The Spatial Temporal of Influenza Disease in Clustering Analysis to Identify Spread Pattern in Thailand

Muttitanon, W.
Faculty of Engineering, Mahidol University, Thailand, E-mail: Wutjanun.mut@mahidol.ac.th

Abstract
The pandemics of influenza in Nonthaburi province was investigated by using autoregression and found the influenza spread pattern by autocorrelation (Moran's I). Population density, temperature, relative humidity, and rainfall are the factors used in the analysis. The influenza quantitative cross-section retrospective research design was employed from 2003-2010. Three seasons are classified as: hot, rainy, and winter season. The study found that influenza outbreaks in the rainy season was R^2=0.45 and population density apparently affected the spread of influenza incidence with statistical significance coefficient (p-value <0.05). From the distribution pattern, the highest Moran's I values were related with the highest population density in 4 sub-districts: Suenyai, Taladkhwun, Bangkhen, and Bangkruay sub-district.

1. Introduction

Influenza is a major cause of death and health problem in every country. The world experienced three influenza outbreaks A/H1N1, A/H3N2 and influenza B. In the past, pandemics of influenza had greatly ruin mankind and economy (Patterson and Pyle, 1991, Johnson and Mueller, 2002 and Rios-Doria and Chowell, 2009). WHO confirmed report of case of illness of new Influenza A/H1N1 form all countries, early of January 2010 amount more than 13,554 deaths (WHO, 2002). In Thailand, the seasonal influenza and pandemic influenza are the major health concerns. The morbidity rate has been reported from CDC-MOPH of Thailand in 2007, in which number of cases reported were 18,368 and at the incidence rate of 29.19, and mortality rate of 0.08 per 100,000 populations. The age group of 0-4 years were highest of morbidity, and 23.57% was highest in agriculture occupation (Bureau Epidemiology, 2007). The outbreaks were often in rainy season. In Thailand there are three seasons, summer season, rainy season and winter season, seasonal influenza is pandemic in winter season. There were researcher who interested in investigating risk area of influenza in Nonthaburi province, especially by using geographic information system (GIS) with generated spatio-temporal and surveillance data. GIS would be a useful tool to study the geographical distribution of health data, and important tool for epidemiology and surveillance Influenza (Patterson and Pyle, 1991 and David and David, 2003). And this research was focused to apply and analyze the relationship of influenza and climate factors alone which are available from the metrology department, to identify the characteristic of pandemic influenza in Thailand. To focus on environmental factor, but not on social-economic factors.

2. Data

The data used in this research:

- Geographic data, scale 1:200,000, from Provincial Public Health Office [NPPHO].
- Influenza statistic reported from Provincial Public Health Office [NPPHO] between 2003 to 2010. Data were classified to 52 sub-districts and three seasons
- Demographic data during 2003 to 2010 in sub-district of Nonthaburi province [NPPHO].
- Climate data from the Meteorological Department of Thailand (TMD), the Kringing techniques, which is a common technique and illustrates geographic movement (Sakai, 2004) used to interpolate climate data to the spatial data, average monthly rainfall, average monthly temperature and average monthly relative humidity.

3. Methodology

The research was into three seasons of the division's season in Thailand. They are summer season from March to May, rainy season from June to October, and winter season from November to February.

- Autoregression analysis: In the first phase, the researcher has set analysis of the relationship of

the factors by autoregression analysis, and obtains the regression model before the factors were analyzed in to identify risk area in the next step. Factors of influence consists of four, such as temperature, rainfall, population density and relative humidity.

- Contiguity Weight Matrix: Determines influencing factor of Influenza incidence. After that, can be used to analyze factors that influence by the model of spatial autocorrelation (Moran's I). Which Moran's I are weighting value by spatial weights matrix that defines a local neighborhood around each geographic unit. Most analysis of spatial autocorrelation adheres to queen contiguity. That refer to what polygons are selected as neighbors for a single target polygons.
- Autocorrelation Analysis: This study use spatial autocorrelation (Moran's I) statistics technique which have a simple and quick way, and measurement quantitative data, the statistics are based on comparing the mean value among adjacent polygons having the same or different values. Usually types of data are display between -1 to 1 index. To description, the possible attribute values associated with polygons would be (Pramod et al., 2006)
- Influenza Risk Zone Classified: For determining an outbreak of influenza risk zone the risk criteria Jenks optimize method is used in this research (Jenks Natural Breaks), also known as the goodness of Variance Fit (GVF), to define natural breaks in the dataset. This classification method groups classes of similar value by minimizing the squared deviations of class means. This is an iterative calculation which starts at an arbitrary class break, compares variance within classes and continues to compare successive class break subtitle minimum variance is found (Delaware Dept. of Transportation, 2005). The goodness of Variance Fit (GVF) is calculated by the following formula (equation 1):

When: $$GVF = \frac{\text{SDAM - SDCM}}{\text{SDAM}}$$

Equation 1

SDAM (squared deviations, array mean)

$$= (\mathbf{X_i} - \overline{\mathbf{X}})^2$$

SDCM (squared deviation, class means)

$$= (\mathbf{X_i} - \overline{\mathbf{Z_c}})^2$$

Within a dataset distribution classification are used to define the 5 classes of probability, very high, high, moderate, low, and very low. The Pattern of Influenza Distribution: to explain the pattern of spread of influenza. The criteria of the degree of spatial relationship were defined by LISA (Local Indicators of Spatial Association) statistic can be interpreted as indicators of local spatial clusters and as diagnostics for local instability (Chuang and Huang, 1992). Spatial autocorrelation can be divided into four categories, corresponding with four quadrants in Moran scatter plot and identify four type of spatial association between an observation and its neighbors as significant Moran index. (Low Cluster/Random/Dispersion Pattern)

4. Result

The spatial autoregression analysis, the Jenks natural breaks optimization method and spatial autocorrelation are used to analyze the spatiotemporal pandemic as the following.

4.1 Spatial Autoregression Analysis

This study were analyzed in four parts, the first was finding the factors those were associated with an outbreak of influenza., the factors that influence defined by the researcher were (population density, rainfall, temperature, and relative humidity), all variables were average from 2003-2010, the second, the third and the fourth were similar but separate analysis of hot, rainy and winter season respectively. The statistical output was calculated from GeoDa 0.9.5 with spatial autoregression model and OLS (Ordinary Least Squares) technique was selected, and must use the queen contiguity weight matric criterion to determine neighboring units. The results of analysis were conducted for comparison of the relationships between several independent variables or predictor variables and dependent variables as detail by sub-districts of Nonthaburi Province. The results of spatial autoregression analysis was showed in Tables 1 followed by a list of four variable names population density, rainfall, temperature and relative humidity, with associated coefficient estimates, standard error, t-statistic and probability. The characteristics of summary of the model list at dependent variable (Influenza incident rate), mean (3.749), R^2 (0.40), standard deviation (1.803) and standard error (1.471). In addition, the number of observations were listed (52), the number of variables in the model (inclusive of the constant term) as (5).

Table 1: Spatial Autoregression Analysis, 2003-2010

Variables	Coefficient	Std.Error	t-statistic	Probability
Constant	238.430	90.806	2.6257	0.01*
Population Density	6.756	2.720	2.4840	0.02*
Rainfall	-0.001	0.006	-0.0780	0.94
Temperature	0.230	0.486	0.4721	0.64
Relative Humidity	-3.388	1.520	-2.2276	0.03*
R-squared : 0.40		Mean : 3.749	S.D. : 1.803	S.E. : 1.471
No of Observations : 52		No of Variable in the Model 5		

*p-value <0.05

Table 2: Spatial Autoregression Analysis Summary,Hot season

Variables	Coefficient	Std.Error	t-Statistic	Probability
Constant	2.055	8.512	0.2414	0.81
Population Density	-8.642	6.693	-1.2911	0.20
Rainfall	-1.141	0.007	-0.0017	0.99
Temperature	-0.105	0.046	-2.3139	0.03*
Relative Humidity	0.0239	0.127	0.1881	0.85
R-squared : 0.16	Mean : 0.366	S.D. :0.3493	S.E. : 0.3374	
No of Observations : 52			No of Variable in the Model 5	

*p-value <0.05

Table 3: Spatial Autoregression Analysis Summary, Rainy season

Variables	Coefficient	Std.Error	t-Statistic	Probability
Constant	76.062	53.921	1.4106	0.165
Population Density	6.755	2.065	3.2718	0.002*
Rainfall	0.008	0.005	1.4568	0.152
Temperature	-0.114	0.206	-0.5508	0.584
Relative Humidity	-1.035	0.748	-1.3845	0.173
R-squared : 0.45	Mean :2.459	S.D. :1.3094	S.E.	
No of Observations : 52		No of Variable in the Model 5		

*p-value <0.05

Table 4: Spatial Autoregression Analysis Summary, Winter season

Variable	Coefficient	Std.Error	t-Statistic	Probability
Constant	4.513	16.617	0.2715	0.79
Population Density	-2.389	1.499	-1.593	0.12
Rainfall	-0.053	0.053	-0.999	0.32
Temperature	-0.432	0.185	-2.329	0.02*
Relative Humidity	0.159	0.334	0.475	0.63
R-squared : 0.24	Mean :0.923	S.D.:0.798 3	S.E. :0.7304	
No of Observations : 52			No of Variable in the Model 5	

*p-value <0.05

There were probability and spatial autoregression coefficient estimate of constant, population density, rainfall, temperature and relative humidity as 238.43, 6.756, -0.001, 0.230 and -3.388 respectively (p-value 0.01, 0.01, 0.94, 0.64, 0.03). In addition, the significance of variable was population density and relative humidity (p-value < 0.05) and the negatives coefficient indicated were rainfall, and relative humidity. Spatial autoregression analysis output in hot season was showed descriptive statistics in Tables 2 which followed by a list of four variable names population density, rainfall, temperature and relative humidity, with associated coefficient estimates, standard error, t-statistic and probability. The summary characteristics of the model list at dependent variable (Influenza incident rate), it was mean (0.336), R^2 (0.16), standard deviation (0.3493), standard error (0.3374). In addition, the number of observations were listed (52), the number of variables in the model. There were probability and spatial autoregression coefficient estimate of constant, population density, rainfall, temperature and relative humidity as 2.055, -8.642, -1.141, -0.105 and 0.0239 respectively (p-value 0.81, 0.20, 0.99, 0.03, 0.85). In addition, the significance of variable was temperature (p-value < 0.05) and the negatives coefficient indicated were population density, rainfall and temperature. The results of spatial autoregression analysis in rainy season was showed descriptive statistics in Tables 3 which followed by a list of four variable names population density, rainfall, temperature and relative humidity, with associated coefficient estimates, standard error, t-statistic and probability. The summary characteristics of the model list at dependent variable (Influenza incident rate), it was mean (2.459), R^2 (0.45), standard deviation (1.3094), standard error (1.0185). In addition, the number of observations were listed (52), the number of variables in the model. There were probability and spatial autoregression coefficient estimate of constant, population density, rainfall, temperature and relative humidity as 76.062, 6.755, 0.008, -0.114 and -1.035 respectively (p-value 0.165, 0.002, 0.152, 0.584, 0.173). In addition, the significance of variable was temperature (p-value < 0.05) and the negatives coefficient indicated were population density, rainfall and temperature. From Table 4 shown the result of spatial autoregression analysis were followed by four variable names population density, rainfall, temperature and relative humidity, with associated coefficient estimates, standard error, t-statistic and probability.

The summary characteristics of the model list at dependent variable (Influenza incident rate), it was mean (0.923), R^2 (0.24), standard deviation (0.7983), standard error (0.7304). In addition, the number of observations were listed (52), the number of variables in the model (inclusive of the constant term) as (5). There were probability and spatial autoregression coefficient estimate of constant, population density, rainfall, temperature and relative humidity as 4.513, -2.389, -0.053, -0.432 and 0.159 respectively (p-value 0.79, 0.12, 0.32, 0.02, 0.63). In addition, the significance of variable was temperature (p-value <0.05) and the negatives coefficient indicated were population density, rainfall and temperature.

4.2 Risk Zone Map Criterion

This section presents the Influenza incident risk zone map criteria. Table 5 shows the average of the predicted incidence of outbreaks of influenza in each class, which were classified to five classes and described as hot season, rainy season, winter season and average the total 2003-2010. These criteria had calculated from the Jenks natural breaks optimization method. Table 6 had presented the Influenza risk zone criteria in different periods of hot season, rainy season, winter season and the sum of Influenza incident rate between 2003 to 2010.

Table 5: Spatial Autoregression Prediction Class Means

Class	Influenza Risk Zone Map Criteria
Class 1	< 3.190
Class 2	3.191 – 4.806
Class 3	4.807 – 5.562
Class 4	5.563 – 6.786
Class 5	> 6.787

Table 6: Influenza Risk Zone Criteria, three seasons and 2003- 2010

Influenza Risk Zone			
Hot	Rainy	Winter	2003-2010
Very Low	Very Low	Very Low	Very Low
-	Low	-	Low
-	Moderate	-	Moderate
-	High	-	High
-	-	-	High

Which the researcher classified into five levels followed as very low, low, moderate, high and very high respectively. In hot season, the risk level was very low risk zone (score <= 3.391), the four risk zone level were found that in rainy season consists such as very low, low, moderate and high (score <= 3.190, 3.191-4.806, 4.807-5.562, 5.563-6.786) respectively. The winter season was found risk zone level as very low (score <=3.190), and the last of period was the sum of Influenza incidence rate prediction 2003-2010 found such as very low, low, moderate, high and very high (score <=3.190, 3.191-4.806, 4.807-5.562, 6.563-6.786 and >6.786). Figure 1 the map had showed outbreak of influenza risk areas in each period of the analysis, is map risk areas during the hot season, rainy season, winter season and the 2003-2010 period. The maps of hot season were all entire area of the Province is just single dot. Therefore, it is the same risk level as well as, and risk zone map was very low. Map risk areas rainy season period, the all entire area of the Province had four pattern color shades, from soft shades to darker shades. The meaning of darker shade had risk of influenza than lighter shades, and darker shade area had a small area to determine at high risk zone map. The other pattern space wider was soft pattern color shade to define the very low, low and moderate risk level. The map of risk areas in winter season period, all entire area of the Province had one pattern color shades, it had very soft shades. The meaning of color shade had Very Low risk of influenza. The other pattern color space wider was soft pattern color shade to define the very low and low risk level of Influenza. And the last one of map risk areas was the 2003-2010 period. The all entire area of the Province had five pattern color shades, from soft shades to darker shades. The meaning of darker shade had risk of influenza than lighter shades, and darker shade area had a small area to determine at high and very high risk zone map. The other pattern color space wider was soft pattern color shade to define the very low, low and moderate risk level.

4.3 Influenza Map Pattern Analysis

In this session authors analyzed the patterns of spread of epidemic of influenza in Nonthaburi Province, using the autocorrelation analysis by geospatial software. The output of analysis was known as Moran's index (Moran I), point of score between -1 to 1, and the interpretation is similar to that of the product moment correlation coefficient. Informally, +1 indicates strong positive spatial autocorrelation (clustering pattern), 0 indicates random spatial autocorrelation, and -1 indicates strong negative spatial autocorrelation (dispersion pattern).

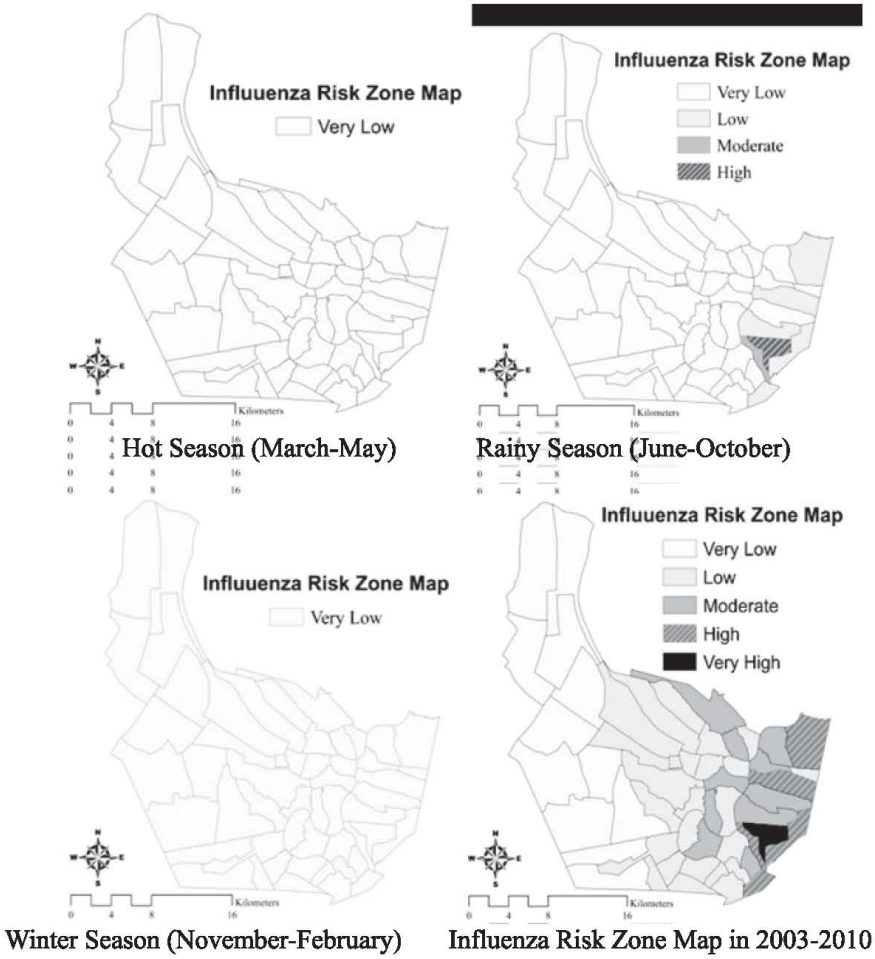

Figure 1: Influenza Risk Zone Map in hot season, rainy season, winter season and 2003-2010

Table 7: Spatial Autocorrelation 2003-2010

Year	Moran Index	Mean	SD	p-value
2003	0.0632	-0.0027	0.0786	0.23
2004	0.1046	-0.0207	0.0693	0.06
2005	0.1727	-0.0178	0.0799	0.04*
2006	0.1471	-0.0264	0.0928	0.05*
2007	0.2604	-0.0251	0.0714	0.01**
2008	0.3139	-0.0130	0.0892	0.01**
2009	0.3800	-0.0078	0.0706	0.01**
2010	0.3962	-0.0230	0.0768	0.01**

Table 8: Spatial Autocorrelation, three seasons and 2003-2010

Year	Moran Index	Mean	SD	p-value
2003-2010	0.3565	-0.0157	0.0890	0.01*
Hot Season	0.2086	-0.0197	0.0800	0.03*
Rain Season	0.3814	-0.0271	0.0668	0.01*
Winter Season	0.2148	-0.0245	0.0831	0.02*

The calculation of weighting was done by queen contiguity weight matrix in 52 sub-district of Nonthaburi Province and analyzed Influenza incidence rate from 2003 to 2010, hot season, rainy season and winter season. From Table 7 and 8 had shown the descriptive statistics of spatial autocorrelation (LISA). The output was analyzed from GeoDa software, which consists of Moran's I Index, mean, standard deviations and p-value, and Moran Index was positive indicates between 0.063 to 0.396 in 2003 to 2010. Maximum Moran's I was 0.3962 in 2010 and minimum was 0.063 in 2003. The p-value was significant in 2005 to 2010, hot season, rainy season and winter season (p-value <0.05).The pattern of Influenza outbreaks in Nonthaburi Province had illustrated a visual map of the distribution of the Influenza was from 2003 to 2010, and the total incidence of Influenza during the years 2003-2010, hot season, rainy season and

winter season. All the map were greater than 0 indicates of Moran Index that interpreted of cluster pattern, which looks like this is almost the same pattern in every year or every season from 2003-2010 except 2003, 2004, no significance. The pattern color shade were similar in each maps in every year, The soft pattern color shade of the maps has Moran Index of not significant, about 5-9 sub-district can be found every year, every season, from 2003 to 2010. The dark color shade of the maps have significant Moran Index and high Moran Index, about 3-4 sub-district can be found every year, every season, from 2003 to 2010. Figure 2 the pattern of Influenza outbreaks had illustrated in a visual map of hot season, rainy season and winter season and the total incidence of Influenza during the years 2003-2010.All the map were greater than 0 indicates of Moran Index that interpreted of cluster pattern, which looks like this is almost the same pattern in every season and all season were significance (p-value <0.05).The pattern color shade were similar in each maps in every season, The pattern soft color shade of the maps have been Moran Index not significant, about 5-9 sub-districts can be found every season and the total of Influenza incidence during the years 2003-2010. The dark pattern color shade of the maps have been Moran Index significant and highly Moran Index, about 3-4 sub-district can be found every season and the total of Influenza incidence during the years 2003-2010.

5. Discussion

These topics of discussion have been split into two parts: The first section was a summary from the analysis of the Influenza risk area in Nonthaburi Province by analyzing the spatial autoregression. The second was a summary from the analysis of the Influenza spread pattern in Nonthaburi Province by analyzing the spatial autocorrelation.

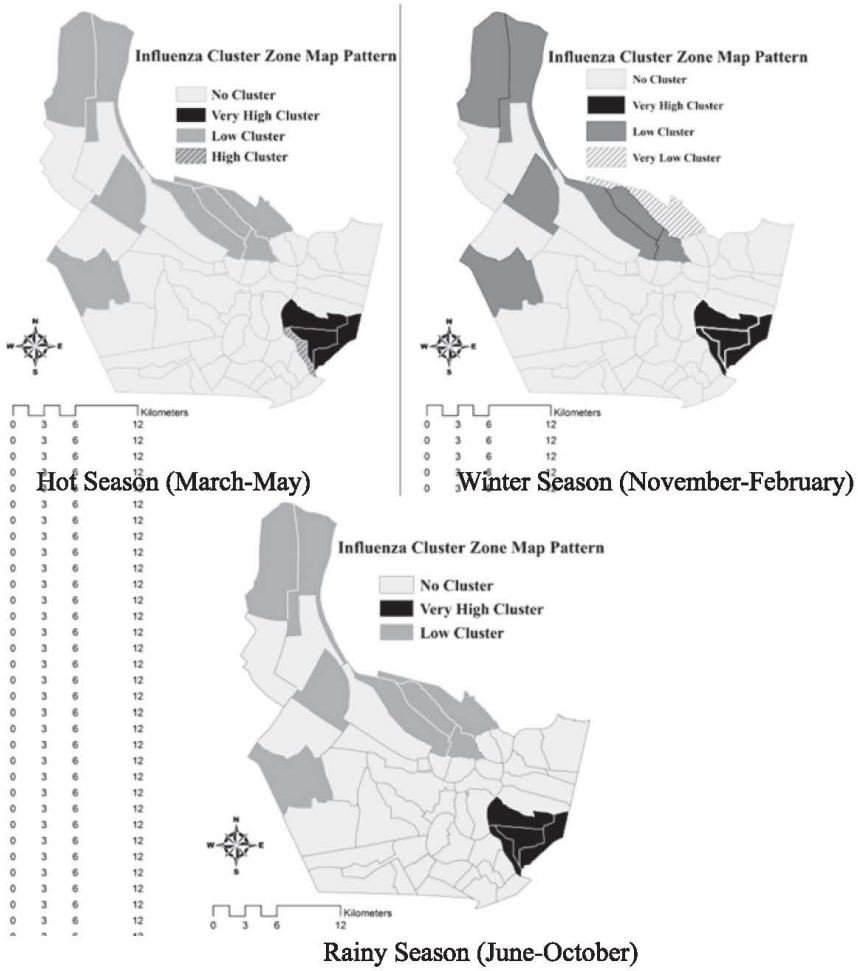

Figure 2: Influenza Spread Pattern hot season, rainy season, winter season and 2003-2010

5.1 Discussion of the Spatial Autoregression Analysis

The influenza risk area was in the area with the largest population density and relative humidity (p-value <0.05), but in difference level of criteria depend on the season. The result of spatial autoregression in hot season found that all sub-districts of Nonthaburi Province were in very low criteria, Influenza incidence rate were 0.33 (per 100,000 population), and all area were 623.6 km^2.The significant influence factor was found temperature (p-value <0.05).In rainy season, the result were found criteria risk level of Influenza risk area as very low, low, moderate and high (population density 371, 2,757, 7,580, 38,830) respectively. Most of the population density was high as 77.81%, two sub-districts, and areas such as 3 $km^2$and Influenza incidence rate 5.40 per 100,000 population. However, it also not found the greatest risk level from this analysis. In addition, the significant influence factor for influenza spread was population (p-value <0.05).For winter season was found the criteria risk level of Influenza risk area as very low (population density 1,534, 3,715, 1,151) respectively. The most of population density was low as 58.05%, 7 sub-districts, and areas such as 53.9 $km^2$and Influenza incidence rate of 0.93 per 100,000 population. The significant influence factor was found temperature (p-value <0.05).The result of analysis in 2003-2010 period found five criteria risk levels of Influenza risk area as very low, low, moderate, high and very high (population density 425, 1,714, 3,233, 7,262, 46,782) respectively. The most of population density was very high as 78.74%, follow by high as 12.22%, one sub-district, area such as 1.3 $km^2$and Influenza incidence rate 9.22 per 100,000 population. The significant influence factors were found population density and relative humidity (p-value < 0.05).

5.2 Discussion of the Spatial Autocorrelation Analysis

Spatial autocorrelation can be divided into four categories, corresponding tofour quadrants in Moran scatter plot and identify four type of spatial association between an observation and its neighbors. The Moran scatter plot of Influenza and population density, two quadrant imply positive spatial association: (High-High, Very High Cluster) where a location with an above-average value, or (Low-Low, Very Low Cluster) where a location with a below-average value. The other two quadrants imply negative spatial association: (High-Low, Low Cluster) where a location with an above-average value, or (Low-High, High Cluster) where a location with a below-average value.The Moran's I value in hot season was 0.2086, rainy season 0.3814, winter season 0.2148 and 2003-2010 period 0.3565. All season and 2003-2010 period were interpreted clustering pattern. The result of spatial autocorrelation in hot season, and had four type of Influenza association. The High-Low association was the most of population density as 78.57%, with coverage area of 1.7 km^2 and Influenza incident rate of 0.11 per 100,000 populations. Following High-High association was population density as 15.94%, coverage area 26.3 km^2 and Influenza incident rate 0.29 per 100,000 populations. The result of spatial autocorrelation in rainy season, and had three type of Influenza association. The High-High association was the most of population density as 92.76%, with coverage area of 28 km^2 and Influenza incident rate of 5.13 per 100,000 populations. Following Low-High association was population density as 4.60%, coverage area 162.4 km^2 and Influenza incident rate 1.62 per 100,000 populations. The result of spatial autocorrelation in winter season, and had found three type of Influenza association. The High-High association was the most of population density as 83.14%, coverage area 28 km^2 and Influenza incident rate 0.79 per 100,000 populations. Following Low-Low association was population density as 12.66%, coverage area 450.2 km^2 and Influenza incident rate 4.09 per 100,000 populations. The result of spatial autocorrelation in 2003-2010 period, and had found three type of Influenza association. The High-High association was the most of population density as 78.91%, coverage area 28 km^2 and Influenza incident rate 6.39 per 100,000 populations.

References

Bureau Epidemiology, 2007, *Annual Epidemiological surveillance Report*, Nonthaburi: Department of Disease Control, Ministry of Public Health. 1, 80-82.

Chuang, K.S., and Huang, H.K., 1992, Assessment of noise in a digital image using the joint-count statistic and the Moran test. *Physics in Medicine and Biology*, 37(2), 357-369.

David, O.S., and David, J.U., 2003, Geographic Information Analysis. New York: *John Wiley & Sons, Inc*. 167-206.

Delaware Department of Transportation, (2005). *Revised Archaeological Predictive Model: 301*

Project Developments[online], available : http://www.deldot.gov/archaeology/historic_pres/us301/pdf/predictive_model/301_pred_model_pred_mod.pdf/ [accessed on 2010 Oct 20].

Johnson, N.P., and Mueller, J., 2002, Global mortality of the 1918–1920 "Spanish influenza pandemic". *Bulletin of the History of Medicine*, 76, 105–115.

Patterson, K., and Pyle, G., 1991, The geography and mortality of the 1918 influenza pandemic. *Bulletin of the History of Medicine*, 65, 4–21.

Pramod, R., Sambidi, R., and Harrison, W., 2006, Spatial Clustering of the U.S. Biotech Industry, in Andersen, P., *Proceedings of 2006 Annual Conference*, Long Beach, California, 23-26 July 2006.

Rios-Doria, D., and Chowell, G., 2009, Qualitative analysis of the level of cross-protection between epidemic waves of the 1918–1919 influenza pandemic. *Journal of Theoretical Biology*, 261, 584–592.

Sakai, T., Suzaku, H., Sasaki, A., Saito, R., Tanabe, N. and Taniguchi, K., 2004, Geographic and Temporal Trends in Influenzalike Illness, Japan, 1992–1999. *Emerging Infectious DiseasesJournal*, 10(10), 1822–1826.

World Health Organization, 2002, *Influenza preparedness for the inevitable in Global defense against the infectious disease*, Geneva: WHO, 68-73.