Statistical Model for Land Surface Temperature Change over Mainland Southeast Asia

Wongsai, N.,[1] Wongsai, S.,[3,4] Lim, A.,[1,2] McNeil, D.[1] and Huete, A. R.[5]
[1]Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, 94000, Thailand, E-mail: naong69@gmail.com, don.mcneil@mq.edu.au
[2]Centre of Excellence in Mathematics, Faculty of Science, Mahidol University, Bangkok 10400, Thailand E-mail: api_45@hotmail.com
[3]Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, 12120, Thailand, E-mail: sangdao.mathstat@gmail.com
[4]Thammasat University Research Unit in Data Learning, Thammasat University, 12120, Thailand
[5]School of Life Sciences, University of Technology Sydney, City Campus, Ultimo NSW, 2007, Australia E-mail: alfredo.huete@uts.edu.au

Abstract
This study presents an alternative statistical methodology for estimating changes in land surface temperatures over mainland Southeast Asia (SEA). The method comprises of seasonal adjusting and autocorrelation filtering of MODIS LST time series obtained from 2000 to 2019 at systematic 45 sample locations. Furthermore, the filtered seasonal-adjusted LST time series were estimated to quantify the decadal change of LST using linear regression model. The long-term dynamic of temperature change was revealed by curve fitting using a spline model with different knots. The overall LST changes in sub-regional and regional scale were estimated using multivariate regression model which adjusted for spatial correlation and aggregated information of LST change from all individual sample locations irrespective of their strength of statistical evidence (p-value). The final result showed that the surface temperature change in the SEA region increases by 0.126 °C/decade. 95% confident interval for increasing ranges between 0.04 to 0.21 °C/decade, which shows evidence of substantial warming surface in this region.

1. Introduction

The earth's surface has warmed by 0.7 – 0.9 °C/century since 1901, but the rate of warming has nearly double since 1975 to 1.5 – 1.8 °C/century (Sánchez-Lugo et al., 2018). Global warming varies substantially over the surface of the earth (Hansen and Lebedeff, 1987). The dynamic of surface temperature has been investigated and estimated by several climate science groups (Hansen et al., 2010) in a global and regional scale. Temperature variations either from satellite- or station-based are typically measured in terms of anomalies relative to some base period, which is one of the several scientific approaches to monitor global or regional temperature change. The estimated trend of temperatures, however, relatively depends on the specific base period, as well as spatially and temporally aggregating surface temperature over specific regions and periods. A baseline period is needed to be defined, and a few important items of concern need to be considered. (i) Choosing one reference period is critical and might affect the trend projection. Using a different reference period as the baseline may cause a change in the variance in a model ensemble, which is often used to estimate the uncertainty in the models. (ii) Although the entire study period can be used as the reference period to compare the estimated trends, the departure values from the average LST value with high variability, especially in remotely sensed data, may be subject to non-normal distribution of data. The extreme satellite-based LST values may be present in the observed data, especially collected from a tropical region with negative influences of atmospheric conditions such as cloud development (Segal et al., 1995 and Ray et al., 2003) and precipitation (Pielke, 2001). Moreover, many previous long-term LST trend studies minimize seasonal variation by focusing on specific periods within a year (e.g., month, growing season) and different times (e.g., day/night time) instead of explicitly accounting for seasonality (Wongsai et al., 2017). The seasonal effect can be mitigated by looking at individual months separately or simply transforming data into a yearly average. However, these approaches might not only substantially reduce the number of observed data, but can also omit the data within incomplete years. In this study, we demonstrate the alternative statistical methods to analyze the long-term change of LST from sample locations over mainland Southeast Asia. All statistical concerns on time series data analysis, including seasonal adjustment and spatiotemporal autocorrelation were addressed and eliminated. The trend of decadal LST was measured using linear regression estimation. Also, the long-term dynamic of surface temperature was represented by curve fitting based on spline function with the different number of knots and their placement.

2. Material and Methods

2.1 Study Area

The study area was the mainland Southeast Asia (SEA) including seven countries: Myanmar, Thailand, Laos, Vietnam, Cambodia, Malaysia, and Singapore, as shown in Figure 1(a). Figure 1(b) displays five sub-regions and the location of 45 sample points, where the LST data were acquired and analyzed. Each sub-region contain nine sample points. The sample locations were horizontally 210 km and vertically 340 km apart according to MODIS sinusoidal grid system. However, some sample points were horizontally distorted to ensure that they reside on the land. Note that one sample point in the second sub-region (in the top-right corner) was located in Yunnan province, China.

2.2 MODIS LST Time Series

Level-3 MODIS/Terra Land Surface Temperature and Emissivity product (MOD11A2) were downloaded from MODIS and VIIRS Land Products Global Subsetting website (ORNL DAAC, 2018). The Collection 6 MODIS LST/E product is an eight-day composite with a 1 km spatial resolution generated from the daily MOD11A1 product by a simple average of LST values over eight days period (Wan and Hulley, 2015). The acquired data contain daytime and nighttime LST (10:00 and 22:00 local over-pass time). However, only daytime LST dataset was used in this study as the example of modeling LST dynamic over the regional scale. As shown in Figure 1(b), there were 45 sample points from which the LST data were obtained, and each sample point covers 7 by 7 LST pixels. LST time series for each pixel consists of 19 years of data observation started from 2000 to 2019, which is 874 observations in total.

2.3 Seasonally Adjusted and Time Independent LST Series

Before modeling long-term LST change, the acquired 45 LST time series were merely averaged by the day of observations to account for spatial heterogeneity of the land surface over 49 km^2 at the sample point.

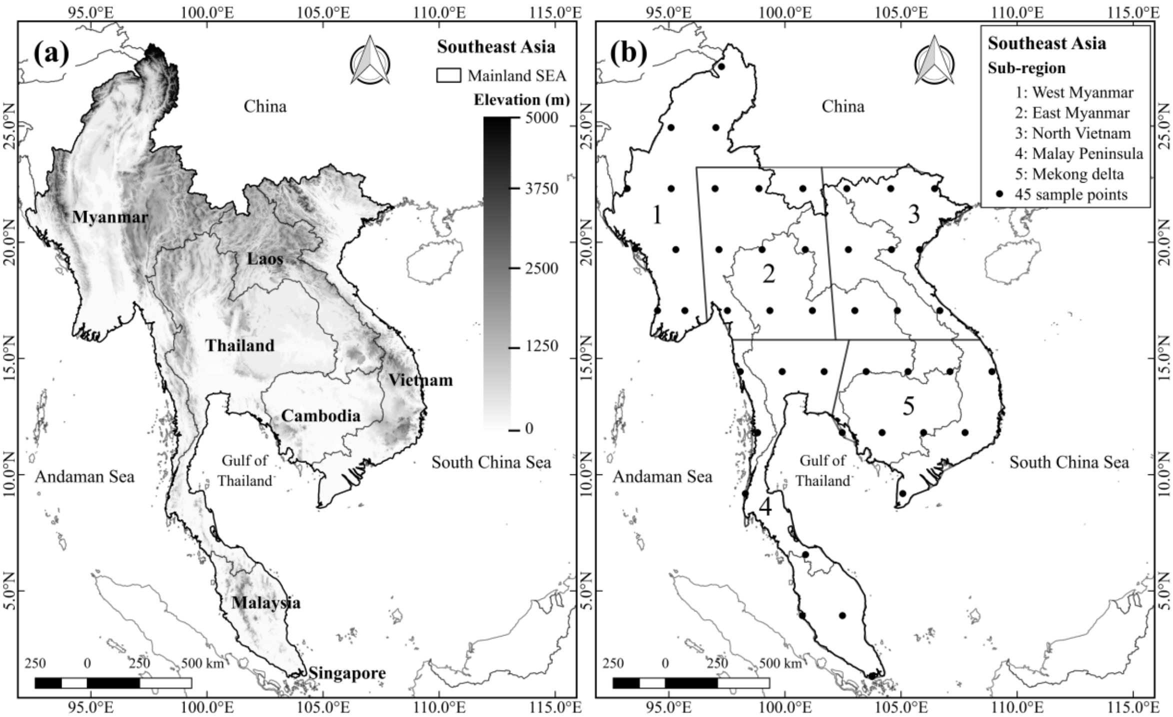

Figure 1: (a) Reginal map of mainland Southeast Asia extended the study area, which covers seven countries. (b) Map of 45 LST data sample points and five sub-regions of the study area

Thus, the spatially smoothed LSTs represented the LST series at a particular sample location. LST time series, in general, consist of seasonal, trend and remainder (or noise) components. The seasonality usually has an enormous influence on long-term LST, resulting in fluctuation and variation in observed surface temperature values. The seasonal effect in LST time series needs to be adjusted before the future trend analysis. We applied the method of seasonal extraction using cubic spline function with annual periodic boundary condition (Wongsai et al., 2017) to detach seasonality from LST time series. A set of eight knots (at Julian day 10, 40, 80, 130, 240, 290, 330 and 360) was applied for extracting seasonality from all 45 spatially smoothed LST time series. However, instead of fitting the derived coefficients from cubic spline function to the LST data using weighted least squares (WLS) regression with a weight matrix constructed from the quality control of LST data, we fitted ordinary least squares (OLS) linear regression model for simplicity. The adjusted r-squared (R^2), which measures the goodness of seasonal component fitted to raw LST data, were reported. Moreover, the seasonally adjusted LST was determined by subtracting the LST series from the seasonal component and then adding the averaged seasonal component.

In time series analysis, correlations in errors from a fitted model are assumed to be independent and identically distributed or stationary. This assumption means that the correlation between these errors at two specified time points depends only on the number of time points separating them (known as lag). Regression-based analysis models usually rely on the assumption that there is no autocorrelation in the residuals because the models often use this error term to assess the goodness of fit of the model. In general, these correlations affect the standard errors of the estimated parameters and need to be considered when fitting the model. Before the trend analysis of seasonally adjusted LST ($LST_{s.adj}$), the autocorrelation of $LST_{s.adj}$ time series was examined using an autoregressive integrated moving average (ARIMA) model (Asteriou and Hall, 2011), which is a generalization of an autoregressive (AR) process. In this study, non-seasonal ARIMA model was used to eliminate the non-stationary (Gagniuc, 2017) of the error term in $LST_{s.adj}$ time series. The non-seasonal ARIMA models are generally denoted as ARIMA(p,d,q) where parameters p is the number of time lags in the autoregressive model, d is the order of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average process, and all three parameters are non-negative integers.

Coefficients of the AR (p) model with different time lags and their standard error (SE) were examined. Subsequently, the $LST_{s.adj}$ time series were filtered using fitted residuals from the selected ARIMA (p,0,0) model with p that significantly accounts for autocorrelated error in the $LST_{s.adj}$ time series. The filtered seasonally adjusted LST can be calculated using the following mathematical formula:

$$\text{Filtered LST}_{s.adj} = \left[\left(\sigma\left(\text{LST}_{s.adj}\right)/\sigma(\epsilon)\right) \times \epsilon\right] + \mu\left(\text{LST}_{s.adj}\right)$$

Equation 1

where,
ϵ = Fitted residual from ARIMA model
σ = Population standard deviation

The additional constant term is the mean of the seasonal-adjusted LST. Moreover, the extreme LST values, which exceed three times the standard deviation from the mean in filtered seasonal-adjusted LST distribution, were removed to ensure normality of data distribution.

2.4 Decadal Trend Analysis

2.4.1 Linear and Spline model

After the influence of seasonality and time correlation in LST time series were eliminated, the OLS estimation then was used to simply fit the stationary seasonal-adjusted LST time series. The simple regression model, which in the form:

$$Y_i = \alpha + \beta x_i + \epsilon_i$$

Equation 2

For x as the time of observation at i = 1, 2,…, n, estimates α and β by minimizing the sum of squared deviations observed from the expected response between time x_1 to x_n (Δx) in LST time series. Consequently, we can obtain the quantity of temperature change (ΔLST) during the Δx period. Also, the adjusted R^2 and p-value were reported. Not only the linear trend was used to estimate the LST change in a decade, but also the polynomial trending, which describes a curved pattern in data, was used to expose the acceleration and the cycle of the LST trend. We used a 2nd derivative spline function (Equation 3) with (i) four knots placed at Julian day of 1, 276, 552, and 828 and (ii) seven knots placed at Julian day of 1, 138, 276, 414, 552, 690, and 828 to estimate the acceleration in square of degrees Celsius ($°C^2$) per decade and to demonstrate the cyclic pulse of surface temperature, respectively.

$$s(t) = c_0 + c_1 t + \sum_{j=1}^{m} c_{j+1} \left(t - T_j\right)_+^2$$

Equation 3

where,
t_+ = Time (t) if $t > 0$ and 0 otherwise
T_j = Specified knots from $j = 1, 2, \ldots, m$
$\left(t - T_j\right)_+^2$ = Derived spline coefficients

Note that a spline function is a piecewise continuous linear function with discontinuities in its derivative at each knot. Once again, the linear regression model was used to fit the derived coefficients of the spline function to the filtered $LST_{s.adj}$ data. We simply took the filtered $LST_{s.adj}$ data as the outcome and the functions t, $(t - T_1)_+^2$, $(t - T_2)_+^2$ … $(t - T_m)_+^2$ as predictors. The curves were then determined and the acceleration of the four knots spline curve was calculated using Equation 4 and Equation 5, respectively.

$$Y = b_0 + b_1 t + c_1 (t - T_1)_+^2 + \ldots + c_i (t - T_i)_+^2$$

Equation 4

where,
i = 1, 2, …, m-2 (m is the number of knots)
b_0 = Regression coefficients of intercept
b_1 = Regression coefficients of observation time
c_i = Derived spline coefficients

$$3 \times c_1(T_3 - T_1) + 3 \times c_2(T_4 - T_2) \times (T_3 - T_2)/(T_4 - T_1)$$

Equation 5

Finally, the curvature lines were plotted for inspection along with the linear trend and their p-value.

2.4.2 Aggregating results

To illustrate the long-term change of surface temperature at each sample point, the decadal ΔLSTs from the linear fitting of filtered $LST_{s.adj}$ time series were presented in both absolute value of temperature change and the z-score. The z-score, in this case, represents the normalized measurement of temperature change without the actual unit which expressed in terms of how many standard deviations from the mean of observed surface temperature. Meanwhile, combining trends from filtered $LST_{s.adj}$ at individual sample point within the same sub-region were reported for the aggregating change of temperature in each sub-region. However, substantial correlations exist between the residual surface temperatures at the sample points within the sub-region. The high correlations between residuals in the spacing sample points suggest that aggregating data can make the stronger conclusions about global warming patterns from sample points and fitting models to the data in a larger area such as sub-regional or regional scale. In this study, multivariate linear regression model (Equation 6) and the standard error of the mean of the multivariate outcomes (Equation 7) are appropriate because they enable fitted parameters for different outcomes to be compared in the presence of such special correlations.

$$f_{ij} = \alpha_i + \beta_j x_i$$

Equation 6

where,
f_{ij} = filtered seasonally-adjusted LST at observation i in sample point j
x_i = observation time elapsed with i = 1, 2, …, n (n = 874: number of observations)
j = 1, 2, …, p (p = 9: members of sample point in a sub-region).

$$SE(mean) = \sqrt{\textstyle\sum_{i,j=1}^{n,p}} V_{ij} / n$$

Equation 7

where, V_{ij} = Variance-covariance matrix of the estimates.

These models provide variance-covariance matrices of estimated temperature increases at the sample points, thus giving confidence intervals for linear combinations of increases in different sample points. Note that estimates of parameters and their standard errors given by multivariate linear regression models are identical to those given by corresponding separately-fitted univariate models which only their covariances are different. But these covariances are needed to obtain accurate confidence intervals for regional trends. However, any row in the response matrix f_{ij} containing a missing value is omitted in multivariate regression model. Thus, the missing LSTs were filled with the fitted values from the regression of filtered $LST_{s.adj}$.

3. Results and Discussions

3.1 Seasonal Patterns

Figure 2 illustrates the seasonal pattern of filtered $LST_{s.adj}$ time series at nine individual sample points in the second sub-region, which covers eastern of Myanmar, northern Thailand, and some part of Laos and China. The lines reveal the seasonal components and plus symbols indicate knots position.

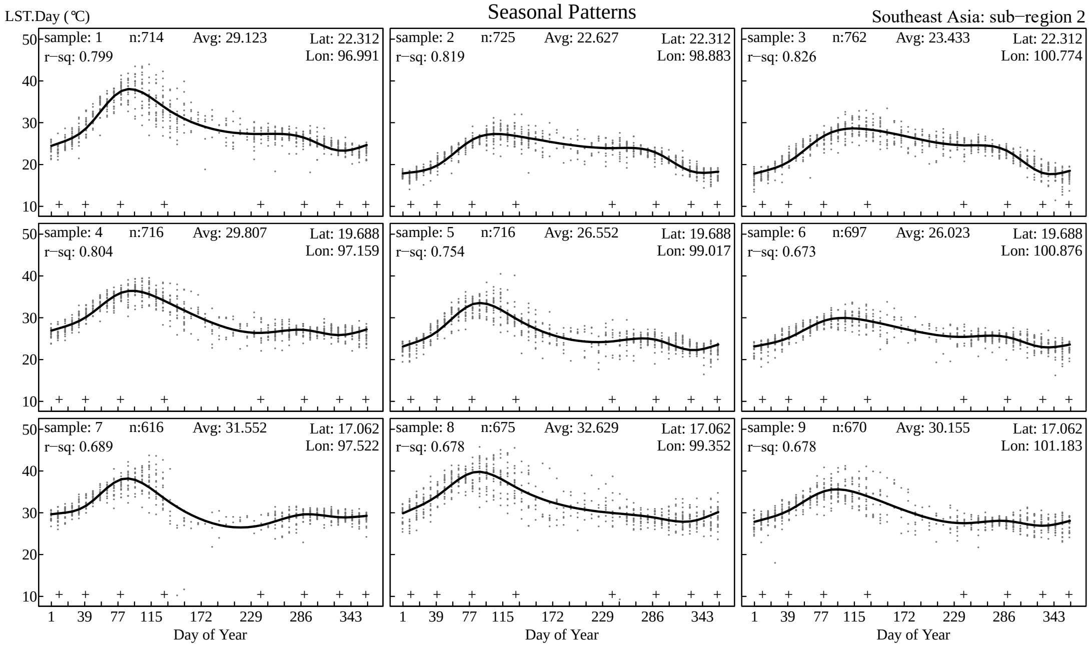

Figure 2: Examples of LST seasonality obtained from nine sample points in the second sub-region

Considering inland area as an example, the mean of LST seasonality in locations at higher latitude was lower than in locations that close to the equator and the high temperature were found between February to May which is the local dry season in this region. However, when looking over the entire study area we found that there was the undistinguishable dry-wet seasonal pattern in the locations that were in closer proximity to the sea. The seasonal patterns of LST in the Malay peninsula, which receives the direct influence from the tropical monsoon climate, have less seasonal variation than from inland SEA, as shown in Figure 4(a). The local tropical weather, also the thermal properties of water surface, may affect the variation of the observed LST that causing low adjusted R^2 in seasonality modeling. Moreover, the magnitude of LST seasonality and its escalation rate at the beginning of the dry season depended on the land cover where the urban land cover has higher average temperature and rising rate (Wongsai et al., 2017).

Although we performed seasonality extraction without considering the quality of remotely retrieved LST data, we found that there was insignificant difference of seasonal patterns between seasonality extraction involving the quality of obtained LST data or not. Despite the ignorance of negative impacts of the variation of surface temperature in various land covers and applying a global set of knots placement, the extracted seasonal components acceptable represented the seasonality for each sample LST time series.

3.2 Long-term LST Trend

Seasonally adjusted LST time series were determined after the process of seasonality extraction, and then temporal correlation was detected and removed from the data series. We found that autoregressive model with time lags as 2 significantly accounts for auto-correlated error in most of the sample $LST_{s.adj}$ time series and thus, the fitted residuals from ARIMA model were used to filter the $LST_{s.adj}$ time series. Figure 3 illustrates the fitted models of nine examples of filtered seasonal-adjusted LST from second sub-region using a simple linear regression model and spline function with equally interval four and seven knots. The "r-sq0" and "r-sq" labels indicate adjusted R^2 from sample linear regression and 7-knot spline models, respectively. Graphs in the last column illustrate a comparison of linear and curved LST trends from three sample points aligned on the same latitude. The pink dots indicate extreme LST values excluded from model fitting. The regression model revealed the estimated amount of surface temperature change in 10-year period. The results from example LSTs in the second sub-region show that most of the LSTs increased with different degrees, whereas the highest increasing temperature (0.574 °C/decade) was found in sample point 5. However, only sample point 5, 6, and 9 have statistically significant increasing temperature (p-values < 0.05). For these three locations, the curvature lines fitted by 4-knot spline model represent more actual forecasting trend which the LSTs have decelerated trends where their linear trends were positive.

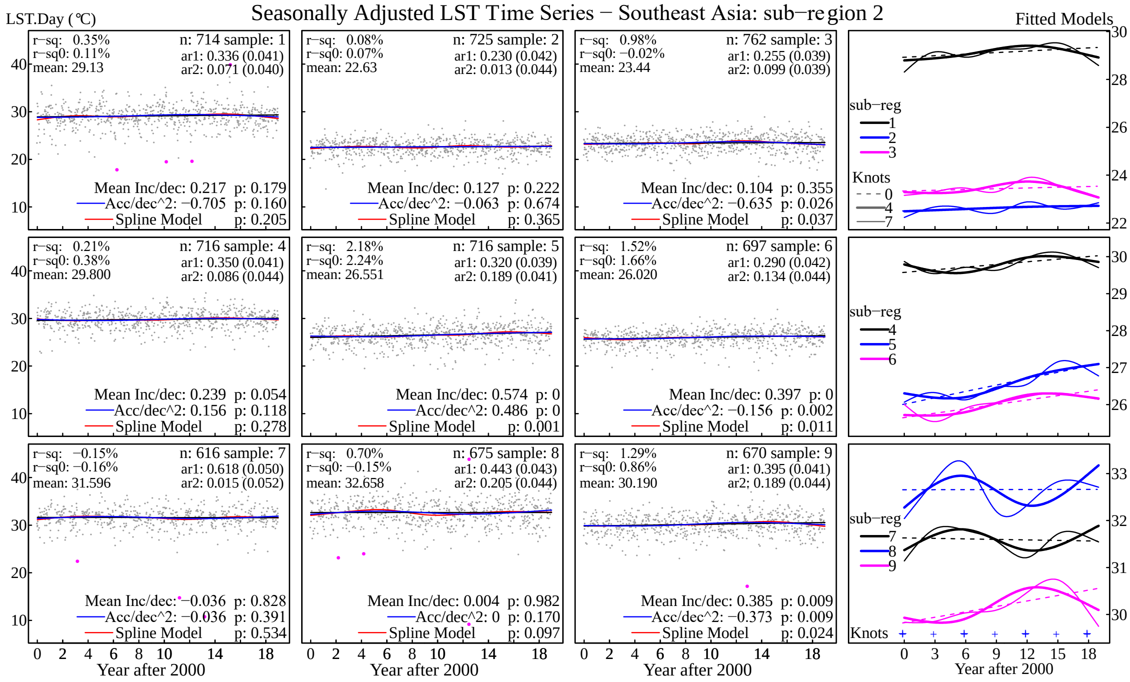

Figure 3: Examples of the fitted linear and curved LST trends using linear regression and spline function with four and seven knots for nine filtered season-adjusted LST time series in sub-region 2

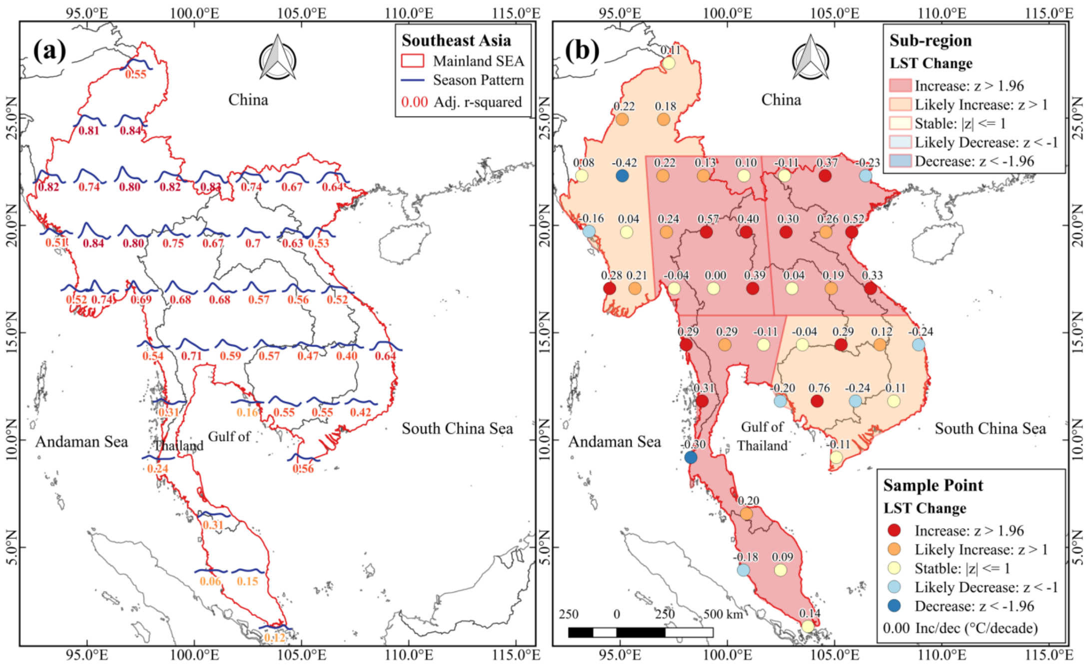

Figure 4: (a) Seasonal pattern of all 45 sample LST time series. (b) The estimated decadal trend of LST at sample points and within sub-regions, which represented as how many standard deviations above or below the normal population mean (z-score)

Moreover, the 7-knot spline model shows evidence of corresponding cyclic pulse. The cycle dynamic of LST provides the indication of other influence that could fluctuate the long-term surface temperature, which needs to be considered. The change of LST for the entire study area presents in Figure 4(b). The ΔLST was varied when considering at individual sample point. Haft of the LST series (22 out of 45 sample points) was more likely increases (z > 1) and 13 LST series were stable (|z| <= 1). Thus, the surface temperatures of all five sub-regions were likely increased when combines information concerning LST increase from all LST time series irrespective of their individual *p*-values. The sub-regional LST increase 1.09, 3.46, 2.50, 2.00, and 1.05 °C/decade for sub-region 1 to 5, respectively. When looking at all 45 sample points on mainland Southeast Asia, the multivariate regression model estimated the increasing of regional LST for 0.126 °C/decade with SE at 0.043 °C/decade. Subsequently, 95% confident interval for this increasing was 0.04 to 0.21 °C/decade. This confident interval range above zero, which provides evidence of substantial warming in this region.

4. Conclusion

The results from successive statistical analysis of example MODIS LST time series presented in this study disclosed the dynamics of LST change and provided evidence of earth surface warming in the study area. The statistical modeling involved (i) seasonal adjusting of the obtained LST time series using cubic spline function with annual periodic boundary condition (Wongsai et al., 2017), (ii) disposing of autocorrelation using data filtering, (iii) estimating of decadal change of time independent seasonally adjusted LST using simple linear regression model, and (iv) gauging the long-term dynamic using quadratic spline models. The magnitudes of temperature change varied in both positive and negative trends, which might depend on the land cover at individual location of the acquired LST data. The curve fitting exposed the cyclic pulse and the acceleration/deceleration of the LST trend. At the sub-regional and SEA regional scale, the multivariate regression model that adjusts for spatial correlation gives the overall LST change, which increased between 0.04 to 0.21 °C/decade at 95% confident interval. The demonstrated statistical methods of LST time series analysis can be applied to the larger area and more density of sample to estimate the changes in land surface temperatures in the global scale with the strength of statistical evidence.

Acknowledgments

This research had been supported by Scholarship Awards for Thai Ph.D. Students under Thailand's Education Hub for Southern Region of ASEAN Countries and Overseas Thesis Research Scholarship from Graduate School, Prince of Songkla University attributed to Noppachai Wongsai. This study was partially funded by the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand, and Thammasat University Research Unit in Data Learning. The constructive comments and suggestions from the anonymous reviewers are very gratefully acknowledged.

Reference

Asteriou, D. and Hall, S. G., 2015, *ARIMA Models and the Box–Jenkins Methodology*. *In Applied Econometrics*, 3rd edition, (Red Globe Press).

Gagniuc, P. A., 2017, *Markov Chains: From Theory to Implementation and Experimentation*, 1st edition, (New Jersey: John Wiley & Sons).

Hansen, J. E. and Lebedeff, S., 1987, Global Trends of Measured Surface Air Temperature. *Journal of Geophysical Research*, Vol. 92, 13345-13372.

Hansen, J. E., Ruedy, R., Sato, M. and Lo, K., 2010, Global Surface Temperature Change. *Reviews of Geophysics*, Vol. 48, RG4004.

ORNL DAAC, 2018, MODIS and VIIRS Land Products Global Subsetting and Visualization Tool. ORNL DAAC, Oak Ridge, Tennessee, USA. Accessed July 27, 2019. Subset obtained for MOD11A2 product at 14.438N, 98.0759E, time period: 2000-02-18 to 2019-07-12, and subset size: 7 x 7 km.

Pielke, R. A. Sr., 2001, Influence of the Spatial Distribution of Vegetation and Soils on the Prediction of Cumulus Convective Rainfall. *Reviews of Geophysics*, Vol. 39, 151-177.

Ray, D. K., Nair, U. S., Welch, R. M., Han, Q., Zeng, J., Su, W., Kikuchi, T. and Lyons T. J., 2003, Effects of Land Use in Southwest Australia: Observations of Cumulus Cloudiness and Energy Fluxes. *Journal of Geophysical Research*, Vol. 108(D14), https://doi.org/10.102-9/2002JD002654.

Sánchez-Lugo, A., Morice, C., Berrisford, P. and Argüez, A., 2018, *State of the Climate in 2017*, *Bulletin of the American Meteorological Society*, Vol. 99(8), S11-S13.

Segal, M., Arritt, R. W., Clark, C., Robin, R. and Brown, J., 1995, Scaling Evaluation of the Effect of Surface Characteristics on Potential for Deep Convection Over Uniform Terrain. *Monthly Weather Review*, Vol. 123, 383-400.

Wongsai, N., Wongsai, S. and Huete A. R., 2017, Annual Seasonality Extraction Using the Cubic Spline Function and Decadal Trend in Temporal Daytime MODIS LST Data. *Remote Sensing*, Vol. 9(12), 1254. https://doi.org/10.3390/rs9-121254.

Z. Wan, S. H. and Hulley, G., 2015, *MOD11A2 MODIS/Terra Land Surface Temperature/ Emissivity 8-Day L3 Global 1km SIN Grid V006*. NASA EOSDIS Land Processes DAAC.