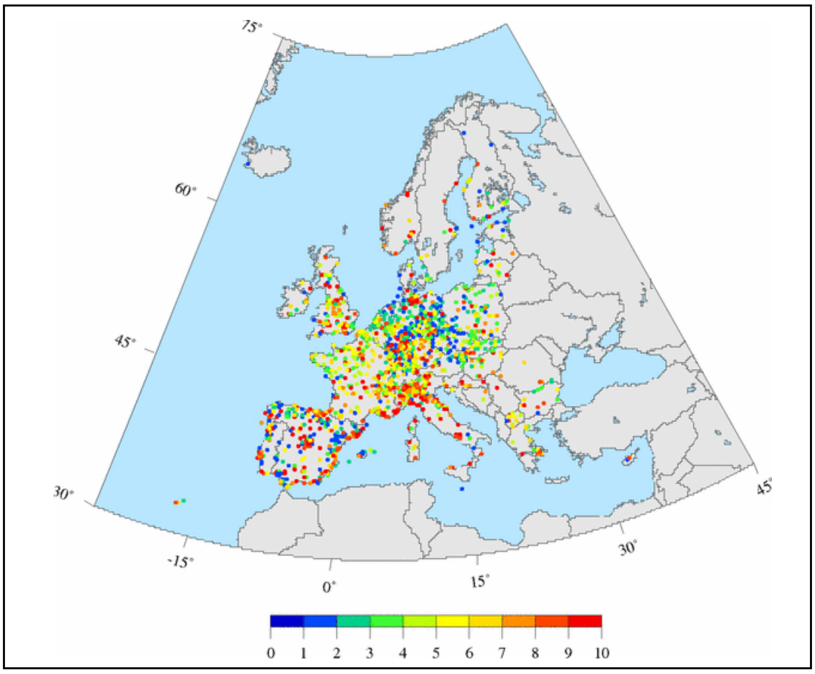
Analysis of station classification



ETC/ACM Technical Paper 2012/17
May 2013

Laure Malherbe, Anthony Ung, Philipp Schneider, Frank de Leeuw

European Topic Centre on Air Pollution and Climate Change Mitigation

The European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM)
is a consortium of European institutes under contract of the European Environment Agency
RIVM UBA-V ÖKO AEAT EMISIA CHMI NILU VITO INERIS 4Sfera PBL CSIC

Front page picture:
This picture represents NO_2 monitoring stations in Europe and the classes that were assigned to them according to the temporal variability of the measured data. The classes go from 1 (blue) to 10 (red) and are specific to each pollutant. Stations displaying low classes (1 to 3) are characteristic of rural background behaviour whereas stations showing high classes (8 to 10) are typical of urban traffic behaviour. As a supplement to AirBase usual classification, this quantitative classification allows pollutant-specific comparison between stations across Europe and can be used both in large scale modelling and mapping studies and for local applications.

Author affiliation:
Laure Malherbe, Anthony Ung: National Institute for industrial Environment and Risks (INERIS), France
Philipp Schneider: Norwegian Institute for Air Research (NILU), Norway
Frank de Leeuw: National Institute for Public Health and the Environment (RIVM), The Netherlands

DISCLAIMER

This ETC/ACM Technical Paper has not been subjected to European Environment Agency (EEA) member country review. It does not represent the formal views of the EEA.

ETC/ACM Technical Paper 2012/17
European Topic Centre on Air Pollution and Climate Change Mitigation
PO Box 1
3720 BA Bilthoven
The Netherlands
Phone +31 30 2748562
Fax +31 30 2744433
Email etcacm@rivm.nl
Website http://acm.eionet.europa.eu/

Analysis of station classification

Background information provided by ETC/ACM
under IP 2012 Task 1.0.1.2

Prepared by:
Laure Malherbe, Anthony Ung (INERIS)
Philipp Schneider (NILU)
Frank de Leeuw (RIVM)

Reviewed by:
Laurence Rouïl

Organisation:
ETC/ACM - EEA

EEA project manager:
Alberto González Ortiz

Contents

Summary

Site classification is a prerequisite for any consistent interpretation and use of air quality assessment data in a wide range of applications. It is closely linked to the question of spatial representativeness, which defines to what extent a measurement is meaningful and useful in a spatial context. There are different ways to classify monitoring sites. The current classification scheme implemented in AirBase is based on two indicators on different scales: the type of area where monitoring stations are located (rural, suburban, urban) and the type of station (background, traffic, industrial) in relation with the emission sources dominating pollution levels. This combination of indicators is interesting for many types of assessments, both in a regulatory framework and in scientific projects. However, one limitation is that some categories, in particular the urban background and suburban background ones, cover a broad range of situations which can be differently appraised depending on the expert estimate. A recent JRC-AQUILA position paper therefore recommends that additional classification schemes be refined or developed to complement AirBase classification and enlarge the amount of available information.

In the framework of the European GMES/MACC program (http://gmes-atmospher.eu), a methodology has been worked out by Météo-France (Joly and Peuch, 2012) to produce a supplementary station classification that could be particularly helpful to modellers who develop data assimilation systems. Stations are classified for each pollutant independently. The classification is based on eight quantitative indicators describing the temporal variability of concentrations. A historical dataset (multiannual time series from AirBase) is required as input. The statistical algorithm also uses AirBase information on station classification (type of area, type of station) as prior knowledge. The output is a pollutant-specific discrete classification going from class 1 to class 10. It is considered as “objective” since it rests on numerical criteria that are uniformly applied over Europe.

The aim of this study was to implement Joly and Peuch (2012) methodology on AirBase stations and analyse the resulting classes in relation with AirBase usual classification and with help of auxiliary variables (population, land cover). This allows better understanding and interpretation of this new classification. Special focus was put on the cities involved in 2012 in the Air Implementation Pilot project, a DG ENV-EEA joint project which gathers several cities across Europe to gain a deeper understanding of cities´ strengths and needs related to the implementation of EU air quality legislation.

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. According to preliminary sensitivity tests, the classification shows good robustness with respect to the training dataset. This robustness appears to be strengthened by the availability of

long multiannual time series. In the following parts of the study, air quality data were taken from AirBase v6 (updated with PM_{10} hourly French data). They consist of hourly time series of NO_2, PM_{10} and O_3 concentrations measured over the period 2002-2010. In agreement with Joly and Peuch (2012) methodology, only the stations qualified as “rural background”, “urban background”, “suburban traffic” and “urban traffic” in AirBase were retained to define the reference Fisher axis for each pollutant. Using those axes, NO_2, PM_{10} and O_3 classes were subsequently assigned to all stations included in AirBase whatever their characteristics (“rural”, “suburban” or “urban”; “background”, ”traffic” or “industrial”), provided the time series fulfilled the necessary conditions to calculate the eight indicators.

An analysis on the European scale shows that the methodology succeeds in separating stations characterized by a background behaviour (classes 1 to 3) from stations showing a more traffic-related behaviour (classes 8 to 10). Most stations falling into low classes are characterized as “rural background” in AirBase metadata. Most stations falling into high classes are characterized as “urban traffic” or “suburban traffic” stations in AirBase. However there is no strict correspondence between those categories and outliers could be identified and plotted on maps. Urban background stations appear to be spread throughout all classes; classes 5 to 7 are the most represented ones. Whereas for NO_2, classes 9 and 10 are mainly or almost entirely composed of traffic sites, for PM_{10} and O_3, they include a noticeable proportion of urban background, suburban background and industrial stations. A range of classes cannot be associated to industrial stations, as these stations are distributed throughout all classes, being classified at times in the same classes as rural and suburban background stations and at others in the same way as traffic stations.

Auxiliary data such as land cover and population density have been considered as well to support this analysis. On average, a class increase from 1 to 5 is associated with an increase of urban density (higher population density, higher fraction of urban fabric) therefore of emission sources. From class 5 to 10 the relationship between both variables is less clear, displaying a different profile according to the pollutant. Additional databases like fine resolution emission inventories could be helpful to bring this analysis forward on a smaller scale.

Differences between NO_2, PM_{10} and O_3 classifications were then investigated. For stations measuring both pollutants, good agreement is observed between NO_2 and PM_{10} classes. The difference is less than 3 points for 90% of the stations. Higher differences, mainly negative (i.e. NO_2 class lower than O_3 class), are observed for a significant number of stations mostly located in coastal and mountainous areas, namely in regions where the transport and chemistry of ozone are influenced by sea-breeze and topography effects. Discrepancies between NO_2 and O_3 classification might also be related to differences in the training datasets used to compute the Fisher axis. Whereas the proportion of background stations is about 60% for NO_2 and PM_{10}, for O_3 it reaches approximately 75%.

Setting focus on the Air Implementation Pilot cities gives a better insight into Joly and Peuch (2012) classification on the urban scale. The analysis of the new classes in relation to AirBase metadata brings out different types of cities, in particular:

- those for which the new classes match AirBase classification quite well with a clear distinction between suburban background, urban background and urban traffic sites (e.g. Berlin);
- those with high urban density and an overall strong influence of traffic: all background and traffic stations are classified between 8 and 10 (e.g. Madrid);
- those for which the "traffic behaviour" of monitored data is less pronounced but with no marked difference between urban background and urban traffic sites (e.g. Dublin, Milan);
- those characterized by low population density and a behaviour typical for background even if differences are observed between urban background and urban traffic sites (e.g. Malmö).

Land cover (fraction of discontinuous and continuous urban fabric) and population density are useful data to interpret the results but they are not sufficient. For a better understanding of the classification at the city level, other variables and parameters should be considered such as high resolution emission inventories, topography (city or street configuration), aerial views,...

On the whole, the classification into ten classes appears to be an added value for qualifying a monitoring station. It offers the possibility of making an objective, quantitative and pollutant specific comparison of the stations, based on air quality data themselves, in that way complementing the Airbase classification. In addition, it can be regularly updated, thus providing information about possible evolutions of the station characteristics. A practical application of this classification in the framework of GMES/MACC2 project was to redefine the set of stations intended for model validation and data assimilation. The ten classes will be used to discard stations that are too specific with respect to the considered models resolution or to gather stations in homogeneous sets according to statistical properties. This complementary classification can also be a beneficial help to a wider community of data providers and users both for scientific projects, reporting obligations and policy-oriented studies on different spatial scales. Experience feedback will be useful to better specify the application framework of this classification in future.

1. Introduction

Station classification is a way to characterize and label a measurement with respect to the station environment, the surrounding emission sources or any other indicator of the main variables influencing concentrations. It represents fundamental information to carry out air quality assessments, allow consistent interpretation of measured data and ensure comparability of observations between different locations and with model results (which are themselves characterized by spatial resolution). Station classification is closely related to the question of spatial representativeness, which defines to what extent a measurement is meaningful and useful in a spatial context (Janssen et al., 2012; Umweltbundesamt, 2007). It is very helpful to modellers who are concerned with using the most appropriate in situ data for the purpose of model evaluation, data assimilation and mapping, and to the health community which needs relevant monitoring data to conduct exposure assessment or epidemiological studies.

There are different ways of classifying monitoring sites which have their strengths and shortcomings depending on the context. A recent JRC-AQUILA position paper on siting criteria, classification and representativeness of air quality monitoring stations includes considerations on that issue[1]. As mentioned in this paper, the current classification scheme used in AirBase is based on two indicators on different scales: the type of area where monitoring stations are located (rural, suburban, urban) and the type of station according to which type of sources dominates the air pollution levels (background, traffic, industrial). The resulting categories combine both indicators as described in the guidance on Exchange of Information (EoI, 2002). The same indicators, referred to as “classification of the area” and “classification of station in relation to predominant emission sources” have been retained in the recent Commission Implementing Decision on EoI and reporting (2011/50/EU, “IPR Decision”) which shall apply from 1.1.2014. This scheme is advantageous given that it is interesting both for modelling and for policy-oriented studies and as such, it has been used in a broad number of applications on the European level and by Member States. However, an issue is that some categories, such as the “urban background” category, cover a wide range of situations, denoting different realities with the same name, and also depend on the expert estimate. Therefore no comparability throughout the Europe can be guaranteed. The JRC-AQUILA paper therefore recommends that additional classification schemes be refined or developed, for example, based on primary air quality data in combination with meta- information describing the station surroundings (e.g. land cover types, population density, emission density within a certain radius, climatological or meteorological data, etc.).

In the framework of the European GMES/MACC program (http://gmes-atmospher.eu), a methodology has been worked out by Météo-France (Joly & Peuch, 2012) to produce a supplementary station classification

[1] Draft version of 15.12.2012 submitted to the AQUILA members

that could be particularly useful to modellers who develop data assimilation systems. The result is an objective and pollutant-specific classification based on the historical concentration data themselves with AirBase classes as a prior knowledge. This new classification rests on numerical criteria that are uniformly applied over Europe, ensuring good comparability between sites. Moreover, it is specific to each pollutant, reflecting the fact that emission sources may have different influences on ambient air concentrations according to the pollutants. It therefore appears as a meaningful complement to the classification currently used in AirBase. It should also help to characterize monitoring sites per pollutant as required by the IPR Decision and to improve the precision of the metadata reported to the European Commission and the EEA. A proper use of this classification still requires some guidance since the new classes are defined as numbers (1 to 10) and have no explicit concrete name like the current AirBase categories.

The purpose of this study is to apply Joly and Peuch (2012) methodology to AirBase stations and analyse the resulting classes with special focus on the cities involved in 2012 in the Air Implementation Pilot project, a DG ENV-EEA joint project which gathers several cities across Europe to gain a better understanding of cities´ strengths and needs related to the implementation of EU air quality legislation. This work is not intended to replace AirBase current categories with new ones but to explain how this new classification can help EEA, data providers and data users to make a better assessment of monitoring sites and investigate the consistency of existing station description. The report is divided in five main parts: part 2 gives a description of the methodology; part 3 shows the robustness of the approach through sensitivity tests; part 4 provides a European-wide analysis of the classification whereas part 5 focuses on the Air Implementation Pilot cities; the different possible uses of the new classes are developed in part 6. Conclusions and remaining issues are exposed in the last section.

2. Description of the method

This section provides the outline of the methodology. More details can be found in Joly and Peuch (2012).

2.1 Principle

The main consideration is that the temporal behaviour of observed concentrations reflects the emission sources and atmospheric processes influencing air quality at a monitoring site and that it therefore makes a good indicator of the measurement representativeness.

The idea is then to classify monitoring sites for each pollutant specifically according to the temporal variability of measured concentrations.

To describe this variability, Joly and Peuch (2012) identified eight indicators related to the various time scale components of a time series:

- Four refer to the diurnal cycle:
  - the daily maximum of the diurnal cycle,
  - the summer/winter ratio of the diurnal cycle daily maxima,
  - the amplitude of the diurnal cycle,
  - the summer/winter ratio of the diurnal cycle amplitude.
- Three refer to the weekly cycle:
  - the weekend effect on the daily mean
  - the weekend effect on the daily maximum
  - the weekend effect on the standard deviation
- The last one characterizes residual variability after the diurnal cycle and the lowest frequencies (presumably related to the meteorological large-scale synoptic conditions) are filtered out.

The classification consists in combining those eight indicators in a linear function that best separates rural background sites from sites influenced by urban and local sources. For each pollutant separately, a linear discriminant analysis (LDA) is applied using AirBase metadata as prior knowledge. The resulting linear function, the Fisher's linear discriminant, defines a direction, the Fisher axis on which any time series summarized by the eight indicators can be projected. A class number between 1 ("very rural") and 10 ("highly influenced") is then attributed depending on the coordinate of the projection. The thresholds delimitating the 10 classes are provided by the 10th to 90th percentiles of the projection coordinates over all stations involved in the axis construction.

2.2 Implementation

For the observation data available in the AirBase database, the methodology has so far been implemented for the following pollutants: NO_2, O_3, SO_2, PM_{10}. For each pollutant, the classification process includes two main stages:

1. the construction of the Fisher axis and the calculation of the bounds delimitating the 10 classes.

   This step requires a training data set where each station is described by the eight indicators and flagged by its metadata as prior knowledge.
   In Joly and Peuch (2012) the training data set is made of rural background sites (labelled "R") and urban background sites together with suburban and urban traffic stations (labelled "U+T"). It is recommended to discard other categories at this stage since they may not facilitate the discrimination between rural and polluted sites.

2. the projection on the Fisher axis to determinate the class number.
   This step can be applied to any time series: all those included in the training data set and any additional time series for which the eight

indicators can be computed. All categories of stations can be considered in this projection.

2.3 Requirements and limitations

At least 8760 hourly values (the equivalent of a full year) are necessary to classify a station. Considering unavoidable data losses, it means that more than one year of data is required for any new (and sometimes on-going) station. Wherever it is possible, the training period should be long enough (several years) to get more robust indicators and provide a more reliable classification. In the same time, it should be short enough to avoid possible drifts and significant changes in the temporal variability.

Some parameters are also set up such as the length of sliding windows to calculate the diurnal cycle or the minimum number of non missing data to determine the diurnal cycle amplitude and the weekend effect.

It should be noted that the methodology has been developed for time series of hourly data. This is the reason why a significant number of monitoring sites could not yet be classified for PM_{10} (only daily measurement values are available in AirBase). To classify those stations, the indicators need to be adapted to daily data.

3. Sensitivity tests and robustness

3.1 Analysis of the classification robustness

The new classification is driven by the hourly time series available in AirBase database. This database is updated on an annual basis to include last validated measurements whereas a station classification should be stable to be useful for users. It is then important to check if the classification is altered or not with new versions of the database. Furthermore the number of available time series in AirBase may evolve for various reasons depending on Member States and because of changes in the monitoring networks. It is then necessary to identify whether the classification is sensitive to an addition or removal of a set of stations.

The robustness test is carried out comparing the results from four experiments based on different sets of data:

- Experiment 1 is a classification obtained with AirBase v5[2] and a training period from 2002 to 2009.

[2] AirBase v5 (http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-3) includes air quality data from 1969 to 2009. AirBase v6 (http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-6) is an update of AirBase v5; it includes data from 1969 to 2010. For experiment 1 and experiment 2, time series of hourly

- Experiment 2 is a classification obtained with AirBase v6 and a training period from 2002 to 2009.
- Experiment 3 is a classification obtained with AirBase v6 and a training period from 2002 to 2010.
- Experiment 4 is a classification obtained with AirBase v6 completed with French database for PM_{10}[3] and a training period from 2002 to 2010.

Comparisons are performed for each pollutant using two types of plots as illustrated hereafter.

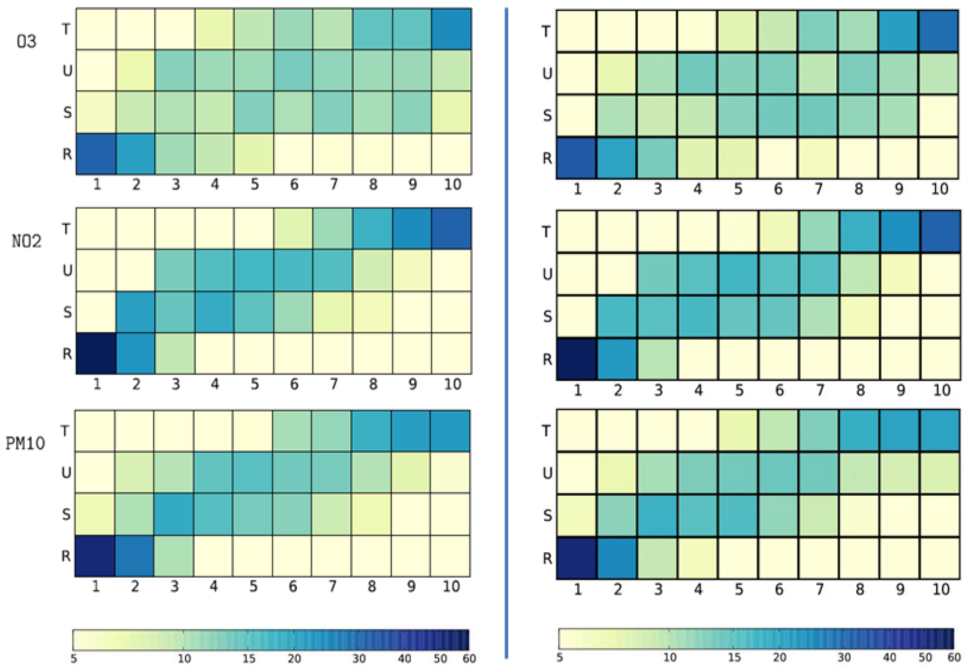


Figure 1: Colour diagrams of class distribution in relation to AirBase metadata for experiment 1 (from Joly and Peuch 2012, left) and experiment 4 (right). From top to bottom: graphics for O_3, NO_2, then PM_{10}.

Colour diagrams such as those displayed on Figure 1 are helpful to visualize the distribution of classes in connection with the current AirBase classification. A colour diagram is provided for each pollutant and shall be read as follows:

- Each line corresponds to a group of stations according to AirBase metadata (R: rural background, S: suburban background, U: urban background, T: urban and suburban traffic).

---

concentrations of NO_2, PM_{10} and O_3 were extracted from AirBase v5 and AirBase v6 respectively over the same period (2002-2009).

[3] Only PM_{10} daily data have to be reported in AirBase; the transmission of hourly data (where they are available) rests on Member States. France only reports daily data; hourly data were thus taken from the French national air quality database.

- On each line, namely for each group, the colour of the first, second, ..., tenth cell represents the proportion of stations from that group that fall in Joly and Peuch (2012) class 1, 2, ..., 10.

For all the experiments, the colour scale has been kept identical. There is little change in the cell colours from an experiment to another, indicating that the frequency distribution of Joly and Peuch (2012) classification is stable.

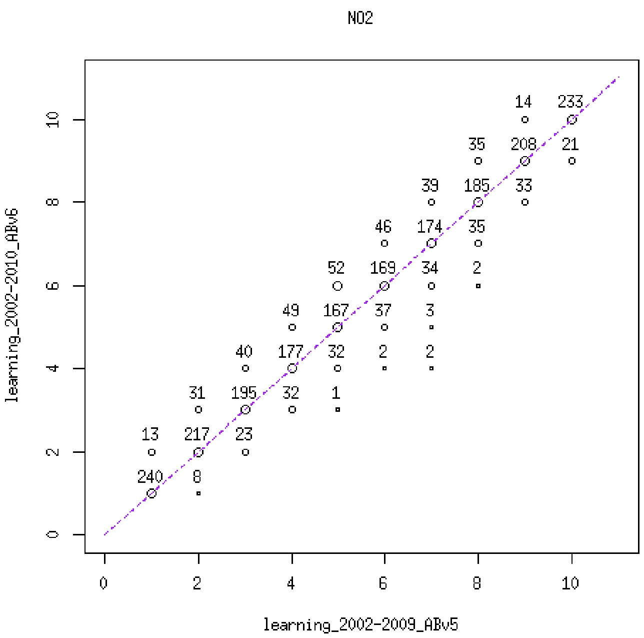


Figure 2: Scatter plot between Joly and Peuch (2012) classes obtained for NO_2 in experiment 1 (x-axis) and experiment 4 (y-axis). Dotted line: bisector.

Scatter plot in Figure 2 highlights differences between classes from two different training data sets. A non-zero distance from the bisector indicates a change in class between both experiments. The number above each point denotes the number of concerned stations. Among 2549 stations measuring NO_2 and common to experiment 1 and experiment 4, 1965 stations keep the same class, 574 stations display a change of 1 class, 8 stations a change of 2 classes, and 2 stations a change of 3 classes. For those last ten stations, less than five years of data are available or the number of data per year is irregular, suggesting that a minimum five-year measurement period and a high data capture make favourable conditions for a robust classification.

For PM_{10} and O_3, the training period has larger influence (certainly because of the dependence on the meteorological variability), implying that for secondary pollutants, the classification may be more sensitive to the considered years. However, the percentage of stations with a change of 3 classes or more is low (less than 2% as shown in Table 1), pointing out an overall good robustness of the classification whatever the pollutant.

Table 1: Statistics on the class differences between experiment 1 and experiment 4

Pollutant	Total number of stations	Perc. of stations with a diff. of 0 class	Perc. of stations with a diff. of 1 class	Perc. of stations with a diff. of 2 classes	Perc. of stations with a diff. of 3 classes	Perc. of stations with a diff. of 4 classes	Perc. of stations with a diff. of 5 classes
NO_2	2549	77.1 %	22.5 %	0.31 %	0.08 %	0 %	0 %
PM_{10}	1342	68.9 %	28.0 %	2.5 %	0.37 %	0.15 %	0 %
O_3	1970	54.9 %	38.3 %	5.5 %	1.1 %	0.15 %	0.10 %

Similar comparisons have been carried out with the results of the four experiments. Most differences of classes between experiments (± 1) are likely due to a slight modification of the Fisher axis and a change in the number of stations used to define the ten classes. As a consequence, it can occur that a station that is projected below the upper bound of e.g. class 4 in one experiment falls above the lower bound of class 5 in another one. Less numerous, larger differences may be the result of significant changes in the indicators, especially for time series including few or incomplete years of measurements: in those cases, the indicators are more sensitive to data updates.

In conclusion, **the classification method can be considered as robust for all pollutants, showing limited sensitivity to the data set and the training period.** Adding new stations or a new period of measurements may shift the classes but do not change drastically the classification.

3.2 Implications for an operational use of the classification

To be able to calculate comparable classes over time and provide meaningful updates of the classification, **the proposed procedure is to define a reference Fisher axis and set fixed class thresholds which will not change with the yearly updating of AirBase.** This axis could be constructed over a 10-year reference period with a selection of stations showing good data quality (well-defined metadata, several years of measurement, high data coverage) and distributed all over Europe.

The reference axis and the fixed thresholds will be used:

- to classify the remaining sites (by projecting the corresponding time series);
- to provide regular updates of the classification (by projecting new or updated time series).

Given the robustness of the classification, as highlighted in section 3.1, and using the most complete and up-to-date available dataset, the Fisher axes obtained in experiment 3 (or 4, as it is the same for these two pollutants) for NO_2 and O_3 and in experiment 4 for PM_{10} are taken as a reference for the present study.

4. First results and analyses on the European Scale

4.1 Purpose

The purpose of this chapter is to provide a European-wide overview of the new complementary classes and give a concrete meaning to the class numbers. For each pollutant (NO_2, PM_{10} and O_3), two types of analysis are performed (subsections 4.2 to 4.4):

- The first one focuses on the statistical distribution of Joly and Peuch (2012) classes in relation to AirBase metadata and other variables indicative of pollution sources. The aim is to help operators to link class numbers (1 to 10) with common parameters describing the station environment.
- The second one is a spatial analysis intended to identify whether station classes are distributed over Europe according to some geographical patterns.

The last subsection (4.5) investigates the dissimilarities between the classifications associated to the three pollutants.

Joly and Peuch (2012) methodology was applied using the following training data sets:

- for NO_2 and O_3, AirBase v6, period 2002-2010;
- for PM_{10}, AirBase v6 updated with French hourly time series, period 2002-2010.

The results are analysed with regard to population and land use variables computed from available European or global databases.

NB: in the following sections, AirBase usual categories, which both account for the type of area and the type of station according to dominant pollution sources (see section 1), are called by their common names ("rural background", "urban background"...), whereas the new classes, which are based on the temporal variability of concentrations, are often referred to using the term "behaviour".

4.2 Results for NO_2

4.2.1 Analysis of the new classes as a function of the station environment

At the European scale the new station classification for NO_2 behaves overall mostly as expected. A good overview of the behaviour of the various classes can be gained by investigating the distribution in frequency of the new NO_2 station classifications with respect to the AirBase metadata. This is shown in Figure 3. For sites labelled as rural background in AirBase, the vast majority of stations falls into classes 1 to 3. The frequency decreases sharply from class 1 to 3 and then slowly levels off until only a very few of these stations are found in the higher classes. The exact opposite behaviour can be found for traffic stations which are hardly ever classified as 1 to 3. The frequency increases slowly for moderate classes from 4 to 6 and then increases rapidly all the way to class 10, which represents by far the most number of stations. Urban background stations show nearly a Gaussian distribution with a very wide spread throughout all classes. The largest number of urban background stations can be found for class 5, whereas the number of stations then drops first slowly and then more rapidly towards both lower and higher values. Hardly any urban background stations are classified in the extreme classes 1 and 10. The patterns for suburban background and industrial stations are not as clear. Whereas suburban background stations are mostly classified in the lower classes from 2 to 6 and are not as prominent in the other classes, industrial stations generally occur in all classes but are classified the most as class 2 with a very gradual drop-off towards higher classes.

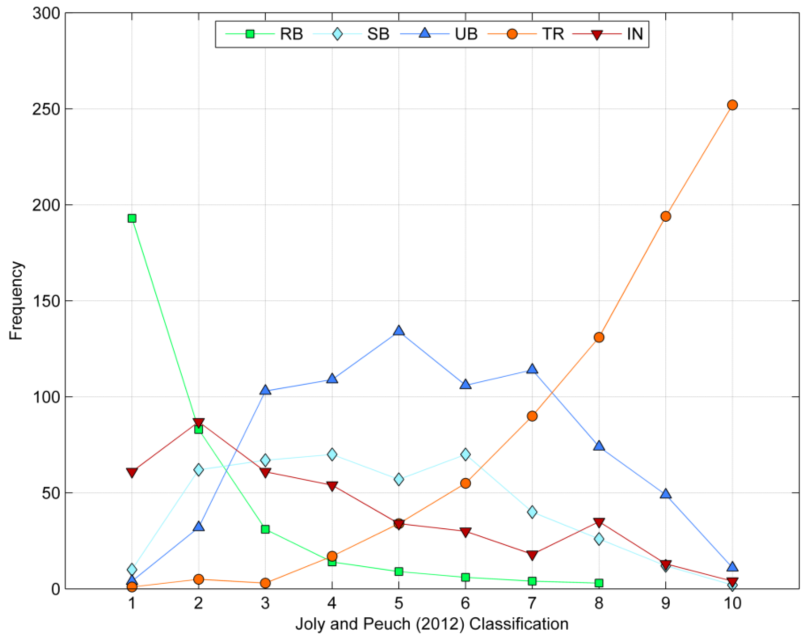


Figure 3: Distribution in frequency of the new NO_2 station classification, categorized by AirBase metadata. RB = Rural background, SB = Suburban background, UB = Urban background, TR = Traffic (all types of area), IN = Industrial (all types of area).

The box plot in Figure 4 shows the means, standard deviations and extreme values of the set of all European stations, categorized by the metadata delivered by AirBase. As would be expected, the new means are generally higher for urban stations than for suburban and rural stations. This is true for background, traffic, as well as industrials stations. However, for several AirBase categories outliers in terms of the new classification can be found. This is particularly true for stations that AirBase considers rural background and that behave more like urban background ones (classes 4 to 6). While the median class for rural background is 1 and even the 75th percentile does not exceed class 2, in several cases the Joly and Peuch (2012) classification algorithm assigns classes of 4 or higher. This behaviour is somewhat curious and needs to be investigated further. For this purpose, the station codes of all the rural background stations that were classified as class 4 or higher are listed in Annex I, Table 4. Similarly, the station codes of the all the urban traffic stations that were classified as class 4 or lower are listed in Annex I, Table 5. Possible reasons for these outliers are also discussed later on in this document and will be further investigated with the data providers.

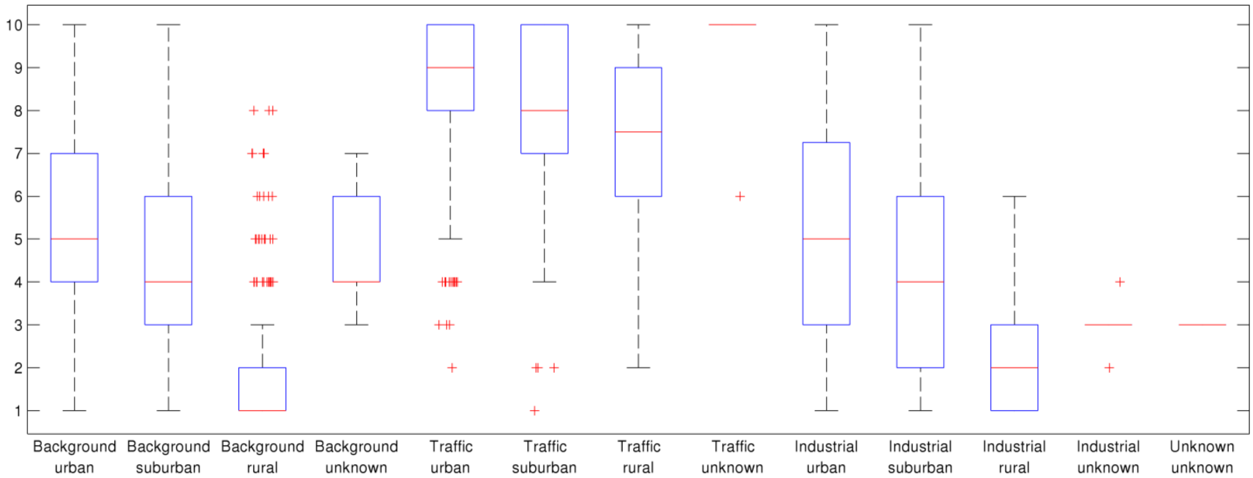


Figure 4: Box plot showing the median (red line), 25th and 75th percentiles (top and bottom of blue rectangle), the extreme values excluding outliers (black whiskers), as well as outliers (red crosses) of the new classification system for NO_2, categorized by AIRBASE station metadata

4.2.2 Dependence on population density

The typical background concentrations of NO_2 are generally highly dependent on the population density in the vicinity of the station. For this reason, an investigation of the relationship between the new station classification and population density was carried out. The population density was derived from the Global Rural-Urban Mapping Project version 1 (GRUMPv1[4]) which builds upon the Gridded Population of the World (GPW) version 3 dataset. The GRUMPv1 dataset provides global gridded data on population density for the year 2000 at a spatial resolution of 30 arc-seconds. For each station stored in AirBase, the corresponding latitude and longitude were extracted and the value of the closest grid cell in the GRUMPv1 dataset was obtained. The resulting dataset was then studied with respect to the dependence of the new classification categories on population density and possible outliers. Consolidating the data by the ten new classes, Figure 5 shows a boxplot of the data highlighting the median, 25th and 75th percentile, the extreme values not considered outliers, and the actual outliers, for each of the ten new classes. As would be expected, classes 1 and 2 are generally associated with very low population densities. From there on, the median population increases with the class

[4] For more information, please visit: http://sedac.ciesin.columbia.edu/data/collection/grump-v1

number, however this increase is not entirely homogeneous (see Figure 6 for more details). Due to the non-Gaussian distribution for each class, quite a few outliers appear for all classes. Whereas the range of outliers for the "mostly rural" classes 1 and 2 is fairly small and generally is less than 4000 persons per square kilometre, the higher classes are associated with significant scatter in the respective outliers. While this would be expected for some of the higher classes such as 8 through 10, it is surprising to also see outlier stations with population densities beyond 15000 persons per square meter for moderately classified stations such as for classes 6 and 7. These are mostly urban background stations in areas of very high population density such as the centre of Paris.

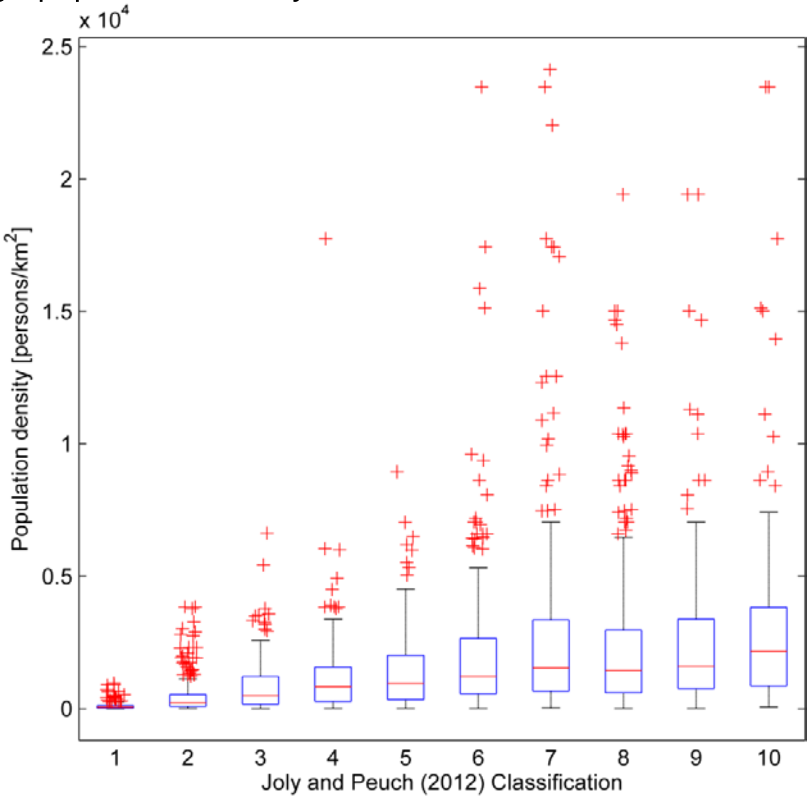


Figure 5: The relationship between the categories of the new NO_2 station classification and population density at the station. This boxplot shows the median (red line), 25th and 75th percentiles (upper and lower limits of the blue box), extreme values excluding outliers (black whiskers), and outliers (red crosses) of population density, categorized by Joly and Peuch (2012) classification.

Due to the scale of the y-axis the change in median values cannot be observed very well in Figure 5. For this reason, Figure 6 shows both the median as well as the mean population density for each class. As would be expected, starting at class 1 both values increase until class 6. However, surprisingly, from class 7 to class 9 the values stay either approximately the same (median) or even decrease rapidly (mean). Only for class 10 is it possible to observe a significant increase again, although the final value for the mean is not much higher than for class 7. There is obviously a discontinuity for the NO_2 classification with respect to the medium- to high classes. It is not entirely clear at this point what causes this behaviour but further analysis should investigate this phenomenon in more detail at the individual station level. Perhaps class 7 behaviour is typical of dense urban areas whereas classes 8 and beyond refer to traffic and industrial influences which are not necessarily correlated with population density.

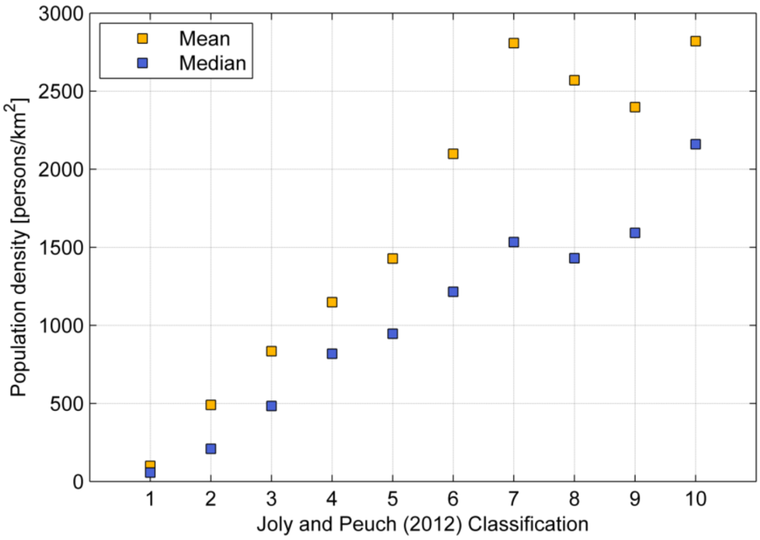


Figure 6: Similar to Figure 5 but only showing the median and mean of population density (outliers included) for each of the ten classes

4.2.3 Dependence on land cover

In order to investigate the impact of the surrounding land cover on the new NO_2 station classification, the CORINE Land Cover data set at 100 m spatial resolution was used. For each of the 2686 stations considered as part of this study, a 1 km x 1 km rectangle centred on the location of the station was extracted from the land cover raster and the fraction of each land cover class was computed for this area.

Figure 7 shows the average fraction (in percent) of the various CORINE land cover classes for each station class identified by the Joly and Peuch (2012) algorithm. It is somewhat surprising to see that for all classes besides 1 the dominating land cover type is CLC112 (Discontinuous Urban Fabric). Looking only at CLC111 (Continuous Urban Fabric) in the first column of the diagram, one can see that its fraction increases continuously from 0.5% for class 1 to 34% for class 10. Interestingly, when examining CLC 112 (Discontinuous Urban Fabric) in second column, the highest fractions are found for class 4 and 5 (both over 50%) and not for the higher classes of 8 and above. As mentioned previously, class 1 is the only class that is not dominated by urban fabric but rather by agricultural areas and forests (CLC211, CLC242, and CLC312 with fraction all over 10%).

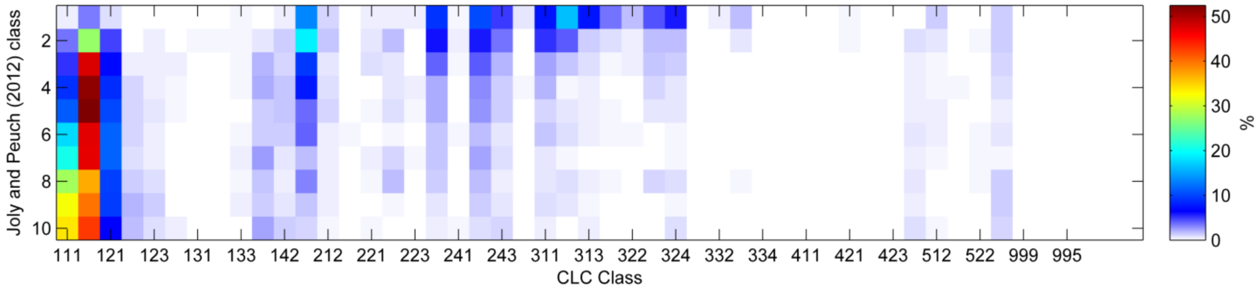


Figure 7: Average percentage of CORINE land cover classes for each of the 10 classes of the Joly and Peuch (2012) scheme that have been assigned to NO_2 monitoring stations. This was computed over all 2686 stations that were studied here. Note that only every second CLC class was labelled here for clarity.

4.2.4 Spatial analysis

A spatial analysis of the new station classification for NO_2 was carried out subsequently. Figure 8 is an overview map of NO_2 monitoring stations and the classes that were assigned to them for NO_2 by the Joly and Peuch algorithm (2012). This graphic is to be compared with Figure 9 in which stations with low classes (1 to 3) or high classes (8 to 10) are represented.

As observed in the previous analysis, low classes are associated with a "background behaviour''. They are distributed throughout most of Europe with the densest occurrence in Germany, Netherlands and the Iberian Peninsula (Figure 9, left). A vast majority of the corresponding stations were originally classified as rural background in the AirBase database. Some discrepancies occur however between both classification systems. They can be more easily seen in Figure 52 (Annex I), which shows a map of only stations which are labelled as rural background in AirBase but are classified as class 4 or higher in the Joly and Peuch (2012) system. Interestingly, most of the stations with the largest discrepancies appear to be located in France. While several of the other large European countries also show a few stations classified as 4 or 5, France has quite a few stations that are classified as 7 and higher. These stations should be investigated further individually.

As can also be seen from previous analysis, high classes of stations are associated with a "traffic behaviour" regarding NO_2 time series. Figure 9 (right graphic) also shows that those stations are distributed throughout Europe. The highest density of these stations can be found in the Po valley in Northern Italy, and Germany. However, a few outliers, here defined as urban traffic stations (from AirBase metadata) with assigned class of 4 or lower, do exist. Their spatial distribution is shown in Figure 54 (Annex I). There appears to be a tendency towards these outliers occurring more in the eastern part of Central Europe. However, in addition two outliers occur also in Spain and one in the United Kingdom, so this pattern is not very strong.

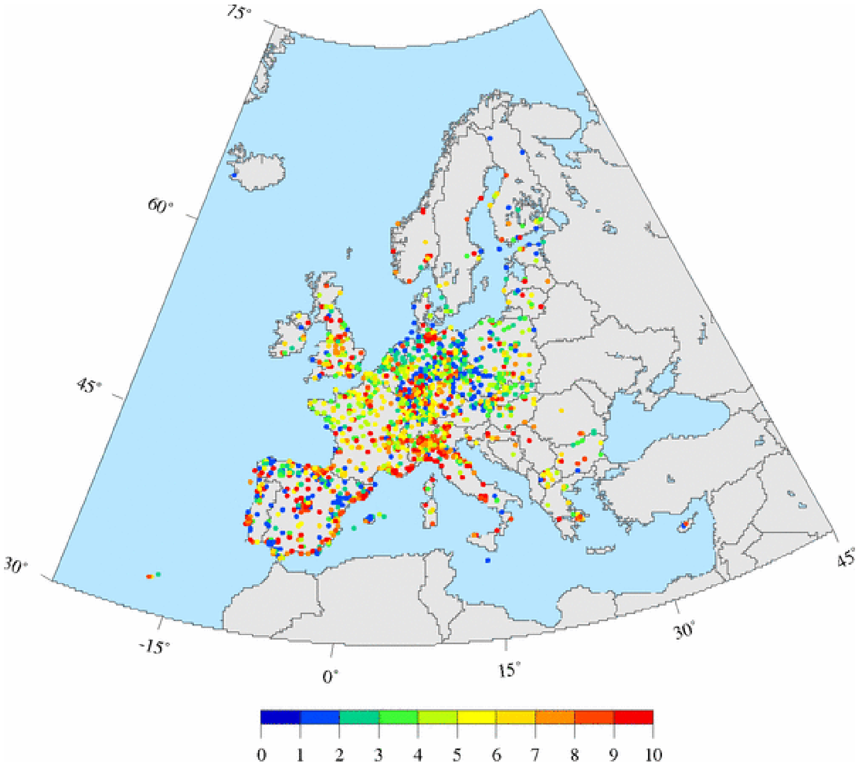


Figure 8: Overview map of NO_2 monitoring stations and the classes that were assigned to them for NO_2 by the Joly and Peuch algorithm (2012)

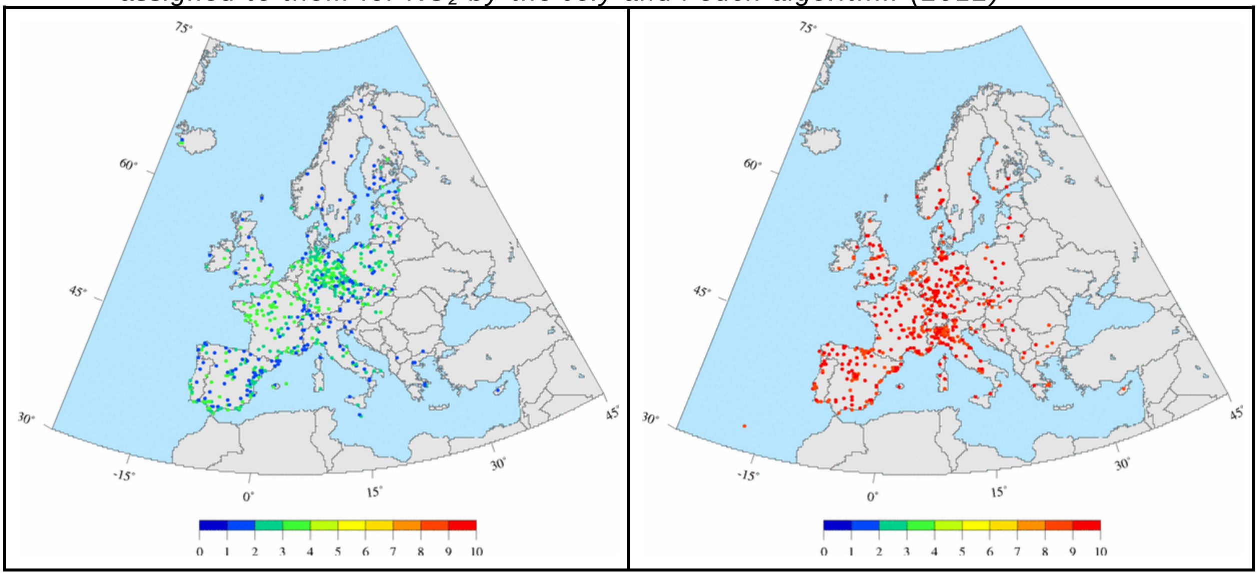


Figure 9: Overview map of NO_2 monitoring stations. Left graphic: 743 stations with low classes (1, 2, 3). Right graphic: 760 stations with high classes (8, 9, 10).

The previously identified outliers in the new NO_2 classification for rural background and urban traffic stations were further studied spatially with regards to the land cover in the vicinity of each station. Figure 52 (Annex I) shows the spatial distribution of the rural background stations that were classified as outliers, i.e. stations that were classified as 4 or higher despite being considered rural background in the AirBase metadata. More importantly, the Figure 53 also shows the fraction of the continuous urban fabric land cover class (CLC111) that occurs within a 1 km x 1 km box centred on the station. While most of the outlier stations have no continuous urban fabric in their vicinity, there are several stations that exhibit surprisingly high values. There are several stations in France that are considered as rural background despite being surrounded by 10% and 20% continuous urban fabric. Even more importantly, two stations in Spain are shown as having 40% and 50% continuous urban fabric in their vicinity, which is quite unrealistic. The new station classification following Joly and Peuch (2012) appears to be performing better as it classifies these two stations as classes 9 and 6 (see Figure 52). A more detailed investigation of the neighbourhood of these stations will be necessary to clearly determine their characteristics.

Similarly, Figure 54 and Figure 55 show the spatial distribution of outliers among the urban traffic stations, i.e. stations that were classified as urban traffic by the AirBase metadata but were assigned classes of 4 or lower, and their respective fraction of the continuous urban fabric land cover class from the CORINE dataset. It can be observed that, despite being considered urban traffic stations, several stations have no fraction of continuous urban fabric whatsoever, primarily in Germany, the Czech Republic, Northern Italy, and Hungary. The newly assigned classes for these stations (as can be seen in Figure 54) appear to fit significantly better as all of these stations are considered to be class 3 or 4 stations. Again, a more thorough investigation of the characteristics of these stations will be necessary in order to clearly determine if the AirBase metadata or the new classification results for these stations are questionable.

4.3 Results for PM_{10}

The same analysis is carried out for PM_{10}. For some monitoring sites (case of measurements performed by gravimetry), even for whole monitoring networks (when only daily average concentrations are reported, e.g. Italy, France), hourly data are not available in AirBase, so the stations could not be classified (see section 2.3). It is therefore planned to adapt the methodology to time series of daily concentration values. In this study French hourly data have been taken from the French national database.

4.3.1 Analysis of the new classes as a function of the station environment

Even if the number of PM_{10} stations is twice lower than for NO_2, the histograms show similar patterns (Figure 10 and Figure 11, to be compared to Figure 3 and Figure 4, respectively). The sharp decrease of the number of rural background stations and the sharp increase of the number of traffic stations from low to high classes illustrate the capacity of the new classification for separating stations characterized by a “rural background behaviour” from stations distinguished by a “traffic behaviour”. From Figure 11 it can be seen that part of the traffic stations falling into low classes are located in rural areas according to AirBase metadata. As for NO_2, urban background stations are spread throughout all classes, classes 5 to 7 being the most represented ones. However, unlike NO_2, a significant fraction of this type of stations falls into class 10.

Whereas for NO_2, classes 9 and 10 are mainly or almost entirely composed of traffic sites, for PM_{10}, they include a noticeable proportion of urban background, suburban background and industrial stations. A range of classes cannot be associated to industrial stations, as these stations are distributed throughout all classes, being classified at times in the same classes as rural and suburban background stations and at others in the same way as traffic stations.
In conclusion, the new classification succeeds in making a clear distinction between stations typical of a “rural background behaviour” (classes 1 to 3) and stations which have a “traffic behaviour” (classes 8 to 10) but there is no well-identified correspondence between AirBase and Joly and Peuch (2012) classifications. One main reason can be the nature of PM_{10} pollution: PM_{10} is not only a primary pollutant but also a secondary pollutant produced from precursors and affected by chemistry and transport processes which are largely driven by the meteorological conditions. Whereas AirBase classification describes the surroundings of the station and the presence of close emission sources if any, the classes derived from Joly and Peuch (2012) algorithm integrate a wider range of influences.

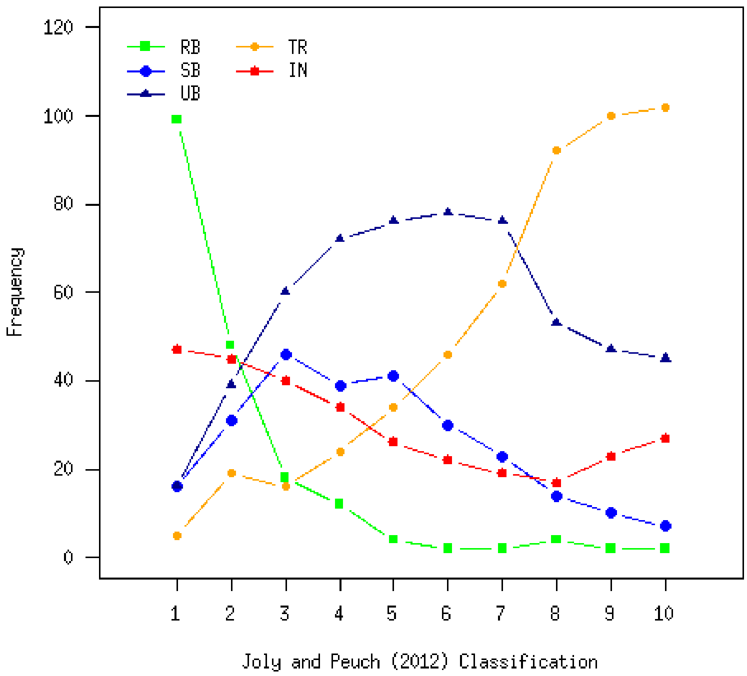


Figure 10: Distribution in frequency of the new PM_{10} station classification, categorized by AirBase metadata. RB = Rural background, SB = Suburban background, UB = Urban background, TR = Traffic (all types of area), IN = Industrial (all types of area).

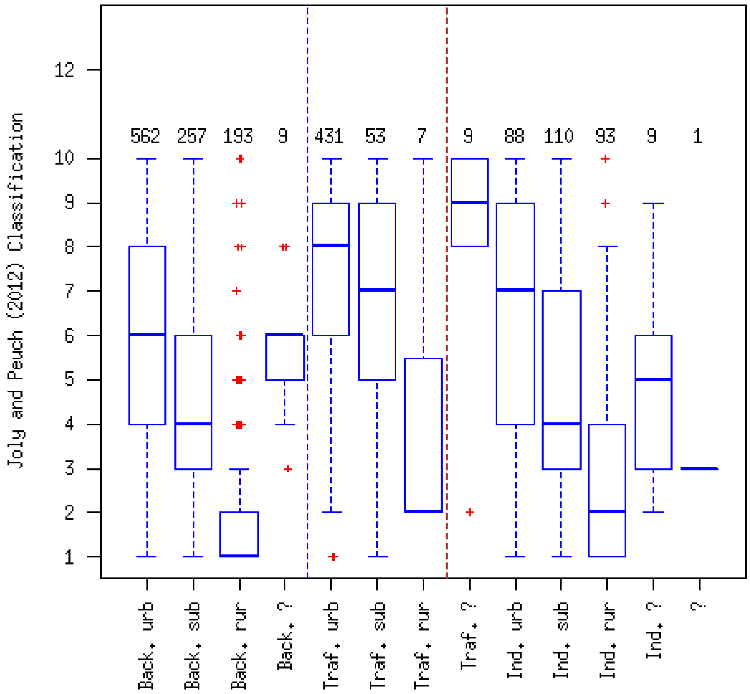


Figure 11: Box plot showing the median (blue line), 25th and 75th percentiles (top and bottom of blue rectangle), extreme values excluding outliers (blue whiskers) and outliers (red crosses) of the new classification system for PM_{10}, categorized by AirBase station metadata. The number of stations per AirBase category is indicated on the graph.

4.3.2 Dependence on population density

As for NO_2, the relationships between the new classes and population density are investigated. Figure 12 shows an increase of population density from class 1 to class 5. From class 5, no specific link between population density and the classes can be highlighted, except for an increase of the mean up to class 9.

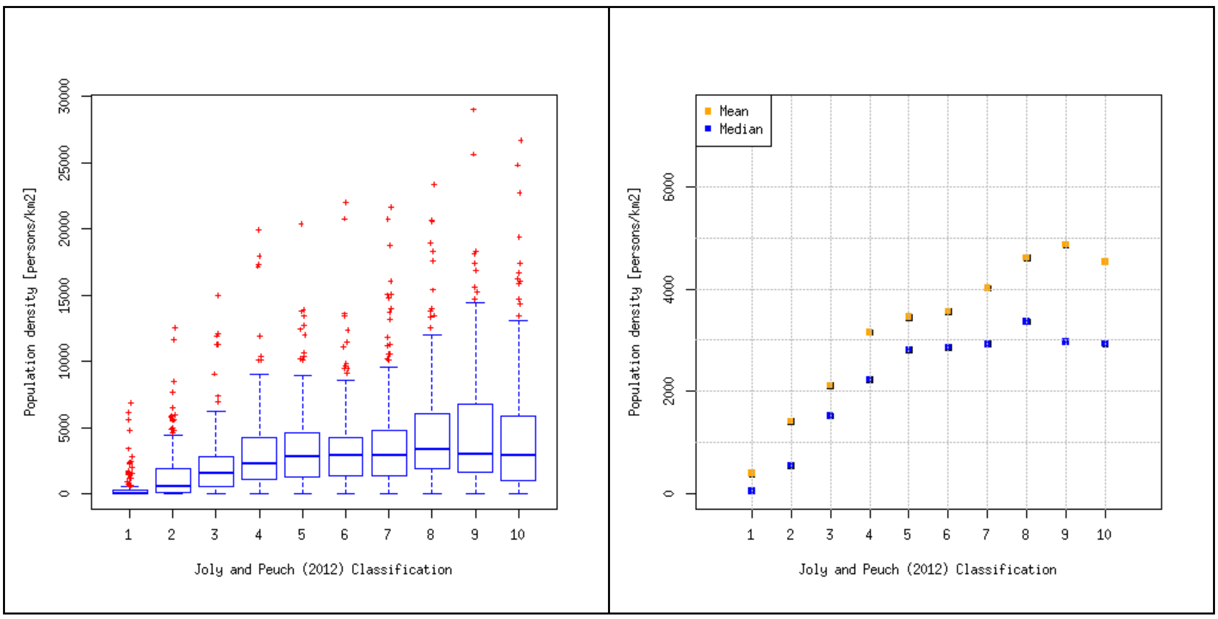


Figure 12: Left: the relationship between the categories of the new PM_{10} station classification and population density[5]. This box plot shows the median (blue line), 25^{th} and 75^{th} percentiles (upper and lower limits of the blue box), extreme values excluding outliers (blue whiskers), and outliers (red crosses) of the rate of population density, categorized by Joly and Peuch (2012) classification. Right: Same graph but only showing the median and mean of the rate of population density for each of the ten classes.

4.3.3 Dependence on land cover

A steady increase of urban fabric can be noticed from classes 1 to 4, followed by a slightly increase until classes 7 or 8 (Figure 13), in a quite similar way to the relationship with population density, especially for the median.

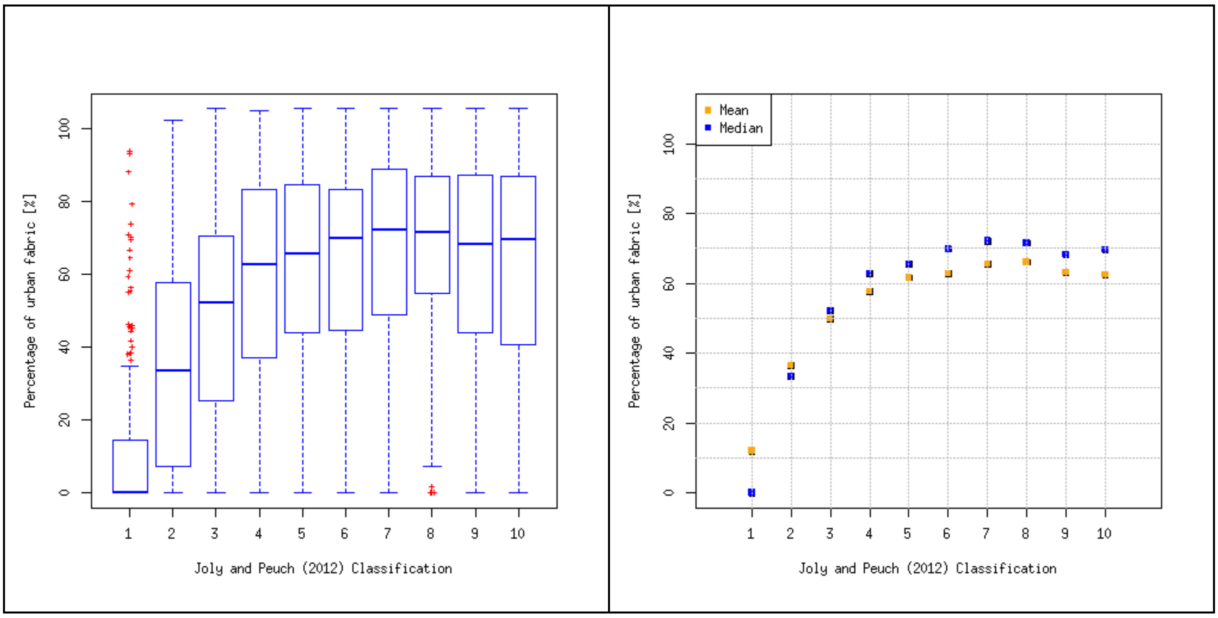


Figure 13: Same graphs as in Figure 12, considering the fraction (%) of urban fabric[6] within a radius of 1km[7] at the station instead of population density.

4.3.4 Spatial analysis

All PM_{10} monitoring stations for which classes could be calculated are mapped on Figure 14. Figure 15 only shows the stations that have been classified as 1 to 3 and 8 to 10. There is not a uniform distribution of classes over Europe. Whereas high classes (8 to 10) can be found in all countries, low classes (1 to 3) are almost absent from France and Eastern Europe. Population density and land cover are not sufficient to explain this uneven distribution. As displayed by Figure 16, the stations belonging to

---

[5] Data (inhabitants/km2) taken from the JRC/EEA population database, http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2

[6] Classes 111 and 112 of CORINE Land Cover database

[7] To bring out the relationship between PM_{10} classification and land use, the fraction of urban fabric was calculated within different averaging distances from the station locations. For this analysis on the European scale a radius of 1 km is considered. In the analysis further performed on the urban scale (see section 5) a radius of 500 m is used for PM_{10}.

those countries (red points) are not preferentially located in highly populated areas. Other auxiliary data, in particular emission inventories, should therefore be considered both on the European and on the local level.

Interesting is also that in Spain and Portugal most stations classified as 1 and 2 are described as rural background stations in AirBase whereas in Germany a significant fraction of them (27%) is made of urban or suburban stations. This could indicate that the definition of urban and suburban background stations is variable according to the country because of different siting criteria or due to different city configurations.

In addition to those results, a list of stations showing discrepancies between Joly and Peuch (2012) classification and the usual classification derived from AirBase metadata is provided in Annex II. Outliers for rural background stations (class of 4 or higher) and outliers for urban traffic stations (class of 4 or lower) show different spatial patterns: the first set of stations is mainly located in France, Belgium, Czech Republic and Poland; the second set is mostly located in Germany, Finland and Spain. Further investigation for those sites will be useful.

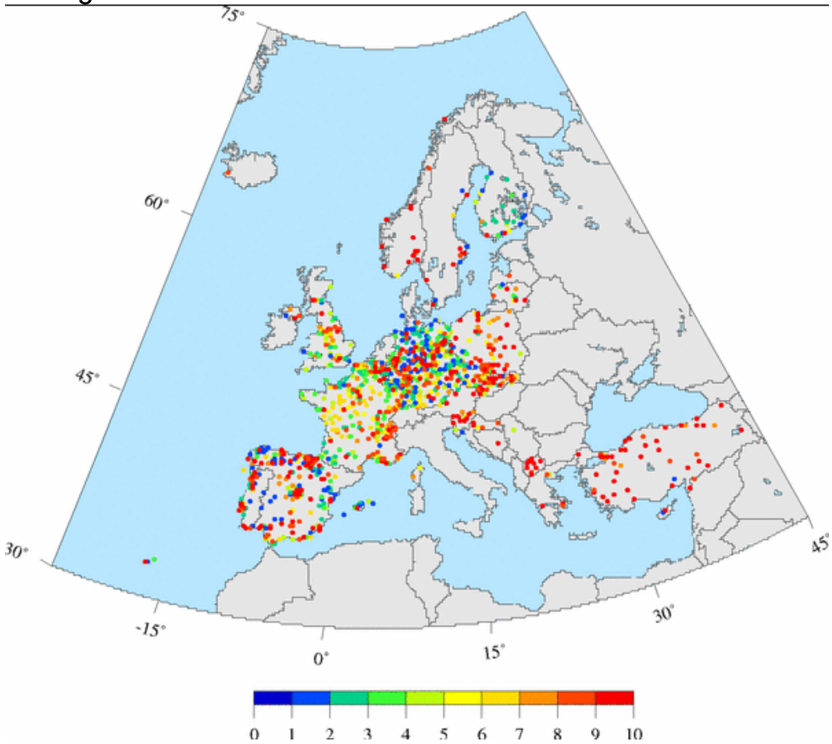


Figure 14: Overview map of PM_{10} monitoring stations and the classes that were assigned to them for by the Joly and Peuch algorithm (2012)

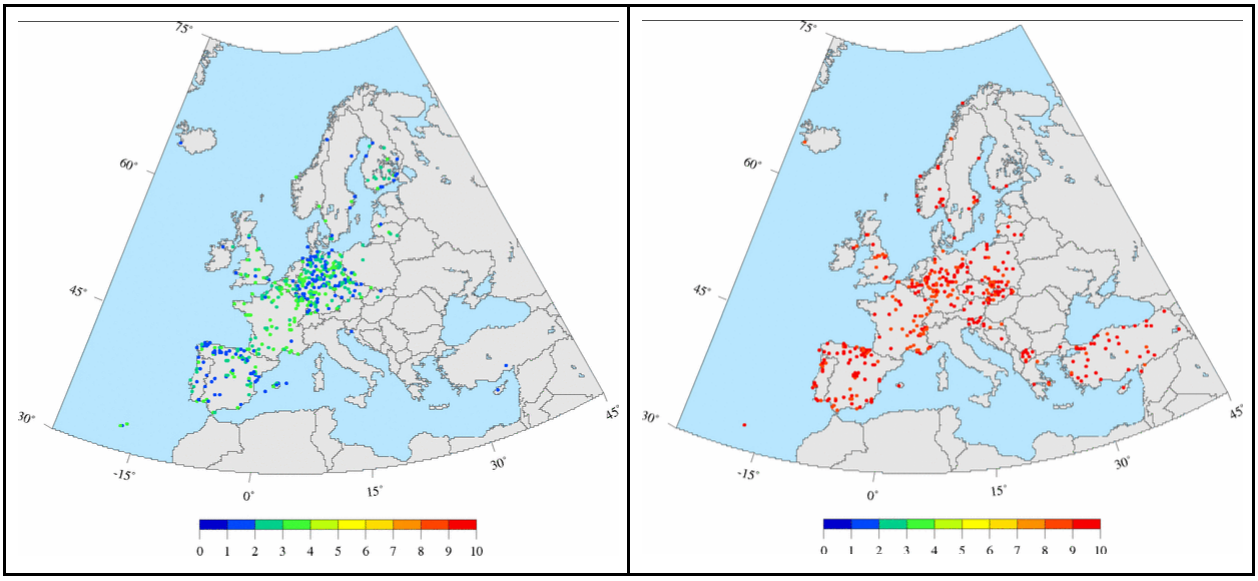


Figure 15: Overview map of PM_{10} monitoring stations. Left graphic: 547 stations with low classes (1, 2, 3). Right graphic: 547 stations with high classes (8, 9, 10).

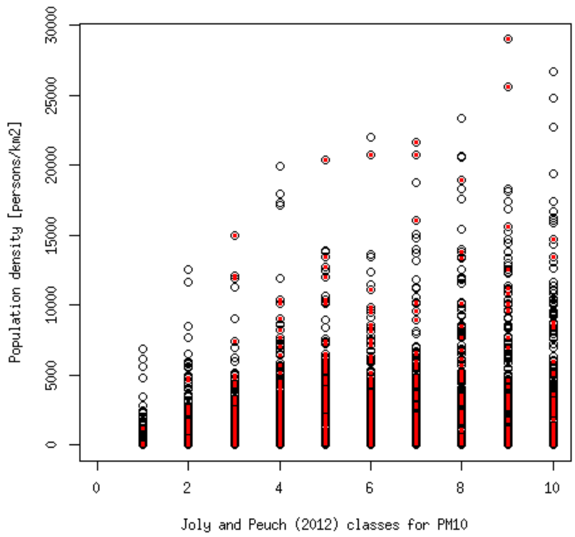


Figure 16: the relationship between the categories of the new PM_{10} station classification and population density. In red: stations located in France, Slovenia, Croatia, Bosnia, Macedonia, Greece and Turkey.

4.4 Results for O_3

4.4.1 Analysis of the new classes as a function of the station environment

As for NO_2 and PM_{10}, rural background stations show a decreasing distribution from class 1 to class 10 whereas an opposite variation is observed for traffic stations (Figure 17 and Figure 18, to be compared to Figures 3 and 10 and Figures 4 and 11, respectively). Urban background stations are spread throughout all classes, with a maximum at class 5. The same applies for suburban background stations with more surprisingly a maximum at class 7. Classes 9 and 10 include a noticeable proportion of urban and suburban background stations.

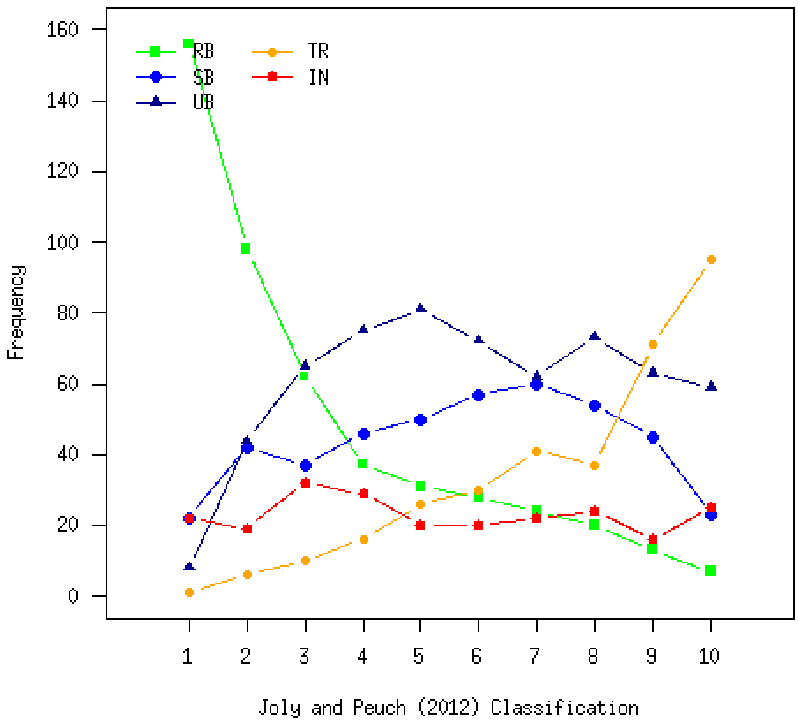


Figure 17: Histograms of the new O_3 station classification, categorized by AirBase metadata. RB = Rural background, SB = Suburban background, UB = Urban background, TR = Traffic, IN = Industrial.

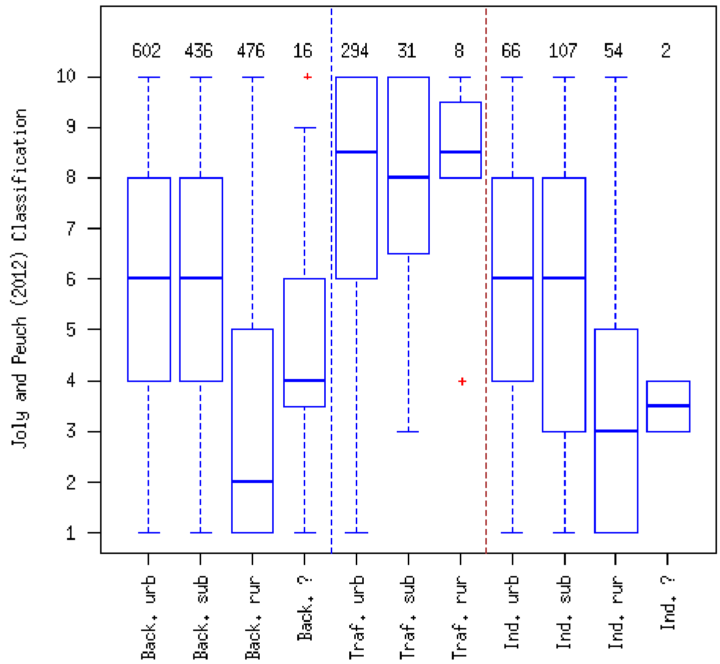


Figure 18: Box plot showing the median (blue line), 25^{th} and 75^{th} percentiles (top and bottom of blue rectangle), extreme values excluding outliers (blue whiskers) and outliers (red crosses) of the new classification system for O_3, categorized by AirBase station metadata. The number of stations per AirBase category is indicated on the graph.

4.4.2 Dependence on population density

The relationship between population density and the new classification (Figure 19) is logically slightly different from the one observed for NO_2 and PM_{10} but some patterns are questionable: population density increases from class 1 to class 6, stabilizes from class 6 to class 8 and raises again from class 8 to class 10.

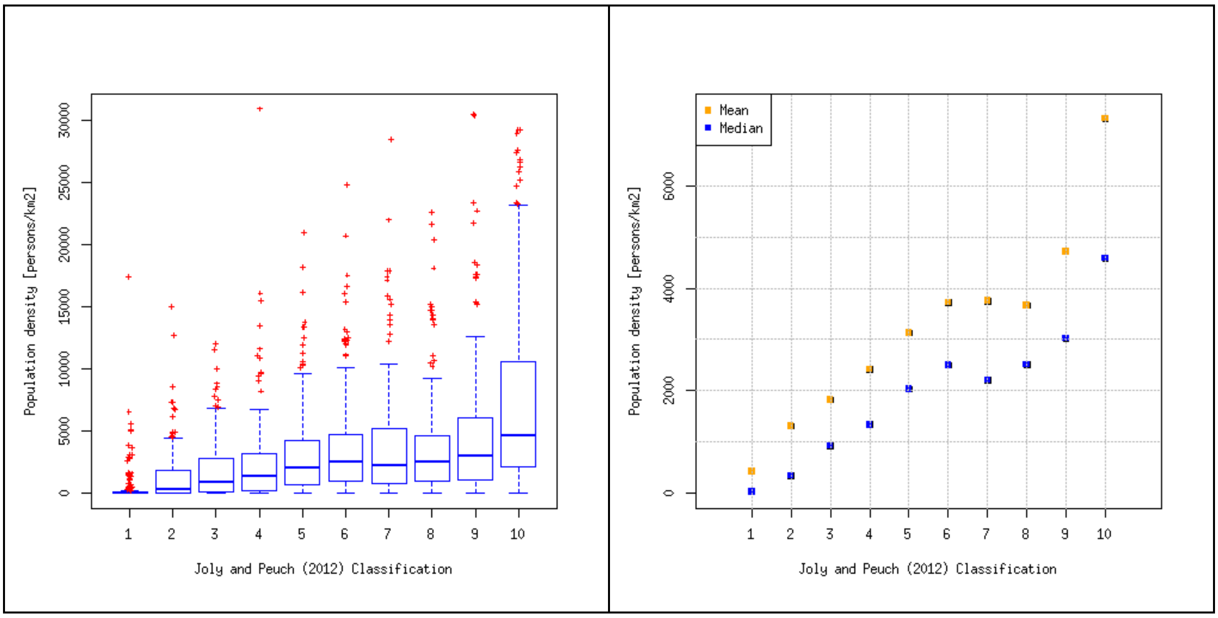


Figure 19: Left: the relationship between the categories of the new O_3 station classification and population density[8]. This box plot shows the median (blue line), 25th and 75th percentiles (upper and lower limits of the blue box), extreme values excluding outliers (blue whiskers), and outliers (red crosses) of the rate of population density, categorized by Joly and Peuch (2012) classification. Right: Same graph but only showing the median and mean of the rate of population density for each of the ten classes.

4.4.3 Dependence on land cover

In a similar way, the relationship between urban fabric and the new classification (Figure 20) shows an increase of urban fabric rate from class 1 to class 5, a stabilization from class 5 to class 8 and a new slighter raise from class 8 to class 10.

[8] See footnote 5.

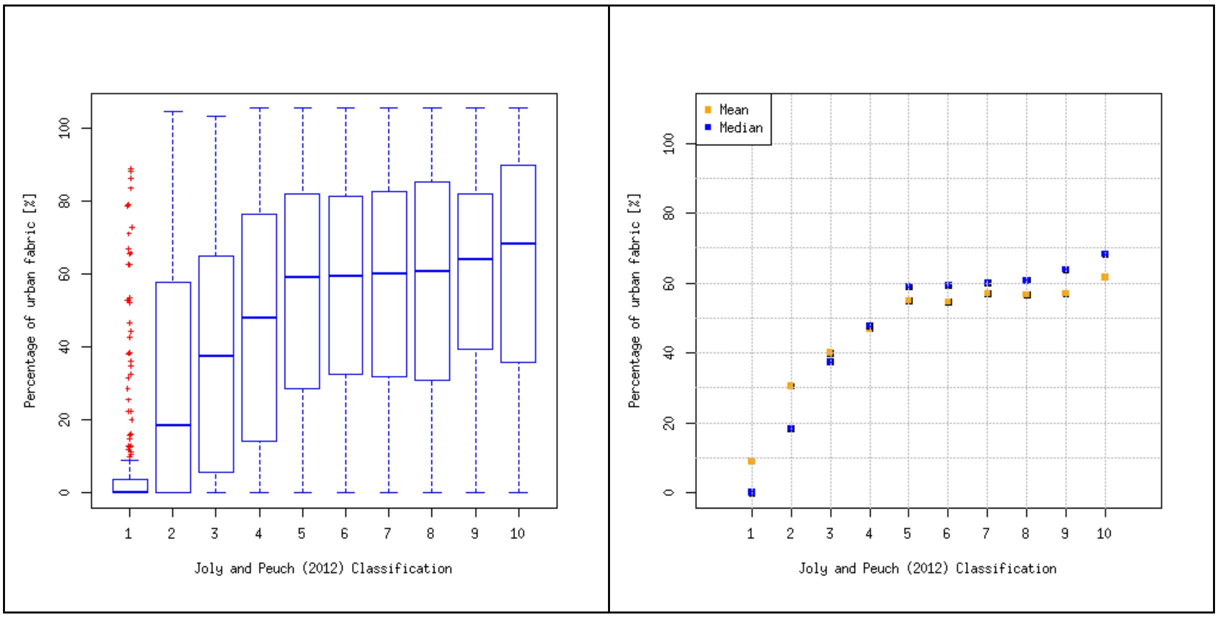


Figure 20: Same graphs as in Figure 12, considering the fraction (%) of urban fabric within a radius of 1km[9] at the station instead of population density.

4.4.4 Spatial analysis

All O_3 monitoring stations for which classes could be calculated are mapped on Figure 21. Figure 22 only shows the stations that have been classified as 1 to 3 and 8 to 10. The spatial distribution of classes 8 to 10 exhibits a strong pattern with a high density of stations in the Alps and Po Valley regions and almost no station in the northern part of Europe.

In addition to those results, a list of stations showing discrepancies between Joly and Peuch (2012) classification and the usual classification derived from AirBase metadata is provided in Annex III. Further investigation is needed for those sites.

[9] see footnotes 6 and 7

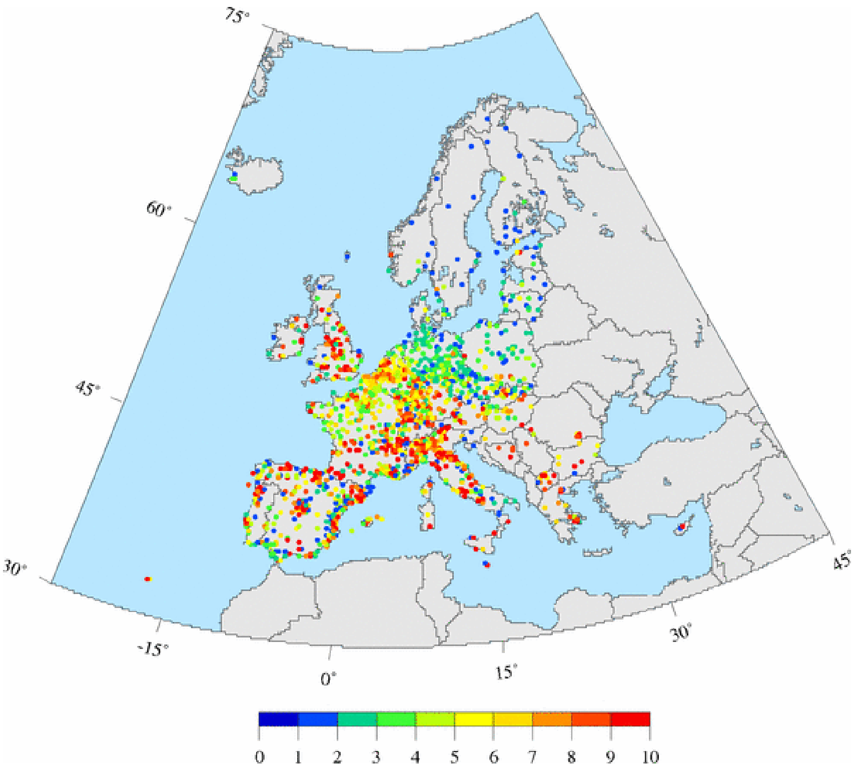


Figure 21: Overview map of O_3 monitoring stations and the classes that were assigned to them for by the Joly and Peuch algorithm (2012).

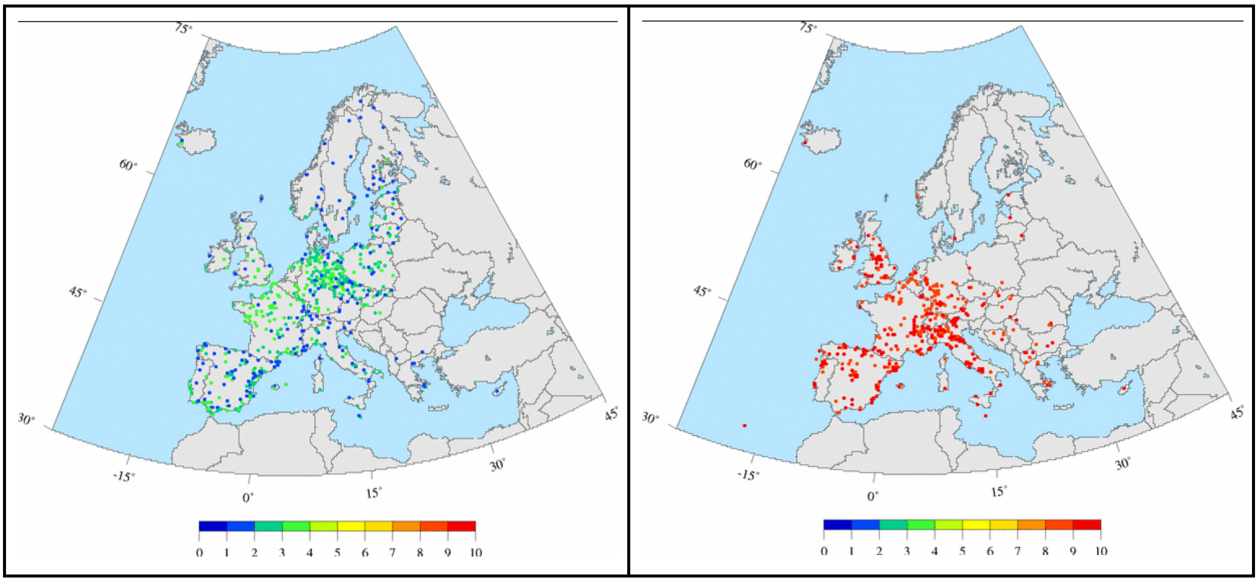


Figure 22: Overview map of O_3 monitoring stations. Left graphic: 628 stations with low classes (1, 2, 3). Right graphic: 628 stations with high classes (8, 9, 10).

4.5 Differences between NO_2, PM_{10} and O_3 classifications

The new classification method is pollutant dependent and makes use of time series of hourly concentrations and AirBase metadata as prior knowledge. This section points out the differences between the classifications obtained for NO_2, PM_{10} and O_3 respectively, thus highlighting stations that require closer analysis of the new classification and further check of AirBase metadata.

4.5.1 Differences between NO_2 and PM_{10} classifications

Figure 23 represents the distribution of the differences between NO_2 and PM_{10} classes. This histogram has a Gaussian shape with a maximum frequency at zero, i.e. a null difference in the classes. For 90% of stations measuring both pollutants, the difference is less than 3 classes. This good agreement (also visible in Table 2) is coherent considering the common origins of both pollutants (road traffic and other fossil fuel combustion). The occurrence of some differences was to be expected as well due to the more local character of NO_2 pollution or the long range transport that can influence PM concentrations. The largest differences are highlighted in Figure 24. This map shows a group of stations in Eastern Europe (blue points) where PM_{10} classes are significantly higher than NO_2 ones.

Table 2: Distribution of the monitoring stations as a function of PM_{10} and NO_2 classes.

	PM_{10} classification									
NO_2	1	2	3	4	5	6	7	8	9	10
1	77	20	5	2	0	0	1	0	0	0
2	21	33	14	9	9	2	2	1	1	0
3	7	21	19	11	21	9	6	3	2	2
4	3	12	26	17	14	13	9	4	6	1
5	4	8	16	22	21	24	16	3	5	5
6	1	3	12	12	17	20	12	13	4	6
7	0	1	5	12	16	18	17	15	15	4
8	0	1	4	6	3	9	10	13	7	12
9	0	1	0	3	4	7	12	12	15	16
10	0	0	1	2	0	1	3	5	10	17

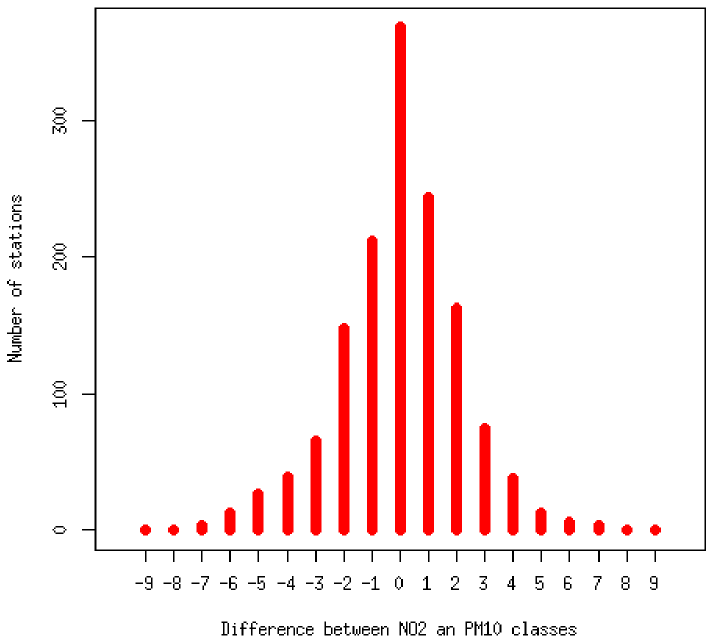


Figure 23: Statistical distribution of the differences between NO_2 and PM_{10} classification (for those stations measuring both pollutants). Differences calculated as: NO_2 class – PM_{10} class.)

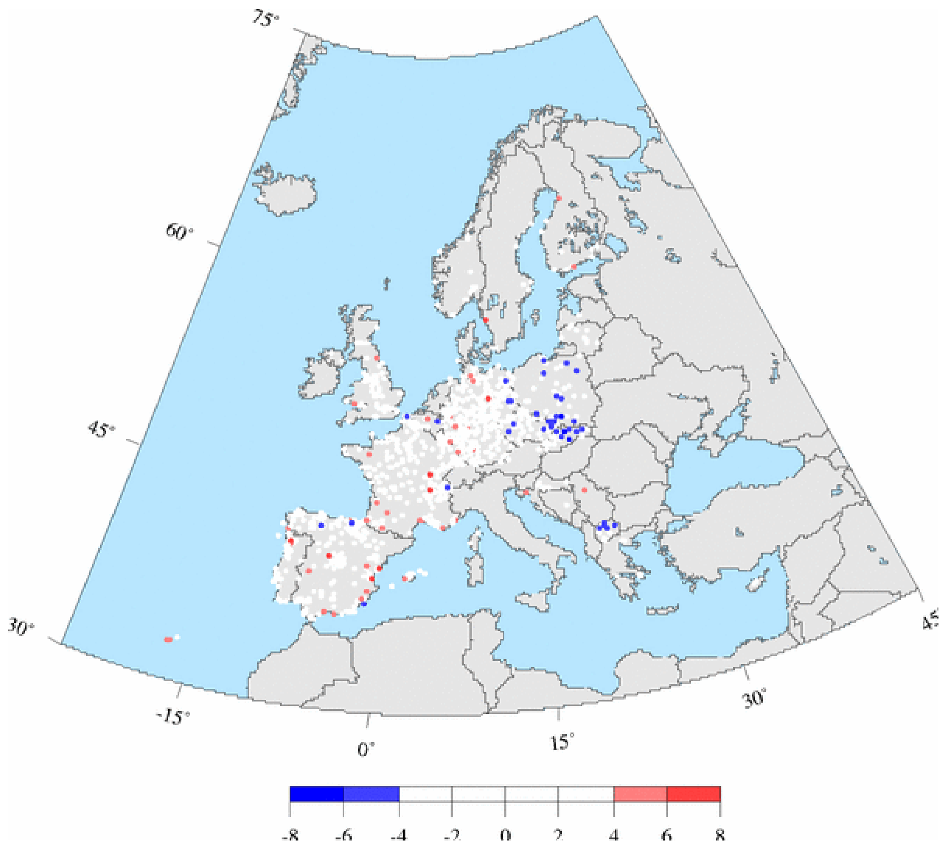


Figure 24: Map of the differences between NO_2 and PM_{10} classification. In colour: monitoring stations with a difference of more than four classes.

4.5.2 Differences between NO_2 and O_3 classifications

Negative differences between NO_2 and O_3 classes (higher classes assigned for O_3) occur more frequently than positive difference (Table 3 and Figure

25). Differences up to 6 classes are observed for a significant number of stations.

Most of the stations with significant differences (Figure 26) are located in coastal zones, Alpine valleys or Mediterranean regions. Those regions are known to be difficult to study. The balance between titration and ozone production is not clear. In coastal zone, ozone production is affected by the land-sea conditions in summer anticyclonic situations. In mountainous regions, the chemistry and transport of ozone are affected by topography effects, which can explain differences between both classifications.
For some stations, additional check of the time series is needed to avoid inconsistencies in the datasets. For example from 2002 to 2010, station AT31496 has only 2 years of O_3 data, and 3 years of NO_2 data.

Table 3: Distribution of the monitoring stations as a function of NO_2 and O_3 classes

	O_3 classification									
NO_2	1	2	3	4	5	6	7	8	9	10
1	99	53	27	14	8	4	4	2	0	0
2	20	42	43	25	24	11	8	10	0	3
3	5	29	29	30	22	24	16	19	7	6
4	4	10	18	27	34	25	27	19	10	8
5	4	9	17	24	25	30	29	23	22	7
6	0	4	12	16	25	23	34	30	21	19
7	1	6	6	17	16	17	25	29	34	27
8	0	3	8	5	7	13	15	21	30	46
9	0	3	1	3	13	16	13	12	26	29
10	1	0	0	0	2	2	6	10	19	32

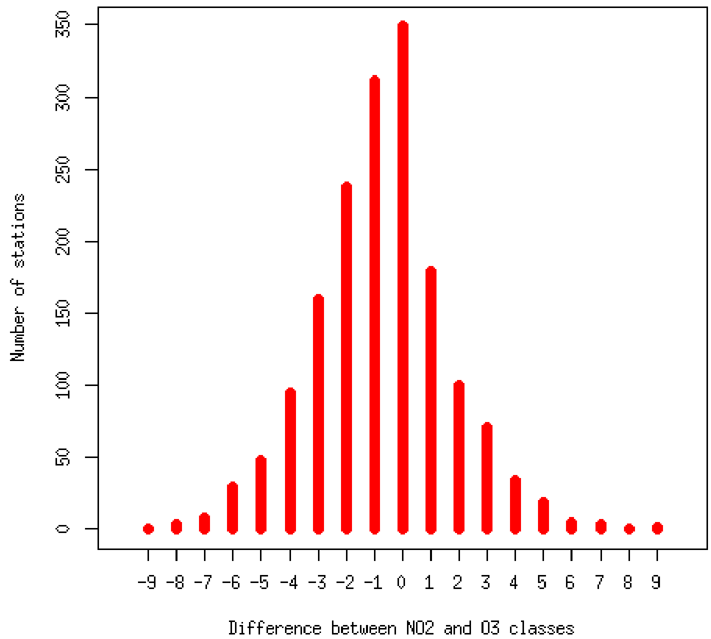


Figure 25: Statistical distribution of the differences between NO_2 and O_3 classification (for those stations measuring both pollutants). Differences calculated as: NO_2 class – O_3 class.

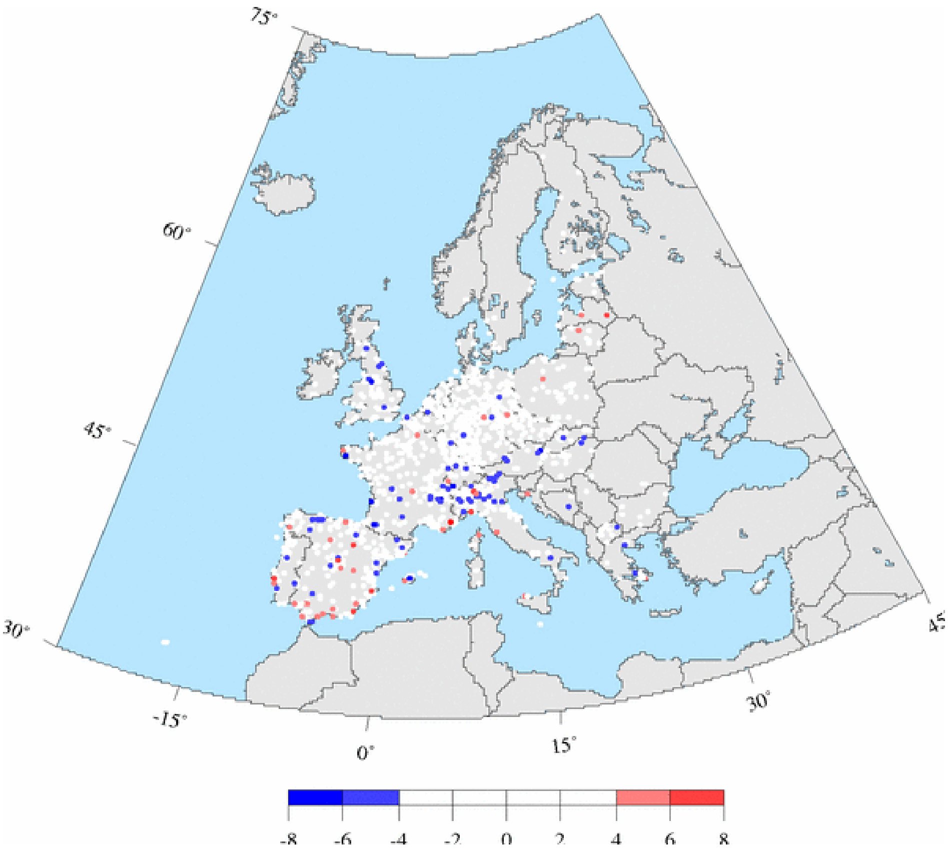


Figure 26: Map of the differences between NO_2 and O_3 classification. In colour: monitoring stations with a difference of more than four classes.

4.5.3 Differences between NO_2, PM_{10} and O_3 classifications with regard to population density and landuse

For NO_2 and PM_{10} (Figure 5, Figure 6, Figure 12, Figure 13), the relationship between station environment and Joly and Peuch (2012) classification is characterized by an increase of population density and urban fabric rate from class 1 to class 7 and by a more variable pattern from class 8 to class 10. For O_3 (Figure 19 and Figure 20), the shape of this relationship is slightly different, with an increase of population density and urban fabric rate from class 1 to class 5 or 6, a stabilization from class 5 or 6 to class 8 and a new rise from class 8 to class 10.

Those results suggest that for NO_2 and PM_{10}, the highest classes (8 to 10) are related to nearby sources and local conditions which population density and land use cannot well capture. For O_3, larger scale influences could contribute to high classes. It should also be mentioned that background stations are in higher proportion for O_3, which may influence the relationship between the classification and the surrounding environment. Ozone pollution is mainly a regional issue and the Air Quality Directive (2008/50/EC) only requires O_3 monitoring at rural, suburban and urban background sites. Ozone background stations have therefore a larger weight in the construction of the Fisher axis and the definition of the classes. For NO_2 and PM_{10}, the proportion of background stations in the training dataset is about 60%; for O_3 it reaches approximately 75%.

5. In-depth analysis for the Air Implementation Pilot cities

In the following sections, the results obtained from applying the Joly and Peuch (2012) station classification algorithm are analysed in detail for several of the European “Air Implementation Pilot cities”. The analysis is carried out for NO_2 and PM_{10} which are most characteristic of urban air quality. A specific focus is set on investigating the station environment. A comparison between all cities is then provided.

5.1 Berlin

5.1.1 NO_2

Figure 27 shows a map of Berlin highlighting the locations of the air quality stations. Each station is coded by colour, indicating the AirBase metadata on station classification, and by number, indicating the new Joly and Peuch (2012) classification. Very good correspondence exists for rural background stations which are all classified as 2 in the new classification system. Only one suburban background station exists in the area and it is classified as 2. As for urban background stations, all of them occur in the city cores of Berlin and have classes ranging from 3 to 5. As would be expected for traffic stations, most of them are classified in very high classes of 9 and 10, although two traffic stations were also classified in classes 5 and 6. Finally, only one industrial station was located within the extent of the map and it was classified as class 2, which is in fact how most industrial stations are classified at the European-level.

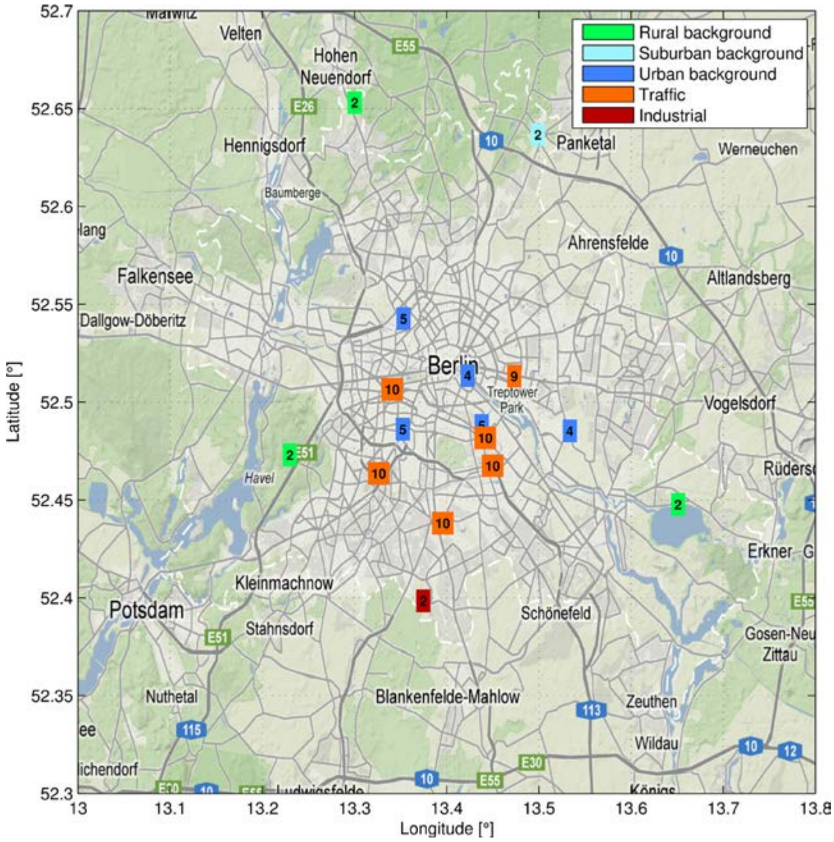


Figure 27: Spatial comparison of the two classification systems in the air quality zone of Berlin. The traditional AirBase metadata classification is shown as coloured boxes whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

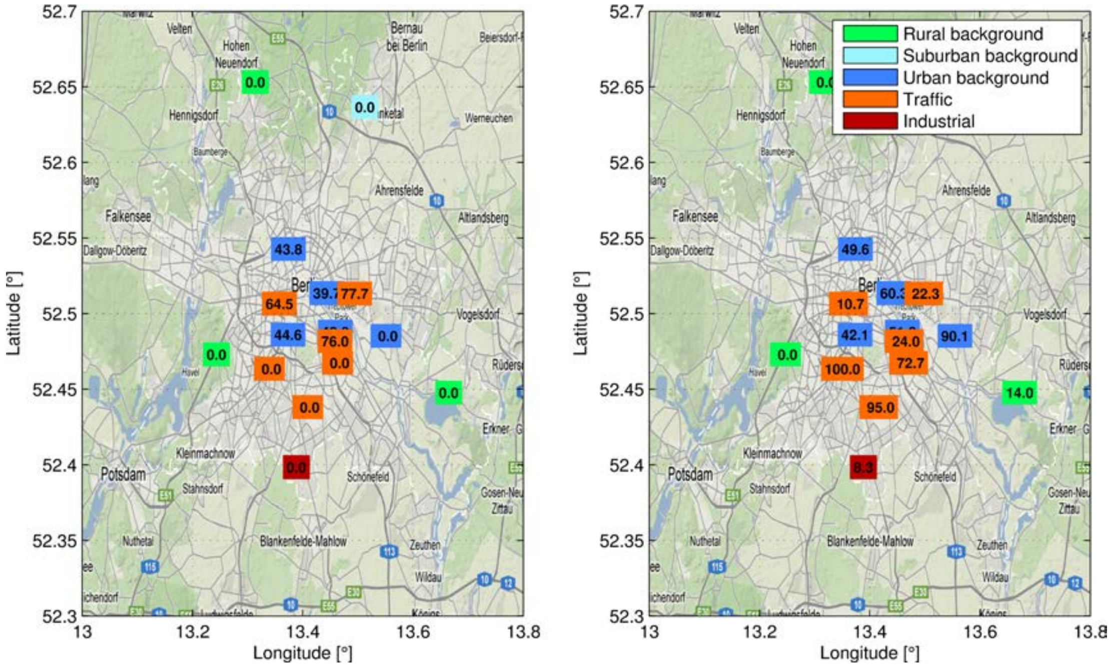


Figure 28: Fraction (in percent) of the land cover classes "Continuous Urban Fabric" (left) and "Discontinuous Urban Fabric" (right) within a 1 km x 1 km box around each NO_2 monitoring station in Berlin. The coloured boxes further indicate the station classification as given by the AirBase metadata.

Figure 28 provides information on the station environment for each station in Berlin, expressed as the percentage of the two urban land cover classes included in the CORINE land cover dataset, namely CLC111 (Continuous Urban Fabric) and CLC112 (Discontinuous Urban Fabric). The fractions were computed for a 1 km x 1 km rectangle around each station. Interestingly, several stations classified as 10 in Figure 27 do not have very high fraction of continuous urban fabric, however they do generally have relatively high fraction of discontinuous urban fabric. All stations classified as 2 have no (or only a very small) fraction of urban land cover classes in their vicinity.

5.1.2 PM_{10}

In the same way as Figure 27, Figure 29 simultaneously displays AirBase metadata on station classification (colours) and the new Joly and Peuch (2012) classes (numbers). As for NO_2, a very good correspondence exists between both classification systems with a low class (2) for all rural background sites, a rather low one (3) for the suburban background site, intermediate classes (5-6) for most urban background stations and high classes (8-9) for most traffic stations. The difference in class values between urban background and urban traffic sites is less pronounced than for NO_2, which is consistent with European-wide results (Figure 10). The rural industrial station is classified as the rural background stations (class 2) suggesting there is no other influence than the industrial source itself, where it is located.

Two sites seem different, with a more “background oriented” behaviour: one urban background site which is classified as 3 and one urban traffic site which is classified as 6. Auxiliary data (Figure 30) do not provide any possible explanation since the density of urban fabric and population density are as high as in other locations. It is to note that both stations are close to each other, which could indicate that this specific behaviour is related to some local features.

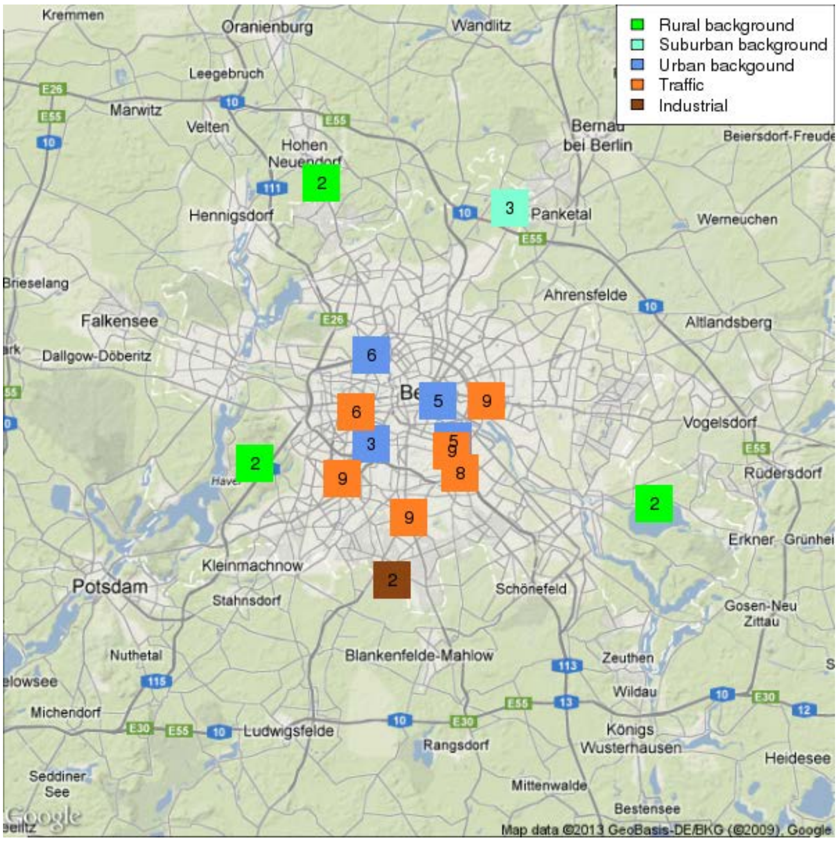


Figure 29: Spatial comparison of the two classification systems in the air quality zone of Berlin. The traditional AirBase metadata classification is represented by colours whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

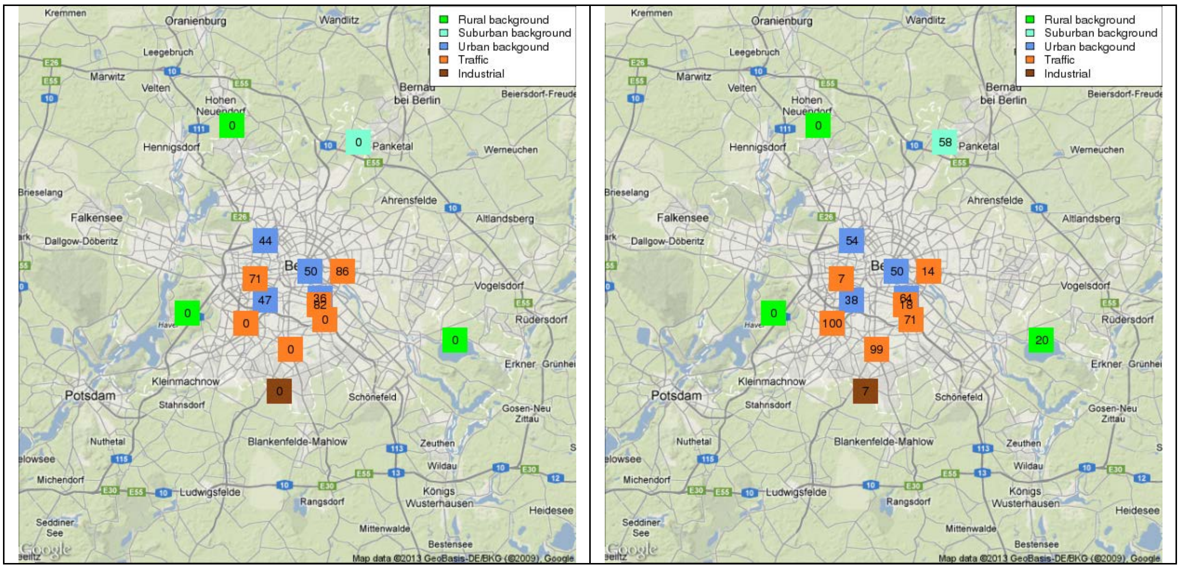


Figure 30: Fraction (in percent) of the land cover classes "Continuous Urban Fabric" (left) and "Discontinuous Urban Fabric" (right) within a circle of 1-km diameter around each PM_{10} monitoring station in Berlin. The colours indicate the station classification as given by the AirBase metadata.

5.2 Dublin

5.2.1 NO_2

Only few air quality stations are located in the city of Dublin. There is no rural background station or industrial station. There is one suburban background station which has been classified as 6, what is slightly higher than what is generally observed at suburban stations (the Europe median class for suburban stations is 4), but still quite reasonable considering that the station is in, or very close to, an industrial complex. There are also two urban background stations, both classified as 7. Finally, two traffic stations occurred in Dublin city and were classified between 6 and 8, which is slightly below the European median classification for traffic stations of 9.

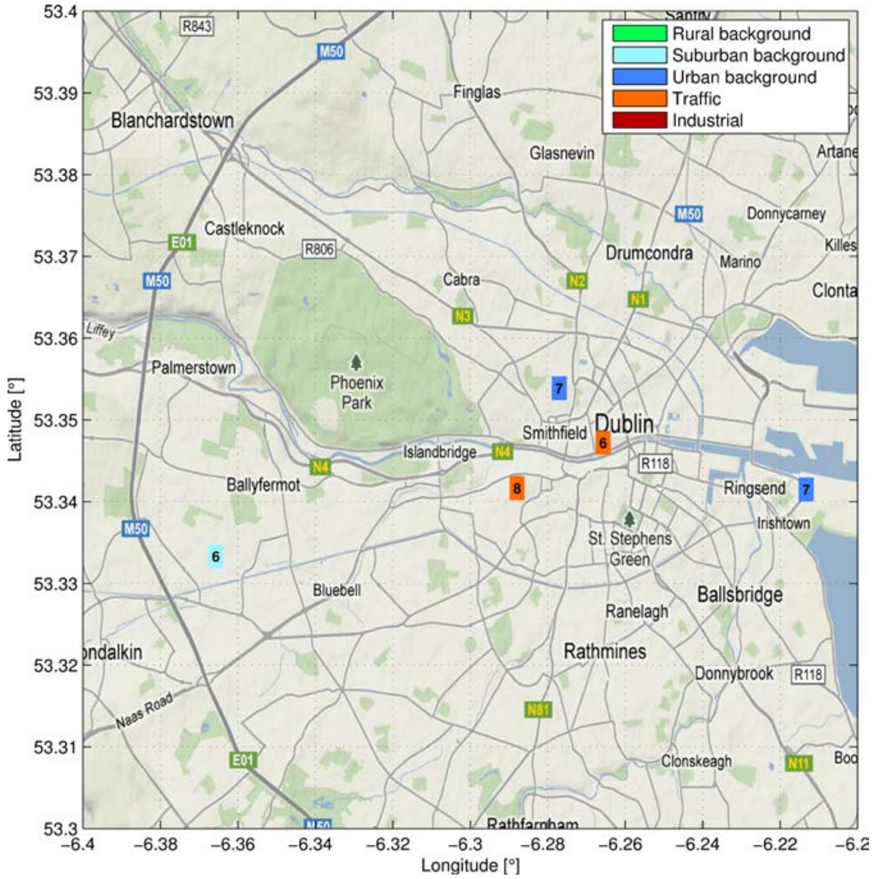


Figure 31: Spatial comparison of the two classification systems in the city of Dublin. The traditional AirBase metadata classification is shown as coloured boxes whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

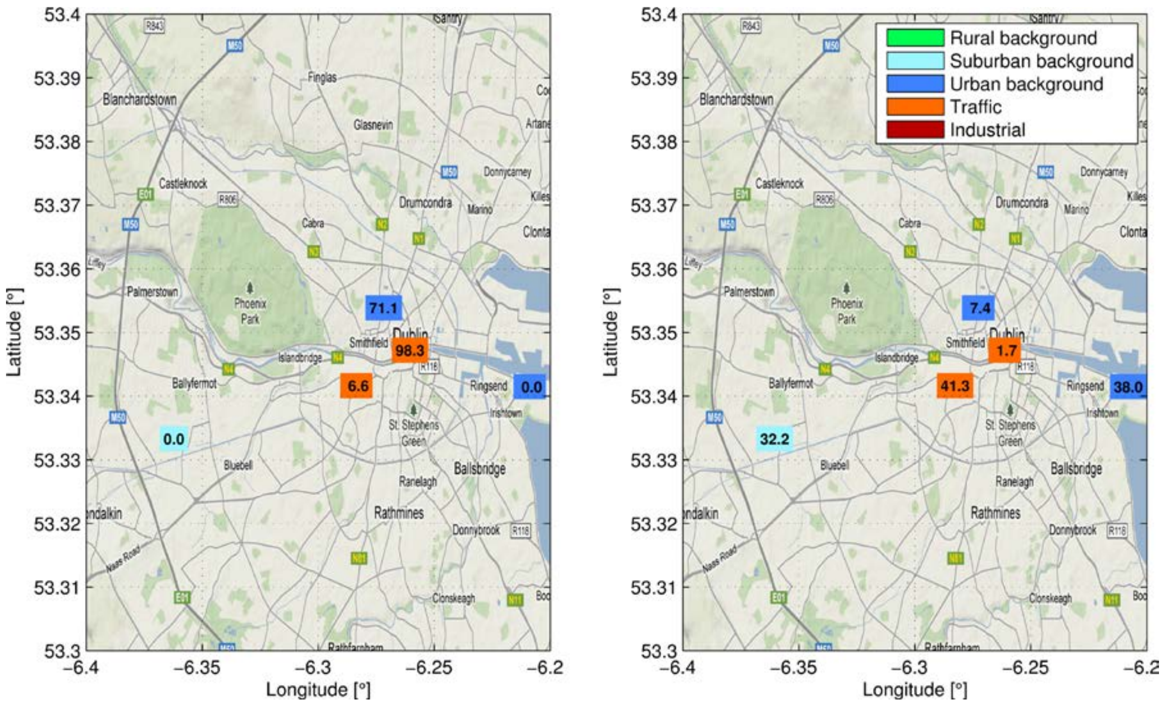


Figure 32: Fraction (in percent) of the land cover classes "Continuous Urban Fabric" (left) and "Discontinuous Urban Fabric" (right) within a 1 km x 1 km box around each NO_2 monitoring station in Dublin. The coloured boxes further indicate the station classification as given by the AirBase metadata.

Figure 32 provides information about the station environment for each station in Dublin, expressed as the percentage of the two urban land cover classes included in the CORINE land cover dataset, namely CLC111 (Continuous Urban Fabric) and CLC112 (Discontinuous Urban Fabric). The traffic stations, which were also classified as 6 and 8 all have high fraction of urban fabric (both continuous and discontinuous). The one suburban background station which was classified as 6 has no continuous urban fabric and only a small fraction of discontinuous urban fabric in its vicinity.

5.2.2 PM_{10}

Available data sets for Dublin do not include any time series of hourly PM_{10} concentrations. Consequently, no analysis of station classification could be performed yet for PM_{10} (see Sections 2.3 and 4.3).

5.3 Madrid

5.3.1 NO_2

A large number of air quality stations have been reported in the air quality zone of Madrid, however no rural background stations are among them. There is one suburban background station, classified as 7, which is significantly above the European median for suburban background stations (4). The situation is very similar for urban background stations, which are all classified between 8 and 10 and as such quite significantly above the European median class (5) for urban background stations. The majority of traffic stations are classified as 10 and are thus, with only three exceptions, classified slightly higher than the European median of 9.

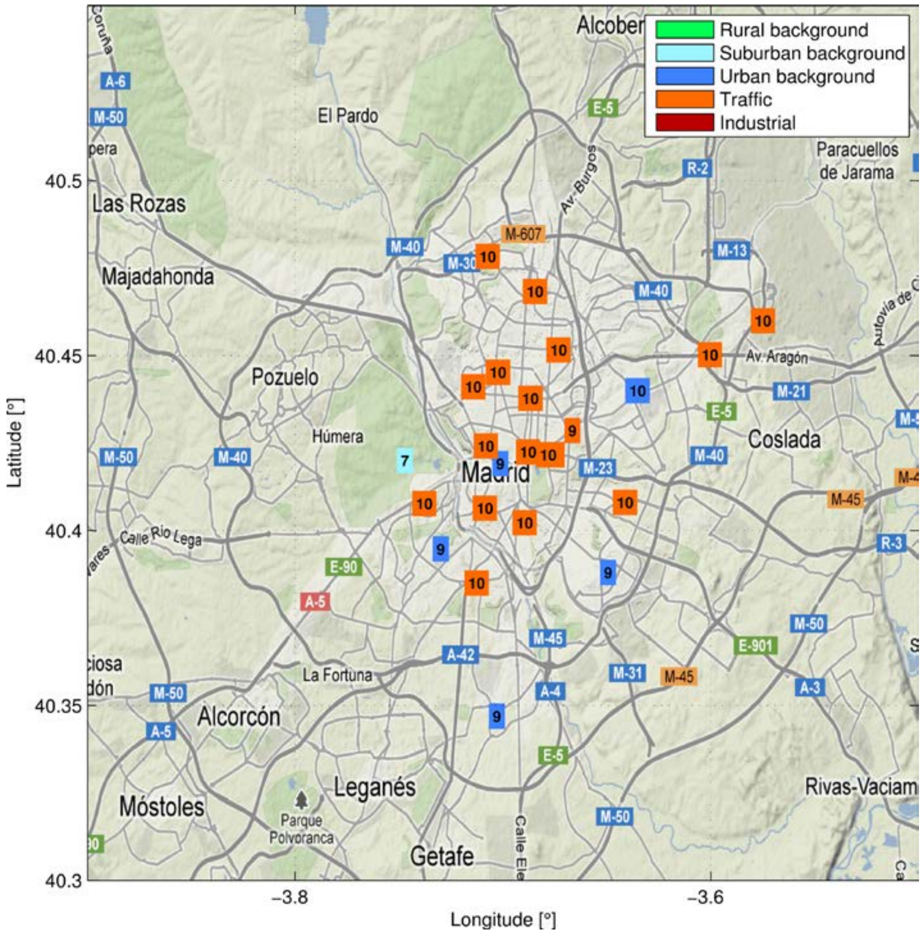


Figure 33: Spatial comparison of the two classification systems in the air quality zone of Madrid. The traditional AirBase metadata classification is shown as coloured boxes whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

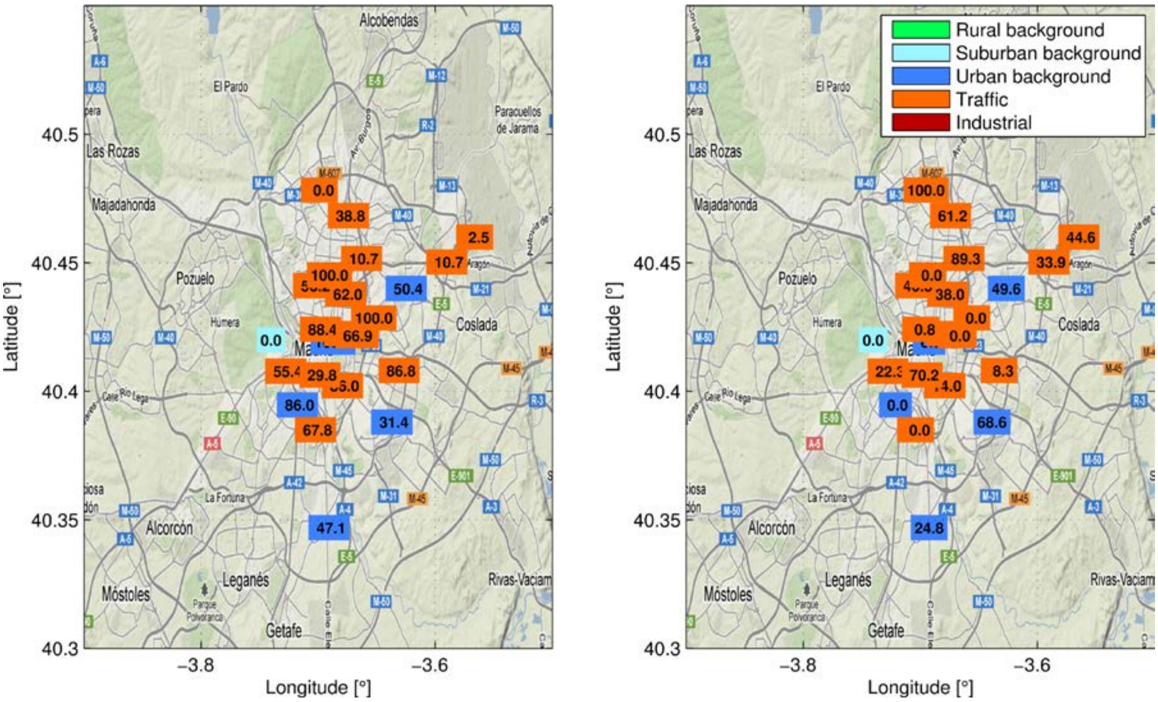


Figure 34: Fraction (in percent) of the land cover classes "Continuous Urban Fabric" (left) and "Discontinuous Urban Fabric" (right) within a 1 km x 1 km box around each NO_2 monitoring station in Madrid. The coloured boxes further indicate the station classification as given by the AirBase metadata.

Figure 34 provides information on the station environment for each station in Madrid, expressed as the percentage of the two urban land cover classes included in the CORINE land cover dataset, namely CLC111 (Continuous Urban Fabric) and CLC112 (Discontinuous Urban Fabric). Many traffic stations in Madrid have been classified as 10, which is consistent with the fraction of land cover in their vicinity. Most of the traffic stations in the central city have very high percentages of continuous urban fabric. Most of the other traffic stations have high fractions of discontinuous urban fabric. If both fractions are taken into account, all urban background stations are completely surrounded by urban fabric. In addition, population density is very high as shown in Annex IV (Figure 60).

5.3.2 PM_{10}

As for NO_2, the few PM_{10} monitoring stations with enough data to be classified according to the Joly and Peuch (2012) classification show a traffic-oriented behaviour with classes significantly above the European median for the background sites: the suburban station is classified as 7 (European median is 4); the urban background stations are classified as 9 (European median is 6) and even 10 for one site close to industrial or commercial units. The urban traffic stations are classified between 8 and 10 (Figure 35).
All sites are located in densely built and populated areas (Figure 60, Annex IV).

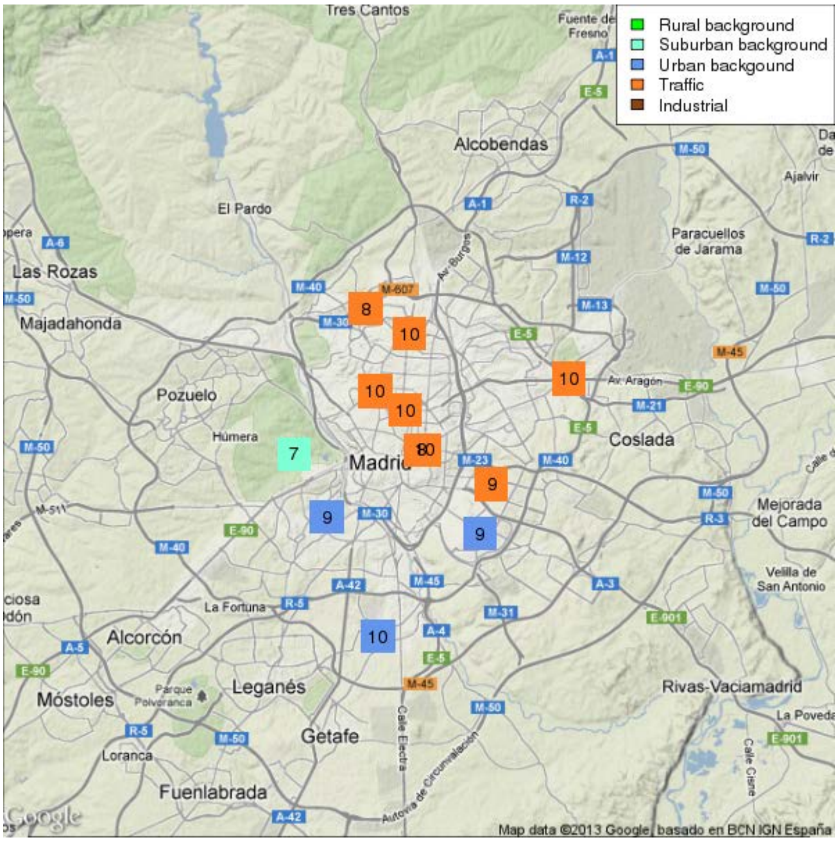


Figure 35: Spatial comparison of the two classification systems in the air quality zone of Madrid. The traditional AirBase metadata classification is represented by colours whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

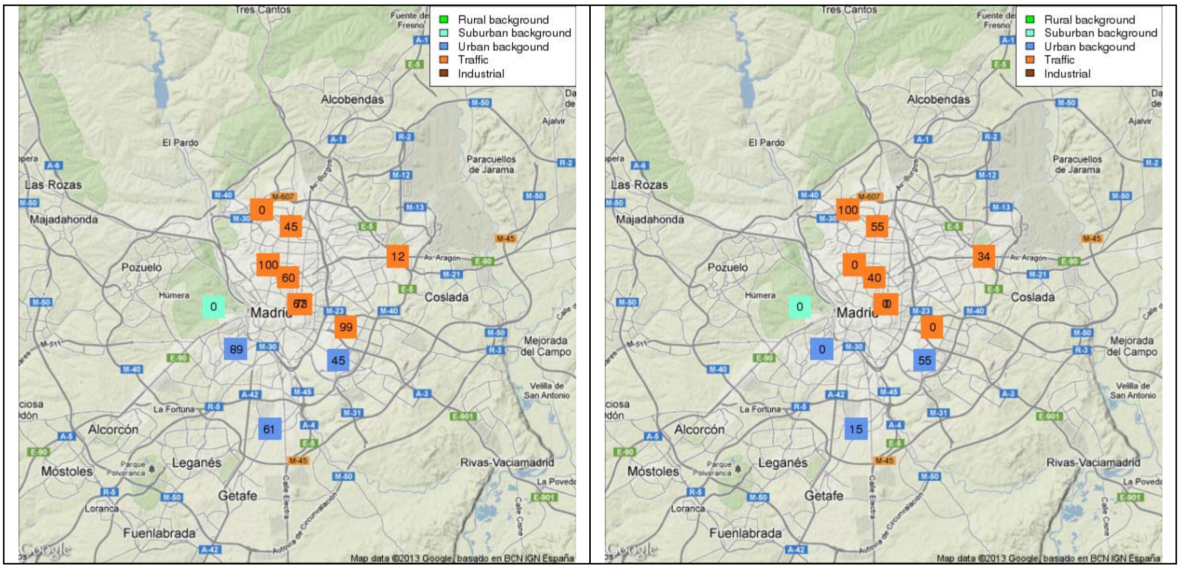


Figure 36: Fraction (in percent) of the land cover classes "Continuous Urban Fabric" (left) and "Discontinuous Urban Fabric" (right) within a circle of 1-km diameter around each PM_{10} monitoring station in Madrid. The colours indicate the station classification as given by the AirBase metadata.

5.4 Malmö

5.4.1 NO_2

Only two stations exist in the city of Malmö, namely one urban background station and one traffic station. With values of 4 and 7, respectively, they are both below the corresponding European medians of 5 and 9.

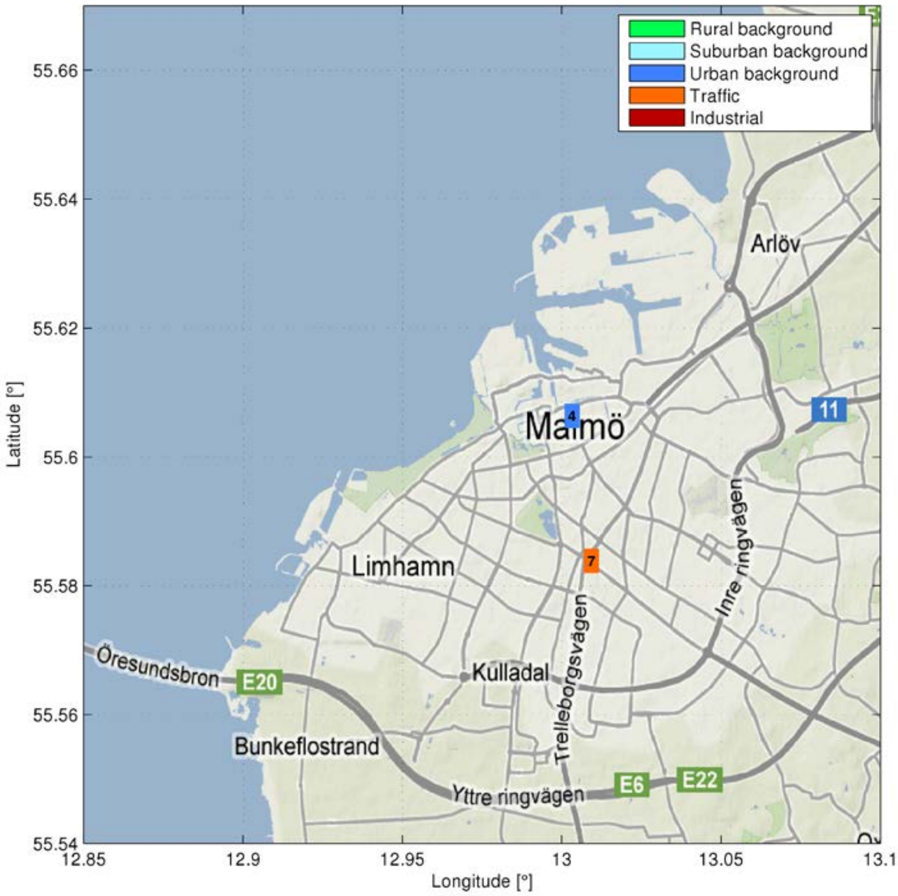


Figure 37: Spatial comparison of the two classification systems in the city of Malmö. The traditional AirBase metadata classification is shown as coloured boxes whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

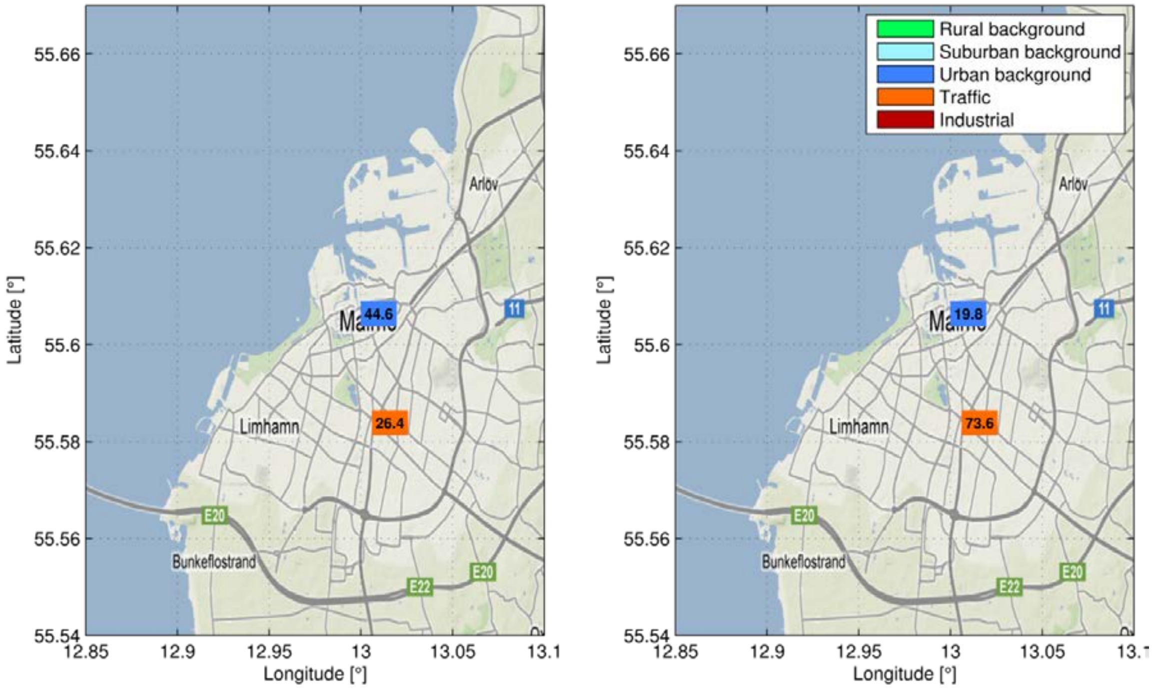


Figure 38: Fraction (in percent) of the land cover classes "Continuous Urban Fabric" (left) and "Discontinuous Urban Fabric" (right) within a 1 km x 1 km box around each NO_2 monitoring station in Malmö. The coloured boxes further indicate the station classification as given by the AirBase metadata.

Figure 38 provides information on the station environment for each station in Malmö, expressed as the percentage of the two urban land cover classes included in the CORINE land cover dataset, namely CLC111 (Continuous Urban Fabric) and CLC112 (Discontinuous Urban Fabric). Nearly half of the land cover in the immediate vicinity of the urban background station is made up of continuous urban fabric. Interestingly the station is still only classified as 4, although a higher value would be expected based on land cover considerations. However, population density is not very high compared to other cities (Figure 60, Annex IV). The majority of the land cover around the traffic station is discontinuous urban fabric. Despite the fact there is only a relatively small fraction of continuous urban fabric, the station was classified as 7.

5.4.2 PM_{10}

Stations measuring NO_2 and PM_{10} are the same. For PM_{10}, both sites are classified far below the European median, showing a strong “background behaviour”, or a weak influence of local sources: the urban background station is classified as 2 (European median is 6); the urban traffic station is classified as 4 (European median is 8) (Figure 35). This result is consistent with rather low observed concentration values - the annual average is around 15 µg/m^3 at the background site over the last years – and limited population density (Figure 60, Annex IV).

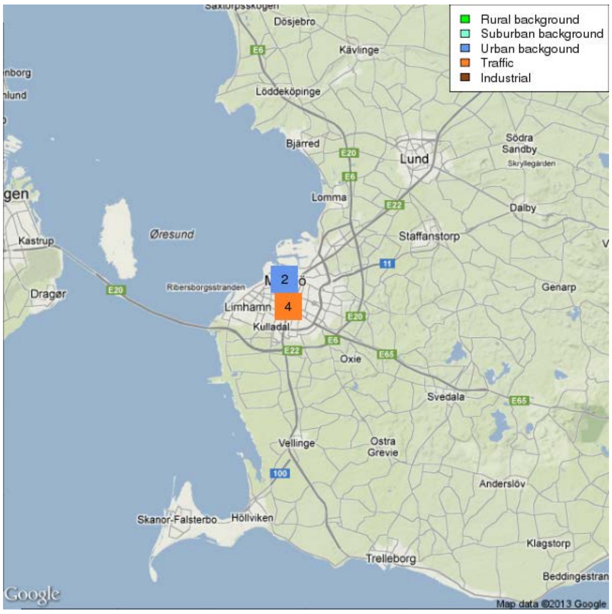


Figure 39: Spatial comparison of the two classification systems in the city of Malmö. The traditional AirBase metadata classification is represented by colours whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

5.5 Milan

5.5.1 NO_2

In the air quality zone of Milan and surrounding area[10] there is no rural background station and only one suburban background station is present. However, several urban background stations exist and with values ranging between 6 and 10 they are classified slightly to significantly higher than the European median of 5. The classification of the traffic stations tends to be around 7 to 8 with only one station at 9 and therefore it tends to be slightly below the European median of 9. No industrial station exists in the greater Milan area.

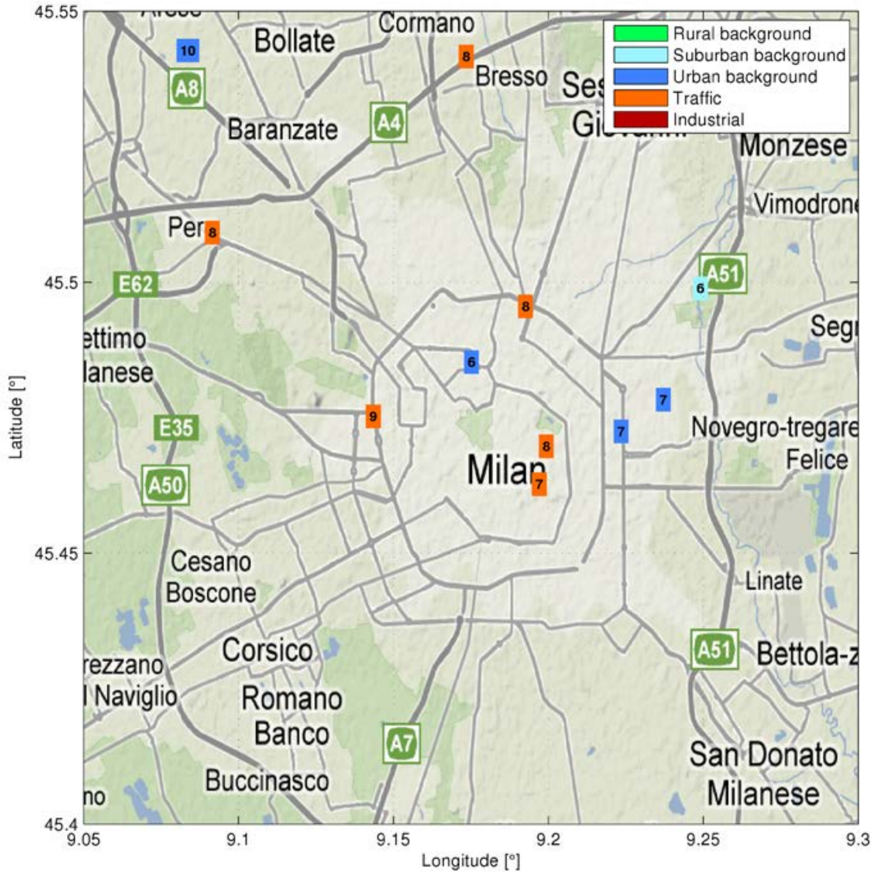


Figure 40: Spatial comparison of the two classification systems in the air quality zone of Milan and surrounding area[10]. The traditional AirBase metadata classification is shown as coloured boxes whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

[10] Since only few stations could be classified in the AQ zone of Milan, sites outside this domain have been considered as well to provide more detailed analysis.

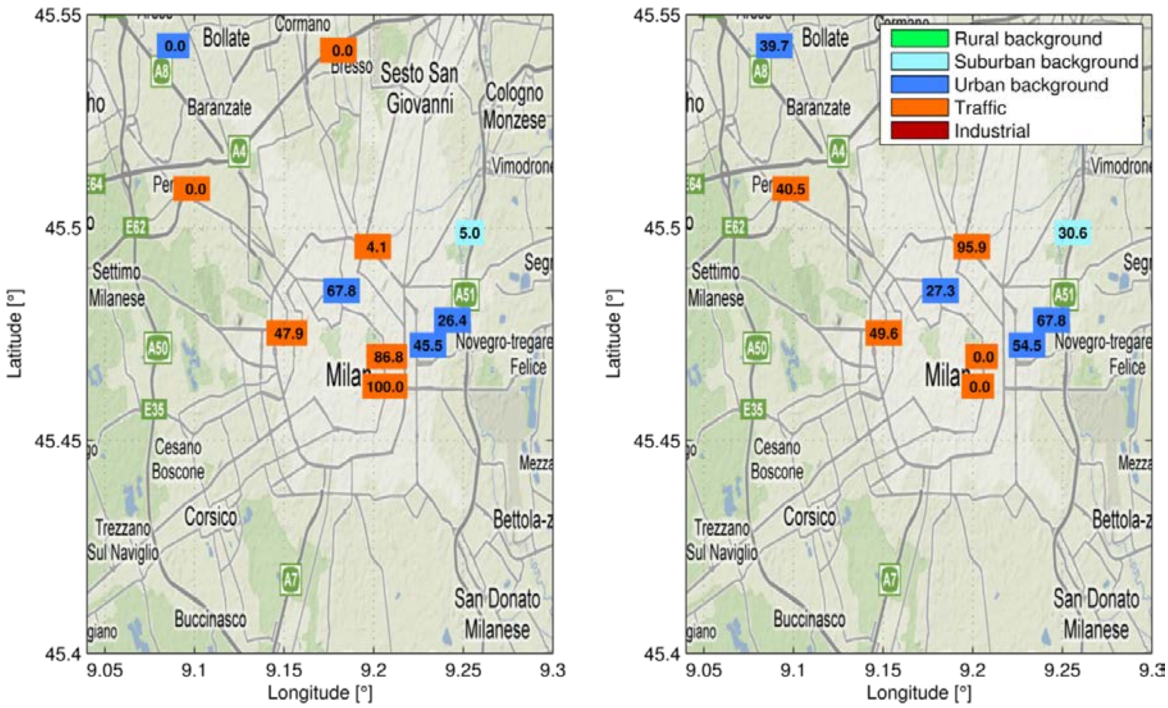


Figure 41: Fraction (in percent) of the land cover classes "Continuous Urban Fabric" (left) and "Discontinuous Urban Fabric" (right) within a 1 km x 1 km box around each NO_2 monitoring station in Milan. The coloured boxes further indicate the station classification as given by the AirBase metadata.

Figure 41 provides information about the station environment for each station in Milan, expressed as the percentage of the two urban land cover classes included in the CORINE land cover dataset, namely CLC111 (Continuous Urban Fabric) and CLC112 (Discontinuous Urban Fabric). All of the inner-city urban background and traffic stations have very high fractions of urban fabric (both continuous and discontinuous depending on station). Several stations, especially in the northern part of the city have been classified as 8 despite having no continuous urban fabric in their vicinity and only a relatively small fraction of discontinuous urban fabric.

5.5.2 PM_{10}

Available data sets for Milan do not include any time series of hourly PM_{10} concentrations. Consequently, no analysis of station classification could be performed yet for PM_{10}.

5.6 Ploiesti

Available data sets for Ploiesti do not include enough data to produce reliable classes (see Section 2.3).

5.7 Prague

5.7.1 NO_2

No rural background station exists in the air quality zone of Prague. As for suburban background stations they are all classified as 3 or 4 and thus very close to the European median of 4. Only two urban background stations exist and with classes of 4 they are both slightly below the European median of 5. Most of the air quality stations in Prague are traffic stations, and their classifications mostly range from 6 to 7 with only two stations classified as 8 and 10. Finally, two industrial stations exists in the area, and with values of 6 and 9 they are both significantly above the European median for industrial stations of 3.

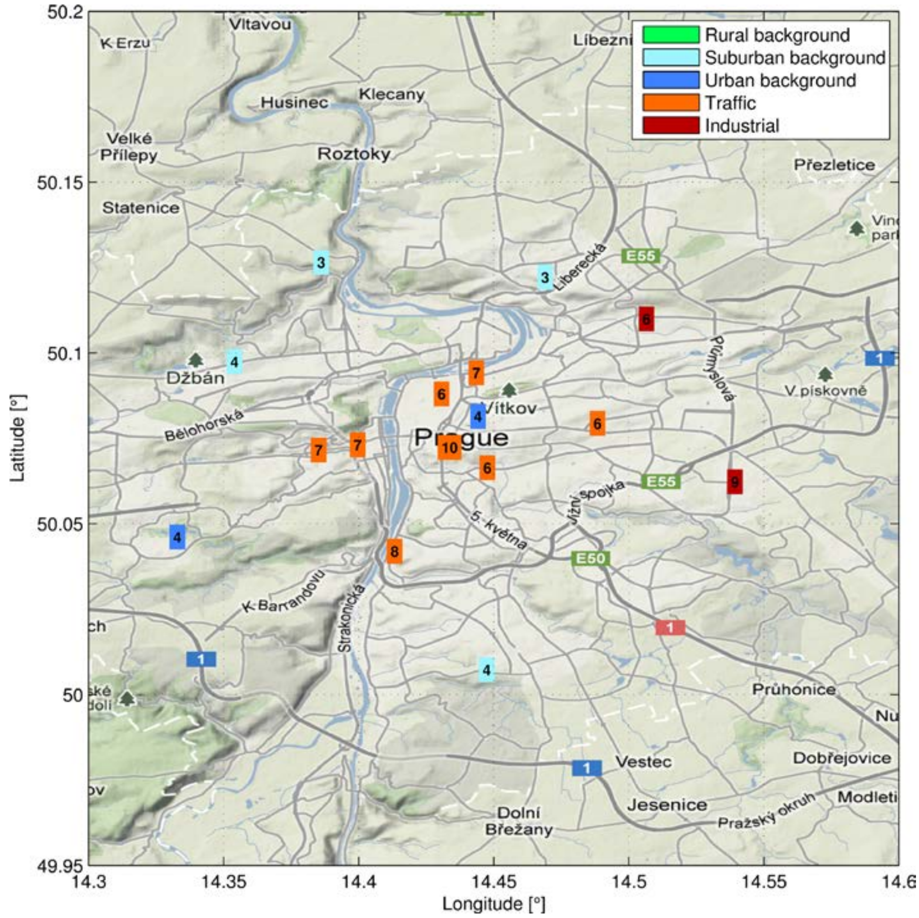


Figure 42: Spatial comparison of the two classification systems in the air quality zone of Prague. The traditional AirBase metadata classification is shown as coloured boxes whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

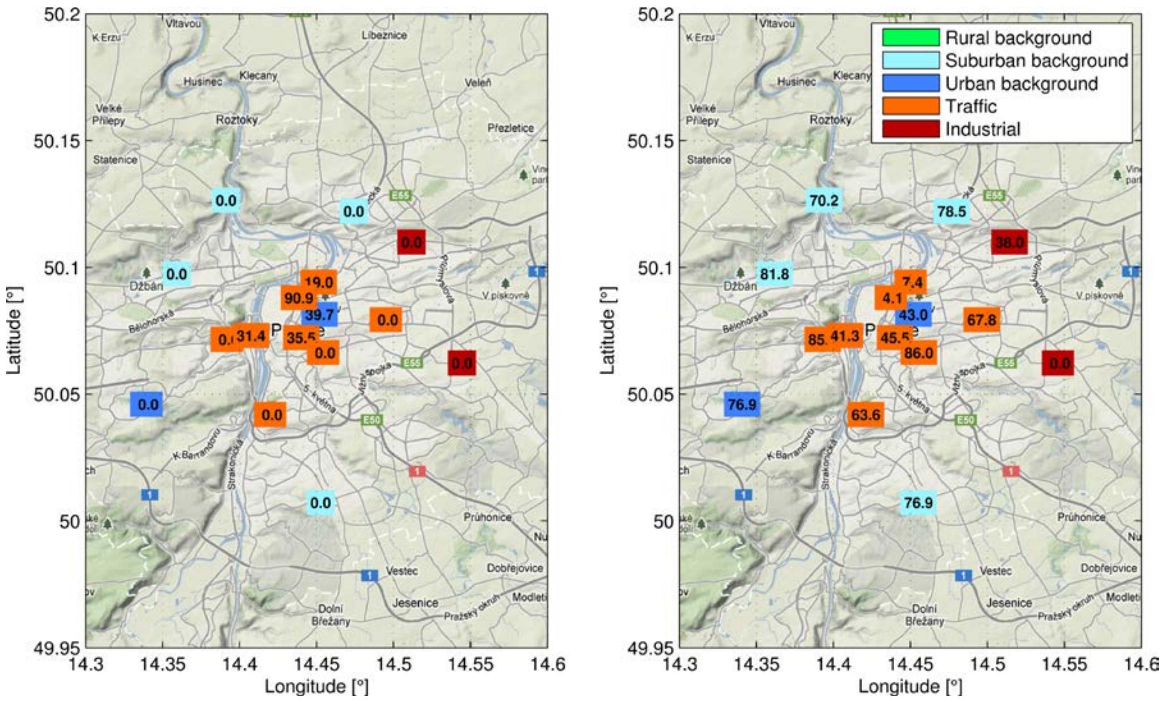


Figure 43: Fraction (in percent) of the land cover classes "Continuous Urban Fabric" (left) and "Discontinuous Urban Fabric" (right) within a 1 km x 1 km box around each NO_2 monitoring station in Prague. The coloured boxes further indicate the station classification as given by the AirBase metadata.

Figure 43 provides information on the station environment for each station in Prague, expressed as the percentage of the two urban land cover classes included in the CORINE land cover dataset, namely CLC111 (Continuous Urban Fabric) and CLC112 (Discontinuous Urban Fabric). As previously seen for other cities and as would be expected, most of the traffic and urban background stations that were classified as 6 or higher have either a large fraction of continuous urban fabric or a large fraction of discontinuous urban fabric. Surprisingly, the industrial station in the very east of the city has been classified as 9 despite having neither any continuous or discontinuous fabric in its vicinity. This is a very good example to show the limitations of land cover indicators to characterize station behaviour.

5.7.2 PM_{10}

PM_{10} classes are characterized by their rather high variability, ranging from 2 (urban background site) to 10 (urban industrial site). Unlike NO_2 classes, there is no clear difference according to AirBase metadata. Classes assigned to urban and suburban background stations range from 4 or 5, except one urban background station which is classified as 2. This station is mainly surrounded by discontinuous urban fabric (Figure 45) and population density (Figure 60, Annex IV) is not different from the rest of the city. Urban traffic stations are classified between 4 and 9. Surprisingly the station classified as 9 has lower fractions of continuous and discontinuous urban fabric than other traffic sites. The industrial station is classified as 10 though it is not surrounded by urban fabric and population density is almost zero. This behaviour could be explained by the impact of industrial emissions or by the influence of traffic close to the facilities. A more thorough analysis of PM_{10} classification is necessary to better interpret those results (number of available data per time series, emission inventories, Google Earth...).

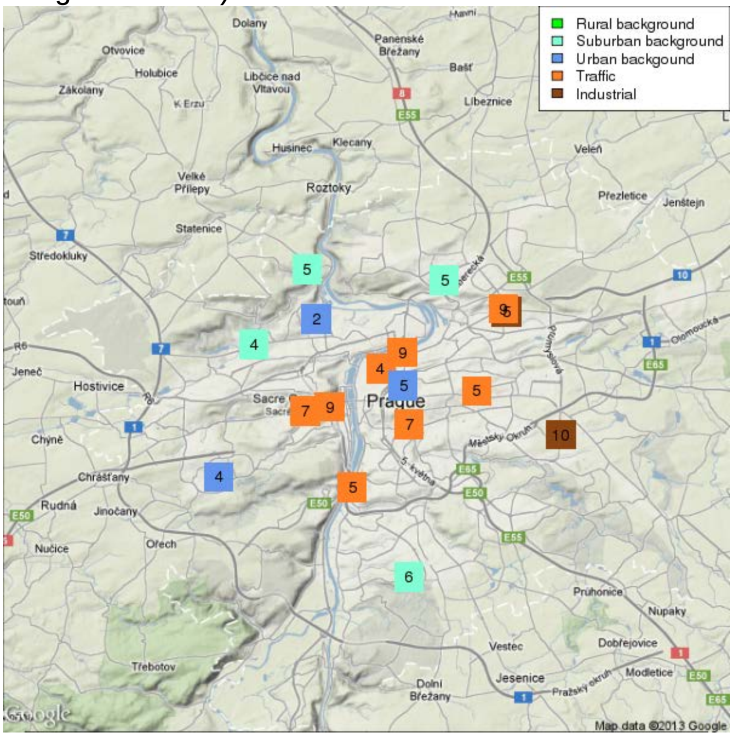


Figure 44: Spatial comparison of the two classification systems in the air quality zone of Prague. The traditional AirBase metadata classification is represented by colours whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

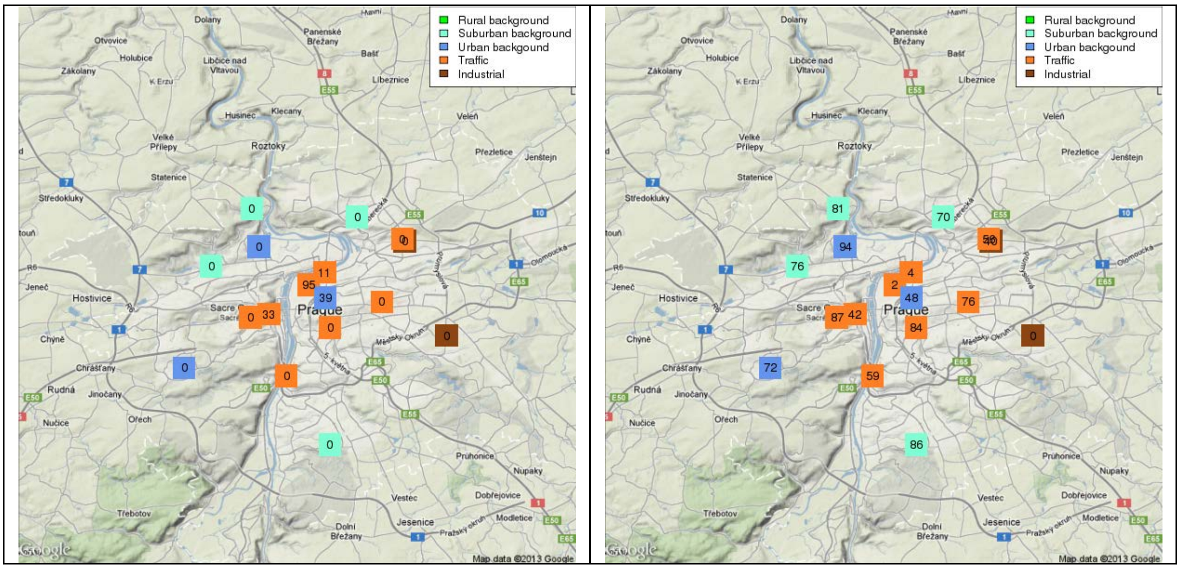


Figure 45: Fraction (in percent) of the land cover classes "Continuous Urban Fabric" (left) and "Discontinuous Urban Fabric" (right) within a circle of 1-km diameter around each PM_{10} monitoring station in Prague. The colours indicate the station classification as given by the AirBase metadata.

5.8 Vienna

5.8.1 NO_2

Quite a few air quality stations are situated in the air quality zone of Vienna. Whereas there is no rural station, several suburban background stations exist and their classification ranges from 2 to 3. As for urban background stations, one of the two matches the European median of 5, whereas the other one is classified slightly below this value. A large number of traffic stations exist in Vienna and with a median class of 7 they are classified quite significantly below the European-scale median of 9. One industrial station is also present and with a value of 5 it is classified slightly above the European median for industrial sites of 3.

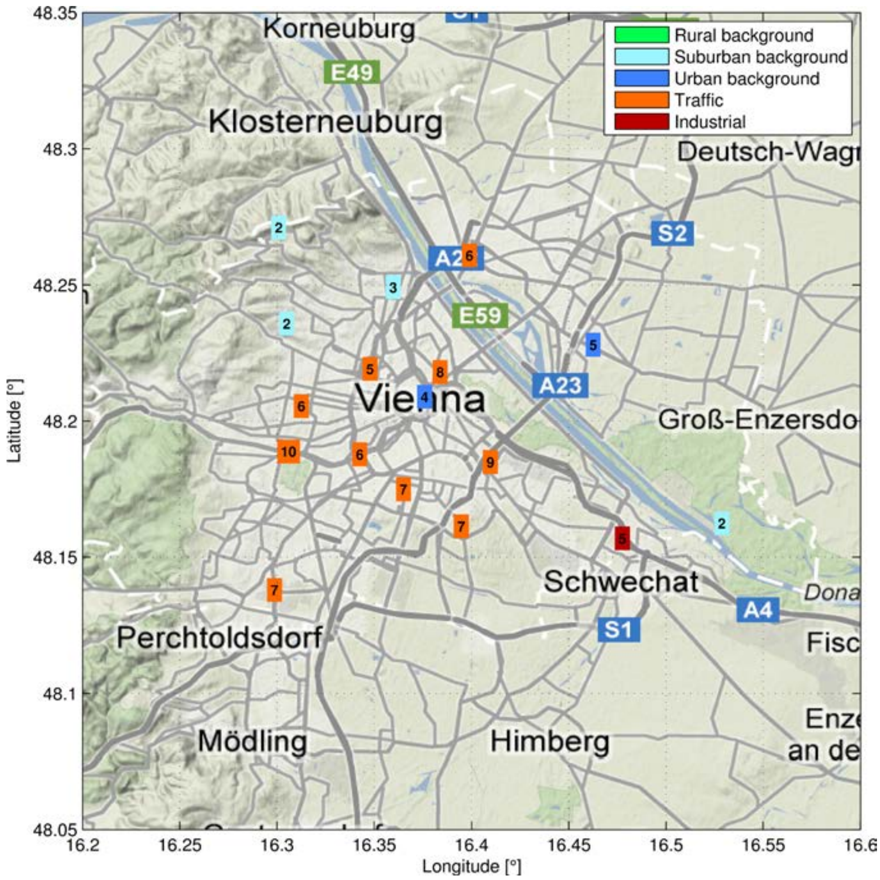


Figure 46: Spatial comparison of the two classification systems in the air quality zone of Vienna. The traditional AirBase metadata classification is shown as coloured boxes whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

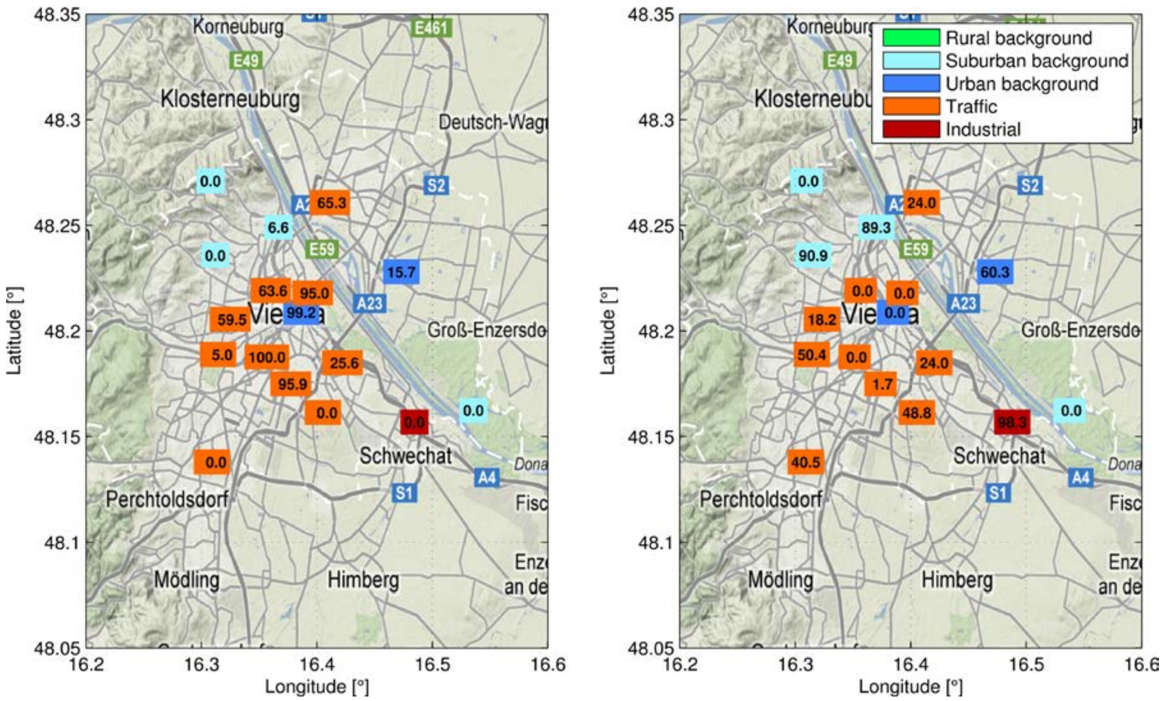


Figure 47: Fraction (in percent) of the land cover classes "Continuous Urban Fabric" (left) and "Discontinuous Urban Fabric" (right) within a 1 km x 1 km box around each station in Vienna. The coloured boxes further indicate the station classification as given by the AirBase metadata.

Figure 47 provides information on the station environment for each station in Vienna, expressed as the percentage of the two urban land cover classes included in the CORINE land cover dataset, namely CLC111 (Continuous Urban Fabric) and CLC112 (Discontinuous Urban Fabric). In Vienna, most of the inner-city traffic stations were classified as 6 or higher. They mostly have very high fractions of continuous urban fabric. Surprisingly, the urban background station in the central city has a fraction of nearly 100% continuous urban fraction in its vicinity, yet the Joly and Peuch (2012) algorithm classified it only as a class 4 station. This is probably due to the station being located in a reduced-traffic and/or in a pedestrian zone of the inner city.

5.8.2 PM_{10}

Few stations could be classified for PM_{10}: one suburban station, classified as 5, and three traffic stations classified between 8 and 10. Those values are slightly above the European medians for urban background and traffic sites (4 and 8 respectively). Land cover and population density do not explain the differences between the traffic sites.

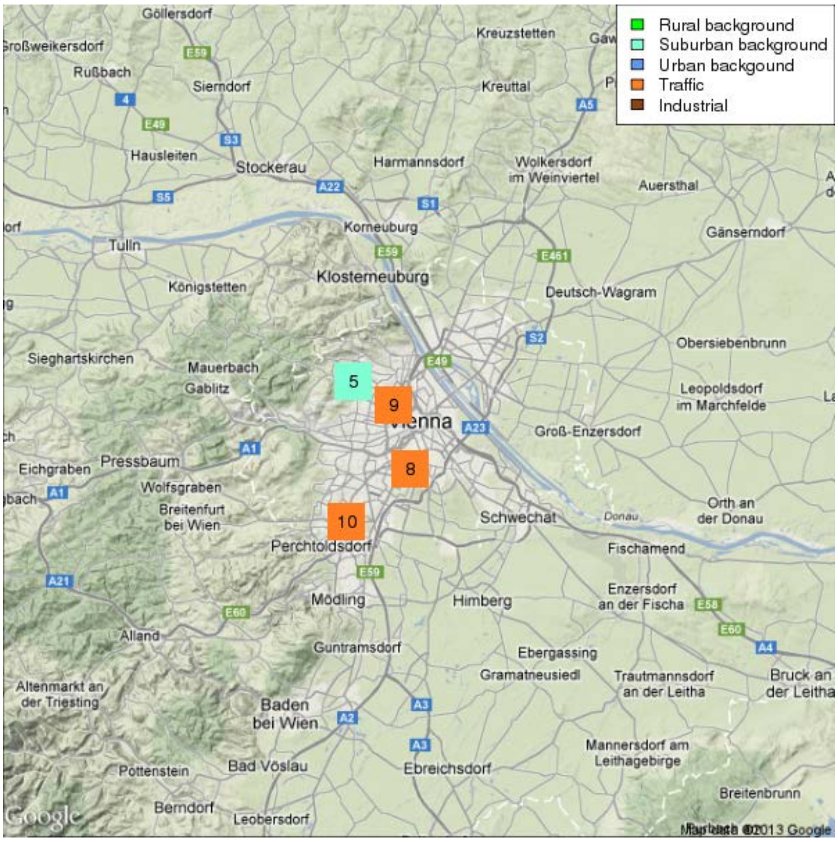


Figure 48: Spatial comparison of the two classification systems in the air quality zone of Vienna. The traditional AirBase metadata classification is represented by colours whereas the Joly and Peuch (2012) classification is given as a number from 1 to 10.

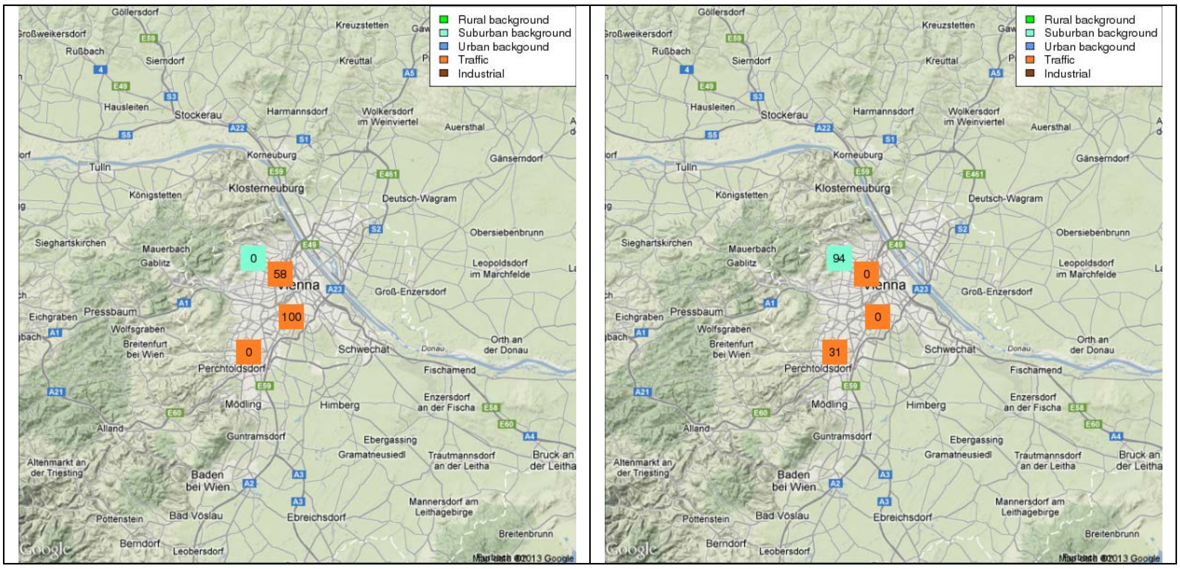


Figure 49: Fraction (in percent) of the land cover classes "Continuous Urban Fabric" (left) and "Discontinuous Urban Fabric" (right) within a circle of 1-km diameter around each PM_{10} monitoring station in Vienna. The colours indicate the station classification as given by the AirBase metadata.

5.9 Comparison between cities

This section illustrates the inter-city variability of NO_2 and PM_{10} classes according to current AirBase classification. This variability is higher for urban and suburban background stations and in general for PM_{10} (Figure 50, Figure 51, Figure 61, Figure 62). Also, the position of each city with respect to the European mean and median class is dependent on the pollutant and the type of station. However, some cities exhibit stable results: Madrid and Malmö are always above and below the European mean/median respectively whereas Berlin is well representative of this European average behaviour

The whole set of graphs is provided in Annex V.

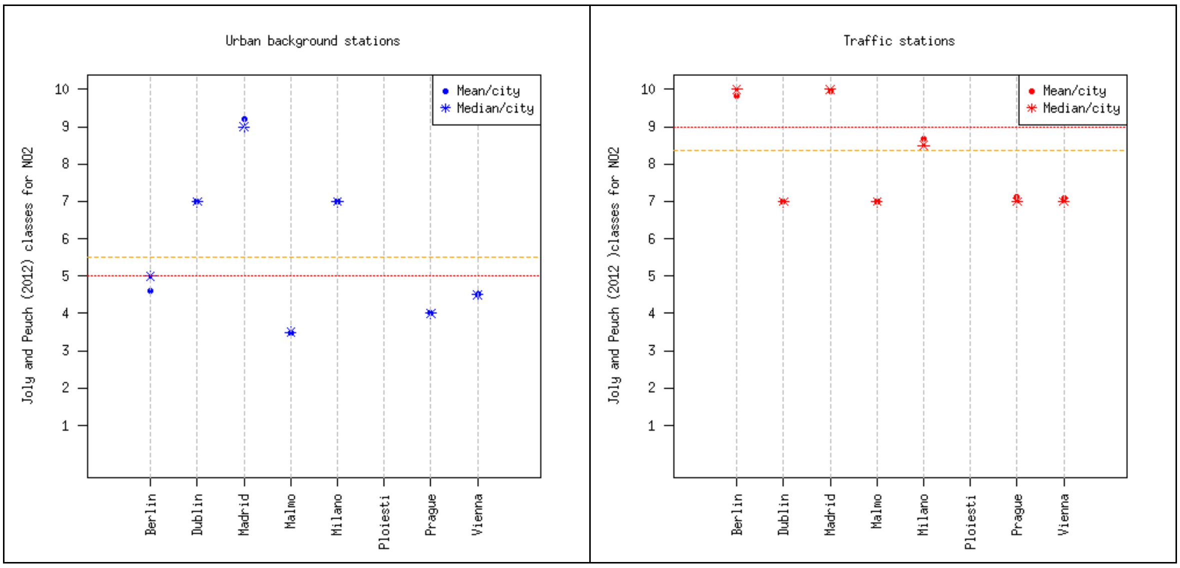


Figure 50: NO_2. Mean and median classes per city for stations classified as urban background (left) and urban traffic (right) according to AirBase metadata. Orange and red horizontal lines: European mean and median classes.

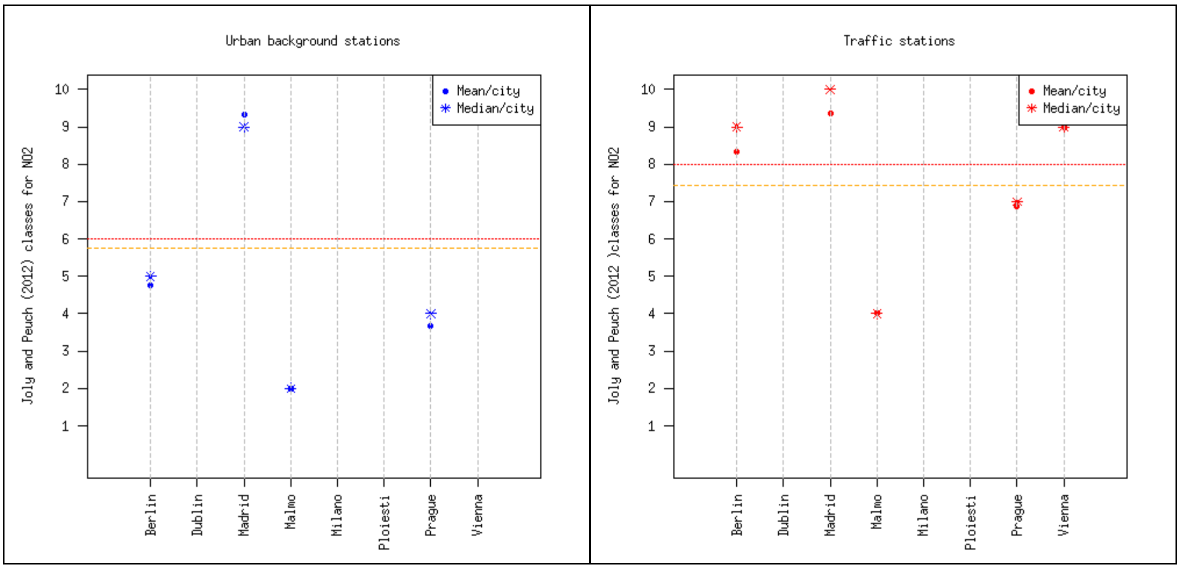


Figure 51: PM_{10}. Mean and median classes per city for stations classified as urban background (left) and urban traffic (right) according to AirBase metadata. Orange and red horizontal lines: European mean and median classes.

6. How to use the new classification

Considering the previous analyses made on the European and local scales, this section explains how the new classification could be used in future as an additional material for data providers and data users.

Use the classification as a tool to define or revise station typology in AirBase

We have seen that Joly and Peuch (2012) methodology succeeds in putting a gradual classification of the typology of the station through an analysis of historical time series. Class 1 is more representative of a rural background station whereas class 10 is representative of a traffic station. This gradual classification is obtained by projecting time series - which are summarized through a combination of eight indicators - on a reference axis, the Fisher axis, calculated from a given training data set. When a new time series is available, it is then interesting to project it on this axis and to make an analysis of the new data in relation with the training data set. Such process will concern newly implemented stations or stations which have not been classified yet in AirBase database or newly measured pollutants for existing stations. The benefit of this methodology will be to help data providers to characterize the typology of the new stations or new measurements in connection with the existing database.

In the same way, the methodology can be helpful for reviewing and where necessary updating the station typology already reported in AirBase in an objective and consistent manner. It should also be useful to adapt the existing typology to each measured pollutant in compliance with the implementing provisions on reporting to be applied from 2014.

Follow the classification in time

We mentioned (§3.2) that the Fisher axis could be defined from a selected data set and be used as reference axis for the classification. An interesting application of this in a long-term perspective is to follow the classification of a station over time. With the reasonable assumption that the reference Fisher axis remains valid for several years, it is then possible to regularly project updated time series on it (summarized by the eight indicators) and see how the classification evolves along the years. The way of updating the indicators still needs to be established; a sliding window could be applied (e.g. 2003-2012, 2004-2013...).
Such information could help data providers to have a better follow-up of the reported information and data users to remove some stations from or include stations into their analysis.

Analyse a station in relation to others

The AirBase classification is a qualitative classification whereas the new classification is based on numerical criteria. The resulting classes can be considered as a measure of the quality of being a background station or a traffic station. It thus makes it easier to draw a comparison between sets of stations located in different places, cities or countries. Applications are numerous. For example, this new information is an added value to understand differences between station types in two cities and compare their air quality indexes.

Use the classification to understand specific stations (coastal...)

As pointed out in section 4.5.2, some stations exhibit particular behaviour for O_3 or NO_2, leading to discrepancies between classifications. Most of them have been identified as located close to mountain or coastal regions but the physical process has not been fully explained yet. The new classification could be used as an additional variable in a more complex analysis to better characterize and understand those specific stations.

Use the classification to make a selection of stations for modelling issues

With its 1-to-10 scale, the new classification is a good indicator of how well monitoring stations are representative of background pollution for a given substance and to which extent concentrations are submitted to local influences. Depending on the modelling domain and the resolution of the model, it can be used to select adequate stations for the purpose of model evaluation or data assimilation. The histograms and maps such as those presented in section 4 can be helpful tools to do this selection.

7. Conclusion

The purpose of this study was to apply Joly and Peuch (2012) methodology to AirBase stations and to analyse the resulting classes with special focus on the cities involved in the Air Implementation Pilot project. The classification depends on the pollutant. For each pollutant, stations are classified into 10 classes using eight indicators related to the temporal variability of concentrations. A linear discriminant analysis (LDA) is applied using AirBase metadata as prior knowledge to determine the function which best separates rural background sites from urban background and traffic stations (the Fisher's linear discriminant).

In section 3, robustness tests reveal that the classification is not very sensitive to an update of Airbase version or to an addition of stations to the training data set (French database for PM_{10} in this study). The differences are rather low, usually not higher than 1 or 2 classes. They are mainly due to a small change in the Fisher axis defined by the LDA and in the bounds delimiting the ten classes. The classification seems to be more sensitive for monitoring stations presenting limited historical data series and low or irregular data capture. Considering those results a procedure for defining a reference Fisher axis from a selection of long-run and well characterized time series is proposed. This reference axis (calculated for each pollutant) could then be used to classify all AirBase stations for which enough data are available, and to follow this classification over time.

In section 4, classifications for O_3, PM_{10} and NO_2 are analyzed with regard to the Airbase classification. For each pollutant, the methodology succeeds in separating stations characterized by a background behaviour from stations showing a more traffic-related behaviour. The classification into 10 classes offers the possibility of making a quantitative comparison of the station typologies based on the pollution data themselves and for each pollutant separately. This represents complementary information in addition to AirBase metadata which qualify the stations according to their environment. Some discrepancies between Joly and Peuch (2012) and AirBase classifications exist however and can be mapped for a spatial analysis. This leads to identify specific stations which need to be investigated more carefully by the data providers. Local discrepancies between the classifications obtained for different pollutants are highlighted as well. In particular a few stations located in complex regions like coastal and mountainous areas appear to be classified differently for O_3 and NO_2. Spatial representations of the classes also point out some disparities between countries.
Auxiliary data such as land cover and population density have been considered as well to support this analysis. On average, a class increase from 1 to 5 is associated with an increase of urban density (higher population density, higher fraction of urban fabric) therefore of emission sources. From class 5 to 10 the relationship between both variables is less clear, displaying a different profile according to the pollutant. Additional

databases like fine resolution emission inventories could be helpful to bring this analysis forward on a smaller scale.

In Section 5 the study focuses on a smaller scale with the cities involved in the Air Implementation Pilot project. The analysis of the new classes in relation to AirBase metadata brings out different types of cities like the following ones:

- those for which the new classes match AirBase classification quite well with a clear distinction between suburban background, urban background and urban traffic sites (e.g. Berlin);
- those with high urban density and an overall strong influence of traffic: all background and traffic stations are classified between 8 and 10 (e.g. Madrid);
- those for which the "traffic behaviour" of monitored data is less pronounced and with no marked difference between urban background and urban traffic sites (e.g. Dublin, Milan);
- those characterized by low population density and a behaviour typical for background even if differences are observed between urban background and urban traffic sites (e.g. Malmö).

NO_2 and PM_{10} classifications are usually consistent but display some differences that could be due to emission sources. Land cover (fraction of urban fabric) and population density are useful data to interpret the results but they are not sufficient. For a better understanding of the classification at the city level, other variables and parameters should be considered like high resolution emission inventories, topography (city or street configuration), aerial views,...

On the whole, the new classification appears to be an added value for qualifying a monitoring station. It does not replace the existing and widely used information about the type of area and type of station but gives the possibility of making an objective, quantitative and pollutant specific comparison of the stations, based on the temporal behaviour of concentrations. It is therefore a complementary approach to AirBase classification that should help data providers and users to further understand and interpret air quality data. A practical application of this classification in the framework of GMES/MACC2 project was to redefine the set of stations intended for model validation and data assimilation. The new classification will be used to discard stations that are too specific with respect to the considered models resolution or to gather stations in homogeneous sets according to statistical properties. In a larger context, both in a short and long-term perspective, this classification may contribute to a more in-depth assessment of monitoring sites.

8. Acknowledgements

The classification methodology was applied using the programmes developed and provided by Météo-France. Mathieu Joly (Météo-France) is acknowledged for his very helpful support.
Also the colleagues in charge of the monitoring networks in the Air Implementation Pilot cities are thanked for their beneficial feedback.

9. References

Decision 2011/850/EU. Commission Implementing Decision of 12.12.2011 laying down rules for Directive 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.

European Commission (DG ENV), 2012. Guidance on the Commission Implementing Decision laying down rules for Directive 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. Draft version of 19 October 2012.

EoI, 2002. Guidance on the Annexes to Decision 97/101/EC on Exchange of Information as revised by Decision 2001/752/EC.

Janssens, S., Dumont, G., Fierens, F., Deutsch, F., Maiheu, B., Celis, D., Trimpeneers, E., Mensink, C., 2011. Land use to characterize spatial representativeness of air quality monitoring stations and its relevance for model validation. Atmospheric Environment, Vol. 59, Issue 1, pages 492-500

Joly, M., Peuch, V.-H, 2012. Objective classification of air quality monitoring sites over Europe. Atmospheric Environment, 47, 111-123.

Umweltbundesamt (Spangl W., Schneider J., Moosmann L., Nagl C.), 2007. Representativeness and classification of air quality monitoring stations (http://www.umweltbundesamt.at/aktuell/publikationen/publikationssuche/publikationsdetail/?pub_id=1684).

Annex I: Station classification for NO_2

Study of the discrepancies between Joly and Peuch (2012) classification and the classification based on AirBase metadata

Table 4: Station codes of the rural background outliers, i.e. rural background stations that were classified as class 4 or higher by the Joly and Peuch (2012) algorithm.

CH0033A	FR33120	FR02005	IT1174A
DEBW004	FR20045	FR01017	IT1178A
DEBY017	FR07002	FR18019	IT1930A
DENW068	FR28011	FR02037	IT1915A
ES1418A	FR11026	FR07001	LI0001A
ES1811A	FR05070	FR27001	PT02002
ES1851A	FR30016	FR11030	
ES1853A	FR21019	FR35006	
FR41003	FR02012	GB0885A	
FR18034	FR08005	IT1233A	

Table 5: Station codes of the urban traffic outliers, i.e. urban traffic stations that were classified as class 4 or lower by the Joly and Peuch (2012) algorithm

CZ0PPLA	ES1867A	HU0031A
DEST057	ES1922A	IT0508A
DEMV002	ES1955A	MK0034A
ES0873A	FI00397	SK0023A
ES1745A	GB0956A	SK0019A

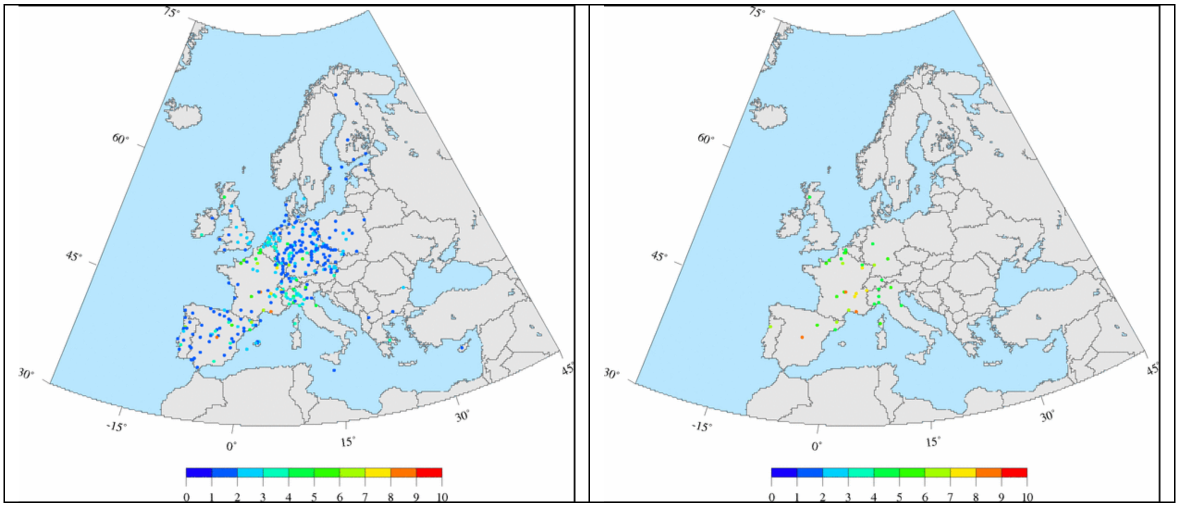


Figure 52: Overview map of NO_2 monitoring stations. Left: stations classified as "rural background" in AirBase with a colour legend indicating the Joly and Peuch (2012) classification.
Right: Outliers in the new NO_2 classification of rural background stations. These are stations that are classified as rural background within the Airbase database but were assigned classes of 4 or higher by the Joly and Peuch (2012) algorithm.

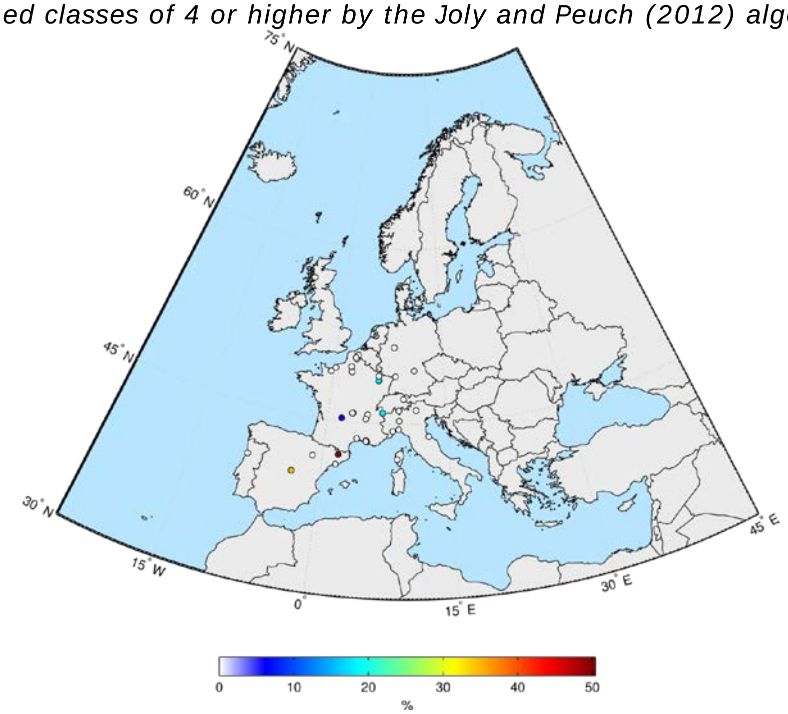


Figure 53: Fraction of the CORINE land cover class CLC111 (Continuous Urban Fabric) for the previously identified "outliers" of the rural background class.

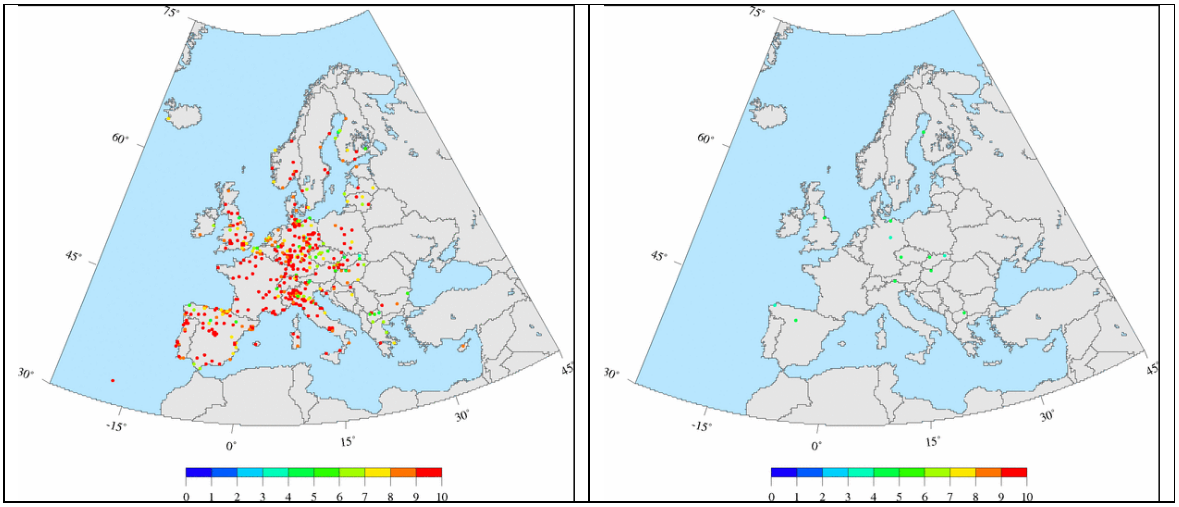


Figure 54: Overview map of NO_2 monitoring stations. Left: stations classified as “urban traffic” in AirBase with a colour legend indicating the Joly and Peuch (2012) classification.
Right: Outliers in the new NO_2 classification of urban traffic stations. These are stations that are classified as urban traffic within the AirBase database but were assigned classes of 4 or lower by the Joly and Peuch (2012) algorithm. Three of the outlier stations in Spain that were listed in Table 5 are located in the Canary Islands and are not shown on this map.

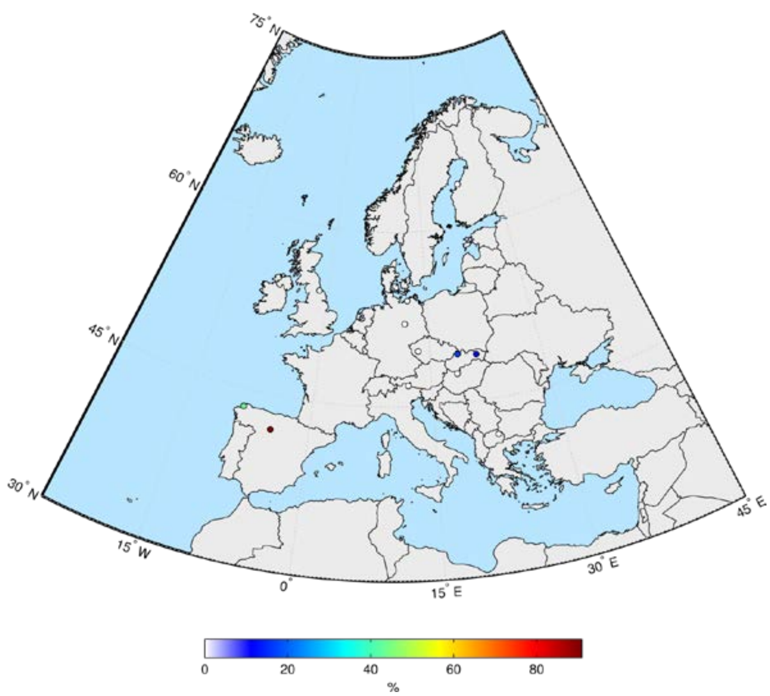


Figure 55: Fraction of the CORINE land cover class CLC111 (Continuous Urban Fabric) for the previously identified "outliers" of the urban traffic class.

Annex II: Station classification for PM_{10}

Study of the discrepancies between Joly and Peuch (2012) classification and the classification based on AirBase metadata

Table 6: Station codes of the rural background outliers, i.e. rural background stations that were classified as class 4 or higher by the Joly and Peuch (2012) algorithm.

BETN073	BETN113	BETN121	BETN132
CZ0TSTD	CZ0TVER	CZ0ZSNV	DEBW004
ES1811A	FR04066	FR04158	ES1418A
FR11030	FR14008	FR18019	FR18039
FR20045	FR21019	FR23124	FR23177
FR33120	PL0062A	PL0094A	PL0243A
PL0349A	PT02002	SI0033A	SK0007R

Table 7: Station codes of the urban traffic outliers, i.e. urban traffic stations that were classified as class 4 or lower by the Joly and Peuch (2012) algorithm

CZ0AREP	CZ0BBNF	DEBB060	DEMV002
DEMV005	DENI073	DENI075	DENI076
DENW177	DERP008	DEST103	ES0873A
ES1132A	ES1137A	ES1239A	ES1296A
ES1421A	ES1482A	ES1741A	ES1742A
ES1745A	ES1922A	ES1926A	ES1953A
ES1955A	FI00074	FI00084	FI00404
FI00446	FI00464	FI00465	FI00483
FI00506	FI00546	FI00549	FI00558
FI00570	FI00574	FI00602	FR10002
FR20002	FR20003	GB0896A	GB0956A
LT00011	PT01046	SE0058A	SE0067A

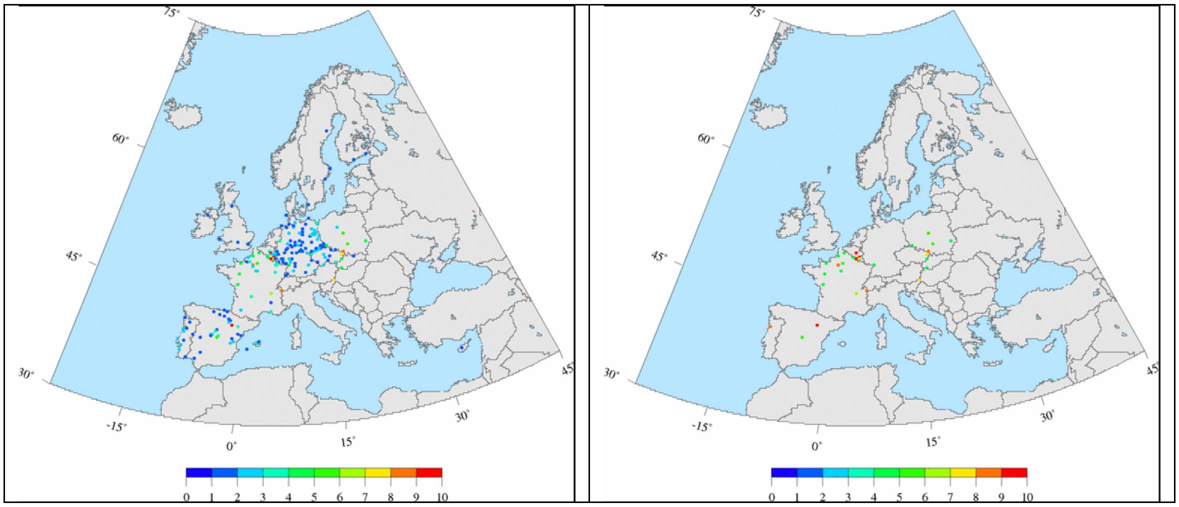


Figure 56: Overview map of PM_{10} monitoring stations. Left: stations classified as "rural background" in AirBase with a colour legend indicating the Joly and Peuch (2012) classification.
Right: Outliers in the new PM_{10} classification of rural background stations. These are stations that are classified as rural background within the Airbase database but were assigned classes of 4 or higher by the Joly and Peuch (2012) algorithm.

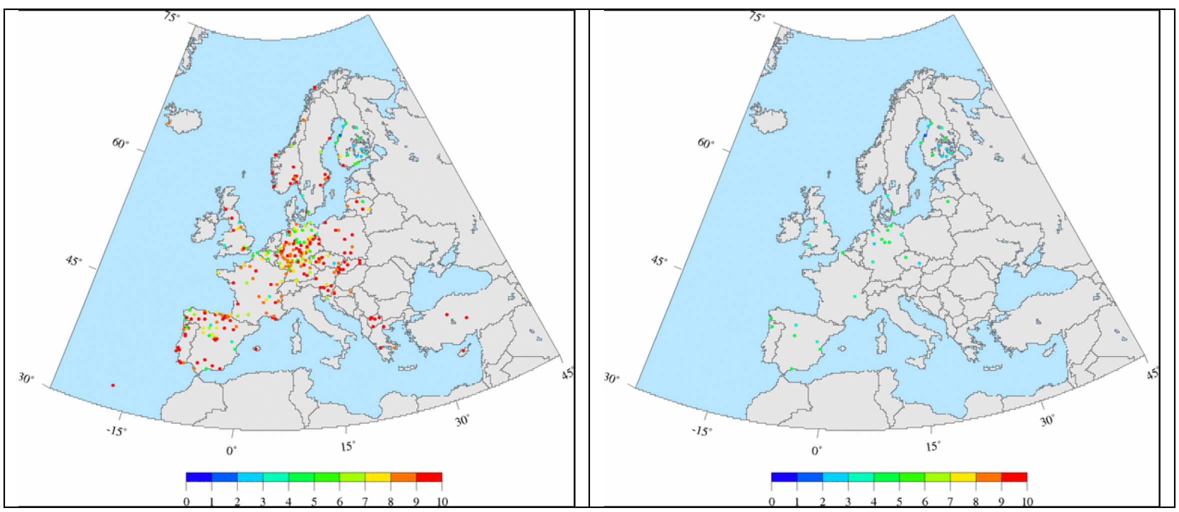


Figure 57: Overview map of PM_{10} monitoring stations. Left: stations classified as "urban traffic" in AirBase with a colour legend indicating the Joly and Peuch (2012) classification.
Right: Outliers in the new PM_{10} classification of urban traffic stations. These are stations that are classified as urban traffic within the AirBase database but were assigned classes of 4 or lower by the Joly and Peuch (2012) algorithm.

Annex III: Station classification for O_3

Study of the discrepancies between Joly and Peuch (2012) classification and the classification based on AirBase metadata

Table 8: Station codes of the rural background outliers, i.e. rural background stations that were classified as class 8 or higher by the Joly and Peuch (2012) algorithm.

AT31496	BETN016	CH0024A	CH0033A
DEHE048	ES0296A	ES1222A	ES1248A
ES1588A	ES1605A	ES1607A	ES1774A
ES1805A	FR03031	FR03048	FR03083
FR11015	FR13012	FR15007	FR17006
FR18019	FR20045	FR31008	FR33120
FR35006	FR41004	GR0110R	IT1121A
IT1174A	IT1233A	IT1418A	IT1464A
IT1519A	IT1736A	IT1865A	IT1927A
IT1982A	NL00235	PL0030A	PT02002

Table 9: Station codes of the urban traffic outliers, i.e. urban traffic stations that were classified as class 4 or lower by the Joly and Peuch (2012) algorithm

AT10001	CZ0CTAB	CZ0HHKB	CZ0PPLA
DEBE044	DEST057	DETH011	DETH020
DETH021	DETH032	ES1184A	ES1429A
ES1479A	ES1613A	ES1643A	ES1745A
ES1759A	ES1849A	ES1901A	IT1282A
LT00033	LT00041	LV000L1	LV00RZ1
PL0252A	PT03097	PT05006	SE0026A
SE0058A			

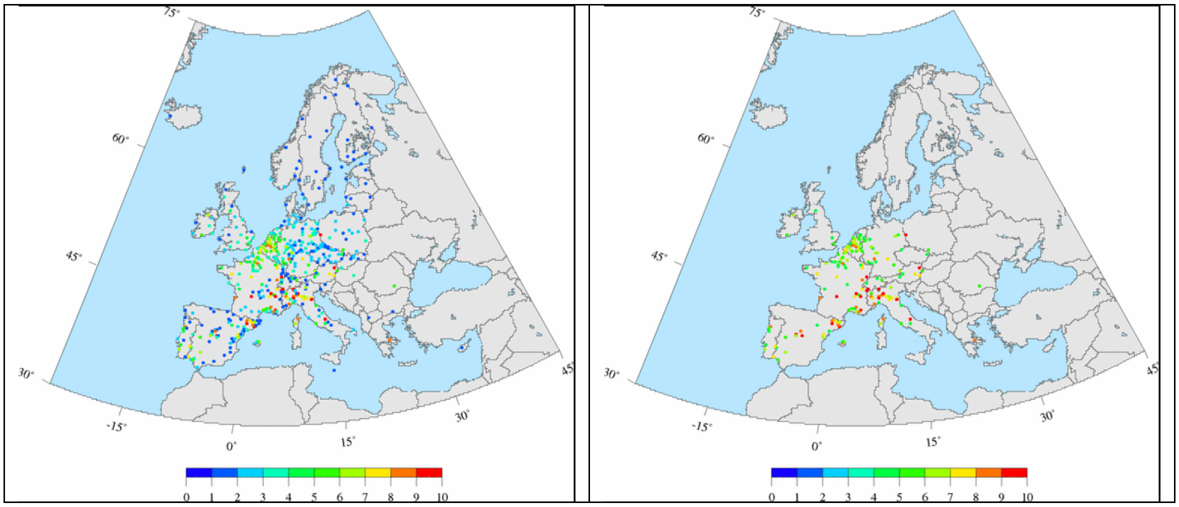


Figure 58: Overview map of O_3 monitoring stations. Left: stations classified as "rural background" in AirBase with a colour legend indicating the Joly and Peuch (2012) classification.
Right: Outliers in the new O_3 classification of rural background stations. These are stations that are classified as rural background within the Airbase database but were assigned classes of 4 or higher by the Joly and Peuch (2012) algorithm.

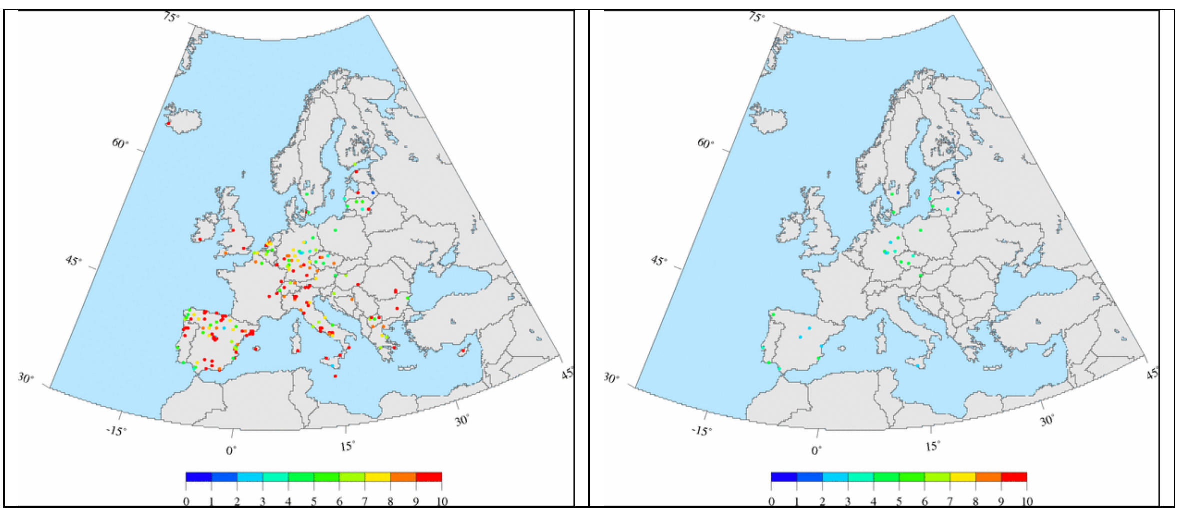
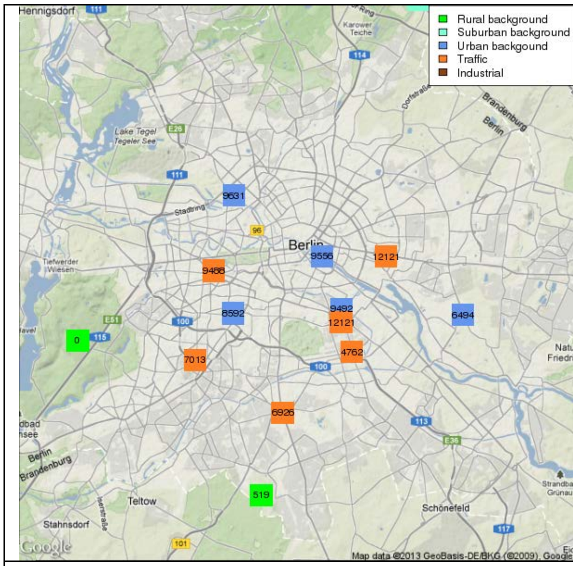
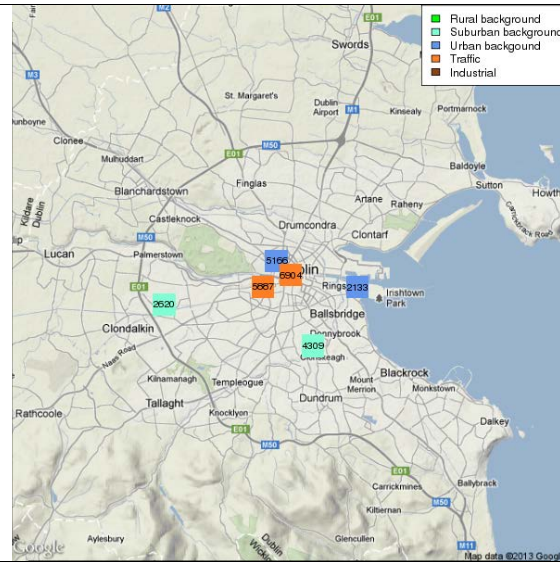


Figure 59: Overview map of O_3 monitoring stations. Left: stations classified as "urban traffic" in AirBase with a colour legend indicating the Joly and Peuch (2012) classification.
Right: Outliers in the new O_3 classification of urban traffic stations. These are stations that are classified as urban traffic within the AirBase database but were assigned classes of 4 or lower by the Joly and Peuch (2012) algorithm.

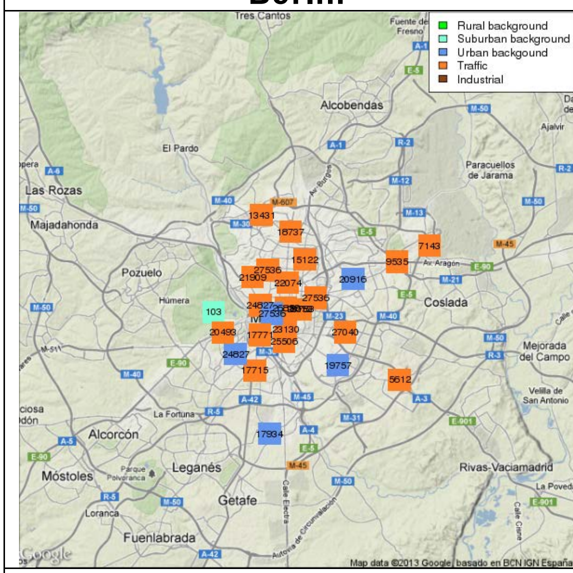
Annex IV: Population density in the Air Pilot cities



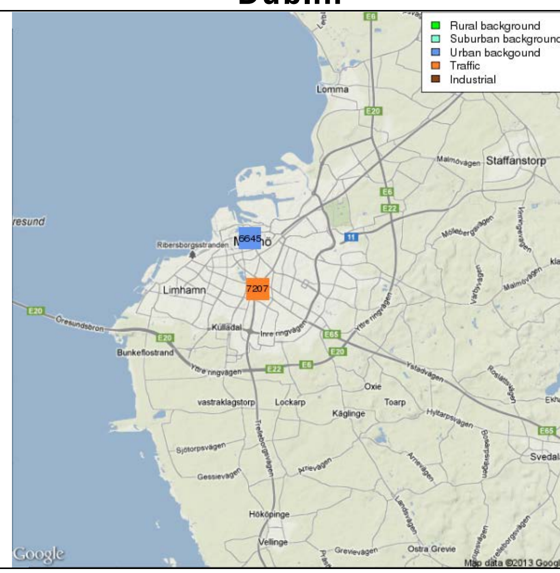
Berlin



Dublin



Madrid



Malmö

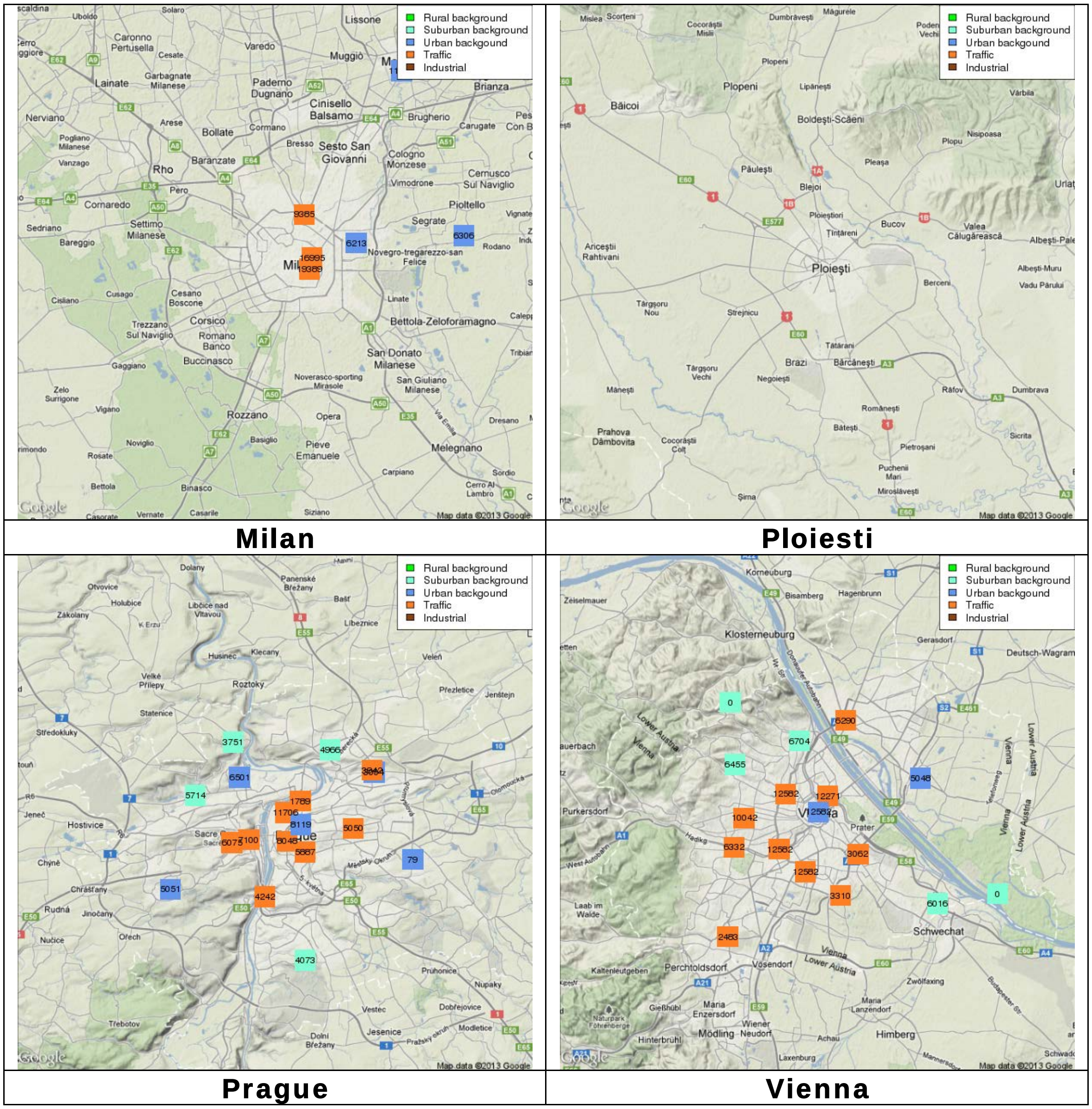
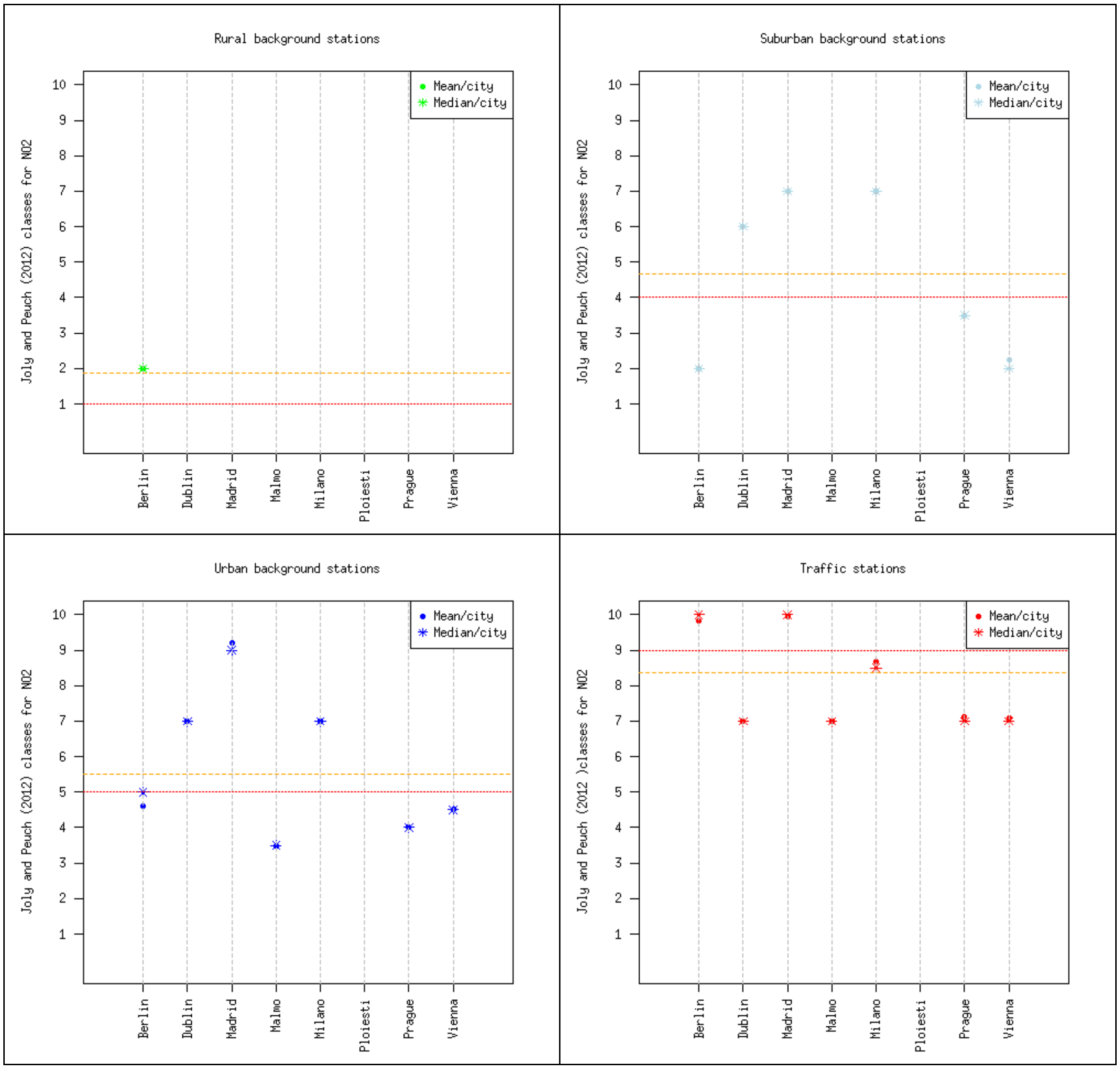


Figure 60: Population density within a diameter of 1 km around the monitoring stations[11] in the Air Pilot cities (all stations for which classes could be calculated for at least one pollutant. As mentioned in the report, not enough data were available to classify Ploiesti stations.). The colours indicate the station classification as given by the AirBase metadata.

[11] calculated from the JRC/EEA population database, http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2

Annex V: Comparison between cities

NO_2



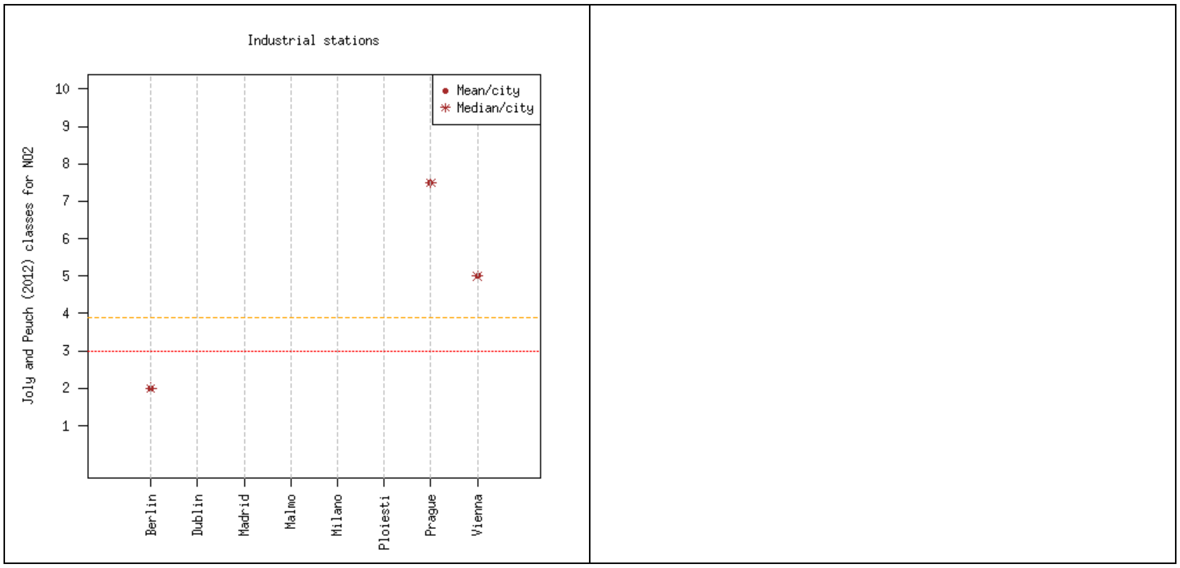
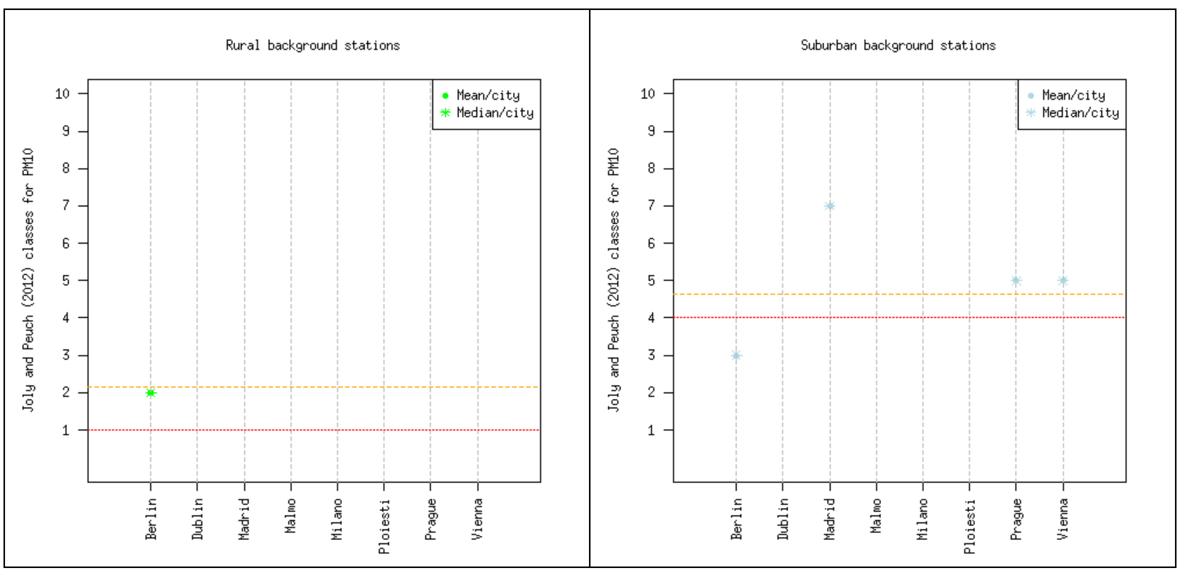


Figure 61: NO_2. Mean and median classes per city for each AirBase category (rural background, suburban background, urban background, traffic and industrial sites). Orange and red horizontal lines: European mean and median classes.

PM_{10}



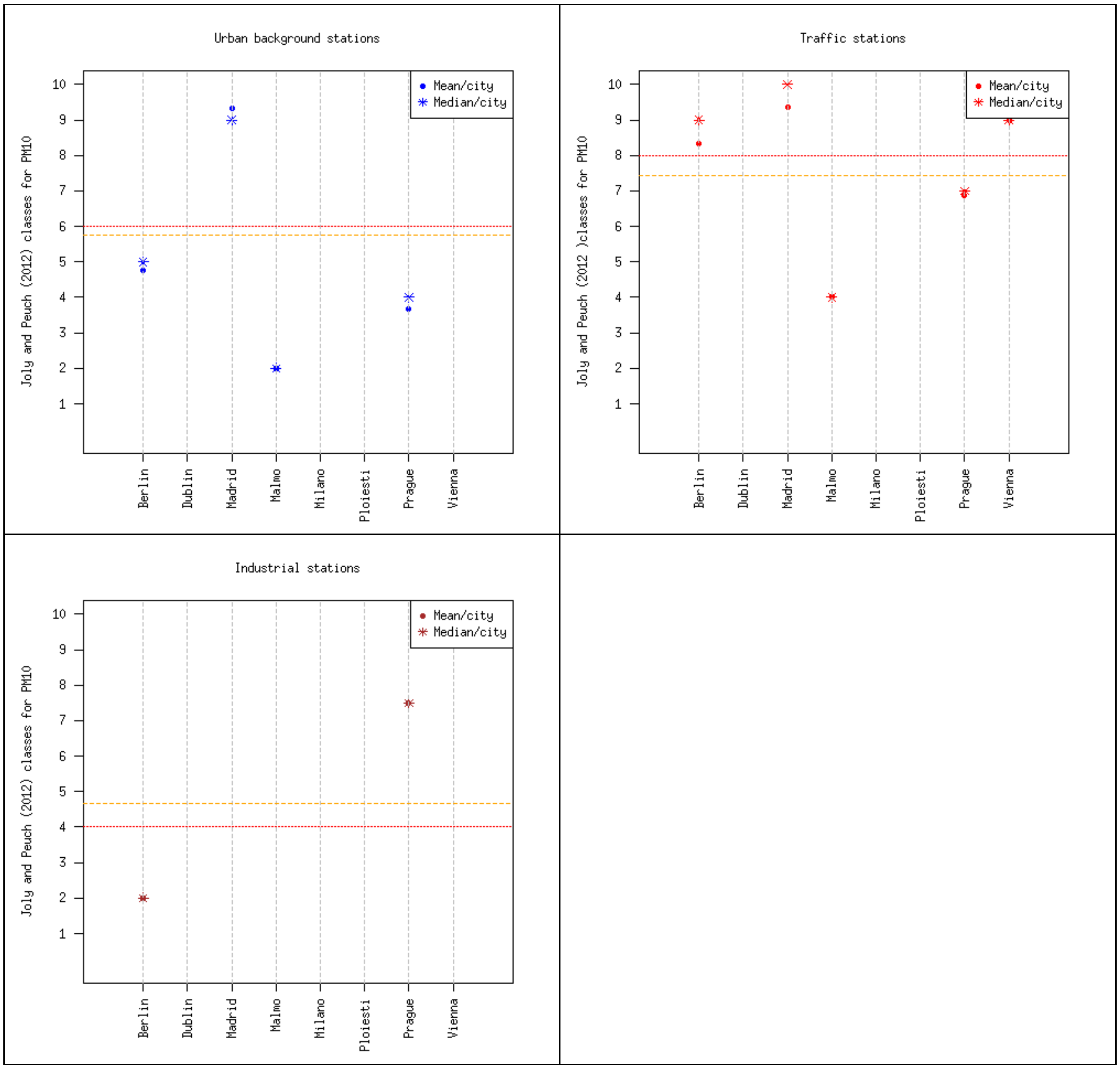


Figure 62: PM_{10}. Mean and median classes per city for each AirBase category (rural background, suburban background, urban background, traffic and industrial sites). Orange and red horizontal lines: European mean and median classes.