Journal of Engineering and Applied Sciences 11 (11): 2387-2395, 2016
ISSN: 1816-949X

Cubic Spline Regression Model and Gee for Land Surface Temperature Trend Using Modis in the Cloud Forest of Khao Nan National Park Southern Thailand During 2000-2015

Anusa Suwanwong and Noodchanath Kongchouy
Department of Mathematics and Statistics, Faculty of Science, Prince of Songkla University,
90110 Songkhla, Thailand

Abstract: Land Surface Temperature (LST) is an important key in climatological and environmental studies. Cubic spline is piecewise polynomials with continuous function and the most successful approximating function. Generalized Estimated Equation (GEE) is used to estimate the parameters of a generalized linear model with longitudinal and other correlated response data. The aim of this study was to use cubic spline regression and GEE to perceive pattern and variation of temperature at Khao Nan during 2000-2015. We downloaded data, for Land Surface Temperatures recorded (LST) by MODIS EARTH Satellites from 2000-2015 in square kilometres grid boxes covering Khao-Nan National Park. The results showed cubic spline regression gave excellent curve fitting for pattern of LST among day time from satellite image at Khao-Nan National Park. Also, cubic spline regression and GEE showed the temperature change around Khao-Nan National Park during 2000-2015 had similar pattern with increasing variation except 2005-2009. In conclusion, cubic spline regression and GEE available to perceive pattern and variation of temperature very well.

Key words: Land surface temperature, cubic spline, regression, GEE, cloud forest, Thailand

INTRODUCTION

Land Surface Temperature (LST) is an important climate variable which is related to surface energy balance and the integrated thermal state of the atmosphere within the planetary boundary layer (Jin and Liang, 2006). The Information of LST was provided by satellite remote sensing. Satellite remote sensing provides efficiency spatially continuous information on near-surface environmental conditions. The information from satellite remote sensing is collected data on environmental variables such as climatic and atmospheric radiation and chemistry, ocean dynamics and productivity properties. The most important is the spatial heterogeneity of satellite observations across large surface area with strong correlation (Prihodko and Goward, 1997). Moderate Resolution Imaging Spectroradiometer (MODIS) consists of Terra and Aqua satellites. MODIS provides data for land cover maps that tell scientists not only whether an area is vegetated but also what kind of vegetation is growing there, separating coniferous forests from deciduous forests or cropland from grassland. MODIS satellite system is very useful in monitoring the situation in the region and the resolution of data is 250-1000 m (Wan and Li, 1997). It can analyze and process several information with properly problems.

In previous studies, linear regression analysis is the most frequently used in climate studies to show trend and pattern of temperature (Krishna, 2014; Brunetti *et al*., 2000). Linear regression model provides a good estimate of daily minimum air temperature during the night while the measured daily maximum Ta and Terra daily land surface temperature daytime is significant during growing season with medium correlation coefficient (Zhu *et al*., 2013). In Hat Yai City, Thailand linear regression has been used to investigate intensity and pattern of land surface temperature (Ruthirako *et al*., 2014). Multiple linear regression were developed and tested the model for maximum and minimum temperature; for Tmax, Terra-night was a better explanatory variable than Aqua-night while Aqua-night provided a better estimation of Tmin than Terra-night (Zeng *et al*., 2015). Random forest models can effectively estimate air temperature over complex terrain regions comparing with multiple linear b regression (Xu *et al*., 2014). Linear regression and time series show spatial distributions of maximum temperature trends in most of the Himalayan region and the middle mountains (Shrestha *et al*., 1999). Cubic spline regression have various used in study of demographic (McNeil *et al*., 1977; Smith *et al*., 2004), medical (Sala *et al*., 2009; Desquilbet and Mariotti, 2010) and community (Shipley *et al*., 2006). GEE is commonly used in large

Corresponding Author: Noodchanath Kongchouy, Department of Mathematics and Statistics, Faculty of Science, Prince of Songkla University, 90110 Songkhla, Thailand

biological studies for application of statistical analysis to (Xie and Paik, 1997; Shih, 1997; Spiess and Hamerle, 1996).

The cloud forest is a rare resource. There are about $380,000^2$ cm or 0.26% around the world. Asia and some country which is close to the equator such as Indonesia, Malaysia and Thailand are found to have enormous cloud forest. In Thailand, except the Doi Inthanon National Park, Khao Nan, Nakhon Si Thammarat.

The objective of this study was used spline cubic regression and GEE for fitting pattern and variation of temperature in the cloud forest of Khao Nan National Park during 2000-2015 using remote sensing data (MODIS). We selected the window size as 10×10 km. Data LST is variable which is used for analyzing data this study.

MATERIALS AND METHODS

Study area: The study area is a part of Khao Nan National Park which is the central area from 8.41'N-8.58'N and 99.56'E-99.74'E and has a total land area of 441^2 cm. The highest peak is about 1,438 m above sea level has 3 highest peaks in this region, namely Khao Nan Yai, San Yean and Khao Tao (Watanasit *et al.*, 2008) (Fig. 1).

Data set and definitions: The satellite data used in this study are Aqua/MODIS data from 2000-2015. MODIS land products, 8, day 1 km land surface temperature product (MOD11A2) were collected from Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC). MOD11A2 provides daytime and night-time LST with 690 observations over 15 year.

LST is the radiative skin temperature of ground. It depends on the albedo, the vegetation cover and the soil moisture. LST influences the partition of energy between ground and vegetation and determines the surface air temperature (Liang *et al.*, 2013).

Preliminary data analysis and sample size: For preliminary data analysis, the outcome variables of interest are LST. Day 1 km (Land Surface Day temperatures). The outcome variable is stored as integers. The temperatures are degrees Kelvin and subtracting 273.15 gives corresponding degrees Celsius. Zeroes correspond to missing data due to insufficient measurement quality. The days are the same for each year has 46 different days from 1-361 every. So there are 690 observations over 15 year. For these temperature outcomes, the data value corresponding to a specified day is created by selecting the day from the 8 day period; when the outcome variable has maximum measurement quality. We code these values as NA when no measurement is recorded at any pixel it means that no measurement in the region in the 8-day period has sufficient measurement quality. The missing data of LST. Day is 1.16% during 15 year.

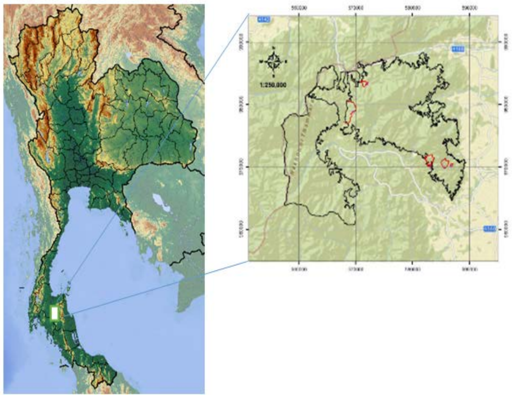

Fig. 1: Map of Khao Nan National Park

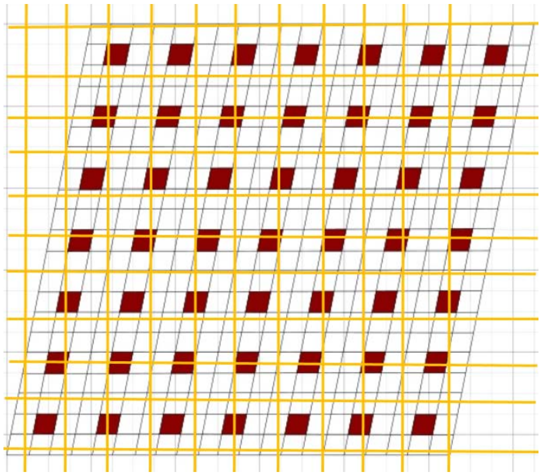

Fig. 2: The sample size of this region

For sample size, we selected grid box 3 cells away to reduce correlation of data. Figure 2 shows each of the 441 1 km^2 pixels in the region, the recorded temperatures can plot by the satellite by the day of the year. These graphs show temperature patterns. When fitting models to these graphs, spatial correlation between sampled pixels is reduced by selecting a sample of 49 from the 441 pixels, separated as far apart as possible in the region.

Statistical analysis: The most popular approximating functions in mathematics is spline function. Spline function is defined as piecewise polynomials of degree n. Knots and the first n-1 derivatives are condition of this function (Wold, 1974). Three aspects are important that is data have to be a set of pairs in real numbers which the values of y should be as measured values, model is represented by mean of continue function based on available data and a criterion of goodness of fit measures the distance of the function to be chosen (Biebler and Wodny, 2013). The data will be fitted by spline cubic regression and the end of any year is followed by the beginning of the next year. Equation 1 for a cubic spline function is:

$$S(t) = a + bt + \sum_{k=1}^{p} c_k (t - t_k)^3_+ \qquad (1)$$

where t denotes time, $t_1 < t_2 < ... < t_p$ are specified knots and $(t\text{-}x)_+$ is t-x for t>x and 0 otherwise. Time series is simply a sequence of numbers collected at regular intervals over a period of time. Mostly, time series methods were used for problems in the physical, environmental sciences and meteorological. This method can be very useful to see how interesting variables changes over the same time period time (Shumway and Stoffer, 2000). The LST data was plotted as time series to perceive their pattern in all of selected pixels over 15 year period time. GEE is a common method which is used to estimate the parameters. When the data has correlation from longitudinal and clustered studies (Hanley *et al*., 2003). Those data was applied by GEE to percieve variation of temperature every 5 year during 2000-2015. Those methods were analyzed by R program version 3.22. GEE was also implemented in R statistical software, using 'geepack' package.

RESULTS AND DISCUSSION

Temperature pattern and variation: Figure 3 shows the temperature pattern for day in a 1 km^2 pixel around latitude 8.4968 and longitude 99.645 for 15-year (June, 2000 to April, 2015). The number of non-missing observations in a day varies averaging 74.64%.

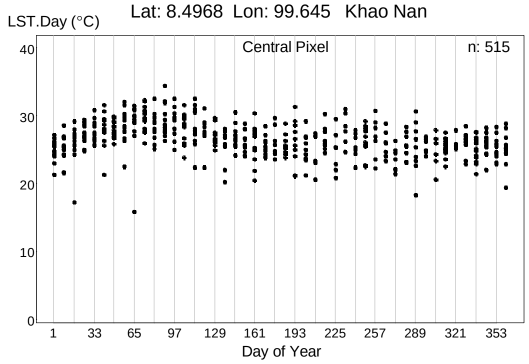

Fig. 3: The temperature pattern for day at central pixel

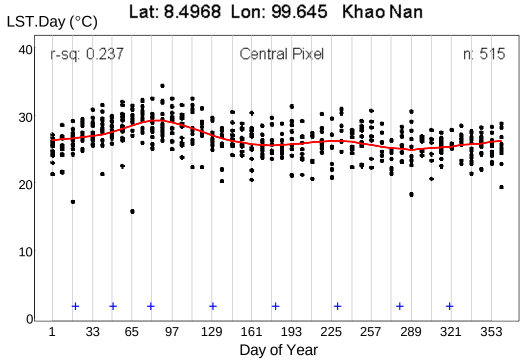

Fig. 4: Temperature pattern is fitted by spline function for each day of year

We fitted cubic spline function by using linear least square regression. The model should provide temperature patterns for each day of the year and this suggests that it should be defined over continuous time. Cubic spline function is given that the end of any year is followed by the beginning of the next year. There are eight knots at days 20, 50, 80, 130, 180, 230, 280 and 320 in fitting the model. The model should be a smooth periodic function with boundary conditions ensuring smooth periodicity. From Eq. 1, The boundary conditions require that s(t) for $t<t_1$ equals s(t) for $t>t_p$ One way of ensuring this is to put b = and make $\sum_{k=1}^{p} c_k (t - t_k)^3_+$ vanish for $t > t_p$, so that:

$$\sum_{k=1}^{p} c_k = 0, \sum_{k=1}^{p} c_k t_k = 0, \sum_{k=1}^{p} c_k t_k^2 = 0, \sum_{k=1}^{p} c_k t_k^3 = 0$$

(Biebler and Wodny, 2013). Adding, these four constraints, the spline function becomes as Eq. 2:

$$s(t) = a + \sum_{k=1}^{p-4} c_k \begin{bmatrix} (t - t_k)^3_+ - a_k (t - t_{p-1})^3_+ + \beta_k (t - t_p)^3_+ \\ -\gamma_k (t - t_{p-1})^3_+ + \delta_k (t - t_p)^3_+ \end{bmatrix}$$

Where:

$$\alpha_k = \frac{(t_p - t_k)(t_{p-1} - t_k)(t_{p-2} - t_k)}{(t_p - t_{p\text{-}3})(t_{p-1} - t_{p-3})(t_{p-2} - t_{p-3})}$$

$$\beta_k = \frac{(t_p - t_k)(t_{p-1} - t_k)(t_{p-3} - t_k)}{(t_p - t_{p\text{-}2})(t_{p-1} - t_{p-2})(t_{p-2} - t_{p-3})}$$

$$\gamma_k = \frac{(t_p - t_k)(t_{p-2} - t_k)(t_{p-3} - t_k)}{(t_p - t_{p\text{-}1})(t_{p-1} - t_{p-2})(t_{p-2} - t_{p-3})}$$

and:

$$\delta_k = \frac{(t_{p-1} - t_k)(t_{p-2} - t_k)(t_{p-3} - t_k)}{(t_p - t_{p\text{-}1})(t_p - t_{p-2})(t_p - t_{p\text{-}3})}$$

Figure 4 shows that the spline function tracks the temperature pattern quite well, although the variation for different years is relatively high and the value for the R^2 is low as 0.237 or 23.7%. Figure 5 and 6 show temperature pattern in all of 49 selected pixels around Khao Nan. Those patterns are similar with higher temperatures in March and August. On pixel 14 averaging 52.46% is the least number of non-missing observations.

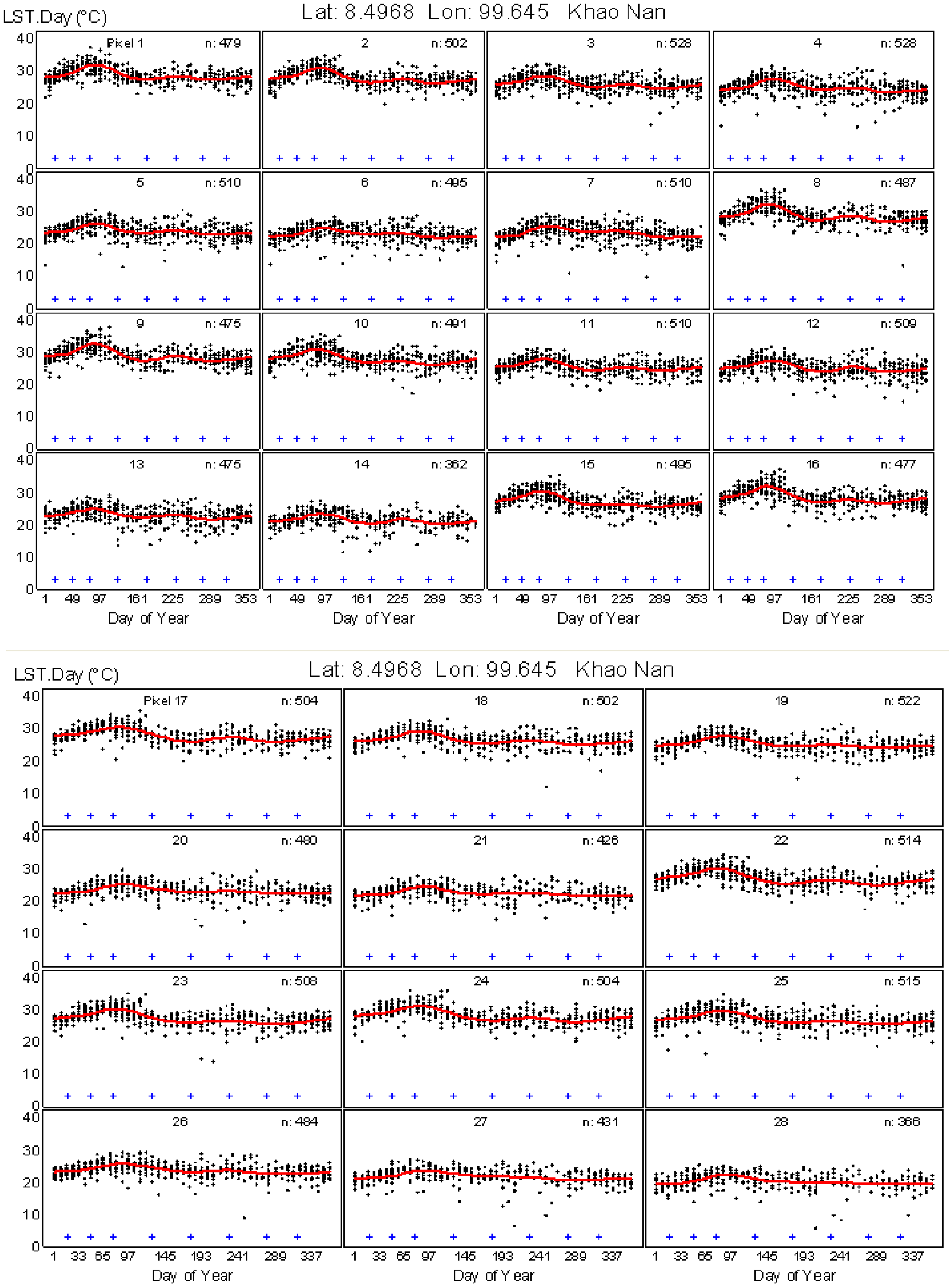

Fig. 5: Temperature pattern for LST Day of pixel 1-11 (Upper) and 17-28 (Lower)

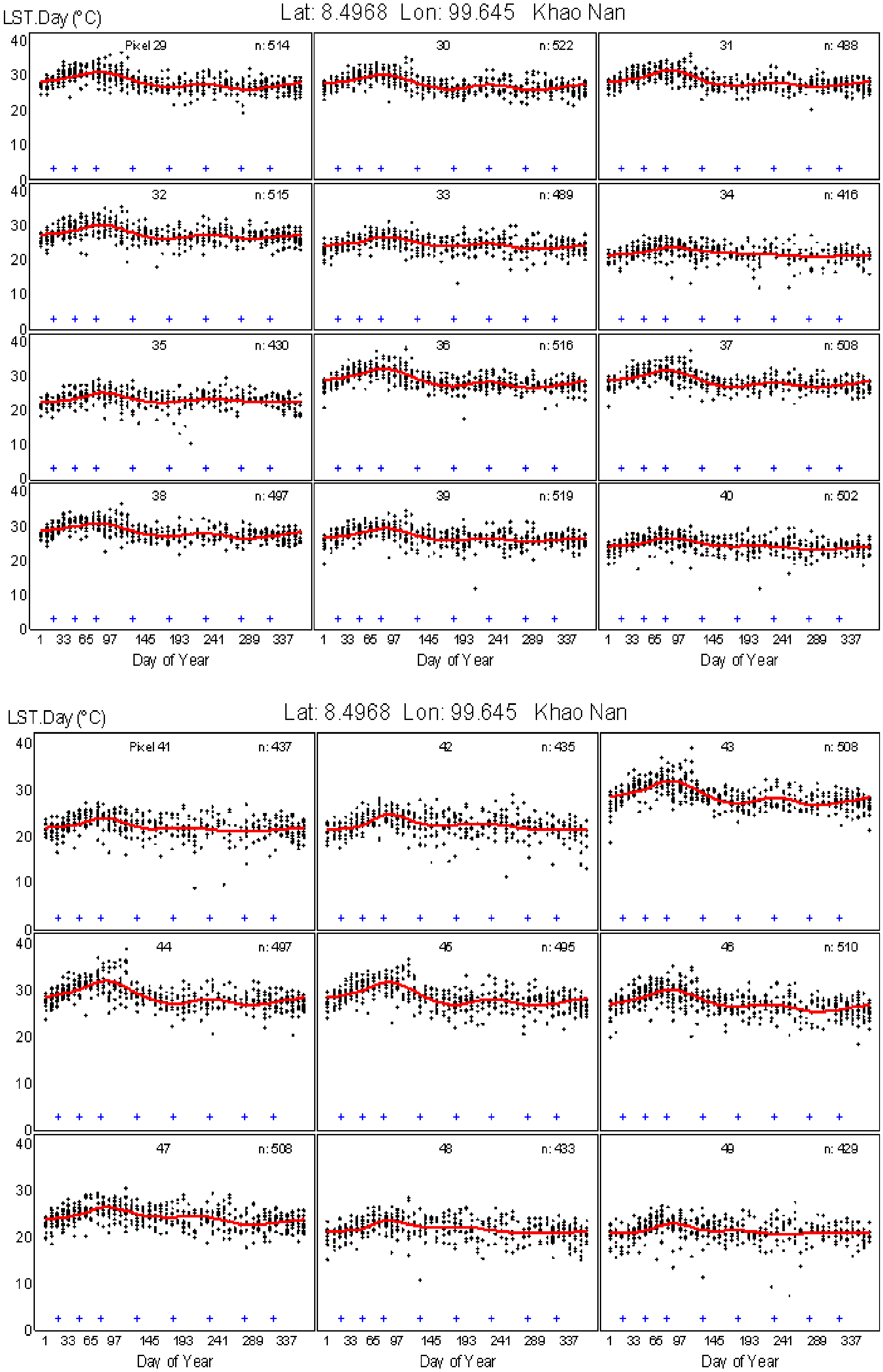

Fig. 6: Temperature pattern for LST Day of pixel 29-40 (Upper) and pixel41-49 (Lower)

The most number of non-missing observations is on pixel 3 and 4 with 76.52%. The seasonally-adjusted temperatures are computed by subtracting the temperature pattern from the data and then adding a constant to ensure that the resulting mean is the same as the mean of the data over the whole period. Fitting a straight line to the adjusted data gives the trend in the time series plot. Figure 7 shows the fitted model (plotted as the red curve) is the trend plus the temperature pattern in a central pixel. Here, the trend is increasing (0.133) but not statistically significant.

Figure 8 show the trends vary and only two pixels are statistically significant with increasing temperature. These data is spatially correlation. GEE is used as a method that takes these correlations into account. Since the data is large we need to allot

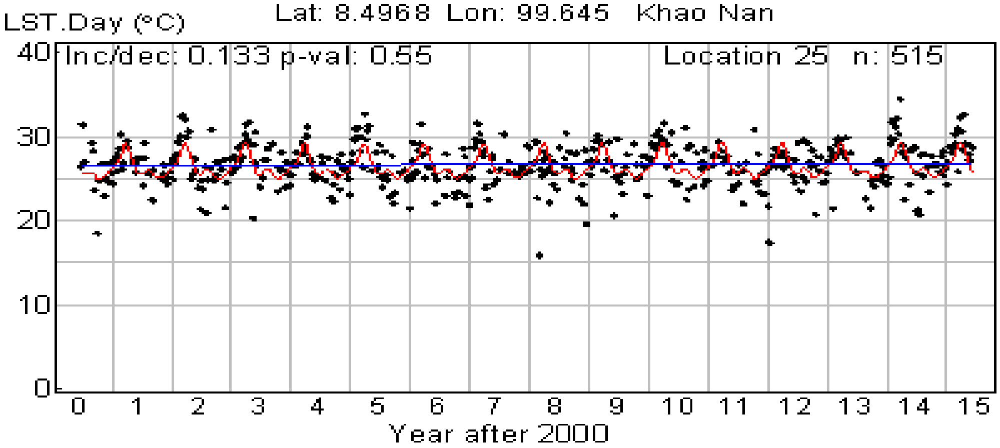

Fig. 7: Time series plot of central pixel over 15 year

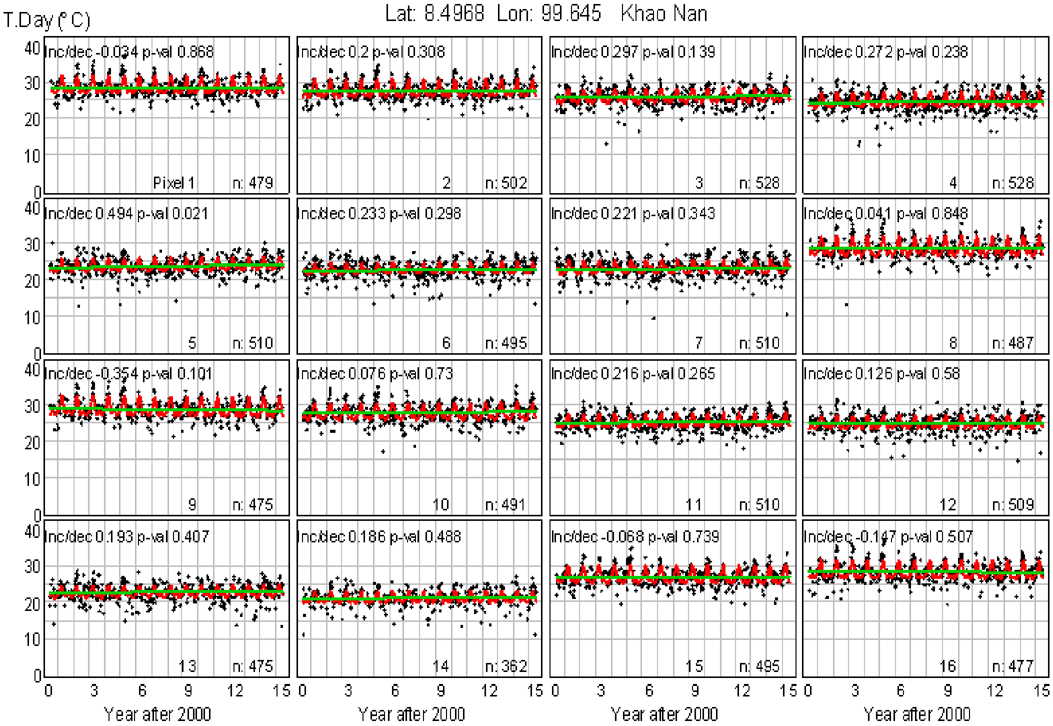

Fig. 8: Continue

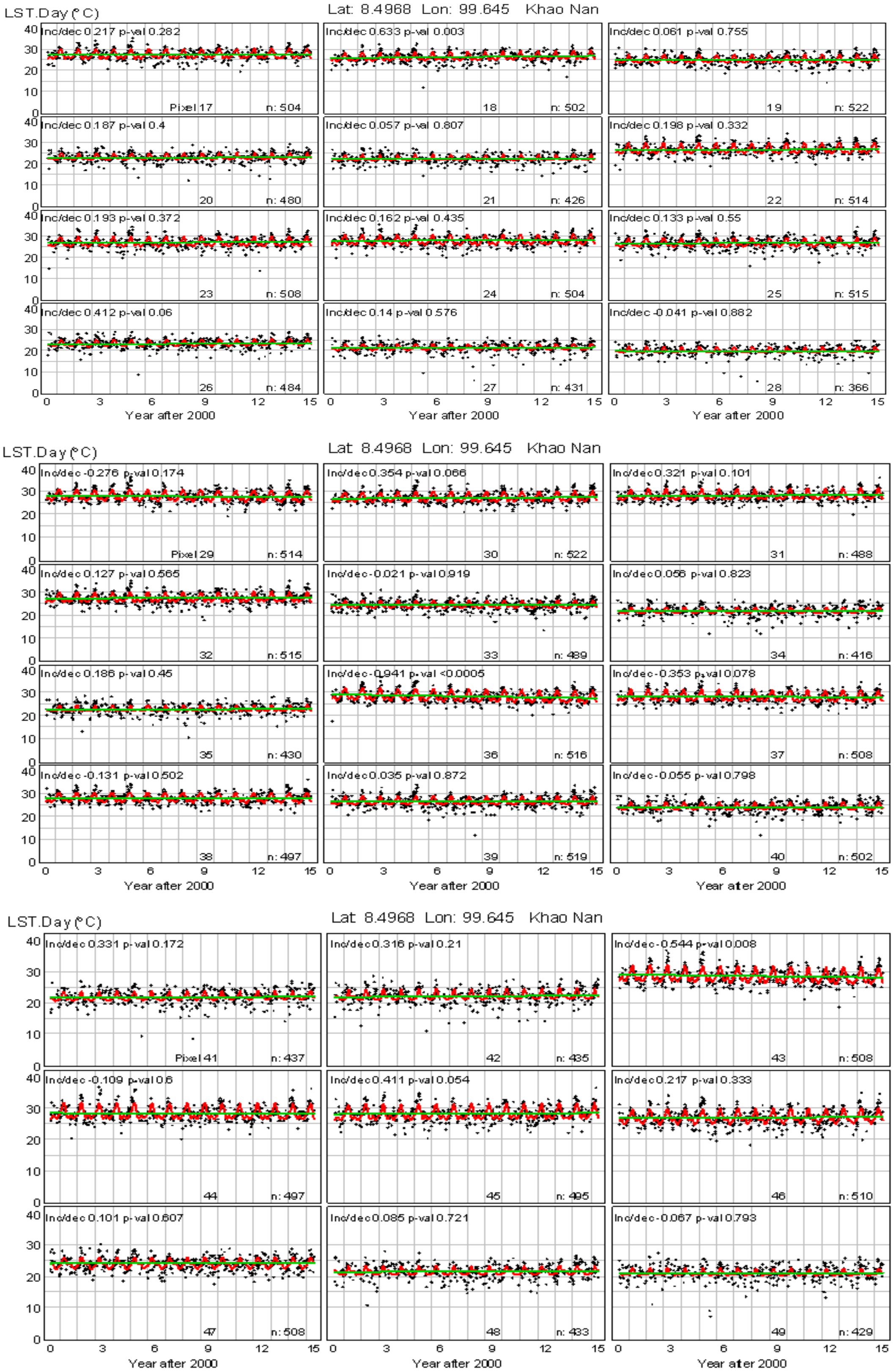

Fig. 8: Time series plots for each pixel

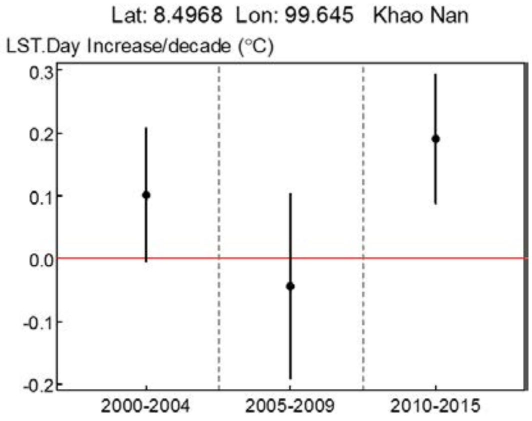

Fig. 9: Temperature variation in 2000-2004, 2005-2009, 2010-2015 and 2005-2009

it every 5 years then analyzed the data as Exchangeable of GEE. Figure 9 shows the temperature has changed over 15 year.

CONCLUSION

The climate in Southern Thailand are 2 seasons such as dry and rain, the dry season is started around day 65 which is at the beginning of March. The maximum and minimum temperatures are 40.29 and 1.07°C, respectively over the 15 year period. The average temperature is 25.55°. Spline function is shown for fitting the temperature pattern in all of the selected pixels. The result shows that those temperatures have similar pattern in all of 49 selected pixels.

Time series shows that only 2 pixels are statistically significant with increasing temperature. The temperature variation has changed increasing except 2005-2009 by using GEE method. Cubic spline regression and GEE are the method which available to perceive pattern and variable of temperature.

ACKNOWLEDGEMENTS

The researchers would like to thank Prof. Don McNeil for his comments and suggestions and Prince of Songkla University for financial support.

REFERENCES

Biebler, K., and M. Wodny, 2013. Splines and Compartment Models. World Scientific Publishing, Singapore,.

Brunetti, M., L. Buffoni, M. Maugeri and T. Nanni, 2000. Trends of minimum and maximum daily temperatures in Italy from 1865 to 1996. Theor. Appl. Climatol., 66: 49-60.

Desquilbet, L. and F. Mariotti, 2010. Dose response analyses using restricted cubic spline functions in public health research. Stat. Med., 29: 1037-1057.

Hanley, J.A., A. Negassa and J.E. Forrester, 2003. Statistical analysis of correlated data using generalized estimating equations: An orientation. Am. J. Epidemiol., 157: 364-375.

Jin, M. and S. Liang, 2006. An improved land surface emissivity parameter for land surface models using global remote sensing observations. J. Clim., 19: 2867-2881.

Krishna, L.V., 2014. Long term temperature trends in four different climatic zones of Saudi Arabia. Intl. J. Appl., 4: 233-242.

Liang, S., X. Li and X. Xie, 2013. Land Surface Observation, Modeling and Data Assimilation. World Scientific Pub. Co., Singapore, ISBN: 9789814472609, Pages: 492.

McNeil, D.R., T.J. Trullell and J.C. Turner, 1977. Spline interpolation of demographic oata. Demography, 14: 245-252.

Prihodko, L. and S.N. Goward, 1997. Estimation of air temperature from remotely sensed surface observations. Remote Sens. Environ., 60: 335-346.

Ruthirako, P., R. Darnsawasdi and W. Chatupote, 2014. Intensity and pattern of land surface temperature in Hat Yai City, Thailand. Walailak J. Sci. Technol., 12: 83-94.

Sala, C., E. Morignat, C. Ducrot and D. Calavas, 2009. Modelling the trend of bovine spongiform encephalopathy prevalence in France: Use of restricted cubic spline regression in age-period-cohort models to estimate the efficiency of control measures. Preventive Vet. Med., 90: 90-101.

Shih, W.J., 1997. Sample size and power calculations for periodontal and other studies with clustered samples using the method of generalized estimating equations. Biometrical J., 39: 899-908.

Shipley, B., D. Vile and E. Garnier, 2006. From plant traits to plant communities: A statistical mechanistic approach to biodiversity. Sci., 314: 812-814.

Shrestha, A.B., C.P. Wake, P.A. Mayewski and J.E. Dibb, 1999. Maximum temperature trends in the Himalaya and its vicinity: An analysis based on temperature records from Nepal for the period 1971-1994. J. Clim., 12: 2775-2786.

Shumway, R.H. and D.S. Stoffer, 2000. Time Series Analysis and its Applications. Springer, New York.

Smith, L., R.J. Hyndman and S.N. Wood, 2004. Spline interpolation for demographic variables: The monotonicity problem. J. Popul. Res., 21: 95-98.

Spiess, M. and A. Hamerle, 1996. On the properties of GEE estimators in the presence of invariant covariates. Biometrical J., 38: 931-940.

Wan, Z. anfd Z.L. Li, 1997. A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data. IEEE. Trans. Geosci. Remote Sens., 35: 980-996.

Watanasit, S., N.N. Anant and A. Phlappueng, 2008. Diversity and ecology of ground dwelling ants at Khao Nan National Park, Southern Thailand. Sonklanakarin J. Sci. Technol., 30: 707-712.

Wold, S., 1974. Spline functions in data analysis. Technometrics. 16: 1-11.

Xie, F. and M.C. Paik, 1997. Generalized estimating equation model for binary outcomes with missing covariates. Biometrics, 53: 1458-1466.

Xu, Y., A. Knudby and H.C. Ho, 2014. Estimating daily maximum air temperature from MODIS in British Columbia, Canada. Intl. J. Remote Sens., 35: 8108-8121.

Zeng, L., B.D. Wardlow, T. Tadesse, J. Shan and M.J. Hayes et al., 2015. Estimation of daily air temperature based on MODIS land surface temperature products over the corn belt in the US. Remote Sens., 7: 951-970.

Zhu, W., A. Lu and S. Jia, 2013. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. Remote Sensing Environ., 130: 62-73.