An automated algorithm to detect timing of urban conversion of agricultural land with high temporal frequency MODIS NDVI data
Bhartendu Pandey**, Qingling Zhang* and Karen C. Seto*

July 8, 2013

Abstract
Urban expansion is one of the major drivers of agricultural lands loss. However, current remote sensing-based efforts to monitor this process are limited to small scale case studies that require much user input. Given the rate and magnitude of contemporary urbanization, there is a need to develop a land change algorithm that can characterize the loss of agricultural land at large scales over long time periods. Moreover, characterizing agricultural land conversion trajectories from remote sensing images is complex due to farm size, climatic variability, changes in cropping patterns, and variations in the rate of development processes. Here we propose an econometric time series approach to identify agricultural land loss due to urban expansion, utilizing high temporal frequency MODIS NDVI data between 2000 and 2010. The algorithm is comprised of two main components: 1) detrending the time series, and 2) testing for the presence of a breakpoint in the detrended time series and estimating the date of the breakpoint. Evaluations of the algorithm with simulated and actual MODIS NDVI data confirm that the method can successfully detect when and where urban conversions of agricultural lands occur. The algorithm is simple, robust, and highly automated, thus is valuable for monitoring agricultural land loss at regional and even global scales.

1. Introduction
We currently live in an era of unprecedented rate and scale of urbanization. Worldwide, the growth of urban areas has transformed landscapes, and in most places the expansion of urban land is occurring at rates faster than the growth of urban population [1, 2] and often on prime farm lands [3-6]. Historical urban conversion of agricultural lands is well documented throughout the world [3-6] and faster rate of urban growth implies concomitant rapid agricultural land conversion to urban land-use types. Urban conversion of agricultural land is a non-linear process and its causal attribution is complicated due to involvement of several time dependent socioeconomic and political factors [7-9]. Statistical analysis for establishing linkages between urban conversion of agricultural lands and various socioeconomic and political processes require multiple observations through time. While multiple observations may be useful in understanding...
the linkages between urban conversion of agricultural lands and socioeconomic processes in regions experiencing gradual land conversions, high temporal frequency observations is especially important in rapidly urbanizing regions [7]. Often the socioeconomic variables are available annually from a variety of sources including statistical yearbooks, whereas information extraction from high temporal frequency remote sensing data on urban land conversion remains a challenge [7, 8]. Earth-observation satellites have generated remotely sensed data for more than 40 years of continuous spatial and temporal coverage that aid in continuous monitoring of land use change process [7, 10, 11]. Despite the long time series and high temporal frequency of many remote sensing datasets, the vast majority of studies on urban growth and concomitant land use changes examine change in two [3, 5, 12] or three time periods [9, 13]. The scarcity of studies involving high temporal frequency mapping to analyze urban growth is attributed to limited number of algorithms [7, 8].

Conventional change detection algorithms utilizing two images have significant limitations in the ability to delineate timing of change, namely that change is estimated to occur in between the two dates. Recognizing these limitations, algorithms for multi-temporal change detection have been developed including, change vector analysis [14], principal component analysis [14, 15], wavelet decomposition [16], cross-correlogram temporal signature matching [17], stacked multi-date composite classification using supervised training algorithms [10, 18], three dimensional segmentation of multi-spectral and multi-temporal images [19], genetic algorithm approach [20], autocorrelation analysis [21], Kalman filtering approach [22], cumulative sum approach [16, 23], Fourier transformation [24]. Still, these methods largely depend on the type of land conversions and their applicability is often limited to specific landscapes. In addition to conventional change detection algorithms, econometric time series analysis methods have also been applied to multi-temporal remote sensing data i.e., multiyear observations [8] and high temporal frequency observations at 16-day [25, 26] and quarterly intervals [27], to delineate the timing of change. The advantage of time series analysis methods is that these are independent of specific threshold requirements or change trajectories [25]. Thresholds determination significantly impedes the applicability of an algorithm to large spatial extents. Trajectory based change detection has been recognized as a useful tool in multi temporal change detection [10, 28, 29]. This method relies on well-defined change trajectories pertaining to specific spectral data and functions only if the empirical change trajectories matches the hypothesized ones [25, 29]. In case of urban
conversion of agricultural lands, the change trajectories are complex, and conventional multi-temporal change detection algorithms cannot be applied [24]. The complexity of agricultural land conversion trajectories is attributed to small farm sizes, climatic variability, changes in cropping patterns through time, and rate of the development process [10, 16, 24]. In this context, a generic change detection approach independent of specific spectral data requirement and change trajectories, can be applied to monitor urban conversion of agricultural lands. The time series analysis method proposed in [25, 26] detects break in trend and seasonal parameters through time and is an example of a generic change detection approach. In this study, we extend this algorithm to pinpoint the timing of urban conversion of agricultural lands in high temporal frequency MODIS NDVI time series and test if these breaks correspond to the actual timing of change for agricultural land conversion using Google Earth imagery as validation.

The goal of this research is to identify annual agricultural land loss due to urban expansion using time series econometric approaches applied to NASA’s Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI). Specifically, we aim to develop an algorithm that enables rapid and automated assessment of the timing of agricultural land loss. Thus this algorithm can be integrated with current monitoring frameworks to provide information on when and where urban conversions of agricultural lands occur. In this study we developed and tested the algorithm across different urbanizing regions in India.

2. Data Description

We use the 250m MODIS MOD13Q1 NDVI dataset for this analysis because of its relatively high temporal resolution. We acquired the MODIS MOD13Q1 NDVI dataset for the period June, 2000 - May, 2010 from http://mrtweb.cr.usgs.gov. The MOD13Q1 NDVI dataset is composited from multi-date observations with the constrained view angle-maximum value composite (CV-MVC) technique at 16-day intervals [30]. The CV-MVC technique reduces the negatively biased noise due to interference of clouds and atmospheric constituents, and constrains the angular variations encountered during the maximum value composite (MVC) selection. The compositing technique produces higher quality datasets but it also aggregates the uncertainties from residual aerosol, cloud and angular effects [30]. Therefore, we applied the adaptive Savitzky-Golay filter using the TIMESAT program [31], to minimize the residual noises in the NDVI time series. The adaptive Savitzky-Golay filtering approach performs better than other algorithms in terms of
flexibility and robustness [3]. In addition, we utilized the pixel reliability parameter included in the MOD13Q1 NDVI dataset during the time series filtering. The pixel reliability parameter ranks each pixel of the composited image into five categories: Good-Marginal-Snow/Ice-Cloudy-No Data. We weighted the NDVI values based on the pixel reliability parameter, assigning 1 for good quality data and 0.5 for marginal quality data. Our filtering approach can effectively minimize the residual noise in the NDVI time series (Fig.1).

3. Automated algorithm to identify the timing of change

In time series analysis, a breakpoint is where the statistical properties such as the mean or variance change in a series of observations [32]. In the econometrics literature, breakpoints are often referred to as structural breaks. The techniques used to identify structural breaks rely on the assumption of stationarity of model parameters. A structural break is said to occur when at least one of the model parameters changes over time. Time series analysis of breakpoints have been utilized in many fields such as medicine [33], climatology [34], oceanography [35], and finance [32] to identify structural shifts in underlying processes. Recently, one of such breakpoint detection techniques, the cumulated sum test, was applied to monitor disturbances in mangrove forests with high temporal frequency satellite data [27]. Here we extend the generic change detection approach proposed in [26] to detect breaks in seasonal and trend components estimated from high temporal frequency MODIS NDVI data, to identify the timing of agricultural land loss with the 16-day MOD13Q1 NDVI time series. We use “breakpoints” and “timing of agricultural land loss/timing of agricultural land conversion” interchangeably throughout the paper.

Our algorithm is comprised of two primary steps: a) detrending the time series, and b) testing for the presence of a breakpoint in the detrended time series and estimating the position of breakpoint (Fig. 2). The methods included in step 2 assume stationarity in the input time series. Therefore, step 1 can be interpreted as a data preprocessing step to impart stationarity in the input time series such that the breakpoint can be accurately determined. Step1 proceeds with an initial additive decomposition of input time series into seasonal, trend and remainder components [36] i.e.,

\[
y_t = S_t + T_t + R_t \quad t = (1, \ldots, n)
\]
where, $Y_t$ is the input time series, $S_t$ is the seasonal component, $T_t$ is the trend component, $R_t$ is the remainder component (total variation in the input data that remains unexplained by $S_t$ and $T_t$) and, $n$ is the total number of observations. We test for the presence of breakpoint ($\tau_1$) in the sum of trend and remainder components ($T_t + R_t$). We employ the supF test [37, 38] to test for the presence of single structural break in a linear model under the null hypothesis of no structural breaks, i.e., $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, such that,

$$T_t + R_t = \alpha_1 + \beta_1 \cdot t, \ t = 1, \ldots, \tau_1$$

$$T_t + R_t = \alpha_2 + \beta_2, \ t = \tau_1 + 1, \ldots, n \quad (2)$$

We fit a linear model for the observations before and after all possible $\tau_1$'s, and compute the error sum of squares ($\text{ESS}_{\tau_1}$). Further, we fit another linear model for $n$ observations and compute the restricted sum of squares ($\text{RSS}_1$). Using the estimated $\text{ESS}_{\tau_1}$ and $\text{RSS}_1$, we compute series of F-statistics ($F_n(\tau_1)$) as,

$$F_n(\tau_1) = \frac{(\text{RSS}_1 - \text{ESS}(\tau_1))}{\text{ESS}(\tau_1)} \cdot (n - 4) \quad (3)$$

We use $F_n(\tau_1)$ to test for the presence of structural breaks with critical values provided in [37]. If structural break exists, we pinpoint the breakpoint ($\tau_1$) by calculating $\tau_1$ below,

$$\tau_1 = \arg\max_{\tau \in [\tau_{1\min}, \ldots, \tau_{1\max}]} F_n(\tau_1) \quad (4)$$

We then apply $\tau_1$ to estimate the piecewise linear trend ($L_t$) using the robust regression based on M-estimation [39] for detrending the time series required by Step 2. If no structural break presents, we estimate a linear trend $L_t$ using robust regression as,

$$T_t + R_t = \alpha + \beta \cdot t, \ t = 1, \ldots, n \quad (5)$$

We calculate the detrended time series as $Y_t - L_t$. Step 2 proceeds with testing for structural breaks in the detrended time series using ordinary least-square-residuals based moving sum (OLS- MOSUM) test [38] (Fig. 2). We employ a three term harmonic model to test and detect the presence of breakpoint in the detrended time series as suggested in [26]. The three term harmonic model is parameterized as,
\[ H_t = \begin{cases} \sum_{k=1}^{3} Y_{1,k} \sin\left(\frac{2\pi k t}{f}\right) + \theta_{1,k} \cos\left(\frac{2\pi k t}{f}\right), & t = 1, \ldots, \tau_2 \\ \sum_{k=1}^{3} Y_{2,k} \sin\left(\frac{2\pi k t}{f}\right) + \theta_{2,k} \cos\left(\frac{2\pi k t}{f}\right), & t = \tau_2 + 1, \ldots, n \end{cases} \]  

where \( f = 23 \) (total number of observations in each year), \( n \) is the total number of observations for all years, \( \tau_2 \) is the estimated breakpoint. If the OLS-MOSUM test indicates presence of any structural break in the detrended data, we fit the three term harmonic model for the observations before and after all \( \tau_2 \)'s, and compute the error sum of squares (ESS\( \tau_2 \)). Then we fit another linear harmonic model for the \( n \) observations and compute the restrict sum of squares (RSS\( \tau_2 \)). Using the estimated ESS\( \tau_2 \) and RSS\( \tau_2 \), we estimate a series of F-statistics (F\( n (\tau_2) \)) as,

\[ F_n(\tau_2) = \frac{(\text{RSS}_{\tau_2} - \text{ESS}(\tau_2))}{\text{ESS}(\tau_2)} \times (n - 12) \]

Finally, we estimate the breakpoint (\( \tau_2 \)) using,

\[ \tau_2 = \arg\max_{\tau_2 \in \{\tau_{2\min}, \ldots, \tau_{2\max}\}} F_n(\tau_2) \]

4. Validation

4.1 Use of simulated NDVI time series to evaluate timing of agricultural land conversion

Simulated data are commonly used in remote sensing for algorithm development and testing [21, 22, 25, 26]. We simulated time series to develop and optimize the change detection algorithm. Our simulation strategy includes estimating the NDVI time series for a pixel representing, a) stable agricultural land, and b) stable urban land. We simulated the NDVI time series for each stable land use using an asymmetric Gaussian function which has the general form:

\[ f(t) = \begin{cases} a \times \left\{\exp\left[\frac{-(t-b)^2}{2c_1}\right], & t > b \\ \exp\left[\frac{-(b-t)^2}{2c_2}\right], & t < b \end{cases} - d \]

where, \( a \) and \( d \) determine the amplitude and \( b \) determines the location of maximum and minimum with respect to time (t). \( c1 \) and \( c2 \) determine the width of the left and right segments of a temporal seasonality curve, respectively. The asymmetric Gaussian function can be used to simulate the time series for one season. However, the MODIS NDVI time series signature for agricultural areas in irrigated and highly productive agricultural areas often shows two distinct
cropping seasons in a given year, “Kharif” (crops planted during the summer and harvested in the fall) and “Rabi” (crops planted during the winter and harvested in the summer). Such agricultural areas are often referred to as “double cropped areas” [40]. The Kharif season extends between June/July and September/October whereas the Rabi season extends between November/December and February/March. We used two asymmetric Gaussian functions to simulate the annual NDVI time series for the two cropping seasons, and one asymmetric Gaussian function to simulate the annual NDVI time series for stable urban area (Fig. 3a). We used an adaptive nonlinear least-squares algorithm to estimate the parameters for each of the asymmetric Gaussian functions, keeping the filtered MODIS NDVI time series as the dependent variable [41].

We simulated the NDVI time series for agricultural land conversion by combining the estimated NDVI time series for stable agricultural land with that of stable urban area. We assumed that the agricultural land conversion is abrupt rather than gradual and hence can be characterized within a 6-month period (Fig. 3b). Accordingly, we simulated the NDVI time series by assigning linear increasing and decreasing weights to the NDVI time series for stable urban and stable agricultural lands, respectively [21, 22]. In this study, we conceptualize agriculture land conversion to have distinct and sequential phases and that each of these is identifiable in a NDVI time-series: a) stable agriculture, b) change event, and c) new non-agricultural land use (Fig. 3b).

We evaluated the performance of the algorithm with simulated NDVI time series for agricultural land conversion. The errors in input time series may result in bias in the breakpoint estimates. Whereas much of the remote sensing accuracy literature focuses on the correct identification of the amount and category of land cover change, significantly less attention is given to the bias introduced when the timing of change is incorrect. When time series remote sensing data are coupled with other time series data for modeling purposes, incorrect identification of when land change occurs will introduce error and bias to the models [8]. In order to evaluate this bias, we generated 40000 NDVI time series for agriculture to urban land conversion with added random noise. We applied our change detection algorithm to all the 40000 simulated time series and estimated the breakpoint, i.e., the Observed, and calculated the error estimate (e_i) as,

\[ e_i = \text{Reference}_i - \text{Observed}_i \]  \hspace{1cm} (10)
Where, $Reference_i$ is the actual timing of change (beginning of the change period). A negative value of $e_i$ indicates that the algorithm estimated the breakpoint after the change whereas a positive value of $e_i$ indicates that the breakpoint is estimated before the actual change. Earlier research efforts at identifying the timing of land cover change used three measures of bias, a) mean, b) skewness, and c) kurtosis [8], which we calculated for each set of $e_i$:

$$\bar{e}_i = \frac{\sum_{d=1}^{N} e_i}{N}$$

(11)

Where, $\bar{e}_i$ is the mean value of the errors, and N is the number of samples.

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

(12)

Where, $\gamma_2$ is the excessive kurtosis, $\mu_4$ is the fourth moment about mean, and $\sigma$ is the standard deviation.

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

(13)

Where, $\gamma_1$ is the skewness, $\mu_i$ is the $i_{th}$ central moment.

### 4.2 Detecting the timing of agricultural land conversion in MODIS NDVI time series dataset (2000-2010)

We implemented a post-classification comparison technique using support vector machines (SVMs) with radial basis function (RBF) kernel [42, 43] to characterize land cover transitions. SVMs are non-parametric binary classifiers that require limited training points and delivers classification accuracy comparable to other machine learning classifiers such as decision trees, neural networks, and others [10, 42, 43]. SVM-based supervised classification has been applied on high temporal frequency MODIS dataset to assess agricultural land abandonment [28]. We applied the SVM-based classification on filtered MODIS NDVI time series (2000-01 and 2009-10) and then compared the classified maps of the two time periods to characterize the change pixels and applied the time series algorithm to detect timing of change (Fig. 4). We selected three
different urban regions in India: Delhi (77.22° E, 28.63° N), Bangalore (77.5° E, 12.97° N) and Ahmedabad (72.56° E, 23.03° N), based on their differences in: a) rate and scale of urban growth, and b) farm sizes in the vicinity of the urban regions.

We selected the training sites in each test site through visual interpretation of temporal Landsat TM/ETM+ images, Google Earth very high resolution (VHR) images (2000-2010) and temporal profiles from the MODIS NDVI time series. We generated a vector layer representing the footprint of a MODIS 250 m pixel and overlaid it over Google Earth and Landsat images. This facilitated in the identification of stable and homogeneous land cover for each MODIS pixel during training site selection. All the training sites were collected as single-pixel samples to prevent autocorrelation effects in multi-pixel samples. We performed the SVM classification in R using caret package that allows for building classification models [44]. We implemented 10-fold cross validation to estimate appropriate values for tuning parameters i.e., σ (scale function) and C (cost value) - controls the complexity of the decision boundary [44].

4.3. Accuracy Assessment of the algorithm to detect timing of agricultural land conversion in MODIS NDVI time series (2000-2010)

We generated 50 random stratified samples in each test site with disproportionate stratification as ground truth to evaluate the results from the automated change detection algorithm. 20 % of such samples were no agricultural to non-agricultural land use transition and the rest 80 % of the samples were agriculture to non-agricultural land use transition. The rationale for using the disproportionate sampling strategy is to evaluate the performance of the automated change detection algorithm to identify the breakpoint rather than the characterization of change type. We followed a double blind procedure to label the year of the breakpoint for all the samples. Each sample was labeled by three remote sensing analysts. In case of disagreement among the remote sensing analysts, we assigned the breakpoint using the majority rule. We analyzed the overall accuracy of the algorithm from the reference samples (n) for which the timing of change was labeled by the analysts. We were also interested in evaluating the algorithm for any consistent bias in the estimated timing of change; therefore we used the series of error (e_i) from all the test sites and estimated the confidence interval for mean value of error, and excessive kurtosis at 95 % confidence interval using basic bootstrap (1000000 replications) with non-parametric ordinary bootstrap method.
5. Results

5.1 Evaluating the algorithm with simulated NDVI time series

The main characteristics of the automated algorithm are revealed by testing its performance under various scenarios using simulated NDVI time series. Our focus here is mainly on two aspects: a) performance of the algorithm under various noise levels (σ) in the simulated NDVI time series for agricultural land conversion and, b) ability of the algorithm to obviate detecting breakpoint in simulated time series for stable land use types i.e., agriculture and urban. The algorithm consistently detected the breakpoint in all the simulated change time series, irrespective of presence or absence of noise (Fig. 5). The OLS-MOSUM test indicates the presence of at least one structural break in all the simulated change time series, at 0.05 level of significance. This implies that irrespective of noise, the algorithm obviates the possibility of any omission errors during breakpoint estimation. In the absence of noise, the algorithm detects the breakpoint in simulated change time series with 100% accuracy. This implies that our three term harmonic model specification is appropriate in context of detecting the timing of agricultural land conversion. Further, in the absence of noise, the algorithm detects no breakpoint in the simulated time series for stable agriculture and urban land use types. The algorithm is thus efficient in avoiding commission errors at least in the absence of noise. Conversely, presence of noise in the simulated time series for stable agriculture and urban area, results into commission errors as the breakpoints are detected in the time series for stable land use types (Table 1). These commission errors are a result of structural break detected by the OLS-MOSUM test. The algorithm is sensitive towards a bandwidth parameter that has to be specified in the OLS-MOSUM test for identifying presence or absence of structural break. The OLS-MOSUM test detects the structural break by analyzing fixed number of residuals in a moving window whose size is determined by the bandwidth parameter $h \in (0, 1)$. In case of agricultural land conversion, a constant bandwidth parameter cannot be used to test the presence of structural break since there are many possible trajectories of agricultural land conversion, for example, episodic, seasonal, abrupt [14]. A possible solution to this problem is iteratively test for the presence of structural break while varying the bandwidth parameter, $h \in (0, 0.01, 0.02, \ldots, 0.99)$ (Fig. 6). In case of NDVI time-series for agricultural land conversion, out of all the MOSUM paths (i.e., the empirical fluctuation processes for all values of bandwidth parameter), at least one of the MOSUM paths
will show a strong shift around the actual timing of change thus rejecting the null hypothesis of parameter constancy. Whereas, in case of NDVI time series for stable land cover, the OLS-MOSUM test will accept the null hypothesis of parameter constancy for all values of bandwidth parameter. In the example presented in Fig. 6, the OLS-MOSUM test reject the alternative hypothesis of presence of a structural break (p-value > 0.05, $h \in (0,1)$) in case of stable agriculture, whereas the OLS-MOSUM test reject the null hypothesis of absence of a structural break at $h = 0.03$ (p-value < 0.05) in case of agricultural land conversion. The iterative testing thus helps in detecting structural breaks and to some extent in differentiating stable land cover with land cover transitions. However, with the presence of noise the stable land cover can also be detected as change leading into commission errors (Table 1).

Another manifestation of noise in the input time series is that it degrades the overall performance of the algorithm while introducing error in the estimated timing of change ($e_i$) (Fig. 5). The overall accuracy i.e., percentage of cases where algorithm determines the breakpoint accurately ($e_i=0$, $i \in (1, 40000)$), decreases with increasing level of noise. We observed that with increasing noise levels, the algorithm tend to identify the breakpoint after the actual timing of change ($\overline{e}_{i} < 0$). The increasing noise level also has an impact on $\gamma_2$. The error distribution tend to be positively skewed with increasing level of noise, $\gamma_1 = -26.64$ ($\sigma = 0.05$) and $\gamma_1 = 4.79$ ($\sigma = 0.3$) (Not shown in Fig. 5). Also, increasing noise levels decreases the degree to which the algorithm detects the break near the actual timing of change, $\gamma_2 = 1232.63$ ($\sigma = 0.05$) and $\gamma_2 = 0.47$ ($\sigma=0.2$).

### 5.2 Applying the algorithm to MODIS NDVI time series for 2000-2010

Here we illustrate an example of a MODIS NDVI time series (2000-2010) for a pixel transitioning from agriculture to non-agriculture land-use type (Fig 7a). The stable agriculture phase in the time-series shows an annual bimodal curve, representing the two cropping seasons. After the change event, the bimodal curve transitions to an annual unimodal curve, representing the new non-agricultural land-use. This example also conforms to our simulated data for the presence of three distinct and sequential phases in agricultural to urban land conversion. The algorithm first removes the trend in the input time series and then tests for presence of structural break in the detrended time series. The trend component along with the remainder component is thus isolated from the seasonal component (Fig. 7b) and tested for presence of structural break. The supF test exhibited a structural break (in 2005) in the trend and thus fits a segmented model.
on the $T_i + R_i$ component using robust regression based on M-estimation, to estimate $L_t$. $L_t$ is further used to detrend the input time series (Fig. 7c). The OLS-MOSUM test indicates presence of structural break in the detrended time series at 0.05 significance level ($h = 0.1$) (Fig. 7d). Further, the algorithm detected the timing of agricultural land conversion in year 2004, i.e., during the change event (Fig 7e). The estimated breakpoint can be validated from temporal Google Earth VHR images in which recent construction activity is noticed in the year 2004. It is worthwhile to mention that in the absence of a detrending routine, the algorithm still detected a statistically significant structural break in the example presented above, but the estimated breakpoint occurred in non-agricultural land-use phase. This emphasizes on the importance of detrending the time series before breakpoint estimation i.e., to impart stationarity in the input time series. This example thus illustrates how the algorithm could be used to detect timing of agriculture land transition from MODIS NDVI time series data.

5.3 Accuracy of the algorithm to detect timing of agricultural land conversion

The accuracy of the algorithm to detect the timing of agricultural land conversion is 70.27%, 69.69 % and, 74.35 %, for Delhi, Ahmedabad, and Bangalore test sites, respectively (Fig. 8). Given 9 possible timing of change (2001-2009), the algorithm successfully identified the exact timing of change with approximately 70 % accuracy. Assuming the gradual nature of agricultural land conversion and allowing for bias of ±2 years, we noted that the accuracy of the algorithm to detect timing of agricultural land conversion is approximately 83 %. The relatively low accuracy to detect exact timing of change in Delhi and Ahmedabad test site is attributed to the pixels exhibiting small scale and gradual agricultural to non-agricultural land conversion. We observed that most of the pixels exhibited rapid agricultural land conversion in the Bangalore test site, thus leading to higher overall accuracy. In our analysis with MODIS NDVI time series dataset, we did not observe any consistent bias associated with the estimated timing of agricultural land conversion across the three test sites. A majority of errors are associated with the change before the actual change, in Delhi and Ahmedabad (mean value of errors is 0.72 and 0.3 for Delhi and Ahmedabad). Whereas the errors are dominated due to the breakpoints detected after the actual change (mean value of errors is -0.33), in Bangalore test site.

We assessed the relative performance of the algorithm and the bias associated with estimated timing of agricultural land conversion, with confidence intervals estimated for mean value of
errors and excessive kurtosis. The confidence intervals estimated using bootstrap resampling (with all samples from the three sites, n= 109) for mean value of errors are 0.54 and -0.1. This implies that the breakpoints are generally estimated before the actual timing of change. This bias as opposed to our findings from the simulation experiments is attributed to small scale and gradual transitions, as in the presence of noise the algorithm is biased towards estimating the breakpoint after the actual change. The confidence intervals estimated using bootstrap resampling (with n = 109) for excessive kurtosis are 5.24 and 0.44. This further suggests that the breakpoints identified by the algorithm are mostly concentrated near the actual date of change relative to a normal distribution.

6. Discussion

With high quality, high temporal frequency MODIS NDVI data, the proposed algorithm accurately identifies the timing of agricultural land loss to urbanization. It is intuitive, requires little manual intervention, and is nearly fully automated. There are several advantages of the automation. First, it enables processing of very large datasets, either in spatial extent or through time. Second, it reduces mistakes due to interpretation or human error. Our results indicate that the algorithm can detect the timing of change with an accuracy of ~ 70%. Our simulation results confirm that in the presence of noise the estimated timing of changes can be either earlier or later than the timing when actual changes occur. However, we found that even in the presence of very high noise levels this bias in the estimated timing of change is still low. For example, at a noise level with \( \sigma = 0.2 \), we found that the bias is approximately \( \pm 1 \) year. Therefore, other factors such as small farm size, changes in cropping patterns and rate of development process, may be attributed to larger bias in the estimated timing of change. Furthermore, our analysis suggests that in case of gradual change, the algorithm may estimate the breakpoint with a larger bias. Thus some of the disagreement between our results and the ground truth data can be attributed to potential noise in the 16-day MODIS NDVI time series which have not been completely corrected for angular effects. However, part of the disagreement is also attributed to a) gradual transitions with prolonged change event phase limiting precise estimation of the breakpoint, and b) limited number of historical high resolution images available in Google Earth to allow analysts to accurately label the timing of change on the ground. The prolonged change event phase in case of gradual urban conversion of agricultural land is due to slow development rates.
Agricultural land conversion in such cases may occur in a step-wise fashion as opposed to gradual changes commonly observed in forest degradation where length of change event phase is characterized by a continual decreasing trend [16, 25]. Small scale development activity may cause agricultural land conversion in part of the pixel in a given time. The next phase of conversion in which the remaining agricultural land is converted (wholly or partly) may occur after a few years. This step-wise land-use transitions result into deviation from the assumption of sequential phases in land conversion process and limit the application of most change detection algorithms. Dealing with such scenarios requires that the algorithm first differentiate between abrupt and gradual change. Further in case of gradual change, sub-pixel analysis is required to compute within pixel agricultural area estimate over time [29]. Nevertheless, our proposed algorithm can also be applied to higher spatial resolution data such as WiFS, AWiFS, Landsat, and others, to alleviate the effects of gradual change and increasing the overall accuracy whereas application of our proposed algorithm on coarse spatial resolution data is still important due to its alarm capability to identify potential agricultural to non-agricultural change regions through time. Although we evaluated the algorithm to identify the timing of urban conversion of agricultural lands, the timing of other historical land-use and land cover transitions such as, deforestation and agricultural expansion, can also be monitored using our proposed algorithm. Our simulation results confirm that the algorithm does not produce any omission errors and to a large extent does not estimate a breakpoint for stable land use. Thus at minimum, the algorithm can be used in conjunction with a two-point change detection method to subsequently identify the timing or date of change. In addition, this highly automated algorithm can be extended for change analysis with other types of time series data with different temporal frequencies, such as daily, 8-day, 10-day, and others.

7. Conclusion
We have presented a simple and robust automated algorithm to identify timing of agricultural land use transitions with high temporal frequency vegetation indices dataset. Our study confirms that the techniques developed in time series econometrics literature can be applied to remote sensing images to detect timing of historical land use transitions such as, agricultural land
conversion. The time series analysis approach used in this study enables automated detection of the timing of agricultural land conversions. The major advantages of the algorithm include a) rapid and automated assessment of agricultural land loss, b) nominal computational requirements, c) reliable estimate of timing of change event, d) applicability to different landscapes. The algorithm presented in this paper is not designed to replace the existing methods of change detection but is intended to supplement the change detection methods to identify the timing of change. The algorithm can be used in an integrated manner to reconstruct land-use and land cover change history.

References:


List of Tables:

Table 1: Commission errors in the form of breakpoints detected in simulated time series of stable agriculture and urban land use type, with noise levels $\epsilon$ (0, 0.05, 0.1, and 0.2)

List of Figures:

Figure 1: Raw and filtered 250 m MODIS NDVI time series at 16-day intervals (2000-2010). The thin dotted line represents the raw NDVI time series from MOD13Q1 NDVI dataset and the solid red line represents the NDVI time series filtered by applying the adaptive Savitzky-Golay filter in the TIMESAT program. The anomalous drops in the raw NDVI time series are filtered.

Figure 2: Overview of the automated change detection algorithm to estimate timing of change.

Figure 3: a) The dotted thin line shows the real NDVI time series whereas the solid red line shows the simulated NDVI time series for stable agriculture area and stable urban area. The temporal signature for the two agricultural seasons i.e. Kharif and Rabi are evident in NDVI time series for the stable agricultural area. b) The simulated time series for agriculture to urban land transition with the change period of 6 months (highlighted in blue color)

Figure 4: Overview of the methodology followed to identify agricultural to non-agricultural land use transitions and subsequently the timing of change.

Figure 5: a) Distribution of errors in the breakpoint estimates for simulated agriculture to non-agriculture land change time series with added random noise ($\mu = 0$, $\sigma \in (0, 0.05, 0.1, \text{and } 0.2)$). b) Mean, Excessive Kurtosis, and Skewness for each set of $e_i$ series with $\sigma \in (0, 0.05, 0.1, \text{and } 0.2)$.

Figure 6: p-values from OLS-MOSUM test for simulated NDVI time series of agricultural land conversion and stable agricultural land use, with bandwidth parameter $(h) \in (0, 0.01, 0.02, \ldots, 0.99)$. The solid red line shows the boundary at 0.05 significance level for presence of structural break.

Figure 7: a) MODIS NDVI time-series ($NDVI_t$) for a pixel transitioning from agriculture to urban land use type, b) Seasonal ($S_t$) and Trend ($T_t$) + Remaider ($R_t$) components obtained from
additive decomposition of NDVI, c) Detrended NDVI time series (NDVI_t - L_t) d) OLS-MOSUM test (h=0.1) detects structural break in the detrended time-series, and e) Series of F-statistics are computed and the breakpoint is estimated in the year 2004.

Figure 8: Distribution of errors in a) Delhi (n = 37), b) Ahmedabad (n=33), c) Bangalore (n=39), and d) All the samples (n = 109).
Table 1: Commission errors in the form of breakpoints detected in simulated time series of stable agriculture and urban land use type, with noise levels $\epsilon (0, 0.05, 0.1, \text{ and } 0.2)$

<table>
<thead>
<tr>
<th>Simulated land use type</th>
<th>Noise level ($\sigma$)</th>
<th>Commission Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>0</td>
<td>0 %</td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.05</td>
<td>15.3 %</td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.1</td>
<td>16.9 %</td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.2</td>
<td>21.4 %</td>
</tr>
<tr>
<td>Urban</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Urban</td>
<td>0.05</td>
<td>19.6 %</td>
</tr>
<tr>
<td>Urban</td>
<td>0.1</td>
<td>19.91%</td>
</tr>
<tr>
<td>Urban</td>
<td>0.2</td>
<td>20.06%</td>
</tr>
</tbody>
</table>
Figure 1:
Figure 3:

(Images of graphs showing NDVI trends in Agriculture and Urban areas, with simulated and real data comparisons. Source: Google Earth.)
Figure 4:
Figure 5:

<table>
<thead>
<tr>
<th>Error Standard Deviation (σ)</th>
<th>Mean</th>
<th>Excessive Kurtosis</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.05</td>
<td>-0.0034</td>
<td>1232.63</td>
<td>-26.64</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.19</td>
<td>11.92</td>
<td>-2.28</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.94</td>
<td>0.47</td>
<td>-0.60</td>
</tr>
</tbody>
</table>
Figure 6:
Figure 7:

1. supF-test indicates a structural break in the linear model
2. Seasonal linear trend is computed (1,1) and NDVI time series is detrended (NDVI\(t+1\)).
Figure 8:

- **a) Distribution of Errors (Delhi)**
  - Overall Accuracy = 70.27%
  - Change detected after actual change

- **b) Distribution of Errors (Ahmedabad)**
  - Overall Accuracy = 69.69%
  - Change detected after actual change

- **c) Distribution of Errors (Bangalore)**
  - Overall Accuracy = 74.35%
  - Change detected after actual change

- **d) Distribution of Errors (All Samples)**
  - Overall Accuracy = 71.55%
  - Change detected after actual change