South African Journal of Geomatics, Vol. 6. No. 3, October 2017

A novel framework for parameter selection of the Autocorrelation Change detection method using 250m MODIS time-series data in the Gauteng province of South Africa

W. Kleynhans[13] , B.P. Salmon[4] , K.J. Wessels[23]

[1] Department of Electrical, Electronic and Computer Engineering, University of Pretoria
[2] Department of Geography, University of Pretoria.
[3] Meraka institute, CSIR, South Africa.
[4] School of Engineering, University of Tasmania

DOI: http://dx.doi.org/10.4314/sajg.v6i3.10

Abstract

Human settlement expansion is one of the most prominent types of land cover change in South Africa. These changes typically occur in areas that are covered by natural vegetation. Methods that can rapidly indicate areas having a high probability of change are very valuable to analysts as this can be used to direct their attention to high probability change areas for further evaluation. MODIS time-series data (8-daily composite) at a resolution of 500 m has been proven to be an effective data source for detecting human settlements in South Africa and it was proposed in Kleynhans et al., 2012 that a Temporal Autocorrelation Change detection method (TACD) be used to detect the formation of new settlements in the Gauteng province of South Africa. In this paper, the TACD that was proposed by Kleynhans et al., 2012 is adapted to be usable with variable sampled temporal resolutions for 250m MODIS data by using a novel framework for parameter selection. The proposed method is applied to variably sampled 250m MODIS time-series data ranging from daily to semi-annually and a comparison of change detection accuracy vs. false alarm rate is done in each instance. Key results indicate that there is little difference in performance between daily sampled and 2-monthly sampled 250m MODIS time-series data for the use case evaluated in this paper.

1. Introduction

The most pervasive form of land cover change in South Africa and many other developing countries around the world is human settlement expansion. In many cases, new human settlements as well as existing settlements expand informally and these expansions occur in areas that were previously covered by natural vegetation. Satellite time-series data has proven to be an effective data source for change detection (Verbesselt et al. ,2010;Lunetta et al., 2006;de Beurs and Henerby, 2005) and in particular, time-series analyses of hyper-temporal satellite data has been successfully applied to land cover change detection in South Africa (Kleynhans et al., 2012; Kleynhans et al., 2015, Salmon et al. 2013; Grobler et al., 2012, Grobler et al., 2013). In Kleynhans et al., 2012 a temporal Autocorrelation change detection (TACD) method was demonstrated to detect the development of new informal settlements in South Africa. The method uses the Autocorrelation Function (ACF) of a MODIS time-series to provide an indication of the level of time-series stationarity (by considering the stability of the time-series mean and variance over time) which is consequently used as a measure of land cover change. This TACD method demonstrated that by using an 8-daily composite MODIS time-series at a resolution of 500 m, new informal settlements could effectively be identified. The objective of this paper is to expand the TACD method developed in Kleynhans et al., 2012 to be usable with MODIS 250 m data having variable temporal resolution. To achieve this, a novel input parameter selection framework had to be developed to enable the method to work for variable temporal resolutions (i.e. daily, weekly etc.) as opposed to the fixed 8-daily temporal resolution used in the original formulation. As a case study, the exact same time period and study area was used as in the original TACD formulation presented in Kleynhans et al., 2012. The two main questions set out to be answered in the study were (1) how does the TACD method applied to daily 250 m data compare to the same area using the 8-daily 500 m data used in the original study and, (2) as the daily time series is available and can be sub-sampled to any temporal frequency, what is the subsequent performance of the method in terms of the change detection accuracy (percentage of change examples correctly detected) and false alarm rate (percentage of no-change examples incorrectly detected as change) as a function of the temporal frequency. For example, would there be a significant difference in the performance of the method when presented with a daily sampled time-series as opposed to a monthly sampled time-series. Although the false alarm rate of the method can be set to any percentage depending on the requirement of the operator, in our use case the false alarm rate requirement was low (< 1%) as the area on which the change algorithm is run is large and the validation of a large number of false alarms could be very costly and time consuming.

2. Data Description

2.1. MODIS data

The time-series data used in this study was derived from the MODerate-resolution Imaging Spectroradiometer (MODIS) instrument. Two time-series cubes were generated using MODIS data and formed the basis for all analytic results presented in this paper. Each cube was constructed as a single HDF5 file containing a surface reflection value for each band at a specific *(x,y)* location of a tile (in our case the h20v11 tile was used) at a particular date. Data for the first cube were obtained from the MCD43A4 product (sampled 8-daily at 500 m resolution) and is exactly the same as in the original study in Kleynhans et al., 2012. Data for the second cube were similar to that of the first cube with the key difference being (1) the spatial resolution was 250 m as opposed to 500 m, (2) the temporal resolution was daily as opposed to 8-daily and (3) only two bands were available as opposed to seven (due to fact that only the first two MODIS bands have a spatial resolution of 250 m). It was shown in Kleynhans et al. (2015), where a similar land cover change was considered, that multiple band combinations as well as vegetation indices including NDVI and EVI were not able to provide significant improvement over that of using only band 1. Band 1 is the red band in the visible spectrum range 620-670 nm and is known to be very sensitive to changes in vegetation. Since band 1 is a common band between the two dataset, the remainder of the analysis was focused on the analysis of band 1 time-series data. Both datasets are bi-directional reflectance distribution function (BRDF) corrected and the time-period for both datasets is 2001/01 to 2008/01. Quality control (QC) flags were not explicitly used in the pre-processing of the time-series data but it should be noted that the study area that was considered does not have prolonged time periods of cloud cover which results in a limited number of missing values (typically less than 4% of samples). In the rare occurrence of missing values, cubic spline interpolation was used to infer these missing values

2.2. Study Area

The study area considered in this study was the Gauteng province of South Africa (figure 1).

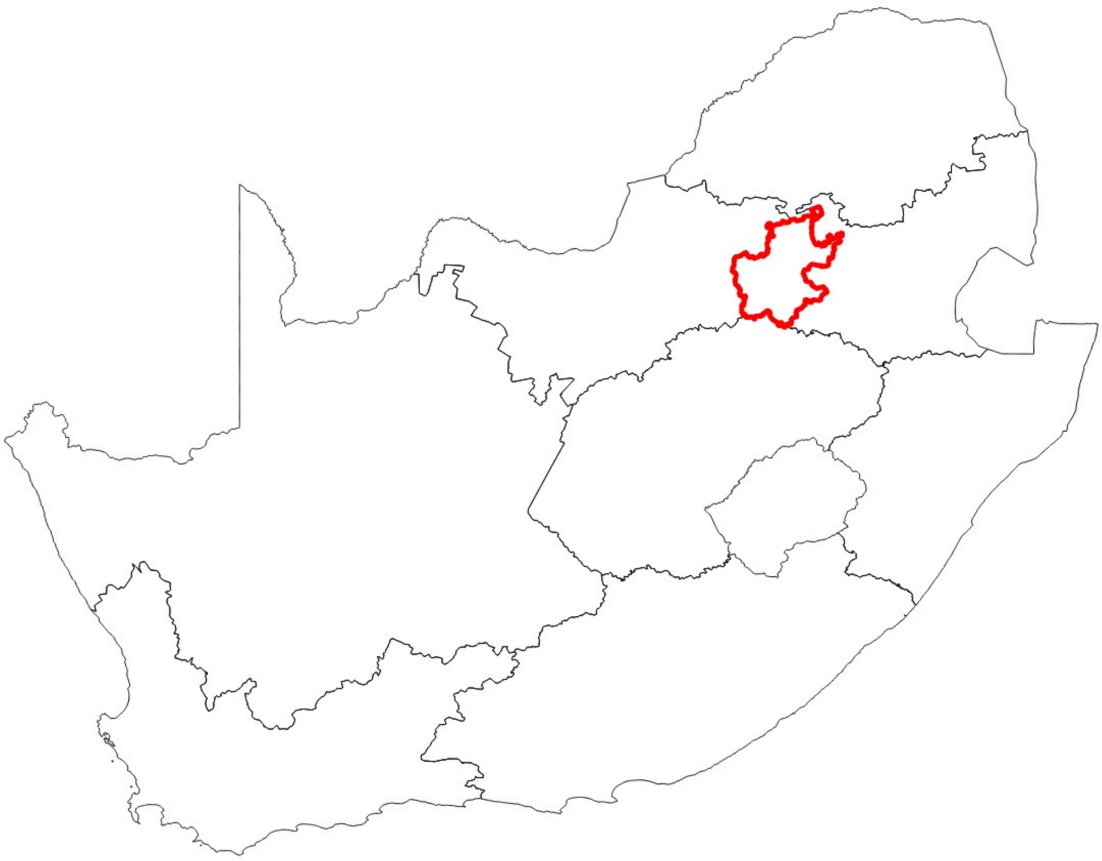

Figure 1: Gauteng province located in the central-northern part of South Africa

A dataset of no-change pixel time-series was identified at both 500 m and 250 m resolution using two steps. First, using visual interpretation, the 500 m pixel grid were overlaid on high resolution images in 2001 and 2008 respectively. The 2008 imagery over the study area were compared to that of 2001 and change areas were rapidly determined. A total of 180 pixels (with reference to the 500 m pixel grid) that changed from natural vegetation to informal settlement developments were identified. For a pixel to have been included as a change example, at least 70% of the pixel had to have changed. Second, as each of the 500 m pixels consisted of exactly four pixels at 250 m resolution, these four pixels were then included in the 250 m change dataset resulting in a total of 720 changed pixels being identified (with reference to the 250 m grid). The procedure to identify no-change pixel examples was similar to the procedure to identify change pixels. A total of 964 no-change pixel examples were identified for the 500 m resolution case with 3856 pixel examples identified for the 250 m resolution case. An example of the aforementioned methodology is shown in figure 2 where a change area is shown overlaid with a 500 m and 250 m grid respectively. The pixel indicated as (1) and (2) in figure 2 corresponds to a 500 m pixel footprint that changed from natural vegetation to settlement whereas (3) and (4) shows the 250 m pixel corresponding to the top left corner of the 500 m pixel that was confirmed to have changed.

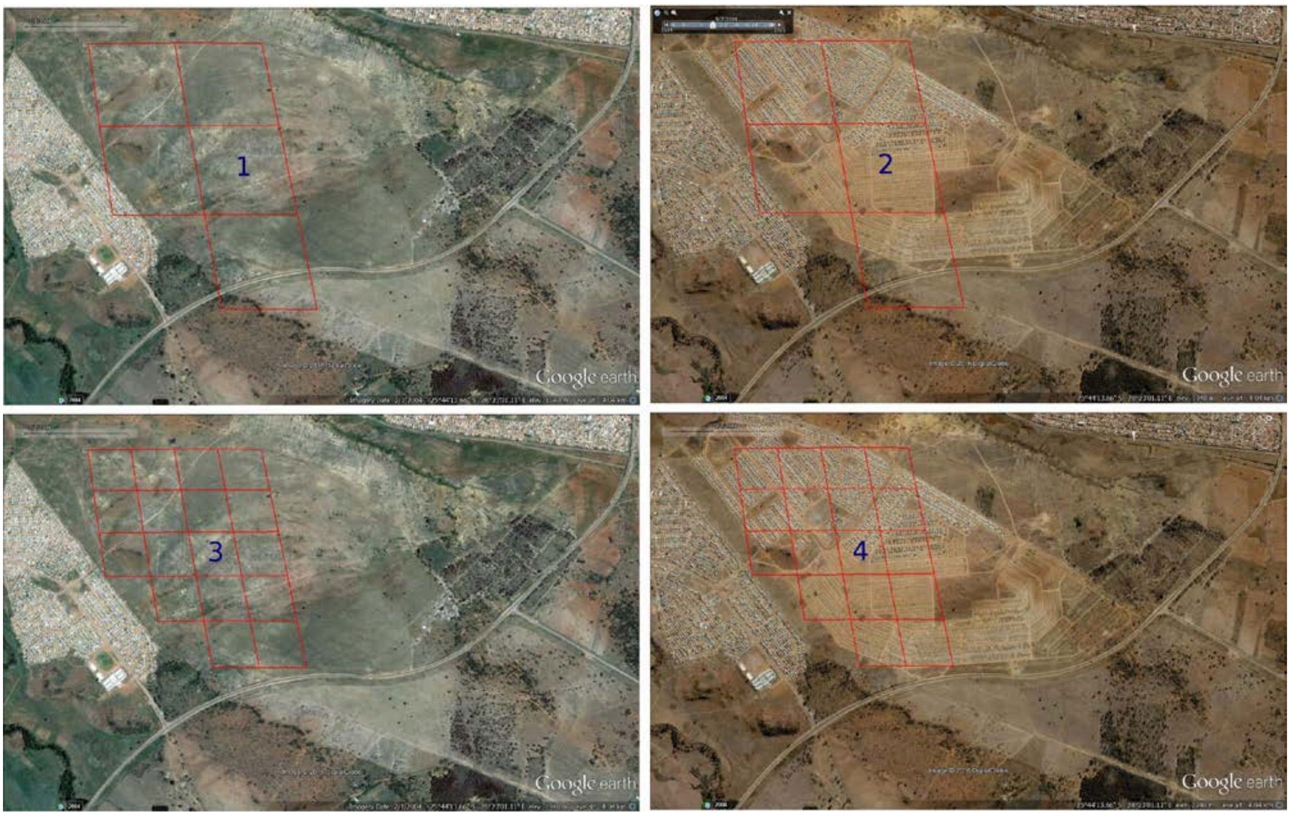

Figure 2: Quickbird imagery showing a change area (courtesy of Google Earth) overlaid with a 500 m and 250 m grid respectively. The pixel indicated as (1) and (2) corresponds to a 500 m pixel that changed from natural vegetation to settlement whereas (3) and (4) shows the 250 m pixel (corresponding to the top left corner of the 500 m pixel) that consequently was also confirmed to have changed.

Table 1: 250m and 500m dataset properties.

	250m	**500m**
Resolution	250m	500m
Bands	2	7
Temporal resolution	Daily	8-daily
# change pixels	720	180
# no-change pixels	3865	964
Use of dataset	Parameter selection and performance evaluation at multiple subsampled frequencies	Used for comparison with the daily sampled 250m dataset.

For simplicity the two datasets will be referred to as the 500 m and 250 m dataset for the remainder of the document. A summary of the datasets is provided in table 1.

From the following sections it will be shown that the primary function of the 250m dataset is to formulate the parameter selection framework (selection of the k and δ^* parameter) by making use of various subsampled temporal frequencies (i.e daily, monthly, quarterly etc.) and to also test the subsequent change detection and false alarm rate performance at each of these subsampled temporal frequencies

3. Temporal ACF Change detection method

The temporal ACF change detection (TACD) method proposed in Kleynhans et al.(2012) with a slight generalization provided in Kleynhans et al. (2015) was developed specifically using an 8-daily 500 m MODIS time-series based on the MCD43A4 product and can be defined as:

$$\delta = \sum_{\tau=1}^{k} \frac{E[(X_n - \mu)(X_{n+\tau} - \mu)}{\mathrm{var}(X)} \quad [1]$$

where X_n is the observation from an arbitrary spectral band at time n, τ is the ACF lag, E denotes the expectation and k is the total number of lags of the ACF to be summed. The mean of X is given as μ and the variance, which is used for normalization, is defined as $var(\boldsymbol{X})$. In essence equation (1) calculates the autocorrelation of the time series $\boldsymbol{X}$ and then sums the first k lags of the autocorrelation output. This summation (δ) is then used as a change metric and compared to a threshold value (δ^*) to determine whether a pixel is classified as change of no change:

$$change = \begin{cases} true & if \ \delta > \delta^* \\ false & if \ \delta < \delta^* \end{cases} \quad [2]$$

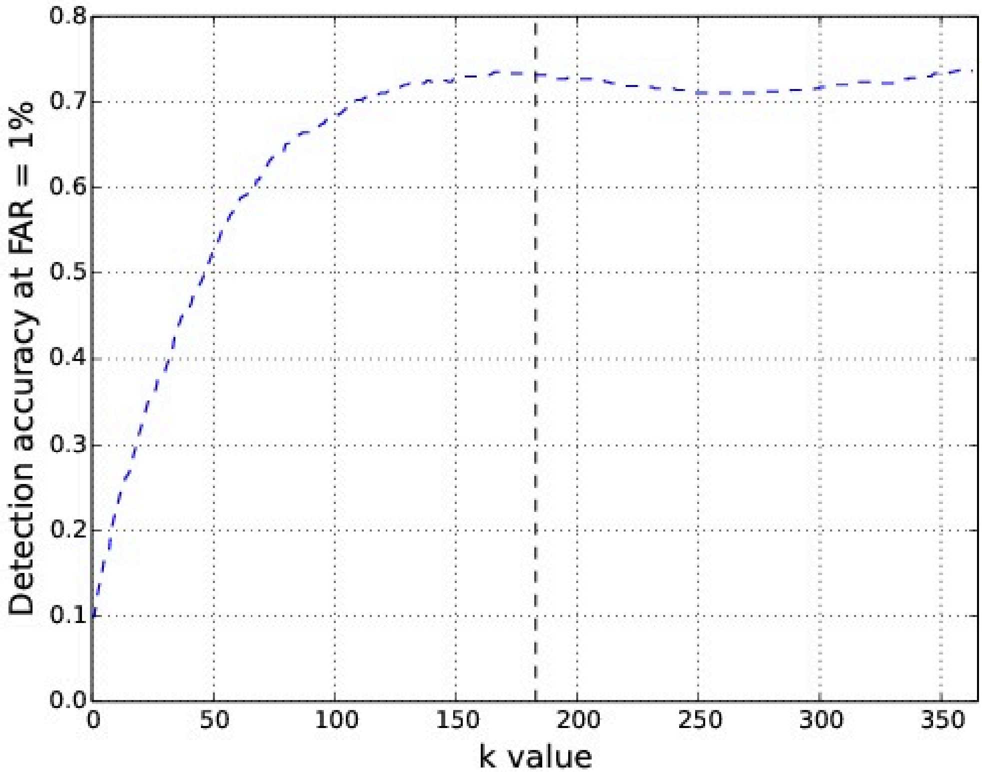

Figure 3: Change detection accuracy at a false alarm rate of 1% for each k value ranging from 1 to 365, vertical line corresponds to k = 183

3.1. Selection of the TACD algorithm parameters to be used on each of the datasets

With reference to equations (1) and (2) it is clear that the k and δ parameters should individually be selected when used with the 250 m and 500 m datasets. For simplicity, the TACD parameters to be used with the 250 m and 500 m datasets will be indicated as TACD250m and TACD500m respectively. It was shown in Kleynhans et al., 2012, where an 8-daily sampled time-series was used that a k value of 23 was optimal. This was determined by sweeping all possible k values and selecting the best performing value. In the current study it was found by empirical evaluation that the optimal k value can more formally be derived as half of the period of the time-series and can be expressed as

$$k = \left\lceil \frac{365}{2S} \right\rceil \quad [3]$$

where S is the temporal acquisition rate of the satellite in days. In the case of 8-daily data k can be calculated from (3) to be k = 23 whereas in the case of daily temporal frequency k = 183. To test the validity of the k value for daily data, the change detection accuracy was evaluated at a false alarm rate of 1% for each k value ranging from 1 to 365, effectively covering the entire period of the annual cycle. The results are shown in figure (3) where it can be seen that the k value shows near

Table 2: Change Detection Accuracy (CDA) , False Alarm Rate (FAR), k and $\delta*$ values for both datasets

Method	**CDA**	**FAR**	k	$\delta*$	**Sampling frequency**
$TACD_{250}$	67.81%	1.1%	183	52.94	Daily
$TACD_{500}$	67.40%	1.1%	23	6.78	8-daily

optimal performance when considering the change detection accuracy at a 1% false alarm rate thus validating the selected value of *k*. The same procedure was also done to validate the selection of *k* using equation (3) for the case of monthly, 2-monthly, quarterly as well as semi-annually temporal acquisition rates and produced similar results. As for the selection of the $\delta*$ parameter, there is a requirement to produce a false alarm rate of ≤ 1% as discussed in section 1, this is due to the fact that the change algorithm is run over large areas and the validation of a large number of false alarms could be costly and time consuming. To determine the threshold value that would yield a false alarm rate in this range, a similar approach was used as in Kleynhans et al. (2015) where the Bayesian decision error was evaluated based on the distribution of the inferred change index (δ) using a training no-change dataset as follows:

$$P(C\,|\,\overline{C}) = \int_{\delta=\delta*}^{\delta*=\infty} p(\delta\,|\,\overline{C}) = 0.01 \qquad [4]$$

where $\delta*$ is the decision threshold value and $P(C|\bar{C})$ is the false alarm rate. It can be seen from (4) that only no-change examples are required to determine the threshold value $\delta*$. The training set that was used constituted 20% of the no-change examples in each of the datasets and was calculated directly using equation (4).

4. Results and discussion

Using the *k* and δ values determined in section 3.1, change was declared on a per pixel basis for each of the datasets. It was found that there was a good correlation between the false alarm rate of the training (1%) versus the unseen datasets (1.1%) which implies that the in-sample and out-of-sample errors based on the value of $\delta*$ were roughly similar indicating good generalization of the method for both datasets. At a false alarm rate of ±1% the change detection accuracy was shown to be 67.81% in the case of TACD250 m with the change detection accuracy of 67.40% at a similar false alarm rate for the TACD500 m case (table 2). From the aforementioned it is clear that the performance of the TACD was virtually identical for both datasets . The next question that arises is how much the temporal frequency affects the performance of the method as the 250 m data was sampled at an eight times higher rate. To answer this question, the daily TACD250 m dataset was sampled 2-daily, 8-daily, monthly, 2-monthly, quarterly and semi-annually and in each of these cases the change detection accuracy at 1% FAR was calculated (table 3). From the results shown in table 3 it is clear that the performance does not significantly change when varying the temporal

Table 3: Change Detection Accuracy (CDA) and False Alarm Rate (FAR) at variable temporal frequencies indicated as TACDi where i is the temporal sampling in days, i.e. TACD8 is every 8'th sample in the TACD250 m dataset.

Method	**CDA**	**FAR**	k	$\delta*$	**Sampling frequency**
$TACD_1$	67.40%	1%	183	53.58	Daily
$TACD_8$	67.54%	1%	23	6.33	8-daily
$TACD_{31}$	68.09%	1%	6	1.14	monthly
$TACD_{62}$	67.82%	1%	3	0.41	2-monthly
$TACD_{122}$	61.88%	1%	2	0.26	Quarterly
$TACD_{183}$	42.27%	1%	1	0.15	semi-annually

sampling frequency between daily and 2-monthly but only starts to significantly reduce when sampling at less than a 2-monthly rate. It might seem counter intuitive that by sampling every two months as opposed to daily that there is no significant performance variation when using the TACD method but this does however make sense when considering the total time-series length and the typical duration of the development of a new or expanding settlement. It was shown in Kleynhans et al. (2012) that the expected duration of the development can vary significantly but typically ranges between 6 months and 24 months. When considering the total length of the study period (7 years) it follows that the study period is significantly longer than the duration of the typical change event evaluated in this study. Even when sampling at a much reduced rate, the underlying time-series is still represented adequately by using more than 6 samples per year with an insignificant loss in performance for our specific use case. This finding has a significant impact on the dataset requirements as the number of images required to generate a daily sampled vs. a 2-monthly sampled time-series varies by a factor of approximately 62.

5. Conclusion

In this paper, an extension is formulated to the temporal autocorrelation change detection (TACD) method proposed in Kleynhans et al. (2012) by formally defining the selection process of one of the parameters as a function of the sampling rate of the time-series. This enables the method to be used with variable temporal resolutions as opposed to the fixed 8-daily temporal resolution that was introduced in the original formulation. The modified algorithm was applied to variably sampled MODIS 250 m time-series data ranging from daily to semi-annually and the change detection accuracy and false alarm rate were computed for each instance. As the algorithm is intended to be run over potentially large areas (regional scale), a primary objective was to ensure that a low false alarm rate should be maintained (≤ 1%) and consequently a threshold was chosen to maintain this false alarm requirement. It was shown that the daily temporal frequency of the 250 m dataset that was used did not play a significant role in improving the change detection accuracy and that even when sampling up to an interval of 2-monthly (i.e 6 yearly observations), there were no performance decrease. There was however a reduction in performance when sampling at a lower rate than 2-monthly. For comparison, the TACD method was applied to both 250m daily sampled and 500m 8-daily sampled MODIS data for the same area and it was found that that near identical change detection accuracy was obtained at a FAR of 1% using both datasets.

REFERENCES

de Beurs, K.M., and G.M. Henerby. 2005. “A statistical framework for the analysis of long image time series.” International Journal of Remote Sensing 26 (8): 1551–1573.

Grobler, T.L., E.R. Ackermann, J.C. Olivier, A.J. van Zyl, and W. Kleynhans. 2012. “Land-Cover Separability Analysis of MODIS Time-Series Data Using a Combined Simple Harmonic Oscillator and a Mean Reverting Stochastic Process.” Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 5 (3): 857 –866.

Grobler, T.L., E.R. Ackermann, A.J. van Zyl, J.C. Olivier, B.P Salmon and W. Kleynhans. 2013. “Using page's cumulative sum test on modis time series to detect land-cover changes.” IEEE Geoscience and remote sensing letters, 10 (2): 332-336

Kleynhans, W., B.P. Salmon, J.C. Olivier, F. van den Bergh, K.J. Wessels, T.L. Grobler, and K.C. Steenkamp. 2012. “Land Cover Change Detection Using Autocorrelation Analysis on MODIS Time-Series Data: Detection of New Human Settlements in the Gauteng Province of South Africa.” Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 5 (3): 777 –783.

Kleynhans, W., B.P. Salmon, K.J. Wessels, and J.C. Olivier. 2015. “Rapid detection of new and expanding human settlements in the Limpopo province of South Africa using a spatiotemporal change detection method.” International Journal of Applied Earth Observation and Geoinformation 40: 74–80.

Lunetta, Ross S., Joseph F. Knight, Jayantha Ediriwickrema, John G. Lyon, and L. Dorsey Worthy. 2006. “Land-cover change detection using multi-temporal MODIS NDVI data.” Remote Sensing of Environment 105 (2): 142–154.

Salmon, B.P., J.C. Olivier, K.J. Wessels, W. Kleynhans, F. van den Bergh, and K.C. Steenkamp. 2013. “Land cover change detection using the internal covariance matrix of the extended Kalman filter over multiple spectral bands.” Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 6 (3): 1079 –1085.

Verbesselt, J., R. Hyndman, G. Newnham, and D. Culvenor. 2010. “Detecting trend and seasonal changes in satellite image time series.” Remote Sensing of Environment 114 (1):106–115.