Time Series Change Detection using Segmentation: A Case Study for Land Cover Monitoring

Varun Mithal, Zachary O'Connor, Karsten Steinhaeuser, Shyam Boriah, Vipin Kumar
Comp. Sci. & Engineering, University of Minnesota
{*mithal,ksteinha,sboriah,kumar*}*@cs.umn.edu,*
oconn399@umn.edu

Christopher S. Potter, Steven A. Klooster
NASA Ames Reseach Center
chris.potter@nasa.gov,
sklooster@gaia.arc.nasa.gov

***Abstract*—Automatic identification of changes in land cover from remote sensing data is a critical aspect of monitoring the planet's ecosystems. We use time series segmentation methodology for detecting land cover changes from Moderate Resolution Imaging Spectroradiometer-based vegetation index. In this paper, we investigate segmentation scores based on difference between models and propose two approaches for normalizing the difference based score. The first approach uses permutation testing to assign a *p-value* to model difference. The second approach builds on bootstrapping methodology used in statistics which estimates the null distribution of complex statistics whose standard errors are not analytically derivable by generating alternative versions of the data by a resampling strategy. More specifically, given a time series with either a single or two segments, we propose a method to estimate the distribution of model difference statistic for each segment. The proposed approach allows normalizing model difference statistic when complex models are being used in the segmentation algorithm. We study the strengths and weaknesses of the two normalizing approaches in the context of characteristics of land cover data such as seasonality and noise using synthetic and real data sets. We show that relative performance of normalization approaches can vary significantly depending on the characteristics of the data. We illustrate the utility of these approaches for detection of deforestation in Mato Grosso (Brazil).**

I. Introduction

Quantifiable knowledge about changes occurring in land cover and land use at a global scale is key to effective planning for sustainable use of diminishing natural resources such as forest cover and agricultural land. Accurate and timely information about land cover and land use changes is therefore of significant interest to earth and climate scientists as well as policy and decision makers. Due to the importance of land cover and land use change detection, it has been a topic of active research in the remote sensing community. The bulk of work in land cover change detection from remote sensing data involves image comparison methods [1], [2] which have well known limitations including need for training data which makes them difficult to apply on a global scale [3].

Recently, global time series data sets, such as Moderate Resolution Imaging Spectroradiometer Enhanced Vegetation Index (EVI), have become publicly available [4] and have been used to identify changes in vegetation cover [3]. Time series-based approaches look at a longer context and therefore can be utilized for providing fine grained information about land cover dynamics that is necessary to quantitatively assess the carbon impact of land cover changes [5]. Hence there is increasing interest in time series-based approaches to change detection in vegetation data [6], [7], [8], [9], [10], [11], [12], [13]. Figure 1 shows illustrative examples from four different locations that have land cover changes. The focus of this paper is on time series segmentation algorithms that can identify such changes as well as the time of change. Moreover, in this paper we focus on time series with either no change or a single change and leave the problem of detecting multiple changes as future research.

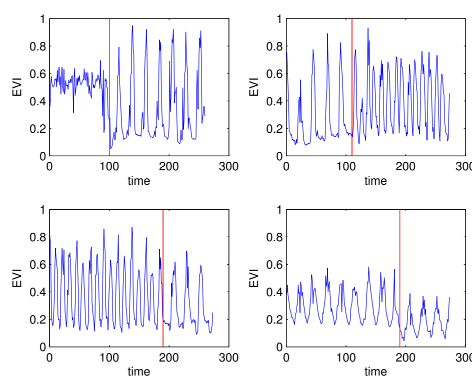


Figure 1. Illustrative examples of EVI time series corresponding to locations where a land cover change occurred.

Segmentation of a time series attempts to divide it into homogeneous subsequences, such that each of these segments are different from each other. A typical segmentation framework involves selecting a model that is used to represent the segment. The algorithm then looks at candidate split points using a search strategy that can be top-down or bottom-up [14]. At each candidate, models are constructed for both segments, and the *segmentation score*, a statistic that indicates the utility of segmenting at that time step is computed. If a given location has not undergone a change, then we expect the models to look very similar and the segmentation algorithm should give the time series a low score. On the other hand, if a land cover change (such as in examples shown in Figure 1) has occurred, the segmentation algorithm should give the time series a high score.

Segmentation algorithms often use the difference between the models of the two segments as a *segmentation score*. As an example, if the model selected is parametric (such as piecewise constant, or piecewise linear model), then the parameters of the model fitted on two segments can be compared (to compare two segments fitted with a linear model the difference between the slopes is used as a statistic). The intuition is that when this statistic (i.e. difference between slopes) has a high value then the two segments are distinct enough to segment the time series.

The same magnitude of the model difference statistic can have different significance for two time series with different characteristics. In the context of land cover monitoring, high inter-annual variability and noise seen in farms, grasslands and tropical areas can cause model difference to be high even in the absence of a land cover change. Therefore, using the magnitude of model difference as segmentation score can lead to large number of false alarms in regions where farms and grasslands are the dominant vegetation. For some of the commonly used models, such as piecewise constant and linear models, the statistic used, such as the difference in piecewise constant approximations and slopes of segments, has a known statistical distribution under the null model and therefore the standard error of the statistic can be analytically computed. This allows using the *t-statistic* for normalization of model difference statistics such as the difference in mean [15] and the difference in slopes [16].

For many domains, including land cover change detection, piecewise linear or constant approximations are inadequate to capture the underlying model [17]. The plots in Figure 1 indicate the following characteristic pattern present in vegetation model changes. There is a recurring annual model that repeats with some variations every year for several consecutive years. When a land cover change occurs, the annual vegetation model changes at the change time (marked as red vertical line). In this case, when a seasonal model has to be used, the standard error for the model difference statistic is not known analytically as in the case of a piecewise constant model [15] or a linear model [16], and therefore *t-statistic* based normalization cannot be used.

A. Our Contributions

- In this paper, we investigate segmentation scores based on the difference between models and propose two approaches for normalizing the difference based score. The first approach is adapted from the $Pete$ algorithm proposed in [18] and uses permutation testing to assign a *p-value* to model difference. More specifically, we generate random permutations of time series, compute the maximum model difference for each random permutation, and use this distribution to assign an empirical p-value to the model difference observed in the original time series. The second approach builds on the bootstrapping methodology used in statistics which estimates the null distribution of complex statistics whose standard errors are not analytically derivable by generating alternative versions of the data by a resampling strategy. More specifically, given a time series with either a single or two segments, we propose a method to estimate the distribution of the model difference statistic for each segment. We show that this distribution can be used as a null distribution of model differences due to intra-segment variations, and the observed model difference statistic between the two segments is normalized with respect to this distribution. The proposed approach thus allows normalizing the model difference statistic when complex models other than piecewise constant and piecewise linear model are being used in the segmentation algorithm.
- We study the strengths and weaknesses of the two normalizing approaches in the context of characteristics of land cover data such as seasonality and noise using synthetic and real data sets. We show that the relative performance of normalization approaches can vary significantly depending on the characteristics of the data.
- We illustrate the utility of these approaches for detection of deforestation in Mato Grosso (Brazil).

II. Earth Science Data

Global remote sensing data sets are available from a variety of instruments at different spatial resolutions as a sequence of global snapshots of measurement values. In principle, the segmentation algorithms discussed can be applied to any geospatial dataset that features regular, repeated observations, consistent image registration and well-defined composite indicators of vegetation. In this study, we used the Enhanced Vegetation Index (EVI), a data product based on measurements taken by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor onboard NASA's Terra satellite [4]. EVI essentially measures the "greenness" signal (area-averaged canopy photosynthetic capacity) as a proxy for the amount of vegetation at a particular location. MODIS algorithms have been used to generate the EVI index at 16-day intervals at 250-meter and 1 km spatial resolution from February 2000 to the present. We process sequences of global snapshots of these indices to construct a time series for each pixel on the globe. Savitzky-golay filtering used in [19] was used with parameters for polynomial degree as 2 and window size as 7.

III. Our approach

In this section we describe the components of our segmentation approach to detect land cover changes from EVI data. A typical segmentation algorithm consists of three components: (1) model choice, (2) search strategy and (3) segmentation score. Table I describes the notations that are used in this paper while discussing different algorithms.

A. Model choice

For describing a time series segment, a simpler generative model for the data in the segment is assumed or chosen from a given set of available choices of models [20]. When this model is fitted to the segments, it provides both a measure for homogeneity as well as a mechanism to compare segments.

We use a nonparametric model described in [17] for vegetation data that accounts for the presence of seasonality with known time period sl. It consists of a seasonal and a natural variation component. The seasonal component is written as $A = (A_1, A_2, ..., A_{sl})$ and can be estimated for time series S as mean of the observations (details in Algorithm 1). Nonparametric models use functions of the input time series as the model and do not approximate the data with other functions. The difference between the observations and the estimated nonparametric model for the time series (i.e. the residuals) are considered the error component.

Algorithm 1 $computemodel(S, sl)$
Calculates A for segment S

Require: time series segment S, season length sl
Ensure: A the model estimate for S
  for $k = 1$ to sl **do**
    $A_k \leftarrow mean(S^k)$
  end for

B. Search strategy

We use a top-down search strategy that computes a segmentation score for each time t. An EVI time series is assigned the maximum segmentation score across all timesteps as the change score and corresponding time t as the change point.

C. Segmentation score

The segmentation score at t indicates the utility of dividing a time series into two segments at t. In this study we investigate the model difference-based segmentation score. In addition, we propose two approaches to normalize the score based on natural variations of the time series. The details of the model difference scoring method and proposed normalizing approaches is presented in the next section.

IV. Different scoring schemes

Here we provide the details of different segmentation scores we use in this study for change detection.

A. Model difference (MD)

In MD we use L1 distance between the models for the two segments as our segmentation score. The intuition behind this is that if a land cover change occurs, then two segments with distinct models are present in the time series. We use L1 distance because it captures changes in shape as well as amplitude, but other distance functions such as L2 norm or correlation can also be used. More formally in the MD scheme, for each candidate time step t we compute the models A_{left} and A_{right} for segments $S_{1,t}$ and $S_{t+1,l}$. The change score for S at t is given as the L1-distance between A_{left} and A_{right} and is referred to as MD_t. We assign the maximum model difference score MD_t over time t as the change score for that time series (MD_{max}).

B. MD with permutation testing (MDPerm)

In this method, we adapt the strategy of the $Pete$ algorithm proposed in [18] for normalizing model difference scores. More specifically, we create random permutations of the data, and for each random permutation of S the highest model difference is computed (referred to as MD^i_{max} corresponding to the i^{th} permutation of S). The p-value assigned to the maximum model difference from actual data (MD_{max}) is the fraction of MD^i_{max}s greater than MD_{max}. If the difference in the models for the time series is significant then it will be higher than the MD^i_{max} and thus the p-value will be low. On the other hand, if the difference is insignificant compared to the differences that can be achieved due to random chance for the time series with similar mean and variability characteristics, some of the MD^i_{max}s will exceed MD_{max} and it will get a higher p-value (i.e. lower significance).

C. MD with bootstrapping (MDBoot)

Random permutation of time series disregards the seasonal structure that is prominent in EVI data. In this section, we propose a framework that can be used to generate a distribution of inter-annual distances within a time series segment and use it to normalize MD scores with respect to the natural variability while retaining the seasonal structure of time series.

1) Main intuition: Figure 2 shows two time series: the first is a stable tropical forest that is resilient to inter-annual fluctuations and the second is a farm that has a significantly higher degree of inter-annual natural variations. EVI time series in Figure 2 have 11 years of data. The model differences corresponding to the $\binom{11}{2}$ pairs of annual segments in the time series are computed and their histograms are plotted adjacent to the time series. The histogram for the forest time series has lower mean and smaller spread while

Notation	Description
S	a time series which is an ordered collection of real values.
sl	the number of time steps in one period of the time series and is referred as season length.
S_t	the value of time series S at time step t.
$S_{i,j}$	a segment i.e. the portion of time series S from time step i to j.
l	the length of a segment and for $S_{i,j}$ is equal to $j-i+1$.
A	an annual segment and is a vector of sl values.
A_k	the value of k^{th} index of A.
S^k	the collection of values S_t of all t that belong to k^{th} season (for example January).

Table I
A TABLE WITH DESCRIPTION OF NOTATIONS.

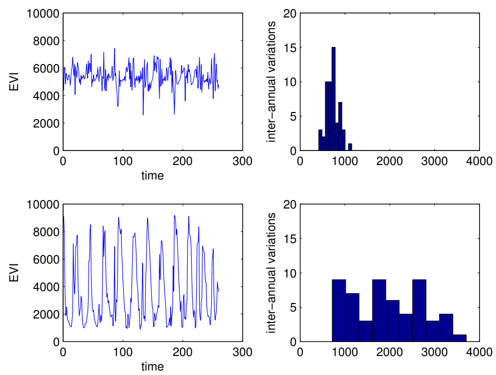


Figure 2. Two different EVI time series with different degree of inter-annual variations. The histograms show the distribution of the pairwise inter-annual L1 distances for the adjacent EVI time series.

the histogram for farm has a higher mean and a wider spread. Therefore significance of model difference scores between two segments in these two different land cover types should be assessed in context of these natural inter-annual variations. A model difference that can belong to the distribution of the pairwise intra-segment differences cannot be considered significant. It is challenging to use this approach due to the small sample size of the segments available for computing the distribution of inter-annual variability. The EVI time series have total of 11 annual segments and therefore the number of annual cycles in a segment will vary between 3 and 10, which is insufficient to robustly estimate mean and standard deviation. To overcome this challenge we propose a bootstrapping method that can be used to estimate the mean and standard deviation of the distribution for the pairwise annual model differences for a segment.

2) Bootstrapping: Bootstrapping is a resampling method that allows for estimating properties of an estimator when sampling from an approximate distribution [21]. If the data is independent and identically distributed, then resampling with replacement can be used for generating alternate versions of data (that could have been observed). Traditional bootstrapping is therefore not suitable for time series because they lack i.i.d. property. Time series have structures such as seasonality and temporal ordering which need to be preserved while generating the alternate versions of data.

Here we describe our bootstrapping framework to generate resamples of annual segments in a time series segment that can be used to compute the distribution for model differences within a segment. The distribution of these model differences is used to normalize the MD score. Under the null hypothesis, i.e., stable recurring annual models with natural inter-annual variations, Algorithm 1 is used to estimate the model A for segment S. To generate a resampled value we randomly choose a value from the set of values in S belonging to season k. This is done for each k from 1 to sl and gives us a resampled annual segment. Similarly, another annual segment is generated. The L1 distance is computed between the two resampled annual segments. This is repeated N times and gives us a distribution of N model differences. The distribution is assumed to be Gaussian and the maximum likelihood estimates for mean and standard deviation are computed. The specifics of the bootstrapping algorithm is presented in Algorithm 2.

Algorithm 2 $computevariability(S, N, sl)$
Calculates μ and σ of $Dist_{null}$ for segment S using Bootstrap

Require: time series segment S, number of bootstrap steps N, season length sl
Ensure: Estimates of μ and σ of $Dist_{null}$
  for $i = 1$ to N **do**
    for $k = 1$ to sl **do**
      $sample1^i_k \leftarrow$ randomly choose from S^k
      $sample2^i_k \leftarrow$ randomly choose from S^k
    end for
    $Dist_{null}(i) \leftarrow sample1_i - sample2_i$
  end for
  $\mu \leftarrow mean(Dist_{null})$
  $\sigma \leftarrow sd(Dist_{null})$

3) Algorithm: In the $MDBoot$ algorithm for each time step t, the time series is partitioned into two segments $S_{1,t}$ and $S_{t+1,l}$. The models A_{left} and A_{right} for the left and right segments are computed and their model difference is calculated. The intra-segment annual variability is computed using the bootstrapping framework discussed. The significance of the model difference is calculated using the distribution of inter-annual variations from the left and right segments. The distributions are assumed to be Gaussian, and the score at time t is computed as the z-statistic of the model difference and the estimates for mean and standard deviation from left and right segments. This gives two z-scores corresponding to the left and right segments, and the maximum of the two scores is used as the segmentation score at time step t. Note that if one of the segments is less than 3*sl in length, then no variability statistics for that segment is computed and the score for that segment is considered to be 0. The values of t range from sl+1 to l-sl, because model computation requires at least one annual segment. Thus no changes occurring in the first and last years can be detected.

Algorithm 3 $MDBoot(S, N, sl)$
Calculates $score$ and cpt for time series S using $MDBoot$

Require: time series S, number of bootstrap steps N, season length sl
Ensure: $score$ and cpt for time series S
  $l \leftarrow length(S)$
  for $t = sl + 1$ to $l - sl$ **do**
    $A_{left} \leftarrow computemodel(S_{1,t}, sl)$
    $A_{right} \leftarrow computemodel(S_{t+1,l}, sl)$
    $modeldiff \leftarrow A_{left} - A_{right}$
    if $t \geq 3sl$ **then**
      $\mu_{left}, \sigma_{left} \leftarrow computevariability(S_{1,t}, N, sl)$
      $score_{t,left} \leftarrow \frac{modeldiff - \mu_{left}}{\sigma_{left}}$
    end if
    if $t \leq l - 3sl$ **then**
      $\mu_{rgt}, \sigma_{rgt} \leftarrow computevariability(S_{t+1,l}, N, sl)$
      $score_{t,right} \leftarrow \frac{modeldiff - \mu_{rgt}}{\sigma_{rgt}}$
    end if
    $score_t \leftarrow max(score_{t,left}, score_{t,right})$
  end for
  $score \leftarrow max_t(score)$
  $cpt \leftarrow argmax_t(score)$

V. Experimental Evaluation

We generate synthetic time series with characteristics similar to those observed in EVI signals of some regions and use them to understand the strengths and weaknesses of the scoring methods discussed earlier. In particular, we present experiments to illustrate the impact of noise and seasonality on the different scoring schemes we study. We also evaluate the different segmentation scores on a validation data set from Mato Grosso in Brazil.

A. Evaluation Methodology

The segmentation algorithms assign a score to each location, and the locations are ranked according to the descending order of their segmentation score. The algorithm flags the top n ranked locations as change events and the lower ranked locations as unchanged. By computing the intersection with the validation data, we find the number of true positives (TP_n), false positives (FP_n), true negatives (TN_n) and false negatives (FN_n). Our evaluation of the performance of the change detection algorithms is based on computation of receiver operating characteristic (*ROC*) curve using the *TPR* and *FPR* [22] and given by:

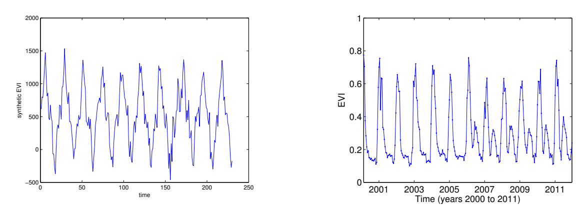
$$\text{TruePositiveRate,}\ \ TPR_n = \frac{TP_n}{TP_n + FN_n}$$

$$\text{FalsePositiveRate,}\ \ FPR_n = \frac{FP_n}{TN_n + FP_n}$$

To compare the relative performance of different techniques, we plot the *ROC* curve for the ranked list of pixels for the values $1 \leq n \leq P$. An ideal change identification algorithm should have a *TPR* of 1 and *FPR* of 0.

B. Synthetic Data

We generate synthetic data using the generative function $A+Bsin^3(wt)+CN(0,1)$ where N(0,1) is a random sample from a standard normal distribution. A changes the mean of the time series, B changes the seasonal component and C changes the amount of noise in data. All synthetic time series have 230 time steps corresponding to 10 years of data with a season length of 23 steps. Figure 3(a) shows a synthetic time series generated using this function.

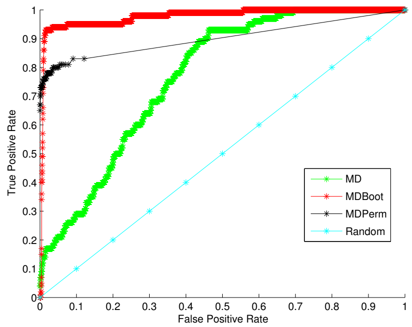


(a) Example of synthetic time series generated.

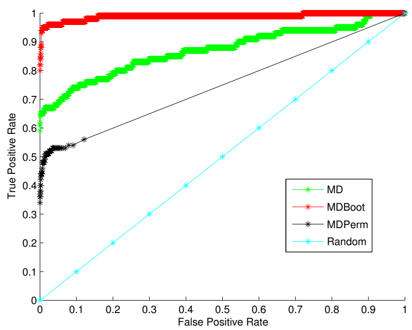
(b) EVI time series that represents a typical change in agriculture cycle in Zimbabwe and other parts of Africa.

Figure 3. Example of synthetic and real time series with high seasonality.

1) Advantage of normalizing MD score: High inter-annual variations and noise is observed in land cover types such as farms and grasslands and tropical areas. MD scores are often high for time series with these characteristics and therefore a large number of false changes are detected if MD scoring is used for areas with farms and grasslands as dominant vegetation. We find that normalizing MD score using permutation testing ($MDPerm$) or by modeling intra-segment variations ($MDBoot$) is useful when segmentation is used for such areas. To illustrate this fact we create a data set of unchanged time series, some of which have high noise (large C). $N1$ is a data set of 5000 unchanged time series that are generated by using A=500, B=750 and varying noise



(a) The ROC curve shows the advantage of normalizing MD score using the two proposed approaches. We see that $MDBoot$ and $MDPerm$ have better performance than MD in presence of time series with differences in noise and change characteristics.



(b) The ROC curve corresponds to performance of the approaches for synthetic time series data with high seasonality. We see that $MDPerm$ has poorer performance in presence of high seasonality in data.

Figure 4. ROC curve for synthetic data. $MDBoot$(*red*), $MDPerm$(*black*) and MD(*green*).

C between 10% to 50% of the amplitude. $P1$ is a set of 100 changed time series that are generated by using A=500, B=750 and noise C=10% of the amplitude. To introduce a change in $P1$, a change time point is chosen between first and last year (i.e. between time step 23 and 207) and amplitude B is decreased by 1% to 100% of the original amplitude for all time series in $P1$. Note that $P1$ data set has changed time series that we want to be detected, and these time series have a moderate noise level. We find that though time series of $N1$ have no actual change in model, they sometimes have MD scores higher than time series in $P1$ which leads to lower accuracy. Figure 4(a) shows the ROC curve for MD, $MDPerm$ and $MDBoot$ for this data set. The ROC curve shows that in presence of noisy time series in data, $MDPerm$ and $MDBoot$ are able to achieve a high TPR at low FPR by distinguishing MD scores due to an actual change in stable time series from MD scores due to noise in unstable time series.

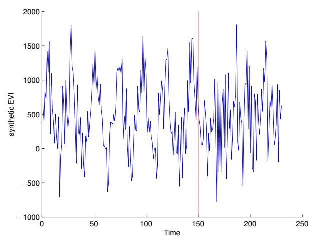
2) Impact of high seasonality in time series: Random permutation of time series destroys the seasonal structure and overestimates the natural variation for land cover types with high seasonal component, and thus $MDPerm$ can assign unusually low scores to these time series (such as Figure 3(b)). To illustrate this fact, we generate a data set $P2$ of high seasonal component using A = 500, B = 1500 and introduced changes between year 2 and year 9 by decreasing the magnitude of B for segment after the change point by 10% to 90%. We use $N1$ as unchanged time series data set. Figure 4(b) shows the ROC curve for MD, $MDPerm$ and $MDBoot$. The ROC curves show that $MDPerm$ *(black curve)* has poorer performance on this data than both MD and $MDBoot$. $MDBoot$ is able to identify these changes because it accounts for seasonal characteristic and resamples on the residuals after removing the seasonal mean. Hence this normalization scheme is particularly useful for identifying changes in farming cycles, given the high seasonality in the associated EVI data.

3) Impact of changes in highly unstable time series: Recall that in data set $N1$, $MDBoot$ is able to significantly outperform MD due to its ability to give higher scores to actual subtle changes in stable time series relative to unchanged, noisy time series. Here we illustrate a weakness of $MDBoot$ in finding changes occurring in noisy time series. We find that for time series with a low signal to noise ratio, $MDBoot$ estimates a high natural variability and thus assign low segmentation scores to actual changes in noisy time series. To quantitatively demonstrate this fact, we generate a data set $P3$ of high noise component using A = 500, B = 750 and C varying between 25% to 50% of amplitude. We introduced changes between year 2 and year 9 by decreasing the magnitude of B for segment after the change point by 10% to 90%. Figure 5(a) is a time series from $P3$ with a change at time step 150. For the unchanged time series data we use the synthetic dataset $N1$. Figure 5(b) shows that $MDBoot$ has poorest performance on this data, while $MDPerm$ is not affected by adding this noise and continues to show an improvement in accuracy over MD. These results show that presence of large amount of noise in the time series data can negate some of the advantages of the normalization scheme used in $MDBoot$.

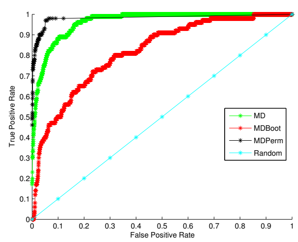
C. Deforestation in Mato Grosso

We present a case study of land cover change detection in Mato Grosso, a state in Brazil which has deforestation events that convert tropical rainforests to agricultural land.

1) Validation Data: Change detection studies are frequently plagued by the lack of accurate and exhaustive ground truth data which forces the evaluation process to be more qualitative in nature. In this study, we have utilized high quality validation data for deforestation generated by an independent source, and are thus able to perform an objective quantitative evaluation. Specifically, we obtained



(a) Example of a synthetic noisy time series with change at time step 150 (red vertical line).



(b) $MDBoot$ (*red curve*) has poorer performance in noisy time series data.

Figure 5. Performance of scoring schemes for noisy time series data.

deforestation boundaries generated by the PRODES (which is considered as the gold standard for ground truth in remote sensing community) for the years 2004 through 2007 for Mato Grosso (state in Brazil). The validation data is in the form of *polygons* which represent the boundaries of deforestation. Our EVI data is georeferenced by the latitude and longitude value for the pixel center. A pixel may have a partial overlap with a polygon. We consider a pixel inside a polygon if the *pixel center* lies inside it, otherwise it is considered outside the polygon. We use the EVI data set for Mato Grosso (Brazil) which has both land cover conversions and unchanged time series. The selected area in Mato Grosso is monitored by PRODES. It consists of about 800,000 locations at 1km spatial resolution. We randomly under-sampled one-tenth of the unchanged time series corresponding to forested areas. The data used has 66,922 unchanged locations and 21,066 locations that are considered deforested by PRODES between the years 2004 to 2007. Since we evaluate different methods using TPR and FPR, undersampling one of the classes does not change the results, but does speed up our evaluation pipeline.

2) Results: Figure 6 shows TPR vs FPR for MD and $MDBoot$ algorithms. We see that both these scores have a significant advantage over random case, showing the utility of the proposed segmentation methodology for land cover change detection problem. We are interested in the recall that algorithms attain at low FPR values and therefore we plot the TPR curve for FPR values till 0.1. We see that $MDBoot$ shows higher TPR than MD at low FPR and the TPR for the two algorithms become equal at FPR of 0.1.

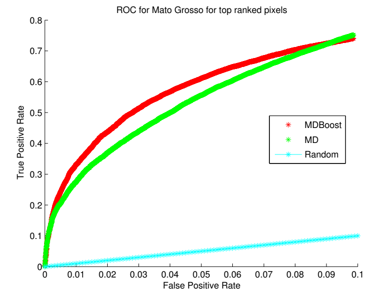


Figure 6. TPR vs FPR for Mato Grosso data till FPR of 10%. $MDBoot$ is *red* curve, MD is *green* curve and *cyan* curve corresponds to random case.

Earlier, we illustrated in the synthetic data experiments that $MDBoot$ does not perform well in presence of high noise. Mato Grosso has tropical rainforests that suffer from poor quality and highly noisy observations due to cloud cover. Thus, we found that $MDBoot$ gives relatively low score to time series with changes that have high noise. We believe that this is responsible for the relatively small improvement of $MDBoot$ over MD for this data. Figure 6 does not show a ROC curve for $MDPerm$. This is because $MDPerm$ assigns significant p-value (p-value of 0) to many unchanged time series that had even a slight variations due to noise or cloud cover. Random permutations evenly spread out this noise and the model estimates for two segments are similar for the permuted time series. Due to this the model difference from the original time series is higher than those from permuted time series, giving a *p-value* of 0 for several unchanged time series.. However, $MDPerm$ has a TPR of 0.8 and FPR of 0.1 at p value $\leq$ 0.01. An FPR of 10% is undesirable as number of changes is usually a small fraction of vast forest cover and it is impossible for end user to look at 10% of data to distinguish actual changes from false positives. We also find that even though PRODES is very effective in mapping land cover changes in Amazon, some of the pixels which show a clear model change in EVI are missed by it. These time series show up as false positives for our segmentation algorithms, but in reality these are deforestation events missed by PRODES.

VI. Conclusion

In this paper, we study a segmentation approach to land cover change identification from EVI time series data and propose two normalization approaches for model difference segmentation score. From our evaluation on synthetic and real data we find that due to the diversity in characteristics of EVI data and the nature of land cover conversions, different normalization approaches have their own strengths and weaknesses.

The proposed normalization methods are described for the model difference statistic of a seasonal nonparametric model. However, this framework is more general and can be used with other model representations such as Fourier and wavelet coefficients. We plan to investigate the utility of this framework with other models in different data sets in our future work. Another research direction is to study other segmentation scores for land cover change detection. As an example, previous segmentation research has used reduction in error to indicate the utility of segmentation of time series into two segments [23], [18]. One approach to normalize the reduction in error based segmentation score is to use relative reduction in error [20]. An interesting study would be to apply the normalization strategies proposed in this paper to reduction in error-based segmentation scores and investigate their performance.

Acknowledgements This research was supported in part by the National Science Foundation under Grants IIS-1029711 and IIS-0905581, as well as the Planetary Skin Institute. Access to computing facilities was provided by the University of Minnesota Supercomputing Institute.

References

[1] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Digital change detection methods in ecosystem monitoring: a review," *International Journal of Remote Sensing*, vol. 25, no. 9, pp. 1565–1596, 2004.

[2] D. Lu, P. Mausel, E. Brondízio, and E. Moran, "Change detection techniques," *International Journal of Remote Sensing*, vol. 25, no. 12, pp. 2365–2401, 2003.

[3] V. Mithal, A. Garg, S. Boriah, M. Steinbach, V. Kumar, C. Potter, S. A. Klooster, and J. C. Castilla-Rubio, "Monitoring global forest cover using data mining," *ACM TIST*, vol. 2, no. 4, p. 36, 2011.

[4] "Land Processes Distributed Active Archive Center," `http://lpdaac.usgs.gov`.

[5] N. Ramankutty, H. K. Gibbs, F. Achard, R. Defries, J. A. Foley, and R. A. Houghton, "Challenges to estimating carbon emissions from tropical deforestation," *Global Change Biology*, vol. 13, pp. 51–66, January 2007.

[6] L. O. Anderson, Y. Malhi, L. E. O. C. Aragão, R. Ladle, E. Arai, N. Barbier, and O. Phillips, "Remote sensing detection of droughts in Amazonian forest canopies," *New Phytologist*, vol. 187, no. 3, pp. 733–750, 2010.

[7] S. Boriah, V. Kumar, M. Steinbach, C. Potter, and S. Klooster, "Land cover change detection: A case study," in *KDD '08: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008, pp. 857–865.

[8] D. Hammer, R. Kraft, and D. Wheeler, "FORMA: Forest monitoring for action—rapid identification of pan-tropical deforestation using moderate-resolution remotely sensed data," Center for Global Development, Working Paper 192, 2009.

[9] R. Kennedy, Z. Yang, and W. Cohen, "Detecting trends in forest disturbance and recovery using yearly landsat time series," *Remote Sensing of Environment*, 2010.

[10] A. Koltunov, S. L. Ustin, G. P. Asner, and I. Fung, "Selective logging changes forest phenology in the Brazilian Amazon: Evidence from MODIS image time series analysis," *Remote Sensing of Environment*, vol. 113, no. 11, pp. 2431–2440, 2009.

[11] J. Kucera, P. Barbosa, and P. Strobl, "Cumulative sum charts-a novel technique for processing daily time series of modis data for burnt area mapping in portugal," in *MultiTemp 2007. IEEE International Workshop on the Analysis of Multi-temporal Remote Sensing Images, 2007.*

[12] R. S. Lunetta, J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy, "Land-cover change detection using multi-temporal MODIS NDVI data," *Remote Sensing of Environment*, vol. 105, no. 2, pp. 142–154, 2006.

[13] D. P. Roy, P. E. Lewis, and C. O. Justice, "Burned area mapping using multi-temporal moderate spatial resolution data–a bi-directional reflectance model-based expectation approach," *Remote Sensing of Environment*, vol. 83, no. 1-2, pp. 263–286, 2002.

[14] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: A survey and novel approach," *Data mining in time series databases*, vol. 57, 2004.

[15] K. Fukuda, H. Stanley, and L. Amaral, "Heuristic segmentation of a nonstationary time series," *Physical Review E*, vol. 69, no. 2, p. 021108, 2004.

[16] V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan, "Mining of concurrent text and time series," in *KDD-2000 Workshop on Text Mining*. Citeseer, 2000.

[17] S. Boriah, V. Kumar, M. Steinbach, C. Potter, and S. Klooster, "Land cover change detection: a case study," in *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2008, pp. 857–865.

[18] K. Vasko and H. Toivonen, "Estimating the number of segments in time series data using permutation tests," in *Data Mining, 2002. ICDM 2002. Proceedings. 2002 IEEE International Conference on*. IEEE, 2002, pp. 466–473.

[19] J. Hird and G. McDermid, "Noise reduction of ndvi time series: An empirical comparison of selected techniques," *Remote Sensing of Environment*, vol. 113, no. 1, pp. 248 – 258, 2009.

[20] V. Guralnik and J. Srivastava, "Event detection from time series data," in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '99. New York, NY, USA: ACM, 1999, pp. 33–42. [Online]. Available: http://doi.acm.org/10.1145/312129.312190

[21] B. Efron and R. Tibshirani, *An introduction to the bootstrap*. Chapman & Hall/CRC, 1993, vol. 57.

[22] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston, MA: Addison-Wesley, 2006.

[23] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series," in *ICDM 2001: Proceedings of the first IEEE International Conference on Data Mining*, 2001, pp. 289–296.