Article

Statistical Analysis of Changes in Sentinel-1 Time Series on the Google Earth Engine

Morton J. Canty 1,*, Allan A. Nielsen 2, Knut Conradsen 2 and Henning Skriver 3

1 Formerly Jülich Research Center (Retired), 52428 Jülich, Germany
2 DTU Compute, The Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby 2800, Denmark; alan@dtu.dk (A.A.N.); knco@dtu.dk (K.C.)
3 DTU Space, The National Space Institute, Technical University of Denmark, Lyngby 2800, Denmark; hs@space.dtu.dk
* Correspondence: mort.canty@gmail.com

Received: 15 November 2019; Accepted: 16 December 2019; Published: 20 December 2019

Abstract: Time series analysis of Sentinel-1 SAR imagery made available by the Google Earth Engine (GEE) is described. Advantage is taken of a recent modification of a sequential complex Wishart-based algorithm which is applicable to the dual polarization intensity data archived on the GEE. Both the algorithm and a software interface to the GEE Python API for convenient data exploration and analysis are presented; the latter can be run from a platform independent Docker container and the source code is available on GitHub. Application examples are given involving the monitoring of anthropogenic activity (shipping, uranium mining, deforestation) and disaster assessment (flash floods). These highlight the advantages of the good temporal resolution resulting from cloud cover independence, short revisit times and near real time data availability.

Keywords: dual polarization SAR; change detection; flood monitoring; deforestation; port activity; uranium mining; Google Earth Engine; Python; Jupyter notebook

1. Introduction

The Sentinel-1a and -1b Synthetic Aperture Radar (SAR) satellite platforms, with spatial resolutions as low as 20 meters depending on acquisition mode, combined revisit times of the order of six days and complete independence from solar illumination and cloud cover, provide an attractive source of Earth observation data for change detection tasks. The Google Earth Engine (GEE) [1,2] includes, in its extensive and up-to-date archive, Sentinel-1 data in dual polarization (vertical transmission and vertical and horizontal reception over land surfaces) multi-look intensity format. The GEE makes available not only near real time data access but also a very powerful application programming interface (API) for processing and visualizing the data. The GEE API is presently written in JavaScript for direct interaction with the web-based GEE Code Editor and in Python for data analysis outside the GEE web environment. Relatively little work [3] has been published till now involving the use of the GEE’s Sentinel-1 archive for change detection. Since single look complex (SLC) data are not included in the archive for technical reasons, commonly used interferometric coherence methods are unavailable. Change detection with GEE Sentinel-1 imagery in forest land cover based on Otsu thresholding [4] and K-means clustering is investigated in [5]. The use of multitemporal SAR for automated monitoring of coastal structures with the COSMOS-SkyMed platform is described in [6], and in [7] long-term flood monitoring in Greece is investigated with time series involving a combination of COSMO-SkyMed, Sentinel-1 and Landsat 8 imagery.

In their original publication on polarimetric SAR data analysis, Conradsen et al. [8] introduced a change detection procedure for multi-look SAR data involving a test statistic for the equality of
polarimetric covariance matrices assumed to follow a complex Wishart distribution. The procedure is capable of determining, on a per-pixel basis, if a change at any prescribed significance level has occurred in two SAR images. Single polarization (power data, dimensionality \( p = 1 \)), dual polarization (for example vertically polarized transmission, vertical and horizontal reception, \( p = 2 \)) and full or quad polarimetric data (all four combinations of vertical and horizontal transmission/reception, \( p = 3 \)) can be analyzed with this method. More recently \([9,10]\) the procedure was extended significantly to deal with multi-temporal polarimetric SAR data in order to determine the points in time at which change occurs by applying omnibus tests and their factorizations. These latter publications describe test statistics and their distributions for time series of full \((2 \times 2)\) or \((3 \times 3)\) covariance matrices. For alternative approaches to polarimetric SAR change detection, see \([11,12]\).

The GEE Sentinel-1 intensity data can only be cast into \(2 \times 2\) covariance matrix format without the complex off diagonal elements, as these are not available in the archive. Since the diagonal form is not Wishart distributed, the originally developed algorithms are not immediately applicable. This restriction was removed in \([13]\) where the theory was extended to include covariance matrices in block diagonal form, the GEE data being a special case. In addition, an interpretation of direction of change based on the so-called Loewner order was recently added to the method \([14,15]\).

In this contribution we first describe briefly, in Section 2 below, the relevant aspects of the multi-temporal Wishart-based change detection algorithm, with emphasis on the dual polarization, diagonal only covariance matrices. In Section 3, details of the analysis software (available on GitHub) are given, including a platform independent graphical interface for convenient data exploration and generation of change maps. Then in Section 4 we present some examples intended to demonstrate the advantages of the algorithm’s application to long time series of SAR imagery extracted from the GEE archive. Some conclusions are drawn in Section 5 and a critical discussion is given in Section 6.

2. Theory

The observed fully polarimetric SAR signals, when expressed in covariance matrix form and multiplied by the (equivalent) number of looks, are complex Wishart distributed for fully developed speckle. This distribution is the multivariate complex analogue of the chi squared distribution and is completely determined by the dimensionality \( p \), the equivalent number of looks \( n \), and \( \Sigma \), the complex variance-covariance matrix. In a time series of \( k \) observations \( \Sigma_i \) of the covariance matrix, the likelihood ratio test statistic \( Q \) for rejecting the null hypothesis that all observations were sampled from the same distribution (no-change hypothesis) in favour of the alternative that at least one change occurred is given by \([9]\)

\[
\ln Q = n(p k \ln k + \sum_{i=1}^{k} \ln |X_i| - k \ln |\sum_{i=1}^{k} X_i|),
\]

where \( X_j = n \Sigma_i \). This is referred to as an omnibus test. The expression \(|\cdot|\) denotes determinant. The probability of finding a smaller value of \(-2p \ln Q\) than an observed value \( z \) can then be shown to be approximately

\[
\Pr(-2p \ln Q \leq z) \simeq \Pr(\chi^2(f) \leq z) + \omega_2(\Pr(\chi^2(f+4) \leq z) - \Pr(\chi^2(f) \leq z)),
\]

where \( \chi^2(f) \) is the chi square probability distribution with \( f \) degrees of freedom and where the parameters \( f \), \( p \) and \( \omega_2 \) are determined by \( p \), \( n \) and \( k \). For the GEE Sentinel-1 imagery, the diagonal-only covariance matrix is not Wishart distributed, however the test statistic \( Q \) is still given by Equation (1) but with special values for the parameters \( f \), \( p \) and \( \omega_2 \), see \([13]\) for the details.

The omnibus test statistic \( Q \) can be factored into a sequence of test statistics \( R_j \) for null hypotheses of the form \( H_{0,j} : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_{j-1} = \Sigma_j \) against the alternative \( H_{1,j} : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_{j-1} \neq \Sigma_j \), thus testing whether the first change occurred between the \((j-1)\)th and \( j \)th observations, \( j = 2 \ldots k \). The likelihood ratio test statistic \( R_j \) is given by
\[ \ln R_j = n(p(j \ln j - (j - 1) \ln(j - 1)) + (j - 1) \ln \left| \sum_{i=1}^{j-1} X_i \right| + \ln |X_j| - j \ln \left| \sum_{i=1}^j X_i \right|) \]  

(3)

and \( Q = \prod_{j=2}^k R_j \). Furthermore the statistics \( R_j \) are stochastically independent under the null hypothesis. Similarly to the omnibus test statistic, the approximate probability distribution of \(-2p_j \ln R_j\) is given by

\[ \Pr(-2p_j \ln R_j \leq z) \simeq \Pr\left(\chi^2(f) \leq z\right) + \omega_2\left(\Pr\left(\chi^2(f+4) \leq z\right) - \Pr\left(\chi^2(f) \leq z\right)\right). \]

The quantities \( f, \rho_j \) and \( \omega_2 \) are functions of \( p, n \) and \( j \) again with particular values for the GEE diagonal-only covariance matrices [13].

Rejection of the null hypothesis for a particular time interval signals a change, but says nothing about the nature of that change. In [14] a characterization of significant change is suggested based on the Loewner order [16], which calculates the definiteness of the difference between the covariance matrices at successive times. Changes are thus classified as positive definite, negative definite or indefinite.

3. Materials and Methods

Software for the sequential SAR change detection algorithm described in the preceding section is provided in Matlab (https://people.compute.dtu.dk/~alan/software.html) and in Python (https://mortcanty.github.io/CRC4Docker/), see also [17]. An additional Python implementation written specifically against the GEE Python API is available on GitHub (https://github.com/mortcanty/EESARDocker) and runs in a Docker container serving Jupyter notebooks to the user’s local browser. The pre-built Docker image can be pulled and run directly from Dockerhub and is platform independent. Since the algorithm runs on the GEE servers, the client-side requirements are light and a Raspberry Pi Docker image is also available. Installation instructions are given in the Github ReadMe file.

Let the index \( \ell \) denote subsequences of the images \( I_m \) beginning with the \( \ell \)th image in the time series and ending with the last, that is,

\[ \ell = 1 : \{I_m, I_{m+1}, \ldots I_k\} \]
\[ \ell = 2 : \{I_m, I_{m+2}, \ldots I_k\} \]
\[ \ell = k - 1 : \{I_{m-1}, I_k\}. \]

Denote with \( Q^\ell \) and \( R^\ell_j \) the test statistics associated with the \( \ell \)th subsequence, see Equations (1) and (3). The evaluation strategy for the GEE API is to pre-calculate, for each realization (observation) \( z^\ell_{ij} \) of \(-2p_j \ln R^\ell_j\), an array of associated \( p \)-values

\[ p_{ij} = 1 - \Pr(-2p_j \ln R^\ell_j \leq z^\ell_{ij}), \quad j = 2 \ldots k - \ell + 1, \quad \ell = 1 \ldots k - 1, \]

and, for realizations \( y^\ell \) of \(-2p \ln Q^\ell\), the \( p \)-value array

\[ p_\ell = 1 - \Pr(-2p \ln Q^\ell \leq y^\ell) \quad \ell = 1 \ldots k - 1. \]

The arrays are calculated via nested iterations over the subsequences.

In a second iteration step a change map is generated by testing the \( p \)-values against the desired significance level. The \( p_{ij} \)-values are only tested within a subsequence if the \( p \)-value \( p_\ell \) rejects the null-hypothesis for the entire subsequence (the omnibus test). This guarantees a fixed false positive probability for all registered changes. If, for a given pixel location, \( p_{ij} \) rejects the null hypothesis, the change map is updated correspondingly and the iteration continues on the subsequence \( I_{m_j} \ldots I_k \). This allows registration of successive changes at a single pixel location, at a spatial resolution of 20 m for the Sentinel-1 data. In a third and final iteration the average covariance matrix over the no-change
observations up to but not including the time of a change is accumulated with the provisional means method, see e.g., [17]. This average is subtracted from the covariance matrix immediately following the signaled change and the positive/negative definiteness (or indefiniteness) of the difference is ascertained to establish the Loewner order. (In [14,15] the opposite convention is adopted: the matrix following change is subtracted from the average prior to change). A particularly efficient method involving matrix pivots [15] is applied which avoids determining, in the case of full $(2 \times 2)$ and $(3 \times 3)$ covariance matrices, the eigenvalues of the complex matrix differences.

A Jupyter notebook (https://jupyter.org/) included with the software displays a simple and convenient graphical interface (Figure 1) allowing easy data exploration of the entire GEE Sentinel-1 archive together with the previewing of change results. Provision is also made for processing the user’s own image collections that have been uploaded to the GEE, e.g., Radarsat-2 quad pol or Sentinel-1 full dual pol data. Change maps and selected spatial subsets of the images can be exported to the user’s assets on the GEE Code Editor or to his or her Google Drive. Detailed instructions are included in a text cell in the notebook.

![Figure 1. Screenshot of the Jupyter notebook graphical interface showing the collection of a time series of Sentinel-1 images over the town of Jülich, Germany. See text for details.](image)

4. Results

One of the characteristics of the sequential omnibus algorithm when applied to long time series is its ability to pinpoint regions of high anthropogenic activity. Figure 2 is a color-coded change map of the total number of significant changes in a sequence of 74 Sentinel-1 images over the NATO Airbase near Geilenkirchen, Germany, acquired from January 2018 through June 2019. The aircraft movements to and from their parking positions on the tarmac are emphasized. Similar observations of frequent,
localized changes are the subject of the first two examples discussed in the following. The other examples involve flood monitoring and deforestation. For investigations of wetlands monitoring with the complex Wishart omnibus method both in full dual polarimetric form as well as with GEE imagery, see [3,18].

4.1. Libyan Maritime Port Activity

In part due to the refugee crisis in the Mediterranean and also because of political instability, ports along the coast of Libya have been the subject of recent attention, many of them having been closed periodically to international traffic. Arrivals and departures of large sea vessels are easily recorded with the algorithm, as can be seen for example from Figure 3 for the port of Tripolis. Positive definite significant differences (arrivals) are in red, negative definite (departures) in green. The number of detected events represents a lower bound on the vessel movements; a single vessel arrival/departure within a six-day interval between consecutive acquisitions will go undetected.

According to marine insurer North P&I Club, (http://www.nepia.com/insights/industry-news/status-of-libyan-ports-starupdatestar/) the Libyan Ministry of Transport closed the port of Tobruk from 27 October 2017 until 22 February 2018, although the nearby oil terminal at Marsa Elhariga remained open. By setting polygons around the two jetty berths at the oil terminal (Figure 4) and measuring the fraction of changed pixels within them, tanker activity was monitored from January 2017 through June 2019, see Figure 5. Two arrivals were recorded during the closure period, namely in the intervals ending 12 December 2017 and 29 January 2018.

One further example: The port of Benghazi was closed in the first part of 2017, reopening on 5 October 2017. This is confirmed in the chart of Figure 6.
Figure 3. Port of Tripolis: First four change maps of 68 in all covering the period 7 May 2018 through 7 June 2019 at six-day intervals. The background is a Sentinel-2 image acquired within the time interval. Changes are shown for a multi-polygon region of interest chosen to exclude noise from water surfaces. The color indicates the Loewner order: red: positive definite, green negative definite.

Figure 4. The oil terminal at Marsa Elhariga, Libya, with region of interest polygons on the tanker jetty docks. Map data ©Google Maps, DigitalGlobe.
4.2. Arms Control and Verification of Non-Proliferation

In the context of international treaties on nuclear non-proliferation, satellite sensors can be used to monitor peaceful nuclear fuel cycle activities, especially to detect undeclared activity at sites which, due to their remoteness, are not easily subject to regular on-site inspection. We first consider an example of a declared, operational uranium site, the McArthur River Uranium Mine in northern Saskatchewan, Canada, see Figure 7. A change profile for the core site can be seen in Figure 8. It is evident that changes are occurring throughout the observation time interval. This is contrasted with a similar profile for the decommissioned Cluff Lake mining site (Figure 7), also located in northern Saskatchewan. This mine ceased uranium production at the end of 2002 when the ore reserves were depleted. Apart from the two spikes in June and early August, much less activity is evident in its change profile. By observing the spatial distribution of change pixels, the spikes can be attributed to environmental changes, most probably in ground moisture following precipitation.
Figure 7. Polygon regions of interest on the Mcarthur River (left) and Cliff Lake (right) Uranium mines.

Figure 8. Fraction of changes at the Mcarthur River (top) and Cluff Lake (bottom) Uranium mines during the spring/summer months of 2017, see Figure 7.
4.3. Flood Monitoring

SAR imagery, because of its insensitivity to cloud cover, can often complement rapid response disaster management procedures, typically in situations involving sudden, large scale flooding. However many of the water detection methods proposed in the literature using reflected SAR radar signals require a considerable amount of user intervention. For example [19] gives an overview of the variety of radar signal/water surface interactions in flooded areas that must be accounted for. These depend on wavelength, polarization, incidence angle, wind/wave conditions, presence of partially flooded vegetation or man-made structures, etc. A recent example of the use of Sentinel-1 SAR imagery for flood monitoring and integration of results into rapid response concepts is given in [20].

In some cases, short-period change detection with polarimetric SAR may offer a simpler alternative to water detection/classification methods. Sudden wide-spread inundation will generally lead to a significant change in the polarimetric matrices characterizing the reflected signal. If the observation interval is short enough, most of the observed changes will be attributable to flooding, and afflicted areas quickly identified.

Cyclone Idai, recorded as the worst weather-based event to ever occur in the southern hemisphere, made landfall near the city of Beira, Mozambique on 15 March 2019 causing widespread inundation in Mozambique and Tanzania and a death toll of more than 1300. Figure 9 shows a sequence of six change maps generated with a GEE time series of Sentinel-1a and Sentinel-1b images. The reduction of reflectance from the water surfaces in both the VV and VH bands corresponds to a negative definite covariance matrix difference (green pixels, center left), the rapid receding of flood waters to a positive definite change (red pixels, center right).

A similar series of change maps is shown in Figure 10 for Golestan province, Iran, where from mid-March to April, 2019 widespread flash flooding affected large parts of the province. Here again one sees the initial flooding (green) followed by receding water (red), clearly delineating the areas most seriously affected.
Figure 9. Buzi district, Mozambique: Six change maps of 15 in all for 16 images covering the period 1 January 2019, through 7 June 2019. Each image has an area of approximately 3000 km$^2$. The gray scale background is the temporal average of the VV band of all 16 images. The maps, read top to bottom, left to right, are for the intervals 18 April–2 March, 2–14 March, 14–20 March, 20–26 March, 26 March–1 April and 1–7 April. Positive definite changes are red, negative definite green and indefinite yellow.
Figure 10. Golestan province, Iran, over the town of Aq Qala. Four change maps of 15 in all covering the period 4 January 2019, through 21 June 2019 are shown. From top to bottom, left to right: March 5 to 17, March 17 to 29, March 29 to April 10, April 10 to 22. Each image covers an area of approximately $25 \times 25$ km$^2$. The gray scale background is the temporal average of the VV band of all 16 images.

4.4. Clear Cut Logging

Another common application for satellite Earth observation is deforestation assessment/control, not only in the Amazon rain forest, but elsewhere. For a multi-sensor and multi-temporal approach to deforestation monitoring see for example [21].

Logging, and especially clear cutting, in Canadian boreal and temperate rain forests [22] continues to be controversial, with the last few percent of remaining old growth under threat. For example in the Nahmint Valley near Port Alberni, Vancouver Island, centuries-old red cedar and Douglas fir trees are still being cut down. Figure 11 shows two Sentinel-2 acquisitions from September, 2017 and June, 2018, where three new clear cuts are visible. A corresponding change map in Figure 12 obtained from a series of 33 Sentinel-1 images from September, 2017 through September, 2018 confirms that the clear cuts can be easily located, both geographically and temporally, by the sequential omnibus method. In [23] an investigation of ALOS-PALSAR polarimetric radar reflectance from reference forest stands and clear cuts in Swedish forests indicates a significant reduction in intensity in the HH and HV bands after cutting. Although this is for L-band, this might be expected to correspond to similar reductions in the VV and VH bands for the GEE Sentinel-1 C-band data, or equivalently to a negative definite change direction. This is confirmed in Figure 13, which also indicates that the bulk of harvesting occurred within just two consecutive 12-day intervals in December, 2017.
Figure 11. Color composite Sentinel-2 images (RGB = spectral bands 8, 3 and 2) northeast of Nahmint Lake, Vancouver Island, acquired 21 September 2017 (left) and 18 June 2018 (right). The long axes of the three fresh cuts range from 300 m to 600 m in length.

Figure 12. Color-coded change map derived from a time series of 33 Sentinel-1 images from the GEE archive over the area of Figure 11. The colors indicate the time of the most recent change: early dark blue, late orange. The series covers the period 7 September 2019 to 26 September 2018. Cutting took place within a short period early in the time series (blue pixels), see Figure 13. Background map data ©MapBox.
5. Conclusions

We have described the fusion of a statistically sound sequential change detection method with the power, convenience and vast data resources of the Google Earth Engine platform. Long time series of pre-processed dual polarization SAR imagery from the GEE archive from all over the globe can be easily accessed and quickly analyzed in a convenient graphical environment. Several illustrative examples were discussed. The software, including source code, is documented and freely available to anyone registered as a GEE user.

6. Discussion

Although we feel that they are outweighed by the advantages mentioned in the preceding Section, there are some drawbacks in the methodology that should be pointed out, and some possible suggestions for future work:

- First of all, and most significantly, the sequential omnibus tests on the GEE are carried out at the nominal scale of the archived Sentinel-1 data (10 m). This is because of the dependence of the Wishart distribution on the equivalent number of looks (ENL). Confining analysis to a single
scale precludes leveraging one of the great advantages of the Earth Engine, namely up-scaling to very large geographical regions. One way to mitigate this in future might be to download representative images with well-developed speckle statistics at different scales and then estimate the ENL values off-line, e.g., with the methods given in [24]. Then those values could be hard wired into the GEE code to allow running the algorithm at coarser scales and on larger scenes.

- The change detection algorithm is purely data driven and unsupervised: The physical cause of detected changes must be inferred from the context. Here, the Loewner order discussed in the text can offer additional information.
- It is our experience that very long time series, typically 75 images or more, can lead to stack overflow on the GEE servers. With typically a 6-day temporal resolution this still allows well over a year of continuous observation at any given location.
- The diagonal-only dual polarization matrix format necessitates resorting to the block diagonal version of the algorithm as discussed in the theory section and in [13]. It would be desirable to have access to the full $2 \times 2$ dual polarization matrix. We understand that the GEE developers are considering ways to ingest single look complex (SLC) Sentinel-1 imagery, which would solve this problem: The multi-look dual polarization matrix format could then be constructed from the SLC data.
- The GEE archive is updated very quickly, the Sentinel-1 images are available within a few days of acquisition. But for timely disaster assessment this may not be good enough. Thus the tools described here will be useful only in situations which are not extremely time critical.

Comparison of the performance of our methodology with other approaches cited here, for example using confusion matrices or ROC curves, presupposes the availability of reliable ground reference data. Such information is often very hard or even impossible to come by, especially in disaster assessment situations where remote sensing data are perhaps the only data source available to assess changes/damages. We have however made considerable effort to provide the computer software tools in a form (Docker container) which guarantees reproducibility, verifiability and ease of use in any field of application for which multi-temporal polarimetric SAR change detection is considered to be useful.

**Author Contributions:** Conceptualization, M.J.C.; Formal analysis, K.C. and A.A.N.; Funding acquisition, A.A.N.; Methodology, A.A.N., H.S. and K.C.; Software, M.J.C.; Supervision, A.A.N. and H.S.; Visualization, M.J.C.; Writing—original draft, M.J.C.; Writing—review & editing, M.J.C., A.A.N., H.S. and K.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors would like to express their thanks to Noel Gorelick and Simon Ilyushchenko at GEE for including, at the authors’ request, the incomplete Gamma function in the API (Noel) and S1 data in untruncated and floating point format (Simon).

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).