The effect of urban variables on air quality; a Dutch study

Erasmus School of Economics

Student Number	**431645**
Name	**Sebastiaan Willemsen**
Course	**Master thesis Urban, Port and Transport economics**
Supervisor	**Anna Bornioli**
Second Assessor	**Jeroen van Haaren**
Date Final Version	**28-02-2021**

This paper analyzes the effect of urban variables on air pollution values in the Netherlands in varying settings namely; a regional level, a municipality level, and an inner-city level. Urban variables are used to indicate the compactness of areas/cities. Data is acquired from governmental supported organizations such as the Dutch Central Bureau of Statistics (CBS), emission registration and luchtmeetnet. OLS regressions show significant positive results from the urban variables population density, land use-built areas and significantly negative land use agriculture and private car ownership at a 1% level on the air pollution values of CO_2 , NO_2 and PM_{10}, also when adding province level fixed effects. Hot Spot analysis and Moran's I show spatial autocorrelation in the dependent variables, significant clusters occur for CO_2 around Amsterdam and for NO_2 and PM_{10} near Rotterdam. Next, the effect of inner-city characteristics on pollution is measured. Inner-city characteristics influence the air dispersity levels, which in turn affects pollution values. Air dispersity is measured by showing the effect of low air dispersity locations, while controlling for wind speed, wind direction, day of week, month, and location of the station. The results showed a positive significant effect of air dispersity on pollution values at a 1% level in all models. Hence, urban planners must consider urban variables, air dispersity levels and inner-city characteristics when designing densely populated cities such as Rotterdam or Amsterdam.

Table of Contents

1. Introduction

The Dutch government has imposed a series of major regulations in order to reduce the amount of nitrogen oxides in 2019 (Rijksoverheid, 2019). Compared to other European countries, the Netherlands had the highest relative pollution of nitrogen per square kilometer (TNO, 2019). This had a deteriorating effect on public health as well as nature (RIVM, ND). Although the Dutch government had implemented a program to contain the pollution of toxic gases, one of the most important judges ruled that these measures were not sufficient (de Graaf & Jans, 2015). Therefore, new measures had to be introduced in a short amount of time. The new measures introduced by the Dutch government had a very large impact for the construction and the agriculture sector. This led to nationwide protests by both sectors and new measures had to be created to satisfy all parties. These short-term measures also include a speed reduction on all highways, funding for cattle farms willing to quit, make renewable stables, and more specific measures for construction. However, nitrogen is not the only issue regarding toxic gases in the Netherlands, CO_2 pollution has also been a major political issue. The Paris agreement, signed by almost all European countries including the Dutch government, is the best example of an attempt to collectively tackle climate change. The goal of the Paris agreements is to stop the temperature from rising. To achieve this goal, the Dutch government has set the objective to reduce CO_2 emissions by 49% in 2030 and 95% in 2050 compared to 1990 (Rijksoverheid, ND). In addition, in a judicial case called Urgenda the judge has decided that the Dutch government must reduce 25% greenhouse gases compared to 1990 (de Graaf & Jans, 2015). This led to a lot of political commotion and because of this the Climate Plan was introduced, which is a 10-year plan to preserve the environment and ensure a healthy living environment. Due to the current corona crisis, the environment thrives mainly because of reduction in (air) traffic (Blum & Neumärker, 2020). However, this also makes it hard to measure whether the reduction in nitrogen and CO_2 is actually because of the policy measures or the corona crises. The good news is that the Dutch government is likely to achieve the short-term goals set out. This paper will focus on long-term reduction of greenhouse gases through urban planning.

This paper tries to establish a relationship between the multiple urban variables and the amount of air pollution per square kilometer in a municipality, region and inner-city environment. As the Netherlands is one of the most densely populated countries in the world, the potential to reduce greenhouse gases by smart urban planning is promising. The three main variables that play a role in this relationship are urban density, the structure of cities and air dispersity.

This paper will formulate an answer to the following research question:

Are compact cities associated with air pollution in the Netherlands?

This question is answered by making use of 3 different hypothesis. First, the analysis will be done on a municipality level for all municipalities in the Netherlands. Compact cities will be defined by urban variables such as population density levels and land use variables. From literature by Neuman (2005) it becomes clear that the concept of the compact city is defined by many characteristics such as: high residential area, urban infrastructure, multimodal transportation, high degrees of accessibility, low open-space ratio and more. Furthermore, Geographic Information Systems (GIS) will be used to show the current situation in terms of pollution for the second hypothesis regarding regional effects. In addition, ArcGIS is used as a statistical tool as well. Also, multivariate fixed-effect regressions will show the effect of compact city development on air pollution. This paper will consider 3 of the most relevant air pollutants namely: CO_2, NO_2 and PM_{10} for hypothesis 1 & 2. All of these are affected differently by the compact city developments. The third hypothesis uses 1-hour NO_2 values to assess urban variables on an inner-city level. This will consider air dispersity as one of the most important reasons for public health issues in inner-city environments. To assist with the research question, the following hypotheses are made.

Hypothesis 1: Land use, population density and private car ownership have a significant effect on emission values

This hypothesis is focused toward the statistical analysis showing the effect of urban variables on air pollution on a municipality level. The urban variables will determine the compactness of the city. The urban variables used to carry out the Ordinary Least Squares (OLS) regressions are land use variables, population density, population density categorical, and private car ownership as the independent variables while average income, average house value, western population share (migrants), distance to amenities and energy consumption represent the control variables. These values were retrieved from the Central Bureau Statistics (CBS). Land use-built areas and population density indicate significant positive results while, land use-agriculture and private car ownership denoted significant negative values. The other compact city determinant population density categorical showed insignificant increasing values when municipalities getting less densely populated. The fixed effect regression with province dummies showed no large difference, signifying strong correlations of the urban variables. However, Fang et al (2015) identified issues with spatial autocorrelation within the dependent and independent variables when using OLS. To show whether this sample contained autocorrelation within the dependent variables, Moran's I and hot spot analyses were carried out in hypothesis 2.

Hypothesis 2: The dependent variables (air pollution values) show significant clusters

To estimate spatial autocorrelation (regional level) of the dependent variables, Moran's I and Hot spot analysis are carried out to show clustering in the sample. The expectation is that there will be positive significant clusters around the Randstad area and possibly negative significant clusters in the North-Eastern part of the Netherlands (Alpkokin, 2004). The results of the hot spot analysis showed significant positive clustered values for CO_2 around Amsterdam at 10% levels, while NO_2 and PM_{10} depicted significant positive clustered values around Rotterdam at a 1% level. In addition, the Z-scores maps of the hot spot analysis reveals a clear spatial autocorrelation in the Netherlands, resulting in a bias for the OLS regression results. The Moran's I statistic confirms this by showing significant results, indicating clustering. More research must be done to account for the spatial autocorrelation in the dependent variables to enhance the OLS results in hypothesis 1.

Hypothesis 3: Inner-city characteristics in Rotterdam and Amsterdam have a significant effect on air quality levels at different wind speeds

This hypothesis is used to estimate the effect of urban variables in an inner-city environment and depict the implications of urban planning on air dispersity. The hypothesis has more implications for urban planners as urban morphology is a determinant for the dispersion of air pollution. This will be confirmed first in the empirical literature section, after which these papers will be compared to data in the Netherlands. Most of the empirical literature uses simulations to analyse the effect of urban morphology on air pollution through air dispersion. Since this is a case study, actual data from measuring stations in Rotterdam and Amsterdam is used to assess the air dispersion levels in the inner-city environment of the Netherlands. City characteristics can be used to explain some of the high pollution levels.

The air dispersion levels in the Netherlands are not comparable to the described building canyons by Yuan et al (2014), but the high density of the buildings paired with the urban morphology of Dutch inner cities might still give a significant effect.. The expectation was that the air dispersion in inner-city environment would have an effect. The OLS results show a positive significant effect of the location denoted as low air disperse compared to the high disperse locations at a 1% level. Regarding the control variables, wind speed, and the city dummy were significant negative at a 1% level, month and time of the day were significant at a 1% level with varying signs. This shows that the identification of low air dispersity locations was solid and city characteristics does have a significant positive effect.

The structure of this paper is the following. First, the literature analysis will explain what research has been done on the topic. The literature review will define the concept of the compact city, show other research linking urban planning and greenhouse gases and define why city structure is a relevant factor for long-term pollution reduction. Next, the variables in the dataset will be explained for every hypothesis independently as well as the sources of these variables will be evaluated. This will be followed by the methodology which will explain what statistical methods were used to compose the results. The results section will analyze and compute different regressions and statistical figures to formulate an answer on the hypothesis and the research question. The conclusion will combine all insights to formulate a final answer to the research question. In addition, the conclusion will provide short policy implications and urban planning recommendations for the Netherlands. Finally, the limitations of the paper are summarized and the recommendations for future research are described.

2. Empirical Literature

The empirical literature will feature prominent papers which contributed to the current knowledge regarding compact cities, air dispersity and air pollution. This will consist of four parts: defining compact cities and introducing sustainable urban forms, current literature in the Netherlands about air pollution/compact cities, papers estimating the impact of urban development on air pollution and the effect of city structure on air dispersion. These four parts together will be the foundation of this research and the statistical methods out of different sources will be used.

2.1 Compact cities and sustainable urban forms

Defining compact cities is not an easy task as denser cities were bound to happen due to the rise in worldwide population size. However, the compact city concept has been around for decades with varying effects on urban planning. There have been many studies to evaluate the performance of compact cities and all these studies have a slightly different description of the compact city. Most of these researches has been done around the 2000s, since then, developments within urban planning has evolved and sustainable urban planning has gained more traction. The sustainable urban form has great similarities with the compact city, but the empirical research will show that the two are not the same. However, sustainability plays an even greater role today, as air pollution threatens the highly dense societies/cities.

Burton (2000) states that the compact city has emerged as a reaction to the need to find more sustainable models for the cities in the developed countries. It has been acknowledged that the compact city has a variety of definitions, in general it means a high-density, mixed-use city, based on an efficient public transport system and more space for walking and cycling (Burton, 2002). This new concept contradicts with the more car-oriented urban sprawl of many modern American cities. Dieleman & Wegener (2004) state that the city of Portland in North America was one of the first to establish growth boundaries. Other North American cities were not using zoning policies to ensure open areas and decrease urban sprawl. Compared to other cities, the public transport use has increased by 60 percent in Portland. Since the United States was rich in open land at the beginning of the century, the compact city was not the priority of urban planners for many cities. Also, Dieleman & Wegener (2004) list the negative aspects of the compact city development in Portland: negative impact on the competitiveness and social balance of cities, house price and rents doubled in the 1990s, inner-city urban renewal projects aimed at medium- and high-income households. On the other hand, Dieleman & Wegener (2004) state other empirical literature that show the house prices have not grown faster than in other comparable cities.

	COMPACT CITY CHARACTERISTICS
1	High residential and employment densities
2	Mixture of land uses
3	Fine grain of land use (proximity of varied uses and small relative size of land parcels)
4	Increased social and economic interactions
5	Contiguous development (some parcels or structures may be vacant or abandoned or surface parking)
6	Contained urban development, demarcated by legible limits
7	Urban infrastructure, especially sewerage and water mains
8	Multimodal transportation
9	High degrees of accessibility: local/regional
10	High degrees of street connectivity (internal/external), including sidewalks and bicycle lanes
11	High degree of impervious surface coverage
12	Low open-space ratio
13	Unitary control of planning of development, or closely coordinated control
14	Sufficient government fiscal capacity to finance urban facilities and infrastructure

Table 1: Compact city characteristics. Source: Neuman (2005)

Neuman (2005) has created a list of 14 compact city characteristics shown in Table 1 above. The characteristics in Table 1 are based on reviews of practice, research, literature (especially Burton 2000; Galster et al. 2001; Song and Knaap 2004), and observation which can be used to guide future research but also can be used to plan and guide a new town. Neuman notes that research around that time were the first to separate the concept of the compact city apart from density. However, the compact city characteristics could describe nearly any city. Furthermore, Neuman shows by making use of empirical literature that the compact city is not directly a sustainable city. Although most of the compact city characteristics are focused toward being sustainable, empirical research shows that for a city to be sustainable, functions and population must be concentrated at higher densities. On the other hand, for a city to be livable, functions and population must be dispersed at lower levels. Neuman (2005) finds that cities do want to be sustainable but do not know how to fit this into their urban planning program. Next, Neuman uses the book 'The Cost of Sprawl' by Burchell et al (2002) to explain the problems with urban sprawl. Most of the studies shown in that book conclude that there are both positives and negatives to this urban planning method. Burchell (2002) concludes that sprawl is more costly than compact city development for both capital and operating costs. Neuman (2005) states the characteristics of the urban sprawl cities (Table 2). Neuman concludes that urban sprawl is a result of cheap land outside the city, advances in transportation, easily available capital to buy property, the rise of real estate developer, mass production of housing, and the always-present image of the single family home as the American dream. Song and Knaap (2004) have conducted a quantitative research analysis on the urban sprawl levels in Portland, Oregon. Their main conclusion is that this area had improved in terms of connectivity, pedestrian access, and density by the newly developed neighborhoods. These newly developed neighborhoods consisted of a multinucleated urban form, a multi-modal transportation system and of mixed-use regional and town centers.

	URBAN SPRAWL CHARACTERISTICS
1	Low residential density
2	Unlimited outward extension of new development
3	Spatial segregation of different types of land uses through zoning
4	Leapfrog development
5	No centralized ownership of land or planning or land development
6	All transportation dominated by privately owned motor vehicles
7	Fragmentation of governance authority of land uses among many local governments
8	Great variances in the fiscal capacity of local governments
9	Widespread commercial strip development along major roadways
10	Major reliance on a filtering process to provide housing for low-income households

Table 2: Urban Sprawl characteristics. Source: Burchell et al, 1998. Neuman (2005).

This thesis will take over the definition and characteristics portrayed by Neuman (2005) and will try to identify urban characteristics for the analysis based on Table 1. Although this will not be possible for every characteristics, some proxy variables could give another insights into the urban planning of a region and the availability of data will have to determine mostly the viability to define the compact city in the Netherlands.

2.2 The Netherlands: compact city development and air pollution

First, the compact city developments over the years will be discussed in this section. After which the Dutch pollution issue will be explained in more detail. Next, the two parts will be linked, giving information into the process behind urban planning in the Netherlands and its relation to pollution.

According to Breheny (1995) the Dutch government has made the compact city a central element of its National Physical Plan since 1991, on advice from the European commission. The Netherlands certainly was not the only country at the time adopting this urban planning strategy. The United Kingdom, Australia and the United States also implemented similar strategies to promote a more sustainable form of urban planning like compact cities (Breheny, 1995).

Alpkokin et al (2004) has analyzed the Netherlands in terms of planning and governmental models toward sustainable urban development. They praised the general policy framework called the Polder Model, which explains the relationships between all stakeholders. It depicts the government interacting with the local authorities and the different ministries at national level. The second relation is between government, non-government organizations and private sector. The last dimension is the relation between the government and the European Union (Alpkokin et al, 2004). In addition, the authors praise the Netherlands for implementing mixed-land use features to reduce the need to travel but also for using instruments to increase the share of public transport. Regarding the environmental planning in the Netherlands, Alpkokin et al (2004) state that the total environmental expenditure cover 1,9% of the total GDP, which is higher compared to the UK (1,4%) or Germany (1,0%). However, at the same time the CO2 pollution from road transport had risen by 35% since 1990 (Alpkokin et al, 2004). The Netherlands used 4 green taxes to compensate the CO_2 pollution from using a motorized vehicle. These are: taxes on passenger cars and motorcycles, motor vehicle tax, tax on heavy good vehicles and environmental taxes. In addition, the Netherlands is one of the first countries to initiate carbon taxes (Alpkokin et al, 2004). In conclusion, this paper states that developing countries can use the Dutch governmental methods to create a sustainable spatial planning policy. Although, differing challenges must be faced in all countries, according to Alpkokin et al (2004), other countries can adopt many of these methods to recreate the success story.

Finally, Priemus (1994) stated the importance of developing the Randstand region before the turn of the century. First, it was stated that the Randstand region can be compared to large dense cities like London, Paris or New York. Next, the importance of networks from the Randstand region to other parts of Europe is explained. It has very strong linkages due to the presence of the Port of Rotterdam and the airport Schiphol. This translates to urban network because of the strategic position of the Randstad as well as the internationally oriented economy. Regions in the Randstand are leading in the fields of distribution and logistics, the financial markets, international organizations and institutions, the degree of education and the work ethic of the population (Priemus, 1994). At the same time, the Dutch government realizes the importance of nature and therefore has created a green heart in the middle of the 4 large cities in the Randstad (The Hague, Rotterdam, Utrecht, and Amsterdam). This area is restricted from urbanization and it should promote a healthier living environment close to the suburbs. Van der Valk (2002) confirms this but also states that the land use changes, and the price of open forest land is incredibly high. Multiple land use and new development concepts are vital for the Netherlands to slow urban sprawl and promote green areas near cities. The road and rail infrastructure around the ring of the green heart is strengthened especially. Even now, Priemus (1994) concluded that the rising car use is one of the main problems to ensure a healthy living environment in the Randstad. In addition, he realized that the road infrastructure had to be reformed and adapt to the rise in car traffic. Also, Priemus (1994) states that the Randstand was saved using bicycles and realized the strategic importance in the upcoming years, also with respect to the environment.

The picture in Figure 1 describes best the pollution situation in the Netherlands. This picture was widely picked up in the Dutch media to explain why immediate political response is necessary to stop the spread of nitrogen. The Dutch nitrogen values are not comparable to any other region in Europe but the Po-valley in Italy. This leads to public health risks in the long-term as well as a deterioration of the preserved nature in the Netherlands. Measurements by the newer Tropomi instrument showed similar results (Veefkind et al, 2012).

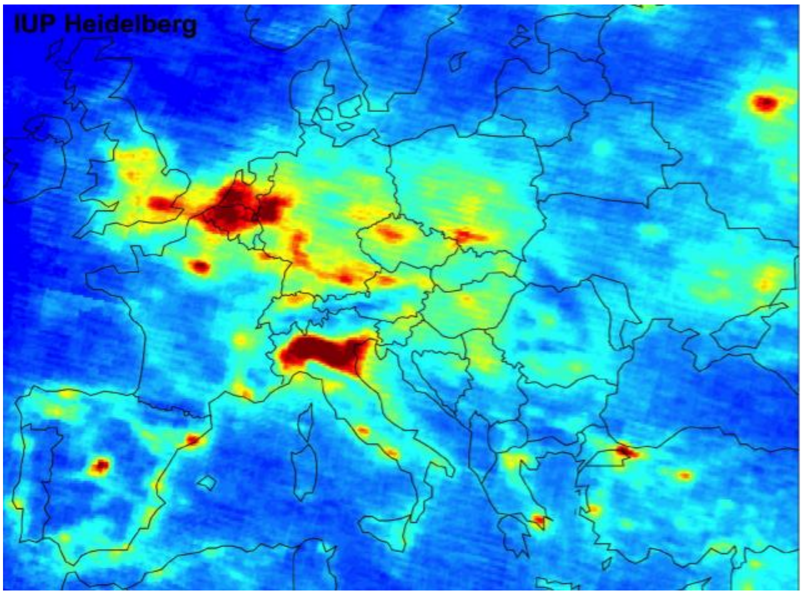


Figure 1: European Mean tropospheric nitrogen dioxide (NO2) vertical column density between January 2003 and June 2004. Source: S. Beirle, U. Platt and T. Wagner of the University of Heidelberg's Institute for Environmental Physics. Almonti (2018).

2.3 Urban development and air quality

This section will discuss relevant literature regarding the relationship between air quality and urban development. These papers are published in trustworthy journals and the researchers have intensively studied this topic. Most of the papers featured have researched fast developing Asian countries. These differ from our research as most of the pollution in these countries is caused by production and the steep rise of intensive car use. It does however offer a very good guide toward the most suitable research methods.

Rodríguez et al (2016) investigated the relationship of urban development and air quality by emphasizing urban fragmentation. They categorize this fragmentation in four city types: dense and continuous (Compact city), dense and fragmented, non-dense and continuous and non-dense and fragmented. They state that non-fragmented cities enhance connectivity, reduce mobility needs and car dependency, and facilitate the use of non-motorized transport modes. Also, continuous cities may induce benefits such as energy savings, reduction of costs of maintenance for energy and transport systems, improvement of quality of life through local services and jobs, and more efficient infrastructure investments. The expectation is that the compact cities produce lower emissions of transport-related pollutants compared to a fragmented city. Rodríguez et al (2016) conclude that the pollutants (NO_2, SO_2 and PM_{10}) are affected differently by urban characteristics due to differences in emission source. Furthermore, they find that high-income areas are associated with lower values of concentration for PM_{10} and SO_2. They claim the results reflect that by designing cities more continuous and non-densely populated cities can improve air quality.

Fang et al (2015) used the Air Quality Index (AQI) to investigate the effect of China's urbanization process of air quality. The AQI represents several social health threatening pollutants. These are categorized on importance to get an overall score for air quality. The methodology of variable selected will be replicated from this research and will be explained in depth further on. Fang et al (2015) identify 12 urbanization variables used as independent variables. Further on the paper uses spatial dependence in the form of Moran's I to estimate whether a variable exhibits significant spatial dependence and heterogeneity. They concluded that, with the exclusion of natural indicators, urbanization plays an important role in determining air quality in China. They found that total population, urbanization rate, automobile density, and the share of secondary industry all had a significant effect on air quality. The paper also finds that by making use of Moran's I neighboring cities have a significant influence on the air quality. However, the results also showed that the relationship of urbanization to air quality was not constant over space, but rather varied. Finally, policy implications are suggested for the next 30 years in China. Fang et al (2015) advises that the Chinese government should control the mega-cities as the high population density is causing a lot of pollution, not only in that regions but also neighboring regions. Also, new policy should consider spatial factors as pollution does not just come from stationary (industry) but also non-stationary (motorized vehicles). Furthermore, the Chinese government must take measures based on the socio-economic status of a regions, as not all regions will experience the same effect of urban development.

Cho & Choi (2014) use a panel dataset from 1996-2009 in from Korea to explain urban development on air pollution, while considering both the spatial distribution of pollutants and the dispersion of pollutants. The panel data is influenced by spatial and panel data. They raise the argument whether compact cities contribute to air pollution reduction, while air pollution is influenced by spatial and temporal changes in urban characteristics as well as by dilution and dispersion processes in the atmosphere. By testing the proximity to the city center on air pollution levels, Cho & Choi (2014) argue that spatial concentration is not the only major contributor. Compact development can enhance the number of green areas in the surroundings of a city, this will increase the dispersion and dilution of pollutants, which can result in lower air pollution levels. The panel data estimations indicate that urban compactness has both positive and negative effects on air quality. Also, Cho & Choi (2014) claim that different policy measures are need for every type of pollutant. NO_2 should receive more local policy as it is associated with high density areas while PM_{10} and O_3, are more affected by regional factors because these are most produced by gas-to-particles conversions (Cho & Choi, 2014).

2.4 Urban Morphology and air dispersion

Urban morphology is a relatively new concept in empirical literature referring to the exact position of buildings or other urban elements. It explains the spatial structure of cities by especially looking at the physical form of streets and buildings, also compared to other cities (Moudon, 1997). The literature that follows will explain the effect on urban morphology on air dispersion.

Hang et al (2009) were one of the first to examine the air dispersion in compact cities where the whole city itself acts as an obstacle to the wind. Pollutant dispersion in compact cities depends on air flow, which in turn is affected by urban morphological characteristics such as building area density, overall city form and street configurations. They simulated three different city forms, with wind coming in from various directions and concluded that urban morphology has a significant effect on air dispersion in city. Moreover, low air dispersion in high-dense cities leads to smog and public health risks. Although the conclusions by Hang et al (2009) are solid, the simulation models of different city forms are quite simple and can be extended further. Hang et al (2012) did extend his research later by assessing the effect of building height on air dispersion. The vortex created by different building heights has a significant effect on air dispersion. Although more simulation options must be conducted for actual policy to be effective, urban morphological characteristics should be considered more often according to Hang et al (2012).

Yuan et al (2014) argue that population density in high-dense megacities has a higher impact on public health than cities of lower population density. Due to the number of high-rise building in these megacities, the pollutants cannot escape the city. This is due to the wind not being able to play a role, which causes low air pollutant dispersion. They show that the proposed mitigation strategies are needed to alleviate the negative effects of air pollution on public health in high density cities. The study by Yuan et al (2014) has modeled different scenarios with respect to the distance in which building should be separated to achieve the best possible air dispersion. They showed, with a real case study that urban morphology is feasible in real urban designs. In addition, they propose mitigation strategies for policy makers to choose from. These strategies include changing only one building, which is more practical than an entire redevelopment of an inner city.

Wang et al (2018) adds to this by stating the importance of trees in these high-dense megacities. They assessed the air dispersion, with various simulation, ranging from different trees but also differing canyon geometries. Due to the low wind speeds, trees can stop the air flow even further and thereby worsen air dispersion. Dense trees in wide canyons are able to reduce pollutants by enhancing outer airflow and canyon vortices. The type of tree used in a certain type of canyon is important. Urban planners can use simulation from this research to find the optimal tree for their inner-city environment (Wang et al, 2018).

3. Data

This section will describe the data used as well as the validity of the sources, any missing observations and transformations. In addition, descriptive statistics will be shown and described. Data used differs for all Hypotheses. Hypothesis 1 & 2 will apply the same dependent variables (CO_2, NO_2 and PM_{10}) but the independent variables (land use, population density and private car ownership) are only utilized in Hypothesis 1. The data for the dependent variables in Hypothesis 1 & 2 is based on yearly municipality level data, estimating pollution values based on local measurements but also on report by refineries, larger companies and agriculture, drinking water supply and more in 2017. The dependent variable in Hypothesis 3 is based on the hourly NO_2 values from inner-city measurement stations in Rotterdam and Amsterdam. Both datasets offer different insights which will be discussed in-depth.

3.1 Dependent variables (Hypothesis 1)

The dependent variables emission values of carbon dioxide (CO_2), nitrogen (NO_2) and particular matter (PM_{10}) are acquired from the Dutch website of emission registration (Emissieregistratie, ND). The goal of this organization is to accurately measure emission on a yearly basis. According to this organization the numbers must be actual, accurate, complete, transparent, comparable, and consistent. This organization has been funded by the ministry of Infrastructure and Water (IenW). However, the actual business and registration of numbers is done by the public institute of health and nature (Rijksinstituut voor Volksgezondheid en Milieu, RIVM). The sources of air pollution are separated in sub-categories: energy usage, garbage disposal, traffic and transport, agriculture, trade, services and government, consumers, construction, chemical industry, drinking water supply, refineries, other industry, sewage and water treatment plants, nature and others. To get a good picture what the main contributors are for the CO_2 pollution, Figure 2 below is created. It shows that almost a quarter of the total CO_2 pollution is caused by traffic and transportation, 23,7% is caused by energy usage, 14.3% by consumers and 13% by other industry. The same figures are made for NO_2 and PM_{10} and can be seen in the Appendix (Figure D and Figure E). Almost two third of the NO_2 pollution is caused by traffic and transportation. Regarding PM_{10}, agriculture plays a somewhat larger role, but other industry, traffic and transportation and consumers remain large (like CO_2). Now, the numbers published by emission registration (Emissieregistratie, ND) cannot be fully accurate, since this organization does not have endless resources, not every region is updated every year. To compensate for this the organization has introduced an uncertainty percentage based on national level. The uncertainty level for carbon dioxide is the lowest (3%), but the PM_{10} and nitrogen gases are higher, respectively 20% and 26%. Nonetheless, this is the best possible source for an dependent variable as it is government funded, multiple sources based and provides detailed information about the source of pollution.

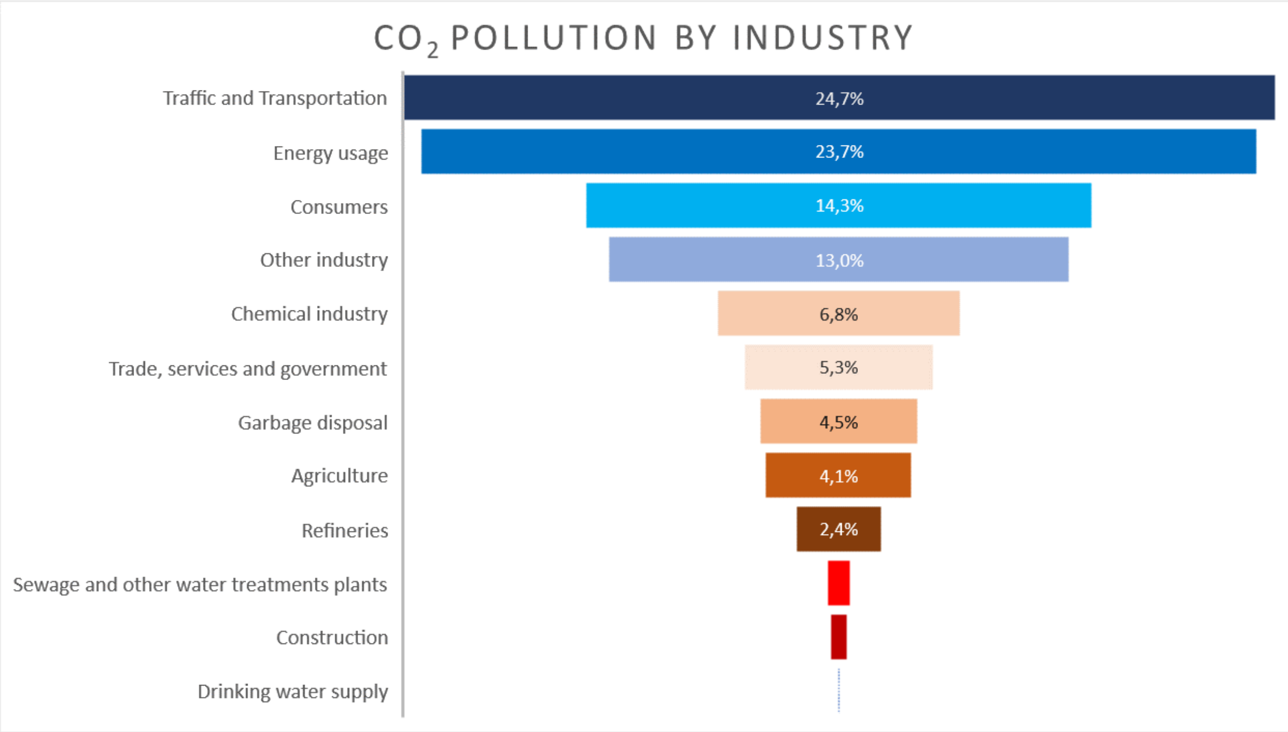


Figure 2: Funnel chart of pollution by industry. Source: Emissieregistratie

For the analysis of hypothesis 1, the focus of the dependent variable will go toward 3 pollutants: Carbon dioxide (CO_2), Nitrogen dioxide (NO_2) and particular matter (PM_{10}). CO_2 is chosen because it is widely popular and one of the most important gases affecting global warming (Lashof & Ahuja, 1990). The causes for CO_2 are mainly transportation, electricity production and industry (Emissieregistratie, ND). The second pollutant chosen for this research was NO_2. Traffic is the main contributor of this pollutant. However, it is also chosen because NO_2 creates smog and continued exposure to NO_2 can lead to a shortened life span of 4 months and is more often present in inner-city environment (Cho & Choi, 2014). In addition, this pollutant has a disruptive effect on the continuation of plants and trees (RIVM, ND). The PM_{10} is the fine pollutant used in this research, it is also known as fine dust. The main contributors of this pollutant were Industry, Energy usage and Refineries. It is chosen because it poses a societal health threat, which can shorten the life span up to 9 months (Maas et al, 2015). The reason to focus on these 3 pollutants is because all of these have a slightly different source and urban characteristics might have a different effect on all of them.

3.2 Independent variables (Hypothesis 1)

The independent variables are acquired from the Dutch Central Bureau of Statistics (CBS) which is the main collector of household and regional data. The largest share of the data has been collected from the "Kerncijfers Wijken en Buurten 2017" (key figures for districts and neighborhoods). This dataset is openly available on the website of CBS and offers detailed data on many regional statistics. This paper considers only the numbers at municipality level, making it possible to be combined with the pollution data. In the end the number of municipalities are 343, excluding the dropped variables due to no observation in the pollution dataset. The reason why the CBS dataset is excellent, is due to the number of detailed independent variables. The ones that will be used to compute the regressions are depicted in Table 3 below. Similarly, to Fang et al (2015) the urbanization is divided into four subsystems: demographic urbanization, economic urbanization, spatial urbanization, and social urbanization. Table 3 shows 14 independent variables chosen to analyze the air quality in different cases of urbanization. All these variables are acquired from the CBS database, but some require some more explanation.

THE SUBSYSTEM OF URBANIZATION	INDEX	ACRONYM INDEX
DEMOGRAPHIC URBANIZATION	Total population	TP
	Western population (with migration background)	WP
	Non-Western population	NWP
ECONOMIC URBANIZATION	Average income	AI
	Total income	TI
	Business locations	BL
	Average house value	AHV
SPATIAL URBANIZATION	Land use	LU
	Population density	PD
	Population density categorical	PDC
	Distance to amenities (schools, doctor, supermarket, daycare)	DA
SOCIAL URBANIZATION	Private car ownership	PCO
	Private cars per unit of km2	PCM
	Energy consumption	EC

Table 3: Acronym explanation of independent variables.

Table 3 above shows all the acronyms for the independent variables. The main variables of interest are land use variables, population density, population density categorical, and private car ownership. The control variables are represented by average income, average house value, western population share (migrants), distance to amenities and energy consumption. These variables align with the independent variables from multiple studies in the empirical literature section (Fang et al, 2015)(Rodríguez et al, 2016). Next, all independent variables are discussed to get some more insight in what they represent precisely. The total population (TP) are all persons registered in the Netherlands with on an address. The Western population (WP) represents all persons migrated from parts of Europe (excluding Turkey), North America, Japan, Oceania, or Indonesia. The Non-Western population (NWP) are all person migrated from Africa, Latin-America, Asia (excluding Indonesia and Japan) and Turkey. The average income (AI) is the average income per person from the people that have a personal income who are participants of private households. Total income (TI) is the average income multiplied by the total population. Business locations (BL) are the (absolute) number of business in a municipality separated in categories namely: Agriculture/fishing (BL:AF), Energy (BL:E), trade/horeca (hotels, restaurants and cafes) (BL:TH), traffic/information/communication (BL:TIC), financial services (BL:F), business services(BL:B), culture/recreation and other services (BL:CRO). The average house value (AHV) is based upon the valuation of real estate (WOZ-waarde) that have at least a value above 0. Land use (LU) is separated in different categories by the CBS dataset namely: traffic areas (rail, road and air) (LU:T), built areas (work, shopping, culture and public amenities)(LU:B), semi-built areas (concrete areas that are not traffic of built areas)(LU:SM), recreational areas (LU:R), agricultural areas (LU:A), forest and open terrain (LU:F). Population density (PD) is the number of inhabitants per square kilometer (km2). The population density categorical (PDC) variable is separated into 5 categories in the CBS dataset based upon the number of address per square kilometer (km2). In case there are more or equal than 2500 addresses this is given the number 1 and can be categorized as a high dense city. The number 2 is given between 1500-2500 addresses per square kilometer (dense city), 3 is assigned in case the number of addresses is between 1000-1500 per square kilometer (larger town), 4 is given to the number of addresses per square kilometer between 500-1000 (medium-sized town) and finally the number 5 is given to the municipalities that have less than 500 addresses per square kilometer (smaller town). The distance to amenities (DA) variable represents the average distance for 90% of the inhabitants to the nearest doctors' office, major supermarket, daycare, or school. The average of these four figures are taken to get an overall number for distance to amenities. Private car ownership (PCO) represents the absolute value of privately owned motorized vehicles in a region. Private cars per square kilometer (PCM) are the number of privately owned motorized vehicles in a region per square kilometer. Energy consumption (EC) shows the average energy use of all house types in kWh (kilowatt hour). These variables represent a proxy for the compact city and special attention will be given toward population density (categorical), private car ownership, and land use. Although, the urban related variables proposed in Rodríguez et al (2016) would offer additional insights, these were not available for the Netherlands.

3.3 Dependent variables (Hypothesis 2)

The dependent variables used are also the emission values for all municipalities in hypothesis 1. These values will be mapped in GIS software to give a better overview of the current situation in the Netherlands in map 1, map 2 and map 3 at the start of the results section.

3.4 Independent variables (Hypothesis 2)

The only independent variable for hypothesis 2 is the distance between municipalities. This is being calculated by the ArcGIS system automatically when running Moran's I or the Hotspot analysis. Also, because it stretches further than comparing municipalities with each other, this method is on a regional scale. Note that the model does not just consider the distance to the neighboring municipality but all municipalities in the model.

3.4 Dependent variables (Hypothesis 3)

The data for hypothesis 3 is yet from another source. The main variable of interest in assessing urban morphology on air pollution, is air dispersion. Municipality level data does not provide sufficient insights into air dispersity on an inner-city level. To make a solid analysis, data on air dispersity levels in inner city environment is required. Air dispersions functions as a proxy for urban morphology/inner-city characteristics (Hang et al, 2009) The data used for this hypothesis is acquired from Luchtmeetnet.nl (Luchtmeetnet, ND). This organization measures air quality levels on a localized level in the Netherlands. They are funded by regional governments and municipalities. The data is publicly available from the website and offers observations until the start of 2020. The goal of this analysis is to observe an effect of dispersity location, depending on the wind variables, time, month, and location. Only the cities of Rotterdam and Amsterdam are used since these both have 10 measuring stations. Other cities like The Hague or Utrecht only have a few, which make comparing them harder. An issue with the data is that not every measuring stations collects all pollutants. Nitrogen (NO_2) is measured by the majority of the measuring stations. CO_2 is not measured at all because this pollutant does not pose a direct health risk. PM_{10} is measured at most of the locations, but the problem with this pollutant is that some measuring stations are near construction sites, which especially pollute

PM_{10}, this could result in a bias in the results. Therefore, Nitrogen (NO_2) is used as the only dependent variable. The data is on an hourly basis, extracted from the 1st of October 2019 until the 31st of December 2019. In addition, three months are chosen to make sure there are enough observations to get significant results and to reduce the standard errors. The measuring stations gave an average value (in microgram per m^3) for NO_2 on an hourly basis. Some values were missing occasionally due to measurement error of the stations.

Table 4 and 5 show the descriptive statistics for all measuring stations used for hypothesis 3. The mean ranges between 20-40 micrograms per cubic metre of NO_2. The dummy variable is shown based on the urban characteristics around the measuring location, which will be explained in more detail later. The minimum almost approaches 0 for every stations while the maximum value for every location ranges between 65-135 micrograms per cubic meter of NO_2. The number of observations differs for every measuring locations since the equipment is very sensitive and fails occasionally. A maximum of 2208 observations could occur, resulting from 3 months (92, October, November and December) multiplied by 24 as these are 1-hour values.

MEASURING LOCATION	MEAN	STANDARD DEVIATION	DUMMY	MINIMUM	MAXIMUM	OBSERVATIONS
SCHIEDAMSE VEST	31,65	14,09	1	4,73	84	2193
HOOGVLIET	26,49	14,73	0	0,2	79,2	2154
PLEINWEG	39,34	17,69	1	4,6	98	2200
ZWARTEWAALSTRAAT	28,20	14,47	0	4,4	80,6	2204
RIDDERKERK	34,54	16,63	1	3,4	135,4	2192
OVERSCHIE	34,21	15,32	0	1	94,8	2054
STATENWEG	42,26	18,73	1	4,8	129,3	2202
SCHIEDAM	32,32	15,33	1	2,1	86,7	2115
MAASSLUIS	30,32	13,82	0*	2,1	85,2	2204
BERGHAVEN	28,68	15,29	0**	0,6	89,8	2117

*Table 4: Mean, standard deviation, minimum and maximum values of NO_2 (microgram/M^3) for all measuring stations in Rotterdam. Sources: Luchtmeetnet.nl & KNMI *The visual characteristics of Maassluis are based on a picture from 2009 and is outdated. **The measuring station was not found on Google Street Maps.*

MEASURING LOCATION	MEAN	STANDARD DEVIATION	DUMMY	MINIMUM	MAXIMUM	OBSERVATION
HAARLEMMERWEG	40,83	18,03	1**	4,4	114,1	2179
NIEUWENDAMMERDIJK	23,39	13,68	0	0,1	67,5	2153
EINSTEINWEG	44,51	21,24	0	2,9	110,9	2117
VAN DIEMENSTRAAT	39,09	18,35	1	4,7	102,7	2204
VONDELPARK	24,97	13,29	0**	3,2	66,9	2203
STADHOUDERSKADE	35,00	15,30	1	4,6	99,6	2173
OUDESCHANS	28,53	15,22	1**	2,7	81,2	2002

JAN VAN GALENSTRAAT	38,33	17,65	1	4,2	118,9	2196
KANTERSHOF	21,89	12,15	0	2,2	65,4	2203
SPORTPARK OOKMEER	23,50	15,21	0	0	77,5	2206

*Table 5: Mean, standard deviation, minimum and maximum values of NO_2 (microgram/M^3) for all measuring stations in Amsterdam. Sources: Luchtmeetnet.nl & KNMI **The measuring station was not found on Google Street Maps.*

3.5 Independent variables (Hypothesis 3)

Hypothesis 3 also has independent variables that affect emission values indirectly. Wind is the main factor in determining the air dispersity of a location and therefore must be considered as independent variables in different forms. Other independent variables used are the hour of the day the emission values are recorded, the month in which the emission value was measured and a city dummy for Rotterdam and Amsterdam. Finally, the measuring locations are added to provide a fixed effect regression.

In addition to pollutant levels, the air dispersity must be measured. This is done by looking at the wind speeds and wind direction. However, the air pollutant measuring station do not include an instrument to measure wind. Therefore, data from the Dutch Weather institute (KNMI) is used (KNMI, ND). The wind speeds measured at Schiphol was used for Amsterdam and there was a wind measurement station in the center of Rotterdam. The wind data is publicly available from the website of KNMI. The wind speed and wind direction are, similar to the emission values, calculated on an hourly basis, giving an average wind speed (meters/second M/S). Although, these do not represent the air dispersity levels at a measurement station, no other weather station was closer to give a better indication. The methodology will discuss what measures are used to estimate air dispersity levels. Like pollutant levels, the wind speeds are measured hourly and extracted from the 1st of October until the 31st of December 2019. Table 6 below shows the descriptive statistics of the wind speed, wind direction in both Rotterdam and Amsterdam. The wind direction is based upon a circle gradient from 0-360, rounding off to the nearest 10. Wind direction can also take the value of 990, which occurs when the wind direction could not be determined with certainty. These values were dropped as they were causing a bias in the interaction term in the OLS regressions. Wind direction is also added to this dataset, as this is also found to be a relevant factor in determining air dispersion (Yuan, 2014). Some of these measuring stations are next to buildings, which might enhance wind direction as a relevant variable. The wind speeds are also rounded off to the nearest 10 per hour, and these numbers are measured by multiplying the meters per second times 10. For example, a value of 80 would indicate an average wind speed of 8 meters/second during that hour.

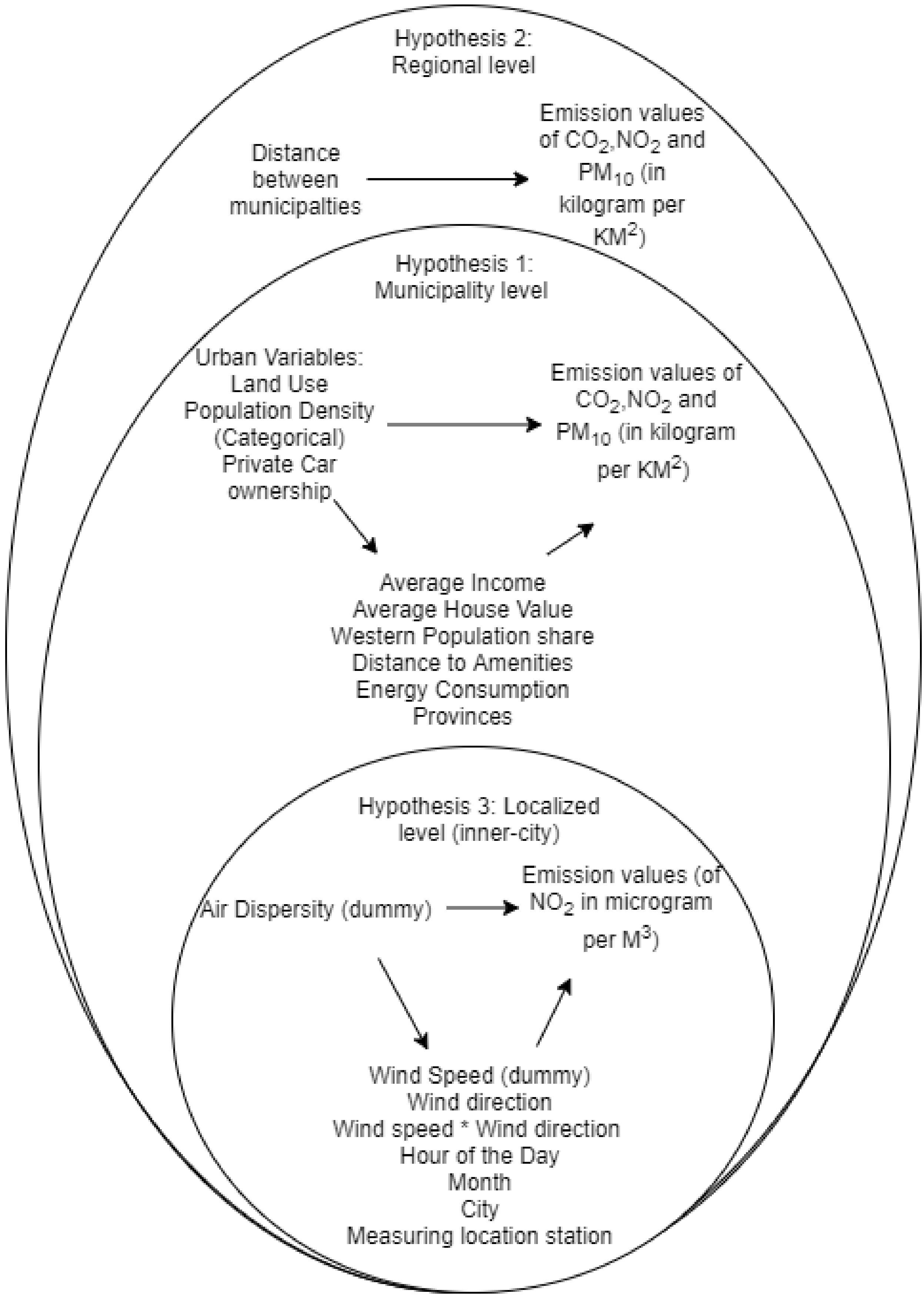
VARIABLES	OBSERVATIONS	MEAN	STANDARD DEVIATION	MINIMUM	MAXIMUM
WIND DIRECTION SCHIPHOL	2,184	178.7	75.5	0	360
WIND SPEED SCHIPHOL	2,184	49.9	25.3	0	130
WIND DIRECTION ROTTERDAM	2,184	178.6	75.2	0	360
WIND SPEED ROTTERDAM	2,184	45.3	24.1	0	120
WIND DIRECTION UNKNOWN	24	990	0	990	990

Table 6: Wind direction and wind speed in Rotterdam and Amsterdam in October, November and December 2019. Source: KNMI. Note: the direction is based on the gradient of a circle (0-360) and an unknown wind direction is defined as 990.

3.6 Connecting the Hypotheses

Diagram 1 below shows the relationship between all the hypotheses based on the scale level. The regional level data stretches throughout the entire Netherlands and indicate the spatial autocorrelation of the dependent variables (CO_2, NO_2 and PM_{10}). The regional level data is, as stated before, similar to the dependent variables used in hypothesis 2, the difference being that the methods used by the GIS systems also considers distance between region as a variable to explain emission values (of neighbouring municipalities). The municipality level data considers the differences in urban variables between municipalities to be the main factor to influence emission values within a municipality. The inner-city level data only regards the hourly NO_2 values as these were most widely available and deemed most important in inner-city environment by Cho & Choi (2014). These measuring locations can indicate the effect of urban planning on the air dispersion within cities in the municipalities.

Diagram 1: Relationship on scale level between all the hypotheses.



4. Methodology

This section will emphasize what statistical methods will be used to carry out the analysis. The methodology for each hypothesis differs and will be explained separately. Sources used in the empirical literature section will be the most important guideline in deciding what research methods are most relevant, also considering the type of data collected.

4.1.1 OLS regression (Hypothesis 1)

The Ordinary Least Squares (OLS) regressions are used as an indicator for hypothesis 1 to test whether the compactness of cities, indicated by the urban variables, have a significant influence on pollution values. The independent variables will be added to see what indirect effect each has on the pollution level. The simple OLS regressions will be made using the following formula. To tackle heteroskedasticity, robust standard errors are used in all the regressions.

$$Y_i = B_0 + B_i X_i + E_i$$

$Y_i = Pollutants\ per\ municipality$

$X_i = Independent\ variables\ per\ municipality$

Next, the fixed effect regression will be made. The fixed effect regression will be on the province level, looking at the within-province level variation. This will be done by adding a dummy variable for provinces in the OLS regression model. The reason why the fixed effect model might give more explanation to the urban development and pollution levels, is twofold. First, provinces themselves make policy beside the National government to reduce air pollution. Second, from literature by Fang et al (2015) it can be noted that dispersion and spatial distance have an effect on air pollution levels. According to Tobler's first law of geography, everything is related to everything else, but near things are more related than distant things (Tobler, 1970). So, by looking at the fixed effects within provinces, the regression will take into account some of the variation caused by spatial distances.

4.1.2 Multicollinearity and correlations (Hypothesis 1)

The first problem regarding the dataset for hypothesis 1 has to do with multicollinearity of the variables. For example, the absolute value of pollution will have a very strong relation with the total population of the region as humans are the number one cause for pollution. To counteract this, all the absolute pollution values per region are divided by the number of square kilometers within that region. This way density still matters with respect to pollution, but the regions are comparable to each other. However, after running the Variance Influence Factor (VIF) on the Ordinary Least Squares (OLS) regression with all variables indicated in Table 3, multicollinearity was still an issue. To make sure the independent variables are viable, all variables that had a VIF > 10 were dropped, as this is the threshold indicator for multicollinearity. The variables that were dropped due to this issue were total income, total population, business locations, private car ownership per KM^2 and non-western population share. Similar to Fang et al (2015), the Spearman correlation coefficients are computed with the independent variables that are left and is shown in Table A in the Appendix. Almost all these correlations were significant at a 1% significance level. The variables that remain are computed in the OLS model shown below. This model is used to run the statistical analysis which will be shown in the results

$$Pollutant\ per\ municipality_i = B_0 + B_1 Land\ use_i + B_2 Population\ density_i + B_3 Population\ density\ categorical_i + B_4\ Private\ car\ ownership_i + B_5\ Average\ income_i + B_6\ Average\ house\ value_i + B_7\ Western\ population\ share_i + B_8\ Distance\ to\ amenities_i + B_9 Energy\ consumption_i + Province_{ij}\ E_i$$

$$i = indicator\ for\ Municipality$$

$$j = Province\ in\ which\ municipality\ is\ located\ in\ (dummy\ 1 = Drenthe, 2 = Flevoland, \dots, 12 = Zuid - Holland)$$

4.1.3 Descriptive statistics (Hypothesis 1)

To give a better overview of the variables that will be used in the regressions, Table 7 below shows the descriptive statistics of all dependent and independent variables. The number of observations differ for average house value (AHV) and Land use (LU) since CBS was missing these in their dataset. However, since both variables are relevant in order to explain pollution, they were kept in the dataset and the total number of observations equals 331. As visible from the mean and standard deviation, the amount of CO_2 is much higher compared to NO_2 or PM_{10}. This is because some areas, like the industrial around surrounding Rotterdam, generate enormous amounts of CO_2. These outliers affect the standard deviation by a large margin. For the independent variables, some interesting things are noted. Land use-built areas (urban developed) (LU: B) has a lot of variation but overall should be a good variable to use in the analysis. On the other hand, energy consumption (EC) likely has too little variance to make a difference. Also, both the Private car variables show similar trends and arguably one must be dropped.

VARIABLES	OBSERVATIONS	MEAN	STD. DEV.	MINIMUM	MAXIMUM
CO_2/KM^2	342	6601626	20637430	142410.5	291290891
NO_2/KM^2	342	10040	15448.6	501.9	182510
$PM10/KM^2$	342	1109	1538.2	48.7	17407
WP	342	4824	11110.1	68	147614
AI	342	25.96	3.6101	19.7	46.7
AHV	334	230	58.9	121	588
LU:T	339	307	270.2	12	2137
LU:B	339	992	1038.5	37	11861
LU:SM	339	132	183.3	3	2077
LU:R	339	292	296.9	32	2697
LU:A	339	5645	6264.8	4	40239
LU:F	339	1382	2078.0	2	19537
PD	342	894.25	1037.9	25	6347
PDC	342	3.342	1.142	1	5
DA	342	.9773	.4059	.475	6.075
PCO	342	22953	28638.7	330	243390
EC	342	3081	328	2160	3990

Table 7: Descriptive Statistics of dependent and independent variables. Source: CBS & Emissieregistratie.nl Note: WP=Western Population (migrated), AI=Average income, AHV=Average House Value, LU=Land use, PDC=Population density categorical, DA=Distance to Amenities, PD= Population Density, PCM=Private Car per KM2, EC=Energy Consumption.

4.2 Spatial autocorrelation; Moran's I and Hot spot analyses (Hypothesis 2)

As indicated before, hypothesis 2 uses the same dependent variables for the analysis as hypothesis 1, so the descriptive statistics in Table 7 are also applicable to this hypothesis for CO_2/KM^2, NO_2/KM^2 and PM_{10}/KM^2.

As Fang et al (2015) warned about in her paper, the OLS might be biased due to spatial autocorrelation. They state that air pollution has a strong regional effect, also influencing neighboring cities or regions. The fixed effect within provinces captures some of this variation. However, this hypothesis gives even a better indication. Spatial autocorrelation creates a bias in the OLS regression, as the samples analyzed must be independent (Fang et al, 2015). To tackle this problem, Moran's I and hot spot analyses will show whether there is spatial autocorrelation and its significance.

Spatial autocorrelation is a fundamental property of all attributes located in space. Global Moran's I is a measure of spatial autocorrelation that indicate whether a variable exhibits significant spatial dependence and heterogeneity at a given scale – in this case, whether a pollutant value does so at the municipality scale (Fang et al, 2015). Global Moran's I can be expressed as:

$$Global\ I = \frac{\sum_{i=1}^{n}\sum_{j\neq i}^{n} W_{ij}(x_i-\bar{x})(x_j-\bar{x})}{S^2\sum_{i=1}^{n}\sum_{j\neq i}^{n} W_{ij}} = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\left(x_i-\bar{x}\right)\left(x_j-\bar{x}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}$$

$$Z = \frac{I-E(I)}{\sqrt{VAR(I)}} = \frac{\sum_{j\neq i}^{n} \mathrm{W}_{ij}(d)(x_j-\bar{x}_i)}{S_i\sqrt{\mathrm{W}_i(n-1-\mathrm{W}_i)/(n-2)}}\quad j\neq i$$

Source: Fang et al (2015)

where, n is the number of municipalities; X_i, X_j is the pollutant value of spatial location i, j; and S represents the standard deviation of the samples. The range of values of global Moran's I is (-1,1). The W_{ij} represents a spatial weight matrix, which measures the observations in the neighborhood for each location. The Z-score is used to analyze the significance of spatial autocorrelation. The ArcGIS features a Moran's I spatial dependence report which include the critical values at which point a cluster is significant. If the Z-value is significant, it means that there is an existence of either high-value or low-value clustering. If I and Z are both close to zero, with a high enough n, it could indicate spatial randomness.

The hot spot analysis is used to show where most spatial autocorrelation occurs. The formula to calculate these hot spots is created by Ord & Getis (1995) and used in many researches, also outside the economic scope. For the analysis, the optimized hot spot analysis is used, to create the best default settings regarding distance and the weight matrix. The formula created is very similar to Moran's I and will not be explicitly shown. An additional advantage of this tool in ArcGIS is that it creates two extra values for every observation with P-values and Z-values. This can be used for significance but also to show the distribution of clustering of pollutant values in the Netherlands.

4.3.1 Health risk scale of Nitrogen dioxide (NO2) (Hypothesis 3)

Hypothesis 3 is created to estimate the effect of inner-city urbanization on pollution. First, it an indicator is needed to estimate the public health risk of nitrogen dioxide (NO_2). The Air Quality Index compiled by the National Ambient Air Quality Standard (EPA, 2010) is used (Table 8). This index was made to reduce respiratory diseases caused by NO_2, especially next to major roadways. The scale is based on a 1-hour NO2 standard at the level of 100 parts per billion (ppb). This is a similar scale as the hourly values used for Hypothesis 3. This index cannot be used for the municipality level data in Hypothesis 1 & 2 since these are not 1-hour level values and these are also retrieved from a wider spectrum of sources. To get an idea how bad the current situation in the inner city in Rotterdam and Amsterdam is, all responses from every station will be counted by the scale in the Air Quality Index. The goal is to observe where the best air quality places are, and where the worst air quality spots are. The expectation is that higher values will be found for low dispersity location located next to major roads. In addition, this indicator is used to assess the current situation in the Netherlands on a inner-level and give some more insights into the reported values.

AIR QUALITY INDEX	IMPACTS ON HEALTH AND ADVICE
GOOD (0-50)	No health impacts are expected when air quality is in this range.
MODERATE (51-100)	Individuals who are unusually sensitive to nitrogen dioxide should consider limiting prolonged outdoor exertion.
UNHEALTHY FOR SENSITIVE GROUPS (101-150)	The following groups should limit prolonged outdoor exertion: • People with lung disease, such as asthma • Children and older adults
UNHEALTHY (151-200)	The following groups should avoid prolonged outdoor exertion: • People with lung disease, such as asthma • Children and older adults Everyone else should limit prolonged outdoor exertion
VERY UNHEALTHY (201-300)	The following groups should avoid all outdoor exertion: • People with lung disease, such as asthma • Children and older adults Everyone else should limit outdoor exertion.

Table 8: Health impact by nitrogen dioxide based on a 1-hour NO_2 standard at the level of 100 parts per billion (ppb) . Source: EPA (Environment Protection Agency, US government, 2010)

4.3.2 Urban morphology and air dispersity (Hypothesis 3)

The methodology used for hypothesis 3 is again different from the first two. As a reminder: Inner-city characteristics have a significant affect air quality levels at different wind speeds, is hypothesis 3. As hypothesis 3 is focused around the city characteristics, municipality level data is not sufficient. The Moran’s I analysis will tell little bit about whether pollutants are transferable over space, but it is hard to link this to the urban city characteristics. Within a municipality there might be many smaller towns, all having different characteristics. Since air dispersity is the variable that links urban characteristics and pollution together, only larger cities will be considered. This is due to the fact that air dispersity in larger cities is affected more by tall buildings. Rotterdam and Amsterdam both have 10 measuring stations measuring the air quality.

To know how the city characteristics look like at the vicinity of the measuring stations, Google Street View is used. A screen shot of every location is taken and can be seen in the Appendix after Table C. The measuring stations are circled in red. However, in 4 cases the stations were not visible on any of the locations on Google Street View and for Maassluis the picture was taken in 2009 and city characteristics likely have changed since then. Based on the visual characteristics of the picture, the description of Luchtmeetnet (ND) of the region, and the population density of the region, a dummy variable of 0 was allocated to high air dispersity regions and a dummy of 1 was allocated to low air dispersity regions. The reasoning behind this can be found in Table C in the Appendix. Table 9 below shows a description of all the different measuring stations. The stations are separated into three types by Luchtmeetnet (ND): S=city background station, V=traffic station and I=industry stations. All of these have different characteristics with respect to the cause of pollution but also air dispersion in their respective places. The dummy shows, based on visual characteristics from the photos, how good the air dispersion is likely to be around the measuring station, with 1 showing little air dispersion and 0 representing more open areas. Table C in the Appendix explains the reasoning behind the dummy variable choices and shows the visual characteristics. Admittedly, more tests must be done to measure these kinds dispersion levels, but these pollution measuring stations do not include wind measures or any other air dispersity tests.

ID	**NAME**	**STATION TYPE**	**DUMMY**	**DESCRIPTION**
		ROTTERDAM		
NL01418	Schiedamse Vest	S	1	Many people live at this location and there are no busy roads, ports, or industrial areas nearby.
NL01485	Hoogvliet	S	0	Many people live at this location and there are no busy roads, ports, or industrial areas nearby.
NL01487	Pleinweg	V	1	At this location, the air quality is largely determined by the emissions from road traffic. Pleinweg is a very busy street in Rotterdam.
NL01488	Zwartewaals traat	S	0	Many people live at this location and there are no busy roads, ports, or industrial areas nearby. High peaks at this station are currently caused by the demolition of a school building next to the measuring point.
NL01489	Nierkerk	V	1	At this location, air quality is largely determined by the emissions from road traffic on the A15 / A16 motorway.
NL01491	Overschie	V	0	At this location, air quality is largely determined by the emissions from road traffic on the A13 motorway.
NL01493	Statenweg	V	1	At this location, the air quality is largely determined by the emissions from road traffic. Bentinckplein is a very busy street in Rotterdam.
NL01494	Schiedam	S	1	Many people live at this location and there are no busy roads, ports, or industrial areas nearby.

NL01495	Maassluis	S/I	0*	At this location, there are mainly houses and air quality is affected by industry and shipping. Construction activities are currently taking place in the immediate vicinity of the monitoring station, which can occasionally lead to high peaks, especially PM_{10}.
NL01496	Berghaven	S	0**	Measuring station in Hoek van Holland
		AMSTERDAM		
NL49002	Haarlemmer weg	S	1**	At this location, many people live on a busy road with about 15,000 vehicles a day. What is special about this measuring station is the exact place where the sampling takes place. This is via a pipe directly on the edge of the parking spaces, after which the suction pipe goes via the ground to a basement with the measuring equipment. The suction height is 2m40. There is a count loop for road traffic at this location.
NL49003	Nieuwendam merdijk	S	0	Many people live at this location and there are no busy roads, ports, or industrial areas nearby.
NL49007	Einsteinweg	V	0	At this location, air quality is largely determined by the emissions from road traffic on the A10 motorway.
NL49012	van Diemenstraa t	V	1	At this location, the air quality is largely determined by the emissions from road traffic. The number of vehicles travelling here is about 15000 per day.
NL49014	Vondelpark	S	0**	Many people live at this location and there are no busy roads, ports, or industrial areas nearby. The Overtoom (closest street) is about 64 meters with an average of about 15000 vehicles per day.
NL49017	Stadhouders kade	V	1	At this location, the air quality is largely determined by the emissions from road traffic. The number of vehicles travelling here is about 22000 per day.
NL49019	Oudeschans	S	1**	Many people live at this location and there are no busy roads, ports, or industrial areas nearby. It is located in the old centre of Amsterdam
NL49020	Jan van GenStat	V	1	At this location, the air quality is largely determined by the emissions from road traffic. The number of vehicles travelling here is about 15000 per day.
NL49021	Kenerson	S	0	Many people live at this location and there are no busy roads, ports, or industrial areas nearby.
NL49022	Sportpark Ookmeer	S	0	Many people live at this location and there are no busy roads, ports, or industrial areas nearby.

*Table 9: Description of all measuring stations in Rotterdam and Amsterdam. Source: Luchtmeetnet.nl Note: V=traffic station, S=city background station and I=industry station. *The visual characteristics of Maassluis are based on a picture from 2009 and is outdated. **The measuring station was not found on Google Street Maps.*

4.3.3 OLS function urban characteristics on pollution level (Hypothesis 3)

Hypothesis 3 is again determined by using a multivariate OLS regression. The relevant variable in this case is the air dispersity dummy defined from Table 9 and Table C (Appendix). The dependent values in this case are the hourly values of NO_2 from all measuring stations (Table 9) and the control variables are wind speed, wind direction, wind speed dummy (>80 M/S), hour of the day dummy, month (October, November or December), city dummy (Amsterdam/Rotterdam) and measuring stations location dummy. The NO_2 values are reported in microgram/M^3 from Luchtmeetnet.nl on an hourly basis. Also, an interaction term between wind speed and wind direction is made to show whether these together could enhance each other. In some cases, the measuring stations might be positioned in such a way that wind from a certain direction is blocked off from a building, the interaction value might pick this up. This function used for the OLS regressions are given below.

$$Pollutant\ value_i = B_0 + B_1 Air\ dispersity_x + B_2 Wind\ speed_{ij} + B_3 Wind\ direction_{ij} + B_4\ Wind\ speed\ dummy + B_5\ Hour\ of\ the\ day_a + B_6\ Month_b + B_7\ City_j + B_8\ Measuring\ location\ station_c + B_9 Wind\ speed_{ij} * Wind\ direction_{ij} + E_i$$

$x = dummy\ high\ dispersity\ (0)\ or\ low\ dispersity\ (1)$

$i = hourly\ values\ of\ wind\ or\ pollutants$

$j = Dummy\ Rotterdam\ (1)\ or\ Amsterdam(0),/(Wind\ Rotterdam\ (0)\ or\ Schiphol\ (1))$

$a = dummy\ hour\ of\ the\ day\ 0 = midnight - 1AM, 1 = 1\ AM - 2\ AM\ etc$

$b = month\ dummy\ 0 = October, 1 = November, 2 = December$

$c = Dummy\ location\ of\ measuring\ station\ 0 = Schiedamse\ Vest, 1 = Hoogvliet\ etc$

5. Results

This section will present the results needed to answer the hypotheses and the general research question. This section will feature numerous graphs and maps to explain how the compact city distribution affect pollution levels. In addition, spatial correlation between densely populated cities will be analyzed. This can help with planning the network connections between cities, especially to tackle pollution caused by transportation. Furthermore, inner-city characteristics will be discussed by analyzing the effect of wind dispersion in Rotterdam and Amsterdam.

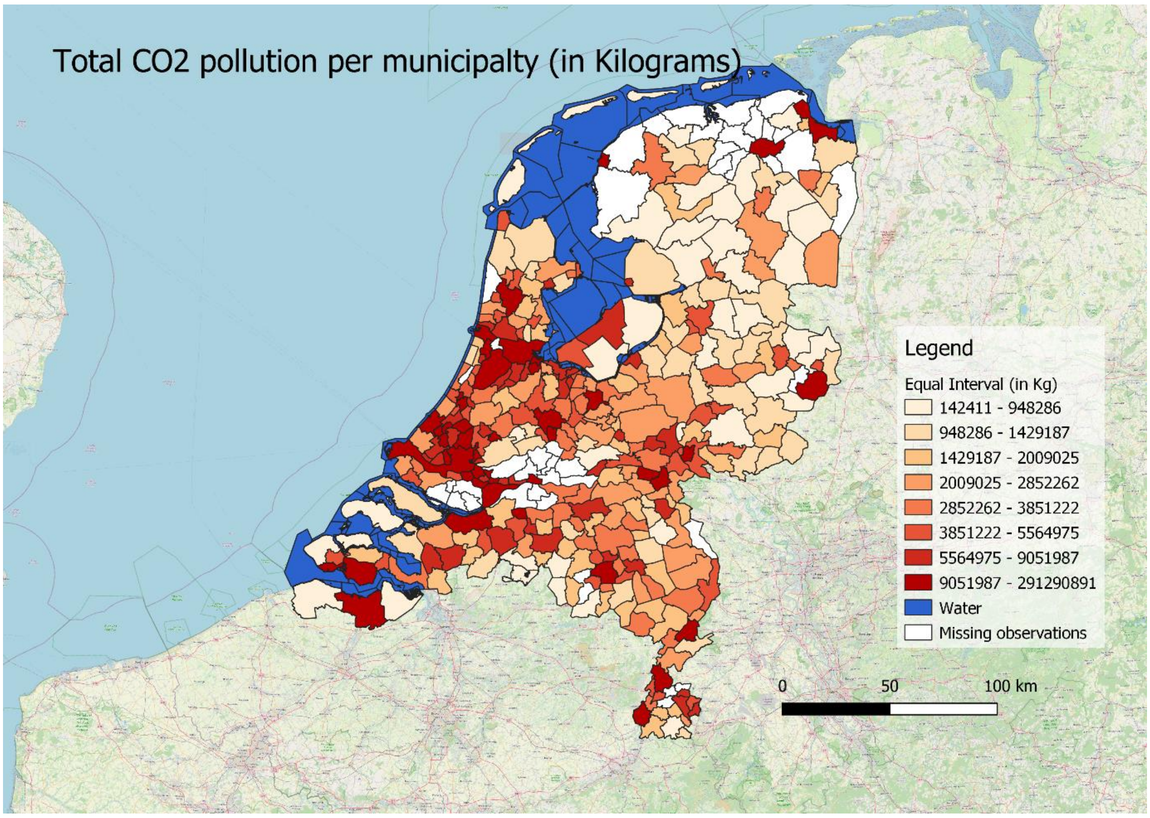
5.1 Maps of pollution values per municipality

This section is ahead from all other results because it gives a good overview of the current situation in the Netherlands. To get these results, the absolute number of CO_2 per municipality is divided by the square kilometers in that respective municipality. In addition, these values are the sum of all different causes identified by emission registration (Emissieregistratie, ND). Before these maps are shown, it must be noted that the classification of the pollution values is based on an equal interval, meaning that every category has about the same observations. This is done because the dataset contains some very high outliers. This is shown in the histograms in the Appendix (Figure A,B & C). These histograms are right-skewed and most of the values are closer to zero than the maximum value. Equal interval also has limitations. For example: these high outliers are relevant, not only because the population density in this regions is high but also because these are the largest cities. In addition, the maps do not show which areas pose a risk to public health. This is also due to the nature of the dataset and the scale used by the indexes that do explain public health risks. This municipality level data is unique because of the addition of the industry within every region. The indexes of public health mainly represent the health risks at a very localized level by using the 1-hour values (EPA, 2010), which mainly are restricted to pollution by consumers and local traffic, but not industry. This is the reason for using equal interval instead of a public health index. More limitations will be discussed in the limitations and recommendations for future research section.

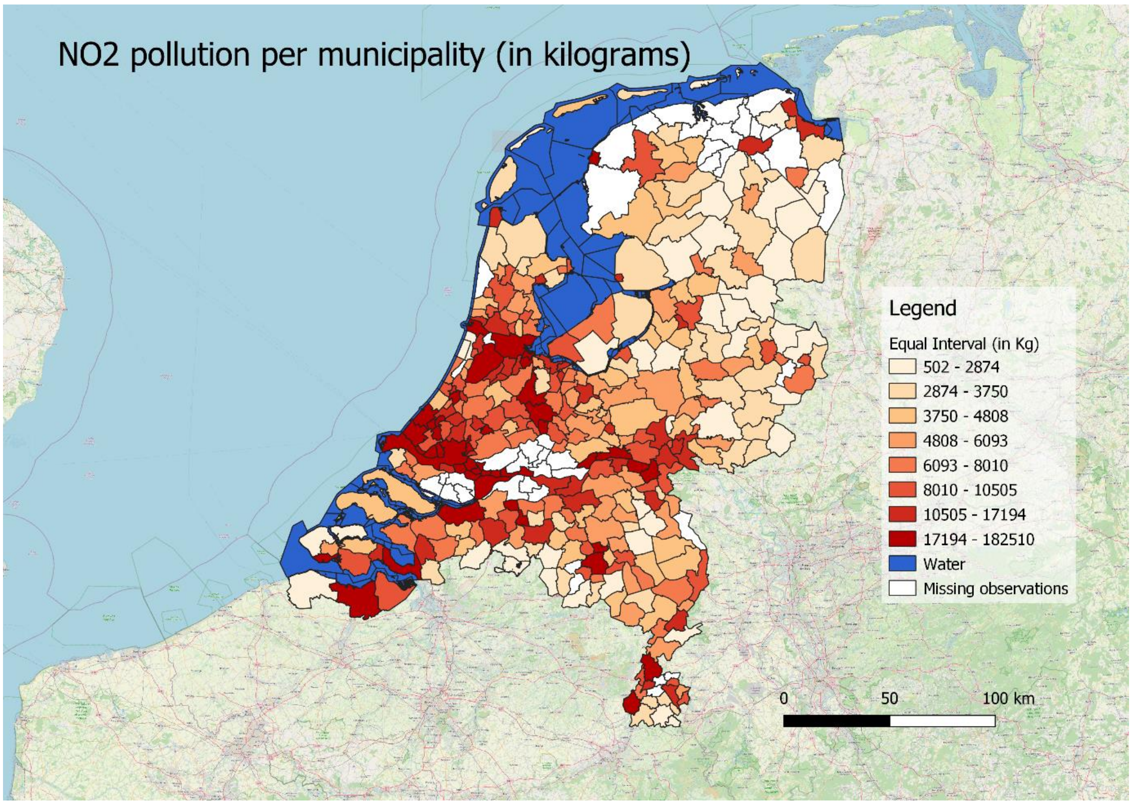
The map of total CO_2 pollution per municipality shows a cluster of the high values in the Randstad area. It also highlights larger cities in the Eastern part of the Netherlands. Nitrogen pollution shows a similar distribution of pollution values compared to CO_2, concentrating mainly around the Randstad area. The PM_{10} pollution values show high values in the Randstad areas but also in Limburg and Noord-Brabant. This is possibly because trucks pollute a large share of PM_{10} and most of the trucking of goods from the port of Rotterdam toward Germany goes through these provinces.

All these maps show a very high values for some regions in Zeeland, this is possibly because a large energy producer is located here. It must be questioned whether such a polluting company should be located in one of the least populated areas in the Netherlands, as this could have a destructive effect on the surrounding nature.

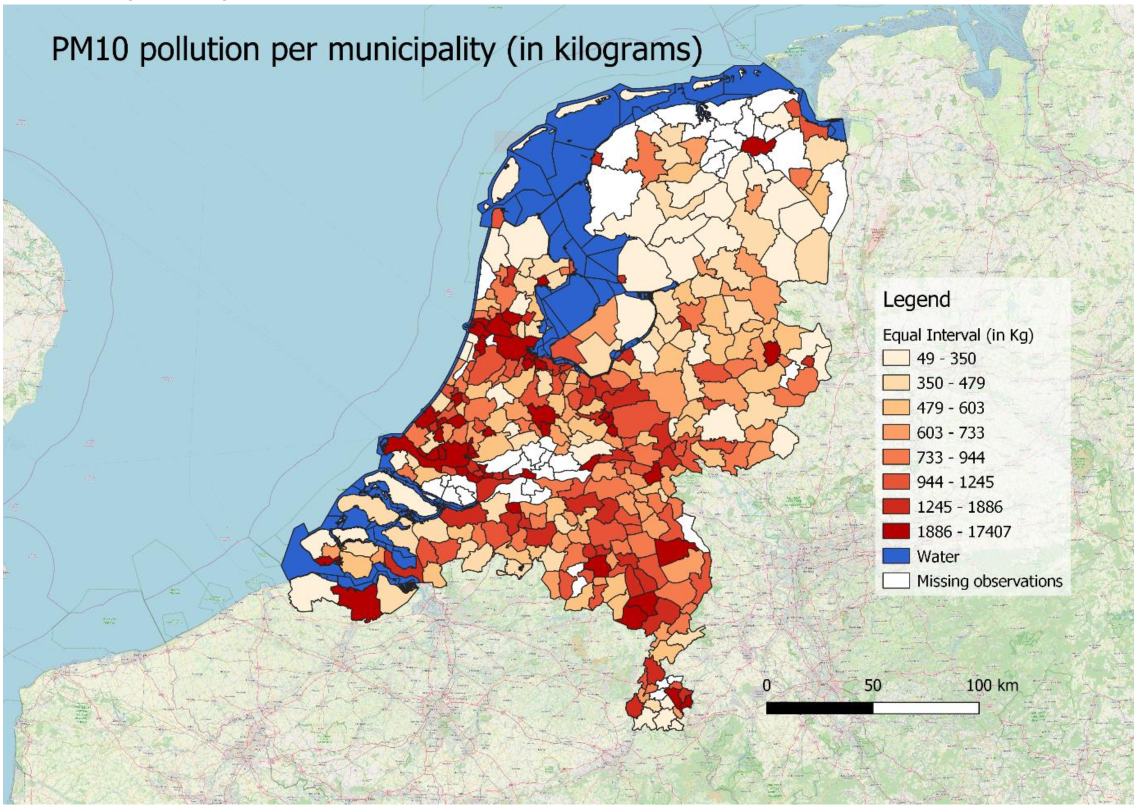
The white spots in every map represent the missing observations. The Emissieregristatie (ND) has values up and until 2017, while the GIS shapefiles were only available from 2018 (CBS). The two datasets were joined based on the name of the municipality. The problem is that some of these municipalities merged in the meantime and changed their municipality name and borders, showing as missing observations in these maps. Future research could update the dataset and add the missing observations to get a more complete map.



Map 1: Map of total (absolute) pollution values CO_2 (in kg),in the Netherlands per municipality using QGIS in 2017. Note: the classification uses an equal interval of about 42 values per category (338 total values). Legend units in kilograms (kg.



Map 2: Map of total (absolute) pollution values NO_2 (in kg) in the Netherlands per municipality using QGIS in 2017. Note: the classification uses an equal interval of about 42 values per category (338 total values). Legend units in kilograms (kg).



Map 3: Map of total (absolute) pollution values of PM_{10} (in kg) in the Netherlands per municipality using QGIS in 2017. Note: the classification uses an equal interval of about 42 values per category (338 total values). Legend units in kilograms (kg).

5.2.1 Analysis of urban variables (Hypothesis 1)

This section will focus on the effect of urban areas within a municipality on the pollution values of CO_2, NO_2 and PM_{10} (Hypothesis 1). This is done by computing OLS regressions, fixed effect regressions and separating the pollution values by industry. These regressions must formulate an answer on hypothesis one.

Before the regressions are carried out, it is important to identify what urban variables to look out for in the regressions. The main variables that would identify a compact city are the land use variables, private car ownership, population density and population density categorical. The other (independent) variables definitely have a certain value as indicated by the studied literature (Fang et al, 2015) (Rodríguez et al, 2016). First, Figure 3 below is created to show that other than population density, the categorical variable population density has been used to separate the different land usage of municipalities. Also, it shows that agriculture and built areas account for the largest share, but also the largest variation. Especially in high density cities, recreational areas, built areas and traffic areas are substantially larger than in other urbanized forms. The expectation is that the regressions will show positive values for all urbanized forms compared to the high density city. As this is a categorical variable, this first category will be the reference. Positive values are expected because the high dense cities represent the compact city, and therefore should generate lower pollutant values per square kilometer than the other more segregated cities. The variable land use and population density categorical are included as variables of interest. Additionally, population density, private car ownership are the other indicators for urbanization/compact cities.

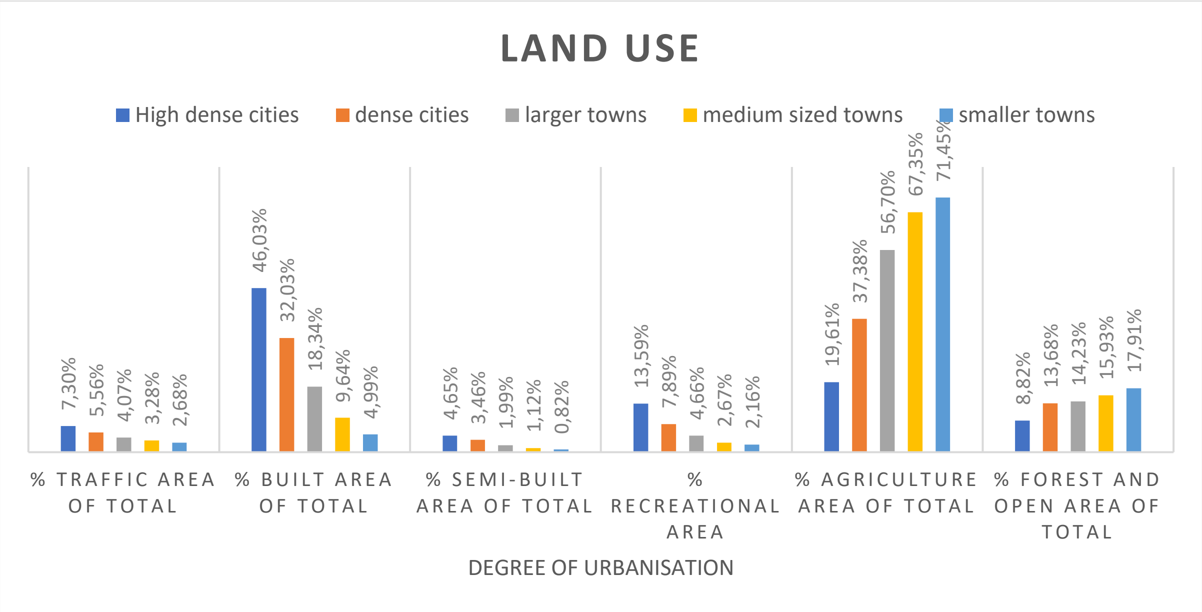


Figure 3: Land use by degree of urbanization (PDC variable). Source: CBS

5.2.2 Regression analysis pollution levels on urban variables (Hypothesis 1)

Table 10 below shows the simple OLS regressions with land use, population density, population density categorical, and private car ownership as the independent variables to see how significant they are, and to observe how they will change when adding other control variables. This regression is mainly for reference, however a few interesting things can already be noted. The only land use variables that show significant variables are from Agriculture and Forest. The only population density variables that show significant values for NO_2 and PM_{10}.

VARIABLES	(1) CO_2/KM^2	(2) NO_2/KM^2	(3) PM_{10}/KM^2
LAND USE TRAFFIC	2,166	14.70*	0.850
	(10,547)	(7.530)	(0.700)
LAND USE BUILT AREAS	3,888	0.306	0.0616
	(3,840)	(2.350)	(0.200)
LAND USE SEMI-BUILT AREAS	15,891	20.32	0.999
	(19,106)	(13.69)	(1.036)
LAND USE RECREATIONAL	4,886	-3.054	-0.463
	(19,363)	(12.24)	(1.015)
LAND USE AGRICULTURE	-646.1**	-0.789***	-0.0356**
	(289.7)	(0.195)	(0.0167)
LAND USE FOREST AND OPEN LAND	-723.7	-0.643*	-0.00407
	(568.7)	(0.353)	(0.0349)
POPULATION DENSITY	1,431	3.028**	0.710***
	(1,602)	(1.200)	(0.255)
POPULATION DENSITY CATEGORICAL = 2	6,203,355	4,020	874.7
	(5,732,822)	(3,761)	(573.9)
POPULATION DENSITY CATEGORICAL = 3	6,320046	1,534	622.2
	(4,357,753)	(3,407)	(517.7)
POPULATION DENSITY CATEGORICAL = 4	6,375,091	2,441	805.6
	(4,254,076)	(3,636)	(581.6)
POPULATION DENSITY CATEGORICAL = 5	7,243,486	2,309	741.1
	(4,751,687)	(3,800)	(604.8)
Constant	-4,196,395	3,623	-364.0
	(4,676,704)	(4,163)	(684.4)
Observations			
R-squared	0.173	0.270	0.214
Province dummies	No	No	No

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10: Regression Land use, Population density and PDC on pollutant values per square kilometer. Source: CBS & Emissieregistratie.nl

Next, the OLS regressions are made including all independent variables and fixed effects on a province level (Table 11). The dependent variable, emission values, differs per column and will be analyzed separately. The CO_2/KM^2 pollutant variable is explained positively significantly by land use-built areas and population density at a 1% significance level. This was expected as the number of people within a square kilometer increases, more pollution CO_2/KM^2 is registered. Similar logic can be applied to land use-built areas. Land use agriculture is negatively significant on a 5% significant level (column 4) and private car ownership at a 1% significance level. As Figure 3 shows, more agriculture land uses coincide with less built areas, and thus with lesser pollution of CO_2/KM^2. Private car ownership is negatively significant at a 1% level due to the fact that high density neighborhoods often mean more use of public transport. In addition, in these high density neighborhoods it is more likely that the average income is lower and people cannot afford to have a car. On the other hand, lower density neighborhoods has more likelihood of wealthier people while not much CO_2/KM^2 is found in these areas. This shows that land use and population density are relevant variables when indicating pollution within municipalities. The categorical variables for population density shows non statistically significant values but increase as the density becomes smaller. The results are similar when fixed effects provinces are added (column 7). NO_2/KM^2 shows the same significant variables when compared to CO_2/KM^2, also similar signs of the relationship (column 5 and 8). The only difference is a positively significant land use traffic variable on a 5% significant level without fixed effects. This completely disappears when adding the province dummy which indicates a spatial relationship. Again, for PM_{10} similar variables are significant, the only difference is a significant average income variable at a 5% significance level (column 6 and 9). This might be due to wealthier municipalities not wanting to live near large industry, which is one of the main contributors of PM_{10} pollution. In addition, residents living in poorer neighborhoods do not have the same options to move away from industry than wealthier residents. The significant variables land use-built areas and population density show a positive sign while the other significant values land use agriculture and private car ownership show a negative sign. In general, all models find a significant positive relationship with land use built areas, population density and significant negative relationship with land use agriculture areas and private car ownership. The R^2 fluctuates between 0.22 and 0.35, which is an understandable result since no meteorological variables are included. Implementing fixed effects (column 7, 8 and 9) resulted in almost identical results, especially for the variables of interest land use, population density, private car ownership, compared to the multivariate OLS models (column 4, 5 and 6). This shows that the presented results are robust and useful. So, the fixed effects improved the model slightly, mainly because the R^2 shows somewhat higher numbers compared to the OLS model.

VARIABLES	(4) CO_2/KM^2	(5) NO_2/KM^2	(6) PM_{10}/KM^2	(7) CO_2KM^2	(8) NO_2/KM^2	(9) PM_210/KM^2
LAND USE TRAFFIC	1,642	13.53**	0.859	559.5	9.244	0.587
	(9,303)	(6.110)	(0.651)	(11,402)	(9.548)	(0.699)
LAND USE BUILT AREAS	11,193***	5.483**	0.344*	12,504***	8.087***	0.493**
	(4,122)	(2.307)	(0.193)	(4,421)	(3.086)	(0.198)
LAND USE SEMI-BUILT AREAS	14,503	19.51	1.228	12,903	12.89	0.843
	(17,891)	(12.05)	(0.987)	(17,258)	(10.44)	(0.923)
LAND USE RECREATIONAL	19,366	6.301	0.419	16,526	4.184	0.305
	(25,709)	(15.95)	(1.331)	(25,655)	(16.21)	(1.342)
LAND USE AGRICULTURE	-867.1**	-0.953***	-0.0482**	-866.3**	-0.907***	-0.0407**
	(341.9)	(0.216)	(0.0190)	(336.7)	(0.221)	(0.0194)
LAND USE FOREST AND OPEN LAND	-1,019	-0.799	-0.0211	-713.8	-0.669	-0.0111
	(794.9)	(0.492)	(0.0443)	(643.6)	(0.409)	(0.0378)
POPULATION DENSITY CATEGORICAL = 2	6,157,543	4,494	812.6	6,256,396	4,004	817.4
	(5,463,938)	(3,692)	(546.3)	(5,194,196)	(3,409)	(579.0)
POPULATION DENSITY CATEGORICAL = 3	7,036,545	2,796	496.6	7,750,039	2,221	509.0
	(4,736,094)	(3,136)	(508.2)	(5,595,755)	(3,625)	(555.0)
POPULATION DENSITY CATEGORICAL = 4	8,187,174	4,848	593.4	9,620,600*	4,545	585.8
	(5,032,990)	(3,494)	(581.8)	(5,638,244)	(4,016)	(633.2)
POPULATION DENSITY CATEGORICAL = 5	9,322,217*	5,055	494.4	1.120e+07*	3,831	455.4
	(5,514,032)	(3,716)	(598.5)	(6,408,548)	(4,889)	(655.9)
POPULATION DENSITY	3,421***	4.214***	0.821***	2,993***	3.593***	0.765***
	(967.3)	(1.127)	(0.271)	(1,016)	(1.009)	(0.252)
AVERAGE INCOME	247,585	71.56	-75.91**	415,485	157.7	-87.54**
	(336,712)	(241.9)	(32.91)	(409,031)	(300.5)	(44.13)
AVERAGE HOUSE VALUE	-14,738	-4.006	2.465	-53,877	-35.08	1.573
	(27,578)	(17.48)	(2.303)	(34,086)	(21.62)	(3.170)
WESTERN POPULATION SHARE	-231.7	-0.158	-0.00199	-206.4	-0.120	-0.00111
	(308.5)	(0.191)	(0.0166)	(336.7)	(0.203)	(0.0177)
PRIVATE CAR OWNERSHIP	-354.4***	-0.244***	-0.0219***	-363.0***	-0.251***	-0.0212***
	(113.2)	(0.0688)	(0.00683)	(113.4)	(0.0653)	(0.00696)
DISTANCE TO AMENITIES	520,327	-73.17	-26.59	906,085	664.2	23.62
	(966,416)	(791.4)	(74.58)	(1,175,456)	(1,126)	(71.50)
ENERGY CONSUMPTION	279.6	-2.677	0.320	2,301	2.813	0.615*
	(3,159)	(3.362)	(0.240)	(6,162)	(3.005)	(0.359)
Constant	-1.169e+07	8,384	163.6	-1.843e+07	-5,149	-540.3
	(1.114e+07)	(13,569)	(1,406)	(1.501e+07)	(10,069)	(1,796)
Observations	331	331	331	331	331	331
R-squared	0.223	0.327	0.316	0.238	0.351	0.325
Fixed Effects (Province dummies)	No	No	No	Yes	Yes	Yes

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 11: Regression of all independent variables on pollution values per square kilometer, also with fixed effects. Source: CBS & Emissieregistratie.nl

To sum up, from the OLS regressions of the total pollution values, it is observed as land use agriculture increases, the number of pollutants significantly decrease. Private car ownership also shows a significant negative relationship. Land use built areas is significantly positive. As the population density increases, as will the pollutants increase significantly positive. The R^2 ranges between 0.22 and 0.35, improving the model slightly when adding fixed effects on a province level. Regarding the hypothesis, it is not rejected as the relevant urban variables show a significant effect at a 1% level.

5.2.2 Regression analysis traffic pollution levels on urban variables (Hypothesis 1)

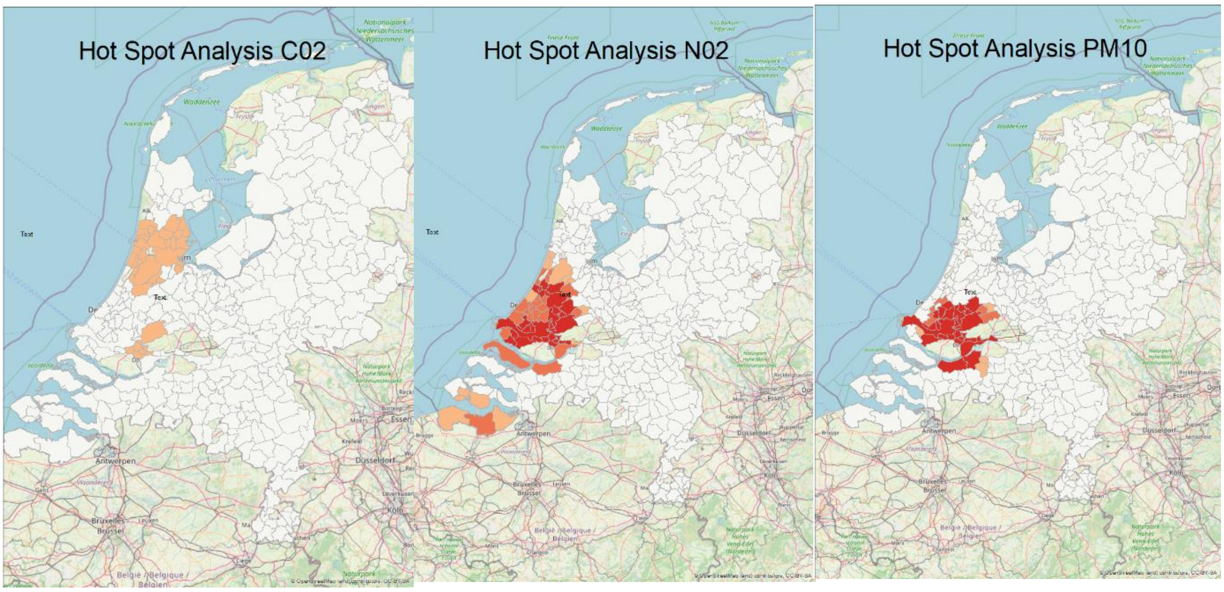
Since traffic and transportation has a large influence on all the pollutants, it would be interesting to see which urban variables affect pollutants caused by traffic and transportation. The regression of all independent variables on the values of pollution from CO_2, NO_2 and PM_{10} is shown in Table B in the Appendix. In this case, the population density categorical variable is negatively significant at a 10% level for all variables, in case of CO_2/KM^2 there are even some categories negatively significant at a 5% level. Population density, land use traffic are positively significant at a 1% level for CO_2/KM^2 while private car ownership and land use agriculture and land use forest are negatively significant at a 1% level. The R^2 for the CO_2/KM^2 regression is almost 0.73, explaining a lot more of the variation in CO_2/KM^2 than the total CO_2/KM^2 in Table 11. Another interesting development is the negative values for the population density category variable, compared to the highest density region 1. This under the condition that the other population density variable is still positive for all pollutants and significant for CO_2/KM^2. This result shows the exact opposite compared to Table 11 with respect to population density, meaning that traffic pollutants decrease as the population density category gets lower. Finally, distance to amenities is positively significant at a 1% level for NO_2 and PM_{10}, indicating that as the distance to public amenities becomes larger, the higher the pollutant values because of for example more car usage. The R^2 for NO_2/KM^2 and PM_{10}/KM^2 is much lower however at respectively, 0.199 and 0.286. This result was expected for PM10, as this pollutant is not primarily caused by traffic, except for container trucking. However, NO_2 is caused by traffic and therefore a higher R^2 and more significant values were expected.

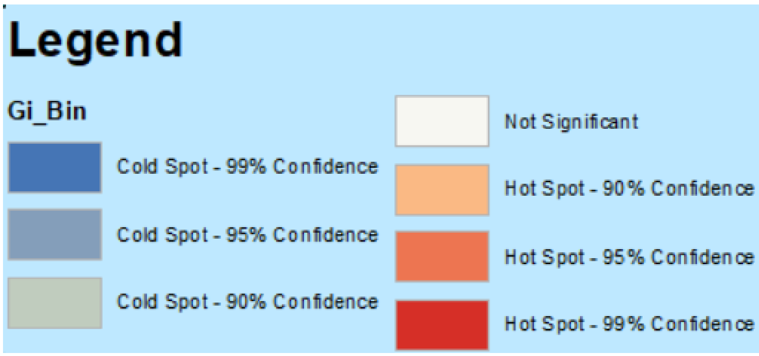
However, both the traffic pollution regression and the total pollution value regression results however fail to capture spatial characteristics. This creates a bias in the OLS estimation results (Fang et al, 2015). Also, an omitted variable bias could also occur as for example, weather variables are not included while they do influence both urban characteristics and the level of pollution values. The next section (Hypothesis 2) will try to analyze some of the spatial characteristics of pollutants.

5.3 Spatial autocorrelation air pollution (Hypothesis 2)

This section will discuss the ArcGIS results regarding spatial autocorrelation and assess whether regional variables can explain some of the variation of pollution values. Fang et al (2015) has identified the importance of spatial autocorrelation in China. She states that the results of the Moran's I model indicate that the air quality of a city can be attributed not only from the city itself, but also that of neighboring cities. Now, as Alpkokin (2004) and Priemus (1994) describe, the Randstad connects the major cities with very strong networks. As population density grows closer, the Randstad could melt into one megacity eventually and cause pollution to neighboring cities.

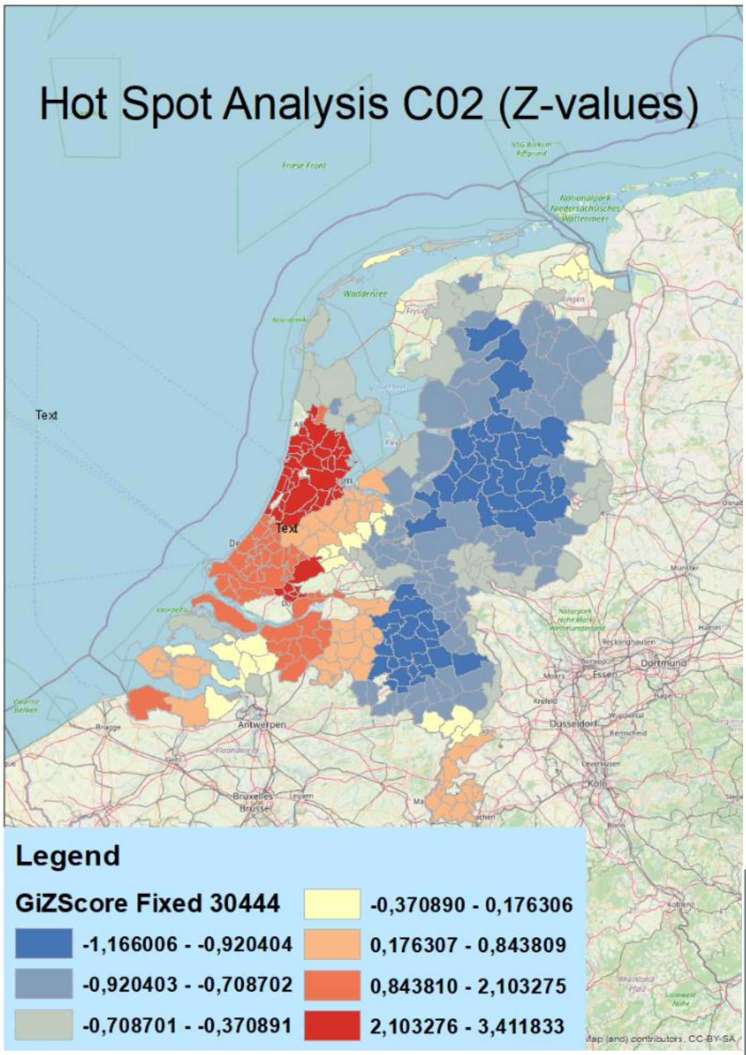
As stated before, the pollutants in this research all come from different sources. CO_2 mainly comes from energy production, transportation, and industry. NO_2 is mainly affiliated with road transportation. The main contributors of the PM_{10} pollutant were industry, energy usage and refineries. The maps created below show the hotspot analysis of all three pollutants. If a cluster is significantly higher compared to municipalities just outside of this cluster, it will become more red in that area. If a cluster is significantly lower than any other cluster or areas, it will become blue. For CO_2 per square kilometer (Map 4 below), significant values (5% level) are observed around Amsterdam and Rotterdam. The hotspot analysis of NO_2 shows a definite cluster around Rotterdam, the Hague and the roads leading to Amsterdam (Map 5 below). This cluster shows significant values on a 1% level, which means that all these areas have very high values because they are clustered in the Randstad. In addition, there are some significant values around an energy plant in Zeeland. PM_{10} shows significant values at a 1% level around Rotterdam (Map 6 below). This makes sense as most of the industry is located close to the port near Rotterdam, which is the main source for PM_{10}. However, as visible from this map, neighboring regions around Rotterdam also shows very significant values and this confirms that like Fang et al (2015) showed in China, neighboring regions are also affected by the industry within another region.





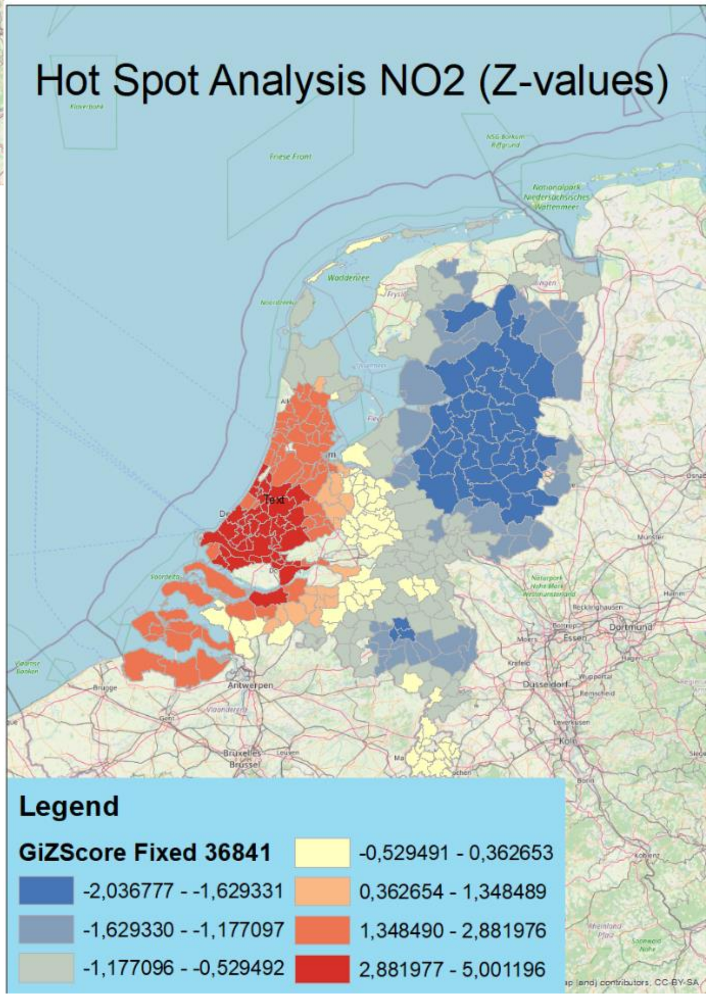
Map 4,5 & 6: Hot spot analyses of CO_2, NO_2 and PM_{10} in the Netherlands using ArcGIS. Source: Emissieregistratie. Note: Legend on the left is applicable to all maps.

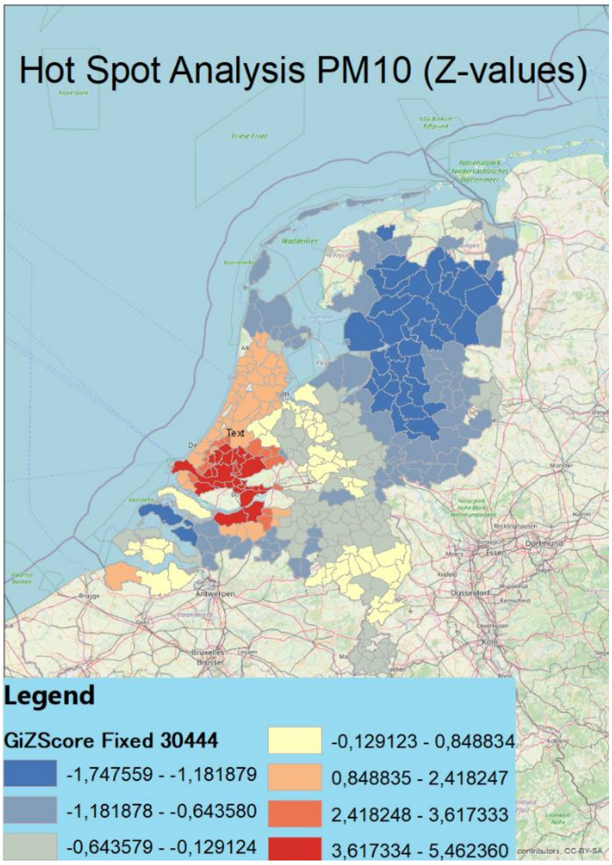
Map 7, 8 & 9 below show the Z-values of the same hotspot analysis in ArcGIS. From the legend, it can be observed how the Z-values are performing in categories and gives a better overview of the presented significant cluster from Map 4,5 & 6. From the Z-values of the hot spot analysis of CO_2 in Map 7, it can be noted that they are not very high in either direction. From Map 4 it was already visible that the values around Amsterdam were barely significant. The NO_2 values on the other hand, were very close to showing a negative cluster in the Eastern part of the Netherlands (Map 8). The PM_{10} Z-values in Map 9 show that the cluster in Limburg and Noord-Brabant is not that bad as initially thought from Map 3. In addition, PM_{10} shows the highest significance value for the cluster in the Randstad (Map 9). These maps clearly show what the effect is of a cluster like the Randstad on the pollution values of surrounding municipalities. For map 7,8 and 9 the classification of natural breaks by Jenks was used because this divided the Z-values evenly and accordingly. There are however also limitations of this method, as from these maps it cannot be observed whether these municipalities posed a public health risk. These limitations will be discussed in more depth in the limitations and recommendations for future research chapter.



Map 7: Hot spot analyses for all Z-values in the Netherlands. using ArcGIS. Source: Emissieregistratie. Note: the classification is based upon the natural breaks by Jenks. Note: The Legend consist of the confidence interval values of 99%, 95% and 90%, blue cold (negative) and red hot (positive).

Map 8: Hot spot analyses for all Z-values in the Netherlands. using ArcGIS. Source: Emissieregistratie. Note: the classification is based upon the natural breaks by Jenks. Note: The Legend consist of the confidence interval values of 99%, 95% and 90%, blue cold (negative) and red hot (positive).





Map 9: Hot spot analyses for all Z-values in the Netherlands. using ArcGIS. Source: Emissieregistratie. Note: the classification is based upon the natural breaks by Jenks. Note: The Legend consist of the confidence interval values of 99%, 95% and 90%, blue cold (negative) and red hot (positive).

Finally, the Moran's I statics is computed in ArcGIS to show whether there is clustering in the sample and how significant the clusters are. Again, the analysis is computed for CO_2, NO_2 and PM_{10} per KM^2 and is based on the equal interval presented in Map 1,2 & 3. The statistical summary of the Moran's I has been shown in the Appendix (Figures F, G and H). The CO_2/KM^2 values are clustered and significant at a 5% level showing a Z-value of 2.24 (P=0.025). The NO_2/KM^2 values are also clustered and significant at a 1% level displaying a Z-value of 6.85 (P=0.000). The PM_{10}/KM^2 values are clustered and significant at a 1% level showing the highest Z-value of 11.10 (P=0.000). This indicates that there is a significant spatial dependence between regions. Industry in the neighborhood of Rotterdam should consider the social health risks of nearby municipalities.

Regarding the compact city, the larger cities in the Randstad (Utrecht, Amsterdam, Rotterdam, and the Hague) are running out of space to build. Rotterdam and the Hague have built so extensively that there is almost no land between them anymore and will merge into each other in the coming years. Agriculture and the green places are getting more scarce in this area. The municipalities must use urban planning to their advantage by using the positive effects of the compact city. Starting with creating an efficient public transport system between all these cities since transportation and the ownership of private vehicles account for a large share of the pollutants. Regarding the hypothesis, it is not rejected, as clusters around Randstad exist for NO_2/KM^2 and PM_{10}/KM^2 around Rotterdam and a cluster of CO_2/KM^2 is shown around Amsterdam.

5.4.1 Air Quality Index inner-city level (Hypothesis 3)

This section will show the results of hypothesis number three. This question is more focused toward the elements in a an inner-city environment as well as urban morphology that affect air pollution or in most cases air dispersion. From the empirical literature section, it can be stated that the urban morphology has an influence on air dispersion (Yuan et al, 2014) (Wang et al, 2018). Therefore, it is assumed also in this case that air dispersity can be represented as a proxy for inner-city characteristics as well as morphology. The following hypothesis is defined to examine the Netherlands.

Hypothesis 3: Inner-city characteristics in Rotterdam and Amsterdam have a significant effect on air quality levels at different wind speeds

Urban morphology can influence air quality/air dispersion by the distance buildings are standing apart, the height of neighboring buildings, even tree types can have an effect (Wang et al, 2018) (Yuan et al, 2014). In addition, urban morphology also explains the historic development of neighborhoods. American streets are in general very wide while as described by Wang et al (2018) Asian inner cities consist of canyons from high buildings. The goal of this section is to measure the effect of these inner-city characteristics in Rotterdam and Amsterdam. Urban planners could use these results to better organize city centers to prevent societal health issues or smog. To get a good overview of the severeness of the current situation, the Air Quality Index is used, which indicates the health risks based on the 1-hour NO_2 values.

The Air Quality Index (Table 12) below shows the number of hours from the 1-hour values at every location. All of the location has a total of about 2200 values. The expectation is that at low air dispersity locations, the number of 1-hour observations which are at moderate or unhealthy, are higher. Table 12 below shows that there were no observations that were unhealthy or very unhealthy within the last quarter of 2019. In addition, the majority of the observations shows a good air quality. In 15,46% of the total observations the air quality was moderate and in 0,134% of the observations the air quality was unhealthy for sensitive groups. Overall, this is quite a good result considering these values were retrieved in inner-city environments and near roads. Comparing this with the air dispersity dummy, it can be found the low air dispersity values (dummy=1) show higher numbers, especially in the moderate category. The expectation is Einsteinweg, located next to the A10 highway. It was denoted as a high air dispersity location, while it does show the most observation in both the moderate and unhealthy for sensitive groups categories. However, this is mainly an indicator and fails to consider urban characteristics/all independent variables. Therefore, the next part will feature the OLS regressions with all independent variables.

AIR QUALITY INDEX	GOOD (0-50)	MODERATE (51-100)	UNHEALTHY FOR SENSITIVE GROUPS (101-150)	UNHEALTHY (151-200)	VERY UNHEALTHY (201-300)	DUMMY A DISPERSITY
SCHIEDAMSE VEST	1946	247	0	0	0	1
HOOGVLIET	1990	164	0	0	0	0
PLEINWEG	1618	582	0	0	0	1
ZWARTEWAALSTRAAT	2033	171	0	0	0	0
RIDDERKERK	1840	343	9	0	0	1
OVERSCHIE	1742	312	0	0	0	0
STATENWEG	1542	648	12	0	0	1
SCHIEDAM	1838	277	0	0	0	1
MAASSLUIS	1997	207	0	0	0	0*
BERGHAVEN	1982	195	0	0	0	0**
HAARLEMMERWEG	1555	615	9	0	0	1**
NIEUWENDAMMERDIJK	2041	112	0	0	0	0
EINSTEINWEG	1291	806	20	0	0	0
VAN DIEMENSTRAAT	1592	609	3	0	0	1
VONDELPARK	2076	127	0	0	0	0**
STADHOUDERSKADE	1820	353	0	0	0	1
OUDESCHANS	1782	220	0	0	0	1**
JAN VAN GALENSTRAAT	1695	496	5	0	0	1
KANTERSHOF	2139	64	0	0	0	0
SPORTPARK OOKMEER	2056	150	0	0	0	0

*Table 12: Air quality scale index by hour value. Source: Luchtmeetnet & EPA Note: *The visual characteristics of Maassluis are based on a picture from 2009 and is outdated. **The measuring station was not found on Google Street Maps. Dummy 1= low air dispersity, 0=high air dispersity. Note: the Dummy Dispersity was taken from the categorisation of table 9.*

5.4.2 Regression analysis inner-city pollution levels (Hypothesis 3)

Table 13 below shows the results from the OLS regressions when estimating the effect of air dispersity on the 1-hour values of NO_2. The NO_2 values are reported in microgram/M^3 from Luchtmeetnet.nl on an hourly basis. More details can be found in Table 4 and Table 5 in the Data section. Column 10 of Table 13 shows the relationship when the relevant independent variable is the only variable to explain NO_2. Air dispersity shows a positive significant result at a 1% level. Since this variable is a dummy, average NO_2 values will increase by 7.75 when the air dispersity is denoted as a low air dispersity location (Table 9, Table C Appendix) compared to when the air dispersity is high. In Column 11, the control variables for wind are added as well as the city dummy (Rotterdam =1). As expected, the wind speed is significantly negative at a 1% level, which means that at higher wind speeds there is almost no NO_2 pollution. The wind speed dummy (above 80 (M/S)) is significantly positive, meaning that if the wind speed is above 80 (dummy =1), the NO_2 value will increase significantly. Wind direction and the interaction value between wind speed and wind direction are insignificant. The city dummy is significantly negative at a 1% level, signifying differences between Rotterdam and Amsterdam. The air dispersity variable is also significantly positive in column 11 at a 1% level, increasing the NO_2 values with 7,81 in a low air dispersity location. Column 12 is added to control for the time of the day, and the month these variables were obtained in. The time of the day shows significant values at a 1% level for almost all possible hours. Compared to the reference category (midnight – 1AM), the largest significant positive results are observed during the traffic peak hours between 8 AM and 10 AM and between 4 PM and 8 PM. This shows that traffic is a relevant variable in explaining the variation in NO_2 at a local scale and time explains much of the variation in NO_2 values. The month variable also shows positive significant values compared to the reference category October. December is much worse compared to October, which most likely has to with more traffic jams in worse weather, leading to more pollution. Again, the air dispersity value is significantly positive at a 1% level in column 12, increasing the NO_2 values with 7,80 in a low air dispersity location. In the final model, a dummy for every location is added, this is not shown as Table 13 would become too long. The air dispersity dummy is still significantly positive at a 1% level with value for NO_2 of 14.76 in a low air dispersity location. Although this seems like a large jump from the other models, the variation in the NO_2 values is now more accurately explained within all separate locations. This is also true for the city dummy, which also made a large jump, but remains significantly negative. The location dummy therefore adds a lot of value, as it makes the other values more accurate, but also because the R^2 increases. The first model (Column 10) has a R^2 of 0.05, Column 11 shows a R^2 of 0.23, the model in Column 12 a R^2 of 0.37 and finally the model with locations dummies (Column 13) has the highest

R^2 of 0.46. This increase signifies that all control variables added, contribute to explaining the variation in the 1-hour NO_2 values.

VARIABLES	(10) NO_2	(11) NO_2	(12) NO_2	(13) NO_2
Air dispersity dummy	7.745***	7.805***	7.804***	14.76***
	(0.162)	(0.146)	(0.132)	(0.394)
Wind Speed		-0.299***	-0.293***	-0.295***
		(0.00946)	(0.00893)	(0.00831)
Wind Direction		-0.00400*	0.000179	-0.000174
		(0.00229)	(0.00214)	(0.00199)
Wind Speed * Wind Direction		-0.0000734	-0.000262***	-0.000241***
		(0.0000475)	(4.55e-05)	(4.27e-05)
City dummy		-0.546***	-0.646***	-8.132***
		(0.147)	(0.133)	(0.381)
Windspeed dummy		4.078***	4.885***	4.898***
		(0.276)	(0.261)	(0.245)
1 AM – 2AM			-2.432***	-2.426***
			(0.383)	(0.382)
2 AM – 3AM			-4.047***	-4.037***
			(0.379)	(0.380)
3 AM – 4 AM			-4.824***	-4.806***
			(0.379)	(0.379)
4 AM – 5 AM			-3.698***	-3.682***
			(0.391)	(0.386)
5 AM – 6 AM			-0.480	-0.456
			(0.409)	(0.390)
6 AM – 7 AM			4.935***	4.959***
			(0.458)	(0.431)
7 AM – 8 AM			9.971***	10.00***
			(0.483)	(0.456)
8 AM – 9 AM			12.41***	12.44***
			(0.488)	(0.462)
9 AM – 10 AM			11.19***	11.20***
			(0.461)	(0.435)
10 AM – 11 AM			10.09***	10.13***
			(0.441)	(0.412)
11 AM – Noon			8.327***	8.342***
			(0.422)	(0.393)
Noon – 1 PM			7.275***	7.279***
			(0.417)	(0.387)
1 PM – 2 PM			7.112***	7.122***
			(0.419)	(0.388)
2 PM – 3 PM			7.023***	7.024***
			(0.428)	(0.395)
3 PM – 4 PM			8.987***	8.982***
			(0.447)	(0.412)

4 PM – 5 PM			11.26***	11.26***
			(0.448)	(0.415)
5 PM – 6 PM			13.20***	13.21***
			(0.435)	(0.406)
6 PM – 7 PM			13.22***	13.23***
			(0.434)	(0.410)
7 PM – 8 PM			12.06***	12.06***
			(0.426)	(0.405)
8 PM – 9 PM			9.926***	9.932***
			(0.414)	(0.397)
9 PM – 10 PM			7.752***	7.755***
			(0.418)	(0.404)
10 PM – 11 PM			5.340***	5.352***
			(0.406)	(0.394)
11 PM – Midnight			3.105***	3.117***
			(0.400)	(0.390)
November			5.615***	5.657***
			(0.159)	(0.145)
December			7.513***	7.549***
			(0.161)	(0.149)
Constant	28.51***	42.98***	33.03***	28.34***
	(0.111)	(0.355)	(0.432)	(0.471)
Location Dummy	No	No	No	Yes
Observations	43,331	43,331	43,331	43,331
R-squared	0.051	0.226	0.367	0.461

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13: OLS regressions of air dispersity dummy on the 1-hour values of NO_2, also with control variables. Source: Luchtmeetnet.nl & KNMI. Note: Air dispersity 1=low dispersity location 0=high dispersity location, city dummy 1=Rotterdam, 0=Amsterdam, Wind speed dummy 1=above 80 (M/S) 0=below and equal to 80 (M/S), month October is reference, hour of the day midnight – 1AM is reference.

The hypothesis described above is not rejected as the inner-city characteristics represented by the air dispersity dummy is significantly positive at a 1% level in all models. This means that if a location were denoted as low disperse, it would add significant explanation, controlling for wind variables, city dummy, time of day, month and the location. The R^2 increases as control variables are added and shows the highest number of 0.46 when the dummies for every location are added.

6. Discussion and Conclusion

This section will formulate an answer on the proposed research question based on the results from all the hypotheses. In addition, policy recommendations for Dutch urban planning will be discussed. Also, a discussion will follow whether the urban morphology features of a city are the solution to the problems being sketched in the introduction. The overall question is as following:

Are compact cities associated with air pollution in the Netherlands?

The results suggest that developing compact cities are associated with increasing pollution values based on the OLS regressions and the Moran's I. This is dependent on the significant urban variables population density, private car ownership and the land use within municipalities. It is also significantly associated with spatial characteristics and thus the location within in the Netherlands. The compact areas around the 'Randstad' showed higher air pollution values around Rotterdam and Amsterdam. Also, urban variables on a local scale can affect pollution values through air dispersity. The compactness of cities is associated with varying levels of air dispersity.

To formulate a comprehensive answer to the research question, the compact city first must be defined. Neuman (2005) is used to identify the most important characteristics of the compact city. Regarding the Netherlands, the literature is focused toward a larger compact city, being the area between Utrecht, Amsterdam, the Hague, and Rotterdam called 'de Randstad' (Priemus, 1994) (Alpkokin et al, 2004). Within the Randstad the networks are strong in terms of industry, traffic, public transport, and trade. This would mean that the compact city stretches further than just the major cities of Rotterdam of Amsterdam.

The methods used to answer the research question are based on three levels: the regressions based on municipality level data to observe what the effect is of urban variables on pollution levels, the regional effect using Moran's I and hotspot analysis and, on a local level the measuring stations within a city analyzing the effect of urban morphology on air dispersion, which in turn affects air pollution levels. The municipality level data has been retrieved from the Dutch Central Bureau of Statistics (CBS) while the pollution values were obtained from Emissieregistratie (ND), a governmental funded data collection organization. These datasets can by combined by municipality to make useful regressions. For the third hypothesis, wind data is retrieved from the Dutch Weather institute (KNMI) and the pollution values for measuring stations in Amsterdam and Rotterdam are retrieved from Luchtmeetnet (ND), also a locally governmental funded data collection organization.

Hypothesis 1: Hypothesis 1: Land use, population density and private car ownership have a significant effect on emission values

The OLS regressions analysing the effect of urban variables on CO_2, NO_2 and PM_{10} per square kilometre show several significant observations. Land use-built areas (positive), land use agriculture (negative), population density (positive) and private car ownership (negative) are significant at a 1% level for almost all models. Population density categorical, showed some significant observations on a 10% level, increasing the pollution values as municipalities got less dense. Comparing these to the compact city characteristics by Neuman (2004) parallels can be observed. Neuman (2004) stated the compact city was characterized by high residential and employment densities, mixture of land uses, and multimodal transportation. These features by Neuman (2004) are very similar to the significant urban variables described from the OLS regressions. Fixed effect on a province level is added to see what the variation is within a province. These results look very similar without the province dummy, strengthening the correlations of the OLS results. The problem of these OLS regressions lie in the spatial autocorrelation of the data. This hypothesis is not rejected as land use, population density and private car ownership show a significant effect at a 1% level.

Hypothesis 2: The dependent variables (air pollution values) show significant clusters

To show how the spatial autocorrelation of the dependent variables looks like, the Moran's I and hot spot analysis were carried out. The Moran's I showed significant clustering at a 5% level for CO_2 and significant clustering at a 1% level for NO_2 and PM_{10}. To visualize these findings, the hot spot analysis maps showed significant clusters at 10% levels around Amsterdam, while NO_2 and PM_{10} showed significant clusters at 1% levels around Rotterdam. This proves that the data contains spatial autocorrelation, creating a bias in the OLS regressions. Although this is regarded as a large issue, it also means that the literature from more than 15 years ago of Alpkokin et al (2004) and Priemus (1994) is still relevant. Clusters in the dependent variables are still centred around the Randstad and thus confirm that the compact city, in this case the "Randstad", is associated with pollution on a regional level. The hypothesis is not rejected based on these findings.

Hypothesis 3: Inner-city characteristics in Rotterdam and Amsterdam have a significant effect on air quality levels at different wind speeds

Empirical literature clearly states that air dispersion is a relevant factor in determining pollution values in inner-city environments. In addition, the literature continues describing how urban morphology influences air dispersion (Yuan et al, 2014) (Hang et al, 2012). However, these studies analyse inner city environments in the Chinese regions based on computer simulations. This paper uses data from actual measuring stations in inner city environments of Rotterdam and Amsterdam, comparing them to suburban environments. The visual characteristics, location of the stations and population density in the nearby region determine whether the station is likely to have low or high air dispersion. Using the hourly wind statistics from KNMI, it can be analysed what the effect is of these locations. The OLS regressions shows significant positive results for the relevant independent variable air dispersity at a 1% level in all models. The R^2 increased when adding control variables, which adds to explaining the variation in the 1-hour NO_2 values. The control variable for wind was as expected significantly negative at a 1% level. Also, the time of the day is significant at a 1% level and shows a higher pollution value at high traffic hour peaks. These results show that the using air dispersity as a proxy for inner-city characteristics as well as urban morphology works. Therefore, this hypothesis is accepted. Meaning that at a local scale the density of the buildings and the historic urban planning are relevant factors in determining air dispersion. Hypothesis 3 is not rejected as the inner-city characteristics represented by the air dispersity dummy is significantly positive at a 1% level in all models.

Regarding policy recommendations, the results show that change is needed to create a healthy living environment in the future. As the clusters are shown around the Randstad, these municipalities should consider working together to reduce air pollutants from transport between these cities. As the traffic hours in hypothesis 3 have proven to generate higher values of NO_2, spreading or avoiding these peaks would be beneficial. Empirical research by Gibson et al (2015) has shown using a natural experiment in Milan that road pricing could reduce the air pollutants and create large welfare gains. These systems could be introduced in high traffic areas in the Randstad region. In addition, the industry and high population density near Rotterdam heavily affects other regions as well as seen in hypothesis 2. Therefore, these municipalities could implement a Coase theorem system whereby permits of CO_2 pollution must be bought from competitors (Coase, 1960). Rose & Stevens (1993) show that the Coase theorem shows efficient outcome when examining the efficiency of alternative assignments of marketable permits for CO_2. Also, the impact of the welfare gains is considered. Finally, population density shows positively significantly pollution values in hypothesis 1. Therefore, creating efficient housing and public transportation is important since private car ownership is significantly negative. This indicates that many car owners drive toward high polluted areas, but do not live there. More efficient public transport could reduce the need for private car ownership and create long-term public welfare gains. Also, a car-free inner-city could reduce air pollution values caused by the inability of pollution to escape because of low dispersity areas. Simulations of these inner-cities can be used to test the effects of wind, also when considering a car-free city centre.

Finally, from the results it can be concluded that compact cities are associated with air pollution levels in the Netherlands. Densely populated areas such as the Randstad show high values of pollution due to industry, large traffic bottlenecks and significant spatial clusters. On the other hand, the air quality index showed few observations that indicated a public health risk. The inner-city characteristics in the Netherlands are significantly affected by the air dispersity levels and this means that urbanisation is affecting pollution. However, the buildings in the Netherlands cannot be compared to the canyons showed by Yuan et al (2014). Hence, urban planners must consider urban variables, air dispersity levels and inner-city characteristics when designing densely populated cities such as Rotterdam or Amsterdam. Policy makers should consider air dispersity: when giving contracts for (high-rise) buildings in dense cities, when industry is located next to highly populated cities and when private car ownership becomes more favourable compared to public transport. Based on the overall findings of this research, advice would be to run air dispersity simulation tests to see how buildings can affect air pollution in the Netherlands. Next, large industrial plants near cities create an unhealthy living environment and urban planners should consider this when expanding a city or approving new industry. Policy makers should avoid significant positive clusters of pollution as this could lower the health of nearby citizens. Urban planners should also convince municipalities to invest more in public transportation as this makes private car ownership more expansive relative to public transport. If policy makers take these things into account, it should improve air pollution in high density cities and work toward creating sustainable living environments for future generations in the Netherlands.

7. Limitations and Recommendations for future research

This section will discuss the limitations of the dataset as well as the used statistical methods and give improvements for future research.

As discussed before, the OLS regressions used to estimate the effect of urban variables on the air pollution levels gave a bias, which was partly resolved by doing Moran's I and hot spot analysis. Fang et al (2015) however described models to estimate urban variables while considering spatial autocorrelation. One way of doing this is by geographically weighted regressions, which is possible in ArcGIS. This regression takes into account the independent values of each neighbor, depending on the distance from each municipality. This is again done by a weight matrix, similar to the one described in the formula for Moran's I. It is however possible to adjust the bandwidth, this can be chosen by minimizing the corrected Akaike Information Criterion (AIC). The geographically weighted regression is used to solve spatial non-stationarity. In addition, Fang et al (2015) use a spatial lag model (SAR) and a spatial error model (SEM) to strengthen their results. The SAR model is an extension of OLS, which corrects spatial dependence problems. The SAR model is appropriate when spatial dependence is suspected in the values of the dependent variable. This has been proven to be the case in our research by the hot spot analysis. The SEM is particularly suitable when spatial dependence is suspected in the error terms, which is the case in our OLS regression.

The maps made for the hot spot analysis show a range of results but fail to show which municipalities pose a public health risk. Although this would be preferable, all indexes that measure health risks are based on 1-hour values of the pollutants (micrograms per cubic metre). The municipality level data is based on a total pollution in kilograms for all industries of a specific year, converted to square kilometer. The municipality level data cannot be converted to fit a certain index. In addition, as shown by the Environment Protection Agency (EPA, 2010) the local inner-city health risks can be very different from an overall health risk at municipality level data since busy streets and other urban characteristics are also a determining factor.

Another limitation was the lack of urban variables in the OLS regressions. Although, the CBS gave some helpful variables, compared to Fang et al (2015) there are some relevant urban variables missing such as urbanization rate and the proportion of secondary industry. Rodríguez et al (2016) also showed that urban fragmentation plays an important role, which has not been a major topic in this paper. Future research could be extended by requesting CBS access to these variables, if available.

The main issue with the data for hypothesis 3 is that the measuring stations are not located in the lowest air dispersity spots in the city. This creates a problem as a location between higher buildings might yield higher pollution values. It is therefore very difficult to make conclusion on the entire city. The recommendation would be to have at least 2 stations near high rise buildings in the city center. Also, the methodology for choosing the high-density area by Google Maps is not ideal. Better would be to physically visit the locations and consider the neighborhood to get a better picture. In addition, wind speeds could be measured at these stations as well to define air dispersity more accurately. In this paper, an overall wind speed for Rotterdam and Amsterdam was used, which could be very different compared to the city center. Compared to empirical literature, it is an improvement that it is possible to test this in practice rather than the simulation by Yuan et al (2014). The local authorities of Rotterdam and Amsterdam could also fund some research simulations to improve the urban planning further on. Finally, it is questionable whether the regression results have much external validity as it is very dependent on a single city or even the location within a city. Generalization of the results is hard as the Netherlands is unique for being flat, which enhances air dispersity. In addition, air pollution of factories outside of the Netherlands could also affect the air pollution due to air dispersity. The compactness aspect discussed by Rodríguez et al (2016) plays an important role in generalizing the results between cities, as they denote key differences between the compactness of cities.

8. Bibliography

Alimonti, G. (2018). Our energy future starts from actual energy limits. In EPJ Web of Conferences (Vol. 189, p. 00003). EDP Sciences.

Alpkokin, P., Kuriyama, K., & Hayashi, Y. (2004). Analyzing urban planning in the Netherlands for dissemination of know-how and experiences. In CODATU XI: world congress: Towards more attractive urban transportation.

Blum, B., & Neumärker, B. (2020). Globalization, Environmental Damage and the Corona Pandemic-Lessons from the Crisis for Economic, Environmental and Social Policy.

Burchell, R. W., Shad, N. A., Listokin, D., Phillips, H., Downs, A., Seskin, S., ... & Gall, M. (1998). The costs of sprawl-revisited (No. Project H-10 FY'95).

Burchell, R., Downs, A., McCann, B., & Mukherji, S. (2005). Sprawl costs: Economic impacts of unchecked development. Island Press.

Burton, E. (2000). The compact city: just or just compact? A preliminary analysis. Urban studies, 37(11), 1969-2006.

Burton, E., Jenks, M., & Williams, K. (Eds.). (2003). The Compact City: A Sustainable Urban Form?. Routledge.

Breheny, M. (1995). The compact city and transport energy consumption. Transactions of the institute of British Geographers, 81-101.

Carmona, M., Heath, T., Tiesdell, S., & Oc, T. (2010). Public places, urban spaces: the dimensions of urban design. Routledge.

Cho, H. S., & Choi, M. J. (2014). Effects of compact urban development on air pollution: Empirical evidence from Korea. Sustainability, 6(9), 5968-5982.

Coase, R. H. (1960). The problem of social cost. In Classic papers in natural resource economics (pp. 87-137). Palgrave Macmillan, London.

Dieleman, F. M., Dijst, M. J., & Spit, T. (1999). Planning the compact city: the Randstad Holland experience. European Planning Studies, 7(5), 605-621.

Dieleman, F., & Wegener, M. (2004). Compact city and urban sprawl. Built Environment, 30(4), 308-323.

Emissieregistratie. (ND). Rijksoverheid. Retrieved on 29-07-2020 from: http://www.emissieregistratie.nl/erpubliek/erpub/weergave/kaart.aspx

EPA. (2010). Air Quality Guide for Nitrogen Dioxide. US Government Environment Protection Agency. Retrieved on 17-08-2020 from: https://www.airnow.gov/sites/default/files/2018-06/no2.pdf

Fang, C., Liu, H., Li, G., Sun, D., & Miao, Z. (2015). Estimating the impact of urbanization on air quality in China using spatial regression models. Sustainability, 7(11), 15570-15592.

Galster, G., Hanson, R., Ratcliffe, M. R., Wolman, H., Coleman, S., & Freihage, J. (2001). Wrestling sprawl to the ground: defining and measuring an elusive concept. Housing policy debate, 12(4), 681-717.

Gibson, M., & Carnovale, M. (2015). The effects of road pricing on driver behavior and air pollution. Journal of Urban Economics, 89, 62-73.

De Graaf, K. J., & Jans, J. H. (2015). The Urgenda decision: Netherlands liable for role in causing dangerous global climate change. Journal of Environmental Law, 27(3), 517-527.

Hang, J., Sandberg, M., Li, Y., & Claesson, L. (2009). Pollutant dispersion in idealized city models with different urban morphologies. Atmospheric Environment, 43(38), 6011-6025.

Hang, J., Li, Y., Sandberg, M., Buccolieri, R., & Di Sabatino, S. (2012). The influence of building height variability on pollutant dispersion and pedestrian ventilation in idealized high-rise urban areas. Building and Environment, 56, 346-360.

Jenks, M., & Jones, C. (Eds.). (2009). Dimensions of the sustainable city (Vol. 2). Springer Science & Business Media.

Johnson, M. P. (2001). Environmental impacts of urban sprawl: a survey of the literature and proposed research agenda. Environment and planning A, 33(4), 717-735.

KNMI. (ND). Koninklijk Nederlands Meteorologisch Instituut. Retrieved on 29-07-2020 from: http://projects.knmi.nl/klimatologie/uurgegevens/selectie.cgi

Lashof, D. A., & Ahuja, D. R. (1990). Relative contributions of greenhouse gas emissions to global

Luchtmeetnet. (ND). Retrieved on 29-07-2020 from: https://www.luchtmeetnet.nl/meetpunten

Maas, R., & Fischer, P. (2015). J Wesseling en F Cassee, 2015, Gezondheidswinst door betere luchtkwaliteit–Is schonere lucht in Nederland mogelijk. Tijdschrift Lucht, september.warming. Nature, 344(6266), 529-531.

Markus, N. (2020). Nu het land deels stilligt, daalt de stikstofuitstoot. Trouw. Retrieved on 24-06-2020 from: https://www.trouw.nl/duurzaamheid-natuur/nu-het-land-deels-stilligt-daalt-de-stikstofuitstoot~b79883f8/

Moudon, A. V. (1997). Urban morphology as an emerging interdisciplinary field. Urban morphology, 1(1), 3-10.

Neuman, M. (2005). The compact city fallacy. Journal of planning education and research, 25(1), 11-26.

Priemus, H. (1994). Planning the Randstad: between economic growth and sustainability. Urban Studies, 31(3), 509-534.

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. Geographical analysis, 27(4), 286-306.

Rijksoverheid (2019). Maatregelen om het stikstofprobleem op te lossen. Retrieved on 24-06-2020 from: https://www.rijksoverheid.nl/onderwerpen/aanpak-stikstof/maatregelen-om-stikstofprobleem-op-te-lossen

Rijksoverheid (ND). Klimaatbeleid. Retrieved on 24-06-2020 from: https://www.rijksoverheid.nl/onderwerpen/klimaatverandering/klimaatbeleid

RIVM (ND). Stikstof. Retrieved on 10-08-2020 from: https://www.rivm.nl/stikstof

Rodríguez, M. C., Dupont-Courtade, L., & Oueslati, W. (2016). Air pollution and urban structure linkages: Evidence from European cities. Renewable and Sustainable Energy Reviews, 53, 1-9.

Rose, A., & Stevens, B. (1993). The efficiency and equity of marketable permits for CO2 emissions. Resource and Energy economics, 15(1), 117-146.

Song, Y., & Knaap, G. J. (2004). Measuring urban form: Is Portland winning the war on sprawl?. Journal of the American Planning Association, 70(2), 210-225.

TNO (2019). Nederland relatief grootste producent stikstof van Europa. Retrieved on 24-06-2020 from: https://nos.nl/artikel/2306387-tno-nederland-relatief-grootste-producent-stikstof-van-europa.html

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography, 46(sup1), 234-240.

Van der Valk, A. (2002). The Dutch planning experience. Landscape and urban planning, 58(2-4), 201-210.

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G. & Van Weele, M. (2012). TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. Remote Sensing of Environment, 120, 70-83.

Wang, C., Li, Q., & Wang, Z. H. (2018). Quantifying the impact of urban trees on passive pollutant dispersion using a coupled large-eddy simulation–Lagrangian stochastic model. Building and Environment, 145, 33-49.

Yuan, C., Ng, E., & Norford, L. K. (2014). Improving air quality in high-density cities by understanding the relationship between air pollutant dispersion and urban morphologies. Building and Environment, 71, 245-258.

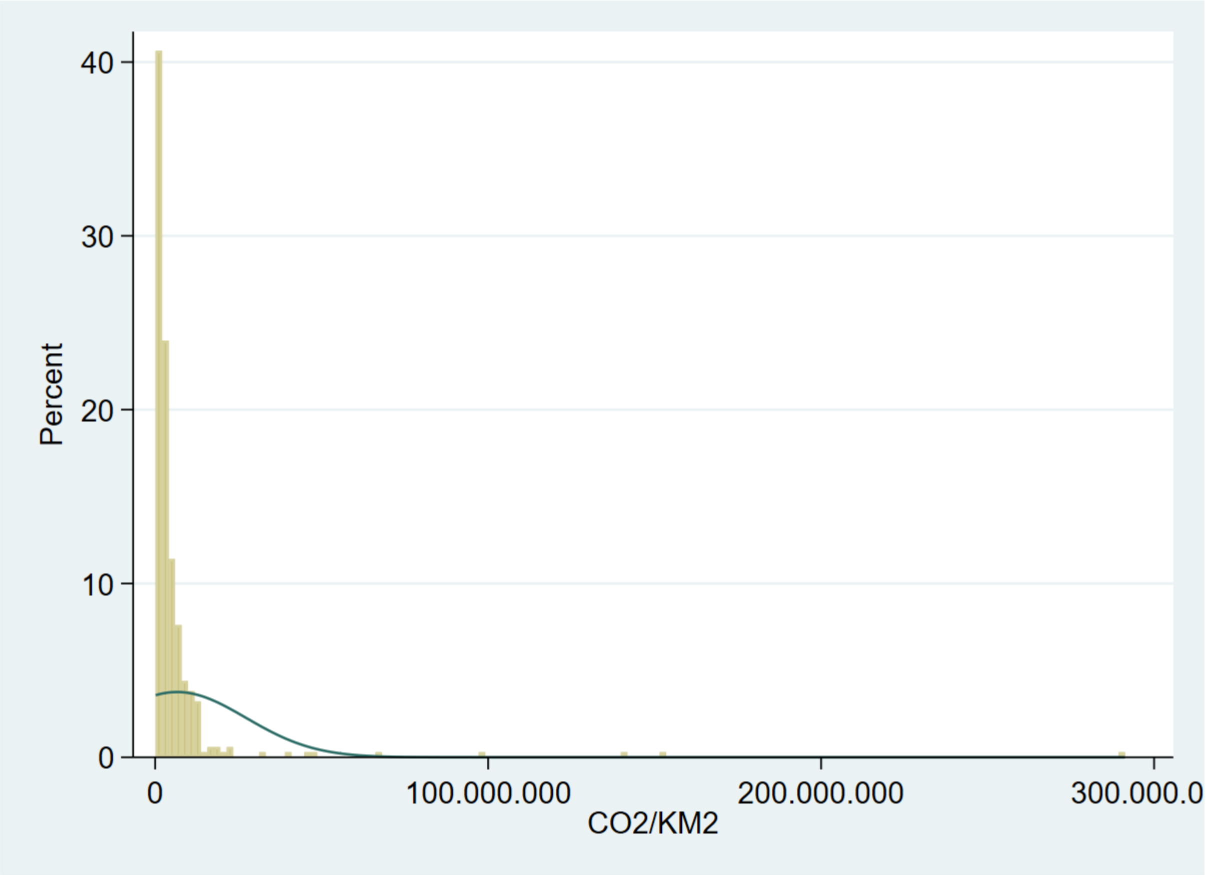
9. Appendix

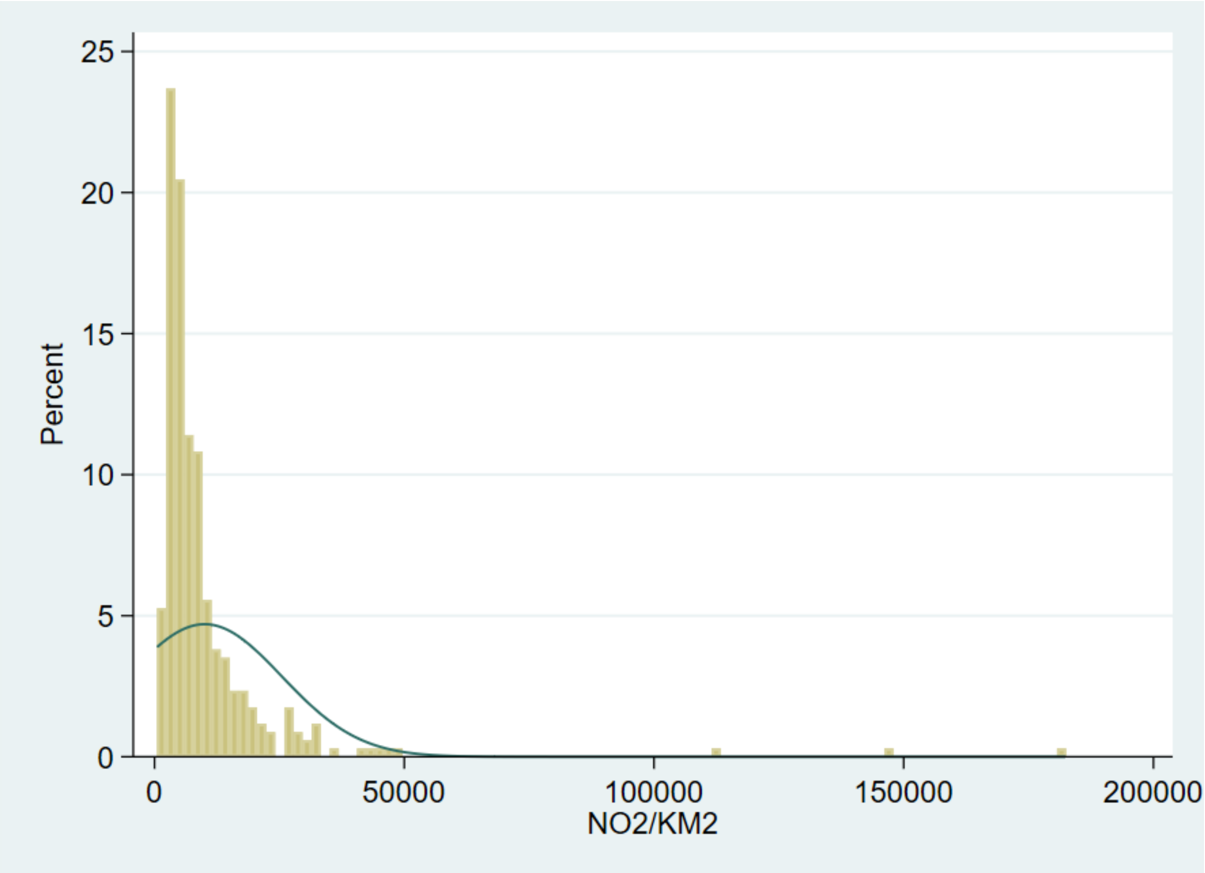
Table A: Spearman correlation coefficients for all independent and dependent variables.

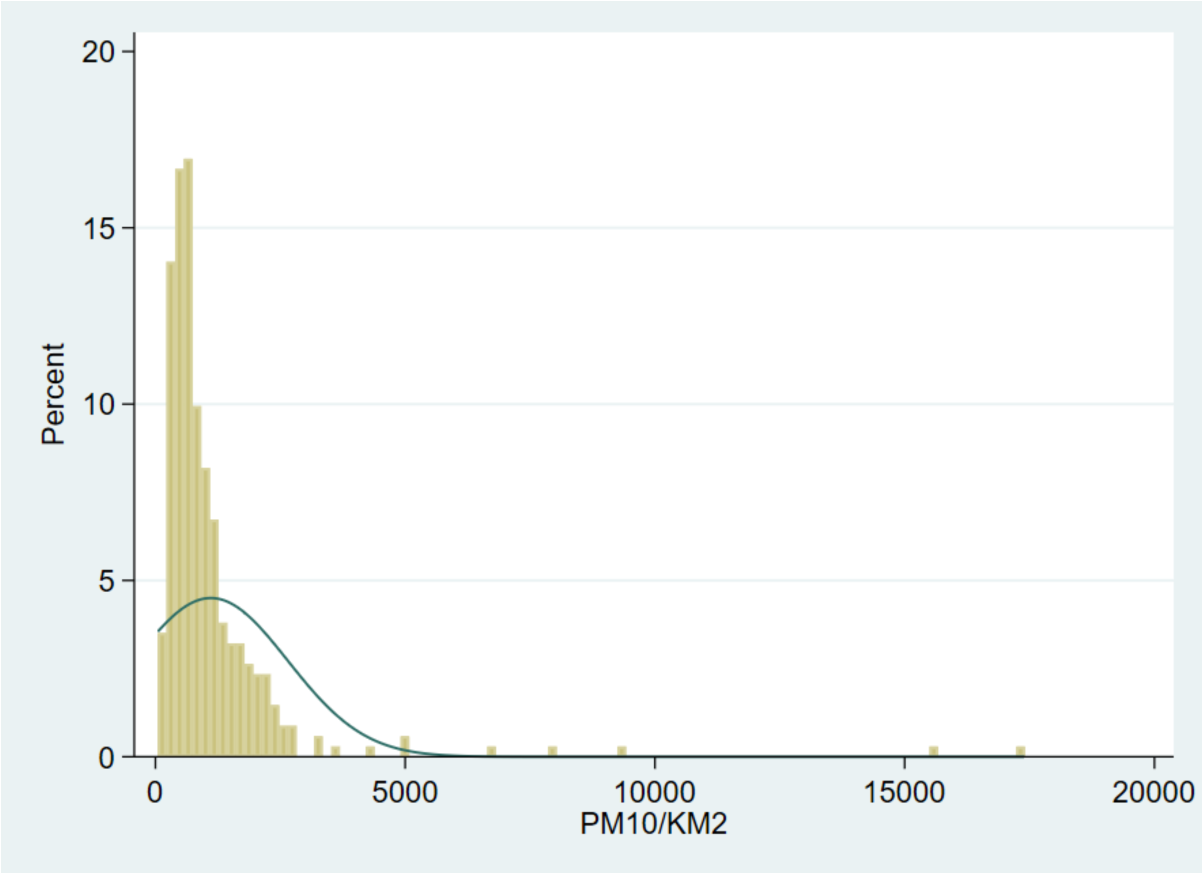
	CO2KM2	NO2KM2	PM10KM2	WPS	AI	AHV	LUT	LUB	LUSM	LUR	LUA	LUF	PDC	DA	PCO	PCM	EC
CO2KM2	1																
NO2KM2	0.9174*	1															
Sig. 2-tailed	0.0000																
PM10KM2	0.8108*	0.7735*	1														
Sig. 2-tailed	0.0000	0.0000															
WPS	0.5811*	0.5036*	0.4622*	1													
Sig. 2-tailed	0.0000	0.0000	0.0000														
AI	0.1615*	0.1489*	-0.0072	0.0654	1												
Sig. 2-tailed	0.0032	0.0066	0.8958	0.2352													
AHV	-0.1959*	-0.1943*	-0.1948*	-0.3350*	0.7124*	1											
Sig. 2-tailed	0.0003	0.0004	0.0004	0.0000	0.0000												
LUT	-0.1435*	-0.1032	-0.1392	0.2970*	-0.3465*	-0.3054*	1										
Sig. 2-tailed	0.0089	0.0607	0.0112	0.0000	0.0000	0.0000											
LUB	0.3373*	0.2808*	0.2850*	0.7661*	-0.2178*	-0.4042*	0.6775*	1									
Sig. 2-tailed	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000										
LUSM	0.2207*	0.2406*	0.1550*	0.5302*	-0.3581*	-0.5047*	0.7385*	0.7619*	1								
Sig. 2-tailed	0.0001	0.0000	0.0047	0.0000	0.0000	0.0000	0.0000	0.0000									
LUR	0.0506	0.0044	-0.0005	0.5731*	-0.1443*	-0.2402*	0.6907*	0.7597*	0.6636*	1							
Sig. 2-tailed	0.3590	0.9363	0.9921	0.0000	0.0086	0.0000	0.0000	0.0000	0.0000								
LUA	-0.5839*	-0.5189*	-0.5060*	-0.1698*	-0.3562*	-0.1094	0.7684*	0.2449*	0.4071*	0.3524*	1						
Sig. 2-tailed	0.0000	0.0000	0.0000	0.0019	0.0000	0.0467	0.0000	0.0000	0.0000	0.0000							
LUF	-0.5229*	-0.5757*	-0.4232*	-0.0416	-0.1633*	0.1120	0.4772*	0.2361*	0.1703*	0.4288*	0.5212*	1					
Sig. 2-tailed	0.0000	0.0000	0.0000	0.4507	0.0029	0.0416	0.0000	0.0000	0.0019	0.0000	0.0000						
PDC	-0.7529*	-0.6591*	-0.6614*	-0.7028*	-0.1442*	0.2134*	0.1366	-0.4097*	-0.1774*	-0.2119*	0.5617*	0.3428*	1				
Sig. 2-tailed	0.0000	0.0000	0.0000	0.0000	0.0086	0.0001	0.0129	0.0000	0.0012	0.0001	0.0000	0.0000					
DA	-0.7691*	-0.6877*	-0.6553*	-0.5421*	-0.2329*	0.1143	0.2833*	-0.2287*	-0.0275	0.0006	0.6867*	0.4215*	0.7855*	1			
Sig. 2-tailed	0.0000	0.0000	0.0000	0.0000	0.0000	0.0377	0.0000	0.0000	0.6180	0.9915	0.0000	0.0000	0.0000				
PCO	0.4111*	0.3685*	0.3482*	0.8417*	-0.1130	-0.3555*	0.5625*	0.9320*	0.6781*	0.7298*	0.1160	0.1108	-0.5346*	-0.3341*	1		
Sig. 2-tailed	0.0000	0.0000	0.0000	0.0000	0.0399	0.0000	0.0000	0.0000	0.0000	0.0000	0.0349	0.0441	0.0000	0.0000			
PCM	0.8505*	0.7776*	0.7424*	0.6075*	0.2453*	-0.1380	-0.3443*	0.2596*	0.0503	0.0049	-0.7429*	-0.5634*	-0.8451*	-0.8486*	0.4141*	1	
Sig. 2-tailed	0.0000	0.0000	0.0000	0.0000	0.0000	0.0119	0.0000	0.0000	0.3618	0.9287	0.0000	0.0000	0.0000	0.0000	0.0000		
EC	-0.3543*	-0.3603*	-0.2184*	-0.4236*	0.2508*	0.5703*	-0.1345	-0.3583*	-0.3448*	-0.3118*	0.1811*	0.2140*	0.4622*	0.3837*	-0.3785*	-0.3442*	1
Sig. 2-tailed	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0143	0.0000	0.0000	0.0000	0.0009	0.0001	0.0000	0.0000	0.0000	0.0000	

*Source: CBS and Emission Registration. Note: * equals significance at a 1% level and the significance level is 2-tailed.*

Figure A, B and C: Histograms of CO_2/KM^2, NO_2/KM^2 and PM_{10}/KM^2 distribution.

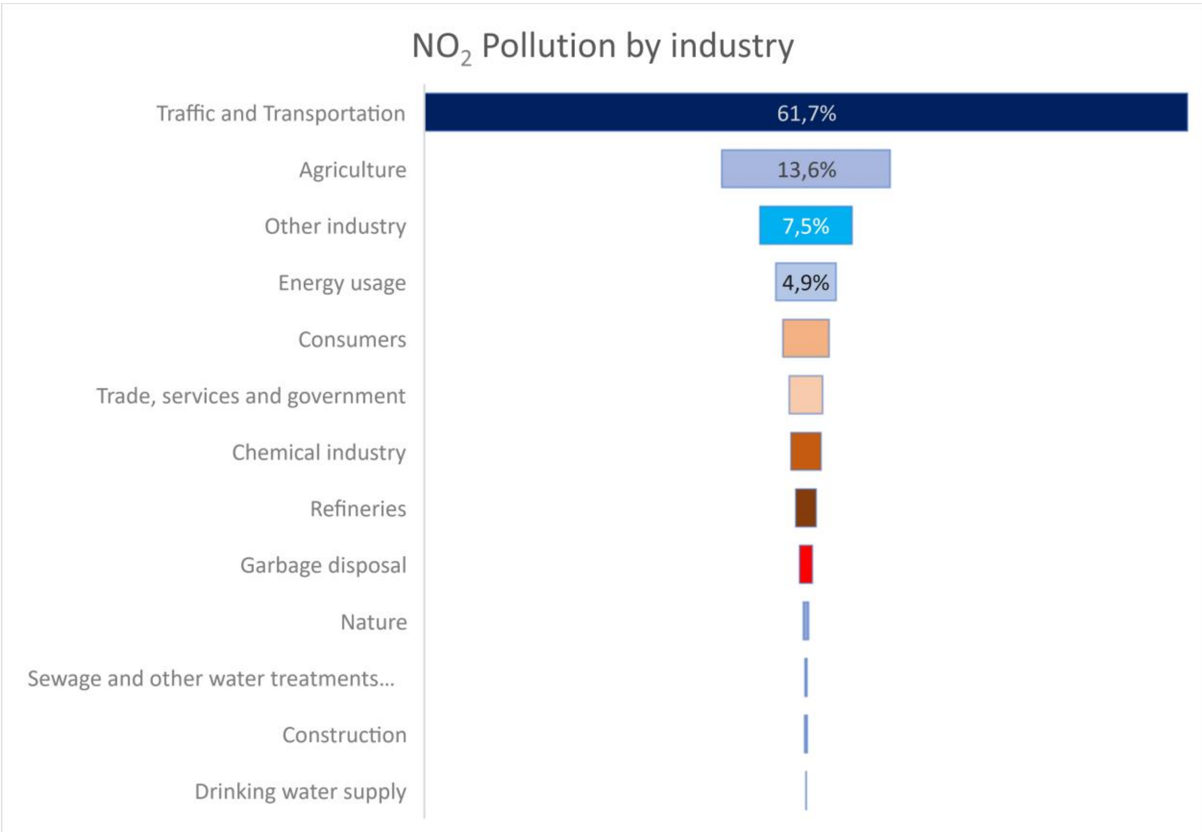


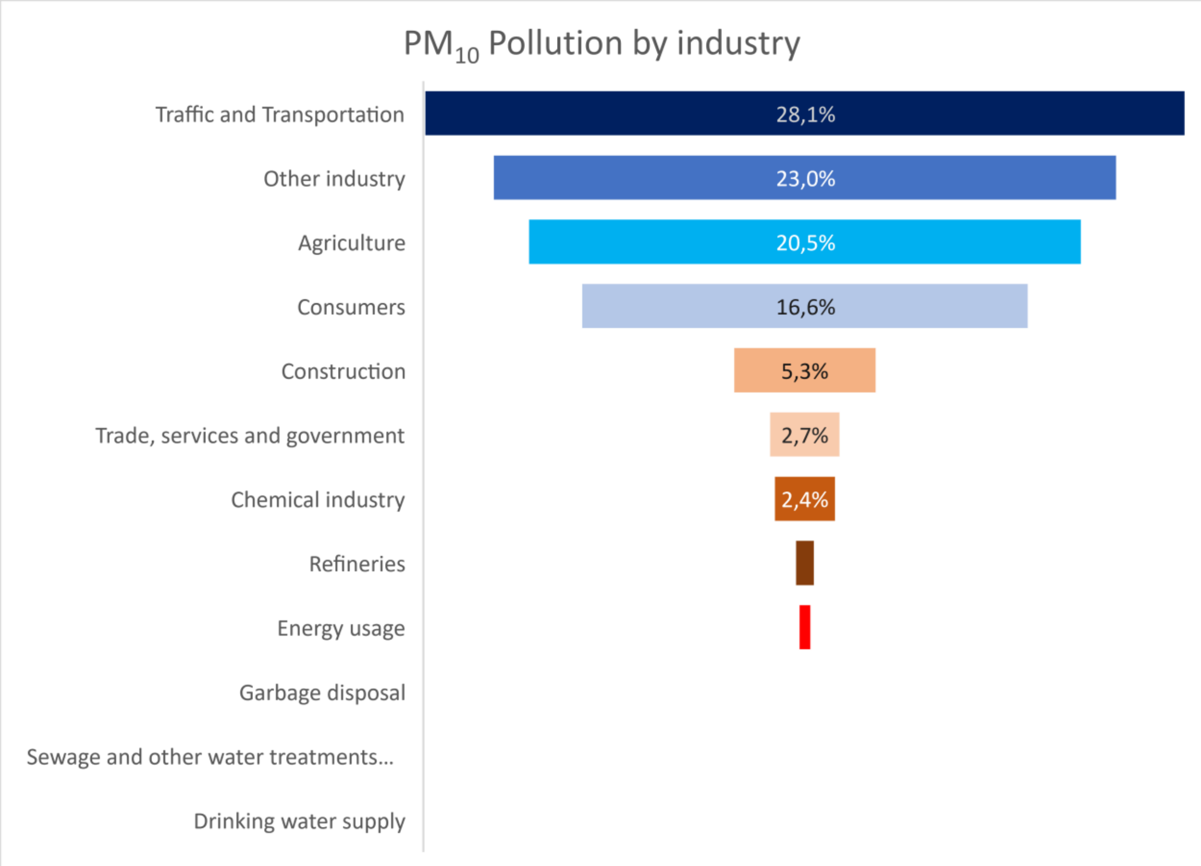




Source: Emissieregistratie.

Figures D and E: Funnel chart of NO_2 and PM_{10} pollution per industry.





Source: Emissieregistratie.

Table B: Regression of all independent variables on traffic and transportation pollution values per square kilometer with fixed effects.

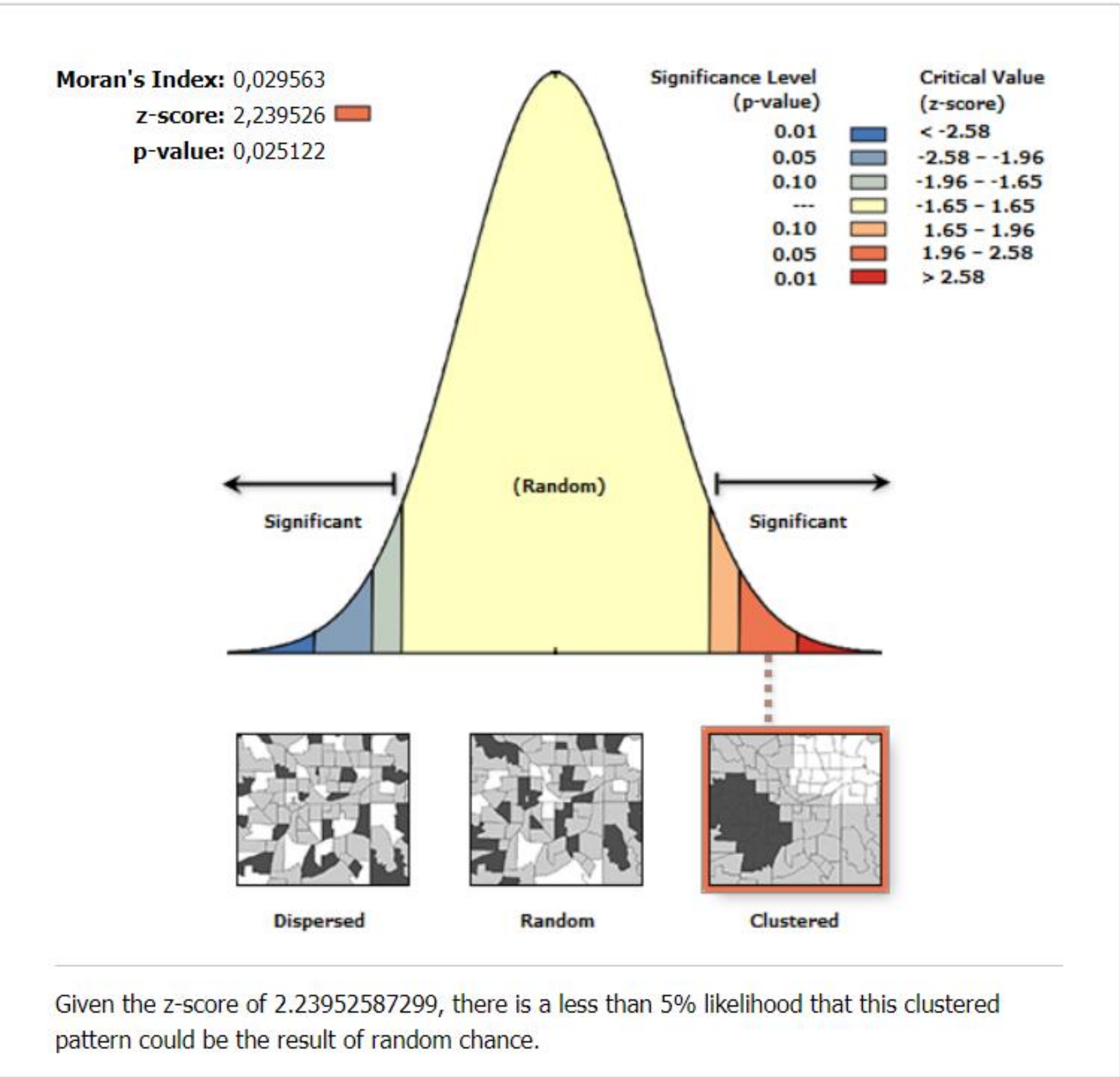
VARIABLES	(20) CO_2/KM^2 Traffic	(21) NO_2/KM^2 Traffic	(22) PM_{10}/KM^2 Traffic
LU:T	5,413***	5.774	0.409
	(678.4)	(10.20)	(0.389)
LU:B	-238.7	0.424	-0.0245
	(171.8)	(2.660)	(0.0958)
LU:SM	-401.3	-5.584	-0.196
	(460.8)	(10.07)	(0.362)
LU:R	-317.5	-5.985	-0.244
	(433.6)	(5.321)	(0.179)
LU:A	-156.9***	-0.166	-0.0131
	(17.63)	(0.185)	(0.00795)
LU:F	-95.77***	-0.295	-0.0134
	(21.28)	(0.357)	(0.0132)
PDC = 2	-397,168	-6,523	-191.8
	(296,964)	(7,043)	(225.5)
PDC = 3	-807,320**	-13,446*	-472.8*
	(356,059)	(7,663)	(256.7)
PDC = 4	-850,735**	-14,274*	-529.3*
	(391,616)	(8,528)	(292.8)
PDC = 5	-940,339**	-17,146*	-643.4**
	(430,420)	(9,473)	(324.6)

PD	799.4***	5.231	0.243**
	(132.4)	(3.871)	(0.118)
AI	58,638	-337.5	-2.761
	(41,996)	(516.6)	(18.49)
AHV	-2,762	53.70	1.770
	(3,097)	(45.08)	(1.644)
WPS	-5.722	-0.103	-0.00326
	(8.225)	(0.226)	(0.00779)
PCO	-10.68***	-0.0268	-0.00103
	(3.642)	(0.0564)	(0.00200)
DA	24.09	6,098***	350.3***
	(90,519)	(1,247)	(82.43)
EC	166.3	-3.353	-0.0872
	(360.4)	(4.054)	(0.149)
Observations	331	331	331
R-squared	0.727	0.199	0.286
Province dummies	Yes	Yes	Yes

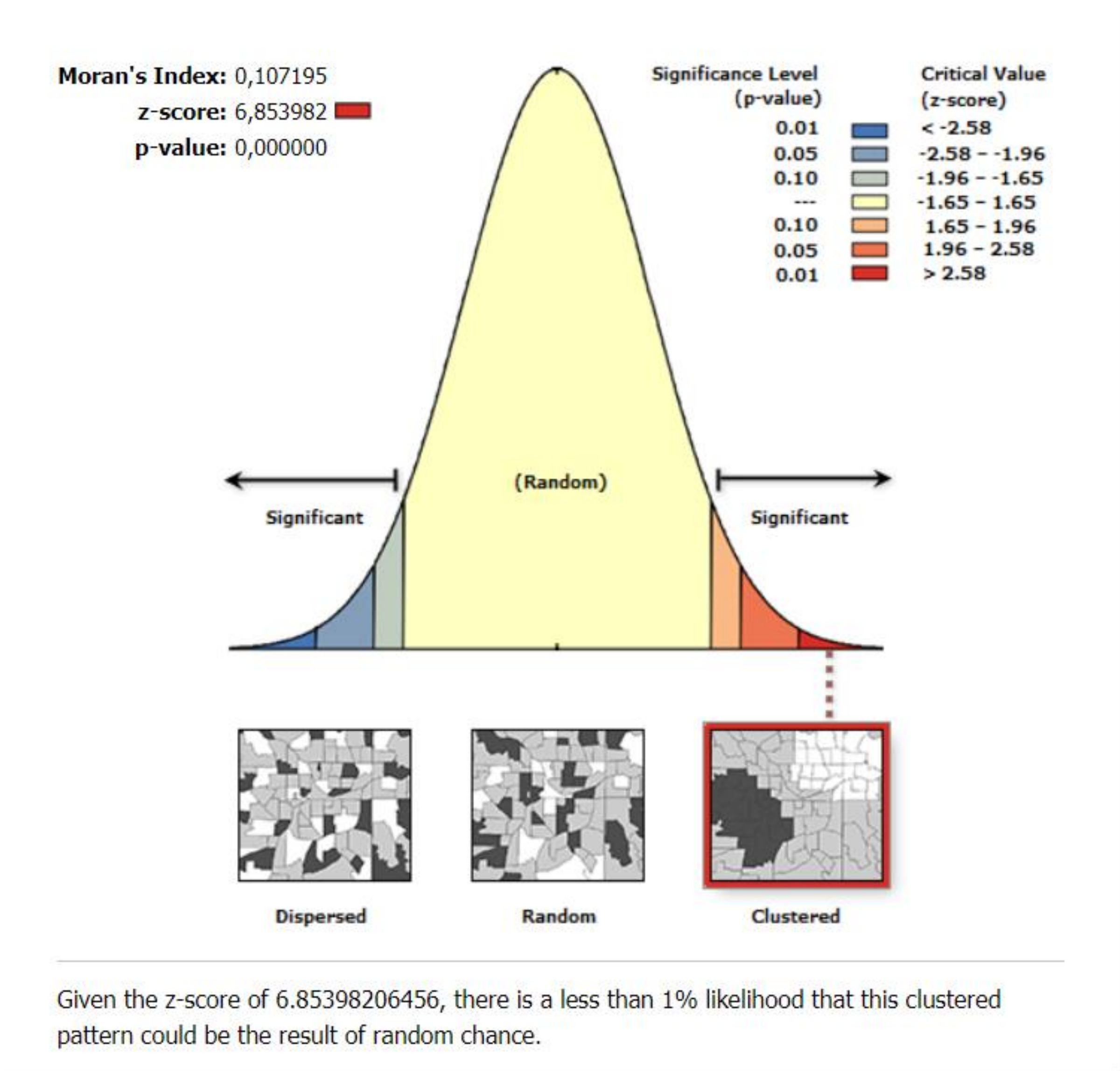
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: CBS & Emissieregistratie.nl Note: LU=Land use (T=Traffic, B= built areas, SM=semi-built areas, R=recreational, A=agriculture and, F=Forest and open land), PDC=Population density categorical. WP=Western Population (migrated), AI=Average income, PCO= Private Car Ownership, AHV=Average House Value, DA=Distance to Amenities, PD= Population Density, EC=Energy Consumption.

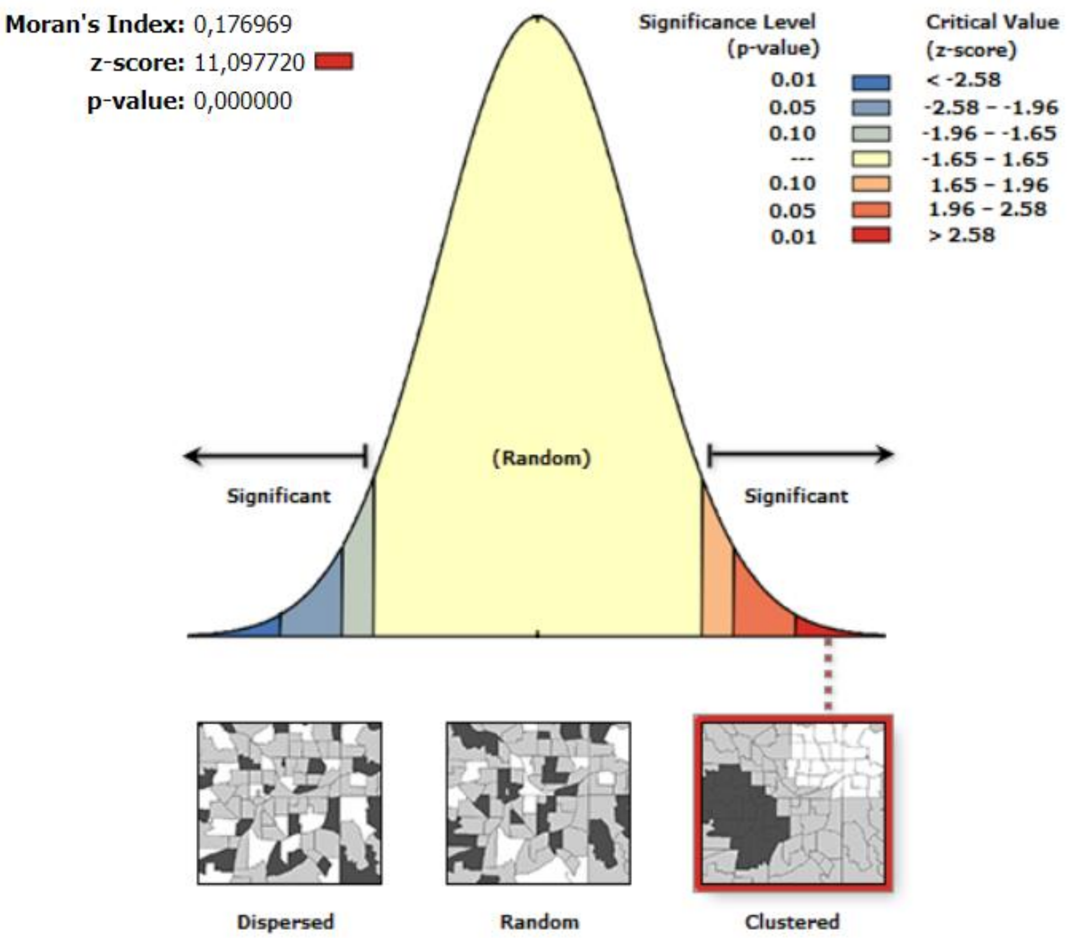
Figure F, G and H: Moran's I spatial autocorrelation report for CO_2/KM^2, NO_2/KM^2 and PM_{10}/KM^2 using ArcGIS.



Moran's I Spatial Autocorrelation report NO2



Moran's I spatial autocorrelation report PM10



Given the z-score of 11.0977199034, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Source: Emissieregistratie.

Table C: Representation of visual characteristics (below) and reasoning behind dummy variable choices.

LOCATION	DUMMY	REASONING
		ROTTERDAM
SCHIEDAMSE VEST	1	Schiedamse Vest is located in the inner-city of Rotterdam. Although it is not surrounded by high-rise buildings, the population density is high as well as nearby traffic.
HOOGVLIET	0	The station in Hoogvliet is located next to a parking place. The surrounding area is mostly open so high air dispersity is expected.
PLEINWEG	1	The measuring station at Pleinweg is located on the South-side of Rotterdam and is next to a busy street. In addition, a four-story building is right next to the both sides of the street.
ZWARTEWAALSTRAAT	0	The Zwartewaalstraat is also on the South-side very close to Pleinweg. However, this station is given a 1 as dummy because of the open area. The houses on the sides will not be able to stop the wind and the flats are too far away.
RIDDERKERK	1	Ridderkerk is located next to a national highway. Although this look like a picture of an open field, on the other side are a lot of trees holding back the high wind speeds. This is also true for the sound barriers next to the highway.
OVERSCHIE	0	Overschie is also located next to a national highway but the difference with Ridderkerk is the distance to the sound barrier. In this case there is a street in between. Although the building will stop some of the wind, the expectation is that this will not be enough at high wind speeds.
STATENWEG	1	Statenweg follows the same reasoning of Pleinweg as to why low air dispersity is expected.
SCHIEDAM	1	The measuring station in Schiedam is located on a parking place next to a very high flat. In addition, it is surrounded by trees which should have an effect on air dispersity.
MAASSLUIS	0*	This measuring station picture is taken in 2009 using Google Street View. On the left side it is clearly visible that construction is going on. Therefore, it is hard to say how much have changed in the urban morphology between 2009 and 2019. It is assumed this area has a high air dispersity, but it is uncertain.
BERGHAVEN (HOEK VAN HOLLAND)	0**	This measuring stations was not found on any of the Google Street View pictures although it could be the square building in the middle. However, since no measuring equipment was seen on this building and this place is an open area, it is assumed that is good air dispersity in this location.
		AMSTERDAM
HAARLEMMERWEG	1**	The description of this measuring station already stated that it is on an unconventional location. The station could not be found on any of the Google Street Map pictures but should be between or on top of the buildings on the right. It is given a 1 because of the high population density in the neighbourhood and the relatively high buildings.
NIEUWENDAMMERDIJK	0	This location is in the suburbs of Amsterdam next to an open area.
EINSTEINWEG	0	The Einsteinweg is next to a national highway but also quite a high flat. In this instance wind direction might play a role, seeing how close the measuring equipment is to the building. It is given the value 0 for the dummy because the cars will be driving next to the station at a

		very close proximity and creating wind speeds as they go. This is way, at high natural wind speeds, the effect is not going to be as large.
VAN DIEMENSTRAAT	1	
VONDELPARK	0**	The measuring station is behind these buildings but could not see it because there is not a road there for Google Street View Cars. It is right next to the Vondelpark, which is an open park with relatively high trees. The expectation will be that this is not enough to stop wind.
STADHOUDERSKADE	1	This location is right next to a busy street, relatively close to the centre Amsterdam. As visible from the picture, there are high buildings in the background and on the right are three-story houses.
OUDESCHANS	1**	Oudeschans in located in the old city-centre of Amsterdam. Although the station is not visible on any of the Google Street View Pictures, it is expected due to the location and the high density of buildings that the location has low air dispersity.
JAN VAN GALENSTRAAT	1	This location has similar characteristics compared to Stadhouderskade and is given a 1 for the same reasons.
KANTERSHOF	0	This location is situated in the Bijlmer suburbs. There are many 4 stories buildings in weird pattern. However, this measuring station is located somewhat on the outside of the neighbourhood and is in the middle of an open area.
SPORTPARK OOKMEER	0	This measuring station is located at the outskirts of Amsterdam. From the picture it is clearly visible that it is next to an open area.

*Source: Google Street View. Note: *The visual characteristics of Maassluis are based on a picture from 2009 and is outdated. **The measuring station was not found on Google Street Maps.*

Schiedamse Vest NL10418 (Dummy=1)

Hoogvliet NL01485 (Dummy=0)

Pleinweg NL01487 (Dummy=1)

Zwartewaalstraat NL01488 (Dummy=0)

Ridderkerk A16 NL01489 (Dummy=1)

Overschie A13 NL01491 (Dummy=0)

Statenweg NL01493 (Dummy=1)

Schiedam Alphons Arienstraat NL01494 (Dummy=1)

Maasssluis NL01495 (Dummy=0)

Berghaven: Hoek van Holland NL01496 (Dummy=0)

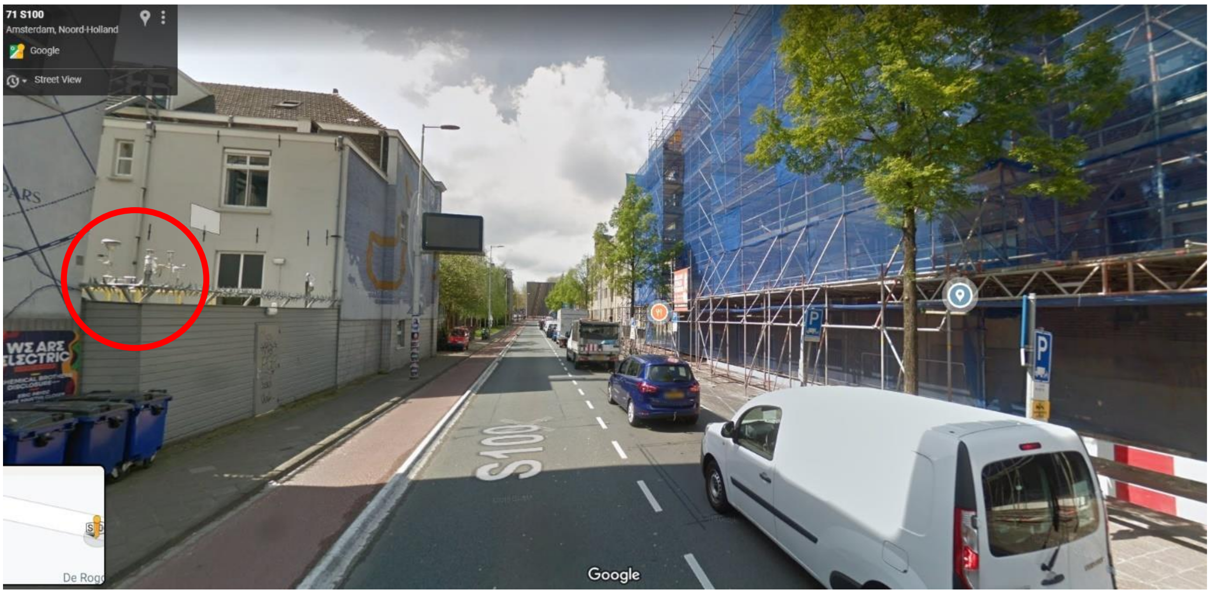
Amsterdam Google Street View Pictures

Haarlemmerweg NL49002 (Dummy=1)

Nieuwendammerdijk NL49003 (Dummy=0)

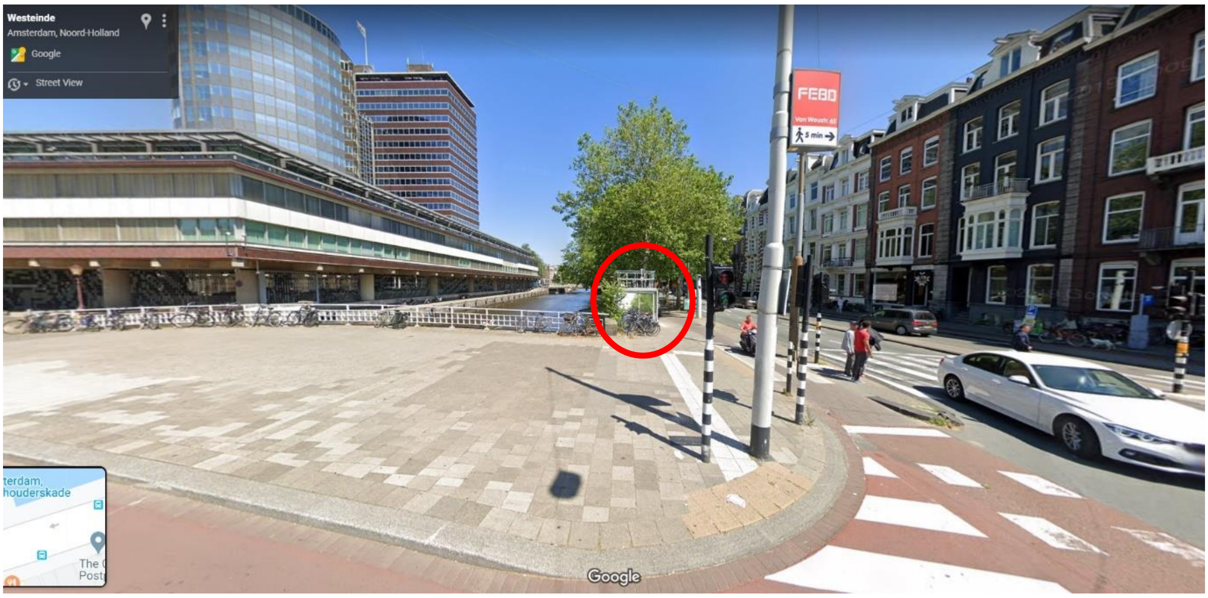
Einsteinweg A10 NL 49007 (Dummy=0)

Van Diemenstraat NL49012 (Dummy=1)



Vondelpark NL49014 (Dummy=0)

Stadhouderskade NL49017 (Dummy=1)



Oudeschans NL49017 (Dummy=1)

Jan van Galenstraat NL49020 (Dummy=1)

Kantershof NL49021 (Dummy=0)

Sportpark Ookmeer NL49022 (Dummy=0)