On-line Access of MERRA Data: From the Perspective of a User

Arlindo da Silva

Global Modeling and Assimilation Office, NASA/GSFC

Arlindo.daSilva@nasa.gov

MERRA User's Workshop
AMS Annual Meeting, Phoenix, AZ
11 January 2009

Outline

- Overview of MERRA collections
- Using the OPeNDAP servers
  - C and Fortran
  - GrADS and the Visualization Tool
  - Python
  - CDAT
  - Perl
  - Matlab, IDL
  - Others: PHP, TCL, NCO, R
- Concluding Remarks

Premise of the talk

- □ While on-line visualization tools play an important role during exploratory data analysis, most in-depth studies require detailed scrutiny by one or more sophisticated analysis/visualization tools.
- □ Researchers by now already have their favorite analysis tools.
  - Meet users at their element.
- □ Review how MERRA data can be accessed from a variety of such applications.
  - Focus will be on OPeNDAP access

Some definitions...

- MERRA data is organized in *collections*, each collection containing a specific group of *variables*:
  - Collection MAI3CPASM contains these variables:
    - slp, ps, u, v, T, q, O_3, etc
  - The term *dataset* will be used to mean *collection*
- For the time period covered by MERRA, each collection comprises of a large number of *files*:
  - Hourly data: one file per day
  - Monthly/diurnal data: one file per month

Authoritative Reference

- When in doubt about units, grid definition, averaging period, etc., consult the MERRA File Spec available from:
  - http://gmao.gsfc.nasa.gov/merra/
- Please let us know about any discrepancy between the File Spec document and metadata found in files/OPeNDAP collections.
  - Michael.Bosilovich@nasa.gov

Resolutions

- 42 Vertical constant pressure levels:□□

  1000, 975, 950, 925, 900, 875, 850, 825, 800, 775, 750, 725, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 150, 100, 70, 50, 40, 30, 20, 10, 7, 5, 4, 3, 2, 1, 0.7, 0.5, 0.4, 0.3, 0.1

- Full horizontal resolution:
  - 2/3° longitude, 1/2° latitude
  - Ranges: [180W,180E)
    [90S,90N]
- Reduced I horizontal resolution (pressure)
  - 1.25° longitude, 1.25° latitude
  - Ranges: [179.375W,179.375E],
    [89.375S,89.375N]

Note: Reduced Grid I like GPCP, but unlike NCEP/ERA lat/lon grids.

Resolutions, cont

- 72 native model layers, top at 0.01 hPa
- Reduced II horizontal resolution (eta coord)
  - 1.25º longitude, 1º latitude
  - Ranges: [180W,180E),
    [90S,90N]

Note: Reduced II grid like in GEOS-4; used in Chemical Transport Modeling colections.

MERRA Collections on OPeNDAP

Available on 3 OPeNDAP servers:

- □ 3D fields for Chemical Transport Modeling
  - ■ http://goldsmr1.sci.gsfc.nasa.gov/dods/
- □ 2D fields: prognostic, diagnostics
  - ■ http://goldsmr2.sci.gsfc.nasa.gov/dods/
- □ 3D fields: analysis and assimilation fields
  - ■ http://goldsmr3.sci.gsfc.nasa.gov/dods/
- □ Each sever has data at several time steps
  - ■ Hourly, 3-hourly, 6-hourly
  - ■ monthly, mean diurnal cycle

QuickTime™ and a
decompressor
are needed to see this picture.

MERRA Collections, cont.

A convenient way to organize the collections is by time steps:

- hourly: fluxes, budget tems (2D)
- 3-hourly: assimilation, physics diags (3D)
- 6-hourly: analysis (3D)
- monthly (2D and 3D)
- mean diurnal (2D and 3D)
- Resolutions:
  - All 2D datasets are given in FULL resolution
  - All 3D datasets are given in REDUCED resolution
    - The only exception is the 6-hourly analysis

Hourly Collections (2D)

	Short		
Nickname	Name	Product	Brief Description
Meteorology	met	MAT1NXSLV	Single level met fields
Surface	sfc	MAT1NXFLX	Surface fluxes
LandSurface	lnd	MAT1NXLND	Land surface diagnostics
Radiation	rad	MAT1NXRAD	Radiation fluxes: sfc & toa
Cloud	cld	MAI1NXINT	Column cloud properties
Budget	bgt	MAT1NXINT	Column budget terms

FULL Horizontal Resolution

3-hourly Collections (3D)

	Short		
Nickname	Name	Product	Brief Description
Assimilation	asm	MAI3CPASM	Instantaneous assimilated state
Clouds	cld	MAT3CPCLD	Cloud properties
MoistPhysics	h2o	MAT3CPMST	Moist physics diagnostics
Radiation	rad	MAT3CPRAD	Cloud/radiation diagnostics
Turbulence	trb	MAT3CPTRB	Turbulence diagnostics
T_Tendency	d_t	MAT3CPTDT	Temperature tendencies
q_Tendency	d_q	MAT3CPQDT	Specific humidity tendencies
uv_Tendency	d_u	MAT3CPUDT	Wind tendencies
O3_Tendency	d_o	MAT3CPODT	Ozone tendencies

Note: Chem transport collections omitted for conciseness.

REDUCED Horizontal Resolution

6-hourly Collection (3D)

	Short		
Nickname	Name	Product	Brief Description
Analysis	ana	MAI6CPANA	Instantaneous analyzed state

FULL Horizontal Resolution

Monthly 2D Collections

	Short		
Nickname	Name	Product	Brief Description
Meteorology	met	MATMNXSLV	Single level met fields
Surface	sfc	MATMNXFLX	Surface fluxes
LandSurface	lnd	MATMNXLND	Land surface diagnostics
Radiation	rad	MATMNXRAD	Radiation fluxes: sfc & toa
Cloud	cld	MAI1NXINT	Column cloud properties
Budget	bgt	MATMNXINT	Column budget terms

Note: these are averages of the hourly collections.

FULL Horizontal Resolution

Monthly 3D Collections

	Short		
Nickname	Name	Product	Brief Description
Analysis	ana	MAIMNPANA	Instantaneous analyzed state
Assimilation	asm	MAIMCPASM	Instantaneous assimilated state
Clouds	cld	MATMCPCLD	Cloud properties
MoistPhysics	h2o	MATMCPMST	Moist physics diagnostics
Radiation	rad	MATMCPRAD	Cloud/radiation diagnostics
Turbulence	trb	MATMCPTRB	Turbulence diagnostics
T_Tendency	d_t	MATMCPTDT	Temperature tendencies
q_Tendency	d_q	MATMCPQDT	Specific hmidity tendencies
uv_Tendency	d_u	MATMCPUDT	Wind tendencies
O3_Tendency	d_o	MATMCPODT	Ozone tendencies

Note: these are averages of the 3- and 6-hourly collections.

REDUCED Horizontal Resolution, except for MAIMNANA

Using the OPeNDAP Servers

- OPeNDAP is a data server architecture that allows users to use data files that are stored on remote computers with their favorite analysis and visualization tools.
- Opening an OPeNDAP file is as easy as entering an OPeNDAP URL into the interface the client software where it expects a local file name to be entered.
- Every MERRA collection that is provided by FTP is also available via OPeNDAP.

Main OPeNDAP Services

Information

This service is activated when the server receives a URL ending with .info.

Returns information about the server and dataset, in human-readable HTML form.

Data Descriptor

This service is activated when the server receives a URL ending with .dds.

This is a text file describing the structure of the variables in the dataset.

Data Attribute

This service is activated when the server receives a URL ending with .das.

This is a text file describing the attributes of each data quantity in that dataset.

OPeNDAP Data

This service is activated when the server receives a URL ending with .dods

This service returns the actual data requested by a given URL.

ASCII Data

This service returns an ASCII representation of the requested data.

Activated when the server receives a URL ending with .asc or .ascii.

Quick Demo: http://goldsmr3.sci.gsfc.nasa.gov/dods/

C/Fortran Programming

- Although there is a C++ OPeNDAP client library, the most commonly used client library is nc-dap, an extension of the NetCDF API that also understands OPeNDAP URLs.
- Your own C/FORTRAN program can be an OPeNDAP client. A C/FORTRAN program that uses standard NetCDF library calls can be linked with the OPeNDAP client library. An OPeNDAP URL can then be used in place of the the file name in the NC_OPEN call.
- There also native implementation of OPeNDAP in Java (NetCDF-Java) and Python (PyDAP).

OPeNDAP Client Applications

- GrADS
- Python
- Perl
- PHP and TCL
- Ferret
- NCL
- Octave
- Matlab
- IDL
- Others: NCO, R

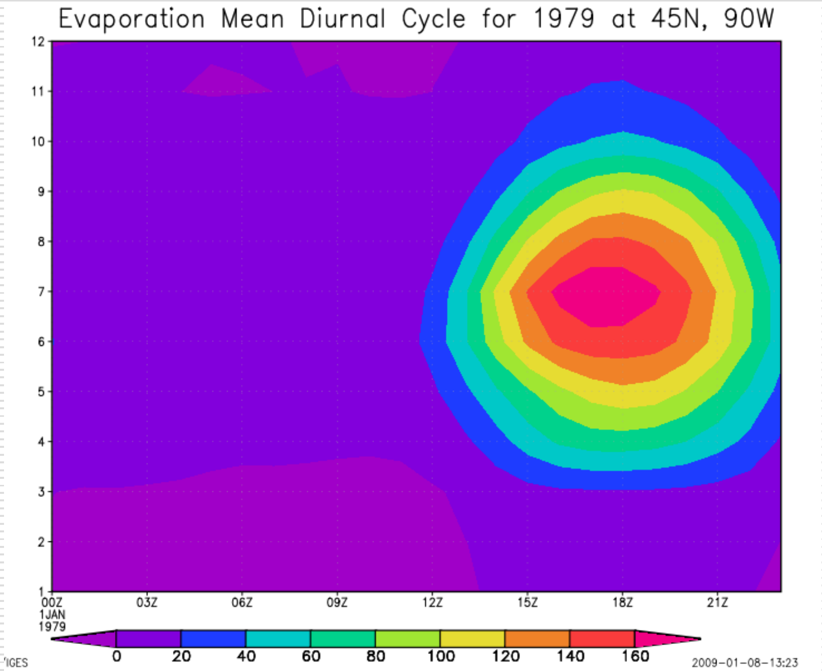
GrADS

- The Grid Analysis and Display System (GrADS) is an interactive desktop tool for easy access, manipulation, and visualization of earth science data
- The format of the data may be either binary, GRIB-1, GRIB-2, NetCDF, OPeNDAP or HDF-4
- GrADS implements a 5-Dimensional data environment: longitude, latitude, vertical level, time, and starting in Version 2, the 5th ensemble dimension.
- GrADS is primarily a CLI tool, but it has capabilities for a GUI as well.

GrADS Ensemble Dimension

- New in Version 2.0
- Extremely useful for dealing with MERRA’s mean diurnal collections
- Example:

```
ga-> merra_diurnal sfc
ga-> set lat 45
ga-> set lon -90
ga-> set e 1 12
ga-> display evap
```



Note: OPeNDAP diurnal collections not working yet.

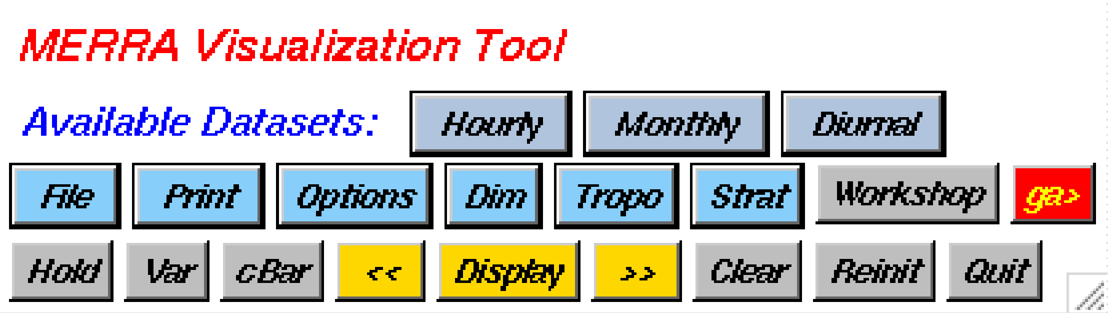
Easy MERRA data access in GrADS

- The following scripts are included in your USB memory stick to facilitate opening the MERRA collections:
  - merra_hourly.gs, merra_3hourly.gs, merra_6hourly.gs
  - merra_monthly2d.gs, merra_monthly3d.gs
- Example. All these are equivalent:

```
ga-> merra_monthly ana
ga-> merra_monthly analysis
ga-> merra_monthly MAIMNPANA
```

MERRA Visualization Tool

- A GrADS-base GUI has been developed to facilitate the hands-on demos this afternoon
  - It is included in your USB memory stick



Quick Demo

Additional GrADS resources

- We have started a MERRA section in the OpenGrADS Cookbooks:

  http://cookbooks.opengrads.org/

- Several of the examples this afternoon will appear in the Cookbooks.

Python

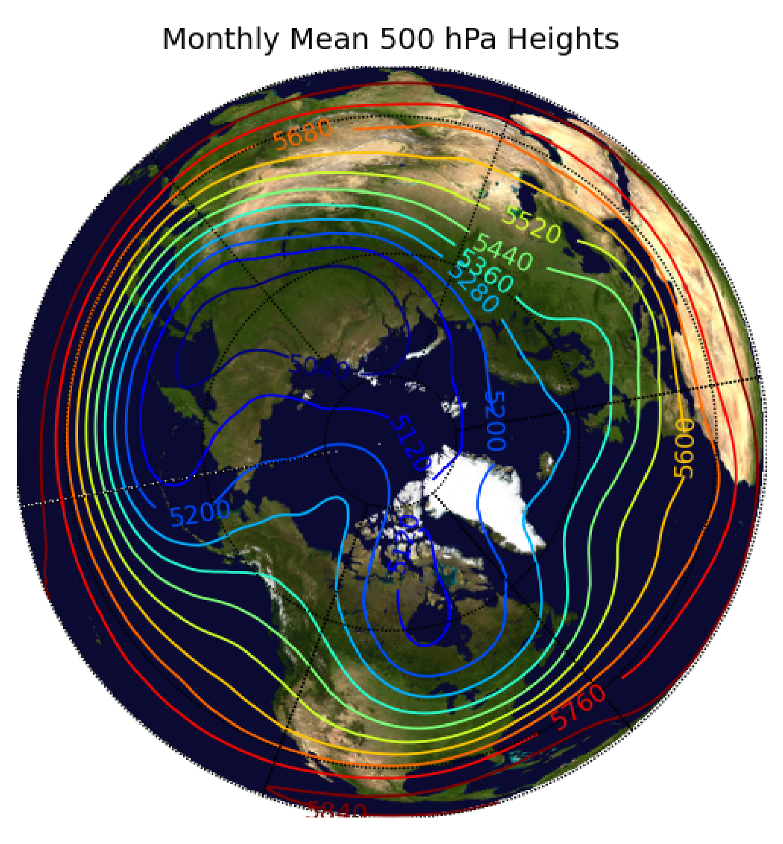
- Python is a general-purpose, high-level programming language.
  - Its design philosophy emphasizes programmer productivity and code readability.
  - Python supports multiple programming paradigms (primarily object oriented, imperative, and functional) and features a fully dynamic type system and automatic memory management, similar to Perl and Ruby.
  - Python is primarily used as a scripting language.
- NumPy and PyLAB are Python packages that offer a large set of software for scientific computing, much of it modeled after MATLAB.

Python OPeNDAP Clients

- PyDAP is a Python implementation of the OPeNDAP.
  - Using pydap as a client you can access access the MERRA collections
  - You can also use it to easily serve your data from a variety of formats.
  - http://pydap.org/
- PyGrADS is an Python interface to GrADS that extends it by providing
  - A more versatile scripting language
  - A large collection of mathematical and statistical software (e.g., “R”)
  - Advanced visualization tools, including 3D volumetric visualization
  - http://opengrads.org/

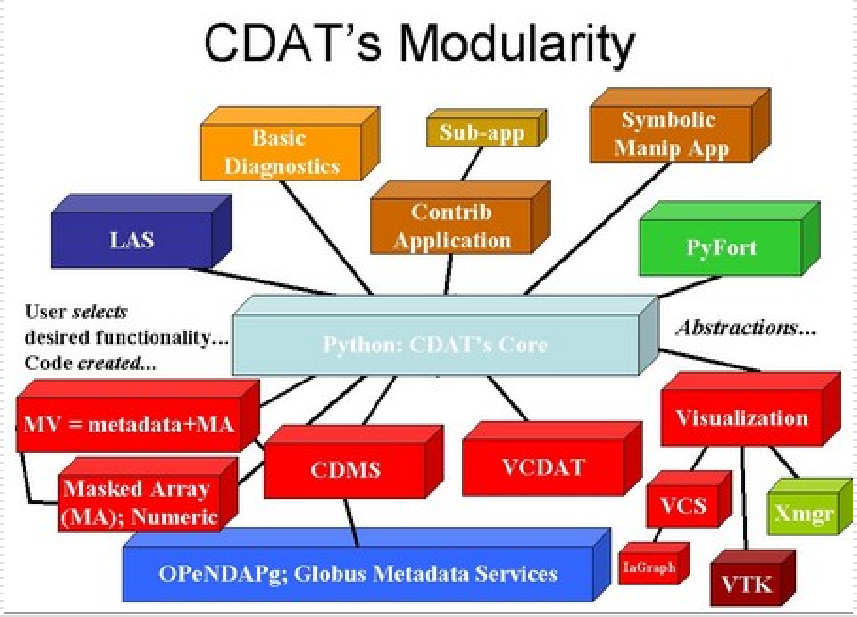
Using PyGRADS to access MERRA data

```
% pygrads -dbl
[1] ga-> . merra_monthly3d asm
[2] ga-> s lev 500
[3] ga-> ga.blue_marble(‘on’)
[4] ga-> ga.basemap(‘npo’)
[5] ga-> ga.contour(‘h’,N=12)
[6] ga-> title('Monthly Mean
    500 hPa Heights')
```



CDAT
Climate Data Analysis Tools

- CDAT makes use of an open-source, object-oriented, easy-to-learn scripting language (Python) to link together separate software subsystems and packages to form an integrated environment for data analysis.



Perl

- Perl is a high-level, general-purpose, interpreted, dynamic programming language
- Perl provides powerful text processing
- It is also used for
  - graphics programming, system administration, network programming, applications that require database access and CGI programming on the Web.
- The Perl Data Language (PDL) adds efficient numerical array capabilities
  - However, scientific computing not as advanced as in Python

Perl OPeNDAP Clients

- Could not find a specific OPeNDAP client
- Gerl: Perl interface to GrADS
  - Similar to PyGrADS but currently not actively maintained
  - There are two basic modules implementing the Perl interface to the GrADS application:
    - Grads.pm: OO interface
    - Gerl.pm: procedural interface
    - http://opengrads.org
  - Example:

```
use Grads::Gerl;
grads {Bin=>"gradsdap", Window=>1 };
ga_ merra_hourly;
display eflux;
quit;
```

QuickTime™ and a
decompressor
are needed to see this picture.

NCL Script Example

```
begin
  ; The URL is so long, break it into two pieces.
    url      = "http://goldsmr2.sci.gsfc.nasa.gov/dods/"
    filename = "MAT1NXFLX” ; hourly surface fluxes

    exists = isfilepresent(url+filename)
    if(.not.exists) then
      print("OPeNDAP test unsuccessful.")
       print("Either file doesn't exist, or NCL does not have OPeNDAP /
       capabilities on this system")
    else
      f = addfile(url + filename,"r")
      variables = getfilevarnames(f)

      print(variable) ; will print a list of variable names on file
    end if
  end
```

QuickTime™ and a
decompressor
are needed to see this picture.

QuickTime™ and a
decompressor

```
yes? SET DATA “http://goldsmr2.sci.gsfc.nasa.gov:80/dods/MAT1NXFLX”
yes? SHADE evap
```

QuickTime™ and a
decompressor
are needed to see this picture.

QuickTime™ and a
decompressor
are needed to see this picture.

Note: There is a MATLAB interface to GrADS along the lines of PyGrADS. To be released soon.

```
ga = grads('gradsdap -bl',0); % Start GrADS

ga.cmd('merra_hourly sfc'); % open MERRA hourly sfc fluxes
disp fh.title;

[evap,g] = ga.expr('evap'); % retrieve variable
fprintf(1,'Evap: min=%f,  max=%f\n', min(min(evap)), max(max(evap)));
pcolor(g.lon,g.lat,evap); shading('interp');
title('Evaporation'); xlabel('Longitude');  ylabel('Latitude');
```

http://www.gomodp.org/modeling-committee/gulf-of-maine-model-interoperability-pilot-project/how-to-set-up-matlab-to-work-with-the-cf-netcdf-java-tools

QuickTime™ and a

NetCDF-Java Based

QuickTime™ and a
decompressor
are needed to see this picture.

Note: There is an IDL interface to GrADS that is currently not being mantained.

Other tools

- There are similar PHP and TCL interfaces to GrADS that would give you access to the MERRA collections on the OPeNDAP server.
- NCO - NetCDF operatars
- R - Statistical package

Concluding Remarks

- Because OPeNDAP provides an extension of NetCDF that understands URLs, there are a number of tools that are OPeNDAP enabled.
- The next release of NetCDF-4 will include OPeNDAP support build in.
- Therefore, accessing the MERRA collections via OPeNDAP is quite visible from the software point of view
- Main questions:
  - Would it scale?
  - Will performance be acceptable?
  - If not now, when?