MODIStsp: an R package for automatic preprocessing of MODIS Land Products time series

L. Busetto[a,*], L. Ranghetti[a]

[a] *Institute for Electromagnetic Sensing of Environment (IREA-CNR) – Via Corti 12, Milano (IT)*

Abstract

MODIStsp is a new R package allowing automating the creation of raster time series derived from MODIS Land Products. It allows performing several preprocessing steps (e.g. download, mosaicing, reprojection and resize) on MODIS products on a selected time period and area. All processing parameters can be set with a user-friendly GUI, allowing users to select which specific layers of the original MODIS HDF files have to be processed and which Quality Indicators have to be extracted from the aggregated MODIS Quality Assurance layers. Moreover, the tool allows on-the-fly computation of time series of Spectral Indexes (either standard or custom-specified by the user through the GUI) from surface reflectance bands. Outputs are saved as single-band rasters corresponding to each available acquisition date and output layer. Virtual files allowing easy access to the entire time series as a single file using common image processing/GIS software or R scripts can be also created. Non-interactive execution within an R script and stand-alone execution outside an R environment exploiting a previously created Options File are also possible, the latter allowing scheduling execution of MODIStsp to automatically update a time series when a new image is available. The proposed software constitutes a very useful tool for the Remote Sensing community, since it allows performing all the main preprocessing steps required for the creation of time series of MODIS data within a common framework, and without requiring any particular programming skills by its users.

Keywords: MODIS, R, time series, preprocessing

1. Introduction

Time series of coarse resolution satellite images (e.g. spatial resolution between 0.1 and 1 km) are currently widely used for monitoring several characteristics of the earth surface. Among their main applications, we can cite the study of vegetation phenology (e.g. Jönsson and Eklundh, 2002; Busetto et al., 2010; Verbesselt et al., 2010), land use/cover mapping and change detection (e.g. Defourny et al., 2007; Klein et al., 2012), the analysis of post-fire vegetation dynamics (e.g. Lhermitte et al., 2007; Gitas et al., 2012), ecological

*Corresponding author
Email addresses: busetto.l@irea.cnr.it (L. Busetto), ranghetti.l@irea.cnr.it (L. Ranghetti)

applications relating animal movement with remotely sensed data (e.g. Bartlam-Brooks et al., 2013), and many others. This research field received a strong boost since the mid 2000's thanks to the availability of free of charge and easy to access data acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) multispectral sensors on-board of NASA's TERRA and AQUA satellites.

MODIS has a viewing swath of 2,330 km and views the entire surface of the Earth every one/two days. Its data is routinely used to derive several products related to radiation budget (e.g. surface reflectance, land surface temperature, albedo), ecosystem variables (e.g. vegetation indexes, leaf area index) and land characteristics (e.g. land cover, thermal anomalies).[1] MODIS Land Products are produced either daily or as temporal composites with different aggregation windows (from 8 days to yearly), and at four nominal spatial resolutions (250, 500, 1000, and 5600 meters (0.05°)). Data at 250, 500 and 1000 m resolution are distributed in adjacent non-overlapping tiles approximately 10 degrees square at the equator to maintain reasonable file sizes. Geographic projection is used for the 5600 m resolution products, while Sinusoidal projection is used for all the others.

Available MODIS Land Products can be searched and downloaded through NASA's REVERB on-line metadata science discovery tool (EOSDIS, 2009) or other on-line tools. Direct access to single datasets is also possible through the LP DAAC on-line Data Pool (`ftp://ladsweb.nascom.nasa.gov/allData/5`). Data corresponding to each acquisition date (or compositing period) and tile is provided as a multiband HDF5 (Hierarchical Data Format) file including several layers, each corresponding to a specific variable (e.g. surface reflectance, acquisition angles, albedo). Additionally, Quality Indicators (QI) (e.g. data acquisition quality, cloud/snow presence) are stored in one or more Quality Assurance (QA) layers using a bit-field representation allowing to store information on several QIs in a single QA layer.

Although very efficient in terms of ease of access and reduction of required storage space, this distribution scheme requires MODIS users to perform several preprocessing steps on the original HDF images to extract the specific data useful for their time series analysis. Typical preprocessing steps include:

1. downloading products as single-date HDF files from the http or ftp distribution archives, for specific dates and tiles;
2. mosaicing, resizing and eventually reprojecting the raster images;
3. extracting the specific layers of interest for the analysis;
4. computing additional layers from the original ones (i.e. extract specific QIs from QA layers or compute Spectral Indexes (SI) from surface reflectance);
5. storing and organizing data so to allow easy access to time series of a given variable/parameter, eventually converting it in an easier-to-use format.

While technically simple, manual execution of these tasks can be rather bothersome, time consuming and

[1] A complete list of available MODIS Land Products can be found at `https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table`.

prone to errors. For this reason, several solutions were developed in the last few years to automatize one or more of the aforementioned steps. These solutions can be roughly categorized in two classes: i) web-based solutions and ii) open-source software scripts or libraries. Among the first category, the MODIS Reprojection Tool (MRT) Web application (`https://mrtweb.cr.usgs.gov`) allows to easily download datasets related to a given temporal and spatial extent, reproject them and saving the results as GeoTiff, HDF or binary files. These services are provided also in the REVERB web-based application (`http://reverb.echo.nasa.gov/reverb`), the NASA metadata and service discovery tool.

While these solutions provide a ready-to-use way to obtain the desired data, they lack the possibility to save the used settings (a fundamental option to schedule batch processing jobs and make analyses reproducible), and allow processing only a limited number of images for each request. To overcome some of these limitations, NASA provides also two off-line cross-platform tools: the java-based MRT stand-alone desktop application (`https://lpdaac.usgs.gov/tools/modis_reprojection_tool`), with the same functionalities of the web application (except the download) and the possibility to store settings in a text file and use it for batch operations, and the HDF-EOS To GeoTIFF Conversion Tool (HEG, `http://newsroom.gsfc.nasa.gov/sdptoolkit/HEG/HEGHome.html`). None of the aforementioned solutions allow however to flexibly extract QA information from MODIS hdf files and to compute Spectral Indexes from reflectance data. Additionally, their resizing and reprojection capabilities are somehow limited.

Third party open-source tools have been also developed in order to add improved functionalities. PyModis (`http://pymodis.fem-environment.eu`) is a collection of Python scripts taking advantage of MRT and GDAL (Geospatial Data Abstraction Library). Compared to MRT and HEG, pyModis also allows to extract QIs (although the user must manually specify the name of the QI of interest or its position within QA layers). PyModis scripts (downloading, mosaicing, converting, extracting quality information) must be run separately by the user and no GUI for setting processing parameters is available.

Several R scripts and libraries are also available. *MODISTools* (Tuck et al., 2014) is a set of functions mainly devoted to extract MODIS time series over specific points and compute summary statistics over time. The main advantage of this package is that it allows users to simply specify a list of coordinates for which MODIS time series are downloaded and saved as simple ASCII files, thus avoiding the need to download and store huge amounts of data as raster files in the case that the analyst is only interested in analyzing specific small locations. However, no spatial processing functions are available and only a limited set of MODIS products is supported.

ModisDownload (Naimi, 2014) is an R script (without GUI) which mirrors MRT functionalities and can be useful for including MODIS preprocessing operations in R applications.

Finally, a more recent R package (*MODIS*; Mattiuzzi, 2016) offers a wider range of functionalities, allowing to download HDF files, preprocess them using MRT or GDAL and extract QI information. It also allows generating time series combining images in a *RasterStack* object (Hijmans, 2015). Its main lacks are the absence of a GUI, which makes the definition of processing parameters somehow difficult, and the lack of functionalities for automating the computation of SIs from surface reflectance data.

While providing most of the functionalities required for an automatic preprocessing chain of MODIS time series, the aforementioned solutions require usually at least some basic programming skills, since they don't provide a user-friendly interface for definition of processing parameters and they don't allow to perform a batch processing without writing at least some lines of code (i.e., from outside the R environment). Moreover, most of them lack the ability to extract and create time series of QIs and none allow to our knowledge to flexibly create time series of SIs from MODIS reflectance layers.

In this manuscript, we present the *MODIStsp* (MODIS Time Series preprocessing) Tool, a new R package devoted to automating the creation of raster time series for variables or parameters derived from MODIS Land Products. Development of *MODIStsp* started from modifications of the `ModisDownload` R script by Hengl (2010), and successive improvements by Naimi (2014). The basic functionalities for download and preprocessing of MODIS data provided by these scripts were gradually enhanced with the aim of developing an application able to perform all the above-mentioned MODIS preprocessing tasks, exploiting a user-friendly GUI interface for parameters selection and providing an easy solution for batch processing to allow also users without any software programming skill to automatically create and constantly update MODIS time series.

For each layer selected by the user, output time series are saved as single-band rasters corresponding to each available acquisition date and resized/reprojected on the required study area. Virtual files facilitating access to the entire time series from ENVI or other image processing/GIS software can be also created. Additionally, output time series can be saved as R *RasterStack* objects including temporal information, allowing easy access and analysis of the preprocessed time series from within R scripts. A comparison between *MODIStsp* functionalities and those of the aforementioned available solutions for MODIS preprocessing is shown in Figure 1.

Users may exploit *MODIStsp* to prepare datasets to be used to perform multitemporal analysis on MODIS data from within R or other GIS/Image processing software or scripting languages. In particular, *MODIStsp* *RasterStack* outputs can be easily used as inputs for functions of other R packages dealing with visualization and analysis of satellite time series datasets. Among others, packages such as *rasterVis* (Perpiñán and Hijmans, 2014) - which provides useful functionalities for visualization and temporal aggregation of multitemporal raster data -, *plotKML* (Hengl et al., 2015) - which allows visualization of satellite time series from within GoogleEarth -, or *bfastSpatial* (Dutrieux and DeVries, 2014) - which performs change detection analysis on time series of MODIS and Landsat imagery - could directly exploit *MODIStsp* outputs, allowing the development of full-fledged analysis solutions (from data discovery and download to final analysis).

Preprocessing Solution	GUI	Batch processing	Download	Mosaic	Resize	Reproject	Resampling (resolution)	Resampling (method)	Extract specific HDF layers	Extract QI data	Compute Spectral Indexres	Convert format	Virtual Time series files
NASA REVERB	V	X	V	V	V only geographic	V only some projections	V only some resolutions	V NN, Cubic, Bilinear	V	X	X	V HDF-EOS; GeoTiff; Raw binary	X
MRT /MRTweb	V	V not in web	X	V	V only geographic	V only some projections	V only some resolutions	X	V	X	X	V HDF-EOS; GeoTiff; Raw binary	X
HEG	V	V	X	V	V only geographic	V only some projections	V only some resolutions	V NN, Cubic, Bilinear	V	X	X	V HDF-EOS; GeoTiff	X
pyModis	X	V	V	V	?	V gdal	V	V	?	Z	X	V GeoTiff	X
MODISTools	X	V	V Few products	X	V	X	X	X	X	X	X	X	V Only on punctual extracted data
ModisDownload	X	V	V	V MRT	X	V MRT	V	V NN, Cubic, Bilinear	V	X	X	V GeoTiff	X link to rts package
MODIS R	X	V	V	V MRT	V	V gdal or MRT	V	V	?	V	X	V	V "R" RasterStack
MODIStsp	V	V	V	V gdal	V	V gdal	V	V NN, mode	V	V	V	V ENVI, GeoTiff	V ENVI, gdal vrt, RasterStack

Figure 1: Main functionalities of *MODIStsp*, compared to those of already available MODIS preprocessing solutions

2. Description of the software

2.1. Installation and dependencies

MODIStsp has been developed completely in the R Language and Environment for Statistical Computing (2015) v. 3.1.3 and is distributed as open source software under the GNU-GPL 3.0 License. Source code can be downloaded at the GitHub repository `https://github.com/lbusett/MODIStsp`. It imports functionalities of several additional R packages (Table 1), which are automatically added to the users' R library during installation. Additionally, it requires availability of GDAL v. > 1.10 (Open Source Geospatial Foundation, 2015), with support for HDF4 raster format.

The package can be easily installed following instructions provided in Supplementary Materials or in the main GitHub page. It was tested on Windows®, Linux® and MacOSX® operating systems.[2]

[2] MODIStsp users are encouraged to submit bug reports and feature requests in the issues section of the MODIStsp GitHub page, at `https://github.com/lbusett/MODIStsp/issues`.

Table 1: List of R packages imported by MODIStsp.

R package	Version	Authors	Usage
bitops	$\geq$ 1.0-6	Dutky and Maechler, 2013	Extraction of QI values from QA bit-field layers
data.table	$\geq$ 1.9.6	Dowle et al., 2015	Aggregation of raster data over spatial features in `MODIStsp_extract`
gdalUtils	$\geq$ 2.0.1.7	Greenberg and Mattiuzzi, 2015	Resize, resample and reprojection of rasters
gWidgetsRGtk2	$\geq$ 0.0-83	Lawrence and Verzani, 2013	GUI creation and management
hash	$\geq$ 2.2.6	Brown, 2013	Creation and use of dictionaries for *MODIStsp* options
plyr	$\geq$ 1.8.3	Wickham, 2015	Easy factors releveling
raster	$\geq$ 2.5-2	Hijmans, 2015	SI and QI computation and saving
RCurl	$\geq$ 1.95-4.8	Lang and the CRAN team, 2016	Download of original MODIS hdf files from NASA server
rgdal	$\geq$ 1.1-8	Bivand et al., 2016	Processing on ESRI shapefiles
rgeos	$\geq$ 0.3-8	Bivand and Rundel, 2016	CRS setting an checking
xts	$\geq$ 0.9.874	Ryan and Ulrich, 2015	Creation of xts objects as outputs of `MODIStsp_extract`
sp	$\geq$ 1.2-2	Pebesma and Bivand, 2005	Spatial operations on vector files and extents
stringr	$\geq$ 1.0.0	Wickham, 2015	Simple string elaborations on file names
XML	$\geq$ 3.98-1.1	Lang and the CRAN Team, 2016	Reading MODIS products characteristics from the `MODIStsp_ProdOpts.xml` file

2.2. Selection of processing options

After installing and loading the package, launching the `MODIStsp` function without additional parameters opens a user-friendly GUI for the selection of processing options required for the creation of the desired MODIS time series (Figure 2). After each successful execution the selected files and parameters are saved, and at the successive execution the GUI automatically retrieves the last used parameters. All processing options can be saved to a users' selected file for later use by pressing the "Save Options" button at the bottom of the GUI. The "Load Options" button allows retrieving previously saved options.

The GUI is organized in seven frames, which allow specifying processing options related to different aspects.

MODIS Product, Satellites and Layers selection: Allows selecting the MODIS product of interest from a drop-down menu, and the MODIS platform to be considered for creation of the time series (TERRA, AQUA or both). After selecting the product, the user can select the MODIS original, QI and SI layers to be processed by pressing the "Select Layers" button, which opens a separate layers' selection panel (Figure 3). The left-hand frame of this panel allows selecting which original MODIS layers should be processed, while

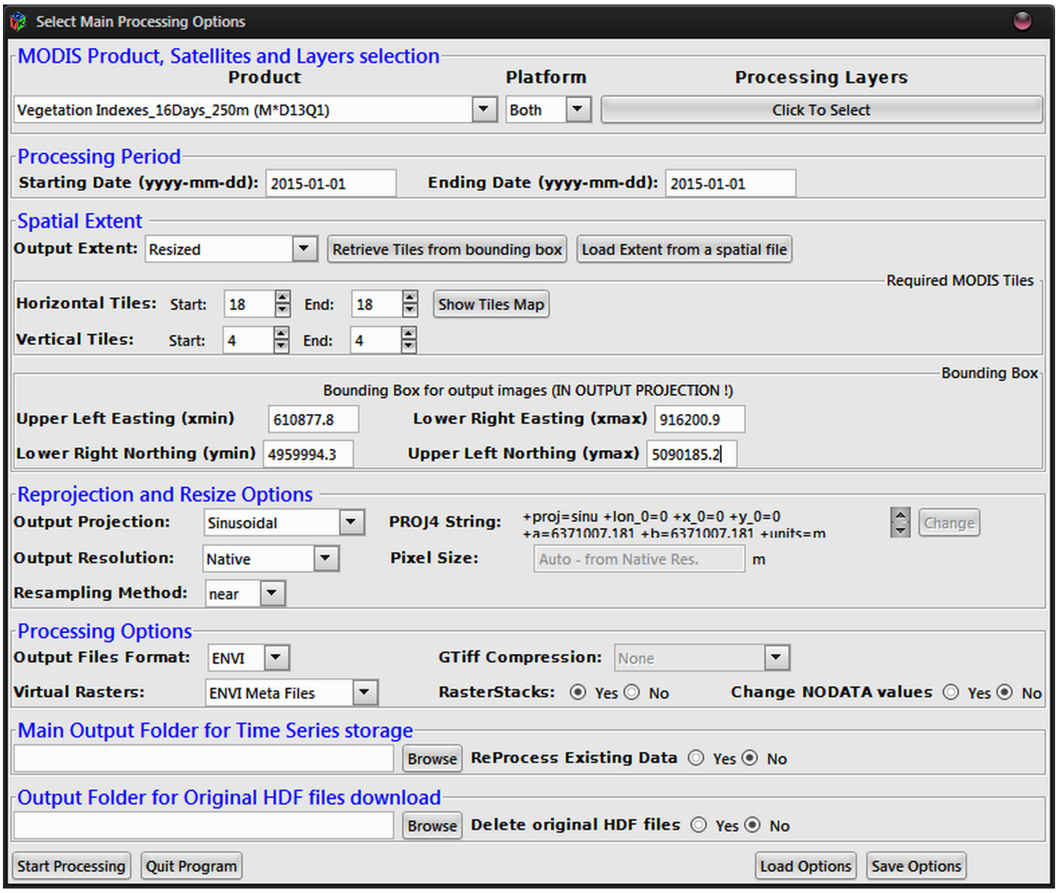

Figure 2: The Graphical User Interface of the MODIStsp Tool

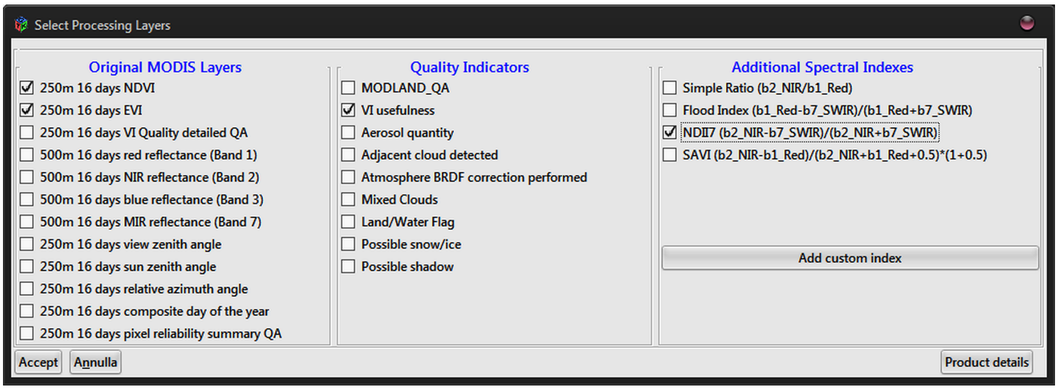

Figure 3: Interface for selection of the MODIS original layers, QIs and SIs for the MOD13Q1 product.

the central frame allows selecting which QIs should be extracted from the QA layers. Finally, for MODIS products containing surface reflectance data, the right-hand frame allows selecting which SIs should be computed. The lists of original MODIS layers, QIs and SIs available for the selected product are retrieved from the `MODIStsp_ProdOpts.xml` XML file distributed with the package. Some of the most common SIs are available for computation by default (Table 2), but users can add custom SIs without modifying *MODIStsp* source code by clicking on the "Add Custom Index" button, which allows specifying the formula of the additional desired SI using a simple GUI interface. The new index is then automatically added to the selection list for all products for which it can be computed.

Table 2: Default SIs available for computation in *MODIStsp*

Acronym	Index name and reference	Index Formula
NDVI	Normalized Difference Vegetation Index (Rouse et al., 1973)	$\frac{b2_{\text{NIR}} - b1_{\text{Red}}}{b2_{\text{NIR}} + b1_{\text{Red}}}$
EVI	Enhanced Vegetation Index (Huete et al., 2002)	$\frac{2.5 * b2_{\text{NIR}} - b1_{\text{Red}}}{b2_{\text{NIR}} + 6 \cdot b1_{\text{Red}} - 7.5 \cdot b3_{\text{Blue}}}$
SR	Simple Ratio (Tucker, 1979)	$\frac{b2_{\text{NIR}}}{b1_{\text{Red}}}$
NDFI	Normalized Difference Flood Index (Boschetti et al., 2014)	$\frac{b1_{\text{Red}} - b7_{\text{SWIR}}}{b1_{\text{Red}} + b7_{\text{SWIR}}}$
NDII7	Normalized Difference Infrared Index – band 7 (Hunt and Rock, 1989)	$\frac{b2_{\text{NIR}} - b7_{\text{SWIR}}}{b2_{\text{NIR}} + b7_{\text{SWIR}}}$
SAVI	Soil Adjusted Vegetation Index (Huete, 1988)	$1 + 0.5 \cdot \frac{b2_{\text{NIR}} - b1_{\text{Red}}}{b2_{\text{NIR}} + b1_{\text{Red}} + 0.5}$
NDSI	Normalized Difference Snow Index (Hall et al., 2002)	$\frac{b4_{\text{Green}} - b6_{\text{SWIR}}}{b4_{\text{Green}} + b6_{\text{SWIR}}}$
NDII6	Normalized Difference Infrared Index – band 6 (Hunt and Rock, 1989)	$\frac{b2_{\text{NIR}} - b6_{\text{SWIR}}}{b2_{\text{NIR}} + b6_{\text{SWIR}}}$
GNDVI	Green Normalized Difference Vegetation Index (Gitelson and Merzlyak, 1998)	$\frac{b2_{\text{NIR}} - b4_{\text{Green}}}{b2_{\text{NIR}} + b4_{\text{Green}}}$
RGRI	Red Green Ratio index (Gamon and Surfus, 1999)	$\frac{b1_{\text{Red}}}{b4_{\text{Green}}}$
GRVI	Green-red ratio vegetation index (Gamon and Surfus, 1999)	$\frac{b1_{\text{Red}} - b4_{\text{Green}}}{b1_{\text{Red}} + b4_{\text{Green}}}$

To facilitate the selection, clicking the "Product Details" button opens a page on the MODIS website providing information on the selected MODIS product and its hdf layers. Finally, the "Start" or "Cancel" buttons send back to the main GUI, either accepting on ignoring the selections.

Temporal Extent: Allows specifying the starting and ending dates to be considered for the creation of the time series.

Spatial Extent: Allows defining the area of interest for the processing. Two options are possible:

1. **Full Tiles Extent**: the user specifies the MODIS tiles to be processed using the "Start" and "End" horizontal and vertical sliders in the "Required MODIS Tiles" frame. A single file covering the total area of the specified tiles is produced for each acquisition date as output.

2. **Resized**: the user specifies the spatial extent of the desired outputs either:
   - by manually entering the coordinates of the Upper Left and Lower Right corners in the "Bounding Box" frame;
   - by clicking the "Load Extent from a Spatial File" button and selecting a raster or vector file. In this case, the bounding box of the selected file is retrieved, converted to the output projection and shown in the "Bounding Box" frame. Required input MODIS tiles are automatically retrieved on the basis of the output extent, and tiles' selection sliders modified accordingly.

Reprojection and Resampling: Allows specifying the options to be used for reprojecting and resampling the MODIS images. In particular:

1. the *Output Projection* can be selected among some predefined values, or manually specified by selecting "User Defined" and entering a valid "Proj4" string in the input dialog box that appears; [3]
2. the *Output Resolution*, *Pixel Size* and *Reprojection Method* menus allow specifying if output images inherit spatial resolution from the original MODIS files, or are resampled to a user-defined resolution. In the latter case, the output spatial resolution must be specified in the measure units of the selected output projection. The resampling method can instead be chosen among Nearest Neighbor and Mode (useful for downsampling purposes). Other resampling methods (e.g., bilinear, cubic) are not currently supported since i) they cannot be used for resampling categorical variables such as the QA and QI layers, and ii) using them on continuous variables (e.g., reflectance, VI values) without performing an a-priori data cleaning would risk to contaminate the values of high-quality observations with those of low-quality ones.

Processing Options: Allows first of all specifying the desired output format. Two of the most commonly formats used in remote sensing applications are available at the moment: ENVI binary and GeoTiff. If GeoTiff is selected, the type of file compression can be specified among "None", "PACKBITS", "LZW" and "DEFLATE". The user can also specify if virtual multitemporal files should be created. Available virtual files formats are ENVI metafiles and GDAL "vrt" files. These virtual files allow access to the entire time series of images as a single file without the need of creating large multitemporal raster images: this proves useful when time series span multiple years or cover large areas. Moreover, virtual files can be easily updated whenever a new image is available, without the need to delete and recreate a huge raster file.

Additionally, the user can select if he desires to save the time series also as R *rasterStack* objects (with temporal information added through the "setZ" method of the *raster* package). This may be useful in order to easily access the preprocessed MODIS data within R scripts.

[3] Whenever the selected MODIS product is changed, the output projection is automatically reset to the native one (Sinusoidal or LatLon according to Product).

Finally, users can select if MODIS NoData values should be kept at original values or changed. Dealing with MODIS NoData can in fact be troublesome since they don't follow a clear standard (for example, NoData value for Band 1 of the MOD09A1 product is set to –28867, while that of MCD43B4 to 32767). By selecting "Yes" in the "Change Original NoData values" check-box, NoData of outputs are set to the largest integer value possible for the data type of each layer (e.g. for 8-bit unsigned integer layers, NoData is set always to 255, for 16-bit signed integer to 32767, and for 16-bit unsigned integer to 65535). Information about the new NoData values is stored both in the output rasters, and in the associated XML files.

Main Output Folder for Time Series Storage: Allows specifying the main folder where time series data will be stored. The "Reprocess Existing Data" check-box allows specifying if images already available should be reprocessed if a new run of *MODIStsp* is launched with the same output folder. If set to "No", *MODIStsp* skips dates for which output files following *MODIStsp* naming conventions are already present. This allows incrementally extending MODIS time series when new data are available, without reprocessing the already processed dates.

Output Folder for Original HDF Storage: Allows specifying the folder where original downloaded MODIS HDF files are stored. The "delete original HDF files" check-box allows specifying if downloaded images must be deleted from the file system at the end of the processing.

2.3. Processing

Upon pressing the "Start" button, *MODIStsp* performs the following main tasks.

1. Retrieve the processing options from the GUI (or the saved RData file in case of non-interactive execution).

2. Retrieve the list of images available for the selected product in the selected time extent, for each tile required to cover the output extent.

3. For each date of acquisition:

    a - download all required hdf images;

    b - for each original hdf layer selected by the user *or* required to compute a selected QI or SI, extract the data from the original MODIS images and resize, reproject and resample it according to processing options. If more than one tile is needed to cover the output extent, virtual mosaics are created before resizing using `gdalbuildvrt` functionalities to avoid creation of large temporary raster files. All the main spatial processing tasks are performed using standard GDAL routines, exploiting R wrappers provided by the *gdalUtils* package (Greenberg and Mattiuzzi, 2015). Results are saved as GeoTiff or ENVI files with *MODIStsp* naming conventions;

c - starting from files created at point b), compute QIs and SIs and save results as GeoTiff or ENVI files. QIs are computed from QA layers using fast bitwise operators available in the *bitops* package (Dutky and Maechler, 2013), using a generalization of the `modis.qc.R` script by Chemin, 2008. Computation of SIs exploits on-the-fly parsing of the indices's' formulae to identify the required input raster files and perform the computation;

d - delete files created at point b) that were required to compute the QI or SI layers but correspond to HDF layers not required by the user.

4. When all dates have been processed, create the virtual raster time series and *RasterStack* RData files if required.

A flow chart of the main processing steps is shown in Figure 4.

2.4. Non Interactive Execution

The `MODIStsp` function can be launched in non-interactive mode by setting the optional `gui` argument to FALSE, and the `options_file` argument to the name of a previously saved Options File. In this case, the GUI is not opened, and processing is executed according to the saved parameters.

This allows exploiting *MODIStsp* functionalities within generic R processing scripts. Specifying also the `spatial_file_path` optional parameter overrides the output extent of the selected Options File, allowing performing the same preprocessing on different extents using a single Options File and looping on an array of spatial files representing the desired output extents (See Supplementary Materials for further details).

2.5. Standalone execution and scheduled processing

Differently from all the other actually existing R packages for MODIS preprocessing, *MODIStsp* was designed so to allow its execution as a standalone application (i.e., outside of an R session), using the `MODIStsp.bat` (Windows) or `MODIStsp.sh` (Linux) launchers available in the `ExtData/Launcher` subfolder of package installation. Double-clicking the files or launching them from a shell starts *MODIStsp* in interactive mode. Non-interactive mode is triggered by adding the `-g` argument to the call, and specifying the path to a valid Options File as `-s` argument.

This feature allows also scheduling a standalone *MODIStsp* non-interactive execution to automatically update MODIS time series of a selected product whenever a new image is available. To do that, the user should simply:

1. open the *MODIStsp* GUI, define the parameters of the processing specifying a date in the future as the "Ending Date", save the processing options and quit the program;

2. schedule non-interactive execution of `MODIStsp.bat` (or `MODIStsp.sh`) as Windows scheduled task (or Linux cron job) according to a specified timing, specifying the path of the saved Options File as `-s` argument (detailed instructions are provided in Supplementary Materials).

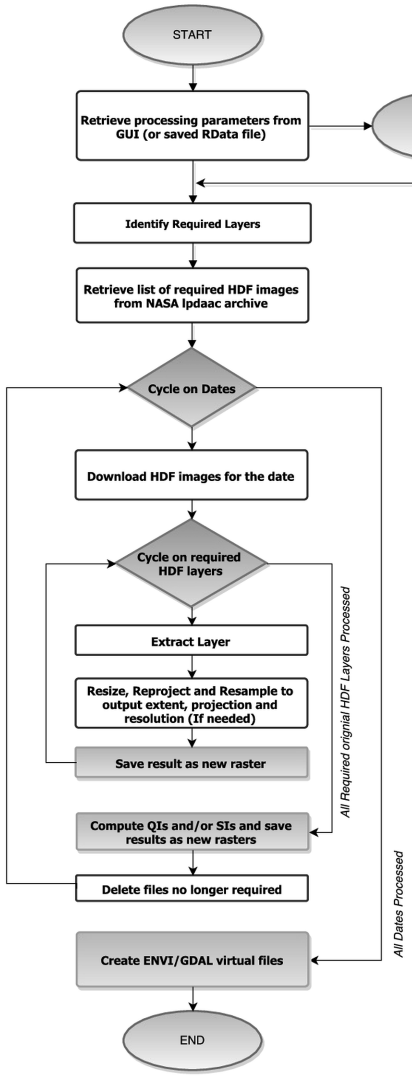

Figure 4: Flow Chart of main processing steps of *MODIStsp*

3. Output Format and Naming Conventions

Output raster files are saved in specific subfolders of the main output folder. A separate subfolder is created for each required original, QI or SI layer. Each subfolder contains one image for each processed date, with the following naming conventions:

ProdCode_Layer_YY_DOY.ext (e.g. `MOD13Q1_SR_2000_065.dat`)

, where *ProdCode* is the code name of the MODIS product from which the image was derived (e.g. MOD13Q1), *Layer* is the short name of the original or derived layer (e.g. b1_Red, EVI, UI), *YY* and *DOY* are the year and DOY (Day of the Year) of acquisition, while *ext* is the file extension (`.tif` for GTiff outputs, or `.dat` for ENVI outputs).

ENVI and/or GDAL virtual time series files and *RasterStack* RData objects are instead stored in the `Time_Series` subfolder if required. Naming convention for these files is as follow:

ProdCode_Layer_StartDOY_StartYear_EndDOY_EndYear_suffix.ext

(e.g. `MOD13Q1_NDVI_49_2000_17_2015_RData.RData`)

, where *suffix* indicates the type of virtual file (`ENVI`, `GDAL` or `RData`), while *StartDOY*, *StartYear*, *EndDOY* and *EndYear* indicate the temporal extent of the time series created.

4. Accessing and analyzing the processed time series from R

Preprocessed MODIS data can be retrieved within R scripts either by accessing the single-date raster files, or by loading the saved *RasterStack* objects. This second option allows accessing the complete data stack and analyzing it using the functionalities for raster/raster time series analysis, extraction and plotting provided for example by the *raster* (Hijmans, 2015) or *rasterVis* (Perpiñán and Hijmans, 2014) packages. *MODIStsp* provides however an efficient function (`MODIStsp_extract`) for extracting time series data at specific locations. The function takes as input a *RasterStack* object with temporal information created by *MODIStsp*, the starting and ending dates for the extraction and a standard R *Sp** object (Pebesma and Bivand, 2005) (or an ESRI® shapefile) specifying the locations (points, lines or polygons) of interest, and provides as output a R *xts* object (Ryan and Ulrich, 2015) containing time series for those locations. If the input *Sp** is of class *SpatialPoints*, the output object contains one column for each point specified, and one row for each date. If it is of class *SpatialPolygons* (or *SpatialLines*), it contains one column for each polygon (or each line), with values obtained applying the function specified as the `FUN` argument (e.g. mean, standard deviation, etc.) on pixels belonging to the polygon (or touched by the line), and one row for each date.

As an example the following code:

```
#Set the input paths to raster and shape file
infile = 'in_path/MOD13Q1_MYD13Q1_NDVI_49_2000_353_2015_RData.RData'
shpname = 'path_to_file/rois.shp'
#Set the start/end dates for extraction
startdate = as.Date("2010-01-01")
enddate = as.Date("2014-12-31")
#Load the RasterStack
inrts = get(load(infile))
# Compute average and St.dev
dataavg = MODIStsp_extract(inrts, shpname, startdate, enddate, FUN = 'mean', na.rm = T)
datasd = MODIStsp_extract (inrts, shpname, startdate, enddate, FUN = 'sd', na.rm = T)
# Plot average time series for the polygons
plot.xts(dataavg)
```

loads a *RasterStack* object containing 8-days 250 m resolution time series for the 2000-2015 period and extracts time series of average and standard deviation values over the different polygons of a user's selected shapefile on the 2010-2014 period. The function exploits rasterization of the input *Sp** object and fast summarization based on the use of *data.table* (Dowle et al., 2015) objects to greatly increase the speed of data extraction with respect to standard R functions. For example, executing the code above on a *RasterStack* object with 512 rows, 1068 columns and 685 bands (dates) to compute average NDVI values for a SpatialPolygons object containing 6 polygons showed that `MODIStsp_extract` takes 1.02 minutes, while a more standard R approach (based on prior subsetting of the time series on the period of interest followed by a call to the `raster::extract` function) takes 4.95 minutes.[4]

5. Summary and Conclusions

In this manuscript, we presented *MODIStsp*, a new R package devoted to automating the creation of raster time series derived from MODIS Land Products data. MODIStsp exploits a powerful and user-friendly Graphical User Interface which allows easy selection of all processing parameters. Besides the product of interest and the temporal and spatial extent of the analysis, users can select which layers of the original MODIS HDF files they want to process, the Quality Indicators to be extracted from MODIS Quality Assurance layers and, for Surface Reflectance products, the Spectral Indexes to be computed from reflectance bands.

Required MODIS HDF files are automatically downloaded from NASA servers and resized, reprojected, resampled and processed according to user's choices. For each desired output layer, outputs are saved as single-band rasters corresponding to each acquisition date available for the selected MODIS product within the specified time period. Virtual files facilitating access to the entire time series can be also created.

Processing parameters can be saved in user-specified options files for later use. This allows both non-interactive execution within an R script and stand-alone execution outside an R environment. The former allows to include MODIStsp functionalities within generic R scripts, while the latter allows scheduling *MODIStsp* execution to automatically update a time series when a new image is available.

Preprocessed time series can be easily accessed from either common image processing/GIS software or from R, allowing successive exploitation of the retrieved data for visualization or scientific analysis.

Although the performed processing tasks are technically straightforward, we believe the developed software to constitute a very useful tool for the Remote Sensing community, since it allows performing all the main preprocessing steps required for the creation of MODIS time series within a standard framework, and without requiring particular programming skills by its users.

Foreseen further developments of *MODIStsp* will concern the development of functions allowing to easily remove low-quality information from the time series on the basis of the QA/QI layers and to perform smoothing of the extracted time series for some of the most used MODIS products (e.g., Vegetation Indexes).

[4]Test conducted on a standard PC, equipped with an Intel® Core™ i7-3770 3.4 GHz processor.

Finally, release of *MODIStsp* on CRAN (Comprehensive R Archive Network - https://cran.r-project.org/) is foreseen.

Acknowledgments

The authors wish to thank the R community for providing the packages exploited by *MODIStsp*, and Dr. Thomas Hengl and Babak Naimi, authors of the `MODISdownload` script from which development of *MODIStsp* started. The authors wish also to thank the two reviewers for their useful suggestions for improving both this manuscript and *MODIStsp*, and in particular Mr. Loïc Dutrieux for pointing out a flaw in implementation of resampling functionalities.

Funding

The research leading to these results was conducted within the ERMES FP7 project (http://www.ermes-fp7space.eu) which received funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement n 606983, and the Space4Agri research project (http://space4agri.irea.cnr.it/) funded and supported by the AQ CNR-Regione Lombardia (CNR, Convenzione Operativa n. 18091/RCC).

References

Bartlam-Brooks, H.L.A., Beck, P.S.A., Bohrer, G., Harris, S., 2013. In search of greener pastures: Using satellite images to predict the effects of environmental change on zebra migration. Journal of Geophysical Research: Biogeosciences 118, 1427–1437. doi:10.1002/jgrg.20096.

Bivand, R., Keitt, T., Rowlingson, B., 2016. rgdal: Bindings for the Geospatial Data Abstraction Library. URL: https://CRAN.R-project.org/package=rgdal. r package version 1.1-8.

Bivand, R., Rundel, C., 2016. rgeos: Interface to Geometry Engine - Open Source (GEOS). URL: https://CRAN.R-project.org/package=rgeos. r package version 0.3-19.

Boschetti, M., Nutini, F., Manfron, G., Brivio, P.A., Nelson, A., 2014. Comparative analysis of normalised difference spectral indices derived from MODIS for detecting surface water in flooded rice cropping systems. PLoS ONE 9. doi:10.1371/journal.pone.0088741.

Brown, C., 2013. hash: Full feature implementation of hash/associated arrays/dictionaries. R package version 2.2.6. URL: http://cran.r-project.org/package=hash.

Busetto, L., Colombo, R., Migliavacca, M., Cremonese, E., Meroni, M., Galvagno, M., Rossini, M., Siniscalco, C., Morra di Cella, U., Pari, E., 2010. Remote sensing of larch phenological cycle and analysis of relationships with climate in the alpine region. Global Change Biology 16, 2504–2517. doi:10.1111/j.1365-2486.2010.02189.x.

Chemin, Y., 2008. modis.qc.R "R" script. URL: http://r-forge.r-project.org/scm/viewvc.php/pkg/RemoteSensing/R/modis.qc.R?root=remotesensing&pathrev=79.

Defourny, P., Vancutsem, C., Bicheron, C., Brockmann, C., Nino, F., Schouten, L., Leroy, M., 2007. GlobCover: A 300M Global Land Cover Product for 2005 Using ENVISAT MERIS Time Series, in: Proceedings of the ISPRS Commission VII Mid-Term Symposium, pp. 8–11.

Dowle, M., Srinivasan, A., Short, T., with contributions from R Saporta, S.L., Antonyan, E., 2015. data.table: Extension of Data.frame. URL: http://CRAN.R-project.org/package=data.table. r package version 1.9.6.

Dutky, S., Maechler, M., 2013. bitops: Bitwise Operations. R package version 1.0-6. URL: http://cran.r-project.org/package=bitops.

Dutrieux, L., DeVries, B., 2014. bfastSpatial: Set of utilities and wrappers to perform change detection on satellite image time-series. URL: https://github.com/dutri001/bfastSpatial, doi:10.5281/zenodo.49693.

EOSDIS, 2009. Earth Observing System ClearingHouse (ECHO) / Reverb Version 10.X [online application]. Greenbelt, MD: EOSDIS, Goddard Space Flight Center (GSFC) National Aeronautics and Space Administrat. URL: https://wist.echo.nasa.gov/api/.

Gamon, J., Surfus, J., 1999. Assessing leaf pigment content and activity with a reflectometer. New Phytologis 143, 105–117.

Gitas, I., Mitri, G., Veraverbeke, S., Polychronaki, A., 2012. Advances in remote sensing of post-fire vegetation recovery monitoringa review, in: Fatoyinbo, L. (Ed.), Remote Sensing of BiomassPrinciples and Applications. InTech, p. 322. doi:10.5772/20571.
Gitelson, A., Merzlyak, M., 1998. Remote sensing of chlorophyll concentration in higher plant leaves. Advances in Space Research 22, 689–692. doi:10.1016/S0273-1177(97)01133-2.
Greenberg, J.A., Mattiuzzi, M., 2015. gdalUtils: Wrappers for the Geospatial Data Abstraction Library (GDAL) Utilities. URL: https://CRAN.R-project.org/package=gdalUtils. r package version 2.0.1.7.
Hall, D.K., Riggs, G.A., Salomonson, V.V., DiGirolamo, N.E., Bayr, K.J., 2002. MODIS snow-cover products. Remote Sensing of Environment 83, 181–194. doi:10.1016/S0034-4257(02)00095-0.
Hengl, T., 2010. Download and resampling of MODIS images. URL: http://spatial-analyst.net/wiki/index.php?title=Download_and_resampling_of_MODIS_images.
Hengl, T., Roudier, P., Beaudette, D., Pebesma, E., 2015. plotkml: Scientific visualization of spatio-temporal data. Journal of Statistical Software 63, 1–25. URL: http://www.jstatsoft.org/v63/i05/.
Hijmans, R.J., 2015. raster: Geographic Data Analysis and Modeling. URL: https://CRAN.R-project.org/package=raster. r package version 2.5-2.
Huete, A., 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment 25, 295–309. doi:10.1016/0034-4257(88)90106-X.
Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83, 195–213. doi:10.1016/S0034-4257(02)00096-2.
Hunt, J., Rock, B., 1989. Detection of changes in leaf water content using Near- and Middle-Infrared reflectances. Remote Sensing of Environment 30, 43–54. doi:10.1016/0034-4257(89)90046-1.
Jönsson, P., Eklundh, L., 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Transactions on Geoscience and Remote Sensing 40, 1824–1832. arXiv:TGRS.2002.802519.
Klein, I., Gessner, U., Kuenzer, C., 2012. Regional land cover mapping and change detection in Central Asia using MODIS time-series. Applied Geography 35, 219–234. doi:10.1016/j.apgeog.2012.06.016.
Lang, D.T., the CRAN team, 2016. RCurl: General Network (HTTP/FTP/...) Client Interface for R. URL: https://CRAN.R-project.org/package=RCurl. r package version 1.95-4.8.
Lang, D.T., the CRAN Team, 2016. XML: Tools for Parsing and Generating XML Within R and S-Plus. URL: https://CRAN.R-project.org/package=XML. r package version 3.98-1.4.
Lawrence, M., Verzani, J., 2013. gWidgetsRGtk2: Toolkit implementation of gWidgets for RGtk2. R package version 0.0-82. URL: http://cran.r-project.org/package=gWidgetsRGtk2.
Lhermitte, S., Verbesselt, J., Verstraeten, W.W., Coppin, P., 2007. Assessing Vegetation Regrowth after Fire Based on Time Series of SPOT-VEGETATION Data, in: 2007 International Workshop on the Analysis of Multi-temporal Remote Sensing Images, IEEE. pp. 1–7. doi:10.1109/MULTITEMP.2007.4293050.
Mattiuzzi, M., 2016. MODIS: MODIS Acquisition and Processing. URL: https://R-Forge.R-project.org/projects/modis/. r package version 0.10-34/r510.
Naimi, B., 2014. ModisDownload: an R function to download, mosaic, and reproject the MODIS images. URL: http://r-gis.net/?q=ModisDownload.
Open Source Geospatial Foundation, 2015. GDAL. 2015. GDAL - Geospatial Data Abstraction Library: Version 1.10. URL: http://www.gdal.org.
Pebesma, E.J., Bivand, R.S., 2005. Classes and methods for spatial data in R. R News 5, 9–13. URL: http://CRAN.R-project.org/doc/Rnews/.
Perpiñán, O., Hijmans, R., 2014. rasterVis. URL: http://oscarperpinan.github.io/rastervis/, doi:10.5281/zenodo.12394. r package version 0.32.
R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.
Rouse, J., Haas, R., Schell, J., Deering, D., 1973. Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351. U.S. Gov. Printing office. Technical Report.
Ryan, J.A., Ulrich, J.M., 2015. xts: eXtensible Time Series. URL: https://R-Forge.R-project.org/projects/xts/. r package version 0.9.7/r875.
Tuck, S.L., Phillips, H.R., Hintzen, R.E., Scharlemann, J.P., Purvis, A., Hudson, L.N., 2014. Modistools - downloading and processing modis remotely sensed data in r. Ecology and Evolution 4, 4658–4668. doi:10.1002/ece3.1273.
Tucker, C., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8, 127–150. doi:10.1016/0034-4257(79)90013-0.
Verbesselt, J., Hyndman, R., Zeileis, A., Culvenor, D., 2010. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. Remote Sensing of Environment 114, 2970 – 2980. URL: http://www.sciencedirect.com/science/article/pii/S0034425710002336, doi:http://dx.doi.org/10.1016/j.rse.2010.08.003.
Wickham, H., 2015. stringr: Simple, Consistent Wrappers for Common String Operations. URL: https://CRAN.R-project.org/package=stringr. r package version 1.0.0.