Satellite Image Time Series Filtering with SITS

Rolf Simoes   *National Institute for Space Research (INPE), Brazil*
Gilberto Camara   *National Institute for Space Research (INPE), Brazil*
Pedro R. Andrade   *National Institute for Space Research (INPE), Brazil*

This vignette describes the time series filtering techniques available in the SITS package. These include the Savitsky-Golay, Whittaker, and Kalman filters, as well as specialised filters for satellite images, which are the envelope filter and the ARIMA filter for cloud removal.

Filtering techniques in SITS

Satellite image time series generally is contaminated by atmospheric influence, geolocation error, and directional effects [Lambin and Linderman, 2006]. Atmospheric noise, sun angle, interferences on observations or different equipment specifications, as well as the very nature of the climate-land dynamics can be sources of variability [Atkinson et al., 2012]. Inter-annual climate variability also changes the phenological cycles of the vegetation, resulting in time series whose periods and intensities do not match on an year to year basis. To make the best use of available satellite data archives, methods for satellite image time series analysis need to deal with *noisy* and *non-homogeneous* data sets. In this vignette, we discuss filtering techniques to improve time series data that present missing values or noise.

The literature on satellite image time series have several applications of filtering to correct or smooth vegetation index data. The `sits` have support for Savitzky–Golay (`sits_sgolay()`), Whitaker (`sits_whittaker()`), envelope (`sits_envelope()`) and, the "cloud filter" based on ARIMA methods (`sits_ndvi_arima_filter()`) filters. The first two filters are commonly used in the literature, while the remaining two have been developed by the authors.

Various somewhat conflicting results have been expressed in relation to the time series filtering techniques for phenology applications. For example, in an investigation of phenological parameter estimation, Atkinson et al. [2012] found that the Whittaker and Fourier transform approaches were preferable to the double logistic and asymmetric Gaussian models. They applied the filters to preprocess MERIS NDVI time series for estimating phenological parameters in India. Comparing the same filters as in the previous work, Shao et al. [2016] found that only Fourier transform and Whittaker techniques improved interclass separability for crop classes and significantly improved overall classification accuracy. The authors used MODIS NDVI time series from the Great Lakes region in North America. Zhou et al. [2016] found that asymmetric Gaussian model outperforms other filters over high latitude boreal biomes, while the Savitzky-Golay model gives the best reconstruction performance in tropical evergreen broadleaf forests. In the remaining biomes, Whittaker gives superior results. The authors compare all

previous mentioned filters plus Savitzky-Golay method for noise removal in MODIS NDVI data from sites spread worldwide in different climatological conditions. Many other techniques can be found in applications of satellite image time series such as curve fitting [Bradley et al., 2007], wavelet decomposition [Sakamoto et al., 2005], mean-value iteration, ARMD3-ARMA5, and 4253H [Hird and McDermid, 2009]. Therefore, any comparative analysis of smoothing algorithms depends on the adopted performance measurement.

One of the main uses of time series filtering is to reduce the noise and miss data produced by clouds in tropical areas. The following examples use data produced by the PRODES project [INPE, 2017], which detects deforestation in the Brazilian Amazon rain forest through visual interpretation. This data set is called `prodes_226_04` and is provided together with the `sits` package. It has 617 samples from a region corresponding to the standard Landsat Path/Row 226/064. This is an area in the East of the Brazilian Pará state. It was chosen because of its huge cloud cover from November to March, which is a significant factor in degrading time series quality. Its NDVI and EVI time series were extracted from a combination of MOD13Q1 and Landsat8 images (to best visualize the effects of each filter, we selected only NDVI time series).

Common interface to SITS filter functions

The SITS package uses a common interface to all filter functions with the `sits_filter`. The function has two parameters: `data` for the dataset to be filtered and `filter` for the filter to be applied. To aid on data visualisation, all bands which are filtered have a suffix which is appended, as shown in the examples below. The following filters are available and described in what follows:

- Savitzky–Golay filter (`sits_sgolay`)

- Whittaker filter (`sits_whittaker`)

- Envelope filter (`sits_envelope`)

- ARIMA filter for cloud removal in NDVI band (`sits_ndvi_arima_filter`)

- Cloud filter (`sits_cloud_filter`)

- Kalman filter (`sits_kalman`)

Savitzky–Golay filter

The Savitzky-Golay filter works by fitting a successive array of $2n+1$ adjacent data points with a d-degree polynomial through linear least squares. The central point i of the window array assumes the value of the interpolated polynomial. An equivalent and

much faster solution than this convolution procedure is given by the closed expression

$$\hat{x}_i = \sum_{j=-n}^{n} C_j \, x_{i+j},$$

where $\hat{x}$ is the the filtered time series, C_j are the Savitzky-Golay smoothing coefficients, and x is the original time series.

The coefficients C_j depend uniquely on the polynomial degree (d) and the length of the window data points (given by parameter n). If $d = 0$, the coefficients are constants $C_j = 1/(2n+1)$ and the Savitzky-Golay filter will be equivalent to moving average filter. When the time series is equally spaced, the coefficients have analytical solution. According to Madden [1978], for $d \in [2,3]$ each C_j smoothing coefficients can be obtained by

$$C_j = \frac{3(3n^2 + 3n - 1 - 5j^2)}{(2n+3)(2n+1)(2n-1)}.$$

In general, the Savitzky-Golay filter produces smoother results for a larger value of n and/or a smaller value of d [Chen et al., 2004]. The optimal value for these two parameters can vary from case to case. In SITS, the user can set the order of the polynomial using the parameter order (default = 3), the size of the temporal window with the parameter length (default = 5), and the temporal expansion with the parameter scaling (default = 1). The following example shows the effect of Savitsky-Golay filter on the original time series.

```
# Take NDVI band of the first sample data set
point.tb <- sits_select_bands(prodes_226_064[1,], ndvi)
# apply SavitzkyGolay filter
point_sg.tb <- sits_filter(point.tb, filter = sits_sgolay())
# plot the series
sits_merge(point_sg.tb, point.tb) %>% sits_plot()
```

Whittaker filter

The Whittaker smoother attempts to fit a curve that represents the raw data, but is penalized if subsequent points vary too much [Atzberger and Eilers, 2011]. The Whittaker filter is a balancing between the residual to the original data and the "smoothness" of the fitted curve. The residual, as measured by the sum of squares of all n time series points deviations, is given by

$$RSS = \sum_i (x_i - \hat{x}_i)^2,$$

where x and $\hat{x}$ are the original and the filtered time series vectors, respectively. The smoothness is assumed to be the measure of the the sum of the squares of the third order differences of the time series [Whittaker, 1922] which is given by

$$SSD = (\hat{x}_4 - 3\hat{x}_3 + 3\hat{x}_2 - \hat{x}_1)^2 + (\hat{x}_5 - 3\hat{x}_4 + 3\hat{x}_3 - \hat{x}_2)^2$$
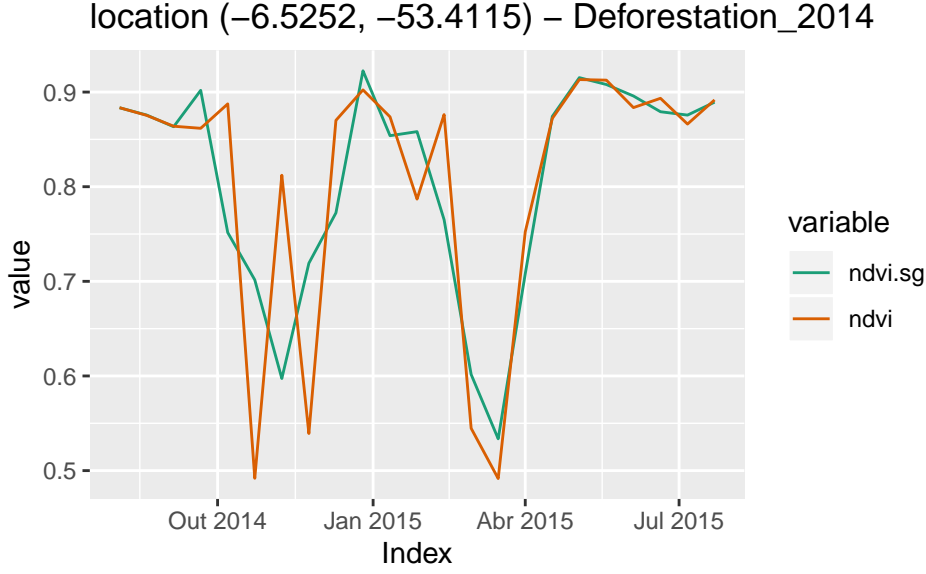$$+ \dots + (\hat{x}_n - 3\hat{x}_{n-1} + 3\hat{x}_{n-2} - \hat{x}_{n-3})^2.$$

Figure 1: Savitzky-Golay filter applied on a one-year NDVI time series.

The filter is obtained by finding a new time series $\hat{x}$ whose points minimize the expression

$$RSS + \lambda SSD,$$

where λ, a scalar, works as an "smoothing weight" parameter. The minimization can be obtained by differentiating the expression with respect to $\hat{x}$ and equating it to zero. The solution of the resulting linear system of equations gives the filtered time series which, in matrix form, can be expressed as

$$\hat{x} = (I + \lambda D^{\mathsf{T}} D)^{-1} x,$$

where I is the identity matrix and

$$D = \begin{bmatrix} 1 & -3 & 3 & -1 & 0 & 0 & \cdots \\ 0 & 1 & -3 & 3 & -1 & 0 & \cdots \\ 0 & 0 & 1 & -3 & 3 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

is the third order difference matrix. The Whitakker filter can be a large but sparse optimization problem, as we can note from D matrix.

Whittaker smoother has been used only recently in satellite image time series investigations. According to Atzberger and Eilers [2011], the smoother has an advantage over other filtering techniques such as Fourier and wavelets as it does not assume signal periodicity. Moreover, the authors argue that is enables rapid processing of large amounts of data, and handles incomplete time series with missing values.

In the `sits` package, the Whittaker smoother has two parameters: `lambda` controls the degree of smoothing and `differences` the order of the finite difference penalty. The default values are `lambda = 1` and `differences = 3`. Users should be aware that increasing `lambda` results in much smoother data. When dealing with land use/land

cover classes that include both natural vegetation and agriculture, a strong smoothing can reduce the amount of noise in natural vegetation (e.g., forest) time series; however, higher values of lambda reduce the information present in agricultural time series, since they reduce the peak values of crop plantations.

```
# apply Whitaker filter to a time series sample for the NDVI band from 2000 to 2016
# merge with the original data
# plot the original and the modified series
point_whit <- sits_filter(point_ndvi, sits_whittaker(lambda = 3))
point_whit %>%
  sits_merge(point_ndvi) %>%
  sits_plot()
```
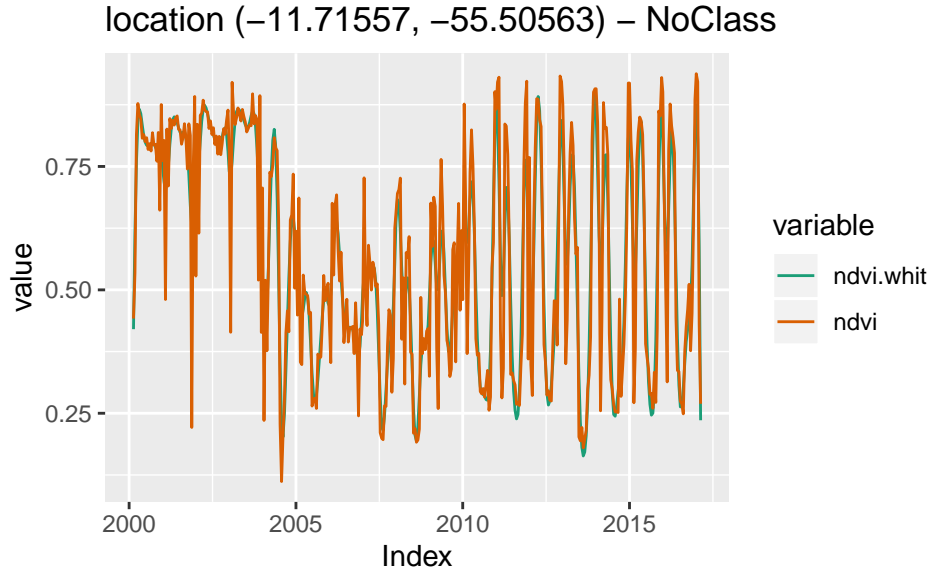

Figure 2: Whittaker smoother filter applied on 16-year NDVI time series

Envelope filter

This filter can generate a time series corresponding to the superior (inferior) bounding of the input signal. This is accomplished through a convoluting window (odd length) that attributes to the point i, in the resulting time series, the maximum (minimum) value of the points in the window. The i point corresponds to the central point of the window. It can be defined as

$$u_i = \max_k(\{x_k : |i - k| \leq 1\}),$$

whereas an lower dilation is obtained by

$$l_i = \min_k(\{x_k : |i - k| \leq 1\}).$$

Here, x is the input time series and, k and i are vector indices.

5

The `sits_envelope()` can combine both maximum and minimum window sequentially. The function can receive a string sequence with "U" (for maximization) and "L" (for minimization) characters passed to its parameter. A repeated sequence of the same character is equivalent to one operation with a larger window. The sequential operations on the input time series produces the final filtered result that is returned.

The envelope filter can be viewed through mathematical morphology lenses, a very common field in digital image processing [Haralick et al., 1987]. Here the operations of "U" and "L" corresponds to the *dilation* and *erosion* morphological operators applied to univariate arrays [Vávra et al., 2004]. Furthermore, the compounds operation of *opening* and *closing* can be obtained by "UL" and "LU", respectively. This technique has been applied on time series analysis in other fields [Accardo et al., 1997] but, for our knowledge, there is no application in satellite image time series literature.

In the following example we can see an application of `sits_envelope()` function. There, we performs the *opening filtration* and *closing filtration* introduced by Vávra et al. [2004]. The correspondent operations sequence are "ULLULUUL" and "LUULULLU".

```
# Take the NDVI band of the first sample data set
point.tb <- sits_select_bands(prodes_226_064[1,], ndvi)
# apply envelope filter (remove downward and upward noises)
point_env1.tb <- sits_filter(point.tb, sits_envelope(operations = "ULLULUUL", bands_suffix
point_env2.tb <- sits_filter(point.tb, sits_envelope(operations = "LUULULLU", bands_suffix

# plot the series
sits_merge(point_env1.tb, point_env2.tb) %>%
    sits_merge(point.tb) %>%
    sits_plot()
```
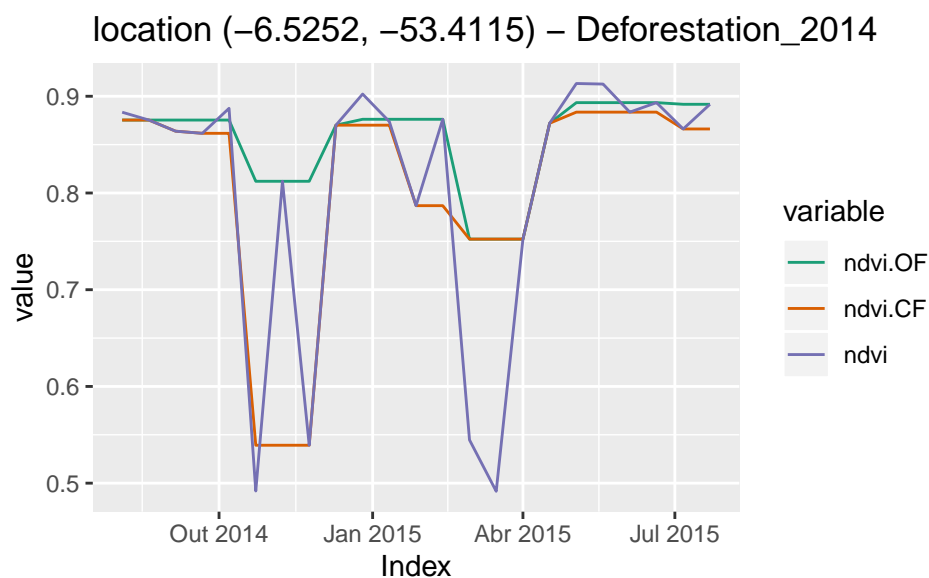

Figure 3: Envelope filter applied on one-year NDVI time series. The examples uses two morfological filters: opening filtration (.OF) and closing filtration (.CF).

ARIMA filter for cloud removal

The cloud filter makes use of the well known autoregressive integrated moving average (ARIMA) model. The algorithm looks to the first order difference time series for points where the value is above a certain threshold. This procedure selects only those points with large variations in the original time series, probably associated with noise. Finally, these points are replaced by the ARIMA correspondent values.

The parameters of the ARIMA model can be set by the user. Please see `arima` for the detailed description of parameters p, d, and q.

```r
# Take the NDVI band of the first sample data set
point <- sits_select_bands(prodes_226_064[1,], ndvi)
# apply ARIMA filter
point_cf <- sits_filter(point, sits_ndvi_arima())
# plot the series
sits_merge(point_cf, point) %>% sits_plot()
```
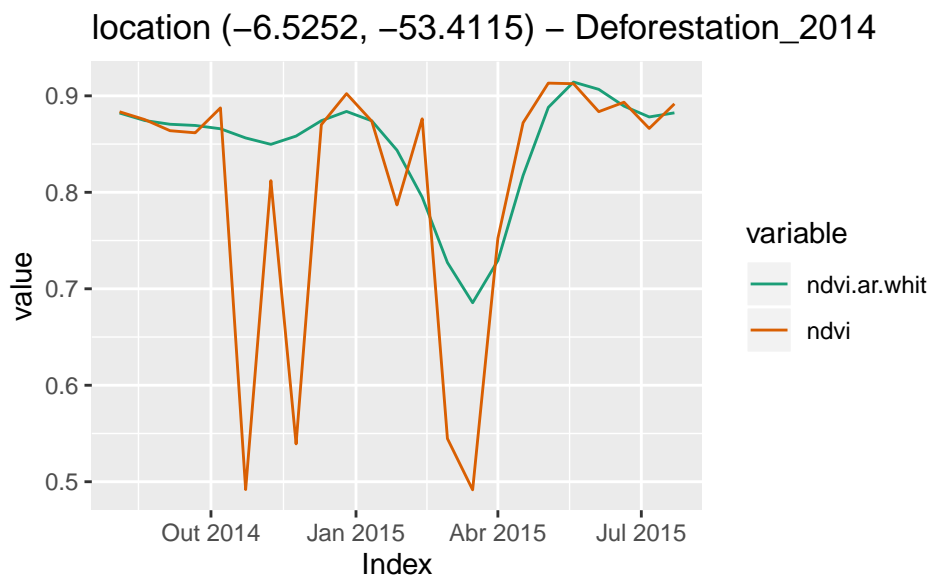

Figure 4: ARIMA cloud filter applied on one-year NDVI time series

References

Agostino Accardo, M Affinito, M Carrozzi, and F Bouquet. Use of the fractal dimension for the analysis of electroencephalographic time series. *Biological cybernetics*, 77(5): 339–350, 1997.

Peter M Atkinson, C Jeganathan, Jadu Dash, and Clement Atzberger. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote sensing of environment*, 123:400–417, 2012.

Clement Atzberger and Paul HC Eilers. Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements. *International Journal of Remote Sensing*, 32(13):3689–3709, 2011.

Bethany A Bradley, Robert W Jacob, John F Hermance, and John F Mustard. A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. *Remote Sensing of Environment*, 106(2):137–145, 2007.

Jin Chen, Per. Jönsson, Masayuki Tamura, Zhihui Gu andBunkei Matsushita, and Lars Eklundh. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment*, 91(3-4):332 – 344, 2004.

Robert M Haralick, Stanley R Sternberg, and Xinhua Zhuang. Image analysis using mathematical morphology. *IEEE transactions on pattern analysis and machine intelligence*, (4):532–550, 1987.

Jennifer N Hird and Gregory J McDermid. Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*, 113(1): 248–258, 2009.

INPE. Amazon Deforestation Monitoring Project (PRODES), 2017. URL www.obt.inpe.br/prodes.

E.F. Lambin and M. Linderman. Time series of remote sensing data for land change science. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7):1926–1928, 2006.

Hannibal H Madden. Comments on the Savitzky-Golay convolution method for least-squares-fit smoothing and differentiation of digital data. *Analytical chemistry*, 50(9): 1383–1386, 1978.

Toshihiro Sakamoto, Masayuki Yokozawa, Hitoshi Toritani, Michio Shibayama, Naoki Ishitsuka, and Hiroyuki Ohno. A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment*, 96(3-4):366–374, 2005.

Yang Shao, Ross S. Lunetta, Brandon Wheeler, John S. Iiames, and James B. Campbell. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. *Remote Sensing of Environment*, 174:258–265, 2016.

F. Vávra, P. Nový, H. Mašková, M. Kotlíková, and A. Netrvalová. Morphological filtration for time series. In *Proceedings of the 3rd International Conference APLIMAT*, pages 983–989, Bratislava, Slovakia, 2004. Slovak University of Technology in Bratislava.

Edmund T Whittaker. On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–75, 1922.

Jie Zhou, Li Jia, Massimo Menenti, and Ben Gorte. On the performance of remote sensing time series reconstruction methods – a spatial comparison. *Remote Sensing of Environment*, 187:367–384, 2016.