

Utilizing Python for the Conversion of GPM HDF5 File to Text File

What is GPM?

The Global Precipitation Measurement Mission (GPM) is an international satellite mission that provides next generation observations of rain and snow worldwide. GPM meteorological data is freely available through the website <https://pmm.nasa.gov/data-access/downloads/gpm> and can be downloaded in HDF5 format. Also see webinars 2 and 3 on <https://pmm.nasa.gov/training> for instructions to download GPM data files.

What is HDF5?

HDF5 is a data model, library, and file format for storing and managing data. It supports an unlimited variety of datatypes and is designed for high volume and complex data. Complex data within HDF5 files can be extracted, manipulated, and converted to various other data formats through the use of Python.

What is Python?

Python is a high-level programming language that is freely available and easily implemented across platforms. Its syntax design emphasizes code readability and allows users to express concepts in fewer lines of code than possible in other available languages.

Download Python

Installing Python from scratch is not a simple task. Many scientific packages require specific versions of Python dependencies. It is hard to keep packages from interacting with each other and to keep them updated. Anaconda makes getting and maintaining these packages quick and easy. Anaconda installs the Python language and 720+ open source packages that are necessary for scientific and numeric analysis. If you do not have Python downloaded already, versions 2.7 and 3.5 of Python can be downloaded for Windows, OSX, and Linux through the following Anaconda link <https://www.continuum.io/downloads>. For this exercise, download Python 3.5 for your specific OS system (if you do not have python already). For Linux this method is strongly recommended for its ease of installation. An installation GUI will guide you through the installation instructions. No installation parameters in the GUI need to be changed.

Downloading the Necessary Packages

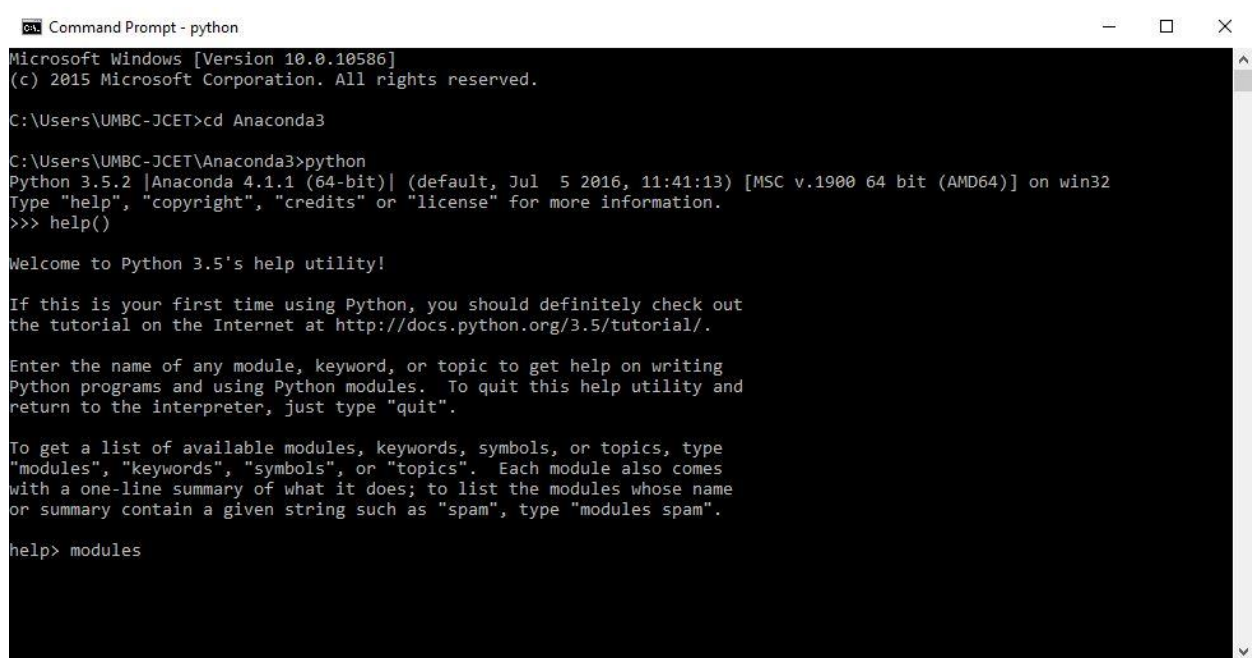
Once Python is installed, it is necessary to make sure that the correct packages are installed for the hdf2txt Python script used in this exercise to function correctly. To run the python scripts you will need the following Python packages to run the code:

1. pip (sometimes installed with python installation)
2. os (automatically installed with python installation)
3. glob (automatically installed with python installation)
4. numpy (sometimes installed with python installation)
5. h5py (sometimes installed with python installation)

If you have downloaded Anaconda's 3.5 version of Python as instructed earlier, all packages should be installed. If you previously had python installed on your computer and are using this version of python,

first check to see if you have the packages above by opening python through the command prompt (figure 1). Python can be opened through the command prompt by using the “cd” command to change to the directory where the python.exe file is stored and then enter “python” at the command line.

Once Python is open, type “help()” to open the help utility. From there type “modules”. This will bring up a list of all available modules. Check the list to see if pip, os, glob, numpy, and h5py are available. Os and glob are standard packages that come with any Python download. If you are missing pip, numpy, or h5py, reference the sections below to install the packages.

A screenshot of a Windows Command Prompt window titled "Command Prompt - python". The window shows the following text:

```
Microsoft Windows [Version 10.0.10586]
(c) 2015 Microsoft Corporation. All rights reserved.

C:\Users\UMBC-JCET>cd Anaconda3

C:\Users\UMBC-JCET\Anaconda3>python
Python 3.5.2 [Anaconda 4.1.1 (64-bit)] (default, Jul  5 2016, 11:41:13) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> help()

Welcome to Python 3.5's help utility!

If this is your first time using Python, you should definitely check out
the tutorial on the Internet at http://docs.python.org/3.5/tutorial/.

Enter the name of any module, keyword, or topic to get help on writing
Python programs and using Python modules.  To quit this help utility and
return to the interpreter, just type "quit".

To get a list of available modules, keywords, symbols, or topics, type
"modules", "keywords", "symbols", or "topics".  Each module also comes
with a one-line summary of what it does; to list the modules whose name
or summary contain a given string such as "spam", type "modules spam".

help> modules
```

Figure 1: Example that demonstrates how to open Python at the command line, and open the help utility and find the list of currently installed packages.

Install and Download Pip

Pip installation instructions are available at the following website <https://packaging.python.org/installing/>. If you have Python 2 >=2.7.9 or Python 3 >=3.4 installed from python.org, you will already have pip, but will need to upgrade to the latest version.

Install Download Numpy or H5py

Numpy, h5py, and other Python packages are available in wheel files at the following website <http://www.lfd.uci.edu/~gohlke/pythonlibs/>. Python wheel files are ideal for installing python packages because they include all of the necessary dependencies for a package to work. For each package available on this website there are six available wheel files. There are two files for each version of python (2.7, 3.4, and 3.5). The first version is for a 32 bit operating system and the second version is for a 64 bit operating system. Select the version that matches with your version of python and operating system. For example, if you have python 3.4 downloaded and a 64 bit OS you would download the wheel in yellow.

- h5py-2.6.0-cp27-cp27m-win32.whl

- h5py-2.6.0-cp27-cp27m-win_amd64.whl
- h5py-2.6.0-cp34-cp34m-win32.whl
- **h5py-2.6.0-cp34-cp34m-win_amd64.whl**
- h5py-2.6.0-cp35-cp35m-win32.whl
- h5py-2.6.0-cp35-cp35m-win_amd64.whl

By double clicking on the wheel you want to use, it will automatically download to your local downloads folder. At the command line, then navigate to your downloads folder and type “pip install your_wheel_file” at the command line (figure 2). This will download the package and all of its dependencies.

```

Command Prompt
Microsoft Windows [Version 10.0.10586]
(c) 2015 Microsoft Corporation. All rights reserved.

C:\Users\UMBC-JCET>cd Downloads
C:\Users\UMBC-JCET\Downloads>pip install h5py-2.6.0-cp34-cp34m-win_amd64.whl

```

Figure 2: Example of downloading a package

=====

Running Python Scripts

For IMERG Data (Level 3 data)

hdf2txt_imerg.py

The hdf2txt_imerg.py script is a Python script that will open a, IMERG HDF5 file and convert it to a text file. Before running this script you need to place all of your HDF5 data files in a single directory. This code can be run by opening the command prompt and “cd” to the directory where all of the HDF5 files are stored. Once in this directory, enter:

For windows machine: full_directory_path_for_the_python_executable hdf2txt_imerg.py

For Linux: python hdf2text_imerge.py

The code will automatically convert datasets in HDF5 files to text files and places them in a folder called “text_files”. Currently the code outputs the latitude, longitude, and surface precipitation. The for loop in the code can be altered to output more datasets.

```

Command Prompt
Microsoft Windows [Version 10.0.10586]
(c) 2015 Microsoft Corporation. All rights reserved.

C:\Users\UMBC-JCET>cd Desktop\testinghdf
C:\Users\UMBC-JCET\Desktop\testinghdf>C:\Users\UMBC-JCET\Anaconda3\python hdf2text.py

```

Running hdf2txt_clip_imerge.py

Sometimes analyzing all of the data in a dataset can be intensive and unnecessary. If you would like to clip the dataset utilize the `hdf2txt_clip_imerg.py` script. Before running this script you need to place all of your HDF5 files in a single directory. This code can be run by opening the command prompt and “`cd`” to the directory where all of the HDF5 files are stored. Once in this directory, use the same commands used to run `hdf2text_imerg.py` (as shown above). The code will begin by asking the user to input their bounding coordinates in terms of north and south latitudes, east and west longitudes. It will then clip the datasets, convert datasets in HDF5 files to text files, and places them in a folder called “`text_files`”. Currently the code outputs the latitude, longitude, and surface precipitation. Also, a file is created in which clipped data are written for each grid point in the file as ‘latitude, longitude, and precipitation’ Table. The for loop in the code can be altered to output more datasets.

For GMI Data (Level 2 data)

hdf2txt_gmi.py

The `hdf2txt_gmi.py` script is a Python script that will open a, GMI swath data in HDF5 format and convert it to a text file. Before running this script you need to place all of your HDF5 data files in a single directory. This code can be run by opening the command prompt and “`cd`” to the directory where all of the HDF5 files are stored. Once in this directory, enter:

For windows machine: `full_directory_path_for_the_python_executable hdf2txt_gmi.py`

For Linux: `python hdf2text_gmi.py`

The code will automatically convert datasets in HDF5 files to text files and places them in a folder called “`text_files`”. Currently the code outputs the latitude, longitude, and surface precipitation. The for loop in the code can be altered to output more datasets.