UNIVERSITÀ DEGLI STUDI DI TRIESTE

Dipartimento di Scienze Chimiche e Farmaceutiche

SOMEnv: un free software tool con Graphical User Interface per l'identificazione di profili ricorrenti di odori tramite elaborazione con algoritmo Self-Organizing Map di dati registrati da IOMS

S. Licen (slicen@units.it)[a], M. Franzon[b], T. Rodani[c], S. Cozzutto[d], P. Barbieri[a]

[a] Dept. of Chemical and Pharmaceutical Sciences, University of Trieste, Via L. Giorgieri 1,34127 Trieste , Italy;
[b] eXact lab s.r.l. , via Beirut, 2 - 34151 Trieste (Italy);
[c] AREA Science Park - Padriciano, 99 34149 Trieste, Italy
[d] ARCO SolutionS s.r.l., spin-off company of the Dept. of Chemical and Pharmaceutical Sciences, University of Trieste, Via L. Giorgieri 1, 34127 Trieste , Italy

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Dipartimento di Scienze Chimiche e Farmaceutiche

UNITA' DI RICERCA IN CHIMICA ANALITICA AMBIENTALE

UNIVERSITA' DEGLI STUDI DI TRIESTE

Prof. Gianpiero Adami

Prof. Pierluigi Barbieri

Prof. Matteo Crosera

Dr. Sabina Licen

UNITA' DI RICERCA IN CHIMICA ANALITICA AMBIENTALE

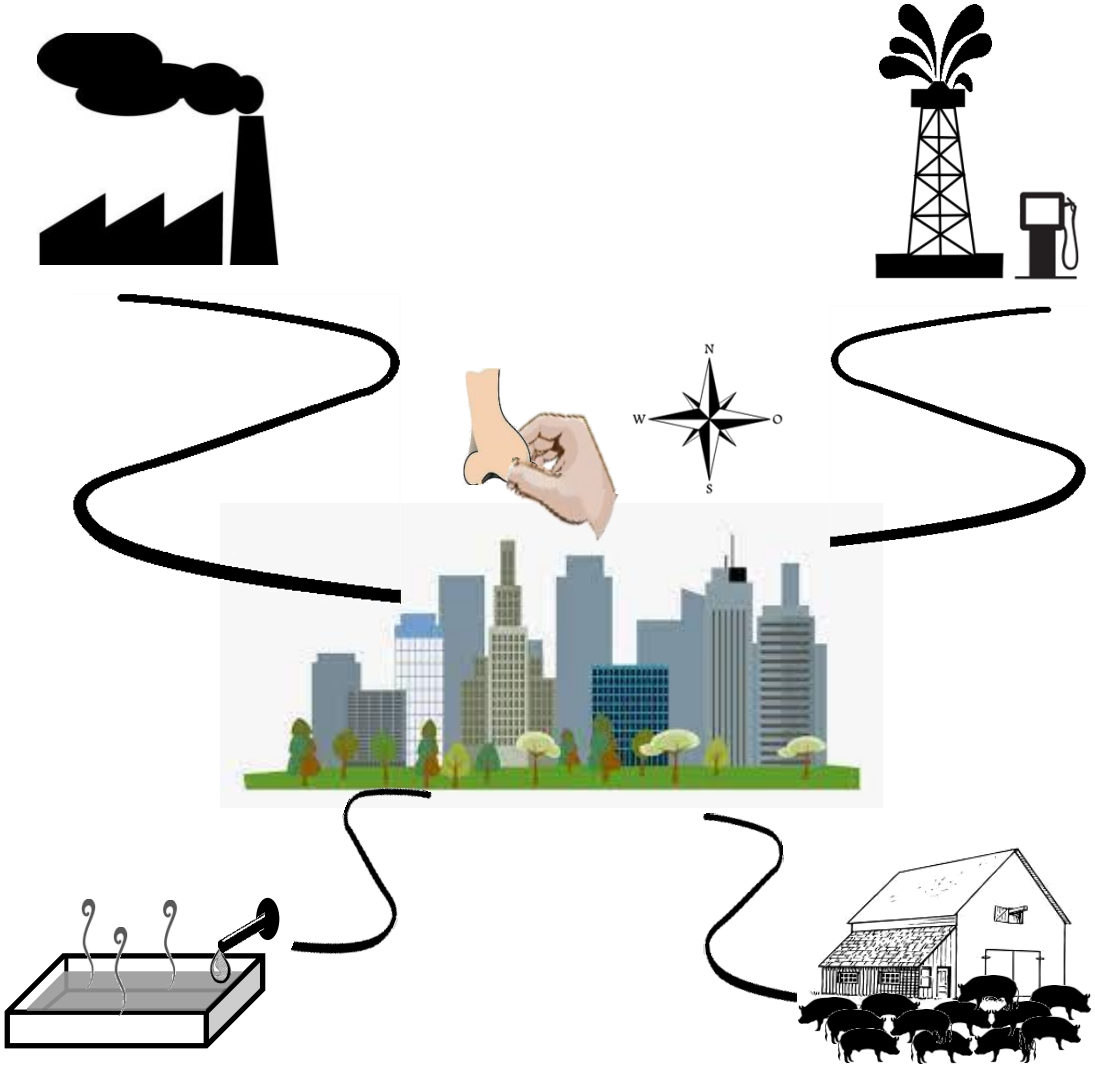
UNIVERSITA' DEGLI STUDI DI TRIESTE

Attività di ricerca principali:

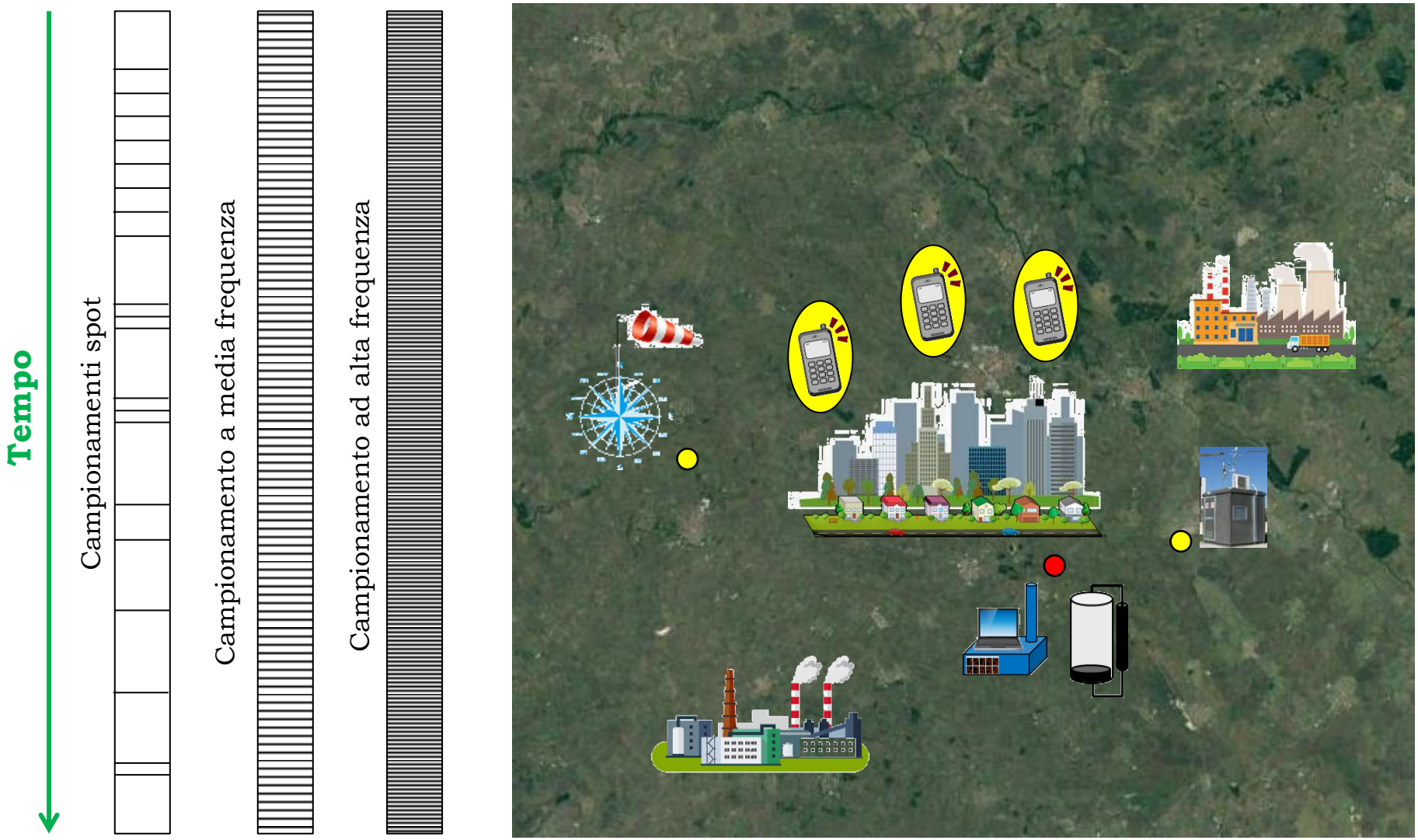
- Caratterizzazione di Composti Organici Volatili nelle valutazioni ambientali, per diagnosi cliniche e qualità della vita;
- Caratterizzazione di aerosol e bioarosol atmosferici e indoor;
- Sviluppo di metodi innovativi per l'analisi multivariata di dati da monitoraggio ambientale;
- Caratterizzazione di nanoparticelle nell'ambiente e nanotossicologia;
- Caratterizzazione di profili elementali di acque e sedimenti marini.

Individuazione della/e sorgente/i di odore

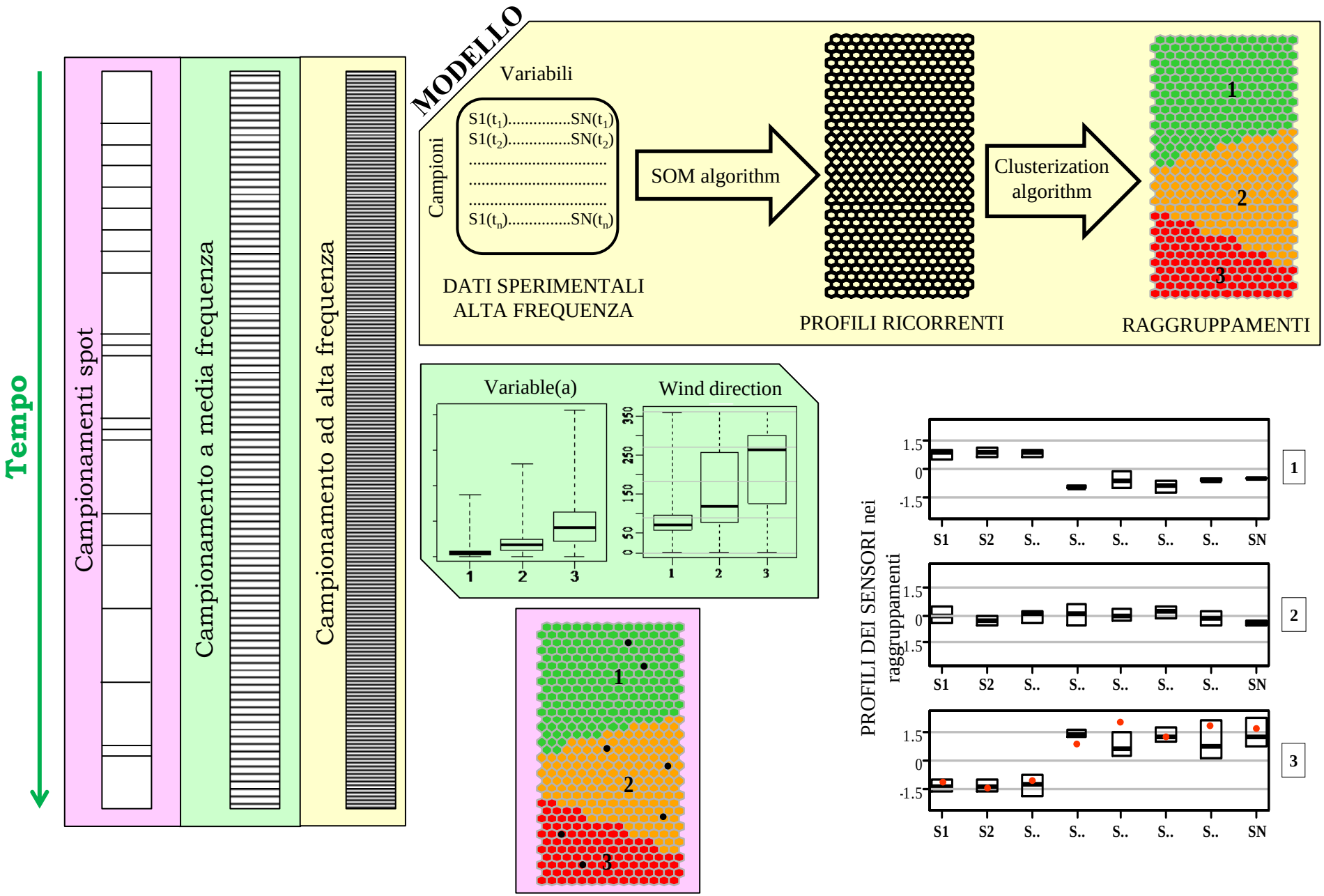
- Quale/i di questi?
- Quando?
- Per quanto tempo?



Dati registrati da IOMS e altri dati/informazioni disponibili



Integrazione dei dati disponibili



Algoritmo Self-Organizing Map (SOM)

- Famiglia delle reti neurali artificiali;
- **Tecnica di analisi multivariata NON supervisionata;**
- Basato sulla comparazione di distanze multidimensionali tra vettori di dati (es. distanza euclidea);
- **Il *noise* viene automaticamente ridotto durante il training del modello;**
- Si possono elaborare **«big data»** (es. monitoraggio annuale dati al minuto = 525’600 campioni);
- **I risultati sono visualizzabili ed esplorabili in mappe bidimensionali**

Il modello **«impara»** dai dati sperimentali

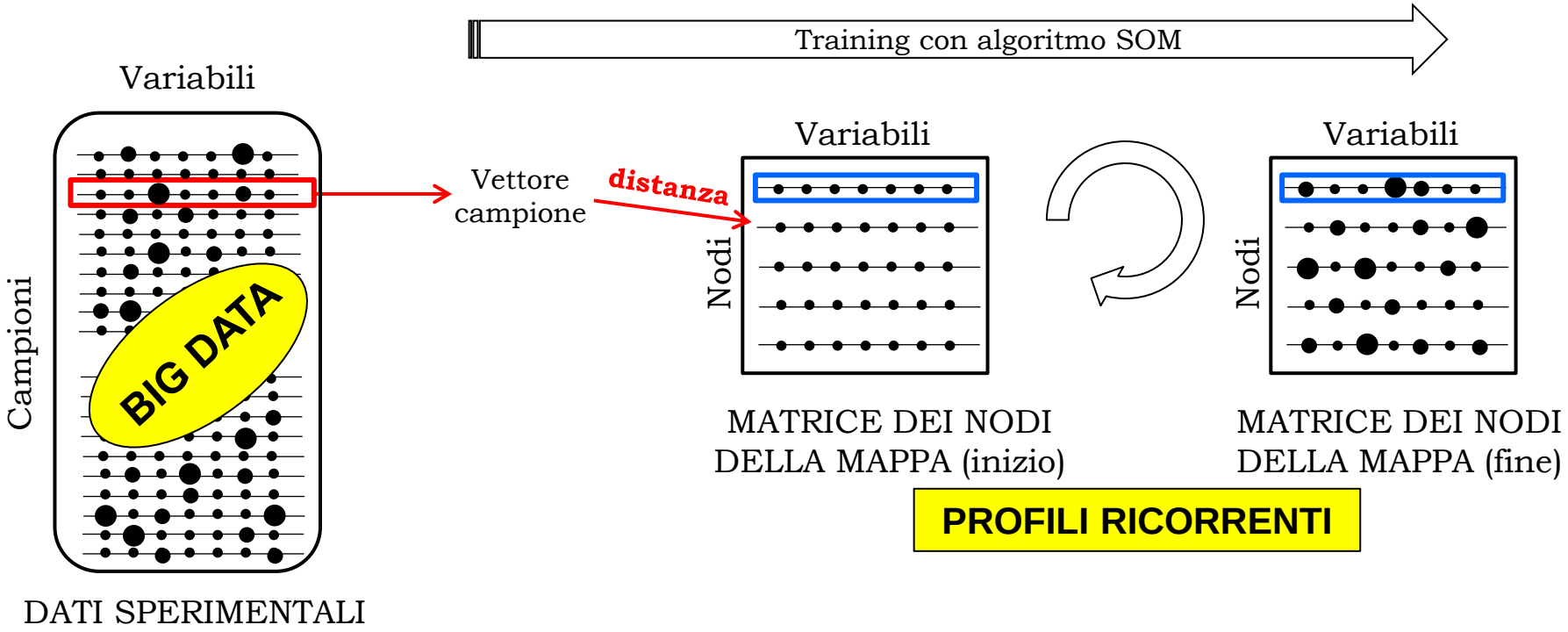
Non necessaria conoscenza a priori sulla **classificazione** dei dati

Può rappresentare anche relazioni **non lineari** tra variabili

Non è necessario pre-processamento per ridurre il rumore

T. Kohonen, Appl. Opt., 1987, 26, 4910. https://doi.org/10.1364/ao.26.004910
J. Himberg, J. Ahola, E. Alhoniemi, J. Vesanto, O. Simula 2001, pp. 38–65. https://doi.org/10.1142/9789812811691_0002

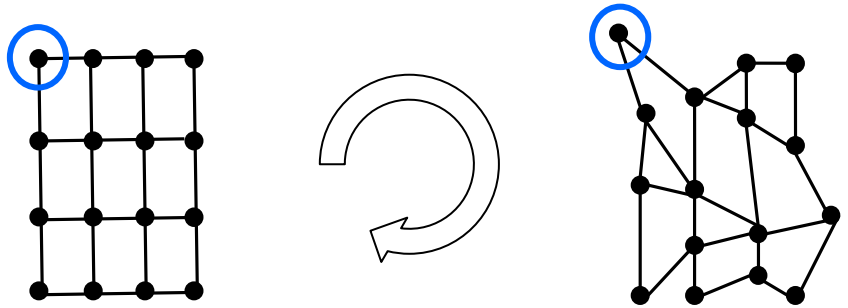
Come funziona l'algoritmo SOM



MANTIENE:

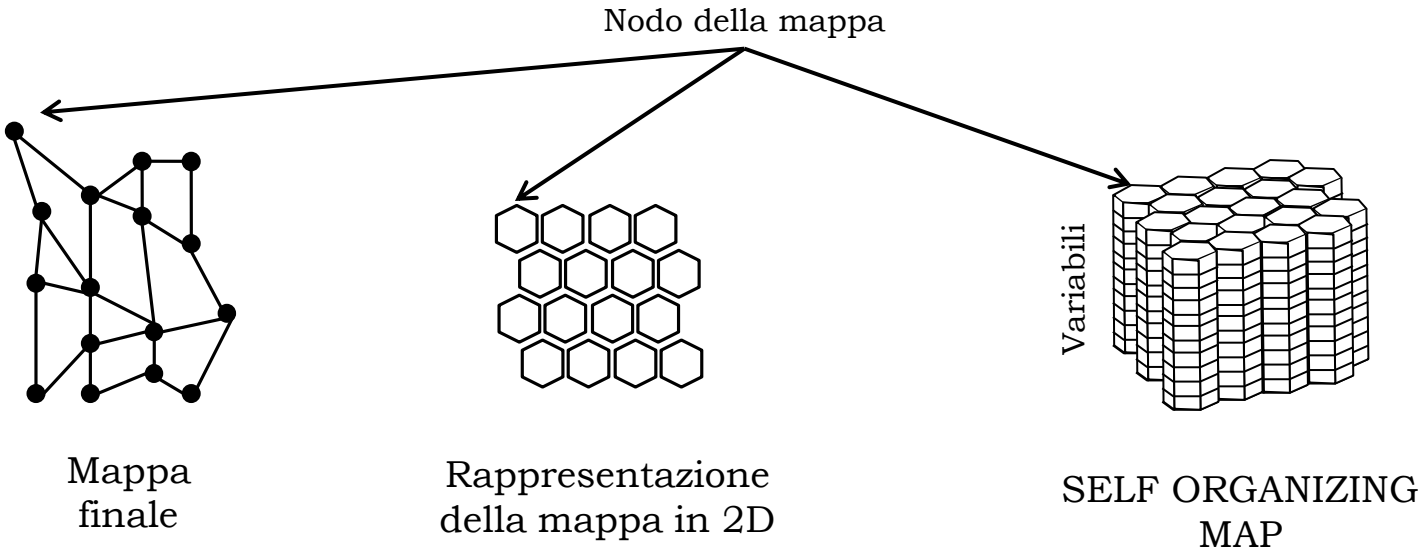
a) **Tutte le variabili sperimentali;**

b) **La variabilità dei dati sperimentali (eccetto il *noise*).**



T. Kohonen, Appl. Opt., 1987, 26, 4910. https://doi.org/10.1364/ao.26.004910

Come funziona l'algoritmo SOM (2)

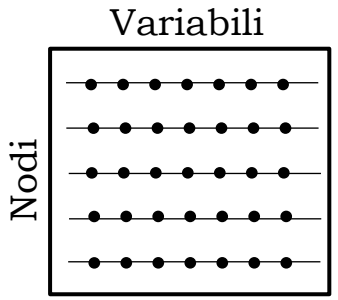


T. Kohonen, Appl. Opt., 1987, 26, 4910. https://doi.org/10.1364/ao.26.004910

Inizializzazione della mappa

L'utilizzatore deve scegliere:

- il **numero di nodi** della mappa;
- il **rapporto tra le dimensioni** della mappa (base x altezza);
- i **valori** delle variabili con cui inizializzare la matrice prima del training.



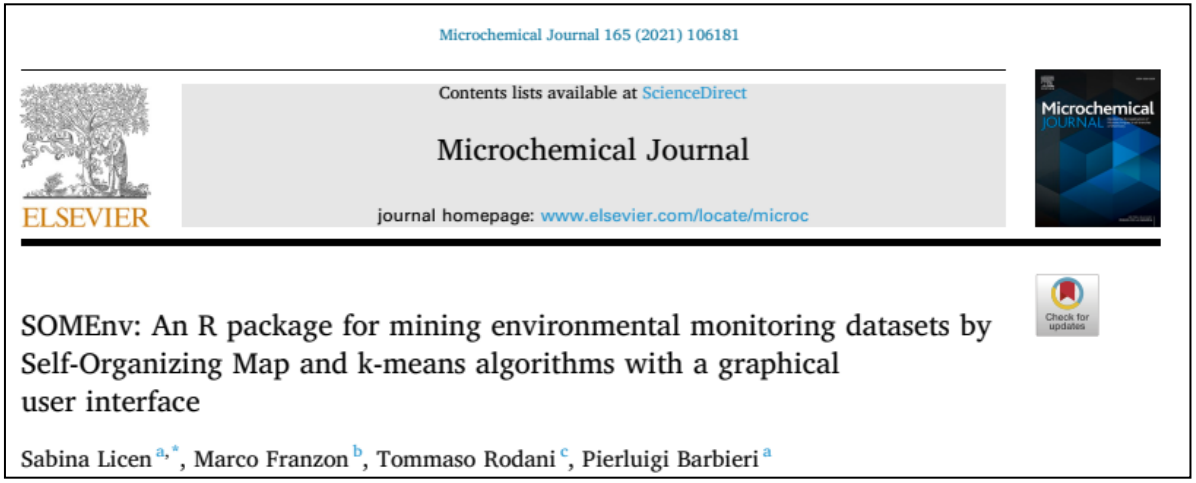
Soluzione:

- regole euristiche di Vesanto, basate su autovalori e autovettori della matrice di dati.

J. Vesanto, Intell. Data Anal., 1999, 3, 111–126 https://doi.org/10.3233/IDA-1999-3203

SOMEnv package

- Basato sull'algoritmo SOM presente nel **kohonen package** di Wehrens, Buydens & Kruisselbrink (2007-2018)
- Regole euristiche di **Vesanto** calcolate automaticamente (fonte: SOMtoolbox per MATLAB)
- Interfaccia Grafica (**GUI**) che si apre nel **browser** predefinito;
- Risultati esplorabili attraverso diversi tipi di **grafici**;
- Facilità di **download** di grafici e risultati.
- ***https://cran.r-project.org/web/packages/SOMEnv/index.html***



Special Issue:

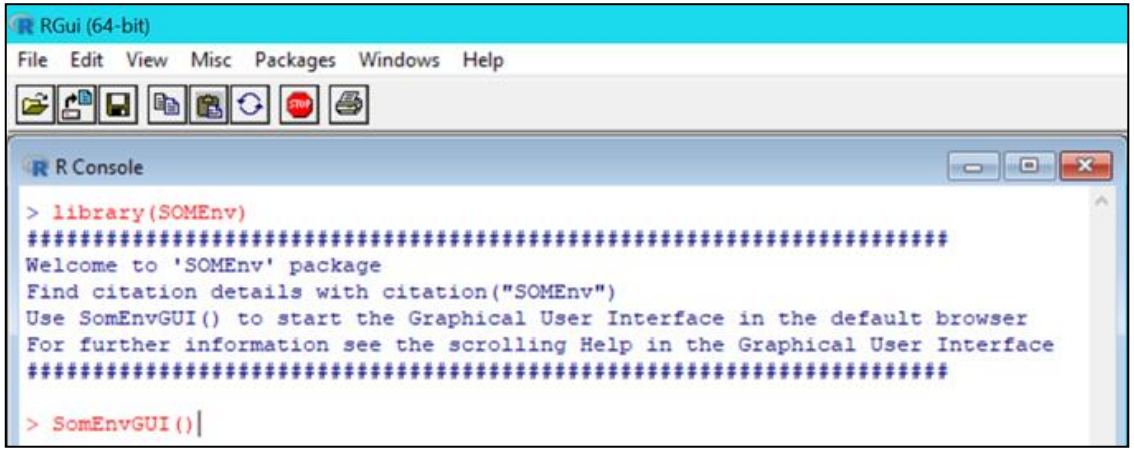
"*Chemometrics in Environmental Analytical Chemistry*"

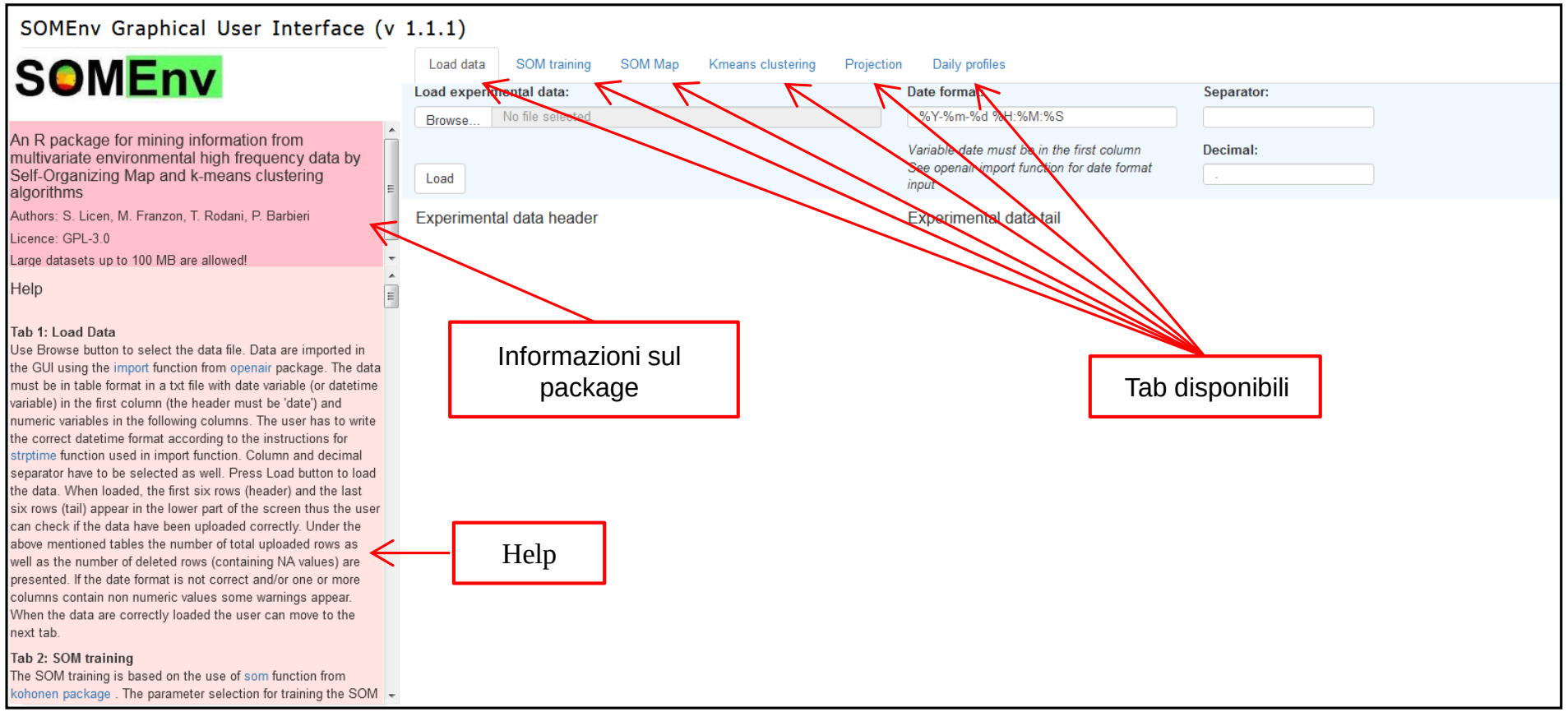
Licen, S., Franzon, M., Rodani, T., Barbieri, P., 2021 Microchem. J. 165. https://doi.org/10.1016/j.microc.2021.106181

SOMEnv package loading

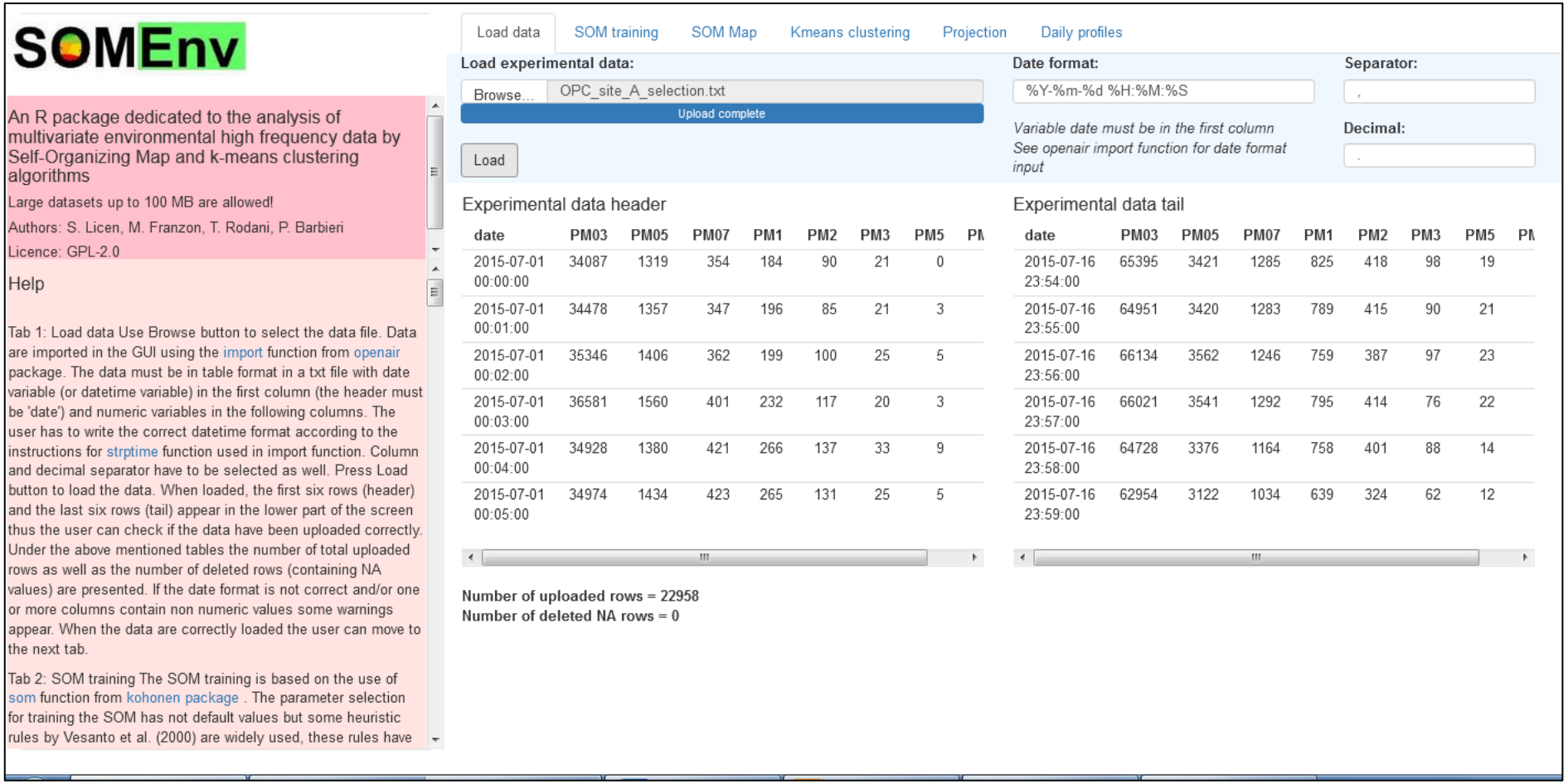
https://cran.r-project.org/

> install.packages("SOMEnv")

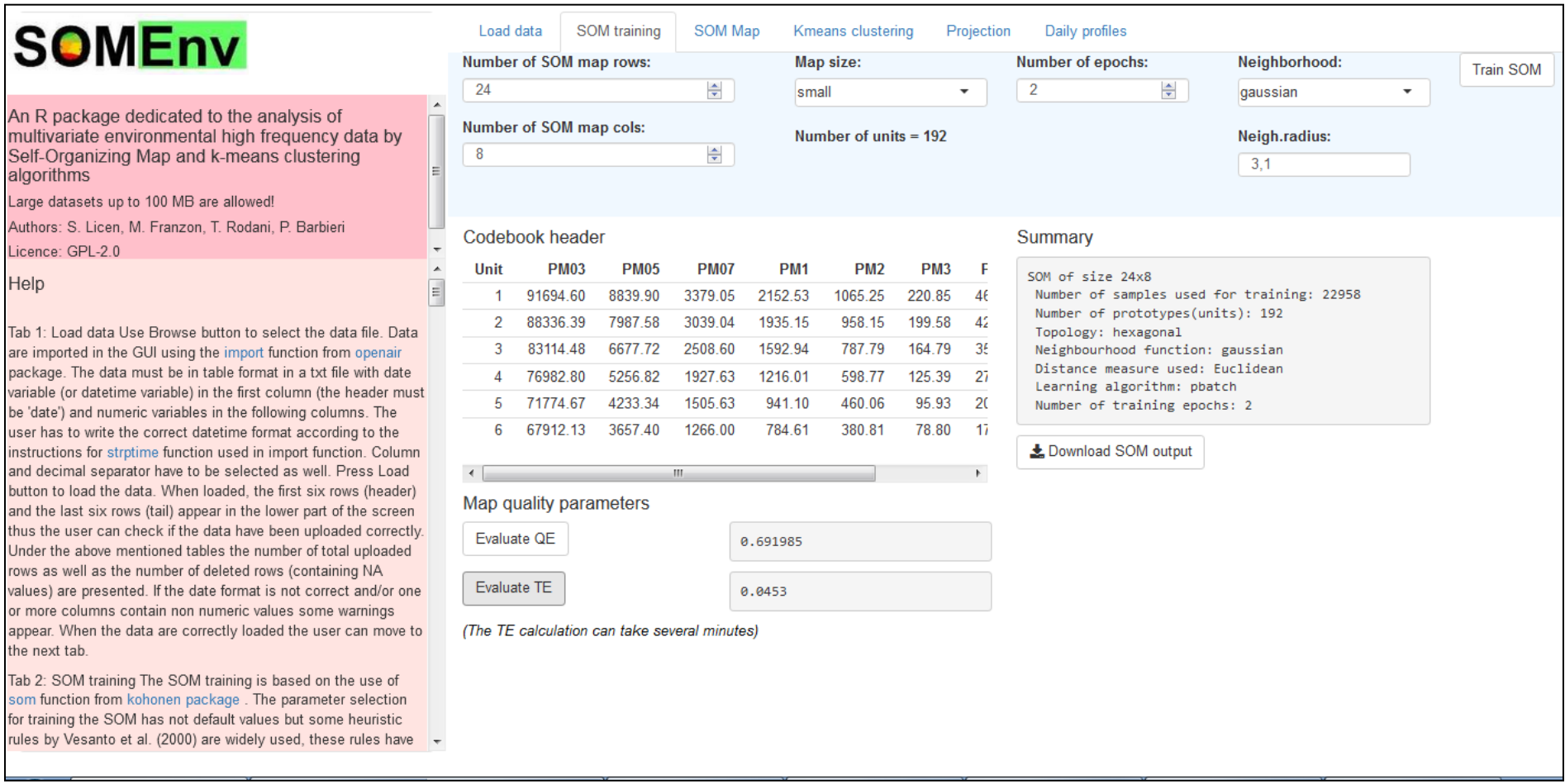




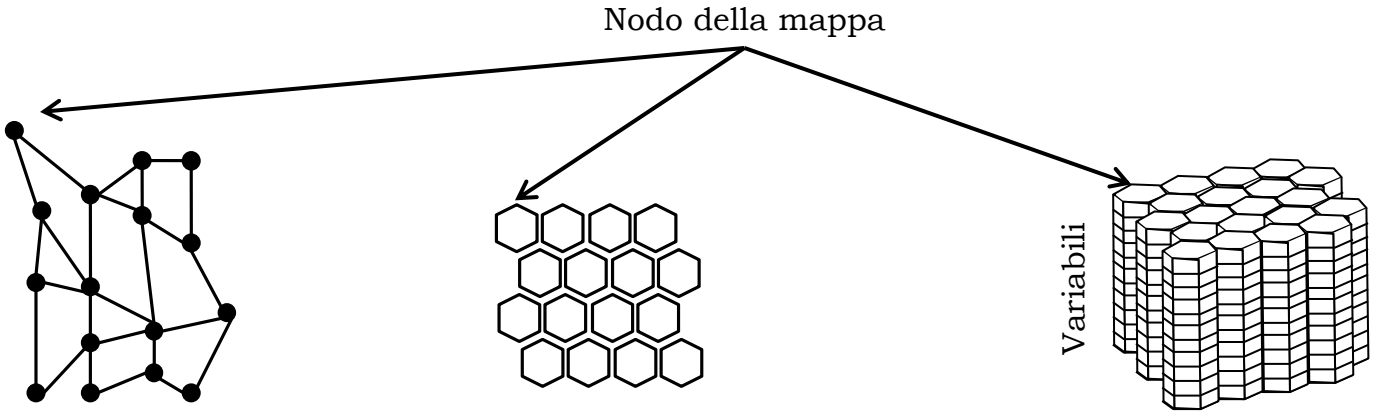
SOMEnv package: Data loading Tab

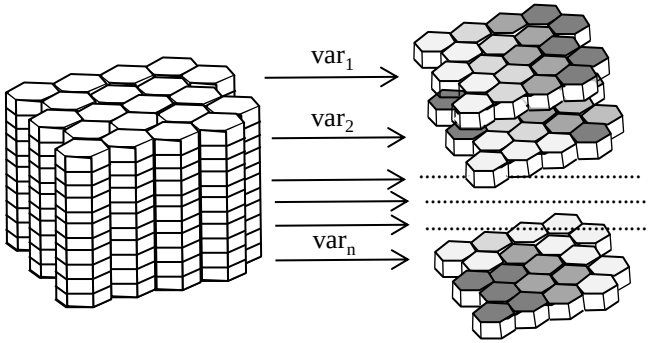


SOMEnv package: SOM training Tab

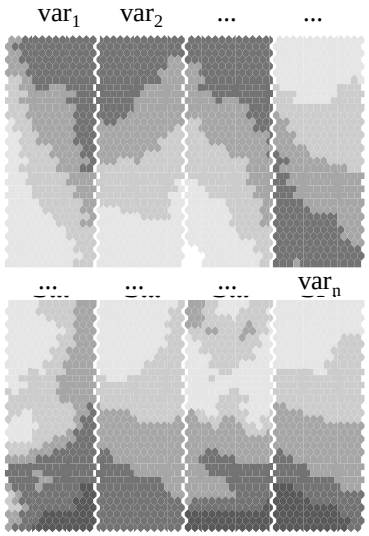


Relazioni tra variabili: esplorare le heatmaps



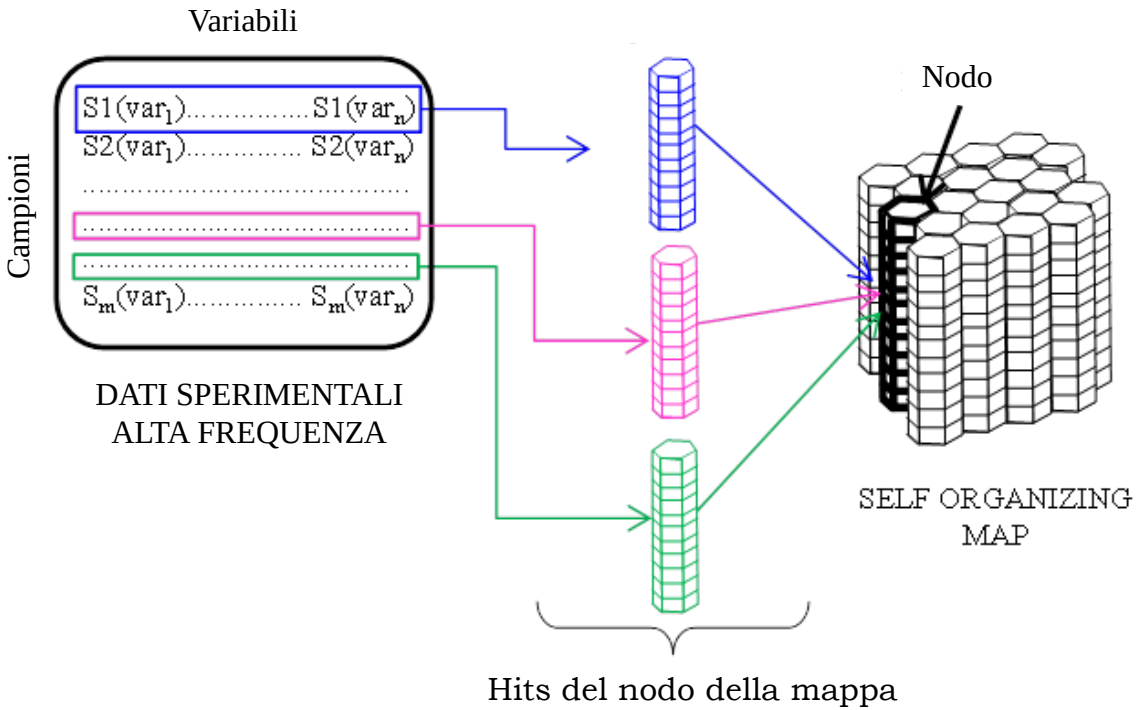


HEATMAP
Distribuzione dei valori modellati di una singola variabile

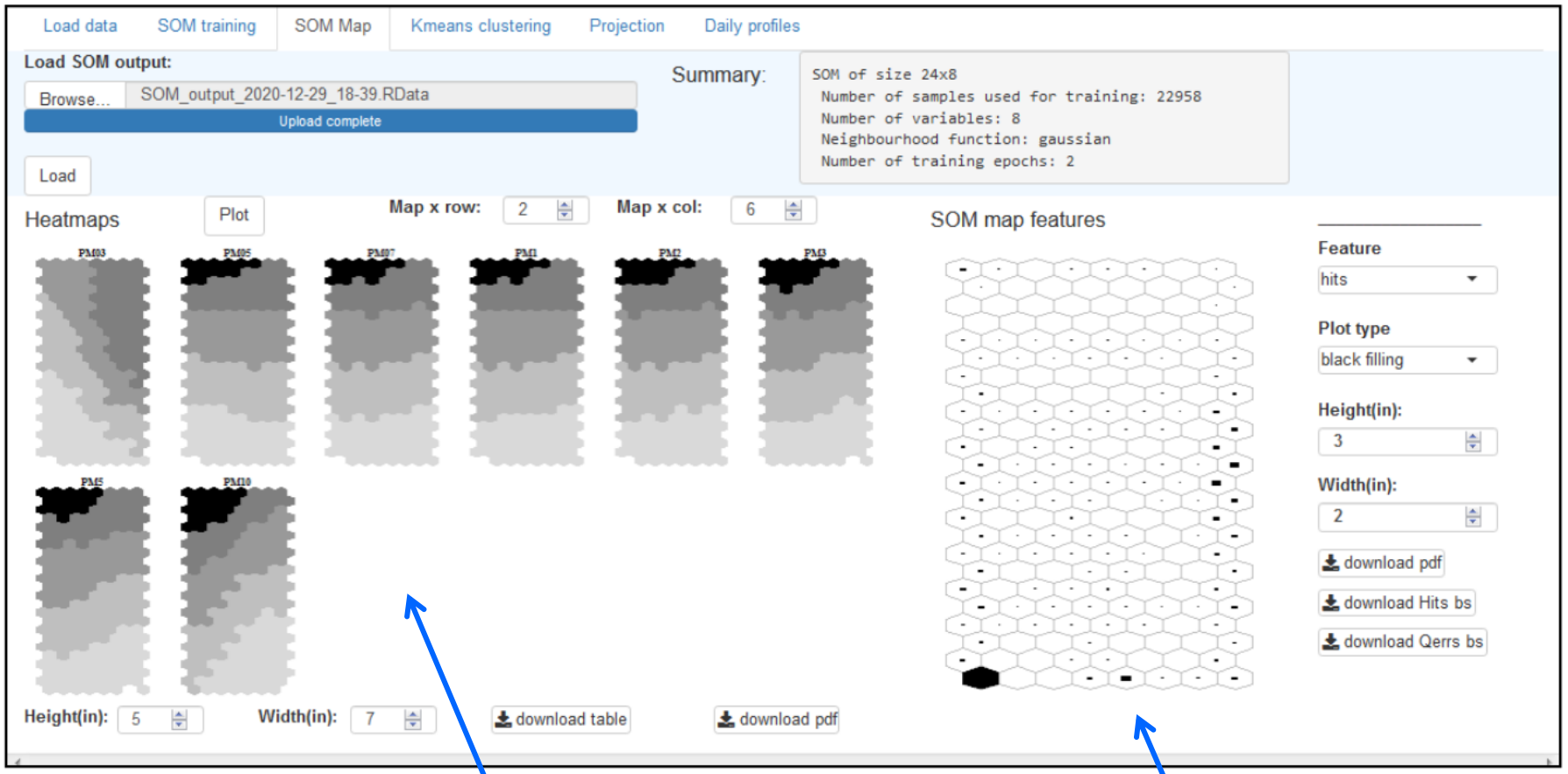


T. Kohonen, Appl. Opt., 1987, 26, 4910. https://doi.org/10.1364/ao.26.004910

Quanti dati sperimentali sono rappresentati da ogni profilo ricorrente? Hits



SOMEnv package: SOM map (visualization) Tab

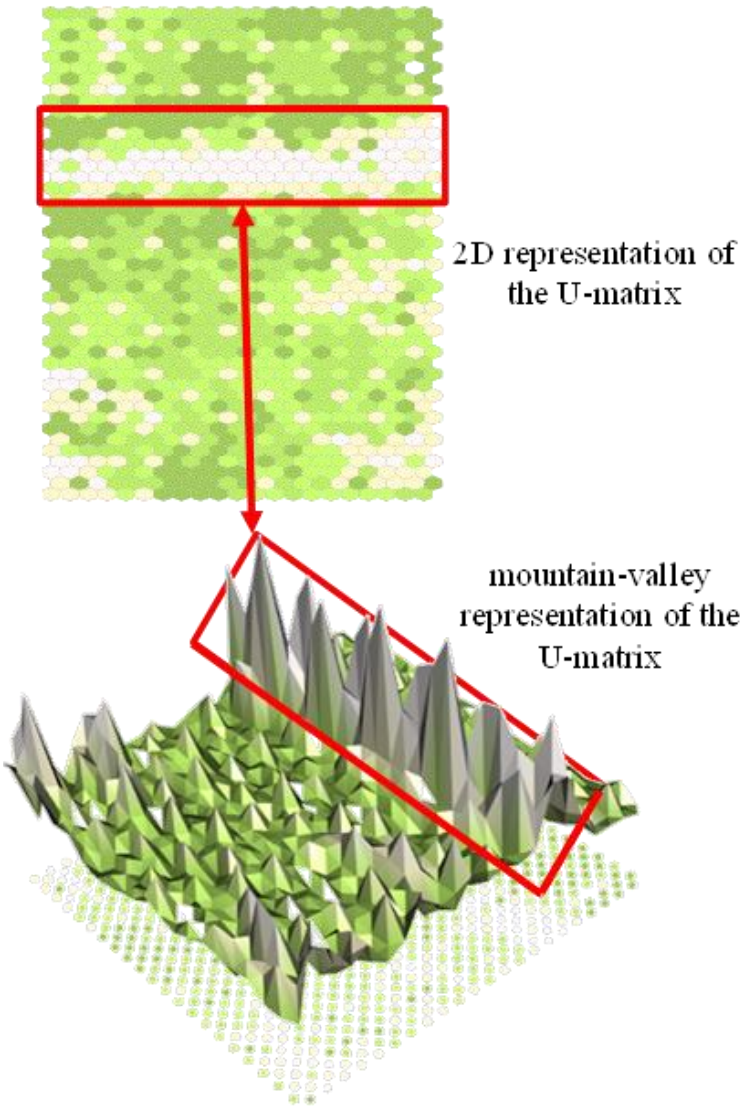


Heatmaps

(scala di grigi da bianco a nero secondo i quartili)

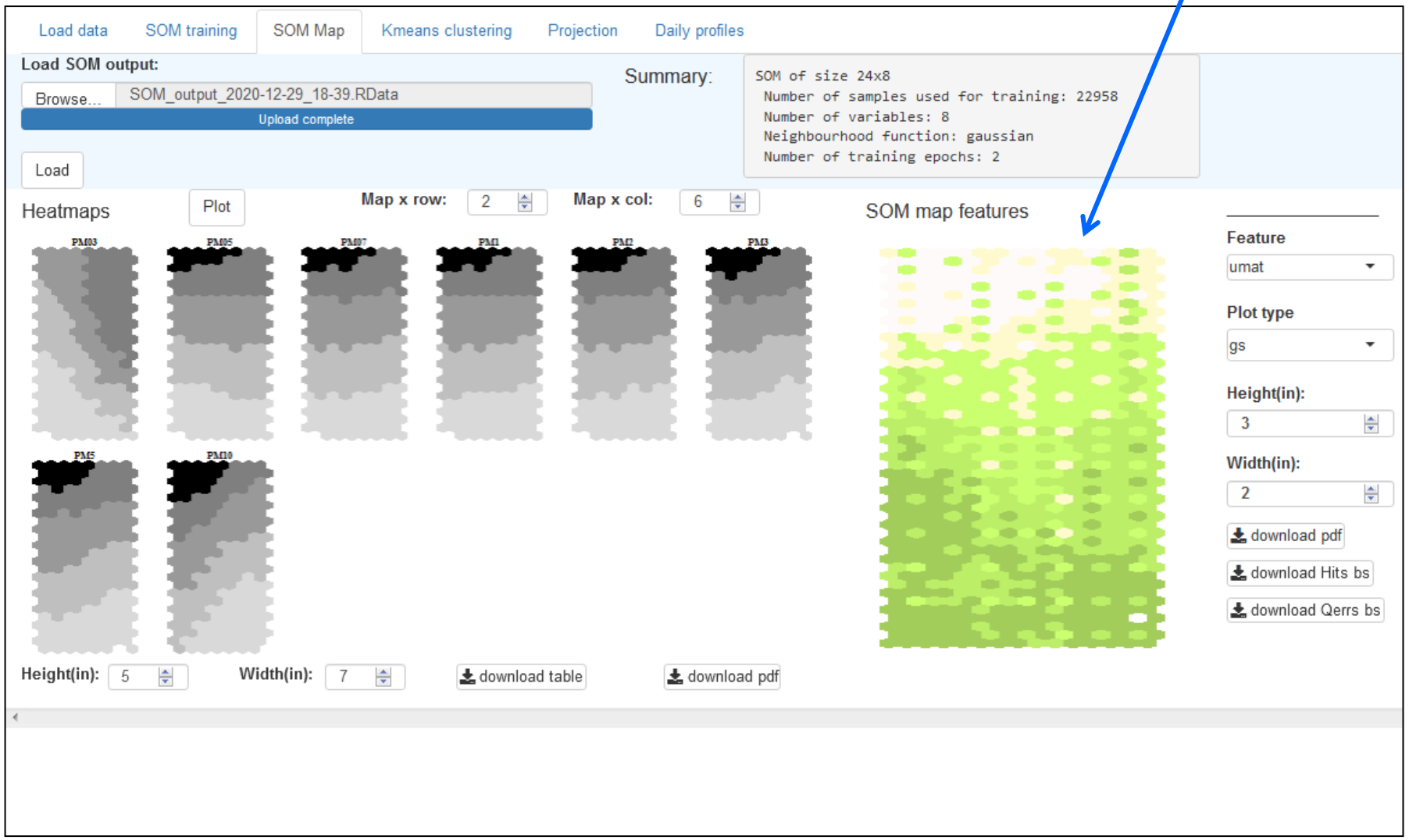
Hits

Quanto sono diversi tra loro i profili ricorrenti? Unified Distance Matrix (U-matrix)

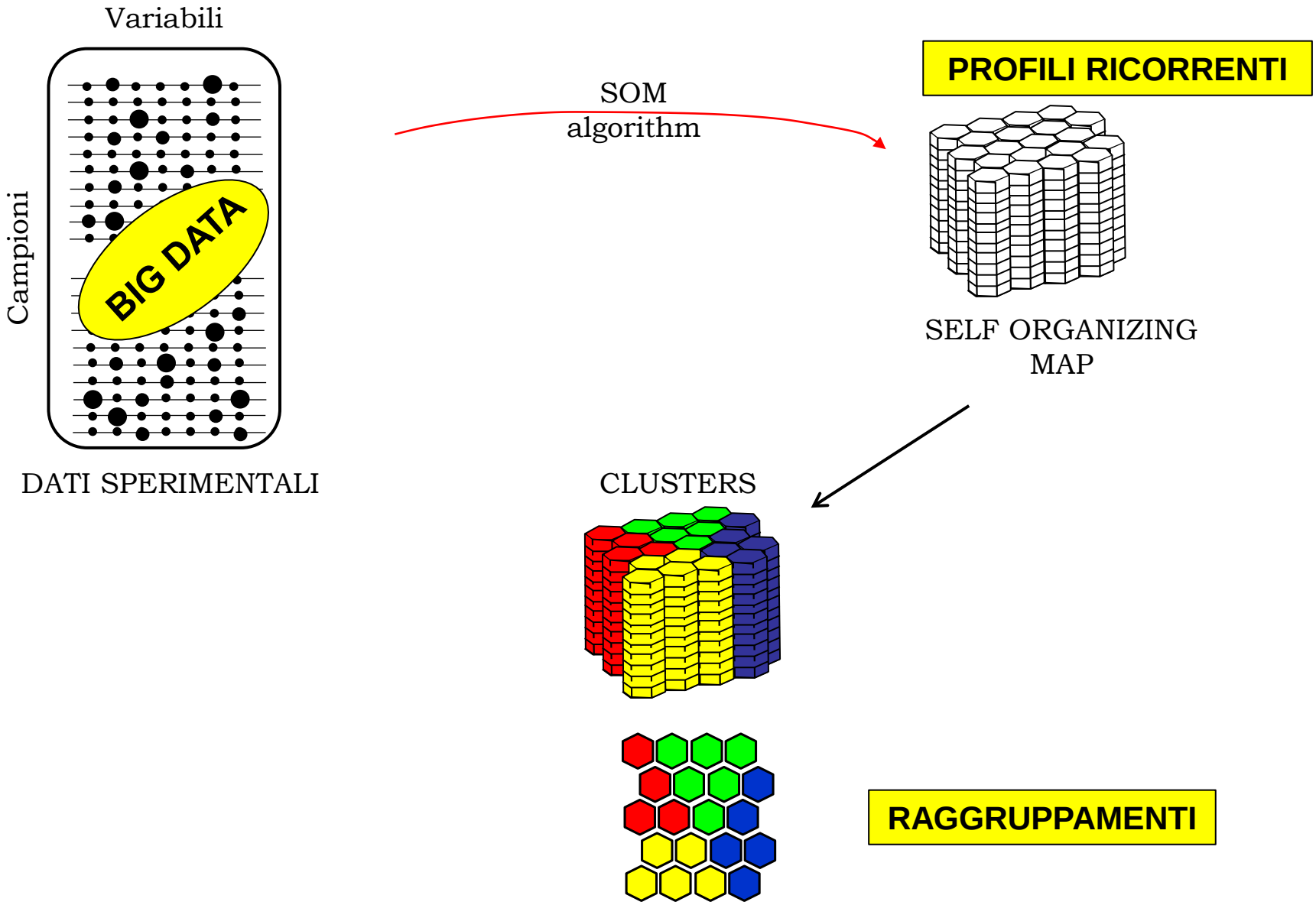


SOMEnv package: SOM map (visualization) Tab

Unified distance matrix



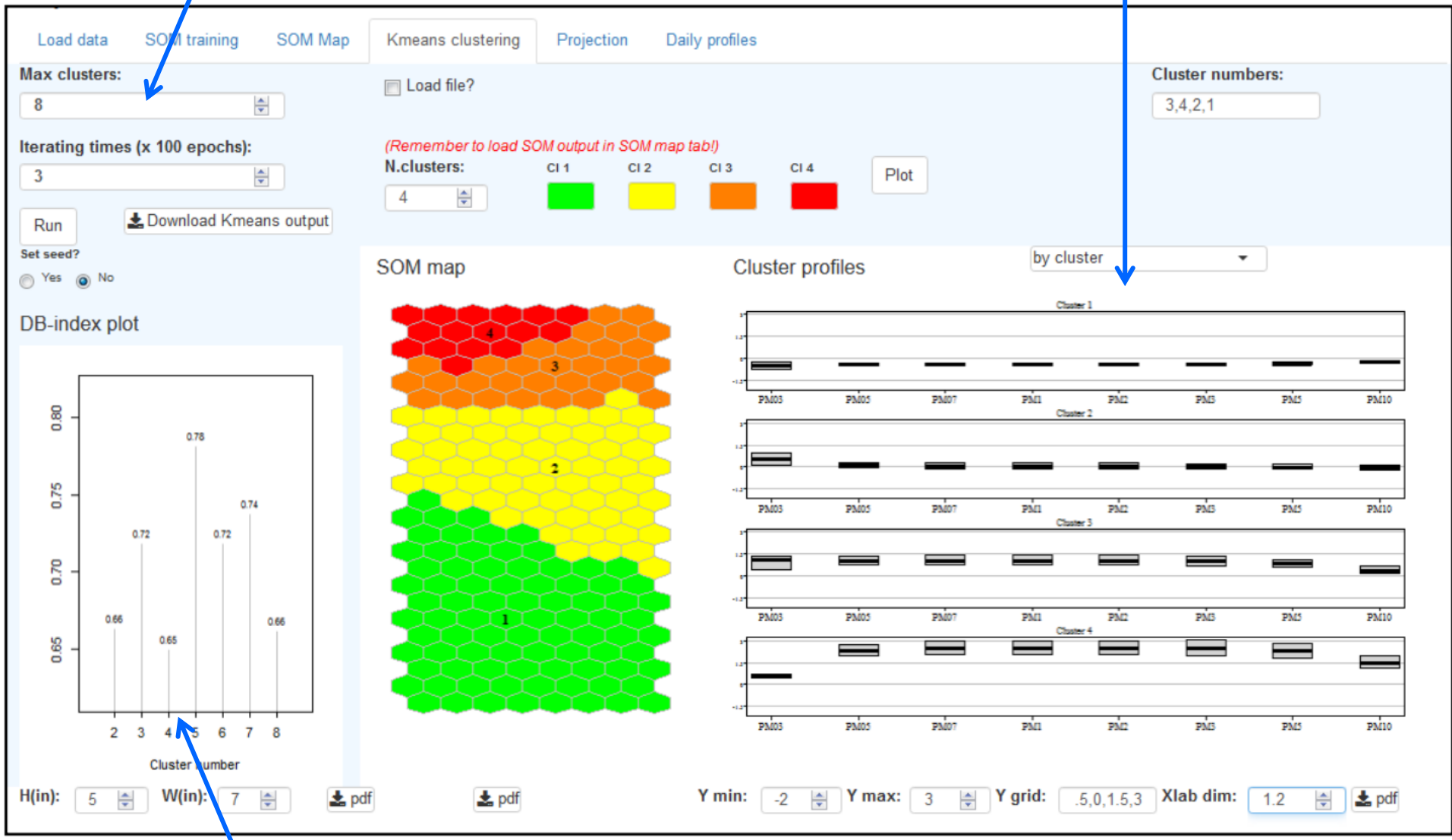
Raggruppamento di profili ricorrenti: k-means clustering



SOMEnv package: K-means clustering Tab

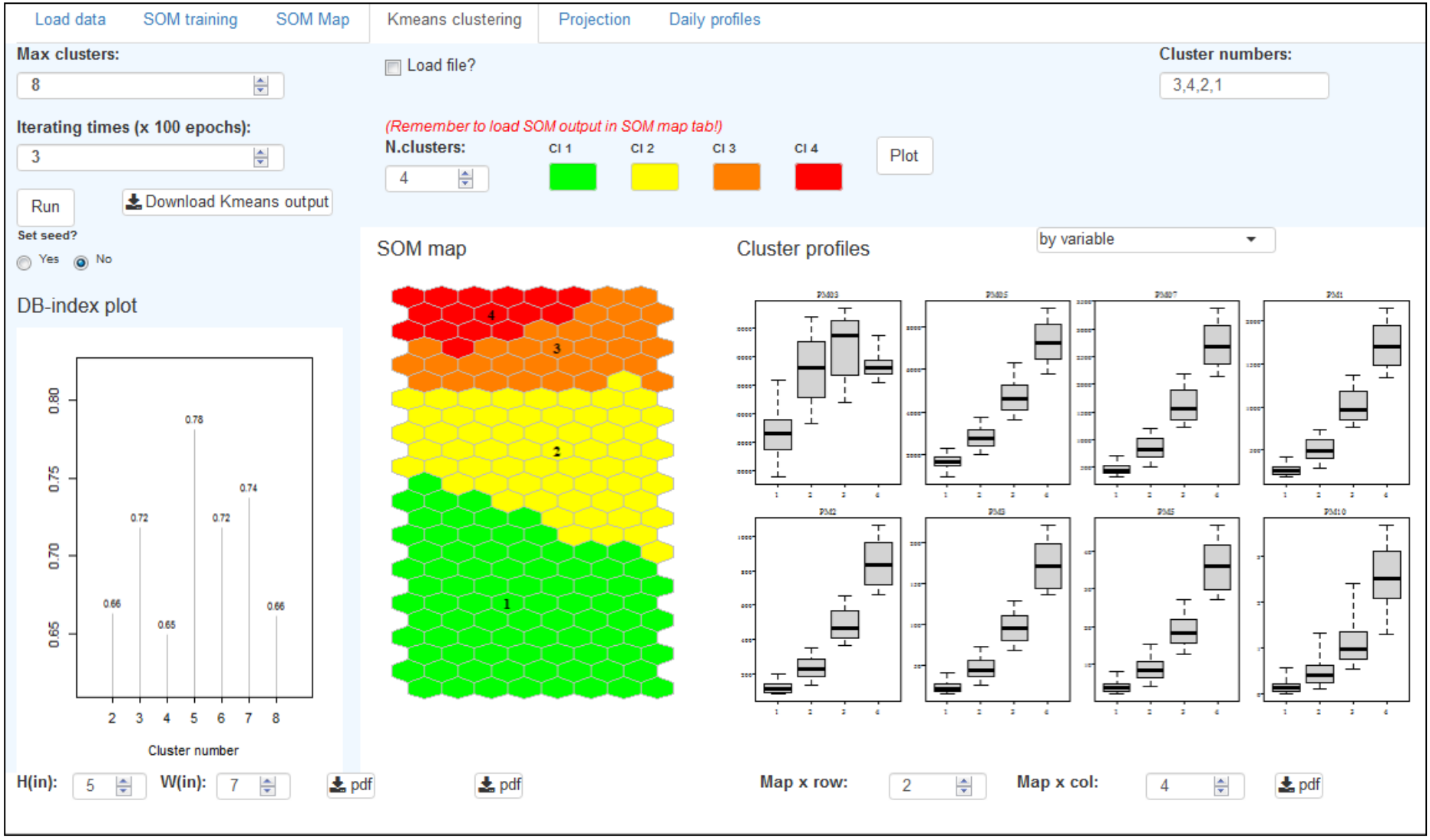
K-means clustering training

Profili delle variabili modellate nei raggruppamenti



Davis Bouldin index: indice di migliore raggruppamento

SOMEnv package: K-means clustering Tab



SOMEnv package: daily profiles Tab

Load data | SOM training | SOM Map | Kmeans clustering | Projection | Daily profiles

Dataset

training

Starting date:

01/07/2015

Ending date:

16/07/2015

Obs/Day

1440

Select sample:

2015-07-07 00:01:00

Units cluster assignment

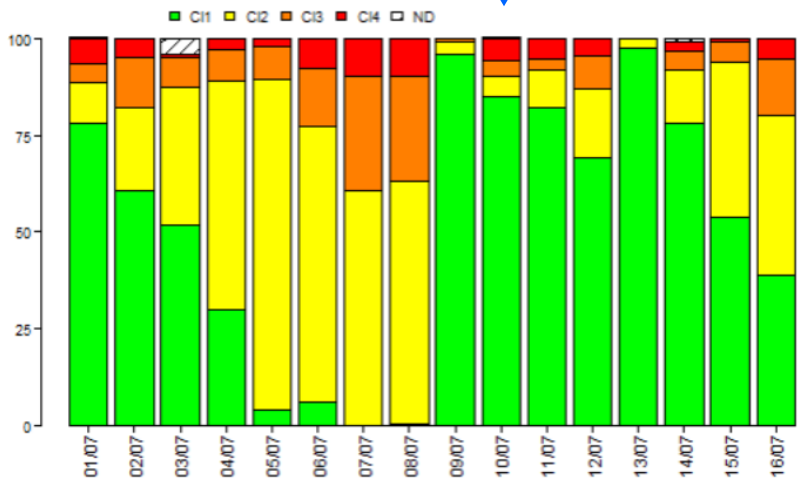
Data cluster assignment

```
BMU = 1
Cluster = 4
Qe = 8.751
```

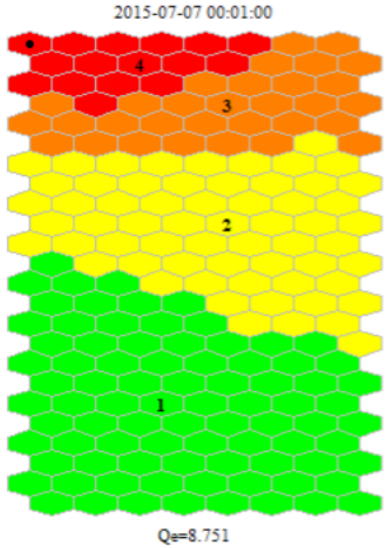
% giornaliera di presenza di ogni raggruppamento

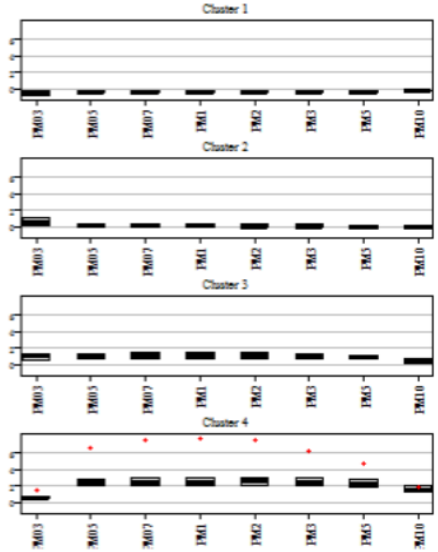
Daily graph

Xlab dim: 1



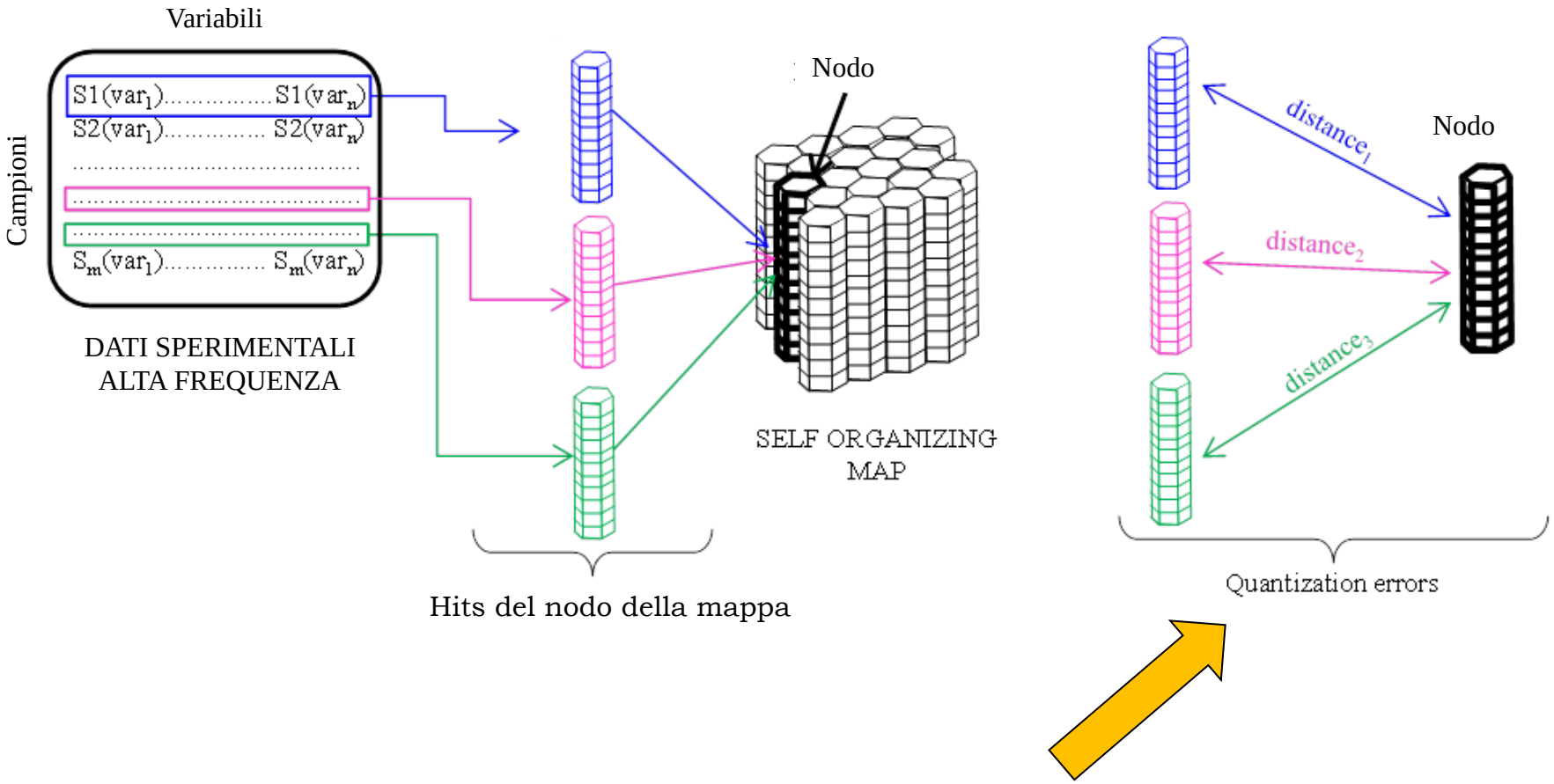
Sample profile





H(in): 5 **W(in):** 7 pdf table overall tab monthly tab pdf

Individuare outliers: Quantization Errors



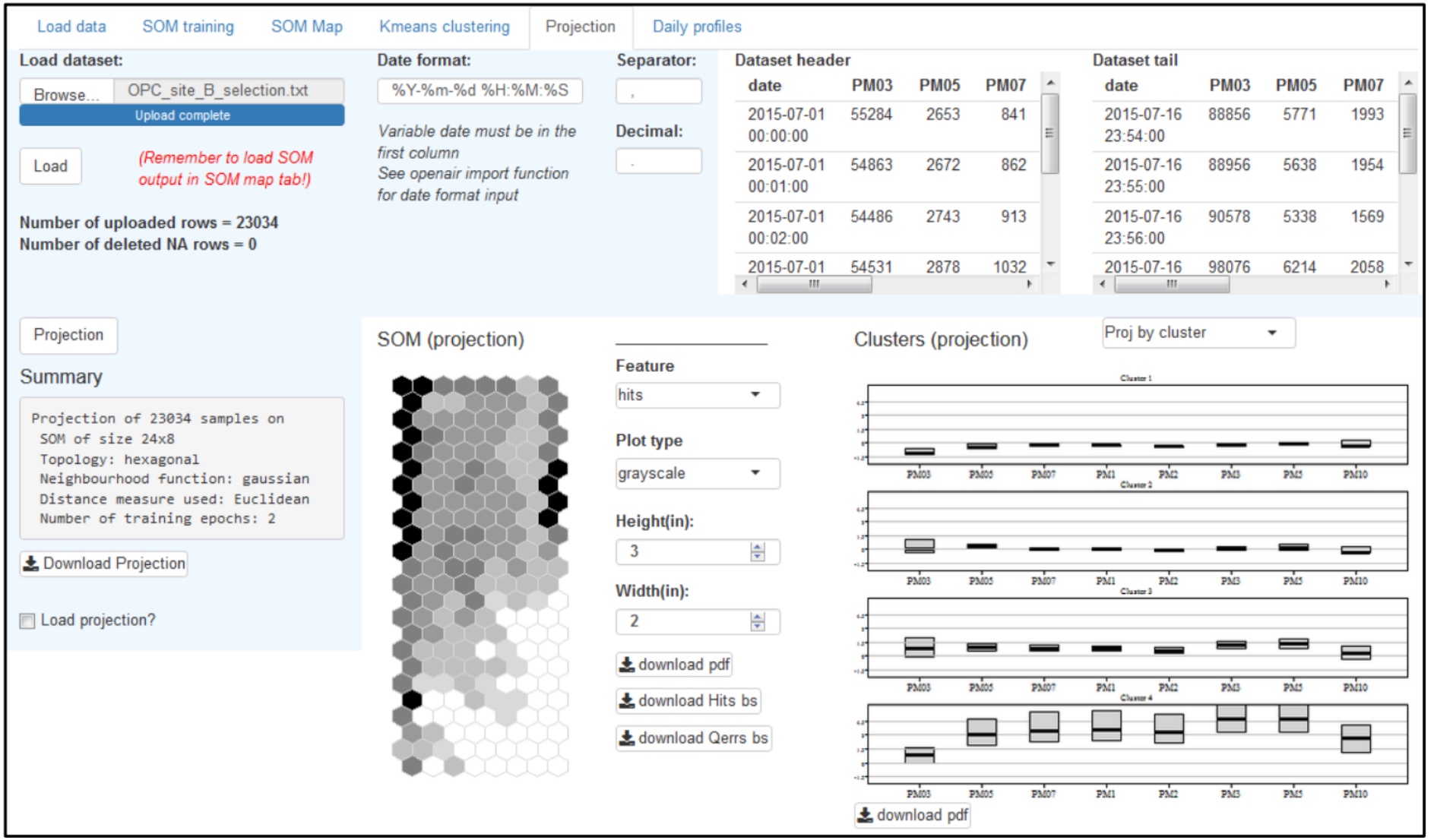
SOMEnv package: daily profiles Tab

esplorazione del singolo dato (vettore campione) registrato

SOMEnv package: projection Tab

Proiettare dati esterni nel modello:

- dati registrati in tempi successivi;
- dati rilevati alle sorgenti



Conclusioni

SOMEnv package per R software è un pacchetto ***free*** che consente di individuare:

(i) profili ricorrenti di odori;

(ii) profilo dell'aria di background;

(iii) frequenze della presenza dei diversi profili di odori;

(iv) possibili outiler o sensor fault;

e di comparare:

(i) i profili ricorrenti con quelli delle sorgenti;

(ii) i raggruppamenti con dati esterni, es. segnalazioni, campionamenti puntuali, dati registrati da altri strumenti.

GRAZIE PER L'ATTENZIONE

slicen@units.it

https://www.researchgate.net/profile/Sabina-Licen-3