Estimation of rainfall spatial interpolation methods in North-East Thailand

W.Thavorntam[1] , N.Wattanakij [1] and C.Mongkolsawat. [1]
Center of Geoinformatics for the Development of Northeast Thailand, Khon Kaen University.
Computer Center Building, Khon Kaen, Thailand, 40002. [1]
Tel :+66 43 348601 Fax : +66 43 348600
Email: watsao@kku.ac.th, nagon.w@gmail.com , charat@kku.ac.th

Abstract :

North-East Thailand is frequently subject to drought due to erratic distribution of rainfall, dry periods within the rainy season, and low soil water holding capacities even the amount of annual rainfall is relatively high. The objective of this study was to evaluate rainfall spatial interpolation methods both determinist and stochastic method and determine the most appropriate method for the region. Monthly rainfall data for the period 1976-2004 from 244 rainfall station distributed in the Northeast region were collected and digitally encoded into GIS database. The spatial interpolation methods in determinist and stochastic method were applied for spatial interpolation of mean annual rainfall in this study period to obtain spatial mean annual rainfall for the whole region. Each method was then evaluated from cross validation for the interpolation quality. The results of the study
indicated that rainfall amounts increased significantly from the southwest to the northeast portions of the study area and the ordinary kriging with spherical model had better performance for interpolation of rainfall within the study area. The estimation of rainfall spatial interpolation methods provides overall insight of rainfall pattern and models for crop growing period.

KEY WORDS: Rainfall , precipitation, drought, Interpolation

1. Introduction

The Northeastern part of Thailand has increasingly been impacted by drought throughout many parts of the region, despite no significant trends in total annual rainfall amounts. The primary causes of drought are associated with an erratic distribution of rainfall and low water holding capacity of the soils of the region, resulting in critical dry spells during the rainy season period, particularly from June to July and the last two weeks of September. The most serious negative consequence from drought is water deficiency for agriculture, the major economic sector in the region, thus impacting human life, property and agricultural production in the region and the country as a whole.

The rainfall period in Northeast Thailand is from May through October and is controlled by the Asia monsoons with two distinct rainfall periods, one from the southwest in the early wet season and the other from the northeast in the latter part. A general pattern of increasing rainfall is found from the southwest to the northeast part of this region, with the greatest inter-annual variability in rainfall occurring in the drier southwest, in which several provinces in this area have been severely drought impacted (Saenjan et al., 1990). Furthermore, unevenly distributed rainfall during the monsoon period has been reported across extensive portions of the southwestern and central parts of the region, with dry spells extending for longer periods in the middle of June (Siripon and Mongkolsawat, 2000).

May studies have been proposed for the spatial interpolation of rainfall to get the optimum parameters required for interpolation methods. Dirks et al (1998) studied interpolation of rainfall data on Norfork Island. Thirteen rain gauges on the island with the area of 35 km^2 were used. The result found kriging provide no significant improvement over simple methods like Inverse Distance, Thiessen and areal mean methods because of the demanding on computation. However, the kriging method was applied to evaluate monthly regional rainfall in the central part of Iran. The two component of rainfall time series, deterministic trends and random component were used for this model. The result of this study will be use for climate and hydrological studies (Karamouz et al, 2007). Geostatistical approaches; simple kriging with varying local means; kriging with an extental drift and colocated cokriging were used to interpolate rainfall incorporate with elevation in Portugal. The prediction performance was compared by using cross validation. The result found that ordinary kriging yields more accurate predictions than linear regression when the correlation between rainfall and elevation is moderate (Goovaerts, 2000).

Studying spatial rainfall pattern is important to many problems such as hydrologic analysis, models for crop growing period and drought. There were several interpolation methods have been used to analyze spatial rainfall pattern and how to choose a method which is the most appropriate and convenience of use is crucial.

Description of the general equation in regionalized variable theory.

The values of attributes at unsampled sites from measure points can be predicted and converted to continuous surface using interpolation methods. There are two methods for interpolation, local deterministic and global methods. Local deterministic methods are Thiessen polygons, inverse distance weighting and thin plate spline. Global interpolators use all data for prediction the whole area and mostly use for examining the effect of global variations. Kriging is the geostatistical methods that required an understanding of the principles of statistical spatial autocorrelation. These methods are applied when the variation of an attributes is irregular. The advantages of geostatistical methods is the spatial variation of any continuous attributes can be better modelled by a stochastic surface rather than a simple mathematical function (Burrough et al, 1998). The attribute is called a regionalized variable which assume that the spatial variation of any variable can be expressed in three major components;

a) A structural component associated with a constant mean value or a constant trend.
b) A random, spatially correlated component.
c) A random noise or residual error term.

Let x be a position in three dimensions, the value of a variable Z at x is given by

$$Z(x) = m(x)+\varepsilon'(x)+\varepsilon'' \qquad \text{.....(1)}$$

Where $m(x)$ is a deterministic function describing the structural component of Z at z
$\varepsilon'(x)$ is the term denoting the stochastic, locally varying, spatially dependent residuals from $m(x)$
ε'' is a residual, spatially dependent Gaussian noise term having zero mean and variance.

The equation (1) can be rewritten by using y(h) instead of $\varepsilon'(x)$ as:

$$Z(x) = m(x)+ \gamma\,(h)+\varepsilon'' \qquad \text{.....(2)}$$

where γ (h) is the semivariance

Intrinsic stationary is the assumption of Ordinary kriging which assumes that the data comes from a random process with a constant mean and an autocorrelation structure depends on the separating direction and distance.The semivariance can be estimated from sample data as shown in equation (3) when the conditions of intrinsic stationary are qualified.

$$\hat{y}(h) = 1/2n \sum_{i=1}^{n}\{z(x_i) - z(x_i + h)\}^2 \qquad \text{....(3)}$$

where n is the number of pairs of sample points of observations of values of attribute z separated by distance h. The experimental variogram is a plot of $\hat{y}(h)$ against h which aim to analyze the autocorrelation structure and fit an appropriate model to the structure. (Burrough, 1986).

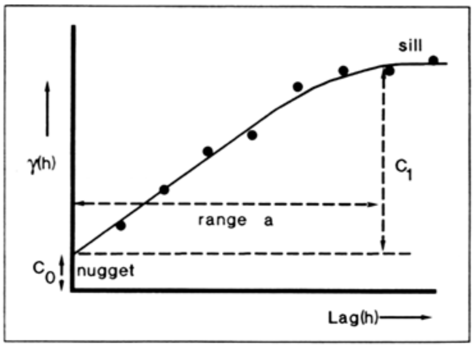


Figure 1. An example of a variogram with range, nugget and sill.
Source : (Burrough, 1986).

Figure 1 shows the example of simple transitional variogram with range, nugget and sill. The lag (h) is in the x axis and y(h) is in the y axis. The curve in the is mathematical model that have been fitted to derived semivariance in order to be able to describe the changing of semivariance with the lag. The first part is Sill, the horizontal part of the curve that levels off. Sill can implies that at these values of the lag there is no spatial dependence between that data point because all estimates of variance of differences are invariant with distance. The second part is range, it is the curve rises from a low value of y(h) to the sill, reaching at the value of h. Range can describes at what distance inter-site difference become spatially dependence. The range can answer the question of how large the window of weight moving average should be. The nugget is the value that cut the y axis at the positive value. This value in the model is an estimate of ε'' and it combines the residual variations that occur over distances much shorter than the sample spacing and that consequently cannot be resolved (Burrough, 1986).

There are several variogram models that are mostly used. The models that have a clear range and sill are called transition model such as spherical model, exponential model and Gaussian model. The spherical model can be used when the nugget variance is important but not too large. The exponential model can be used when the curve increases continuously to the range. In addition, when the variation is very smooth and the nugget variance ε'' is very small comparing with ε' , the Gaussian Model can be used in this case (Burrough et al, 1998). The Linear model is known as non-transition models because the semivariogram does not have a sill and it can be used in the universal kriging.

2. Objective

The objective of this study is to evaluate rainfall spatial interpolation methods both determinist and stochastic method and determine the most appropriate method in North-East Thailand.

3. Study Area

The study area encompassed most of North-East Thailand, with an approximate area of 60,000 km^2, between 14° 18′ N to 18° 15′ N and 102° 22′ E 104° 50′ E (Fig. 2). Most of the area is dominated by rice paddy fields and other crops, with isolated patches of remnant forest. The topography of the area is gentle undulating terrain and small hills, and the area is underlain by a thick sequence of Mesozoic rock, the Maha Sarakam Formation which consists of sandstone, siltstone and interbedded rock salts. The soils are inherently low in fertility and have sandy textures with low water holding capacities.

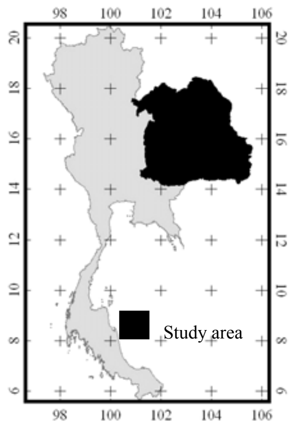


Figure 2. Study area.

4. Methodology

4.1 Data collection

Monthly rainfall data from 308 Thailand Meteorological Department stations, distributed throughout the study area as shows in figure 3, were collected for the years 1976 to 2004 and digitally encoded into a GIS database. A preliminary analysis of the rainfall data was made to examine rainfall frequency distributions and spatial autocorrelation patterns over different locations of the study area. The mean annual rainfall of each station was calculated for the 29 years, and the median value of annual rainfall data was selected as the representative value of each station during the study period.

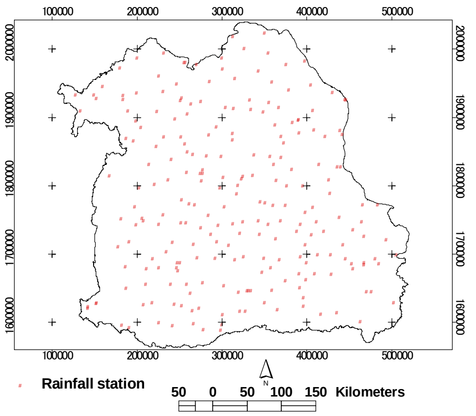


Figure 3. Rainfall stations in the study area.

4.2 Statistical Exploration of data

Exploring rainfall data before using the interpolations techniques is important for selecting the appropriate parameters for interpolation model. The statistical of rainfall data in the region was explored by these methods;

1) Testing frequency distribution of data to explore the univariate distribution of a dataset.
2) Using Normal QQ plot to check for normality of a dataset.
3) Identify global trends in a dataset.
4) Analyze the spatial dependencies in a dataset and exploring the spatial autocorrelation by using Semivariogram/Covariance cloud.

4.3 Deterministic methods for spatial interpolation

This study used two deterministic methods to create surfaces from measure points based on the extent of similarity or the degree of smoothing. The Deterministic methods that have been chosen for this study are;

1) Inverse Distance Weighted (IDW)

The IDW predict a value for any unmeasured location using the surrounding measured values which the closer measured values to the prediction location will have more influence than those farther apart (Johnston et al, 2001).

2) Radial Basis Function

Radial Basis Function methods are exact interpolation technique which the surface have to go through every measured value. These Radial Basis Functions were used to create a surface; thin-plate spline, spline with tension and regularized spline.

4.4 Stochastic methods for spatial interpolation

Ordinary kriging was applied in this study for stochastic methods. The considerable parameters for this method are;

1) Transformation of data is applied to make it to normal distribution.
2) Modeling semivariogram and covariance function is studied to find the most accurate model.
3) Adjust lag size to get the optimum distance.
4) Perform cross-validation and validation to see the result from the model.

5. Results & Discussion

5.1 29 year rainfall record

Examples of the spatial and temporal variability in annual rainfall are shown in Fig. 4 for three zones, one in the southwest (SW), one in the central (C), and one in the northeast (NE) portions of the study area. Rainfall amounts in the northeast are much higher than in the southwest, throughout the 1976 to 2004 record, primarily a result of the heavy rains from the northeast that N-E Thailand receives during the September and October monsoon period. The southwest area is seen to have a strong decline in precipitation in the last 5 years 2000-2004. In contrast, there are smaller declines in precipitation in the central and northeast areas and there are no significant long term trends in mean annual precipitation in all three zones. Overall, annual rainfall varied from minimums near 800 mm, in the southwest and western parts of the region, to a maximum of over 3000 mm in the northern part.

5.2 Statistical Exploration of data

The frequency distribution of annual rainfall at the 308 stations shows the mean values of annual rainfall to be near 1300 mm (Fig. 5). Mean annual rainfall values in the 1000 mm to 1200 mm range occurred most frequently (highest frequencies), primarily in the central part of the study area. High annual rainfalls between 1900 mm to 2100 mm occurred with less frequency and primarily along the Mekong River in the northeast portion of the study area.

The spatial autocorrelation of rainfall in the study area was examined with the semivariogram and covariance cloud. The different pairs of sample location were examined by measure the distance between two locations and the semivariogram cloud was created by plotting half the difference squared between the values at the locations. High spatial variation of rainfall in the stations were located in the southern and the northern parts of the study area and relatively low variation in rainfall were found throughout the central, northwest, and southeast stations of the region (Fig. 6).

The normal QQ plot is created by plotting data values versus the value of a standard normal where their cumulative distributions are equal (Johnston et al, 2001). We used the QQ plot to compare the distribution of rainfall data to a standard normal distribution. The plot is close to a straight line can be assumed that it is close to normal distribution. From figure 7a, point data were plotted to creating a straight line except the points data in the northeast and southwest of the region. From figure 7b, the plot was closer to a straight line after the data was transformed to make it similar to a normal distribution,. When the data did not exhibit a normal distribution, Transformation of the data may be necessary before using kriging interpolation techniques.

Looking for global trends is important to make an appropriate model to create a smooth continuous surface. For geostatistical method, trend could be removed before modeling the semivarigogram/covariance for more accurate. The direction of trend can be identified from figure 8. There was global trend in this study area because a curve was not flat on the projected plane. The trend strongest influence was from the southwest to the northeast.

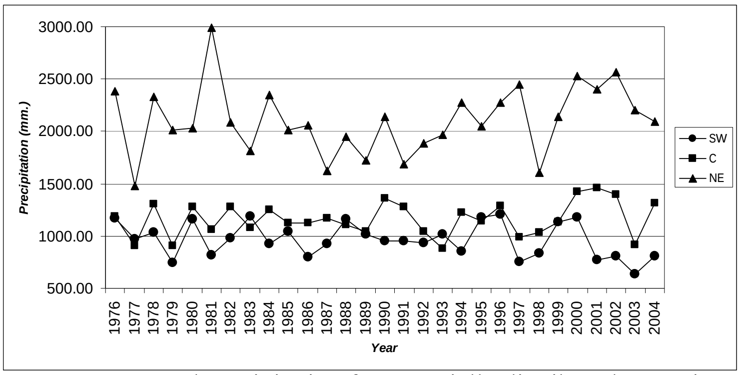


Figure 4: Mean annual precipitation for 3 spatially distributed zones in N-E Thailand.

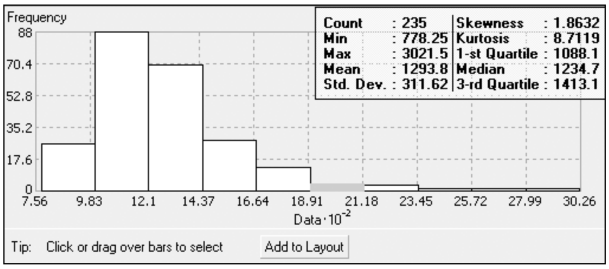


Figure 5: The frequency distribution of annual rainfall

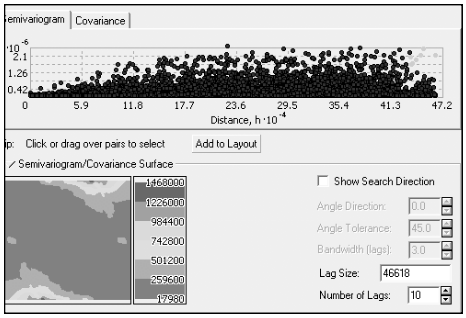


Figure 6: Semivariogram/covariance cloud of rainfall in the region

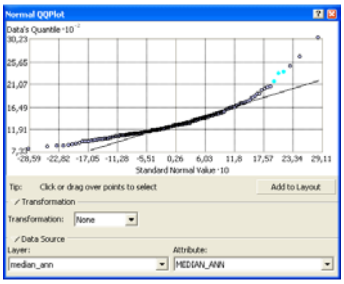


Figure 7a: The Normal QQ plot

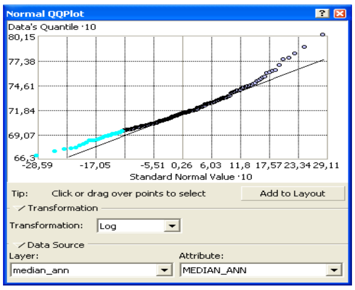


Figure 7b: The Normal QQ plot after transformation.

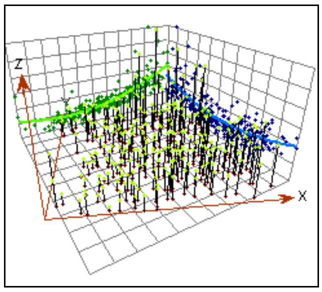


Figure 8. Trend Analysis.

5.3 Deterministic methods for spatial interpolation

There are two selected spatial interpolation methods for deterministic; IDW and Radial Basis Function and the result are listed below;

1) **Inverse Distance Weighted (IDW)**

The power parameter and the neighborhood search are influenced the surface from this method. This study used power 2 as the optimal power for prediction. The selected shape of the neighborhood was circle because there were no directional influences on the weight of the data. The prediction errors from the method are -7.219 for Mean error and 166.4 for RMS error. Figure 9 shows the spatial pattern of rainfall, obtained by IDW method, shown the increasing of mean annual rainfall from the southwest to the northeast parts of the region over the 29 year period was evident.

2) **Radial Basis Function**

There were three basis functions used in this study ; thin-plate spline, spline with tension and regularized spline. The prediction errors from thin-plate spline were -1.311 for Mean error and 187.6 for RMS error. The prediction errors from spline with tension were -4.461 for Mean error and 157.8 for RMS error. For regularized spline, mean error and RMS error were -5.13 and 157.9 respectively. Figure 10a, figure 10b and figure 10c shows the spatial pattern of mean annual rainfall in the study area from thin-plate spline, spline with tension and regularized spline respectively.

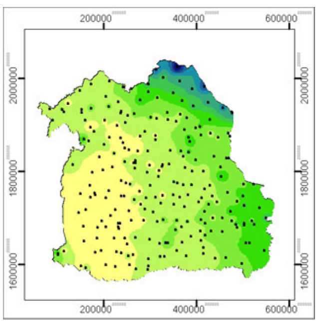


Figure 9. IDW method.

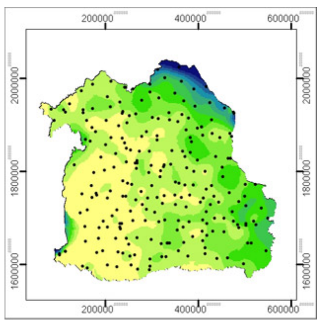


Figure 10a.Thin-plate spline

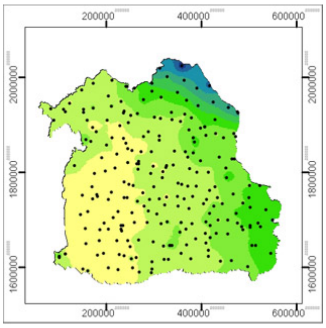


Figure 10b. Spline with tension

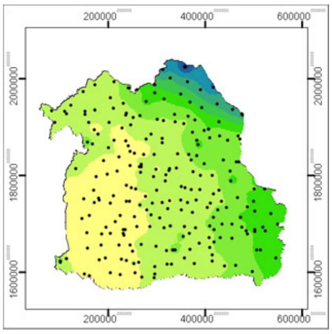


Figure 10c.R egularized spline.

5.4 Stochastic methods for spatial interpolation

This study applies ordinary kriging as a stochastic method for interpolation rainfall. There are several conditions to adjust to get the optimal model and most accuracy and the result of each condition are as follow;

1) Transformation of data and trend removal did not apply for the first test. The model used in semivariogram was spherical. Figure 11a shows the semivariogram of spherical model. Semivariogram/covariance modeling was used to determine the best fit model. The x-axis on the semivariogram graph is the distance from the center of the cell to the center of the semivariogram surface and the y-axis is the semivariance shows semivariogram value. The value of calculated semivariogram represents from lower values in blue color to higher values in red color. The yellow line shows fitted semivariogram model and red dots are empirical semivariogram values. From this figure, the semivariogram started low at a small distance and increased as distance increase. From the Semivariogram/Covariance surface in figure 11b, the semivariance increased from the southwest to south and the northeast to north direction.

Figure 11c shows cross-validation from this model. The Cross validation computes the residuals for all data points, and the self-consistency of the variogram is tested by a mean difference of residuals of zero and a variance of 1 (Burrough et al, 1998). The dashed line is the 1:1 line shows where the data should fall. High rainfall was expected to be less while low rainfall was expected to be more. This is a property of kriging that tends to underpredict large values and overpredict small values (ArcGis help, 2002). The blue line is the line of best fit with the equation of regression function was y = 0.751x + 305.695 which given below the plot. The prediction errors from cross validation can help to find the most accurate prediction model. For an accurate model, the mean error should be close to 0, the root-mean-square error and average standard error should be as small as possible. Average standard errors should be close to root-mean-squared. The root-mean-square standardized error should be close to 1 (Johnston et al, 2001). From the spherical model, mean error was -1.377, root-mean-square error and average standard error were 152.2 and 100.8 respectively. The root-mean-square standardized error was 1.697. To check an un biases of prediction, the mean prediction error should be near 0. The root-mean-square error was greater than the average standard error means there was an under estimating the variability in the prediction which was agree with the value of root-mean-square standardized error was greater than 1. These results were used for comparing models.

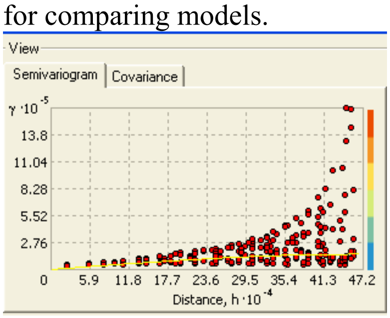


Figure 11a.The semivariogram of spherical model.

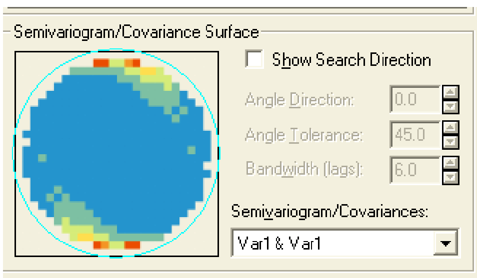


Figure11b. The semivariogram/covariance surface

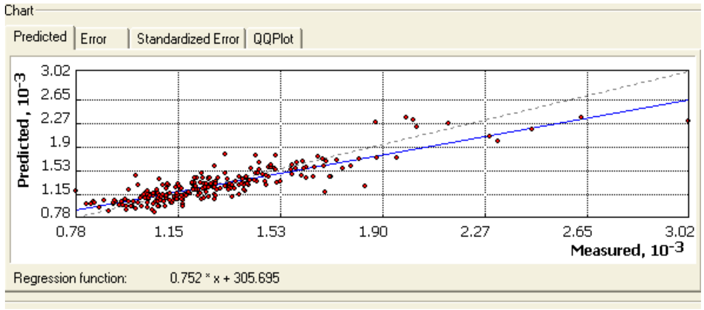


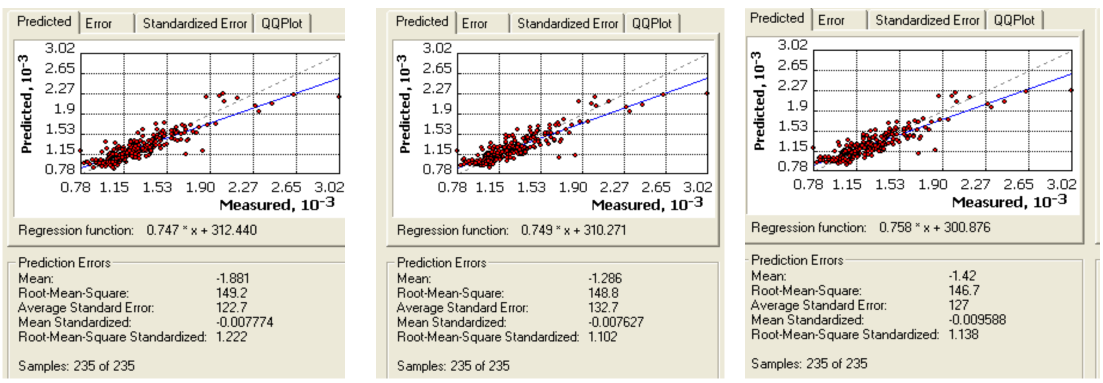
Figure11c. The cross validation and the prediction chart for median rainfall in rainy period

2)Transformation of data and trend removal did not apply but lag size was adjusted. The model used was still spherical. The cross validation from this model was show in figure 13a. The prediction errors were; mean error is -1.861, root-mean-square error and average standard error were 149.2 and 122.7 respectively. The root-mean-square standardized error was 1.222. This model gives the better result comparing with the first test.

3) Transformation of data did not apply but lag size was adjusted. The 2nd order of trend removal was used for this second test. The model was still spherical. The cross validation from this model is show in figure 13b. The prediction error were; mean error was -1.285, root-mean-square error and average standard error were 148.8 and 132.7 respectively. The root-mean-square standardized error was 1.102. This model gives the better result comparing with the first and the second test.

4) Transformation of data did not apply but lag size was adjusted for the fourth test. The 2nd order of trend removal was used. Model was changed to exponential. The cross validation from this model is show in figure 13c. The prediction error were; mean error was -1.42, root-mean-square error and average standard error were 146.7 and 127 respectively. The root-mean-square standardized error was 1.138. For overall prediction error statistics, changing model from spherical to exponential did not provide a better prediction. Figure 12 shows cross validation and the prediction chart for the second, third and fourth test respectively.

From the Ordinary kriging method with spherical model, adjusted lag size and transformation of data was applied, the region-wide spatial pattern of rainfall revealed an overall mean annual rainfall that increased from the southwest to the northeast parts of the region over the 29 year period (Fig. 13).



(a) (b) (c)

Figure 12. The cross validation and the prediction chart for the second, third and fourth model respectively.

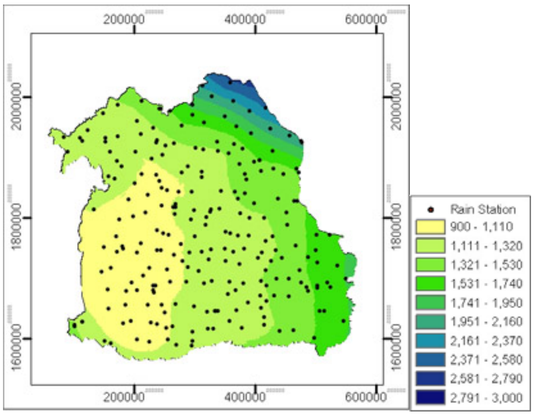


Figure 13. The spatial pattern of mean annual rainfall in 29 years period.

6.Conclusion

The spatial interpolation of mean annual rainfall in 29 years period was studied using interpolation methods; IDW, radial basis function and Ordinary kriging. The performance of each method was evaluated by cross validation. There were significant differences depending on the interpolations used. Radial Basis Function (Thin-plate spline) was the better efficient deterministic interpolation methods with mean error -1.311. For Ordinary kriging methods (Stochastic), spherical model, adjusted lag size and transformation of data was the most representative stochastic interpolation methods with mean error, Root-Mean-Square, Average Standard error and Root-Mean-Square Standardized were -1.28, 148.8, 132.7 and 1.02 respectively. The spatial pattern form this interpolation method show rainfall amount in the 29 years period in the northeast of Thailand increased significantly from the southwest to the northeast portion of the region. The estimation of rainfall spatial interpolation methods provides overall insight of rainfall pattern and models for crop growing period.

References

Burrough, P. A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. *Oxford University Press Inc., New York.*

Burrough, P. A. & McDonnell, R. A., 1998. Principles of Geographic Information Systems. *Oxford University Press Inc., New York.*

Dirks, K. N., Hay, J.E., Stow, C.D. and Harris, D.1998. High-resolution studies of rainfall on Norfork Island Part II: Interpolation of rainfall data. *Journal of Hydrology. Vol 208. Issues 3-4 July 1998, Pp 187-193.*

Goovaerts, P.2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology. Vol 228. Issues 1-2, 21 February 2000, Pp 113-129..*

Johnston, K., Ver Hoef, M. J., Krivoruchko, K. and Lucas, N. 2001.Using ArcGIS Geostatistical Analyst.*ESRI.*

Karamouz, M., Torabi, S. and Araghinejad, S.2007. Case study of monthly regional rainfall evaluation by spatiotemporal geostatistical method. *Journal of Hydrologic Engineering. Vol 12. No.1, January 1, 2007.*

Saenjan, P., Garnier B.J., and Maclean,P.A., 1990, Pattern of Wet Season Rainfall in Northeast Thailand. *Proceeding of the Seminar on Remote Sensing and GIS for Soil and Water Management*. Khon Kaen University, 180-202.

Siripon, K. and Mongkolsawat, C., 2000, Spatial and Temporal Analysis of Rainfall Pattern in Northeastern Thailand : Application of GIS. *Journal of Remote Sensing and GIS Association of Thailand*, Vol 1(1), 1-18.