



## **Global Modeling and Assimilation Office**

*GMAO Office Note No. 3 (Version 1.2)*

## **The MERRA-Land Data Product**

Release Date: 9/10/2012

---

**Global Modeling and Assimilation Office  
Earth Sciences Division  
NASA Goddard Space Flight Center  
Greenbelt, Maryland 20771**

This page intentionally left blank.

# The MERRA-Land Data Product

Document maintained by Rolf Reichle (GMAO)

This document should be cited as

Reichle, R. H., 2012: The MERRA-Land Data Product. GMAO Office Note No. 3 (Version 1.2), 38 pp, available from [http://gmao.gsfc.nasa.gov/pubs/office\\_notes](http://gmao.gsfc.nasa.gov/pubs/office_notes).

Approved by:

---

Michele M. Rienecker                      Date  
Head, Global Modeling and Assimilation Office  
Code 610.1, NASA GSFC

## REVISION HISTORY

<b>Version Number</b>	<b>Revision Date</b>	<b>Extent of Changes</b>
1.0	2/24/2012	Baseline
1.1	4/12/2012	Corrected ECS short names for “const_2d_mld_Nx” and “tag1_2d_mld_Nx” files
1.2	9/10/2012	Converted to GMAO Office Note No. 3 (version 1.2). Updated “References” and “Web Resources”. Replotted Figure C.6-5 with legend consistent with that of Figure C.6-6.

## TABLE OF CONTENTS

<b>1. INTRODUCTION .....</b>	<b>1</b>
<b>2. FORMAT AND FILE ORGANIZATION.....</b>	<b>3</b>
2.1 DIMENSIONS .....	3
2.2 VARIABLES .....	4
2.3 GLOBAL ATTRIBUTES .....	6
<b>3. TIME-AVERAGED PRODUCTS.....</b>	<b>7</b>
<b>4. GRID STRUCTURE .....</b>	<b>8</b>
<b>5. FILE NAMING CONVENTIONS .....</b>	<b>9</b>
5.1 FILE NAMES.....	9
5.2 EARTH SCIENCE DATA TYPES (ESDT) NAME.....	11
<b>6. SUMMARY OF THE MERRA-LAND FILE COLLECTION .....</b>	<b>12</b>
<b>7. METADATA .....</b>	<b>18</b>
7.1 EOSDIS METADATA .....	18
7.2 CF METADATA .....	18
<b>8. SAMPLE SOFTWARE.....</b>	<b>19</b>
8.1 HDF-EOS EXAMPLE .....	19
8.2 HDF (NON EOS) EXAMPLE.....	21
<b>APPENDIX A: SURFACE REPRESENTATION.....</b>	<b>24</b>
<b>APPENDIX B: LAND SURFACE BUDGETS.....</b>	<b>25</b>
<b>APPENDIX C: ADDITIONAL NOTES FOR MERRA-LAND .....</b>	<b>27</b>
C.1 SPIN-UP AND EXPERIMENT PERIOD .....	27
C.2 SURFACE METEOROLOGICAL FORCING .....	27
C.3 PRECIPITATION CORRECTIONS .....	28
C.4 TOTAL LAND WATER STORAGE .....	29
C.5 SNOW COVER FRACTION.....	30
C.6 UPDATE OF REICHLER ET AL. 2011 SCIENCE RESULTS .....	31
<b>REFERENCES .....</b>	<b>37</b>
<b>WEB RESOURCES .....</b>	<b>38</b>

# 1. Introduction

This document describes the **MERRA-Land** reanalysis, a supplemental land surface data product of the Modern-Era Retrospective analysis for Research and Applications (MERRA).

The standard MERRA data product is a NASA atmospheric reanalysis for the satellite era using Version 5 of the Goddard Earth Observing System model (GEOS-5) and its associated data assimilation system (DAS), Version 5.2.0 (Rienecker et al., 2008). MERRA focuses on historical analyses of the hydrological cycle on a broad range of weather and climate time scales and places the NASA EOS suite of observations in a climate context. MERRA covers the period 1979-present, continuing as an ongoing climate analysis as resources allow. Details regarding MERRA can be found in the MERRA File Specification Document (Lucchesi, 2012), peer-reviewed literature (e.g., Rienecker et al., 2011 and other papers in the Journal of Climate MERRA Special Collection, <http://journals.ametsoc.org/page/MERRA>), and at <http://gmao.gsfc.nasa.gov/research/merra/>. MERRA data can be downloaded from the M-DISC (<http://disc.sci.gsfc.nasa.gov/mdisc/>).

**MERRA-Land** is a land-only (“off-line”) replay of the MERRA land model component with two key changes from MERRA:

1. MERRA-Land precipitation forcing is based on merging a gauge-based data product from the NOAA Climate Prediction Center with MERRA precipitation (Appendix C.3).
2. The Catchment land surface model used in MERRA-Land was updated from the “MERRA” version (used in GEOS-5.2.0) to the “Fortuna-2.5” version (used operationally in GEOS-5.7.2 since Aug 2011).

These two changes lead to various science improvements that are documented, for a *preliminary* version of the MERRA-Land data product, in Reichle et al. (2011) and, for the official MERRA-Land data product, in Appendix C.6.

**Please acknowledge any use of MERRA-Land data by citing (Reichle et al., 2011).**

The Concurrent Versions System (CVS) repository tag of the GEOS-5 land data assimilation system (LDAS) used to generate the MERRA-Land data product is “GEOSldas-MERRALand” (CVS module: “LDASsa\_m2”).

The MERRA-Land data are available online through the Goddard Earth Sciences (GES) Data and Information Services Center (DISC), alongside MERRA data. The disk cache and interface tools in this Modeling DISC (MDISC) system leverage new DISC capabilities in serving active data archives. More information can be found at <http://mdisc.sci.gsfc.nasa.gov>. Detailed documentation of the data access procedures can be found at <http://gmao.gsfc.nasa.gov/research/merra>.

The MERRA-Land data are available as hourly averages of land surface fields at a horizontal

resolution of  $2/3^\circ$  longitude by  $1/2^\circ$  latitude (section 4) **from 1 January 1980** onwards (Appendix C.1). Daily and monthly averages are also available through MDISC sub-setting tools.

A complete list of the MERRA-Land output fields can be found in Section 6. Sections 2-5 and 7-8 provide information on the general file format and organization, including file names, metadata, and sample software. Appendix A discusses the surface representation in MERRA, Appendix B provides information on the MERRA land surface budgets, and Appendix C includes additional notes for MERRA-Land, including spin-up, surface meteorological forcing data and precipitation corrections, differences between the MERRA-Land and MERRA data products, and an update of the (Reichle et al., 2011) science results.

## 2. Format and File Organization

GEOS-5 files are in HDF-EOS format, which is an extension of the Hierarchical Data Format Version 4 (HDF-4), developed at the National Center for Supercomputing Applications <http://www.hdfgroup.org/>. HDF-EOS provides additional capabilities over HDF-4, but does not prevent the use of the files as standard HDF-4 files (de Witt, 1996; Gross, 1997).

Each GEOS-5 file contains a single HDF-EOS grid, which in turn contains a number of geophysical quantities that we refer to as “fields” or “variables.” For MERRA-Land only two-dimensional fields are provided. Soil temperature profiles, for example, are stored as different fields for each layer.

The variables are created using the **GDdefield** function from the HDF-EOS GD (Grid) API, which stores them as HDF Scientific Data Set (SDS) arrays, so that they can be read with standard HDF routines. In addition to the geophysical variables, the files have SDS arrays that define dimension scales (or coordinate variables). There are two distinct scales for each dimension, which ensures that a wide variety of graphical display tools can interpret the data. In particular, there are dimension scales that adhere to the [CF](#) conventions, as well as ones that adhere to the [COARDS](#) conventions.

Due to the size of the MERRA archive, all data are compressed with a GRIB-like method that is invisible to the user. This method does degrade the precision of the data, but every effort has been made to ensure that differences between the product and the original, uncompressed data are not scientifically meaningful. Once the precision has been degraded, the files are written using the standard (internal) gzip deflation available in HDF-4.

EOS Core System (ECS) metadata and other information are stored as global attributes.

### 2.1 Dimensions

GMAO HDF-EOS files contain two sets of dimension scale (coordinate) information. One set of dimensions is defined using the **SDsetdimscale** function of the standard HDF SD interface. This set of scales has an attribute named “units,” set to an appropriate string defined by the CF and COARDS conventions that can be used by applications to identify the dimension. The other set of dimension scales is created using the **GDdefield/GDwritefield** functions.



Table 2.1-1. Dimension Variables Contained in GMAO HDF-EOS Files

Name	Description	Type	<i>units</i> attribute
XDim:EOSGRID	Longitude	float32	degrees_east
YDim:EOSGRID	Latitude	float32	degrees_north
Height:EOSGRID (3D only)	pressure or layer index	float32	hPa or layer
TIME:EOSGRID	minutes since first time in file	float32	minutes
XDim	Longitude, in degrees east	float64	N/A
YDim	Latitude, in degrees north	float64	N/A
Height (3D only)	pressure or level indices	float64	N/A
Time	seconds since 00:00Z on 1 January 1993	float64	N/A

The 32-bit dimension variables have a "units" attribute that makes them COARDS-compliant, while the 64-bit dimension variables satisfy ECS requirements.

## 2.2 Variables

Variables are stored as SDS arrays, even though they are defined with the HDF-EOS **GDdefield** function. As a result, one can use the SD interface of the HDF library to read any variable from the file. The only thing one must know is the SDS variable name listed in Table 6-1 and the number of dimensions (the rank). One can quickly list the variables in the file by using common utilities such as *ncdump* or *hdp*. Both utilities are distributed by the HDF Group with the HDF-4 library. In [Section 8](#) we present sample code for reading one or more data fields from this file. The *short names* for all variables in all GMAO data products are listed in Section 6. A glossary with a brief description of each variable is available in the separate GEOS-5 Variable Definition Glossary, available on the GMAO web page (see Web Resources).

Each variable has several useful metadata attributes. Many of these attributes are required by the [CF](#) and [COARDS](#) conventions, while others are specific for GMAO products. The following table lists required attributes. Other attributes may be included for internal GMAO use and can be ignored.

Table 2.2-1 Metadata attributes associated with each SDS.

Name	Type	Description
_FillValue	float32	Floating-point value used to identify missing data. Normally be set to 1e15. Required by CF.
missing_value	float32	Same as _FillValue. Required for COARDS backwards compatibility.
valid_range	float32, array(2)	This attribute defines the valid range of the variable. The first element is the smallest valid value and the second element is the largest valid value. Required by CF. In MERRA files these are set to -/+ _FillValue.
long_name	String	A brief description of the variable contents taken from the <i>Description</i> column of the tables in section 6.
standard_name	String	An ad hoc description of the variable as required by <a href="#">COARDS</a> . It approximates the standard names as defined in an early version of CF conventions. (See References).
units	String	The units of the variable. Must be a string that can be recognized by UNIDATA's Uunits package.
scale_factor	float32	If variable is packed as 16-bit integers, this is the <code>scale_factor</code> for expanding to floating-point. Currently data are not packed, thus value is 1.0
add_offset	float32	If variable is packed as 16-bit integers, this is the offset for expanding to floating-point. Currently, data are not packed, thus value is 0.0.

### 2.3 Global Attributes

In addition to SDS arrays containing variables and dimension scales, global metadata is also stored in GMAO HDF-EOS files. Some metadata are required by the CF/COARDS conventions, some are present to meet ECS requirements, and others as a convenience to users of GMAO products. A summary of global attributes present in all MERRA and MERRA-Land files is shown in Table 2.3-1.

Table 2.3-1 Global metadata attributes associated with each SDS.

	<b>Type</b>	<b>Description</b>
Conventions	character	Identification of the file convention used, currently “CF-1.0”
title	character	Experiment identification: “MERRA-Land”
history	character	CVS tag of this release. CVS tags are used internally by the GMAO to designate versions of the system.
institution	character	“NASA Global Modeling and Assimilation Office”
source	character	CFIO Version (CFIO is the GMAO’s IO layer)
references	character	GMAO website address
comment	character	As required
HDFEOSVersion	character	Version of the HDF-EOS library used to create this file.
StructMetadata.0	character	This is the GridStructure metadata that is created by the HDF-EOS library.
CoreMetadata.0	character	The ECS inventory metadata.
ArchivedMetadata.0	character	The ECS archive metadata

### 3. Time-averaged Products

The MERRA-Land file collection only contains time-averaged products.

Time-averaged collections may contain either hourly or monthly means, but not mixtures of these. (Files with daily mean values are also available through the MDISC subsetting tools.) Each time-averaged collection consists of a continuous sequence of data averaged over the indicated interval and time stamped with the central time of the interval. For hourly data, for example, these times are 00:30 GMT, 01:30 GMT, 02:30 GMT, etc. Monthly files represent averages for the calendar months, accounting for leap years. For monthly means, each file contains a single month.

Each hourly collection consists of a set of daily files, with the date as part of the [file name](#), and each daily file contains 24 hourly averages per field. For collections of monthly or seasonal means each month or season is in a separate file, and file names also include a date in the file name.

## 4. Grid Structure

MERRA-Land fields are produced in the tile space that is associated with the native horizontal grid of the Finite-Volume dynamical core, with a resolution of  $2/3^\circ$  longitude by  $1/2^\circ$  latitude. See Appendix A for additional information on the GEOS-5 tile space. The MERRA-Land data product is provided as gridded fields.

The GEOS-5 MERRA *native resolution* gridded output is on a global horizontal grid, consisting of **IMn=540** points in the longitudinal direction and **JMn=361** points in the latitudinal direction. The horizontal native grid origin, associated with variables indexed ( $i=1, j=1$ ) represents a grid point located at ( $180^\circ\text{W}, 90^\circ\text{S}$ ). Latitude ( $\varphi$ ) and longitude ( $\lambda$ ) of grid points as a function of their indices ( $i, j$ ) can be determined by:

$$\begin{aligned}\lambda_i &= -180 + (\Delta\lambda)_n(i-1), \quad i = 1, \text{IMn} \\ \varphi_j &= -90 + (\Delta\varphi)_n(j-1), \quad j = 1, \text{JMn}\end{aligned}$$

Where  $(\Delta\lambda)_n = 2/3^\circ$  and  $(\Delta\varphi)_n = 1/2^\circ$ . For example, ( $i = 271, j = 181$ ) corresponds to a grid point at ( $\lambda = 0, \varphi = 0$ ).

For land surface applications it is important to note that the “grid points” mentioned above should be considered as the center points of land surface “grid cells”. For example, the “grid cell” associated with the grid point at ( $\lambda = 0, \varphi = 0$ ) extends from  $0.3333^\circ\text{W}$  to  $0.3333^\circ\text{E}$  in longitude and from  $0.25^\circ\text{S}$  to  $0.25^\circ\text{N}$ . The MERRA grid is therefore different from the grid used in the Global Soil Wetness Project 2 (GSWP-2) in which the Greenwich Meridian ( $\lambda = 0$ ) is a grid cell boundary. All land surface tiles within the same grid cell (Appendix A) experience the same surface meteorological forcing.

## 5. File Naming Conventions

Each GEOS-5 product file has a complete file name identified in the EOSDIS metadata as "LocalGranuleID". EOSDIS also requires eight-character abbreviated naming indices for each Earth Science Data Type (ESDT). In MERRA and MERRA-Land, each file collection has a unique ESDT index. The ESDT index convention is described in section 5.2.

### 5.1 File Names

Similar to the convention for MERRA files, the complete file names for the GEOS-5 MERRA-Land product consist of six dot-delimited nodes:

***runid.runtype.config.collection.timestamp.extension***

The node fields, which vary from file to file, are defined as follows:

#### ***runid***

All MERRA *Mainstream* runs are identified by **MERRASV<sub>v</sub>**, where the numeric qualifiers **S** and **V<sub>v</sub>** denote the production *Stream* and the *Version* numbers. MERRA was run in three production *Streams*, each covering approximately a third of the MERRA period. For MERRA-Land files, the *runid* identifies the MERRA *Stream* and *Version* that was used to construct the surface meteorological forcing data (Appendix C).

#### ***runtype:***

All MERRA-Land file names have a ***runtype*** designation **prod**.

**prod** = Standard product

#### ***config:***

MERRA-Land is a land-only “replay” (model integration) of the MERRA land surface model component with the modifications identified in section 1. The configuration is therefore described as a “simulation”.

**simul** = Simulation. Surface meteorological forcing data are prescribed as discussed in section 1 and Appendix C.

#### ***collection:***

All MERRA and MERRA-Land data are organized into file *collections* that contain fields with common characteristics. These collections are used to make the

data more accessible for specific purposes. Collection names are of the form *freq\_dims\_group\_HV*, where the four attributes are:

*freq*: time-independent (**cnst**) or time-average (**tavgF**), where *F* indicates the frequency or averaging interval and can be any of the following:

- 1** = Hourly
- M** = Monthly mean
- U** = Monthly-diurnal mean

*dims*: **2d** for MERRA-Land collections with only 2-dimensional fields

*group*: A three-letter mnemonic for the type of fields in the collection.

**mld** = MERRA-Land

*HV*: Horizontal and vertical grid.

- H* = **N** Native (2/3° x 1/2°) horizontal resolution
- V* = **x** Horizontal-only data (surface, single level, etc.) ; *dims* must be **2d**

***timestamp***:

The *timestamp* node defines the date and time associated with the data in the file. It has the form *yyyymmdd* for daily-aggregated files of hourly averages, *yyyymm* for monthly-mean files.

- yyyy** - year string (e.g. , "2002")
- mm** - month string (e.g., "09" for September)
- dd** - day of the month string (if needed)

***extension***:

The file name *extension* is **hdf**.

***EXAMPLE***:

**MERRA300.prod.simul.tavg1\_2d\_mld\_Nx.20020915.hdf**

This is an example for the name of a MERRA-Land file from the production (“prod”) segment of the MERRA-Land simulation (“simul”) using surface meteorological forcing data constructed from the original version of the third (most recent) MERRA stream (“MERRA300”). The data are hourly time averages (“tavg1”), two-dimensional (“2d”), MERRA-Land (“mld”) fields, at native horizontal resolution (“Nx”). The file contains 24 hourly averages per field for 15 September 2002 and is in “hdf” format.

## 5.2 Earth Science Data Types (ESDT) Name

To accommodate EOSDIS toolkit requirements, all MERRA and MERRA-Land files are associated with a nine-character ESDT. The ESDT is a short (and rather cryptic) handle for users to access sets of files. In MERRA and MERRA-Land the ESDT is used to identify the *Mainstream collections* and consists of a compressed version of the collection name of the form:

**MCTFHVGGG**

where

**C:** Configuration

**S** = Simulation

**T:** Time Description:

**T** = Time-averaged

**C** = Time-independent

**F:** Frequency

**1** = Hourly

**M** = Monthly mean

**U** = Monthly-Diurnal mean

**H:** Horizontal Resolution

**N** = [Native](#) (2/3° longitude by 1/2° latitude).

**V:** Vertical Location

**X** = Two-dimensional

**GGG:** Group

**MLD** = MERRA-Land land surface variables



## 6. Summary of the MERRA-Land File Collection

The GEOS-5 MERRA-Land product consists primarily of one “file collection”, with group label “mld”, for time-varying land surface fields (hourly averages) at the “native” grid resolution (Table 6-1). In addition, a single file with land surface constants is provided at the “native” grid resolution (Table 6-2). The file names follow the conventions described in [Section 5](#). The approximate data product volume is 0.11 Gbytes/day and 1.3 Tbytes for the entire MERRA-Land collection of hourly files. All numbers refer to the (internally) compressed file sizes.

The MERRA-Land “mld” collection of time-varying land surface fields is designed to include all fields that are present in the standard MERRA “lnd” collection and a few additional fields of interest for land surface research and applications, notably: (i) the 6-layer soil temperature profile; (ii) the gridded land surface temperature, and (iii) the gridded soil moisture in volumetric units ( $\text{m}^3 \text{m}^{-3}$ ) as well as in dimensionless units of degree of saturation (or wetness), including the total profile soil moisture.

All fields are computed on the GEOS-5 “Land” tiles (Appendix A) and then averaged (area-weighted) to grid space, except for “vegetation greenness fraction” (GRN), which is weighted by tile area and by leaf area index (LAI) in the averaging step from tile to grid space. Surface temperature at the tile level is itself computed as a spatial average across sub-tile hydrological regimes as outlined below.

The surface and soil temperature fields require some additional explanation. TSURF is the spatial average across the grid element of the temperature in direct contact with the atmosphere, i.e., the temperature from which the outgoing longwave radiation, sensible heat flux, and latent heat flux are computed. Specifically, TSURF is computed for each tile as the weighted average of the sub-tile temperatures:

$$\text{TSURF} = \text{TPSNOW} * \text{FRSNO} + \text{TSAT} * \text{FRSAT} + \text{TUNST} * \text{FRUNST} + \text{TWLT} * \text{FRWLT}.$$

and subsequently averaged to grid space (excluding lake and land ice temperatures). TPSNOW is the temperature of the topmost snow layer and FRSNO is the snow cover fraction. The non-snow surface temperature components (TSAT, TUNST, and TWLT) are prognostic model variables with surface heat capacities of  $70,000 \text{ J m}^{-2}$  for (tropical) broadleaf evergreen forest tiles and  $200 \text{ J m}^{-2}$  for all other tiles; different heat capacities were assigned for reasons associated with MERRA skin temperature generated during production. For tropical forest land tiles, the TSAT, TUNST, and TWLT fields are therefore tied to approximately the top 5 cm of soil, whereas for all other tiles the effective soil depth associated with TSAT, TUNST, and TWLT is negligible.

Because the sub-tile area fractions are different for each tile and vary dynamically and independently of the sub-tile temperatures, the above equation for TSURF is only an approximation for gridded and/or time-averaged data. In particular, it does *not* apply for monthly-averaged data. Therefore, the correctly determined TSURF is available as a separate field in the “mld” file collection. The grid-cell-averaged depth associated with the TSAT,

TUNST, and TWLT fields is provided as the temporally constant field DZTS (Table 6-2). Note that the sub-tile temperatures are set to no-data values when the corresponding sub-tile area fractions vanish.

The fields TSOIL1 through TSOIL6 provide the spatially-averaged temperature profile within the soil, starting at the topmost soil temperature layer (TSOIL1). The corresponding soil temperature layer thicknesses are the same for all tiles and are provided in the fields DZGT1 through DZGT6 (Table 6-2). Note that the vertical profile of temperature represented by TSOIL1 through TSOIL6 should be “shifted downward” by DZTS.

The soil moisture variables are provided in units of degree of saturation (fraction) and volumetric soil moisture ( $\text{m}^3 \text{m}^{-3}$ ) for three different averaging depths. The surface soil moisture (SFMC and GWETTOP) is the average soil moisture for the top DZSF=0.02 m of the soil. This surface soil moisture layer depth is identical for all land tiles and grid cells.

The root zone soil moisture variable, RZMC, ostensibly refers to the average amount of water in a nominal “root zone” of 1 meter depth. In some areas of the globe, however, the bedrock depth used was shallow enough to require (for computational purposes) the definition of a shallower root zone depth, as small as 0.75 meters. The layer depth associated with root zone moisture content is provided as the temporally constant field DZRZ (Table 6-2). Note that the layer associated with the root zone soil moisture is a “nested” thickness and includes the top 0.02 m surface soil moisture layer; this reflects the fact that the Catchment model does not have a traditional ‘layer-based’ structure (Koster et al., 2000; Ducharne et al., 2000). Because the RZMC variable is intrinsic rather than extrinsic ( $\text{m}^3 \text{m}^{-3}$  rather than  $\text{kg m}^{-2}$ ), for most applications the exact layer depth associated RZMC and GWETROOT will not be an issue.

The soil moisture in the entire soil profile down to the bedrock is given as the “profile” soil moisture (PRMC and GWETPROF). The associated “depth-to-bedrock” is provided as the grid-cell average constant DZPR. Again, for reasons related to the Catchment model’s structure, this thickness is a “nested” thickness that includes the soil moisture associated with the root zone reservoir.

The land surface constants file (Table 6-2) also provides the wilting point in units of degree of saturation (WPWET, fraction), volumetric soil moisture (WPMC,  $\text{m}^3 \text{m}^{-3}$ ) and equivalent mass of water (WPEMW,  $\text{kg m}^{-2}$ ) in the full soil moisture profile, assuming the full profile (not just the nominal root zone) is at the wilting point. In the Catchment model, wilting points are constant with depth. Only the soil moisture associated with the surface reservoir (SFMC, GWETTOP) can dry out below the wilting point through bare soil evaporation. The plants’ ability to extract soil water ceases when the soil moisture in the root zone reservoir and/or the total profile reaches the wilting point.

The “plant available” soil moisture in each reservoir can be computed in units of degree of saturation or in volumetric soil moisture units by subtracting the wilting point (in the appropriate units) from the soil moisture variables. Furthermore, to obtain the total “plant available” water in the soil column, first compute the total “profile” soil moisture PREMW in units of equivalent water mass ( $\text{kg m}^{-2}$ ); this can be approximated by subtracting the total snow mass from the total

land water storage, that is,

$$\text{PREMW} = \text{TWLAND} - \text{SNOMAS}.$$

This approximation neglects the water stored in the canopy interception reservoir, which is typically well below  $1 \text{ kg m}^{-2}$ . The total plant available water in units of equivalent depth of water ( $\text{kg m}^{-2}$ ) can then be computed by subtracting the wilting point WPEMW (in units of equivalent mass of water) from PREMW. Note that assuming a density of  $1,000 \text{ kg m}^{-3}$  for liquid water, units of equivalent mass of water ( $\text{kg m}^{-2}$ ) are identical to units equivalent *depth* of water in millimeters.

While soil moisture is available in units of degree of saturation (or wetness), field capacity is not used in the GEOS-5 Catchment land surface model and is therefore not available with the MERRA-Land data.

Variable names in Tables 6-1 and 6-2 refer to HDF names. Note that the land surface constants provided with MERRA-Land (Table 6-2) apply equally to MERRA. Additional details on the variable definitions may be found in the GEOS-5 Variable Definition Glossary, available on the GMAO web page (see Web Resources).

Table 6-1 MERRA-Land “mld” file collection

ECS short name: **MST1NXMLD**  
 ECS long name: MERRA-Land 2d land surface diagnostics  
 Name: tavg1\_2d\_mld\_Nx  
 Characteristics: **Time averaged, single level, at native resolution**  
 Dimensions: longitude: **540**, latitude: **361**  
 Times: **0:30, 1:30, 2:30, 3:30, 4:30, ... GMT**

Variable Name	Description	Units
GRN	Vegetation greenness fraction (LAI-weighted)	Fraction
LAI	Leaf area index	m <sup>2</sup> m <sup>-2</sup>
GWETPROF*	Total profile soil wetness	Fraction
GWETROOT	Root zone soil wetness	Fraction
GWETTOP	Top soil layer wetness	Fraction
PRMC*	Total profile soil moisture content	m <sup>3</sup> m <sup>-3</sup>
RZMC*	Root zone soil moisture content	m <sup>3</sup> m <sup>-3</sup>
SFMC*	Top soil layer soil moisture content	m <sup>3</sup> m <sup>-3</sup>
TSURF*	Mean land surface temperature (incl. snow)	K
TPSNOW	Top snow layer temperature	K
TUNST	Surface temperature of unsaturated (but non-wilting) zone	K
TSAT	Surface temperature of saturated zone	K
TWLT	Surface temperature of wilting zone	K
TSOIL1*	Soil temperature in layer 1	K
TSOIL2*	Soil temperature in layer 2	K
TSOIL3*	Soil temperature in layer 3	K
TSOIL4*	Soil temperature in layer 4	K
TSOIL5*	Soil temperature in layer 5	K
TSOIL6*	Soil temperature in layer 6	K
PRECSNO	Surface snowfall	kg m <sup>-2</sup> s <sup>-1</sup>
PRECTOT	Total surface precipitation	kg m <sup>-2</sup> s <sup>-1</sup>
SNOMAS	Snow mass	kg m <sup>-2</sup>
SNODP	Snow depth	m
EVPSOIL	Bare soil evaporation	W m <sup>-2</sup>
EVPTRNS	Transpiration	W m <sup>-2</sup>
EVPINTR	Interception loss	W m <sup>-2</sup>
EVPSBLN	Sublimation	W m <sup>-2</sup>

<b>Variable Name</b>	<b>Description</b>	<b>Units</b>
RUNOFF	Overland runoff	$\text{kg m}^{-2} \text{s}^{-1}$
BASEFLOW	Baseflow	$\text{kg m}^{-2} \text{s}^{-1}$
SMLAND	Snowmelt over land	$\text{kg m}^{-2} \text{s}^{-1}$
QINFIL*	Soil water infiltration rate	$\text{kg m}^{-2} \text{s}^{-1}$
FRUNST	Fractional unsaturated (but non-wilting) area	Fraction
FRSAT	Fractional saturated area	Fraction
FRSNO	Fractional snow-covered area	Fraction
FRWLT	Fractional wilting area	Fraction
PARDF	Surface downward PAR** diffuse flux	$\text{W m}^{-2}$
PARDR	Surface downward PAR** beam flux	$\text{W m}^{-2}$
SHLAND	Sensible heat flux from land	$\text{W m}^{-2}$
LHLAND	Latent heat flux from land	$\text{W m}^{-2}$
EVLAND	Evaporation from land	$\text{kg m}^{-2} \text{s}^{-1}$
LWLAND	Net downward longwave flux over land	$\text{W m}^{-2}$
SWLAND	Net downward shortwave flux over land	$\text{W m}^{-2}$
GHLAND	Downward heat flux into top soil layer	$\text{W m}^{-2}$
TWLAND	Total water stored in land reservoirs	$\text{kg m}^{-2}$
TELAND	Energy stored in all land reservoirs	$\text{J m}^{-2}$
WCHANGE	Total land water change per unit time	$\text{kg m}^{-2} \text{s}^{-1}$
ECHANGE	Total land energy change per unit time	$\text{W m}^{-2}$
SPLAND	Spurious land energy source	$\text{W m}^{-2}$
SPWATR	Spurious land water source	$\text{kg m}^{-2} \text{s}^{-1}$
SPSNOW	Spurious snow energy source	$\text{W m}^{-2}$

\*Denotes variables that are new in MERRA-Land and not available in the MERRA "lnd" collection.

\*\*PAR = Photosynthetically Active Radiation

Table 6-2 MERRA-Land “land constants” file collection

ECS short name: **MSC0NXMLD**  
 ECS long name: MERRA-Land 2d land surface constants  
 Name: const\_2d\_mld\_Nx  
 Characteristics: **Constant at native resolution**  
 Dimensions: longitude: **540**, latitude: **361**

<b>Variable Name</b>	<b>Description</b>	<b>Units</b>
DZSF	Thickness of soil layer associated with SFMC and GWETTOP	m
DZRZ	Thickness of soil layer associated with RZMC and GWETROOT	m
DZPR	Thickness of soil layer associated with PRMC and GWETPROF (“depth-to-bedrock” in the Catchment model)	m
DZTS	Thickness of soil layer associated with TSAT, TUNST, and TWLT	m
DZGT1	Thickness of soil layer associated with TSOIL1	m
DZGT2	Thickness of soil layer associated with TSOIL2	m
DZGT3	Thickness of soil layer associated with TSOIL3	m
DZGT4	Thickness of soil layer associated with TSOIL4	m
DZGT5	Thickness of soil layer associated with TSOIL5	m
DZGT6	Thickness of soil layer associated with TSOIL6	m
WPWET	Soil wilting point (in degree of saturation units)	Fraction
WPMC	Soil wilting point (in volumetric units)	m <sup>3</sup> m <sup>-3</sup>
WPEMW	Soil wilting point (in units of equivalent mass of total profile water)	kg m <sup>-2</sup>

*\*Variable names used in column “Description” are defined in Table 6-1.*

## 7. Metadata

GEOS-5 gridded output files include two types of metadata. When using the HDF-EOS library and tools, the EOSDIS metadata are used (Doplick, 1997). Other utilities such as [GrADS](#) or [FERRET](#) use the CF metadata (NOAA, 1995).

### 7.1 EOSDIS Metadata

The EOSDIS toolkit only uses the EOSDIS metadata. EOSDIS identifies two major types of metadata, collection and granule.

Granule metadata is the "table of contents" information stored on the data file itself. The EOSDIS granule metadata include:

- File name (local granule ID)
- Grid structure
- Number of time stamps stored in the file
- Number of vertical levels for each variable in this file
- Names of variables in this file
- Variable format (32-bit floating point, 16-bit integer, etc.)
- Variable storage dimensions
  - time, latitude and longitude for 2-d fields
- "Missing" value for each variable.

### 7.2 CF Metadata

When [GrADS](#) or [FERRET](#) are used to view GEOS-5 gridded data sets, the application uses the CF metadata embedded in the data products. These metadata include the following information:

- Space-time grid information (dimension variables)
- Variable names and descriptions
- Variable units
- "Missing" value for each variable.

The grid information and units comply with the CF conventions. The variable names and descriptions are only loosely based on an early version of CF conventions.

## 8. Sample Software

The following example illustrates the use of the standard HDF library or the ECS HDF-EOS library to read GEOS-5 products. The program shown below accepts as command line arguments a file name and a field name. It opens the file, reads the requested field at the first time, computes an average for this field, and prints the result to standard output. There are two versions of this program. The first version uses the HDF-EOS library to read the file. The second version uses the standard HDF library to read the file. Electronic copies of these programs can be obtained from the Operations section of the GMAO [web page](#).

### 8.1 HDF-EOS Example

```

/*****
/* This program demonstrates how to read a field from a GMAO HDF-EOS
/* product using the HDF-EOS library. It will take a file name and
/* field name on the command line, read the first time of the given
/* field, calculate an average of that time and print the average.
/*
/* usage: avg <file name> <field name>
/*
/* Rob Lucchesi
/* rob.lucchesi@nasa.gov
*****/

#include "hdf.h"
#include "mfhdf.h"
#include <stdio.h>

main(int argc, char *argv[]) {

int32 sd_id, sds_id, status;
int32 sds_index;
int32 start[4], edges[4], stride[4];
char *fname, *vname;
float32 *data_array;
float32 avg, sum;
int32 i;
int32 file_id, gd_id;
int32 xdim, ydim, zdim, len;

if (argc != 3) {
    printf("Usage: avg <filename> <field> \n");
    exit (-1);
}

```



```

fname = argv[1];
vname = argv[2];

/* Open the file (read-only) */

if ( (file_id = GDopen (fname, DFACC_RDONLY)) < 0) {
    printf ("Could not open %s\n", fname);
    exit(-1);
}

/* Attach to the EOS grid contained within the file. */
/* The GMAO uses the generic name "EOSGRID" for the grid in all products. */

if ( (gd_id = GDattach (file_id, "EOSGRID")) < 0) {
    printf ("Could not open %s\n", fname);
    exit(-1);
}

status = GDget(file_id, xdim, ydim, zdim),

/* Set positioning arrays to read the entire field at the first time. */

start[0] = 0;
start[1] = 0;
start[2] = 0;
start[3] = 0;

stride[0] = 1;
stride[1] = 1;
stride[2] = 1;
stride[3] = 1;

edges[0] = 1;
edges[1] = zdim;
edges[2] = ydim;
edges[3] = xdim;

len = xdim*ydim*zdim;
data_array = (float32 *)malloc(len);

/* Read the data into data_array */

status = GDreadfield (gd_id, vname, start, stride, edges, data_array);
printf ("Read status=%d\n", status);

```

```

/* Calculate and print the average */

sum=0.0;
for (i=0; i<len; i++) sum += data_array[i];
avg = sum/(float32)len;

printf ("Average of %s in 3 dimensions is=%f\n",vname,avg);

/* Close file. */

status = GDdetach (gd_id);
status = GDclose (file_id);

}

```

## **8.2 HDF (non EOS) Example**

```

/*****
/* This program demonstrates how to read a field from a GMAO HDF-EOS          */
/* product using the HDF library (HDF-EOS not required). It will take        */
/* a file name and field name on the command line, read the first time      */
/* of the given field, calculate an average of that time and print the average.*/
/*                                                                            */
/* usage: avg <file name> <field name>                                     */
/*                                                                            */
/* Rob Lucchesi                                                            */
/* rob.lucchesi@nasa.gov                                           */
/*                                                                            */
*****/

#include "hdf.h"
#include "mfhdf.h"
#include <stdio.h>

main(int argc,char *argv[]) {
int32 sd_id, sds_id, status;
int32 sds_index;
int32 start[4], edges[4], stride[4];
char *fname, *vname;
float32 *data_array;
float32 avg, sum;

```

```

int32 i, xdim, ydim , zdim, len;

if (argc != 3) {
    printf("Usage: avg <filename> <field> \n");
    exit (-1);
}

fname = argv[1];
vname = argv[2];

/* Open the file (read-only) */

if ((sd_id = SDstart (fname, DFACC_RDONLY)) < 0) {
    printf ("Could not open %s\n",fname);
    exit(-1);
}

/* Find the index and ID of the SDS for the given variable name and get its dimensions. */

if ((sds_index = SDnametoindex (sd_id, vname)) < 0) {
    printf ("Could not find %s\n",vname);
    exit(-1);
}

sds_id = SDselect (sd_id,sds_index);
status = GDget(file_id,xdim,ydim,zdim),

/* Set positioning arrays to read the entire field at the first time. */

start[0] = 0;
start[1] = 0;
start[2] = 0;
start[3] = 0;

stride[0] = 1;
stride[1] = 1;
stride[2] = 1;
stride[3] = 1;

edges[0] = 1;
edges[1] = zdim;
edges[2] = ydim;
edges[3] = xdim;

len = xdim*ydim*zdim;
data_array = (float32 *)malloc(len);

```

```
/* Read the data into data_array */

status = SDreaddata (sds_id, start, stride, edges, (VOIDP) data_array);
printf ("read status=%d\n",status);

/* Calculate and print the average */

sum=0.0;
for (i=0; i<len; i++) sum += data_array[i];
avg = sum/(float32)len;
printf ("Average of %s in 3 dimensions is=%f\n",vname,avg);

/* Close file. */

status = SDendaccess (sds_id);
status = SDend (sd_id);
}
```

## Appendix A: Surface Representation

In GEOS-5 the surface below each atmospheric column consists of a set of tiles that represent various surface types. Tiles can be of four different types: Ocean, Land, (land) Ice, or Lake, as illustrated in Figure A-1. In each grid box a single Ice tile represents those areas covered by permanent ice. Similarly a single Lake tile represents continental areas covered permanently by water. Other continental areas (non Lake or Ice) can be further subdivided into tiles that represent parts of the grid box in different hydrological catchments, defined according to the Pfafstetter (1989) system. Each of these is, in turn, divided into sub-tiles (not shown in the figure) that represent the wilting, unsaturated (but non-wilting), saturated, and snow-covered fractions of the tile. These sub-tile fractions vary with time and are predicted by the model based on the hydrological state of the catchment and its fine-scale topographic statistics. Details of the land model, including the partitioning into sub-tiles, can be found in Koster et al. (2000). MERRA-Land replays only the Land tiles (that is, excluding Lake and Ice tiles), consistent with the MERRA “lnd” file collection.

The area fraction of each surface type within each grid cell can be found in the MERRA 2d constants file collection (MAC0NXASM or “const\_2d\_asm\_Nx”).

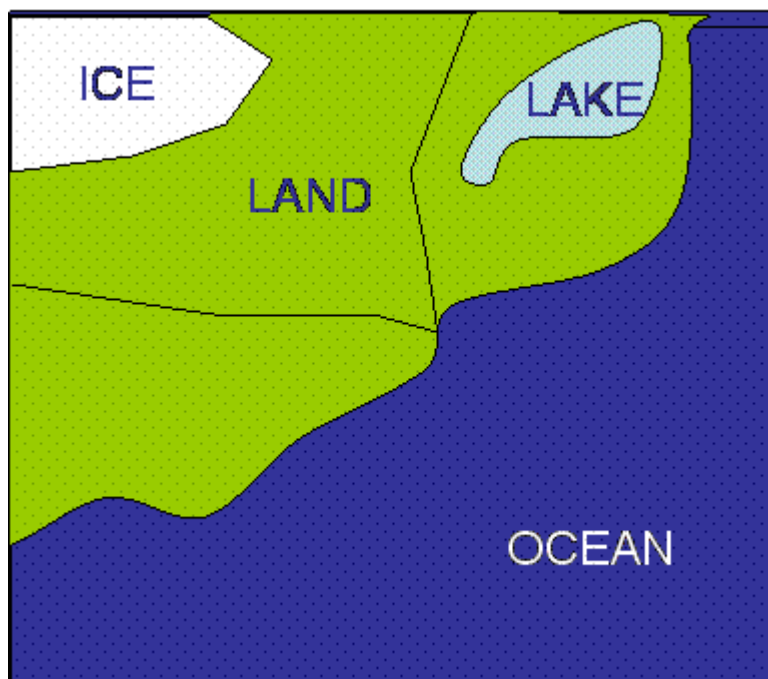


Figure A-1. Land surface representation in GEOS-5.

## Appendix B: Land Surface Budgets

Complete budgets for total land water storage and total land energy storage are accessible in the [tavgl\\_2d\\_mld\\_Nx](#) collection, but note that these budgets are not saved at the full precision (section 2).

**Total Land Water Budget**– The land surface water balance equation can be written

$$\frac{\partial W}{\partial t} = P_l + P_s - E_L - R_L + \mathfrak{R}_W$$

where  $W$  is the total water held in all land surface reservoirs (comprising the soil, the interception reservoir, and the snowpack),  $P_l$  and  $P_s$  are the liquid rain and “snowfall” rates, respectively,  $E_L$  is the total evapotranspiration rate,  $R_L$  is the total runoff–surface (or overland) plus baseflow, and  $\mathfrak{R}_W$  is a spurious water source (nonzero for MERRA due to logistical issues with the land-atmosphere interface but identical to zero for MERRA-Land). The corresponding collection-variable equation, using quantities stored in [tavgl\\_2d\\_mld\\_Nx](#), is

$$\text{WCHANGE} = \text{PRECTOT} - \text{EVLAND} - \text{RUNOFF} - \text{BASEFLOW} + \text{SPWATR}.$$

See Section 6 for variable definitions. Note that all of these quantities are values per unit *land* area, in  $\text{kg m}^{-2} \text{s}^{-1}$  – i.e., not weighted by fractional land area of the grid cell and excluding any contributions from lake, land ice, or ocean tiles. Note also that WCHANGE, PRECTOT, EVLAND, RUNOFF, BASEFLOW and SPWATR are defined only at grid cells for which the land fraction is non-zero and are set to the undefined value elsewhere. (In contrast to EVLAND, the MERRA field EVAP includes evaporation from all surface types, including land, lake, land-ice, and ocean.) Note that the precipitation rate in MERRA-Land differs from that found in MERRA collections due to the correction of the precipitation forcing with a gauge-based precipitation data product (Appendix C.3). Again, for MERRA-Land, the “spurious” water term SPWATR is always zero.

In the [tavgl\\_2d\\_mld\\_Nx](#) collection, the evapotranspiration is separated into components according to its sources as follows:

$$E_L = E_{tr} + E_{bs} + E_{int} + E_{snow},$$

where  $E_{tr}$  is the transpiration rate,  $E_{bs}$  is the evaporation from bare soil surfaces,  $E_{int}$  is the evaporation from the land’s interception reservoir (e.g., water droplets sitting on leaves after a rainstorm), and  $E_{snow}$  is the water that sublimates from the snowpack. The corresponding collection-variable equation is

$$\text{EVLAND} = \frac{1}{L_v} (\text{EVPTRNS} + \text{EVPSOIL} + \text{EVPINTR}) + \frac{1}{L_s} \text{EVPSBLN}.$$

where  $L_v$  and  $L_s$  are the latent heat of vaporization and the latent heat of sublimation, respectively. Following GEOS-5.7.2, MERRA-Land uses  $L_v=2.4665\text{e}6 \text{ J kg}^{-1}$  (at  $15^\circ\text{C}$ ) and  $L_s=2.8002\text{e}6 \text{ J kg}^{-1}$ , whereas MERRA (GEOS-5.2.0) used  $L_v=2.4548\text{e}6 \text{ J kg}^{-1}$  and  $L_s=2.8368\text{e}6 \text{ J kg}^{-1}$ .

**Total Land Energy Budget** –The balance equation for total land surface energy can be written:

$$\frac{\partial \varepsilon}{\partial t} = SW_L + LW_L - SH_L - L_v E_L - L_f \Delta SWE + \mathfrak{R}_L,$$

where  $\varepsilon$  is the total heat content (in the soil, canopy, and snowpack) relative to liquid water.  $SW_L$  is the net shortwave radiation,  $LW_L$  is the net longwave radiation,  $L_v$  is the latent heat of vaporization (from liquid),  $E_L$  is the total evaporation from the land surface,  $SH_L$  is the sensible heat flux from the land surface,  $L_f$  is the latent heat of fusion, and  $\Delta SWE$  is the change in the snow water equivalent (through addition of frozen precipitation falling on the surface or removal of snow through melt or sublimation). The term  $L_f \Delta SWE$  therefore accounts for the snowmelt energy and the added energy needed to evaporate from solid rather than liquid water. The spurious snow energy source  $\mathfrak{R}_L$  is explained below.

In terms of “mld” collection variables the energy budget is:

$$\begin{aligned} \text{ECHANGE} = & \text{SWLAND} + \text{LWLAND} - \text{SHLAND} - L_v \text{EVLAND} \\ & - L_f \text{PRECSNO} - \text{SPLAND} - \text{SPSNOW}. \end{aligned}$$

All of these quantities are computed over land only and are *not* weighted by fractional land area.  $\text{SPLAND}$ , in analogy to  $\text{SPWATR}$ , is associated with the coupled land-atmosphere interface in MERRA and is always zero for MERRA-Land. The term  $\text{SPSNOW}$  contains “spurious” snow-related energy sources and sinks associated with several small accounting inconsistencies across the coupled models. When, for example, the same amount of snow falls at  $-20^\circ\text{C}$  in one region and at  $0^\circ\text{C}$  in a second region, more energy is needed in the first region to melt the snow, because in the first region, energy is needed first to warm the snow up to  $0^\circ\text{C}$ . The atmospheric model does not distinguish between the energy content of snow falling at  $-20^\circ\text{C}$  and that falling at  $0^\circ\text{C}$ , whereas the land model does account for this energy difference. To rectify this inconsistency between the land and atmosphere models, the “negative energy” of the colder snow (i.e., the energy deficit relative to  $0^\circ\text{C}$  snow) is “invented” and added to the snow’s internal energy as soon as it hits the surface. It is therefore implicitly included in  $\text{ECHANGE}$ .

## Appendix C: Additional Notes for MERRA-Land

This section discusses additional details regarding the MERRA-Land data product, including the spin-up procedure, surface meteorological forcing data, and science validation. The discussion in this section also highlights differences between the MERRA-Land and MERRA data products, as well as differences between the official MERRA-Land data product and the preliminary version discussed in Reichle et al. (2011).

### *C.1 Spin-up and Experiment Period*

MERRA-Land does *not* use the 1 January 1979 restart from the MERRA product because the Catchment model version and the precipitation forcing of MERRA-Land differ from that of MERRA, which implies that MERRA-Land and MERRA each have their own, unique soil moisture climatology.

Because the gauge-based precipitation product used for MERRA-Land is not available for 1978, it is difficult to construct an adequate restart file for 1 January 1979. **MERRA-Land data are therefore only available from 1 January 1980 onwards.**

The 1 January 1980 restart file for the MERRA-Land data product was obtained by cycling two times through a 10-year period (1979-1988), followed by a 1-year integration for 1979. Specifically, the MERRA-Land system was first integrated from a “cold” restart for 2 January 1979 to 2 January 1989. The resulting 2 January 1989 restart file was used as the 1 January 1979 restart file for a second 10-year integration, which produced another restart file for 1 January 1989. This restart file was used as the 1 January 1979 restart for a 1-year integration to 1 January 1980, which generated the 1 January 1980 restart file for the final MERRA-Land data product (1980-present).

### *C.2 Surface Meteorological Forcing*

Surface meteorological forcing data are from MERRA as described in Reichle et al. (2011), except that (i) the hourly MERRA precipitation was corrected with the NOAA Climate Prediction Center “Unified” (CPCU) precipitation product and (ii) photosynthetically active radiation from MERRA was also used. See Appendix C.3 for details regarding the precipitation corrections.

MERRA-Land forcing is based on the MERRA surface meteorological data from the MERRA “mainstream”, that is,

1 January 1979 to 31 December 1992:	stream 1 (“Jan79”)
1 January 1993 to 31 December 2000:	stream 2 (“Jan89”)
1 January 2001 to present:	stream 3 (“Jan98”)



Specifically, MERRA-Land was forced with hourly

- air temperature, humidity and wind (TLML, QLML, ULML, VLML) at the lowest atmospheric model level (HLML) from the MERRA “flx” file collection,
- surface pressure (PS) from the MERRA “slv” file collection
- downward and net shortwave radiation at the surface (SWGDN, SWLAND), downward longwave radiation at the surface (LWGAB), photosynthetically active direct and diffuse radiation (PARDR, PARDF) from the MERRA “rad” and “lnd” file collections, and
- corrected precipitation forcing (see Appendix C.3).

Note that MERRA-Land uses the direct and diffuse components of the photosynthetically active radiation (PARDIR and PARDIF) from MERRA directly, whereas the preliminary product of Reichle et al. (2011) was forced with PARDIR and PARDIF given by constant fractions of the MERRA downward shortwave radiation (PARDIR=PARDIF=0.25\*SWGDN).

### ***C.3 Precipitation Corrections***

The global gauge-based NOAA Climate Prediction Center “Unified” (CPCU) precipitation product (Chen et al., 2008) is available from [ftp://ftp.cpc.ncep.noaa.gov/precip/CPC\\_UNI\\_PRCP/GAUGE\\_GLB](ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB). For MERRA-Land, we used the CPCU “V1.0” product for the period 1979-2005 and the CPCU “real-time” (“RT”) for the period 2006-present.

There are two main reasons for the switch from the GPCP that was used for the preliminary MERRA-Land product of Reichle et al. (2011) to the CPCU precipitation product that is used for the final MERRA-Land product. First, the CPCU product is available with latency on the order of days so that the MERRA-Land data stream can be updated to the present along with the MERRA data. This is not possible with the GPCP product because its usual latency is on the order of months. Second, the CPCU precipitation product is nominally at 0.5° and daily resolution and was found to improve the mid-latitude skill of the MERRA-Land soil moisture and runoff estimates beyond that shown in Reichle et al. (2011). See Appendix C.6 for details.

Because at 0.5° the spatial resolution of the CPCU precipitation product is similar to that of the background MERRA data that are being corrected, the precipitation correction algorithm was modified from that described in Reichle et al. (2011). We started the precipitation corrections by re-gridding the CPCU data to the MERRA grid (rather than mapping the MERRA precipitation to the grid of the precipitation product as in Reichle et al. (2011)). Scaling factors were then computed on the MERRA grid, separately for each MERRA grid cell and for each day in the experiment period, by dividing the CPCU daily precipitation by the daily-accumulated precipitation from the corresponding 24 MERRA hourly estimates. Note that in this step the CPCU “daily” precipitation totals refer to *different* 24-hour periods based on the

geographic location (see CPCU documentation for details). For each grid cell we therefore carefully selected the MERRA hourly files that contribute to the “daily” accumulation in the CPCU data for this particular grid cell. Finally, the MERRA hourly estimates of total precipitation (PRECTOT), snowfall (PRECSNO), and convective precipitation (PRECCONV) from the MERRA “flx” file collection were multiplied with the corresponding scaling factors to obtain the corrected MERRA precipitation forcing that was used for MERRA-Land.

Whenever, for a given grid cell and a given day, MERRA has zero precipitation but CPCU has non-zero precipitation the scaling factor cannot be computed (division by zero). In this case, we added the “missing” CPCU precipitation evenly to the three MERRA hourly estimates from midnight to 3am local time. Nighttime was selected to avoid inconsistencies with the MERRA shortwave downward radiation. By contrast, in the preliminary MERRA-Land product of Reichle et al. (2011) this “missing” precipitation was ignored.

### ***C.4 Total Land Water Storage***

MERRA’s total land water storage variable (TWLAND) occasionally contains negative values, typically in desert areas, because of a suboptimal reference baseline used for computing TWLAND from the Catchment model’s soil moisture excess and deficit variables. This has been changed in MERRA-Land.

Generally, for each land surface tile TWLAND is computed as follows:

$$TWLAND = MAXSOILWAT - CATDEF + SRFEXC + RZEXC + SNOMAS + CAPAC$$

where CATDEF, SRFEXC, and RZEXC are the Catchment model’s soil water prognostic variables (catchment deficit, surface excess, and root zone excess), SNOMAS is the total snow water equivalent, and CAPAC the total interception water (all variables are in units of  $\text{kg m}^{-2}$ ). The difference in the definition of the TWLAND variables output by MERRA and MERRA-Land lies in the reference baseline used for soil water, MAXSOILWAT:

$$\textbf{MERRA:} \quad \quad \quad MAXSOILWAT = CDCR2$$

$$\textbf{MERRA-Land:} \quad \quad \quad MAXSOILWAT = CDCR2 / (1 - WPWET)$$

where CDCR2 is the maximum water holding capacity of the land surface tile ( $\text{kg m}^{-2}$ ) and WPWET is the corresponding wilting point, in units of degree of saturation. Note that the issue is entirely confined to the "TWLAND" output field, which is only offset by a value that is constant in time, albeit different for each tile and each grid cell.

### ***C.5 Snow Cover Fraction***

The snow model component of the Catchment land surface model uses an important parameter, WEMIN, which describes the minimum snow water equivalent that must be present per unit surface area before the model considers the surface to be snow covered. If the amount of snow (SNOMAS) on a given tile is insufficient to cover the entire tile with at least an amount of WEMIN, the snow cover fraction (FRSNO) is the fraction of the tile area that would be covered by an amount equal to WEMIN, that is,  $FRSNO = \min( 1., SNOMAS/WEMIN )$ . Note that the snow depth (SNODP) is the actual depth of the snow in just the snow-covered fraction.

Between the MERRA (GEOS-5.2.0) and the MERRA-Land (GEOS-5.7.2) versions of the Catchment model, changes were made to WEMIN and to the maximum depth of the uppermost snow layer (DZ1MAX) to improve the modeled albedo and the stability of the surface calculation when snow is present (not shown). Specifically, MERRA-Land uses WEMIN=26 kg m<sup>-2</sup> and DZ1MAX=0.08 m, whereas MERRA used WEMIN=13 kg m<sup>-2</sup> and DZ1MAX=0.05 m. Primarily due to the change in WEMIN, the snow cover fraction (FRSNO) values in the MERRA-Land data product are thus consistently lower than those of MERRA. Preliminary results indicate that MERRA snow cover fraction estimates are in better agreement with MODIS observations than those of MERRA-Land (not shown). For applications of MERRA-Land that require snow cover fraction estimates, it may thus be necessary to bias-adjust the FRSNO field accordingly.

## C.6 Update of Reichle et al. 2011 Science Results

In this section we present the key figures of Reichle et al. (2011), updated with results for the official MERRA-Land data product as described in this document.

Figure C.6-1 documents the climatological difference between the CPCU gauge-based precipitation product used for the official MERRA-Land product and the GPCPv2.1 pentad product used for the preliminary product of Reichle et al. (2011). The CPCU product has generally lower precipitation amounts than GPCP, which leads to lower soil moisture values and lower latent heat flux values (see below).

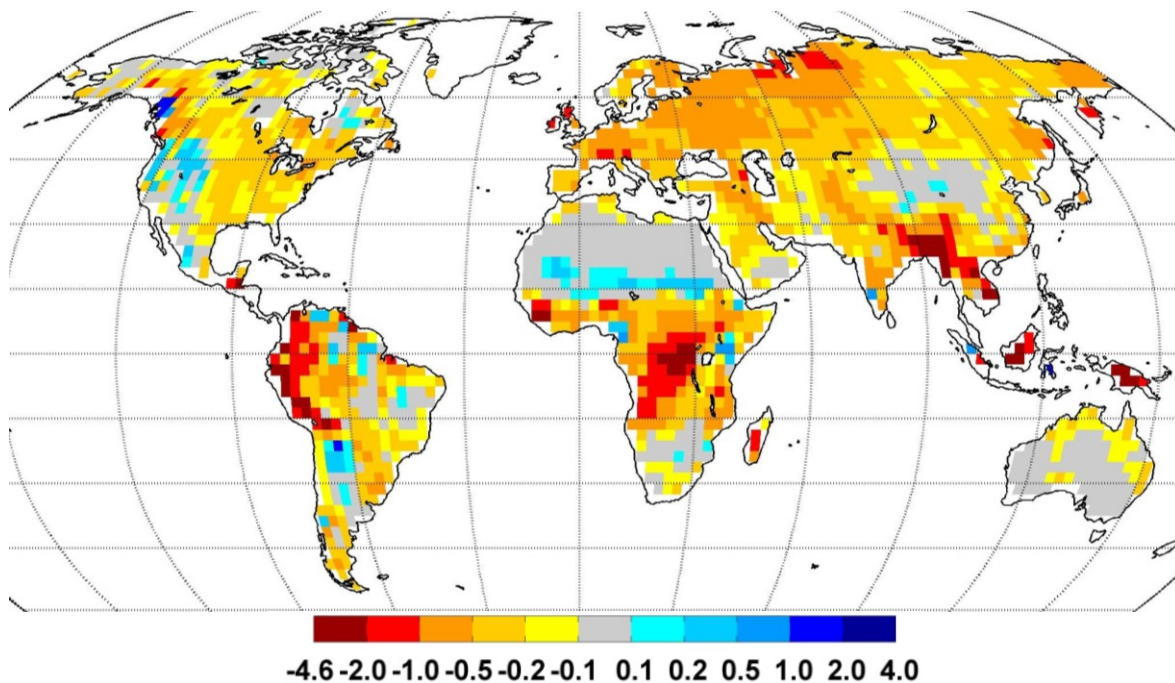


Figure C.6-1. CPCU minus GPCP (v2.1 pentad) precipitation difference ( $\text{mm d}^{-1}$ ) averaged over the period 1981-2008. The global mean difference is  $-0.4 \text{ mm d}^{-1}$ .

Figure C.6-2 corresponds to Figure 4(c) of Reichle et al. (2011) and shows that the interception ratio of the final MERRA-Land data product is very similar to that of the preliminary product of Reichle et al. (2011).

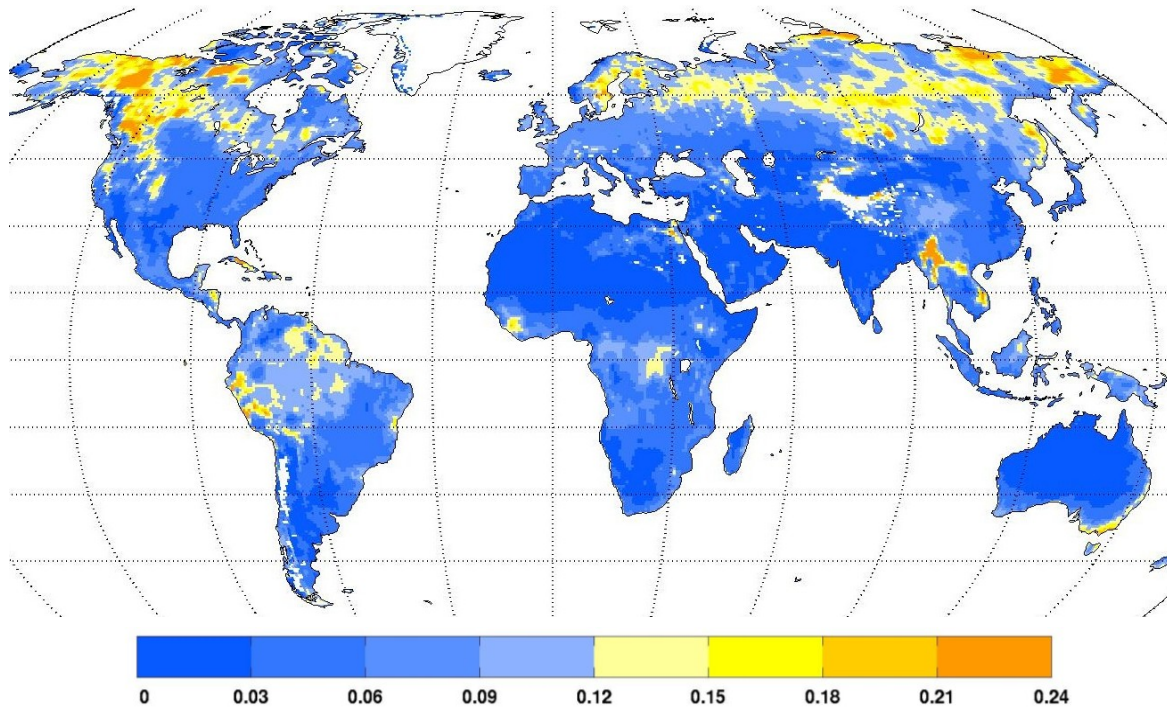


Figure C.6-2. 2003-2007 average interception loss fraction (dimensionless) for MERRA-Land. Global average is 0.06. Same as Figure 4(c) of Reichle et al. (2011) except that MERRA-Land has been replaced with the official MERRA-Land data product described in this document.

Figures C.6-3 and C.6-4 correspond to Figures 5(b) and 6(b), respectively, of Reichle et al. (2011). Due to the lower values of precipitation in the CPCU product relative to the GPCPv2.1 pentad product, the final MERRA-Land product has lower latent heat flux values and lower soil moisture values than the preliminary product of Reichle et al. (2011).

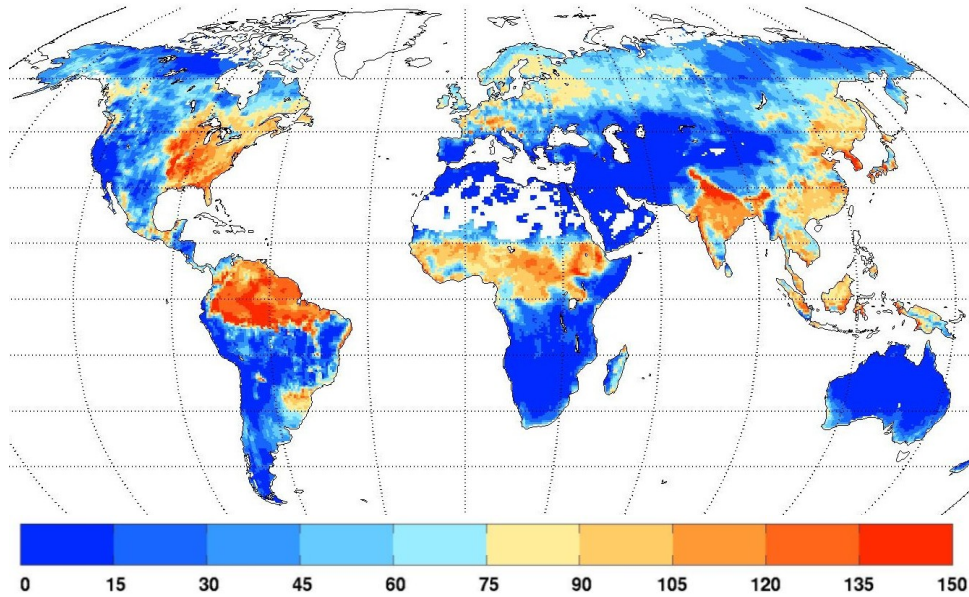


Figure C.6-3. Average latent heat flux ( $\text{W m}^{-2}$ ) for August 1994 for MERRA-Land. Global average is  $49 \text{ W m}^{-2}$ . Same as Figure 5(b) of Reichle et al. (2011) except that MERRA-Land has been replaced with the official MERRA-Land data product described in this document.

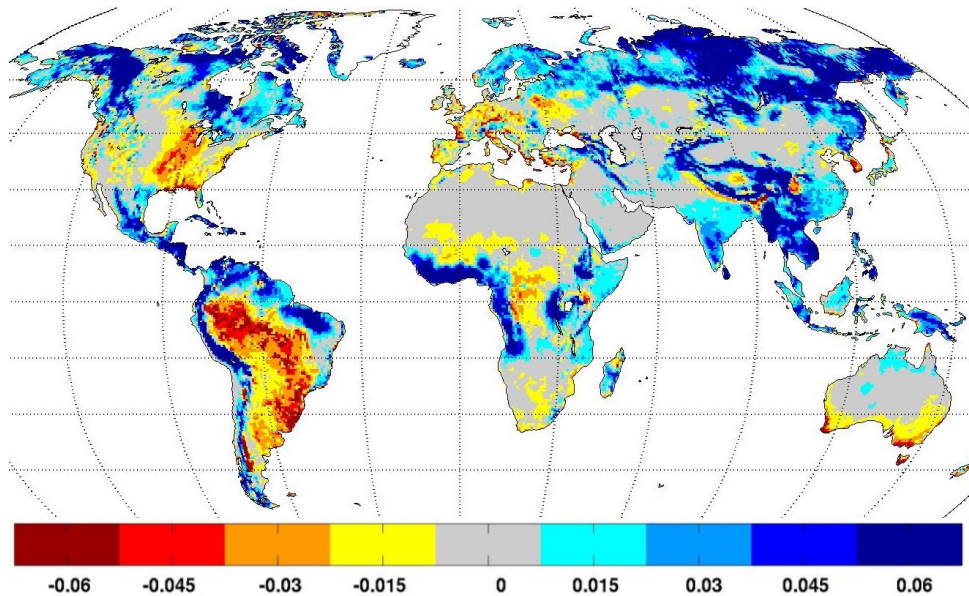


Figure C.6-4. Annual average root zone soil moisture ( $\text{m}^3 \text{ m}^{-3}$ ) differences (1981-2008) for MERRA minus MERRA-Land. Same as Figure 6(b) of Reichle et al. (2011) except that MERRA-Land has been replaced with the official MERRA-Land data product described in this document.

Figure C.6-5 corresponds to Figure 7 of Reichle et al. (2011) and shows that the soil moisture skill vs. SCAN in situ measurements is significantly higher in the final MERRA-Land product when compared to the preliminary MERRA-Land product of Reichle et al. (2011).

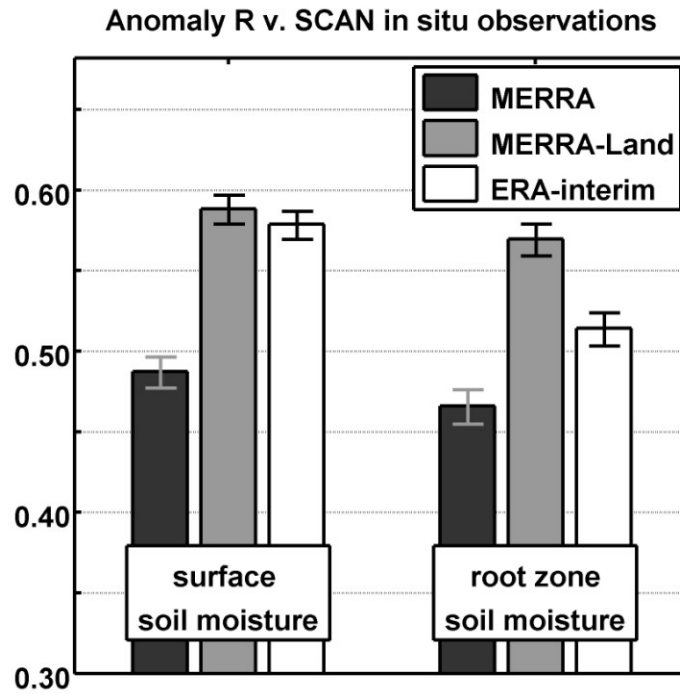


Figure C.6-5. Skill (pentad anomaly R; dimensionless) of MERRA, MERRA-Land, and ERA-Interim estimates (2002-2009) versus SCAN in situ surface and root zone soil moisture measurements. Error bars indicate approximate 95% confidence intervals. Same as Figure 7 of Reichle et al. (2011) except that MERRA-Land has been replaced with the official MERRA-Land data product described in this document.

Figure C.6-6 corresponds to Figure 8 of Reichle et al. (2011) and shows that the runoff skill is slightly higher overall in the official MERRA-Land product when compared to the preliminary MERRA-Land product of Reichle et al. (2011). The figure also corrects minor errors in processing that also affect the skill values for MERRA and ERA-Interim in three basins (Missouri at Ft. Randall Dam, Arkansas at Ralston, and Missouri at Garrison Reservoir).

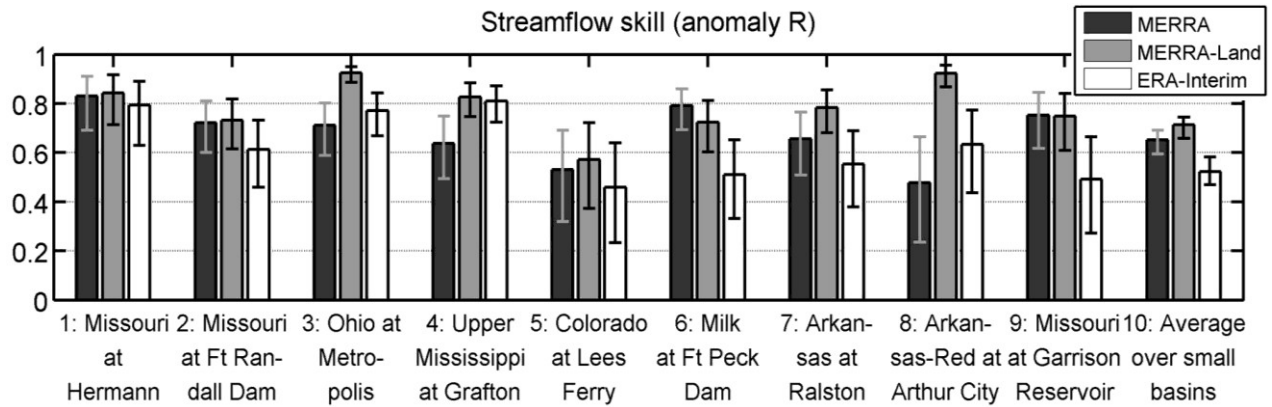


Figure C.6-6. Seasonal anomaly time series correlation coefficients (dimensionless) for runoff estimates from MERRA, MERRA-Land, and ERA-Interim for the basins and time periods listed in Table 1 of Reichle et al. (2011). Same as Figure 8 of Reichle et al. (2011) except that MERRA-Land has been replaced with the official MERRA-Land data product described in this document.



Figure C.6-7 corresponds to Figure 9(a) of Reichle et al. (2011) and shows that on average the snow depth skill vs. CMC estimates is similar for the official MERRA-Land product and the preliminary MERRA-Land product of Reichle et al. (2011). At very high latitudes, however, the skill of the final MERRA-Land snow depth estimates is lower than that of the preliminary product, which is likely due to the lower quality of the CPCU precipitation estimates relative to the GPCPv2.1 data.

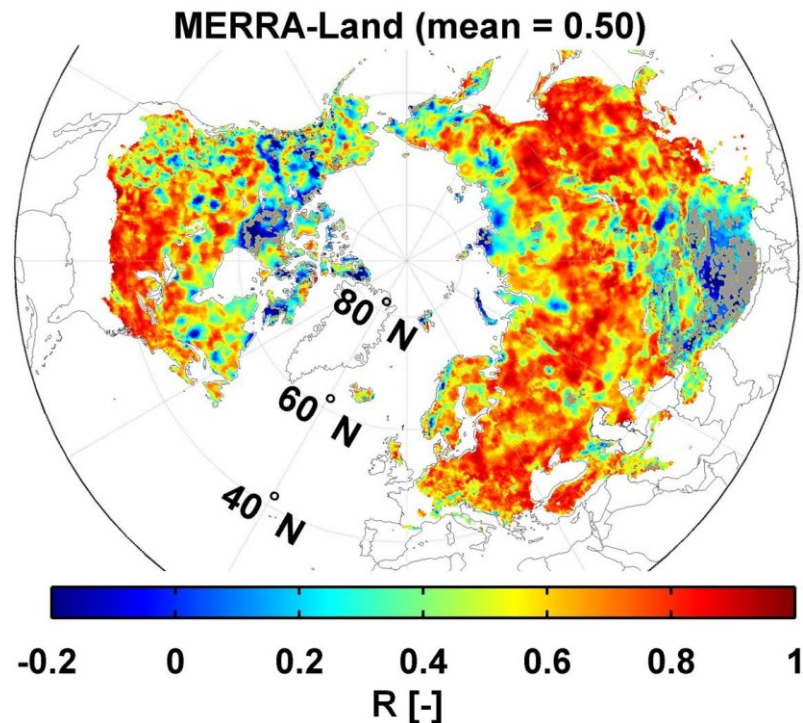


Figure C.6-7. Skill (pentad anomaly R) of MERRA-Land snow depth versus CMC estimates (September 1998 – September 2009). R values that are not statistically different from zero at the 5% significance level are shown in gray. Global average is R=0.5. Same as Figure 9(a) of Reichle et al. (2011) except that the official MERRA-Land data product described in this document is used.

## References

- Chen, M., and Coauthors (2008), Assessing objective techniques for gauge-based analyses of global daily precipitation, *J. Geophys. Res.*, 113, D04110, doi:10.1029/2007JD009132.
- deWitt, S. (1996), Writing HDF-EOS Grid Products for Optimum Subsetting Services. ECS Technical Paper 170-TP-007-001.
- Dopplick, T. (1997), The Role of Metadata in EOSDIS. EOSDIS Technical Paper 160-TP-013-001.
- Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model 2. Parameter estimation and model demonstration, *J. Geophys. Res.*, **105**, 24823–24838, doi:10.1029/2000JD900328.
- Gross, C. (1997), B.0 Earth Science Data Model for the ECS Project. EOSDIS Technical Paper 420-TP-015-001.
- Koster, R. D., M. J. Suárez, A. Ducharne, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a GCM, Part 1, Model Structure. *J. Geophys. Res.*, **105**, 24809-24822, doi:10.1029/2000JD900327.
- Lucchesi, R. (2012), File Specification for MERRA Products. GMAO Office Note No. 1 (Version 2.3), 82 pp, available from [http://gmao.gsfc.nasa.gov/pubs/office\\_notes](http://gmao.gsfc.nasa.gov/pubs/office_notes).
- NOAA (1995), Conventions for the Standardization of NetCDF Files (COARDS). [http://ferret.wrc.noaa.gov/noaa\\_coop/coop\\_cdf\\_profile.html](http://ferret.wrc.noaa.gov/noaa_coop/coop_cdf_profile.html)
- Pfaffstetter, O. (1989), Classification of hydrographic basins: coding methodology, unpublished manuscript, Departamento Nacional de Obras de Saneamento, August 18, 1989, Rio de Janeiro; available from J.P. Verdin, U.S. Geological Survey, EROS Data Center, Sioux Falls, South Dakota 57198 USA. See, for example: Verdin, K. L. and J. P. Verdin, (1999), A topological system for delineation and codification of the Earth's river basins, *J. Hydrology*, 218, 1-12, or: <http://gis.esri.com/library/userconf/proc01/professional/papers/pap1008/p1008.htm>
- Reichle, R. H., and Coauthors (2011), Assessment and enhancement of MERRA land surface hydrology estimates, *J. Clim.*, 24, 6322-6338, doi:10.1175/JCLI-D-10-05033.1.
- Rienecker, M. M., and Coauthors (2011), MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, *J. Clim.*, 24, 3624-3648, doi:10.1175/JCLI-D-11-00015.1.
- Rienecker, M.M., and Coauthors (2008), The GEOS-5 Data Assimilation System - Documentation of Versions 5.0.1, 5.1.0, and 5.2.0. *Technical Report Series on Global Modeling and Data Assimilation 104606*, v27. Available at [http://gmao.gsfc.nasa.gov/pubs/docs/GEOS5\\_104606-Vol27.pdf](http://gmao.gsfc.nasa.gov/pubs/docs/GEOS5_104606-Vol27.pdf)

## Web Resources

GMAO web site: <http://gmao.gsfc.nasa.gov/>

GMAO Data Products page: <http://gmao.gsfc.nasa.gov/products/>

CF Standard Description: <http://cf-pcmdi.llnl.gov/>

The HDF Group: HDF-4 information: <http://www.hdfgroup.org/>

GEOS-5 Variable Definition Glossary, available from:  
<https://gmao.gsfc.nasa.gov/products/documents/>

NASA, 1999: HDF-EOS Standards: <http://earthdata.nasa.gov/data/references/>