Mapping and attributing normalized difference vegetation index trends for Nepal

Nir Y. Krakauer1*, Tarendra Lakhankar1, José D. Anadón2

1 Department of Civil Engineering and NOAA-CREST, The City College of New York, New York, USA; tlakhankar@ccny.cuny.edu
2 Department of Biology, Queens College, New York, USA; Jose.Anadon@qc.cuny.edu
* Correspondence: mail@nirkakauer.net; Tel.: +1-212-650-8003

Abstract: Global change affects vegetation cover and processes through multiple pathways. Long-time series of surface land surface properties derived from satellite remote sensing give unique abilities to observe these changes, particularly in areas with complex topography and limited research infrastructure. Here, we focus on Nepal, a biodiversity hotspot where vegetation productivity is limited by moisture availability (dominated by a summer monsoon) at lower elevations and by temperature at high elevations. We analyze normalized difference vegetation index (NDVI) from 1981 to 2015 semimonthly, at 8 km spatial resolution. We use a random forest (RF) of regression trees to generate a statistical model of NDVI as a function of elevation, land use, CO2 level, temperature, and precipitation. We find that NDVI has increased over the studied period, particularly at low and middle elevations and during fall (post-monsoon). We infer from the fitted RF model that the NDVI linear trend is primarily due to CO2 level (or another environmental parameter that is changing quasi-linearly), and not primarily to temperature or precipitation trends. On the other hand, interannual fluctuation in NDVI is more correlated with temperature and precipitation. RF accurately fits the available data and shows promise for estimating trends and testing hypotheses about their causes.

Keywords: random forest; regression tree; carbon fertilization; land cover change; climate change

1. Introduction

Vegetation is being impacted globally by widespread stressors and changes, including land conversion to anthropogenic uses, climate change leading to heat and moisture stress, CO2 fertilization, nitrogen deposition, and the spread of pest and invasive species. While methods such as in situ inventories, atmospheric trace gas measurement, and numerical modeling can provide valuable insights into quantifying and attributing impacts [1–5], remote sensing of the land surface offers a unique avenue for observing change in vegetation cover over large areas and timespans of days to decades. Normalized difference vegetation index (NDVI), based on the relative surface reflectance in red and near infrared wavelengths, is well correlated with cover of healthy vegetation [6], and regional and global products based on different satellite sensors are available [7,8]. NDVI is negatively affected by drought in warm regions [9–11] but has increased in response to warming in many temperate and Arctic areas, which has resulted in longer growing seasons there [12–14]. For East and Central Asia, NDVI was found to have increased from roughly 1982 to 1996, due to longer growing seasons, and then decreased from 1997 to 2006, due to worsening aridity [15,16]. Urbanization, irrigation and fertilization can also change NDVI [17,18]. Since the trends in most places are small in magnitude, the remote sensing data need to be carefully processed to remove artifacts due to, for example, degradation of the satellite sensors over the course of a mission [19].
Here, we study trends in vegetation cover in Nepal (located at 26-31 °N, 80-89 °E) a least developed country where the majority of the population is engaged in agriculture and is highly vulnerable to climate change. Nepal is a biodiversity hotspot, due in part to the wide topographic and climatic range found over relatively short distances, ranging from the Indo-Gangetic plain in the south to the Himalayan peaks and the Tibetan Plateau to the north [20]. There have previously been several studies of NDVI trends in the broader region. [21] found a generally increasing trend in spring NDVI over the Hindu Kush-Himalayan region between 1982 and 2006. [22] analyzed NDVI seasonality between 1982 and 2006 for the Himalayas region, including Nepal, finding that the start date of the growing season trended earlier while the end date did not change. [23] conducted cluster analysis to identify patterns in mean and maximum warm-season NDVI between 2001 and 2016, finding mostly increasing trends (greening), with decreasing trends (browning) most common between 4 and 5 km elevation. Similarly, a study of NDVI trends in Yarlung Zangbo Grand Canyon Nature Reserve, Tibet between 1999 and 2013 [24] found that greening was concentrated at the lower elevations, below 3 km. Delayed green-up in Alpine grasslands of the western Tibetan plateau may be due to declines in spring precipitation [25,26].

Limited research has been carried out on NDVI trends specifically in Nepal. [27] mapped forest types in Manaslu Conservation Area and computed NDVI trends and correlations with temperature and precipitation from a nearby meteorological station for 2000-2008. [28] found a significant increasing trend in warm-season NDVI over the Koshi River Basin over 1982-2006, though with a decline between 1994 and 2000.

This research, therefore, has two main objectives. First, we describe trends in NDVI in Nepal by elevation and season based on a long-term remote-sensing data product. Second, we attempt to attribute trends and interannual variability to changes in climate, land use, and CO₂ concentration.

2. Methods

2.1. Data

2.1.1. NDVI

NDVI data was obtained from the NDVI3g.v1 time series, an update of the earlier NDVI3g.v0 [29] which provides NDVI values twice a month on a 1/12 degree (approximately 8 km) grid from July 1981 to December 2015. This dataset is derived from measurements by over a dozen Advanced Very High Resolution Radiometers (AVHRR) that orbited on different satellites for parts of this time period. It has been extensively processed to correct for artifacts resulting from causes such as instrument and orbit drift and volcanic eruptions so to be suitable for analysis of climate change. The formula for NDVI is \( \frac{(\text{NIR} - \text{RED})}{(\text{NIR} + \text{RED})} \), where NIR refers to reflectance in the AVHRR near-infrared band (channel 2, 0.725-1.1.0 \( \mu \)m) and RED to reflectance in the AVHRR red band (channel 1, 0.58-0.68 \( \mu \)m) [30]. Missing or suspect data in NDVI3g.v1 is flagged and filled in either by spline interpolation or from an average of other years.

Despite quality control steps used to derive this product, we found occasional NDVI values that were quite different from those at adjacent time periods, and which are therefore likely to be due to satellite instrument or viewing condition artifacts [31]. We therefore smoothed the NDVI3g.v1 series by subtracting the median seasonal cycle, applying a 3-point median filter, and adding the seasonal cycle back.

2.1.2. Elevation

Elevation was obtained at 3 arcsecond (approximately 80 m) resolution from the United States Geological Survey (USGS) and World Wildlife Foundation (WWF) Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales (HydroSHEDS) project. HydroSHEDS is derived from spaceborne radar images with extensive quality control and corrections of artifacts [32]. We
computed the average and standard deviation of elevation over each 1/12° grid cell in Nepal as possible predictors of NDVI. Pixel-mean elevations ranged from 60 to over 6000 m (Figure 1a), with an average of 2078 m.

2.1.3. Climate

Monthly mean temperature and precipitation were obtained on a 1° grid. Precipitation was from the Global Precipitation Climatology Center (GPCC) combined Full Version 7 and Monitoring Version 4 product, which is based on quality-controlled data from thousands of stations globally (including data contributed by Nepal through the World Meteorological Organization) that is interpolated to fill in gaps in coverage [33–35]. Temperature was from the Berkeley Earth (BEST) dataset, which uses several times more station records compared to other gridded temperature data sets. Station records undergo automated quality control, and are weighted using geostatistics methods to produce spatial fields [36].

The 1° resolution of these available products does not fully resolve the topography-driven climate variability in Nepal. We mitigated this shortcoming, to some extent, by applying downscaling adjustments. We downscaled the temperature data within each 1° cell to 1/12° by applying a lapse rate of 6 K per km as an additive adjustment. This lapse rate was approximately that inferred from the temperature difference between adjacent BEST pixels, which was found to be approximately the same for all seasons. We downscaled the precipitation data within each 1° cell to 1/12° by applying a multiplicative conversion factor based on the higher-resolution gridded APHRODITE product, available for 1951-2007 in a public version at 0.25° resolution and, courtesy of the Nepal Department of Hydrology and Meteorology, at the original 0.05° interpolation resolution [37]. These adjustment factors were the same for each month, so that time trends were not affected, and preserved the 1° mean values from GPCC and BEST.

The mean temperature and precipitation obtained, along with the elevation, for each 1/12° NDVI pixel are shown in Figure 1.

2.1.4. Land cover

Land cover classifications for 1990, 2000, and 2010 were obtained from the International Centre for Integrated Mountain Development (ICIMOD) [38]. These were generated using public domain Landsat Thematic Mapper 30 m images and an object-based classification algorithm, and validated and refined using aerial photographs and field observations. The land cover classes were forest, shrub, grass, agricultural, barren, lake, river, snow/glacier, and urban. We computed the percentage in each cover category for each 1/12° pixel and year. Land cover was imputed by pixel and year via linear interpolation between the three available years. Before 1990 and after 2010, we assumed the land cover to have stayed constant at the earliest/latest available value.

The obtained land cover classification showed that forest (at almost half the area) and agriculture (at about a quarter) were the dominant categories (Table 1). Agriculture dominated in the lowest elevations in the south, while forest dominated in the middle elevations, and grassland and snow and glaciers were found at high elevations in the north (Figure 2). Forest cover decreased by several percentage points between 1990 and 2000 while agricultural and barren areas increased, before stabilizing between 2000 and 2010 (Table 1).

Table 1. Area coverage (%) of Nepal by land cover category and year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Forest</th>
<th>Shrub</th>
<th>Grass</th>
<th>Farm</th>
<th>Barren</th>
<th>Lake</th>
<th>Snow/Ice</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>45.2</td>
<td>2.2</td>
<td>11.7</td>
<td>25.3</td>
<td>6.8</td>
<td>0.6</td>
<td>7.9</td>
<td>0.2</td>
</tr>
<tr>
<td>2000</td>
<td>41.7</td>
<td>2.4</td>
<td>11.4</td>
<td>27.7</td>
<td>9.5</td>
<td>0.5</td>
<td>6.5</td>
<td>0.3</td>
</tr>
<tr>
<td>2010</td>
<td>42.1</td>
<td>2.3</td>
<td>10.5</td>
<td>27.3</td>
<td>8.6</td>
<td>0.5</td>
<td>8.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>
2.1.5. CO₂ concentration

Yearly carbon dioxide dry air mixing ratios from Mauna Loa, Hawai’i (at 20° north, close to Nepal’s latitude) was obtained from the United States Government’s Earth System Research Laboratory, Global Monitoring Division. These are transformed to logarithms and can represent the impact of increasing local carbon dioxide levels on plant gas exchange and carbon fixation. This time series is also highly correlated to the summed anthropogenic greenhouse gas forcing [39], the global warming trend [40], and other monotonic trends over recent decades, such as that of global population [41,42].

2.2. Regression model

We expect the relationship of NDVI with such variables as temperature, precipitation and elevation to be nonlinear. There also may well be interactions between potential explanatory variables, e.g. the effect of precipitation increase could vary depending on elevation and season. Random forest (RF) of regression trees [43] is a method of empirically constructing a predictive model that is well suited for handling such complexity. As such, RF has been used in many environmental mapping applications, including for wetland cover from radar imagery [44], water table dynamics and depth to groundwater [45,46], and ecosystem light use efficiency [47]. Here, the training data were half of the available 1,429,443 NDVI values from Nepal (1955 pixels for 828 time periods, excluding 12% interpolated data). The half the available data not used to train the model (test data) provided a test of its ability to capture the NDVI patterns consistently found by remote sensing. The RF model run included 100 regression trees, and other parameter settings were kept at default values from the R randomForest package [48].

The predictors in the RF model fell into the following categories: (1) Interannually constant seasonal and geographic factors: month of year, pixel longitude, pixel latitude, pixel mean elevation, pixel standard deviation of elevation. (2) Mean temperature (°C) for 0-0.5, 0.5-1.5, 1.5-3, 3-6, 6-12, 12-24, and 24-48 months prior to the end of each semimonthly period. (3) Precipitation rate (mm month⁻¹) over the same periods as for temperature. (4) Land cover: percent of pixel in each of 8 land cover categories. (5) Logarithm of atmospheric CO₂ concentration. The total number of predictors in the model was therefore 5 + 7 + 7 + 8 + 1 = 28.

We then used the fitted model to predict NDVI trends if only one factor (temperature, precipitation, land use, or CO₂) changes with time, while the others are held at average values for the period (in the case of temperature and precipitation, seasonally specific averages). To the extent that these factors have different time histories so that their effects of NDVI can be separated by the RF model, this would allow us to estimate how much of the NDVI trends and interannual variability over Nepal can be attributed to each factor.

2.3. Analysis

For each grid cell and (semi-monthly) time of year, we computed the NDVI trend as the slope obtained from linear regression, using only non-interpolated values. Trends were similarly computed for the RF-predicted NDVI. The Nash-Sutcliffe coefficient [49], applied to different transformations of the data, is used to quantify how well the RF models fit test data based on either the full set of forcings or sets that include only one time-changing factor. NSC_all is based on the raw observed and modeled NDVI values (using only the test data), NSC_detrend is based on the same NDVI values, but after subtracting the linear trend, which also removes the mean seasonal and spatial NDVI patterns from consideration. NSC_trend is based on the linear trends (calculated using both training and test data), and measures how well the RF model is able to capture the observed trend across space and season.

Figure 3 shows graphically the overall workflow followed, including the relationship between data processing and modeling.

As one way of visualizing the results, we compute the time series of mean NDVI over Nepal by averaging across the pixels at each (bi-monthly) timestep, using either the filled-in NDVI or the RF
predictions. The mean seasonal cycle is subtracted to obtain a deseasonalized time series for computing trends and detrended variability. We also compute mean NDVI amounts and trends by season and elevation, using a smoothing spline [50] to estimate the mean elevation dependence.

3. Results

Mean NDVI was near 0 in the dry, cold high-elevation north of Nepal, which has little vegetation. It was somewhat lower in the low plains near the India border compared to middle elevations, peaking at about 0.65 over the elevation range 500-2000 m (Figure 4).

The time series of mean NDVI is dominated by the seasonal cycle, though interannual variability is also evident (Figure 5). The mean NDVI seasonal cycle nationally and at elevations under ∼4000 m was more influenced by moisture than by temperature, with peak values immediately after the monsoon in early October. NDVI declines through winter, as water availability decreases, and reaches a nadir in late April (Figure 5b; Figure 6a). At higher elevations (4000-5000 m), the peak occurs earlier (late September) and the lowest values are in late February, consistent with a greater role of cold temperature in controlling vegetation cover.

NDVI overall showed an increasing linear trend, averaging $0.448 \times 10^{-3}$ units per year, over the period of record. This trend varied across seasons and elevations, however. It was strongest in late August through October, near the annual peak, and in the lower elevations, below 2000 m (Figure 6b). NDVI at 4000-5000 m actually showed a slight declining trend.

The RF model was able to represent the NDVI patterns seen extremely well (Figure 7a), with NSC$_{all}$ of 0.959 for the test data. However, most of this reflects skill at capturing the mean annual cycle (Figure 7b), rather than trends and year-to-year variability, so that even with no interannually-varying forcings, the NSC$_{all}$ was still 0.941. Out of the single forcings, CO$_2$ and precipitation contributed most to improving NSC$_{all}$, and temperature contributed least (Table 2).

Interannual variability in NDVI after detrending was captured less well by the RF model (Figure 7c), with NSC$_{detrend}$ of 0.288 for the test data. As expected, with no interannually-varying forcing, none of this variability was captured. Out of the single forcings, precipitation contributed most to NSC$_{detrend}$, with temperature playing a smaller role that was similar in importance to those of the slowly varying CO$_2$ and land use (Table 2).

The pattern of trends in NDVI across location and season was better captured by the RF model, with NSC$_{trend}$ of 0.793. With no interannually varying forcing, the trend was uniformly zero, as opposed to the positive mean value actually seen, resulting in slightly negative NSC$_{trend}$. Out of the individual forcings, CO$_2$ as well as land cover, which both changed quasi-linearly with time, explained the trends best, with trends in temperature and precipitation showing smaller positive NSC$_{trend}$ (Table 2).

Averaged over all Nepal, the mean NDVI trend from the fitted RF model was $0.422 \times 10^{-3}$ per year, just 5% less than the $0.448 \times 10^{-3}$ calculated from the observation dataset. (Including all pixels and months, also those missing from the observations, increases the mean trend calculated from the fitted RF model slightly to $0.449 \times 10^{-3}$ per year.) This mean trend is indicated as being due essentially to rising CO$_2$ level, which by itself raises NDVI $0.509 \times 10^{-3}$ per year. Land cover has a net negative influence on NDVI trend, while precipitation and temperature trends have small positive influences (Table 2).

Figure 8 shows the modeled NDVI trend by season and elevation, which can be compared with the observation-based trend in Figure 6b. CO$_2$ change is the dominant factor at most of the affected seasons and elevations. Land use change seems to have the largest impacts, which vary by season, above 2000 m. Precipitation trends impact NDVI primarily during the monsoon season of summer and early fall. Temperature trends increase NDVI is the spring and fall around 3000-4500 m.
Table 2. Nash-Sutcliffe coefficients and mean trend for random-forest (RF) predictions of NDVI. The full RF model includes all time-varying factors. The other predictions are with either interannually constant factors or only one time-varying factor. The mean trend is in $10^{-3}$ units per year. For comparison, the mean trend calculated from observations is $0.448 \times 10^{-3}$ per year.

<table>
<thead>
<tr>
<th></th>
<th>NSC\textsubscript{all}</th>
<th>NSC\textsubscript{detrend}</th>
<th>NSC\textsubscript{trend}</th>
<th>Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full RF</td>
<td>0.959</td>
<td>0.288</td>
<td>0.793</td>
<td>0.422</td>
</tr>
<tr>
<td>Constant</td>
<td>0.941</td>
<td>-0.000</td>
<td>-0.068</td>
<td>0.000</td>
</tr>
<tr>
<td>CO\textsubscript{2}</td>
<td>0.946</td>
<td>0.053</td>
<td>0.344</td>
<td>0.509</td>
</tr>
<tr>
<td>Land cover</td>
<td>0.945</td>
<td>0.034</td>
<td>0.238</td>
<td>-0.135</td>
</tr>
<tr>
<td>Precipitation</td>
<td>0.946</td>
<td>0.102</td>
<td>0.043</td>
<td>0.020</td>
</tr>
<tr>
<td>Temperature</td>
<td>0.943</td>
<td>0.035</td>
<td>0.051</td>
<td>0.026</td>
</tr>
</tbody>
</table>
Figure 1. Pixel mean (a) elevation, (b) temperature, (c) precipitation.
Figure 2. 2010 land cover map of Nepal.

Figure 3. Overall structure of data and analysis workflow.
Figure 4. Mean NDVI by pixel over Nepal.
Figure 5. (a) Time series of mean Nepal NDVI. (b) Mean seasonal cycle of NDVI. (c) NDVI anomaly (subtracting the mean seasonal cycle) and the least-squares linear fit to it.
Figure 6. (a) Mean NDVI (unitless) and (b) 1981-2015 linear trend in NDVI ($10^{-3}$ per year) for Nepal, by season and elevation.
Figure 7. Same as Figure 5, but using mean NDVI as predicted by the random forest regression model.
Figure 8. Modeled 1981-2015 linear trend in NDVI ($10^{-3}$ per year) for Nepal, by season and elevation, with (a) all forcings, (b) only CO$_2$ level change, (c) only land use change, (d) only precipitation change, (e) only temperature change.
4. Discussion

We find an overall increasing trend in NDVI for Nepal over 1981-2015. The RF analysis suggests that this trend is not primarily due to changes in climate (temperature and precipitation), but correlates best with increasing atmospheric CO$_2$ level, although precipitation and temperature are more important in explaining interannual NDVI variability. Similarly, ecosystem models suggest that most of the observed increase in the seasonal amplitude of atmospheric CO$_2$, indicative of increasing plant growth in the Northern Hemisphere over recent decades, is due to CO$_2$ fertilization, with climate and land use changes playing secondary roles [5]. An increasing trend in global net primary productivity, particularly in tropical forests, has also been identified based on satellite imagery for 1982-1999 and attributed primarily to CO$_2$ fertilization [2]. On the other hand, in boreal areas, warming has been a major driver of longer growing seasons and higher productivity [51,52], while in arid and semiarid areas, moisture availability is a primary modulator of vegetation growth [15,53–55]. The small positive impact of precipitation on NDVI trends, concentrated during and after the summer monsoon, is consistent with the increasing trend in monsoon precipitation found for recent decades over much of Nepal [56], although dry spells have also increased [57,58].

Land cover change is inferred to have made a negative contribution to the NDVI trend. This is plausible insofar as the land cover data show net decrease in forest area and increase in agricultural area, where forest often has higher and more constant NDVI than agricultural land [59]. Although the importance of anthropogenic land cover change as a driver of ecological impacts is widely recognized [60–62], studies of global and regional NDVI trends over recent decades have generally concentrated on climate drivers without explicitly accounting for the contribution of land cover change. An RF model offers one method to distinguish the influence of all these factors, which operate simultaneously around the world.

The work presented here has several significant limitations. The NDVI trends attributed to CO$_2$ level in the RF model could well also reflect contributions from other factors that have been changing quasi-linearly and whose quantitative evolution was not incorporated in the RF model. These may include, for example, N fertilization due to direct application and deposition, irrigation, change in crops planted, and management of grasslands and forests, and changes in cloud and aerosols, For most of these terms, more work is needed to understand how they are changing over different parts of Nepal. As well, the quality of some of the inputs used could be improved. Land cover change could be evaluated from Landsat imagery before 1990 and after 2010. A precipitation product at higher resolution, which could be based on remote sensing calibrated to available weather stations, would better resolve the sharp elevation and orographic gradients within Nepal and thus help clarify the impact of moisture stress [63–65].

5. Conclusions

We find that NDVI has increased over the studied period in Nepal, consistent with global trends. Increases were uneven, and concentrated at low and middle elevations and during fall (post-monsoon). We infer from the fitted RF model that the NDVI linear trend is primarily due to CO$_2$ level (or another environmental parameter that is changing quasi-linearly), and not primarily to temperature or precipitation trends. On the other hand, interannual fluctuation in NDVI is more correlated with temperature and especially precipitation. RF accurately fits the available data and shows promise for estimating trends in spatiotemporal remote sensing data such as gridded NDVI and testing hypotheses about their causes.

Acknowledgments: The authors gratefully acknowledge support from USAID IPM II under the project “Participatory biodiversity and climate change assessment for integrated pest management in the Chitwan-Annapurna Landscape, Nepal” and from NOAA under grants NA11SEC4810004 and NA15OAR4310080. All statements made are the views of the authors and not the opinions of the funders or the U.S. government.

Author Contributions: All authors conceived and designed the data analysis and contributed analysis tools. NYK performed the data analysis and wrote the paper.
Conflicts of Interest: The authors declare no conflict of interest.

References


