Recibido: 26/02/2018
Aceptado: 06/06/2018

REVISTA DE TELEDETECCIÓN
Asociación Española de Teledetección
(2018) 51, 19-31
ISSN 1133-0953
EISSN 1988-8740
https://doi.org/10.4995/raet.2018.9749

TISSBERT: A benchmark for the validation and comparison of NDVI time series reconstruction methods

Julien, Y.*, Sobrino, J. A.

Global Change Unit, Image Processing Laboratory, University of Valencia, P.O. Box 22085, E-46071 Valencia, Spain.

Abstract: This paper introduces the Time Series Simulation for Benchmarking of Reconstruction Techniques (TISSBERT) dataset, intended to provide a benchmark for the validation and comparison of time series reconstruction methods. Such methods are routinely used to estimate vegetation characteristics from optical remotely sensed data, where the presence of clouds decreases the usefulness of the data. As for their validation, these methods have been compared with previously published ones, although with different approaches, which sometimes lead to contradictory results. We designed the TISSBERT dataset to be generic so that it could simulate realistic reference and cloud-contaminated time series at global scale. To that end, we estimated both cloud-free and cloud-contaminated Normalized Difference Vegetation Index (NDVI) statistics for randomly selected control points and each day of the year from the Long Term Data Record Version 4 (LTDR-V4) dataset by assuming different statistical distributions. The best approach was then applied to the whole dataset, and validity of the results were estimated through the Kolmogorov-Smirnov statistic. The dataset elaboration is described thoroughly along with how to use it. The advantages and drawbacks of this dataset are then discussed, which emphasize the realistic simulation of the cloud-contaminated and reference time series. This dataset can be obtained from the authors upon demand. It will be used in a next paper to compare widely used NDVI time series reconstruction methods.

Key words: NDVI, gap-filling, reconstruction, dataset, comparison.

TISSBERT: una referencia para la validación y la comparación de métodos para la reconstrucción de series temporales de NDVI

Resumen: En este trabajo se presenta la base de datos titulada *Time Series Simulation for Benchmarking of Reconstruction Techniques* (TISSBERT) con el propósito de ofrecer una herramienta para la validación y la comparación de métodos para la reconstrucción de series temporales. Tales métodos se usan de manera rutinaria para la estimación de características de la vegetación a partir de datos obtenidos por teledetección óptica, donde la presencia de nubes disminuye su utilidad. En cuanto a su validación, estos métodos se han comparado con otros publicados anteriormente, aunque desde perspectivas diferentes, lo cual conduce a resultados contradictorios. La base de datos TISSBERT se ha diseñado como una herramienta genérica para una simulación realista a escala global de series temporales de referencia o contaminadas por nubes. Para ello, se estimaron estadísticas de *Normalized Difference Vegetation Index* (NDVI) con y sin contaminación de nubes para unos píxeles de control seleccionados de manera aleatoria, y para cada día del año, usando la base de datos *Long Term Data Record Version* 4 (LTDR-V4), y probando con varias distribuciones estadísticas. La mejor metodología se aplicó al conjunto de la base de datos, y la validez de

To cite this article: Julien, Y., Sobrino, J. A. 2018. TISSBERT: A benchmark for the validation and comparison of NDVI time series reconstruction methods. *Revista de Teledetección***,** 51, 19-31. https://doi.org/10.4995/raet.2018.9749

* *Corresponding author: yves.julien@uv.es*

los resultados se comprobó con la prueba de Kolmogorov-Smirnov. La elaboración de la base de datos se describe detalladamente así como la manera de usarla. Finalmente, se analizan las ventajas y los inconvenientes de la base de datos TISSBERT, los cuales enfatizan la simulación realista de series temporales de referencia y con contaminación nubosa. Esta base de datos se puede obtener gratuitamente de los autores, y se usará en un futuro para comparar métodos usuales de reconstrucción de series temporales de NDVI.

Palabras clave: NDVI, relleno de huecos, reconstrucción, base de datos, comparación.

1. Introduction

Due to both its mathematical simplicity and its wide availability, the Normalized Difference Vegetation Index (NDVI) is routinely used to monitor vegetation from space (Rouse *et al.,* 1974; Tucker, 1979). Indeed, by retrieving information in red and near-infrared wavelengths, this index provides information on vegetation greenness, and by proxy on productivity. However, when we observe vegetation from space, the more than occasional presence of clouds decreases the usefulness of the data. For example, in temperate areas, the estimation of vegetation phenology phases from NDVI time series, such as start and end of season, is complicated by the often persisting cloud cover in spring and autumn seasons. This has strong implications for the characterization of vegetation activity, and also for the retrieval of key phenological parameters, considered widely as indicators of vegetation response to climate change. This is also the case for other biophysical parameters such as sea (SST) or land surface temperatures (LST) estimated through thermal infrared data (see Julien *et al.,* 2006).

Of course, many approaches have been designed to overcome this problem (vanDijk *et al.,* 1987; Viovy *et al.,* 1992; Roerink *et al.,* 2000; Chen *et al.,* 2004; Beck *et al.,* 2006; Ma and Veroustraete, 2006; Hird and McDermid, 2009; Julien and Sobrino, 2010; Cho and Suh, 2013; Ke *et al.,* 2013; Lin *et al.,* 2014; Michishita *et al.,* 2014; Moreno *et al.,* 2014; Xiao *et al.,* 2015; Xu *et al.,* 2015; Yang *et al.,* 2015; Zhou *et al.,* 2015). These approaches, termed time series reconstruction –or gap-filling– methods, aim at providing alternative data for cloud-contaminated observations. In the case of vegetation, since it is assumed to change little over a few days (Holben, 1986), the time series reconstruction methods can offer an estimation of the actual value of the retrieved NDVI parameter below the cloud cover.

However, validation of these time series reconstruction methods is not straightforward: in the optical spectrum, no data can be retrieved from below the clouds by a satellite. Therefore, in the absence of ground truth, previous works have relied on indirect validation. First, studies have used a visual assessment of the improvement over pre-existing methods (vanDijk *et al.,* 1987; Viovy *et al.,* 1992; Roerink *et al.,* 2000; Chen *et al.,* 2004; Ma and Veroustraete, 2006; Beck *et al.,* 2006; Julien and Sobrino, 2010; Cho and Suh, 2013; Ke *et al.,* 2013; Lin *et al.,* 2014; Xiao *et al.,* 2015; Xu *et al.,* 2015; Yang *et al.,* 2015). More recently, authors have moved towards the simulation of cloud-free and cloud-contaminated time series (Hird and McDermid, 2009; Michishita *et al.,* 2014; Moreno *et al.,* 2014; Zhou *et al.,* 2015).

While the simulation of cloud-free time series is usually carried out through averaging several years of data (Hird and McDermid, 2009; Michishita *et al.,* 2014; Moreno *et al.,* 2014; Zhou *et al.,* 2015), the simulation of cloud-contaminated time series is more problematic. For example, some studies have introduced different levels of noise in the data (Michishita *et al.,* 2014; Moreno *et al.,* 2014), while realistic bi-monthly cloud frequencies have been introduced lately (Zhou *et al.,* 2015). On another hand, Hird and McDermid (2009) has tried to introduce realistic cloud-contaminated values in the simulated time series, by using NDVI values from actual cloudy observations.

Therefore, the main objective of this work is to provide a framework for the validation of time series reconstruction techniques, as well as for the intercomparison of different techniques. This framework consists mainly in the elaboration of a global dataset allowing for the simulation of

NDVI time series flexible enough to be used for most time series reconstruction techniques, and representative of both vegetation and cloud expected temporal behaviors. More specifically, this paper provides the following contributions:

- an explanation of how the dataset, termed TISSBERT for "Time Series Simulation for Benchmarking of Reconstruction Techniques", is built (Section 2),
- a complete description of the resulting dataset (Section 3),
- a description of how to use this dataset to generate both reference and test time series (Section 4),
- a discussion of the underlying assumptions made to build this dataset, and its differences with previous simulation techniques (Section 5).

2. Data and Methods

2.1. Data

We used the Long Term Data Record Version 4 (LTDR-V4) dataset (Pedelty *et al.,* 2007), which provides global daily acquisitions at 0.05° spatial resolution (roughly 5 km at the Equator). These data were acquired by the Advanced Very High Resolution Radiometer (AVHRR) instruments onboard the National Ocean and Atmospheric Administration (NOAA) satellite series. When starting this work, the time span for the available data was from July 1981 to December 2013. We therefore downloaded all these available data, although a few dates (22 in total) with obvious georeference errors had to be removed from the dataset.

We used channel 1 (0.6 µm, red) and 2 (0.8 µm, near infrared) data to retrieve daily NDVI (Rouse *et al.,* 1974; Tucker, 1979), and Quality Assessment (QA) information to build a land/sea mask, to identify valid observations, and to distinguish between cloudy and cloud-free data. NDVI values range from 0 to 1 for land areas, with values close to 0 corresponding to deserts and 1 to densely vegetated areas. Negative NDVI values correspond to water (lakes and oceans) and clouds. Since our purpose is to compare and/or validate reconstruction methods everywhere, our study area is the emerged surfaces of the Earth, excluding Antarctica and inland Greenland, where snow and ice prevent the growth of any vegetation.

To obtain relevant time series for the evaluation and comparison of reconstruction techniques, we need to simulate NDVI time series for both cloud-free and cloudy cases, and for all emerged areas. In order to test and validate different approaches for the elaboration of the TISSBERT dataset, we selected randomly 10 000 pixels belonging to different International Geosphere-Biosphere Programme (IGBP) land cover classes (black dots in Figure 1). Thus, a pixel by pixel retrieval of statistics was chosen.

2.2. Time window for statistic retrieval

A preliminary analysis (not presented here) of the data showed us that cloud-free observations during the more than 30 years of the LTDR-V4 time span were not enough to retrieve confidently statistics

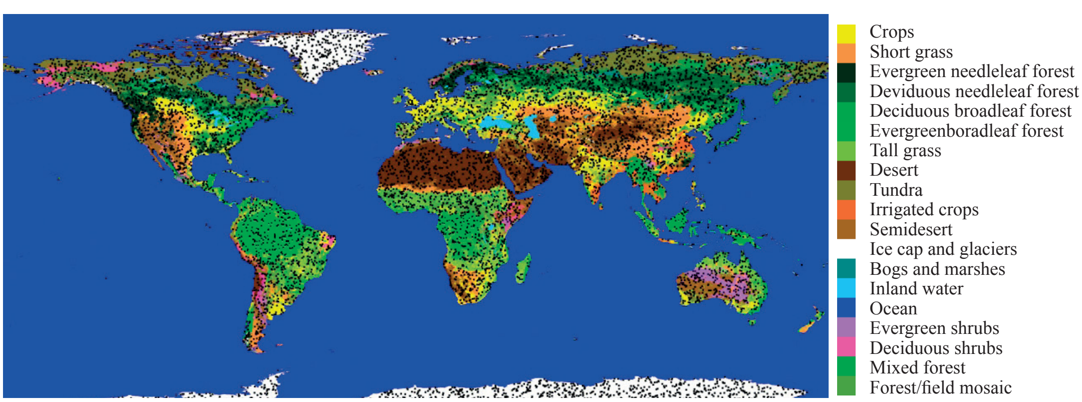

Figure 1. Localization of randomly selected pixels for preliminary analysis (black dots).

for a large portion of the emerged areas. However, the percentage of successful statistics retrieval became satisfactory for most areas when grouping daily observations from a same pixel within a 5-day or more time window. When increasing this time-window to values above 5 day, a small improvement in the percentage of successful statistical retrieval was observed, although only for pixels located on the edge of polar areas where data are lacking during wintertime. This small improvement came at the cost of a largely increased processing time, so we settled for a 5-day time window.

Therefore, all the TISSBERT statistics were retrieved pixel by pixel using all valid observations from the whole LTDRV4 dataset within a 5 days sliding time window over a given DOY (Day Of Year), i.e. from DOY-2 to DOY+2. For leap years, 29 February data were not considered in the analysis.

2.3. Statistical distribution

In order to model cloud-free and cloud contaminated NDVI distributions, we tested four different classic distributions: Gaussian (or normal), Cauchy, Log-normal, and Gamma. We thus tested for each random pixel and over each DOY 5-day window these four statistical distributions, for cloud contaminated NDVI data first, and then for cloud-free data. The parameters for each of these four statistical distributions were retrieved from the estimation of their respective moments (average, standard deviation, etc.), and agreement between modeled distributions and LTDR-V4 NDVI value distribution was tested using the Kolmogorov-Smirnov statistical test. This test returns values close to 0 to indicate a perfect agreement between compared distributions, while a value of 1 corresponds to a perfect disagreement. In opposition with normality tests for example, this test is not a test to determine the adequate distribution to represent the data. We finally assessed the suitability of distributions by comparing their Kolmogorov-Smirnov score. Additionally, we tested if modeled NDVI values for a given DOY were in agreement with LTDR-V4 observed values by estimating a range agreement statistic, with values ranging from 0 (all modeled values fall outside the range of LTDR-V4 observed values) to 1 (all modeled values fall within the range of LTDR-V4 observed values).

However, for the Gaussian model of cloud-free data, a first retrieval of mean and standard deviation values led to unstable statistics, with large oscillations through the year, due mainly to the presence of outliers. Therefore, we also estimated Gaussian mean and standard deviation for cloud-free observations, by assuming a discrete normal distribution of the data. In that Discrete Gaussian model, the mean value was set to the median value of the pixel data for all available years over a time window of 5 days. As regards the standard deviation value, we ordered all valid values, for each pixel and 5-day time window, and we identified the two NDVI values corresponding to 34.1% of the valid NDVI values around the median value (i.e. the values corresponding to 15.9% and 84.1% of the total of ordered NDVI values). The standard deviation value was then set to the smallest of the difference between these two values and the median value. This approach allowed to increase the stability of the retrieved statistics through the year, so this method was chosen for the retrieval of cloud-free statistics.

When no cloud-free observations were available over the 5-day time window, cloud-free statistics were estimated from the 5 highest cloud-contaminated NDVI values. Since the effect of clouds on NDVI is to decrease the NDVI value, the highest observed NDVI values correspond to the observations with the lowest amount of clouds, and may provide useful information for the time series. This was needed for a few areas (highest mountain ranges) where the cloud mask might have been overconservative. Finally, we also estimated the pixel by pixel probability of cloud presence as the percentage of cloudy observations among valid data for the considered 5-day time window.

By applying the best distribution model to the data, we obtain 5 different images of 7200 by 3600 pixels, for each DOY, corresponding to cloud presence probability (CloudProba), cloudy mean NDVI value (CloudAvge), cloudy standard deviation NDVI value (CloudStdv), cloud-free mean NDVI value (ClearAvge), and cloud-free standard deviation NDVI value (ClearStdv). These data are available from the authors upon request.

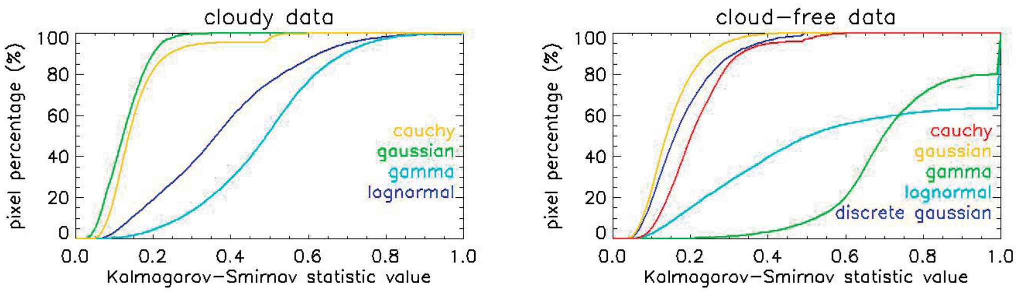

Figure 2. Histogram of Kolmogorov-Smirnov test values for all pixels of Figure 1 for different statistical distribution of NDVI values.

3. Results

3.1. Best distribution

Figure 2 shows the results of the Kolmogorov-Smirnov statistics for all tested distributions. This figure shows that the best model for cloudy NDVI value distribution is the Gaussian model, with an average Kolmogorov-Smirnov value of 0.17, closely followed by the Cauchy model, while the Log-normal and Gamma models perform poorly. In the case of cloud-free data, the previous models behave similarly, while the discrete Gaussian model ranks second, close to the best model (Gaussian), with average Kolmogorov-Smirnov values of 0.21 and 0.19 respectively.

Figure 3 presents the results obtained for the range agreement statistic for the two best performing models in both cloudy and cloud-free cases. As regards the case of cloudy values, the Gaussian model still performs best, with an average range agreement value above 0.9. As for cloud-free values, the Gaussian model performs worse, while the Discrete Gaussian model reaches a mean range agreement value about 0.98. Since range agreement values were clearly higher in the case of the Discrete Gaussian model, it was selected to model cloud-free TISSBERT data. Moreover, we observed that such distribution model led to lower and more stable (less noisy) standard deviation values through the year. As for cloudy observations, the Gaussian model clearly performed best, and therefore was used for the TISSBERT dataset.

Land cover influence on best model selection was also analyzed, although the best performing model was the same (Gaussian) for all IGBP classes, with only small variations of the Kolmogorov-Smirnov statistic. Finally, we analyzed if the selection of the best model was time-dependent. To that end, we selected for each random pixel and each DOY the best performing model for NDVI value distribution, and determined the percentage of pixels for which each method corresponded to the best model (Figure 4). This shows that the Gamma distribution performs poorly for all DOYs, while the Gaussian model performs best for all DOYs and is quite stable throughout the year. The Log-normal model tends to perform better in the period corresponding to June to August, increasing its pixel percentage in detriment of the Cauchy model.

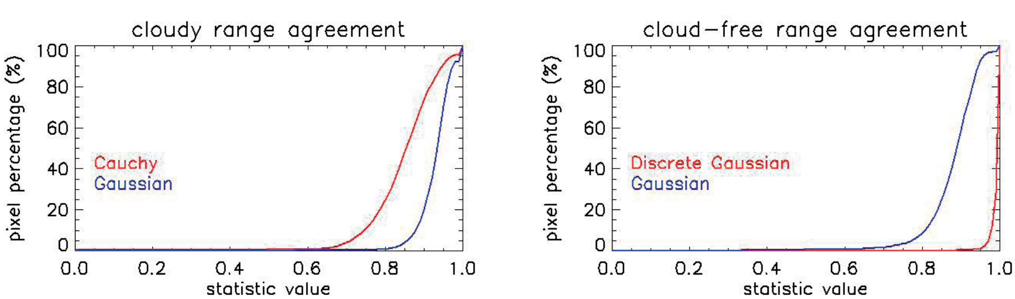

Figure 3. Histogram of range agreement values for all pixels of Figure 1 for different statistical distribution of NDVI values.

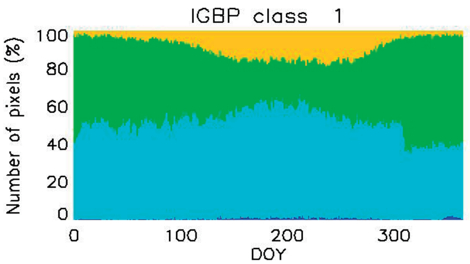

Figure 4. Percentage of pixels for which the best statistical model for cloudy NDVI values is Gamma (dark blue), Gaussian (light blue), Cauchy (green) and Log-normal (yellow) through the year.

3.2. Examples of TISSBERT data

Figure 5 presents TISSBERT parameters (cloudProba, clearAvge, clearStdv, cloudAvge, cloudStdv) values for DOYs 1 (1st of January), 91 (1st of April), 182 (1st of July), 274 (1st of October). Summer NDVI values for temperate areas show the expected lower cloud probability and higher cloud-free and cloud-contaminated NDVI average values, with lower cloud-free and higher cloud-contaminated standard deviations. In deserts and in the tropics, all parameters are roughly stable throughout the year, with small NDVI standard deviation values in all cases, and low and high NDVI average values for deserts and tropical forests respectively. South-East Asia, with its monsoonal regime, shows a higher probability of clouds for DOY 182, while the other parameters show comparatively little change throughout the year. Notably, polar areas show no available data during winter. Cloud-free standard deviations show higher values for the Northern temperate areas in winter, due to the lower number of available cloud-free observations.

4. Using TISSBERT data

From the TISSBERT dataset, we can generate synthetic cloud-free and cloudy time series to validate and compare NDVI time series reconstruction methods. Cloud-free time series can be generated by using the ClearAvge harmonic formulation only, while cloud-only time series can be generated by using a gaussian distributed random function with average CloudAvge, and standard deviation CloudStdv. A cloud occurence time series can be generated from the CloudProba data, and used to combine the cloud-free and cloud-only time series into a cloud contaminated time series. For authors interested in assessing the influence of NDVI retrieval error on their methods, the cloud-free time series should be generated by using a gaussian distributed random function with average ClearAvge, and standard deviation ClearStdv.

Therefore:

$$NDVI_{clear}(t) = ClearAvge(t) \quad (1)$$

$$NDVI_{clear}^{noisy}(t) = ClearAvge(t) + randN_1(t) \times ClearStdv(t) \quad (2)$$

$$NDVI_{cloud}(t) = ClearAvge(t) + randN_2(t) \times ClearStdv(t) \quad (3)$$

where randN1(t) and randN2(t) are two independent gaussian random vectors of the desired length.

Finally, we can simulate our reference (ref) and cloud-contaminated (test) time series as follows:

$$NDVI_{ref}(t) = NDVI_{clear}(t) \quad (4)$$

$$NDVI_{test}(t) = NDVI_{clear}(t) \times (randU(t) > CloudProba(t)) + \\ + NDVI_{cloud}(t) \times (randU(t) < CloudProba(t)) \quad (5)$$

where randU(t) is a random uniform vector of the desired length.

When interested in assessing also the influence of the noise in NDVI retrieval value, the reference and cloud-contaminated time series should be simulated as follows:

$$NDVI_{ref}^{noisy}(t) = NDVI_{clear}^{noisy}(t) \quad (6)$$

$$NDVI_{test}^{noisy}(t) = NDVI_{clear}^{noisy}(t) \times (randU(t) > CloudProba(t)) + \\ + NDVI_{cloud}(t) \times (randU(t) < CloudProba(t)) \quad (7)$$

To determine the needed extension of the time series to retrieve reliable statistics, we estimated the standard deviation between the reference NDVI and the noise contaminated reference NDVI for a given number of years (from 1 to 20). We then computed the percentage of pixels for which the relative change in standard deviation between a given year and the preceding was below 1%. This 1% figure corresponds to the maximum acceptable standard deviation change from a given period and a 1-year larger period. As a compromise between stable statistics and reasonable computation time, we conclude that a time series of 15 simulated years from the TISSBERT dataset is well-suited to study time reconstruction method performance.

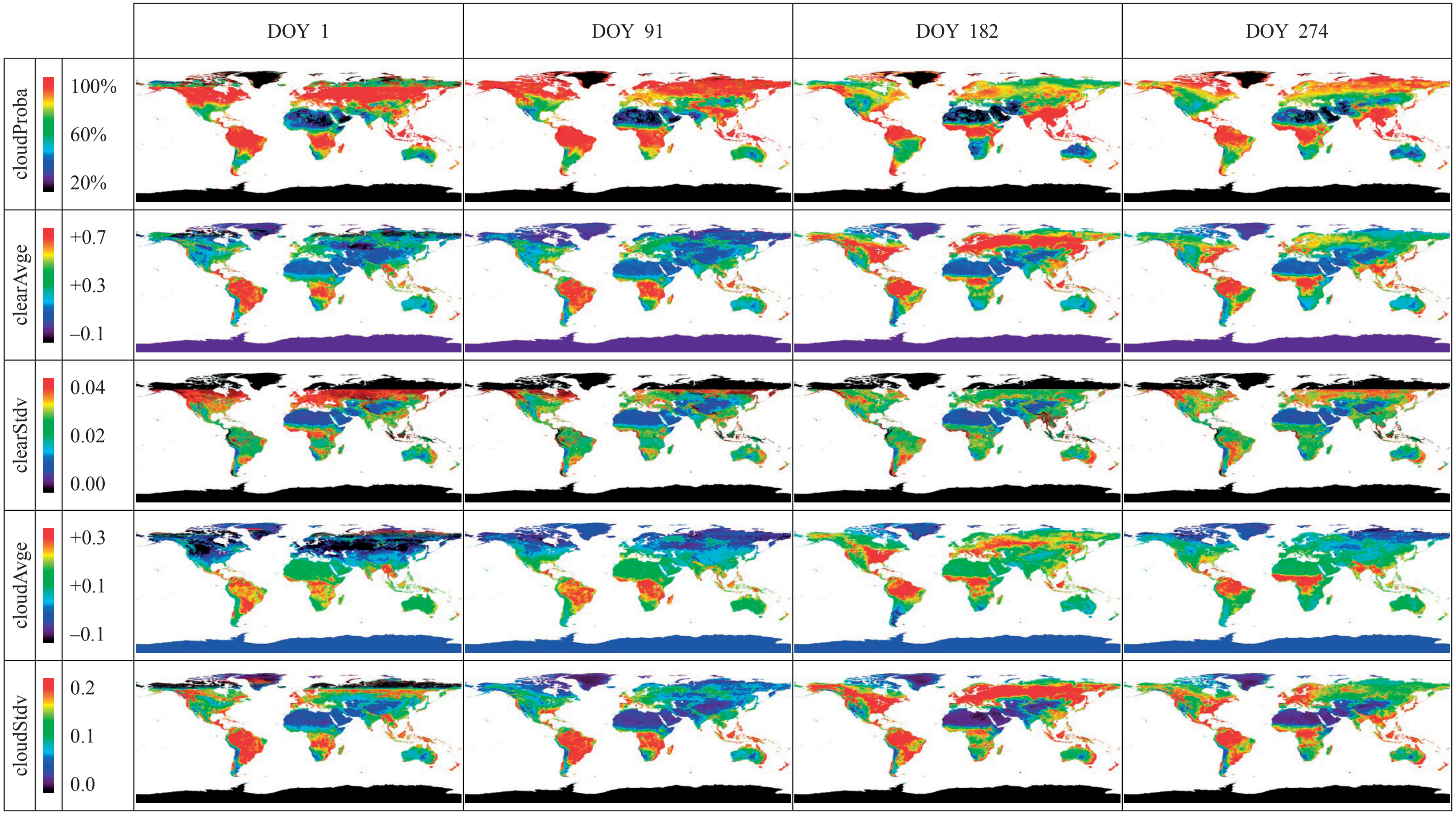

Figure 5. TISSBERT parameters (cloudProba, clearAvge, clearStdv, cloudAvge, cloudStdv) values for DOYs 1 (1st of January), 91 (1st of April), 182 (1st of July), 274 (1st of October). All values are in NDVI units, except for cloudProba parameter, in percentage.

Figure 6 illustrates the retrieval of TISSBERT parameters for a deciduous forest pixel in Northern Canada (42.725°N, 77.525°W). In this figure, blue crosses correspond to observations labeled as cloudy in the LTDR-V4 dataset, while orange crosses correspond to cloud-free observations. The dark blue continuous line indicates cloud contaminated average DOY NDVI (CloudAvge), while the dotted blue line indicates one standard deviation around this average (CloudStdv). Similarly, red continuous and dotted lines indicate average and standard deviation DOY NDVI for cloud-free data (respectively ClearAvge and ClearStdv). The black dotted line indicates the probability of cloud (CloudProba), while the black crosses indicate the simulated TISSBERT data following Equation 7 ($NDVI_{test}^{noisy}$). Figure 6 shows higher standard deviations in summer for cloudy observations, while standard deviations for cloud-free observations decreases in this period, as commented for Figure 5. The probability of retrieving cloudy observations also diminishes in summer, while it is close to 1 in winter time. The simulated TISSBERT time series following Equation 7 mirrors the behavior of NDVI values for this pixel with almost no cloud-free observations during winter, and a higher amplitude variation for the rest of the year.

5. Discussion

Here, we first discuss the quality of the TISSBERT dataset, then its underlying assumptions, and finally compare this dataset to the methods used in the literature to validate time series reconstruction techniques.

5.1. Quality of the dataset

To assess the quality of the TISSBERT dataset, we retrieved for each pixel the percentage of DOYs for which not enough data were available for the retrieval of TISSBERT parameters (Figure 7). This retrieval was more problematic for polar areas in both cloud-free and cloudy cases, due to the absence of data during winter. It was also problematic for a few desert areas (eastern Sahara and central Arabic Peninsula) in the case of cloudy NDVI values, and mountainous and tropical areas in the case of cloud-free NDVI values. However, these problems are directly inherited from the LTDR-V4 dataset, and alternative data sources

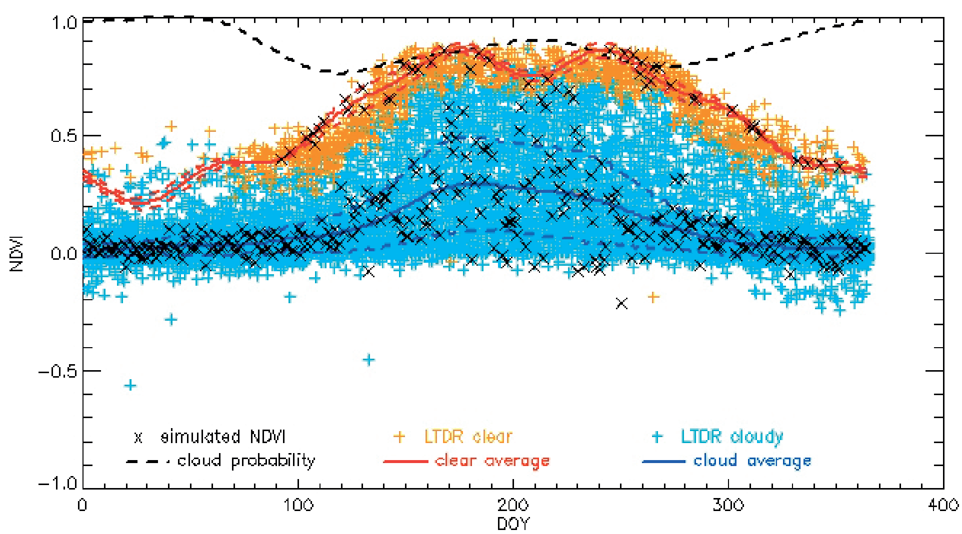

Figure 6. Retrieval of TISSBERT parameters for a deciduous forest pixel in Northern Canada (42.725°N, 77.525°W). Blue crosses correspond to observations labeled as cloudy in the LTDR-V4 dataset, while orange crosses correspond to cloud-free observations. The dark blue continuous line indicates cloud contaminated average DOY NDVI, while the dotted blue line indicates one standard deviation around this average. Similarly, red continuous and dotted lines indicate average and standard deviation DOY NDVI for cloud-free data. The black dotted line indicates the probability of cloud, while the black crosses indicate the simulated TISSBERT data following Equation 7.

(see below) present similar flaws, resulting directly of local climate characteristics. Therefore, the TISSBERT dataset can be considered as representative of LTDR-V4 data for most regions of the globe, although retrieved statistics are less reliable in tropical and some desert areas. During winter time, the LTDR-V4 dataset has no valid NDVI data for all polar areas (Scandinavia, Northern Canada and Russia, Southern South America), since shorter sun hours there during this period do not allow for the retrieval of visible and near infrared information. As a result, in polar areas, the estimation of TISSBERT statistics in wintertime is flawed, and may result in unreliable results for this period. However, in the absence of an alternative data source, there is no way this can be improved. Finally, due to the highest noise in the case of the ClearStdv parameter, we do not recommend its use (see below).

5.2. Underlying assumptions

The choice of a 5-day time window for the estimation of pixel DOY statistics was made as a compromise between the number of available cloud-free observations and the expected stability of the observed vegetation during this period. This 5-day time window is shorter than the period usually set for compositing algorithms: 10-day for Vegetation data (see for example Swinnen and Veroustraete, 2008), or 15-day for Global Inventory Modeling and Mapping Studies (GIMMS) data (Tucker *et al.,* 2005).

By summarizing the LTDR-V4 dataset into DOY statistics, we implicitly assume that the vegetation has been stable during more than 30 years. This is clearly a strong assumption, since vegetation has been shown to have experienced changes in many studies, through both land use (Sobrino and Julien, 2011) and phenology changes (Julien and Sobrino, 2009). However, due to the coarse spatial resolution of the dataset, this assumption holds for most pixels. Moreover, the approach chosen here (median value and discrete estimation of standard deviation) allows to decrease the influence of outliers, and therefore the resulting statistics correspond preferably to stable states of the vegetation.

Additionally, NDVI interannual variability, due to variations in phenology for example, has not been considered here. However, this artificially increases DOY standard deviation for cloud-free observations, especially during vegetation green-up and senescence. However, as stated before, we do not recommend the use of this parameter for studying the influence of noise in NDVI clear-sky value estimation, which can perfectly be studied by using a classical white noise approach.

In this paper, we have also assumed a perfect cloud-masking of the LTDRV4 data. Due to the small number of channels of the AVHRR instrument, cloud detection is more difficult than with more recent instruments. However, the choice of the median and discrete standard deviation estimation of normal distribution parameters allows to screen out the eventual mislabeling of the cloud algorithm, although this could modify slightly the values of all retrieved statistics.

By estimating cloud-free average NDVI value from the highest 95% of cloudy labeled observations when no cloud-free observations are available, biases can appear in areas where clouds are persistent for long periods, as well as a higher instability in the temporal profile. This assumption is based on the fact that some labeled

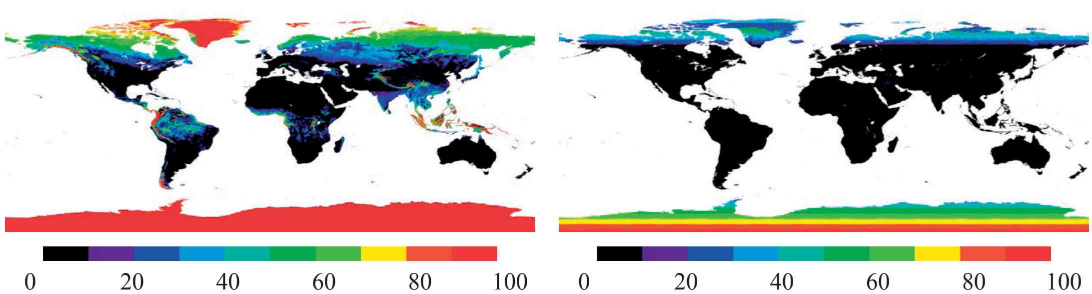

Figure 7. Percentage of DOYs for which the estimation of TISSBERT parameters was unsuccessful in the cases of cloud-free (left) and cloudy (right) data due to the lack of valid observations during a given 5-day time window.

cloudy observations still have high NDVI values, and that some useful information can be extracted from these cloudy labeled observations.

Another drawback of the LTDR-V4 dataset is the influence of the orbital drift effect on the data. We have shown in another study (Sobrino and Julien, 2016) that although this effect exists, it has a low influence on the retrieved NDVI trends. Moreover, and once more, the chosen approach for the retrieval of the normal distribution parameters diminishes the influence of the orbital drift effect, by not taking into account the more extreme values.

Due to the spatial resolution of the LTDR-V4 data (0.05°, roughly 5 km at the Equator), the TISSBERT dataset may not be optimal for the characterization of time series reconstruction methods in the case of high resolution data (100 m or lower). This is due to the spatial averaging of heterogeneous land covers, which results in a decrease of the NDVI amplitude for given land covers (see for example Munyati and Mboweni, 2012). For example, NDVI values of 0.8 are almost inexistent in the TISSBERT data, while these can be found easily in Landsat images (spatial resolution of 30 m) of cultivated areas, as a consequence of differences in both spatial resolution and spectral characteristics.

5.3. Differences with previous methods

As mentioned in the introduction section, this paper is not the first one to provide a method to compare NDVI time series reconstruction methods nor to try and estimate its accuracy. As a matter of fact, recent papers (Hird and McDermid, 2009; Michishita *et al.,* 2014; Moreno *et al.,* 2014; Zhou *et al.,* 2015) have all introduced simulations to provide an accuracy assessment of the performance of the presented methods, as well as to compare them with previous ones.

Simulation of cloud-free data is generally carried out through the averaging of real data over several years (Hird and McDermid, 2009; Michishita *et al.,* 2014; Moreno *et al.,* 2014; Zhou *et al.,* 2015), since this provides a realistic shape of the annual vegetation signal. However, it can be conducted using all available acquisitions, or only the acquisitions flagged as cloud-free. Even though the latter case will provide a better reference time series, the effect of undetected clouds may be significant, especially in areas with regular cloud cover. The use of the median approach in the TISSBERT dataset allows to lower this effect. On the other hand, an ensemble approach has been used by Geng *et al.* (2014) to build a reference time series, which uses the average of the reconstructed time series with different methods as a reference time series. However, this approach biases the estimation of the error in the reconstruction, since some methods use an upper envelope approach to detect cloud-contaminated data, and will therefore always provide a reconstructed time series above the reference one.

As for the simulation of cloud-contaminated data, we also find three different approaches: noise introduction, random gap introduction, and time-realistic introduction of gaps. Noise introduction resides in the addition to the reference time series of random noise of different amplitudes to the reference time series (Michishita *et al.,* 2014; Moreno *et al.,* 2014). Since cloud-contamination tends to decrease NDVI values, this approach, although valid to indicate the method sensitivity to noise, is inadequate to estimate method errors, since roughly half the generated cloud-contaminated values are above the values of the reference time series. As for the random gap introduction used by Hird and McDermid (2009), it included realistic cloud-contaminated NDVI values, while the temporal distribution of cloud contaminated values remained far from realistic. Finally, a more time-realistic introduction of gaps has been provided recently by Zhou *et al.* (2015), through the determination at global scale of a gap proportion in the data, taking into account the timing and possibility of persisting cloud cover. However, this approach filled the simulated cloud-contaminated pixels with an invalid value, which makes it easy for the reconstruction method to identify clouds, and was restricted to bimonthly time series. Nonetheless, this approach (Zhou *et al.,* 2015) is, with this study, the only approach which attempted to provide a temporally realistic distribution of cloud-contaminated data.

Of course, for methods based on the knowledge of time series cloud flags, the actual value of the cloud contaminated NDVI is irrelevant, since only the cloud-free data are considered for the reconstruction. Nonetheless, mislabeling of the acquisition can always occur, and both omission and commission errors may influence the reconstruction of the time series.

Recently, spatial-temporal methods for time series reconstruction have been developed (Poggio *et al.,* 2012; Cho and Suh, 2013; Lin *et al.,* 2014; Weiss *et al.,* 2014; Xu *et al.,* 2015). Although the TISSBERT dataset only includes temporal information, it could be extended to provide spatial information in combination to the temporal information described above, to allow for the validation and intercomparison of such methods.

6. Conclusions and perspectives

We have introduced here the TISSBERT dataset to the scientific community, through a thorough description of its elaboration, as well as an explanation of how to use it. Due to its daily temporal resolution, the TISSBERT dataset also allows for the validation and benchmarking of reconstruction techniques whatever the temporal resolution of the planned application, by random or composite selection of the data within the chosen time window.

As the temporal length of available NDVI data increases, alternative daily datasets from more recent instruments (MODIS) may become available. In such case, a rerun of the TISSBERT approach could be carried out to improve the quality of the dataset.

Apart from its main purpose of time series reconstruction benchmarking, the TISSBERT dataset allows for a range of applications in real-time monitoring of NDVI data. For example, it could be used as a complementary tool to detect cloud contamination of the observed data, by testing if the observed NDVI value falls within the expected distribution of values at a given level of confidence. It could also be used to detect automatically deviations of newly acquired data from TISSBERT data, and therefore labeling the corresponding area as possible land cover change. Another simple application of this dataset is the a priori estimation of the probability of cloud-contamination for a given location and DOY.

Since the compilation of the TISSBERT dataset is time consuming, and because the aim of this paper is to provide a framework for the comparison of time series reconstruction techniques, this dataset is available upon request to the authors. A global comparison of the performances of most widely used time series reconstruction techniques will be carried out in a near future.

Acknowledgement

This work was supported by the Spanish Ministerio de Economía y Competitividad (CEOS-SPAIN2, project ESP2014-52955-R and SIM, project PCIN-2015-232). The authors also thank NASA for the free access to the LTDRV4 data.

Referencias

Beck, P., Atzberger, C., Hogda, K.A., Johansen, B. Skidmore A. 2006. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sensing of Environment*, 100, 321-334. https://doi.org/10.1016/j.rse.2005.10.021

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L. 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sensing of Environment*, 91, 332-334. https://doi.org/10.1016/j.rse.2004.03.014

Cho, AR., Suh, M.S. 2013. Detection of contaminated pixels based on the short-term continuity of NDVI and correction using spatio-temporal continuity. *Asia-Pacific Journal of Atmospheric Sciences*, 49(4), 511-525. https://doi.org/10.1007/s13143-013-0045-7

Geng, L., Ma, M., Wang, X., Yu, W., Jia, S. and Wang, H. 2014. Comparison of eight techniques for reconstructing multi-satellite sensor time-series NDVI data sets in the Heihe river basin, China. *Remote Sensing*, 2014, 6, 2024-2049

Hird, J.N., McDermid, G.J. 2009. Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*, 113, 248-258. https://doi.org/10.3390/rs6032024

Holben, B.N. 1986. Characteristics of maximum-value composite image from temporal AVHRR data. *International Journal of Remote Sensing*, 7, 1417-1434. https://doi.org/10.1080/01431168608948945

Jönsson, P., Eklundh, L. 2004. TIMESAT - A program for analyzing time-series of satellite sensor data. *Computers and Geoscience*, 30, 833-845. https://doi.org/10.1016/j.cageo.2004.05.006

Julien, Y., Sobrino, J.A. 2009. Global land surface phenology trends from GIMMS database. *International Journal of Remote Sensing*, 30(13), 3495-3513. https://doi.org/10.1080/01431160802562255

Julien, Y., Sobrino, J.A. 2010. Comparison of cloud-reconstruction methods for time series of composite NDVI data. *Remote Sensing of Environment*, 114, 618-625. https://doi.org/10.1016/j.rse.2009.11.001

Julien, Y., Sobrino, J.A. 2012. Correcting Long Term Data Record V3 estimated LST from orbital drift effects. *Remote Sensing of Environment*, 123, 207-219. https://doi.org/10.1016/j.rse.2012.03.016

Julien, Y., Sobrino, J.A., Verhoef, W. 2006. Changes in land surface temperatures and NDVI values over Europe between 1982 and 1999. *Remote Sensing of Environment*, 103, 43-55. https://doi.org/10.1016/j.rse.2006.03.011

Ke, L., Ding, X., Song, C. 2013. Reconstruction of time series MODIS LST in central Qinghai-Tibet plateau using geostatistical approach. *IEEE Geoscience and Remote Sensing Letters*, 10(6), 1602-1606. https://doi.org/10.1109/LGRS.2013.2263553

Lin, C.H., Lai, K.H., Chen, Z.B., Chen, J.Y. 2014. Patch-based information reconstruction of cloud-contaminated multitemporal images. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1), 163-174. https://doi.org/10.1109/TGRS.2012.2237408

Ma, M., Veroustraete, F. 2006. Reconstructing pathfinder AVHRR land NDVI timeseries data for the Northwest of China. *Advances in Space Research*, 37, 835-840. https://doi.org/10.1016/j.asr.2005.08.037

Michishita, R., Jin, Z., Chen, J., Xu, B. 2014. Empirical comparison of noise reduction techniques for NDVI time-series based on a new measure. *ISPRS Journal of Photogrammetry and Remote Sensing*, 91, 17-28. https://doi.org/10.1016/j.isprsjprs.2014.01.003

Moreno, A., García-Haro, F.J., Martínez, B., Gilabert, M.A. 2014. Noise Reduction and Gap Filling of fAPAR Time Series Using an Adapted Local Regression Filter. *Remote Sensing*, 6, 8238-8260. https://doi.org/10.3390/rs6098238

Munyati, C., Mboweni, G. 2012. Variation in NDVI values with change in spatial resolution for semi-arid savanna vegetation: a case study in northwestern South Africa. *International Journal of Remote Sensing*, 34(7), 2253-2267. https://doi.org/10.1080/01431161.2012.743692

Pedelty, J., Devadiga, S., Masuoka, E., Brown, M., Pinzon, J., Tucker, C., *et al.* 2007. Generating a long-term land data record from the AVHRR and MODIS instruments. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2007, 1021-1025, https://doi.org/10.1109/IGARSS.2007.4422974

Poggio, L., Gimona, A., Brown, I. 2012. Spatio-temporal MODIS EVI gap filling under cloud cover: an example in Scotland. *ISPRS Journal of Photogrammetry and Remote Sensing*, 72, 56-72. https://doi.org/10.1016/j.isprsjprs.2012.06.003

Roerink, G.J., Menenti, M., Verhoef, W. 2000. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing*, 21(9), 1911-1917. https://doi.org/10.1080/014311600209814

Rouse, J.W., Haas, R.H., Scheel, J.A., Deering, D.W. 1974. Monitoring Vegetation Systems in the Great Plains with ERTS. *3rd Earth Resource Technology Satellite (ERTS) Symposium Proceedings*, Vol. 1, 48-62.

Sobrino, J.A. Julien, Y. 2011. Global trends in NDVI derived parameters obtained from GIMMS data. *International Journal of Remote Sensing*, 32(15), 4267-4279. https://doi.org/10.1080/01431161.2010.486414

Sobrino, J.A., Julien, Y. 2016. Exploring the validity of the Long Term Data Record V4 database for land surface monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 99, 1-8, https://doi.org/10.1109/JSTARS.2016.2567642

Swinnen, E., Veroustraete, F. 2008. Extending the SPOT-VEGETATION time series (1998-2006) back in time with NOAA-AVHRR data (1985-1998) for Southern Africa. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2), 558-572. https://doi.org/10.1109/TGRS.2007.909948

Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127-150. https://doi.org/10.1016/0034-4257(79)90013-0

Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A. Pak, E.W., Mahoney, R., Vermote, E.F., El Saleous, N. 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, 26(20), 4485-4498. https://doi.org/10.1080/01431160500168686 van Dijk, A., Callis, S., Sakamoto, C. and Decker, W. 1987. Smoothing vegetation index profiles: An alternative method for reducing radiometric disturbance in NOAA/AVHRR data. *Photogrammetric Engineering and Remote Sensing*, 53, 1059-1067.

Viovy, N., Arino, O., Velward, A. 1992. The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series *International Journal of Remote Sensing*, 13, 1585-1590. https://doi.org/10.1080/01431169208904212

Weiss, D.J., Atkinson, P.M., Bhatt, S., Mappin, B., Hay, S.I., Gething, P.W. 2014. An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 98, 106-118. https://doi.org/10.1016/j.isprsjprs.2014.10.001

White, M.A., De Beurs, K.M., Didan, K., Inouye, D. W., Richardson, A.D., *et al.* 2009. Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. *Global Change Biology*, 15, 2335-2359. https://doi.org/10.1111/j.1365-2486.2009.01910.x

Xiao, Z., Liang, S., Wang, T., Liu, Q. 2015. Reconstruction of satellite-retrieved land-surface reflectance based on temporally-continuous vegetation indices. *Remote Sensing*, 7, 9844-9864. https://doi.org/10.3390/rs70809844

Xu, L., Li, B., Yuan, Y., Gao, X., Zhang, T. 2015. A temporal-spatial iteration method to reconstruct NDVI time series datasets. *Remote Sensing*, 7, 8906-8924. https://doi.org/10.3390/rs70708906

Yang, G., Shen, H., Zhang, L., He, Z. and Li, X. 2015. A moving weighted harmonic analysis method for reconstructing high-quality SPOT VEGETATION NDVI time-series data. *IEEE Transactions on Geoscience and Remote Sensing*, 53(11), 6008-6021. https://doi.org/10.1109/TGRS.2015.2431315

Zhou, J., Jia, L. and Menenti, M. 2015. Reconstruction of global MODIS NDVI time series: Performance of Harmonic ANalysis of Time Series (HANTS). *Remote Sensing of Environment*, 163, 217-228. https://doi.org/10.1016/j.rse.2015.03.018