

# **ATLAS (AuTomedated Lidar Analysis Software) manual**

Author(s): Nikolaos Siomos

Last Update: 11/23/22

Version: 0.1

## Table of Contents

1	Installation.....	3
2	Dependencies.....	3
3	Execution.....	4
4	The configuration file.....	8
4.1	Section: Lidar.....	9
4.2	Section: Channels.....	10
4.3	Channel Nomenclature.....	16
5	The settings file.....	17
5.1	Section: general.....	17
5.2	Section: converter.....	19
5.3	Section: preprocessor.....	22
5.4	Section: quicklooks.....	24
5.5	Section: rayleigh_fit.....	27
5.6	Section: telecover.....	29
5.7	Section: polarization_calibration.....	31
6	The radiosonde files.....	34

# 1 Installation

ATLAS can be cloned through git. In order to do so:

- Git must be installed in the computer
- ATLAS can be cloned from the stable repository: <https://github.com/nikolaos-siomos/ATLAS>
- Open a terminal and cd to the folder you want to place ATLAS to
- run: `git clone https://github.com/nikolaos-siomos/ATLAS.git`

Alternatively you can just download ATLAS from the link above.

# 2 Dependencies

The current version of ATLAS has been developed and tested in python 3.9. The following python packages should be installed:

- numpy – version 1.24.0
- pandas – version 1.5.2
- xarray – version 2022.12.0
- netcdf4 – version 1.6.2
- scipy – version 1.9.3
- matplotlib – version 3.6.2
- h5netcdf – version 1.1.0

Note that other package versions (or higher versions) might also coincidentally work.

The recommended ways to install the dependencies and run ATLAS is currently either through an Anaconda environment (tested with Miniconda) or with poetry.

## 2.1 Poetry installation

The dependencies are automatically collected with poetry as long as the dependency file `pyproject.toml` is present in the downloaded ATLAS project folder.

An installation example follows for Linux systems.

- Download and install the latest poetry and add it to your path (instructions here <https://python-poetry.org/docs/>)
- Create an environment for python 3.9 (if not already created by poetry)
- Go to the ATLAS project folder and install the dependencies e.g. with `poetry install`

- Before running ALTAS make sure that the python 3.9 environment is activated.

## 2.2 Anaconda installation

The dependencies for the current version were collected with conda – version 4.12.0

An installation example follows for Linux systems (it should also work for Windows and MAC systems – if problems arise the users are encouraged to provide feedback to the EARLINET forum):

- Download and install the latest [Miniconda](#) for python 3.9
- Activate the **base** environment (usually it is auto-activated by default)
- Create a new environment to install the dependencies of ATLAS (let's call it **atlas\_box** here):
  - `conda install -n atlas_box python=3.9`
- It is highly recommended to install **mamba** at this stage that greatly speed ups conda.
  - `conda install -n base mamba`
  - If mamba is not installed, please replace the word “mamba” with the word “conda in all commands below. Test were run with mamba – version 0.15.3
- Install all dependencies using mamba in the **atlas\_box** environment
  - `mamba install -n atlas_box numpy=1.24 pandas=1.5.2 xarray=2022.12.0 netcdf4=1.6.2 scipy=1.9.3 matplotlib=3.6.2 h5netcdf=1.1.0`
  - If one or more of the packages are not available in the specific version then try installing the version the package manager recommends e.g.:
  - `mamba install -n atlas_box numpy pandas xarray netcdf4 scipy matplotlib h5netcdf`
- Before running ALTAS make sure that the **atlas\_box** environment is activated. It is best to close the terminal after installation of the dependencies and open a new one to execute ATLAS

## 3 Execution

Execution of ATLAS is quite straightforward given that the necessary input files (settings file, configuration file) are properly set.

In general all necessary input and output paths can be provided as command-line arguments (see the argument list below). The default folder structure is used in the example bellow to make the definition of paths a bit more clear. Let's assume the following two folders:

*/mydrive/ATLAS*: contains the ATLAS software downloaded from Git. The `__master__.py` is the main executable and it is located at */mydrive/ATLAS/program/\_\_master\_\_.py*

*/anotherdrive/lidar\_data*: the input lidar data and the configuration and settings file are placed here

*/anotherdrive/radiosondes*: the input radiosonde ascii data

The default folder structure that ATLAS expects is given below. Directories are marked with bold and files with bold italic:

- **/mydrive**
  - **/ATLAS**
    - **/program**
      - ***\_\_master\_\_.py*** – main executable
- **/anotherdrive**
  - **/radiosondes** – must exist, see section 6 for the filename format
  - **/lidar\_data**
    - ***/settings\_file.ini*** – must exist, see section 5
    - ***/config\_file.ini*** – must exist, see section 4
    - **/nrm** – normal measurement, optional (must include a pol. Calibration for PollyXT systems)
    - **/tlc** – telecover measurement with 4 sectors, optional
      - **/north** – north sector files
      - **/east** – east sector files
      - **/south** – south sector files
      - **/west** – west sector files
    - **/tlc\_rin** – telecover measurement with 2 rings, optional
      - **/inner** – inner ring files
      - **/outer** – outer ring files
    - **/pcb** – polarization calibration measurement, optional (ignored for PollyXT systems – nrm folder is used instead)
      - **/+45** – files recorded with the calibrator at +45° (or equivalent, e.g. HWP at 22.5°)
      - **/-45** – files recorded with the calibrator at -45° (or equivalent, e.g. HWP at 22.5°)
    - **/drk** – common dark measurement for all QA tests – optional

Basic execution example assuming anaconda was used for the installation of the python environment:

- Activate the **atlas\_box** environment using:

- `conda activate atlas_box`
- Get the absolute path to the `__master__.py` file, eg. `/mydrive/ATLAS/__master__.py`
- We will need at least the following argument:
  - `--parent_folder /mydrive/lidar_data/ --radiosonde /mydrive/radiosondes --file_format licel`
- Run the following command:
  - `python /mydrive/ATLAS/__master__.py --parent_folder /mydrive/lidar_data/ --radiosonde /mydrive/radiosondes --file_format licel`
- For a PollyXT system substitute **licel** with **polly\_xt**

It is possible to run ATLAS also from the Spyder IDE. For execution using Spyder:

- Press **Ctrl + F6** OR go to **Run --> Configuration per File**
- Activate **Run file with custom configuration**
- Tick the **Command line options**
- Add all necessary arguments inside the box:
  - `--parent_folder /mydrive/lidar_data/ --radiosonde /mydrive/radiosondes --file_format licel`
  - For a PollyXT system substitute **licel** with **polly\_xt**
- Process using **F5** OR **Run File** like a normal python script

Sometimes it is more convenient to use different paths and/or filenames for the configuration file and the settings file. If for example they were all placed directly under the `/mydrive` folder and `config_file.ini` and `settings_file.ini` were renamed to `config_file_polis.ini` and `settings_file_polis.ini`:

- We would need additionally the following arguments:
  - `--config_file /mydrive/config_file_polis.ini`
  - `--settings_file /mydrive/settings_file_polis.ini`
- And the execution command becomes:

- `python /mydrive/ATLAS/__master__.py --parent_folder /mydrive/lidar_data/  
--radiosonde /mydrive/radiosondes --file_format licel --config_file /mydrive/config_file.ini  
--settings_file /mydrive/processing_options.ini`
- For a PollyXT system substitute **licel** with **polly\_xt**

### The output

Status messages and errors will be displayed on the terminal screen while ATLAS is processing. ATLAS is divided in three parts, the raw file converter, the preprocessor, and the visualizer. The output of the first two modules is one netcdf file per module and QA test processed. The output of the visualizer is a set of plots per channel and per QA test. By default all netcdf files are placed in the **netcdf** folder inside the parent folder. Likewise the plots will be placed in the **plots** folder inside the parent folder. Those two folders will be created by ATLAS if they do not exists. Alternatively, the user can define separate paths for the output of each module using specific arguments that are explained in the list of arguments paragraph.

The list of ATLAS arguments.

According to the default folder structure all paths are set with respect to a user provided parent folder path. Providing a path explicitly always overrides its default value. If the parent folder is not provided then all other paths must be explicitly defined. The file format is a mandatory argument, it must be provided.

-f or --parent\_folder

The path to the parent folder that is the basis of the default folder structured. It is mandatory unless all the other paths are explicitly provided.

--dark\_folder

The path to the dark folder. Defaults to: a drk folder inside the parent folder

--rayleigh\_folder

The path to the rayleigh fit measurement folder. Defaults to: a nrm folder inside the parent folder

--telecover\_sectors\_folder

The path to the telecover folder that contains the sector files. Defaults to: a tlc folder inside the parent folder

--telecover\_rings\_folder

The path to the telecover folder that contains the ring (inner/outer) files – if any. Defaults to: a tlc\_rin folder inside the parent folder)

--pcb\_cal\_p45\_folder

The path to the polarization calibration +45 folder. Defaults to: a pcb/+45 folder inside the parent folder

--pcb\_cal\_m45\_folder

The path to the polarization calibration -45 folder. Defaults to: a pcb/-45 folder inside the parent folder

--pcb\_cal\_stc\_folder

The path to the polarization calibration folder for a calibration with a single calibrator position. Defaults to: a pcb/stc folder inside the parent folder

--radiosonde

The path to the radiosonde file or the radiosonde folder. Provide either a folder path containing ascii radiosonde files or a netcdf file with the radiosonde data in the expected ATLAS (SCC)



format. If a folder path is provided, the radiosonde file that is closest to the measurement within 12h will be selected if more than 1 files are provided inside

`--radiosonde_filename`

The absolute path to the radiosonde file. Use it to provide directly an already generated netcdf file in the expected ATLAS (SCC) format. Cannot be provided together with the radiosonde

`--converter_out`

The path to the folder where the converted netcdf files will be placed. This optional argument can be used if the folder must be placed out of the `parent_folder`. Defaults to:  
`parent_folder/netcdf`

`--preprocessor_out`

The path to the folder where the preprocessed netcdf files will be placed. This optional argument can be used if the folder must be placed out of the `parent_folder`. Defaults to:  
`parent_folder/netcdf`

`--visualizer_out`

The path to the folders where the plots and ascii folders will be created. This optional argument can be used if these two folders must be placed out of the `parent_folder` (default)

`-c` or `--config_file`

The path to the configuration file that contains the necessary metadata. This optional argument can be used if the configuration file must be placed out of the `parent_folder` (default)

`-s` or `--settings_file`

The path to the settings file that contains the processing options to run ATLAS. This optional argument can be used if the configuration file must be placed out of the `parent_folder` (default)

`--file_format`

Raw lidar file format. Currently only `licel` and `polly_xt` are supported. Choose one of:

- `licel`: for systems that use `licel` recorders
- `polly_xt`: for systems that use `polly_xt` recorders

`--isday`

- Defaults to: False. If set to True all the vibrational Raman and the fluorescence signals will be excluded in the preprocessing stage

`--quick_ran`

- Defaults to: False. If set to True the converter and the preprocessing modules will not be called if the algorithm detects output files already produced by them for a specific measurement. This mainly saves time during execution.

#### --process

- The user can choose specific QA test(s) to process. Use any of:
  - ray: Rayleigh Fit
  - tlc: Telecover Test
  - pcb: Polarization Calibration
  - off: Nothing will be processed

Defaults to: ray, tlc, pcb

#### --process\_qck

- Choose which quicklooks to process. It must be a subset of **process**. Note that each visualizer module (ray, tlc, pcb) can create their own quicklooks. Choose among
  - ray: Rayleigh Fit
  - tlc: Telecover Test
  - pcb: Polarization Calibration
  - off: Nothing will be processed

Defaults to: ray

## 4 The configuration file

The *configuration\_file.ini* sample file is provided inside the *./templates* folder. The file consists of two main sections, Lidar and Channels. The file is structured according to the following format:

```
[Lidar]
variable_1 = <single_element_1>
variable_2 = <single_element_2>
.....

[Channels]
variable_A = <element_A1>, <element_A2>, ....., <element_An>
variable_B = <element_B1>, <element_B2>, ....., <element_Bn>
.....
```

According to the example, the first section consist of variables that take a single value as input while the later consist of variables that take coma separated values as input. For the latter, the number of values per variable is always equal to the number of signal channels that the user wants to process.

Some fields are optional. The raw file parser attempts to fetch all available metadata from the raw files. Default values are provided when the user either leaves them empty or does not include them in the configuration file at all. For example:

```
variable_1 = <single_element_1>
variable_2 =
variable_3 = <single_element_3>
```

Is equivalent to:

```
variable_1 = <single_element_1>
variable_3 = <single_element_3>
```

The variable\_2 will not be taken at all into account and default values will be used instead. Please not that some of the fields are mandatory and there are others that are partly mandatory in a scene that using the default values can be quite error prone. More details on the variables themselves are provided the following sections.

## 4.1 Section: Lidar

### Mandatory Variables

**lidar\_name:** The lidar name. This field will be reported in the plots.

*Example: lidar\_name = POLIS*

**lidar\_id:** The lidar id according to the EARLINET DB.

*Example: lidar\_id = mun*

### Optional Variables

**file\_format:** Define the parser for the raw files (Defaults to: licel). Currently licel and polly\_xt are supported.

*Example: file\_format = polly\_xt*

**altitude:** The altitude of the station above sea level (in meters).

*Example: altitude = 535*

**location:** The station location. This field will be displayed in the plots.

*Example: location = Munich*

**latitude:** The station latitude (in degrees).

*Example: latitude = 48.148*

**longitude:** The station longitude (in degrees).

*Example: longitude = 11.573*

**zenith\_angle:** Zenith angle of the lidar (in degrees, 0 at zenith, 90 at horizon).

*Example: zenith\_angle = 41*

**azimuth\_angle:** Azimuth angle of the lidar (in degrees, North = 0, E = 90).

*Example: azimuth\_angle = 0*

## 4.2 Section: Channels

All variables provided for this section should include exactly as many channels as the `channel_id` variable. If a variable is not relevant for a specific channel (e.g. `dead_time` for analog channels) fill with “\_”. Note that the number of channels provided has to be equal or less than the total number of channel in the raw input files. A subset of the raw channels can be provided in case the user doesn’t want to process all of them. For Licel systems, the Licel channel ID in combination with the laser identifier in the header (see the Licel manual) are used in order to link with the raw binary files. For PollyXT systems there is no specific identifier per channel. The user can use the channel index (first channel --> 1, second channel --> 2, etc) to link with the raw netcdf files. If nothing is provided, all PollyXT channels available in the raw files will be processed. For example, `channel_id = 1, 2, 4` means that the 3<sup>rd</sup> channel ordered as in the raw netcdf file will not be processed. It is therefore important that all input files **have exactly the same channels** when processing a measurement with ATLAS (e.g. same channels for all QA tests). This is also a good practice for Licel systems.

A detailed explanation of all the variables is provided below with examples. For all examples we have assumed the same system of 12 channels (6 physical) including analog and photon detection at 355 parallel and cross, 387 Raman, 532 parallel and cross, and 607 Raman. Please note that all example values presented here correspond to a fictional system and **should do not be blindly taken and applied to an actual system**.

### Mandatory Variables

**telescope\_type**: The telescope configuration. Choose among:

- n: near range
- f: far range
- x: single telescope

*Example: `telescope_type = x, x, x, x, x, x, x, x, x, x, x, x`*

**channel\_type:** The channel type. Choose one among:

- p: co-polar linear analyzer
- c: cross-polar linear analyzer
- t: total (no depolarization)
- o: co-polar circular analyzer
- x: cross-polar circular analyzer
- v: vibrational Raman
- r: rotational Raman
- a: Cabannes (HSRL)
- f: fluorescence

*Example: channel\_type = p, p, c, c, v, v, p, p, c, c, v, v*

**channel\_subtype:** The channel subtype. Choose among:

- r: Signal Reflected from a PBS
- t: Signal Transmitted through a PBS
- n: N<sub>2</sub> Ramal line
- o: O<sub>2</sub> Ramal line
- w: H<sub>2</sub>O Ramal line
- c: CH<sub>4</sub> Ramal line
- h: High Rotational Raman
- l: Low Rotational Raman
- a: Mie (aerosol) HSRL signal
- m: Molecular HSRL signal
- b: Broadband Fluorescence
- s: Spectral Fluorescence
- x: No specific subtype

*Example: channel\_subtype = r, r, t, t, n, n, t, t, r, r, n, n*

**dead\_time:** The dead time of the photon counting channels (in ns). (for analog channels set \_).

*Example: dead\_time = \_, 3.571, \_, 4.545, \_, 3.704, \_, 3.167, \_, 3.846, \_, 3.846*

**trigger\_delay\_bins:** The trigger delay per channel in bins. Provide negative values if the recording starts before the Q-switch (pre-triggering) and positive values if the recording starts after the Q-switch.

*Example: trigger\_delay\_bins = -2020, -2011, -2020, -2011, -2020, -2011, -2025, -2014, -2025, -2014, -2025, -2014*

#### Mandatory Variables for Licel – Ignored for PollyXT

**channel\_id:** IDs of each channel according to the raw file header. For licel provide the licel id per channel that is going to be include. Currently only BT and BC channels are supported. For PollyXT systems all channels are included by default. Provide only if you want to process specific channels

*Example: channel\_id = BT0, BC0, BT1, BC1, BT2, BC2, BT3, BC3, BT4, BC4, BT5, BC5*

**laser:** The ascending laser number according to the licel manual. Use to link with the licel file. The channel\_id is not a unique identifier for licel channels. A single channel can synchronize with more than one lasers. This variable is ignored when processing PollyXT files

*Example: laser = 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1*

#### Mandatory Variables in the Future

**scc\_id:** IDs of each channel according to the SCC configuration. In the future, linking with the HOI will be done through the scc\_id (currently optional).

*Example: scc\_id = 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630*

### Partly Optional Variables

These variables take default values. It is highly recommended though to fill them in because might not be valid for the system.

**dead\_time\_correction\_type:** The type of dead time correction. Choose among:

- 0: Non Paralyzable
- 1: Paralyzable

Defaults to: 0 for photon channels.

*Example: dead\_time\_correction\_type = \_, 0, \_, 0, \_, 0, \_, 0, \_, 0, \_, 0*

**emitted\_wavelength:** The wavelength of the originally emitted laser beam per channel. A very rough guess process is applied by default using that detected channel information that it is valid **ONLY for Nd:Yag lasers with elastic and vibrational Raman channels**. Please provide the values explicitly if your system does not fall in this category. Defaults to:

- 355 for channels with detected wavelength <520nm
- 532 for channels with detected wavelength <1000nm
- 1064 for channels with detected wavelength >=1000nm

If accurate molecular depolarization ratio calculations are required this variable should be filled in with the actual emitted wavelength with sub-nanometer accuracy.

*Example: emitted\_wavelength = 354.6, 354.6, 354.6, 354.6, 354.6, 354.6, 532.2, 532.2, 532.2, 532.2, 532.2*

**detected\_wavelength:** The wavelength of the detected signal according to the applied Interference Filter of each channel (in nm). This variable is provided by default from the raw file metadata for both Licel and PollyXT. However, if accuracy is required for molecular calculations the exact central wavelength should be provided.

*Example: detected\_wavelength = 354.6, 354.6, 354.6, 354.6, 386.5, 386.5, 532.2, 532.2, 532.2, 532.2, 607.5, 607.5*



**channel\_bandwidth:** Interference filter Bandwidth (in nm, Defaults to: 1nm). The default value is a very rough guess and it can easily lead to high uncertainty in the molecular depolarization value. Exact values provided from the manufacturers should be used instead.

Example: `channel_bandwidth = 0.5, 0.5, 0.5, 0.5, 1.0, 1.0, 0.5, 0.5, 0.5, 0.5, 1.0, 1.0`

**background\_low\_bin:** Starting bin of the background correction averaging range. Defaults to:

- The 600<sup>th</sup> bin before the end of the profile if the `trigger_delay_bins < 400`
- The 100<sup>th</sup> bin if the `trigger_delay_bins >= 400`

Using default values is currently risky. Please make sure that no spikes are appearing in the pretrigger range that would affect the background correction or manually provide this variable.

Example: `background_low_bin = 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100`

**background\_high\_bin:** ending bin of the background correction averaging range. Defaults to:

- The 100<sup>th</sup> bin before the end of the profile if the `trigger_delay_bins < 400`
- The 300<sup>th</sup> bin if the `trigger_delay_bins >= 400`

Using default values is currently risky. Please make sure that no spikes are appearing in the pretrigger range that would affect the background correction or manually provide this variable.

Example: `background_high_bin = 500, 500, 500, 500, 500, 500, 500, 500, 500, 500, 500, 500`

#### Optional Variables (include only to override the raw file metadata)

**acquisition\_mode:** The mode of the recorded signals per channel. Choose among:

- 0: analog
- 1: photon counting

Example: `acquisition_mode = 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1`

**bins:** The total bins of the recorded signals per channel.

Example: `bins = 8192, 8192, 8192, 8192, 8192, 8192, 4096, 4096, 4096, 4096, 4096, 4096`

**laser\_polarization:** The laser polarization of the recorded signals per channel. Choose among:

- 1: linear
- 3: circular

*Example: laser\_polarization = 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1*

**laser\_shots:** The number of acquired laser shots per channel. Not recommended to provide it manually. If fetched from the metadata, this variable is different per file which is more accurate than providing a constant value here.

*Example: laser\_shots = 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200*

**data\_acquisition\_range:** The Data Acquisition Range of each analog channel [mV]. Use \_ for photon channels

*Example: data\_acquisition\_range = 100, \_, 100, \_, 20, \_, 100, \_, 100, \_, 20, \_*

**analog\_to\_digital\_resolution:** The analog to digital resolution (in bits) of each analog channel. Use \_ for photon channels.

*Example: analog\_to\_digital\_resolution = 16, \_, 16, \_, 12, \_, 16, \_, 16, \_, 12, \_*

**range\_resolution:** The range resolution of each channel [m]. It will be used to calculate the Sampling Rate.

*Example: range\_resolution = 3.75, 3.75, 3.75, 3.75, 3.75, 3.75, 7.5, 7.5, 7.5, 7.5, 7.5, 7.5*

**pmt\_high\_voltage:** The high voltage [V] provided to the detection unit per channel.

*Example: pmt\_high\_voltage = 850, 850, 870, 870, 900, 900, 840, 840, 860, 860, 920, 920*

**laser\_repetition\_rate:** The Laser Repetition Rate [Hz]

*Example: laser\_repetition\_rate = 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20*

### 4.3 Channel Nomenclature

Based on the input provided in the configuration file and the header of the raw files an ID is created per channel in ATLAS and accompanies it during processing and exporting. The ID consists of the following 8 characters:

- first 4 characters correspond to the **detected\_wavelength** variable rounded up to the nearest integer and filled with leading zeros wherever applicable (e.g. 0355, 0607, 1064 etc)
- the 5<sup>th</sup> character corresponds to the **telescope\_type** variable (refer to the respective table)
- the 6<sup>th</sup> character corresponds to the **channel\_type** variable (refer to the respective table)
- the 7<sup>th</sup> character corresponds to the acquisition mode and can be “a” for analog or “p” for photon. The respective information is provided by the **acquisition\_mode** variable
- the 8<sup>th</sup> character corresponds to the **channel\_subtype** variable (refer to the respective table)

For example, channel **0355ftpr** would correspond to a far field photon counting co-polar channel at 355nm detected in the reflected path with respect to the PBS

In another example, channel **0387nvan** would correspond to a near field analog vibrational Raman channel at 387nm the corresponds to scattering from N<sub>2</sub> molecules.

Following the 12 channels defined in the examples of section 4.2, we would get the following channels in the ATLAS plots and netcdf files:

0355xpar, 0355xppr, 0355xcat, 0355xcpt, 0387xvan, 0387xvpn, 0532xpat, 0532xppt, 0532xcar, 0532xcpr, 0607xvan, 0607xvpn

## 5 The settings file

The *settings\_file.ini* sample file is provided inside the *./templates* folder. Similarly to the *config\_file.ini* it is also structured in sections and shares a similar syntax. The main difference is that the variables of the settings file do not have a specific number of values per section.

The file consists of seven main sections: converter, preprocessor, quicklooks, rayleigh\_fit, telecover, and polarization\_calibration. Note that none of these variables is truly mandatory. Unless provided they take default values that are explained below. However the settings file must still exist.

## 5.1 Section: converter

**debug:** If set to True, debugging files will be generated in the `<path_to_the_converter_folder>/debug` folder. Note that the path to the converter folder Defaults to: `<parent_folder>/netcdf` but it can be different if manually provided by the user. These included the metadata gathered from the configuration file, the licel header, and the combination of the two. Defaults to: False

*Example: debug = True*

**trim\_overflows:** This options determines how overflow values in the raw input files will be treated if detected. Choose among:

- 0: the algorithm will stop and provide a diagnostic error (default)
- 1: the files containing at least one overflow value will be screened out
- 2: overflows will be interpolated (use with care and only only a few bins per profile have overflow values)
- 3: overflows will not be excluded, use this only for debugging purposes

*Example: trim\_overflows = 2*

**files\_per\_sector:** The number of telecover files per sector (integer). If provided, the telecover files must be placed in a single folder (this should be `<path_to_parent_folder>/tlc` according to the default folder structure). An automated assignment of the telecover files in different sectors will be attempted serially assuming the following temporal sequence of sectors: north – east – south – west. Note that the telecover test can have more than 1 rounds.

*Example: files\_per\_sector = 5*

**files\_per\_ring:** The number of telecover files per ring (integer). If provided, the telecover files must be placed in a single folder (this should be `<path_to_parent_folder>/tlc_rin` according to the default folder structure). An automated assignment of the telecover files in different sectors will be attempted serially assuming the following temporal sequence of sectors: inner - outer. Note that the telecover test can have more than 1 rounds.

*Example: files\_per\_ring = 5*

**rsonde\_skip\_header:** Radiosonde parser option. Number of lines to skip at the beginning of the radiosonde ascii file. Defaults to: 1 (1 line reserved for header info)

*Example: rsonde\_skip\_header = 6 (skip the first 6 lines)*

**rsonde\_skip\_footer:** Radiosonde parser option. Number of lines to skip at the end of the radiosonde file. Defaults to: 0 (no footer assumed)

*Example: rsonde\_skip\_footer = 10 (skip the last 10 lines)*

**rsonde\_delimiter:** Radiosonde parser option. The delimiter that separates columns in the radiosonde file choose one of:

- S: space
- C: comma

Defaults to: S.

*Example: rsonde\_delimiter = C*

**rsonde\_column\_index:** Radiosonde parser option. The index (integer) of the column that contains the Height, Pressure, Temperature, and Relative Humidity (optional) information. For example, rsonde\_columns = 1, 3, 2, 6 means:

- Height: 1<sup>st</sup> column in the radiosonde file
- Pressure: 3<sup>rd</sup> column
- Temperature: 2<sup>nd</sup> column
- Relative Humidity: 6<sup>th</sup> column

The relative humidity column is OPTIONAL and can be omitted. Defaults to: 2 1 3

*Example: rsonde\_column\_index = 2, 1, 3, 5*

**rsonde\_column\_units:** Radiosonde parser option. The units of Height, Pressure, Temperature, and Relative Humidity (optional) columns in the radiosonde file. The number of values must be the same as in **rsonde\_column\_index**. Supported units for:

- height: m\_asl (default), km\_asl, m\_agl, km\_agl
- pressure: hPa (default), Pa, atm
- temperature: K (default), C, Cx10

- relative humidity: percent (default), fraction

If the Height units are in agl (above ground level) then the station altitude must be provided in the **rsonde\_geodata**

*Example: rsonde\_column\_units = m\_asl, hPa, C, percent*

**rsonde\_geodata:** The radiosonde station latitude, longitude, and altitude. For example: rsonde\_geodata = 35.33, 25.18, 39.0 means:

- Latitude: 35.33°
- Longitude: 25.18°
- Altitude: 39.0 m

Defaults to None for all values.

*Example: rsonde\_geodata = 48.25, 11.55, 492.0*

**rsonde\_station\_name:** The name of the station where the radiosonde measurement was performed. Defaults to: None.

*Example: rsonde\_station\_name = Muenchen-Oberschleissheim*

**rsonde\_wmo\_number:** The WMO number of the radiosonde station. Defaults to: None

*Example: rsonde\_wmo\_number = 10868*

**rsonde\_wban\_number:** The WBAN number of the radiosonde station. Defaults to: None

*Example: rsonde\_wban\_number =*

## 5.2 Section: preprocessor

**vertical\_trimming:** If set to True then bins above a certain altitude (20km by default) will be removed. Can speed up computations and save RAM. Defaults to: True

*Example: vertical\_trimming = False*

**vertical\_limit:** The maximum altitude in km above which no calculations will be performed. Solar background calculations are performed prior to vertical signal trimming to enable background calculations up to the maximum signal range. Defaults to: 20km

Example: `vertical_limit = 25.0`

**channels:** Provided it to process only selected channels. By default, no channel is excluded if this variable is not provided. The channel name must follow the nomenclature of section 4.3.

Example: `channels = 0355xcar, 0355xcpr`

In this case only the these 2 channels will be processed

**exclude\_telescope\_type:** Provide it to entirely exclude all channels of a specific **telescope\_type**. By default, no channel is excluded if this variable is not provided.

Example: `exclude_telescope_type = x, n`

This will exclude all channels with x and n values provided in their **telescope\_type**, that is all near field and unspecified channels

**exclude\_channel\_type:** Provide it to entirely exclude all channels of a specific **channel\_type**. By default:

- if `isday = False` --> no channel is excluded if this variable is not provided
- if `isday = True` --> `exclude_scattering_type = v, f` (vibrational Raman and fluorescence channels excluded)

Example: `exclude_scattering_type = v, r, o, x`

This will exclude all channels with v, r, o, and x values provided in their **channel\_type**, that is all vibrational and rotational Raman, co- and cross- circular polarized channels

**exclude\_acquisition\_mode:** Provide it to entirely exclude all channels of a specific **acquisition\_mode**. By default, no channel is excluded if this variable is not provided.

Example: `exclude_acquisition_mode = a`

This will exclude all channels with 0 values provided in their **acquisition\_mode**, that is all analog channels.

**exclude\_channel\_subtype:** Provide it to entirely exclude all channels of a specific **channel\_subtype**. By default, no channel is excluded if this variable is not provided.



*Example: `exclude_channel_subtype = w`*

This will exclude all channels with w values provided in their **channel\_subtype**, that is all water vapor channels.

## 5.3 Section: quicklooks

The following variables:

- **channels**
- **exclude\_telescope\_type**
- **exclude\_channel\_type**
- **exclude\_acquisition\_mode**
- **exclude\_channel\_subtype**

where defined in section 5.2 and are also applicable here. The usage is the same with the only exception that exclusions take place in the visualization phase and only for the quicklooks

**use\_log\_scale:** If set to True, a logarithmic scale will be used for the visualization of the *z*\_axis (signal levels). It affects the selection of the default *z*\_lims. Defaults to False.

*Example: use\_log\_scale = True*

**use\_distance:** If set to False, the *y* axis units of the quicklook will correspond to the altitude in meters above sea level. If set to True, the *y* axis units corresponds to the distance from the system. This variable determines the units of the following variables:

- **y\_lims**
- **z\_max\_zone**
- **smoothing\_range**

Defaults to: True

*Example: use\_distance = False*

**x\_lims:** The *x* axis limits (lower and upper). Use two integers corresponding to the first and last temporal samples (not date!) that will be plotted in the quicklooks. Use 1 to start from the first sample. If values below 1 or above the total number of samples are used, they will be ignored

*Example: x\_lims = 100, 300*

Assuming a Rayleigh Fit measurement with 1000 samples, the quicklook with start cover the zone between the 100<sup>th</sup> and the 300<sup>th</sup>.

**x\_tick:** The x axis finest major tick in number of samples. Defaults to: automatic selection.

*Example:  $x\_tick = 2$*

**t\_lims:** The x axis limits in time units. Use the following format hh:mm for both limits (not meant to be used for day-long quicklooks). Defaults to: automatic selection.

*Example:  $t\_lims = 17:38, 21:56$*

**t\_tick:** The major tick for the time axis (same as x axis but with different units). Defaults to: automatic selection

*Example:  $t\_tick = 5$*

**y\_lims:** The y axis limits (altitude/distance) in km (lower and upper). If values below 0 or above the maximum altitude/distance are used, they will be ignored. Defaults to: 0., 14.

*Example:  $y\_lims = 0., 10.$*

**y\_tick:** The y axis finest major tick in km. Defaults to: 1km

*Example:  $y\_tick = 0.5$*

**z\_lims:** The colorscale limits of the normalized range-corrected signal normalized with the mean value in the **z\_max\_zone** region. This setting is useful if for example the normalization takes place in a very strong aerosol layer making all other layers difficult to discern. Defaults to:

- $z\_lims = 0., 1.$  when **use\_log\_scale** = False and
- $z\_lims = 1E-5, 1.$  when **use\_log\_scale** = True

*Example:  $z\_lims = 0., 0.9$  (assuming **use\_log\_scale** = False)*

**z\_max\_zone:** Provide a zone (min and max altitude/distance) in km. The maximum range-corrected signal value encountered in the zone will be used for the normalization of the range-corrected signal for the quicklook. Particularly useful in order to avoid scaling the colors with a cloud. Defaults to: 0.1, 2.

*Example:  $z\_min\_zone = 0.5, 1.5$*

**z\_min\_zone:** Provide the zone (min and max height/distance) in km. The minimum range-corrected signal value encountered in the zone will be used as lower z\_axis limit in case the **z\_lims** are not explicitly provided and **use\_log\_scale** is set to True. Defaults to 2., 10.

*Example: z\_min\_zone = 5., 8.*

**smooth:** If set to True, a sliding average smoothing filter will be applied on the signals across y axis for better visualization. The **smoothing\_exponential**, **smoothing\_exponential**, and **smoothing\_window** will be ignored if **smooth** is set to False. Defaults to: False

*Example: smooth = True*

**smoothing\_range:** Set the first and last altitude/distance boundaries in km where smoothing should be applied. If they exceed the actual signal boundaries the actual boundaries will be used instead. Defaults to: 0.05, 14.

*Example: smoothing\_range = 0.02, 10.*

**smoothing\_window:** The smoothing window in the first and last bin of the smoothing region, in m. The widow progressively changes between from the first to the last value. Use the same value twice to apply a constant window. Defaults to: smoothing\_window = 50., 500.

*Example: smoothing\_window = 5., 300.*

**smoothing\_exponential:** This variable is ignored if the upper and lower values of **smoothing\_window** are the same. Choose one of:

- True: a smoothing window that exponentially increases with altitude/distance will be applied.
- False: a smoothing window that exponentially increases with altitude/distance will be applied

Defaults to: True.

*Example: smoothing\_exponential = False*

**dpi:** The dots per inch (dpi) resolution of the exported figures. Defaults to: 300 dpi

*Example: dpi = 100*

**color\_reduction:** If set to True, and provided that Imagemagick is installed, an adaptive color reduction will be applied to save space when exporting the figures. Defaults to True. A warning will be raised if Imagemagick is not installed.

*Example: color\_reduction = False*

## 5.4 Section: rayleigh\_fit

The following variables:

- **channels**
- **exclude\_telescope\_type**
- **exclude\_channel\_type**
- **exclude\_acquisition\_mode**
- **exclude\_channel\_subtype** (b, s, a, w, c)

where defined in section 5.2 and are also applicable here. The usage is the same with the only exception that exclusions take place in the visualization phase and only for the Rayleigh fit test. Respective default values are provided in parenthesis.

**use\_distance:** If set to False, the x axis units of the Rayleigh fit will correspond to the altitude in meters above sea level. If set to True, the x axis units corresponds to the distance from the system. This variable determines the units of the following variables:

- **normalization\_range**
- **x\_lims**
- **smoothing\_range**

Defaults to: True

*Example: use\_distance = False*

The following variables:

- **smooth** (True)
- **smoothing\_range** (0.050, 15.)
- **smoothing\_window** (100., 100.)
- **smoothing\_exponential** (False)
- **dpi** (300)
- **color\_reduction** (True)

where defined in section 5.3 and are also directly applicable here. Respective default values are provided in parenthesis.

**use\_lin\_scale:** If set to True, a linear scale will be used for the y axis (signal axis). If set to False a logarithmic scale will be applied instead. Defaults to False:

Example: `use_lin_scale = True`

**normalization\_region:** The lower and upper limits of the region used for normalizing the signal in the Rayleigh fit. If `use_distance` is called, the limits correspond to distance. Defaults to: 8.5, 9.5

Example: `normalization_region = 4., 6.`

**x\_lims:** Set the x axis (altitude/distance) limits in km (lower and upper). If values below 0 or above the maximum altitude/distance are used, they will be ignored. Defaults to: 0., 14.

Example: `x_lims = 0., 24.`

**x\_tick:** The x axis finest major tick in km. Defaults to: 1km

Example: `x_tick = 2.`

**y\_lims:** The y axis (signal) limits (lower and upper) of the normalized range-corrected signal in  $\text{m}^{-1} \text{sr}^{-1}$  (pseudo-units). It is recommended to skip this variable and use the automatic selection because channels in different wavelengths have different optimal limits. Defaults to: automatic selection

Example: `y_lims = 5E-8, 5E-5`

## 5.5 Section: telecover

The following variables:

- **channels**
- **exclude\_telescope\_type**
- **exclude\_channel\_type**
- **exclude\_acquisition\_mode**
- **exclude\_channel\_subtype**

where defined in section 5.2 and are also applicable here. The usage is the same with the only exception that exclusions take place in the visualization phase and only for the Rayleigh fit test

The following variables:

- **use\_distance** (True)
- **x\_lims** (0., 2.4)
- **x\_tick** (0.2)
- **y\_lims** (automatic selection)
- **normalization region** (1.8, 2.2)
- **smooth** (True)
- **smoothing\_exponential** (False)
- **smoothing\_range** (0.50, 2.5)
- **smoothing\_window** (100., 100.)
- **dpi** (300)
- **color\_reduction** (True)

where defined in section 5.4 and are also applicable here. The default values are not always the same and are provided in parenthesis. Note that:

- The **use\_distance**, **use\_lin\_scale**, **x\_lims** and **x\_tick** variables correspond to the x axis that is the same for all the 3 subplots of the telecover test.
- The smoothing options affect only the 2<sup>nd</sup> and 4<sup>th</sup> subplot of the telecover test (normalized signals in the near range and their respective deviations from the mean)



- The **y\_lims** variable units should be the same with the units of the 2<sup>nd</sup> telecover subplot (normalized signals in the near range).

**use\_non\_rangecor:** If set to True, the non range corrected signals will be used for the left subplot of the telecover test. If set to False the range corrected signals will be used instead. Defaults to: False

*Example: use\_non\_rangecor = True*

**use\_last:** If set to True an additional black line with the difference between the first and the last sector is added in the 4<sup>th</sup> subplot (deviations) of an interleaved telecover test e.g.  $N2 - N1$ . If the telecover is not interleaved the difference line will be visible even if **use\_last** is set to False.

*Example: use\_last = True*

## 5.6 Section: polarization\_calibration

The following variables:

- **use\_distance** (True)
- **dpi** (300)
- **color\_reduction** (True)

where defined in section 5.4 and are also applicable here. The default values are not always the same and are provided in parenthesis.

The following variables:

- **smooth\_calibration** (True)
- **smoothing\_exponential\_calibration** (False)
- **smoothing\_range\_calibration** (1., 15.)
- **smoothing\_window\_calibration** (100., 100.)

operate in exactly the same way with the smoothing options defined in section 5.3. The only difference is that smoothing is applied only on the signals used for the ratios in the calibration subplot (left) of the polarization calibration figures.

The following variables:

- **smooth\_rayleigh** (True)
- **smoothing\_exponential\_rayleigh** (False)
- **smoothing\_range\_rayleigh** (1., 14.)
- **smoothing\_window\_rayleigh** (100., 100.)

operate in exactly the same way with the smoothing options defined in section 5.3. The only difference is that smoothing is applied only on the signals used for the ratios in the Rayleigh subplot (right) of the polarization calibration figures.

**ch\_r**: Provide here channel names from the available ones with the reflected **channel\_subtype** (e.g. 0355xpar). The number of reflected channels must be the same as the number of the respective transmitted channels provided by **ch\_t** and by the GHK parameters. By default, all available reflected channels will be matched to all available transmitted channels that share the same telescope type,

detection mode, and wavelength. WARNING! The field is mandatory if non-default GHK values are applied.

*Example: `ch_r = 0355xppr, 0532xcpr`*

**ch\_t:** Provide here channel names from the available ones with the transmitted **channel\_subtype** (e.g. 0355xcat). The number of reflected channels must be the same as the number of the respective transmitted channels provided by **ch\_r** and by the GHK parameters. By default, all available reflected channels will be matched to all available transmitted channels that share the same telescope type, detection mode, and wavelength. WARNING! The field is mandatory if non-default GHK values are applied.

*Example: `ch_t = 0355xcpt, 0532xppt,`*

**calibration\_region:** The lower and upper limits of the region used for  $\Delta 90$  calibration. If `use_distance` is called, the limits correspond to distance. Defaults to: 2., 4. km

*Example: `calibration_region = 3., 4.`*

**rayleigh\_region:** The lower and upper limits of the region used for the comparison with the Rayleigh atmosphere. Defaults to: 8.5, 9.5

*Example: `rayleigh_region = 8.5, 9.5`*

**x\_lims\_calibration:** The x axis (altitude/distance) limits in km (lower and upper) for the pol. calibration subplot. If values below 0 or above the maximum altitude/distance are used, they will be ignored. Defaults to: 0, 15

*Example: `x_lims_calibration = 0.1, 6.`*

**x\_lims\_rayleigh:** The x axis (altitude/distance) limits in km (lower and upper) for the Rayleigh comparison subplot. If values below 0 or above the maximum altitude/distance are used, they will be ignored. Defaults to: 0, 15

*Example: `x_lims_rayleigh = 0.1, 8.`*

**x\_tick\_calibration:** The x axis finest major tick in km for the pol. calibration subplot. Defaults to: 2.

*Example: `x_tick_calibration = 1.`*

**x\_tick\_rayleigh:** The x axis finest major tick in Km for the Rayleigh plot. Defaults to: 2.

*Example:  $x\_tick\_rayleigh = 1$ .*

**y\_lims\_calibration:** The y axis limits (lower and upper) of the gain ratios at  $\pm 45^\circ$ . Used for the pol. calibration subplot. Defaults to: automatic selection

*Example:  $y\_lims\_calibration = 0, 0.1$*

**y\_lims\_rayleigh:** The y axis limits (lower and upper) of the volume depolarization ratio. Used for the Rayleigh subplot. Defaults to: automatic selection

*$y\_lims\_rayleigh = 0, 0.1$*

**K:** The K value for each channel pair. Defaults to: 1 for all channels

*Example:  $K = 1.05, 1$ .*

**G\_R:** The G value for the reflected channel of the pair. Defaults to: 1 for all channels (assuming no correction for the emission and the receiver)

*Example:  $G\_R = 1., 1$ .*

**G\_T:** The G value for the transmitted channel of the pair. Defaults to: 1 for all channels (assuming no correction for the emission and the receiver)

*Example:  $G\_T = 1., 1$ .*

**H\_R:** The H value for the reflected channel of the pair. Defaults to:

- 1 for all co-polar (p) reflected channels
- -1 for all cross-polar (c) reflected channels

(assuming no correction for the emission and the receiver)

*Example:  $H\_R = 0.94, -0.98$*

**H\_T:** The H value for the transmitted channel of the pair. Defaults to: 1 or -1 for all co-polar (p) and cross-polar (c) transmitted channels, respectively (no receiver optics + emitted pcb. state correction)

- 1 for all co-polar (p) transmitted channels
- -1 for all cross-polar (c) transmitted channels

(assuming no correction for the emission and the receiver)

*Example:  $H_T = -0.94, 0.98$*

**R\_to\_T\_transmission\_ratio:** The transmission ratio of the R to T channels per pair ( $T_R/T_T$ ).  $T_R = 1$  and/or  $T_T = 1$  if no filter was applied during calibration. Defaults to: 1 for all pairs

*Example:  $R\_to\_T\_transmission\_ratio = 8., 0.15$*

## 6 The radiosonde files

This section includes some more details about the radiosonde files and how ATLAS is handling them. Please refer also to section 5.2 and specifically to the following variables:

- **rsonde\_skip\_header**
- **rsonde\_skip\_footer**
- **rsonde\_delimiter**
- **rsonde\_column\_index**
- **rsonde\_column\_units**
- **rsonde\_geodata**
- **rsonde\_station\_name**
- **rsonde\_wmo\_number**
- **rsonde\_wban\_number**

The file format is general ascii. The parser is flexible and can handle headers and footers, two different delimiters (comma and space), custom column order, and custom units.

### The Radiosonde folder

This folder should contain the radiosonde files. At this stage only ascii files can be parsed. Note that more than one files can be provided here. ATLAS scans all files and selects the one that it is closest in time with the Rayleigh measurement. Depending on the application, it can be more convenient to use a specific radiosonde folder where all radiosonde data is included and provided it as a command line argument when running ATLAS. Note that the time difference between the Radiosonde and the lidar measurement **cannot be larger than 18 hours**.

### Filename Format

The name of the radiosonde file must follow a specific format so that it is recognized by ATLAS. It must start with the date and end with .txt Whatever comes after the date and before the extension is not important for parsing but can be helpful for sorting purposes for the user:

**yyyymmdd\_hhmm**<add\_whatever\_you\_want\_here>.**txt**

Example: 20230219\_1200\_Munich.txt

This radiosonde file corresponds to 19.02.2023 at 12:00 UTC.