

Using MATLAB to Read HDF5 Files for GPM data files

The NASA Goddard Precipitation Processing System, <http://pps.gsfc.nasa.gov>

Written by Rachael Kroodsma (rachael.a.kroodsma@nasa.gov)

Updated: 27 July 2021

This document provides examples for reading HDF5 files from the GPM mission into MATLAB. IMERG and Level 1C GMI files are used as examples.

There are three examples given for how to read in data.

- 1.) Read one variable at a time. Requires knowing the variable name and data structure. PPS THOR tool is very useful for looking at the hierarchy of the data file to see the variables contained in the files.
- 2.) Read several variables. Requires knowing the variable name and data structure but condenses code by not having several lines of read functions.
- 3.) Read several variables with no knowledge of data structure or variable name. This example can be applied to any arbitrary HDF5 file. However, it is not always recommended since it will read in all variables in the file which can be slow and may not be necessary for your application.

Example MATLAB code below:

```
%% Read GPM HDF5 files into MATLAB
% Written by Rachael Kroodsma, NASA GSFC, PPS
% Updated July 27, 2021

clear

fileName1 = '3B-HHR.MS.MRG.3IMERG.20210101-S000000-E002959.0000.V06B.HDF5';
fileName2 = '1C.GPM.GMI.XCAL2016-C.20210101-S013133-E030407.038882.V05A.HDF5';

% get information about file
h5infoFile1 = h5info(fileName1);
h5infoFile2 = h5info(fileName2);
% display all metadata (if desired)
h5disp(fileName1);
h5disp(fileName2);

%% 1.) Read one variable at a time

% IMERG
Lat = h5read(fileName1, '/Grid/lat');
% 1C reading variable that's in a subgroup
```

```

SCLat = h5read(fileName2,'S1/SCstatus/SCLatitude');

%% 2.) Read several variables with known names

% HDF5 variable names (including group and subgroup names)
InVar = {'/Grid/lat','/Grid/lon','/Grid/precipitationCal'};
% MATLAB variable names corresponding to HDF5 variables
Var = {'Lat','Lon','precip'};

for i = 1:length(Var)
    eval([Var{i}, '= double(h5read(["',fileName1,'"','"',InVar{i},'"']);]);
end

%% 3.) Read several variables without knowing the variable names
% This can be applied to any arbitrary HDF5 file without knowing any of the
% group or dataset names. Code will read in all variables in the file and
% name the variable in this format: 'Group_Subgroup_DatasetName'

% choose to read IMERG or 1C file
% h5infoFile = h5infoFile1;
% fileName = fileName1;
h5infoFile = h5infoFile2;
fileName = fileName2;

% get number of groups
groupLength = length(h5infoFile.Groups);

for g = 1:groupLength
    groupName = h5infoFile.Groups(g).Name;

    % read in any subgroups
    for g2 = 1:length(h5infoFile.Groups(g).Groups)
        subgroupName = h5infoFile.Groups(g).Groups(g2).Name;
        for i = 1:length(h5infoFile.Groups(g).Groups(g2).Datasets)
            datasetName = h5infoFile.Groups(g).Groups(g2).Datasets(i).Name;
            pathName = [subgroupName '/' datasetName];
            eval([strjoin(strsplit(pathName(2:end),'/'),'_'),' = double(h5read(["',fileName,'"','"',pathName,'"']);]);
        end
    end

    % read in the Groups
    for i = 1:length(h5infoFile.Groups(g).Datasets)
        datasetName = h5infoFile.Groups(g).Datasets(i).Name;
        pathName = [groupName '/' datasetName];
        eval([strjoin(strsplit(pathName(2:end),'/'),'_'),' = double(h5read(["',fileName,'"','"',pathName,'"']);]);
    end
end

```