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 3, 2019

Abstract

Forest ecosystems REDD+ process order of time series SAR

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 SAR time series.

Misses overview about established methods. Especially the percentile range

2 Method

2.1 Recurrence Plots

Recurrence plots (RP) have been proposed by [**?**]. 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 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

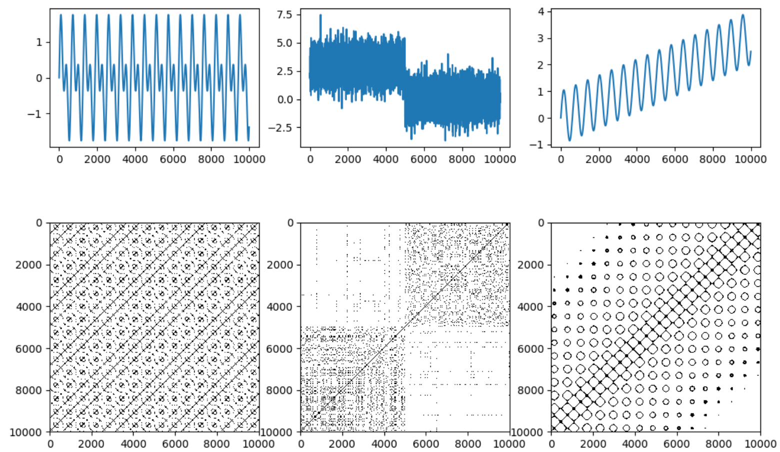

Figure 1: Recurrence Plots for a sine wave, a step function with noise, and a sine wave with trend.

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) [**?**]. 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 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 overview of RQA measures see Table 1 and for an in depth discussion [?]. All of the results have been produced using the Julia RecurrenceAnalysis.jl package [?].

Name	Formula	Interpretation
Recurrence Rate	$RR = \frac{\sum_{i,j=1}^{N} (R_{i,j})}{N^2}$	Probability that a state of the system recurs.
Determinism	$DET = \frac{\sum_{l=lmin}^{N} (lP(l))}{\sum_{l=1} (lP(l))}$	Measures the predictability of the signal.
Average Length of Diagonal Structues	$dlavg = \frac{\sum_{l=lmin} (lP(l))}{\sum_{l=lmin}^{N} (P(l))}$	
Maximum Length of Diagonal Structures	$L_{max} = \max(l_i)$	
Divergence	$DIV = \frac{1}{L_{max}}$	
Entropy of diagonal structures	$ENTR = \sum_{i=1}^{N} P(l_i) * log(P(l_i))$	
Trend	$TREND = \frac{\sum_{\tau=1}^{\tilde{N}} (\tau - \tilde{N}/2)(RR_\tau - \langle RR_\tau \rangle)}{\sum_{\tau=1}^{\tilde{N}} (\tau - \tilde{N}/2)}$	
Laminarity	$LAM = \frac{\sum_{v=vmin} (vP(v))}{\sum_{v=1}^{N} (vP(v))}$	
Trapping time	$TT = \frac{\sum_{v=vmin} (vP(v))}{\sum_{v=vmin}^{N} (P(v))}$	
Maximum Length of vertical structures	$V_{max} = \max_{i=1,\ldots,N} (v_i)$	
Entropy of vertical structures	$VENTR = \sum_{i=1}^{N} P(v_i) * log(P(v_i))$	
Mean recurrence time	$MRT = \frac{\sum_{i=1}^{N} (w_i P(w_i))}{\sum_1^N (P(w_i))}$	
Recurrence time entropy	$RTE = \sum_{i=1}^{N} P(w_i) * log(P(w_i))$	
Number of the most probable recurrence times	$NMPRT = \max w_i$	

Table 1: Overview of Recurrence Quantification measures.

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.

3.1 Preprocessing

We use the pyroSAR python package [?] 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 [?] 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 [?]. The preprocessing also included radiometric terrain flattening after [?] which results in γ^0 backscatter values.

3.2 Ingestion into the ESDL datacube

We use the spatialist python package [?] to coregister all scenes in one raster grid in the DEM geometry after geocoding to achieve a subpixel coregistration precision which 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.

From Sentinel-1 data are for every time step cross-polarized as well as co-polarized data available. 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. There are only "200" scenes with values in every variable.

4 Experimental Results

4.1 Separability analysis

Figure 3 shows the recurrence plots of a neighbourhood of examplary forest and deforestation areas. The upper figures shows the time series of a 7x7 matrix

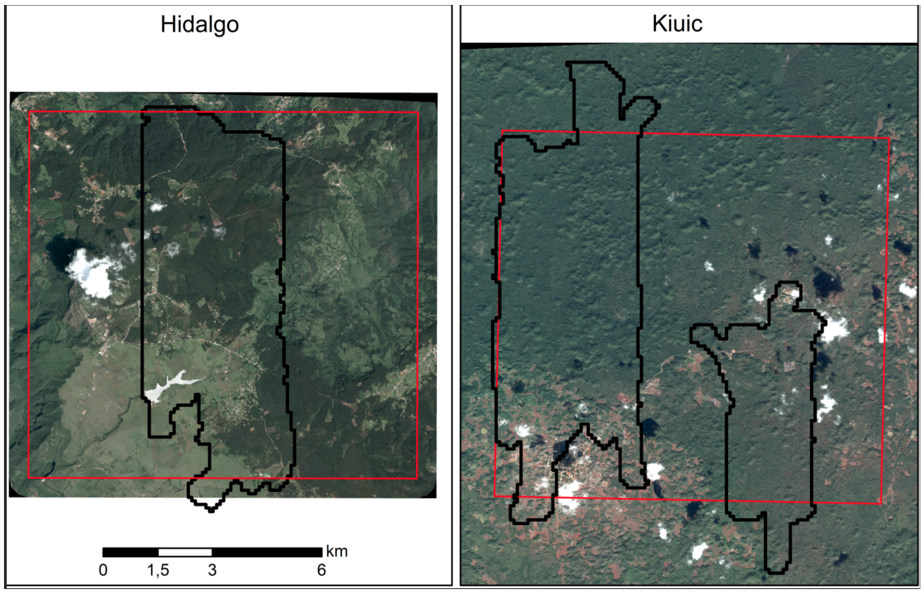

Figure 2: The testsites are located in Mexico and are dominated by temperate forests (Hidalgo) and tropical dry forest(Kiuic).

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.

Figure 4 shows the RQA TREND statistic from two year VH data from march 2017 till march 2019. For this time frame are 187 time steps available. The red polygons are deforestations which happened between october 2017 and july 2018 and the green are stable forest areas. Most of the deforested areas are clearly distinguishable as a distinct black area compared to the surrounding grey. The not detected deforestation areas are too late in the sensing period. So that the backscatter change due to the deforestation did not have a long enough effect on the TREND statistic.

Figure **??** shows the histograms for the RQA trend metric and the percentile range for the deforestation areas in red and the stable forest areas in green.

Show that the TREND metric is enhancing the separability of deforested areas and stable forests. Quantify it.

Is there a recommendation, which threshold we should use for the percentile range?

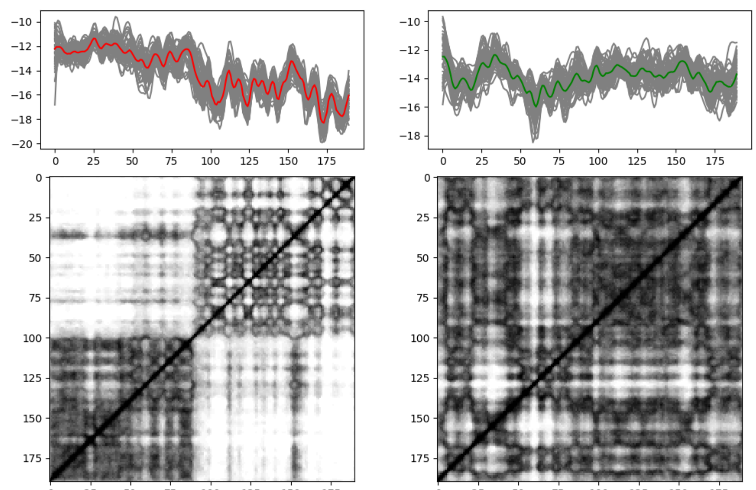

Figure 3: Recurrence Plots for a 3x3 matrix of deforestated(left) and stable forest(right) pixels. The gray lines in the above figure are all pixels and the colored line is the mean of these pixels.

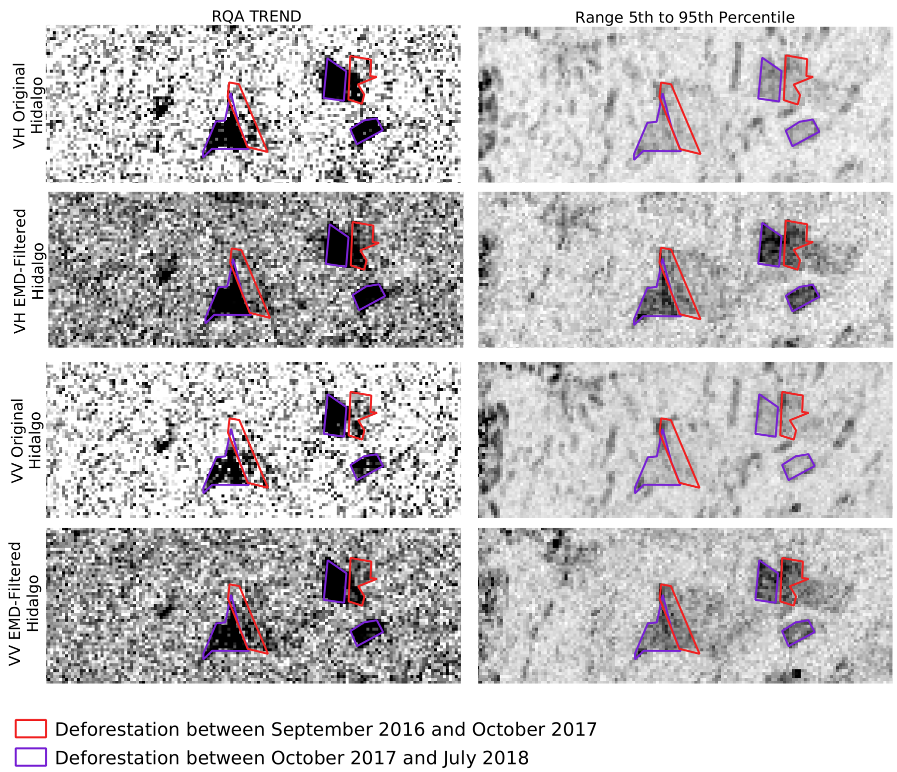

Figure 4: Comparison of the RQA trend and the range between the 5th and 95th temporal percentile. The deforested areas are better visible in the RQA trend metric compared to the percentile range and the deforestation is better visible in the cross-pol data. Using the EMD-Filter enhances the distinguishability of the deforestation areas, but it increases also the number of stable forest pixels which have similar values than deforested areas.

5 Discussion

Do I need to do a comparison against optical deforestation maps? Will do a comparison against the percentile range.

Should I show the deforested area in the testsite figure? Yes please. If I am showing the testsite figure at all.

6 Summary and Conclusion

In this paper we showed, that the use of the inherent order of a time series conveys information, which can be used to better map deforestation.

Acknowledgment

This work was funded by the DLR in the Sentinel4REDD project (FKZ:50EE1540) and the DFG project HyperSense. This work uses Copernicus Sentinel data 2017-2019.