*National Aeronautics and Space Administration
Goddard Earth Science Data Information and
Services Center (GES DISC)*

README Document for MODIS Collection 6.1 Equal-Angle Three-Hourly Cloud Regimes

Last Revised June 29, 2021

Goddard Earth Sciences Data and Information Services Center (GES DISC)
http://disc.gsfc.nasa.gov
NASA Goddard Space Flight Center
Code 610.2
Greenbelt, MD 20771 USA

Prepared By:

Lazaros Oreopoulos

Name
Climate and Radiation Laboratory
GSFC Code 613

Nayeong Cho

Name
Climate and Radiation Laboratory
GSFC Code 613

Jackson Tan

Name
Climate and Radiation Laboratory
GSFC Code 613

July 8, 2021

Date

Reviewed By:

Jerome Alfred

Reviewer Name
GES DISC
GSFC Code 610.2

July 8, 2021

Date

Goddard Space Flight Center
Greenbelt, Maryland

Revision History

Revision Date	Changes	Author

Table of Contents

1.0 Introduction

This document provides basic information for using the Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 6.1 Equal-Angle Three-Hourly Cloud Regime product. This product is a discrete classification of cloud fields at the mesoscale as observed by the MODIS sensors aboard the Terra and Aqua satellites. Derived by applying the *k*-means clustering algorithm to joint histograms of cloud top pressure and cloud optical thickness, the cloud regimes represent different atmospheric systems based on their cloud signatures.

1.1 Dataset/Mission Instrument Description

The MODIS instrument is a cross-track scanning radiometer with 36 spectral bands of wavelengths ranging from 0.4 µm to 14.4 µm. There are two MODIS instruments in operation, one aboard the Terra satellite and one aboard the Aqua satellite. The Terra satellite flies in a sun-synchronous orbit with equator-crossing times of 10:30 AM and PM local time, while the Aqua satellite flies, as part of the A-Train constellation, in a sun-synchronous orbit with equator-crossing times of 1:30 AM and PM local time. Compared to previous generations of radiometers, the additional spectral channels in MODIS provide unique capability in retrieving cloud properties globally. These properties include cloud top pressure, cloud top temperature, cloud optical thickness, cloud effective radius, and cloud phase, available at the sensor footprint level or Level-2 products (Platnick et al. 2015) and gridded level or Level-3 products (Platnick et al. 2017).

Cloud Regimes provide an effective way of classifying atmospheric systems into a discrete set of states. First applied to the International Satellite Cloud Climatology Project (ISCCP) cloud observations (Jakob and Tselioudis 2003; Rossow et al. 2005), Cloud Regimes identify recurring patterns in the distribution of cloud properties at the mesoscale and categorize them into a manageable number of categories. In grouping together cloud patterns that are similar, Cloud Regimes can overcome inconsequential differences arising from minor variations and measurement uncertainties, and capture the essence of underlying mesoscale phenomena.

This Cloud Regimes product is derived from ISCCP-like joint histograms of Cloud Top Pressure (CTP) and Cloud Optical Thickness (COT) from Collection 6.1 of the MODIS atmospheric products. Compared to the previous version (Collection 6), Collection 6.1 contains corrections to the underlying MODIS channel radiances (Level-1b data) used to perform the retrievals. These affected cloud masking, which in turn had some impact on cloud property retrievals, though the overall bearing on the cloud property products involved in the derivation of the Cloud Regimes was minor. See https://atmosphere-imager.gsfc.nasa.gov/documentation/collection-61.

One key difference of this product from previous MODIS Cloud Regime products is the use of the ISCCP HGG equal-area grid at the 3-h timescale to facilitate further comparison and analysis with ISCCP products (Young et al. 2018). This grid is approximately 110 km (~1° at the equator) in size, which is more spatially representative as a function of latitude. Since there are no official MODIS products that provide joint histograms on this grid, we perform custom gridding of the Level-2 (pixel-level) cloud and aerosol observations from the MOD04_L2 (MYD04_L2) and MOD06_L2 (MYD06_L2) Terra (Aqua) products at every 3-h window centered on the nominal time. The sampling follows that of the MODIS Level-3 1° joint histogram (MOD08_D3 and MYD08_D3 products): (1) using only pixels with successful CTP and COT retrievals (hence limited to sunlit hours); (2) using every fifth COT retrieval in both the along-track and cross-track direction but with an offset in the former to avoid a defective sensor; and treating both nominal and PCL ("partially cloudy") retrievals separately. The Cloud Regimes are then interpolated from the 110 km equal-area grid to a 1° equal-angle grid using the nearest neighbor approach.

The custom-gridded ISCCP-like joint histograms, consisting of 7 CTP classes and 6 COT classes, were subjected to a *k*-means clustering procedure to derive MODIS Cloud Regimes. We selected only joint histograms formed from at least 120 pixels, which we deemed to provide a representative spatial sampling of the grid cell. Furthermore, the sun-synchronous orbits of Terra and Aqua naturally results in an over-representation of observations at higher latitudes; therefore, we subsampled the joint histograms constructed from Level-2 data based on their latitude, such that their meridional sampling is latitudinally representative, before ingestion into the clustering algorithm. Using observations from the period of 2003 to 2018, the *k*-means clustering produces 11 Cloud Regimes, with grid cells that are completely clear forming a 12th "clear sky" Cloud Regime; details for the choice of 11 clusters are discussed in Cho et al. (2021). All globally available grid cell joint histogram were then assigned to the closest cluster centroid based on minimum Euclidean distance to identify the Cloud Regime. This process is also applied joint histograms from outside the clustering period, which allows us to extend the record without repeating the clustering process.

The MODIS Collection 6.1 Equal-Angle Three-Hourly Cloud Regime product and its documentation are available at https://disc.gsfc.nasa.gov/datacollection/MODIS_CR_Equal_Angle_3h_1.0.html.

1.2 Data Disclaimer

1.2.1 Data Citation and Acknowledgment

If you use this data in a publication, please acknowledge the project. An example acknowledgment statement is:

“We acknowledge the MODIS Cloud Regime Team at NASA Goddard Space Flight Center for making the MODIS Cloud Regime data available.”

Please also cite Cho et al. (2021) and the data product:

Cho, N., Tan, J., & Oreopoulos, L. (2021). Classifying planetary cloudiness with an updated set of MODIS Cloud Regimes, Journal of Applied Meteorology and Climatology, https://doi.org/10.1175/JAMC-D-20-0247.1.

Cho, N., J. Tan, L. Oreopoulos (2021), MODIS Collection 6.1 Equal-Angle Three-Hourly Cloud Regimes, Version 1.0, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [Data Access Date], https://doi.org/10.5067/MEASURES/MODISCR/EQANG3H/DATA301

1.2.2 Contact Information

Please contact Lazaros Oreopoulos (lazaros.oreopoulos@nasa.gov) for any queries about this product.

2.0 Data Organization

The data consists of Cloud Regime identification at every three-hour interval in equal-angle grid cells (1°) compiled into yearly files.

2.1 File Naming Convention

The product is organized into yearly files. The file naming convention is:

MODIS_CR.EqualAngle_3h.C61.V1.0.L3.YYYY.nc4

where YYYY is the 4 digit year number. For example, the file containing data for the year 2015 is MODIS_CR.EqualAngle_3h.C61.L3.2015.nc4. The “C61” refers to MODIS Collection 6.1 from which this product is derived; the “V1.0” refers to the Version 1.0 of this product; the “L3” refers to this product’s Level-3 nature.

2.2 File Format and Structure

The MODIS_CR_Equal_Area_3h files are in NetCDF-4 format. NetCDF is a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data that was developed by UCAR/Unidata (http://doi.org/10.5065/D6H70CW6) https://www.unidata.ucar.edu/software/netcdf/. This product is compliant with the Climate and Forecast (CF) convention, version 1.7.

All variables in the files are placed in the root group. For geophysical variables, the spatial dimensions are given by the dimensions "longitude" and "latitude" while the temporal dimension is given by the dimension "time". See Sec. 3.0 for more details.

2.3 Key Science Data Fields

The variables that provide the Cloud Regime indices are "MODIS_Terra_CR" and "MODIS_Aqua_CR", representing observations from each of the two MODIS instruments in orbit. The variable that provides the mean joint histogram of each regime is "centroids".

3.0 Data Contents

The following subsections provide more information on the MODIS Collection 6.1 Equal-Angle Three-Hourly Cloud Regimes, their dimensions, and their geolocation.

This product is derived from interpolating the 110-km equal-area MODIS CRs (MODIS Collection 6.1 Equal-Area Three-Hourly Cloud Regimes products) to 1°. This equal-area MODIS CR product can be accessed at https://doi.org/10.5067/MEASURES/MODISCR/EQAR3H/DATA301.

3.1 Data Set Attributes (File Metadata)

In addition to SDS arrays containing variables and dimension scales, global metadata is also stored in the files. Some metadata are required by standard conventions, some are present to meet data provenance requirements and others as a convenience to users of the MODIS Collection 6.1 Equal-Angle Three-Hourly Cloud Regime product. A summary of global attributes present in all files is shown in Table 1.

Global Attribute	Description	Type
ShortName	Short name for data set	string
LongName	Long, descriptive name for data set	string
title	A title for the data set	string
VersionID	Data version	string
history	A record of modifications to the data	string
Format	File format of the data (NetCDF4)	string
ProcessingLevel	The processing Level (Level 3)	string
IdentifierProductDOIAuthority	Authority through which DOI can be resolved	string
IdentifierProductDOI	Digital object identifier	string

Conventions	The metadata conventions followed in the file, CF-1.7	string
source	Platforms/Instruments and any other factors related to the origin of the data product or products	string
MapProjection	Applies to gridded data.	string
DataSetQuality	Overall assessment of quality of data	string
ContactPersonName	Name of point of contact regarding the data	string
ContactPersonEmail	Email address for contact person name	string
institution	Where the data were produced	string
InputDataProducts	Input data to the product of interest	string
InputDataProductVersion	Input data version	string
GranuleID	The file name	string
references	Published or web-based information on the data	string
ProductionDateTime	Date and time the current file was produced	string
RangeBeginningDate	Start date of the data in the file	string
RangeBeginningTime	Start UTC time of the data	string
RangeEndingDate	End date of the data in the file	string
RangeEndingTime	End UTC time of the data	string
SouthernmostLatitude	Southernmost latitude of global grid of data set	string
NorthernmostLatitude	Northernmost latitude of global grid of data set	string
WesternmostLongitude	Westernmost longitude of global grid of data set	string
EasternmostLongitude	Easternmost longitude of global grid of data set	string
Description	Short, informative descriptive statement	string
comment	Short, informative descriptive statement	string

Table 1: A summary of global attributes in the files.

3.2 Variable Data Attributes

Table 2 gives the attributes used by the variables.

Variable Attribute	**Description**	**Type**
long_name	Long, descriptive name for variable	string
units	Units of the variable	string
_FillValue	Value used to indicate missing data	string
valid_range	Range of valid values of data in data field	same type as data field

coordinates	The coordinates (dimensions) of the data field	same type as data field
comments	Miscellaneous information about the variable	string
bounds	Bounds of the variable	string

Table 2: summary of the attributes used by the variables.

3.3 Geolocation Fields

Table 3 gives the geolocation fields for the data.

Geolocation variable	Description	Type
longitude	longitude of the center of the cell	32-bit floating-point
latitude	latitude of the center of the cell	32-bit floating-point
longitude_bounds	longitude bounds of the cell	32-bit floating-point
latitude_bounds	latitude bounds of the cell	32-bit floating-point
time	UTC hour and days since 2000-03-01 00:00:00Z	32-bit floating-point

Table 3: A summary of the geolocation variables used by the variables.

3.4 Dimensions

Table 4 gives the dimensions used by the variables in the dataset.

Dimensions	Description	Size
longitude	Longitude of the center of the cell	360
latitude	Latitude of the center of the cell	180
time	UTC hour and days since 2000-03-01 00:00:00Z	undefined
edge	Lower limit, upper limit for longitude/latitude bounds	2
COT	Cloud optical thickness dimension	6
CTP	Cloud top pressure dimension	7
cr	Index of Cloud Regime	12

Table 4: A summary of the dimensions associated with the variables.

3.5 Data Fields

Table 5 summarizes the data fields in this product.

Data Field Name	Long_Name/Description	Type	Dimensions	Undefined Value	Units

MODIS_Terra_CR	MODIS Terra Cloud Regime index	16-bit integer	longitude, latitude, time	-99	1
MODIS_Aqua_CR	MODIS Aqua Cloud Regime index	16-bit integer	longitude, latitude, time	-99	1
centroids	global mean of all 2003-2019 joint histograms assigned to CR	32-bit floating-point	COT, CTP, cr	-999.0	%

Table 5: Summary of the data fields in this product.

As described in Sec. 1.1, Cloud Regimes identify recurring patterns in the distribution of cloud top pressure and cloud optical thickness and categorize them into a manageable number of categories. This product identifies 11 Cloud Regimes from the *k*-means clustering process, plus a 12th Cloud Regime from cloudless joint histograms. A brief physical description of the CRs, adapted from Table 1 of Cho et al. (2021) is provided below:

CR1: Optically thick vertically developed clouds with high cloud tops concentrated in the maritime continent and other areas of the Intertropical Convergence Zone.

CR2: Spatially correlated with CR1 in the deep tropics, but with optically thinner clouds.

CR3: Vertically developed storm clouds with mostly midlatitude presence, but also representing weaker tropical convection.

CR4: High cloud outflows associated with subtropical intrusions of midlatitude fronts and weaker non-organized tropical convection.

CR5: Mostly high latitude oceanic storms.

CR6: Mostly high latitude midlevel clouds with wide presence over northern lands and areas of high elevation.

CR7: The optically thickest and highest of midlatitude and subtropical stratus.

CR8: Oceanic stratus at midlatitudes, western subtropical coasts of major continents, and Arctic, with lower cloud tops than CR7.

CR9: Subtropical and midlatitude oceanic stratocumulus of modest optical thickness.

CR10: Optically thin oceanic shallow convection with extensive geographical coverage and notable presence of coincident high clouds.

CR11: Omnipresent regime with shapeless centroid pattern indicating mixtures of high- and low-level clouds of small cloud fraction. This regime can arguably be split into subregimes.

CR12: Clear sky (cloudless) regimes. Joint histograms from this regime are not included in the clustering process.

Additional information on the physical interpretation and properties of the cloud regimes in this and associated CR products can be found in Cho et al. (2021).

4.0 Options for Reading the Data

4.1 Command Line Utilities

4.1.1 ncdump

The ncdump tool can be used as a simple browser for HDF data files, to display the dimension names and sizes; variable names, types, and shapes; attribute names and values; and optionally, the values of data for all variables or selected variables in a netCDF file. The most common use of ncdump is with the –h option, in which only the header information is displayed.

ncdump [-c|-h] [-v ...] [[-b|-f] [c|f]] [-l len] [-n name] [-d n[,n]] filename
Options/Arguments:
[-c] Coordinate variable data and header information
[-h] Header information only, no data
[-v var1[,...]] Data for variable(s) <var1>,... only data
[-f [c|f]] Full annotations for C or Fortran indices in data
[-l len] Line length maximum in data section (default 80)
[-n name] Name for netCDF (default derived from file name)
[-d n[,n]] Approximate floating-point values with less precision filename File name of input netCDF file
(https://www.unidata.ucar.edu/software/netcdf/workshops/2011/utilities/Ncdump.html)

4.1.2 HDFView

HDFView is a Java based graphical user interface created by the HDF Group which can be used to browse HDF files. The utility allows users to view all objects in an HDF file hierarchy which is represented as a tree structure. Additional information about HDFView can be found at https://support.hdfgroup.org/products/java/hdfview/ and for HDF at https://portal.hdfgroup.org/display/support

4.2 Tools/Programming

The product files can be read and queried using the NetCDF4 library and tools maintained by Unidata (http://www.unidata.ucar.edu/software/netcdf/). Support for reading NetCDF is offered in many programming languages, including Python, Matlab, IDL, C/C++ and Fortran. NetCDF4 files are legal HDF5 files with additional bookkeeping information managed by the NetCDF4 library. It is therefore possible to inspect and copy data out of the NetCDF4 files by using the HDF5 utilities and libraries maintained by the HDF Group (https://www.hdfgroup.org/products/hdf5_tools/index.html) or by using the HDF5 interface in your favorite programming language. However, the two libraries should not be considered fully interchangeable.

4.2.1 Python

The following code snippet shows how to read the variable lat, lon, and MODIS_Terra_CR from the dataset with the name "filename". Also shown are some basic information about the size of the variables arrays.

```
import netCDF4

nc_file = netCDF4.Dataset(filename ,mode='r',format='NETCDF4')

#read in the variables
lat = nc_file.variables['latitude'][:]
lon = nc_file.variables['longitude'][:]
cr_terra = nc_file.variables['MODIS_Terra_CR'][:]

# print out the minimum, maximum, and dimensions for the three variables
print("-- lat  Min/Max values", lat[:].min(), lat[:].max())
print("lat.shape:", lat.shape)
print("-- lon Min/Max values:", lon[:].min(), lon[:].max())
print("lon.shape:", lon.shape)
print("-- cr_terra Min/Max values:", cr_terra [:].min(), cr_terra [:].max())
print("cr_terra.shape:", cr_terra.shape)
```

5.0 GES DISC Data Services

If you need assistance or wish to report a problem:
Email: gsfc-dl-help-disc@mail.nasa.gov
Voice: 301-614-5224
Fax: 301-614-5268
Address:
Goddard Earth Sciences Data and Information Services Center NASA Goddard Space Flight Center Code 610.2 Greenbelt, MD 20771 USA

5.1 How To Articles

The GESDISC web site contains many informative articles under the "How To Section", "FAQ" (frequently asked questions), "News", "Glossary", and "Help" . A sample of these articles includes:

Earthdata Login for Data Access

How to Download Data Files from HTTPS Service with wget

How to Obtain Data in NetCDF Format via OPeNDAP

Quick View Data with Panoply

How to Read Data in NetCDF Format with R

How to Read Data in HDF-5 or netCDF Format with GrADS

6.0 More Information

For more information on the MODIS sensors, see https://modis.gsfc.nasa.gov/.

For more information on the MODIS Atmosphere retrievals, including cloud retrievals and the changes in Collection 6.1, see https://atmosphere-imager.gsfc.nasa.gov/products.

For information on the ISCCP counterpart to this MODIS Collection 6.1 Equal-Area Three-Hourly Cloud Regimes, see https://isccp.giss.nasa.gov/wstates/hggws.html.

7.0 Acknowledgments

The product is part of an effort to "unify" the MODIS Cloud Regimes with the ISCCP Weather States and is funded by the NASA MEaSUREs program.

8.0 References

Cho, N., Tan, J., & Oreopoulos, L. (2021). Classifying planetary cloudiness with an updated set of MODIS Cloud Regimes, Journal of Applied Meteorology and Climatology, https://doi.org/10.1175/JAMC-D-20-0247.1.

Cho, N., J. Tan, & L. Oreopoulos (2021), MODIS Collection 6.1 Equal-Angle Three-Hourly Cloud Regimes, Version 1.0, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/MEASURES/MODISCR/EQANG3H/DATA301.

Jakob, C., & G. Tselioudis (2003), Objective identification of cloud regimes in the Tropical Western Pacific. Geophysical Research Letters, 30(21), 2082. https://doi.org/10.1029/2003GL018367.

Platnick, S., S. Ackerman, M. King, G. Wind, K. Meyer, P. Menzel, R. Frey, R. Holz, B. Baum, & P. Yang (2015), MODIS atmosphere L2 cloud product (06_L2), NASA MODIS Adaptive Processing System, Goddard Space Flight Center, http://doi.org/10.5067/MODIS/MOD06_L2.006, http://doi.org/10.5067/MODIS/MYD06_L2.006.

Platnick, S., M. King, & P. Hubanks (2017), MODIS Atmosphere L3 Daily Product, NASA MODIS Adaptive Processing System, Goddard Space Flight Center, http://doi.org/10.5067/MODIS/MOD08_D3.061, http://doi.org/10.5067/MODIS/MYD08_D3.061.

Rossow, W. B., G. Tselioudis, A. Polak, & C. Jakob (2005), Tropical climate described as a distribution of weather states indicated by distinct mesoscale cloud property mixtures. Geophysical Research Letters, 32(21), L21812. https://doi.org/10.1029/2005GL024584.

Young, A. H., K. R. Knapp, A. Inamdar, W. Hankins, & W. B. Rossow (2018), The International Satellite Cloud Climatology Project H-Series climate data record product, Earth Syst. Sci. Data, 10, 583–593, doi:10.5194/essd-10-583-2018.