Tracking Vegetation Change in Urban Gardens using NDVI data: Socioeconomic and Built Environment Characteristics

The determinants of garden greening

Master Thesis

Vrije Universiteit Amsterdam

School of Business and Economics

Master of Science in Spatial, Transport and Environmental Economics

Aron Uidam

Student number: 2740084

Supervisor: E. Koomen

29 June 2025

Abstract

Urban greening is the addition of urban green space (UGS) in cities. Private residential gardens can play a crucial role in increasing the amount (UGS) in cities, which mitigates climate change impacts and improves public health. This thesis aims to uncover the drivers of garden greening in the city of Breda in the Netherlands. This is done by analyzing changes in garden-level vegetation densities in the years 2016, 2017, 2018 and 2020, with the help of high-resolution NDVI (Normalized Difference Vegetation Index) data. Over 45,000 gardens were analyzed and multiple methods of greening detection were developed, to uncover underlying mechanics in garden greening.

In order to explain variation in garden greening, neighborhood-level socioeconomic indicators were combined with garden-specific characteristics in multiple econometric models. This thesis implemented OLS regression, logistic regression and spatial models, to account for spatial autocorrelation. Spatial models were used, as the Lagrange Multiplier (LM) tests revealed the presence of significant spatial lag and error dependence. A spatial SAR, SEM and SARAR model were therefore implemented to correct for the spatial autocorrelation.

The results show that socioeconomic factors, such as average income and education level, do impact the amount of greening in gardens, though the influence is generally quite small but significant. Some neighborhood-level data also showed conflicting results between models, where for instance the percentage of males decreased the likelihood of small NDVI increases, while larger increases were more common. This highlights that different definitions of greening can provide insights into potential motivations for garden greening. Finally, this thesis identifies points of improvements, which municipalities can for example inform people of creative solutions to green gardens.

Table of contents

1. Introduction

The world population is increasing rapidly. In 2022, the milestone of 8 billion inhabitants was reached, and this number is estimated to grow to 9 billion by 2037 (UN, 2022). The Dutch population has grown by over 100.000 people per year in the past decade, largely due to migration (CBS, n.d.-a). This increased migration pressure requires the construction of many new houses, which is why the Netherlands is aiming to build 981.000 between 2021 and 2030. Part of these houses has to be built in or around existing cities (VRO, 2023). This often comes with reductions in other types of land use (PBL, 2023). Urbanization and the densification of cities is likely to decrease the amount of urban green space, particularly in poorer neighborhoods (de Vries et al., 2020).

1.1 Urban Green Space

These reductions in urban green space result in a call for more urban green space via urban 'greening'. Urban greening can be summarized as the planning, management and implementation of UGS (Fruth et al., 2019), where vegetation is added in urban areas. UGS is currently unequally distributed between and across cities with regard to quantity and quality (de Vries et al., 2020). This is because neighborhoods with limited UGS in streets, generally also offer less private garden area (de Vries et al., 2023). However, according to Heidt & Neef (2008), UGS provides ecological, economical and social services. Ecological services can entail habitat for birds and higher overall biodiversity (Mason, 2000). Economical services entail higher property values and reduced energy costs for cooling due to lower ambient temperatures (Sander et al., 2010; Saphores & Li, 2012). These lower temperatures are reached because of evapotranspiration, where water is evaporated which reduces the air temperature, and because trees shade buildings from the sun (Heidt & Neef, 2008). Social services can be anything from relaxation to a more beautiful aesthetic, shade, stress reduction, food production and environmental stewardship (Langemeyer et al., 2018). Social services can also entail significant health benefits, where people with more green in their garden can enjoy up to 20% reductions in the prevalence of certain diseases. De Vries et al. (2025) showed that the prevalence of 16 out of 21 studied diseases significantly decreased among people with large green gardens.

1.2 Shrinking urban green space

Decreases in UGS area is the result of two forces that work in the opposite direction: Urban sprawl must be limited to reduce the total area required for human activity, and preserve green space outside the urban area. However increasing urbanization therefore demands increased density within the urban core, which in turn reduces the amount of UGS (Balikçi et al., 2022). Balikçi et al. (2022) for example found that the amount of UGS decreased in both Amsterdam and Brussels. Even though both cities aim to protect existing UGS, this does not always prevent UGS from making way for housing.

One way to nevertheless increase the amount of UGS in dense cities, is to encourage and incentivize people to green their private garden around their house. In this context, a garden is a privately owned outdoor space directly adjacent to a home, typically used for recreation, planting or decoration. Currently, 60% of gardens in the Netherlands are paved (Mulder, 2020). There is therefore ample room to increase the amount of UGS in private gardens (van Heezik et al., 2012). One such initiative to increase UGS is the NK tegelwippen (Dutch program of tile lifting), which has been introduced nationally and was first held in 2021 (NK tegelwippen, 2024). This program encourages people to remove tiles from their gardens and spread awareness about the benefits of UGS (Pisman et al., 2021). This contest turns urban greening into a collective effort by counting the amount of tiles removed. On top of this, the NK tegelwippen raises awareness about the numerous values urban green can supply, which fosters environmental stewardship (Langemeyer et al., 2018).

1.3 Health benefits of green

There is a particularly large literature regarding the correlation between UGS and public health, similar to de Vries et al. (2025). This literature explores the correlation between residential greenness and maternal health and pregnancy outcomes. Banay et al. (2017) found that there are correlations between more greenness and higher birthweight and lower odds of antenatal depression. James et al. (2015) investigated how neighborhood greenness affects various health outcomes. They found that there is a positive association between green areas and physical activity, especially under children. On top of this, there were fewer obese or overweight people, mental health was improved, the chances of cardiovascular disease was smaller and lower mortality was recorded in greener areas. Literature does not agree on all these topics however, as some find contradictory results. James et al. (2015) do highlight however that these outcomes are also partly confounded. More physical activity for example improves birth outcomes, reduces mortality and cardiovascular diseases and improves mental health. Therefore it is difficult to infer from this that UGS directly causes all these benefits, but a clear association is observed, indicating that people are generally healthier in greener environments (James et al., 2015).

1.4 Study area

Breda is the study area of this thesis and is a city that has participated in the NK tegelwippen since the first edition (year). Breda is a city located in the province of North-Brabant in the south of the Netherlands. The city is home to 188.834 inhabitants, and this amount continues to increase (CBS, n.d.-b). Breda is the 8th largest municipality in the Netherlands regarding population size. With a population density of 1496 people per km^2, Breda is fairly densely populated, but there are more dense cities in the Netherlands (CBS, n.d.-b). The municipality of Breda is actively striving to become a more sustainable and green city (Gemeente Breda, n.d.). The municipality focuses on reducing carbon emissions, reducing energy consumption, better management of waste, but also increasing the amount of rainwater infiltration into the soil, the amount of green areas in the city, the amount of green on roofs and building façade and the amount of green in gardens (Gemeente Breda, n.d.). For this last point, Breda has been participating in the 'NK tegelwippen' since 2021 (Knook, 2024). Since 2020, the municipality has also provided subsidies for tile removal and greening of gardens (Marcelis, 2021). The aim of the city is to become what they call 'a city in a park' (Gemeente Breda, n.d.).

Breda has been fairly successful in the NK tegelwippen competition and has finished in the top 5 municipalities regarding how many tiles were removed per 1,000 residents during the last three years. In total, Breda has now officially removed 863,191 tiles (Knook, 2024). This equates to roughly 78,500 m^2 of area since 1 m^2 is covered by 11 tiles (NK-tegelwippen, n.d.). Furthermore, the municipality of Breda has provided subsidies to people who replace tiles with vegetation in their gardens since 2017 (Gemeente Breda, n.d.). Figure 1 shows the different neighborhoods in Breda.

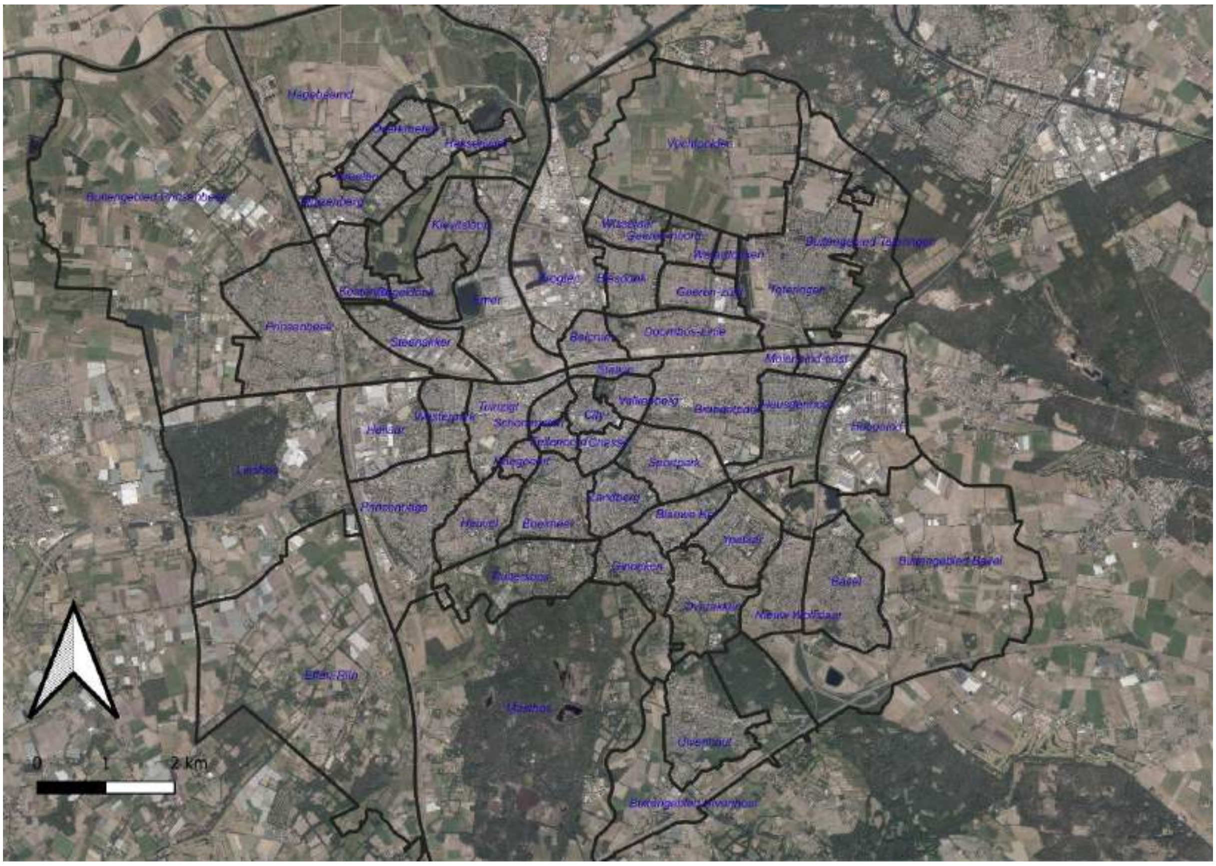

Figure 1. Study area

1.5 NDVI (Normalized Difference Vegetation Index)

An often employed measure to measure vegetation changes is the NDVI score. This vegetation index has previously been used as an indicator changing vegetation cover of natural and agricultural land (de Jong et al., 2011)(Bai et al., 2008)(Gandhi et al., 2015)(Yeager et al., 2023)(de Vries et al., 2020). This research paper employs the use of the NDVI-score to assess how much green was present in gardens during various years. Previous studies with NDVI data often used low spatial resolutions of for example 8 km (de Jong et al., 2011) (Bai et al., 2008), but more recent satellite data can have resolutions of 250 to 1,000 meter (Pettorelli et al., 2005). De Jong et al. (2011) for example used this low resolution to calculate worldwide NDVI anomalies. Other studies use NDVI data to detect vegetation densities or changes on a smaller scale, such as per city (Ju et al., 2021), neighborhood (de Vries et al., 2020) or garden (Yeager et al., 2023). Whereas earlier studies primarily used satellite data, these newer studies may use drone/UAV imagery or aerial photographs from planes (Torres-Sánchez et al., 2014; Yeager et al., 2023).

This study exploits high resolution NDVI data from aerial photographs. This way, it is possible to detect changes in vegetation cover within gardens. These changes in NDVI values in gardens will be used to determine the amount of greening or 'browning' in gardens, where vegetation is either added or removed.

1.6 Research question

The research question of this thesis is:

To what extent have gardens in Breda become greener over time, and what neighborhood- and garden characteristics can explain variations in garden greening?

This question will be answered with the help of the following sub-questions:

1. What has the average change in NDVI values been across all gardens.
2. How can differences in NDVI value changes and greening be explained by demographic-, socioeconomic- and urban characteristics?
3. Is there spatial autocorrelation in garden greening, and how does it affect the results?

Data from the NK tegelwippen shows that many tiles have been removed (NK-tegelwippen, n.d.). Therefore, an increase in the amount of UGS in gardens in Breda could be measurable and observable on aerial photographs. Nevertheless, even though according to the data from ‘NK tegelwippen’ many tiles have been removed in Breda, it is unclear how many tiles are concurrently added into the garden as well. It has not yet been investigated to what extent gardens in Breda have become more green, therefore it is as of yet unclear whether or not gardens have actually become greener on average.

2. Literature Review

2.1 Greenery and demographics

The following section will provide a brief introduction into the existing literature, with regards to greenery in gardens and in surrounding areas. First of all, it will be investigated how demographics can influence the amount of greenery in a city, neighborhood or in a private garden.

A study in Santiago, Chile revealed that tree cover density was associated with socioeconomic variables at the neighborhood scale. They showed that areas with high socioeconomic status had higher tree cover compared to areas with medium to low socioeconomic status. They argued that this could be a result of worse pruning, irrigation, pest control or fertilizer application in areas with lower socioeconomic status (Guevara et al., 2024). The city of Santiago is however much different compared to the Netherlands, also in regard to climate. The city averages 275 mm of rainfall per year and has short rainy periods (Climates to travel, n.d.). This is much different compared to the Netherlands, where annual rainfall averages 851 mm per year (Bessembinder et al., 2023). It is therefore unclear if similar patterns are observable in Breda, as irrigation might not be as much of a limiting factor in greening a garden in the Netherlands.

A meta-analysis in the United States by Gerrish & Watkins (2018) also found that cities with higher income inequality also showed more inequality with regards to urban tree cover. Richer neighborhoods generally had higher urban tree cover compared to low income neighborhoods. More importantly, this study also found that studies that did not control for spatial autocorrelation, showed much stronger evidence for inequality in urban tree cover. The fact that results are different when not controlling for spatial autocorrelation, implies that these models are biased and inconsistent. According to Gerrish & Watkins (2018), studies can control for spatial autocorrelation with spatial error or spatial lag models in order to attain unbiased and consistent estimations. Another study in the United States by Casey et al. (2017) studied how differences in demographics affected changes in NDVI values across 59,483 census tracts. They found that neighborhoods with a higher percentage of Asian, Black, and Hispanic inhabitants had lower NDVI values in the base year compared to neighborhoods which consisted primarily of White people. Furthermore, they used the Index of Concentration at the Extremes (ICE), which measures the concentrations of poverty and affluence in neighborhoods (Casey et al., 2017). They found that poorer neighborhoods, with lower initial NDVI values, showed more greening compared to other neighborhoods. Nevertheless, these neighborhoods remained less green compared to more affluent neighborhoods. They found that some minority neighborhoods saw slight increases in NDVI over time, but this was inconsistent. Overall, they found that racial minorities and low-income individuals live in less green neighborhoods, and these conditions have not improved (Casey et al., 2017). The study also corrects for spatial autocorrelation, similar to what Gerrish & Watkins (2018) recommended.

Murtagh & Frost (2023) conducted a survey among 1,000 working adults in the UK. They focus on front gardens and explored what motivations people have for gardening and linked this to the time spent gardening and the green coverage in the garden. They found that intrinsic motivations, such as pleasure or creating room for wildlife, were the strongest drivers behind front gardening, while aesthetic was also important. Murtagh & Frost (2023) also differentiated between genders, and found that intrinsic motivations were important to women, but not for men. Men with higher incomes also spent less time gardening. For men, the aesthetic was more important, while education and the intrinsic values were not significant. Furthermore, they found that larger front gardens were more green. The importance of intrinsic and aesthetic motivations however indicates that not all variation can be captured with sociodemographic variables. This study won't be able to capture these motivations in the sociodemographic variables.

2.2 Lack of greenery in gardens

Stobbelaar et al. (2021) investigated what potential limitations are as to why people do not invest in a green garden, despite the benefits that UGS provides. They used a qualitative research approach with in-depth interviews with garden owners in the Netherlands. The interviews were semi-structured, meaning that people were able to give open answers as to what might prevent them from greening their garden. They explored people's attitudes towards gardening and what potential barriers are towards greening gardens. This was done by asking about their current gardening practices, their view on climate adaptation and motivations of greening gardens. This provides deeper insights into the motivations and influences that couldn't be captured in revealed preference data, such as socioeconomic neighborhood data.

Stobbelaar et al., (2021) found that there is a wide variety of reasons as to why people may not green their garden. The first is a lack of awareness of the ecological services a garden might provide and a lack of awareness about how to effectively green your garden. Secondly, some people simply do not feel the urgency to green their garden. Third, people may prefer a low-maintenance garden that isn't dirty or time-consuming to maintain. Fourth, there could be practical limitations to people not being able to maintain a garden. This could be anything from being physically incapable or not having the skills, money or time to maintain a garden. Finally, social norms also strongly influence people's effort to green a garden. It has been shown that the gardens of neighbors dictate to what extent people green their own gardens (Hunter & Brown, 2012). Neighbors with and aesthetically pleasing gardens are clustered, therefore adjacent neighbors also take better care of their garden (Hunter & Brown, 2012). The reason to not green a garden is therefore often due to a combination of personal preferences, practical limitations or a lack of awareness about the benefits a green garden provides.

In order to combat some of these misconceptions, the people in Breda are encouraged to green their garden. Websites such as 'bredastadineenpark.com', 'bredanationalparkcity.nl' and the website of the municipality of Breda (Gemeente Breda, n.d.), provide information about the ambitions of the city to become a City in a Park. These websites aim to educate people about the benefits of greenery in gardens, with the aim to increase the amount of greenery in gardens.

3. Method

3.1 Data

3.1.1 Data sources

NDVI data for the Netherlands, including Breda, is gathered from the aerial images from PDOK (Publieke Dienstverlening Op Kaart)(PDOK, n.d.-b). These images were pre-processed by the department of Spatial Economics at the VU Amsterdam. This final dataset contains NDVI values for 44,697 gardens in Breda. The aerial photographs are supplied by PDOK and have a spatial resolution of 0.25 m^2, indicating that each pixel captures an area of 25 by 25 cm on the surface of the earth. The aerial photographs consist of both a RGB color layer, as well as an infrared layer (PDOK, n.d.-b). The core statistics for neighborhoods and districts in the Netherlands is sourced from the Dutch bureau for statistics, which will be referred to as CBS (CBS, 2025). Income data on neighborhood level is also available at CBS (CBS, 2023). The aerial pictures from (PDOK, n.d.-b) are made during the summer, therefore plants have developed leaves and vegetation should be easily identifiable on aerial photographs. This study specifically uses the aerial pictures from 2016, 2017, 2018 and 2020, as will be elaborated on in section 3.3.

3.1.2 Data preparation

Since the initial dataset covered all property areas, the dataset first had to be selected to only include gardens. This was done by overlaying the property areas with the ‘CBS bodemgebruik’ map, which contains the land-use for the Netherlands in 2017, and is available at PDOK (PDOK, 2022). From this dataset, the layer with residential land-use was used, to only include garden areas in residential areas, as the initial dataset also contained observations which were not located in a garden. Because this layer contains data from 2017, some newly built properties were not included in the final analysis. This is an advantage for this analyses, as this way properties that were not yet present in 2017 are excluded from the analysis. This facilitates a more fair comparison over time, especially since building plots were often also labeled as gardens before construction had begun. There was however variance in the amount of gardens per neighborhood. For example, in the old city center of Breda, there are barely any gardens and many buildings here are not residential use, therefore the amount of observations in these neighborhoods declined more compared to the suburbs at the edge of the city. This step reduced the amount of NDVI observations of from 55.834 to 49.405. On top of this, areas with a size larger than 2,500 m^2 and smaller than 20 m^2 were removed from the dataset, as these area sizes were generally classified as gardens, even though they were not. Areas larger than 2500 m^2 were often not gardens, as aerial photographs showed that these areas were often street sections or pastures. Areas smaller than 20 m^2 were often noise. The aerial photographs showed that this were often parking spots that had an NDVI value assigned or small areas which according to the aerial photographs were not gardens. It is possible that some smaller gardens may have been removed because of this, however the value of 20 m^2 was chosen to minimize this amount. This reduced the amount of observations to 45,792. Lastly, a visual inspection was performed of the data attributes. The aim of this inspection was to identify oddly shaped or misplaced observations. Inspection was performed using the same aerial photographs employed during the NDVI harmonization process. During this inspection, it was found that many roads, parks, buildings, car parks and industrial areas had NDVI values assigned. These areas were consequently removed manually from the analyses. 684 observations were manually removed this way.

Since data from CBS was aggregated at the neighborhood level, neighborhoods with too few observations were removed, due to missing data. These neighborhoods included Moleneind-Oost, Liesbos, Hagebeemd, Emer and Krogten. Mastbos was also removed, as the results from this neighborhood were not representative of regular land change due to a large land redevelopment which skewed the results. Appendix B Figures B1 and B2 show the redevelopment taking place. As shown in Figure 1, Krogten, Emer and Moleneind-Oost are industrial areas, while Liesbos, Mastbos and Hagebeemd are rural areas. The final dataset contains data on 44.697 individual residential gardens.

3.2 Greening

3.2.1 Greening models

This study employs three greening models to capture different interpretations of what 'greening' is, since there is not single definition. This approach was also taken by de Vries et al. (2020), who also formulated different definitions of greenery, such as the green area, garden greenness and street greenery at multiple distance thresholds. This study is concerned with changes in greenery and NDVI values however. Therefore the following definitions of 'greening' are formulated:

- Model 1 quantifies greening as a change in NDVI scores for each garden between the first observation year (2016) and the last observation year (2020).
- Model 2 classifies a garden as greened, when it's NDVI value has surpassed 0.1.
- Model 3 classifies a garden as greened, when it's NDVI value increases by 0.1 in one year. The NDVI value should be at least 0.1 after greening.

For model 2 and 3, NDVI values can not decrease to below 0.1, once the NDVI value has surpassed 0.1. Otherwise the garden will not be classified as greened. This is done to prevent including gardens which were only green for 1 year. By only including gardens that remain green, measurement inaccuracies and outliers are partly removed. These two requirements are at the basis of the classification of green gardens. Between these models, model 1 is based on a continuous scale, while model 2 and 3 are binary (greened or not greened). Therefore model 1 is analyzed using OLS (Ordinary Least Squares), while model 2 and 3 require a logistic regression due to the binary outcome (Harrell, 2015).

A binary logistic regression is used when one wishes to determine how predictor variables relate to a binary response variable (Harrell, 2015), which is greening in the case of model 2 and 3. The coefficients estimated in the regression will be the log-odds for a unit change in the predictor variable. The odds and probability will however also be converted into odds and a probability to allow for easier interpretation.

The odds are the ratio between probabilities, where either something does or does not happen, relative to an outcome (Sperandei, 2014). A coefficient of 1 indicates that both outcomes are equally probable, in other words, the variable doesn't affect chances of greening. A coefficient above 1 indicates an increase in the chances of greening to occur, whereas a coefficient below 1 indicates a decrease. The probability is the ratio between the amount of times greening takes place relative to the total amount of potential greening events (Sperandei, 2014). A coefficient of 0.5 indicates a 50/50 chance of greening happening. A coefficient above this again increases the chance, below this the chance decreases.

These three calculations combined will give good insight into the amount of greening in gardens, as they all approach and measure greening in different ways. The first model does not count the number of gardens that have been greened. In other words, it won't be exactly clear what percentage of gardens have been greened, but it does calculate the net amount of greening. This calculation also measures whether or not the greening efforts of some households is offset by the removal of vegetation by other households. The other two models will be used to measure how many people increase the amount of greenery in their garden to an NDVI score above 0.1. Model 3 provides insight into how many gardens in a certain neighborhood have experienced significant greening during 1 year. This calculation focuses on gardens which have undergone a large greening project, which significantly increased the NDVI.

3.2.2 Classification of greening

The choice for a threshold value of 0.1 for when a garden is classified as greened in model 2 and 3, was based on a study by Gandhi et al. (2015), who found that NDVI values of between 0.2 and 0.3 are generally shrub or grassland. Rainforests yield the highest NDVI scores of around 0.6 (Gandhi et al., 2015). On the other hand, areas with no vegetation or water yield negative NDVI values (Pettorelli et al., 2005), while unvegetated land yields NDVI scores around zero (Gandhi et al., 2015). Based on these findings, this study assumes that NDVI scores greater than 0.1 are classified as green gardens, while NDVI scores below this are regarded as not green enough. A slightly higher value, such as 0.15, can be chosen for stricter selection. Since this paper is aims to capture most types of greening in gardens, the cutoff value is purposely kept low in order to prevent excluding too many gardens with only limited space or ability for greening. In the studied sample, an NDVI value of 0.1 represents a garden which is about half vegetated and half paved. Examples of this are a house with a driveway and grass lawn or a paved garden with some trees. Driveways or other types of paved areas located on private property are also considered when determining the NDVI value. Since they are often unvegetated, driveways reduce NDVI values. See appendix A Figure A1 for some examples of gardens with an NDVI value of 0.1, to illustrate the minimum greenness threshold used in classification.

3.3 Accuracy assessment

The final dataset with only gardens was used to determine the accuracy of the NDVI classification. Accuracy assessments are usually done to test for the accuracy of the classification results (Balikçi et al., 2022). The classification accuracy is generally seen as the degree to which the classification accuracies are in agreement with the observed data (Foody, 2002). For this study, the classification accuracy is the degree to which measured NDVI changes can also be observed on aerial photographs. This accuracy assessment was done with the help of aerial photographs from the same year as when the classification took place. This allows for direct comparison between measured and observed changes in NDVI values. As previously discussed however, NDVI values are harmonized in order to correct for exogenous variation in NDVI values which do not relate to changes in land cover or productivity (Bai et al., 2008). The harmonization process affects the NDVI value assigned to a garden, therefore the amount of greenery visible on the aerial photographs is not directly related to the harmonized NDVI value. Therefore, during the accuracy assessment, only visible changes in land cover were classified as more or less green. For example, a grass lawn that has turned yellow due to a rainfall deficiency will not be classified as less green, because the harmonization process should have corrected for this since all gardens should be slightly less green.

The assessment was done according to the method from Balikci et al. (2022), however only 40 points are taken for each level of urban greening. The changes in urban green are defined as follows:

1. Less green (ΔNDVI < -0.1)(ΔNDVI < -0.05)
2. Equally green (-0.1 < ΔNDVI < 0.1)(-0.05 < ΔNDVI < 0.05)
3. More green (ΔNDVI > 0.1)(ΔNDVI > 0.05)

The Greek letter Δ denotes a change, ΔNDVI therefore refers to the change in NDVI score between two years.

The accuracy assessment is performed on two different threshold values. The first level requires an increase/decrease in NDVI score of at least 0.1. In other words, a garden is not labeled as more green or less green unless the NDVI score changes by at least 0.1. The second level requires an increase or decrease of at least 0.05. This is done to test for noise in the data, where for example a garden in classified as more green even though no changes in the amount of vegetation are visible. These accuracy assessment are done for multiple different seasons, from 2016 to 2017, 2017 to 2018, 2018 to 2020 and 2020 to 2022. The accuracy assessment is consequently done, based on changes in the level of urban green. This is in contrast to the method by Balikçi et al. (2022) who focus on the type of land cover, rather than changes in land cover, for their accuracy assessment. Since this study focuses on detecting changes in vegetation quantities in gardens, the accuracy assessment aims to uncover if changes in NDVI scores are reflected in actual land use change. Land use change in this case refers to the transition from a paved garden to a green garden or vice versa. Actual land use change will be determined based on aerial photographs. Changes in NDVI score should consequently be reflected in changes in vegetation density and amount in gardens. If a garden shows significant signs of greening or the removal of vegetation, then a garden should also be classified as more or less green based on the threshold values defined above. If a garden shows no signs of changes in the amount of vegetation, then the garden should be classified as equally green. This assessment will be performed visually. If there are not visible changes on the aerial photographs, then the garden will be classified as equally green. If there is visible greenery added to the garden, then it will be classified as more green. If visible greenery is removed, then the garden will be classified as less green.

The primary aim of the accuracy assessment is to determine the quality of the NDVI measurement and harmonization processes. Accuracy assessments can however also be useful in determining what errors or limitations might hinder more accurate classification of NDVI values in gardens (Foody, 2002).

3.3.1 Accuracy assessment results

	Accuracy assessment ΔNDVI > 0.1											
	More green				Equally green				Less green			
Year transition	2016-2017	2017-2018	2018-2020	2020-2022	2016-2017	2017-2018	2018-2020	2020-2022	2016-2017	2017-2018	2018-2020	2020-2022
More green	38	28	28	12	2	3	3	7	0	0	0	1
Equally green	2	12	11	3	38	36	36	33	6	0	0	35
Less green	0	0	1	0	0	1	1	0	34	40	40	4
Producer accuracy	*0.95*	*0.7*	*0.7*	*0.8*	*0.95*	*0.9*	*0.9*	*0.825*	*0.85*	*1*	*1*	*0.1*
Overall Accuracy	2016-2017	2017-2018	2018-2020	2020-2022								
	0.9167	0.867	0.867	0.575								

Table 1. Accuracy assessment with an NDVI threshold value of 0.1. For the more green 2020-2022 column, not enough greened gardens were present, therefore the accuracy is based on a lower number of observations.

The accuracy assessment results for ΔNDVI of 0.1 are shown in Table 1. The classification accuracy for ΔNDVI of 0.1 is quite high, while the accuracy for ΔNDVI of 0.05 is much lower, as shown in table C1 in Appendix C. This shows that with a ΔNDVI of 0.05 is too low, and quite a bit of noise and misclassification occurs. The reasons for these misclassifications will be discussed in chapter 5. The results of ΔNDVI of 0.05 are shown in Table C1.

Nevertheless, 88.3% of gardens was properly classified with the 0.1 threshold value for the classification of a more or less green garden. This is above the 85% mark, which is often proposed as being the minimum accuracy of land cover classification (Foody, 2002)(Thomlinson et al., 1999). Therefore the threshold of 0.1 found during this accuracy assessment is deemed adequate for the aim of this study.

NDVI data was also collected for the years 2022 and 2023. The classification of NDVI values between 2020 and 2022 showed troubling results however. All gardens that were classified as more green were gardens that were either completely bare in the prior year or were located on building sites. These observations were therefore not representative of regular garden greening, but were part of unfinished gardens which became green in the consecutive year. On top of this, compared to previous years, there were very few gardens in which regular greening took place. The data for the years 2022 and 2023 was therefore discarded, also because of the low accuracy in the 'less green' column from the accuracy assessment. Gardens were in 90% of cases classified as less green, while aerial photographs showed no change in vegetation cover. See Figure 2 also visualizes the drop in NDVI scores between 2020 and 2022.

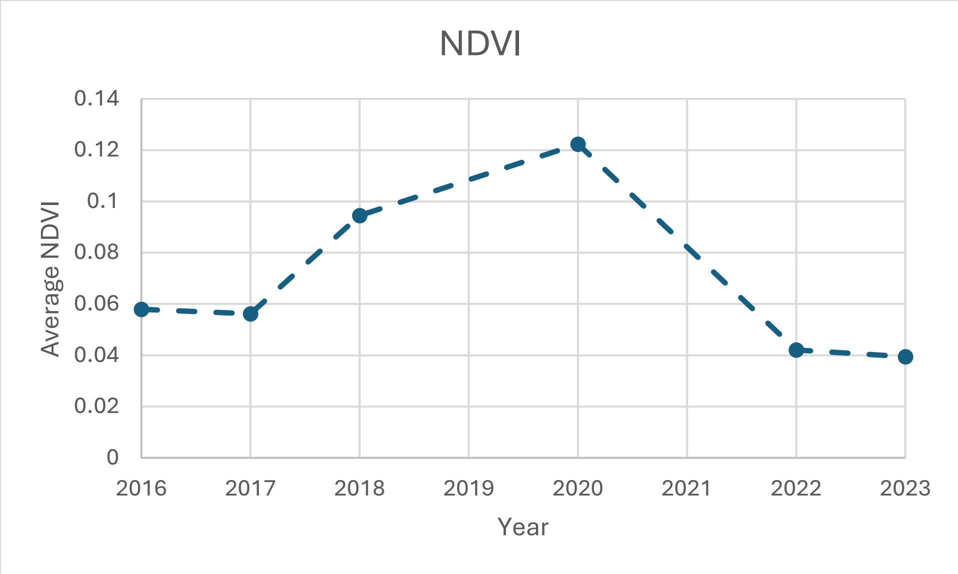

Figure 2. Average NDVI scores measured across all gardens for all years with observations.

On the basis of these results, it was determined that the NDVI value should increase by at least 0.1 in order for a garden to be classified as greened. Furthermore, all ΔNDVI values greater than 0.2 were removed from the dataset, as these observations were generally the result of a garden being renovated in one year and therefore completely bare. It is therefore not clear to what extent gardens have become greener compared to the year before the garden became bare, therefore it is not possible to state that the garden has become greener compared to before the garden became bare.

As an additional test to see if dense areas are classified equally well compared to less dense areas, a randomly chosen densely populated neighborhood was compared to a less dense neighborhood. The dense neighborhood 'Tuinzigt' was here compared to the less dense neighborhood of 'Ruitersbos'. The results from this analysis showed that the accuracy of the more green gardens in Tuinzigt was particularly poor, with only about 55% of gardens correctly classified as greened, while in the Ruitersbos, the accuracy was similar to the accuracy found during the previous accuracy assessment. This may introduce a bias, where gardens in denser neighborhoods, often closer to the city center, may suffer more from measurement inaccuracies. However, the smaller size of gardens in denser neighborhoods also made detecting vegetation changes more challenging. Therefore the classification accuracy could also be lower due to human error.

3.4 Spatial autocorrelation

Since different neighborhoods differ with regards to demographics as will be shown in section 3.5, it can be suspected that there may be some form of spatial autocorrelation present. This was also found by Gerrish & Watkins (2018), where studies that considered and corrected for spatial autocorrelation, either with the help of spatial lag or spatial error models, their coefficient for income became smaller. This indicates that without these models, there is omitted variable bias, where observations are not independently distributed, but rather affect each others. Because of this, a test for spatial autocorrelation was also conducted in this study. The amount of spatial autocorrelation can be tested for with the help of 'Moran's I', which in this case tests whether greened gardens are randomly distributed or not.

It was found that there is statistically significant spatial autocorrelation present in this dataset. For model 1, the Moran's I statistics was 0.32. For model 2, the Moran's I statistic was 0.13. For model 3, the statistic was 0.15. All these statistics had a p-value of $< 2.2e\text{-}16$ and are therefore statistically different from zero. Moran's I being different from zero means that observations are not independently distributed from each other. This suggests that the numbers found without correcting for spatially correlated errors would be incorrect (Gerrish & Watkins, 2018).

Spatial econometrics uses techniques that are designed to incorporate spatial dependence among observations into the model (LeSage, 2008). The OLS regression models assumes that observations are independently distributed, i.e. one observation does not impact the other. When observations are not independently distributed however, the regression coefficients will become biased and inconsistent (LeSage, 2008). As shown by The Moran's I statistics above and literature such as Gerrish & Watkins (2018), OLS estimates will be biased, and therefore spatial models are required. These models correct for correlation within the independent variables or the error term. Similar to many studies performed by Gerrish and Watkins (2018), the model must incorporate either spatial lag and/or spatial error correction, depending on which spatial models are significant during the study.

3.4.1 Spatial regression models

A number of spatial regression models exist. In a spatial autocorrelation model (SAR), the dependent variable (greening) is related to the dependent variable in neighboring gardens (LeSage, 2008). The SAR model corrects for the spatial lag. In spatial autocorrelation models, the scalar parameter ρ indicates the strength of the spatial lag present among observations (LeSage, 2008). This parameter is a decay factor, where gardens further away are less impacted by another garden (LeSage, 2008).

Another spatial regression model is the spatial error model (SEM). This model corrects for spatial dependence in the error term, but does not consider spatial dependence in the dependent variable, which is greening in this study. Both the SAR and SEM models capture spatial autocorrelation. However, SAR models capture this autocorrelation within the dependent variable while SEM captures autocorrelation in the unexplained part of the dependent variable, which is the error term (Ni & Zhang, 2024).

One model which combines both the SEM and SAR models is the spatial autoregressive model with spatial autoregressive disturbances (SARAR) model. This model accounts for both spatial autocorrelation in the dependent variable and in the error term (Ni & Zhang, 2024). This model should capture the spatial spillovers present in this model. In contrast to separate SAR and SEM models which may overestimate the spatial effects, this model should distribute spatial dependence more realistically between both domains of spatial dependence (Ni & Zhang, 2024). This model should therefore give accurate results in this study, as this model corrects for both types of spatial dependence present in the data. It is however only possible to compare OLS regressions with the spatial models, as the SARAR model does not support a binary dependent variable.

3.4.2 Results of the test between the spatial lag and spatial error models

In order to calculate the amount of autocorrelation within the models, Rao's score was calculated for the models in RStudio. The null hypothesis is that there is no spatial dependence, while the alternative hypothesis states that there is some form of spatial dependence (Anselin, 2001). The outcome of the test for spatial autocorrelation indeed showed the presence of spatial autocorrelation in the studied sample. For model 1, the RS-test is significant for both the spatial lag and error models, indicating the presence of both a spatial lag as well as a spatial error. For this reason, the results must be corrected for both the spatial lag as well as the spatial error.

For model 2 and 3, the coefficients are negative and insignificant. This could be because of the logistic regression technique used for model 2 and 3, since the adjusted lag and error terms are only significant in OLS versions of models 2 and 3. Based on these results, however, the spatial models will only be applied to model 1.

Greening model	Adjusted Error	Adjusted Lag	SARMA
Model 1 (OLS)	2913.3	334.75	6024.9
Model 2 (logistic)	-1562.1	-2764.2	-1537
Model 3 (logistic)	-3807.1	-5531.3	-3785.1

Table 1. All variables are significant for model 1 with a p-value of 2.2e-16 for all variables. For model 2 and 3, the p-value is 1, which indicates no significance at all.

3.5 Explanatory variables

The explanatory variables in this study consist of the variables from "Kerncijfers Wijken en Buurten 2024" (CBS, 2025) and income data is sourced from CBS as well and contains income data for the year 2020 per neighborhood (CBS, 2023). The variables 'NDVI 2016' and 'Garden area' were derived from the NDVI harmonization process and Kadaster data available at PDOK (PDOK, n.d.-a).

The control variables used in this study are:

1. NDVI 2016: Corrects for the baseline NDVI value in 2016. This coefficient helps determine how greening depends on NDVI values in the base year.
2. Garden area: The coefficient measures how a 100 m2 increase in garden area, affects the amount of greening. Murtagh & Frost (2023) found that larger front garden sizes are more green than smaller gardens according to the survey. It is therefore expected that less greening takes place in larger gardens. For model 3, the coefficient is expected to be negative, since a large increase in NDVI scores in a larger garden would require more greening effort.
3. Percentage households with children: Murtagh & Frost (2023) found that people with many children, also spend more time outside gardening. They also argue however, that people with children are younger and therefore might have recently bought a home. Therefore the garden was yet to be shaped into an acceptable form, increasing the likelihood of greening to occur. Hussain et al. (2014), found something similar, where a green hedge provides people with a sense of comfort, privacy and safety, which most parents prefer for children. This could be a reason for young adults with children to green their garden. It is therefore expected that more greening takes place in neighborhoods with more children.
4. Percentage people aged 65 and older: Murtagh & Frost (2023) argued that older people have a mature garden, in which fewer changes are required. This could explain why model 3 has a negative coefficient, where fewer large greening projects are undertaken by older people, as their garden is already in an acceptable state. Furthermore, they argue that some older people have less time for gardening due to having to care for others.
5. Average Income: A study in the United States by Troy et al. (2007) found that garden expenditures increased with income and they therefore consider greenery in gardens to be a normal good. Another study in Laten America cities by (Ju et al., 2021) found mixed results, where cities with higher socioeconomic status (SES) often had less greenery. They found a positive correlation between greening and SES over entire cities, but the reverse was true when looking at sub-cities, which are sections of cities. Yeager et al. (2023) conducted a study in the USA, where they combined NDVI data on greenness of gardens with survey data on socioeconomic characteristics of households. They found that income is positively correlated with NDVI values in gardens and within a 500 meter radius. The study the UK by Murtagh & Frost (2023) however argue that based on their earlier research, garden expenditures do not directly relate to the amount of greenery in a garden. Instead, people with lower incomes invest less money in more expensive plants that subsequently can die. Garden expenditures therefore may increase with income, but that does not necessarily result in greener gardens.
6. Urbanity: Higher urbanity generally results in less greenery according to de Vries et al. (2020), who conducted a study with NDVI data in the Netherlands. They found that high-urban

municipalities are generally less green compared to low-urban municipalities. Therefore, more urban areas will likely see less greening.

7. Cars per household: Research from Great Britain by Bates & Leibling (2012), indicates that many front gardens are transformed into paved driveways, in order to create parking space for the increasing amount of cars per household. Therefore it is expected that a larger number of cars per household decreases the amount of greening that takes place.
8. Percentage males: Grampp (1990) (adopted from (Murtagh & Frost, 2023)) found that males see gardens more as a project, where the aim is to keep the garden neat and controlled. Furthermore, Gu et al. (2024) found that women care more about plant type, variation and structure, while men care more about garden size, plant coverage and fence material. This study argues similarly to Murtagh & Frost (2023), that men care more about the overall image of the garden, while women care more about the details in gardens (Gu et al., 2024). Based on these findings, neighborhoods with a higher proportion of males could see more large-scale garden changes, while there are relatively fewer small scale changes to gardens as women dominate this type of gardening.
9. Percentage highly educated: Murtagh & Frost (2023) found that education had a positive effect on the amount of greening, even though this result was only significant for women. A higher education is therefore expected to result in slightly more greening.
10. Percentage immigrants: this variable was created by taking the sum of Western-immigrants and non-Western immigrants. Casey et al. (2017) and Duncan et al. (2013) found that neighborhoods with a large proportion of non-white inhabitants are often less green. Therefore, less greening might take place in neighborhoods with more immigrants. The percentage of immigrants will be used as a proxy for non-white people. However, it is not certain that immigrants are all non-white people.

These explanatory variables can be grouped in two categories: Socio-demographic characteristics and built environment characteristics. Socio-demographic characteristics would be: Percentage households with children, percentage people aged 65 and older, average income, percentage males, percentage highly educated and percentage immigrants. Built environment characteristics are: NDVI 2016, garden area, urbanity and cars per household.

The 56 neighborhoods in Breda exhibit distinct spatial characteristics. The population densities vary between 59 to 13,088 per km^2, the percentage of married people varies between 12 and 46%, the percentage of elderly (65+) varies between 5 and 37%, the percentage of single person households varies between 20 and 74%, the percentage males varies between 45 and 55% and average yearly income varies between €31,200 and €37,600 per person (CBS, 2023, 2025, n.d.-b). The differences in the socioeconomic characteristics of the neighborhoods in Breda will be used as a proxy for how different neighborhood compositions, affect the amount of greening.

3.5.1 Elastic Net variable selection

Variable selection is primarily based on which variables were relevant in other literature, because this allowed for the validation of the results found in this study. Nevertheless, the elastic net variable selection method was used to get an indication of which variables are important in what models. The elastic net was first introduced by (Zou & Hastie, 2005). It is a method which combines both ridge and lasso regression. The elastic net provides a approach to deal with variable selection, especially when there are many predictors that are highly correlated (see section 3.6 for the test of collinearity among predictors). The Elastic Net regression combines the benefits of both Ridge regression as well as Lasso regression. The Ridge regression technique shrinks variable coefficients towards 0, while Lasso performs variable selection by shrinking some variables to exactly 0. These methods standardize the coefficients of the variables (Zou and Hastie, 2004), making them easy to compare regarding how important they are in determining the amount of greening in gardens.

The parameters of the models were defined as follows. To optimize for both variable selection (Lasso) and predictive accuracy, the L1 and L2 penalties were balanced in the Elastic Net model. The regularization strength (λ) was calculated with the help of 10-fold cross validation with the help of the 'lambda. 1se' command in RStudio. The optimal value of lambda is generally calculated this way (Zou & Hastie, 2005). This way, the largest value of λ is selected which is at most 1 standard error away from the minimum error. This larger value of λ was selected, because lower values of λ ('lambda.min' in RStudio) resulted in no variable selection. Furthermore, the variable that controls the balance between Ridge and Lasso in the Elastic Net regression, α, was set at 0.5. This ensures that both Ridge and Lasso regularization techniques are employed, which yields the Elastic Net regression and leverages the advantages of both (Zou & Hastie, 2005).

3.5.2 Elastic Net results

The elastic net results are shown in Table 2. Model 1 has the most variables that remain included. The amount of variables in model 3 is surprisingly low, indicating that enough variation is explained by the 3 selected variables. For this reason the variables 'Age 65+' and '% 40% lowest incomes' were removed, to show what other variables were of interest. This is shown in the column 'Model 3.1'. Model 3.1 highlights some other variables of interest that affect greening in model 3.

The different outcomes between each model are likely explained by different drivers for different types of greening. In model 3, the variables 'NDVI 2016' and '% age 65 +' are particularly large coefficients. This implies that older people and people with lower incomes are less likely to make large changes to the greenery in their gardens. Furthermore, some demography variables are quite large. Model 1 includes more explanatory variables, indicating that the benefits of additional complexity are larger than the cost of the increased elastic net penalty (Zou & Hastie, 2005). Overall, the variables 'average income', 'NDVI 2016' and 'garden area' are included in all models, therefore these are some of the variables that will be included. These coefficients don't allow for inference however, as ridge and lasso introduces bias to reduce variance via the bias-variance trade-off. Furthermore, elastic net has a grouping effect, where highly correlated variables are either all included or excluded (Zou & Hastie, 2005), which affects what variables are kept in the model.

Variable	Model 1	Model 2	Model 3	Selected
Number of inhabitants	-3.52E-07		-7.45E-07	2
Average income	7.54E-06	5.08E-05	-4.60E-05	3
NDVI 2016	-0.37982	-0.36986	0.393333	3
Address density			2.37E-06	1
% Age 0-14			9.08E-05	1
% Age 15-24		-0.00046		1
% Age 45-64	-0.00053	NA		1
% Age 65+		7.21E-05	-0.00364	2
% Antilles/Aruba			0.002011	1
% hh with children	-0.00053			1
Rental Housing		-0.00065	-1.00E-04	2
Owner occupied		0.000515	7.28E-05	2
% Moroccan	-0.0009			1
% Not married	0.000216			1
% Surinam	-0.00107	-0.00811		2
% Turkey			-0.00059	1
% Western Immigrants	-0.00045	-0.0033		2
Properties after 2000	-8.66E-05			1
Urbanity	6.10E-05			1
Property value	6.79E-05			1
Garden area	1.58E-09	-3.70E-09	-1.53E-09	3
% 20% highest income	2.15E-05	0.00018		2
% 40% lowest income			-0.00023	1

Table 2. Elastic Net results. Some cells are empty, as these were not selected in the Elastic Net variable selection technique.

3.6 Testing for multicollinearity

Due to the large amount of variables in the CBS data on neighborhood characteristics (CBS, 2025), it is important to test for multicollinearity among these variables. Multicollinearity is a phenomena, where two linear variables are significantly correlated, not only with the dependent variable, but also with each other (Young, 2018, cited by Shrestha, 2020)(Schober et al., 2018). This results in inflated Standard Errors (SE) (Harrell, 2015). Multicollinearity can also increase the variance of the coefficients of the variables. This can make the coefficients unstable according to (Keith (2019), cited by Shrestha (2020). This phenomena was also observed in our data, where the coefficient for the percentage of owner owned properties was positive, but became negative when the percentage of rental properties was included in the regression. This indicates the presence of significant collinearity between these variables, which makes sense, as the percentage of owner owned properties is nearly directly dependent on the percentage of rental properties in a neighborhood.

It is therefore important to test how highly multicollinear variables affect the coefficients of other variables. This will make further interpretation of the final coefficients easier, as highly correlated coefficients return inconsistent results. In order to test for correlation between variables, the Pearson correlation coefficient is often used (Schober et al., 2018). The Pearson correlation coefficient ranges between -1 and +1, where a positive value indicates a positive correlation, while a negative value indicates a negative correlation. A coefficient value of 0 indicates no correlation between the variables (Schober et al., 2018), meaning that they are independently distributed. Furthermore, the Variance Inflation Factor (VIF) will be used, as this can quantify the amount of collinearity between variables in a regression (Young, 2018). According to a study by Craney & Surles (2002), a VIF of greater than 5 is often considered as the highest allowable value. They argue however that a VIF cutoff value of 2 is more appropriate when you do not want more than half of the variation in a variable to be explained by another variable. This thesis therefore only allows for VIF values up to 5.

The Pearson correlation coefficients for the explanatory variables included in this thesis are shown in Figure 3. Furthermore, an extended Pearson's correlation matrix was calculated for a wider range of variables, to provide insight into how variables relate to each other. This is shown in appendix F Figure F1. As shown in Figure F1, there is heavy multicollinearity among variables, indicating that variables should be carefully selected, based on multicollinearity among variables.

Figure 3 shows that the highest absolute correlation coefficients are -0.64 and 0.63, outside of the correlation coefficients for the variable '% immigrant'. According to (Schober et al., 2018), this can be interpreted as moderate collinearity, while all other coefficients show weak signs of collinearity. This shows that multicollinearity between variables is kept to a minimum. The variable '% immigrant' was nevertheless included in the final regression, since the higher multicollinearity scores didn't change the direction of any of the other variables, which means that the other variables didn't become positive instead of negative or vice versa after the inclusion of the variables.

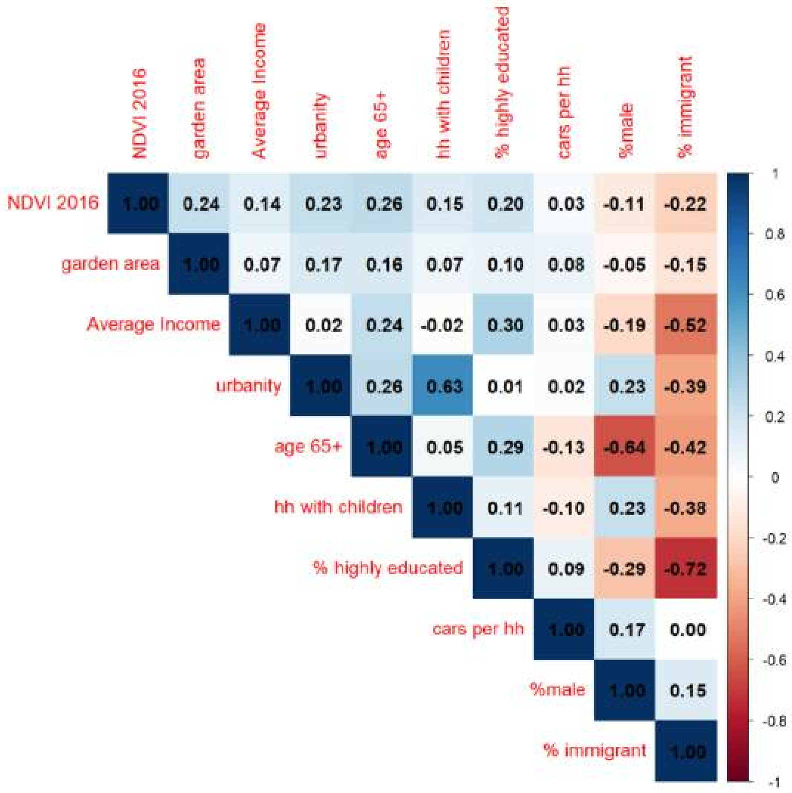

Figure 3. Pearson's correlation coefficients of the selected variables.

Since the extended Pearson's correlation index in Figure F1 in Appendix F shows that there are many variables with strong collinearity with many variables having a correlation coefficient of greater than 0.7 (Schober et al., 2018), it was decided to thoroughly test for this multicollinearity. AIC (Akaike Information Criterion) value will be used as a measure of model performance, while punishing model complexity (Harrell, 2015). The AIC value is expressed as

$$AIC = 2V - 2\log(L) ,$$

here V is the number of parameters included in the model, while L is the likelihood function that describes how well the model explains data (Troy et al., 2007), which in this case is the greening of gardens. The aim is to minimize the AIC value of a model. This is done by balancing between the amount of parameters included in the model, and the likelihood function. The likelihood function should be maximized, while keeping the amount of variables low. This way, the estimated number of parameters in a regression is reduced (Harrell, 2015). The most appropriate spatial model will also be chosen based on the highest likelihood value and lowest AIC value.

3.7 NDVI change analysis

The NDVI value of a garden is calculated in the following manner. The NDVI exploits the fact that chlorophyll in plants absorbs visible light, while near-infrared (NIR) is reflected by healthy vegetation (Casey et al., 2017). The formula for the NDVI is:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

NIR contains data from the Near InfraRed band, while RED contains data from the red band, which contains the visible light (Pettorelli et al., 2005). The formula indicates that NDVI values can range from -1 to +1 and are therefore normalized. Healthy plants return a high NDVI value, since healthy plants reflect NIR and absorb RED. Therefore the amount of ‘light’ in the NIR band is high, while the amount of light in the red band is much lower, resulting in a high NDVI value. Unhealthy plants absorb less red light and therefore have a lower NDVI score. A high NDVI score indicates that there is lots of plant growth and high plant density. Healthy and productive vegetation has the highest NDVI score. A lower NDVI score on the other hand indicates unvegetated land or unhealthy vegetation. This could for example be urban areas or barren land (Lenney et al., 1996).

3.7.1 Harmonized NDVI

NDVI values need to be harmonized, to correct for the presence of cloud cover, haze, the time of the growing season and atmospheric conditions (de Jong et al., 2011), but also for fluctuations in rainfall, sunshine or large scale atmospheric disturbances (Bai et al., 2008). These factors can cause increases or decreases in NDVI values, without reflecting actual changes in land cover or productivity (Bai et al., 2008). This thesis also uses harmonized NDVI data, to enable comparison of NDVI values across years.

4. Results

The area weighted NDVI value changes across neighborhood are shown in Figure 4. These NDVI scores are corrected for the area of gardens, where larger gardens weigh more heavily compared to smaller gardens. All neighborhoods demonstrated signs of greening. Overall, the area-weighted NDVI scores have increased by 0.054 across all gardens. In Figure D2 in appendix D the NDVI values of neighborhoods in 2016 are shown. The NDVI values in 2016 are also mapped in appendix D in Figure D1. The highest recorded increase was 0.12 in City, which is the old city center of Breda. The lowest increase in NDVI scores was in Effen-Rith, which is a more rural neighborhood, as shown in Figure 1. Figure 5 shows the average increase in NDVI values per neighborhood in Breda on a map. This map shows clearly that more greening took place in more urban and central neighborhoods, while limited greening took place in more rural neighborhoods at the city's edge. Furthermore, compared to Figures D1 and D2 in appendix D, Figures 4 and 5 show how greening primarily took place in neighborhoods with low initial NDVI values.

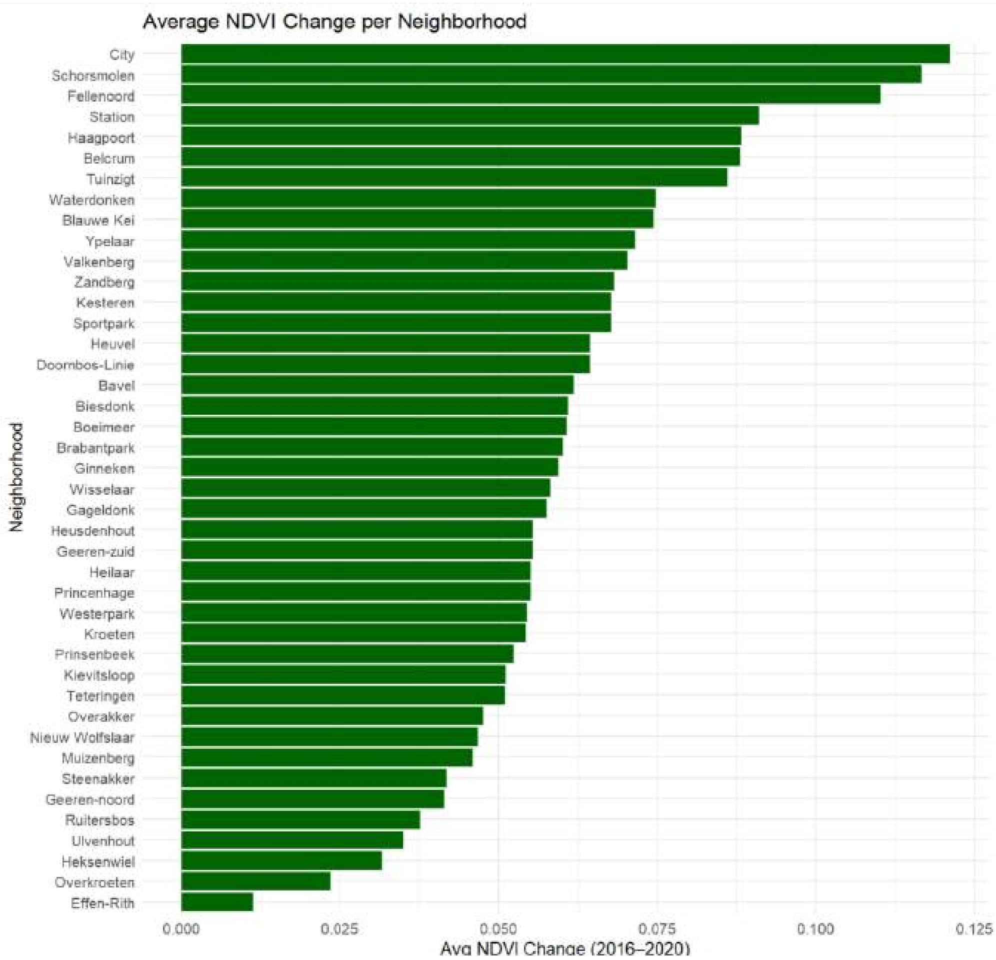

Figure 4. Weighted NDVI score increases between 2016 and 2020.

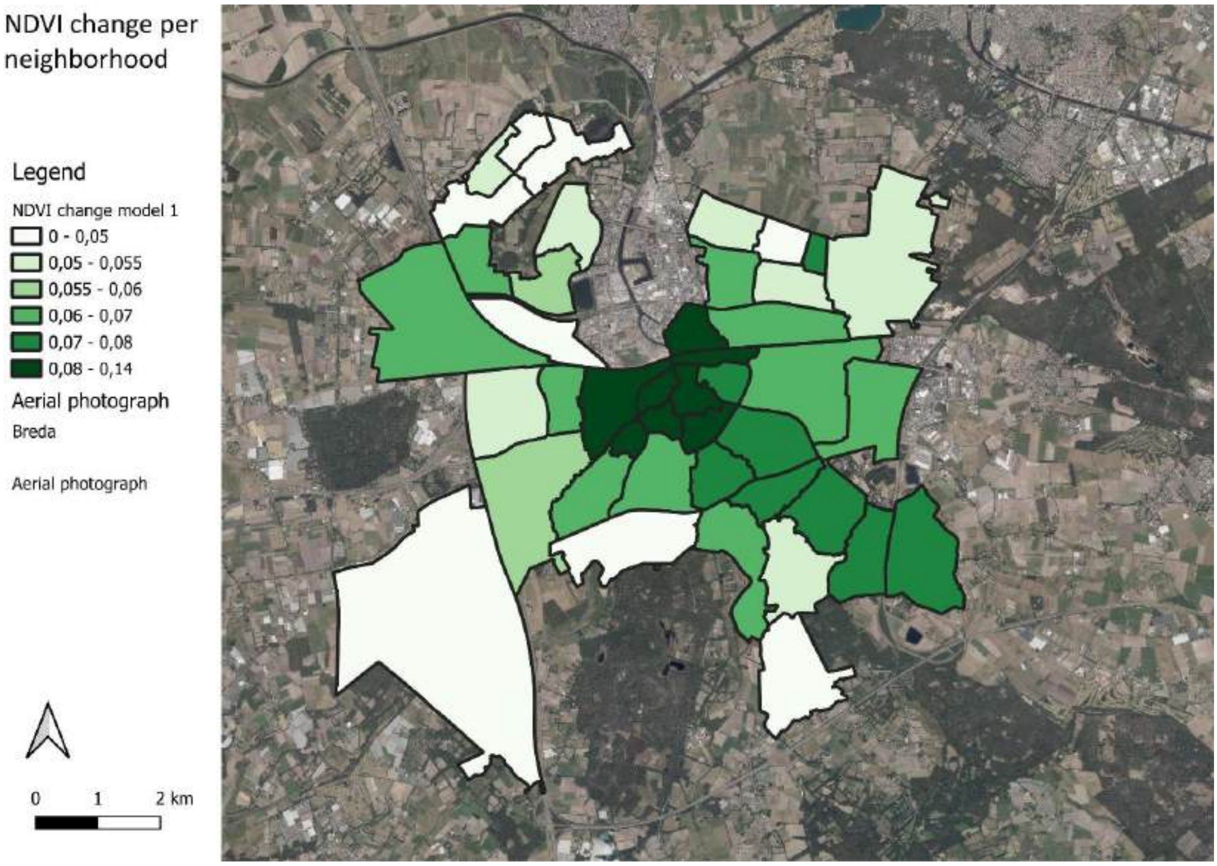

Figure 5. NDVI change per neighborhood between 2016 and 2020 as visualized in RStudio

Figure 6 shows a transition matrix. The categories are classified as follow:

1. NDVI value below zero
2. NDVI value between 0 and 0.1
3. NDVI value above 0.1

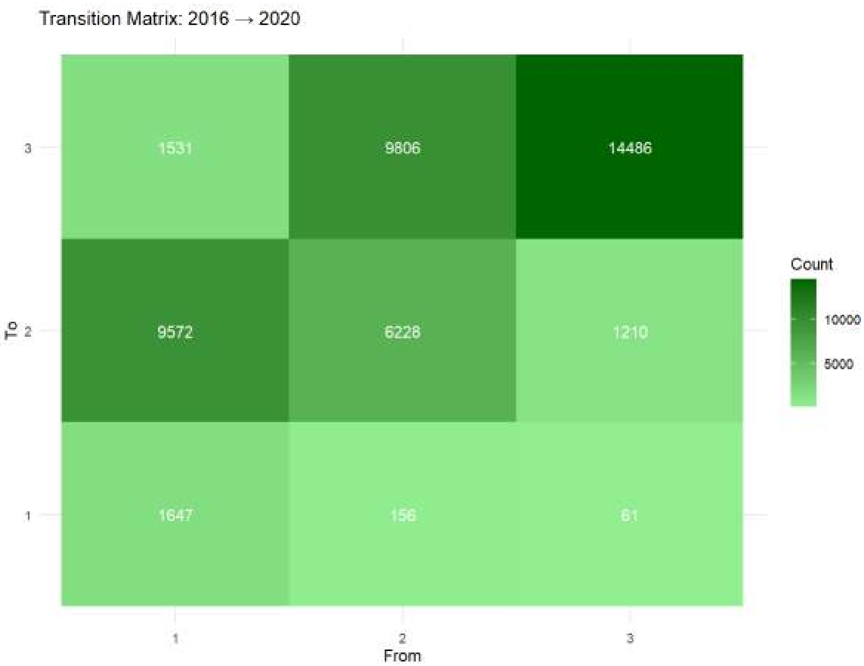

Figure 6. Garden transition matrix from 2016 to 2020.

2016-2017			
Transition matrix	**No vegetation**	**Sparse vegetation**	**Dense vegetation**
No vegetation	13182	3944	116
Sparse vegetation	2431	14451	1220
Dense vegetation	409	4775	15306
2017-2018			
Transition matrix	**No vegetation**	**Sparse vegetation**	**Dense vegetation**
No vegetation	9524	6016	482
Sparse vegetation	1585	11217	10368
Dense vegetation	370	1417	14855
2018-2020			
Transition matrix	**No vegetation**	**Sparse vegetation**	**Dense vegetation**
No vegetation	3206	7648	625
Sparse vegetation	429	10872	7349
Dense vegetation	117	2325	23263

Table 3. Transition matrixes decomposed for individual years, the left side is the classified category in the first year, top row is the classified category for the second year. The numbers

As shown in the transition matrix in Figure 6, many gardens transitioned from 1 to 2 and from 2 to 3. In particular gardens with NDVI values below 0 in 2016 saw nearly all gardens increase to NDVI levels above 0 and even above 0.1. Table 3 shows that most gardens experienced the highest NDVI increases in 2017 to 2018. This jump is also visible in Figure 2.

4.1 Regression results

This section will summarize the results from the data and validates these findings with data from existing literature. First of all, we discuss the average increase in NDVI in model 1. Then model 2 with the amount of gardens with an NDVI value that exceeded 0.1 without a decrease in any of the following years will be discussed. Lastly, model 3 will be discussed, where NDVI values had to increase by at least 0.1 to above an NDVI value of 0.1, without a decrease in any of the following years.

4.1.1 Base Regression results

Base regressions

	Dependent variable:		
	Model 1	Model 2	Model 3
	OLS	*logistic*	*logistic*
	(1)	(2)	(3)
NDVI 2016	-0.039***	-0.248***	0.472***
	(0.0002)	(0.013)	(0.017)
Area (100 m^2)	0.002***	-0.071***	-0.041***
	(0.0001)	(0.007)	(0.008)
% hh with children	-0.001***	0.013***	0.016***
	(0.00004)	(0.002)	(0.003)
% age 65 +	-0.0002***	0.008**	-0.053***
	(0.0001)	(0.003)	(0.004)
Average income	0.00002***	0.001***	0.0005***
(per €1,000)	(0.00000)	(0.0001)	(0.0002)
Urbanity	0.002***	0.078***	0.189***
1 = high, 5 = low	(0.0003)	(0.018)	(0.024)
% immigrant	-0.0001***	-0.003	0.001
	(0.00004)	(0.002)	(0.003)
Cars per hh	-0.008***	0.142	-0.760***
	(0.002)	(0.116)	(0.161)
% male	-0.001***	-0.051***	0.019
	(0.0003)	(0.014)	(0.018)
% highly educated	0.001***	0.012***	0.003
	(0.00003)	(0.002)	(0.002)
Constant	0.124***	0.098	-2.689***
	(0.014)	(0.750)	(0.975)
Observations	44,697	44,697	44,697
R^2	0.392		
Adjusted/McFadden R^2	0.392	0.020	0.044
Log Likelihood		-24,638.940	-15,851.190
Akaike Inf. Crit.	-148100	49,299.880	31,724.390
Residual Std. Error	0.046 (df = 44686)		
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4. Base regression results (hh = household). The highest VIF (Variance Inflation Factor) in these models is 4.93 for the variable 'immigrant'.

Model 2	**log-odds**	**odds**	**probability**	**Interpretation**
NDVI 2016	-0.25	0.779	0.438	Decreases
Garden area	-0.071	0.931	0.482	Decreases
% hh with children	0.014	1.014	0.503	Slight increase
% age 65+	0.009	1.009	0.502	Slight increase
Average income	0.001	1.001	0.500	No effect
Urbanity	0.086	1.090	0.521	Increase
Cars per hh	*0.124*	*1.132*	*0.531*	*Increase*
% male	-0.047	0.954	0.488	Decrease
% immigrant	*-0.03*	*0.970*	*0.493*	*Slight decrease*
% highly educated	0.014	1.014	0.503	Slight increase
constant	-0.311	0.733	0.423	

Table 5. Interpretation of the logistic regression coefficients of model 3. Coefficients in Italic are insignificant.

Model 3	**log-odds**	**odds**	**probability**	**Interpretation**
NDVI 2016	0.472	1.603	0.616	Increases
Garden area	-0.041	0.960	0.490	Slight decrease
% hh with children	0.016	1.016	0.504	Slight increase
% age 65+	-0.053	0.948	0.487	Slight decrease
Average income	0.0004	1.000	0.500	No effect
Urbanity	0.187	1.206	0.547	Increase
Cars per hh	-0.758	0.469	0.319	Decrease
% male	*0.018*	*1.018*	*0.504*	*Slight increase*
% immigrant	*0.001*	*1.001*	*0.500*	*No effect*
% highly educated	*0.002*	*1.002*	*0.500*	*No effect*
constant	-2.615	0.073	0.068	

Table 6. Interpretation of the logistic regression coefficients of model 3. Coefficients in Italic are insignificant.

4.1.2 Discussion base regression results

The base regression results are shown in Table 4. The regression results can be interpreted as the following. For the 'percentage variables', a 1% increase or decrease results in a change in the rate of greening, the size of the coefficient in Table 4 for model 1. For the other variables, a one unit increase results in a increase in NDVI values the size of the coefficient. For model 2 and 3, the odds and probability in Table 5 and 6 can be used as an indicator for how the coefficients affect the amount of greening. The 'interpretation' provides an overview of the economic significance of each variable. A slight increase/decrease is an maximum change of 0.01. Otherwise the variable is interpreted as increase or decrease.

First of all, garden area has a small positive effect in model 1, while the effect is negative and quite large in model 2 and 3. This could indicate 2 things. First of all, larger gardens are generally more green, as shown Appendix E Figure E1. Because of this, gardens are not able to be 'greened' anymore. Second, due to the larger size, large gardens require more effort to be greened, therefore they might be less often greened.

An increase in income by €1,000 has no effect on greening in model 2 and 3, while it has a small impact in model 1, as shown in Table 4, 5 and 6. This is in line with research by Murtagh & Frost (2023) who also found that income has only a small positive effect on how green a garden is. The result is however in contrast to other research, where it was found that income increases greening (Ju et al., 2021; Troy et al., 2007; Yeager et al., 2023). Only the study by Murtagh & Frost (2023) was conducted in Europe however.

For model 1 and 2, the NDVI value in 2016 had a negative effect on the particular types of greening, indicating that more green gardens are more difficult to make any greener. For model 2, greening also can't occur when the NDVI value was above 0.1 in the first year, decreasing the chance of greening for green gardens. For model 3, the coefficient is positive, as an higher NDVI score in 2016 makes it easier for the NDVI score to exceed 0.1. It is therefore easier for already fairly green gardens to meet these criteria, as these gardens will automatically meet the criteria when the NDVI increases by 0.1. Model 3 could be positive due to most gardens in which the NDVI value is able to increase by 0.1, have an NDVI value far below 0.1. It would be more difficult to increase and NDVI score from 0.11 to 0.21 compared to 0.01 to 0.11, as the former would require much higher vegetation densities in the garden.

For households with children, the coefficients are only positive for model 2 and 3. This indicates that neighborhoods with many children often see NDVI values increase to above 0.1, indicating that people with children more often seek green gardens. This confirms the results found by Hussain et al. (2014) and Murtagh & Frost (2023). The coefficient for model 1 is very small and therefore has no economical significance, which means that the coefficient is too small to matter in a meaningful way.

Regarding the percentage of people aged 65 and older, the coefficient in model 3 is larger compared to the other 2 models. This indicates that in neighborhoods with many elderly, gardens less often see large increases in greenery. In other words, older people therefore less frequently make large changes to the amount of greenery in their garden. Murtagh & Frost (2023) found that age didn't really affect the amount of green in gardens. This could explain the mixed results found in this thesis. This thesis however does find that large vegetation cover increases are even more unlikely with more older people in a neighborhood.

The level of urbanity has a positive correlation in all three models, although only model 1 and 3 are significant. Model 3 is interesting, as is has quite a large coefficient. This coefficient indicates that people in less urban areas, which is indicated by a higher value for urbanity, more often see NDVI increases in their garden of 0.1 or greater. More rural gardens are thus much more likely to be greened.

This could explain why (de Vries et al., 2020) found more urban areas are less green. On the other hand, as shown in section 4.1, most greening took place in the city center. Therefore one would assume that more urban areas see more greening. Section 4.1 doesn’t correct for NDVI values in 2016 however. More urban neighborhoods had much lower initial vegetation cover, therefore much more room for improvement was possible. The lack of additional explanatory variables in section 4.1 therefore gives an incomplete view the main drivers behind greening.

The number of cars per household decreases the amount of greening in model 1 and 3, while model 2 is insignificant. This is in line with other research, where it is argued that greenery has to make way for the parking of cars (Warhurst et al., 2014).

The percentage of males is negatively associated with model 1 and 2, while it positively affects greening in model 3. The negative values of model 1 and 2 indicate that males are less likely to add small areas of green to gardens, while the positive value in model 3 indicates that neighborhoods with a high proportion of males are more likely to significantly green their garden. This is seemingly an odd relationship, but as argued in section 3.5, men care more about the larger image of what a garden looks like, while women care more about smaller details (Grampp, 1990; Gu et al., 2024; Murtagh & Frost, 2023). This could explain why neighborhoods with a larger proportion of males have a elevated chance of large greening projects to occur in gardens.

Lastly, it was decided to include the % immigrants. Papers from other countries indicate that people with a non-white skin color often have less greenery in the direct vicinity of their garden (Casey et al., 2017; Duncan et al., 2013). Our results however show that the percentage of immigrants has minimal effect on greening in gardens. Even though non-white people therefore often live in less green neighborhoods, these initial results indicate that this does not extent to the greening of gardens. In section

4.1.3 Spatial econometric models

Due to the presence of both a spatial lag and a spatial error component model 1, both models are estimated. The results of the SAR and SEM models are shown in Table 7, the results of the SARAR model are shown in Table 8. All spatial models will be compared to the OLS regression to see how spatial dependence affects the coefficients.

Comparison of spatial error model

	Dependent variable:		
	Greening (Model 1)		
	OLS	*spatial error*	*spatial autoregressive*
	Model 1	Error	lag
	(1)	(2)	(3)
NDVI 2016	-0.039***	-0.042***	-0.036***
	(0.0002)	(0.0003)	(0.0002)
Area (100 m^2)	0.002***	0.001***	0.002***
	(0.0001)	(0.0001)	(0.0001)
% hh with children	-0.001***	-0.001***	-0.0005***
	(0.00004)	(0.0001)	(0.00003)
% age 65 +	-0.0002***	-0.0001	-0.00004***
	(0.0001)	(0.0001)	(0.00000)
Average income	0.00002***	0.00002***	0.00002***
(per €1,000)	(0.00000)	(0.00000)	(0.00000)
% immigrant	-0.0001***	-0.0001*	-0.00001***
	(0.00004)	(0.0001)	(0.00000)
Urbanity	0.002***	0.003***	0.003***
(1 = high, 5 = low)	(0.0003)	(0.001)	(0.0003)
Cars per hh	-0.008***	-0.006*	-0.006***
	(0.002)	(0.003)	(0.002)
% male	-0.001***	-0.001**	-0.001***
	(0.0003)	(0.0004)	(0.00004)
% highly educated	0.001***	0.001***	0.001***
	(0.00003)	(0.0001)	(0.00002)
Constant	0.124***	0.122***	0.081***
	(0.014)	(0.020)	(0.001)
Observations	44,697	44,697	44,697
R^2	0.392		
Adjusted R^2	0.392		
Log Likelihood		76,777.930	75,523.630
$sigma^2$		0.002	0.002
Akaike Inf. Crit.	148096	-153,529.900	-151,021.300
Residual Std. Error	0.046 (df = 44686)		
F Statistic	2,886.409*** (df = 10; 44686)		
Wald Test (df = 1)		6,346.698***	3,270.062***
LR Test (df = 1)		5,436.071***	2,927.490***
Note:			*p**p***p<0.01

Table 7. results of the spatial error and lag models compared to the regular OLS model

Variable	Model 1	SARAR 1
(Intercept)	0.123801	0.160689
NDVI 2016	-0.03936	-0.04181
Area (100 m^2)	0.002356	0.00092
% hh with children	-0.00069	-0.00082
% age 65 +	-0.00024	-0.00027
Average income	2.19E-05	1.92E-05
% immigrant	-0.00014	-0.00024
Urbanity	0.002455	0.002569
Cars per hh	-0.00784	-0.00777
% male	-0.00087	-0.00119
% highly educated	0.000727	0.000745
rho		-0.17396
lambda		0.500016
AIC	-148100	-153950
Log-likelihood		76990

Table 8. SARAR model results compared to the OLS model

4.1.4 Spatial model comparison

Similar to LeSage & Pace (2008) and Ni & Zhang (2024), the log-likelihoods of the different models will be compared, to determine what spatial model best fit the data. The AIC score will also be used, as this score punishes model complexity, which log-likelihood doesn't do. Based on the log-likelihood and AIC scores, the SARAR model explains model 1 best. Furthermore, the spatial error model performs better compared to the spatial lag model. These results indicate that correcting for both a spatial autocorrelation in the error term and the dependent variable yields the best results, which was already found in the meta-analysis by (Gerrish & Watkins, 2018). Overall, while the spatial models do explain some variation in the data, as most coefficients slightly increase or decrease in size, they do not significantly affect the conclusions of this thesis.

4.2 Policy implications

Based on these results, a few areas of improvement have been identified. First of all, according to (Stobbelaar et al., 2021) behavioral failures are often a limitation to greening. Many incentives are already undertaken to tackle these problems, however there is a lack of creative solutions on these websites, which will be discussed in section 4.2.1. Second, this thesis showed that neighborhoods with low initial vegetation cover generally showed larger increases in vegetation cover. Greening campaigns should therefore focus on the less green neighborhoods, as these allow for the largest improvements in vegetation cover. Third, since the number of cars per household directly seems to affect the amount of greening in gardens, measures should either be taken to reduce the amount of vehicles in gardens, or people could be informed about ways in which their driveway can become more green. According to the study by Bates & Leibling (2012), cars are more frequently parked outside nowadays, therefore garden space is often converted into paved driveways (Warhurst et al., 2014). This is similar to the findings in this study, where households with more cars were less likely to green their garden.

4.2.1 Creative greening solutions

It is therefore important to actively encourage people to green driveways and pathways within their garden (Beleri & Kotyal, 2024). Higher permeability and vegetation cover can be achieved by replacing conventional tiles with permeable tiles, such as green grids or permeable pavers. Such paved surfaces contain spaces for vegetation to grow thru. Furthermore, pebbles or water permeable paving blocks can be used, as these also allow for rainwater infiltration (Beleri & Kotyal, 2024; Sha et al., 2021). Beleri & Kotyal (2024) do highlight however that compressive strength of surfaces should be considered, as not all surfaces allow for traffic.

These permeable paving types allow for water infiltration, which reduces excess runoff. These types of paving should have primarily be used on low-speed roads, parking lots or sidewalks, due to the reduced strength of this paving (Sha et al., 2021). These characteristics also make it useful for parking spots in gardens (Beleri & Kotyal, 2024). It can both serve as a driveway, while also benefitting the environment. Furthermore, nature-based solutions can be used for pathways in gardens. This could be anything form wooden planks to grass pavers and green grids (Beleri & Kotyal, 2024).

The municipality of Breda already provides subsidies for rainwater infiltration measures (Gemeente Breda, 2025), but the website lacks creative design solutions to facilitate this. The municipality could therefore improve their (digital) educational and informational campaigns by promoting creative solutions to both have a functional and ecological garden.

5 Discussion

5.1 Limitations

During the data preparation in QGIS, it was found that sometimes rows of houses with garden covered by large streets showed large variation in NDVI values between years. These observations were consequently removed, as these changes in NDVI values were not visible on the aerial photographs, since the gardens were only partly visible to begin with. Nevertheless, there are still gardens left in which NDVI values are affected by surrounding vegetation or overlapping vegetation from other gardens. Furthermore, it was also found that many non-garden areas within residential areas had an NDVI value assigned. This could be anything from parking lots to street sections. These areas had to be manually removed, although it is almost impossible to detect all these areas. Therefore some observations might not relate to NDVI data from a garden, although this amount of observations is kept to a minimum in this thesis.

Gerrish & Watkins (2018) argue that grass voluntarily grows on non-impervious surfaces, while tree cover is more influenced by humans in urban areas. Therefore they argue that inequity tree cover in cities, might be more pronounced than inequity in overall vegetation, including grass. This study measures the overall amount of vegetation, which might explain why income doesn't really affect the amount of vegetation cover, while tree cover is affected. Further research could therefore consider tree canopy cover rather than NDVI values, as this might yield different results.

As shown in the section 4.1.1, the McFadden R^2 values of model 2 and 3 are particularly poor, especially compared to model 1. These R^2 values are however not comparable to the OLS R^2. According to (Hemmert et al., 2018), a McFadden R^2 value between 0.2 and 0.4 can be considered a good fit. Model 2 and 3 therefore don't have a good fit according to this metric. These low fits are however comparable to the study by Murtagh & Frost (2023), who conducted a survey under 1,000 respondents in England about the properties of their front gardens. Their R^2 values also ranged from 0.06 to 0.11, depending on the dependent variable of interest. Furthermore, Murtagh & Frost (2023) have access to a wide range of other garden-level variables, such as education, income, age and gender, which allows for more accurate use of explanatory variables, since this thesis relies on neighborhood level data for many variables. It is therefore not unusual to find low R^2 values for this type of study. Data from the CBS alone is not enough to explain why people may or may not green their garden as many unobservable factors might affect it (Murtagh & Frost, 2023). Inclusion of more explanatory variables did increase the McFadden R^2 values however.

Due to model 2 and 3 requiring a logistic regression, it was not possible to estimate model 2 and 3 with the same spatial models as used in model 1. Attempts to use alternative models that functioned similarly did not work for this thesis, although a more experienced RStudio user might be able to fit logistic spatial models to model 2 and 3. Model 1 was also the model with the strongest spatial dependencies, as shown by the Moran's I scores in section 3.4. It is therefore expected that the conclusions for model 2 and 3 will not change when correcting for spatial autocorrelation. Because of this, inference is possible for model 1 with the SARAR model, while model 2 and 3 are biased due to the spatial dependency. Nevertheless, it is suspected that the conclusions for model 2 and 3 do not change, since they did not change for model 1.

5.1.1 Accuracy Assessment limitations

The accuracy assessment accuracies are not as high compared to the accuracies obtained by (Balikçi et al. (2022). This can be caused by the smaller discrepancies between the different types of land cover. While Balikçi et al. (2022) differentiate between water, greenspace and urban, this study aims to differentiate between different densities of greenery in gardens. Gardens are particularly difficult to calculate accurate NDVI measurements for, since part of the ground can be blocked by swimming pools and sunscreens. These types of objects in gardens can also vary year by year, complicating accurate measurement changes of NDVI values in gardens. Primarily smaller gardens were affected by this, as objects can cover a significant part of the green space in a garden. Larger gardens were generally classified better as these are relatively less affected by objects such as sunscreens or swimming pools. The small scale of some gardens makes them susceptible to large changes in NDVI scores year on year, which introduces noise in the dataset. In section 6.1, a solution is proposed to combat this, where satellite imagery is not collected during the summer.

5.2 Model discussion

The models all approach greening in a different way, to give deeper insights into the main drivers behind why some gardens are greened, while other are not. Model 3 only classifies gardens as greened, when there is an increase in NDVI score of 0.1 or greater and when the NDVI score after greening is greater than 0.1. This is the most selective model, where only gardens with large increases in NDVI scores are included. Compared to the other models, it is possible to visually confirm the increase in NDVI scores, as these increases are clearly visible on the aerial photographs. The downside compared to model 1 and 2 is however, that smaller increases go unnoticed. This model does not suffer from the problem of outliers however, as the accuracy assessment of ΔNDVI > 0.1 shows that gardens are classified as more green with 78% accuracy.

Model 3 is not with problems however. For example, gardens which go thru a renovation are often bare during one year. The following year, the garden is green again, which significantly increases the NDVI score, which consequently results in the garden being classified as 'greened'. The downside of this is that a garden might have not greened compared to before the renovation, but only compared to during the renovation when the soil was bare. This results in a situation where a garden is properly classified as 'more green', without actually increasing the amount of green area in the garden. Model 3 therefore isolates gardens that have undergone significant change, without always necessarily resulting in more greenery.

In order to evade this problem, model 2 was introduced. Here, no large increases in NDVI scores were mandatory, which prevents only selecting gardens which have been (partly) renovated. Now, smaller types of garden greening will also be considered in the analysis, as garden greening does not always increase the NDVI score by more than 0.1 in one year. Recall that during the accuracy assessment, NDVI scores of 0.1 were required in order to exclude noise in the dataset, where a garden in classified as ‘greened’, without actually showing sings of greening. Therefore, one could expect that this approach introduces noise in the dataset. In order to limit the noise, greening should now be permanent. By excluding gardens that only became more green during one year and less green during the following, noise could be limited. Noise in this case is often a scenario where a garden shows significant increases in NDVI values during one year and a consequent decrease in the following year, even though no significant changes are visible in the garden. The pitfall of both model 2 and 3 is that for the year 2020, it is unclear if the measurement is correct, as data for the next years was deemed to be inaccurate.

6 Conclusion

This thesis contributes to the growing body of literature, in which garden-level NDVI data is used to determine small-scale changes in vegetation cover. The study area is the city of Breda in the Netherlands. NDVI data on 45,000 gardens over multiple years is assessed using both conventional regression techniques, as well as spatial regression models. This is done to identify the main drivers that affect greening behavior in gardens.

It was found that the NDVI values in gardens in Breda have on average increased by 0.056. The NDVI value in first year did impact the amount of greening however, where gardens with high NDVI values in 2016 saw less greening. Socioeconomic variables, such as income and education, had minor effects on greening. Demographic variables played a larger role. The percentage of males generally had a negative effect on the amount of greening, while larger greening interventions became more likely. This suggests that gender plays a role in the type of greening, where males are more likely to undertake large greening projects. Neighborhoods with a higher proportion of older residents (aged 65+) saw fewer large-scale greening interventions. The amount of immigrants had no meaningful effect on greening. Urban characteristics played a role as well, where less urban areas saw more greening, while higher car ownership rates negatively affected the amount of greening, likely because people use garden space as a (paved) driveway.

Spatial dependency was present in both the dependent variable, as well as in the unexplained part of the models, as shown by the Lagrange Multiplier tests. For these reasons, spatial econometric models were employed, to correct for spatial autocorrelation. The spatial models did affect the results, but did not alter the conclusions of the study.

This study confirms many results found in other studies, and based on this key areas of improvement for municipal policy are identified. The findings underscore that socioeconomic and demographic characteristics can have different impacts, depending on how greening is defined. This provides deeper insight into the reasons why some gardens might become more green, while others remain stagnant, based on a data-driven analysis.

6.1 Further research

Depending on the aim of the study, it might be more sensible to use aerial photographs from winter or spring, in order to detect land cover change. This way, the effect of trees have on NDVI values in gardens is less pronounced. This study didn't allow for detection of land cover below trees. Large trees sometimes covered large areas of gardens, which made it impossible to determine the type of land cover below the trees. When for example investigating the effectiveness of the 'NK tegelwippen', tree cover hinders accurate assessment of tile cover. Furthermore, during winter and the beginning of spring, people are less likely to have put any kind of sunshade up or an inflatable swimming pool in the garden. Aerial images for the Netherlands which were taken during the spring are also provided by PDOK (n.d.) and have a higher resolution of 8cm, making them even more suitable for precise estimation of NDVI values. These 8 cm images do not have a InfraRed (CIR) layer however (PDOK, n.d.-b), therefore NDVI classification is not possible. Torres-Sánchez et al. (2014) have developed and tested alternative Vegetation Indexes however. They found that these indexes yielded around 90% accuracies in classifying vegetation in crop fields. These indexes do depend on vegetation being green however, but that could work for detecting grass cover, as this in generally green. The 8 cm resolution would allow for a more accurate accuracy assessment, as it will become easier to differentiate between different types of land cover visually.

A pitfall of this study is the use of neighborhood-level data on socioeconomic and built environment characteristics. Optimally, the garden-level NDVI data should be accompanied by garden-level data on socioeconomic and built environment characteristics. This could be done by combining survey data with NDVI data, to more accurately assess how the results from the survey relate to changes in vegetation in gardens. Due to the aggregation of the socioeconomic characteristics on neighborhood level, it was not sensible to correct for autocorrelation in the explanatory variables, as these were the same across all gardens within a neighborhood.

Literature

Anselin, L. (2001). Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference*, *97*(1), 113–139. https://doi.org/10.1016/S0378-3758(00)00349-9

Bai, Z. G., Dent, D. L., Olsson, L., & Schaepman, M. E. (2008). Proxy global assessment of land degradation. *Soil Use and Management*, *24*(3), 223–234. https://doi.org/10.1111/j.1475-2743.2008.00169.x

Balikçi, S., Giezen, M., & Arundel, R. (2022). The paradox of planning the compact and green city: analyzing land-use change in Amsterdam and Brussels. *Journal of Environmental Planning and Management*, *65*(13), 2387–2411. https://doi.org/10.1080/09640568.2021.1971069

Banay, R., Bezold, C., James, P., Hart, J., & Laden, F. (2017). Residential greenness: current perspectives on its impact on maternal health and pregnancy outcomes. *International Journal of Women's Health*, *Volume 9*, 133–144. https://doi.org/10.2147/IJWH.S125358

Bates, J., & Leibling, D. (2012). *Spaced Out: Perspectives on parking policy*. https://www.racfoundation.org/wp-content/uploads/2017/11/spaced_out-bates_leibling-jul12.pdf

Beleri, P. S., & Kotyal, K. (2024). Eco-friendly paving solutions in landscape design. In P. Vyas, D. Yadav, E. Rajput, Nasima, & D. S. Macchindra (Eds.), *Insights of Floriculture and Landscaping* (Vol. 3, pp. 119–133). Stella international Publication.

Bessembinder, J., Bintanja, R., van Dorland, R., Homan, C., Overbeek, B., Selten, F., & Siegmund, P. (2023). *KNMI'23 klimaatscenario's*. https://www.knmi.nl/kennis-en-datacentrum/achtergrond/knmi-23-klimaatscenario-s

Casey, J., James, P., Cushing, L., Jesdale, B., & Morello-Frosch, R. (2017). Race, Ethnicity, Income Concentration and 10-Year Change in Urban Greenness in the United States. *International Journal of Environmental Research and Public Health*, *14*(12), 1546. https://doi.org/10.3390/ijerph14121546

CBS. (n.d.-a). *Bevolkingsgroei: Hoe snel groeit de Nederlandse bevolking?* Retrieved June 19, 2025, from https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/bevolkingsgroei/groei

CBS. (n.d.-b). *Inwoners per gemeente*. Retrieved April 22, 2025, from https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/regionaal/inwoners

CBS. (2023). *KVK 2020*. https://www.cbs.nl/-/media/_excel/2023/35/kvk2020-wb2023.xlsx

CBS. (2025, March 27). *Kerncijfers wijken en buurten 2024*. https://www.cbs.nl/nl-nl/maatwerk/2025/13/kerncijfers-wijken-en-buurten-2024

Climates to travel. (n.d.). *Climate in Santiago (Chile)*. Retrieved June 17, 2025, from https://www.climatestotravel.com/climate/chile/santiago

Craney, T. A., & Surles, J. G. (2002). Model-Dependent Variance Inflation Factor Cutoff Values. *Quality Engineering*, *14*(3), 391–403. https://doi.org/10.1081/QEN-120001878 de Jong, R., de Bruin, S., de Wit, A., Schaepman, M. E., & Dent, D. L. (2011). Analysis of monotonic greening and browning trends from global NDVI time-series. *Remote Sensing of Environment*, *115*(2), 692–702. https://doi.org/10.1016/j.rse.2010.10.011

de Vries, S., Baliatsas, C., Verheij, R., & Dückers, M. (2025). Domestic gardens and morbidity: Associations between private green space and diagnosed health conditions in the Netherlands. *Environment International*, *199*, 109450. https://doi.org/10.1016/j.envint.2025.109450

de Vries, S., Buijs, A. E., & Snep, R. P. H. (2020). Environmental Justice in The Netherlands: Presence and Quality of Greenspace Differ by Socioeconomic Status of Neighbourhoods. *Sustainability*, *12*(15), 5889. https://doi.org/10.3390/su12155889

De Vries, S., Kramer, H., Smits, M., Spijker, J., Verheij, R., Dückers, M., Baliatsas, C., & Wiegersma, S. (2023). *Een groene tuin, een gezonde tuin? Onderzoek naar het belang van privégroen bij huis voor de gezondheid van burgers*. https://edepot.wur.nl/586989

Duncan, D. T., Kawachi, I., White, K., & Williams, D. R. (2013). The Geography of Recreational Open Space: Influence of Neighborhood Racial Composition and Neighborhood Poverty. *Journal of Urban Health*, *90*(4), 618–631. https://doi.org/10.1007/s11524-012-9770-y

Fruth, E., Kvistad, M., Marshall, J., Pfeifer, L., Rau, L., Sagebiel, J., Soto, D., Tarpey, J., Weir, J., & Winiarski, B. (2019). Economic valuation of street-level urban greening: A case study from an evolving mixed-use area in Berlin. *Land Use Policy*, *89*, 104237. https://doi.org/10.1016/j.landusepol.2019.104237

Gandhi, G. M., Parthiban, S., Thummalu, N., & Christy, A. (2015). Ndvi: Vegetation Change Detection Using Remote Sensing and Gis – A Case Study of Vellore District. *Procedia Computer Science*, *57*, 1199–1210. https://doi.org/10.1016/j.procs.2015.07.415

Gemeente Breda. (n.d.). *Duurzaam Breda*. Retrieved April 24, 2025, from https://www.breda.nl/duurzaam

Gemeente Breda. (2025). *Subsidie voor duurzaam omgaan met regen*. https://www.breda.nl/subsidie-voor-duurzaam-omgaan-met-regen

Gerrish, E., & Watkins, S. L. (2018). The relationship between urban forests and income: A meta-analysis. *Landscape and Urban Planning*, *170*, 293–308. https://doi.org/10.1016/j.landurbplan.2017.09.005

Grampp, C. (1990). *The meaning of gardens* (M. Francis & R. T. Hester, Eds.). MIT Press.

Gu, T., Han, C., Zhao, M., Zhang, L., Yang, Z., & Zhang, J. (2024). Is there any difference between the owners and the public in their visual impact assessments?——A case study of the front garden of multi-storey residential buildings. *PLOS ONE*, *19*(1), 15–16. https://doi.org/10.1371/journal.pone.0296519

Guevara, B. R., Uribe, S. V., de la Maza, C. L., & Villaseñor, N. R. (2024). Socioeconomic Disparities in Urban Forest Diversity and Structure in Green Areas of Santiago de Chile. *Plants*, *13*(13), 1841. https://doi.org/10.3390/plants13131841

Harrell, F. E. (2015). Regression Modeling Strategies. In *Regression Modeling Strategies* (2nd ed., pp. 78–79). Springer. https://link-springer-com.vu-nl.idm.oclc.org/book/10.1007/978-3-319-19425-7#overview

Heidt, V., & Neef, M. (2008). Benefits of Urban Green Space for Improving Urban Climate. In *Ecology, Planning, and Management of Urban Forests* (pp. 84–96). Springer New York. https://doi.org/10.1007/978-0-387-71425-7_6

Hemmert, G. A. J., Schons, L. M., Wieseke, J., & Schimmelpfennig, H. (2018). Log-likelihood-based Pseudo- R^2 in Logistic Regression. *Sociological Methods & Research*, *47*(3), 507–531. https://doi.org/10.1177/0049124116638107

Hunter, M. C. R., & Brown, D. G. (2012). Spatial contagion: Gardening along the street in residential neighborhoods. *Landscape and Urban Planning*, *105*(4), 407–416. https://doi.org/10.1016/j.landurbplan.2012.01.013

Hussain, M. R. M., Tukiman, I., Zen, I. Hj., & Shahli, F. M. (2014). The Impact of Landscape Design on House Prices and Values in Residential Development in Urban Areas. *APCBEE Procedia*, *10*, 316–320. https://doi.org/10.1016/j.apcbee.2014.10.059

James, P., Banay, R. F., Hart, J. E., & Laden, F. (2015). A Review of the Health Benefits of Greenness. *Current Epidemiology Reports*, *2*(2), 131–142. https://doi.org/10.1007/s40471-015-0043-7

Ju, Y., Moran, M., Wang, X., Avila-Palencia, I., Cortinez-O'Ryan, A., Moore, K., Slovic, A. D., Sarmiento, O. L., Gouveia, N., Caiaffa, W. T., Aguilar, G. A. S., Sales, D. M., De Pina, M. D. F. R. P., Coelho, D. M., & Dronova, I. (2021). Latin American cities with higher socioeconomic status are greening from a lower baseline: evidence from the SALURBAL project. *Environmental Research Letters*, *16*(10). https://doi.org/10.1088/1748-9326/ac2a63

Keith, T. Z. (2019). *Multiple Regression and Beyond*. Routledge. https://doi.org/10.4324/9781315162348

Knook, O. (2024, November 13). Breda eindigt op vierde plek bij NK Tegelwippen, ruim 320.000 tegels gewipt. *BredaVandaag*. https://www.bredavandaag.nl/algemeen/algemeen/432638/breda-eindigt-op-vierde-plek-bij-nk-tegelwippen-ruim-320-000-

Langemeyer, J., Camps-Calvet, M., Calvet-Mir, L., Barthel, S., & Gómez-Baggethun, E. (2018). Stewardship of urban ecosystem services: understanding the value(s) of urban gardens in Barcelona. *Landscape and Urban Planning*, *170*, 79–89. https://doi.org/10.1016/j.landurbplan.2017.09.013

Lenney, M. P., Woodcock, C. E., Collins, J. B., & Hamdi, H. (1996). The status of agricultural lands in Egypt: The use of multitemporal NDVI features derived from landsat TM. *Remote Sensing of Environment*, *56*(1), 8–20. https://doi.org/10.1016/0034-4257(95)00152-2

LeSage, J. P. (2008). An Introduction to Spatial Econometrics. *Revue d'économie Industrielle*, 19–44. https://doi.org/10.4000/rei.3887

LeSage, J. P., & Pace, R. K. (2008). SPATIAL ECONOMETRIC MODELING OF ORIGIN-DESTINATION FLOWS*. *Journal of Regional Science*, *48*(5), 941–967. https://doi.org/10.1111/j.1467-9787.2008.00573.x

Marcelis, H. (2021). Breda stelt 100.000 euro extra beschikbaar voor groene daken en onttegelen. *Breda Vandaag*.

https://www.bredavandaag.nl/algemeen/algemeen/345175/breda-stelt-100-000-euro-extra-beschikbaar-voor-groene-daken-en

Mason, C. F. (2000). Thrushes now largely restricted to the built environment in eastern England. *Diversity and Distributions*, *6*(4), 189–194. https://doi.org/10.1046/j.1472-4642.2000.00084.x

Mulder, F. (2020, May 6). Een tuin is een aanval. *De Groene Amsterdammer No 19*. https://www.groene.nl/artikel/een-tuin-is-een-aanval?utm_source=De%20Groene%20Amsterdammer&utm_campaign=bf2a323c39-Dagelijks-2020-05-07&utm_medium=email&utm_term=0_853cea572a-bf2a323c39-70899937&fbclid=IwAR1rc0IgOUySfjpDrWRVnneRBGa7wPrlpAI8RFQVWdkz4L6jBlxh45vM6iQ

Murtagh, N., & Frost, R. (2023). Motivations for urban front gardening: A quantitative analysis. *Landscape and Urban Planning*, *238*, 104835. https://doi.org/10.1016/j.landurbplan.2023.104835

Ni, L., & Zhang, D. (2024). Using a spatial autoregressive model with spatial autoregressive disturbances to investigate origin-destination trip flows. *PLOS ONE*, *19*(6), e0305932. https://doi.org/10.1371/journal.pone.0305932

NK tegelwippen. (2024). *NK tegelwippen*. https://www.nk-tegelwippen.nl/

NK-tegelwippen. (n.d.). *Veelgestelde vragen*. Retrieved April 10, 2025, from https://www.nk-tegelwippen.nl/veelgestelde-vragen/

PBL. (2023). *Natuur in en om de stad*. https://www.pbl.nl/uploads/default/downloads/pbl-2023-natuur-in-en-om-de-stad-5023.pdf

PDOK. (n.d.-a). *Dataset: Kadastrale Kaart*. Retrieved April 26, 2025, from https://www.pdok.nl/introductie/-/article/kadastrale-kaart

PDOK. (n.d.-b). *Dataset: PDOK Luchtfoto RGB (Open)*. Retrieved April 24, 2025, from https://www.pdok.nl/introductie/-/article/pdok-luchtfoto-rgb-open-

PDOK. (2022). *Nieuwe service voor CBS Bodemgebruik 2017*. https://www.pdok.nl/-/nieuwe-service-voor-cbs-bodemgebruik-2017

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. Chr. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, *20*(9), 503–510. https://doi.org/10.1016/j.tree.2005.05.011

Pisman, A., Vanacker, S., Bieseman, H., Vanongeval, L., Van Steertegem, M., Poelman, L., & Van Dyck, K. (2021). *Ruimterapport 2021*. https://www.vlaanderen.be/publicaties/ruimterapport-vlaanderen-rura-een-ruimtelijke-analyse-van-vlaanderen

Sander, H., Polasky, S., & Haight, R. G. (2010). The value of urban tree cover: A hedonic property price model in Ramsey and Dakota Counties, Minnesota, USA. *Ecological Economics*, *69*(8), 1646–1656. https://doi.org/10.1016/j.ecolecon.2010.03.011

Saphores, J.-D., & Li, W. (2012). Estimating the value of urban green areas: A hedonic pricing analysis of the single family housing market in Los Angeles, CA. *Landscape and Urban Planning*, *104*(3–4), 373–387. https://doi.org/10.1016/j.landurbplan.2011.11.012

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Sha, A., Liu, Z., Jiang, W., Qi, L., Hu, L., Jiao, W., & Barbieri, D. M. (2021). Advances and development trends in eco-friendly pavements. *Journal of Road Engineering*, *1*, 1–42. https://doi.org/10.1016/j.jreng.2021.12.002

Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39–42. https://doi.org/10.12691/ajams-8-2-1

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica*, 12–18. https://doi.org/10.11613/BM.2014.003

Stobbelaar, D. J., van der Knaap, W., & Spijker, J. (2021). Greening the City: How to Get Rid of Garden Pavement! The 'Steenbreek' Program as a Dutch Example. *Sustainability*, *13*(6), 3117. https://doi.org/10.3390/su13063117

Torres-Sánchez, J., Peña, J. M., de Castro, A. I., & López-Granados, F. (2014). Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. *Computers and Electronics in Agriculture*, *103*, 104–113. https://doi.org/10.1016/j.compag.2014.02.009

Troy, A. R., Grove, J. M., O'Neil-Dunne, J. P. M., Pickett, S. T. A., & Cadenasso, M. L. (2007). Predicting Opportunities for Greening and Patterns of Vegetation on Private Urban Lands. *Environmental Management*, *40*(3), 394–412. https://doi.org/10.1007/s00267-006-0112-2

UN. (2022). *Day of 8 billion*. https://www.un.org/en/dayof8billion

van Heezik, Y. M., Dickinson, K. J. M., & Freeman, C. (2012). Closing the Gap: Communicating to Change Gardening Practices in Support of Native Biodiversity in Urban Private Gardens. *Ecology and Society*, *17*(1), art34. https://doi.org/10.5751/ES-04712-170134

VRO. (2023, October 13). *Veel potentie om aan de rand van stad en dorp een straatje bij te bouwen*. https://www.volkshuisvestingnederland.nl/onderwerpen/woondeals/nieuws/2023/10/13/veel-potentie-om-aan-de-rand-van-stad-en-dorp-een-straatje-bij-te-bouwen

Warhurst, J. R., Parks, K. E., McCulloch, L., & Hudson, M. D. (2014). Front gardens to car parks: Changes in garden permeability and effects on flood regulation. *Science of The Total Environment*, *485–486*, 329–339. https://doi.org/10.1016/j.scitotenv.2014.03.035

Yeager, R., Browning, M. H. E. M., Breyer, E., Ossola, A., Larson, L. R., Riggs, D. W., Rigolon, A., Chandler, C., Fleischer, D., Keith, R., Walker, K., Hart, J. L., Smith, T., & Bhatnagar, A. (2023). Greenness and equity: Complex connections between intra-neighborhood contexts and residential tree planting implementation. *Environment International*, *176*, 107955. https://doi.org/10.1016/j.envint.2023.107955

Young, D. S. (2018). *Handbook of Regression Methods*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315154701

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Appendix

Appendix A

 

Figure A1. Gardens with an NDVI value of precisely 0.1. Gardens are visually clearly partly paved, partly green. The left top picture shows, similar to the bottom left picture, a garden with a terrace and some vegetation in the form of bushes or trees. The upper right picture shows a garden with a terrace, pool and driveway, along with a large green lawn, small trees and a hedge. The bottom picture shows a garden that is partly covered by trees. It is therefore unclear what land cover under the tree looks like, but the garden is still assigned an NDVI value of 0.1. The NDVI-score for each garden was calculated as an average for each garden. Therefore some parts may be very green, yet the NDVI score remains low due to the simultaneous presence of paved areas.

Appendix B

Figure B1. Redevelopment of an old estate. Left side of the picture is the situation in 2016, while the right side of the picture in the situation in 2020. As visible on the picture, vegetation was (temporarily) replaced by bare sand.

Figure B2. Some of the vegetation had returned in 2023.

Appendix C

				Accuracy assessment ∆NDVI > 0.05			**Reference class**		
	More green			Equally green			Less green		
User class	2016-2017	2017-2018	2018-2020	2016-2017	2017-2018	2018-2020	2016-2017	2017-2018	2018-2020
More green	8	7	10	0	0	0	0	0	0
Equally green	32	33	30	40	38	40	12	14	9
Less green	0	0	0	0	2	0	28	26	31
Producer accuracy	*0.2*	*0.175*	*0.25*	*1*	*0.95*	*1*	0.7	0.65	0.775
	2016-2017	2017-2018	2018-2020		More green	Equally green	Less green		
Overall accuracy	0.6333333	0.591667	0.675		21%	99%	71%		

Table C1. accuracy assessment with an NDVI threshold value of 0.05. Gardens that had classified more green are particularly poorly classified. 2020 to 2022 was excluded since the results in the other accuracy assessment were already below standard, therefore this assessment would have the same conclusion due to the even smaller threshold of 0.05.

Appendix D

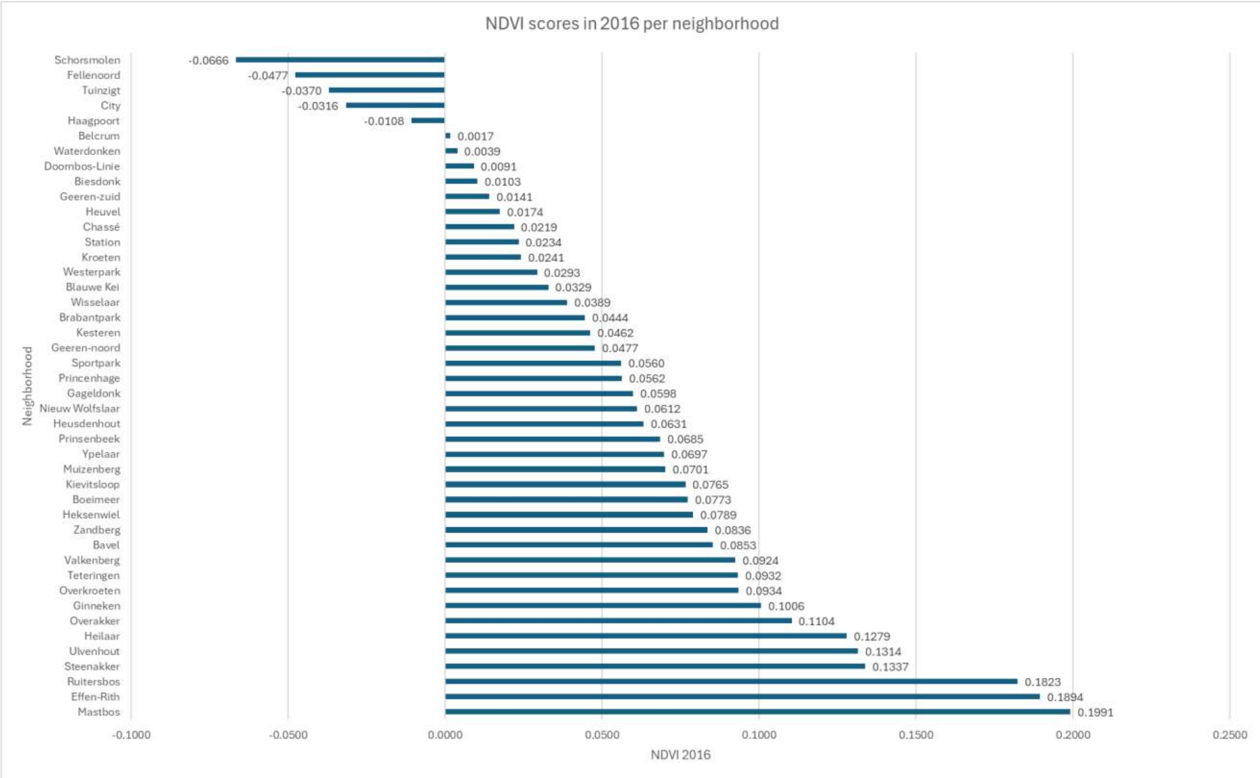

Figure D1. NDVI scores per neighborhood in 2016

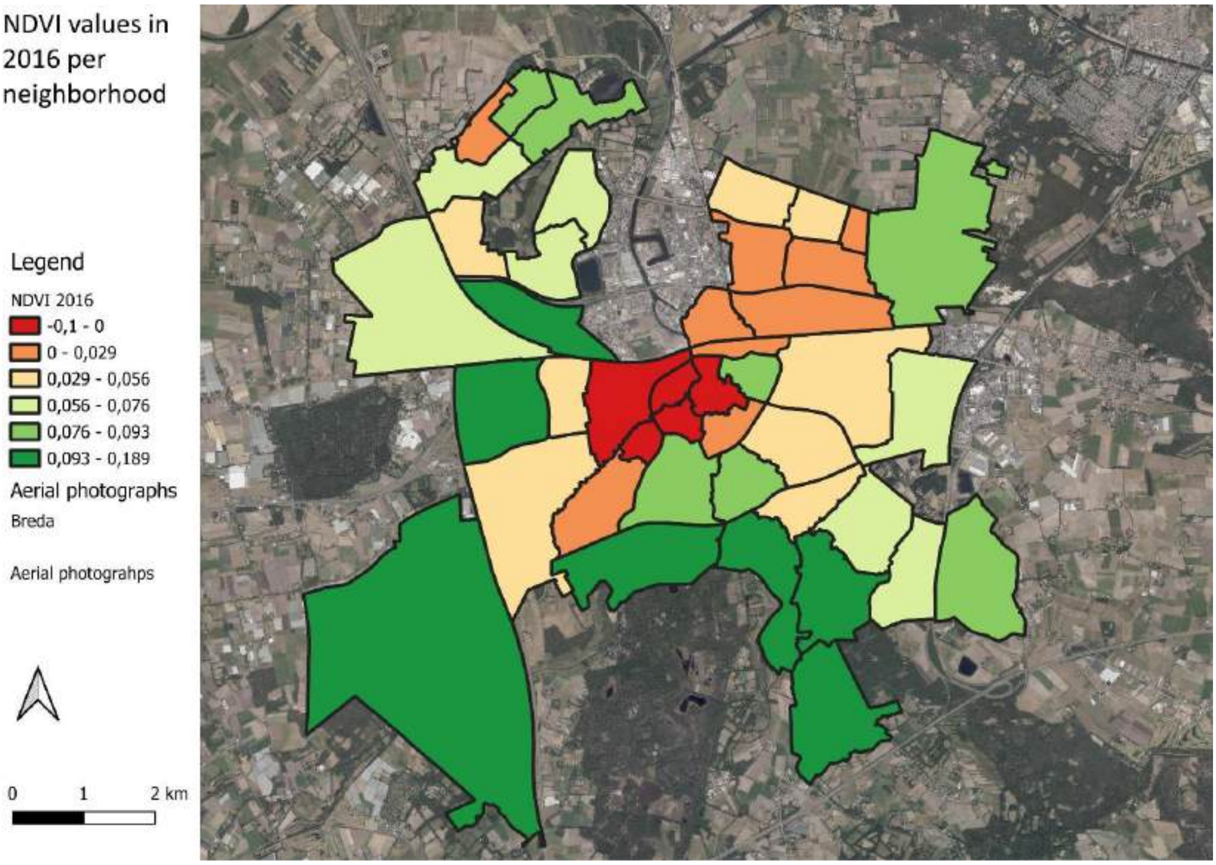

Figure D2. NDVI values per neighborhood in 2016 mapped

Appendix E

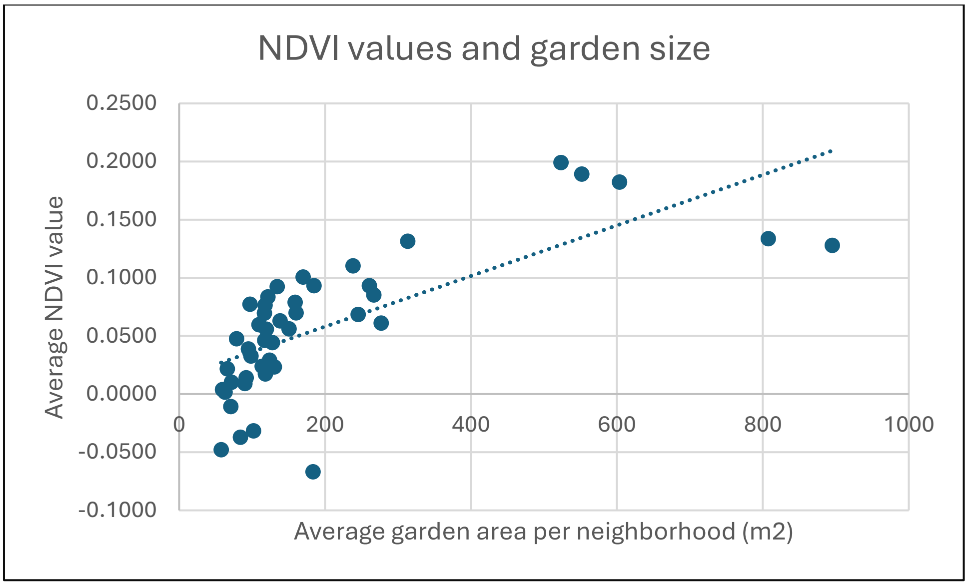

Figure E1. NDVI values in 2016 is positively associated with the average garden size in a neighborhood, indicating that larger gardens are more green.

Appendix F

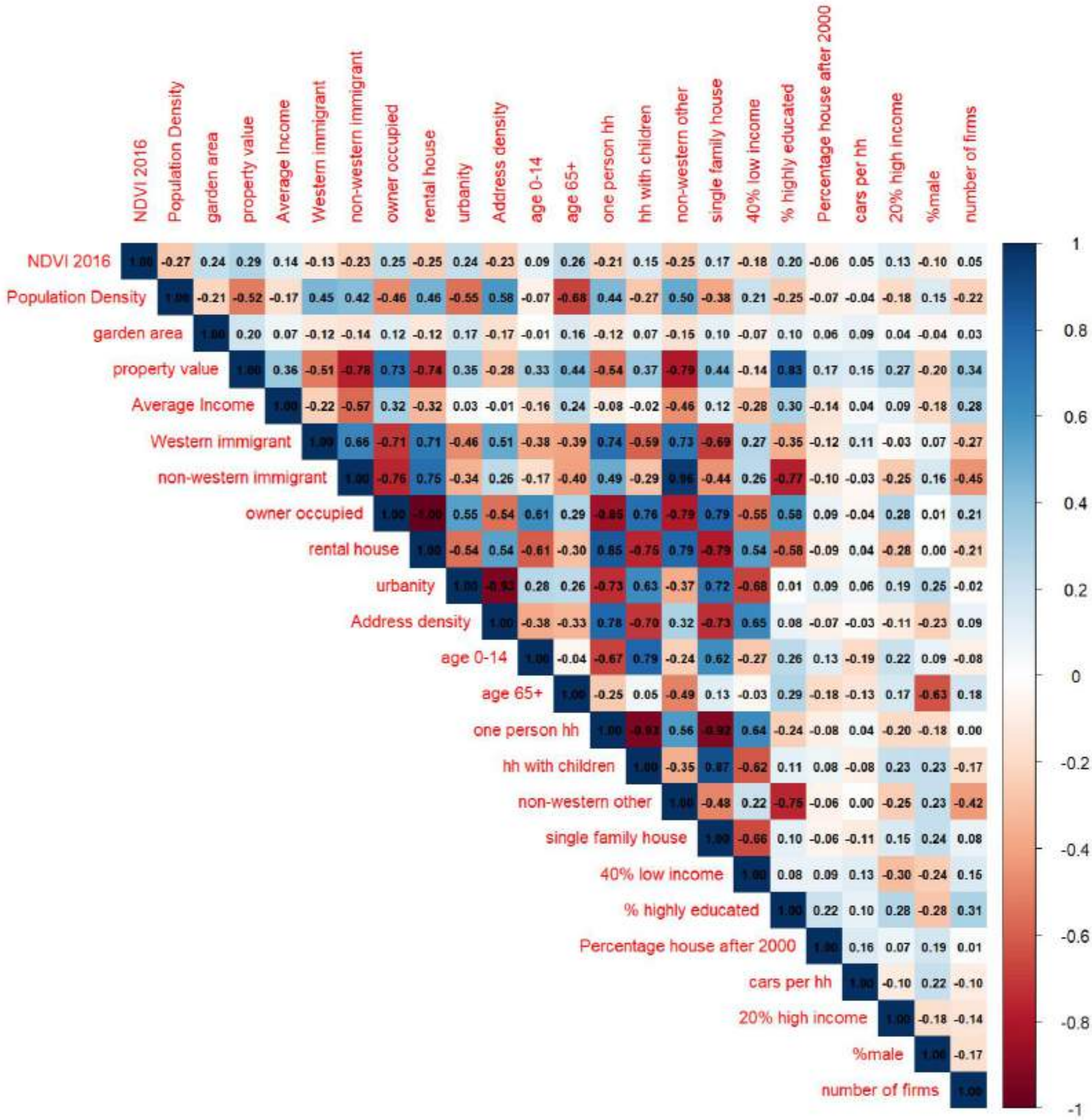

Figure F1. Pearson Correlation Coefficients for a large number of variables