mtg2snap.py

This python script was developed to transform the disseminated MTG FCI files to a set of NC4 files that could be read and processed with ESA SNAP software.

syntax

GUI:

```
[python] mtg2snap.py [config]
```

cmdline:

```
[python] mtg2snap.py config date_1 slots_1 res_1 [ date_2 slots_2 res_2 … ]
```

Requirements

This script was developed using Python 3.12.0 on MS Windows 10, 64bit. It was not tested on any other MSWin version, on Linux or on any other OS, nor was it tested with any older Python release.

In addition to basic Python distribution it needs following packages to be installed:

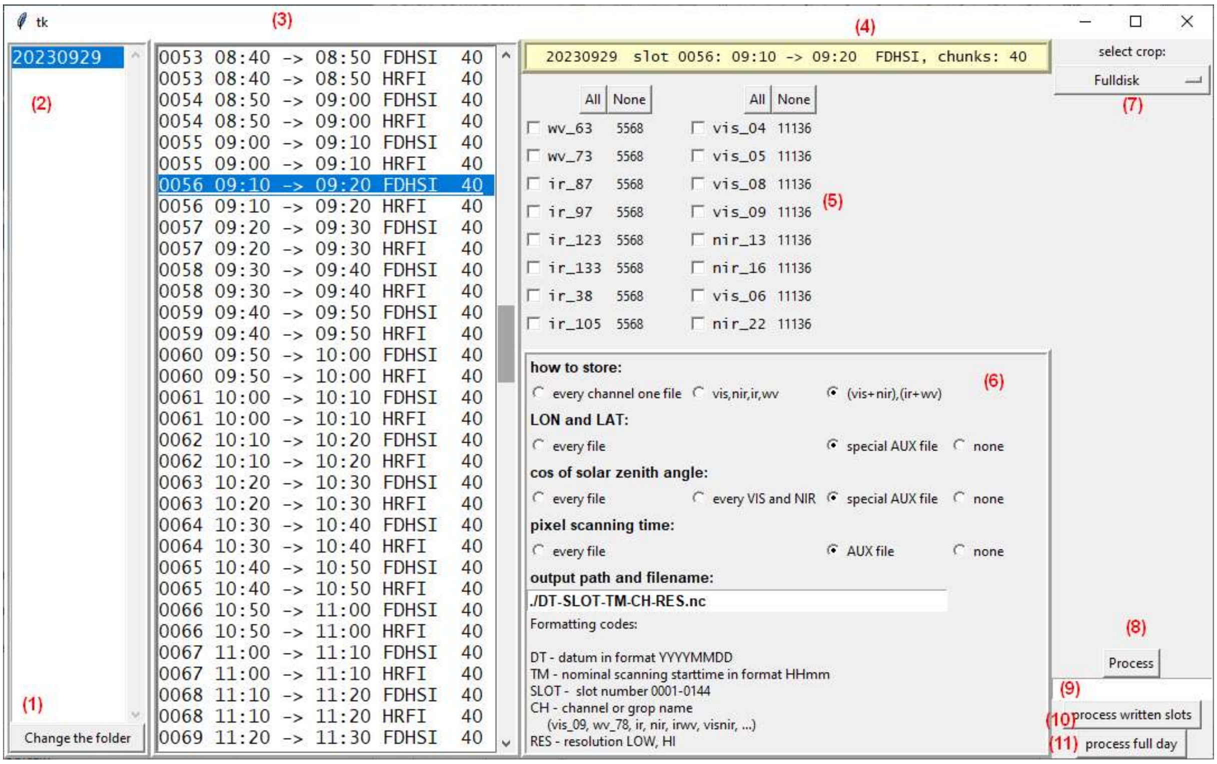
package	version used	why
netCDF4	1.6.5	the data are transformed from NC4 to NC4, just different structure
numpy	1.26.1	the NetCDF data are transformed from/to numpy arrays and numpy functionality is heavily relied upon
hdf5plugin	4.3.0	Eumetsat used some advanced compression algorithms that are unavailable in the basic NetCDF. Without this package you would be able to process only files without “DIS” in their name.
pandas	2.1.2	some database simulation, I used it to store some info about chunks and channels

executing

This script can be executed in two ways. As GUI-based or as command-line.

GUI

If this script is executed without any parameter or with one parameter (will be discussed later), it opens GUI and can be controlled through it. The GUI looks like this:



```
20230929
```

```
0053 08:40 -> 08:50 FDHSI  40
0053 08:40 -> 08:50 HRFI   40
0054 08:50 -> 09:00 FDHSI  40
0054 08:50 -> 09:00 HRFI   40
0055 09:00 -> 09:10 FDHSI  40
0055 09:00 -> 09:10 HRFI   40
0056 09:10 -> 09:20 FDHSI  40
0056 09:10 -> 09:20 HRFI   40
0057 09:20 -> 09:30 FDHSI  40
0057 09:20 -> 09:30 HRFI   40
0058 09:30 -> 09:40 FDHSI  40
0058 09:30 -> 09:40 HRFI   40
0059 09:40 -> 09:50 FDHSI  40
0059 09:40 -> 09:50 HRFI   40
0060 09:50 -> 10:00 FDHSI  40
0060 09:50 -> 10:00 HRFI   40
0061 10:00 -> 10:10 FDHSI  40
0061 10:00 -> 10:10 HRFI   40
0062 10:10 -> 10:20 FDHSI  40
0062 10:10 -> 10:20 HRFI   40
0063 10:20 -> 10:30 FDHSI  40
0063 10:20 -> 10:30 HRFI   40
0064 10:30 -> 10:40 FDHSI  40
0064 10:30 -> 10:40 HRFI   40
0065 10:40 -> 10:50 FDHSI  40
0065 10:40 -> 10:50 HRFI   40
0066 10:50 -> 11:00 FDHSI  40
0066 10:50 -> 11:00 HRFI   40
0067 11:00 -> 11:10 FDHSI  40
0067 11:00 -> 11:10 HRFI   40
0068 11:10 -> 11:20 FDHSI  40
0068 11:10 -> 11:20 HRFI   40
0069 11:20 -> 11:30 FDHSI  40
```

```
20230929  slot 0056: 09:10 -> 09:20  FDHSI, chunks: 40
```

☐ wv_63 5568 ☐ vis_04 11136
☐ wv_73 5568 ☐ vis_05 11136
☐ ir_87 5568 ☐ vis_08 11136
☐ ir_97 5568 ☐ vis_09 11136
☐ ir_123 5568 ☐ nir_13 11136
☐ ir_133 5568 ☐ nir_16 11136
☐ ir_38 5568 ☐ vis_06 11136
☐ ir_105 5568 ☐ nir_22 11136

how to store:
○ every channel one file ○ vis,nir,ir,wv ● (vis+nir),(ir+wv)

LON and LAT:
○ every file ● special AUX file ○ none

cos of solar zenith angle:
○ every file ○ every VIS and NIR ● special AUX file ○ none

pixel scanning time:
○ every file ● AUX file ○ none

output path and filename:
./DT-SLOT-TM-CH-RES.nc

Formatting codes:

DT - datum in format YYYYMMDD
TM - nominal scanning starttime in format HHmm
SLOT - slot number 0001-0144
CH - channel or grop name
(vis_09, wv_78, ir, nir, irwv, visnir, ...)
RES - resolution LOW, HI

Legend:

1. button for working directory change (you must go to the directory where MTG images reside)
2. list of dates for which are MTG data present in the current working directory
3. list of slots for the date selected in (2) present in the current working directory
4. description of slot selected in (3)
5. list of channels available in the slot and their selection for future processing
6. additional settings of the processing
7. crop selection for the future processing
8. button to process the selected single slot
9. editbox for writing list of slots you want to be processed
10. button for processing slots written into the editbox (9)
11. button for processing all available slots of the selected date from (2)

The workflow is generally this (I will use indexes from previous list):

1. select the working folder using the button (1) and the “open folder” dialog which appears after you press the button.
2. After you select the correct folder, in the (2) you should see list of dates for which chunks are present. Select the date you want (only single date can be selected)
3. in (3) you can see the list of available slots, their nominal times, resolution and number of chunks available. Select the slot you want (only single can be selected)
4. in (4) you can see the same description of the slot you have selected, in (5) there is list of available channels divided into groups “ir+wv” and “vis+nir”. This division is purely due to different resolution. You can select channels for processing using their respective

checkbuttons. If you want process all channels in the group, you an use the “All” button, or you can clear the selection with the “None” button.

5. in (6) select how to store the channels and additional info:
    a. “how to store” – there are 3 possibilities:
        i. every channel is stored in separate file
        ii. channels are stored in groups depending on the prefix in their name. There will be separate file with vis channels, with nir channels, with ir channels and with wv channels.
        iii. channels are stored in groups due to their resolution. There will be file with ir and wv channels together and file with vis and nir channels together
        iv. the fourth possibility was impossible because I was not able to store arrays of different dimensions into a single nc4 file
    b. “LON and LAT” – where to store the geographical coordinates:
        i. every file created will contain lon/lat respective to the dimension
        ii. special AUX file will be created for each dimension, which means that there will be two AUX files, one for ir+wv and one for vis+nir channels. This is independent on selection in point 5.
        iii. they will not be saved at all
    c. “cos of solar zenith angle”
        i. every file – like 5.b.i
        ii. every vis and nir – similar to the previous one, only that long infrared channels will be omitted, as this value may seem unnecessary for them.
        iii. special AUX file - like 5.b.ii, if there is lon/lat stored, they will be together.
        iv. it will not be saved at all
    d. “pixel scanning time” – this was a special request from MS. He wanted to be able to compare scanning time with the time of other observations or other data.
    e. “output path and filename”
    It is possible to somehow affect the output path and filename.

    If you want to add some folder, you can use slash (/) for directory separator even on MSWin, as python can handle it correctly. If the path does not exist, the script tries to create it forcefully.

    Regarding the name format here are codes that will be later replaced with real values:
        i. DT – nominal date in format YYYYMMDD, the same as in (2)
        ii. TM – nominal date of scan start in format HHmm
        iii. SLOT – slot index from the range 1-144 in the format nnnn with leading zeros
        iv. RES – resolution LOW or HI (FDHSI -> LOW, HRFI -> HI)
        v. CH – it depends on selection in 5.a:
            1. if every channel is stored in its file, then CH is replaced with channel name
            2. if groups ir, wv, vis and nir are used, CH is replaced with one of this group
            3. if compound groups “ir+wv” and “vis+nir” are used, CH is replaced with “irwv” or “visnir”
6. in (7) select what part of the disk you want to be processed, either the fulldisk or some crop. The list is predefined, will be discussed later

7. If you want to process just the selected slot, press button (8) and wait for surprise
8. If you want to process several slots for the selected date, you can write them into the editbox (9). It is similar to printing where you select pages. You can use list with commas and intervals with “-“. E.g you can use this: 1-7,60-120,138,140
   When writing an interval, both limits are processed. If you write “1-3”, slots 1, 2 and 3 will be processed. Notice that slot indexing is 1-based! After you wrote the slot list, press the button (10) and wait for surprise.

   The selection in listbox (3) will be ignored in term that it is not forcibly included into the list in case you omit it. It serves only as a source to determine the resolution and channels. If you selected FDHSI slot, only FDHSI slots are processed, if you selected HRFI, only HRFI slots are processed.
9. If you want to process the whole day, you can either write slots 1-144 and press the button (10), or you can just press the button (11) and wait for surprise.

When processing, the configuration is saved into some file. The default name is “**mtg2snap_RES.cfg**”, where the code RES is replaced with the resolution processed, either FDHSI or HRFI. This file is located in the same directory as the python script itself, regardless of the current working directory. The file contains configuration used in the process in JSON format. If the file already exists, it is replaced and previous info is lost. You can change the prefix in the filename from “msg2snap” to another text if you append one parameter to the script execution command. E.g. if you call the script as “`mtg2snap.py MyCatHatesMe`” and process some FDHSI slot(s), the configuration is stored into “**MyCatHatesMe_FDHSI.cfg**”. Here is some example of the file:

```
{
    "channels": {
        "wv_63": 1,
        "wv_73": 1,
        "ir_87": 1,
        "ir_97": 1,
        "ir_123": 1,
        "ir_133": 1,
        "ir_38": 1,
        "ir_105": 1,
        "vis_04": 1,
        "vis_05": 1,
        "vis_08": 1,
        "vis_09": 1,
        "nir_13": 1,
        "nir_16": 1,
        "vis_06": 1,
        "nir_22": 1
    },
    "division": "typ",
    "lonlat": "aux",
    "cossolar": "aux",
    "timechan": "aux",
    "output": "./pok/DT-SLOT-TM-CH-RES.nc",
    "crop": "MSG B region"
}
```

In the “`channels`” section there is selection for every channel: 1=process, 0=do not process.

The “`division`” value determines how channels will be stored:

- “every” – every channel its own file
- “typ” – separate file for VIS, IR, NIR, WV
- “group” – separate files for VIS+NIR and for IR+WV

The “`lonlat`”, “`cossolar`” and “`timechan`” sections determine how will the LON+LAT, solar zenith angle cosinus and the pixel scanning time stored:

- “every” – every file contains the info
- “vis” – only VIS and NIR files contain the info (valid only for “cossolar”!)
- “aux” – info is stored to AUX file
- “none” – info is not stored

The “`output`” section contains the data output format and path used

The “`crop`” session contains identification of the image crop used.

crop definitions

Crop definitions are located in the file **mtg2snap_crops.json** and as the extension implies, they are in JSON format.

There are two sections: “`crops`“ and “`cropcode`“. In the first on ethere are the definitions and in the second on ethere are abbreviations of those with complicated names.

Each crop definition contains the name (must be unique through the “`crops`“ section) and a set of four coordinates: [x_left, x_right, y_top, y_bottom]. The image coordinates are based on the smallest type of data (IR/WV low res, 5568*5568pxl). Upper-left corner has coordinates[x=0, y=0], x from left to right, y from top to bottom. There is a vis_08 image resampled added for convenience if you want to create your own definitions. The crop values are recalculated when processing other resolution image (i.e. multiplied by 2 or 4).

The crop name may be any general unicode text. In the case there is a difficulty in using the name in command line, you can create an abbreviation and add it to the “`cropcode`“ section. The value related with the abbreviation must be the exact name from “`crops`“ section.

In the script configuration file in the “`crop`” parameter you can use either the crop name or the cropcode, whichever is more convenient.

executing as command-line

When executing as command-line, use following syntax

```
python mtg2snap.py config date_1 slots_1 res_1 [ date_2 slots_2 res_2 … ]
```

where

- config is the prefix of configuration file
- date_n is the date in format YYYYMMDD
- slots_n is the slot list as wrttten in the editbox (9)
- res_n is the resolution, “FDHSI”, “HRFI”, “ALL”, “BOTH”. When used “ALL” or “BOTH”, both resolutions are processed

When the script is called with more than one parameter, it does not open the GUI and runs as command-line app.

The script must be called directly from the directory where the original images are stored. All the data to be processed in the single call must be stored in this directory. It does not move to another directory during the process. The output path in the configuration file must be either absolute or must be relative to the working directory.

E.g. if you call the script with parameters

```
SAHARA 20231201 “1-144” FDHSI 20231201 “40-105” HRFI
```

the script will:

- look for the config file **SAHARA_FDHSI.cfg** in its HOME directory (where the script is stored)
- process data from 1st December 2023, all slots, low res
- look for the config file **SAHARA_HRFI.cfg** in its home directory
- process data from 1st December 2023, slots 40 to 105 incl., high res

The script searches for slots which belong to the given list. If there is any slot missing, there will be no error produced.

NOTES:

1. When running as GUI app, the config file is just written/rewritten. Its data is never read.
2. When running as commandline app, the config file is just read, it is never rewritten, as there is no reason for any change.
3. When selecting some slot in GUI, the channel list is always newly created and cleared. If you select some slot, select channels for processing and then select another slot, the channel selection is cleared and you must again select those channels.
4. There may be memory problems when displaying the HRFI full globe data. Better do not try to store the LON/LAT data into the same file as channel data. Although it is made successfully, SNAP was unable to open the file due to memory problems, even on the PC with 16GB RAM. Thee should be no problem with some smaller crops.
5. Please, do not expect any brutal speed. When converting the whole globe, on PC with 16GB RAM and Intel i7-7800 it took me about 8 minutes to convert the full dataset, FDHSI and HRFI together.
6. Any questions and complains direct at pavel.hampl@chmi.cz