CLOUD BASED SPATIO-TEMPORAL ANALYSIS OF CHANGE IN SEQUENCES OF SENTINEL IMAGES

Allan A. Nielsen[a], Morton J. Canty[b], Henning Skriver[c] and Knut Conradsen[a]

[a]Technical University of Denmark, DTU Compute – Applied Mathematics and Computer Science
DK-2800 Kgs. Lyngby, Denmark
[b]Heinsberger Str. 18, D-52428 Jülich, Germany
[c]Technical University of Denmark, DTU Space – National Space Institute
DK-2800 Kgs. Lyngby, Denmark

1. INTRODUCTION

An important task in remote sensing Earth observation involves the detection of changes which may signal for example environmentally significant events. The Sentinel-1 synthetic aperture radar (SAR) and the Sentinel-2 as well as the Landsat optical/visible-infrared spaceborne platforms, with spatial resolutions of the order of 10-20-30 meters and revisit times of the order of days, provide an attractive source of data for change detection tasks. Specifically, the SAR imagery provide complete independence from solar illumination and cloud cover. A convenient source of such data is the Google Earth Engine which gives near real time data access and which has an application programming interface for the access and for processing the data. Here we make available open-source automatic change detection software and for optical data also automatic radiometric normalization software for both cloud and local processing.

The theory sections of this contribution are very similar (nearly identical) to sections in [1]. In this contribution, we exclude examples on radiometric normalization and include new developments in both stand-alone and cloud software implementation and we give new examples.

2. CHANGE DETECTION IN SAR DATA

In [2] a change detection procedure for multi-look polarimetric SAR data [3] is described involving a test statistic (and its factorization) for the equality of polarimetric covariance matrices following the complex Wishart distribution. The procedure is capable of determining, on a per-pixel basis, if and when a change at any prescribed significance level has occurred in a time series of SAR images. The procedure may also be applied to collections of pixels (segments, patches, fields). Single polarization (power data, dimensionality $p = 1$), dual polarization (for example vertically polarized transmission, vertical and horizontal reception, $p = 2$) and full or quad polarization (all four combinations of vertical and horizontal transmission/reception, $p = 3$) can be analyzed.

The term multi-look in SAR imagery refers to the number of independent pixels (termed the equivalent number of looks, ENL) of a surface area that have been averaged in order to reduce the effect of speckle, a noise-like consequence of the coherent nature of the signal transmitted from the sensor. The observed signals in the covariance representation, when multiplied by the equivalent number of looks, are complex Wishart distributed. This distribution is the multivariate complex analogue of the well-known chi squared distribution.

The complex Wishart distribution is completely determined by the parameters p (dimensionality), ENL, and $\mathbf{\Sigma}$ (the variance-covariance matrix). Given two observations of the same area at different times, one can set up a hypothesis test in order to decide whether or not a change has occurred between the two acquisitions. The null hypothesis, H_0, is that $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, i.e., the two observations were sampled from the same distribution and no change has occurred, and the alternative (change) hypothesis, H_1, is $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$. Since the distributions are known, a likelihood ratio test can be formulated which allows one to decide to a desired degree of significance whether or not to reject the null hypothesis. Acceptance or rejection is based on the test's p-value, which in turn may be derived from the (approximately known) distribution of the test statistic when $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ ("under H_0" in statistical parlance).

For analysis of the situation with data from two time points, $k = 2$ below, see [4, 5, 6, 7]. In [8] the authors describe bi-temporal region-based change detection for polarimetric SAR images by means of mixtures of Wishart distributions.

If we have data from more than two time points, $k > 2$, the procedure sketched can be generalized to test a hypothesis that all of the k pixels (or patches) are characterized by the same $\mathbf{\Sigma}$,

$$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_k (= \mathbf{\Sigma})$$

against the alternative (H_1) that at least one of the $\mathbf{\Sigma}_i$, $i = 1, \ldots, k$, is different, i.e., that at least one change has taken place.

For the logarithm of the omnibus likelihood ratio test statistic Q for testing H_0 against H_1 we have, see [2]

$$\ln Q \quad = \quad n\{pk \ln k + \sum_{i=1}^{k} \ln |\boldsymbol{X}_i| - k \ln |\boldsymbol{X}|\}.$$

Here n is ENL, the $\boldsymbol{X}_i = n\hat{\mathbf{\Sigma}}_i$ (i.e., ENL times the observed covariance matrix) follow the complex Wishart distribution, $\boldsymbol{X}_i \sim W_C(p, n, \mathbf{\Sigma}_i)$, and $\boldsymbol{X} = \sum_{i=1}^{k} \boldsymbol{X}_i \sim W_C(p, nk, \mathbf{\Sigma})$. Also, if the hypothesis is true, $\hat{\mathbf{\Sigma}} = \boldsymbol{X}/(kn)$. $Q \in [0, 1]$ with $Q = 1$ for equality.

The probability of finding a smaller value of $-2 \ln Q$ is approximated by ($z = -2 \ln q$, where q is the actually observed value of Q)

$$P\{-2 \ln Q \leq z\} \quad \simeq \quad P\{\chi^2((k-1)f) \leq z\},$$

i.e., the probability of change at some time point. $f = 9$ for quad pol, $f = 4$ for dual pol, $f = 2$ for dual pol diagonal only.

Furthermore this test can be factored into a sequence of tests involving hypotheses of the form $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ against $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$,

$\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}_3$ against $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 \neq \mathbf{\Sigma}_3$, and so forth. More specifically, to test whether the first $1 < j < k$ complex variance-covariance matrices $\mathbf{\Sigma}_i$ are equal, i.e., given that

$$\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_{j-1}$$

then the likelihood ratio test statistic R_j for testing the hypothesis

$$H_{0,j} : \mathbf{\Sigma}_j = \mathbf{\Sigma}_1 \text{ against } H_{1,j} : \mathbf{\Sigma}_j \neq \mathbf{\Sigma}_1$$

is given by, see [2]

$$\begin{aligned} \ln R_j &= n\{p(j\ln j - (j-1)\ln(j-1)) \\ &\quad + (j-1)\ln|\sum_{i=1}^{j-1}\boldsymbol{X}_i| + \ln|\boldsymbol{X}_j| - j\ln|\sum_{i=1}^{j}\boldsymbol{X}_i|\}. \end{aligned}$$

Finally, the R_j constitute a factorization of Q such that $Q = \prod_{j=2}^{k} R_j$ or

$$\ln Q = \sum_{j=2}^{k} \ln R_j.$$

The probability of finding a smaller value of $-2\ln R_j$ is approximated by ($z_j = -2\ln r_j$, where r_j is the actually observed value of R_j)

$$P\{-2\ln R_j \leq z_j\} \simeq P\{\chi^2(f) \leq z_j\},$$

i.e., the probability of change at time point j with no previous change.

The tests are statistically independent under the null hypothesis. In the event of rejection of the null hypothesis at some point in the test sequence, the procedure is restarted from that point, so that multiple changes within the time series can be identified. For details including better approximations to the distributions of Q and R_j under the null hypotheses, see [2], visualization of change, and (some of) the software developed, see [9].

Since the omnibus method can detect not only if changes occur but also, within the temporal resolution of an image sequence, when they occur, long time series of frequent acquisitions over relevant sites are of special interest. One convenient source of such data is the Google Earth Engine[1] (GEE) [10] which ingests Sentinel-1 data (C-band, multi-looked VV/VH or HH/HV) as soon as they are made available by the European Space Agency (ESA) and provides an easy-to-use application programming interface (API) for accessing and processing the data.

3. CHANGE DETECTION AND RADIOMETRIC NORMALIZATION IN OPTICAL DATA

With respect to optical/visible-infrared imagery, a data-driven, statistical approach to change detection is provided by the iteratively reweighted multivariate alteration detection (IR-MAD) algorithm [11, 5]. This method applies iterated canonical correlation analysis (CCA) to geometrically co-registered multispectral images from two time points before calculating band-wise differences. The CCA orders the image bands according to similarity (measured by correlation), rather than spectral wavelength. The differences between corresponding pairs of canonical variates are termed the MAD variates. Specifically, a MAD variate Z is

$$Z = \boldsymbol{a}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{Y}$$

where $\boldsymbol{X}$ represents the m-dimensional image at time point 1, $\boldsymbol{Y}$ represents the m-dimensional image at time point 2, and $\boldsymbol{a}$ and $\boldsymbol{b}$ are the eigenvectors from the CCA. Thus $\boldsymbol{a}^T\boldsymbol{X}$ is a canonical variate for time point 1 and $\boldsymbol{b}^T\boldsymbol{Y}$ is a canonical variate for time point 2. We have m uncorrelated canonical variates (CVs) with mean value zero and variance one from both time points, the correlation between corresponding pairs of CVs is ρ (termed the canonical correlation which is maximized in CCA), and we have m uncorrelated MAD variates with mean value zero and variance $2(1-\rho)$.

In each iteration the values of each image pixel j are weighted by one minus the current estimate of the change probability and the image statistics (mean and covariance matrices) are re-sampled. Since the MAD variates for the no-change observations are approximately Gaussian and uncorrelated, the sum of their squared values (after normalization to unit variance)

$$C^2 = \sum_{i=1}^{m} \frac{Z_i^2}{2(1-\rho_i)}$$

ideally follows a chi squared distribution with m degrees of freedom, $C^2 \sim \chi^2(m)$. The probability of finding a smaller value of C^2 is approximated by (c^2 is the actually observed value of C^2)

$$P\{C^2 \leq c^2\} \simeq P\{\chi^2(m) \leq c^2\}.$$

Small P-values favour rejection of the no-change hypothesis, so for each iteration, $1 - P\{\chi^2(m) \leq c^2\}$ is used to weight each pixel to gradually reduce the influence of the change observations on the MAD transformation. Iterations continue until the canonical correlations stop changing (or a maximum number of iterations is reached).

Furthermore, canonical correlation analysis is invariant to linear and affine transformations, a fact that can be used to perform automatic relative radiometric normalization of the two multispectral images [12, 1]. This is not pursued further here.

4. SOFTWARE

The authors have made available the necessary change detection software for interaction with the GEE on the open-source repository Github[2]. The client-side programs run in a local Docker container serving a simple Flask web application. Apart from the Docker engine[3] and a browser, no software installation is required whatsoever. After the user has been authenticated to the Earth Engine, he or she can carry out the following tasks: 1) run the IR-MAD algorithm on Sentinel-2 (or Landsat) bi-temporal imagery, 2) perform relative radiometric normalization in batch mode on an image sequence, 3) run the sequential omnibus algorithm on Sentinel-1 dual polarization image time series, 4) export imagery to his or her Earth Engine assets folder or to Google Drive for further processing or visualization.

JavaScript code[4] to run both the Wishart omnibus and the IR-MAD methods directly in the GEE code editor/playground is also available. The Wishart omnibus code also generates an MP4 movie showing where and when change occurred.

As a recent development, a Docker-based interface to the GEE for the Wishart omnibus algorithm is made available.[5] It talks to the GEE servers from a Jupyter notebook and is more flexible than the

[1] https://earthengine.google.com/ and https://developers.google.com/earth-engine/

[2] https://github.com/mortcanty/earthengine/

[3] https://docs.docker.com/

[4] http://fwenvi-idl.blogspot.de/

[5] http://fwenvi-idl.blogspot.com/2018/07/jupyter-notebook-interface-for.html

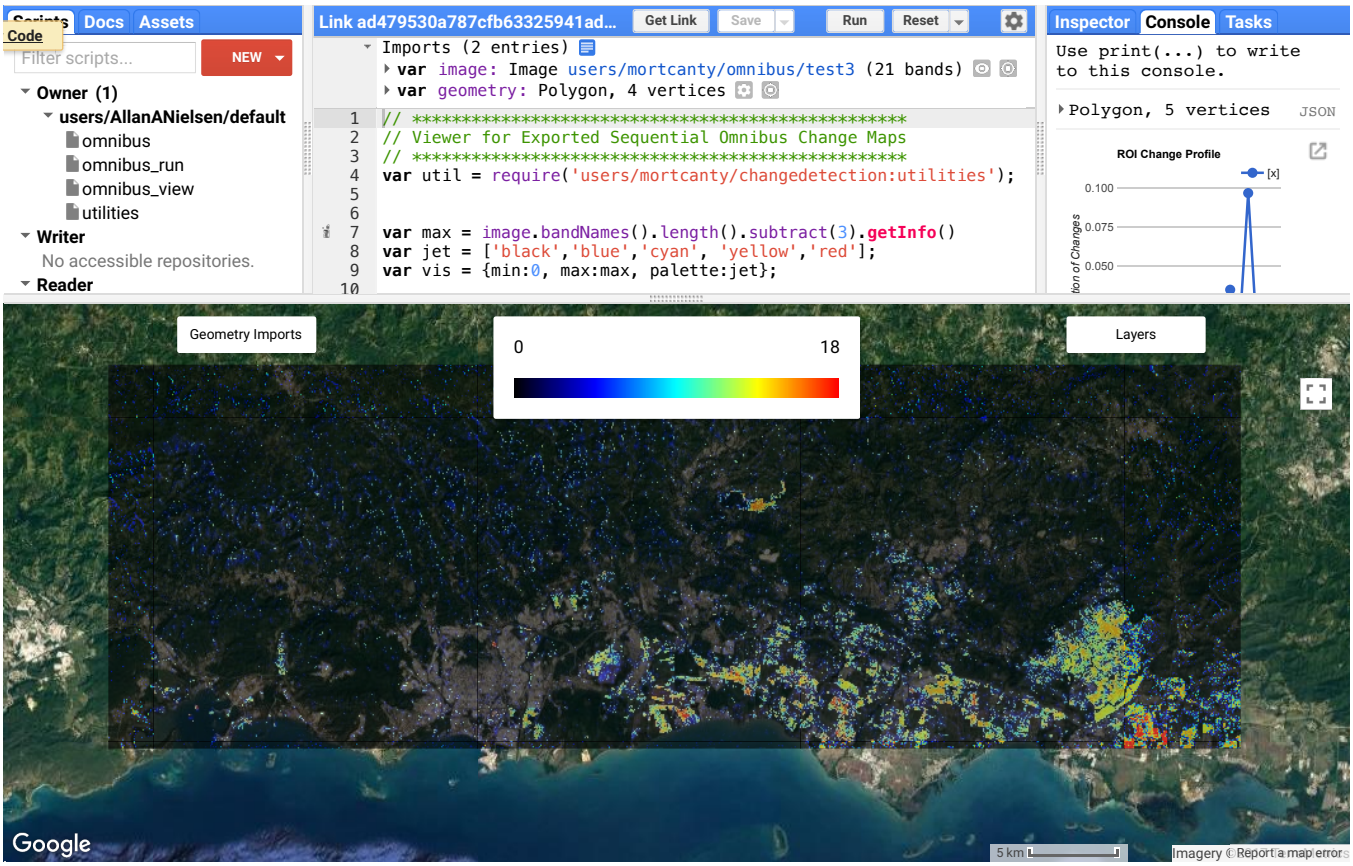

Fig. 1. Sequential omnibus change map for a region in southern Puerto Rico, showing the time of the most recent change (black none, blue early, red late). The time series consisted of 19 Sentinel-1 images from April to October 2017. Hurricane Maria struck on 20 September.

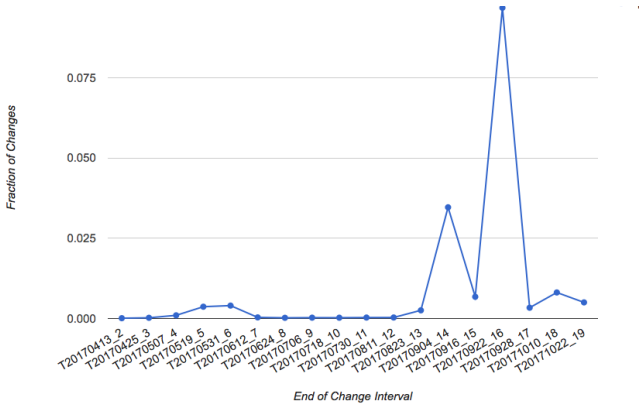

Fig. 2. Fraction of changed pixels in the south-eastern part of the change image shown in Figure 1. The peak occurs for the interval ending 22 September 2017, hurricane Maria struck on 20 September.

web interface, since the user is in a universal interactive Python programming environment.

Software is available also for local processing,[6] see [9]. Tutorials on how to install software and to do both the polarimetric SAR and the optical data processing locally on your own hardware are available on Github.[7,8] As another recent development, computer implementation work has been done within the Horizon 2020 project DataBio[9] DLV-732064 funded by the European Union (command-line and GUI executables[10] for Windows and Linux based on our Matlab code and on extended code from [13], a version for small images which fit into memory and a line-by-line version for big data exist), see proceedings from this meeting (first author Behnaz Pirzamanbein).

5. EXAMPLES

To illustrate, the Sentinel-1 multi-temporal VV/VH based change map in Figure 1 displays the color-coded time intervals in which the most recent changes in the 2017 hurricane Maria catastrophe in Puerto Rico occurred. Figure 2 shows the fraction of changed pixels which peaks in the interval ending on 22 September 2017. Maria made landfall in Puerto Rico on 20 September 2017. The change maps can be viewed interactively in the GEE Code Editor.[11]

Changes in one of several wildfires (the so-called Tubbs Fire[12] which took place on 9-30 October 2017 between Calistoga and Santa

[6] https://people.compute.dtu.dk/alan/software.html

[7] https://mortcanty.github.io/src/tutorialsar.html
[8] https://mortcanty.github.io/src/tutorial.html
[9] https://www.databio.eu/
[10] https://github.com/BehnazP/DataBio/
[11] https://code.earthengine.google.com/9374d69f4b0e3c11a7a14a9581f858d0
[12] https://en.wikipedia.org/wiki/Tubbs_Fire

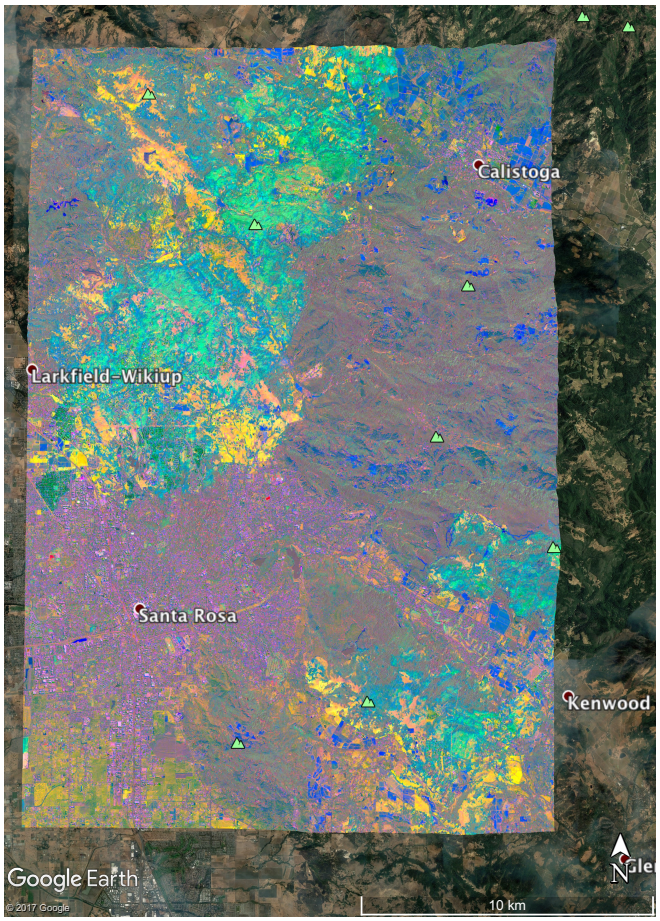

Fig. 3. The Tubbs Fire north of Santa Rosa, California, October 2017 (top-left; the bottom-right shows part of a larger fire around Kenwood). IR-MAD change variates associated with three greatest canonical correlations shown as RGB, burned areas in dark green (built-up areas), lighter green (mostly wooded) and bright yellow (mostly non-wooded), other non-fire related change mostly in blue (for example near Calistoga), and pale yellow (south of Santa Rosa). All variates are stretched over $\mp$ 16 no-change standard deviations.

Rosa and in which nearly 150 km^2 burned) in the northern California wine areas Napa Valley and Sonoma Valley are detected. The burned areas depicted in green (built-up and wooded areas) and bright yellow (non-wooded areas) in Figure 3 (where another fire down towards Kenwood is visible also) match well with published fire maps.[13,14] The Sentinel-2 images were acquired on 5 October and 1 November (bracketing the fire), only the four 10 m bands 2, 3, 4 and 8 were analyzed.

6. CONCLUSIONS

Examples based on both Sentinel-1 dual polarization synthetic aperture radar data and Sentinel-2 optical data show the usefulness of the generic, automatic change detection techniques sketched. Note, that for the optical change detection method, because of the orthogonality between the change variates, different types of change can be discriminated between.

The introduction of software for automated change analysis with polarimetric SAR as well as optical image data available to run either on your own hardware or to anyone authenticated to run on the Google Earth Engine is expected to be extremely useful to both researchers and practitioners. Generic, automatic techniques as these are expected to be useful in many other application areas also (other than natural disasters) where the study of spatio-temporal dynamics is important.

[13] http://abc7news.com/maps-a-look-at-each-north-bay-fire/2517694/
[14] http://fire.ca.gov/

7. REFERENCES

[1] M. J. Canty and A. A. Nielsen, "Spatio-temporal analysis of change with Sentinel imagery on the Google Earth Engine," in *ESA Conference on Big Data from Space (BiDS)*, pp. 126–129, Toulouse, France, 28-30 Nov 2017, https://doi.org/10.2760/383579.

[2] K. Conradsen, A. A. Nielsen, and H. Skriver, "Determining the points of change in time series of polarimetric SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 3007–3024, 2016, https://doi.org/10.1109/TGRS.2015.2510160.

[3] J. J. van Zyl and F. T. Ulaby, "Scattering matrix representation for simple targets," in *Radar Polarimetry for Geoscience Applications*, F. T. Ulaby and C. Elachi, Eds. Artech, Norwood, MA, 1990.

[4] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 1, pp. 4–19, 2003, https://doi.org/10.1109/TGRS.2002.808066.

[5] M. J. Canty, *Image Analysis, Classification, and Change Detection in Remote Sensing, With Algorithms for ENVI/IDL and Python*, Taylor and Francis, Third revised edition, 2014.

[6] A. A. Nielsen, K. Conradsen, and H. Skriver, "Change detection in full and dual polarization, single- and multi-frequency SAR data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 8, pp. 4041–4048, 2015, https://doi.org/10.1109/JSTARS.2015.2416434.

[7] V. Akbari, S. N. Anfinsen, A. P. Doulgeris, T. Eltoft, G. Moser, and S. B. Serpico, "Polarimetric SAR change detection with the complex Hotelling-Lawley trace statistic," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 3953–3966, 2016, https://doi.org/10.1109/TGRS.2016.2532320.

[8] W. Yang, X. Yang, T. Yan, H. Song, and G.-S. Xia, "Region-Based Change Detection for Polarimetric SAR Images Using Wishart Mixture Models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6746–6756, 2016, https://doi.org/10.1109/TGRS.2016.2590145.

[9] A. A. Nielsen, K. Conradsen, H. Skriver, and M. J. Canty, "Visualization of and software for omnibus test based change detected in a time series of polarimetric SAR data," *Canadian Journal of Remote Sensing*, vol. 43, no. 6, pp. 582–592, 2017, https://doi.org/10.1080/07038992.2017.1394182.

[10] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Tau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sensing of Environment*, vol. 202, pp. 18–27, 2017, https://doi.org/10.1016/j.rse.2017.06.031.

[11] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 463–478, 2007, https://doi.org/10.1109/TIP.2006.888195.

[12] M. J. Canty and A. A. Nielsen, "Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation," *Remote Sensing of Environment*, vol. 112, no. 3, pp. 1025–1036, 2008, https://doi.org/10.1016/j.rse.2007.07.013.

[13] N. Falco, P. R. Marpu, and J. A. Benediktsson, "A toolbox for unsupervised change detection analysis," *International Journal of Remote Sensing*, vol. 37, no. 7, pp. 1505–1526, 2016, https://doi.org/10.1080/01431161.2016.1154226.