1. User manual

Basically, the software can be divided in two blocks. The first block is the code for database creation, and the second block is only related to wind resource assessment. Outputs of the software are also included in the second block.

Since the software has not been deployed and the user interface was not designed, future users will need to launch the code from the scripts. This means users are required to have minimum IT skills for its use. As explained, the software was written in *Python*, version 2.7 to be precise. All the work was developed under a *Linux OS* environment, in this case *Ubuntu 14.04 LTS* was chosen. All the instructions are based on the use of the same operating system.

1.1 Software installation

First at all, it is mandatory to install the same database, *MySQL*. This could be done by the use of *Ubuntu Software Center.* Installing the *MySQL Workbench* it may be useful for beginners to database programming. At *MySQL* installation, the password for user "root" will be set by first time. It is recommended to maintain the username and host name as given by default, "root" and "localhost" respectively. Also, the password must be kept.

Once the database software is installed, some *Python* libraries must be installed. The *Python* version must be the same; if that's not the case it will be necessary to install it. Here is the list of the *Python* packages included in the software; *pandas, numpy, scipy, MySQLdb, netCDF4, xml, gzip, openpyxl, windrose, matplolib* and *basemap*. All these packages are available through *PiPY*, the online repository for *Python*.

One of the packages, *MySQLdb*, is especially important. This is the bridge between *Python* and *MySQL*. By default, the data directory is set to a folder created during installation. It is possible to change the data directory. The change is very convenient due to future DB size, thus operating system and DB can be placed in different hard drives. Setting a new data directory for *MySQL* is not straightforward. It is necessary to edit a few files from the operating system.

1.2 User interaction

Once the software is installed, there would not be any required programming skill with *Python*. However, since there is no user interface, it will be necessary to run *Python* scripts from the command line, as well as a very basic knowledge of *Unix/Linux*. Before running any script, all the database scripts must be kept in the same folder. It is the same case for the wind assessment scripts. Database scripts and wind assessment scripts are completely independent. Thus, it is better to separate both script groups into different folders.

Once the command prompt is opened, the directory must be changed to the folder containing the scripts. This is done by simply writing:

cd path/to/the/folder

After, the script can be summoned easily by writing:

python name_of_script.py

These steps are show in Figures 1 and 2. There are only two scripts to remember, one for each software block; database creation and wind assessment. Once any of these two scripts are called, the processes will start asking the user to make some choices and/or set some parameters as shown in Figures 2 and 4.

The script name to be called for starting the database creation is *Insertion_process.py.* On the other hand, the script name to begin the wind assessment is *sat_options.py.* It could be helpful to follow the flowcharts from Figures 6 and 8 for database creation and wind assessment respectively.

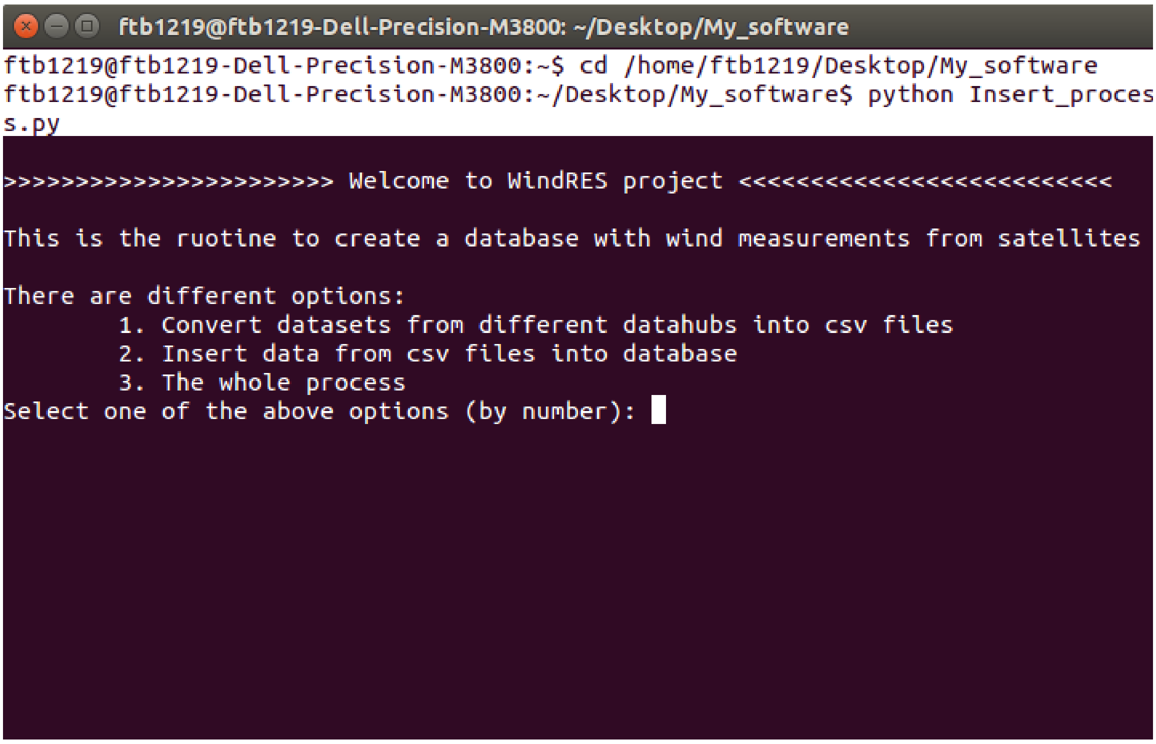

Figure 1. Screen shoot with the software window for creating the database.

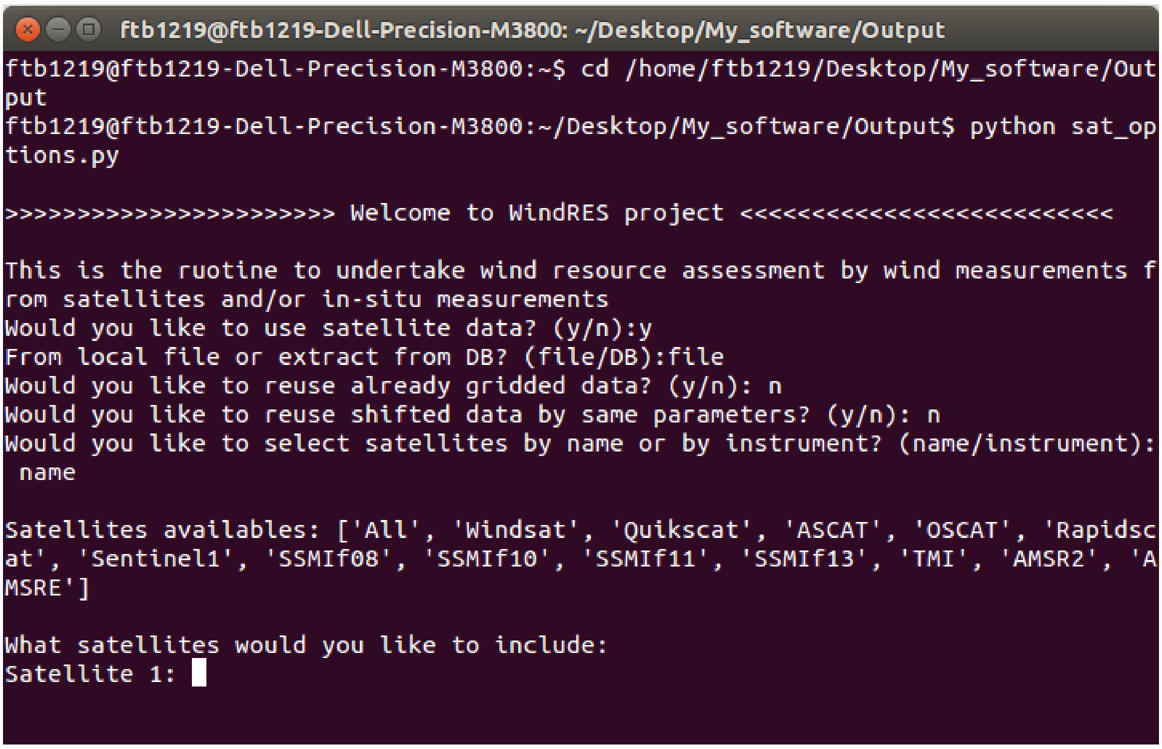

Figure 2. Screen shoot with the software window for the wind assessment.

After any of those two scripts are run, the user will only interact with the command line window for taking choices and decisions, setting some parameters and inserting path to specified files. The software has been designed to be easily used from the command line. There is an exception, every time an output, such as a map, graph, or wind rose, is reached, a new window will be displayed. That window is created by the *Python* package *matplotlib.* A few options will appear on the window such as zoom, grid and more important save file as a picture. An example of an output window is shown in Figure 3. Furthermore, in Figure 3 is shown the example of the software interface for the output.

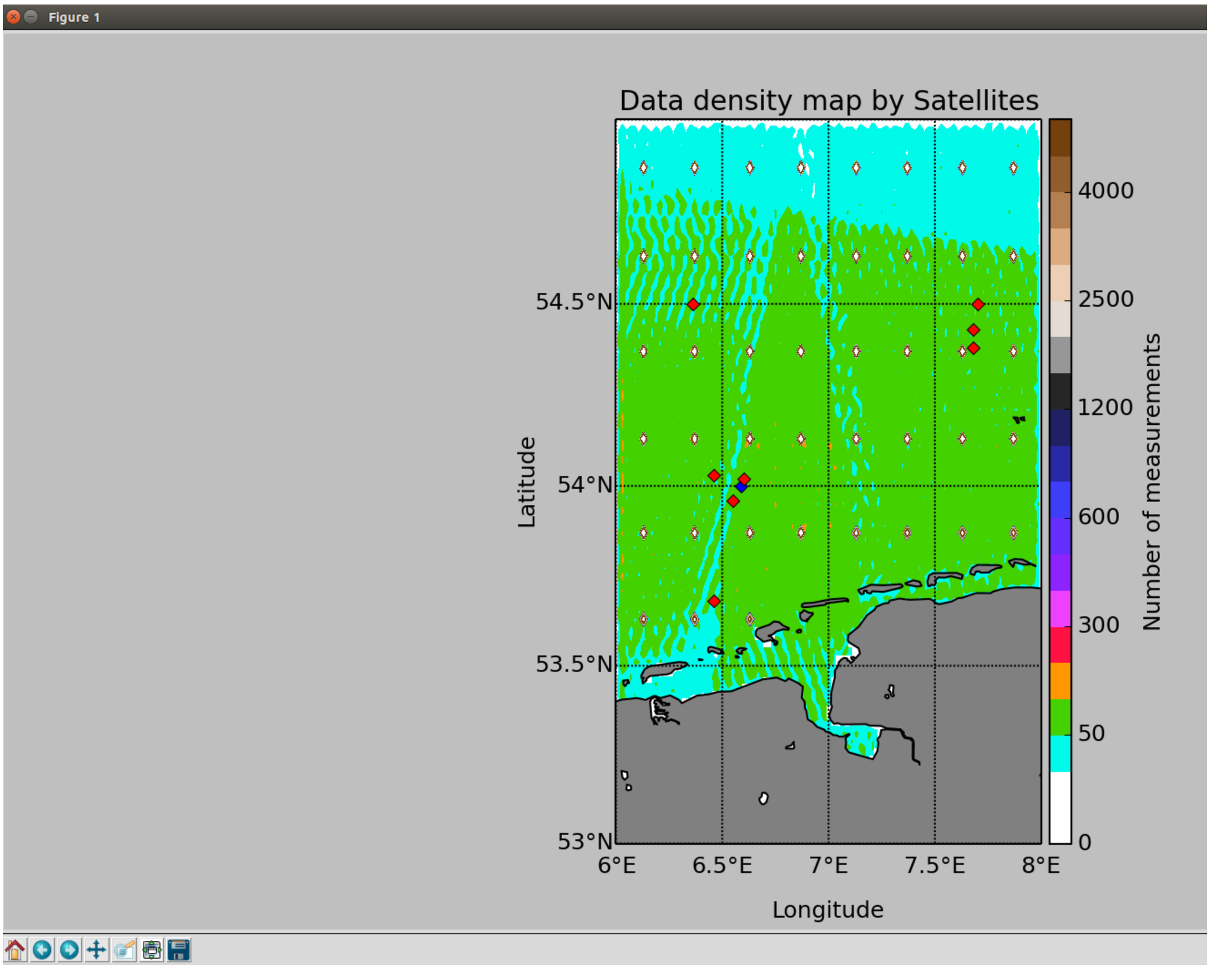

Figure 1. Software output window example, in this case for a data density map.

```
Would you like to use satellite data? (y/n):y
From local file or extract from DB? (file/DB):file
Would you like to reuse filtered data by same parameters? (y/n): y
Full, absolute path to file with applied filters: All_instant.csv
At what height had the winds been lifted?(m): 100

What basis was chosen long-term or instant basis?(long/instant): instant

Set area:
        Longitude 1 (-180 to 180):6
        Longitude 2 (-180 to 180):8
        Latitude 1 (-90 to 90):53
        Latitude 2 (-90 to 90):55
Would you like to set period?(y/n)y

Do you want to include date interval?(y/n)y
Starting date (dd/mm/yyyy):01/01/2000
End date (dd/mm/yyyy):31/12/2016

Do you want to include time interval?(y/n)n

What year/s would you like to include (All to select all, None to unselect all):

Year 1: None

What months would you like to include: ['january', 'february', 'march', 'april',
 'may', 'june', 'july', 'august', 'september', 'october', 'november', 'december'
, 'All', 'None']
Month 1: None

What day basis would you like to choose? ['day', 'night', 'All', 'None']
Day basis: None
Would you like to include in-situ measurements?(y/n)n

>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>> Calculating statistics >>>

>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>> Statistics already calculated >>>

Select outputs: ['Weibull_curve', 'Charts', 'Wind_rose', 'Parameters', 'Wind map
', 'Regression', 'Export', 'All']
Output: Wind map
Output:

Where you want to study wind resource? Choose coordinates:
        Longitude (-180 to 180):7
        Latitude (-90 to 90):54

>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>> Calculating raster layer >>>
```

Figure 4. Software interface used to achieve the output from Figure 3.6.

1.3 Database

Although satellite data is available online, freely for many devices, the datasets are not arranged in the same format. Each data hub has its own format. Hence, there it was necessary to transform the data to a standard format. Furthermore, because of the nature of the different devices, this new format had to be capable of including all instruments; radiometers, scatterometers and SAR. Altimeters and GNSS-R were discarded because of their narrow swaths. The main difficulty with this was the combination of different resolutions.

The first step was to set a format. For wind resource assessment, the more data that is included in an analysis, the better the accuracy. Hence, the DB was devised to host the data from complete missions when possible. This meant a very large amount of data would need to be manipulated, i.e. satellite Big Data. For that reason, and in order to decrease the amount of data to compute simultaneously, the DB was composed of 1,200 tables. These tables represent a UTM square as shown in Fig. 5. Exceptions located in the *X* and *V* rows were removed and divided in a similar form as the rest of the rows.

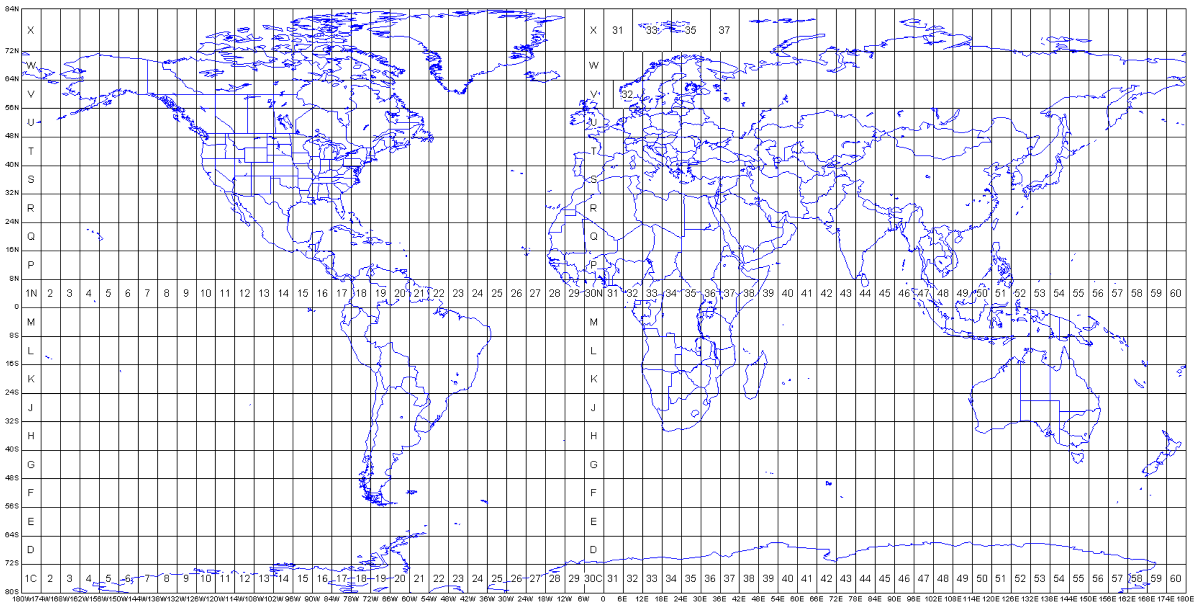

Figure 5. World map with UTM division.

Satellite datasets are arranged by daily products for radiometers and scatterometers. Each product includes the different measurable parameters for the entire globe. Where the satellite was not observing, non-valid or sentinel values are found. This fact leads to an unnecessary multiplication of data. The case for Sentinel-1 is different since their datasets are arranged by single images. Thus, the first scripts to be written were designed to read each kind of satellite dataset, removing the sentinel data and inserting the remaining data into the appropriate table of the DB.

As explained previously each device has its own resolution. In order to place data from different satellites into the same table, four different fields for geolocation were added. These are low and high longitude, plus low and high latitude. Along with these fields, Satellite name, Type of Instrument, Date and Time, Wind speed and direction were also included. Once all data was inserted into the DB, the scripts to read and select data from the DB were written. Since the tool was designed to be focused in small areas, there was no need to work with all tables at once. Thus, the tool is capable of selecting the data according to the location and date, and export it to a *CSV* file. The benefits of this process are minimum work with the DB and again a decrease in the amount of unnecessary data being processed. As a result, the speed at which the whole tool operates was kept as high as possible.

1.3.1 Database creation

Before the database, the first step is to download the satellites datasets. It is recommended to download the full mission of a satellite before inserting the data into the database. Datasets can be kept in the same folder or directory tree. After, the datasets could be transformed into a new format, *CSV* files. At this point datasets are ready for insertion into database. Parallel to satellite data transformation, the database with its 1200 tables must be created. Scripts for all the explained processes were created. The whole DB creation process is shown in Figure 6.

Even when the code was designed to have high performance, some routines are slow. Scripts were written to handle only one satellite mission every time. Thus, the risk of process interrupted, and therefore data loses, was reduced. However, an advanced user may try to modify the code to automate the process or to work in parallel. For any user, here there are some recommendations when preparing a computer to host the DB:

- Hard drive larger than 3 TB.
- Memory RAM larger than 32 GB.
- Process each satellite mission separately or independently. Satellite datasets can be processed in parallel strings, but insertion into DB must be a single string.
- Build indexes only when all satellite data has been inserted into DB.

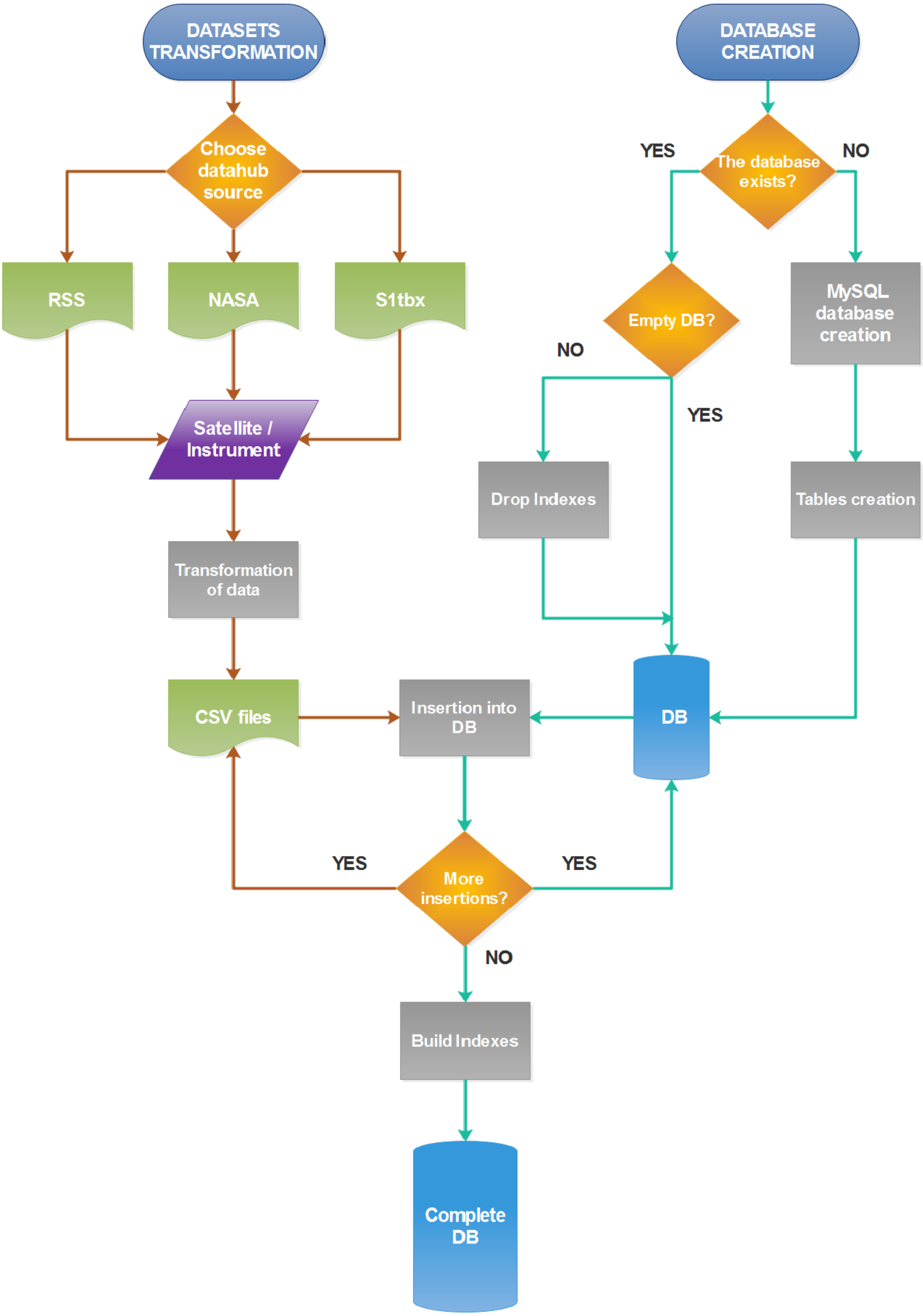

Figure 6. Full process for the creation of the database. Brown lines represent satellite data; green lines represent database management.

1.4 Wind resource assessment design

In order to explain all of the capabilities and structure of the software, the whole analysis process is explained. The process is also represented in Fig. 8 to ease comprehension.

1. **Working with or without satellite data**. Although the goal was to use satellite data, the software is able to do a wind assessment with only in situ or satellite measurements or both at the same time. However, users need to provide in situ data. At this point, the software only reads in situ data in one format of data table. Extra scripts were written to create a specific template according to user preferences. Thus, users can transform their data in order to be readable by the tool.

2. **Acquiring satellite data**. As explained, satellite data is contained in a DB. Once the target area is set, the tool exports a file containing all satellite wind measurements for the specified location. Since the tool was designed for small areas, a maximum of four UTM squares were allowed as an area definition. The software works with a coordinate system in degrees; from −180° to 180° for longitudes, and from −90° to 90° for latitudes.

3. **Filtering data.** This filtration removes missing values, and it is done according to some optional parameters that the user can choose, such as:

    - Rain rate.
    - Satellites to work with.
    - Instruments (or type of devices) to work with.
    - Atmospheric conditions.
    - Overlaps between ground and satellite measurements.
    - Period of study. This option is very flexible since it is not just a selection of the start and end date. Different temporal analyses can be undertaken; by year, by months of one year or by the same month but different years, by day and night and also by selected time for different days.

4. **Setting grid or resolution.** This option depends entirely on the user's selection of satellites because the resolution on the wind map is exactly the same as the resolution of the satellite data. By default the highest resolution is

chosen automatically. For example, SAR satellites have the best resolution with pixels of 2km x 2km, whereas the most common resolution is 25km x 25km. Satellite data is adapted to the grid.

In this process, a problem was considered, which was named "data spreading". This problem consists of the increase in the number of measurements when low-resolution data is fitted to the high-resolution grid. Since low-resolution pixels cover a larger area than high-resolution pixels, the low-resolution data could be represented in many high-resolution pixels. Fig. 8 is a representation of data spreading. If every high-resolution pixel contains the values from low-resolution data, the low-resolution data is multiplied by a factor of (low resol. / high resol.)^2, this factor is 25 in Fig 7. Furthermore, this fact will automatically be a loss of resolution and, when represented, high-resolution pixels will not be observed.

In order to solve data spreading, every single low-resolution pixel was inserted in only one high-resolution pixel from the grid. A single pixel in microwave remote sensing represents the average wind speed over that area. The maximum representativeness of the measurement is expected to be the centre of such a pixel. Hence, the high-resolution pixel hosting the measurement from a low-resolution pixel is the one closest to the centre of the low-resolution pixel. This method leads to the production of cells or pixels with a huge amount of data called HDDC (High Data Density Cells).

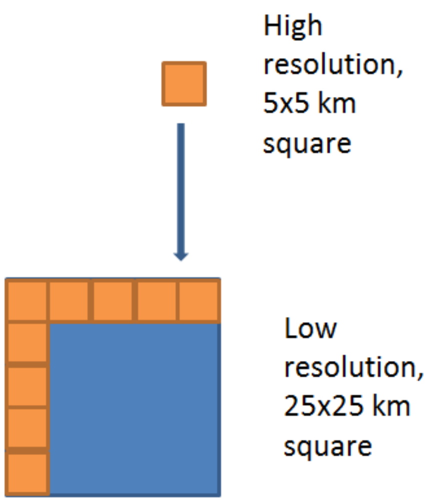

Figure 7. Representation of data spreading.

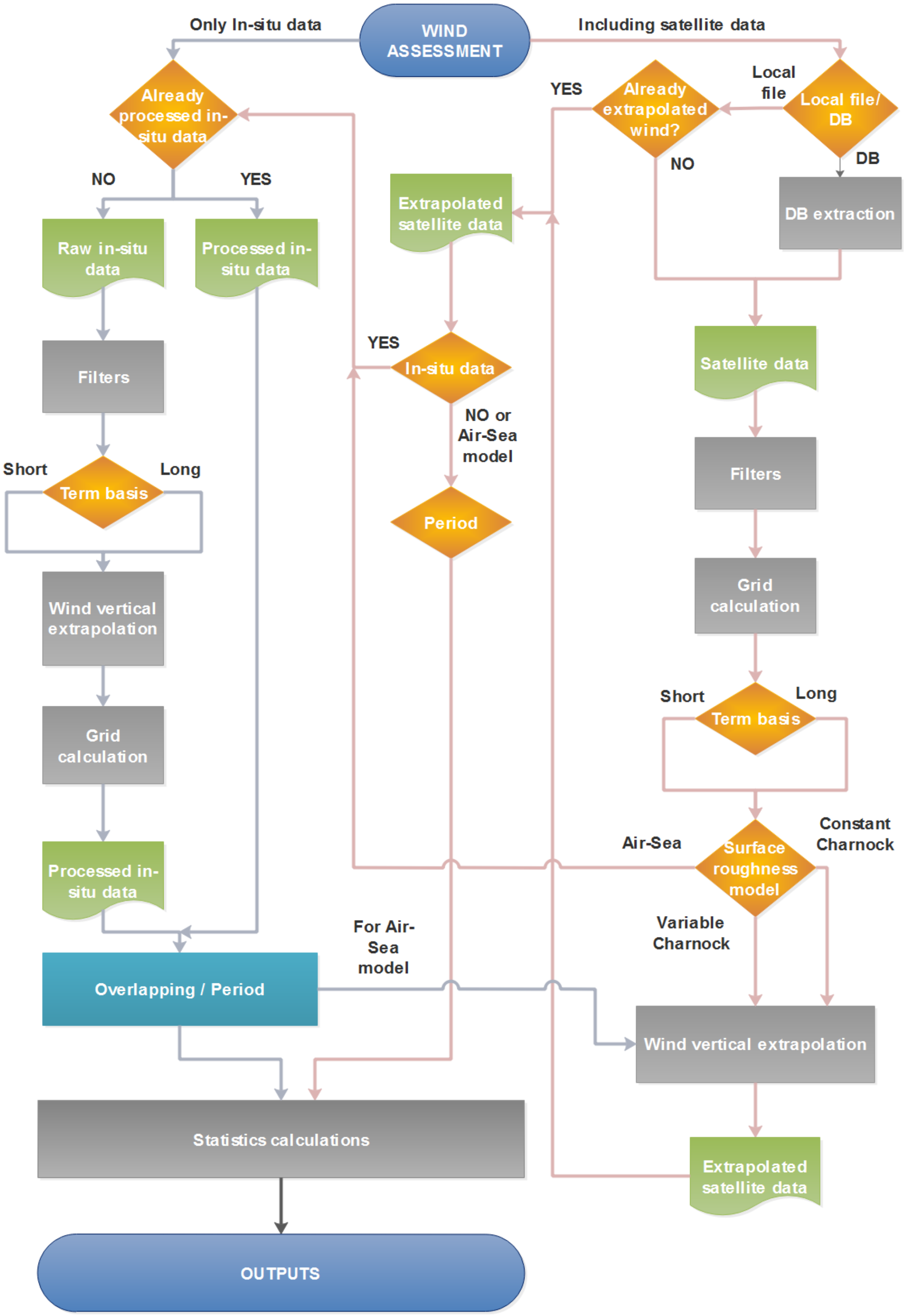

Figure 8. Flowchart of the software structure. Grey line represents in situ data, red lines means satellite data.

5. **Vertical wind extrapolation.** Here winds at hub height are calculated. The user must choose the hub height. It is possible to work on a long or short-term basis. There are three different models for shifting the wind; from simple to complex data analysis:

    - Iteration with constant Charnock parameter
    - Variable Charnock parameter and friction velocity
    - Air-sea interaction models to calculate surface roughness

6. **Choosing output parameters.** Users can select what it is going to be represented. The software can do comparisons between different years, months or day or night; and so, observe yearly seasonal and diurnal variations. Hence, the software is capable of producing different maps. Otherwise, the user may select to make one wind map with all available data and higher data density. It is also possible to select those satellite measurements that overlap in time with in-situ measurements and so be able to undertake a fair comparison and correlation between satellites and ground devices.

7. **Calculating statistics.** According to which parameters were chosen in the previous step not only are the mean, minimum, maximum wind speed and standard deviation calculated but also Weibull parameters. At this point, it could easily calculate power density, turbulence intensity, etc. However, these were not included in an analysis as explained in chapter 4. The reason for this is that the most important parameter to be studied is the wind speed.

8. **In-situ measurements**. There is an option to include in-situ measurements inside the area under study, and so be able to do an accuracy study. This data will be processed under the same parameters as the satellite data. In order to use in-situ data, it is necessary to adapt the data to the templates used by the software thereby allowing the analysis software to read it.

1.5 Outputs

As explained previously, at the moment the main output is a wind map. However, in the future, it could be possible to include more outputs such as energy yield, turbulence intensity, extreme winds and risk calculations. If the user selects a set of coordinates inside the map, the software can release the following:

- Wind roses.
- Weibull curves.
- Bar charts of comparison for wind speed, temperature, pressure, and humidity. The last three parameters were included for in-situ measurements only.
- CSV file with all parameters for each pixel of the map.
- CSV file is pre-formatted to be an input for a GIS tool. This is useful when a user needs a better layout of the wind map.
- Regression analysis for wind speed at calculated height and surface roughness between satellites and in-situ devices.

1.6 Access to satellite data

Instrument	Data access	Type of data
ERS 1&2 Envisat ASAR	website https://earth.esa.int Go to data access/browse data product Earth topic = ocean and coast Mission = Envisat Instrument = ASAR Typology = Radar imagery	Processing level = 0 / 1 / 1B / 2
NSCAT	ftp://podaac-ftp.jpl.nasa.gov/OceanWinds/	L2 product
QuikSCAT	RSS website http://www.remss.com/	L2 product
ASCAT	ftp://podaac-ftp.jpl.nasa.gov/OceanWinds/	L2 product
	RSS website http://www.remss.com/	L2 product
SSM/I	RSS website http://www.remss.com/	10 m above sea level Daily / 3-day /

		weekly / monthly data
WindSat		
AMSR		L2 product
TMI		
OSCAT	ftp://podaac-ftp.jpl.nasa.gov/OceanWinds/	Level 2B product
Sentinel-1	https://sentinel.esa.int/web/sentinel/sentinel-data-access/access-to-sentinel-data/	L0 or L1 products
RapidScat	ftp://podaac-ftp.jpl.nasa.gov/OceanWinds/	L2 product