Assessing the spatial representativeness (SR) of air quality sampling points – Sensitivity and feasibility tests for a tiered approach – Final Report

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

ED 11492 | Task 1 Issue number 5 | Date 21/12/20

Customer:

Report for European Commission - DG Environment

Customer reference:

Ares (2018) 4920320

Confidentiality, copyright and reproduction:

Contact:

Beth Conlan, Gemini Building, Fermi Avenue, Harwell, Didcot, OX11 0QR, UK

T: +44 (0) 1235 753 480

E: beth.conlan@ricardo.com

Author:

Hans Hooyberghs (VITO), Leonor Tarrason (NILU), Stijn Janssen (VITO) and Joana Soares NILU)

Approved by:

Beth Conlan

Date:

21 December 2020

Ref: ED 11492

Ricardo is certified to ISO9001, ISO14001, ISO27001 and ISO45001

Disclaimer: *The conclusions contained within this report may be amended in light of outcomes from discussions with relevant stakeholders.*

Table of Contents

1 Introduction

1.1 Challenges and aims of this project

This document reports on activities in Task 1 of the project **"Assessing the spatial representativeness of air quality sampling points",** with specific contract number **070203/2018/793545/SFRA/ENV.C.3**, which is service request 5 under Framework Contract number ENV.C.3/FRA/2017/0012 dated 17th December 2018. The aims and objectives of the project can be summarised as:

- To support the development of recommendations on methodologies to assess spatial representativeness of sampling points, with a view of supporting improvements in air quality modelling
- To collate authoritative air quality and air emission maps, and compile these to produce bottom-up composite maps of air pollution in Europe.
- To provide support to the assessment of the application of the criteria for selecting sampling points in Member States laid out in EU ambient air quality legislation, by carrying out an overview of the existing network.

To tackle these objectives, the project is structured around three different Tasks and this document reports on Task 1, which focusses on the development of recommendations for the assessment of spatial representativeness for specific assessment needs. In a first phase of Task 1 a literature review was conducted to take stock of the recent developments in this field. This resulted in a Tiered framework to classify and structure spatial representativeness assessment methodologies per assessment need identified in the implementation of the Ambient Air Quality Directive (2008/50/EC, as amended by Commission Directive (EU) 2015/1480) (AAQD). In this report we now describe the results of a subsequent sensitivity analysis that tested various approaches and methodologies to eventually come to conclusions and recommendations for the assessment of spatial representativeness.

1.2 The need for sensitivity and feasibility tests

The Literature Review has identified three domains for which additional guidance is required (Maiheu and Janssen 2019). The first domain deals with the spatially explicit spatial representativeness (SR) area of a monitoring station. Some important longstanding open issues related to this assessment need concern the choice of similarity criteria and related threshold values, and the difference between a contiguous versus a non-contiguous approach. Secondly, under the "Commission Implementing Decision on the reciprocal exchange of information and reporting on ambient air quality" (2011/850/EU, further referenced as IPR (Article 12 and Annex II), Member States are obliged to report the area, population and road length for every air quality zone, where an exceedance of an environmental objective of the AAQD has occurred, such as of a limit or target value. Guidance on how to report these exceedance situation indicators can help improve the comparability of the resulting information. The third and final domain concerns the evaluation of the monitoring sampling points in the overall monitoring network design and the use of measurement data from sampling points for model calibration and model validation.

The Literature Review furthermore describes how different approaches and methodologies are currently used for those different domains or assessment needs. The variety of approaches described has also been identified in the joint "FAIRMODE and AQUILA Inter-comparison Exercise on Spatial Representativeness of Air Quality Monitoring Sites" led by the JRC, further on referred to as IE (Kracht et al. 2017). Both reports identified the need for additional sensitivity studies to provide further insight in the differences and overlaps between the methodologies regarding the assessment needs highlighted in the previous paragraph. As a first step in the process, the Literature Review has put forward a Tiered approach to organise the methodologies according to their complexity and the assessment needs they serve. Additional sensitivity studies regarding the relation between the different Tiers have been introduced. Important aspects of consideration include how more complex methods (higher Tiers) can inform less complex methods (lower Tiers) and providing guidance on fitness for purpose of the approaches in the different Tiers. Furthermore, specific issues regarding exposure assignment or spatial variability in the presence of street canyons have been raised.

The results of the sensitivity tests have led to recommendations on methodological choices regarding the assessment of SR, the reporting of exceedance situation indicators and evaluation of a monitoring network design.

1.3 Setup of sensitivity and feasibility tests

This report describes the results of a set of sensitivity and feasibility tests to evaluate SR assessment methodologies. Sensitivity tests are applied in three distinct cities: Oslo, Krakow and Antwerp. These cities have been chosen as they present a variation in air quality situations for NO_2 and PM_{10} pollution and meteorological conditions, and a spread in geographic zones in Europe (North, East and West of Europe). Moreover, the lead authors of the document are acquainted with assessing and modelling the air pollution in these three cities. Note that in this study we mainly focus on the pollutants NO_2 and PM_{10} as their assessment in an urban context is most challenging.

The Tiered Framework approach has been used to shape the sensitivity study. For each of the cities, use is made of methodologies from all available Tiers. Developing new methods and / or repeating existing methods with modified parameters or for other domains is outside the scope of the current project. Obviously, this puts some limitations on the overall robustness of the sensitivity tests, but we highlight in the discussion to what extent the results can be generalized.

The report is structured as follows: all methodologies used in the sensitivity analysis are described in Chapter 2. Only a brief description highlighting the most important methodological aspects is provided, and we refer the reader to the relevant literature for more technical details. The following Chapters describe the results of the sensitivity tests with respect to methodological options for the different assessment needs. Chapter 3 focuses on the SR areas of monitoring sampling points. Building on these results, Chapter 4 deals with the exceedance situation indicators of the area, population and road length in exceedance of the annual EU limit values. Chapter 5 adds the time aspects to the analysis and presents some feasibility tests focussing on the design of monitoring networks. The relation between the different Tiers is assessed throughout these Chapters.

Based on the results of these sensitivity tests the report goes on in Chapter **Error! Reference source not found.** to provide initial recommendations on methodological choices to assess SR and how to apply the Tiered approach in practice.

2 Overview of Tiered approaches for spatial representativeness

This Chapter provides an overview of the different methodological approaches used in the sensitivity tests. In the first phase of this project a Tiered approach was put forward to organise various methodologies described in the literature and/or used in practical applications according to assessment needs and increasing complexity (Maiheu and Janssen 2019). In general terms the Tiers range from expert opinion (Tier 1), to the use of proxy data or specific measurement campaign data (Tier 2), and to the application of fit-for-purpose modelling applications (Tier 3) (see Section 2.3.1). In a fourth Tier, the modelling applications are complemented with additional measurement data to further improve the quality of the modelling applications. In the Literature Review, the Tiered approach was then further refined for each of the assessment needs summarized in the introduction. A summary of this analysis is given in Table 1 below.

Table 1: SR assessment methods in different Tiers per assessment need (source: Literature Review (Maiheu and Janssen 2019))

<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>

A comprehensive overview of all available methodologies per Tier and per assessment need is given in the Literature Review (Maiheu and Janssen 2019). In the remaining part of this Chapter we will describe the methodologies that are available in the three test cities and can be used in the sensitivity tests.

2.1 Tier 1 – Expert analysis

Tier 1 groups all the methods that solely rely on expert opinion and relatively simple “distance to source” considerations.

For the estimation of spatial representativeness (SR) areas, we rely on the definitions put forward in the AAQD. For different sampling points types typical length or area scales are specified and these can be used to estimate a (minimum) SR area. According to the AAQD, traffic and industrial stations should “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, where feasible”. Furthermore, urban background locations “shall, as a general rule, be representative for several square kilometres.” Considering the sampling points are placed by competent authorities in the EU Member States according to the provisions of the AAQD, these values provide indications for the (minimal) size of the SR area that air quality monitoring sampling points should have, including under a Tier 1 approach. In the sensitivity tests, the length scales provided by the AQD are used in a simple GIS-approach in which the spatial representativeness area is determined by a fixed-radius zone around the measurement station. The only parameter in this analysis is the radius of the circle.

For the estimation of the **exceedance situation indicators** use can also be made of the official data reported by the respective Member States to the EU EIONET Central Data Repository (CDR, https://cdr.eionet.europa.eu/). In this Report we refer to this information as “CDR information”. This information, provided by Member States, is considered as expert opinion at least for the cases studies in this report: the cities Antwerp, Oslo and Krakow.

For the **network design**, we rely on the station classification as assigned by the network managers and reported in the CDR. The application of siting criteria and sampling point classification in Member States was evaluated in Tarrason et al (2020), as part of Task 3 in this project. National competent authorities need to follow Annex III of the AQD for the macro and microscale siting criteria of their sampling points. The Member States' and European Commission's Common Understanding of the Commission Implementing Decision laying down rules for Ambient Air Quality Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of information and reporting on ambient air (Decision 2011/850/EU) (IPR guidance) provides guidance for station classification based on direct measurements and local knowledge at site. Although over 60% of the countries indicate that they have available information on the spatial extent of representative area for their monitoring network, none of them reports this information. This lack of reporting in the CDR is an indication of the complexity and difficulties associated with the calculation of the spatial representativeness (SR) of sampling points.

An overview of the various methodologies used per assessment cluster and per city is given in Table 2 below.

Table 2: Overview of Tier 1 methodologies used by the national competent authorities included in this sensitivity analysis

Tier 1	SR area	Exceedance Situation Indicators	Network design
Antwerp	AQD length scales	CDR information	AQD siting and classification
Oslo	AQD length scales	CDR information	AQD siting and classification
Krakow	AQD length scales	CDR information	AQD siting and classification

2.2 Tier 2 – Additional monitoring data and use of proxies

Tier 2 goes a step further and groups all the methods which combine expert opinion with proxy data related to source and dispersion characteristics. In addition to this proxy data approach Tier 2 also comprises specific monitoring campaigns designed to assess the spatial (or temporal) gradients in the concentration levels. So pure measurement approaches without the involvement of modelling techniques are also classified under the Tier 2 approach. Finally, time series trend data are used to support monitoring site objective classification which can be applied as Tier 2 method for network design evaluation.

In the remaining part of this Section we will describe the methodologies that are available in the three test cities and can be used in the sensitivity tests. For convenience, the methods are grouped by type (proxy based, methods based on sampling campaigns and objective classification for network design). Developing new methods and / or repeating existing Tier methods with modified thresholds or for other domains is outside the scope of the current project.

2.2.1 Proxy-based Tier 2 methods

For the purpose of calculating exceedance situation indicators, in the joint FAIRMODE and AQUILA Spatial Representativeness Intercomparison Exercise (IE) which focussed on Antwerp, results of 11 SR methods were compared, three of them being a Tier 2 approach based on proxy data: a land use regression method carried out by VITO, an emission based methodology carried out by ISSeP & AWAC and an application of the Spangl-classification method carried out by VMM (Kracht et al. 2017). The results of these three methods are readily available and can be used in this analysis.

It should be mentioned however that all methods applied in the IE relied on input data for the year 2012, whereas the methods for the other Tiers in this report are based on measurements and model results for 2018. When comparing these Tier 2 results with the results of other Tiers, we thus observe a combined effect of different years and different methodologies. We expect the effect of the different years to be rather small. Although there have been decreases in the emissions in the 2012 to 2018 period, the spatial pattern of the emissions and the land use proxies are rather constant, especially in the vicinity of the three sampling points under consideration.

Note that all free parameters of the IE methods have been fixed by the participating teams. Oliver Kracht deliberately did not provide any specific guidance on these methodological options during the IE. As a result, participating teams were required to make their own choices which are still reflected in the final results and therefore complicate the comparison of various results of the IE. A sensitivity analysis considering the underlying parameters of these methods falls outside the scope of the current task.

In the IE, the **SR area** of three air quality measurement sampling points in Antwerp was determined, including two urban background sampling points and one traffic station. A comparison of these three Tier 2 approaches applied to three sampling points in Antwerp sought to reveal differences amongst methods for different station types as well as the overall potential of these Tier 2 methods to assess SR areas. The three Tier 2 approaches of the IE can also be used to estimate the **exceedance situation indicators**. It should be taken into account that for the three Antwerp sampling points investigated in the IE only the traffic station has an exceedance of the limit values set in the AQD. An overview of the available Tier 2 methodologies is given in Table 3. Below in Section 2.2.1.1 to 2.2.1.3 we provide a brief introduction to the three methodologies applied in the IE and direct the reader to (Kracht et al. 2017) and references therein for further details.

To our knowledge, no proxy-based Tier 2 methods have been applied for Krakow and Oslo. As repeating existing Tier methods for other domains falls outside the scope of the project, no results are available for these cities for Tier 2.

Table 3: Overview of proxy-based Tier 2 methodologies used in this sensitivity analysis

Tier 2	SR area	Exceedance Situation Indicators	Network design
Antwerp	- Land cover proxy (IE) - Emission proxy (IE) - Spangl-methodology (IE)	- Land cover proxy (IE) - Emission proxy (IE) - Spangl-methodology (IE) - Passive sampling campaigns	Objective classification
Oslo	NA	NA	Objective classification
Krakow	NA	NA	Objective classification

2.2.1.1 Land use regression methodology

VITO determined the SR of the measurement sampling points by similarity in land cover proxy values. The methodology is based on a relationship or 'trend function' between the long term averaged pollutant concentration observed in all available air quality monitoring sampling points (AQMS) in the region and a land cover indicator β. This β indicator, based on the Copernicus Land Monitoring Service's Coordination of Information on the Environment (CORINE) Land Cover (CLC)[1], is optimized to explain as much as possible the variations observed in the AQMS network. The β map is available at 4 km x 4 km resolution.

The SR area is composed of these 4 x 4 km blocks. Allowing a user defined deviation in concentration level of a specific AQMS, the trend function is used to assess a corresponding variation in β. In the IE a deviation in concentrations of 15% was used. The SR area for the AQMS is then determined as the set of grid cells, for which (i) the β value is within this AQMS specific β interval and that (ii) form a contiguous area neighbouring the AQMS.

Note that the application of this methodology required a few specific choices with respect to the definition of the SR area. The concentration threshold of 15% as well as the contiguous approach was adopted by the VITO modelling team during the IE exercise.

2.2.1.2 Emissions methodology

The SR methods used by ISSEP & AwAC (Belgium) are based on emission data. Depending on the type of station a specific approach is adopted:

- For traffic sites, all streets are classified into three pollution levels depending on road emissions (NO_x emissions when dealing with NO_2-concentrations and primary PM_{10} emissions for PM_{10}-concentrations) and on how the traffic lanes are enclosed by surrounding buildings. The SR area is evaluated within a 500 m radius around the AQMS and extends to all road segments with the same emission level.
- For background sites, total emissions (NO_x emissions when dealing with NO_2-concentrations and primary PM_{10} emissions for PM_{10}-concentrations) of each pollutant are first disaggregated into 100x100 m² cells, then re-aggregated through a spatially moving sum with a circular window of radius 1 km. The SR area extends to all points with total emission values similar to those at the AQMS ± a tolerance.

2.2.1.3 Spangl-Classification method

VMM (Belgium) applied a classification methodology developed by UBA under a service contract for the European Commission (Spangl et al. 2007). The methodology considers emissions from road traffic, domestic heating and industrial emissions, and dispersion conditions for all AQMS in the network. Population density is used as a proxy for domestic heating and CORINE land cover data for dispersion conditions. The surrounding of the AQMS is divided into smaller sub-areas, each of which is classified. For each station, the classified domain is a 1100x1100 m² grid with mesh size 100m (and the classification is thus limited to a domain of 1.21 km²). The similarity in classifications of the sub areas and the AQMS is then quantified. Finally, the SR area is calculated as the set of sub areas for which the weighted sum of a similarity indicator is above a given threshold.

2.2.2 Specific monitoring campaigns

As an alternative to the Tier 2 methods that focus on proxy data (land use, emissions), methods relying on (large-scale) sampler campaigns could also be used. An example considers the passive sampler network of the Flemish Environmental Agency (VMM) in Antwerp. Twenty passive diffusion tubes measure the NO_2-concentrations at locations which are poorly represented in the official network (complex traffic locations, tunnel exits, urban parks…). The samplers are biweekly replaced, and for each two-weeks period a calibration using passive tubes distributed at the telemetric sampling points is applied. By combining all biweekly measurements, indicative annual mean concentrations for an additional twenty locations are composed. This campaign is however not used by local experts to assess the exposure

[1] https://land.copernicus.eu/pan-european/corine-land-cover

indicators, since only a few additional locations are sampled. A reliable Tier 2 method should be based on a large-scale sampling campaign in which the distance between the samplers is in the same order of magnitude as the length scale of the spatial variation of the underlying air pollutant. For NO_2, this implies that hundreds or thousands of samplers should be distributed among a city. An example of such a campaign is the Curieuzeneuzen citizen science campaign, in which NO_2 measurements took place at almost 20,000 locations in Flanders during May 2018. 1131 of those locations where situated in the Antwerp air quality zone. Acquiring such an extensive dataset requires a citizen science approach, which entails a large-scale scientific and logistic effort. Typical tasks carried out in the Curieuzeneuzen project include an advertising campaign to promote the project, a selection procedure to identify the most optimal locations using model results, a logistic operation to distribute and collect the samplers, a large-scale operation to read out the samplers and a data-processing task (Meysman et al. 2020, in preparation)[2]. Obviously, all these tasks also come with monetary and labour costs which should be considered in the overall applicability of this Tier 2 methodology.

The resulting dataset cannot directly be used to assess the exceedance situation indicators, as the time-limited character of the measurements prohibits compliance checking against annual mean concentrations. To circumvent this issue, a model approach relying on the passive sampler results and measurement data for an entire year has been developed (De Craemer et al. 2019)[3]. This "model" is an example of a Tier 2 approach, which solely relies on a dense passive sampler campaign and measurements of the official telemetric network. We will use the results of this Tier 2 approach for Antwerp in assessing the **population exceedance situation indicator.**

No sampling-based Tier 2 methods have been applied for Krakow and Oslo, because, to our knowledge, no reliable large-scale (> 100 samplers) campaigns are available for these cities.

[2] A video on https://curieuzeneuzen.be/in-english/ provides an overview of these tasks of the Curieuzeneuzen project.
[3] Note that the legal status of this approach is uncertain, because it is not clear whether the data complies with the strict definition of "model estimates" set in the AAQD nor with the quality requirements proposed by the AAQD for indicative measurements (De Craemer et al. 2019).

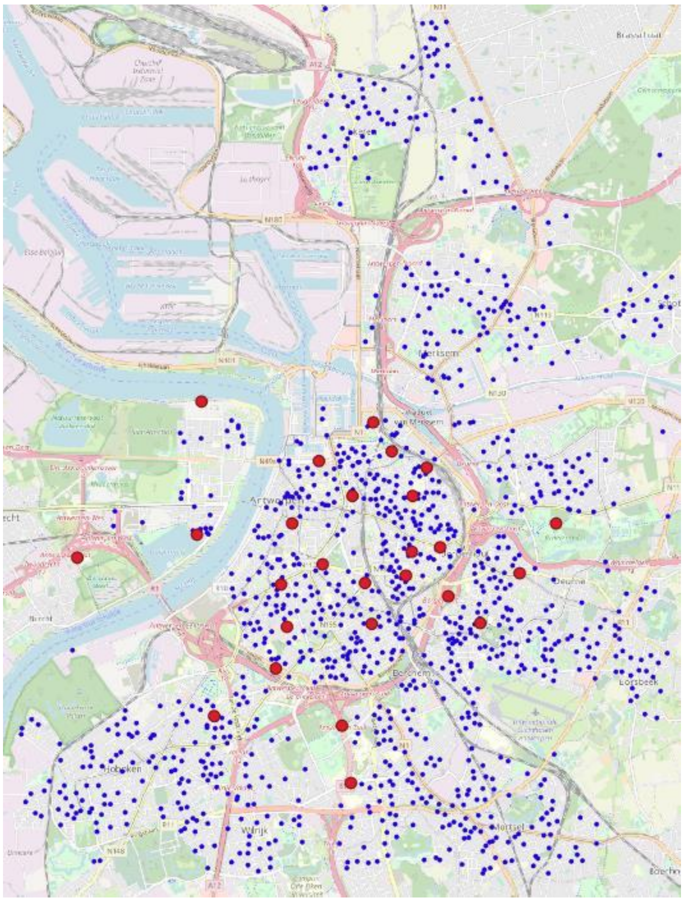

Figure 1: Locations of the passive samplers for the campaigns in Antwerp. Red dots indicate the locations of the VMM samplers, the blue dots indicate the locations of the participants in the Curieuzeneuzen citizen science campaign.

2.2.3 Objective classification

For monitoring design purposes, objective classification methods have been proposed to determine sampling points characterisation based on the analysis of their specific time series. The two methods are the one developed by Joly and Peuch (2012) and the one proposed by Soares et al (2018). Both methods for station classification are: a) objective by being based entirely on the existing time series of observations at each sampling point; b) pollutant-specific as it applies to each sampling point; and c) provide an independent approach to "fingerprint" the station/sampling point in relation to the spatial representativeness. They provide a complementary independent approach to test the sampling point siting criteria and classification as identified following Annex III of the AQD and the requirements under the Implementing Decision.

The Joly and Peuch (2012) method was developed in the context of the precursors of the Copernicus Atmospheric Monitoring Service (CAMS) and is currently used in CAMS to select the sampling points that are to be used in the data assimilation routines used by the service (L. Rouïl, pers.comm.). The method classifies sampling points in 10 different classes based on their statistical timeseries behaviour. It is considered “objective” since it rests on numerical criteria and can be uniformly applied for all timeseries data from existing sampling points over the whole of Europe. The Joly and Peuch (2012) algorithm makes use of a linear discriminant analysis. It requires the selection of a training dataset to compute the so-called Fisher axis which is then used as a reference to classify any station measuring the considered pollutant. The statistical algorithm also uses information on station classification (type of area, type of station) as prior knowledge. The method proposed was further applied by Malherbe et al. (2013) to test the station classification of the reported sampling points in the EEA metadata in 2013 for NO_2, PM_{10} and O_3. The results showed that the Joly and Peuch (2012) methodology was in good agreement with respect to the standard classification. However, Malherbe et al documented some important short-comings of the method, because it did not manage to provide a clear correspondence from the 10 classes classification and the classification currently used in the context of European air quality legislation. It was also difficult to explain outliers in the classification. They showed that the Joly and Peuch methodology succeeds in separating sampling points characterized by a background behaviour (classes 1 to 3) from stations/sampling points showing a more traffic-related behaviour (classes 8 to 10). Background sampling points mostly fall into the lowest classes (below 3) whereas suburban and especially urban traffic sampling points mostly fall into the highest classes (above 6). This means that the method describing the temporal variability of concentrations succeed in differentiating rural background sampling points from those influenced by traffic. However, the methodology failed to describe industrial sampling points and urban background sampling points appeared to be spread throughout all classes; classes 5 to 7 are the most represented ones. Malherbe et al concluded that additional work for objective classification, using land cover, population and emission information would be necessary to make use of the Joly and Peuch classification in air quality management applications.

The clustering method by Soares et al. (2018) shows more promising results for evaluating the siting and classification of a given monitoring network based on the actual measurement data. It analyses the level of similarity or dissimilarity of air concentration data from all sampling points and allows the identification of possible outsiders. The method has been successfully tested for all traffic and industrial sampling points in Europe in 2017 in Tarrason et al, (2020). It is presented in Section 2.3.3. of this report and forms the basis for the evaluation of Tier 2 and Tier 3 methods applied in the feasibility tests of Chapter 5 for monitoring design and model validation purposes.

2.3 Tier 3 – Use of modelling data

Tier 3 groups all methods which rely on the use of suitable and fit-for-purpose air quality models. An important remark in this definition concerns the description of a modelling chain as being suitable or fit for purpose. In the context of these sensitivity tests we assume the models can describe the relevant spatial and temporal variations of the observed concentrations in the related air quality zone. A general discussion about model qualification as fit-for-purpose is deemed outside the scope of the current task and is something FAIRMODE should further investigate.

For the three selected cities, we make use of the available high-resolution modelling chains. In Krakow and Antwerp, the ATMO-Street model chain has been used, in Oslo the EPISODE application is applied. The former model is described in Section 2.3.2, the latter one in 2.3.3.

For the evaluation of the network design, a cluster analysis is applied. This methodology is described in Section 2.3.4.

Table 4: Overview of Tier 3 methodologies used in the sensitivity analysis

Tier 3	SR Area	Exceedance Indicators	Situation	Network design
Antwerp	ATMO-Street	ATMO-Street		Cluster analysis
Oslo	EPISODE	EPISODE		Cluster analysis
Krakow	ATMO-Street	ATMO-Street		NA

2.3.1 Fit for purpose models

Models or modelling applications as a whole are considered as a Tier 3 (or Tier 4) approach when they are fit-for-purpose. In the framework of FAIRMODE various aspects of such a fit-for-purpose definition have been suggested over time. These elements can be summarized as:

- The modelling application should comply with the FAIRMODE Modelling Quality Objective (MQO) and fulfil the Modelling Quality Indicator (MQI) criterium MQI < 1 (FAIRMODE Guidance Document vs3.2, Janssen et al, 2020)
- The modelling application should be able to meaningfully reproduce what is observed in the ambient atmosphere, regardless of the spatial scale and within the tolerance margins of the Modelling Quality Objective (FAIRMODE Recommendations, Thunis et al., 2019). As a consequence, all available observations in the modelling domain (or air quality zone) have to be used in the calculation of the MQI.
- The previous requirement does not explicitly set a criterium for the spatial scale of the modelling application. Based on the sensitivity analysis of this study presented further on in this report a minimum model resolution of 100m for NO_2 and 1km for PM_{10} seems to be appropriate.
- To describe all relevant concentration gradients, the modelling application should explicitly take into account emissions from the important sources in the air quality zone.
- For NO_2 the inclusion of road increments in the modelling chain seems appropriate, especially when SR areas of traffic stations are assessed or when an indicator with road length in exceedance is estimated. Street canyon effects are specific contribution to these road increments and have an important local impact on NO_2 concentrations. For the assessment of the number of people exposed to exceedances, the effect of street canyons is still under debate.
- As an additional requirement it is recommended that the spatial heterogeneity of the modelling application results is tested versus the observed heterogeneity. A possible methodology to perform such a test is via a semi-variogram analysis. Another possible methodology is the use of hierarchical clustering analysis.

We consider that the definition of a model's fitness-for-purpose requires further discussion, notably within the FAIRMODE modelling community, but the elements mentioned above can serve as a basis for this debate. However, this shouldn't prevent the community from relying on validated modelling applications, or so called Tier 3 approaches, to assess SR of monitoring stations and the estimation of exceedance situation indicators.

2.3.2 ATMO-Street

The ATMO-Street model chain (Lefebvre et al. 2013a) consists of the land-use based interpolation model RIO determining background concentrations (Janssen et al. 2008), the bi-gaussian plume dispersion model IFDM accounting for the impact of local emissions from road traffic, shipping and industry (Lefebvre et al. 2011), and the street canyon module OSPM that calculates the in-street increment resulting from street canyon effects (Berkowicz et al. 1997; Berkowicz 2000a). The model calculates hourly concentration values at irregular spaced receptors, which are subsequently gridded to a regular raster with a 10m resolution. The model chain is visualized in Figure 2. As details on the residential emissions are not available at the level of the individual buildings, these emissions are not included in the IFDM-OSPM model core. Pollution related to the residential sector is however considered in the background concentrations. The model results thus incorporate the residential emissions on a lower resolution (order of several kilometres, the resolution of the RIO model) than the traffic and industrial emissions. This approximation is only of minor importance for NO_2 but can have a significant influence on the concentrations field for PM_{10} in the immediate vicinity of strong residential sources.

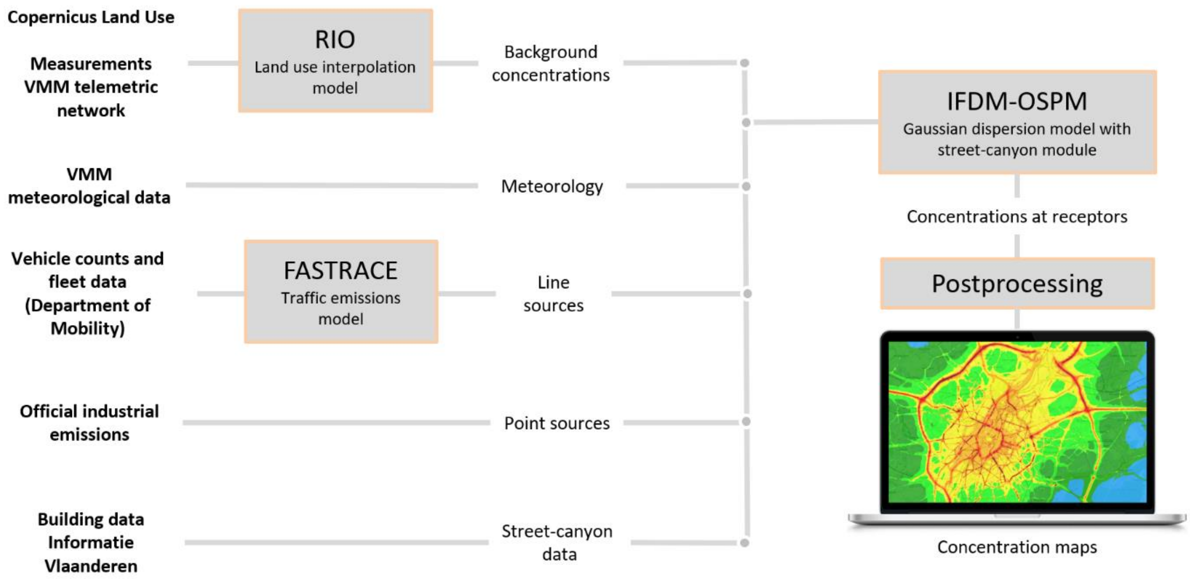

Figure 2: Schematic overview of the ATMO-Street model chain.

Model simulations were carried out with the ATMO-Street model for Krakow and Antwerp for the year 2018. Both simulations start from annual emissions, which are distributed over the hours of the year using temporal profiles. For the traffic sector these profiles are specified per road type (urban, rural and highway), but all shipping and industrial emissions use the same time profile. Although yearly mean concentrations are not much influenced by these choices, it is an important caveat when discussing the correlation between the time series for model output.

For the Antwerp case study, the official ATMO-Street map produced by the Flemish Environmental Agency (VMM) has been used[4]. All input stems from official datasets of the Flemish authorities. The regional background model RIO has been set up using the official measurements of the telemetric network of the Flemish Environmental Agency and the Copernicus Corine Land Cover as land-use input. The Gaussian dispersion and street canyon model use the official total road traffic emissions data of the Department of Mobility, combined with official point source emissions data for industry and line source emissions for shipping. Building data has been retrieved from the official building dataset for Flanders (Informatie Vlaanderen). Meteorological data has been composed by the Belgian Interregional Environment Agency (IRCEL - CELINE) by assimilating Copernicus C3S ERA5 reanalysis data (Copernicus Climate Change 2017) with local measurements, yielding wind and temperature fields with a 1km resolution.

[4] Yearly mean concentrations are visualised on https://www.vmm.be/data/stikstofdioxide-no2-jaargemiddelde.

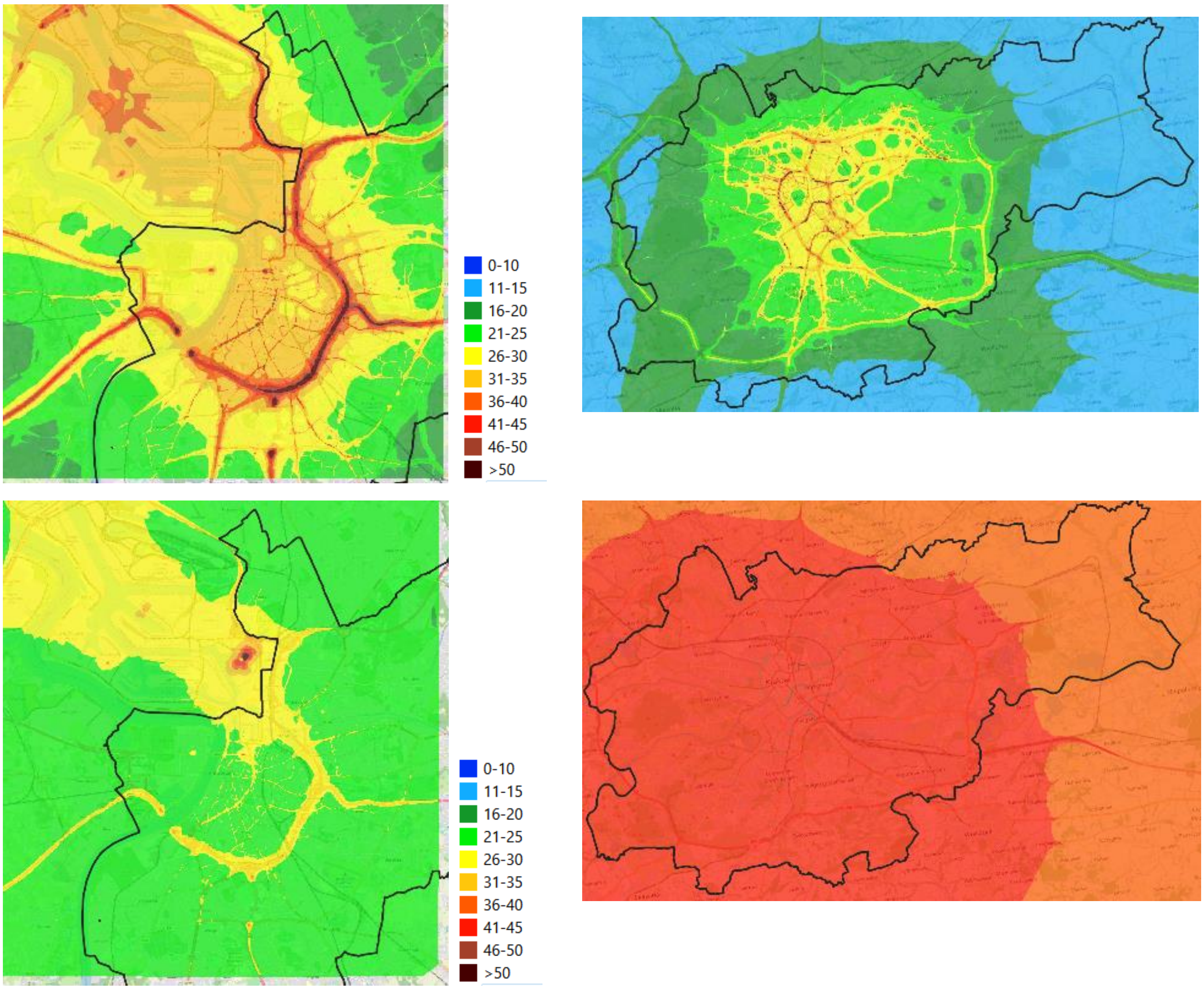

Figure 3: ATMO-Street model results for Antwerp (left) and Krakow (right). The top figures visualise the yearly mean NO_2-concentration (in µg/m^3), the lower show the yearly mean PM_{10}-concentration (in µg/m^3). The black line indicates the extent of the air quality zone for Antwerp and Krakow. The colour scale is shared between all figures.

The Krakow model run has been set up by VITO, relying on input of the Krakow city administration. RIO uses measurements of the official telemetric monitoring network and population density disaggregated with Copernicus Corine Land Cover (Gallego 2010) for the land use proxy. Traffic emissions are composed by combining traffic volumes and fleet information of the Krakow city administration with COPERT emission factors. Building data is taken from the 3D model of the Central Office of Geodesy and Cartography (available at www.geoportal.gov.pl). Meteorological input stems from the Copernicus C3S ERA5 reanalysis data (Copernicus Climate Change 2017).

The final yearly mean NO_2 and PM_{10} concentration maps for Antwerp and Krakow are visualized in Figure 3.

2.3.3 EPISODE model

The EPISODE model is a combined 3D-Eulerian and Lagrangian air pollution dispersion model for urban and local scale applications (Slørdal, Walker and Solberg, 2003; Slørdal, McInnes and Krognes, 2008). It consists of a Eulerian 3D grid model with embedded subgrid Gaussian and Lagrangian models, which take care of the dispersion from different type of sources (point, line, and area sources). The Eulerian part of the model consists of a numerical solution of the atmospheric (mass) conservation equation of the pollutant species in a three-dimensional Eulerian grid. The Lagrangian part consists of separate subgrid-models for line- and point-sources. The line source model is an integrated Gaussian type model, while the point source model is a Gaussian puff trajectory model. Point sources are for example stack emissions from industry. Line sources are typically emissions from traffic. Area sources are emissions dispersed in space as for example the emissions from domestic heating in a city.

The model is typically used in Norway to calculate air pollution concentrations in cities and urban areas from multiple emission sources such as road traffic, shipping, domestic heating and industry. The model calculates hourly average concentrations as gridded values and in a set of irregularly placed receptor points. The output of the model in hourly frequency is used for calculating long-term average concentrations such as annual averages and other statistical parameters such as percentiles, on the basis of the hourly data. EPISODE has been applied for the calculation of the concentration of airborne species such as SO_2, CO, NO_X, NO_2, PM_{10} and $PM_{2.5}$. Calculations of NO_2 are based on a simplifying assumption of photochemical equilibrium between NO, NO_2 and O_3 for each time step. For urban scale application, deposition is not a consideration.

The model simulations are carried out in 1x1km using the EPISODE model and then emissions are incorporated as line sources so that the final resolution of the results is 100x100m. In order to create maps at 100 m resolution, the model domain is populated with a large number of receptor points. These receptor points are placed with higher density near roads, out to the extent of the road link influence distance (400 m), the distance to which the line source model is applied. Outside of this region receptor points are placed every 500 m in a regular grid as these sample only from the grid model. The mapping process consists of pre-processing of receptor points and post-processing for creating the maps. It should be mentioned that street canyon situations are not modelled with the approach chosen in EPISODE.

The model then calculates concentrations at all the mapping receptor points and saves the annual mean concentrations, the number of exceedances above the prescribed limit value and the related percentiles for each limit value. The model also saves the same type of concentration data for each model grid.

The EPISODE model calculates concentrations at the receptor points by adding line source and grid model concentrations. No interpolation of the gridded concentrations is applied, often leading to clearly visible ‘grid shapes’ in the receptor point concentration data. To obtain smoother variations in the map, related to gridded concentrations, the receptor data is post-processed. The gridded concentration fields are interpolated, using a cubic spline interpolation, at all receptor points. The original gridded concentrations are then subtracted from all receptor points and the interpolated gridded concentrations are added back. This creates a smooth concentration surface for the grid model contribution but does not change the line source contribution. The new receptor point data is then linearly interpolated to a 20 m sub-grid throughout the entire model domain creating a high-resolution map. This interpolated sub-grid is then aggregated into 100 m grids by taking the mean of the sub-grids. Maximum sub-grid values are also calculated for each 100 m grid but are not used in the maps. In this way the 20m sub-grid interpolation is used as a numerical integration method to determine the means in the 100 m mapping grids. Further detail can be found in Denby et al. (2014). Results from the EPISODE model are given in Figure 4.

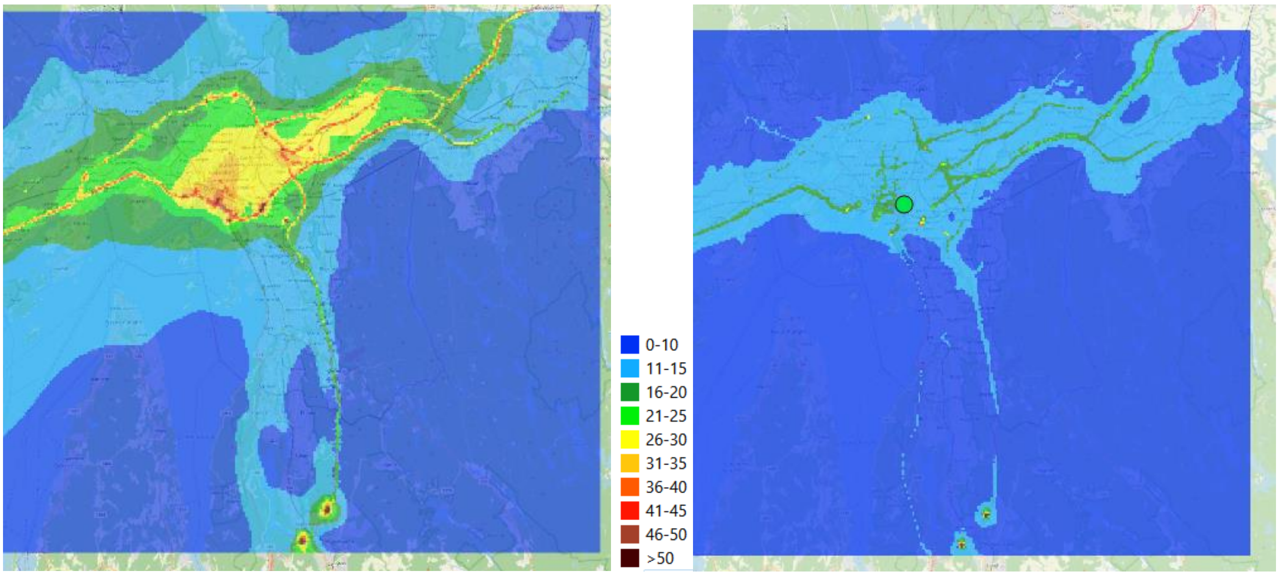

Figure 4: Model results for Oslo for 2015. The left figure visualises the yearly mean NO_2-concentration (in µg/m^3), the right shows the yearly mean PM_{10}-concentration (in µg/m^3). The colour scale is shared between both figures.

The model has been thoroughly validated in Norwegian applications as part of the Bedre Byluft (Better City Air) program since 2000. An evaluation of the EPISODE model was also carried out in FAIRMODE and the results are documented in Janssen et al. (2017). The results from these modelling applications are comparable with those of state-of-art models used in Europe for air policy applications and the model chain satisfies the demands of the FAIRMODE Model Quality Objective (MQO).

2.3.4 Cluster analysis

The hierarchical clustering methodology proposed by Soares et al. (2018) provides a screening tool to evaluate measurement and modelling air quality data. The methodology helps us establish the level of similarity (or dissimilarity) of the different air quality values. It requires an air quality dataset covering a long-term time series, preferably, at least one year of data. The data can be either from monitoring sampling points or from model calculations. The methodology uses three different dissimilarity metrics: (1) a correlation variable 1-R, where R is the Pearson linear correlation coefficient (Solazzo and Galmarini, 2015), (2) the Euclidean distance, EuD (Soares et al, 2018), and (3) multiplication of metric (1) and (2). The metric based on correlation, (1-R), assesses dissimilarities associated with the changes in the temporal variations in concentration, while the metric based on the Euclidian distance (EuD) assesses dissimilarities on the basis of magnitude of the concentration over the time period of the analysis. The multiplication of these two metrics (1-R) x EuD allows assessing correlations in terms of both time variation and pollution levels.

The results from the hierarchical clustering methodology will be presented here either as dendrograms or as coloured spatial clusters. Dendrograms are 2D representations of the clustering process showing the pattern of linkages between the data series while clustering occurs, as well as their level of dissimilarity. This graphical representation is useful to visualize the results for monitoring results. From the dendrograms, it is possible to identify the dissimilarity level at which sampling points cluster together, distinguishing groups of sampling points. The clusters can then be displayed geographically, colour-coded according to the cluster the sampling points have been allocated to, and shape-coded according to the station classification. This way of visualization serves as a quick screening tool to envisage how the different sampling points cluster together at different levels of dissimilarity.

It should be noted that the choice of threshold for the allowed levels of dissimilarity is essential to the conclusions driven by this type of analysis. The level of dissimilarity chosen for a given application will determine the level where the dendrograms start branching into smaller clusters. The selection of the threshold needs to be carefully balanced choosing the branching that shows a clear division between different groups of data. The highest level of branching shows the clusters that have the highest similarity, within a single cluster, at that level of similarity. The clustering can also show outliers, or data points with specific behaviour that differ from the rest of the data points.

The methodology has three interesting uses that will all be tested in this feasibility analysis. These are a Tier 2 approach to monitoring network design, a Tier 3 approach to model validation and calibration, and a Tier 3 approach to monitoring design.

1) **Tier 2 approach to Monitoring network design** - With monitoring data alone, the clustering methodology allows a screening evaluation of the siting and classification of a given monitoring network based on the actual measurement data. It serves to analyse the level of similarity or dissimilarity of air concentration data from all sampling points. It allows the identification of sampling points with similar behaviour and those data points with specific behaviour that differ from the rest of the data sampling points. These different behaviours may be well justified in terms of specific conditions around the sampling points, such as influence from specific sources or specific meteorological dispersion conditions. In this sense, the identification of "outliers" in the clustering analysis is not an excluding exercise but a screening approach to identify different behaviours that need to be further investigated.
2) **Tier 3 approach to Modelling calibration and validation -** The clustering methodology can be used for validation and calibration purposes of modelling results. This is because the clustering methodology can be used with modelled results as well as with monitoring data. The comparison of the modelled and the monitoring data results from the clustering analysis provides an additional evaluation of the model performance that considers both the spatial and temporal variability within the area covered by the given datasets. When used with modelling results, the clustering methodology is to be used in the same grid points where the monitoring sampling points are located. The modelling grid data is then clustered following the Soares et al (2018) approach according to their magnitude and temporal variability. If the clustering of the modelling grid data corresponds to the clustering of the monitoring data, this provides indication of the quality of the modelling results. The comparison of dendrograms for model and monitoring data provides a complementary approach to model validation that considers both spatial and temporal variability, thus adding to the current approach to Model Quality Objectives (MQO).
3) **Tier 3 approach to Monitoring network design** Tier 3 involves the use of modelling data in addition to sampling point data. Once a set of modelling results has been considered fit-for-purpose through model validation, following the analysis of dendrograms as explained above, the modelling results can be analysed further to support monitoring network design. The hierarchical approach is then used for all grid points in the modelling domain to identify clusters of similarity with respect to modelled air quality situations. These clusters help identify areas with different air quality levels and are visualised geographically as coloured clusters. Comparison with the siting and classification of existing sampling points can help to identify redundancies in the monitoring network as well as gaps, and to provide an independent evaluation of the sampling point classification. In this way, the hierarchical clustering methodology can be used to actively support monitoring network design activities at local and national level.

2.4 Tier 4 – Combined use of modelling and observation data

In a Tier 4 approach, model results are complemented with dedicated measurement data to further improve the quality of the model output. In many Tier 3 modelling applications data from fixed measurement sampling points are already taken into account. So, in the Tier 4 approach, additional data are expected to be delivered by dedicated sensor networks or measurement campaigns. The merging process of these additional data with the Tier 3 models is obtained via so called data fusion or data assimilation procedures. In these data fusing processes, the quality of the additional measurement data, particularly where low cost sensors are used, might be an issue. In measurement campaigns with high quality monitoring devices or passive sampling techniques the measurement uncertainty is rather well known but when low cost sensors are deployed uncertainty can be large. In some of the data fusion algorithms the quality/uncertainty of the measurement data can be taken into account by allocating specific weights to the observations according to their inherent uncertainty.

At present there are no operational examples of such Tier 4 methods in the three case study cities that can be used in the sensitivity analysis. There are only results available from some research projects which are investigating the potential of sensor data in combination with modelling results in Antwerp. A result from the Dencity research project[5] is given in Figure 5 below. Observations from 3 (synthetic) sensors in the city of Antwerp are assimilated in the ATMO-street map, giving rise to modified concentration values in the vicinity of the sensor locations. This methodology is based on an Ensemble Kalman filter approach and assimilates hourly values based on given uncertainties of both the sensors and the ATMO-Street maps. However, the system is not yet deployed in an operational context.

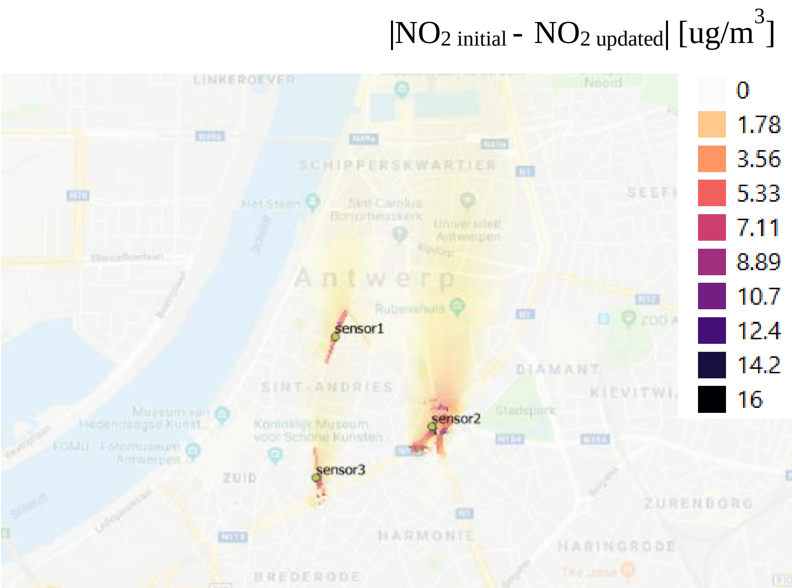

Figure 5: Example of a Tier 4 approach for Antwerp: sensor data is used to locally correct the ATMO-Street map

An example of an operational Tier 4 approach is presented for Oslo in Figure 6 that also shows how modelled air quality maps can be improved using a Tier 4 approach. The work of Schneider et al (2017) used 24 deployed low-cost sensors (AQMesh nodes) in Oslo and combines the results with the EPISODE model and geospatial data fusion techniques. An important limitation of these type of Tier 4 approaches is that they currently operate at campaign experiment level, so that the data is only available for short periods of time. For the purpose of spatial representativeness calculations, the use of longer time periods, preferably 1-year is advisable in order to secure the temporal representativeness of the results under different meteorological and source emission conditions.

[5] https://www.imeccityofthings.be/en/projects/dencity-more-sensors-in-the-city

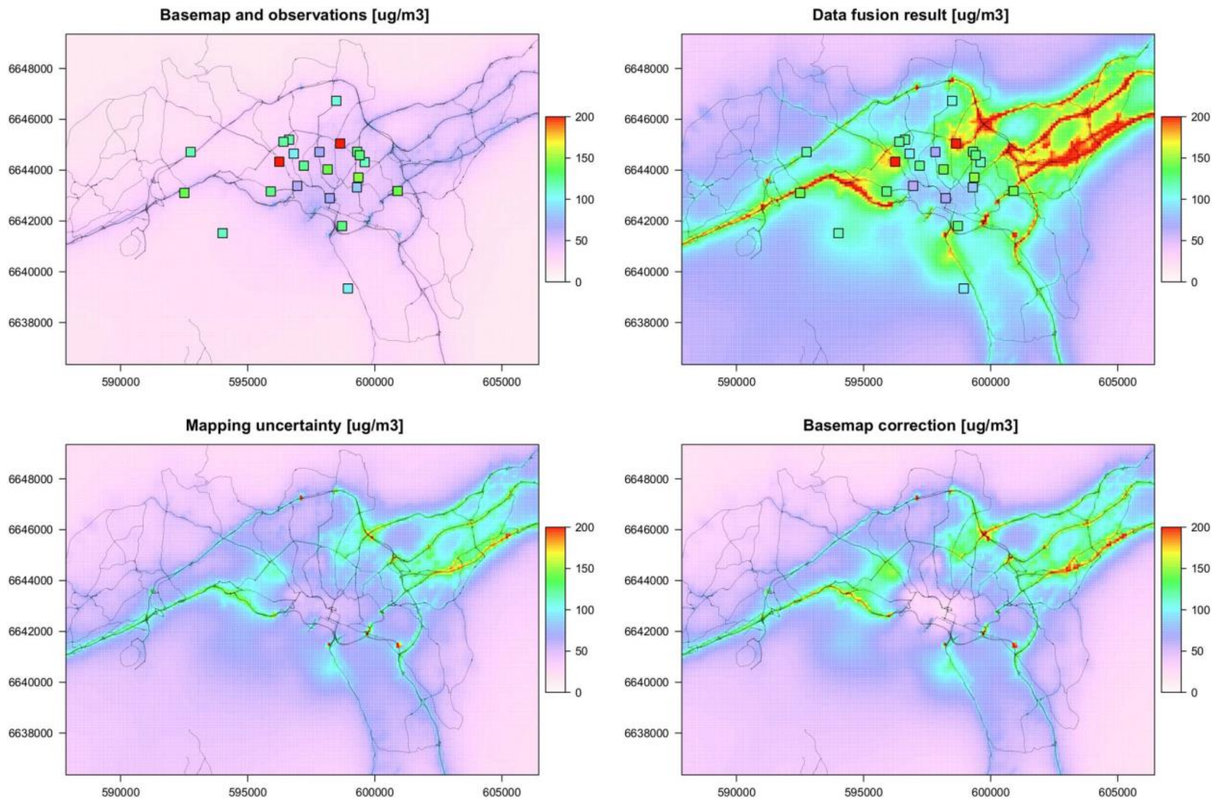

Figure 6: Example of the data fusion process combining observations from a low-cost sensor network with a modelled basemap, here shown for NO_2 on 6 January 2016 at 9:00 UTC. The top left panel shows both input datasets required for the methodology, namely the temporally constant model-derived basemap and the observations from the low-cost sensor network. Note that the color scale applies equally to the basemap and the observations, indicating that the observations exhibit significantly higher NO_2 concentrations than the basemap. The top right panel shows the result of the data fusion process as well as the observations from the low-cost sensor platforms for reference, indicating that the fused concentration field is much more in line with the observations. The bottom left panel gives the absolute uncertainty associated with the data fusion, while the bottom right panel shows the correction that the data fusion applied to the basemap to achieve the result. All concentrations shown in units of µg m^{-3}. The faint thin black lines indicate the major roads in Oslo for reference. Coordinates in UTM32N/WGS84 (Schneider et al. 2017).

3 Sensitivity tests for spatial representativeness area of monitoring sampling points

Before focussing on sensitivity tests for specific assessment needs, we set out some general concepts of the SR area of monitoring sampling points. According to the IPR Article 12 and Annex II, Member States are obliged to report the area, population and road length in exceedance for every air quality zone, where an exceedance of an environmental objective of the AAQD has occurred. Since the reporting obligations are driven by exceedances observed at station locations, there is a clear link between the exceedance situation indicators and the SR area of monitoring sampling points. This analysis will therefore help to make deliberate choices in the sensitivity tests for the specific assessment needs in Chapter 4 and 5.

As stated in the Literature Review (Maiheu and Janssen 2019), the exceedances of the environmental objectives of the AAQD on NO_2 and PM_{10} either concern directly, or can indirectly linked to annual averaged pollutant concentrations. Consequently, the temporal dimension is by definition fixed for these indicators and the related SR similarity criteria. In this chapter, we therefore focus on sensitivity tests dealing with spatial variations of annual mean concentrations. The goal of the analysis is twofold. First, we want to analyse the behaviour of an SR area of a monitoring station with respect to: choice of similarity criteria and related threshold values and the difference between a contiguous versus a non-contiguous approach. These are methodological options in the SR assessment which would benefit from further guidance to establish a harmonized SR assessment framework. Further, special attention is devoted to an intercomparison of the different Tiers and how they impact the overall quality of the SR assessment process.

3.1 Methodology

To tackle the first objective of this analysis, i.e. investigate methodological options in the SR definition, we could rely on a wide variety of SR assessment methodologies as described in Chapter 2. However, at the end of the Intercomparison Exercise process (Kracht et al, 2017), there was a general consensus amongst the participating teams that any methodology relying on a fit-for-purpose modelling application is preferred over the simpler expert opinion or proxy data based approaches. Even the lead author of the Tier 2 UBA methodology (Spangl et al, 2007), stated at the end of the IE process that proxy based approaches lead to results which are by construction inferior to model based results which explicitly take into account our best available knowledge of emission sources and dispersion and transformation processes in the atmosphere. In addition, it was concluded that many of the Tier 2 methods require a substantial effort to collect all relevant input data. This input data is also needed to run an atmospheric dispersion model and therefore it was already recommended at the end of the IE to focus on model based (or Tier 3 in this context) approaches for SR assessment. For this reason, the first part of the methodological sensitivity analysis will be solely based on the Tier 3 results. In a second part, a comparison of different Tier results will be carried out to illustrate the strengths and weaknesses of the Tier levels.

The Tier 3 analysis starts from gridded mean concentration maps for 2018: 10m resolution ATMO-Street maps for Antwerp and Krakow, and 100m resolution maps of EPISODE for Oslo. For each monitoring location, the related SR area is defined as the grid cells in the air quality zone which satisfy a similarity criterion concerning the *modelled* yearly mean concentration. Stated another way, the SR area is defined as the collection of grid cells for which the difference between the *modelled* concentration in the grid cell and the *modelled* concentration in the sampler location is smaller than a threshold value. By using the modelled concentration for both locations, we correct for the model bias[6]. For all three case studies, we

[6] The analysis has been repeated using the difference between the *modelled* concentration at each grid cell and the *measured* concentration at the monitoring location. Although the details of the results differ, the main conclusions of the analysis remain unchanged. Obviously, this conclusion only holds for fit-for-purpose models with a small bias. If larger biases are present in the model, this might hamper the results.

have applied the method to all available measurement sampling points in the air quality zone related to the city centres of Antwerp[7], Oslo and Krakow. In all cases we limited the maximal extent of the SR area to the spatial extent of the air quality zone in which the sampling locations are located; Obviously the areas are also limited by the extent of the model results if applicable.

There are two major options to delineate the SR area: a contiguous or a discontiguous approach. The merits of these approaches have been much debated in the report of the IE exercise and subsequent FAIRMODE meetings. In a contiguous approach, there are two requirements which must be met by a grid cell to belong to the SR area. The grid cell should of course satisfy the similarity criterion (i.e. the difference in annual averaged concentration between the grid cell and the sampling location should be smaller than a threshold value). In addition, there should exist an uninterrupted path of grid cells that satisfy the similarity criterion from the grid cell to the monitoring location. In the discontiguous approach, the second requirement is not required, and all the grid cells in the air quality zone which satisfy the similarity criterion are retained. Figure 7 visualises both options and some of the sensitivity tests will further highlight the striking differences between both approaches.

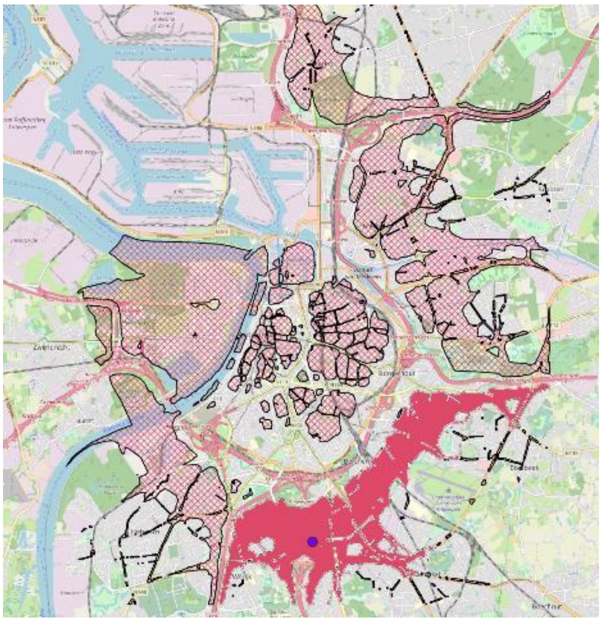

Figure 7: Illustration of the concept of contiguous vs discontiguous SR areas. The figure shows all the locations which satisfy the similarity criterion with a threshold of 10% for the station Antwerpen-Groenenborg. The station location is indicated with the blue dot. The fully shaded polygon represents the contiguous area. The discontiguous zone is composed of both the fully shaded area and the hatched area, which represents all the locations where the similarity criterion is met.

For a similarity criterion based on annual averaged concentration we can either apply an absolute or a relative difference in concentrations. In the former option, the SR area is formed by all locations for which the absolute concentration difference is smaller than the threshold, while for the latter option the relative difference (i.e. the absolute difference divided by the modelled concentration at the sampling location) is used. In section 3.2 various threshold values are tested for both types of similarity criteria (absolute or relative). A broad range of thresholds has been considered, from very low thresholds in the order of the model uncertainty level (relative threshold 5%, absolute threshold 1µg/m^3) to very large thresholds (50% / 15 µg/m^3) for which most of the domain will be covered for an urban background station. These values have been considered both for NO_2 and PM_{10} concentrations.

[7] We thus neglect the zone 'Port of Antwerp'. As the main goal of the analysis is to provide insight in the exposure and exceedance of annual mean thresholds, this zone is less important. There are only a few inhabitants, and no exceedances have been reported in recent years.

In summary, there are three degrees of freedom which are being investigated in this sensitivity analysis: whether to allow discontiguous zones or not; the choice between an absolute or relative threshold; and the actual value of the threshold in the similarity criterion. On top of that the three different cities, their topologies and pollution levels and the different modelling chains (with different final resolutions) also provide variability to further test the generality of all findings.

3.2 Results

In this section we discuss the results of the sensitivity tests for assessing an SR area based on various methodological options. We start with a visual, qualitative description of the SR areas for a few interesting locations and afterwards quantify the results for all available sampling points in the zone.

First of all, it should be noted that there are major differences between the SR areas for NO_2 and PM_{10}, and between the areas for urban background and traffic sampling points. The analysis will thus be split into these subcategories. Our main goal is to highlight the key messages of the analysis, and we thus only provide results for selected sampling points and selected parameter values. All other results (other sampling points, other threshold values) have been analysed thoroughly, and they back-up the main findings presented in this report.

3.2.1 Qualitative analysis

Figure 8 and Figure 9 map the SR areas for a selected urban background and traffic station for both NO_2 and PM_{10}. In Antwerp, we have opted for the urban background station R811 – Antwerpen Groenenborger, located in a suburban area in the south of the city centre, and for the traffic station R804 – Ring, located just besides the busiest highway in Antwerp. In Oslo, the urban background station Akebergveien is used together with the traffic station Hjortnes, positioned beside one of the main highways. In Krakow, we identified the suburban station PL501A located in the south of the city centre and the traffic station PL641 is in the busy street canyon Dietla street, located in the city centre. In each of the cases, a relative threshold value of 10% is applied. This choice is based on a first estimate according to expert opinion. In section 3.2.3, a quantitative analysis focuses on the optimal choice for the threshold values.

The area of representativeness for the contiguous option is often limited to the station and its immediate surroundings, while the discontiguous approach yields larger areas in which also locations further away from the station are included. At first sight, the results for the contiguous method seem confined to specific regions in the domain, both for the urban background and the traffic sampling points.

For traffic sampling points, the contiguous SR area is limited to the station and its immediate vicinity. Note that in some cases the SR areas are barely visible because they only encompass a few model grid cells. Other locations that are strongly influenced by traffic (e.g. locations near other highways or in other street canyons) are included in the discontiguous areas, but they are excluded from the contiguous areas because they are separated by locations with lower concentrations. These effects are observed for PM_{10} and NO_2.

A similar effect is observed on a larger scale for the urban background sampling points. This is most clearly visible for NO_2 for the urban background sampling points in Antwerp and Krakow. In both cases, the domain is divided in several subdomains by major highways which provide “concentration barriers” for the contiguous area. As a result, the SR area for the station in the south of Antwerp is limited to the zone between the highway in the south-west and the highway in the east of the domain. Although the concentrations at the other sides of the highway are similar, the contiguous area can never include them because of the “concentration barriers” shielding them from the station location. The effects do not only arise for NO_2, but also for PM_{10}, as can be seen on Figure 9 for Antwerp and Oslo. In both cases the locations in- and outside of the ring roads are shielded from each other by concentration barriers. Finally, note that these findings are observed in all three case studies, and for modelling chains with and without street canyon contributions. As the difference between using either a contiguous or discontiguous approach is compelling, the effect is quantified and discussed it in more detail in the next section 3.2.2.

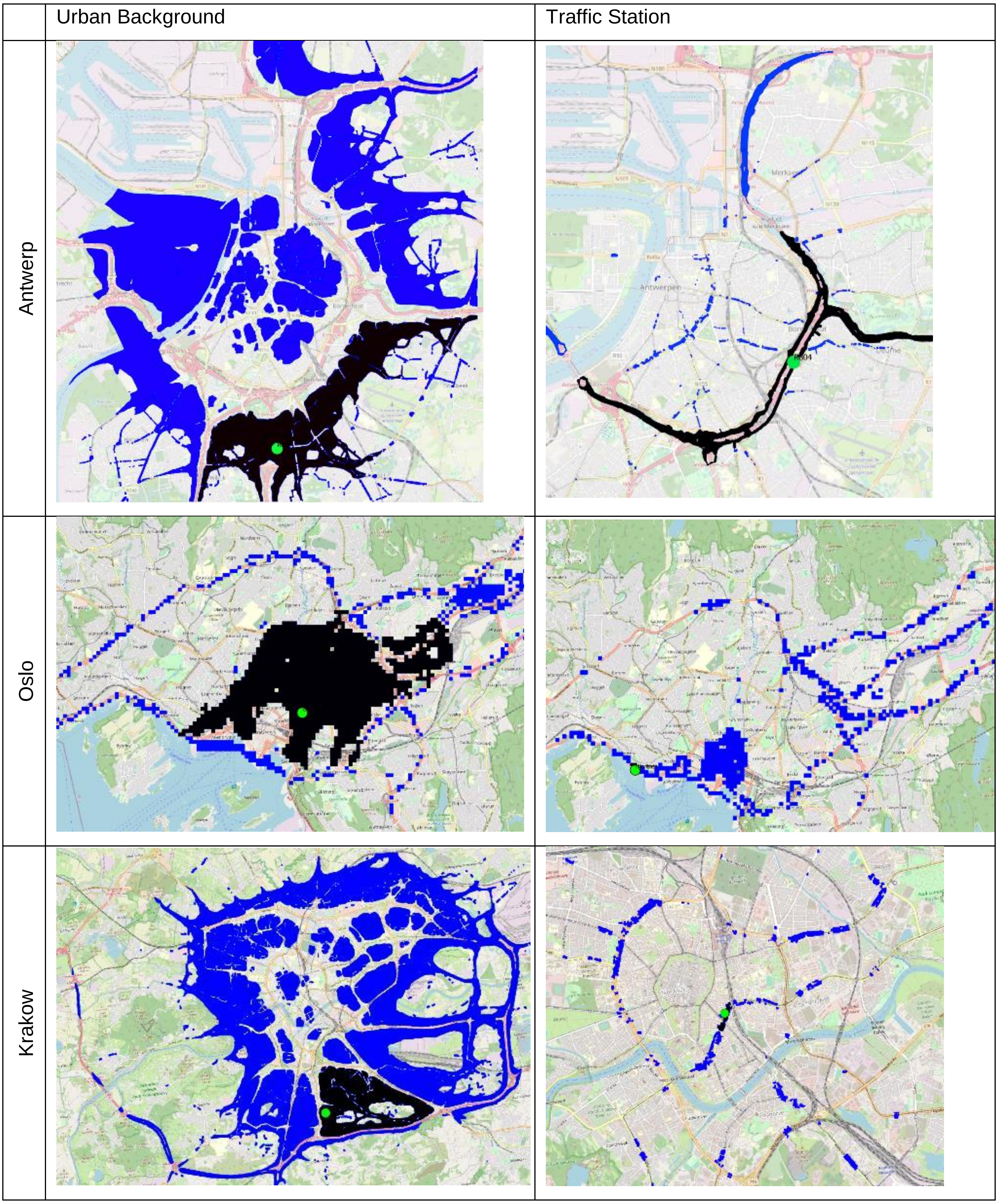

Figure 8: SR areas for NO_2, for an urban background station (left) and a traffic station (right). In Krakow this refers to a street canyon location, in Antwerp and Oslo it refers to a station close to a major highway. The contiguous representativeness area is represented by the black area, the related discontiguous area consists of the combination of the black and the blue area. The station under consideration is always indicated with a green dot. For all figures, a relative threshold of 10% is applied.

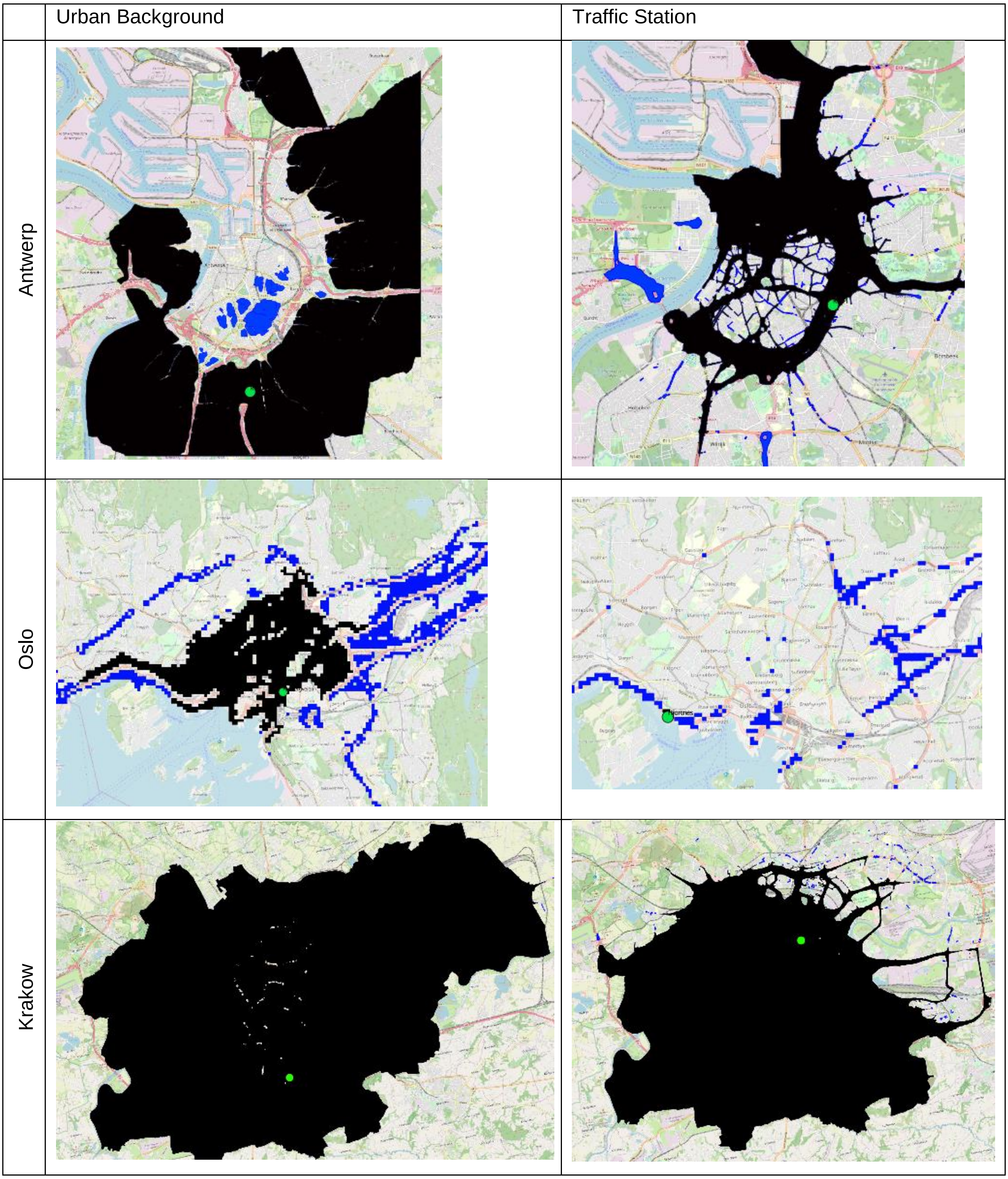

Figure 9: SR areas for PM_{10}, for an urban background station (left) and a traffic station (right). In Krakow this refers to a street canyon location, in Antwerp and Oslo it refers to a station close to a major highway. The contiguous representativeness area is represented by the black area, the related discontiguous area consists of the combination of the black and the blue area. The station under consideration is always indicated with a green dot. For all figures, a relative threshold of 10% is applied.

Another feature of an SR area concerns the typology of the locations included in a single SR area (at least when considering the discontiguous option). When a similarity criterion based solely on an annual mean concentration is applied, the influence of other topographical features such as the proximity to buildings or emission sources is not directly taken into account. As a consequence, the SR area consists of locations with different typologies. The effect is illustrated in Figure 10 for the traffic station in Antwerp. Both pictures are taken at locations which belong to the same SR area of the traffic station when a 10% threshold is applied (both for PM_{10} and NO_2). One is a street canyon (modelled with the OSPM model), the other is an open highway (modelled by the IFDM Gaussian dispersion model). Although similar annual mean concentration levels occur at both locations, they represent completely different air pollution contexts. Note that the effects also occur in the urban background SR area, which contains both open urban background locations and street canyons located in more rural locations. One could avoid this mixing of typologies in one SR area, by extending the primary similarity criterion with a secondary criterion related to, for example, typology. In such a case the street canyon and the open street location could be considered as different locations not representative of each other. However, this clearly complicates the SR analysis, and requires answers to difficult questions such as “What is a street canyon?”, “What about half open locations?” etc., and most probably does not lead to any additional insights. We therefore suggest using as simple as possible similarity criteria, and therefore limit to criteria that are only based on the modelled concentrations.

Figure 10: Illustration of two locations that belong to the same SR area of the traffic station R804 – Ring in Antwerp when a relative threshold of 10% is applied. Source: Google Maps.

The above discussion about canyons is only relevant when models explicitly consider street canyon effects. The absence of street canyons in the Oslo model chain also causes some other differences between the three case studies. When considering the modelled NO_2 concentrations presented in Figure 3 and Figure 4 it can be seen that in the city centre of Antwerp and Krakow the spatial variation of the modelled concentrations is very high (see Figure 3). The concentrations vary over very short distances, and very high gradients are encountered in and near street canyon locations. In contrast there is less variation in the modelled concentrations in the city centre of Oslo. Consequently, the SR areas in Oslo contain much less heterogeneity in comparison with the city centre of the other two cities.

When comparing Figure 8 and Figure 9 the difference between particulate matter (PM) and nitrogen dioxide (NO_2) also becomes apparent. Although the same relative threshold has been applied, the SR areas are much smaller for NO_2 in comparison with PM_{10}. These results are related to the nature of both pollutants and are especially influenced by the smaller concentration gradients for PM compared to NO_2 (which is in turn caused by the stronger dependence of NO_2 on traffic sources)[8].

In the analysis we have limited the extent of the SR areas to the extent of the air quality zone under consideration. This deliberate and pragmatic choice is in line with the IPR requirements where SR areas and related indicators must be reported per air quality zone. Also, when this spatial limitation is dropped, the discontiguous SR areas could become very large: for example, the 10% threshold PM_{10} SR area for the urban background station in Antwerp would encompass an area of 10,900 km^2 or 80% of the Flanders

[8] Note that the traffic results for PM_{10} in Antwerp and Krakow should be treated with care, as the modelling chain underestimates the repeated resuspension of dust in street canyons. Moreover, residential emissions are only considered via the background concentration, and potential local hotspots due to the e.g. wood or coal burning are not explicitly modelled.

region. Even when the SR area is limited to the air quality zone, it can still be very large, or contain locations located far away from each other. These large areas, for instance, occur for very large air quality zones, or for discontiguous air quality zones. An example of the latter one is the air quality zone “Cities with more than 50,000 inhabitants, except Ghent and Antwerp” in Flanders, Belgium. As its name suggests, the zone combines the city centres of several medium-sized cities. It contains cities both in the west, centre and the east of the Flemish Region. An example of its SR area is shown in Figure 11. The contiguous SR area in black is clearly only limited to the immediate surroundings of the air quality station, while the discontiguous approach in blue adds urban background locations in and around other cities as well. According to the contiguous option, an urban background station would be required in each of these cities to cover all the locations thus invoking the need for many additional air quality monitoring locations.

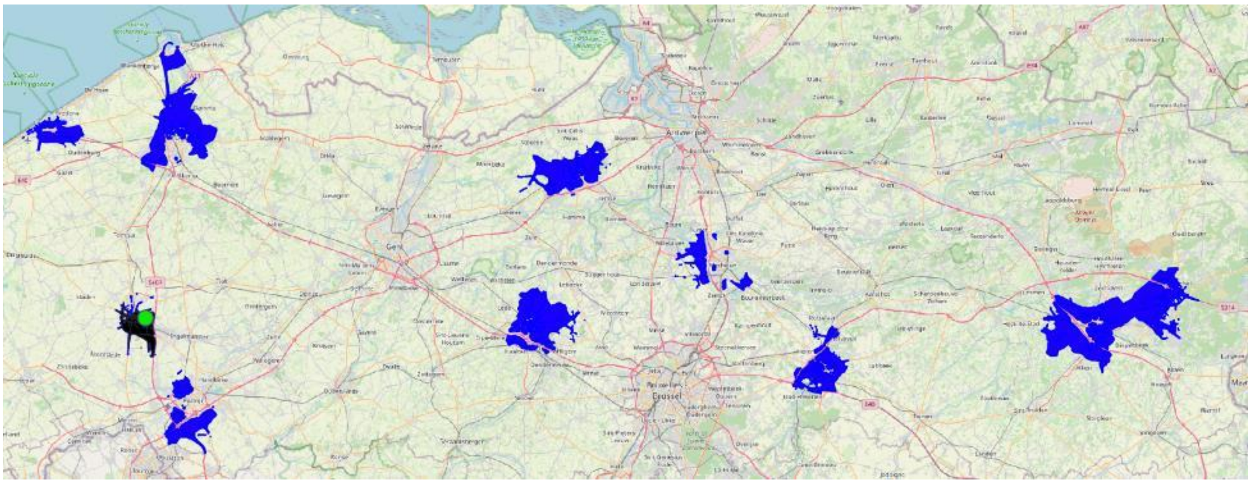

Figure 11: SR areas for NO_2, for an urban background station in the zone “Cities with more than 50.000 inhabitants except Antwerp and Ghent”. The contiguous SR is represented by the black area, the discontiguous area consists of the black and the blue area. The station under consideration is indicated with a green dot. A relative threshold of 10% is applied.

3.2.2 Quantitative analysis: contiguous versus discontiguous option

In the previous section we demonstrate the differences between a contiguous and a discontiguous approach. It was argued that in the contiguous option SR areas can be too small because they are hindered by concentration barriers, while the discontiguous option results in areas that can appear quite disconnected. In this section, these qualitative remarks are substantiated with a quantitative analysis of the size of the zones, and meanwhile some other drawbacks of the contiguous approach are highlighted.

Figure 12 quantifies the size of the NO_2 SR areas as a function of the threshold value for urban background and traffic sampling locations. For each city, we show results for similar sampling points. The graphs demonstrate that very small SR areas are derived for traffic sampling points if only the contiguous approach is applied. Depending on the traffic station under consideration, at least a 10 to 30% threshold must be applied before the SR areas extends further than the immediate surroundings (order of hundreds of meters, which is in line with the AAQD requirements) of the station. When the discontiguous option is applied, there is a more gradual increase, and even for small thresholds (<10%) locations further away from the sampling location are included.

The size of contiguous SR areas is moreover much more sensitive to the threshold than the discontiguous areas. While the size of the latter increases gradually and smoothly with increasing thresholds, the former observes jumps in the size of the area when concentration barriers are passed. The effect is visualized in Figure 13, which shows the SR areas for the urban background station PL501A (Bujaka) in Krakow for two thresholds. Although the threshold only increases slightly from 12.5% to 15%, the extent of the SR area increases dramatically as the concentration barriers are crossed. These effects are of course absent for the discontiguous option, as only the similarity criterion and its related threshold value is important for this option.

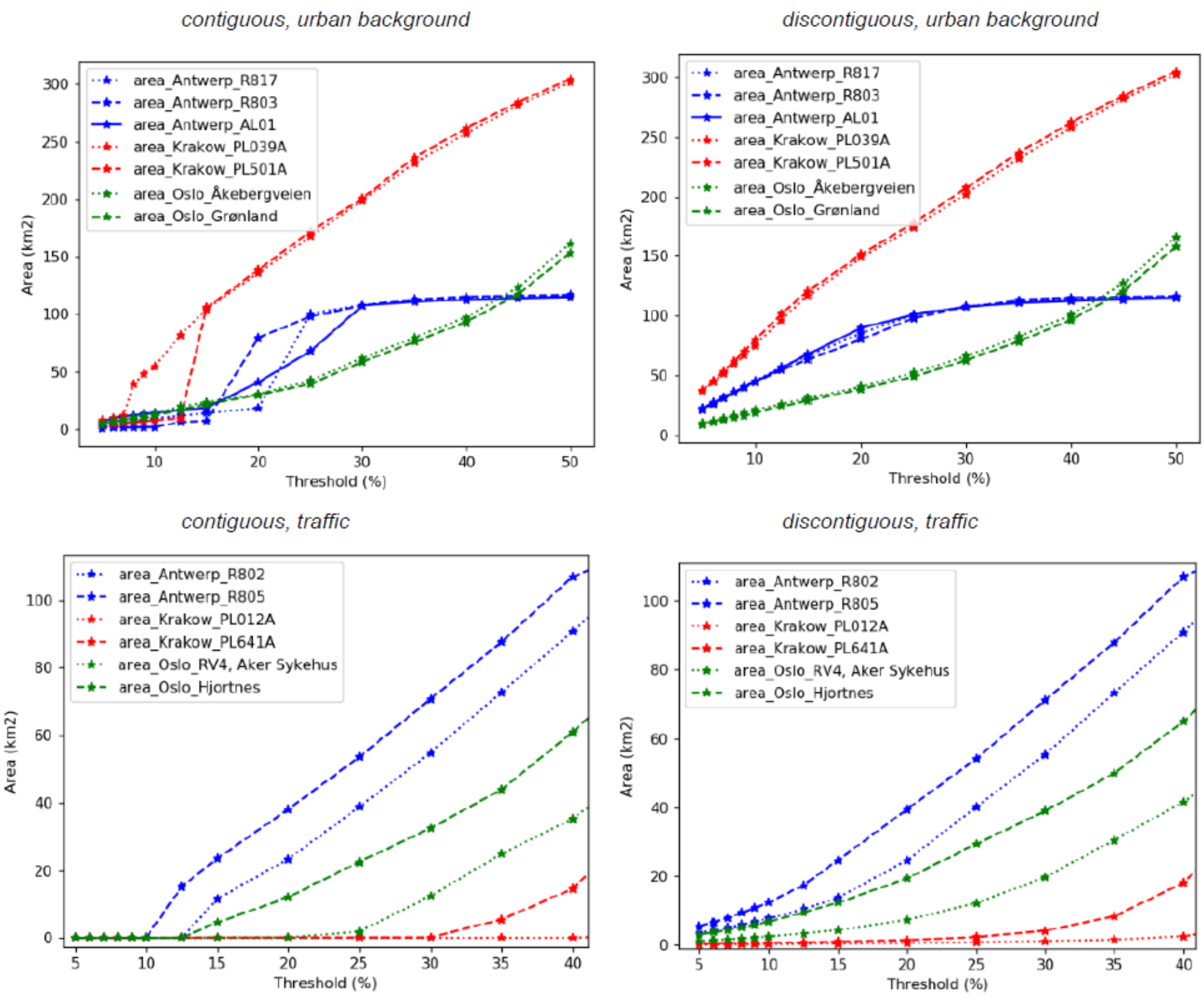

Figure 12: Comparison of the contiguous (left) and the discontiguous approach (right) for NO_2. The figures show the size of the SR areas as a function of the relative threshold. The top row provides results for selected urban background sampling points, the middle and bottom row provides results for traffic sampling points. The bottom row provides a zoom on the results for small thresholds for the traffic stations. The colouring refers to the city: Antwerp (blue), Krakow (red) or Oslo (green).

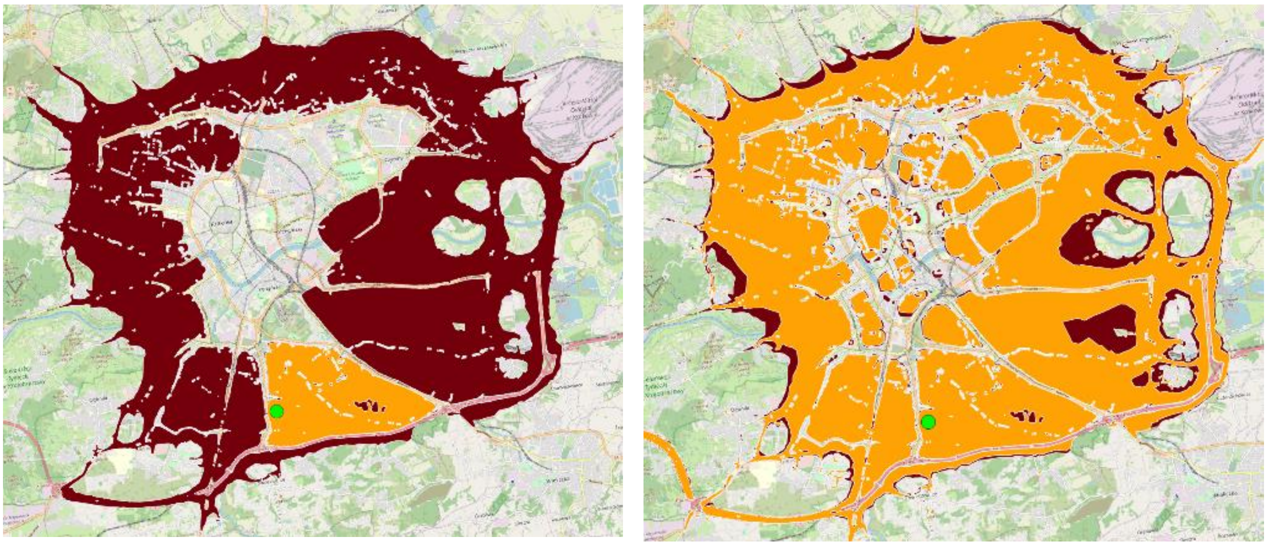

Figure 13: Illustration of the sensitivity of the contiguous (left) and discontiguous (right) SR areas to minor changes in the threshold. The figure shows results for the station PL501 (location indicated with green dot) in Krakow. The orange shaded area indicates the extent of the SR areas when a relative threshold of 12.5% is applied, the brown shading delineates the area that is added to the SR area when the threshold is 15%. The figures show results for NO_2.

Furthermore, when the same threshold is applied to different urban background sampling points, the resulting SR areas differ in size when applying the contiguous option, while they are much more consistent when using the discontiguous approach. The effect becomes especially clear when focusing on the urban background sampling points in Antwerp (see Figure 12). When applying a 20% threshold, there are vast differences in the size of the contiguous SR areas, because for one of the sampling points all the concentration barriers are already exceeded (R803, Park Spoor Noord), while for other sampling points some (AL01, Linkeroever) or all (R817, Groenenborgerlaan) of the barriers are still present. When the threshold is increased to 25%, the behavior is entirely different. The concentrations barriers are now breached for R803 and R817, while there are still some barriers for AL01. For the discontiguous option, the size of the SR areas is similar for all these sampling points, and the behaviour of the three sampling points is similar for the 20% and 25% threshold.

We have thus identified several drawbacks of the contiguous approach. Although the exact details of the results strongly depend on the city typology and the air quality data used in the analysis, the general effects arise in all three case study cities. As the model data and the network layouts used in this analysis span a wide range (10m street canyon versus 100m open street, highways very close or much further away from the city center), it is safe to assume that the drawbacks of the contiguous option would arise in any SR assessment exercise. All these issues mainly arise for NO_2. When solely focusing on particulate matter, the drawbacks of the contiguous approach are of less importance, as the analysis of the size of the SR areas in Figure 14 indicates. For particulate matter, there are simply less concentration bridges in the domain, although they are not completely absent (see for instance the example for the urban background station Åkebergveien in Oslo in Figure 8). Note that both graphs in Figure 14 look very similar although small differences appear, especially for low thresholds.

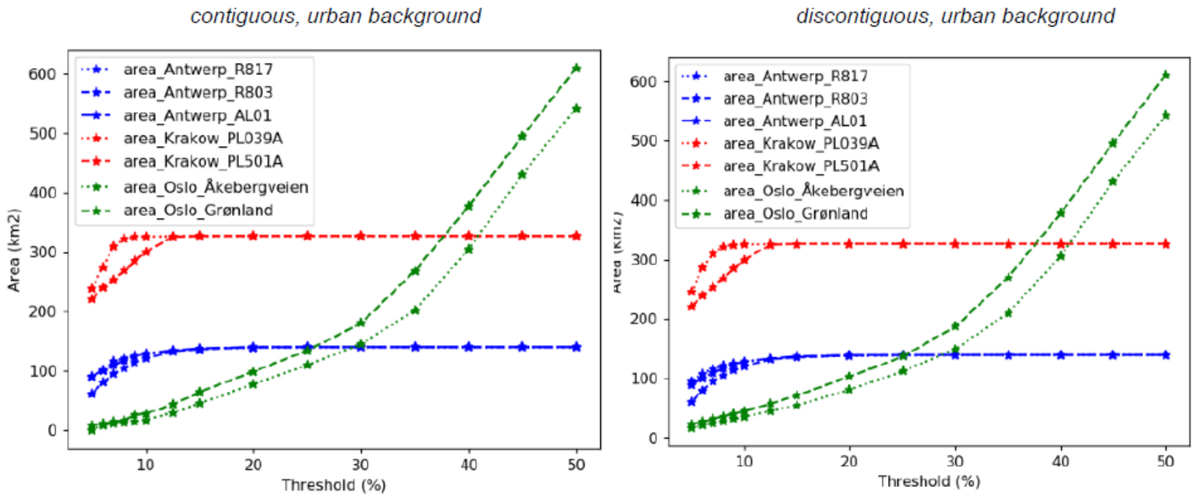

Figure 14: Comparison of the contiguous (left) and the discontiguous approach (right) for PM_{10}. The figures show the size of the SR areas as a function of the relative threshold for selected urban background sampling points. The colouring refers to the city Antwerp (blue), Krakow (red) or Oslo (green).

3.2.3 Quantitative analysis: Similarity criteria and threshold analysis

Another important open question concerns the choice of similarity criterion and the optimal threshold value. The graphs in Figure 12 and Figure 14 show the relationship between the size of the SR areas and the relative threshold that has been used. We only focus on the discontiguous approach (right figures) in this section.

Choosing an optimal or preferable threshold for a spatial representative analysis is of course a complex issue, but the results in Figure 12 and Figure 14 provide some insights to support the decision. Under a discontiguous approach, the SR areas increase gradually in size, and, consequently, the exact value of the threshold is less important than its order of magnitude. For instance, the difference in the SR area is rather limited when applying a 12% or a 15% threshold value. If the thresholds are too small, the SR areas are limited to locations with concentrations that are very similar to the (modelled) concentration at the station locations. The SR area increases gradually as the threshold increases, and, at some point, the SR areas start to include almost the entire air quality zone. In this case locations are included that are no longer similar to the (modelled) concentration at the position of the station. An optimal value for the thresholds clearly lies in between those two extremes. Comparing the results for the different cities in Figure 12 and Figure 14, the authors suggest an optimum value lies in the range from 10% to 20%.

As an alternative to the relative thresholds, absolute thresholds (in concentration units as opposed to using percentages for relative thresholds) can also be applied, as shown in Figure 15. For NO_2, the difference between both approaches is limited because the concentrations are similar in the different case studies[9] (and hence also the relative thresholds are similar). A 10% threshold thus roughly amounts to the same absolute difference in NO_2 for all three case study cities. This is not the case for PM, however, because there are large differences observed in the PM_{10} urban background concentrations. As a result, the graphs for PM_{10} differ strongly when considering absolute (Figure 15, right hand side) or relative thresholds (Figure 14, right hand side). Based on these findings, we recommend using relative thresholds instead of absolute thresholds, as it is much easier to compare different cities when using the relative approach.

[9] Note that in other cities (where the urban background NO_2-concentrations may be lower or higher), the graphs could have a different behaviour.

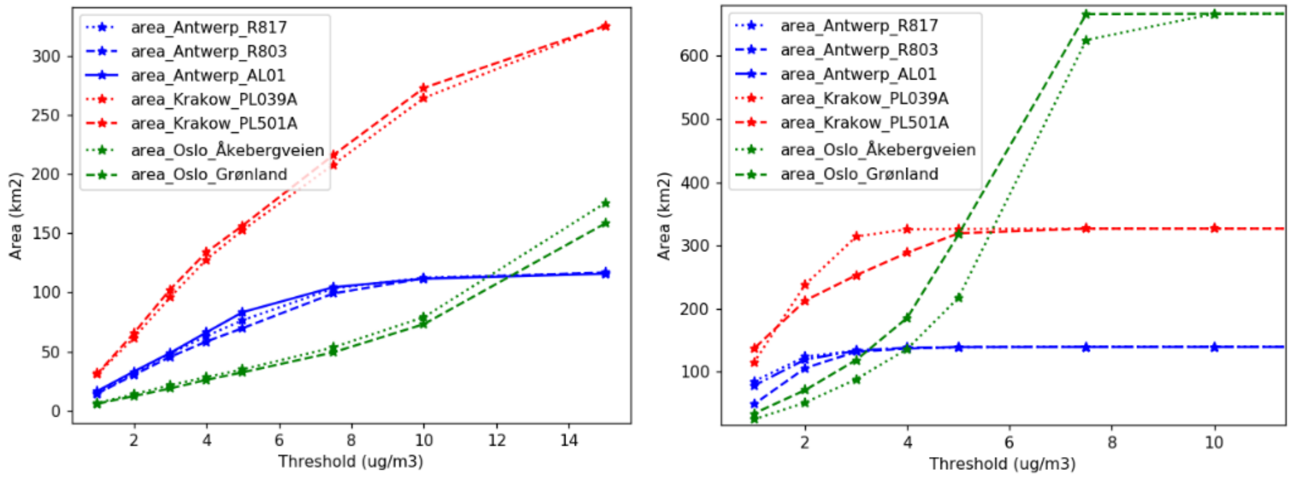

Figure 15: Results for absolute thresholds (in µg/m³) for some urban background sampling points, using the discontiguous approach. The left figure shows the size of the SR areas as a function of the absolute thresholds for NO_2, the right figure shows similar results for PM_{10}.

3.3 Comparison between the different Tiers

Relying on the insights obtained in the Tier 3 based sensitivity analysis, we now compare the results of the different Tiers. Since Tier 2 results are only available for Antwerp, we only provide results for this case study. For clarity, the analysis is limited to one urban background station (R817) and one traffic street canyon station (R802). Selected SR areas for all Tiers are mapped in Figure 18 (urban background, NO_2), Figure 16 (traffic station, NO_2) and Figure 20 (urban background, PM_{10}). We do not provide results for PM_{10} at the traffic sampling points as the model results for Antwerp are known to underestimate the repeated recirculation of particles in the canyon (Berkowicz 2000b; Lefebvre et al. 2013b). As this is the case for most models, it means that assessing SR for PM_{10} at traffic sites is currently very difficult, if not impossible with state of art methods (as used in Tier 3 or 4). Additionally, Figure 17 and Figure 19 quantitatively compares the size of the SR areas for different parameterisations of the Tiers, respectively for NO_2 and PM_{10}

Tier 1 results include the fixed radius method and a delineation of the minimal SR of a traffic station according to the AAQD. The AAQD stipulates that “the air sampled is representative of air quality for a street segment no less than 100 m length”. For urban background sampling points the requirements of the AAQD are less precise (“representative for several square kilometres”), and we thus only report the fixed radius method for urban background locations. For each station type, we have selected three options for the fixed radius. For the traffic station, 100m, 500m and 1km have been retained, while for the urban background we used somewhat larger radii: 1km, 5km, 10km. Note that these parameter choices are the same for both pollutants.

3.3.1 NO_2 traffic station

When the SR areas of the three Tiers are compared, a vast range of results is observed. Particularly for the traffic station (Figure 16), there is not much agreement between the different Tiers.

The Tier 3 SR areas combine street canyon locations with those close to major roads and highways (bottom row in Figure 16). The size of the areas of course increases if the threshold increases, but for all threshold values there is a large spatial heterogeneity. Although the Tier 1 fixed radius method presents much smoother results, its results are only to a limited extent useful for delineating a correct SR area for the traffic station (left figure on top row in Figure 16). The smallest radius (100m) on the one hand leads to a very small zone, while larger radii mix all types of locations, including locations that have a very distinct concentration according to the model results. The 1000m radius area for instance combines several street canyons with an urban park (Antwerp Zoo) with much lower concentrations. In addition, applying Tier 1 and the minimum requirement of the AAQD to cover a street segment of 100 metres leads to a very small area, because it is based on the *minimal* area required by the AAQD (right figure on top row in Figure 16).

The results of Tier 2 are strongly influenced by the underlying methodology (middle row in Figure 16). The land use proxy method relies on CORINE land use on a 4 by 4 km resolution, and thus includes the entire city centre in the SR area, without considering urban parks, water bodies etc (middle figure on middle row in Figure 16). The emissions method correctly identifies the locations of the street canyons around the traffic station, but it does only so for a limited area (500m x 500m) (left figure on middle row in Figure 16). Finally, the results of the Spangl method are very difficult to interpret (right figure on middle row in Figure 16). In general, it seems that the Tier 2 approach does not bring much added value in comparison with Tier 1 for the traffic sampling points.

This qualitative description is substantiated by the analysis of the size of the SR areas in Figure 17 (right plot). The results of the land use method, a 5km radius and the 35% threshold of Tier 3 are in line with each other and are much larger than the other options. On the other hand, the areas obtained with the 100m and 500m radius, the emissions and the Spangl method are much smaller than the other methods. In between lie the 1km radius and various options for the relative threshold applied to Tier 3. Also, the Tier 2 results seem to contribute only limited additional value in comparison with the Tier 1 approaches with medium radii.

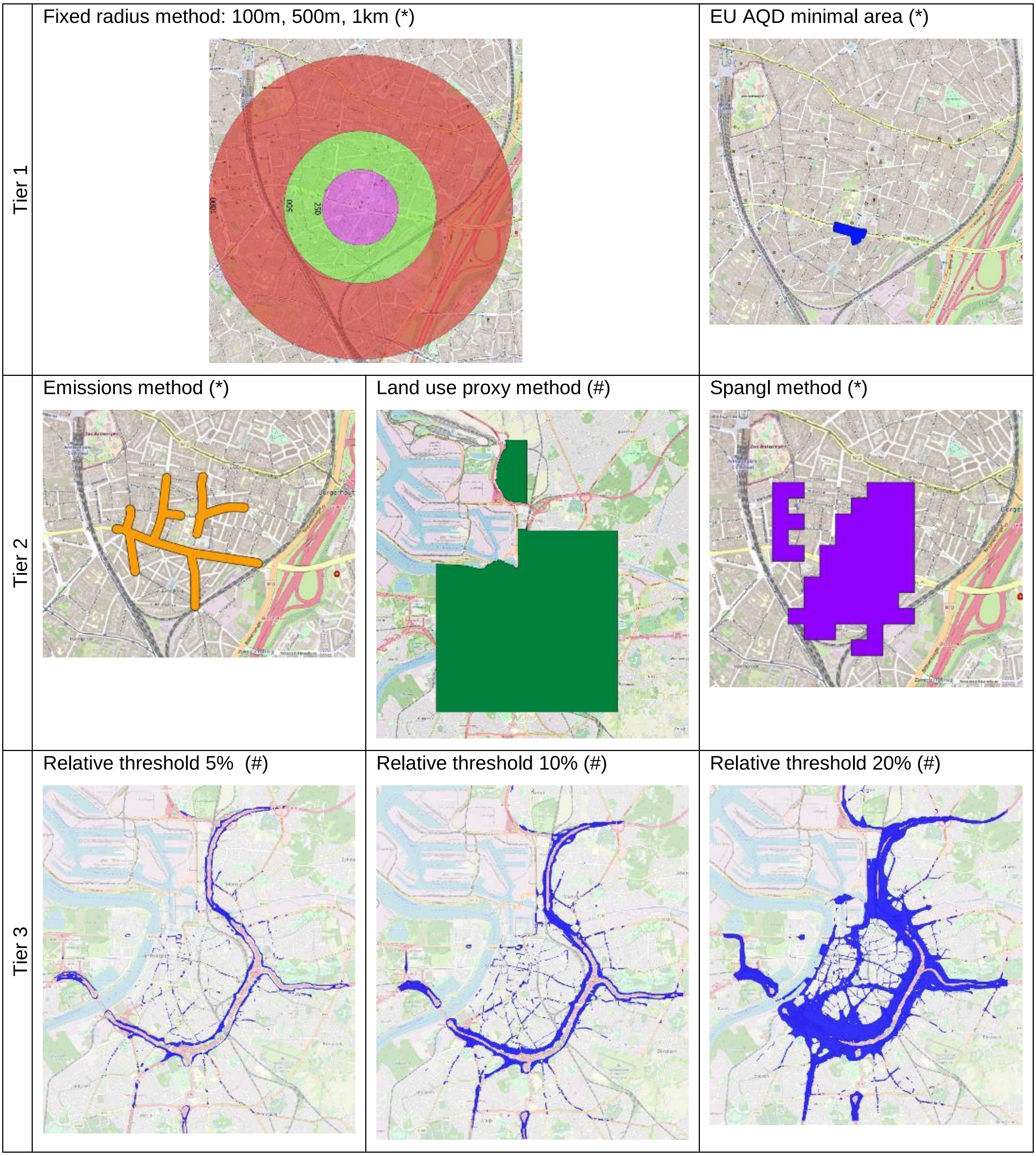

Figure 16: Comparison of the SR areas for the street canyon traffic station Borgerhout Straatkant (R802) in Antwerp. As the areas vary greatly in size between the methods, a small (*) and a large (#) extent have been used. Results for NO_2. More details are provided in the main text.

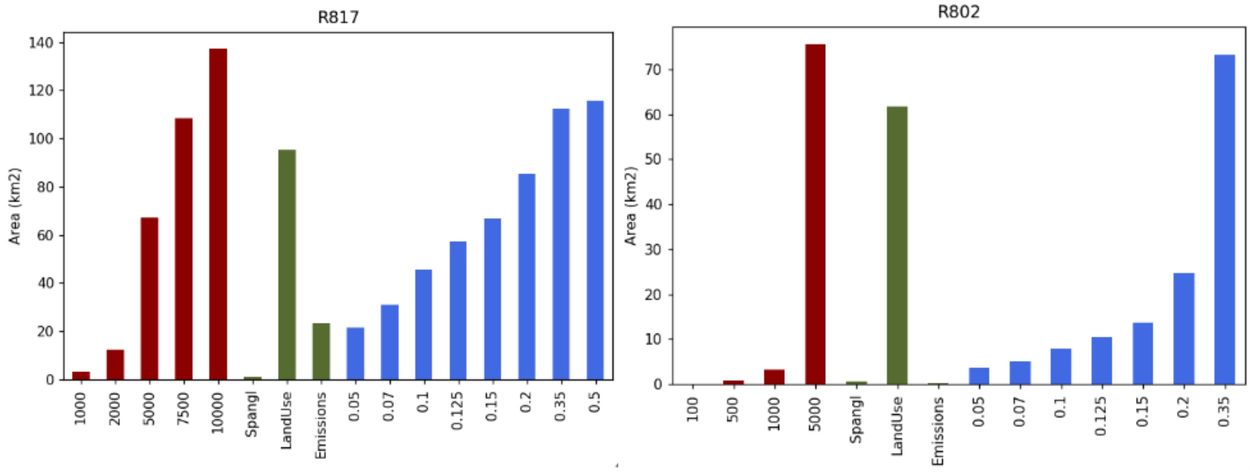

Figure 17: Size of the SR areas for NO_2, for an urban background station (left) and a street canyon station (right) according to the three Tiers considered in this analysis. In red the results for the fixed radius method of Tier 1, with the labels indicating the radius (in m). In green the three Tier 2 methods, with the labels indicating the type of methodology. The results for Tier 3 are shown in blue, with the labels indicating the relative thresholds values.

3.3.2 NO_2 urban background station

For urban background locations, there is more of an agreement for NO_2, at least when leaving out some extreme methods (Figure 18).

The Tier 1 approaches with the small fixed radii (< 1km) yield very small SR areas (top figure in Figure 18). On the other hand, the Tier 1 approaches with larger fixed radii provide zones that mix all types of locations. The 5km zone includes both urban background locations and some very busy parts of the Ring Road, for which the concentrations are much larger than those at the urban background station. When a 10km radius is applied, the entire city centre, including the Ring Road and many street canyons are also included in the zone. The Tier 1 fixed radius approaches thus fail at some point in collecting locations of similar representativeness.

There is much more agreement between the Tier 2 (middle row on Figure 18) and Tier 3 methods (bottom row on Figure 18). Only the Spangl method stands out since it is limited to a 1100x1100 m^2 area around the sampling locations (right figure on middle row of Figure 18), and thus yields a SR area that is much smaller than most of the other methods. The other Tier 2 and the Tier 3 methods yield SR areas that form a broad band around the city centre of Antwerp. The exact details of the band depend on the nature of the methodology and thresholds applied, but the main features are the same. Hence, some of the approaches in Tier 2 are providing additional value over the Tier 1 results for the urban background sampling locations.

The qualitative description is once more backed up by the quantitative analysis of the size of the SR areas (see Figure 17, left plot). The Spangl method and 1000m fixed radius approach stand out because of the very small area. For all other approaches, the size of the areas falls in the same range, with the exact details depending on the choice of the parameters.

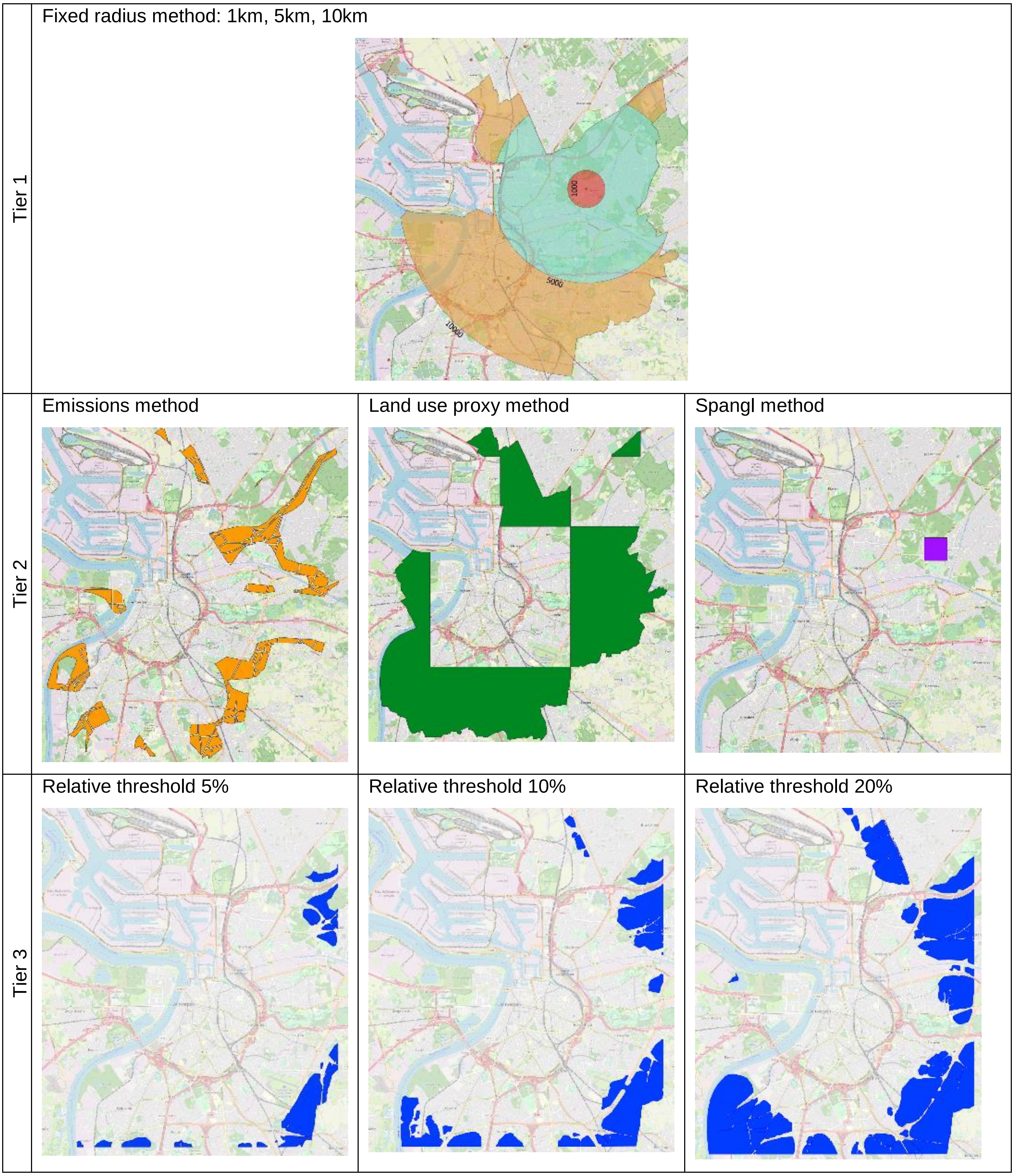

Figure 18: Comparison of the SR areas for the urban background station Schoten (R811) in Antwerp. Results for NO_2. More details are provided in the main text.

3.3.3 PM_{10} urban background station

Finally, urban background SR areas for PM_{10} are presented in Figure 20. Results of all Tier methods can roughly be classified in two groups. The first group contains approaches which yield a rather small SR area. These include the Spangl method (with a limited 1100x1100 m^2 area around the station, right figure on middle row of Figure 20), the emissions method (which only includes a broad band around the city centre, like the results for NO_2, left figure on middle row of Figure 20) and the Tier 1 methods with small radii (smaller than 5 km, top figure on Figure 20). For this entire group, the size of the SR area is clearly underestimated.

The second group combines the Tier 3 approaches (for all thresholds, bottom row on Figure 20), the Tier 2 land use method (middle figure on middle row of Figure 20) and the Tier 1 methods with radii larger than 5 km. The results for the Tier 1 methods are of course the same as the one for NO_2, but in contrast to the situation for NO_2, the mixing of the different locations is not that much of a problem, as the concentration gradients are much smaller for particulate matter. The Tier 2 land use approach adds the entire air quality zone to the SR area of this urban background station. Also, for the Tier 3 methods with a relatively large threshold (25%), almost the entire zone is included in the SR area. For smaller thresholds, the size of the area of course decreases, but it still encompasses a large part of the air quality zone, even for the smallest threshold considered (5%).

To summarize, it seems that both Tier 1 methods with large radii and Tier 2 methods based on land use provide realistic estimates for the urban background PM_{10} SR areas. The qualitative description is once more backed up by the quantitative analysis of the size of the SR areas (see Figure 19). The two groups of approaches are easily detected[10], just as the similarity between the size for the Tier 3, Tier 2 land use and Tier 1 with large radii approaches.

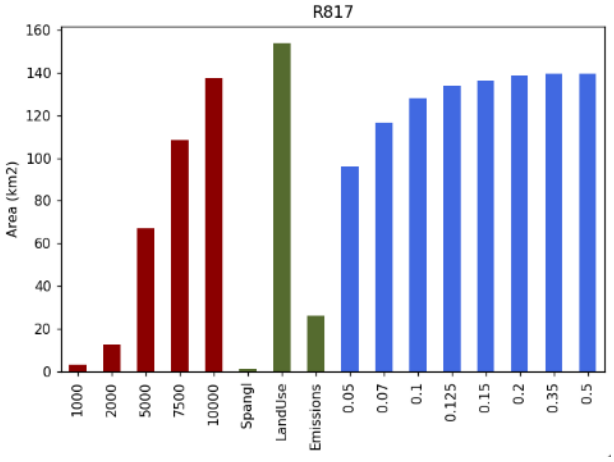

Figure 19: Size of the SR areas for PM_{10} for an urban background station according to the three Tiers considered in this analysis. In red the results for the fixed radius method of Tier 1 are considered, with the labels indicating the radius (in m). In green the three Tier 2 methods are highlighted, with the labels indicating the type of methodology. The results for Tier 3 are shown in blue, with the labels indicating the relative thresholds.

[10] Note that the threshold for the Tier 1 method can vary continuously and that at some point there will be a continuous transition from the first to the second group.

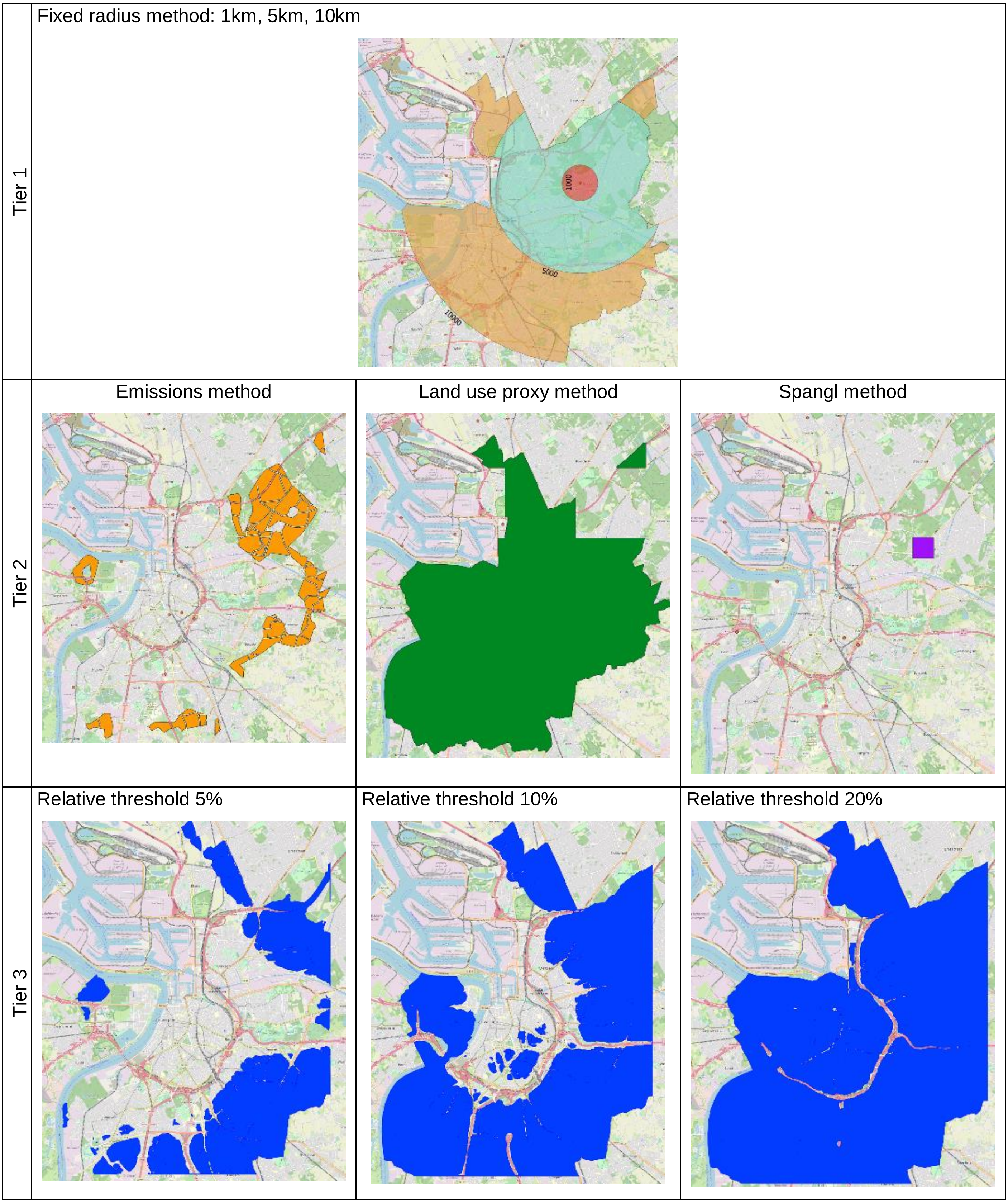

Figure 20. : Comparison of the SR areas for the urban background station Schoten (R811) in Antwerp. Results for PM_{10}. More details are provided in the main text.

4 Sensitivity tests for exceedance situation indicators

According to the IPR (Art. 12.3), for each air quality zone in which an exceedance is reported, the information made available shall also include information on the area of exceedance, the number of people exposed and the road length in exceedance (i.e. the exceedance situation indicators). Since these reporting obligations are driven by exceedances observed at station locations, there is a clear link between these exceedance situation indicators and the spatial representativeness of monitoring sampling points discussed in Chapter 3. The SR area of sampling locations in exceedance can help to estimate the exceedance situation indicators. However, the SR area may not cover the whole air quality zone and ultimately an assessment for the entire air quality zone must be reported. The application of a (Tier 1) expert opinion "extrapolation" would be an improvement over the declaration of the whole zone whilst use of a modelling technique (within Tier 3 or Tier 4) would be recommended as a more precise method to derive the exceedance situation indicators. Therefore, in this section we again first rely on the more comprehensive Tier 3 approach to evaluate the main features of these indicators. At the end of this section, we make a comparison with lower Tier approaches.

This section focuses on the sensitivity of these exceedance situation indicators with respect to the air quality data used, the calculation method and the applied threshold values. We focus on exceedances of the annual mean limit value of 40 µg/m^3 for NO_2 and PM_{10}. For NO_2, there are exceedances in every case study city, while for PM_{10} the limits are only exceeded in Krakow. We therefore mostly focus on the results for NO_2, but also provide some insights into PM_{10} exceedances for Krakow.

Before presenting the results in Section 4.2, some methodological considerations related to the three indicators and the three different Tiered approaches are presented in the next Section 4.1.

4.1 Methodology

4.1.1 Area in exceedance

It is straightforward to calculate the area in exceedance based on the Tier 3 model results by retaining all the grid cells where the modelled concentration is larger than the threshold under consideration. Tier 3 modelled based approaches are also available for the three cities. Therefore, these Tier 3 results will be used to analyse the sensitivity of a number of methodological options in the calculation of the area in exceedance situation indicators. We will illustrate the sensitivity of the results to the threshold value and to whether street canyons are included in the methodology. To test the latter option, we compare model results for Antwerp and Krakow including and excluding the street canyon increment, by comparing ATMO-Street results with and without the OSPM model (see section 2.3.1). These results are presented in Section 4.2. For Antwerp, results from the three Tiers are available thus a comparison between the different methodologies can be made as well. These results are presented in Section 4.3.

Deriving the area in exceedance from the fixed radius Tier 1 methods and the Tier 2 proxy-based methods is not as straightforward as for Tier 3. Both Tiers provide information on the station locations, and not on the entire domain of the air quality zone. We therefore use the size of the SR area of the sampling points in exceedance for the area in exceedance within the air quality zone. Obviously, this approach has its limitations since other hotspots can occur in the air quality zone apart from the measurement station locations. In Antwerp, for example only the street canyon station R802 is in exceedance of the limit values of the AAQD[11] and only this station is retained in this Tier 1 and Tier 2 analysis.

Besides the fixed radius results, Tier 1 also comprises expert opinion results based on the data reported by the Member States in the EU EIONET Central Data Repository (https://cdr.eionet.europa.eu/). The Member States are obliged to report the exceedance situation indicators when an air quality zone is in exceedance. Not all Member States do, however, report all the requested quantities, and some mention that it is impossible to estimate the requested values at the moment. For example, Belgium stipulates that

[11] Note that also measurements at the Ring and Belgiëlei telemetric station and several (yearlong) passive sampler locations indicate exceedances of the annual mean limit value for NO_2 . With the exception of the Ring Station (which lies at a non-publicly accessible location), all the other exceedances are reported as well. We omit these locations because there is no Tier 2 analysis available. Note that no PM_{10} exceedances have been measured and reported.

"Detailed information on exceedance area and/or road length are currently not possible to estimate", while Norway reports the total size of the air quality zone with the additional remark "Overestimation of the surface Area (conform to zone area). Unfortunately, there are no detailed information about the surface area or the road length available by the time of delivery." Poland provides an estimate for the area and the population exposed to exceedances but does not provide any information on the road length in exceedance. The data retrieved from the CDR is summarized in Table 5.

Table 5: Exceedance statistics reported by the Member States for the three case study cities as retrieved from the EU EIONET Central Data Repository (https://cdr.eionet.europa.eu/) on March 30th 2020. Reported data are for NO_2 and PM_{10} respectively.

City	Area (km²)	Population (% of AQ zone inhabitants)	Road length
Antwerp	Not provided	Not provided	Not provided
Krakow	23.9 / 156.6	28.7 / 64	Not provided
Oslo	Full size of AQ zone: 1575	Not provided	Not provided

4.1.2 Population exposed to exceedances

The estimated population exposure for Antwerp has been based on best available official local population data. We therefore use the official population dataset ("officiële statistiek van de bevolking") as retrieved from ProvincieInCijfers (https://provincies.incijfers.be/dashboard). The dataset provides the number of inhabitants per statistical sector (the smallest official administrative entity in Belgium), which is too coarse to be coupled with air quality data with a 10m resolution. We therefore spread the inhabitants of a statistical sector uniformly over all the buildings in the statistical sector. Building data has been retrieved from the official building dataset for Flanders (Informatie Vlaanderen). For Krakow and Oslo, the population data is based on the 250m resolution population layer of the Global Human Settlement Index (https://ghsl.jrc.ec.europa.eu/). For each of the 250m grid cells, the population in the grid cell has been spread over all the buildings in the grid cell. For Krakow, building data is taken from the 3D model of the Central Office of Geodesy and Cartography (available at www.geoportal.gov.pl), while for Oslo the buildings of OpenStreetMaps (www.openstreetmaps.org) have been used[12].

In the Tier 3 approach we assess how many inhabitants are living at a location at which the modelled concentration exceeds the EU limit value. There are however several options to couple the population data per building polygon to the gridded air quality maps. The simplest method uses the centroid of the building. This will be henceforth referred to as the "centroid" method. All inhabitants living in buildings for which the modelled concentration at the centroid of the building is higher than the threshold, are then retained. The highest roadside concentrations are of course omitted in this methodology. Therefore, we also assess the influence of retaining the entire building polygon and using the maximal overlapping modelled concentration. This will be henceforth referred to as the "building maximum" method. Inhabitants are thus retained if the polygon of the building in which they live overlaps at least with one grid cell for which the modelled concentrations exceeds the threshold. This method favours the high roadside concentrations[13], which contrasts sharply with the centroid method that favours somewhat lower concentrations further away from the roadside. Note that several other options exist to calculate the population exposure (e.g. using the original resolution of the population data), but we limit the analysis to these two options to illustrate the importance of the methods used to calculate the population exposed to exceedances. Results of this Tier 3 sensitivity analysis are presented in Section 4.3.

For the Tier 1 expert opinion approach it was necessary to rely on the results reported in the EU EIONET Central Data Repository for which data on population exposed to exceedances is only available for

[12] All these processing steps have their own caveats and assumptions, but analyzing the effect of different population datasets falls outside the scope of the current project.
[13] The method is applied by the Flemish Environmental Agency (VMM) to calculate the official population exposure in Flanders.

Krakow (see Table 5). The total population of Krakow as reported in the EU EIONET CDR was found to differ (by 2.5%) from the total population in the JRC population dataset applied in the Tier 3 approach and therefore only the percentage of the population living at locations exceeding the limit values of the AAQD (for NO_2 and PM_{10}), as indicated by the Polish authorities in the Central Data Repository, has been used in the comparison exercise.

For the fixed radius Tier 1 and the proxy-based Tier 2 methods the results from the SR areas for the station Borgerhout R802 in Antwerp were used. The population living inside this SR area is considered as the population exposed to exceedances of the relevant limit values of the AAQD. We only use the methodology based on the centroid of the buildings for the Tier 1 and Tier 2 analysis.

Additionally, for Tier 2 we present an estimate of the number of people exposed to exceedances based on two dedicated passive sampling campaigns (see paragraph 2.2.2). In each of these campaigns, the concentration at several sampler locations in the air quality zone has been measured. To estimate the population exposure in the air quality zone based on the results of the sampling campaign, it is assumed that the overall concentration distribution measured by the set of samplers is comparable to the concentration distribution of the entire city. In other words, assessing the fraction of people exposed to a certain concentration interval based on a discrete sample set will only match an assessment based on a concentration map with full spatial coverage when the concentration distribution modelled is similar to the one measured with samplers. This turns out to be an important precondition when using monitoring campaigns for population exposure assessment. For small sampling campaigns (e.g. the VMM campaigns) large deviations are to be expected, but for larger well-designed campaigns (e.g. the CurieuzeNeuzen campaign) a better match is expected. If both distributions match, the percentage of the population in exposure is equal to the percentage of the sampler locations in exceedance.

Comparison of the various Tiered approaches is presented in paragraph 4.3.

4.1.3 Road length in exceedance

The analysis of the road length in exceedance focuses on Krakow and Antwerp because these are the case studies in which street canyons are considered.

The IPR does not stipulate which type of roads must be considered when calculating the road length in exceedance (only roads with motorized traffic, pedestrian streets in a city centre, bicycle paths along a highway etc…?). We have explored the importance of the road network used in the computation of the road length in exceedance by considering several options for the road map. A first option concerns the road map used in the modelling application as a basis for the emission calculations. For reasons of computational resources, only the most important roads in the network have been retained in the air quality modelling chain. The total road length according to the model input thus underestimates the actual total road length in the domain, and therefore OpenStreetMaps (OSM) has also been investigated. The OSM roadmap contains all the accessible roads in the domain, including footpaths, untarred tracks and bicycle lanes. To eliminate all these minor roads, two selections have been composed. In the first selection, only living streets[14], residential roads, motorways, trunk roads, and primary, secondary and tertiary roads have been included (OSM selection 1). In a second selection, service roads have been included as well (OSM selection 2). The first selection is in line with the selection of roads used in the Krakow model set-up (although some of the roads have only a limited traffic load), while the second selection is based on expert opinion of the authors.

To couple the road map consisting of line shapes with the gridded air quality maps, the road map is discretized into 10m segments. For each of these segments, the concentration at the middle of the segment is used as the concentration of the segment. The total road length in exceedance is equal to the total length of the segments for which this concentration exceeds the threshold.

To compare the different Tiered results the fixed radius Tier 1 and the Tier 2 method applied for the station Borgerhout R802 are once more used. The road length inside this SR area is considered as the length in

[14] According to the definitions used by the OpenStreetMaps initiative, a living street is a street designed primarily with the interests of pedestrians and cyclists in mind. These roads have lower speed limits, and special traffic and parking rules compared to streets tagged as residential. For more information, see https://wiki.openstreetmap.org/wiki/Tag:highway%3Dliving_street.

exceedance of the EU limit values. We only use the model input road network for these Tiers as opposed to OSM.

4.2 Results based on a Tier 3 approach

As mentioned previously, we first discuss in this Section the results of the sensitivity tests according to the Tier 3 approach. This comprehensive methodology is best suited to evaluate methodological options and general sensitivities. Afterwards, the results are compared in Section 4.3 with the lower Tier approaches, using the insights of the Tier 3 analysis.

In general, the Tier 3 analysis indicates that the exceedance situation indicators are very sensitive to minor changes in the methodology and the input data. This behaviour is expected as indicators based on the exceedance of thresholds are especially sensitive to underlying assumptions. The effect is especially important for NO_2, and to a lesser extent for PM_{10}. For particulate matter, the limit values are only exceeded in Krakow and the background concentrations are already exceeding the limit values. We illustrate the sensitivity of the indicators by focussing on the effects of modifying some input datasets (type of road network etc.) and methodologies (combination of population data and concentrations, etc.) as described above in Section 4.1. Obviously, exceedance situation indicators will be sensitive to other inputs and assumptions as well (e.g. discretization of the road map, population data, etc.), but our objective here is to illustrate the main concept of the sensitivity of the results. All results are normalized with respect to one of the choices to facilitate the comparison between the case studies.

4.2.1 Street canyons

It is well known that street canyon effects can result in a significant increase in pollutant concentrations. This is especially true for NO_2. As residences in urban areas are often adjacent to locations influenced by the street canyon increments, these additional increments are of importance for the population exposure. However, based on the air quality maps collected in the context of the FAIRMODE Composite Mapping initiative, inclusion of street canyon increments in operational air quality models is not common practice these days in Europe. Figure 21 visualizes the impact of explicitly including street canyon effects in the calculation of the exceedance situation indicators. Only results for Antwerp and Krakow are presented.

In Krakow, there are almost no modelled exceedances if the street canyons are not included in the model. As a result, there is an immense difference between relative results from including or excluding street canyon contributions. In Antwerp there are NO_2 exceedances outside the street canyons (these are in open locations close to the Ring Road), but there is still a difference of 20% in the area in exceedance between when canyons are included or excluded. For the population and road length in exceedance, we used the centroid method mentioned above and the road map used in setting up the model. When considering the number of people exposed to exceedances, the difference is even larger, because most of the exceedances of the non-street canyon results occur at locations with a lower population density (in Antwerp this area is a small band around the Ring Road). For the road length indicator, the difference in road length indicated to be in exceedance between including or excluding street canyons is estimated to be 25%.

To summarize, adding the street canyon locations results in a significant difference for all the exceedance statistics. However, it should be noted that operational street canyon modelling comes at an additional computational cost and requires extra 3D building information as an input parameter to the model.

In the remainder of the analysis, we will focus on the results including the street canyons contributions for Antwerp and Krakow. For Oslo, street canyon contributions are not considered in the model, and we should thus realize that the modelled exceedances for Oslo will underestimate the actual values.

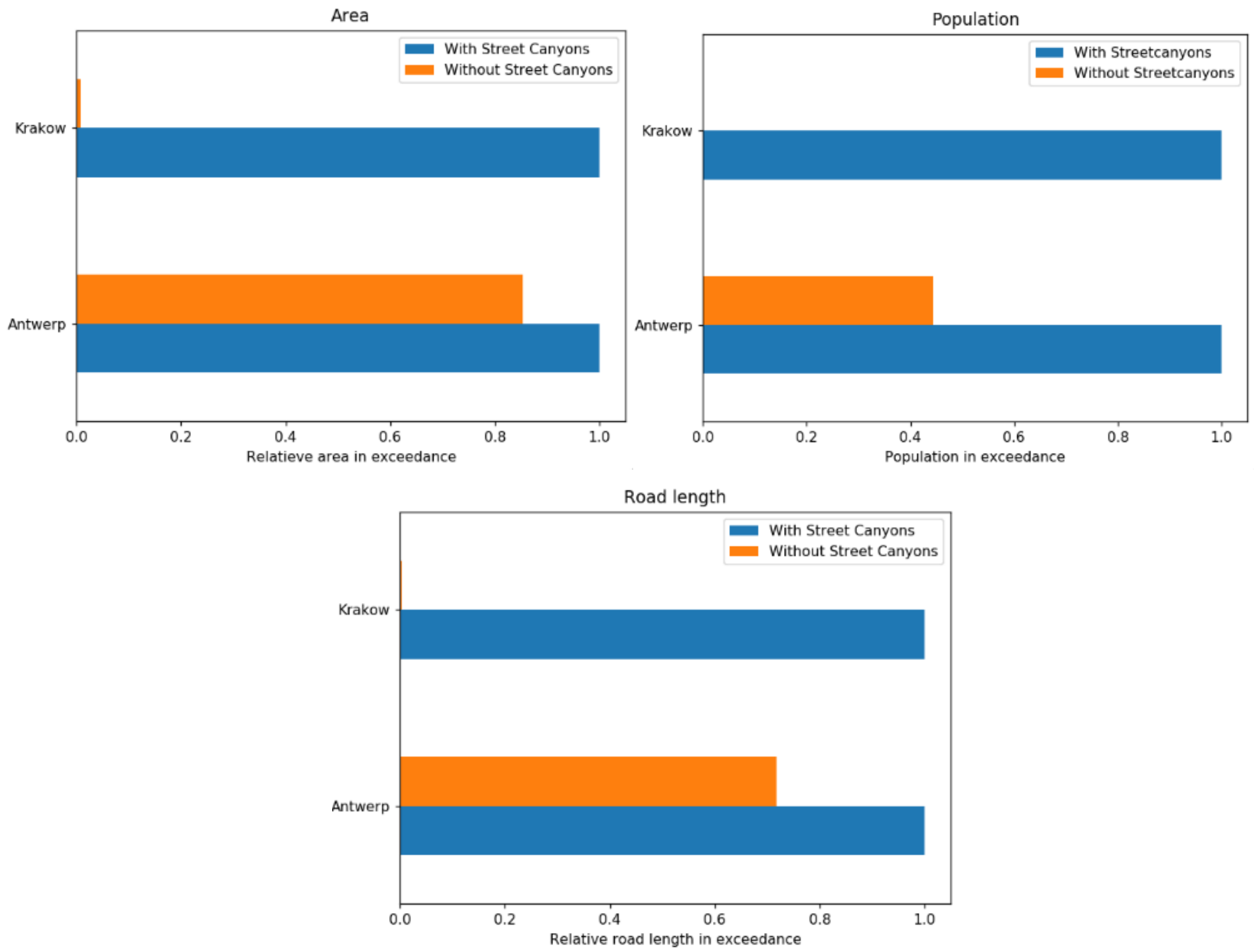

Figure 21: Sensitivity of the exposure indicators for NO_2 to the inclusion of street canyons. Results are normalized with respect to the value for the standard ATMO-Street map including street canyons. Top left: area in exceedance. Top right: population living at locations exceeding the limit value. Bottom: road length in exceedance.

4.2.2 Population and road length input data sets

Figure 22 shows the sensitivity of the population and road length exceedance situation indicators with respect to changes in the input data or assessment methodology.

The graph presenting the number of people exposed to exceedances visualizes the effect of the two methodologies assigning population to gridded model output as outlined in Section 4.1.2. The number of people exposed to exceedances obtained using the centroid method is of course lower than the value obtained using the building maximum method. There is a significant difference between both approaches in Antwerp and Krakow. In both cases the population indicator is almost three times larger when the building maximum method is applied. This effect is partially caused by the street canyon contribution, as this increment doesn't reach the centroid of the buildings (as it is limited to the actual street canyon and thus stops at the façade of the building).

The effect is however not only important for air quality maps in which street canyon contributions are explicitly modelled, as the results for Oslo indicate. Although the difference between both approaches is smaller in the Oslo analysis, the exposure is still more than 50% larger when the building maximum method is applied. For simplicity (and due to computational resource and time constraints), we use the centroid method for the remainder of the sensitivity tests.

Finally note that these effects are of less importance when considering population exposed to exceedances of the annual limit value of the AAQD for PM_{10}. PM_{10} concentration gradients are in general less pronounced, even if canyon effects are considered. In this sensitivity analysis the only exceedance of the annual mean limit value is observed in Krakow. Since the background concentration is already exceeding the limit values in most of the air quality zone, the impact of the different methodologies is limited. It is difficult to draw general conclusions based on the actual threshold and the cities under consideration, but extra tests with modified thresholds indicate that the population exposure to PM_{10} is only very sensitive to the calculation method if the background concentration is close to the threshold value.

Figure 22 illustrates how the road length in exceedance depends on the actual input road link data. In Antwerp, the road map used in the model only includes the most important roads in the network. Consequently, the road length in exceedance is larger for the (sub)sets of the OSM roads. When considering the full OSM road input, the road length in exceedance is two times larger than the results using the model road map. In Krakow, the road map used in the model coincides with the first sub-selection of the OSM data, and both results thus lead to the same road length in exceedance. The total length in exceedance increases if a larger OSM dataset is used.

The impact of the type of road link data is even important for PM_{10}. As (almost) the entire air quality zone exceeds the limit value in Krakow, the road exceedance parameter is (almost) equal to the total road length in the domain, and thus solely influenced by the input road map.

This seems to be an obvious conclusion, but further guidance is required if such a parameter is to be reported, in order to ensure comparability between different zones.

For simplicity (and due to computational resource and time constraints), we use the model road map in the remainder of the sensitivity tests.

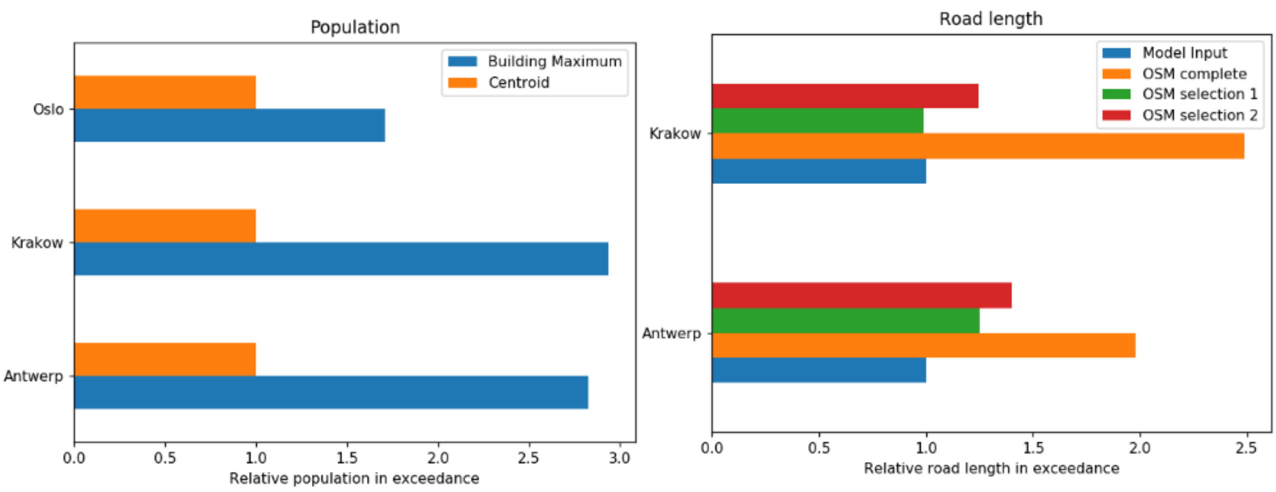

Figure 22: Sensitivity of the exceedance situation indicators to the calculation method for the population (left) and road length (right) in exceedance for NO_2. Results are always normalized with respect to the value for the standard methodology. More details are provided in the main text.

4.2.3 Limit values

Finally, we assess the sensitivity of the exceedance situation indicators with respect to changes in the threshold or limit value. We first focus on NO_2. By changing the threshold of 40 µg/m^3 with a relative offset the robustness of the indicator estimates is investigated. These sensitivity tests provide insight into both the impact of model uncertainty and the general sensitivity of threshold-based indicators.

Figure 23 provides the results for the three NO_2 exceedance situation indicators in the three cities. All results have been normalized with respect to the results for the standard threshold (40 µg/m^3). For NO_2, the three exceedance situation indicators are sensitive to changes in the threshold value. The effect for the area in exceedance is similar in the three case study cities. Decreasing the model threshold by 20% (8µg/m^3), multiplies the size of the area in exceedance by a factor of three to four. Similarly, increasing the threshold by 20%, the area in exceedance falls back to 30% of the original area in exceedance. The sensitivity of the number of people exposed to exceedances to the threshold value is even larger. In Antwerp, decreasing the threshold by 20%, yields an increase in the number of people exposed to exceedances by a factor of six. Because densely populated areas in a broad band around the Ring Road are added to the locations in exceedance, the number of people exposed to exceedances increases drastically. In the other case study cities, the factor is somewhat smaller, but still significantly larger than the factor for the area in exceedance. This observation clearly points towards a large sensitivity of this exceedance situation indicator to its underlying data sets and assumptions. The road length in exceedance, on the other hand, is somewhat less sensitive to changes in the threshold. In summary, all three NO_2 exceedance situation indicators are sensitive to changes around the limit value.

The effects differ for PM_{10} and they depend to a large extent on the overall background levels compared to the limit value. In Krakow, the exceedance situation indicators are very sensitive to small changes in the thresholds or modelled values, and, for most of the offset range, the area in exceedance is either non-existent or equal to the entire air quality zone. Figure 24 shows the results for the area in exceedance, but the other indicators follow a similar pattern. The underlying reason for this behavior is a combination of the small gradient in PM-concentrations in the domain, and the value of the background concentration, which is very close to the threshold value. Consequently, an increase of 10% of the threshold causes all grid cells to be below the threshold, and all exceedances to disappear. On the other hand, if the threshold is decreased by only 5%, all grid cells exceed the threshold and the entire air quality zone thus exceeds the limit values. For the other two cities, within the range of thresholds offsets that has been considered, the sensitivity is very limited, simply because the urban background concentration is much lower than the threshold value. We can therefore conclude that the area in exceedance (and all related parameters) for PM_{10} is very sensitive to the actual value of the threshold, but only in cities for which the background concentration is close to the threshold value.

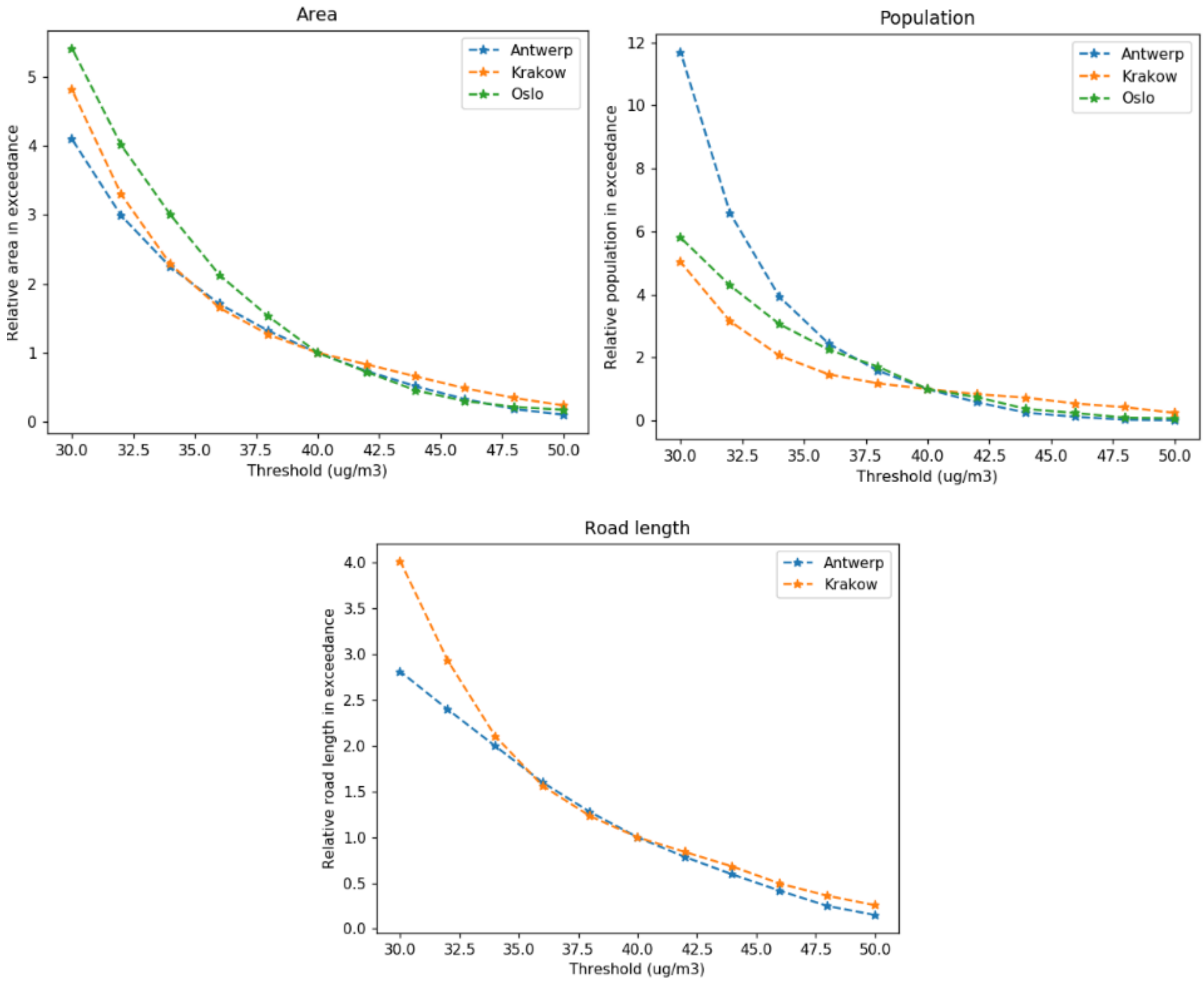

Figure 23: Sensitivity of the exposure indicators for NO_2 to the threshold of the AAQD. Results are normalized with respect to the value for the standard threshold (40 µg/m³). Top left: area in exceedance. Top right: population living at locations exceeding the limit value. Bottom: Road length in exceedance.

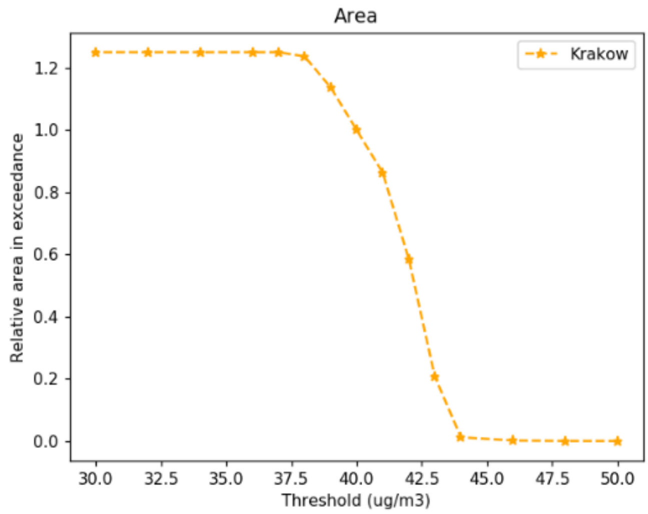

Figure 24: Sensitivity of the area in exceedance for PM_{10} to the threshold of the AAQD. Results are normalized with respect to the value for the standard threshold (40 µg/m³).

4.3 Comparison between different Tiers

The previous Section illustrated the sensitivity of the exceedance situation indicators to small changes in underlying input data and assessment methodologies. These sensitivities were already large based on well-established Tier 3 methods. We thus expect even larger differences between the different Tiers, and even between the methods within the same Tiers.

For Tier 3 in this comparison, we make use of the results including street canyons (if available), population exposure based on the centroid method, and road lengths based on the road map used in the model.

First, we focus on the Tier 1 expert opinion as reported in the EU EIONET Central Data Repository. As mentioned in Table 5, the reported data is far from complete. Two of the Member States linked to the cities under consideration do not report the area in exceedance (or mention the impossibility to correctly report these statistics), while data is even sparser for the number of people exposed to exceedances. This issue already indicates the difficulty faced in this Tier 1 approach, since many local experts opt not to report explicit values based on the measurement data. When exceedance situation indicators are reported, their values differ strongly from the model results, as indicated in Figure 25. In Oslo, the whole zone is reported as exceeding the limit values, while only a few grid cells exceed the thresholds according to the model results. The Norwegian government also explicitly mentions that the area in exceedance is overestimated. In Krakow, the reported area in exceedance is larger than the modelled area for NO_2, while it is smaller for PM_{10}. These results are also reflected in the number of people exposed to exceedances indicator.

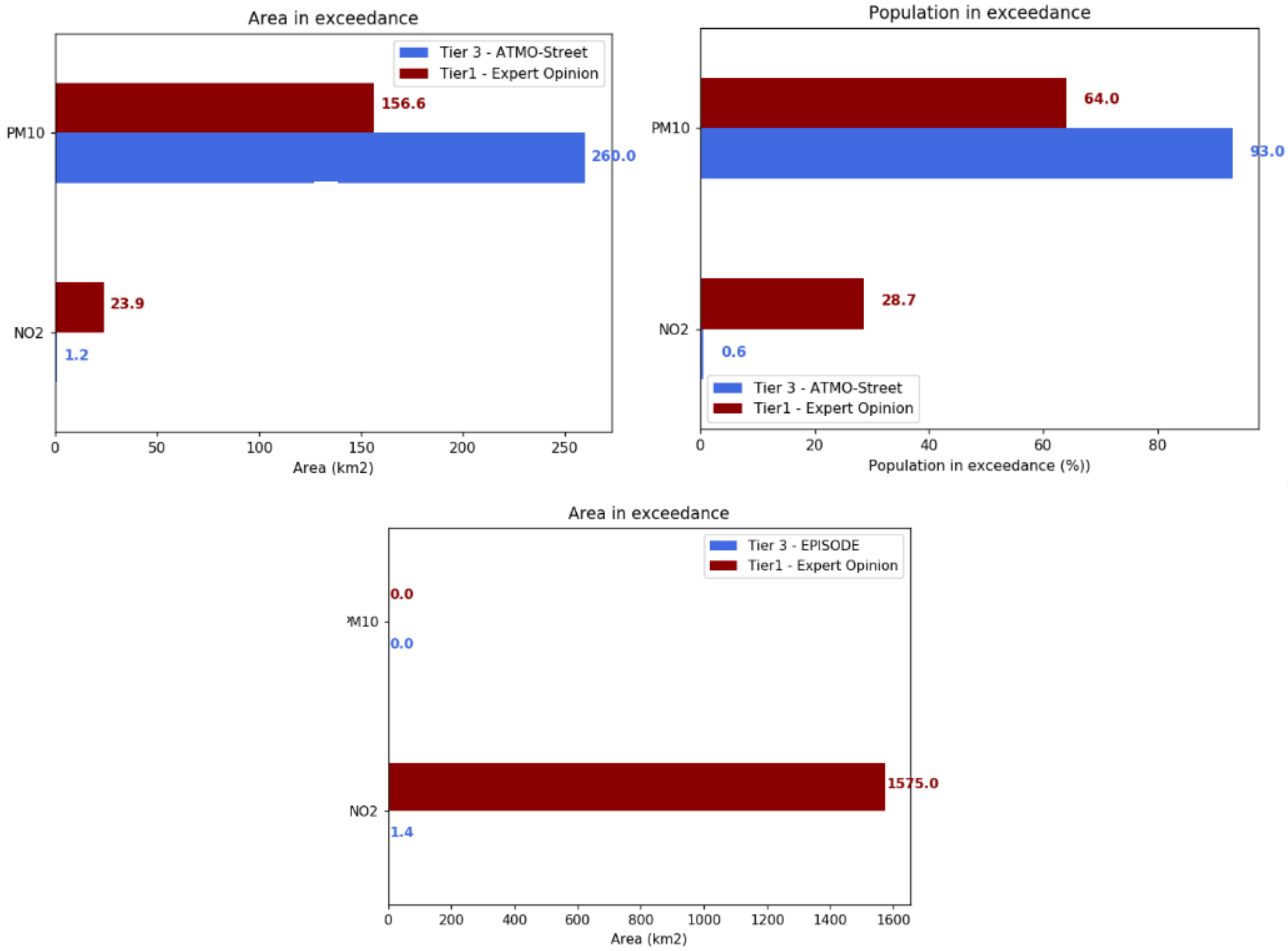

Figure 25: Comparison of the exposure indicators reported by the countries (expert opinion) with those obtained using the ATMO-Street/EPISODE simulations. The top row provides results for Krakow, the bottom row for Oslo.

Figure 26 compares the NO_2 exceedance situation indicators for the three Tiers. As mentioned before results for all three Tiers are only available in Antwerp. For Tier 1 and Tier 2 these results are based on the SR area for the traffic, street canyon station R802. For the number of people exposed to exceedances, we also provide the results of the two Tier 2 methods based on the passive sampler campaigns of VMM and the CurieuzeNeuzen campaign and results for the two options to calculate the population exposure for Tier 3 (building and centroid method, introduced in section 4.1.2). Each of the methods is discussed in more detail below.

The Tier 1 result collected via the CDR only reports exceedances for a very limited area. This method is based on the minimal SR area for the traffic station cited in the AAQD, and thus only the 100m road segment in the immediate neighborhood of the air quality station R802 is considered. The method is labeled as ‘100m street segment’ in Figure 26. The Tier 1 methods based on the fixed radius span a larger range of exceedance areas. For the smallest radii (< 500m), the area in exceedance is much smaller than the area obtained in the Tier 3 approach. When larger radii are considered, the results are more in line with the model results. Note however that the spatial layout of the locations in exceedance of the limit value differs strongly from the one obtained in the Tier 3 approach. For example, if we focus on the approach with a radius of one kilometer, the size of the area in exceedance is in the same order of magnitude as the one for the Tier 3 model results, but the map of the locations in exceedance is entirely different (see Figure 27). Although the Tier 1 results can provide some insight into the size of the area in exceedance in Antwerp, there is no basis to assume that similar radii provide reasonable results for the size of the area in exceedance for other cities.

As for the SR area analysis, the results of the Tier 2 methods based on proxy data (emissions, land use and Spangl method) span a very wide range of results. The land use proxy method classifies the entire city centre as part of the SR area, and thus in exceedance of the limit values. The area in exceedance is thus very large in comparison with the other methods. The two other Tier 2 methods (the emissions method and the Spangl method) limit the SR area to the immediate surroundings of the traffic station, and thus yield a limited area in exceedance.

When focusing on the number of people exposed to exceedances, the Tier 2 methods based on the CurieuzeNeuzen sampling results provides reasonable results. The Tier 2 method based on the small VMM sampling campaign however significantly overestimates the number of people exposed to exceedances when compared with the Tier 3 results. The goal of this campaign is to sample a few very specific locations which are both difficult to model and have high concentrations, which of course leads to a much higher estimate for the population in exceedance than the Curieuzeneuzen campaign and the Tier 3 methods. Tier 2 methods based on sampling campaigns thus only provide added value if the number of samplers is large enough and if the distribution of the sampling locations mimics the concentration field in the domain under consideration. Note that in the CurieuzeNeuzen campaign model results have been used to select the optimal measurement locations.

In summary, we conclude that the Tier 1 methods and Tier 2 methods based on proxy data provide very varied estimates for the exceedance situation indicators depending on the actual choice of the method and the thresholds. Only specific choices for the thresholds lead to results that are in the same order of magnitude as the Tier 3 model results, whereas the other choices lead to a large under- or overestimation of this benchmark. Although the uncertainty on the Tier 3 model results is quite large (as indicated in the previous section), some Tier 1 and Tier 2 methods provide results that have a much larger spread than the largest uncertainty interval noticed in the sensitivity studies of the Tier 3 model results. Both Tier 1 and Tier 2 methods based on land use proxy methods thus seem of limited use when reporting the exceedance statistics required by the AAQD and the IPR. On the other hand, Tier 2 methods based on sampler campaigns could provide an alternative to estimate the population exposure if the number of sampling locations is large enough and if the sampling locations are carefully selected to correctly measure the concentration field in the air quality zone under consideration. It is unclear whether such methods can be used to estimate the area and road length in exceedance, as methods to calculate these indicators based on large-scale sampling campaigns are not currently available. Note that such campaigns require a substantial investment.

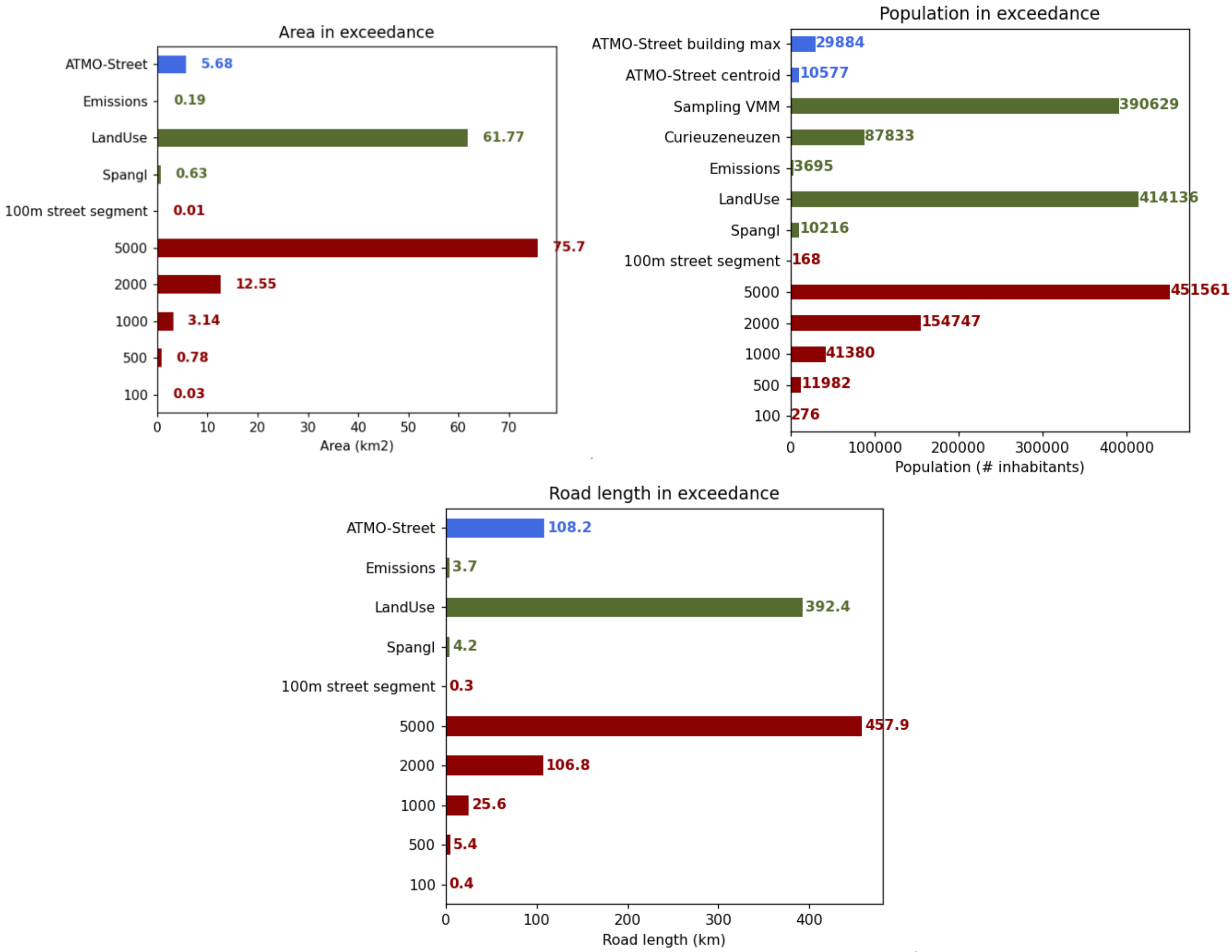

Figure 26: Comparison of the NO_2 exceedance situation indicators obtained using the different Tiers for Antwerp. Top left: area in exceedance. Top right: population living at locations exceeding the limit value. Bottom: Road length in exceedance. The colours indicate the Tiers of the results: Tier 1 (red), Tier 2 (green), or Tier 3 (blue). More details and further explanation of the labels: see main text.

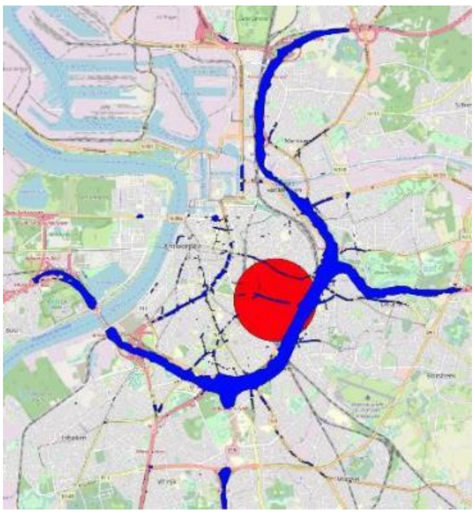

Figure 27: Comparison of the area in exceedance of the threshold values for NO_2. The map shows results for a fixed radius method with radius equal to 1 kilometre (red) and the results of the ATMO-street model chain (blue).

5 Feasibility tests for network design & model validation

The current air quality monitoring network in Europe has been designed inter alia following the Tier 1 methodologies in accordance with the requirements under Annex III of the AAQD and the IPR (Article 12 and Annex II). However, significant gaps have been identified in the evaluation and reporting of the area of representativeness of the monitoring sampling points (Interim Task 3 report- Tarrasón et al, 2020). A possible reason for this, is the lack of guidance and tested methodologies to independently assess the quality of the current monitoring network and the spatial representativeness of its sampling points. The tiered approach tested in this report is intended to provide some of this necessary guidance. This chapter presents feasibility tests in Antwerp and Oslo on the potential of using higher level Tier approaches as screening methods to determine the quality of the monitoring network. In particular, the potential of the hierarchical clustering methodology to evaluate and support monitoring network optimization either as a Tier 2 or as a Tier 3 is presented here, showing how the approach can help answer the following three questions:

- Are there any redundancies in the monitoring network in the city area? (where more than one monitoring station is located within the same SR area)
- Are there any significant outliers in the monitoring network?
- Are all hot-spot areas covered by the current monitoring network?

The feasibility tests are carried out both for a Tier 2 approach, using only monitoring data, and for a Tier 3 approach, when both monitoring and modelling data are available in the same area. If only monitoring data is available, the application of the clustering methodology tested represents a Tier 2 approach, similar to the objective classification of Joly et Peuch (2012) but with the additional value that it allows the identification of possible redundancies and outliers in network. If both monitoring and modelling data are available, the application of the clustering methodology allows a Tier 3 approach for both model validation and monitoring network design purposes.

In order to carry out these feasibility tests, we need access to both air quality monitoring and modelling data over a long-term timeseries, preferably one year, with hourly resolution. The combined analysis of modelled results and observations can provide useful insights on network design and for model validation purposes, but it might be demanding in terms of the amount of data to be compiled and analysed. In this chapter, we show how a simple evaluation using a hierarchical clustering methodology applied to modelled and observed air quality concentrations can help identify both redundancies and gaps in the monitoring network design.

In the first section, a short summary of the hierarchical clustering methodology is introduced. In the second section, the feasibility tests are presented first for monitoring data to derive conclusions on the network design (Tier 2), then, in the third section with modelled data, the clustering methodology is used in a Tier 3 evaluation for model validation purposes. In the final section, feasibility tests are shown using the methodology for a Tier 3 cluster application to support the evaluation of spatial representativeness (SR) of sampling points, and through the SR clustering contribute to monitoring network design.

5.1 Hierarchical Clustering Methodology

The hierarchical clustering methodology proposed by (Soares et al. 2018) determines the level of similarity between the monitoring data from different sampling points. Hierarchical clustering assumes that the data contains a level of (dis)similarity and groups the station records based on characteristics of the actual data. As explained in chapter 2, with the presentation of the different methodologies, this approach goes beyond other objective classification methods as it can be used beyond sampling point characterization purposes. The method can be applied to monitoring data, as well as modelling results and it can be used for sampling point characterization purposes, for model validation purposes, for characterization of spatial representativeness purposes and for monitoring design purposes.

The analysis of data with hierarchical clustering requires a temporal coverage of 75% of the period analysed, here one year, and that there are no gaps over 168 hours, as recommended by Solazzo and Galmarini (2015). To establish the level of similarity (or dissimilarity) of the different air quality measurements across the monitoring network, three dissimilarity metrics are used: (1) 1-R, where R is the Pearson linear correlation coefficient (Solazzo and Galmarini, 2015), (2) the Euclidean distance, EuD (Soares et al, 2018), and (3) multiplication of metric (1) and (2). The Euclidian distance (EuD) is calculated as indicated in Equation (1).

$$EuD_{x,y} = \sqrt{\sum_{j=1}^{j}(x_j - y_j)^2} \quad (1)$$

where *x* and *y* are two different time series, *j* is jth record of the time series.

The metric based on correlation, (1-R), assesses dissimilarities associated with the changes in the temporal variations in concentration, while the metric based on the Euclidian distance (EuD) assesses dissimilarities on the basis of magnitude of the concentration over the time period of the analysis. The multiplication of these two metrics (1-R) x EuD allows assessing correlations in terms of both time variation and pollution levels. For simplification purposes, in this chapter, we will show only results from the combined metric because it provides information both on the temporal variability and the magnitude of the air concentrations.

The outcome of the methodology indicates the (dis)similarity of sampling points, at a certain level of the chosen metric, across the whole data set. The higher the dissimilarity level between a station or cluster of sampling points, the more dissimilar those pairs are. The dissimilarity level value depends on the metric chosen. The values calculated based on the correlation metric can vary from 0 to 2, with values over 1 meaning that the data samples anti-correlate. The values for the other two metrics are varying from zero to a value that strongly depends on the average concentration levels and the number of records analysed. The outcome of the analysis also strongly depends on the quality of the reported data, so that random errors in the observations can potentially change the results (Soares et al., 2018).

There are two outcomes of the analysis that are used for visualization of the results: dendrograms and ranking sampling points from most to least dissimilar. Dendrograms are 2D representations of the clustering process showing the pattern of linkages between the data series while clustering occurs, as well as their level of dissimilarity. This graphical representation is the easiest way to visualize the results and to identify the dissimilarity level at which sampling points cluster together, distinguishing groups of sampling points. The clusters can then be displayed geographically, colour-coded according to the cluster the sampling points have been allocated to, and shape-coded according to the sampling point classification. In addition, a ranking of the sampling points is done according to the level they have cluster for the first time, indicating their level of dissimilarity in relation to the remaining sampling points available for the analysis. Both visualization systems are used for the results of the feasibility tests reported here.

5.2 Tier 2 objective classification of monitoring networks

The potential of the hierarchical clustering methodology to be used as a screening evaluation of the siting and classification of a given monitoring network has been tested here in two cities: Antwerp and Oslo. For this feasibility study, monitoring data was available from sampling points of the air quality networks in Antwerp and Oslo, continuously measuring NO_2 and PM_{10}. Table 6 shows the number of sampling points which data fulfils the temporal-coverage requested by the methodology, for each monitoring network. For Antwerp, the data is available for 2018 and in Oslo for 2015.

Table 6: Number of sampling points and classification type available for the analysis for the monitoring network in Antwerp for 2018 and in Oslo for 2015.

Pollutant	City	No. sampling points	Classification
NO_2	Antwerp	22	industrial (14), traffic (3), background (4)
	Oslo	8	traffic
PM_{10}	Antwerp	17	industrial (10), traffic (3), background (4)
	Oslo	8	traffic (7), background (1)

This hierarchical clustering methodology has been applied to these two areas. It serves to analyse the level of similarity or dissimilarity of air concentration data from all sampling points and allows the identification of sampling points with similar behaviour and those with specific behaviour that differs from the rest of the data sampling points, the so-called “outliers”. The results shown here comprise a ranking of sampling points and a series of dendrograms representing the clustering process based on the 1-RxEuD metric. The ranking of sampling points conveys the same message as the dendrograms, but in a simpler way. In both cases, sampling points are singled out in terms of their high or low level of dissimilarity.

- High level of dissimilarity usually indicates the sampling points are related to unique sources types (industry, airports, tec..) or alternatively it can indicate potential QA/QC issues with the measurements.
- Low level of dissimilarity indicates potential redundancies in the network, meaning that the sampling points are potentially measuring similar sources, thus one of them could be relocated to another area or hot spot if legal requirements allow for this.

The metric 1-RxEuD was chosen for the dendrograms because it reflects both temporal profile and magnitudes dissimilarities between the data records. The dendrograms will also exhibit a dashed line, that indicates the level of dissimilarity chosen to single out unique sampling points (cluster 2-red, 3-orange, 4-blue) and sampling points that present a higher level of similarity (cluster 1, black). Together with the dendrograms, in all the results presented below, there is a map displaying the sampling point location represented by a circle that is colour-coded similar to what is shown in the dendrograms to visualize which sampling points cluster at the level of dissimilarity chosen.

5.2.1 Antwerp monitoring network

The dendrogram describing the clustering process for sampling points measuring NO_2 in Antwerp based on the 1-RxEuD metric is shown in Figure 28a. The level of dissimilarity chosen is 14,800 µg/m^3 for the 1-RxEuD metric. The unit of this metric is the same as concentrations because correlations (1-R) have no units while the (EuD) Euclidian distance is measured in µg/m^3. The Euclidian distance is a cumulative value of the difference between the concentrations at two different sampling points. Since we have used hourly concentrations for a whole year (a total of 8760 values), the level of dissimilarity chosen here allows for an average difference of 1,68 µg/m^3 between the NO_2 concentrations of any given pair of sampling points. Any sampling points showing smaller differences will be clustered together, any pair with larger differences will be shown out of the cluster. The choice of the level of dissimilarity is important to allow the interpretation of the results and should be adjusted for the application in place, considering how different are air quality levels both in magnitude and in temporal and spatial variability. The spatial location of the sampling points is provided in Figure 28b, showing how the different clusters are distributed in the city area. The ranking of sampling points based on the same metric is provided in Table 7, with short names to allow the identification of the sampling points in the map in Figure 28b.

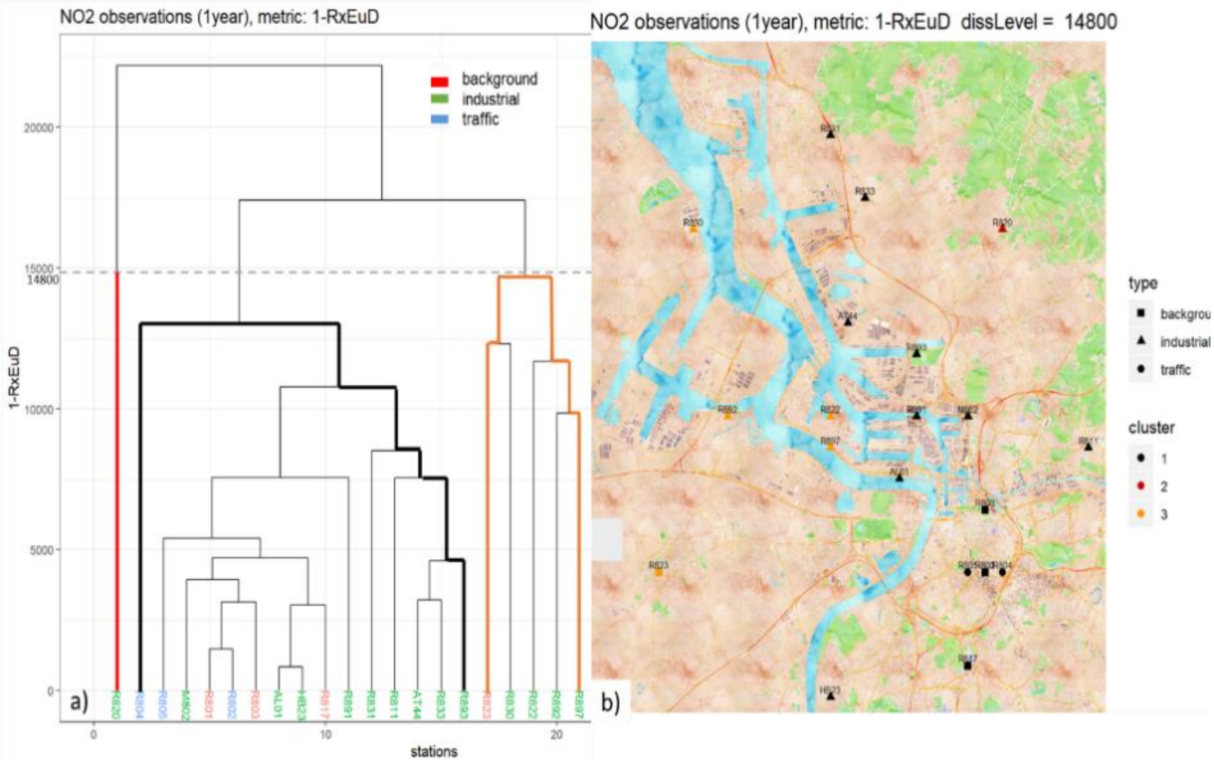

Figure 28: a) dendrogram for clustering of sampling points measuring NO_2 in Antwerp, in 2017, considering 1-RxEuD metric; the dashed line represents the level chosen to indicate unique and sampling points that present a higher level of similarity; and the coloured lines represent the sampling points clustered for that level of dissimilarity; b) spatial distribution of the clusters at the same level of dissimilarity, same colour code.

The dissimilarity analysis from the clustering results identify the sampling points in Antwerpen Ring (R804) and Kapellen (R820) as very different from the rest of the sampling points in the Antwerp monitoring network. Antwerpen Ring, R804 is located very close to busiest highway. Kapellen, R820, on the other hand, is classified as an industrial station, although it is located at a street canyon location at approximately 10km from the main industrial sources in the port of Antwerp and shows a very different behavior that the other industrial sampling points. Because of the very different nature of the sources affecting these sampling points, it is of no surprise that they are very dissimilar to the rest of the network. The analysis in Figure 28 shows how the rest of the sampling points in the network cluster as three different groups, to a large extent following the given classification of the sampling points according to Tier 1, but modifying it with additional air quality regime information. The Tier 1 classification is given in the figure as “background”, “traffic” or “industrial sites”, while the Tier 2 classification here shows how these different sampling points cluster according to different air quality variability regimes. It is interesting to note how the predominantly industrial cluster (the “yellow cluster – number 3”) is mostly situated at the left side of the river, while the urban cluster (the “black cluster- number 1”) is located in the right side, in agreement with the source and air dispersion conditions in the area.

Table 7: Ranking of sampling points measuring NO_2 in Antwerp, in 2018, from largest to smallest dissimilarity.

NO_2 sampling point code	Sampling point name	1-RxEuD metric (Unit: µg/m^3)
R820	Kapellen	22195
R804	Antwerpen Ring	13046
R823	Beveren	12309
R830	Doel	12309
R822	Antwerpen Polderdijk	11691
R892	Kallo Sluis	9852
R897	Antwerpen Scheldelaan	9852
R831	Berendrecht	8516
R891	Antwerpen Scheurweg	7570
R811	Schoten	7547
R805	Antwerpen Belgiëlei	5403
R893	Antwerpen Ekersedijk	4599
M802	Antwerpen Luchtbal	3952
AT44	Antwerpen Ardamstraat	3211
R833	Stabroek	3211
R803	Antwerpen Park Spoor Noord	3147
R817	Antwerpen Groenenborgerlaan	3033
R801	Antwerpen Borgerhout straatkant	1477
R802	Antwerpen Borgerhout achtergrond	1477
AL01	Antwerpen Linkeroever	847
HB23	Hoboken Karmeliet	847

The dendrogram describing the clustering of sampling points measuring PM_{10} in Antwerp based on the 1-RxEuD metric is shown in Table 8.

Figure 29a; the level of dissimilarity chosen is 4200 µg/m^3 to be able to distinguish the variability of the network measured concentration in Antwerp. The ranking of sampling points is presented in. The clustering analysis of sampling points shows that these are clearly more similar than those for NO_2, relating to the fact that PM_{10} concentrations are much more evenly distributed in Antwerp due to their origin from a large variety of sources and showing smaller gradients in PM_{10} concentrations in the domain.

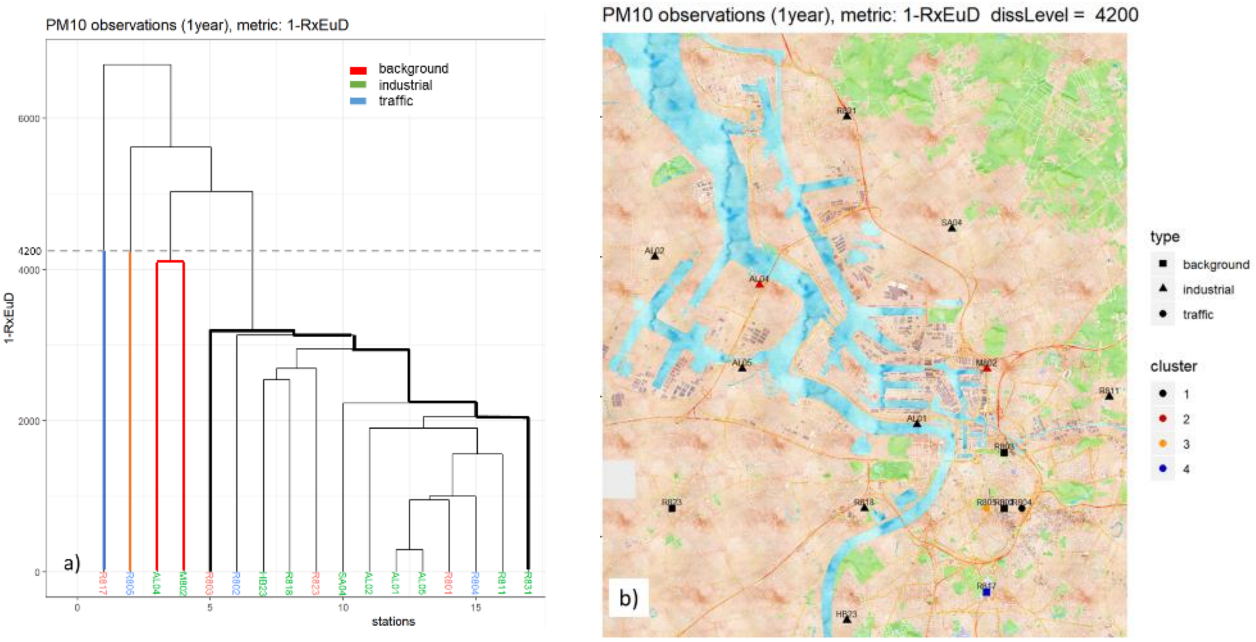

Figure 29: a) dendrogram for clustering of station measuring PM_{10} in Antwerp, in 2018, considering 1-RxEuD metric; the dashed line represents the level chosen to indicate unique and sampling points that present a higher level of similarity; and the coloured lines represent the sampling points clustered for that level on dissimilarity; b) spatial distribution of the clusters at the same level of dissimilarity, same colour code.

Table 8: Ranking of sampling points measuring PM10 in Antwerp, in 2018 from largest to smallest dissimilarity.

PM_{10} sampling point code	Sampling point name	1-RxEuD metric (Unit: µg/m^3)
R817	Antwerpen Groenenborgerlaan	6709
R805	Antwerpen Belgiëlei	5613
AL04	Kallo Liefkenshoektunnelstraat	4104
M802	Antwerpen Luchtbal	4104
R803	Antwerpen Park Spoor Noord	3202
R802	Antwerpen Borgerhout straatkant	3133
R823	Beveren	2689
HB23	Hoboken	2539
R818	Antwerpen Burchtse Weel	2539
SA04	Hoevenen	2234
R831	Berendrecht	2055
AL02	Doel	1900
R811	Schoten	1557
R804	Antwerpen Ring	1000
R801	Antwerpen Borgerhout achtergrond	952
AL01	Antwerpen Linkeroever	294
AL05	Kallo Sluis	294

5.2.2 Oslo monitoring network

The dendrogram describing the clustering process for sampling points measuring NO_2 in Oslo based on the 1-RxEuD metric is shown in Figure 30a; the level of dissimilarity chosen is 2411 µg/m^3.much lower than in the city of Antwerp because the air pollution concentrations of NO_2 are lower in Oslo but also because we have less variability and fewer street-canyon sampling points in this city. As indicated before, the choice of the threshold is essential to the clustering results and needs to be adapted to the area under analysis. The spatial location of the sampling points is provided in Figure 28b, showing how the different clusters are distributed in the city area. The ranking of sampling points is provided in Table 9 with short names to allow the identification of the sampling points in the Oslo city maps.

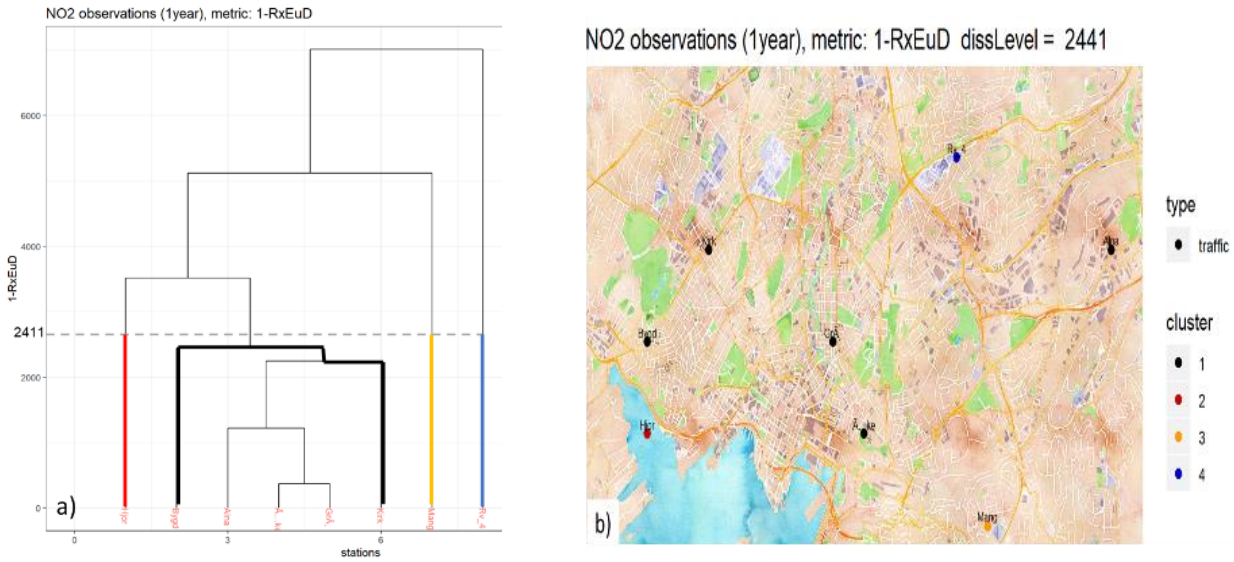

Figure 30: a) dendrogram for clustering of sampling points measuring NO_2 in Oslo considering 1-RxEuD metric; the dashed line represents the level chosen to indicate unique and sampling points that present a higher level of similarity; and the coloured lines represent the sampling points clustered for that level on dissimilarity; b) spatial distribution of the clusters at the same level of dissimilarity, same colour code.

Table 9: Ranking of sampling points measuring NO_2 in Oslo, from largest to smallest dissimilarity, for 2015

NO_2 sampling point code	Sampling point name	1-RxEuD metric (Unit: µg/m^3)
Rv4	Rv4 Aker sykehus	7014
Mang	Manglerud	5122
Hjor	Hjortnes	3516
Bygd	Bygdøy_Alle	2440
Krk	Kirkeveien	2244
Alna	Alnabru	1219
GrÅ	Grønland	369
Åke	Åkebergveien	369

Two sampling points are clearly ranked as the most similar sampling points in the Oslo monitoring network:Grønland and Åkebergveien, for all metrics. These sampling points seem to have a very similar temporal profile and magnitude variation throughout 2015, and therefore could be considered as potentially redundant. Considering the three most dissimilar sampling points, Rv4 Aker sykehus, Manglerud and Hjortnes are singled out as the most dissimilar sampling points in the network. Manglerud is unique in terms of temporal variability. The other two sampling points are located at highly trafficked locations, with Rv4 Aker Sykehus located by the motorway connecting Oslo to other municipalities surrounding the city area (thus the main traffic for work-home commuting) and Hjortnes located close to the motorway connecting several ports in Norway (a motorway characterized by heavy-duty traffic).

The dendrogram describing the clustering process for sampling points measuring PM_{10} in Oslo based on the 1-RxEuD metric is shown in Figure 31a; where the level of dissimilarity chosen is 3294 µg/m³. The ranking of sampling points are presented in Table 10.

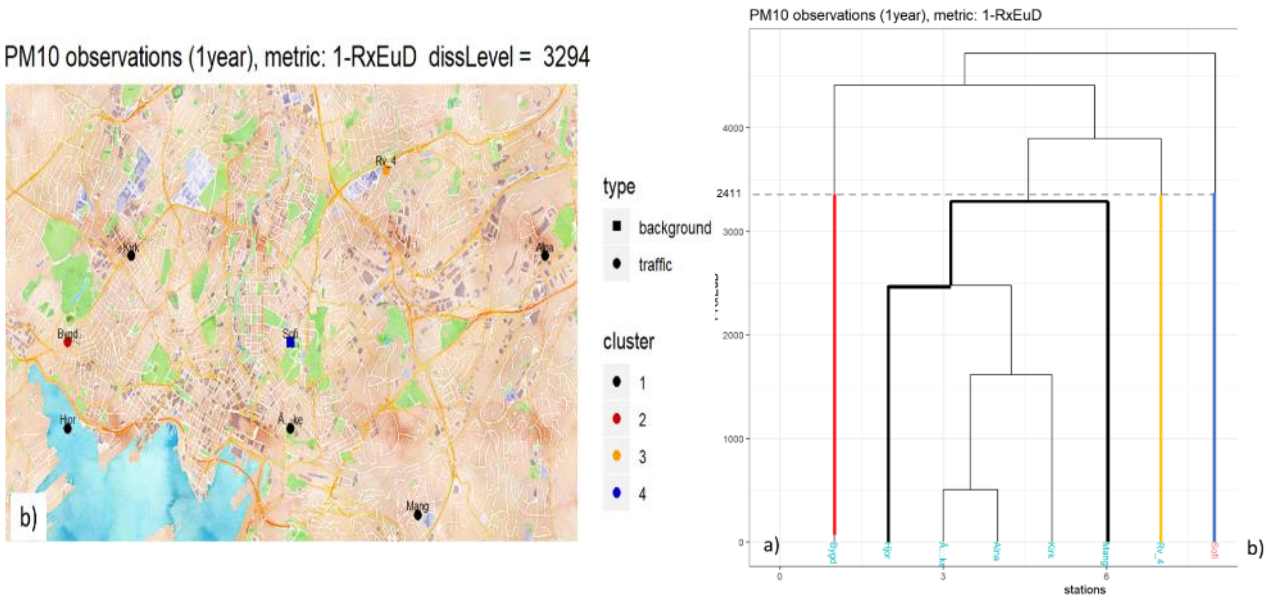

Figure 31: a) dendrogram for clustering of sampling points measuring PM_{10} in Oslo considering 1-RxEuD metric; the dashed line represents the level chosen to indicate unique and sampling points that present a higher level of similarity; and the coloured lines represent the sampling points clustered for that level on dissimilarity; b) spatial distribution of the clusters at the same level of dissimilarity, same colour code.

Table 10: Ranking of sampling points measuring PM_{10} in Oslo, from largest to smallest dissimilarity, for 2015

PM_{10} sampling point code	Sampling point name	1-RxEuD metric (Unit: µg/m³)
Sofi	Sofienbergparken	4720
Bygd	Bygdøy Alle	4414
Rv4	Rv4 Aker sykehus	3897
Mang	Manglerud	3293
Hjor	Hjortnes	2479
Krk	Kirkeveien	1616
Alna	Alnabru	507
Åke	Åkebergveien	507

Two sampling points are clearly ranked as the most similar sampling points within the network, Alnabru and Åkebergveien. These sampling points seem to have a very similar temporal profile and magnitude variation throughout 2015, and therefore could be considered as potentially redundant. The analysis indicates that the top three most dissimilar sampling points are Sofienbergparken, Bygdøy Alle and Rv4 Aker sykehus. Sofienbergparken is a background station that seems to represent very different profiles from the remaining sampling points.

The results from the hierarchical clustering for Oslo provides insights on the monitoring network in line with local knowledge. It also provides an interesting analysis of possible redundancies in the monitoring system that could be useful when considering an extension or a revision of the current monitoring network in the city. The system is easy to use and depends only on access to monitoring data with sufficient coverage for the year under study.

The hierarchical clustering methodology is simple to apply but requires local expert knowledge to understand the characteristics of the sampling points. This is particularly important, as the methodology can point out to outliers that are not due to specificities of the station but rather to the quality of the reported data not being optimal, either due to reporting mistakes or data collection setbacks. The levels of dissimilarity identified by the hierarchical clustering methodology are more a qualitative measure than quantitative. The results are dependent on the choice of the threshold levels for dissimilarity and therefore it is important to evaluate how distinct are the levels of dissimilarity throughout the clustering process and pin-point possible outliers and possible redundancies in the network, based on a correct selection of the threshold levels.

5.3 Tier 3 dendrograms for model validation and calibration

The hierarchical clustering methodology can be used for model validation and calibration purposes if modelling results are available in a certain region. In this case, the methodology is applied both to monitoring and to modelling results. The analysis of the modelling results is done in the same position where the sampling points are located. The comparison of the modelled and the monitoring data results from the clustering analysis provides then an evaluation of the model performance that considers both the spatial and temporal variations and provides a way to determine how well the modelling application reproduces the temporal variability within the area covered by the given datasets.

The clustering methodology is applied to the model results extracted at the station location to compare if the model results present the same clustering as the observations. If so, the model is considered “fit-for-purpose” and can be used further to guide the location of the new sampling points. In this sense, the model validation analysis is a first requirement for further analysis of monitoring design. Examples on how to use the clustering methodology for model validation purposes are given below for the cities of Antwerp and Oslo.

5.3.1 Model validation – feasibility test in Antwerp

The clustering results using 1-RxEuD metric based on model and observation are shown below. As it can be seen from Figure 32, the model and the observations provide similar clustering results at higher dissimilarity levels. This is an indication that the spatial and temporal variability of the model reproduces relatively well the observations of NO_2 in the city domain.

The dendrograms based on modelling results and observations are typically different starting with the levels of dissimilarity. Typically, the clustering analysis based on measurements have a larger range on the levels of dissimilarity and the model tends to show more homogenous fields, mostly because temporal profiles for sources are split according to categories and not individual sources. Here, one should focus on the clustering at higher dissimilarity levels. Comparing the dendrograms based on measurements and model results indicates some overlap in the clusters that are singled out for both clustering analyses. The largest differences between both methodologies are observed for the industrial sampling points located close to the main sources (shipping and industry) in the Port of Antwerp (e.g. station R897, R822). Emissions from these sources are approximated for the model (e.g. yearly emissions reported by the companies are distributed uniformly over all hours, shipping emissions are broadly estimated)[15]. The sampling points that are not showing the same degree of clustering in both analyses can be considered as potential “room for improvement” in the model, and they clearly correlated with known issues in the model.

[15] The main goal of the model exercise is providing (annual mean) concentrations for all locations in Flanders where people are living. Since the population density of the port area is very low, the concentrations in the Port have not been modelled using the best available datasets (e.g. for shipping).

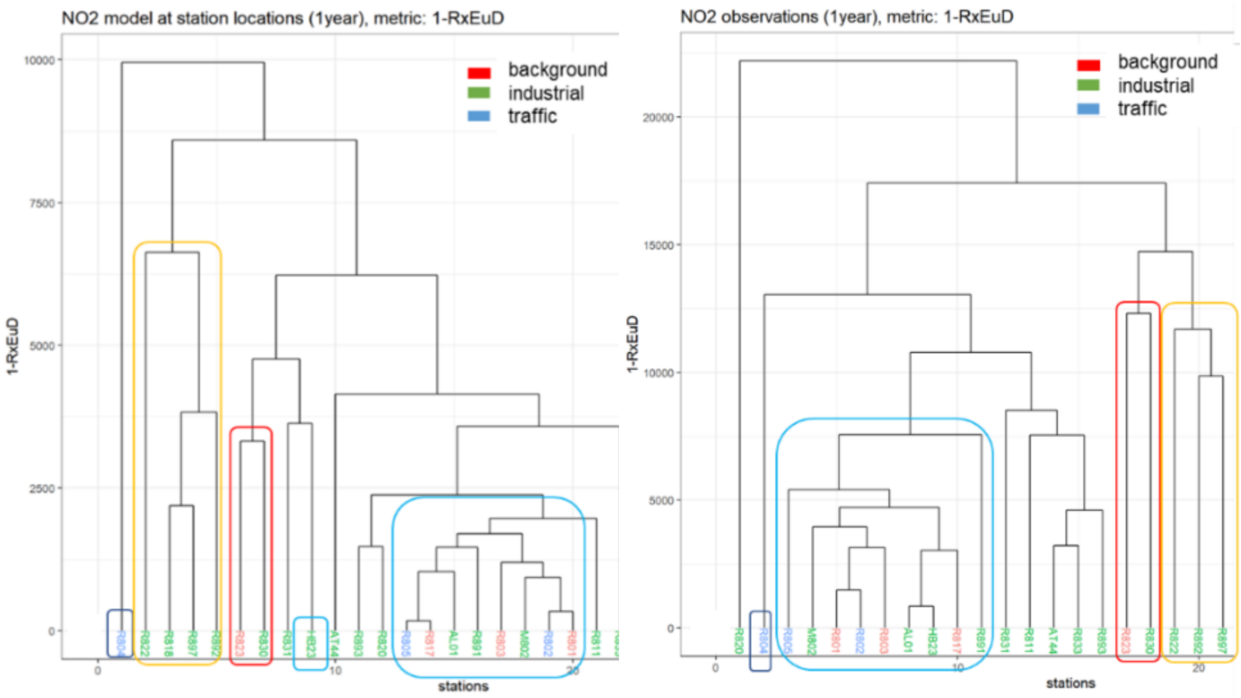

Figure 32: Dendrogram for clustering of station measuring NO_2 in Antwerp, in 2018, considering 1-RxEuD metric, based on model results at sampling locations (left) and based on the measurements (right).

There is good agreement shown in the dendrograms from the modelled and measurement data for sampling points in the (sub)urban locations. For instance, the sampling points R801, R802 and R803 are similar in both the model and the measured dendrograms, as are the sampling points in the rural areas west of the port of Antwerp (R823, R892). This is a good indication that the model is performing relatively well in the urban context and can be used for the next step of the assessment that is to check if potential new sampling points would benefit the network or not. In the second step of the analysis, we therefore mostly focus on the air quality zone concerning the city centre of Antwerp.

For PM_{10}, the results of the analysis are not as clear as for NO_2. Figure 33 shows the dendrograms for the clustering analysis based on observations and modelling results at station locations. The clustering results show that there is very little agreement between the two datasets, showing shortcomings in the model performance and making it less reliable to assess if potential new sampling sites would benefit the current monitoring network. These shortcomings in the model are due to several issues including dust-blown sources from construction sites which are poorly represented in the model, missing wood burning emissions mostly important for the suburban areas and the difficulty in accounting for repeated resuspension of particulate matter in street canyons.

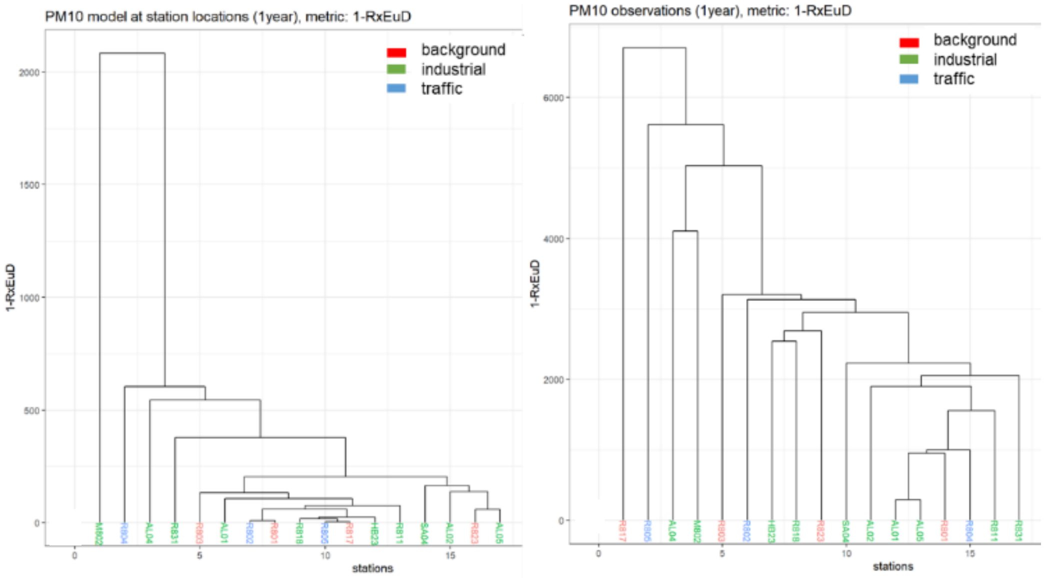

Figure 33: Dendrogram for clustering of station measuring PM_{10} in Antwerp, in 2017, considering 1-RxEuD metric based on modelling results at sampling locations (left) and measurements (right).

5.3.2 Model validation – feasibility test in Oslo

The same analysis was done in the city of Oslo. All sampling points for NO_2 are classified as traffic-oriented. The clustering results are presented in Figure 34 for both measurement and modelling results. The dendrograms mirror each other and show the same type of clustering of (dis)similarities for all the sampling points. This is an indication of the model's capability to reproduce the observed NO_2 levels and their temporal variability in the city of Oslo, and thus an additional characterization of the modelling application as "fit-for-purpose".

For PM_{10}, the results of model validation in Oslo are presented in Figure 35. In Oslo, the urban background station in Sofienbergparken showed a clearly different behavior in the monitoring dendrograms than all the other traffic-oriented sampling points. However, the (dis)similarity dendrograms in the modelling results do not show the same results. The complexity of different sources affecting PM_{10} concentrations is reflected in the fact that the modelling results show more difficulties reproducing the temporal variability of the observations. These results are similar to the ones in Antwerp, where the modelling application also showed less capability to reproduce the observed PM_{10} variability than that of NO_2. This is an indication both of the complexity of the PM_{10} air quality situation and of the stringency of the dendrogram analysis as an indicator of model performance.

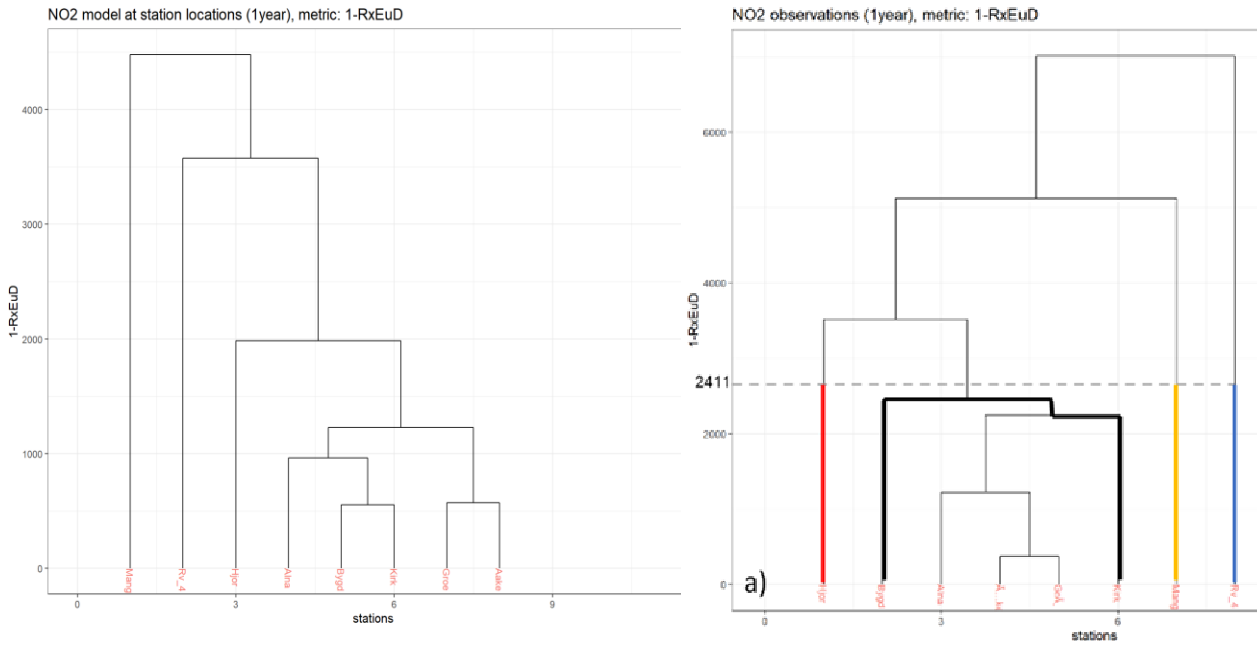

Figure 34: Dendrogram for clustering of station measuring NO_2 in Oslo, in 2015, considering 1-RxEuD metric based on modelling results at sampling locations (left) and measurements (right).

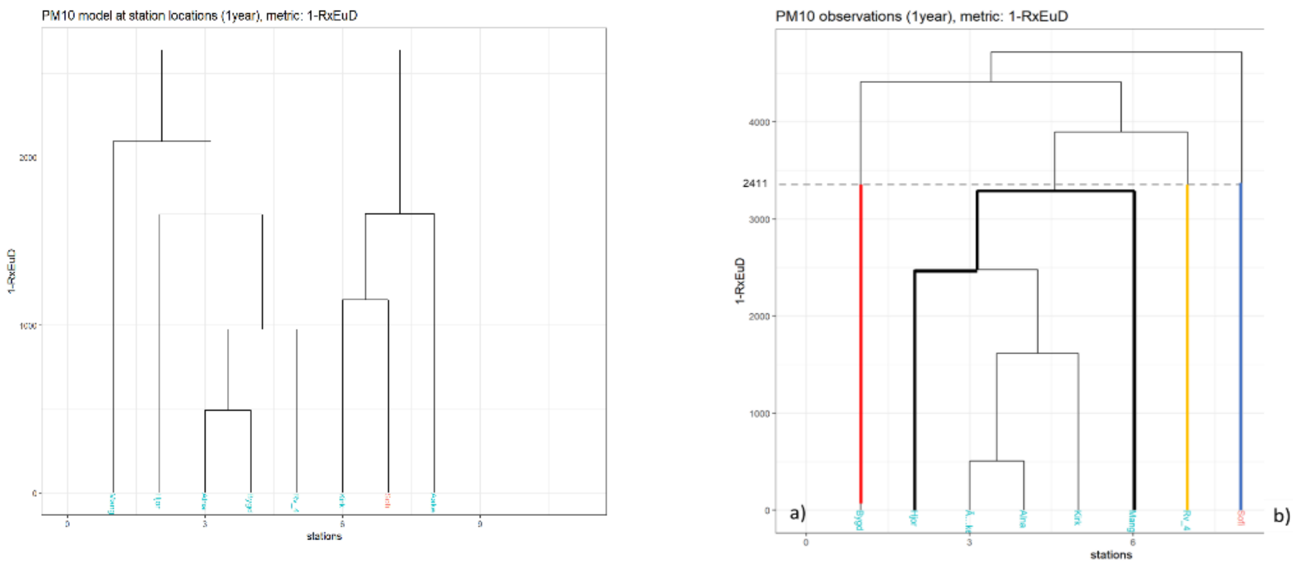

Figure 35: Dendrogram for clustering of station measuring PM_{10} in Oslo, in 2015, considering 1-RxEuD metric based on modelling results at sampling locations (left) and measurements (right).

As shown in the examples above, the clustering methodology provides an additional evaluation of the model performance that considers both the spatial and temporal variability within the area covered by the given datasets. The use of the temporal variability in the assessment of model validation represents an addition to the current approach to Model Quality Objectives (MQO) that may be interesting to evaluate further. The use of the temporal variability in the assessment of monitoring network design allows the identification of clusters with similar air quality regimes to guide the siting of sampling points in representative areas.

5.4 Tier 3 clusters for extensions of monitoring network design

Once a model is considered “fit-for-purpose” based on the analysis of the dissimilarity dendrograms for a specific component or by other relevant means, the modeling results can be useful to support monitoring network design activities. Modelling results can support the evaluation of possible new locations of sampling points and can support the design of future extensions of monitoring networks.

In the following we present how the use of hierarchical clustering applied to modelling results can support in the design of revisions and extensions of current monitoring networks. The first example, in Antwerp, shows the potential of using dissimilarity dendrograms with model results at the grid locations of the new sampling points and how this can help in the design of an extension of the current monitoring network in the city. The second example, in Oslo, shows the use of clusters as a complementary approach to the revision/extension of monitoring networks and indicates how these clusters can serve to identify the area of representativeness of the sampling points.

5.4.1 Extension of monitoring network – feasibility tests in Antwerp

Two years ago, the Flemish Environmental Agency (VMM) identified several locations for a passive sampler campaign. These locations have been chosen based on expert opinion, with two goals in mind. Some of the locations are chosen to further underpin the representativeness of the current measurement sampling points, while other locations have been added to sample a few specific locations (e.g. tunnel exits, and street canyons close to the main highways). A sub-selection of 18 additional locations has been selected, as identified in Figure 36.

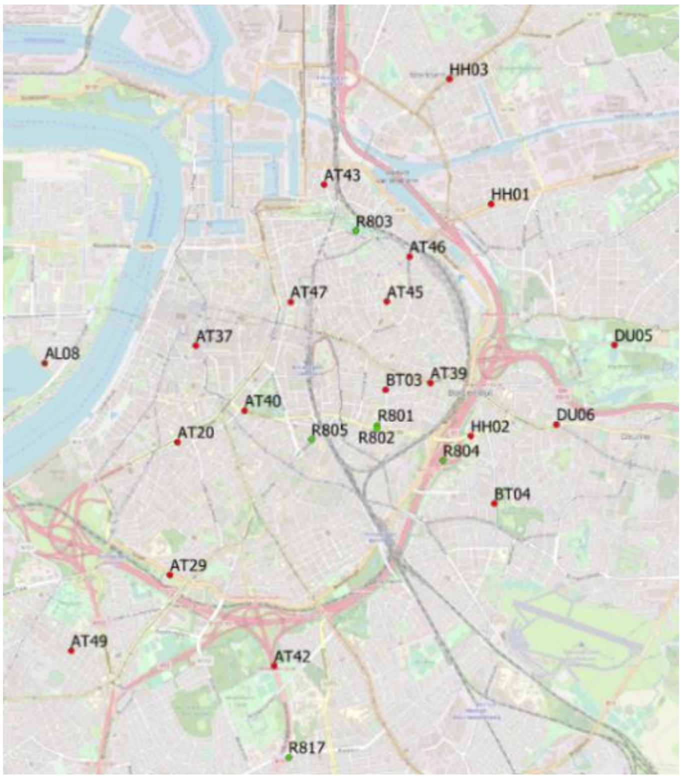

Figure 36: Location of the new sampling sites (red) and the station locations (green) in the air quality zone Antwerpen Centrum

To evaluate the capabilities of the proposed new locations for sampling points, we have carried out a feasibility test using hierarchical clustering. Since the modelling results in Antwerp were considered fit-for-purpose for NO_2, the clustering methodology has been applied using the model results both at existing and new monitoring sampling points.

shows the results from the hierarchical clustering methodology based on the modelling results. The dendrogram shows that samplers AT49, AT45 and AT37 cluster at low levels of dissimilarity with existing monitoring sites (highlighted by the black cluster in Figure 37). This indicates that these samplers are probably not bringing any additional benefit to the air quality monitoring network already in place in Antwerp. These samplers are located at urban background and street canyon locations within the Antwerp Ring Road, which are also covered by other measurement sampling points (R801, R802, R803, R805). Similarly, two sampler locations with very high pollution (DU06, HH04) are in the same cluster as the station R804 (purple cluster on Figure 37, which is located very close to the main highway (and with the same orientation towards this source as the sampling locations). All these locations are therefore already covered by the current monitoring network.

On the other hand, AT42 is visibly the outlier of the analysis (blue cluster on Figure 37). This new sampling site is located very close to the tunnel exit of a principal highway (linking Antwerp and Brussels). The location is thus very specific and deliberately added to the current analysis because local experts considered it as poorly sampled. As there is no public access at this site it is not suitable for compliance assessment under the AAQD siting criteria. In addition to AT42, there is a very large cluster of new sampling sites (AT40, AT39, HH01, AT20, AT46 and HH03, i.e. the green cluster on Figure 37) that are located in busy street canyons. The analysis indicates that this type of site is missing from the current monitoring network and thus has potential to add value to the city monitoring network.

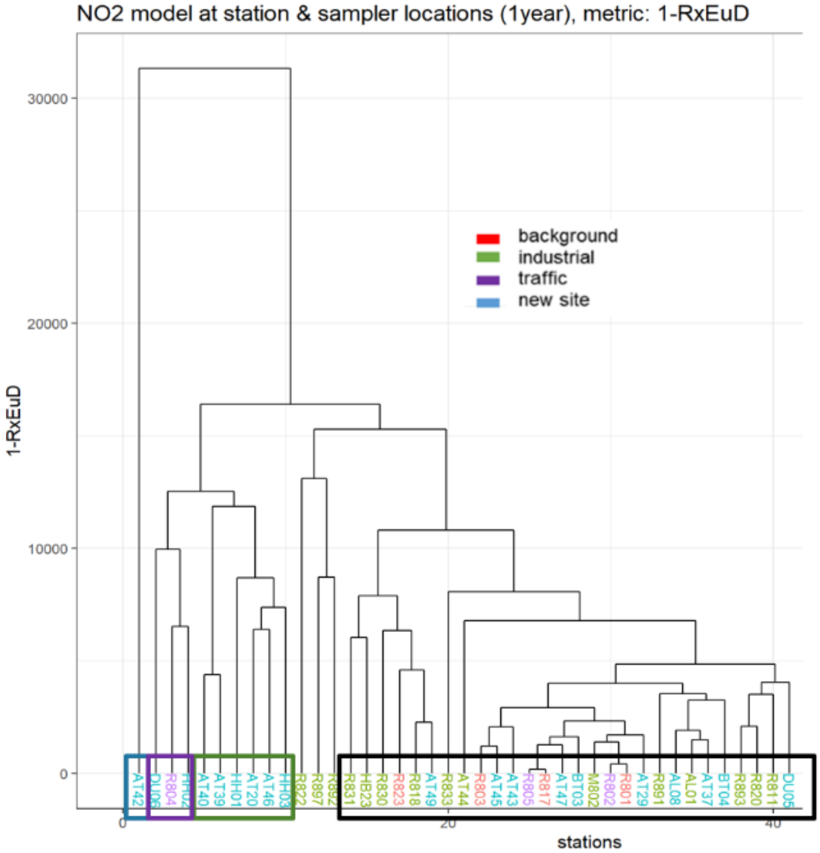

Figure 37: Dendrogram for NO_2 in Antwerp, in 2018, considering 1-RxEuD metric, based on modelling results extracted at station location and at potential new sampling sites location. New sites are given in light blue.

The feasibility tests in Antwerp show how the use of dendrograms visualizing the dissimilarity classes can help when preparing an extension of the monitoring network in a city. The methodology can support the evaluation of optimal new locations for new sampling points, screening the proposed new locations to avoid redundancies and secure good coverage of different air quality situations, thus contributing to the optimisation of the monitoring network.

5.4.2 Extension of monitoring network – feasibility tests in Oslo

In this feasibility study for the city of Oslo, no specific new sampling locations are identified. Instead, the analysis of possible extensions of the monitoring network in Oslo are done on the basis of modelling results alone. The hierarchical clustering methodology is applied to all grid points in the Oslo modelling domain, in order to identify (dis)similarity clusters in the air quality regimes in the city area. For this assessment, it was assumed that each grid cell in the modelling domain is a potential station and the number of clusters requested is the same as the number of sampling points available for 2015. These clusters are, in principle, representing areas with specific temporal profile and concentration levels, meaning unique sources and meteorological conditions, e.g., up- or downwind of prominent sources. The example below is for the city of Oslo, where 9 sampling points were active in 2015. Ideally, there should be a monitoring sampling point for each of the identified clusters, where the cluster area would be the area of representativeness of each monitoring sampling point.

Figure 38 shows 9 different areas of similarity for NO_2 concentrations in 2015. The figure in the right shows a larger domain of the city while the left picture shows the clustering around the city center where most sampling points are located. The black dots represent the actual location of the current monitoring sampling points with NO_2 sampling points. As it can be seen, the current sampling points are placed in 3 different clusters in the large domain and 4 clusters in smaller domain. The choice of the domain is relevant to understand whether there are or not areas of interest in the urban domain that are not covered by the current sampling points. In the case for Oslo, the current monitoring network for NO_2 has sampling points in the most relevant clusters, where people live. The figure shows some of the possible redundancies in the current monitoring network in Oslo as already identified with the dendrograms in section 5.2. for the two stations. Grønland and Åkerbergveien in cluster “2”. It also identifies two cluster areas, where it would be appropriate to locate monitoring sampling points – in cluster “3”, “4” and “7” where currently there is no sampling point representing the cluster/air quality regime area. In addition, the figure helps identifying similarity clusters, that can be used as an indication of the area of spatial representativeness of the sampling points inside the clusters.

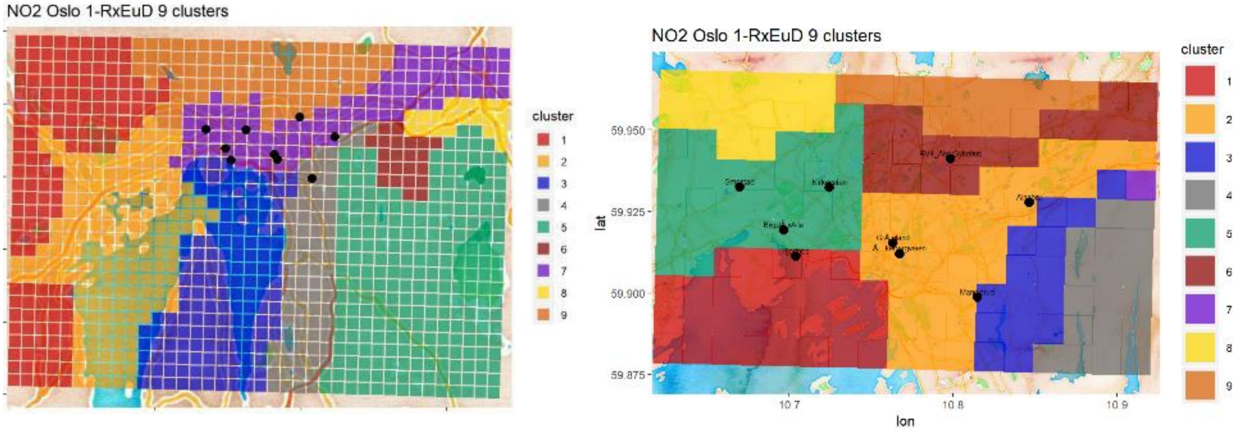

Figure 38: SR Clusters for NO_2 in Oslo in 2015.The position of the monitoring sampling points are given as black dots. The number of clusters is 9 in connection to the nine sampling points currently operating in Oslo. Left panel , 9 clusters in the large Oslo Domain, right panel, 9 clusters in the city centre.

The number of clusters can be increased to consider within which areas in Oslo possible new sampling points for NO_2 could be located. By increasing the number of possible sampling points to 15, the exercise allows to also identify areas of (dis)similarity inside the clustering in Figure 38 – the current monitoring network. The exercise serves to identify particular hot spot areas that can be candidate locations for additional monitoring. In Figure 39, the clustering analysis of the model results is provided for 15 possible new monitoring stations. The numbers in the clusters refer to the original clusters in Figure 38. So, for instance, the original cluster “2” with four sampling points is now divided in two, the “blue” and the “yellow” clusters” justifying the placement of the “Alnabru” sampling point and consistent with the results of the dendrograms that showed this sampling point location with high level of dissimilarity. The analysis of Figure 39 indicates again the value of adding sampling points in the original clusters “3”, “4” and “7” and confirms the latter as a very specific location that may be worth considering for future monitoring sampling point locations.

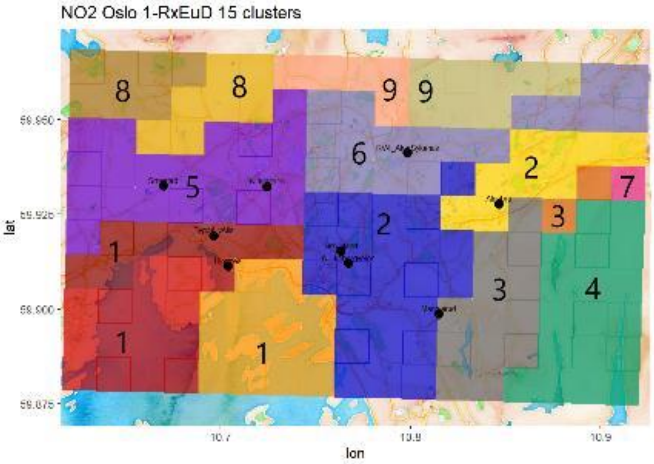

Figure 39: Possible extension of the monitoring network to 15 sampling points and corresponding additional SR Clusters for NO_2 in Oslo. The numbering in the clusters corresponds to the ones in Figure 40, left panel.

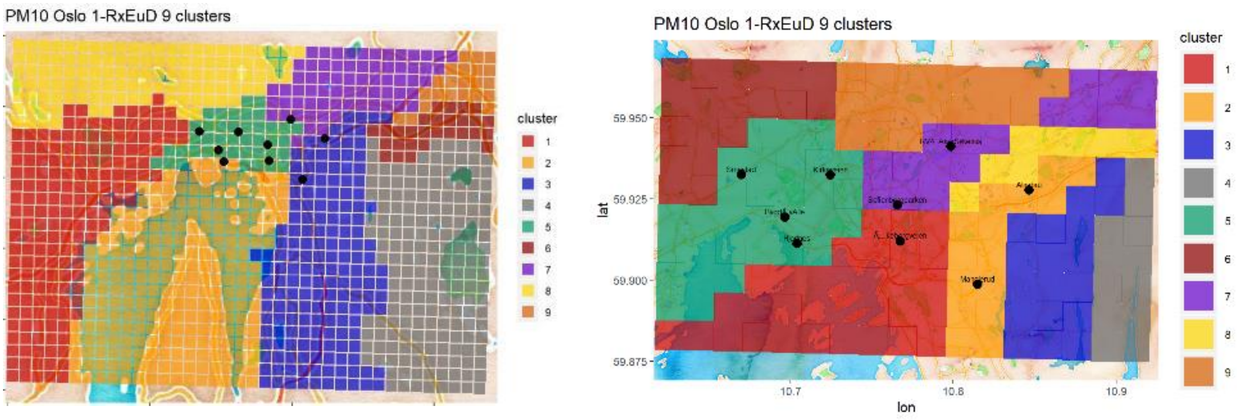

Figure 40: SR Clusters for PM_{10} in Oslo in 2015.The position of the monitoring sampling points are given as black dots. The number of clusters is 9 in connection to the nine sampling points currently operating in Oslo. Left panel, clusters in the large Oslo domain, right panel, clusters in the city center.

Figure 40 shows the nine cluster regions for the operational PM_{10} sampling points in Oslo in 2015. The figure shows that the current monitoring network covers three clusters in the large modelling domain and four clusters in the city center. The clusters of spatial representativeness for PM_{10} are different than for NO_2 as it should be expected. There are some similarities as the clusters along of the fjord, with clusters around the coastline contrasted with inland, and different population density across the domain, reflecting the variation in type and number of sources. Cluster number “7” is the one with largest level of dissimilarity – as shown when adding the number of clusters to 15 or 25 as shown in Figure 41. The analysis identifies two areas where possible additional sampling points could be considered: cluster “8”, the yellow cluster, and again cluster “3”, the blue cluster, where people live but is currently poorly represented with the existing monitoring sampling points.

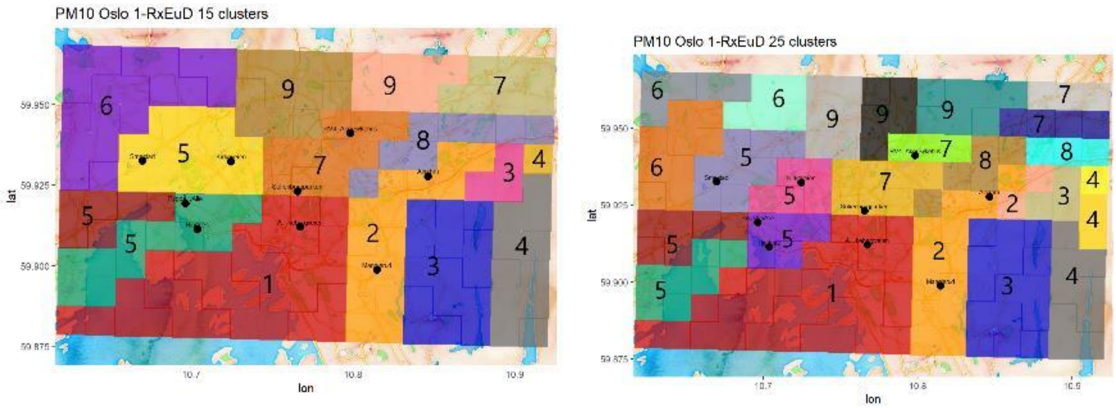

Figure 41: Possible extension of the monitoring network to 15 (left) or 25 (right) sampling points and corresponding additional SR Clusters for PM_{10} in Oslo. The numbering in the clusters corresponds to the ones in Figure 40 left panel.

The feasibility tests show that the methodology can also be used to assess possible locations for new sampling points using model results. The use of the hierarchical clustering for given locations is easy to use and for uses as in the example in Antwerp, it requires only a yearly set of hourly data in given locations, so that the data storage needs are manageable. For the use of the cluster data as done in Oslo, to investigate the extension of the monitoring network using modelling grid data, the data storage needs are much larger, because the yearly set of hourly data is necessary at each grid point in the modelling domain. The methodology can also be applied to observational data from campaigns to explore for new locations where to set-up new sampling points, in this case the request for data is less extensive than for the colour cluster of representativeness. Despite the large need of data for carrying out the cluster analysis as in the Oslo example, the analysis is useful for assessing area of representativeness for a monitoring sampling point. This type of clustering analysis if complemented with information such as population density and road distance would enable the clustering methodology to be used for exposure and exceedance situation indicator purposes.

6 Conclusions

In this final section, we formulate the main conclusions for the assessment of spatial representativeness (SR) of monitoring sampling points in the context of the estimation of exposure and exceedance situation indicators, monitoring network design, and the calibration and validation of modelling. The methodology for undertaking SR assessment is an open issue which has been discussed in the air quality monitoring and modelling community for a while, and was also identified in the FAIRMODE/AQUILA inter-comparison exercise (Kracht et al. 2017). In this project, we have addressed the assessment of spatial representativeness by proposing a Tiered framework approach. This approach recognises the need for a flexible framework to carry out SR assessment that distinguishes the use of different methods depending on the purpose of the assessment.

The recommendations below have been drafted in order to provide pragmatic and flexible solutions for future assessments of SR. The recommendations have been discussed during a number of events with the relevant European air quality communities (Air Quality Expert Group, FAIRMODE, AQUILA, EIONET, CAMS, IPR). A summary of the feedback collected is presented in Janssen et al, (2020).

We have found in this project that although reporting of EU air quality assessments in Member States largely complies with the provisions in the AAQDs, there are significant gaps in coverage and inconsistencies in reporting of data related to sampling point SR. We also noted that for the topic of SR, there is a systematic lack of consensus and limited guidance on how to assess and subsequently report to the EC. In this project, we have aimed to support the reporting practice of sampling point SR through:

- Improved understanding of the concepts of SR in terms of methodological guidelines to specify as clearly as possible the definition of SR to be used for different purposes
- Added clarity on the classification of different methods for different purposes with the proposal of a Tiered approach where higher tiers deliver more robust or accurate results but also imply the need for more resources or more technically demanding approaches.
- Identification of possible methods and key issues that need to be addressed under further guidance on the tools/methods and resources needed to implement the different Tiers.

The conclusions below present findings on the proposed Tiered approach for the three assessment purposes defined for EU air quality reporting, namely, exceedance situation indicators, network design and model validation. In order to assist the reader, short summaries of the work contained in the earlier sections of this report are provided where considered necessary to aid understanding of the conclusions drawn.

6.1 Tiered approach

Discussions within the air quality expert community and analysis of the CDR database have made it clear that Member States are still mostly relying on monitoring information and expert opinions for the estimation of the SR of monitoring stations, of the exceedance situation indicators to be reported under the IPR and the design of monitoring networks. It was also recognized in the past that this approach has clear limitations on the accuracy of the results, and that different practices may hamper comparison amongst air quality zones and Member States. Limited inputs are currently made from modelling results. Although stakeholder feedback indicates modelling capability is there it is not utilized for IPR reporting by many Member States at present. During these stakeholder interactions (e.g. the Air Quality Expert Group, see Janssen et al. 2020) Member States recognized the potential of more complex model based approaches and expressed a clear willingness and interest to further support this type of assessments.

This project proposes a Tiered framework approach for the assessment of SR for each of the 3 assessment purposes identified in the AAQDs. The classification of the methods currently used for determination of exposure and exceedance situation indicators, network design and model validation in the Tiered approach is intended to increase the common understanding of the concepts of SR. In this framework higher tiers deliver more robust or accurate results but also imply the need for more demanding approaches. In this way, it is expected that the use of a Tiered approach for documenting the methodologies used could inform a higher degree of comparability across Members States and allow assessments that involve comparisons across the EU.

The Tiered framework approach is summarised in Table 11 classifying the different SR assessment methods in different Tiers per assessment purpose based on the conclusions from the Literature Review of Maiheu and Janssen (2019).

The feedback from the stakeholder community (Janssen et al. 2020) was positive to the adoption of this Tiered approach to frame future guidance and evaluation of the assessment of SR. The community responded that there is good coverage of current practices within the Tiered structure and that it provides clarity for the classification of methods and is flexible enough for new methods to be included. In particular, the air quality expert community recommended considering the inclusion of approaches based on machine learning or other statistical methods also under Tier 2 in future iterations.

Table 11: SR assessment methods in different tiers per assessment need (replicated from Table 1).

<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>

6.2 Exceedance situation and exposure indicators

At present lower Tiered approaches are commonly used to support the evaluation of sampling point representativeness for the purpose of evaluating both exceedance situation and exposure indicators over a given area. However, as demonstrated in the sensitivity analysis of this report, higher Tier approaches provide more reliable and robust results, although they may be more demanding to implement. To ensure better comparability across Member States, all approaches, regardless their Tier level, would benefit from a series of guidance choices and clarifications.

The e-Reporting process setup under Commission Decision 2011/850/EU (the so-called IPR Decision) requires exceedance situation indicators to be derived for the entire air quality zone under consideration. Almost by definition this requires an extrapolation of the exceedance(s) observed at the monitoring station(s) to other locations within the zone to assess the related exceedance situation indicators requested in the IPR Decision (area, number of people exposed, road length). Further to the analysis presented in Sections 3 and 4 of this report, fit-for-purpose models (Tier 3 or Tier 4) are, in our view, the most useful tool to perform this task as they contain our best available understanding of the relevant emission sources and dispersion and transformation processes in the atmosphere and have been validated using independent measurements. Lower Tier approaches should be used with care in this extrapolation process to assess exposure and exceedance situation indicators.

It remains an open question how street canyon effects should be included to indicate the number of people exposed to exceedances. In a static population exposure assessment, an overlay is made between concentration maps and population at home addresses. Depending on the assessment method the canyon increments can contribute to the final exposure result. When these exposure maps are used for health impact assessments there might be arguments to not include the canyon effects for the time being. Currently available concentration-response functions make a link between observed health effects and (urban) background concentration levels typically used in the epidemiologic studies (Maiheu et al 2017). The inclusion of canyon increments would therefore disturb the applicability of these health impact relations. However, the results of a static population exposure analysis can be further improved using a more dynamic approach. In such an approach canyon effects and on-road concentrations should be included since people spend some time in traffic, and inhabited buildings can be located right next to the streets. Based on these dynamic exposure estimates, updated and more refined concentration-response functions could be derived for more detailed health impact assessment.

It should be noted that exceedance situation indicators are very sensitive to changes in the methodology and the input data. The main recommendation from this work and subsequent expert consultations is to develop detailed guidance on the actual calculation method for exceedance situations. Such guidance plus the requirement of reporting and documenting the applied methodologies (e.g. how to combine population data with air quality data) and the input datasets (e.g. which type of road link data to use) is very important for the comparability of results from different countries and different cities. Expert consultation indicated that the indicator of road length in exceedance is bringing little added value and its relevance is questioned.

6.3 Recommendations for technical clarifications in the definition of spatial representativeness of monitoring sampling points

The estimation of an SR area of a monitoring site requires specific choices in the methodological approach. Based on the Tier 3 analysis presented in this report we have built up three recommendations, that in principle are valid for all tiers, and are put forward to further guide and harmonize the derivation of SR areas.

6.3.1 Contiguity

The SR area of a monitoring site can be defined as a contiguous area connected to the location of the station. In this case, any location in the SR area is connected to the monitoring site via a path that fulfils the similarity criterion. In an alternative discontiguous definition the SR area can be composed of disconnected subareas, each of them fulfilling the similarity criterion. The difference is illustrated for an urban background NO_2 station in Antwerp in Figure 42.

As demonstrated in the sensitivity analysis reported in section 3.2.2, a contiguous approach for identifying the SR area of a monitoring station is not a useful approach in practice. The use of a contiguous approach introduces several practical problems:

- Contiguous areas are overly sensitive to small variations in threshold values, model, and measurement uncertainty.
- Very often contiguous areas are limited in size.
- Very often, similar locations in an air quality zone are disconnected from areas that would otherwise be deemed similar in SR, by physical obstacles such as a major road or an opening in a street canyon.

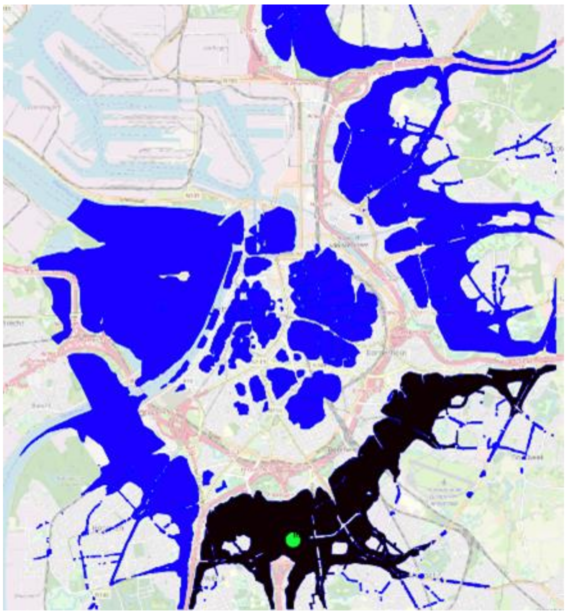

Figure 42: Contiguous (black) and discontiguous (black+blue) SR area of an urban background NO2 station (green dot) in Antwerp (replicated from Figure 8a).

It is therefore useful to opt for a discontiguous approach in the assessment of the SR area for air quality monitoring sampling points, as this will enable a more robust and consistent approach to determine the spatial representativeness of the sampling points. Allowing for discontiguity in SR areas is also compatible with the AAQD (Directive 2008/50/EC, Annex III.B.1(f)) requirement for the macroscale siting of sampling points that shall, where possible, also be representative of similar locations not in their immediate vicinity.

When a discontiguous approach is adopted the request for the definition of a maximum spatial extent of the SR area naturally appears. With a view to the general approach taken to reporting under the AAQDs per air quality zone, we suggest that the SR area is limited to the spatial extent of the air quality zone used in the AAQD assessment process. For urban or industrial locations, the AAQD air quality zone can be limited in area whereas for more rural locations, the zones are larger and consequently the discontiguous area can span large parts of a region or a country. By using the air quality zone as a pragmatic boundary for the SR area, the arbitrary choice of a maximum distance to the sampling location (e.g., 50 km) is avoided.

The consultations with the expert community showed a tendency in favor of a discontiguous approach for the calculation of SR of sampling points. However, further guidance in the details of implementing, and testing of the implications, of this approach in other European cities was recommended.

6.3.2 Similarity Criterion

The similarity criterion defines how locations within a SR area are “similar” to the location of the monitoring site. In the Literature Review it was illustrated that some SR methodologies make use of complex and multiple similarity criteria. This complexity was driven by scientific arguments to capture many effects of spatial representativeness of monitoring sites. On the other hand, Member States require transparent and practical tools to assess SR of monitoring sampling points.

Therefore, it is recommended to keep the similarity criterion as simple as possible so that it can be easily and consistently applied. For the AAQD exceedance situation indicators it is therefore recommended to use the related temporal aggregation, i.e. annual average concentrations, as the main similarity criterion. Exceedances of e.g. daily limit values or percentiles are much more difficult to derive although eventually also these indicators have to be assessed in the context of the AAQD. Tier 3 tools provide a solid basis for such a more complex assessment.

We do not recommend bringing in other secondary similarity criteria and context related information such as urban topology or proximity to important sources. This would hamper the practical implementation of the spatial representativeness methodology. However, in specific applications, such as for network design and for validation of modelling results, the inclusion of time information could be useful. This was demonstrated in the feasibility tests reported here for Antwerp and Oslo (see Section 5).

6.3.3 Threshold values

To apply a similarity criterion a specific threshold value should be selected. For example, when annual mean concentrations are used as a basis for the SR assessment, a quantification of the tolerance level or threshold of allowed deviation from that concentration level is required. Based on the sensitivity analysis (see Section 3.2.3), we recommend applying a relative threshold in the range of 10 to 20 %. Smaller thresholds are incompatible with the intrinsic uncertainty related to the assessment method, regardless if it is based on measurement or model data. Larger thresholds would allow too much variation within the SR area. The analysis showed that there is a preference for relative thresholds (as %) whereas absolute thresholds (as a concentration value in µg/m³) hinder the comparison of results from different cities and regions. For low concentrations, for example in rural areas, a relative value can become problematic since it becomes too small and a fixed absolute threshold value (e.g. 2 µg/m³) would be more suitable for these lowest concentration ranges.

The threshold value could also be defined in relation to the measurement uncertainty as is done in the definition of the FAIRMODE Modelling Quality Objective and it can even be pollutant dependent. However, given the relatively low maturity of current SR assessment methodologies and the lack of common understanding of the concepts, it is deemed that a transparent and simple threshold value is at present more useful to air quality practitioners than a quantity which is fully underpinned by scientific arguments.

As the size of the spatial representative areas vary smoothly with the threshold (at least in the discontiguous approach), selecting a specific threshold in the range from 10 to 20 % does not introduce too much uncertainty in the analysis. For practical reasons we therefore propose a conservative but transparent threshold of 20 % for NO_2 and PM_{10}.

The discussion with the expert community concluded that a combined threshold function based on a relative value and an absolute threshold for low concentrations might be useful. Whilst a clear majority of the expert community supports the use of a 20 % threshold value as a starting point, it was still recommended to gather more evidence about this value. The community also recommended to test different thresholds further in other European cities and regions. Such testing activities can be coordinated within the FAIRMODE network.

6.4 Network design and modelling validation

The design of the air quality network in Europe follows the AAQDs. The results from the Task 3 report (Tarrason et al., 2020) showed that Member States in many cases use Tier 1 approaches to determine the siting and classification of the air quality monitoring network design. The use of higher Tier approaches involves independent screening methods to help qualify the quality of the monitoring network. This report presented a series of feasibility tests on the use of a hierarchical clustering methodology to support Tier 2 and Tier 3 approaches for monitoring network design and model validation purposes.

The Soares et al. (2018) clustering methodology tested here allows for a screening evaluation of the siting and classification of a given monitoring network based on the actual measurement data. When applied only to monitoring data, it constitutes a Tier 2 approach. This hierarchical clustering methodology serves to analyse the level of similarity or dissimilarity of air pollutant concentration data from all sampling points. It also allows the identification of sampling points with similar behaviour and those with specific behaviour that differs from the rest of the data sampling points, the so-called “outliers”. In this sense, the clustering methodology represents an independent screening of the monitoring network classification carried out in Tier 1. The clustering methodology at Tier 2 is simple to use and allows for additional expert evaluation of the sampling point classification to identify redundancies and gaps in the monitoring network. Because of this possibility for actual feedback on the Tier 1 classification, this clustering methodology is more valuable than the previous Tier 2 approaches from Joly and Peuch (2012) that only provided an alternative sampling point classification and was difficult to relate to the classification in Tier 1.

The clustering methodology also can be used for validation and calibration purposes of modelling results, supporting Tier 3 approaches. In this case, the methodology is applied both to monitoring and to modelling results in the same position where the sampling points are located. The comparison of the modelled and the monitoring data results from the clustering analysis provides an evaluation of the model performance that considers both the spatial and temporal variations and provides a way to determine how well the modelling application reproduces the temporal variability within the area covered by the given datasets.

The similarity clusters can be visualised in the form of dendrograms, as tested here for Oslo, and provide an alternative procedure to model validation. The use of dendrograms for model validation requires a manageable amount of data and expert evaluation to set and explain the different dissimilarity levels. It is a good candidate to become a complementary approach to the current Model Quality Objectives (MQO) promoted by FAIRMODE for model validation purposes.

As a Tier 3 approach, the hierarchical clustering can be applied to modelling results to further support possible extensions of the monitoring network, identifying redundancies in the monitoring network as well as gaps. This Tier 3 approach to monitoring network design involves the use of fit-for-purpose modelling data for all grid points in the modelling domain to identify clusters of similarity with respect to modelled air quality regimes. The feasibility study in Antwerp and Oslo has shown how to use both dendrograms and geographically coloured clusters to evaluate the position of possible new sampling points in a network. The number of clusters may be constrained by the total number of sampling points in the monitoring network to allow cost-effective evaluations given realistic budget constraints. It can also be used to identify an optimal siting and location of the given number of sampling points. This approach allows for the description of spatial representative areas in the modelled domain but requires a considerable amount of data to be stored and processed. It should be noted that although this final type of SR clustering analysis can be valuable both for monitoring design and for exposure calculation applications, it also involves significant data requirements. To further assess the feasibility of the approach, it needs to be further tested and evaluated in different locations.

Under consultation with the expert community, the proposed clustering method was considered a valuable approach for understanding specific aspects of network design. It was acknowledged as an interesting additional approach to the evaluation of monitoring networks as it can be useful to identify gaps, outliers, and possible enhancements to the existing network. If further tested by national experts, it may prove a valuable approach to combine modelling information for monitoring design purposes. The FAIRMODE and AQUILA communities could work together to enable further testing of the clustering approaches for monitoring design applications.

The proposed clustering method or similar approaches have the potential to bring added value to a model evaluation exercise and may be used as an additional modelling quality performance test. However, these approaches should be broadly tested in various locations by national experts before they can be endorsed for widespread use under the AAQDs. The FAIRMODE and AQUILA communities could work together to enable further testing of the clustering approaches for model validation.

6.5 Suitability of Tiered methodologies

The sensitivity study reported here compared different Tiered methods for sampling points in Antwerp, Oslo and Krakow. The wide variety in the output clearly illustrates the complexity of the problem and the lack of comparability amongst some of the lower Tier approaches. This is consistent with the results of previous Inter-comparison Exercises (Kracht et al, 2017). However, the analysis carried out in this study allows an initial evaluation of the suitability of different Tiers.

Tier 1 methodologies might be applicable for the assessment of the SR area of rural and urban background sampling points. At these locations, concentration gradients vary more smoothly, and expert opinion can give a reasonable first order estimate. When exceedances are reported in these stations, a conservative Tier 1 approach is to assume the entire air quality zone is in exceedance of the limit value and derive the related exceedance situation indicators under this assumption. For industrial or traffic sampling points, a more comprehensive Tier 2 or higher approach would be required.

Tier 2 methods based on proxy data can complement and support Tier 1 expert opinion and contribute to a further harmonization and comparability of results amongst air quality zones. Proxy based Tier 2 methods can be used to assess the SR area of rural and urban background stations although the outcome should be handled with care. Population density, land use information or emission maps can be used as proxy data and spatial drivers in the assessment method. Areas close to major (point and line) sources should be excluded from the SR area. Although Tier 2 methods are not recommended for the more complex traffic stations in an urban setting, the SR area and its related exceedance situation indicators could be estimated based on a selection of locations with similar traffic intensities and comparable street canyon configuration. Again, such an assessment should be handled with care.

Tier 2 methods based on sampler campaigns could provide an alternative if the number of sampling locations is large enough and if the sampling locations are carefully selected to correctly measure the (potentially heterogeneous) concentration field in the air quality zone under consideration. This might be the case for the delineation of the area of SR and the estimation of the number of people exposed to exceedances, whereas it is unclear whether such campaigns can be used to estimate the area and road length in exceedance as no such method has yet been implemented. It should be noted that such campaigns require a substantial investment. However, such sampling campaigns could also inform model development and support the achievements to arrive at a fit-for-purpose Tier 3 approach.

The location of additional monitoring sampling points may be identified with the help of the hierarchical clustering methodology presented here which is also useful to identify redundancies in the network as well as outliers when used as a Tier 2 approach.

Tier 3 approaches are generally recommended for the evaluation of the exceedance situation indicators. For the assessment of SR indicators, it is in general recommended to use fit-for-purpose models (Tier 3) over Tier 1 or Tier 2 methodologies based on expert judgment or proxy data. Air quality models represent our best available understanding of the spatial dimension of the sources of air pollution and the dispersion and transformation processes in the atmosphere. In lower Tiers these mechanisms are at best mimicked by proxy data or simple distance rules.

For monitoring network design purposes, in combination with modelling results as a Tier 3 approach, the use of hierarchical clustering provides additional insights on model validation and guidance for evaluation and optimization of monitoring network design. It allows the definition of SR area clusters that may be used as a basis for identifying gaps in the monitoring network design and can support the calculation of exceedance situation indicators. An evaluation of the potential of the hierarchical clustering methodologies for use in monitoring network design activities and to support exceedance calculations could be carried out as a follow-up from the results of this study.

To account for shortcomings in modelling applications, it is always recommended to complement models with additional measurement data in a **Tier 4** approach. Such a data fusion process will remove biases and capture unexpected spatial and temporal variations.

All these recommendations related to the Tiered approaches are summarized per assessment need in Table 12.

		SR area of sampling points (based on annual mean concentrations)	Exceedance Situation indicators (area, no. of people, road length)	Design of monitoring network	Sampling points for model calibration and validation
Tier 1	Expert Opinion	• Only for (urban) background sampling points • Not recommended for traffic sampling points • Unclear for industrial sampling points	• Not recommended	• Significant gaps related to the evaluation of "representative area" of sampling points	• Significant gaps related to the evaluation of "representative area" of sampling points
Tier 2	Proxy Information	• Only for (urban) background sampling points • Not for traffic sampling points • Unclear for industrial sampling points	• Not recommended	• Screening methods for sampling point classification • Clustering methodology - Use of dendrograms to identify redundancies and outliers	• Screening method for sampling classification • Clustering methodology - Use of dendrograms to identify redundancies and outliers
	Sampling campaigns	• For all sampling points, if the campaign is well-designed and contains enough sampling locations	• Can be used for number of people exposed to exceedances, if the campaign is well-designed and contains enough sampling locations • Unclear for other indicators due to an absence of available methods.	• Can be effective to support screening methods depending on design of the campaign	• Can be effective to support screening methods depending on design of the campaign
Tier 3	Geographically explicit, comprehensive fit-for-purpose modelling	• For all sampling points, if the model is fit-for-purpose	• For all indicators, but sensitive to methodologies and model errors	• Hierarchical clustering - SR clusters can be used to identify network redundancies and gaps • (-) Data demanding (hourly data in high resolution) • (+) Can support spatial representativeness analysis for purposes beyond monitoring design	• Clustering methodology - provides additional evaluation of temporal variability • (+) Use of dendrograms to QA/QC model performance
Tier 4	Modelling complemented with dedicated measurements	• For all sampling points, if the methodology is fit-for-purpose	• For all indicators, but sensitive to methodologies and model errors	• Can be useful when combined methodology is fit-for-purpose	• Can be useful when combined methodology is fit-for-purpose

Table 12: Summary of the suitability of the proposed Tiered SR methodologies per assessments purpose. In green, recommended approaches, in orange, approaches with identified caveats, in red, not recommended approaches for specific assessment purpose. More details are provided in the main text.

7 Bibliography

Berkowicz R (2000a) OSPM - A Parameterised Street Pollution Model. Environ Monit Assess 65:323–331 . doi: 10.1023/A:1006448321977

Berkowicz R (2000b) OSPM - A Parameterised Street Pollution Model. Environ Monit Assess 65:323–331 . doi: 10.1023/A:1006448321977

Berkowicz R, Hertel O, Larsen S, Sørensen N, Nielsen M (1997) Modelling traffic pollution in streets. Natl Environ Res Institute, Roskilde, Denmark 10129:20

Copernicus Climate Change C (2017) ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS).

De Craemer S, Vercauteren J, Fierens F, Lefebvre W, Meysman F (2019) Using large-scale NO2 data from citizen science for air quality compliance and policy support. EarthArXiv

Gallego FJ (2010) A population density grid of the European Union. Popul Environ 31:460–473 . doi: 10.1007/s11111-010-0108-y

Janssen S, Dumont G, Fierens F, Mensink C (2008) Spatial interpolation of air pollution measurements using CORINE land cover data. Atmos Environ 42:4884–4903 . doi: 10.1016/j.atmosenv.2008.02.043

Janssen S, Tarrason L, Green J (2020), Record of meetings with expert stakeholders - Assessing the spatial representativeness of air quality sampling points, Service Request 5 under Framework Contract ENV.C.3/FRA/2017/0012

Kracht O, Santiago JL, 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 - Outcomes of the FAIRMODE/AQUILA intercomparison exercise

Lefebvre W, Van Poppel M, Maiheu B, Janssen S, Dons E (2013a) Evaluation of the RIO-IFDM-street canyon model chain. Atmos Environ 77:325–337 . doi: 10.1016/j.atmosenv.2013.05.026

Lefebvre W, Van Poppel M, Maiheu B, Janssen S, Dons E (2013b) Evaluation of the RIO-IFDM-street canyon model chain. Atmos Environ 77:325–337 . doi: 10.1016/j.atmosenv.2013.05.026

Lefebvre W, Vercauteren J, Schrooten L, Janssen S, Degraeuwe B, Maenhaut W, de Vlieger I, Vankerkom J, Cosemans G, Mensink C, Veldeman N, Deutsch F, Van Looy S, Peelaerts W, Lefebre F (2011) Validation of the MIMOSA-AURORA-IFDM model chain for policy support: Modeling concentrations of elemental carbon in Flanders. Atmos Environ 45:6705–6713 . doi: 10.1016/j.atmosenv.2011.08.033

Maiheu Bino, Wouter Lefebvre, Heather Walton, David Dajnak, Stijn Janssen, Martin Williams, Lisa Blyth, Sean Beevers (2017), Improved Methodologies for NO2 Exposure Assessment in the EU, Study accomplished under the authority of the European Commission, DG-Environment under service contract 070201/2015/SER/717473/C.3

Maiheu B, Janssen S (2019) Literature Review - Assessing the spatial representativeness of air quality sampling points, Service Request 5 under Framework Contract ENV.C.3/FRA/2017/0012

Meysman F, De Craemer S, Lefebvre W, Vercauteren J, Sluydts V, Dons E, Hooyberghs H, Van den Bossche J, Trimpeneers E, Fierens F, Huyse H (2020) Citizen science reveals the spatial structure of NO2 traffic-related air pollution

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

Solazzo, E. and S. Galmarini. (2015), “Comparing apples with apples: Using spatially distributed time series of monitoring data for model evaluation”. Atmos. Environ., Vol. 112, pp. 234-245.

Spangl W, Schneider J, Moosmann L, Nagl C, Booker J, Borowiak A, Bouallala S, Chow JC (2007) Representativeness and classification of air quality monitoring stations

T: +44 (0) 1235 753000

E: enquiry@ricardo.com

W: ee.ricardo.com