File Specification for MERRA-2 Stratospheric Composition Reanalysis of Aura MLS (M2-SCREAM)

Document Version: 1.0
Date: June 21st 2022

Table of Contents

1. Introduction

This document provides a brief description of the output collections from the MERRA-2 Stratospheric Composition Reanalysis of Aura MLS (M2-SCREAM) produced at NASA's Global Modeling and Assimilation Office. These data are generated by assimilating MLS and OMI retrievals into the GEOS Constituent Data Assimilation System (CoDAS) driven by meteorological fields from MERRA-2. M2-SCREAM assimilates hydrochloric acid (HCl), nitric acid (HNO_3), stratospheric water vapor (H_2O), nitrous oxide (N_2O) and ozone with a system equipped with a version of the GEOS general circulation model and a stratospheric chemistry model, StratChem. Assimilated fields are provided globally at 0.5˚ by 0.625˚ resolution at three-hourly frequencies from 2004/09/01 to 2021. Assimilation uncertainties for each of the assimilated constituents are calculated from the CoDAS statistical output (Wargan et al., 2022) and provided as global full-resolution three-dimensional monthly files.

2. System configuration and data description

2.1 System configuration

M2-SCREAM is produced by assimilation of MLS Version 4.2 retrieved stratospheric constituent profiles and OMI total ozone with a CoDAS configuration that combines the GEOS general circulation model, version "Icarus-3_2_p9", and the StratChem stratospheric chemistry model. Temperature, winds, surface pressure and tropospheric water vapor from MERRA-2 are used to force the GEOS model using the *replay* method (Orbe et al. 2017). HCl, HNO_3, H_2O, N_2O, and ozone are assimilated using the three-dimensional variational analysis (3DVar) with Incremental Analysis Update (IAU) as described in Wargan et al., (2020) and Wargan et al., (2022). M2-SCREAM has been generated in three production streams covering periods between September 2004 and August 2010, between September 2010 and August 2015, and from September 2015 onwards.

2.2 Details of spatial and temporal resolution and coordinate systems

The three-dimensional grid structure in the M2-SCREAM output collections is the same as that in other GEOS products, including MERRA-2. Data are provided on a longitude/latitude grid on 72 vertical hybrid sigma-p layers between the surface and 0.01 hPa, although generally only stratospheric constituent data are recommended for scientific use. The horizontal resolution is 0.5˚ by 0.625˚ (latitude by longitude). There are 576 and 361 longitude and latitude points running from 180˚W to 179.375˚E and from 90˚S to 90˚N, respectively. The vertical resolution in most of the stratosphere is 1.1–1.2 km. The hybrid sigma-p layers are surface pressure-following. The pressure at the model top is constant PTOP=1 Pa = 0.01 hPa. The pressures of layer boundaries can be calculated by starting at PTOP and summing the successive layer thicknesses, DELP provided in the output collections. The representative mid-layer pressures are calculated

as arithmetic averages of the pressures of corresponding layer boundaries. For convenience the representative mid-layer pressures are provided in the output collections (the “PL" fields). Note that the unit of pressure used in the GEOS output is Pa. The indexing for the GEOS vertical coordinates is top to bottom.

2.3 Vertical shift correction

A coding error, identified after the reanalysis was completed for the period 2004 – April 2021, resulted in an upward shift of the assimilated water vapor, HCl, HNO_3, and N_2O fields by half the model layer, or approximately 0.5 km. A correction was applied to the affected fields in post-processing. It was determined that the corrected fields are unbiased with respect to the assimilated data although the initial error and the correction procedure added a small additional uncertainty to these fields. That uncertainty has been quantified using a separate assimilation experiment and included in the uncertainty information provided with the reanalysis (see section 3). See Wargan et al. (2022) for details.

2.4 Data location

The M2-SCREAM output can be accessed from Goddard Earth Sciences Data and Information Services Center (https://disc.gsfc.nasa.gov).

3. Format and File organization

Two collections are provided to the users: (1) instantaneous composition and meteorological fields at a three-hourly resolution and (2) monthly assimilation uncertainties. M2-SCREAM three-hourly data files are provided in NetCDF-4 format. The file organization is very similar to that used in other GMAO products, such as MERRA-2. Monthly uncertainties associated with the M2-SCREAM assimilated variables are also provided in NetCDF-4 format. These are computed from the CoDAS internal diagnostic output binned into 10˚ by 10˚ longitude latitude grid cells and mapped back onto the M2-SCREAM native model grid. The uncertainty fields in the regions with no observations, such as most of the troposphere, are filled with missing values. In addition to assimilation uncertainties, (2) includes time-independent uncertainty estimates associated with the vertical shift correction of the HCl, HNO_3, H_2O and N_2O fields (see Section 2.3) for the period between October 2004 and April 2021. These additional fields are not included in the uncertainty files with time stamps starting May 2021. All uncertainties are provided as 1-sigma (one standard deviation).

3.1 Dimensions

Each of the two M2-SCREAM collections contains variables that define the dimensions of longitude, latitude, vertical level (for all variables except surface pressure), and time. Dimension variables have an attribute named “units,” set to an appropriate string defined by the CF and

COARDS conventions (NOAA 1995) that can be used by applications to identify the dimension. The M2-SCREAM dimension variables are listed in Table 1.

Table 1. Dimension variables contained in M2-SCREAM files

Name	**Description**	**Type**	***Units* attribute**
lon	Longitude	double	degrees_east
lat	Latitude	double	degrees_north
lev	Layer index or globally averaged pressure (for monthly data)	double	layer
time	Time	int	Minutes or months (for monthly data)

3.2 Variables

Variables in the three-hourly assimilated output and the monthly uncertainty collections are stored as HDF-5 dataset objects. M2-SCREAM uses the "classic" NetCDF data model and does not use any of the extensions supported by NetCDF-4 and the underlying HDF-5 format. This allows applications written to read NetCDF files to easily read variables without having to modify code. Variables available in the M2-SCREAM output are listed in Section 4 along with information about sizes and dimensions. Monthly uncertainties are stored in a separate collection in NetCDF format.

Each variable has several metadata attributes. Many of these attributes are required by the CF and COARDS conventions, while others are specific for GMAO products. Table 2 lists required attributes and those that are useful to the users. Other attributes may be included for internal GMAO use and can be ignored.

Table 2. Metadata attributes associated with each variable in the three-hourly output. Attributes that also exist in the monthly analysis uncertainty output are in bold.

Name	**Type**	**Description**
long_name	**String**	**A brief description of the variable contents**
units	**Char string**	**The units of the variable**
_FillValue	**32-bit float**	**Floating-point value used to identify missing data.**
missing_value	**32-bit float**	**Same as _FillValue. Required for COARDS backwards compatibility.**
scale_factor	32-bit float	If variable is packed as 16-bit integers, this is the scale_factor for expanding to floating-point. Currently data are not packed, thus value is 1.0.
add_offset	32-bit float	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.
standard_name	String	Same as long_name
valid_range	32-bit float 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 M2-

		SCREAM files these are set to -/+ _FillValue.

3.3 Global metadata

Global metadata are stored in the three-hourly M2-SCREAM assimilation collection as NetCDF-4 global attributes. Some of these metadata are required by CF/COARDS conventions and others represent additional information as a convenience to the M2-SCREAM users. All global metadata are of type character. Table 3 summarizes the most important global metadata attributes.

Table 3. Global metadata attributes (type: character).

Name	**Description**
_NCProperties	Library versions
Filename	Filename of this granule
GranuleID	The same as filename
comment	additional information
ShortName	Product short name used by G-DISC
LongName	Description of product type.
title	The same as LongName
Format	“NetCDF-4/HDF-5” or “NetCDF”
Conventions	Identification of the file convention used, currently “CF-1”
VersionID	Release version
SouthernmostLatitude	“-90.0”
NorthernmostLatitude	“90”
WesternmostLongitude	“-180.0”
EasternmostLongitude	“179.375”
LatitudeResolution	“0.5”
LongitudeResolution	“0.625”
RangeBeginningDate	Date corresponding to the first timestep in this file.
RangeBeginningTime	Time corresponding to the first timestep in this file.
RangeEndingDate	Date corresponding to the last timestep in this file.
RangeEndingTime	Time corresponding to the last timestep in this file.
institution	"NASA Global Modeling and Assimilation Office"
source	Data assimilation and model version or source of uncertainty estimates
ProcessingLevel	“Level 4”
DataSetQuality	Brief summary of data quality and references
post_processing	Compression and vertical shift correction applied for data prior to May 2021
ProductionDateTime	Production date & time of this granule.
IdentifierProductDOIAuthority	"https://doi.org/"
IdentifierProductDOI	Unique Digital Object Identifier
ProjectFunding	“NASA Modeling Analysis and Prediction”

PI	Principal Investigator
references	Description and validation publication
ContactPersonAddress	Contact email
history	Production date & time of this granule.

4. File Naming Conventions

Three-hourly instantaneous files containing assimilated constituent fields and replayed meteorological variables are named
M2SCREAM.inst3_3d_asm_met_chm_Nv.**yyyymmdd_hhhh**z.nc4,
where **yyyymmdd** and **hhhh** reflect the four-digit year, month, day, and hour of the date whose contents are reported in the file.

Monthly uncertainties are stored in files named
M2SCREAM.uncertainties.**yyyymm**.nc4,
where **yyyymm** denotes the four-digit year and month corresponding to the file.

5. Available Data

Data are provided in netCDF format in the following two collections:

inst3_3d_asm_met_chm_Nv: *assimilated constituents and replayed meteorological fields*
Frequency: *Daily, containing 1 daily value*
Dimensions: *longitude=576, latitude=361, time=1*
Granule Size: *~240 MB*
Filename: M2SCREAM.inst3_3d_asm_met_chm_Nv.yyyymmdd_hhhhz.nc4 *where **yyyymmdd** and **hhhh** reflect the four- digit year, month, day and hour of the date whose contents are reported in the file.*
Short name: GMAO_M2SCREAM_INST3_CHEM
doi: 10.5067/7PR3XRD6Q3NQ

Science Variables

Name	***Dim***	***Description***	***Units***
DELP	tzyx	Pressure thickness	Pa
EPV	tzyx	Ertel's potential vorticity	K m^2 kg^{-1} s^{-1}
H	tzyx	Mid-layer heights	m
HCL	tzyx	Hydrochloric acid (assimilated)	mol mol^{-1}
HNO3	tzyx	Nitric acid (assimilated)	mol mol^{-1}
HNO3COND	tzyx	Condensed nitric acid	mol mol^{-1}
N2O	tzyx	Nitrous oxide (assimilated)	mol mol^{-1}
O3	tzyx	Ozone (assimilated)	ppmv
OMEGA	tzyx	Vertical pressure velocity	Pa s^{-1}
PL	tzyx	Mid-layer pressure	Pa
PS	tyx	Surface pressure (replayed)	Pa
QV	tzyx	Specific humidity (assimilated/replayed)	kg kg^{-1}

T	tzyx	Air temperature (replayed)	K
U	tzyx	Eastward wind (replayed)	m s^{-1}
V	tzyx	Northward wind (replayed)	m s^{-1}

M2SCREAM.uncertainties: *assimilation uncertainties, additional uncertainties from shift correction*[(a)]

Frequency: *Monthly, containing 1 monthly value*
Dimensions: *longitude=576, latitude=361, time=1*
Granule Size: *~572 MB, ~*343 *MB (see Section 2.3)*
Filename: M2SCREAM.uncertainties.yyyymm.nc4 where **yyyymm** *reflects the four-digit year and month.*
Short name: GMAO_M2SCREAM_MONTH_UNCERT
doi: 10.5067/7XRIJO9OP8PE

Science Variables

Name	***Dim***	***Description***	***Units***
uncert_h2o	tzyx	Specific humidity uncertainty	mol mol^{-1}
uncert_hcl	tzyx	Hydrochloric acid uncertainty	mol mol^{-1}
uncert_hno3	tzyx	Nitric acid uncertainty	mol mol^{-1}
uncert_n2o	tzyx	Nitrous oxide uncertainty	mol mol^{-1}
uncert_o3	tzyx	Ozone uncertainty	ppmv
uncert_rep_h2o[(a)]	tzyx	Specific humidity uncertainty from shift correction	mol mol^{-1}
uncert_rep_hcl[(a)]	tzyx	Hydrochloric acid uncertainty from shift correction	mol mol^{-1}
uncert_rep_hno3[(a)]	tzyx	Nitric acid uncertainty from shift correction	mol mol^{-1}
uncert_rep_n2o[(a)]	tzyx	Nitrous oxide uncertainty from shift correction	mol mol^{-1}
PL	tzyx	Average mid-layer pressure	Pa

[(a)]Between October 2004 and April 2021

6. Contact

Krzysztof Wargan, Global Modeling and Assimilation Office, Code 610.1, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. (krzysztof.wargan-1@nasa.gov)

7. Acknowledgement of Funding

This work has been supported by NASA's Modeling Analysis and Prediction

8. References

NOAA, 1995: Conventions for the Standardization of NetCDF Files (COARDS). http://ferret.wrc.noaa.gov/noaa_coop/coop_cdf_profile.html

Orbe, C., Oman, L. D., Strahan, S. E., Waugh, D. W., Pawson, S., Takacs, L. L., & Molod, A. M. (2017). Large-scale atmospheric transport in GEOS replay simulations. *J. Adv. Mod. Earth Sys.*, 9. https://doi. org/10.1002/2017MS001053

Wargan, K., Weir, B., Manney, G. L., Cohn, S. E., & Livesey, N. J. (2020). The anomalous 2019 Antarctic ozone hole in the GEOS Constituent Data Assimilation System with MLS observations. *Journal of Geophysical Research: Atmospheres*, 125, e2020JD033335. https://doi.org/10.1029/2020JD033335

Wargan, K., Weir, B., Manney, G. L., Cohn, S. E., K.E. Knowland, P. Wales & Livesey, N. J. (2022). M2-SCREAM: A Stratospheric Composition Reanalysis of Aura MLS data with MERRA-2 transport. Earth Sys. Sci. Dat., *submitted*.