International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:10, 2008

Empirical Statistical Modeling of Rainfall Prediction over Myanmar

Wint Thida Zaw and Thinn Thu Naing

***Abstract*—**One of the essential sectors of Myanmar economy is agriculture which is sensitive to climate variation. The most important climatic element which impacts on agriculture sector is rainfall. Thus rainfall prediction becomes an important issue in agriculture country. Multi variables polynomial regression (MPR) provides an effective way to describe complex nonlinear input output relationships so that an outcome variable can be predicted from the other or others. In this paper, the modeling of monthly rainfall prediction over Myanmar is described in detail by applying the polynomial regression equation. The proposed model results are compared to the results produced by multiple linear regression model (MLR). Experiments indicate that the prediction model based on MPR has higher accuracy than using MLR.

***Keywords*—**Polynomial Regression, Rainfall Forecasting, Statistical forecasting.

I. INTRODUCTION

RAINFALL information is important for food production plan, water resource management and all activity plans in the nature. The occurrence of prolonged dry period or heavy rain at the critical stages of the crop growth and development may lead to significant reduce crop yield. Myanmar is an agricultural country and its economy is largely based upon crop productivity. Thus rainfall prediction becomes a significant factor in agricultural countries like Myanmar.

A wide range of rainfall forecast methods are employed in weather forecasting at regional and national levels. Fundamentally, there are two approaches to predict rainfall. They are empirical and dynamical methods.

The empirical approach is based on analysis of historical data of the weather and its relationship to a variety of atmospheric and oceanic variables over different parts of the world. The most widely use empirical approaches used for climate prediction are regression, artificial neural network, stochastic, fuzzy logic and group method of data handling.

In dynamical approach, predictions are generated by physical models based on systems of equations that predict the evolution of the global climate system in response to initial atmospheric conditions. The dynamical approaches are implemented using numerical weather forecasting method.

Wint Thida Zaw is with the Umiversity of Computer Studies, Yangon, Myanmar (corresponding author to provide phone: 95-1-661713; fax: 95-1-665686; e-mail: wintthida@gmail.com).

Thinn Thu Naing is with the Umiversity of Computer Studies, Yangon, Myanmar (corresponding author to provide phone: 95-1-661713; fax: 95-1-665686; e-mail:ucsy21@most.gov.mm).

In this paper, rainfall prediction model is implemented with the use of empirical statistical technique, MPR. We use 37 years (1970-2006) datasets of the global climate data such as sea surface temperature (SST), India Ocean Dipole (IOD), Southern Oscillation Index (SOI), Oceanic Nino Index (ONI) and premonsoon month rainfall amount over 17 Myanmar weather stations as local weather data. The model forecasts monthly rainfall amount in summer monsoon season (in mm). The resulted rainfall amounts are intended to help farmers in making decision concerning with their crop.

It is possible to predict monthly rainfall amount with one month ahead with acceptably accuracy. The experimental results show that there is a close agreement between the predicted and actual rainfall amount.

II. RELATED WORK

Accurate and timely weather forecasting is a major challenge for the scientific community. Rainfall prediction modeling involves a combination of computer models, observation and knowledge of trends and patterns. Using these methods, reasonably accurate forecasts can be made up. Several recent research studies have developed rainfall prediction using different weather and climate forecasting methods.

Regression is a statistical empirical technique and is widely used in business, the social and behavioral sciences, the biological sciences, climate prediction, and many other areas.

N. Sen [1] has presented long-range summer monsoon rainfall forecast model based on power regression technique with the use of Ei Nino, Eurasian snow cover, north west Europe temperature, Europe pressure gradient, 50 hPa Wind pattern, Arabian sea SST, east Asia pressure and south Indian ocean temperature in previous year. The experimental results showed that the model error was 4%.

S. Nkrintra, et al. [2] described the development of a statistical forecasting method for SMR over Thailand using multiple linear regression and local polynomial-based nonparametric approaches. SST, sea level pressure (SLP), wind speed, EiNino Southern Oscillation Index (ENSO), IOD were chosen as predictors. The experiments indicated that the correlation between observed and forecast rainfall was 0.6.

T. Sohn, et al. [3] has developed a prediction model for the occurrence of heavy rain in South Korea using multiple linear and logistics regression, decision tree and artificial neural network. They used 45 synoptic factors generated by the numerical model as potential predictors.

M. T. Mebrhatu et al. [4] modeled for prediction categories of rainfall (below, above, normal) in the highlands of Eritrea. The most influential predictor of rainfall amount was the southern Indian Ocean SST. Experimental results showed that the hit rate for the model was 70%.

H. Hasani et al. [5] proposed human height prediction model based on multiple polynomial regression that was used successfully to forecast the growth potentials of height with precision and was helpful in children growth study.

Vaccari et al. [6] modeled plant motion time-series and nutrient recovery data for advanced life support using multi variable polynomial regression.

The following sections present an approach to the development of rainfall forecasting system. Firstly, second-order polynomial regression is discussed in section 3. Next, how to forecast precipitation with the use of MPR is described in section 4. Some experimental results are reported in section 5. Finally, the paper is concluded in section 6.

III. SECOND-ORDER POLYNOMIAL REGRESSION

Regression is a statistical empirical technique that utilizes the relation between two or more quantitative variables on observational database so that an outcome variable can be predicted from the others.

Polynomial regression produces a polynomial describing the relationship between any set of inputs and corresponding output. Polynomial regression model which contains more than two predictor variables is called MPR. If a MPR model which contains n predictors expressing with the first and second and third powers of the predictors, this polynomial model is called a third-order polynomial model with n predictor variable.

The polynomial regression can become numerically unstable, especially if the degree is high or the function domain is not centered at zero. If the polynomial degree is too low, it will not be able to represent the complexity of the function being learned.

Thus, the statistical relationship between rainfall amount and other climatic data is searched with the use of second-order MPR equation which contains added terms and nonlinear cross product interaction of n predictors expressing with the first and second power of the predictors.

The second-order MPR used in this study is in the form of equation (1) where y_t represents the predictand, representing SMR at time t. $x_{1t}, x_{2t} \ldots x_{nt}$ are predictors representing premonsoon month temperature, rainfall, IOD, SOI, SST at time t. $\beta_{1t}, \beta_{2t}, \ldots$ are least square regression parameters estimated from the observed data. ***e*** denotes the error that is difference between predicted and actual values.

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_n x_{nt} + \beta_{n+1} x_{1t}^2 + \beta_{n+2} x_{1t} x_{2t} + \ldots + \beta_{2n} x_{1t} x_{nt} + \beta_{2n+1} x_{2t}^2 + \beta_{2n+2} x_{2t} x_{3t} + \ldots + \beta_{3n-1} x_{2t} x_{nt} + \ldots + \beta x_{nt}^2 + e \quad (1)$$

The second-order MPR model used in this study is in the matrix form of

$$Y = X\beta + e \quad (2)$$

where

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{n1} & x_{11}^2 & x_{11}x_{21} & \ldots & x_{11}x_{n1} & x_{21}^2 & x_{21}x_{31} & x_{21}x_{41} & \ldots & x_{21}x_{n1} & \ldots & x_{n1}^2 \\ 1 & x_{12} & x_{22} & \ldots & x_{n2} & x_{12}^2 & x_{12}x_{22} & \ldots & x_{12}x_{n2} & x_{22}^2 & x_{22}x_{32} & x_{22}x_{42} & \ldots & x_{22}x_{n2} & \ldots & x_{n2}^2 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1t} & x_{2t} & \ldots & x_{nt} & x_{1t}^2 & x_{1t}x_{2t} & \ldots & x_{1t}x_{nt} & x_{2t}^2 & x_{2t}x_{3t} & x_{2t}x_{4t} & \ldots & x_{2t}x_{nt} & \ldots & x_{nt}^2 \end{pmatrix} \quad (3)$$

$$Y^T = (y_1 \quad y_2 \quad \ldots \quad y_t) \quad (4)$$

$$\beta^T = (\beta_0 \quad \beta_1 \quad \ldots) \quad (5)$$

The least square estimates of the regression parameters are

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (6)$$

The least square prediction $\hat{y}$ is

$$\hat{Y} = X\hat{\beta} \quad (7)$$

IV. RAINFALL FORECASTING USING MPR

The general processes of forecasting rainfall amount are

1. Collecting data
2. Reduction explanatory predictors
3. Building model using stepwise regression procedure
4. Validity check

Data are collected first. Data are separated into training data and test data.

Then the predictors which have high intercorrelation with others are reduced because the presence of many highly intercorrelated explanatory variables may substantially increase the sampling variation of the regression coefficients, and not improve, or even worsen the models' predictive ability. The correlation of variables is calculated using pearson correlation formula.

The next step of the reduction explanatory predictors is the building model with the use of training data. At this stage, predictors subset selection procedure is employed to identify the model containing small group of predictors which can give an optimal prediction. We use the stepwise regression technique. The stepwise selection procedure is the combination of forward selection and backward elimination. After adding one predictor to the subset or excluding one predictor from the subset, we find the least squares estimates of $\hat{\beta}$ from equation (6). We calculate the least square prediction $\hat{y}$ from equation (7) for the model. We compute RMSE using equation (8). This procedure is repeated until RMSE can be reduced significantly by adding another predictor or removing the predictor to the model.

Finally, the model built over training period is tested with test data to verify how much accuracy the model can give. The overview of forecasting system is illustrated in Fig. 1.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (8)$$

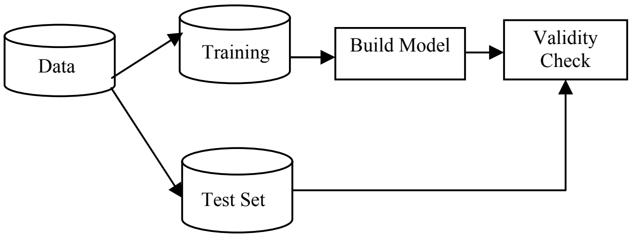

Fig. 1 Overview of the forecasting system

V. EXPERIMENTAL RESULTS

We have conducted experiments to evaluate the accuracy of rainfall prediction. The prediction results will be reported in this section. To measure the quality of the output, the forecasted rainfall amount is compared with actual rainfall amount. Experiments and graphs are reported for Pathein rain gauge station located in lower Myanmar and Magwe station located in upper Myanmar where rain fall prediction is needed more for agriculture planning and management.

For experiments, regional rainfall amount taken from rain gauge stations over Myanmar and large scale data such as East India SST, SOI, ONI taken from [7],[8],[9] are used as predictors. The following table shows the details of the predictors used for rainfall prediction.

TABLE I DETAIL OF PREDICTORS

Predictor	Symbol	Period	Location
Rainfall	P_1	Previous month	Predicted Station
ONI	P_2	MAM	5°N-5°S 120° 170°W
ONI	P_3	AMJ	5°N-5°S 120°-170°W
Temperature	P_4	Previous month	Predicted Station
IOD	P_5	April	90°E-110°E 10°S-0°N
IOD	P_6	May	
East India SST	P_7	April	50°E 70°E 10°S-10°N
East India SST	P_8	May	
SOI	P_9	Previous two months	17.5°S 149.6°W, 12.4°S 130.9°E
SOI	P_{10}	Previous month	
SST	P_{11}	Previous two months	15°-20° N 85°-90° E
SST	P_{12}	Previous two months	10°-15° N 85°-90° E
SST	P_{13}	Previous two months	5°-10° N 85°-90° E
SST	P_{14}	Previous two months	5°-10° N 90°-95° E
SST	P_{15}	Previous two months	5°-10° N 95°-100° E

Data over 1990-2005 are used as training data. Data of 2006 and 2007 are used as testing data. Myanmar receives its annual rainfall mainly in raining season which starts from June and ends in September.

Table 2 describes the predictors used in prediction and the error over training period in forecasting monthly rainfall amount during rainy season.

TABLEII SELECTED MODEL SKILLS FOR PATHEIN AND MAGWAE STATION OVER TRAINING PERIOD

Station	Month	RMSE	Predictors Used
Pathein	June	0.6	P_2, P_4, P_8, P_9
	July	0.4	P_3, P_7, P_4, P_{15}
	August	0.2	P_{10}, P_4, P_8, P_9
	September	1.3	P_1, P_{11}, P_3, P_9
Magwe	June	2	P_4, P_{10}, P_{14}, P_{15}
	July	27	P_2, P_{11}, P_{14} ,P_{15}
	August	5	P_2, P_4, P_5, P_{14}

The model evaluation is performed using test data of 2006. The trained model is tested with the use of 2006 test data and the test result is compared to the actual rainfall amount.

Fig.2 illustrates 2006 monthly rainfall prediction amount compared with actual rainfall amount in June, July, August and September in Pathein rain gauge station which is the area of rice grown most.

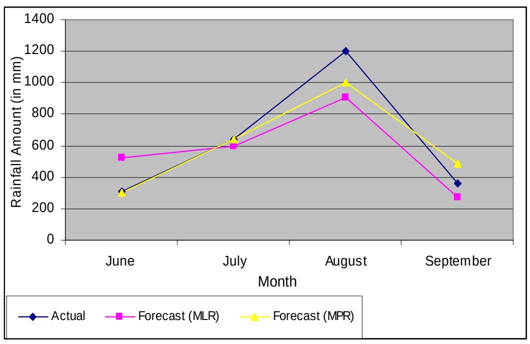

Fig. 2 Comparison between actual and prediction Pathein monthly rainfall amount with test data of 2006

For Pathein station, the prediction accuracy is 99% in training period and 86% with test data of 2006 with the use of MPR.

Fig. 3 shows the comparison between forecast and actual rain amount with test data of 2006 for Magwe rain gauge station where groundnut and beans grown most.

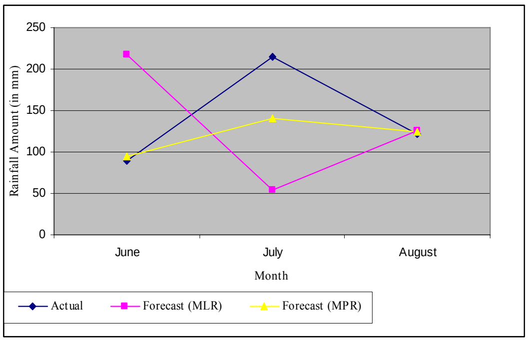

Fig. 3 Comparison between actual and prediction Magwae monthly rainfall amount with test data of 2006

From these experimental results, it is observed that using MPR method achieves closer agreement between actual and estimated rainfall than using MLR in the prediction of rainfall over Magwae rain gauge stations.

VI CONCLUSION

In Myanmar, rainfall is a critical factor in farm management and water resource management. Second-order MPR rainfall prediction model based on 15 predictors using stepwise regression approach has been developed. As a result of several experiments, the predicted rainfall amount is close to the actual value. The model results are intended to use in the areas such as crop planting and yield, water management and reservoir control.

In this implementation of MPR model, all possible subsets of predictors have not been examined. Thus, it can't grantee that the selected model built based on stepwise selection technique can offer the smallest prediction error.

REFERENCES

[1] N. Sen, " New forecast models for Indian south-west Monsoon season Rainfall", *in Current Science*, vol. 84, No. 10, May 2003, pp.1290-1291.
[2] S. Nkrintra, et al., "Seasonal Forecasting of Thailand Summer Monsoon Rainfall", *in International Journal of Climatology*, Vol. 25, Issue 5, American Meteorological Society, 2005, pp. 649-664.
[3] T. Sohn, J. H. Lee, S. H. Lee, C. S. Ryu, "Statistical Prediction of Heavy Rain in South Korea", *in Advances in Atmospheric Sciences*, Vol. 22, No. 5, 2005, pp.703-710.
[4] M. T. Mebrhatu, M. Tsubo, S. Walker, "A Statistical Model for Seasonal Rainfall Forecasting over the Highlands of Eritrea", *in International Crop Science Organization*.
[5] H. Hasani, M. Zokaei, A. Amidi,"A New Approach to Polynomial Regression And Its Application to Physical Growth of Human Height".
[6] D.A. Vaccari, J. Levri, "Multivariable Empirical Modeling of ALS Systems Using Polynomials," *in Life Support and Biosphere Science*, 1999, vol. 6 pp. 265-271.
[7] http://www.cru.uea.ac.uk/.
[8] http://www.cpc.noaa.gov/.
[9] http://www.jamstec.go.jp/