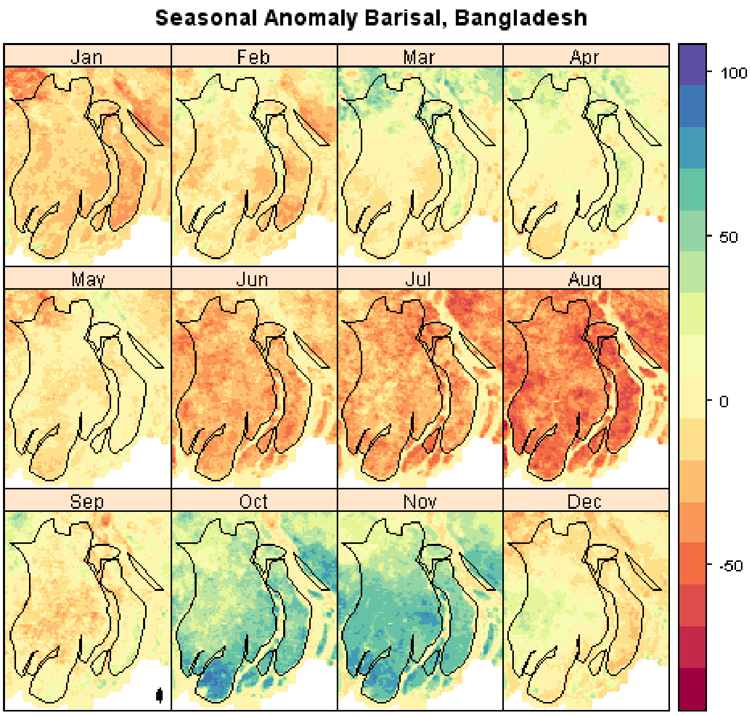
Time Series Analysis of Vegetation Indices (NDVI) with R

Use of `ndvits` package : from a global to a local scale analysis.

DRAFT - not yet for distribution



Example - Tutorials

Romain Frelat, Bruno Gérard
August 8, 2011

Abstract

The objective of this document is to present the `ndvits` package. NDVI Time Series extraction and analysis project use the R plateform to disseminate automated methods to analyse NDVI data. The routines extract automatically NDVI time series from different satellites (AVHRR, Spot Vegetation, MODIS) and provide tools to display and analyze them. Phenological metrics can be computed for every season as well as vegetation anomaly maps.

First, we present the dataset and the advantage of remote sensing data. Algorithms to produce anomaly maps and calculate metrics are explained in this section.

Second, we work on a practical example : a multi scale analysis accross the study sites of the CGIAR's Systemwide Livestock Programme.

Contents

Introduction

This document presents the R package `ndvits` and how to use it. It has been developed under the CGIAR Systemwide Livestock Programme (SLP) project : *Optimizing livelihood and environmental benefits from crop residues in smallholders crop-livestock systems in sub-Saharan Africa and South Asia: regional case studies.*

The main goal of this project is to understand the trade-offs of crop residue uses in smallholders crop-livestock systems. A large survey is addressed to small-scale farmers in order to understand the actual systems and conceive ways to conduct a sustainable intensification. Data from West, East and Southern Africa as well as from South Asia are collected thanks to a cooperation of CGIAR research centers.

In this context, adding an information about vegetation from remote sensing is very interesting to get an overview of the history of vegetation dynamics. Remote sensing (RS) is a source of temporal and spatial information which are precious to understand the processes, dynamics and disturbances of ecosystems. The Normalized Difference Vegetation Index (NDVI) is the most widely used vegetation index and is freely and easily available on internet.

To get the best of these data, an open-source tool to analyze NDVI data has been developed. We choose to use the platform R for its flexibility and its already existing libraries for statistics as well as spatial and time series analysis.
Comparable to `Timesat` [13], this R package is more flexible and allow you to design and conduct your own analysis.

In this document, we (1) presente the methods used for a NDVI analysis; and (2) comment examples of applications of the package `ndvits`. The analysis considers NDVI time series at differend scales. First, we are comparing time series from a global perspective, then we focus more on countries and extract annual phenological metrics (regional scale) and we finish with mapping anomalies on a restricted area (local scale). In appendix, more practical problems are discussed.

This document assumes that you already had an exposure to R . For a better display, all the R-script are presented in blue and console output are in red in the text.

1 Materials and methods

1.1 Remote sensing and vegetation index

1.1.1 Normalized Difference Vegetation Index (NDVI)

Remote sensing is nowadays a key tool to characterize and study agricultural landscapes. Its main advantages are its capacity to cover large area - and particularly remote area, difficult to access - as well as its relatively low cost in term of money and time.

According to Vincikova et al. (2010) [20], NDVI is the most widely used index to study vegetation. It is based on vegetation reflectance. Green plants has a very low reflectance in the visible and mainly in the blue and the red part of the spectrum ($0.43 - 0.66\mu m$). On the contrary, vegetation has high reflectance in Near Infra Red (NIR : $0.7 - 1.1\mu m$) with a "red edge" to distinguish healthy plant from dry one (see Figure 1).

NDVI is calculated with the formula first described by Tucker C. in 1979 [15]:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

The values of NDVI vary between -1 and 1. Low values around 0 are signal of bare soil (low or no vegetation) and values of 0.7 or larger for dense vegetation, forest. It is a good representation of vegetation growth and vigor at the land surface.

NDVI datasets are freely available on internet. On the following, we decided to focus only on two of them, coming from two different satelitte : the Advanced Very-High-Resolution Radiometer and SPOT Vegetation.

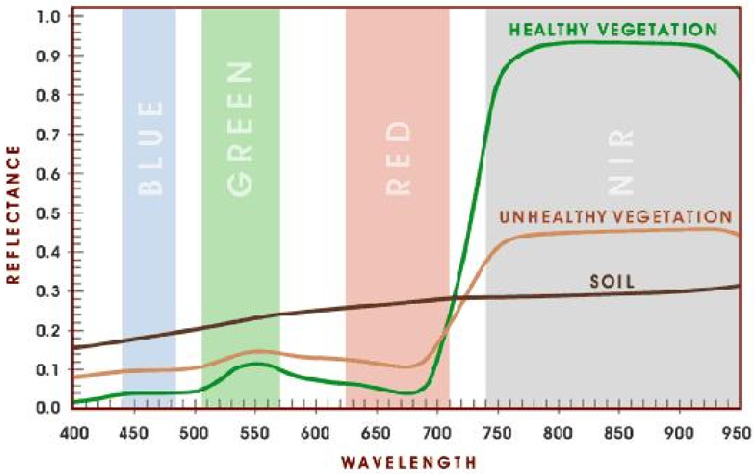


Figure 1: Reflectance spectra of vegetation and soil (source Utah State University[1]).

1.1.2 The Global Inventory Modeling and Mapping Studies (GIMMS)

Global Inventory Modeling and Mapping Studies (GIMMS)[16] data come from the National Oceanic and Atmospheric Administration Advanced Very-High-Resolution Radiometer (NOAA-AVHRR). Maps are freely available at `http://www.landcover.org/data/gimms/`.

[1] `http://extension.usu.edu/nasa/htm/on-target/near-infrared-tutorial/`, accessed 25/07/2011

Maps are projected on a conic projection of Albers Equal Area, ellipsoid Clarke 1866 with a WGS84 datum. The spatial resolution is 8km. NDVI values are coded with 16 Bits with a multiplication factor of 10000 (values are between -10000 and 10000). Maps are issued bimonthly using the Maximum Value Composite over a 15-day period. That's mean, for each pixel, the map keeps only the maximum value over the 15 days.

The main advantage of this dataset is its long term data availability, from July 1981 through December 2006. Despite its relatively low spatial resolution (one pixel represents $8km^2$), GIMMS datasets provide the longest time series available and it is the only one able to capture trends at decadale scale. That's why, it is often preferred in studies which are focus on the history of the vegetation and its trend - especially climate change studies.

However, according to Vancutsem et al. 2008 [17] and comparing to new high resolution satellite systems, AVHRR have a poor geometric accuracy and limited radiometric calibrations that introduces spatial and temporal inconsistencies in the time series.
Moreover, few periods are subject to low consistency values : NDVI signal decline in mid-1991, due to the Pinatubo eruption, during the second half of 1994 and during most of 2000 , an unusually early or late overpass times are observed. Data from these periods should be used with caution.

1.1.3 SPOT Vegetation (VGT)

Spot Vegetation data are freely distributed by VITO at `http://free.vgt.vito.be/` [1].

Maps have a spatial resolution of 1 km and are in a geographic projection with a WGS84 datum. NDVI value are coded in 8 Bit with numbers from 0 to 255. Data are available from April 1998 and continues to date The data considered is "S10-composited" : a 10-day maximum value composite.ata is $1km^2$. The 10 day maximum value composite (MVC) is used to minimize the noise and especially the problem of clouds.

Although the spatial and temporal resolution is better than GIMMS, time series are for the moment hardly longer than a decade. VGT dataset, together with new high resolution satellites, will be prefered for precision works such as land cover classification or prediction studies.

Spot vegetation is known to have a high spectral and spatial consistency. The NDVI images are radiometrically calibrated and atmospherically corrected. Time series extracted from VGT are consistent - no spatial and time errors - and comparable - consistent spatialy and radiographically.

1.1.4 More dataset and consistency issue

More dataset are available on internet such as Moderate Resolution Imaging Spectroradiometer (MODIS), Sea-viewing Wide Field-of-view Sensor (SeaWiFS) or LandSAT Enhanced Thematic Mapper Plus (LandSAT ETM+). Function to handle these nomenclatures are not ready yet. For the moment, you should make an index file (comparable to `Timesat` input, see Appendix B for more details).

According to Brown et al. 2006 [3], only one dataset should be analyzed at a time. Differences between sensors, such as different Spectral Response Functions, induce differences of NDVI values. Comparison can be done between anomalies but merging datasets is relatively difficult and common source of errors.

1.2 Noise reduction in time series : the Savitzky-Golay filter

NDVI time series can be very noisy. The noise is induced by atmospheric conditions variability as well as because of the Bidirectional Reflectance Distribution Function. Dust, cloud or ozone decrease the near infrared reflectance so noise is generally admitted to be sudden drops in the time series. A good filter should keep the upper envelope of the time series and delete the spurious drops (see Figure 2).

The first filter is the Maximum Value Composite (MVC)[12] which is already applied in the GIMMS and VGT datasets. It processes NDVI images over a time period - 10 days for VGT, 15 days for GIMMS - and keep only the maximum values. This methods deletes the negative noise coming from short cloudy period or atmospheric constituents.

However, according to Hird et al. (2009) [11], an other noise reduction methods should be applied on NDVI time series. Different methods have been developed to deal with this problem. Among them, some methods use a fitting function, such as asymmetrical Gaussian function fitting (Jonsson and Eklundh 2002) or double logistic function (Beck et al. 2006). Others use mathematical tools like Fourier smoothing (Sellers et al. 1996) or filtering techniques such as mean-value iteration filter (Ma and Veroustraete 2006) or Best Index Slope Extraction (Viovy et al. 1992).
The most widely used is the modified Savitzky-Golay filter, described by Chen et al. 2004 [4]. The idea is to use iteratively a polynomial function to fit the time series locally.

The comparison conducted by Hird et al. (2009) [11] explains that depending on the area and what is the goal of the analysis, filters are more or less reliable. For the moment, only the Savitzky-Golay filter is implemented in the `ndvits` package. We used an already available function from the Nonlinear Time Series Analysis (TISEAN) package [10]. Nonetheless, the full-iterative method explained by Chen et al. 2004 is not yet functional. For the moment, only a one-way Savitzky-Golay filter is applied to time series. For each point, a new value is computed according to a polynomial function fitted with the surrounding points (within a moving window). The parameters by default used for NDVI time series are a a polynomial function degree 2 and a moving windows of 5 points backward and 5 points forward.

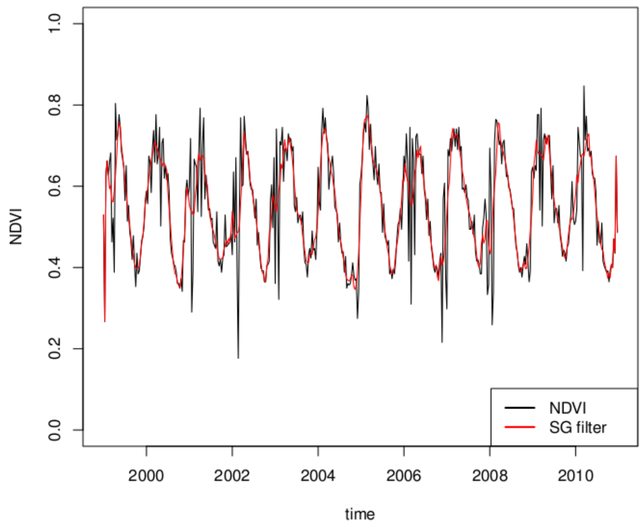


Figure 2: Savitzky-Golay filter on a NDVI time series from Samhamo Chiumia, Malawi.

1.3 Seasonal-Trend decomposition procedure based on LOESS (STL)

The Seasonal-Trend decomposition method based on LOcally wEighted regreSion Smoother (LOESS) is first described by Cleveland et al. 1990 [5]. It is a method available in the basic R package `ts`.

The time series is decomposed in two components : the seasonal and the trend (see Figure 5). The seasonal component is the mean annual signal : for each date, it is the mean of NDVI values of the different years. Then the trend is computed with the reminder of the seasonal component and at the end, the remainder component is the residuals from the seasonal plus trend fit.

STL decomposition doesn't make any assumption while breaking down the time series in two components. However, it prevents the detection of changes within time series. With NDVI data, the remainder part is often very high. The decomposition model doesn't explain so well the variability of the time series. Nevertheless, the STL method is an easy and quick way to get an overview of the signal periodicity.

1.4 Phenological metrics

Phenology is the study of the timing of recurring biological cycles related to climate. With NDVI time series, it is possible to extract metrics corresponding to critical points in the growing season. Phenological metrics are usefull to caracterize agricultural system and to study ecosystem responses to climate change.

There is no consensus yet on how to calculate metrics, every method has its limitations. However, it provides an independent measure to detect interannual variability in vegetation phenology.

The method to compute phenological metrics is inspired from a threshold-based technique developed by Brown et al. (2010).[2]. The main advantage of this method is its ability to handle a double cropping season without making any assumption on the ndvi signal (without using a modeling function). Even if there are some limitations, these metrics give an idea of the timing and the greenness of the growing season.

We proceed in 3 main steps :

1. Detecting the number of minimums and maximums with the mean signal.
   The mean signal over the year is computed (see Figure 3.A). Then we detect the minimums and the maximums. A maximum is detected if it is a maximum over a window of 6 measures and if its value is higher than the mean of the mean signal. Similarly, a minimum is detected if it is a minimum over a window of 6 measures and if its value is lower than the mean of the mean signal. Then, a routine check if minimums and maximums are one after another and if there are the same number of maximums and minimums.

   If the global mean of the time series is higher than 0.7 or smaller than 0.2, further calculations are stopped. The signal is similare to bare soil or forest and metrics became very difficult to compute, not reliable anymore.

2. Detecting annual minimums and maximums.
   The main issue is to handle the full season whenever it starts or stops. It is often that seasons straddles two years. Our solution is to start the annual time series two months before the global minimum detected during step 1. If two minimums (two seasons), we take into account the

minimum of them. For each year, we consider a time series with a length of 16 months to be sure to capture the full season (see figure 3.B) The annual minimums/maximums are detected within a two months windows around the minimums/maximums of the mean signal (detected during the first step, see Figure 3.A).

3. Calculating the phenological metrics.
   Every year, four metrics are extracted from the annual time series :

   - SOS : time for the Start Of the growing Season, also referred as green-up date
     We use a threshold method. The date of the start of the season is detected when the NDVI value has increase by 50% of the distance between the "left" minimum and the maximum, measured from the left minimum. The value of the threshold can be customized by the user.
   - EOS : time for the End Of the growing Season
     Similarly than SOS, the date of the end of the season is detected when the NDVI value has increase by 50% of the difference between the maximum and the "right" minimum, measured from the right minimum. The value of the threshold can be customized by the user.
   - LOS : Length Of the growing Season, $LOS = EOS - SOS$.
     Please be aware that LOS is a number of period.
   - cumNDVI : integral under the NDVI curve during the growing season, also called integrated NDVI.
     There isn't a consensus on the way to compute the cumulative NDVI, the minimum value is still in question. For the moment, we compute the full integral under the curves (comparable to the large seasonal integral in Timesat). Hopefully, soon there will be an option to consider either 0.1 as minimum (to delete bare soil effect) or the mean of the ndvi values during SOS and EOS (to take into account only the peak of the growing season).

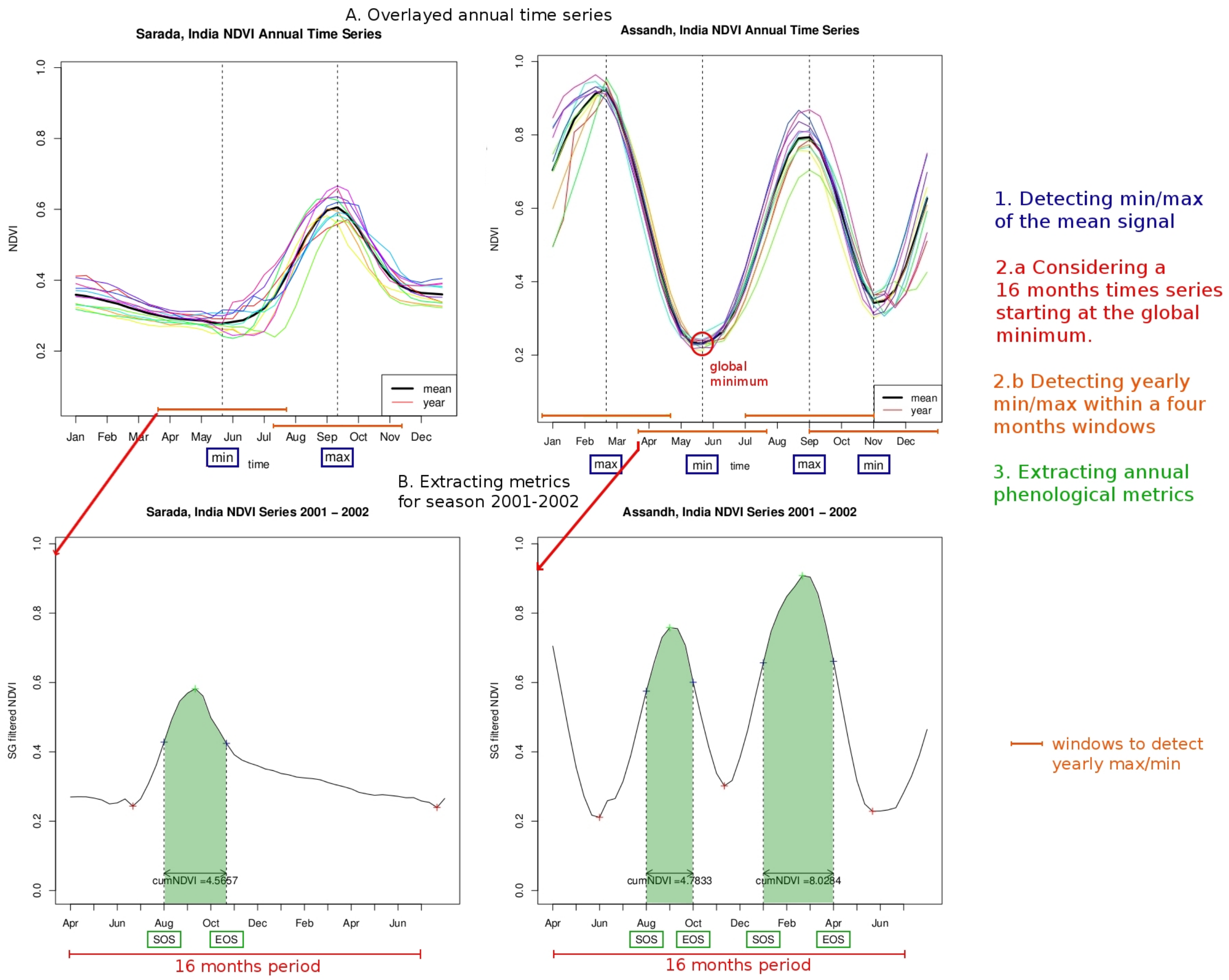


Figure 3: Method to compute phenological metrics in the case of one (left side) or two (right side) growing seasons.

1.5 Anomaly maps

Mapping anomalies of phenological metrics would be the best but it is too demanding in computation time for large area. That's why, we focused on two indicators to compute anomaly maps. The first one is annual maximum value of NDVI. The second one is the mean of the NDVI signal over the growing season.

Nevertheless, both of these methods takes into account the gross data, contaminated with clouds. For the maximum anomalies the effect of clouds is limited since we take into account only the maximum value. However, for the seasonal anomalies, it is subject to discussion. We consider that clouds have a constant effect over the years.

1.5.1 Maximum annual anomaly

Maximum anomaly focus on the value of the seasonal peak. It can be related with the maximum greenness of a year.

First, the map of the maximum of each year is computed. Then, the mean over the years of the computed maximums is calculated.

The anomaly map comes from the difference between the maximum of a specific year and the mean of the maximums. Maps are saved automatically in geotiff coded with 16 bits per pixel and a multiplication factor of 10000 to get value between -10000 and 10000.

The map of global mean over the maximum per year as well as the standard deviation map are also saved in the same way (over 16 bits, with a multiplication factor of 10000). The global mean of the maximums shows the spatial differences of the maximum NDVI. The standard deviation map shows how much the maximum of ndvi varies over the years within the map.

1.5.2 Seasonal anomaly

Computing seasonal anomaly put the emphasis on the differences of NDVI signal during a given period, we often focus on the growing season.

For each year, the mean of the NDVI is computed over the given period. Then, the global mean over the year is calculated for this period.

Anomaly maps comes from the difference between the annual mean of the period and the global mean. If the difference is positive, it means the season was "greener" than usual. On the contrary, if it is negative, the season was drier than normal. These values are computed for each pixel and the output maps of anomaly are coded with 16 bits, with a multiplication factor of 10000 to get values between -10000 and 10000.

The map of the global mean over the years for the given season as well as the standard deviation are also saved in the same way (coded in 16 bits, with a multiplication factor of 10000 applied). The global mean maps shows the distribution of the greenness over the given period and spatial differences of the mean signal for the season, identifies greener area than other. The standard deviation map shows how much seasons vary over the years within the map.

1.6 Breaks For Additive Seasonal and Trend (BFAST)

This method was proposed and developed by Verbesselt et al. (2010). [19, 18]. Verbesselt considers three main classes of changes in terrestrial plant ecosystems :

- phenological changes : changes detected in the seasonal component due to climate fluctuation or change in land cover type.
- abrupt changes : changes detected in the trend component caused by disturbances such as deforestation, floods, fires or insect attacks.
- gradual changes : long term change such as land degradation, very difficult to detect.

Time series are decomposed and iteratively fitted to a piecewise linear trend model and seasonal model. The seasonal model can be "harmonic"[19] or linear ("dummy")[18] With NDVI data, better results are obtain with harmonic model.

Changes are detected in both components (trend and seasonal) with an iterative procedure. By default, changes occurring within 2 years are not detected.

Compare to phenological approches which exploit the information contained in the shape of the seasonal growth cycle, the main innovation of this method is to take into account the whole time series to detect changes. The method is implemented in the R-package `Bfast`[19].

2 Results

The main goal of this section is to show and explain how the package `ndvits` works. The following analysis is focused on sites of the Systemwide Livestock Programme's project. There are examples of classical analysis which can be conducted with the `ndvits` package at different scales.

The emphasis was on how to call the function as well as how to interpret the results. However, this is not a manual, all the functions are not detailled. Please refer to the `ndvits` manual for further details.

2.1 Comparison of the NDVI time series in the African sites : a global scale analysis

A global scale comparison is suitable for having on overview of the situation. At this stage, we use GIMMS data to get an overall picture of the history of the vegetation in the different sites of Africa.

The long term availability of GIMMS is its main advantage and will allow us to capture the full picture since 1982. Its poor spatial resolution is in most of the case a limiting factor. For our study, it is not a big issue since we want to capture a global image and compare very distant sites and capture only global signal. Moreover, it will allow us to do the full analysis in the same time : the images are smaller and the full map of Africa can be load in R .

2.1.1 Extracting long term time series

```
#loading the package
library(ndvits)

#defining the local variables
shape="SLP_Africa"
shapedir="C:/DirShape/SLP_Africa.kml"
ndvidirectory="C:/DirGIMMS/"
region="AF"
Ystart=1982
Yend=2006
outfile="TSGIMMS_SLPAf.txt"
outfile2="SLPAf_GimmsSTL.pdf"
outfile3="SLPAf_GimmsTS.pdf"

#making a buffer around the individual points
newshape="SLP_Africa_Buffer"
dirshape = "C:/NewDirShape/"
pointtobuffer(shape,shapedir,ndvidirectory,region, Ystart,
    shapeext="kml", outshape = newshape, outdir = dirshape,
    type= "GIMMS")

#extracing time series
signal=TimeSeriesAnalysis(newshape, dirshape, ndvidirectory, region, Ystart, Yend, outfile, outfile2,
    outfile3, ext="shp", type="GIMMS", title="SLP african study sites - GIMMS NDVI Time Series", nb = 9)
```

On the R console, the following message will appear :

```
choose between one of the grouping factor available :
c("Name", "Descriptio", "coords.x1", "coords.x2", "coords.x3",
"coords.x1.", "coords.x2.")
 :
```

Please type "name" to merge NDVI time series by name. In our case, the signal of the 9 points from the same buffer (see appendix D, Figure 23) will be merge into one computing the average.

The input file is a kml file from Google Earth with one point per site (`shapedir = "C:/DirShape/SLP_Africa.kml"`). To have a better idea of the vegetation in the area and reduce the noise, we create a buffer around the points by calling the function `pointtobuffer` (see appendix D, Figure 23) .

Then, the main function `TimeSeriesAnalysis` is called to extract time series from the buffer. After a certain computation time, time series are finally extracted.

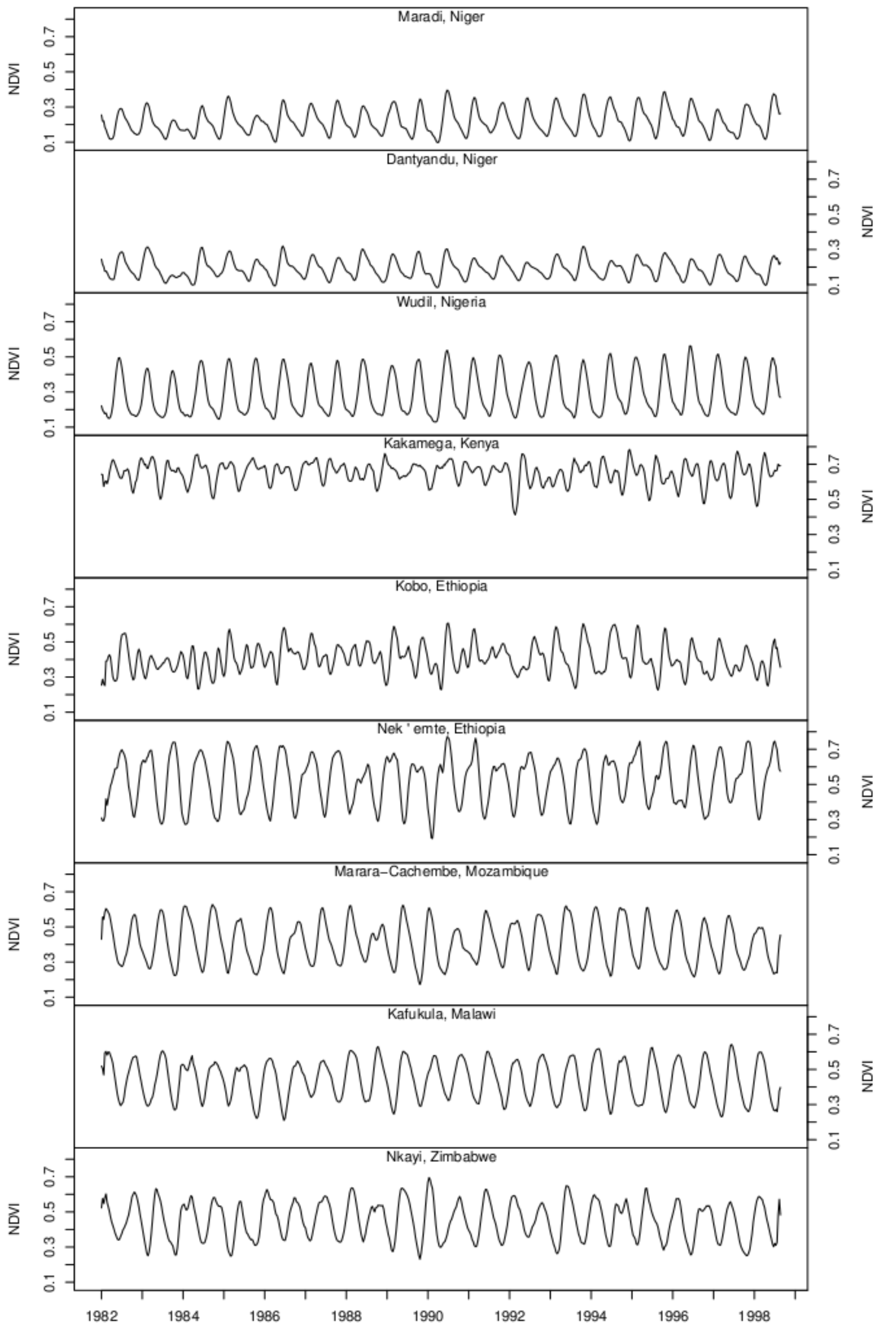


Figure 4: NDVI time series from African study sites of SLP project (saved in the file `output3`)

From Figure 4, we can have an idea of the vegetation in the different region as well as some ideas of the agricultural systems.

There are two very noisy signal, one in the region of Kakamega in Kenya and the other one in Kobo, Ethiopia. It can be explained by the very diverse land cover in the considered area.
For example, Kakamega is the home to a forest National Reserve. The NDVI values, very high in this region, are affected by this forest component. The non-seasonality is explained by a multi-component signal, coming from a very heterogeneous land cover. This effect is increased also by the the low resolution of the images, all the pixels are mixed pixels with a mixed signal.

Apart from these two signals, the seven others are clearly periodic. All of them have only one growing season : we observe only one peak per year.

In term of global NDVI signal, we can see that West African sites (especially Niger) has lower NDVI values than Southern African sites for example. These statements are in agreement with climatology knowledge. For example, Niger is a north from the Sub-Saharan Belt and has a hot and dry climate.

Moreover, we can obeserve that, compare to other sites, the length of the season seems shorter in two study sites in Niger : the peaks are narrow. In the two sites in Niger, the region of Maradi and Dantyandu, croppping season are short and during the rest of the year the landscapes seems dry (low NDVI values).

From these time series, we can also discuss the inter annual variability but it is easier to do with the decomposed time series.

2.1.2 Seasonal-Trend decomposition

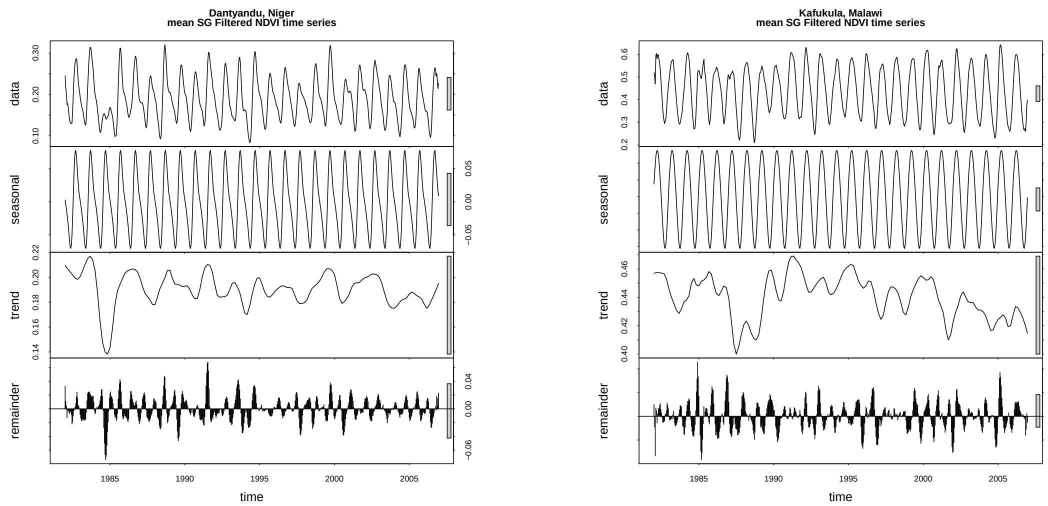


Figure 5: Seasonal-Trend decomposition of Dantyandu and Kafukula NDVI time series (saved in the file `output2`)

The seasonal component of the STL decomposition give an idea of the growing season pattern, excluding the yearly variation, while the trend component can show the inter annual variability

In the example shown Figure 5, we can see that Niger and Malawi have their growing season and their "dry" season at the opposite time. This is perfectly normal since the two sites are on opposite sides of the equator.

In Dantyandu, Niger, the trend component shows pretty well the big drought of 1985. The trend component of Kafukula is decreasing from year to year. In the original time series, we can observe that the level of the minimum values are decreasing.

However, the gray bar on the side of the graph, shows the scale of the different components. The scale of the trend component is very small so it should be interpretated too deeply. Generally, the remainder component is still very important. The decompostion doesn't explain totally the variability of the NDVI signal.

2.1.3 Comparison of the variability intra-region

```
WA_sites=c("Maradi, Niger", "Dantyandu, Niger", "Wudil, Nigeria")
EA_sites=c("Kakamega, Kenya", "Kobo, Ethiopia", "Nek ' emte, Ethiopia")
SA_sites=c("Marara-Cachembe, Mozambique", "Kafukula, Malawi", "Nkayi, Zimbabwe")

OverlayTS(signal2[rownames(signal2)\%in\%EA_sites,], Ystart, 36, "SLP East Africa study sites -
    GIMMS NDVI Time Series", "SLPEA_GIMMS.pdf")
OverlayTS(signal2[rownames(signal2)\%in\%SA_sites,], Ystart, 36, "SLP Southern Africa study sites -
    GIMMS NDVI Time Series", "SLPSA_GIMMS.pdf")
OverlayTS(signal2[rownames(signal2)\%in\%WA_sites,], Ystart, 36, "SLP West Africa study sites -
    GIMMS NDVI Time Series", "SLPWA_GIMMS.pdf")
```

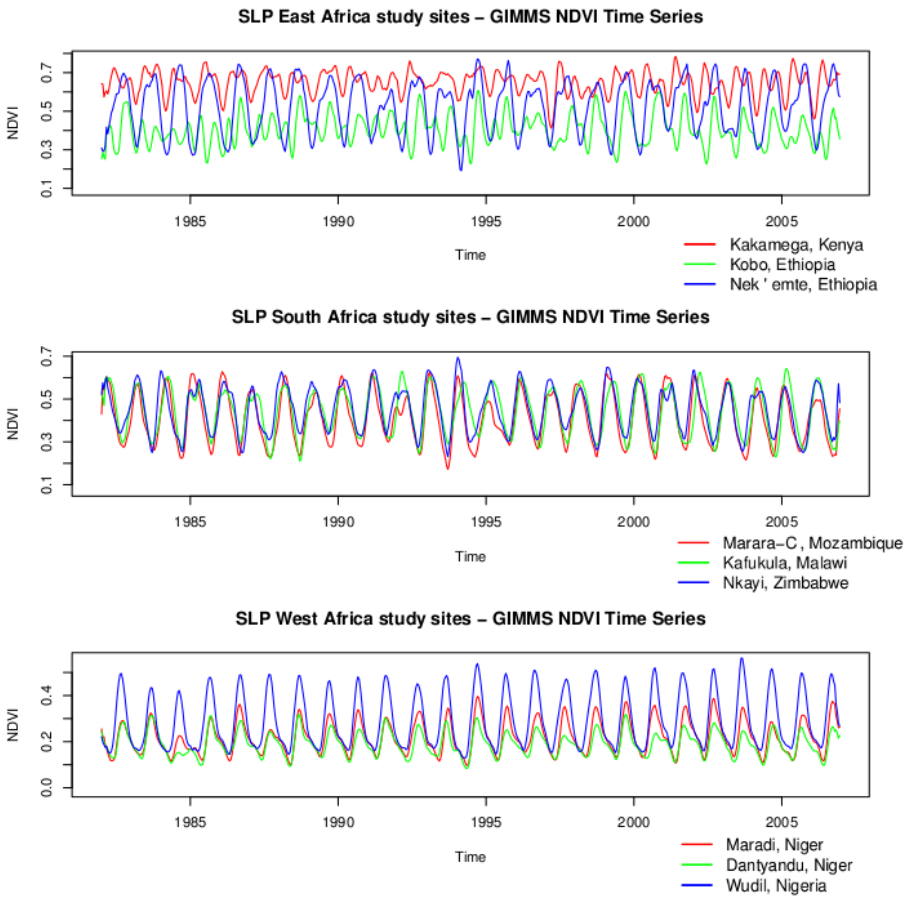


Figure 6: Variability intra-region

We can see clearly the vegetation in East Africa is very heterogeneous, mainly due to the two "noisy" time series in Kobo and Kakamega.

At the contrary, the seasons in Southern Africa and West Africa are more similar with nearly the same growing season in the 3 study sites of the region.

From Figure 6, it is easy to compare NDVI signals and especially the starting time of the season and the value of NDVI.

For example, in West Africa, the season start earlier in Wudil than in Dantyandu and Maradi, the peak is higher (around 0.5) and during the dry season, the ndvi value is also higher (around 0.2). Vegetation seems more dense than in the two study site in Niger. We can also find some differences between the two site in Niger : the growing seasons are greener in Maradi than in Dantyandu. But in both sites, the big drought in 1983 is clearly visible.

We can do the same kind of analysis in Southern Africa. Nkayi in Zimbabwe as often a higher value of NDVI than the region of Kafukula in Malawi or Marara-Cachembe in Mozambique. Nevertheless, we can observe that perturbation doesn't occur at the same time in all the sites. For example, in 1989, Kafukula and Marara-Cachembe had a small spring season while it didn't affect Nkayi. In 1992, Nkayi and Marara-Cachembe knew a bad season but not Kafukula. More recently, in 2000, Nkayi was the only study sites to face a drought while Kafukula and Marara-Cachembe knew a normal growing season.

2.2 Analysis of phenological metrics at regional scale level

The aim of this section is to analyze phenological metrics. To detect metrics in a desirable way, time series have to be as consistent as possible and free of noise. That's why, we choose to use Spot Vegetation data for this analysis.

2.2.1 Calculating phenological metrics with one or two growing seasons : a study focus on South Asian study sites

```
#loading the package
library(ndvits)

#defining the local variables
shape="SLPSAs_Buffer"
shapedir = "/dirShape/"
ndvidirectory="/dirVGTdata/"
region=""
Ystart=1999
Yend=2010
outfile="TSVGT_SPLSAs.txt"
outfile2="SLPSAs_VGTSTL.pdf"
outfile3="SLPSAs_VGTTS.pdf"
outmetricspdf="MetricsSAs.pdf"
outmetricstxt="MetricsSAs.txt"
period=36

#Extracting time series
signalSAs=TimeSeriesAnalysis(shape, shapedir, ndvidirectory, region, Ystart,
    Yend, outfile, outfile2, outfile3, type="VITO_VGT",
    title="SLP South Asia study sites - VITO NDVI Time Series")

#Calculating phenological metrics
sas=AnPhenoMetrics(signalSAs,outmetricstxt,outmetricspdf,Ystart,period=period)
```

The function `AnPhenoMetrics` computes automatically phenological metrics for the different time series. Depending on the area, farmers in South Asia grow crops once or twice a year. Our way to calculate metrics can handle a second cropping season (see section 1.4). To understand how does it works, we focus on two regions of India : Sarada in Rajastahn with a single cropping season between August and October and Assandh in Haryana with a second cropping season between January and April (see Figure 7).

Figure 7 and 8 are extracted from the file `outmetricspdf`. Similarly table 1 and 2 are extracted from the file `outmetricstxt`.

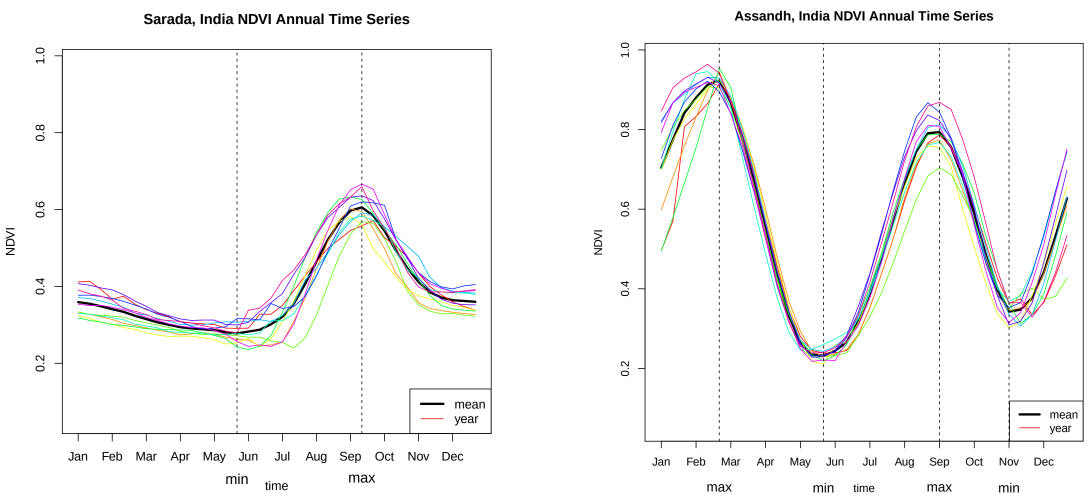


Figure 7: Annual mean signal in the case of a single (left side) and double (right side) cropping season

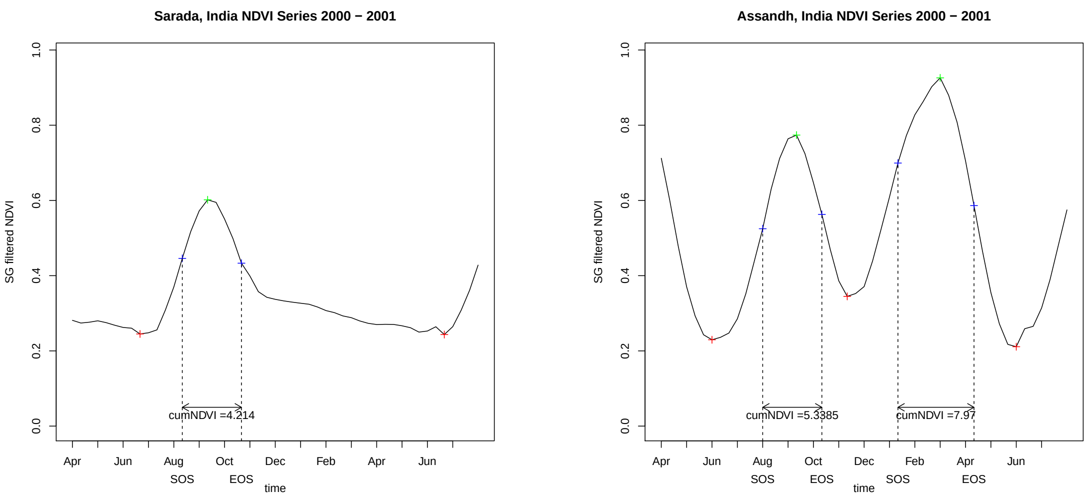


Figure 8: Phenological metrics for the season 2001-2002 in the case of a single (left side) and double (right side) cropping season

Table 1: Calculated phenological metrics in Sarada, Rajasthan, India

Year	Max	SOS	EOS	LOS	cumNDVI
1999	0.57	23	32	9	5.102
2000	0.6	23	30	7	4.214
2001	0.58	22	30	8	4.566
2002	0.57	25	30	5	3.137
2003	0.63	22	31	9	5.6
2004	0.58	23	31	8	4.678
2005	0.59	23	32	9	5.284
2006	0.62	24	31	7	4.496
2007	0.64	22	31	9	5.655
2008	0.67	23	31	8	5.184

Table 2: Calculated phenological metrics in Assandh, Haryana, India

	Max_1	SOS_1	EOS_1	LOS_1	$cNDVI_1$	Max_2	SOS_2	EOS_2	LOS_2	$cNDVI_2$	$cNDVI_{tot}$
1999	0.79	22	29	7	5.46	0.95	3	11	8	7.11	12.57
2000	0.77	22	29	7	5.34	0.93	2	11	9	7.97	13.31
2001	0.76	22	28	6	4.78	0.91	1	10	9	8.02	12.81
2002	0.7	23	29	6	4.45	0.95	4	10	6	5.57	10.02
2003	0.79	22	29	7	5.72	0.95	2	10	8	7.36	13.09
2004	0.77	22	28	6	4.83	0.93	1	10	9	8.47	13.31
2005	0.81	22	29	7	5.71	0.92	1	10	9	7.99	13.7
2006	0.87	22	28	6	5.39	0.93	0	10	10	9.07	14.47
2007	0.84	21	29	8	6.44	0.92	1	10	9	8.23	14.67
2008	0.81	22	28	6	5.08	0.96	1	10	9	8.58	13.66

All the metrics displayed Figure 8 are saved Table 1 and 2 under the year 2000. As explained section 1.4, even if a season lasts until the following year, metrics are saved under the year of the global minimum. For example, the second season of Assandh takes place from January to April 2001 but its metrics are saved as the second season of the year 2000.

In Sarada, the season start at the period 23 which corresponds, with 3 values per months, to the second third of August. It ends at period 30, last third of October. Dotted lines reprensent these two metrics in Figure 8 . Similarly, SOS and EOS for the two seasons in Assandh are plotted in Figure 8 and saved table 2. The second season, from second third of January to second third of April, seems to be the most important one with a maximum of 0.93. Cumulative NDVI (cumNDVI) is the full intergrale under the curve from SOS to EOS.

Globally, the EOS seems very stable from year to year. SOS is most of the time very constant but it is sometimes late.

Exploring the table, we can observe that the season 2002 - 2003 seems very dry season for both sites. In Sarada, the season started one month late (period 25 compare to a mean around 22-23). It causes a short growing season of 5 periods (less than two months) and a low cumNDVI (3.1) . In Assandh, the process is nearly similar. The first growing season is a bit delayed (starts period 23) and the second one is very late (starts period 31). The lenght of the second season, the most important one in term of cumNDVI, is one month shorter than usual. As a result, cumNDVI in Sarada and Assandh for the season 2002 - 2003 are low. It means that these two region were confronted with a bad growing season mainly due to a delay of the green-up date.

This is only a descriptive analysis. For more statistical tests, we should restrict the study to simple cropping season to be able to compare metrics.

2.2.2 Statistical analysis of phenological metrics, a study focus on Southern Africa

Extracting time series and phenological metrics

```
#loading the package
library(ndvits)

#defining the local variables
shape="SLPSA_Buffer"
shapedir = "/media/FreeAgent GoFlex Drive/Shape/SLP/"
ndvidirectory = "/media/FreeAgent GoFlex Drive/GrossData/Vito/Africa/"
region = "Africa"
Ystart = 1999
Yend = 2010
outfile = "TSVGT_SPLSA.txt"
outfile2 = "SLPSA_VGTSTL.pdf"
outfile3 = "SLPSA_VGTTS.pdf"

#extracing time series
signalSA = TimeSeriesAnalysisSouthern Africa : (shape, shapedir, ndvidirectory, region, Ystart, Yend, outfile, outfile2,
    outfile3, type = "VITO_VGT", title="SLP Southern Africa study sites - VITO NDVI Time Series")

#calculating phenological metrics
sa = AnPhenoMetrics(signalSA, "MetricsSA.txt", "MetricsSA.pdf", Ystart, period = 36)
```

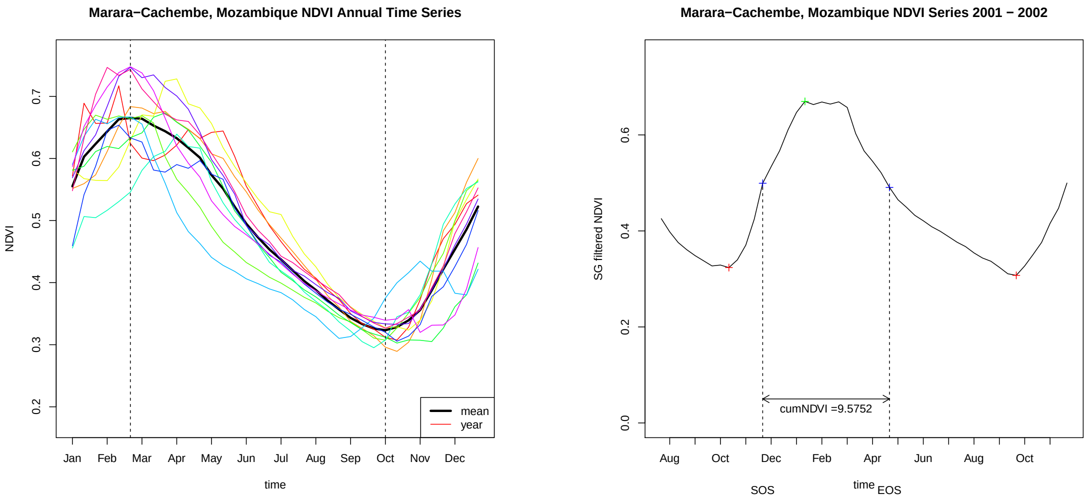


Figure 9: Output of `AnPhenoMetrics` function

The figure 9 show the output of `AnPhenoMetrics` function.

The graph on the left shows the annual NDVI time series overlaid as well as their mean (in black). We can see how seasons vary and for example, we observe that around the minimum date, the NDVI values for the different years are similar. While it is very heterogeneous around the maximum date. The peak of the growing season depends yearly while the "dry" season is more constant as well as the start of the season.

The graph on the right part show the season starting in December 2001 and finishing late April 2002. Even if the growing season occurs most of the time during 2002, the metrics will be saved for the year 2001. All the metrics for the different years are saved in the `outfile3`.

Once the metrics computed, an analysis to compare metrics between site and year can be conducted with R .

Transformation of data and exploration

We decided to focus on two metrics : cumNDVI and SOS. The first one is related to the Net Primary Production and the yield of the cropping season. The second one is an indicator of the start of the season.

The aim is to understand the influence of the different sites and the years on these two metrics. Therefor, we proceed a linear regression analysis. We supposed the year as a qualitative variables, because of the unpredictableness of the growing season and the sudden drought happening This hypothesis can be checked on the left side of Figures 10 and 11, showing the distribution of the variables over the years.

The two explicative variables, sites and year, are qualitative and supposed independent. The main test will be an analysis of variance to make a global comparison of means of the whole data (ANOVA).

```
#Transformation of variables
cumNDVI=as.numeric(sa\$cumNDVI)
SOS=as.numeric(sa\$SOS)
SOS=ifelse(SOS>18,SOS-36,SOS)
year=as.factor(rep(sa\$year,3))
site=as.factor(c(rep(sa$names[1],
    length(sa$year)), rep(sa$names[2],
    length(sa$year)), rep(sa$names[3],
    length(sa$year))))

#Exploration of the cumNDVI
#calculating the mean per group
tapply(cumNDVI, site, mean)
#plotting distribution
boxplot(cumNDVI~site)
pp=16:18
plot(year, cumNDVI, col=rainbow(3)
    [as.numeric(site)], pch=pp[as.numeric(site)])
legend("topright", levels(site),
    col= rainbow(3),pch=pp)
interaction.plot(year,site,cumNDVI,
    ylab="cumNDVI")

#Exploration of the SOS
boxplot(SOS~site)
plot(year,SOS,col= rainbow(3)[as.numeric(site)],
    pch=pp[as.numeric(site)])
legend("topleft", levels(site), col= rainbow(3),
    pch=pp)
```

First the data is transformed and two vectors are created with the two interesting metrics. We transformed the variable SOS to become continue. Originally, it is a number indicating the period of the year when the start of the season is detected. This number is not continue (34, 35 is followed by 0, 1). Because the SOS is always around the start of the year, we changed high values to be negative one. The two explicative variables are considered as factor : `as.factor()`.

The exploration of variables is a very important step to know how is distributed the variables.

`tapply` compute the mean per group while `boxplot` and `plot` are more visual.

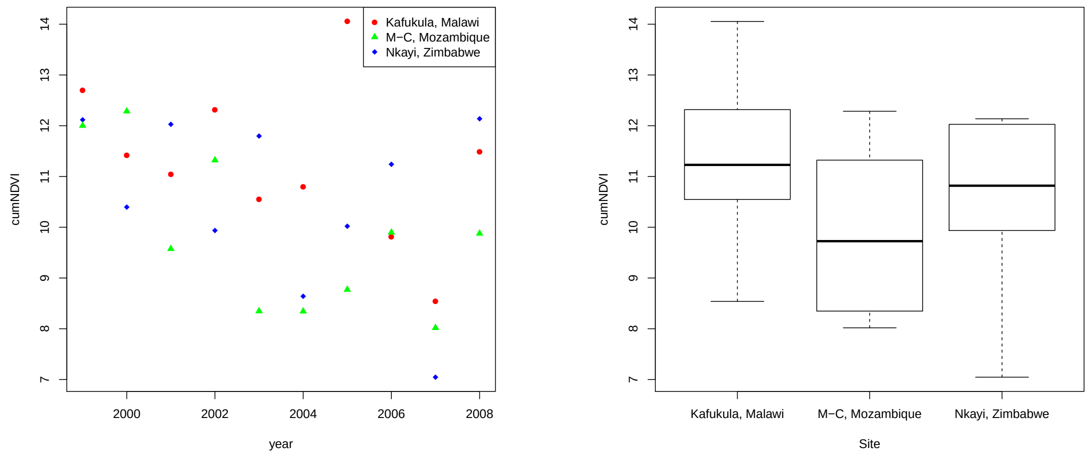


Figure 10: Distribution of cumNDVI over the years and the sites

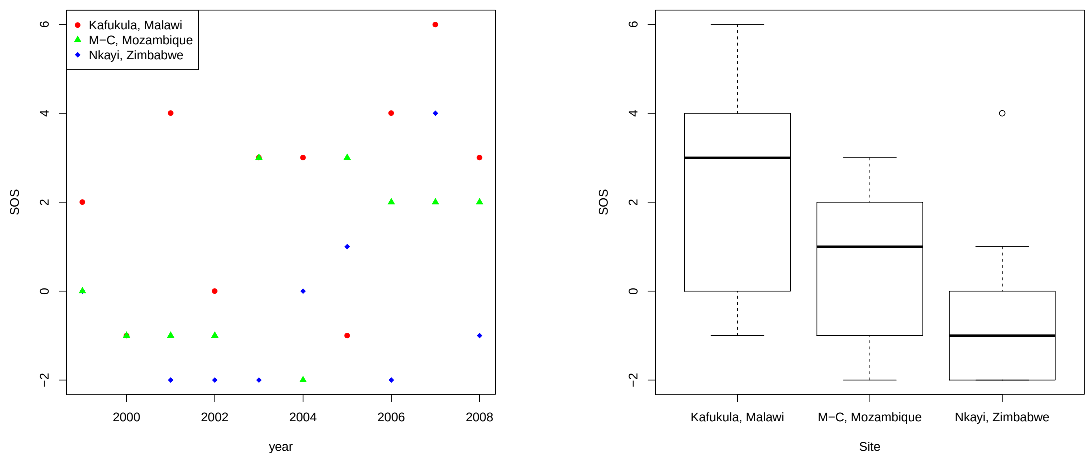


Figure 11: Distribution of SOS over the years and the sites

Test of normality with cumNDVI

```
#basic test of normality and homogeneity :
#normally distributed population?
shapiro.test(cumNDVI)


        Shapiro-Wilk normality test
data:  cumNDVI
W = 0.9738, p-value = 0.6463


#homogeneity of variances?
bartlett.test(cumNDVI, year)
bartlett.test(cumNDVI, site)


        Bartlett test of homogeneity of variances
data:  cumNDVI and year
Bartlett's K-squared = 7.9464, df = 9, p-value = 0.5396

        Bartlett test of homogeneity of variances
data:  cumNDVI and site
Bartlett's K-squared = 0.0938, df = 2, p-value = 0.9542
```

The two main hypothesis that have to be checked before doing a linear model are :

- the residuals (or the variable) follow a low-normal
- the variance are homogene, equal across the population

There are others hypothesis but must be check after the model is created.

Shapiro–Wilk test is a test of normality. The null hypothesis is that the sample came from a normally distributed population. In our case, P-value is more than 0.05 so we can't reject H0, cumNDVI follow a normal distribution.

Bartlett's test checks the homoscedasticity, it means if the variances are equal across samples. With two P-value higher than 5%, we can conclude that the 3 sites, as well as the 10 different years are from populations with equal variances.

Linear Modelling

```
lm1=lm(cumNDVI~site)
anova(lm1)
```

```
Analysis of Variance Table
Response: cumNDVI
          Df Sum Sq Mean Sq F value Pr(>F)
site       2 10.191  5.0953  1.9949 0.1556
Residuals 27 68.963  2.5542
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

ANOVA compare the variability explained by the model to the residual variability. The first model is built with the variable `site`. The variability explained by the model is the variability inter sites, the differences of the average of cumNDVI beween different sites. The residual variability is the variability intra site. ANOVA make an average comparison of these two variability. The F test of the ANOVA make the following null hypothesis : "There isn't difference between the cumNDVI average of the different sites".

In our case, the P-value here is higher than 0.05, we accept H0 : there isn't a significative difference between sites in term of cumNDVI.

In the following, the same kind of analysis is made with different illustrative variables.

```
lm2=lm(cumNDVI~year)
anova(lm2)
```

```
Analysis of Variance Table
Response: cumNDVI
          Df Sum Sq Mean Sq F value  Pr(>F)
year       9 41.104  4.5671  2.4007 0.04938 *
Residuals 20 38.049  1.9024
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
lm3=lm(cumNDVI~site+year)
anova(lm3)
```

```
Analysis of Variance Table
Response: cumNDVI
          Df Sum Sq Mean Sq F value  Pr(>F)
site       2 10.191  5.0953  3.2922 0.06047 .
year       9 41.104  4.5671  2.9510 0.02423 *
Residuals 18 27.858  1.5477
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The linear model `lm2` makes a linear regression of cumNDVI over the years. The p-value of ANOVA test is lower than 0.5. There is a significant differences of cumNDVI between the years. It seems that the variable year is more important that the variable site to explain cumNDVI.

In the model `lm3`, we build a model taking into account both variables. First, all the variability explained by `site` is took into account. Then, the added variability explained by `year` is retain from the residuals. In our case, the explained sum of the squares (SS) of the variable year (41.104) is the same than in the model `lm2`. It means that none of the variability of cumNDVI explained by year can be explained by site : the two variables are not correlated.

To know which model to keep (`lm2` or `lm3`), we run an ANOVA test between the two models. It will tell us if the parameters added by the variables `site` (lower degree of freedom) is worth in term of the variability explained.

```
anova(lm2,lm3)
```

```
Analysis of Variance Table
Model 1: cumNDVI ~ year
Model 2: cumNDVI ~ site + year
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     20 38.049
2     18 27.858  2    10.191 3.2922 0.06047 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The p-value of the ANOVA is higher than 0.5 . Both model are not significantly different. The degree of freedom lost by taking into account the variable site doesn't worth the few variability explained. So the best model to explain `cumNDVI` is `lm2`, taking into account only the year. The coefficient of determination R^2 is the proportion of variability in a data set that is accounted for by a statistical model. For `lm2`, $R^2 = \frac{SS_{model}}{SS_{total}} = 41.10441.104 + 38.049 = 0.519$ The model explains only 52% of the total variability of cumNDVI.

Interpretation of the p-value make sense only if all the hypothesis of the linear modeling are respected.

```
#Verification of the hypothesis
plot(lm2,1)
plot(lm2,4)
```

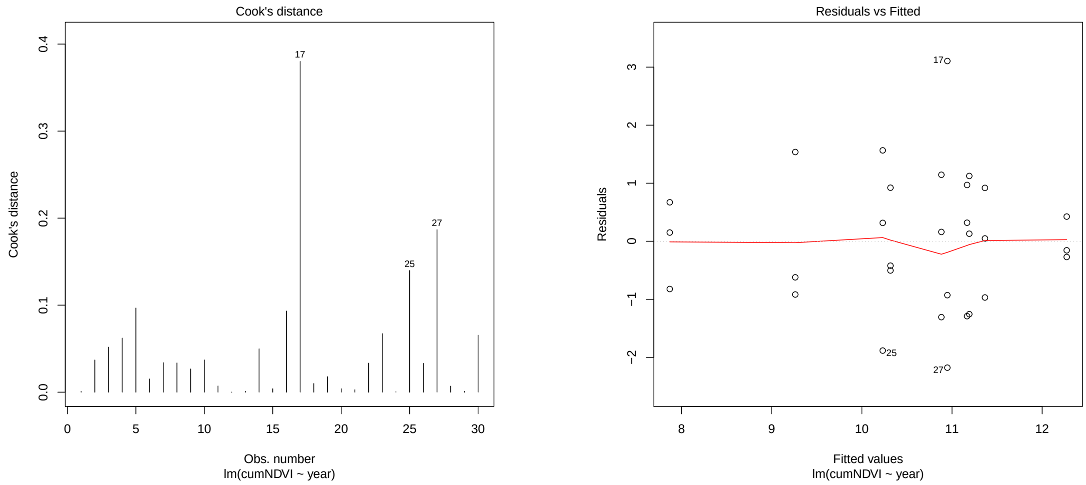


Figure 12: Verification hypothesis cumNDVI

From the right graph in Figure 12, the following hypothesis can be verified :

- The model seems correct.
- Residuals are independent from the explicated variable.
- Residuals are centered around 0.

From the left graph in Figure 12, we can check that there is no aberrant values : all the Cook's distance are lower than 0.5.

A completely similar analysis is done with the variable SOS.
Contrary to the variable cumNDVI, SOS is significantly correlated to the site (`lm1`, p-value = 0.017) but not to the year (`lm2`, p-value = 0.33). It is something that could be observed in Figure 11 : the mean of SOS in different sites are really different. The annual variability is less important than the variability inter sites. The start of the season is kind of constant over years for each sites.

If both site and year are took into account (`lm3`), the loose of degree of freedom is not worth the gain in explained variability (p-value = 0.12). So the best model to explain SOS is `lm1`.

Checking the assumption made by the linear modeling, we can say that

- normality of the error distribution (`shapiro.test(residuals(lm1))`, p-value = 0.48 > 0.05.
- independence of the errors, residuals are independent from the predicted value (Figure 13)
- model seems correct : linearity of the relationship between dependent and independent variables (Figure 13)
- constant variance across the sites (`bartlett.test(SOS,site)` p-value = 0.79).
- no outlier : Cook'sdistance lower than 0.5 for all the measures (Figure 13)

```
lm1=lm(SOS~site)
lm2=lm(SOS~year)
lm3=lm(SOS~site+year)
#ANOVA
anova(lm1)
```

```
Analysis of Variance Table
Response: SOS
          Df  Sum Sq Mean Sq F value  Pr(>F)
site       2  39.467 19.7333  4.7276 0.01736 *
Residuals 27 112.700  4.1741
```

```
anova(lm2)
```

```
Analysis of Variance Table
Response: SOS
          Df Sum Sq Mean Sq F value Pr(>F)
year       9 54.167  6.0185  1.2283 0.3327
Residuals 20 98.000  4.9000
```

```
anova(lm3)
```

```
Analysis of Variance Table
Response: SOS
          Df Sum Sq Mean Sq F value   Pr(>F)
site       2 39.467 19.7333  6.0683 0.009674 **
year       9 54.167  6.0185  1.8508 0.127306
Residuals 18 58.533  3.2519
```

```
anova(lm1,lm3)
```

```
Analysis of Variance Table
Model 1: SOS ~ site
Model 2: SOS ~ site + year
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     27 112.700
2     18  58.533  9    54.167 1.8508 0.1273
```

```
#Verification of the hypothesis
shapiro.test(residuals(lm1))
```

```
shapiro.test(residuals(lm1))
        Shapiro-Wilk normality test
data:  residuals(lm1)
W = 0.9679, p-value = 0.4842
```

```
bartlett.test(SOS,site)
```

```
bartlett.test(SOS,site)
        Bartlett test of homogeneity of variances

data:  SOS and site
Bartlett's K-squared = 0.4715, df = 2, p-value = 0.79
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
plot(lm1,1)
plot(lm1,4)
```

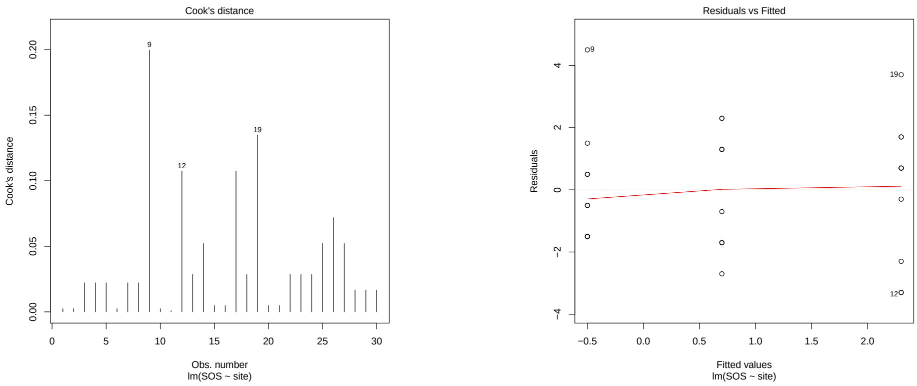


Figure 13: Verification hypothesis Start of Season

2.3 Mapping vegetation anomalies : a study focus on the district of Mzimba in Northern Province of Malawi

2.3.1 Extracting time series and metrics

```
#loading the package
library(ndvits)

#defining the local variables
ndvidirectory="/dirClipData/Vito_Mzimba/"
region="Mzimba"
Ystart=1999
Yend=2010
region="Mzimba"
shape="SLP_Mzimba"
shapedir=system.file("extdata/shape/", package="ndvits")

#Extracting time series
signal=TimeSeriesAnalysis(shape, shapedir, ndvidirectory, region, Ystart,
    Yend,"TS_Mzimba.txt","STL-Mzimba.pdf","MzimbaTS.pdf")
OverlayTS(signal,Ystart,36,"","Mzimba.pdf")

#Calculated annual phenological metrics
Mzimba=AnPhenoMetrics(signal,"Mzimba-Metrics.txt","Mzimba-Metrics.pdf",1999,36)

#printing table of cumNDVI and maximum
cNDVI=as.data.frame(Mzimba$cumNDVI)
rownames(cNDVI)=Mzimba$year
colnames(cNDVI)=Mzimba$names
print(cNDVI)
Max=as.data.frame(Mzimba$M)
rownames(Max)=Mzimba$year
colnames(Max)=Mzimba$names
print(Max)
```

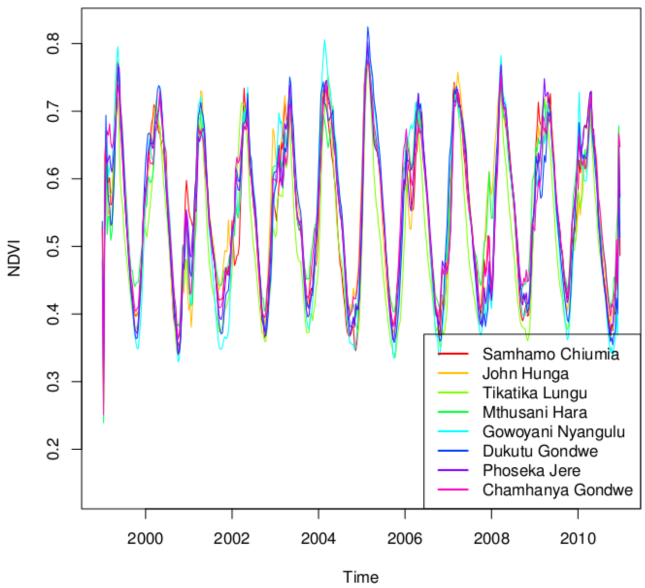


Figure 14: Overlay NDVI time series in the region of Mzimba, Malawi

Figure 14 gives an overview of the NDVI time series of the 8 villages of the study in the region of Mzimba. Because the area is small, it is interesting to see the spatial distribution of the NDVI signal. First, a map of basic statistics of the region of intersest is done.

2.3.2 Mapping local statistics

```
#computing local statistics over the period 1999 - 2010
res=maplocalstat(ndvidirectory, region, Ystart, Yend, type="VITO_CLIP", outname="Mzimba",
    shapefile=shape, shapedir=shapedir)

#saving the mean and sd maps
savemap(res$mean, outname = "MzimbaMean9910",outext="png", title="1999 - 2010 mean of NDVI
    in Mzimba", shapefile=shape, shapedir=shapedir, label="Village")
savemap(res$sd, outname = "MzimbaSD9910",outext="png", title="1999 - 2010 standard
    deviation of NDVI in Mzimba", shapefile=shape, shapedir=shapedir, label="Village", pal="OrRd")

#saving the min and max maps together
list=paste("Mzimba", c("max","min"),"9910.tif",sep="")
multimap(list, names=c("max","min"), outname="Mzimba9910minMax", org=c(2,1), outext="png",
    title="1999 - 2010 min/max of NDVI values in Mzimba", shapefile=shape, shapedir=shapedir)
```

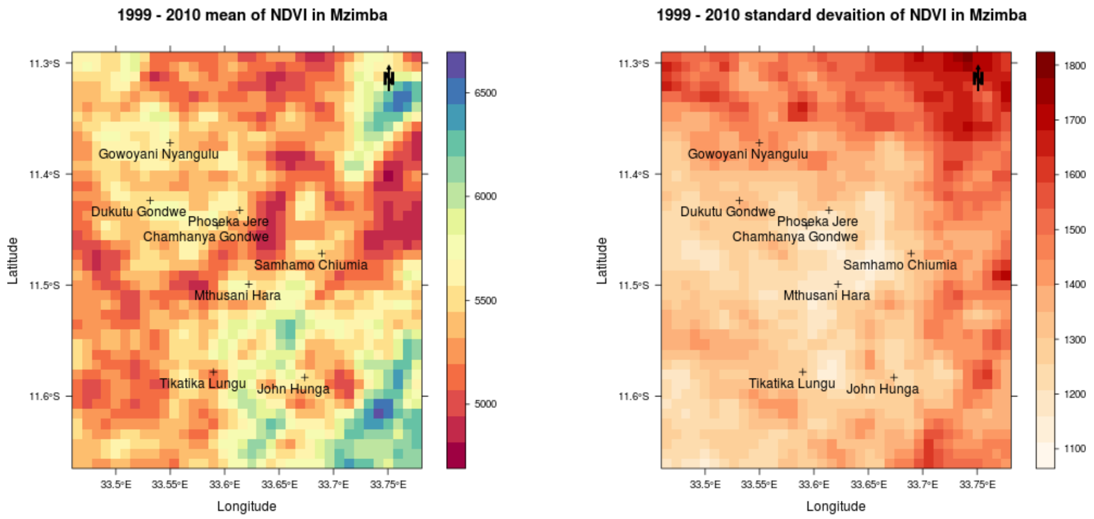


Figure 15: Mean and standard deviation of NDVI values in Mzimba region

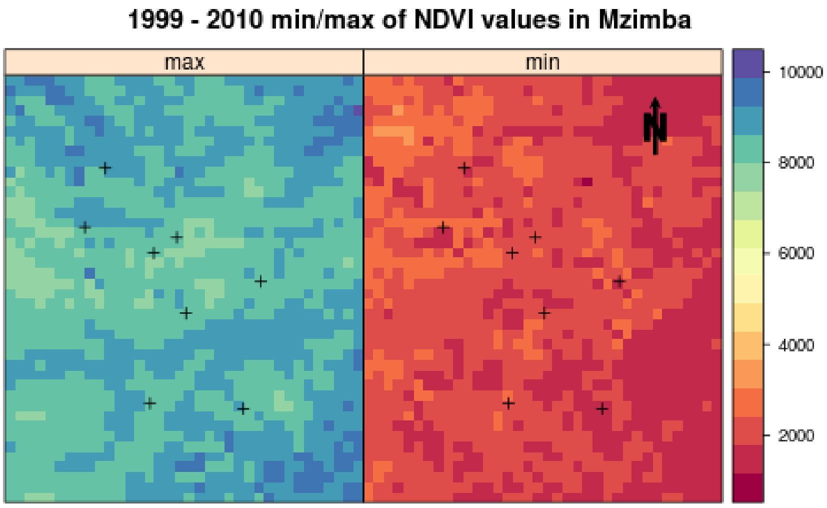


Figure 16: Minimum and maximum of NDVI values in Mzimba region

The mean NDVI 15 reflects the greenness of the area between 1999 and 2010. The green/blue area has a high mean NDVI (0.65), it is most probalby a forest area. All the other seems more or less equal, around 0.5 - 0.55.

The standard deviation is higher for the high ndvi mean. The distribution is very special, with a high standard deviation at north, east.

The combination mean ndvi signal and standard deviation is mainly due to land cover. That's why we did see the GLC map [6] see Figure 17.

Figure 16 gives an idea of the maximum and the minimum per pixel during the period 1999-2010. The period is wide and these two indicators, especially the minimum, can be influenced by noise.

In the area of Mzimba, the mean NDVI is between 0.5 and 0.65. The maps prensented Figure 15 show some interesting patterns.

We can, for example, observe that around the cities of T. Lungu, M. Hara and S. Chiumia the mean NDVI values is low (0.5) and the standard deviation as well (0.11)

The region of the North East and South east have a high ndvi.

It can easily be linked with the land cover.

" The inter-annual variation in the integrated NDVI represents the stability of crop growth. Higher inter annual variability means more vulnerable to changes in weather leading to significant changes in the levels of accumulated biomass and crop yield. Recurring rainfall deficiency or delayed sowings tend to reduce the biomass levels of agricultural crops. "NNRMS

We can study more in details the feature of the maximum per year, computing anomaly maps.

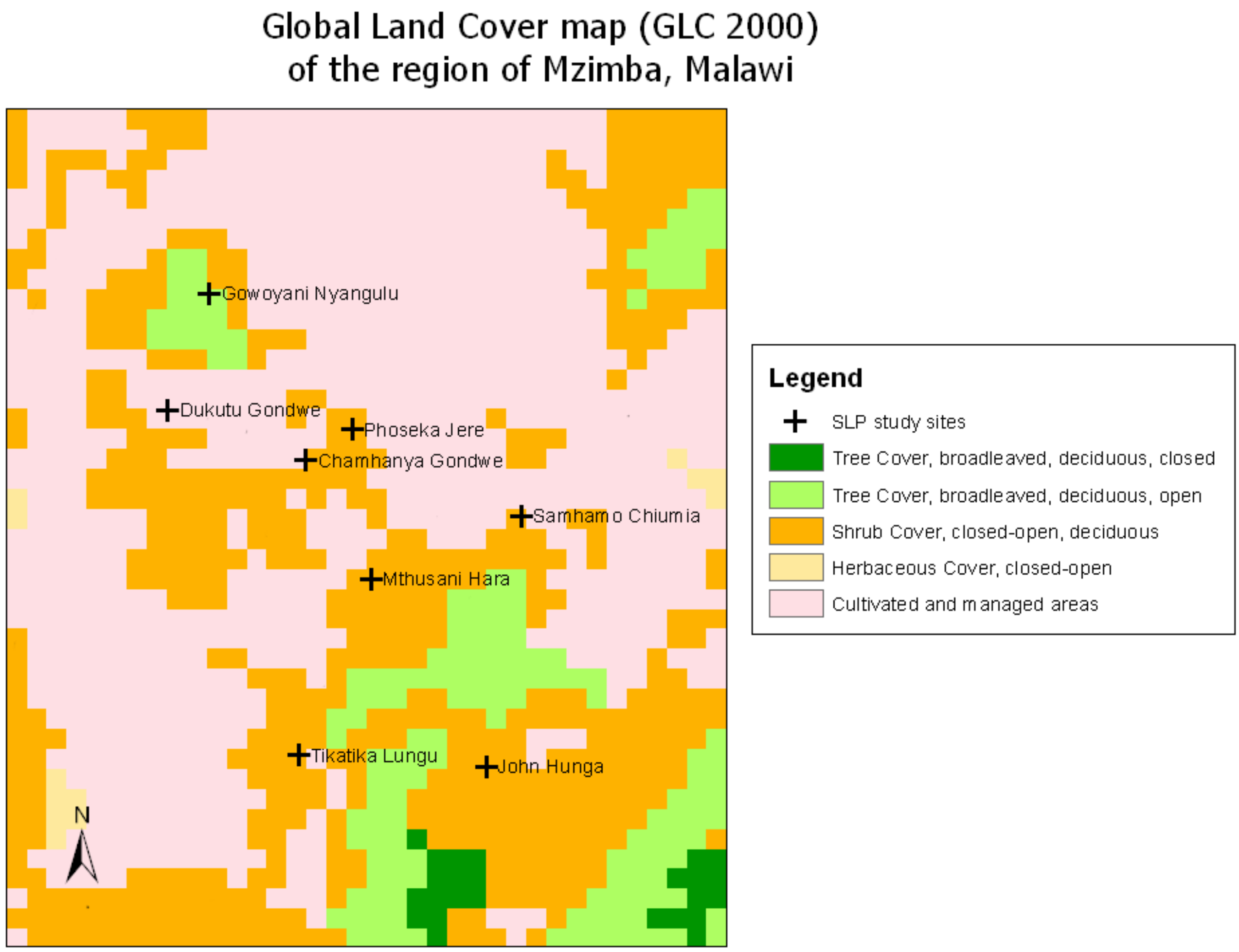


Figure 17: Map from the GLC project covering the area of Mzimba

2.3.3 Mapping maximum anomaly

```
MaxAnomaly(ndvidirectory, region, 1999, 2010, "Mzimba", outext="show",
    shapefile=shape, shapedir=shapedir, label="Village")
listname=paste("MzimbaM",1999:2010,".tif", sep="")
multimap(listname, as.character(1999:2010), outname="MzimbaMax9910", org=c(4,3),
    title="Maximum Anomaly Mzimba", shapefile=shape, shapedir=shapedir, outext="png")
```

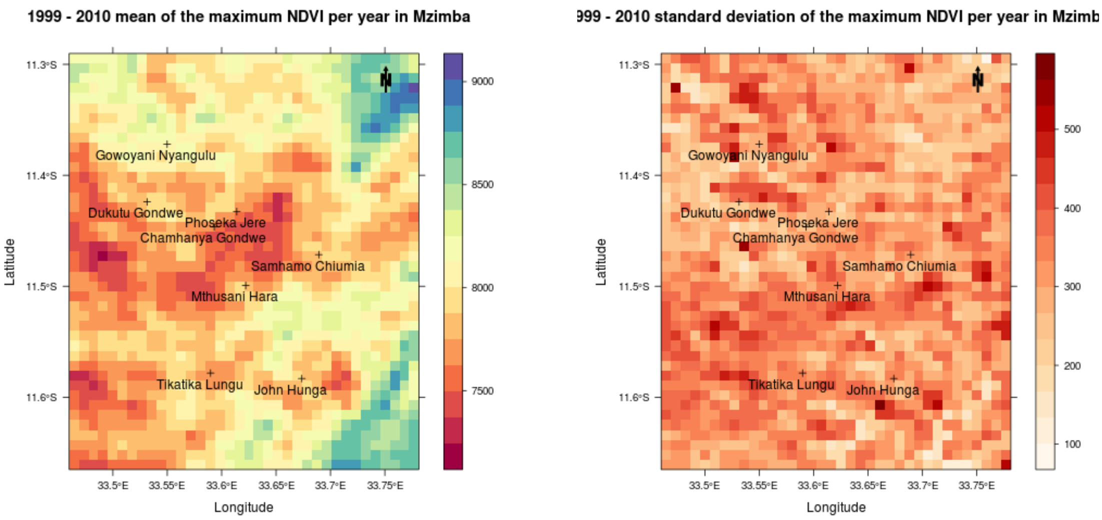


Figure 18: Detection of metrics

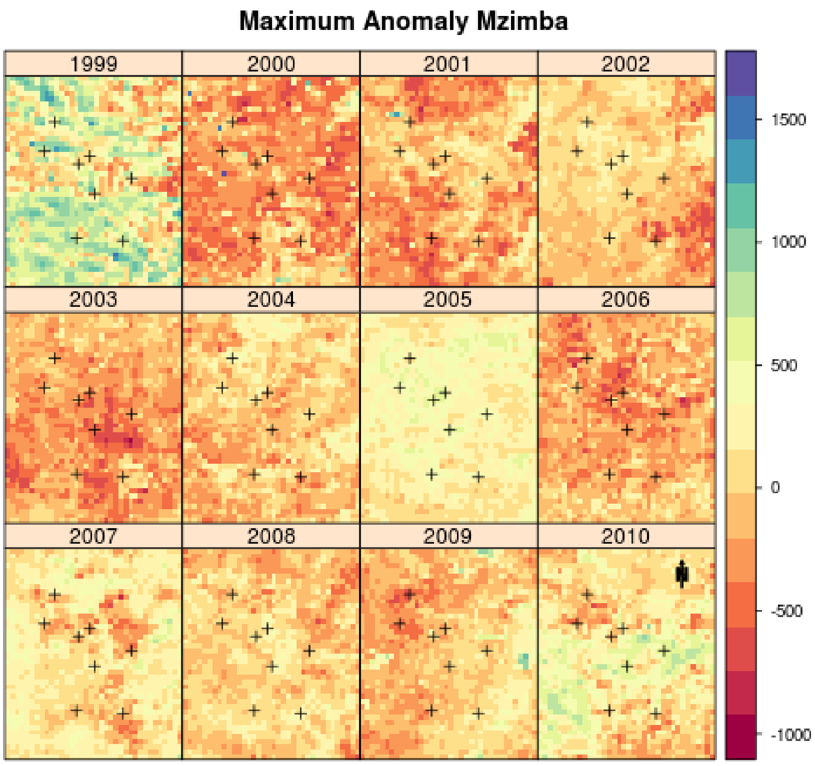


Figure 19: Mzimba c'est toi le roi

Table 3: Max per site and corrected year

	S. Chiumia	J. Hunga	T. Lungu	M. Hara	G. Nyangulu	D. Gondwe	P. Jere	C. Gondwe
2000	0.71	0.707	0.717	0.728	0.734	0.738	0.722	0.73
2001	0.681	0.73	0.692	0.7	0.722	0.713	0.674	0.675
2002	0.734	0.709	0.712	0.706	0.735	0.726	0.716	0.688
2003	0.72	0.723	0.692	0.723	0.712	0.751	0.739	0.71
2004	0.743	0.724	0.751	0.708	0.806	0.712	0.745	0.737
2005	0.775	0.773	0.766	0.775	0.795	0.825	0.802	0.774
2006	0.699	0.724	0.703	0.72	0.712	0.727	0.726	0.68
2007	0.743	0.758	0.728	0.717	0.727	0.737	0.724	0.732
2008	0.757	0.747	0.729	0.756	0.782	0.766	0.769	0.756
2009	0.726	0.72	0.695	0.718	0.713	0.7	0.748	0.697

Less information, seems less diverse spatially, 2005 less high but that's it while season information is distributed.

2.3.4 Mapping period anomaly

```
PeriodAnomaly(ndvidirectory, region, 1999, 2010, period=3:19, outname="Mzimba",
    outext="png", shapefile=shape, shapedir=shapedir, label="Village")
listname=paste("MzimbaP",1999:2010, ".tif", sep="")
multimap(listname, as.character(1999:2010), outname="MzimbaPeriod9910", org=c(4,3),
    title="Period Anomaly Mzimba", shapefile=shape, shapedir=shapedir, outext="png")
```

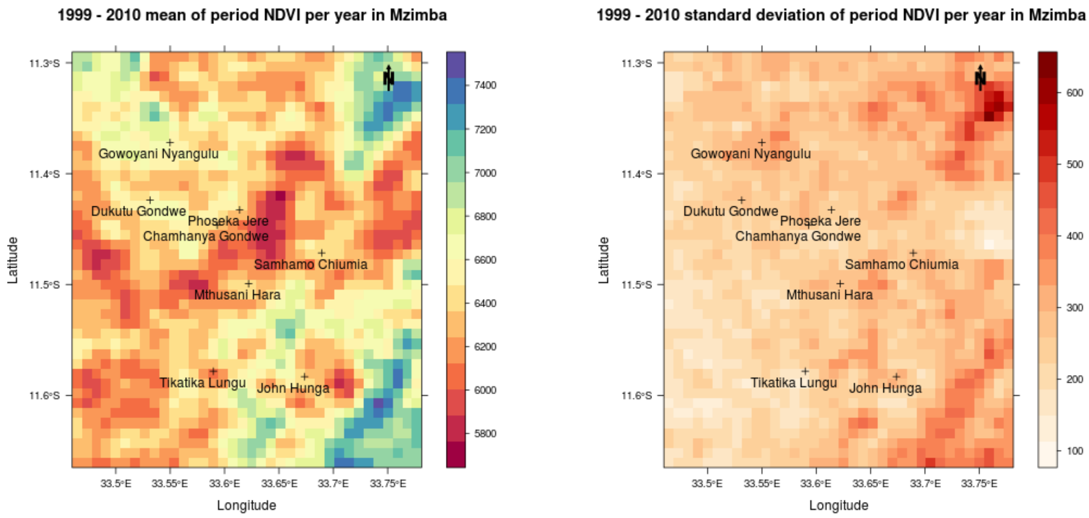


Figure 20: Mean (left) and SD (rigth) NDVI values on the period over the year

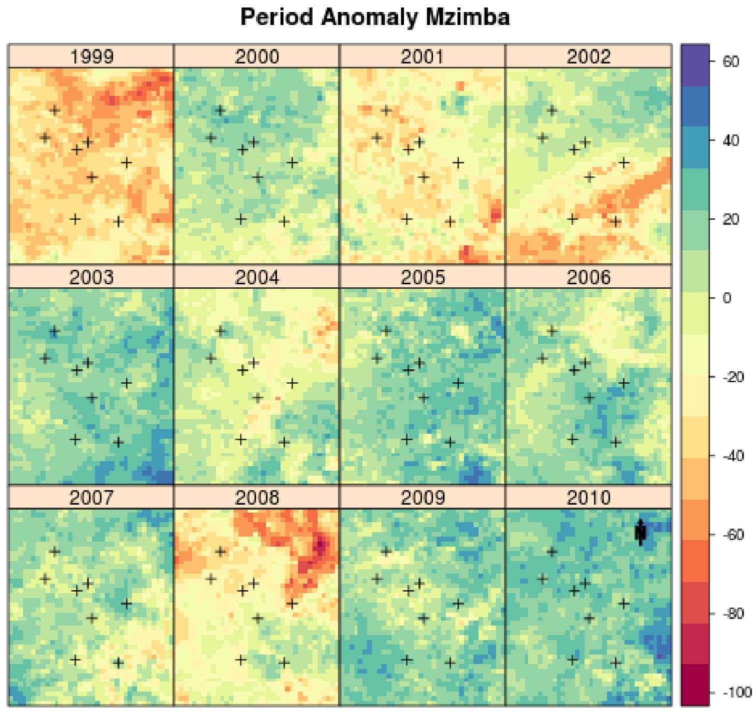


Figure 21: Period Anomaly in Mzimba

Table 4: cumNDVI per site and per year (corrected year)

	S. Chiumia	J. Hunga	T. Lungu	M. Hara	G. Nyangulu	D. Gondwe	P. Jere	C. Gondwe
2000	13.525	12.155	9.942	10.922	13.495	13.253	13.524	12.808
2001	12.055	6.084	7.472	10.649	11.547	11.882	11.677	11.391
2002	7.103	8.94	9.053	7.669	10.687	10.82	10.615	11.273
2003	12.695	14.954	11.156	12.282	12.917	13.364	12.83	12.498
2004	9.887	11.446	8.77	11.711	12.672	13.309	11.365	11.328
2005	11.654	11.609	7.462	11.045	12.567	11.442	10.899	11.105
2006	13.601	12.937	10.487	13.017	14.831	13.479	13.716	14.467
2007	10.694	10.932	8.491	10.071	11.836	12.074	10.049	9.382
2008	9.34	14.365	8.271	12.151	8.98	10.129	9.43	9.366
2009	12.064	12.695	10.718	13.091	11.975	11.877	11.396	10.902

For example 2002 as weel as 2008 is very interesting.

3 Discussion

Analying NDVI time series gives relevant information on the state of vegetation over a large area. This information is more and more used in climate change studies, biodiversity, wildlife ecology and agriculture.

`ndvits` package allow any scientist to extract easily informations from the free NDVI data. The phenological metrics are often reffered as Land Surface Phenology (LSP) in the litterature.

According to Ganguly et al. 2010 [7], phenology can be used as a diagnostic of ecosystem response to global change as well as a tool for study of habitat and biodiversity. Moreover, vegetation phenology is linked to terrestrial carbon cycle and shift is timing of leaf development (SOS) senescence (EOS) and changes in growing season length (LOS). are precious informations for agricultural and climate change studies.

monitor Land Surface Phenology (LSP) Satellite- based observations of phenology are often referred to as land surface phenology (LSP) because satellite sensors integrate land surface processes and the reflectance of electromagnetic radiation from vegetation over large areas Hanes2010

new indicator ? but metrics not very relevant in all the region. For sure it add a new information and a reliable esay and very cheap.

a better understanding of the vegetation of the studied area.

Computing variogram to study the spatial heterogeneity of the vegetation cover, knowing that crop sites are more heterogeneous than natural vegetation, forest. [8, 14]

"Growing importance of NDVI time series analyses in climate change, biodiversity, wildlife ecology" hird 2009 monitoring vegetation conditions used to study the biophysical features of vegetation monitoring tool for the vegetation health and dynamics : easy temporal and spatial comparison

"studying past and future changes in phenology and related ecosystem processes (e.g., water, energy, and carbon fluxes)."Hanes2010

assess the photo- synthetic activity of the land surface

Mapping Vulnerability to drought : NDVI based analysis leading to agricultural drought vulnerability "exposure, sensitivity and adaptive capacity to the reduced soil moisture availability in spatial and temporal dimensions" NNRMS : Remote sensing in Climate Change

NDVI integrated over growing period has been found to be an excellent measure to study the patterns of net primary productivity and crop production. Very strong relation between grain yields and time integrated NDVI has been widely reported (Quamby et al., 1993, Rasmussen 1992, Prasad et al., 2006) NNRMS

Information related to land surface phenology is important for a variety of applications. For example, phenology is widely used as a diagnostic of ecosystem response to global change. In addition, phenology influences seasonal scale fluxes of water, energy, and carbon between the land surface and atmosphere. Increasingly, the importance of phenology for studies of habitat and biodiversity is also being recognized. While many data sets related to plant phenology have been collected at specific sites or in networks focused on individual plants or plant species, remote sensing provides the only way to observe and monitor phenology over large scales and at regular intervals. Ganguly 2010

Because vegetation phenology affects terrestrial carbon cycling across a wide range of ecosystem and climate regimes. shifts in the timing of bud burst, leaf development, senescence, and changes in growing season length have been widely studied in the context of ecosystem responses to climate change Ganguly 2010

”Moderate resolution satellite remote sensing provides global high temporal frequency measurements of land surface properties and is therefore well suited for monitoring seasonal-to-decadal patterns and trends in regional-to-global phenology ”Ganguly 2010

”retroactively estimate past changes in seasonal vegetation growth and forecast future changes over large areas” Hanes2010

ouverture mixel : mixed pixel sensors integrate intra-pixel spatial heterogeneity information is lost high resolution Modis [7]

new indicators more precise

Modis LAI and LST product [9]

link with village survey data? PCA ?

4 Conclusion

The `ndvits` package is still under development. This beta-version is released and enable to make basics analysis. The user is encourage to combined the developed function with his own analysis.

While computing time series and analysing time series in R make sense, calculating maps is more challenging. However, R can handle this issue and the computing time is comparable with the majority of the GIS software. Nevertheless, the authors advice the use of quantumGIS to produce high quality maps.

5 Aknowledgements

This study was supported by the CGIAR Systemwide Livestock Programme. The autors are grateful to the Vlaamse instelling voor technologisch onderzoeck (VITO) and the Global Land Cover Facility (GLCF) for providing already processed NDVI dataset from satellite data.

References

[1] F. Baret, E. Bartholomé, P. Bicheron, G. Borstlap, L. Bydekerke, B. Combal, and J. Derwae. VGT4Africa User Manual. *[online]*, 2006.

[2] M. E. Brown, K. de Beurs, and A. Vrieling. The response of African land surface phenology to large scale climate oscillations. *Remote Sensing of Environment*, 114(10):2286–2296, Oct. 2010.

[3] M. E. Brown, J. E. Pinzón, K. Didan, J. T. Morisette, and C. J. Tucker. Evaluation of the Consistency of Long-Term NDVI Time Series Derived From AVHRR , and Landsat ETM + Sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7):1787–1793, 2006.

[4] J. Chen. 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, June 2004.

[5] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. STL: A seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statisitcs*, 6(1):3–73, 1990.

[6] S. Fritz, E. Bartholomé, A. Belward, A. Hartley, H. Eva, P. Mayaux, S. Bartalev, R. Latifovic, P. S. Roy, S. Agrawal, W. Bingfang, X. Wenting, J.-f. Pekel, C. Giri, S. Mücher, E. D. Badts, R. Tateishi, J.-l. Champeaux, and P. Defourny. Harmonisation , mosaicing and production of the Global Land Cover 2000 database. 2003.

[7] S. Ganguly, M. a. Friedl, B. Tan, X. Zhang, and M. Verma. Land surface phenology from MODIS: Characterization of the Collection 5 global land cover dynamics product. *Remote Sensing of Environment*, 114(8):1805–1816, May 2010.

[8] S. Garrigues, D. Allard, F. Baret, and J. Morisette. Multivariate quantification of landscape spatial heterogeneity using variogram models. *Remote Sensing of Environment*, 112(1):216–230, Jan. 2008.

[9] J. M. Hanes and M. D. Schwartz. Modeling land surface phenology in a mixed temperate forest using MODIS measurements of leaf area index and land surface temperature. *Theoretical and Applied Climatology*, Nov. 2010.

[10] R. Hegger, H. Kantz, and T. Schreiber. Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos*, 9(2):413–435, 1999.

[11] J. Hird and G. Mcdermid. Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*, 113(1):248–258, Jan. 2009.

[12] B. N. Holben. Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing*, 7(11):1417–1434, 1986.

[13] P. Jonsson and L. Eklundh. TIMESAT a program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, 30(8):833–845, Oct. 2004.

[14] Y.-P. Lin, H.-J. Chu, C.-L. Wang, H.-H. Yu, and Y.-C. Wang. Remote Sensing Data with the Conditional Latin Hypercube Sampling and Geostatistical Approach to Delineate Landscape Changes Induced by Large Chronological Physical Disturbances. *Sensors*, 9(1):148–174, Jan. 2009.

[15] C. J. Tucker. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2):127–150, 1979.

[16] C. J. Tucker, J. E. Pinzón, M. E. Brown, D. Slayback, E. Pak, R. Mahoney, E. Vermote, and N. Saleous. An Extended AVHRR 8-km NDVI Data Set Compatible with MODIS and SPOT Vegetation NDVI Data. *International Journal of Remote Sensing*, 26(20):4485–4498, 2005.

[17] C. Vancutsem, J. Pekel, C. Evrard, F. Malaisse, and P. Defourny. Mapping and characterizing the vegetation types of the Democratic Republic of Congo using SPOT VEGETATION time series. *International Journal of Applied Earth Observation and Geoinformation*, 11(1):62–76, Feb. 2009.

[18] J. Verbesselt, R. Hyndman, G. Newnham, and D. Culvenor. Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, 114(1):106–115, Jan. 2010.

[19] J. Verbesselt, R. Hyndman, A. Zeileis, and D. Culvenor. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment*, 114(12):2970–2980, Sept. 2010.

[20] H. Vinciková, M. Hais, J. Brom, J. Procházka, and E. Pecharová. Use of remote sensing methods in studying agricultural landscapes – a review. *Journal of Landscape Studies*, 3(April):53 – 63, 2010.

A Installation of R and the needed packages

R is free, open-source software that runs on Mac OS, Windows, and Linux platforms. Download links for all versions can be found at the official R Project website (`http://www.r-project.org`). After downloading R, you should install the program as appropriate for your operating system.

To run the entire functions, you also need to install the following packages :

`gstat, rgdal, RTisean, RColorBrewer`

You can also get ride of any nomenclature and just give a list of files with the option `type=FILES` while extracting time series.

The easiest way is to open the R console and write :

```
install.packages(c("gstat","rgdal","RTisean","RColorBrewer"))
```

But before using RTisean package (the Savitzky-Golay filter), you need to install the `TISEAN` executables.

1. download TISEAN-3.0.1 binaries from `http://www.mpipks-dresden.mpg.de/~tisean/Tisean\_3.0.1/index.Hml`
2. unpack them to a local folder

At the first call to a function in this package, you will be asked to specify the local directory where TISEAN binaries resides. Your answer will be stored in a settings file in your home directory, and used in all future functions calls.

For windows 7, download TISEAN_3.0.0-windows.zip sur le lien : `http://www.mpipksdresden.mpg.de/~tisean/windows\_3.0.0.html`.

B Nomenclature handle by `ndvits` functions

The function to extract NDVI time series (TimeSeriesAnalysis) has a new parameter: type which indicates the type of data (VITO, GIMMS) as explain in the following:

`ndvits` can handle three different nomenclatures. To facilitate explanation, the variable part of the names are between [ ].

- Our nomenclature : **type=VITO_CLIP**
  After a process of clipping the images in a particular region, we rename the files as following :
  `[REGION][YYYY]M[MM]P[D].tif`
  with :
  s `REGION` : the region as entered while clipping data.
  `YYYY` corresponging to the year (four digits)
  `MM` corresponging to the months (two digits)
  `D` corresponding to the period (0, 1 or 2).

- FreeVGT VITO's nomenclature : **type=VITO_VGT**
  `NDV_[YYYYMMDD]_[REGION]_Extract.tif`

with :
YYYY corresponging to the year (four digits)
MM corresponging to the months (two digits)
DD corresponging to the date of the period : ”01”, ”11” or ”21” (10-day mesures).
REGION: the region selected while extracting maps
NB: You can clip data using the function `ClipVGTVito` *and get you files will be renamed as explained.*

- GIMMS nomenclature (from Global Land Cover Facility) : **type=GIMMS**
  texttt[RE][YY][MMM][DDD].[SAT]-VIg.tif
  with:
  RE identifies the continent (AF for Africa,AZ for Australia and New Zealand, EA for Eurasia, NA for North America and SA for South America and Central America)
  YY corresponging to the year (two last digits).
  MMM abbreviation of the month (three letters).
  DDD identifies the composite period, 15a denotes the days 1-15 of the month, and 15b denotes the days from 16 to the end of the month.
  SAT identifies the satellite, NOAA-7, 9, 11, 14, 16, or 17 from which the data originated.

You can also get ride of any nomenclature and just give a list of files with the option `type=FILES` while extracting time series.

- **type = FILES** : `ndvidirectory` should contain a file respecting the following format :

  ```
  N
  path\image1.tif
  path\image2.tif
  ...
  path\imageN.tif
  ```

  with N the number of image files. The list of images should be in chronological order, and include the complete path to the file, unless the files are located in the working directory from which R is run. Only the N first images in the file will be processed.

- **type = TEXT** : `ndvidirectory` should contain a file respecting the following format : very usefull not to re-extract time series

C Customize the colors of the map

Every function making maps as output has a variable `pal` indicating the name of the palette of colors used to make the map. The palettes comes from the package `RColor-Brewer`.
By default, `pal = "Spectral"`.
You can customize your maps using the different palettes available (see Figure 22).

```
library(RColorBrewer)
display.brewer.all()
```

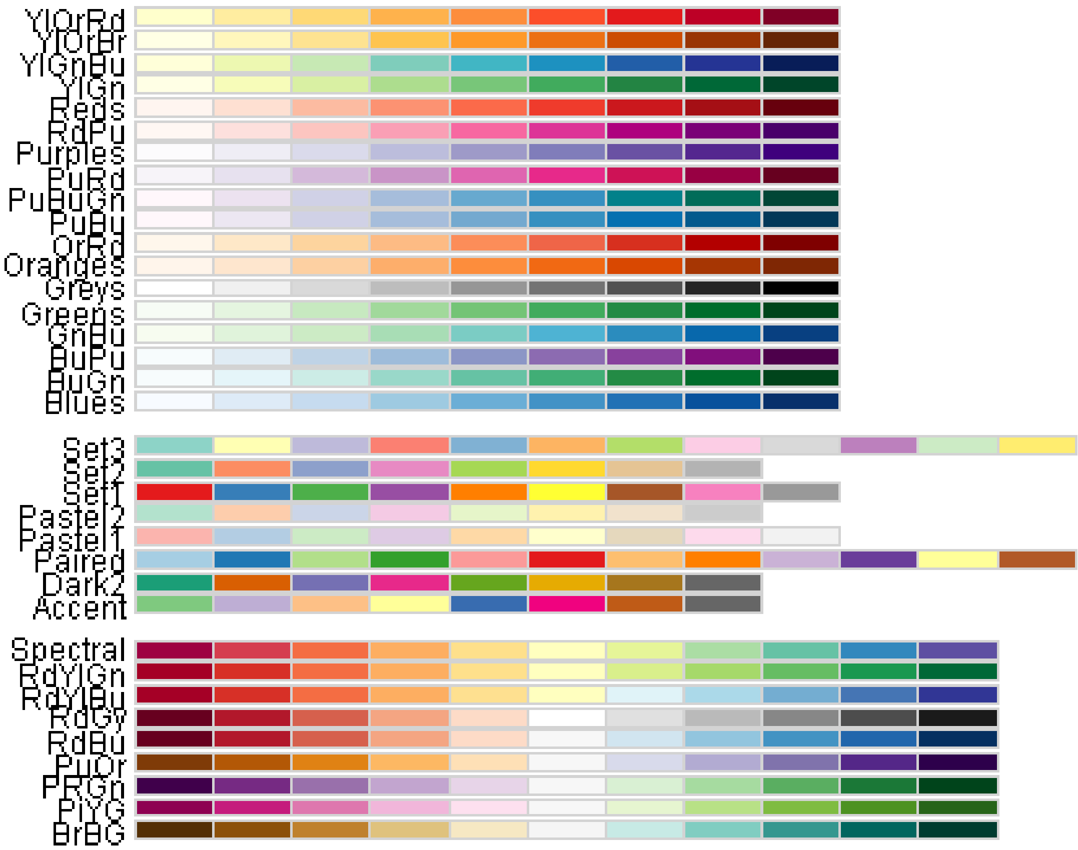


Figure 22: Colors available for the maps

NB : For an output which may be printed in black and white, it is convenient to use a palette with a single shade, for example, `pal = "Greys"`.

D Creating buffer from multipoints shapefile

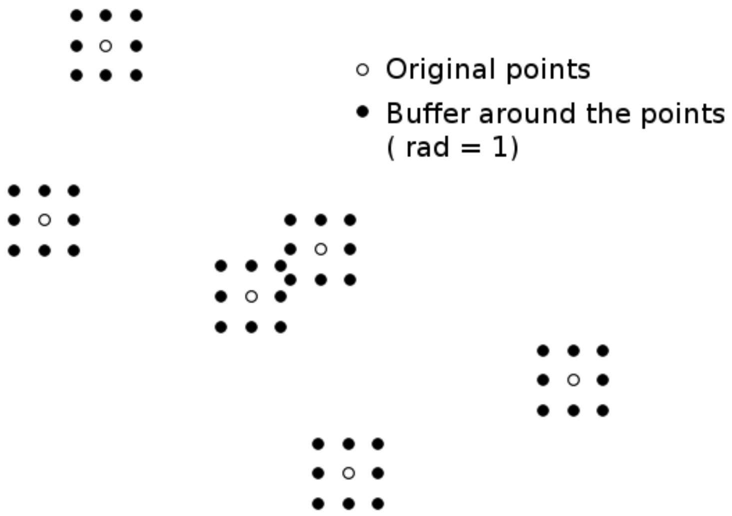


Figure 23: Creation of buffer with the function `pointtobuffer`

E Known problems and their solutions

- Getting error reading a "kml" : `shape` and `shapedir` may be not set correctly.
  `shape` is the name of the layer in the kml file (often the name of the file but can be something else : check in Google Earth)
  `shapedir` is the entire direction to the kml file + the kml file : `shapedir="D:Dir name.kml"`
  Be sure to specify `ext="kml"` in the function

- got the message : `"0 lines read.  It makes no sense to continue.  Exiting!"` You use a polygon which doesn't include any center of pixel.
  To understand this error, it is important to know that time series of polygons are extracted for the center of every pixels in the polygon (figure 3). Your error comes from your shapefile with a polygon without centers (figure 4).

  To solve this error, please use a dataset with a better spatial resolution or draw a bigger polygon.

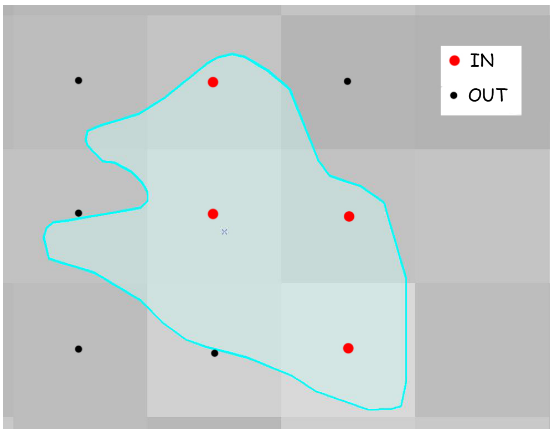


Figure 24: Example of how time series are extracted in a polygon

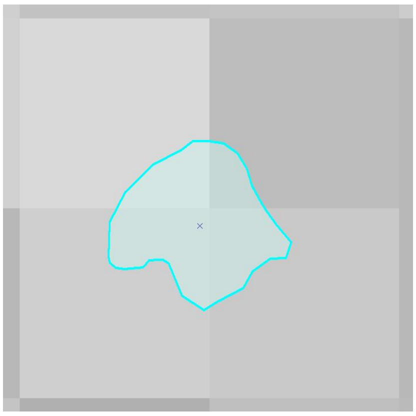


Figure 25: Polygon not matching with any center of pixels

- got the message "statistics not supported by this driver", what does that mean?
  It is a message from `readRGAL` function while loading the NDVI images. It has no effect on the analysis It just mean that you read data from an external hard drive but it doesn't affect the calculation of R . It is just a bit slower.

- A new error? Please ask Bruno Gerard[2] or Romain Frelat[3], we may have an answer for you and the error may appear soon in this list!

[2] b.gerard@cgiar.org
[3] rfrelat@yahoo.com