SPATIO-TEMPORAL ANALYSIS OF CHANGE WITH SENTINEL IMAGERY ON THE GOOGLE EARTH ENGINE

Morton J. Canty
Heinsberger Str. 18
D-52428 Jülich, Germany

Allan A. Nielsen
Technical University of Denmark
Applied Mathematics and Computer Science
DK-2800 Kgs. Lyngby, Denmark

1. INTRODUCTION
A characteristic task in remote sensing Earth observation involves the registration of changes which may signal environmentally significant events. The Sentinel-1 synthetic aperture radar (SAR) and the Sentinel-2 optical/visible-infrared space-borne platforms, with spatial resolutions of the order of 10-20 meters and revisit times of the order of days, provide an attractive source of data for change detection tasks, the SAR imagery especially providing 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 the processing the data. Here we make open-source automatic change detection software and for optical data also automatic radiometric normalization software available.

2. CHANGE DETECTION IN SAR DATA
In [1] a change detection procedure for multi-look polarimetric SAR data [2] 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. 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 observations (termed the equivalent number of looks, ENL) of a surface pixel 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 representations, when multiplied by the 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 $\Sigma$ (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 $\Sigma_1 = \Sigma_2$, i.e., the two observations were sampled from the same distribution and no change has occurred, and the alternative (change) hypothesis, $H_1$, is $\Sigma_1 \neq \Sigma_2$. 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.

For analysis of the situation with data from two time points, $k = 2$, see [3, 4, 5, 6]. In [7] 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 are characterized by the same $\Sigma$ (the null hypothesis $H_0$),

\[ H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_k (= \Sigma) \]

against the alternative ($H_1$) that at least one of the $\Sigma_i$, $i = 1, \ldots, k$, 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 [1])

\[ \ln Q = n \{ pk \ln k + \sum_{i=1}^{k} \ln |X_i| - k \ln |X| \}. \]

Here $n$ is ENL, the $X_i = n \Sigma_i$ (i.e., ENL times the observed covariance matrix) follow the complex Wishart distribution, $X_i \sim W_C(p, n, \Sigma_i)$, and $X = \sum_{i=1}^{k} X_i \sim W_C(p, nk, \Sigma)$. Also, if the hypothesis is true (”under $H_0$” in statistical parlance), $\Sigma = X/(knQ)$. $Q \in [0,1]$ 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)

\[ P[-2 \ln Q \leq z] \approx P\{ \chi^2((k-1)f) \leq z \}; \]

for $f = 9$ for quad pol, $f = 4$ for dual pol, $f = 2$ for dual pol diagonal only. The no-change probability is $1 - P(\chi^2((k-1)f)) \leq z$.

Furthermore this test can be factored into a sequence of tests involving hypotheses of the form $\Sigma_1 = \Sigma_2$ against $\Sigma_1 \neq \Sigma_2$, $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_{j-1} = \Sigma_j \neq \Sigma_i$, and so forth. More specifically, to test whether the first $1 < j < k$ complex variance-covariance matrices $\Sigma_i$ are equal, i.e., given that

\[ \Sigma_1 = \Sigma_2 = \cdots = \Sigma_{j-1} \]

then the likelihood ratio test statistic $R_j$ for testing the hypothesis

\[ H_{0, j} : \Sigma_j = \Sigma_1 \text{ against } H_{1, j} : \Sigma_j \neq \Sigma_1 \]

is given by (see [1])

\[ \ln R_j = n \{ p(j \ln j - (j-1) \ln (j-1)) \]

\[ + (j-1) \ln \sum_{i=1}^{j-1} X_i | + \ln |X_j| - j \ln \sum_{i=1}^{j} X_i | \}. \]

Proc. of the 2017 conference on Big Data from Space (BiDS’17) doi:10.2760/383579

28–30 November 2017

Toulouse, France
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 $R_j$)
\[ P(-2 \ln R_j \leq z_j) \approx P(\chi^2(f) \leq z_j). \]
The no-change probability is $1 - P(\chi^2(f) \leq z_j)$.

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 also on better approximations to the distributions of $Q$ and $R_j$ under the null hypotheses, see [1, 8].

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 (GEE) [9] which ingests Sentinel-1 (and Sentinel-2) data 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 (e.g., Sentinel-2 or Landsat) imagery, a data-driven, statistical approach to change detection is provided by the iteratively reweighted multivariate alteration detection (IR-MAD) algorithm [10, 4]. This method applies iterated canonical correlation analysis (CCA) to multispectral images from two time points before performing band-wise differences. The CCA orders the image bands according to similarity (correlation), rather than spectral wavelength. The differences between corresponding pairs of canonical variates are termed the MAD variates. Specifically, a MAD variate $Z$ is
\[ Z = a^T X - b^T Y \]
where $X$ represents the $m$-dimensional image at time point 1, $Y$ represents the $m$-dimensional image at time point 2, and $a$ and $b$ are the eigenvectors from the CCA. Thus $a^T X$ is a canonical variate for time point 1 and $b^T Y$ 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 $\rho$ (termed the canonical correlation which is maximized in CCA), and we have $m$ uncorrelated MAD variates with variance $2(1 - \rho)$.

In each iteration the values of each image pixel $j$ are weighted by $w_j$, which is the current estimate of the no-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)
\[ C^2 = \sum_{i=1}^{m} \frac{Z_i^2}{2(1 - \rho_i)} \]
will ideally follow 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 $C^2$)
\[ P(C^2 \leq c^2) \approx P(\chi^2(m) \leq c^2). \]

Hence the no-change probability used as weight $w_j$ in the iterations is $1 - P(\chi^2(m) \leq c^2)$. Iterations continue until the canonical correlations stop changing (or a maximum number of iterations is reached).

This procedure establishes an increasingly better background of no-change against which to detect significant change. 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 [11]. A threshold is set on the no-change probability (typically 95%) to identify invariant pixels in each scene. Their intensities are then regressed against each other band-wise to determine normalization coefficients. Because we have uncertainty in both variables here, we use orthogonal regression (as opposed to ordinary regression which places all uncertainty on the response variable). Again, the GEE is an ideal platform for accessing and processing (e.g., Sentinel-2 or Landsat) data in near real time.

4. CLOUD SOFTWARE

The authors have made available the necessary change detection software for interaction with the GEE on the open-source repository GitHub². The client-side programs run in a local Docker container serving a simple Flask web application. Apart from the Docker engine 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 polarimetric image time series, 4) export imagery to his or her Earth Engine assets folder or to Google Drive for further processing or visualization.

(Software is available also for local processing. 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 Github³,⁴.)

5. EXAMPLES

To illustrate, the Sentinel-1 multi-temporal change map in Figure 1 displays the color-coded time intervals in which the most recent changes in the 2016 growth period in an agricultural area southwest of Winnipeg, Manitoba, Canada, occurred. The yellow and red areas (seasonally late changes) will mostly correspond to grain harvesting. The change maps can be viewed interactively in the GEE Code Editor.

Figure 2 is a change frequency map showing shipping activity at the port of Tripoli, Libya, for a time series of 28 Sentinel-1 images. Heavy activity is concentrated to the northwest in the inner harbor.

²https://github.com/mortcanty/earthenGINE
³https://docs.docker.com
⁴https://mortcanty.github.io/src/tutoralsar.html
⁵https://mortcanty.github.io/src/tutorial.html
⁶https://code.earthengine.google.com/14d818dc83bed52608adf477999c76f8
⁷https://code.earthengine.google.com/5b543ad81805801d4c86a499b4171a8
Fig. 1. Sequential omnibus change map for a region southwest of the city of Winnipeg, Manitoba, Canada, showing the time of the most recent change (black none, blue early, red late). The time series consisted of 19 Sentinel-1 images from May through October, 2016.

Fig. 2. Sequential omnibus change map for the port of Tripoli, Libya, showing the frequency of changes. The time series consisted of 28 Sentinel-1 images from April through December, 2016.

The golden yellow signal in the Sentinel-2 bi-temporal change map of Figure 3 shows part of the large area devastated by a major forest fire southeast of Coimbra, Portugal, which broke out on June 17, 2017. Note that the IR-MAD method clearly discriminates changes due to agriculture (in blue and cyan). The extreme flooding caused by hurricane Harvey in August, 2017 is apparent in the IR-MAD change map of Figure 4 (green signal). The heaviest rains fell between initial landfall near Houston, Texas, on August 26, continuing until August 29. We interpret the color graduation from green to blue at the edges of the flooding signal as reflecting receding floodwaters by August 30, the time of the second acquisition. Note that the IR-MAD method clearly discriminates irrelevant changes due to cloud and cloud shadows (in red and dark gray).

Finally, Figure 5 illustrates relative radiometric normalization using two Landsat-7 ETM+ images. The first image (June 26, 2001) is used as reference, the second (August 29, 2001) as target, the target is normalized to the reference. Note, that the amount of change between the two acquisitions is considerable due to agricultural harvesting. Note also, that there is a clear difference in intensities especially.

1https://code.earthengine.google.com/a1f9a4a55783c0e958941e56f150594c
2https://code.earthengine.google.com/b19e90e6713448ec862e512cc8595b24
3https://code.earthengine.google.com/5f0c16f7922e9a7629971b7e393d00a8
Fig. 5. Relative radiometric normalization of two Landsat-7 ETM+ images (at-sensor radiances expressed in digital numbers) acquired over the town of Jülich, Germany, on June 26 and August 29, 2001. Top segment: target image August 29, middle segment: reference image June 26, bottom segment: radiometrically normalized target image. Bands 4, 5 and 7 are shown in RGB composite linearly stretched from 0 to 250. The 30m non-thermal bands 1, 2, 3, 4, 5 and 7 were processed with the IR-MAD transformation to determine the invariant pixels.

6. CONCLUSIONS

Examples based on both Sentinel-1 dual polarimetry 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. Also, for optical data an automatic radiometric normalization scheme is sketched and illustrated. The examples shown cover different application areas: agriculture, surveillance/remote monitoring of port traffic and natural disasters, here forest fire and flooding.

Generic, automatic techniques as these are expected to be useful in many other application areas also where the study of spatio-temporal dynamics is important. The introduction of software 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 researchers and practitioners alike.

7. REFERENCES


Proc. of the 2017 conference on Big Data from Space (BiDS’17) Toulouse, France 28–30 November 2017