Intercomparison Exercise of Spatial Representativeness Methods

Antonio Piersanti, Giuseppe Cremona, Gaia Righini, Lina Vitali
ENEA

FAIRMODE spatial representativeness workshop, 22 June 2017

ENEA
Italian National Agency for New Technologies, Energy and Sustainable Economic Development

Starting point…..

"A point measurement is representative of the average in a larger area (or volume) if the probability that the squared difference between point and area (volume) measurement is smaller than a certain threshold more than 90% of the time.
The maximum tolerable difference has to be assessed for every individual problem; it should not be smaller than the uncertainty of the measurement."

Nappo et al, 1982.

We are familiar with models! so we assumed model concentrations as "measurements" and developed a procedure for recursively comparing concentration time series.

At each time step t_i,
given the fixed site of interest with concentration $C(X_{site}, Y_{site}, t_i)$,
at each grid point in the model computation domain, with concentration $C(x, y, t_i)$,

the **relative difference between concentration values, ΔC/C,** is computed:

$$\frac{|C(X_{site}, Y_{site}, t_i) - C(x, y, t_i)|}{C(X_{site}, Y_{site}, t_i)}$$

Blanchard et al. (1999), Janssen el al. (2008), FAIRMODE SR survey (Denby and Balaguer, 2011), AQD quality target for measurements

and compared with a threshold: **0.2 (20%)**.
The **concentration similarity occurs when ΔC/C < 0.2**.

A concentration similarity frequency function $f_{site}(x,y)$ counts **positive occurrences of concentration similarity on the selected time interval N_t**:

$$f_{site}(x,y) = \frac{\sum_{i=1}^{N_t} flag}{N_t}, \qquad where\ flag = \begin{cases} 1, & \frac{|C(X_{site}, Y_{site}, t_i) - C(x, y, t_i)|}{C(X_{site}, Y_{site}, t_i)} < 0.2 \\ 0, & \frac{|C(X_{site}, Y_{site}, t_i) - C(x, y, t_i)|}{C(X_{site}, Y_{site}, t_i)} > 0.2 \end{cases}$$

Intercomparison Exercise of Spatial Representativeness Methods

Finally, the count of positive occurences of similarity on the time interval N_t is compared with a threshold: 0.9 (**90 % of time instants**) (Nappo et al., 1982).

On the time interval N_t,
the fixed site of interest with concentration $C(X_{site}, Y_{site}, t_i)$ is representative of the grid point with concentration C(x, y) if
$F_{site}(x,y) > 0.9$

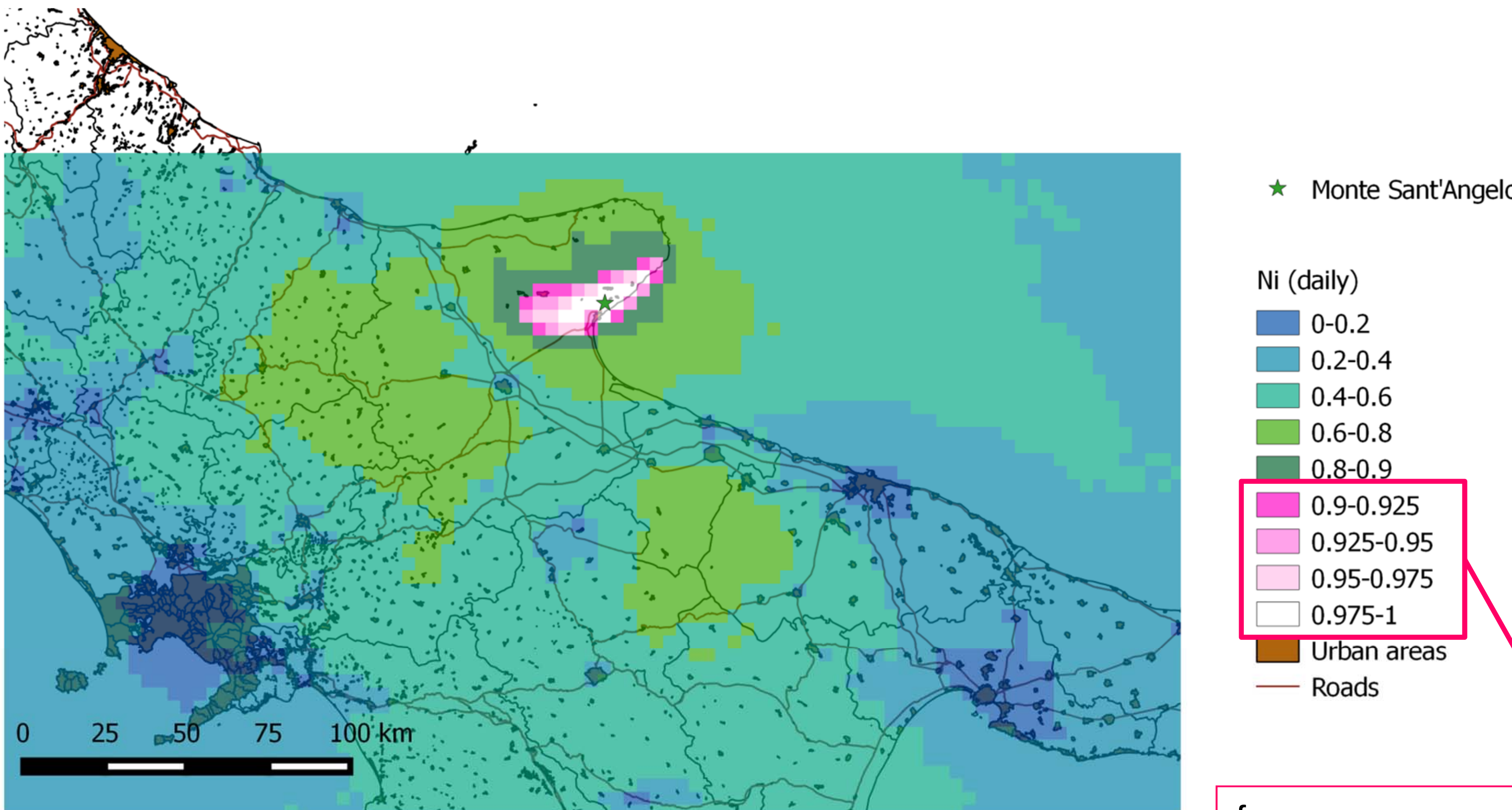
The area of representativeness of the fixed site (x_{site}, y_{site}) is the union of the grid cells (x, y) where F_{site} (x, y) > 0.9.

0.7	0.89	0.91	0.89	0.8
0.85	0.91	0.93	0.9	0.9
0.92	0.98	1	0.95	0.85
0.93	0.98	0.97	0.89	0.83
0.85	0.86	0.95	0.89	0.78

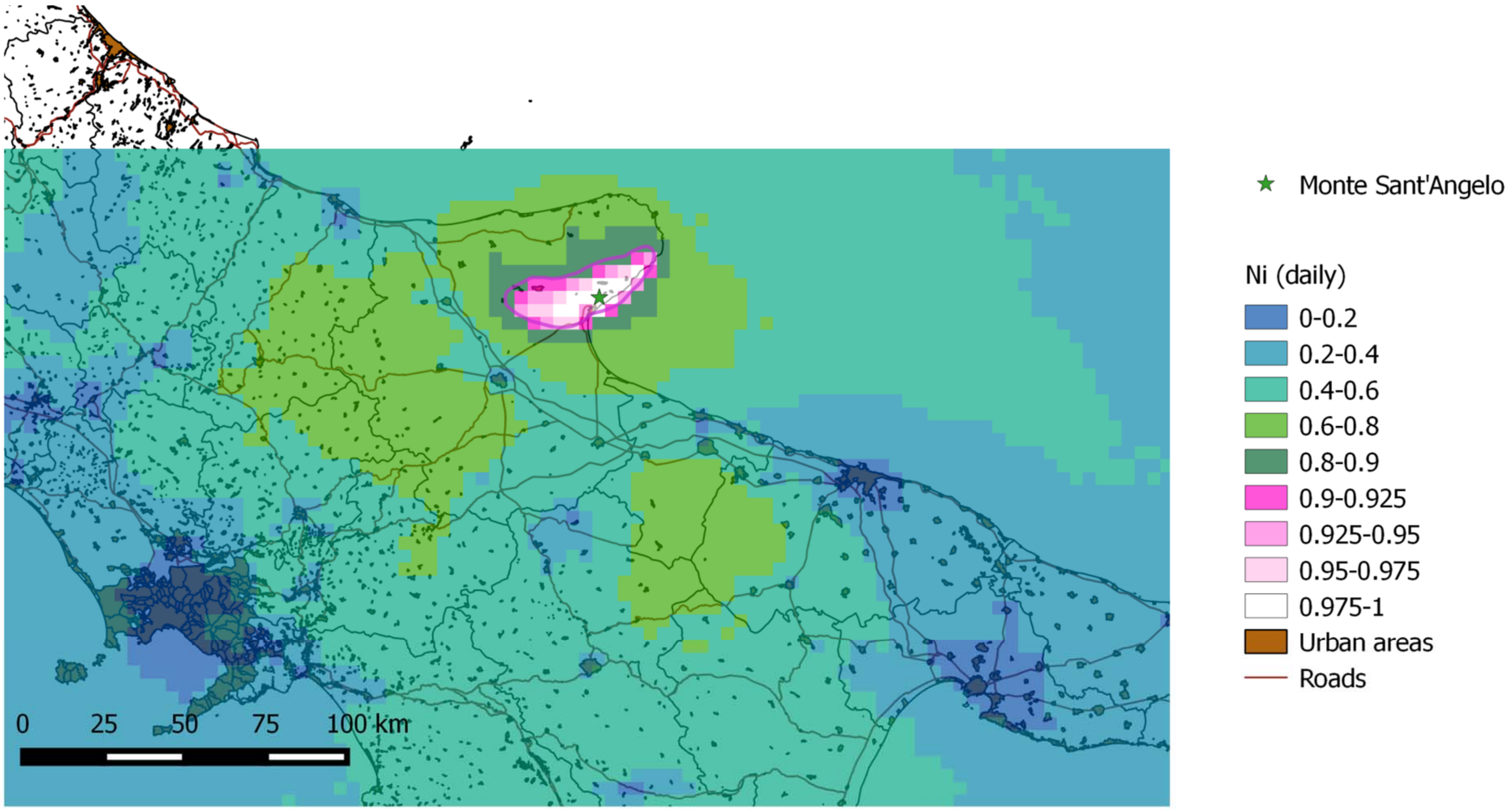
(grey)	Area of representativeness
1	Cell where the monitoring site is located

ENEA

Intercomparison Exercise of Spatial Representativeness Methods



frequency occurrences of concentration similarity ($\Delta C/C<0.2$) >0.9


Monte Sant'Angelo
Ni (daily)
0-0.2
0.2-0.4
0.4-0.6
0.6-0.8
0.8-0.9
0.9-0.925
0.925-0.95
0.95-0.975
0.975-1
Urban areas
Roads
0
25
50
75
100 km

Intercomparison Exercise of Spatial Representativeness Methods

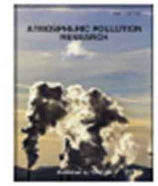
Atmospheric Pollution Research 6 (2015) 953–960

HOSTED BY
ELSEVIER

Contents lists available at ScienceDirect

Atmospheric Pollution Research

journal homepage: http://www.journals.elsevier.com/locate/apr



Original article

Spatial representativeness of air quality monitoring stations: A grid model based approach

CrossMark

Antonio Piersanti*, Lina Vitali, Gaia Righini, Giuseppe Cremona, Luisella Ciancarella

ENEA – National Agency for New Technologies, Energy and Sustainable Economic Development, Bologna, Italy

A R T I C L E I N F O

Article history:
Received 16 January 2015
Received in revised form
29 April 2015
Accepted 30 April 2015
Available online 12 October 2015

Keywords:
Chemical transport model
Monitoring networks
Spatial representativeness
Rural background

A B S T R A C T

A methodology for quantifying areas of spatial representativeness of air quality monitoring station is here proposed, exploiting the wide spatial and temporal coverage of chemical transport models results. The method is based on the analysis of time series of model concentrations, extracted at monitoring sites and around, by means of a Concentration Similarity Function (CSF). The method was tested on AMS-MINNI model results, covering Italy and three reference years (2003, 2005, 2007), for assessing the spatial representativeness of PM2.5 and O_3 rural background monitoring stations. The CSF methodology shows good performances in describing both the extension and the shape of representativeness areas, taking into account the difference between pollutants and the dependence on averaging time and temporal interval of concentration data. Results show a large variability in the size and shape of the selected stations in Italy, ranging from 220 to 4500 km^2. This confirms the importance of carrying out ad-hoc analyses on monitoring stations, as general a priori classifications and qualitative assessments of spatial representativeness are not able to fully capture the complexity of different territorial contexts.

ENEA

STRENGTHS

- BASED ON 4D MODEL CONCENTRATIONS FIELDS
  - not limited to fixed monitoring locations
  - all kind of pollutants, individually
  - works on time series --> based on annual time series of short-time averages (e.g. hourly NO2, daily PM10, max of daily 8-hour running means of O3…): fit for SR assessment of short-time averages.
  - precise definition of the SR shape

- FLEXIBLE PROCEDURE
  - F_{site} is a 3D raster field --> SR delimitation depends on the selected threshold
  - inside the SR area, the variability of F_{site} allows second-level assessment («high SR» vs «low SR»)

WEAKNESSES

- BASED ON 4D MODEL CONCENTRATIONS FIELDS
  - resolution is bound to model results
  - quality of results depends on the quality of the model: meteo and emission input, chemistry, physics, data assimilation….
  - works on time series --> based on annual time series of short-time averages (e.g. hourly NO2, daily PM10, max of daily 8-hour running means of O3…): fit for SR assessment of short-time averages.
  - to be applied to traffic stations: needs microscale modeling + some knowledge of background concentrations….

CSF METHOD APPLIED TO ANTWERP DATASET

The calculation methodology is in 3 steps:

- **Application of the the CSF function**
  For each of the 11 studied stations, the CSF function value (i.e. the indicator of concentration similarity) was calculated on every virtual station, using the time series of concentrations (hourly for NO2, daily maximum of 8 hours running average for O3, daily average for PM10). The threshold on the difference of the single time step concentration was 20%.

- **Inverse Distance Weighting (IDW) interpolation of the CSF point values**.
  Statistical interpolation methods, i.e. kriging, were tried but distribution data resulted not satisfactory.

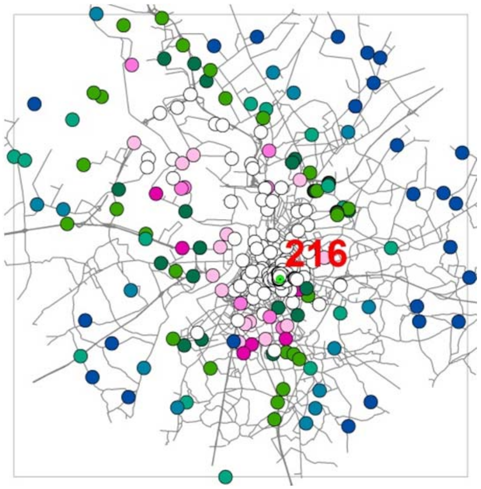
- **Representativeness area assessment**
  Contouring the area where the CSF values are greater than the threshold of 0.9 (i.e. frequency of positive occurrences of concentration similarity greater than 90%).

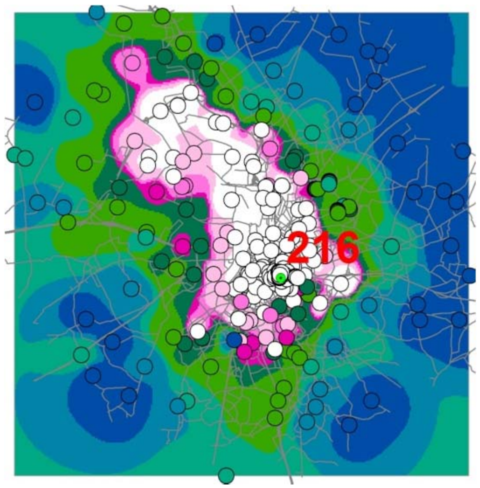
Station 216

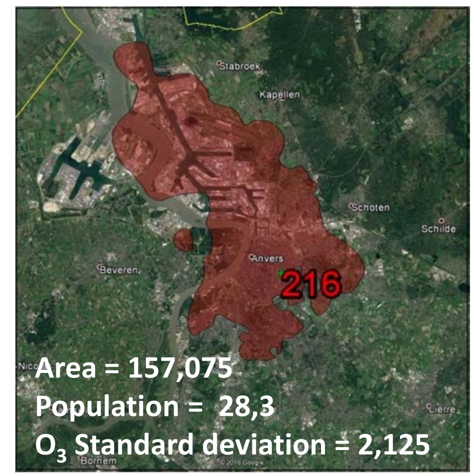
Type
Urban Traffic SC

Coordinates
X = 154396
Y = 211055

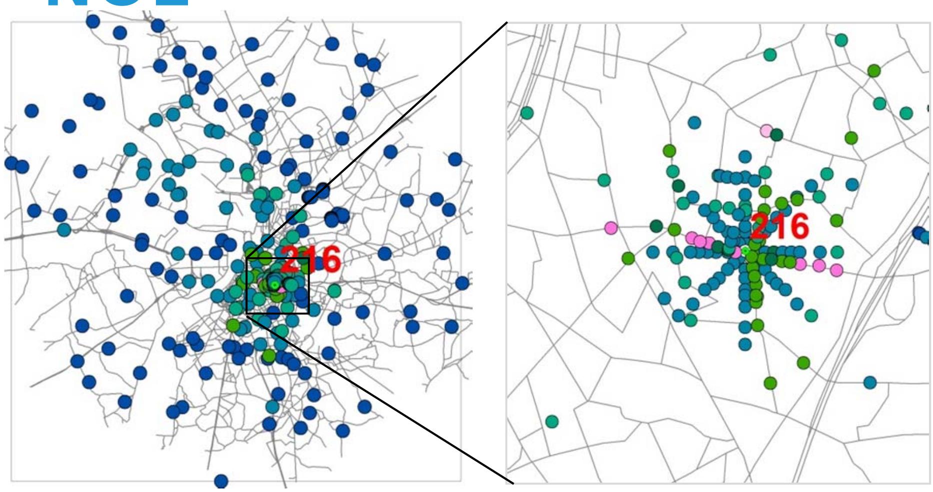
PM10

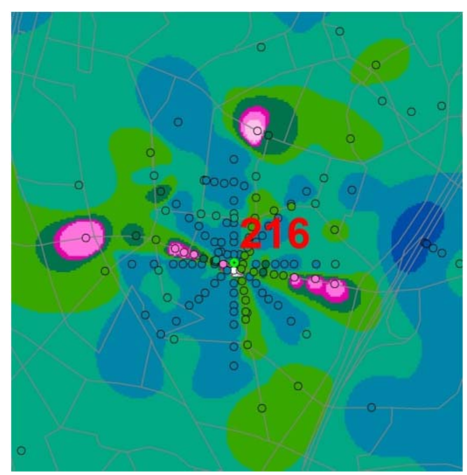


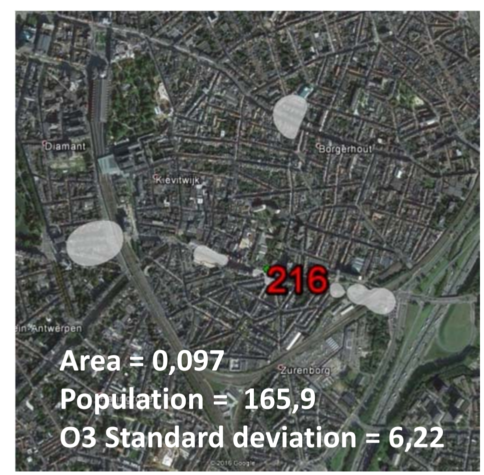




NO2





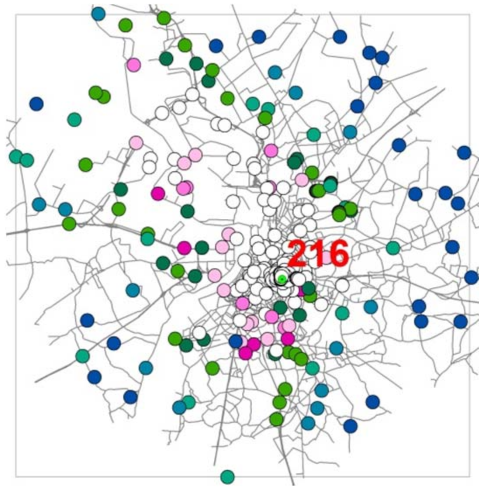


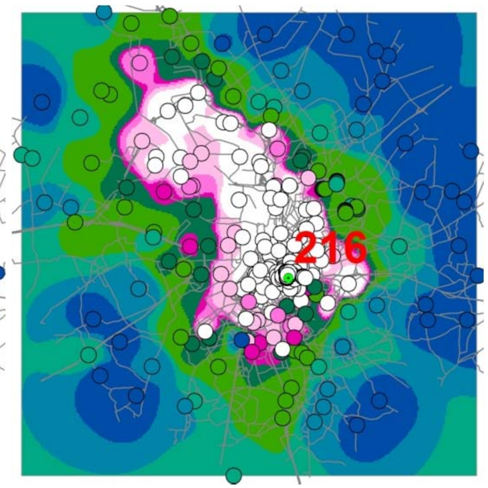
PM10

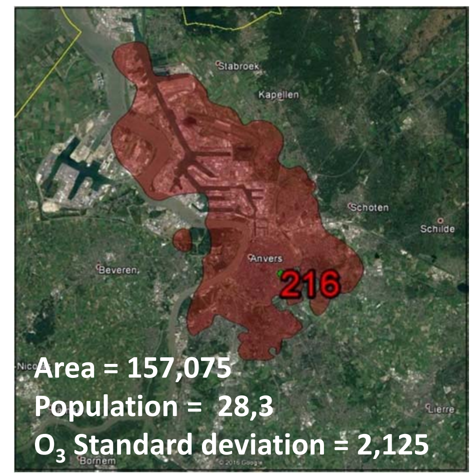
Station 216

Type
Urban Traffic SC

Coordinates
X = 154396
Y = 211055



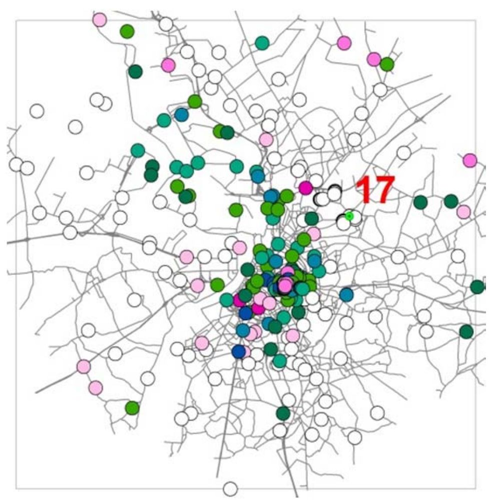


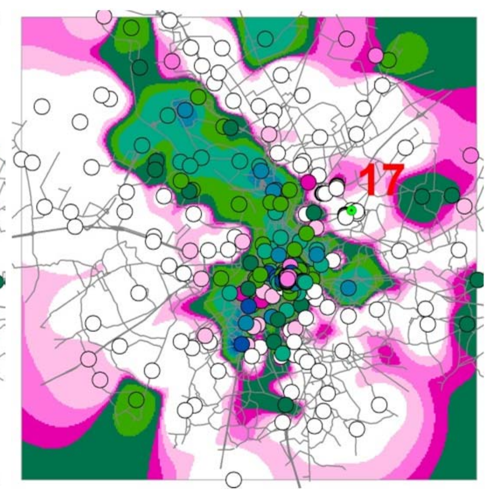


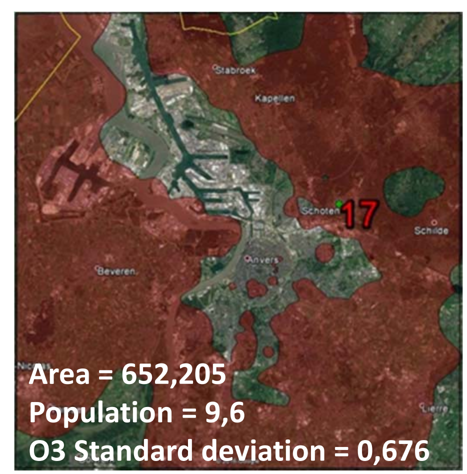
Station 17

Type
Urban Background

Coordinates
X = 158560
Y = 215807







Scope, objectives and typical use of the selected spatial representativeness (SR) method

1) What is the **scope** and the detailed **objectives** of your SR method used in the exercise?

Calculate an area of representativeness of a given georeferenced monitoring station

2) In which **context** do you typically use this method?

Provide SR for the National special purpose monitoring network

Provide Italy-Regions with a method for e-reporting request of SR

Data assimilation / fusion

3) Are there **other SR methods** that you would typically use in your work on SR assessments?

No (but we have tried some others: Janssen et al. 2008, Righini et al. 2014)

4) How does the use of your method(s) relate to local / regional / national / EU-wide **regulatory and / or legal obligations**?

Provides Regions (in charge of e-reporting) with a method for quantifying SR. Regions are free to adopt it or not

ENEA

Maturity and fitness to purpose of the SR method used in the exercise

1) How many **years of experience** do you have with the specific SR method used in the exercise?

Since 2013, 4 years

2) How many **years of experience** do you have with evaluating SR in general (including experience with other methods?

Same as above

3) How would you rate the **maturity of the SR method** you have used in the exercise?
(This may reach from "rather experimental" to "well established" – please also comment on the fitness to purpose of you method.)

Well established for modellers: very simple mathematical assumptions so easy to use, but hourly model concentration around monitoring stations, therefore adequate model setups, are needed.
The method answers precisely to the request of the SR area on monitoring stations.

4) Is it possible to **apply your method by other institutes** using the tools you have developed?
(e.g.: Are your tools available to others? Is there a copyright concern? What is the level of difficulty and necessary skills for their implementation?)

Yes. The method is published, and the software tool (based on NCL language) is available for the community.

Similarity criteria & definition of Spatial Representativenes (1)

1) Please summarize the underlying **definition of SR** you have used in the exercise.

Nappo, C. J., Caneill, J. Y., Furman, R. W., Gifford, F. A., Kaimal, J. C., Kramer, M. L., Lockhart, T. J., Pendergast, M. M., Pielke, R. A., Randerson, D. , Shreffler, J. H.,Wyngaard, J. C., 1982. The Workshop on the Representativeness of Meteorological-Observations, June 1981, Boulder, Colorado. Bulletin of the American Meteorological Society 63, 761–764.

"a point measurement is representative of the average in a larger area (or volume) if the probability that the ~~squared~~ difference between point and area (volume) measurement is smaller than a certain threshold more than 90% of the time"

2) Please summarize the underlying **similarity criteria & threshold parameters** you have used.

Similarity = difference between point measure at station and concentration value around: 20%

Number of time instants of similarity : 90% of available time instants

3) Are there **other SR definitions** and / or **similarity criteria** you would typically use in your work on SR?

Other SR definitions: maybe, after a first testing phase of different methods in 2013 we selected this one and no more tried anything else.
Other similarity criteria: probably, our two choices on thresholds are fixed (based on literature), it would be nice to test the sensitivity to other choices, based on scientific consensus.

ENEA

Similarity criteria & definition of Spatial Representativenes (2 – some details)

1) Are the boundaries of your spatial SR areas constrained **exactly**, or did you add some additional **buffers or safety factors**?

"Almost" exactly: the boundaries are drawn via contouring the raster layer values of f_{site}. So the precision depends on the model resolution.

2) Can SR areas of different stations **overlap** or are they considered to be **exclusive** by principal?

They can overlap.

3) Are your similarity criteria applied **one sided** or **two sided**?

(i.e.: Are you evaluating deviations only towards higher values, or towards both higher and lower values?)

Two sided.

4) Within your estimated SR areas: is spatial representativeness guaranteed for locations of **all station types**, or only for locations of **station types identical** to the type of the central station?

(e.g.:

Within the SR areas estimated for the urban background stations Schoten and Antwerpen-Linkeroever: is spatial representativeness guaranteed for locations of all station types? Or for locations of background station type only?

Within the SR area estimated for the urban traffic station Borgerhout: is spatial representativeness guaranteed for locations of all station types? Or for locations of traffic station type only?)

Not clear….However, the method does not depend on station type. Being based on gridded model results, it has the limitations given by the model resolution.

Input data

1) Please summarize which part of the **input dataset** you have used in the exercise.

time series of model concentrations at all available locations (virtual stations)

2) Did you use **additional data**, not contained in our dataset?
(e.g., Street View pictures, maps from other sources, etc.)

No

3) How suitable did you find the **Antwerp dataset** for your method? / How suitable would you rate your method to be for this type of dataset?

Unfortunately the Antwerp dataset was not well fit for our method, as we use model concentrations on the whole calculation grid, for each model time step. Our method certainly is not so good for the Antwerp dataset, as it can say nothing in the areas where hourly concentrations are not available. The areas we have drawn are largely arbitrary, we did not work on a good interpolation of similarities on the selected points, as it is not part of our tool.

4) Did you **miss** any type of data / information in this dataset?

Yes, hourly concentrations on the model domain.

5) How does the dataset of the exercise compare to the **data you would more typically use** for you work on SR?

See above.

THANK YOU

antonio.piersanti@enea.it