



Agenzia nazionale per le nuove tecnologie,
l'energia e lo sviluppo economico sostenibile

SOM: minimum number of days for the annual average

Felicita Russo¹, Daniela Barbero², Gianni Tinarelli² e Maria Gabriella Villani¹

1-ENEA, Italian National Agency for New Technologies, Energy and Sustainable Economic Development, SSPT-MET-INAT, 40129 Bologna, Italy.
2-ARIANET, Via Gian Giacomo Gilino, 9, 20128 Milan, Italy

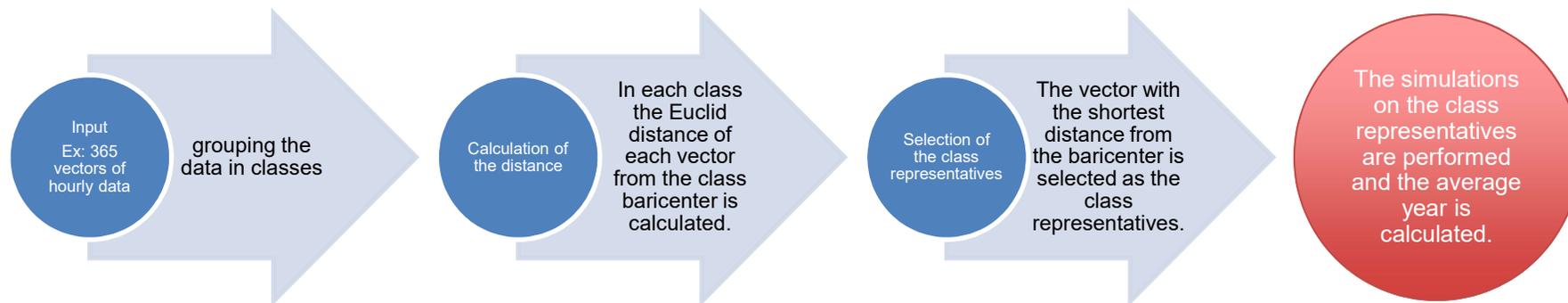


1101 0110 1100
0101 0010 1101
0001 0110 1110
1101 0010 1101
1111 1010 0000



Self Organizing Maps (SOM)

The **SOM** algorithm is a **classification method** based on **neural networks**. It allows the classification of different types of data. NCAR is using this method to classify meteorological regimes (using WRF 3D output for example) or typical days (using meteorological station measurements). We use a SOM algorithm provided by **ARIA Technologies SA** with which Arianet is closely collaborating.



Once the class representatives are simulated with PMSS, to calculate annual means and percentiles two ways of reconstructing the annual average are available:

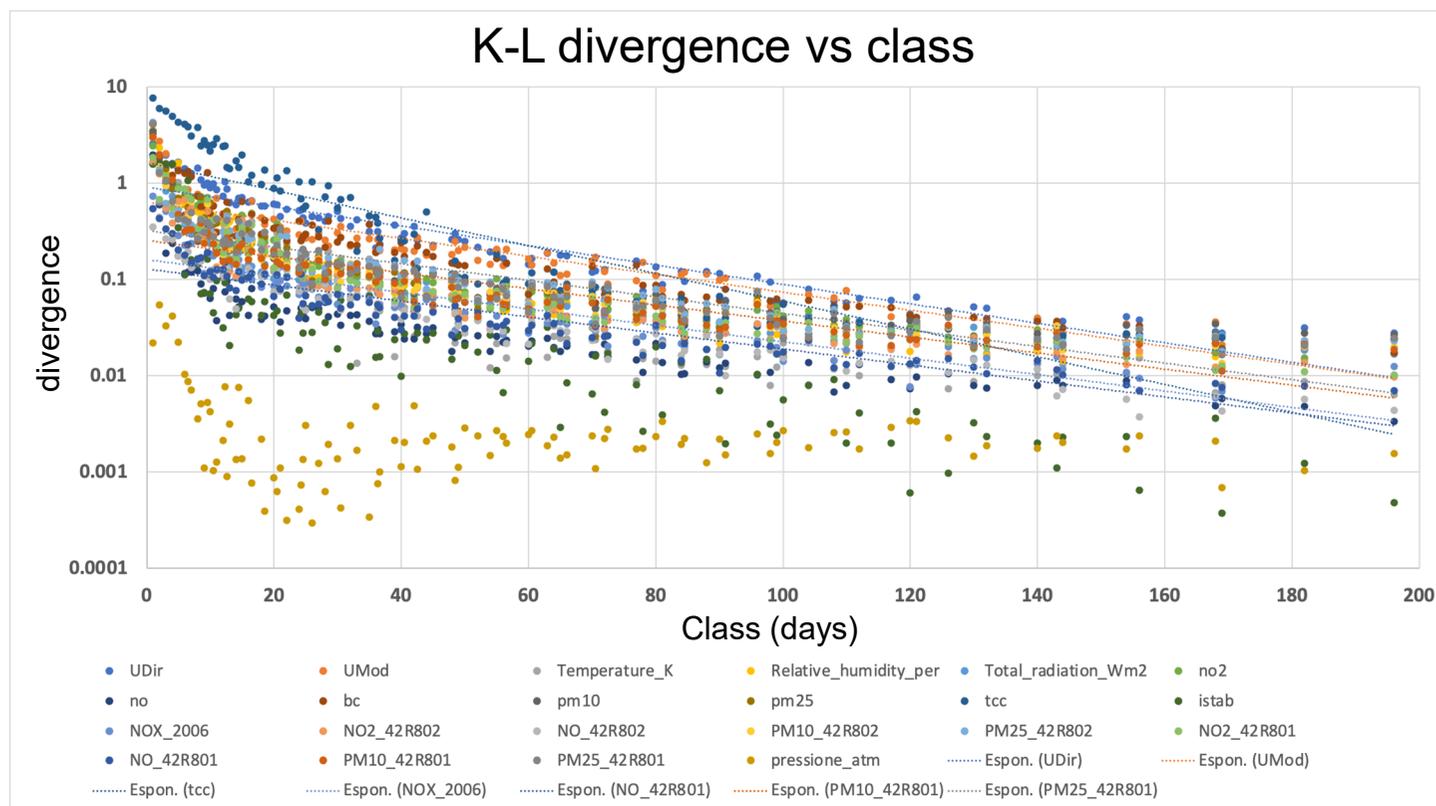
BEST: the average is calculated by averaging the simulations on class representative days.

POND: each day of the year is constructed as weighted average of the class representatives and the yearly average is calculated.

SOM classification for Anversa

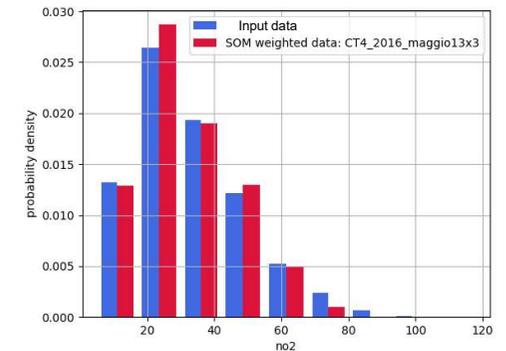
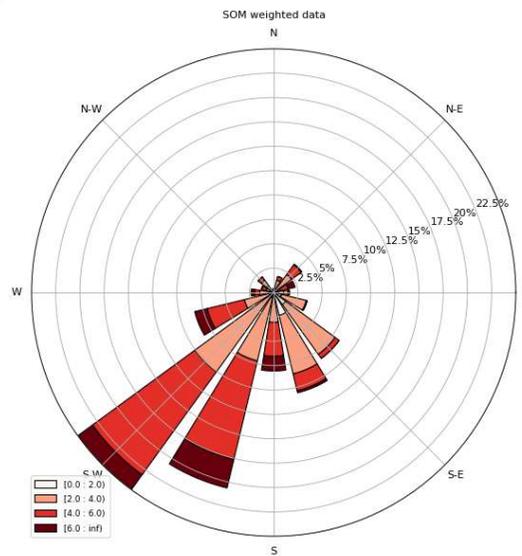
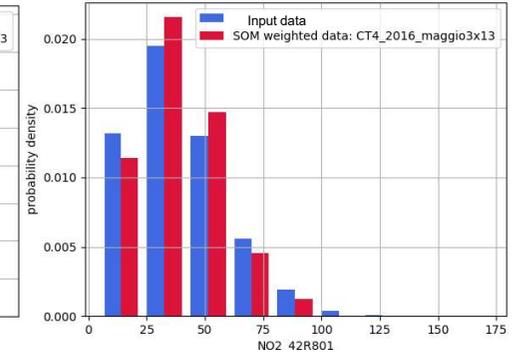
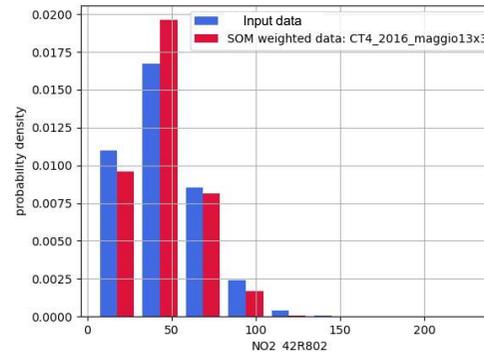
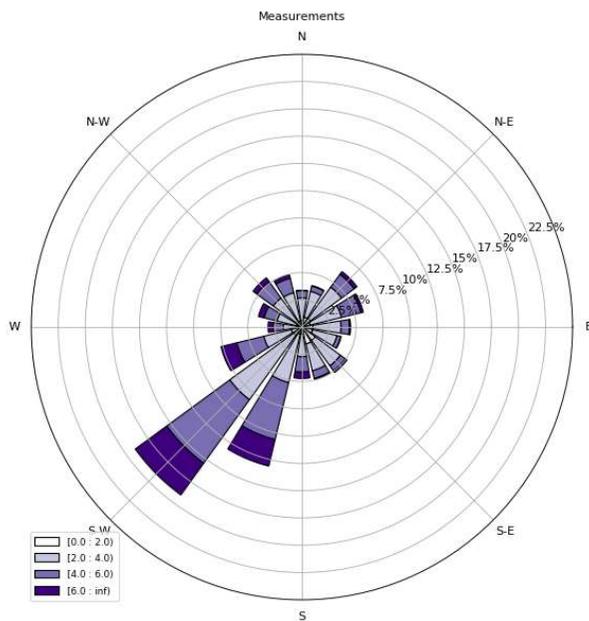
In our first inspection of the tool we used all the available quantities as input to the algorithm.

The choice of the class with minimum divergence is subjective, but if we consider 0.1 as a divergence acceptably small, any class around 40 or larger can be considered acceptable.



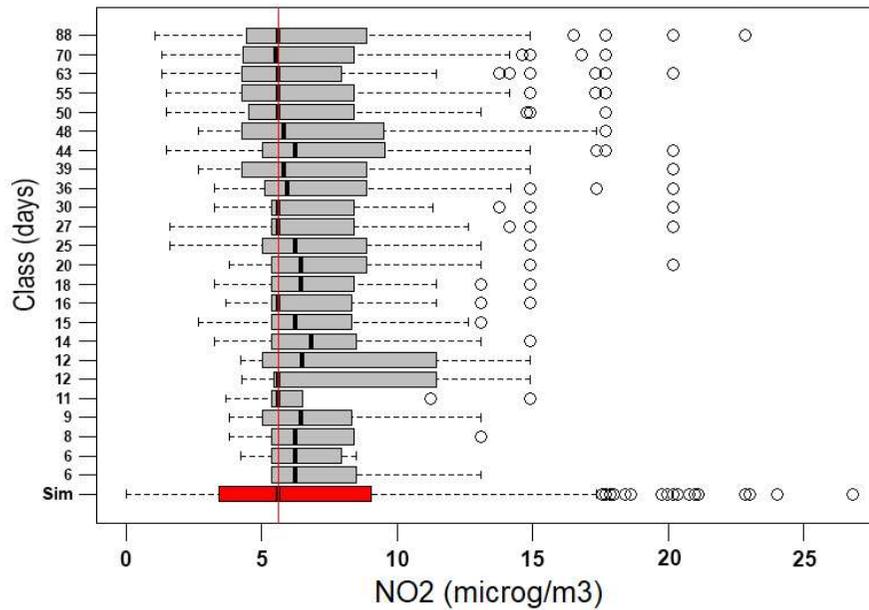
Automatic reconstruction with 39 class representatives

- POND postprocessing -

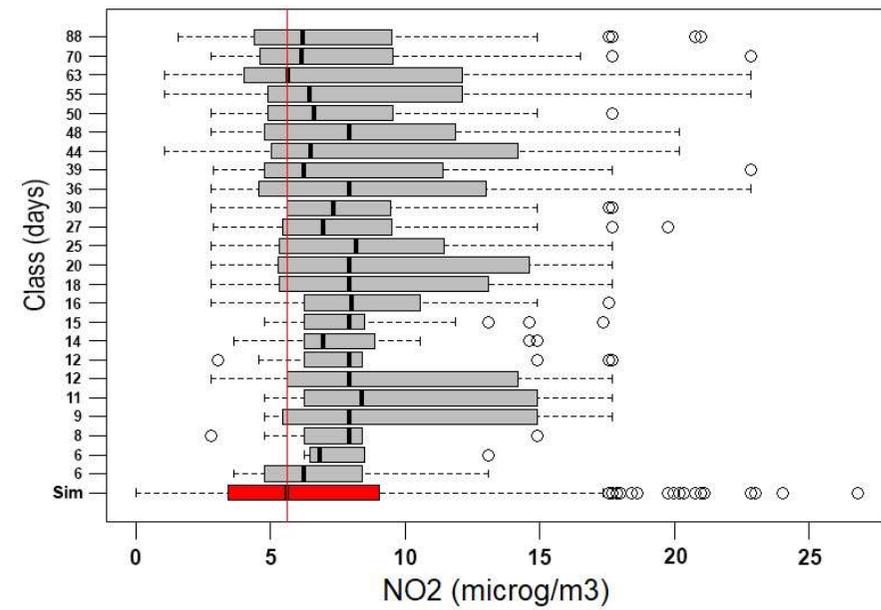


Influence of the input data on the yearly average - analysis at the traffic station, with BEST -

Extended dataset SOM statistical comparison



Reduced* dataset SOM statistical comparison



Better convergence using the extended dataset.



* Wind speed, wind direction, RH, T, NO2 background, NO2 from stations. 6

SOM performances - analysis at the traffic station -

NO₂ results for 39 days

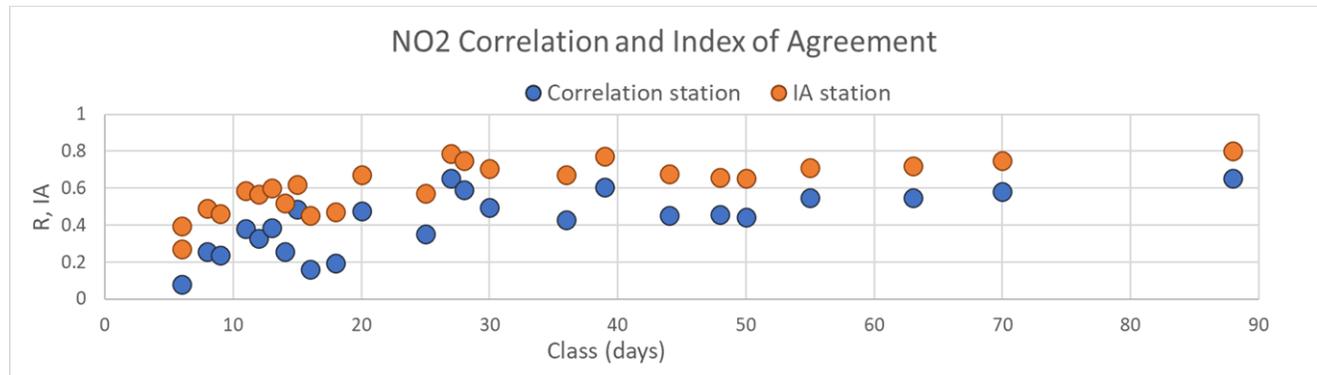
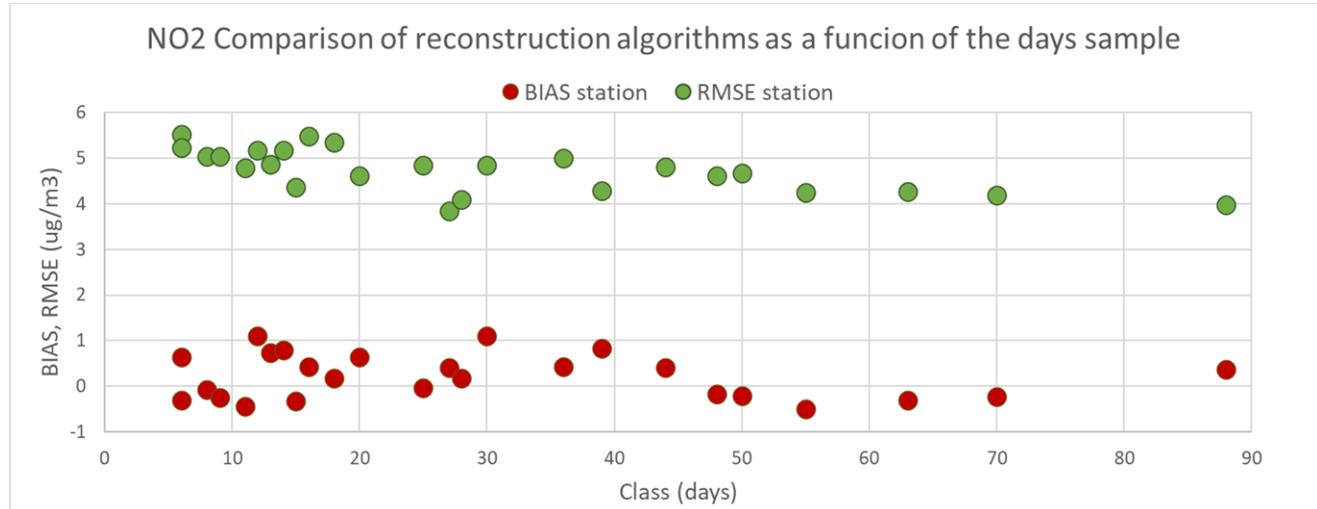
($\mu\text{g}/\text{m}^3$):

BIAS = 0.42

RMSE = 4.8

R = 0.60

IA = 0.77



SOM performances

- analysis on the full simulation domain -

NO2 results for 39 days

($\mu\text{g}/\text{m}^3$):

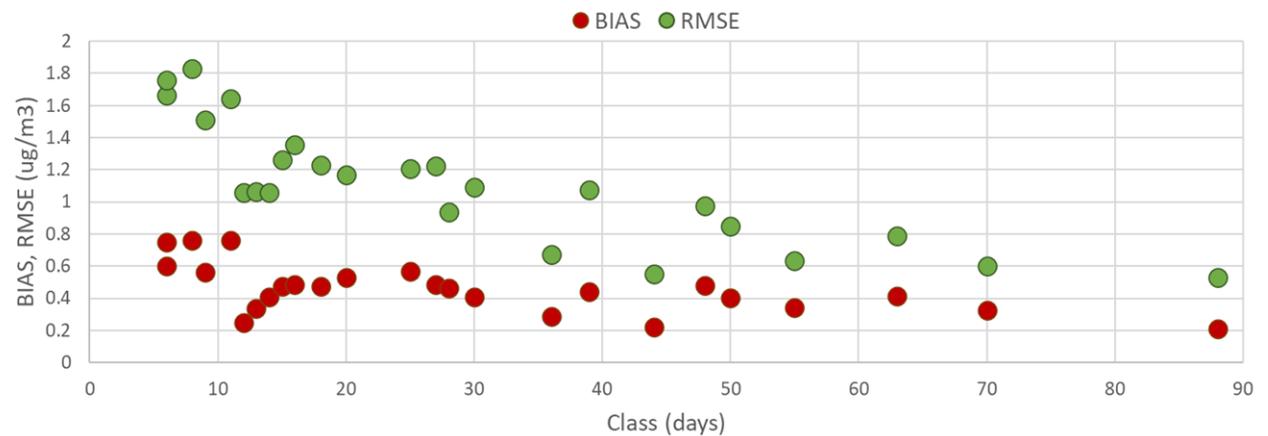
BIAS = 0.44

RMSE = 1.1

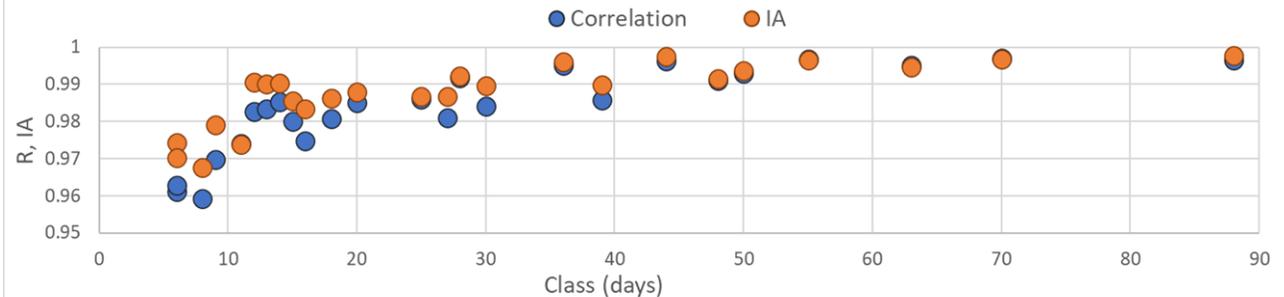
R = 0.99

IA = 0.99

NO2 Comparison of reconstruction algorithms as a function of the days sample



NO2 Correlation and Index of Agreement



Preliminary conclusions

- We performed the full year simulation with PMSS in the CT4 intercomparison exercise.
- We studied the SOM algorithm to find a minimum number of representative days for the annual mean concentration of NO₂, comparing the resulting sampled simulations with the average concentration simulated with PMSS. We found that:
 - The choice of the subset of days is highly subjective.
 - We found that there is no clear and definitive indication of which set of data is best.
 - A larger set of input parameters gives better prediction of the year average.
- The computation time needed to perform these tests was significantly larger than the computation time necessary for our yearly simulation. Therefore, we continue to prefer to perform a full year simulation.