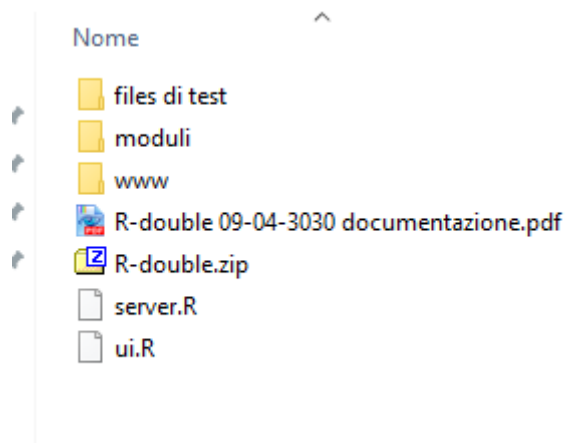


Documentazione R-DOUBLE – ver. 09-04-2020

R-double è una applicazione Shiny-R orientata a facilitare l'analisi di serie di dati e la loro correlazione. L'applicazione gira all'interno di Rstudio oppure in un server Shiny. L'applicazione utilizza le seguenti librerie che devono essere installate in R studio: **dplyr , openair, ggplot2, shinythemes, DT, BaTFLED3D, gridExtra, ggpubr, tidyverse, data.table.**

Installazione e lancio applicazione

Scompattare il file zip in una directory, una volta decompresso la struttura dei files deve essere la seguente:



All'interno di Rstudio aprire il file ui.r e lanciare lo script.

Funzioni principali

1. Import di serie di dati in formato csv da file multipli, possibilità di generare un dataset di lavoro combinando dati da più files csv.
 - Possibilità di aggregazione dei dati importati in medie orarie, giornaliere, mensili.
 - Export in formato csv del dataset di lavoro.
2. Analisi delle serie importate con:
 - Sommario statistico quartili
 - Rilevamento valori anomali
 - Istogramma della distribuzione delle frequenze
 - Box plot dei quartili
 - Analisi variazione nel tempo grafica con possibilità di export dei dati:
 - Settimana oraria tipo

- Giorno tipo
- Mese tipo
- Giorno settimanale tipo

3. Analisi della correlazione tra due serie con:

- Scatter plot delle due serie con retta di regressione lineare
- Box plot di comparazione dei dati di posizione statistici
- Grafico di comparazione delle due serie.
- Calcolo indici di correlazione: Pearson, Kendall, RMSE, NMRSE
- Sommario statistico delle due serie con calcolo delle differenze per ogni quartile.
- Export delle elaborazioni in formato PDF.

Import di file multipli in formato csv e costituzione di un dataset di lavoro.

L'applicazione è stata pensata per integrare serie di dati da file csv multipli in un unico dataset di lavoro sul quale verranno effettuate le elaborazioni.

Il primo passo consiste nell'importare un file csv.

Per default sono impostate le opzioni 'importa nomi campi' e 'serie temporale sulla prima colonna' che comportano di avere il primo campo in un formato data e/o ora e di importare le intestazioni delle colonne del file csv come nomi dei campi.

Premendo sul pulsante 'Browse...' viene aperta una finestra sul file-system che consente la scelta del file da caricare. Scegliendo un file e premendo 'Apri' il file viene caricato, la barra celeste che indica il caricamento cresce fino a che compare la scritta 'Upload complete'.



A questo punto è conveniente aprire il file csv anche con un editor di testo (es. Notepad) per verificare come è strutturato. L'applicativo per importare correttamente il file csv deve sapere il separatore dei valori, il carattere usato come decimale, se i valori sono racchiusi da apici.

Attenzione: Il campo data-ora deve essere il primo campo del dataset

Select a csv file

Browse... SMART19_FINALE.csv

Upload complete

☒ Importa nomi campi ☐ Elimina maggiori di

☒ Serie temporale sulla prima colonna

☐ Elimina dati negativi

1000

Apici

☒ Nessuno

☐ Doppio

☐ Singolo

Separatore

☐ Virgola

☒ Punto e virgola

☐ Tab

Carattere decimale

☐ Virgola

☒ Punto

Formato Data

☒ Italiano DD/MM/AAAA H:M

☐ Italiano DD-MM-AAAA H:M:S

☐ Italiano DD/MM/AAAA H:M:S

☐ Inglese AAAA-MM-DD H:M:S

☐ Inglese AAAA/MM/DD H:M:S

☐ Inglese AAAA-MM-DD H:M:S

☐ Inglese AAAA-MM-DD

☐ Definito da utente

Formato utente

%d/%m/%Y %H:%M

Aggregazione temporale

☒ Nessuna aggregazione

☐ dati orari

☐ dati giornalieri

☐ dati mensili

☒ Aggrega solo con 75% validi

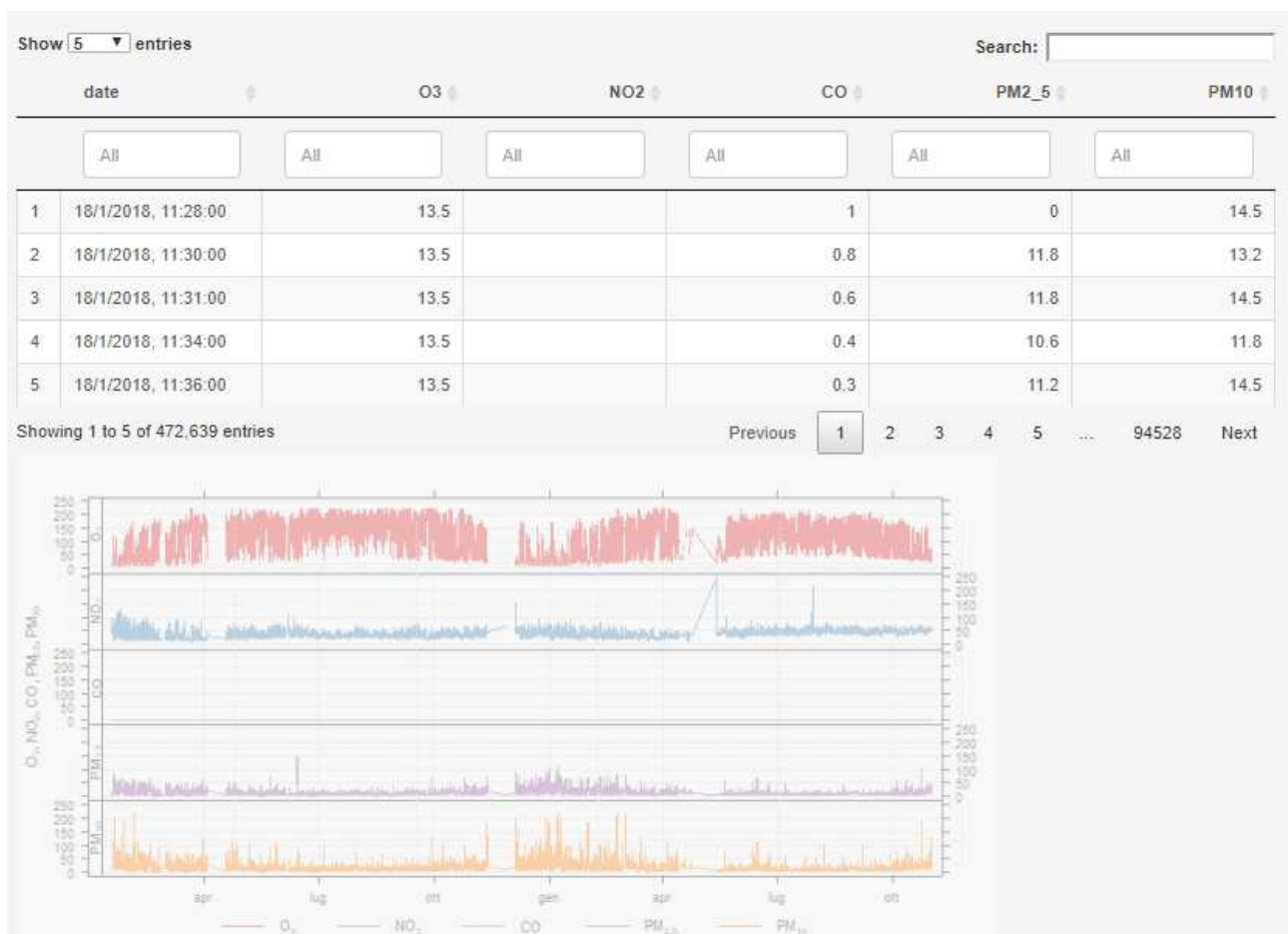
☒ Arrotonda

Decimali arrotondamento

1

Un altro aspetto fondamentale è l'impostazione del campo dato-ora che deve essere il primo della serie, occorre sceglierla nel tab 'Formato data'. Se l'impostazione della data non è tra quelle impostate è possibile inserirne una personalizzata scegliendo l'opzione 'Definito da utente' con il corretto set dei parametri giorno %d, mese %m, anno %Y, ora %H, minuti %M. E' altresì importante la corretta impostazione del carattere separatore della data e dell'ora.

Quando il settaggio della struttura del csv è corretto l'applicativo è in grado di caricare il file mostrando i primi 5 record ed una anteprima grafica di tutte le serie importate:



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.**

Il dataset di lavoro

Il dataset di lavoro è la base dei dati utilizzata per effettuare le elaborazioni. Per default i dati sono ordinati per data_ora, è comunque possibile ordinare i dati per una qualsiasi delle colonne. E' inoltre possibile filtrare il dataset inserendo un range di valori specifico in una o più colonne:



The screenshot shows a web interface titled "Dataset di lavoro". It includes a "Reset dataset" button and a "Download dati lavoro" button. Below these, there's a "Show 24 entries" dropdown and a "Search:" input field. The table has columns: data_ora, O3, NO2, CO, PM2_5, and PM10. Each column has a filter dropdown. The NO2 filter is active, showing a range from 28.3 to 53.4. The table displays 6 rows of data.

| | data_ora | O3 | NO2 | CO | PM2_5 | PM10 |
|---|---------------------|------|------|-----|-------|------|
| 1 | 18/1/2018, 00:00:00 | 37.4 | 28.3 | 0.3 | 15.8 | 24.4 |
| 2 | 19/1/2018, 00:00:00 | 49.3 | 53.4 | 0.5 | 21.3 | 37.1 |
| 3 | 20/1/2018, 00:00:00 | 25.7 | 41.6 | 0.5 | 22.2 | 30.6 |
| 4 | 21/1/2018, 00:00:00 | 51.2 | 39.6 | 0.5 | 17.7 | 23.6 |
| 6 | 23/1/2018, 00:00:00 | 21.7 | 47.9 | 0.4 | 18.5 | 23.2 |

Il dataset di lavoro può essere esportato in formato csv premendo il pulsante 'Download dati lavoro'

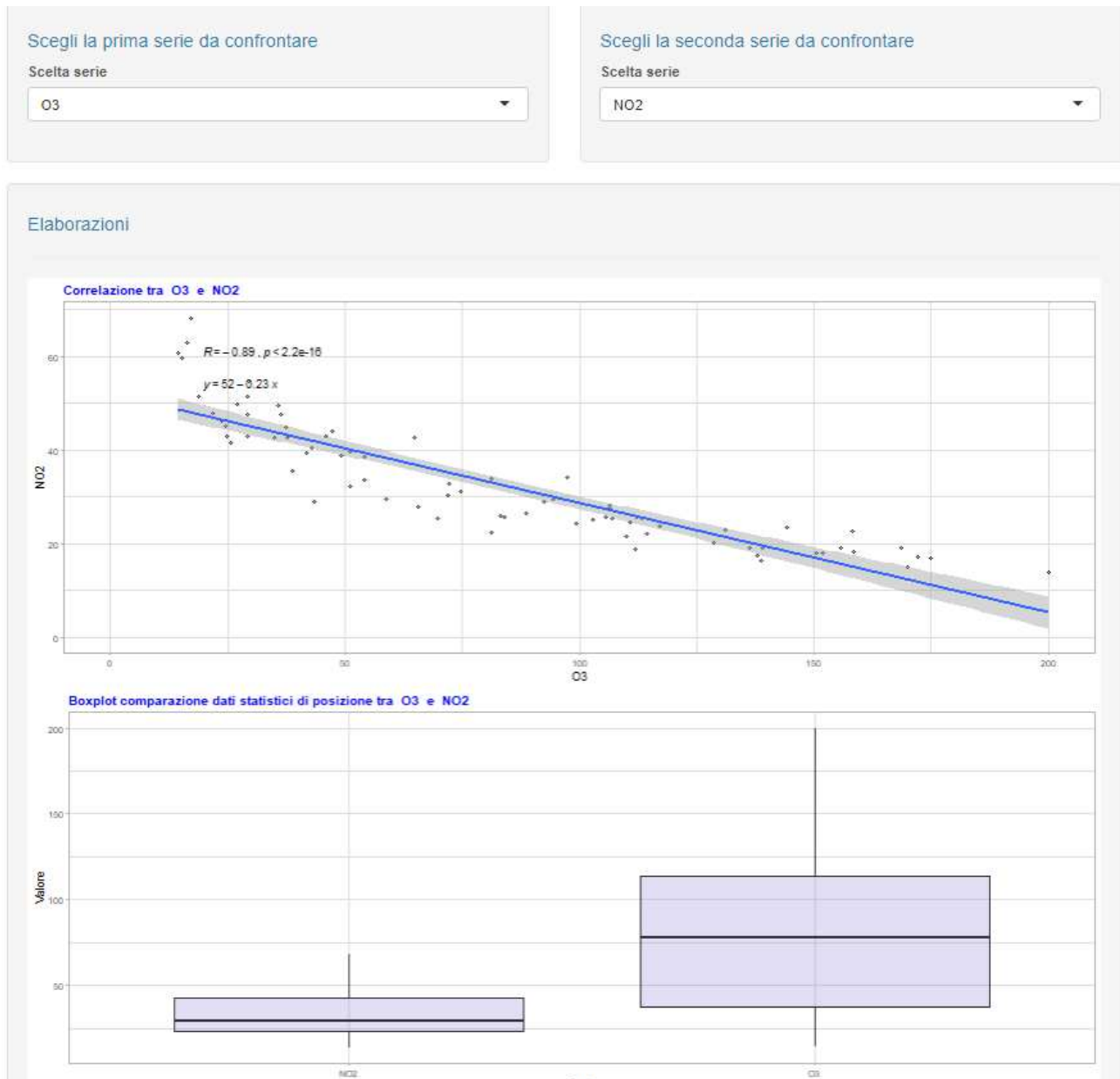
Analisi singola serie

Scegliendo una delle serie dalla list box l'applicazione propone una serie di elaborazioni:

- Sommario statistico quartili
- Rilevamento valori anomali
- Istogramma della distribuzione delle frequenze
- Box plot dei quartili
- Analisi variazione nel tempo grafica con possibilità di export dei dati:
 - Settimana oraria tipo
 - Giorno tipo
 - Mese tipo
 - Giorno settimanale tipo

Analisi di correlazione

in questa pagina vengono fornite alcune elaborazione grafiche e numeriche per analizzare la correlazione di due serie. Per default vengono selezionate la prima - che costituisce il riferimento - e la seconda serie del Dataset di lavoro.



La pagina può essere esportata in formato PDF premendo il pulsante Download in alto a sinistra.