

ESTIMATING RELIABILITY OF DISTURBANCES IN SATELLITE TIME SERIES DATA BASED ON STATISTICAL ANALYSIS

Z.-G. Zhou^a, P. Tang^{b,*}, M. Zhou^a

^a Academy of Opto-electronics (AOE), Chinese Academy of Sciences (CAS), Beijing 100094, China;

Key Laboratory of Quantitative Remote Sensing Information Technology, CAS – (zgzhou, zhoumei)@aoe.ac.cn

^b Institute of Remote Sensing and Digital Earth (Radi), Chinese Academy of Sciences (CAS), Beijing 100101, China –
tangping@radi.ac.cn

Commission III, WG III/3

KEY WORDS: Multi-temporal, Anomaly Detection, Change Detection, Disturbance Detection, Confidence Level, Time Series

ABSTRACT:

Normally, the status of land cover is inherently dynamic and changing continuously on temporal scale. However, disturbances or abnormal changes of land cover — caused by such as forest fire, flood, deforestation, and plant diseases — occur worldwide at unknown times and locations. Timely detection and characterization of these disturbances is of importance for land cover monitoring. Recently, many time-series-analysis methods have been developed for near real-time or online disturbance detection, using satellite image time series. However, the detection results were only labelled with “Change/ No change” by most of the present methods, while few methods focus on estimating reliability (or confidence level) of the detected disturbances in image time series. To this end, this paper propose a statistical analysis method for estimating reliability of disturbances in new available remote sensing image time series, through analysis of full temporal information laid in time series data. The method consists of three main steps. (1) Segmenting and modelling of historical time series data based on Breaks for Additive Seasonal and Trend (BFAST). (2) Forecasting and detecting disturbances in new time series data. (3) Estimating reliability of each detected disturbance using statistical analysis based on Confidence Interval (CI) and Confidence Levels (CL). The method was validated by estimating reliability of disturbance regions caused by a recent severe flooding occurred around the border of Russia and China. Results demonstrated that the method can estimate reliability of disturbances detected in satellite image with estimation error less than 5% and overall accuracy up to 90%.

1. INTRODUCTION

Normally, the status of land cover is inherently dynamic and changing continuously at the temporal scale. However, disturbances (anomalies or abrupt changes) of land cover — caused by nature, biogenic factors or anthropogenic activities — occur worldwide at unknown time and locations. Detecting and estimating disturbances of land cover are important for resource managers to monitor land cover dynamics over large areas, especially where access is difficult or hazardous (Pickell et al., 2014). Satellite remote sensing images provide consistent observation of land cover over large areas at high frequencies, allowing monitoring the status of land covers (Hutchinson et al., 2015). Therefore, detecting disturbances of land cover in satellite image time series is crucial for risk warning and assessment (Verbesselt et al., 2012).

Recently, several time series analysis methods have been developed for near-real-time or online detection of disturbances in satellite image time series. For example, some online change detection methods use Extended Kalman Filter (Kleynhans et al., 2011) and Gaussian Process (Chandola and Vatsavai, 2011). A few near-real-time disturbance detection methods are based on harmonic models (Verbesselt et al., 2012; Zhou et al., 2014; Zhu et al., 2012), and nonlinear least square (Anees and Aryal, 2014). Generally, these methods consist of two main steps, i.e., fitting historical time series with some model and then detecting disturbances in new data according to discrepancies between true observations and predictions of the model.

However, despite the utility of these disturbance detection methods, the detection results are usually only flagged with Disturbance/Not disturbance, and they could not tell the reliability (or confidence level, e.g., 90%) of individual detected disturbance. As disturbances of land cover at different areas may have different significances, the reliability of each of the detected disturbances would be a basis for assessing risk grades and distributions.

To this end, this study proposes a method for estimating reliability of disturbances detected in satellite image time series based on statistical analysis. The method was validated with a case study for estimating reliability of unexpected flooding areas detected in MODIS image time series.

2. METHOD

The method consists of three steps: time series modelling, time series forecasting and disturbance detection, and reliability estimation. The time series data is extracted from satellite image time series pixel by pixel.

2.1 Segment and Model a Time Series

A seasonal time series Y_t ($t = t_1, \dots, t_n$) can be decomposed into three components — trend, season and residuals — by a season-trend model. Many times a satellite time series data does not varies regularly and may have different patterns at different time intervals. Break points between pairs of these patterns or time intervals can be estimated using Bai and Perron (2003) (BP),

* Corresponding author

adapted in the Breaks For Additive Season and Trend (BFASST) approach based on iterative regression with piecewise season-trend models (Verbesselt et al., 2010).

The estimated break points, supposed with number b , separate Y_t into $P = b + 1$ segments. For each segment, at time interval from $t^{(p-1)}$ to $t^{(p)}$, $p = 1, \dots, P$, the time series data can be regressed with a season-trend model:

$$Y_t = \alpha_p + \beta_p t + \sum_{k=1}^K a_{p,k} \sin\left(\frac{2\pi kt}{f} + \delta_{p,k}\right) + \varepsilon_t, \quad t^{(p-1)} \leq t < t^{(p)} \quad (1)$$

where α and β are respectively the intercept and slope of linear trend component, a_k , δ_k and K are respectively the amplitude, phase and order of harmonic season component, f is the frequency of time series data (e.g., 23 observations in a year) (Verbesselt et al., 2010). The season-trend model can be estimated by:

$$\begin{aligned} y_t &= \mathbf{x}_t^T \boldsymbol{\beta} + \varepsilon_t \\ \mathbf{x}_t &= [1, t, \sin(2\pi t/T), \cos(2\pi t/T), \dots, \sin(2\pi Kt/T), \cos(2\pi Kt/T)]^T \\ \boldsymbol{\beta} &= [a_1, a_2, \gamma_1 \cos(\delta_1), \gamma_1 \sin(\delta_1), \dots, \gamma_K \cos(\delta_K), \gamma_K \sin(\delta_K)]^T \end{aligned} \quad (2)$$

where $\boldsymbol{\beta}$ is parameters vector which can be estimated using ordinary least squares (OLS).

Among the piecewise models, the last one (i.e., $p = P$) can represent the recent pattern of the time series data, of which the time interval $(t^{(P-1)} \leq t \leq t_n)$ is called stable history period (Verbesselt et al., 2012). Therefore, the recent pattern of the time series data can be represented by the season-trend component of the last fitted model (for an illustration please see the time series data from 2009 to 2013 in Figure 1, here $t_n = 2013$):

$$\hat{Y}_t = Y_t - \varepsilon_t = \mathbf{x}_t^T \boldsymbol{\beta}, \quad t^{(P-1)} \leq t \leq t_n \quad (3)$$

2.2 Detect Disturbances in New Time Series Data

Under normal conditions — if there is no land cover disturbance or anomaly — the pattern of satellite time series data would persist, i.e., the last season-trend model in the stable history period would hold true for new time series data (i.e., $t > t_n$). Otherwise, if a disturbance appears in the new data, the disturbance data will, to some extent, deviate from the expected

pattern, i.e., the last fitted season-trend model $\mathbf{x}_t^T \boldsymbol{\beta}$ (Equation 3). For illustrations, please see the time series data from 2013 to 2014 in Figure 1 and Figure 2.

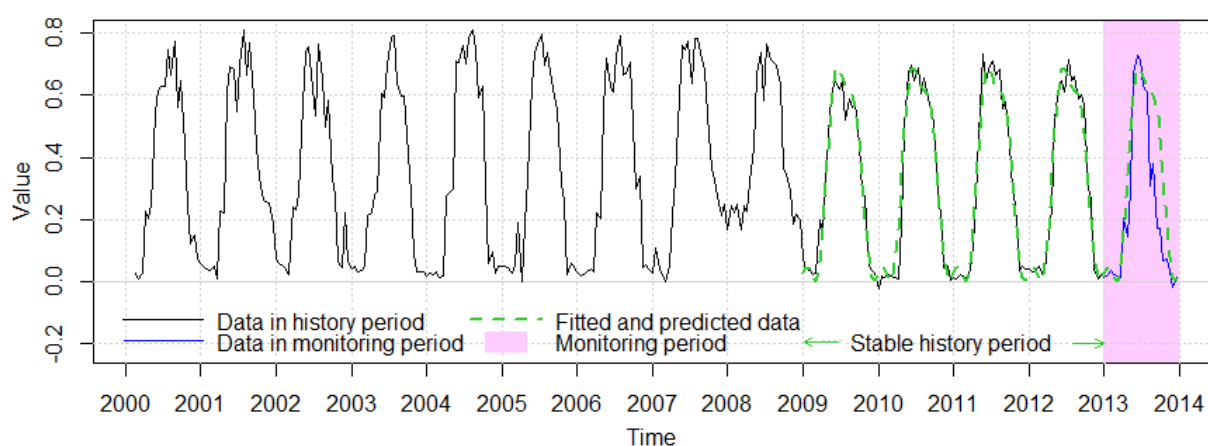


Figure 1. Illustration of segmenting and modelling a time series data before 2013 and forecasting data in 2014 with season-trend model.

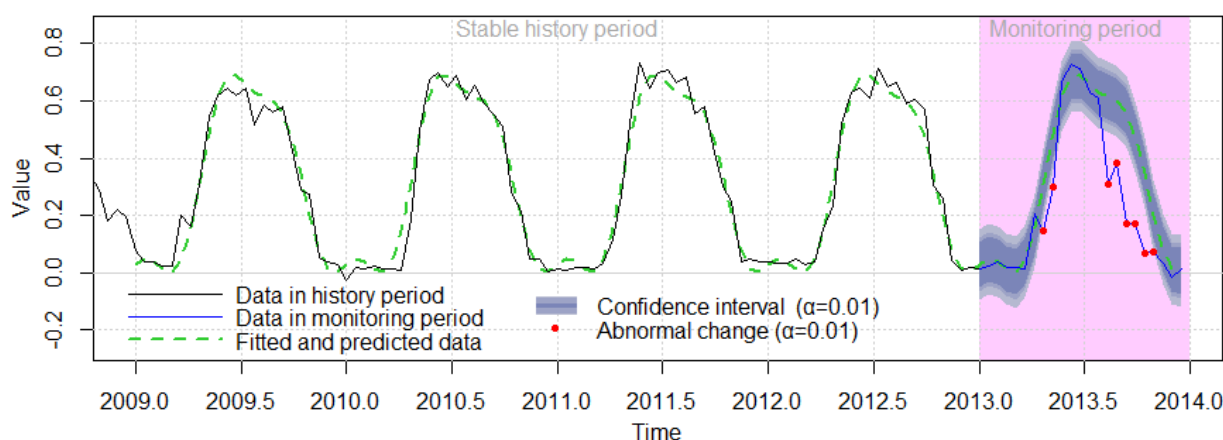


Figure 2. Illustration of detecting disturbances (anomalies or abnormal changes) according to the discrepancies between observations and predictions.

Therefore, disturbances in the new time series data can be determined based on discrepancies between true observations ($Y_t, t > t_n$) and the predictions of the last season-trend model. The predictions is generated by:

$$\hat{Y}_t = \mathbf{x}_t^T \boldsymbol{\beta}, t > t_n \quad (4)$$

A disturbance is detected at the Significance Level α (e.g., 0.01) if Y_t is beyond a Confidence Interval (CI) of the predictions \hat{Y}_t :

$$Y_t \notin CI_t: [\hat{Y}_t - z_{[\alpha/2]}\sigma, \hat{Y}_t + z_{[\alpha/2]}\sigma] \quad (5)$$

where σ is standard deviation of the regression residuals (ε_t) for the last season-trend model, $z_{[\alpha/2]}$ is cut-off z-value so that the area to its right under the standard normal curve is $\alpha/2$, given that ε_t follows the normal distribution $N(0, \sigma)$. The desired Significance Level α is not determined by the time series data but is set empirically. The smaller the Significance Level is (e.g., 0.01), the bigger the cut-off z-value is (e.g., 2.576) and less disturbances will be detected, and vice versa. Table 1 lists some commonly used Significance Levels and the corresponding cut-off z-values.

2.3 Estimate Reliability of Each Disturbance

The reliability of each detected disturbance can be estimated by measuring its Confidence Level (CL). The CL of a disturbance indicates how much reliability or confidence to put on this disturbance. It is defined as:

$$CL_t = 1 - \int_{|z_t|}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1 - P[z > |z_t|] \quad (6)$$

$$z_t = \frac{(Y_t - \hat{Y}_t) - u}{\sigma} \quad (7)$$

where u and σ are respectively the mean value and standard deviation of the regression residuals (ε_t) for the last season-trend model (see Equation 3), z_t is called observed z-value which is hypothesized to follow standard normal distribution, z is a continuous random variable representing possible results of z_t , and $P[z > |z_t|]$ is the probability of obtaining a result of z that is more extreme than what was actually observed of z_t .

The greater the absolute value of observed z-value ($|z_t|$), the more likely it is a small probability event, and we are more confident to take it as a disturbance. Some commonly used CL or reliability and the corresponding observed z-values are listed in Table 1.

Significance Level α	Cut-off z-value $z_{[\alpha/2]}$	Observed z-value $ z_t $	Reliability / CL (%)
0.2	1.282	0.842	80
0.1	1.645	1.282	90
0.05	1.960	1.645	95
0.01	2.576	2.326	99
0.001	3.291	3.090	99.9
0.0001	3.891	3.719	99.99

Table 1. Some common Significance Level and the corresponding cut-off z-value, observed z-value, and reliability.

3. DATA AND EXPERIMENT

The study area covers the Tongjiang section of Heilongjiang River (between southeast Russia and northeast China, covering about 6,000 km² with roughly 120 km by 50 km. There is no significant land cover change except that a big flood occurred in summer 2013. The riverbank was broke on August 23, causing severe flooding over large areas (Figure 3).

The experimental data are 16-day composited 250m spatial resolution Normalized Difference Vegetation Index (NDVI) images from satellite Terra/MODIS (23 images per year). NDVI indicates vegetation status with higher values for green vegetation (e.g., 0.2 – 0.8) and lower values for non-vegetation and particularly water bodies (e.g., -0.2 - 0.2). The NDVI image time series covering the Tongjiang section of the Heilongjiang River of China spanning from February 2000 to 2014 were collected from and subset on the LAADS Web (<http://ladsweb.nascom.nasa.gov>). It should be noted that cloud-removal or data-smooth was not conducted because after our inspection, we found that for transient flooding areas in MODIS NDVI images, the corresponding Vegetation Index Quality information in Quality Assurance (QA) dataset tends to be flagged with ‘Cloud’. The left images in Figure 4 show some of the MODIS NDVI images.

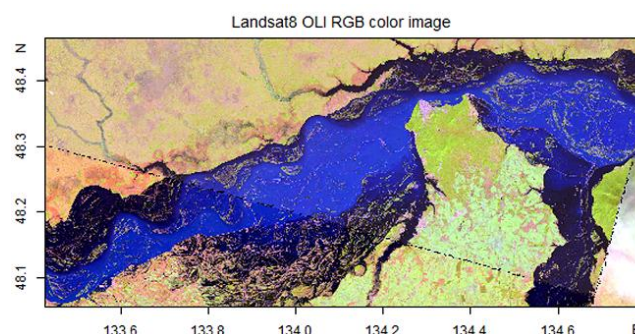


Figure 3. The reference 30-meter Landsat-8 OLI image showing the severe flood after riverbank break.

Some satellite images with higher spatial resolution (30-meter) were selected as reference images. The multi-spectral ETM+ and OLI images from Landsat-7/8 covering the study area from 2012 to 2013 were collected from the Earth Explorer website (<http://earthexplorer.usgs.gov>). The reference image for the severe flooding period in 2013 after mosaic is shown in Figure 3. The dates of the reference images during the severe flooding are 2013/09/18 and 2013/09/27, in the 17th compositing period of the MODIS NDVI images (see Figure 4(e)).

To validate the results of anomaly detection from MODIS NDVI image time series, a reference map of anomaly region from the reference images was produced. At first, for distinct classes in each image (i.e., water, forest, cropland, vegetative wetland, built-up area, cloud and cloud shadow, residual gaps), we visually selected more than 5% pixels in each class as sample data. Then, each of the four images is classified using a supervised classification method Support Vector Machine (SVM) (Pal and Mather, 2005), resulting in two water maps for the year 2012 and 2013. Finally, the reference anomaly region was derived from the two water maps by subtracting the water map in 2012 from the water map in 2013. The reference images before 2012 were not used here because there was not big flood for this period so that we use the image in 2012 to represent all the images before 2012.

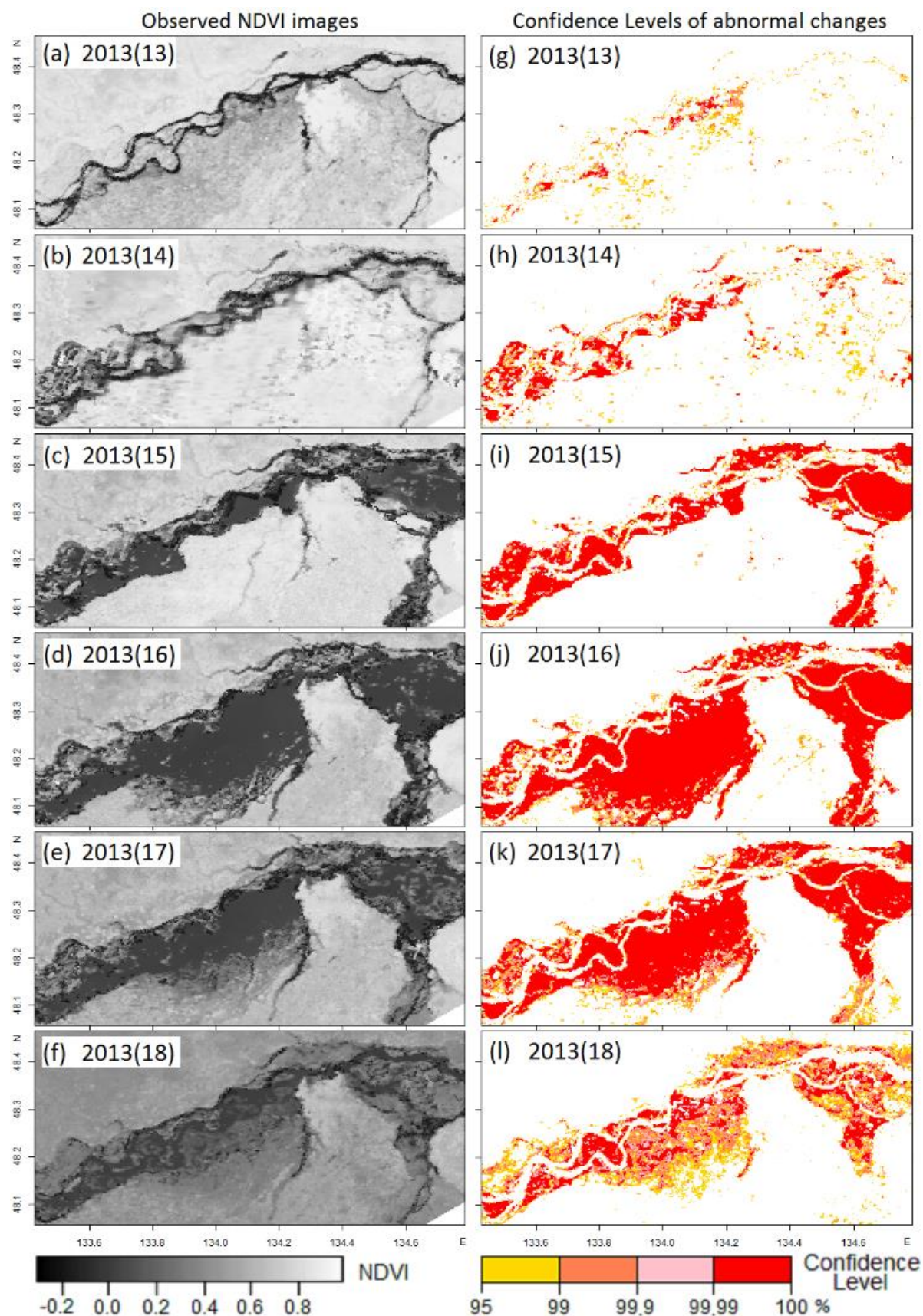


Figure 4. The MODIS NDVI image time series (a-f) during flooding period and the estimated reliability of the detected disturbance regions (unexpected flooding areas) (g-l). The number in parenthesis indicates the sequence number of the 23 16-day compositing period.

As is shown in Figure 4(a-f), in the study area, a big flood occurred during the summer of 2013 (from 2013(13) to 2013(18), 13 indicates the 13th image of 23 images per year). In this experiment, the 23 images in 2013 are taken as new image time series for disturbance detection and reliability estimation, while all the image time series before 2013 (not shown in this paper) are used for modelling. The disturbances to be detected are mainly unexpected flooding areas (in contrast to permanent and seasonal flooding areas).

4. RESULTS AND DISCUSSION

Disturbances in each of the 23 images were detected and the reliability (Confidence Level) of each disturbance was estimated. Figure 4 shows the MODIS NDVI images during the flooding period from 2013(13) to 2013(18) and the estimated reliability of the detected disturbances in each image. Comparing to the left images, the detected disturbance regions clearly show the dynamic spatial-temporal changes of the unexpected flooding areas, including the coming (a), expanding (b-d) and receding (e-f) of the flood. Note that the values of CL are actually continuous; they are sliced here just for contrast, red colour for highest level (>99.99%) and yellow for lowest (>95%). Due to difficulties to distinguish between flooding and humid vegetation areas especially as the flood recedes, some detected disturbances have low CL (e.g., yellow areas in Figure 4(g, l)), and we have low confidence to say these areas were already or still flooded at that moment. The estimated reliabilities can reflect the possibilities we are able to identify the flooding areas in the MODIS NDVI images.

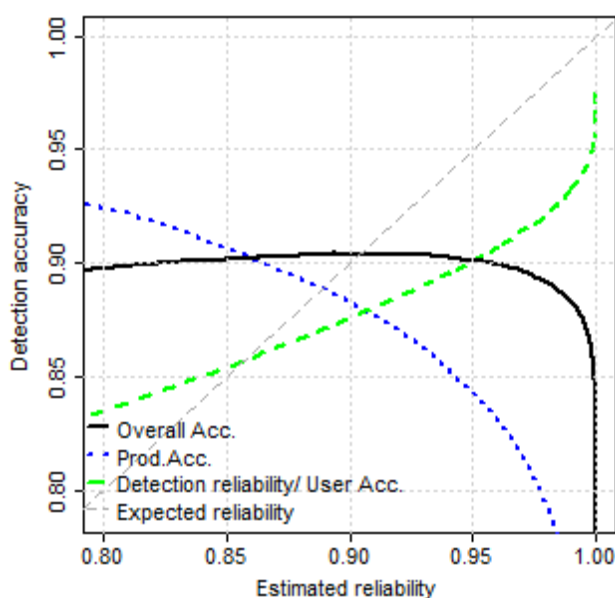


Figure 5. The detection accuracies and the detection reliabilities versus the estimated reliabilities.

The results demonstrate the effectiveness of the method, i.e., estimating reliability of disturbances detected in satellite image time series. The robustness of the method is intrinsic because, for one thing, the recent pattern of time series data is automatically identified using BFAST to avoid influences of past disturbances. For another, the higher the variability (σ) of time series data, the broader the Confidence Interval (CI) will be and the Confidence Levels (CL) of disturbances will achieve low values (see Equation 6 and 7). In this way, false detections would be properly

reduced and reliabilities would be greatly improved. The results also reveal an advantage of the method that it can eliminate expected changed areas (e.g., seasonal river water areas) from disturbance regions (e.g., unexpected flooded areas) by analysing full temporal information of satellite time series.

However, since the method detects disturbances in each image based on the discrepancies between predications and observations in that individual image, the method is sensitive to single/discontinuous strong noises and in fact, some detected disturbance may be a strong noise caused by cloud-covering or aerosol for example. Therefore, cloud-removing or noise reduction should be conducted before using the method to detect disturbances and estimate their reliabilities.

5. CONCLUSION

This study proposed a statistical analysis method for estimating reliability of disturbances detected in satellite image time series. The method detects disturbances based on discrepancies between true observations and predictions of a regression model. At the meantime, the reliability of each detected disturbance is estimated based on the statistical distributions of the discrepancies. The method was validated by a case study to estimate reliability of disturbance regions caused by severe flooding using MODIS NDVI image time series.

By analysing full temporal information, the method can detect disturbances in satellite image time series with high accuracies and meanwhile estimate reliability of each disturbance with low error. The method is independent of land cover types and it can be applied to various kinds of seasonal satellite image time series. However, it should be noted that the method is sensitive to single/discontinuous strong noise, so that some noise reduction is recommended to do beforehand. The method is now being improved to resistant to single strong noise by taking advantage of adjacent or continuous temporal information of the time series data.

ACKNOWLEDGEMENTS

This work was funded by the Project Y50B10A15Y supported by the Innovation Program of Academy of Opto-Electronics (AOE), Chinese Academy of Science (CAS). We are grateful to the anonymous reviewer and the WG Chairs for their careful critiques, comments, and suggestions, which helped us to considerably improve the paper.

REFERENCES

- Anees, A., Aryal, J., 2014. Near-Real Time Detection of Beetle Infestation in Pine Forests Using MODIS Data. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 7(9), pp. 3713-3723.
- Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18(1), pp. 1-22.
- Chandola, V., Vatsavai, R.R., 2011. A Gaussian Process Based Online Change Detection Algorithm for Monitoring Periodic Time Series, *Proceedings of the 2011 SIAM International Conference on Data Mining*. SIAM, pp. 95-106.
- Hutchinson, J., Jacquin, A., Hutchinson, S., Verbesselt, J., 2015. Monitoring vegetation change and dynamics on US Army

training lands using satellite image time series analysis. *Journal of environmental management*, 150(0), pp. 355-366.

Kleynhans, W., Olivier, J.C., Wessels, K.J., Salmon, B.P., van den Bergh, F., Steenkamp, K., 2011. Detecting Land Cover Change Using an Extended Kalman Filter on MODIS NDVI Time-Series Data. *IEEE Geoscience and Remote Sensing Letters*, 8(3), pp. 507-511.

Pal, M., Mather, P., 2005. Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26(5), pp. 1007-1011.

Pickell, P.D., Hermosilla, T., Coops, N.C., Masek, J.G., Franks, S., Huang, C., 2014. Monitoring anthropogenic disturbance trends in an industrialized boreal forest with Landsat time series. *Remote Sensing Letters*, 5(9), pp. 783-792.

Verbesselt, J., Hyndman, R., Zeileis, A., Culvenor, D., 2010. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment*, 114(12), pp. 2970-2980.

Verbesselt, J., Zeileis, A., Herold, M., 2012. Near real-time disturbance detection using satellite image time series. *Remote Sensing of Environment*, 123(0), pp. 98-108.

Zhou, Z., Tang, P., Zhang, Z., 2014. A method for monitoring land-cover disturbance using satellite time series images, SPIE Asia Pacific Remote Sensing. International Society for Optics and Photonics, pp. 9260381-9260386.

Zhu, Z., Woodcock, C.E., Olofsson, P., 2012. Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment*, 122(0), pp. 75-91.