Overview

These directories contain observational data sets intended for the evaluation of clouds produced by large-scale models. We have chosen a subset of the standard MODIS Level-3 Monthly products, reformatted the data as netCDF 3.4 files, and combined observations from Terra and Aqua to produce a single file for each month in which both platforms are available.

The files are derived from MODIS Collection 5 data products and follow the CF conventions. Each data set has a local attribute indicating the name of the data set in the original HDF files. We also expand on the MODIS HDF naming convention: files beginning MOD are from the Terra platform, files beginning MYD from the Aqua platform, and files beginning MCD are averages of the Terra and Aqua, constructed (as described below) is the same way that MODIS data processing averages over each month.

These observations may be most directly compared to large-scale model fields by using the "MODIS simulator" available as part of the CFMIP Observation Simulator Package (COSP) version 1.2 or later. In most applications comparisons would be made to the the combined (MCD) files.

We are currently working on a manuscript describing, among other things, systematic differences between these observations and ISCCP, the use of these observations in the evaluation of large-scale models, and the implementation of the ISCCP simulator.

For further information please contact Robert Pincus (Robert.Pincus@noaa.gov).

Files for individual platforms (Terra, Aqua)

To produce the monthly-average files for each platform we reformat a subset of the available fields from HDF-EOS to netCDF. Along the way we rename most of the data sets; the original names appear in the HDF_variable_name local attribute for each variable. We also remove most metadata (much of which is used to be able to control the production of the HDF files) while adding metadata so that the files comply with the CF conventions.

We convert the joint histogram of cloud top pressure p_c and optical thickness τ from integer counts C to the fractional cloudiness c in each category, i.e.

$$c(p_c, \tau) = c_r \times C(p_c, \tau) / \sum_{p_c, \tau} C(p_c, \tau) \tag{1}$$

where the mapping between variables and dataset names is

Variable	dataset
$C(p_c, \tau)$	Optical_Thickness_vs_Cloud_Top_Pressure (in)
c_r	Cloud_Fraction_Retrieval_Total_Mean
$c(p_c, \tau)$	Optical_Thickness_vs_Cloud_Top_Pressure (out)

We produce two estimates of high, middle, and low cloud fraction, defined as the fraction of total pixels with $p_c \leq 440\ hPa$, $440 < p_c \leq 680\ hPa$, and $p_c > 680\ hPa$, respectively. The first estimate corresponds to the "retrieval cloud fraction" in that it represents the number of pixels for which cloud optical properties are retrieved, and it is obtained by summing over the joint histogram:

$$
\begin{aligned}
c_{h,r} = &\quad \sum c(p_c \leq 440, \tau) \quad (2)\\
c_{l,r} = &\quad \sum c(p_c > 680, \tau) \quad (3)\\
c_{m,r} = &\quad c_r - c_h - c_l \quad (4)
\end{aligned}
$$

The second set of estimates of high, middle, and low cloud fraction correspond to the "mask cloud fraction," i.e. those pixels judged likely to be filled with cloud, approximated as the fraction of pixels for which cloud top pressure has been retrieved. (This is a superset of the pixels for which cloud optical properties are available.) The distribution of cloud top pressures is available as a histogram $C_m(p)$ in the MODIS Level 3 data. Unfortunately, the boundaries of this histogram do not correspond to the thresholds used in the joint histogram of p_c and τ, but are rather spaced evenly in units of 100 hPa. We assume that cloud are uniformly distributed within each bin of cloud top pressure and estimate the histogram counts for high, middle, and low cloud as

$$
\begin{aligned}
C_{h,m} &= \sum C_m(p_c \leq 440) \approx \sum C_m(p_c \leq 400) + 0.4C_m(400 < p_c \leq 500) \quad (5)\\
C_{l,m} &= \sum C_m(p_c > 680) \approx \sum C_m(p_c > 700) + 0.2C_m(600 < p_c \leq 700) \quad (6)\\
C_{m,m} &= \sum C_m(440 < p_c \leq 680) \quad (7)\\
&\approx C_m(500 < p_c \leq 600) + 0.6C_m(400 < p_c \leq 500) + 0.8C_m(600 < p_c \leq 700) \quad (8)
\end{aligned}
$$

Cloud fraction variables $c_{i,m}$ are computed assuming that cloud top pressure has been determined for every pixel identified as cloudy by the cloud mask:

$$
c_{i,m}(p_c) = c_m \times C_{i,m}(p_c) / \sum_{p_c,\tau} C_m(p_c) \quad (9)
$$

In practice, these two steps are combined. Variables names for these derived data sets are as follows

Variable	dataset
$c_{h,r}$	Cloud_Fraction_Retrieval_High_Mean
$c_{m,r}$	Cloud_Fraction_Retrieval_Mid_Mean
$c_{l,r}$	Cloud_Fraction_Retrieval_Low_Mean
c_m	Cloud_Fraction_Mask_Total_Mean
C_m	Cloud_Top_Pressure_Day_Histogram_Counts
$c_{h,m}$	Cloud_Fraction_Mask_High_Mean
$c_{m,m}$	Cloud_Fraction_Mask_Mid_Mean
$c_{l,m}$	Cloud_Fraction_Mask_Low_Mean

Combining Aqua and Terra files

We compute a single monthly-mean file by averaging together observations from Terra and Aqua consistent with way the MODIS data is aggregated to Level 3 at other steps. For variables χ related to cloud fraction and cloud-top pressure the combined dataset is a simple linear average of the Terra and Aqua values for the month. The following variables use linear averaging:

χ
Cloud_Fraction_Mask_Total_Mean
Cloud_Fraction_Mask_High_Mean
Cloud_Fraction_Mask_Mid_Mean
Cloud_Fraction_Mask_Low_Mean
Cloud_Fraction_Retrieval_Total_Mean
Optical_Thickness_vs_Cloud_Top_Pressure
Cloud_Fraction_Retrieval_High_Mean
Cloud_Fraction_Retrieval_Mid_Mean
Cloud_Fraction_Retrieval_Low_Mean
Cloud_Fraction_Retrieval_Liquid_Mean
Cloud_Fraction_Retrieval_Ice_Mean
Cloud_Top_Pressure_Total_Mean

Variables related to cloud optical properties are averaged on a per-pixel basis, so that days with a greater number of cloudier observations count more heavily towards the monthly mean. Since the number of individual observations (i.e. pixels) may differ between platforms, we compute the average value for these parameters as

$$\chi = (\chi_T \times N_T + \chi_A \times N_A)/(N_T + N_A) \tag{10}$$

where the subscripts T and A refer to the Terra and Aqua values and N is the number of observations:

χ	N
Cloud_Optical_Thickness_Total_Mean	Cloud_Retrieval_Total_Pixel_Counts
Cloud_Optical_Thickness_Total_MeanLog10	
Cloud_Optical_Thickness_Liquid_Mean	Cloud_Retrieval_Liquid_Pixel_Counts
Cloud_Optical_Thickness_Liquid_Uncertainty_in_Mean	
Cloud_Optical_Thickness_Liquid_MeanLog10	
Cloud_Optical_Thickness_Liquid_Uncertainty_in_MeanLog10	
Cloud_Particle_Size_Liquid_Mean	
Cloud_Particle_Size_Liquid_Uncertainty_in_Mean	
Liquid_Path_Mean	
Liquid_Path_Uncertainty_in_Mean	
Cloud_Optical_Thickness_Ice_Mean	Cloud_Retrieval_Ice_Pixel_Counts
Cloud_Optical_Thickness_Ice_Uncertainty_in_Mean	
Cloud_Optical_Thickness_Ice_MeanLog10	
Cloud_Optical_Thickness_Ice_Uncertainty_in_MeanLog10	
Cloud_Particle_Size_Ice_Mean	
Cloud_Particle_Size_Ice_Uncertainty_in_Mean	
Ice_Path_Mean	
Ice_Path_Uncertainty_in_Mean	

Pixel_Counts variables and the histogram of cloud top pressure are omitted from the combined Terra/Aqua files, and a global attribute "Conventions" is added with the text value "CF-1.4".

Further information

For further information please see

netCDF	`http://www.unidata.ucar.edu/software/netcdf/`
CF conventions	`http://cf-pcmdi.llnl.gov/`
COSP	`http://cfmip.metoffice.com/COSP.html`