A robust anomaly based change detection method for time-series remote sensing images

Yin Shoujing¹, Wang Qiao¹, Wu Chuanqing¹, Chen Xiaoling², Ma Wandong¹, Mao Huiqin¹
¹Satellite Environment Center, Ministry of Environmental Protection, Beijing, China
²State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

E-mail: shoujingy@163.com

Abstract. Time-series remote sensing images record changes happening on the earth surface, which include not only abnormal changes like human activities and emergencies (e.g. fire, drought, insect pest etc.), but also changes caused by vegetation phenology and climate changes. Yet, challenges occur in analyzing global environment changes and even the internal forces. This paper proposes a robust Anomaly Based Change Detection method (ABCD) for time-series images analysis by detecting abnormal points in data sets, which do not need to follow a normal distribution. With ABCD we can detect when and where changes occur, which is the prerequisite condition of global change studies. ABCD was tested initially with 10-day SPOT VGT NDVI (Normalized Difference Vegetation Index) times series tracking land cover type changes, seasonality and noise, then validated to real data in a large area in Jiangxi, south of China. Initial results show that ABCD can precisely detect spatial and temporal changes from long time series images rapidly.

1. Introduction

A mountain of multi-scale time-series remote sensing images have been accumulated with the development of remote sensing technology, which play an important role in global change research as they have recorded detail information about land cover changes. However, there are not only abnormal changes caused by human activities and emergencies (e.g. fire, drought, insect pest etc.) recorded in the time-series remote sensing images, which are the focus of global change research, but also changes caused by vegetation phenology, climate changes, weather condition variation and sensor aging. Yet, challenges occur in analyzing the features of global environment changes and even the internal driving factors.

Therefore, few methods have been proposed to detect changes based on time-series remote sensing images at present. Most methods are for bi-temporal images analysis or based on them. These methods can be classified into two categories. One is based on image classification. Single-date images are classified separately and compared by pairs [1]. It’s easy to implement but time-consuming, and more priori knowledge is needed. The other is based on image transformation or mathematical combination of time series images. In some cases, a group of single-date images spanning a period are combined together and compared to another period using Change Vector Analysis (CVA) [2, 3]. In some other cases, images spanning the same month or season from different years are used as input, then image
transformation techniques such as Principal Component Analysis (PCA) [4], CVA [5], wavelets [6] (or Fourier transforms [7] are used to highlight change information. But it’s scene dependent and the result is difficult to interpret with threshold needed to detect changed area. Single-band quantitative parameters such as Normalized Difference Vegetation Index (NDVI), instead of multi-spectral images, are mostly used for establishing change trajectories [8].

These methods have their own limitations because they depend entirely on the specific image series analyzed. For example, images should follow specific distribution or the amount of images scenes should be enough to satisfy for enough statistical sampling. And it’s hard to interpret when and where changes occur for long time series analysis with these techniques.

As an important remote sensing parameter, NDVI can sensitively reflect the change of vegetation growth condition, biochemical property, and the ecosystem parameters. It can also reflect the comprehensive situation of the land cover type according to the pixel on the image [9, 10]. NDVI is mostly used for time-series analysis in global change research, especially in land use and land cover changes [11-15].

Land cover change is a relatively rare occurrence over large areas and during long period. Thus, the data acquired when land cover changes can be taken as anomaly or outlier, which is defined as an observation which appears to be inconsistent with the remainder of the data series [16] and deviates so much from other observations as to arouse suspicion that it is generated by a different mechanism [17]. A robust Anomaly Based Change Detection method (ABCD) is proposed here for time-series images analysis by detecting abnormal points in NDVI datasets, which needn’t follow normal distribution. With ABCD we can detect when and where changes occur, which is the prerequisite condition of global change study.

2. Methodology

Outlier detection methods, arising from statistics, have been suggested for numerous applications, such as credit card fraud detection, clinical trials, voting irregularity analysis, network intrusion, severe weather prediction, geographic information systems, and other data-mining tasks. Outlier detection methods can be divided into parametric (statistical) methods, such as Nair test, Grubbs-Test, Dixon-Test and T-Test, and nonparametric methods that are model-free, such as Box-Plot and Median Absolute Deviation (MAD) [18].

Sequence length of consecutive NDVI time series is often no more than 20 years, which is a small sample in statistics. And NDVI or its difference sequence hasn’t been proved to follow normal distribution in temporal domain. Thus, MAD is selected as it’s a nonparametric method for small sample statistic and there is no strict limitation for the number of outliers, which is necessary for land cover change detection.

MAD, firstly proposed by Hampel [19] and improved by Rousseeuw and Crous [20], is defined as follows:

\[
MAD = k \cdot \text{median}_i \left( | x_i - \text{median}_j(x_j) | \right)
\]

(1)

When k=1.4826, MAD is equal to standard deviation for data series of normal distribution.

Steps of MAD outlier detection are as follows for a variable \( x \) with length of \( n \):

**Step 1:**
Get \( \{d(x)\} \) array by formula:

\[
d(x) = \{|x_0 - m(x)|, \ldots, |x_i - m(x)|, \ldots, |x_n - m(x)|\}
\]

(2)

Here \( m(x) \) is the median of the array.

Then, \( M_d \), median of \( \{d(k)\} \), and \( MAD = 1.4826 \cdot M_d \), median absolute deviation of \( \{d(k)\} \) can be inferred.

**Step 2:**
Get statistical value \( L \) for each variable \( x \) which is to be detected by following formula,
\[ L = \frac{|x - m(x)|}{b_n \cdot \text{MAD}} = \frac{0.6745|x - m(x)|}{b_n \cdot M_d} \tag{3} \]

in which \( b_n \) is rectifying coefficient given by Rousseeuw & Croux.

**Step 3:**
Variable \( x \) with \( L > 3 \) can be considered to be outlier, which is similar to Layida Rule. Generally, outlier is detected by \( L > t_\alpha \) for given confidence \( \alpha \), where \( L \) follows t-distribution with \( t - 1 \) degree of freedom.

# 3. Experiments

## 3.1. Data preparation

Jiangxi province (Figure 1), the location of Poyang Lake Watershed, locates in southeast of China, where has undergone great changes with the construction of Poyang Lake ecological economic region and the construction of Three Gorges Dam. Products of SPOT Vegetation NDVI (VGT NDVI) s10 from January 1999 to December 2008 (Figure 2) are used for change detection, and Landsat TM images acquired in Sep. 29, 2002 and Oct. 18, 2003 are used for accuracy assessment.

![Figure 1. Location of Jiangxi Province.](image1)

![Figure 2. Mean NDVI of Jiangxi Province between 1999 and 2008.](image2)

## 3.2. Data Processing

Linear interpolation and Savitzky-Golay filtering are performed on the VGT NDVI s10 products to remove the noises of the datasets. Then the processed data is aggregated into monthly averaged NDVI time series. Normal distribution testing is performed on NDVI time series and NDVI differential time series and the results show that not all pixels follow normal distribution. Thus the Anomaly Based Change Detection method (ABCD) is proposed here and its technical steps are as follows:

1. Construct an array \( M \).
   
   \[ M = \{M(m, y) | m = 1, \ldots, 12, y = 1999, \ldots, 2008\} \tag{4} \]

   Here, \( m \) is month, and \( y \) is year. Then 12 sub-arrays are constructed by the same month of the 10 years month mean NDVI, For January (M01) as an example:

   \[ M01 = M(1) = \{M(1, y) | y = 1999, \ldots, 2008\} \]. \tag{5}
(II) Compute the difference of the monthly averaged sub-array to get NDVI change matrix $D$.

$$D = \{D(m, y) | M = 1, ..., 12, y = 1999, ..., 2008\}$$

(6)

$D$ indicates the NDVI variation between the same two months of neighbor years. For example, $D(1,2000) = \text{abs}(M(1,2000) - M(1,1999))$ indicates the absolute difference between January 1999 and January 2000.

(III) Perform MAD outlier detection on $D$ to get anomaly array $A(m, y)$. At the same time, perform threshold segmentation on $D$ to get the suspected change array $A_1(m, y)$. Abnormal values are assigned the value of 1 and 0 else for both arrays.

(IV) For each pixel, if there are more than three months with value 1 for both arrays in the same year, it’s regarded to be change point. It’s based on the hypothesis that land cover type changes no more than once in a year.

3.3. Results

Confidence level of 95% is selected for outlier detection and twice standard deviation is selected for threshold segmentation. Result of land cover changes from 1999 to 2008 in each year is shown in Figure 3. Changes in different years are presented with different colors, showing the spatial-temporal dynamic changes processes during the research period. As shown in the figure 3, regions undergoing changes are mainly in urban areas, forest areas, and wetland areas around the lake and along the rivers.

![Figure 3. Land cover changes from 1999 to 2008 using ABCD](image)

![Figure 4. Land cover changes between 2002 and 2003 in Nanchang. a) Using NDVI; b) using ETM+.](image)

Real land cover changes between 2002 and 2003 in Nanchang urban area are obtained using Landsat ETM+ images and used for accuracy assessment. Comparison of land cover change using NDVI and ETM+ is shown in Figure 4 and accuracy assessment results are shown in Table 1.

As is shown in the table 1, the overall accuracy for the change detection reaches 98.49% in the urban area of Nanchang, with the false detection rate of changed area of 16.34% and Kappa coefficient of 0.83. But for the Poyang Lake wetland area, false changes are detected as it’s influenced by the seasonal water cover.
Table 1 Accuracy assessment of change detection

|               | Unchanged | Changed | False detection rate % |
|---------------|-----------|---------|------------------------|
| Unchanged     | 9460      | 75      | 9535                   | 7.87                   |
| Changed       | 76        | 389     | 465                    | 16.34                  |
|               | 9536      | 464     | 10000                  |                        |
| Omission detection rate % | 7.97   | 16.16   |                        |                        |

Overall accuracy=0.9849  Kappa=0.8295

There are two factors possibly affecting the accuracy assessment: (a) cross-scale assessment using ETM+ data for VGT NDVI may introduce errors as mixed pixels exist when ETM+ image is aggregated into the same scale of VGT NDVI; (2) two single-date ETM+ images may not reflect the real changes across the year 2002 and 2003, which may be explored by NDVI time series. But the overall accuracy is incredible.

4. Conclusion
The availability of continuous NDVI time series from SPOT VGT, AVHRR and MODIS provides convenience for the interannual land cover change detection over large areas. The anomaly based land cover change detection method using NDVI time series proposed in this paper shows incredible results. And the advantages of ABCD method can be concluded as follows: (1) the capability of direct interpretation for annual change product, (2) robust results, (3) rapid computation. And the availability of monthly data shows the capacity of capturing the actual time of land cover changes. What’s more, the method can also be used to other data such as environment monitoring data, surface temperature, and suspended solids concentration in water body, etc.

The disadvantages of the ABCD method include: (1) False changes may be detected in sensitive areas undergoing irregular changes such as wetland areas, as in MAD method there are only 50% of the input data can be regarded as outlier points at most; (2) segment threshold partly influences the detection accuracy. Mixed pixels because of the low spatial resolution of SPOT VGT NDVI consist of both changed and unchanged objects, which may bring puzzles to the determination of changed/unchanged area. Further work is needed on accuracy assessment for ABCD method.

References
[1] Cardille J A and Foley J A. 2003 Remote Sens. Environ. 87 551-62.
[2] Lambin E F and Strahler A H 1994 Remote Sens. Environ. 48 231-44.
[3] Lambin E F and Strahler A H 1994 Int. J. Remote Sens. 15 2099-119.
[4] Eastman J R and Fulk M 1993 Photogram. Eng. Remote Sens. 59 991-6
[5] Bayarjargal Y, Karnieli A, Bayasgalan M, Khudulmur S, Gandush C and Tucker C J 2006 Remote Sens. Environ. 105 9–22.
[6] Blackburn B Aand Ferwerda J G 2008 Remote Sens. Environ. 112 1614-32.
[7] Westra T and De Wulf R R 2007 Int. J. Remote Sens. 28 1595-610.
[8] Gong J Y, Sui H G, Ma G R and Zhou Q M 2008 Int. Arch. Photogram. Rem. Sens. Spatial Inform. Sci. (Beijing) Vol XXXVII Part B7 P 757.
[9] Liu D S, Kelly M and Gong P 2006 Remote Sens. Environ. 101 167-80.
[10] Oetter D R, Cohen W B and Bernmetche M 2000 Remote Sens. Environ. 72 139-55.
[11] Xiao X M, Boles S and Fmlidng S. 2006 Remote Sens. Environ. 100 95-113.
[12] Chen J, Jonsson P, Tamura M, Gu Z, Matsushita B and Eklundh L. 2004 Remote Sens. Environ. 91 332-44.
[13] Myneni R B, Keeling C D, Tucker C J, Asrar, G and Nemani R R 1997 Nature 386 6626 695–702.
[14] Townshend J R G 1992 IGBP (Stockholm, Sweden)
[15] Tucker C J and Sellers P J 1986 Int. J. Remote Sens. 7 1395-416.
[16] Johnson R1992 Applied Multivariate Statistical Analysis (Prentice Hall)
[17] Hawkins D 1980. Identification of outliers. Monographs on Applied Probability and Statistics.
[18] Maimon O. and Rockach L 2005 Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers (Kluwer Academic Publishers)
[19] Hampel F R, Ronchetti E M, Rousseeuw P J and Stahel W J 1986 Robust Statistics. The Approach Based on Influence Functions (New York: Wiley)
[20] Rousseeuw P J and Croux C 1993 J. Amer Stat Association 88 1279-81