International Top Level Forum on Rapid Change Phenomena in tropical Cyclones, Hai Kuo, China, 5-9 November 2012

A Study on the Linear Regression Model for the Seasonal Tropical Cyclone Genesis Frequency in the Western North Pacific

Kamol P.N. Sakolnakhon[1], KiRyong Kang[2], KiSeon Choi[2]

1) Numerical Weather Prediction Sub-Bureau, Weather Forecast Bureau,
Thai Meteorological Department, 4353 Sukhumvit Road, Bangna, Bangkok, 10260
Thailand .
2) National Typhoon Center, Korea Meteorological Administration,

1622-1, Hannam-ri, Namwon-eup, Seogwipo, Jeju, 699-942, Korea

Abstract

This paper studied the seasonal tropical cyclone genesis frequency in the North West Pacific used the linear regression model for predicting. The data set used an eight climate variables: the sea surface temperate (SST.4), the sea surface temperature in the Western Pacific (SST.W), the temperature at 500hPa, the outgoing long wave radiation (OLR), the thermal wind at 850hPa, the thermal wind at 200hPa, the sea level pressure at Darwin at 1000 hPa and Tahiti based on June, July, August and September (JJAS) during period 1974-2004. The method developed applying a multiple linear regression technique used to investigate the tropical cyclone genesis. The preceding data was a different current year choose calibrations of the simulation model. The verification of simulation model based on during the period 2005-2009. The cross-validations were changed by the sea surface temperature in each month showed a good prediction which a quiet close of the month of research in May and a weak prediction in January. The significance of the sea surface temperature was changed in each month that is the importance of the tropical cyclone genesis frequency in the region. Therefore, the simulation model exhibits potential for application for forecasting the seasonal tropical cyclone genesis frequency in the Western North Pacific.

Key words: The regression model, Seasonal Tropical cyclone genesis, Tropical cyclone model.

1. Introduction

A Pacific typhoon or tropical storm is a tropical cyclone that develops in the Western North Pacific Ocean. In this area is no official typhoon seasons as tropical cyclones form throughout the year. The tropical storm frequency between May to December, and they are low frequency information between January and April. The tropical cyclones are winds blowing around a central area of low atmospheric pressure. The tropical cyclones are called hurricanes or typhoons in the Northern hemisphere, and the winds blow in an anti-clockwise circle. The tropical storms are known as cyclones in the Southern hemisphere and the wind direction are clockwise circle. Cyclones develop over warm seas near the Equator. The air heated by the sun rises very swiftly which creates areas of very low pressure. In a cyclone to develop, the sea surface must have a temperature of at least 26.5ºC. When warm air rises from the sea and condenses into clouds, the massive amounts of heat are released. The result of this mixture of heat and moisture is often a collection of thunderstorms, and also which a tropical storm can develop [1][3]. Therefore, the sea surface temperature is one of importance variable to support tropical cyclone genesis and it had been examined by many authors. The relationship of statistics and seasonal tropical cyclone genesis in the Western North Pacific is obviously interesting that a paper used a regression model for the western North Pacific tropical cyclone intensity [2][3], and a paper studies the sea surface temperature in El Nino-Sothern Oscillation (ENSO) on tropical cyclone intensity[4][5]. Some paper used the regression equation to predict typhoon track, a technique for adjusting dynamical tropical cyclone motion forecasts is extended to the Hurricane and Typhoon Tracking then the results were able to reduce the systematic longitudinal/latitudinal errors at all levels [7][8].

The paper’s influence of the El Niño–Southern Oscillation (ENSO) on tropical cyclone intensity in the Western North Pacific basin is examined. Accumulated cyclone energy (ACE), constructed from the best-track dataset for the region for the period 1950–2002, and other related variables are analyzed. ACE is positively correlated with ENSO indices. These and other statistics of the inter annually varying tropical cyclone distribution are used to show that there is a tendency in El Niño years toward tropical cyclones that are both more intense and longer-lived than in La Niña years [4]. The paper suggests that ACE is correlated approximately as strongly with ENSO indices up to six months later (northern winter), as well as simultaneously. It appears that not all of this lead–lag relationship is easily explained by the autocorrelation of the ENSO indices, though much of it is. Inter annual variations in the annual mean lifetime, intensity, and number of tropical cyclones all contribute to the ENSO signal in ACE [15].

A statistical model is a multiple linear regression technique for predicting the intensity of tropical cyclones in the Bay of Bengal. The model parameters are determined from the database of 62 cyclones that developed over the Bay of Bengal during the period 1981–2000. The parameters selected as predictors are initial storm intensity, intensity changes during the past 12 hours, storm motion speed, initial storm, latitude position, vertical wind shear averaged along the storm track, vorticity at 850 hPa, divergence at 200 hPa and sea surface temperature (SST) [5][6]. This paper suggests that when the model is tested with the dependent samples of 62 cyclones, the forecast skill of the model for forecasts up to 72 hours is found to be reasonably good. The average absolute erro rs (AAE) are less than 10 knots for forecasts up to 36 hours and maximum forecast error of order 14 knots occurs at 60 hours and 72 hours. When the model is tested with the independent samples of 15 cyclones (during 2000 to 2007), the AAE is found to be less than 13 knots (ranging from 5.1 to 12.5 knots) for forecast up to 72 hours [13][15].

The paper used a new approach for forecasting the summer rainfall over the middle to lower reaches of the Yangtze River valley (YRV) and north China that were to define DY as the difference in any variable between the current year and the preceding year. YR is denoted as the seasonal mean precipitation rate over the middle to lower reaches of the YRV. After analyzing the DY of the atmospheric circulations in the boreal winter/spring that was associated with the DY of the YR, the predictors were identified. The forecast model for the DY of the YR was established and then applied to forecast the YR [9][10].

The paper was developed a multiple linear regression model for the seasonal prediction of the summer tropical cyclone genesis frequency (TCGF) in the western North Pacific using the three teleconnection patterns. These patterns are representative of the Siberian High Oscillation (SHO) in the East Asian continent, the North Pacific Oscillation in the North Pacific, and Antarctic Oscillations (AAO) near the Australia during the boreal spring (April–May). The statistical model is verified through the two analyses: (a) statistical method of cross validation and (b) differences between the high TCGF years and low TCGF years that is verification of the statistical model. From this paper suggests that the high TCGF years are characterized by the following anomalous features: Three anomalous teleconnection patterns such as anticyclonic circulation (positive SHO phase) in the East Asian continent, the pressure pattern like “north-high and south-low” in the North Pacific, and cyclonic circulation (negative AAO phase) near the Australia were strengthened during the period from boreal spring to boreal summer. Thus, anomalous trade winds in the tropical western Pacific (TWP) were weakened by anomalous cyclonic circulations that located in the subtropical western Pacific (SWP) in both hemispheres [13].

The statistical models for seasonal prediction of tropical cyclone frequency in the mid-latitudes of East Asia. A multiple linear regression model (MLRM) established to investigate the seasonal prediction of the summer tropical cyclones (TC) frequency in the mid-latitudes of East Asia and then analyzed its validity using large-scale environments. The 850-hPa geopotential heights of the preceding April in the open ocean east of the Philippines and in the Bering Sea were used as independent variables. In the low-frequency years predicted by the MLRM, there was a larger amount of sea ice around the Sea of Okhotsk during the preceding spring and its cooling effect continued into the summer [12]. In addition, topographic and geographic effects around the Sea of Okhotsk that results in the easy formation of cold air created an anomalous cold high over this region in the summer. As a result, the northerlies from an anomalous cold high around the Sea of Okhotsk caused cold surface air temperature anomalies in the mid-latitudes of East Asia, which played an important role in preventing a western Pacific subtropical high from advancing toward the mid-latitudes of East Asia. Eventually, these environments led to a reduced summer TC frequency in the mid-latitudes of East Asia [14].

In this study, we attempt to predict the seasonal tropical cyclones for the seasonal tropical cyclone genesis frequency in the Western North Pacific with climate factors focus on JJAS during period 1974-2004 by the linear regression model. The preceding year runs calibration of the model and verification model during 2005-2009. The cross-validations were changed by the sea surface temperature and also investigated model by statistics variables.

2. Data and Methodology

The data set used the climate data from any center: the Distributed Data Bases (DDB) of Japan Meteorological Agency (JMA), the National Oceanic and Atmospheric administration

(NOAA) and the annual tropical cyclone report from the joint typhoon warning center (JTWC). The method has been used eight climate parameters to investigate the seasonal tropical cyclone genesis frequency (SST.4, SST.W, the temperature at 500hPa, OLR, the thermal wind at 850hPa, the thermal wind at 200hPa, the sea level pressure at Darwin at 1000 hPa, the sea level pressure at Tahiti at 1000 hPa), let briefing the climate parameters to use:

2.1 The sea surface temperature the sea or ocean temperatures greater than 26.5 °C through a depth of at least 50 meters are generally favorable for the formation and sustaining of tropical cyclones. In general the higher SST that was the strongest storm and greater chance of genesis.

2.2 The outgoing long wave radiation Climatology of outgoing long wave radiation (OLR) provides information about the regions for which significant deep convection can be expected throughout the year.

2.3 The temperature at 500hPa Temperature determines the energy-level distribution of the particles in a system: the particle velocity distribution, the properties of the equilibrium electromagnetic radiation of bodies, the spectral density of the radiation and the total energy radiated per unit volume.

2.4 The thermal wind is the most fundamental and significant dynamical balance controlling the large-scale circulation of the atmosphere and ocean. It is a consequence of hydrostatic and geotropic balance, and relates horizontal buoyancy gradients to changes in the horizontal wind with height.

2.5 The sea level pressure is a force per unit area exerted by the air above the surface of the Earth. Standard sea-level pressure is defined as 1 atmosphere (atm.) equal 101.35 kilopascals, but pressure varies with elevation and temperature. It is usually measured with a mercury barometer (hence the term barometric pressure), which indicates the height of a column of mercury that exactly balances the weight of the column of atmosphere above it. It may also be measured using an aneroid barometer, in which the action of atmospheric pressure in bending a metallic surface is made to move a pointer. Atmospheric pressure at the earth's surface varies from place to place and with time. Of particular importance are the non periodic variations associated with the onset, development, and dissolution of slowly moving high-pressure regions called anticyclones and vast, relatively fast-moving vortices, or cyclones, in which low pressure prevails. The climate, weather usually measure the sea-level surface pressure difference between at Tahiti and Darwin to consider a year of El Nino and La Nin will come.

In climate parameters, they have many parameters concern tropical cyclone genesis. However, in this research, we should be selected eight parameters to use because of they are importance relationship tropical cyclone genesis frequency more other parameters that reason we are briefing of each parameter on above, and then we selected nine parameters to run tests from the same technique.

The model developed applying a multiple linear regression technique to use in the experiments which using eight climate variables in June, July, August and September (JJAS) based on 1974-2004. The preceding data were a different current year chose calibrations of the simulation model. In this experiment selected eight variables based on June, July, August and September (JJAS). The general a multiple linear regression equation is following:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + \varepsilon \qquad \text{(Eq. 1.1)}$$

Where $i = 1,2,3..,k$, Y is the dependent variable, X is the independent variable, β_0 is the intercept, β_i is slop or regression variable and ε is error variable.

The experiment design in table 1.1: Ex-1 is SST.4, Ex-2 is SST.W, Ex-3 is the temperature at 500hPa, Ex-4 is OLR, Ex-5 is the thermal wind at 850hPa, Ex-6 is the thermal wind at 200hPa, Ex-7 is the Sea Level Pressure at Darwin at 1000 hPa, Ex-8 is the Sea Level Pressure at Tahiti at 1000 hPa. DY is defining the difference any variances between current year (Model-1) and the preceding year (Model-2) for the forecast model of the Western North Pacific tropical cyclone genesis in seasonal are following table 1.1.

Table 1.1: The experiment design

Experiments	File name	Description	Boundary
Ex-1	SST.4	Avg. August of the sea surface temperature	5°S - 5°N, 160°E - 150°W
Ex-2	SST.W	Avg. August of the sea Surface temperature in the Western Pacific	Equator - 15°N, 130°E - 150°E
Ex-3	Temp_500hPa	Avg. August in the temperature at 500hPa	
Ex-4	OLR	Avg. June of the Outgoing Long Wave Radiation	Equator, 160E-160W
Ex-5	U_850hPa	Avg. August of the thermal wind at 850hPa	
Ex-6	U_200hPa	Avg. June, July, August, September of the thermal wind at 200hPa	
Ex-7	SLP_D	Avg. June, July, August, September of the sea level pressure in Darwin 1000 hPa	Darwin
Ex-8	SLP_T	Avg. June, July, August, September of the sea level pressure in Tahiti 1000 hPa	Tahiti
Model-1	Ex-1, Ex-2, Ex-3, Ex-4, Ex-5, Ex-6, Ex-7, Ex-8		A current year
Model-2	Ex-1, Ex-2, Ex-3, Ex-4, Ex-5, Ex-6, Ex-7, Ex-8		A preceding year
DY	A simulation model was calibrated		

3. Results

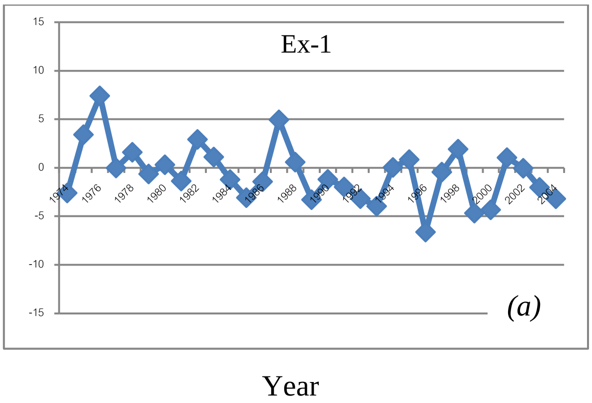
3.1. Regression Model From the multiple linear regression equation 1.1, we determined variable in the equation 1.1 by calculating with calculators or statistics software to help. Therefore, we can get the coefficient of multiple determination (R^2) is 0.6871, the correlation coefficient is 0.8289, the intercept(β_0) is 3,250.1908, the regression variables are $\beta_1 = 5.2123, \beta_2 = 7.0103, \beta_3 = -13.5323, \beta_4 = -0.0027, \beta_5 = 0.0866, \beta_6 = -0.4076, \beta_7 = -0.6701, \beta_8 = 3.7079$ and ε is error variable. In this experiment a small error, then defied ε is zero, and also we represent all variable in equation 1.1, then we can get equation 1.2.

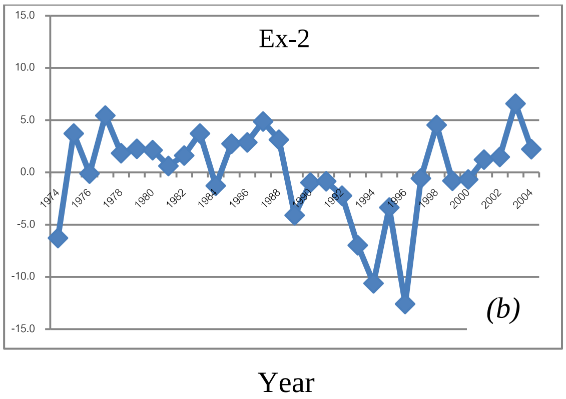
$$Y = 3,250.1908 + 5.2123X_1 + 7.0103X_2 - 13.5323X_3 - 0.0027X_4 + 0.0866X_5 - 0.4076X6 - 0.6701X7 + 3.7079X8 \qquad \text{(Eq. 1.2)}$$

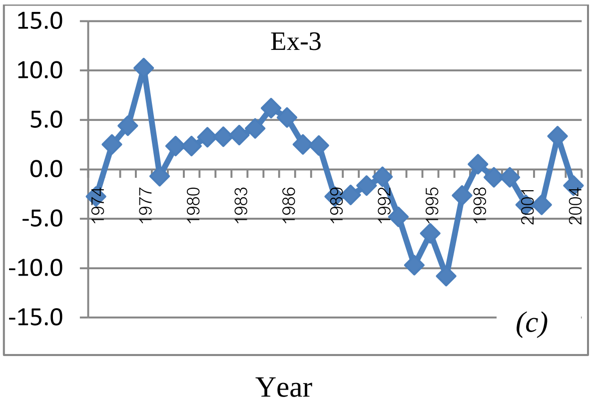
Where $i = 1,2,3..,8$, Y is the dependent variable, X is the independent variable are X_1 is average the sea surface temperature in August, X_2 is average the sea surface temperature in

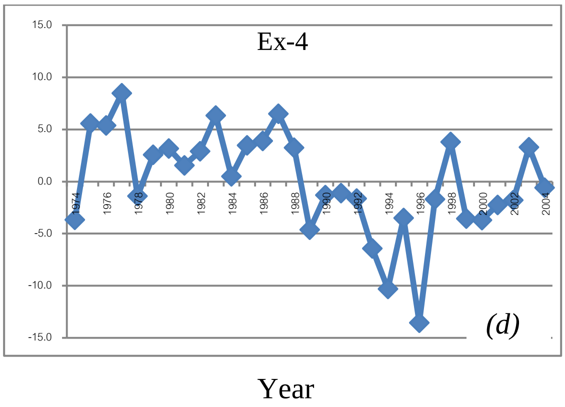
August at the Western Pacific, X_3 is average the temperature at 500hPa in August, X_4 is average the outgoing long wave radiation in June, X_5 is average the thermal wind at 850hPa in August, X_6 is average the thermal wind at 200hPa in JJAS, X_7 is average the sea level pressure at Darwin 1000 hPa in JJAS, X_8 is average the sea level pressure at Tahiti 1000 hPa in JJAS.

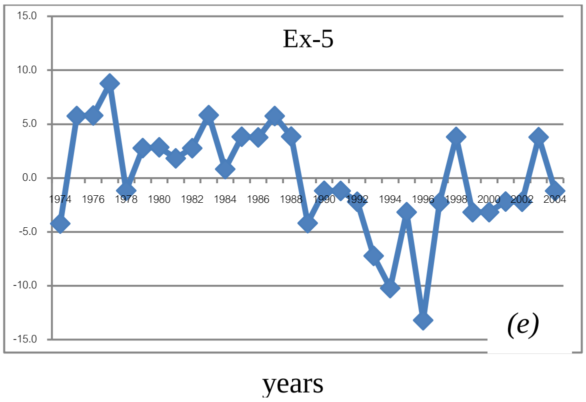
From the equation 1.2, we represent each the dependent variable related to observations are following figure 1.1. The results showed that a correlation with seasonal tropical cyclones during JJAS, the mostly variables suggest that have a quiet over predicts than observations in 1975-1988 and 2002-2004, and a quiet under prediction than observations during 1994-1996. The results suggestion that EX-1 is a better simulation than other that showed a quiet close. Therefore, the sea surface temperature is important in the typhoon genesis than others.

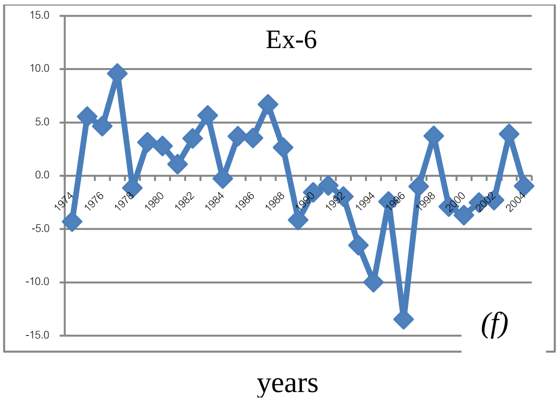












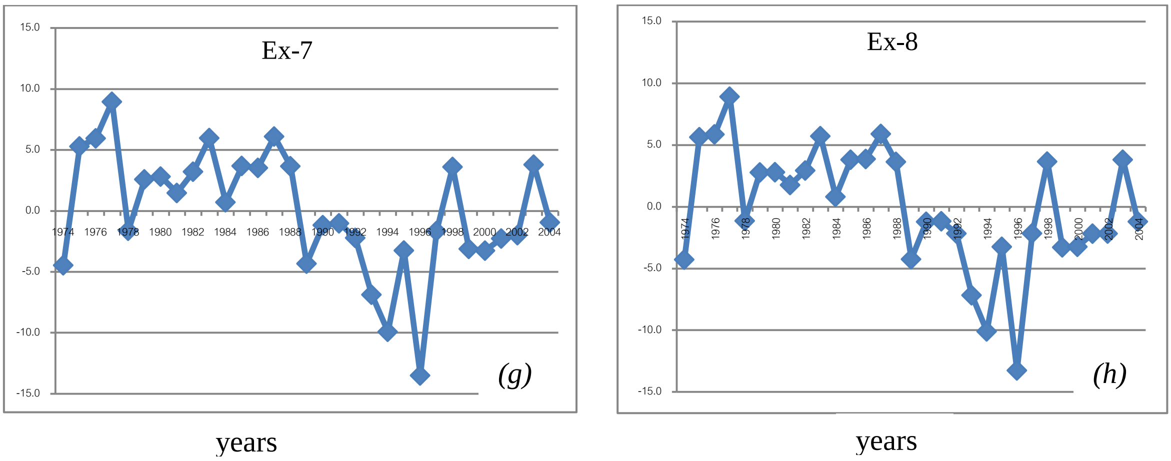


Figure 1.1: The time series of each dependent related to observations based on during the period 1974-2004 showed error values in *(a)-(h)*.

We represent all variables in a multiple linear regression equation 1.2 then we can get the results related to observations showed was a quit over predicted than observations in 1977(7), 2003(7), 1985(5), 1986(5), 1983(3), 2004(3), 1979(2), 1980(1), 1981(1), 1984(1), 1987(1), 1988(1), 1991(1), 1998(1) and 2000(1), and a quit under predicts than observations in 1994(10), 1996(8), 1995(6), 1974(3), 1976(3), 1989 (3), 1993(3) and 2001(2) are following table 1.2, and the time series errors of model related to observations in figure 1.2

Table 1.2: The simulation Model-1 related to observations based on 1974-2004.

Year	Observation	Model_1	Relative Error (%)	MAE	(-) under Predict (+) over Predict
1974	35	32	-11	3	-3
1975	25	25	0	0	0
1976	25	22	-11	3	-3
1977	22	29	22	7	7
1978	32	29	-9	3	-3
1979	28	30	7	2	2
1980	28	29	3	1	1
1981	29	30	5	1	1
1982	28	28	0	0	0
1983	25	28	9	3	3
1984	30	31	2	1	1
1985	27	32	16	5	5
1986	27	32	16	5	5
1987	25	26	5	1	1
1988	27	28	5	1	1
1989	35	32	-8	3	-3
1990	32	32	0	0	0
1991	32	33	2	1	1
1992	33	33	0	0	0
1993	38	35	-11	3	-3
1994	41	31	-32	10	-10

1995	34	28	-20	6	-6
1996	44	36	-25	8	-8
1997	33	33	0	0	0
1998	27	28	2	1	1
1999	34	34	0	0	0
2000	34	35	3	1	1
2001	33	31	-6	2	-2
2002	33	33	0	0	0
2003	27	34	21	7	7
2004	32	35	11	3	3

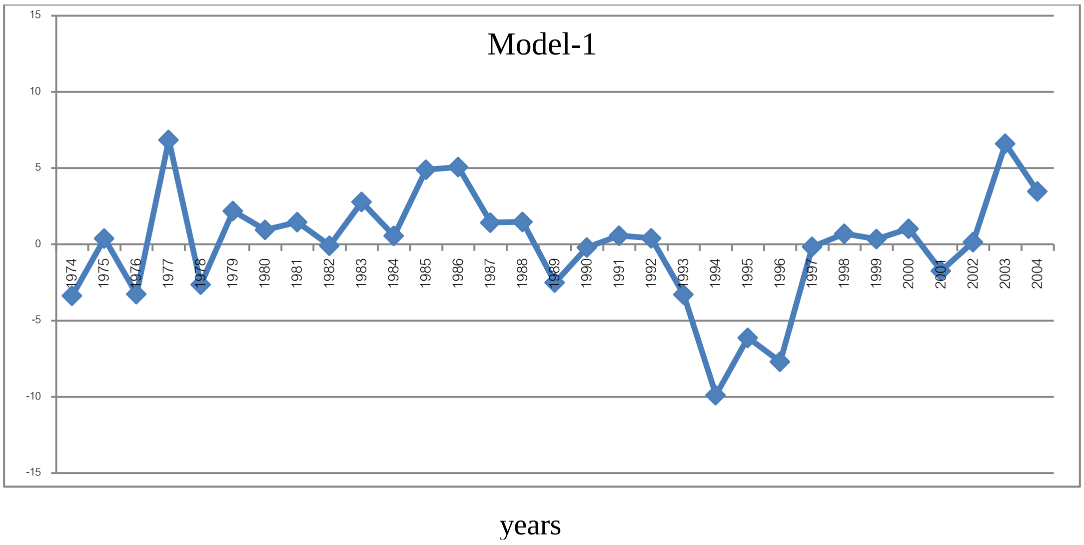


Figure 1.2: The times series of the simulation model errors related to observations.

3.2. Calibrations and Verification

a. Calibrations The preceding is Model-2 run for calibration Model-1 then we can get a forecasting model to predict the seasonal tropical cyclone genesis in the area to study that is DY. However, the simulation model must be investigated a quality model therefore we defied some technical to investigate as:

1. The simulated annual tropical cyclone frequency
   Let $\Delta y = y_1 - y_2$ is different forecasting in Model-1 and Model-2 in a forecast model DY, and let $DY = \Delta y + y_0$, where y_0 is observations,
2. A comparison the simulation model to observations so call the relative error:
   Let $\frac{y-y_0}{\overline{y_0}} x100\%$ where $y = \Delta y + y_0$ then $\Delta y = y - y_0$, $\bar{y}_0$ is the average of the observations.
3. The mean absolute error (MAE)
   Let $|y - y_0|$,
4. The average mean absolute error (MAE)

   Let $|y - y_0|/N$,

5. The average relative root mean square error of the simulation (RMSE).

   Let $\sqrt{\frac{1}{N}\sum_{i=1}^{k}(y-y_o)^2/\bar{y}_0}$

   Where N is the length of the sample,
6. The average absolute root mean square error of the simulation (RMSE)

   Let $\sqrt{\frac{1}{N}\sum_{i=1}^{k}(y-y_o)^2}$

   Where N is the length of the sample,

7. The coefficient regression determination (R^2),
8. The correlation coefficient,
9. The threat score (TS).

Table 1.3: A calibration statistics model process

Year	Obs	Model_1 (y_1)	Model_2 (y_2)	$\Delta y = y_1 - y_2$	$DY = \Delta y + y_0$	Relative error (%)	(+) over Pre. (-) under Pre.	MAE
1974	35	32	32	0	35	0	0	0
1975	25	25	25	0	25	0	0	0
1976	25	22	22	0	25	0	0	0
1978	22	30	29	1	23	3	1	1
1977	32	29	29	0	32	0	0	0
1979	28	30	30	0	28	0	0	0
1980	28	29	29	0	28	0	0	0
1981	29	30	30	0	29	0	0	0
1982	28	28	28	0	28	0	0	0
1983	25	28	28	0	25	0	0	0
1984	30	30	30	0	31	0	0	0
1985	27	32	32	0	27	0	0	0
1986	27	32	32	0	27	0	0	0
1987	25	26	26	0	23	0	0	0
1988	27	28	28	0	27	0	0	0
1989	35	32	32	0	35	0	0	0
1990	32	33	33	0	32	0	0	0
1991	32	34	34	0	31	0	0	0
1992	33	34	34	0	32	0	0	0
1993	38	35	35	0	38	0	0	0
1994	41	31	31	0	41	0	0	0
1995	34	28	28	0	34	0	0	0
1996	44	37	37	0	43	0	0	0
1997	33	33	33	0	33	0	0	0
1998	27	27	27	0	28	0	0	0
1999	34	34	34	0	34	0	0	0
2000	34	35	35	0	34	0	0	0
2001	33	32	31	1	34	3	1	1
2002	33	33	33	0	33	0	0	0
2003	27	33	33	0	28	0	0	0
2004	32	35	35	0	32	0	0	0

2005	25	32	31	1	26	3	1	1
2006	27	33	32	1	28	3	1	1
2007	26	29	28	1	27	3	1	1
2008	27	37	37	0	27	0	0	0
2009	28	29	28	1	29	3	1	1

The multiple linear regressions established to forecast tropical cyclone in the Western North Pacific in equation 1.2 related to observations during 1974-2009 was fugure 1.3 *(i)*.. The results showed the coefficient regression determination is (R^2) is 0.9943, the correlation coefficient is 0.9972, the avg. mean absolute error (MAE) is 0.2979, the root mean square error (RSME) is 0.5458 and the threat score (TS) is 0.7148 was fugure 1.3 *(j)*. The significant the simulation model and observations during calibration showed a good correlation and a quiet close, see figure 1.3.

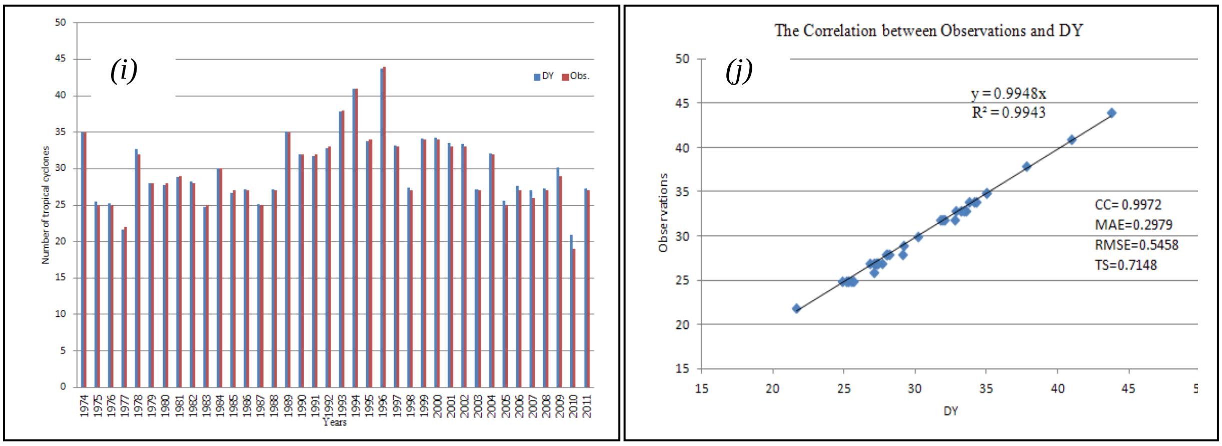


Figure 1.3: The time series errors of DY related to observations during 1974-2009 in *(i)*, and the correlation between observations and DY in *(j)*.

b. Verification models The simulation model that was calibration and using predicts the tropical cyclone in the season during 2005-2009 for verification. The model showed that a number tropical cyclone is over prediction in 2005(1), 2006(1), 2007(1) and 2009(1), the relative error 3% in 2005, 2006, 2007 and 2009, the correlation coefficient is 0.9231. The simulation model predicts a quiet over prediction, but it still quit close observations. Therefore, the forecast model exhibits potential for application in forecasting the seasonal tropical cyclone genesis in the Western North Pacific are following figure 1.4.

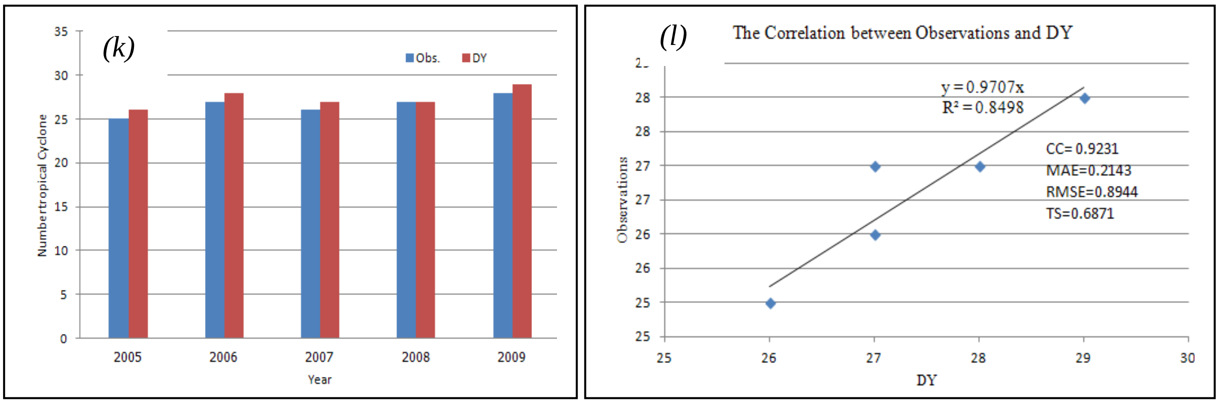


Figure 1.4: The verification model during period 2005-2009, the simulation model (red) related observations (blue) in *(k)*, and the correlation between observed and DY in *(l)*.

3.3. The simulations model by the cross-validation test The cross-validation experiments design were changed by the sea surface temperature in another month based on climate parameters in JJAS 1974-2004: DY_SST(DEC) is the sea surface temperature in December, DY_SST(JAN) is the sea surface temperature in January, DY_SST(FEB) is the sea surface temperature in February, DY_SST(MAR) is the sea surface temperature in March, DY_SST(APR) is the sea surface temperature in April, DY_SST(MAY) is the sea surface temperature in May, DY-75 is a script in 1974, DY+U50 is increasing the thermal wind at 50hPa and DY+U30 is increasing the thermal wind at 30hPa are following table 1.4.

Table 1.4: The experiments design for cross –validation

Step to work	Experiments	File name	Description
Step-1	DY_SST(DEC)	SST.4	Avg. December of the sea surface temperature
	DY_SST(JAN)	SST.4	Avg. January of the sea surface temperature
	DY_SST(FEB)	SST.4	Avg. February of the sea surface temperature
	DY_SST(MAR)	SST.4	Avg. March of the sea surface temperature
	DY_SST(APR)	SST.4	Avg. April of the sea surface temperature
	DY_SST(MAY)	SST.4	Avg. May of the sea surface temperature
Step-2	DY_75	Script 1974	1975-2009
Step-3	DY+U50	Ex-1, Ex-2, Ex-3, Ex-4, Ex-5, Ex-6, Ex-7, Ex-8, U50hPa	
	DY+U30	Ex-1, Ex-2, Ex-3, Ex-4, Ex-5, Ex-6, Ex-7, Ex-8, U30hPa	

The results in Step-1 of DY_SST (DEC) showed the over prediction is the relative error are 1975(1), 1984(1), 1988(1), 1998(1), 2000(1), 2001(1), 2005(1), 2006(1), 2007(2), 2008(1), 2009(1). The under predict is 1977(-1), are following table 1.7. The statistics were investigated of equations: the coefficient regression determination (R^2) is 0.9890, the correlation coefficient is 0.9945, avg. MAE is 0.4338, RSME is 0.6587 and TS is 0.5 are following table 1.5.

DY_SST (JAN) showed the over prediction are 1974(1), 1982(1), 1984(1), 1986(1), 1987(1), 1989(1), 1999(1), 1997(1), 1999(1), 2000(1), 2002(1) 2007(1), 2008(1), 2001(2), 2006(1), 2009(2). The under prediction are 1983(-1), 1988(-1), 1995(-1) and 1998(-1), are following table 1.7. The coefficient regression determination (R^2) is 0.9797, the correlation coefficient is 0.9897, avg. MAE is 0.6632, RSME is 0.8144 and TS is 0.2414 are following table 1.5.

DY_SST (FEB) showed the over prediction are 1974(1), 1976(1), 1982(1), 1984(1), 1986(1), 1987(1), 1989(1), 1994(1), 1997(1), 1999(1), 2000(1), 2002(1) 2004(1), 2005(1), 2007(1), 2008(1), 2001(2), 2006(2), 2009(2). The under prediction is 1995(-1). The under prediction are 1983(-1), 1988(-1), 1995(-1) and 1998(-1), are following table 1.7. The coefficient regression determination (R^2) is 0.9804, the correlation coefficient is 0.9901, avg. MAE is 0.7061, RSME is 0.8403 and TS is 0.2857 are following table 1.5.

DY_SST (MAR) showed the over prediction are 1976(1), 1982 (1), 1984(1), 1986(1), 1989(1), 1994(1), 1997(1), 1999(1), 2000(1), 2002(1) 2004(1), 2005(1), 2007(1), 2008(1), 2001(2), 2006(2), 2009(2). The under prediction is 1995(-1), are following table 1.7. The coefficient regression determination (R^2) is 0.9843, the correlation coefficient is 0.9921, avg. MAE is 0.6087, RSME is 0.7803 and TS is 0.3333 are following table 1.5.

DY_SST (APR) showed the over prediction are 1976(1), 1978(1), 1986(1), 1994(1), 2000(1), 2001(1), 2002(1), 2005(1), 2006(1), 2007(1), 2008(1), and 2009(2). The under prediction is 1995(-1), are following table 1.7. The coefficient regression determination (R^2) is 0.9890, the correlation coefficient is 0.9945, avg. MAE is 0.4568, RSME is 0.6759 and TS is 0.4694 are following table 1.5.

DY_SST (MAY) showed the over prediction are 1976(1), 2000(1), 2001(1), 2006(1), 2007(1), 2008(1), and 2009(1). The under prediction is 1980(-1), 1983(-1), 1992(-1), 1995(-1), are following table 1.7. The coefficient regression determination is (R^2) is 0.9909, the correlation coefficient is 0.9954, avg. MAE is 0.3672, RSME is 0.6059 and TS is 0.5319 are following table 1.5.

Table 1.5: The statistical parameters of cross–validation test

Statistics	DY_SST (DEC)	DY_SST (JAN)	DY_SST (FEB)	DY_SST (MAR)	DY_SST (APR)	DY_SST (MAY)	DY
R^2	0.9890	0.9797	0.9804	0.9843	0.9890	0.9909	0.9972
Correlation	0.9945	0.9897	0.9901	0.9921	0.9945	0.9954	0.9986
Avg. MAE	0.4338	0.6632	0.7061	0.6089	0.4568	0.3672	0.2979
RSME	0.6587	0.8144	0.8403	0.7803	0.6759	0.6059	0.5458
TS	0.5000	0.2414	0.2857	0.3333	0.4694	0.5319	0.7143

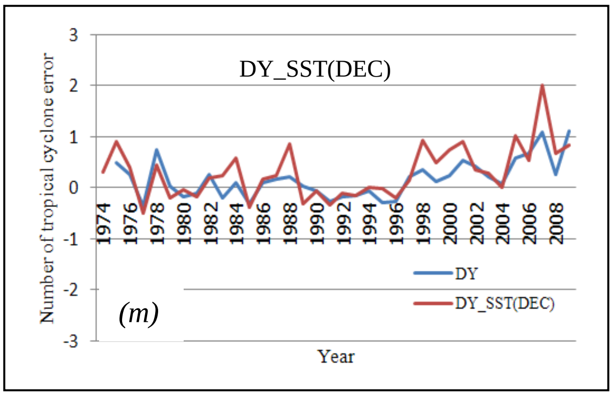
The Step-2 experiment was scripted in 1974, the results of DY_75 showed the over predictions are 1978(1), 2001(1), 2002(1), 2003(1), 2005(1), 2006(1), 2007(2) and 2009(2), see table 1.7. The statistics was investigated of equations; the coefficient regression determination (R^2) is 0.9791, the correlation coefficient is 0.9895, avg. MAE is 0.3947, RSME is 0.6283 and TS is 0.4286 are following table 1.6.

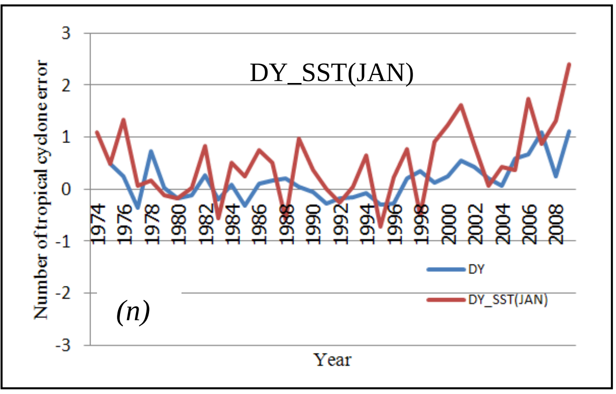
In Step-3 run test 9 climate parameters only show, the results of DY+U50 showed the over predictions are 2007(1). The statistics were investigated of equations: the coefficient regression determination (R^2) is 0.9980, the correlation coefficient is 0.9989, avg. MAE is 0.1563, RSME is 0.3953 and TS is 0.9459 that is a good model to predict are following table 1.6.

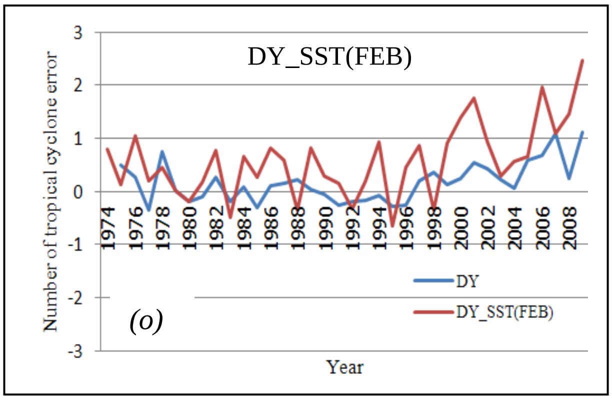
The results of DY+U30 showed the over predictions are 1978(1), 2001(1), 2005(1), 2006(1), 2007(2), 2008(1) and 2009(1). The coefficient regression determination (R^2) is 0.9918, the correlation coefficient is 0.9972, avg. MAE is 0.2979, RSME is 0.5458 and TS is 0.7143 are following table 1.6.

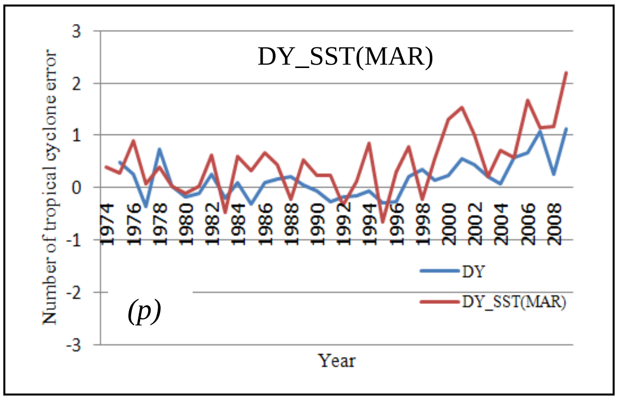
Table 1.6: The experiments design for cross -validation

Statistics	DY_75	DY+U50	DY+U30
R^2	0.9791	0.9980	0.9918
Correlation	0.9895	0.9989	0.9959
Avg. MAE	0.3947	0.1563	0.3141
RSME	0.6283	0.3953	0.5604
TS	0.4286	0.9459	0.6744









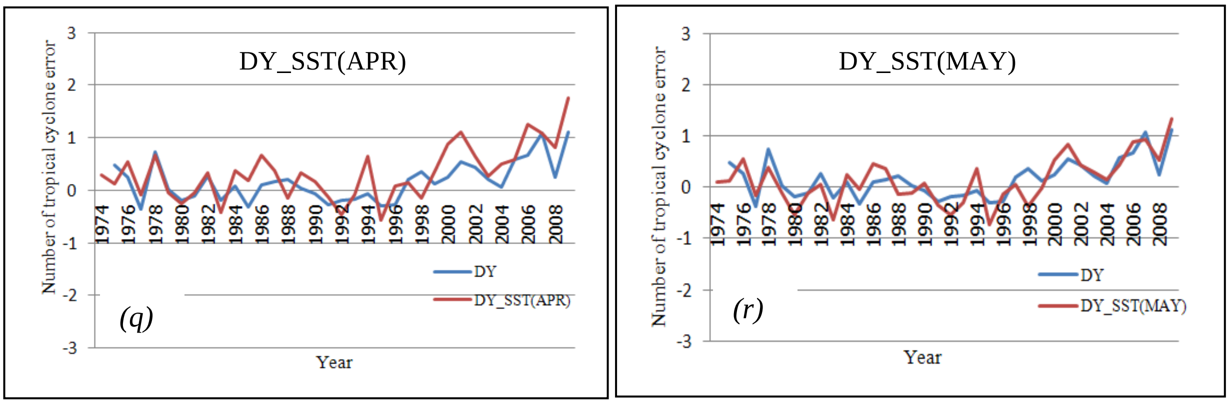


Figure 1.5: The time series errors of DY related to Step-1(DY_SST(DEC), DY_SST(JAN), DY_SST(FEB), DY_SST(MAR), DY_SST(APR) and DY_SST(MAY)) during 1974-2009 in figure *(m) – (r).*

Table 1.7: The simulation model related to observations showed errors.

Year	DY_SST (DEC)	DY_SST (JAN)	DY_SST (FEB)	DY_SST (MAR)	DY_SST (APR)	DY_SST (MAY)	DY_75	DY_U50	DY_U30	DY
1974	0	1	1	0	0	0	0	0	0	0
1975	1	0	0	0	0	0	-1	0	0	0
1976	0	1	1	1	1	1	0	0	0	0
1977	-1	0	0	0	0	0	0	0	0	0
1978	0	0	0	0	1	0	1	0	1	1
1979	0	0	0	0	0	0	0	0	0	0
1980	0	0	0	0	0	-1	0	0	0	0
1981	0	0	0	0	0	0	0	0	0	0
1982	0	1	1	1	0	0	0	0	0	0
1983	0	-1	0	0	0	-1	-1	0	0	0
1984	1	1	1	1	0	0	0	0	0	0
1985	0	0	0	0	0	0	0	0	0	0
1986	0	1	1	1	1	0	-1	0	0	0
1987	0	1	1	0	0	0	0	0	0	0
1988	1	-1	0	0	0	0	0	0	0	0
1989	0	1	1	1	0	0	0	0	0	0
1990	0	0	0	0	0	0	0	0	0	0
1991	0	0	0	0	0	0	0	0	0	0
1992	0	0	0	0	0	-1	-1	0	0	0
1993	0	0	0	0	0	0	-1	0	0	0
1994	0	1	1	1	1	0	0	0	0	0
1995	0	-1	-1	-1	-1	-1	0	0	0	0
1996	0	0	0	0	0	0	-1	0	0	0
1997	0	1	1	1	0	0	0	0	0	0
1998	1	-1	0	0	0	0	0	0	0	0
1999	0	1	1	1	0	0	0	0	0	0
2000	1	1	1	1	1	1	0	0	0	0
2001	1	2	2	2	1	1	1	0	1	1
2002	0	1	1	1	1	0	1	0	0	0
2003	0	0	0	0	0	0	1	0	0	0
2004	0	0	1	1	0	0	0	0	0	0
2005	1	0	1	1	1	0	1	0	1	1
2006	1	2	2	2	1	1	1	0	1	1
2007	2	1	1	1	1	1	2	1	2	1
2008	1	1	1	1	1	1	0	0	1	0
2009	1	1	1	1	1	1	1	0	1	1

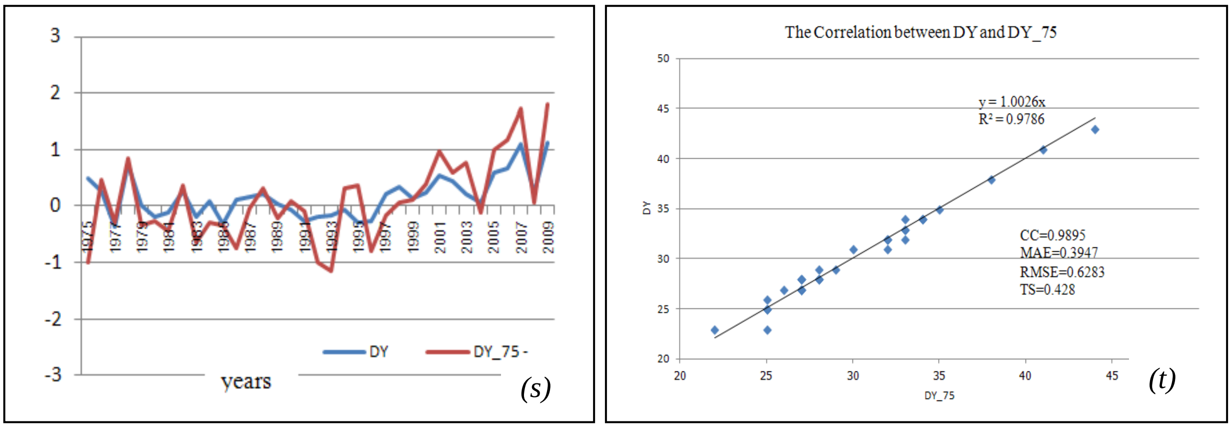


Figure 1.6: The time series of DY_75 (red) related to DY (blue) in *(s)*, and the correlation between DY to DT_75 in *(t)*.

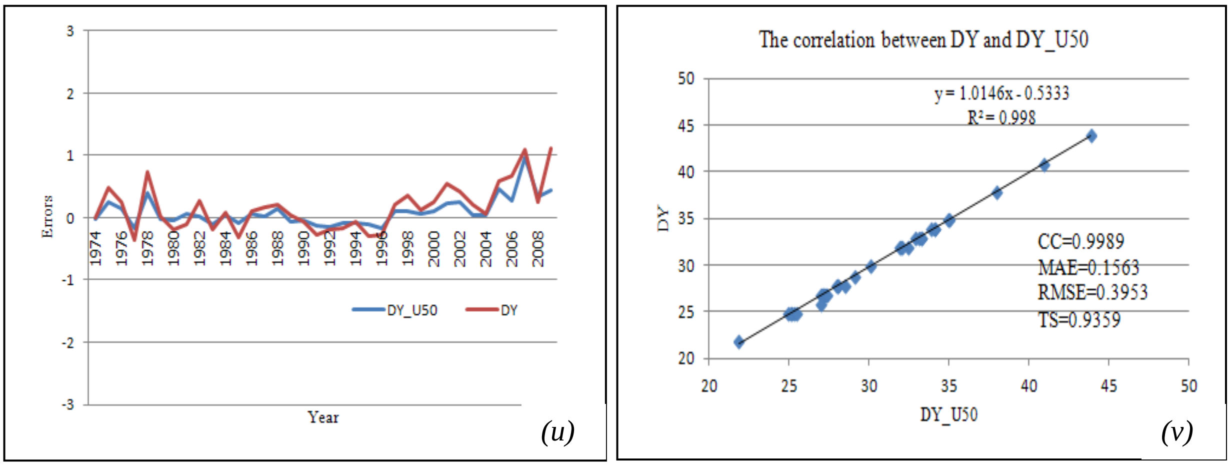


Figure 1.7: The time series of DY_U50 (red) related to DY (blue) in *(u)*, and the correlation between DY to DT_U50 in *(v)*.

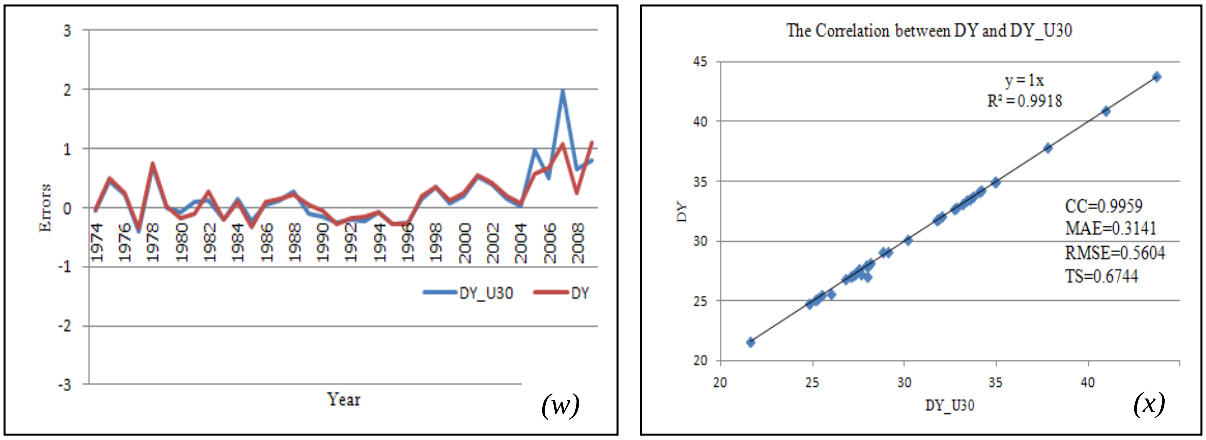


Figure 1.8: The time series of DY_U30 (red) related to DY (blue) in *(w)*, and the correlation between DY to DT_U30 in *(x)*.

7. Discussion and Concluding Remarks

This paper studied the linear regression model using the eight climate parameters (see the table 1.1) to predict the seasonal tropical cyclone genesis in the Western North Pacific. The statistics variables should be investigated the model. The results showed over predict in 1997 and 2003(7) and also the weak prediction in 1994 (-10), a high the relative error in 1977 and 2003 (~22%) and low the relative error in 1994(-32%), the MAE is 10 and avg. MAE is 2.62, are following table 1.2. The experiments used technical as the preceding year selected to run calibration the simulation model in the current year based on 1974-2004 then the relative error is 3% in 1978, 2001, 2005, 2006, 2007 and 2009 and MAE is 1, are following table 1.3. We also used the statistical method to investigate the simulation model, then showed the coefficient regression determination is (R^2) is 0.9972, the correlation coefficient is 0.9986, avg. MAE is 0.2979, the RSME is 0.5458 and also TS is 0.7148. The verification of the simulation model used data during the period 2005-2009 then showed the relative error 3% in 2005, 2006, 2007 and 2009, the correlation coefficient is 0.9596 then the simulation model predicts is quiet close observations. The results suggest that the technique run preceding year calibration the simulation model, then it did reduce error prediction and mad correctly the linear regression model to predict seasonal tropical cyclone genesis frequency in the Western North Pacific.

The technique cross-validations were changed by the sea surface temperature in each month based on JJAS 1974-2004. The results were DY_SST (DEC), DY_SST (JAN), DY_SST (FEB) and DY_SST (MAR) a quiet over prediction tropical cyclones in 2001(2), 2006(2) and 2007(2), are following table 1.7. The statistics, investigate each the simulation model that the results of the sea surface temperature in May showed the correctly prediction and quit close DY, and also have the statistics variable are the correlation coefficient is 0.9954 and TS is 0.5319. The sea surface temperature in January showed a weak correlation that was the correlation coefficient is 0.9897 and TS is 0.2414. The time series of the simulation are following figure 1.6. We were comparing the cross-validations simulations with DY then the sea surface temperature in DY showed a good the statistics variables more than the statistics variables in cross-validations. From the results suggest that the sea surface temperature in each month was changed by cross-validations testing, then it has affected the seasonal tropical cyclone genesis.

Step-2, the experiment used an eight climate parameters to run tests by script in 1974, the result of simulation model showed a quiet over prediction. The maximum of MAE(2) in 2007, and min. MAE(1), are following table 1.6. The correlation coefficient is 0.9895 and TS is 0.4286, are following table 1.6. From this technique by script some year runs tests, then it should be error occur more than old technique. From the results suggest that the preceding year should be run calibration must be not more differently years.

Step-3, the experiment uses the climate parameters, run tests by input some the climate parameter in the simulation model that is the thermal wind at 50hPa and 30hPa based on JJAS 1974-2004. The linear regression equation was established of forecasting the tropical cyclone and using the statistics to investigate. The results of DY+U50 showed a close of observations that have statistics values to investigate: avg. MAE is 1 is following table 1.7, the correlation coefficient is 0.9989 and TS is 0.9459. The results of DY+U30 showed quiet close observations less than DY+U50, and then its have statistics values: max. MAE is 1 is following table 1.7, the correlation coefficient is 0.9959 and TS is 0.6744 are following table 1.6. From the results suggest that the technical increasing the climate parameters as the thermal wind in the linear regression do not always successfully then must be consider another the climate variables.

Therefore, in this study, the technical improved the linear regression model that enough reason to make the linear regression model correctly prediction of the seasonal tropical cyclone genesis frequency in the Western North Pacific which was related the eight climate variables and observations. The results suggest that the linear regression model includes the climate parameters can used as an effective tool in the operational tropical cyclone genesis frequency.

8. Acknowledgements: This research was supported by the National Typhoon Center (NTC) of the Korea Meteorological Administration and performed during the participations of the Research Fellowship of the ESCAP/WMO Typhoon Committee Training Research Coordination Group offered by the Korea Meteorological Administration (KMA) in 2010. We would like to thank all other staffs of NTC/KMA for their help during the staying period 1 September to 27 November 2010.

References

[1] Baik, J. –J. and J. –S., Paek, 1998: A climatology of sea surface temperature and the maximum intensity of western North Pacific tropical cyclones. J. Meteor. Soc. Japan, 76, 129-137.

[2] Baik, J.-J., and J.-S. Paek, 2000: A neural network model for prediction typhoon intensity. J. *Meteor. Soc.* Japan, 78, 857–869.

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

[4] Camargo, S. L., and A. H. Sobel, 2005: Western North Pacific tropical cyclone intensity and ENSO. J. Climate, 18, 2996–3006.

[5] Camargo, S. L., K. A. Emanuel, and A. H. Sobel, 2007: Use of genesis potential index to diagnose ENSO effects on tropical cyclone genesis. J. Climate, 20, 4819–483.

[6] Chan, J. C. L., 2000: Tropical cyclone activity over the western North Pacific associated with El Nin˜ o and La Nin˜ a events. *J. Climate, 13, 2960–2972*.

[7] Chia, H. H., and C. F. Ropelewski, 2002: The interannual variability in the genesis location of the tropical cyclones in the Northwest Pacific. *J. Climate, 15, 2934–2944*.

[8] Chu, J.-H., 1994: A regression model for the western North Pacific tropical cyclone intensity forecasts. NRL Memo. Rep. 7541-94-7215, Naval Research Laboratory, 33 pp. [Available from Naval Research Laboratory, 7 Grace Hopper Ave., Monterey, CA 93943-5502.]

[9]Fan, K. (2007): North Pacific sea ice cover, a predictor for the Western North Pacific Typhoon Frequency, *Science China, 50D,* 1251–1257.

[10] Fan, K., M. J. Lin, and Y. Z. Gao (2008): Forecasting the North China's Summer Rainfall (in Chinese). *Science China, 38D*, 1452–1459.

[11] Gilchrist and Ronald Charles (1987): Statistical Adjustment of the Hatrack Typhoon Track Forecasting Model.

[12] Hart, R. E., 2003: A cyclone phase space derived from thermal wind and thermal asymmetry. *Mon. Wea. Rev.*, 131, 585-616.

[13] Ki-Seon Choi, Ja-Yeon Moon, Do-Woo Kim and Pao-Shin Chu (2010). Seasonal prediction of tropical cyclone genesis frequency over the western North Pacific using teleconnection patterns, *Theoretical and Applied Climatology, Vol. 100, Numbers 1-2, 191-206.*

[14] Ki-Seon Choi, Do-Woo Kim and Hi-Ryong Byun (2010) Statistical model for seasonal prediction of tropical cyclone frequency in the mid-latitudes of East Asia, *Theoretical and Applied Climatology, Vol. 102, Numbers 1-2, 105-114.*

[15] S. D. Kotal, S. K. Roy Bhowmik, P. K. Kundu and Ananda Das Kumar (2007), A statistical cyclone intensity prediction (SCIP) model for the Bay of Bengal, *Journal of Earth System Science, Volume 87 / 1978 - Volume 119 / 2010.*