Using Multiple Regression to Explain Spatial Variation of Annual Rainfall: A Case Study of the Northeast of Thailand

Siripon Kamontum
Geo-informatics and Space Technology Development Agency (GISTDA)
120 Moo 3 Building B, 6th & 7th Floors, Chaeng Wattana Rd., Lak Si
Bangkok 10210, THAILAND
Fax (+66) 0-2143-9594 E-mail: siripon@gistda.or.th

Abstract

Spatial variation of rainfall is fundamental information required in flood and drought management, and also needed for crop cultivation planning. If variation of rainfall over the space is precisely explained, water resources can be managed effectively leading to welfare of people in the area. Thus, an objective of this study was to explain spatial variation of annual rainfall of the Northeast of Thailand using multiple regression. In this study, the dependent variable was mean annual rainfall calculated from rainfall data collected during 1951 to 2012 from 240 rain gauge stations provided by the Meteorology Department and the Royal Thai Irrigation Department of Thailand. The independent variables were elevation, distance from the South China Sea (SCS), and x and y coordinates. Dummy variables were land cover types (forest, urban, agricultural and water body areas) and windward areas near LAOS PDR. Next, the independent variables were interacted with each other, and then interacted these terms with dummy variables. Backward and forward stepwise regressions were then applied. The result showed that by using x(Elevation), x(Distance from the SCS), x/(Elevation), y/(Distance from the SCS), y^2, and (Elevation/Distance from the SCS)LAOS in the regression model, 75.27% of variation in mean annual rainfall of the Northeast of Thailand is explained. The derived regression model also complies with regression assumption including linearity, normality, constant variance and multicollinearity. Regression is a powerful tool for rainfall estimation because various types of data can be applied as input of the model.

Key words: rainfall, regression

1. Introduction

The Northeastern of Thailand, covering 168,854 km^2 (41,724,732.1 acres), is the largest region in this country. Principal land use of the region is agricultural land (major economic crops are rice, sugar cane, and cassava). Due to an inadequacy of irrigation system, approximately 80% of the agricultural area depends solely on rainfall. Some areas in this region are susceptible to drought, but some other areas are prone to flooding. Thus, spatial pattern of rainfall distribution in this region is uneven. If variation of rainfall over the space is precisely explained, water resources can be managed effectively leading to welfare of people in the region.

Variability of monsoon rains over Thailand was studied by Singhrattna, N. et al. (2005) and it was found that pacific sea surface temperatures (SSTs), in particular, El Nino-Southern Oscillation (ENSO), had a negative relationship with the summer monsoon rainfall in recent decades.

SST is one of important factors applied in rainfall prediction. Colman, A. et al. (1998) found that 2 predictors, (1) the 30N-30S portion of the third covariance-based EOF of Atlantic SST for all seasons, and (2) the first EOF of Pacific SST for Dec-Jan-Feb, delivered substantial forecast skill of multiple regression for March-April-May 1998 rainfall prediction in northeast Brazil. Greischar, L. and Hastenrath, S. (2000) used (1) an index of Jan SST in the equatorial Pacific, (2) an index of Jan SST in the tropical Atlantic, 30N-30S, (3) an index of Nov-Dec-Jan SST in the tropical Atlantic, 30N-30S, (4) Oct-Jan precipitation at 27 stations, and (5) an index of Jan meridional surface wind component over the tropical Atlantic, 30N-30S, to predict Mar-Apr-May-Jun 2000 rainfall in northeast Brazil using stepwise and multiple regression.

Lachniet, M.S. and Patterson W.P. (2005), utilized correlation and stepwise regression to evaluate physical controls on the stable isotope values of Panamanian rain and surface water. Variables were latitude, longitude, distance from Caribbean sea, $\delta^{18}O$, δD, deuterium-excess (d_x), stream length above sampling site, estimated mean annual precipitation, and pH.

According to Croke, B.F.W. et al. (2001), elevation was a strong control on precipitation quantity. The researchers created a linear regression model using elevation as the independent variable to predict total annual rainfall of Upper Murrumbidgee catchment, Australia.

There are several studies applying multiple regression for rainfall prediction and independent variables are different according to approach of researchers, such as cloud top temperature (Vicente, G.A. and Scofield, R.A., 1998 and Hu, J. et al., 2006), evaporation (Yin, X. and Nicholson, S.E., 2002), and Normalized Difference Vegetation Index (NDVI) (Foody, G.M., 2003).

Although some of the above mentioned independent variables are not available in the study area, a multiple regression model can be created based on X, Y coordinates, elevation, distance from South China Sea, and land use. An objective of this study is to create a regression model to explain spatial variation of mean annual rainfall of the Northeastern Thailand.

2. Study Area

The Northeastern Thailand is located from 14° 14' to 18° 27'N latitude and 101° 15' to 105° 35'E longitude. The region is a saucer shaped plateau with gently rolling low hills known as Khorat Plateau. Flat topped mountains ring the plateau on the west and south, and Mekong River draws the north and east border of the region, which is a border line between Thailand and Laos as well. Topography of the study area is illustrated in Figure 1. Average elevation of the region is 200 meters above mean sea level.

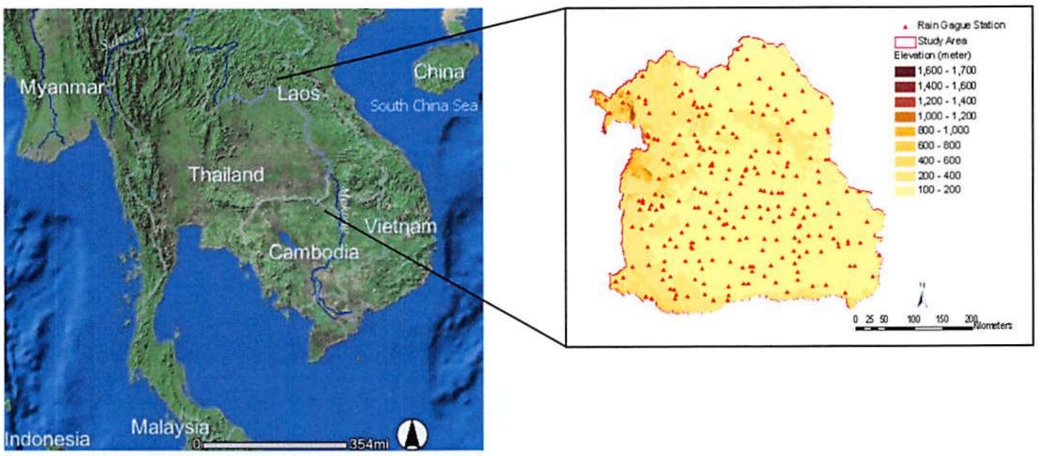


Figure 1. Topography of the study area and rain gauge locations

There are 3 seasons in the study area: summer (Mar. – May), rainy (Jun. – Oct.), and winter (Nov. – Feb.). The rainy season is influenced by south-west monsoon wind and storms from South China sea. The former brings small quantity of rain because the mountains at the west and south block this wind, when the later (Typhoon, Tropical storm, and Tropical depression) are so strong that moisture from the South China is carried through Vietnam and Laos and causes large amount of rainfall through out the region. For this reason, area in the east seems to have higher precipitation quantity than the west.

3. Methods

3.1 Data and Variables

Rainfall data collected during 1951 to 2000 from 240 rain gauge stations (n = 240) were provided by the Meteorology Department and the Royal Thai Irrigation Department, Thailand. Mean annual rainfall of each station was a dependent variable in this study.

In order to explain spatial variation of mean annual rainfall of the Northeastern Thailand, X and Y coordinates of rain gauges were used as independent variables, and coordinate system used in this study was based on UTM projection, WGS84 datum. Unit of X and Y coordinate was meter.

Generally, elevation is a strong control of rainfall quantity (the higher the elevation, the larger the amount of rainfall). For this reason, elevation of rain stations was added as one of independent variables. By utilizing contour data provided by Khon Kaen University, Thailand, DEM was created and the elevation of the rain stations was defined using ArcGIS. Unit of the elevation was meter.

Because rainfall quantity in the study area was mostly contributed by storms from the South China, distance from the sea was applied as one of independent variables (Areas near the sea tend to have more rain). By using ArcGIS, buffers of coast line along the sea were created in kilometer unit, and distance from the South China was assigned to each rain gauge.

Land use may be another factor controlling rainfall quantity. For example, forest areas have larger quantity of water recycling than urban areas, and the former tend to have more rain than the later. Therefore, 4 land use types – forest, urban, agriculture, and water body & wetland, were applied in the study as dummy variables.

Hence, variables in this study were:

- Mean annual rainfall (millimeter): Dependent variable
- X coordinate (meter): Independent variable
- Y coordinate (meter): Independent variable
- Elevation (meter): Independent variable
  Assumption: Higher elevation, larger rainfall quantity.
- Distance from the South China (kilometer): Independent variable
  Assumption: Areas close to the sea tend to have more rain.
- Forest area: Dummy variable
  Assumption: Forest areas tend to have large amount of rainfall.
- Urban area: Dummy variable
  Assumption: Urban areas tend to have small amount of rainfall.
- Agricultural area: Dummy variable
  Assumption: Agricultural areas tend to have small rainfall quantity.
- Water body & Wetland: Dummy variable
  Assumption: Water bodies and Wetlands tend to have large amount of rainfall.

3.2 Regression Analysis

3.2.1 Data Preparation:

According to an assumption that relationship between y and x is linear, scatter plots of y and each independent variable were created to explore the linearity. If the relationship was not linear, data transformation was needed, and the transformation applied in this study were $1/x$, $\sqrt{x}$, $\ln(x)$, and $\log(x)$. The transformed data that gave the linear relationship were added as one of independent variables.

In this study, relationship between elevation and rainfall, and distance to the South China and rainfall were not linear. Thus, the data were transformed and 1/Elev and 1/Dist gave linear relationship as shown in Figure 2.

Next, independent variables were interacted with each other, and then interacted these terms with dummy variables. Finally, input all of these terms in NCSS.

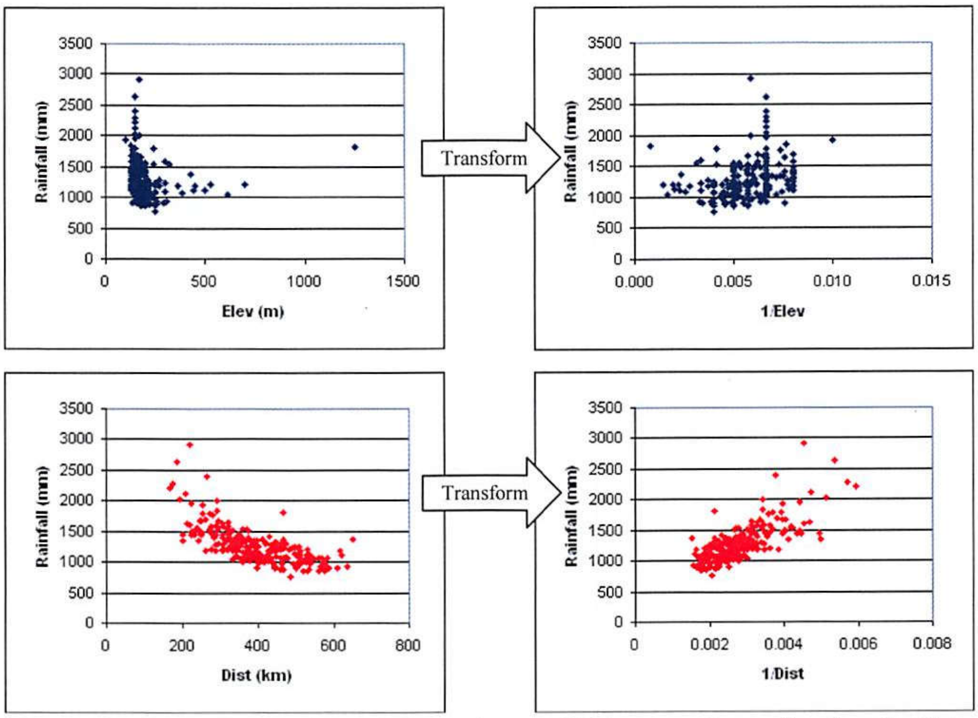


Figure 2. Transformation of Elevation and Distance from the South China

3.2.2. Variable Selection:

There were 86 regressors (including interaction terms) in this study. Therefore, variable selection routine had to be performed. First, backward stepwise regression was run, 28 independent variables were selected and R^2 was 0.77. After ran multiple regression with those variables, severe multicollinearily problem occurred. Next, forward stepwise regression was run, 4 independent variables were chosen, and R^2 was 0.64. Then these variables were plugged in multiple regression and no multicollinearity occurred. For this reason, forward stepwise method was applied for variable selection in this study.

3.2.3 Outlier and Assumption Checking:

- To check if outliers existed in the data set, Cook's D values in Regression Diagnostic Section were investigated. If Cook's D > 1.0, data were potential outliers.
- To prove normally distributed residuals, these output were investigated:
  (1) The Normal Probability Plot: If all residuals fell within the confidence bands, residuals were normally distributed.
  (2) The Normal Assumption Section: If null hypothesis of normality was accepted, residuals were normally distributed.

  If the distribution of residuals was not normal, then trend surface of residuals was plotted to explore if they were randomly distributed. If big residuals showed in a specific location, physical characters of the location was then inspected and finally created a new dummy variable. Next, interacted this dummy with the existing variables and performed forward stepwise and multiple regression.
- From Residual VS Predicted plot and Residual VS Predictors plots, if those plots showed rectangular shape, variance of residual was constant.
- In Multicollinearity Section, if any variable had a Variance Inflation Factor (VIF) bigger than 10, multicollinearity was exist. Then some variables were removed and re-run multiple regression until there was no multicollinearity problem.

3.2.4 Reliability of Regression:

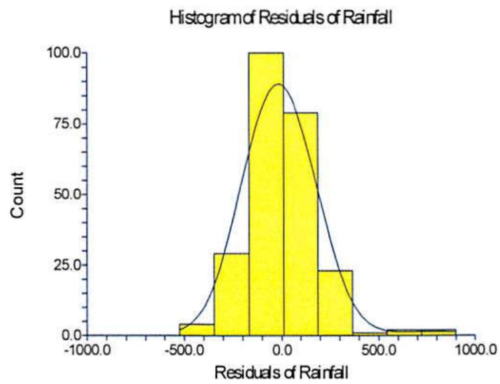
- Significance tests of the regression coefficients were proved in the Regression Equation Section. In this study H_0: $\beta_i = 0$ and confidence level was 95%.
- R^2 (Coefficient of determination) represents variation in mean annual rainfall explained by the independent variables in the regression model.
- F-ratio in the Analysis of Variance Section was applied to validate the overall strength of the model (H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$).

4. Results

After input 86 independent variables, forward stepwise regression selected 4 independent variables: Elev, 1/(Elev), 1/(Dist), and Water/($Dist^2$) *(Note: Elev = Elevation, Dist = distance from the South China, and Water is a dummy variable representing water bodies and wetlands)*. After applied these variables in multiple regression, R^2 of the model is 0.64 but residuals are not normally distributed as shown in the following Normality Tests Section. All null hypotheses of normality are rejected at 95% confidence level. Moreover, histogram of residuals and normality of probability plot in Figure 3 are evidence of the non-normality.

Normality Tests Section

Test Name	Test Value	Prob Level	Reject H0 At Alpha = 5%?
Shapiro Wilk	0.9425	0.000000	Yes
Anderson Darling	1.4750	0.000836	Yes
D'Agostino Skewness	5.7167	0.000000	Yes
D'Agostino Kurtosis	5.3534	0.000000	Yes
D'Agostino Omnibus	61.3395	0.000000	Yes



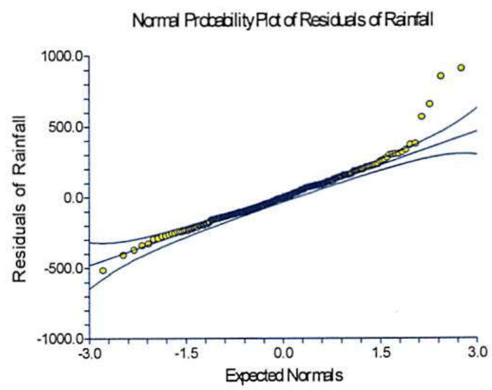


Figure 3. Non-normally distributed Residuals

In order to check spatial distribution of residuals, a residual surface was generated as illustrated in Figure 4. From the residual map, there is an area in the northeast (next to the Mekong and near large forest area in Laos) that value of mean annual rainfall is extremely under predicted. Large quantity of rainfall in this small area may be influenced by the Mekong and big forest area in Laos. Hence, a dummy variable (named Laos) was created to identify this area. Then, the Laos variable was interacted with other existing variables, and forward stepwise regression was run.

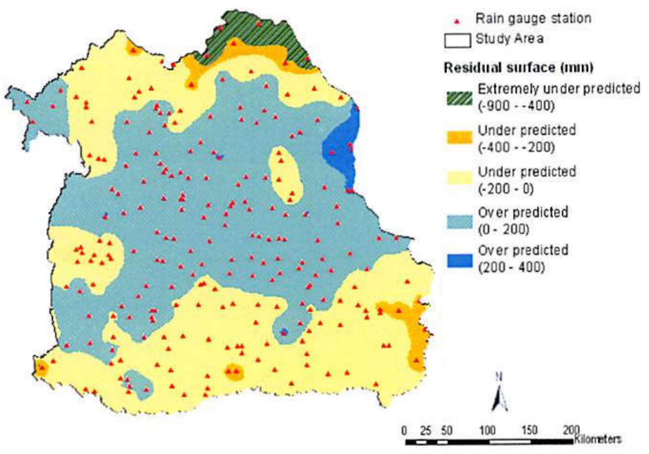


Figure 4. Residual Surface of Mean Annual Rainfall

The forward stepwise regression selected 10 independent variables : 1/(Dist), XY, X(Elev), X(Dist), X/(Elev), X/(Dist), Y(Dist), Y/(Dist), Y^2, and (Elev/Dist)Laos. These variables were then applied in multiple regression and an output shows that R^2 is better (0.86) and residuals are normally distributed as shown in the following Normality Tests Section. All assumptions that residual distribution is normal cannot be rejected. However, the Normal Probability Plot of Residuals in Figure 5 portrays that residual distribution is not perfectly normal because several residuals lie outside the confidence bands.

Normality Tests Section

Test Name	**Test Value**	**Prob Level**	**Reject H0 At Alpha = 5%?**
Shapiro Wilk	0.9917	0.192885	No
Anderson Darling	0.2854	0.627070	No
D'Agostino Skewness	-1.2502	0.211215	No
D'Agostino Kurtosis	1.9041	0.056898	No
D'Agostino Omnibus	5.1887	0.074696	No

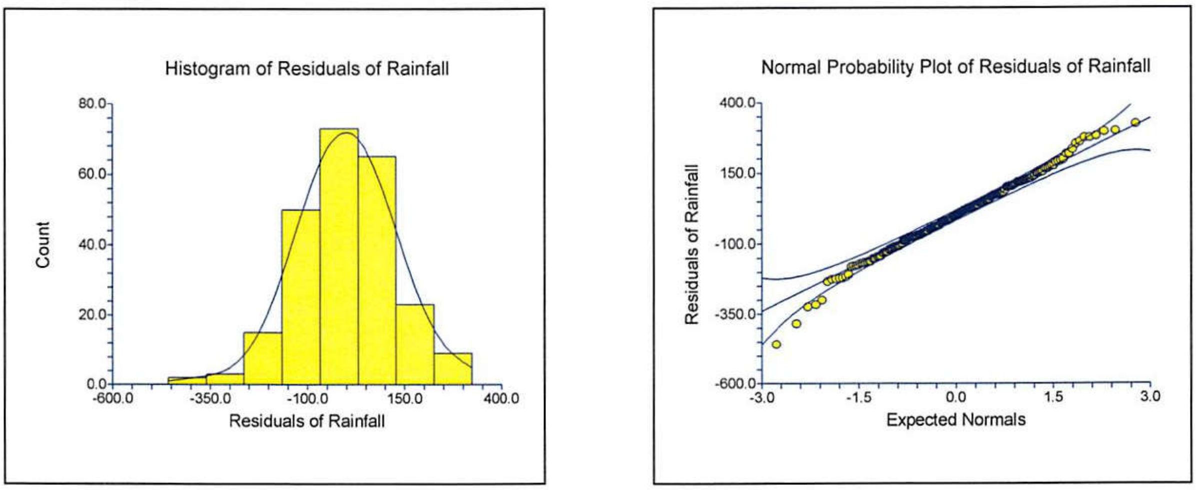


Figure 5. Imperfect Normal Distribution of Residuals

Although R^2 of this model is higher (0.83) than the previous (0.64) that did not include the Laos dummy variable, this model has a severe multicollinearity problem as shown in the following Multicollinearity Section. There are 6 variables having big VIF.

Multicollinearity Section

Independent Variable	Variance Inflation Factor	R2 Versus Other I.V.'s	Tolerance	Diagonal of X'X Inverse
XY	0.0000			0
X(Elev)	0.0000			0
X(Dist)	50.5224	0.9802	0.0198	2.877008E-16
X/(Elev)	19.3993	0.9485	0.0515	8.464323E-08
X/(Dist)	6774.0069	0.9999	0.0001	9.335535E-05
Y(Dist)	6874.8443	0.9999	0.0001	1.052469E-15
Y/(Dist)	16603.9470	0.9999	0.0001	2.460118E-05
YY	0.0000			0
(Elev/Dist)Laos	1.6018	0.3757	0.6243	0.8528636
1/(Dist)	39602.3015	1.0000	0.0000	2.432929E+08

In order to solve the multicollinearity problem, some variables have to be eliminated and accept the smaller R^2.

- First, 1/(Dist) was deleted due to its largest VIF (39602). R^2 of this new model is 0.81 (slightly smaller) and residuals are still normally distributed as shown in the following Normality Tests Section. However, the severe multicollinearity still exists (in the following Multicollinearity Section).

Normality Tests Section

Test Name	Test Value	Prob Level	Reject H0 At Alpha = 5%?
Shapiro Wilk	0.9920	0.217303	No
Anderson Darling	0.3491	0.475017	No
D'Agostino Skewness	-0.9907	0.321823	No
D'Agostino Kurtosis	2.0385	0.041499	Yes
D'Agostino Omnibus	5.1370	0.076649	No

Multicollinearity Section

Independent Variable	Variance Inflation Factor	R2 Versus Other I.V.'s	Tolerance	Diagonal of X'X Inverse
XY	0.0000			0
X(Elev)	0.0000			0
X(Dist)	50.4999	0.9802	0.0198	2.875728E-16
X/(Elev)	18.7521	0.9467	0.0533	8.181919E-08
X/(Dist)	761.2948	0.9987	0.0013	1.049171E-05
Y(Dist)	2570.3695	0.9996	0.0004	3.934975E-16
Y/(Dist)	459.1326	0.9978	0.0022	6.802722E-07
YY	0.0000			0
(Elev/Dist)Laos	1.5018	0.3342	0.6658	0.7996584

- Second, Y(Dist) was remove because of its biggest VIF and plugged the rest variables in the multiple regression. R^2 of the new model is 0.78 and residuals are still normally distributed, but the severe multicollinearity still exists.
- Third, eliminated X/(Dist) and ran multiple regression. R^2 of the new model is 0.77 , distribution of residuals is normal, and multicollinearity is a mild problem as shown in the following sections.

Normality Tests Section

Test Name	Test Value	Prob Level	Reject H0 At Alpha = 5%?
Shapiro Wilk	0.9965	0.879267	No
Anderson Darling	0.2128	0.854126	No
D'Agostino Skewness	0.5248	0.599743	No
D'Agostino Kurtosis	1.1519	0.249365	No
D'Agostino Omnibus	1.6022	0.448825	No

Multicollinearity Section

Independent Variable	Variance Inflation Factor	R2 Versus Other I.V.'s	Tolerance	Diagonal of X'X Inverse
XY	0.0000			0
X(Elev)	0.0000			0
X(Dist)	17.6074	0.9432	0.0568	1.002657E-16
X/(Elev)	16.3444	0.9388	0.0612	7.131407E-08
Y/(Dist)	34.2021	0.9708	0.0292	5.067545E-08
YY	0.0000			0
(Elev/Dist)Laos	1.2424	0.1951	0.8049	0.6615242

- Forth, deleted Y/(Dist) due to its largest VIF and ran multiple regression. R^2 of the new model is 0.74. Although there is no multicollinearity in the model, residuals are not normally distributed (as shown in the following section). Thus, Y/(Dist) cannot be eliminated.

Normality Tests Section

Test Name	Test Value	Prob Level	Reject H0 At Alpha = 5%?
Shapiro Wilk	0.9796	0.001544	Yes
Anderson Darling	0.7241	0.058995	No
D'Agostino Skewness	3.4667	0.000527	Yes
D'Agostino Kurtosis	2.2208	0.026363	Yes
D'Agostino Omnibus	16.9501	0.000209	Yes

Multicollinearity Section

Independent Variable	Variance Inflation Factor	R2 Versus Other I.V.'s	Tolerance	Diagonal of X'X Inverse
XY	0.0000			0
X(Elev)	0.0000			0
X(Dist)	0.0000			0
X/(Elev)	14.4748	0.9309	0.0691	6.315642E-08
YY	0.0000			0
(Elev/Dist)Laos	1.0952	0.0869	0.9131	0.5831459

Right now, independent variables are XY, X(Elev), X(Dist), X/(Elev), Y/(Dist), Y^2, and (Elev/Dist)Laos. Because Y/(Dist) cannot be removed from the model, another variable that has strong correlation with Y/(Dist) must be deleted. According to the following Correlation Matrix Section, Y/(Dist) has strong correlation with XY. Thus, XY is deleted and the rest variables are put in the multiple regression analysis.

Correlation Matrix Section

	Y/(Dist)	YY	st)Laos (Elev/Di	Rainfall
XY	0.7579	0.0738	0.1082	0.6508
X(Elev)	0.4387	0.0784	0.0124	0.4768
X(Dist)	-0.1468	-0.7132	-0.1510	0.0375
X/(Elev)	0.5250	-0.1673	0.0467	0.5137
Y/(Dist)	1.0000	0.6252	0.2987	0.7571
YY	0.6252	1.0000	0.2724	0.3602
(Elev/Dist)Laos	0.2987	0.2724	1.0000	0.5335
Rainfall	0.7571	0.3602	0.5335	1.0000

- Finally, plug in X(Elev), X(Dist), X/(Elev), Y/(Dist), Y^2, and (Elev/Dist)Laos to the multiple regression analysis. R^2 of the model is 0.75 and residuals are normally distributed. Moreover, there is no multicollinearity problem anymore as shown in the following sections (output with more details are shown in Appendix).

Normality Tests Section

Test Name	**Test Value**	**Prob Level**	**Reject H0 At Alpha = 5%?**
Shapiro Wilk	0.9941	0.469973	No
Anderson Darling	0.3307	0.513484	No
D'Agostino Skewness	1.6443	0.100118	No
D'Agostino Kurtosis	0.5708	0.568104	No
D'Agostino Omnibus	3.0295	0.219861	No

Multicollinearity Section

Independent Variable	**Variance Inflation Factor**	**R2 Versus Other I.V.'s**	**Tolerance**	**Diagonal of X'X Inverse**
X(Elev)	0.0000			0
X(Dist)	0.0000			0
X/(Elev)	8.0505	0.8758	0.1242	3.512594E-08
Y/(Dist)	8.3518	0.8803	0.1197	1.237438E-08
YY	0.0000			0
(Elev/Dist)Laos	1.1479	0.1289	0.8711	0.611208

This regression model also complies with regression assumption, such as linearity, normality, constant variance and multicollinearity :

- Linearity: According to plots of rainfall versus the independent variables in the Appendix, the plots show linear relationship between dependent variable and regressors.
- Normality: Residuals are normally distributed as shown in the Normal Probability Plot and Normality Tests Section in the Appendix.

- Constant variance: According to Residuals vs. Predicted plot and Residuals vs. Predictor Plots (in the Appendix), the plots show a rectangular shape. Therefore, variance of residuals is constant.
- Outliers: Following the Regression Diagnostics Section in the Appendix, Cook's D values are less than 1.00, thus there is no outlier in this data set.
- Multicollinearity: All VIFs are less than 10. From Eigenvalues of Centered Correlation, all condition numbers are smaller than 100. For these reasons, multicollinearity is not a problem.

The model for mean annual rainfall (mm) estimation is:
807.875957957589+ 4.31041044320538E-06*X(Elev)-1.61265829155383E-06*X(Dist)+ .098650207451646*X/(Elev)+ 6.63301935873058E-02*Y/(Dist)-3.58865152795949E-11*YY+ 1352.62690056273*(Elev/Dist)Laos

Intercept of the model is around 807.8760 mm.
When the remaining independent variables are held constant:

If X(Elev) increases 1 unit, estimated mean annual rainfall increases ≈ 4.3104 mm.
If X(Dist) increases 1 unit, the mean annual rainfall decreases ≈ 0.0000016 mm.
If X/(Elev) increases 1 unit, the mean annual rainfall increases ≈ 0.0987 mm.
If Y/(Dist) increases 1 unit, the mean annual rainfall increases ≈ 0.0663 mm.
If YY increases 1 unit, the mean annual rainfall decreases ≈ 0.000000000036 mm.
If (Elev/Dist)Laos increases 1 unit, the mean annual rainfall increases ≈1352.6269 mm

Next, reliability of the model is validated using Standard error of β, F-ratio, ANOVA, R^2, and Adjusted R^2

- Standard error of β
  H_0: $\beta_i = 0$
  The null hypothesis is tested via t-test at 95% confidence level. The following Regression Equation Section indicates that coefficients of Intercept, X/(Elev), Y/(Dist), and (Elev/Dist)Laos are different from zero but those of X(Elev), X(Dist), and YY are not different from zero.

Regression Equation Section

Independent Variable	**Regression Coefficient b(i)**	**Standard Error Sb(i)**	**T-Value to test H0:B(i)=0**	**Prob Level**	**Reject H0 at 5%?**	**Power of Test at 5%**
Intercept	807.8760	205.2812	3.935	0.0001	Yes	0.9750
X(Elev)	0.0000	0.0000	0.000	1.0000	No	0.0500
X(Dist)	0.0000	0.0000	0.000	1.0000	No	0.0500
X/(Elev)	0.0987	0.0285	3.456	0.0007	Yes	0.9308
Y/(Dist)	0.0663	0.0169	3.915	0.0001	Yes	0.9737
YY	0.0000	0.0000	0.000	1.0000	No	0.0500
(Elev/Dist)Laos	1352.6269	119.0679	11.360	0.0000	Yes	1.0000

- F-ratio, ANOVA, and R^2
  H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$
  According to the following Analysis of Variance Section, the null hypothesis is rejected at 95% confidence level (H_0 would be rejected up to 99.99% confidence level). Therefore, approximately 75.27% of the variation in mean annual rainfall is explained by the independent variables.

Analysis of Variance Section

Source	DF	R2	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1		3.889348E+08	3.889348E+08			
Model	6	0.7527	1.645119E+07	2741865	118.208	0.0000	1.0000
Error	233	0.2473	5404513	23195.34			
Total(Adjusted)	239	1.0000	2.18557E+07	91446.45			

- Adjusted R^2

  Adjusted R^2 =0.7464

  The adjusted R^2 is slightly smaller than the R^2 because sample size of this study is large (n=240).

Conclusion

- By applying X(Elev), X(Dist), X/(Elev), Y/(Dist), Y^2, and (Elev/Dist)Laos in the regression model, 75.27% of variation in mean annual rainfall of the Northestern Thailand is explained.

- The dummy variable identifying an area near the Mekong and big forest area in Laos enhances performance of the model.

- To improve accuracy of the model, other factors influencing quantity of rainfall (such as physical characters of study area and neighbor countries) should be studied.

- Regression is a powerful tool for rainfall estimation because various types of data can be applied as input of the model.

References

Colman, A. Davey, M., Harrison, M., and Evans, A., 1998. "Prediction of March-April-May 1998 Rainfall in Northeast Brazil Using Input from Multiple Regression, Discriminant Analysis and an Atmospheric Global Circulation Model". 5 April 2006. <http://grads.iges.org/ellfb/Mar98/co2.html>.

Croke, B.F.W., Gilmour, J.K. and Newham, L.T.H. 2001. "A Comparison of Rainfall Estimation Techniques". 5 April 2006. < http://eprints.anu.edu.au/archive/00000634/ >.

Ebdon, D. 1977. *Statistics in Geography: Apractical Approach.* Oxford: Basil Blackwell.

Edwards, A.L. 1984. *An Introduction to Linear Regression and Correlation.* 2nd ed. New york: W.H. freeman and Company.

Foody, G.M., 2003. "Geographical Weighting as a Further Refinement to Regression Modeling: An Example Focused on the NDVI – Rainfall Relationship". *Remote Sensing of Environment*, 88, 283-293.

Freund, R.J. and Minton, P.D. 1979. *Regression Methods, A Tool for Data Analysis.* New York: Marcel Dekker.

Greischar, L. and Hastenrath, S. 2000. "Multiple Regression and Discriminant Analysis to Predict Mar-Apr-May-Jun 2000 Rainfall in Northeast Brazil". 5 April 2006. < http://grads.iges.org/ellfb/Mar00/greischar.oo.htm>.

Hu, J., Chen, J., Liu, G., Tsai, M. and Chang, M. 2006. "Over-ocean rainfall retrieval from TRMM/TMI data during the Typhoon season". 5 April 2006. < http://ams.confex.com/ams/Annual2006/techprogram/paper_103044.htm >.

Lachniet, M.S. and Patterson W.P. 2005. "Use of correlation and stepwise regression to evaluate physical controls on the stable isotope values of Panamanian rain and surface waters". 24 April 2006.
< http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V6C-4HNS8NN-1&_user=10&_coverDate=11%2F28%2F2005&_rdoc=1&_fmt=summary&_orig=browse&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=784e915c867ad16ec26c6167096952b7 >.
Rogerson, P.A. 2005. *Statistical Methods for Geography.* Thousand Oaks: Sage.
Sen, A. and Srivastava M. 1990. *Regression Analysis: Theory, Methods and Applications.* New York: Springer-Verlag.
Silk, J. 1979. *Statistical Concepts in Geography.* London: George Allen & Unwin.
Singhrattna, N., Rajagopalan, B., Kumar, K.K., and Clark, M. 2005. "Interannual and Interdecadal of Thailand Summer Monsoon Season". 5 April 2006.
<http://sciencepolicy.colorado.edu/admin/publication_files /resource-1701-2005.09.pdf>.
Vicente, G.A. and Scofield, R.A. 1998. "Satellite Rainfall Estimates in Real Time for Applications to Flash Flood Watches And Warnings, Heavy Precipitation Forecasting and Assimilation on Numerical Weather Prediction Models". 5 April 2006.
<http://www.srh.noaa.gov/topics/attach/html/ssd98-18.htm>.
Yin, X. and Nicholson, S.E., 2002. "Interpreting Annual rainfall from the Level of Lake Victoria". *Journal of Hydrometeorology*, 3, 406-416.