

# The Forecasting Model with Climate Variables of the Re-emerging Disease Rate in Elderly Patients

NIPAPORN CHUTIMAN<sup>1</sup>, PANNARAT GUAYJARERNPANISHK<sup>2</sup>,  
MONCHAYA CHIANGPRADIT<sup>1</sup>, PIYAPATR BUSABABODHIN<sup>1</sup>,  
SAOWANEE RATTANAWAN<sup>1</sup>, BUTSAKORN KONG-IED<sup>1,\*</sup>

<sup>1</sup>Data Science and Sustainable Agriculture Research Unit, Department of Mathematics,  
Faculty of Science, Mahasarakham University, Maha Sarakham 41150, THAILAND

<sup>2</sup>Faculty of Interdisciplinary Studies, Nong Khai Campus, Khon Kaen University,  
Nong Khai 43000, THAILAND

\*Corresponding author; butsakorn.k@msu.ac.th

*Abstract:* This research forecasted the incidence rate per 100,000 elderly population with food poisoning, pneumonia, and fever of unknown origin in Khon Kaen Province and Roi Et Province in the northeastern part of Thailand. In the study, the time series forecasting with Box-Jenkins Method (SARIMA model) and Box-Jenkins Method with climate variables, i. e total monthly rainfall, maximum average monthly temperature, average relative humidity, minimum average monthly temperature and average temperature (SARIMAX model) was performed. The study results revealed that the forecasting accuracy was closely similar to the model without the climate variables in the combined analysis although such climate variables had relationship with the incidence rate per 100,000 elderly population with food poisoning, pneumonia, and fever of unknown origin. Therefore, the appropriate model should be the SARIMA model because it is easier for analysis but with higher forecasting accuracy than the SARIMAX model.

*Key-Words:* re-emerging disease rate, SARIMA, SARIMAX

Received: March 18, 2021. Revised: July 20, 2021. Accepted: July 29, 2021. Published: August 4, 2021.

## 1 Introduction

At present, the population structure of Thailand has changed into aging societies. According to the 2020 Report of Ministry of Social Development and Human Security, all Thai elderly people at the age of 60 years upward were 11,627,132 (17.57%) of all population of the country. The largest number of elderly population were in the northern part with 3,684,395 elderly people and it is estimated that in the next 20 years the number of the elderly people at the age of 60 years upward will be at 30 percent. Regarding the changing conditions of the climate, the trend is toward more seriousness which has effects on the elderly people's wellbeing. In overall, most disasters of flood, windstorm, and drought mostly occur in the northern part at 43 percent of Thailand. If the climate change trend is more serious, it will have effects on the potential of the elderly people who are the main population of communities. The study on forecasting the re-emerging disease rate of the elderly people due to climate change is very important because this group of population is vulnerable and is likely

affected by the disaster dangers. Moreover, this issue is a problematic issue which should be paid full attention from related agencies for social preparation concerning elderly people in order to develop the policies in accordance with lifestyles of the elderly people. The holistic study enables the trend estimation for welfare arrangement to support the aging societies with completion and effectiveness.

Several researchers study on forecasting the incidence rates by using the Auto Regressive Integrated Moving Average (ARIMA) Model. [1] studied on evaluation and management of infectious and non-infectious diseases by using the ARIMA Model. In addition, [2] proposed the ARIMA Model to forecast the number of patients with epidemic diseases, and found the model accuracy at 92.1%. [3] forecasted the incidence rate of influenza in children at Wuhan, China by using the ARIMA Model and the Seasonal Auto Regressive Integrated Moving Average (SARIMA) Model. [4] studied the process of the self identification ResNet ARIMA algorithm or the

SIRA algorithm. [5] forecasted the number of injured in road accidents in northeast Algeria by using the ARIMA Model.

The forecasting with the ARIMA Model or the SARIMA Model uses only the data of the studied diseases. However, if influence of other variables are needed for combined study, the Auto Regression Integrated Moving Average with external variables (ARIMAX) Model is used to form the forecasting model. [6] studied on forecasting the incidence rate of malaria in Bhutan by using the ARIMAX Model to co-analyze the climate variables. In addition, [7] studied on forecasting the incidence rate of hand, foot and mouth disease in Guangdong, China by using the ARIMAX Model with joint analysis of climate variables.

The purpose of this research is to study on forecasting the re-emerging rate in the elderly people in Khon Kaen Province and Roi Et Province in the middle northeastern part of Thailand. Khon Kaen Province is a large smart city with the number of the elderly people on the second rank of the northeastern part whereas Roi Et is becoming a smart city with the main focus on health. Therefore, the researcher studies on forecasting the incidence rate of diseases in relation to climate changes i.e. food poisoning, pneumonia, and fever of unknown origin because these diseases are often found in the elderly people in the northeastern part of Thailand.

## 2 Data Preparation

This research emphasized on studying the time series forecasting method with Box-Jenkins Method by using the incidence rate per 100,000 elderly population at the ages of 60 years upward with food poisoning, pneumonia, and fever of unknown origin in Khon Kaen Province and Roi Et Province in the monthly classification during 2013 –2020. The study was on the time series forecasting method by using climate variables in the combined analysis with the incidence rate per 100,000 elderly population with re-emerging diseases, i.e.

- $X_1$ : total monthly rainfall,
- $X_2$ : maximum average monthly temperature,
- $X_3$ : average relative humidity,
- $X_4$ : minimum average monthly temperature,
- and  $X_5$ : average temperature.

## 3 Methods

### 3.1 Time Series Forecasting Method with Box-Jenkins Method

The Box-Jenkins method [8]-[10] is the forecasting method which uses stationary time series data or time series data converted to stationary data. SARIMA model is the product of seasonal and non-seasonal polynomials and is designated by SARIMA  $(p, d, q) (P, D, Q)_s$ , where  $(p, d, q)$  and  $(P, D, Q)$  are non-seasonal and seasonal components, respectively which defined as

$$\Phi(B^s)\varphi(B)(1-B^s)^D(1-B)^d y_t = \Theta(B^s)\theta(B)\varepsilon_t,$$

where  $\Phi$  is autoregressive (AR) parameter of seasonal component

$\varphi$  is autoregressive (AR) parameter of non-seasonal component

$\Theta$  is moving average (MA) parameter of seasonal component

$\theta$  is moving average (MA) parameter of non-seasonal component

$B$  is backward operator:  $B(y_t) = y_{t-1}$

$(1-B^s)^D$  is  $D^{th}$  seasonal difference of season  $S$

$(1-B)^d$  is  $d^{th}$  non-seasonal difference

$\varepsilon_t$  is an independently distributed random variable

$P$  and  $p$  are the orders of AR components

$Q$  and  $q$  are the orders of MA

components

$D$  and  $d$  are difference terms

### 3.2 Time Series Forecasting Method with Box-Jenkins Method with the Joint Analysis of Climate Variables

The forecasting model was determined with time series data and related variables in the joint analysis in the models of Auto Regression Integrated Moving Average with external variables (ARIMAX)  $(p,d,q)$  or Seasonal Auto Regressive Integrated Moving Average with external variables (SARIMAX)  $(p,d,q) (P, D, Q)_s$  which defined as

$$\Phi(B^s)\varphi(B)(1-B^s)^D(1-B)^d y_t = \beta_i x_i + \Theta(B^s)\theta(B)\varepsilon_t,$$

where  $x_i$  is explained variable,

$\beta_i$  is a parameter of  $x_i$ .

The procedure of the model analysis is as follows.

1. The relationship between the incidence rate per 100,000 elderly population with the studied

diseases and the related climate variables was analyzed by  $X_1$ : total monthly rainfall,  $X_2$ : maximum average monthly temperature,  $X_3$ : average relative humidity,  $X_4$ : minimum average monthly temperature, and  $X_5$ : average temperature.

2. The model with seasonal variables and the model without seasonal variables were analyzed to find the appropriate model.

3. The model hypotheses obtained from Step 1 were tested on errors with normal distribution by using the statistical test of Jarque Bera, on errors with means in zero by using  $t$ -test, on errors with constant variance by using White's Test, and on errors with independence by using the statistical test of Durbin-Watson.

4. Bayesian Information Criterion (BIC) of the models passing the hypothesis testing in Step 2 was calculated.

5. The most appropriate model was selected on the basis of the least BIC in Step 4, and the model appropriateness was tested by using Ljung-Box Q Test.

### 3.3 Criteria for the Effectiveness Comparison of the Forecasting Models

In the test of the forecasting accuracy, the mean absolute percent error (MAPE) was used.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100$$

where

$y_t$  represents the data real value at a time  $t$

$\hat{y}_t$  represents the data forecasting value at a time  $t$

$n$  represents the total number of the data

### 4 The Results

The relationship was analyzed between the incidence rate per 100,000 elderly population with food poisoning, pneumonia, and fever of unknown origin. The related climate variables were  $X_1$ : total monthly rainfall,  $X_2$ : maximum average monthly temperature,  $X_3$ : average relative humidity,  $X_4$ : minimum average monthly temperature, and  $X_5$ : average temperature. The analyzing results are illustrated in Table 1.

Table 1. The correlation coefficient between the incidence rate per 100,000 elderly population and the related climate variables in Khon Kaen Province and Roi Et Province

Province	Re-emerging Disease	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
Khon Kaen	food poisoning	0.161 (0.138)	0.250* (0.014)	0.038 (0.716)	0.209* (0.041)	0.230* (0.024)
	pneumonia	-0.161 (0.117)	-0.147 (0.154)	-0.188 (0.067)	-0.284** (0.005)	-0.238** (0.002)
	fever of unknown origin	0.463** (<0.005)	0.033 (0.746)	0.522** (<0.005)	0.355** (<0.005)	0.252* (0.013)
Roi Et	food poisoning	0.295** (0.004)	-0.041 (0.694)	0.267** (0.009)	0.211* (0.039)	0.179 (0.080)
	pneumonia	-0.084 (0.413)	-0.181 (0.077)	0.033 (0.752)	-0.152 (0.140)	-0.238* (0.019)
	fever of unknown origin	0.521** (<0.005)	-0.109 (0.292)	0.647** (<0.005)	0.423** (<0.005)	0.233* (0.023)

Remarks: 1. \* Statistical significance at 0.05.  
 2. \*\*Statistical significance at 0.01.  
 3. The values in brackets are p-value.

According to Table 1, in Khon Kaen Province, the total monthly rainfall ( $X_1$ ), the average relative humidity ( $X_3$ ), the minimum

average monthly temperature ( $X_4$ ) and the average temperature ( $X_5$ ) have linear relationship in the same direction as the incidence rate per 100,000

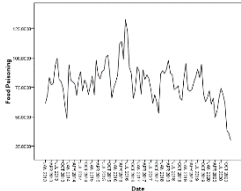
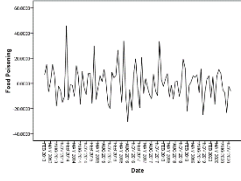
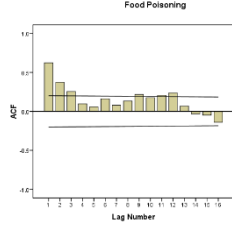
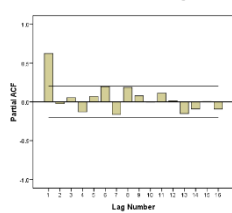
elderly population with fever of unknown origin. Meanwhile, the maximum average monthly temperature ( $X_2$ ), the minimum average monthly temperature ( $X_4$ ) and the average temperature ( $X_5$ ) have linear relationship in the same direction as the incidence rate per 100,000 elderly population with food poisoning. However, the minimum average monthly temperature ( $X_4$ ) and the average temperature ( $X_5$ ) have linear relationship in the opposite direction to the incidence rate of 100,000 elderly population with pneumonia.

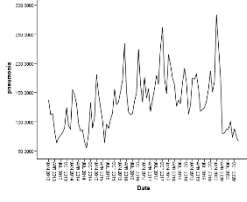
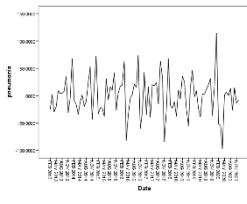
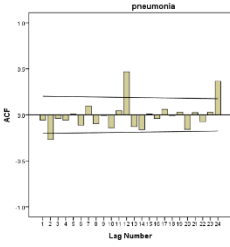
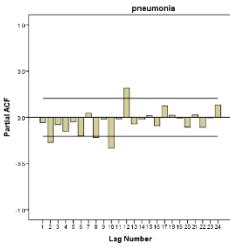
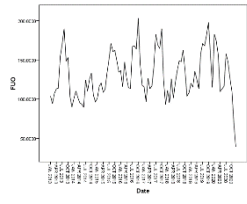
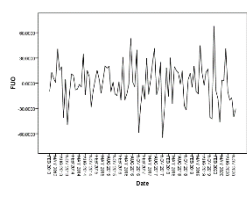
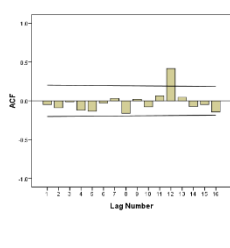
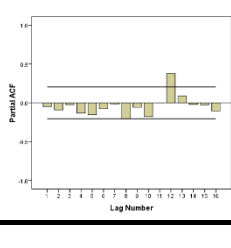
In Roi Et Province, the total monthly rainfall ( $X_1$ ), the average relative humidity ( $X_3$ ) and the minimum average monthly temperature ( $X_4$ ) have linear relationship in the same direction as the

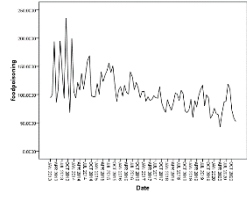
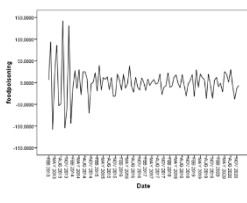
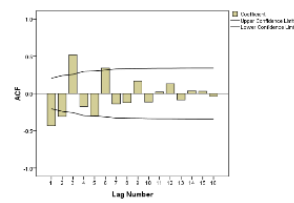
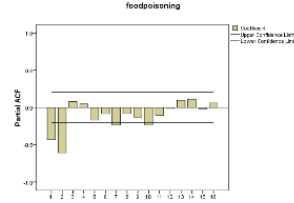
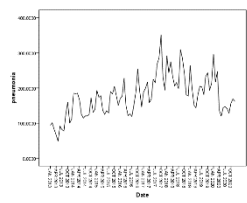
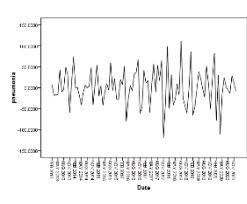
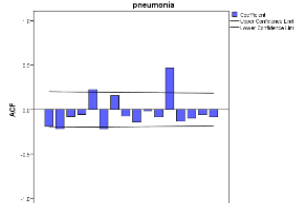
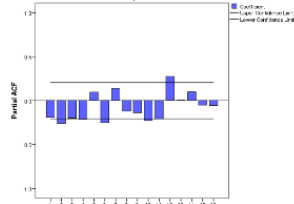
incidence rate per 100,000 elderly population with food poisoning. The total monthly rainfall ( $X_1$ ), the average relative humidity ( $X_3$ ), the minimum average monthly temperature ( $X_4$ ) and the average temperature ( $X_5$ ) have linear relationship in the same direction as the incidence rate per 100,000 elderly population with fever of unknown origin. Nevertheless, the average temperature ( $X_5$ ) has linear relationship in the opposite direction to the incidence rate of 100,000 elderly population with pneumonia.

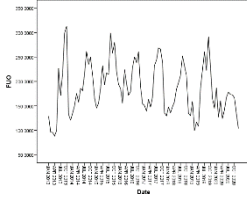
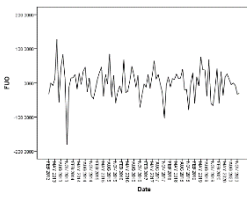
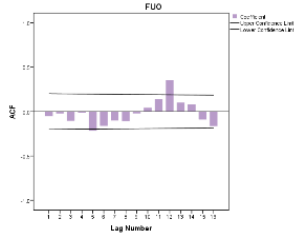
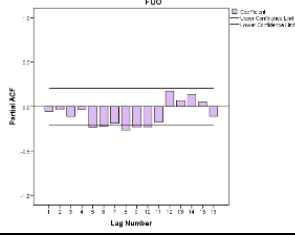
Table 2. illustrates the testing results about the stationary trend and seasonal to find out the appropriate model.

Table 2. The testing results of the stationary trend and seasonal to find out the appropriate model

Province	Re-emerging Disease	Stationary Inspection	Trend and seasonal Inspections
Khon Kaen	food poisoning	 <p>The data was non-stationary and converted into stationary by using the first difference.</p> 	<p>ACF</p>  <p>PACF</p> 

Province	Re-emerging Disease	Stationary Inspection	Trend and seasonal Inspections
Khon Kaen	pneumonia	 <p data-bbox="566 622 954 728">The data was non-stationary and converted into stationary by using the first difference.</p> 	<p data-bbox="981 324 1045 358">ACF</p>  <p data-bbox="981 672 1061 705">PACF</p> 
	fever of unknown origin	 <p data-bbox="566 1321 954 1422">The data was non-stationary and converted into stationary by using the first difference.</p> 	<p data-bbox="981 1019 1045 1052">ACF</p>  <p data-bbox="981 1366 1061 1400">PACF</p> 

Province	Re-emerging Disease	Stationary Inspection	Trend and seasonal Inspections
Roi Et	food poisoning	 <p>The data was non-stationary and converted into stationary by using the first difference.</p> 	<p>ACF</p>  <p>PACF</p> 
	pneumonia	 <p>The data was non-stationary and converted into stationary by using the first difference.</p> 	<p>ACF</p>  <p>PACF</p> 

Province	Re-emerging Disease	Stationary Inspection	Trend and seasonal Inspections
Roi Et	fever of unknown origin	 <p>The data was non-stationary and converted into stationary by using the first difference.</p> 	<p>ACF</p>  <p>PACF</p> 

According to Table 2, the non-stationary data were found in the three diseases in Khon Kaen Province and Roi Et Province so these data were converted into stationary by using the first difference. Regarding the inspections of trend and

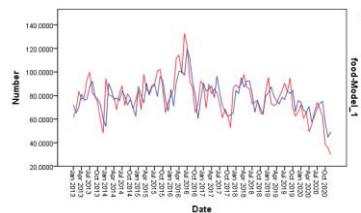
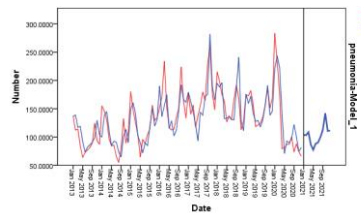
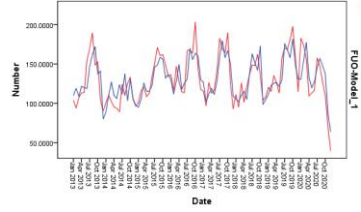
seasonal, there were trend and seasonal. The possibility of the models were analyzed which assumptions were valid and the results are shown in Table 3. The best model was determined by the minimum *MAPE*, as shown in Table 4.

Table 3. The analysis of the model possibility

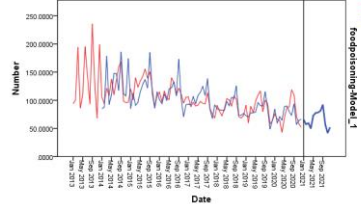
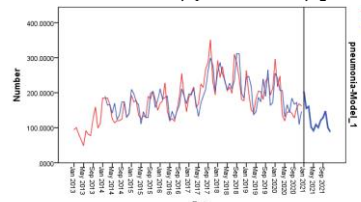
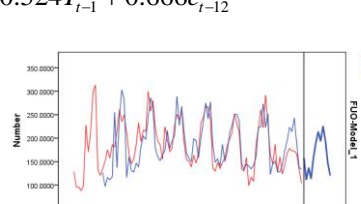
Province	Re-emerging Disease	Models	BIC	Ljung – Box Q (p-value)	MAPE
Khon Kaen	food poisoning	SARIMA(1,0,0)(2,0,0) <sub>12</sub>	5.107	13.245 (0.583)	12.070
		SARIMAX(1,0,0)(2,0,0) <sub>12</sub> with X <sub>2</sub>	5.162	13.445 (0.568)	12.048
		<b>SARIMAX(1,0,0)(2,0,0)<sub>12</sub> with X<sub>4</sub></b>	<b>5.163</b>	<b>13.590 (0.557)</b>	<b>11.999</b>
		SARIMAX(1,0,0)(2,0,0) <sub>12</sub> with X <sub>5</sub>	5.166	12.959 (0.605)	12.121
	pneumonia	<b>SARIMA(1,0,0)(1,0,1)<sub>12</sub></b>	<b>6.643</b>	<b>9.481 (0.851)</b>	<b>14.385</b>
		SARIMAX(1,0,0)(1,0,1) <sub>12</sub> with X <sub>4</sub>	6.793	12.219 (0.662)	15.410
		SARIMAX(1,0,0)(1,0,1) <sub>12</sub> with X <sub>5</sub>	6.793	11.485 (0.718)	15.563
	fever of unknown origin	SARIMA(1,0,0)(1,0,1) <sub>12</sub>	6.067	8.093 (0.920)	11.966
		SARIMAX(1,0,0)(1,0,1) <sub>12</sub> with X <sub>1</sub>	6.116	9.553 (0.848)	11.879
		SARIMAX(1,0,0)(1,0,1) <sub>12</sub> with X <sub>3</sub>	6.119	9.514 (0.849)	11.898
<b>SARIMAX(1,0,0)(1,0,1)<sub>12</sub> with X<sub>4</sub></b>		<b>6.017</b>	<b>8.470 (0.904)</b>	<b>11.227</b>	
SARIMAX(1,0,0)(1,0,1) <sub>12</sub> with X <sub>5</sub>		6.074	10.045 (0.817)	11.363	

Province	Re-emerging Disease	Models	BIC	Ljung – Box Q (p-value)	MAPE
Roi Et	food poisoning	<b>SARIMA(3,0,0)(0,1,1)<sub>12</sub></b>	<b>6.471</b>	<b>11.511 (0.645)</b>	<b>15.601</b>
		SARIMAX(3,0,0)(0,1,1) <sub>12</sub> with X <sub>1</sub>	6.475	12.239 (0.587)	15.659
		SARIMAX(3,0,0)(0,1,1) <sub>12</sub> with X <sub>3</sub>	6.476	11.956 (0.610)	15.646
		SARIMAX(3,0,0)(0,1,1) <sub>12</sub> with X <sub>4</sub>	6.477	12.050 (0.602)	15.621
	pneumonia	<b>SARIMA(3,0,0)(0,1,1)<sub>12</sub></b>	<b>7.269</b>	<b>10.078 (0.815)</b>	<b>13.701</b>
		SARIMAX(0,1,2)(0,1,1) <sub>12</sub> with X <sub>5</sub>	7.269	9.341 (0.859)	13.725
	fever of unknown origin	<b>SARIMA(1,0,0)(0,1,1)<sub>12</sub></b>	<b>7.287</b>	<b>23.222 (0.108)</b>	<b>13.152</b>
		SARIMAX(1,0,0)(0,1,1) <sub>12</sub> with X <sub>1</sub>	7.334	23.010 (0.113)	14.066
		SARIMAX(1,0,0)(0,1,1) <sub>12</sub> with X <sub>3</sub>	7.349	23.374 (0.104)	14.157
		SARIMAX(1,0,0)(0,1,1) <sub>12</sub> with X <sub>4</sub>	7.351	24.064 (0.088)	14.018
SARIMAX(1,0,0)(0,1,1) <sub>12</sub> with X <sub>5</sub>		7.352	23.257 (0.107)	14.145	

Table 4. The best model classified according to the provinces and the re-emerging diseases

Province	Re-emerging Disease	Model	Forecast Model
Khon Kaen	food poisoning	SARIMAX(1,0,0)(2,0,0) <sub>12</sub> with X <sub>4</sub>	$\hat{Y}_t = 71.481 + 0.732Y_{t-1} + 0.302Y_{t-12} + 0.282X_4$ 
	pneumonia	SARIMA(1,0,0)(0,1,1) <sub>12</sub>	$\hat{Y}_t = 4.793 + 0.849Y_{t-1} + 0.996Y_{t-12} + 0.910e_{t-12}$ 
	fever of unknown origin	SARIMAX(1,0,0)(1,0,1) <sub>12</sub> with X <sub>4</sub>	$\hat{Y}_t = 0.746Y_{t-1} + 0.997Y_{t-12} + 4.527X_4$ 



Province	Re-emerging Disease	Model	Forecast Model
Roi Et	food poisoning	SARIMA(3,0,0)(0,1,1) <sub>12</sub>	$\hat{Y}_t = -7.257 - 0.206Y_{t-2} + 0.508Y_{t-3} + 0.866e_{t-12}$ 
	pneumonia	SARIMA(0,1,2)(0,1,1) <sub>12</sub>	$\hat{Y}_t = -1.0935 + 0.2793e_{t-1} + 0.3905e_{t-2} + 0.8252e_{t-12}$ 
	fever of unknown origin	SARIMA(1,0,0)(0,1,1) <sub>12</sub>	$\hat{Y}_t = 0.524Y_{t-1} + 0.666e_{t-12}$ 

### 5 Conclusion and Discussion

Although the climate variables have relationship with the incidence rate (per 100,000 elderly population) of food poisoning, pneumonia, and fever of unknown origin; in the analysis for the appropriate model to forecast the incidence rate, only the variable of the minimum average monthly temperature ( $X_4$ ) were found in the forecasting model with the incidence rate of food poisoning and fever of unknown origin in Khon Kaen Province. Even though the forecasting model with climate variables is appropriate model, the forecasting accuracy is closely similar to the forecasting accuracy of the model without climate variable in the combined analysis. Therefore, the appropriate model should be the SARIMA model because it is easier for analysis

In addition, the forecasting model of the incidence rate in Roi Et Province does not use climate variables.

### Acknowledgements:

This research project was financially supported by Thailand Science Research and Innovation Fund

(TSRI) 2021 and Mahasarakham University. The authors would like to thank the associate editor and the referees.

### References:

- [1] R. Sato, Disease Management with ARIMA model in Time Series, *Einstein*, Vol. 11, No.1, 2013, pp. 128 – 131.
- [2] Y. Pan, M. Zhang, Z. Chen, M. Zhou and Z. Zhang, An ARIMA based model for forecasting the patient number of epidemic disease, *2016 13<sup>th</sup> International Conference on Service Systems and Service Management (ICSSSM)*, 2016, pp. 1-4.
- [3] Z. He and H. Tao, Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study, *International Journal of Infectious Diseases*, Vol. 74, 2018, pp.61-70.
- [4] P. Khanarsa, A. Luangsodsas and K. Sinapiromsaran, Self-Identification ResNet-ARIMA Forecasting Model, *WSEAS Transactions on Systems and Control*, Vol. 15(21), 2020, pp. 196-211.

- [5] F. Merabet, H. Zeghdoudi, On Modelling Seasonal ARIMA Series: Comparison, Application and Forecast (Number of Injured in Road Accidents in Northeast Algeria), *WSEAS Transactions on Systems and Control*, Vol. 15(25), 2020, pp. 235-246.
- [6] K. Wangdi, P. Singhasivanon, T. Silawan, S. Lawpoolsri, N. J White, J. Kaewkungwal, 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, *Malaria Journal*, Vol. 9, 251, 2010, doi:10.1186/1475-2875-9-251.
- [7] Z. Du, L. Xu, W. Zhang, D. Zhang, S. Yu, Y. Hao, Predicting the hand, foot, and mouth disease incidence using search engine query data and climate variables: an ecological study in Guangdong, China, *BMJ open*, Vol. 7 No. 10, 2017, doi:10.1136/bmjopen-2017-016263.
- [8] G. E. P. Box, G. M. Jenkins, G.C. Reinsel, *Time Series Analysis: Forecasting and Control*. 3<sup>rd</sup> edition. New Jersey: Prentice Hall, 1994.
- [9] J.D. Cryer, K.S. Chan, *Time series analysis: With applications in R*. New York. Springer-Verlag. 2008.
- [10] J. Wang, Y.H. Du, X.T. Zhang, *Theory and application with seasonal time series*. 1<sup>st</sup> ed. Nankai. Nankai University Press. 2008

### **Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)**

This article is published under the terms of the Creative Commons Attribution License 4.0

[https://creativecommons.org/licenses/by/4.0/de.ed.en\\_US](https://creativecommons.org/licenses/by/4.0/de.ed.en_US)

### **Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)**

Nipaporn Chutiman has organized the research. Pannarat Guayjarenpnpanishk, Monchaya Chiangpradit were responsible for the data and statistic. Piyapatr Busababodhin, Saowanee Rattanawan have analyzed the results of Section 4. Butsakorn Kong-ied carried out the manuscript.

### **Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself**

This research project was financially supported by Thailand Science Research and Innovation Fund (TSRI) 2021 and Mahasarakham University.