OzFlux Standard QA/QC Instructions

Level 1 (L1): From Spreadsheet to nc file

What happens at this stage of processing?

Data from the logger is ingested into the program and output and saved as an nc file (i.e. transform an excel file to an nc file). The contents and structure of the nc file are configured in 'control files' which need to be customized by the user before processing.

Before Processing: Setting up an L1 control file

Before processing, you will need to customize your _control files_. Control files are text files that essentially 'control' how the program reads information. The Level 1 control file dictates how the datalogger data is read and stored in the nc file. For example, in the L1 control file you will need to specify the location of the excel data (_See Fig 1._), the name of the spreadsheets where the data will be taken from _(See Fig 2)_, the variable names from column headings of the data and what they will be renamed to in the nc file. As this system is customized, it is easy for users with diverse 'raw' datalogger sets to streamline them into a standard 'OzFlux' template, which is the same across all sites.

```
[Files]
    [[L1]]
        xlFilePath = ../../Sites/HowardSprings/2012/ ——— File path for 'raw' excel data
        xlFileName = HowardSprings_2012_L1_20120215.xls ——— Name for excel data
        xlHeaderRow  = 2 •—Row number where variable names are stored
        xl1stDataRow = 5 •—Row number where data first begins
        ncFilePath = ../../Sites/HowardSprings/2012/ ——— File path for generated 'nc' file
        ncFileName = HowardSprings_2012_L1.nc ——— Name for 'nc' file
```

Fig 1. Example of part of the top section of an L1 control file with predefined file paths

```
[Variables]
    [[xlDateTime]]                                    nc file variable name
        [[[xl]]]
            Sheet     = Flux                          Excel spreadsheet name
            Name      = TIMESTAMP                      Excel variable name
        [[[Attr]]]
            Description = "Date/time in Excel format"  Excel data attributes
            Units     = "days"                         (description of what the data is)
    [[Fn_NR]]
        [[[xl]]]
            Sheet     = EBal
            Name      = Rn_Avg
        [[[Attr]]]
            Description = "Net radiation (NRLite)"
            Units     = "W/m2"
    [[Fld]]
        [[[xl]]]
            Sheet     = EBal
            Name      = Fld
        [[[Attr]]]
            Description = "Down-welling long wave (Kipp&Zonen)"
            Units     = "W/m2"
```

Fig 2. Excel variables are defined in an L1 control file, including they should be renamed in an nc file

How to:

1. Go to 'console' and type 'python qc.py'. Ensure your start up directory (Edit – Settings – Startup dir) is directed to the folder with your script files. This will bring up a GUI (see below).
2. Click 'Load Datalogger data' (see below, pink).
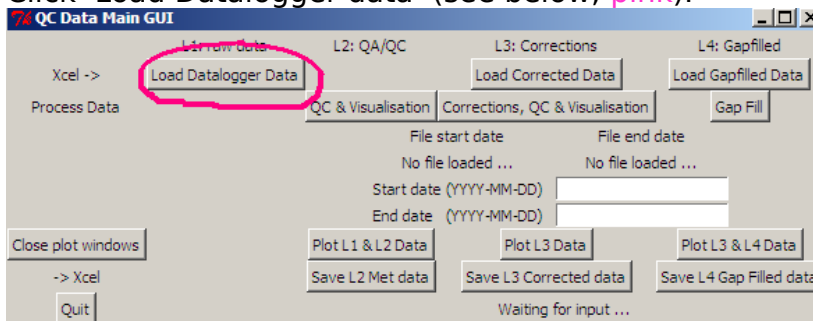
3. A new window will appear asking you to open a file. Navigate to your L1 control file and click 'open'. In the console window, you should be able to see the data being read and ingested by the program. Check that all data reads 'INFO' *(See Fig 3)*. If data comes back reading 'ERROR' this is because it couldn't find the data. You may need to go back to you control file and check that nothing was misspelt (including uppercase/lowercase of words).

Fig 3. Ensure data reads 'INFO' in the console window

Level 2 (L2): Adding Range Checks, Diurnal Checks, Excluding Dates, Excluding Hours

What happens at this stage of processing?

It is important to remember that the data was converted from an excel to nc file in the previous L1 stage – this data has not been quality controlled. In L2, the L1 nc file is run through a number of quality control checks which are specified and customised by the user through an L2 control file. The advantage of this process is having a systematic method of controlling values (not doing it manually).

Valid quality control sections are:

RangeCheck: reject data less than "Lower" or greater than "Upper"
DiurnalCheck: reject data outside $Avg(H.h) +/- NumSd*Stdev(H.h)$
ExcludeDates: reject data between specified dates
 ExcludeHours: reject data at specified times between the specified date range
 Linear: apply a linear correction to the data within a date range,
 Correction: apply an offset to a wind direction measurement to correct for alignment

Before Processing: Setting up an L2 control file
As was done L1, you will need to set up an L2 control file including the locations of data, what you want the names of the data saved as etc. During this stage of the QA/QC process, you will need to specify numerical ranges for the data. The basic quality control during this stage is _Range Check_: upper and lower limits considered to be reasonable for any given variable are specified. Data within these user defined ranges is kept and data outside these ranges is rejected. Rejected data are 'flagged' with an associated number. This flag value describes why the data was rejected (i.e. 1: missing from L1 spreadsheet, 2: failed range check, 5: failed diurnal check etc.). These flag values are available on a separate word document.

After the L2 process is complete, there is the first opportunity to view the data visually through graphs. These are described below.

How to:

1. The GUI should still be open. Click 'QC & Visualisation' (see below, pink).
2. A new window will appear asking you to open a file. Navigate to your L2 file (which you should have already customised) and click open. In the console window, you should be able to see the data being read and ingested by the program. Check that all data reads 'INFO' .
3. To visualise data, click 'Plot L1 & L2 Data' (see below, blue). Plots will appear (as defined towards the bottom of the L2 control file). You will note that there are two forms of data plotted on these graphs (See Fig 4.)
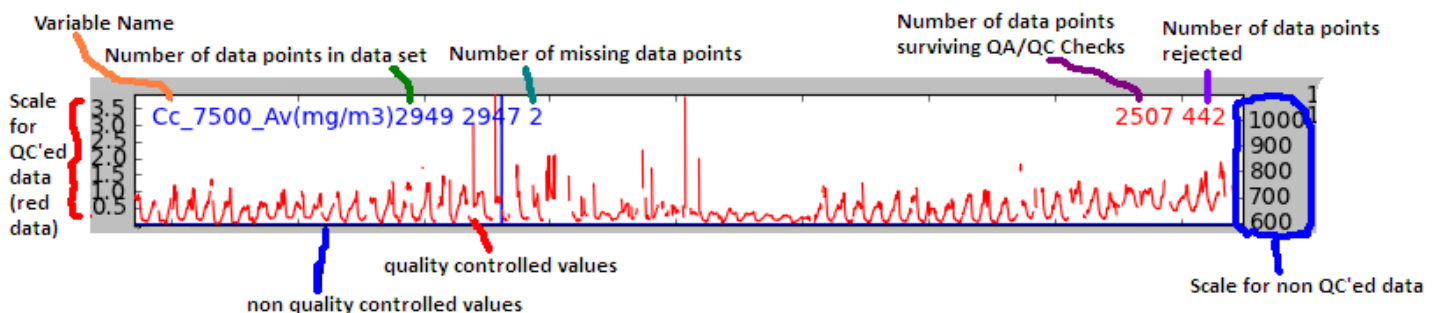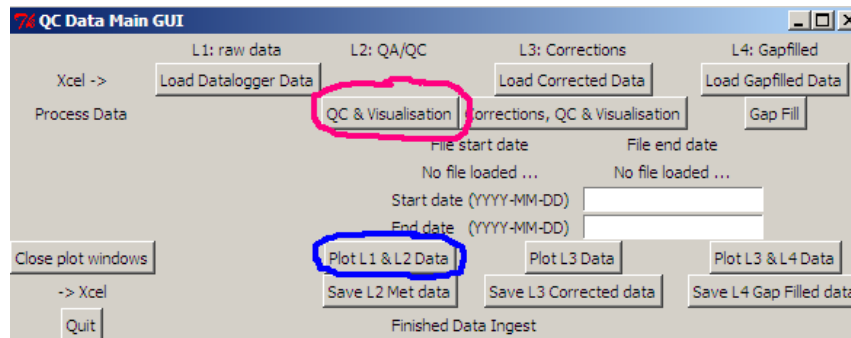

Fig 4. Archetypal graph for any given variable (in this case Cc_7500_Av).

The red data is the quality controlled data, and it's associated scale on the right. The blue data is non-quality controlled data, and its associated scale on the left.

The graphs to the right of the major graphs (See Fig 5.) include a diurnal graph and 'flagged data' histogram for the specified period. The diurnal graph (red) represents the *maximum*, green the *average* and blue the *mimimum* values measured averaged over the given time period. The histogram on the right is a quick way to diagnose how much and why the data was rejected (if applicable). In *Figure 5,* the histogram details that over 400 data points (see scale) have been rejected because of data flag 4 (i.e. values failed LI-7500 diagnostic check).
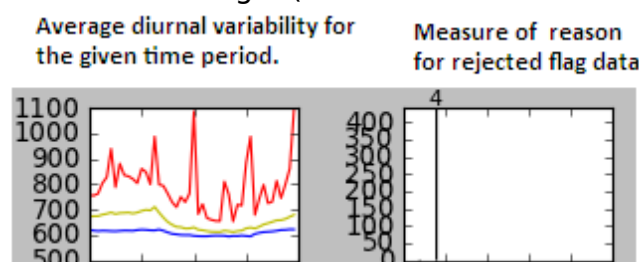

4

Fig 5. Ancillary graphs to the right of the main graph showing average daily ranges (left) and flagged data points (right)

Level 3 (L3): Merging like-data and range checking

What happens at this stage of processing?

By this stage, you will have quality controlled data for each variable. On most towers, there are many instruments that measure the same thing (i.e. HMP and CSAT both measure temperatures and humidities). Some instruments, for some reason or another, drop out occasionally (i.e. CSAT and Li-7500 are frequently disturbed due to rain). Consequently, there will be times where one instrument is measuring and the other isn't. Therefore, it is necessary to merge these two data into the same series in an attempt to fill any gaps. Once merged, QA/QC checks (range checks, diurnal checks etc.) are applied once again to ensure that the resultant merged data is quality controlled.

Before Processing: _Setting up an L3 control file_
As was done in the L1 and L2 control files, you will need to go through the same process of defining file paths and file names. You will note that you have the ability to merge two data series at this level. 'MergeSeries' _(See Fig 6)_ allows the user to group 'like' data and merge them together. When merging, it is important to have the 'Source' data arranged in order of preference; the program will use the first variable (Ah_HMP_02), if there are gaps in the first source of data, it will attempt to fill these gaps with your second preference (Ah_7500_Av) , so on (Ah_HMP_01) and so forth.

Fig 6. An example of L3 control file defining 'merge series'

Range checks, diurnal checks etc. are then applied to the data.

Other checks that can be done at this stage of analysis includes:
L3 check: INSERT WHAT L3 check does
Fingerprint analysis: rectangular plots of variables which colour shows the value, time is in x-axis and DOY is in y-axis. Fingerprint graphs are useful to visually address patterns of the changes of the variables throughout the year.

How to:

    1. The GUI should still be open. Click 'Corrections, QC & Visualisation' (see below, pink).

2. A new window will appear asking you to open a file. Navigate to your L3 file (which you should have already customized) and click open. In the console window, you should be able to see the data being read and ingested by the program.

3. To visualise data, click 'Plot L3 Data' (see below, blue). Plots will appear (as defined towards the bottom of the L3 control file). Unlike the L2 plots, the L3 plots only show quality controlled data (all in red). You will notice that the scale on the left and right hand side of the graphs are the same. Diurnal graphs and QC Flag graphs still exist to the right hand side of these graphs.
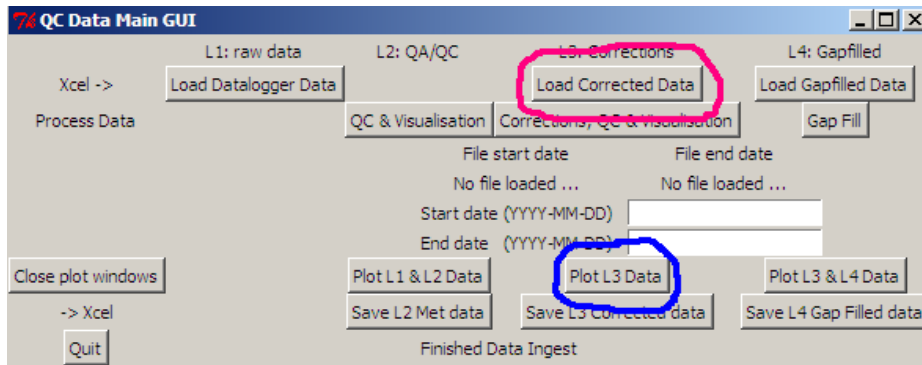
4. Type 'l3check.py'
5. Type 'fingerprint.py'

Level 4 (L4): Gap-filling

What happens at this stage of processing?

Once L3 has been completed, what we are left with is *real* field measurements that have been quality controlled. However, during instances where instruments have failed, and there was no redundancy in measurements, we are left with unavoidable gaps in the data. This requires us to fill the gaps using proxy data. The data gaps can be inferred by various means including radiation, meteorology, or in some cases, some data can be filled using data from other nearby flux towers (i.e. savanna sites).

Before Processing: *Setting up an L4 control file*
Insert here
How to:

1. Insert here