SPATIAL REPRESENTATIVENESS

French contribution (LCSQA/INERIS)

(Laurent Létinois, Laure Malherbe, Maxime Beauchamp)

FAIRMODE technical meeting

Athens, 19-21 June 2017

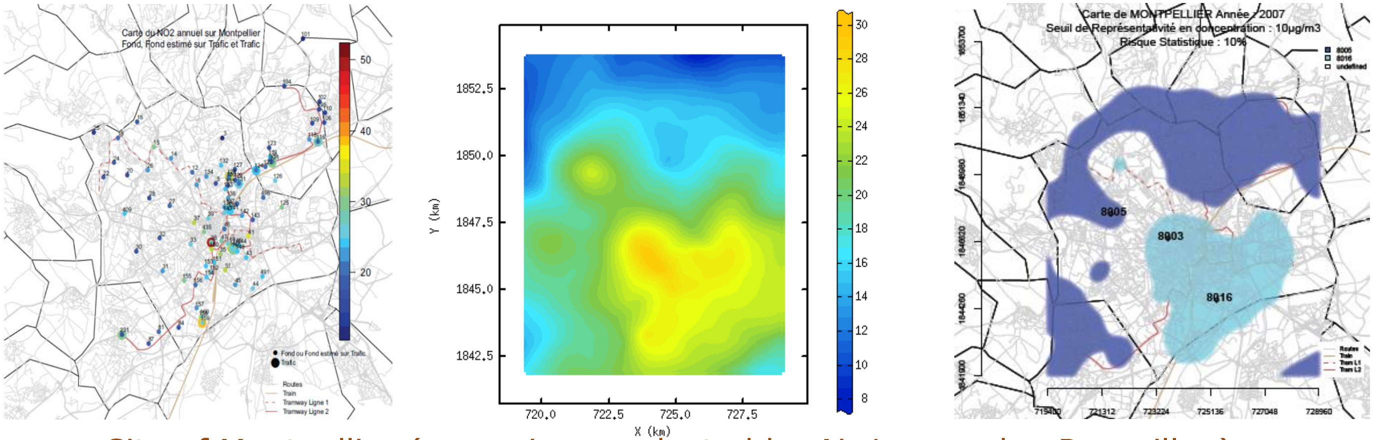
Methodology applied in this intercomparison exercise :

- ✓ Based on a study from Atmo Normandie (Cori, 2005; Bobbia et al., 2008) and further developed by the LCSQA - the French NRL – (Cárdenas et al., 2007; Beauchamp et al., 2011, 2012) for local use by the regional air quality monitoring associations (AASQAs).
- ✓ Requires a spatial estimation of concentrations with corresponding uncertainty as input → kriging + kriging error standard deviation.

Sampling points: several periods during the year 2007

Estimation map of NO2 annual mean concentrations: kriging with NOx emissions as external drift

Representativeness area for sites 08005 and 08016



City of Montpellier (campaign conducted by Air Languedoc-Roussillon)
Annual mean concentrations of background NO_2. 2007.

FAIRMODE meeting, May 2012

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?

Scope: air quality assessment by measurement methods under Directives 2008/50/CE & 2004/107/CE

Objectives:

- Checking compliance with the Directives as regards spatial representativeness of the sampling points
- Providing information on station representativeness in the AQ reporting
- Supporting the monitoring strategy:
  - ✓ Does the monitoring network provide a good coverage of the considered domain?
  - ✓ Can all locations be related to a monitoring site? Are there redundancies? Lacks?

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

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

Until now this methodology has been applied to **NO_2 annual mean concentrations in urban areas, using data from passive sampling surveys.** Applied for the first time to PM_{10} and modelling data in this exercise.

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

Other methods may be applied but not "typically".

- ✓ One study on the SR of PM_{10} traffic monitoring stations based on measurement campaigns (a few measurement sites along and across the road) → experimental assessment of the spatial variations of PM_{10} concentration.
- ✓ Studies on station classification:
  - on the European scale : application of Joly & Peuch methodology (ETC/ACM 2013)
  - on the national scale : classification based on PCA

  → qualitative information about station SR.
- ✓ AASQAs: local studies on SR (monitoring campaigns, use of land cover…)

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

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

French guidelines on station siting : the AASQAs are requested to assess the spatial representativeness of their monitoring points 1) before locating a new station ; 2) every 5 years for existing stations.

Objectives: checking compliance with AQD requirements and fulfilling IPR obligations.

No specific methodology is prescribed. The kriging-based method is proposed as a possible approach.

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?

   About 6 years including development and applications but no regular use.

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

   About 6 years but not a regular activity.

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.)

   Beyond the experimental stage but not completely established yet. Some issues arouse during this exercise (See slide on limitations).

   Need to test the methodology in other contexts (other pollutants, other statistics, rural areas…)

Maturity and fitness to purpose of the SR method used in the exercise (2)

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?)

Two steps :

1) Production of a concentration map with its associated uncertainty

Kriging from passive sampling campaigns: R scripts (internal use).

Kriging from urban modelling outputs: C++ & R scripts (internal use and distribution to AASQAs on demand).

Distribution to other institutes: not currently envisaged.

2) Computation of the representativeness area from the kriging results

Generic R script applicable to any kriging gridded output (internal use and distribution to AASQAs on demand).

Distribution to other institutes: possibly, in the framework of FAIRMODE.

- Condition: reference to LCSQA with any published results.
- User friendliness: easy to use.

Similarity criteria & definition of Spatial Representativenes (1)

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

The representativeness area of a station S_0 located in x_0 is defined as:

$$A_0 = \{x \ / \ |Z(x) - Z(x_0)| < \delta\} \quad (1)$$

- $Z(x)$: unknown concentration at a location x
- $Z(x_0)$: observed concentration at station S_0 (measurement error not taken into account)
- δ: maximum tolerated deviation of concentration (in $\mu g.m^{-3}$ or %) with respect to the station

Considering the annual mean concentration as the realization of a random function, definition, (1) can be expressed in terms of expectation.
→ *Z(x)* is replaced by its kriging estimate, noted $Z^K(x)$. The estimation error is pragmatically assumed to be normally distributed.

Given the statistical risk, noted η_r, that a point be wrongly included in A_0, definition (1) finally becomes:

$$A_0 = \left\{x \quad / \left|Z^K(x) - Z(x_0)\right| < \boldsymbol{\delta} - \sigma_K(x) \times q_{1-\frac{\eta r}{2}}\right\} \quad (2)$$

$\sigma_K(x)$ *is* the kriging standard deviation at location x and $q_{1-\frac{\eta r}{2}}$, the quantile of order $1-\frac{\eta r}{2}$ of the normal distribution.

Concentration mapping step

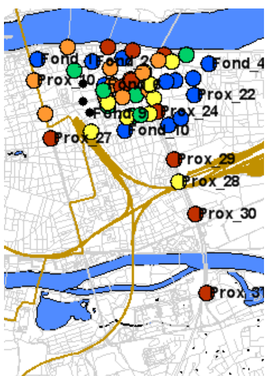
- Previous studies: kriging applied to passive sampling campaigns + auxiliary variables (step 1 or steps 1 & 2)

  1. Background pollution mapping

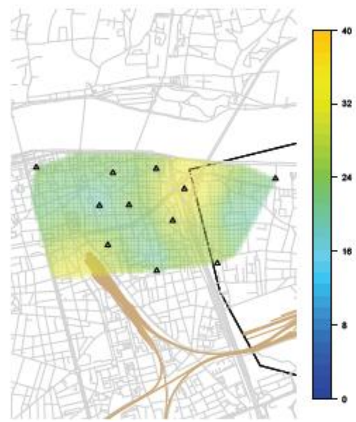
  background sampling sites + auxiliary variables → external drift kriging

  2. Addition of an estimated traffic-related increment in roadside locations

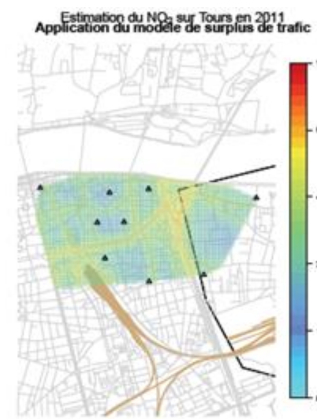
  traffic-related sampling sites (along and across roads) + auxiliary variables → statistical model for traffic-related increment



City of Tours. NO_2. Passive sampling survey conducted by Lig'Air around a traffic monitoring station. Measurement period: all the year 2011.



Background pollution: kriging with NO_x emissions and population density as external drift.



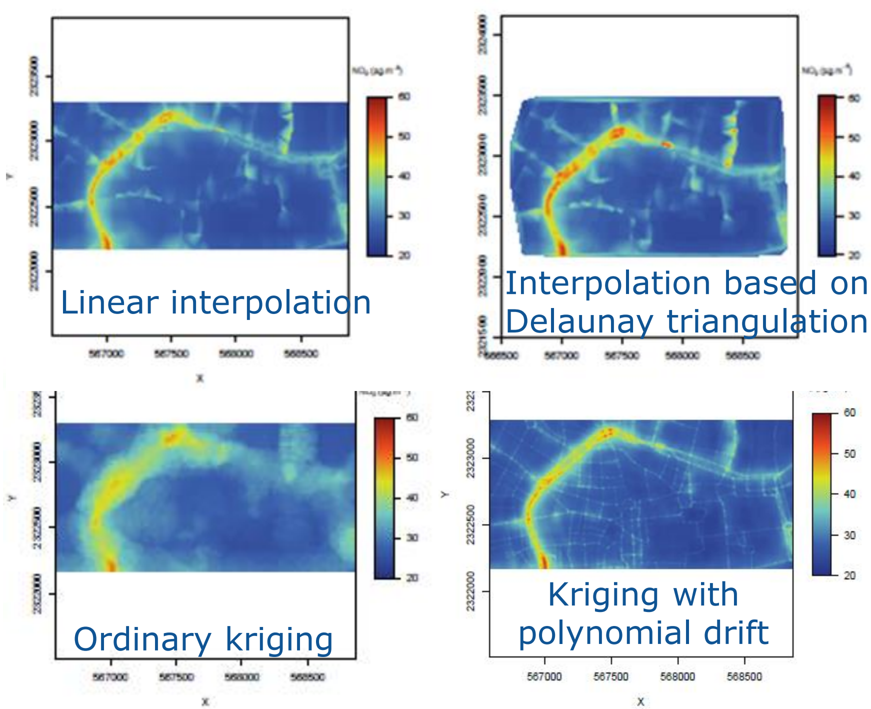
Background + traffic-related pollution

FAIRMODE meeting,January 2014

Concentration mapping step

- 2016: geostatistical developments to interpolate concentrations from a set of background and traffic-related data, taking the emission inventory and traffic linear emissions along the road network into account (kriging with a polynomial external drift) (Beauchamp et al., 2016).

This method can be applied to passive sampling campaigns mixing background and traffic sites but also to urban modelling outputs → <u>approach used in this exercise</u>.



Interpolation from ADMS-Urban output data (irregular grid) to a fine resolution grid. Comparison of different interpolation methods.

Similarity criteria & definition of Spatial Representativenes (1)

2) Please summarize the underlying **similarity criteria & threshold parameters** you have used.
3) Are there **other SR definitions** and / or **similarity criteria** you would typically use in your work on SR?

No other SR definition

Similarity criteria:

Previous studies (NO_2 annual mean): δ = 10 µg.m^{-3} and η_r=10% (based on sensitivity tests)

This study: some adjustments had to be done → new sensitivity tests to find the best compromise between realistic SR areas and reasonable values for δ and η_r.

Intercomparison Exercise of Spatial Representativeness Methods

Pollutant	Type of concentration	Reference measurement data	Main input data for the mapping	Auxiliary data for the mapping	Mapping method	Output maps	Maximum deviation (δ)	Statistical risk (η_r)	Representativeness area	Comment
NO_2 PM_{10}	Annual mean	Virtual station data (with noise)	Virtual passive sampling data (for NO_2 only) IFDM-RIO-OSPM original output data (before interpolation on a grid)	Gridded emissions (NO_x for NO_2, PM_{10} for PM_{10}) (to account for background pollution) Road emissions (NO_x for NO_2, PM_{10} for PM_{10}) and distance to the roads (to account for traffic related pollution)	Kriging with external polynomial drift	Estimated annual mean concentration Kriging standard deviation	10 µg.m^{-3} (Beauchamp et al., 2011) **30%** (maximum allowed uncertainty for NO_2 modelling in the region of the annual LV, cf. Dir. 2008/50/EC.) 50% (maximum allowed uncertainty for PM_{10} modelling in the region of the annual LV, cf. Dir. 2008/50/EC.)	10% 12,5% **15%** 20% 30% 40%	Estimated: - before processing intersection zones (possible overlap between SR areas) - after processing intersection zones (no overlap)	δ=10 µg.m^{-3} and η_r =10% are the values retained by Beauchamp et al. (2011). For this exercise, the values for δ and η_r and were readjusted after additional tests.

⇨ **δ = 30% of the measured concentration at the station and η_r =15% for both NO_2 and PM_{10}.**

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

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

No buffer added. Safety factor: included in the accepted risk η_r that the concentration at point inside the SR area differs from the station measurement by more than δ.

2) 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, see definition slide 8.

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

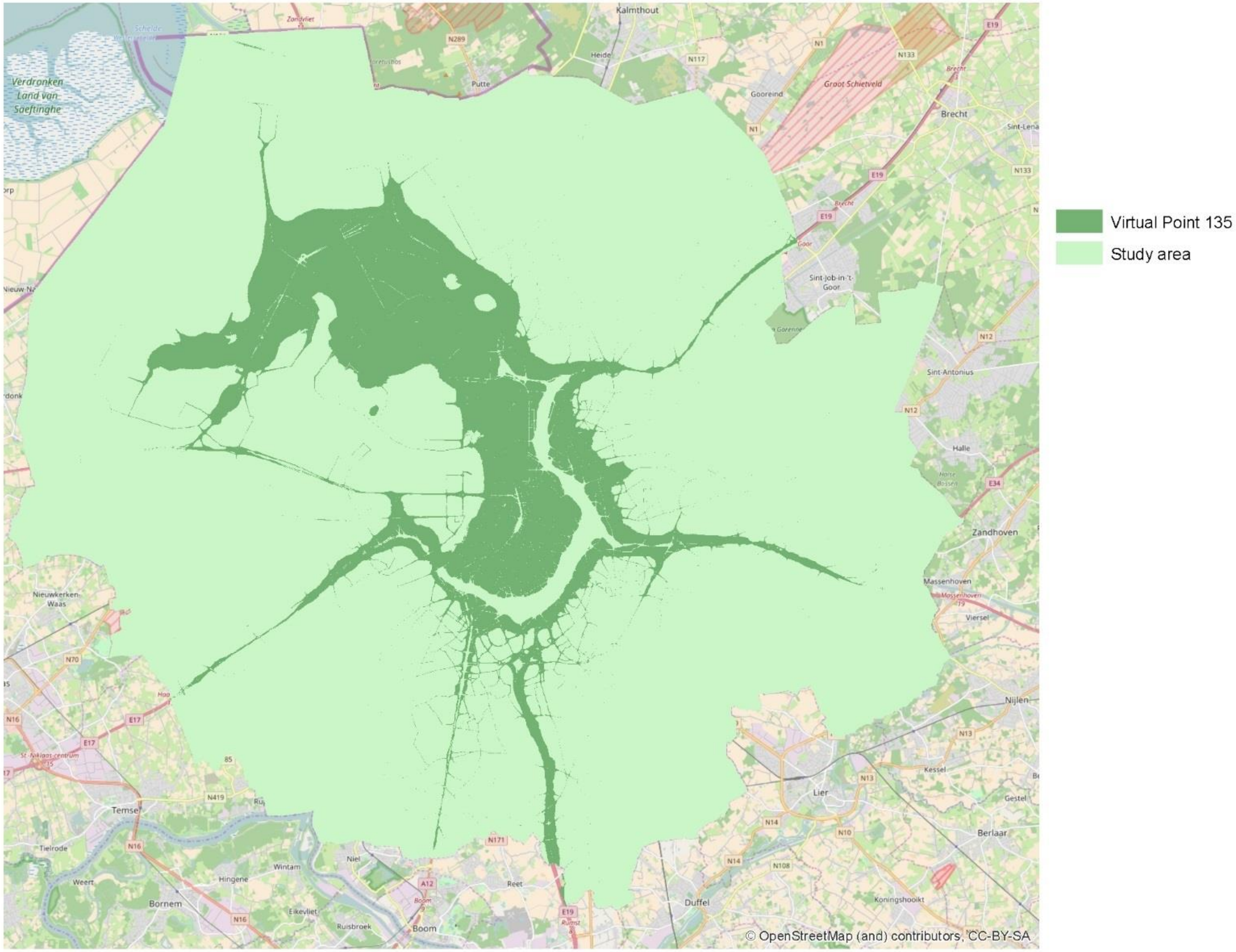
3) Can SR areas of different stations **overlap** or are they considered to be **exclusive** by principal?

Both:

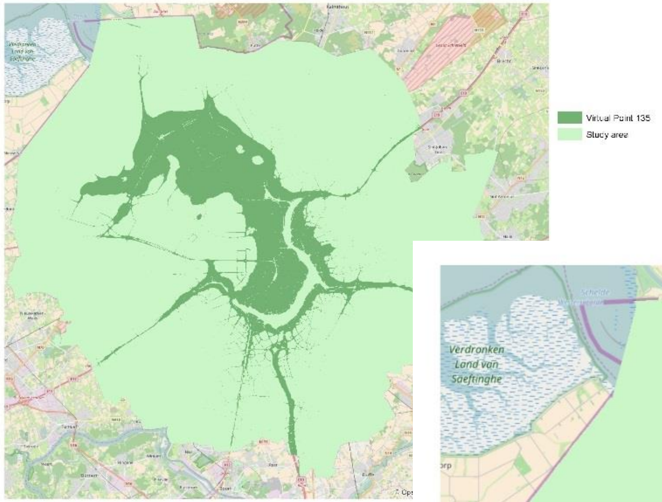
1) the SR areas are first calculated independently for each monitoring station.

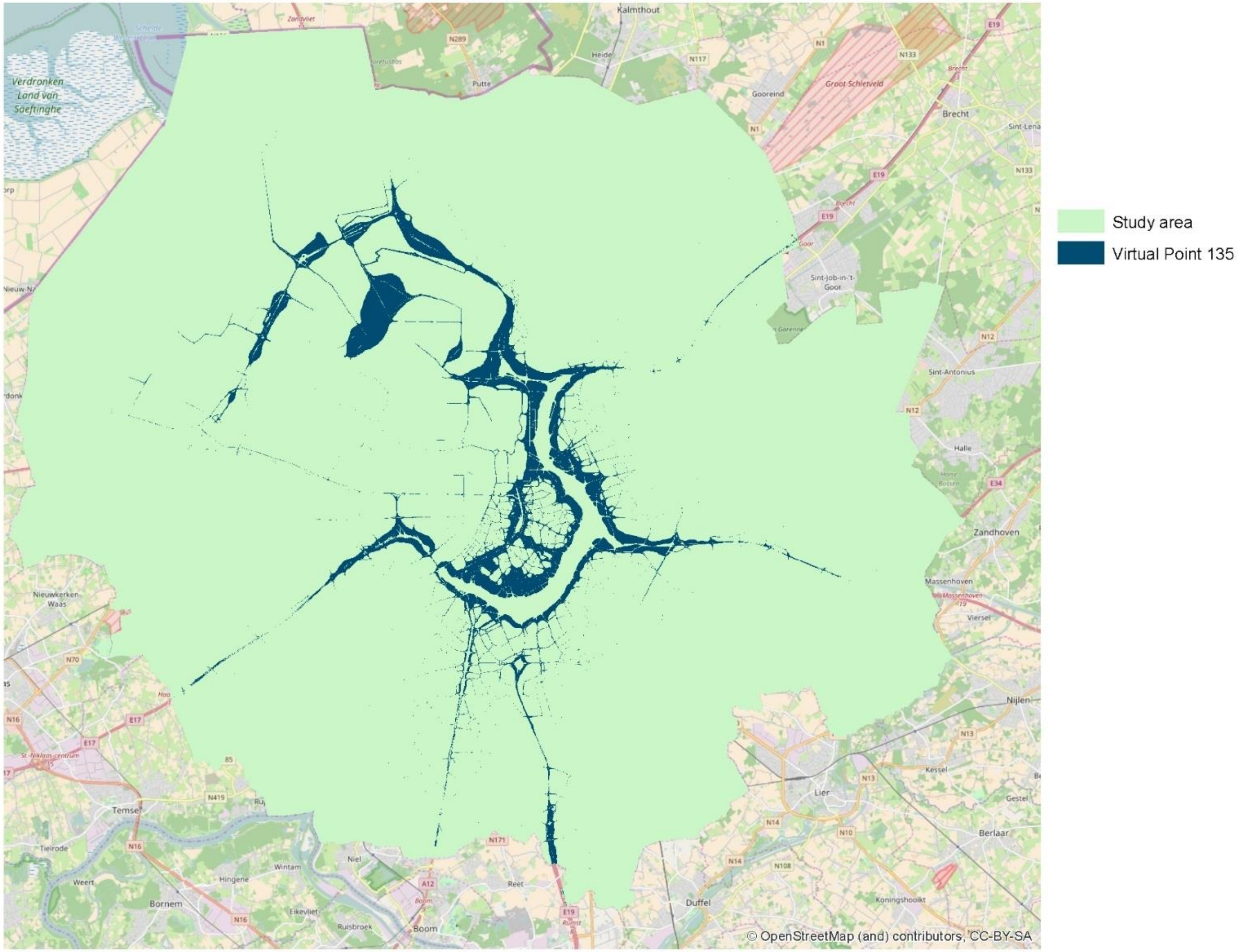
2) the monitoring network is then considered in its whole: points located in intersection zones are linked to one station only. Objective: determining by which monitoring site each point is best represented.

Different criteria previously tested to deal with overlaps. Criterion retained: minimum concentration difference with respect to the station.



Example:
NO_2, estimated representativeness area for VS135 <u>before</u> removing overlaps





Example:
NO_2, estimated representativeness area for VS135 <u>after</u> removing overlaps

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

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?)

According to our definition, spatial representativeness within the SR area is guaranteed at a given statistical risk (to account for the map uncertainty) whatever the type of location.

However, further investigation based on local expertise would be necessary to check the relevance of the delimited zones (especially for the traffic station).

Input data

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

   See table

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

   No additional data from external sources.

   Use of the original modelling output data (i.e. before interpolation) provided by VITO (data not included in the initial data set sent to participants).

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?

   Suitable dataset

Pollutant	Type of concentration	Reference measurement data	Main input data for the mapping	Auxiliary data for the mapping	Mapping method	Output maps	Maximum deviation (δ)	Statistical risk (η_r)	Representativeness area	Comment
NO_2 PM_{10}	Annual mean	Virtual station data (with noise) - including at the telemetric sites (after discussion with JRC)-	Virtual passive sampling data (for NO_2 only)	**Gridded emissions** (NO_x for NO_2, PM_{10} for PM_{10}) (to account for background pollution) **Road emissions** (NO_x for NO_2, PM_{10} for PM_{10}) and **distance to the roads** (to account for traffic related pollution)	Kriging with external polynomial drift	Estimated annual mean concentration Kriging standard deviation	10 µg.m-3 (Beauchamp et al., 2011) 30% (maximum allowed uncertainty for NO_2 modelling in the region of the annual LV, cf. Dir. 2008/50/EC.) 50% (maximum allowed uncertainty for PM_{10} modelling in the region of the annual LV, cf. Dir. 2008/50/EC.)	10% 12,5% 15% 20% 30% 40%	Estimated: - before processing intersection zones (possible overlap between SR areas) - after processing intersection zones (no overlap)	δ=10 µg.m-3 and η_r =10% are the values retained by Beauchamp et al. (2011). For this exercise, the values for δ and η_r and were readjusted after additional tests.
			IFDM-RIO-OSPM original output data (before interpolation on a grid)							

Input data

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

No missing data.

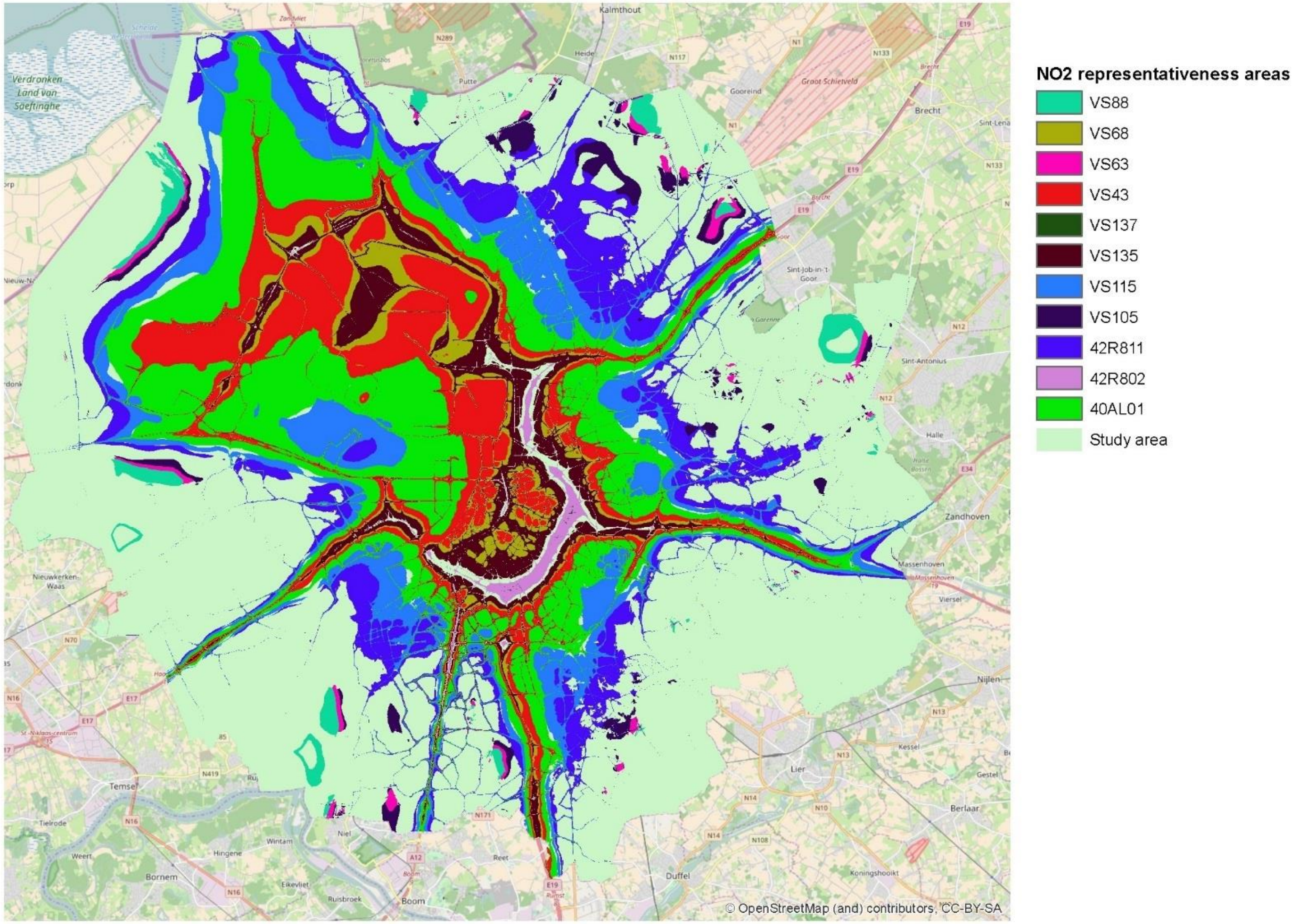
Unclear instruction: the exact list of virtual monitoring stations to consider for SR assessment

A more detailed characterisation of the virtual sampling points/original modelling points (background location traffic location, transition between traffic & background …?) would have been helpful for the geostatistical modelling. Not enough time to examine the input data in detail.

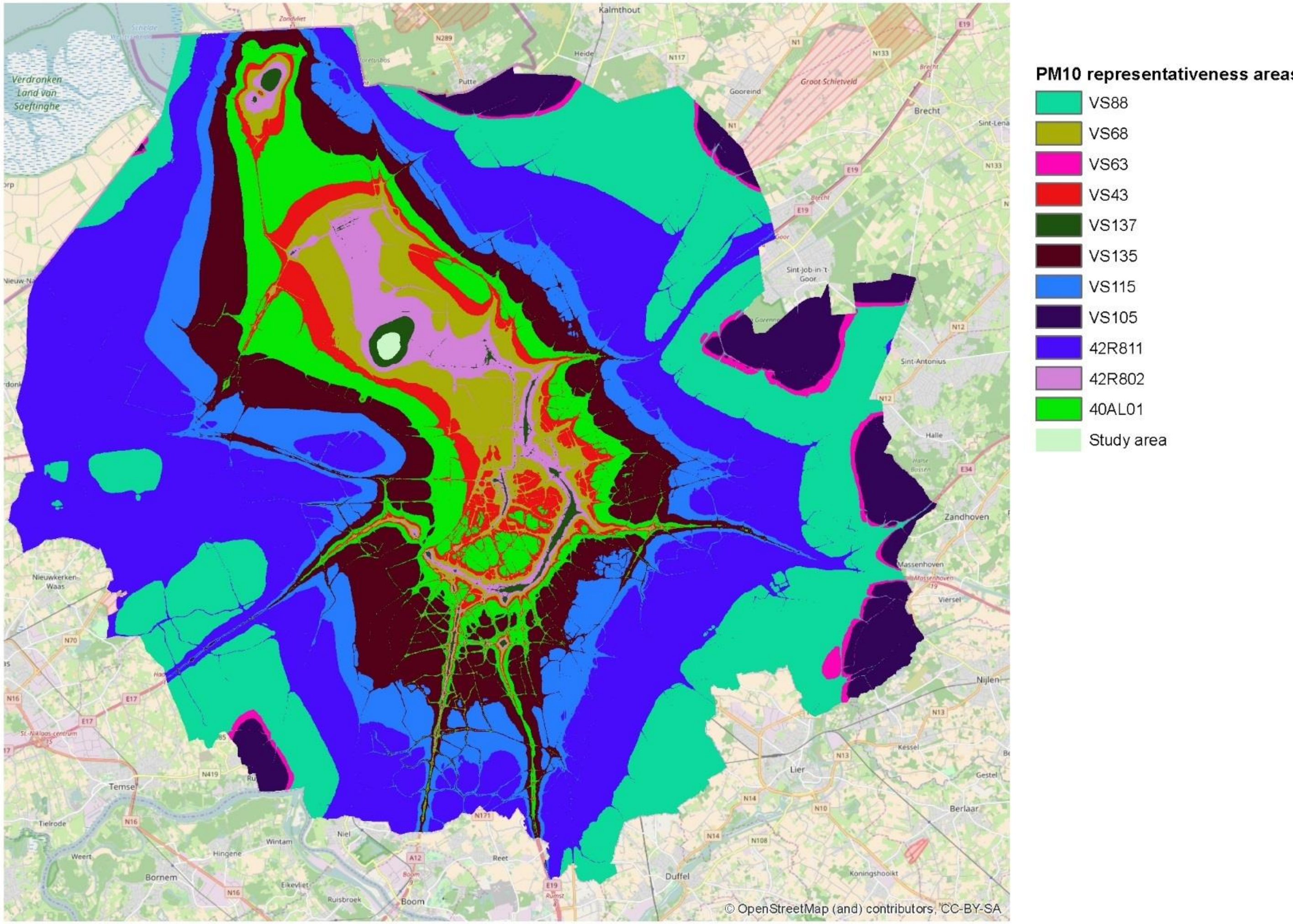
4) How does the dataset of the exercise compare to the **data you would more typically use** for you work on SR?

- Passive sampling data:
  - number: more numerous than in usual sampling surveys.
  - type of location for the sampling points: not enough information to answer
- Modelling data: overall similar. But background and street canyons concentrations are calculated by two different models, IFDM and OSPM, possibly at the same or at close locations (→ 2 datasets to deal with for the interpolation), whereas AMDS-Urban or SIRANE commonly used in France provide only one dataset (including all contributions).
- Auxiliary data: comparable

Results



SR areas for NO_2, annual mean



SR areas for PM_{10}, annual mean

Conclusion

- **The air quality maps obtained with the virtual passive sampling data were not precise enough (high kriging standard deviation) to properly delimit representativeness areas**, even with an increased statistical risk (from 10% to 40%).

- The kriging standard deviation was logically lower using the modelling data (much more points as input in the kriging). **A similarity criterion in % gave more realistic representativeness areas than a criterion in µg.m^{-3} which tended to produce too large areas for stations measuring low values.**

- **For NO_2, some stations were found to be outside their SR area.** This can be explained by the modulation of the similarity criterion according to the map uncertainty (slide 8, formula 2), even in regions where the standard kriging deviation is moderate. This problem was only partly solved by increasing the statistical risk from 10% to 15%. **Excessive sensitivity to the kriging standard deviation may thus be a limitation of the methodology.** More research would be necessary to investigate this issue.

- IFDM and OSPM modelling data were processed together without fine analysis of their spatial distribution in the city. **A preliminary selection of IFDM and OSPM modelling points according to the location type and the street network would have probably improved the kriging and provided more accurate maps.** More realistic SR areas may have resulted as well from such improvements.