GriddingMachine.jl: A database and software for Earth system modeling

Yujie WANG

Why a new database?

Data owner experience

- Store the data somewhere
- Direct users to the URL on request
- … …
- Where is my data?

Data user experience

1. Online search
2. Read the paper just to find URL
3. (Register and) Download the data
4. Regrid the data if needed
5. Contact the data owner if steps above do not work…

- … …
- Repeat steps 1–5

Datasets for ESMs: Sources

- Various FTP sites
- Various HTTP sites

Datasets for ESMs: Formats

- Binary
- CSV
- Excel
- GeoTiff
- NetCDF
- Txt
- ...

Datasets for ESMs: Number Types

- Byte
- Integer
- Float
- May be scaled (not mentioned in the dataset file itself)
- May have offset (not mentioned in the dataset file itself)

Datasets for ESMs: Map Orientations

- Some from S to N, and some from N to S
- Some from -180 to 180, and some from 0 to 360
- Some represent the middle point of a pixel (360 * 180)
- Some represent the edges of a pixel (361 * 181 or 360 * 181)

Datasets for ESMs: Coverages

- Some exclude north pole
- Some exclude south pole

Datasets for ESMs: Map Projections

- Cylindrical (WGS84)
- Pseudo-cylindrical

Barriers and Challenges for Users

Repeatedly doing the following

- Download the data
- Rename and remember the file path
- Remember the data label in the dataset
- Use different tools to use the datasets
- Scale and convert the data and units
- What if they only need a point data?

Need for Standardized Datasets and Easier Access

Step 1: Collect the Datasets

GriddingMachine.jl is a byproduct CliMA Land

Acknowledgments:

- Myself
- Renato Braghiere
- Marcos Longo
- Russell Doughty
- Philipp Köhler
- Christian Frankenberg
- Dataset owners

Step 2: Reprocess the Datasets

- Download source: CaltechDATA, Fluo FTP, functions
- Format: NetCDF
- Number type: Float
- Orientations: 180W to 180E, 90S to 90N (center of a grid)
- Coverage: Global
- Projection: Cylindrical (WGS84)

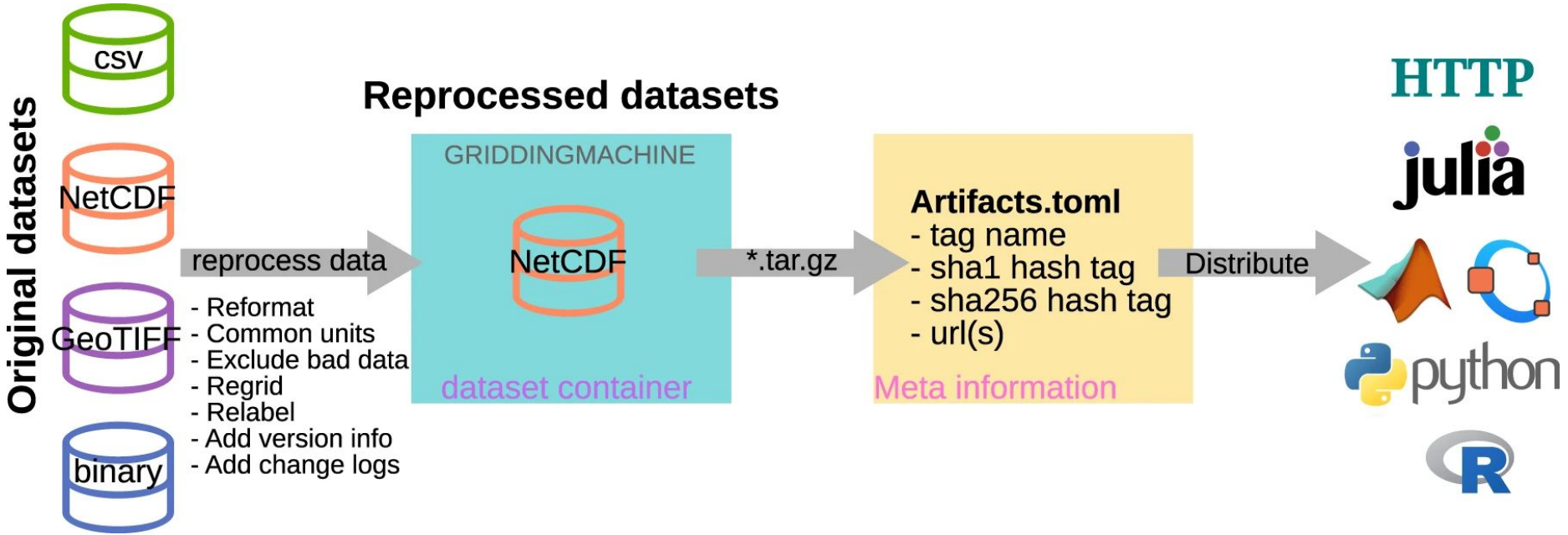
Step 3: Julia Artifacts

A Julia artifact:

- Empty file: GRIDDINGMACHINE
- Data file: LABEL_EXTRALABEL_IX_JT_(YYYY)_VK.nc

Meta info in Artifacts.toml

- Artifact name, e.g., LAI_MODIS_2X_8D_2020_V1
- SHA1, SHA256
- Downloading urls
- ...


Original datasets
csv
NetCDF
GeoTIFF
binary
reprocess data
- Reformat
- Common units
- Exclude bad data
- Regrid
- Relabel
- Add version info
- Add change logs
Reprocessed datasets
GRIDDINGMACHINE
NetCDF
dataset container
*.tar.gz
Artifacts.toml
- tag name
- sha1 hash tag
- sha256 hash tag
- url(s)
Meta information
Distribute
HTTP
julia
python
R

https://github.com/CliMA/GriddingMachine.jl/issues/62

These datasets are supposed to be updated on annually basis:

Dataset type	LABEL	EXTRALABEL	IX	JT	YEAR	VK	Reference	Change logs
Gross primary productivity	GPP	MPI_RS	2X	1M, 8D	2001-2019	V1	Tramontana et al. (2016)	4,9
	GPP	VPM	5X, 12X	8D	2000-2019	V2	Zhang et al. (2017)	1,4
Leaf area index	LAI	MODIS	2X, 10X, 20X	1M, 8D	2000-2020	V1	Yuan et al. (2011)	1,4,9
Latent heat flux	LE	MPI_RS	2X	1M, 8D	2001-2015	V1	Jung et al. (2019)	4,9
Solar induced chlorophyll fluorescence	SIF	TROPOMI_683, TROPOMI_683DC	1X, 2X, 4X, 5X, 12X	1M, 8D	2018-2020	V2	Köhler et al. (2020)	1,8
	SIF	TROPOMI_740, TROPOMI_740DC	1X, 2X, 4X, 5X, 12X	1M, 8D	2018-2021	V1	Köhler et al. (2018)	1,8
	SIF	OCO2_757, OCO2_757DC, OCO2_771, OCO2_771DC	5X	1M	2014-2020	V3	Sun et al. (2017)	1,8

Dataset type	LABEL	EXTRALABEL	IX	JT	YEAR	VK	Reference	Change logs
Biomass	BIOMASS	ROOT	120X	1Y	-	V1	Huang et al. (2021)	1,2
	BIOMASS	SHOOT	120X	1Y	-	V2	Santoro et al. (2021)	1,2
Canopy height	CH	-	20X	1Y	-	V1	Simrad et al. (2021)	1
	CH	-	2X	2Y	-	V2	Boonman et al. (2020)	1,3,4,5
Clumping Index	CI	-	240X	1Y	-	V1	He et al. (2012)	1,3,4
	CI	-	2X	1Y	-	V1	He et al. (2012)	1,3,4,6
	CI	-	2X	1Y	-	V2	Braghiere et al. (2019)	1
Elevation	ELEV	-	4X	1Y	-	V1	Yamazaki et al. (2017)	1,4
Land mask	LM	-	4X	1Y	-	V1	ERA5	1
Leaf nitrogen content	LNC	-	2X	1Y	-	V1	Butler et al. (2017)	
	LNC	-	2X	1Y	-	V2	Boonman et al. (2020)	1,3,4,5
Leaf phosphorus content	LPC	-	2X	1Y	-	V1	Butler et al. (2017)	
Plant functional type	PFT	-	2X	1Y	-	V1	Lawrence and Chase (2007)	1,4,7
Solar induced luminescence	SIL	-	20X	1Y	-	V1	Köhler et al. (2021)	1,3,4
Specific leaf area	SLA	-	2X	1Y	-	V1	Butler et al. (2017)	
	SLA	-	2X	1Y	-	V2	Boonman et al. (2020)	1,3,4,5
Soil color	SC	-	2X	1Y	-	V1	Lawrence and Chase (2007)	1,4,6
Soil hydraulic parameters	SOIL	SWCR	12X	1Y	-	V1	Dai et al. (2019)	1,3,6
	SOIL	SWCR	120X	1Y	-	V1	Dai et al. (2019)	1,3
	SOIL	SWCS	12X	1Y	-	V1	Dai et al. (2019)	1,3,6
	SOIL	SWCS	120X	1Y	-	V1	Dai et al. (2019)	1,3
	SOIL	VGA	12X	1Y	-	V1	Dai et al. (2019)	1,3,6
	SOIL	VGA	120X	1Y	-	V1	Dai et al. (2019)	1,3
	SOIL	VGN	12X	1Y	-	V1	Dai et al. (2019)	1,3,5,6
	SOIL	VGN	120X	1Y	-	V1	Dai et al. (2019)	1,3,5
	SOIL	KSAT	100X	1Y	-	V2	Gupta et al. (2021)	1,2,3,4,5
Surface area	SA	-	1X	1Y	-	V1	Lawrence and Chase (2007)	1,4,6
	SA	-	2X	1Y	-	V1	Lawrence and Chase (2007)	1,4
Tree density	TD	-	2X	1Y	-	V1	Crowther et al. (2017)	1,3,4,6
	TD	-	120X	1Y	-	V1	Crowther et al. (2017)	1,3,4
Maximum carboxylation rate	VCMAX	-	2X	1Y	-	V1	Smith et al. (2019)	1,2,4
	VCMAX	-	2X	1Y	-	V2	Luo et al. (2019)	1,8
Wood density	WD	-	2X	1Y	-	V1	Boonman et al. (2020)	1,3,4,7

Step 4: Dataset Distribution

Traditional ways

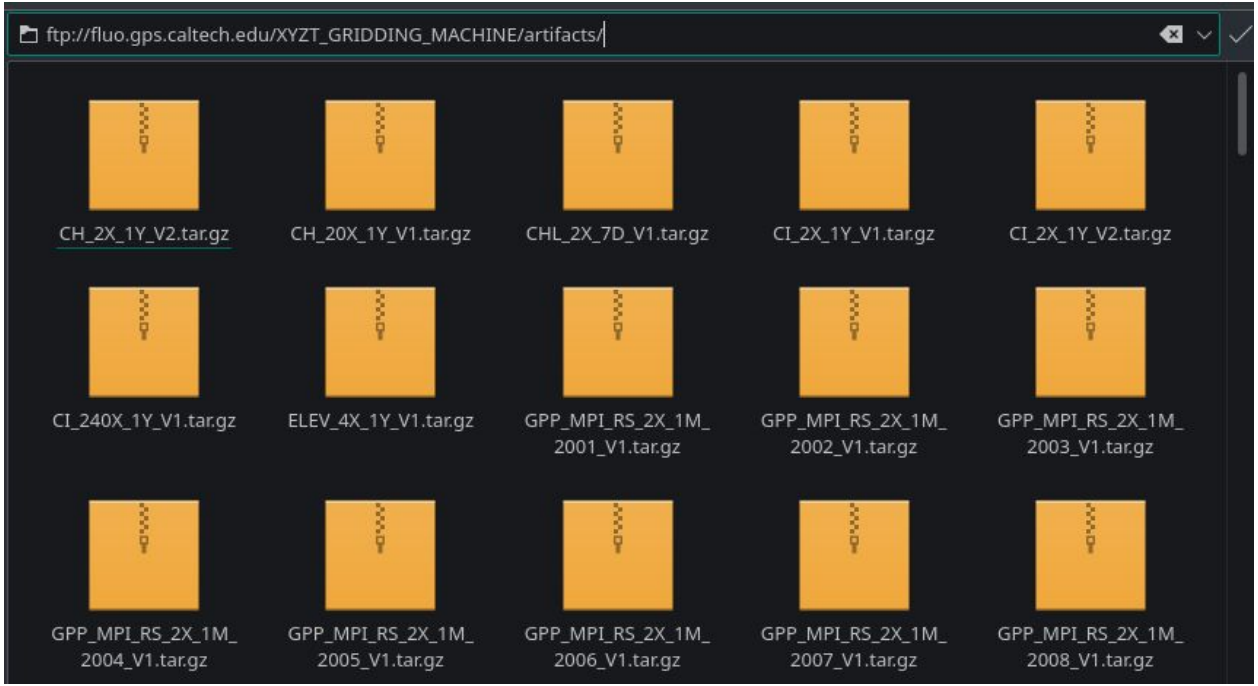
- FTP
- HTTP

New ways

- Direct Julia, Matlab, Octave, Python, and R supports

User Interfaces to Datasets

FTP


ftp://fluo.gps.caltech.edu/XYZT_GRIDDING_MACHINE/artifacts/
CH_2X_1Y_V2.tar.gz
CH_20X_1Y_V1.tar.gz
CHL_2X_7D_V1.tar.gz
CI_2X_1Y_V1.tar.gz
CI_2X_1Y_V2.tar.gz
CI_240X_1Y_V1.tar.gz
ELEV_4X_1Y_V1.tar.gz
GPP_MPI_RS_2X_1M_2001_V1.tar.gz
GPP_MPI_RS_2X_1M_2002_V1.tar.gz
GPP_MPI_RS_2X_1M_2003_V1.tar.gz
GPP_MPI_RS_2X_1M_2004_V1.tar.gz
GPP_MPI_RS_2X_1M_2005_V1.tar.gz
GPP_MPI_RS_2X_1M_2006_V1.tar.gz
GPP_MPI_RS_2X_1M_2007_V1.tar.gz
GPP_MPI_RS_2X_1M_2008_V1.tar.gz

HTTP (CaltechDATA 2129)

Artifacts of GriddingMachine.jl (v0.2) for land modeling

Dataset

2021-09-27

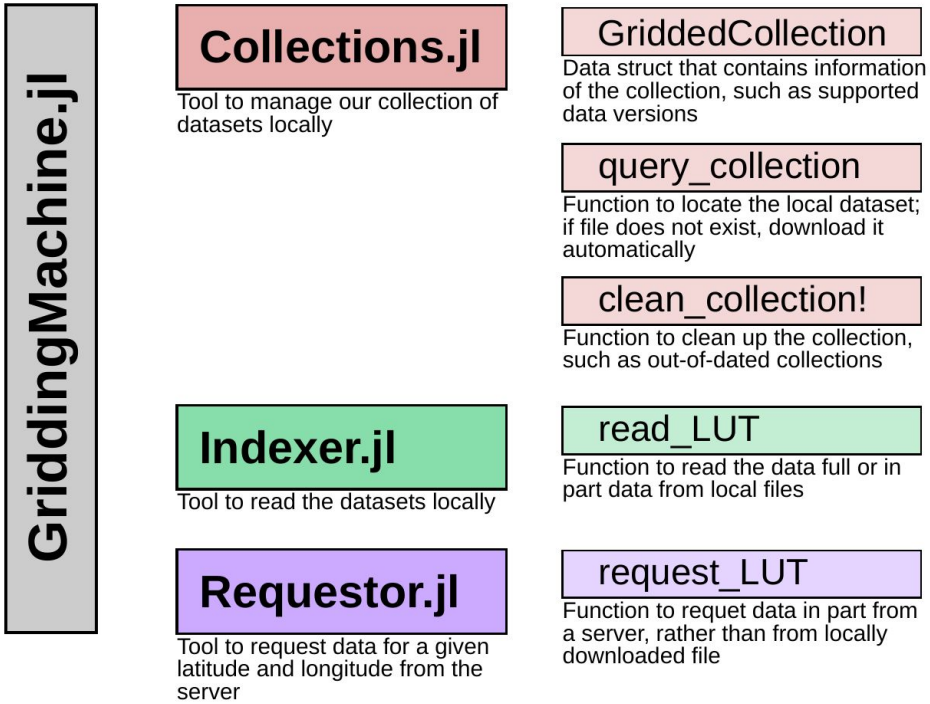
CaltechDATA

Download

Filename	Size	
CH_2X_1Y_V2.tar.gz	572.4 kB	Download
CH_20X_1Y_V1.tar.gz	3.6 MB	Download
CI_2X_1Y_V1.tar.gz	468.0 kB	Download
CHL_2X_7D_V1.tar.gz	7.0 MB	Download
CI_2X_1Y_V2.tar.gz	836.9 kB	Download

What if the dataset gets updated?

Julia: GriddingMachine.jl



Julia: GriddingMachine.jl

https://github.com/CliMA/GriddingMachine.jl

```
# To install
using Pkg; Pkg.add("GriddingMachine");
# To use
using GriddingMachine.Collections;
file_path = query_collection("VCMAX_2X_1Y_V1");
```

https://github.com/CliMA/GriddingMachine.jl/issues/62

These datasets are supposed to be updated on annually basis:

Dataset type	LABEL	EXTRALABEL	IX	JT	YEAR	VK	Reference	Change logs
Gross primary productivity	GPP	MPI_RS	2X	1M, 8D	2001-2019	V1	Tramontana et al. (2016)	4,9
	GPP	VPM	5X, 12X	8D	2000-2019	V2	Zhang et al. (2017)	1,4
Leaf area index	LAI	MODIS	2X, 10X, 20X	1M, 8D	2000-2020	V1	Yuan et al. (2011)	1,4,9
Latent heat flux	LE	MPI_RS	2X	1M, 8D	2001-2015	V1	Jung et al. (2019)	4,9
Solar induced chlorophyll fluorescence	SIF	TROPOMI_683, TROPOMI_683DC	1X, 2X, 4X, 5X, 12X	1M, 8D	2018-2020	V2	Köhler et al. (2020)	1,8
	SIF	TROPOMI_740, TROPOMI_740DC	1X, 2X, 4X, 5X, 12X	1M, 8D	2018-2021	V1	Köhler et al. (2018)	1,8
	SIF	OCO2_757, OCO2_757DC, OCO2_771, OCO2_771DC	5X	1M	2014-2020	V3	Sun et al. (2017)	1,8

Dataset type	LABEL	EXTRALABEL	IX	JT	YEAR	VK	Reference	Change logs
Biomass	BIOMASS	ROOT	120X	1Y	-	V1	Huang et al. (2021)	1,2
	BIOMASS	SHOOT	120X	1Y	-	V2	Santoro et al. (2021)	1,2
Canopy height	CH	-	20X	1Y	-	V1	Simrad et al. (2021)	1
	CH	-	2X	2Y	-	V2	Boonman et al. (2020)	1,3,4,5
Clumping Index	CI	-	240X	1Y	-	V1	He et al. (2012)	1,3,4
	CI	-	2X	1Y	-	V1	He et al. (2012)	1,3,4,6
	CI	-	2X	1Y	-	V2	Braghiere et al. (2019)	1
Elevation	ELEV	-	4X	1Y	-	V1	Yamazaki et al. (2017)	1,4
Land mask	LM	-	4X	1Y	-	V1	ERA5	1
Leaf nitrogen content	LNC	-	2X	1Y	-	V1	Butler et al. (2017)	
	LNC	-	2X	1Y	-	V2	Boonman et al. (2020)	1,3,4,5
Leaf phosphorus content	LPC	-	2X	1Y	-	V1	Butler et al. (2017)	
Plant functional type	PFT	-	2X	1Y	-	V1	Lawrence and Chase (2007)	1,4,7
Solar induced luminescence	SIL	-	20X	1Y	-	V1	Köhler et al. (2021)	1,3,4
Specific leaf area	SLA	-	2X	1Y	-	V1	Butler et al. (2017)	
	SLA	-	2X	1Y	-	V2	Boonman et al. (2020)	1,3,4,5
Soil color	SC	-	2X	1Y	-	V1	Lawrence and Chase (2007)	1,4,6
Soil hydraulic parameters	SOIL	SWCR	12X	1Y	-	V1	Dai et al. (2019)	1,3,6
	SOIL	SWCR	120X	1Y	-	V1	Dai et al. (2019)	1,3
	SOIL	SWCS	12X	1Y	-	V1	Dai et al. (2019)	1,3,6
	SOIL	SWCS	120X	1Y	-	V1	Dai et al. (2019)	1,3
	SOIL	VGA	12X	1Y	-	V1	Dai et al. (2019)	1,3,6
	SOIL	VGA	120X	1Y	-	V1	Dai et al. (2019)	1,3
	SOIL	VGN	12X	1Y	-	V1	Dai et al. (2019)	1,3,5,6
	SOIL	VGN	120X	1Y	-	V1	Dai et al. (2019)	1,3,5
	SOIL	KSAT	100X	1Y	-	V2	Gupta et al. (2021)	1,2,3,4,5
Surface area	SA	-	1X	1Y	-	V1	Lawrence and Chase (2007)	1,4,6
	SA	-	2X	1Y	-	V1	Lawrence and Chase (2007)	1,4
Tree density	TD	-	2X	1Y	-	V1	Crowther et al. (2017)	1,3,4,6
	TD	-	120X	1Y	-	V1	Crowther et al. (2017)	1,3,4
Maximum carboxylation rate	VCMAX	-	2X	1Y	-	V1	Smith et al. (2019)	1,2,4
	VCMAX	-	2X	1Y	-	V2	Luo et al. (2019)	1,8
Wood density	WD	-	2X	1Y	-	V1	Boonman et al. (2020)	1,3,4,7

Matlab

https://github.com/CliMA/GriddingMachine.jl#other-language-supports

```
% To install

urlwrite('https://github.com/Yujie-W/octave-griddingmachine/raw/main/GriddingMa
chine.mltbx', 'GriddingMachine.mltbx');

matlab.addons.toolbox.installToolbox('GriddingMachine.mltbx');

delete('GriddingMachine.mltbx');

% To use

file_path = query_collection('VCMAX_2X_1Y_V1');
```

Octave

https://github.com/CliMA/GriddingMachine.jl#other-language-supports

```
% To install

pkg install "https://github.com/gnu-octave/pkg-json/archive/v1.5.0.tar.gz";

pkg install
"https://github.com/Yujie-W/octave-griddingmachine/archive/v0.1.0.tar.gz";

% To use

pkg load griddingmachine

file_path = query_collection('VCMAX_2X_1Y_V1');
```

Python

https://github.com/CliMA/GriddingMachine.jl#other-language-supports

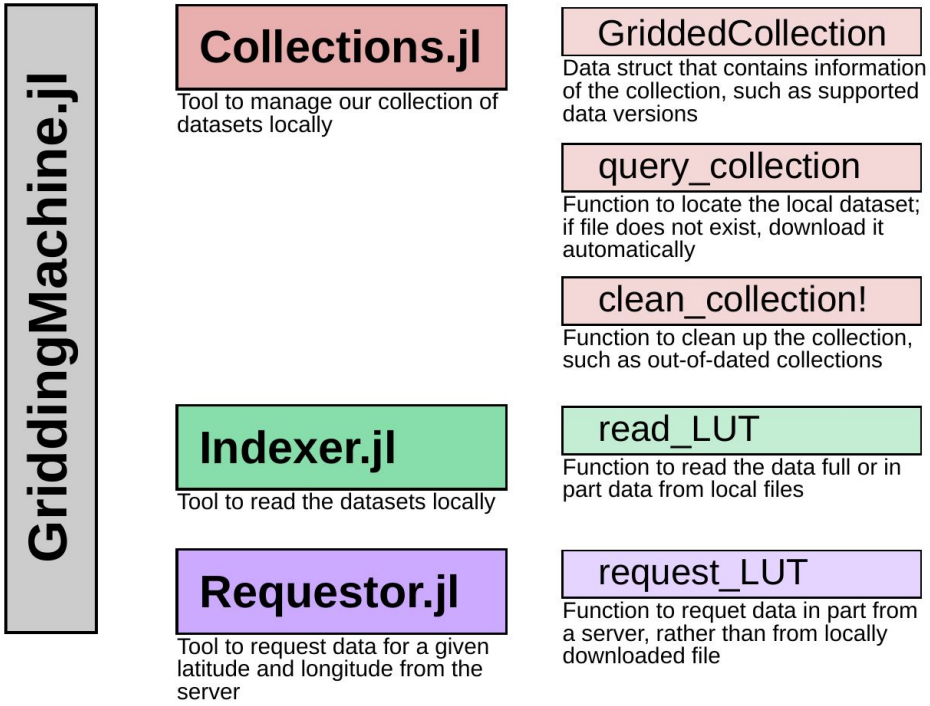
```
# To install

pip install python-griddingmachine

# To use

from griddingmachine import query_collection

file_path = query_collection("VCMAX_2X_1Y_V1");
```

R

https://github.com/CliMA/GriddingMachine.jl#other-language-supports

```
# To install

library("devtools");

install_github("Yujie-W/r-griddingmachine");

# To use

library("griddingmachine");

file_path = query_collection("VCMAX_2X_1Y_V1");
```

User Interface to Request Point Data



GriddingMachine.jl framework

Step 1: GriddingMachine.jl as Web Host

Input parameters

- Mandatory
  - Artifact name
  - Latitude
  - Longitude
- Optional
  - User name (no wait time)
  - Cycle index (point or vector)
  - Interpolation

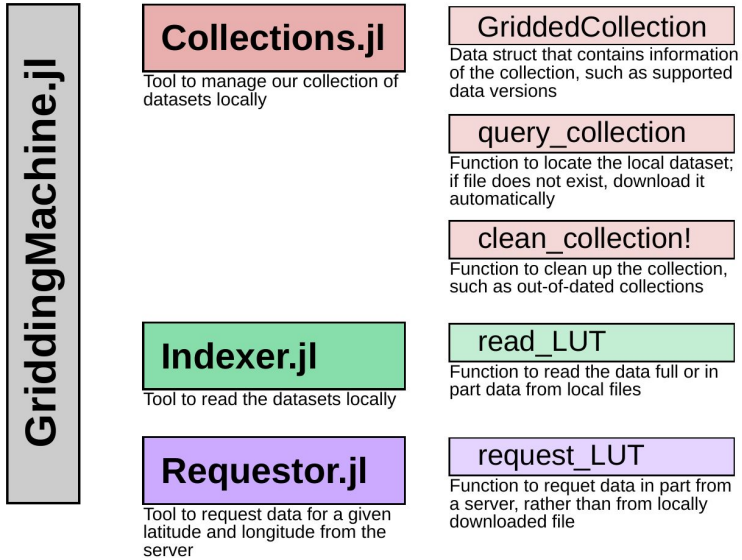
Step 2: Send Request to Server

Julia Example

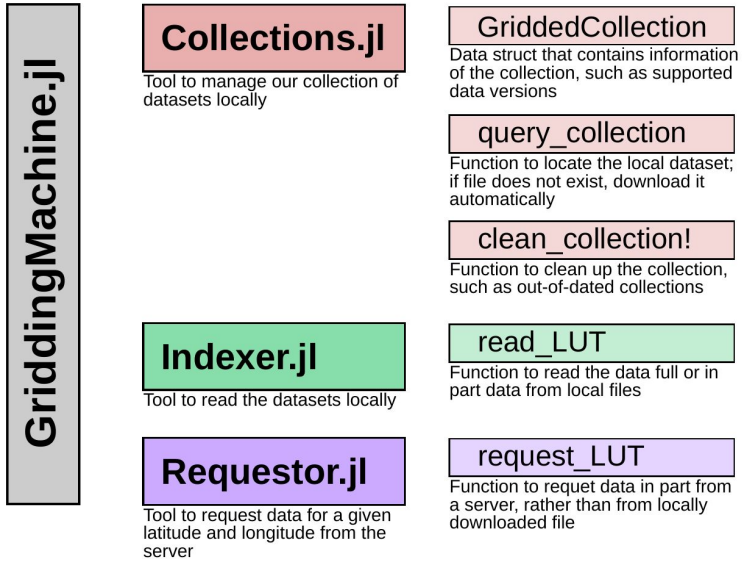
```
using GriddingMachine.Requestor;

request_LUT("LAI_MODIS_2X_8D_2017_V1", 30.5, 115.5);

request_LUT("LAI_MODIS_2X_8D_2017_V1", 30.5, 115.5;
interpolation=true);

request_LUT("LAI_MODIS_2X_8D_2017_V1", 30.5, 115.5, 8);
```

Step 3: Server Prepares Data

Julia server calls GriddingMachine to read the data and prepares a JSON file to return



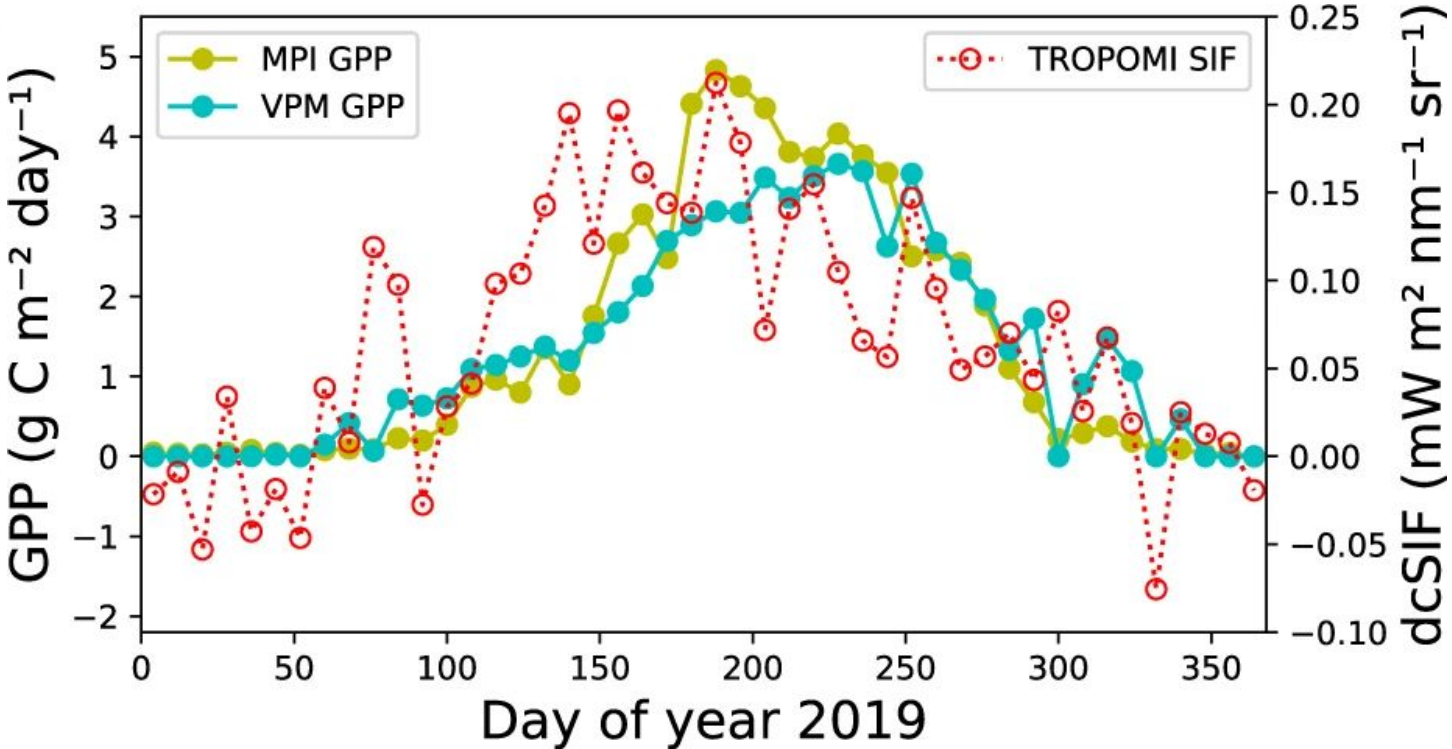
Step 4: Translate JSON Back to Data



Julia Example

```
using GriddingMachine.Requestor;

request_LUT("LAI_MODIS_2X_8D_2017_V1", 30.5, 115.5);

request_LUT("LAI_MODIS_2X_8D_2017_V1", 30.5, 115.5;
interpolation=true);

request_LUT("LAI_MODIS_2X_8D_2017_V1", 30.5, 115.5, 8);
```

```
# Julia script
using GriddingMachine.Requestor;
# load FLUXCOM and VPM GPPs and TROPOMI DC SIF
gpp_mpi,_ = request_LUT("GPP_MPI_RS_2X_8D_2019_V1", 40.0329, -105.5464);
gpp_vpm,_ = request_LUT("GPP_VPM_12X_8D_2019_V2", 40.0329, -105.5464);
sif_dc,_ = request_LUT("SIF_TROPOMI_740DC_12X_8D_2019_V1", 40.0329, -105.5464);
# plot the comparison
```



Works for Other Languages as well

- Matlab
- Octave
- Python
- R

```
% Matlab script
% Install the toolbox
url = 'https://github.com/Yujie-W/octave-griddingmachine/raw/main/GriddingMachine.mltbx';
urlwrite(url, 'GriddingMachine.mltbx');
matlab.addons.toolbox.installToolbox('GriddingMachine.mltbx');
delete('GriddingMachine.mltbx');
% Use the toolbox
update_GM();
art_name = 'VCMAX_2X_1Y_V1';
file_path = query_collection(art_name);
[vcmax,error] = request_LUT(art_name, 35.1, 115.2);
[vcmax,error] = request_LUT(art_name, 35.1, 115.2, 'interpolation', true);
```

```
% Octave script
% Install the package
pkg install "https://github.com/gnu-octave/pkg-json/archive/v1.5.0.tar.gz";
pkg install "https://github.com/Yujie-W/octave-griddingmachine/archive/v0.1.1.tar.gz";
% Use the package
pkg load griddingmachine;
update_GM();
art_name = 'VCMAX_2X_1Y_V1';
file_path = query_collection(art_name);
[vcmax,error] = request_LUT(art_name, 35.1, 115.2);
[vcmax,error] = request_LUT(art_name, 35.1, 115.2, 'interpolation', true);
```

```
# Python script
from griddingmachine import update_GM, query_collection, request_LUT;
update_GM();
art_name = "VCMAX_2X_1Y_V1";
file_path = query_collection(art_name);
vcmax,error = request_LUT(art_name, 35.1, 115.2);
vcmax,error = request_LUT(art_name, 35.1, 115.2, interpolation = True);
```

```
# R script
# Install the package
library(devtools);
install_github("Yujie-W/r-griddingmachine");
# Use the Package
library("griddingmachine");
update_GM();
art_name <- "VCMAX_2X_1Y_V1";
file_path <- query_collection(art_name);
results <- request_LUT(art_name, 35.1, 115.2);
results <- request_LUT(art_name, 35.1, 115.2, interpolation = TRUE);
```

scientific data

Check for updates

OPEN

ARTICLE

GriddingMachine, a database and software for Earth system modeling at global and regional scales

Yujie Wang[1]✉, Philipp Köhler[1], Renato K. Braghiere[2,3], Marcos Longo[2,4], Russell Doughty[1,5], A. Anthony Bloom[2] & Christian Frankenberg[1,2]

Land and Earth system modeling is moving towards more explicit biophysical representations, requiring increasing variety of datasets for initialization and benchmarking. However, researchers often have difficulties in identifying and integrating non-standardized datasets from various sources. We aim towards a standardized database and one-stop distribution method of global datasets. Here, we present the GriddingMachine as (1) a database of global-scale datasets commonly used to parameterize or benchmark the models, from plant traits to vegetation indices and geophysical information and (2) a cross-platform open source software to download and request a subset of datasets with only a few lines of code. The GriddingMachine datasets can be accessed either manually through traditional HTTP, or automatically using modern programming languages including Julia, Matlab, Octave, Python, and R. The GriddingMachine collections can be used for any land and Earth modeling framework and ecological research at the regional and global scales, and the number of datasets will continue to grow to meet the increasing needs of research communities.

“Java” style: process once and use everywhere

The GriddingMachine packages, hopefully, will benefit the ESM and ecology communities

Planned Features for GriddingMachine.jl

From GriddingMachine v0.1

- Download and grid MODIS data (GEE as an alternative)
- Download and regrid ERA5 data
- Regrid the datasets

New Functionality (for v0.3+, TBD)

- Query ungridded “raw” data within an area of interest

Acknowledgments

- Christian Frankenberg
- Philipp Köhler
- Renato Braghiere
- Russell Doughty
- Marcos Longo
- Anthony Bloom
- … …
- Dataset owners