The European Commission’s science and knowledge service
Joint Research Centre

Spatial Representativeness Evaluation by Point-Centred Variography
and links to the FAIRMODE / AQUILA SR Intercomparison Exercise

Oliver Kracht & Michel Gerboles

18th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
HARMO
9th - 13th October 2017
Bologna - Italy
**Outline & Context**

**Spatial Representativeness (SR)**

**Most basic definition:**
- The *representativeness area* is described by the set of all locations where the concentration of a pollutant does not differ from the measurements at the central point (monitoring station) by more than a certain threshold.

**A geostatistical approach:**
- Classical geostatistical analysis would describe the *spatial correlation structure* of the whole concentration field in terms of the *variogram*.
- The *point-centred variogram* is based on the average of squared concentration *differences* observed in pairs formed between a *particular central point* and the set of *all other points* in the domain.

**Context:**
- FAIRMODE activities on spatial representativeness (SR).
- FAIRMODE / AQUILA *Intercomparison Exercise (IE)* of Spatial Representativeness Methods (SR-IE).
A geostatistical approach to SR:

Traditional versus Point-Centered Semivariance

**Traditional semivariance**

\[
\gamma(h) = \frac{1}{2} \frac{1}{N_h} \sum_{i=1}^{N_h} \left[ Z(s_i) - Z(s_i + h) \right]^2
\]

Half the average of the squared deviations between all paired observations at distance h.

**Point-centered semivariance**

\[
\gamma_{cp}(h) = \frac{1}{2} \frac{1}{N_{cp,h}} \sum_{i=1}^{N_{cp,h}} \left[ Z(s_{cp}) - Z(s_{cp} + h) \right]^2
\]

Half the average of the squared deviations within pairs formed by a single central point (cp) and all other observations at distance h from this cp.

Point-centered variography places a monitoring station in the context of the local or regional air quality pattern.
Besides that: same terminology

Semivariance $y(h)$ in $(\mu g/m^3)^2$

- $C_0 + C_1$
- Sill $(C_0 + C_1)$: Limit of the variogram at infinite lag distances
- Range $(a)$ of spatially correlated measurements
- Nugget $(C_0)$

Lag distance $(h)$

Graph showing the relationship between lag distance and semivariance.
**Context / Case study:**

**FAIRMODE / AQUILA Intercomparison Exercise of Spatial Representativeness Methods**

- **Performed by 11 different groups**, but on the same **shared dataset** (prepared by VITO).
- Existing stations for PM$_{10}$ (n=15), NO$_2$ (n=18) and O$_3$ (n=3)
- Dataset based on outputs from the **RIO-IFDM-OSPM model chain** for the region of **Antwerp** (year 2012).
- **Virtual stations** (n=341) from hourly model data
- **Gridded model data** (annual means, 5x5m$^2$)
- Emissions
- Population density
- Building heights
- CORINE land cover

**Spatial representativeness** estimates for:

- PM$_{10}$ and NO$_2$ at one traffic station
- PM$_{10}$, NO$_2$ and O$_3$ at two urban background stations
Case study data:

- **n=341 receptor points** ("virtual stations") from hourly model data
- **aggregated to 14-day averages** (i.e. to emulate diffusive samplers)
- **classified into street canyon (SC) and non-street canyon (non-SC)** locations
**Workflow:**

- Aggregate modelled time series of virtual receptors to different integration time scales (shown here: only 14-day averages).
- Log-transformation of concentration values.
- Calculate the point-centered variogram clouds.
  - all data pairs formed between the central point and the other virtual receptors
  - 15 equidistant lag classes
  - cutoff-distance of 14315 m corresponds to one third of the diagonal of the bounding box of the total Antwerp modelling domain.
- Three variations:
  - all receptors (“all”)
  - street canyons only (“SC”)
  - non-street canyons only (“non-SC”)
- Fit point-centered variogram models to the clouds.
  - Using a spherical variogram model
  - Evaluate and filter out (remove) singular model fits (non-convergence)
- Define the semivariance at the limits of spatial representativeness.
  - Threshold values for the maximum relative deviation of concentrations permissible
  - 25% ($\text{PM}_{10}$), 15% ($\text{NO}_2$), 15% ($\text{O}_3$) at the 2$\sigma$-level
  - To obtain these thresholds, used the DQO of European Directive 2008/50/EC as a proxy.
- Invert the variogram models to estimate the distance of spatial representativeness (dist.SR).
Fitting the Variogram Models (examples)

Omnidirectional variogram clouds

SR thresholds

\[ \gamma(h_{SR}) = \frac{1}{2} \left( \ln \left(1 + \frac{DQQ}{2}\right) \right)^2 \]

PM\(_{10}\)  NO\(_2\)  Ozone

25 %  15 %  15 %

15 lag-classes

Spherical Variogram Model

\[ \gamma_{cp}(h) = C_0 + C_1 \left[ 1.5 \frac{h}{a} - 0.5 \left( \frac{h}{a} \right)^3 \right] \quad \text{if} \quad 0 \leq h \leq a \]

\[ \gamma_{cp}(h) = C_0 + C_1 \quad \text{if} \quad h > a \]
Exception handlings:

- Required semivariance threshold might not be reached within the range of the variogram.
- In such cases we chose the distance of spatial representativeness to equal the value of the range parameter.
- Other interpretations are conceivable ("infinite SR" ?)
**Time Series of SR-distance estimates:**

\( \text{dist}_{SR} \) in [km] for station Linkeroever (urban background site)

Estimated \( \text{dist}_{SR} \) tends to be larger when only receptor points of corresponding station types are considered for the analysis (as expected; more homogeneous).
Time Series of SR-distance estimates:

- Linkeroever (v7: noSC)
- Schoten (v17: noSC)
- Borgerhout-Straatkant (v216: SC)

Note the different scales of the Y-axis: $\text{dist}_{\text{SR}}(\text{PM}_{10}) > \text{dist}_{\text{SR}}(\text{O}_3) > \text{dist}_{\text{SR}}(\text{NO}_2)$
Time Series of SR-distance estimates from evaluating the full dataset

Might trigger questions like: Does one need more dense NO\textsubscript{2} observations in summertime?
How do the estimates from the Point-Centered Variography compare to the outcomes of the SR Intercomparison Exercise?

with contributions from:
José Luis Santiago & Fernando Martin (CIEMAT), Antonio Piersanti, Giuseppe Cremona, Gaia Righini & Lina Vitali (ENEA), Kevin Delaney (EPA IE), Bidroha Basu & Bidisha Ghosh (TCD), Wolfgang Spangl & Christine Brendle (FEA-AT), Jenni Latikka (FMI), Anu Kousa (HSY), Erkki Pärjälä (City of Kuopio), Miika Meretoja (City of Turku), Laure Malherbe, Laurent Letinois & Maxime Beauchamp (INERIS), Fabian Lenartz(ISSeP), Virginie Hutsemekers (AwAC), Lan Nguyen & Ronald Hoogerbrugge (RIVM), Kristina Eneroth & Sanna Silvergren (City of Stockholm), Hans Hooyberghs, Peter Viaene, Bino Maiheu & Stijn Janssen (VITO), David Roet (VMM)
Size and Location of estimated SR areas (NO₂ at site v17)

**ENA: site v17**
- Estimated SR area: 1.7 km²
- NO₂ level: [Graph showing concentration levels]

**EPAIE: site v17**
- Estimated SR area: 27.4 km²
- NO₂ level: [Graph showing concentration levels]

**SLB: site v17**
- Estimated SR area: 19.9 km²
- NO₂ level: [Graph showing concentration levels]

**VITO: site v17**
- Estimated SR area: 289 km²
- NO₂ level: [Graph showing concentration levels]

**FEA-AT: site v17**
- Estimated SR area: 180 km²
- NO₂ level: [Graph showing concentration levels]

**FI: site v17**
- Estimated SR area: 58.6 km²
- NO₂ level: [Graph showing concentration levels]

**VMM: site v17**
- Estimated SR area: 1.21 km²
- NO₂ level: [Graph showing concentration levels]

**INERIS: site v17**
- Estimated SR area: 66.9 km²
- NO₂ level: [Graph showing concentration levels]

**ISSEPAWAC: site v17**
- Estimated SR area: 71.5 km²
- NO₂ level: [Graph showing concentration levels]

**SR areas for NO₂ at site v17**

- Median = 43 km²
Interim Conclusion from the IE have been:

- The Spatial Representativeness Areas estimated by the different participants are quite diverse.
- The results in particular reveal an enormous scattering of the extent and position of the estimated polygons.
- This diversity of results deserved a closer look behind the scenes.
How to compare the PCV results with the IE?

- Subject to site specific conditions and to different SR approaches, SR areas can have quite complex, irregular and even discontinuous shapes.
- In contrast, the point-centered variogram method (as presented here) delivers on single value (distance of spatial representativeness: dist$_{SR}$).
- From a conceptual point of view, the latter corresponds to the conception of a simple circular shaped area of representativeness.
- We need to accept, that this is likely oversimplified.
- In order to compare the results, we recalculate the PCV dist$_{SR}$ to the surface area equivalent of a corresponding circle.
Overview:
results from the Intercomparison
SR areas for NO₂ at site v7

- Background: 0.08 km²
- Median: 8 km²

SR areas for O₃ at site v7

- Median: 104.7 km²

SR areas for PM₁₀ at site v7

- Median: 34.8 km²

Results from PCV

- PCV results tend to be on the lower end of the conceivable SR scale

- PCV delivers rather strict SR estimates
**Conclusions (1)**

Depending on the spatial scale of the investigation, the **Point-centered Variogram** places a monitoring station in the context of the local or regional air quality pattern.

The Point-centred Variogram does not, however, serve as a substitute for the **traditional variogram** in the sense that geostatistical methods like kriging require a model fitted for the traditional variogram.

**Point-centred Variography** can on the other hand provide valuable information with regard to the **spatial representativeness** of air quality monitoring sites.

We may also obtain information about the **temporal variation of SR**.

However, a comparison with results obtained by other **spatial representativeness approaches** or based on different conceptualizations is not necessarily simply one-to-one.
**Way forward:**

The concept of a **single spatial representativeness distance** (dist.SR) value implies the assumption of a **radially symmetric area** of spatial representativeness. This corresponds to the use of an **omni-directional variogram**.

The omni-directional approach is **probably overly simplified** and more detailed information (i.e. about the **anisotrophy** of the variogram) could be extracted from the data.

In **future developments** it would be **recommendable** to extend the evaluation by applying **directional variograms**. Disadvantage could be the **limited number of data-pairs** available in the individual directional sectors.

Thank you for your attention!
Participants agreed that the discrepancies observed in this exercise require further efforts towards the quantitative definition of the concept of “the area of representativeness” and in eliminating unnecessary differences in the methodologies.

In the second part of the workshop it was therefore more intensively discussed, if for the aim of harmonization the concept of spatial representativeness would require a paradigm shift in its definition:

1) What are the future needs for harmonization and for establishing a common frame of reference?
2) Is there a future need for standardization, too?
3) Beyond standardization, should the regulators / political bodies make the use of standards mandatory?
4) Would it conversely be preferable to have at disposal a set of transparent definitions and practical guidelines, but maintaining the freedom of choosing the most appropriate procedures for the different particular purposes and applications?

It was found consensus amongst participants that currently it would not (yet) be reasonable to start discussing about (2) or (3), but that for the mid term future the efforts of the experts community should be directed towards (4) first.