# R-DOUBLE – ver. 09-04-2020

'R-double' is a Shiny-R application oriented to facilitate the analysis of data series and their correlation. The application runs inside Rstudio or on a Shiny server.

The application uses the following libraries which must be installed in R studio::

dplyr , openair, ggplot2, shinythemes, DT, BaTFLED3D, gridExtra, ggpubr, tidyverse, data.table.

Setup:

Unzip the zip file into a directory, once decompressed the file structure must be the following:
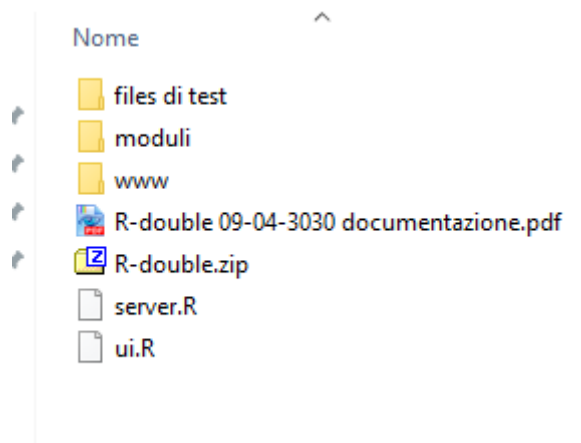
Within Rstudio open the ui.r file and launch the script.

Main functions

1. Import of data series in csv format from multiple files, possibility to generate a working dataset by combining data from multiple csv files.
   - Possibility of aggregation of the imported data in hourly, daily, monthly averages.
   - Export the working dataset in csv format.
2. Analysis of imported series with:
   - Statistical summary quartiles
   - Abnormal values detection
   - Histogram of frequency distribution
   - Quartile box plot
   - Graphical variation analysis over time with possibility of data export:

- Hourly week type
- Typical day
- Typical month
- Typical weekly day

3. Analysis of the correlation between two series with:
   - Scatter plot of the two series with linear regression line
   - Comparative box plot of statistical position data
   - Comparison chart of the two series.
   - Calculation of correlation indices: Pearson, Kendall, RMSE, NMRSE
   - Statistical summary of the two series with calculation of the differences for each quartile.
   - Export of the elaborations in PDF format.

Import of multiple files in csv format and creation of a working dataset.

The application has been designed to integrate data sets from multiple csv files into a single working dataset on which the processing will be carried out.

The first step is to import a csv file.

By default the 'import field names' and 'time series on the first column' options are set, which entail having the first field in a date and / or time format and importing the column headings of the csv file as field names.

Pressing the 'Browse ...' button opens a window on the file system that allows you to choose the file to be loaded. By choosing a file and pressing 'Open' the file is loaded, the light blue bar indicating the upload increases until the word 'Upload complete' appears.

At this point it is convenient to open the CSV file also with a text editor (eg Notepad) to check how it is structured. The application to correctly import the csv file must know the separator of the values, the character used as decimal, if the values are enclosed in quotes.

Another fundamental aspect is the setting of the data-time field which must be the first in the series, it must be chosen in the 'Date format' tab. If the date setting is not among those set, you can enter a personalized one by choosing the 'User defined' option with the correct set of parameters day% d, month% m, year% Y, hour% H, minutes% M . It is also important to correctly set the date and time separator character.

When the setting of the csv structure is correct the application is able to load the file showing the first 5 records and a graphic preview of all the imported series:
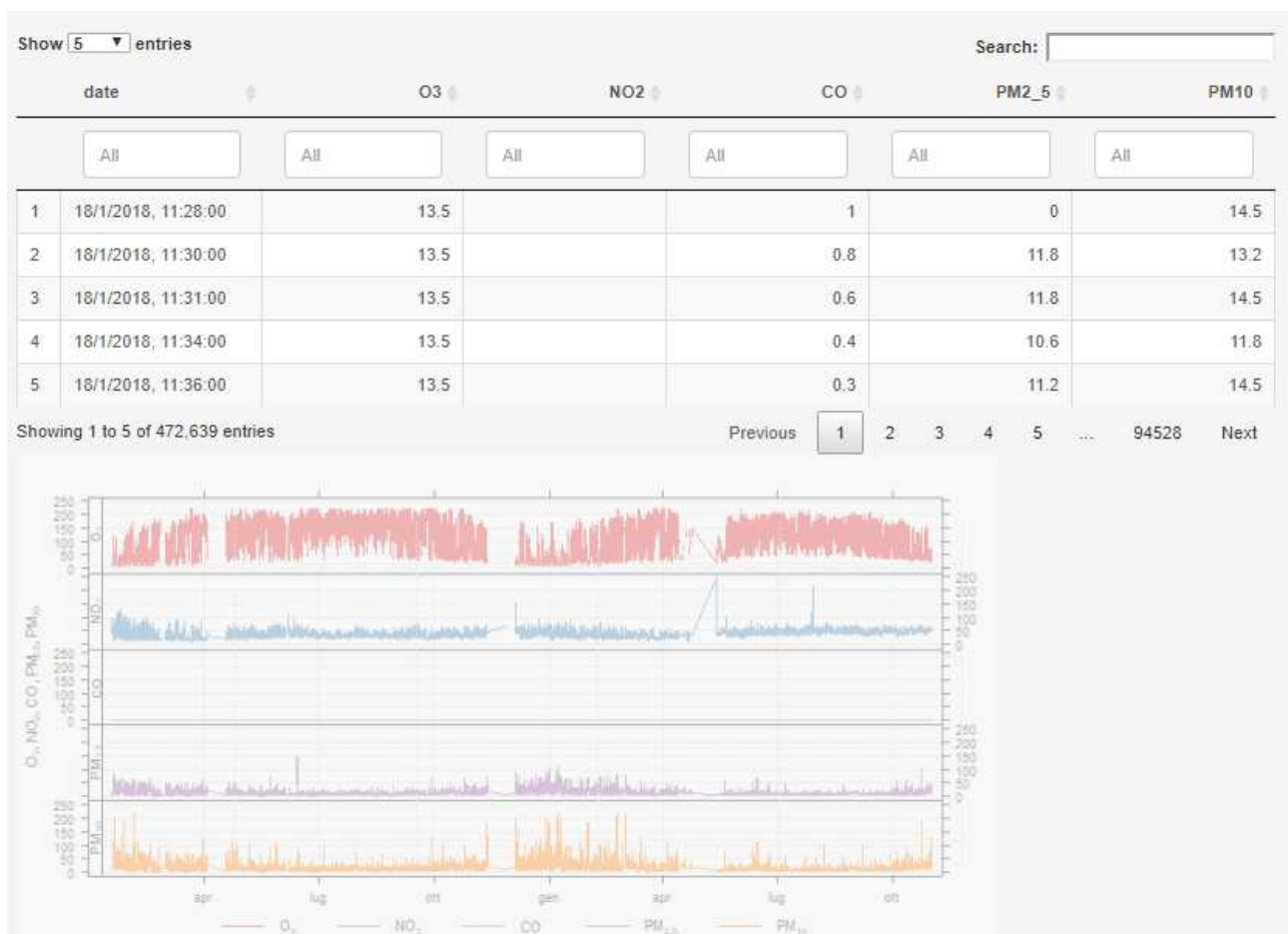
A questo punto è possibile operare in modo preliminare sulla serie dei dati in vari modi.
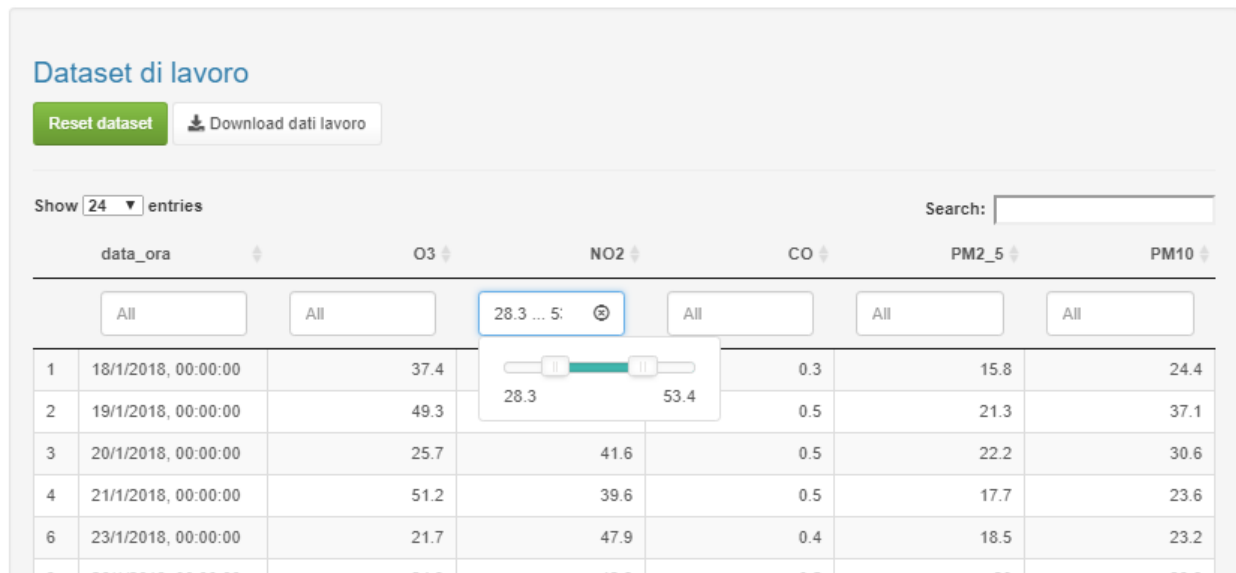
1. E' possibile integrare i dati puntuali in medie orarie, giornaliere, mensili. Per serie molto grandi è anche un modo per velocizzare le elaborazioni.
2. E' possibile eliminare numeri negativi oppure eliminare numeri superiori ad un certo valore.
3. E' possibile arrotondare i valori ad un numero di decimali a piacere.

Dopo questa fase il primo set di dati è pronto, premendo il pulsante rosso 'Aggiungi a dataset di lavoro' il subset acquisito viene copiato nel Dataset di lavoro che rappresenta la base di dati sulla quale verranno effettuate le operazioni.

L'operazione di import csv può essere ripetuta n volte in modo da integrare molteplici fonti di dati. **I dati vengono integrati sul campo della data-ora pertanto sarà opportuno una volta scelta una aggregazione temporale (es media oraria) mantenerla per tutti i dataset importati.**

The working dataset

The working dataset is the database used to carry out the processing. By default the data is sorted by date_time, however you can sort the data by any of the columns. It is also possible to filter the dataset by inserting a specific range of values in one or more columns:

The working dataset can be exported in csv format by pressing the button 'Download job data'

Single series analysis

By choosing one of the series from the list box, the application offers a series of elaborations:

- Statistical summary quartiles
- Abnormal values detection
- Histogram of frequency distribution
- Quartile box plot
- Graphical variation analysis over time with possibility of data export:
- Hourly week type

- Typical day
- Typical month
- Typical weekly day

Correlation analysis

this page provides some graphic and numerical processing to analyze the correlation of two series. By default, the first - which constitutes the reference - and the second series of the work dataset are selected.
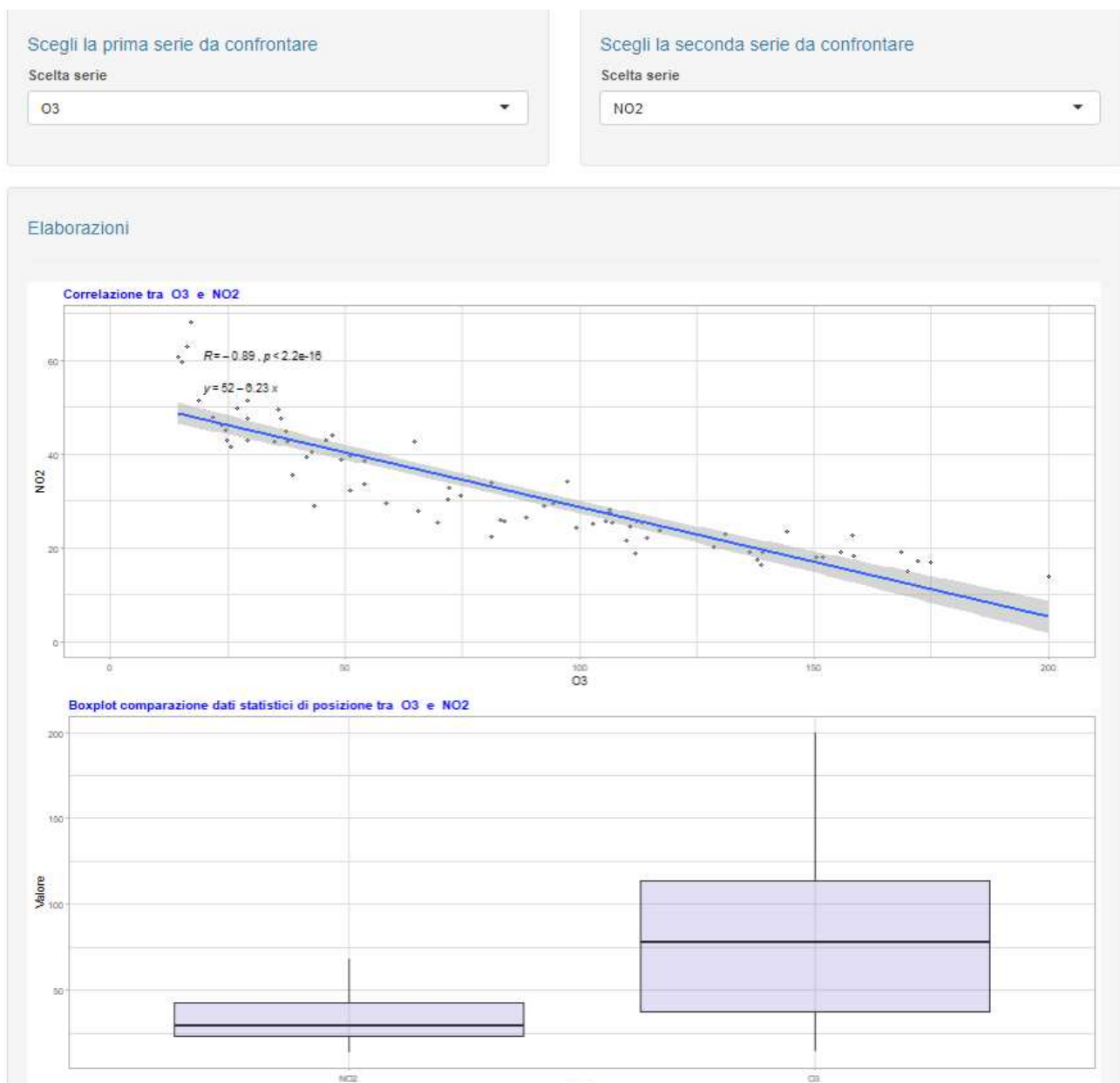
The page can be exported in PDF format by pressing the Download button at the top left.