International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:8, No:11, 2014

Forecasting Rainfall in Thailand: A Case Study of Nakhon Ratchasima Province

N. Sopipan

Abstract**—**In this paper, we study the rainfall using a time series for weather stations in Nakhon Ratchasima province in Thailand by various statistical methods to enable us to analyse the behaviour of rainfall in the study areas. Time-series analysis is an important tool in modelling and forecasting rainfall. The ARIMA and Holt-Winter models were built on the basis of exponential smoothing. All the models proved to be adequate. Therefore it is possible to give information that can help decision makers establish strategies for the proper planning of agriculture, drainage systems and other water resource applications in Nakhon Ratchasima province. We obtained the best performance from forecasting with the ARIMA Model$(1,0,1)(1,0,1)_{12}$.

Keywords**—**ARIMA Models, Exponential Smoothing, Holt-Winter model.

I. Introduction

RAINFALL estimates are an important component of water resources applications, for example, in designing drainage systems and irrigation, an accurate estimate of rainfall is needed. There are also concerns with producing valid estimates using appropriate methods. In order to develop a comprehensive solution to the forecasting problem, including addressing the issue of uncertainty in predictions, a statistical model must be employed.

Many methods and approaches for formulating forecasting models are available in the literature. This paper deals exclusively with a time series forecasting model, in particular, the Auto Regressive Integrated Moving Average (ARIMA). These models were described by Box and Jenkins [1].

The Box-Jenkins approach possesses many appealing features. It allows the manager who has only data on the quantity of rainfall for the previous years to forecast future the amount of future rainfall without having to search for other related time series data, for example, temperature. The Box-Jenkins approach also allows for the use of several time series, for example, temperature, to explain the behavior of another series, for example rainfall, if these other time series data are correlated with a variable of interest and if there appears to be some cause for this correlation.

The Box-Jenkins (ARIMA) model has been successfully applied in various water and environmental management applications.

Other successful automatic forecasting methods in practice are based on exponential smoothing models. There are a variety of such models, each having a property that provides forecasts which are weighted averages of past observations with recent observations given relatively more weight than older observations. The name "exponential smoothing" reflects the fact that the weights decrease exponentially as the observations get older (Hyndman [2]).

Winter generalized the Holt exponential method to handle seasonality in time series data by introducing a third factor into time series analysis which is seasonal smoothing. Seasonal smoothing is an estimated value of seasonal growth ratereflecting the seasonal pattern of time series data [3]. Since the time series data for rainfall show a certain seasonality, the Holt-Winters Exponential Smoothing model is suitable for further investigation.

We adapt two approaches in modeling and forecasting rainfall in this paper. The next section is the materials and methods based on the Box Jenkins model building technique and exponential smoothing methods. The empirical methodology and model estimation results are given in Section III. In Section IV, we give an evaluation model for forecasting rainfall. The conclusion is given in Section V.

II. Materials and Methods

In this paper, we use two methods to forecast rainfall which are: the auto regressive integrated moving average (ARIMA) and Holt-Winter's method which uses the exponential smoothing method.

Because the data for rainfall is seasonal, we will briefly describe these particular models.

A. The Auto Regressive Integrated Moving Average (ARIMA)

The main stages in setting up a forecasting ARIMA model includes model identification, estimation of the model parameters and diagnostic checking for the appropriateness of the identified model for modeling and forecasting.

Model Identification is the first step in this process. The data was examined to check for the most appropriate class of ARIMA processes through selecting the order of the consecutive and seasonal differencing required to make the series stationary, as well as specifying the order of the regular and seasonal auto regressive and moving average polynomials necessary to adequately represent the time series model.

The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are the most important elements of time series analysis and forecasting. The ACF measures the amount of linear dependence between observations in a time series that are separated by a lag k. The PACF plot helps to determine how many auto regressive terms

N. Sopipan is lecturer at Nakhon Ratchasima Rajabhat University, Nakhon Ratchasima, 30000, Thailand (e-mail: nopsopipan@gmail.com).

are necessary to reveal one or more of the following characteristics: time lags where high correlations appear, seasonality of the series, trends either in the mean level or in the variance of the series.

The general model introduced by Box and Jenkins includes autoregressive and moving average parameters as well as differencing in the formulation of the model. The three types of parameters in the model are: the autoregressive parameters (p), the number of differencing passes (d) and moving average parameters (q). The Box-Jenkins model is summarized as ARIMA (p, d, q). In addition to the non-seasonal ARIMA (p, d, q) model, introduced above, we could identify seasonal ARIMA (P, D, Q) parameters for our data. These parameters are: seasonal autoregressive (P), seasonal differencing (D) and seasonal moving average (Q).

The seasonal ARIMA denoted by ARIMA(p,d,q)(P,D,Q) is given as (Abraham et.al. [4]):

$$\begin{aligned}&(1-\phi_1 B-\phi_2 B-...-\phi_p B)(1-\Phi_1 B^{12}-\Phi_2 B^{12}-...-\Phi_P B^{12})\\&(1-B)^d(1-B^{12})^D Y_t=\beta_0+(1-\theta_1 B-\theta_2 B-...-\theta_q B)\\&(1-\Theta_1 B^{12}-\Theta_2 B^{12}-...-\Theta_Q B^{12})e_t\end{aligned} \tag{1}$$

where

$(1-\phi_1 B-\phi_2 B-...-\phi_p B)$ is the autoregressive part of order p (AR(p)) ,

$(1-\Phi_1 B^{12}-\Phi_2 B^{12}-...-\Phi_P B^{12})$ is the seasonal autoregressive part of order P (ARs(P)) ,

$(1-\theta_1 B-\theta_2 B-...-\theta_q B)$ is the moving average part of order q (MA(q)) ,

$(1-\Theta_1 B^{12}-\Theta_2 B^{12}-...-\Theta_Q B^{12})$ is the seasonal moving average part of order Q (MAs (Q)) ,

$(1-B)^d$ is differencing of order d (I (d)) ,

$(1-B^{12})^D$ is seasonal differencing of order D (Is (D)) ,

β_0 is constant,

12 is the period of the seasonal pattern appearing with each month of the year.

The idea behind the seasonal ARIMA is to look at what are the best explanatory variables to model a seasonal pattern. Details of such a model can be found in Box and Jenkins [1].

This model can be multiplied out and used for forecasting after the model parameters are estimated, as we discuss below.

After choosing the most appropriate model (step 1 above) the model parameters are estimated (step 2) by using the least square method. In this step, values of the parameters are chosen to make the Sum of the Squared Residuals (SSR) between the real data and the estimated values as small as possible. In general, the nonlinear estimation method is used to estimate the above identified parameters to maximize the likelihood (probability) of the observed series given the parameter values. The methodology uses the following criteria in the estimation of the parameters.

In order to check step three, the residuals from the fitted model should be examined against adequacy. This is usually done by correlation analysis through the residual ACF plots and the goodness-of-fit test by means of Chi-square statistics χ^2. If the residuals correlate, then the model should be refined as in step one above. Otherwise, the autocorrelations are white noise and the model is adequate to represent our time series.

After the application of the previous procedure for a given time series, a calibrated model can be developed which encloses the basic statistical properties of the time series into its parameters (step four). For example, the developed model, as shown in (1) can be multiplied out and the general model is written in terms of Y_t.

B. The Holt-Winters Method

The method used for the load forecast is based on time series and only takes into consideration the history of the consumption in order to establish a pattern in the past that might be useful and similar with the present load curves. This technique uses exponentially decreasing weights as the observations get older. Recent observations are given relatively more weight in forecasting than the older observations. Exponential Smoothing is used to generate the smoothed values in order to obtain estimates.

Some time series data exhibit cyclical or seasonal patterns that cannot be effectively modeled using the polynomial model. Several approaches are available for the analysis of this data. The methodology of the Day-Ahead forecast presented in this paper was introduced by Holt and Winters and is generally known as the Winters' method; in this case, a seasonal adjustment is made to the linear trend model.

Two types of adjustments are used, namely the multiplicative and the additive model. This paper uses a multiplicative model and the estimated values are calculated as in the following equation:

$$Y_t=(\beta_0+\beta_1 t)S_t+\varepsilon_t \tag{2}$$

The forecast ℓ step ahead follows this equation:

$$\hat{Y}_t(\ell)=[\hat{\beta}_0(n)+\hat{\beta}_1(n)\ell]\hat{S}_{t+\ell}(t+\ell-s) \tag{3}$$

$$\hat{\beta}_0(n)=\alpha_1\frac{Y_t}{\hat{S}_t(t-s)}+(1-\alpha_1)[\hat{\beta}_0(t-1)+\hat{\beta}_1(t-1)] \tag{4}$$

$$\hat{\beta}_1(t)=\alpha_2[\hat{\beta}_0(t)-\hat{\beta}_0(t-1)]+(1-\alpha_2)\hat{\beta}_1(t-1) \tag{5}$$

$$\hat{S}_t(t)=\alpha_3\frac{Y_t}{\hat{\beta}_0(t)}+(1-\alpha_3)\hat{S}_t(t-s) \tag{6}$$

where $\alpha_1,\alpha_2,\alpha_3$ are the discount factors (constants) which must be chosen for the smallest sum of the squared forecast errors.

International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:8, No:11, 2014

III. Empirical Methodology and Model Estimation Results

A. Data

The historical rainfall dataset is for Nakhon Rachasima province. The data was collected from 15 stations by the Hydrology and Water Management Center for Lower Northeastern Region $\{I(t); t = 1, 2, \ldots, 96\}$ and it consists of 8 years of monthly data from April 2005 to March 2013. As shown in Fig. 1, the rainfall time series is subject to considerable short-term variations despite increasing long-term trends. Seasonality is an important property exhibiting certain cyclical or periodic behaviors in time series data. Generally, a period of 1 year (12 months) is considered for monthly data such as rainfall.

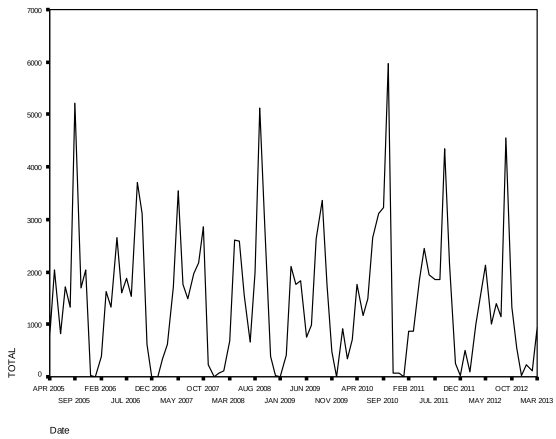

Fig 1 The historical rainfall dataset for Nakhon Rachasima province from April 2005 to March 2013

The Autocorrelation Function Plot of the monthly increment of rainfall in Fig. 2 shows that rainfall increment has relatively higher autocorrelations at lag levels of multiple 12, which are lags 12, 24, 36, and 48. This is evidence for the certain seasonality of rainfall increment dataset with the period of 12 months [5].

In order to improve predictability, seasonal models such as ARIMA(p,d,q)(P,D,Q) and Holt-Winter's model will be fitted to each dataset, then the fitted models will be used to make an out-of-sample forecast. The whole dataset from 2005 to 2013 is used as an example to illustrate the model fitting process.

B. ARIMA(p,d,q)(P,D,Q)

Since the data is for the monthly rainfall, Fig. 1 shows that there is a seasonal cycle of the series and the series is stationary. The ACF and PACF of the original data, as shown in Fig. 2, show that the rainfall data is stationary.

As shown in Fig. 3, the ACF and PACF for the seasonalized rainfall data are almost stable which supports the assumption that the series is stationary in both the mean and the variance. Therefore, an ARIMA (p, 0, q) $(P, 0, Q)_{12}$ model could be identified.

After the ARIMA model was identified above, the p, q, P and Q parameters need to be identified for our model. In Fig. 3, we have one autoregressive (p) and one moving average (q) parameter and the ACF has exponential decay starting at lag 12. Similarly, the PACF has an exponential decay starting at lag 12. Since we have identified for the rainfall data our tentative model will be ARIMA (1, 0, 1)$(1, 0, 1)_{12}$.

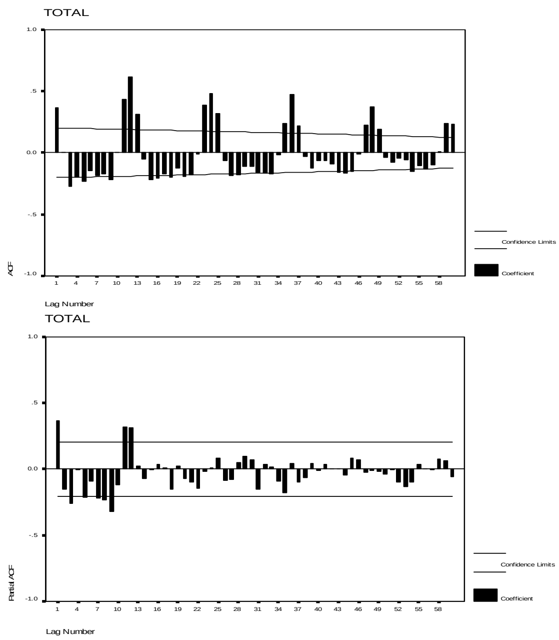

Fig. 2 The ACF and PACF Plot of the monthly increment of rainfall from April 2005 to March 2013

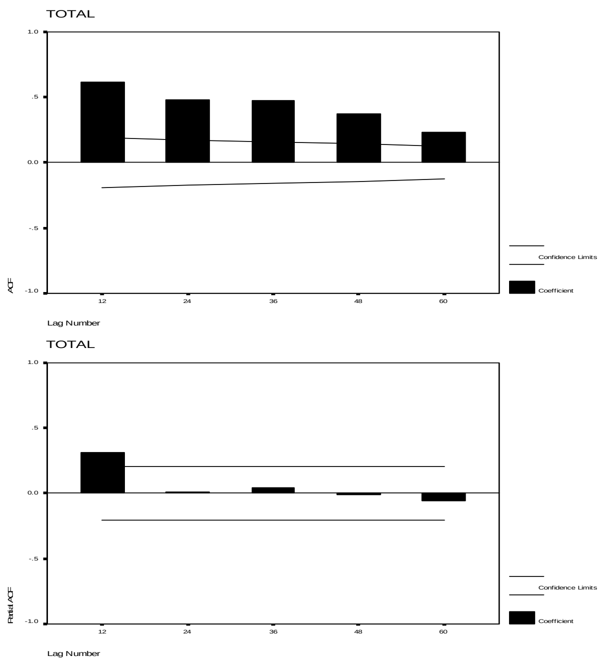

Fig. 3 The ACF and PACF Plot for the seasonalized rainfall from April 2005 to March 2013

International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:8, No:11, 2014

In order to make sure that this model is representative of our data and could be used to forecast the upcoming rainfall data, we need to test the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) for the residual errors resulting from fitting such model to our data. Fig. 4 shows the residual errors for the ARIMA fitted model. It is clear that we have no observable pattern in the residual errors which shows that the model could be used to represent our data.

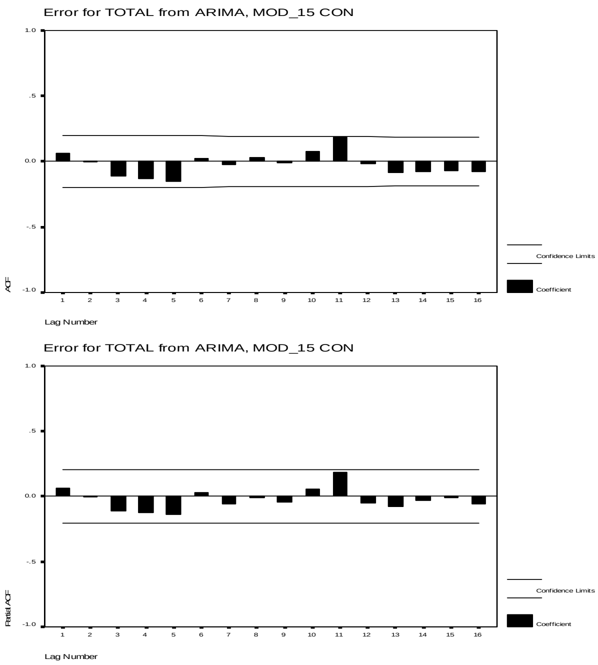

Fig. 4 The ACF and PACF for the residual errors resulting from fitting such a model to rainfall from ARIMA (1, 0, 1) (1, 0, 1)$_{12}$

After we fitted an ARIMA (1, 0, 1) (1, 0, 1)$_{12}$ for our data we need to estimate the parameters values for our model, as shown in (1). As a rule of thumb, in ARIMA modeling we need to minimize the sum squared of the residuals which need to be minimized between the forecasted and existing values:

$$(1-\phi_1 B)(1-\Phi_1 B^{12})Y_t = \beta_0 + (1-\theta_1 B)(1-\Theta_1 B^{12})e_t$$

The sum squared of the residuals for the model was 67944822.2 and the values of the parameters are shown in Table I as follows:

TABLE I
PARAMETERS ESTIMATOR IN ARIMA (1,0,1)(1,0,1)$_{12}$

Parameter	Parameter value	T-ratio	P-value
ϕ_1	-0.87577	-4.73	0.0000083
θ_1	-0.79573	-3.39	0.0010452
Φ_1	0.97970	34.87	0.0000000
Θ_1	0.72024	4.02	0.0011866
β_0	1468.19460	4.87	0.0000047

After the model parameters have been estimated, we are able to forecast the upcoming rainfall data.

C. Holt-Winters Exponential Smoothing

Winter generalized the Holt Exponential Smoothing method to handle seasonality in time series data by introducing a third factor seasonal smoothing, into the time series analysis. Seasonal smoothing is an estimated value of the seasonal growth rate reflecting the seasonal pattern of the time series data. Since rainfall time series data show seasonality, Holt-Winters Exponential Smoothing model is fitted to investigate whether the amount of error would be improved in comparison with the ARIMA model.

Least Square Estimate (LSE) is used to decide the appropriate values for the weighting factors α_1, α_2 and α_3 such that the MSE of the forecasted values is minimized. The optimal values of these parameters are summarized in Table II.

TABLE II
PARAMETERS ESTIMATOR IN HOLT-WINTER EXPONENTIAL SMOOTHING MODEL

Parameter	Parameter value
Seasonal indices:	
1	400
2	500
3	600
4	700
5	800
6	900
7	1000
8	1100
9	1200
10	100
11	200
12	300
α_1	0.7
α_2	0.04
α_3	0.46
$\hat{\beta}_0(0)$	1485.445583
$\hat{\beta}_1(0)$	-2.57708

After the model parameters have been estimated, we are able to forecast the upcoming rainfall data.

IV. EVALUATION MODEL FOR FORECASTING RAINFALL

In order to compare the predictability of the ARIMA model and the Holt-Winter model between April 2005 and March 2013,

Three general error measures, the Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), and Mean Absolute Error (MAE) are used to measure the predictability of the ARIMA model and the Holt-Winter model.

The less the error measures, the more accurate the prediction is ARIMA(1,0,1)(1,0,1)$_{12}$. Table III summarizes the error measures for these two models. Table III shows that ARIMA(1,0,1)(1,0,1)$_{12}$ gives the best performance for forecasting rainfall. Fig. 5 shows the time series plot of actual rainfall, according to ARIMA's forecast of rainfall and Holt-Winters forecast of rainfall.

TABLE III
MEASURE OF ERROR IN FORECASTING WITH THE ARIMA AND HOLT-WINTER'S MODELS

Error measures	ARIMA(1,0,1)(1,0,1)$_{12}$	Holt-Winter
MAPE	3.23622%	11.95938%
MSE	775427.300	1694700.000
MAE	616.918	943.194

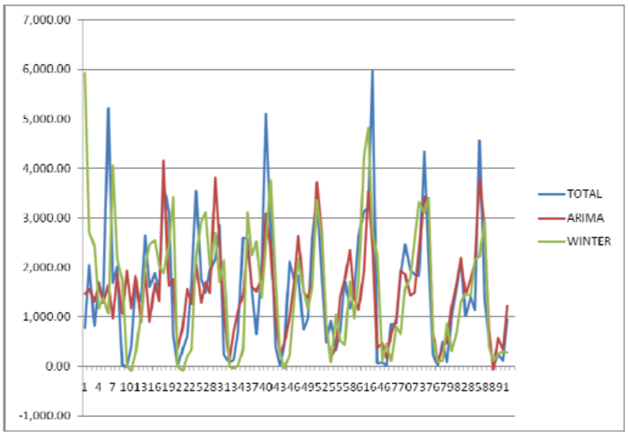

Fig. 5 Comparison of forecast of rainfall with actual rainfall values

V. CONCLUSION

The time-series analysis is an important tool in modelling and forecasting rainfall. Two time series analysis techniques were applied in this study, the Holt-Winter's exponential smoothing and the Box-Jenkins' modeling method (ARIMA). Both models were found to be adequate for forecasting. Diagnostic checking confirms the adequacy of the models.

The ARIMA (1, 0, 1) (1, 0, 1)$_{12}$ model give us information that can help decision-makers establish strategies, priorities and the proper use of water resources in Nakhon Ratchasima Province. This information is not appropriate for the prediction of the exact amount of monthly rainfall data. Therefore, individual monthly rainfall data should not be used in decision-making by depending on our model. However, an intervention time series analysis can be tested to see if we can improve our model performance in forecasting the peak values of rainfall data.

ACKNOWLEDGMENTS

This research is (partially) supported by the Research and Development Institute, Nahkon Ratchasima Rajabhat University, Thailand.

REFERENCES

[1] Box, G.E.P. and G.M. Jenkins, 1976. Time Series Analysis: Forecasting and Control. Revised Edn., Hoden-Day, San Francisco.
[2] Hyndman, R.J. and Khandakar, Y. 2008. Automatic Time Series Forecasting: The forecast Package for R. Journal of statistical software. http://www.jstatsoft.org/.
[3] Winters, P., 1960. Forecasting sales by exponentially weighted moving averages. Management Science 6, 324-342.
[4] Abraham, Bovas & Ledolter , Johannes. (1983). Statistical Methods for Forecasting. New York: John Willey & Sons.
[5] Tsay, R.S. (2005). Analysis of financial time series, Second Edition, John Wiley & Sons, Inc., Hoboken, NJ.