Thesis for Master's degree

Diagnosis of tropical cyclone track patterns in WRF using the fuzzy *c*-means clustering method

Fuzzy *c*-means clustering 방법을 이용한

WRF 모의 태풍의 진로 군집 분석

February 2013

Hong, Sunyoung

School of Earth and Environmental sciences

Seoul National University

Abstract

This study investigated the variability of simulated tropical cyclone (TC) tracks over the western North Pacific (WNP) from the regional climate model (RCM) to compare to the observed TC tracks using the fuzzy clustering analysis. To understand variability of TC tracks specifically, TC tracks are classified to six representative patterns with fuzzy c-means clustering method.

The optimum cluster number is determined by cross checking four objective validity measures. First, 519 observed TC tracks formed from 1982 to 2010 during July to October are classified to six clusters and 549 simulated TC tracks are classified based on the clusters for observed TC tracks. These clusters are described as TCs landfalling countries of East Asia (e.g., East China, Taiwan, Korea and Japan) (C1), TCs affecting Japan with long trajectories (C2), early recurving tracks passing the east of Japan (C3), TCs moving the easternmost region over the WNP (C4), TCs over the South China Sea (SCS) with short straight trajectories (C5) and TCs moving across the Philippines with straight trajectories (C6).

In comparison between observed track clusters and simulated clusters, C1 and C2 have remarkable differences in the percentage (number). The C1 of simulated TC tracks decreased significantly from 19.5% (101) to 13.1% (72)

and C2 increased most from 15% (79) to 24% (132). These differences in percentage of C1 and C2 lead to the variability in spatial distribution of TC genesis frequency and track density. For the simulation, TC genesis frequency decreased in the Philippine Sea (PS), where the genesis region of TCs for C1 and increased in the southeastern part of the WNP where the TCs for C2 are mainly formed. The track density of simulated TC is lower than observation in the PS and East China Sea but higher in the eastern part of WNP. The other clusters (C3–6) are successfully simulated in the percentage and spatial distribution.

These significant discrepancies in C1 and C2 can be explained by large-scale environments. The membership coefficient weighted composites for the TC genesis day are conducted to investigate the relationship between variability of dominant track patterns (e.g., C1 and C2) and large-scale circulation. The C1 is the cluster related to the La Niña and simulated TCs of C1 are formed less because low level cyclonic wind is underestimated in the PS and vertical wind shear is overestimated than those of observation. But for C2, El Niño related cluster, the genesis of simulated TCs are enhanced by stronger positive relative vorticity and monsoon trough extending to the genesis region (10°–20°N, 145°–160°E). Moreover, simulated TCs of C2 recurve earlier because of eastern retreating for the North Pacific subtropical High (NPSH) than observation.

Therefore, simulated TC tracks of C2 are steered more eastward than tracks of observation.

The results show that the WRF model has good performance for simulating TC tracks but distinctive variability in C1 and C2. This study is expected to improve the ability of regional climate model further.

Keywords : tropical cyclone, track clustering, regional climate model

Student number : 2009-22516

Table of Contents

List of Figures

(black : observation, red : model)

Figure 9. Monthly mean number of TCs for (a)─(f) six hard clusters and (g) all TCs (white : observation, black : model)

Figure 10. seven sub-regions located in WNP area. (The Philippines, Vietnam, South China, East China, Taiwan, Korea and Japan)

Figure 11. The membership coefficient weighted composites of the SST anomaly (shading), 850hPa wind anomaly (vector) based on the day of TC genesis (a) NCEP and (b) WRF for C2. Only significant values at the 90% confidence level are drawn.

Figure 12. The monthly positions of monsoon trough during June-November from Atkinson [1971]

Figure 13. The membership coefficient weighted composites of the 850hPa zonal wind (shade) on the day of TC genesis of (a) NCEP and (b) WRF for C2

Figure 14. The membership coefficient weighted composites of the tropospheric layer-mean flows (stream line) and 850hPa positive relative vorticity (shade) on the day of TC genesis of (a) NCEP and (b) WRF for C2. The red contour is the 5880gpm of the 500hPa geopotential height composite. Black solid line is mean track

Figure 15. TC recurving frequency of (a) RSMC and (b) WRF for C2 in unit grid (5° × 5°) per year during TC season (JASO) in 1982–2010.

Figure 16. Same as Figure 10 (a, b) for C1

Figure 17. The membership coefficient weighted composites of the vertical wind shear (shade) on the day of TC genesis of (a) NCEP and (b) WRF for C1. Solid line is mean track.

List of Table

1. Introduction

TC has been one of the most catastrophic natural disasters to the landfall countries for a long time period. Strong gusts accompanied by heavy rain cause tremendous economic damages and victims. To prevent TC induced damages, forecasting seasonal TC activity is regarded as very important issue to the TC-prone countries in every year. TC landfall is highly dependent on their tracks, it is tried to understand long-term variability of TC tracks [e.g., Chan, 1985; Harr and Elsberry, 1991 and Ho et al., 2004].

Many scientists have tried to improve the ability of global and regional climate models (RCM). It is most essential to simulate observed TC activity appropriately for evaluating the ability of model. In previous studies, climate variability of simulated TC activities (i.e., TC genesis, structure and intensity) is examined [e.g., Knutson et al., 2007; Stowasser et al., 2007; Au-Yeung and Chan, 2012]. Au-Yeung and Chan [2012] reproduced TC tracks formed in 1997−1998 which is one of the strongest El Niño/Southern Oscillation events. However in the simulation of TC tracks, analyses are carried out only for the individual TC track or a few tracks for short time periods. The climatological variability of TC tracks can’t be explained by simulation for individual TC track because of its numerous possible trajectories over the WNP basin.

Moreover, it is too complicated to define boundaries of TC tracks due to its ambiguity [Kim, 2005]. To understand variability of TC tracks, classifying various TC tracks into representative patterns is essential. Through the specific track patterns, the dominant factor affecting on the climatological variability of TC activity can be figured out. Therefore, classifying with representative track patterns are needed to investigate long term changes of TC track in the basin.

To examine the ability of RCM, observed TC tracks over WNP reproduced by WRF are classified with six representative patterns using fuzzy clustering method (FCM). For evaluating the simulation ability of TC track, each pattern will be compared to observation and the relationship between variability of track patterns and large-scale environments will be discussed.

2. Data and Method

2.1. Data

In this study, TC data for evaluating the simulation ability of WRF is obtained from the best track data sets issued by Regional Specialized Meteorological Center-Tokyo (RSMC) Tropical Cyclone Center. The data sets consist of names, longitude and latitude positions, minimum surface pressures, and maximum wind speeds of TCs every 6 hours. TCs are classified into four categories according to 10 min averaged maximum sustained wind speeds (v_{max}) : tropical depression (TD, v_{max} <17ms^{-1}), tropical storms (TS, 17ms^{-1} ≤ v_{max} <25ms^{-1}), severe tropical storms (STS, 25 m s^{-1} ≤ v_{max} < 33ms^{-1}), and typhoons (TY, v_{max} ≥ 33ms^{-1}). These categories are from defined by World Meteorological Organization. This study will consider TCs over TD with life time more than 48hours. For analyzing large-scale environment and model initialization, National Centers for Environmental Prediction–National Center (NCEP) / Department of Energy (DOE) Reanalysis-Ⅱ (R-2) daily data set and National Oceanic Atmospheric Administration (NOAA) weekly optimal interpolation sea surface temperature (OISST) data which have 2° × 2° horizontal resolutions [Kanamitsu et al., 2002] are used. This study is focused

on WNP region where 30% of TCs global total [Park, 2008]. The analyzed period is set from 1982 to 2010 during TC season (July−October, JASO).

2.2. WRF model and experimental design

2.2.1. Model Configuration

The simulation results of TC activity over WNP discussed in this study is obtained from Jin et al. [2012]. The TCs are simulated with global and regional Weather Research and Forecasting (WRF) model version 3.3 [Skamarock et al., 2008]. WRF is non-hydrostatic model and mesoscale numerical weather prediction system designed by National Center for Atmospheric Research (NCAR) for operational forecasting and atmospheric research.

Global WRF has limitations of TC resolving capability for 150-km horizontal resolution, one way nesting is applied to regional climate model (RCM) with 50-km horizontal resolution over the western North Pacific (WNP) region (Fig. 1). The model top level is 50hPa and there are 27 vertical layers from the surface and the top level. The integral time step for simulation is 600s in the global WRF model and 200s in the nested WRF RCM. The physical parameterization schemes and detection and tracking criteria for TC simulation is defined as follows :

1) Local minimum sea level pressure

2) Maximum relative vorticity at 850hPa > 4.9 X 10^{-5} s^{-1}

3) Maximum wind speed at surface > 17 m s^{-1}

4) Warm core criterion : $\Delta T_{ave}=\Delta T_{300}+\Delta T_{500}+\Delta T_{700}\geq 1.7K$

5) Maximum wind speed at 850hPa > that at 300hPa

6) Duration of all above conditions > 2 days

This criteria is used same with in previous studies [Oouchi et al., 2006; Camargo et al., 2007; Cha et al., 2011] except for the strengthened criteria for the surface wind speed and the warm core criterion : average of temperature deviations at 300, 500 and 700hPa being 17 ms^{-1} and 1.7K. The low boundary SST is prescribed every 24 hours for the simulation and the time period of simulation is JASO for 29 years from 1982 to 2010.

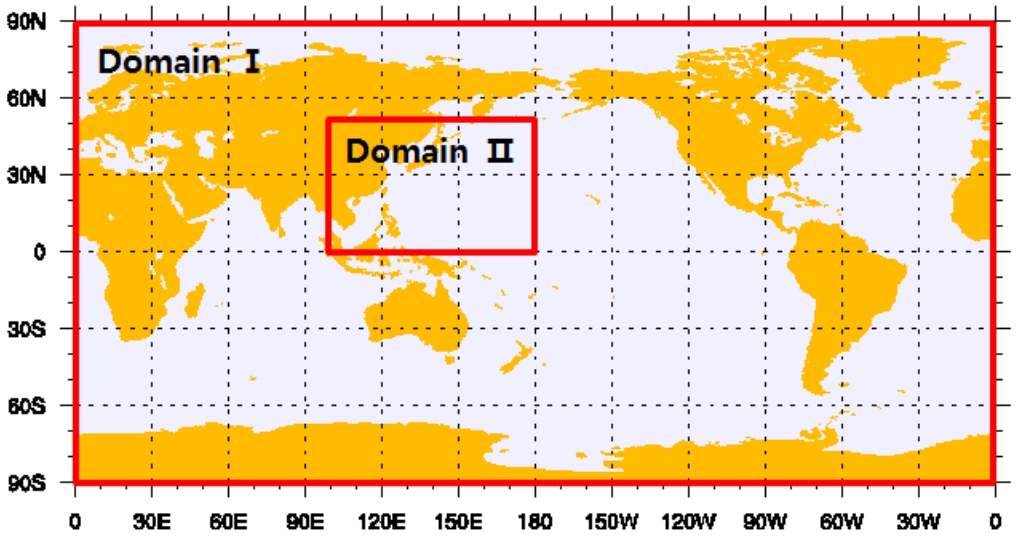


Figure 1. Global (Domain Ⅰ) and regional (Domain Ⅱ) for WRF model.

2.2.2. Climatology of simulated TCs

The first requirement for examining the climatological variability of the model-simulated TC activity is the appropriate simulation for observed climatology. Comparing between observation and model, the model reasonably reproduces seasonal TC numbers formed in the WNP. The 17.9 TCs are observed per year and 18.9 TCs are simulated in the model. This discrepancy between observation and model can be seen by investigating the spatial distribution of TC genesis and track climatology.

The advantageous location for TC formation over the WNP is in the intertropical convergence zone (ITCZ) and in the PS and SCS where monsoon trough located [e.g., Gray, 1968; Ramage, 1974; Chia and Ropelewski, 2002]. Figure 2a and 2b shows the spatial distribution of TC genesis frequency over the WNP for observation and model. TC genesis frequency is defined how many TCs are formed in unit grid per year. In this study, unit grid is set by 5° × 5° latitude—longitude, horizontally. The climatological genesis frequency of simulated TCs has notable similarities with that of observed TCs (Figs 2a vs 2b). The simulated TCs are formed near the ITCZ and region where monsoon trough formed. However in the Fig. 2a, the observed TCs are formed most in the PS (10°−20°N, 125°−135°E) and maximum values for the simulated TC genesis frequency is revealed in the south eastern part of WNP (10°−20°N,

145°−160°E) and the SCS (10°−20°N, 110°−120°E). The genesis frequency of simulated TCs decreased than that of observed TCs in the PS (Fig. 2b).

Figs. 2c and 2d show the climatology of TC track density for observation and model. TC track density denotes how many unit times (6hrs) TCs stay in unit grid (5° × 5°) per year. It represents the climatologically concentrated region for TC passage and its duration over the WNP. In the spatial distribution of TC track density, simulation result shows similar pattern to the observation remarkably. Although, the simulated track density is lower in the PS and East China Sea (ECS) but higher in the south eastern part of WNP (10°−20°N, 145°−160°E) compared to its observed counterpart (Figs 2c vs 2d).

The model tends to simulate less TCs in the PS but more in the south eastern part of WNP comparing to the observation. However, it needs to specific analysis for TC tracks because the goal of this study is to understand the variability of simulated TC activity, especially TC tracks, and estimate the simulation capability of the model. Including numerous possible tracks in the climatological distribution, representative track patterns with distinctive characteristics are essential to understand variability of simulated TC tracks in this study.

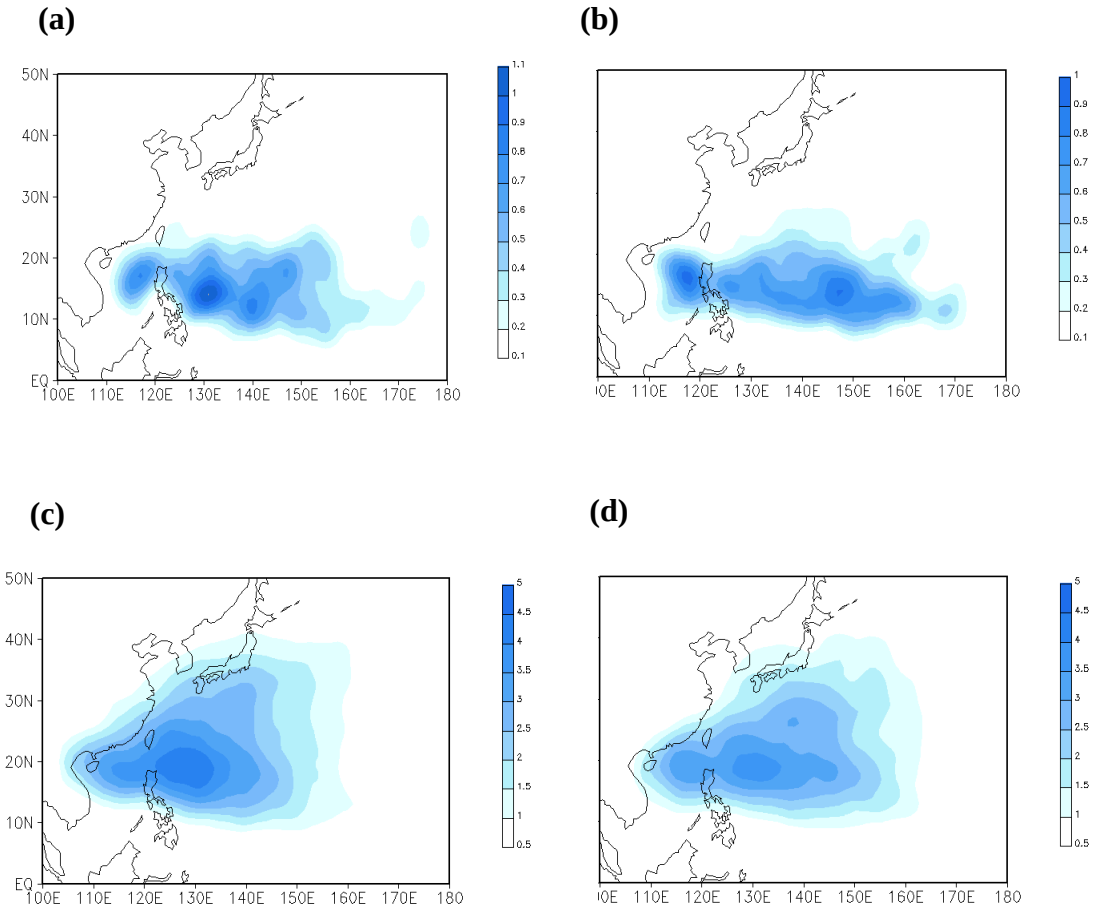


Figure 2. Climatology of TC genesis frequency for (a) RSMC, (b) WRF and climatology of TC track density for (c) RSMC, (d) WRF per unit grid (5° × 5°) during TC season (JASO) in 1982–2010.

2.3. Pattern Classification of TC tracks

2.3.1. Classifying TC tracks with fuzzy clustering method

In this study FCM is applied to classify TC track patterns. FCM is known for appropriate technique to classify data containing vague boundaries between clusters like TC tracks [Kim et al., 2011]. The algorithm is used same as in Bezdak [1981] and Kim et al. [2011]. The algorithm and detailed explanations of FCM are written in Appendix A−C.

In the FCM, each data object belongs to all clusters with different membership coefficients. The membership coefficient indicates how closely the data object is located from the cluster center varying from 0 to 1. Stronger belongingness of the data object to the cluster is denoted with higher membership coefficient. Therefore, the cluster center is defined as membership coefficient-weighted mean of all data objects in the FCM [Kim et al., 2011].

Because the result of FCM is highly dependent on the number of cluster center, it is crucial to determine optimum cluster number. To find optimum cluster number of TC track patterns objectively, four scalar validity measures are applied [Kim et al., 2011]. Those are partition coefficient [Bezdek, 1981], partition index [Bensaid et al., 1996], separation index [Xie and Beni, 1991] and Dunn index [Dunn, 1973]. The formulas and detailed explanations for these

indices are noted in Appendix D. The partition coefficient measures the degree of fuzzy cluster overlapping but, other indices measure compactness and separation of clusters. Therefore, larger partition coefficient represents better clustering, while smaller values mean more optimal in other indices (e.g., partition index, separation index, Dunn index). Figure 3a−3d shows four scalar validity measures for selecting the optimum number of TC tracks. In Figure 3a, the partition coefficient is the maximum at cluster number 2. And the partition index (Fig. 3b) and the separation index (Fig. 3c) show monotonic decrease with cluster number increase. However, in the Dunn index (Fig. 3d), the value is the minimum at cluster number six. According to the result, optimum cluster number six is taken for fuzzy clustering of RSMC TC tracks and this is based on the fuzzy clustering of simulated TC tracks. Figure 4 shows the six fuzzy clusters (C1−C6) of observed TC tracks during the TC season (JASO) obtained by the FCM. The color depth of TC tracks denotes the membership coefficient to the clusters and it is shown that the tracks with deep color are close to the cluster center. The bold line denotes the mean track of each cluster.

Kim et al. [2011] presented seven fuzzy clusters classified by FCM for TCs over the WNP. However, C3 of Kim et al. [2011] is not occurred in this study due to the difference in the analysis period. Because of difference in analysis period, total analyzed TCs are 855 for Kim et al. [2011] but 519 TCs for this

study. Furthermore, it is referred that C1−C3 are fuzzier than C4−C6. TCs of C1−C3 are distributed in the average region of all TCs climatologically [Kim et al., 2011]. In this study, TC tracks of C1−C3 are relocated when input TC data are adjusted. For this reason, C3 of Kim et al. [2011] is not occurred.

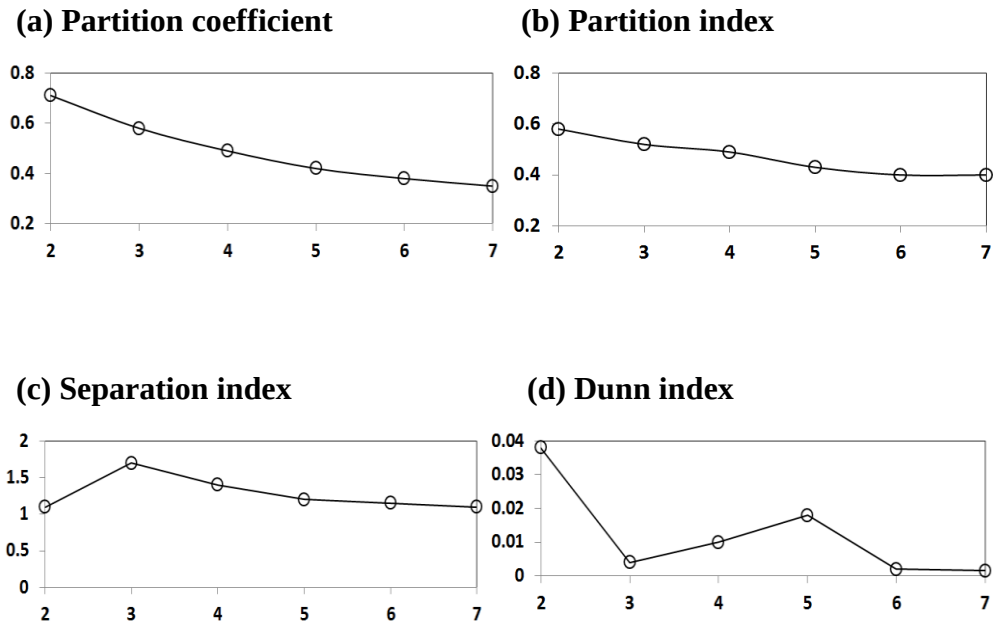


Figure 3. Response of four scalar validity measures to an increase in the number of cluster: (a) partition coefficient, (b) partition index, (c) separation index, (d) Dunn index.

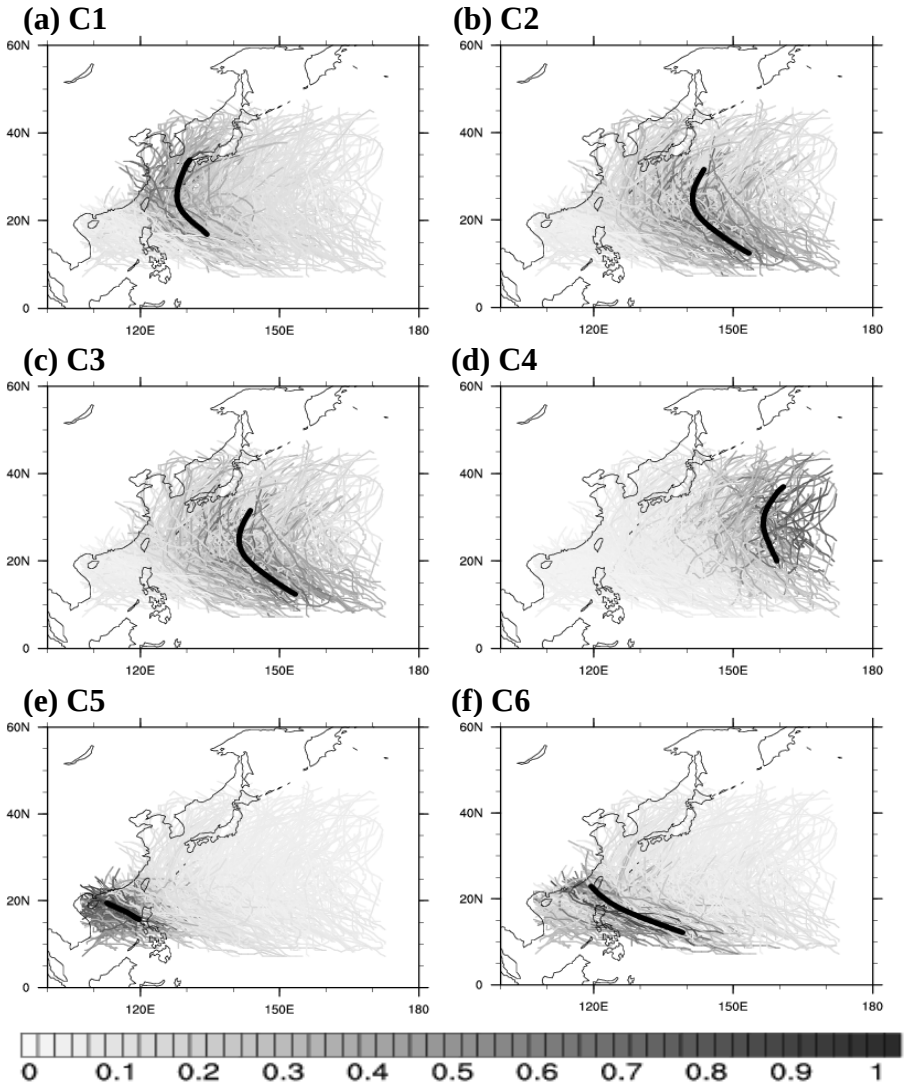


Figure 4. The six fuzzy clusters of observed TC tracks during the TC season (JASO). The color depth of tracks represents the membership coefficients. The bold line denotes the mean track of the cluster.

3. Track patterns of simulated TC

3.1. Spatial distribution of track patterns

The six optimum clustering number of TC tracks is determined by FCM from 519 observed TCs and 549 simulated TCs for the analysis period (C1−C6). To compare cluster by cluster for the observed TCs and those of simulation, the observed track cluster centers are fixed in classifying simulated TC tracks. Figs. 5 shows representative track patterns of observed TC tracks obtained by FCM and the simulated track clusters are shown in the Figs. 6. Each cluster includes all membership coefficients of whole TCs related to the distance from cluster centers. It indicates probability of cluster center and distinguishes geographical characteristics of track patterns. The bold black line represents the mean track of each cluster. In the Figs. 5 and 6, each TC is set down by a cluster where membership coefficient is the largest. For practical purpose, analyses will be based on the hard clusters hereafter. These hard clusters still exit having membership coefficient information of each cluster.

Track patterns of observation consist of four recurving tracks (Figs. 5a–d) and two straight moving tracks (Figs 5e and 5f). In the bottom-right of the panels, percentage and number (parentheses) of each cluster are noted. 19.5%

(101/519) of TCs belong to the C1 as second highest percentage. The C1 includes the TCs develop around the northernmost part of the PS, recurve in the East China Sea (ECS) and hit the East Asian countries (e.g., East China, Taiwan, Korea and Japan) (Fig. 4a). 15.2% (79/519) TCs of C2 with longest trajectories develop south eastern region of the WNP and move rather straight northwestward before recurving south of Japan (Fig. 4b). 16.8% (87/519) recurving tracks of C3 locates offshore east of Japan with irregular shapes (Fig. 4c). 11.3% (59/519) TCs of C4 also recurve in the east of Japan but they move farther to the east (Fig. 4d). C3 and C4 both include recurving tracks over the open ocean east of Japan. However, TCs of C3 are mostly oriented from north so recurve quickly as developing but TCs of C4 move from the central North Pacific. C5 has 20.2% (105/519) straight moving tracks in the South China Sea (SCS) with the highest percentage and landfall on northern Vietnam and the southern China coastal regions (Fig. 4e). Seventeen percent (88/519) TCs of C6 also move straight travel across the Philippines with long westward trajectories and hit similar region to those of C5 (Fig. 4f).

The track patterns of simulated TCs have remarkable similarities with those of observation. In comparison between observation and model, remarkable discrepancies are turned up in the TC tracks of C1 and C2. The simulated C1 has the lowest percentage decreased from 19.5% to 13.1%. In contrast, C2 has

the highest percentage increased from 15.2% to 24.0%. In the difference in number of C1, 29 simulated TCs are decreased than those of observation during TC season for 29 years (1982−2010) in other words, one TC per year. And simulated C2 increased 33 TCs than observed C2 during 29 years in other words, 1.8 TCs per year. It is comparable to the difference in the climatology of seasonal total TC number and of observation and model. In the observation, the seasonal mean TC number of the year is 17.9 but 18.9 for the model. Therefore the variability of simulated C1 and C2 can be noted to dominant clusters in the variability in the climatology of TC activity.

Dominant changes of C1 and C2 are coincidently confirmed in the spatial distribution of TC track density. In the PS, where TCs of C1 mostly migrate, simulated track density decreased than those of observation but increased in the eastern part of WNP where TCs of C2 mainly move.

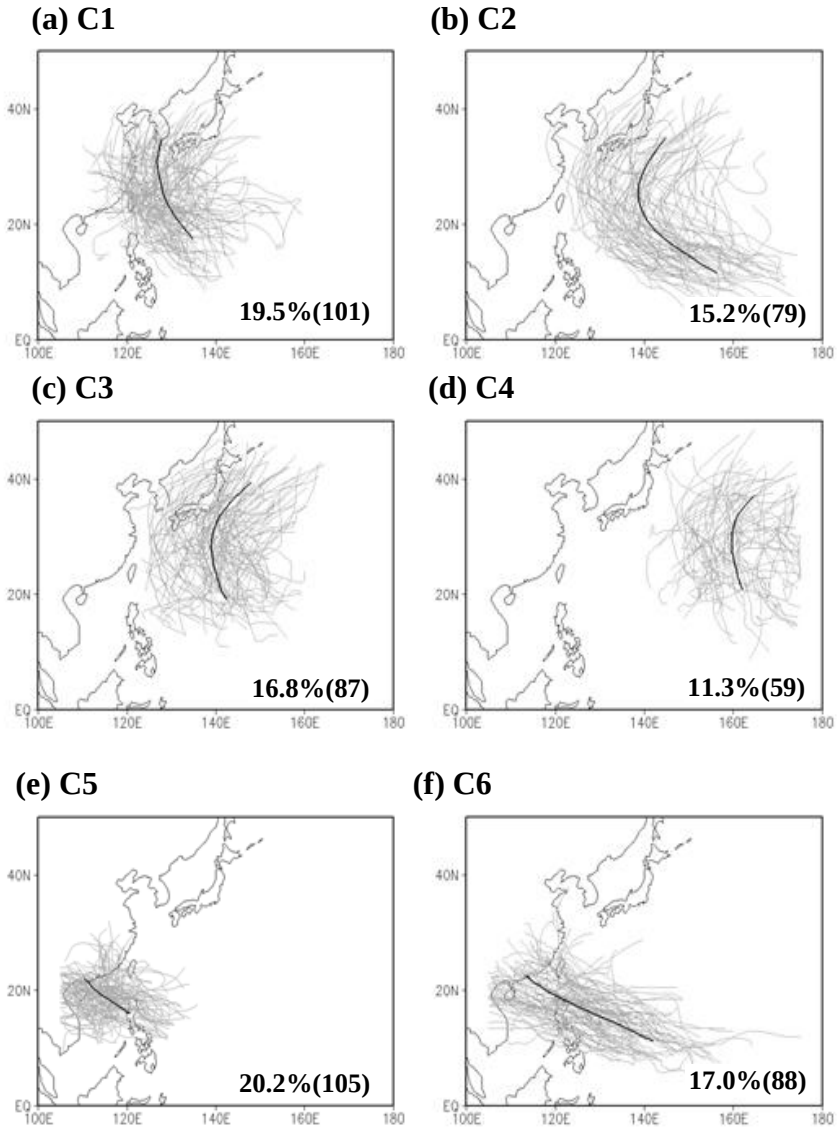


Figure 5. The six fuzzy clusters of observed TC tracks (C1─6) assigned to six clusters where the membership coefficient is the largest. The percentage of cluster (number) is displayed on the bottom-right corner of the map.

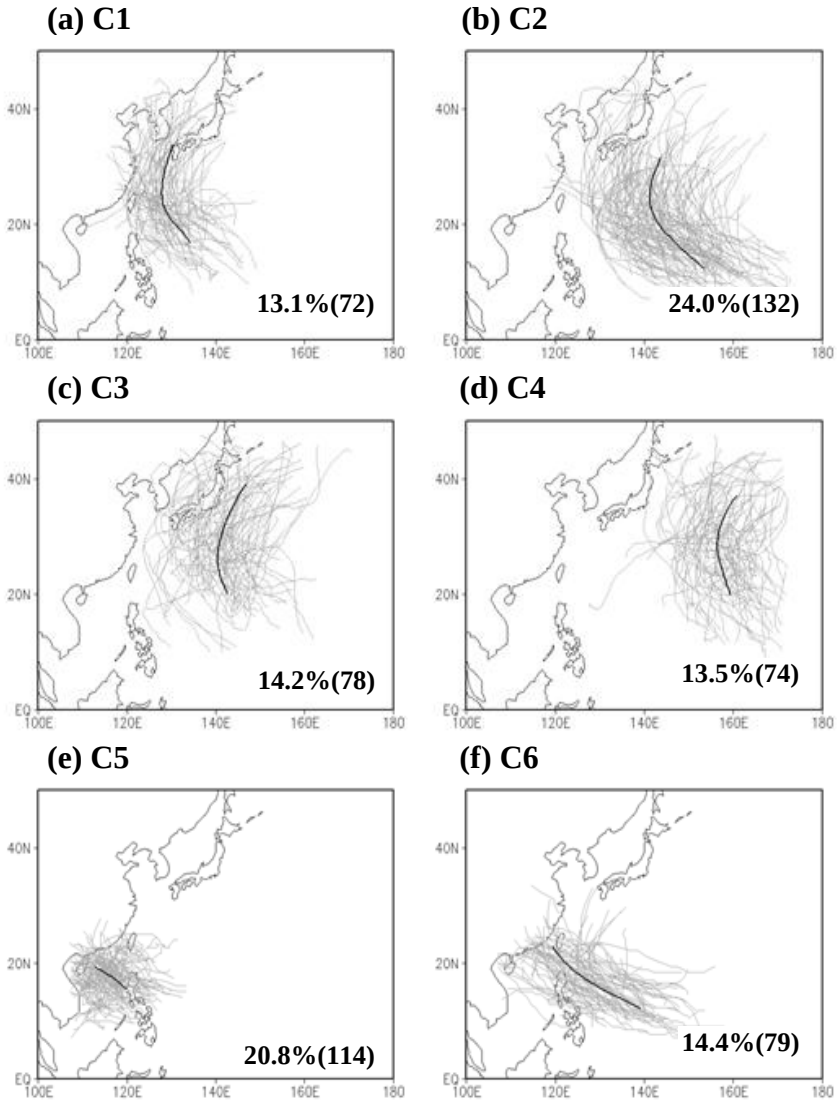


Figure 6. Same as Figure 5 (a–f) for simulated TC tracks.

3.2 Mean properties

For comparing each cluster between observation and model specifically, mean properties of clusters are examined. Table 2 shows mean properties such as minimum central pressure, maximum wind speed, life time and genesis location of each cluster for observation and model. The mean minimum central pressure is 959.1hPa for model (961.6hPa for observation), the mean maximum wind speed is 36.0m/s (34.7m/s), the mean life time is 4.4days (6days) and mean genesis location is 15.73°N, 141.14°E (15.96°N, 140.73°E) for 549 (519) TCs. The mean values of simulated track clusters are remarkably similar to those of observation. Although, there is difference in the life time between observation and model. This is because that TC intensity is maximized when v_{max} is more than $33ms^{-1}$ and less than $44ms^{-1}$ but underestimated for minimum sea level pressure under 960hPa (figure not shown).

It is referred in previous studies [e.g., Camargo and Sobel, 2005; Chan, 2008 and Kim et al., 2011] that TCs formed in the south and east region can be stronger than any other TCs having potential power to travel far northwest. In the table, clusters formed in south and east region (e.g., C2, C4 and C6) relatively have longer life time and stronger intensity than clusters formed in north and west (e.g., C1, C3 and C5). Among the stronger clusters (e.g., C2, C4

and C6), C2 has TC tracks with the longest trajectories and the longest life time for both observation and model. On the other hand, C5 is representative cluster of this hypothesis having the lowest life time, the weakest intensity with shortest trajectories for both observation and model. These similarities in mean values of each cluster show that simulated track clusters are well classified having their distinctive characteristics with those of observed track clusters.

	Observation					Model				
	Minimum Central pressure (hPa)	Maximum Wind speed (m/s)	Life time (days)	Genesis location		Minimum Central pressure (hPa)	Maximum Wind speed (m/s)	Life time (days)	Genesis location	
				Latitude (degree east)	Longitude (degree north)				Latitude (degree east)	Longitude (degree north)
C1	958.9	35.9	6.2	17.70	134.52	956.9	36.5	4.6	16.85	134.35
C2	938.4	43.1	7.5	11.58	156.13	952.0	38.2	4.7	12.37	153.40
C3	965.7	33.0	6.0	19.15	142.17	960.2	34.3	4.3	20.13	142.71
C4	972.3	30.9	5.1	20.86	162.20	964.2	33.3	4.4	19.97	159.37
C5	979.9	26.9	4.5	16.00	121.16	966.6	34.6	3.7	12.15	139.14
C6	952.6	39.0	6.7	11.40	141.59	956.5	38.2	4.7	15.64	119.73
All TCs	961.6	34.7	6.0	15.96	140.73	959.1	36.0	4.4	15.73	141.14

Table 1. Mean values of the minimum central pressure, maximum wind speed, life time and genesis location for the TCs in six hard clusters and all TCs.

3.3. The genesis frequency difference of clusters

To investigate variability in spatial distribution of TC genesis for the six simulated track clusters, the TC genesis frequency difference between observed and simulated TCs (model minus observation) is examined (Figs. 7a−g). In the Fig. 7g, TC genesis frequency of all TCs increased in the SCS (10°−20°N, 115°−125°E) and the south eastern part of WNP (10°−20°N, 140°−160°E) and decreased in the ECS (15°−25°N, 120°−130°E) and the PS (10°−20°N, 130°−140°E) respectively. Looking into each cluster, there have slight differences between observation and model in the C3, C4 and C6 (Figs. 7c, d and 7f) but remarkable differences are shown for the C1, C2 and C5 in the genesis difference map (Figs. 7a, b and 7e).

In the genesis difference map of C1, mostly negative values are distributed over the WNP and the simulated genesis frequency of C1 is significantly decreased in the ECS and PS. In contrast, those of C2 have mostly positive values and intensely increased in the south eastern part of WNP. These discrepancies are coincident with the climatology of TC genesis frequency. In the PS, where simulated TCs of C1 mainly formed, the simulated genesis frequency decreased. And the simulated genesis frequency increased in the south eastern part of WNP where mean genesis region for TCs of C2. The

remarkable distinctions in TC genesis frequency of C1 and C2 can be informed as dominant changes in the variability of simulated genesis frequency.

Although similar in the percentage between observation and model (20.2% for observation vs 20.8% for model), TC genesis frequency of C5 is clearly increased in the SCS and decreased in the PS (Fig. 7e). The dipole pattern in the SCS and the PS is shown in the difference map and this means that the genesis location of C5 is shifted westward from PS to SCS without total TC number change.

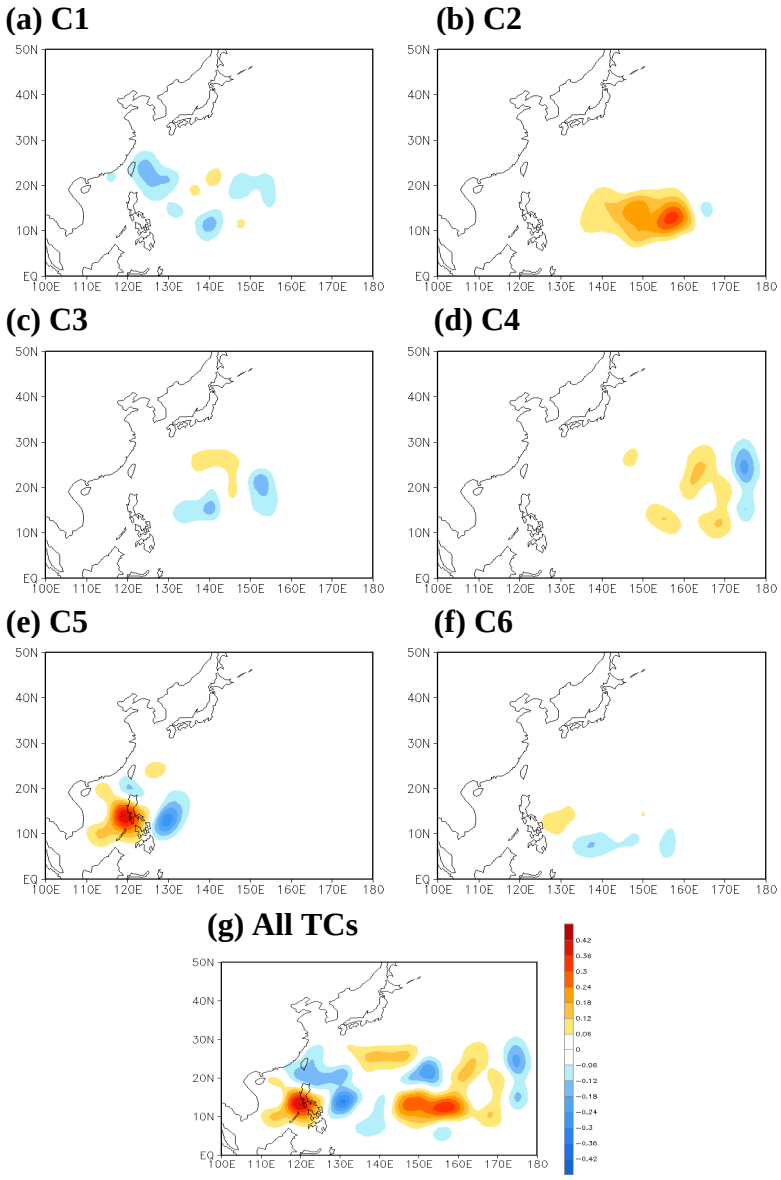


Figure 7. Difference (WRF minus RSMC) of TC genesis frequency in unit grid (5° × 5°) per year during TC season (JASO) in 1982–2010.

3.4. Interannual and Seasonal variability

TC activities vary at a wide variety of time scales, ranging from diurnal to interdecadal [Ho et al., 2004] and there is distinguishable seasonality not only for TC genesis but also for movement of TC. For climatological sense, TCs moving straight revealed in early of late TC season. In contrast, recurving TCs are shown in the peak season. Because this study focuses on the seasonal TC activity for the long-term period, interannual and seasonal time series of six clusters between observation and simulation are examined (Figs 7a, b, Figs 8a–g).

In the figure 7a, all simulated TCs have similar interannual variability to observed TCs and correlation coefficient between observation and model is 0.5. According to the each cluster, correlation coefficient of interannual variability between observation and model is insignificant except for C2 (figure not shown). This is because all TC tracks are classified to six clusters regardless of time concept. FCM is considered the shape, length of TC tracks and distance between individual track and cluster center. However, only C2 shows significant correlation coefficient between observation and model (0.68). Time series of simulated TC genesis follow reasonably those of observed TCs. The significant correlation between observed and simulated C2 is due to the El Niño

event. Kim et al. [2011] found that the tracks of C2 are related to the El Niño phenomenon. The significant relationship between interanuual variations in TC genesis frequency and El Niño / Southern Oscillation (ENSO) phenomenon over the WNP were shown in the previous studies [Chan, 1985; Chen et al.,1985 and Chia and Ropelewski, 2002]. The relationship with large-scale circulations will be discussed in next section.

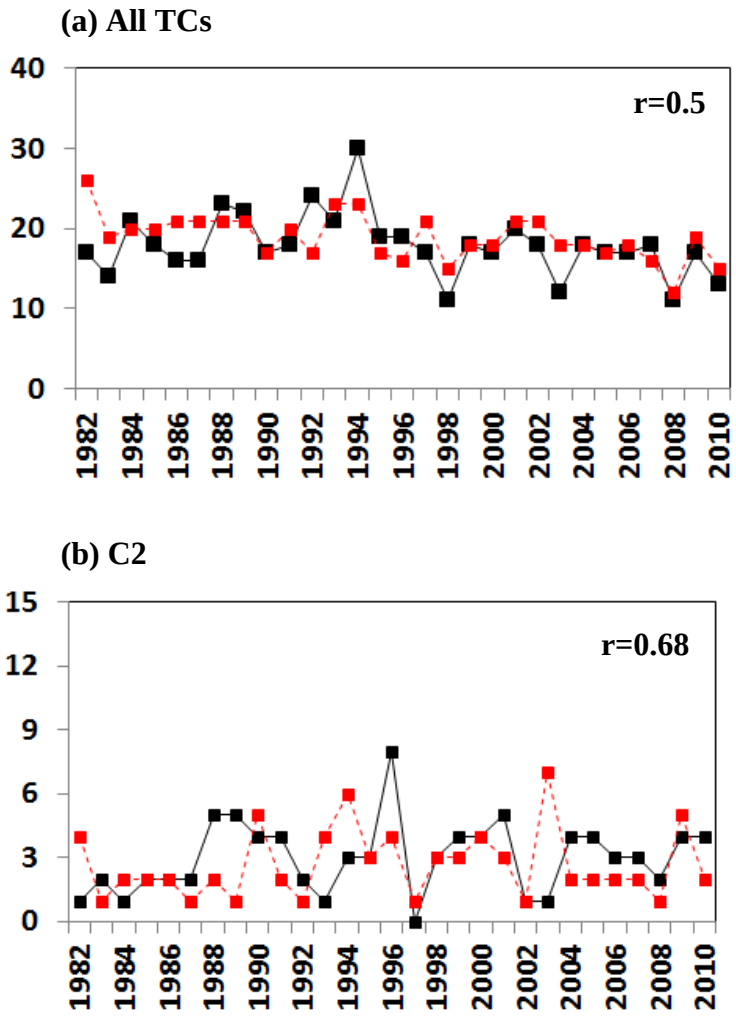


Figure 8. Time series of TCs for (a) all TCs and (b) C2 (black : observation, red : model).

There is distinguishable seasonality not only for TC genesis but also for movement of TC. The peak season for TC activity over the WNP covers summer and fall [Neumann, 1993]. For climatological sense, TCs moving straight revealed in early of late TC season in contrast, recurving TCs are shown in the peak season. And mean genesis location moves northward during summer, but retreats to the southeast in September. The seasonal variations of TC activity over WNP are highly related to climatological location of the monsoon trough in the Philippine Sea and NPSH retreating [Chia and Ropelewski, 2002].

Figure 9 shows monthly number per year for each cluster (Figs.9a−f) of observation and model. The clusters with recurving tracks (e.g., C1, C3, C4) are higher genesis frequency in August and September except for C2. TCs of C2 increase monotonically from July to October. TCs of C5 are mostly formed in the PS where deep monsoon trough locates. Therefore C5 shows the highest monthly number during the peak TC season when monsoon trough is developed most. TCs of C4 migrating in the open ocean are inactive in June. On the contrary to C4, C5 and C6 including straight-moving TCs are active in June.

As shown in the comparison between observation and model, the monthly distribution of all simulated TCs is similar to those of observation except for October (Fig. 9g). The TCs formed in October are overestimated in the model

than any other months. Comparing cluster by cluster, the simulated TCs of C1 decreased from July to October but those of C2 increased during the TC season. The simulated C5 with short straight-moving TCs remarkably increased in October. As mentioned above, the C5 is the cluster related to the monsoon trough and simulated TC genesis region of C5 is moved from the PS to the SCS. In the difference of monthly TC number in October may be associated with shift in TC genesis location. The monthly distributions of the simulated clusters (C3, C4 and C6) have similarities with those of observation.

The discrepancy in monthly distribution of C1 and C2 leads the seasonal variation of TC activity over WNP. The variability of monthly distribution for cluster can be explained by investigating the thermodynamic factors related to the C1 and C2. So, the variability of dominant clusters (C1 and C2) and relationship with large-scale environments will be discussed in the following section.

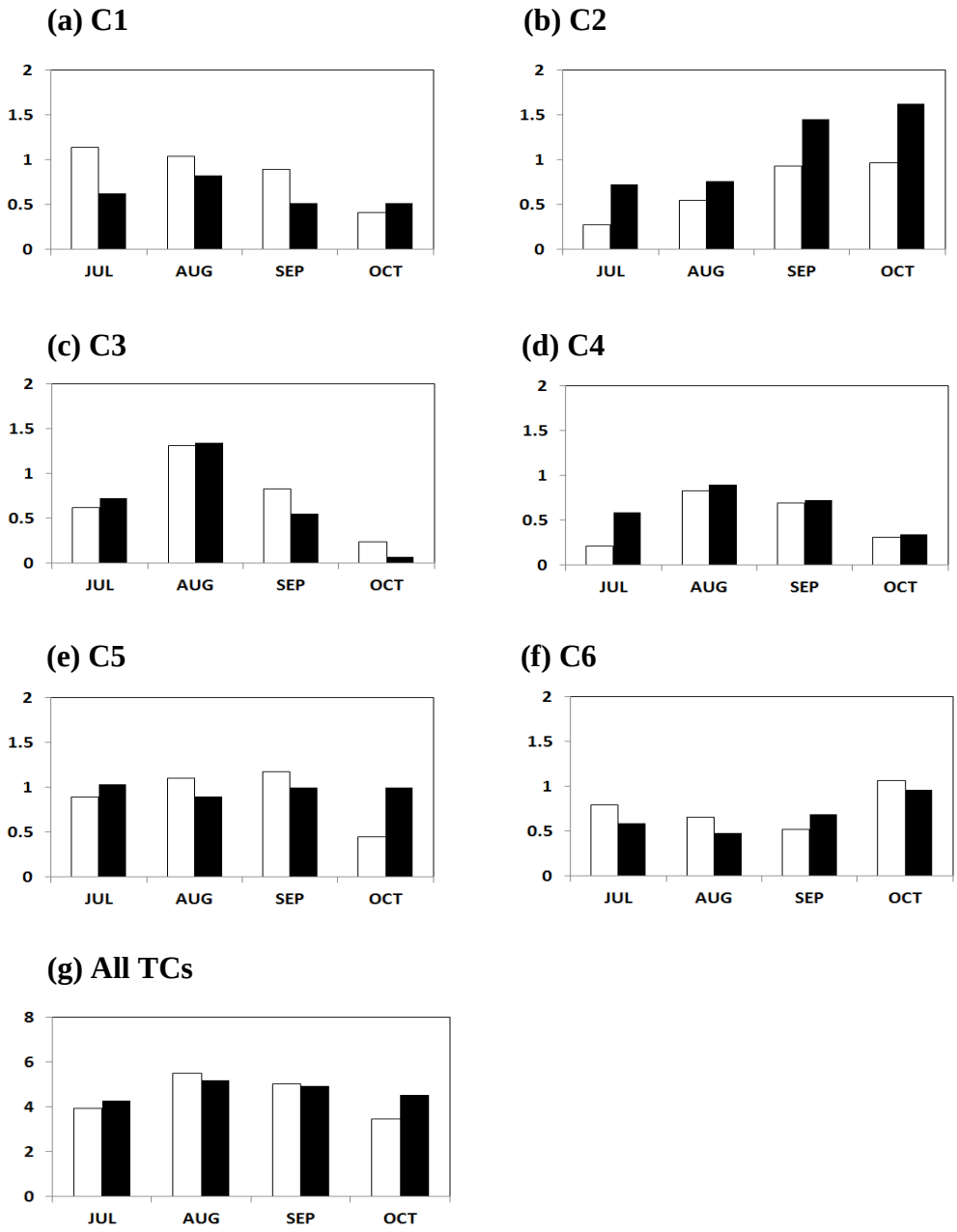


Figure 9. Monthly mean number of TCs for (a)–(f) six hard clusters and (g) all TCs (white : observation, black : model).

3.5. Landfalls of simulated TCs

Enormous economic damages and victims caused by TCs are determined by its landfall and it is highly dependent on the trajectories of TC. Therefore, seasonal forecasting of TC activity especially TC tracks is to determine the TC-prone regions possibly damaged by TCs. For estimating potential ability in seasonal forecasting of TC tracks, examining landfall regions of TCs is essential.

TCs formed in the WNP influence on costal region for Asian countries. For discussing by the countries, the Asia is divided into seven subareas based on the national boundaries to examine TC landfalls over the WNP. Figure 10 shows seven sub-regions those are Vietnam, Philippines, Taiwan, East China, South China, Korea and Japan. However for large territory of China, it is divided north and south along the 25°N line artificially (i.e., East China and South China).

It is different in the specific landfall region according to the TC track patterns. TCs of C1 mostly hit the East Asian countries such as East China, Taiwan, Korea and Japan. TCs of C2 influence on the south eastern Japan, C3 have TC tracks recurving to the East Asia particularly, Korea and Japan. TCs for C4 pass over open ocean east of Japan so, C4 does not landfall on the countries. TCs of C5 and C6 mostly hit on the Vietnam, South China and Philippines.

Table 3 shows the number of landfall TCs for the six clusters of observation and model. In this study, the landfall of TC is defined whether the TC center located on the coastal line and it is over-counted whenever one TC hit different countries. The simulated TCs of C1 mainly affecting on the East Asia (e.g., Korea, Japan, East China and Taiwan) decreases 47 TCs (among a total of 69 decreased ~ 68%) than those of C1. Among the countries hit by TCs of C1, simulated landfall TCs on the East China and Taiwan remarkably decreased than those of observation: 23 TCs decreased in the East China (among a total of 38 decreased ~ 60%) and 15 TCs decreased in the Taiwan (among a total of 18 decreased ~ 83%). However, simulated landfall TCs of C2 mostly hit the Japan are similar to observed landfall TCs despite of increase in TC number from 79 (observation) to 132 (model). This is because the most simulated TC passages of C2 concentrate east sea of Japan. The simulated landfall TCs of C3 and C4 have similarities with those of observation. Regardless of number difference, landfall TCs of C5 are less than observed TCs in the Vietnam (23 TCs decreased) but more in the Philippines (13 TCs increased) and landfall TCs of C6 increased 26TCs in the Vietnam because simulated TCs tracks of C5 and C6 have westward shorter lengths than those of observation. This result shows that most landfall TCs are reasonably reproduced except for landfall TCs of C1. The variability of simulated TCs for C1 influences on the distribution of TC landfall

on the East Asian countries.

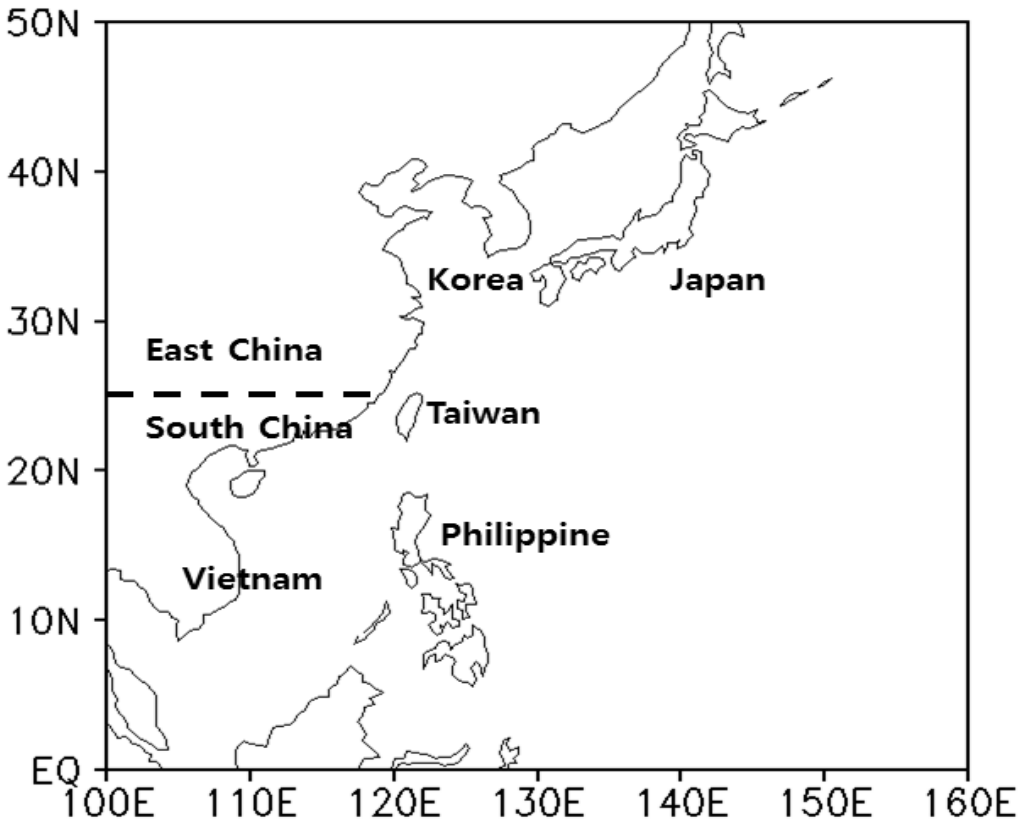


Figure 10. Seven sub-regions located in WNP area. (The Philippines, Vietnam, South China, East China, Taiwan, Korea and Japan).

	Cluster / Country	C1	C2	C3	C4	C5	C6	All
Observation	**Korea**	26	7	8	0	0	0	41
	Japan	24	25	41	1	0	0	91
	East China	46	2	1	0	24	25	98
	South China	9	0	0	0	65	41	115
	Taiwan	32	0	0	0	9	25	66
	Philippine	8	0	0	0	42	54	104
	Vietnam	0	0	0	0	50	32	82
Model	**Cluster / Country**	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**All**
	Korea	21	5	5	0	0	0	31
	Japan	21	29	35	2	0	1	88
	East China	22	5	1	0	16	16	60
	South China	2	0	0	0	50	35	87
	Taiwan	17	1	0	0	12	18	48
	Philippine	6	0	0	0	55	34	95
	Vietnam	0	0	0	0	27	6	33

Table 2. The number of landfall TC on Asian countries.

4. Relationship with Large-scale environments

4.1. Relationship between C1 and large-scale circulation

4.1.1. SST and low-level circulation

Gray [1978] and Mc Bride [1995] have investigated thermodynamic and dynamic factors making favorable conditions for TC formation over the WNP. These include SST providing midlevel moisture and conditional instability, positive low-level vorticity related to formation of the monsoon trough and weak vertical wind shear for developing TC. To investigate large-scale environmental influences in the region where genesis frequency of simulated TCs decreasing, membership coefficient weighted composites are performed on the day of TC genesis same with in Kim et al. [2011].

Figure 11 shows the composite anomalies of SST (shades) and wind at 850hPa (vectors) of observation (Fig. 11a) and model (Fig. 11b) for C1. Shades and vectors are denoted that the anomalies are at the 90% significant level. And black dot indicates mean genesis location of TCs for C1. The negative SST anomalies distribute over the central and eastern Pacific which is appeared during La Niña event (Fig. 11a and 11b). However in the mid-latitudes, the

positive SST anomalies are appeared from the Yellow Sea to the Kuroshio extension. It is noted in Kim et al. [2011] that may be caused by a low-level anti-cyclonic wave response and TCs mainly move to the East Asia due to this anti-cyclonic wave. In the simulated C1, positive SST anomalies in the mid-latitudes are underestimated than observed C1 and simulated TCs moving toward the East Asia decreased than observation.

The favorable effect of cyclonic circulation at the low-level for TC formation is discussed in many studies. Chia and Ropelewski [2002] noted that the positive relative vorticity makes TC occurrence in the PS where low-level cyclonic flows exist. And it is referred that low-level southwesterly winds to the south of monsoon trough axis give positive impact on the tropical cyclogenesis. In the Figure 11a, there are significant low-level cyclonic flows around the mean genesis location of observed TCs. Southwesterly low-level winds are advected to the mean genesis location in the observation. However, simulated C1 shows very weak low level flows around mean genesis location where monsoon trough distributes (Fig. 11b). Underestimated low-level cyclonic flow for simulated C1 forms relatively less TCs than those of observation over the PS.

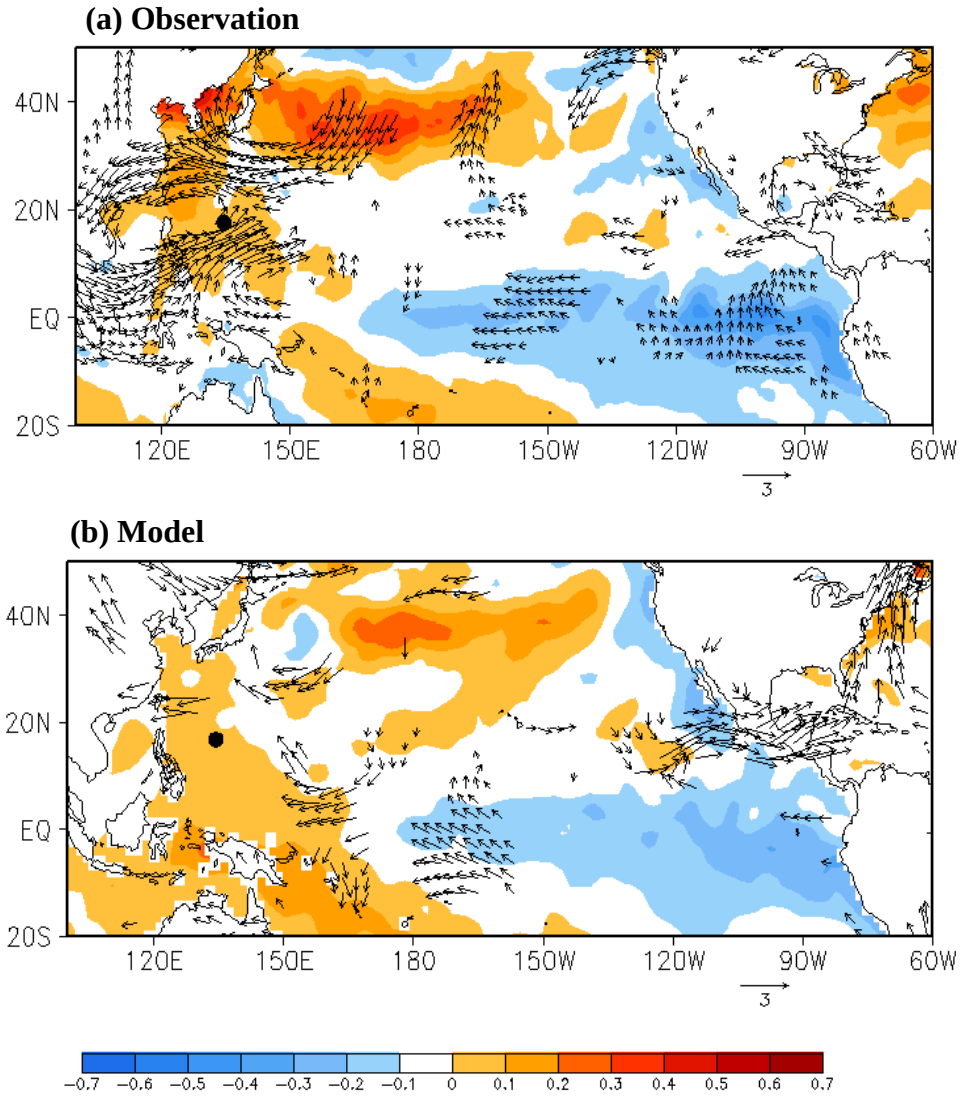


Figure 11. The membership coefficient weighted composites of the SST anomaly (shading), 850hPa wind anomaly (vector) based on the day of TC genesis (a) observation and (b) model for C1. Only significant values at the 90% confidence level are drawn.

4.2.2. Vertical wind shear

It is known that strong vertical wind shear have negative correlation with the TC formation and intensification in previous studies [e.g., Gray, 1968; Zehr, 1992; Corbosiero and Molinari, 2002]. The wind shear transfers the warm upper-level air out of the TC center and raises the pressure at the low level and it makes the TC to tilt downshear. It causes the TC can't maintain its vertical structure in the region where vertical wind shear is strong. Therefore, strong vertical wind shear blocks the TC development and intensification at all stage of TC life cycle including its genesis. To investigate the relationship with variability of decreasing simulated TCs for C1 and unfavorable atmospheric factor for TC formation, composite analysis for vertical wind shear is conducted.

Figure 11 shows the membership coefficient weighted composites of vertical wind shear between 200hPa and 850hPa for (a) observed and (b) simulated C1. The black dot represents the mean genesis location of TCs for C1 same in Figure 10. In the model, vertical wind shear is overestimated than observation near the mean genesis region. Zehr [1992] found that tropical cyclone cannot sustain themselves in environments where the vertical wind shear between 200hPa and 850hPa is greater than 12.5ms^{-1}. Comparing observation and model, the vertical wind shear is 8–10ms^{-1} around the mean genesis location in the observation but 14–16 ms^{-1} in the model (Fig. 11a vs Fig. 11b). Therefore

overestimated vertical wind shear resists the formation and the development of simulated TCs for C1 over the PS.

Moreover, vertical wind shear in the mid-latitude, especially in East Asian countries (e.g., East China, Korea and Japan), is remarkably stronger than observation. This may be the factor which weakens simulated TCs of C1 in the mid-latitude and blocks simulated TCs moving northward to the East Asia. This difference in the vertical wind shear of C1 influences on decreasing in number of simulated TCs during the TC season and less TC landfalling on the East Asian countries.

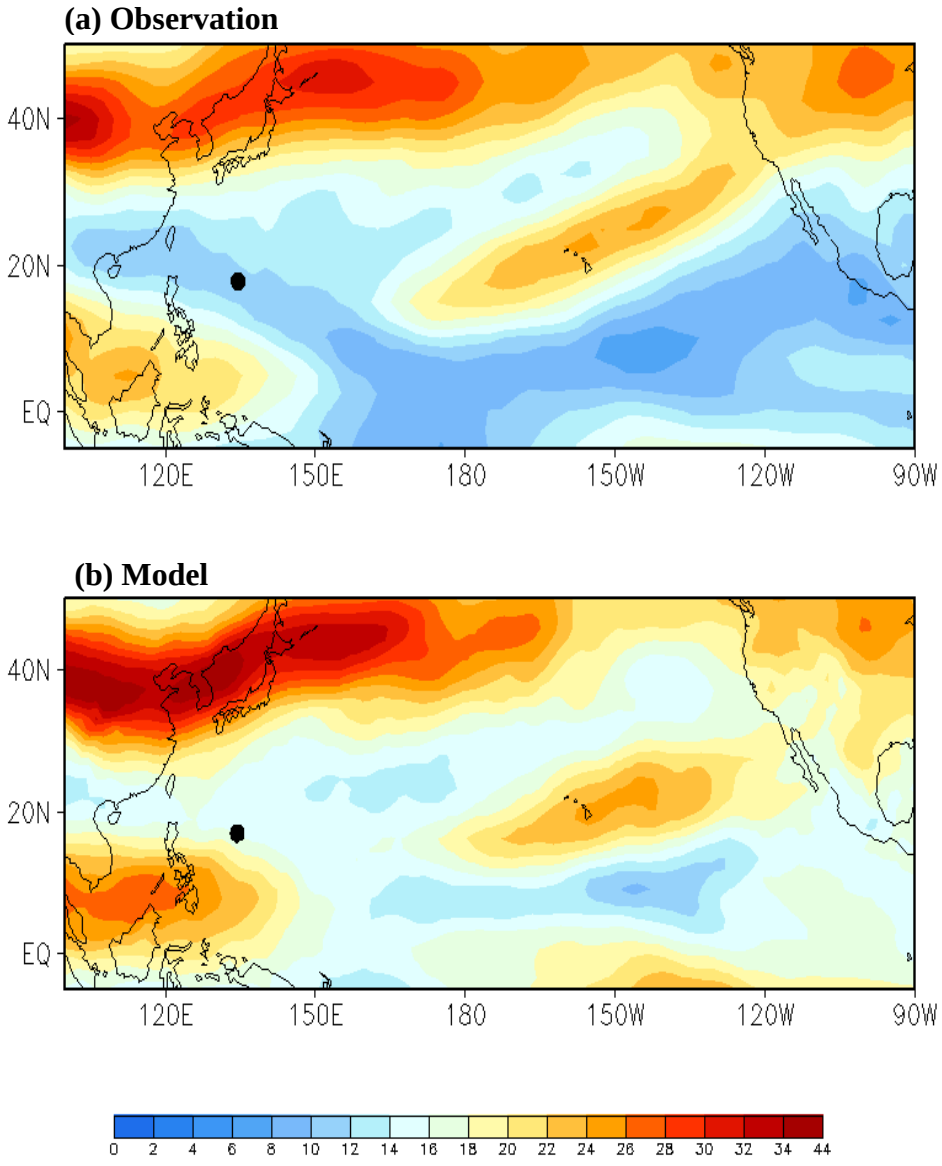


Figure 12. The membership coefficient weighted composites of the vertical wind shear (shade) on the day of TC genesis of (a) NCEP and (b) WRF for C1.

4.2. Relationship between C2 and large-scale circulation

4.2.1. SST and low-level circulation

The atmospheric and oceanic environments related to the variability of C2 are analyzed in this section. Figure 13 shows the composite anomalies of SST and low-level wind of (a) observation and (b) model for C2 same in the Figure 10. Positive SST anomalies are shown in the range of 0.1℃ to 0.5℃ over the equatorial central and eastern Pacific, which means C2 is the track pattern related to the El Niño. The distribution of SST anomalies of simulated C2 is similarly reproduced to those of observation.

The low-level cyclonic winds are significant around the mean TC genesis location in both observation and model. These cyclonic winds cause the shear vorticity favoring TC genesis. Furthermore, the anomalous westerly winds distributing south of mean genesis location is shown where the zonal SST gradient is strong. These flows are stronger in the boundary region between the positive and negative SST anomalies and TCs of C2 are mainly formed there.

In the midlatitude, south easterly low-level winds are shown north-west of mean genesis location of simulated TCs. These flows help TCs for C2 move to the south east sea of Japan after formation. However in the observation, these south easterly flows are relatively weaker than the model and anomalous north

easterly winds are shown in the south-west sea of Japan. These disturb observed TCs moving to the south-eastern part of Japan. Therefore, the track density of simulated TCs for C2 is higher than observation in the south east of WNP. The large-scale circulations related to TC steering will be discussed in the following section.

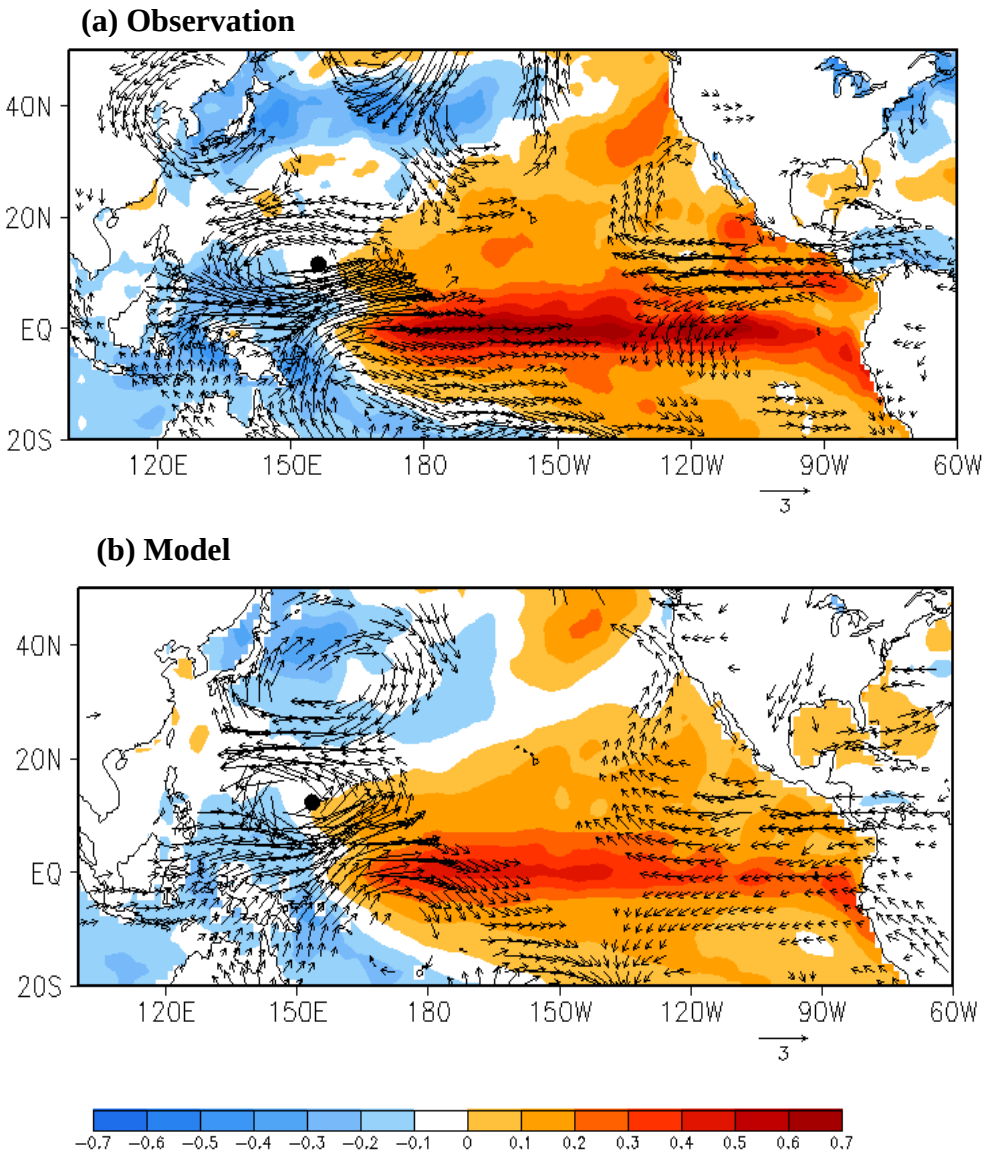


Figure 13. Same as Figure 11 (a, b) for C2.

4.2.2. The monsoon trough

The monsoon trough is one of the most important factors for tropical cyclogeneis over the WNP [Gray, 1968; Ramaze, 1974; Ritchie, 1995; Briegel and Frank, 1997; Chia and Ropelewski, 2002]. Low-level equatorial westerly or southwesterly winds from monsoon flow and the subtropical easterly trade winds from ITCZ derive to the large-scale cyclonic vorticity in the monsoon trough. Mentioned above, the cyclonic vorticity gives favorable condition for TC formation. TCs form near the eastern end of the monsoon trough and ITCZ between the monsoon westerlies and the easterly trade winds.

Atkinson [1977] mentioned mean axis of the monsoon trough in the WNP has a strong seasonality and it shows monthly mean position of monsoon trough during June to November (Fig. 14). The monsoon trough locates highest latitude in August and lowest in November however axis of trough extends easternmost in November. It is closely related to the NPSH retreating during autumn season [Lau and Li, 1984; Wang and Wu, 1997]. When the NPSH retreats from the SCS to the PS, the monsoon westerly winds go through the SCS and PS [Frank, 1987]. So, the movement of the monsoon trough with North Pacific subtropical High (NPSH) extending is closely related to the monthly mean genesis position.

To understand the relationship between variability of TCs for C2 and

monsoon trough, 850hPa zonal wind is composited on the TC genesis day (Fig 15a and b). The low-level zonal winds can show the confluent zone from the location of monsoon westeries and easterly trade winds where favoring TC genesis and the monsoon trough axis can be found north of low-level westeries. So, positive values denote the westerly zonal winds and negative values mean easterly zonal winds. In the region for zero value represents the eastern boundary of monsoon trough axis. Low level zonal flows over the SCS and PS shows well-developed monsoon shear line associated with the high cyclonic vorticity developing the TC [Ritchie and Holland, 1999]. In the Figure 15a and b, eastern boundary of the monsoon trough locates near the 160°E where TCs of C2 are generated mostly. However, the simulated zonal winds are stronger than observation in the SCS and PS and the gradient of low-level zonal wind for simulated C2 is stronger around the confluent zone. Larger zonal wind gradient between westerly and easterly winds induces stronger cyclonic vorticity for cyclogenesis in the confluent region. Lastly, simulated TC can be formed easier than observation in the confluence zone. This is coincident with increasing simulated TC genesis frequency for C2 near the 160°E.

After formation, TCs are steered by south easterly wind and move along the west boundary of subtropical ridge in the midlatitude. The west end of easterly wind between 20°N and 30°N represents the west boundary of subtropical ridge

axis. Therefore, the west end of subtropical ridge shows the recurving point of TCs for C2. In the simulation, west boundary of easterly wind locates eastern than observation. This indicates that simulated TCs can recurve eastern than those of observation (Fig 15a vs 15b). Strong westeries in the model are affected by weaken subtropical ridge. The retreating NPSH and its seasonality related to TC recurving for C2 will be investigated in next section.

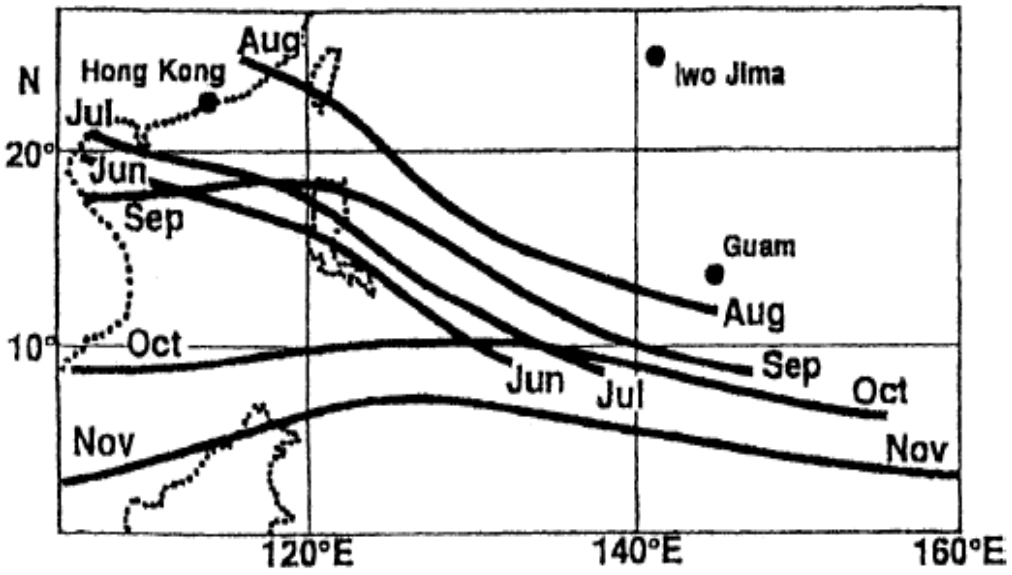


Figure 14. The monthly positions of monsoon trough during June-November from Atkinson [1971].

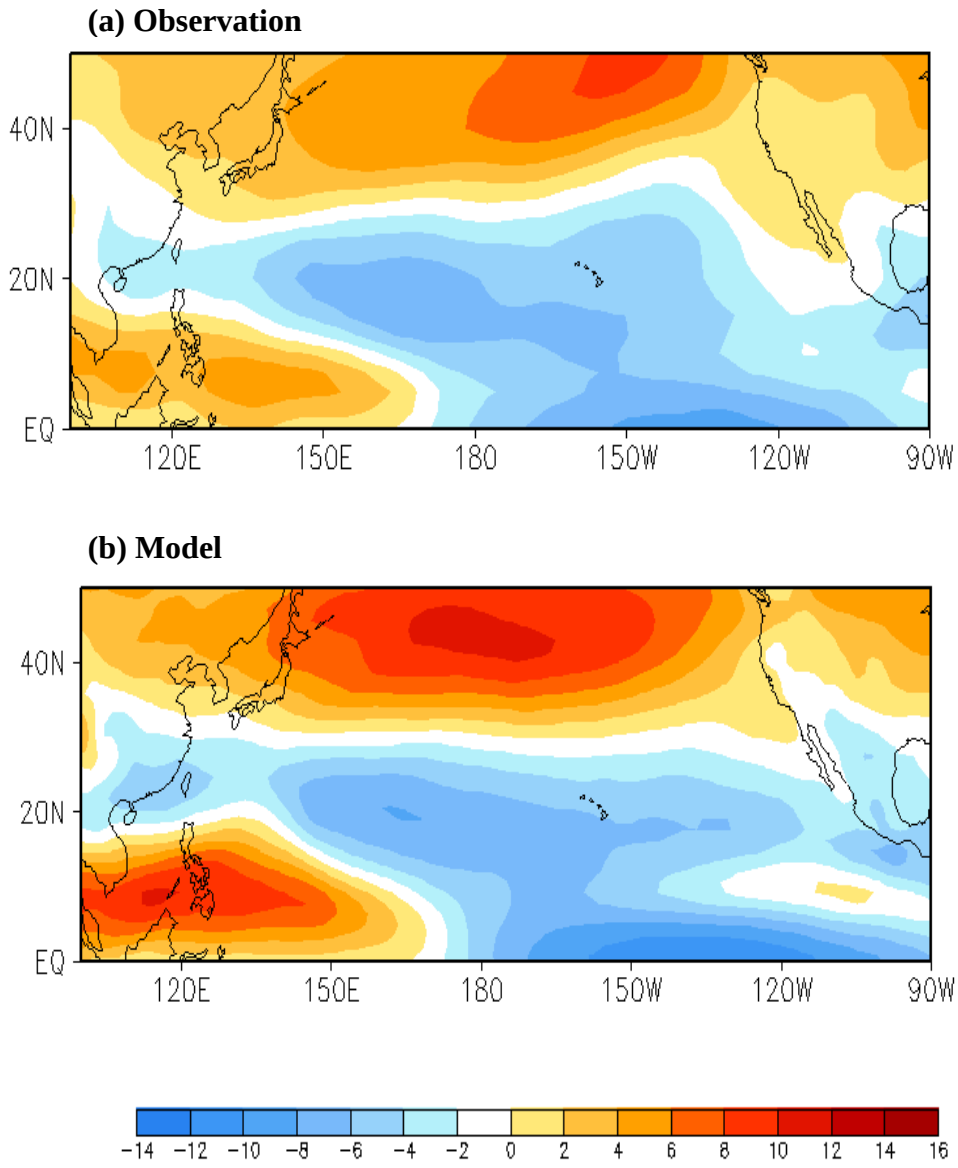


Figure 15. The membership coefficient weighted composites of the 850hPa zonal wind (shade) on the day of TC genesis of (a) NCEP and (b) WRF for C2.

4.2.3. Steering circulation

It has known that a tropical cyclone move in response to the other environmental systems for over a century. Holland [1993] mentioned 70 to 90 percent of TC motion is influenced by environmental steering. As TC moving in few days, atmospheric and oceanic environments keep interacting with movement of TC. The most approach to the study of TC motion is to consider the average wind through a deep layer of the atmosphere. This tropospheric layer mean flows in other word, the steering flow is calculated as pressure-weighted averaged horizontal winds from 850hPa to 200hPa level. To investigate relationship between steering flow and movement of TC, membership coefficient weighted composites on TC genesis day is conducted. For interacting between environmental flow and TC motion, examining composite map between large-scale circulation and the day of TC genesis is not perfect to explain the entire mean track of each cluster. However, composites of individual events for long period can show synoptic environmental characteristics maintained for several days climatologically. Therefore, composites on the TC genesis day represent large-scale background related to mean track of TC [Kim et al., 2011].

In the Figure 16, the composite map of steering flow (streamline) and NPSH (red contour) drawn as 5880gpm line at 500hPa for (a) observed and (b)

simulated C2 are shown. Shades denote the positive low-level relative vorticity and black line represents the mean track of TCs for C2. Comparing model to the observation, simulated low-level relative vorticity is overestimated in the TC mean genesis region. This is related to the strong low-level westerlies with intensified monsoon trough and it gives favorable condition to form more TCs than observation in the model (Fig. 16a vs 16b).

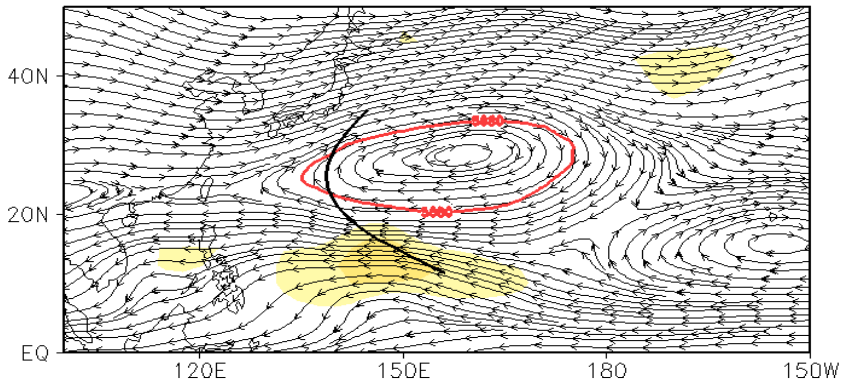
Another factor contributing the TC motion over the WNP can be proposed by the NPSH. TC moves along the south (e.g. straight moving TC) and the west (e.g. recurving TC) boundaries of NPSH after its formation. The NPSH expands westernmost in August and starts retreating in September. Therefore, TC trajectories are dependent on seasonal change of NPSH. TCs move straight in early summer and recurve eastern in late summer and autumn.

In the Fig. 16, TCs of C2 move along the west fringe of NPSH both in observation and model. However, the composite of simulated NPSH retreats eastward than observation and this causes TCs of C2 recurving earlier (eastern). It is confirmed in the Figure 17 showing TC recurving frequency. The recurving location is checked by the grid point where TC locates westernmost except for start and end points of its trajectories. In the observation, TC recurving frequency is the maximum in the ECS and simulated TCs mostly recurve north eastern PS so, simulated TCs recurve eastward than those of

observation. This discrepancy influences on the climatology of simulated TC track density increasing in the eastern WNP therefore, C2 is dominant cluster in variability in climatology of TC track density.

The mean location of NPSH expands toward west and retreats east seasonally. Therefore, regarding seasonal variation of NPSH, the variability in NPSH of C2 is related to the monthly distribution change of C2 TCs, as shown in the Fig 9b. The simulated TCs for C2 are increased in September and October than observation. This may be suggested that the NPSH retreating eastward in autumn season may make monsoon trough extend eastward and TC genesis of simulated C2 increased in September and October. And TCs formed in the east part of WNP may be steered toward eastern in September and October because of NPSH retreating eastern in autumn season.

(a) Observation



(b) Model

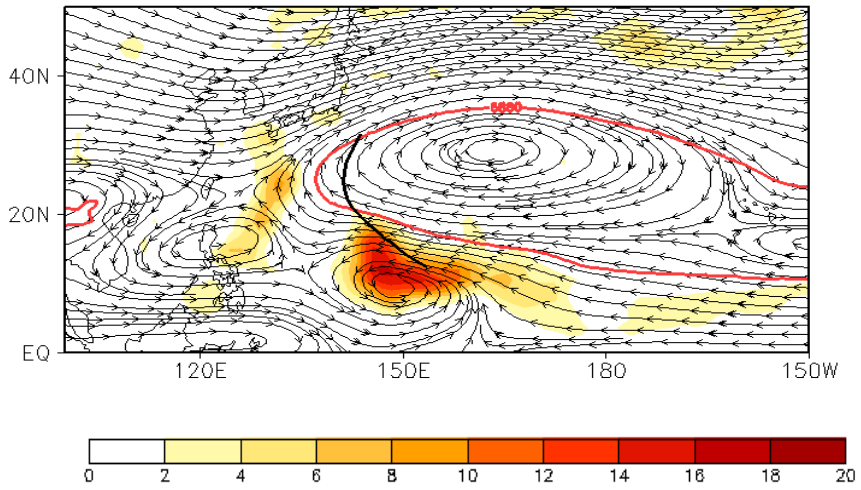
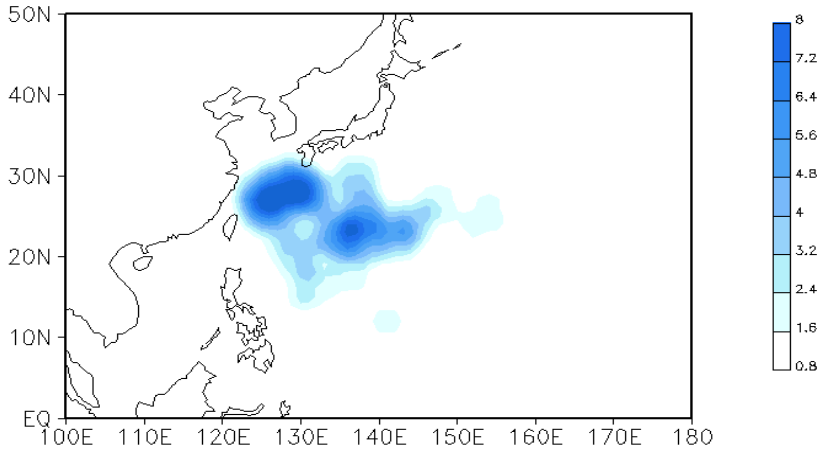


Figure 16. The membership coefficient weighted composites of the tropospheric layer-mean flows (stream line) and 850hPa positive relative vorticity (shade) on the day of TC genesis of (a) NCEP and (b) WRF for C2. The red contour is the 5880 gpm of the 500hPa geopotential height composite. Black solid line is mean track.

(a) Observation



(b) Model

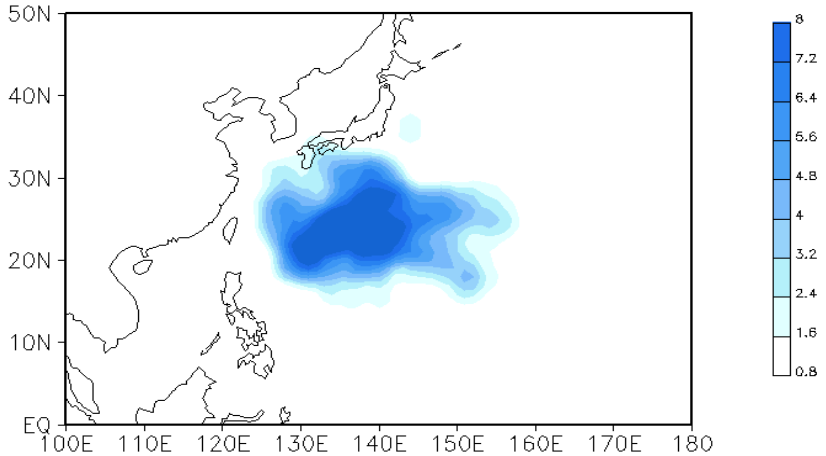


Figure 17. TC recurving frequency of (a) RSMC and (b) WRF for C2 in unit grid (5° × 5°) per year during TC season (JASO) in 1982–2010.

5. Summary and discussion

In this dissertation, the ability of regional climate model to reproduce the TC activity over the WNP is examined. Focused on the TC tracks, fuzzy clustering c-means method is used for investigating variability of TC tracks specifically and the relationship between variability of TC tracks and large-scale circulation is also discussed.

TCs are simulated with global and regional WRF model and WRF RCM is designed one way nested with 50-km horizontal resolution. For evaluating experiment, best track data from RSMC is used during TC season (JASO) for the period from 1982 to 2010. The model reproduced TCs similarly with a spatial pattern to the observation. However in climatology of TC genesis frequency and track density, there is difference in the maximum location. TC genesis frequency decreases in the PS but increases south-eastern of WNP and SCS. For climatology of track density, it decreases in PS but increases in south-eastern WNP.

For analyzing variability of TC tracks specifically, the 519 observed TC tracks are classified to six representative patterns using four validity objective measures (e.g., partition coefficient, partition index, separation index and Dunn index). The 549 simulated TCs are classified with same six clusters on the basis

of observation for comparing cluster by cluster. The simulated track clusters are similar to those of observation in pattern.

Comparing each cluster between observation and model, TCs for C1 decreased most from 101 to 72 and C2 TCs increased most from 79 to 132 in the model. Other simulated track patterns are similar to the observed track patterns. Each cluster has distinctive characteristics in mean values (e.g., central pressure, maximum wind speed, life time and mean genesis location) and model follows characteristics of observed track patterns. In spatial distribution, there are discrepancies between observation and model in TC genesis frequency. TC genesis frequency of all simulated TCs increased in the SCS and south eastern part of WNP but decreased in the PS. The simulated TCs for C1 are less formed in the PS but C2 TCs are formed more in the south eastern part of WNP in the model. TC genesis frequency of simulated C5 is higher in the PS but lower in the SCS. It shows the shift of genesis location with dipole pattern. Therefore, C1 and C2 are dominant clusters in variability of spatial distribution. For demonstrating interannual variability, the correlation coefficient between all TCs of model and observation is 0.5. However, only C2 has significant correlation coefficient between model and observation as 0.68. In the monthly distribution, all TCs are similar between observation and model except for October. The monthly number of simulated TCs for C1 decreases most in July.

For C2, monthly number of model is more than observation in September and October currently.

To understand relationship between dominant variability of TC track patterns and large-scale environment, membership coefficient weighted composites of thermodynamic and dynamic factors are applied for the TC genesis day. In anomaly composite of SST, C1 has negative SST anomalies in the East Pacific and it is La Niña related cluster. Weaker low-level cyclonic winds of simulated C1 around the mean genesis location give unfavorable condition for TC formation. And stronger vertical wind shear in the mean genesis location blocks the simulated TCs forming and developing than observation.

In contrast, positive SST anomalies are revealed in the East Pacific for C2. It means C2 is the cluster related to El Niño event. The monsoon trough is the important factor associated with TC formation in the WNP. The monsoon trough changes by season and it makes favorable condition for TC forming with relative vorticity and south westerly low-level wind to the its axis. For simulated C2, monsoon trough is stronger than observation and extends to the mean genesis location. Monsoon trough extending eastern most is associated with NPSH retreating in autumn season. This is shown in monthly number of simulated C2 increased in September and October remarkably. In addition, relative vorticity is strong around the mean genesis location of C2 and it

supports to form more TCs than observation.

The 70 to 80 percent of TC movement is influenced by tropospheric mean flow as steering flow. Another dynamic factor affecting TC motion over WNP is NPSH. The NPSH develops most in summer, retreats in autumn and recurving TCs move along the south and west boundary of NPSH. For simulated C2, NPSH retreats eastward than observation and it means that TCs recurve earlier (eastward) than observation. This is coincident with TC recurving frequency map of observed and simulated C2. TC In the model, recurving of C2 concentrates eastward than observed C2. In 5880gpm line composite of simulated C2, NPSH retreats eastward in autumn and it is reason for increasing number of simulated C2 in autumn (e.g. September and October) currently. And retreating NPSH helps maintain monsoon trough having its intensity and extend eastward.

The C1 and C2 are dominant clusters in reproducing the variability of seasonal TC activity (e.g. TC genesis frequency and TC track density) for a long time period. The bias of WRF RCM in simulating TC activity especially, TC tracks and relationship with large-scale environments are understood through this study by pattern classification. This examination helps improve the reproducing ability of WRF RCM and this results can contribute WRF RCM has better skill for forecasting the seasonal TC activity over the WNP further.

APPENDIX

A. The algorithm of Fuzzy *c*-means Clustering Method

$$J = \sum_{i=1}^{C} \sum_{k=1}^{K} (\mu_{ik})^m \, \|\boldsymbol{x}_k - \boldsymbol{c}_i\|^2, \qquad \text{(A1)}$$

where $\mu_{ik} = \left[\sum_{j=1}^{C} \left(\frac{\|X_k - c_i\|^2}{\|X_k - c_j\|^2} \right)^{\frac{2}{m-1}} \right]^{-1}$ and $c_i = \frac{\sum_{k=1}^{K} (\mu_{ik})^m X_k}{\sum_{k=1}^{K} (\mu_{ik})^m}$

FCM algorithm is on the basis of minimizing an object function ***J***. The $\boldsymbol{x}_k$ is the ***k***th data object and $\boldsymbol{c}_i$ is the *i*th cluster center, *C* is the number of cluster and *K* is the number of data objects. $\| \ \|$ indicates any vector norm representing the distance between cluster center and the data object. This study used the 2-norm (Euclidean norm) which is commonly used in the FCM. m is the fuzziness coefficient which represents overlapping degree between clusters. $\boldsymbol{\mu}_{ik}$ is the membership coefficient which means that how strong the ***k***th data belongs to the *i*th cluster. It varies from 0 to 1 depending on the distance between ***k***th data object and *i*th cluster center ($\|\boldsymbol{x}_k - \boldsymbol{c}_i\|^2$). To minimize the ***c***-means functional, J, it is subject to two constraints $\boldsymbol{\mu}_{ik} \geq 0$ and $\sum_{i=1}^{C} \boldsymbol{\mu}_{\boldsymbol{ik}} = 1$ $(i = 1, \dots, C, k = 1, \dots, K)$. The fuzziness coefficient, m denotes the degree of overlap between

clusters. As m is close to 1, $\boldsymbol{\mu}_{ik}$ converge to 0 for the objects far from a cluster center or 1 for those close to a cluster center, which implies less fuzziness (i.e. clearer cut) between clusters. Here m is set to 2 which is a common value used in the FCM. FCM allows the ambiguity of the data, therefore the cluster center ($\boldsymbol{c}_i$) represents membership coefficient-weighted mean of all data objects. Equal data lengths are required for all data objects, each track is interpolated into M segments. In this study M=20 segment is chosen (4times × 5days) with consideration for the mean TC life time formed in WNP.

B. Initialization of TC tracks for fuzzy clustering

To make data lengths equal, this study interpolate a TC track into M segment (M+1 data points) with equal length by leaving out time information. Despite this artificial interpolation eliminate the information of velocity, track shape is the most critical information for the FCM. The distances between 6-hourly segments of the best track data is defined as:

$$dist_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \text{ for } i=1,\ldots,N-1$$

Where (x_i, y_i) are ith longitude and latitude of the TC and N is the number of 6-hourly observed TC locations. The length of interpolated segments is edist $= \frac{1}{M}\sum_i^N dist_i$, where M is the number of interpolated segments. The interpolated positions $(\tilde{x}, \tilde{y})$ are calculated as :

$$\tilde{x}_j = x_1, \tilde{y}_j = y_1 \text{ for } j=1,$$

$$\tilde{x}_j = x_N, \tilde{y}_j = y_N \text{ for } j=M+1,$$

C. Iterative method for minimizing fuzzy c-means functional

Given the data set **x**, choose the number of clusters 1<C<K, the weighting exponent m>1, the termination tolerance ε>0, and the partition matrix **U**;

$$U=\begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,C} \\ \mu_{2,1} & \mu_{2,2} & \dots & \mu_{2,C} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{K,1} & \mu_{K,2} & \cdots & \mu_{K,C} \end{bmatrix} \quad \text{(C1)}$$

Initialize the partition matrix $U^{(0)}$ randomly.

Repeat following steps for l=1,2,….. until $\left\|J^{(l)} - J^{(l-1)}\right\| < \varepsilon$

Step 1. Compute the cluster centers:

$$\boldsymbol{c}_i{}^{(l)} = \frac{\sum_{k=1}^{K}\left(\mu_{ik}{}^{(l-1)}\right)^m \boldsymbol{x}_k}{\sum_{k=1}^{K}\left(\mu_{ik}{}^{(l-1)}\right)^m}, 1 \le i \le C \quad \text{(C2)}$$

Step 2. Update the partition matrix:

$$\mu_{ik}{}^{(l)} = \left[\sum_{j=1}^{C}\left(\frac{\left\|\boldsymbol{x}_k - \boldsymbol{c}_i{}^{(l)}\right\|^2}{\left\|\boldsymbol{x}_k - \boldsymbol{c}_j{}^{(l)}\right\|^2}\right)^{\frac{2}{m-1}}\right]^{-1} \quad \text{(C3)}$$

[Abony and Feil, 2007; Kim, 2011].

D. Validity scalar measure for the optimum cluster number

The partition coefficient [Bezdek, 1981] measures the amount of overlapping between the clusters. It is computed as:

$$\text{Partition coefficient} = \frac{1}{K}\sum_{i=1}^{C}\sum_{k=1}^{K}{\mu_{ik}}^{2} \quad \text{(D1)}$$

Partition coefficient is proportional to the entire average overlap between the fuzzy subsets inversely. The handicap of this index is that it is only based on the membership coefficients. Therefore, it is lacking in the direct connection to the geometrical properties of the object data.

The partition index [Bensaid et al., 1996] is given by the sum of the compactness ratio to the separation.

$$\text{Partition index} = \sum_{i=1}^{C}\frac{\sum_{k=1}^{K}{\mu_{ik}}^{m}\|\boldsymbol{X}_k - c_i\|^2}{\sum_{k=1}^{K}\mu_{ik}\sum_{j=1}^{K}\|\boldsymbol{X}_k - c_i\|^2} \quad \text{(D2)}$$

The partition validates compactness and separation of the clusters. The compactness is denoted with the mean of the distance between the data objects and the cluster center which is membership coefficient weighted. The separation is represented by the sum of the distances from a cluster center to all other cluster centers.

The separation index [Xie and Beni, 1991] is noted by the ratio of the compactness to the separation, similar to the partition index. The difference

with partition index is that it is defined as the minimum distance between the cluster centers. The separation index is defined as:

$$\text{Separation index} = \frac{\sum_{i=1}^{C}\sum_{k=1}^{K}\mu_{ik}{}^{m}\|X_k - c_i\|^2}{K min_{i,j}\|c_j - c_i\|^2} \quad \text{(D3)}$$

The Dunn index [Dunn, 1973] represents the ratio of shortest distance between the two objects belonging to each cluster and the largest distance between the two objects belonging to the same cluster. This is originally applied to hard partitions. However, it is applied after being assigned to a cluster where its membership coefficient is largest in this study. It is computed as:

$$\text{Dunn index} = min_{1\le i\le C}\left\{min_{1\le j\le C}\left\{\frac{min_{\boldsymbol{X}_i\in C_i,\boldsymbol{X}_j\in C_j}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|}{max_{1\le k\le C}\{max_{\boldsymbol{X}_i,\boldsymbol{X}_j\in C}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|\}}\right\}\right\}. \quad \text{(D4)}$$

Reference

Abony, J, and B. Feil, 2007: Cluster Analysis for Data Mining and System Identification. Birkhauser Basel, 303pp.

Atkinson, G. D., 1977: Forecasters guide to tropical meteorology. U.S. Air Force Tech. Rep. 240, Air Weather Services [Military Airlift Command (MAC)], 360pp.

A. Y. M. Au-Yeung and J. C. L. Chan, 2012: Potential use of a regional climate model in seasonal tropical cyclone activity predictions in the western North Pacific, Clim. Dyn., 39 (3−4) pp 783−794

Bensaid, A. M., L. O. Hall, J. C. Bezdek, L. P. Clarke, M. L. Silbiger, J. A. Arrington, and R. F. Murtagh, 1996: Validity-guided (re)clustering with applications to image segmentation. IEEE Trans. Fuzzy Syst., 4, 112−123.

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

Briegel, L. M., W. M. Frank, 1997: Large-scale influences on tropical cyclogenesis in the western North Pacific. Mon. Wea. Rev., **125**, 1397−1413.

Cha, D. H., C. S. Jin, D. K. Lee, and Y. H. Kuo, 2011: Impact of intermittent spectral nudging on regional climate simulation using Weather Research

and Forecasting model. J. Geophys. Res., 116, doi: 10.1029/2010jd015069.

Carmargo, S. J., and A. H. Sobel, 2005: Western North Pacific tropical cyclone intensity and ENSO. J. Climate, 18, 2996−3006.

Carmargo, S. J., A. W. Robertson, S. J. Gaffney, P. Smyth and M. Ghill, 2007a: Cluster analysis of typhoon tracks. Part Ⅰ: General Properties. J. Climate, 20, 3635−3653.

Carmargo, S.J., A. W. Robertson, S. J. Gaffney, P. Smyth and M. Ghill, 2007b: MJO effects. Geochem. Geophys. Geosy., 9, Q06V05, DOI:10..1029/2007GC001861

Chan, J. C. L.,1985: Tropical cyclone activity in the northwest Pacific in relation to the El Niño/Southern Oscillation phenomenon. Mon. Wea. Rev., 113, 599−606

Chan, J. C. L., and K. S. Liu, 2004: Global Warming and western North Pacific typhoon activity from and observational perspective. J. Clim., 17, 4590−4602.

Chen, T.-C., and J.-M. Chen, 1995: An observational study of the South China Sea monsoon during the 1979 summer : Onset and life cycle. Mon. Wea. Rev., 123, 2295-2318

Chen, T.-C., S.-P. Weng, N. Yamazaki, and S. Kiehne, 1998: Interannual variation in the tropical cyclone formation over the western North

Pacific. Mon. Sea. Rev., 126, 1080−1090

Chia, H.-H., and C. F. Ropelewski, 2002: The interannual variability in the genesis location of tropical cyclones in the Northwest Pacific. Journal of Climate, 15, 2934−2944

Chu, P.-S., 2004: ENSO and tropical cyclone activity, in Hurricanes and typhoons: Past, present, and potential, edited by R. J. Murnane and K. B. Liu, pp. 297−332, Columbia Univ. Press: New York.

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

Dong, K., 1988: El Niño and tropical cyclone frequency in the Australian region and the northwest Pacific. Aust. Meteor. Mag., 36, 219−255.

Dunn, J. C., 1973: A Fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters. Cybernet. Syst., 3, 32−57.

Frank, W. M., 1987: Tropical cyclone formation. A Global View of Tropical Cyclone, R. L. Elsberry et al., Eds., Naval Postgraduate School, 53−90.

Gray, W. M., 1968: Global view of the origin of tropical disturbance and storms. Mon. Wea. Rev., 96, 669−700.

Gray, 1979: Hurricanes: Their formation, structure and likely role in the general

circulation. Meteorological Society, 155−218

Jin, C.-S., C.-H. Ho, J.-H. Kim, D.-K. Lee, D.-H. Cha, S.-W. Yeh, 2012: Critical role of northern off-equatorial sea surface temperature forcing associated with central Pacific El Niño in more frequent tropical cyclone movements toward East Asia, J. Clim., (accepted)

Knutson TR, Sirutis JJ, Carnet ST, Held IM, Tuleya RE, 2007; Simulation of the recent multidecadal increase of Atlantic hurricane activity using un 18-km grid regional model. Bull Am Meteorol Soc 88(10), 1549−1565

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

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

Kim, J. -H., 2005: A study on the seasonal typhoon activity using the statistical analysis and dynamic modeling, Ph.D. thesis, 187pp, Seoul National Univ., Korea.

Kim, H. -S., J..-H. Kim, C. -H. Ho, and P. -S. Chu, 2011: Pattern classification of typhoon tracks using the fuzzy c-means clustering method, Journal of Climate, 24(2), 488−508

M. Stowasser, Y. Wang and K. Hamilton, 2006: Tropical cyclone changes in

the western North Pacific in a global warming scenario. Journal of Climate, 20, 2378−2396.

Neumann, C. J., 1993: Global overview. Global Guide to Tropical Cyclone Forecasting, World Meteor. Org., 1.1–1.56.

Oouchi, K., J. Yoshimura, H. Yoshimura, R. Mizuta, S. Kusunoki, and A. Noda, 2006: Tropical cyclone climatology in a global-warming climate as simulated in a 20-km-mesh global atmospheric model : Frequency and intensity analysis, J. Meteorol. Soc. Jpn., 84, 259−276

Park, D.-S. R., 2008: Decadal Oscillation of summertime tropical cyclone genesis frequency between the south China sea and the Philippine sea, Unpublished master's thesis, Seoul National University

Ritchie, E. A., and G. J. Holland, 1999: Large-scale pattern associated with tropical cyclogenesis in the western Pacific. Mon. Wea. Rev., 127, 2027−2043.

Skamarock, W. C., and Coauthors, 2008: A description of the advanced research WRF version 3. NCAR Tech. Note TN-475+STR, 88pp.

Wang, B., and J. C. L. Chan, 2002: How strong ENSO events affect tropical storm activity over the western North Pacific. J. Climate, 15, 1643−1658.

Wu, M. C., W. L. Chang, and W. M. Leung, 2004: Impacts of El-Niño Southern Oscillation events on tropical cyclone landfalling activity in the western

North Pacific. J. Climate, 17, 1419−1428.

Xie, X. L, and G. A. Beni, 1991: Validity measure for fuzzy clustering. IEEE Trans. PAMI., 3, 841-846.

Zehr, R.M., 1992: Tropical cyclogenesis in the Western Pacific. NOAA Tech. Rep. NESDIS 61, 181pp.

국문초록

본 연구는 지역기후모델을 이용하여 재생산된 북서태평양 태풍의 진로를 관측과 비교를 통해 모델의 모의성능을 알아보고, 대규모 순환장과의 상관성에 대하여 논의하였다. 모의된 태풍 진로의 변화를 구체적으로 알아보기 위해 퍼지군집분류 방법을 이용하여 6개의 대표적인 유형으로 분류하였다.

최적의 군집 개수는 4개의 객관적인 검증 지수를 통해 결정되었으며, 1982년부터 2010년 기간 동안 7월에서 10월에 발생한 519개의 관측 태풍의 진로는 6개의 대표적인 유형으로 분류하였고, 549개의 모의 태풍의 진로는 관측 태풍 진로의 군집을 바탕으로 하여 분류하였다. 이 6개의 군집은 동아시아 상륙하는 유형(C1), 긴 진로를 갖고 일본에 영향을 주는 유형(C2), 빨리 전향하여 일본 동쪽을 통과하는 유형(C3), 북서태평양의 가장 동쪽으로 이동하는 유형(C4), 남중국해에서 짧게 직진하는 유형(C5) 그리고 필리핀을 가로질러 직진으로 이동하는 유형(C6)로 나뉘었다.

관측태풍과 모의태풍의 진로 군집을 비교하였을 때, C1과 C2가 가

장 현저한 비율(개수) 변화를 보였다. C1은 19.5%(101)에서 13.1%(72)로 줄었으며 C2는 15%(79)에서 24%(132)로 증가하였다. 이러한 차이는 태풍의 발생빈도와 진로밀도의 공간분포의 변화에서도 두드러진다. 모의결과에서 태풍 발생빈도는 C1 태풍의 평균 발생 지역인 필리핀해에서 줄었고, C2의 주 발생지역인 북서태평양 남동부에서 증가하였다. 모의 태풍의 진로밀도는 필리핀해와 동중국해에서 관측보다 낮고, 북서태평양 동부에서 높았다. 나머지 군집들에서는 모델이 관측과 유사하게 모의하였다.

위에서 보았던 C1과 C2의 유의한 차이는 대규모 순환장으로 설명되었다. 위의 진로 군집의 변화와 대규모 순환장과의 관계를 살펴보기 위해 태풍 발생일과 합성분석이 수행되었다. C1은 라니냐 관련 군집으로서 필리핀해에서 과소 모의된 저고도의 저기압성 바람과 과대 모의된 연직바람시어로 인해 C1의 모의태풍의 발생이 줄었다. 하지만 엘니뇨와 관련된 C2에서는 강한 양의 상대와도와 몬순골의 확장으로 인해 모의 태풍의 발생이 강화되었다. 게다가 C2의 모의태풍은 북태평양 고기압의 후퇴로 빨리 전향하였다. 그 결과 모의 태풍은 관측보다 더 동쪽으로 이동하였다.

본 연구의 결과는 WRF 모델이 태풍의 진로를 모의하는데 있어 우

수한 성능을 가지고 있다는 것을 보여주었다. 다만 C1과 C2에서의 변동성을 이해함으로써 향후에는 모델의 성능을 개선하는데 기여할 것이다.

주요어 : 태풍, 진로, 군집 분류, 지역기후모델

학번 : 2009-22516