

# modisfast: An R package for fast and efficient access to MODIS, VIIRS and GPM Earth Observation data

1 August 2024

## Summary

**modisfast** is an R package that allows for easy and fast downloads of various Earth Observation (EO) data, including Moderate Resolution Imaging Spectroradiometer (MODIS) Land products, Visible Infrared Imaging Radiometer Suite (VIIRS) Land products, and Global Precipitation Measurement mission (GPM) products. Based on the Open-source Project for a Network Data Access Protocol (OPeNDAP) standard framework, it enables users to apply filters (spatial, temporal, and dimensional) directly at the downloading phase, supports parallelized downloads, and streamlines data import into R. Therefore, **modisfast** offers R users a cost-effective, time-efficient, and energy-saving approach to accessing a set of EO datasets.

## Statement of need

Data from Earth Observation satellites are a crucial and increasingly valuable resource for monitoring and understanding our planet, especially in the context of global change. EO data from the U.S. federal government National Aeronautics and Space Administration (NASA) are among the richest and longest-standing in the field. Iconic and widely used NASA EO data collections include the MODIS Land products (Justice et al. 2002), the VIIRS products - which continue the legacy of MODIS (Román et al. 2024), and the GPM mission products (Skofronick-Jackson et al. 2017) (in collaboration with the Japan Aerospace Exploration Agency (JAXA)). Collectively, these products have provided essential data for over 20 years, enabling the study of Earth's dynamics, including (but not limited to) global land cover, vegetation health, land surface temperature, rainfall patterns, burned areas. They support research in climate change, disaster response, biodiversity, ecosystem monitoring, ecology, epidemiology, and more.

Despite the increasing availability and utility of EO data, accessing them presents several challenges. Users often encounter issues such as multiple data sources, data complexity, large file sizes, and the need for advanced technical skills to process the information. These data are typically distributed as multidimensional layers over extensive areas, making accessibility and processing difficult, especially when large time series are required. This problem is particularly acute in developing regions where internet infrastructure can be limited. The complexity of accessing EO data often leads to siloed data processing workflows, separating data extraction from pre-processing and analysis, thereby hindering transparent and reproducible open science practices. While tools like Google Earth Engine offer some solutions, they also have limitations, such as requiring the use of new programming languages and proprietary software. Altogether, these barriers hinder the full potential of Earth observation data to support global research and decision-making. Efforts to simplify and streamline access to this data, while maintaining an open-source and open-science framework, are essential to overcoming these obstacles and maximizing the benefits of satellite data for all.

Here, we introduce **modisfast**, an open-source R package (R Core Team 2024) designed to simplify, streamline, and accelerate the download and import of MODIS, VIIRS, and GPM time series for R users. **modisfast** allows users to download and import only the specific chunks of satellite data needed, filtered spatially,

temporally, and dimensionally, optimizing data download and processing. Additionally, downloads can be parallelized for increased efficiency. **modisfast** thus facilitates access to EO data for R users (especially in regions with limited internet infrastructure), promotes digital sobriety, supports the open-source software movement, and enables embedding data extraction within complex and holistic data workflows in R, fostering transparency and reproducibility in the context of Open Science. Importantly, the foundational framework of **modisfast** (see section ??foundational-fmwrk)) ensures the long-term sustainability and cost-free nature of the package.

## Main features

### Typical workflow

The typical workflow to access and import MODIS, VIIRS or GPM data in R with **modisfast** involves the following steps :

1. Defining the parameters of interest as natural R objects,
2. Login to NASA EOSDIS EarthData with the function `mf_login()`,
3. Building the URL(s) of the dataset(s) of interest with the function `mf_get_url()`,
4. Downloading the dataset(s) with the function `mf_download_data()` (the preferred and default output format is the widely used NetCDF format, although other formats can be defined),
5. Importing the dataset(s) in R as a SpatRast object of **terra** library (Hijmans 2024) with the function `mf_import_data()`.

This workflow is graphically summarized in figure Figure 1.

Other functions available in the package include :

- `mf_list_collections()` to list all the EO data collections available through **modisfast**,
- `mf_list_variables()` to list the available layers for a given collection.

### Special features

To further simplify and fasten data download and import in R, **modisfast** offers several interesting features :

- Ability to download data for multiple regions of interest simultaneously (see dedicated vignette);
- Ability to download data for multiple time periods of interest simultaneously (see dedicated vignette) ;
- Ability to import the downloaded data as a *Virtual Raster Dataset* (VRT) (useful for high-volume data, e.g. country or continental scale data at high spatial resolution) (see documentation of function `mf_import_data()`).

### Available data collections

Currently **modisfast** supports download of 87 data collections, extracted from the following three meta-collections :

- MODIS land products made available by the NASA / USGS LP DAAC (source server : <https://opendap.cr.usgs.gov/opendap/hyrax/>) ;

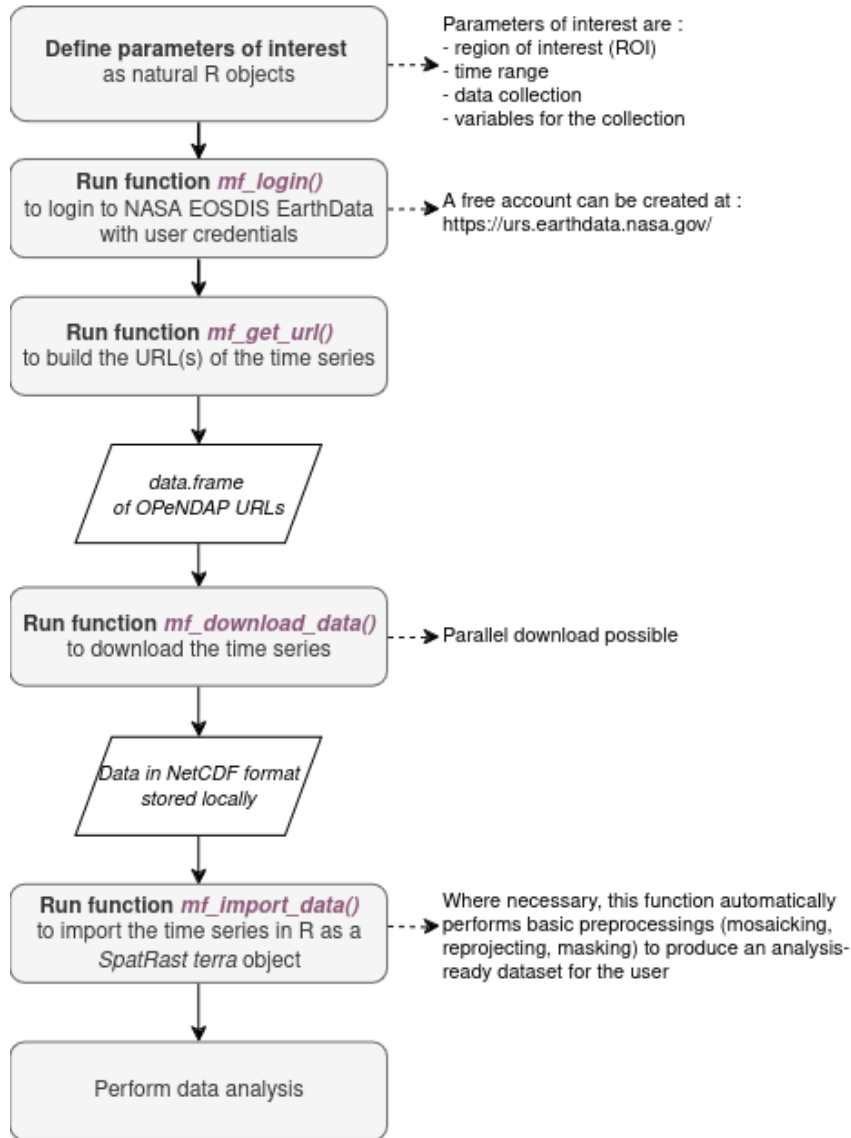


Figure 1: Workflow for data download and import with `modisfast`.

- VIIRS land products made available by the NASA / USGS LP DAAC (source server : <https://opendap.cr.usgs.gov/opendap/hyrax/>)
- Global Precipitation Measurement (GPM) made available by the NASA / JAXA GES DISC (source server : [https://gpm1.gesdisc.eosdis.nasa.gov/opendap/hyrax/GPM\\_L3/](https://gpm1.gesdisc.eosdis.nasa.gov/opendap/hyrax/GPM_L3/)).

The full list of available data collections can be accessed with the function `mf_list_collections()`.

## Foundational framework of `modisfast`

Technically, `modisfast` is a programmatic interface to several NASA OPeNDAP servers (<https://www.opendap.org/>). OPeNDAP is the acronym for “Open-source Project for a Network Data Access Protocol” and designates both the software, the access protocol, and the corporation that develops them. The OPeNDAP is designed to simplify access to structured and high-volume data, such as satellite products, over the Web. It is a collaborative effort involving multiple institutions and companies, with open-source code, free software, and adherence to OGC standards. It is widely used by NASA, which partly finances the project. A key feature of OPeNDAP is its capability to apply filters at the data download process, ensuring that only the necessary data is retrieved. These filters, specified within a URL, can be spatial, temporal, or dimensional. Nevertheless, OPeNDAP URLs are not trivial to build. `modisfast` constructs these URLs based on the filters that are specified by the user through standard R objects.

A key feature of `modisfast` is that it builds on top of robust, sustainable, and cost-free foundations for both the data provider (NASA) and the software (R, OPeNDAP, the `tidyverse` (Wickham et al. 2019) and GDAL (GDAL/OGR contributors 2024) suite of packages). This stability guarantees the long-term reliability of the package.

## Comparison with similar packages

There are other R packages available for accessing MODIS data, which may be more suitable if your requirements differ. These include :

- `MODIS`
- `MODISrsp`
- `MODIStools`
- `appeears`

`modisfast` is particularly suited for retrieving MODIS, VIIRS OR GPM data **over long time series** and **over areas**, rather than short time series and points.

## Installation

The released version of `modisfast` can be installed from CRAN in R with :

```
install.packages("modisfast")
```

or the development version (to get a bug fix or to use a feature from the development version) with :

```
if(!require(devtools)){install.packages("devtools")}
devtools::install_github("ptaconet/modisfast")
```

## Example

Accessing and importing MODIS data in R with `modisfast` is a simple workflow, as shown in the example below.

1/ First, load the packages and define the parameters of interest for you data query (Region Of Interest (ROI), time frame, data collection, and bands for the collection) :

```
# Load the necessary packages
library(modisfast)
library(sf)
library(terra)

## ROI (format sf type POLYGON, with an 'id' column). Here, we provide the bounding box of Madagascar
roi_mada <- st_as_sf(data.frame(id = "madagascar", geom = "POLYGON((41.95 -11.37,51.26 -11.37,51.26 -26.85,41.95 -26.85,41.95 -11.37))"))

## Time range (two dates, i.e. the first and the last)
t_range <- as.Date(c("2023-01-01","2023-12-31"))

## Collection.
# run mf_list_collections() to get an exhaustive list of collections available
coll <- "MOD13A3.061" # Here we choose MODIS/Terra Vegetation Indices Monthly 1 km

# Variables for the collection
## run mf_list_variables("MOD13A3.061") to get an exhaustive list of variables available for the collection
bands <- c("_1_km_monthly_NDVI")
```

2/ Then, download the data executing sequentially the functions `mf_login()`, `mf_get_url()`, and `mf_download_data()` :

```
## Login to Earthdata servers with your EOSDIS credentials.
# To create an account (free) go to : https://urs.earthdata.nasa.gov/.
log <- mf_login(credentials = c("username","password")) # set your own EOSDIS username and password

## Get the URLs of the data
urls <- mf_get_url(
  collection = coll,
  variables = bands,
  roi = roi_mada,
  time_range = t_range
)

## Download the data. By default the data is downloaded in a temporary directory, but you can specify a destination
res_dl <- mf_download_data(urls, parallel = T)
```

3/ And finally, open the data in R as a `terra::SpatRaster` object using the function `mf_import_data()` :

```
r <- mf_import_data(
  path = dirname(res_dl$destfile[1]),
  collection = coll,
  proj_epsg = 4326
)
```

You may now go ahead with any useful analysis, e.g. plot the data with (see Figure 2 for the output) :

```
plot(r, col = rev(terrain.colors(20)))
```

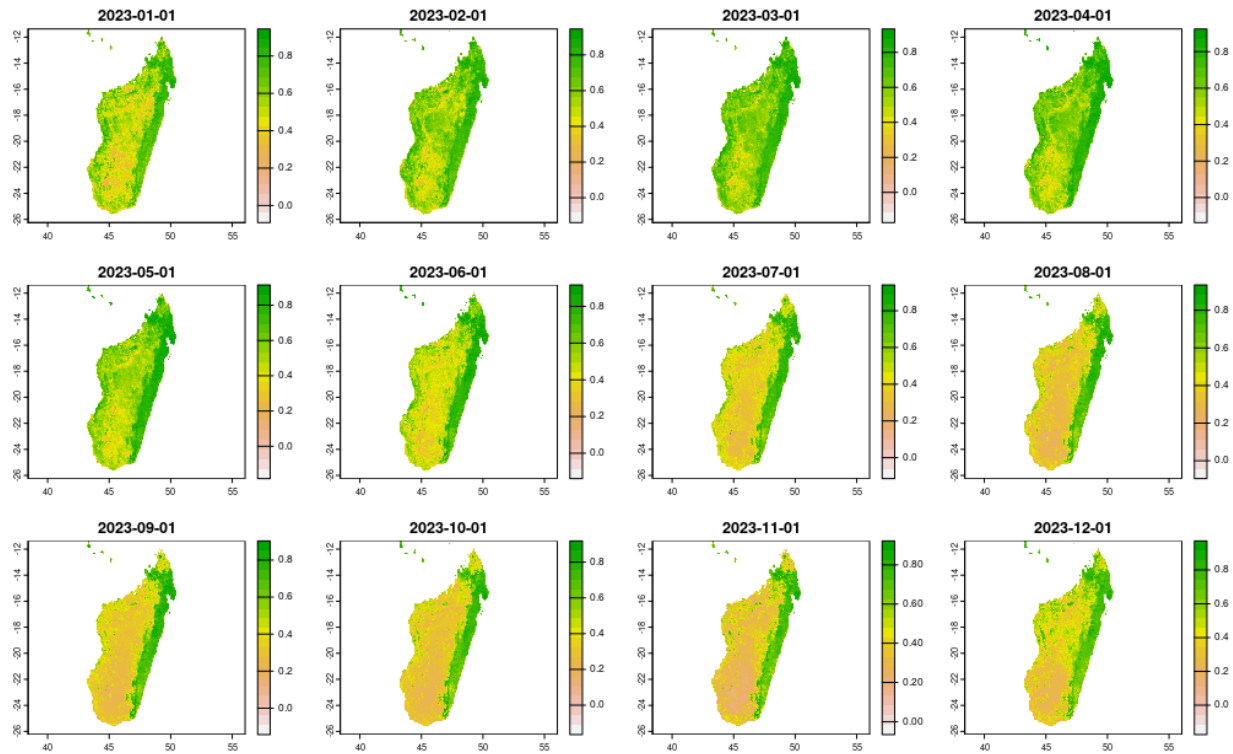


Figure 2: Time series of monthly MODIS NDVI over Madagascar for the year 2023, retrieved with `modisfast`.

`modisfast` provides three long-form documentations and examples to learn more about the package :

- a Get started vignette describing the core features of the package;
- a Get data on several regions or periods of interest simultaneously vignette detailing advanced functionalities of `modisfast` (for multi-time frame or multi-regions data access);
- a Full use case vignette showcasing an example of use of the package in a scientific context (here: landscape epidemiology).

## Future work

Future development of the package may include access to additional data collections from other OPeNDAP servers and support for a variety of data formats as they become available from data providers through their OPeNDAP servers. Furthermore, the creation of an RShiny application (Chang et al. 2024) on top of the package is being considered, as a means of further simplifying data access for users with limited coding skills.

## Acknowledgements

We thank NASA and its partners for making all their Earth science data freely available, and implementing open data access protocols such as OPeNDAP. `modisfast` heavily builds on top of the OPeNDAP, so we thank the non-profit OPeNDAP, Inc. for developing the eponym tool in an open and collaborative way.

We also thank the contributors that have tested the package, reviewed the documentation and brought valuable feedbacks to improve the package : Florian de Boissieu, Julien Taconet, Annelise Tran.

This work has been developed over the course of several research projects :

- the REACT 1 project with funding from the French National Research Institute for Sustainable Development (IRD),
- the REACT 2 project funded by the *l'Initiative - Expertise France*,
- the ANORHYTHM project (ANR-16-CE35-008) funded by the French National Research Agency (ANR),
- the DIV-YOO project (ANR-23-CE35-0005) funded by the French National Research Agency (ANR).

## References

- Chang, Winston, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. 2024. *Shiny: Web Application Framework for r*. <https://CRAN.R-project.org/package=shiny>.
- GDAL/OGR contributors. 2024. *GDAL/OGR Geospatial Data Abstraction Software Library*. Open Source Geospatial Foundation. <https://doi.org/10.5281/zenodo.5884351>.
- Hijmans, Robert J. 2024. *Terra: Spatial Data Analysis*. <https://CRAN.R-project.org/package=terra>.
- Justice, C. O., J. R. G Townshend, E. F Vermote, E Masuoka, R. E Wolfe, N Saleous, D. P Roy, and J. T Morisette. 2002. “An Overview of MODIS Land Data Processing and Product Status.” *Remote Sensing of Environment* 83 (1): 3–15. [https://doi.org/10.1016/S0034-4257\(02\)00084-6](https://doi.org/10.1016/S0034-4257(02)00084-6).
- R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Román, Miguel O., Chris Justice, Ian Paynter, Peter B. Boucher, Sadashiva Devadiga, Arthur Endsley, Angela Erb, et al. 2024. “Continuity Between NASA MODIS Collection 6.1 and VIIRS Collection 2 Land Products.” *Remote Sensing of Environment* 302: 113963. <https://doi.org/10.1016/j.rse.2023.113963>.
- Skofronick-Jackson, Gail, Walter A. Petersen, Wesley Berg, Chris Kidd, Erich F. Stocker, Dalia B. Kirschbaum, Ramesh Kakar, et al. 2017. “The Global Precipitation Measurement (GPM) Mission for Science and Society.” *Bulletin of the American Meteorological Society* 98 (8): 1679–95. <https://doi.org/10.1175/bams-d-15-00306.1>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.