Thai Journal of Mathematics
Special Issue: The 17th IMT-GT ICMSA 2021
Pages 106–116 (2022)
http://thaijmath.in.cmu.ac.th

ISSN 1686-0209

Land Surface Temperature Prediction in Chiang Mai Province Thailand Using MODIS LST Data

Khodeeyoh Kasoh[1], **Salang Musikasuwan**[1,*] **and Rattikan Saelim**[1]

[1] *Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani, Thailand 94000*
e-mail : 6220320002@psu.ac.th (K. Kasoh); salang.m@psu.ac.th (S. Musikasuwan); rattikan.s@psu.ac.th (R. Saelim)

Abstract The temperature increase is one of the indicators of global warming. Therefore, Land Surface Temperature (LST) trends can be used to identify climate change. The objectives of this study were (i) to analyze the trend of LST change in Chiang Mai province (ii) to investigate the suitable models for predicting LST in Chiang Mai, Thailand. The observation data used in this study were obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) LST Data on the National Aeronautics and Space Administration (NASA) website and referred to as LST MODIS. The data were collected every 8 days from January 1, 2001, to December 27, 2020 (920 observations).The data were split into 70%-30% proportions for training and checking datasets, respectively. In this study, the simple linear regression was used to analyze trends of the average LST change over 20 years. It is found that, the average LST in Chiang Mai province has been slightly increasing around 0.0184 degrees Celsius per year. The autoregressive integrated moving average (ARIMA) model has been applied for predicting LST, and the Root Mean Squared Error (RMSE) and coefficient of determination (R-squared) were used to measure the model performance. The results showed that ARIMA(2,0,0) model had the smallest RMSE for both training and checking data sets. In addition, all fitted ARIMA models can describe the LST with R-squared ranging from 0.6404 - 0.7871.

MSC: 49K35; 47H10; 20M12
Keywords: land surface temperature; climate change; autoregressive integrated moving average

Submission date: 31.01.2022 / Acceptance date: 31.03.2022

1. Introduction

The changes in Land Surface Temperature (LST) on sub-continental or regional sizes reveal distinct characteristics. In addition, the regional climate is more complex than the global climate since it is impacted by ocean-atmospheric circulation, land cover, and feedback processes. Thus, the regional climate is important for the environment and economic output [1].

*Corresponding author.

Published by The Mathematical Association of Thailand.

Time series analysis is used to develop mathematical models for the purpose of computing statistics from data on climatic variables using the autoregressive integrated moving average (ARIMA) model. Since the 1970s, time series analysis has rapidly advanced in theory and practice for the purpose of predicting and controlling various climatic factors such as precipitation, temperature, and others [2]. In 2014, Wang et al. [3] proposed the predictive model for monthly precipitation at the Lanzhou station in Lanzhou, China, using the enhanced ARIMA model. The findings indicated that the revised model is much more accurate than the seasonal model, with a mean residual of 9.41 mm and a forecast accuracy of 21%. Later in 2015, Bari et al. [4] forecasted monthly precipitation for Sylhet, Bangladesh, using the ARIMA model. They discovered that the ARIMA(0,0,1)(1,1,1) technique was the most efficient for forecasting future precipitation with a 95% confidence interval. El-Mallah et al. [5] used the Box-Jenkins method to predict the annual warming trend in 2016 and found that ARIMA(3,1,2) and ARIMA(3,2,3) were capable of predicting non-seasonal linear and quadratic trend models with results that followed their predicted patterns with correlation values of around 80% for both models. Morever, in 2017 Wongsai et al. [6] presented the analysis of annual seasonality extraction using the cubic spline function and decadal trend in temporal daytime Moderate Resolution Imaging Spectroradiometer (MODIS) LST data. Later in 2018, Sharma et al. [7] analyzed LST data from 2000 to 2015 to determine seasonal variations in the Kathmandu Valley of Nepal. They discovered that the patterns were significantly associated with altitude (p-value $< .01$). In the same year, Ruchiraset and Tantrakarnapa [8] conducted the study about time-series modeling of pneumonia admissions and its association with air pollution and climate variables in Chiang Mai Province, Thailand. As Chiang Mai Province faced with the variation of climate change, the study of trend and predicting model for temperature would be considered to conduct and analyze.

Chiang Mai locates in the northern Thailand. It is one of Thailand's major cities with 696 kilometers north of Bangkok. Its landscape is a mountain-rimmed basin [9]. Chiang Mai is one of cities that faced with extremely air pollution. There were some studies focused on this city. In 2012, Gou et al. [10] reported that particulate matter and ozone are the principal ambient contaminants in Chiang Mai. Suwanprasit [11] had analyzed the changes in land use and LST across Mueang Chiang Mai District, Thailand using satellite photos from Landsat TM and ETM+. The findings demonstrated that during the research period, the city's land usage changed dramatically, the maximum LST values found at bare ground area, and lowest LST values found at the forest, farm, and water resource classes. The temperature difference between cities and suburbs was 1 - 2C in 1994 and 5-8C in 2014.

Chiang Mai is the top destination for domestic and foreign tourists, and the surface temperature is one of the important factors that tourists use for making their decisions to visit Chiang Mai. The short-term predictive model would be suggested for both tourists and tourism agencies. Hence, this study has been conducted to analyze the trend of LST change in Chiang Mai Province using MODIS LST from the NASA website and investigated the suitable models for predicting short-term LST in Chiang Mai Province. The rest of this paper is organized as follows: Section 2 explains the data and methods used in this study. The results of this study have been presented in Section 3. Finally, Section 4 describes our discussion and conclusion.

2. MATERIALS AND METHODS

2.1. CONCEPTUAL FRAMEWORK

The first step in our investigation was to get LST data from the NASA website. Because the original LST's unit was Kelvin and there were missing values, it was necessary to undertake data management. After that, the LST season was explored by averaging LST on the same day of the year. Following that, a simple linear regression was used to determine the trend of LST change in Chiang Mai. The ARIMA was used as a predictive model to fit the original LST. Finally, the model's performance was assessed by using the root mean square error (RMSE) and the coefficient of determination (R-squared).

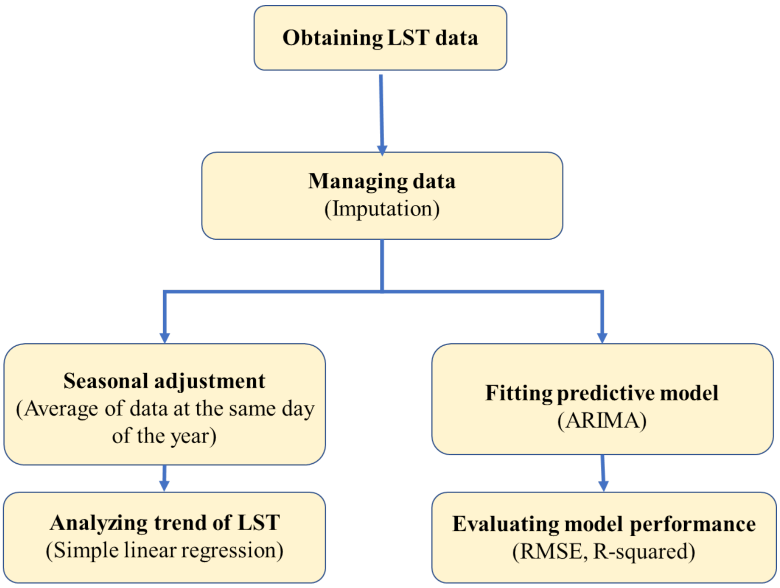

FIGURE 1. Conceptual Framework

2.2. STUDY AREA

Chiang Mai is the biggest city in northern Thailand. It is located at latitude 18.793867 and longitude 98.997116 and covering the area of 20,170 square kilometers. It is divided into 25 districts. There are 1,682,164 people living in the city, with 742,489 households. Chiang Mai has three seasons: winter (November to February), summer (March to May), and the rainy season (June to October) [8]. In this study, 9 regions in Chiang Mai province were selected (black dot in Figure 2). Each region comprising 49 pixels in 77 arrays as shown in the right panel of Figure 2.

2.3. DATA COLLECTION AND EXPLORATION

The LST data were obtained from MODIS on the NASA website using MOD11A2 product, which was collected every 8 days during January 1, 2001, and December 27, 2020 (in total, 920 data observations). LST data of 9 regions have been considered for this study. The LST units were converted to degrees Celsius by subtracting 273.15 from the Kelvin values. Figure 2 showed the time-series plots of LST for each region over 20 years. The time series plot of the 9 regions and the overall average (of all 9 regions) have been presented in Figure 3.

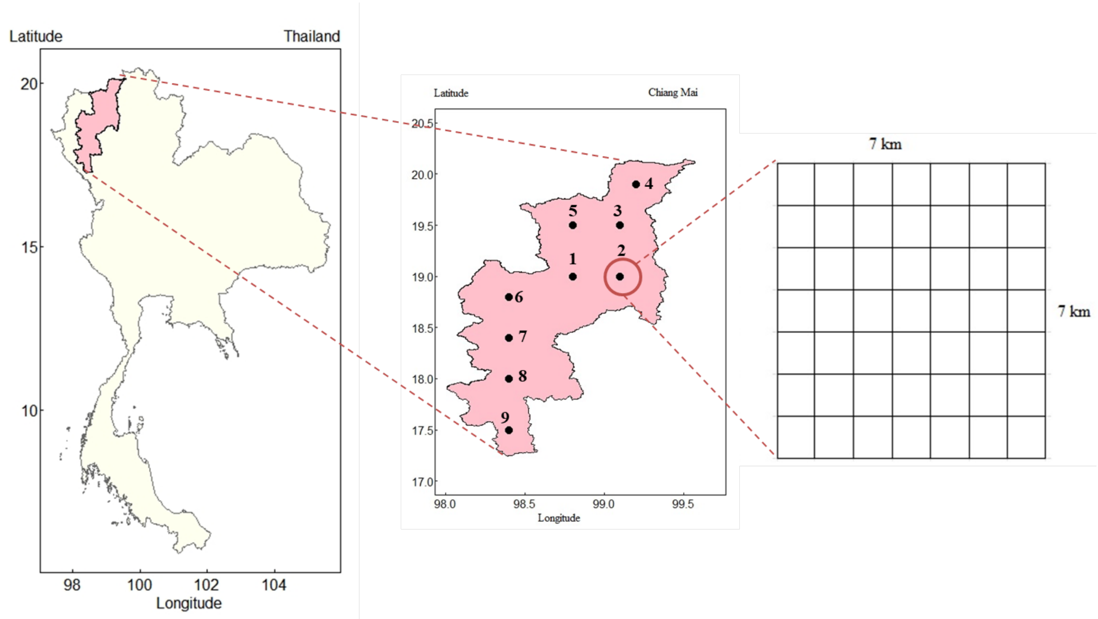

FIGURE 2. Study area for LST in urban area of Chiang Mai Province

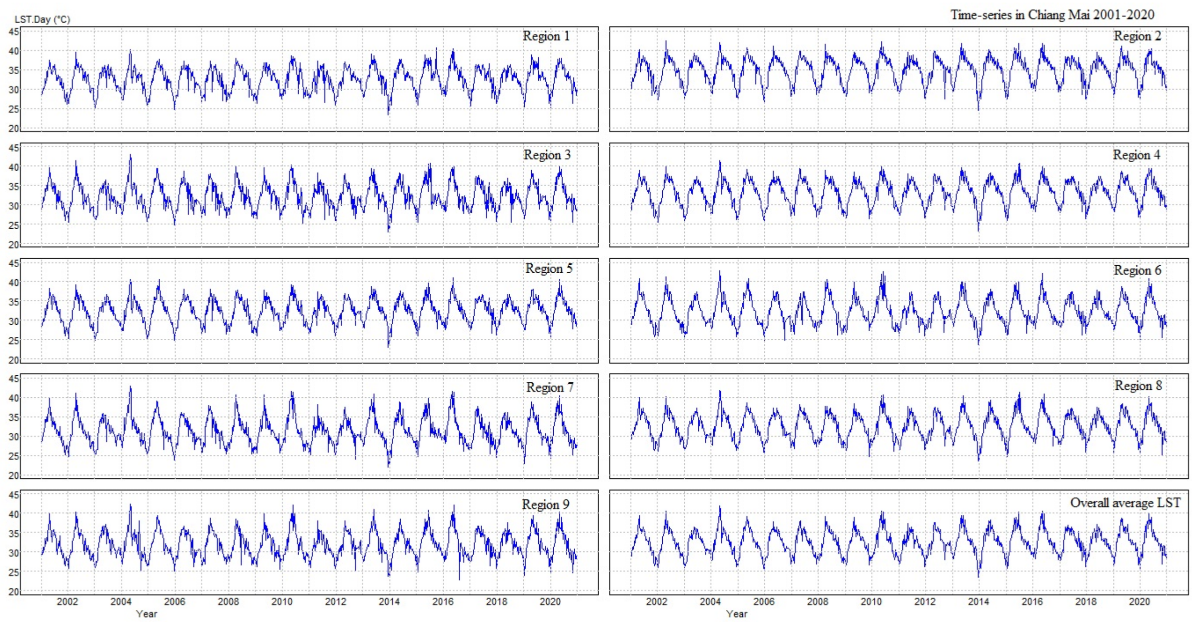

FIGURE 3. Time-series plots of LST for 9 regions and overall average LST

There are some missing values in the downloaded data. In this study, mean substitution has been selected for dealing with missing values. To construct the predicting models, the data have been divided into 70%-30% proportions for training and checking data sets for the 9 regions and the overall average LST. Training dataset has been used for constructing the predictive models, while checking dataset has been used for evaluating and validating the predictive models by considering the RMSE and R-squared.

2.4. METHODS

Simple linear regression was fitted to investigate the temperature trends. The simple linear regression model takes the following form (2.1):

$$\hat{y}_t = \beta_0 + \beta_1 t + \epsilon, \tag{2.1}$$

where $\hat{y}_t$ is the fitted LST, β_0 is the intercept, β_1 is the regression coefficient, t is the time (t=1,2,3,...,920) and ϵ is the error term.

Time series models have been built using stationary variables that have the same mean and variance across time. In principle, ARIMA models are the best models for forecasting a time-series. However, fitting a suitable model, estimating the parameters, and validating the model are all part of the process [12]. The best prediction model for all 9 regions and their average turned out to be ARIMA(2,0,0) whose general equation is: as shown in Equation 2.2.

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, \tag{2.2}$$

where $\hat{y}_t$ is the predicted LST value at observation t, ϕ_1,ϕ_2 are coefficients of the lag variables, y_{t-1} and y_{t-2}, respectively. ϵ_t is the value not explained by the model.

After obtaining the appropriate models, the predicted value is calculated from the training and checking datasets and evaluating those models using RMSE as shown in Equation 2.3.

$$RSME = \sqrt{\sum_{t=1}^{n} \frac{(y_t - \hat{y}_t)^2}{n}}, \tag{2.3}$$

where n is the defined as the number of predicted data, t is the defined time, y_t is the observation at time t, and $\hat{y}_t$ is the predicted value. Finally, the final models were tested using R-squared to determine whether they fit well enough for the training dataset. Equation 2.4 shows the formula for calculating the R-squared value.

$$R^2 = 1 - \sum_{t=1}^{n} \frac{(\hat{y}_t - y_t)^2}{(y_t - \overline{y}_t)^2}, \tag{2.4}$$

where $\hat{y}_t$ is the predicted LST value at observation t, y_t is LST value, $\overline{y}_t$ is the mean of LST value.

3. RESULTS

3.1. SEASONAL PATTERNS ANALYSIS

To eliminate the effect of the seasonality, the seasonal adjustment has been performed. The stationary of LST can be checked at this process. The time series plot of original data (blue lines) and seasonal patterns (red curve) in Chiang Mai over 20 years were shown in Figure 4. While Figure 5 presented the time series plot of seasonal adjusted LST. After perfoming seasonal adjustment, all data were satisfied for applying with ARIMA model. LST increased slightly in all regions, including the overall average LST, after fitting the basic linear regression for each location.

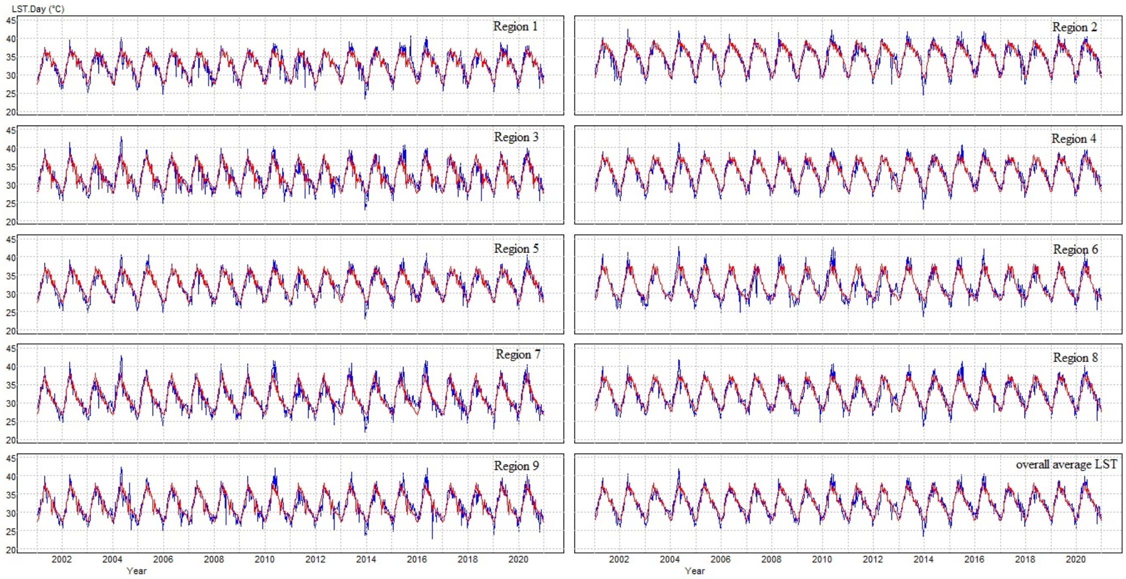

FIGURE 4. Time-series plots and seasonal patterns

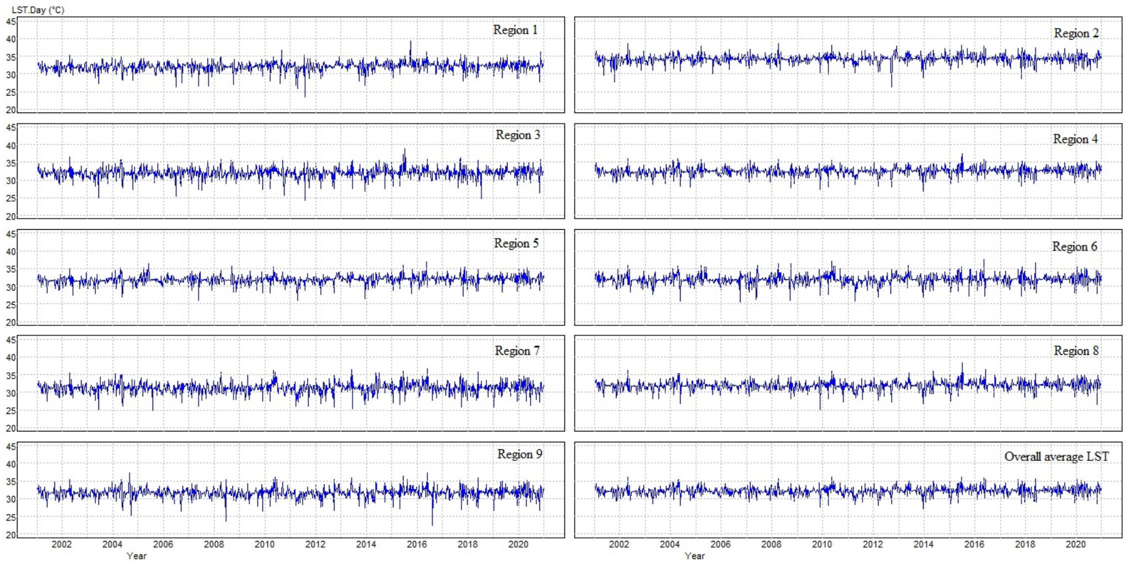

FIGURE 5. Time-series plots of LST after applied seasonal adjusted.

3.2. LAND SURFACE TEMPERATURE TREND ANALYSIS

After fitting the simple linear regression for each location, it was discovered that LST increased slightly in all regions, including the overall average LST. Each blue points represents original LST and red linear line represents the trend of LST for each region as showed in Figure 6.

3.3. LAND SURFACE TEMPERATURE PREDICTING MODELS

To construct the predictive models, the seasonally adjusted of training datasets will be used to create the ARIMA model by considering the graph of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to determine the q and p parameters of ARIMA model as shown in Figure 7.

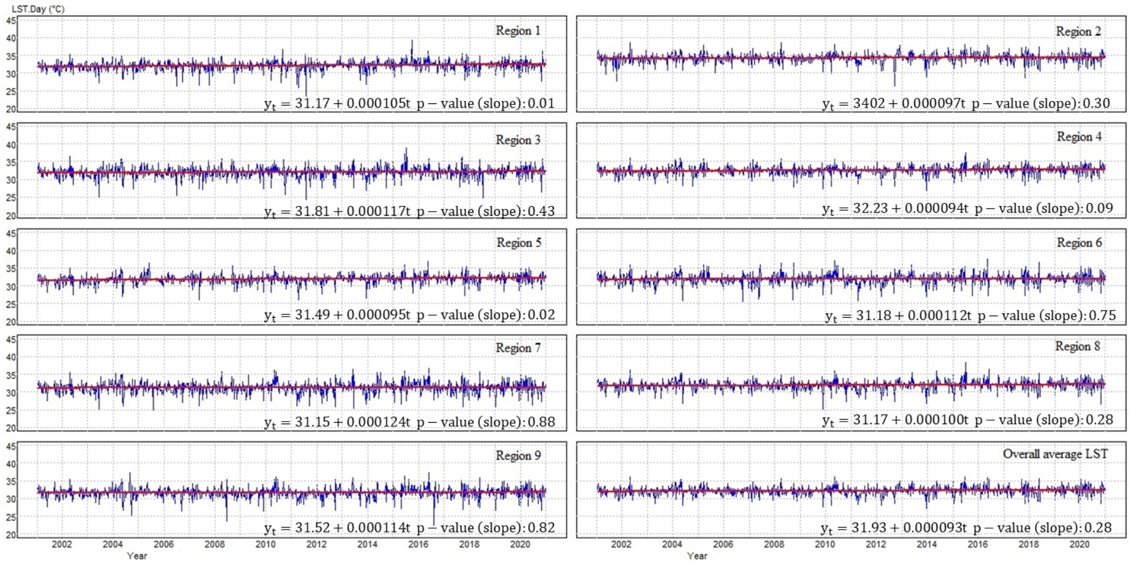

FIGURE 6. The red line showed the trend of LST change in Chiang Mai for 9 regions and overall average LST.

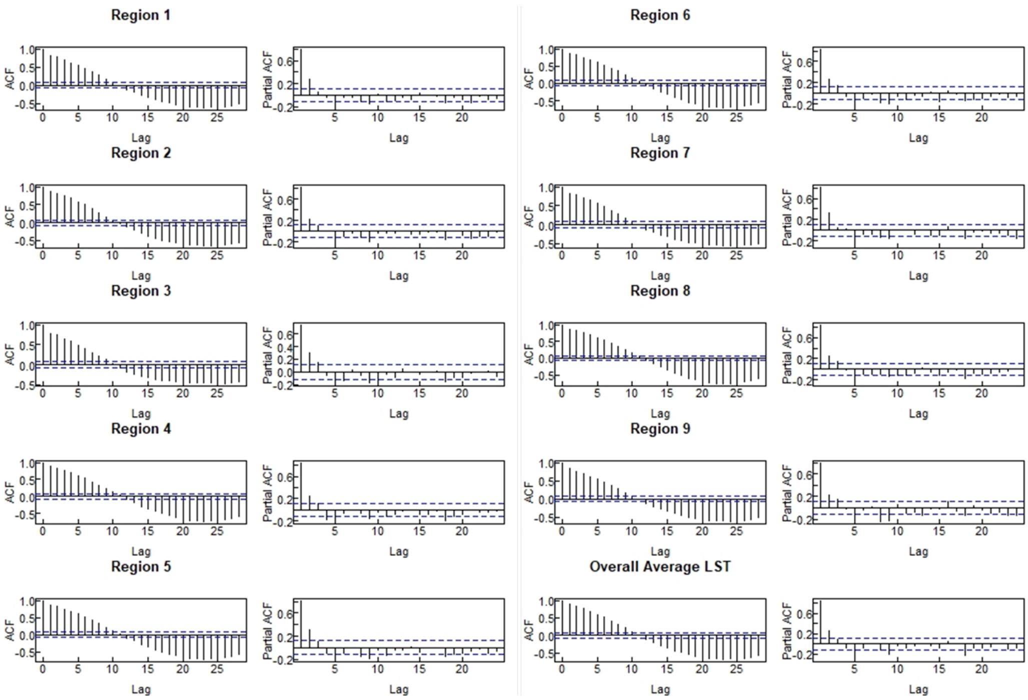

FIGURE 7. The graph of ACF and PACF for 9 regions and overall average LST

It was found that ARIMA(2,0,0) was the most suitable for dataset in all regions. The equations of ARIMA(2,0,0) for 9 region and overall average LST was given in Table 1. To evaluate the performance of the obtained models, the results of each model will be compared with testing and checking datasets (original datasets) using RMSE. Table 2 showed the obtained RMSE values for all regions and overall average LST. In fact, the obtained RMSE for training dataset for all regions varied between 1.40 and 1.96 degree Celsius, and for checking dataset varied between 1.51 and 2.07 degree Celsius. In addition, all fitted ARIMA models can describe the LST with R-squared ranging from 0.6404 to 0.7871.

Table 1. The equation of ARIMA (2,0,0) for each region and overall average LST

Region	Equation of ARIMA(2,0,0)
1	$\hat{y}_t$=3.8537+0.5501y_{t-1}+0.3304y_{t-2}
2	$\hat{y}_t$=3.2307+0.6108y_{t-1}+0.2959y_{t-2}
3	$\hat{y}_t$=4.5602+0.4996y_{t-1}+0.3576y_{t-2}
4	$\hat{y}_t$=2.6698+0.6491y_{t-1}+0.2697y_{t-2}
5	$\hat{y}_t$=2.9785+0.5693y_{t-1}+0.3380y_{t-2}
6	$\hat{y}_t$=3.0152+0.6433y_{t-1}+0.2630y_{t-2}
7	$\hat{y}_t$=3.2677+0.5176y_{t-1}+0.3778y_{t-2}
8	$\hat{y}_t$=2.8779+0.6327y_{t-1}+0.2783y_{t-2}
9	$\hat{y}_t$=4.1001+0.6561y_{t-1}+0.2148y_{t-2}
overall average LST	$\hat{y}_t$=2.6691+0.6440y_{t-1}+0.2735y_{t-2}

Table 2. TABLE 2 The RMSE for 9 region and overall average LST

Region	RMSE		R-squared
	Training	Checking	
1	1.64776	1.72194	0.7023
2	1.57787	1.65628	0.7558
3	1.95550	2.07311	0.6404
4	1.46398	1.51291	0.7904
5	1.52230	1.58928	0.7533
6	1.63139	1.66319	0.7662
7	1.80424	1.92773	0.7293
8	1.52475	1.65968	0.7726
9	1.77237	1.95997	0.6989
Overall Average LST	1.40722	1.53105	0.7871

The plots of predicted LST against original data for all regions using the ARIMA(2,0,0) model have been presented in Figure 8. Each blue line represents original LST, the red line represents predicted LST for training dataset and the green line represent predicted LST for checking dataset.

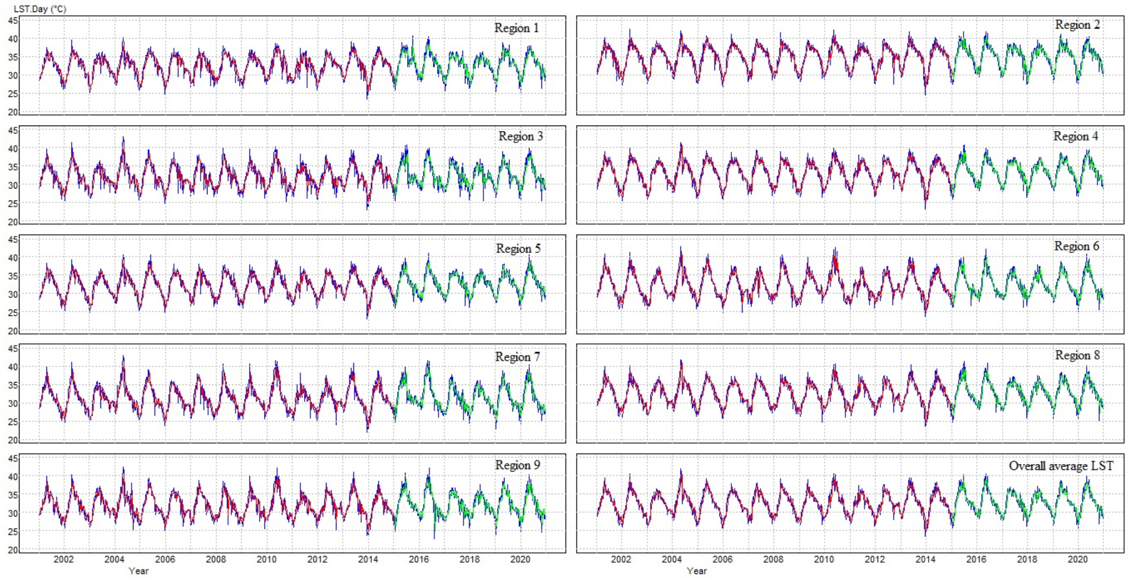

FIGURE 8. The plots of predicted against original values of LST using ARIMA(2,0,0) for region 1-9 and overall average LST

4. DISCUSSION AND CONCLUSION

The objectives of this study were to analyze the trend of LST change in Chiang Mai Province and investigate the suitable models for predicting land surface temperature in Chiang Mai, Thailand, by using RMSE as a measurement. The 8-day observation data were obtained from the NASA website.

By analyzing the data used in this study, the LST in Chiang Mai showed the maximum of 41.893 degrees Celsius, the minimum of 23.35 degrees Celsius, and the average of 32.61 degrees Celsius.

The simple linear regression had been used to analyze trend of the average LST change over 20 years. From the analysis, it was found that the LST of all regions approximately increase over 20 years. As an overall LST in Chiang Mai Province, the average LST has been slightly increasing around 0.0184 degrees Celsius each year. It should be noticed that region 2 has a greater LST than the other regions. This could be due to the fact that it covers the Chiang Mai city area, which is densely populated with high-rise buildings and has a high level of commuting. As a result, this research will provide evidence to policymakers so that they are aware of the impact of climate change in Chiang Mai.

In this study, the LST observations had been divided into 2 partitions as 70%:30% for training and checking datasets, respectively. To investigate the suitable predictive models for the LST, the ARIMA model has been applied with training dataset. The RMSE and R-squared were used to measure the performance of the models. The results showed that ARIMA(2,0,0) model had the smallest RMSE while testing with both training and checking datasets. It can be suggested that our final ARIMA(2,0,0) model was suitable for predicting LST in Chiang Mai Province. Furthermore, the model derived from the average of overall LST can be used to represent the entire province of Chiang Mai. Noted that, this study assessed only LST data in Chiang Mai Province, so the finding of this study did not provide a general conclusion for other locations. For future work, it would be suggested that machine learning techniques can be applied for developing the LST predictive model to compare the performance with ARIMA.

Acknowledgements

This work was partially supported by Faculty of Science and Technology, and Graduate School, Prince of Songkla University, Pattani Campus, Thailand. We would also like to express our gratitude to the referees for their helpful suggestions, which significantly improved the manuscript.

References

[1] G. Yan, D. Wen-Jie, R. Fu-Min, Z. Zong-Ci, H. Jian-Bin, Surface air temperature over China with CMIP5 and CMIP3, Advances in Climate Change Research 4 (3) (2013) 145-152.

[2] K.H. Machekposhti, H. Sedghi, A. Telvari, H. Babazadeh, Modeling climate variables of rivers basin using time series analysis (case study: Karkheh river basin at Iran), Civil Engineering Journal 4 (1) (2018) 78-92.

[3] H.R. Wang, C., X. Lin, J. Kang, An improved ARIMA model for precipitation simulations, Nonlinear Processes in Geophysics 21 (6) (2014) 1159-1168.

[4] S.H. Bari, M.T. Rahman, M.M. Hussain, S. Ray, Forecasting monthly precipitation in Sylhet city using ARIMA model, Civil and Environmental Research 7 (1) (2015) 69-77.

[5] E.S. El-Mallah, S.G. Elsharkawy, Time-series modeling and short term prediction of annual temperature trend on Coast Libya using the Box-Jenkins ARIMA Model, Advances in Research 6 (5) (2016) 1-11.

[6] N. Wongsai, S. Wongsai, A.R. Huete, Annual seasonality extraction using the cubic spline function and decadal trend in temporal daytime MODIS LST data, Remote Sensing 9 (12) (2017) 1-17.

[7] I. Sharma, A. Ueranantasun, P. Tongkumchum, Modeling of land surface temperatures to determine temperature patterns and detect their association with altitude in the Kathmandu Valley of Nepal, Chiang Mai University Journal of Natural Sciences 17 (4) (2018) 275-288.

[8] A. Ruchiraset, K. Tantrakarnapa, Time series modeling of pneumonia admissions and its association with air pollution and climate variables in Chiang Mai province, Thailand, Environmental Science and Pollution Research 25 (33) (2018) 33277-33285.

[9] N. Kammuang-Lue, P. Sakulchangsatjatai, P. Sangnum, P. Terdtoon, Influences of population, building, and traffic densities on urban heat island intensity in Chiang Mai City, Thailand, Thermal Science 19 (2) (2015) 445-455.

[10] Y. Guo, K. Punnasiri, S. Tong, Effects of temperature on mortality in Chiang Mai city, Thailand: a time series study, Environmental Health 11 (1) (2012) 1-9.

[11] C. Suwanprasit, Pattern of land-use change and urban heat island during 20 years in Chiang Mai, Thailand, International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management 41 (17) (2017) 535-542.

[12] R. Anokye, E. Acheampong, I. Owusu, E. Isaac Obeng, Time series analysis of malaria in Kumasi: using ARIMA models to forecast future incidence, Cogent Social Science 4 (1) (2018) 1-13.