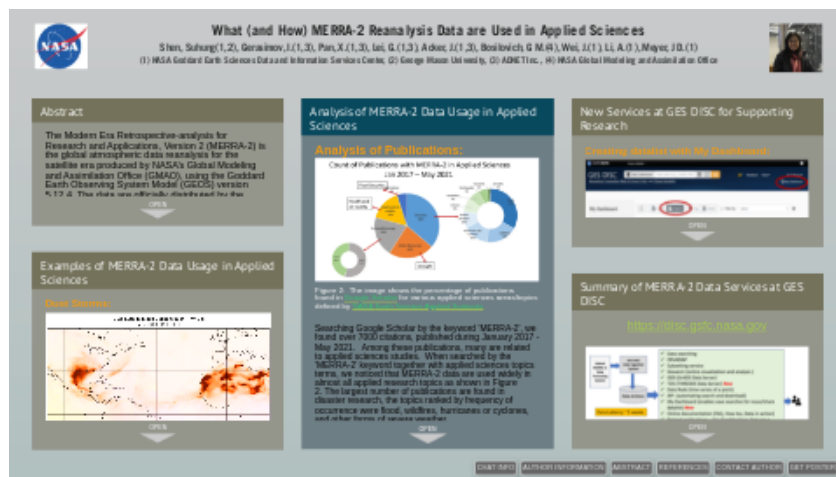
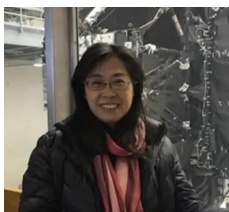


What (and How) MERRA-2 Reanalysis Data are Used in Applied Sciences

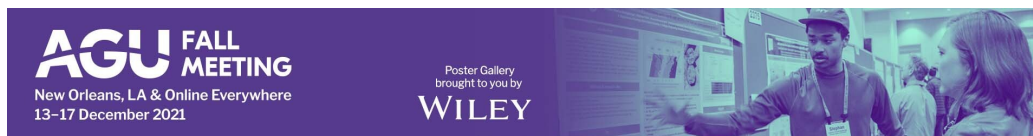


Shen, Suhung^(1,2), Gerasimov, I.^(1,3), Pan, X.^(1,3), Lei, G.^(1,3), Acker, J.^(1,3),
 Bosilovich, G. M.⁽⁴⁾, Wei, J.⁽¹⁾, Li, A.⁽¹⁾, Meyer, J. D.⁽¹⁾

(1) NASA Goddard Earth Sciences Data and Information Services Center, (2) George Mason University,
 (3) ADNET Inc., (4) NASA Global Modeling and Assimilation Office



PRESENTED AT:



ABSTRACT

The Modern Era Retrospective-analysis for Research and Applications, Version 2 (MERRA-2) is the global atmospheric data reanalysis for the satellite era produced by NASA's Global Modeling and Assimilation Office (GMAO), using the Goddard Earth Observing System Model (GEOS) version 5.12.4. The data are officially distributed by the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC). MERRA-2 data have been widely used by the Earth sciences and application community. Since MERRA-2 data were released in early 2016, the number of registered data users has grown steadily from 1,252 in 2016 to 6477 in 2020. By the end of October 2021, ~16 petabytes (over 360 million files) of data have been distributed to more than 18,900 users. Searching in Google Scholar, we have found over 7,000 articles, published between Jan 2017 and May 2021, involving the use of MERRA-2 data.

In this presentation, we will discuss the preliminary findings from a review of selected literature that uses MERRA-2 data in applied sciences. The current analytic and interoperable data services at GES DISC are listed, such as the on-the-fly subset and analysis service, NASA Giovanni; THREDDS Data Server (TDS); and Python Jupyter notebooks. In addition, we will introduce two new services for supporting the open sciences: **My Dashboard** and **Related Publications**.

Click to see all MERRA-2 data collections (<https://disc.gsfc.nasa.gov/datasets?project=MERRA-2>)

EXAMPLES OF MERRA-2 DATA USAGE IN APPLIED SCIENCES

Dust Storms:

[VIDEO] https://res.cloudinary.com/amuze-interactive/video/upload/vc_auto/v1638230222/agu-fm2021/30-EA-4D-D4-45-E4-89-5B-91-32-F6-F0-48-02-D4-E9/Video/dust-animated-gif_uk17e4.mp4

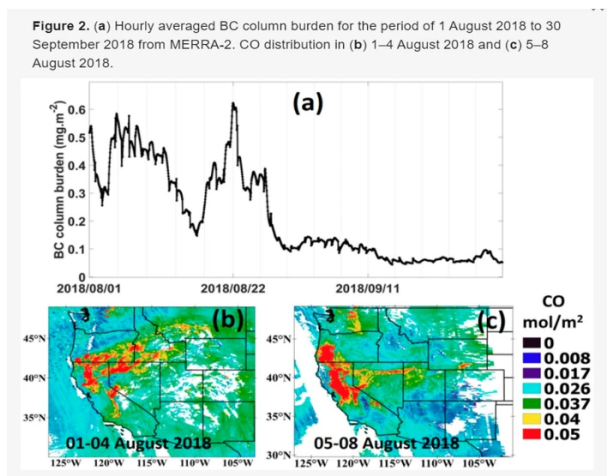
Figure 1: This video shows the transport of the Godzilla dust cloud from the Sahara Desert to the US in June 2020. The image shows the MERRA-2 PM_{2.5} surface dust mass concentration, unit=kg/m³

Wildfires:

Lerato (<https://doi.org/10.3390/atmos12010011>)Shikwambana (<https://doi.org/10.3390/atmos12010011>), et al (2021)

(<https://doi.org/10.3390/atmos12010011>) investigated smoke from large wildfires that may travel long distances and have a harmful effect on human health, environment, and climate in other areas. This study analyzed the spatial-temporal distribution of black carbon (BC) and carbon monoxide (CO) and the vertical distribution of smoke with data from MERRA-2, Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), and TROPOspheric Monitoring Instrument (TROPOMI) for the following data collections:

"M2T1NXAER", "M2TMNXSLV", "M2TMNXFLX", "S5P_L2_CO____", "AIRS2RET"



Renewable Energy:

Maria

(https://research.chalmers.se/publication/512775/file/512775_Fulltext.pdf)Taljegard

(https://research.chalmers.se/publication/512775/file/512775_Fulltext.pdf) et al.

(https://research.chalmers.se/publication/512775/file/512775_Fulltext.pdf)(2019)

have modeled different regions in Europe with different conditions for renewable energy (wind, solar, and hydropower) by using the hourly MERRA-2 meteorological data for the solar condition [M2T1NXRAD_5.12.4] and wind [M2T1NXSLV_5.12.4]).

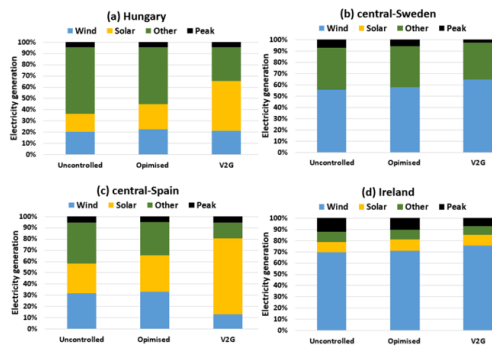


Figure3: Electricity generation by power sources for the four regions (a) Hungary, (b) Ireland, (c) Spain and (d) Sweden.

Tropical Cyclones:

Kevin Hodges, et al. (<https://journals.ametsoc.org/view/journals/clim/30/14/jcli-d-16-0557.1.xml>)(2017) have identified and tracked tropical cyclones (TCs) using MERRA-2 along with five other reanalysis datasets and compared them with those from the IBTrACS best-track archive.

TABLE 2.
The POD for the NH and SH for the direct matching method applied to the reanalysis tracks (cf. section 3a) and the POD and FAR for the NH and SH based on the objective detection method [cf. section 3b(2)].

	ERA-I	JRA-25	JRA-55	NCEP-CFSR	MERRA	MERRA-2
				POD		
NH direct match	0.95	0.95	0.95	0.95	0.95	0.95
NH objective	0.60	0.76	0.80	0.70	0.51	0.67
SH direct match	0.93	0.93	0.94	0.93	0.90	0.93
SH objective	0.76	0.84	0.87	0.83	0.61	0.79
				FAR		
NH objective	0.28	0.16	0.29	0.36	0.21	0.36
SH objective	0.60	0.43	0.58	0.58	0.54	0.63

ANALYSIS OF MERRA-2 DATA USAGE IN APPLIED SCIENCES

Analysis of Publications:

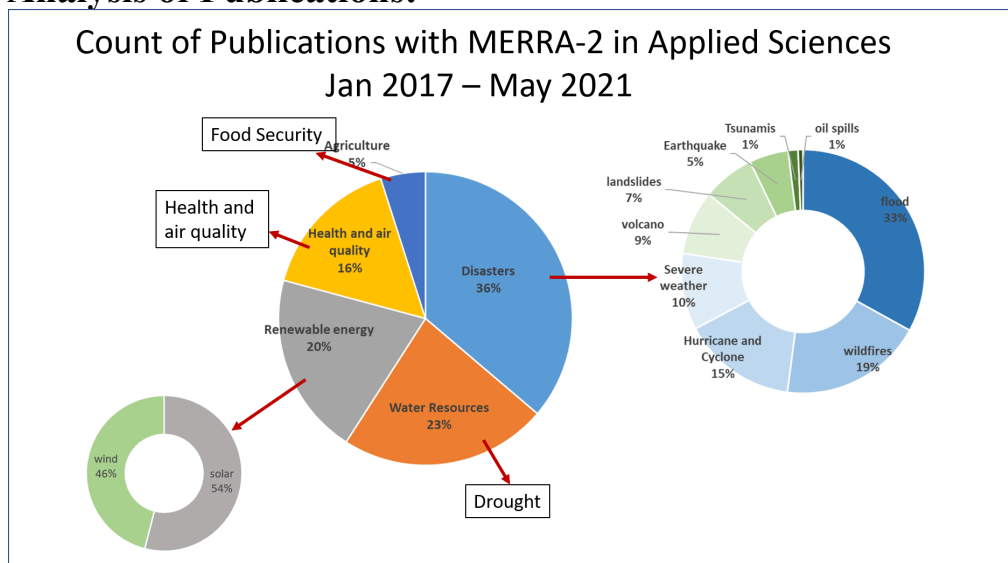


Figure 2: The image shows the percentage of publications found in [Google Scholar](https://scholar.google.com) (<https://scholar.google.com>) for various applied sciences areas/topics defined by [NASA](https://appliedsciences.nasa.gov/) (<https://appliedsciences.nasa.gov/>) [Earth Science Applied Sciences](https://appliedsciences.nasa.gov/EarthScienceAppliedSciences) (<https://appliedsciences.nasa.gov/>)

Searching Google Scholar by the keyword 'MERRA-2', we found over 7000 citations, published during January 2017 - May 2021. Among these publications, many are related to applied sciences studies. When searched by the 'MERRA-2' keyword together with applied sciences topics terms, we noticed that MERRA-2 data are used widely in almost all applied research topics as shown in Figure 2. The largest number of publications are found in disaster research, the topics ranked by frequency of occurrence were flood, wildfires, hurricanes or cyclones, and other forms of severe weather.

To further improve the GES DISC data services, our interests let us investigate the most popular data collections and variables that have been used in various applied science areas. The preliminary analysis was done with publications that acquired data through the NASA Giovanni system, where all MERRA-2 monthly and selected sub-daily data are available. More than 200 publications referencing Giovanni services published from January 2016 to September 2021, were collected. To analyze these publications, we came up with a set of terms for each application area/topic. Then, we applied Natural Language Processing extraction of these terms from titles and abstracts for all collected publications and associated each publication with the corresponding topics and research areas.

Figure 3 shows, in the "Air Quality" research topic, the top five data collections used in Giovanni are M2TMNXAER, M2T1NXAER, M2TMNXFLX, M2TMNXADG, and M2TMNXRAD. Looking into the variables, we found that the 'black carbon surface mass concentration' is used most frequently, followed by 'dust surface mass concentration', and 'surface wind speed'.

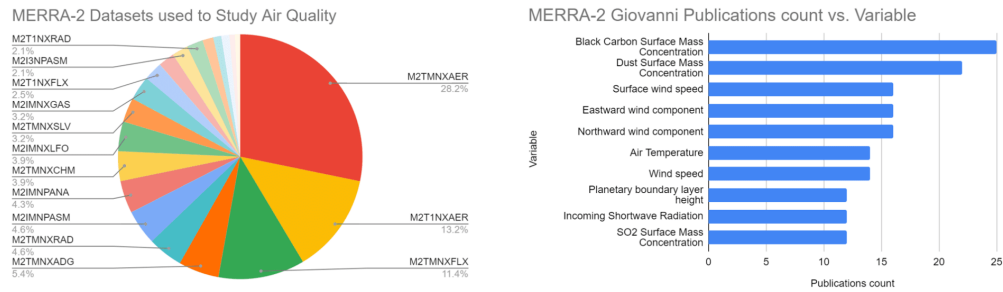


Figure 3: The left plot shows the number of publications grouped by the data collections in the "Air Quality" research topic. The right plot shows the top ten MERRA-2 variables used in the collected publications.

MERRA-2 Data Distributions:

Data usage is analyzed with the distribution metrics shown in Figure 4. The number of users has grown each year. Since 2000, about 4 million files have been distributed to more than 1000 users each month. A significant amount of files (18%) have been distributed to commercial users. The variable distribution metrics indicate that the wind at 50 meters, surface pressure, the temperature at 2 meters, and wind at 10m are of the most interest to commercial users.

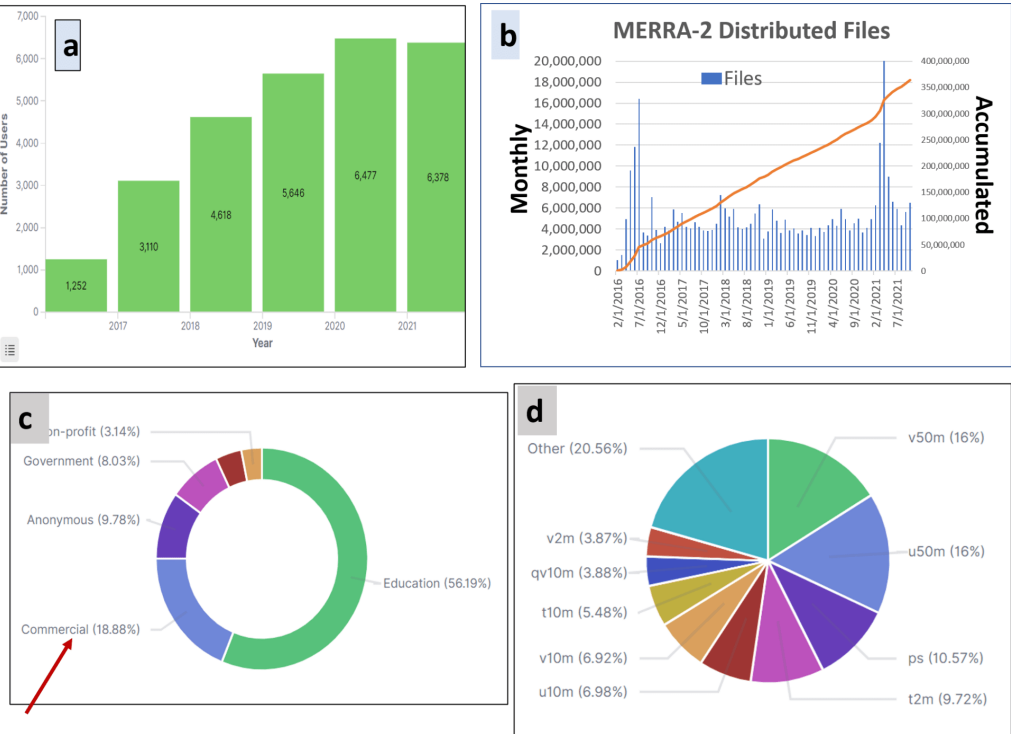


Figure 4: MERRA-2 data distribution metrics from 2016 to October 2021: a) annual unique users; b) monthly and accumulated total distributed data files; c) percentiles grouped by user affiliation; d) the top accessed variables by commercial users.

NEW SERVICES AT GES DISC FOR SUPPORTING RESEARCH

Creating datalist with My Dashboard:

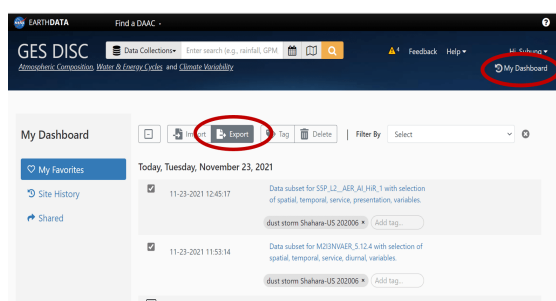


Figure 5: Example of the interface of "My Dashboard"

Questions often received from users are similar to "Which data I should use?", "Which service I can use to get what I am interested in?". We understand that finding the most relevant data and documentation from various sources, as well as working through the data downloading service may take a significant time. Recently, the GES DISC has released a new service, called **My Dashboard** as shown in Figure 5. It enables a user to:

- save selected variables and subset options from various data collections
- save the selected documentation
- form a **datalist** and export it to a JSON file

This saved JSON file could be:

- imported back to the system later for accessing similar data at different regions or time periods which will save time to go through searching and selection again
- shared with other scientists to conduct a similar study, which supports open science in terms of finding and accessing data

Use Case: Creating a datalist to investigate the June 2020 dust storm transported from Sahara to US

In June 2020, a vast expanse of dust, known as the Saharan Air Layer (SAL), was observed. It is also called the "Godzilla dust cloud" as well as the "most significant event in the past 50 years" by some experts. To evaluate the air pollution of this event, we were interested in viewing the dust PM_{2.5} as the surface mass concentration, the column mass density, and the dust vertical profile. We have created a datalist as shown below:

1. selected variables from two MERRA-2 collections:

M2T1NXAER.5.12.4 (<https://doi.org/10.5067/KLICLTZ8EM9D>)

DUSMASS25 = dust surface mass concentration - pm 2.5

DUCMASS25 = dust column mass density - pm 2.5

M2I3NVAER.5.12.4 (10.5067/LTVB4GPCOTK2)

DU001, DU002, DU003 = dust mixing ratio (bin 001, 002, 003) [for *dust bin size* range: 0.1-1, 1-1.8, 1.8-3, 3-6, 6-10 (μm), respectively]

• **Date Range:** 2020-06-165 to 2020-07-02

•**Region:** -120.0, -8.0, 8.0, 50.0 (Search and Crop)

2. selected satellite observation data from Sentinel-5P TROPOMI for evaluating the reanalysis data

S5P_L2_AER_AI_HiR (10.5270/S5P-0wafvaf)

aerosol_index_354_388 = Aerosol index from 388 and 354 nm)

•**Date Range:** 2020-06-25 to 2020-06-26

•**Region:** -95.5, 12.1, -65.0, 41.6

3. go to 'My Dashboard ==> Site History', mark the relevant items as 'My Favorites'

4. go to 'My Favorites', tag each item as 'dust storm Shahara-US 202006' as in Figure 7

5. select all relevant items, click on 'Export' to save to a JSON file

Sample plots using the downloaded data are shown in Figures 1, and 6, 7, 8.

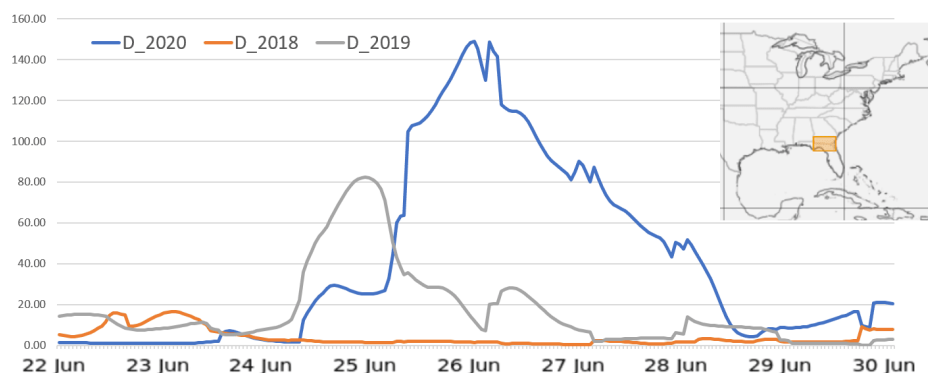


Figure 6: MERRA-2 MERRA-2 Dust Surface Mass Concentration - PM 2.5 [mg/m3], M2T1NXAER v5.12.4, Region 85.1W, 29.5N, 81.4W, 31.9N. The concentration in June 22 - July 2 of 2020 (blue) is much larger than for other years, such as 2018 (orange), 2019 (gray) in the plot.

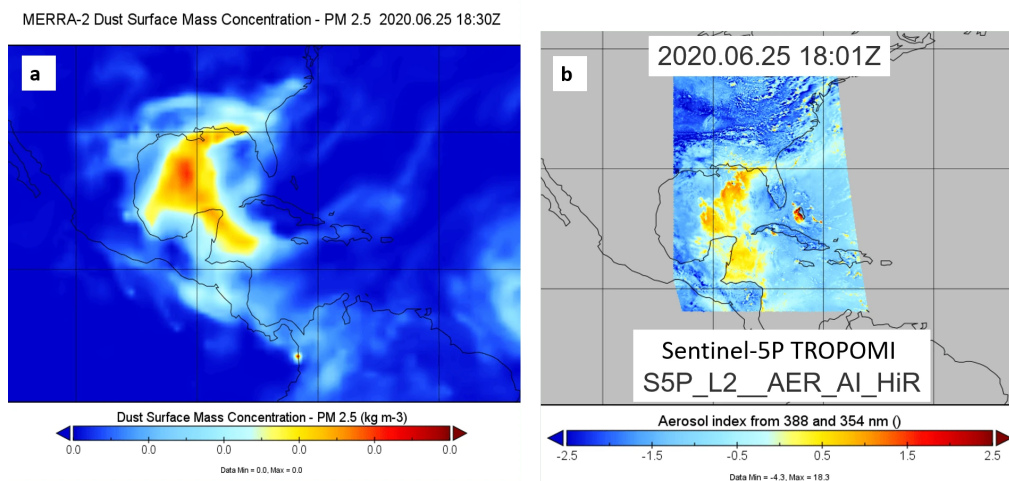


Figure 7: a) MERRA-2 dust surface mass concentration -PM2.5, 2020.06.25 18:30Z; b) Sentinel-5P TROPOMI Aerosol Index at 354-388 nm, 2020.06.25 18:01Z. The comparison indicates that the assimilated dust location in MERRA-2 matches well with the satellite observation.

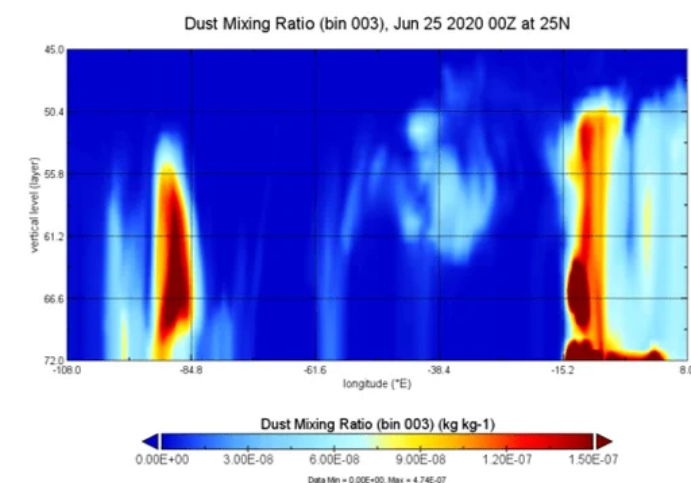


Figure 8: Image of longitude-vertical cross-section map at 25°N from MERRA-2 3-hourly dust mixing ratio (bin 003), M2I3NVAER.5.12.4 (10.5067/LTVB4GPCOTK2)_DU003 for June 25, 2020, 00Z. The dust was observed at model level 55 (~ 4km) and above.

Learning Data Usage from Related Publications and Data DOI Citations:

Related Publications: This is a list of publications collected from journals by GES DISC, that use one or more variables in the data collection.

This is another new service at GES DISC supporting research and open sciences. It allows one to explore easily how the data collection was used in various research areas. The link in the list points to the online publication resource.

From a dataset landing page, click on the 'References' tab for a list of Related Publications as shown in Figure 9.



Figure 9: Example of the list of 'Related Publication' on the dataset landing page of M2T1NXSLV_12.5.4

Advantages of Citing Data with DOI in Publication:

a) It helps the reader to know exactly which data are used in the research work.

- b) It is easier to find the dataset landing page for downloading data and obtaining data information using the data DOI.
- c) It will give credit to the data producers.
- d) It has a better chance to be collected by GES DISC and be listed in the 'Related Publications' if data DOI is cited. This will increase the visibility of the research work

How to Find and Cite Data DOI:

All MERRA-2 (and the majority of data at GES DISC) have data DOIs, which can be found on the dataset landing page under "Data Citation". For example, the data DOI of M2T1NXSLV_5.12.4 is

10.5067/VJAFPLI1CSIV (<https://doi.org/10.5067/VJAFPLI1CSIV>).

Click the above DOI to view how to cite the data.

SUMMARY OF MERRA-2 DATA SERVICES AT GES DISC

<https://disc.gsfc.nasa.gov> (<https://disc.gsfc.nasa.gov>)

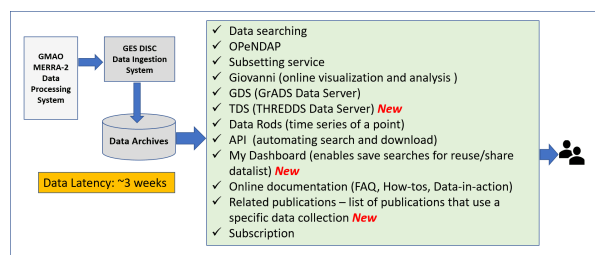


Figure 10: MERRA-2 Data flow and the list of data services at GES DISC

• Find documentation and all data access methods from the **dataset landing page** as shown in Figure 11. Click here (<https://doi.org/10.5067/VJAFPLIICSIV>) to see the dataset landing page for the collection: M2T1NXSLV_5.12.4

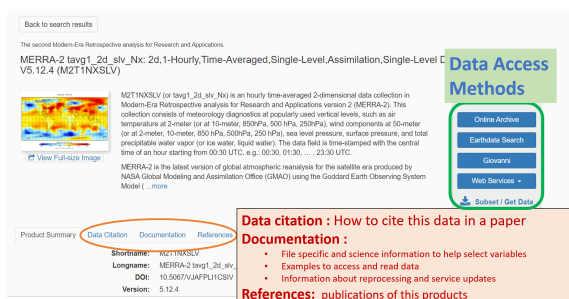


Figure 11: Example of the dataset landing page

• Read MERRA-2 Data Access – Quick Start Guide (<https://disc.gsfc.nasa.gov/information/documents?title=MERRA-2%20Data%20Access%20%E2%80%93%20Quick%20Guide>)

• Read more online help documentation (How- (<https://disc.gsfc.nasa.gov/information/howto?keywords=%22MERRA-2%22&page=1>) tos (<https://disc.gsfc.nasa.gov/information/howto?keywords=%22MERRA-2%22&page=1>) and FAQs (<https://disc.gsfc.nasa.gov/information/faqs?keywords=%22MERRA-2%22&page=1>)) to select the best data access methods, e.g.

1. Downloading the subset of a file with only interested variables and/or region (<https://disc.gsfc.nasa.gov/information/howto?keywords=Level%203&title=How%20to%20use%20the%20Level%203%20and%204%20Subsetter%20and%20R>

2. Downloading the daily mean (<https://disc.gsfc.nasa.gov/information/howto?keywords=daily&title=How%20to%20Download%20MERRA-2%20Daily%20Mean%20Data>)

3. Getting the automated data searching and downloading with the usage of API (<https://disc.gsfc.nasa.gov/information/howto?keywords=MERRA-2%20API&title=How%20to%20Use%20the%20Web%20Services%20API%20for%20Subsetting%20MERRA-2%20Data>)

4. Downloading the time series in a single file for a grid point or a region (<https://disc.gsfc.nasa.gov/information/howto?keywords=TDS&page=1>)

5. Finding examples of Python (<https://disc.gsfc.nasa.gov/information/howto?keywords=python&page=1>) jupyter (<https://disc.gsfc.nasa.gov/information/howto?keywords=python&page=1>) notebooks (<https://disc.gsfc.nasa.gov/information/howto?keywords=python&page=1>)

Question on data access, email to

GES DISC Help Desk

(gsfc-dl-help-disc@mail.nasa.gov)

Question on science, email to

MERRA-2 science team

(merra-questions@lists.nasa.gov)

AUTHOR INFORMATION

Suhung Shen^{1,2}, Irina Gerasimov^{1,3}, Xiaohua Pan^{1,3},
Guang-Dih Lei^{1,3}, James, Acker^{1,3}, Michael G Bosilovich⁴,
Jennifer Wei¹, Angela Li¹, David J Meyer¹

1 NASA Goddard Earth Sciences Data and Information
Services Center (GES DISC)

2 George Mason University

3 ADNET Inc.

4 NASA Global Modeling Assimilation Office

ABSTRACT

The Modern Era Retrospective-analysis for Research and Applications, Version 2 (MERRA-2) is the global atmospheric data reanalysis for the satellite era produced by NASA's Global Modeling and Assimilation Office (GMAO), using the Goddard Earth Observing System Model (GEOS) version 5.12.4. The data are officially distributed by the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC). MERRA-2 data have been widely used by the Earth sciences and application community. Since MERRA-2 data were released in early 2016, the number of registered data users has grown steadily from 1,252 in 2016 to 6477 in 2020. By the end of October 2021, ~16 petabytes (over 360 million files) of data have been distributed to more than 18,900 users. Searching in [Google Scholar](https://scholar.google.com/) (<https://scholar.google.com/>), we have found over 7,000 articles, published between January 2017 and May 2021, involving the use of MERRA-2 data. The figure shows the numbers for various application areas in which the MERRA-2 data have been used, covering almost all of the application areas defined in [NASA Applied Sciences](http://appliedsciences.nasa.gov) (<http://appliedsciences.nasa.gov>). The largest number of articles are found in disaster research, with the subcategories ordered in flood, wildfires, hurricanes and cyclones, and other forms of severe weather.

In this presentation, we will discuss the preliminary findings from a review of the selected literature that uses MERRA-2 data in applied sciences. The current analytic and interoperable data services at GES DISC are listed, such as the on-the-fly subset and analysis service, NASA Giovanni; THREDDS Data Server(TDS); and Python Jupyter notebooks. In addition, we will introduce two new services for supporting the open sciences: **My Dashboard** and **Related Publications**.

Click to see all MERRA-2 data collections (<https://disc.gsfc.nasa.gov/datasets?project=MERRA-2>)

REFERENCES

Shikwambana, L.; Kganyago, M. Observations of Emissions and the Influence of Meteorological Conditions during Wildfires: A Case Study in the USA, Brazil, and Australia during the 2018/19 Period. *Atmosphere* **2021**, *12*, 11. <https://doi.org/10.3390/atmos12010011>

Maria Taljegard , Lisa Göransson, Mikael Odenberger and Filip Johnsson, Electric vehicles as power and energy provider for the European electricity system - an electricity systems modelling study, 32nd Electric Vehicle Symposium (EVS32) Lyon, France, May 19 - 22, 2019

Kevin Hodges (https://journals.ametsoc.org/search?f_0=author&q_0=Kevin+Hodges), Alison Cobb (https://journals.ametsoc.org/search?f_0=author&q_0=Alison+Cobb), and Pier Luigi Vidale (https://journals.ametsoc.org/search?f_0=author&q_0=Pier+Luigi+Vidale), How Well Are Tropical Cyclones Represented in Reanalysis Datasets?, J. of Climate, Vol. 30, No.14, pp 5243–5264, DOI: <https://doi.org/10.1175/JCLI-D-16-0557.1> (<https://doi.org/10.1175/JCLI-D-16-0557.1>)