Julian, 1994), QBO (Ebdon, 1960) and ENSO, including El Nino and La Nina occurrence (Chan, 1985a; Chu, 2004; Wang and Chan, 2002). In addition, real-time TC prediction, which is of great interest to the general public, will be developed via collaboration with the Hong Kong Observatory.

5. Dynamic models (e.g., MM5 and WRF) have experienced rapid advancements in the last few decades and are now broadly used for TC tracks and intensity prediction. Given the knowledge (e.g., rules, structures and clusters) unraveled by the data mining methods and the development of the dynamic models, the data mining methods and the dynamic model will mutually enrich each other in a complementary, integrated manner. The findings of this study can be used to improve the dynamic model by fine tuning the parameters and verifying the model outputs.

6. This thesis focuses on how TC dynamics are influenced by large-scale circulation (e.g., the subtropical high, monsoons, and westerlies). TC internal dynamics and the interaction between TCs and large-scale circulation are excluded due to a lack of

observations inside TCs and our insufficient understanding of the mechanisms related to these processes. Further studies on these aspects are required, given the increasing spatial and temporal availability of observations coming from radar, satellites and other forms of reconnaissance.

7. ENSO exerts a significant impact on TC activity (e.g., TC intensity, genesis and track) in WNP. The ENSO indices will therefore be considered as potential variables that influence TC landfall and recurvature in the future research. El Nino Modoki (i.e., Central Pacific El Nino) (Ashok *et al.*, 2007; Kao and Yu, 2009; Kim *et al.*, 2009b; Weng *et al.*, 2007) has been studied by meteorologists in last few years. The modulation of El Nino Modoki on TC activity will be taken into account in further investigation.

Reference

Aamodt, A. and E. Plaza, 1994: Case-Based Reasoning - Foundational Issues, Methodological Variations, and System Approaches. *Ai Communications*, **7**, 39-59.

Aberson, S. and J. Franklin, 1999: Impact on hurricane track and intensity forecasts of GPS dropwindsonde observations from the first-season flights of the NOAA Gulfstream-IV jet aircraft. *Bulletin of the American Meteorological Society*, **80**, 421-427.

Aberson, S. D. and C. R. Sampson, 2003: On the predictability of tropical cyclone tracks in the Northwest Pacific basin. *Monthly Weather Review*, **131**, 1491-1497.

Adem, J. and P. Lezama, 1960: On the motion of a cyclone embedded in a uniform flow1. *Tellus*, **12**, 255-258.

Archambault, H., D. Keyser, and L. Bosart, 2009: The Extratropical Flow Response to Recurving Western North Pacific Tropical Cyclones. *Climate Prediction S&T Digest*, 1.

Ashok, K., S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, 2007: El Nino Modoki and its possible teleconnection. *J. Geophys. Res*, **112**, C11007.

Bao, C. L. and J. C. Sadler, 1983: The Speed of Recurving Typhoons over the Western North Pacific-Ocean. *Monthly Weather Review*, **111**, 1280-1292.

Bender, M., R. Tuleya, and Y. Kurihara, 1985: A numerical study of the effect of a mountain range on a landfalling tropical cyclone. *Monthly Weather Review*, **113**, 567-583.

Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The operational GFDL coupled hurricane-ocean prediction system and a summary of its performance. *Monthly Weather Review*, **135**, 3965-3989.

Berg, W., T. L'Ecuyer, and C. Kummerow, 2006: Rainfall climate regimes: The relationship of regional TRMM rainfall biases to the environment. *Journal of Applied Meteorology and Climatology*, **45**, 434-454.

Bhowmik, S., S. Kotal, and S. Kalsi, 2005a: An empirical model for predicting the decay of tropical cyclone wind speed after landfall over the Indian region. *Journal of Applied Meteorology*, **44**, 179-185.

Bhowmik, S. K. R., S. D. Kotal, and S. R. Kalsi, 2005b: An empirical model for predicting the decay of tropical cyclone wind speed after landfall over the Indian region. *Journal of Applied Meteorology*, **44**, 179-185.

Biernacki, C., G. Celeux, and G. Govaert, 2002: Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**, 719-725.

Bister, M. and K. Emanuel, 1997: The genesis of Hurricane Guillermo: TEXMEX analyses and a modeling study. *Monthly Weather Review*, **125**, 2662-2682.

Blackwell, K. G., J. Holmes, R. A. Wade, and S. K. Kimball, 2006: Collapsing precipitation cores in open-eyewall hurricanes at landfall. *Bulletin of the American Meteorological Society*, **87**, 1310-1311.

Bove, M., J. O'Brien, J. Eisner, C. Landsea, and X. Niu, 1998: Effect of El Ni o on US landfalling hurricanes, revisited. *Bulletin of the American Meteorological Society*, **79**, 2477-2482.

Brand, S. and J. W. Blelloch, 1973: Changes in the Characteristics of Typhoons Crossing the Philippines. *Journal of Applied Meteorology*, **12,** 104-109.

Brand, S. and J. W. Blelloch, 1974: Changes in the Characteristics of Typhoons Crossing the Island of Taiwan. *Monthly Weather Review*, **102,** 708-713.

Bray, J. H. and S. E. Maxwell, 1985: *Multivariate analysis of variance*. Sage Publications, Inc.

Breiman, L., J. Freidman, R. Olshen, and C. Stone, 1984a: *Classification and regression trees*. Chapman & Hall/CRC.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone, 1984b: Classification and regression trees. Wadsworth & Brooks. *Cole, Pacific Grove, California, USA*.

Brettschneider, B., 2008: Climatological hurricane landfall probability for the United States. *Journal of Applied Meteorology and Climatology*, **47,** 704-716.

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

Burpee, R., J. Franklin, S. Lord, R. Tuleya, and S. Aberson, 1996: The impact of Omega dropwindsondes on operational hurricane track forecast models. *Bulletin of the American Meteorological Society*, **77,** 925-933.

Burroughs, L. D. and S. Brand, 1973: Speed of Tropical Storms and Typhoons After Recurvature in the Western North Pacific Ocean. *Journal of Applied Meteorology*, **12,** 452-458.

Camargo, S., A. Robertson, S. Gaffney, and P. Smyth, 2004: Cluster analysis of western North Pacific tropical cyclone tracks. 250–251.

Camargo, S. J. and A. H. Sobel, 2005: Western North Pacific tropical cyclone intensity and ENSO. *Journal of Climate*, **18,** 2996-3006.

Camargo, S. J., A. W. Robertson, A. G. Barnston, and M. Ghil, 2008: Clustering of eastern North Pacific tropical cyclone tracks: ENSO and MJO effects. *Geochemistry Geophysics Geosystems*, **9,** -.

Camargo, S. J., A. W. Robertson, S. J. Gaffney, P. Smyth, and M. Ghil, 2007a: Cluster analysis of typhoon tracks. Part II: Large-scale circulation and ENSO. *Journal of Climate*, **20,** 3654-3676.

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

Cameron, J. and J. Allen, 2005: Climate change and Capical. *The finance of climate change: a guide for governments, corporations and innovators*, k. Tang, Ed., Risk Books, a Division of Incisive Financial Publishing Ltd

Carr, L. E. and R. L. Elsberry, 1990: Observational Evidence for Predictions of Tropical Cyclone Propagation Relative to Environmental Steering. *Journal of the Atmospheric Sciences*, **47,** 542-546.

Carr, L. E. and R. L. Elsberry, 1995: Monsoonal Interactions Leading to Sudden Tropical Cyclone Track Changes. *Monthly Weather Review*, **123,** 265-290.

Chan, J. C. L., 1984: An Observational Study of the Physical Processes Responsible for Tropical Cyclone Motion. *Journal of the Atmospheric Sciences*, **41,** 1036-1048.

Chan, J. C. L., 1985a: Tropical Cyclone Activity in the Northwest Pacific in Relation to the El-Nino Southern Oscillation Phenomenon. *Monthly Weather Review*, **113,** 599-606.

Chan, J. C. L., 1985b: Identification of the Steering Flow for Tropical Cyclone Motion from Objectively Analyzed Wind Fields. *Monthly Weather Review*, **113,** 106-116.
Chan, J. C. L., 1995a: Prediction of Annual Tropical Cyclone Activity over the Western North Pacific and the South China Sea. *International Journal of Climatology*, **15,** 1011-1019.
Chan, J. C. L., 1995b: Performance of Global and Regional Nwp Models in Their Prediction of Typhoon Nat (1991). *Weather and Forecasting*, **10,** 400-410.
Chan, J. C. L., 2000: Tropical cyclone activity over the western North Pacific associated with El Nino and La Nina events. *Journal of Climate*, **13,** 2960-2972.
Chan, J. C. L., 2005a: The physics of tropical cyclone motion. *Annual Review of Fluid Mechanics*, **37,** 99-128.
Chan, J. C. L., 2005b: Interannual and interdecadal variations of tropical cyclone activity over the western North Pacific. *Meteorology and Atmospheric Physics*, **89,** 143-152.
Chan, J. C. L., 2006: Comment on "Changes in tropical cyclone number, duration, and intensity in a warming environment". *Science*, **311,** -.
Chan, J. C. L., 2010: Movement of Tropical Cyclones. *Global Perspectives on Tropical Cyclones : From Science to Mitigation*, J. C. L. Chan and J. D. Kepert, Eds., World Scientific.
Chan, J. C. L. and W. M. Gray, 1982: Tropical Cyclone Movement and Surrounding Flow Relationships. *Monthly Weather Review*, **110,** 1354.
Chan, J. C. L. and R. T. Williams, 1987: Analytical and Numerical-Studies of the Beta-Effect in Tropical Cyclone Motion .1. Zero Mean Flow. *Journal of the Atmospheric Sciences*, **44,** 1257-1265.
Chan, J. C. L. and K. K. W. Cheung, 1995: Fourier components of the circulation associated with tropical cyclone motion. *21st Conference on Hurricanes and Tropical Meteorology*, 12-14
Chan, J. C. L. and X. D. Liang, 2003: Convective asymmetries associated with tropical cyclone landfall. Part I: f-plane simulations. *Journal of the Atmospheric Sciences*, **60,** 1560-1576.
Chan, J. C. L., W. M. Gray, and S. Q. Kidder, 1980: Forecasting Tropical Cyclone Turning Motion from Surrounding Wind and Temperature-Fields. *Monthly Weather Review*, **108,** 778-792.
Chan, J. C. L., F. M. F. Ko, and Y. M. Lei, 2002: Relationship between potential vorticity tendency and tropical cyclone motion. *Journal of the Atmospheric Sciences*, **59,** 1317-1336.
Chang, C.-P., Y. Zhang, and T. Li, 2000: Interannual and Interdecadal Variations of the East Asian Summer Monsoon and Tropical Pacific SSTs. Part I: Roles of the Subtropical Ridge. *Journal of Climate*, **13,** 4310-4325.
Chang, S. W. J., 1982: The Orographic Effects Induced by an Island Mountain Range on Propagating Tropical Cyclones. *Monthly Weather Review*, **110,** 1255-1270.
Charney, J. and A. Eliassen, 1964: On the Growth of the Hurricane Depression. *Journal of the Atmospheric Sciences*, **21,** 68-75.
Chen, G., 1999: The subtropical high. The Droughts and Floods in Summer in China and Background Fields. Z. Zhao, Ed., China Meteorological Press, 45-52.

Chen, G., 2009: Interdecadal variation of tropical cyclone activity in association with summer monsoon, sea surface temperature over the western North Pacific. *Chinese Science Bulletin*, **54,** 1417-1421.

Chen, L., Z. Lin, and Y. Li, 2004: Research advances on tropical cyclone landfall process. *Acta Meteorologica Sinica*, 05.

Chen, L. S. and Y. H. Ding, 1979: *Introduction to Tropical Cyclones in the Western Pacific*. Science Press.

Chen, P., M. P. Hoerling, and R. M. Dole, 2001a: The origin of the subtropical anticyclones. *Journal of the Atmospheric Sciences*, **58,** 1827-1835.

Chen, T. C. and J. H. Yoon, 2000: Interannual variation in Indochina summer monsoon rainfall: Possible mechanism. *Journal of Climate*, **13,** 1979-1986.

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. *Weather and Forecasting*, **24,** 646-666.

Chen, Y. and M. Yau, 2001: Spiral bands in a simulated hurricane. Part I: Vortex Rossby wave verification. *Journal of the Atmospheric Sciences*, **58,** 2128-2145.

Chen, Y., H. Zhang, R. Zhou, and H. Wu, 2001b: Relationship between the ground surface temperature in Asia and the intensity and location of subtropical high in the western Pacific. *Chinese Journal of Atmospheric Sciences*, **25,** 515-522.

Cheng, C., N. Hsu, and C. Wei, 2008: Decision-tree analysis on optimal release of reservoir storage under typhoon warnings. *Natural Hazards*, **44,** 65-84.

Cheung, N. K. W., 2006: The roles of ENSO on the occurrence of abruptly recurving tropical cyclones over the Western North Pacific Ocean Basin. *Adv. Geosci.*, **6,** 139-148.

Ching, S. E., J. C. L. Chan, and E. S. T. Lai, 2000: Tropical cyclone landfall along the South China Coast. *24th Conference on Hurricanes and Tropical Meteorology/10th Conference on Interaction of the Sea and Atmosphere*, **643,** 274-275

Choi, K., B. Kim, C. Choi, and J. Nam, 2009: Cluster analysis of Tropical Cyclones making landfall on the Korean Peninsula. *Advances in Atmospheric Sciences*, **26,** 202-210.

Choi, K. S. and B. J. Kim, 2007: Climatological Characteristics of Tropical Cyclones Making Landfall over the Korean Peninsula. *Asia-Pacific Journal of Atmospheric Sciences*, **43,** 97-109.

Choi, K. S., B. J. Kim, D. W. Kim, and H. R. Byun, 2010: Interdecadal variation of tropical cyclone making landfall over the Korean Peninsula. *International Journal of Climatology*, **30,** 1472-1483.

Chu, J., C. Sampson, A. Levine, and E. Fukada, 2002: The Joint Typhoon Warning Center tropical cyclone best-tracks, 1945–2000. *Naval Research Laboratory, Reference Number NRL/MR/7540-02-16.*

Chu, P. S., 2004: ENSO and tropical cyclone activity. *Hurricanes and Typhoons: Past, Present, and Potential*, R. J. Murnane and K.-B. Liu, Eds., Columbia Press, 297-332.

Craig, G. and S. Gray, 1996: CISK or WISHE as the mechanism for tropical cyclone intensification. *Journal of the Atmospheric Sciences*, **53,** 3528-3540.

DeMaria, M., 2009: A simplified dynamical system for tropical cyclone intensity prediction. *Monthly Weather Review*, **137,** 68-82.

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

DeMaria, M. and J. Kaplan, 1999: An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Weather and 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). *Weather and Forecasting*, **20,** 531-543.

Dempster, A., N. Laird, and D. Rubin, 1977: Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39,** 1-38.

Deng, A., N. L. Seaman, G. K. Hunter, and D. R. Stauffer, 2004: Evaluation of Interregional Transport Using the MM5-SCIPUFF System. *Journal of Applied Meteorology*, **43,** 1864-1886.

DeSarbo, W. and W. Cron, 1988: A maximum likelihood methodology for clusterwise linear regression. *Journal of classification*, **5,** 249-282.

Diercks, J. W. and R. A. Anthes, 1976: A Study of Spiral Bands in a Linear Model of a Cyclonic Vortex. *Journal of Atmospheric Sciences*, **33,** 1714-1729.

Ding, Y. H., 1992: Summer Monsoon Rainfalls in China. *Journal of the Meteorological Society of Japan*, **70,** 373-396.

Ding, Y. H., 2007: The variability of the Asian summer monsoon. *Journal of the Meteorological Society of Japan*, **85B,** 21-54.

Ding, Y. H. and J. C. L. Chan, 2005: The East Asian summer monsoon: an overview. *Meteorology and Atmospheric Physics*, **89,** 117-142.

Dobos, P. H. and R. L. Elsberry, 1993: Forecasting Tropical Cyclone Recurvature .1. Evaluation of Existing Methods. *Monthly Weather Review*, **121,** 1273-1278.

Dong, M., L. Chen, Y. Li, and C. Lu, 2010: Rainfall Reinforcement Associated with Landfalling Tropical Cyclones. *Journal of the Atmospheric Sciences*, **67,** 3541-3558.

Donnelly, J., S. Roll, M. Wengren, J. Butler, R. Lederer, and T. Webb, 2001: Sedimentary evidence of intense hurricane strikes from New Jersey. *Geology*, **29,** 615.

Ebdon, R. A., 1960: Notes on the wind flow at 50 mb in tropical and sub-tropical regions in January 1957 and January 1958. *Quarterly Journal of the Royal Meteorological Society*, **86,** 540-542.

Efron, B., 1969: Student's t-test under symmetry conditions. *Journal of the American Statistical Association*, **64,** 1278-1302.

Elsberry, R., 1987: *A global view of tropical cyclones*. University of Chicago Press.

Elsberry, R., 2004: Monsoon-related tropical cyclones in East Asia. *East Asian Monsoon*, 463–498.

Elsberry, R. and N. P. School, 1987: *A global view of tropical cyclones*. University of Chicago Press.

Elsberry, R. L., 1990: International Experiments to Study Tropical Cyclones in the Western North Pacific. *Bulletin of the American Meteorological Society*, **71,** 1305-1316.

Elsberry, R. L., T. D. B. Lambert, and M. A. Boothe, 2007: Accuracy of Atlantic and Eastern North Pacific Tropical Cyclone Intensity Forecast Guidance. *Weather and Forecasting*, **22,** 747-762.

Elsner, J. B., 2003: Tracking hurricanes. *Bulletin of the American Meteorological Society*, **84,** 353-356.

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

Emanuel, K., 1986: An air-sea interaction theory for tropical cyclones. Part I: Steady-state maintenance. *Journal of the Atmospheric Sciences*, **43,** 585-604.

Emanuel, K., 1988: The maximum intensity of hurricanes. *Journal of the Atmospheric Sciences*, **45,** 1143-1155.

Emanuel, K., 1991: The theory of hurricanes. *Annual Review of Fluid Mechanics*, **23,** 179-196.

Emanuel, K., 1997: Some aspects of hurricane inner-core dynamics and energetics. *Journal of the Atmospheric Sciences*, **54,** 1014-1026.

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

Emanuel, K., C. DesAutels, C. Holloway, and R. Korty, 2004: Environmental Control of Tropical Cyclone Intensity. *Journal of the Atmospheric Sciences*, **61,** 843-858.

Evans, J. L. and K. McKinley, 1998: Relative timing of tropical storm lifetime maximum intensity and track recurvature. *Meteorology and Atmospheric Physics*, **65,** 241-245.

Evans, J. L., G. J. Holland, and R. L. Elsberry, 1991: Interactions between a Barotropic Vortex and an Idealized Subtropical Ridge. Part I: Vortex Motion. *Journal of the Atmospheric Sciences*, **48,** 301-314.

Everitt, B., 1988: A finite mixture model for the clustering of mixed-mode data. *Statistics & Probability Letters*, **6,** 305-309.

Everitt, B. and D. Hand, 1981: *Finite mixture distributions*. Chapman & Hall.

Fayyad, U. and K. Irani, 1992: The attribute selection problem in decision tree generation. JOHN WILEY & SONS LTD, 104-104.

Fayyad, U. and P. Stolorz, 1997: Data mining and KDD: Promise and challenges. *Future Generation Computer Systems*, **13,** 99-115.

Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth, 1996: From data mining to knowledge discovery in databases. *AI magazine*, **17,** 37.

Fiorino, M. and R. L. Elsberry, 1989: Some aspects of vortex structure related to tropical cyclone motion. *Journal of Atmospheric Sciences*, **46,** 975-990.

Fitzpatrick, M. E., 1992: Tropical Cyclone Motion and Recurvature in TCM-90, Colorado State University, 93.

Fitzpatrick, P. J., 1997: Understanding and forecasting tropical cyclone intensity change with the Typhoon Intensity Prediction Scheme (TIPS). *Weather and Forecasting*, **12,** 826-846.

Flatau, M., W. H. Schubert, and D. E. Stevens, 1994: The role of baroclinic processes in tropical cyclone motion: The influence of vertical tilt. *Journal of the Atmospheric Sciences*, **51,** 2589-2601.

Ford, D. M., R. L. Elsberry, P. A. Harr, and P. H. Dobos, 1993: Forecasting Tropical Cyclone Recurvature .2. An Objective Technique Using an Empirical Orthogonal Function Representation of Vorticity Fields. *Monthly Weather Review*, **121,** 1279-1290.

Friedl, M. and C. Brodley, 1997: Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, **61,** 399-409.

Fudeyasu, H., S. Iizuka, and T. Matsuura, 2006: Impact of ENSO on landfall characteristics of tropical cyclones over the western North Pacific during the summer monsoon season. *Geophysical Research Letters*, **33,** -.

Fujibe, F. and N. Kitabatake, 2007: Classification of surface wind fields of tropical cyclones at landfall on the Japan main islands. *Journal of the Meteorological Society of Japan*, **85,** 747-765.

Gaffney, S. and P. Smyth, 1999: Trajectory clustering with mixtures of regression models. *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM.

Gaffney, S., A. Robertson, P. Smyth, S. Camargo, and M. Ghil, 2007: Probabilistic clustering of extratropical cyclones using regression mixture models. *Climate Dynamics*, **29,** 423-440.

Gaffney, S. J., 2004: Probabilistic curve-aligned clustering and prediction with regression mixture models, University of California, Irvine, 281.

Genesereth, M., 2010: Data Integration: The Relational Logic Approach. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, **4,** 1-97.

Gentemann, C., F. Wentz, C. Mears, and D. Smith, 2004: In situ validation of Tropical Rainfall Measuring Mission microwave sea surface temperatures. *Journal of Geophysical Research*, **109,** C04021.

Gentry, M. S. and G. M. Lackmann, 2010: Sensitivity of simulated tropical cyclone structure and intensity to horizontal resolution. *Monthly Weather Review*, **138,** 688-704.

George, J. E. and W. M. Gray, 1977: Tropical Cyclone Recurvature and Nonrecurvature as Related to Surrounding Wind-Height Fields. *Journal of Applied Meteorology*, **16,** 34-42.

Girden, E. R., 1992: *ANOVA: Repeated measures*. Sage Publications Newbury Park, CA.

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. *Weather and Forecasting*, **25,** 587-593.

Goldman, J. L. and T. Ushijima, 1971: Intensity Changes in Hurricanes after Landfall. *Bulletin of the American Meteorological Society*, **52,** 790-&.

Guard, C., 1977: Operational Application of a Tropical Cyclone Recurvature/Non-Recurvature Study Based on 200 MB Wind Fields. Fleet Weather Central/Joint Typhoon Warning Center Fpo San Francisco 96630.

Hampson, S. and D. Volper, 1986: Linear function neurons: Structure and training. *Biological Cybernetics*, **53,** 203-217.

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*. Morgan Kaufmann.

Harr, A. and J. Chan, 2004: Monsoon impacts on tropical cyclone variability. *Review Topic B3e: Tropical cyclones. Naval Postgraduate School Press.*

Harr, P. and R. Elsberry, 2000: Extratropical transition of tropical cyclones over the western North Pacific. Part I: Evolution of structural characteristics during the transition process. *Monthly Weather Review*, **128,** 2613-2633.

Harr, P., R. Elsberry, and T. Hogan, 2000: Extratropical transition of tropical cyclones over the western North Pacific. Part II: The impact of midlatitude circulation characteristics. *Monthly Weather Review*, **128,** 2634-2653.

Harr, P. A. and R. L. Elsberry, 1991: Tropical Cyclone Track Characteristics as a Function of Large-Scale Circulation Anomalies. *Monthly Weather Review*, **119,** 1448-1468.

Harr, P. A. and R. L. Elsberry, 1995a: Large-Scale Circulation Variability over the Tropical Western North Pacific .2. Persistence and Transition Characteristics. *Monthly Weather Review*, **123,** 1247-1268.

Harr, P. A. and R. L. Elsberry, 1995b: Large-Scale Circulation Variability over the Tropical Western North Pacific .1. Spatial Patterns and Tropical Cyclone Characteristics. *Monthly Weather Review*, **123,** 1225-1246.

Harr, P. A., R. L. Elsberry, and J. C. L. Chan, 1996: Transformation of a Large Monsoon Depression to a Tropical Storm during TCM-93. *Monthly Weather Review*, **124,** 2625-2643.

Harris, R. J., 1994: *ANOVA: An analysis of variance primer*. FE Peacock Publishers.

Heming, J. and J. Goerss, 2010: Track and Structure Forecasts of Tropical Cyclones. *Global Perspectives on Tropical Cyclones : From Science to Mitigation*, J. Heming and J. Goerss, Eds., World Scientific.

Heming, J., J. Chan, and A. Radford, 1995: A new scheme for the initialisation of tropical cyclones in the UK Meteorological Office global model. *Meteorological Applications*, **2,** 171-184.

Heymsfield, G. M., J. B. Halverson, J. Simpson, L. Tian, and T. P. Bui, 2001: ER-2 Doppler radar investigations of the eyewall of Hurricane Bonnie during the convection and moisture experiment-3. *Journal of Applied Meteorology*, **40,** 1310-1330.

Hill, E. and W. Malkin, 1965: Weather Note: Recurvature of Hurricane Cleo, 1964, and Associated 500-MB. Streamline Analysis. *Monthly Weather Review*, **93,** 565.

Ho, C. H., J. J. Baik, J. H. Kim, D. Y. Gong, and C. H. Sui, 2004: Interdecadal changes in summertime typhoon tracks. *Journal of Climate*, **17,** 1767-1776.

Hodanish, S. and W. M. Gray, 1993: An Observational Analysis of Tropical Cyclone Recurvature. *Monthly Weather Review*, **121,** 2665-2689.

Hodyss, D. and E. Hendricks, 2010: The Resonant Excitation of Baroclinic Waves by the Divergent Circulation of Recurving Tropical Cyclones. *Journal of the Atmospheric Sciences*, **67,** 3600-3616.

Holland, G., 1993a: Global guide to tropical cyclone forecasting. *WMO, Geneva.* _.

Holland, G. J., 1983a: Tropical Cyclone Motion - Environmental Interaction Plus a Beta-Effect. *Journal of the Atmospheric Sciences*, **40,** 328-342.

Holland, G. J., 1983b: Tropical Cyclone Motion: Environmental Interaction Plus a Beta Effect. *Journal of the Atmospheric Sciences*, **40,** 328-342.
Holland, G. J., 1984: Tropical Cyclone Motion - a Comparison of Theory and Observation. *Journal of the Atmospheric Sciences*, **41,** 68-75.
Holland, G. J., 1993b: Tropical Cyclone Motion. *global guide to tropical cyclone forecasting*, G. Holland, Ed., WMO.
Holland, G. J. and M. Lander, 1993: The Meandering Nature of Tropical Cyclone Tracks. *Journal of the Atmospheric Sciences*, **50,** 1254-1266.
Holland, G. J. and G. S. Dietachmayer, 1993: On the interaction of tropical cyclone scale vortices. III: Continuous barotropic vortices. *Quarterly Journal of the Royal Meteorological Society*, **119,** 1381-1398.
Holland, G. J. and Y. Q. Wang, 1995: Baroclinic Dynamics of Simulated Tropical Cyclone Recurvature. *Journal of the Atmospheric Sciences*, **52,** 410-426.
Holland, G. J., Y. Q. Wany, and M. Lander, 1995: The Meandering Nature of Tropical Cyclone Tracks - Reply. *Journal of the Atmospheric Sciences*, **52,** 289-290.
Holliday, C. R. and A. H. Thompson, 1979: Climatological characteristics of rapidly intensifying typhoons. *Monthly Weather Review*, **107,** 1022.
Hong, X., S. Chang, S. Raman, L. Shay, and R. Hodur, 2000: The interaction between Hurricane Opal (1995) and a warm core ring in the Gulf of Mexico. *Monthly Weather Review*, **128,** 1347-1365.
Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, **8,** 38-55.
Jarvinen, B. and C. Neumann, 1979: Statistical forecasts of tropical cyclone intensity for the North Atlantic basin. *NWS NHC-10, NOAA Tech. Memo*, 22.
Jiang, H. Y. and J. B. Halverson, 2008: On the differences in storm rainfall from Hurricanes Isidore and Lili. Part II: Water budget. *Weather and Forecasting*, **23,** 44-61.
Joint Typhoon Warning Center (JTWC), 1988: *Annual tropical cyclone report*, 216 pp.
Juneng, L., F. T. Tangang, C. J. C. Reason, S. Moten, and W. A. W. Hassan, 2007: Simulation of tropical cyclone Vamei (2001) using the PSU/NCAR MM5 model. *Meteorology and Atmospheric Physics*, **97,** 273-290.
Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, and J. Woollen, 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77,** 437-471.
Kantardzic, M., 2003: *Data mining: concepts, models, methods, and algorithms*. Wiley-IEEE Press.
Kao, H. Y. and J. Y. Yu, 2009: Contrasting eastern-Pacific and central-Pacific types of ENSO. *Journal of Climate*, **22,** 615-632.
Kaplan, J. and M. Demaria, 1995: A Simple Empirical-Model for Predicting the Decay of Tropical Cyclone Winds after Landfall. *Journal of Applied Meteorology*, **34,** 2499-2512.
Kaplan, J. and M. DeMaria, 2001: On the decay of tropical cyclone winds after landfall in the New England area. *Journal of Applied Meteorology*, **40,** 280-286.

Kaplan, J., M. DeMaria, and J. A. Knaff, 2010: A Revised Tropical Cyclone Rapid Intensification Index for the Atlantic and Eastern North Pacific Basins. *Weather and Forecasting*, **25,** 220-241.

Kasahara, A., 1959: A Comparison Between Geostrophic and Non-Geostrophic Numerical Forecasts of Hurricane Movement with the Barotropic Steering Model. *J. Meteor.*, **16,** 371-384.

Kepert, J. D., 2010: Track and Structure Forecasts of Tropical Cyclones. *Global Perspectives on Tropical Cyclones : From Science to Mitigation*, J. Heming and J. Goerss, Eds., World Scientific.

Kim, D. K., K. R. Knupp, and C. R. Williams, 2009a: Airflow and Precipitation Properties within the Stratiform Region of Tropical Storm Gabrielle during Landfall. *Monthly Weather Review*, **137,** 1954-1971.

Kim, H. M. and B. J. Jung, 2009: Singular Vector Structure and Evolution of a Recurving Tropical Cyclone. *Monthly Weather Review*, **137,** 505-524.

Kim, H. M., P. J. Webster, and J. A. Curry, 2009b: Impact of shifting patterns of Pacific Ocean warming on North Atlantic tropical cyclones. *Science*, **325,** 77.

Kim, J. and F. J. Kohout, 1975: Analysis of variance and covariance: subprograms ANOVA and ONEWAY. *Statistical package for the social sciences*, 398-433.

Knaff, J. A., 2009: Revisiting the maximum intensity of recurving tropical cyclones. *International Journal of Climatology*, **29,** 827-837.

Knaff, J. A., C. R. Sampson, and M. DeMaria, 2005: An Operational Statistical Typhoon Intensity Prediction Scheme for the Western North Pacific. *Weather and Forecasting*, **20,** 688-699.

Knutson, T. R., J. L. McBride, J. Chan, K. Emanuel, G. Holland, C. Landsea, I. Held, J. P. Kossin, A. K. Srivastava, and M. Sugi, 2010: Tropical cyclones and climate change. *Nature Geoscience*, **3,** 157-163.

Ko, F. M. F. and J. C. L. Chan, 2000: Non-barotropic processes in tropical cyclone motion Part III: Potential vorticity advection. *24th Conference on Hurricanes and Tropical Meteorology/10th Conference on Interaction of the Sea and Atmosphere*, **643,** 106-107

Ko, K. and H. Hsu, 2006: Sub-Monthly Circulation Features Associated with Tropical Cyclone Tracks over the East Asian Monsoon Area during July-August Season. *J Meteorol*, **84,** 871-889.

Ko, K. and H. Hsu, 2009: ISO modulation on the submonthly wave pattern and recurving tropical cyclones in the tropical western North Pacific. *Journal of Climate*, **22,** 582-599.

Kolodner, J., 1993: *Case-Based Reasoning*. Morgan Kaufmann.

Kong, Z. L., D. F. Wang, J. S. Pan, S. F. Hao, Z. H. Zhang, and Y. Wu, 2008: The Development and Application of Marine System Based on GIS. *Itess: 2008 Proceedings of Information Technology and Environmental System Sciences, Pt 1*, **1259,** 58-63

Koperski, K. and J. Han, 1995: Discovery of spatial association rules in geographic information databases. Springer, 47-66.

Koperski, K., J. Adhikary, and J. Han, 1996: Spatial Data Mining: Progress and Challenges-Survey paper. Citeseer.

Kossin, J. P., B. D. McNoldy, and W. H. Schubert, 2002: Vortical swirls in hurricane eye clouds. *Monthly Weather Review*, **130,** 3144-3149.

Kossin, J. P., S. J. Camargo, and M. Sitkowski, 2010: Climate Modulation of North Atlantic Hurricane Tracks. *Journal of Climate*, **23,** 3057-3076.

Krishnamurti, T. and D. Subrahmanyam, 1982: The 30-50 Day Mode at 850 mb During MONEX. *Journal of Atmospheric Sciences*, **39,** 2088-2095.

Krishnamurti, T. N., H. S. Bedi, K. S. Yap, D. Oosterhof, and G. Rohaly, 1992: Recurvature Dynamics of a Typhoon. *Meteorology and Atmospheric Physics*, **50,** 105-126.

Kumar, K. V., A. Bhattacharya, and C. Subramanyam, 1998: Coastal morphological influence for tropical cyclone track deviation along Andhra coast: GIS and remote sensing-based approach. *Current Science*, 955-958.

Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The tropical rainfall measuring mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, **15,** 809-817.

Kummerow, C., J. Simpson, O. Thiele, W. Barnes, A. Chang, E. Stocker, R. Adler, A. Hou, R. Kakar, and F. Wentz, 2000: The status of the Tropical Rainfall Measuring Mission (TRMM) after two years in orbit. *Journal of Applied Meteorology*, **39,** 1965-1982.

Kuo, H. C., R. Williams, and J. H. Chen, 1999: A possible mechanism for the eye rotation of Typhoon Herb. *Journal of the Atmospheric Sciences*, **56,** 1659-1673.

Kuo, H. L., 1965: On Formation and Intensification of Tropical Cyclones Through Latent Heat Release by Cumulus Convection. *Journal of the Atmospheric Sciences*, **22,** 40-63.

Kurbalija, V., M. Ivanovic, and Z. Budimac, 2009: Case-based curve behaviour prediction. *Software-Practice & Experience*, **39,** 81-103.

Kurihara, Y., 1976: On the development of spiral bands in a tropical cyclone. *Journal of Atmospheric Sciences*, **33,** 940-958.

Lage, 1982: Forecasting tropical cyclone recurvature using an empirical orthogonal function representation of the synoptic forcing, Department of Meteorology, Naval Postgraduate School, 77.

Lander, M. A., 1994: Description of a Monsoon Gyre and Its Effects on the Tropical Cyclones in the Western North Pacific during August 1991. *Weather and Forecasting*, **9,** 640-654.

Lander, M. A., 1996: Specific Tropical Cyclone Track Types and Unusual Tropical Cyclone Motions Associated with a Reverse-Oriented Monsoon Trough in the Western North Pacific. *Weather and Forecasting*, **11,** 170-186.

Larson, J., Y. Zhou, and R. Higgins, 2005: Characteristics of landfalling tropical cyclones in the United States and Mexico: Climatology and interannual variability. *Journal of Climate*, **18,** 1247-1262.

Lau, K., G. Yang, and S. Shen, 1988: Seasonal and intraseasonal climatology of summer monsoon rainfall over East Asia. *Monthly Weather Review*, **116,** 18-37.

Lavrac, N., S. Džeroski, and U. Fayyad, 1997: Knowledge discovery in databases: An overview. *Inductive Logic Programming*, Springer Berlin / Heidelberg, 1-16.

Lee, J., J. Han, and K. Whang, 2007: Trajectory clustering: a partition-and-group framework. ACM, 593-604.

Leftwich, P. W., 1980: An Analysis of Predictions of Recurvature of Atlantic Tropical Cyclones. *Bulletin of the American Meteorological Society*, **61,** 1126-1126.

Lehmiller, G., T. Kimberlain, and J. Elsner, 1997: Seasonal prediction models for North Atlantic basin hurricane location. *Monthly Weather Review*, **125,** 1780-1791.

Lehodey, P., M. Bertignac, J. Hampton, A. Lewis, and J. Picaut, 1997: El Nino Southern Oscillation and tuna in the western Pacific. *Nature*, **389,** 715-718.

Lenzerini, M., 2002: Data integration: a theoretical perspective. *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, ACM.

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

Leung, Y., J. H. Ma, and W. X. Zhang, 2002: A new method for mining regression classes in large data sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23,** 5-21.

Leung, Y., J. C. Luo, J. H. Ma, and D. P. Ming, 2006: A New Method for Feature Mining in Remotely Sensed Images. *Geoinformatica*, **10,** 295-312.

Leung, Y., M. H. Wong, K. C. Wong, W. Zhang, and K.-S. Leung, 2011: A novel web-based system for tropical cyclone analysis and prediction. *International Journal of Geographical Information Science (in press)*.

Lewis, B. M. and H. F. Hawkins, 1982: Polygonal Eye Walls and Rainbands in Hurricanes. *Bulletin of the American Meteorological Society*, **63,** 1294-1301.

Li, W. W., C. W. Yang, and D. L. Sun, 2009: Mining geophysical parameters through decision-tree analysis to determine correlation with tropical cyclone development. *Computers & Geosciences*, **35,** 309-316.

Li, X. F. and B. Wang, 1994: Barotropic Dynamics of the Beta-Gyres and Beta-Drift. *Journal of the Atmospheric Sciences*, **51,** 746-756.

Li, Y., L. Chen, and X. Xu, 2004a: The characteristics of sub-synoptic scale circulation of tropical cyclones sustaining overland. *Acta Meteor Sinica*, **62,** 257-268.

Li, Y., L. Chen, and J. Wu, 2004b: The diagnostics analysis on the characteristics of large scale circulation corresponding to the sustaining and decaying of tropical cyclone after it's landfall. *Acta Meteor Sinica*, **62,** 167-179.

Li, Y., L. Chen, and S. Zhang, 2004c: Statistical characteristics of tropical cyclone making landfalls on China [J]. *Journal of Tropical Meteorology*, **1**.

Li, Y. S. and J. C. L. Chan, 1999: Momentum Transports Associated with Tropical Cyclone Recurvature. *Monthly Weather Review*, **127,** 1021-1037.

Liu, K. and M. Fearn, 2000: Reconstruction of prehistoric landfall frequencies of catastrophic hurricanes in northwestern Florida from lake sediment records. *Quaternary Research*, **54,** 238-245.

Liu, K., C. Shen, and K. Louie, 2001: A 1000-year history of typhoon landfalls in Guangdong, southern China, reconstructed from Chinese historical documentary records. *Annals of the Association of American Geographers*, **91,** 453-464.

Liu, K. S. and J. C. L. Chan, 2003: Climatological characteristics and seasonal forecasting of tropical cyclones making landfall along the South China coast. *Monthly Weather Review*, **131,** 1650-1662.

Liu, K. S. and J. C. L. Chan, 2008: Interdecadal variability of western North Pacific tropical cyclone tracks. *Journal of Climate*, **21,** 4464-4476.

Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17,** 321-333.

Lorenz, E. N., 1969: Three approaches to atmospheric predictability. *Bull. Amer. Meteor. Soc*, **50,** 345?49.

Louie, K. and K. Liu, 2003: Earliest historical records of typhoons in China. *Journal of Historical Geography*, **29,** 299-316.

Lu, R., 2001: Interannual Variability of the Summertime North Pacific Subtropical High and its Relation to Atmospheric Convection over the Warm Pool. *Journal of the Meteorological Society of Japan*, **79,** 771-783.

Lu, R. and B. Dong, 2001: Westward Extension of North Pacific Subtropical High in Summer. *Journal of the Meteorological Society of Japan*, **79,** 1229-1241.

Lu, X. Y., X. Z. Zhang, and J. N. Chen, 2007: The relationship between the interdecadal variability of East Asian summer monsoon's movement and the spatial distribution pattern of the summer rainfall in East China *Atmospheric and Environmental Remote Sensing Data Processing and Utilization Iii: Readiness for Geoss*, **6684,** R6840-R6840.

Lyons, S. W., 2004: US tropical cyclone landfall variability: 1950-2002. *Weather and Forecasting*, **19,** 473-480.

MacAfee, A., 1997: Canadian Hurricane Centre Forecaster's Workstation. *22nd Conf. on Hurricanes and Tropical Meteorology*, Fort Collins, CO, Amer. Meteor. Soc., 137-138.

Macdonald, N. J., 1968: The evidence for the existence of Rossby like waves in the hurricane vortex. *Tellus*, **20,** 138-150.

MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. California, USA, 14.

Madden, R. A. and P. R. Julian, 1971: Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific. *Journal of the Atmospheric Sciences*, **28,** 702-708.

Madden, R. A. and P. R. Julian, 1972: Description of Global-Scale Circulation Cells in the Tropics with a 40-50 Day Period. *Journal of the Atmospheric Sciences*, **29,** 1109-1123.

Madden, R. A. and P. R. Julian, 1994: Observations of the 40-50-Day Tropical Oscillation - a Review. *Monthly Weather Review*, **122,** 814-837.

Malkus, J. and H. Riehl, 1960: On the dynamics and energy transformations in steady-state hurricanes. *Tellus*, **12,** 1-20.

Mandal, M., U. C. Mohanty, P. Sinha, and M. M. Ali, 2007: Impact of sea surface temperature in modulating movement and intensity of tropical cyclones. *Natural Hazards*, **41,** 413-427.

Mao, J. and G. Wu, 2008: Influences of Typhoon Chanchu on the 2006 South China Sea summer monsoon onset. *Geophys. Res. Lett.*, **35.**

Marks, F. and L. Shay, 1998: Landfalling tropical cyclones: Forecast problems and associated research opportunities. *Bulletin of the American Meteorological Society*, **79,** 305-323.

Marks, F. D., P. G. Black, M. T. Montgomery, and R. W. Burpee, 2008: Structure of the eye and eyewall of Hurricane Hugo (1989). *Monthly Weather Review*, **136,** 1237-1259.

Masser, I., H. Campbell, and M. Craglia, 1996: *GIS diffusion: The adoption and use of geographical information systems in local government in Europe*. CRC Press.
McBride, J. L. and G. J. Holland, 1987: Tropical-Cyclone Forecasting: A Worldwide Summary of Techniques and Verification Statistics. *Bulletin of the American Meteorological Society*, **68,** 1230-1238.
McLachlan, G. and K. Basford, 1988: Mixture Models: Inference and Applications to Clustering, Marcel Dekker. *New York*.
McLachlan, G. and T. Krishnan, 1997: The EM Algorithm and Extensions. Wiley, New York.
McLachlan, G. and D. Peel, 2000: *Finite mixture models*. Wiley-Interscience.
McLachlan, G. J., R. Bean, and D. Peel, 2002: A mixture model-based approach to the clustering of microarray expression data. *Bioinformatics*, **18,** 413.
Medlin, J. M., S. K. Kimball, and K. G. Blackwell, 2007: Radar and rain gauge analysis of the extreme rainfall during hurricane Danny's (1997) landfall. *Monthly Weather Review*, **135,** 1869-1888.
Miller, H. and J. Han, 2009: Geographic data mining and knowledge discovery-an overview. *Geographic data mining and knowledge discovery*, CRC Press, Taylor and Francis Group, 1-26.
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.
Mohanty, U. C. and A. Gupta, 2008: Deterministic Methods for Prediction of Tropical Cyclone Tracks. *Modelling and Monitoring of Coastal Marine Processes*, 141-170.
Montgomery, M. T. and C. Lu, 1997: Free waves on barotropic vortices. Part I: Eigenmode structure. *Journal of the Atmospheric Sciences*, **54,** 1868-1885.
Montgomery, M. T., V. A. Vladimirov, and P. V. Denissenko, 2002: An experimental study on hurricane mesovortices. *Journal of Fluid Mechanics*, **471,** 1-32.
Nakamura, J., U. Lall, Y. Kushnir, and S. J. Camargo, 2009: Classifying North Atlantic Tropical Cyclone Tracks by Mass Moments*. *Journal of Climate*, **22,** 5481-5494.
Naumann, F., U. Leser, and J. C. Freytag, 1999: Quality-driven integration of heterogeneous information systems. Citeseer, 447-458.
Neter, J., W. Wasserman, and M. H. Kutner, 1990: *Applied linear statistical models: regression, analysis of variance, and experimental designs*. Irwin Homewood, IL.
Neumann, C. J. and J. R. Hope, 1972: Performance Analysis of the HURRAN Tropical Cyclone Forecast System. *Monthly Weather Review*, **100,** 245-255.
Ngan, K. W. and J. C. L. Chan, 1995: Tropical cyclone motion - Steering vs propagation. *21st Conference on Hurricanes and Tropical Meteorology*, 23-25.
O'Brien, J., T. Richards, and A. Davis, 1996: The effect of El Ni o on US landfalling hurricanes. *Bulletin of the American Meteorological Society*, **77,** 773-774.
O'Shay, A. J. and T. N. Krishnamurti, 2004: An examination of a model's components during tropical cyclone recurvature. *Monthly Weather Review*, **132,** 1143-1166.
Ooyama, K., 1969: Numerical Simulation of Life Cycle of Tropical Cyclones. *Journal of the Atmospheric Sciences*, **26,** 3-&.
Pearson, K., 1894: III. *Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London, A*, **185,** 71-110.

Pielke Jr, R., J. Gratz, C. Landsea, D. Collins, M. Saunders, and R. Musulin, 2008: Normalized hurricane damage in the United States: 1900–2005. *Natural Hazards Review*, **9,** 29.

Powell, M., D. Bowman, D. Gilhousen, S. Murillo, N. Carrasco, and R. Fleur, 2004: Tropical cyclone winds at landfall. *BULLETIN-AMERICAN METEOROLOGICAL SOCIETY*, **85,** 845-852.

Powell, M. D., 1982: The transition of the hurricane frederic boundary-layer wind-field from the open gulf of mexico to landfall. *Monthly Weather Review*, **110,** 1912-1932.

Powell, M. D. and S. H. Houston, 1998: Surface wind fields of 1995 Hurricanes Erin, Opal, Luis, Marilyn, and Roxanne at landfall. *Monthly Weather Review*, **126,** 1259-1273.

Powell, M. D., P. P. Dodge, and M. L. Black, 1991: The landfall of hurricane hugo in the carolina-surface wind distribution. *Weather and Forecasting*, **6,** 379-399.

Quinlan, J., 1987: Decision trees as probabilistic classifiers. Morgan Kaufmann Publishers, 31.

Quinlan, J., 1993: *C4. 5: programs for machine learning*. Morgan Kaufmann.

Ramsay, H. A. and L. M. Leslie, 2008: The Effects of Complex Terrain on Severe Landfalling Tropical Cyclone Larry (2006) over Northeast Australia. *Monthly Weather Review*, **136,** 4334-4354.

Raymond, D. and H. Jiang, 1990: A theory for long-lived mesoscale convective systems. *Journal of the Atmospheric Sciences*, **47,** 3067-3077.

Reading, A., 1990: Caribbean tropical storm activity over the past four centuries. *International Journal of Climatology*, **10,** 365-376.

Reasor, P. D., M. T. Montgomery, F. D. Marks Jr, and J. F. Gamache, 2000: Low-wavenumber structure and evolution of the hurricane inner core observed by airborne dual-Doppler radar. *Monthly Weather Review*, **128,** 1653-1680.

Reed, R. J., W. J. Campbell, L. A. Rasmussen, and D. G. Rogers, 1961: Evidence of a Downward-Propagating, Annual Wind Reversal in the Equatorial Stratosphere. *J. Geophys. Res.*, **66,** 813-818.

Rennick, M., 1999: Performance of the Navy's tropical cyclone prediction model in the western North Pacific basin during 1996. *Weather and Forecasting*, **14,** 297-305.

Reynolds, C., M. Peng, and J. Chen, 2009: Recurving tropical cyclones: Singular vector sensitivity and downstream impacts. *Monthly Weather Review*, **137,** 1320-1337.

Ricciardulli, L. and F. J. Wentz, 2004: Uncertainties in sea surface temperature retrievals from space: Comparison of microwave and infrared observations from TRMM. *J. Geophys. Res*, **109**.

Riehl, H., 1972: Intensity of Recurved Typhoons. *Journal of Applied Meteorology*, **11,** 613-615.

Riehl, H. and R. J. Shafer, 1944: The recurvature of tropical storms. *Journal of the Atmospheric Sciences*, **1,** 42-54.

Roddick, J. and M. Spiliopoulou, 1999: A bibliography of temporal, spatial and spatio-temporal data mining research. *ACM SIGKDD Explorations Newsletter*, **1,** 34-38.

Roddick, J., K. Hornsby, and M. Spiliopoulou, 2001: An updated bibliography of temporal, spatial, and spatio-temporal data mining research. *Temporal, Spatial, and Spatio-Temporal Data Mining*, 147-163.

Rodgers, E., W. Olson, J. Halverson, J. Simpson, and H. Pierce, 2000: Environmental forcing of supertyphoon Paka's (1997) latent heat structure. *Journal of Applied Meteorology*, **39,** 1983-2006.

Rossby, C., 1939: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. *J. Mar. Res*, **2,** 38–55.

Rossby, C., 1948: On displacements and intensity changes of atmospheric vortices. *J. Mar. Res*, **7,** 71.

Sampson, C. R. and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting System (Version 3.2). *Bulletin of the American Meteorological Society*, **81,** 1231-1240.

Saunders, M. A., R. E. Chandler, C. J. Merchant, and F. P. Roberts, 2000: Atlantic hurricanes and NW Pacific typhoons: ENSO spatial impacts on occurrence and landfall. *Geophys. Res. Lett.*, **27,** 1147-1150.

Schumacher, R. S., T. J. Galarneau, and L. F. Bosart, 2010: Distant Effects of a Recurving Tropical Cyclone on Rainfall in a Midlatitude Convective System: A High-Impact Predecessor Rain Event*. *Monthly Weather Review*, **139,** 650-667.

Shapiro, L. J., 1992: Hurricane Vortex Motion and Evolution in a 3-Layer Model. *Journal of the Atmospheric Sciences*, **49,** 140-153.

Shay, L. K., 2010: Air-Sea Interactions. *Global Perspectives on Tropical Cyclones: From Science to Mitigation* J. C. L. Chan and J. D. Keprt, Eds., World Scientific Publishing Company

Shekhar, S. and S. Chawla, 2003: Spatial databases: a tour. *Upper Saddle River, New Jersey*, **7458.**

Shekhar, S., P. Zhang, and Y. Huang, 2010: Spatial data mining. *Data Mining and Knowledge Discovery Handbook*, 837-854.

Shen, W., I. Ginis, and R. Tuleya, 2002: A numerical investigation of land surface water on landfalling hurricanes. *Journal of the Atmospheric Sciences*, **59,** 789-802.

Silk, J., 1981: The analysis of variance. Geo Abstracts, University of East Anglia.

Simpson, A. and H. Riehl, 1981a: *The hurricane and its impact*. Louisiana State University Press.

Simpson, J., E. Ritchie, G. Holland, J. Halverson, and S. Stewart, 1997: Mesoscale interactions in tropical cyclone genesis. *Monthly Weather Review*, **125,** 2643-2661.

Simpson, J., J. B. Halverson, B. S. Ferrier, W. A. Petersen, R. H. Simpson, R. Blakeslee, and S. L. Durden, 1998: On the role of "hot towers" in tropical cyclone formation. *Meteorology and Atmospheric Physics*, **67,** 15-35.

Simpson, R. and H. Riehl, 1981b: *The hurricane and its impact*. Louisiana State University Press Baton Rouge.

Sinclair, M., 2002: Extratropical transition of southwest Pacific tropical cyclones. Part I: Climatology and mean structure changes. *Monthly Weather Review*, **130,** 590-609.

Soden, B. J., C. S. Velden, and R. E. Tuleya, 2001: The impact of satellite winds on experimental GFDL hurricane model forecasts. *Monthly Weather Review*, **129,** 835-852.

Sokal, R. R. and F. J. Rohlf, 1969: Biometry: the principles and practice of statistics in biological research. W. H. Freeman, San Francisco.

Sui, C., P. Chung, and T. Li, 2007: Interannual and interdecadal variability of the summertime western North Pacific subtropical high. *Geophysical Research Letters*, **34,** L11701.

Sun, S. and M. Ying, 1999: Subtropical high anomalies over the western pacific and its relations to the asian monsoon and sst anomaly. *Advances in Atmospheric Sciences*, **16,** 559-568.

Tao, S. and L. Chen, 1987: A review of recent research on the East Asian summer monsoon in China. *Monsoon Meteorology*, 60.

Tao, Z., B. Tian, and W. Huang, 1994: Asymmetry structure and torrential rain of landing typhoon. *Journal of Tropical Meteorology*, 01.

Thorncroft, C. and S. Jones, 2000: The extratropical transitions of Hurricanes Felix and Iris in 1995. *Monthly Weather Review*, **128,** 947-972.

Tuleya, R., M. Bender, and Y. Kurihara, 1984: A simulation study of the landfall of tropical cyclones using a movable nested-mesh model. *Mon. Wea. Rev*, **112,** 124–136.

Tuleya, R. E. and Y. Kurihara, 1977: Numerical-Simulation of Landfall of Tropical Cyclones. *Bulletin of the American Meteorological Society*, **58,** 109-109.

Tuleya, R. E. and Y. Kurihara, 1978: A Numerical Simulation of the Landfall of Tropical Cyclones. *Journal of the Atmospheric Sciences*, **35,** 242-257.

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. *Weather and Forecasting*, **6,** 244-253.

Vickery, P., 2005: Simple empirical models for estimating the increase in the central pressure of tropical cyclones after landfall along the coastline of the United States. *Journal of Applied Meteorology*, **44,** 1807-1826.

Wang, B. and X. Li, 1992: The beta drift of three-dimensional vortices- A numerical study. *Monthly Weather Review*, **120,** 579-593.

Wang, B. and X. F. Li, 1995: Propagation of a Tropical Cyclone in a Meridionally Varying Zonal Flow - an Energetics Analysis. *Journal of the Atmospheric Sciences*, **52,** 1421-1433.

Wang, B. and Z. Fan, 1999: Choice of South Asian summer monsoon indices. *Bulletin of the American Meteorological Society*, **80,** 629-638.

Wang, B. and J. C. L. Chan, 2002: How strong ENSO events affect tropical storm activity over the Western North Pacific. *Journal of Climate*, **15,** 1643-1658.

Wang, B. and T. Li, 2004: East Asian monsoon-ENSO interactions. *East Asian Monsoon*, 177–212.

Wang, B., R. G. Wu, and T. Li, 2003: Atmosphere-warm ocean interaction and its impacts on Asian-Australian monsoon variation. *Journal of Climate*, **16,** 1195-1211.

Wang, B., Z. W. Wu, J. P. Li, J. Liu, C. P. Chang, Y. H. Ding, and G. X. Wu, 2008: How to measure the strength of the East Asian summer monsoon. *Journal of Climate*, **21,** 4449-4463.

Wang, Y., 2001: An Explicit Simulation of Tropical Cyclones with a Triply Nested Movable Mesh Primitive Equation Model: TCM3. Part I: Model Description and Control Experiment*. *Monthly Weather Review*, **129,** 1370-1394.

Wang, Y., 2002: Vortex Rossby waves in a numerically simulated tropical cyclone. Part I: Overall structure, potential vorticity, and kinetic energy budgets. *Journal of the Atmospheric Sciences*, **59,** 1213-1238.

Wang, Y. and G. J. Holland, 1996a: The beta drift of baroclinic vortices. Part I: Adiabatic vortices. *Journal of the Atmospheric Sciences*, **53,** 411-411.

Wang, Y. and G. J. Holland, 1996b: Tropical cyclone motion and evolution in vertical shear. *Journal of the Atmospheric Sciences*, **53,** 3313-3332.

Wang, Y. and G. J. Holland, 1996c: The beta drift of baroclinic vortices. Part II: Diabatic vortices. *Journal of the Atmospheric Sciences*, **53,** 3737-3756.

Wang, Y. and C. C. Wu, 2004: Current understanding of tropical cyclone structure and intensity changes - a review. *Meteorology and Atmospheric Physics*, **87,** 257-278.

Weng, H., K. Ashok, S. K. Behera, S. A. Rao, and T. Yamagata, 2007: Impacts of recent El Nino Modoki on dry/wet conditions in the Pacific rim during boreal summer. *Climate Dynamics*, **29,** 113-129.

Wentz, F. J., P. Ashcroft, and C. Gentemann, 2002: Post-launch calibration of the TRMM Microwave Imager. *IEEE Transaction on Geoscience and Remote Sensing*, **39,** 415-422.

Wentz, F. J., C. Gentemann, D. Smith, and D. Chelton, 2000: Satellite measurements of sea surface temperature through clouds. *Science*, **288,** 847.

Williams, R. T. and J. C. L. Chan, 1994: Numerical-Studies of the Beta-Effect in Tropical Cyclone Motion .2. Zonal Mean Flow Effects. *Journal of the Atmospheric Sciences*, **51,** 1065-1076.

Willoughby, H., 1978: A possible mechanism for the formation of hurricane rainbands. *Journal of Atmospheric Sciences*, **35,** 838-848.

Willoughby, H. E., F. D. Marks Jr, and R. J. Feinberg, 1984: Stationary and moving convective bands in hurricanes. *Journal of the Atmospheric Sciences*, **41,** 3189-3211.

Wong, M. L. M. and J. C. L. Chan, 2004: Tropical cyclone intensity in vertical wind shear. *Journal of the Atmospheric Sciences*, **61,** 1859-1876.

Wong, M. L. M. and J. C. L. Chan, 2006a: Land friction influences tropical cyclone motion. *Bulletin of the American Meteorological Society*, **87,** 564-565.

Wong, M. L. M. and J. C. L. Chan, 2006b: Tropical cyclone motion in response to land surface friction. *Journal of the Atmospheric Sciences*, **63,** 1324-1337.

Wong, M. L. M. and J. C. L. Chan, 2007: Modeling the effects of land-sea roughness contrast on tropical cyclone winds. *Journal of the Atmospheric Sciences*, **64,** 3249-3264.

Wong, M. L. M., J. C. L. Chan, and W. Zhou, 2008: A simple empirical model for estimating the intensity change of tropical cyclones after landfall along the south China coast. *Journal of Applied Meteorology and Climatology*, **47,** 326-338.

Woodcock, F., 1995: Tropical Cyclone Work Station to Support Warning-response Management: ATCW Project 1/91. *Australian Journal of Emergency Management, The*, **10,** 21.

Wroe, D. R. and G. M. Barnes, 2003: Inflow layer energetics of Hurricane Bonnie (1998) near landfall. *Monthly Weather Review*, **131,** 1600-1612.

Wu, C.-C. and H.-J. Cheng, 1999: An Observational Study of Environmental Influences on the Intensity Changes of Typhoons Flo (1990) and Gene (1990). *Monthly Weather Review*, **127,** 3003-3031.

Wu, C. C., 2001: Numerical simulation of Typhoon Gladys (1994) and its interaction with Taiwan terrain using the GFDL hurricane model. *Monthly Weather Review*, **129,** 1533-1549.

Wu, C. C. and K. A. Emanuel, 1993: Interaction of a baroclinic vortex with background shear: Application to hurricane movement. *Journal of the Atmospheric Sciences*, **50,** 62-62.

Wu, C. C. and Y. H. Kuo, 1999: Typhoons affecting Taiwan: Current understanding and future challenges. *Bulletin of the American Meteorological Society*, **80,** 67-80.

Wu, C. C., K. H. Chou, H. J. Cheng, and Y. Q. Wang, 2003: Eyewall contraction, breakdown and reformation in a landfalling typhoon. *Geophysical Research Letters*, **30,** -.

Wu, L. G. and B. Wang, 2004: Assessing impacts of global warming on tropical cyclone tracks. *Journal of Climate*, **17,** 1686-1698.

Wu, M., W. Chang, and W. Leung, 2004: Impacts of El Nino-Southern Oscillation events on tropical cyclone landfalling activity in the western North Pacific. *Journal of Climate*, **17,** 1419-1428.

Yang, F., H. L. Pan, S. K. Krueger, S. Moorthi, and S. J. Lord, 2006: Evaluation of the NCEP Global Forecast System at the ARM SGP Site. *Monthly Weather Review*, **134,** 3668.

Yang, M., D. Zhang, and H. Huang, 2008: A modeling study of Typhoon Nari (2001) at landfall. Part I: Topographic effects. *Journal of the Atmospheric Sciences*, **65,** 3095-3115.

Yang, R., J. Tang, and D. Sun, 2011: Association Rule Data Mining Applications for Atlantic Tropical Cyclone Intensity Changes. *Weather and Forecasting*, **0,** null.

Yohannes, Y. and P. Webb, 1999: *Classification and regression trees, CART: a user manual for identifying indicators of vulnerability to famine and chronic food insecurity*. Intl Food Policy Research Inst.

Yu, C.-K. and L.-W. Cheng, 2008: Radar Observations of Intense Orographic Precipitation Associated with Typhoon Xangsane (2000). *Monthly Weather Review*, **136,** 497-521.

Yu, Z. F., X. D. Liang, H. Yu, and J. C. L. Chan, 2010: Mesoscale vortex generation and merging process: A case study associated with a post-landfall tropical depression. *Advances in Atmospheric Sciences*, **27,** 356-370.

Yuan, J. N., Y. Y. Huang, and C. X. Liu, 2008: A Simulation Study of the Influence of Land Friction on Landfall Tropical Cyclone Track and Intensity. *Journal of Tropical Meteorology*, **14,** 53-56.

Yuan, Z., Y. Hui, W. Dong-fa, and T. Wei-ping, 2007: Factors affecting the track of tropical cyclones after landfall in eastern China. *Journal of Tropical Meteorology*, 03.

Zar, J. H., 1974: *Biostatistical analysis*. Vol. 620, Prentice-Hall Englewood Cliffs, NJ.

Zhang, H., C. Xiao-peng, K. Feng-qin, and C. Shou-xin, 2007: Observational Analysis for a Torrential Rain Caused by a Landfall Typhoon [J]. *Plateau Meteorology*, **5.**

Zhang, J. and S. Yu, 1998: A diagnostic study on the relationship between the assembling of low frequency waves in the pacific ocean and the abnormality of the subtropical high. *Advances in Atmospheric Sciences*, **15,** 247-257.

Zhou, T., R. Yu, J. Zhang, H. Drange, C. Cassou, C. Deser, D. L. R. Hodson, E. Sanchez-Gomez, J. Li, N. Keenlyside, X. Xin, and Y. Okumura, 2009: Why the Western Pacific Subtropical High Has Extended Westward since the Late 1970s. *Journal of Climate*, **22,** 2199-2215.