Creation of a Landsat 8 satellite imagery collection, Collection 2, consisting of 40 scenes showing the equatorial and sub-equatorial Andes during the four epochs of the period 1980-2021.

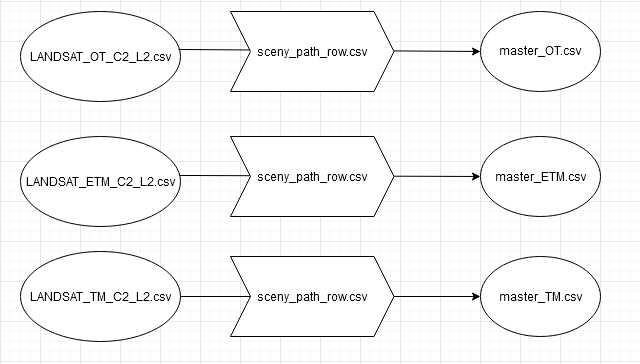
Landsat Metadata Tool documentation

Author: Julia Trzpil

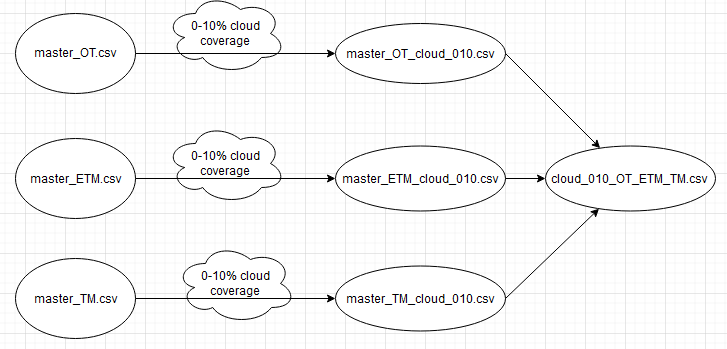
1. Code repository

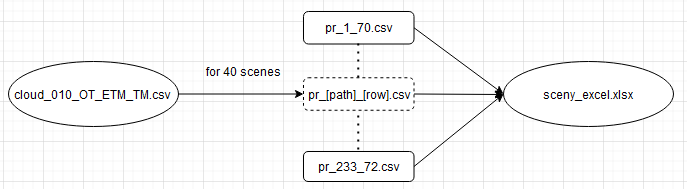
The project of metadata parser and this documentation will be available for download at a public GitHub repository available at <https://github.com/jtrzpil/andy_cent>. It can be download either by Git client by running ‘git clone <https://github.com/jtrzpil/andy_cent>’ in command line, or clicking on ‘download ZIP’ on the GitHub site. Apart from that, all project files are also located on the external drive which will be handover at the date of project submission.

1. Used technologies  
     
   The main language used for extracting data and operating on large .csv files was Python3 with Pandas framework. Also Jupyter notebook was used for the development process. The project and data files are stored on GitHub repository.
2. Data Source of metadata Landsat scenes  
     
   The Landsat scenes were downloaded from USGS Bulk Metadata Service available at <https://www.usgs.gov/core-science-systems/nli/landsat/bulk-metadata-service>. The metadata was part of Landsat Collection 2, level 2 for the following Sensors:  
   a) Landsat 8 Operational Land Imager (OLI)/ Thermal Infrared Sensor (TIRS)  
   b) Landsat 7 Enhanced Thematic Mapper Plus (ETM+)  
   c) Landsat 4-5 Thematic Mapper (TM)
3. Data extraction for given Landsat scenes  
     
   Based on the content of wrs2\_ncn.shp file the scenes path and row identifiers were extracted to a csv file called sceny\_path\_row.csv. Sequentially andy\_OT\_ETM\_TM.py should be run from the repository. It takes the path and row csv file and saves the data from 3 masterdata csv downloaded from USGS to new csv files for each sensor.  
   All of the columns are copied to their respective files.



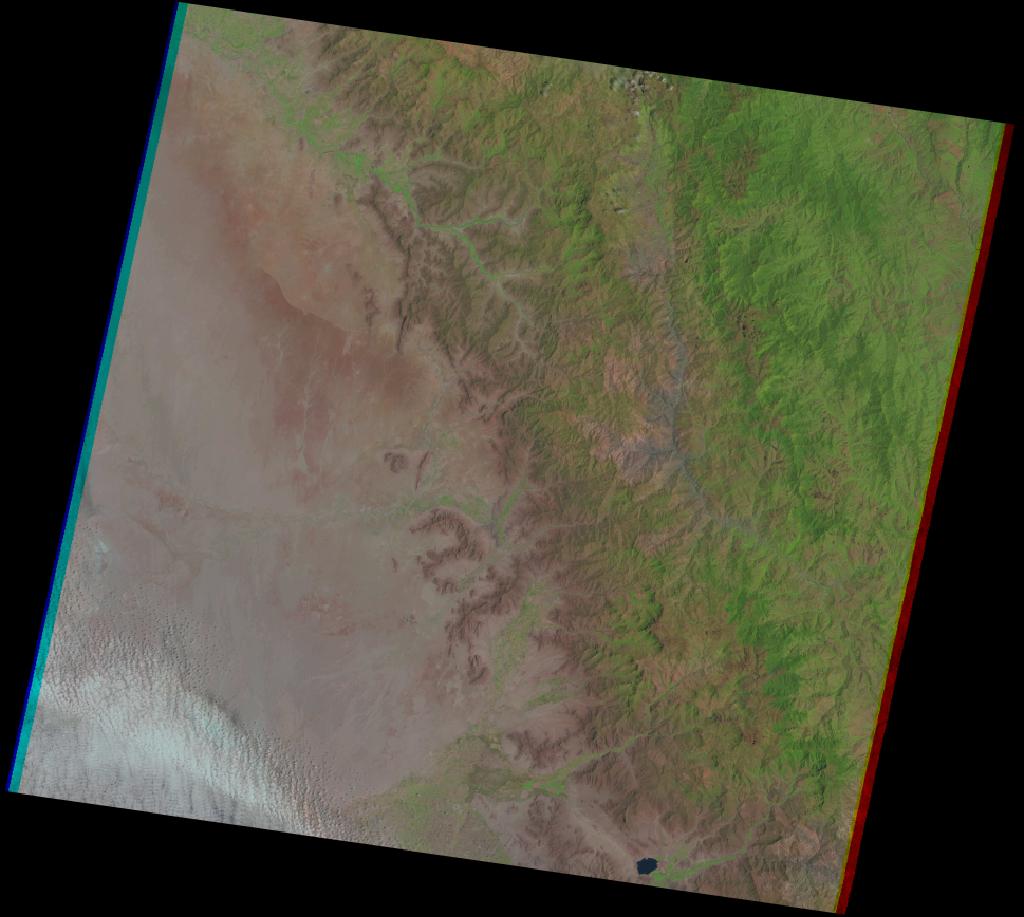
3. Data preparation for quality analysis

In cloud\_cover\_0\_10.py each of the master data csv files for given scene was filtered to only feature scenes with Scene Cloud Cover below 10% resulting in 3 new .csv files. In cloud\_cover\_0\_10\_OT\_ETM\_TM.py all the 3 files were merged into one .csv file.   
  


Then in scenes\_cloud\_0\_10.py, for each scene a new file was generated with data from each sensor. Finally, all 40 scenes were merged into one excel file with 40 sheets – all\_filtered\_scenes.xlsx (E:\datasets\user\_data\results\all\_filtered\_scenes.xlsx).  
 

1. Quality assurance of selected images based on previews

Satellite image previews are available on <https://landsatlook.usgs.gov/> as .jpg. The URL can be found both under the ‘Browse Link’ column in the original and filtered Landsat metadata files. Based on the previews from Landsatlook site, the images were analysed manually regarding the cloud coverage. The results were saved in the merged excel file – final\_sets.xlsx (E:\datasets\user\_data\results\final\_sets.xlsx).



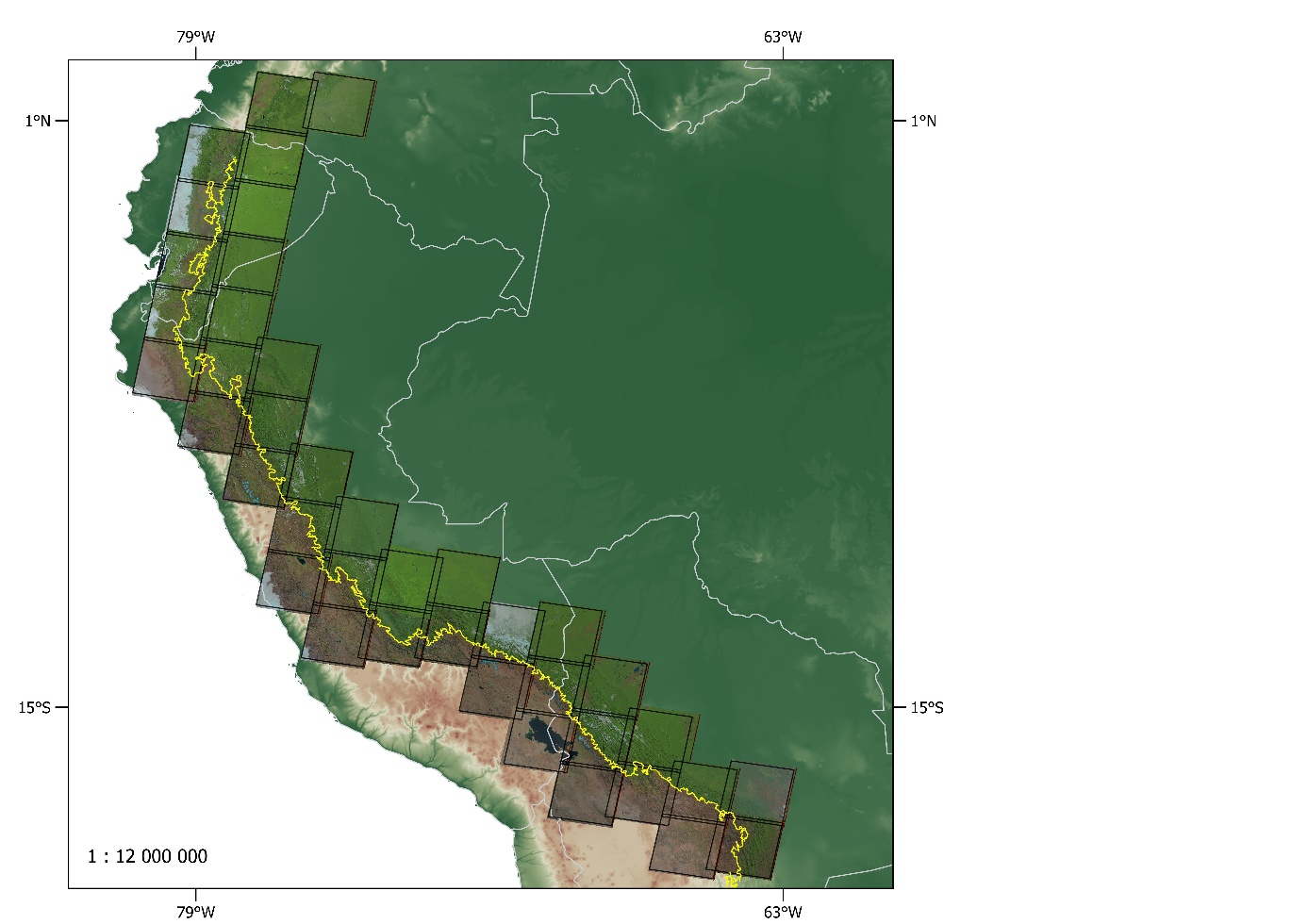
Picture 1. Sample preview of a scene

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| WRS Path | WRS Row | CeNT | Done | Removed from selection |
| 1 | 70 |  | v |  |
| 1 | 71 | v | v |  |
| 1 | 72 |  | v |  |
| 2 | 69 |  |  | x |
| 2 | 70 | v | v |  |
| 2 | 71 |  |  | x |
| 3 | 69 | v | v |  |
| 3 | 70 |  | v |  |
| 4 | 68 |  |  | x |
| 4 | 69 | v | v |  |
| 5 | 68 |  |  | x |
| 5 | 69 | v | v |  |
| 6 | 67 |  |  | x |
| 6 | 68 | v | v |  |
| 6 | 69 |  | v |  |
| 7 | 66 |  | v |  |
| 7 | 67 | v | v |  |
| 7 | 68 |  | v |  |
| 8 | 59 |  |  | x |
| 8 | 64 |  |  | x |
| 8 | 65 | v | v |  |
| 8 | 66 | v | v |  |
| 9 | 59 |  |  | x |
| 9 | 60 |  |  | x |
| 9 | 61 |  |  | x |
| 9 | 62 |  |  | x |
| 9 | 63 |  |  | x |
| 9 | 64 |  | v |  |
| 9 | 65 | v | v |  |
| 10 | 60 |  | v |  |
| 10 | 61 |  | v |  |
| 10 | 62 |  | v |  |
| 10 | 63 |  | v |  |
| 10 | 64 |  | v |  |
| 231 | 72 |  | v |  |
| 231 | 73 |  | v |  |
| 232 | 72 | v | v |  |
| 232 | 73 |  |  | x |
| 233 | 71 | v | v |  |
| 233 | 72 | v | v |  |

Table 1. Scenes used in the project.

1. Final satellite image selection based cloud coverage and lightning conditions  
     
   Previews where the border of the forest was visible were chosen, then the Sun Azimuth column was compared based on that, a set was created with similar sun azimuth with maximum difference 3**°.**  
     
   Not all scenes however could be analysed in this way. Some scenes had cloud coverage above 10% meaning that cloud\_cover\_0\_10\_OT\_ETM\_TM.py couldn’t generate a csv with filtered data. Therefore the scenes needed to be manually searched on earthexplorer.usgs.gov and based on that a list of good images was prepared. Subsequently metadata was fetched from master\_OT.csv; master\_ETM.csv; master\_tm.csv based on the ‘Browse link’ column.
2. QGIS project visualisation based on results.  
     
   The final visualisation is in form of a QGIS project. The project is located on external drive (

Second project (E:\project2.qgz) shows data availability. Each of those scenes has an attribute (field “number” in attribute table) presenting how many satellite images made it through quality assurance and create a set that could be analysed and compared.



Picture 2. Area of research preview in QGIS