IMPROVING TIME SERIES ANOMALY DETECTION BASED ON EXPONENTIALLY WEIGHTED MOVING AVERAGE (EWMA) OF SEASON-TREND MODEL RESIDUALS

Zeng-Guang Zhou [1,2], Ping Tang [1,]*

[1] Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China
[2] Key Laboratory of Quantitative Remote Sensing Information Technology, Academy of Opto-electronics, Chinese Academy of Sciences, Beijing 100094, China

ABSTRACT

Continuous anomaly detection in satellite image time series is important for studying spatial-temporal processes of land cover changes. In another study, we proposed a method based on Z-scores of Season-Trend model Residuals (Z-STR), which can continuously detects anomaly regions in image time series. However, Z-STR only detects anomalies with large shifts but cannot detect anomalies with small or gradual shifts. To improve the effectiveness of continuous anomaly detection, in this study, we propose a new method based on Exponentially Weighted Moving Average (EWMA) of Season-Trend model Residuals (EWMA-STR). Experiment for detecting spatial-temporal anomaly regions caused by severe flooding validated the better performance of EWMA-STR than Z-STR for time series anomaly detection. By using information of all historical time series data, EWMA-STR can detects anomalies with small or gradual shifts as well as anomalies with large or abrupt shifts.

Index Terms— Temporal analysis, land cover, change detection, disturbance detection, time series analysis

1. INTRODUCTION

Anomalies in satellite image time series can reflect anomalous changes of land cover caused by natural or human activities such as flood, forest fire, deforestation, etc. Continuous anomaly detection in satellite image time series is important for studying spatial-temporal dynamic changes or processes of land cover [1].

In another study [1], we proposed a method based on Z-scores of Season-Trend model Residuals (Z-STR) for continuous anomaly detection. Although it can continuously detects anomaly areas in satellite image time series, it can only detects anomalies with large shifts or changes rather than small or gradual shifts, because Z-STR only uses individual data for anomaly detection. The Exponentially Weighted Moving Average (EWMA) control chart [2] is very effective in detecting small shifts in time series process. It uses information from all samples and detects much smaller process shifts than a normal control chart would.

Therefore, in order to detect more anomalies with small or gradual shifts, this study proposes a method based on EWMA of Season-Trend model Residuals (EWMA-STR). It was tested and compared to the Z-STR with an experiment for continuously detecting spatial-temporal anomaly regions caused by a severe flooding, using MODIS Normalized Difference Vegetation Index (NDVI) image time series. Results demonstrated the better performance of EWMA-STR than Z-STR for detecting both small and large anomalies.

2. METHOD

2.1. Residuals extraction from a season-trend model

As described in [1, 3], a seasonal time series y_t can be decomposed into components of trend, season, and residuals by a season-trend model, which can be written as a standard linear regression form:

$$\begin{aligned} y_t &= \mathbf{x}_t^T \boldsymbol{\beta} + r_t \\ \mathbf{x}_t &= [1, t, \sin(2\pi 1 t/\mathrm{T}), \cos(2\pi 1 t/\mathrm{T}), \ldots, \sin(2\pi \mathrm{K} t/\mathrm{T}), \cos(2\pi \mathrm{K} t/\mathrm{T})]^T \\ \boldsymbol{\beta} &= [a_1, a_2, \gamma_1 \cos(\delta_1), \gamma_1 \sin(\delta_1), \ldots, \gamma_\mathrm{K} \cos(\delta_\mathrm{K}), \gamma_\mathrm{K} \sin(\delta_\mathrm{K})]^T \end{aligned} \tag{1}$$

where y_t is the observations values of the time series, T is the cycle or period of the time series (e.g, T=23 if there are 23 observations for a year), $\boldsymbol{\beta}$ is a vector of unknown parameters in which γ_k, δ_k, and K are respectively the amplitude, phase, and highest order of the harmonic season component.

Then, the residuals, r_t, can be directly extracted from the standard linear regression season-trend model:

$$r_t = y_t - \mathbf{x}_t^T \boldsymbol{\beta} \tag{2}$$

See Fig.1(b) for an illustration of season-trend model residuals extracted from the NDVI time series in Fig.1(a).

2.2. Detect anomaly based on EWMA of model residuals

Anomalies in the seasonal time series can be enhanced by the season-trend model residuals (see the residuals in 2008 and 2013 in Fig.1(b)). For the season-trend model residuals r_t, a recursion form of the EWMA of Season-Trend model Residuals (EWMA-STR) is:

$$z_t = \lambda r_t + (1-\lambda) z_{t-1} \tag{3}$$

where z_t is EWMA statisitcs, λ is a weighting factor ($\lambda \in (0,1]$) which is chosen by the user to determine how older

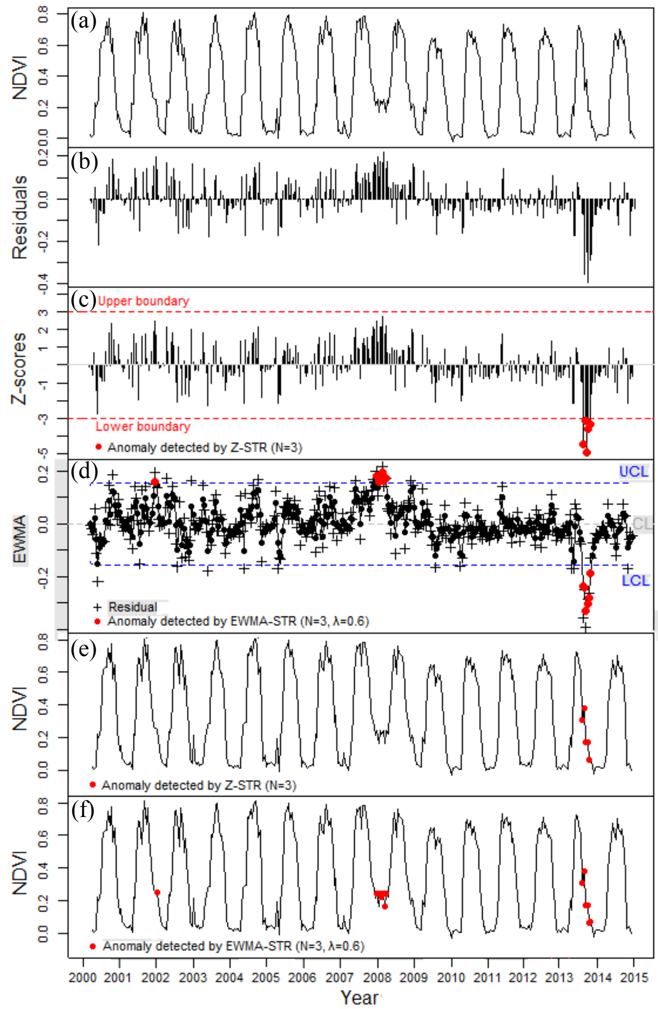

Fig. 1. Illustrations of anomaly detection in NDVI time series based on Z-sores of and EWMA of Season-Trend model Residuals (STR).

data points affect the mean value compared to more recent ones. A value of λ=1 implies that only the most recent measurement influences the EWMA (degrades to Shewhart chart on which the Z-STR is based). Thus, a large value of λ (closer to 1) gives more weight to recent data and less weight to older data; a small value of λ (closer to 0) gives more weight to older data.

The boundaries of EWMA chart, Upper Control Limit (UCL) and Lower Control Limit (LCL), are calculated as:

$$UCL = \mu + N \cdot \sigma\sqrt{\lambda/(2-\lambda)}$$
$$LCL = \mu - N \cdot \sigma\sqrt{\lambda/(2-\lambda)} \qquad (4)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are estimations of the mean value μ and standard deviation σ of the season-trend model residuals time series which is supposed following a normal distribution. As r_t may contain anomalies data, here we also use some robust estimations as in . $\hat{\mu}$ equals the average of r_t after eliminating both 5% highest and 5% lowest data. $\hat{\sigma}$ is calculated by $\sigma = \overline{|r_t|} \cdot \sqrt{\pi/2}$. N is a user defined factor which is usually set equal 3 or selected from the standard normal distribution.

For a time series process (i.e., r_t) being monitored by the EWMA chart, the process is out of control when the EWMA statistics (i.e., z_t) shifts beyond the Control Limits (i.e., UCL and LCL). Therefore, anomalies in the residuals time series data can be detected by a rule that:

$$z_t \ is \begin{cases} positive\ anomaly, & if\ z_t > UCL \\ negative\ anomaly, & if\ z_t < LCL \end{cases} \qquad (5)$$

See Fig.1(d) and Fig.1(f) for an illustration of EWMA-STR (N=3, λ=0.6) and the corresponding anomaly detection results. EWMA chart with a medium λ would sensitive to both small/gradual drift and big/abrupt change in the time series process. Therefore, EWMA-STR detects both small shifts in 2008 and abrupt changes in 2013. For comparison, Fig.1(c) shows the Z-STR (N=$z_{[\alpha/2]}$=3). Only the abrupt changes in 2013 are detected as shown in Fig.1(e).

3. EXPERIMENT AND RESULTS

3.1. Study area and data

The same as in [4], the study area is an area of 6,000 km^2 with roughly 120 km by 50 km around a part of the Tongjiang section of Heilongjiang River (see Fig.2(a)). The data is the 250-meter 16-day composited (23 images per year) MODIS NDVI images, spanning from Feb 2000 to Feb 2015.

The anomaly areas of interest in the NDVI images are also the unexpected inundated areas caused by a riverbank break (see Fig.2(d)) due to severe flooding in summer 2013.

For validation, some Landsat ETM+ and OLI 30-meter multi-spectral images covering the study area were collected as reference images. The reference anomaly areas derived from the reference images were produced by Supporting Vector Machine (SVM), as describe in [4]. The date of the reference anomaly areas corresponds to the 17th compositing period of the MODIS NDVI images (see Fig.2(e)).

3.2. Methods for experiment

The EWMA-STR and the ZSTR methods were used to test and compare the performance of continuous anomaly detection in satellite time series images. For EWMA-STR, the smoothing factor λ was set equal 0.8 (more sensitive to big or abrupt changes). For both methods, we set N=3 and only negative anomalies were considered, because only the anomalies caused by unexpected inundations, which lead to dramatic decrease of NDVI values, were of interest.

3.3. Results and discussion

Fig.2 lists the MODIS NDVI images during the flooding period in 2013 and the results of anomaly areas detected by EWMA-STR and Z-STR. Like the results of the spatial-temporal changes of unexpected inundated areas detected by Z-STR, the EWMA-STR results also show clearly the spatial-temporal continuous anomaly areas in the NDVI images.

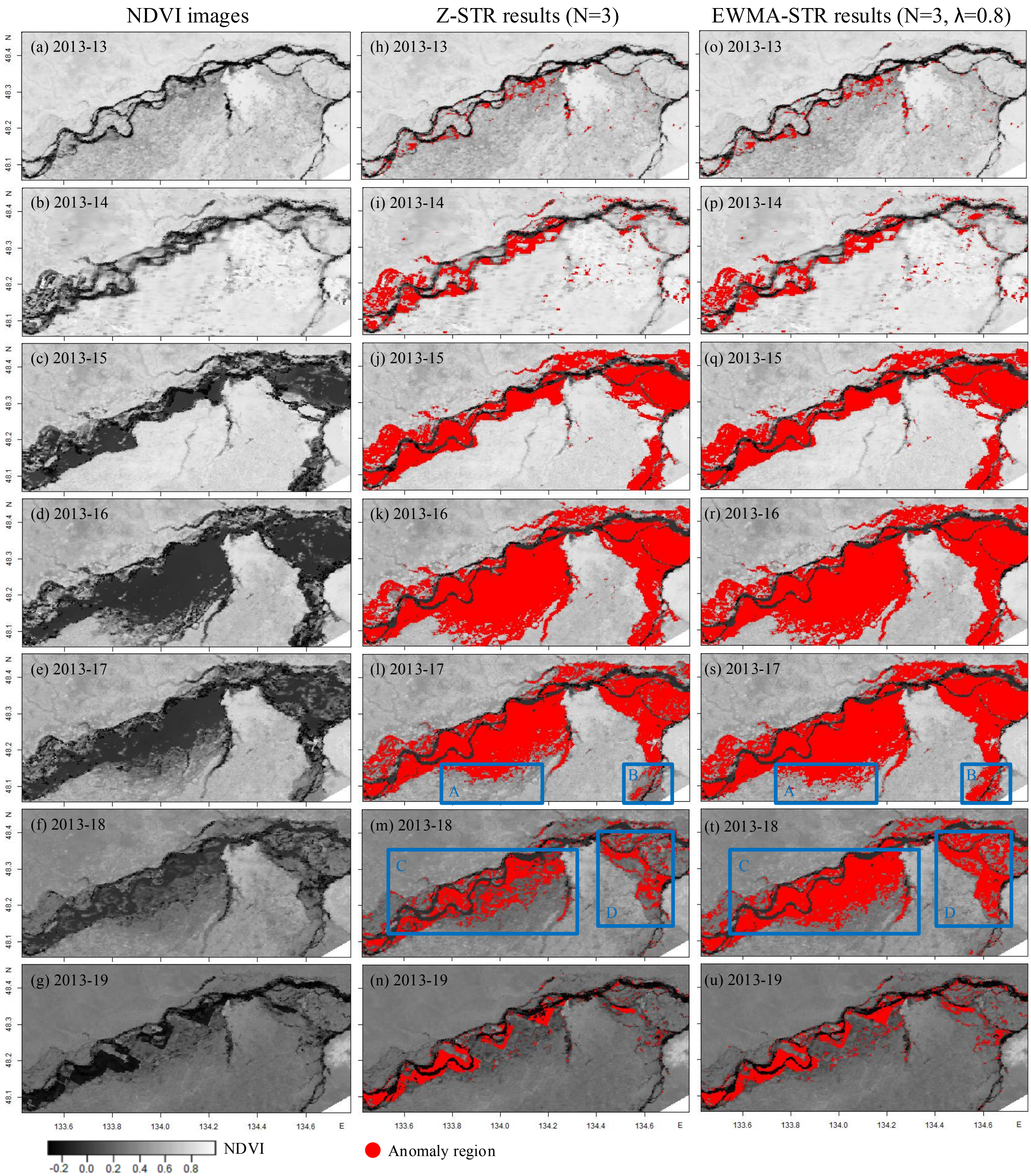

Fig. 2. (a-g) MODIS NDVI images during the flooding period, (h-n) anomaly regions (unexpected inundated areas) detected by Z-STR, and (o-u) anomaly regions detected by EWMA-STR. The numbers after the years indicate the sequence numbers of the 23 compositing periods.

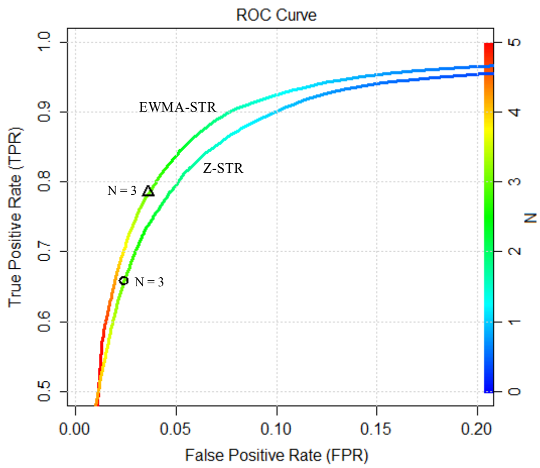

Fig. 3. ROC curves of the EWMA-STR (λ=0.8) and Z-STR methods on detection of anomaly areas in the 2013-17 NDVI image.

However, as shown in the blue rectangles in Fig.2(l), (m), (s), and (t), the EWMA-STR results have more anomaly areas than the Z-STR results. The probable reasons may be as follows. The time periods of the MODIS NDVI images in these figures (i.e., 2013-17 and 201-18) are just after the time period of the severe flooding (i.e., 2013-16) as shown in Fig.2(d), (k), and (r). The receding of flood after the big flooding leaded to gradual recovery of vegetated land cover and increase of NDVI values. Compared to the dramatic decrease of NDVI values in the NDVI image in 2013-16, the increases of NDVI values in the images in 2013-17 and 2013-18 caused relatively small shifts of NDVI time series from expected seasonal pattern. Therefore, the EWMA-STR could detect these small shifts but the Z-STR could not detect them.

To compare the performance of the two methods, we employed the receiver operating characteristic (ROC) curve (see Fig.3). The ROC curve is a plot of true positive rate (i.e., detection rate) versus false positive rate (i.e., false detection rate) when changing the discrimination factors (here it is N). Since the ROC curve more approaching to the upper left corner shows better detection performance, it can be seen from Fig.3 that the EWMA-STR method with λ=0.8 performances better than the Z-STR method for detecting anomaly areas in the NDVI image in 2013-17.

For quantitative comparison, Table 1 lists the confusion matrix and accuracies of the detected results. The Producer Accuracies for the two methods when N=3 are corresponding to the True Positive Rates as marked by triangle and circle in Fig.3. From Table 1 it can be seen that, for anomaly detection, the Producer Accuracy of EWMA-STR (78.48%) is almost 13 percent higher than that of Z-STR (65.81%) but the User Accuracies are almost the same (93.41% vs 94.73%). It means that, with a similar detection precision, the EWMA-STR has a much higher detection rate (i.e., it detects much more anomalies) than the Z-STR. It also can be seen that the Overall Accuracy (Kappa coefficient) is improved from 85.03% to 89.30% (from 0.67 to 0.77).

Table 1. Confusion matrix and accuracies of the anomaly areas in the 2013-17 NDVI image detected by the two methods (N=3).

Method	Detection	Reference (Pixels)		Accuracy (%)			Kappa Coef.
		Anomaly	Other	Producer	User	Overall	
EWMA-STR (λ=0.8)	Anomaly	34592	2441	**78.48**	93.41	**89.30**	**0.77**
	Other	9487	64927	96.38	87.25		
Z-STR	Anomaly	29010	1613	**65.81**	94.73	**85.03**	**0.67**
	Other	15069	65755	97.61	81.36		

The results demonstrated that the EWMA-STR method not only can continuously detect anomaly regions in satellite image time series, but also has better performance of anomaly detection than that of the Z-STR method. In particular, the EWMA-STR can detect more anomalies having small shifts, but Z-STR only detect anomalies with big changes.

4. CONCLUSION

This paper proposed a method, EWMA-STR, to improve the effectiveness of continuous anomaly detection in satellite image time series. EWMA-STR was validated and compared to the Z-STR method through an experiment to continuously detect anomalies caused by severe flooding. The experiment demonstrated the better performance of EWMA-STR than Z-STR for continuous anomaly detection in satellite image time series. Unlike Z-STR which is based on individual data, EWMA-STR use information from all historical data so that it can detects much smaller shifts of time series than Z-STR would. The effectiveness of the EWMA-STR having different smoothing factors will be analyzed in further study.

5. ACKNOWLEDGEMENTS

This work was funded by the project 181811KYSB20160001 supported by the Bureau of International Co-operation Chinese Academy of Sciences and the project Y3SG1500CX supported by the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences.

6. REFERENCES

[1] Z.-G. Zhou and P. Tang, "CONTINUOUS ANOMALY DETECTION IN SATELLITE IMAGE TIME SERIES BASED ON Z-SCORES OF SEASON-TREND MODEL RESIDUALS," in *IEEE International Geoscience and Remote Sensing Symposium*, Beijing, 2016.

[2] S. Roberts, "Control chart tests based on geometric moving averages," *Technometrics,* vol. 1, pp. 239-250, 1959.

[3] J. Verbesselt, A. Zeileis, and M. Herold, "Near real-time disturbance detection using satellite image time series," *Remote Sensing of Environment,* vol. 123, pp. 98-108, 2012.

[4] Z.-G. Zhou and P. Tang, "ESTIMATING RELIABILITY OF DISTURBANCES IN SATELLITE TIME SERIES DATA BASED ON STATISTICAL ANALYSIS," in *23th International Society for Photogrammetry and Remote Sensing*, Prague, 2016.