Comparison of the residual interpolation optimized for ozone (RIO) estimates and actual $PM_{2.5}$ measurements within the municipality of Utrecht

Author: I. Batoukhtine

Department: MSc Applied Data Science, Utrecht University

Course: Master Thesis

Coordinator: E. van Kesteren

Date: July 2, 2021

ABSTRACT

Current air-quality maps in The Netherlands are produced on a 4x4 km² spatial resolution by a interpolation model, called RIO. This model is based on a low resolution measurement network, which misses air-quality on micro-scale. However, deployment of low-cost sensors has helped in producing detailed air-quality maps in the recent years. One of the entities that implemented a low-cost sensor network, through the citizen science project called Snuffelfiets, is the Province of Utrecht. These sensors measure particulate matter with particles smaller than 2.5 micrometer (PM2.5). This research paper investigated the quality of the observations done by low-cost sensors of the Snuffelfiets project and compared these to the estimated $PM_{2.5}$ concentrations from the RIO model. First, the raw dataset is cleaned to remove unreliable observations. Observations are then aggregated within a 1x1 km² grid cell on a weekly basis between 06:00 and 20:00 on Monday through Friday and the mean $PM_{2.5}$ concentration is assigned to the corresponding grid cells. The calculated $PM_{2.5}$ means are then compared to the referenced RIO concentrations using the one-sampled t-test. The outcome is used to produce t-score and p-value maps, which show if there is a statistical difference between the low-cost sensors and the RIO model. Results did not show any spatial pattern between the weeks of analysis, which was due to the irregularities in temporal and spatial distribution from the Snuffelfiets observations. In addition, the quality of the Snuffelfiets sensors combined has been investigated against the advanced RIVM sensors. The analysis showed no systematic bias between the low-cost sensors and the advanced sensors.

1 INTRODUCTION

Air pollution is a global health problem and growing in salience in the face of urbanization. Exposure to air pollution has been linked to cardiovascular and pulmonary diseases, neurodevelopmental disorders, respiratory diseases and neurodegenerative diseases (Landrigan, 2017; Cesaroni et al., 2014; WHO, 2014). One of the most toxic chemical compounds found in the air and responsible for these health problems is Nitrogen Dioxide (NO_2). This is dangerous due to its high reactivity associated with it being a free radical and is part of particulate matter with particles smaller than 2.5 micrometer: $PM_{2.5}$ (Chen et al., 2007; Fazlzadeh et al., 2021). What is more, NO2 is also one of the most salient compounds in urban contexts as it is mainly generated by combustion processes, such as fuel-based motors in cars.

To improve the public health and the monitoring and assessment of air quality, The European Commission published a directive which limits the Nitrogen Dioxide (NO_2), Particulate Matter ($PM_{2.5}$ and PM_{10}) and Carbon Dioxide (CO_2) values emitted (EC, 2008). Hourly average emissions of NO_2 are limited to 200 µg/m³ (not to be exceeded more than 18 times a calendar year), while yearly average emissions of NO_2 and $PM_{2.5}$ are limited to 40 µg/m³ and 20 µg/m³ respectively. The World Health Organization even advises to limit the yearly average for $PM_{2.5}$ at 10 µg/m³ (WHO, 2005). In addition, the directive imposes that EU members must assess annual ambient air quality in all air quality zones and agglomerations on their territory. To comply to this directive, The Ministry of Public Health and Environment (RIVM) monitors the air quality at 44 automatic monitoring sites, which collect data on an hourly basis, spread over the Netherlands. They are divided into five categories: (1) regional; (2) urban; (3) heavy traffic; (4) industrial and (5) Other. The different categorical sites monitor different air quality variables, for example, particulate matter, nitrogen dioxide and ozone concentration. Using the measured values of particulate matter at the regional and urban locations, the RIVM implemented a new interpolation model, called RIO[1], to estimate the air quality for the Netherlands on a 4 x 4 km² resolution (RIVM, 2014; Janssen, 2008). This model uses the Corine Land Cover (CLC) 2000 map, which

[1] Residual Interpolation optimized for Ozone (RIO)

has a spatial resolution of 100 x 100 m², to incorporate the local character of air pollution in the interpolation model (RIVM, 2014; Janssen, 2008). This links specific (statistical) properties of the air pollution to land use patterns at the same local scale which are described by a so called land use indicator (Janssen, 2008). This results in a final air pollution map of 4 x 4 km², which accounts for the local character of air pollution based on the CLC map.

Providing more detailed spatial mapping of air-quality is limited due to the low density of advanced monitoring stations, only 44 in the Netherlands. The installation of particular monitoring stations tends to be very expensive, require regular maintenance and the monitored value is only representative in a small surrounding area (Borrego et al., 2016; Kumar et al., 2015). Therefore, air-quality maps often lack variation of air quality on a micro-scale, as this depends predominantly on local emission sources, atmospheric flow conditions and street topology (Britter and Hanna, 2003; Bossche, 2015). However, in the recent years a growing trend has been developed to use low-cost mobile air-quality sensors in mapping local variation for air-quality (Borrego et al., 2016; Bossche, 2015; Hu, 2021; Hankey and Marshall, 2015). Although the collection through low-cost sensors produces lower quality data, they can be deployed at a high number of locations simultaneously, which allows for high-resolution air quality mapping (Kumar et al., 2015). However, these datasets contain substantial data gaps (Schneider et al., 2017) and large amounts of data are required to represent the range of possible meteorological and traffic conditions (Bossche, 2015).

One of the entities that deployed the use of low-cost air-quality sensors in The Netherlands is the Province of Utrecht. Through a citizen science project called Snuffelfiets (https://snuffelfiets.nl/), participants received a mobile air-quality monitoring kit, which measured $PM_{2.5}$, humidity, temperature, atmospheric pressure and time of measurement and uploaded this to an online portal. The project went live in 2018 and since has acquired over 25 million observations within Utrecht (Hendricx, 2021). Initially the project was developed to monitor green cycling routes, but the amount of observations collected brought up the interest for use in detailed spatial mapping.

Previous research has produced detailed air-quality maps through the use of data fusion between low-cost sensors and model estimates (Schneider et al., 2017; Hasenfratz, 2015; Gressent et al., 2020). However, these studies are based on a combination of mobile and fixed sensors and have been pre-calibrated before being deployed. As the Snuffelfiets sensor kits are mobile and uncalibrated before deployment, it is of importance to investigate the produced quality of the Snuffelfiets observations before they can be used in data fusion. In this paper, the performance of the Snuffelfiets observations is compared to the estimated RIO concentrations of $PM_{2.5}$ within the Municipality of Utrecht. The aim is to use the observations done by the low-cost air-quality sensors and investigate if there is a statistical difference between the measured $PM_{2.5}$ concentrations and the estimated $PM_{2.5}$ concentrations produced by the RIO model. The materials used in this study are presented in Section 2, the methods applied are described in Section 3, the results and differences between the two maps are presented in Section 4, followed by a discussion and conclusion in Section 5.

2 MATERIALS

2.1 SNIFFING BIKE

Since 2018, the Province of Utrecht started collecting mobile measurements of particulate matter with the help of volunteered cyclists. Participants received a compact mobile monitoring kit to measure different air quality values at high temporal resolution (every ten seconds) during their cycling trips, which is uploaded in real time using Narrow Band IOT (NB-IOT) or LTE-M to the online data platform

of Civicity as weekly CSV files (Civicity, 2019). Measurements for $PM_{2.5}$ are done using the MCERTS[2]-certified Sensirion SPS30 sensor. This sensor measures the mass concentration of $PM_{2.5}$ within the range of 0 – 1000 µg/m³ with a precision of ± 10 µg/m³ between 0 -100 µg/m³ and ± 10% between 100 – 1000 µg/m³ and can detect particles ranging from 0,3 - 2,5 μm (Sensirion, 2021). Before being uploaded to the online platform, each measurement is linked to its geographical location in EPSG:426 CRS and time of acquisition using GPS and internet time.

2.2 RIVM

In collaboration with RIVM, hourly estimated $PM_{2.5}$ values are obtained in text format. The text files contain the geographical point coordinates for the Province of Utrecht (within the boundary box in latitude, longitude going clockwise from top left: 52.257450, 4.881910; 52.25799, 5.74604; 51.90746, 5.74324; 51.90693, 4.88585) in Amersfoort RD New coordinate reference system (CRS) and the average hourly $PM_{2.5}$ estimates corresponding to the point coordinate. Originally the RIO model produces estimates on a 4x4 km² spatial resolution, however the RIVM was able to extract estimates on a 1x1 km² resolution. Therefore, the text files contain point coordinates on a 1x1 km² spatial resolution with a total of 2.400 values per text file.

3 METHODS

3.1 STUDY DESIGN

Data collection is done for the Municipality of Utrecht (latitude 52.083333, longitude 5.166667, area of 1.560 km²) starting from the 6th of January 2020 at 12:00 until the 3rd of February 2020 11:59. The hourly average $PM_{2.5}$ estimates from the RIO model are given as point coordinates on a 1 x 1 km² resolution (points are spaced 1 km from each other). Analysis is done on the aggregated mean $PM_{2.5}$ concentrations from the RIO point estimates within a 1x1 km² grid cell (Figure 1). Same is done for all snuffelfiets observations that fall within a 1 x 1 km² grid cell. The observations within a grid cell are aggregated and the mean $PM_{2.5}$ concentration of all observations is assigned to the intersecting grid cell.

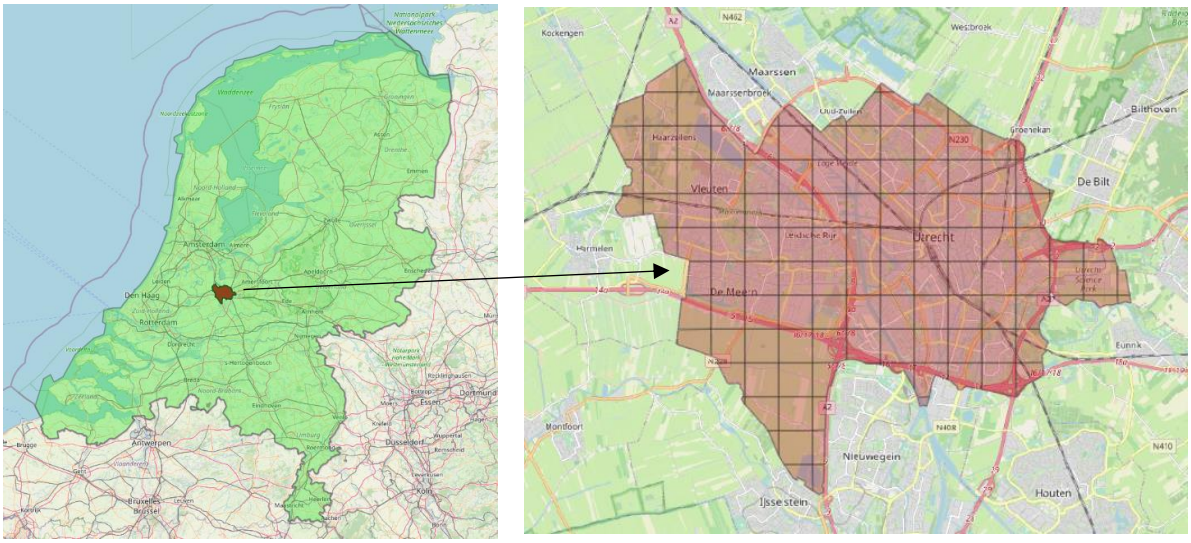

Figure 1: Study area of the municipality of Utrecht within The Netherlands including the 1 x 1 km² grid surface

Data exploration of the Snuffelfiets observations showed that the spatial and temporal distribution differs over days. Most of the cycling activity is registered on Monday through Friday between 06:00 and 20:00, with low to no activity in the night hours. To cover the majority of the grid cells within the

[2] the Environment Agency of England & Wales (EA) Monitoring Certification Scheme

municipality with mean $PM_{2.5}$ values, while keeping the temporal resolution as low as possible, four different weeks are used in the analysis. For each week, observations are aggregated that are registered between 06:00 and 20:00 on Monday through Friday. This results in a mean $PM_{2.5}$ concentration that is aggregated over 70 available hours a week.

3.2 DATA PROCESSING

The Snuffelfiets dataset is cleaned and processed using Python 3.8.6. Through data exploration, incorrect $PM_{2.5}$ observations were detected for different sensors. For example, the dataset contained values within the range 0 – 5.630 ug/m³, which are outside the detection range of 1.000 ug/m³ from the Sensirion SPS30 sensor. These faulty detections can significantly affect the statistics (e.g., the average and standard deviation), resulting in overestimated or underestimated values during analysis (Kwak, 2017). Therefore, outliers are excluded from the analysis by defining a study scope. The following criteria are set for the observations: (1) based on expert knowledge from RIVM, $PM_{2.5}$ values need to lie withing the range of 0.5 - 150 ug/m³ (52.533 observations excluded). Observations outside this range are to be considered as measurement error(s); (2) the registered speed of observations needs to lie within the range of 5 – 45 km/h (354.936 observations excluded). Observations higher than 45 km/h are not considered to be measured by a cyclist and observations lower than 5 km/h are not considered to be part of a cycling trip, as speeds lower than 5 km/h could indicate that the bike is stored inside and (3) Observations need to be located within the geographical boundaries of the municipality of Utrecht (520.575 observations excluded), which are given by the Central Bureau of Statistics (CBS, 2021).

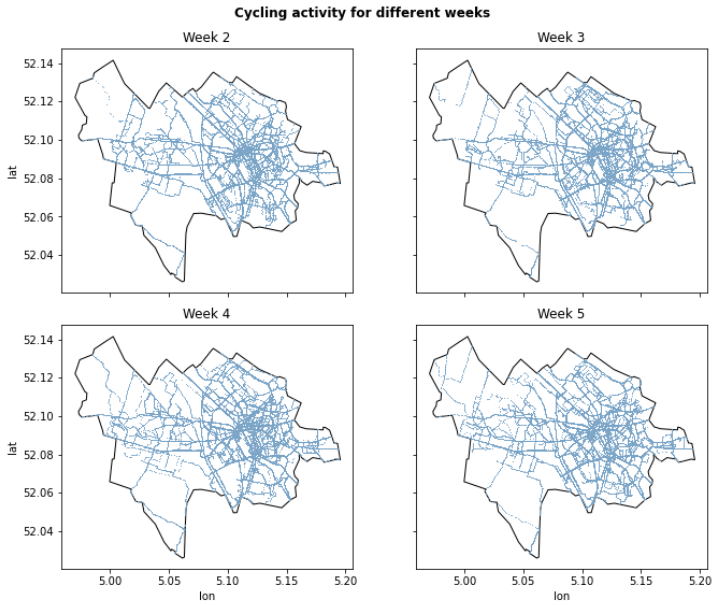

Figure 2: Cycling activity for the municipality of Utrecht for each given week after data processing

Further analysis on the mean $PM_{2.5}$ values registered by each sensor showed observations are present within the dataset that are random in registration. As individual observations are assigned to a specific cycling trip through the 'trip_sequence' variable, trips with <= 20 observations, or a good three minutes, are excluded from the dataset. Trips containing <= 20 observations are considered as noise and therefore are not used in further analysis (27.333 observations excluded). The statistics for the cleaned dataset are shown in Table 1. Trimming the dataset and excluding outliers has reduced the original dataset from 1.112.960 to 157.584 observations, which means only 14,2% of the original Snuffelfiets data is considered as valid for this particular analysis. The corresponding cycling activity for each week is given in Figure 2.

Table 1: Overview of Snuffelfiets dataset after data processing

Week	Start date and hour	Stop date and hour	Observations	Mean $PM_{2.5}$	Range $PM_{2.5}$	St. Dev
2	06-01-2020 12:00	10-01-2020 20:00	36.850	8.90 ug/m²	[1 – 103] ug/m³	5.67 ug/m³
3	13-01-2020 06:00	17-01-2020 20:00	32.521	6.51 ug/m³	[1 - 127] ug/m³	5.54 ug/m³
4	20-01-2020 06:00	24-01-2020 20:00	38.031	16.56 ug/m³	[1 - 124] ug/m³	12.54 ug/m³
5	27-01-2020 06:00	31-01-2020 20:00	33.978	5.63 ug/m³	[1 – 54] ug/m³	4.40 ug/m³

3.3 COMPARISON

The aim of this study is to test whether the mean of the aggregated Snuffelfiets observations is the same as the aggregated mean of the RIO estimates. Mapping the results, based on the Student's t-statistic (Student, 1908), is already used in the field of neuroscience, where mean values for brain activity are compared between two brain activity maps (Miyauchi, 1989; Duffy, 1981; Hassainia, 1994). However, no literature has been found for this application in the field of geoscience. Therefore this study will adapt the significance probability mapping (SPM) technique and produce t-value and p-value maps between the means of the Snuffelfiets observations and the means of the RIO estimates within the municipality of Utrecht. First, the pre-processed Snuffelfiets observations are aggregated on a 1x1 km² grid on Monday through Friday between 06:00 and 20:00. Next, multiple statistics are calculated for each grid for use in analysis; (1) mean $PM_{2.5}$ value; (2) maximum $PM_{2.5}$; (3) minimum $PM_{2.5}$; (4) number of observations; (5) mean relative humidity measured by the Snuffelfiets sensor kit (6) amount of unique sensors contributing to the mean $PM_{2.5}$ concentration; (7) amount of unique hours (Monday-Friday) contributing to the aggregated $PM_{2.5}$ concentrations; (8) mean temperature and (9) mean atmospheric pressure. As previous literature (Li, 2017; Yang, 2017) showed that meteorological variables have a positive or negative correlation with $PM_{2.5}$ concentration, it is of interest to investigate if there is a correlation between different independent variables and the t-score. For this, the significance of statistics five through nine (as statistics one through four are directly correlated with the t-score) will be tested with the use of a linear regression model based on the ordinary least squares method. From the dataset with the RIO estimates the mean $PM_{2.5}$ is calculated, as the available dataset doesn't contain raw observations. Finally, the Student's t-statistic and corresponding p-value is calculated for each 1 x 1 km² grid cell and the two topographic maps are produced for each week.

3.3.1 One-sample t-test

For each 1x1 km² grid cell the following null hypothesis is tested: the mean $PM_{2.5}$ concentration from the Snuffelfiets observations is the same as the mean $PM_{2.5}$ concentration from the RIO model. As no raw data is available from the RIO model, no standard error can be calculated for the RIO model. Therefore, a one-sampled t-test is used, where the mean $PM_{2.5}$ concentration of the Snuffelfiets observations is tested against the referenced mean $PM_{2.5}$ concentration from the RIO model. If there is a difference between the Snuffelfiets mean $PM_{2.5}$ concentration and the referenced mean $PM_{2.5}$ concentration from the RIO model, the null hypothesis is rejected and the alternative hypothesis that the Snuffelfiets mean $PM_{2.5}$ concentration is significantly different from the referenced mean $PM_{2.5}$ concentration from the RIO model is accepted:

$$H_0: \mu_e = \mu_o$$

$$H_a: \mu_e \neq \mu_o$$

μ_0 = Mean from the aggregated Sniffing bike observations

μ_e= Reference mean $PM_{2.5}$ concentration from the RIO model

The t-score, which takes into account the amount of observations done within a 1x1 km² grid cell, is calculated using formula (1). This score uses the standard error of the population mean to indicate the precision of the Snuffelfiets observations mean value. A large sample size 'n' will result in a smaller standard error of the mean and therefore a higher t-score. High t-scores indicates that the Snuffelfiets mean $PM_{2.5}$ concentration is significantly different from the referenced mean $PM_{2.5}$ concentration from the RIO model, while lower t-scores indicate that the Snuffelfiets mean $PM_{2.5}$ concentration is similar to the referenced $PM_{2.5}$ concentration from the RIO model:

$$t = \frac{\bar{x} - \mu_e}{\frac{s}{\sqrt{n}}} \quad (1)$$

$\bar{x}$ = Sample mean
μ_e = Referenced RIO value
s = sample standard deviation
n = number of observations

3.3.2 Snuffelfiets data reliability

At the start of the citizen science project of the Snuffelfiets, no calibration of the low-cost sensors is performed. As calibration of low-cost air quality sensors require a controlled environment, where sensors measure at a constant time interval and location (Wang et al., 2019; Patra et al., 2021; Zimmerman et al., 2018; Chu et al., 2020), calibration of the sensor is out of the scope of this study. However, low-cost sensors are easily affected by environmental parameters, such as temperature and relative humidity (Wang et al., 2019). Therefore, the lack of calibration can result in random and/or systematic bias produced by the low-cost sensors. To interpret the results of this study, it is therefore needed to assess the quality of the Snuffelfiets observations against advanced sensors from the RIVM measurement station. Systematic errors of the Snuffelfiets observations are investigated for hourly aggregated $PM_{2.5}$ concentrations of all sensors combined. The hourly aggregated $PM_{2.5}$ concentrations are compared to the hourly $PM_{2.5}$ concentrations of the two official RIVM measurement stations, where $PM_{2.5}$ is being measured, within Utrecht: NL10643 (Utrecht-Griftpark) and NL10636 (Utrecht-Kardinaal de Jongweg). These are depicted in Figure 3.

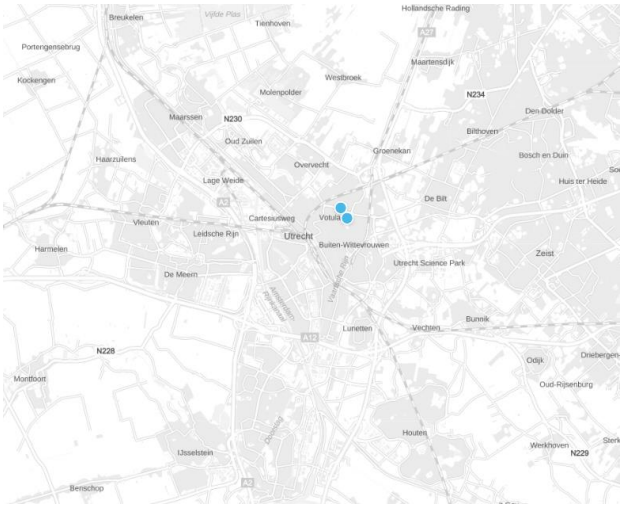

Figure 3: Official $PM_{2.5}$ RIVM measurement stations

4 RESULTS

Results are divided into three sections. First, the t-score and p-value map, where significant differences within grid cells, are presented. Second, the correlation of different independent variables on the t-score is discussed. Finally the Snuffelfiets observations are compared to the official RIVM measurements for data reliability.

4.1 T-scores and p-values

T-score and p-value maps are produced to locate grid cells where significant differences are observed between the Snuffelfiets mean $PM_{2.5}$ concentrations and the referenced RIO $PM_{2.5}$ concentrations. These are presented in Figure 4. Week 2 shows that the RIO model significantly underestimated the mean $PM_{2.5}$ value for the largest part of the municipality (49 grid cells) where t-scores are higher than 2 and are as high as 13.8 with p-values < 0.0001. The absolute mean $PM_{2.5}$ differences vary between 0.44 and 11.95 ug/m³. For the area where no significant difference between the two mean values is observed (31 grid cells) the t-score varies between -2 and 2. These grid cells correspond with p-values > 0.05. Finally a small area (18 grid cells) with t-scores between negative 2 and negative 23 shows that the mean $PM_{2.5}$ concentrations are significantly overestimated by the RIO model with t-scores as low as -23 with p-values < 0.0001. Mean $PM_{2.5}$ difference varies between -0.48 and -5.62 ug/m³. In contrary to the 2nd week, the t-score map for week 3 overestimates the largest part of the municipality (46 grid cells) where t-scores vary between -2 and -32.6 with $PM_{2.5}$ difference varies between -0.35 and -4.41 ug/m³. No significant difference is found for 38 grid cells. Just a small area is underestimated by the RIO model with 19 grid cells. The mean $PM_{2.5}$ difference varies between 0.65 and 8.74 ug/m³. For week 4 the largest area of the municipality is considered significantly underestimated with 58 grid cells containing t-scores that vary between 2 and 27.5. The mean $PM_{2.5}$ difference varies between 0.95 and 30.46. No significant difference is observed for 30 grid cells with mean $PM_{2.5}$ differences that vary between -6.90 and 15.9. The 19 overestimated grid cells have t-scores between -2 and -51.9 with mean $PM_{2.5}$ differences between -0.94 and -13.39. One extreme outlier is observed with a t-score of -51.9, which indicates an extreme overestimation by the RIO model. However, this particular grid cell only contained observations from a single unique sensor on a single day, which could explain the difference. For week 5 the largest area of the map is considered significantly overestimated with 59 grid cells containing t-scores that vary between -2 and -144.2 and mean $PM_{2.5}$ differences between -0.30 and -4.97 ug/m³. Some extreme overestimations are observed with t-scores < -50. These grid cells contained observations from a single sensor within a single day, which could explain the big difference observed between the referenced RIO $PM_{2.5}$ concentrations and the mean $PM_{2.5}$ Snuffelfiets concentrations. No significant difference is observed for 29 grid cells with mean $PM_{2.5}$ differences that vary between -1.23 and 3.72 ug/m³. The 21 underestimated grid cells have t-scores between 2 and 42.5 with mean $PM_{2.5}$ differences between 0.46 and 9.12 ug/m³.

From the t-score maps, no systematic spatial pattern could be detected between the weeks for area's where underestimation, overestimation or no significant difference is observed. Week 2 and 4 seem to share a pattern where the mean $PM_{2.5}$ concentrations are significantly underestimated from the referenced RIO $PM_{2.5}$ concentration, while week 3 and 5 show a dominantly overestimated map. For grid cells where significant differences are found, the mean $PM_{2.5}$ difference falls predominantly within the $\pm$10 ug/m³ range.

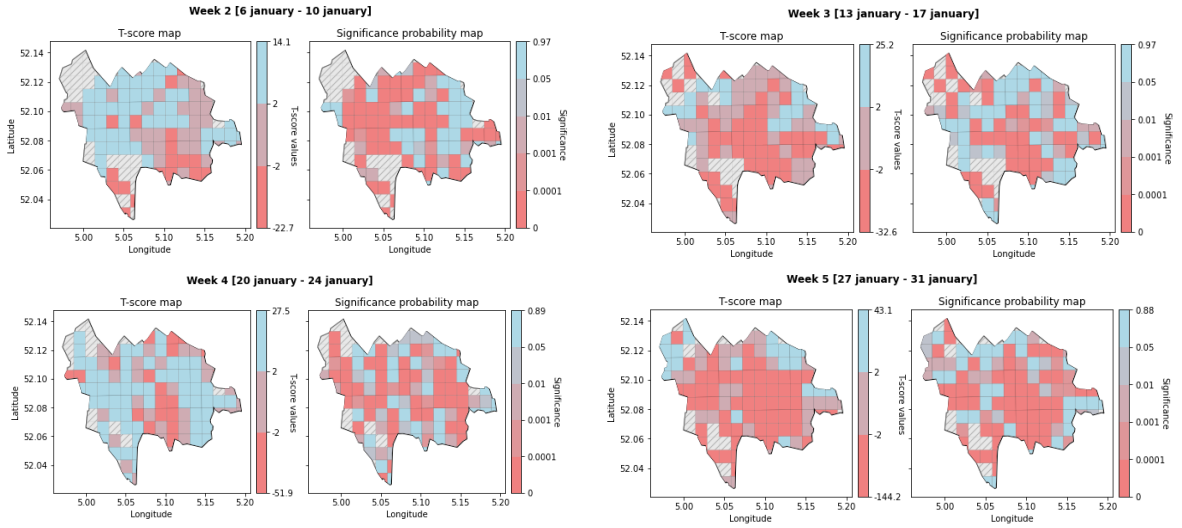

Figure 4: T-scores and corresponding p-values for the four different weeks. Top left: 06 January - 10 January 2020; Top right:13-17 January 2020; Bottom left: 20-24 January 2020 and Bottom right: 27-31 January 2020

Finally, the amount of unique hours that contribute to the aggregated mean $PM_{2.5}$ concentrations from the Snuffelfiets observations are analysed (Figure 5). The referenced $PM_{2.5}$ concentrations from the RIO model are available for all 70 hours that are used for analysis, however the aggregated $PM_{2.5}$ concentration, which is calculated from the Snuffelfiets observations, varies between 1 and 65 unique hours. Week 2 showed 50% of the grid cells containing < 20 unique hours with a maximum of 64 unique hours. Week 3 contained < 17 unique hours for 50% of the grid cells with a maximum of 64 unique hours for a single grid cell. Week 4 contained < 20.5 unique hours for 50% of the grid cells with a maximum of 64 unique hours for a single grid cell and week 5 contained < 17 unique hours for 50% of the grid cell with a maximum of 65 unique hours for a single grid cell.

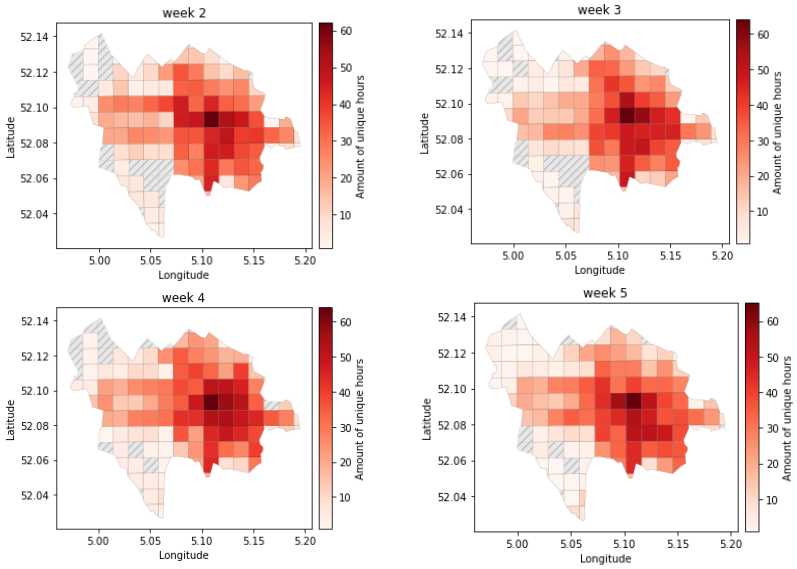

Figure 5: Amount of unique hours contributing to each 1x1 km² grid cell for each week

4.2 Independent variables

As no spatial pattern between the t-score maps is detected, other independent variables are investigated to find a possible correlation with the t-scores. Five of the nine calculated statistics, as four give information about the $PM_{2.5}$ concentrations or the standard error and are directly correlated with the t-score, are used for finding this relationship: (1) the mean relative humidity; (2) number of unique sensors; (3) mean temperature; (4) number of unique hours and (5) mean atmospheric pressure. The correlation plots are presented in Figure 6. For each plot a regression line with a 95% confidence interval is shown.

For week 2, a large negative correlation is observed between the mean temperature and the mean atmospheric pressure against the t-score. However, the mean relative humidity, the number of unique sensors and the number of unique hours show a small positive correlation with the t-score. Linear regression based on the ordinary least squares method is performed to test if the variables have a significant influence on the t-score. This resulted in a model with a R^2 adjusted of 0.102 with none of the variables having any significance. For week 3 a large negative correlation is observed between the mean temperature and the t-score, while the other variables don't seem to show any correlation with the t-score. The Linear regression model resulted in a R^2 adjusted of 0.267 with mean relative humidity, mean temperature and mean atmospheric pressure being significant with coefficients of -0.41, -3.84 and 0.069 respectively, while the other variables are insignificant. Statistics for week 4 showed some extreme values with mean temperature registrations of 78 °C and 162 °C, mean humidity values of 0% and 21% and mean pressure values larger than 2 bar. These observations are due to measurement errors within the sensors and are excluded from analysis. The resulting plots are shown in Figure 6. The linear regression model resulted in a R^2 adjusted of 0.169 with the mean relative humidity, mean temperature and mean atmospheric pressure being significant with coefficients of -0.69, -3.42 and 0.08 respectively, while the other variables are insignificant. Week 5 shows a small positive correlation between the mean humidity and the mean temperature against the t-score, while the other variables don't seem to indicate a relationship with the t-scores. The linear regression model resulted in a R^2 adjusted of 0.214 with the mean relative humidity, mean temperature and mean atmospheric pressure being significant with coefficients of 1.20, 4.60 and -0.14 respectively, while the other variables are insignificant.

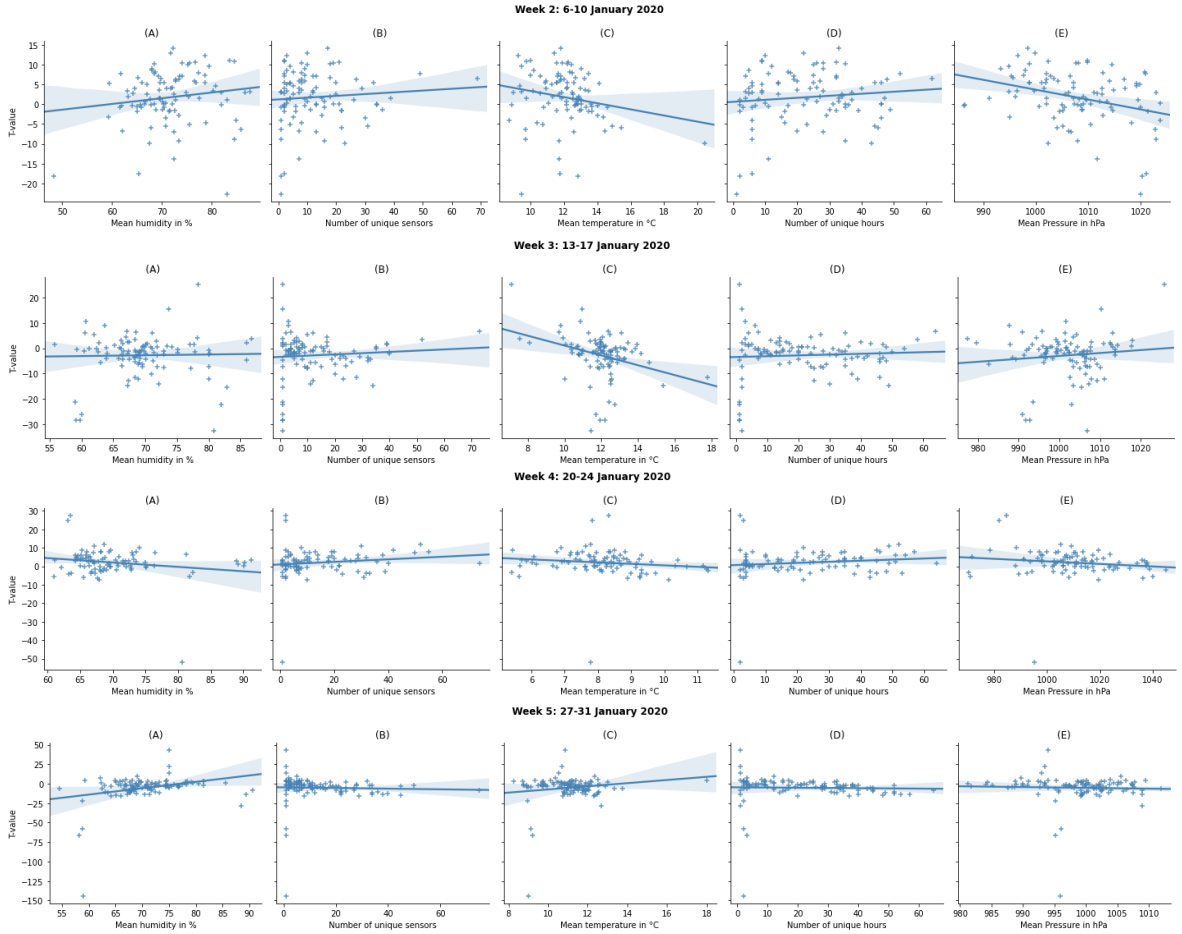

Figure 6: Correlation between the independent variables and the t-value: (A) Observations within a grid cell; (B) Mean humidity in %; (C) Number of unique sensors and (D) mean temperature. First row: 06-10 January 2020; Second row: 13-17 January 2020; Third row: 20-24 January 2020 and fourth row: 27-31 January 2020

4.3 Data reliability

The hourly $PM_{2.5}$ concentrations of the Snuffelfiets observations and the official RIVM stations are presented in Figure 7. For the hourly mean $PM_{2.5}$ concentrations of the Snuffelfiets observations, the 95% confidence band is included in the figure. Compared to station 'NL10636' the hourly $PM_{2.5}$ concentrations range between -3 and 5 ug/m³ for week 2, between -3 and 5 ug/m³ for week 3, between -7 and 8 ug/m³ for week 4 and between $\pm$5 ug/m³ for week 5. Compared to station 'NL10643' the hourly $PM_{2.5}$ concentrations range between $\pm$4 ug/m³ for week 2, between $\pm$5 ug/m³ for week 3, between $\pm$6 ug/m³ for week 4 and between -2 and 5 ug/m³ for week 5. However, differences between the two RIVM stations are also observed, where the hourly $PM_{2.5}$ concentrations range between $\pm$4 ug/m³ for week 2, between $\pm$4 ug/m³ for week 3, between $\pm$5 ug/m³ for week 4 and between $\pm$4 ug/m³ for week 5.

Weather conditions at the weather station 'The Bilt' are used to investigate differences where $PM_{2.5}$ concentrations differ from the official RIVM measurement stations. For week 04, where higher $PM_{2.5}$ concentrations are observed, the relative humidity was the highest between 95% and 100%, which is an indication of rainfall or fog. Relative humidity for the other weeks ranged between 70% and 95%. Temperature for week 04 was lowest between 0 C° and 6 °C, while for the other weeks the temperature ranged between 6 °C and 12°C. For week 05, where the Snuffelfiets observations seem to be systematically lower than the official RIVM measurements for a couple of days, no different conditions for temperature or relative humidity could be found in data from the official measurement station.

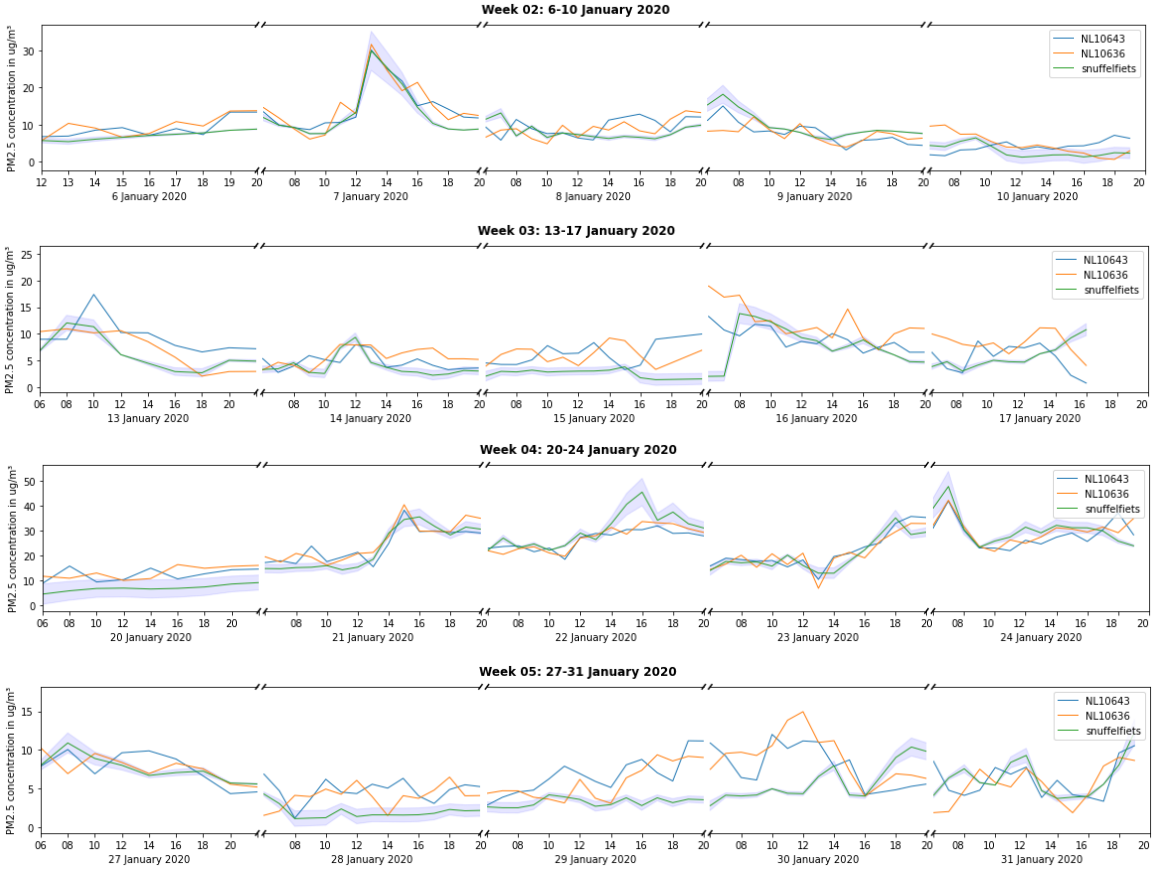

Figure 7: Hourly $PM_{2.5}$ concentrations Snuffelfiets vs RIVM measurement stations

5 CONCLUSION AND DISCUSSION

The study is the first in the geoscience domain that investigated the difference in $PM_{2.5}$ concentration between low-cost sensors and the RIO model based on the one-sampled t-test. First, the corrected data is used to produce t-score and p-value maps for week 2-5 in January 2020. Based on the spatial distribution of the cycling activity, most activity takes place around the city centre, comparable spatial patterns were expected for this area. However, not all 1x1 km² grid cells had equal temporal distribution from the Snuffelfiets observations. Most of the 1x1 km² grid cells (50%) contained calculated $PM_{2.5}$ concentrations from data collected within less than 20 unique hours, while the referenced $PM_{2.5}$ concentration from the RIO model is based on all 70 available hours from official RIVM measurement stations for that week. The produced t-score and p-value maps can show significant differences for areas, while these are not based on a comparable scale. Grid cells where $PM_{2.5}$ concentrations are calculated based on low observations or unique hour count don't show the full averaged $PM_{2.5}$ concentration of the entire week. Therefore, fluctuations due to local emission sources and there temporal variability (Janssen, 2008) are missed by the Snuffelfiets. It is also reasonable that Snuffelfiets activity is lower with (local) bad weather conditions and therefore the actual $PM_{2.5}$ concentrations of that moment are missed in the analysis.

Second, individual variables are investigated for correlation with the calculated t-scores. The mean humidity, the number of unique sensors, the mean temperature, the number of unique hours and the mean atmospheric pressure are used in a linear regression model based on the ordinary least squares method, but all showed poor performance with calculated R^2 adjusted values smaller than 0.267. As literature suggested (Li, 2017; Yang, 2017), the meteorological variables temperature, relative humidity and atmospheric pressure showed a significant correlation with the t-scores. However, the number of unique sensors and unique hours did not show any significance. Based on this analysis, no

minimum amount of unique sensors or unique hours can be determined for an accurate estimation of the t-scores within the 1x1 km² grid cells.

Third, the Snuffelfiets observations are compared to the official RIVM measurement stations. Analysis showed that the Snuffelfiets observations follow the trend of the official RIVM measurement stations within a margin of $\pm$ 6 ug/m³ and no systematic bias could be detected. Considering that the Sensirion SPS30 sensors have a accuracy of $\pm$ 10 ug/m³, the Snuffelfiets observations performed good in comparison to the official RIVM measurement stations. However, due to uncertainty in the spatial domain, it is unknown how these sensors actually perform. As Chu et al. (2020) mentioned in their study, low-cost observations are influenced by high relative humidity. This is observed for week 04 (20 – 24 January), where relative humidity was measured above 95%. Data quality could be improved by calibrating the Sensirion SPS30 sensors in a controlled environment (Wang et al., 2019; Patra et al., 2021; Zimmerman et al., 2018; Chu et al., 2020). However, it should be noted that even two official measurement stations, that are placed 200m apart from each other, show noticeable differences in hourly $PM_{2.5}$ concentrations. It should be noted that the combined Snuffelfiets sensors are compared to single advanced RIVM sensors. Single comparison between low-cost sensors and advanced sensors would be possible if the low-cost sensors are measuring at a fixed location and time interval. Future studies should incorporate a set of fixed sensors to assess the quality of the existing mobile sensors.

In conclusion, the Snuffelfiets Sensirion SPS30 sensors are adequate for measuring air-quality at a 1x1 km² resolution. However, the biggest obstacle for use in air-quality mapping is the inconsistency in temporal resolution. For better performance the sensors should measure at a pre-defined location and time interval. This would mean that the Snuffelfiets sensors shouldn't be mounted on bicycles, but rather be mounted on, for example, lighting poles throughout the municipality. By removing the temporal resolution problem, this study showed that the combined performance of the Sensirion SPS30 sensors is comparable to the advanced RIVM stations. Finally, this study showed that the Snuffelfiets observations are usable for data fusion. As previous literature resulted in promising results (Schneider et al., 2017; Hasenfratz, 2015; Gressent et al., 2020), further research should focus on extracting RIO model estimates on a higher spatial resolution of maximum 250m and use this in data fusion with the Snuffelfiets observations, which could generate improved air-quality maps, better estimation of the individual exposure and general air-quality monitoring.

6 REFERENCES

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods*, *16*(2), 270–301. https://doi.org/10.1177/1094428112470848

Borrego C., Costa A.M., Ginja J., Amorim M., Coutinho M., Karatzas K., Sioumis Th., Katsifarakis N., Konstantinidis K., De Vito S., Esposito E., Smith P., André N., Gérard P., Francis L.A., Castell N., Schneider P., Vianah M., Penzam M. (2016) Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise, *Atmospheric Environment,* 147, 246-263. https://doi.org/10.1016/j.atmosenv.2016.09.050

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

Britter, R.E., Hanna, S.R. (2003). Flow and dispersion in urban areas. *Annual Review of Fluid Mechanics*, 35, Pages 469 – 496. DOI: 10.1146/annurev.fluid.35.101101.161147

Centraal Bureau voor Statistiek (CBS). (2021). CBS gebiedsindelingen. Retrieved from: https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/cbs-gebiedsindelingen

Cesaroni, G., Forastiere, F., Stafoggia, M., Andersen, Z.J., Badaloni, C., Beelen, R., Caracciolo, B., de Faire, U., Erbel, R., Eriksen, K.T., Fratiglioni, L., Galassi, C., Hampel, R., Heier, M., Hennig, F., Hilding, A., Hoffmann, B., Houthuijs, D., Jockel, K.H., Korek, M., Lanki, T., Leander, K., Magnusson, P.K., Migliore, E., Ostenson, C.G., Overvad, K., Pedersen, N.L., Penell, J., Pershagen, G., Pyko, A., Raaschou-Nielsen, O., Ranzi, A., Ricceri, F., Sacerdote, C., Salomaa, V., Swart, W., Turunen, A.W., Vineis, P., Weinmayr, G., Wolf, K., de Hoogh, K., Hoek, G., Brunekreef, B., Peters, A. (2014) Long term exposure to ambient air pollution and incidence of acute coronary events: pro- spective cohort study and meta-analysis in 11 European cohorts from the ESCAPE Project. *Br Med J,* 348-412.

Chen et al., (2007). Outdoor Air Pollution: Nitrogen Dioxide, Sulfur Dioxide, and Carbon Monoxide Health Effects. *The American Journal of the Medical Sciences*, 333(4), 249-256

Chu, HJ., Ali, M.Z. & He, YC. (2020). Spatial calibration and $PM_{2.5}$ mapping of low-cost air quality sensors. *Scientific Reports*, 10**,** 22079. https://doi-org.proxy.library.uu.nl/10.1038/s41598-020-79064-w

Civicity. (2019). Inzicht in schone fietsroutes. Retrieved from https://civity.nl/wp-content/uploads/2019/12/brochure-Snuffelfiets-Civity.pdf

Duffy, F.H., Bartels, P.H., Burchfiel, J.L. (1981). Significance probability mapping: an aid in the topographic analysis of brain electrical activity. *Electroencephalography and Clinical Neurophysiology*, 51 (5), 455-462

European Commission (EC). (2008). DIRECTIVE 2008/50/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 21 May 2008 on ambient air quality and cleaner air for Europe. Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008L0050#d1e89-30-1

Fazlzadeh, M., Rostami, R., Yusefian, F., Yunesian, M., Janjani, H. (2021). Long term exposure to ambient air particulate matter and mortality effects in Megacity of Tehran, Iran: 2012–2017. https://doi.org/10.1016/j.partic.2021.01.017

Gressent, A., Malherbe, L., Colette, A., Rollin, H., Scimia, R. (2020). Data Fusion for Air Quality Mapping Using Low-Cost Sensor Observations: Feasibility and Added-Value. *Environment International*, *143*, 105965. https://doi.org/10.1016/j.envint.2020.105965

Hankey, S., Marshall, J.D. (2015). On-bicycle exposure to particulate air pollution: particle number, black carbon, PM2.5, and particle size. *Atmospheric Environment*, 122, 65-73. https://doi.org/10.1016/j.atmosenv.2015.09.025

Hasenfratz, D., Saukh, O., Walser, C., Hueglin, C., Fierz, M., Arn, T., Beutel, J., Thiele, L. (2015). Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive and Mobile Computing*, 16,268-285. https://doi.org/10.1016/j.pmcj.2014.11.008

Hassainia, F., Petit, D. & Montplaisir, J. (1994). Significance probability mapping: The final touch int-statistic mapping. *Brain Topography,* 7, 3–8. https://doi-org.proxy.library.uu.nl/10.1007/BF01184832

Hendricx, W., Weeseling, J., Huitema, M., Janssen, G. (2021). Monitoring van de luchtkwaliteit .. op de fiets!

Hu, H., Chen, Q., Qian, Q., Lin, C., Chen, Y., Tian, W. (2021). Impacts of Traffic and Street Characteristics on the Exposure of Cycling Commuters to PM2.5 and PM10 in Urban Street Environments. *Building and Environment* 195, 107476. https://doi.org/10.1016/j.buildenv.2021.107745

Janssen, S., Fierens, F., Dumont, G., Mensink, C. (2008). RIO: A NOVEL APPROACH FOR AIR POLLUTION MAPPING.

Janssen, S., Dumont, G., Fierens, F., Mensink, C. (2008). Spatial interpolation of air pollution measurements using CORINE land cover data. *Atmospheric Environment*, 42(20), 4884-4903. https://doi.org/10.1016/j.atmosenv.2008.02.043

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Sabatino, S., Bell, M., Norford, L., Britter, R. (2015). The rise of low-cost sensing for managing air pollution in cities. *Environment International*, 75, 199-205. https://doi.org/10.1016/j.envint.2014.11.019

Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W. (2004). Applied linear statistical models (5th ed.). Boston: McGraw-Hill/Irwin.

Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: management of missing values and outliers. *Korean journal of anesthesiology*, *70*(4), 407–411. https://doi.org/10.4097/kjae.2017.70.4.407

Landrigan, P. J. (2017). Air pollution and health. *The Lancet Public Health*, 2(1), e4-e5. https://doi.org/10.1016/S2468-2667(16)30023-8

Li, X., Feng, Y. J. and Liang, H. Y. (2017). The impact of meteorological factors on PM2.5 variations in Hong Kong. IOP Conference Series Earth and Environmental Science, 78, 012003-012014. DOI: 10.1088/1755-1315/78/1/012003.

Miyauchi, T., Hagimoto, H., Saito, T., Endo, K., Ishii, M., Yamaguchi, T., Kajiwara, A., Matsushita, M. (1989). EEG background activity in patients with dementia of the Alzheimer type--with special reference to analysis by t-statistic significance probability mapping (SPM) in Alzheimer's disease and senile dementia. *Psychiatria et Neurologia Japonica,* 91(4), 244-259. https://europepmc.org/article/med/2798586

Patra, S.S., Ramsisaria, R., Du, R. Wu, T., Boor, B.E. (2021) A machine learning field calibration method for improving the performance of low-cost particle sensors. *Building and Environment*, 190, 107457. https://doi-org.proxy.library.uu.nl/10.1016/j.buildenv.2020.107457

Rijksinstituut voor Volksgezondheid en Milieu (RIVM). (2014). Verbeterde actuele luchtkwaliteitskaarten - Validatie interpolatiemethode RIO Nederland.

Schneider, P., Castell, N., Vogt, M., Dauge, F.R., Lahoz, W.A., Bartonova, A. (2017). Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environment International*, 106, 234-247. https://doi.org/10.1016/j.envint.2017.05.005

Sensirion. (2021). Datasheet Sensirion SPS30. Retrieved from: https://www.sensirion.com/en/environmental-sensors/particulate-matter-sensors-pm25/

Student. (1908). The Probable Error of A Mean. *Biometrika*, 6(1), 1–25. https://doi-org.proxy.library.uu.nl/10.1093/biomet/6.1.1

Wang, Y., Du, Y., Wang, J., Li, T. (2019). Calibration of a low-cost PM2.5 monitor using a random forest model. *Environment International*, 133, 105161. https://doi.org/10.1016/j.envint.2019.105161

World Health Organization (WHO). (2005). WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide – Global update 2005 – Summary of risk assessment.

WHO. (2014). WHO media centre. 7 million premature deaths annually linked to air pollution. Retrieved from http://www.who.int.proxy.library.uu.nl/mediacentre/news/releases/2014/air-pollution/en/

Yang, Q., Yuan, Q., Li, T., Shen, H., Zhang, L. (2017). The Relationships between $PM_{2.5}$ and Meteorological Factors in China: Seasonal and Regional Variations. *Int J Environ Res Public Health*, 14(12), 1510. doi: 10.3390/ijerph14121510.

Zimmerman, N., Presto, A., Kumar, A., Gu, J., Hauryliuk, A., Robinson, E., Robinson, E., Subramanian, R. (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmospheric Measurement Technicques*, 11, 291–313. https://doi.org/10.5281/zenodo.1146109