MODELING OF SATELLITE DATA TO IDENTIFY THE SEASONAL PATTERNS AND TRENDS OF VEGETATION INDEX IN KATHMANDU VALLEY, NEPAL FROM 2000 TO 2015

Ira Sharma, Attachai Ueranantasun*, Phattrawan Tongkumchum

Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, 181 Charoenpradit Rd., Rusamilae, Muang Pattani, Pattani 94000

Graphical abstract

Abstract

Normalised difference vegetation index (NDVI) data were analysed to identify the seasonal patterns and the time series trends of vegetation in Kathmandu. The data were managed in three steps: reordering, removal of unreliable values and validating. A cubic spline function was used to examine annual seasonal patterns that revealed regular seasonal peaks (day 225 to 280) and troughs (day 50 to 81) of vegetation and start of greening from April and of browning from November. Linear regression models were fitted to seasonally adjusted NDVI, which statistically showed 40.70% of the grid cells had a significant increase and 24.71% of it had decreasing trends. To adjust for autocorrelation, generalized estimating equations (GEE) were fitted to the data for whole area that showed, the overall vegetation has been significantly declining at a rate of -0.005°C and -0.006°C per decade for 2000-2004 and 2010-2015 respectively. The recent period of decline is alarming for a growing city like Kathmandu.

Keywords: Satellite data, normalised difference vegetation index, cubic spline function, linear model, generalized estimating equations

© 2018 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

Vegetation is a general term for plant life, referring to the ground cover, and is the most abundant biotic element on Earth. It is essential for sustaining the ecological system of the Earth and serves critical functions at all possible spatial scales [1]. Therefore, quantifying the time trends of types, extents and characteristics of vegetation is of utmost importance for resource management, addressing the climatic issues among others. The worldwide studies of vegetation show that it has been changing in location dependent manners [2, 3]. The changes in vegetation on the land surface affect climates at both regional and global regions from short to extended periods of time [4]. Most of the factors that cause climate change are correlated with vegetation [3, 5, 6]. Therefore, a study of quantitative changes of vegetation is important for assessing climate change related issues. However, reliable and complete data are challenging to obtain, especially in low-income countries where the field data inventory system is not yet properly in place. In this situation, the remote sensing or satellite data provide the best alternative, and one of the most common types of satellite based data is the Normalised Difference Vegetation Index (NDVI) from Moderate Resolution Imaging Spectro-radiometer (MODIS) [2, 7, 8, 9, 10].

MODIS is a sensor, fitted aboard the Terra and Aqua satellites by the National Aeronautics and Space Administration (NASA), and it monitors environmental changes due to fire, vegetation, temperature, earthquakes, droughts and floods of
Earth [10]. Normalised Difference Vegetation Index (NDVI) is created based on MODIS remote sensors capturing the spectral behaviour of vegetation. The theory behind these sensing data is that vegetation reacts differently to different parts of the electromagnetic spectrum (including visible light). The electromagnetic waves are typically absorbed in the red and blue wavelengths, so reflected light retains the green wavelengths, with strong reflection also in the near infrared (NIR) wavelengths [11]. Based on this, the NDVI is calculated as a normalised ratio of the NIR and red bands. In every grid, each observation time is 16-day period. Therefore, in every year, the number of observation times is 23 and, consequently, a total of 345 times for 15 years period. NDVI for each observation time can be computed as follows,

$$\text{NDVI}_i = \frac{(\text{NIR}_i - \text{RED}_i)(\text{NIR}_i + \text{RED}_i)}{2}$$ \hspace{1cm} (1)

where $i$ denotes an observation time (1, 2, 3, ..., 345). $\text{NDVI}_i$ is the NDVI value of the observation $i$, while $\text{NIR}_i$ is the NIR reflectance and $\text{RED}_i$ is the RED reflectance of the observation $i$, respectively.

NDVI has been found more reliable [12] than other data types and equally useful for study purposes in either local [13] regional [7] or global scales [3]. The analysis of large areas is common in prior research studies [2, 6, 9, 13] in which NDVI has been applied for detecting changes. The local changes need to be analysed in a relatively small area for benefit of the local government or the local people. In a macro-level spatial variation analysis, often the local level changes have been overlooked. Additionally, an analysis of remote sensing data over a smaller area for understanding the seasonal patterns and trends in detail is still seldom pursued.

A country where the vegetation data are considered fairly significant for environmental issues is Nepal, as the Department of Forest Research and Survey has reported that 40.36% of its total land areas is still covered by forests [14]. Regarding the vegetation in Nepal, most studies have been regional [15, 16] or on the national level [17, 18]. Studies have been carried out using remote sensing data to investigate the changes in land covers [16, 18, 19] and farming or grazing areas [15, 17] and the associated factors. However, assessment of changes in vegetation as a natural resource, in particular for Nepal, is still lacking. Moreover, the current and accurate data on vegetation index, especially including historic time series has not been widely available. It is also difficult to survey the vegetation in Nepal because of complications in its geography. This lack of vegetation studies and relevant data is also evident for Kathmandu valley, a part of the mountain range in middle Nepal. Because Kathmandu valley is the fastest growing urban region, the vegetation changes are necessary to assess for urban planning and environmental concerns. Therefore, a preliminary step is to understand the interactions of vegetation and other factors, and to identify the patterns and trends through temporal and spatial analysis techniques. This study was aimed to identify the inter-annual temporal trends and intra-annual seasonal patterns using NDVI as remote sensing data for Kathmandu valley from 2000 to 2015 by using appropriate statistical methods.

### 2.0 METHODOLOGY

#### 2.1 The Data

The Kathmandu valley, covering an area of 900 km², consists of the three major districts, Kathmandu, Bhaktapur and Lalitpur, with the highest population density (>4000/ km²) in Nepal. Kathmandu valley has a warm temperate climate with dry and cold winters [14]. The temperature is highest (>30°C) in April and May, and the lowest (<1°C) in December and January. There is a heavy monsoon period in middle of the year. It has three main annual seasons, summer, rainy fall and winter. Regarding vegetation, the valley consists of mostly the temperate varieties. The plants shed off their leaves during winter, give sprout from March, and become fully canopy loaded in June [20].

![Figure 1 Study area in Kathmandu (demarke as parallelogram at the center)](image)

The NDVI data were downloaded from MODIS’s website for data subset [21] for the study area around Kathmandu valley as shown in Figure 1. The coordinates of the central point of the study area are 27.595°N and 85.394°E. Regarding the format of NDVI data from MODIS, the area around the central point was obtained with 250×250 m² grid as spatial
resolution. The covered area extended on all four sides 10 km away from the center (East, West, North and South). As a result, the study area was automatically generated to cover 20.25×20.25 or 410.0625 km², with 6561 grid cells (81×81) of 250×250 m² each. The downloaded data were specified for a period from 2000, the starting point of MODIS service, to 2015. The NDVI data were recorded every 16 days. For each grid cell, there were around 23 observations per year, accumulating to maximally 345 observations over the 15 years. Some observations were missing due to the sensor’s technical problems, and the actual total count of observations for each grid cell was typically below this maximum.

The raw data for each selected NDVI grid cell were divided by 10000 to adjust the values to the range from −1 to 1. The negative values up to 0 correspond to water. The values from 0 to 0.1 indicate soil, rocks or concrete, snow land and barren land. The low positive values (0.2 to 0.4) mean shrubs and grass land. Values close to 1 (0.6 and above) are detected for forests [22]. Therefore, the greater the NDVI value is, the denser the vegetation is in the area.

![Figure 2](image2.png)

**Figure 2** The plot of NDVI from the central grid cell showing unreliable low indices

As an example, the total scatter plot of NDVI data across the central grid is displayed in Figure 2. Here, n represents the total number of observations and every dot on each vertical grid line represents one observation value on the same recording period (23 periods). Therefore, every vertical line on the x-axis displays 15 NDVI values corresponding to a particular day in each of the 15 years, in a consecutive manner. This plot starts from Julian day 1 and ends on day 365. However, the data needed to be further organized before going to analysis due to three problems. Firstly, during the raining season or between days 160 and 260, sparse NDVI values could be seen. This scarcity could affect the analysis for determining seasonal patterns, which needs uncensored data. The second problem was rapid increases in NDVI within a short time periods (within a few weeks or months), which was virtually impossible. This was attributed to growth of plants and these NDVI observations were considered unreliable (an example is illustrated by an arrow between two encircled dots in Figure 2). The final problem for these data was that, on a heavily clouded or wet day, NDVI might be perturbed by obstruction of the sensor by clouds or water vapour, and the data would need to be validated. The NDVI data then needed to be cross compared with another MODIS signal at the same location and time, to confirm it.

![Figure 3](image3.png)

**Figure 3** The plot showing unreliable (cross marks) and doubtful (blue dots) data for the central grid cell

To solve the aforementioned problems, the data were cleaned in the following steps. First, the data were reordered to start from day 190 and end on day 189 as shown in Figure 4 to move scarcity to beginning and end of the period. This made the data more continuous for further analysis. Second, the data points were deemed unreliable if they were greater than 0.02 or lower than −0.02, and consequently were removed to reduce the level of fluctuation, while still maintaining the normality assumption of residuals in data. Those data points are shown with cross marks in Figure 3. Finally, to check NDVI’s validity, Land Surface Temperature (LST) was selected to confirm that all the signals were valid. For the same location and the same observation day, if LST was not realistic, for example the LST went missing or the temperature could not be detected, it confirmed the MODIS did not work properly. Thus, other measured signals of that day, including NDVI, were doubtful. On these grounds, 23 NDVI values were identified as doubtful and were duly removed. These doubtful data are depicted as bigger dots in Figure 3. In total 59 unreliable and doubtful NDVI values were detected and removed. The remaining 238 NDVI observations for this grid after removal of doubtful and unreliable values were used for the analysis. The same procedure was carried out for each grid cell of the study area. R Statistical Programming version 3.2.1 [23] was implemented for data management and graphical displays.
2.2 The Statistical Methods

There were three main steps of data analysis in this study. First, the NDVI data, at individual grid cell level, were used for analysing the intra-annual seasonal patterns. The second step was trend analysis for each grid cell, for which, the data were seasonally adjusted and linear models were fitted to identify trends of individual grid cells. Finally, the data for the whole area were divided into three 5-year periods (2000 to 2004, 2005 to 2009 and 2010 to 2015) for analysis and comparison of the overall NDVI changes in a smaller time frame. To clarify the data structure in this study, the three main notations of the data are defined. Firstly, \( j \) denotes an observation time after omitting the observations of unreliable and doubtful data for each grid cell. Therefore, it varies among different grid cells and takes the maximum of 345, but usually has the value below the maximum (\( j = 1, 2, 3, ..., n \), where \( n =< 345 \)). Secondly, \( q \) denotes an observation time after omitting unreliable and doubtful data for 5-year period for each grid cell with the maximum of 115 observations, so \( q = 1, 2, 3, ..., m \), where \( m =< 115 \). Lastly, \( r \) denotes a grid cell, considered a cluster for the analysis. In this study, then, \( r = 1, 2, 3, ..., N \), where \( N = 49 \).

This study used the cubic spline function to find the seasonal pattern of NDVI. This pattern is calculated by fitting a cubic spline function to the 15-year data combined and plotted in one-year format. This enables the extraction of the seasonal pattern in a year without concerning the long-term trend. Furthermore, the function satisfies a special boundary condition where the functions beyond the first and last knot points are linear with the same slope. Thus, the cubic spline function is a combination of cubic and linear terms. In this study, the cubic spline took the form,

\[
S = \alpha + bt + \sum_{k=1}^{\ell} c_k (t - t_k)^3
\]  

(2)

where, \( S \) is the spline function, \( t \) is time in Julian day, and \( \alpha, b \) and \( c_k \) are the coefficients of the model. \( t_k \) is the location of a knot, while \( t_1 < t_2 < ... < t_p \) are the specified knots and \( (t - t_k) + \) is the positive part of \( (t - t_k) \) or max \( \{0, (t - t_k)\} \).

The data were, then, seasonally adjusted to stabilize the mean. Because of the additive form of the seasonal component retrieved from the cubic spline function, the seasonal adjustment can be realized by subtracting the original values from seasonal component and then adding the difference between the means of seasonal component and seasonally adjusted NDVI [24], which is presented as,

\[
y_j = (x_j - S_j) - (x_j - S_j) + \bar{x}
\]  

(3)

here, \( y_j \) is the seasonally adjusted NDVI at observation \( j \), while \( x_j \) is NDVI and \( S_j \) is the seasonal component value extracted from cubic spline function at observation \( j \), respectively. \( x_j - S_j \) and \( \bar{x} \) are the means of \( x_j - S_j \) and \( x_j \), respectively.

The seasonally adjusted NDVI data were further used for detecting the time series trend over the 15 year period. Therefore, linear models were fitted separately to the seasonally adjusted data at every grid cell, to extract the trends. The form of a linear model is,

\[
y_j = \beta_0 + \beta_1 t_j + \epsilon_j
\]  

(4)

here, \( y_j \) is the seasonally adjusted NDVI, \( t_j \) is the time at each observation and \( \epsilon_j \) is the error term of the data, for each observation \( j \), respectively. \( \beta_0 \) is the intercept of the linear equation and \( \beta_1 \) is the coefficient of the time \( t_j \).

To account for the overall change in the study area, trends for the whole study area were computed. However, the data for a linear trend in each grid cell had spatial correlation within the cell. To tackle this problem, the Generalized Estimation Equations (GEE) were applied in this study. GEE is an extension of linear model that is specially designed for correlated data [25]. Furthermore, from the fitted models explaining the NDVI changes for all 15 years, it was difficult to distinguish the details of trends within this period. Therefore, the data were divided into three periods of 5 years each (2000-2004, 2005-2009 and 2010-2015) and fitted with GEE models separately. To display the results, 95% confidence intervals were calculated for each sub-period and plots were produced to show the changes in three time frames for NDVI in Kathmandu valley. The equations for GEE and related equations [25, 26] can be explained as follows. The generalized linear model can take the form of,

\[
E(Y_{qr}) = \mu_{qr}, \mu_{qr} = g^{-1}(T_{qr}B)
\]  

(5)

where \( Y_{qr} \) is a vector of seasonally adjusted NDVI at observation time \( q \) in a grid cell \( r \). \( E(Y_{qr}) \) or \( \mu_{qr} \) is an expected value of \( Y_{qr} \), \( g^{-1} \) is an inverse link function of \( T_{qr} \), a matrix of observation days and \( B \), a vector of regression coefficients.
Taking into account all 49 grid cells together for the data of 5-year period, the GEE or quasi-score equation to estimate $B$ is as follows,

$$\sum_{r=1}^{N} \left( \frac{\partial \mu_r}{\partial B} \right)^T V_r^{-1} (y_r - \mu_r) = 0 \quad (6)$$

Here, $\left( \frac{\partial \mu_r}{\partial B} \right)^T$ is a transposed matrix of partial derivatives, where $\mu_r$ is a vector of expected values of NDVI at a grid cell $r$, $y_r$ is a vector of NDVI data at a grid cell $r$, and $V_r^{-1}$ is the inverse of the variance-covariance matrix of NDVI.

All data analysis and graphical displays were in R Statistical Programming version 3.2.1 [23].

### 3.0 RESULT AND DISCUSSION

#### 3.1 Result

##### 3.1.1 Seasonal Pattern from Cubic Spline Function

Eight knots were selected to fit the cubic spline function, shown as plus signs at the bottom of Figure 4, at the position of 15, 40, 70, 120, 150, 200, 230 and 350 days. The model gave the coefficient of determination ($R^2$) equal to 50%. The thin line in Figure 4 is the spline fitted before removal of unreliable and doubtful values, while thick line is the spline fit after rearrangement of data and removal of unreliable and doubtful values.

##### 3.1.2 Trends for Individual Grid Cells from Linear Model

For each grid cell, a linear regression model was fitted for the whole 15-year seasonally adjusted data. To illustrate the result, 12 grid cells were selected as representatives of three types of changes: an increase (e.g. grid cells, 242, 2086, 2574, 2580, 4942, 4948 and 5430), a decrease (e.g. grid cells, 248, 730, 736 and 2092) and a no-change (e.g. grid cell 5436), which are shown in Figure 5(a) to 5(l). The annual seasonal fluctuation cycle of NDVI, obtained from the spline function was added back for plotting and is shown as a wavy red line, along with a straight green line depicting the NDVI patterns and trends over 15 years. The dots in the figure are year-wise data points. The crosses are unreliable and doubtful data, removed before the cubic spline fit. The increasing or decreasing trend (Inc/dec) per decade and respective p-values from linear regression are shown in each picture. Here, $n$ represents the number of observations in each plot.

![Figure 4 Spline curves before (thin line) and after (thick line) removing unreliable and doubtful values](image)

The results from all grid cells showed NDVI ranging from 0.3 to 0.9. The seasonal pattern was roughly unchanged for every year as shown in Figure 5. The linear trends illustrated a distinct variation between different years and locations among the grid cells. Moreover, 40.7% of the cells had statistically significant rise, while significantly declining trends were seen in 24.7% of the cells. Out of all grid cells, only 1.2% of the cells did not show any change in NDVI while 33.4% grid cells had changes but were not significant at all. Hence, the results showed a mixed picture of increasing and decreasing trends in NDVI, by grid cell location in the study area.

##### 3.1.3 Trend for the Whole Area and Confidence Interval Plots from GEE

Finally, the GEE were fitted to the data divided into three time periods. The time series plots were drawn to illustrate the trend from GEE in each period. The rates per decade change are -0.005, -0.003 and -0.006 during 2000-2004, 2005-2009 and 2010-2015, respectively. The rate of change showed statistically significant association in only the first and the last periods (p values 0.050 and 0.018 respectively) and the overall decline throughout 15 years period is evident in the graph (Figure 6 (a), (b) and (c)).
As the results from GEE models, 95% CI plots of NDVI for all 49 grids were drawn to observe the confidence levels of the change. In Figure 7, the CI for the period of 2005-2009 is crossing a zero value. Therefore, NDVI can be considered statistically unchanged, coinciding with the p value. However, in other two periods of 2000-2004 and 2010-2015, NDVI showed a statistically declining trend with negative CIs. As shown in Figure 7, the rate of change for the years 2000 to 2004 is ranging from -0.010 to 0 per decade, while that for the years 2010 to 2015 is ranging from -0.011 to -0.001 per decade.

Regarding the distribution of the data, NDVI is found slightly left skewed. However, this distribution is acceptable in this study because GEE can be used with the condition of the distribution assumption being relaxed [27].

![Figure 5](https://example.com/figure5.png)

**Figure 5** Time series plot of vegetation showing increasing, decreasing and no-change trends in 12 different grids.

### 3.2 Discussion

In this study, the seasonal pattern of vegetation showed the highest level in the rainy season during days 241 to 257 (September) and gradually declined to the lowest level in the winter season during days 50 to 97 (February to April). Actually the winter begins from December, but the vegetation declines from the end of the rainy season to the end of January. This time variation of NDVI (decline) by season might be due to the time taken for plant defoliation till the minimum temperature day. The greening of vegetation started from day 97 (April) to the time when the atmospheric temperature adequately warmed up in the summer, while the rising of temperature began from mid of February (after day 45). The seasonal pattern of NDVI growth, however, presented a slight gap between seasons and the vegetation growth, can be another inherent topic for further study.

During the rainy season, the plant growth and refoliation is much favoured by high humidity, temperature and rainfall. Later the growth ceases with heavy defoliation and physiological dormancy after the onset of dry and cold winter season. Therefore, the seasonal pattern of NDVI, as seen in the results, fluctuated driven by these phenomena. This result is particularly helpful in agricultural sector to understand the annual climate response of vegetation or crops, especially to understand the
existing phenological characteristics of the area. This summer-winter vegetation character is well supported by the previous studies where there is higher growth of vegetation in warmer seasons than in winter [8, 12]. Also, it is scientifically proved that, a relatively cooler air temperature reduces the plant metabolic rates including its growth [27]. These prior reports corroborate the results in this study.

In the trend analysis, the statistical results showed that significantly increasing or decreasing grid cells were numerically close to each other and the rest were insignificant on the level of individual grid cells. After GEE, the trends in vegetation changes for the whole study area could be assessed. The 95% confidence interval plots of vegetation showed significant declines in the periods 2000-2004 and 2010-2015, while the mid period showed no change and the rate of decline was the highest in recent years. The Global NDVI trend studied during 1982-2012 showed an increasing trend in many parts of the world including India and Southeast China [3]. The similar result was seen in Tibetan plateau of China during 2000 to 2009 [13]. Nepal lies in between these land blocks and may have similar pattern overall. However, as Kathmandu is a growing densely populated city, it may have been locally affected by several other factors, such as high density of population, resource exploitation, pollution and unplanned urbanization causing the decline of vegetation. In addition, Uddin et al. [28] have explained that the overall vegetation in Nepal is in a state of decline over the past few years. This is consistent with the declining pattern of vegetation found in this current study. Additionally, the method of data rearrangement and cleaning significantly contributed to obtaining much improved results from cubic spline fitting, to get the seasonal patterns. Otherwise, the same spline fitting technique before data management showed a lower NDVI scale even in the rainy season. This method of managing data can also be applicable to other types of noisy data with periodic censoring.

Hence it was found that the overall vegetation around Kathmandu valley is declining in recent years, at different rates by time period, while the seasonal patterns show no remarkable changes. Further investigation is still required to understand the potential reasons behind this seasonal pattern and the trend.

The limitation of this study is that it covers only a part of Nepal. With extended study areas, a complete picture of the vegetation changes for the country could be revealed. Also, this study includes only one indicator from a variety of remote sensing data. Therefore, the inclusion of more indices of satellite data can provide more information regarding changes in vegetation and other related factors in the region.

Figure 6 The trend of overall NDVI (from GEE model) during 2000-2004 (a), 2005-2009 (b) and 2010-2015 (c)

Figure 7 Confidence Interval plots of the overall NDVI trends in three time frames
4.0 CONCLUSION

This study showed the changing pattern and trend of NDVI in three ways. Firstly, the seasonal pattern at grid cell level showed the local level annual changes, and secondly, the trend of individual grid cells indicated the changes in small, grid level areas and the proportions of each change. Finally, the detail of trends for the whole study area in three time periods were investigated. In addition, a comparison of rates of changes in those time segments was carried out. Hence, it was clear that the vegetation in this study site had different rates of decline in different time frames, since 2000.

The spatio-temporal changes of vegetation are a serious threat to ecosystems today. Therefore, this is an alarming signal for the policy makers around Kathmandu valley and measures to prevent a further vegetation decline should be taken. Technically, the study concludes that applying spline function fit and linear models along with GEE help successfully analyse the seasonal patterns and time profiles of changes in NDVI. Some other environmental factors could be added to predict probable causes of vegetation decline. This study indicates that simple yet effective approaches to time series data for assessing spatial and temporal changes in urban vegetation at a local scale can provide basic information for urban planners and anthropogenic studies.

Acknowledgement

This work was supported by the Higher Education Research Promotion and the Thailand’s Education Hub for Southern Region of ASEAN Countries Project Office of the Higher Education Commission (grant number, TEH-AC 018/2015. We acknowledge the Department of Mathematics and Computer Science of Prince of Songkla University for providing facilities for this study. We are grateful to Professor Don McNeill for his immense guidance during the study.

References

DOI: http://dx.doi.org/10.1029/2003GL018251.


