Journal of
Thai Interdisciplinary Research
Volume 12, Number 3, Pages 1 - 6

A comparative study of forecasting models for the number of Malaria patients in Phanom District, Surat Thani Province

Sujaree Damsri[1,*] and Ketsuda Maneewong[1]

[1]Department of Mathematics, Faculty of Science and Technology,
Suratthani Rajabhat University, Surat Thani 84100, Thailand

Abstract

The purpose of this study was to compare 5 forecasting models: Ratio-to-Trend, Holt-Winters Exponential Smoothing, Regression Dummy Variables, Theta, and Combined Forecasting. These methods were used for forecasting the number of Malaria patients in Phanom District, Surat Thani Province. The monthly time series data were collected from Surat Thani Provincial Health Office during January 2011 through December 2015 and divided into two groups. The first group of data covered the period of time from January 2011 to December 2014 and has been used for constructing the forecasting models. The secondary group of data covered the period of time from January 2015 to December 2015 and has been used for checking the accuracy of the forecasting models. The accuracy characteristics that we used were Mean Absolute Deviation (MAD) and Mean Square Error (MSE). The smallest values of MAD and MSE indicate the optimal forecasting model.

Based on the study, the optimal model was found to be Combined Forecasting and the equation of the forecasting model was $\hat{Y}_t = 0.907F_{1t} + 0.093F_{2t}$; $t = 1, 2, ..., 12$ where each F_{1t} and F_{2t} represents the single forecast at time *t*, from the Regression Dummy Variables model and Holt-Winters Exponential Smoothing model, respectively.

Keywords: forecasting model, Malaria, Surat Thani Province, combined forecasting model

Article history: Received 11 January 2017, Accepted 11 May 2017

1. Introduction

Forecasting is the process of making predictions of the future based on past and present data. The categories of forecasting are qualitative and quantitative. Qualitative forecasting is the approach based on judgments and opinions [1]. Quantitative forecasting can be applied when two conditions are satisfied: numerical information about the past was available and some aspects of the past patterns will continue into the future [2]. Many years ago, forecasting techniques have been continuously developed due to their usefulness in planning and decision making. Both the short-term and long-term forecasting techniques were a tool to obtain future information for planning. It is widely accepted that forecasting plays a key role both in the public and private sectors [3]. At the present time, forecasting techniques are used to predict the number of patients in epidemic diseases [4, 5, 6, 7, 8]. The suitable forecasting model depends on many factors such as the amount of historical data available, the experience of the forecaster, the degree of accuracy desirable, the time period of forecasting etc.

Malaria is caused by five species of parasites belonging to the genus *Plasmodium*: *P. falciparum*, *P. vivax*, *P. malariae*, *P. knowlesi* and *P. ovale*. Of these, *P. falciparum* and *P. vivax* are the most important [9]. Major *Plasmodium* that are spread from one person to another by the bite of female *Anopheles* mosquitoes in Thailand were *P. falciparum* and *P. vivax* [9, 10, 11]. From 1991 to 2002, the incidence of malaria in 14 southern provinces of Thailand has been present as important health problem especially in Ranong, Krabi, Surat Thani and Yala Province [10]. In 2015, Surat Thani Province is the region with the ninth Malaria patients of country especially in Phanom district has the highest number of malaria cases in Surat Thani province [11, 12].

In this paper, we constructed 5 forecasting models to determine the optimal models for the number of Malaria patients in Phanom District, Surat Thani Province. Forecasting values that we obtain will be used as the information in the preliminary plan for disease prevention and control.

2. Materials and methods

2.1 Data collection

The data under investigation were the secondary time series data from Phanom District. These time series data contain the number of Malaria patients and were collected by Surat Thani Provincial Health Office from January 2011 to December 2015. The data were monthly time series which amounts to 60 values that we divided into two groups. The data collected from January 2011 to December 2014

DOI 10.14456/jtir.2017.14

*Corresponding author; e-mail: melinda.sao@hotmail.com

amounts to 48 values, are considered to be the first group. The data collected from January 2015 to December 2015 amounts to 12 values, are considered to be the second group.

2.2 Exploring data pattern

The identification and understanding of historical patterns in the data is a major factor influencing the selection of a forecasting model [13, 14]. There are four general patterns of time series data: horizontal, trend, seasonal and cyclical [1, 13]. We investigated the data pattern by graphing the time series data and the autocorrelation for various lags of a time series which is called correlogram. When time series data fluctuate around a constant level or mean, a horizontal pattern exists. When time series data grow or decline over an extended period of time, a trend pattern exists. When time series data exhibit rises and/or falls that were not over an extended period, a cyclical pattern exists. When time series data display the pattern of change that repeats itself year after year, a seasonal pattern exists [13, 15, 16].

2.3 Constructing the forecasting model

We used the data from January 2011 to December 2014 for constructing 5 forecasting model. Considering the model to constructing the forecasting model is as follows: appropriate for the pattern of data and used to predict the number of patients in epidemic diseases [4, 17, 18]. We selected 5 forecasting models: Ratio-to-Trend, Holt-Winters Exponential Smoothing, Regression Dummy Variables, Theta, and Combined Forecasting by using Minitab software. The equations of forecasting models are as follows:

(1) Ratio-to-Trend [4, 15]

The forecasting equation used in Multiplicative Ratio-to-Trend is as follows:

$$\hat{Y}_t = b_0 b_1^{\,t} \hat{S}_i \,; i = 1, 2, ..., 12 \tag{1}$$

(2) Holt-Winters Exponential Smoothing [13]

This method has three parameters; α, β, γ where the values of these parameters are between 0 to 1. The four equations used in Multiplicative Holt-Winters' Exponential Smoothing are as follows:

The exponentially smoothed series:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{2}$$

The trend estimate:

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{3}$$

The seasonality estimate:

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s} \tag{4}$$

The forecasting model:

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p}\,. \tag{5}$$

Where:

L_t means the new smoothed value or the current level estimate,

α means the smoothing constant for the level,

Y_t means the actual value in period t,

β means the smoothing constant for trend estimate,

T_t means the trend estimate,

γ means the smoothing constant for seasonality estimate,

S_t means the seasonality estimate,

p means the period to be forecasted into the future,

s means the length of seasonality,

$\hat{Y}_{t+p}$ means the forecast for p periods into the future.

(3) Regression Dummy Variables Models

Regression Dummy Variables Models are regression models for time series with seasonal pattern that involve dummy variables. A seasonal model for monthly data with the time trend in multiplicative model is given below [13, 15]:

$$\ln \hat{Y}_t = \ln a_0 + \ln a_1 t + \ln a_2 t^2 + \ln b_1 S_1 + \ln b_2 S_2 + \cdots + \ln b_{i-1} S_{i-1}$$

(6) where:

$\hat{Y}_t$ means the forecast value for time period *t*,

$a_0, a_1, a_2, b_1 ..., b_{i-1}$ means the coefficients to be estimated,

S_i means the dummy variable which is 1 for month *i*; 0 otherwise.

(4) Theta

The Theta model is based on the concept of modifying the local curvatures of the time series. This change is obtained by a coefficient, called Theta-coefficient (θ). When the value of Theta-coefficient equals zero, the time series is equivalent to a linear regression line. When the Theta-coefficient takes a positive value, then the time series is dilated [19]. In this paper, we used Theta model for Theta-coefficient equals 0 or 2. The steps of Theta model are given below [20, 21]:

Step 1: Seasonality testing by *t*-test for autocorrelation function.

Step 2: Deseasonalisation by the classical decomposition method.

Step 3: Decomposition of the time series into two theta lines, $\theta = 0$ and $\theta = 2$. The equations are as follows: [9, 10].

The theta line for $\theta = 0$:

$$\hat{Y}_{n/h}(0) = \hat{a}_0 + \hat{b}_0(n + h - 1) \tag{7}$$

The theta line for $\theta = 2$:

$$\hat{Y}_{n/h}(2) = \alpha \sum_{i=0}^{n-1} (1-\alpha)^i Y_{n-i,2} + (1-\alpha)^n Y_{1,2} \tag{8}$$

Step 4: Combination two theta lines with equal weight.

The forecasting model:

$$\hat{X}_{n/h} = \frac{\hat{Y}_{n/h}(0) + \hat{Y}_{n/h}(2)}{2}$$

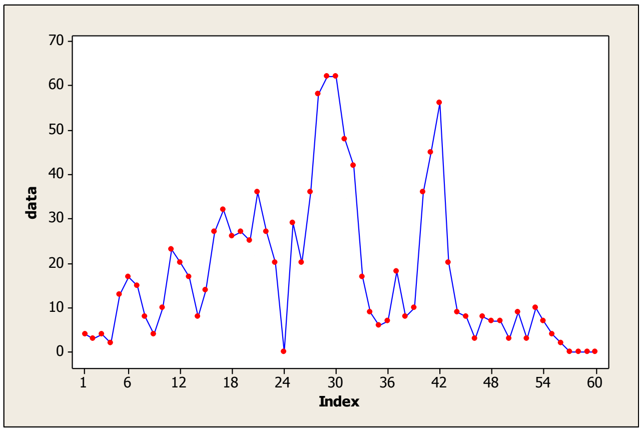

Figure 1 Plot of the number of Malaria patients

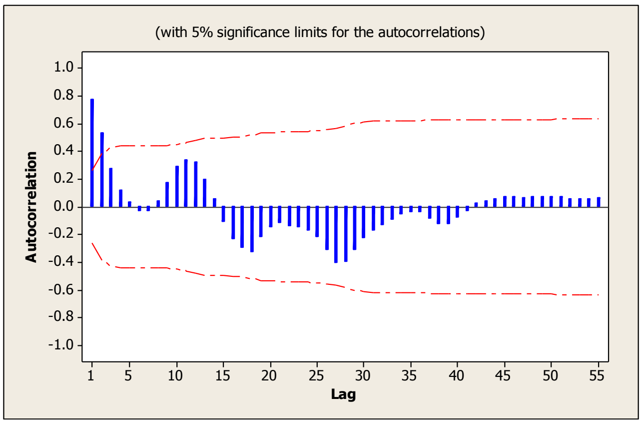

Figure 2 Autocorrelation function for data

$$\hat{X}_{n/h} = \alpha X_n + (1-\alpha)\hat{X}_n + \frac{1}{2}\hat{b}_0(h-1+\frac{1}{\alpha}) \qquad (9)$$

Where:

$$\hat{a}_0 = \bar{X} - \hat{b}_0\frac{(n-1)}{2},$$

$$\hat{b}_0 = \frac{6}{n^2-1}\left[\frac{2}{n}\sum_{t=1}^{n} tX_t - (n+1)\bar{X}\right]$$

$\hat{X}_{n/h}$ means the h-step forecast ahead,

n means the amount of observation,

α means the smoothing constant.

Step 5: Reseasonalisation.

(5) Combined Forecasting [16]

$$\hat{Y}_t = W_1F_{1t} + W_2F_{2t} + ... + W_iF_{it} \qquad (10)$$

Where:

W_i means the weight of the forecasting technique i,

F_{it} means the forecast value for forecasting technique i at time period t.

2.4 The comparative analysis for the optimal model for forecasting

MAD and MSE are the characteristics for forecasting accuracy that we used to compare the models and choose the optimal model for forecasting. The forecasting model which had the smallest values of MAD and MSE gives the optimal forecasting model. The equations of forecasting accuracy are given as follows: [13]

$$MAD = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right| \qquad (11)$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2 \qquad (12)$$

where:

Y_t means the actual value in time period t,

$\hat{Y}_t$ means the forecasted value for time period t.

3. Results and discussion

3.1 Data pattern

From Figure 1, the patterns of the number of Malaria patients showed that there exist both trend and seasonal variations because the series grew over an extended period of time and a repeated calendar pattern can be observed.

Figure 2 shows the same result of data pattern as Figure 1 because of autocorrelation coefficients (r_k) were typically large for the first several time lags and then gradually drop toward zero as the number of periods increased. This means that the number of Malaria patients seems to indicate a trend and seasonal variation.

3.2 Constructing the forecasting models

In this section, we present the equations of five forecasting models which are given in Table 1.

Table 1 Equations of each forecasting model

Forecasting models	Equations
Ratio-to-Trend	$\hat{Y}_t = 1.33(1.02)^t \hat{S}_i ; i = 1, 2, ..., 12$
Holt-Winters Exponential Smoothing	$\hat{Y}_t(p) = [17.52 + p(-1.08)]\hat{S}_i ; i = 1, 2, ..., 12, p = 1, 2, ...$ when $\alpha, \beta, \gamma = 0.1$ $\hat{S}_1 = 0.881, \hat{S}_2 = 0.485, \hat{S}_3 = 0.766, \hat{S}_4 = 1.412, \hat{S}_5 = 1.816, \hat{S}_6 = 1.924, \hat{S}_7 = 1.306,$ $\hat{S}_8 = 0.966, \hat{S}_9 = 0.747, \hat{S}_{10} = 0.575, \hat{S}_{11} = 0.698, \hat{S}_{12} = 0.42$
Regression Dummy Variables	$\ln \hat{Y}_t = 0.775 + 0.176t - 0.00329t^2 + 0.284S_1 - 0.319S_2 + 0.054S_3 + 0.450S_4 + 1.01S_5 + 1.05S_6 + 0.696S_7 + 0.278S_8 - 0.061S_9 - 0.303S_{10} - 0.014S_{11}$
Theta	$\hat{X}_{n/h} = \left[\alpha X_n + (1-\alpha)\hat{X}_n + 7.33165 \times 10^{-2}(h - 1 + \frac{1}{\alpha})\right] \times \hat{S}_i$ when h = 1,2,3,… and $\alpha = 0.3$ $\hat{S}_1 = 0.834, \hat{S}_2 = 0.423, \hat{S}_3 = 0.667, \hat{S}_4 = 1.67, \hat{S}_5 = 1.857, \hat{S}_6 = 1.873,$ $\hat{S}_7 = 1.374, \hat{S}_8 = 1.07, \hat{S}_9 = 0.56, \hat{S}_{10} = 0.725, \hat{S}_{11} = 0.676, \hat{S}_{12} = 0.272$
Combined Forecasting	$\hat{Y}_t = 0.907F_{1t} + 0.093F_{2t} ; t = 1, 2, ..., 12$

Table 2 Forecasting accuracy

Forecasting model	Forecasting accuracy	
	MAD	MSE
Ratio-to-Trend	8.49	106.56
Holt-Winters Exponential Smoothing	6.54	59.18
Regression Dummy Variables	1.41	4.21
Theta	13.30	237.80
Combined Forecasting	1.24	3.37

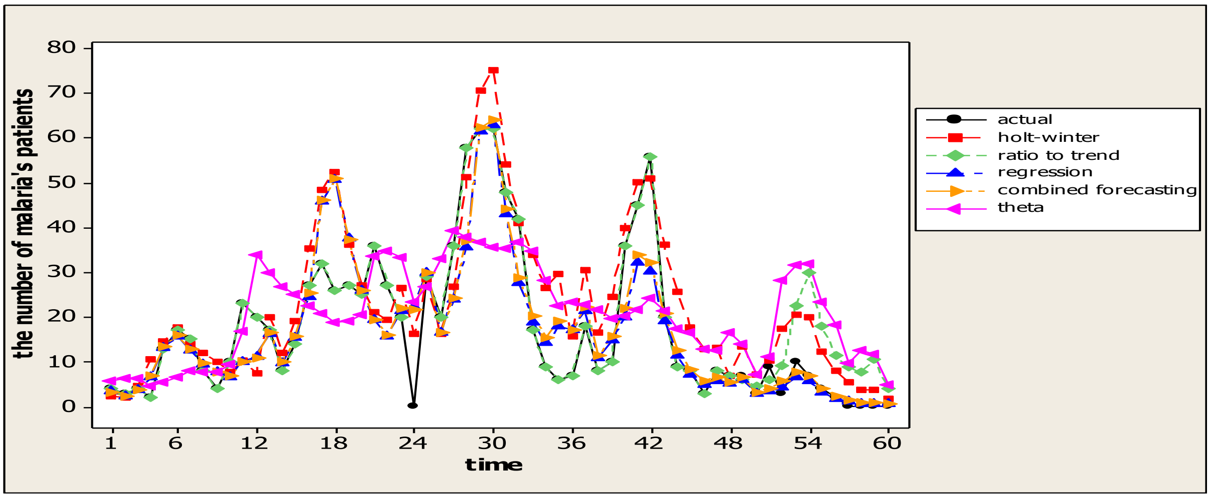

Figure 3 Comparison between actual values and forecasted values

3.3 The comparative analysis for the optimal for forecasting model

The values of forecasting accuracy are shown in Table 2 and the comparison between actual values and forecasted values are shown in Figure 3.

From Table 2, Combined Forecasting showed the lowest values of forecasting accuracy that means Combined Forecasting was the optimal model which corresponded with Manmin [3] who said that Combined Forecasting is the most accurate model for forecasting and corresponded with Taesombut [15] who said that Combined Forecasting is a method to make good prediction.

Forecasting by times series techniques is only study the difference of the data which depend on times. In fact, there are many factors that are important for the spread of disease which complicate and differ in each area. In case of some factors has changed, the number of patients may be mistaken from the forecasting. This means that the forecasting equation should usually examined by statistical analysis if forecasting values are not reasonable, we can construct new forecasting model or adjust the original forecasting model by collecting more data.

4. Conclusions

In the study of the number of Malaria patients in Phanom District, Surat Thani Province, we compare 5 models: Ratio-to-Trend, Holt-Winters Exponential Smoothing, Regression Dummy Variables, Theta, and Combined Forecasting. Two accuracy characteristics are used for forecasting: MAD and MSE. The results of the study show that Combined Forecasting was the best method to forecasting Malaria patients. The forecasted values from this model would be a useful guidance for timely prevention and control measures to define the Malaria control operation area which consists of 2 phases: transmission area (A) and non-transmission area (B) [22].

Acknowledgements

We would like to thank Surat Thani Provincial Health Office for providing the necessary information. We also would like to thank the Department of Mathematics, Faculty of Science and Technology, Suratthani Rajabhat University and Department of Mathematics, Faculty of Science, Kasetsart University for providing the facilities to carry out the research.

References

[1] Makridakis S, Wheelwright SC. **Forecasting Methods for Management.** 5th ed. New York: John Wiley & Sons; 1989.

[2] Hyndman RJ, Athanasopoulos G. **Forecasting: principle and practice [Internet].** OTexts; 2016 [cited 11 November 2016]. Available from: https://www.otexts.org/fpp/

[3] Manmin M. **Time series and forecasting.** Bangkok: Four Printing; 2006.

[4] Sanguanrungsirikul D, Chiewanantavanich H, Sangkasem MA. A Comparative Study to Determine Optimal Models for Forecasting the Number of Patients Having Epidemiological-Surveillance Diseases in Bangkok. **KMUTT Research and Development Journal.** 2015; **38**: 36-55.

[5] Wongkoon S, Pollar M, Jaroensutasinee M, Jaroensutasinee K. Predicting DHF incidence in Northern Thailand using time series analysis technique. **Int. J. Biol. Life Sci.** 2008; **4**: 117-121.

[6] Wangdi K, Singhasivanon P, Silawan T, Lawpoolsri S, White NJ, Kaewkungwal J. Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: A case study in endemic districts of Bhutan. **Mala J.** 2010; **9**: 251.

[7] Gomez-Elip A, Otero A, Herp MV, Anuirre-Jaime A. Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997-2003. **Mala J.** 2007; **6**: 1-10.

[8] Nithiworaset K. Forecasting model of Malaria incidence in Ubonratchatani Province by using monthly case report and weather factors. **[Thesis].** Khon Kaen, Thailand: Khon Kaen University; 2010.

[9] World Health Organization. **World malaria report 2015 [Internet].** Geneva, Switzerland: World Health Organization; 2015 [cited 4 April 2017]. Available from: http://www.who.int/malaria/publications/world-malaria-report-2015/report/en/

[10] Rojanawatsirivet C, Konchom S. Malaria incidence pattern in the 14 Southern Provinces of Thailand: 1991-2002. **Disease Control Journal**. 2004; **30**: 299-305.

[11] Bureau of Vector Borne Diseases. **Annual report 2015**. Bangkok, Thailand: Department of Disease Control Ministry of Public Health; 2015.

[12] The Center of Vector-Borne Diseases Control 11.3, Surat Thani. **The Operational Evaluation of Prevention and Control Vector-Borne Diseases.** Surat Thani, Thailand: The Office of Prevention and Control Disease Region 11; 2011.

[13] Hanke JE, Reitsch AG, Wichern DW. **Business Forecasting.** 7th ed. United State of America: Prentice-Hall; 2001.

[14] Suwanwongse S. **Quantitative forecasting Technique: Time Series Analysis.** 1st ed. Nakhon Pathom: Mahidol University Press; 2013.

[15] Taesombut S. **Quantitative forecasting.** 1st ed. Bangkok: Kasetsart University; 2006.

[16] Chandrachai A. **Forecasting Technique for Management.** 1st ed. Bangkok: Chulalongkorn University Press; 2015.

[17] Janmuang N. **Descriptive epidemiology of the communicable disease and forecasting important communicable disease in migrant workers, Songkhla province.** [Thesis]. Songkhla, Thailand: Songkhla Rajabhat University; 2013.

[18] The Office of Prevention and Control Disease Region 1. **Malaria prediction by time series analysis in Mae Hong Son province, 2013 [Internet].** ChiangMai, Thailand: Department of Disease Control Ministry of Public Health; 2013. [cited 1 May 2017]. Available from: http://1.10.141.27/epidpc10/view.php?ct_id=92

[19] Assimakopoulos V, Nikolopoulos K. The theta model: a decomposition approach to forecasting. **Int. J. Forecast.** 2000; **16**: 521-530.

[20] Hyndman RJ, Billah B. Unmasking the Theta model. **Int. J. Forecast.** 2003; **19**: 287-290.

[21] Damsri S, Payakkapong P, Hirunwong A. A Comparative Study of Decomposition Methods in Time Series Data Forecasting with Seasonal and Nonseasonal Variation. **Kasetsart Journal: Social Sciences.** 2007; **28**: 221-227.

[22] Bureau of Vector Borne Diseases. **The Malaria Control Operational Guideline for Public health personnel 2009** Bangkok, Thailand: Department of Disease Control Ministry of Public Health; 2015.