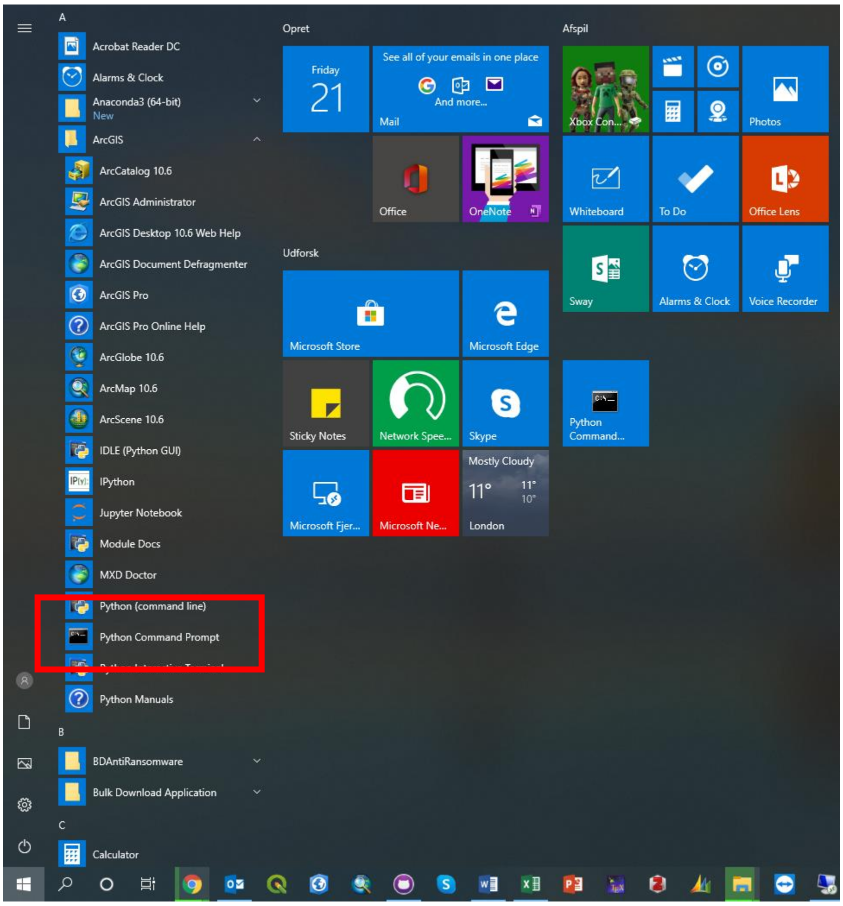
Setting up Python for batch downloading Landsat scenes

By Penny How – how@asiaq.gl

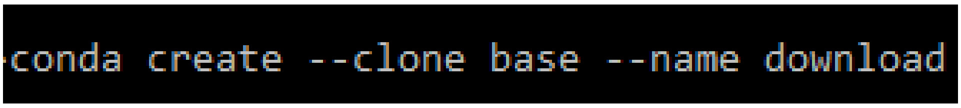
This set of instructions is for downloading Landsat scenes from the EarthExplorer repository.

1. Setting up Python to facilitate downloading

We first need to create Conda environment to hold the packages we need to batch download Landsat scenes. To do this, find Python Command Prompt which will be on the Start menu either under Anaconda3:

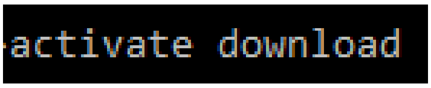


This will open a black box console where you can type text. The text printed before the cursor signifies the Conda environment that is active (in the brackets) followed by the active file directory. We firstly want to create a new Conda environment, by copying the base environment holding all the key Python packages. To do this, enter the following command:



```
conda create --clone base --name download
```

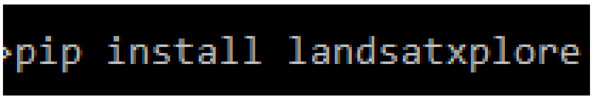
By entering this command we are stating to create a new Conda environment, which is a copy of the base environment, and we are going to call it 'download'. Once entered, you will see many statements showing the environment being constructed. Once finished, you will see your cursor prompt return. To activate this environment, type:



```
activate download
```

You will now see the activated Conda environment change from the base environment to your newly made environment.

Now we have to install the package needed for our batch downloading to work. This package is called landsatxplore, which was developed to search, download and retrieve data from EarthExplorer Landsatxplore is available to download through pip, an easy installation method which requires just one command:



```
>pip install landsatxplore
```

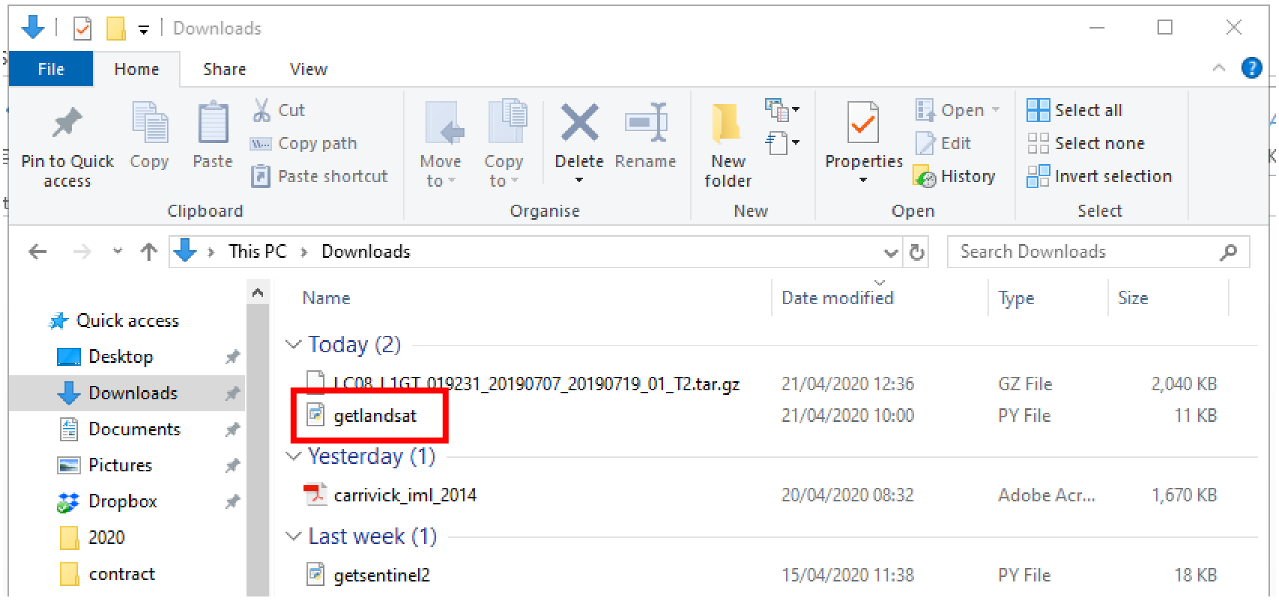
We now have a Conda environment called 'download' and has the landsatxplore package installed in it.

2. Getting the Python script and setting it up for downloading

Now we have our activated Conda environment, we need to get our Python script for batch downloading and choose a folder where our initial downloads will be inputted. First, make a copy of the Python script from the P: drive directory:

P:\B35_Remote_Sensing\B35-01_Projektstyring-og-Markedsføring\B35-01-004_Strategi-og-Metode\Python tutorials\Batch downloading with Python\landsat\getlandsat.py

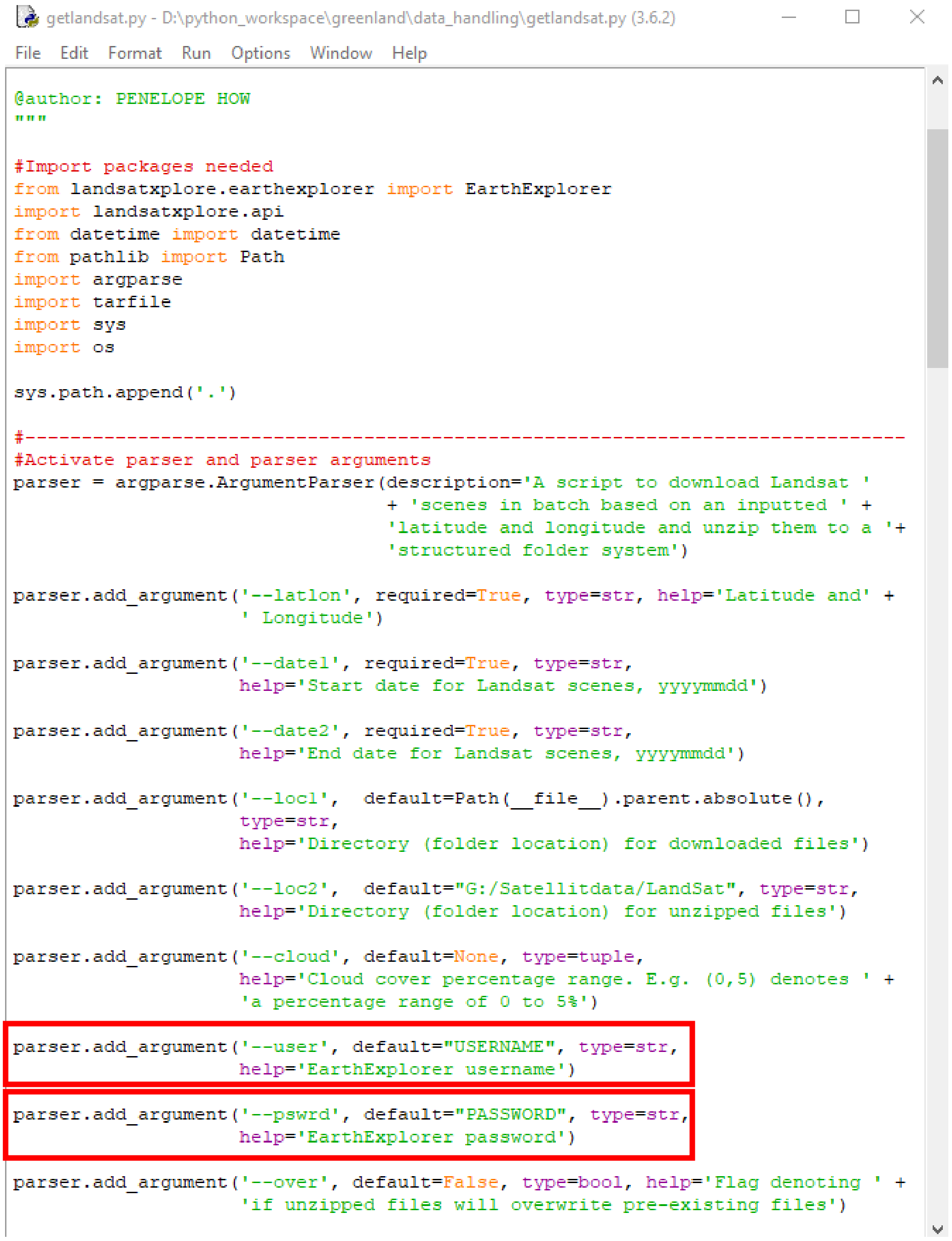
And place this file in a folder where you want your Landsat imagery to be initially downloaded to. A suitable location would be the Downloads folder, for example.



Name	Date modified	Type	Size
LC08_L1GT_019231_20190707_20190719_01_T2.tar.gz	21/04/2020 12:36	GZ File	2,040 KB
getlandsat	21/04/2020 10:00	PY File	11 KB
carrivick_iml_2014	20/04/2020 08:32	Adobe Acr...	1,670 KB
getsentinel2	15/04/2020 11:38	PY File	18 KB

To download Landsat images from the EarthExplorer online repository, we need to have login details (a username and password) to access EarthExplorer. If you do not have an account with EarthExplorer, you will need to make one at the web address https://ers.cr.usgs.gov/login/.

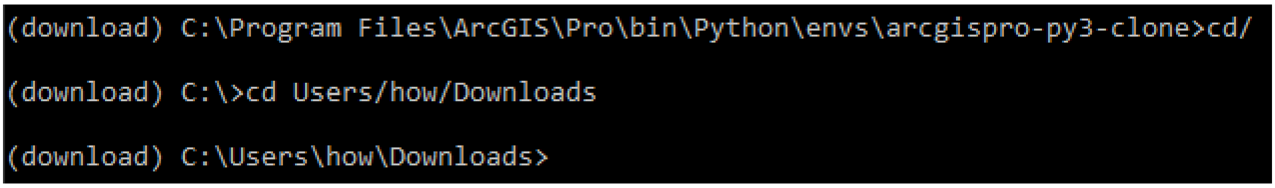
Your EarthExplorer login details need to be inputted into the getlandsat.py file, so that Landsat imagery can be retrieved from the repository through your EarthExplorerer account. Open the getlandsat.py file by double-clicking on it. This should open a Python console where you can scroll through the code and edit it. Scroll to the lines shown below; these should be near to the top of the script.


getlandsat.py - D:\python_workspace\greenland\data_handling\getlandsat.py (3.6.2)
File Edit Format Run Options Window Help

```
@author: PENELOPE HOW
"""

#Import packages needed
from landsatxplore.earthexplorer import EarthExplorer
import landsatxplore.api
from datetime import datetime
from pathlib import Path
import argparse
import tarfile
import sys
import os

sys.path.append('.')

#-------------------------------------------------------------------------------
#Activate parser and parser arguments
parser = argparse.ArgumentParser(description='A script to download Landsat '
                                 + 'scenes in batch based on an inputted ' +
                                 'latitude and longitude and unzip them to a '+
                                 'structured folder system')

parser.add_argument('--latlon', required=True, type=str, help='Latitude and' +
                    ' Longitude')

parser.add_argument('--date1', required=True, type=str,
                    help='Start date for Landsat scenes, yyyymmdd')

parser.add_argument('--date2', required=True, type=str,
                    help='End date for Landsat scenes, yyyymmdd')

parser.add_argument('--loc1',  default=Path(__file__).parent.absolute(),
                    type=str,
                    help='Directory (folder location) for downloaded files')

parser.add_argument('--loc2',  default="G:/Satellitdata/LandSat", type=str,
                    help='Directory (folder location) for unzipped files')

parser.add_argument('--cloud', default=None, type=tuple,
                    help='Cloud cover percentage range. E.g. (0,5) denotes ' +
                    'a percentage range of 0 to 5%')

parser.add_argument('--user', default="USERNAME", type=str,
                    help='EarthExplorer username')

parser.add_argument('--pswrd', default="PASSWORD", type=str,
                    help='EarthExplorer password')

parser.add_argument('--over', default=False, type=bool, help='Flag denoting ' +
                    'if unzipped files will overwrite pre-existing files')
```

Currently the EarthExplorer username and password are set to ‘USERNAME’ and ‘PASSWORD’. These are invalid log in details that we need to edit. The two commands highlighted are where the terms ‘USERNAME’ and ‘PASSWORD’ need to be replaced with our own EarthExplorer username and password in order to gain permission to download. In each command, replace the word ‘USERNAME’ with your username (for the first command) and the word ‘PASSWORD’ with your password (for the second command). Remember that there should be quotation marks around your username and password, and the words should therefore be highlighted in green (as above). After you are modified the script, save it. For each time we run this script, your user details will always be used to log into EarthExplorer now.

Close the Python script and look at your command line again. We now have to navigate to the directory our Python script is in using command line, which is where we will run this Python script. Use the command *‘cd’* to navigate to a file directory in command line, inputting the folder name after to navigate into a given folder.



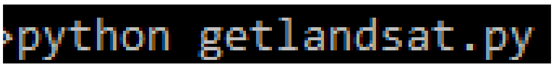
```
(download) C:\Program Files\ArcGIS\Pro\bin\Python\envs\arcgispro-py3-clone>cd/

(download) C:\>cd Users/how/Downloads

(download) C:\Users\how\Downloads>
```

In this example, we first type the command *'cd/'* to navigate back up the C: directory, then *'cd Users/how/Downloads'* to navigate to the Downloads folder in my user space (user 'how'). If you are unsure of the contents of the directory you are in, type '*dir'* to print the contents of the directory.

We are now ready to run the Python script and start downloading our images.

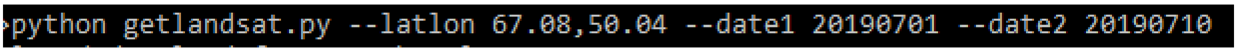
3. Running the Python script

To run the script, we type the command *'python getlandsat.py'* into the command line, signifying to open a Python window and run the getlandsat.py script through it.



```
>python getlandsat.py
```

There are a couple of variables we need to define in order to download and retrieve our desired satellite imagery. There are three mandatory variables that are needed in order to make the script run – the area of interest (AOI) you wish to download scenes from (as either a list of tile ids, or a geojson file containing a polygon of your AOI), the start date for the download, and the end date for the download – which will look something like this:

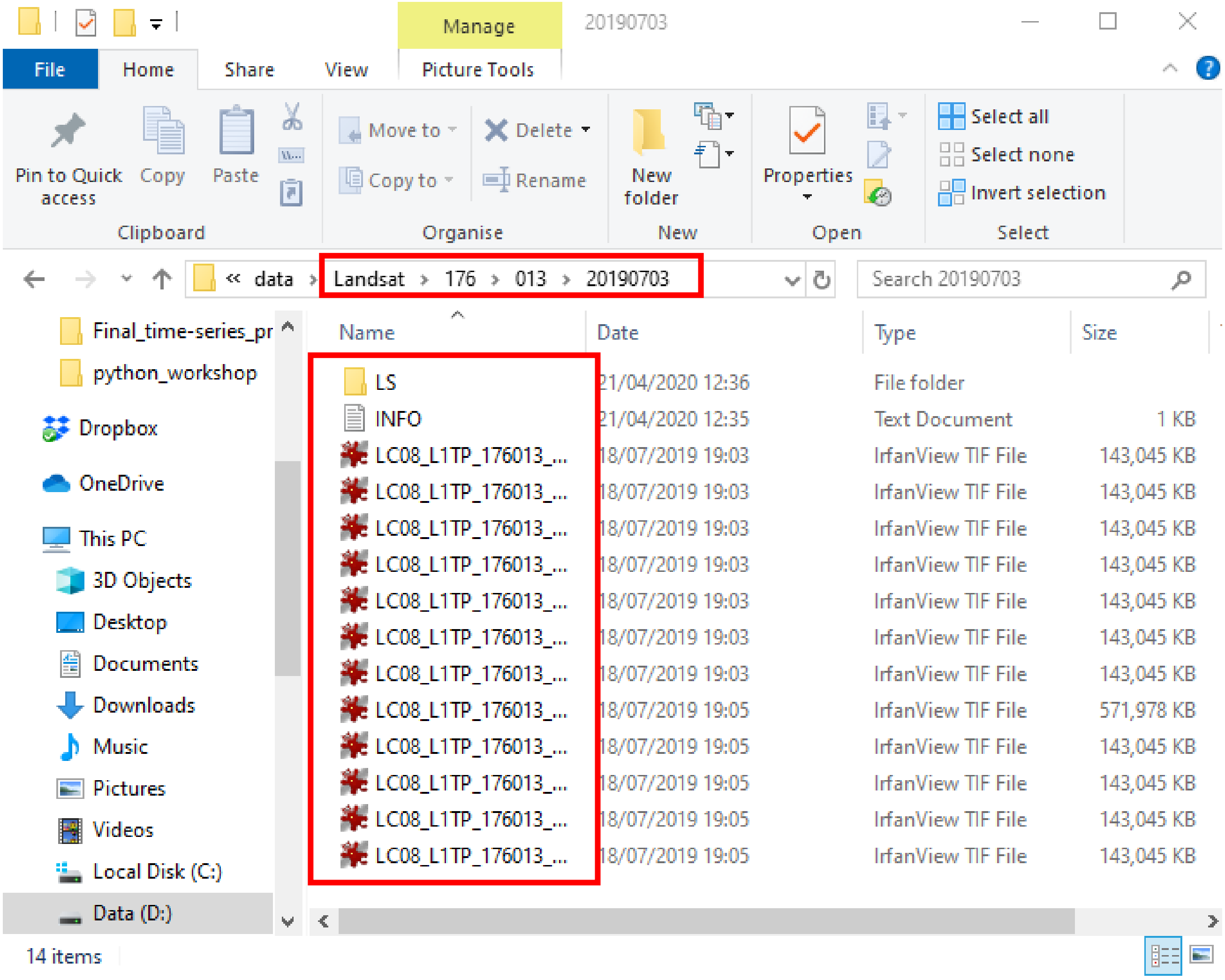


```
>python getlandsat.py --latlon 67.08,50.04 --date1 20190701 --date2 20190710
```

In this example, we are looking for scenes that correspond with the position 67°08' (latitutde) and 50°04' (longitude) (which is the Jakobshavn Isbræ catchment) between the dates 01/07/2019 and 10/07/2019. When you hit enter, a number of steps happen which are printed in the command line as they occur:

A. Your log in details are used to sign into EarthExplorer (which is where the Landsat scenes are held online). A print statement will signify if your login details are successful
B. A list of scenes satisfying your inputted parameters will be retrieved
C. Each product will be checked to see if it is valid product type. This mainly applied to Landsat 8 imagery, ensuring that only OLI-TIRS combined products are downloaded and not OLI-only or TIRS-only products
D. Each product will then be checked to see if it has already been downloaded. If a product already exists offline then it will move to the next product. If a product has not been downloaded previously then a request is sent to the EarthExplorer repository to download the product
E. One downloaded, each product is unzipped and moved into a common folder directory structured as *Landsat >> path >> row >> acquisition date.*

In each downloaded scene folder (i.e. after *'acquisition date'*), the band .tif files are in the top level and the remaining files (e.g. auxillary files, cloud mask files) can be found in the *'LS'* folder. Additionally, a readme file is generated with each downloaded scene called *'INFO.txt'*, which contains the metadata for the scene, the date the scene was downloaded, and who downloaded it (i.e. your Asiaq username).



« data › Landsat › 176 › 013 › 20190703

Name	Date	Type	Size
LS	21/04/2020 12:36	File folder	
INFO	21/04/2020 12:35	Text Document	1 KB
LC08_L1TP_176013_...	18/07/2019 19:03	IrfanView TIF File	143,045 KB
LC08_L1TP_176013_...	18/07/2019 19:03	IrfanView TIF File	143,045 KB
LC08_L1TP_176013_...	18/07/2019 19:03	IrfanView TIF File	143,045 KB
LC08_L1TP_176013_...	18/07/2019 19:03	IrfanView TIF File	143,045 KB
LC08_L1TP_176013_...	18/07/2019 19:03	IrfanView TIF File	143,045 KB
LC08_L1TP_176013_...	18/07/2019 19:03	IrfanView TIF File	143,045 KB
LC08_L1TP_176013_...	18/07/2019 19:03	IrfanView TIF File	143,045 KB
LC08_L1TP_176013_...	18/07/2019 19:05	IrfanView TIF File	571,978 KB
LC08_L1TP_176013_...	18/07/2019 19:05	IrfanView TIF File	143,045 KB
LC08_L1TP_176013_...	18/07/2019 19:05	IrfanView TIF File	143,045 KB
LC08_L1TP_176013_...	18/07/2019 19:05	IrfanView TIF File	143,045 KB
LC08_L1TP_176013_...	18/07/2019 19:05	IrfanView TIF File	143,045 KB

Above is a downloaded and unzipped scene from our example download in this document – this is a scene from 67°08', 50°04' on 03/07/2019. Note the folder structure and the contents of the directory – each band .tif file, the LS folder containing all other scene data, and the INFO.txt file containing the metadata file.

4. Adapting the input variables to refine your scene downloads

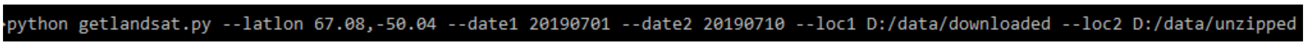
Above is an example of where we download Landsat scenes using the three mandatory variables – the latitude/longtitude, start date and end date. There are other optional variables that can be inputted to refine the scene downloads and modify the output locations, which are summarised in the table below.

Variable	Name	Notes	Mandatory/ optional variable?	Default input
--latlon	Area of interest	Latitude, longitude position denoting the area of interest	Mandatory	N/A
--date1	Start date	Start date for scene download	Mandatory	N/A
--date2	End date	End date for scene download	Mandatory	N/A

--loc1	Download folder location	Folder location to download zip files to	Optional	Directory that the Python script is in*
--loc2	Unzip folder location	Folder location to unzip files to in the set folder structure	Optional	'G:/Satellitdata/LandSat'
--cloud	Cloud cover	Cloud cover percentage range; for example, '(0,5)' denotes a percentage range of 0% to 5%	Optional	None
--user	EarthExplorer username	Username for logging in to EarthExplorer	Optional	'guest'
--pswrd	EarthExplorer password	Password for logging in to EarthExplorer	Optional	'guest'
--over	Overwrite flag	Flag denoting whether files that exist should be overwritten with newly downloaded version	Optional	False
--prodctrl	Product control flag	Flag denoting whether product control is implemented (i.e. L8 OLI/TIRS combined products will be downloaded only)	Optional	True

* pathlib.Path(__file__).parent.absolute()

For example, if we want to redefine the downloaded folder location and the unzipped folder location for the Landsat downloads, we can run the script with the *'--loc1'* and *'--loc2'* variables, in addition to our three mandatory variables:



```
python getlandsat.py --latlon 67.08,-50.04 --date1 20190701 --date2 20190710 --loc1 D:/data/downloaded --loc2 D:/data/unzipped
```

In this instance, the zip files from EarthExplorer will be downloaded to our directory defined by *'--loc1'*, and unzipped to our directory defined by *'--loc2'*.