MODIStsp: A tool for automaic preprocessing of MODIS time series

Lorenzo Busetto (busetto.l@irea.cnr.it) Luigi Ranghetti(ranghetti.l@irea.cnr.it)

2015-06-13

1 Introduction

MODIStsp is a novel "R" package allowing to automatize the creation of time series of rasters derived from MODIS Land Products data. It allows to perform several preprocessing steps on MODIS data available within a given time period.

Development of MODIStsp started from modifications of the ModisDownload "R" script by Thomas Hengl (2010), and successive adaptations by Babak Naimi (2014). The basic functionalities for download and preprocessing of MODIS datasets provided by these scripts were gradually incremented with the aim of:

- Developing a standalone application allowing to perform several preprocessing steps (e.g., download, mosaicking, reprojection and resize) on all available MODIS land products by exploiting a powerful and user-friendly GUI front-end;
- Allowing the creation of time series of both MODIS original layers and additional Quality Indicators (e.g., data acquisition quality, cloud/snow presence, algorithm used for data production, etc.) extracted from the aggregated bit-field QA layers
- Allowing the automatic calculation and creation of time series of several additional Spectral Indexes starting form MODIS surface reflectance products

All processing parameters can be easily selected with a user-friendly GUI, although non-interactive execution exploiting a previously created Options File is possible. Stand-alone execution outside an "R" environment is also possible, allowing to use scheduled execution of MODIStsp to automatically update time series related to a MODIS product and extent whenever a new image is available.

2 Installation

MODIStsp was developed completely in the "R" Language and Environment for Statistical Computing (R Development Core Team 2014) 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

To Install MODIStsp:

1. Within "R", install the "gWidgetsRGtk2" package

   ```
   install.packages('gWidgetsRGtk2')
   ```

   when asked, request to install GTK

2. Install the package from GitHub. (You'll need to have the "devtools" package installed and loaded)

   ```
   install.packages('devtools')
   require('devtools')
   install_github('lbusett/MODIStsp', ref = 'devel')
   ```

2.1 Dependencies

MODIStsp exploits functionalities of several additional "R" packages. A complete list of required packages is shown in Table I.

"R" package	Version	Authors
gWidgetsRGtk2	>= 0.0-54	(Lawrence and Verzani 2013)
rgdal	>= 0.8-16	(Bivand and Rundel 2014)
plyr	>= 1.8.1	(Wickham 2011)
gdalUtils	>= 0.3.1	(Greenberg and Mattiuzzi 2014)
XML	>= 3.98-1.1	(Lang 2013a)
RCurl	>= 1.95-4.1	(Lang 2013b)
Stringr	>= 0.6.2	(Wickham 2012)
hash	>= 2.26	(Brown 2013)
raster	>= 2.3.40	(Hijmans 2014)
rgeos	>= 0.3-8	(Bivand, Keitt, and Rowlingson 2014)

Additionaly, MODIStsp **requires availability of the GDAL - Geospatial Data Abstraction Library v. >1.10, with support for HDF4 raster format**.

3 Running the tool

3.1 Interactive mode

To run the tool in interactive mode, load the package and launch the MODIS_tsp function, with no parameters

```
library(MODIStsp)
MODIStsp()
```

This opens a GUI from which processing options can be specified and eventually saved (or loaded). A description of the differentp rocessing parameters to be selected is reported in the "Selection of Processing parameters" subsection at the end of this vignette

3.2 Non-Interactive mode (Using a previously saved options file)

MODIStsp can be also launched in non-interactive mode by setting the optional "GUI" parameter to FALSE, and the "Options_File" parameter to the path of a previously saved Options file. This allows to exploit MODIStsp functionalities within generic "R" processing scripts

```
library(MODIStsp)
# --> Specify the path to a valid options file saved in advance
options_file = "X:/yourpath/youroptions.RData"   from the GUI
MODIStsp(gui = FALSE, options_File = options_file)
```

Specifying also the "spatial_file_path" parameter overrides the output extent of the selected Options File. This allows to perform the same preprocessing on different extents using a single Options File, by looping on an array of spatial files representing the desired output extents.

For example:

```
# Create a character array containing a list of shapefiles (or other spatial files)
extent_list = list.files("X:/path/containing/some/shapefiles/", "\\.shp$")

# loop on the list of spatial files and run MODIStsp using each of them to automatically
# define the output extent (A separate output folder is created for each input spatial file).

for (single_shape in extent_list) {
  MODIStsp(gui = FALSE, options_File = "X:/yourpath/youroptions.RData",
           spatial_file_path = single_shape )
}
```

4 Selection of processing parameters through the GUI

When running MODIStsp in Interactive Mode, a user-friendly GUI allows selection of all processing options required for the creation of the desired MODIS time series

The main available processing options are described briefly in the following:

4.1 MODIS Product, Satellites and Layers

Allows to select the MODIS product of interest from a drop-down menu. The user can also select which MODIS sensors should be considered for download and creation of the time series ("Terra", "Aqua" or both). After selecting the product, pushing the "Select Processing Bands" button opens the "Select Processing Layers" GUI panel (Figure 2), from which the user **must** select which MODIS original and/or derived QI and SI layers should be processed:

1. The left-hand frame allows to select which original MODIS layers should be processed
2. The central frame allows to select which Quality Indicators should be extracted from the original MODIS Quality Assurance layers.
3. For MODIS products containing surface reflectance data, the right-hand frame allows to select which additional Spectral Indexes should be computed.[1]

Some of the most commonly used Spectral Indexes are available for computation by default (Table II).

Index Acronym	Index name and reference
NDVI	Normalized Difference Vegetation Index (Rouse et al. 1973)
EVI	Enhanced Vegetation Index (Huete et al. 2002)
SR	Simple Ratio(Tucker 1979)]
NDFI	Normalized Difference Flood Index (Boschetti et al. 2014)
NDII7 (NDWI)	Normalized Difference Infrared Index – Band 7 (Hunt and Rock 1989)
SAVI	Soil Adjusted Vegetation Index (Huete 1988)
NDSI*	Normalized Difference Snow Index ((Hall et al. 2002)
NDII6*	Normalized Difference Infrared Index – band 6 (Hunt and Rock 1989)
GNDVI*	Green Normalized Difference Vegetation Index (Gitelson and Merzlyak 1998)
RGRI*	Red Green Ration index (Gamon and Surfus 1999)
GRVI*	Green-red ratio vegetation index (Tucker 1979)

[1]The lists of original MODIS layers, QIs and Sis available for the selected product are automatically retrieved from the "MODIStsp_Products_Opts" XML file distributed with the package in /ExtData subfolder.

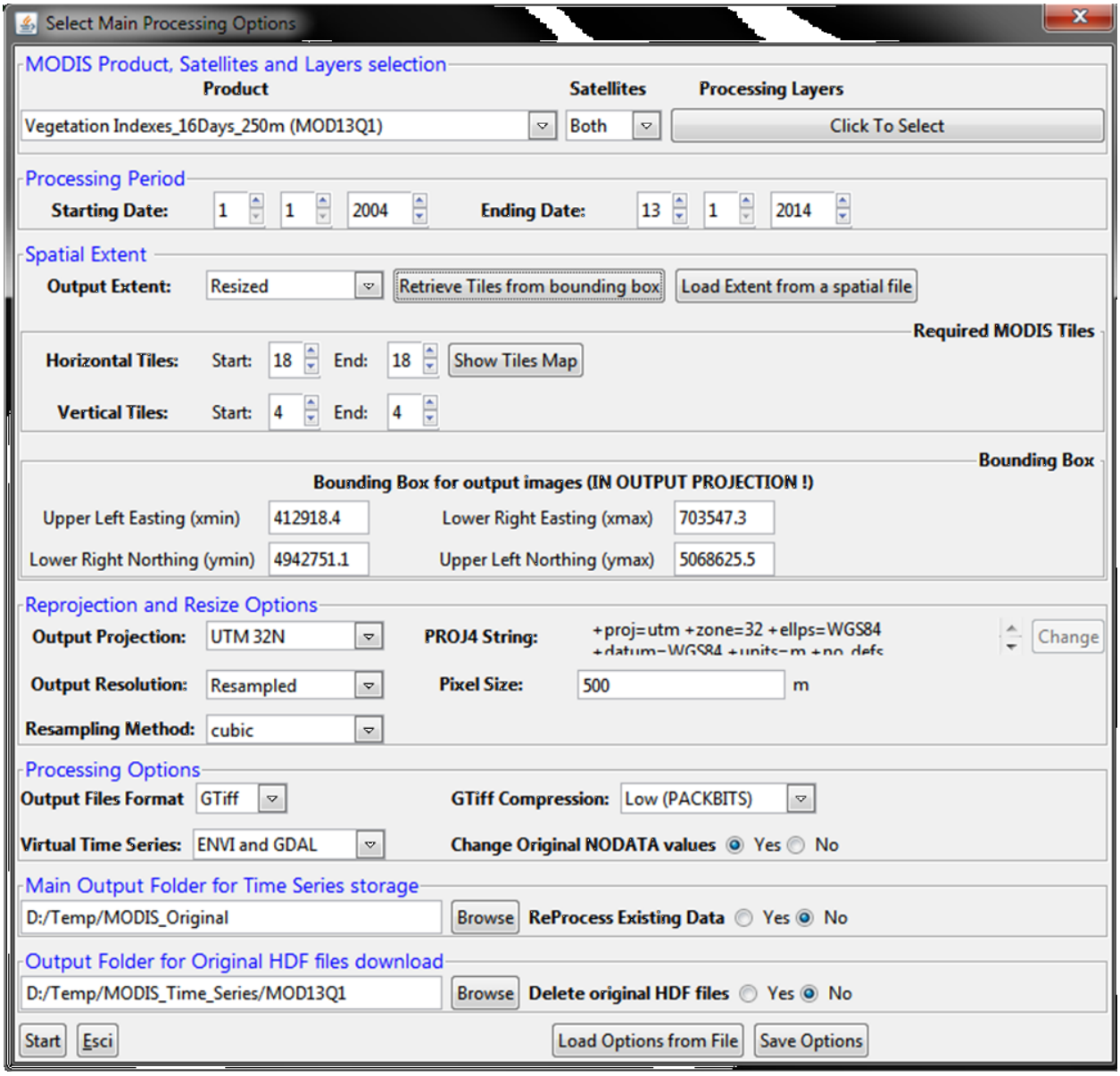

Figure 1: The MODIStsp main GUI

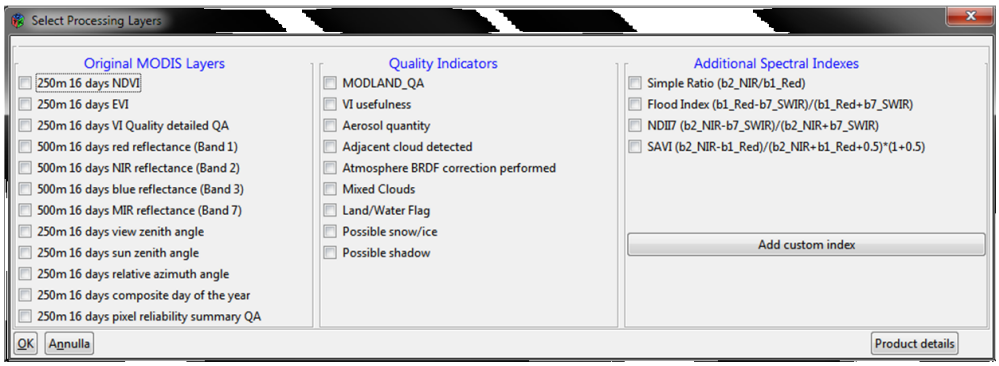

Figure 2: The "Select Processing Layers" GUI

Users can however **specify other SIs to be computed without modifying MODIStsp source code** by clicking on the ***"Add Custom Index"*** button, which allow to provide info related to the new desired SI using a simple GUI (Figure 3) interface.

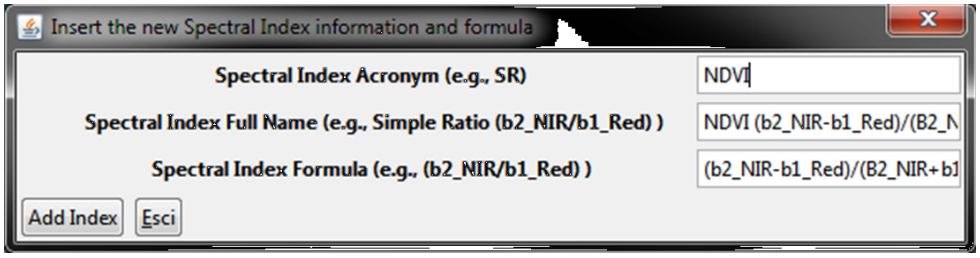

Figure 3: The GUI for insertion of additional Spectral Indexes

Provided information (e.g., correct bandnames, computable formula, etc. . .) is automatically checked, and the new index added in the list of available ones for all products allowing its computation at the next MODIStsp execution.

4.2 Processing Period

Allows to specify the starting and ending dates (dd/mm/yyyy) to be considered for the creation of the time series.

4.3 Spatial Extent

Allows to define the area of interest for the processing. Two main options are possible: .

1. *Full Tiles Extent*: the user must specify which MODIS tiles he would like to process using the "Start" and "End" horizontal and vertical sliders in the "Required MODIS Tiles" frame. During processing, data from the different tiles is mosaiced, and a single file covering the total area is produced for each acquisition date (Note: pressing the "show map" button, a representation of the MODIS tiles grid is shown to facilitate the selection).
2. *Resized*: the user can specify the spatial extent of the desired outputs either;
    a. **Manually inserting** the coordinates of the Upper Left and Lower Right corners of the area of interest in the "Bounding Box" frame. *Coordinates of the corners must be provided in the coordinate system of the selected output projection.*

b. **pressing the "Load Extent from a Spatial File" and selecting a raster or vector spatial file**. In this case, the bounding box of the selected file is retrieved, converted in the selected output projection, and shown in the "Bounding Box" frame.Required input MODIS tiles are also automatically retrieved from the output extent, and the tiles selection sliders modified accordingly.

4.4 Reprojection and Resize

Allows to specify the options to be used for reprojecting and resizing the MODIS images. In particular:

1. The **"Output Projection"** menu allows to either select one of the pre-defined output projections or specify a user-defined one by selecting "User Defined" and then inserting a valid "Proj4" string in the pop-up window. Validity of the Proj4 string is automatically checked, and error messages issued if the check fails.
2. The **"Output Resolution"**, **"Pixel Size"** and **"Reprojection Method"** menus allow to specify whether output images should inherit their spatial resolution from the original MODIS files, or be resampled to a user-defined resolution. In the latter case, output spatial resolution must be specified in the measure units of the selected output projection. Resampling method can be selected among those supported by the "gdalwarp" routine (http://www.gdal.org/gdalwarp.html).

4.5 Processing Options

Allows first of all to specify the **format desired for the output images**. 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 also specified among "None", "PACKBITS", "LZW" and "DEFLATE".

The user can then specify if **virtual multitemporal files** should be created. 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. Available virtual files formats are ENVI metafiles and GDAL "vrt" files.

Finally, users can select if the **NoData values** of MODIS layers should be kept at their original values, or changed to those specified within the "MODIStsp_Products_Opts" XML file. By selecting "Yes" in the "Change Original NODATA values" checkbox, NoData of outputs are set to the largest integer value possible for the data type of the processed layer (e.g., for 8-bit unsigned integer layers, NoData is set always to 255, for 16-bit signed integer layers to 32767, and for 16-bit unsigned integer layers to 65535). .

4.6 Main Output Folder for Time Series Storage

Allows to specify the main folder where the pre-processed time series data will be saved stored. The **"Reprocess Existing Data"** checkbox allows to specify 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 the MODIStsp naming conventions are already present in the output folder. This allows to incrementally extend MODIS time series without reprocessing already available dates.

4.7 Output Folder for Original HDF Storage

Allows to specify the folder where downloaded **original MODIS HDF files** are stored. The **"delete original HDF files"** checkbox allows to specify if the downloaded images should be deleted from the file system at the end of the processing. To avoid accidental file deletion, this is always set to "No" by default, and a warning is issued before execution whenever the selection is changed to "Yes".

5 Processing

Upon pressing the "Start" button, the main processing routine is launched (MODIStsp_process). This routine 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. Connect to the lpdaac http MODIS distribution archive, and retrieve the list of HDF images of the selected MODIS product available for the time period selected by the user, for each tile required to cover the selected study area.
3. For each identified date of acquisition;
    a. Download all required hdf images.
    b. For each original hdf layer i) selected by the user, or ii) required to compute a selected QI or SI layer, extract the data from the original MODIS images and resize and reproject it to the selected output projection, extent and resolution. If more than one tile is needed to cover the output extent, all required tiles are automatically mosaicked before resizing (gdalbuildvrt functionalities are used to avoid creating large temporary raster files). All the main spatial processing tasks are performed using standard GDAL routines, exploiting the wrappers for "R" provided in the "gdalUtils" package (Greenberg and Mattiuzzi 2014). Results are saved as raster GeoTiff or ENVI files using MODIStsp naming conventions (See Section 5).
    c. Starting from files created at point b), compute required QI and/or SI layers and save the results as GeoTiff or ENVI files. Quality Indicators are computed from original QA layers using bitwise operators available in the "bitOps" package (Dutky and Maechler 2013), using a generalization of the "modis.qc.R" script by Yann Chemin (Chemin 2008). Computation of SI layers exploits on-the-fly parsing of the indexes' formulae to identify the required input raster files and perform the computation.
    d. Delete raster files created at point b) used to compute the QI or SI layers but that correspond to original HDF layers not originally selected for processing.
4. When all dates have been processed, create the virtual time series files if required.

6 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 processed original MODIS layer, Quality Indicator or Spectral Index. Each subfolder contains one image for each processed date, created according to the following naming conventions:

"Product_Code"_"Layer"_"YYYY"_"DOY"."ext" *(e.g.,MOD13Q1_NDVI_2000_065.dat)*

Product_Code is the code name of the MODIS product from which the image was derived (e.g., MOD13Q1), **Layer** is a short name describing the dataset (e.g., b1_Red, NDII, UI), **YYYY** and **DOY** corresponds to the year and DOY (Day of the Year) of acquisition of the original MODIS image, and ext is the file extension (.tif for GTiff outputs, or .dat for ENVI outputs).

If requested, ENVI and/or GDAL virtual files are stored in the "Time_Series" subfolder. Naming convention for virtual files is:

"Product_Code"_"Layer"_"StartYYYY"_"StartDOY"_"EndYYYY"_"EndDOY"."ext"

(e.g., MOD13Q1_NDVI_49_2000_17_2015.dat)

StartYYYY, **StartDOY**, **EndYYYY** and **EndDOY** indicate the starting and ending years and DOYS of the time series.

7 Standalone execution and scheduled processing

MODIStsp can be executed as a standalone application using the MODIStsp.bat (for Windows) or MODIStsp.sh (for Linux) batch execution scripts available in the "MODIStsp/ExtData/launcher" subfolder of the package installation. Double-clicking the files or launching them from a shell without parameters launches 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

- **In Linux:** `yourpath_to_MODIStsp.bsh/MODIStsp -g -s "/yourpath/youroptions.RData"`

(launch `yourpath_to_MODIStsp.bsh/MODISts -h` for details).

- **In Windows:** `yourpath_to_MODIStsp.bat\MODIStsp -g -s "X:/yourpath/youroptions.RData"`

(launch `yourpath_to_MODIStsp.bat\MODIStsp -h` for details).

Standalone non-interactive execution easily allows to automatically update the time series of a selected product over a given study area whenever a new MODIS image is available. To do that, the user must simply:

1. Open the MODIStsp GUI, define the parameters of the processing specifying a date in the future as the "Ending Date" and save the processing options. Then 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 time schedule, specifying the path of a previously saved Options file as additional argument:
   - **In Linux:** edit your crontab by opening a terminal and typing

     ```
     crontab -e
     ```

     Then add an entry for the MODIStsp.bsh For example, if you want to run the tool every day at 23.00, add the following row:

     ```
     0 23 * * * /yourpath_to_MODIStsp.bsh/MODIStsp -g -s "/yourpath/youroptions.RData"
     ```

   - **In Windows:** create a Task following these instructions; add the path of the MODIStsp.bat launcher as Action (point 6), and specify `-g -s "X:/yourpath/youroptions.RData"` as argument.

7.1 Adding links to desktop/Start menu for standalone execution

Links to the MODIStsp.bat or MODIStsp.sh standalone launchers can be created automatically launching from R the function `MODIStsp_install_launcher()`

- **In Linux:** this creates a desktop entry (accessible from the menu in the sections "Science" and "Geography"), and a symbolic link in a known path (default: /usr/bin/MODIStsp).
- **In Windows:**: this creates a link in the Start Menu and optionally a desktop shortcut.

See `?install_MODIStsp_launcher` for details and path customisations.

References

Bivand, Roger, and Colin Rundel. 2014. "geos: Interface to Geometry Engine - Open Source (GEOS). R package version 0.3-8." http://cran.r-project.org/package=rgeos.

Bivand, Roger, T. Keitt, and B Rowlingson. 2014. "rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.8-16." http://cran.r-project.org/package=rgdal.

Boschetti, M., F. Nutini, G. Manfron, P.A. Brivio, and A. Nelson. 2014. "Comparative Analysis of Normalised Difference Spectral Indices Derived from MODIS for Detecting Surface Water in Flooded Rice Cropping Systems." *PLoS ONE* 9 (2). doi:10.1371/journal.pone.0088741.

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

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

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

Gamon, J.A., and J.S. Surfus. 1999. "Assessing leaf pigment content and activity with a reflectometer." *New Phytologis* 143 (1): 105–17. doi:10.1046/j.1469-8137.1999.00424.x.

Gitelson, Anatoly A., and Mark N. Merzlyak. 1998. "Remote sensing of chlorophyll concentration in higher plant leaves." *Advances in Space Research* 22 (5): 689–92. doi:10.1016/S0273-1177(97)01133-2.

Greenberg, Jonathan Asher, and Mattia Mattiuzzi. 2014. "gdalUtils: Wrappers for the Geospatial Data Abstraction Library (GDAL) R package version 0.3.1." http://cran.r-project.org/package=gdalUtils.

Hall, Dorothy K, George A Riggs, Vincent V Salomonson, Nicolo E DiGirolamo, and Klaus J Bayr. 2002. "MODIS snow-cover products." *Remote Sensing of Environment* 83 (1-2): 181–94. doi:10.1016/S0034-4257(02)00095-0.

Hengl, T. 2010. "Download and resampling of MODIS images." http://spatial-analyst.net/wiki/index.php?title=Download/_and/_resampling/_of/_MODIS/_images.

Hijmans, R.J. 2014. "raster: raster: Geographic data analysis and modeling. R package version 2.3-0." http://cran.r-project.org/package=raster.

Huete, A, K Didan, T Miura, E.P Rodriguez, X Gao, and L.G Ferreira. 2002. "Overview of the radiometric and biophysical performance of the MODIS vegetation indices." *Remote Sensing of Environment* 83 (1-2): 195–213. doi:10.1016/S0034-4257(02)00096-2.

Huete, A.R. 1988. "A soil-adjusted vegetation index (SAVI)." *Remote Sensing of Environment* 25 (3): 295–309. doi:10.1016/0034-4257(88)90106-X.

Hunt, J.R., and B. Rock. 1989. "Detection of changes in leaf water content using Near- and Middle-Infrared reflectances." *Remote Sensing of Environment* 30 (1): 43–54. doi:10.1016/0034-4257(89)90046-1.

Lang, Duncan Temple. 2013a. "XML: Tools for parsing and generating XML within R and S-Plus. R package version 3.98-1.1." http://cran.r-project.org/package=XML.

———. 2013b. "RCurl: General network (HTTP/FTP/...) client interface for R. R package version 1.95-4.1." http://cran.r-project.org/package=RCurl.

Lawrence, M., and J. Verzani. 2013. "gWidgetsRGtk2: Toolkit implementation of gWidgets for RGtk2. R package version 0.0-82." http://cran.r-project.org/package=gWidgetsRGtk2.

Naimi, B. 2014. "ModisDownload: an R function to download, mosaic, and reproject the MODIS images." http://r-gis.net/?q=ModisDownload.

Rouse, J.W.J., R.H. Haas, J.A. Schell, and D.W. Deering. 1973. *Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351. U.S. Gov. Printing office.* Edited by S.C.

Freden, E.P. Mercanti, and M.A. Becker. *Third Earth Resources Technology Satellite-1 Symposium- Volume I: Technical Presentations.* NASA.

Tucker, Compton J. 1979. "Red and photographic infrared linear combinations for monitoring vegetation." *Remote Sensing of Environment* 8 (2): 127–50. doi:10.1016/0034-4257(79)90013-0.

Wickham, Hadley. 2011. "The Split-Apply-Combine Strategy for Data." *Journal of Statistical Software* 40 (1): 1–29. http://www.jstatsoft.org/v40/i01/.

———. 2012. "stringr: Make it easier to work with strings.. R package version 0.6.2." http://cran.r-project.org/package=stringr.