The design and evaluation of high-resolution personal environmental exposure assessment for global scale studies

MSc Thesis Lars Groeneveld
Student nr: 5519152
37.5 ECTS
August 2019 – April 2020

First supervisor: Prof. dr. D.J. Karssenberg
Second supervisor: dr. M. Lu
Faculty of Geosciences, Utrecht University

Utrecht University

Abstract

The impact of environmental variables like air pollution and greenness on health is studied using personal human exposure assessment. Studies rely on of static exposure assessment, only using the residential location of a population, or apply data-rich techniques, which rely on GPS or portable air pollution measurement devices to measure the exposure for often a small population over a limited time span. Long-term population wide-exposure assessments have often low data availability because of a lack of human mobility data. In this study, a method is established for a high-resolution personal exposure model for NO_2 and greenness in a sparse space-time activity data situation. Downscaling methods and exposure assessment techniques are developed and evaluated to study environmental exposures in Utrecht and Madrid. To establish a high-resolution personal exposure assessment, population datasets are downscaled using OpenStreetMap building data from Utrecht and Madrid. Because there is no global human mobility data set available, human mobility is simulated for three social-economic groups, commuters, homemakers and students, with different mobility patterns. Activity is represented in zones of potential activity and using weighted buffers to represent how likely a place is visited. The exposure assessment shows that commuters have the highest NO_2 exposure and lowest NDVI exposure and commuters and students have a lower range of exposure compared to homemakers. When population data is downscaled there is an negligible difference in residential exposure between the two population datasets. Together with the considerable difference in the mean error between the different downscaling methods, this indicates that the location of the residences is more important than the resolution of the population data. Sensitivity analysis shows that there is only a minor effect (max 2%) in changing the weights of the buffers while changing the size of the buffers influences the exposure up to 18%. This study indicates that downscaling of environmental information is of major importance for exposure assessments while the weights used for different buffer sizes has relatively limited effect on calculated exposure values.

Table of Contents

Digital supplement

1. Introduction

Personal human exposure assessments are used in epidemiological studies to identify the impact of environmental exposures on physical and mental health. Long-term exposure to air pollution negatively affects public health. Air pollution leads to an increase in mortality and cardiovascular and respiratory diseases (Hoek et al., 2013). Approximately 3 million people die due to ambient air pollution annually (World Health Organization, 2016). Green space accessibility is associated with physical (Richardson et al., 2011; James et al., 2015) and mental health benefits (Alcock et al., 2014), because green space promotes physical activity and decreases stress. Moreover, vegetation has a reducing effect on air pollution, noise levels and heat exposure (e.g. James et al., 2015; Nowak et al., 2006).

Personal exposure assessments can be measured using different techniques. Data-rich techniques use continuous measurements using portable air pollution measurement devices installed at or close to the individual (Dons et al., 2011). Alternatively, GPS tracking devices are used to track a person's activity and exposures are read from air pollution maps (Dias & Tchepel, 2014) or satellite data (Almanza et al., 2012) at the route followed by individuals. These methods, however, can only be applied to a relatively small population of individuals and often for a period of only a few weeks (Dons et al., 2011). Long-term population-wide exposure assessment studies face the problem of low data availability because personal space-time data is not available. These assessments often use a low spatial resolution, e.g. Liu et al. (2017) uses a 1 x 1 kilometre spatial resolution.

However, when using a low spatial resolution the spatial patterns of air pollution and greenness concerning different parts of a city or the distance from a road are not taken into consideration. The spatial variability of pollutants plays an important role in personal exposure models because intra-urban air pollution is characterized by rapid decay from emitting sources (Hewitt, 1991). To capture the high spatial variability, it is important to use high-resolution Land Use Regression (LUR) models. LUR models are based on monitored levels of pollutants and use variables like traffic and land use to predict pollution levels at each location using regression models (Ryan & Lemasters, 2007). Schmitz et al. (2019) published an air pollution concentration map with a resolution of 5 x 5 m for the Netherlands. Despite the increase in the resolution of the LUR, population space-time data is still at a low resolution. Previous studies used home locations for calculating personal exposure to air population, however with this method human activity patterns are not represented which could lead to over- or underestimation of air pollution exposure (Schmitz et al., 2019). Therefore, human activity patterns preferably need to be included in exposure studies to better represent the exposure of individuals. Exposure models using human activity use residential and working locations to represent space-time data (Park & Kwan, 2017). Lu et al. (2019) developed an agent-based model to represent space-time activity patterns for different social-economic groups. However, an agent-based method uses too much computational power and needs too detailed input data and is therefore not suitable for the approach used in this study. Almost all exposure assessments study intra-urban exposure assessments, so the exposure inside the city itself. However, in these assessments the agglomeration of the city is not studied. Therefore these exposure assessments neglect commuters who live in towns/cities surrounding the major city. A problem of these methods is that high spatial resolution air pollution models together with information about residential and working locations are needed. However, there is often limited geographically referenced data available at the individual level, and often the resolution of the data is coarse. In almost all studies the living locations of the studied persons is already known or the study is on a local scale in which an addresses database is used for residential locations. For a global exposure assessment study this information is however not available and therefore a global population dataset is needed. Tatem et al. (2011) used several kilometre scale gridded

population datasets for a population at risk of disease assessment and found variations in the number of population at risk of more than 10% between the different population datasets. In recent years higher resolution population datasets are developed with 3 arcsecond and 250 metre resolution but these datasets are not used in exposure assessment studies and the effect on the exposure between different resolution global datasets is not known when used for downscaling.

Most exposure studies focus on a single study area and the data is optimized for their study area. For a global exposure assessment it is not practical to adjust all parameters for each city and use city specific datasets, therefore globally available datasets have to be used, the effect of different resolutions and the quality of the data between different cities is however not yet studied.

Most greenness exposure assessments quantify the amount of green space using vegetation indices, such as the Normalized Difference Vegetation Index (NDVI) (e.g. Crouse et al., 2017; Gascon et al., 2016; Herrera et al., 2018; Kollányi & Prohászka, 2019). In almost all exposure assessments only surrounding or residential greenness is measured; this is the amount of vegetation in the surrounding of a person's residence (Fong et al., 2018). This method is questioned because people work or recreate away from home (James et al., 2015). Several studies showed that for mobility-dependent exposure assessments, mainly based on pollution assessments, only using a static location leads to the so-called neighbourhood effect averaging problem. Therefore, a static greenness exposure assessment will over- or underestimate personal exposure for residences in respectively higher and lower than average areas (Kwan, 2018).

In this MSc thesis personal exposure to air pollution and greenness will be studied for two cities, Utrecht in the Netherlands and Madrid in Spain. The main objectives are to establish and evaluate a method for personal exposure assessment of air pollution and greenness based on high-resolution air pollution and greenness data but sparse space-time activity data. Secondly, the differences in distribution of personal exposure between socio-economic groups is studied. The main research question is: How can high-resolution personal exposure be simulated in sparse space-time data situations and what is the difference in exposure between social-economic groups? The following sub questions will be answered; (1) how can population data be downscaled and what is the effect of the use of different population datasets on exposure assessment, (2) what is the impact of different resolutions of population data on the static personal exposure, (3) how can human activity be represented in exposure assessment for different social-economic groups and what is the difference in exposure between these social-economic groups, (4) what is the difference between static exposure and personal exposure values, (5) what is the difference in air pollution and greenness exposure between intra-urban and city agglomerate areas for Utrecht and Madrid? Exposure to air pollution will be studied for NO_2 and greenness maps will be calculated using a NDVI analyses using Google Earth Engine. The population data will be downscaled using a building dataset and for the Netherlands the population datasets will be compared with a Dutch governmental statistical population dataset of the CBS. Space-time activity data will be simulated for three social-economic groups; homemakers, commuters and students.

2. Literature Review

2.1 Air Pollution

2.1.1 Physical and mental health effect of air pollution

There are several different traffic-related pollutants that can negatively affects individuals health, e.g. NO_x, CO, SO_2, volatile organic compounds (VOCs) and particulate matter (PM) (Bernstein et al., 2004; Han & Naeher, 2006). In exposure assessments particulate matter is defined as particles which have a negative impact on human health. PM is a pollutant, studied in many exposure assessments and consists of a mixture of solid and/or liquid particles suspended in the air consisting of particles with different sizes (nm - µm) and different chemical compositions (Brunekreef & Holgate, 2002; Kampa & Castanas, 2008). Several studies indicated an increase in overall mortality with exposure to PM (e.g. Brunekreef & Holgate, 2002; Mannucci et al., 2015).

Nitrogen dioxide (NO_2) is one of the main traffic-related air pollutants. Short-term exposure to high concentration can cause airway responsiveness and a decline of the lung function (Han & Naeher, 2006). WHO (2006) guidelines indicate that NO_2 concentration higher than 200 µg/m^3 has significant health effects. Long-term exposure to lower concentrations will also lead to health problems such as reduced immunity and respiratory infections (Han & Naeher, 2006). Giving symptoms as nose and throat irritation or bronchoconstriction (Kampa & Castanas, 2008). The annual mean guideline proposed in WHO (2006) indicates a maximum NO_2 concentration of 40 µg/m^3, however, they also indicate that lower concentrations can already have a health effect on children. An UK cohort study found that long term exposure to NO_2 leads to an increase in the chance of heart failure (Atkinson et al., 2013). NO_2 is highly correlated with other pollutants, particularly fine particulate matter. Brunekreef & Holgate (2002) made an assessment that therefore NO_2 can serve as a surrogate for all traffic-related pollutants.

2.1.2 Land use regression models

Land use regression (LUR) models explain the spatial variability using measured pollutant concentrations at sampling points as dependent variable and predictor variables as independent variable using multiple regression equations (Ryan & Lemasters, 2007). In most studies the predictor variables are based on one or multiple of the following attributes: road infrastructure, traffic density, population density, land cover, elevation and climate (Hoek et al., 2008; Ryan & Lemasters, 2007; Schmitz et al., 2019). Several LUR models have been developed for various areas in the world. In Europe, the European Study of Cohorts for Air Pollution Effects (ESCAPE) developed several LUR models for European countries which explained 86% of the variability of NO_2 concentrations (Beelen et al., 2013; Eeftens et al., 2012). LUR pollutant concentration levels calculated with a yearly mean concentration can be considered as long term average values, because the spatial contrasts are relatively stable over the years (Schmitz et al., 2019). Schmitz et al. (2019) used the ESCAPE LUR models together with higher resolution predictor variables to create high resolution air pollution concentration maps with a resolution of 5 metre.

2.2 Greenness

2.2.1 Physical and mental health effect of greenness

The influence of green space on physical activity is typically studied on adults in cross-sectional analyses using surveys to determine physical activity and using NDVI or land use datasets to measure neighbourhood greenness (James et al., 2015). The main conclusion from these studies is that greenness has a moderately positive association with physical activity (Fong et al., 2018; James et al., 2015). James et al. (2017) measured activity using GPS devices and found a strong nonlinear positive relation between

greenness and physical activity. The level of physical activity increased most when NDVI values were higher than 0.6. Also, for children, a positive relation between greenness and outdoor physical activity was found. Grigsby-Toussaint et al. (2011) found that an increase of 0.1 on the NDVI greenness approximately increased outdoor activity by 3 minutes a day. Klompmaker et al. (2018) and Ord et al. (2013) found no indications that the distance to a park is associated with physical activity or being overweight. However, a study by Maas et al. (2008) including a survey of almost 5000 Dutch people concluded that the amount of green space in the neighbourhood is scarcely related to physical activity. An explanation for this reverse outcome in the Netherlands could be that in general when a neighbourhood is greener, facilities are located further away and in greener neighbourhoods there is often a private car parking (den Hertog et al., 2006). However, agricultural lands increase the amount of green space availability which shows a positive relation to self-reported health (de Vries et al., 2003) and negatively to morbidity (Maas et al., 2009).

Both Fong et al. (2018) and James et al. (2015) found a positive relationship between green space and mental health by reviewing literature. A higher neighbourhood greenness is associated with higher self-rated mental health and lower depression prevalence. Several studies indicated that visual contact with green spaces, even through a window, can already have a positive effect on stress reduction and attention restoration (Ekkel & de Vries, 2017).

Greenness exposure is associated with an increase in birth weight (Fong et al., 2018; James et al., 2015). Hystad et al. (2014) found that an interquartile increase of residential NDVI increased birth weight and decreased the chance of preterm birth. Two studies associated higher levels of NDVI with a lower risk of cardiovascular mortality (Crouse et al., 2017; Vienneau et al., 2017). However, James et al. (2016) did not find this association.

2.2.2 Environmental variables greenness

In greenness exposure assessment studies several different environmental variables are used to represent exposure to green space. The most common environmental variables used to represent green space are NDVI and land-use or land cover databases (Fong et al., 2018). Only a few studies use different metrics for greenness, for example using surveys to measure the subjective perception of greenness (Weimann et al., 2015).

2.2.2.1 Normalized Difference Vegetation Index

The Normalized Difference Vegetation Index is an index used to determine the density of green on land. The principle behind NDVI is the difference in reflection between visible light and near-infrared from plants. Chlorophyll, the pigment in plant leaves, strongly absorbs all visible light (400 – 700 nm) for photosynthesis, for NDVI calculations the reflectance of red light is measured. The near-infrared light (700 – 1100 nm) is strongly reflected by the cell structure of the leaves (Weier & Herring, 2000). Therefore, more vegetation leads to a larger difference between the reflection of visible light and near-infrared light. Mathematically NDVI is written as:

$$NDVI = \frac{NIR - VIS}{NIR + VIS} \quad (1)$$

Where NIR is the near-infrared light and VIS is the visible light. NDVI values range between -1 and +1. Negative values indicate water bodies, values around zero bare soil and rock. NDVI values close to 1 indicate a high density of green leaves.

2.2.2.2 Land use/cover dataset

Land use and land cover datasets are maps showing different categories of land use. A potential advantage of land use and land cover datasets is that they may provide information about the quality and usefulness of the green space. But the often coarse spatial resolution of these datasets means that small scale vegetation patches as gardens and trees are not included in these datasets (James et al., 2015). Secondly, to generate these maps often several input datasets are used (Hazeu et al., 2014). Therefore, the categories and resolution of different land-use datasets are often not consistent due to a difference in input data or classification method. Finally, a temporal drawback of these datasets is that these datasets are often only updated every few years, leading to a potential temporal mismatch in the exposure assessment (James et al., 2015).

2.3 Exposure assessment

In general, there are two methods for exposure assessment, air monitoring and biological measurements. Air monitoring can again be divided in two groups: direct exposure in which a portable device measures personal exposure (Dons et al., 2011). The most direct greenness exposure measurement is using pedestrians videos to measure the greenness exposure (Hong et al., 2019). The second group are indirect exposure assessments in which measuring stations are used together with LUR models for air pollution and NDVI for example for greenness. These variables are combined with activity patterns to get a personal exposure assessment. Air pollution monitoring gives pollutant concentrations, thus the amount of pollutant per unit volume of air, but does not give information about how much of the concentration is found inside the human body. With biological monitoring the dose of a pollutant inside the human body is measured. This gives a better indication of health effects. Biological monitoring is mainly used for measuring the health effect for working environments at an individual level (Watson et al., 1998). Biological monitoring is also not applicable for greenness exposure assessments because greenness doesn't affect the human body directly.

Studies with the most direct exposure variable can however not be used for large scale exposure assessment. Therefore, larger scale exposure assessments use LUR or NDVI maps to quantify the exposure. These assessment studies use in general two approaches in their exposure assessment techniques, static and dynamic exposure assessments. In static exposure assessments static population data is used to calculate the residential exposure, hereby assuming that a person's activity only takes places at their residence. In dynamic exposure assessments human space-time patterns are added in the exposure assessment (Beckx et al., 2009). In most air pollution exposure assessments dynamic exposure techniques are used to study the exposure, because most studies are cohort studies in which the health effect of the studied population is measured using activity patterns. In greenness exposure studies mainly static exposure techniques are used, so only residential locations are used and human activity patterns are often neglected.

2.3.1 Static exposure assessment

In greenness exposure assessments, where greenness is expressed using NDVI values, exposure is mainly calculated as the mean NDVI value within a radius around the residence, the residential greenness (e.g. Gascon et al., 2015; Klompmaker et al., 2018; Ord et al., 2013). Several studies change all negative NDVI values to zero. The negative values represent blue space and would therefore negatively affect the mean green space value (Klompmaker et al., 2018).

Land use and land cover datasets are used in two different methods to quantify exposure to green spaces. The method used most is to measure the proportion of green space. The green space proportion is calculated by selecting the classes representing green space in an administrative boundary or a certain radius around a residence (Klompmaker et al., 2018; Vienneau et al., 2017). Another method is to measure the closest distance from a residence to green space, e.g. a park, using the road network (Klompmaker et al., 2018).

Agriculture contributes to most of the green space for most Dutch neighbourhoods (de Vries et al., 2003). Ekkel & de Vries (2017) suggest that therefore the access to agricultural areas is important in greenness exposure studies. They also conclude that cumulative opportunity indicators like the NDVI tend to show more consistent and better association with health indicators than residential proximity metrics. Also Klompmaker et al. (2018) and Ord et al. (2013) found that strongest associations for green space is given by NDVI surrounding greenness. NDVI can be used as a valid measure in quantifying the levels of residential greenness and are highly correlated with expert ratings of greenness (Rhew et al., 2011).

2.3.2 Dynamic exposure assessment

Activity pattern data together with the studied environmental variable are the main variables in personal exposure assessments (Steinle et al., 2013). The detail of activity patterns is based on the methodology of the exposure assessment. Most exact data is available in exposure assessments using GPS measurements. These studies have information on home and work locations and exact individual activity patterns for often a small population of individuals (e.g. Dias & Tchepel, 2014; Dons et al., 2011). Less detailed information is used in cohort studies (e.g. Dadvand et al., 2017; Ntarladima et al., 2019) in which the study population is known including their address. These studies often use time activity diaries or questionnaires to simulate the activity patterns (Steinle et al., 2013). For example, Ntarladima et al. (2019) used the OViN dataset a study on the mobility in the Netherlands. Or the American National Human Activity Pattern Survey (NHAPS) surveying almost 10,000 people used in Klepeis et al. (2001).

In data-sparse cases there is no information about precise living/working locations and/or travel routes. Therefore, different methods are needed to calculate exposure in the microenvironments visited during a day. Agent-based models are used to model human space-time activity because each agent will have its activity based on a set of probabilistic rules (e.g. Lu et al., 2019; Park & Kwan, 2017; Yang et al., 2018). Lu et al. (2019) studied two different social-economic groups, homemakers and bike commuters, using different space-time tracks in a data-sparse situation. The exact distribution of commuters and homemakers was not known and therefore they assume that all residence are homemakers or commuter, assuming no difference in the spatial pattern. Agent-based modelling is however not applicable in this study because for agent-based modelling often more input data is needed and secondly it needs a lot of computing power to calculate individual personal exposure on large populations.

The study by Dadvand et al. (2015; 2017) is currently one of the few greenness exposure assessments which uses human activity patterns. Patterns are represented as the main visited microenvironments for schoolchildren. Using a buffer of 20 metres around school location and travelling route between home and school and a buffer of 200 and 500 metres around the home location the NDVI greenness is measured. Zhang et al., 2018 used a questionnaire to represent activity data to study the impact of greenspace exposure using individual activity spaces. Buffer sizes of 100 metre and 500 metre were used for respectively home, school and work locations and travel routes.

Ntarladima et al. (2019) defined the microenvironments into the four main activities of children on a day (being at home, playing in the neighbourhood, travelling to/from school or being at school). For each activity, the area around the residence in which the activity is expected to take place is defined. Personal exposure is then calculated as the average concentration in the activity zone. A study by the European Commission proposed that accessibility is defined as within a 15-minute walk, corresponding to 500 metres on foot and 300 metres "as the crow flies" (European Commission, 2003). However, in previous described studies the residential locations were known and for global scale assessments these detailed datasets are not available.

Therefore, a global population dataset should be used together with a building dataset to estimate precise living locations. There are several global population datasets, e.g. the WorldPop dataset with a 3 arcsecond (~100 metres) resolution (Lloyd, 2017) or the 250 metre resolution JRC GHS dataset (Schiavina et al., 2019). Also, higher resolution datasets are available, but they lack total global coverage. For example, the 1 arcsecond (~30 metres) resolution high resolution settlement layer (Facebook Connectivity Lab and CIESIN, 2016) which only covers a few countries in the world. To downscale a population dataset to individual buildings a building dataset is needed. OpenStreetMap is an open source platform containing geographical information which can be used for buildings and road networks. Kloog et al. (2018) checked the validity of OSM data for an environmental exposure study focussing on road networks and concluded that OSM data is reliable. Liu et al. (2017) used OSM road data and POI data to downscale population census data.

3. Methods

3.1 Study areas

In this study exposure assessment is performed on two areas; the municipality of Utrecht in the Netherlands and the municipality of Madrid (Figure 1). To study the effect of intra-urban and agglomerate areas an area surrounding the city is added to the analyses. The municipality of Utrecht has an area of approximately 99.2 km² and a population of approximately 350,000, the total study area around Utrecht has an area of approximately 437 km² and a population of approximately 700,000. The municipality of Madrid has an area of 604 km² and a population of approximately 3,2 million, the total study area around Madrid has an area of approximately 1500 km² and a population of approximately 4,9 million.

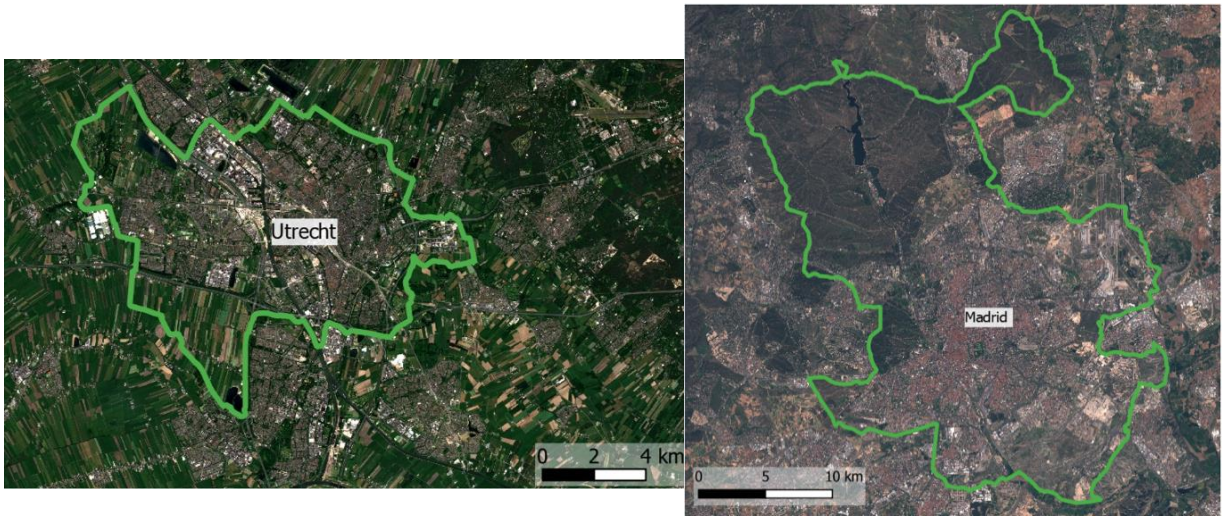


Figure 1. Study areas with the municipalities outlined in green. Left: Utrecht, The Netherlands right: Madrid, Spain. Source: Sentinel-2 (2017), processed by the European Space Agency (ESA).

The schematic overview of the workflow (Figure 2) was applied to both study areas. The three blocks at the top of the figure (orange) show the input datasets. Population data sets were resampled to a 25-metre resolution using a nearest neighbourhood resampling method and were then downscaled using residential buildings from the OSM building dataset. For the Netherlands, the downscaled population maps and the original population maps were compared with the CBS statistical population dataset. The downscaled population was used together with the exposure variable maps as input in the static exposure assessment. For the dynamic exposure assessment also space-time activity data was needed. The space-time activity data consist of activity patterns, the time spent per activity, and a spatial component which indicates where activities can take place. The spatial locations are obtained from the OSM building dataset which consists of the residential, school and working locations for both study areas.

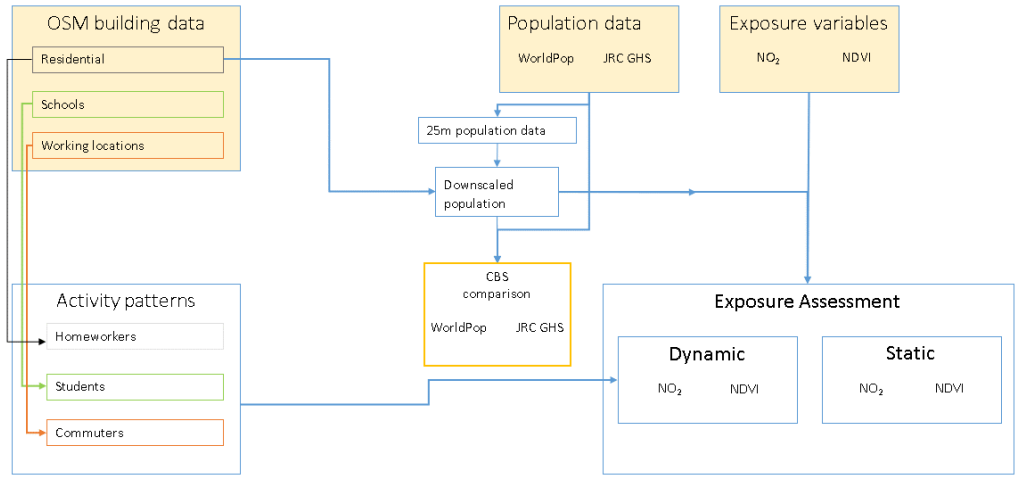


Figure 2. Schematic overview of the workflow.

3.2 Environmental variables

3.2.1 Air pollution

The air pollution maps for both Utrecht and Madrid consists of a yearly average NO_2 map from 2017. For the Netherlands, the air pollution data consisted of the yearly average NO_2 concentration measured in 78 NO_2 measurement stations. For Spain, the air pollution data consist of the yearly average NO_2 concentration measured in 380 station. The measurements from these stations are interpolated using several buffer predictors (e.g. total length of highway) and point predictors (e.g. population data) using a random forest model, for detailed information see (Lu et al., in press).

3.2.2 Greenness

Google Earth Engine (Gorelick et al., 2017) was used to process and calculate NDVI maps from the Sentinel-2A dataset with a resolution of 10 metres. From the Sentinel database, images from the months March till June were selected. These months were selected because the summer drought does not influence the greenness of the vegetation and winter conditions are gone. From this sub-selection, the image with the lowest cloudy pixel percentage was selected. For Utrecht, a Sentinel-2A image of 26/05/2017 was selected and for Madrid, the best image available was on 09/04/2017.

A Normalized Difference Vegetation Index (NDVI) was computed for both study areas using Sentinel 2 band 4 (red, 664.5 nm) and band 8 (near-infrared, 835.1 nm). The negative NDVI values were set to zero and the map was resampled to a 25 metres resolution to be compatible with the population data.

3.3 Exposure variables

3.3.1 Downscaling of census data

To study the effect of different population datasets two different datasets were used in this study. The 2015 JRC GHS population dataset (Schiavina et al., 2019) with a resolution of 250 metres and the 2017 WorldPop population dataset (Lloyd, 2017) with a resolution of 3 arcsecond (Netherlands: ~70 metres).

Using data from OpenStreetMap the population datasets were downscaled to higher resolution population maps (25-metre resolution). The OpenStreetMap dataset consists of vector-based individual building information downloaded using Geofabrik (http://download.geofabrik.de). Two methods were used in the selection of the buildings from the OSM dataset. The first method was to select only buildings with a residential tag.

Not all residential buildings are tagged as residential buildings in OSM, especially in Madrid, therefore the second method added all unclassified buildings with an area larger than 20 m^2 inside a residential area in the OSM land use layer to the building selection (Appendix A). The residential building polygons were transformed to point features, representing the centre of the building using QGIS 3 (QGIS Development Team, 2019) and rasterized to a 25-metre resolution raster using the "Point to Raster" function in ArcMap 10 (ESRI, 2012), pixel values showing the number of buildings per pixel. Raster calculations are performed in the Python programming language with the PCRaster library (Karssenberg et al., 2010).

The population of a single original pixel was only divided over the 25-metre pixels containing buildings inside the original pixel. The downscaled population was calculated on every original pixel using equation 2, here given for the GHS resolution:

$$population_{25} = population_{250} * \frac{number\ of\ houses_{25}}{number\ of\ houses_{250}} \quad (2)$$

Where $population_{25}$ and $population_{250}$ is the number of people living in respectively a 25-metre or 250-metre pixel. N$umber\ of\ houses_{25}$ is the number of houses located in a single 25-metre pixel and $number\ of\ houses_{250}$ is the sum of all houses in the 250-metre pixel. This equation, however, will not downscale population from 250-metre pixels without buildings. Therefore, the population that was located in pixels without buildings was equally divided over all buildings in the map using:

$$Downscaled\ population_{25} = population_{25} + \left(total\ number\ of\ 'homeless' / \frac{total\ number\ of\ houses}{number\ of\ houses_{25}}\right) \quad (3)$$

Where $total\ number\ of\ 'homeless'$ is the total number of people located in building free pixels. $Total\ number\ of\ houses$ indicates the number of houses in the entire study area.

To compare the accuracy of the population maps and the two methods for downscaling, the downscaled population maps of the Netherlands were compared with the 2017 CBS population dataset, the dataset consist of square polygons of 100 by 100 metre (Centraal Bureau voor de Statistiek, 2019). Also the original population datasets were compared to the CBS data to see the effect of downscaling. For the comparison the "zonal statistics" function in QGIS 3.4 was used, which calculates the sum of all pixels inside each CBS polygon.

3.3.2 Human activity patterns

Because this study aims to study large scale population-wide exposure assessment, human activity patterns cannot be modelled as space-time tracks for all individuals. Therefore, human activity was represented using zones of activity. The zones represent an area around a specific location in which the activity most likely occurs. The following zones were used;

- Activity at home
- Activity at work or study
- Commuting

These zones were used for three different social-economic groups with a distinct mobility pattern for each of the following groups, commuters, homemakers and students.

For homemakers the only activity zone is activity at home, so homemakers stay throughout the day at the home location. Individual study and work locations for students and commuters are not known, therefore potential study and work locations were selected using OSM data. Potential study locations include all educative related buildings tags, except buildings tagged as kindergartens (Appendix A). Potential working locations include all commercial buildings in the OSM database (Appendix A). The microenvironment for commuting consists of all roads in the OSM dataset. Students and commuters were assumed to stay at home for 15 hours, to travel 1 hour and to work/study 8 hours a day.

3.4 Personal exposure assessment

To study the effect of downscaling and the effect of different resolutions of population data on personal exposure a static exposure assessment was performed. The downscaled population was then used together with the space-time activity data to create the social-economic groups for the personal exposure assessment.

3.4.1 Static exposure assessment

To study the effect of the resolution of population data on static exposure assessment, population data was homogeneously divided over 3 different raster sizes. Firstly, the total population in the entire map was calculated and evenly divided over every 25-metre resolution cell, this represents the natural environmental variable distribution, this scenario will be mentioned in this paper as the natural distribution. Secondly, the study area was divided in square grids of 1 by 1 kilometre, for each grid the total population inside the grid was calculated and then divided over the cells in the grid. Thirdly, the original population data was resampled to 25 metres using a nearest neighbour resampling technique. Finally, to study the effect of downscaling the downscaled population distribution was added, the total distribution of these different exposure assessments scenarios were compared. The effect of downscaling was calculated by calculating the ratio of the mean exposure of the natural distribution versus the mean exposure using the downscaled population.

3.4.2 Personal exposure using activity patterns

In general, exposure to the environmental variable can be calculated using (Watson et al., 1998):

$$E_i = \sum_{j}^{J} C_j t_{ij} \tag{4}$$

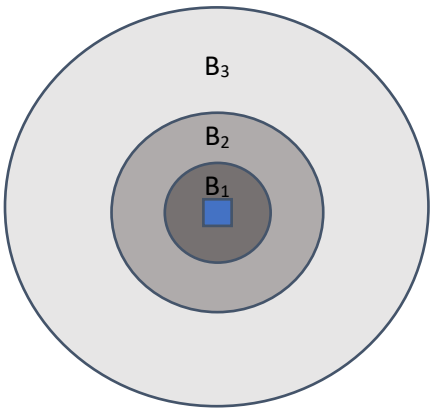
Where E_i is the personal exposure, for air pollution given as (µg/m^3), for person i. C_j is the environmental concentration for microenvironment j, t_{ij} is the total time that person i spends in microenvironment j. J is the total number of microenvironments visited by person i. For microenvironment j it is assumed that the concentration of C_j is spatially homogenous over the 25-metre resolution grid. As not all ambient air pollution infiltrates a building a constant indoor proportion R was used for all indoor microenvironments in the air pollution exposure assessment. The R of a building is normally affected by factors such as infiltration, ventilation but also warm and cold seasons (Rivas et al., 2015; Yang et al., 2004). In this study a constant R factor if 0.7 was used, based on studies by WHO (2010) and Yang et al. (2004).

The indoor microenvironments for home, work and school locations were represented by a 50 metre circular buffer. This buffer is increased to 100 metre for the greenness exposure assessment. The commuting microenvironment for the air pollution assessment was defined as the NO_2 concentration at the road pixels only for the road pixels located inside

the zone of activity. For the greenness exposure assessment, a buffer of 50 metres was added around each road pixel to calculate the average NDVI value within this buffer (Dadvand et al., 2015).

The static exposure assessment in Section 3.4.1 was used to study the effect of different levels of detail in the population data on the spatial distribution of personal exposure. Besides the spatial pattern, it is also important to study the distribution of personal exposure over the population. This was studied using the different activity pattern groups described in Section 3.3.2. Like in Lu et al. (2019), the spatial distribution of residences in each group is not known and therefore the assumption was made that all residences are part of the chosen social-economic groups. Like in Ntarladima et al. (2019) it is not known to which specific school or work location a person is travelling, therefore zones around each residence was used in which all locations are used, a distance weighting was added in this study due to the large study area.

The exposure was calculated as the average exposure for the specific microenvironment in the zone and weighted according to the weight of the corresponding buffer. A distance weighting was added based on the idea that for space-time activity patterns, it is more likely that a location close by is visited in comparison with a location at a larger distance. In this study, this means that there is a higher chance that a person spent time in a microenvironment close by than further away. The distance weighting consisted of three circular buffers around each residence with a size of 1000, 2000 and 100000 metres around each residence. The weighting of these buffers was 100%, 90% and 50% respectively. For example, in the 2000 metre buffer only schools or working locations are used in the range of 1000 – 2000 metre (Figure 3).



$$\widetilde{B_1} = \frac{\sum C_{b1}}{N_{B_1}} * f_1 \quad (5)$$

$$\widetilde{B_2} = \frac{\sum C_{b2} - \sum C_{b1}}{N_{B_2} - N_{B_1}} * f_2 \quad (6)$$

$$\widetilde{B_3} = \frac{\sum C_3 - \sum C_2}{N_{B_3} - N_{B_2}} * f_3 \quad (7)$$

$$C_{school} = \sum_{i=1}^{3} \tilde{B}_i \Big/ \left(\sum_{i=1}^{3} N_{B_i} * f_i \right) \quad (8)$$

Figure 3 Schematic representation of the three different buffer sizes with the formulas used to calculate the average exposure per buffer $\tilde{B}$ and for example the exposure for schools C_{school} for each residence. The blue box represents a house with the three buffers of 1000, 2000 and 10000 metre around it. $\sum C_i$ is the sum of the exposure for all pixels in buffer i.N_{B_i} is the count of all pixels in buffer i. f_i is the weighting for buffer i.

The individual exposure was calculated by expanding equation 4 for all primary activities associated with the corresponding socio-economic group. For example, the individual exposure of a student is calculated using:

$$E_{student} = \frac{C_{home} * t_{home} + C_{school} * t_{school} + C_{commuting} * t_{commuting}}{t_{home} + t_{school} + t_{commuting}} \quad (9)$$

C_{school} was calculated for every school building inside the circular buffer and weighted by the distance from the residential location (Figure 3). $C_{commuting}$ was calculated in the same buffer but only for roads inside this buffer, because commuting takes place on roads. For greenness, the buffer around a road was increased because greenness is visible further away while travelling on a road.

3.5 Sensitivity analyses

To study the importance of a distance decay weighting and the effect of different ring sizes on the exposure assessments sensitivity analyses were performed for both variables. In the sensitivity analyses six situations were assessed (Table 1). For each situation, one variable was adjusted while the other variables were fixed.

Table 1 Variable situations for the weighting of the buffers (%) and the buffer size (m) used in the sensitivity analyses. Scenario 2 is the standard scenario.

Testing variable	Weight per buffer (%)			Buffer (m)			Scenario
	1st	2nd	3th	1st	2nd	3th	#
Weighting	100	100	100	1000	2000	10000	1
	100	*90*	*50*	*1000*	*2000*	*10000*	*2*
	100	50	25	1000	2000	10000	3
	100	50	10	1000	2000	10000	4
Buffer size	100	90	50	1000	2000	10000	2
	100	90	50	1000	2000	5000	5
	100	90	50	300	2000	10000	6

4. Results

4.1 Environmental variables

4.1.1 Air pollution map

The spatial patterns of air pollution concentrations (NO_2, µg/m^3) in Utrecht and Madrid (Figure 4) show that the mean NO_2 concentration for the entire study area is lower in Utrecht than Madrid, 21.29 and 28.47 µg/m^3 respectively. For the municipality of Utrecht and Madrid the mean NO_2 concentrations are 26.91 and 32.90 µg/m^3 respectively. Both cities show a spatial pattern of high concentrations in the city with increased concentrations on roads. The NO_2 concentration for Utrecht is in the range of 15.24 - 42.64 µg/m^3 and for Madrid it is in the range of 8.21- 60.62 µg/m^3.

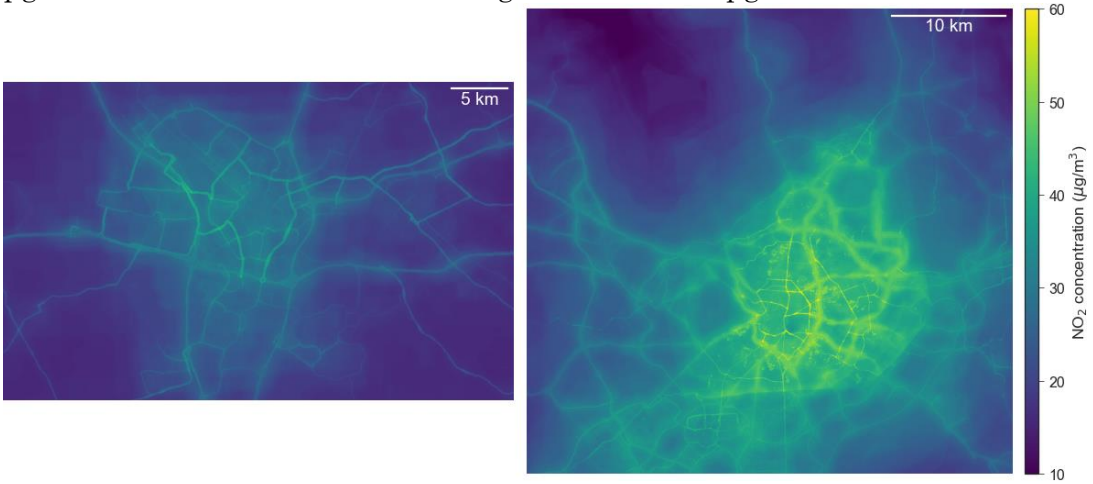


Figure 4. Spatial pattern of NO_2 concentrations (µg/m^3) for Utrecht (left) and Madrid (right). See Figure 1 for study area overview.

4.1.2 NDVI map

The spatial NDVI patterns for Utrecht and Madrid both show a general trend of a low NDVI value in the city and higher NDVI value in the agglomerate areas (Figure 5). The average NDVI value for the total map of Utrecht is higher compared to the total map of Madrid with a mean NDVI value of 0.66 and 0.41 respectively. The municipality of Madrid is characterized by mainly larger areas of higher NDVI and large, almost continuous, zones of low NDVI values (<0.1), with a mean NDVI concentration of 0.39. The municipality of Utrecht is characterized by low NDVI values in the city centre, but in residential areas a diverse pattern of low and high NDVI values is visible. The mean NDVI value of the municipality of Utrecht is 0.53.

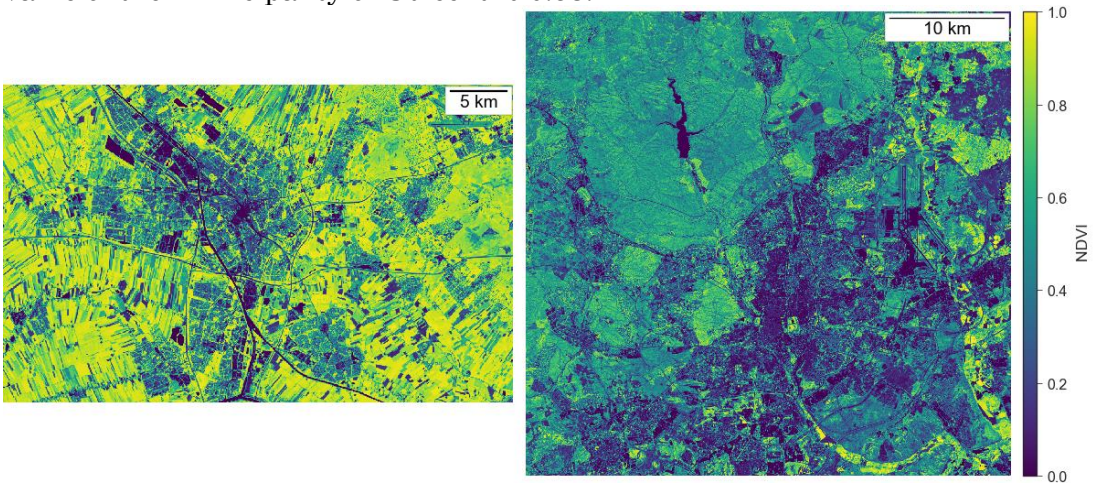


Figure 5. Spatial pattern of the Normalized difference vegetation index (NDVI) for Utrecht (left) and Madrid (right). See Figure 1 for study area overview.

4.2 Exposure variables

4.2.1 Downscaling

The spatial distribution of Utrecht and Madrid (Figure 6) show that population density in the municipality of Madrid is higher compared to the municipality of Utrecht. A subset of the centre of both cities shows that every pixel contains population (Figure 6 row 2). When the population is downscaled the number of pixels containing population decreases (Figure 6 row 3) and this shows the actual distribution of houses in the city.

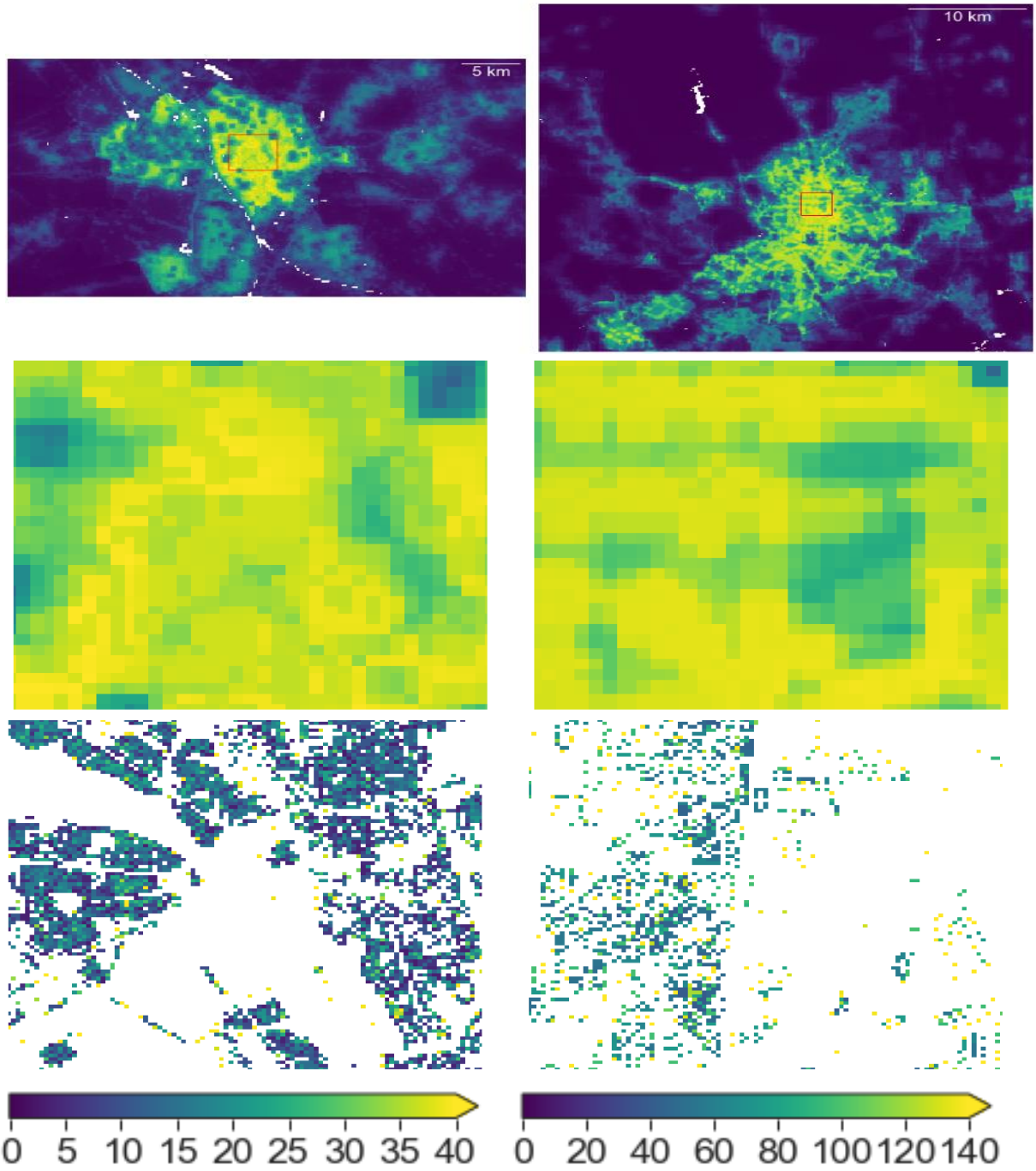


Figure 6 WorldPop Population maps for Utrecht (left) and Madrid (right). Top: Original WorldPop population map, white areas no data. Centre: 250 x 250 metre subset of the red square (top). Bottom: Same area as centre panels showing the downscaled population. Colorbar is applicable to all panels in the same column.

The original population datasets compared to the reference CBS population data the WorldPop dataset has a lower mean error compared to the original GHS dataset, 26.22 versus 29.93 respectively. The original GHS dataset underestimates the reference population. In both original population datasets the data does not show the trend of the 1:1 line (Figure 7). When the original datasets are downscaled the accuracy of the datasets compared to the reference CBS data increases and more points are located around the 1:1 line. There is still a difference in the accuracy between the two different population datasets for Utrecht (Figure 7). The downscaling of the WorldPop dataset better represents the reference CBS data than downscaling using the GHS population dataset. This is best visible in the logarithmic hexbin plots, showing that the high number of counts of the WorldPop data is more clustered compared to the GHS data where the points are more scattered (Appendix B). The accuracy of only using the residential buildings or adding the unclassified buildings inside residential land use areas is almost the same comparing the r^2 and the RMSE. When the mean error for each CBS grid is calculated, it shows that the residential buildings downscaling method has a lower error per grid compared to the land-use area, 0.16 versus 1.52 respectively for the WorldPop population dataset. This result is also found for the GHS, with a mean error of 7.47 versus 8.90.

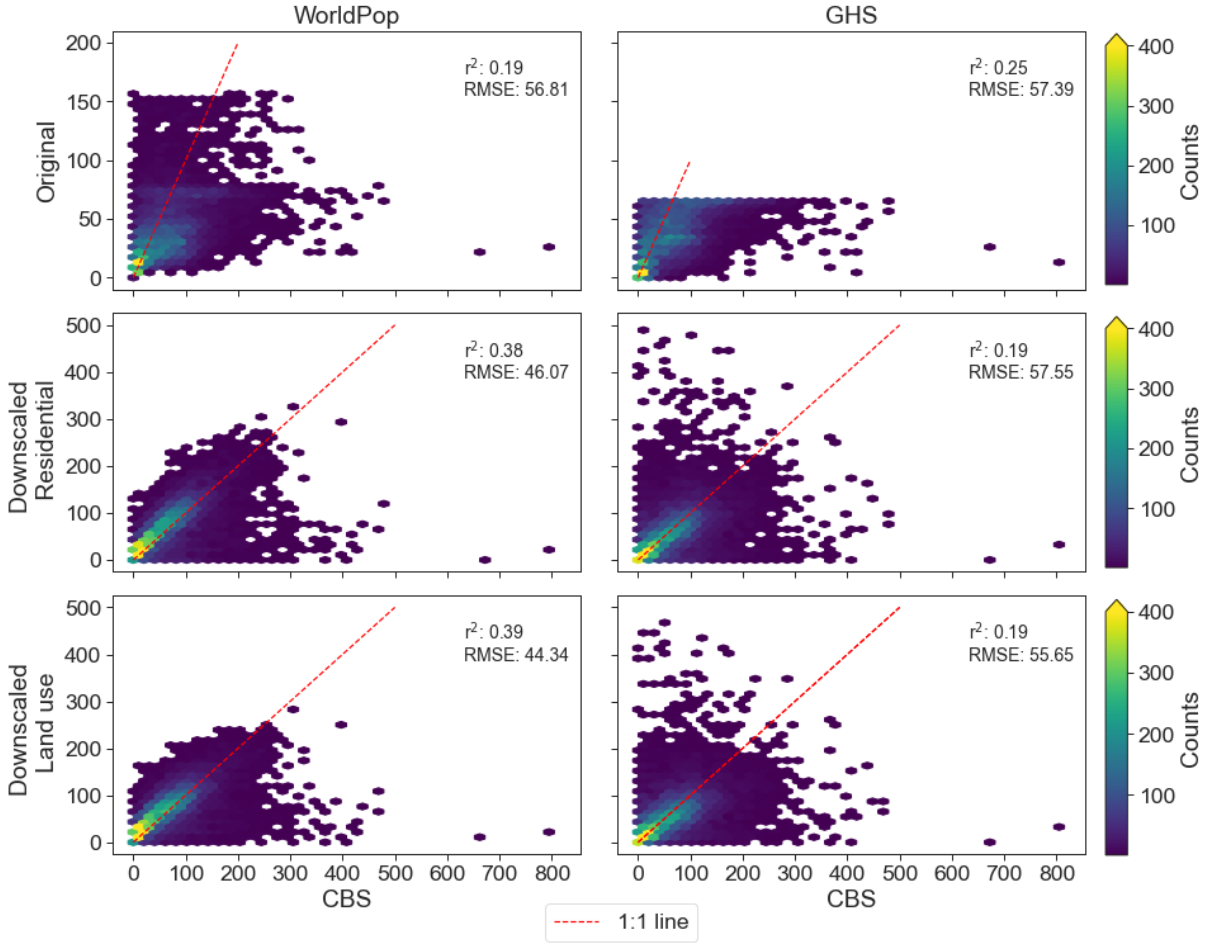


Figure 7 Hexbin plot of the reference population data (CBS) values on the x-axis versus the population data of the WorldPop (left) and GHS dataset (right) for Utrecht. For the logarithmic plot see Appendix B. The upper row (original) shows the original population datasets. The centre row (residential) shows the downscaled population datasets using buildings tagged as residential. The lower row (land use) shows the population datasets downscaled using the land use method, so the unclassified buildings inside land use areas classified as residential are added. The colorbar shows how many points are located inside each hexagon, each point has a resolution of 20 persons, all 10283 samples are used.

4.3 Exposure assessment

4.3.1 Static exposure

The effect of different population datasets and the effect of different resolutions in population data is represented in Figure 8 by using two different population datasets and by dividing the population data using different grids. The first column in the figures illustrates the natural NO_2 pattern. The distributions of NO_2 (Figure 8A, row 1 & 2) show that both datasets have the same positive skewed distributional shape. With a peak of ~38,000 residence exposed to 16.3 µg/m³. Despite the difference in the two population datasets the median (19.83 µg/m³) NO_2 exposure is the same. When the population datasets are dived in grids of 1 kilometre [B] and 25 metre [C] the median NO_2 exposure between the WorldPop and GHS dataset are not similar, median NO_2 of 19.83 versus 20.23 µg/m³ (Figure 8B) and 19.82 versus 22.18 µg/m³ (Figure 8C).

Both the 25-metre grid and the downscaled population show population data on 25-metre resolution, the median exposure is, however, increased by 2.91 µg/m³ using the downscaled population. Secondly, the median NO_2 exposure using the downscaled population is the same for both population datasets (22.74 µg/m³). In the downscaled distribution a multimodal distribution is visible with 2 peaks. Most residences are exposed in the range of 18 - 21 µg/m³, then a decrease is visible around 23 - 24 µg/m³ and then a second large group of residences has an NO_2 exposure in the range of 25 - 31 µg/m³.

The distributions for the municipality of Utrecht (Figure 8 row 3 & 4) has a higher minimum NO_2 concentration (17.41 µg/m³) and the median NO_2 concentrations are higher for all scenarios with an increase of approximately 7 µg/m³. For the natural spatial NO_2 pattern (Figure 8A), the main peak is around 28 µg/m³ with some smaller peaks at low concentrations and a median exposure of 27.45 µg/m³. The other scenarios show an almost symmetric distribution with a distinct peak. Just as in the total study area scenario an increase in the median NO_2 exposure is visible for the downscaled scenario compared to the 25-metre grid scenario. Just as in Figure 8A, the median of the downscaled scenarios (Figure 8D) are the same between the different population datasets. For the downscaled scenario the median of the distribution (28.62 µg/m³), is located near the peak of the distribution at 29.25 µg/m³ and ~15,000 residence.

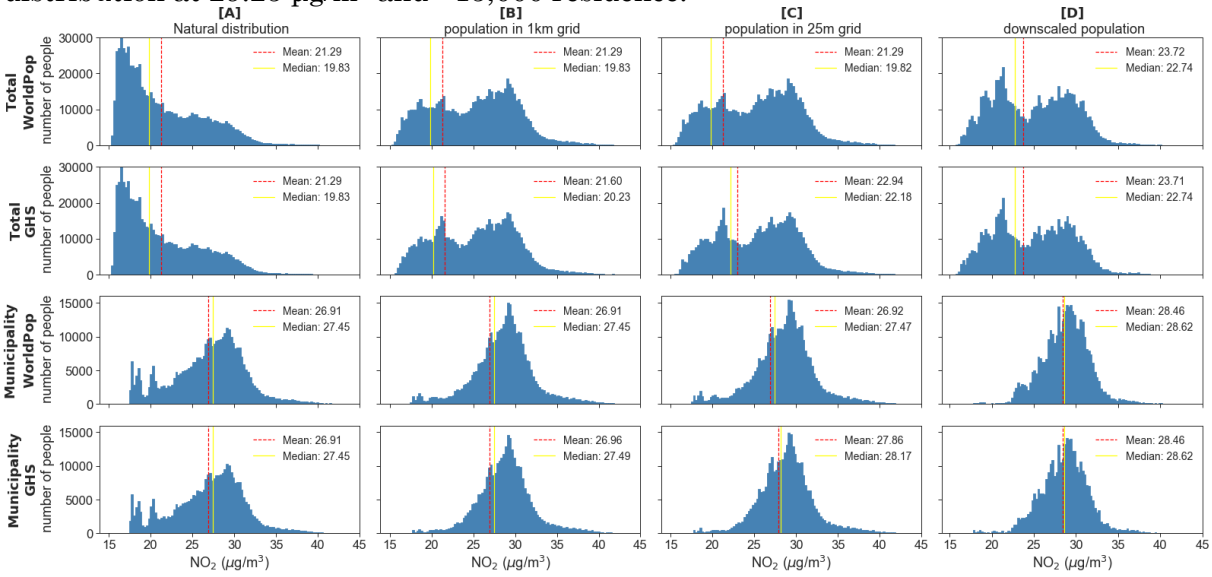


Figure 8. Static NO_2 exposure distribution (µg/m³) over the population of Utrecht. Shown for the two population datasets (WorldPop and GHS). The first two rows show the distribution over the total study area, the last two rows show the distribution only for the municipality of Utrecht. The first three columns show the total number of residences evenly divided over the area but at different grid sizes, total map [A], 1 km grid [B] and 25m grid[C]. The last column shows the NO_2 distribution using the downscaled population [D]. The yellow line is the median exposure and the red dotted line represents the mean concentration.

The static NO_2 exposure distribution for the residences in Madrid (Figure 9) show a completely different shape for all scenario's compared to the distribution of Utrecht. The distributions for the total study area (Figure 9A, row 1 & 2) show that there is no distinctive peak but a gentle increase in population exposed until 24 µg/m^3 and then the number of exposed people declines slowly until a small increase at ~42 µg/m^3. The smaller grids distributions (Figure 9B, C, D) can be described as negatively skewed distributions with one distinctive peak around ~42 µg/m^3. The 1-kilometre and 25-metre grid scenarios (Figure 9B & C) in Madrid also show a few units difference for the median between the two different population datasets. The median exposure between (Figure 9C) and the downscaled scenario (Figure 9D) is increased with 12.45 µg/m^3. In the downscaled scenario the median is almost identical (40.08 and 40.06 µg/m^3). The peak of the exposure is at 42.56 µg/m^3 exposing ~25,4000 persons.

The distribution for the municipality of Madrid (Figure 9 row 3 & 4) show that the exposure for the lower NO_2 concentrations declined and the minimum NO_2 concentration increased to 12.23 µg/m^3. The median NO_2 exposure increases on average with 5.8 µg/m^3 for the WorldPop scenario's and for the GHS scenarios the median increases with 6 µg/m^3 on average. In the downscaled scenario (Figure 9D) the distribution is characterized by a small range with one distinctive peak at 42.57 µg/m^3.

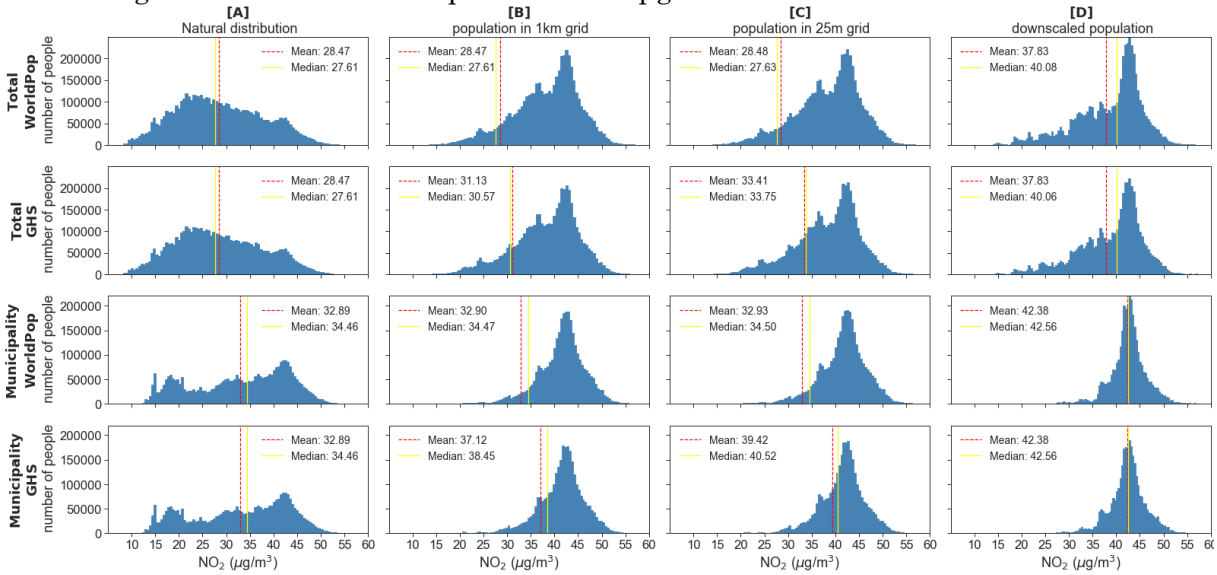


Figure 9. Static NO_2 exposure distribution (µg/m^3) over the population of Madrid. Shown for the two population datasets (WorldPop and GHS). The first two rows show the distribution over the total study area, the last two rows show the distribution only for the municipality of Madrid. The first three columns show the total number of residences evenly divided over the area but at different grid sizes, total map [A], 1 km grid [B] and 25m grid[C]. The last column shows the NO_2 distribution using the downscaled population [D]. The yellow line is the median exposure and the red dotted line represents the mean concentration.

Both Utrecht and Madrid show that the median concentrations for the WorldPop population dataset scenario [A] and [B] are the same. Comparing the median exposure in scenario [A] and [B] with scenario [C] the median exposure decreases with 0.02 µg/m^3 in Utrecht and increases with and 0.02 µg/m^3 in Madrid. While for the GHS dataset the median increases with 2.3 µg/m^3 for Utrecht and even a larger increase is found in Madrid with an increase of 6 µg/m^3.

The natural NDVI distribution (Figure 10A row 1 & 2) show an almost reversed distribution in comparison with the NO_2 distribution (Figure 8A row 1 & 2). The distribution is negatively skewed with the peak of the exposure at 0.94 exposing ~85,300 persons. The distributions for the smaller grid sizes (Figure 10B & C) show almost the same distributional shape as in the total scenario (Figure 10A). The mean and median values for the WorldPop dataset are almost the same for the first three columns (0.66 ± 0.01 and 0.77 ±0.01). In the 1-kilometre grid the difference between the median NDVI values for the WorldPop and GHS dataset are still negligible (-0.01). In the 25-metre grid scenario the difference between the WorldPop and GHS dataset increases.

The distribution of the downscaled population scenario (Figure 10D) shows a positive skewed bell-shaped distribution, with a decrease in the exposure at a NDVI value of 0.49. Compared to the natural NDVI distribution (Figure 10A) the median NDVI values decreases with 0.36. Both population datasets have the same median NO_2 exposure.

For the municipality of Utrecht (Figure 8A row 3 & 4) the shape of distributions are the same as for the total study area, however the height of the peaks of the NDVI values > 0.8 is halved. Therefore, the median NDVI value decreases for all scenarios. For the total map and 1-kilometre grid scenario the decrease in the median NDVI values is the biggest with a decrease of 0.21. For the downscaled scenario the median NDVI values is 0.35.

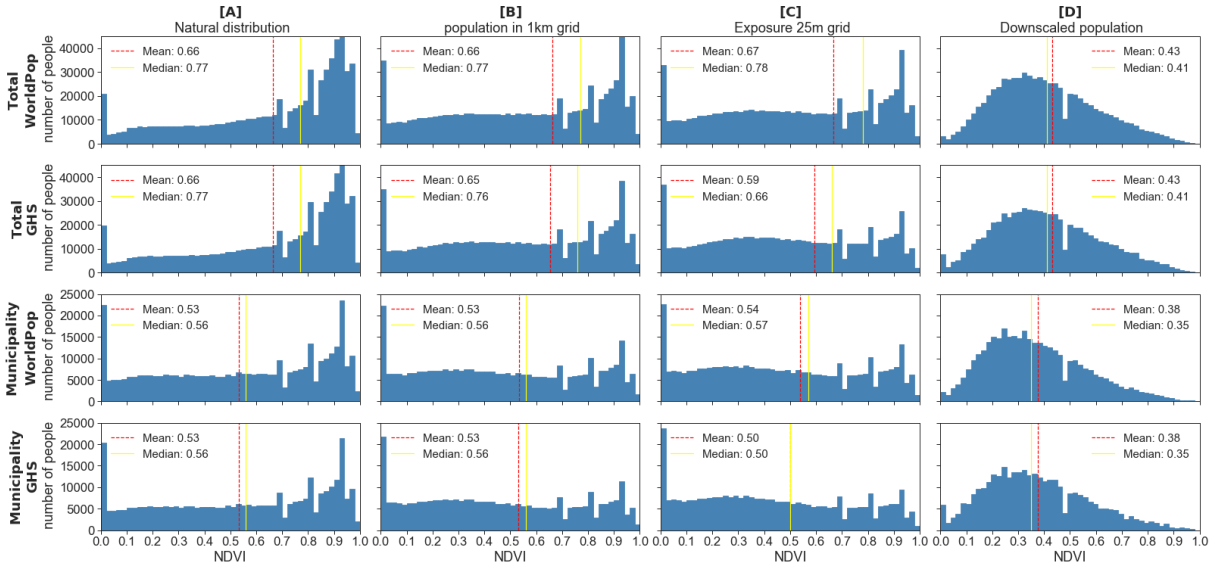


Figure 10. Static NDVI exposure distribution over the population of Utrecht. Shown for the two population datasets (WorldPop and GHS). The first two rows show the distribution over the total study area, the last two rows show the distribution only for the municipality of Utrecht. The first three columns show the total number of residences evenly divided over the area but at different grid sizes, total map [A], 1 km grid [B] and 25m grid[C]. The last column shows the NO_2 distribution using the downscaled population [D]. The yellow line is the median exposure and the red dotted line represents the mean concentration.

The natural NDVI exposure distribution for the total study area in Madrid (Figure 11 row 1 & 2) show a peak at a NDVI value of 0 and is further characterised by an increase in the number of people exposed to NDVI around 0.1 and 0.6. The smaller grid scenario's (Figure 11B, C & D) are positively skewed distributions. The 1 kilometre and 25 metre scenario have a peak at 0.0 and the number of persons exposed gradually decreases with an increase in NDVI. The peak of downscaled distribution (Figure 11D) is between 0 - 0.1 exposing 2,3 million, an increase in the NDVI values shows a quick decrease in the number of people exposed.

For all the scenarios in the municipality of Madrid (Figure 11 row 3 & 4), the shape of the distributions are almost the same as the shape of the total study area distributions (Figure 11 row 3 & 4). Just as in the total study area (Figure 11 row 1) the median NDVI value for the WorldPop dataset does not change for the first three columns. For the GSH dataset the median NDVI values decreases for the smaller grid scenarios in both the total study area as in the municipality. In the municipality the median NDVI value decrease from 0.42 [A] to 0.31 [B] and 0.24 [C]. In the downscaled scenario (Figure 8D) the peak of the distribution is still located between 0.0 - 0.1 but the number of people exposed decreases to 1.5 million. The difference in the median NDVI value is only 0.01 (0.12 versus 0.11).

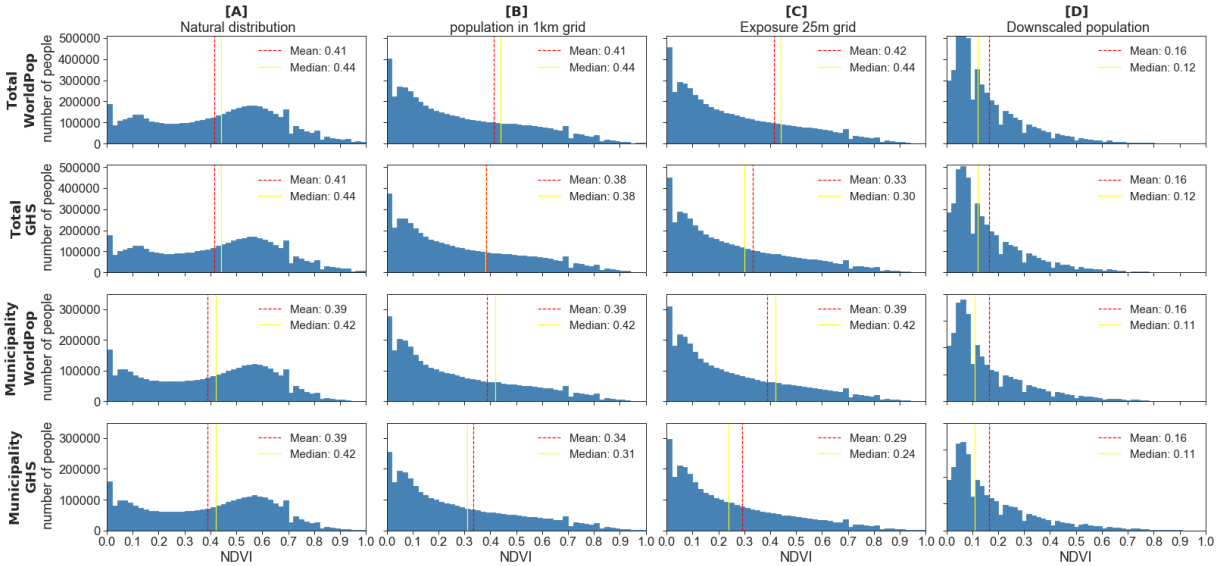


Figure 11. Static NDVI exposure distribution over the population of Madrid. Shown for the two population datasets (WorldPop and GHS). The first two rows show the distribution over the total study area, the last two rows show the distribution only for the municipality of Madrid. The first three columns show the total number of residences evenly divided over the area but at different grid sizes, total map [A], 1 km grid [B] and 25m grid[C]. The last column shows the NO_2 distribution using the downscaled population [D]. The yellow line is the median exposure and the red dotted line represents the mean concentration.

4.3.2 Personal exposure

The spatial distribution of E_{NO_2} shows in general that the highest exposure is found in the municipality of Utrecht and Madrid especially near the major roads (Figure 12 & Figure 13). The general trend for homemakers is that E_{NO_2} decreases when moving away from the municipality, with increased E_{NO_2} for residences near a road and living in the smaller cities (Figure 12). This trend is also visible for students and commuters, however, the decrease in E_{NO_2} when moving away from the municipality and moving away from roads is more gradually. The difference between the E_{NO_2} of students and commuters versus homemakers is that in general in the municipalities and near roads homemakers have a higher E_{NO_2}. While in the city agglomerate areas students and commuters have a higher E_{NO_2}.

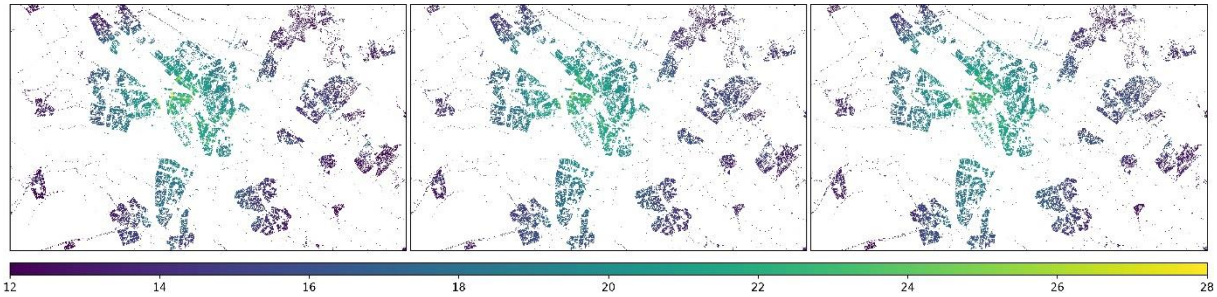


Figure 12. NO_2 exposure (µg/m³) Utrecht for homemakers (left), students (centre) and commuters (right). The map shows the exposure for each location based on the downscaled building map. See Appendix C for larger images.

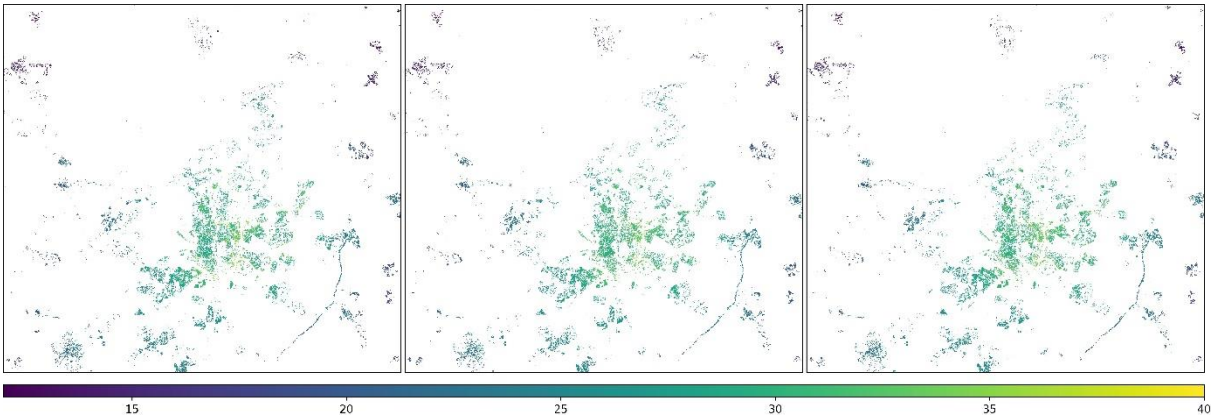


Figure 13 NO_2 exposure (µg/m³) Madrid for homemakers (left), students (centre) and commuters (right). The map shows the exposure for each location based on the downscaled building map. See Appendix C for larger images.

The spatial distribution of E_{NDVI} (Figure 14 & Figure 15) shows a reverse trend compared to the spatial distribution of E_{NO_2}. Residences in agglomerate areas are exposed to higher levels of NDVI compared to residences in the municipality. Commuters and students in the municipality of Utrecht have a lower NDVI exposure than homemakers and the exposure of commuters and students in agglomerate cities are higher than for homemakers.

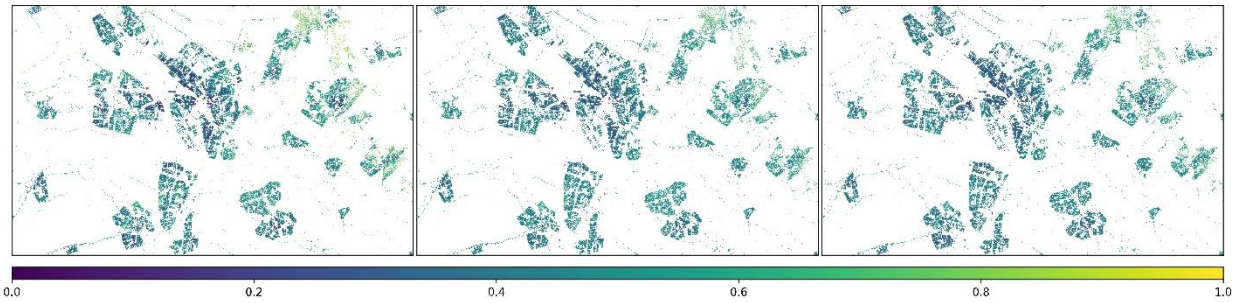


Figure 14 NDVI exposure Utrecht for homemakers (left), students (centre) and commuters (right). The map shows the exposure for each location based on the downscaled building map. See Appendix C for larger images.

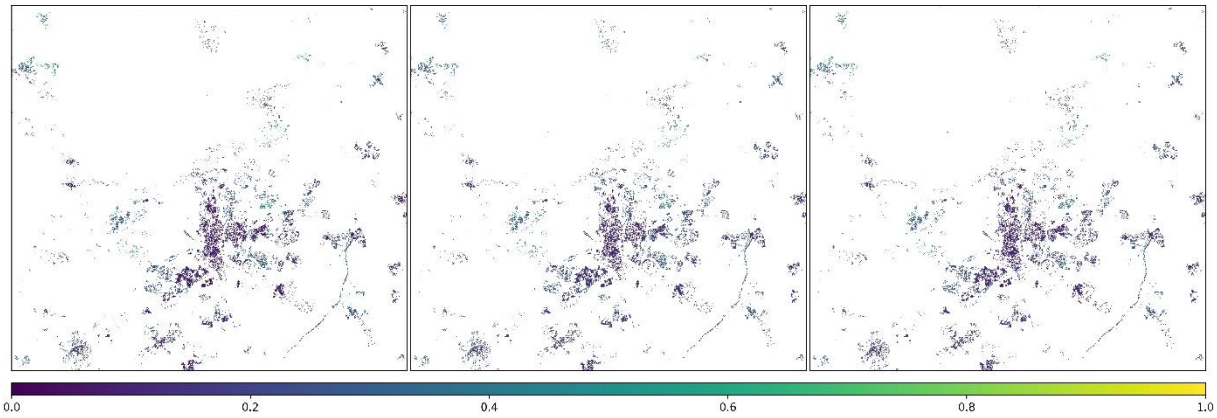


Figure 15 NDVI exposure Madrid for homemakers (left), students (centre) and commuters (right). The map shows the exposure for each location based on the downscaled building map. See Appendix C for larger images.

Students and commuters have the same time patterns for their activity, the only difference between the groups is the location of school and work. In Figure 16 the difference between these social-economic groups is shown, see Appendix D for all difference plots. The maximum NO_2 difference between commuters and students is 1.92 µg/m^3 and -2.22 µg/m^3 for Utrecht and Madrid respectively. The median difference is the same for both cities 0.06 µg/m^3. The maximum NDVI difference is -0.1 and -0.11 for Utrecht and Madrid respectively. The median change is -0.02 and -0.01. The spatial pattern shows that in Utrecht and Madrid commuters have a higher NO_2 exposure compared to students mainly for the agglomerate areas, the NO_2 of students is higher in the city centre. For the NDVI students exposure is mainly larger for residences in the municipality.

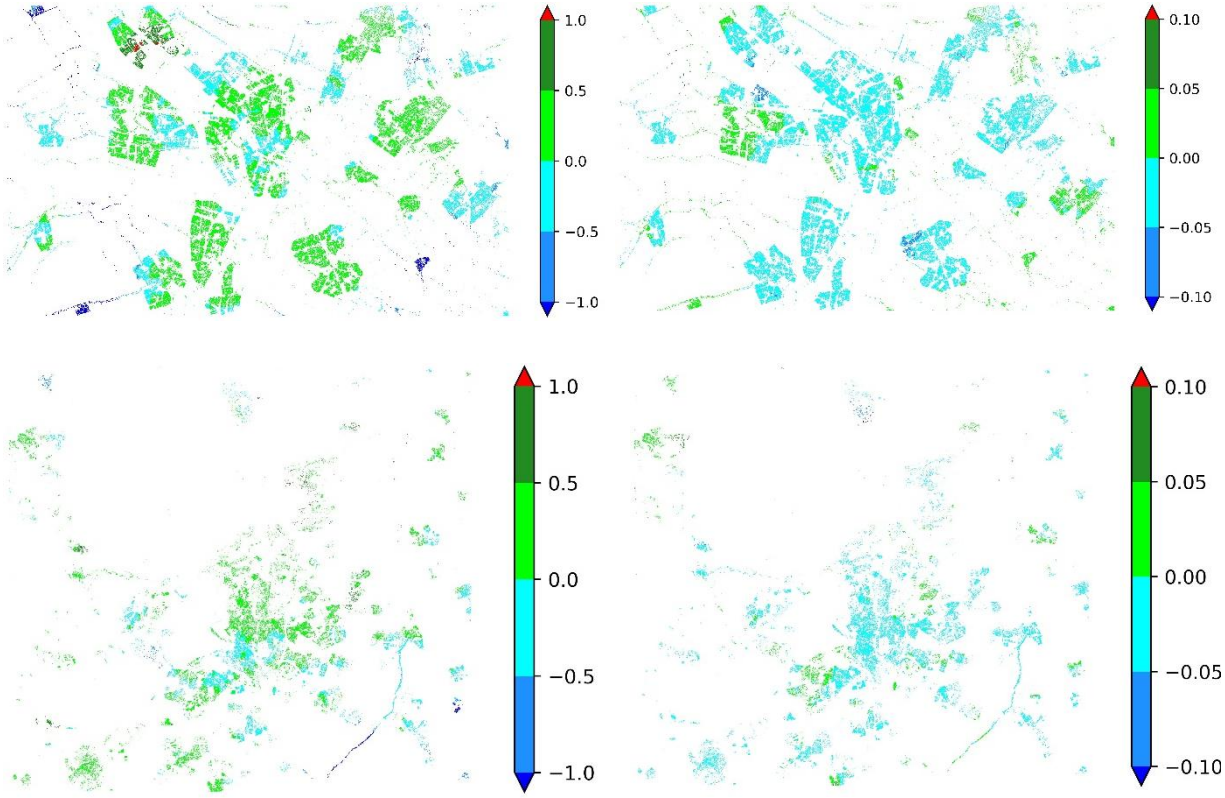


Figure 16 The difference between commuters and homemakers (left) and commuters and students (right) (E_{NO_2} of commuters - E_{NO_2} of students) for Utrecht (upper) and Madrid (lower) for NO2 µg/m^3 (left) and NDVI (right). See Appendix D for all difference plots.

The E_{NO_2} (µg/m^3) distribution for the different social-economic groups show that for Utrecht and Madrid the median E_{NO_2} is highest for commuters (16.56 and 28.62 µg/m^3) (Figure 17) (Appendix E). The median E_{NO_2} for homemakers is in both cities the group with the lowest E_{NO_2}(15.93 and 28.08 µg/m^3). The E_{NO_2} of all the social-economic groups are highly correlated (Pearson correlation efficient: 0.99 or 1.0). The more gradual decrease in the E_{NO_2} for homemakers compared to students and commuters is also visible in the distribution, this is shown by the greater interquartile range of homemakers (5.3 µg/m^3) compared to students and commuters (4.8 µg/m^3).

The social-economic group with the highest exposure to NDVI is different for Utrecht and Madrid (Figure 18). In Utrecht, homemakers are exposed most to greenness (0.49), while in Madrid students have the highest exposure to greenness (0.22). Commuters are in Utrecht and Madrid the group with the lowest exposure to greenness, 0.47 and 0.21 respectively.

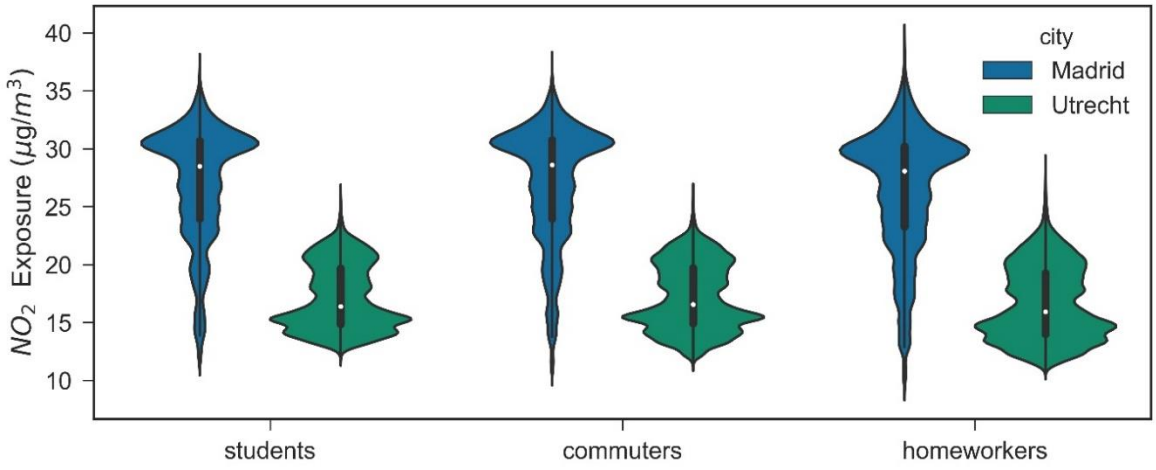


Figure 17. NO_2 exposure (µg/m^3) for students, commuters and homemakers in Utrecht (green) and Madrid (blue). Each boxplot shows the distribution of NO_2. White dot represents the median. The black bar shows the interquartile range and the black line shows the 1.5x interquartile range.

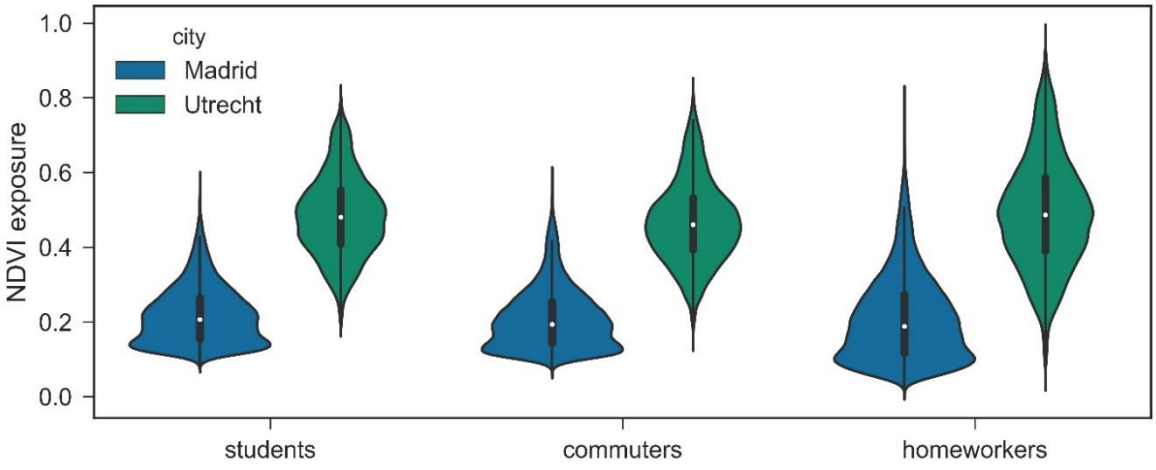


Figure 18. NDVI exposure for students, commuters and homemakers in Utrecht (green) and Madrid (blue). Each boxplot shows the distribution of NO_2. White dot represents the median. The tick black bar shows the interquartile range and the thin line shows the 1.5x interquartile range.

4.4 Sensitivity analysis

The sensitivity analysis in this study only affects the exposure of students and commuters in Utrecht and Madrid, exposure of homemakers is not changed because only the buffer weighting is changed and the size of the buffers. See Appendix E for median exposures of students, commuters and homemakers for all scenario's used in the sensitivity analysis.

4.4.1 Decay factor

The sensitivity analysis for the different weighting of the buffers shows that there is only a minor effect on the exposure for NO_2 (Figure 19) and NDVI (Figure 20) in both Utrecht and Madrid (Appendix E). Comparing scenario 1 (no weighting) and scenario 4 (strongest weighting), the E_{NO_2} changes with -2.3% for students and -1.6% for commuters in Utrecht. In Madrid the change is negligible, 0% and +0.4% for students and commuters respectively. For both cities a decrease in weighting of the buffers increases the difference between commuters and students, but the difference of these groups compared to homemakers decreases. The E_{NDVI} is less affected than the E_{NO_2} with only a 1% increase for students and commuters in Utrecht. In Madrid the E_{NDVI} changes with -1% for students and +1% for commuters.

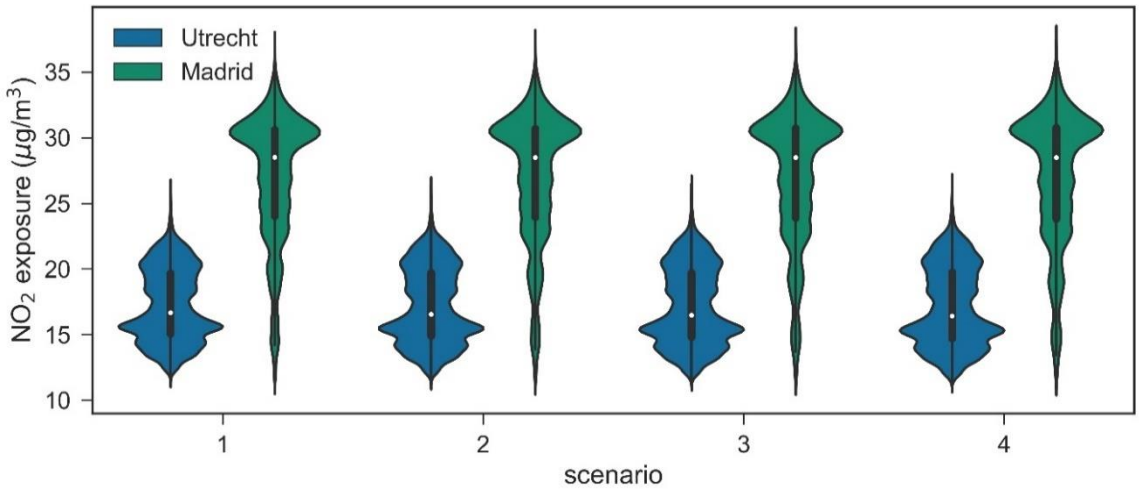


Figure 19 Sensitivity analysis of the effects of weighting of the buffers to students NO_2 exposure (µg/m³) for Utrecht (blue) and Madrid (green). Each boxplot shows the distribution of NO_2, the white dot represents the median, the tick black bar shows the interquartile range and the thin line shows the 1.5x interquartile range. The violin plots are grouped on the different scenarios (Table 1).

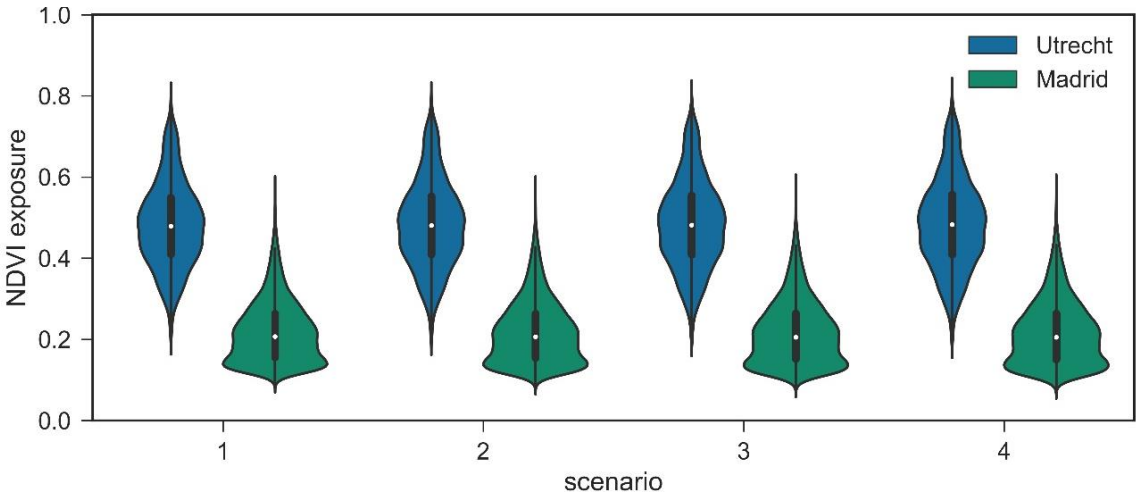


Figure 20 Sensitivity analysis of the effects of weighting of the buffers to students NDVI exposure for Utrecht (blue) and Madrid (green). Each boxplot shows the distribution of NO_2, the white dot represents the median, the tick black bar shows the interquartile range and the thin line shows the 1.5x interquartile range. The violin plots are grouped on the different scenarios (Table 1).

4.4.2 Buffer size

The sensitivity analyses for the different buffer sizes shows a difference in the sensitivity for NO_2 and NDVI (Figure 21 & Figure 22) (Appendix E). In the NO_2 exposure there is a difference in the sensitivity between both cities, in Utrecht there is no significant difference between the scenarios compared to Madrid. The median NO_2 exposure changes with -0.3% for scenario 5 and 1.7% for scenario 6 compared to the standard exposure in Utrecht. While in Madrid the median exposure changes with -18.2% and -18.5% for scenario 5 and 6 respectively. All residences in Madrid have a decrease in their NO_2 exposure ranging between 1.95 and 7.35 and 1.76 and 7.76 µg/m^3 for scenario 5 and 6 respectively, residential locations in the municipality of Madrid show a higher reduction to exposure for both scenarios, especially residential locations near major roads.

In the NDVI exposure there is no significant difference between the different scenarios for Utrecht and Madrid. The difference in spatial pattern for changing the buffer size for NDVI shows no clear trend for scenario 6. For scenario 5 there is no trend for the city agglomerate but in both municipalities the NDVI exposure shows a general decrease.

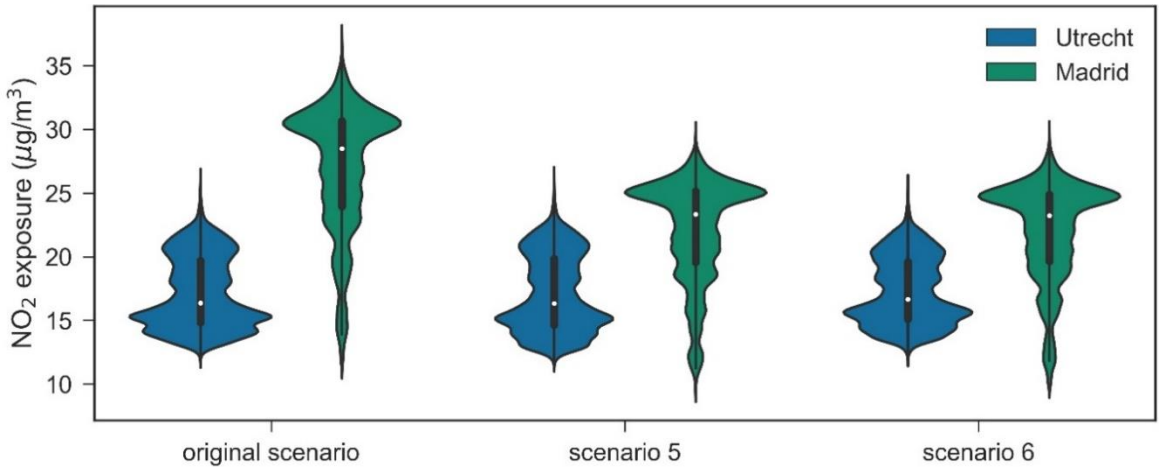


Figure 21 Sensitivity analysis of the effects of different buffers sizes of students NO_2 exposure (µg/m^3) for Utrecht (blue) and Madrid (green). Original scenario (#2) uses the standard buffer sizes, scenario 5 has a decrease of the furthest buffer to 500 metre and for scenario 6 the smallest buffer was decreased to 300 metre (Table 1).Each boxplot shows the distribution of NO_2, the white dot represents the median, the tick black bar shows the interquartile range and the thin line shows the 1.5x interquartile range.

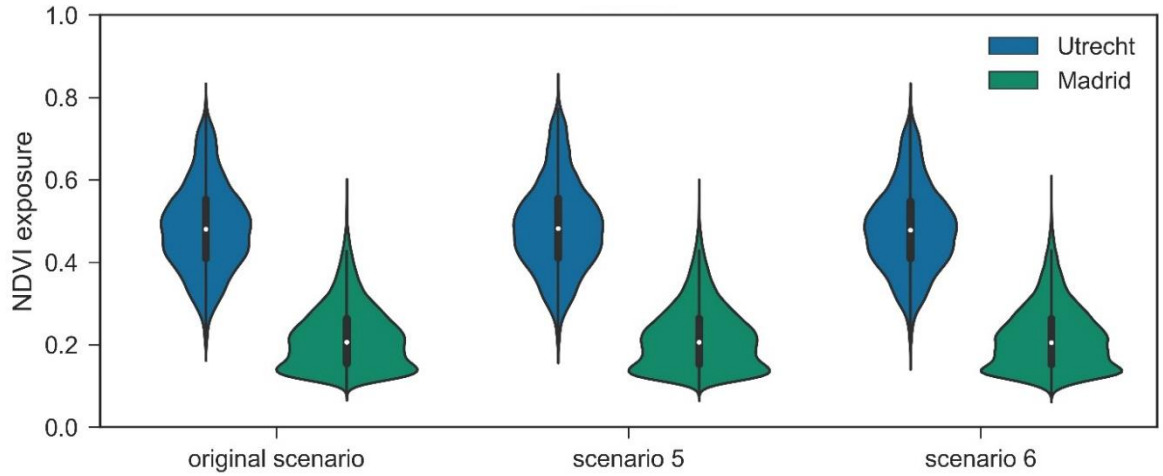


Figure 22 Sensitivity analysis of the effects of different buffers sizes of students NDVI exposure for Utrecht (blue) and Madrid (green). Original scenario (#2) uses the standard buffer sizes, scenario 5 has a decrease of the furthest buffer to 500 metre and for scenario 6 the smallest buffer was decreased to 300 metre (Table 1).Each boxplot shows the distribution of NO_2, the white dot represents the median, the tick black bar shows the interquartile range and the thin line shows the 1.5x interquartile range.

4.5 Summary exposure assessment

For all different exposure techniques the mean and median NO_2 exposure are higher in Madrid than in Utrecht (Table 2). Also the range of NO_2 concentration is larger in Madrid than Utrecht. The difference between the natural NO_2 distribution and the downscaled static exposure shows that for Utrecht the median exposure increases with 2.9 µg/m^3 while in Madrid the increase is 12.47 µg/m^3. The NDVI exposure is highest in Utrecht for all different techniques. The median NDVI exposure decreases when natural NDVI distribution is compared with the static downscaled scenario for both cities, in Utrecht the NDVI decreases with 0.36 and in Madrid with 0.32 (Table 3). Roads are the microenvironment with the highest exposure for NO_2 and NDVI in Utrecht and Madrid.

Commuters are the social-economic group with the highest NO_2 exposure and the lowest NDVI exposure in Utrecht and Madrid. In Utrecht have homemakers the lowest exposure to NO_2 and the highest exposure to NDVI. Also in Madrid have homemakers the lowest NO_2 exposure, the highest NDVI exposure is for students, however, the median exposure only changes 0.01.

Table 2 Statistical overview and ratio's for air pollution concentrations (NO_2, µg/m^3) for the static natural distribution and the downscaled exposure scenarios (Section 4.3.1). The activities road, school, work and home as well as the personal exposure for students and commuters (Section 4.3.2). Ratio is the mean exposure of a column divided by the static natural distribution mean exposure.

	Static natural	Static downscaled	Road	School	Work	Home[1]	$E_{student}$	$E_{commuter}$
				Utrecht				
Mean	21.30	23.72	24.46	17.25	17.35	16.61	17.15	17.18
1SD	4.75	4.51	3.52	2.22	2.35	3.16	2.80	2.87
Median	19.84	22.74	24.00	12.89	11.89	15.93	16.38	16.55
Min	15.24	15.41	16.85	16.74	16.91	10.79	11.89	11.43
Max	42.64	41.72	30.94	21.37	21.45	28.77	26.33	26.38
Ratio	1.00	1.11	1.15	0.81	0.81	0.78	0.81	0.81
				Madrid				
Mean	28.47	37.83	38.09	26.49	26.65	26.50	26.98	27.03
1SD	9.10	7.42	6.82	4.30	4.39	5.19	4.90	4.95
Median	27.61	40.08	14.29	13.29	11.94	28.08	28.50	28.62
Min	8.21	13.41	14.60	27.59	28.07	9.39	11.49	10.63
Max	60.62	57.57	46.19	33.01	32.75	39.63	37.17	37.33
Ratio	1.00	1.33	1.34	0.93	0.94	0.93	0.95	0.95

[1] Home exposure is equal to $E_{homemaker}$

Table 3 Statistical overview and ratio's for the static natural distribution and the downscaled exposure scenarios (Section 4.3.1). The activities road, school, work and home as well as the personal exposure for students and commuters (Section 4.3.2). Ratio is the mean exposure of a column divided by the static natural distribution mean exposure.

	Static natural	Static downscaled	Road	School	Work	Home[1]	$E_{student}$	$E_{commuter}$
				Utrecht				
Mean	0.66	0.43	0.55	0.46	0.41	0.49	0.48	0.47
1SD	0.29	0.20	0.08	0.07	0.07	0.15	0.11	0.11
Median	0.77	0.41	0.53	0.46	0.40	0.49	0.48	0.46
Min	0.00	0.00	0.38	0.30	0.27	0.05	0.18	0.15
Max	1.00	0.98	0.80	0.67	0.73	0.96	0.81	0.83
Ratio	1.00	0.65	0.83	0.69	0.62	0.74	0.73	0.70
				Madrid				
Mean	0.41	0.16	0.28	0.23	0.20	0.21	0.22	0.21
1SD	0.24	0.15	0.05	0.04	0.05	0.11	0.08	0.08
Median	0.44	0.12	0.19	0.12	0.10	0.19	0.21	0.19
Min	0.00	0.00	0.27	0.23	0.19	0.02	0.08	0.07
Max	1.00	0.91	0.50	0.53	0.46	0.81	0.59	0.60
Ratio	1.00	0.40	0.67	0.55	0.48	0.50	0.52	0.50

[1] Home exposure is equal to $E_{homemaker}$

5. Discussion

In this study methods for personal exposure assessment using high-resolution air pollution and greenness data but spare space-time activity data were proposed and evaluated for Utrecht and Madrid. Firstly, it is important to understand the effect of population data on exposure assessments; what is the accuracy of different resolution population datasets, how can population be downscaled, does this improve the accuracy and what is the impact on the static exposure assessments. Secondly, the difference between static exposure assessments and personal exposure assessments are discussed as well as the difference between social-economic groups in and between both cities. For further research the impact of different parameters are discussed like the size of the buffers and adding weights to the buffers.

5.1 Population data

In an exposure assessment study one has to select a population dataset to represent population density distribution. There are several population datasets available with different temporal and spatial resolutions and not all datasets cover the whole world. In this study the WorldPop (3 arcsecond) and JRC GHS (250 metre) population dataset were used because of their global high resolution characteristics. An advantage of the WorldPop dataset is the higher temporal and spatial resolution compared to the GHS dataset.

Comparing both datasets with the Dutch population reference dataset of the CBS showed that the mean error of the WorldPop dataset (26.22) is lower compared to the GHS dataset (29.93). Therefore the higher resolution WorldPop population dataset was used in the personal exposure assessment. The population was downscaled due to the high mean error per pixel. The papers of Liu & Long (2016) and Liu et al. (2017) use OSM street and POI data to spatialize population census data over parcels, a parcel is an area surrounded by streets. This method however does not downscale the population data to individual buildings but to individual parcels. Because of the high spatial variability of NO_2 (Hewitt, 1991) OSM building data was used to downscale population to individual buildings.

Two methods were used in the selection of buildings from OSM, the r^2 (0.39) and RMSE (44.34) of the land-use area downscaling method indicate that this method represents the reference CBS population data better than only using the residential buildings (r^2 0.38, RMSE 46.07) (Figure 7). The mean error, however, indicates that the mean difference between the CBS grids is substantially lower when only the residential buildings are used for downscaling (0.16 versus 1.52) (Figure 7). The higher accuracy could be explained by the way (residential) buildings are imported in OSM. In the Netherlands, all buildings are imported from the BAG (Basisregistratie Adressen en Gebouwen); therefore (almost) all residential buildings are known in OSM. In Spain, there is no import of the typology of buildings and therefore only using buildings tagged as residential could lead to an underestimation of the number of residential buildings in the study area. The comparison with the reference population data in Utrecht showed that the mean error decreased with 99% when downscaling was used. Therefore, the population data was downscaled for the personal exposure assessment for both cities. The downscaling method was different between the two studies. For Utrecht only the residential buildings were used in the analyses because of the lower mean error. For Madrid the land-use area selection method was used because the number of buildings in the residential method was low and using the land-use area method the number of buildings increased with 78% (Appendix A).

5.2 Exposure assessment

5.2.1 Static exposure

The effect of grid size of the static exposure is different for NO_2 and NDVI. For the NO_2 the general shape of the distribution remains the same across for the three grid size scenarios, while for the NDVI in Utrecht the downscaled scenario shows a positive skewed bell-shaped distribution compared to a negatively skewed distribution for the other scenarios (Figure 10). This bell-shaped distribution is the result of only using the locations of the residences. Most residences are not located on low NDVI values (< 0.1) or high NDVI values (> 0.9). And in the 1-kilometre and 25-metre grid scenarios, population is present on every pixel. However, this distinctive shape was only found for the NDVI in Utrecht. It could be caused by the strong negatively skewed natural NDVI distribution in Utrecht which is characterized by high NDVI values.

The main difference between the NO_2 exposure in the downscaled scenario for Utrecht and Madrid (Figure 8 & 9) is that in Utrecht the distribution is characterized by two peaks while Madrid has one distinctive peak. The two peaks in Utrecht are caused by the difference between agglomerate cities and the municipality of Utrecht, representing respectively the peak of the lower exposure and the peak with the higher exposure. For both studies areas the exposure for NO_2 was highest in the municipality. The difference between the municipality and the agglomerate area for Utrecht (26%) is comparable with the 23% difference found in Rotko et al. (2001), the difference in Madrid was only 6%. The difference in the shape of the distribution between the municipality and agglomerate areas was not found in Madrid because in Madrid the spatial extent of lower NO_2 areas is more limited than in Utrecht. Moreover, the influence of the municipality is stronger for both the number of residences and the larger part of the size compared to the total study area (Figure 4). This is also shown in the smaller difference between the distributions of the total study area and the municipality (Figure 9).

The homemakers exposure calculated in the personal exposure assessment using a 100-metre buffer for NDVI is also a static exposure. Many exposure studies use a buffer for residential exposure assessments and therefore the mean NDVI exposure calculated with the 100-metre buffer could be used for literature comparison. Klompmaker et al. (2018) found a mean NDVI value of 0.49 using a 100-metre buffer for a cross-sectional study in the Netherlands, this is similar to the mean NDVI using a 100-metre buffer in this study (0.49). Dadvand et al. (2017) calculated the median home NDVI using a 100-metre buffer for the city of Barcelona 0.16 which is comparable with the NDVI in Madrid (0.19)

The static exposure assessment shows that the mean NO_2 is higher in Madrid (28.47 $\mu g/m^3$) than in Utrecht (21.30 $\mu g/m^3$) and the mean NDVI is lower in Madrid (0.41) than in Utrecht (0.66) (Table 2 & 3). The ratio of the mean NO_2 exposure of the downscaled population and the mean NO_2 exposure for the natural exposure is larger than 1.0 for both cities (Table 2). This means that the population lives on locations which have a higher NO_2 concentration than is expected from the natural spatial distribution. The ratio is larger for Madrid (1.33) than for Utrecht (1.11) indicating that besides the higher mean NO_2 concentration in Madrid the spatial distribution of residences in the city is less favourable for NO_2 concentration. This implies that persons in Madrid are more exposed to NO_2 compared to Utrecht. For the NDVI, the ratio is smaller than 1.0 for both cities with a ratio of 0.65 for Utrecht and 0.40 for Madrid (Table 3). So in both cities the residential locations are less exposed to greenness than could be expected using only the mean NDVI. So the mean concentration of NO_2 is higher and the mean NDVI is lower in Madrid which increases the health risks compared to Utrecht. Adding the spatial distribution of residential locations in both cities, increases the health risk for the population the most in Madrid.

So when only the mean value of an environmental variable is used in exposure studies the calculated exposure will give a general idea of the exposure for the population inside the study area. This will however not indicate the real exposure; for example two cities with the same mean value of the environmental variable considered could have different health risks depending on the spatial distribution of residences.

The WHO proposed a guideline of a maximum annual mean NO_2 concentration of 40µg/m^3 (WHO, 2006). The downscaled exposure assessment shows that for 213 inhabitants in Utrecht and approximately 2,7 million inhabitants in Madrid the NO_2 concentrations exceed this guideline.

Tatem et al. (2011) found that different gridded population datasets resulted in an uncertainty of more than 10% between disease assessments. In the static exposure assessment, the median exposure varies between the gridded population datasets in the 1-kilometre grid and using the original dataset at 25-metre resolution in the range of 1 to 32% (Figure 8, 9, 10 and 11 B & C). The impact of different population datasets on the assessed exposure is insignificant when the datasets are downscaled using OSM data. The insignificant difference in the static NO_2 and NDVI exposure between the WorldPop and GHS population datasets could indicate that the location of residences is more important than the resolution of the population dataset. Therefore, the resolution of population data has less influence on the exposure assessment than the spatial distribution of residences in a city.

5.2.2 Personal exposure

The personal exposure assessment in this study seems to indicate that the residential location of a person is more important than the social-economic group the person belongs to. The range of exposures varies between 11 and 29 µg/m^3 based on the residential location. The difference between homemakers versus commuters and students varies between -2.9 and 2.4 µg/m^3 (Table 2). The larger range of exposure values between residential locations could be explained by the fact that the largest part of the day is still spent at home for students and commuters (15 h/day). In the personal exposure assessment, the median E_{NO_2} of homemakers almost corresponds to the median exposure in the static downscaled exposure assessment, except that in the static exposure assessment the indoor proportion R is excluded. Therefore the median personal exposure of homemakers is 0.7 times lower than in the downscaled static exposure assessment, this is similar to Lu et al. (2019).

Due to the difference in using the indoor proportion between the static exposure and personal exposure assessment, the exposure calculated in static exposure assessments overestimates the exposure compared to the personal assessments. The median static exposure does not represent the actual exposure for residences, because most time is spent inside. If the indoor proportion is included in the static exposure calculation of the natural NO_2 distribution the median exposure is 2.04 and 8.75 µg/m^3 lower in Utrecht and Madrid compared to the personal exposure of homemakers. This shows that the median exposure of the downscaled static exposure assessment better correlates with the personal exposure assessment than the static natural distribution assessment.

The spatial patterns for Utrecht and Madrid shows the same expected trends for both environmental variables, the NO_2 decreased from the city centre towards the agglomerate areas and the NDVI increased from the city centre towards agglomerate areas. The higher NO_2 concentration in the urban areas compared to the agglomerate areas is in line with several other studies (e.g. Gurram et al., 2015; Krämer et al., 2017; Rijnders et al., 2001; Rotko et al., 2001).

The study of Lu et al. (2019) also used social-economic groups with a distinct space-time pattern for the city of Utrecht. The study showed several similarities but also some differences were found. The trend of decreasing NO_2 while moving away from the city centre is similar. And in both studies commuters are the social-economic group with the highest exposure. There is a difference in the magnitude of NO_2 exposure with the study of Lu et al. (2019), the median NO_2 exposure was higher for commuters (21.35 $\mu g/m^3$) and homemakers (20.64 $\mu g/m^3$) compared to the results in this study (16.56 and 15.93 $\mu g/m^3$). This difference could be explained by the differences in the study area between the studies, Lu et al. (2019) only used the municipality of Utrecht and therefore it did not include the lower NO_2 concentrations in the agglomerate areas. When only the exposure for the municipality is calculated in this study the exposure for homemakers (20.06 $\mu g/m^3$) is comparable to the result of Lu et al., the exposure for commuters (23.57 $\mu g/m^3$) is however higher (Appendix E). A reason for the higher exposure could be the fact that in this study the average NO_2 concentration of all routes is calculated while in Lu et al. only the route to work is used.

Dadvand et al. (2017) is one of the few NDVI personal activity studies using space-time activity data, a NDVI exposure for students was calculated for the city of Barcelona. The median exposure for schools and commuting is 0.17 and 0.11 respectively. In this study the median exposure for school is 0.12 and commuting is 0.19 for Madrid (Table 3). The difference with the median NDVI values for Utrecht (0.30 and 0.38 respectively) are larger (Table 3). A reason for the higher commuting NDVI than school NDVI in this study could be that the median NDVI value of all roads are used, so also rural roads, while Dadvand et al. only uses the direct commuting road from home to school. The larger difference between Utrecht and Barcelona are caused by the higher NDVI values in Utrecht than Barcelona in general.

When the exposure assessment is performed only on the municipality of Utrecht the NO_2 exposure increases and the NDVI decreases (Appendix E). Which is expected from the static exposure assessment (Table 2 & 3). The exposure of students and commuters changes most with an increase in the NO_2 exposure of 44% and 42% the exposure of homemakers increases with 26%. The NDVI exposure decreases with 14%, 12% and 13% for students commuters and homemakers respectively.

The sensitivity study on the parameters of the microenvironments shows that in general changing the buffer size of the personal exposure assessment has a larger effect than changing the weighting of the buffers. However, for Utrecht, the change in buffer size is smaller than the change in the weighting of the buffers. A factor which could control this is the smaller size of the study area in Utrecht so that the size of the buffer is of less influence. This small increase in Utrecht is even stronger in the exposure assessment for only the municipality of Utrecht and in scenario 6 the exposure even decreases slightly (Appendix E).

In the decay factor sensitivity analyses, the general trend is that students and commuters show the same trend comparing the strongest decay factor with the no weighting scenario, the magnitude of this increase or decrease changes. But in the NDVI exposure assessment for Madrid, the exposure for students decreases with 1% and for commuters increases with 1%, it is however not clear how this change occurred.

The input data for the personal exposure in this study consists of a yearly average NO_2 map, so there is no temporal variability in the NO_2 concentrations in this study. Therefore NO_2 concentrations during peak hours are not used for commuting which leads to lower NO_2 concentrations during commuting (Zhang et al., 2011). Secondly, short-term exposure to high NO_2 concentrations can lead to significant health effects (WHO, 2006).

5.3 Future research

The (almost) equal exposure in the static exposure assessment shows that the population data has no influence on the static exposure. However, if population data is downscaled the mean error decreases and together with the large difference between the static downscaled exposure values and the natural static exposure values this shows that further exposure assessments should downscale their population data. The comparison with the reference data shows that a higher spatial resolution population dataset better represents the reference data, therefore it is best to use the highest spatial and most recent resolution population dataset available. But downscaling the population dataset decreased the mean error with 99%. Therefore, it is important to downscale the gridded population data with the location of buildings from OSM. The static exposure assessments showed that the location of residential buildings is more important than the distribution of population over the study area. Therefore, the spatial distribution of residential buildings inside a study area is more important than the population data used. Population data can best be downscaled to buildings that are tagged as residential in OSM. However, not all cities have enough detailed information in OSM, for these cities adding unclassified buildings inside residential land use areas increases the number of buildings for downscaling.

The median NO_2 exposure of students and commuters has a maximum difference of 0.6 $\mu g/m^3$ compared to homemakers; also the exposure is within the minimum and maximum exposure of homemakers. Therefore, when using annual mean NO_2 concentrations and representing activity with buffers, the addition of social-economic groups has no direct beneficial effect for estimating the health risk differences between social-economic groups. However, when using a higher temporal resolution of the environmental variable the peak exposure could be calculated for different social-economic groups. To better represent all citizens more social-economic groups could be added with more extensive time schedules and different transportation methods (Beckx et al., 2009). Since most people who live in suburban cities work in the main city, the size and shape of the buffers could fluctuate especially when using different transportation methods and large city agglomerates. The detail of OSM will determine how many social-economic groups can be added with unique spatial data for each group.

The method used in this study could be extended for multiple cities around the globe. An advantage of using microenvironments with buffers is that computational problems in data storage and long run time, found in Lu et al. (2019) are not a problem to apply this method for multiple cities worldwide. An important factor which needs further research is the determination of the buffer size when using cities of different sizes, especially if multiple transport methods are added. If the buffer size is not large enough there could be for example no school or work location for a residence which leads to an incomplete activity exposure. The effect of the weighting of the buffers is neglectable and could, therefore, be removed in further studies. The exposure calculated in the simpler static exposure assessment is able to predict quite accurately the exposure in the personal exposure assessment. Therefore the static exposure could be used to gain a quick overview of the exposure in many cities.

6. Conclusion

The objective of this study was to establish a method for personal exposure assessment based on high-resolution exposure variable data but sparse space-time activity data. The main data needed for a personal exposure assessment is a high-resolution population map with a level of detail that the number of residences per building is known. There are several global gridded population datasets available with different spatial and temporal resolutions. Using these population datasets without further downscaling the magnitude of exposure in a static exposure assessment can vary with 32%. Therefore, it is necessary to downscale these population datasets with OpenStreetMap data. The WorldPop population dataset which has a higher spatial resolution than the GHS population dataset has a lower mean error when comparing the data with Dutch population reference values. However, for a static exposure assessment when population data is downscaled with the OSM residential buildings the difference between population datasets with different resolutions is negligible. This indicates that the location of residential buildings is more important than the number of inhabitants predicted by the population dataset used. OSM data consist of volunteered geographic information and therefore the accuracy and detail of the data changes between locations. Therefore, extracting only the buildings tagged as residential could lead to an underestimation of the residential locations for a study area.

The personal exposure assessment in this study is based on three social-economic groups, homemakers, students and commuters. Their activities are represented by the microenvironments home, work, school and road. Zones of activity were used for each microenvironment to represent human activity patterns. The zones of activity were established with different weightings for each zone. The distance decay weighting is used to represent the idea that a location closer by is more likely to be visited than one further away.

Commuters are the social-economic group with the highest exposure to NO_2 in Utrecht and Madrid, followed by students whose exposure is slightly lower. In both study areas, homemakers have the lowest median NO_2 exposure. The exposure to NO_2 is, however, higher for homemakers near major roads. Also, the range of exposure is higher compared to students and commuters, because the exposure is not an average of the exposure at home, work and commuting microenvironments. The residential location is the most important factor in the exposure assessment as most time is spent at home for all groups. Commuters are also the group with the lowest NDVI exposure. The sensitivity analyses showed that using different weights the median exposure is only affected up to 2 %; therefore the added value of the distance decay function could be questioned. The effect of different buffer sizes is much stronger for the NO_2 exposure up to 18% change between different buffer sizes. However, for NDVI there was no significant change.

In this study, only three social-economic groups are represented and the difference between commuters and students is only their work or study location. Therefore, this method can be expanded for more social-groups or more detailed space-time activity data with different transportation methods. For future studies, the effects of different buffers sizes should be investigated.

References

Alcock, White, Wheeler, Fleming, & Depledge. (2014). Longitudinal effects on mental health of moving to greener and less green urban areas. *Environmental Science and Technology*, *48*(2), 1247–1255. https://doi.org/10.1021/es403688w

Almanza, Jerrett, Dunton, Seto, & Ann Pentz. (2012). A study of community design, greenness, and physical activity in children using satellite, GPS and accelerometer data. *Health and Place*, *18*(1), 46–54. https://doi.org/10.1016/j.healthplace.2011.09.003

Atkinson, Carey, Kent, van Staa, Anderson, & Cook. (2013). Long-Term Exposure to Outdoor Air Pollution and Incidence of Cardiovascular Diseases. *Epidemiology*, *24*(1), 44–53. https://doi.org/10.1097/EDE.0b013e318276ccb8

Beckx, Int Panis, Arentze, Janssens, Torfs, Broekx, & Wets. (2009). A dynamic activity-based population modelling approach to evaluate exposure to air pollution: Methods and application to a Dutch urban area. *Environmental Impact Assessment Review*, *29*(3), 179–185. https://doi.org/10.1016/j.eiar.2008.10.001

Beelen, Hoek, Vienneau, Eeftens, Dimakopoulou, Pedeli, … de Hoogh. (2013). Development of NO 2 and NO x land use regression models for estimating air pollution exposure in 36 study areas in Europe - The ESCAPE project. *Atmospheric Environment*, *72*, 10–23. https://doi.org/10.1016/j.atmosenv.2013.02.037

Bernstein, Alexis, Barnes, Bernstein, Bernstein, Nel, … Williams. (2004). Health effects of air pollution. *Journal of Allergy and Clinical Immunology*, *114*(5), 1116–1123. https://doi.org/10.1016/j.jaci.2004.08.030

Brunekreef, & Holgate. (2002). Air pollution and health. *Lancet*, *360*(9341), 1233–1242. https://doi.org/10.1016/S0140-6736(02)11274-8

Centraal Bureau voor de Statistiek. (2019). *Statistische gegevens per vierkant en postcode 2018-2017-2016-2015*. 1–32. Retrieved from https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/kaart-van-100-meter-bij-100-meter-met-statistieken

Crouse, Pinault, Balram, Hystad, Peters, Chen, … Villeneuve. (2017). Urban greenness and mortality in Canada's largest cities: a national cohort study. *The Lancet Planetary Health*, *1*(7), e289–e297. https://doi.org/10.1016/S2542-5196(17)30118-3

Dadvand, Nieuwenhuijsen, Esnaola, Forns, Basagaña, Alvarez-Pedrerol, … Sunyer. (2015). Green spaces and cognitive development in primary schoolchildren. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(26), 7937–7942. https://doi.org/10.1073/pnas.1503402112

Dadvand, Sunyer, Alvarez-Pedrerol, Dalmau-Bueno, Esnaola, Gascon, … Nieuwenhuijsen. (2017). Green spaces and spectacles use in schoolchildren in Barcelona. *Environmental Research*, *152*(September 2016), 256–262. https://doi.org/10.1016/j.envres.2016.10.026

de Vries, Verheij, Groenewegen, & Spreeuwenberg. (2003). Natural environments - Healthy environments? An exploratory analysis of the relationship between greenspace and health. *Environment and Planning A*, *35*(10), 1717–1731. https://doi.org/10.1068/a35111

den Hertog, Bronkhorst, Moerman, & van Wilgenbur. (2006). De Gezonde Wijk. Een onderzoek naar de relatie tussen fysieke wijkkenmerken en lichamelijke activiteit. *Medicine*, *23*(2S), 36–43.

Dias, & Tchepel. (2014). Modelling of human exposure to air pollution in the urban environment: a GPS-based approach. *Environmental Science and Pollution Research*, *21*(5), 3558–3571. https://doi.org/10.1007/s11356-013-2277-6

Dons, Int Panis, Van Poppel, Theunis, Willems, Torfs, & Wets. (2011). Impact of time-activity patterns on personal exposure to black carbon. *Atmospheric Environment*, *45*(21), 3594–3602. https://doi.org/10.1016/j.atmosenv.2011.03.064

Eeftens, Beelen, De Hoogh, Bellander, Cesaroni, Cirach, … Hoek. (2012). Development of land use regression models for PM 2.5 , PM 2.5 absorbance, PM 10 and PM coarse in 20 European study areas; Results of the ESCAPE project. *Environmental Science and Technology*. https://doi.org/10.1021/es301948k

Ekkel, & de Vries. (2017). Nearby green space and human health: Evaluating accessibility metrics. *Landscape and Urban Planning*, *157*, 214–220. https://doi.org/10.1016/j.landurbplan.2016.06.008

Environmental Systems Research Institute (ESRI). (2012). *ArcGis release 10*. Redlands, CA.

Facebook Connectivity Lab and CIESIN. (2016). High Resolution Settlement Layer (HRSL). *Source Imagery for HRSL © 2016 DigitalGlobe*. Retrieved from ciesin.columbia.edu/data/hrsl

Fong, Hart, & James. (2018). A Review of Epidemiologic Studies on Greenness and Health: Updated Literature Through 2017 Compliance with Ethical Standards Conflict of Interest HHS Public Access. *Current Environmental Health Reports*, *5*(1), 77–87. https://doi.org/10.1007/s40572-018-0179-y

Gascon, Cirach, Martínez, Dadvand, Valentín, Plasència, & Nieuwenhuijsen. (2016). Normalized difference vegetation index (NDVI) as a marker of surrounding greenness in epidemiological studies: The case of Barcelona city. *Urban Forestry and Urban Greening*, *19*, 88–94. https://doi.org/10.1016/j.ufug.2016.07.001

Gascon, Mas, Martínez, Dadvand, Forns, Plasència, & Nieuwenhuijsen. (2015). Mental health benefits of long-term exposure to residential green and blue spaces: A systematic review. *International Journal of Environmental Research and Public Health*, *12*(4), 4354–4379. https://doi.org/10.3390/ijerph120404354

Gorelick, Hancher, Dixon, Ilyushchenko, Thau, & Moore. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Grigsby-Toussaint, Chi, & Fiese. (2011). Where they live, how they play: Neighborhood greenness and outdoor physical activity among preschoolers. *International Journal of Health Geographics*, *10*, 1–10. https://doi.org/10.1186/1476-072X-10-66

Gurram, Stuart, & Pinjari. (2015). Impacts of travel activity and urbanicity on exposures to ambient oxides of nitrogen and on exposure disparities. *Air Quality, Atmosphere and Health*, *8*(1), 97–114. https://doi.org/10.1007/s11869-014-0275-6

Han, & Naeher. (2006). A review of traffic-related air pollution exposure assessment studies in the developing world. *Environment International*, *32*(1), 106–120. https://doi.org/10.1016/j.envint.2005.05.020

Hazeu, Schuiling, & Dorland. (2014). Landelijk Grondgebruiksbestand Nederland versie 7 (LGN7). *Alterra-Rapport*.

Herrera, Markevych, Berger, Genuneit, Gerlich, Nowak, … Radon. (2018). Greenness and job-related chronic stress in young adults: A prospective cohort study in Germany. *BMJ Open*, *8*(6), 1–12. https://doi.org/10.1136/bmjopen-2018-021599

Hewitt. (1991). Spatial variations in nitrogen dioxide concentrations in an urban area. *Atmospheric Environment. Part B. Urban Atmosphere*, *25*(3), 429–434. https://doi.org/10.1016/0957-1272(91)90014-6

Hoek, Beelen, de Hoogh, Vienneau, Gulliver, Fischer, & Briggs. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, *42*(33), 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057

Hoek, Krishnan, Beelen, Peters, Ostro, Brunekreef, & Kaufman. (2013). Long-term air pollution exposure and cardio-respiratory mortality: A review. *Environmental Health: A Global Access Science Source*, *12*(1). https://doi.org/10.1186/1476-069X-12-43

Hong, Tsin, van den Bosch, Brauer, & Henderson. (2019). Urban greenness extracted from pedestrian video and its relationship with surrounding air temperatures. *Urban Forestry and Urban Greening*, *38*(January), 280–285. https://doi.org/10.1016/j.ufug.2019.01.008

Hystad, Davies, Frank, Loon, Gehring, Tamburic, & Brauer. (2014). Residential Greenness and Birth Outcomes: Evaluating the Influence of Spatially Correlated Built-Environment Factors. *Environmental Health Perspectives*, *122*(10), 1095–1102. https://doi.org/10.1289/ehp.1308049

James, Banay, Hart, & Laden. (2015). Erratum to: A Review of the Health Benefits of Greenness. *Current Epidemiology Reports*, *2*(3), 218–218. https://doi.org/10.1007/s40471-015-0044-6

James, Hart, Hipp, Mitchell, Kerr, Hurvitz, … Laden. (2017). GPS-based exposure to greenness and walkability and accelerometry-based physical activity. *Cancer Epidemiology Biomarkers and Prevention*, *26*(4), 525–532. https://doi.org/10.1158/1055-9965.EPI-16-0925

Kampa, & Castanas. (2008). Human health effect of air polution-Enviro Polution-08.pdf. *Environmental Pollution*, *151*, 362–367. https://doi.org/10.1016/j.envpol.2007.06.012

Karssenberg, Schmitz, Salamon, de Jong, & Bierkens. (2010). A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. *Environmental Modelling and Software*, *25*(4), 489–502. https://doi.org/10.1016/j.envsoft.2009.10.004

Klepeis, Nelson, Ott, Robinson, Tsang, Switzer, … Engelmann. (2001). The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. *Journal of Exposure Analysis and Environmental Epidemiology*, *11*(3), 231–252. https://doi.org/10.1038/sj.jea.7500165

Klompmaker, Hoek, Bloemsma, Gehring, Strak, Wijga, … Janssen. (2018). Green space definition affects associations of green space with overweight and physical activity. *Environmental Research*, *160*(May 2017), 531–540. https://doi.org/10.1016/j.envres.2017.10.027

Kloog, Kaufman, & de Hoogh. (2018). Using open street map data in environmental exposure assessment studies: Eastern Massachusetts, Bern region, and South Israel as a case study. *International Journal of Environmental Research and Public Health*, *15*(11). https://doi.org/10.3390/ijerph15112443

Kollányi, & Prohászka. (2019). *Greenness Indicator for Spatial and Settlement Planning Based on NDVI and LAI Indicators*. *6*.

Krämer, Koch, Ranft, Ring, Kramer, Koch, … Behrendt. (2017). Traffic-Related Air Pollution Is Associated with Atopy in Children Living in Urban Areas. *Epidemiology*, *11*(1), 64–70.

Kwan. (2018). The neighborhood effect averaging problem (NEAP): An elusive confounder of the neighborhood effect. *International Journal of Environmental Research and Public Health*, *15*(9). https://doi.org/10.3390/ijerph15091841

Liu, X., & Long. (2016). Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environment and Planning B: Planning and Design*, *43*(2), 341–360. https://doi.org/10.1177/0265813515604767

Liu, Z., Xie, Tian, & Gao. (2017). GIS-based analysis of population exposure to PM2.5 air pollution—A case study of Beijing. *Journal of Environmental Sciences*, *59*, 48–53. https://doi.org/10.1016/j.jes.2017.02.013

Lloyd. (2017). High resolution global gridded data for use in population studies. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(4W2), 117–120. https://doi.org/10.5194/isprs-archives-XLII-4-W2-117-2017

Lu, Schmitz, Hoogh, & Kai. (n.d.). *Evaluation of different methods and data sources to optimise mapping of NO2 at a global scale*.

Lu, Schmitz, Vaartjes, & Karssenberg. (2019). Activity-based air pollution exposure assessment: Differences between homemakers and cycling commuters. *Health and Place*, *60*(September), 102233. https://doi.org/10.1016/j.healthplace.2019.102233

Maas, J., Verheij, De Vries, Spreeuwenberg, Schellevis, & Groenewegen. (2009). Morbidity is related to a green living environment. *Journal of Epidemiology and Community Health*, *63*(12), 967–973. https://doi.org/10.1136/jech.2008.079038

Maas, Jolanda, Verheij, Spreeuwenberg, & Groenewegen. (2008). Physical activity as a possible mechanism behind the relationship between green space and health: A multilevel analysis. *BMC Public Health*, *8*, 1–13. https://doi.org/10.1186/1471-2458-8-206

Mannucci, Harari, Martinelli, & Franchini. (2015). Effects on health of air pollution: a narrative review. *Internal and Emergency Medicine*, *10*(6), 657–662. https://doi.org/10.1007/s11739-015-1276-7

Nowak, Crane, & Stevens. (2006). Air pollution removal by urban trees and shrubs in the United States. *Urban Forestry and Urban Greening*, *4*(3–4), 115–123. https://doi.org/10.1016/j.ufug.2006.01.007

Ntarladima, Vaartjes, Grobbee, Dijst, Schmitz, Uiterwaal, … Karssenberg. (2019). Relations between air pollution and vascular development in 5-year old children: A cross-sectional study in the Netherlands. *Environmental Health: A Global Access Science Source*, *18*(1), 1–12. https://doi.org/10.1186/s12940-019-0487-1

Ord, Mitchell, & Pearce. (2013). Is level of neighbourhood green space associated with physical activity in green space? *International Journal of Behavioral Nutrition and Physical Activity*, *10*, 1–8. https://doi.org/10.1186/1479-5868-10-127

Park, & Kwan. (2017). Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. *Health and Place*, *43*(October 2016), 85–94. https://doi.org/10.1016/j.healthplace.2016.10.002

QGIS Development Team. (2019). *QGIS Geographic Information System*. Open Source Geospatial Foundation Project.

Rhew, Stoep, Kearney, Smith, & Dunbar. (2011). Validation of the Normalized Difference Vegetation Index as a Measure of Neighborhood Greenness. *Annals of Epidemiology*, *21*(12), 946–952. https://doi.org/10.1016/j.annepidem.2011.09.001

Rijnders, Janssen, van Vliet, & Brunekreef. (2001). Personal and outdoor nitrogen dioxide concentrations in relation to degree of urbanization and traffic density. *Environmental Health Perspectives*, *109*(SUPPL. 3), 411–417. https://doi.org/10.2307/3434789

Rivas, Viana, Moreno, Bouso, Pandolfi, Alvarez-Pedrerol, … Querol. (2015). Outdoor infiltration and indoor contribution of UFP and BC, OC, secondary inorganic ions and metals in PM2.5 in schools. *Atmospheric Environment*, *106*, 129–138. https://doi.org/10.1016/j.atmosenv.2015.01.055

Rotko, Kousa, Alm, & Jantunen. (2001). Exposures to nitrogen dioxide in EXPOLIS-helsinki: Microenvironment, behavioral and sociodemographic factors. *Journal of Exposure Analysis and Environmental Epidemiology*, *11*(3), 216–223. https://doi.org/10.1038/sj.jea.7500162

Ryan, & Lemasters. (2007). A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhalation Toxicology*, *19*(SUPPL. 1), 127–133. https://doi.org/10.1080/08958370701495998

Schiavina, Freire, & MacManus. (2019). GHS population grid multitemporal (1975-1990-2000-2015), R2019A. *European Commission, Joint Research Centre (JRC)*. https://doi.org/10.2905/0C6B9751-A71F-4062-830B-43C9F432370F

Schmitz, Beelen, Strak, Hoek, Soenario, Brunekreef, … Karssenberg. (2019). Data Descriptor : High resolution annual average air pollution concentration maps for the Netherlands. *Nature Publishing Group*, *6*, 1–12. https://doi.org/10.1038/sdata.2019.35

Steinle, Reis, & Sabel. (2013). Quantifying human exposure to air pollution-Moving from static monitoring to spatio-temporally resolved personal exposure assessment. *Science of the Total Environment*, *443*, 184–193. https://doi.org/10.1016/j.scitotenv.2012.10.098

Tatem, Campiz, Gething, Snow, & Linard. (2011). The effects of spatial population dataset choice on estimates of population at risk of disease. *Population Health Metrics*, *9*, 1–14. https://doi.org/10.1186/1478-7954-9-4

Vienneau, de Hoogh, Faeh, Kaufmann, Wunderli, & Röösli. (2017). More than clean air and tranquillity: Residential green is independently associated with decreasing mortality. *Environment International*, *108*(July), 176–184. https://doi.org/10.1016/j.envint.2017.08.012

Watson, Bates, & Kennedy. (1998). Assessment of human exposure to air pollution: methods, measurements, and models. In *Air pollution, the automobile, and public health*. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/33eba485-e1e3-4748-9358-0d66ef86bcc3

Weier, & Herring. (2000). Measuring Vegetation (NDVI & EVI). Retrieved from https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php

Weimann, Rylander, Albin, Skärbäck, Grahn, Östergren, & Björk. (2015). Effects of changing exposure to neighbourhood greenness on general and mental health: A longitudinal study. *Health and Place*, *33*, 48–56. https://doi.org/10.1016/j.healthplace.2015.02.003

World Health Organization. Regional Office for Europe. (2010). *WHO guidelines for indoor air quality: selected pollutants*. Retrieved from https://apps.who.int/iris/handle/10665/260127

World Health Organization. (2006). *WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide : global update 2005 : summary of risk assessment*. Retrieved from https://apps.who.int/iris/handle/10665/69477

World Health Organization. (2016). Ambient Air Pollution: A global assessment of exposure and burden of disease. *World Health Organization*. Retrieved from https://apps.who.int/iris/handle/10665/250141

Yang, L., Hoffmann, Scheffran, Rühe, Fischereit, & Gasser. (2018). An Agent-Based Modeling Framework for Simulating Human Exposure to Environmental Stresses in Urban Areas. *Urban Science*, *2*(2), 36. https://doi.org/10.3390/urbansci2020036

Yang, W., Lee, & Chung. (2004). Characterization of indoor air quality using multiple measurements of nitrogen dioxide. *Indoor Air*, *14*(2), 105–111. https://doi.org/10.1046/j.1600-0668.2003.00216.x

Zhang, K., Batterman, & Dion. (2011). Vehicle emissions in congestion: Comparison of work zone, rush hour and free-flow conditions. *Atmospheric Environment*, *45*(11), 1929–1939. https://doi.org/10.1016/j.atmosenv.2011.01.030

Zhang, L., Kwan, Chen, Lin, & Zhou. (2018). Impacts of individual daily greenspace exposure on health based on individual activity space and structural equation modeling. *International Journal of Environmental Research and Public Health*, *15*(10). https://doi.org/10.3390/ijerph15102323

Appendices

Appendix A OSM selection

Table 4 Overview of the number of OSM features for Utrecht and Madrid. The code column shows the SQL code used to only select tags classified as residential, educative or work.

OSM feature	QGIS selection expression	Number of buildings	
		Utrecht	Madrid
Residential Utrecht	"type" in ('apartments' , 'farm' , 'house' , 'houseboat' , 'residence' , 'residential' , 'retail;apartments' , 'school;apartments' , 'shop;appartments' , 'terrace')	181,352	-
Residential Madrid	"type" IN ('aparments', 'apartament', 'apartaments', 'apartments', 'apartments:level5', 'appartment', 'bungalow', 'cabin','detached', 'dormitory' ,'family_house', 'farm', 'farmhouse', 'flats', 'house', 'house;barn', 'house;carport', 'house;yes', 'houseboat', 'semidetached_house' ,'terrace', 'yes;apartments' ,'residential')	-	62,315
Residential + Land use >20m^2		208,931	110,668
Educational	"type" IN ('college' , 'school' , 'university')	186	1148
Work	"type" in ('commercial' , 'industrial' , 'kiosk' , 'office' , 'retail' , 'supermarket' , 'warehouse')	6349	8255
Roads	All roads	184,598	596,506

Appendix B Reference population versus gridded population datasets

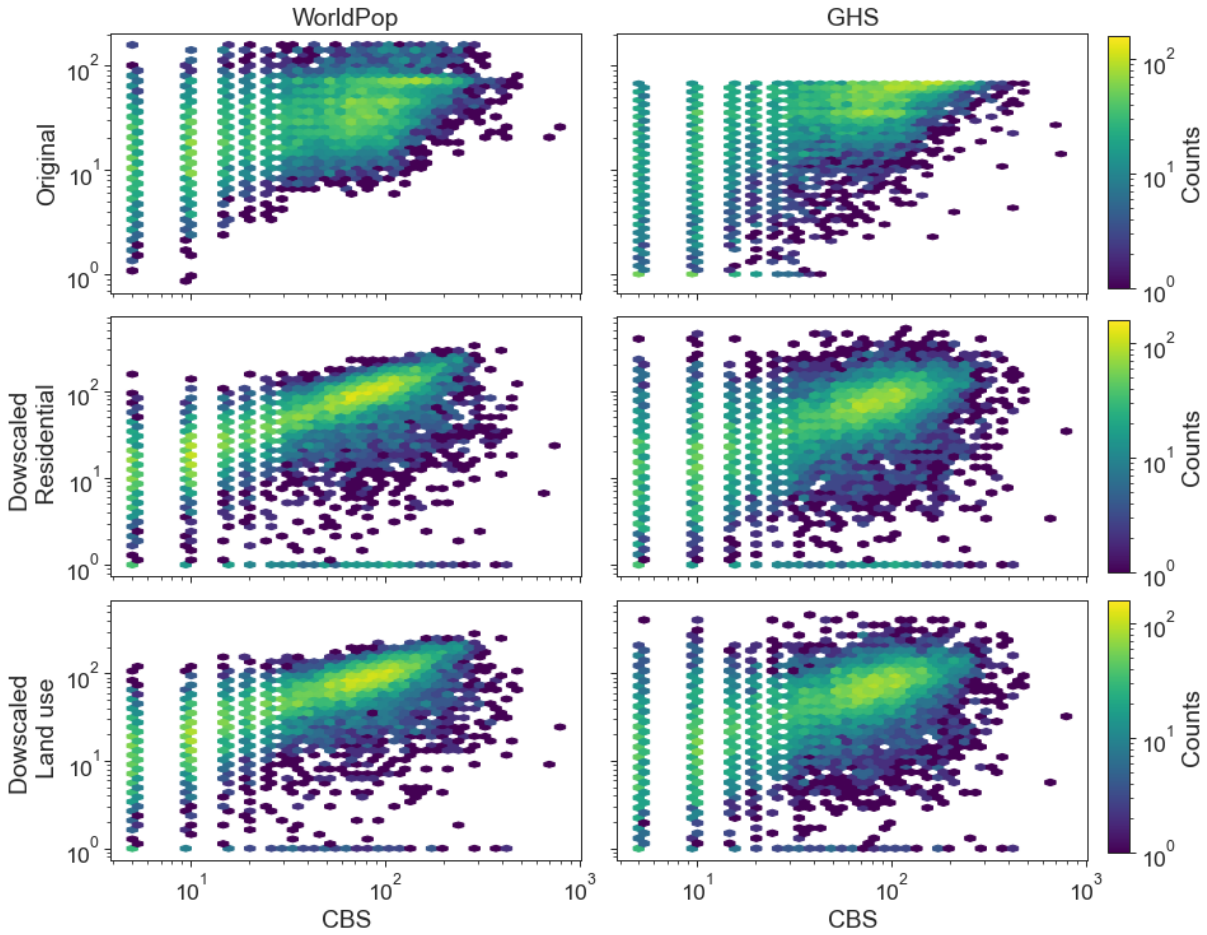


Figure 23 Logarithmic hexbin plot of the reference population data (CBS) values on the x-axis versus the population data of the WorldPop (left) and GHS dataset (right) for Utrecht. The upper row (original) shows the original population datasets. The centre row (residential) shows the downscaled population datasets using buildings tagged as residential. The lower row (land use) shows the population datasets downscaled using the land use method, so the unclassified buildings inside land use areas classified as residential are added. The colorbar shows how many points are located inside each hexagon, each point has a resolution of 20 persons, all 10283 samples are used.

Appendix C Personal exposure maps

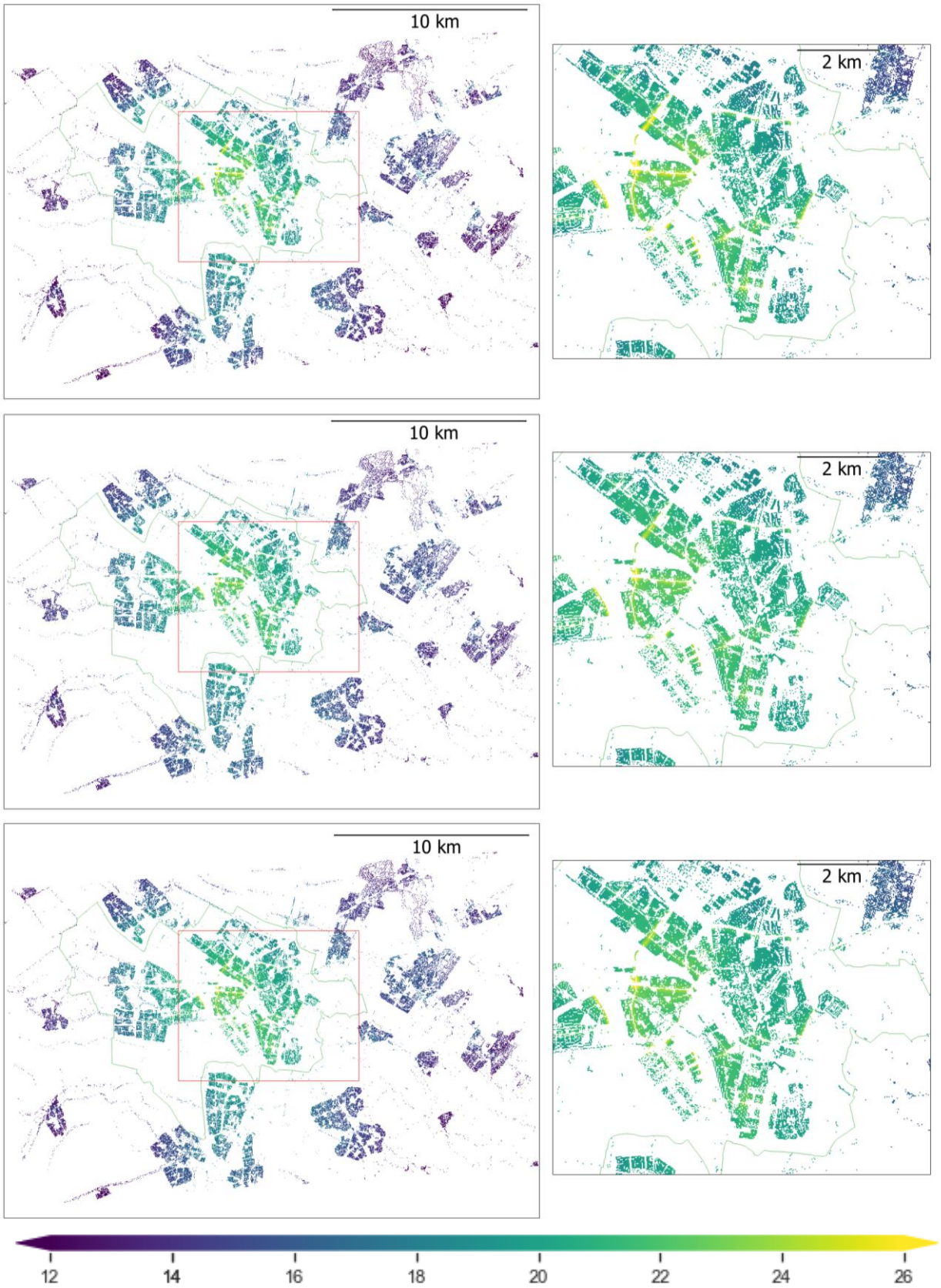


Figure 24 The NO_2 exposure (µg/m³) for Utrecht for homemakers (upper), students (centre) and commuters (bottom). The left column shows the exposure for each location based on the downscaled building map. The municipality is outlined in green, the red square shows the area of the zoomed area in the right column. The colorbar is applicable to all panels, minimum exposure is 10.79 µg/m³, maximum is 28.77 µg/m³.

Figure 25 The NO_2 exposure (µg/m³) for Madrid for homemakers (upper), students (centre) and commuters (bottom). The left column shows the exposure for each location based on the downscaled building map. The municipality is outlined in green, the red square shows the area of the zoomed area in the right column. The colorbar is applicable to all panels, minimum exposure is 9.39 µg/m³, maximum is 39.63 µg/m³.

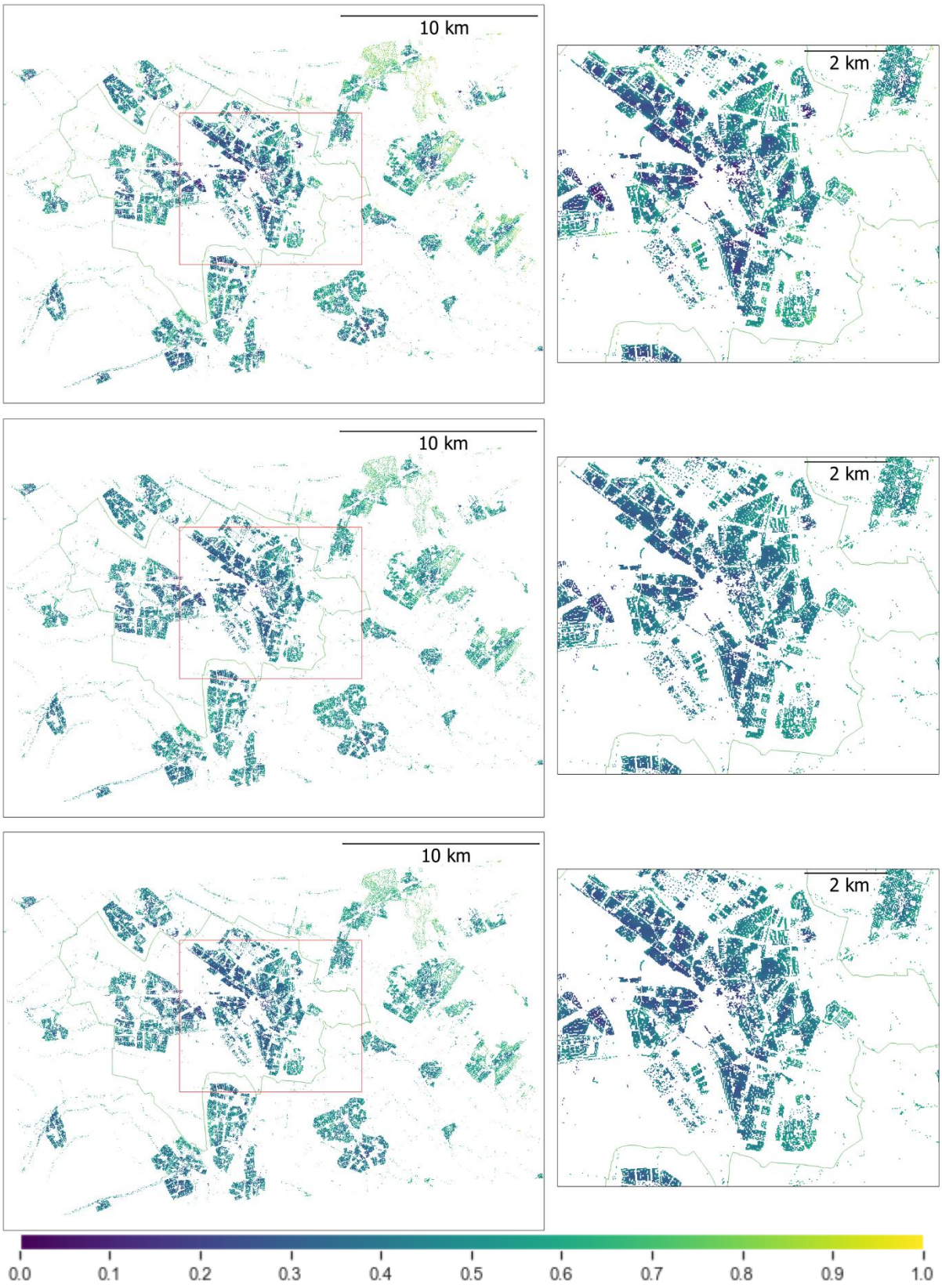


Figure 26 The NDVI exposure for Utrecht for homemakers (upper), students (centre) and commuters (bottom). The left column shows the exposure for each location based on the downscaled building map. The municipality is outlined in green, the red square shows the area of the zoomed area in the right column. The colorbar is applicable to all panels.

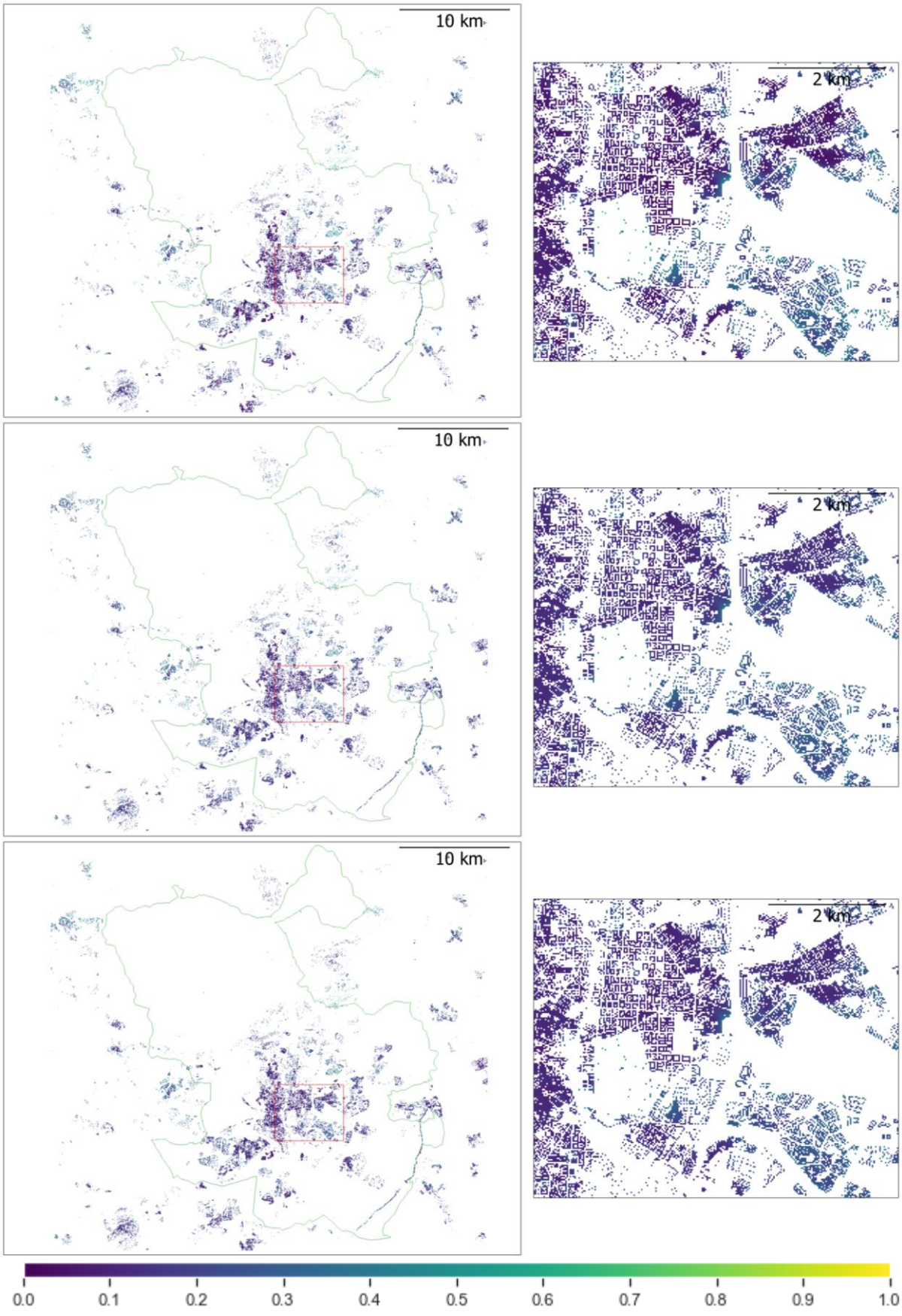


Figure 27 The NDVI exposure for Madrid for homemakers (upper), students (centre) and commuters (bottom). The left column shows the exposure for each location based on the downscaled building map. The municipality is outlined in green, the red square shows the area of the zoomed area in the right column. The colorbar is applicable to all panels.

Appendix D Difference maps personal exposure

NO_2 Utrecht

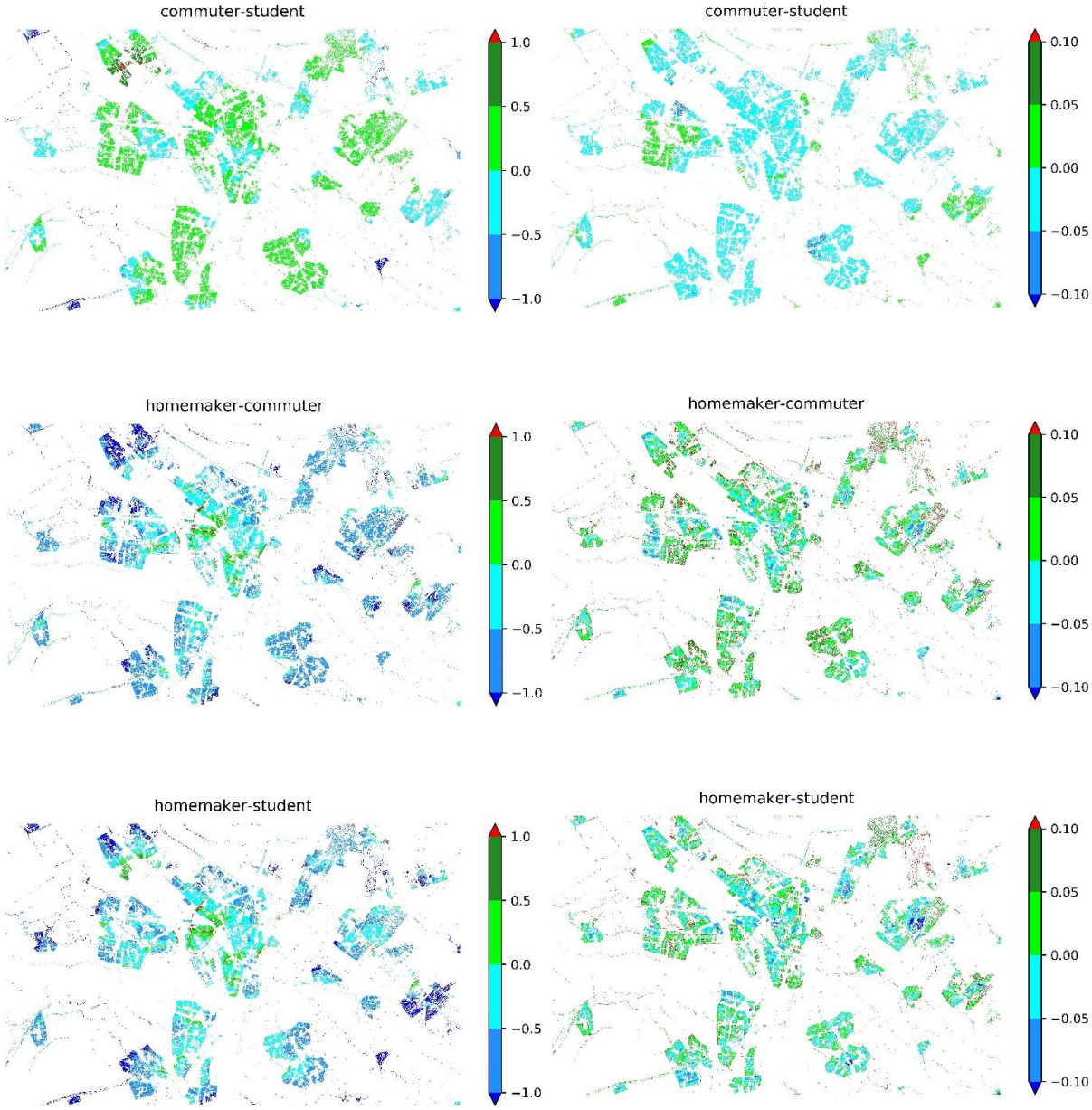


Figure 29 Difference in NO_2 exposure (µg/m³) between the social-economic groups for Utrecht.

NDVI Utrecht

Figure 28 Difference in NDVI exposure between the social-economic groups for Utrecht.

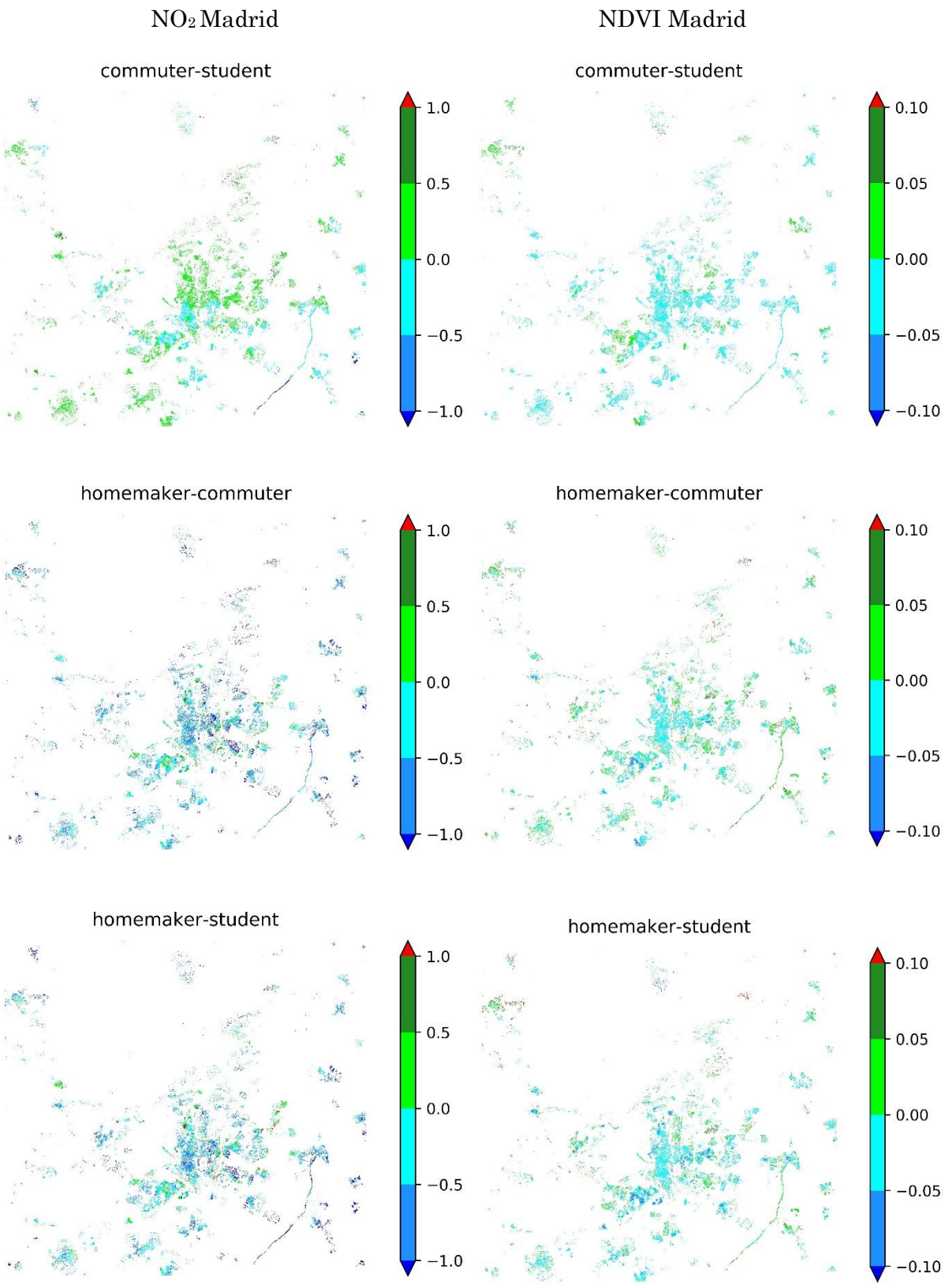


Figure 30 Difference in NO_2 exposure (µg/m³) between the social-economic groups for Madrid.

Figure 31 Difference in NDVI exposure between the social-economic groups for Madrid.

Appendix E Personal exposure

Table 5 Median personal exposure for NO_2 and NDVI. Scenario 2 is the standard scenario used in the results section. The exposure for homemakers is the same for each scenario. See Table 1 for the definition of all scenarios.

		Utrecht		Madrid	
Social-economic group	Scenario #	E_{NO_2} (µg/m^3)	E_{NDVI}	E_{NO_2} (µg/m^3)	E_{NDVI}
Student	1	16.5519	0.4789	28.5014	0.2067
	2	16.3797	0.4806	28.4968	0.2062
	3	16.2870	0.4819	28.4992	0.2055
	4	16.1672	0.4831	28.4938	0.2050
	5	16.3364	0.4817	23.3108	0.2060
	6	16.6567	0.4779	23.2156	0.2050
Commuter	1	16.6748	0.4593	28.5712	0.1932
	2	16.5552	0.4606	28.6176	0.1940
	3	16.4837	0.4618	28.6454	0.1946
	4	16.4138	0.4626	28.6796	0.1952
	5	16.5493	0.4607	23.4109	0.1950
	6	16.6802	0.4622	23.2842	0.1923
Homemaker		15.9260	0.4868	28.0801	0.1884

Table 6 Median personal exposure for NO_2 and NDVI for the total study area in Utrecht and for only the municipality of Utrecht. See Table 1 for the definition of all scenarios.

		Utrecht total study area		Utrecht municipality	
Social-economic group	Scenario #	E_{NO_2} (µg/m^3)	E_{NDVI}	E_{NO_2} (µg/m^3)	E_{NDVI}
Student	1	16.55	0.4789	23.53	0.4126
	2	16.37	0.4806	23.57	0.4124
	3	16.2	0.4819	23.58	0.4120
	4	16.16	0.4831	23.61	0.4114
	5	16.33	0.4817	23.63	0.4113
	6	16.65	0.4779	23.51	0.4122
Commuter	1	16.67	0.4593	23.52	0.4047
	2	16.55	0.4606	23.53	0.4058
	3	16.48	0.4618	23.53	0.4080
	4	16.41	0.4626	23.54	0.4088
	5	16.54	0.4607	23.58	0.4049
	6	16.68	0.4622	23.49	0.4086
Homemaker		15.92	0.4868	20.06	0.4246