18th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
9-12 October 2017, Bologna, Italy

SPATIAL REPRESENTATIVENESS EVALUATION BY POINT CENTRED VARIOGRAPHY

Oliver Kracht and Michel Gerboles

European Commission - Joint Research Centre (JRC), Ispra, Italy

Abstract: Within the FAIRMODE cross-cutting activity group on spatial representativeness a geostatistical technique based on point-centered semi-variograms has been proposed, which can be used to derive valuable information about the spatial representativeness of individual air quality monitoring sites. Whereas classical geostatistical analysis describes the spatial correlation structure of a concentration field in terms of the variogram, point centred variography on the other hand 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. It thereby places a monitoring station in the context of the local or regional air quality pattern. We demonstrate how a mathematical inversion of the point centred variogram can provide estimates about the extent of the spatial representativeness area. The application of this approach is tested on a set of modelling data from the city of Antwerp, which was later used for the FAIRMODE / AQUILA intercomparison exercise of methods for the assessment of the spatial representativeness.

Key words: *point centred variography, spatial representativeness, geostatistics, air pollution, air quality monitoring, measurements, modelling*

INTRODUCTION

Commonly used definitions for the spatial representativeness of an air quality monitoring site are established on an evaluation of the similarity of pollutant concentrations around this point. Hence, in its 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. Whereas in this context classical geostatistical analysis would describe the spatial correlation structure of the whole concentration field in terms of the variogram, the point centred variography 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.

The point centred variography thus places a monitoring station in the context of the local or regional air quality pattern. It thereby enables systematic evaluation of the spatial relationship between point observations of pollutant concentrations at a particular monitoring site and the corresponding concentration field within its immediate and / or wider environment. In a final step, a mathematical inversion of the point centred variogram can be linked to the data quality objectives of the European Directive 2008/50/EC, thus providing information about the extent of the spatial representativeness area.

MAHEMATICAL FRAMEWORK

The Point Centred Semivariance

The point centred experimental semivariance is defined as the average of squared differences of within data pairs formed between a particular central point *(cp)* and all other points in the domain that are separated from this central point by a lag distance *h*:

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

where $N_{cp,h}$ is the total number of data pairs formed with the central point at lag distance h, and $Z(s_{cp})$ and $Z(s_{cp}+h)$ are the values of Z at the corresponding locations (s_{cp}) and $(s_{cp}+h)$.

As for the traditional experimental variogram, the lag distance h can be accompanied by a tolerance interval to create distance classes $\bar{h}_j$. For each lag class, the point centred experimental semivariance is then estimated to:

$$\hat{\gamma}_{cp}\left(\bar{h}_j\right)=\frac{1}{2}\frac{1}{N_{h,cp}}\sum_{N_{cp,h}}\left[Z\left(s_{cp}\right)-Z\left(s_{cp}+h\right)\right]^2 \quad \forall\, h\in\bar{h}_{j,cp} \tag{2}$$

Likewise a point centred variogram cloud can be created that collects the individual point-pair contributions to the final point centred variogram. If n is the total number of observations within a spatial dataset, the full point centred variogram cloud consists of $N_{full\ cloud,\ pc}$ point pairs according to:

$$N_{full\ cloud,\ pc}=(n-1) \tag{3}$$

Comparing the traditional variogram and the point centred variogram it should be noted that different types of variograms are needed for different purposes. For its scope of applications, the point centred variogram $\gamma_{cp}(h)$ does not in fact serve as a substitute for the traditional variogram $\gamma(h)$ in the sense that geostatistical methods like kriging require a model for the traditional variogram. Rather than this, the aim of the point centred variogram is to provide additional information and a clearer description of the spatial continuity around a central reference point.

Interrelation between the Point Centred Variogram and Spatial Representativeness

In the following we will establish a link between the information provided by point centred variography and these limits of the spatial representativeness area. In fact, most of the commonly used definitions of spatial representativeness are based on the similarity of concentrations of a specific pollutant around a monitoring site. In this way the representativeness area is defined as the area where the concentration $z(x_i)$ at locations x_i does not differ from the concentration $z(x_{cp})$ measured at the monitoring station located at x_{cp} (central point) by more than a specified threshold Δz.

The point centred semivariance in effect provides a measure of dissimilarity between the pollutant concentrations observed at different locations and the corresponding reference concentration observed at the central point x_{cp}. Let h_{SR} be the lag distance at the limits of spatial representativeness around the central point x_{cp} of a point centred variogram, and $z(x_{cp}+h_{SR})$ the pollutant concentration at locations positioned at this limit. The semivariance at the limits of spatial representativeness can then be calculated to be

$$\gamma(h_{SR})=\frac{1}{2}\left(z(x_{cp})-z(x_{cp}+h_{SR})\right)^2=\frac{1}{2}\left(z(x_{cp})-\left(z(x_{cp})+\Delta z_{threshold}\right)\right)^2 \tag{4}$$

where $\Delta z_{threshold}$ is the maximum permissible deviation of concentrations within the limits of spatial representativeness. This relationship can then be reduced to:

$$\gamma(h_{SR})=\frac{1}{2}\left(\Delta z_{threshold}\right)^2 \tag{5}$$

which immediately provides the relevant threshold value for $\gamma(h_{SR})$ in absolute units of the semivariance. The lag distance h_{SR} can then be computed by inverting the corresponding semivariance model function obtained beforehand from a fit to the experimental data.

APPLICATION STUDY

The application of the point centred variography approach has been tested on a set of modelling data from the city of Antwerp. This dataset contains information at a very high spatial (street level) and temporal resolution for three main pollutants (PM_{10}, NO_2 and Ozone), over the whole city. The underlying model results, among other features comprising gridded time series for a number of 341 virtual receptor points, have been prepared by VITO (Belgium) by applying the RIO-IFDM-OSPM model chain (e.g., Lefebvre et al. 2013). Furthermore, the FAIRMODE (Forum for Air Quality Modelling in Europe) cross-cutting activity group on SR in cooperation with AQUILA (the European Network of Air Quality Reference Laboratories) recently concluded an intercomparison exercise on spatial representativeness methods, which was also based on sharing this dataset (Kracht et al. 2016). A basic overview of the Antwerp modelling domain and the example of the annual average concentration field for NO_2 is provided in **figure 1**.

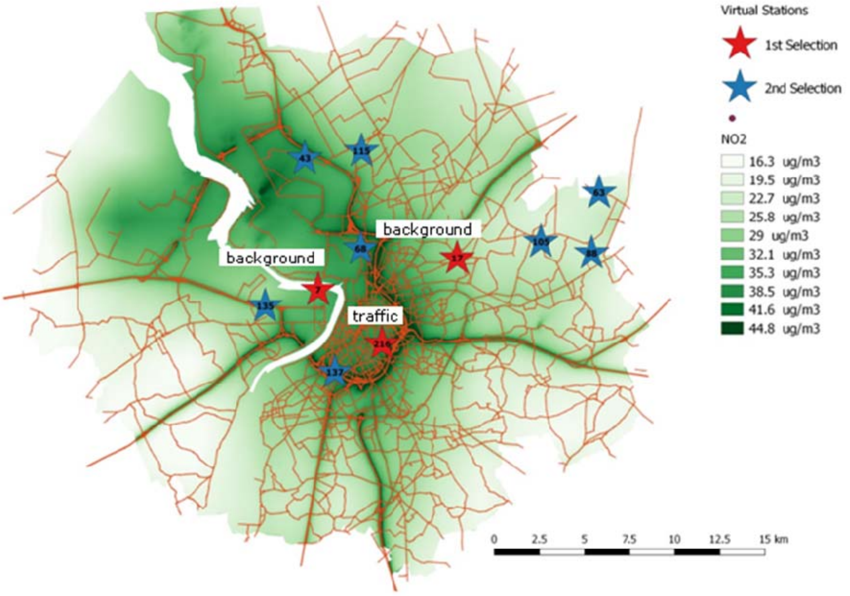

Figure 1. Overview of the Antwerp modelling domain, showing the annual average concentration field for NO_2 (green colours). A basic road network (brown lines) is drawn for orientation. Red stars highlight the three monitoring stations Linkeroever (cp7), Schoten (cp17) and Borgerhout (cp216).

From this dataset, the following three monitoring sites have been selected for closer evaluation:

1) As an example for the traffic sites:
   - Borgerhout II (Belgium Lambert 72 coordinates: 154396 / 211055): virtual point cp216

2) As examples for the urban background sites:
   - Antwerpen-Linkeroever (Belgium Lambert 72 coordinates: 150865 / 214046): virtual point cp7
   - Schoten (Belgium Lambert 72 coordinates: 158560 / 215807): virtual point cp17

RESULTS

The algorithms for point centred variography have been applied to the aggregated time series of the Antwerp dataset (time series of 14-day averages of PM_{10}, NO_2 and Ozone, and 1-day averages of PM_{10}). Following the model fitting of point centred variograms, individual estimates for the limits of spatial representativeness (**dist.SR**) have been calculated by inverting the fitted variogram model functions. **Table 1** exemplifies a summary statistics of the estimates obtained for the 14-day averages time series. In this table, a comparison is also made between results obtained by considering all 341 virtual receptor points (columns denoted as **_all**), and results obtained by using only the 241 non-street-canyon points for the evaluation of virtual monitoring stations cp7 and cp17 (columns denoted as **_noSC**), and only the 100 street-canyon points for the evaluation of virtual monitoring station cp216 (columns denoted as **_SC**).

Table 1. Summary statistics of estimated limits of spatial representativeness (dist.SR) obtained from the inversion of point centred variograms.

PM_{10} (based on 14-day average concentrations, ΔPM_{10}-threshold = 25%)						
dist.SR	**cp7_all**	**cp7_noSC**	**cp17_all**	**cp17_noSC**	**cp216_all**	**cp216_SC**
min	3822 m	5976 m	1381 m	1836 m	0 m	1325 m
1st quartile	6739 m	8729 m	2074 m	2518 m	1063 m	1863 m
median	**7457 m**	**10864 m**	**2670 m**	**3251 m**	**1925 m**	**2586 m**
3rd quartile	9477 m	12413 m	3530 m	4880 m	4015 m	4334 m
max	12928 m	14278 m	8720 m	7101 m	9634 m	10606 m
criterion used	**cp7_all**	**cp7_noSC**	**cp17_all**	**cp17_noSC**	**cp216_all**	**cp216_SC**
estimated from threshold	62%	65%	92%	81%	100%	100%
estimated from range	19%	4%	8%	19%	0%	0%
NA because dist.SR > cutoff	19%	31%	0%	0%	0%	0%
NO_2 (based on 14-day average concentrations, ΔNO_2-threshold = 15%)						
dist.SR	**cp7_all**	**cp7_noSC**	**cp17_all**	**cp17_noSC**	**cp216_all**	**cp216_SC**
min	87 m	148 m	0 m	45 m	0 m	0 m
1st quartile	116 m	218 m	52 m	87 m	0 m	0 m
median	**161 m**	**273 m**	**69 m**	**130 m**	**0 m**	**0 m**
3rd quartile	210 m	391 m	120 m	178 m	0 m	0 m
max	385 m	679 m	175 m	237 m	0 m	0 m
criterion used	**cp7_all**	**cp7_noSC**	**cp17_all**	**cp17_noSC**	**cp216_all**	**cp216_SC**
estimated from threshold	100%	100%	100%	100%	100%	100%
estimated from range	0%	0%	0%	0%	0%	0%
NA because dist.SR > cutoff	0%	0%	0%	0%	0%	0%
Ozone (based on 14-day average concentrations, ΔO_3-threshold = 15%)						
dist.SR	**cp7_all**	**cp7_noSC**	**cp17_all**	**cp17_noSC**	**cp216_all**	**cp216_SC**
min	0 m	0 m	131 m	143 m	0 m	0 m
1st quartile	505 m	772 m	223 m	203 m	0 m	387 m
median	**1111 m**	**929 m**	**298 m**	**262 m**	**180 m**	**658 m**
3rd quartile	2068 m	1627 m	455 m	452 m	298 m	1261 m
max	3491 m	3103 m	783 m	723 m	1086 m	4365 m
criterion used	**cp7_all**	**cp7_noSC**	**cp17_all**	**cp17_noSC**	**cp216_all**	**cp216_SC**
estimated from threshold	100%	100%	100%	100%	100%	100%
estimated from range	0%	0%	0%	0%	0%	0%
NA because dist.SR > cutoff	0%	0%	0%	0%	0%	0%

DISCUSSION

For the two background sites at cp7 and p17, median values for the spatial representativeness distance of PM_{10} extend between 2277 m (cp17_all for daily PM_{10}) and 10864 m (cp7_noSC for 14-day average PM_{10}). The median value for PM_{10} for the traffic site cp216 ranges between 1529 m (cp216_all for daily PM_{10}).and 2586 m (cp216_SC for 14-day average PM_{10}). For Ozone 14-day averages the estimated limits of spatial representativeness for the two background sites cp7 and p17 have median values between 262 m (cp17_noSC) and 1111 m (cp7_all). For NO_2 the estimated limits of spatial representativeness are clearly shorter than for PM_{10} and Ozone. Particularly for the traffic site cp216 a zero distance of spatial representativeness was found.

As a general observation, the estimated values for the limits of spatial representativeness are larger when variograms are based on data which are restricted to the corresponding station area types (_noSC for the background stations at cp7 and p17, and _SC for the traffic station at cp216), as compared to those results obtained by considering all virtual monitoring points simultaneously. This was anticipated, as the set of monitoring points becomes more homogeneous when street canyon and non-street canyon sites are distinguished from another. The only exceptions are the cases of Ozone for the background stations cp7 and p17, where the limits of spatial representativeness are a little smaller for the groups cp7_noSC and cp17_noSC as compared to the groups cp7_all and cp17_all.

With regard to the integration time-scales, the estimated distances of spatial representativeness tend to be higher for the PM_{10} data based on 14-day averages than for PM_{10} based on daily values (daily values have not been investigated for NO_2 and Ozone; they can thus not be compared).

In summary, the three virtual monitoring stations can consistently be ranked for all three pollutants: The distance of spatial representativeness tends to be highest for virtual station cp7 (corresponding to the urban background station Antwerpen-Linkeroever), second highest for virtual station cp17 (corresponding to the urban background station Schoten) and lowest for virtual station cp216 (corresponding to the traffic station Borgerhout-Straatkant).

SUMMARY & CONCLUSIONS

Depending on the spatial scale of the investigation, point centred variography places a monitoring station in the context of the local or regional air quality pattern. It thereby enables systematic evaluation of the spatial relationship between point observations of pollutant concentrations at this monitoring site and the corresponding concentration fields within its immediate and / or wider environment. Point centred variography can thus provide valuable information with regard to the spatial representativeness of the air quality monitoring site. 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.

Time series of spatial representativeness results have been inferred from the Antwerp dataset for three selected monitoring station locations. With regard to the transferability and generalisation of results, it needs to be pointed out that in this exercise the evaluation of spatial representativeness was specifically done from the methodological perspective of the point centred variography. A comparison with results obtained by other spatial representativeness approaches or based on different conceptualizations is not necessarily simply one-to-one. It should rather be anticipated that the integration of information obtained by different spatial representativeness methodologies requires a certain degree of technical effort and of expert knowledge to be applied.

A set of methodological recommendations has been summarized in a recent JRC technical report (Kracht et al. 2017) that can be used for planning further developments of this method. These proposals for further developments do specifically include suggestions for (i) possible variations of the underlying type of the variogram (directional variogram, relative variograms), (ii) modifications of the variogram model functions, (iii) the criteria deployed for defining the limits of the spatial representativeness area, (iv) the numerical procedures, and (v) the pre-treatment and selection of datapoints.

REFERENCES

Kracht, O., Hooyberghs, H., Lefebvre, W., Janssen, S., Maiheu, B., Martin, F., Santiago, J.L., Garcia, L. and Gerboles, M. (2016): FAIRMODE Intercomparison Exercise - Dataset to Assess the Area of Representativeness of Air Quality Monitoring Stations, 267 pp, *JRC Technical Reports,* ***102775***, EUR 28135 EN, ISBN 978-92-79-62295-3 (PDF), DOI 10.2790/479282.

Kracht, O., Solazzo, E. and Gerboles, M. (2017): Point Centred Variography to Assess the Spatial Representativeness of Air Quality Monitoring Sites: Application to the Datasets of the FAIRMODE Intercomparison Exercise of Spatial Representativeness. 135 pp. *JRC Technical Reports*, ***107479***, EUR 28726 EN, ISBN 978-92-79-71654-6 (PDF), DOI 10.2760/64951.

Lefebvre, W., Van Poppel, M., Maiheu, B., Janssen, S., and Dons, E (2013): Evaluation of the RIO-IFDM-street canyon model chain, *Atmospheric Environment*, ***77***, 325-337.