Land Surface Temperature Changes in Songkhla, Thailand from 2001 to 2018

Rattikan Saelim[1,2]*, Salang Musikasuwan[1,2] and Nasuha Chetae[1]

[1] Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani, 94000, Thailand

[2] Centre of Excellence in Mathematics, CHE, Bangkok 10400, Thailand

* Corresponding author. E-mail address: rattikan.s@psu.ac.th

Received: 31 October 2019; Revised: 12 February 2020; Accepted: 18 February 2020

Abstract

Agriculture is one of the most important factors contributing to the economy of Songkhla province in Thailand. Since the agriculture is highly dependent on the climate and hence on the temperatures, it was the aim of this study to investigate the trends and model the Land Surface Temperature (LST) from January 2001 to December 2018 of Songkhla province. Firstly, simple linear regression was applied and it was found that LST has increased approximately 0.3312 degrees Celsius during the last 18 years. After that the data were divided by 70:30 split to training and testing sets. Then for the predictive model, multiple linear regression and ARIMA (p,d,q) models were fit. Among the possible choices of (p,d,q) parameters in ARIMA, (3,0,0) performed the best. Further, according to their respective root mean squared errors (RMSE), namely 1.3334 and 1.3248, the ARIMA (3,0,0) performed slightly better than multiple linear regression in the training set. However, multiple linear regression performed slightly better than ARIMA (3,0,0) in the test data, with respective RMSEs 1.3249 and 1.3489. In other words, ARIMA gave a better fit, but linear regression gave better predictions. It is worth noting that the performance of a model type varies depending both on context and on the proportions of training and testing sets, so this case study demonstrates a model comparison approach but the results do not allow a generally applicable conclusion of ranking the model types.

Keywords: Land Surface Temperature, temperature prediction, Remote Sensing, Linear Regression, ARIMA

Introduction

Temperature is one of the key variables used to describe climate change. Long-term changes in temperature at a fixed location, say, over years, decades, and centuries, is indicative of climate change. In terms of food security, modern humans rely on agriculture. However, there are many studies confirming that agriculture is highly dependent on the climate (e.g., Parry & Carter, 1989; Bachelet, Brown, Böhm, & Russell, 1992; Parry, Rosenzweig, Iglesias, Livermore, & Fischer, 2004; Kang, Khan, & Ma, 2009; Valizadeh, Ziaei, & Mazloumzadeh, 2014; Hatfield & Prueger, 2015; Iizumi & Ramankutty, 2015; Powell & Reinhard, 2016). For example, a study on how climate change effects the wheat production (Valisadeh, et al., 2014) found that reduced wheat growth season length in various climate change scenarios could be caused by an increase in temperature. There is evidence (Somparn et al., 2004) showing that heat could reduce production rates by the cattle and buffalo industries in the northeastern region of Thailand. Moreover, an increased temperature exhibits a large impact on grain yield (Hatfield & Prueger, 2015). In fact, it is observed that maize grain yield is greatly reduced by above normal temperatures during the grain-filling period.

In the last 5 decades, there is a significant country-wide warming of 1.30 degrees Celsius in Thailand (Limsakul et al., 2019). This warming agrees with other studies (Sharma & Babel, 2014) showing a significant increase in the annual number of warm days and warm nights.

Songkhla (Figure 1) is a province in southern Thailand connecting to four other provinces, one country (Malaysia), and the gulf of Thailand on the east coast (Figure 1). Songkhla's economy depends mainly on industry and agriculture, whose contributions to the Gross Domestic Product (GDP) in 2016 were 20.01% and 14.63%, respectively (Office of Commercial Affairs Songkhla, 2018). The purposes of this paper are (i) to investigate the trends in LST in Songkhla over the last 18 years, and (ii) to compare the suitability of some conventional models. In this case, the model types tested are multiple linear regression and ARIMA (p,d,q) models.

Data

The data used in this study were obtained from the NASA website at https://modis.ornl.gov/data.html. Each time series has time (as date) and Land Surface Temperature (Day-LST) by 1-square-kilometer-pixels, from 2001 to 2018. The 8-day Day-LST averaging is already done by NASA to the daily LST. Hence, this 8-day time series has 46 observations each year, or 828 observations over 18 years. In other words, one can picture this data set as 828 layers of 11,025-pixel sheets (our study area covering mainly the Songkhla province). Now, the 11,025 pixels were averaged at each particular point of time, in order to have just one time series (Figure 2) for analysis.

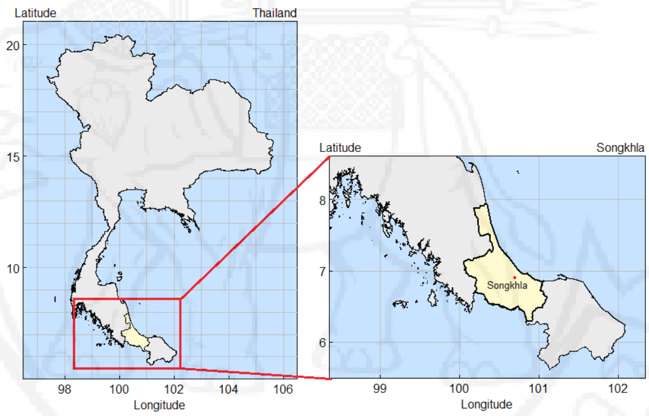

Figure 1 Study area: Songkhla province in southern Thailand.

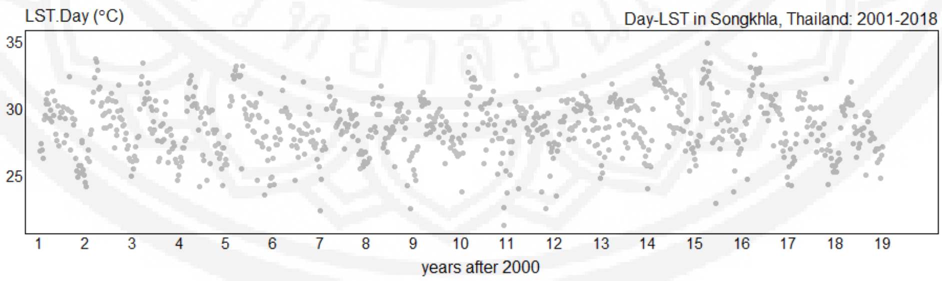

Figure 2 Average 8-day Day-LST at each point of time from 2001 to 2018.

Methods and Results

Seasonal adjustment

Seasonal adjustment is a method designed to smooth out the data in order to make it easier to observe cyclical patterns, trends, or other non-seasonal movements. According to NASA data manipulation, the observation (after averaging) of LST starts on the 1st of January each year. The following observations will be at every 8 days until day 361 of the year and then start again on the next 1st of January. Since temperature on earth depends on both latitude and time of the year, in this case the seasonal pattern or trend is obtained by averaging the observations on the same day of each year, i.e., 1, 9, 17, …, 353, and 361, as shown in Figure 3 (top). For the seasonal adjustment, we first difference the original time series with its trend (average value of temperatures each year at the same day of year), then subtract the result by its average and add back the average of the original time series to obtain the seasonally adjusted Day-LST shown in Figure 3 (bottom).

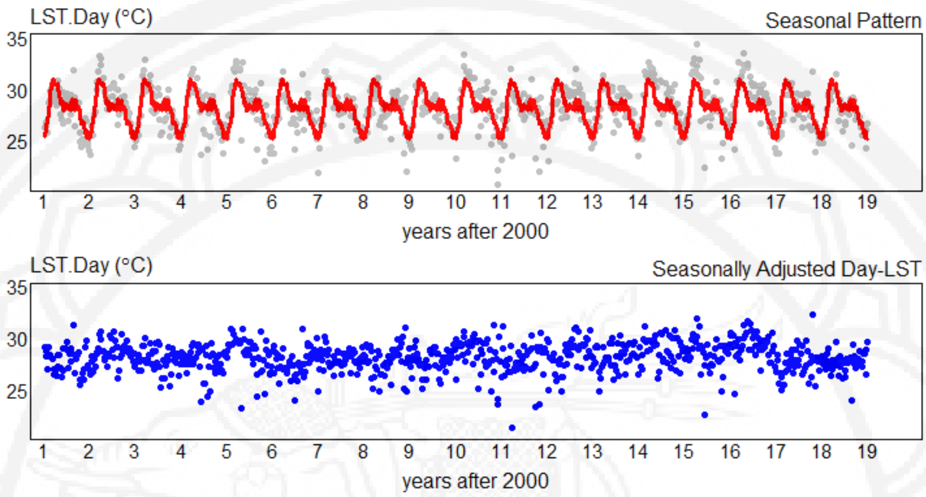

Figure 3 Top: Day-LST (grey dots) and Seasonal pattern (red line). Bottom: Seasonally adjusted Day-LST.

Tendency

The long term trend of Day-LST is obtained by fitting with simple linear regression, with seasonally adjusted Day-LST as the dependent variable and time of observation as the independent variable. With p-values less than 0.05 for both intercept, 29.2500, and slope, 0.0004, the model becomes y = 29.2500 + 0.0004t. The increase over 828 points of time is, by simple calculation, 0.3312 degrees Celsius in the last 18 years or 1.84 degrees Celsius per century.

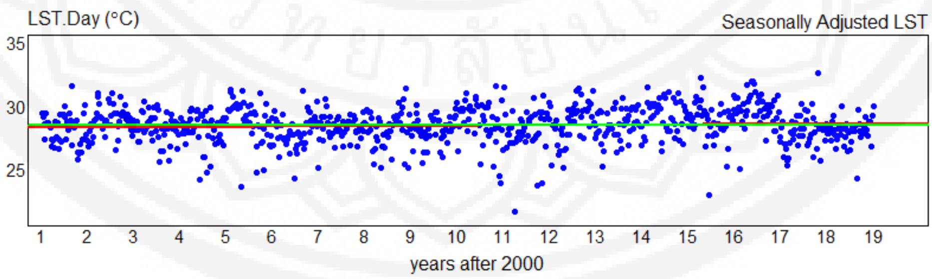

Figure 4 Trend in Day-LST obtained from $y(t) = 29.2500 + 0.0004t$ (red), showing a very small difference from the overall mean (green).

Predictions

It is always questioned whether other factors (besides time) could explain the temperature in the next time step. A very first thought and simple conjecture is the temperature at the previous time point. Then, the next question is how far back (with what time lag) that the previous temperature would contribute to the forecasted future temperature. This can be addressed with multiple linear regression (Allison, 1999, chapter 4, Venables, & Ripley, 2002, section 14.5) and with ARIMA (Venables, & Ripley, 2013, section 14.2) models. Before fitting the models, we split the data in conventional 70:30 proportions to training and testing sets. The models will be fit to the training set and tested for predictions in the testing set they have not "seen" before. The performance of the models will be measured by Root Mean Square Error (RMSE), namely, the square root of the mean of the squared error.

As detailed in Ripley and Venables, 2002, section 14.5, the multiple linear regression takes its factors (predictors) as another variables. For example, taking temperature as an outcome variable, its predictors could be, such as, time, normalized difference vegetation index (ndvi), elevation, or even its own time-lag variables. While the predictors of ARIMA(p,d,q) model could only be its own time-lag variables and/or the errors of time-lag variables. There are many literatures with their analysis on comparing ARIMA and regression model, see e.g., Kinney, 1978; Krä mer, 1986; Stergiou, 1991; Stergiou, Christou, & Petrakis, 1997; Ediger, Akar, & Uğurlu, 2006; Hassan, 2014; Miswan, Said, & Anuar, 2016; Murat, Malinowska, Gos, & Krzyszczak, 2018.

Multiple Linear Regression

As the time points are 8 days apart, we first consider a 4 step time lag (approximately one month) for the predictors used as inputs in multiple linear regression. It was found that only the lags 1, 3, and 4 were statistically significant and the model is

$$y(t) = 17.9643 + 0.1841\, y(t-1) + 0.1180\, y(t-3) + 0.0850\, y(t-4), \qquad (1)$$

with RMSE 1.3334. However, when the model (1) is used to predict the testing set it surprisingly performs a little better with RMSE 1.3249. The output values from the model (1) for both training and testing sets are depicted in Figure 5.

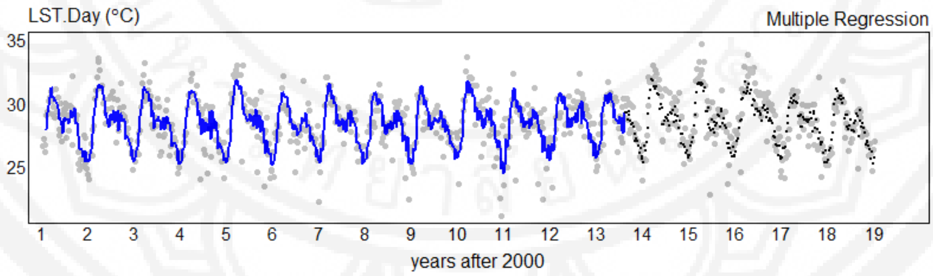

Figure 5 True Day-LST (grey dots) and values output by the model (1) over the training period (blue line) and the test period (black dots).

ARIMA (p,d,q)

It is confirmed by Figure 4 and the result from simple linear regression above that the seasonally adjusted Day-LST contains (almost) no trend and is mainly seasonal. In other words, the stationarity assumption is appropriate. The choice of (p,d,q) was made by ACF and PACF of the data. The ACF, Figure 6 (top) tails off while the PACF Figure 6 (botom) cuts off after lag 3, suggesting the model to be ARIMA(3,0,0). However, since the ACF does not perfectly tail off, several models according to various values of p, d, and q, from 1 to 3 are tested using RMSE as a forecasting accuracy criteria. It confirms that still the model ARIMA(3,0,0) or simply AR(3) is the best model. Moreover, the p-value from the Box-Ljung test is greater than 0.05 so the model does not show a lack of fit. The RMSEs in training and testing were 1.3248 and 1.3489, respectively.

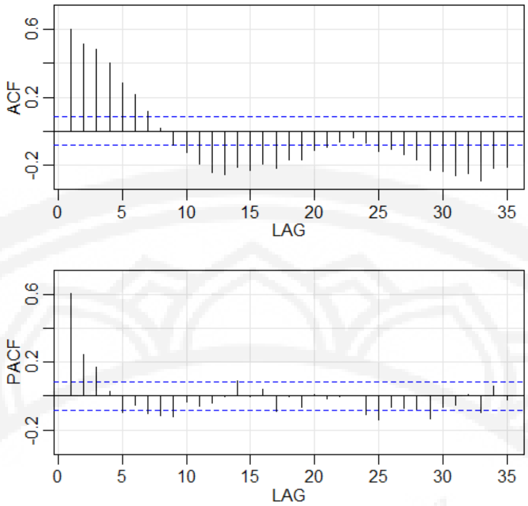

Figure 6 ACF (top) and PACF (bottom) of seasonally adjusted Day-LST up to lag 35.

In fact, the ARIMA (3,0,0) gives

$$y(t) = 18.8312 + 0.1871\ y(t-1) + 0.0456\ y(t-2) + 0.1248\ y(t-3). \qquad (2)$$

The outputs from ARIMA (3,0,0) are illustrated in Figure 7, where the magenta line and black dots show the fitted values in training period and forecasts in the testing period.

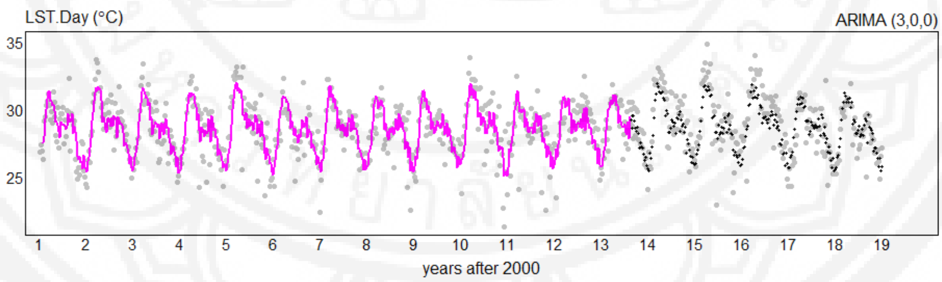

Figure 7 True Day-LST (grey dots) and outputs from the ARIMA model (2) over the training (magenta line) and testing (black dots) periods.

Discussion and Conclusion

Simple linear regression gave a positive trend with p-value less than 0.05, indicating that the temperature has increased and the increase is statistically significant. However, the increase of 0.3312 degrees Celsius in 18 years or 1.84 degrees Celsius in a century could be considered small. In other words, it can be said that the crop yields or the growing season period in Songkhla would not be affected by the Day-LST increase. In terms of forward prediction, it turns out that the multiple linear regression and ARIMA (3,0,0) both needed the same number of predictors with lags 1, 3, and 4, and lags 1, 2, and 3, respectively. Moreover, their performances did not differ much.

Acknowledgements

We would like to thank the Centre of Excellence in Mathematics, Commission on Higher Education (CHE), Ministry of Education, Ratchathewi, Bangkok, 10400 Thailand for supporting this project.

References

Allison, P. D. (1999). *Multiple regression: A primer.* Thousand Oaks, California: Pine Forge Press.

Bachelet, D., Brown, D., Böhm, M., & Russell, P. (1992). Climate change in Thailand and its potential impact on rice yield. *Climatic change, 21*(4), 347-366.

Ediger, V. Ş., Akar, S., & Uğurlu, B. (2006). Forecasting production of fossil fuel sources in Turkey using a comparative regression and ARIMA model. *Energy Policy*, *34*(18), 3836-3846.

Hassan, J. (2014). ARIMA and regression models for prediction of daily and monthly clearness index. *Renewable Energy*, *68,* 421-427.

Hatfield, J. L., & Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and climate extremes, 10,* 4-10.

Iizumi, T., & Ramankutty, N. (2015). How do weather and climate influence cropping area and intensity? *Global Food Security, 4,* 46-50.

Kang, Y., Khan, S., & Ma, X. (2009). Climate change impacts on crop yield, crop water productivity and food security–A review. *Progress in natural Science, 19*(12), 1665-1674.

Kinney Jr, W. R. (1978). ARIMA and regression in analytical review: An empirical test. *Accounting Review*, *51,* 48-60.

Krämer, W. (1986). Least squares regression when the independent variable follows an ARIMA process. *Journal of the American Statistical Association*, *81*(393), 150-154.

Limsakul, A., Kachenchart, B., Singhruck, P., Saramul, S., Santisirisomboon, J., & Apipattanavis, S. (2019). Updated basis knowledge of climate change summarized from the first part of Thailand's Second Assessment Report on Climate Change. *Applied Environmental Research, 41*(2), 1-12.

Miswan, N. H., Said, R. M., & Anuar, S. H. H. (2016). ARIMA with regression model in modelling electricity load demand. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *8*(12), 113-116.

Murat, M., Malinowska, I., Gos, M., & Krzyszczak, J. (2018). Forecasting daily meteorological time series using ARIMA and regression models. *International agrophysics*, *32*(2), 253-264.

Office of Commercial Affairs Songkhla. (2018). *Marketing Data in 2017.* Thailand: Songkhla.

Parry, M. L., & Carter, T. R. (1989). An assessment of the effects of climatic change on agriculture. *Climatic Change, 15*(1-2), 95-116.

Parry, M. L., Rosenzweig, C., Iglesias, A., Livermore, M., & Fischer, G. (2004). Effects of climate change on global food production under SRES emissions and socio-economic scenarios. *Global environmental change, 14*(1), 53-67.

Powell, J. P., & Reinhard, S. (2016). Measuring the effects of extreme weather events on yields. *Weather and Climate extremes, 12,* 69-79.

Sharma, D., & Babel, M. S. (2014). Trends in extreme rainfall and temperature indices in the western Thailand. *International journal of Climatology, 34*(7), 2393-2407.

Somparn, P., Gibb, M. J., Markvichitr, K., Chaiyabutr, N., Thummabood, S., & Vajrabukka, C. (2004). Analysis of climatic risk for cattle and buffalo production in northeast Thailand. *International journal of biometeorology, 49*(1), 59-64.

Stergiou, K. I. (1991). Short-term fisheries forecasting: comparison of smoothing, ARIMA and regression techniques. *Journal of Applied Ichthyology*, *7*(4), 193-204.

Stergiou, K. I., Christou, E. D., & Petrakis, G. (1997). Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods. *Fisheries Research*, *29*(1), 55-95.

Valizadeh, J., Ziaei, S. M., & Mazloumzadeh, S. M. (2014). Assessing climate change impacts on wheat production (a case study). *Journal of the Saudi Society of Agricultural Sciences, 13*(2), 107-115.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S-PLUS* (4th Ed.). New York: Springer-Verlag.