Assessing the spatial representativeness of air quality sampling points – Literature Review

Service Request 5 under Framework Contract **ENV.C.3/FRA/2017/0012**

Specific Contract: **07.0203/2018/793545/SFRA/ENV.C.3**

Report for European Commission - DG Environment

Ares (2018) 4920320

11492

Customer:

Report for European Commission - DG Environment

Customer reference:
Ares (2018) 4920320

Confidentiality, copyright & reproduction:

Ricardo reference:
Ref: ED 11492 – Final

Contact:

Joanne Green, Ricardo.

t: +44 (0)1235 75 3450
e: Jo.Green@ricardo.com

Author:
Bino Maiheu
Stijn Janssen

Approved By:
Beth Conlan

Date:
18 December 2019

Table of contents

1 Introduction

1.1 Challenge and aims of this project

Monitoring networks to measure air pollution are at the core of air quality policy. To assess air quality, report compliance, estimate population and ecosystem exposure and to validate and calibrate air quality models, it is crucial that a monitoring network is configured so that it is capable of providing a representative assessment.

The spatial representativeness (SR) of monitoring stations is at the basis of configuring monitoring networks. The evaluation of the SR of monitoring stations is essential where monitoring networks are used to estimate the number of people and extent of ecosystems exposed to the air pollution measured by a monitoring station and therefore to estimate the health and ecosystem impact of air pollution. It is also implicit for all other applications of monitoring sites in the Ambient Air Quality Directive 2008/50/EU[1] (AAQD). The AAQD and its implementing provisions leave room for interpretation of SR and methods for its evaluation and hence SR of air quality monitoring sites remains an issue for which at present there appears to be no standardised approach.

The SR of an air quality monitoring site can be broadly viewed as a spatial extent over which an air quality concentration can be considered similar to its observation at the site. Using such a description as a preliminary definition for the concept of SR immediately poses several fundamental questions. For example, how to define or interpret "*spatial extent*", what does "*similar*" mean and in what way does this concept differ depending on the specific assessment need? It was realised, as long ago as the 1990's, that it was very difficult to derive a fully objective definition for "representativeness", because data are representative for a specific application (Wieringa, 1996). Different studies highlight that the representativeness of stations (also their classification) varies with the pollutant species considered, so what might be a representative area (or classification) for the station for one pollutant may not be the same for another pollutant at the same station. This variability is inherent to the nature of sources and activities causing pollution and dispersion of pollutants in the atmosphere as a result of meteorological, physical and chemical conditions.

Different studies and reviews have expressed the need for harmonisation in the methods for evaluating SR as the estimates of this concept can vary enormously between methodologies. The initial report by Spangl et al. (2007) clearly pointed out challenges related to SR assessment and put forward a first practical assessment methodology. Since then considerable effort has been put into providing guidance and testing methods for establishing spatial representativeness of monitoring sites in the FAIRMODE[2], AQUILA[3] and CAMS[4] communities. An overview of the past efforts is given in Appendix 1 and this project is a continuation of those efforts. In the most recent initiative, an intercomparison exercise (IE) organised within FAIRMODE and AQUILA (Oliver Kracht et al., 2017) involved several research groups who compared their estimates of the SR of three distinct monitoring sites in the city of Antwerp, Belgium. Figure 1 below, taken from this IE, illustrates the variation in SR estimates. Despite these efforts, final guidance relating to assessment and reporting under the EU AAQD is not available. Currently, in the absence of guidance on SR to accompany the implied SR needs of the EU ambient air quality legislation there is no common view on how to calculate SR of monitoring stations.

As a broad range of different methods to assess SR exist across Europe, estimates of population exposure derived from calculations of SR of monitoring stations cannot be readily

[1] https://eur-lex.europa.eu/eli/dir/2008/50/oj and the amendment https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32015L1480
[2] https://fairmode.jrc.ec.europa.eu/
[3] https://ec.europa.eu/jrc/en/aquila
[4] https://atmosphere.copernicus.eu/

compared across Member States and sometimes not even across air quality zones in the same Member State.

In this project, the overall goal is to provide an overview of the existing challenges associated with the determination of spatial representativeness, identify ambiguities and provide guidance and recommendations for its assessment in the context of the AAQD.

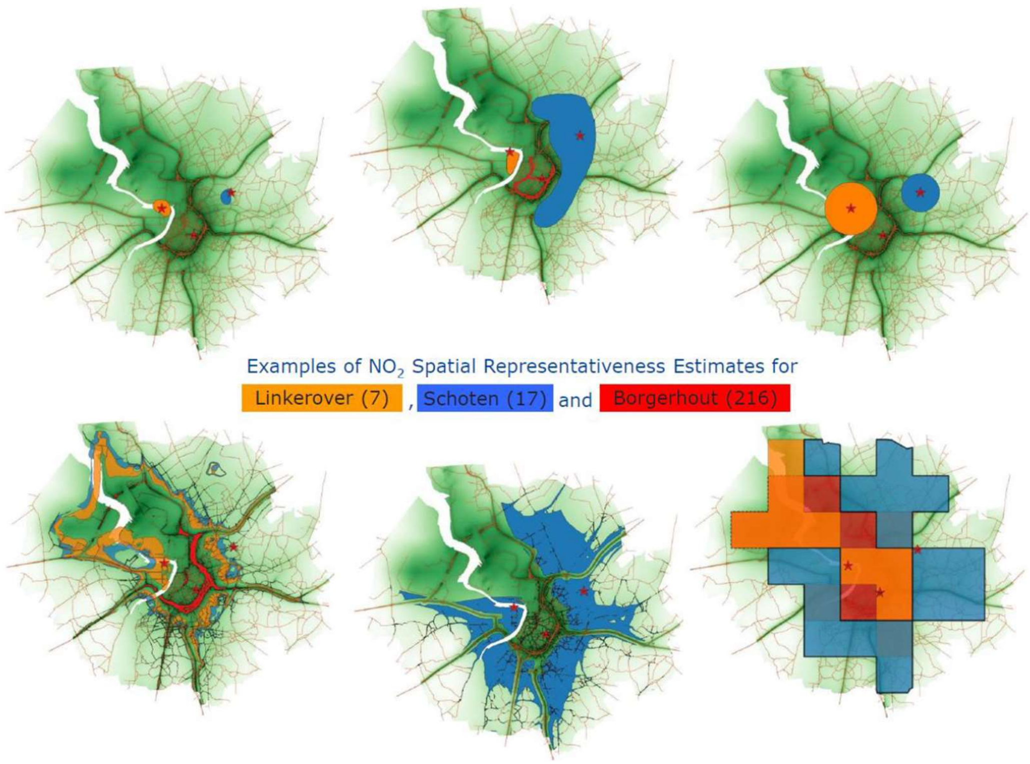

Figure 1: Results of six different modelling teams when asked to calculate the SR of three stations (Linkeroever, Schoten, Borgerhout) in Antwerp, Belgium. Source: (Oliver Kracht et al., 2017)

1.2 The context of the Ambient Air Quality Directive

Despite the current lack of a common view or established guidance to assess SR of air quality monitoring sites, the importance of such guidance cannot be underestimated as it could provide important improvements to the current Ambient Air Quality Directive and its subsequent implementing decision 2011/850/EU[5] and the IPR[6] guidance[7].

Overall reporting guidance for the Member States is provided in the IPR guidance part 1, however for the "*Evaluation of representativeness*", this guidance document states that *"There is no definition of the spatial representativeness of monitoring stations in the AQ legislation yet. FAIRMODE is in the process of developing tools for its quantitative assessment*" and provides a link to the related FAIRMODE cross cutting activity[8]. Currently, the IPR guidance document also refers to the 2007 study by the Austrian UBA (Spangl et al., 2007) and the CAMS station classification paper by Joly and Peuch (2012), and indicates that final recommendations will be introduced into the IPR guidance following further work by the

[5] https://eur-lex.europa.eu/eli/dec_impl/2011/850/oj
[6] Implementing Provisions for Reporting
[7] See https://aqportal.discomap.eea.europa.eu/toolbox-for-e-reporting/guidance-on-the-commision-ipr-decision/ and the latest IPR guidance part 1 v2.0.1 https://www.eionet.europa.eu/aqportal/doc/IPR%20guidance_2.0.1_final.pdf which is the "Member States' and European Commission's Common Understanding of the Commission Implementing Decision 2011/850/EU (or IPR guidance part 1 v2.0.1)(pdf)".
[8] http://fairmode.jrc.ec.europa.eu/cca.html

AQUILA and FAIRMODE communities. None of these documents deals comprehensively with methods to calculate spatial representativeness in relation to reporting requirements in the AAQD. Spatial representativeness related reporting opportunities for Member States within e-Reporting dataflows include:

- Dataflow B under (2011/850/EU, Article 6) requests information on zones and agglomerations.
- Dataflow D under (2011/850/EU, Articles 8 and 9), requests information on assessment methods for both fixed and indicative measurements, requests an evaluation of representativeness (see 2011/850/EU, ANNEX II - (D)) and a classification of the stations/areas and network for both the local dispersion situation, the regional dispersion situation, the network type (local, urban, regional, national) as well as a station and an area classification.
- Dataflow G under (2011/850/EU, Article 12) requests information on the attainment of environmental objectives, in particular the area of exceedance and the number of people exposed.

It is important to note that in many cases reporting information is conditional and only mandatory when or if available.

In the absence of further guidance, there is room for interpretation of the provisions of the AAQD with regard to the assessment needs listed above. This has been discussed recently, and an overview of related instances of such lack of clarity is listed in Table 6 of (Nagl et al., 2019). The most relevant of such instances where there is a need for further guidance in the context of spatial representativeness are listed below.

With respect to the **macroscale siting** of sampling points directed at the protection of human health (AAQD[9], Annex III B):

- The requirement is to sample the "*areas within zones and agglomerations where the highest concentrations occur to which the population is likely to be directly or indirectly exposed for a period which is significant in relation to the averaging period of the limit value(s)*" as well as the request to provide data on the "*levels in other areas within the zones and agglomerations which are representative of the exposure of the general population*".
    - **Guidance need 1**: As (Nagl et al., 2019) states, there is no definition given for "*the exposure of the general population*", furthermore it is unclear what "*significant*" means in this context.
    - **Guidance need 2**: It is also unclear what is meant by "*representative*" as this could imply knowledge on activity patterns. "*Exposure*" to air quality can be measured in a dynamic or static way, as discussed in (Maiheu et al., 2017). A dynamic exposure assessment typically tracks an individual's movement pattern throughout the day and accumulates the concentration values they are exposed to during that time at different locations. In a static or address based exposure assessment, a population map is typically compared with a pollutant concentration assessment to derive "*exposure*", with population density acting as a proxy for an individual's true exposure pattern. The latter method is more commonly used in health impact assessments. To summarise, more guidance is required on what "representative" means to account for an individual's actual exposure to air pollution on a dynamic basis.
    - **Guidance need 3**: An additional complication arises in the presence of street canyons, where the concentration pattern is discontinuous. For street canyons,

9 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008L0050&from=EN

it is possible that concentrations at building front facades are significantly higher than concentrations at the rear of the buildings. A question then arises in terms of what concentration values to assign to the population when calculating exposure.

- The requirement that sampling points shall be *"be sited in such a way that the air sampled is representative of air quality for a street segment no less than 100 m length at traffic-orientated sites and at least 250 m × 250 m at industrial sites"*
  - **Guidance need 4**: Implies that these fixed length scales offer an acceptable level of spatial variability, however the amount of spatial variability is impacted by location/source characteristics and hence is not a fixed quantity.
- In paragraph C under Annex III – B.1, when siting for urban background locations, it is required that sampling points at such locations should be representative for several square kilometres, with their levels not dominated by a single source.
  - **Guidance need 5**: Again, the level of spatial variability implied by a representative area of "*several square kilometres*" is not a fixed quantity. In addition, there is a lack of guidance as to what "*dominated*" means, these provisions are not precisely quantified.

With respect to the **microscale siting** of sampling points (AAQD Annex III, C), it is required that "*the inlet probe shall not be positioned in the immediate vicinity of sources in order to avoid the direct intake of emissions unmixed with ambient air*", it should be *"some metres away from buildings, balconies, trees and other obstacles and at least 0.5 m from the nearest building in the case of sampling points representing air quality at the building line"* and "*for all pollutants, traffic-orientated sampling probes shall be at least 25 m from the edge of major junctions and no more than 10 m from the kerbside*."

- **Guidance need 6**: "*immediate vicinity*" is not strictly defined, and the requirement to position the inlet "*some metres away from*" various obstacles is not quantified.
- **Guidance need 7**: There is a lack of knowledge about whether the requirement that traffic stations are no more than 10 m from the kerbside is always compatible with the requirement for them to be representative for a street segment no less than 100 m as indicated above.

With respect to the criteria for determining the **minimum numbers of sampling points** for fixed measurements of concentrations (AAQD Annex V): section A.1. provides a minimum number of sampling points as a function of population total in the agglomeration or zone and depending on whether the maximum concentrations exceed the upper assessment threshold (UAT) or whether maximum concentrations are between upper and lower assessment thresholds.

- **Guidance need 8**: There are no guidelines on how and where to determine the maximum pollution level in the zone. The location of the maximum concentration may not be in the most obvious area. For instance, the high density CurieuzeNeuzen[10] monitoring campaign in Flanders indicated that the maximum concentration for the whole of Flanders was found in a rather unexpected location at a busy intersection on the N715 road near Houthalen-Helchteren, in a rather rural area without exceedances reported by the regular network.
- **Guidance need 9**: It is unclear whether the minimum number of stations determined from the AAQD requirements is always sufficient to fulfil requirements for

[10] https://curieuzeneuzen.be/

representativeness and model validation. For example, is a single traffic station enough to report for a medium sized town or agglomeration?

Many of the remaining instances of lack of clarity listed in Table 6 of (Nagl et al., 2019), refer to a **lack of quantification** in the IPR guidelines (DG-Environment, 2018), e.g.:

- In the recommendations for assessing the local dispersion situation (Table 9 in the IPR guidance document) a street canyon is defined as continuous/compact buildings along both sides of the street over more than 100 m with an average ratio of height of buildings to width of street > 0.5. Besides there being no provision for canyons with ratio < 0.5 as indicated in (Nagl et al., 2019), the concept of a street canyon also poses considerable (model-dependent) challenges and guidance is required to help determine when to account for canyons in a modelling assessment.
- In Table 12 in the IPR guidance document, the station classification criteria for traffic, industrial, background stations are formulated as a function of “proximity to a source”, however there is no quantification as to what distances apply or perhaps more importantly, what distances can be considered relevant.

Given the points raised above and the fact that it is not fully clear how individual Member States interpret or implement these, there is clear room for improving guidance. It should however also be noted that some of the flexibility allowed by the Directive enables Member States to have the discretion and the possibility to implement the requirements that is most appropriate for their specific circumstances. In Task 3 of this project, an overview of current practice in the Member States in respect of siting criteria is being investigated.

1.3 Spatial representativeness for different applications

The FAIRMODE/AQUILA intercomparison exercise has been a significant step forward in bringing clarity to the difficulties of interpreting and calculating SR. One of the key outcomes was expressed as the need for a paradigm shift in the definition of spatial representativeness. The idea of spatial representativeness as being a single property of a monitoring site should be abandoned and instead the aim should be to distinguish between SR definitions, methods, objectives and purposes for performing an SR assessment (Oliver Kracht et al., 2017).

Within this Service Contract, this line of thought was adopted, and the ***assessment needs*** were clearly described in the statement of work as follows:

a) Estimate of the surface area where the level was above the environmental objective,

b) Estimate of the length of road where the level was above the environmental objective,

c) Estimate of the total resident population in the exceedance area,

d) Facilitate the configuration of a representative monitoring network,

e) Identify sampling points that are suitable for model calibration and validation,

f) Determine the spatial variability within the “area of representativeness”.

To account for the suggested paradigm shift and the statement of work in this Service Contract, the assessment needs expressed above will provide the review framework for the methodologies that will be discussed in this work. As will be made clear in the next chapters, a stratification according to the specific assessment needs will help bring clarity in the definitions and proposed methodologies in order to improve the guidance for the implementation of the AAQD.

1.4 Important interrelated concepts

Several important interrelated concepts are introduced in this section as these play a major role in understanding the differences between the methodologies discussed. These methodologies will be reviewed below in Chapter 3 and are essential to find a way to harmonise the understanding of spatial representativeness and to address the highlighted ambiguities above. Different ***assessment needs*** will require different approaches in terms of spatial representativeness assessment.

1.4.1 The spatial representativeness area

The concept of a spatial representativeness area around a monitoring station constitutes an explicitly delineated geographical area for which the observed air quality metric at the monitoring station can be considered representative. This consideration therefore involves a similarity criterion (i.e. how do we establish whether or not a particular location is similar to the observed metric) as well as a tolerance level (i.e. how much deviation do we allow from the observed metric).

When looking closer at the ***assessment needs***, one can note that the first 3 (***a. estimate of surface area***, ***b. estimate of road length*** and ***c. estimate of population exposure***) require the estimation of a geographical area. Interestingly, one of the outcomes of the FAIRMODE/AQUILA IE (Oliver Kracht et al., 2017) as presented by S. Janssen at the FAIRMODE plenary meeting in February 2019, was the observation that the concept of a "***spatial representativeness area***" can be seen as a first step forward in the common understanding of station representativeness and the delineation of a geographical area was as such also the key requirement for the intercomparison exercise. Estimation of a surface area or the length of road clearly implies a geographical extent or area, but also exposure can be easily assessed once the SR area of a monitoring station is known when an overlay with a population density map is made. Note that EU wide population maps are freely available[11].

In the context of the other assessment needs, the concept of a geographical area might equally help to improve the general understanding. For model validation (assessment need ***e***), the SR area (or the characteristic length scale of the area) of a monitoring station may help to define whether the station can be used in the validation exercise by comparing the size of the SR area with the model resolution. For example, a station situated near a busy traffic lane or industrial plant may not be representative for the wider area and therefore have a spatial representativeness area of e.g. 100x100 m^2. Comparing such stations to e.g. chemical transport models with a model resolution of say 4 x 4 km^2 is therefore not ideal as it cannot be expected that the air quality model is able to resolve features of or gradients in the concentration field down to this scale.

Even for network design (assessment need ***d***) the concept of SR area might be a useful starting point. A monitoring network in combination with its (total) SR area will clearly point out the spatial coverage of the network as a whole, indicate where blind spots are present and where large overlaps in SR area occur between neighbouring stations.

Next to an explicit geographical definition of the spatial representativeness area, a more implicit or qualitative definition may be adopted as well to the concept of a spatial representativeness area. In the AAQD, a classification of stations is required in relation to predominant emission sources in accordance with the macro scale siting criteria. Guidance to this is provided in (DG-Environment, 2018), Table 12, distinguishing between Industrial, Traffic and Background stations. In addition, several criteria are listed for area classification (Urban, Suburban and Rural) as well.

---

[11] https://ghsl.jrc.ec.europa.eu/ghs_pop.php

Station and area classification is a way to assign a monitoring station to one of the categories mentioned above, based on the characteristics and properties of a dominant source/sector (station classification) or a particular environment or geography (area classification). Though classification may seem qualitative in nature, quantitative approaches can be applied to automate derivation of a station class and the macroscale siting criteria relate station classification to an area of representativeness. A station classified as an “urban traffic” station will likely be representative for the geographical area close to the location of the roads in the city. Similarly, a rural background station may be representative for remote locations far away from any emission sources. While an exact delineation of such a qualitative area is obviously not provided via a station classification by itself, it is easy to see how an explicit geographical area would improve the interpretation of differences between the properties of stations assigned to the same category. For example, two stations classified as “industrial” may have a very different area of representativeness when the underlying industrial emissions are very different in nature. An area of representativeness based upon a spatio-temporal assessment of the concentrations which accurately reflect these emissions can aid in interpreting such differences.

The concept of a geographical area of representativeness implies that there is some methodology to represent the **spatio-temporal variability** of the concentrations, or the environmental objective being assessed[12]. Concentration values in ambient air can vary significantly in space and time, depending on the nature of the activity and distance to the emission sources, the diurnal cycle night/day, rush-hour, season, or the environmental dispersion conditions (terrain roughness, mountains/valleys, urban features such as street canyons, presence of buildings, vegetation elements, screens etc.). When assessing spatial representativeness of monitoring stations, methods used to reflect or capture the spatio-temporal variability will therefore have to be fit-for-purpose, meaning that they should be able to represent the required features and gradients in the concentration field as accurately as required to meet quality objectives.

1.4.2 Similarity criterion

Another important concept to introduce here is that of the **similarity criterion**. Despite the broad differences in methodologies that exist today to assess SR, both qualitatively or quantitatively, one thing which all methods do have in common is the fact that they try to quantify the similarity between air quality metrics at geographically distinct locations. Classification methods will assign stations to the same categories if their properties and attributes (e.g. time-series characteristics) are similar, if locations can be thought to belong to the same spatial representativeness area of a particular monitoring site, and if some properties and characteristics (e.g. concentration levels, relationship/proximity to a dominant source, area type) are similar enough. A similarity criterion expresses this similarity mathematically via a similarity criterion.

Such a criterion is used to delineate a geographical area in which the air quality metric considered deviates only by a given **tolerance level** from the observed value. Despite this common and fairly simple concept, there is a large variety in the interpretation of “similarity”.

Different approaches can be adopted to define a similarity criterion, for example:

- As an absolute deviation, i.e. the air quality metric is not allowed to differ by more than +/- xx µg/m^3 from the value at the monitoring site.
- As a relative criterion, e.g. the air quality metric is not allowed to differ more by more than +/- yy % from the value at the monitoring site.

[12] Not all environmental objectives directly refer to concentration values in µg/m^3, some use related metrics based on percentile values or number of exceedance days, AOT40 etc.

- As a combination of both absolute and relative criteria (as proposed by the IE, §10.4), e.g. as the max of [± xx µg/m^3] and [± yy % of the observed concentration].
- A given maximum range in the similarity criterion, i.e. next to fulfilling one of the criteria above, the distance to the observation location should be less than a given value (e.g. 100 km). Such an additional criterion is considered to account for transport of pollutants in the atmosphere, ensuring that the concentrations within the area of representativeness can physically (via transport in the atmosphere) originate from the same emission sources. One can therefore also ask whether spatial contiguousness should or should not be a prerequisite in the similarity criterion.
- Next to purely spatial considerations, temporal information may be included as this may also be representative for different underlying emission patterns. If one purely defines an SR area based on the spatial pattern of the air quality metric such as an annual averaged NO_2 concentration, there may be no distinction between concentrations arising from e.g. road traffic or industrial sources if both happen to give rise to similar values for the annual averages. The question arises therefore whether or not it is desirable to break up the SR area to account for this difference in underlying emission pattern. Metrics reflecting differences in the timeseries such as the variance of the hourly values or the temporal correlation coefficient between the values observed at the monitoring location and within the SR area may account for this.
- Finally, the approach for defining the similarity criterion could take into account the observation uncertainty, similar to the way in which the observation uncertainty is included in the FAIRMODE DELTA model benchmarking tool. For example, the tolerance level could be set to be larger than the observation uncertainty.

The different possible ways in which the similarity criterion can be defined will be a subject of sensitivity studies later in this project. An important realisation from the FAIRMODE/AQUILA IE is the concept of a **primary similarity criterion** which relates directly to the metric under consideration. For instance, when using an annual averaged PM_{10} concentration at a particular monitoring site to test compliance with the limit value, the similarity criterion should primarily relate to the level of acceptable deviations from the annual averaged PM_{10} value observed at that monitoring site location rather than any other characteristic of the observations at the location. Additional criteria may be added, such as requirements for temporal correlation of hourly values, but the primary criterion should be defined upon the metric required for compliance checking itself. This primary similarity criterion should allow the formulation of transparent and above all pragmatic definitions. Whether or not it would be useful to consider additional criteria to fully define SR for the given purpose/assessment need depends on technical feasibility and will be explored in the sensitivity studies later in this project.

1.4.3 The tolerance level

When a similarity criterion is established or chosen, a particular threshold or **tolerance level** value has to be adopted. This implies that within the area of spatial representativeness, one allows a certain spatio-temporal variability to remain, or in other words, the concentration metrics can differ from the observed values at the monitoring station by a given amount. This tolerance level will significantly influence the delineation of the SR area. A higher tolerance level will make the SR area lager, whereas a stricter tolerance level, will make the SR area smaller.

Conceptually, classification methods to some extent also can be thought to employ a certain tolerance level, as the attributes and properties used to decide in which class to put a monitoring station will not be identical for all stations belonging to the same class but differ by some amount.

In the IE, a tolerance level was not agreed beforehand, which explains the divergence in the final outcome of the exercise.

1.5 Outline of this report

Following this introduction, we present a literature review of various methods assessing spatial representativeness in Chapter 2. In Chapter 3, we review the specific methodological requirements for specific purposes of spatial representativeness and refer to the literature review to discuss applicability of the methods and the room for further clarification related to spatial representativeness in the current IPR guidance for different applications. In Chapter 4 the methods are further structured according to a tiered approach per assessment need as a way to provide guidance in order to address some of the ambiguities and acting as a framework for further recommendations. Finally, conclusions are presented in Chapter 5, followed by references and appendices containing more detailed and technical information, including an overview of past efforts regarding harmonizing and formulating recommendations for the assessment of spatial representativeness.

2 Literature review

2.1 Literature search

This section presents a review of recent literature dealing with spatial representativeness of monitoring stations. We briefly discuss a number of attributes for each methodology to enable us to capture the essence and evaluate them with regards to the framework proposed in the FAIRMODE/AQUILA intercomparison exercise (Oliver Kracht et al., 2017).

These attributes are:

- The main **outcome** of the methodology: a descriptive assessment of spatial representativeness, a geographical SR area or a classification
- The **method** to establish the spatio-temporal variability
- The **spatial scale** of the assessment
- The method of assessment of the **temporal variability** and the associated scale
- The way in which **similarity** was established
- **Pollutants** the methodology was applied to
- The overall **goal** of the study and how it refers to the assessment needs listed above
- And some editorial **remarks**

Table 1 contains all underlying methodologies which were used in the FAIRMODE/AQUILA intercomparison exercise, however, in more recent years several further interesting papers have been published which offer additional insights.

The main distinction in the technical outcome is whether the method allows to delineate an explicit geographical area of representativeness from which the surface area, total length of road or total resident population in exceedance of the environmental objective can be derived. This distinction was clearly made already in (Spangl et al., 2007) where:

- Station **classification** aims to categorise monitoring locations into groups with common characteristics, separating them from other groups, with other common features.
- Assessment of **station representativeness** aims to delineate areas of the concentration field with similar characteristics.

In the columns for spatial and temporal variability, we typically list the spatial and temporal resolution of the underlying methodology used to establish the spatio-temporal variability.

Table 1 : Overview table with references and attributes per methodology

Reference	Outcome	Spatio-temporal variability method	Spatial scale	Temporal variability	Similarity	Pollutants	Goal of SR assessment	Notes
(Cosemans et al., 1997)	Descriptive assessment, monitoring network optimisation	IFDM Gaussian modelling using detailed industrial emissions	Gridded to 500 x 500 m	Modelling for 11 different selected meteo years, aggregated to median and maximum, 98 percentile, and daily average value	Qualitative, concentration field is used	SO_2, NO_x	Determine the optimal siting of air quality monitoring stations around five oil refineries in Antwerp, mainly focussed on identifying extreme concentration values	
(Blanchard et al., 1999)	SR area	Gaussian puff model	50x50 grid points, spaced 1 km	Twelve hour pseudo mean concentrations	20% in concentration levels	PM_{10}	Determination of scales of transport and SR during the 1995 monitoring study in San Joaquin valley, CA.	
(Scaperdas and Colvile, 1999)	Descriptive assessment of SR	CFD model	Model at 5 m resolution	Distinct meteo conditions (wind directions)	Qualitative	passive tracer i.e. a generic pollutant which does not react chemically.	Determine the area of influence for a monitor situated at a cross road of two street canyons	
(Vardoulakis et al., 2011a, 2005)	classification and SR areas	Dedicated sampling campaign using passive samplers; complemented with dispersion modelling (STREET-SRI, OSPM, AEOLIUS)	Street canyon level and in-situ monitoring	Short sampling campaign	Assessed using statistical indicators	NO_2, O_3	Assessment of site representativeness, optimisation of monitoring location	

Reference	Outcome	Spatio-temporal variability method	Spatial scale	Temporal variability	Similarity	Pollutants	Goal of SR assessment	Notes
(Spangl et al., 2007)	classification and SR areas	GIS proxy data	Depending on input data used, typically down to 100 - 250 m (CORINE).	Time aggregations as required by the AAQD, no temporal variability	NO_2, PM_{10} annual mean +/- 5 µg/m^3 PM_{10} annual 90.4 percentile of day average +/- 8 µg/m^3 O_3 annual 93.2 percentile of max8h +/- 9 µg/m^3	NO_2, PM_{10}, O_3	Classification and SR areas	
(Ott et al., 2008)	Descriptive assessment of SR	Passive samplers and Kriging using spherical semi-variogram	In-situ	3-week average	Pearson correlation and coefficient of divergence (COD) combined with ANOVA test to test for equal means	PM_{10}, $PM_{2.5}$	Capture spatial variability of PM in Iowa city Deploy effective sampling strategy	
(Lozano et al., 2009)	Descriptive assessment of SR	In situ monitoring Inverse distance weighted interpolation (IDW)	In-situ	Two sampling campaigns	n/a	NO_2, O_3	Determine optimal locations for NO_2 and O_3 monitors.	IDW has been shown to be unsuited for urban applications. IDW could be considered a tier 1 method
(Nguyen et al., 2009)	classification	Fixed distance (as applied in IE)	street location: 100m, urban background 1 km	Hourly timeseries	PCA on hourly timeseries, diurnal variation and pollution roses. Classification by analysis of PC loading plot.	NO, NO_2, CO, PM, O_3, NH_3 and SO_2	Classification as input for network optimisation	
(Karabelas and Sarigiannis, 2008; Sarigiannis and Saisana, 2007)	n/a	Satellite imagery and CORINE land cover and Kriging	1 km	n/a	Linear regression	NO_2, SO_2, O_3, CO, $PM_{2.5}$, PM_1 and PM_{10} combined	Optimisation of air quality monitoring network	Include CAPEX/OPEX into cost function for optimal siting

Reference	Outcome	Spatio-temporal variability method	Spatial scale	Temporal variability	Similarity	Pollutants	Goal of SR assessment	Notes
(Wu et al., 2010)	n/a	Ordinary Kriging with different covariance models	Unclear, Ordinary Kriging based on O3 network in France	Hourly	Network reduction via RMSE of Kriging predictions of a "subnetwork" w.r.t. the observations	O_3	Examine how well a subset of O_3 monitors over France can represent the concentration field.	
(Henne et al., 2010)	classification	Emission proxy data and Lagrangian Particle Dispersion Modelling	7x7 km meteo fields,	3-hourly	Based upon calculation of residence time in Lagrangian dispersion model: catchment area	O_3, NO_2	Identify factors determining station representativeness	Included surface deposition considerations for O_3 Discusses suitability of Lagrangian transport models for complex terrain.
(Joly and Peuch, 2012)	classification	n/a	n/a	Hourly timeseries	Linear Discriminant Analysis on timeseries derived metrics and percentile classification	O_3, NO_2, NO, PM_{10}, SO_2	Identification of stations for calibration/validation in regional scale models	Mainly aimed at separating rural from urban stations
(Janssen et al., 2012)	SR area	Land use regression modelling using CORINE and residual Kriging	Gridded results at 4x4 km^2	Annual maps	20% deviation	NO_2	Determine area of representativeness	Use relationship between NO_2 and LC derived parameter to define SR area
(Vincent and Stedman, 2013) (Ricardo-AEA)	classification	Expert opinion and source apportionment plots generated using the PCM model	1x1 km	Annual source apportionment for 2011	n/a	NO_x, SO_2, PM_{10}, $PM_{2.5}$, O_3	Classifying UK network stations	
(Righini et al., 2014)	SR area	Emission "variability maps"	Gridded results at 4x4 km^2	n/a	Natural breaks ("jenks") classification	$PM_{2.5}$, PAH, As, PM_{10}	Detect spatial representativeness of selected Italian monitoring stations	Emissions are used as a surrogate of concentrations, only works for primary pollutants, not advised for urban areas

Reference	Outcome	Spatio-temporal variability method	Spatial scale	Temporal variability	Similarity	Pollutants	Goal of SR assessment	Notes
(Martin et al., 2014)	SR area	WRF-CHIMERE CTM, corrected via residual Kriging with spherical variogram	Gridded results at 9x9 km^2	Time aggregations as required by the AAQD, no temporal variability	Depending on the pollutant, factor of 1.2 (20%) or factor of 2 for lowest concentration values	NO_2, SO_2, O_3 and PM_{10}	Quantifying SR areas of rural background stations; evaluate station redundancy and network coverage	Similarity criterion depends on concentration
(Piersanti et al., 2015)	SR area	AMS-MINNI CTM	Gridded results at 4x4 km^2	Hourly model timeseries	Concentration Similarity Function (CSF) using 20% deviation threshold	$PM_{2.5}$, O_3	Quantifying SR areas of rural background stations in Italy	
(Duyzer et al., 2015)	Descriptive assessment of SR	Urban dispersion model URBIS with separate calculation for canyons (CAR)	Gridded results at 10x10 m	Annual averaged	n/a	NO_2	Derive exposure curves, assess compliance	Authors state: the results of measurements at traffic stations are not useful to assess exposure of the general population. They often indicate only the exposure of a (sometimes very) small fraction of the population.
(Diegmann et al., 2015)	Descriptive assessment of SR	MISKAM CFD model	Gridded results at 1-5m resolution	Annual mean	Percentage deviation	NO_2, NO_x	Assess spatial representativeness of traffic-oriented sites	
(Barrero et al., 2015)	classification	n/a	n/a	Hourly PM_{10} measurements	k-means clustering in 4 groups based on time series properties	PM_{10}		
(Vitali et al., 2016)	SR area	Lagrangian dispersion model	100 m	Hourly	Concentration Similarity Function (CSF) using 20% threshold.	PM_{10}	Assess representativeness area of industrial station	
(Tapia et al., 2016)	classification	n/a	n/a	Hourly data from 1998 - 2012	4 classes are introduced based on characteristics of cumulative frequency distribution	O_3	Using metric based on Gini index to reflect the equality of the frequency distribution	Suited for O_3 as relies on absence of extreme outliers

Reference	Outcome	Spatio-temporal variability method	Spatial scale	Temporal variability	Similarity	Pollutants	Goal of SR assessment	Notes
(Soares et al., 2018)	classification and SR area	GEM-MACH CTM simulations	Gridded results at 2.5 km	Hourly modelled timeseries	Hierarchical clustering and degree of similarity defined by Euclidian distance and pearson correlation	NO_2, SO_2	Evaluating network redundancy	
(Gupta et al., 2018)		Use GIS proxy info (LUR)					Air quality monitoring design optimisation, however targeted improving exposure assessment in LURs	Still interesting as the methodology could apply
(Beauchamp et al., 2018; Bobbia et al., 2008)	SR area	Kriging with external drift model	Proxies of 150 m (traffic), 500 m for background; or using high resolution dispersion model	n/a	10 µg/m^3 and a statistical risk of 10 % that the concentration of any point within the SR area deviates by more than the tolerance	NO_2, PM_{10}	Taking estimation uncertainty into account in the SR area determination	Distinguish between area of similarity and area of exceedance; formulate probabilistic framework for area of exceedance (using Kriging) Discuss a way to deal with overlap
(Hao and Xie, 2018)	SR area	WRF-CALPUFF	Gridded results at 2 km	4 months of daily averages SO_2 and NO_2	Using Pearson correlation with cut-off of 0.7	SO_2, NO_2	Optimize the network size and site locations simultaneously for Shijiazhuang city, China	Using genetic algorithm in optimisation, to maximize coverage of the total SR area; and minimize the overlap
(Rivas et al., 2019; Santiago et al., 2013)	SR area	Weighted CFD - RANS modelling	Gridded results at 1 – 5 m resolution for 7.7 km × 5.4 km domain	Aggregations using meteo statistics	20% of annual averaged value	NO_2, NO_x	Assess spatial representativeness of each air quality monitoring station of the Pamplona network in Spain.	

Reference	Outcome	Spatio-temporal variability method	Spatial scale	Temporal variability	Similarity	Pollutants	Goal of SR assessment	Notes
(Rodriguez et al., 2019)	SR area	PMSS microscale model	Gridded results at 3 m resolution covering 10x12 km^2 of Paris	Hourly gridded model values for period of 10 days	Area in which temporal correlation is high and NRMSE low, thresholds by iterative analysis	NO_x and PM_{10}	Assess SR areas around five urban background and five traffic-oriented monitoring-sites of the AIRPARIF network during ten days in March 2016	Distinguish and compare homogeneity area (20%) deviation from their definition of similarity
(Li et al., 2019)	SR area	Combination of mobile and distributed stationary samplers	Results aggregated in 50x50m grid cells around the monitoring site	Aggregations of sampling campaign during winter and summer 2016 - 2017	Statistical distribution test (Wilcoxon rank-sum test) using p value of 5% as threshold for similarity	NO_2, UFP and PM_1	Assess spatial variability around monitoring stations in Pittsburgh, PA.	Interestingly are looking at number of restaurants to describe pollutant variation, though unsuccessfully

An older paper by (Levy and Hanna, 2011) provides an overview of studies in the US between 2000 and 2007 discussing particulate matter variability in the urban environment using monitors. Table 1 in that reference (reproduced here in Appendix 3) lists additional ways to define the concept of spatial heterogeneity, using correlation, land-use regression, factor analysis, comparisons across means and percentiles and ANOVA. These methods are consistent with those mentioned above.

2.2 Discussion

Here we provide a general discussion on the outcome of the literature review. We have structured it around the different attributes which were used to summarise the literature in Table 1 and discuss their relevance for the AAQD assessment needs given in Section 1.3. We discuss the methodology outcomes, the different ways of establishing spatial variability and accounting for temporal variability as well as establishing the similarity between locations.

In addition to the discussion on the different attributes used to summarise methods found in the literature, it is important to realise than any available methodology to assess SR should also be evaluated with respect to practical and pragmatic criteria related to available input data, fitness-for-purpose and applicability. Table 2 describes these criteria.

Table 2 – Evaluation criteria for SR assessment methodologies

Criterion	Description
Data	
Data availability	The input data needed for the methodology should in principle be readily available for every location (in particular urban areas) in a Member State, and across Europe if to be widely adopted.
Fitness-for-purpose	
Temporal resolution	A large temporal variability is present in ambient air concentration levels. Not all methods represent short term variability, e.g. some methods only produce annual averages. Temporal resolution should be sufficient to capture observed temporal variability of the air quality metric.
Spatial resolution	In urban environments, the spatial variability of some pollutant concentrations is very high, hence the level of spatial resolution of the method should be sufficient to capture observed spatial gradients of the air quality metric.
Spatial coverage	The methodology should enable production of results at e.g. a national or regional level depending on the area being assessed, and cover stations representative of all area types e.g. urban, rural within the area being assessed.
Applicability	
Technical result	Is the method a classification method (qualitative), or does it yield an area of representativeness (quantitative)?
Simplicity	Methods should preferably be straightforward to understand and apply. The end users will be a diversified group with different educational and professional backgrounds throughout the MS. It is advisable, also for acceptability, that the method is widely accessible.
Applicability for SR purposes	The concept of SR is serving a wide variety of purposes: exposure assessment, model validation, data assimilation, network design. It should be advantageous if the method can serve as many purposes. As mentioned before the introduction of a SR area might be an intermediate step to serve the application domains.

2.2.1 Methodology outcome and purpose

The following observations can be made regarding the methodologies described in the literature:

- Different methodologies are used, ranging from dedicated monitoring campaigns using both mobile and fixed measurements, to detailed air quality modelling at different scales and using different modelling techniques.
- Many publications aim to deliver either an explicit area of representativeness or a classification. Interestingly, there are methods which combine both, yielding a more complete picture (Soares et al., 2018).
- A large part of the effort documented in Table 1 is targeted at station classification, optimising air quality networks and assessing station redundancy.
- When determining the spatial variability around monitoring stations, any model-based approach (either via GIS modelling or a full air quality model) may have bias relative to the true concentrations. Hence, there are methods to account for bias when delineating an area of representativeness. Several authors use co-kriging to remove the spatial bias between the observations at the monitoring sites and a model-based assessment.
- Most references are documented in scientific literature, only a few reports by Member States are found outside the scientific literature where there is a direct reference to spatial representativeness in the context of e-Reporting which means there is limited evidence related to common practice in the reporting community.
- One author (Beauchamp et al., 2018) introduces the concept of observation uncertainty into the SR area assessment via Kriging. Interestingly, the same author correctly states that there should be a clear distinction between an area of similarity and an area of exceedance. This is an important realisation and may be dealt with in different ways, depending on the spatial resolution of the approach. For a modelling approach which does not fully capture the spatial variability, a probabilistic framework can be introduced to estimate the probability of exceedance. However, when using high resolution modelling approaches, the area of exceedance or area of similarity can be derived from the high resolution maps (provided that comprehensive emission information and sufficiently high spatial resolution is used). It should be mentioned here however that exceedances of limit values in the AAQD are hard thresholds, so it is unclear if and how an exceedance probability could be interpreted in a legal context.

2.2.2 Establishing the spatio-temporal variability

Depending on the pollutant and the air quality metric considered, it is well known that there can be a high degree of spatial variability. Sharp gradients in the concentration field can be observed especially in urban areas, but also near industrial sources. A striking example of this spatial variability was recently noted in Copenhagen, where the traffic station at the H.C. Andersens Boulevard (Copenhagen/1106) experienced a decrease of about 8 µg/m^3 in the annual averaged NO_2 concentration as a result of moving its location 2.7 m further away[13] from the inner traffic lane (Ellerman et al., 2018). In Antwerp, the CurieuzeNeuzen monitoring campaign established the spatial variability in NO_2 concentration for a whole city, showing that the NO_2 monthly mean as measured by passive samplers can differ by a factor of two between neighbouring streets.

To account for spatial variability it is important that the methods employed to establish the

[13] Note that this relocation in 2016 was to compensate for a previous rearrangement of the traffic lanes in 2010, which effectively brought the traffic 2.7 m closer to the monitoring location. The relocation in 2016 was therefore to restore the original monitoring conditions before 2010.

spatial pattern for the AQ metric provide data at a spatial resolution consistent with the expected level of spatial variability of the AQ metric in a given environment. Figure 2 provides an example showing that the observed variability dramatically increases already at short distances. When attempting to capture this variability using model-based approaches, regional scale models or models which employ a continuous open-line dispersion methodology without accounting for urban structure (in particular the occurrence of street canyons) can fail to resolve the spatial variability established by monitoring. A model-based semi-variogram can only reproduce the observed structure of differences in air pollutant concentrations when explicitly accounting for the street canyon increment and thus the strong inhomogeneity and discontinuities in the urban concentration field.

Similar considerations can be made for temporal variability. Some model-based approaches for example derive annual averaged concentrations, or more subtly use annual emission values, disaggregated with fixed time factors, while others estimate shorter timescale variability. Temporal resolution should be sufficient to capture observed temporal variability of the air quality metric. Hence it is important that any approach establishing the spatio-temporal variability should be fit-for-purpose (see further in the discussion under Section 2.2).

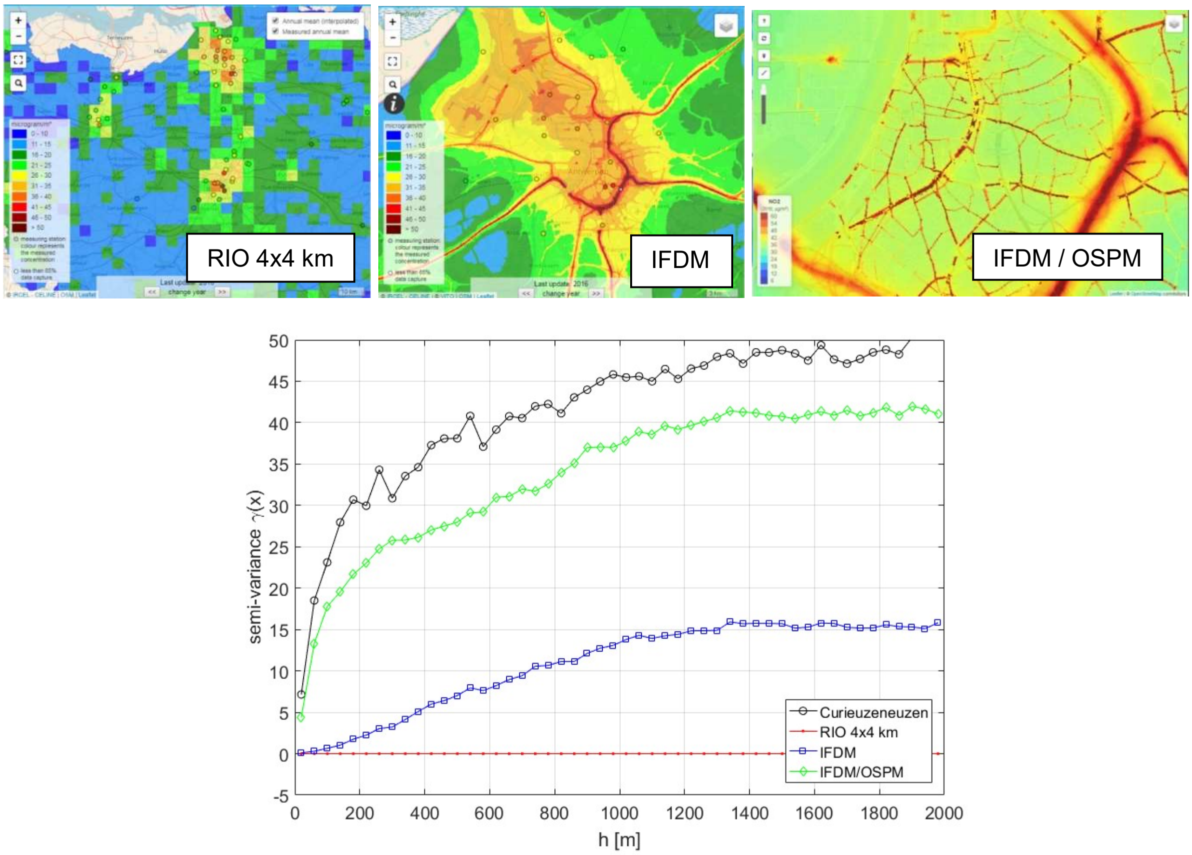

Figure 2: Illustration of different model based approaches attempting to capture the spatial variability of the CurieuzeNeuzen dataset for Antwerp (https://curieuzeneuzen.be/). The top figure shows the maps for Antwerp for 3 different model assessments, a regional scale model at 4x4 km resolution, a Gaussian dispersion model, only taking into account open-line dispersion from roads and on the right hand side, the same model, complemented with a street canyon module (OSPM), which allows to estimate the impact of reduced ventilation between buildings in the urban environment. Underneath, the semi-variogram is shown, which gives a measure of the spatial variability as a function of separation distance (horizontal axis). The latter gives the Euclidean distance between 2 points in space, hence the left hand side of the horizontal axis are any two points which are close together, the right hand side gives any two points which are far apart. The vertical axis gives the variance of the concentration difference between the observations which are separated on average by the distance given on the horizontal axis.

Table 3 below summarises the ways identified in the literature (see Table 1) to establish spatio-temporal variability around monitoring stations. We provide references for the methodologies along with an assessment of the criteria listed above. Low cost sensor networks are included as a potentially interesting future methodology. Methods which allow for the delineation of a SR area have been considered and not methods primarily intended to deliver a station classification, although literature suggests that an explicit delineation of a SR area can aid in understanding station classification. Methodologies are grouped into:

- **Obstacle resolved modelling (CFD)**: In obstacle resolved modelling, flow around buildings is explicitly represented via the solving of fluid dynamics equations on a 3D grid which accounts for the buildings. The technique offers a very detailed view of the flow and pollutant dispersion patterns and allows to explicitly model the impact of vegetation elements, screens and other obstacles. It does however come at a significant computational burden and is rarely used on areas larger than a few streets. Examples of patterns obtained by CFD modelling and the level of detail obtained with this type of method is given below in Figure 3, clearly showing CFD modelling is able to resolve concentration gradients inside the street canyons. For practical application, the so-called Reynolds Averaged Navier-Stokes approach (RANS) is typically used whereby a steady state solution is sought for a given set of boundary conditions. Meteorological statistics are often applied to produce annual averaged or hourly results. Atmospheric stability effects are rarely looked at and usually calculations are done for a neutral atmosphere.

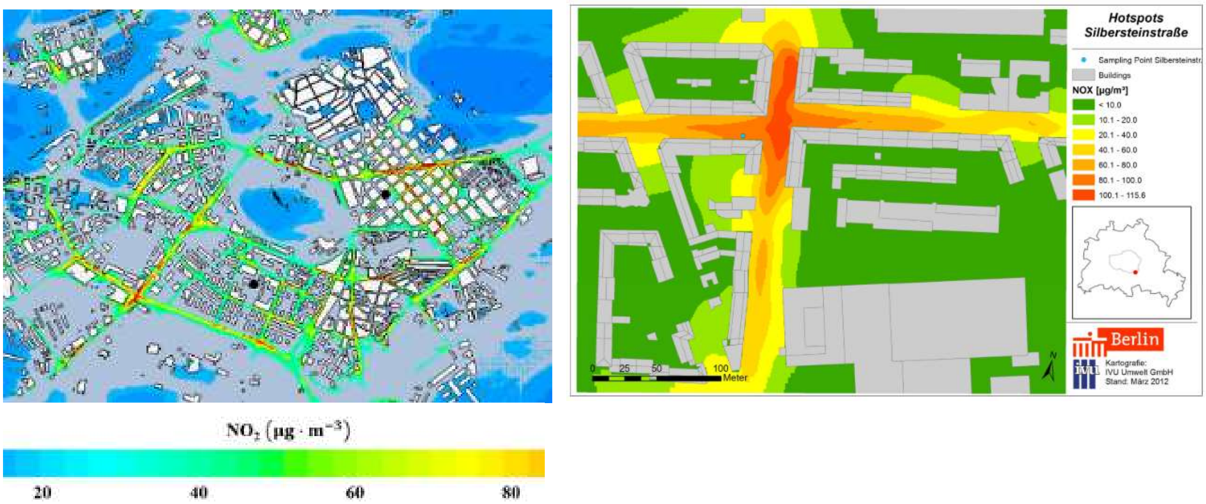

Figure 3: Examples of CFD modelling, left (Rivas et al., 2019), right: (Diegmann et al., 2015). On the right hand side concentration gradients inside the street canyons are resolved.

- **Local scale dispersion modelling (Gaussian, Lagrangian):** Typically, Gaussian or Lagrangian modelling, approaches are used potentially including street canyon parameterisations. Important to note here is that the emission sources are modelled explicitly for example as line sources (e.g. traffic, shipping) or point sources (e.g. industrial stacks). These models are well suited to resolve concentration gradients from such sources, but typically lack the capacity to explicitly model the flow around obstacles. Such effects are usually included via parameterisation e.g. of a street canyon increment. Examples of application of local scale dispersion models are given below. Local scale dispersion models are very widely spread in policy support and the evaluation of local air quality action plans. They often account well for effects of atmospheric stability and (depending on the modelling approach) allow for generating temporal information (i.e. hourly concentration fields). Most local scale dispersion models are able to represent the fast NO_x-O_3 chemistry, but usually lack a comprehensive description of long-range transport and transformation processes.

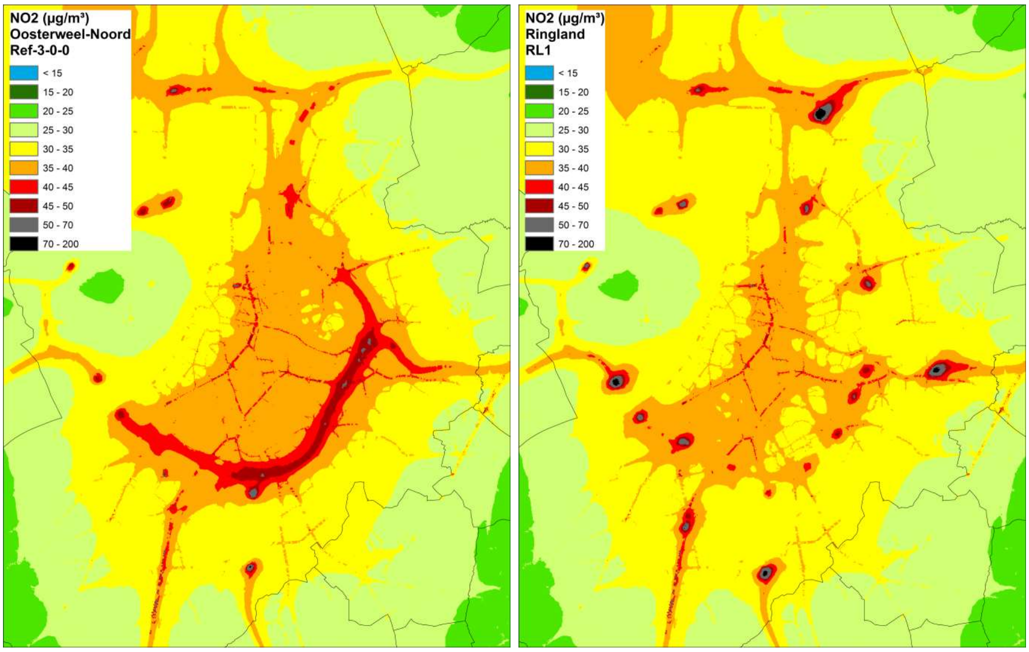

Figure 4: Top: IFDM-OSPM NO_2 concentration map for the city of Antwerp, Belgium. The figure on the left shows a particular future urban development scenario (2020), the figure on the right shows the impact on top of the scenario covering the entire ring road as proposed by a local action group (Ringland). Source: VITO / Ringland (www.ringland.be)

- **Chemical transport modelling**: Chemical transport modelling (CTM) is an established technique to represent large-scale concentration patterns, accounting for long-range transport and transformation phenomena. The spatial resolution of the modelling approach is typically limited to the ~km scale. Though the technique does not represent intra-urban variation well, it does generate consistent regional or continental scale concentration fields which typically act as background concentrations for more localised assessments. Emissions are typically aggregated to a raster and urban properties are represented via an aggregated roughness length assigned to the model grid cells. The Copernicus Atmospheric Monitoring Service (CAMS) currently operates[14] an ensemble of CTMs for Europe at a resolution of 10 km.

[14] https://atmosphere.copernicus.eu/

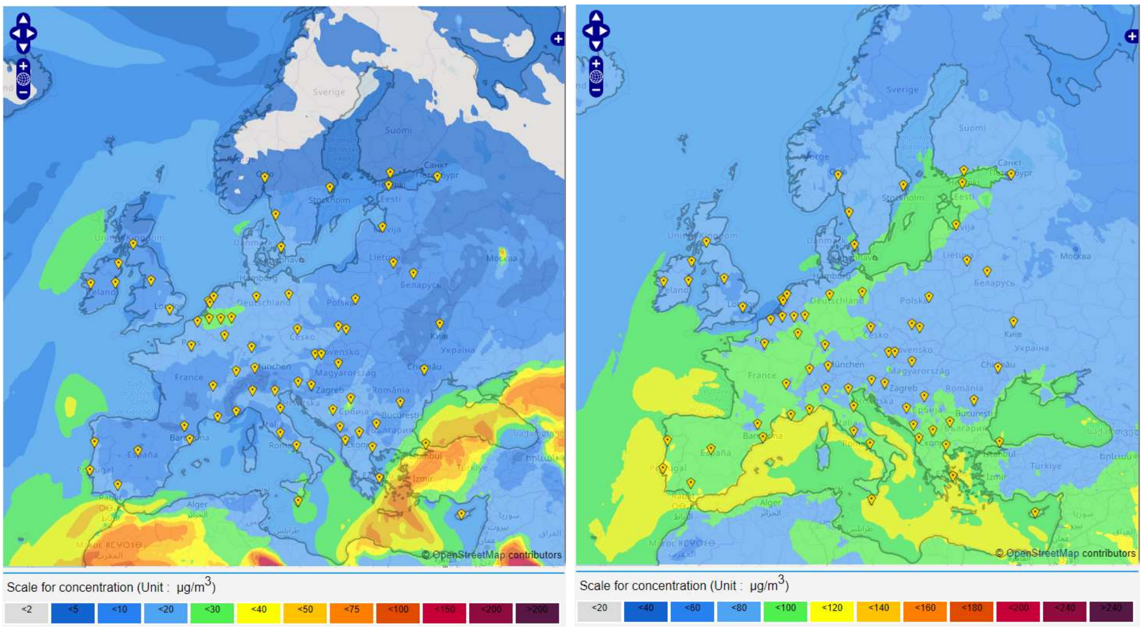

Figure 5 : Example forecast maps for daily mean PM_{10} and daily mean O_3 concentration from the CAMS Chemical Transport Model ensemble, see http://www.regional.atmosphere.copernicus.eu/

- **GIS-based methods, Land use regression, emission proxies**: Representing the spatio-temporal variability via proxy parameters is also a widespread practice where, especially in epidemiological communities, land use regression models are widely employed. In such approaches, the concentration patterns are represented by geographical information system (GIS) based proxy data and some form of regression and/or geostatistical interpolation technique such as (co-)Kriging. Guidelines for this type of modelling methodology have been published in the framework of the TRANSPHORM project (Denby, 2010). Typical proxy parameters that are used include “distance to road”, “total length of road in a given buffer”, and population density. These techniques can only approximate the gradients from local sources (via statistical relationships with GIS proxy data).

- **Passive sampling**: This involves the deployment of so-called passive samplers, which typically measure an average in-situ concentration value integrated over an extended time period. Recently, the use of passive samplers on a large scale has become popular for citizen science projects. Being inexpensive, passive sampling campaigns allow for very dense sampling campaigns but lack the capacity for very fine scale time-resolved measurements.

- **Mobile monitoring campaigns** employ portable monitoring equipment installed on a mobile platform and allow transects of concentration values to be obtained. Though the technique allows for very detailed representation of spatio-temporal concentration patterns along streets or bicycle paths, results should be interpreted cautiously, as they can be distorted due to issues of sensor response time and effects due to the position of the mobile platform in traffic (e.g. directly behind strongly polluting vehicles). From this perspective, there may be a need for developing quality assurance techniques for aggregating mobile data.

In addition to these more technical methods, "**expert opinion**" is often used in practice. In such an approach, an expert judgement is used based upon a mixture of qualitative and quantitative considerations about proximity of sources, dispersion conditions and siting in general.

Table 3 summarises the groups of methodologies derived from the literature and outlined above in terms of the evaluation criteria from Table 2 and in terms of applicability from the perspective of the assessment needs listed in Section 1.3.

Table 3: Methods for capturing spatio-temporal variability

Methodology	Data availability	Fitness for purpose				Applicability
		Spatial	Temporal	Coverage	Simplicity	Assessment needs
Obstacle resolved modelling (CFD) **(Diegmann et al., 2015; Rivas et al., 2019; Rodriguez et al., 2019; Santiago et al., 2013; Scaperdas and Colvile, 1999)**	Requires detailed emissions and 3D building information	Allows for very detailed obstacle resolved concentration fields at high spatial resolution (e.g. 1 m), therefore well suited for resolving strong gradients around buildings and other obstacles.	Usually a steady state approach, so need meteorological statistics or scaling to generate annual averages, or hourly values	Limited area (e.g. 5 x 5 km)	Requires expert knowledge	• Can be used to delineate SR areas in urban (traffic-related) settings or near industrial sources for pollutants with strong local variability, mainly NO_2 / NO_x, but also PM, depending on completeness of emission inventory (assessment needs a) to c) and f)). • Given the very demanding computational nature of CFD calculations, the modelling domain remains limited and the method therefore is mainly targeted at assessing intra-urban variability. Nevertheless, very recently such methods have been applied to entire cities (Rivas et al., 2019) • Not well suited to derive SR areas for percentile air quality metrics, unless relationships with annual averages can be found. • Not applicable for classification. • Suited to account for microscale environment and dispersion conditions where a complex environment geometry is relevant and thus can facilitate configuration of a representative monitoring network (assessment need d)).
Local scale dispersion modelling (Gaussian, Lagrangian) **(Blanchard et al., 1999; Cosemans et al., 1997; Duyzer et al., 2015; Vardoulakis et al., 2011b, 2005; Vitali et al., 2016)**	Requires detailed emissions down to line and point source level and 3D building information	Resolution depends on emissions considered (e.g. ~10m to ~100 m), may include parameterised canyon effects	Typically, hourly resolution is available	Can extend to an urban agglomeration or region, less suited for national or regional scales.	Requires expert knowledge	• Suited for assessment needs a) to f), depending on the completeness of emission inventory and accounting for street canyons. • Especially suited for assessment need b since line sources are explicitly taken into account. • Due to the hourly time resolution, also suited to derive SR areas for the percentile based metrics for all pollutants. • Can aid in the configuration of network (assessment need d)) via local scale source apportionment and SR area (see Section 1.4.1). • Less well suited for microscale environment dispersion conditions as dispersion at microscale is parameterised.

						• Allows for source apportionment for primary emissions and therefore possible to study representativeness as a function of emission sectors.
Chemical Transport Modelling **(Martin et al., 2014; Piersanti et al., 2015; Soares et al., 2018)**	Requires comprehensive emission inventory as well as 4D meteorological data and boundary conditions. CAMS CTM data freely available.	Resolution from continental scale to ~1-2 km CAMS at 10 km EU-wide.	Hourly	National - Regional coverage	Requires expert knowledge, though CAMS data is freely available	• Not able to resolve urban scale concentration gradients and thus not suited to assess SR in these areas. • Mainly used to quantify SR areas for rural background areas. • Not suited for assessment need b) (length of road) only a) and c), also to a lesser extent d) and e). • Some authors use CTMs for optimisation of monitoring networks (assessment need d)) (Soares et al., 2018). • Allow for source apportionment and therefore study representativeness as a function of emission sectors, these considerations are valuable for facilitating the configuration of a representative network.
GIS-based methods, Land use regression, emission proxies **(Beauchamp et al., 2018, 2011; Bobbia et al., 2008; Janssen et al., 2012; Karabelas and Sarigiannis, 2008; Righini et al., 2014; Sarigiannis and Saisana, 2008; Spangl et al., 2007)**	Requires observations and GIS proxy data, both are typically available	Resolution dependent on input data, ~100 m –5 km	Depends on application, can be deployed directly on metric required e.g. Kriging interpolation of 90.4 percentile value for PM_{10}, or required metric can be computed based on higher time resolution application	Depends on coverage of proxy data but can cover all scales. Usually adopted GIS proxies (CORINE, open street map etc. are available EU-wide).	Requires GIS knowledge, but straightforward to apply given GIS expertise.	• Approximate the spatial variability via proxy information; where most approaches typically try to mimic dispersion relations from air quality models. • Depending on the scale resolved by the proxy information, method is suited to derive SR areas and therefore deliver assessment needs a and c, though, estimation of the length of road in exceedance (assessment need b) is less obvious. • Given the statistical nature, typically all pollutants and metrics can be covered • Does not allow for source apportionment and are therefore only to a lesser degree suited to facilitate configuration of a representative network (assessment need d)).

Passive sampling**(Ott et al., 2008; Vardoulakis et al., 2011b, 2005)**	Requires dedicated campaigns	In-situ concentration values; need to be complemented by spatial modelling to have full coverage	Usually shorter campaigns relative to averaging time of the annual environmental objectives, (a few weeks) however corrections can be applied to yield e.g. annual averaged metrics.	Limited to local area, unless in exceptional cases such as CurieuzeNeuzen	Requires expert handling and calibration of sampler tubes. Citizen science is a proven option here.	• Generally not suited to estimate the SR area as in-situ data does not provide a full areal picture unless complemented by modelling. • Well suited to facilitate the configuration of a representative network when campaigns have been set up specifically for this purpose e.g. (Vardoulakis et al., 2005). • Can be used to determine residual spatial variability within the area of representativeness. • Availability of monitoring technique and aggregation time depends on pollutant.
Mobile monitoring campaigns **(Li et al., 2019)**	Requires dedicated campaigns Citizen Science	High spatial and temporal resolution (~ 10 m / seconds), but aggregation needed to yield useful results in this context. Requires response time correction.		Only a snapshot relative to the AAQD reporting metrics being on annual timescales.		
Low cost sensor networks	Requires dedicated campaigns Citizen Science	As above		As above	As above	• May in future be used for similar purposes as more mature monitoring-based assessments, but currently insufficient quality.

With respect to the evaluation of methodologies presented in Table 3, some conclusions can be drawn and further discussion points drafted:

- Geographically explicit approaches are required to delineate a SR area to meet the requirements for the first 3 assessment needs set out in Section 1.3. Such methodologies are either based on air quality modelling or GIS modelling by means of proxy data. It is clear that where air quality modelling based approaches require significant expertise (independent of the type of modelling), GIS based methods may be more straightforward to apply.
- Assessment need b), looking to quantify the total length of road in exceedance of the environmental objective suggests approaches where line sources are treated explicitly, or modelling is carried out at high spatial resolution.
- Studies that attempt to quantify the total length of road in exceedance, or (related) the representativeness of traffic stations argue that either the results of measurements at traffic stations are not useful to assess exposure of the general population (Duyzer et al., 2015), or indicate that the requirement that sampling points at traffic-oriented sites should be representative of air quality for a street segment no less than 100 m seems almost impossible (Diegmann et al., 2015). This indicates the need for very high resolution modelling to be able to resolve the variability. The authors mention the additional complication of inhomogeneous traffic emissions along the street.

  When assessing spatial representativeness for NO_2/NO_x, chemical transport models are typically only used when dealing with rural background locations (Martin et al., 2014; Piersanti et al., 2015; Soares et al., 2018). Due to the limited spatial resolution of this approach, however, CTMs will fail to capture the spatial gradients near roads or other strong emission sources and may therefore underestimate for example the area in exceedance. This is illustrated by Figure 6. Indeed, when a CTM is used to derive an area of representativeness around a rural station, it will for example explicitly resolve a highway intersecting this SR area and may therefore overestimate the representativeness area or underestimate the area in exceedance as illustrated.

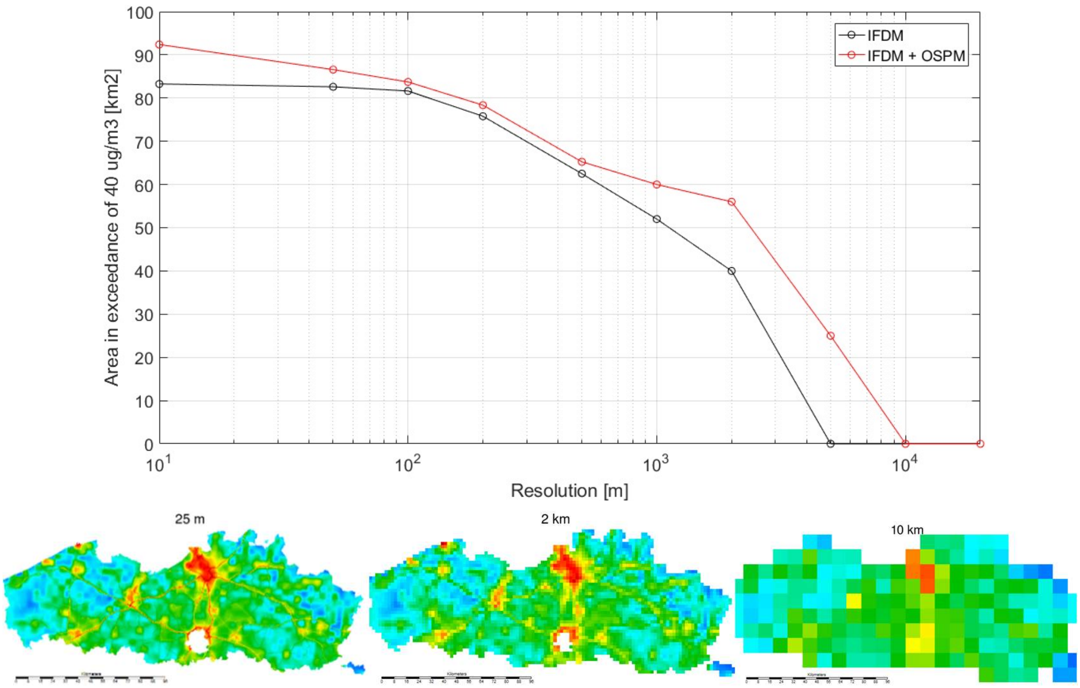

Figure 6: Illustration of the dependence of the area in exceedance of 40 µg/m³ NO_2 annual mean concentration in Flanders on spatial scale starting from the highest resolution map (25 m) and gradually degrading the resolution to 10 km. The figure above shows the total area in exceedance as calculated with the VITO ATMO-Street model taking street-canyons into account via OSPM (red line) or not (black line).

- In the literature presented, no mention was made of SR areas of O_3 concentrations at high spatial resolution. Clearly, O_3 is a regional scale pollutant, but due to the fast O_3-NO_x chemistry, there may be significant gradients close to roads. (Lin et al., 2016) for example deployed 30 passive sampler sites in Edinburgh for NO_2 and O_3 and investigated the temporal stability of the intra-urban spatial contrasts. They found O_3 to vary significantly between these sites, ~45 µg/m^3. (Sadighi et al., 2018) measured the intra-urban variability of O_3 in a ~20x20 km area in Riverside, near LA. The study analysed the median of differences between pairs of observations from calibrated low-cost sensor platforms (21 in total) and found the median of these differences to be between 4.4 µg/m^3 and 18.6 µg/m^3.
- With respect to assessment need c) estimating the total resident population in the exceedance area, it is clear that in addition to an estimation of the geospatial area of representativeness, population data is required. Determining population in an exceedance area can be challenging, but detailed assessments are able to resolve discontinuities in the urban areas (e.g. CFD modelling or street-canyon modelling). When exceedances are situated in street canyons, how does one assign population in neighbouring houses to these exceedances? Should either the street side or concentration at the rear of a building be taken, or some average? (Diegmann and Pfäfflin, 2016) describe a segment-based (SBE) methodology which can be used to assign population in a lower resolution population grid to detailed street level concentration estimates and have applied this for Berlin. Given the improved capacity within the community in terms of high resolution modelling, this is a discussion point which needs to be considered further.

- It is interesting to note that in the literature considered, though most of the attempts at optimisation of monitoring networks remain qualitative in their discussion and analysis, attempts have been made to look at this as a formal (spatial) optimisation problem. (Sarigiannis and Saisana, 2007) for example minimise a cost function including CAPEX/OPEX to this end and use GIS based methods to establish the spatial patterns. In such formal optimisation studies, the application of complex modelling using CTM, or urban scale modelling (depending on the approach) may not be suited due to computational restrictions. Care must be taken however when using GIS based approaches as the monitoring design and the GIS based approach may influence each other. A particularly interesting analysis was presented in (Wu et al., 2017) in which the effect of monitoring network design in Edinburgh on land use regression models for estimating residential NO_2 concentrations was investigated. The study suggested a lack of spatial contrast in the LUR modelled pollution surface and argued that dispersion models (in this particular case ADMS-Urban) are shown to be useful tools for designing monitoring networks. It is also worth mentioning that more recently, the concept of SR area was applied for optimisation of urban monitoring networks in cities in China (Hao and Xie, 2018).
- Dedicated sampling campaigns are an interesting way for assessing spatial variability as many studies in the summary table above indicate. Most notably, the CurieuzeNeuzen campaign in Flanders, Belgium explored the potential of such passive samplers in a large citizen science experiment[15]. Here 20,000 passive samplers were distributed to volunteers, asking them to suspend the tubes for a period of a month (end of April 2018 – end of May 2018), after which returning them for further analysis. Next to using passive sampler readings at fixed locations, it is possible to move the monitoring equipment. Two approaches are possible (Gillespie et al., 2017):
    - *continuously mobile monitoring*, where the monitoring equipment is moved throughout the duration of the study; (Joris, 2016; Peters et al., 2014; Van den Bossche et al., 2015)
    - so-called *peripatetic monitoring*, where mobile equipment is deployed at specific sites for short time periods before moving to another site (Gillespie et al., 2017)
- Currently, low cost sensors are not yet perceived ready for applications that require high accuracy and can therefore only provide coarse information about the observed concentration levels (see

[15] https://curieuzeneuzen.be

Table 4).

Table 4 below discusses some additional elements concerning the applicability of monitoring-based methods for assessing the spatial variability (mainly for assessment need e).

Table 4: Table summarising the applicability of monitoring based approaches for establishing SR around monitoring stations.

Methodology considered	Papers/references (not exhaustive)	Applicability for assessment needs
Passive sampler campaigns	(Hagenbjörk et al., 2017; Vardoulakis et al., 2011b); CurieuzeNeuzen Flanders	• NO_2 Palmes tubes have successfully been applied in multiple studies. • Usually shorter campaigns compared to. averaging time of the annual environmental objectives however corrections can be applied. • Not applicable for percentile values. • Can be deployed on large scale via citizen science experiments (e.g. CurieuzeNeuzen), however, significant effort regarding organization required.
Mobile monitoring campaigns	(Gillespie et al., 2017; Li et al., 2019; Peters et al., 2013; Van den Bossche et al., 2015; Van Poppel et al., 2013)	• Difficulties interpreting results from mobile campaigns, response time corrections needed when using continuous monitoring, or some way of aggregating repeated results. • Only a snapshot relative to AAQD environmental objectives on annual timescales.
Low cost sensor networks	(Badura et al., 2018; Castell et al., 2017a; Sadighi et al., 2018; Spinelle et al., 2015)	• Significant challenges associated with sensor robustness and measurement repeatability (Castell et al., 2017b; Spinelle et al., 2015). • Issues related to sensitivity and low signal/noise ratios under ambient concentrations, chemical interference (in particular for NO_2 and O_3 using electrochemical sensors), transient effects in response to changes in relative humidity and dependence of the calibration on environmental conditions in general. These issues make low cost sensor use currently less obvious for detailed monitoring of spatial variations within an area of representativeness. For $PM_{2.5}$ the impact of high humidity is particularly problematic and high relative errors are seen at concentrations below 20-30 µg/m^3 (Badura et al., 2018). • Considering the current uncertainties, a concern would be the ability to distinguish sensor issues from true spatial variability. • Potential for very explicit measurement of temporal variability via high monitoring frequency (sometimes minute or second-basis).

2.2.3 Establishing similarity

In this section the second important aspect in the literature summary table is discussed, namely the ways in which the different methodologies establish similarity. Below a summary is provided of different methodologies along with a discussion of relevance from the perspective of the assessment needs. As all of these methods are purely based on mathematical techniques, the criteria used to evaluate the methodologies for capturing the spatio-temporal variability are not applicable here. However, Table 5 indicates what if any spatial model is applied.

Table 5 : List of methodologies for establishing similarity between two geographically distinct locations.

Similarity Methodology	Description	Spatial model
Euclidean distance	The two sites should not be separated by more than a certain distance; hence the SR area will be a circular buffer around the monitoring location	n/a
Fixed absolute threshold; (Spangl et al., 2007)	The value of the air quality metric is only allowed to differ by some absolute number (e.g. +/- 5 µg/m^3) in the SR area	• Can be applied to different model resolutions, doesn't require time resolved modelling.
Fixed relative threshold; (Blanchard et al., 1999), (Rivas et al., 2019), (Martin et al., 2014)	The value of the air quality metric under consideration is only allowed to differ by a fixed percentage (e.g. +/- 20 %) in the SR area	• Can be applied to different model resolutions, doesn't require time resolved modelling.
Linear Discriminant Analysis (LDA) (Joly and Peuch, 2012)	Timeseries of concentration data at monitoring locations are analysed for certain time series characteristics yielding a set of 8 characteristic indicators.	• n/a, the method described is purely an objective classification method.
Principle Component Analysis (PCA) (Nguyen et al., 2009)	A dimensionality reduction is performed on timeseries of hourly data for the selected station, resulting in loadings per station on 2 principle components axes. These 2D loadings are used for classifying the stations in rural, urban background and traffic.	• n/a, fixed distance taken
Concentration Similarity Function (CSF) (Nappo et al., 1982; Piersanti et al., 2015; Vitali et al. 2016)	Imposes the requirement on the two locations that their so-called *frequency function* is above 90%, indicating that for 90% of the samples in the timeseries, the relative difference between the concentration values is less than 20%	• Required timeseries of gridded model output (CTM or Lagrangian model) • Or (as in the IE) inverse distance weighted interpolation (IDW) applied.
(Temporal) Pearson correlation coefficient and Euclidean distance; (Soares et al., 2018)[16]	• Apply timeseries filter to isolate specific frequency components in the timeseries • Can be used for network design → assessing potential station redundancy	Have used GEM-MACH CTM to provide gridded timeseries concentrations.
(Temporal) Pearson correlation coefficient and NRMSE[17]; (Rodriguez et al., 2019)	Based on timeseries of simulated concentrations	PMSS (Parallel Micro-Swift-Spray) high resolution air quality model at 3 m horizontal resolution.
Kriging based approach; (Bobbia et al., 2008), (Beauchamp et al., 2018)	Expands the criterion of an absolute threshold to a probabilistic approach by requiring that the **expectation value** of the difference between the metric at the observed value be smaller than a given delta. This is demonstrated to be rewritten into the requirement that the difference between the (co)-Kriging estimate and the observed value be smaller than the given	Emission based co-variates (both local and regional scale).

[16] In (Soares et al., 2018), the authors complement the 1-R metric, R being the Pearson correlation coefficient with a Euclidean distance between locations.
[17] Normalized Root Mean Squared Error

	delta corrected for the Kriging error variance.	
Wilcoxon rank-sum test[18]; (Li et al., 2019)	Using a statistical test to establish whether a grid cell is statistically different ($p<0.05$) from a reference cell.	Using mobile monitoring results in 50 m cells.
Coefficient of Divergence (COD) (Ma et al., 2019)	Employs coefficient of divergence together with satellite imagery (AOD) to establish representativeness Used to evaluate whether there are redundant observations.	Satellite imagery of AOD

The clear challenge in the formulation of the similarity criterion is there is a huge variety of methods documented in literature.

- A wide variety of methodologies exist. For methods aiming to establish an explicit spatial representativeness, methods can be distinguished which employ a fixed absolute or relative threshold on the metric, and methods which additionally include more fine-grained temporal information. Examples of the latter are the so-called concentration similarity function, and methods which use the Pearson correlation coefficient r.
- Typically, methods that attempt to identify sampling points suited for calibration/validation benefit from objective classification or clustering methods which are based on the properties of the observations, predominately temporal aspects are considered, potentially being complemented with additional GIS-based or distance related metrics. In such methods, monitoring sites are classified into distinct types indicative for different sampling conditions by means of supervised or unsupervised learning algorithms and different methods have been applied in literature: principle component analysis[19] (PCA), linear discriminant analysis[20] (LDA), hierarchical clustering[21] methods. Applying the concept of a spatial representativeness area as presented above in this case presents some difficulties also:
  - It may not fully capture the temporal aspects which are of importance in a validation exercise, unless temporal aspects are included in the similarity criterion employed to define the area.
  - The method to derive the area of representativeness may be model-based in itself (either via proxy GIS information or via explicit high-resolution modelling) and therefore the validation exercises is not fully independent anymore.
  - Not every model is designed to simulate every aspect of the observations. Certain models, for example the Dutch OPS[22] model, are only able to generate long term averaged metric such as annual mean and can therefore not be validated for short term temporal characteristics.
- (Joly and Peuch, 2012) mention that data assimilation procedures may be improved by selection of monitoring stations that are representative of geographical areas related to the spatial resolution of the models, but no further discussion was given as to how to calculate these areas. Indeed, in several publications it is found that objective classification methods do not always allow a clear separation between classes. It was discussed already in (Spangl et al., 2007) and emphasised in (Joly and Peuch, 2012)

[18] Also known as the Mann-Whitney U test
[19] https://en.wikipedia.org/wiki/Principal_component_analysis
[20] https://en.wikipedia.org/wiki/Linear_discriminant_analysis
[21] https://en.wikipedia.org/wiki/Hierarchical_clustering
[22] https://www.rivm.nl/operationele-prioritaire-stoffen-model

that the inclusion of too many parameters in the classification may lead to an over-categorisation of the sites with too many subgroups to be practically useful. Therefore it could be argued that such classification methods could be complemented with more spatially explicit emission related information, such as SR areas around the stations which account for the nearby local emissions and would therefore allow to discriminate between for example two stations labelled as “industrial”, but are representative for an entirely different emission pattern and have substantially different SR areas.

- In addition to the methods discussed above, the Kriging-based approach mentioned in (Beauchamp et al., 2018) deserves attention as this approach allows to integrate a model-based assessment (irrespective of the spatial scale[23] and method: GIS based or air quality modelling). It can be considered a probabilistic extension[24] of the requirement for a fixed value-based similarity criterion:

$$|Z(x) - Z(x_0)| < \delta \rightarrow \ E(|Z(x) - Z(x_0)|) < \delta$$

This could potentially be seen as a more advanced way of defining the similarity criterion, but taking a more probabilistic approach which allows to treat uncertainties and account for probability of exceedance in the SR area.

- These considerations are no longer valid when looking for reducing redundancy in monitoring networks or optimizing networks. Here, more elaborate definitions for the similarity criterion can be employed, complemented by a spatio-temporal assessment model. Such approaches employ e.g. temporal correlation coefficients (Soares et al., 2018) or the so-called concentration similarity function. As the aim of the network optimisation is highly specific, it is difficult to draw unified conclusions. Similar considerations are valid for identifying stations suited for validation as this depends largely on the purpose of the validation. What aspects of the model are under scrutiny e.g. temporal aspects, spatial aspects, a combination of both, the distribution of concentration values, the extremes. The differences in purpose may explain to some degree also the range of similarity assessment methods in the literature.

There are also large differences in opinion on what maximum range should be allowed to establish spatial representativeness. (Spangl et al., 2007) argue to take 100 km as maximum range, whereas (Martin et al., 2014) suggest 200 km to prevent underestimation of SR area in case of remote stations.

[23] Assuming that for a local/urban scale assessment, the spatial pattern is adequately resolved by the method used to estimate the spatio temporal variability.

[24] In the publication this definition is further improved by introducing the statistical risk that the concentration of any points within the representativeness area deviates by more than delta from the concentration measured in x0. The way we discussed this in this text merely serves the purpose of highlighting the probabilistic approach that this methodology brings to the table.

3 Methodological requirements for different applications

3.1 Introduction

In this section we present the practical requirements for determining spatial representativeness for the different applications introduced in Chapter 1 in order to bring in some focus and structure and to account for the paradigm shift that was suggested by the IE.

The SR intercomparison exercise report was presented as a modular approach towards better SR characterisation. In this modular approach, four different aspects need to be made clear. These are:

- The **purpose** of evaluating SR, including the **legislative requirements** from the perspective of the AAQD.
- The **set of metrics / characteristics** required for the purpose/application.
- **Context** related definitions of SR metrics.
- Fitness of **technical methods** for estimating a particular SR metric.

Below, we present, for each of the assessment needs, the metrics and characteristics for the given purpose as well as context related definitions of the SR metric, if applicable. These concepts provide the theoretical framework for discussing the applicability of different methodologies. We present what we are aiming to quantify and what information related to spatial representativeness is required. Presenting and agreeing first on the needs from a theoretical point of view should better guide the community towards consensus, but also help the EC in formulating guidelines for the recommendations.

3.2 Estimate of the spatial area where the level was above the environmental objective

3.2.1 Legislative requirements

The purpose of this assessment need is to infer geographical areas (area, km^2) where the air quality metrics are in exceedance of the limit values. As mentioned in the guidance document (DG-Environment, 2018), where environmental objectives for the protection of human health have been exceeded, estimates of the total area, population and, where applicable, road length exposed to levels above the environmental objective shall be reported for each zone.

The following reporting requirements have been deducted for PM_{10}, NO_2, $PM_{2.5}$ and O_3 from the AAQD:

- the area in which the **annual averaged PM_{10}** concentration exceeds **40 µg/m^3**.
- the area in which the **90.4th percentile value** of the **PM_{10} daily means** exceeds **50 µg/m^3**
- the area in which the **annual averaged $PM_{2.5}$** concentrations exceed **25 µg/m^3**
- the area in which the **93.2th percentile** value of the **daily maximum of the 8-hour moving average O_3** concentrations exceeds **120 µg/m^3**
- the area in which the **daily maximum of the 8-hour moving average O_3** concentrations exceeds **120 µg/m^3**
- the area in which the **annual averaged NO_2** concentrations exceed **40 µg/m^3**

In addition, there is an objective for NO_2 hourly means corresponding to the area in which the 99.8th percentile value of the NO_2 hourly means exceeds 200 µg/m^3. However, given the assessment of representativeness for short term values is notoriously difficult and there is no clear relationship between the NO_2 annual averaged value and the 99.8th percentile, we will not consider it further.

3.2.2 Set of metrics and characteristics required

The listed environmental objectives for this purpose require:

- The estimation of a **geographically explicit area**. Indeed, the guidance document (DG-Environment, 2018) clearly states "Associated geometry information (GIS data) shall also be provided", which implies the delivery of this information e.g. as polygon shapes.
- The capability to estimate for PM_{10}, $PM_{2.5}$ and NO_2 the annual mean concentrations, as well as the listed percentile values and moving averages (in case of O_3). In the case of PM_{10}, the required percentile values may be translated to effective values of the annual mean concentrations to simplify the estimation, however such statistical relationships may differ from region to region and depend on the underlying variability of the emissions. The uncertainty of using such relationships should therefore be considered if such an approach is adopted.

Since the environmental objectives relate to particular **time aggregations**, the time aggregation of relevance in the assessment of the SR is the time aggregation of each environmental objective e.g. the SR of the annual mean for PM_{10} at a particular location would be relevant while the SR of hourly mean PM_{10} at that location would not.

3.2.3 Context related definitions of the metrics

For this particular assessment need, the time aggregation of the metrics required in the context of the AAQD is defined and the need for reporting a geographically explicit area is required. This means that the similarity criterion should primarily be concerned with the spatial variation of the air quality metric and should not be concerned with the temporal aspects of how the metric is constructed[25]. Similarity criteria which adopt a relative or absolute threshold, or a combination thereof as mentioned in the IE can be considered adequate if the method to establish the spatio-temporal variability is fit for purpose and adequately reflects the relevant emission/meteo information.

A consensus should be sought as to whether this primary similarity criterion, as it is put by the FAIRMODE IE (see (O. Kracht et al., 2017) p 53), is sufficient for this assessment need. This will be further elaborated in the sensitivity studies as there is currently no consensus regarding whether or not to use a fixed absolute threshold, a relative one or a combination thereof. (Spangl et al., 2007) argues for a fixed value[26], whereas several other authors argue for a relative value of 20 %, potentially depending on the values of the metrics themselves (Martin et al., 2014).

This specific assessment need requires estimation of an area in exceedance above the environmental objective. Such an area is closely related to an area of representativeness, but it is important to realise than an area *in exceedance* is not the same as an area *of similarity*. This issue is discussed in depth by (Beauchamp et al., 2018) bringing in the concept of exceedance *probability*. Nevertheless, in the legislative requirements, the exceedance of an environmental objective is formulated as a hard threshold, whereas the concepts of similarity and representativeness suggest a range or tolerance.

Furthermore, from the perspective of the AAQD, when a monitoring station is in exceedance of the limit value, the area of the whole zone in which it is located could be considered to be

[25] For example, by imposing a similarity criterion on the daily averages used to calculate the 90.4 percentile value for PM_{10}, or on the 8hr sliding mean values for O_3 used to calculate the corresponding metric.

[26] The concentration similarity thresholds for the representative area have been set as ±5 % of the total concentration range observed in Europe. These values, based on AirBase data for 2002 to 2004 :

- NO_2: Annual mean value at the monitoring station ± 5 µg/m3
- PM_{10}: Annual mean value at the monitoring station ± 5 µg/m3
- PM_{10}: Annual 90.4 percentile of daily mean values at the monitoring station ± 8 µg/m3
- Ozone: annual 93.2 percentile of daily maximum 8-hour mean values at the monitoring station +/- 9 µg/m3.

in exceedance. This enables the reporting of the area of exceedance in the absence of more detailed information on the SR of that particular exceedance. This would be a pragmatic and more conservative approach but may tend to deviate considerably from what is obtained when accounting for the true spatial variability of the AQ metric which might result in more confined areas where limit values are exceeded. Guidance is therefore required to overcome this issue as reporting a whole zone in exceedance based upon a single measurement is considerably different to having a spatially explicit assessment of the metric which accounts for the areas in exceedance within the zone, especially when population exposure is taken into account. It is also important to mention that the AAQD requires that measurement data is collected in those locations where the highest concentrations occur. However, it is very likely that in many situations this requirement is not fulfilled. The recent CurieuzeNeuzen measurement campaign in Flanders, Belgium pointed out that many local hotspots remain undetected by the fixed monitoring network and as such many more zones are likely to be in exceedance.

An assessment by zones allows MSs to organise policies and measures by regional/local authorities, assigning responsibility for local actions. Having a finer spatial assessment of the actual area in exceedance, explicitly accounting for the spatio-temporal variability, rather than considering the whole area as such in exceedance, will likely be beneficial for formulating such local action plans and enable more targeted measures to address the exceedance. Consistency between the areas used for planning and the delineation of the air quality zones is therefore important.

3.2.4 Fitness of technical methods

As this assessment need effectively requires the estimation of a geographically explicit area, the use of monitoring campaigns alone (not complemented by some form of modelling, be it GIS modelling or comprehensive air quality modelling) is insufficient, unless a very dense coverage can be obtained. In addition, dedicated monitoring campaigns, be it using passive samplers or via mobile measurements are often conducted during a short period and can therefore not reflect the annual aggregations as needed by the legislative requirements unless corrections are applied. Similarly, metrics which involve aggregations other than annual means may be difficult to reflect by measurements alone unless a full temporally explicit timeseries can be obtained. A dense geographical coverage in combination with temporally explicit measurements is a prospect which is promised by air quality sensors, however, to date issues with data quality of such devices still remain.

Obtaining such a geographically explicit area as required can however be done via GIS modelling (e.g. land use regression) as well as air quality modelling (using different techniques). When estimating the area in exceedance, the use of detailed air quality modelling may render some ambiguities obsolete (see discussion under section 1.2) as it provides a best estimate of the geographically explicit picture of the air quality and more importantly, quantifies the distance and dispersion relations between sources and receptors as well as contributions of individual sources.

Assuming the model application is of sufficient quality and fit for purpose (see below), the estimation of the area in exceedance may be done directly from the model results (preferably calibrated using the monitoring stations). In this sense, there is no need any longer to resort to determination of areas of representativeness around the monitoring stations and deal with issues such as whether or not the SR areas should be mutually exclusive or be contiguous or not. The determination of the area in exceedance can be provided based on the maps output by the model, directly yielding the required metrics listed under 3.2.1.

Important considerations when estimating such a geographical explicit area are:

- The methods employed for NO_2 should ideally possess the ability to resolve strong roadside gradients as well as account for street canyon effects. Hence, the spatial resolution of the models and the level at which traffic (as well as industrial) emissions are considered, should allow resolution of these gradients. This is clearly illustrated by

Figure 6 above, where a significant dependence of the area in exceedance of the 40 μg/m^3 on the spatial resolution can be observed.

- For PM_{10} and $PM_{2.5}$, the roadside gradients are smaller as road traffic is a less dominant contributor to the total concentrations, but resuspension of road dust should be taken into account via modelling.
- It is well known that O_3 concentrations near roads are lower due to the reaction with NO, hence, though sizeable road side gradients of O_3 may exist (and are observed in literature, see previous chapter), for the purpose of evaluating the total area where the O_3 standards are exceeded, a lower resolution modelling application (e.g. via chemical transport modelling) may be sufficient as these exceedances will more likely occur in non-urbanised areas.

As illustrated above in Figure 2, fitness for purpose of models involves the ability of accurately representing the spatial gradients in the concentration field and the variability. Semi-variograms were given as one possible way to assess this, but guidance is needed here.

3.3 Estimate of the length of road where the level was above the environmental objective

3.3.1 Legislative requirements

The purpose of this assessment need is very similar to the one stated under 3.2. There is however a subtle difference in the sense that we require here to estimate the total length of road in exceedance.

Clearly, these requirements refer to **kerbside** concentrations here as the AAQD does not require monitoring on the road itself (i.e. on the driving lanes, except where pedestrians have access to the central reservation), as noted in 2008/50/EC Annex III, under A.2.c.

Other than that, we can repeat the same metrics as listed in 3.2.1 with “total area” replaced by “total length”.

3.3.2 Set of metrics and characteristics required

Also, for the metrics and context related definitions we can repeat the same metrics and characteristics as under 3.2.2 with the only caveat that they should be available on line segments or street level.

3.3.3 Context related definitions of the metrics

See 3.2.3.

3.3.4 Fitness of technical methods

Evaluating the total length of road in exceedance requires the assessment technique to resolve the roads as well as the urban structure which governs dispersion from those roads (street canyons). This can be done via 3D obstacle resolved modelling (CFD) on a limited area, or via parametric models such as OSPM (Berkowicz, 2000), which account for the street canyon concentration increment given a number of geometrical parameters such as street width, building height, distance to nearest crossing.

In case of a central reservation when pedestrian access is possible, the technical methods based on modelling require the ability to resolve such detailed features and more “simple” street canyon parametrisation schemes may not be adequate. More detailed, obstacle resolved methods or in-situ monitoring for such specific circumstances may be required.

Considerable uncertainty may occur in such detailed assessments due to a lack of comprehensive traffic information requiring, for example, the use of generic vehicle fleet composition at national level, as opposed to the true fleet composition for the city of interest.

In the context of future so called “Smart cities”, where monitoring of traffic flows and composition are envisioned to be much more extensive, these uncertainties may be addressed.

3.4 Estimate of the total resident population in the exceedance area

3.4.1 Legislative requirements

The requirements under 3.2.1 can largely be repeated here, however, when assessing the total resident population, an overlay with a population map is required where typically the population is summed for the area of exceedance as derived in the first assessment need above.

An important consideration here is that the legislative requirement may be different from requirements in the context of health impact assessments. In (Maiheu et al., 2017), a first attempt was made to propose methodologies consistent with health impact assessments. In this case the methodology is required to be compatible in spatial scale with concentration response functions used to derive the health outcomes. Here, the AAQD and guidance documents stipulate simply to report the total population exposed which is typically based on residential address.

3.4.2 Set of metrics and characteristics required

In principle, the overlay is fairly straightforward given that population data and the concentration fields are available. Given the requirement to estimate the total population, in addition to the requirements under 3.2.1, a detailed population map (preferably address-level) is required here. By overlaying this population map with a spatially explicit area in exceedance (polygon shape), the total number of residents living in that area can be reported.

3.4.3 Context related definitions of the metrics

See previous sections.

3.4.4 Fitness of technical methods

The same considerations are valid as in 3.2.4.

An additional important issue is related to the use of very high resolution air quality assessments in this context. This was already mentioned when discussing the ambiguities in the current guidelines (see section 1.2) and is more clearly illustrated below in Figure 7. In these circumstances, the question arises how to assign the population adjacent to street canyons, in particular for situations where at the front side of the building, an exceedance may occur while there is none on the back side of the building. Clearly, this ambiguity arises from the discontinuity of the concentration field in the urban environment and will obviously lead to differences in the technical methods and hence differences in the SR assessment.

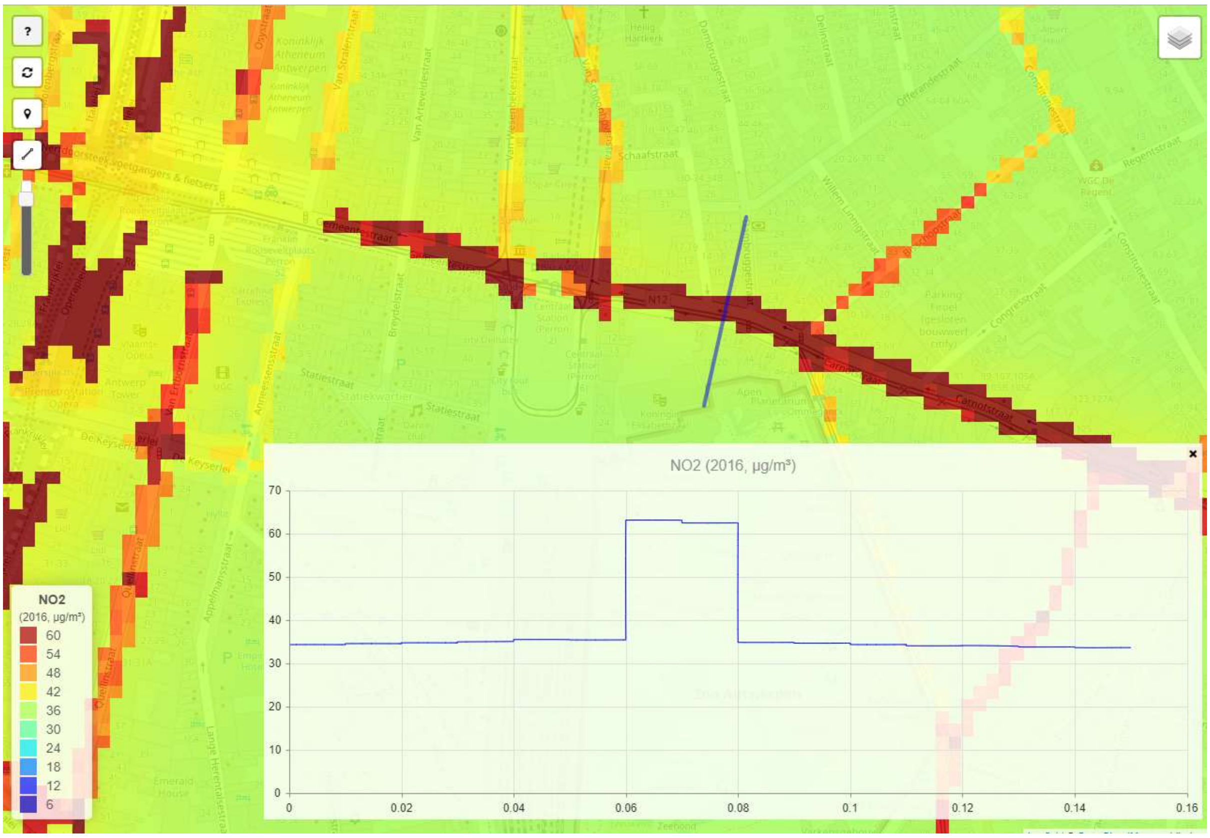

Figure 7: Example of the issue of where to assign population. In the detailed model, an exceedance of the 40 µg/m³ LV for NO_2 is modelled in the street canyon, whereas at the backside of the houses adjacent to the canyon, no exceedance is found.

3.5 Facilitate the configuration of a representative monitoring network

3.5.1 Legislative requirements

Member States are required to subdivide their territory into air quality zones, primarily intended for air quality management. The assessment requirements for each zone depend on whether in the preceding years, certain assessment thresholds, the upper assessment threshold (UAT) and the lower assessment threshold (LAT) were exceeded. Both are lower than the limit value and expressed as a percentage thereof. For example, for the NO_2 annual average limit value of 40 µg/m³, the UAT is at 80 % (32 µg/m³) and the LAT at 65 % (26 µg/m³), see Annex II of 2008/50/EC - A. If the UAT of a certain metric is exceeded, the most intensive assessment requirements apply for this pollutant; if the LAT is exceeded, but UAT is not, slightly less intensive assessment requirements are prescribed; if the levels are everywhere below LAT the least intensive requirements apply[27].

In addition to the total number of sampling points required by the Directives, being able to defensibly estimate the area, road length and population exposed to concentrations exceeding prescribed limit values is a critical requirement for reporting.

As well as defining the number of stations, there will be requirements on the monitoring network from a model validation point of view. To be able to adequately evaluate the performance of models in the context of reporting, it is necessary to have both enough and a representative sample of stations to perform a statistically significant validation. Evaluating the

[27] See https://ec.europa.eu/environment/air/pdf/guidanceunderairquality.pdf, Figure 3 and Table 1 on p 10.

model performance with a single traffic station for an entire urban area may not yield sufficient information to assess whether or not the model is fit for purpose. Such requirements are currently being developed in the framework of CEN working group 43 on Model Quality Objectives.

3.5.2 Set of metrics and characteristics required

In the guidance on assessment under the EU Air Quality directives the monitoring objectives are stated (see[28], 4.3.4.1 on p 41) to determine compliance with air quality limit values, to assess exposure addressing both the highest levels and levels representative of those to which the general population is exposed. The network design process considers 3 topics (see 4.3.4.2 of the guidance document):

- Station classification: given the fact that a monitoring network has only a limited number of stations, monitoring stations must in some way represent other pollution/exposure situations in the zone. This holds for numerous small scale situations, but also background concentration levels in the city. Hence, an initial step toward network design is station and area classification, in other words, defining traffic stations, industrial stations, background stations as well as defining an urban area, a suburban area or a rural area.
- Number of stations: the guidance as to how to determine the minimum number of sampling points are given in 2008/50/EC – Annex V, under A.1 . This differs by pollutant and depends *inter alia* on the total population of the agglomeration under consideration. The final network design also has to consider other aspects such balance of PM_{10} and $PM_{2.5}$ monitoring points, balance of site classification types amongst others.
- Location of stations: to meet the requirements, both the determination of the maximum 'hotspot' concentration in a zone as well as levels to which the general population is exposed are required. The former was mentioned in section 1.2 as one of the ambiguities in the current legislation and therefore additional guidance as to how to obtain this maximum concentration location would help evaluate more precisely whether or not a zone has surpassed any of the assessment thresholds (or the limit value itself).

The current guidance document mentions a step-wise process (see p 42) to network design. This process clearly involves both station classification as well as assessment of the spatial distribution of the concentration levels. A more quantitative approach to the distribution of monitoring stations within a zone may be formulated as an optimisation problem (see examples in Table 1) with an overall objective to find optimal locations for monitoring sites in order to:

- Reduce redundancy in the network, thus optimising investment and operational expenditure related to monitoring stations
- Fill in gaps in the coverage of the network in order to adequately sample the territory, various location types and population exposure.

Clearly, the configuration of a representative network also involves considerations other than adequately sampling the territory. Set up and operational costs also play a role and can be considered as part of the optimisation.

Even though the configuration of a representative network is a different application than determining the surface area, length of road or total resident population in the exceedance area, at its core the problem requires consideration of a spatial representativeness area around the stations. In this case, an actual area of similarity is required as we are looking at the total coverage of the territory by the network of monitoring stations. In fact, the optimisation problem described above can then be thought of simply as finding a layout for the air quality monitoring network which maximises the coverage made up by the sum of the SR areas of

[28] https://ec.europa.eu/environment/air/pdf/guidanceunderairquality.pdf

the individual network stations and at the same time minimises the overlap between the areas of representativeness amongst the stations in the network.

3.5.3 Context related definitions

As mentioned above, the concept of spatial representativeness may be interpreted here both as qualitative, referring to station classification, and as quantitative in relation to the assessment of a spatial representativeness area.

The design of a representative network does not need to be restricted to the metrics related to the limit and target values of the AAQD as listed under 3.2.1. This is why for this assessment need the definition of the similarity criterion should also consider temporal information.

3.5.4 Fitness of technical methods

Both station classification methods as well as methods to establish the spatial pattern of concentrations are required for this assessment need. Different classification methods were discussed at length above in Chapter 2 and section 2.2.3 in particular. A crucial point with these methods is their ability to separate between traffic / industrial and background stations. For methods establishing the spatial pattern, we can refer to the discussion under 3.7.4

3.6 Identify sampling points that are suitable for model calibration and validation

3.6.1 Legislative requirements

The purpose of this assessment need is to select monitoring sites which can be used for model calibration and validation. This may be extended to data assimilation as well, in which modelled estimates are combined with observations, taking into account their relative uncertainties and where it is equally important to have a good match between the scale of the model and the "scale" or representativeness of the observations.

Currently, there are on-going standardisation efforts with regard to model calibration and validation. The process originated from the FAIRMODE model quality objectives, and the DELTA Benchmarking tool[29]. This process has delivered a set of model quality indicators and objectives which may be used to asses air quality modelling applications when used in the context of reporting. The harmonisation of the methodologies to assess model performance developed in FAIRMODE is now part of a standardisation process in CEN Working Group 43 on Model Quality Objectives.

The current legislation states the minimum required number of stations for different assessment regimes (2008/50/EC Annex V – A.1) but it is clear that zones and the monitoring networks are not defined with validation or calibration in mind. For example, for agglomerations up to 1 million inhabitants, 3 monitoring stations are required for pollutants except PM if the maximum concentrations exceed the UAT, with the additional requirement to include at least one urban background station and one traffic station. The minimum requirements in the AQDD are therefore not able to yield a statistically significant sample to represent the spatial pattern generated due to traffic.

Finally, it should be mentioned that in the DELTA tool, there is no provision to explicitly account for the limited spatial representativeness[30] of monitoring stations. In the list of open issues, there is a clear need to quantify the impact of limited spatial representativeness on the measurement uncertainty which is taken into account in the construction of the model quality indicators.

[29] https://aqm.jrc.ec.europa.eu/index.aspx
[30] https://fairmode.jrc.ec.europa.eu/document/fairmode/WG1/Guidance_MQO_Bench_vs2.1.pdf

3.6.2 Set of metrics and characteristics required

For validation and calibration purposes, it is important that the spatio-temporal variability and hence the “scale” of the observations match those of the model and vice-versa.

Depending on the nature of the modelling technique, very different scales may be captured, going from the global or continental scale down to an obstacle resolved microscale level. It is not meaningful to compare for example NO_2 concentrations of a roadside traffic station exhibiting a high degree of variability with model results at a regional scale (typically obtained by chemical transport models at a resolution of a few km) as both observations capture very different aspects of the pollution. A validation exercise may in this particular case erroneously conclude that the air quality model is underestimating the concentrations at the traffic location, whereas the model was not intended to resolve this scale in the first place. Also, calibration or assimilation of a coarser scale model using such a road-side station may in fact overcompensate for the negative bias and introduce large overestimations elsewhere in the domain. The characteristics required are therefore primarily qualitative in nature, and dependent on the purpose and type of modelling being undertaken.

The purpose of validating a model can differ quite significantly. A validation exercise may want to investigate specific aspects of a model’s ability to represent the observed concentrations and as such may focus on temporal or spatial aspects, or more detailed aspects looking at the way in which a model is able to capture the distribution of observed values, and hence percentile values. For more detailed studies, validation exercises may aim at assessing certain aspects of the model, such as its ability to capture roadside gradients, or simulate downwash from industrial stacks in the recirculation zones behind large buildings. Determining whether or not sampling points are suited for model calibration and validation depends on the aim of the validation.

From the perspective of the assessment needs this can also be the way in which the similarity criterion is defined, or rather how temporal information or information about the spatial distribution is included in the definition of the metric. For example, for a model providing estimates of the length of road in exceedance of an environmental objective, traffic stations with appropriate representativeness are most suited to calibrate or validate such a model.

3.6.3 Context related definitions

The context related definition of the spatial representativeness area, and thus the way in which the similarity is established, will depend on the purpose and nature of the validation exercise. More specifically the definition of the metric should be driven by what information is required with regard to the statistical distribution of the observations to be represented in the validation or integrated into the model. This can be:

- The annual mean for a spatial validation on annual averages.
- Temporal aspects to check if the model and observations are representative of the same temporal variability. For example, it does not make sense to require models which use annual averaged daily traffic counts to be able to represent concentration values at particular hours. Such a match could be encoded by the way in which the similarity criterion is defined.

A model with a high spatial resolution can still be compared to a station with a large SR, as long as the length scale of the station is comparable or larger than the model length scale. However, in most cases such a model application will be of little use since the model will be able to resolve concentration variations which do not occur in practice (e.g. setup a CFD model for a rural area without any nearby sources).

3.6.4 Fitness of technical methods

Given the qualitative nature of the characteristics required, **classification methods** as discussed in Chapter 2 and section 2.2.3 can provide a good way of providing a station selection. For temporal aspects, the way of establishing similarity should clearly be temporally explicit and allow to discriminate between different levels of temporal variability.

Concerning the spatial aspects, it should be shown that the monitoring stations are representative for a similar spatial scale as the model under investigation. A way to test this can be the use of semi-variograms as illustrated in Figure 2. Though, this will not allow to select individual stations suited for calibration or validation. Here, again the concept of an area of spatial representativeness can be valuable as a way to:

- Understand differences between stations belonging to the same class as delivered by the classification method.
- Estimate a concrete area in which the concentrations only differ by the measurement uncertainty and assess how this scale matches with the resolution of the model which is being evaluated.

For example in the (Joly and Peuch, 2012) method, the classification is introduced on the basis of fixed percentile thresholds. However, though the approach works for separating rural and urban stations (with the aim of validation/calibration of regional scale models) it is not straightforward to have a clean separation for example between suburban and urban sites. Data assimilation procedures may be improved by selecting the monitoring sites that are representative of geographical areas related to the spatial resolution of the models.

3.7 Determine the spatial variability within the “area of representativeness”

3.7.1 Legislative requirements

This assessment need refers to the requirements under 2008/50/EC Annex III, section C, microscale siting of sampling points, where a number of requirements have to be met. Depending on the type of station, these include the following:

- The flow around the inlet sampling probe shall be unrestricted (free in an arc of at least 270°) without any obstructions affecting the airflow in the vicinity of the sampler (normally some metres away from buildings, balconies, trees and other obstacles and at least 0,5 m from the nearest building in the case of sampling points representing air quality at the building line)
- The inlet probe shall not be positioned in the immediate vicinity of sources in order to avoid the direct intake of emissions unmixed with ambient air,
- For all pollutants, traffic-orientated sampling probes shall be at least 25 m from the edge of major junctions and no more than 10 m from the kerbside.,

This particular assessment need requires some further thought in the sense that the spatial variability, once the area of representativeness is determined, is by definition known as it is given by the applied similarity criterion and its tolerance level. Especially when arguing that an approach attempting to capture the SR area of a monitoring site should aim to capture the full spatial/temporal variability(depending on the context), there should not be any remaining variability present larger than what is allowed by the similarity criterion.

To some extent this refers also to a discussion regarding whether or not SR areas should be allowed to be discontiguous or in contrast require strict contiguity. For example, the discontinuous nature of the urban fabric, with possible large differences between road-side and backyard or urban “background” concentrations will naturally introduce discontiguous zones. Therefore, a spatial representativeness area for an urban background location should probably naturally reflect this as it may otherwise only be representative up to the next street canyon, which goes against the notion of having an urban background concentration. Fit-for-purpose assessment methods should therefore be able to resolve this variability and exclude those areas (contiguous or not) which fall outside of the similarity criterion.

Given the inherent limitations of model based assessment of SR areas, this assessment need can be seen as a validation or quality check of the SR area determined. We can therefore interpret this assessment need from that perspective and look for ways to evaluate the residual variability within a spatial representativeness area established by the macroscale siting criteria.

3.7.2 Set of metrics and characteristics required

At the moment, there is no predefined or harmonized way for Member States to analyse this remaining microscale variability and how it should be quantified. Properties of a statistical distribution and/or timeseries of the concentrations or air quality metrics inside the SR area are required. These can be expressed as the variance or standard deviation, quantiles, mean, maximum, median or the statistical frequency distribution as a whole. A number of graphical means to represent this variability can be used, for example, box plots, histograms, timeseries plots, etc. In addition, providing concentration maps or maps or dashboards visualising the observed values can provide insight into the variability within the area of representativeness.

3.7.3 Context related definitions

As this particular assessment need starts from a predefined area of representativeness and aims to quantify the remaining variability, there is no need here to define additional specific SR metrics as for the previous assessment needs. The same metrics will hold when quantifying the variability within the spatial representativeness area.

3.7.4 Fitness of technical methods

Methodologies suited for this purpose will necessarily require a very detailed spatial resolution able to resolve effects of vegetation, screens and obstacles on the flow and the concentrations. Microscale CFD modelling as described in 2.2.2 is a suitable methodology for this. However, any model based approach is subject to uncertainties or an incomplete description, for example Vardoulakis et al.(2005) determined the spatial variability of air pollution in the vicinity of a very busy traffic station in Paris and found that relatively simple dispersion models failed to properly treat effects of differential street canyon height and urban vegetation.

Models remain approximations of reality and at present, there are gaps in which high resolution air quality modelling fails to account for the true spatial variability, for example:

- The spatial pattern of $PM_{10}/PM_{2.5}$ concentrations at local scale due to the lack of accurate emission inventories which properly account e.g. for residential heating or resuspension
- Significant lacks in traffic intensity data and fleet composition which is required for the application of road-side dispersion models.
- Dynamic traffic effects (structural traffic jams, stop and go in front of red lights) may not be adequately reflected by the model.

Therefore, the assessment of microscale variability may also benefit greatly from detailed monitoring campaigns around the site of interest, see

Table 4 for an overview of different approaches.

4 A tiered approach as a framework for guidance recommendations

4.1 Introduction

With the aim of formulating recommendations on how to provide guidance for the determination of spatial representativeness, we will introduce different tiers. These effectively present Member States with a roadmap towards better understanding of spatial representativeness and how to deal with some of the flexibility allowed in the AAQD. This tiered approach will necessarily be specific for the assessment needs as described in the previous chapter and attempt to address the different guidance needs listed under 1.2. The tiers are also designed to provide an increasing level of complexity to allow for varying modelling and monitoring capabilities and access to input data in different Member States.

The idea for a tiered approach was already given in the intercomparison exercise (O. Kracht et al., 2017), questioning whether or not "*it would be necessary and reasonable to define an order of preference for the selection of methods in an application*". The aim of this tiered approach is not merely an attempt to classify the models and/or methodologies to assess spatial representativeness. It is also an attempt to move forward the understanding of the issues by organising methods according to their complexity and aligning them with the different assessment needs. Much of the current confusion and discussion related to assessing spatial representativeness stems from an observation-focussed approach, starting at the monitoring stations and seeking to derive areas of representativeness for these measurements. Using an approach such as a fixed buffer radius (as some methodologies in the IE have documented) around a monitoring location, provides limited data to inform the development of an air quality plan. In the end, the ultimate aim for Member States should be to improve air quality at those locations where it is necessary and for this, understanding the spatio-temporal variability and concentration patterns in greater detail is beneficial.

It should also be clear that this tiered approach does not aim to give a specific quality assessment for individual methodologies or models. Some of the flexibility in the AAQD is intentionally there to allow member states to freely apply the tools and methodologies suited for their particular situation, as long as they are fit for purpose. It is therefore not a question of a specific methodology being "good" or "bad" but rather what we can learn from it for the different applications. As such the discussion in the previous chapter has helped us in specifying the limitations and capabilities of the different methodologies, thus allowing us to establish their fitness for each application or their "fitness-for-purpose".

We will see below, that this tiered approach succeeds in addressing the guidance needs listed in 1.2. Many of these needs are related to quantification and to distance from sources, both for macroscale siting criteria as well as for the microscale criteria. As there are different ways to accomplish this, it seems natural to introduce a system of different tiers related to the comprehensiveness of establishing spatial representativeness and the way in which different emission sources, dispersion conditions and transport phenomena are quantified. As discussed in different locations in the previous chapters, such a hierarchy, though perhaps natural for ways of establishing the spatio-temporal variability, can also be applied to classification methods as we shall see below.

Several additional aspects were considered when drafting this tiered approach :

1. It is important to take Member States capabilities into account. Not every Member State possesses the capacity or resources to provide comprehensive high-resolution air quality assessments. It is the intention that the accuracy of the methods increases with

each tier and therefore MS are encouraged to use as higher a tier as possible in their assessment of spatial representativeness.

2. The tiers should reflect progressively more elaborate data requirements, going from basic to a comprehensive set of input data needs. As above, MS are encouraged to use as higher a tier as possible with regard to the data available to them but that all tiers are deemed as compliant options for assessment.
3. The different tier levels will progressively add detail and accuracy in the assessment of the spatial representativeness area, so that the more advanced tier methods will build on the previous ones.
4. The ultimate aim is a full characterisation of the complete spatio-temporal variability of the concentration field. Once this is known, the derivation of the different assessment needs becomes straightforward. Already in (Spangl et al., 2007) it was indicated that modelling would be the optimal method for determining representativeness. However, a tier based on comprehensive modelling can never be the ultimate tier level as one has to recognise limitations in the models and their input data.
5. At the same time, it should be recognised that monitoring in itself, though not capable of capturing the full, geographically explicit spatio-temporal variability, is an effective means of informing an expert opinion in a Tier 1 approach. However, as stated in the introduction, it should be clear that only so much can be learned from such an approach.
6. The feasibility of the tiered approach for application by Member States is to be tested through a series of sensitivity studies. These sensitivity studies will investigate the accuracy of the spatial representativeness assessment for the different tiers and their applicability in different areas.

By using such a tiered approach for each assessment need, the recommendations will recognise limitations in Member State resources for performing complex air quality model simulations or implementing advanced statistical analysis routines by promoting the uptake of easily accessible, appropriate techniques. As stated in the IE report, there is a growing availability of high-resolution air quality modelling tools and expertise that can be widely applied across Europe. It is important also to draw on experience in applying certain methods by Member States.

4.2 Defining the tier levels

Four tier levels have been defined for different assessment needs and guided by the following principles:

- **Tier 1** is based on the characterisation of the monitoring site. It includes an expert opinion that provides a qualitative assessment of spatial representativeness made on the basis of local knowledge of the monitoring site and relatively simple “distance to source” considerations. Such a qualitative assessment is sometimes complemented with additional in-situ or mobile monitoring either of air quality or meteorology condition to better understand the spatial representativeness.
- **Tier 2** will add source and dispersion related information into the assessment of spatial representativeness of a monitoring site. The method in Tier 2 will be based on the combination of monitoring site characterisation with proxy data, which can be geographical via GIS data or temporal via time series analysis to determine the variability of AQ concentrations in the area surrounding the monitoring site, without using AQ dispersion modelling.
- **Tier 3** is based on the use of air quality dispersion modelling to link information on sources and dispersion conditions around the monitoring site adding possibly also

information of long-range transport influences to the site. The use of air quality modelling provides comprehensive fit-for-purpose and geographically explicit information, adding to the accuracy of the spatio-temporal variation of the concentration field near a monitoring site. Tier 3 provides then an explicit spatial representativeness area based on modelling information, with the recognised limitations that models and their input data may have.

- **Tier 4** represents a more comprehensive approach as it combines and complements the modelling information with additional detailed observations around the monitoring site. This method helps improve the accuracy of the results reducing the uncertainty inherent to any modelling application with the help of additional observations. Typically, the additional observations are collected via dedicated monitoring campaigns with a high sampling density, allowing to fully characterise (within the limits of the observational uncertainty) the spatio-temporal variability. Tier 4 results in the most accurate characterisation of spatial representativeness in the current guidance.

4.3 Methodology classification

In this section we propose mapping of different methodologies to the different tier levels for each assessment need. We build on the analysis and evaluation presented in previous chapters and classify in Table 6 the methodologies per assessment need in the different tier levels.

Given the difficulty in interpreting the last assessment need (see 3.7) regarding the estimation of the spatial variability inside the area of representativeness, it has been omitted from this tier classification but is referred to in more detail in the discussion in the next sections.

The assessment needs for the estimation of surface area in exceedance, estimation of length of road in exceedance and the estimation of total resident population in the area of exceedance should be considered independently of one another as there are some subtle differences which need to be accounted for, making a single analytical method unsuitable, for example whether or not street canyon concentrations have to be taken into account in the estimation of resident population exposure. Also, there will be different requirements and recommendations for models able to generate a geographically explicit area or (parametric) models which can determine roadside or street canyon concentrations.

Table 6. : SR assessment methods in different tiers per assessment need.

<table>
<tr><th></th><th>Estimation of surface area in exceedance</th><th>Estimation of total resident population in area of exceedance</th><th>Estimation of length of road in exceedance</th><th>Facilitation of configuration of representative network</th><th>Identify sampling points suitable for calibration and validation</th></tr>
<tr><td>Tier 1
Expert Opinion</td><td colspan="2">Fixed radius e.g. (Castell-Balaguer and Denby, 2012)</td><td>Fixed length</td><td>Classification based on expert opinion and station classification</td><td>Expert assignment of station siting and type</td></tr>
<tr><td>Tier 2
Proxy Information</td><td colspan="3">Methods relying on proxy data and distance relations to estimate source emissions and dispersion conditions. E.g. (Henne et al., 2010; Janssen et al., 2012; Righini et al., 2014; Spangl et al., 2007)</td><td colspan="2">Objective station classification based on time series or GIS proxy data (Joly and Peuch, 2012; Nguyen et al., 2009)</td></tr>
<tr><td rowspan="2">Tier 3
Geographically explicit, comprehensive fit-for-purpose modelling</td><td colspan="3">Comprehensive and fit-for-purpose local scale modelling: line source modelling, parametric street box models (OSPM, CAR, …), obstacle resolved modelling (CFD), (Rivas et al., 2019; Santiago et al., 2013)</td><td rowspan="2">Determine gaps in the network coverage taking into account the SR areas of the stations, e.g. (Soares et al., 2018)</td><td rowspan="2">Geographically explicit models applied for objective classification. (typical SR length scale based on independent modelling)</td></tr>
<tr><td colspan="2">Comprehensive and fit-for-purpose regional scale modelling: regional scale Eulerian models e.g. (Martin et al., 2014)</td><td></td></tr>
<tr><td>Tier 4
Modelling complemented with dedicated measurements</td><td colspan="5">Modelling complemented with passive sampler campaigns, mobile monitoring, e.g. (Hagenbjörk et al., 2017; Li et al., 2019; Vardoulakis et al., 2011b, 2005). In the future sensor observations (Sadighi et al., 2018) might be used as well if sensor uncertainty is properly defined.</td></tr>
</table>

For the estimation of the area, population and road length exceeding, the tiered approach builds upon an assessment of the SR area which gradually increases in complexity. Starting with expert opinion related to station classification (Tier 1), to GIS based proxy data and distance rules (Tier 2) towards a modelling methodology (Tier 3) which, in principle, comprises our state-of-the-art understanding of source information and dispersion characteristics. It is clear that for this Tier 3 approach, the fitness for purpose of the modelling tools is a very essential and critical condition to be met. As mentioned before, the modelling application is in this context considered as the “best possible” understanding and description of the air quality concentration patterns. However, for the time being, it is known that no model is perfect and always comes with a certain level of uncertainty. In the final approach (Tier 4), this modelling information is complemented with detailed measurement data to account for this uncertainty and any imperfection in the air quality models and its input data.

To support design of a representative network and identification of sampling points suitable for calibration and validation, we have identified methods (beyond Tier 1) that produce a classification based on timeseries analysis or proxies. Such classification methods in fact provide a **qualitative** estimate of spatial representativeness, in relation to a label or classification for the station such as “urban background”, “traffic”, “regional background”. The nature of the classification is governed by the type of methodology applied. This indicates that configuration of a representative network would benefit from having such a geographically explicit area of representativeness for each station. Based on this reasoning, methodologies which complement the classification with a geographical area are classified as Tier 3 (for example (Piersanti et al., 2015; Soares et al., 2018).

The application of “modelling” in Tier 3 to identify sampling points suitable for calibration/validation of “models” may be inappropriate due to circularity of the process. Models are only limited representations of reality and are dependent on the limited input information. It is possible to combine different spatial scale modelling techniques to identify suitable locations for different station classifications. The use of very high resolution models becomes very interesting as they account for the full variability in the urban environment (Santiago et al., 2013). Likewise, the use of local scale dispersion models, applied to a larger domain allow the identification of locations which are unaffected by local sources, and can therefore be used to validate/calibrate more coarse scale model such a chemical transport model.

Caution should be exercised in the application of these techniques, underlining the importance of scientific expertise and understanding. Due consideration should be given to:

- How the background concentrations are derived and represented within a more detailed modelling approach (and in what way they are related to the lower resolution model under scrutiny).
- To what extent the more detailed model application accurately reflects the spatial variability of the concentrations.
- Whether the chosen model is sufficiently fit for purpose e.g. the nature of relevant sources is being appropriately represented with time varying emissions profiles and met data.

For all assessment needs, the Tier 4 is a comprehensive and complementary suite of both modelling and monitoring approaches, recognising the benefits of both forms of evidence. Modelling will intrinsically always be limited by an imperfect description of reality, while measurements – when not accounting for their spatial representativeness - will always be limited to representing a specific location at a specific point in time. We refer here for example to (Vardoulakis et al., 2005) who have used very detailed monitoring and modelling around a specific traffic location in Paris to assess its representativeness and research a more representative location.

4.4 Addressing the guidance needs

The added value of the tiered approach should become clear in the context of improving guidance as discussed in Section 1.2. The Table 7 below provides an overview of how.

Table 7: Correspondence table indicating how the tiered approach will allow to address the different needs listed under Section 1.2.

Id (§1.2)	Short description of the guidance need with respect to SR	Addressing by proposed tiered approach
1, 2	Lack of definition of exposure and unclear how to interpret what “representative” in this context refers to	As such this guidance need refers to the difference between static and dynamic exposure and is not a subject of this overview.
3	Assignment of population to areas in exceedance in presence of discontinuous canyons	For this particular need, it is not so much the categorisation of the methods in a tiered approach which will aid in resolving this ambiguity, but rather the aspiration mentioned under 4.1 to aim for a full categorisation of the spatio-temporal variability of the pollutant concentrations, resolving explicitly the differences between backyard and front-side concentrations. Sensitivity studies will have to be performed to provide insight into differences between the approaches (see Chapter 5)
4	Better insight in what spatial variability is allowed for by the scales of 100 m for traffic-oriented and 250 m for industrial sites.	Length scales can be explicitly dealt with via dispersion or microscale modelling as it paints an explicit picture of the concentration patterns, i.e. road-side gradients are explicitly accounted for, and street canyons resolved. In addition, the flexibility allowed in statements such as “in the immediate vicinity” (e.g. under AAQD Annex III, C) can be addressed explicitly. Therefore, an adoption of higher tier methods, will aid in quantifying the distance and proximity relations. When begin complemented with additional sampling campaigns in a highest tier method, this will further aid in quantifying possible effects which are not captured by the modelling such as variation in traffic emissions along the streets or erratic emission patterns in industrial area’s which are not captured by the reported emissions from industry.
5	Regarding how to define urban background stations as not being “dominated” by a single source	Comprehensive air quality modelling allows to quantify individual source contributions explicitly via scenario assessment. Requirements for example such as paragraph C under the AAQD Annex III – B.1. stating that urban background locations should not be “dominated by a single source” can therefore be explicitly modelled and the contribution of such sources quantified, as opposed to having an expert opinion or using proxy GIS data to represent the distance relations. In addition, it is mentioned in the literature that in an urban setting, the distinction between urban background, traffic and industrial stations is not always straightforward and cannot always be uniquely defined by an objective classification method. The way in which such methods are integrated in the tiered system as discussed above, adding the concept of an explicit spatial representativeness area to better discriminate different station types can therefore be a way to better understand these classification issues.

		Also here, additional monitoring in Tier 4 can aid in capturing erratic effects, not well represented by the models.
6	Unclear what “immediate vicinity” and “some meters” away from means in the microscale criteria	Similar considerations hold in terms of improved quantification, starting from current practice rules of thumb to explicit quantification of the flow around the inlet, taking into account very localised effects for example of vegetation or flow obstacles.
7	Compatibility of microscale and macroscale requirements for traffic stations	Again, here an explicit resolving of roadside gradients via high resolution modelling, potentially complemented by dedicated monitoring campaigns will improve the understanding of whether both the microscale (not more than 10 m away from the road) and the macroscale (representative for 100m) is fulfilled w.r.t. expert judgement or modelling via proxy data which not always adequately captures the gradients.
8	Determination of maximum concentrations within air quality zone	Going from expert considerations of where we expect the highest concentration, to a set of distance relationships using GIS data, to a full comprehensive modelling approach which will show geographically explicitly where the maximum is expected. Also, this follows nicely the approach presented in 4.3.
9	Compatibility between minimum number of stations required in assessment requirements w.r.t. LAT / UAT.	Having an explicit way to estimate the spatial variability around each station on top of a station classification (as the first step in configuration of a representative monitoring network, see Section 3.5) and as proposed in the higher tier methods for this assessment need, will help to determine locations that are interesting to sample in a more dedicated validation campaign or a fixed monitoring network.

It should also be mentioned that there is support from literature for this frame of thought. Duyzer et al.,(2015) for example present an interesting perspective regarding network design. Station siting issues are subject to continuous debate. In addition to considerations such as the ones above, further practical implementation issues on the ground may need to be considered. For instance, it may not be very practical to move monitoring stations and thereby interrupt a continuous time series. The authors recommend however the use of model calculations to compensate for microscale siting issues related to strong road-side concentration gradients of street canyon sampling locations.

From Table 7 it becomes clear that the tiered approach helps in addressing some of the guidance needs. However, to consolidate this and better inform the guidance recommendations for each tier level that we will formulate in the later stage in this project, a number of sensitivity studies are proposed. These will build upon further work undertaken within the FAIRMODE intercomparison exercise and also address some specific issues raised in this review. In the next chapter, we present a brief outline of the sensitivity studies planned.

5 Proposed sensitivity studies

In order to concretise guidance in the different tier levels, a number of sensitivity studies are proposed which will be further elaborated in the next phases of this project. These studies will partly follow the recommendations formulated by Oliver Kracht et al. (2017) and will serve a number of different purposes :

- informing lower tier approaches;
- informing guidance on fitness for purpose of the approaches in the tiers;
- addressing specific issues raised in this document, e.g. regarding exposure assignment or spatial variability in the presence of street canyons

These will be further discussed in the sections below.

5.1 Informing lower tier approaches

The purpose of these sensitivity studies is to better understand to what extent a lower tier approach differs from more advanced approaches and as such better constrain what we can expect each tier level to deliver. In addition, here we will pick up some of the points discussed in the IE. Studies which will comprise of :

- Comparing the spatial representativeness areas as delineated by a simple Tier 1 approach (fixed buffer size) to higher tier approaches using data obtained from the cities for Krakow, Antwerp and Oslo.
- It will be informative to establish, starting from the detailed air quality maps and the metrics derived from them in the context of compliance checking (area in exceedance, total length of road in exceedance, population exposed), what “fixed” buffer sizes (in a Tier 1 approach) will yield comparable values as well as how the specific geometrical assumptions on the buffer (e.g. contiguity) influence it’s estimation. These studies can easily be undertaken if detailed high resolution air quality modelling results are available as for Antwerp, Krakow and Oslo, including estimations of street canyon effects.

There is also significant added value to further elaborate and understand different ways to characterise the similarity criterion discussed in Section 2. How does the parameterisation of the similarity criteria and threshold values influence the estimation of SR areas? Blanchard et al. (1999) analysed sensitivity by changing the criteria of concentration similarity of PM_{10} from 20% to 10% and the spatial representativeness area was reduced about half of those obtained with the 20% criteria. Pay et al., (2014) carried out a test of the sensitivity of the threshold (5, 10, 15, 20%) for several pollutants to maxima discrepancy concluding that 20% for all the pollutants could be a conservative selection. Again, based upon existing high-resolution modelling results for Krakow, Antwerp and Oslo for NO_2, $PM_{2.5}$, O_3 and PM_{10}, such studies can be performed. In these studies the following will be considered :

- How urban structure influences the SR area (by comparing different cities) and to what extent recommendations can be generalised.
- In what way SR areas are influenced by requiring spatial contiguity and/or exclusivity. (i.e. whether or not SR areas around neighbouring stations are allowed to overlap or not)

In relation to this final point, we will consider the possibility of including additional functionality in the FAIRMODE composite mapping viewer which will empower member states to perform such sensitivity analyses as well (see Figure 8). Providing tools to delineate areas of representativeness based on the available air quality maps in the composite mapping viewer may in fact also be an effective way of providing additional guidance. This is illustrated below.

Such functionality would be out of scope in this project but may prove a useful addition in the future.

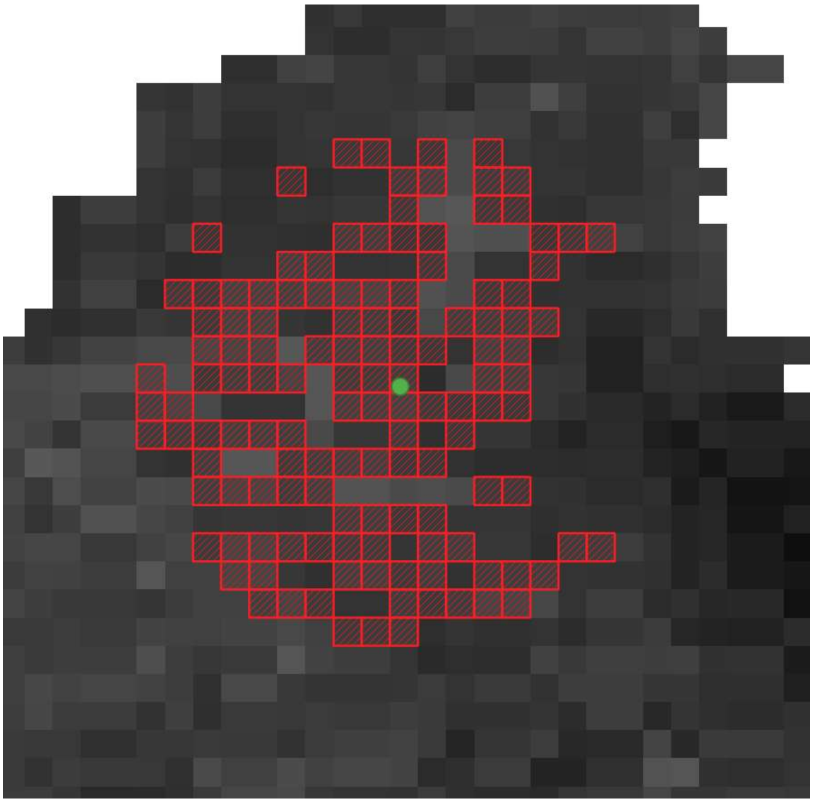

Figure 8 : Illustration of possible additional functionality in the FAIRMODE composite mapping viewer. A user would be allowed to click on a given location and an area would be highlighted for which the concentrations in the maps selected only deviate by a given percentage/absolute value (configurable).

5.2 Informing fitness for purpose of the approaches in the tiers

When adopting air quality modelling as a Tier 3 approach, there should be a framework for assessing a model's fitness for purpose, and guidance will have to be developed for that. A potential methodology was already discussed to evaluate and quantify the fitness for establishing the spatial variability as required by the assessment needs. We refer here to the discussion on semi-variograms, see Figure 2. Using this as a starting point, the following should be discussed/researched further in sensitivity studies :

- How/if, this can be generalised ? Are observed semi-variograms similar? How do they differ between scales? Can guidelines be formulated for their use in establishing fitness for purpose ?
- What are the limitations of the use of semi-variograms? While they are certainly an interesting way to assess whether a model is able to capture the spatial variability, their interpretation at very short distances can be ambiguous given the discontinuity of the concentration fields in the street canyons (see also Figure 7).

This is certainly an area where the further sensitivity studies can be valuable. However, they will be limited in scope due to the limited amount of high density monitoring/modelling data available. The recent CurieuzeNeuzen campaign data with 20000 passive sampling points is suitable for a sensitivity study, but this study will be limited to Flanders and NO_2 long term averaged concentrations.

5.3 Addressing specific issues

Furthermore, a number of more detailed issues will have to be addressed in relation to the guidance needs listed and discussed in the first chapter:

1. When it comes to determining the number of people exposed to concentrations in excess of a limit value, in particular for NO_2, discussion and/or sensitivity studies and ultimately, guidance is needed on how to assign population in buildings to discontinuous concentration assessments, explicitly accounting for street canyon effects.
    - Detailed high resolution maps will be used for the cities of Antwerp and Krakow, comparing the number of people exposed at address level with different methodologies of assigning, such as street side vs. backyard, average or some more advanced methodology such as the SBE method discussed in (Diegmann and Pfäfflin, 2016), see discussion under 2.2.2..
    - The extent to which differences occur will also depend on the differences in modelling approach.
2. Sensitivity studies on the compatibility of the microscale and macroscale requirements for the siting of traffic stations. Here, in principle it is possible to:
    - Use existing 3D microscale simulations, or output of parametrised models such as an Operational Street Pollution Model (OSPM) and analyse the concentration gradients that result from different characteristics of the urban structure and built environment (such as ventilation openings[31] or the lack thereof)
    - Analyse results of mobile monitoring campaigns such as “Meet Mee Mechelen[32]”. which can help understanding the extent over which concentration levels vary along the streets and also account for changing emissions along the streets. However, such data is scarce and does not necessarily cover regulated pollutants (as opposed e.g. to Black Carbon, which is fairly easily measured with portable monitoring equipment).
    - Given the possible scope of such sensitivity studies and the available data, only a qualitative discussion will be possible based on existing data/model results

[31] In other words, openings in continuous building facades through which “fresh” air flow may enter in the street canyon.

[32] https://mechelen.meetmee.be/kaart

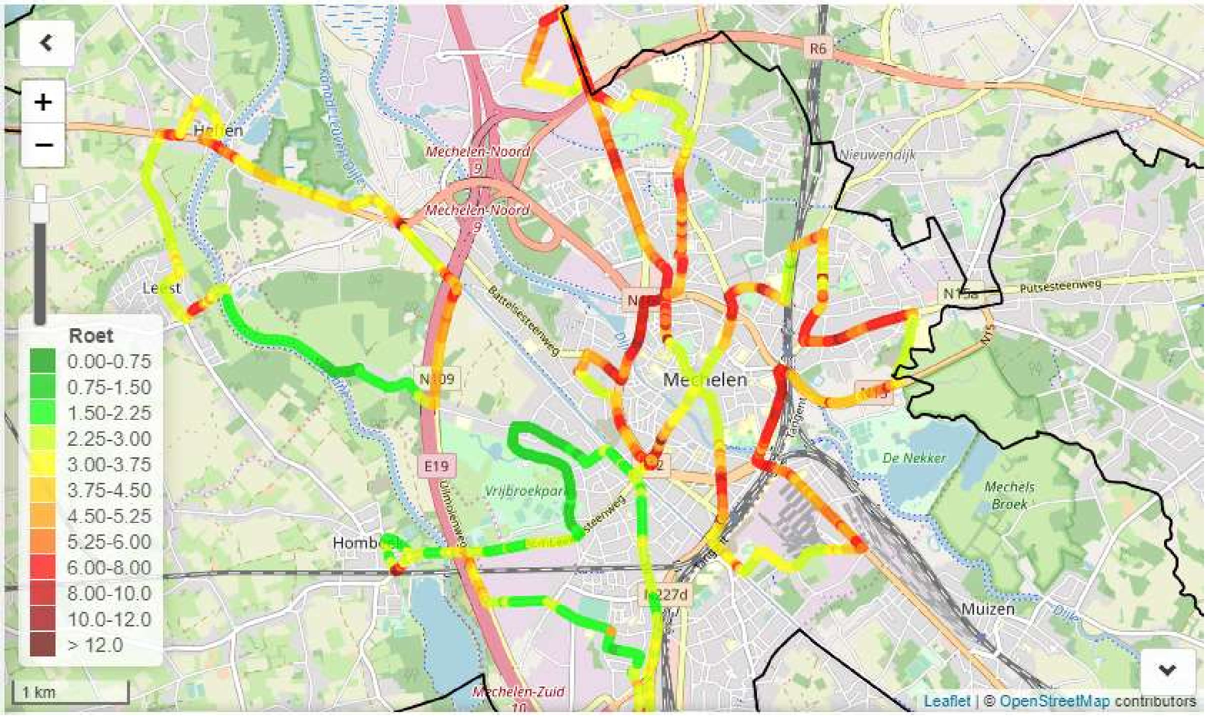

Figure 9 : Example of mobile transects of Black Carbon (BC) as measured in the "Meet mee mechelen" campaign in Flanders, Belgium. The transects give an impression of the spatial variability along street segments.

6 References

Badura, M., Batog, P., Drzeniecka-Osiadacz, A., Modzel, P., 2018. Evaluation of Low-Cost Sensors for Ambient PM 2.5 Monitoring . J. Sensors 2018, 1–16. https://doi.org/10.1155/2018/5096540

Barrero, M.A., Orza, J.A.G., Cabello, M., Cantón, L., 2015. Categorisation of air quality monitoring stations by evaluation of PM<inf>10</inf> variability. Sci. Total Environ. 524–525, 225–236. https://doi.org/10.1016/j.scitotenv.2015.03.138

Beauchamp, M., Malherbe, L., de Fouquet, C., Létinois, L., 2018. A necessary distinction between spatial representativeness of an air quality monitoring station and the delimitation of exceedance areas. Environ. Monit. Assess. 190, 441. https://doi.org/10.1007/s10661-018-6788-y

Beauchamp, M., Malherbe, L., Létinois, L., 2011. Application de méthodes géostatistiques pour la détermination de zones de représentativité en concentration et la cartographie des dépassements de seuils.

Berkowicz, R., 2000. OSPM - A Parameterised Street Pollution Model. Environ. Monit. Assess. 65, 323–331. https://doi.org/10.1023/A:1006448321977

Blanchard, C.L., Carr, E.L., Collins, J.F., Smith, T.B., Lehrman, D.E., Michaels, H.M., 1999. Spatial representativeness and scales of transport during the 1995 integrated monitoring study in California ' s San Joaquin Valley. Atmos. Environ. 33, 4775–4786.

Bobbia, M., Cori, A., Fouquet, C. De, 2008. Représentativité spatiale d'une station de mesure de la pollution atmosphérique Spatial representativeness of an air pollution measurement. Pollut. Atmosphérique 197, 63–75.

Castell-Balaguer, N., Denby, B., 2012. A survey to elicit expert opinion on the spatial representativeness of ground based monitoring data.

Castell, N., Dauge, F.R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., Bartonova, A., 2017a. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environ. Int. 99, 293–302. https://doi.org/10.1016/j.envint.2016.12.007

Castell, N., Dauge, F.R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., Bartonova, A., 2017b. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environ. Int. 99, 293–302. https://doi.org/10.1016/j.envint.2016.12.007

Cosemans, G., Dumont, G., Roekens, E., Kretzschmar, J.G., 1997. Optimal siting of air quality monitoring stations around five oil refineries. Trans. Ecol. Environ. 13, 381–390.

Denby, B., 2010. Spatially distributed source contributions for health studies: comparison of dispersion and land use regression models. - TRANSPHORM Deliverable D2.2.4.

DG-Environment, 2018. Member States' and European Commission's Common Understanding of the Commission Implementing Decision laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of informati.

Diegmann, V., Pfäfflin, F., 2016. Quantifizierung der verkehrsbezogenen populationsgewichteten Feinstaubexposition 2013 für die Modellregionen Berlin und Brandenburg.

Diegmann, V., Pfäfflin, F., Breitenbach, Y., Lutz, M., Rauterberg-Wulff, A., Reichenbächer, W., Kohlen, R., 2015. Spatial representativeness of Air Quality Samples at Hot Spots, in: HARMO 15, 15th Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes. Madrid.

Duyzer, J., van den Hout, D., Zandveld, P., van Ratingen, S., 2015. Representativeness of air quality monitoring networks. Atmos. Environ. 104, 88–101. https://doi.org/10.1016/j.atmosenv.2014.12.067

Ellerman, T., Nygaard, J., Nøjgaard, J.K., Nordstrøm, C., Brandt, J., Christensen, J., Ketzel, M., Massling, A., Bossi, R., Jensen, S.S., 2018. The Danish Air Quality Monitoring Programme.

Annual Summary for 2007, Scientific Report from DCE – Danish Centre for Environment and Energy.

Geiger, J., Malherbe, L., Mathe, F., Ross-Jones, M., Sjoberg, K., Spangl, W., Stacey, B., González, A., Frank De Leeuw, O., Borowiak, A., Galmarini, S., Gerboles, M., De Saeger, E., 2013. Assessment on siting criteria, classification and representativeness of air quality monitoring stations.

Gillespie, J., Masey, N., Heal, M.R., Hamilton, S., Beverland, I.J., 2017. Estimation of spatial patterns of urban air pollution over a 4-week period from repeated 5-min measurements. Atmos. Environ. 150, 295–302. https://doi.org/10.1016/j.atmosenv.2016.11.035

Gupta, S., Pebesma, E., Mateu, J., Degbelo, A., 2018. Air quality monitoring network design optimisation for robust land use regression models. Sustain. 10, 1–27. https://doi.org/10.3390/su10051442

Hagenbjörk, A., Malmqvist, E., Mattisson, K., Sommar, N.J., Modig, L., 2017. The spatial variation of O3,NO,NO2and NOx and the relation between them in two Swedish cities. Environ. Monit. Assess. 189. https://doi.org/10.1007/s10661-017-5872-z

Hao, Y., Xie, S., 2018. Optimal redistribution of an urban air quality monitoring network using atmospheric dispersion model and genetic algorithm. Atmos. Environ. 177, 222–233. https://doi.org/10.1016/j.atmosenv.2018.01.011

Henne, S., Brunner, D., Folini, D., Solberg, S., Klausen, J., Buchmann, B., 2010. Assessment of parameters describing representativeness of air quality in-situ measurement sites. Atmos. Chem. Phys. 10, 3561–3581. https://doi.org/10.5194/acp-10-3561-2010

Janssen, S., Dumont, G., Fierens, F., Deutsch, F., Maiheu, B., Celis, D., Trimpeneers, E., Mensink, C., 2012. Land use to characterize spatial representativeness of air quality monitoring stations and its relevance for model validation. Atmos. Environ. 59, 492–500. https://doi.org/10.1016/j.atmosenv.2012.05.028

Joly, M., Peuch, V., 2012. Objective classification of air quality monitoring sites over Europe. Atmos. Environ. 47, 111–123. https://doi.org/10.1016/j.atmosenv.2011.11.025

Joris, V.D.B., 2016. Towards high spatial resolution air quality mapping : a methodology to assess street-level exposure based on mobile monitoring.

Karabelas, A., Sarigiannis, D., 2008. Optimal air pollution monitoring network configuration, LIFE05 ENV/GR/000214 Deliverable 6.1. Thessaloniki.

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., Gerboles, M., 2017. First outcomes of the fairmode & aquila intercomparison exercise on spatial representativeness, in: HARMO 2017 - 18th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Proceedings.

Kracht, Oliver, Santiago, J.L., Martin, F., Piersanti, A., Cremona, G., Vitali, L., Delaney, K., Basu, B., Ghosh, B., Spangl, W., 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., Gerboles, M., 2017. Spatial Representativeness of Air Quality Monitoring Sites.

Levy, J.I., Hanna, S.R., 2011. Spatial and temporal variability in urban fine particulate matter concentrations. Environ. Pollut. 159, 2009–2015. https://doi.org/10.1016/j.envpol.2010.11.013

Li, H.Z., Gu, P., Ye, Q., Zimmerman, N., Robinson, E.S., Subramanian, R., Apte, J.S., Robinson, A.L., Presto, A.A., 2019. Spatially dense air pollutant sampling: Implications of spatial variability on the representativeness of stationary air pollutant monitors. Atmos. Environ. X 2, 100012. https://doi.org/10.1016/j.aeaoa.2019.100012

Lozano, A., Usero, J., Vanderlinden, E., Raez, J., Contreras, J., Navarrete, B., El Bakouri, H., 2009. Design of air quality monitoring networks and its application to NO2 and O3 in Cordova, Spain. Microchem. J. 93, 211–219. https://doi.org/10.1016/j.microc.2009.07.007

Ma, M., Chen, Y., Ding, F., Pu, Z., Liang, X., 2019. The Representativeness of Air Quality Monitoring Sites in the Urban Areas of a Mountainous City. J. Meteorol. Res. 33, 236–250. https://doi.org/10.1007/s13351-019-8145-7

Maiheu, B., Lefebvre, W., Walton, H., Dajnak, D., Janssen, S., Williams, M., Blyth, L., Beevers, S., 2017. Improved Methodologies for NO2 Exposure Assessment in the EU (No. Report nr. 2017/RMA/R/1250), Study accomplished under the authority of the European Commission, DG-ENV Service Contract 070201/2015/SER/717473/C.3.

Martin, F., Fileni, L., Palomino, I., Vivanco, M.G., Garrido, J.L., 2014. Analysis of the spatial representativeness of rural background monitoring stations in Spain. Atmos. Pollut. Res. 5, 779–788. https://doi.org/10.5094/apr.2014.087

Martín, F., Santiago, J.L., Kracht, O., García, L., Michel, G., 2015. FAIRMODE Spatial representativeness feasibility study.

Nagl, C., Spangl, W., Buxbaum, I., 2019. Sampling points for air quality Representativeness and comparability of measurement in accordance with Directive 2008/50/EC on ambient air quality and cleaner air for Europe.

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. Bull. Am. Meteorol. Soc. 63, 761–764.

Nguyen, P.L., Hoogerbrugge, R., Arkel, F. Van, 2009. Evaluation of the representativeness of the Dutch national Air Quality Monitoring Network.

Ott, D.K., Kumar, N., Peters, T.M., 2008. Passive sampling to capture spatial variability in PM10-2.5. Atmos. Environ. 42, 746–756. https://doi.org/10.1016/j.atmosenv.2007.09.058

Peters, J., Theunis, J., Van Poppel, M., Berghmans, P., 2013. Monitoring PM10 and ultrafine particles in urban environments using mobile measurements. Aerosol Air Qual. Res. 13, 509–522. https://doi.org/10.4209/aaqr.2012.06.0152

Peters, J., Van den Bossche, J., Reggente, M., Van Poppel, M., De Baets, B., Theunis, J., 2014. Cyclist exposure to UFP and BC on urban routes in Antwerp, Belgium. Atmos. Environ. 92, 31–43. https://doi.org/10.1016/j.atmosenv.2014.03.039

Piersanti, A., Vitali, L., Righini, G., Cremona, G., Ciancarella, L., 2015. Spatial representativeness of air quality monitoring stations: A grid model based approach. Atmos. Pollut. Res. 6, 953–960. https://doi.org/10.1016/j.apr.2015.04.005

Righini, G., Cappelletti, A., Ciucci, A., Cremona, G., Piersanti, A., Vitali, L., Ciancarella, L., 2014. GIS based assessment of the spatial representativeness of air quality monitoring stations using pollutant emissions data. Atmos. Environ. 97, 121–129. https://doi.org/10.1016/j.atmosenv.2014.08.015

Rivas, E., Santiago, J.L., Lechón, Y., Martín, F., Ariño, A., Pons, J.J., Santamaría, J.M., 2019. CFD modelling of air quality in Pamplona City (Spain): Assessment, stations spatial representativeness and health impacts valuation. Sci. Total Environ. 649, 1362–1380. https://doi.org/10.1016/j.scitotenv.2018.08.315

Rodriguez, D., Valari, M., Payan, S., Eymard, L., 2019. On the spatial representativeness of NO X and PM 10 monitoring-sites in Paris, France. Atmos. Environ. X 1, 100010. https://doi.org/10.1016/j.aeaoa.2019.100010

Sadighi, K., Coffey, E., Polidori, A., Feenstra, B., Lv, Q., Henze, D.K., Hannigan, M., 2018. Intra-urban spatial variability of surface ozone in Riverside, CA: Viability and validation of low-cost sensors. Atmos. Meas. Tech. 11, 1777–1792. https://doi.org/10.5194/amt-11-1777-2018

Santiago, J.L., Martín, F., Martilli, A., 2013. A computational fluid dynamic modelling approach to assess the representativeness of urban monitoring stations. Sci. Total Environ. 454–455, 61–72. https://doi.org/10.1016/j.scitotenv.2013.02.068

Sarigiannis, D.A., Saisana, M., 2008. Multi-objective optimization of air quality monitoring. Environ.

Monit. Assess. 136, 87–99. https://doi.org/10.1007/s10661-007-9725-z

Sarigiannis, D.A., Saisana, M., 2007. Multi-objective optimization of air quality monitoring. Environ. Monit. Assess. 136, 87–99. https://doi.org/10.1007/s10661-007-9725-z

Scaperdas, A., Colvile, R.N., 1999. Assessing the representativeness of monitoring data from an urban intersection site in central London, UK. Atmos. Environ. 33, 661–674. https://doi.org/10.1016/S1352-2310(98)00096-X

Soares, J., Makar, P.A., Aklilu, Y., Akingunola, A., 2018. The use of hierarchical clustering for the design of optimized monitoring networks. Atmos. Chem. Phys. 18, 6543–6566. https://doi.org/10.5194/acp-18-6543-2018

Spangl, W., Schneider, J., Moosmann, L., Nagl, C., 2007. Representativeness and classification of air quality monitoring stations. Umweltbundesamt Rep.

Spinelle, L., Gerboles, M., Villani, M.G., Aleixandre, M., Bonavitacola, F., 2015. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. Sensors Actuators, B Chem. 215, 249–257. https://doi.org/10.1016/j.snb.2015.03.031

Tapia, O., Escudero, M., Lozano, A., Anzano, J., Mantilla, E., 2016. New classification scheme for ozone monitoring stations based on frequency distribution of hourly data. Sci. Total Environ. 544, 1–9. https://doi.org/10.1016/j.scitotenv.2015.11.081

Van den Bossche, J., Peters, J., Verwaeren, J., Botteldooren, D., Theunis, J., De Baets, B., 2015. Mobile monitoring for mapping spatial variation in urban air quality: Development and validation of a methodology based on an extensive dataset. Atmos. Environ. 105, 148–161. https://doi.org/10.1016/j.atmosenv.2015.01.017

Van Poppel, M., Peters, J., Bleux, N., 2013. Methodology for setup and data processing of mobile air quality measurements to assess the spatial variability of concentrations in urban environments. Environ. Pollut. 183, 224–233. https://doi.org/10.1016/j.envpol.2013.02.020

Vardoulakis, S., Gonzalez-Flesca, N., Fisher, B.E.A., Pericleous, K., 2005. Spatial variability of air pollution in the vicinity of a permanent monitoring station in central Paris. Atmos. Environ. 39, 2725–2736. https://doi.org/10.1016/j.atmosenv.2004.05.067

Vardoulakis, S., Solazzo, E., Lumbreras, J., 2011a. Intra-urban and street scale variability of BTEX, NO 2 and O 3 in Birmingham, UK: Implications for exposure assessment. Atmos. Environ. 45, 5069–5078. https://doi.org/10.1016/j.atmosenv.2011.06.038

Vardoulakis, S., Solazzo, E., Lumbreras, J., 2011b. Intra-urban and street scale variability of BTEX, NO2 and O3 in Birmingham, UK: Implications for exposure assessment. Atmos. Environ. 45, 5069–5078. https://doi.org/10.1016/j.atmosenv.2011.06.038

Vincent, K.J., Stedman, J.R., 2013. A review of air quality station type classifications for UK compliance monitoring.

Vitali, L., Morabito, A., Adani, M., Assennato, G., Ciancarella, L., Cremona, G., Giua, R., Pastore, T., Piersanti, A., Righini, G., Russo, F., Spagnolo, S., Tanzarella, A., Tinarelli, G., Zanini, G., 2016. A Lagrangian modelling approach to assess the representativeness area of an industrial air quality monitoring station. Atmos. Pollut. Res. 7, 990–1003. https://doi.org/10.1016/j.apr.2016.06.002

Wieringa, J., 1996. Does representative wind information exist? J. Wind Eng. Ind. Aerodyn. 65, 1–12. https://doi.org/10.1016/S0167-6105(97)00017-2

Wu, L., Bocquet, M., Chevallier, M., 2010. Optimal reduction of the ozone monitoring network over France. Atmos. Environ. 44, 3071–3083. https://doi.org/10.1016/j.atmosenv.2010.04.012

1 Appendix – Overview of past harmonisation efforts

In this appendix, we provide a brief overview of initiatives that were initiated within the FAIRMODE, AQUILA and CAMS communities to harmonize the common understanding and definitions and the SR concept.

2011

- During the FAIRMODE meeting in Norrköping, Sweden, it was concluded in SG1-WG2 that a consensus table is required on spatial representativeness, obtained through expert elicitation (contribution by B. Denby). A survey was organised within the FAIRMODE community to collect expert based length scales for spatial representativeness of NO_2, PM_{10}, $PM_{2.5.}$ and O_3 stations in background, traffic and industrial sites for various aggregation periods.

2012

- In the CAMS community, a paper is produced by (Joly and Peuch, 2012) describing an objective classification of air quality monitoring sites over Europe. The approach however is mainly targeted at site classification and identification of appropriate monitoring sites for regional scale model validation and data assimilation.
- At the FAIRMODE meeting in Utrecht, the Netherlands (SG1-WG2, led by B. Denby) was a dedicated workshop on SR:
    - The expert elicitation was summarised by N. Castell (NILU), with main conclusions that expert opinions differed a lot (e.g. 7 – 40 km for rural $PM_{2.5}$ daily averages; 20 to 245 m for near-source locations), and that there is a clear need for a scientific objective methodology. See further also Table 8 for results of the expert survey.
    - It was concluded that the concept of a circular area of representativeness is not applicable.
    - Several modelling teams presented a wide range of methodologies and views: using passive sampler surveys (L. Malherbe, INERIS), practices in the UK (K. Vincent, AEAT (now Ricardo)), the method by W. Spangl (Spangl et al., 2007), using land use data (VITO S. Janssen, D. Roet, VMM), using CFD modelling (F. Martín, CIEMAT), etc. The workshop Illustrated that different modelling teams in Europe have a very different understanding of the SR concept and its practical assessment.

2013

- A JRC-AQUILA working group comprised of several experts published a position paper on “Assessment on siting criteria, classification and representativeness of air quality monitoring stations”, the so-called SCREAM – paper (Geiger et al., 2013). This paper points out that :
    - “Since air quality assessment is mainly based on monitoring at distinct locations, it is necessary to extend this point information to spatial information”.
    - “So far, a definition of the spatial representativeness of monitoring stations is still missing in the AQ legislation and there is a need to develop tools for its quantitative assessment.”

2015

- A new FAIRMODE survey was organised by F. Martín et al. in an attempt to move this forward. This effort, documented in (Martín et al., 2015), resulted in the inception of an intercomparison exercise (IE) aiming at exploring the strengths and weaknesses of different contemporary approaches for computing the spatial representativeness area by applying them to a joint example case study (Antwerp, Belgium). The report contained a bibliographical review of studies on spatial representativeness published in scientific journals or technical reports.

2015 - 2016

- O. Kracht (JRC) led an intercomparison exercise for Antwerp for which VITO delivered the necessary data to accommodate all the submitted methodologies in the survey. 11 teams participated in the exercise.

2017

- A dedicated workshop was held in Athens back-to-back with the FAIRMODE technical meeting. to discuss the output of the Antwerp IE. Again, there were no firm conclusions, but the need was expressed for :
    - Sensitivity analysis on parameter values used in the similarity criteria.
    - Sensitivity analysis on the choice of additional criteria (i.e. should SR area be contiguous or not?)
    - How SR methods can be used to find optimal station position.
    - The community should work towards guidelines, however this objective likely requires first establishing a common framework for SR definitions and SR similarity criteria, and for harmonising the related terminologies.

2 Appendix - FAIRMODE Expert elicitation exercise

In Castell-Balaguer and Denby (2012), the question put to the expert audience was to provide a "radius of representativeness" for PM_{10}, PM_{25}, NO_2 and O_3 monitoring data at different averaging periods (one hour, one day and one year) for different (ill-defined) station types : rural background, suburban background, urban background, traffic and industrial. The survey question was as follows :

> *"For what horizontal area surrounding a monitoring station (represented by a circular diameter) do you consider the given station classification to be representative, for the given averaging period?"*

The concept "representative" was defined here as being indicative of the measured concentration not varying more than approximately 20 % within the given representative area. It is instructive to view the different ranges for such a hypothetical circular diameter which were returned by the review. These are given below here in the table (adjusted from (Castell-Balaguer and Denby, 2012)).

Table 8 : Indication of ranges for a circular "diameter of representativeness" provided by a survey put to an expert audience. The table indicate the min and maximum for this diameter provide by the 7 respondents in the survey and this for PM_{10}, $PM_{2.5}$, NO_2 and O_3 for 3 different averaging periods and 5 station types. Table adjusted from : (Castell-Balaguer and Denby, 2012).

		PM_{10} Averaging period			$PM_{2.5}$ Averaging period		
		One hour	One day	One year	One hour	One day	One year
Station classification	Rural background	2.5 - 30 km	5 - 30 km	10 - 50 km	5 - 30 km	7 - 40 km	10 - 50 km
	Suburban background	1 - 10 km	2.5 - 10 km	3 - 20 km	2 - 12 km	5 - 15 km	5 - 20 km
	Urban background	200 m - 8 km	300 m - 9 km	400 m - 20 km	200 m - 12 km	300 m - 15 km	400 m - 20 km
	Traffic	15 - 50 m	20 - 250 m	20 m - 2.5 km	15 - 250 m	20 m - 1 km	20 m - 2.5 km
	Industrial	50 m - 1 km	50 m - 5 km	20 m - 10 km	50 m - 3 km	50 m - 8 km	50 m - 15 km

		NO_2 Averaging period			O_3 Averaging period		
		One hour	One day	One year	One hour	One day	One year
Station classification	Rural background	1 - 30 km	2.5 - 30 km	10 - 30 km	5 - 100 km	10 - 100 km	20 - 100 km
	Suburban background	500 m - 2 km	1 - 5 km	1 - 10 km	2 - 15 km	5 - 20 km	3 - 30 km
	Urban background	200 m - 2 km	300 m - 3 km	400 m - 5 km	200 m - 15 km	300 m - 20 km	400 - 25 km
	Traffic	5 m - 50 m	10 - 100 m	10 - 250 m	10 - 500 m	50 m - 1 km	100 m - 2 km
	Industrial	50 m - 200 m	50 m - 1 km	50 m - 3 km	100 m - 4 km	100 m - 9 km	100 m - 20 km

3 Appendix – Overview by Levy and Hanna, 2011 for $PM_{2.5}$

Levy and Hanna (2011) provide an overview of methods used across studies to evaluate variability when only considering monitoring observations for $PM_{2.5}$. The table below lists additional methodologies to account for spatial heterogeneity (based upon papers from 2000 – 2007).

Table 1
Summary of findings from New York City fine particulate matter exposure studies.

Study	Avg. time	Number of monitors	Distribution of monitors	Concept of spatial heterogeneity	Key findings
(Ito et al., 2007)	24-h	30 FRM, 24 TEOM[a]	Urban/Suburban	Correlation, CV of means	High monitor–monitor correlation (>0.9), ~10% spatial variation in mean
(Ross et al., 2007)	3-year	62	Urban/Suburban	Land-use regression	Traffic within 300–500 m buffer explained 37–44% of variance across models; statistical models imply ~ 5 μg/m³ gradient across NYC
(Lall and Thurston, 2006)	24-h	3	Urban/Rural (Manhattan)	Correlation, factor analysis	Correlation between rural monitor and NYC monitor of 0.82; traffic factor contributes 5–6 μg/m³ in NYC, similar to rural-NYC difference
(Maciejczyk et al., 2004)	24-h	6 (mobile + fixed)	Urban (Bronx)	Comparison of medians/means	Median concentrations across monitors within 20%, difference in means significant for all sites in summer, some sites in spring/fall, none in winter
(Ito et al., 2004)	24-h	3	Urban (Bronx/Queens)	Correlation, factor analysis	High monitor–monitor correlation (>0.9), traffic factor contributes 2.5–6.2 μg/m³ across sites
(Restrepo et al., 2004)	24-h	4 (mobile + fixed)	Urban (Bronx)	Comparison of means	No significant difference across mobile van and 3 fixed sites in South Bronx (magnitude of mean difference = 0.4–1.5 μg/m³)
(DeGaetano and Doherty, 2004)	Hourly	20	Urban/Suburban/Rural	Factor analysis, comparison across percentiles	Most between-station correlations >0.85, 90% of variation across monitors explained by a single principal component with similar weight for each monitoring site. Concentrations consistently higher in lower Manhattan, with secondary maximum in upper Manhattan/Bronx.
(Bari et al., 2003)	Hourly, 24-h	2	Urban (Manhattan/Bronx)	Correlation	Correlation for hourly data of 0.79, 24-h data of 0.96
(Lena et al., 2002)	12-h	7	Urban (Bronx)	ANOVA, comparison of means	Intersite differences account for 24% of total variance, average concentrations range 11 μg/m³ across sites
(Kinney et al., 2000)	8-h	4	Urban (Harlem)	ANOVA, comparison of means	Intersite differences account for 14% of total variance, average concentrations range 10.5 μg/m³ across sites
(Qin et al., 2006)	24-h	5	Urban/Suburban (Bronx/Queens/NJ)	Correlation, comparison of means, factor analysis	3 NYC sites correlated 0.91–0.93, average concentrations range 0.7 μg/m³ across sites, traffic factors contribute 1.5–2.5 μg/m³ across sites
(Venkatachari et al., 2006)	Hourly	2	Urban (Bronx/Queens)	Correlation, CD	Correlation of 0.77, coefficient of divergence (CD) of 0.15 (moderate spatial heterogeneity)

[a] FRM = Federal Reference Method, TEOM = Tapered Element Oscillating Microbalance.

RICARDO

ricardo.com