Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 10, Number 5 (2014), pp. 765-768
© Research India Publications
http://www.ripublication.com

Predicting Run off Using A Multiple Linear Regression Equation In Sakaeo, Thailand

J. Mekparyup[1], K. Saithanu[2] and W. Laipradith[3]

[1,2,3]Department of Mathematics, Faculty of Science, Burapha University
169 Muang, Chonburi, Thailand
[1]jatupat@buu.ac.th, [2]ksaithan@buu.ac.th, [3]wow_waru@hotmail.com

Abstract

The objectives of the present study were to estimate runoff in Sakaeo, Thailand, by using a multiple linear regression equation. Runoff and meteorological factors were collected from Hydrology and Water Management Center for Eastern Region since 2010 to 2013 for building equation. The result found that the regression equation for estimation the runoff was runoff=65.93+0.08919 Rainfall + 0.07887 Rainfall 1+0.926RH_3 with standard error of estimation 9.56705 and adjusted coefficient of determination 0.607.

Mathematics Subject Classification: 62J05

Keywords: multicollinearity, variance inflation factor, best subset method

Introduction

In rural areas of Thailand, the fresh water is essential to agriculture because the water supply can not reach these areas, especially in areas that lack of civilization. One of the main occupations in Sakaeo is agriculture so irrigation system is important to farm. The mainly water that culturists use is runoff occurred by water resources such as Prasatueng Cannal, Prapong River, Bangpakong River, etc. Therefore, the management of runoff is very considerable to agriculture in Sakaeo.

Materials And Methods

The data, runoff, temperature, rainfall and relative humidity, were collected from Hydrology and Water Management Center for Eastern Region since 2010 to 2013. Simple correlation coefficients (R) were firstly calculated to identify relationship among these data.

Building The Multiple Regression Equation

The multiple linear regression equation to estimate monthly runoff was generated by regression model [1][2][3] as Equation (1).

$$\begin{aligned} Runoff &= \beta_0 + \beta_{1i}T + \beta_{1i-1}T_1 + \beta_{1i-2}T_2 + \beta_{1i-3}T_3 + \beta_{2i}R + \beta_{2i-1}R_1 + \beta_{2i-2}R_2 \\ &+ \beta_{2i-3}R_{i-3} + \beta_{3i}H + \beta_{3i-1}H_1 + \beta_{3i-2}H_2 + \beta_{3i-3}H_3 + \varepsilon \end{aligned} \qquad (1)$$

The model is compose of one dependent variable; *Runoff*=Monthly runoff, and twelve independent variables; *T*=Maximum monthly temperature, T_j=Maximum monthly temperature in time lag *j*, *R*=Monthly rainfall, R_j=Monthly rainfall in time lag *j*, *H*=Monthly relative humidity, H_j=Monthly relative humidity in time lag *j*, (*j*=1, 2, 3) and ε = Error of regression model.

Checking Asuumptions For Multiple Regression Analysis

After obtained the best fitted multiple regression equation, the assumptions checking for multiple regression analysis was proceeded. There are four assumptions to be tested; (I) normality of the error distribution using Anderson-Darling statistic by Equation (2) [4];

$$AD = -n - \sum_{i=1}^{n} \frac{2i-1}{n}\left[ln F(Y_i) + ln(1 - F(Y_{n+1-i}))\right] \qquad (2)$$

(II) independence of the errors using Durbin-Watson statistic by Equation (3) [5];

$$DW = \sum_{i=1}^{n} (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2 \Big/ \sum_{i=1}^{n} \hat{\varepsilon}_i^{\,2} \qquad (3)$$

(III) homoscedasticity (constant variance) of the errors using Breusch-Pagan statistic by Equation (4) [6];

$$BP = \frac{SSR^*}{2} \div \left(\frac{SSE}{n}\right)^2 \qquad (4)$$

where SSR^* = Sum of squares in regression between e_j^2 = *j*th residual and x_{ij}, *SSE* = Sum of squares in regression error between y_j and x_{ij}; (IV) multicollinearity among predictor variables using Variance Inflation Factor (VIF) Equation (5)

$$VIF_j = \frac{1}{1 - R^2_{j|others}} \qquad (5)$$

where $R^2_{j|others}$ = Multiple coefficient of determination between x_{ij} and all x_i. After tested all assumptions, the comparison between the real values and the estimated values of the runoff from the obtained multiple regression equation was plotted.

Validation Between Predicted Value And Observed Runoff

Finally, the comparison between the real values and the predicted values of runoff considering by the obtained multiple regression equation was then created time series plot and scatter plot when checking all assumptions of multiple regression analysis was determined.

Results And Discussion

These was highly positive correlation between *Runoff* and R (R=0.561, p-value=0.000), *Runoff* and R_1 (R=0.645, p-value=0.000) and *Runoff* and H (R=0.598, p-value=0.000) which was the same previous studies [2][7]. According to the best subsets method, R (rainfall), R_1 (rainfall in time lag 1) and H_3 (relative humidity in time lag 3) were selected to build the possible multiple linear regression equation for estimation runoff as Equation (6) with the Mallow C-p=6.6, S=9.5671, R^2=0.641 and R^2_{adj}=0.607.

$$\widehat{Runoff} = -65.93 + 0.089191R + 0.07887R_1 + 0.926H_3 \tag{6}$$

After obtained the equation, (I) the test of normality was determined again and was found that hypothesis testing of normality was satisfied with AD=0.439 (p-value=0.277). (II) The test of independence of the errors was tested by Durbin-Watson statistic value (DW=1.78276, DL=1.318) so the errors were independent. (III) The test of homoscedasticity of error variation was tested by Breush-Pagan statistic (BP=5.5227, p-value=0.011) so the error variances were constant. (IV) Test of multicollinearity: the VIF values of R, R_1 and H_3 were calculated and the results were as 1.4, 1.2 and 1.2 consequently. All VIF values were less than 5 then there was no relationship among independent variables in multiple regression equation [8]. After all assumptions were validated, plotting between predicted runoff by Equation (6) and the real values was compared by graph of time series and scatter plot as Figure (1a) and Figure (1b) respectively. It was shown that the both graphs were closely plotted with the correlation coefficient 0.801.

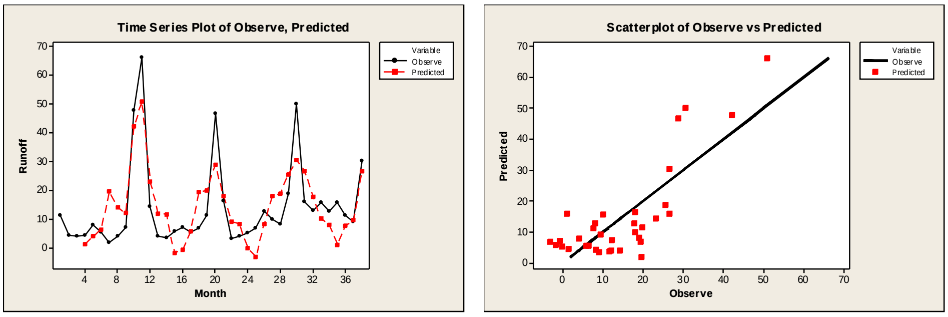

Figure 1: Comparison between observed and predicted runoff (a) Time series plot, (b) Scatter plot

Discussion

The predictors used to predict the monthly runoff in Sakaeo were Rainfall (*R*), Rainfall in time lag 1 (R_1) and Relative humidity in time lag 3 (H_3) with adjusted coefficient of determination (R^2_{adj}) 0.607 and the standard error of the estimation (S) 9.5671. The accuracy of estimation was shown by comparing the graph between the observe values and the predicted values from the multiple linear regression equation.

Acknowledgement

We are grateful to Hydrology and Water Management Center for Eastern Region, Thailand for kindly providing all data.

References

[1] Zhao, Q.H., Liu, S.L., Deng, L., Dong, S.K., Wang, C., & Yang, J.J., 2012, “Assessing the damming effects on runoff using a multiple linear regression model: A case study of the Manwan Dam on the Lancang River,” Procedia Environmental Sciences, 13, 1771-1780.

[2] Schär, C., Vasilina, L., Pertziger, F., & Dirren, S., 2004, “Seasonal runoff forecasting using precipitation from meteorological data assimilation systems,” Journal of Hydrometeorology, 5(5), 959-973.

[3] Archer, D. R., & Fowler, H. J., 2008, “Using meteorological data to forecast seasonal runoff on the River Jhelum, Pakistan,” Journal of Hydrology, 361(1), 10-23.

[4] Lewis, P.A.W., 1961, “Distribution of the Anderson-Darling Statistic,” The Annals of Mathematical Statistics, 32(4), 1118-1124.

[5] Durbin, J., & Watson, G.S., 1951, “Testing for Serial Correlation in Least Squares Regression II,” Biometrika, 38(2), 159-177.

[6] Breusch T.S., & Pagan, A.R., “A Simple Test for heteroscedasticity and Random Coefficient Variation,” Econometrica, 47(5), 1287-1294.

[7] Yang, Y., & Tian, F., 2009, “Abrupt change of runoff and its major driving factors in Haihe River Catchment, China,” Journal of Hydrology, 374(3), 373-383.

[8] Kutner, M.H., Christopher, J.N., & Neter, J., 1996, “Applied liner regression models, 4th ed,” McGraw-Hill/Irwin , USA.