

---

## **Spatial representativeness evaluation of air quality monitoring sites by point-centred variography**

---

**Oliver Kracht\***

European Commission – Joint Research Centre,  
Directorate C – Energy, Transport and Climate,  
Air and Climate Unit, Via E. Fermi 2749,  
I-21027, Ispra, Italy

Email: [oliver.kracht@gmx.de](mailto:oliver.kracht@gmx.de)

and

International Atomic Energy Agency – Vienna International Centre,  
P.O. Box 100, 1400 Vienna, Austria

\*Corresponding author

**Michel Gerboles**

European Commission – Joint Research Centre,  
Directorate C – Energy, Transport and Climate,  
Air and Climate Unit, Via E. Fermi 2749,  
I-21027, Ispra, Italy

Email: [michel.gerboles@ec.europa.eu](mailto:michel.gerboles@ec.europa.eu)

**Abstract:** We propose a geostatistical technique based on point-centred semi-variograms that can be used to derive valuable information about the spatial representativeness of air quality monitoring sites. Whereas classical geostatistics describes the spatial correlation structure of a concentration field in terms of the variogram, point-centred variography is based on the average of squared concentration differences observed in pairs formed from 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 of the extent of the representativeness area of a monitoring site. The application of this approach is tested on a set of modelling data from the city of Antwerp, which was used for the FAIRMODE/AQUILA intercomparison exercise of methods for the assessment of spatial representativeness.

**Keywords:** point-centred variography; spatial representativeness; geostatistics; air pollution; air quality monitoring; measurements; modelling.

**Reference** to this paper should be made as follows: Kracht, O. and Gerboles, M. (2019) 'Spatial representativeness evaluation of air quality monitoring sites by point-centred variography', *Int. J. Environment and Pollution*, Vol. 65, Nos. 1/2/3, pp.229–245.

**Biographical notes:** Oliver Kracht was a Post-Doctoral Researcher and contracted scientist at the European Commission Joint Research Centre in Ispra (Italy) from 2012 to 2017, where he worked on geostatistical methods for the evaluation of ambient air quality data. He graduated from the Department of Geosciences at the University of Bochum (Germany). After an employment at the University of Lausanne (Switzerland), he moved to the Swiss Federal Institute of Aquatic Science and Technology. In 2007, he obtained his PhD in Environmental Engineering from the ETH Zurich. He later worked as a PostDoc at the University of Bern (Switzerland), and as a Marie Curie Fellow at the Geological Survey of Norway. Since 2018 he is working as a technical officer at the International Atomic Energy Agency (IAEA) in Vienna (Austria).

Michel Gerboles graduated in Analytical Chemistry from the University of Bordeaux (FR). He joined the Joint Research Centre of the European Commission in 1990, as quality officer of the European Reference Laboratory for Air Pollution. His activities range from the improved calibration of primary reference methods for air quality monitoring to the organisation of proficiency testing and intercomparison exercises in support of EU air quality policy. His current research interests include the validation of novel inexpensive gas sensor technologies, and the development of automatic methods for the extraction of quality indicators from large ambient air monitoring data sets.

This paper is a revised and expanded version of a paper entitled ‘Spatial representativeness evaluation of air quality monitoring sites by point-centred variography’ presented at the 18th International conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Bologna, Italy, 9–12 October 2017.

---

## 1 Introduction

The key question about the spatial representativeness (SR) of ambient air quality measurements is to what extent a point observation at a particular monitoring site can be extended to conclude about the air pollutant concentrations around it. From a regulatory point of view, the European Commission has worked intensively on the implementation of a harmonised programme for the monitoring of air pollution. The European Directives 2008/50/EC and 2004/107/EC aim to improve the quality of measurements and data collection, to ensure that the information on air pollution is sufficiently representative and comparable across the community. However, though these directives include several considerations about the order of magnitude of the SR of a monitoring site, no detailed provisions on the methods for assessing SR are provided. Also in the scientific literature, no well-established procedures have been identified so far (Kracht et al. 2017b).

Commonly used definitions for the spatial representativeness of an air quality monitoring site are usually based 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 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.

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.

## 2 Mathematical framework

### 2.1 The point centred semivariance

The point centred experimental semivariance is defined as one half of the average of squared differences 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  $h$ :

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

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 as:

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

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

$$N_{full\ cloud, cp} = (n - 1) \quad (3)$$

For practical reasons, it can be helpful to introduce a maximum spatial separation distance  $h_{cutoff}$  up to which point pairs are included into the point centred semivariance estimates. The number of point pairs in the variogram cloud then reduces according to:

$$N_{cloud} \leq N_{full\ cloud, cp} \text{ with } N_{cloud} = f(N_{full\ cloud, up}, h_{cutoff}) \quad (4)$$

In common geostatistical applications, the experimental semivariance values are often approximated by a simple continuous model function in which the semi-variance  $\gamma$  is described as a function of lag distance  $h$ . Such a model fit is referred to as the theoretical variogram. In this context the Gaussian, the exponential or the spherical variogram models are the most commonly used. The spherical model [equation 5(a)] is often considered the best choice when spatial autocorrelation decreases to a point after which it becomes zero.

$$\begin{aligned} \gamma(h) &= C_0 + C_1 \left[ 1.5 \frac{h}{a} - 0.5 \left( \frac{h}{a} \right)^3 \right] & \text{if } 0 \leq h \leq a \\ \gamma(h) &= C_0 + C_1 & \text{if } h > a \end{aligned} \quad (5a)$$

For the point centred variogram, equation 5(a) straightforwardly reformulates to:

$$\begin{aligned} \gamma_{cp}(h_{cp}) &= C_0 + C_1 \left[ 1.5 \frac{h_{cp}}{a} - 0.5 \left( \frac{h_{cp}}{a} \right)^3 \right] & \text{if } 0 \leq h_{cp} \leq a \\ \gamma_{cp}(h_{cp}) &= C_0 + C_1 & \text{if } h_{cp} > a \end{aligned} \quad (5b)$$

which represents the point centred spherical variogram model.

The parameters of the spherical model are the nugget  $C_0$ , the partial sill  $C_1$ , and the range  $a$ . The nugget variance  $C_0$  represents the variability of the observations at small distances (tending towards 0). The empirical nugget variance is unknown since it is the value of the theoretical variogram at the origin. The nugget parameter  $C_0$  is thus estimated by extrapolating the variogram towards  $h = 0$ . From this point, the semivariance increases until the full sill variance  $C_0 + C_1$  is reached at a lag distance called the range ( $a$ ). The range provides the distance beyond which semivariances remain constant. Up to this distance, observations of the regionalised variable in the sampling locations are correlated, beyond this distance they must be considered to be spatially independent. Note specifically that the term partial sill is used to denote  $C_1$ , whereas the term sill denotes  $C_0 + C_1$ .

In the context of the assessments presented in this exercise, we focus on the use of the spherical model. However, it can be useful to also evaluate the use of alternative variogram models. For example the use of a power model might be considered if the established variograms do not have well defined sills.

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, in fact, the point centred variogram  $\gamma_{cp}(h)$  does not 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.

## 2.2 *Interrelation between the point centred variogram and spatial representativeness*

In the following we will demonstrate the link between the information provided by point centred variography and the 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  $\Delta z$ .

The point centred semivariance in effect provides a measure of the 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 as

$$\gamma(h_{SR}) = \frac{1}{2} \left( z(x_{cp}) - z(x_{cp} + h_{SR}) \right)^2 = \frac{1}{2} \left( z(x_{cp}) - \left( z(x_{cp}) + \Delta z_{threshold} \right) \right)^2 \quad (6)$$

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

$$\gamma(h_{SR}) = \frac{1}{2} \left( \Delta z_{threshold} \right)^2 \quad (7)$$

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.

When the point centred semivariance is calculated for log-transformed data, the threshold value  $\gamma(h_{SR})$  needs to be determined in a slightly different way. In this case equation (7) needs to be modified to:

$$\begin{aligned} \gamma(h_{SR}) &= \frac{1}{2} \left( \ln(z(x_{cp})) - \ln(z(x_{cp} + h_{SR})) \right)^2 \\ &= \frac{1}{2} \left( \ln(z(x_{cp})) - \ln(z(x_{cp}) + \Delta z_{threshold}) \right)^2 \end{aligned} \quad (8)$$

which can be converted to:

$$\gamma(h_{SR}) = \frac{1}{2} \left( \ln \left( \frac{z(x_{cp}) + \Delta z_{threshold}}{z(x_{cp})} \right) \right)^2 = \frac{1}{2} \left( \ln \left( 1 + \frac{\Delta z_{threshold}}{z(x_{cp})} \right) \right)^2 \quad (9)$$

By introducing  $\Delta_r z_{threshold}$ , which is the maximum relative deviation of concentrations permissible within the limits of spatial representativeness, this relationship can then be further reduced to:

$$\gamma(h_{SR}) = \frac{1}{2} \left( \ln \left( 1 + \frac{\Delta_r z_{threshold} \cdot z(x_{cp})}{z(x_{cp})} \right) \right)^2 = \frac{1}{2} \left( \ln(1 + \Delta_r z_{threshold}) \right)^2 \quad (10)$$

In order to establish a suitable reference value for  $\Delta_r z_{threshold}$  for the purpose of this exercise we can refer to the data quality objectives (DQO) provided in Annex 1 of the European Directive 2008/50/EC on ambient air quality and cleaner air for Europe. For

fixed measurements the uncertainties (expressed at a 95% confidence level) of the assessment methods are given as:

**Table 1** Data quality objectives used as a default input for the similarity criterion thresholds

$PM_{10}$	$NO_2$	Ozone
25 %	15 %	15 %

We need to consider that these DQOs are expressed at the 95% confidence level, which corresponds to approximately two times the standard deviation ( $2\sigma$ -level), whereas the variogram is conventionally providing semivariance values corresponding to the  $1\sigma$ -level.

Considering this necessary conversion, the logarithmic semivariance threshold value for the lag distance corresponding to the limit of spatial representativeness ( $h_{SR}$ ) is thus finally calculated as:

$$\gamma(h_{SR}) = \frac{1}{2} \left( \ln \left( 1 + \frac{DQO}{2} \right) \right)^2 \quad (11)$$

### 2.3 Inversion of the spherical variogram model

In order to extract information about spatial representativeness, we need to solve the spherical model equation for the priori unknown lag distance  $h$ , at which a certain semivariance  $\gamma(h)$  is reached. For example,  $\gamma(h)$  might then be set to a predefined limit value  $\gamma(h_{SR})$  to still accept spatial representativeness. From equation [5(a)] (describing the spherical variogram model) we consider that if  $0 \leq h \leq a$  the semivariance  $\gamma(h)$  is given by:

$$\gamma(h) = C_0 + C_1 \left[ 1.5 \frac{h}{a} - 0.5 \left( \frac{h}{a} \right)^3 \right] \quad (12)$$

which can be transformed to:

$$\frac{\gamma(h) - C_0}{C_1} = 1.5 \frac{h}{a} - 0.5 \left( \frac{h}{a} \right)^3 \quad (13)$$

Equation (13) can then easily be rearranged into the general form of the cubic equation:

$$0.5 \left( \frac{h}{a} \right)^3 - 1.5 \frac{h}{a} + \frac{\gamma(h) - C_0}{C_1} = 0 \quad (14)$$

And then into the depressed cubic form:

$$h^3 - 3a^2 * h + 2a^3 \frac{\gamma(h) - C_0}{C_1} = 0 \quad (15)$$

In the general case, a cubic equation with real coefficients has three solutions, some of which may equal each other if they are real, and two of which may be complex non-real numbers. We can find these solutions by following Cardano's method (Gerolamo Cardano, 1545). For doing this, we first define:

$$p = -3a^2 \quad (16)$$

and

$$q = 2a^3 \frac{\gamma(h) - C_0}{C_1} \quad (17)$$

and the discriminant  $D$  as being:

$$D = \left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2 \quad (18)$$

In the case of  $D < 0$  the cubic equation has three real roots. By using the trigonometric method these three solutions can be found to be:

$$\begin{aligned} h_1 &= 2 \cdot \sqrt{\left|\frac{p}{3}\right|} \cdot \cos\left(\frac{\phi}{3}\right) \\ h_2 &= 2 \cdot \sqrt{\left|\frac{p}{3}\right|} \cdot \cos\left(\frac{\phi}{3} + \frac{2}{3} \cdot \pi\right) \\ h_3 &= 2 \cdot \sqrt{\left|\frac{p}{3}\right|} \cdot \cos\left(\frac{\phi}{3} + \frac{4}{3} \cdot \pi\right) \end{aligned} \quad (19)$$

where  $\phi$  is defined as:

$$\phi = \arccos\left(-\frac{q}{2 \cdot \left(\left|\frac{p}{3}\right|\right)^{\frac{3}{2}}}\right) \quad (20)$$

From these three solutions  $h_1, h_2, h_3$  one particular solution  $h_{select}$  that is applicable for the inverted variogram model needs to be selected. For this selection, the criterion

$$0 \leq h_{select} \leq a \quad (21)$$

is applied. This is justified, because we started these derivations by assuming equation [5(a)], which defines the spherical variogram model for exactly this range of  $h$  values.

In the case of  $D > 0$  the cubic equation has one real root and two conjugated complex roots. In the case of  $D = 0$  the cubic equation has three real roots, one of which is duplicated. However, in our application inverting the spherical variogram model we have not encountered any cases of  $D \geq 0$  so far. For brevity, the solutions for the cases of  $D > 0$  and of  $D = 0$  are thus not presented here.

Note that as an alternative procedure, an option for the numerical inversion of the variogram model can be used. In the same way as for the analytical solution, this numerical solution needs to be restricted within the boundaries of 0 and the range  $a$ .

If, for the evaluation of spatial representativeness, an inversion of the variogram model is used, but the required semivariance  $\gamma(h_{SR})$  is not reached within the range of the variogram (i.e., the variogram's total sill is smaller than the required  $\gamma(h_{SR})$  value), a specific exception handling needs to be applied. In such cases we chose to consider the

distance of spatial representativeness to equal the value of the variogram's range parameter  $a$  [see equations 5(a) and 5(b)].

### 3 Numerical tools

All computer codes used in this study were developed in the R environment (R Core Team, 2016). In order to extend the necessary capacities for geostatistical techniques, projections and spatial analyses, we included functionalities from the packages 'sp' (Pebesma and Bivand, 2016; Bivand et al., 2013), 'gstat' (Pebesma, 2004), 'rgdal' (Bivand et al., 2016), 'raster' (Hijmans, 2015). For data manipulation, filtering and working with time series we also used packages 'data.table' (Dowle et al., 2015) and 'zoo' (Zeileis and Grothendieck, 2005).

### 4 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<sub>10</sub>, NO<sub>2</sub> 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 for the model year 2012 (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 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<sub>2</sub> are provided in Figure 1.

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

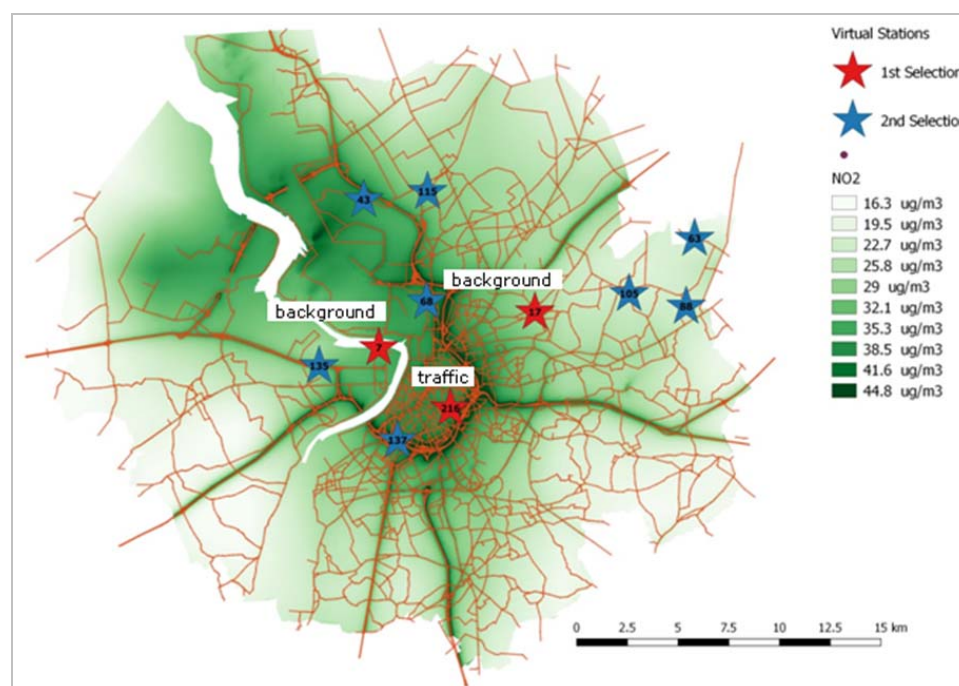
- 1 As an example for the traffic sites:
  - Borgerhout-Straatkant (Belgium Lambert 72 coordinates: 154396/211055): virtual point cp216.
- 2 As examples for the urban background sites:
  - Antwerpen-Linkeroever (Belgium Lambert 72 coordinates: 150865/214046): virtual point cp7.
  - Schoten (Belgium Lambert 72 coordinates: 158560 / 215807): virtual point cp17.

The choice of these three example sites was motivated by the fact that same set of stations were used by the FAIRMODE/AQUILA group in the course of the recently concluded FAIRMODE/AQUILA intercomparison exercise of methods for the assessment of spatial representativeness (Kracht et al., 2017b). It goes without saying that



besides of its relation to the station area type, the extent of spatial representativeness will depend on the pollutant and on the station type, but also on the individual spatial placement of the station and on the integration time scale.

**Figure 1** Overview of the Antwerp modelling domain, showing the annual average concentration field for NO<sub>2</sub> (green background colours) (see online version for colours)



Notes: A basic road network (brown lines) is drawn for orientation. Red stars highlight the three selected monitoring stations ('Virtual Stations, 1st Selection') Linkeroever (cp7), Schoten (cp17) and Borgerhout-Straatkant (cp216).

## 5 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<sub>10</sub>, NO<sub>2</sub> and ozone, and 1-day averages of PM<sub>10</sub>). During a first exploration we tried to establish variogram model fits using a point centred spherical variogram model [equation 5(b)] applied to the immediate pollutant concentrations (μg/m<sup>3</sup>). However, the success rate for achieving acceptable model fits on this direct scale was not satisfactory and for a larger proportion of the data no model fit was found by the automatic algorithm at all. We thus decided to apply a log transformation to the concentration values first. Working on the log-scale indeed improved the success rate and quality of the model fits immediately (Kracht et al., 2017a).

**Table 2** Summary statistics of estimated limits of spatial representativeness ( $dist_{SR}$ ) obtained from the inversion of point centred variograms established at the three monitoring stations Linkeroever (cp7), Schoten (cp17) and Borgerhout-Straatkant (cp216) for the modelling year 2012. Columns denoted as *\_all* are considering all 341 virtual receptor points, columns denoted as *cp7\_noSC* and *cp17\_noSC* are considering only the 241 non-street-canyon points, columns denoted as *cp216\_SC* are considering only the 100 street-canyon points

<i>NO<sub>2</sub> (based on 14-day average concentrations, <math>\Delta NO_2</math>-threshold = 15%)</i>						
	<i>cp7_all</i>	<i>cp7_noSC</i>	<i>cp17_all</i>	<i>cp17_noSC</i>	<i>cp216_all</i>	<i>cp216_SC</i>
<i>dist<sub>SR</sub></i>						
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
<i>median</i>	<i>161 m</i>	<i>273 m</i>	<i>69 m</i>	<i>130 m</i>	<i>0 m</i>	<i>0 m</i>
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
<i>Criterion used</i>						
Estimated from threshold ( $h_{SR}$ )	100%	100%	100%	100%	100%	100%
Estimated from range ( $a$ )	0%	0%	0%	0%	0%	0%
NA because $dist_{SR} > h_{cutoff}$	0%	0%	0%	0%	0%	0%
<i>Ozone (based on 14-day average concentrations, <math>\Delta O_3</math>-threshold = 15%)</i>						
<i>dist<sub>SR</sub></i>						
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
<i>median</i>	<i>1,111 m</i>	<i>929 m</i>	<i>298 m</i>	<i>262 m</i>	<i>180 m</i>	<i>658 m</i>
3rd quartile	2,068 m	1,627 m	455 m	452 m	298 m	1,261 m
max	3,491 m	3,103 m	783 m	723 m	1,086 m	4,365 m
<i>Criterion used</i>						
Estimated from threshold ( $h_{SR}$ )	100%	100%	100%	100%	100%	100%
Estimated from range ( $a$ )	0%	0%	0%	0%	0%	0%
NA because $dist_{SR} > h_{cutoff}$	0%	0%	0%	0%	0%	0%
<i>PM<sub>10</sub> (based on 14-day average concentrations, <math>\Delta PM_{10}</math>-threshold = 25%)</i>						
<i>dist<sub>SR</sub></i>						
min	3,822 m	5,976 m	1,381 m	1,836 m	0 m	1,325 m
1st quartile	6,739 m	8,729 m	2,074 m	2,518 m	1,063 m	1,863 m
median	7,457 m	10,864 m	2,670 m	3,251 m	1,925 m	2,586 m
3rd quartile	9,477 m	12,413 m	3,530 m	4,880 m	4,015 m	4,334 m
max	12,928 m	14,278 m	8,720 m	7,101 m	9,634 m	10,606 m
<i>Criterion used</i>						
Estimated from threshold ( $h_{SR}$ )	62%	65%	92%	81%	100%	100%
Estimated from range ( $a$ )	19%	4%	8%	19%	0%	0%
NA because $dist_{SR} > h_{cutoff}$	19%	31%	0%	0%	0%	0%

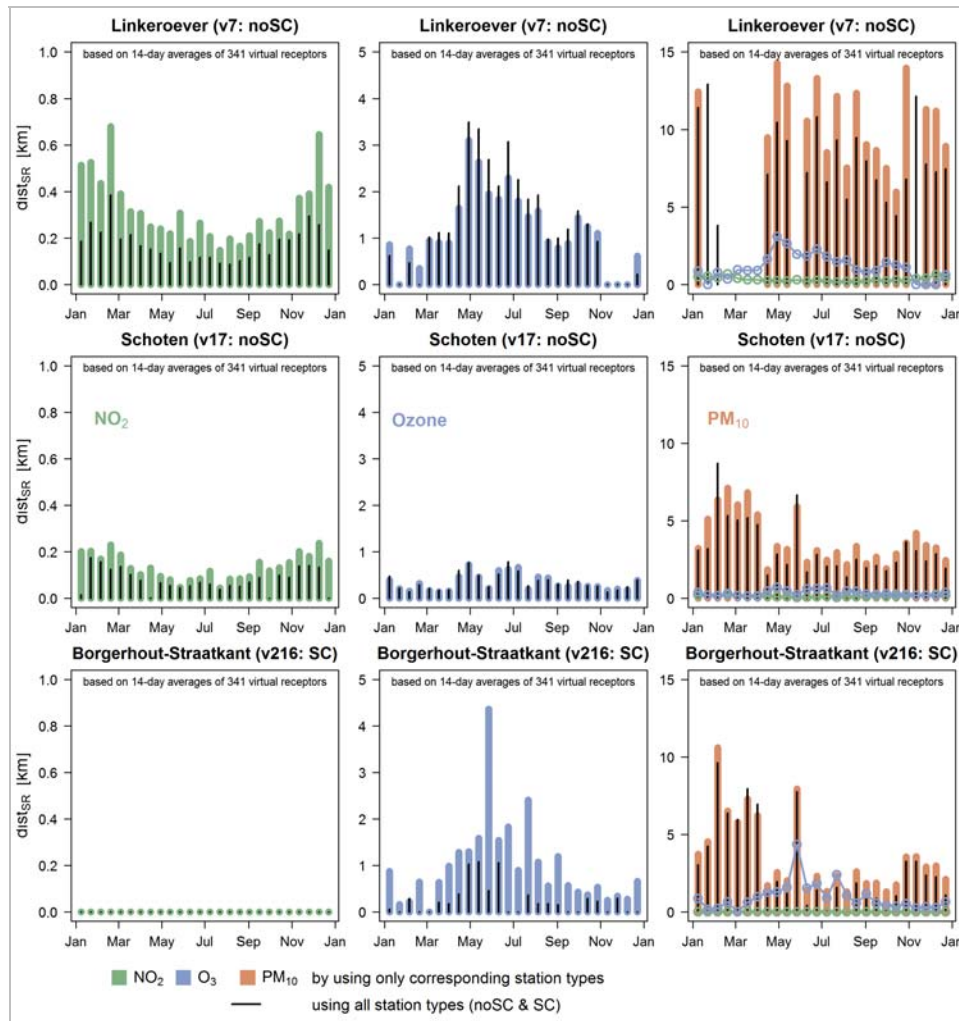
**Table 2** Summary statistics of estimated limits of spatial representativeness ( $dist_{SR}$ ) obtained from the inversion of point centred variograms established at the three monitoring stations Linkeroever (cp7), Schoten (cp17) and Borgerhout-Straatkant (cp216) for the modelling year 2012. Columns denoted as *\_all* are considering all 341 virtual receptor points, columns denoted as *cp7\_noSC* and *cp17\_noSC* are considering only the 241 non-street-canyon points, columns denoted as *cp216\_SC* are considering only the 100 street-canyon points (continued)

<i>PM<sub>10</sub>-daily (based on daily average concentrations, <math>\Delta PM_{10}</math>-threshold = 25%)</i>						
	<i>cp7_all</i>	<i>cp7_noSC</i>	<i>cp17_all</i>	<i>cp17_noSC</i>	<i>cp216_all</i>	<i>cp216_SC</i>
$dist_{SR}$						
min	854 m	1,242 m	600 m	750 m	0 m	0 m
1st quartile	4,272 m	5,433 m	1,585 m	1,883 m	142 m	1,257 m
median	5,864 m	7,124 m	2,277 m	2,653 m	1,529 m	2,348 m
3rd quartile	7,972 m	9,610 m	3,797 m	4,585 m	3,829 m	4,445 m
max	14,149 m	14,295 m	9,217 m	11,378 m	13,704 m	14,193 m
Criterion used						
Estimated from threshold ( $h_{SR}$ )	63%	63%	88%	82%	98%	98%
Estimated from range ( $a$ )	20%	15%	12%	18%	0%	0%
NA because $dist_{SR} > h_{cutoff}$	17%	22%	0%	0%	2%	2%

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 2 shows different 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 simultaneously considering all 341 virtual receptor points (columns denoted as *\_all*), and results obtained by restricting the analysis to only those virtual receptor points of types matching to the respective station types of the corresponding central points. For this latter case, the 241 non-street-canyon points have exclusively been used for the evaluation of the two urban background sites Linkeroever and Schoten (columns denoted as *cp7\_noSC* and *cp17\_noSC*) and the 100 street-canyon points have exclusively been used for the evaluation of the traffic site Borgerhout-Straatkant (columns denoted as *cp216\_SC*).

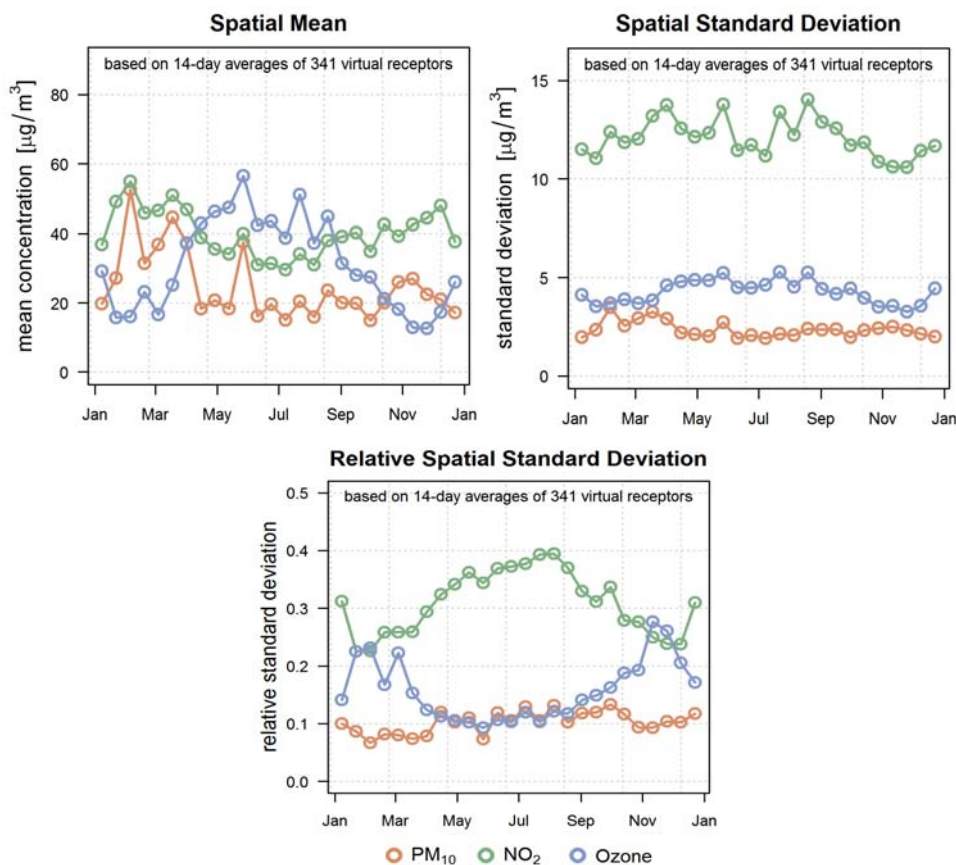
A more detailed breakdown of these results is presented in Figure 2, which presents the individual spatial representativeness distance estimates as annual 14-day averages time series. Again, results obtained by considering simultaneously all 341 virtual receptor points are compared with those obtained by restricting the analysis to the 241 non-street-canyon points or the 100 street-canyon points. It should be noted that because of the different orders of magnitude of the spatial representativeness ranges obtained, a different axis scaling had to be used for each of the three sets of graphical panels representing  $NO_2$ ,  $O_3$  and  $PM_{10}$ .

**Figure 2** Annual time series of the spatial representativeness limits ( $dist_{SR}$ ) for the two urban background sites Linkeroever (v7) and Schoten (v17), and for the traffic site Borgerhout-Straatkant (v216). Green ( $NO_2$ ), blue ( $O_3$ ) and brown bars ( $PM_{10}$ ) show results calculated by exclusively using corresponding station types: the 241 non-street-canyon points matching for Linkeroever and Schoten, and the 100 street-canyon points matching for Borgerhout-Straatkant (see online version for colours)



Notes: Superimposed black bars indicate results obtained by using all 341 virtual receptor points. Note that the x-axes for  $NO_2$  (left),  $O_3$  (centre) and  $PM_{10}$  (right) are on different scales. For better comparison, line charts of the  $NO_2$  and  $O_3$  results have therefore been additionally overlain onto bars of the  $PM_{10}$  panels (right).

**Figure 3** Time series of spatial mean, spatial standard deviation, and relative spatial standard deviation of the 14-day average values of 341 virtual monitoring points for the modelling year 2012 (see online version for colours)



Notes: These metrics reflect the overall means, the total standard deviations and total relative standard deviations of concentrations of all virtual monitoring points within the full spatial extent of the model domain as can be obtained for each 14-day timestep.

## 6 Discussion

For the two background sites at cp7 and p17, median values of the spatial representativeness distance of PM<sub>10</sub> range from 2,277 m (cp17\_all for daily PM<sub>10</sub>) to 10,864 m (cp7\_noSC for 14-day average PM<sub>10</sub>). The median value for PM<sub>10</sub> for the traffic site cp216 ranges from 1,529 m (cp216\_all for daily PM<sub>10</sub>) to 2,586 m (cp216\_SC for 14-day average PM<sub>10</sub>). For ozone 14-day averages the estimated limits of spatial representativeness for the two background sites cp7 and p17 present median values between 262 m (cp17\_noSC) and 1,111 m (cp7\_all). For NO<sub>2</sub> the estimated limits of spatial representativeness are clearly shorter than for PM<sub>10</sub> and ozone. Notably the traffic site cp216 was found to have a zero distance of spatial representativeness.

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 expected, 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 in the two cases of ozone for the background stations cp7 and cp17, where the median limits of spatial representativeness are a little smaller for the restricted groups cp7\_noSC and cp17\_noSC as compared to the full set groups cp7\_all and cp17\_all.

Another important observation is the very short extent of spatial representativeness of NO<sub>2</sub> at the traffic site (Borgerhout-Straatkant, cp216), which always appeared to be 0 m, regardless of using the full set or the restricted set of virtual receptors. This result clearly makes sense, because NO<sub>2</sub> is the most heterogeneous of the three air pollutants considered in this study, resulting into huge NO<sub>2</sub> gradients particularly at the traffic site.

With regard to the integration time-scales, the estimated distances of spatial representativeness tend to be higher for the PM<sub>10</sub> data based on 14-day averages than for PM<sub>10</sub> based on daily values (daily values have not been investigated for NO<sub>2</sub> and ozone; they can thus not be compared). This observation corresponds well with our general expectations about the positive correlation between integration time-scales and the extent of the spatial representativeness areas.

Time series of spatial representativeness distance estimates (Figure 2) reveal a clearly systematic pattern. The extent of spatial representativeness for NO<sub>2</sub> is clearly larger during winter time, while it gets smaller during summer time. Ozone reveals an opposite behaviour, having a larger extent of spatial representativeness during summer, and a smaller extent of spatial representativeness during winter. A less clear pattern, however, is observed for PM<sub>10</sub>, although a small decrease of spatial representativeness can be observed in summer at stations Schoten and Boggerhout-Straatkant (apart from an isolated peak around the end of Mai). This decrease is likely caused by a decrease of PM10 concentrations during summer.

It can be useful to compare these observations to the seasonal evolution of some basic statistical properties of the full set virtual monitoring points. For this purpose, Figure 3 summarises annual time series of spatial mean, spatial standard deviation, and relative spatial standard deviation of the 14-day average values of 341 virtual monitoring points. These metrics reproduce the overall means, the total standard deviations and total relative standard deviations of the three pollutant concentrations within the full spatial extent of the model domain. Seasonal variations are not clearly pronounced for the absolute spatial standard deviations. In contrast, a strong mirror-inverted like progression of the relative standard deviations for NO<sub>2</sub> and for ozone can be observed. This evolution of the relative standard deviations is in turn mainly caused by the corresponding annual pattern of absolute concentration variations: While NO<sub>2</sub> concentrations tend to be higher in winter time and lower in summer time, their relative standard deviation is lower in winter and higher in summer. For ozone it is the opposite, with a higher relative standard deviation in winter, and a lower relative standard deviation in summer. With regards to the spatial representativeness, it is finally driven by the nature of its definition, that a concentration field with a high relative standard deviation (i.e., with steeper concentration gradients) will have more clearly separated (i.e., less extended) areas of spatial representativeness. In turn, a concentration field with a lower relative standard deviation (i.e., with shallow

spatial gradients) will have less clearly separated areas of spatial representativeness, which consequently have larger extents.

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). Again, this finding corresponds with our general expectations about the relations between spatial representativeness and station types.

## **7 Summary and 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, however, does not 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 conceptualisations is not necessarily simply one-to-one. Rather, it should 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 summarised in a recent JRC technical report (Kracht et al. 2017a) that can be used for planning further developments of this method. These proposals for further developments do specifically include suggestions for:

- 1 possible variations of the underlying type of the variogram (directional variogram, relative variograms)
- 2 modifications of the variogram model functions
- 3 the criteria deployed for defining the limits of the spatial representativeness area
- 4 the numerical procedures
- 5 the pre-treatment and selection of data-points.

In this present study we have applied the concept of a spatial representativeness distance ( $dist_{SR}$ ), which in turn implies the assumption of a radially symmetric area of spatial representativeness. This corresponded to the use of an omni-directional variogram.

However, this approach is likely oversimplified and more detailed information (i.e., about the anisotropy of the variogram) could be extracted from the data (point 1 of the above list). In future developments it would thus be recommendable to extend the evaluation by applying directional variograms. Example given, this could be done by building directional variograms for the  $0^{\circ}$ – $90^{\circ}$ ,  $90^{\circ}$ – $180^{\circ}$ ,  $180^{\circ}$ – $270^{\circ}$ , and  $270^{\circ}$ – $360^{\circ}$  angular classes. A finer classification, although possibly desirable, might on the other hand be limited by the availability of data-points

With reference to the point 2, it should be noted that in this present study we have focused on the use of a spherical variogram model. However, it is not intrinsically obvious that the spherical model is the optimal choice and it could be worth evaluating different variogram models as an alternative. A more detailed discussion on the use of alternative variogram models can be found in Kracht et al. (2017a), which also demonstrates an inversion of the exponential and of the Gaussian variogram model (mathematical inversion in order to extract information about spatial representativeness).

## References

- Bivand, R.S., Keitt, T. and Rowlingson, B. (2016) *rgdal: Bindings for the Geospatial Data Abstraction Library*, R Package version 1, pp.1–9 [online] <https://CRAN.R-project.org/package=rgdal>.
- Bivand, R.S., Pebesma, E.J. and Gomez-Rubio, V. (2013) *Applied Spatial Data Analysis with R*, 2nd ed., Springer, NY [online] <http://www.asdar-book.org/>.
- Directive 2004/107/EC of the European Parliament and of the Council of 15 December 2004 relating to Arsenic, Cadmium, Mercury, Nickel and Polycyclic Aromatic Hydrocarbons in Ambient Air, Official Journal of the European Union L 152/1 of 11.6.2008.
- Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on Ambient Air Quality and Cleaner Air for Europe, Official Journal of the European Union L 152/1 of 11.6.2008.
- Dowle, M., Srinivasan, A., Short, T. and Lianoglou, S. with contributions from Saporta, R. and Antonyan, E. (2015) *Data.table: Extension of Data.frame*, R package version 1.9.6 [online] <https://CRAN.R-project.org/package=data.table> (accessed 9 October 2017).
- Hijmans, R.S. (2015) *Raster: Geographic Data Analysis and Modeling*, R package version 2.5-2 [online] <https://CRAN.R-project.org/package=raster> (accessed 9 October 2017).
- 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*, JRC Technical Reports, 267p, 102775, EUR 28135 EN, ISBN 978-92-79-62295-3 (PDF), DOI: 10.2790/479282.
- Kracht, O., Solazzo, E. and Gerboles, M. (2017a) *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*, RC Technical Reports, 135p, 107479, EUR 28726 EN, ISBN 978-92-79-71654-6 (PDF), DOI: 10.2760/64951.
- Kracht, O., Santiago, J.L., Martin, F., Piersanti, A., Cremona, G., Righini, G., Vitali, L., Delaney, K., Basu, B., Ghosh, B., Spangl, W., Brendle, C., Latikka, J., Kousa, A., Pärjälä, E., Meretoja, M., Malherbe, L., Letinois, L., Beauchamp, M., Lenartz, F., Hutsemekers, V., Nguyen, L., Hoogerbrugge, R., Eneroth, K., Silvergren, S., Hooyberghs, H., Viaene, P., Maiheu, B., Janssen, S., Roet, D. and Gerboles, M. (2017b) *Spatial Representativeness of Air Quality Monitoring Sites: Outcomes of the FAIRMODE/AQUILA Intercomparison Exercise*, JRC Technical Reports 108791, 402p, EUR 28987 EN, ISBN: 978-92-79-77218-4, DOI: 10.2760/60611.
- Lefebvre, W., Van Poppel, M., Maiheu, B., Janssen, S., and Dons, E. (2013) ‘Evaluation of the RIO-IFDM-street canyon model chain’, *Atmospheric Environment*, Vol. 77, pp.325–337.



- Pebesma, E.J. and Bivand, R.S. (2005) 'Classes and methods for spatial data in R', *R News*, Vol. 5, No. 2, [online] <http://cran.r-project.org/doc/Rnews/> (accessed 9 October 2017).
- Pebesma, E.J., (2004) 'Multivariable geostatistics in S: the gstat package', *Computers and Geosciences*, Vol. 30, No. 7, pp.683–691.
- R Core Team (2016) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria [online] <http://www.R-project.org/> (accessed 9 October 2017).
- Zeileis, A. and Grothendieck, G. (2005) 'zoo: S3 infrastructure for regular and irregular time series', *Journal of Statistical Software*, Vol. 14, No. 6, pp.1–27, DOI: 10.18637/jss.v014.i06.