

LPJ-PROSAIL
**User's Guide for the Global dynamic simulated imaging spectroscopy data products for
spaceborne imaging spectroscopy missions.**

Authors: Bryce Currey^{1,2}, Ben Poulter¹, Leo Calle^{1,3}

Associations:

1. Biospheric Sciences Laboratory, NASA Goddard Space Flight Center
2. Land Resources and Environmental Sciences, Montana State University
3. Earth System Science Interdisciplinary Center, University of Maryland

Contact Information:

Bryce Currey:

bryce.currey@montana.edu

Benjamin Poulter:

benjamin.poulter@nasa.gov

Leonardo Calle:

lcalle@umd.edu

Dataset Citation

Poulter, Benjamin, Bryce Currey, Leonardo Calle, Alexey N. Shiklomanov, Cibele H. Amaral, E. N. Jack Brookshire, Petya Campbell, Adam Chlus, Kerry Cawse-Nicholson, Fred Huemmrich, Charles E. Miller, Kimberley Miner, Zoe Pierrat, Ann M. Raiho, David Schimel, Shawn Serbin, William K. Smith, Natasha Stavros, Jochen Stutz, Phil Townsend, David R. Thompson, and Zhen Zhang. 2023. "Simulating Global Dynamic Surface Reflectances for Imaging Spectroscopy Spaceborne Missions: LPJ-PROSAIL." *Journal of Geophysical Research: Biogeosciences* 128(1):e2022JG006935. doi: 10.1029/2022JG006935.

1. Dataset overview

Imaging the reflectivity of the land surface (soils, vegetation, water/ice/snow) offers insights into its health, composition, and productivity. This process, known as imaging spectroscopy or hyperspectral remote sensing, improves the resolution of detail of which we can measure Earth's biodiversity, function, and heterogeneity. By the end of this decade, numerous polar-orbiting (i.e., global coverage), full-range visible and shortwave near-infrared (VSWIR, i.e., 400-2500 nm spectral range) spaceborne missions will have been launched to capture this information in unprecedented detail, including NASA's highly anticipated Surface Biology and Geology (SBG) mission (Cawse-Nicholson et al. 2021) and the ESA's Copernicus Hyperspectral Imaging Mission for the Environment (CHIME; Celesti et al. 2022).

These VSWIR, spaceborne imaging spectrometers will measure the ratio of incoming solar radiance to reflected radiance from a large number of nearly contiguous, narrow spectral bands across a portion of the electromagnetic spectrum, signifying a great advancement over previous earth observation satellites. These detailed measurements of material's spectral reflectance can identify and quantify their albedo, spectrally resolved reflectivity, and infer their chemical and physical properties (Goetz et al., 1985). Particularly powerful are the capabilities of imaging spectroscopy at being able to observe and quantify plant traits, such as chlorophyll content, water stress, and nitrogen and other nutrients across space (Asner and Martin 2016; Féret et al. 2019; Kokaly et al. 2009; Serbin et al. 2014; Singh et al. 2015; Wang et al. 2020), and also enable these retrievals through time (Cheng, Riaño, and Ustin 2014; Chlus and Townsend 2022). This information can be used to monitor vegetation function and health, assess crop yields, and track

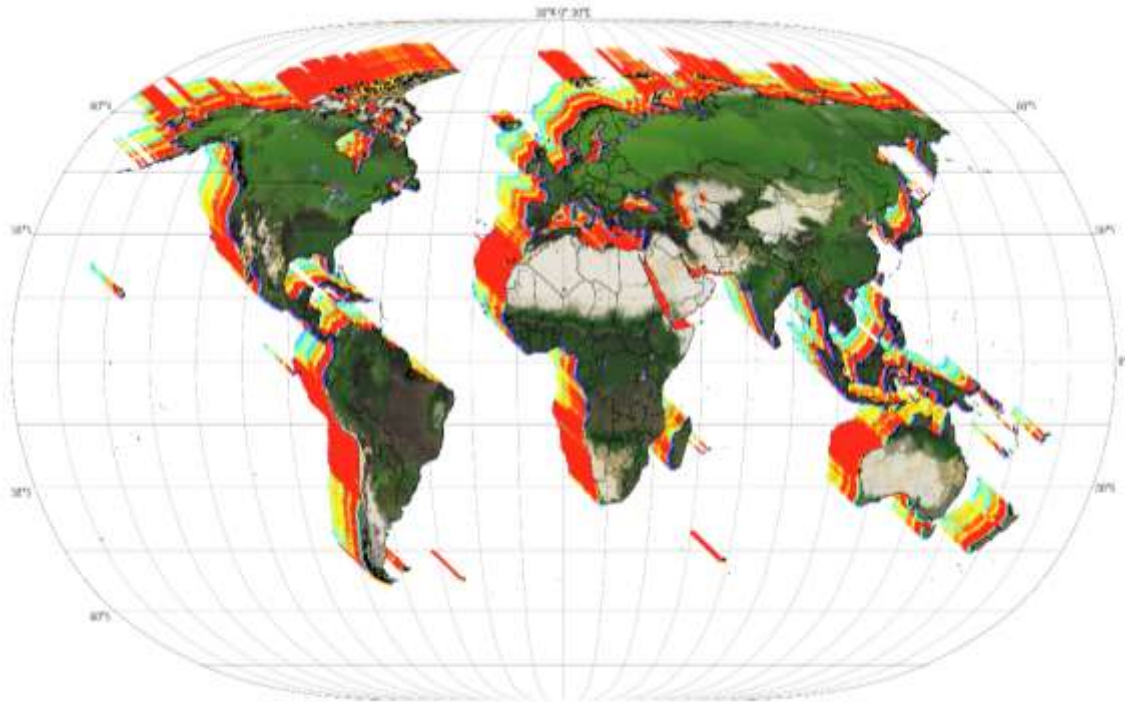


Figure 1. Example of a global imaging spectroscopy data cube. RGB image derived from LPJ-PROSAIL simulated reflectance data.

the effects of climate change on vegetation and ecosystem biodiversity (Gamon et al. 2019; Jetz et al. 2016; Rocchini et al. 2022; Schweiger and Laliberté 2022).

These and other missions alike promise to more accurately gather detailed reflectance data through this remote sensing process of imaging spectroscopy. However, up until launch (even shortly after), these missions depend on simulated datasets and predictive modeling to assess potential uncertainties, create workflows and algorithms, and generally provide the community with an understanding of the possibilities of these data. **In this context, we introduce a dynamic land surface model that generates reflectance spectra ranging from 400 to 2,500 nm, at a 10 nm resolution, covering the entire global land area on a daily time scale known as LPJ-PROSAIL.**

Up front, we must strongly emphasize that these data are simulated data and should under no circumstances be used to derive empirical understanding or ecological conclusions. The purpose of these data is for designing workflows, testing algorithms, and other pre-launch activities or activities in which testing of large spectroscopy datasets are needed.

2. Summary of approach to creating data

The basis of LPJ-PROSAIL is two coupled models, LPJ and PROSAIL. The first, the Lund-Potsdam-Jena (LPJ; LPJ-wsl v2.0; Calle and Poulter 2021) model, is a dynamic global vegetation model (DGVM) that provides a prognostic understanding of how vegetation communities interact with the environment and each other, taking into account impacts from climate, soil resources, and texture, and land use change (Sitch et al. 2003). The LPJ DGVM divides the globe into 50km² grid cells and estimates how different kinds of plant functional types (PFT's; e.g., tropical trees, temperate grasses, etc.) photosynthesize and compete for resources. DGVMs typically simulate thousands of years of historical conditions until the model is “spun up” and then can match contemporary vegetation and carbon stocks and fluxes with a sufficient degree of accuracy (Prentice et al. 2007).

The second model, PROSAIL, is a radiative transfer model (RTM), which is also a mathematically-based model that is used to understand how energy (light) interacts with a medium, in this case, soil and vegetation (Baldocchi et al. 1985; Verhoef 1998; Yuan et al. 2017). PROSAIL itself is a combination of two models, a leaf-level model, PROSPECT (from the French Propriétés Spectrales, or *Spectral Properties*; Jacquemoud and Baret 1990), and SAIL (Scattering by Arbitrarily Inclined Leaves; Verhoef and Bach 2007). In simple terms, PROSPECT simulates the reflectance from a single leaf and SAIL translates the leaf level reflectance to the entire tree canopy scale, incorporating information like leaf area, leaf angle, and soil reflectance properties. Combining these two models, LPJ-PROSAIL simulates the plant functional types and each's suite of plant traits and feeds that into PROSAIL to simulate canopy reflectance.

Finally, secondary, downstream outputs from LPJ-PROSAIL are a simulated top-of-atmosphere (TOA) radiance data product and derived trait products from the reflectance products. The Radiance product is created by running the LPJ-PROSAIL reflectances through the ISOFIT

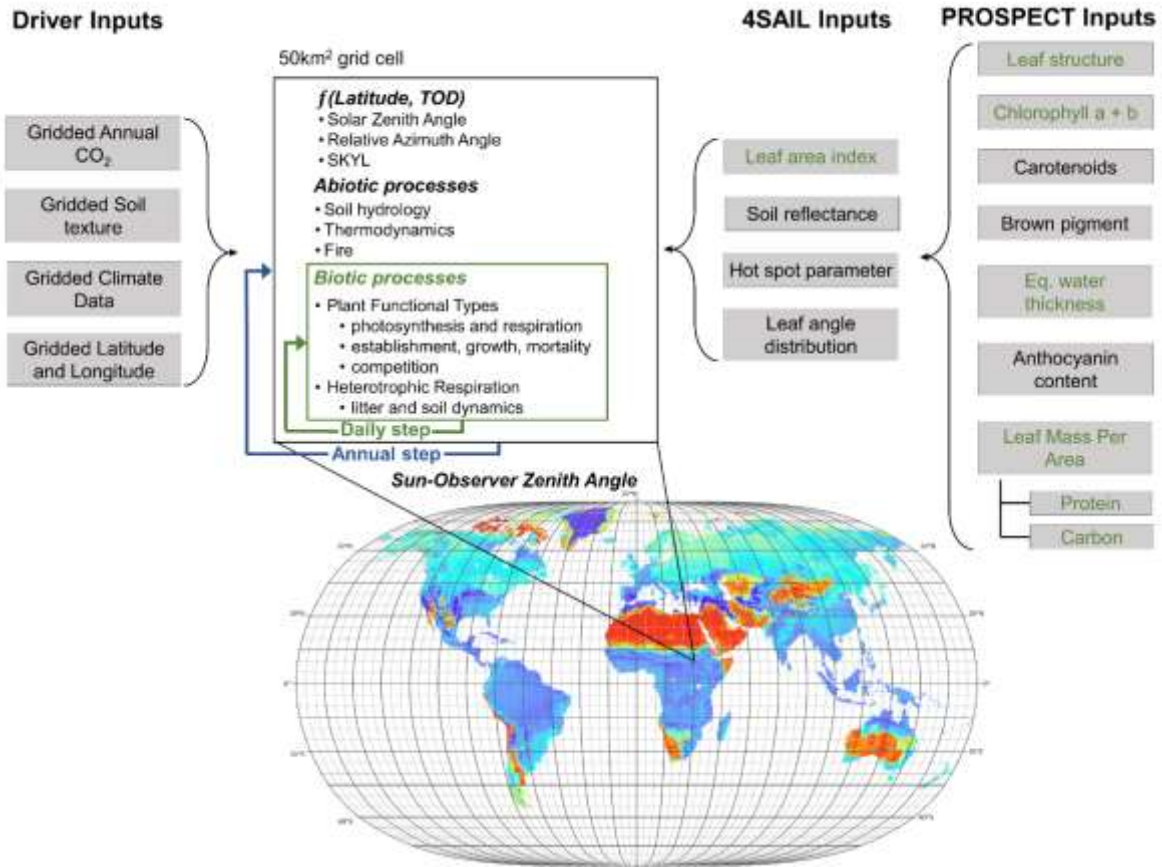


Figure 2. Workflow diagram of the LPJ processing architecture and linkage with PROSPECT-PRO and 4SAIL. Green parameters are dynamic within the LPJ model. Modified from Poulter et al., 2023.

optimal estimation technique. We encourage readers to see Thompson et al. 2018 and visit the GitHub repository (<https://github.com/isofit/isofit>) if they wish to learn more about ISOFIT. As for the estimated trait products, those are created by comparing regression-based coefficients at each wavelength to the LPJ-PROSAIL reflectances (See the ATBD and Serbin et al., 2014 for more details).

2.1. Inputs

The LPJ DGVM requires at minimum daily or monthly gridded temperature, precipitation, cloud cover, CO₂, soil type. For climate, data from the Climate Research Unit (Harris et al. 2020) is typically used, soil texture fractions are obtained from the Harmonized World Soil Database (FAO 2012) and regridded to 0.5°. Global mean annual atmospheric CO₂ time-series data are obtained from NOAA. Other ancillary data such as land use (Hurtt et al. 2020) or other remote sensing data such as data from the Soil Moisture Active Passive (SMAP) mission (Entekhabi et al. 2014). The use of the ancillary inputs depends on the version of the data (see below for versioning).

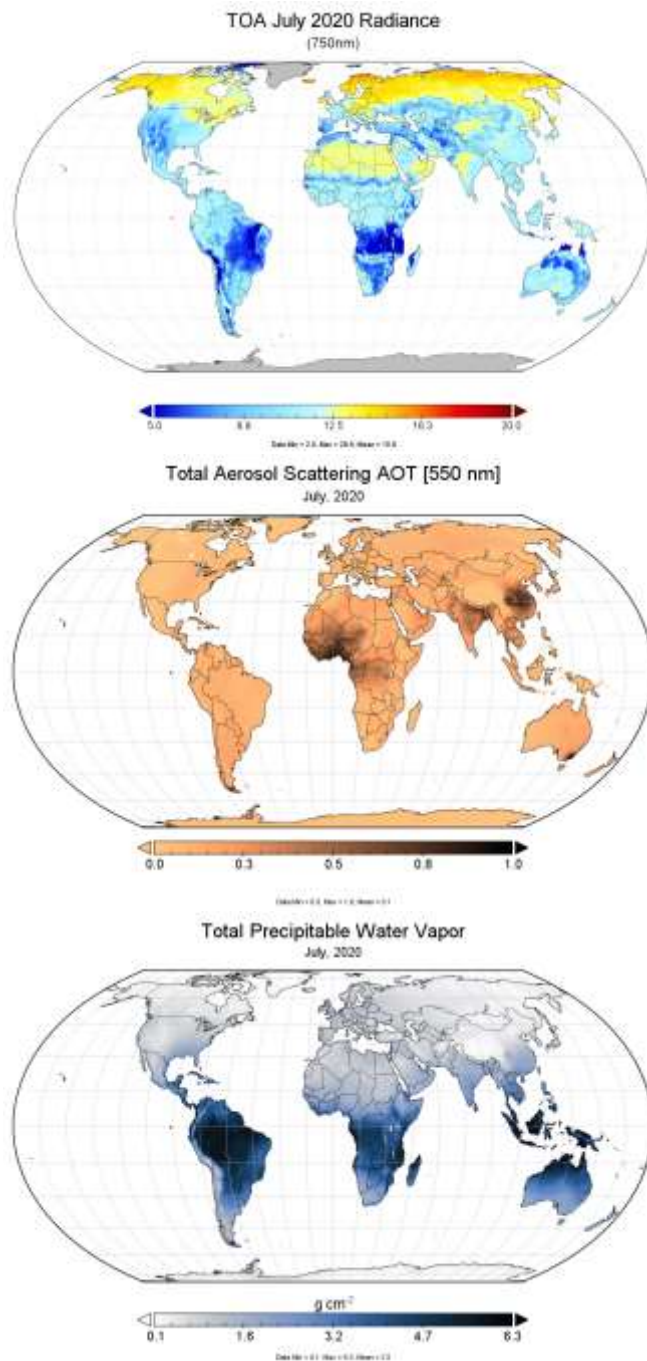


Figure 3. (top) TOA radiance data from LPJ-PROSAIL at 750 nm for July 2020. (middle) MERRA2 aerosols and (bottom) MERRA2 water vapor. Together, these data make up the atmospheric state for any given pixel for ISOFIT.

PROSAIL requires specific absorption coefficients (SACs), which are empirically derived and obtained from the original PROSPECT model (Jacquemoud and Baret 1990). Beyond that, the necessary coefficients are either parameterized as static constants or dynamically derived from the model itself (see Poulter et al. 2023, and the Algorithm Theoretical Basis Document (ATBD) for more information). The original version of LPJ-PROSAIL used PROSPECT-5B (Féret et al. 2008) as the basis of the leaf RTM, but recently we have added PROSPECT-PRO (Féret et al. 2021) to improve fidelity and dynamic parameterization in LPJ-PROSAIL.

Unique to LPJ-PROSAIL are hardcoded spectra from various surfaces, those being snow and woody-branch material. These are used in the reflectance mixing scheme when vegetation is minimal or absent. In earlier versions of the model, the mixing scheme was very rudimentary but has since been improved in later versions. However, more research should be conducted in this realm regarding understanding how multi-surface spectra scale at increasing spectral scales. Also, for the Directional Reflectance stream, (i.e., top-of-canopy reflectance; see below), the diffuse and direct solar irradiance spectra are hardcoded in (again, we refer readers to Poulter et al., 2023).

As for the TOA radiances, inputs to the ISOFIT routine are 1) an instrument model, 2) a surface model, and 3) an atmospheric state. The instrument parameterization we use is the “best estimate” of the capacity that an SBG like instrument would have. The instrument model defines the

instrument’s error producing components, such as the instrument’s signal to noise ratio, smile, tilt, etc. The surface model we use is a generic terrestrial surface model, which in essence simply provides a compilation of possible surface reflectances. Finally, for the atmospheric state, we use MERRA2 aerosol optical thickness (AOT) data and total precipitable water vapor for the same day or month as the reflectance data (Figure 3).

2.2. Outputs

The main outputs from LPJ-PROSAIL are five different reflectance streams. The first four are the standard outputs from the SAIL model (Verhoef and Bach 2007), and the fifth is unique to LPJ-PROSAIL (but see also Shiklomanov et al. 2021):

1. Bidirectional Reflectance (BDR)
2. Bi-hemispherical Reflectance (BHR)
3. Hemispherical-Directional Reflectance (HDR)
4. Directional-Hemispherical Reflectance (DHR)
5. Directional Reflectance (DR).

The nomenclature for the first four reflectances are defined as incoming—outgoing/viewing reflectance. These reflectance streams change given the incident and viewing angles for the directional streams or are agnostic to these angles for the hemispherical streams. For a more in-depth understanding of these angles, we refer readers to the 4SAIL publication (Verhoef and Bach 2007) and Schaepman-Strub et al. 2006 for reflectance angles more generally.

The 5th reflectance stream, dubbed “Directional Reflectance”, is a linear combination of Directional-Hemispherical Reflectance and Bidirectional Reflectance. The weighting of each stream depends on the fraction of diffuse to direct solar illumination, which is hardcoded into the model but is itself a function of wavelength given a general atmospheric composition. The more direct solar irradiance, the more heavily Bidirectional Reflectance is weighted. This occurs at longer wavelengths, as the fraction of direct irradiance increases due to *decreased* scattering effects from the relative size of the atmospheric particles (Figure 4). Conversely, at lower wavelengths, there is more scattering, and as such, Directional-Hemispherical Reflectance is given more weight. We consider DR a simulation of “what a satellite observes”, and for this reason, DR can also be thought of as the closest proxy to top-of-canopy reflectance. All five of these reflectance streams considered an individual science datasets layer and are provided within each collection.

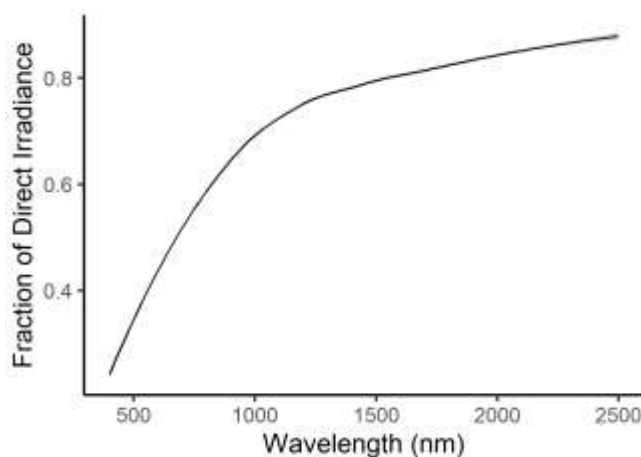


Figure 4. Fraction of direct versus diffuse irradiance across wavelength. Around 650 nm the proportion of diffuse and direct are equal.

The current, final output from LPJ-PROSAIL are the post-processed TOA radiances. Forthcoming, we will translate the top-of-canopy reflectance stream into various vegetation trait products as a secondary output, similar to the level 2+ products from various satellite missions. Initially, we will simulate the core trait products of the SBG mission, namely top-of-canopy chlorophyll, top-of-canopy nitrogen (N), leaf mass per area (LMA) and equivalent water thickness (EWT), with the possibility to host other products as determined by the SBG or CHIME user communities. Each simulated trait map will be hosted as individual science dataset layers provided in the collection.

3. Collections

There are three primary types of collections: 1) reflectances, 2) radiances, and 3) derived vegetation traits from the reflectances. Each collection is global in extent (longitudes from -180 to 180 degrees East and latitudes from -90 to 90 degrees North) and unprojected using the EPSG code 4326 coordinate reference system and WGS84 datum. The temporal granularity is either daily or monthly, depending on version, and the spatial resolution is either at the native resolution as LPJ (0.5 degree, ~50km²) or downscaled to a higher resolution (~1km). For more information regarding downscaling procedures, see the ATBD. Finally, all files are in NetCDF version 4 format with the dimensions: [lon, lat, layer, time]. For the reflectance and radiance collections, the layer dimension is the band wavelengths (typically 211 in size). For other datasets, the layer is a single dimension that represents the dataset itself (i.e., a vegetation trait layer).

3.1. 0.5-Degree Monthly Reflectances

Long name	LPJ-PROSAIL L2 Global Simulated Dynamic Surface Reflectance
Short name	LPJ_L2_SSREF
Granule details	
- Science dataset layers	5 (each reflectance stream)
- Dimensions	720 x 360 x 211 x 12 (lon x lat x bands x time)
- Pixel size	~ 50km ²
- Data type	16-bit Integer (i.e., “short” format)
- Units	Unitless
- Valid range	0-10,000
- Scaling factor	0.0001

3.2. 0.5-degree Monthly TOA Radiances

Long name	LPJ-PROSAIL L2 Global Simulated Dynamic TOA Radiance
Short name	LPJ_L2_SRAD
Granule details	
- Science dataset layers	1
- Dimensions	720 x 360 x 211 x 12 (lon x lat x bands x time)
- Pixel size	~ 50km ²
- Data type	16-bit Integer (i.e., “short” format)

- Units	Unitless
- Valid range	0-10,000
- Scaling factor	0.01

3.3. 0.5-degree Monthly Vegetation Traits

Long name	LPJ-PROSAIL L3 Global Simulated Vegetation Traits	
Short name	LPJ_L3_SVT	
Granule details		
- Science dataset layers	4 (each trait)	
- Dimensions	720 x 360 x 12 (lon x lat x time)	
- Pixel size	~ 50km ²	
- Data type	32-bit Float	
- Units	<ul style="list-style-type: none"> - Chlorophyll (ug/cm²) - Nitrogen (%) - LMA (g/m²) - EWT (cm) 	
- Valid range	<ul style="list-style-type: none"> - Chlorophyll = [0, 100] - Nitrogen = [0, 10] - LMA = [0,1500] - EWT = [0, 2] 	
- Scaling factor	None	

3.4. Short name naming convention

The short name naming conversion is as follows: “DDD_LL_SPPP.VVV”

DDD	LPJ to abbreviate LPJ-PROSAIL
LL	LL indicating the equivalent processing level of the data (L2 or L3)
S	S indicating simulated data
PPP	PPP indicating the product (REF = reflectance, RAD = radiance, VT = vegetation traits).
VVV	VVV flags the version of the data.

4. Versioning

The current versioning structure of the data (as of this draft) of the dataset is as follows. Of note, some of these data products are not yet available but this is the planned versioning structure. If the data is not available, reasoning is provided below.

4.1. Version 1 (V001).

V001 is the original data from the publication of Poulter et. al., 2023. This data is output only as a single year of data for the year 2020. This version contains only LPJ_L2_SSREF surface reflectances.

4.2. Version 2 (V002).

V002 data is the original data from the publication but spanning from 2000 through to the current year (although will probably be depreciated after the year 2023). This version contains only LPJ_L2_SSREF surface reflectances.

4.3. Version 2.1 (V021).

V021 data represents an internal improvement in the architecture of LPJ-PROSAIL, resulting in improved simulated reflectances (see ATBD and Currey et al., In prep). V021 data span 2000-current year. This version contains both LPJ_L2_SSREF surface reflectances and LPJ_L2_SRAD TOA radiances.

4.4. Version 3 (V003).

V003 data represents an increase in the spatial resolution of the data (see ATBD). V003 data spans a shorter time frame 2020-current year given the enormous size of these data. This version currently is only available for LPJ_L2_SSREF surface reflectances.

4.5. Version 4 (V004).

V004 data will represent an increase in the temporal resolution of the data, from daily to monthly (see ATBD). V004 data spans a shorter time frame 2020-current year given the large size of these datasets.

Data Knowledge

5.1. Important considerations

These data are simulated data, and as such, they should not be used to deduce ecological information for decision making. The intended purpose of these data is for testing various algorithms, workflows, and data management practices. Furthermore, like all process-based model output, these data are based on many assumptions across a global context, and as such, the fidelity of these data with in-situ data is moderate.

Of note, there are currently no LP_L3_SVT datasets because global trait algorithms and models have yet to be developed or deployed yet. Similarly, given the simulated nature of these data products, there are no associated QA/QC data with these data, other than limited point-level evaluations as presented in Poulter et al. 2023.

5.2. Related datasets

A notable alternative to these data is the NASA Ames Global Hyperspectral Synthetic Data (AGHSD, pronounced “August”) that were created based on 1km MODIS data for 2019. More information can be found in Wang et al. 2023 and on the website: [AGHSD Portal \(nasa.gov\)](https://ghrsdportal.jpl.nasa.gov/).

6. Examples of Application

Example R code of conducting various common operations on the LPJ-PROSAIL data. This example was created using R markdown, the script which can be found in the linked GitHub repo at the bottom of this example.

6.1. Example Spatial Operations.

Overview

This is a simple example of how to use various geospatial operations like `crop()`, `mask()`, `extract()`, and `reproject()`.

Packages

```
library(tidyverse)
library(terra)
library(ncdf4)
library(readr)
library(ggplot2)
library(tidyterra)
library(RColorBrewer)
library(rworldmap)
library(ggpubr)
```

LPJ path and dataset

```
# We use LPJ_L2_SSREF.V021 for this example
dp <- '~/path/to/data/'
git <- '~/path/to/gitRepo/Global_trait_PLSRs/'
lpj.nc <- 'lpj-prosail_levelC_DR_Version021_m_2020.nc'
```

NEON data

```
# In this example, we'll extract spectra from each NEON site.
# You can download from here: https://www.neonscience.org/field-sites/explore-field-sites or in line in R.
# You can also pull from the Github repo
# Note: You might need to Run R in admin mode to use download.file().
# Note 2: This file gets updated regularly, but the version below is the version on GitHub.

# neon <-
# 'https://www.neonscience.org/sites/default/files/NEON_Field_Site_Metadata_20230309.csv'
# download.file(neon, '~/Current Projects/SBG/LPJ/GlobaL_trait_PLSRs/data/')

neon <- read_csv(file.path(git, 'data/NEON_Field_Site_Metadata_20230309.csv'))
```

Extract the data

```
lpj.array <- nc_open(file.path(dp, lpj.nc))
# To extract all lons, lats, and bands, but select only one month, we will use
# the start and count functionality of ncdf library.
# the dimensions of the data are lon, lat, wavelength, and time. To select all
# values from a dimension, we set start = 1 and count = -1 (for all values).
# In the final dimension placeholder, we select 5 for the 5th month, and 1
# since we only want that one value.

start = c(1,1,1,5)
count = c(-1,-1,-1,1)
rfl <- ncvar_get(lpj.array, 'DR', start, count) # 'DR' is the reflectance stream.
dim(rfl) # should be [720, 360, 211] representing [lon, lat, band].
```

```
## [1] 720 360 211
# To plot as a raster in R, Lon and Lat need to be transposed as such:
rfl <- aperm(rfl, perm = c(2,1,3))
```

Convert to a SpatRaster

```
lpj.r <- rast(rfl, crs = crs('EPSG:4326'), ext = c(-180,180,-90,90))
names(lpj.r) <- paste0('band', seq(400,2500,10))
```

```
# Examine the data
```

```
ggplot() +
  geom_spatraster(data = lpj.r, aes(fill = band650)) +
  scale_fill_gradientn(colours = rev(brewer.pal(11, 'Spectral')), limits = c(0,1),
na.value = 'transparent') +
  theme_void() +
  labs(title = 'May, 2020. Wavelength = 650 nm')
```

May, 2020. Wavelength = 650 nm

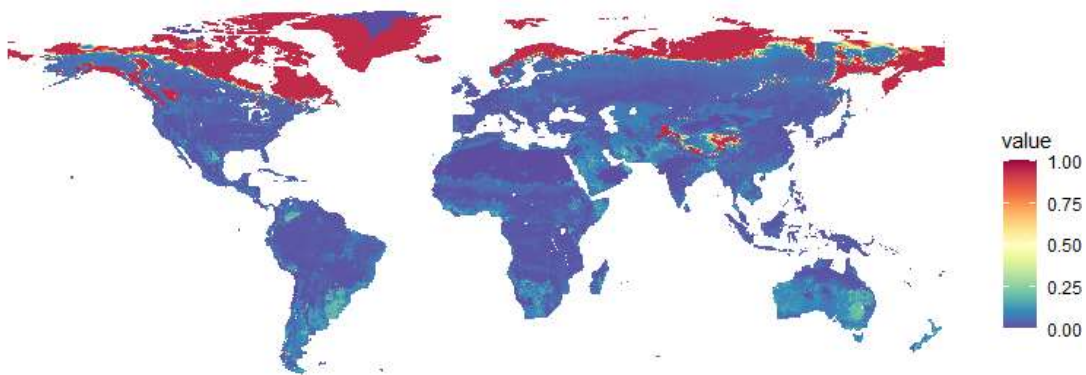


Figure 5. Example reflectance at 650 nm (just before the red edge)

Simple set of spatial operations

Crop()

```
# Next we'll perform a simple set of spatial operations to make a nice RGB figure of the US,
```

```
# First, obtain spatvector of the US. We'll start with a shapefile of the world from rworldmap packages,
```

```
# and subset it to the US. (We'll also convert it to a SpatVector using vect().)
```

```
world <- getMap()
USA <- vect(subset(world, NAME == "United States"))
```

```
# Crop LPJ to that vector
```

```
LPJ.NA <- crop(lpj.r, USA)
plot(LPJ.NA$band400, smooth = T, main = 'North America LPJ Reflectance, 400 nm')
plot(USA, add = T)
```

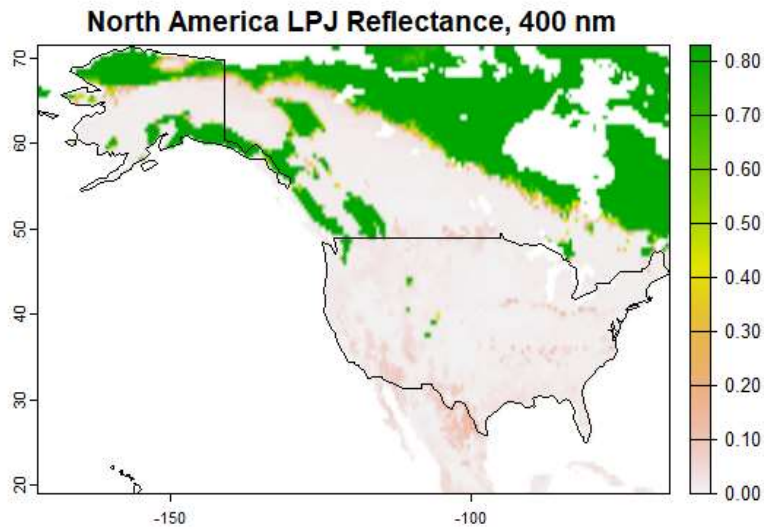


Figure 6. Plot example using terra::plot() instead of ggplot of North America.

Mask()

```
LPJ.USA <- terra::mask(LPJ.NA, USA)
plot(LPJ.USA$band780, smooth = T, main = 'US LPJ Reflectance, 780 nm')
plot(USA, add = T)
```

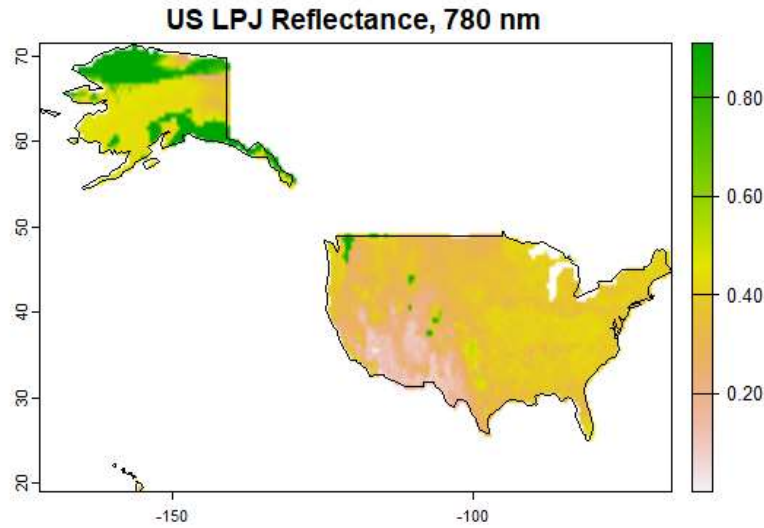


Figure 7. Same as Figure 6 but with only the U.S. displayed.

Extract()

```
# Extract the points from a spat vector (we'll plot later).
neon.points <- vect(neon, geom=c('field_longitude', 'field_latitude'), crs =
'EPSG:4326')
neon.rf1 <- terra::extract(LPJ.USA, neon.points)
```

```

# Set the neon site ID as the ID column
neon.rfl$ID <- neon$field_site_id
head(neon.rfl)
# Convert to long format for ggplot
neon.rfl.plot <- pivot_longer(neon.rfl, -ID, names_to = 'Wavelength', values_to =
'Reflectance') %>%
  tidyr::separate(Wavelength, into = c(NA, "Wavelength"), sep = "(?<=[a-zA-
Z])(?=[0-9])", convert = T)
head(neon.rfl.plot)
## # A tibble: 6 × 3
##   ID      Wavelength Reflectance
##   <chr>      <int>      <dbl>
## 1 ABBY         400      0.0092
## 2 ABBY         410      0.0095
## 3 ABBY         420      0.0097
## 4 ABBY         430      0.0098
## 5 ABBY         440      0.0101
## 6 ABBY         450      0.0109

```

Reproject()

```

# Lets convert to NAD83 for display
LPJ.USA.nad <- project(LPJ.USA, 'EPSG:4269')
USA.nad <- project(USA, 'EPSG:4269')
neon.points.nad <- project(neon.points, 'EPSG:4269')

```

Plot RGR and spectra from NEON sites together

To make an RGB plot, you must know the band index Location of the R, G, and B channels.

Well use Blue = 440, Green = 550, and Red = 660

```

wl <- seq(400,2500,10)
red <- which(wl == 660)
green <- which(wl == 550)
blue <- which(wl == 440)

```

```

spatialPlot <- ggplot() +
  geom_spatraster_rgb(data = LPJ.USA.nad, r = red, g = green, b = blue,
max_col_value = 0.3) +
  geom_spatvector_label(data = neon.points.nad, aes(label = field_site_id),
alpha = 0.5) +
  geom_spatvector(data = USA.nad, alpha = 0, linewidth = 1, color = 'black') +
  scale_fill_manual(na.value = 'transparent') +
  coord_sf(crs = 4269) +
  theme_classic(base_size = 20) +
  theme(plot.margin = margin(1, 0, 1, 0, "cm")) +
  labs(x = 'Longitude', y = 'Latitude')

```

```

spectraPlot <- ggplot(neon.rfl.plot) +
  geom_line(aes(x = Wavelength, y = Reflectance, group = ID, color = ID)) +
  scale_color_viridis_d() +
  theme_bw(base_size = 20) +
  theme(legend.position = c(0.99, 0.99), # Position at the top right
legend.justification = c(1, 1), # Justify plot's corner at that
position
legend.background = element_blank(), # Remove the background

```

```

legend.box.background = element_rect(color = "black", size = 0.5),
legend.text = element_text(size = 15),
plot.margin = margin(0, 6, 0, 6, "cm")

annotate_figure(
  ggarrange(spatialPlot, spectraPlot,
    nrow = 2, ncol = 1,
    heights = c(1, 1), widths = c(1,0.5)),
  top = text_grob("LPJ-PROSAIL spectra extract from NEON sites, May 2020", size =
25, face = "bold", color = "Black")
)

```

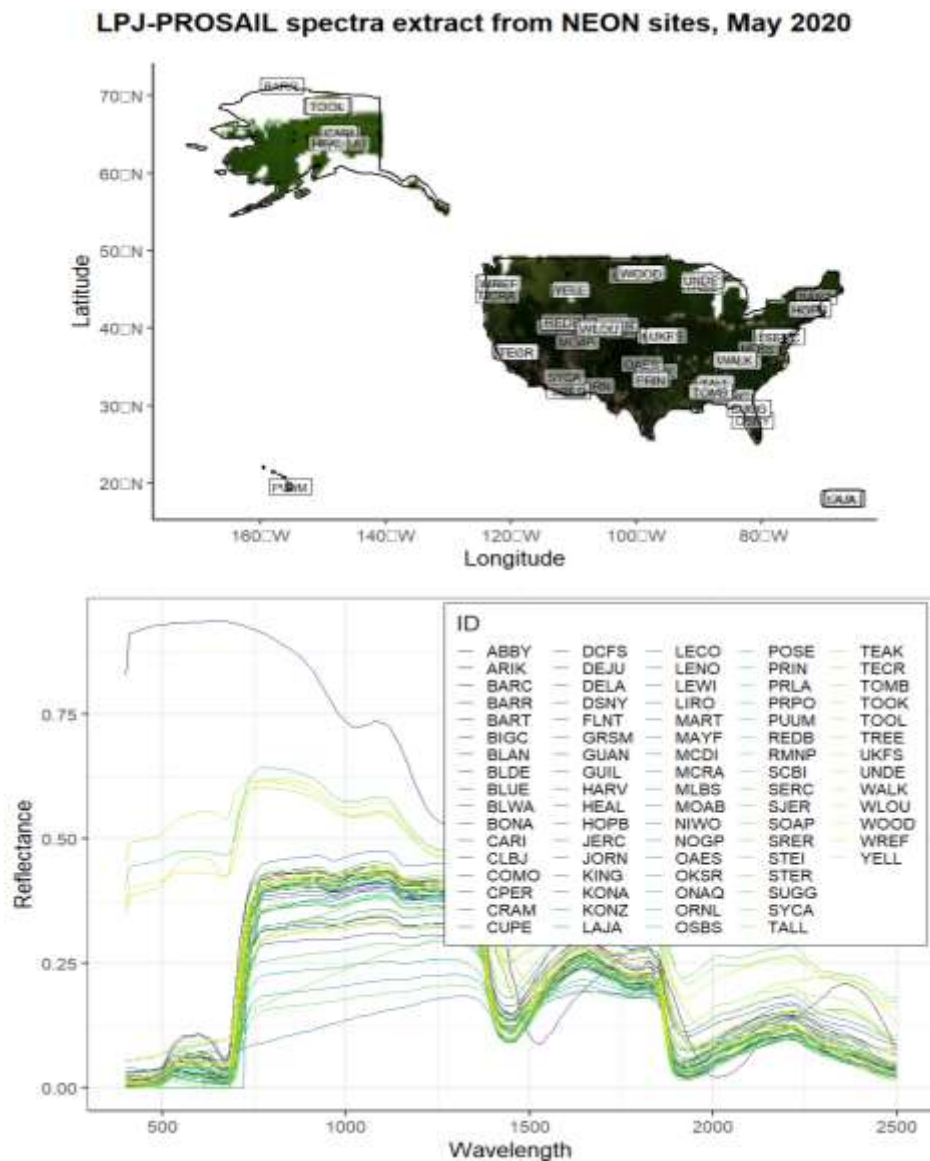


Figure 8. (top) NEON sites displayed with an RGB plot of the U.S. derived from LPJ PROSAIL spectra. (bottom) Extracted LPJ-PROSAIL spectra for each NEON locations.

6.2. Applying PLSR Trait coefficients

This is a simple example of how to map PLSR coefficients across the LPJ-PROSAIL directional reflectance data output. Created in R markdown in the Github repo

Packages

```
library(dplyr)
library(terra)
library(ncdf4)
library(readr)
library(ggplot2)
library(tidyterra)
library(RColorBrewer)
```

Data path and dataset

```
dp <- '~/path/to/data/'
lpj.nc <- 'lpj-prosail_levelC_DR_Version021_m_2020.nc'
coefficients <- '~/Global_trait_PLSRs/Example_LPJ-PROSAIL_PLSR_coefficients.csv'
```

Functions

```
# Resample the bands of lpj-prosail to the same band centers as the coefficient data
wl.interp <- function(y, wavelength) {
  out <- approx(x = seq(400,2500,10), y = y, xout = wavelength,
               method = "linear", rule = 2)[[2]]
  return(out)
}

# Function that maps PLSR coefficients across the LPJ-PROSAIL dataset using
# the terra 'app' function.
trait.map <- function(raster, coeffs, intercept = 0, coeffs_wl = seq(400,2500,10),
na.rm = F) {
  require(terra)

  # first check if the coefficient wavelengths and lpj prosail have the same bands
  if (dim(raster)[3] != length(coeffs_wl)) {
    # if not, resample bands
    raster <- app(raster, function (y, w) wl.interp(y, coeffs_wl))
  }

  # applying the coefficients
  traitmap <- app(raster, function (x) sum(x * coeffs, na.rm) +
as.numeric(intercept))
  return(traitmap)
}
```

Extract the data

```
lpj.array <- nc_open(file.path(dp, lpj.nc))
# To extract all lons, lats, and bands, but select only one month, we will use
# the start and count functionality of ncdf library.
# the dimensions of the data are lon, lat, wavelength, and time. To select all
# values from a dimension, we set start = 1 and count = -1 (for all values).
# In the final dimension placeholder, we select 7 for the 7th month, and 1
```



```
# since we only want that one value.
start = c(1,1,1,7)
count = c(-1,-1,-1,1)
rfl <- nvar_get(lpj.array, 'DR', start, count) # 'DR' is the reflectance stream.
dim(rfl) # should be [720, 360, 211] representing [Lon, Lat, band].
## [1] 720 360 211
# To plot as a raster in R, Lon and Lat need to be transposed as such:
rfl <- aperm(rfl, perm = c(2,1,3))
```

Convert to a SpatRaster

```
lpj.r <- rast(rfl, crs = crs('EPSG:4326'), ext = c(-180,180,-90,90))
lpj.r[lpj.r>0.7] <- NA # remove permanent snow features
lpj.r[lpj.r==0] <- NA # set zero values to NA
names(lpj.r) <- paste0('band', seq(400,2500,10))

# Examine the data
ggplot() +
  geom_spatraster(data = lpj.r, aes(fill = band650)) +
  scale_fill_gradientn(colours = rev(brewer.pal(11, 'Spectral')), limits =
c(0,0.5), na.value = 'transparent') +
  theme_void() +
  labs(title = 'Wavelength = 700 nm')
```

Wavelength = 700 nm

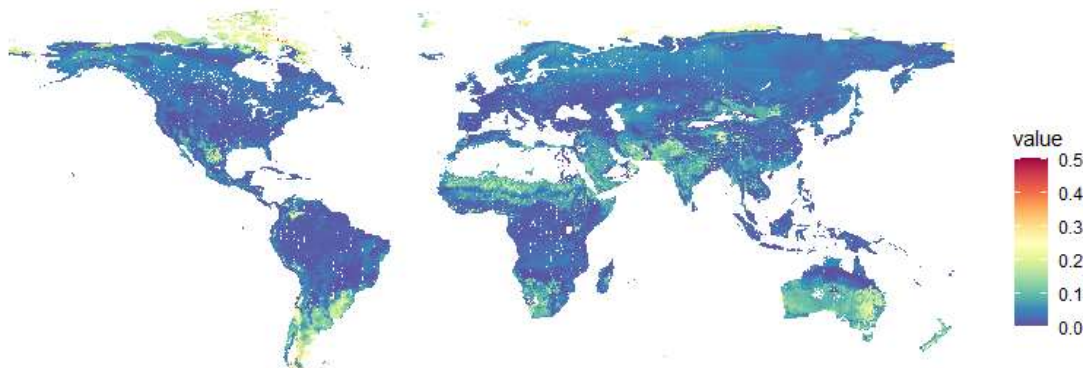


Figure 9. Example of 1 band (700 nm) from LPJ-PROSAIL.

Apply the coefficients

```
coeffs <- read_csv(coefficients)
head(coeffs)
## # A tibble: 6 × 5
##   coeff      lma      n      p      sla
##   <chr>    <dbl> <dbl> <dbl> <dbl>
## 1 Intercept 0.300 18.5  1.23 11.4
## 2 400      0.0100 -0.486 -0.0182 -0.296
## 3 410      0.0102 -0.497 -0.0202 -0.292
## 4 420      0.0103 -0.501 -0.0220 -0.295
```

```
## 5 430      0.0104 -0.504 -0.0236 -0.298
## 6 440      0.0105 -0.508 -0.0251 -0.303
# A simple LMA map made using a PLSR between TRY data and LPJ-prosail data
# In this dataset, the intercept is the first coefficient
lma <- trait.map(raster = lpj.r, coeffs = coeffs$lma[-1], intercept = coeffs$lma[1])

ggplot() +
  geom_spatraster(data = lma) +
  scale_fill_gradientn(colors = c("wheat2", "darkgreen"), limits = c(0.25, 0.4),
na.value = 'transparent') +
  theme_void() +
  labs(title = 'LPJ-PROSAIL estimated LMA (g/g)')
```

LPJ-PROSAIL estimated LMA (g/g)

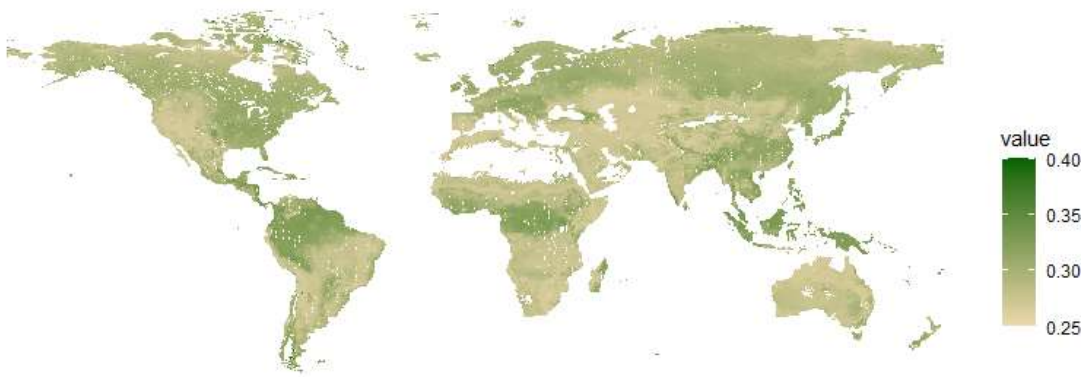


Figure 10. Example of apply PLSR model coefficients to LPJ PROSAIL data. This is an example of what a LPL3SVI product might look like in the future.

More examples of how to use this data can be found here: https://github.com/Green-Currey/Global_trait_PLSRs

7. References

- Asner, Gregory P., and Roberta E. Martin. 2016. "Spectranomics: Emerging Science and Conservation Opportunities at the Interface of Biodiversity and Remote Sensing." *Global Ecology and Conservation* 8:212–19. doi: 10.1016/j.gecco.2016.09.010.
- Baldocchi, D. D., B. A. Hutchison, D. R. Matt, and R. T. McMillen. 1985. "Canopy Radiative Transfer Models for Spherical and Known Leaf Inclination Angle Distributions: A Test in an Oak-Hickory Forest." *Journal of Applied Ecology* 22(2):539–55. doi: 10.2307/2403184.

- Calle, Leonardo, and Benjamin Poulter. 2021. “Ecosystem Age-Class Dynamics and Distribution in the LPJ-Wsl v2.0 Global Ecosystem Model.” *Geoscientific Model Development* 14(5):2575–2601. doi: 10.5194/gmd-14-2575-2021.
- Cawse-Nicholson, Kerry, Philip A. Townsend, David Schimel, Ali M. Assiri, Pamela L. Blake, Maria Fabrizia Buongiorno, Petya Campbell, Nimrod Carmon, Kimberly A. Casey, Rosa Elvira Correa-Pabon, Kyla M. Dahlin, Hamid Dashti, Philip E. Dennison, Heidi Dierssen, Adam Erickson, Joshua B. Fisher, Robert Frouin, Charles K. Gatebe, Hamed Gholizadeh, Michelle Gierach, Nancy F. Glenn, James A. Goodman, Daniel M. Griffith, Liane Guild, Christopher R. Hakkenberg, Eric J. Hochberg, Thomas R. H. Holmes, Chuanmin Hu, Glynn Hulley, Karl F. Huemmrich, Raphael M. Kudela, Raymond F. Kokaly, Christine M. Lee, Roberta Martin, Charles E. Miller, Wesley J. Moses, Frank E. Muller-Karger, Joseph D. Ortiz, Daniel B. Otis, Nima Pahlevan, Thomas H. Painter, Ryan Pavlick, Ben Poulter, Yi Qi, Vincent J. Realmuto, Dar Roberts, Michael E. Schaepman, Fabian D. Schneider, Florian M. Schwandner, Shawn P. Serbin, Alexey N. Shiklomanov, E. Natasha Stavros, David R. Thompson, Juan L. Torres-Perez, Kevin R. Turpie, Maria Tzortziou, Susan Ustin, Qian Yu, Yusri Yusup, and Qingyuan Zhang. 2021. “NASA’s Surface Biology and Geology Designated Observable: A Perspective on Surface Imaging Algorithms.” *Remote Sensing of Environment* 25.
- Celesti, M., M. Rast, J. Adams, V. Boccia, F. Gascon, C. Isola, and J. Nieke. 2022. “The Copernicus Hyperspectral Imaging Mission for the Environment (Chime): Status and Planning.” Pp. 5011–14 in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*.
- Cheng, Tao, David Riaño, and Susan L. Ustin. 2014. “Detecting Diurnal and Seasonal Variation in Canopy Water Content of Nut Tree Orchards from Airborne Imaging Spectroscopy Data Using Continuous Wavelet Analysis.” *Remote Sensing of Environment* 143:39–53. doi: 10.1016/j.rse.2013.11.018.
- Chlus, Adam, and Philip A. Townsend. 2022. “Characterizing Seasonal Variation in Foliar Biochemistry with Airborne Imaging Spectroscopy.” *Remote Sensing of Environment* 275:113023. doi: 10.1016/j.rse.2022.113023.
- Entekhabi, D., S. Yueh, P. E. O’Neill, K. H. Kellogg, A. Allen, R. Bindlish, M. Brown, S. Chan, A. Colliander, and W. T. Crow. 2014. “SMAP Handbook Soil Moisture Active Passive.” *Jet Propulsion Laboratory, California Institute of Technology: Pasadena, CA, USA*.
- FAO. 2012. “Harmonized World Soil Database v1.2.”
- Féret, J. B., G. le Maire, S. Jay, D. Berveiller, R. Bendoula, G. Hmimina, A. Cheraiet, J. C. Oliveira, F. J. Ponzoni, T. Solanki, F. de Boissieu, J. Chave, Y. Nouvellon, A. Porcar-Castell, C. Proisy, K. Soudani, J. P. Gastellu-Etchegorry, and M. J. Lefèvre-Fonollosa. 2019. “Estimating Leaf Mass per Area and Equivalent Water Thickness Based on Leaf Optical Properties: Potential and Limitations of Physical Modeling and Machine Learning.” *Remote Sensing of Environment* 231:110959. doi: 10.1016/j.rse.2018.11.002.

- Féret, Jean-Baptiste, Katja Berger, Florian de Boissieu, and Zbyněk Malenovský. 2021. “PROSPECT-PRO for Estimating Content of Nitrogen-Containing Leaf Proteins and Other Carbon-Based Constituents.” *Remote Sensing of Environment* 252:112173. doi: 10.1016/j.rse.2020.112173.
- Féret, Jean-Baptiste, Christophe François, Gregory P. Asner, Anatoly A. Gitelson, Roberta E. Martin, Luc P. R. Bidet, Susan L. Ustin, Gueric le Maire, and Stéphane Jacquemoud. 2008. “PROSPECT-4 and 5: Advances in the Leaf Optical Properties Model Separating Photosynthetic Pigments.” *Remote Sensing of Environment* 112(6):3030–43. doi: 10.1016/j.rse.2008.02.012.
- Gamon, J. A., B. Somers, Z. Malenovský, E. M. Middleton, U. Rascher, and M. E. Schaepman. 2019. “Assessing Vegetation Function with Imaging Spectroscopy.” *Surveys in Geophysics* 40(3):489–513. doi: 10.1007/s10712-019-09511-5.
- Harris, Ian, Timothy J. Osborn, Phil Jones, and David Lister. 2020. “Version 4 of the CRU TS Monthly High-Resolution Gridded Multivariate Climate Dataset.” *Scientific Data* 7(1):109. doi: 10.1038/s41597-020-0453-3.
- Hurt, George C., Louise Chini, Ritvik Sahajpal, Steve Froking, Benjamin L. Bodirsky, Katherine Calvin, Jonathan C. Doelman, Justin Fisk, Shinichiro Fujimori, Kees Klein Goldewijk, Tomoko Hasegawa, Peter Havlik, Andreas Heinemann, Florian Humpenöder, Johan Jungclaus, Jed O. Kaplan, Jennifer Kennedy, Tamás Krisztin, David Lawrence, Peter Lawrence, Lei Ma, Ole Mertz, Julia Pongratz, Alexander Popp, Benjamin Poulter, Keywan Riahi, Elena Shevliakova, Elke Stehfest, Peter Thornton, Francesco N. Tubiello, Detlef P. van Vuuren, and Xin Zhang. 2020. “Harmonization of Global Land Use Change and Management for the Period 850–2100 (LUH2) for CMIP6.” *Geoscientific Model Development* 13(11):5425–64. doi: <https://doi.org/10.5194/gmd-13-5425-2020>.
- Jacquemoud, S., and F. Baret. 1990. “PROSPECT: A Model of Leaf Optical Properties Spectra.” *Remote Sensing of Environment* 34(2):75–91. doi: 10.1016/0034-4257(90)90100-Z.
- Jetz, Walter, Jeannine Cavender-Bares, Ryan Pavlick, David Schimel, Frank W. Davis, Gregory P. Asner, Robert Guralnick, Jens Kattge, Andrew M. Latimer, Paul Moorcroft, Michael E. Schaepman, Mark P. Schildhauer, Fabian D. Schneider, Franziska Schrod, Ulrike Stahl, and Susan L. Ustin. 2016. “Monitoring Plant Functional Diversity from Space.” *Nature Plants* 2(3):1–5. doi: 10.1038/nplants.2016.24.
- Kokaly, Raymond F., Gregory P. Asner, Scott V. Ollinger, Mary E. Martin, and Carol A. Wessman. 2009. “Characterizing Canopy Biochemistry from Imaging Spectroscopy and Its Application to Ecosystem Studies.” *Remote Sensing of Environment* 113:S78–91. doi: 10.1016/j.rse.2008.10.018.
- Poulter, Benjamin, Bryce Currey, Leonardo Calle, Alexey N. Shiklomanov, Cibele H. Amaral, E. N. Jack Brookshire, Petya Campbell, Adam Chlus, Kerry Cawse-Nicholson, Fred Huemmrich, Charles E. Miller, Kimberley Miner, Zoe Pierrat, Ann M. Raiho, David Schimel, Shawn Serbin, William K. Smith, Natasha Stavros, Jochen Stutz, Phil Townsend, David R. Thompson, and Zhen Zhang. 2023. “Simulating Global Dynamic

- Surface Reflectances for Imaging Spectroscopy Spaceborne Missions: LPJ-PROSAIL.” *Journal of Geophysical Research: Biogeosciences* 128(1):e2022JG006935. doi: 10.1029/2022JG006935.
- Prentice, I. Colin, Alberte Bondeau, Wolfgang Cramer, Sandy P. Harrison, Thomas Hickler, Wolfgang Lucht, Stephen Sitch, Ben Smith, and Martin T. Sykes. 2007. “Dynamic Global Vegetation Modeling: Quantifying Terrestrial Ecosystem Responses to Large-Scale Environmental Change.” Pp. 175–92 in *Terrestrial Ecosystems in a Changing World, Global Change — The IGBP Series*, edited by J. G. Canadell, D. E. Pataki, and L. F. Pitelka. Springer Berlin Heidelberg.
- Rocchini, Duccio, Maria J. Santos, Susan L. Ustin, Jean-Baptiste Féret, Gregory P. Asner, Carl Beierkuhnlein, Michele Dalponte, Hannes Feilhauer, Giles M. Foody, Gary N. Geller, Thomas W. Gillespie, Kate S. He, David Kleijn, Pedro J. Leitão, Marco Malavasi, Vítězslav Moudrý, Jana Müllerová, Harini Nagendra, Signe Normand, Carlo Ricotta, Michael E. Schaepman, Sebastian Schmidlein, Andrew K. Skidmore, Petra Šímová, Michele Torresani, Philip A. Townsend, Woody Turner, Petteri Vihervaara, Martin Wegmann, and Jonathan Lenoir. 2022. “The Spectral Species Concept in Living Color.” *Journal of Geophysical Research: Biogeosciences* 127(9):e2022JG007026. doi: 10.1029/2022JG007026.
- Schaepman-Strub, G., M. E. Schaepman, T. H. Painter, S. Dangel, and J. V. Martonchik. 2006. “Reflectance Quantities in Optical Remote Sensing—Definitions and Case Studies.” *Remote Sensing of Environment* 103(1):27–42. doi: 10.1016/j.rse.2006.03.002.
- Schweiger, Anna K., and Etienne Laliberté. 2022. “Plant Beta-Diversity across Biomes Captured by Imaging Spectroscopy.” *Nature Communications* 13(1):2767. doi: 10.1038/s41467-022-30369-6.
- Serbin, Shawn P., Aditya Singh, Brenden E. McNeil, Clayton C. Kingdon, and Philip A. Townsend. 2014. “Spectroscopic Determination of Leaf Morphological and Biochemical Traits for Northern Temperate and Boreal Tree Species.” *Ecological Applications* 24(7):1651–69. doi: 10.1890/13-2110.1.
- Shiklomanov, Alexey N., Michael C. Dietze, Istem Fer, Toni Viskari, and Shawn P. Serbin. 2021. “Cutting out the Middleman: Calibrating and Validating a Dynamic Vegetation Model (ED2-PROSPECT5) Using Remotely Sensed Surface Reflectance.” *Geoscientific Model Development* 14(5):2603–33. doi: 10.5194/gmd-14-2603-2021.
- Singh, Aditya, Shawn P. Serbin, Brenden E. McNeil, Clayton C. Kingdon, and Philip A. Townsend. 2015. “Imaging Spectroscopy Algorithms for Mapping Canopy Foliar Chemical and Morphological Traits and Their Uncertainties.” *Ecological Applications* 25(8):2180–97. doi: 10.1890/14-2098.1.
- Sitch, S., B. Smith, I. C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J. O. Kaplan, S. Levis, W. Lucht, M. T. Sykes, K. Thonicke, and S. Venevsky. 2003. “Evaluation of Ecosystem Dynamics, Plant Geography and Terrestrial Carbon Cycling in the LPJ Dynamic Global

- Vegetation Model.” *Global Change Biology* 9(2):161–85. doi: 10.1046/j.1365-2486.2003.00569.x.
- Thompson, David R., Vijay Natraj, Robert O. Green, Mark C. Helmlinger, Bo-Cai Gao, and Michael L. Eastwood. 2018. “Optimal Estimation for Imaging Spectrometer Atmospheric Correction.” *Remote Sensing of Environment* 216:355–73. doi: 10.1016/j.rse.2018.07.003.
- Verhoef, Wout. 1998. “Theory of Radiative Transfer Models Applied in Optical Remote Sensing of Vegetation Canopies.” Ph.D., Wageningen University and Research, Netherlands.
- Verhoef, Wout, and Heike Bach. 2007. “Coupled Soil–Leaf–Canopy and Atmosphere Radiative Transfer Modeling to Simulate Hyperspectral Multi-Angular Surface Reflectance and TOA Radiance Data.” *Remote Sensing of Environment* 109(2):166–82. doi: 10.1016/j.rse.2006.12.013.
- Wang, Weile, Jennifer Dungan, Vanessa Genovese, Yohei Shinozuka, Qiguang Yang, Xu Liu, Benjamin Poulter, and Ian Brosnan. 2023. “Development of the Ames Global Hyperspectral Synthetic Data Set: Surface Bidirectional Reflectance Distribution Function.” *Journal of Geophysical Research: Biogeosciences* 128(6):e2022JG007363. doi: 10.1029/2022JG007363.
- Wang, Zhihui, Adam Chlus, Ryan Geygan, Zhiwei Ye, Ting Zheng, Aditya Singh, John J. Couture, Jeannine Cavender-Bares, Eric L. Kruger, and Philip A. Townsend. 2020. “Foliar Functional Traits from Imaging Spectroscopy across Biomes in Eastern North America.” *New Phytologist* 228(2):494–511. doi: 10.1111/nph.16711.
- Yuan, Hua, Yongjiu Dai, Robert E. Dickinson, Bernard Pinty, Wei Shangguan, Shupeng Zhang, Lili Wang, and Siguang Zhu. 2017. “Reexamination and Further Development of Two-Stream Canopy Radiative Transfer Models for Global Land Modeling.” *Journal of Advances in Modeling Earth Systems* 9(1):113–29. doi: 10.1002/2016MS000773.