Using the ESDL data cube for Sentinel-1 time series for deforestation mapping with Recurrence Metrics

Felix Cremer, Felix Glaser, Mikhail Urbazaev, and Christian Thiel

July 18, 2019

1 Introduction

The tropical forest ecosystems stabilise the world climate, the protection of the biodiversity and for the well being of a vast amount of the global population. In the last decade remote sensing technologies have played a substantial role in the consistent, reliable and timely information gathering about forest cover changes. With the REDD+ mechanism the use of remote sensing to monitor and map deforestation and degradation processes increased. Forest/Non-Forest maps are mostly operationally if they are based on optical sensors. Especially in the tropics, the optical imaging is hindered by clouds. In this study we propose a deforestation mapping approach based on Synthetic Aparture Radar (SAR) time series which takes the order of the time series into account to generate yearly deforestation maps.

For the data handling, we are using the Earth System Data Laboratory (ESDL) data cube. This is the first time, that the ESDL data cube is used with data of very high resolution of 20mx20m. The ESDL data cube eases the data handling by semi automatically multi threading the data analysis and by handling the data management, so that not the whole data set has to be loaded into the working memory.

2 Method

2.1 Recurrence Plots

Recurrence plots (RP) have been proposed by [1]. They are a method to visualize the recurrences of a time series. They are defined as follows:

$$R_{i,j} = \theta(\epsilon - |x_i - x_j|), i, j = 1, ..., N$$

hereby ϵ is a threshold value which indicates up to which distance two time steps are viewed as similar. θ is the Heaviside function which sets everything below

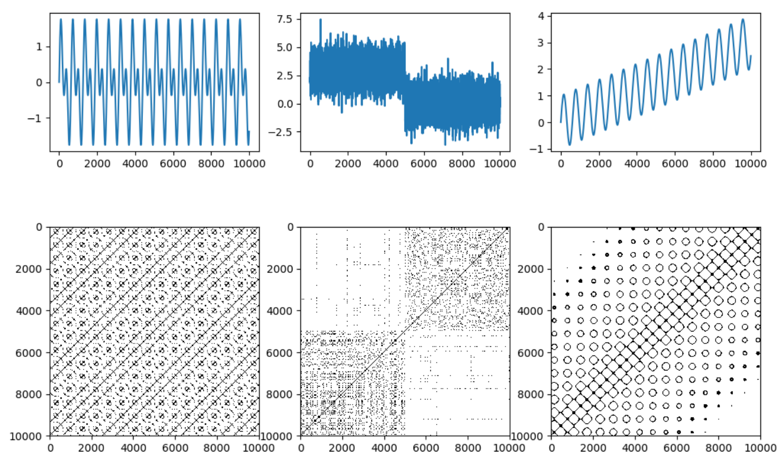

Figure 1: Recurrence Plots for a sine wave, a step function with noise, and a sine wave with trend.

zero to zero and every positive value to one. N is the number of time steps. This leads to a quadratic matrix with black dots where the time steps are similar to each other and white dots where they are distinct. The main diagonal is always black, because every time step is similar to itself. It is a nonlinear data analysis tool. Figure 1 shows example Recurrence plots of a sum of two sine waves with different frequencies, a step function from three to zero with an overlaid white noise with standard deviation 1 and third a sine wave with overlaying trend. For the composition of two different frequencies we can see a regular pattern with distinguished diagonals which are indicating the frequency. In the noisy step function we see four distinct quadrants in the recurence plot. In the two quadrants near the main diagonal every point is randomly similar to other points in this part of the time series with a high probability. In the other two quadrants, the probability is low, that two points are similar to a point in the other part of the step function. In the third example, we see a clear pattern, but these patterns are fading out to the edge of the recurrence plot. This is due to the difference of the values at the beginning and the end of the time series. Therefore we can use this pattern as an indicator for a trend in the time series.

These visual patterns can be quantified using recurrence quantification analysis (RQA) [2]. The simplest measure is the recurrence rate (RR) which is the number of recurrences in a recurrence plot devided by the squared number of time steps. It measures the density of the recurrence points in a RP. Another

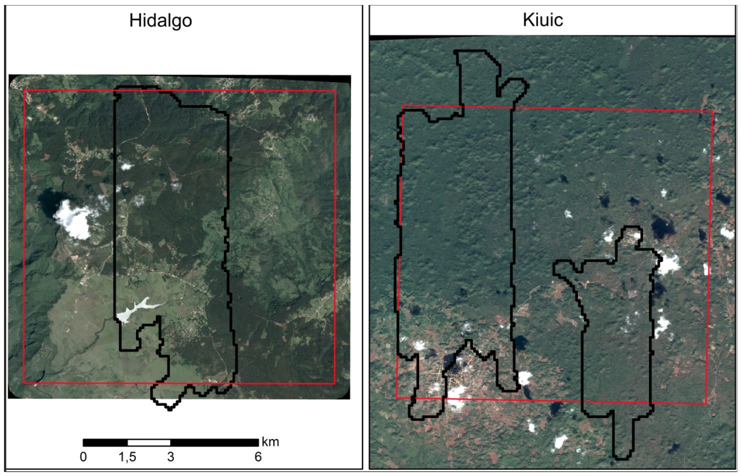

Figure 2: The testsites are located in Mexico and are dominated by temperate forests (Hidalgo) and tropical dry forest(Kiuic).

measure is the trend. It is defined as

$$TREND = \frac{\sum_{\tau=1}^{\tilde{N}} (\tau - \tilde{N}/2)(RR_\tau - \langle RR_\tau \rangle)}{\sum_{\tau=1}^{\tilde{N}} (\tau - \tilde{N}/2)}.$$

It is a linear regression coefficient over the recurrence rate of the diagonals in comparison to their distance to the main diagonal. It indicates if the process is drifting. For an in depth discussion of different RQA metrics see [3]. All of the results have been produced using the Julia RecurrenceAnalysis.jl package which is part of DynamicalSystems.jl [4].

3 Data preparation

We test the separability of stable forest and deforestation on two testsites in Mexico. One is mostly covered by temperate forests in central Mexico, the other is situated on the Yucatan peninsula and is covered by tropical dry forests. Figure 2 show very high resolution Pliades data of the two testsites. In the following, we are focusing on the Hidalgo testsite in central Mexico.

3.1 Preprocessing

We use the pyroSAR python package [5] to handle the raw data management and the data processing chain. This package allows to set a region of interest and returns a consistently preprocessed stack of SAR scenes. For this preprocessing, we use the SNAP software [6] in version 6. The single time steps are multilooked to a 10 m x 10 m pixel spacing. The orthorectification is based on the original orbit state vectors and the 30 m SRTM digital elevation model [7]. The preprocessing also included radiometric terrain flattening after [8] which results in γ^0 backscatter values.

3.2 Speckle Filter

We have reduced the speckle noise and other unwanted short-term perturbations of the data using the filter described in [9]. The filter is based on the Empirical Mode Decomposition a data adaptive alternative to the fourier transformation which can handle non-stationary data. Each pixel is separately decomposed into intrinsic mode functions (IMFs) of different temporal frequencies. In order to reduce the speckle, the two IMFs with the highest temporal frequencies are removed. This results in a nonparametric image transform that fully preserves the geometric resolution and has a similar speckle suppression than the Quegan 5x5 filter [10].

3.3 Ingestion into the ESDL datacube

We use the spatialist python package [11] to coregister all scenes in one raster grid in the DEM geometry after geocoding to achieve a subpixel coregistration precision. This is of eminent importance when the pixels are investigated in the temporal domain only. This operation returns a list of geotiff files for all time steps. For the data ingestion we convert the geotiff files into netcdf files via gdalwarp. We then use an own data provider which is a subclass of the NetCDFCubeSourceProvider class. Currently this provider is avaliable in a fork of the esdl-core package at
`https://github.com/felixcremer/esdl-core/blob/S1Prov/esdl/providers/sentinel1.py` and we plan to release it as an own submodule.

For every time step cross-polarized as well as co-polarized data are available from Sentinel-1. For this study we seperated the data into the ascending and descending orbits. In the future we would like to give the possibility to seperate the data due to the relative orbit of the acquisition, because the different orbits have different local incidence angles and this leads to different backscatter properties.

Since we want to keep the irregular time series of the underlying data intact, we are using a time axis with one day intervals. One day intervals are not yet working, because the generation script assumes intervals of multiple days. See issue url for further information. There are only scenes with values in every variable. For further processing of the data, the time axes could be rearranged to only contain valid data, but we found that for pure multitemporal analyses it is faster to use the original stack and implement the analysis functions so that they skip missing values.

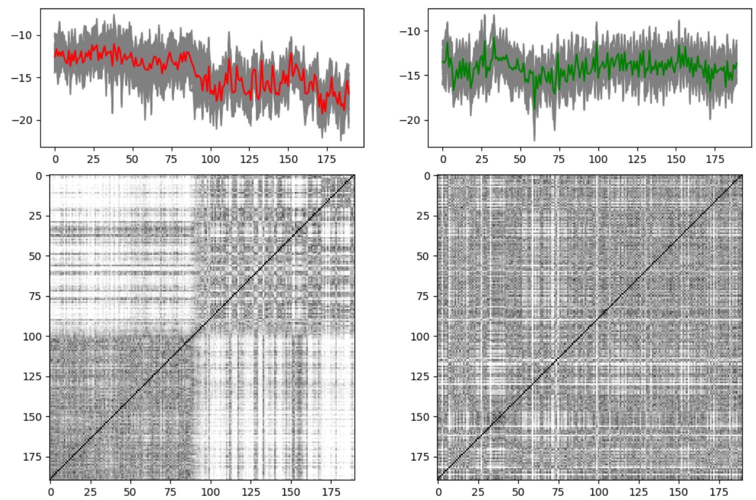

Figure 3: Recurrence Plots for a 3x3 matrix of deforestated(left) and stable forest(right) pixels. The gray lines in the above figure are all pixels and the colored line is the mean of these pixels.

4 Experimental Results

Figure 3 shows the recurrence plots of a neighbourhood of examplary forest and deforestation areas. The upper figures shows the VH time series of a 7x7 matrix of a deforested area (left) and a stable forest (right) with the mean of these pixels in red respectively green. The bottom figures show the grayscale of the corresponding recurrence plots. Black points are similar in every pixel of the neighborhood and white points in none. There is a clear drop in the backscatter time series of the deforested area, but then the signal is going up to a similar level than before the deforestation. In the Recurrence Plot this change is visible as large white areas in the corners of the recurrence plots. In the forest, the signal is stable which is visible in the recurrence plot as a white noise, because of the small differences in the backscatter in every time step.

Figure 4 shows the RQA TREND statistic from two year VH data from march 2017 till march 2019 in comparison to the range between the multitemporal 5th to 95th percentile (prange).. For this time frame are 187 time steps available. The red polygons are deforestations which happened between september 2016 and october 2017 and the violet ones deforestations between October 2017 and July 2018. The later deforestations which are fully in our sensing period are distinctly visible as clear black areas in the RQA TREND metric. The EMD filter reduces the pixel inside the deforestation areas which are having a

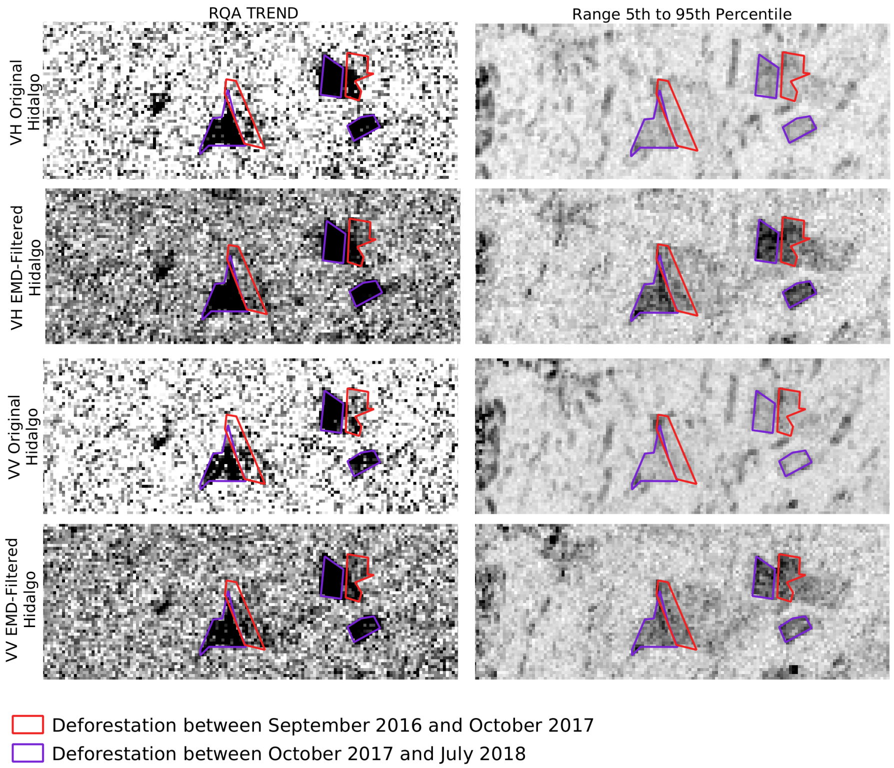

Figure 4: RQA TREND statistic. The red polygons are deforestations between October 2017 and July 2018 and the green polygons are stable forest areas. These polygons have been select from VHR Pleiades data.

low RQA TREND metric especially in the left polygon in VV data. . In the prange metric from the original data is no clear difference visible between the deforested areas and the surrounding stable forest areas. Applying the EMD filter makes the deforestations visible. The earlier deforestations are also partly visible, because the deforestations happened in the beginning of the sensing period. Therefore the 95th percentile by the still standing forest and the 5th percentile is dominated by the already deforested signal. The prange metric gains especially from the temporal denoising by the EMD filter. Overall, the cross-pol data better suited to map deforestations than co-pol data.

5 Discussion

In this work, we showed, that the RQA TREND metric is a good indicator for deforestations. We compared the RQA TREND to the range between the 5th and the 95th multitemporal percentile. Because the RQA TREND is also using the order of the time series to derive information about the underlying land cover, it gives a clearer picture about deforestations.

The ESDL data cube enhances the work with large stacks of Sentinel-1 data by allowing to work with data on disk in contrast to only data which would fit into RAM. Furthermore, the ESDL infrastructure makes it very easy to use multithreading over the whole data stack. This will make it possible, to redo the analysis on larger regions without having to deteriorate the spatial or temporal resolution.

6 Outlook

We are going to tidy up the ESDL data cube generation scripts and are publishing them as an standalone python package. Building on this, we want to integrate the ESDL data cube generation into pyroSAR, so that the user has an easy access point to use Sentinel-1 data in the ESDL infrastracture. This is especially relevant in the discussion about analysis ready data, which will further increase the uptake of SAR data.

We are going to redo the analysis now that it is possible to use shapefiles together with the datacube. Until now, we had to export our results, to analyse them further with in situ data. Furthermore, we want to quantify the separability enhancement between deforestation and stable forest for RQA TREND compared to the prange. And we are going to extend the analysis to the testsite in Yucatan.

Acknowledgment

This work uses Copernicus Sentinel data 2017-2019.

References

[1] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *EPL (Europhysics Letters)*, vol. 4, no. 9, p. 973, 1987.

[2] J. P. Zbilut and C. L. Webber, "Embeddings and delays as derived from quantification of recurrence plots," *Physics Letters A*, vol. 171, no. 3, pp. 199 – 203, 1992.

[3] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5, pp. 237 – 329, 2007.

[4] G. Datseris, "Dynamicalsystems.jl: A julia software library for chaos and nonlinear dynamics," *Journal of Open Source Software*, vol. 3, p. 598, mar 2018.

[5] J. Truckenbrodt, F. Cremer, I. Baris, and R. Kidd, "pyrosar a python framework for large-scale sar satellite data processing," 2017.

[6] "Esa snap – esa sentinel application platform." [Version 6.0.2].

[7] N. J. P. Laboratory, "Nasa shuttle radar topography mission united states 1 arc second version 3," 2013.

[8] D. Small, "Flattening gamma: Radiometric terrain correction for sar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 3081–3093, Aug 2011.

[9] F. Cremer, M. Urbazaev, C. Berger, M. Mahecha, C. Schmullius, and C. Thiel, "An image transform based on temporal decomposition," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, pp. 537–541, April 2018.

[10] S. Quegan and J. J. Yu, "Filtering of multichannel sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 2373–2379, Nov 2001.

[11] J. Truckenbrodt, F. Cremer, and I. Baris, "spatialist a python module for spatial data handling," 2019.