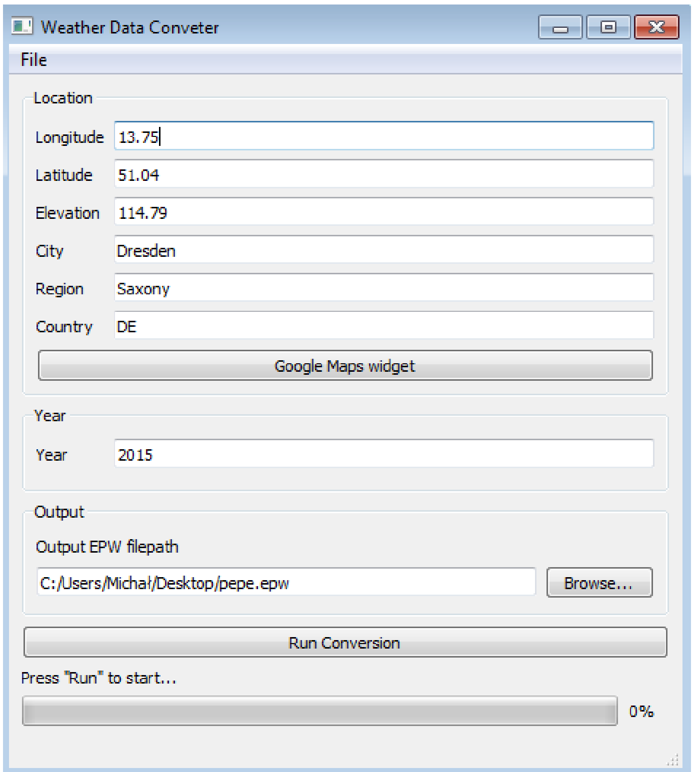
Weather Data Converter

User manual



Michal Marek Drzazga

Technische Universität Dresden
Faculty of Architecture
Institute of Building Climatology

Dresden, April 2018

Table of Contents

1 Introduction

1.1 What is Weather Data Converter?

Weather Data Converter is a software tool developed at the Institute of Building Climatology of Technical University of Dresden. It automatically fetches the hourly climate data for a given location provided by the DWD – Deutscher Wetterdienst on its FTP server. Then it converts it to a EPW file format that is supported by programs for building physics analysis such as Therakles, Delphin, as well as other programs that use the EnergyPlus core. This is done according to [1].

1.2 What is the goal of this user manual?

The goal of this manual is to provide the user with information on the structure of the interface, but most importantly to describe the underlying processes of the data conversion that the program undertakes.

2 Getting and running the program

2.1 Getting the program

The program is available as the source code at the GitHub platform. It can be accessed either via website (https://github.com/drzamich/WeatherDataConverter) and downloaded as the .zip file or using the git command:

```
git clone https://github.com/drzamich/WeatherDataConverter
```

2.2 Running the program

The program has been written and tested in the programming language Python, version 3.6.3. To run it, user has to have the compiler installed on their machine. Furthermore, the following libraries have to be installed in order for the program to compile:

- PyQt5
- pvlib
- numpy
- pandas
- PrettyTable

In order for the Google Maps widget, that makes the input of the location easier, to run, user has to put his own Google Maps API key into the code. In the 13th line of the map.html file, the string *<YOUR API KEY>* has to be replaced with the personal API key, that can be obtained at [2].

After getting the source code, to run the program, run the script Main.py.

3 Structure of the user interface

The program is a simple application with a main window and two secondary dialogs.

3.1 Main window

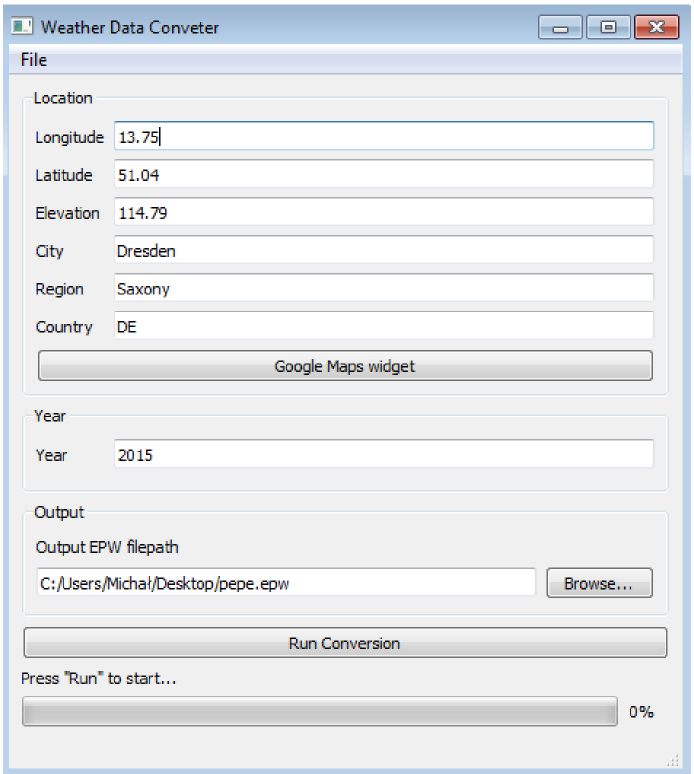


Figure 3.1.1: Main Window of the program

In the Main Window, user can define the most important parameters required for the data conversion process. The location defined by the longitude, latitude and the year for which the data is needed. Other parameters: elevation, city, region and country are arbitrary and are saved in the final output file but don't affect the conversion process itself.

In the middle part user can set the output path of the .epw file. The output directory can be set in the settings so that it's not necessary to define it with every run of the program.

At the bottom of the window, the "Run Conversion" button initializes the conversion process. The progress bar underneath notifies user of its current stage.

3.2 Map widget

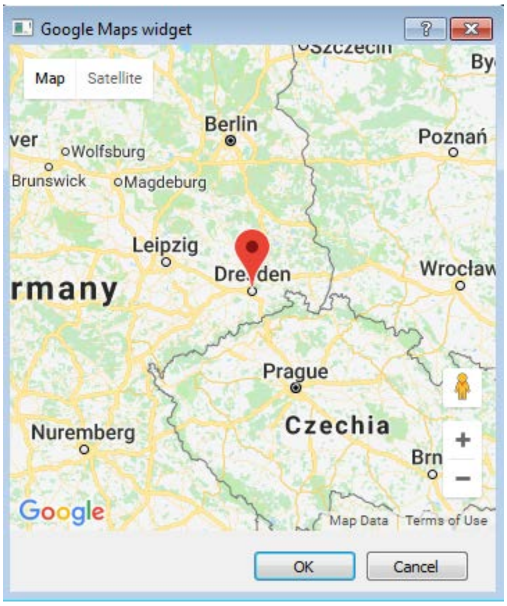


Figure 3.2.1: Google Maps widget

The Google Maps widget makes it easier for the user to input the location data in the Main Window fields. After placing a marker in the specific location and clicking “OK”, the fields in the Main Window are updated automatically. Please note, that ‘City’ and ‘Region’ fields may not be correct for locations outside of Germany. They are, however, not important for the conversion process and may be corrected manually.

The internet connection is required in order for the Google Maps widget to work properly.

3.3 Settings dialog

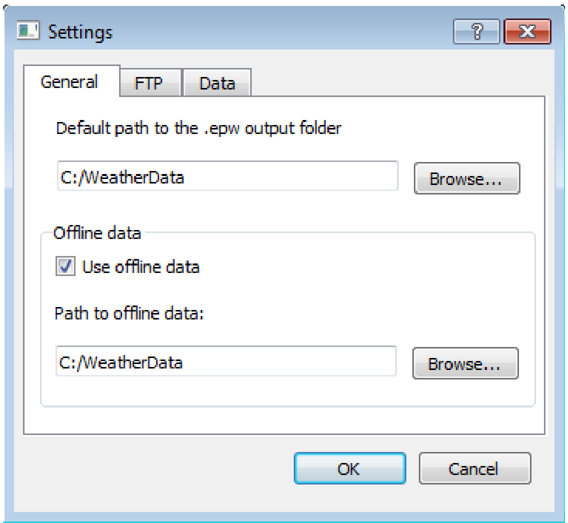


Figure 3.3.1: Settings dialog, *General* tab

In the settings dialog, accessible from the Main Window -> File tab -> Settings or using *Ctrl+S* keyboard shortcut, user can define various parameters of the conversion process. It is divided in three tabs.

On the *General* tab, the general parameters such as default .epw file output path and path to offline data is defined. User can also set here whether the program should fetch data from the FTP server or use the data stored on the hard drive.

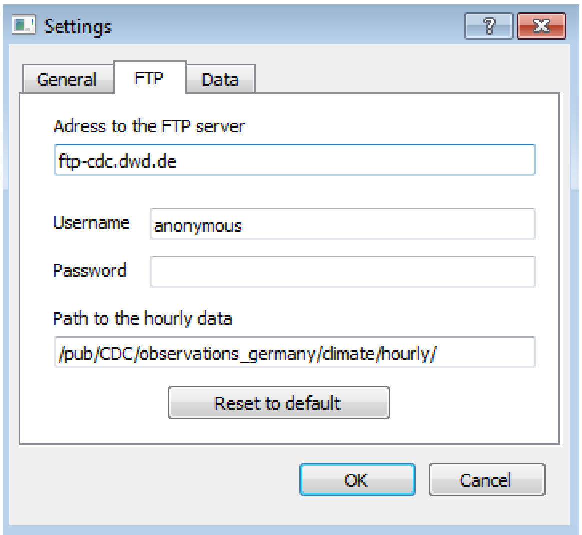


Figure 3.3.2: Settings dialog, *FTP* tab

On the *Settings* tab, the information about connection with the DWD FTP server is stored. The fields here should not be edited, unless the user really knows what to do. Incorrect data can be reset to defaults upon clicking the “Reset to default” button.

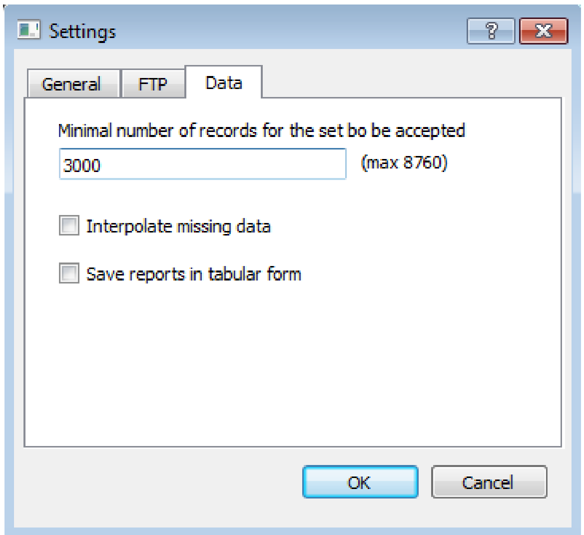


Figure 3.3.3: Settings dialog, *Data* tab

On the *Data* tab the parameters important for the conversion process itself are defined. The *minimal number of records…* controls the acceptance limit of the data set. The following check boxes control whether the missing values are to be exchanged with interpolated values or not and if the files with data in the reports directory are to be saved in the tabular form (takes longer computational time).

4 Conversion process

The conversion process, based on the input parameters: longitude, latitude and year, fetches the weather data provided by the DWD and converts it to the .epw file format, saving easy-to-evaluate report files for checking purposes along.

4.1 Localization and structure of the weather data

The program uses hourly weather data stored on the DWD FTP server, marked as *historical,* meaning that is has been checked for errors. The data is subdivided in eight categories – one for each climate element – that are stored in the sub-folders:

- air_temperature – for air temperature and relative humidity data
- cloudiness – for cloudiness data
- precitipation – for precipitation data
- pressure – for pressure data
- soil_temperature – for soil temperature data
- solar – for radiation data
- sun – for sunshine duration data
- wind –for wind speed and direction data

The data can also be downloaded from the server and stored on the hard drive in order to allow the running of the program without internet connection. The localization of the offline data is to be set in the settings window. The folder path has to point at the directory storing the eight sub-folders.

The location of the weather data, whether online of offline, will be in the following parts of the manual referenced to as *Repository*.

4.2 Creating station set

The first step in the conversion process is to choose the weather stations from which the weather data will be taken. The deciding factor is the distance from the station to the specified location. As there is a different number of stations collecting data for each climate element, it can happen that the data for a location will be taken from eight different stations.

Firstly, the program enters a sub-folder and opens the .txt file storing stations characteristics such as longitude, latitude, name, ID, starting and ending dates of operation etc. Based on that, it creates a preliminary list of stations. Then it scans the sub-folder for .zip files containing weather data and compares the list of files with the stations list. As the .zip files have stations ID stored in them, program deletes from the stations list those for which there is no .zip available. Finally, using the Haversine formula, it calculates the distance from the specified location to the station, taking into consideration only stations operating in the defined year. The station for which the distance is lowest, is in the end saved. The process is repeated eight times, once for each climate element.

In choosing the stations, program does not take into consideration those listed in the forbidden_stations.txt file in the *programdata/* directory. If for whatever reason user wants to exclude a station from search, they can add a new line to the file containing the following syntax:

```
year        station ID      climate element
```

with <TAB> as delimiter. It's possible to exclude a station for all years (and/or) all climate elements by writing `all` in the respective field.

4.3 Reading data

When the set of stations is defined, program enters each sub-folder of the Repository and opens the .zip file corresponding to the chosen station. Then it opens the .txt file storing weather data inside. It deletes all the entries that do not correspond to the specified year, as well as columns that do not store important data:

- for air_temperature: all but 3rd and 4th
- for cloudiness: all but 4th
- for precitipation: all but 3rd
- for pressure: all but 3rd and 4th
- for soil_temperature: all but from 3rd to 8th
- for solar: all but from 3rd to 7th
- for sun: all but 3
- for wind: all but 3rd and 4th

The data is saved as the list object, making it easier to handle in the following steps.

4.4 Converting and calculating data

Extracted raw data needs to be converted and recalculated before writing to the output .epw file.

4.4.1 Removing duplicates and inserting missing dates

Duplicated entries are removed from the set. Hours which are missing from the set are added with values in all the data columns marked as missing. If the number of existing entries does not exceed the minimal number defined in settings, the process is interrupted to choose new station for the specific climate element. The station for the current year and climate element is added to the forbidden list and will not be taken into consideration in the future conversion processes.

4.4.2 Data interpolation

If the user wants the missing values to be filled, a process of interpolation is undertaken. There are two interpolation approaches used in the conversion process:

a) direct interpolation – missing values are calculated based on the nearest neighboring available data so that there is a smooth transition with no gaps in the data set. This approach is applied for the following climate elements:
    - cloudiness
    - precipitation
    - pressure
    - wind
b) interpolation by average – missing value is an average of values in a +24h and -24h time distance from the missing one This approach is applied for the following climate elements:
    - air temperature
    - relative humidity
    - soil temperature
    - solar radiation

4.4.3 Stripping leap year

If the year for which the conversion process is made is a leap year (there are 8784 records in a data set), it's modified so that it has as many records as the normal year. The records after 28th of February get one day added to their time stamp (for example 10:00 29th February 2016 becomes 10:00 1st March 2016 etc.). The last 24 entries are removed from the data set.

4.4.4 Calculating air temperature data

Based on the dry bulb temperature and relative humidity, the dew point temperature is calculated using the equation obtained from [3].

4.4.5 Calculating solar radiation data

Radiation parameters required by the EPW format that are of highest importance for building physics simulations are Direct Normal Radiation and Diffuse Horizontal Radiation. Both in [W/m2]. Data from weather stations provide us with hourly sums of diffuse solar radiation and hourly sums of total solar incoming radiation with respect to the horizontal plane, in [J/cm^2*h].

Firstly the conversion to the appropriate units has to be made. This is done by multiplying the original value with the factor of $\frac{10\,000}{3\,600}$. The value of Diffuse Horizontal Radiation is then known.

The value of Direct Normal Radiation is calculated using the *disc* method of *irradiance* module of *pvlib* library – a community supported tool that provides a set of functions and classes for simulating the performance of photovoltaic energy systems [3]. The method uses the DISC model to calculate Direct Normal Irradiance taking Global Horizontal Irradiance, zenith angle and day of the year as input parameters [4].

4.4.6 Converting cloudiness data

Cloudiness data obtained from the stations is given in oktas. The EPW format requires the cloudiness data to be given in tenths. The conversion follows the pattern shown in the table below:

okta [1/8]	**tenth [1/10]**
0	0
1	1
2	2
3	4
4	5
5	6
6	7
7	9
8	10

Figure 4.4.6.1: Pattern for conversion oktas to tenths

4.4.7 Converting wind data

Converting wind direction units from Grad to degrees by multiplying by 0.9 factor.

4.4.8 Calculating horizontal infrared radiation intensity

Based on the dew point temperature calculated in a previous step and sky cover (cloudiness) data, values of horizontal infrared radiation intensity are calculated based on the equation given in [4].

Besides conversion and calculation steps mentioned above, all the values are converted to the units required by the EPW file format.

4.5 Writing the .epw file

When all the values are calculated, the program continues to put them all together in an EPW file.

4.5.1 EPW file header

Firstly the header lines are prepared. The information about the location is given in this line. The Design Conditions header is left blank, as well as Typical/Extreme Periods header. Then the average ground temperatures for every first day of all months of the year are calculated and written in the header with values of soil conductivities, soil densities and soil specific heats left blank.

In the comment section of the header, information about set of stations used for data extraction is saved.

4.5.2 Hourly weather data

Finally, 8760 lines with weather data for each hour of the year are written in the file. Each line consists of 35 fields separated by comas, containing information about weather conditions. Not all of the fields are filled, as some are not used in Energy Plus calculations or are not possible to determine basing on the data from the weather stations.

Field number	Field name	Provided?
1	Year	Yes
2	Month	Yes
3	Day	Yes
4	Hour	Yes
5	Minute	Yes
6	Data Source and Uncertainty Flags	Yes
7	Dry Bulb Temperature	Yes
8	Dew Point Temperature	Yes
9	Relative Humidity	Yes
10	Atmospheric Station Pressure	Yes
11	Extraterrestial Horizontal Radiation	No
12	Extraterrestrial Direct Normal Radiation	No
13	Horizontal Infrared Radiation Intensity	Yes
14	Global Horizontal Radiation	No
15	Direct Normal Radiation	Yes
16	Diffuse Horizontal Radiation	Yes
17	Global Horizontal Illuminance	No
18	Direct Normal Illuminance	No
19	Diffuse Horizontal Illuminance	No
20	Zenith Luminance	No
21	Wind Direction	Yes
22	Wind Speed	Yes

23	Total Sky Cover	Yes
24	Opaque Sky Cover	No
25	Visibility	No
26	Ceiling Height	No
27	Present Weather Observation	No
28	Present Weather Codes	No
29	Precipitable Water	No
30	Aerosol Optical Depth	No
31	Snow Depth	No
32	Days Since Last Snowfall	No
33	Albedo	No
34	Liquid Precipitation Depth	No
35	Liquid Precipitation Quantity	No

Figure 4.5.2.1: Listing of fields provided in the EPW file

For existing data, the uncertainty flags are modified according to the following pattern:

Field number	Field name	Data Source Flag - missing data	Data Source Flag - existing data	Uncertainty flag
7	Dry Bulb Temperature	?	A	7
8	Dew Point Temperature	?	E	7
9	Relative Humidity	?	A	7
10	Atmospheric Station Pressure	?	A	7
13	Horizontal Infrared Radiation Intensity	?	E	0
14	Global Horizontal Radiation	?	A	0
15	Direct Normal Radiation	?	D	0
16	Diffuse Horizontal Radiation	?	D	0
17	Global Horizontal Illuminance	?	A	0
18	Direct Normal Illuminance	?	A	0
19	Diffuse Horizontal Illuminance	?	A	0
20	Zenith Luminance	?	A	0
21	Wind Direction	?	A	0
22	Wind Speed	?	A	0
23	Total Sky Cover	?	A	0
24	Opaque Sky Cover	?	A	0
25	Visibility	?	A	0
26	Ceiling Height	?	A	0
27	Present Weather Observation	?	A	0
28	Present Weather Codes	?	A	0
29	Precipitable Water	?	A	0
30	Aerosol Optical Depth	?	A	0
31	Snow Depth	?	A	0
32	Days Since Last Snowfall	?	A	0

Figure 4.5.2.2: Pattern for data source and uncertainty flags.

The file is saved to the path specified by the user.

4.6 Creating reports

Upon finishing the conversion process, program generates files that are saved in the *reports/* directory in the folder which name corresponds to the time of finishing the conversion process. Files stored there allow user to check the correctness of the output data. In the *00_report.txt* file the list of stations used for data extraction, as well as number of missing values for each climate element is saved. In the file *00_missing_values.txt* time stamps of entries, where missing values occurred for each climate element, are saved. There are also two folders created: *raw_data* and *converted_data.* The first one consists of files with listed values extracted from the weather stations. User can see which of those were missing – they are marked as -999. In the *converted_data* directory the interpolated and additionally calculated values are saved. When stored in tabular form, the names of columns starting with (r) indicate, that the field value is taken raw from the station (interpolated when missing).

5 References:

[1] https://bigladdersoftware.com/epx/docs/8-3/auxiliary-programs/energyplus-weather-file-epw-data-dictionary.html#energyplus-weather-file-epw-data-dictionary

[2] https://developers.google.com/maps/

[3] https://en.wikipedia.org/wiki/Dew_point#Calculating_the_dew_point

[4] https://bigladdersoftware.com/epx/docs/8-3/auxiliary-programs/energyplus-weather-file-epw-data-dictionary.html#field-horizontal-infrared-radiation-intensity

[5] https://pvlib-python.readthedocs.io/en/latest/

[6] http://pvlib-python.readthedocs.io/en/latest/_modules/pvlib/irradiance.html

6 Appendix

The appendix contains sample report files resulting from data extraction.

Data was extracted for following input parameters (Dresden, Germany):

- Longitude: 13.75 E
- Latitude: 51.06 N
- Year: 2015

6.1 Content of file ***00_report.txt***

```
Data extraction executed at 2018-04-17 22:34:18.588672
Chosen year: 2015
Latitude: 51.06
Longitude: 13.75

List of most favourable weather stations:
+------------------+-------+-----------+----------+---------+---------+-------------------+
|  ClimateElement  |   ID  | DateStart | DateEnd  |   Lat.  |   Lon.  |    StationName    |
+------------------+-------+-----------+----------+---------+---------+-------------------+
| air_temperature  | 01051 |  20070301 | 20161231 | 51.0240 | 13.7737 |  Dresden-Strehlen |
|    cloudiness    | 01048 |  19750701 | 20161231 | 51.1280 | 13.7543 | Dresden-Klotzsche |
|  precipitation   | 01051 |  20070301 | 20161231 | 51.0240 | 13.7737 |  Dresden-Strehlen |
|     pressure     | 01048 |  19750701 | 20161231 | 51.1280 | 13.7543 | Dresden-Klotzsche |
| soil_temperature | 01048 |  19760409 | 20161231 | 51.1280 | 13.7543 | Dresden-Klotzsche |
|      solar       | 01048 |  19960901 | 20171031 | 51.1280 | 13.7543 | Dresden-Klotzsche |
|       sun        | 01048 |  19740801 | 20161231 | 51.1280 | 13.7543 | Dresden-Klotzsche |
|       wind       | 01048 |  19730101 | 20161231 | 51.1280 | 13.7543 | Dresden-Klotzsche |
+------------------+-------+-----------+----------+---------+---------+-------------------+
Successfully extracted data for air_temperature
There are 0 missing entries in the original data set

Successfully extracted data for cloudiness
There are 7 missing entries in the original data set

Successfully extracted data for precipitation
There are 0 missing entries in the original data set

Successfully extracted data for pressure
There are 0 missing entries in the original data set

Successfully extracted data for soil_temperature
There are 112 missing entries in the original data set

Successfully extracted data for solar
There are 178 missing entries in the original data set

Successfully extracted data for sun
There are 0 missing entries in the original data set

Successfully extracted data for wind
There are 3 missing entries in the original data set
```

6.2 Content of file *00_missing_values.txt*

```
air_temperature

cloudiness
2015022012
<5 lines cut>
2015122407

precipitation

pressure

soil_temperature
2015110104
<110 lines cut>
2015111610

solar
Missing values for soil temperature at 2 cm depth were omitted
2015010110
<172 lines cut>
2015122015

sun
Missing values between 21:00 and 2:00 were omitted

wind
2015022513
2015022514
2015022515
```

6.3 Content of file *raw_data/air_temperature.txt*

```
+------------+---------------+-----------------+
|    Date    | Air Temp. [*C] | Rel. humid. [%] |
+------------+---------------+-----------------+
| 2015010100 |      0.8      |       99.0      |
| 2015010101 |      0.8      |       99.0      |
| 2015010102 |      0.8      |       99.0      |
| 2015010103 |      0.8      |      100.0      |
| 2015010104 |      0.9      |      100.0      |
<8750 lines cut>
| 2015123119 |      0.8      |       70.0      |
| 2015123120 |      0.2      |       85.0      |
| 2015123121 |      0.2      |       89.0      |
| 2015123122 |      0.1      |       91.0      |
| 2015123123 |     -0.1      |       92.0      |
+------------+---------------+-----------------+
```

6.4 Content of file *converted_data/air_temperature.txt*

```
+------------+--------+---------+----------+-----------+------+
|    Date    | (r)Air | (r)Rel. | Dry Bulb | Dew Point | Rel. |
|            | Temp[C]|Humid [%]| Temp [C] | Temp [C]  |Hum[%]|
+------------+--------+---------+----------+-----------+------+
| 2015010100 |  0.8   |   99.0  |   0.8    |    0.7    |  99  |
| 2015010101 |  0.8   |   99.0  |   0.8    |    0.7    |  99  |
| 2015010102 |  0.8   |   99.0  |   0.8    |    0.7    |  99  |
| 2015010103 |  0.8   |  100.0  |   0.8    |    0.8    | 100  |
| 2015010104 |  0.9   |  100.0  |   0.9    |    0.9    | 100  |
<8750 lines cut>
| 2015123119 |  0.8   |   70.0  |   0.8    |   -4.0    |  70  |
| 2015123120 |  0.2   |   85.0  |   0.2    |   -2.0    |  85  |
| 2015123121 |  0.2   |   89.0  |   0.2    |   -1.4    |  89  |
| 2015123122 |  0.1   |   91.0  |   0.1    |   -1.2    |  91  |
| 2015123123 |  -0.1  |   92.0  |   -0.1   |   -1.2    |  92  |
+------------+--------+---------+----------+-----------+------+
```