บทความวิจัย

MONTHLY RAINFALL AMOUNT FORECASTING IN KANCHANADIT DISTRICT, SURAT THANI PROVINCE BY STATISTICAL FORECASTING TECHNIQUES

Orawan Suebsen* and Supachai Damkam

Mathematics Faculty of Science and Technology, Surat Thani Rajabhat University, Surat Thani, 84100

*E-mail: orawan.sue@sru.ac.th

Received: 30-07-2022

Revised: 17-11-2022

Accepted: 25-11-2022

ABSTRACT

The research aims to study the appropriate forecasting model for the rainfall in Kanchanadit, Surat Thani Province. The data used in this research was the average monthly rainfall from Surat Thani Meteorological Station from January 2013 to February 2020 (86 values). The time series data were divided into two categories. The first 74 values from January 2013 until February 2019 were used for the modeling by the Decomposition Method, Triple Exponential Smoothing: Winter's Method, and Box – Jenkins Method. On the other hand, the last 12 values from March 2019 to February 2020 were used check the forecasting models' accuracy via the determination of the mean absolute percentage error. The research findings were as follows: 1) The Box–Jenkins method was the most suitable method that has been used for this time series. 2) The forecasting model for the rainfall in Kanchanadit, Surat Thani Province, could be written as $\hat{Y}_t = Y_{t-12} - 0.915\ (Y_{t-12} - Y_{t-24}) - 0.655\ (Y_{t-24} - Y_{t-36}) - 0.444\ (Y_{t-36} - Y_{t-48})$

Keywords: Decomposition Method, Triple Exponential Smoothing: Winter's Method, Box – Jenkins Method

Introduction

Surat Thani Province is an agricultural province that relies on natural rainwater for agriculture. Most of the population engage in agriculture, such as farming, gardening, and uses land to cultivate about 45% of the total area (Surat Thani Land Development Station, 2019) which require rainwater to reduce the risk of damage crops; therefore, it is necessary to organize a good planting system. One must know the variability and distribution characteristics of rain. Rainwater is an important factor for agriculture. The most important constraint is having sufficient water for planting and plant growth.(Office of the Meteorological Department, 2018) However, whether it rains or not is an uncontrollable factor. Rainfall is also unstable and varies from area to area. In addition, rainwater is important for agriculture. Therefore, rainfall forecasting is essential for agriculture. (Suwanwong, 2013) Kanchanadit District, It is a good plan for Surat Thani province, It is worth paying attention to the study. The amount of rainfall that has fallen is uncertain. Rainfall can be predicted by analyzing trends in rainfall changes. For plant cultivation, planning and finding solutions to problems that will occur with the rain that tends to change for the reasons mentioned above, the researcher was interested in studying the rain forecasting in Kanchanadit District, Surat Thani Province, which is useful in agriculture. Soil and water conservation in engineering, such as irrigation projects and reservoirs Projects to build dams, etc. can also be applied to solve flooding problems and provide guidelines for the further development of water resources.

Purpose of the study

1. Create a monthly rainfall forecast model for Kanchanadit District Surat Thani Province
2. Compare the three methods of forecasting rainfall, namely the decomposition method, the exponential smoothing method with Winter's method, and the Box-Jenkins method, which uses the mean absolute percentage error (MAPE) criterion.

Methods

1. Collect a monthly rainfall dataset. Kanchanadit District, Surat Thani Province collecting 86 values from January 2013 to February 2020. (Office of the Meteorological Department, 2018). The data source was the weather forecast bureau, meteorological department, Ministry of Information and Communication Technology, Surat Thani Province.
2. Take the monthly time series data from January 2013 to February 2020 and divide the data into two sets. (Ruangchaisiwawet, 2011)

The first dataset contains the monthly rainfall recorded from January 2013 to February 2013. Forecast models are based on this dataset.

The second dataset contains the monthly precipitation collected during March 2019 and February 2020.

3. Generate prediction models with three predictions from the 1 dataset.
4. Use the forecast values obtained from item 2 to calculate the accuracy of the forecast by comparing then with the actual data. Between March 2019 and February 2020. The accuracy was measured by the mean percentage error of the absolute value.
5. Take the mean of the absolute percent error from the time series into account. Compare the forecasting approaches for time series.

Results and Discussions

This dataset consisted of trends (T) and seasonal variations (S), as it can be seen in Figure 1, but these methods produce ambiguous results, for accuracy and comprehension. They are consistent. Therefore, the trend component test and the seasonal component test were used as follows:

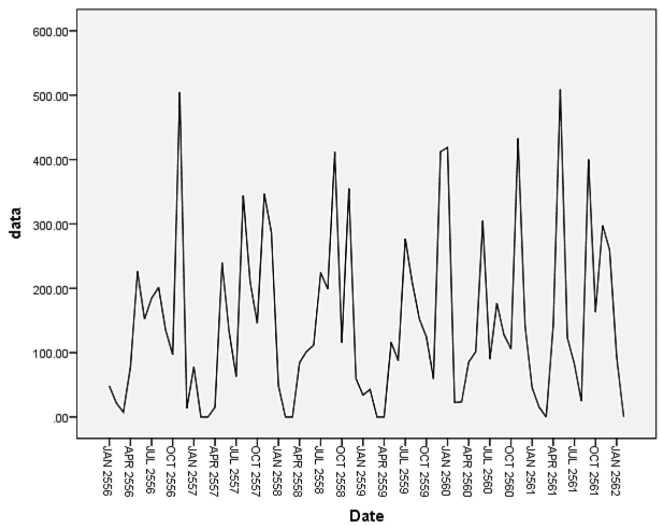

Figure 1 The movement characteristics of time series data

1. Trend component testing

As shown in Figure 2, the trend component was created using Run Test (Suwanwong, 2013). By the Run Test method, Hypotheses are as follows, H_0 : time series has no trend component and H_1 : time series has trended component. At the significance level of .05, the *p*-value was found to be .010, which was less than the significance level, and was therefore H_0 was rejected. In conclusion, such time series contains trend component.

Runs Test

	data
Test Value[a]	113.40
Cases < Test Value	37
Cases >= Test Value	37
Total Cases	74
Number of Runs	27
Z	-2.575
Asymp. Sig. (2-tailed)	.010

a. Median

Figure 2 Component testing of rainfall trends

2. Seasonal Component Test

As shown in Figure 3, the seasonal components were calculated using Kruskal-Wallis (Ketiam, 2005) By the Kruskal-Wallis method, the hypotheses are as follows: H_0 : The time series has no seasonal variance components; H_1 : The time series has seasonal variance components. At the significance level of .05, the *p*-value was found to be .010, which was less than the significance level and rejected H_0. It was concluded that the time-series data contained seasonal variations.

Test Statistics[a,b]

	data
Chi-Square	42.987
df	11
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable:

MONTH, period 12

Figure 3 Seasonal component test of rainfall

3. Create forecast models

3.1 That the time series contains trend components (Ketiam, 2005). So we can generate trend equations by finding trend line equations. By creating a trend line using the least-squares method, the equation of a straight line trend was formed as follows:

$$Y_t = (b_0 + b_1 t)\hat{S}_t \qquad \text{.....................................(1)}$$

Where Y_t is an observation value at time t
b_0 is the mean of the data
b_1 is the slope of the data
$\hat{S}_t$ is the estimated seasonal variation at time t.

Short term forecast

The equation for forecasting was $Y_t = (18.885t)\hat{S}_t$

When $b_1 = 18.885$

Table 1 Coefficients of model Forecast calculation example

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		Std. Error
MONTH, period 12	18.885	3.868	.499	4.882	.000
(Constant)	27.708	28.100		.986	.327

March 2019 rainfall forecast

the equation $Y_t = (18.885t)\hat{S}_t$

$Y_{75} = (18.885(3))(0.032)$

So $Y_{75} = 1.813$

Therefore, the March 2020 rainfall forecast is 1.813 millimeters.

Table 2 shows actual and forecasted rainfall. By the decomposition method, the mean absolute percentage error (MAPE) was 1.789.

Table 2 shows the mean absolute percent error (MAPE) values of the time series of the forecasted values compared to the actual values by the decomposition method.

Month/year	Actual rainfall (mm)	Forecasted rainfall (mm)	The absolute value of the error relative to the actual value.
Mar 19	0.0	1.813	-
Apr 19	53.6	32.482	0.394
May 19	242.1	131.628	0.456
Jun 19	139.8	112.290	0.197
Jul 19	107.8	143.431	0.331
Aug 19	139.0	198.972	0.431
Sep 19	298.7	268.715	0.100
Oct. 19	153.7	157.501	0.025
Nov 19	85.2	453.901	4.327
Dec. 19	19.5	301.405	14.457
Jan 20	0.0	13.597	-
Feb 20	15.5	3.890	0.749
	$\sum$		21.467
	MAPE		**1.789**

3.2 Exponential smoothing method with Winter's method

A forecasting equation of rainfall in Kanchanadit District, Surat Thani Province was created using Winter's method of exponential smoothing.

Below is a Multiplicative Seasonality Model

$$\hat{Y}_{t+m} = (a_t + b_t(m))\hat{S}_{t-p+m} \qquad \text{.................................(2)}$$

$$a_t = \frac{\alpha Y_t}{\hat{S}_{t-p}} + (1-\alpha)\left[a_{t-1} + b_{t-1}\right]$$

$$b_t = \gamma(a_t - a_{t-1})\,(1-\gamma)\,b_{t-1}$$

$$\hat{S}_t = \frac{\delta Y_t}{a_t} + (1-\delta)\hat{S}_{t-p}$$

Where $\hat{Y}_{t+m}$ is the observation at time t

a_t is the level of the data or smoothing part

b_t is the trend

$\hat{S}_t$ is the seasonal part

m is the number of periods to be forecasted forward

p is the number of seasons p=12 when the data is monthly.

α is the smoothing constant between the data and the forecast value, and **$0 \le \alpha \le 1$**

γ is the smoothing constant between the actual trend and the trend estimate, and **$0 \le \gamma \le 1$**

δ is the smoothing constant between the actual season value and seasonal estimate, and **$0 \le \delta \le 1$**

The Winter method is analyzed using the multiplication pattern α, δ and γ between 0.001-1.0 by trying to change the α, δ and γ values in increments of 0.01, and then selected the α, δ and γ that give the lowest SSE values the, and values of Win's method. The appropriate parameters are α = 0.001, γ = 0.001 and δ = 0.001.

There is a forecasted value m, the forward time unit, the forecast at time t, and the monthly data are:

$$\hat{Y}_{t+m} = (a_t + b_t(m))\hat{S}_{t-p+m} \qquad \text{.................................(2)}$$

Short term forecast

When, a_t = 07.708 and b_t = 18.885 are obtained by the least-squares trend line generation method of the discrete forecasting method. Fromm's data analysis, The winter season index ($\hat{S}_t$) is obtained when (t = 1, 2, 3, ..., p) which is equal to the seasonality index obtained by the modular forecasting method. As show in Table 3.

Table 3 winter season index of rainfall.

t	1	2	3	4	5	6
$\hat{S}_t$	0.622774	0.335709	0.027191	0.432075	1.488382	1.016865
t	7	8	9	10	11	12
$\hat{S}_t$	1.055471	1.291599	1.718774	0.901464	1.958707	1.150991

Forecast calculation example

March 2019 rainfall forecast (data 75)

Into the equation $\hat{Y}_{t+m} = (a_t + b_t(m))\hat{S}_{t-p+m}$

Represent $\hat{Y}_{75} = (a_{74} + b_{74}(3))\hat{S}_{63}$

$\hat{Y}_{75} = 1.530$

Therefore, the rainfall forecast for March 2020 is 1.530 mm. Rainfall Forecast for April 2019 (Data 76)

Table 4 shows the actual rainfall and the actual rainfall from the forecast. The mean absolute percentage error (MAPE) in winter was 1.497.

Table 4 shows the mean absolute percent error (MAPE) of the time series for the forecasted values compared to the actual values with the winter method.

Month/year	Actual rainfall (mm)	Forecasted rainfall (mm)	The absolute value of the error relative to the actual value.
Mar 19	0.0	1.530	-
Apr 19	53.6	32.633	0.168
May 19	242.1	140.504	0.249
Jun 19	139.8	115.236	0.026
⋮	⋮	⋮	⋮
Dec. 19	19.5	260.840	12.376
Jan 20	0.0	11.765	-
Feb 20	15.5	12.691	0.419
	$\sum$		17.966
	MAPE		**1.497**

3.3 Box - Jenkins Time Series Analysis Method

3.3.1 Determine the time series model

From Figure 4, it was found that when r_k is intermittently high, there is an undulating movement that is repeated for every 12 K values. This could lead to the conclusion that the time series data influences the seasonal variations involved. Therefore, such time series must be transformed so that they are free from the influence of seasonal variation using the seasonal difference method. By assigning D = 1, or the number of times the seasonal difference is equal to 1, the result is shown in Figure 4 as follows:

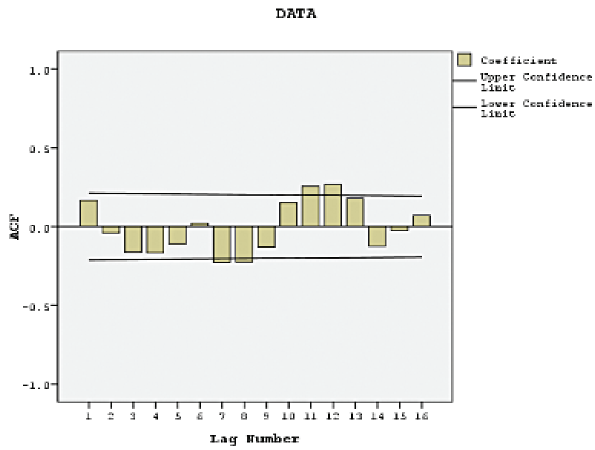

Figure 4 The determination and correlogram of r_k

From Figure 5, it was found that when the aforementioned time series was modified by the seasonal difference method, it could be seen that r_k It decreases rapidly when K increases. This shows that the time series data was free from the influence of trends. The above time series is now stationary. However, because the original time series includes the trend's influence, Therefore, the model was set up in the form of ARIMA (p, d, q) $(P, D, Q)_s$ Model determination was based on the results of ARIMA analysis. Which used the following model selection method:

1. Choose a model with the lowest RMSE value.
2. Choose a model with the least number of parameters.
3. A model that accepts the assumption that the parameter value is 0 is not chosen, which results in the following:

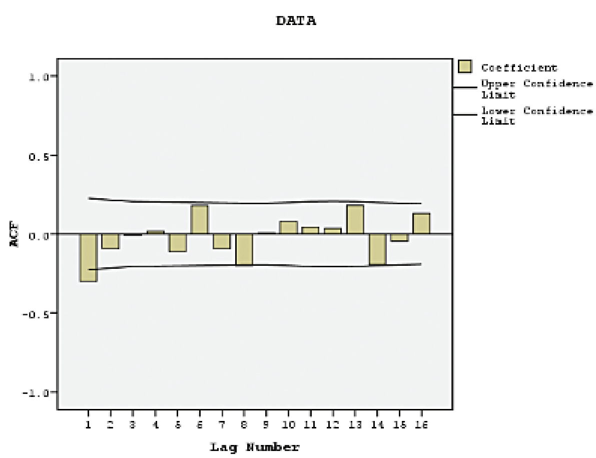

Figure 5 The values and correlogram of r_k after finding the difference to eliminate the influence of the seasons.

From Table 5, it was concluded that the model ARIMA(0,0,0)(3,1,0)$_{12}$ It is the best model with the lowest RMSE value, so the model is chosen ARIMA(0,0,0)(3,1,0)$_{12}$ which the general form of SARIMA (p,d,q) There is an equation

$$(1 - B)^d (1 - B^L)^D (1 - \phi_1 B - \phi_2 B^2 - .... - \phi_p B^p)(1 - \phi_L B^L - \phi_{2L} B^{2L} - .... - \phi_{PL} B^{pL}) y_t$$
$$= \theta_0 + (1 - \theta_L B - \theta_2 B^2 - .... - \theta_q B^q)(1 - \theta_L B^L - \theta_L B^{2L} - .... - \theta_{qL} B^{pL}) \varepsilon_t$$

Where B called backshift, where $B^p y_t = y_{t-p}$ และ $B^p \varepsilon_t = \varepsilon_{t-q}$

θ_0 is a constant

ϕ_i is the parameter of AR at i ; i =1,2,3, ..., p

θ_i is the parameter of MA at i ; i =1,2,3, ..., p

Therefore, the equation as

$$\hat{Y}_t = Y_{t-12} - \phi_1(Y_{t-12} - Y_{t-24}) - \phi_2(Y_{t-24} - Y_{t-36}) - \phi_3(Y_{t-36} - Y_{t-48}) \quad ....(4)$$

Table 5 defines the model in the form ARIMA(p,d,q)(P,D,Q)$_s$ Using the following model selection method:

1. Selecting a Model with the Least RMSE Value

ARIMA(p, d, q)(P, D, Q)$_s$	RMSE
ARIMA(0,0,1)(0,1,1)$_{12}$	124.976
ARIMA(0,0,0)(2,1,1)$_{12}$	124.976
ARIMA(0,0,0)(2,1,2)$_{12}$	125.441
ARIMA(0,0,0)(2,1,0)$_{12}$	129.983
ARIMA(0,0,0)(3,1,0)$_{12}$	**124.109**

3.4 Parameter estimation

Least Squares Parameter Estimation was performed using a statistical package. This will produce the results shown in Figure 6.

From Table 6 - 7, the estimated value was ϕ_1 = -0.915, ϕ_2 = -0.655 and ϕ_2 = -0.444

When substituting the values into the equation

$$\hat{Y}_t = Y_{t-12} - 0.915(Y_{t-12} - Y_{t-24}) - 0.655(Y_{t-24} - Y_{t-36}) - 0.444(Y_{t-36} - Y_{t-48}) \quad ....(5)$$

Table 6 Model Description

Model Description	
Model Type	ARIMA(0,0,0)(3,1,0)

Table 7 the approximation of parameter ϕ_1, ϕ_2 and ϕ_3 in ARIMA model parameters

ARIMA model parameters		Estimate	SE	t	Sig.
AR, Seasonal	Lag 1	-.915	.159	-5.756	.000
	Lag 2	-.655	.226	-2.903	.005
	Lag 3	-.444	.223	-1.990	.049
Seasonal Difference	1				

Table 8 shows actual and forecast rainfall. Perform by using the Box-Jenkins time series analysis method. The mean absolute percentage error (MAPE) was 1.157.

Table 8 shows the mean absolute percent error (MAPE) values of the time series for the forecasted values compared to the actual values by the Box Jenkins time-series analysis method.

Month/year	Actual rainfall (mm)	Forecasted rainfall (mm)	The absolute value of the error relative to the actual value.
Mar 19	0.0	19.35	-
Apr 19	53.6	85.11	0.588
May 19	242.1	152.92	0.368
Jun 19	139.8	170.94	0.223
⋮	⋮	⋮	⋮
Nov 19	85.2	321.02	2.768
Dec 19	19.5	185.81	8.529
Jan 20	0.0	136.80	-
Feb 20	15.5	41.05	1.648
	$\sum$		13.891
	MAPE		**1.157**

Table 9 The comparison of the mean absolute percent error (MAPE) values of three methods forecasting the rainfall in Kanchanadit, Surat Thani Province

Methods	MAPE
The decomposition method.	1.789
The winter method	1.497
The Box Jenkins time-series analysis method.	1.157

Table 9 shows the mean absolute percent error (MAPE) values of three forecasting models. The Box Jenkins time-series analysis method was suitable for this forecasting model for the rainfall in Kanchanadit, Surat Thani Province.

Conclusions

Using results from a study on the monthly rainfall forecast in Kanchanadit District, Surat Thani Province using statistical forecasting techniques, the following issues can be discussed: The Box-Jenkins Method: This was the most appropriate method for this time series. This is consistent with research by Ngamsuk (2012), a study of time series forecasting by comparing traditional methods and the Box-Jenkins method, a case study of the number of accidents in Thailand, and Kiratiwiboon's (2014) study on a rainfall forecasting model in Songkhla Province. The Box-Jenkins Method was the most appropriate method for this time series. The results of this research could be used to plan future crop production for farmers in Surat Thani province. The forecast models are suitable for time series with similar motion characteristics. However, rainfall may not depend solely on time factors. For future research, researchers should consider other factors in the predictive model study. Also, when more recent rainfall data was available. In applying the research results, the researcher should modify the model to suit their data. Furthermore, climate change should be closely monitored as a factor in data analysis. Therefore, the researcher must use caution when analyzing the data.

Acknowledgement

Thank you to Surat Thani Rajabhat University for funding research, and to the Surat Thani Meteorological Station for analyzing monthly rainfall data in Kanchanadit District, Surat Thani Province.

References

Ketiam, S. (2005). Forecasting techniques. (2nd ed), Songkhla: Thaksin University. (In Thai)

Kiratiwiboon, W. (2014). A model for forecasting rainfall in Songkhla province. **Thaksin University Journal**, **17**(1), 40-48. (In Thai)

Ngamsuk, W. (2011). **Time Series Forecasting by the Comparison of classical and Box-Jenkins Methods Case Study the Number of accidents in the country**. (Master's Thesis). Department of Mathematics Faculty of Science Burapha University, Chonburi. (In Thai)

Office of the Meteorological Department. (2018). **Surat Thani rainfall**. Retrieved from https://www.tmd.go.th [2019, 16 Dec].

Ruangchaisiwawet, P. (2011). Forecasting **monthly rainfall of Lampang Province by Statistical forecast techniques**. (Master's thesis). Department of Statistics, Faculty of Science, Kasetsart University, Bangkok. (In Thai)

Suwanwong. S. (2013). **Quantitative Forecasting Techniques**: Time Series. Bangkok: Mahidol University.

Surat Thani Land Development Station. (2019). **Information of Surat Thani Province**. Retrieved from http://r11.ldd.go.th/station/sni/ [2019, 2 Dec].