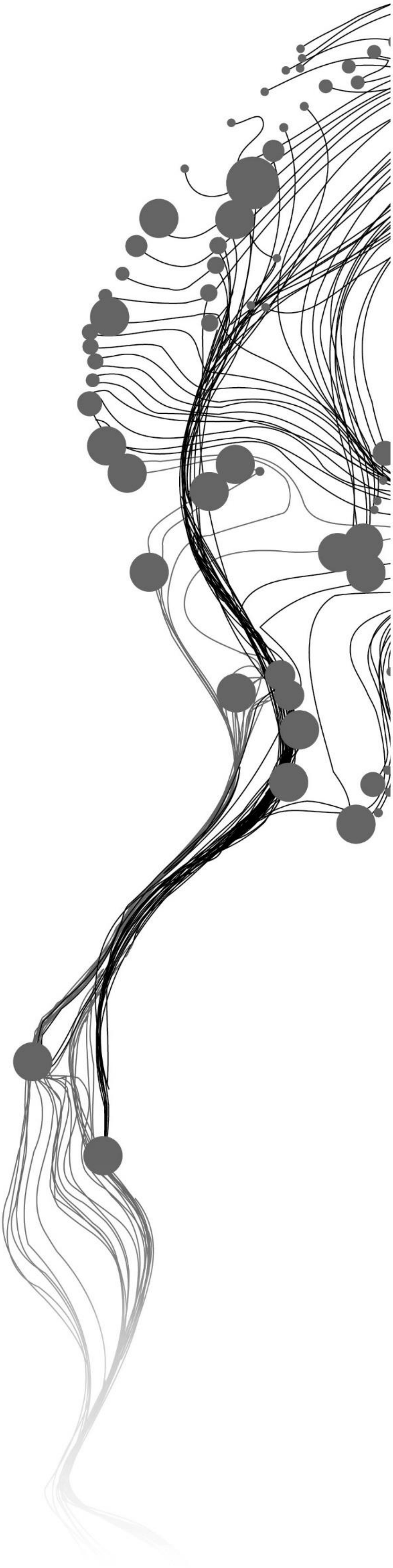
THE IMPACT OF DIFFERENT TYPES OF URBAN GREEN ENVIRONMENTS ON PROPERTY VALUE (ALKMAAR, NETHERLANDS)

MAJEDEH FALLAHIANBIZHAN
July, 2024

SUPERVISORS:
Dr. M.N. Koeva
Prof.dr. M. Kuffer

ADVISOR:
Dr. E. Simao Da Graca Dias



THE IMPACT OF DIFFERENT TYPES OF URBAN GREEN ENVIRONMENTS ON PROPERTY VALUE (ALKMAAR, NETHERLANDS)

MAJEDEH FALLAHIANBIZHAN
Enschede, The Netherlands, July 2024

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Urban Planning and Management

SUPERVISORS:
Dr. M.N. Koeva
Prof.dr. M. Kuffer

ADVISOR:
Dr. E. Simao Da Graca Dias

THESIS ASSESSMENT BOARD:
Prof. Dr. K. Pfeffer (Chair)
Marion Plegt, MSc. (External Examiner, Kadaster)

ABSTRACT

The rate of urbanization and population growth is increasing rapidly worldwide. It is estimated that a significant portion of the world's population will live in urban areas by 2050. According to this phenomenon, the demand for properties has increased in cities to accommodate the growing population that pressures the property market. Simultaneously, various factors should be considered when estimating the value of properties. This research aims to investigate the effects of different types of green environments on property value.

The factors affecting the property value are categorized into three groups: locational, physical, and environmental characteristics of the property. Among different categories affecting property value, one of the most critical factors is the distance to green environments. However, there are different types of green environments with specific characteristics across the cities that are not in similar conditions. Thus, the main research problem lies in understanding how and to what extent different types of green environments, along with other factors, affect property value. This research provides empirical evidence on how different types of urban green environments influence property value, helping urban planners and policymakers make informed urban development decisions.

Regarding the methodology, this research elaborated on the quantitative property valuation method to find the effects of different types of green environments on property values. First, urban green environments are classified based on size, height, density, type of vegetation, and services they provide. Then, property value prediction models are constructed by combining two-dimensional (2D) factors, for instance, the size of the property and distance to CBD, with three-dimensional (3D) factors, such as property visibility and orientation. For 2D data, three methods, Random Forest (RF), Ordinary Least Square (OLS), and Geographically Weighted Regression (GWR), are applied. In addition, for the 3D data, the OLS method is executed for modelling.

Comparing the results of applying the OLS, GWR, and RF methods illustrates that the RF explains 83.1% of the property value variation based on the adjusted R-square value, which is higher than the OLS and GWR. Hence, RF is the most suitable method to predict the property value of Alkmaar. The RF illustrates that there is a non-linear correlation between the property value and different types of green environments. For instance, the size of the green environment factor is the most important factor in the model. When the size of green environments in a distance of 25m around the properties increases from 600m^2 to 800m^2, the property value decreases significantly.

The OLS model of property value by the 3D factors has the adjusted R square value of 0.169, meaning that the model explains the 16.9% of the property value variation. The vegetation in front of the building and the view of vegetation variables are two important factors that positively correlate with the property value. The importance of different types of urban green environments is different in each of the RF and OLS models. The RF and OLS models are also validated by k-fold cross-validation using the actual value and predicted propety value by models. The percentage error of the RF model was 12.1%, while the percentage error of the OLS model was 17.01%, indicating that the developed model with 2D data is more accurate than the model with only 3D data.

ACKNOWLEDGEMENTS

I am grateful to all the people who have supported and guided me throughout this journey. First and foremost, I would like to express my sincere gratitude to my supervisors, Dr. Mila Koeva and Dr. Monika Kuffer, for their guidance, support, and constructive suggestions for my MSc research thesis. Each discussion with them was fruitful and motivated me a lot to dive into the issues. I also would like to express my appreciation to my advisor, Dr. Eduardo Simao Da Graca Dias, who gave me warm help and guidance from the start to the end. Without their guidance, this research could not be completed.

Second, I would love to thank my beloved parents, *Madar* and *Pedar*, for their endless understanding, support, and love during the different steps of my life. Their sacrifices and encouragement have been the foundation of my success.

Third, thanks to my lovely and wonderful sisters, *Mae* and *Mohan,* who were always there for me and kept me company. Their constant encouragement has inspired me to strive for excellence.

In the end, my best love, *Amir*. He has been my rock, providing me with the strength and motivation to overcome every challenge. I am grateful to *Amir* for believing in me even when I doubted myself. His endless support and patience paved the way to success for me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIC	Akaike information criterion
ANN	Artificial Neural Network
ANOVA	Analysis of variance
CNN	Convolutional Neural Network
DFM	Digital Feature Model
DSM	Digital Surface Map
DTM	Digital Terrain Model
GE	Green Environments
GEE	Google Earth Engine
GWR	Geographically Weighted Regression
HPM	Hedonic Price Model
k-NN	K-nearest neighbour algorithm
LOD	Level of Details
MAE	Mean absolute error
MDI	Mean Decrease in Impurity
MRT	Mass Rapid Transit
NDVI	Normalized difference vegetation index
OLS	Ordinal Least Squares
RF	Random Forest
RMSE	Root Mean Square Error
R^2	R Square
SAR	Spatial Regression Model
SSR	Residual Sum of Square
SVF	Sky View Factors
SVM	Support Vector Machine
SVR	Support Vector Regression
URT	Urban Rail Transit
VIF	Variance Inflation Factor
VI	Vegetation Index

1. INTRODUCTION

1.1. Background and justification

Urbanization rates are increasing rapidly worldwide, and it is estimated that 70% of the world's population will live in urban areas by 2050 (United Nations, 2022). As population growth and urbanization continue, cities face several challenges, such as pressure on housing markets (Nijskens et al., 2019). Pressure on housing markets is addressed from two points of view. This issue is examined from a financial market stability or social and economic perspective, which considers housing affordability, city competitiveness, and social segregation (Nijskens et al., 2019). Following the migration of new people to cities, housing demand will also change, which affects the property value (Wang et al., 2017).

The factors affecting property values are categorized into three groups: structural, locational, and environmental (Wittowsky et al., 2020). Structural factors are the main characteristics of the property, including the size and age of the property. In contrast, locational factors refer to the geographical location of property and proximity to urban facilities, consisting of distance to CDB, workplaces, transportation, schools, and other facilities (Sirmans et al., 2005; Wittowsky et al., 2020). Environmental factors refer to the surrounding landscape near the property, such as water and green spaces, urban green vegetation, and environmental quality (Sirmans et al., 2005; Wittowsky et al., 2020). Even though all these factors are mainly two-dimensional (2D), there are three-dimensional (3D) factors to consider in property values (Hui et al., 2012; Ying et al., 2021). Within the development of 3D modelling in recent years, more researchers found 3D factors important in affecting property values. Ying et al. (2021) explored the sky view factor, property orientation, and sunlight are essential 3D factors affecting property values.

Among all the common factors affecting property values, distance to urban parks and green areas is one of the most critical factors (Ludwig et al., 2021; Panduro & Veie, 2013). Panduro & Veie (2013) and Wolch et al. (2014) found that urban green environments improve public health, reduce air pollution, and decrease city temperatures. Parks and natural areas may also allow people to do outdoor activities, increasing social integration and interaction among people (Peschardt et al., 2012). Consequently, urban residents are more inclined to reside near green spaces. Additionally, the real estate market in developed countries demonstrated that people are willing to pay more for properties near green spaces (Jim & Chen, 2010; Luttik, 2000). However, green spaces are not equally distributed in cities, and there is heterogeneity in different types of green environments (Panduro & Veie, 2013). As a result, the distance of properties to green areas positively correlated with the property value (Liebelt et al., 2019; Piaggio, 2021).

There are also a few studies showing that proximity to green areas has adverse effects on property value. Troy & Grove (2008) found that proximity to parks negatively affects property value in cities with a high rate of robbery and crime. Green spaces with poor management and maintenance, for example, in terms of standard of upkeep and quality, may negatively impact local property value (Troy & Grove, 2008). Furthermore, the property will be exposed to the risk of wildfire, which will negatively affect property value (Donovan et al., 2007). Therefore, the relationship between urban green environments and property value is context-dependent. Based on the reviewed literature, Figure 1 indicates the positive and negative impacts of green environments.

In conclusion, various factors affect property value, and distance to green spaces is one of the most critical factors. However, green spaces are not equally distributed in cities, and there is heterogeneity in different types of green environments (Panduro & Veie, 2013).

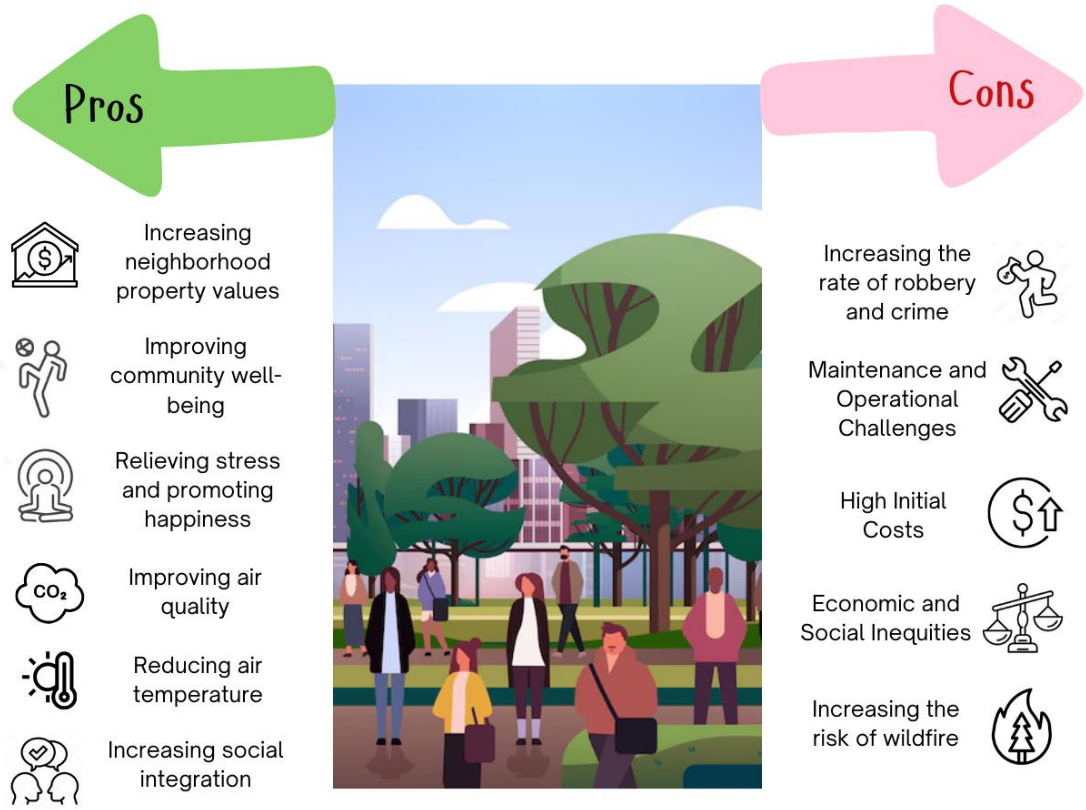


Figure 1: The positive and negative impacts of green environments based on the reviewed literature

Source: Author, 2024.

1.2. Research problem

Following the increased housing demand in the Netherlands, the housing market planned to accelerate housing construction in urban areas with high demand to provide for their needs (Nijskens et al., 2019). In this situation, urban planners and city governments should consider different factors affecting the value of properties to estimate the value of new construction. One of the most critical factors among different categories affecting property value is parks and green spaces (Wittowsky et al., 2020). However, there are different types of green spaces with specific characteristics that are not in similar conditions, especially in the cities in the Netherlands, which are mostly green and have different types of green environments. Thus, the main research problem is how different types of green environments and other essential variables affect property value to predict the value of future construction properties precisely.

1.3. Research gap

Most of the existing literature on the impact of green environments on property value considers all the green spaces of the same type, such as parks in scales of neighbourhood parks and urban parks (Panduro & Veie, 2013). However, there is limited research on how different types of green environments, such as dense or scattered green spaces, affect property value. Moreover, the measurement and definition of green environments in different research varies, which makes it complex to compare the results of the effects of green environments on property value (Waltert & Schläpfer, 2010).

Green environments exist in heterogeneity in a wide variety of types around the cities (Morano et al., 2019). Individuals perceive green environments as heterogeneous, varying significantly in quality and quantity. For instance, they distinguish between different types of green spaces and value each type based on the services provided by green spaces (Peschardt et al., 2012). These types are from small lawns and grasses to large parks and urban forests.

Different classification approaches lead to numerous classes of green spaces. Various studies have divided green spaces according to their size, services, and distance to property (Chen et al., 2023; Dell'Anna et al., 2022). Chen et al. (2023) suggested that the main characteristics of green space should be considered to find the impact of green spaces on property value. They found that all sorts of green spaces, including parks, private golf courses, or gardens, do not have the equivalent positive effect on property value.

Even though some studies focused on types of green spaces, the definition of each type and their approach to categorizing green spaces need to be clearly defined. In addition, they have not considered which type of green environment has the highest correlation with property value. This research classifies green environments based on a selected approach. Then, it finds the relationship between each type and property value to fill in the gap in the research on the effects of different types of green environments on property value. This research aims to help urban planners and policymakers understand the importance of different urban factors affecting property value for new construction. Moreover, this study provides a scientific basis for planners and policymakers involved in urban green environments and concerned about greening cities and housing affordability.

1.4. Research objectives and questions

1.4.1. General objective

This research aims to analyse the impact of different types of urban green environments on property value in the city of Alkmaar.

1.4.2. Sub-objective and research question

Sub-objective 1: To identify the 2D and 3D factors relating to urban green environments and property value by literature review.

Q1: What are common factors that affect property value based on a literature review?
Q2: What are the different types of urban green environments?
Q3: What is a suitable classification approach for urban green environments?

Sub-objective 2: To classify different types of urban green environments affecting property value.

Q1: What method is suitable for classifying different types of urban green environments?
Q2: What are the different types of urban green environments in the study area?
Q3: What is the geographical pattern of each type of green environment in the study area?

Sub-objective 3: To develop a model to estimate property value based on the combination of different types of urban green and other common factors affecting property value.

Q1: What are common methods used for modelling property value?
Q2: What method is suitable for developing a model for analyzing the impact of the different types of urban green environments on property value?

Q3: What is the relationship between each type of urban green environment and property value in the developed model?
Q4: How effective is the model to generalize property value prediction?

1.5. Conceptual framework

Figure 2 shows the conceptual framework of the research. Different 2D and 3D factors are independent variables affecting the property value. However, there are different types of green environments that can be categorized by a specific classification approach that might affect property value. This research combines the 2D and 3D factors with types of green environments and finds how they affect property value as the dependent variable. Meanwhile, this process is the basis for the development of a property value prediction model.

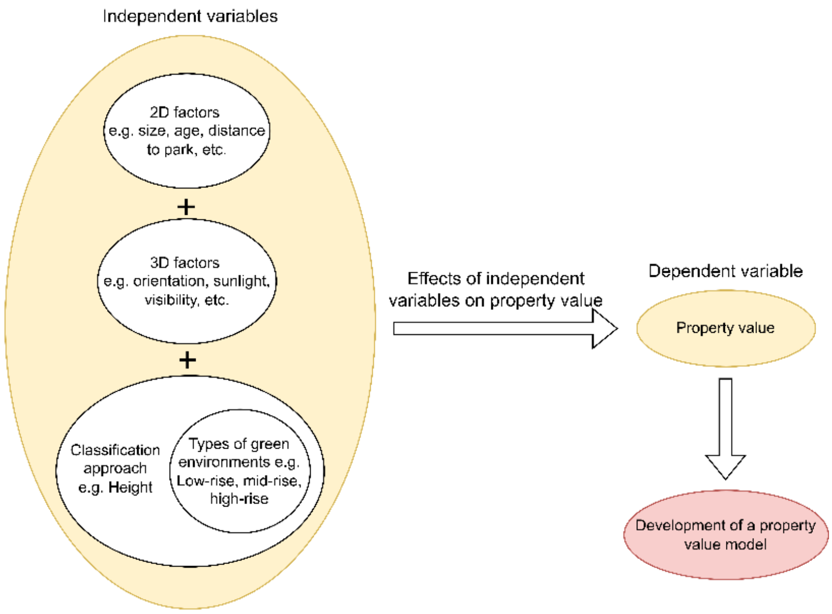


Figure 2: Conceptual framework

Source: Author, 2024.

1.6. Summary

First, the background and justification of the research, research problem, and gap are explained in this chapter. Afterwards, based on the research problem, the research's main aim and questions regarding each sub-objective are mentioned. In the end, the conceptual framework visualizing the expected relationship between the variables is constructed.

2. LITERATURE REVIEW

2.1. General factors affecting property value

In property value literature, the factors affecting property value are mainly categorized into three groups: structural, locational, and environmental factors (Sirmans et al., 2005; Wittowsky et al., 2020; Wu et al., 2015). The factors in each of these categories have specific effects on property value. In the following, several literature and studies are reviewed to find the effects of various factors in each category in detail.

A variety of factors affecting property value are in the locational classification group. Liang et al. (2018) chose 13 locational factors in their research to investigate their effects on property value at the community level. This research shows that proximity to locational facilities such as transport systems, medical facilities, educational facilities, and commercial facilities substantially affects the real estate market. Another study in Hangzhou, China, also illustrated that property value improves when the property is surrounded by kindergartens, high schools, and college institutions (Wen et al., 2014). They explored that adding one kindergarten located one kilometer from the neighbourhood raises the cost of housing by 0.300%.

Wittowsky et al. (2020) investigated the factors affecting property value in the City of Dortmund, Germany. They noticed that the presence of amenities and public transportation stops are insufficient to explore their effects on property value. At the same time, accessibility to these services is notable and should be considered. As a result of this research showed that proximity to amenities, especially restaurants and parks, are two crucial locational factors, and dwelling characteristics such as living area, plot size, number of floors, and house condition are also essential physical characteristics affecting property value.

The literature shows correlations between distance to transportation infrastructure and property value. In Athens, Greece, proximity to different transportation infrastructures directly impacts property value (Efthymiou & Antoniou, 2013). However, based on the type of transportation system, this correlation is positive or negative in this city. For instance, Efthymiou & Antoniou (2013) determined that metro, tram, suburban railway, and bus stations positively affect property value, while national rail stations, airports, and ports negatively affect property value due to noise issues. On the other hand, Yang et al. (2020) discovered that Urban Rail Transit (URT) positively impacts nearby property value. URT aims to mitigate various urban problems (e.g., traffic congestion) and improve transportation mobility opportunities in adjacent areas, raising the value of properties there.

Another considerable factor affecting property value in the category of locational factors is the distance to the Central Business District (CBD). Wu et al. (2015) illustrated that properties located in the proximity of CBD have better accessibility to a wide range of amenities and services, including shopping centres, working places, and public transportation, which will positively affect property value. Moreover, Cao et al. (2019) investigated the spatial variation of housing value in Singapore. To conclude, they found that age and the floor area of the housing units are two substantial physical factors affecting housing value. On the other hand, the distance to the nearest park, the distance to CBD, and the distance to the nearest Mass Rapid Transit (MRT) station are important locational factors in this process.

The natural urban landscape, from the category of environmental factors, positively affects property value by improving the city's quality and providing insightful leisure sites for people. Wen et al. (2017) researched on the effects of the Grand Canal landscape on property value in Hangzhou, China. The results revealed that a 1% increase in the property distance from the canal would result in a 0.0016% decrease in property value. People are more willing to pay for houses near the natural landscape to live there and benefit from their advantages.

Air pollution, in the category of environmental factors, is also one of the main concerns in cities, especially large cities. Degraded air quality has adverse effects on human health. Saptutyningsih et al. (2013) investigated the effects of air quality on property value. They explored the fact that the concentration of air pollutants (e.g., CO and NO_2) is negatively correlated with property value. The value of properties located in areas with high levels of air pollution is lower than in areas with less air pollution. Moreover, along with the distance to parks, the Vegetation Index is another variable in the environmental category. Zambrano-Monserrate et al. (2021) examined the effects of the interaction between distance to green spaces and the Normalized Difference Vegetation Index (NDVI) of green spaces with property value and explored that NDVI positively contributes to the property value.

According to the reviewed literature, 3D urban factors, such as the sky view factor (SVF), property orientation, height, and sunlight, contribute to property value. Due to the impact of property orientation on natural light and energy efficiency, it is a significant factor affecting property value. Wu et al. (2015) investigated the correlation between property orientation and property value. They discovered that properties with a south-facing orientation are more expensive than other orientations since they receive more natural light than others. Moreover, Fleming et al. (2018) indicated that people are willing to pay for houses receiving more sunlight yearly. They found that each extra daily hour of sunlight exposure is associated with a 2.6% increase in house sale prices in Wellington, New Zealand.
In addition, the SVF is an urban geometry 3D factor that has not been studied in much of the literature on property value. This factor has been widely used recently in relation to air temperature and urban heat islands (Zheng & Li, 2022). SVF measures the openness of the sky, which is mainly influenced by the height and density of buildings in an urban area. Zheng & Li (2022) demonstrated that by increasing SVF from 0.05 to 0.45, the indoor temperature increased approximately 10 °C at 16:00 and 4 °C throughout the night for each month. Thus, SVF and sunlight should be considered when determining the property value. Table 1 and Table 2 summarize 2D and 3D factors affecting property value acquired from the reviewed literature.

Table 1: The list of 2D factors affecting property value based on the literature review

Classification	**Factor**	**Description**	**Reference**
Structural characteristics	Age of property	Referring to how long the property is built. Calculating by subtracting the construction year from the analysis year	Wu et al. (2015), Cao et al. (2019)
	Area	Square meters of the living area	Wittowsky et al. (2020), Wu et al. (2015)
	House type	Detached house, semidetached house, terraced house, corner house, and apartment	Wittowsky et al. (2020)
Locational characteristics	Distance to park	The distance to the closest public parks	Piaggio (2021), Kim et al. (2015), Wittowsky et al. (2020), Cao et al. (2019)
	Distance to the business area	The distance to the closest industrial and commercial areas	Liang et al. (2018), Cao et al. (2019)
	Distance to the educational centre	Distance to the closest kindergartens and primary schools	Liang et al. (2018), Wu et al. (2015)
	Distance to the health care centre	Distance to the closest hospital and small medical cares	Liang et al. (2018)
	Distance to the city centre	Distance to the CBD	Wu et al. (2015), Kim et al. (2015), Cao et al. (2019)
	Distance to the public transportation station	Distance to the closest public transport stations	Efthymiou & Antoniou (2013), Liang et al. (2018)
	Distance to the train station	Distance to the closest train station	Efthymiou & Antoniou (2013), Yang et al. (2020)

	Distance to amenities such as restaurants	Distance to restaurants	Wittowsky et al. (2020)
Environmental characteristics	Distance to water	Distance to the closest water bodies	Wen et al. (2017)
	NDVI	Referring to the level of vegetation cover around the property	Zambrano-Monserrate et al. (2021)
	Air pollution	Referring to the level of air pollution around the property	Saptutyningsih et al. (2013)

Source: Author, 2024.

Table 2: The list of 3D factors affecting property value based on the literature review

Factors	Description	Reference
Sky view factor	Referring to the openness of the sky around the property (influenced by the height and density of buildings around)	Zheng & Li (2022)
Property orientation	Referring to whether the property receives natural lights. A dummy variable (If the orientation is south, southwest, and southeast, 1; otherwise, 0)	Wu et al. (2015)
Sunlight	This refers to whether the property is in shadow at a specific time during the day and receives more sunlight. A Dummy variable (If property not in shadow, 1; otherwise, 0)	Fleming et al. (2018)

Source: Author, 2024.

2.2. General methods to quantify the impact of factors on property value

The literature on property value analysis shows two principal research trends: hedonic price modelling (HPM) and machine learning algorithms for developing property value prediction models. The HPM is widely applied to estimate property value based on the variables that affect property value (Wittowsky et al., 2020; Wu et al., 2015). The theory underlying the HPM is that the value of properties is comprised of various factors (e.g., structural, environmental, and locational attributes). In HPM, the property value is considered the dependent variable, and the other identified factors are the explanatory variables (Wing & Chin, 2003). Based on this theory, how explanatory variables contribute to property value is analysed. The hedonic price regression model shows the portion of the value that is determined by each property characteristic. It can be estimated through the coefficient of the regression (importance of indicators) while all other factors are held constant (Sirmans et al., 2005).

Various research has focused on analysing the correlation between various factors and property value using HPM (Chen et al., 2023; Jim & Chen, 2010; Wüstemann & Kolbe, 2015). For instance, Chen et al. (2023) exerted an HPM to explore the association between property characteristics and value. They employed two functional forms of HPM: linear regression and semi-log regression. Similarly, Wen et al. (2017) researched the effects of the Grand Canal landscape on property value in Hangzhou, China. They constructed hedonic price and spatial econometric models to examine this correlation.

The HPM uses different methods, such as Ordinary Least Squares (OLS) regression, the Artificial Neural Network (ANN), and Spatial Econometric models (Usman et al., 2021). The most widely used method for estimating property value is conventional HPM using OLS (Saptutyningsih et al., 2013; Wittowsky et al., 2020). HPM also requires an appropriate functional form to find the correlation between property value and other independent factors (Usman et al., 2020). The functional form in HPM refers to the mathematical equation that captures the relationship between property value as a dependent variable and other independent variables (Sirmans et al., 2005). Function forms include linear, quadratic, semi-log, log-log, and Box-Cox transformation (Wüstemann & Kolbe, 2015).

The hedonic price models have limitations in assumptions and estimations (Usman et al., 2020; Cao et al., 2019). For instance, two major issues related to HPM are Omitted Variable Bias and Collinearity from compound variables, leading to inconsistent estimation and modelling (Mosammam et al., 2017). However, recent studies have proposed statistical methods to handle these limitations. Econometricians usually apply Spatial Econometric models, such as the spatial regression model (SAR), spatial error (SEM), Durbin (SDM) and autocorrelation (SAC) model, and Geographically Weighted Regression (GWR) to capture the effect of spatial factors on property value (Efthymiou & Antoniou, 2013). Recent studies on property value modelling have extensively employed GWR as a statistical alternative for HPM (Cao et al., 2019; Liang et al., 2018). GWR is a multiple regression model that can be applied to precisely assess the spatial variability of the relationship between dependent and independent variables (Liang et al., 2018). Cao et al. (2019) employed two models of OLS and GWR to find the effects of distance to CBD and the distance to the nearest MRT station on property value. Comparing the results of the two models showed that the GWR model performed much better than the traditional hedonic regression model. In fact, the GWR model can effectively support the spatial variation of public transportation in this research so that more strong results will be acquired through this model. Similarly, in another research, Liang et al. (2018) executed HPM based on the OLS and GWR models to find the effect of locational indicators on property value. The authors discovered that GWR is more reasonable in explaining the relationship between the property value and the explanatory variables compared to OLS.
However, recent studies have highlighted the advantage of machine learning algorithms over HPM. One of the main limitations of HPM is that it only assumes a linear relationship between the independent variables and property value. Meanwhile, machine learning algorithms aim to examine more complex and non-linear correlations between variables, which makes the property value model more accurate (Ho et al., 2021; Hoang & Tran, 2021). Hu et al. (2019) applied six machine learning algorithms, including random forest regression (RFR), extra-trees regression (ETR), gradient-boosting regression (GBR), support vector regression (SVR), multi-layer perceptron neural network (MLP-NN) and K-nearest neighbour algorithm (k-NN) to establish a property value prediction based on several locational, environmental, and structure factors. Selim (2009) indicated that an artificial neural network (ANN) is an improved alternative for predicting property value in Turkey by comparing prediction performance between conventional hedonic price modelling and ANN. Yao et al. (2016) also integrated a convolutional neural network (CNN) with random forests (RF) and proposed a deep-learning-based framework to map housing value in Shenzhen. Furthermore, previous studies compared the results of several machine-learning algorithms to find the best algorithm for modelling housing value. Park & Kwon Bae (2015) tested the performance of several machine-learning algorithms used in property value modelling and indicated that a machine-learning algorithm could enhance the predictability of property value. Ho et al. (2021) applied three machine learning algorithms, SVM), RF, and GBM, to predict property value. They found that RF and GBM are more accurate in predicting property value than SVM. However, the choice of algorithm depends on several factors, such as the size of the data set, the computing power of the equipment, and the availability of waiting time for the results (Ho et al., 2021).

2.2.1. Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR)

Based on the literature, one of the common methods of property value analysis is OLS and GWR. These two models are described in more detail in the following. First, the OLS regression method estimates the effects of different independent factors on property value as a dependent variable. The OLS equation is:

$$y_i = \beta_0 + \Sigma\beta_k x_{ik} + \varepsilon_i \quad (1)$$

Where y_i is the property value, β_0 represents the intercept value, β_k represents the coefficient of the corresponding variable to be estimated, x_{ik} represents the corresponding independent variable, and ε_i is the error term showing the difference between the actual value and the predicted one.

Second, the GWR is a multiple regression model that can be applied to precisely assess the spatial variability of the relationship between dependent and independent variables (Liang et al., 2018). GWR is a linear regression model, the same as OLS, but GWR handles spatial variability in the model. Based on Fotheringham, Brunsdon, and Charlton (2002), the GWR equation is:

$$y_i = \beta_0(u_i, v_i) + \Sigma\beta_k(u_i, v_i)x_{ik} + \varepsilon_i \tag{2}$$

Where y_i is the property value of the location of i, β_0 represents the intercept value at the location of i, β_k represents the coefficient of the corresponding variable to be estimated, (u_i, v_i) refers to the spatial coordinates of the sample point i, x_{ik} represents the K^{th} attribute for location i, and ε_i is the random error.

The GWR regression represents a weighted matrix W_i in estimating the regression coefficient. The observations closest to the i location in the weight function offer the most accurate estimate of the coefficient at i. Where the d_{ij} represents the distance between the location of i to nearby observation j, W_{ij} is the weight for the K^{th} variable at the location of i, and b is the bandwidth, which can be specified either by a fixed distance or by a fixed number of nearest neighbours (Lu et al., 2011).

2.2.2. Random Forest (RF)

One of the most common machine learning algorithms to develop prediction models is RF, a classification and regression tree collection. In RF, decision trees are constructed using random training datasets and random subsets of variables to build a prediction model. Then, the results of each decision tree are aggregated and averaged to explore the best outcome for the random forest prediction. Since the RF model consists of different decision trees, it provides higher accuracy results than a model with only one decision tree (Speiser et al., 2019).

The key features of RF are Bootstrapping, Feature Randomness, and Ensemble Averaging. A distinct bootstrap sample of the data is used to train each tree, meaning the same data is not applied for each decision tree in the model. Feature randomness also illustrates that a random subset of features is considered when splitting a node in the tree. This will increase the diversity among the trees and decrease the correlation among them. In the end, the prediction results of each decision tree are averaged, making the final random forest model prediction. It should also be noted that the RF method resists overfilling well and performs more effectively with large datasets. Figure 3 shows the structure of the RF method.

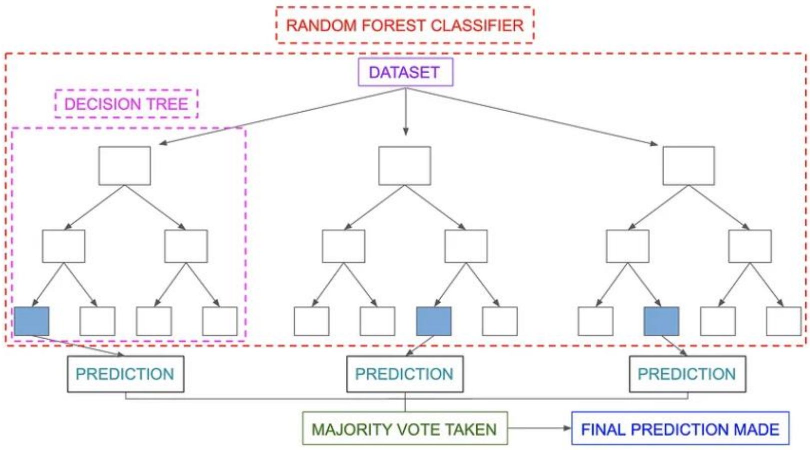


Figure 3: Random forest method structure

Source: (Belgiu & Drăgu, 2016)

2.3. General types of urban green environments

Urban green environments are heterogeneously distributed all around the cities. This heterogeneity can be measured from various perspectives, such as size, density, and quality of vegetation. Through each classification approach, different types of green environments are provided. The following discusses the classification approaches and general types of urban green environments based on the literature review.

Urban green environments can be classified based on the services they provide. People value green environments according to the services they offer (Peschardt et al., 2012). For instance, parks offer more recreational activities, while attached green spaces primarily make the building environment more beautiful. People prefer to reside near each of these green spaces to take advantage of the services. Panduro & Veie (2013) researched the effects of eight categories of green space according to their services to find their effects on property value in Alburg, Denmark. These eight groups were parks, lakes, nature, churchyards, sports fields, common areas, agricultural fields, and green buffers. As a result of this research, people do not value these types of urban green environments equally.

Classification based on the function of urban green environments is another approach to classification. Green environments serve different functions, such as recreational parks, biodiversity conservation areas, urban agriculture sites, and cultural or historical landmarks (Panduro & Veie, 2013). Degerickx et al. (2020) also similarly categorized urban green spaces based on providing similar functions related to specific ecosystem services. They explored that a botanical garden is fascinating from an ecological, educational, and scientific standpoint compared to a lawn, which primarily functions as a playground for children.

Urban green spaces provide multiple ecosystem services. The contribution of each type of urban green space differs in terms of ecosystem service. To explore this differentiation, Derkzen et al. (2015) published a list of eight classes of green spaces based on their importance in ecosystem services in Rotterdam, Netherlands. First, they defined six significant ecosystem services of air purification, carbon storage, noise reduction, run-off retention, cooling, and recreation to evaluate the effects of each type of green environment on these ecosystem services. Then, they classified green spaces based on these ecosystem services into types of trees: woodland, tall shrubs, short shrubs, herbaceous, gardens, water, and others.

Physical characteristics of urban green environments are also significant in classification. Zambrano-Monserrate et al. (2021) investigated how different characteristics of green spaces, such as the area of green space, affect property value. Through this research, they divided green environments such as parks, forests, and cemeteries based on their sizes into small, medium, and large. Moreover, they also explored that green environments can be categorized based on their size and the activities they accommodate. For instance, parks are designed for short walks and walking a dog, while medium parks are great green spaces for people to do various activities such as exercise or take a walk. Large size parks are also designed for families and friends to do several activities together. In similar research, Czembrowski & Kronenberg (2016) classified green environments according to size. They indicated that the effects of proximity to green spaces on property value differ in green spaces of different sizes. They also found that extensive forests and large parks are the most important and, along with small forests, positively influence property value. In another research related to this approach, Jiao & Liu (2010) classified green environments based on the size and service level in Shanghai into three groups of city level green spaces (such as parks above 10ha), district-level (parks of 4-10 ha), and community-level green spaces (under 10 ha green spaces covered by trees and lawn). Similarly, Chen et al. (2023) explored that the size and shape of green spaces are two main characteristics of green space that affect property value, and these can be measured by a Landscape Shape Index (LSI). They found a positive effect of the size of green spaces on property value. Moreover, they discovered the importance of the shape of green spaces on housing value. They suggested that when the LSI of green space is more significant than 1.3, which means that the green space has an irregular shape and is more complete than a simple shape, this green space positively affects the value of adjacent houses.

Researchers developed a new classification system called "Local Climate Zones (LCZ)". This classification system categorizes buildings and vegetation in urban and rural environments based on specific characteristics (density and height) to present local climate zones (Kaloustian & Bechtel, 2016; Stewart et al., 2014). In the LCZ classification system from the height classification approach, urban green environments are divided into high-rise, midrise, and low-rise vegetation. On the other hand, from the density point of view in the LCZ classification system, four classes of dense trees, scattered trees, bushes, and low plants are suggested.

In a similar study, Mathey et al. (2021) researched classifying green environments based on the height and density in cities to improve the quality and quantity of green environments. They found three classes of vegetation height: i.e., low height (grass, meadows, shrubbery, herbs of low height ≤ 1 m), medium height (shrubbery, herbs, hedges, bushes, and small trees > 1 m to ≤ 3 m) and high vegetation (medium and high trees > 3 m). Furthermore, from the density point of view, they categorized urban greenery into three classes: low-volume, medium-, and high-volume vegetation. Moreover, Gupta et al. (2012) conducted a study to measure the green environments in cities. They classified green spaces based on height into four categories: very high, high, moderate, and low-rise green spaces. On the other hand, they categorized green environments based on density into four classes: dense vegetation, grass, low vegetation, and open spaces without vegetation. As a result of this research, urban green areas with high-rise low density and low-rise low density have good quality compared to other categories. To classify green environments based on density, NDVI has been applied to distinguish vegetated and non-vegetated (Gupta et al., 2012). Faryadi and Taheri (2009) also applied NDVI measurement to classify urban green environments based on the vegetation density cover to evaluate their environmental quality in Tehran.

Classification based on the vegetation type in urban green environments is another approach to classification. Degerickx et al. (2020) researched to optimize the design and management of green spaces regarding their ecosystem services. Through this research, they categorized green environments into three main categories: trees, shrubs, and herbaceous plants. Similarly, Mathey et al. (2021) classified urban green environments based on the type of vegetation into four classes: deciduous, evergreen trees, shrubs, and decorative lawns.

The final significant approach to green space classification explored through the literature was qualitative. Stessens et al. (2020) identified quietness, spaciousness, cleanliness and maintenance, and the feeling of safety as essential factors in classifying green spaces. They conducted a survey among Brussels, Belgium, residents to investigate their perception of urban green spaces from each considered qualitative approach to classify urban green environments. In the end, different classification approaches and types of urban green environments are summarized in Table 3.

Table 3: Summary of classification approaches and types of urban green environments based on the literature

Classification approach	Types	Reference
Service	Parks, lakes, nature, churchyards, sports fields, common areas, agricultural fields, and green buffers	Panduro & Veie (2013), Jiao & Liu (2010), Mathey et al. (2021)
Function	Recreational parks, biodiversity conservation areas, urban agriculture sites, and cultural or historical gardens	Panduro & Veie (2013), Degerickx et al. (2020)
Ecosystem services	Woodland, tall shrubs, short shrubs, herbaceous, gardens, water	Derkzen et al. (2015)
Size	Small, medium, and large size	Zambrano-Monserrate et al. (2021), Chen et al. (2023), Czembrowski & Kronenberg (2016)
Density	Dense vegetation, grass and low vegetation, and open spaces without vegetation	Stemwart & Oke (2012), Kaloustian & Bechetl (2016), Mathey et al. (2021)
Height	High-rise, midrise, and low-rise green environment	Chen et al. (2023), Stemwart & Oke (2012), Kaloustian & Bechetl (2016), Mathey et al. (2021)
Type of vegetation	Deciduous, evergreen trees, shrubs, and decorative lawns.	Degerickx et al. (2020), Mathey et al. (2021)
Quality of green spaces	Measurement of the level of, for instance, quietness, spaciousness, cleanliness and maintenance, and the feeling of safety	Canters et al. (2020)

Source: Author, 2024.

2.4. Summary

First, different 2D and 3D factors affecting property value are explored by literature review in this chapter. Besides, several classification approaches have been discovered that provide different types of urban green environments through each classification approach. In addition, three OLS, GWR, and RF methods are reviewed in detail to model the property value.

3. METHODOLOGY

3.1. Study area

The study area for this research is the city of Alkmaar Figure 4, which is located in the North of the Netherlands in the province of Noord-Holland. Alkmaar is a medium-sized city with 111.766 inhabitants (CBS, 2023). According to Dutch statistics, the municipality of Alkmaar contained a total of 53831 residential properties till the end of September 2023 (Centraal Bureau voor de Statistiek, 2023). However, following the population growth in recent years, the housing demand also increased in Alkmaar.

Based on the reviewed documents published by the municipality of Alkmaar, following the population growth, the municipality plans to construct 15000 homes for approximately 30000 residents along the bank of the Noordhollandsch Canal, which is named the “Alkmaar Canal Program”. According to the “Environmental Vision 2040 of Alkmaar (Omgevingsvisie Alkmaar 2040)” report, the city should transform the industrial area along the banks of the Noordhollansch Canal into a high-density multi-functional (housing, business, recreation) urban area. In addition, the 2040 vision pictures the compact city of Alkmaar, focusing more on city densification than city expansion (Gemeente Alkmaar, 2017). Thus, the rate of construction of new properties in Alkmaart is increasing.

In addition to constructing new properties along the banks of the Noordhollansch canal plan, there is another plan for greening the city named “Alkmaar Greenery Policy Plan 2017-2027 (Beleidsplan Groen 2017-2027)”. According to this plan, the new green areas will be constructed in neighbourhoods that lack greenery or where extra greenery is desirable for rainwater infiltration or the prevention of heat stress. As part of Alkmaar’s green policy, a separate plan elaborating on the tree structure was released in 2017 (Gemeente Alkmaar, 2017). Based on the “Tree Structure Plan for Alkmaar 2017-2027” report, even though Alkmaar already has much greenery, the city is planning to implement this plan of planting new green environments to enhance the liveability of cities.

To conclude, according to the reviewed documents and reports from the municipality of Alkmaar, urban green environments are improving across the city. Besides, the Alkmaar property market has changed in recent years, leading to planning to build new houses to meet the housing demand in Alkmaar.

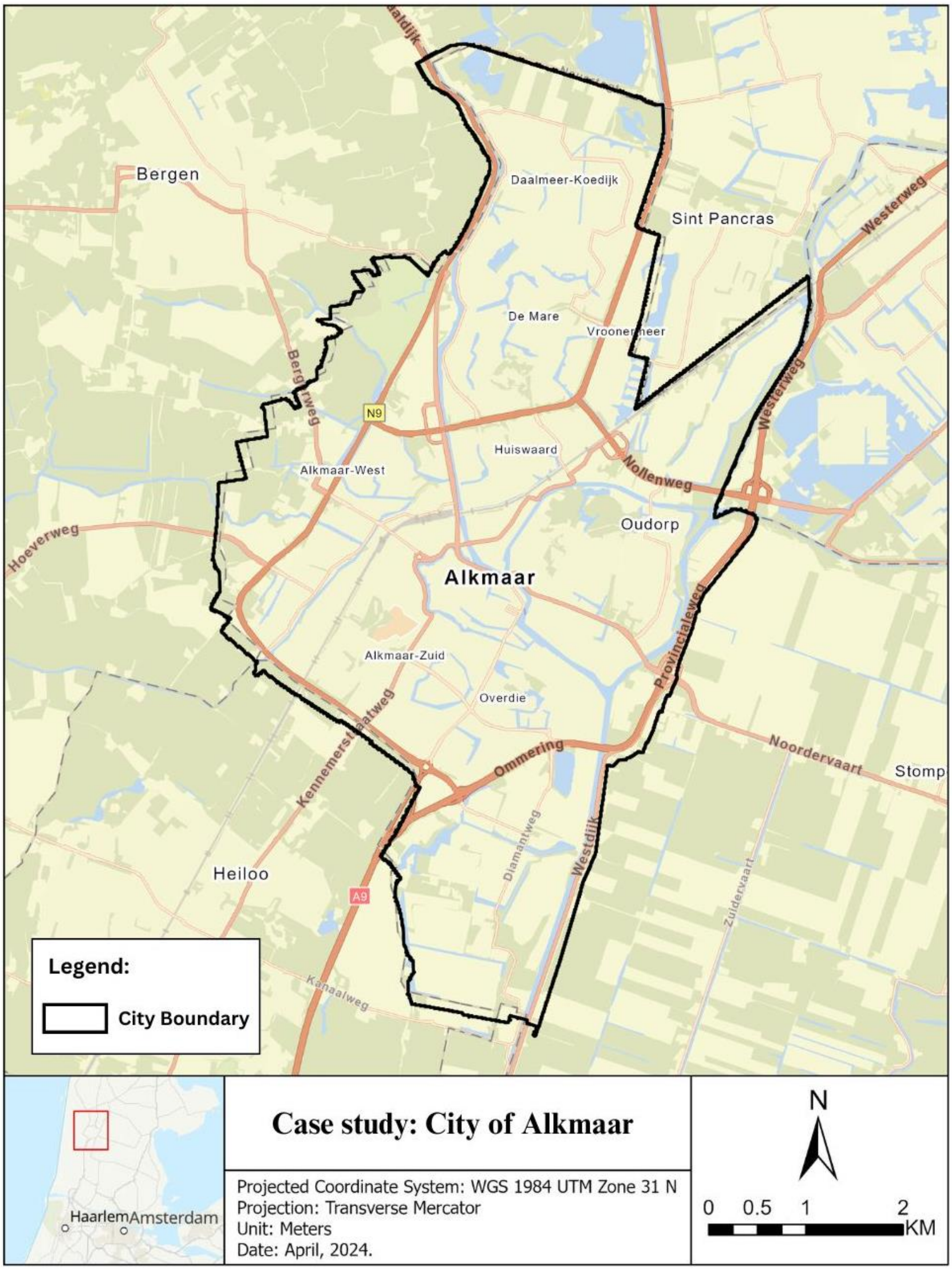


Figure 4: Case Study, City of Alkmaar

Source: Author, 2024.

3.2. Overal approach and method

This section describes the quantitative methods of the research that have been applied to address each research sub-objective. Figure 5 shows the overall methodological flowchart of the research. The flowchart is divided into three sections based on the sub-objectives, which will be elaborated precisely in the following.

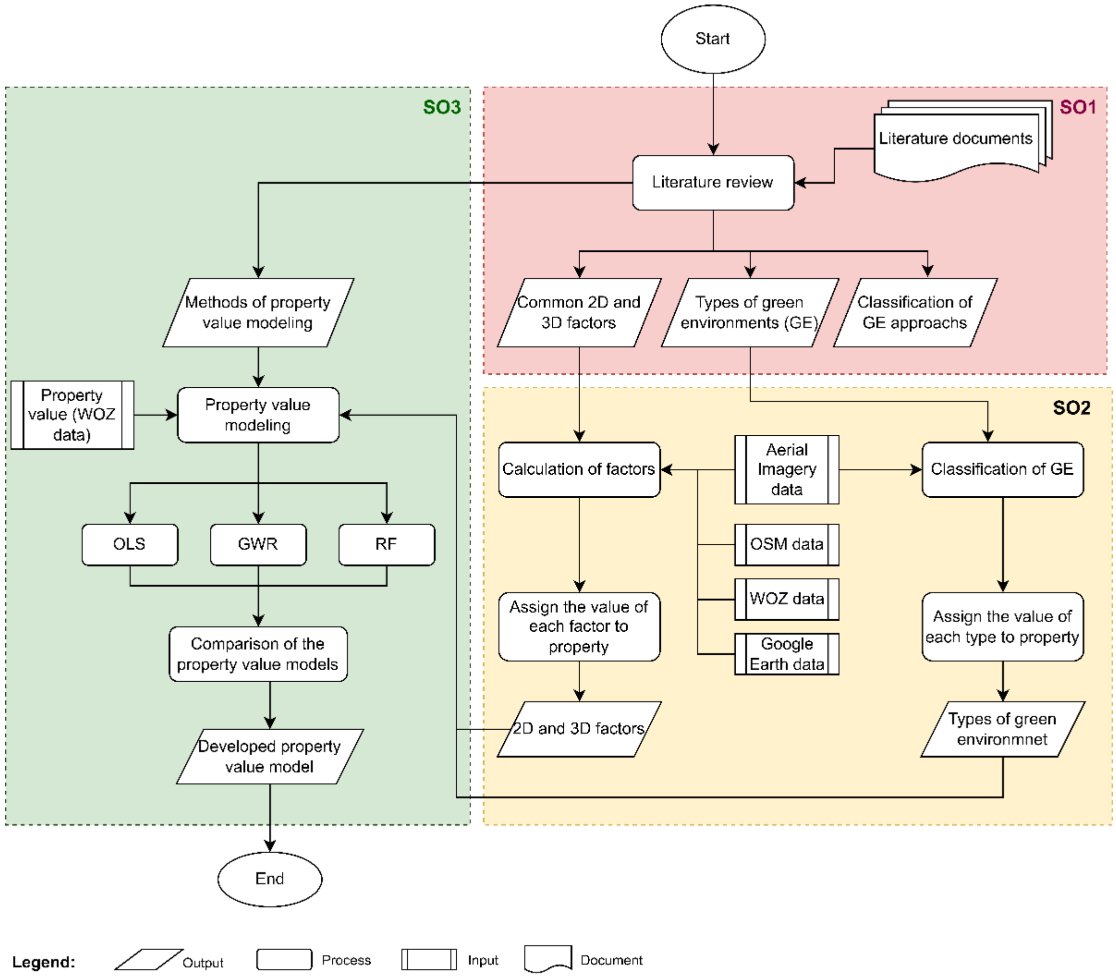


Figure 5: The overall methodological flowchart of research

Source: Author, 2024.

3.3. Factors affecting property value and types of green environments

The 2D and 3D factors affecting property value are described as a result of the literature review. In addition, the common approaches for classifying green environments are explored. According to data availability and study area analysis, the final list of factors and types of green environments is provided. Table 4 and Table 5 illustrate the exact definition of the factors, their categories, and an overview of the required data and sources for analysis. Moreover, Table 6 demonstrates the final list of 3D factors and their main definition in this research. All the required data for analysis are secondary data without any restrictions on downloading them. After downloading the data from the sources, the data is corrected by removing any mistakes in the dataset or missing values such as null data on property value.

Table 4: The final list of the types of green environments and required data

Factor	Category	Definition	Data/ Unit	Source
Distance to parks	Classification based on services	Calculating the distance of properties to the nearest parks	Location of the main entrance of parks (.shp)	https://www.pdok.nl/
Distance to agricultural lands		Calculating the distance of properties to the agricultural lands	Location of agricultural lands (.shp)	https://www.pdok.nl/
Distance to recreational green spaces		Calculating the distance of properties to the nearest recreational green spaces	Location of recreational green spaces (.shp)	https://www.pdok.nl/
Distance to greenhouses		Calculating the distance of properties to the nearest greenhouses	Location of greenhouses (.shp)	https://www.pdok.nl/
Distance to allotment		Calculating the distance of properties to the nearest allotment	Location of allotments (.shp)	https://www.pdok.nl/
Distance to sports fields		Calculating the distance of properties to the nearest sports fields	Location of sports fields (.shp)	https://www.pdok.nl/
Density of green environment	Classification based on the density	Calculating the mean NDVI value in the distance of 25m around the properties as the density of vegetation	Aerial imagery data-25-cm resolution (tiff)	https://www.pdok.nl/
Height of vegetation	Classification based on the height	Calculating the mean height of vegetation in the distance of 25m around properties	DSM (tiff) DTM (tiff)	https://geotiles.nl/
Size of green environment	Classification based on the size	Calculating the mean size of vegetation in the distance of 25m around properties	Aerial imagery data-25-cm resolution (tiff)	https://www.pdok.nl/
Type of vegetation	Classification based on the type	Calculating the percentage of each type of vegetation (grass, shrubs, evergreen, and deciduous trees) in the distance of 25m around the properties	Location of trees (point), map of type of vegetation (tiff)	https://www.pdok.nl/ https://www.alkmaar.nl/

Source: Author, 2024.

Table 5: The final list of the 2D factors and required data

Factor	**Definition**	**Data/unit**	**Source**
Property value	WOZ property value	Property value (Euro/m^2)	WOZ, 2022.
Age of properties	Age of properties from the year of construction till 2024	Age (number)	WOZ, 2022.
Size of property	The area of the property (square meter)	Area (number (m^2))	WOZ, 2022.
Type of property	The types of property from the 5 categories of Detached houses, Semi-detached houses, Corner houses, and Terraced house	Categorized data/ Detached house=1 Semi-detached house=2 Corner house=3 Terraced house=4 Apartment=5	WOZ, 2022.
Distance to the nearest business areas	Calculating the distance of properties to the closest industrial and commercial area	Location of industrial and commercial areas (.shp)	OSM, 2024.
Distance to school	Calculating the distance of properties to primary and secondary schools	Location of primary and secondary schools (.shp)	OSM, 2024.
Distance to the healthcare centres	Calculating the distance of properties to hospitals	Location of hospitals (.shp)	OSM, 2024.
Distance to CBD	Calculating the distance of properties to the central business district	Location of CBD (.shp)	OSM, 2024.
Distance to road	Calculating the distance of properties to roads with a max speed of 50 km/h	Location of roads with a max speed of 50 km/h (.shp)	OSM, 2024.
Distance to the bus stations	Calculating the distance of properties to the nearest bus stations	Location of bus stations (vector)	OSM, 2024.
Distance to the train stations	Calculating the distance of properties to the nearest train stations	Location of train stations (.shp)	OSM, 2024.
Distance to the amenities	Calculating the distance of properties to the nearest restaurants	Location of restaurants (.shp)	OSM, 2024.
Distance to water	Calculating the distance of properties to water canals in the city	Location of water canals (.shp)	OSM, 2024.
Air pollution	Calculating the amount of No_2 in the city	Sentinel-5p satellite image (tiff)	https://earthengine.google.com/

Source: Author, 2024.

Table 6: The final list of the 3D factors

Factor	Definition
View to sky	The percentage of the property view allocated to sky
View to water	The percentage of the property view allocated to water
View to vegetation	The percentage of the property view allocated to vegetation
View to road	The percentage of the property view allocated to roads
View to building	The percentage of the property view allocated to other buildings
Property orientation	A dummy variable (If the orientation is south, southeast, southwest, and southeast, 1; otherwise, 0)
Sunlight	The building is not in shadow at a specific time during the day and receives more sunlight. A Dummy variable (If property not in shadow, 1; otherwise, 0)
Height of vegetation	Calculating the mean height of vegetation in the distance of 25m around properties

Source: Author, 2024.

3.4. Methods of classification of green environments

This section introduces the methods for the classification of green environments based on density, height, type of vegetation, and size of green environments. The method of classification of green environments based on the services they provide will be discussed in section 3.5.2. Figure 6 illustrates the diagram of classification directly.

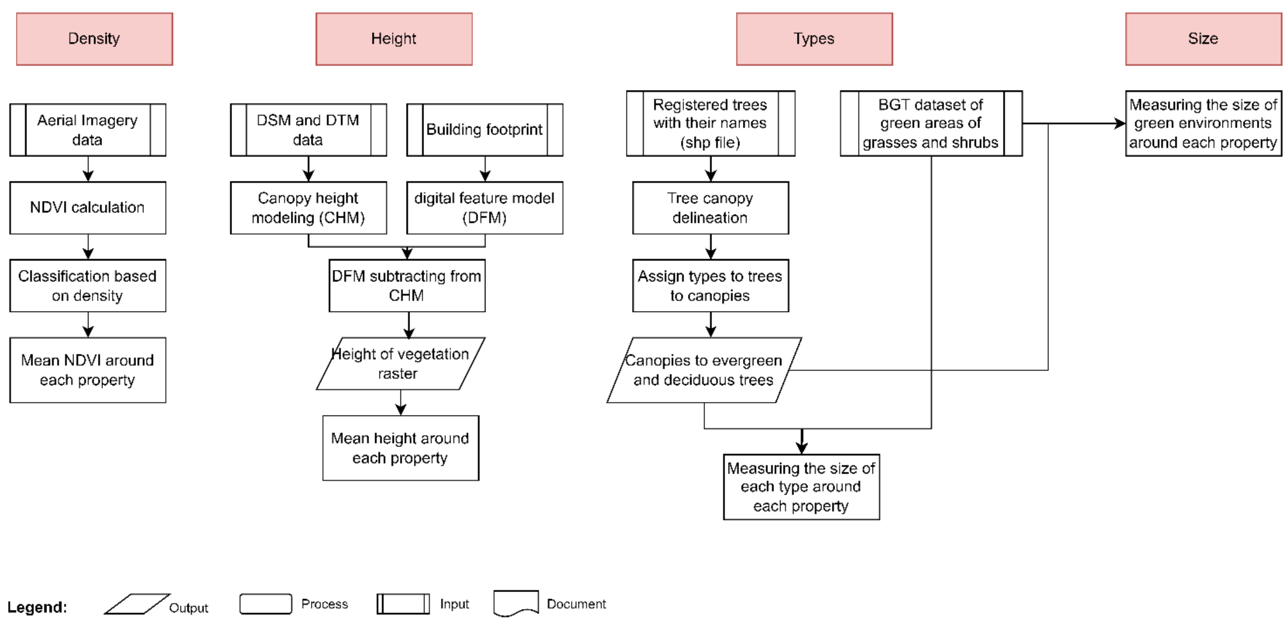


Figure 6: The diagram of classification of different types of green environments

Source: Author, 2024.

3.4.1. Classification of green environments based on the Density

To classify green environments based on density, NDVI is calculated (Aryal et al., 2022; Mathey et al., 2021). This index measures the level of vegetation cover around properties in the city of Alkmaar. NDVI is performed on a multispectral analysis of very high-resolution aerial images. NDVI value is between -1 to +1, and it can be calculated by this formula:

$$NDVI \ = (NIR - RED)/(NIR - RED) \tag{3}$$

NDVI value closer to 1 shows more vegetation cover compared to a lower NDVI value (Zambrano-Monserrate, 2021). This study uses aerial imagery data with four bands of RGB+NIR bands with a resolution of 25 centimeters (from the PDOK website[1]). This image is dated 8th July 2023, with less than 10% cloud. Faryadi & Taheri (2009) divided the NDVI map into 5 classes (Table 7): Thick vegetation, dense vegetation, medium vegetation, scarce vegetation, and no vegetation are the types of green vegetation. However, this research considers the value of NDVI as the density of a green environment, meaning that a higher NDVI value means the green environment is denser.

Table 7: Classification of green environment based on the level of density

Level of Density	NDVI (-1,1)	Categorization value
Thick vegetation	0.61<NDVI<1	1
Dense vegetation	0.451<NDVI<0.6	2
Medium vegetation	0.31<NDVI<0.45	3
Scarce vegetation	0.151<NDVI<0.3	4
No vegetation	-1<NDVI<0.15	5

Source: Faryadi and Taheri, 2009.

3.4.2. Classification of green environments based on the height

To classify the urban green environments based on height, their height is measured using the Digital Surface Map (DSM) and Digital Terrain Model (DTM). DSM and DTM are obtained from the Actueel Hoogtebestand Nederland 4 (AHN 4) data repository with a 0.5-meter resolution.

In the data cleaning process, there are a number of No-data pixels (missed data) for DTM and DSM. This problem is solved by filling in these pixels with the NoData Fill tool from GDAL in QGIS. Thus, the input data is completed. Then, based on the Canopy Height Model (CHM), the DSM is subtracted from DTM to find the height of features above the ground (Figure 7). A Raster Calculator is used to calculate this in the QGIS environment.

In the next step, the building footprint from OSM data (OSM was easily accessible compared to other data sources) and the AHN4-point cloud data are acquired to find the height of the building. The lidar points are classified into classes of buildings, ground, unassigned, water, and reserved. The point clouds of the building footprint are extracted and converted into a raster showing the height of the buildings (Digital Feature Model (DFM)). By subtracting the DFM from CHM, the height of the urban green environments of Alkmaar was found. The method of categorizing is Natural Breaks (Janks); this method accounts for non-uniform distributions, giving an unequal class width with varying frequency of observations per class.

[1] https://www.pdok.nl/

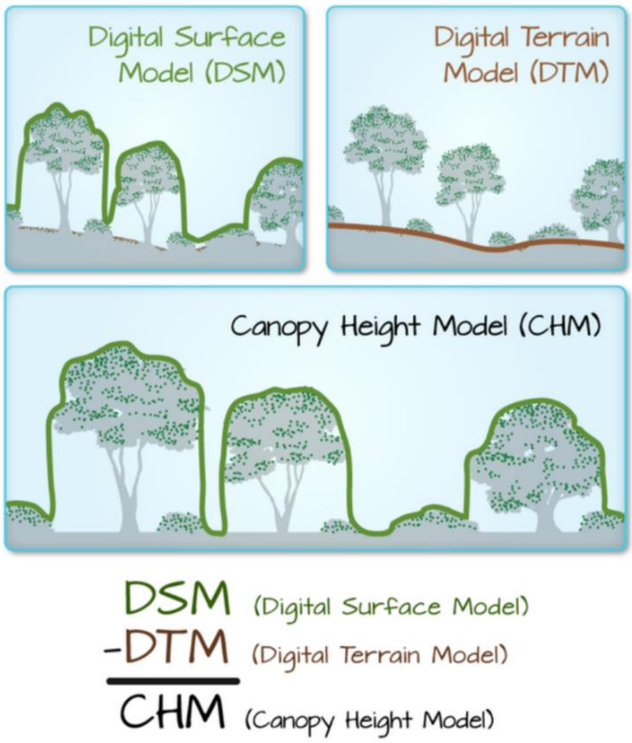


Figure 7: Canopy height model calculation.

Source: https://www.earthdatascience.org/

3.4.3. Classification of green environments based on the type of vegetation

To classify the green environments based on the type of vegetation, four classes of evergreen trees, deciduous trees, grasses, and shrubs are considered. According to the BGT dataset on the PDOK website, the placement of two types of shrubs and grass are explored, while evergreen and deciduous tree categories are missing. Thus, the canopy of trees is first detected and then classified into two types of trees.

3.4.3.1. Tree canopy detection

This study delineates tree canopy using the NDVI and the CHM. The level of vegetation cover in the city is identified by calculating the NDVI, and the vertical height of the tree canopy to the ground is also found by the calculation of CHM. The tree canopy is detected by the combination of NDVI and CHM. The NDVI threshold mask (NDVI>0.3) is applied to keep only areas with this level of vegetation. On the other hand, height thresholding (CHM>2m) is also applied to find only trees and other high-rise vegetation. The combination of these two layers delineates the 2D boundary of tree canopies.

3.4.3.2. Tree classification

The data on the location of trees is acquired from the Municipality of Alkmaar. It is the list of trees registered and managed by the Municipality. This point data (SHP file) is analysed in the ArcGIS Pro environment, and their names are mentioned in the attribute table of tree points. Based on the name of the trees, they are categorized into deciduous and evergreen tree classes. Then, this map intersects with the tree canopy map, and these two categories are assigned to the trees on the classified map. In the end, the four types of grasses, shrubs, evergreen trees, and deciduous trees around the properties are explored. To find their effects on the property, the percentage of each type around the property is calculated.

3.4.4. Classification of green environments based on the size

The last approach to classifying green environments in this research is size. By merging all the types of green environments from section 3.4.3, the size of green environments scattered across the city of Alkmaar is explored. This classification calculates the area of the green environments at a specific distance from the properties to find its effect on property value.

3.4.5. Property analysis buffer zone

To find the correlation of different types of green environments on the property value, a 25m buffer zone is established around each property (as a point in the WOZ dataset) (Figure 8). The factors regarding the type of green environments are calculated in this buffer zone and then assigned to the respective property. Different radius buffer zones of 25m, 50m, and 100m are examined to find which one is more effective. The results illustrated that a radius of 25m is suitable for exploring the effects of green environments in this distance around properties on their values (Appendix 1). A 25m radius buffer zone makes it possible to evaluate the effect of the green environment surrounding the property more locally and in more detail, while those above the 25m radius size might overlook these effects on property value.

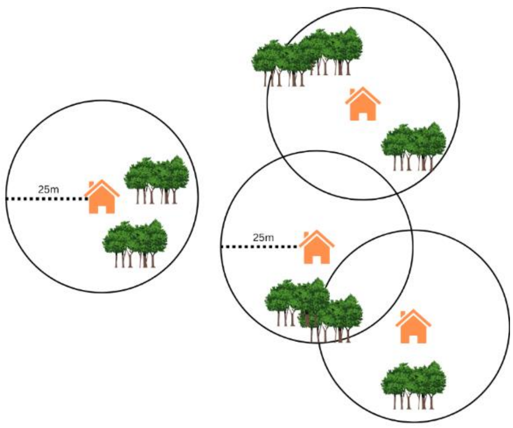


Figure 8: A visualization of 25m buffer zone of properties

Source: Author, 2024

3.5. Methods of calculating 2D factors on property value

This section includes the methods to calculate the 2D factors. First, the property points are acquired from the WOZ dataset, which consists of the property value (as the dependent variables), age, size, and property types (as physical characteristic variables). The Municipality publishes the property value data (WOZ) every year, and the WOZ data used in this research is from 2022. However, there is missing and incorrect information. For instance, there are no existing values for several properties in the city, which are deleted in the data cleaning process. In addition, there are other 2D independent factors, and the methods for calculating these variables are discussed in the following. Moreover, these factors are calculated and visualized through ArcGIS Pro.

3.5.1. Global Moran's Index

Global Moran's Index is measured to analyse the property value variation across the city of Alkmaar. It is a measure of spatial autocorrelation that is widely applied to assess whether the value of variables and location are spatially correlated (Song et al., 2011). This indicates that a variable's distribution pattern is clustered, dispersed, or randomly distributed in an area. This index aims to explore whether high-value or low-value properties are clustered or dispersed across Alkmaar, which is crucial in the property value analysis.

Global Moran's Index ranges from -1 (indicating a high negative correlation) to +1 (indicating a high positive correlation). 0 value in the middle of the range also means the value of the variable is uncorrelated with the spatial location of the variable. The significance of the Global Moran's Index is evaluated by the z-score and p-value. The z-score measures the standard deviation of the results from the mean. The p-value also shows the probability that the observed pattern was created randomly (Song et al., 2011).

3.5.2. Spatial analysis

According to the table of indicators, the distance to business areas, CBD, bus stations, train stations, restaurants, and water canals needs to be measured. Furthermore, the green environments are classified based on the services they provide into parks, recreational green space, agricultural lands, greenhouses, allotments, and sports fields. Distance to these types is another factor that is used to measure their effects on property value. Thus, the Euclidean distance method is determined to analyse these factors spatially. This method is widely accepted in many research studies and is straightforward in measuring proximity. The Euclidean distance method delineates a straight line as the shortest distance between two points in an Euclidean space and calculates the length of the line (Del Giudice, 2023).

3.5.3. Air pollution calculation

Based on the literature, air quality is one of the important factors affecting property values. Nitrogen Dioxide pollutants (NO_2) is one of the most common air pollutants. The amount of NO_2 in the city of Alkmaar is extracted by Python language from the Google Earth Engine using the Sentinel-5P Data (taken on 10th April 2023) (Appendix 9). The higher the No_2 pollutant, the greater the air pollution situation in the area.

3.6. Methods of calculating 3D factors on property value

This section elaborates on the methodology for quantifying 3D factors that affect the property value. These factors consist of property visibility (view to sky, vegetation, rive, and buildings), property orientation, sunlight, building height, and the height of green environments around the property.

3.6.1. Visibility

The 3D factors of the research are quantified by creating a 3D model of the study area with remote sensing techniques. For 3D modelling of the city, according to the reviewed studies related to measuring the view of the properties (Lee et al., 2020; Ying, 2019), Level of Details 1 (LOD1) is employed. Then, a 3D model of the city is created using data on the height of vegetation, building height, location of rivers, and city terrain data. Afterwards, by defining the location of the property that the observer is in, the 3D factors are analysed. It should also be noted that no data is available to quantify the 3D factors for all the properties in Alkmaar. Thus, 30 properties are randomly selected to measure the 3D factors and find their correlation with property value. One of the main reasons behind this number of sample points is the time limitations. Since the data for each factor is obtained manually, the 30 properties can be managed to be researched within the timeframe.

The 3D factors regarding the visibility are calculated and visualized in CityEngine. CityEngine software is widely used in various research studies that represent the 3D model of cities by integrating the characteristics of the buildings and the spatial information. For analysis of visibility, the Viewshed analysis tool in CityEngine is applied to construct a viewshed index to measure the visibility of properties toward the sky, vegetation, buildings, water, and roads (Lee et al., 2020; Ying, 2019).

Following the calculation of the visibility factor, it is assumed that the observer is in the middle of the building, in front of the window, with a horizontal angle of view of 120 and a vertical angle of view of 60 degrees, looking outside directly. In this situation, the height of the building is also considered as another 3D factor affecting the property value model. Therefore, this research selects the properties for 3D factor analysis from neighbourhoods with high-rise buildings. Even though the city of Alkmaar has a low number of high-rise buildings, two neighbourhoods, Ovedie East and Overdie West[2], with high-rise buildings, are chosen. In fact, there are only a few high-rise buildings in the city of Alkmaar[3], and there are 5 properties with a height between (35-45 meters) in these two neighbourhoods, and the rest are approximately 8-10 meters. Then, 15 properties in each neighbourhood are randomly manually selected to explore the effects of 3D factors on the value of these properties. In addition to these factors, the height of green environments around the properties (as a type of green environment) is also important. Thus, the mean value of the height of vegetation around the properties is also considered as a 3D factor in this research.

The 3D factors are first considered at the three distances of 25, 50, and 100m from the properties and then compared. Among these distances, the distance view of 25m is chosen since this distance view considers the effect of visibility factors on property value more locally and also in more detail, while a 50m and 100m distance size might overlook these effects on property value.

Figure 9: The proportion visibility by layers in the total view of the observer in the distance of 25m

Source: Author, 2024.

The Viewshed tool provided the share of each type of view in the total outside view of the observer in a specific property. Figure 9 shows the proportion of each type of view contributing to the total view of the observer; for instance, the view to vegetation is 15% of the total outside view of the property. Then, the details of all of these factors are imported into an Excel file to find their correlation with property value.

3.6.2. Sunlight and Orientation

Property orientation and sunlight are two other 3D factors in this study. Property orientation is considered as a dummy variable (if the property is toward the south, southeast, and southwest, it is valued as one; otherwise, 0). For the sunlight analysis, the "Sun Shadow Volume" tool in ArcGIS Pro is applied to assess whether the building is in shadow at a specific time during the day and receives more sunlight or not. The steps for measuring these factors are described below.

[2] https://allecijfers.nl/wijk/overdie-alkmaar/
[3] https://skyscraperpage.com/cities/maps/?cityID=2423

First, the 3D model of the Ovedie East and Overdie West neighbourhoods is built and imported into the ArcGIS Pro. The considered day is the 2nd of February, 2022, which is the middle of winter with minimum sunlight and long shadow in the sky. This can be the worst sunlight analysis scenario. Based on the results, each building is in shadow at different hours a day. For instance, few buildings are in the shadows at 8:30 in the morning. For this research, the sun shadow volume is considered as a dummy variable; properties not in shadow at 8:30 AM are valued at 1, and the rest are valued at 0 and imported to the attribute table of selected properties.

3.7. Methods of modelling the property value prediction

The data on the factors and types of green environments are in different measurement units, which initially causes spatial data variation. Before running the three OLS, GWR, and RF models, the data are normalized through the max-min standardisation method in the range of 0 to 1. This process reduces the data dispersion and ensures that variables are on the same variation interval, which makes the model perform more accurately. The standardisation equation is:

$$Standardisation\ = (Value - lowest\ value)/(Highest\ value - lowest\ value) \tag{4}$$

3.7.1. Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR)

The hedonic price modelling by the OLS method is applied through multiple regression in SPSS. All the selected factors affecting the property value in the last steps are considered independent variables in the OLS modelling, and the property value is the dependent variable. After the standardisation of the data, ensuring that the factors are in the appropriate analysis format, the OLS regression is run. This method indicates the correlation between each factor and property value in the OLS model. Moreover, Durbin-Watson, F-test, and VIF assumptions are also tested to ensure this model is a great fit for predicting the property value.

Moreover, the GWR method is used along with OLS in this study to capture the effects of spatial factors on property values. Here, Python coding language is executed to perform the GWR model. At first, the predictors are standardided to have a mean of 0 and a standard deviation of 1. Then, the Akaike information criterion (AICc) method is applied to select the final bandwidth GWR. In the GWR model, each weight matrix is formed by the number of data points determined by bandwidth. In addition, the Fixed Gaussian function is selected as the spatial kernel type.

3.7.2. Random Forest (RF)

As the most common machine learning algorithm, RF is applied to develop prediction property value models. The Scikit-learn Python module is used to import the Random Forest Regressor to build the RF model (Appendix 10 and 11). The data is split into two sets of training and testing data in the ratio of 80 to 20. In fact, 80% of the data is used to train the model, and 20% of the data to test it. The cross-validation method is applied to calculate the appropriate value of n_estimators and the random_state (max-depth) in the modelling. First, the range of potential values of 50 to 150 with step 20 for n_estimator and 10 to 50 with step 8 for ‘max_depth” defines. Afterwards, through the cross-validation method, the performance of the model for each parameter is measured. Then, based on the highest cross-validation score, the ‘n_estimator=100’ and “max_depth=42” as optimal value for the current Random Forest model were chosen.

3.8. The methods for property value model validation

After modelling the property value through the method, the model should be validated to find its accuracy and how well it performs in predicting the property value. In fact, the validation represents the model's performance by comparing the predicted values with the actual values.

The property value model is created based on the property value dataset from WOZ data and not the actual value of properties from the real estate market, which is more accurate. WOZ property value is only based on the specific characteristics of the property. In contrast, the real-estate market value is the actual property value influenced by the different economic factors, market demand, and supply. However, the actual property value for all the properties of Alkmaar was not available in this research. Only the actual value of around 30 properties was the actual property value was also obtained from the Funda website[4] which is a real state platform showing the value of available properties to buy or rent in the Netherlands. Since the actual property values were selected randomly and manually from the Funda website, the number of 30 properties makes it possible to conduct the research within the timeframe. Four statistical metrics that are used for the validation of models in this research are introduced below.

- **R^2/ adjusted R^2**

R^2 is a measure of the amount of variability in one variable that is shared by the other. In fact, R^2 illustrates the proportion of variance for the property value that is defined by the 2D and 3D factors and types of green environments as independent variables in the model. R^2 value ranges from 0 to 1. The value 0 means that no amount of the property value variability is derived from independent variables (no correlation). However, the value 1 indicates that the independent variables define all the variations of the property value (complete correlation) (Field, 2018). R^2 assumes that all the independent variables affected the result of the model, while adjusted R^2 prevents the variables that have no effects on the performance of the model. The adjusted R^2 should be the same or very close to the value of R^2 to generalize the model.

- **Root Mean Squared Erro (RMSE)**

RMSE measures the average difference between the predicted values by the property model and the actual property value. In fact, RMSE is a factor in evaluating the model's accuracy by measuring how well the model can predict the property value. The lower the RMSE number, the better fit the model is.

- **Mean Absolute Error (MAE)**

Mean Absolute Error (MAE) is also the measure of error between the predicted property value and the actual property value without considering their direction.

- **K-Fold Cross-validation**

One of the other most common methods that is used in this research is the K-fold cross-validation to estimate the error between the predicted property value and the actual property value. In this approach, the dataset should randomly be divided into k fords of similar size. Then, a classifier using K-1 folds is trained, and the error value is calculated by testing the classifier in the remaining fold. Ultimately, the k-fold cross-validation is the average of the estimation error in each fold (Rodríguez et al., 2010). K-fold cross-validation has the benefit of using all sample data for both training and validation, with each test subsample being used only once (Barrow & Crone, 2013). Error percentage can also be applied to measure the degree of deviation with the equation below:

[4] https://www.funda.nl/

$$Error\ percentage = (The\ standart\ deviation\ of\ the\ predicted\ value) / (\text{The}\ mean\ property\ valye\ of\ the\ samples) \quad (5)$$

3.9. Summary

This chapter represents the methods applied to resolve the research sub-objectives. First, different factors affecting property value and approaches for classifying urban green environments are explored based on the literature review. Second, the methods of classification of green environments based on different approaches are mentioned. Third, the methodology of 2D factors and 3D factors analysis are discussed. In the end, the methods of developing OLS, GWR, and RF models for property value evaluation are introduced.

4. RESULTS

4.1. Property value variation in Alkmaar based on the literature

As the share of the population living in cities in the Netherlands increased, the demand for housing in this country increased (Nijskens et al., 2019). According to the reviewed literature and documents, the average WOZ value of Alkmaar increased from 2022 to 2023. It changed from 318,002 euros in 2022 to 347,874 euros in 2023[5]. However, the city's transaction price differs from the property value published by the municipality. In the following, the transaction changes of Alkmaar in recent years are analysed.

The transaction price is the purchase price of all existing owner-occupied homes sold yearly (Centraal Bureau voor de Statistiek, 2023). Several local real estate reports from Alkmaar are reviewed to explore the variation in transaction price. The reviewed documents demonstrate that after house prices reached a provisional peak in the summer of 2022, a period followed in which the housing market gradually cooled down. This is mainly due to the increased mortgage interest rates, which also reduces the sentiment on the housing market and results in lower demand for buying new homes and, ultimately, falling house prices[6]. Thus, the average transaction price of Alkmaar decreases from 398,000 euros in 2022 to 371,000 euros in 2023 by 9.5% (NVM, 2024). Moreover, the average transaction price per square meter from 2022 to 2023 increases from 3,829 euros/m^2 meter to 3,965 Euros/m^2.

According to Table 8 and Figure 10, the two neighbourhoods of Zuid and Veroonermeer have the highest transaction prices compared to other properties, while the transaction price per square meter in Centrum is the highest among all other neighbourhoods in Alkmaar. One of the reasons behind this is that the size of properties in the centre is small, and there is a high demand for limited spaces in the centre compared to other neighbourhoods that are located in the city surrounded by large-sized properties. Based on Figure 11, even though the average transaction decreases from 2022 to 2023, there are fluctuations in different seasons. In the following, the fluctuation of transaction prices of different housing types in recent years is analysed based on the available documents.

[5] https://walterliving.com/city/alkmaar
[6] https://www.ing.nl/zakelijk/economie/nederland/

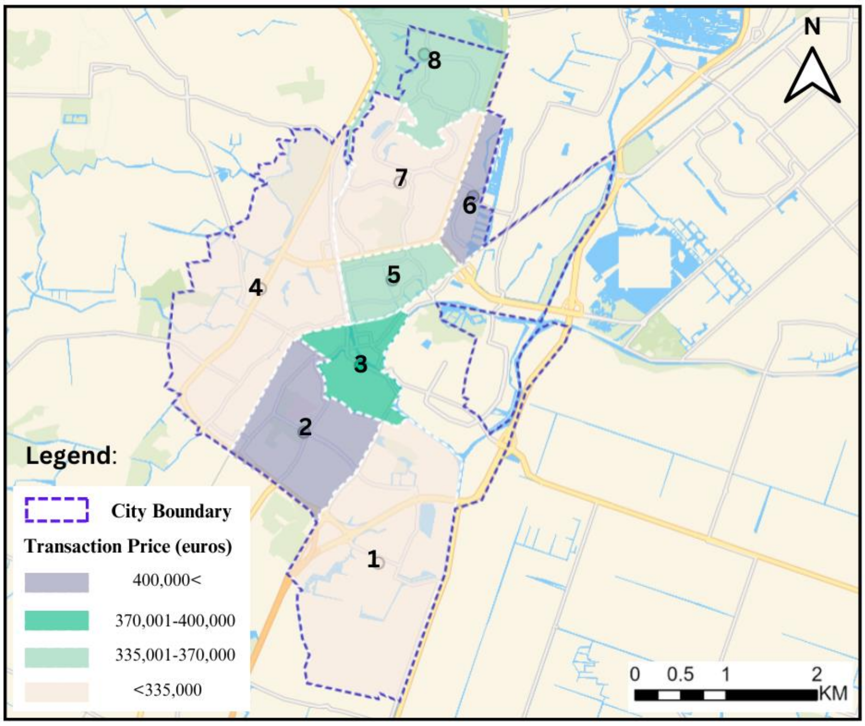


Figure 10: Neighbourhood No. in Alkmaar
Source: https://walterliving.com/city/alkmaar

Table 8: Transaction price per neighbourhood in 2023.

Neighbourhood No.	Neighbourhood name	Transaction price (Euros)	Transaction price square meter
1	Overdie	331000 Euros	3900 Euros/M²
2	Zuid	431000 Euros	4400 Euros/M²
3	Centrum	375000 Euros	4600 Euros/M²
4	West	318000 Euros	3800 Euros/M²
5	Huiswaard	339000 Euros	3500 Euros/M²
6	Veroonermeer	468000 Euros	3700 Euros/M²
7	De Mare	321000 Euros	3300 Euros/M²
8	Daalmeer/Koedijk	361000 Euros	3500 Euros/M²

Source: https://walterliving.com/city/alkmaar

According to Figure 11, the transaction prices of different types of houses are totally different, especially detached houses, which have higher transaction prices compared to others each year. Even though the average transaction price of houses decreases from 2022 to 2023, this trend is different for each type. Moreover, a slight transaction price increase for all types can be observed at the end of 2023. The fluctuations of houses after the summer of 2022 to the middle of 2023 are more moderate, except for detached houses. Detached houses experience significant fluctuation with a noticeable peak and drop from close to 800,000 euros to 650,000 euros. This category has the most considerable variability among all types in these years. The transaction price fluctuation of the three types of corner, apartment, and terraced houses are fairly near one another.

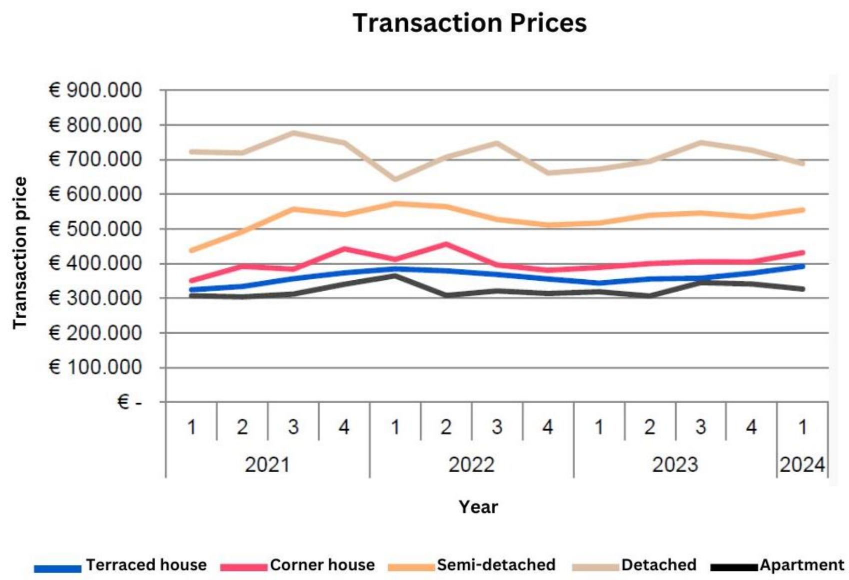


Figure 11: Transaction price based on the housing type in different seasons from 2021 to 2024

Source: (NVM, 2024)

Figure 12 shows that the transaction price development had a decreasing general trend from 2022 to 2023. However, this trend was also different among different types of houses. For instance, the transaction price development of semi-detached houses reached -15% in the spring of 2023, while this development was around 10% in the detached category. This trend gradually recovered with positive growth for the rest of 2023. The detached house category had the highest irregular fluctuation compared to other types, with sharp declines and also increases. One of the reasons behind the high increase in transaction price development of detached houses, while other properties had negative growth, might be that detached houses are mostly large and private properties, which made these places more desirable during the post-pandemic period when people preferred more private and spacious living environments.

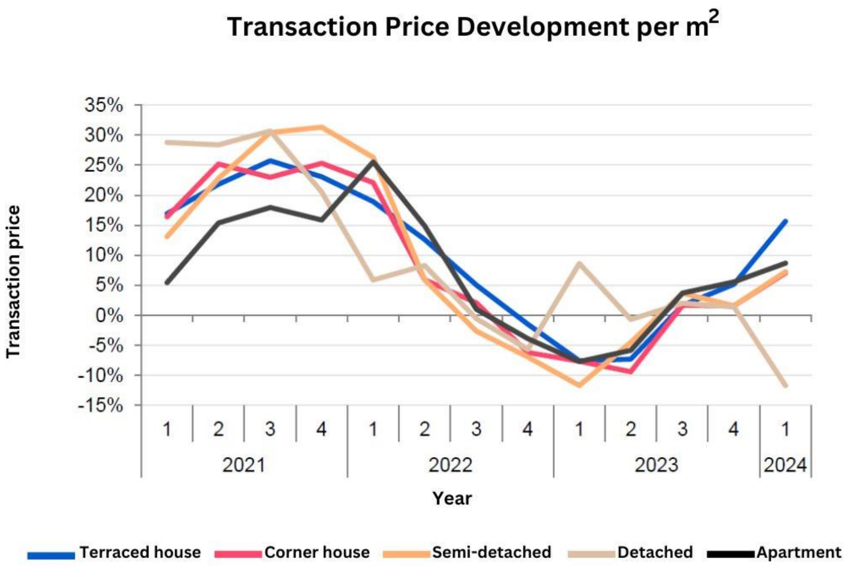


Figure 12: Transaction price development per m2 based on the housing type in different seasons from 2021 to 2024. Source: NVM, 2024.

Appendix 2 shows the property value distribution in Alkmaar. It illustrates that high-value properties (euro/m²) are located in the city centre, and medium-value properties (3000-3660 euros/m²) are clustered in the northern part of the City. This research also has calculated Global Moran's Index to identify the spatial pattern of property value in Alkmaar. Based on the generated report in Appendix 3, the Global Moran's Index is 0.29, showing a positive spatial autocorrelation. It means that similar property values are possibly clustered. Moreover, the Z-score of 8.751 and p-value of 0.000 indicate a clustered spatial pattern and a positive autocorrelation among property values. In fact, there is less than a 1% likelihood that the observed clustered spatial pattern is a result of random chance.

4.2. Different types of green environments

After exploring different types of urban green environments, five approaches -density, size, height, type of vegetation, and services- are chosen to classify green environments. Service and size are the most significant approaches to green environment classification based on the literature. On the other hand, there is limited research on the classification of green environments regarding the height, density, and type of vegetation. Therefore, these approaches are selected to classify green environments in Alkmaar and then find their effects on property value.

4.2.1. Classification based on type of vegetation

Figure 13 is an example of the classification of green environments based on the type of vegetation into four classes: grass, shrubs, evergreen trees, and deciduous trees. This classification is the result of the BGT dataset of vegetation on the PDOK website and the characteristics of the registered trees by the municipality across the city. The geographical pattern of this type of green environment will be discussed in section 5.1.

Figure 13: Classification of urban green environments based on the type of vegetation in Alkmaar.

Source: Author, 2024.

4.2.2. Classification based on height

Figure 14 is an example of the classification of green environments based on the height of vegetation into four classes: high-rise, medium-rise, low-rise, and no-vegetation classes. These classes are the result of CHM for finding the height of vegetation in the city. No vegetation areas consist of roads and buildings. Medium-rise green environments are mostly shrubs, bushes, and small trees. However, the high-rise vegetation areas are the tall trees allocated to the canopy of trees. The tree canopy also has the highest vegetation value. Figure 15 is an example area showing the detected tree canopies with a resolution of 25cm in Alkmaar. It shows that tree canopies consist of diverse vegetation heights, starting with the highest height value to the lowest.

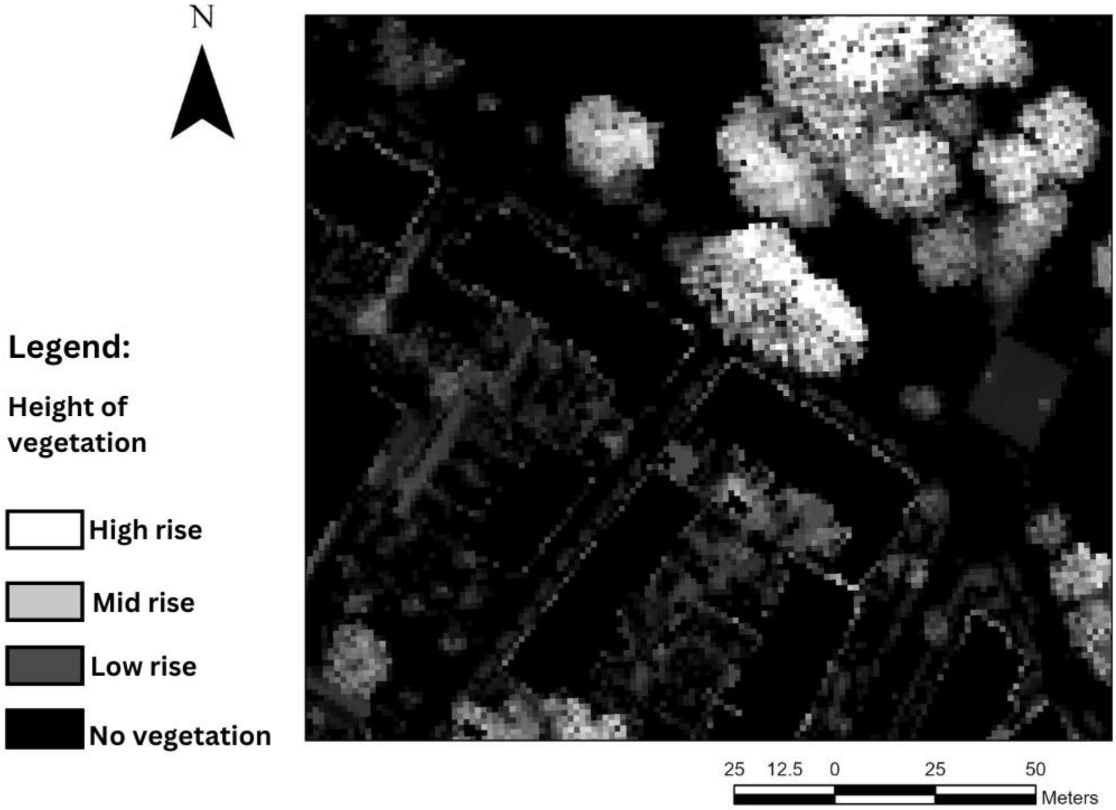


Figure 14: Classification of urban green environments based on the height of vegetation in Alkmaar.

Source: Author, 2024.

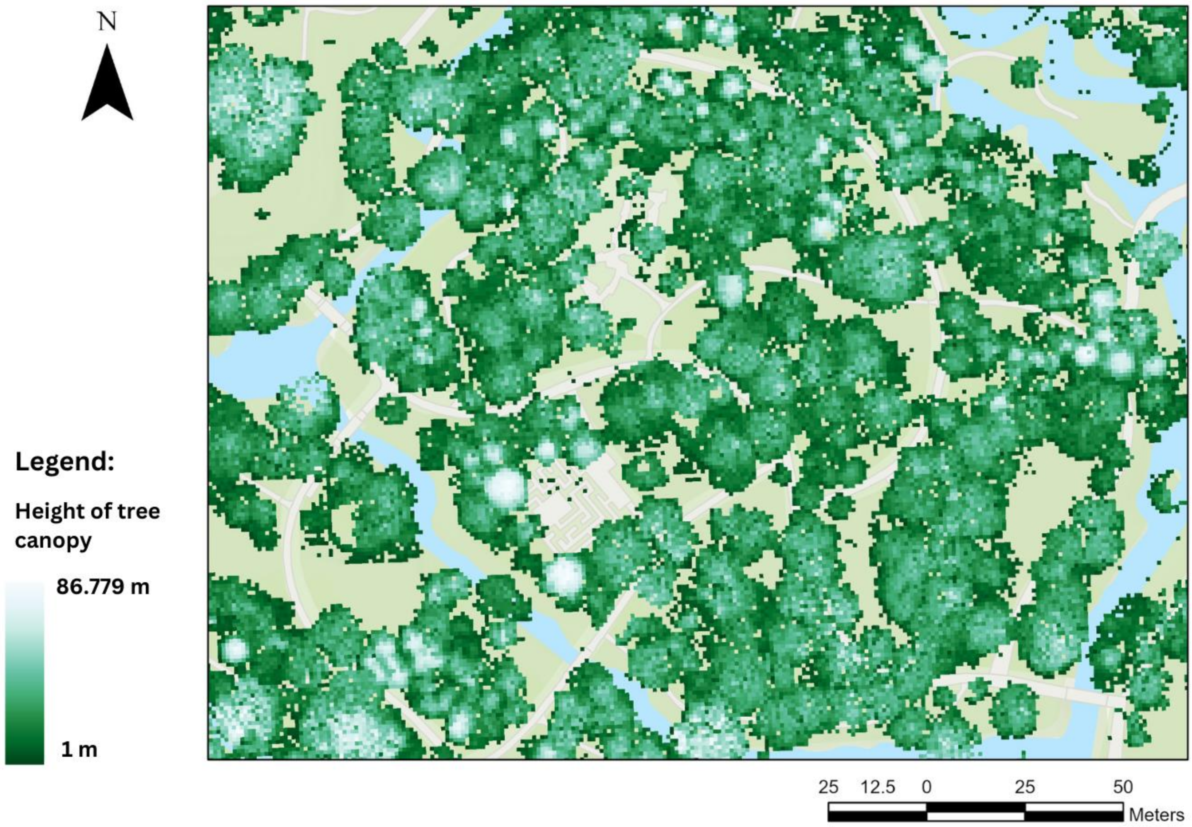


Figure 15: An example area with the detected tree canopies in Alkmaar.

Source: Author, 2024.

4.2.3. Classification based on density

Figure 16 presents an example of the classification of urban green environments based on the density level in Alkmaar. The map is divided into five classes: thick vegetation, dense vegetation, medium-density vegetation, scarce-density vegetation, and no vegetation areas. This classification is from the NDVI calculation method.

4.2.4. Classification based on size

According to the NDVI map (Figure 16), which illustrates the distribution of all the green environments in Alkmaar, the size of green spaces around each property (as a point) is significantly different. After measuring the size of green environments, they are categorized into three categories: small (less than a quarter of a circle around a property), medium, and large size vegetation (more than three-quarters of a circle around a property). A number of properties are surrounded (at a distance of 25m) by the large size of green environments, while there is no vegetation around other properties.

Figure 16: Classification of urban green environments based on the level of density in Alkmaar (NDVI map).

Source: Author, 2024.

4.2.5. Classification based on services

The map in indicates green environment classification based on the services they provide in Alkmaar. The BGT dataset on the PDOK website provided a map of classified green spaces by the services they provide. Green environments are classified based on specific service types, including sports fields, greenhouses, allotments, parks, recreational green spaces, and agricultural lands in Alkmaar. This research explores how distance to each of these classes as a factor affects the property value (Appendix 5). The geographical pattern of this type of green environment will be discussed in section 5.1.

4.3. Spatial analysis of 2D factors

The 2D factors from the category of locational characteristics of the property, such as distance to CBD and distance to train station, are provided in Appendix 6,7 and 8. In addition, air pollution is another important factor in the environmental characteristics affecting property value. The spatial distribution of No_2, as an air pollutant, is in Appendix 9, which shows the spatial distribution of air pollution in Alkmaar.

4.4. Property value modelling

This section aims to develop a model to estimate property value based on combining different types of urban green environments and other 2D and 3D factors affecting property value. Due to the data limitations in measuring the 3D factors for all the properties in Alkmaar, two models for predicting property value are developed. The first is based on the combination of 2D factors and type of green environments, and the second is based on the 3D factors affecting the property value. In the end, the predicted property value is compared with the actual property value to validate the models.

4.4.1. Property value modelling by 2D factors and types of green environments

The value of 2D factors and types of green environments in the distance of 25m of each property is measured and then assigned to the property. OLS, GWR, and RF methods are applied to find the correlation between the calculated factors and the property value. First, the three methods are compared to find the most suitable method for developing a property value prediction model. Then, the impacts of different types of green environments and other factors on the property value are analysed.

4.4.1.1. OLS Model

The hedonic price modelling using the OLS method is executed through multiple regression in SPSS. After the first run, the significance of all the factors except the percentage of shrubs and the percentage of deciduous trees around the properties as two types of green environments was lower than 0.5. Therefore, these two types of green environments were excluded, and the model was executed once again. Through the new OLS model, all the variables have significance values lower than 0.05, indicating that all of these independent variables are statistically significant in the model.

Table 9: The summary of the OLS model

Model	*R*	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*	*Durbin-Watson*
(Constant)	0.404	0.163	0.162	788.465	0.798

Source: Author, 2024.

Table 10: Correlation table of OLS model

Model	*Standardized coefficient*	*t*	*Sig.*	*Pearson's correlation coefficient*	*Collinearity statistics*	
(Constant)	*Beta*				*Tolerance*	*VIF*
Size of property	0.009	1.982	<0.05	-0.023	0.807	1.238
Age of property	-0.044	-8.752	<0.001	0.094	0.708	1.412
Distance to schools	-0.106	-17.050	<0.001	-0.115	0.459	2.178
Distance to roads	-0.103	-20.401	<0.001	-0.047	0.698	1.433
Distance to hospitals	-0.168	-29.835	<0.001	-0.163	0.561	1.781
Distance to amenities	-0.177	-21.590	<0.001	-0.218	0.265	3.772
Distance to bus stations	0.103	21.946	<0.001	0.102	0.813	1.230
Distance to business	0.072	10.208	<0.001	0.048	0.355	2.821
Distance to water	-0.054	-11.283	<0.001	-0.012	0.784	1.275
Distance to parks	-0.067	-11.393	<0.001	-0.028	0.516	1.938
Distance to sports fields	-0.139	-14.779	<0.001	0.072	0.200	4.989
Distance to agricultural lands	-0.265	-26.224	<0.001	0.206	0.175	5.701
Distance to allotment	0.180	18.659	<0.001	0.145	0.192	5.205
Density of vegetation	0.134	13.275	<0.001	-0.163	0.177	5.665
Height of vegetation	-0.030	-1.976	<0.001	-0.055	0.445	2.245
The percentage of grasses	0.013	1.450	<0.05	-0.158	0.212	4.706
The percentage of deciduous trees	-0.021	-2.883	<0.05	-0.050	0.911	1.098

Source: Author, 2023.

Table 11: ANOVA test analysis

Model	*Sum of squares*	*df*	*Mean Square*	*F*	*Sig.*
Regression	5674930401.8	25	226997216	365.1	0.000
Residual	29147999747.1	46886	621678.1		
Total	34822930149	46911			

Source: Author, 2023.

According to Table 9, the R square value is 0.163, meaning that the OLS model can explain 16.3% of the variation in property value. This value is not significant enough to generalize the model. However, other assumptions, including Durbin-Watson, F-test, and VIF, are tested to ensure this model is a great fit to predict the property value. First, the Durbin-Watson value, which represents the independence of error, is lower than one (Table 9). In fact, the Durbin-Watson value is far away from the optimal value of two (no autocorrelation), which means that the error residuals in this OLS model are not independent and need a strong positive correlation.

Second, the F-test in the ANOVA table analysis (Table 11) is more than one, meaning that the variance explained by the model (SS_M) is significantly greater than the error within the model (SS_R). To be more specific, this test indicates that the OLS regression model is better at predicting the property value than using the mean value (Table 10). The last assumption for checking the generalization of the model is collinearity diagnostics by calculating the VIF value. The VIF values should be below 10, and the tolerance statistics should be above 0.2. Based on the collinearity Statistics, the VIF values of a few variables, including air pollution, distance to CBD, distance to train stations, distance to agriculture, distance to the greenhouses, and size of vegetation, are above 10, indicating that there is a significant multicollinearity among these independent variables and these variables are excluded from the model.

Table 10 demonstrates the contribution of each variable in predicting the property value by model. Among the different types of green environments, the standardized coefficient (Beta) value of the density of vegetation was 0.134. In fact, by increasing the density of vegetation (one square meter) in the distance of 25m surrounding the properties and holding the other variables constant, the property value increases by 0.135 standardized coefficients in Alkmaar. Regarding the classification of green environments based on service, distance to agricultural lands and sports fields had the highest negative correlation with property value in Alkmaar.

In conclusion, based on all of the assumption tests above, the OLS model does not significantly fit the data and can not be generalized to the property value prediction for the city of Alkmaar. Thus, the GWR regression model is executed to find the best property value model.

4.4.1.2. GWR Model

Table 12 illustrates the descriptive statistics of the coefficients of the variables in the GWR model. Beta is the coefficient of the variables in the model. Based on the limitations of finding the significance of these variables in GWR modelling, it is assumed that all the variables are significant in the same way as the OLS model. According to Table 12, the size of the green environment variable has the highest negative coefficient value, meaning that this factor has a more substantial and varied effect on property values when standardized. However, there is a significant difference between Beta min and Beta coefficient max. It implies that there are spatial differences in how the size of green environments affects property values across the Alkmaar. several areas have positive effects, while other areas have negative effects. This correlation is similar to the correlation between the percentage of deciduous trees as a type of vegetation around the properties.

Among the classifications based on the service factors, the first is that the distance to recreational green areas is more significant compared to the other variables, with a standardized coefficient of 161.410. This is a positive link, meaning that property value increases when the property is located close to recreational green areas. Afterwards, the distance to agricultural lands variable with the negative standard coefficient value of 92.654 is significant. This means that increasing the distance from agricultural lands enhances the value of properties. It should also be mentioned that among these two types of green environments, the Beta standard deviation value of distance to agricultural land is higher than the distance to the recreational green areas, meaning that there was more considerable variability in the effect of distance to agricultural lands on the property value across Alkmaar. The spatial effects of distance to agricultural land on property value are totally different.

Table 12: Descriptive statistics of GWR estimation coefficients

Factor	β mean	β min	β max	β standardized	β standard deviation
Size of property	-310.737	-3321.14	2407.616	0.007	677.297
Age of property	-119865.92	-5835455575	769895885	-1422.431	27853344.554
Air quality	199.0396	-71320.3	486528.5	-1.131	4095.415
Distance to schools	-35.894649	-136752.3	13842.3527	-0.285	2322.762
Distance to roads	-29.7866	-4752.38	3540	-0.061	515.763
Distance to hospitals	-2.47924	-12214.5	135777.1	-0.477	2447.999
Distance to amenities	81.22543	-15324.2	12726.7	-0.374	1824.040
Distance to CBD	-252.577	-7927883	20565707	-391.486	548453.942
Distance to train stations	-331.018	-15925581.73	5533028	-66.942	389710.106
Distance to bus stations	185.5993	-57563.1	57908.4	0.370	3116.568
Distance to business	2137.415	-1399850	3851584	7.062	84518.595
Distance to water	11.4532035	-32400.86458	18751.2838	-0.089	1427.412
Density of vegetation	6.006602	-3434.97	3131.916	0.076	489.976
Distance to parks	143.4493	-12593.7	9995.913	-0.127	1640.825
Distance to sports fields	621.5497	-833551	678832.1	-9.547	59177.063
Distance to recreational green spaces	-8417.35	-4968952	2293015	161.410	280948.407
Distance to greenhouses	6684.565	-1989926	4177650	68.473	204847.472
Distance to agricultural lands	2211.443	-3014577	3231764	-92.654	301245.567
Distance to allotments	1211.544	-2054247	2240845	33.027	158091.326
Height of vegetation	26.57692	-4027.57	3889.655	-0.017	499.996
The percentage of grasses	17339.56	-3647733064	3089766975	465.503	30851676.410
The percentage of deciduous trees	-368614	-18991691016	10585499285	-3437.543	141035251.136
Size of green environments	-32961.8	-5322924626	6284167063	-12337.524	53149206.197
Local R^2	0.579	0.004	0.872	-	0.137
Intercept	-81942.134	-248817056	1473714880	1270.544	21893514.695
AICc	701682.619				
Adjusted R^2	0.681				

Source: Author, 2024.

The R square value of the GWR model is 0.675, meaning that 67.5% of the variance in the property values can be estimated by the variables in the GWR model (Table 13). This is a strong correlation between independent variables and property value as the dependent variable. The adjusted R square value is also 0.681, indicating the GWR model is well-fitted, and independent variables meaningfully explain the spatial distribution of property value in Alkmaar. Figure 17 illustrates the distribution of the local R^2 GWR model of Alkmaar. According to this figure, the local R^2 value of the GWR model varied from 0.237 to 0.872, and the mean local R^2 is 0.569. Moreover, Figure 17 shows that the local R^2 of properties in the city centre and southern area of Alkmaar are lower than the surroundings and northern areas of Alkmaar. In fact, by moving toward the north area or city surroundings, the local R^2 value increases, indicating that the simulation of the GWR model in these areas is better than that of the city centre. In fact, the explanatory independent variables, specifically in the GWR model, strongly influence the property value in these areas compared to other areas in Alkmaar. However, few other properties with low local R^2 were randomly among properties with high Local R^2 values.

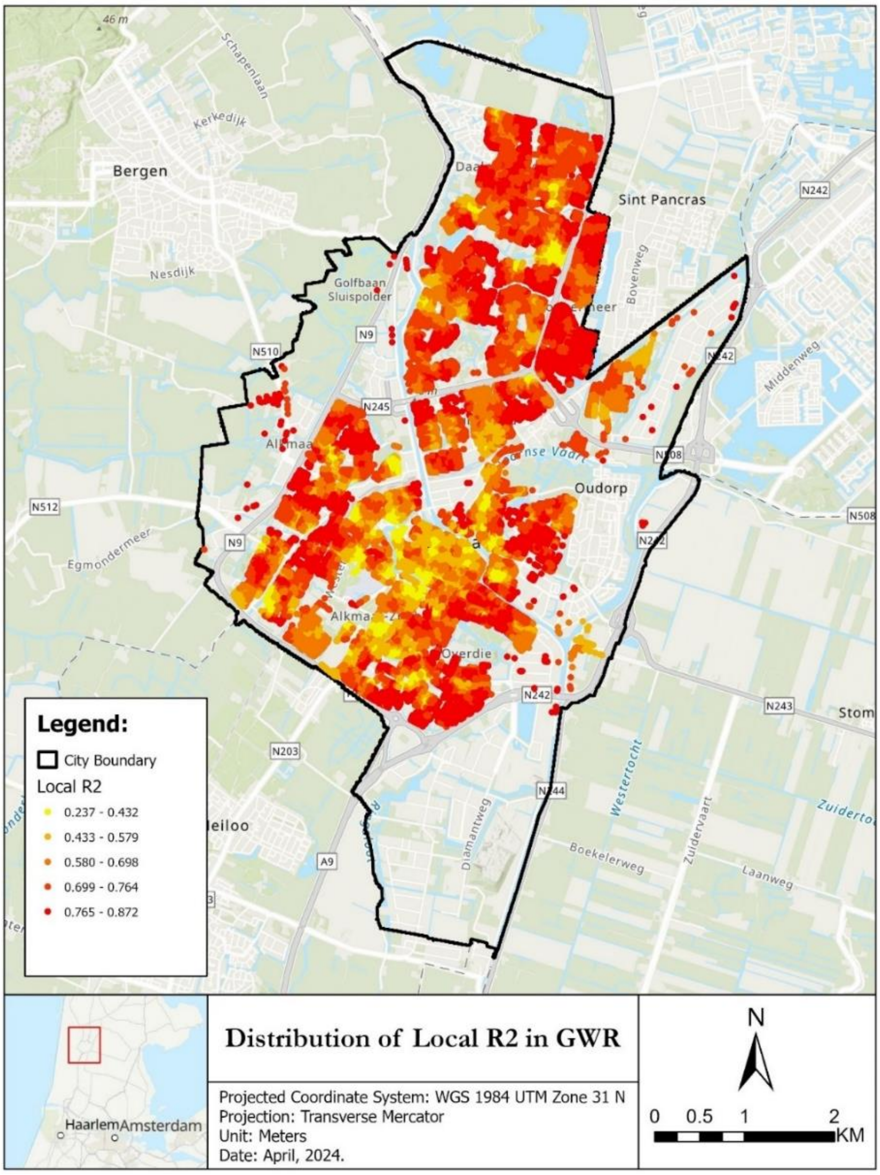


Figure 17: The distribution of local R^2 in GWR

Source: Author, 2024.

The OLS and GWR models are compared using an ANOVA test (Table 13). The adjusted R^2 has increased from 0.162 in OLS to 0.681 in GWR, meaning that the GWR model explains the variation in property value more than the OLS model. Moreover, the residual sum of squares (SS_R) in GWR to predict property value has decreased significantly compared to OLS. In conclusion, GWR is more well-fitted than OLS in predicting the property value in Alkmaar.

Table 13: The results of OLS and GWR models comparison

Model	***R square***	***Adjusted R square***	***SS_R***
GWR	0.675	0.681	26258167935.2
OLS	0.163	0.162	29147999747.1

Source: Author, 2024.

4.4.1.3. Random Forest Model

The Random Forest model is applied to explore the effects of the 2D factors and different types of green environments on the property value. The correlation between different indicators is explored using Pearson's correlation. The correlation matrix heatmap (Appendix 12) shows the coefficients of correlation between the variables in the model. This figure illustrates a high correlation between the distance to the CBD and the distance to the train station variables since the city had two train stations close to the central business district. In this situation, the distance to the train station factor was omitted among the variables affecting property value.

Table 14: The results of the Random Forest model performance

R-square	0.858
Adjusted R-square	0.831
MAE	173.91
RMSE	327.86
Pearson R	0.926

Source: Author, 2024.

Table 14 provides the metrics to evaluate the performance of the RF model. The model explains 83.1% of the property value variation based on the adjusted R-square value. Comparing the R-square values of the OLS, GWR, and RF models, the RF model has the highest value, meaning that this model fits the data well to predict the property value of Alkmaar. MAE, measuring the accuracy of the model, illustrates that the average difference between the actual WOZ property value and the predicted value by the model is 173.91 (Euro/ m^2). An RMSE of 327.86 represented the standard deviation of the prediction errors. These results imply that there is a strong correlation between the test and trained data in the RF model. In fact, the 2D factors and types of green environment variables used in the RF model are significant determinants of the property value prediction model in the city of Alkmaar.

Feature Importance Analysis:

Two methods of Mean Decrease in Impurity (MDI) and Permutation Importance were used to assess the importance of each variable (2D factors and types of green environments) in the property value prediction model. In the MDI Impurity method, one independent variable is omitted in each decision tree, and then the Gini index (Impurity) is measured. In the Permutation Importance method, each variable's value is shuffled randomly, and the difference in the model's performance is calculated. A higher Gini index value and Permutation Impurity value show that the variable is of higher importance in the model. Table 15 represents the results of feature importance based on these two methods. The size of the property, distance to CBD, and distance to the hospital are the most important 2D variables in the model that affect the property value in Alkmaar. Among different types of green environments, the density of vegetation around the properties is a substantial factor in predicting the property value. Moreover, among different types of green environments based on the services, distance to recreational green areas and also greenhouses are two critical features in the property value model. The results of the feature importance analysis are illustrated in a figure in Appendix 13.

Table 15: Feature importance analysis results

Factor	***Mean Decrease in Impurity***	***Permutation Importance***
Size of property	0.281	0.245
Distance to CBD	0.101	0.069
Age of property	0.082	0.071
Distance to greenhouses	0.056	0.037
Distance to hospital	0.035	0.028
Distance to recreational centres	0.031	0.026
Distance to allotment	0.025	0.009
Distance to amenities	0.025	0.031
Density of vegetation	0.021	0.028
Distance to business centres	0.021	0.027
Distance to agricultural lands	0.019	0.002
Distance to park	0.019	0.009
Distance to school	0.018	0.013
House type 2 (semi-detached house)	0.017	0.019
Distance to sports centres	0.018	0.007
Height of vegetation	0.017	0.021
House type 5 (Apartment)	0.015	0.008
Air pollution	0.014	0.011
Distance to water	0.013	0.004
Size of vegetation	0.012	0.006
Distance to the bus station	0.012	0.008
Size of grass (Type1)	0.012	0.009
Distance to road	0.015	0.009
Size of Deciduous trees (Type2)	0.008	0.016
House type 4 (Terraced house)	0.003	0.015
Size of shrubs (Type3)	0.002	0.000
House type 3 (Corner house)	0.002	0.008
House type 1 (Detached house)	0.000	0.000

Source: Author, 2024.

Partial dependence plot:

After finding the importance of different variables in the RF model, the partial dependence plot is applied to interpret the changes in the property value prediction model more precisely. These plots explain the nonlinear correlation between the property value and a specific variable. This section selected six variables that significantly influence the Gini index (Impurity)- three from the 2D factors and three from the types of green environments- for further analysis. The plot of the rest of the variables is in Appendix 14. In the following, the partial dependence plots of selected figures are described.

Figure 18 is the first partial dependence plot showing the relation between the size of green environments (m^2) and property value (Euro/m^2). Besides, Figure 19 illustrates the nonlinear correlation between the property value and the mean height of vegetation (trees, grasses, and shrubs) at a distance of 25m around the properties. Figure 21 is also a partial dependence plot showing the relation between the percentage of a deciduous tree area at a distance of 25m around the properties with property value (Euro/m^2). The nonlinear relationship illustrates that overall, there is a positive correlation between the size of deciduous trees as a type of vegetation and property value. These partial dependence plots are discussed in section 5.2.

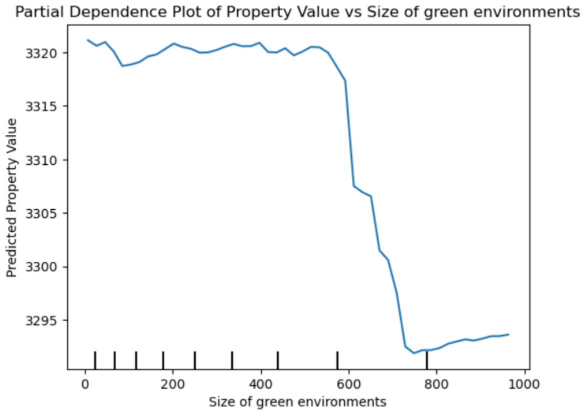


Figure 18: The partial dependence plot of property value vs size of green environments around the property
Source: Author, 2024.

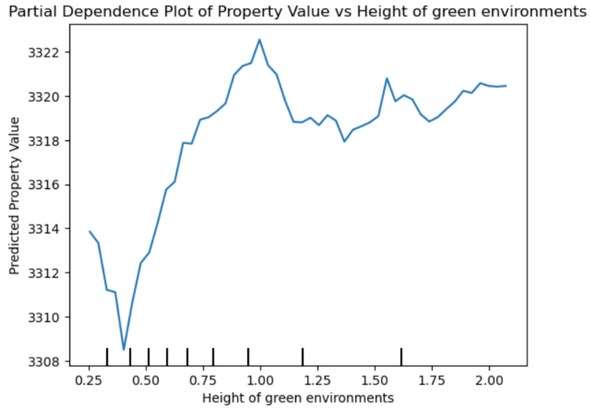


Figure 19: The partial dependence plot of property value vs height of green environments around the property
Source: Author, 2024.

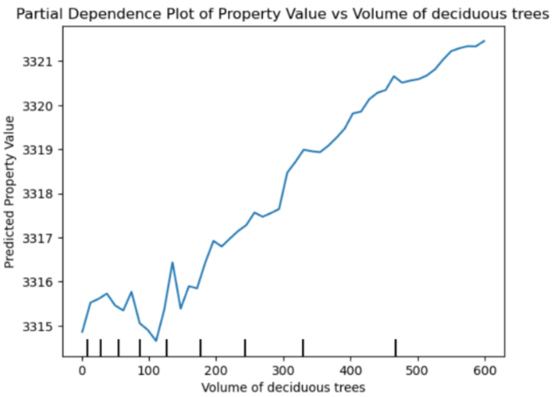


Figure 21: The partial dependence plot of property value vs size of deciduous trees around the property
Source: Author, 2024.

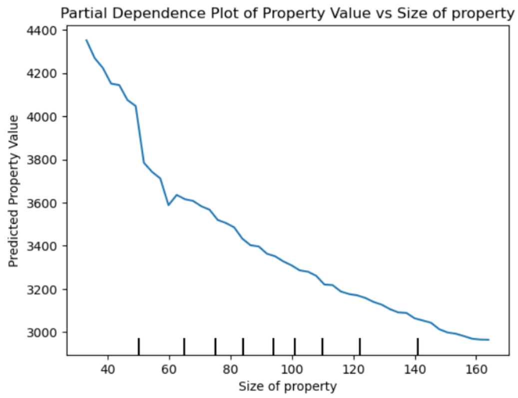


Figure 20: The partial dependence plot of property value vs size of the property
Source: Author, 2024.

Figure 20 is a partial dependence plot illustrating the relation between the Size of properties (m^2) as a 2D factor with property value (Euro/m^2). The nonlinear negative relationship indicates that as the size of the property increases, the predicted property value decreases. In addition, Figure 22 is a partial dependence plot illustrating the non-linear correlation relationship between the distance to amenities (m) as a 2D factor with property value (Euro/m^2). In the end, Figure 23 demonstrates the relation between the distance to CBD (m) with property value (Euro/m^2).

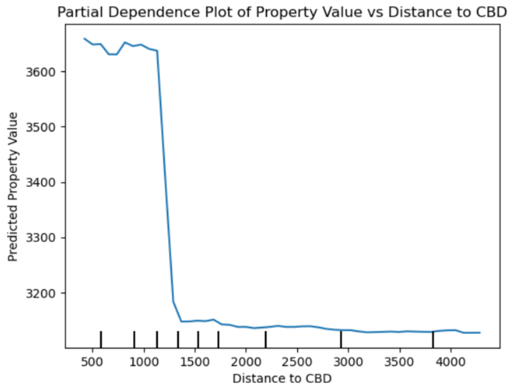


Figure 23: The partial dependence plot of property value vs distance to CBD
Source: Author, 2024.

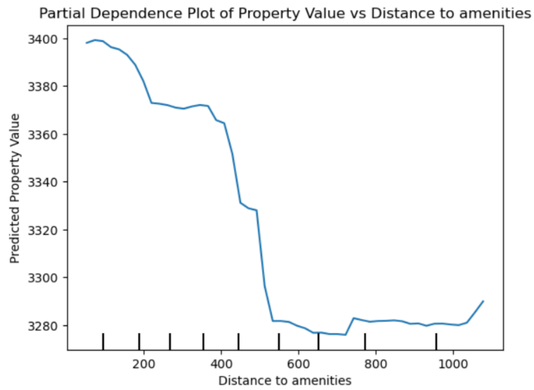


Figure 22: The partial dependence plot of property value vs distance to amenities
Source: Author, 2024.

Comparison of GWR, OLS, and RF models:

Table 16 shows the comparison of the three GWR, OLS, and RF model results. The R-square value of the RF model is higher than the two others. It indicates that the combination of 2D factors and different types of green environments strongly explain the variations in property value compared to the OLS and GWR models. Therefore, the RF is the most suitable method for modelling the property value in Alkmaar.

Table 16: The results of comparing OLS, GWR, and RF models

Model	***R square***	***Adjusted R square***
OLS	0.163	0.162
GWR	0.675	0.681
RF	0.871	0.871

Source: Author, 2024.

4.4.1.4. Modeing validation: K-fold cross-validation

After simulating the property value prediction model through random forest, the next step is to test the model with the testing dataset. In this section, two validation is conducted. First, the model is tested using 20% testing data for all properties. Second, the model is tested by comparing the model's predicted value with the actual property value of 30 properties randomly selected.

The K-fold cross-validation method is applied to validate the model. This method selects the k value of 10 to validate the model in this research. First, the dataset is split into 10 folds, 9 for training the data and only 1 for testing the data. This process is repeated 10 times, and each time results in the corresponding correct rate. In the end, the performance of the RF model is evaluated by calculating the average of the correct rate of all of the 10 steps. The final k-fold cross-validation value of the RF model was 1285.439 Euro/m², with an error percentage of 8.5 %, meaning that the model predicts the property value with an 8.5% error range.

The second RF model validation is by comparing the predicted property value with the actual property value obtained from a real estate platform. Since there are a small number of sample points for this validation, the LOOCV validation method (Leave-One-Out Cross-Validation) is applied to validate the performance of the model by 30 property data. In this method, each data point is selected once to be the test set, and the remaining data points are the training set to fit the model. In the end, the average estimation error in each field is 1640.031 Euro/m² with an error percentage of 12.1%, meaning that the RF model can predict the actual value with a 12.1% error range.

4.5. Property value model by 3D factors

This section aims to develop a model to estimate property value based on the 3D factors affecting property prices. The factors considered in this section consist of view to the road, buildings, vegetation, water, sky, property orientation, property sunlight, property height, and height of vegetation around the properties. Due to the data limitations for measuring the 3D factors for all the properties, this modelling is applied to 28 randomly selected properties.

Property value modelling is performed in SPSS using the OLS model. Table 17 demonstrates the results of the OLS model. The R square and adjusted R square values in the OLS model based on the selected factors are 0.177 and 0.169, meaning that the model explains the 16.9% of the property value variation. Even though this value is not significant enough to generalize the OLS model for the whole city, this correlation is similar to the OLS model in the previous steps. Table 18 illustrates the contribution of each factor in the prediction of property value.

In the first step, the effects of all the factors as independent variables are measured on the property value as a dependent variable. However, the significance of orientation and view to water variables are above 0.05, meaning that these variables have no significant effect on property value and should be excluded from the model.

Then, the Durbin-Watson value and Collinearity Statistics are analysed to test the OLS model's generalization. According to Table 17, the Durbin-Watson value is above one and close to two, meaning that the error residuals in the current OLS model do not need any autocorrelation, and errors are independent. In addition, according to Collinearity Statistics (Table 18), the VIF value of all the variables was below 1, meaning that variables are not correlated and there is no multicollinearity in the model.

The formula to calculate the property value based on the 3D factors is summarized below:

$$Property\ value: 2666.478984 - (0.195 * View\ to\ Sky) - (10.097 * View\ to\ Building) + (16.402 * View\ to\ Vegetation) + (2.405 * View\ to\ Road) + (7.534 * Building\ Height) - (172.897 * sunlight) + (847.037 * Vegetation\ Height) \tag{6}$$

Table 17: Summary of the OLS model

Model	***R***	***R-Square***	***Adjusted R-Square***	***Std. Error of the Estimate***	***Durbin-Watson***
(Constant)	0.421	0.177	0.169	935.145	1.846

Source: Author, 2024.

Table 18: The correlation table of the OLS model

Model	*Standardized coefficient*	*t*	*Sig.*	*Pearson's correlation coefficient*	*Collinearity statistics*	
	Beta				*Tolerance*	*VIF*
(Constant)		3.698	<0.05			
View to Sky	-0.004	-1.005	<0.05	-0.450	0.265	9.480
View to Building	-0.149	-6.252	<0.05	-0.091	0.217	8.513
View to Vegetation	0.133	10.341	<0.05	0.209	0.269	3.715
View to Road	0.280	7.064	<0.05	0.072	0.218	4.591
Building Height	0.054	2.244	<0.05	0.045	0.824	1.214
Sunlight	-0.092	-3.365	<0.05	0.006	0.650	1.539
Vegetation Height	0.384	12.558	<0.05	0.378	0.823	1.215

Source 1: Author, 2024.

The LOOCV validation method (Leave-One-Out Cross-Validation) is applied to validate the performance of the model. In this method, each data point is selected once to be the test set, and the remaining data points are the training set to fit the OLS model. This process is repeated for every data point, and the property value and the one were recorded. In the end, the results of each validation were average to validate the performance of the model (Appendix 15).

For this validation, the actual property value of 28 properties is obtained from the Funda website[7] which is a real state platform in the Netherlands. This data is acquired randomly and manually to calculate the accuracy of the model. The real state value is different from the value of the properties (WOZ data) since it reflects market conditions such as supply, demand, or market fluctuations. Thus, the LOOCV method was applied to validate the model using the predicted property value from the OLS model and the actual property value.

According to Figure 24, only the predicted value of three properties (points No. 20, 22, 26) is higher than the actual value. Moreover, the MAE of half of the predictions (Appendix 15) is less than 1000 Euros. However, one of the largest errors was -1623.317 Euros since it was a property in a building surrounded by high-rise vegetation and also had a very good view of the sky, which led to a high predicted property value. On the other hand, this property has a lower actual value due to its small size and location in the surrounding city, which are ignored in the 3D factor property model.

There is also another property with an MAE value of 2373.734 Euros, showing that the actual property value is higher than the predicted property value. It is a building in a dense area close to the city centre. It is surrounded by high-rise buildings, has high views of buildings, and is surrounded by less vegetation, resulting in low predicted property value. However, it is close to the city centre and is within good distance of different facilities and parks, which results in a large actual property value. This error demonstrates that other important factors affecting property value are not in the current property value developed by only 3D factors. In the end, the average standard deviation is 674.846 Euros, and the error percentage of the LOOCV is 17.01%, indicating that the property model based on the 3D factors predicts the property value 83% times accurately.

This research also compared the predicted property value of the 28 selected properties (the OLS model) with the predicted value from the developed model by 2D and types of green environments factors (RF model in section 4.4.1.3). The average predicted property value of the OLS model is 3290.367 Euros, while the average predicted property value of the RF model is 3651.873 Euros. However, the actual value of these 28 properties is 3965.214 Euros. These results illustrate that 2D factors and types of green environments

[7] https://www.funda.nl/

are important factors in estimating the property value compared to the 3D factors, which should be considered significantly in property value modelling.

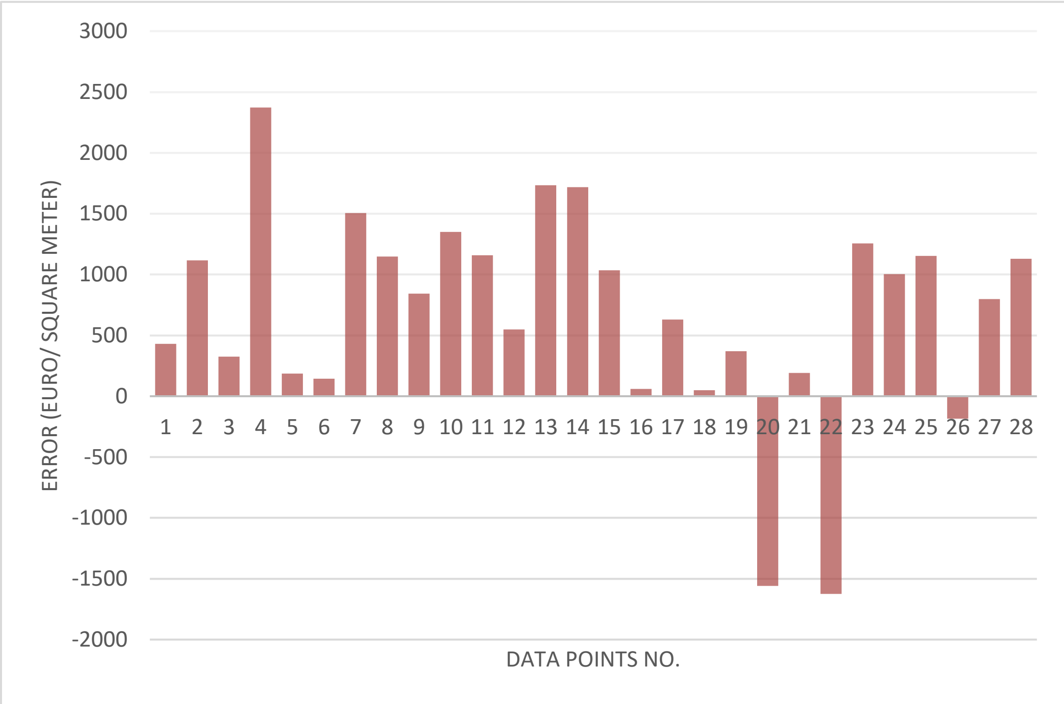


Figure 24: The error between predicted value and actual property value

Source: Author, 2024.

4.6. Summary

The chapter is the result of the implementation of the research workflow. After the development of three models through the OLS, GWR, and RF methods, the RF is selected as the most suitable method for modelling the property value in Alkmaar. This model illustrates that there is a non-linear correlation between different types of green environments and property values. Besides, an OLS model based on 3D factors is developed to predict property values. In the end, the developed models are validated by comparison with actual property value.

5. DISCUSSIONS

In this chapter, the main results and findings of the research are discussed in more detail in the context of the reviewed literature. Section 5.1 reflects the approaches to the classification of green environments in the city of Alkmaar. Section 5.2 summarizes the findings of property value modelling by the combination of 2D factors and types of green environments represented in section 4.4 of the research. Section 5.3 of this chapter also summarizes the results of property value modelling using 3D factors and compares this model with the previous one. In the end, the main limitations of this research are discussed.

5.1. Reflection on the classification of urban green environments and geographical pattern of types

After reviewing the literature, five approaches for the classification of urban green environments in the city of Alkmaar are selected. First was a classification of urban green environments based on their services. People prefer to reside near green spaces to take advantage of the services they provide (Panduro & Veie, 2013; Mathey et al., 2021). They are classified into five classes: sports fields, agricultural lands, parks, allotments, greenhouses, and recreational green spaces in Alkmaar. To analyse the geographical pattern of this type of green environment in Alkmaar, agricultural lands are located on the outskirts of the city in less urbanized areas. Parks are scattered evenly inside the city, showing properties are mostly close to parks. Besides, most of the sports fields and recreational green spaces are located in the city's surroundings; however, sports fields are clustered in Alkmaar. In the end, there are only a small number of greenhouses and allotment green environments inside the city. This research explores how distance to each of these classes as a factor affects the property value (Appendix 4).

The second classification is based on the height of urban green environments, and the third one is based on the density of vegetation. High-rise and dense vegetation increases the environmental quality around the properties, which will increase the property value (Kaloustian & Bechtel, 2016). However, these types of green environments provide more shadow on properties, which might negatively affect the property value (Gupta et al., 2012).

The city centre of Alkmaar consists of mostly scarce-density vegetation. This is due to the morphological pattern of the city centre of Alkmaar, which is characterized by dense buildings and lacks space for vegetation. Most of the green environments in Alkmaar have a medium level of density. However, small portions of thick and dense vegetation are also observed in Alkmaar. Especially in the northern part of the city outskirts where larger green spaces and most of the agricultural lands are located. The scarce-density vegetation is also around the water canals and roads in the city. Meanwhile, medium-density vegetation is inside the neighbourhoods. To conclude, the areas with different levels of vegetation density are interconnected in Alkmaar. Thus, the average level of density should be considered around the properties to find the effect of this type of green environment on property value.

Regarding the height factor, no vegetation areas consist of roads and buildings. Medium-rise green environments are mostly shrubs, bushes, and small trees. These types are found along the river, small roads, and also housing gardens. Low-rise green environments are scattered throughout the city, and there are mostly grasses, agricultural lands, and sports fields. Meanwhile, the high-rise vegetation areas are the tall trees allocated to the canopy of trees. The Tree canopies also consist of diverse vegetation heights, starting with the highest height value to the lowest. In conclusion, there are green environments with different heights around the properties in the whole city, which should be considered as a factor affecting the property value. Thus, these two classification approaches are critical in exploring how they affect the value of property in Alkmaar.

The fourth approach is the size of green environments surrounding the properties. In fact, proximity to green spaces is an influential factor affecting property value, but these effects differ in various sizes of green spaces. As Czembrowski & Kronenberg (2016) indicated, distance to parks above 10ha in size is the most significant factor affecting the property value positively. The size of green spaces around each property is significantly different across the city. A number of properties, especially on the city outskirts, are surrounded (at a distance of 25m) by large green environments, while there is no vegetation around properties in the city centre. Thus, the size of green environments in Alkmaar should be considered to determine whether or not it affects property value.

The last classification is based on the type of vegetation, such as deciduous, evergreen trees, shrubs, and decorative lawns (Mathey et al., 2021). This research classifies green environments into four classes: grasses, shrubs, evergreen trees, and deciduous trees in Alkmaar. Deciduous trees dominate the city, covering a significant portion of Alkmaar but cluster in a few areas. On the other hand, there is a limited number of evergreen trees compared to the deciduous trees in Alkmaar. Evergreen trees are scattered without any specific pattern in the city. Moreover, Shrubs are mostly located along the water canals in the city, and grasses are scattered throughout the area in large patches. It should also be noted that trees and grasses are also around the shrubs. Thus, Alkmaar is covered with diverse types of vegetation specified in different parts of the city, making this factor critical in the property value analysis.

5.2. Reflection on the results property value modelling by the 2D factors and types of green environments

Three models of OLS, GWR, and RF are applied to find the effects of 2D factors and types of green environments on the property value together. There are limitations to exploring the data for all the 3D factors. Thus, they are analysed separately and are reflected in section 5.3. In the following, each property value model and the most effective types of green environments on the property value in the model are discussed.

First, the OLS model, as the most common method of property value analysis (Wittowsky et al., 2020; Wu et al., 2015), is applied. According to the adjusted R square value, this model explains 16.3% of property value variation. However, each type of green environment has different correlations with the property value in this model. The size of green environments at a distance of 25m has negative effects on property value. The main reason is that most of the expensive properties are located in the city centre of Alkmaar, which has mostly small-size vegetation. This result is in conflict with what Czembrowski & Kronenberg (2016) explored. They demonstrated a positive correlation between property value and the size of green space in their research.

The second factor with the highest positive correlation is the distance to recreational green spaces, meaning that people would like to locate around this type of green environment to benefit from this service (Panduro & Veie, 2013). However, distance to sports fields and agricultural lands negatively correlated with property value. Even though it is clear that people do not value these types of green environments (classified based on services) equally (Panduro & Veie, 2013), the negative correlation between distance to sports fields and property value is surprising. The main reason behind this is that sports fields are mostly located in the low-density areas in the suburbs of Alkmaar, where there is less accessibility to other facilities.

In addition, the value of properties in areas with high-density and high-rise vegetation is higher than in areas with lower density and height. According to Table 10 in section 4.4, by increasing the density of vegetation (1 square meter) in the distance of 25m surrounding the properties and holding the other variables constant, the property value increases by 0.135 standardized coefficients in Alkmaar. This finding has not been explored in any of the reviewed literature before. It might be because people value properties in areas with vegetation to take advantage of the environmental quality or views (Sirmans et al., 2005). It should be noted that OLS is a linear regression, while the effects of density and height of green environments might be to a specific point, which is analysed in other models. The variables related to the type of vegetation category do not have any specific correlation with property value in the OLS model.

Different assumptions are checked for the generalization of the OLS model. The results indicate that the OLS model does not fit the data and can not be generalized to predict the property value. However, in various studies, the OLS model performance to predict the property value was high (Wittowsky et al., 2020; Wu et al., 2015). This might be due to the fact that there are nonlinear correlations between most of the variables, which are considered linear in the OLS model.

Second, the GWR model is applied, which considers the spatial variability of factors (Cao et al., 2019; Liang et al., 2018). The adjusted R square value is 0.681 (it was 0.162 in the OLS model), indicating the GWR model is well-fitted, and independent factors meaningfully explain the spatial distribution of property value compared to the OLS. This differentiation in the performance of these two models might be because the GWR model considers the spatial variability between Alkmaar data in the model. It is consistent with the results of Cao et al. (2019). The degree of influence of the independent variables, such as distance to sports fields, distance to schools, or the size of green spaces around properties in the GWR model, varies in different neighbourhood areas around the city. One of the main reasons is that a number of factors have a local influence on property value. For instance, distance to parks might have a strong positive effect on property value in the city surroundings compared to the city centre. This might be due to the fact that the city centre of Alkmaar is compact, consisting of a high density of buildings and scarce vegetation.

The GWR model illustrates that the size of the green environment variable has the highest negative coefficient value, meaning that this factor has a more substantial and varied effect on property values when standardized. However, it implies that there are spatial differences in how the size of green environments affects property values across Alkmaar, with areas having a positive effect and also other areas having a negative effect. This correlation is similar to the correlation between the presence of deciduous trees as a type of vegetation around a property with its value.

Among the classifications based on the service factors, the distance to recreational green areas factor is more significant than the other factors in the GWR model. It is a positive correlation, meaning that property value increases when the property is located close to recreational green areas (Panduro & Veie, 2013). Afterwards, the distance to agricultural lands factor with the negative standard coefficient value of 92.654 is significant. Even though the main correlations between each type of green environment and property value of the GWR model are similar to the OLS model, the GWR model implies that there is spatial differences in the effects of each factor on property value across Alkmaar. This correlation in each neighbourhood is different from the other by considering the local variations (Efthymiou & Antoniou, 2013).

Third, the RF model is applied, showing that RF explains 83.1% of the property value variation based on the adjusted R square value, which is higher than the other two methods. This might be due to the variant of non-linear variables, such as the density of green environments around the properties. This finding was consistent with the study of Ho et al. (2021). The RF model investigates the effects of all the factors, especially the types of green environments, on the property value. Based on the feature importance analysis results (Appendix 1), the importance of each type of green environment on the property value model is different. In the following, the partial dependence plots of six factors with high feature importance in the RF model (section 4.4.) are discussed.

Figure 18 shows the nonlinear correlation between the size of green environments and property value. It illustrates that once the size of green environments reaches 600 m^2 in the distance of 25m around the properties, the value of the properties drops significantly, and once it reaches 800 m^2, the property value does not change in relation to this variable. This is because properties surrounded by large green environments are on the outskirts of Alkmaar, where the value of properties is low.

Figure 19 illustrates the nonlinear correlation between the property value and the mean height of vegetation (trees, grasses, and shrubs) at a distance of 25m around the properties. The property value increases considerably as the mean height of vegetation rises from approximately 0.25 to 1m. After the peak of 1m, by increasing the height of vegetation around the properties, the property value starts to fluctuate, showing that the effects of the height of vegetation on property value are not strictly monotonic. It can be concluded that the height of vegetation might positively affect the property value only to a certain point due to the

positive effects, such as improving air quality or enhancing the aesthetic appeal, and beyond this certain point, the correlation fluctuates.
The size of a deciduous tree at a distance of 25m around the properties also has an overall positive nonlinear correlation with property value (Euro/m²) (Figure 21). However, the variation of the property value from 0 to 600m is not that much significant. In addition, there are slight fluctuations in the correlation curve, showing the effects of this variable, which is not constant. These results illustrate that policymakers and city planners should consider trees around the properties in future planning to enhance the property value in Alkmaar.
According to the Figure 20, there is a non-linear negative correlation between the size of properties (m²) as a 2D factor with property value (Euro/m²). This figure indicates that as the size of the property increases, the predicted property value decreases. This is due to the fact that the high-value properties are located in the city centre of Alkmaar, where most of the properties are small. The initial sharp decrease in property value is when the size of properties increases from 40m² to around 60m². However, after the size of 60m², the value declined gradually.
In addition, the relationship between the distance to CBD (m) as a 2D factor and property value (Euro/m²) is shown in Figure 23. The nonlinear relationship indicates that once the distance to CBD reaches around 1200m, the predicted value of properties suddenly drops significantly. It drops from around 3600 (Euro/m²) to 3200 (Euro/m²). Afterwards, the property value is approximately steady and does not change in relation to this variable. This is because properties located near the CBD have better access to different facilities, such as public transportation, services, and shopping centre in Alkmaar, making the properties around them more valuable.
In the end, Figure 22 indicates that properties at lower distances to amenities (such as shopping centres and restaurants) in Alkmaar have higher property value compared to others. The main reason behind this is that properties near these amenities offer more convenient services to people, making these properties more desirable for them. The sharpest drop in the predicted property value (from around 3380 (Euro/m²) to 3280 (Euro/m²)) when the distance of properties from the amenities reaches 400m. Afterwards, the property value fluctuates gradually. In fact, after this distance, the distance to amenities variable is no longer significant, and other factors might play more important roles in property value prediction.
These findings are not explored in any of the reviewed research. As an accurate and more consistent, the RF model predicted the property value by combining 2D factors and types of green environments more precisely. However, the RF model had an error percentage of 12.1%, meaning that the RF model can predict the actual value with a 12.1% error range.

5.3. Reflection on the results of 3D factors

According to the results in section 4.5, view to sky, building, vegetation, road, sunlight, building height, and vegetation height are the most significant 3D factors affecting the property value in Alkmaar. View to water is not significant in the developed model, which could be because Alkmaar contains various water canals. Moreover, property orientation also has no significant effect on property value, which conflicts with the results of Wu et al. (2015), showing that buildings toward the south, southeast, and southwest, which receive more sunlight, are more valuable than others. This might be because of local reasons in the city that should be studied further.
Due to data limitations, this research could not combine 3D factors with 2D factors and types of green environments together to find their effect on property value. Thus, the property model was only created by the 28 random properties. Since there are limited samples, the OLS model is chosen to model the property value. The Adjusted R square value of the current model is 0.177, meaning that this model explains the 17.7% property value variation. This is not as significant as the RF model developed in the previous section. Comparing the effects of these 3D factors on property value in the model, vegetation height has the highest positive correlation with the Standardized coefficient of 0.384. It means that by increasing the 1m height of

trees and canopies in front of the property, its value will increase by 38.4% in Alkmaar, which is a significant high correlation. View to vegetation variable also positively affected property value as a 3D factor. This might be because people would like to benefit from the environmental quality or aesthetic value of green environments around their properties (Sirmans et al., 2005). In other words, by considering these two factors, urban green vegetation has a value-added effect on property within a distance of 25m in Alkmaar.

Other factors are sunlight and view to building factors, which are negatively correlated with property value. By increasing the number of buildings around the properties, they receive less natural light, and the aesthetic appeal of properties will decrease, which will affect the value of the property (Fleming et al., 2018). This correlation means that increasing the view of the property to another building will decrease the value of the property. Also, if the building is in the shadow at 8:30 AM, the value of the property is lower than the ones that are not in the shadow at that time.

In addition, the value of properties with high rise was higher than low rise properties. It might be because the properties with more stories are able to see farther than those with lower levels. However, the view to the sky has a negative low correlation with property value in the model, which is surprising. This might be possible because by increasing the height of vegetation, which has a positive correlation with property value, the sky might be blocked, and there would be less view of the sky.

In the end, based on the generalization tests, the Durbin-Watson value is above one and close to two, meaning that errors in the model is independent. Moreover, the error percentage of the LOOCV is 17.01%, indicating that the property model based on the 3D factors predicts the property value 83% times accurately. The findings of this section illustrated that even though the property model by 3D factors is effective in predicting property value (R^2=0.421), the model by 2D factors and types of green environments explains the property variation more accurately (R^2=0.871).

5.4. Limitations

One of the main limitations of this study was the time limitation and the lack of available data. First, since there was a lack of available data and difficulties in data acquisition regarding the 3D factors for all the properties, the property value prediction model based on 3D factors was developed with limited samples acquired manually. Second, due to the time limitation, other important types of green environments that affect property value, such as the shape of green environments, were not considered during the period of this study. Third, this research explored the effects of types of green environments with other factors on property value from WOZ data and not the real estate value of properties, which is more accurate. Fourth, the Euclidean distance method was applied for the spatial analysis of the 2D factors; however, this method ignores the physical barriers to calculate the distances. The city of Alkmaar also consists of numerous water canals that block the accessibility of properties. Even though Euclidean distance is not an accurate method, the researcher could not focus on alternative methods due to the time limit.

In the validation step, only actual value data of around 30 properties were explored to validate the RF model. However, a large number of actual property value samples could improve the model validation analysis. Moreover, the WOZ data used in this study was acquired in 2022, and there were no data available for 2023. On the other hand, the calculation of all the factors and types of green environments was based on data from 2024.

In the end, the shadow volume analysis for the sunlight factor was implemented only by considering the buildings and not the height of vegetation to make shadows on buildings. This was due to the limitations of the software in the measurement of shadow volume. However, the height of vegetation in front of the properties was a separate factor in the property value prediction model.

6. CONCLUSIONS AND RECOMMENDATIONS

This chapter presents the main conclusions of the current scientific research and recommendations for further research. The conclusions are demonstrated based on the research results regarding each sub-objective. Then, the recommendations for the future are discussed.

6.1. Conclusion

This research presents a model to predict the property value by considering various types of green environments along with other factors affecting the property value. First, through the literature, all the significant types of green environments and their classification approaches are investigated. Afterwards, the green environments in the city of Alkmaar are classified based on size, height, density, service, and type of vegetation. Then, three OLS, GWR, and RF models are performed to find the most suitable modelling method. Since RF has the highest accuracy in predicting the property value based on the types of green environments and 2D factors compared to other methods, it is selected as the most suitable method for modeling. On the other hand, considering the limitations, another OLS model is also performed to find the effects of 3D factors on the property value. This research provides a scientific basis for urban planners and city policymakers involved in urban green environments and is concerned about greening cities and housing affordability in the municipality of Alkmaar.

6.1.1. Research sub-objective one

The first sub-objective is to identify the 2D and 3D factors relating to urban green environments and property value by literature review. First, all the significant factors affecting the value of the properties are explored through reviewing the recent studies. These factors are categorized into three groups: structural, locational, and environmental characteristics of properties. Even though the studies mostly focused on 2D factors, a number of 3D factors, such as sky view factors, property visibility, sunlight, and orientation, have been identified as affecting property value.

Most of the existing literature considers green environments as parks related to property value studies. Hence, further research has been conducted to find other approaches for the classification of green environments and also different types from each classification. They are mostly categorized quantitatively by considering the services and functions of the green environment as the most suitable approach since people mostly prefer to reside near them to take advantage of their services. However, other types of urban green environments have also been discovered based on density, height, size, shape, and type of vegetation. For instance, the classification of urban green environments based on density provided five categories: thick vegetation, dense vegetation, medium vegetation, scarce vegetation, and no vegetation.

6.1.2. Research sub-objective two

The second sub-objective is to apply a method to classify different types of urban green environments affecting property value. Five classification approaches are based on the services, height, density, size, and vegetation types within a distance of 25m around the properties. The services approach is selected since it is the most important and common classification approach affecting property value in the literature. They are classified into parks, recreational green spaces, agricultural lands, greenhouses, allotments, and sports fields.

The four other approaches are size, density, height, and types of vegetation in the distance of 25m around the properties. The main reason behind selecting these approaches is that, at first, there is limited research on the effects of these factors on property value. Second, the data regarding the calculation of each of these factors to measure their effects is more accessible due to the limitations of this city. Third, the city of

Alkmaar is a green city in the Netherlands, and its properties are surrounded by different types of green environments. In recent years, the housing construction rate has accelerated due to the increasing housing demand in Alkmaar. Thus, finding the effects of these types of urban green environments on property value helps policymakers in the process of property valuation.

6.1.3. Research sub-objective three

The third sub-objective is to develop a model to estimate property value based on the combination of different types of urban green and other common factors affecting property value. Hedonic price modelling and machine learning algorithms, as two common methods for modelling property value, are selected for modelling. This research explores the combination of 2D factors with types of green environments on property value separately from the effects of 3D factors on property value. In fact, due to the data limitation for 3D factors, finding their effects on property value is conducted by considering around 30 properties. In contrast, the effects of 2D factors and types of green environments together on property value are calculated for all the available properties in Alkmaar.

First, the OLS regression, the most widely used method for the estimation of the property value, is conducted. Since OLS assumes only linear regression between different and property variables, the GWR model was also conducted, considering spatial variability between variables. Comparing the results of these two models using the ANOVA test, the adjusted R^2 increased from 0.162 in OLS to 0.681 in GWR, meaning that the GWR model explains the variation in property value more than the OLS model. In conclusion, GWR is more well-fitted than OLS. However, RF was also applied as a non-linear and common machine learning algorithm to find the most suitable method for this study. The results illustrate that the RF model explains 83.1% of the property value variation based on the adjusted R-square value, concluding that RF is the most suitable method for developing a model for analysing the impact of the different types of urban green environments on property value in Alkmaar.

The RF model demonstrates a non-linear correlation between variables and property value. For instance, the correlation between the property value and the height of the green environment indicates that the property value increases considerably as the mean height of vegetation rises from approximately 0.25 to 1m. After the peak of 1m, by increasing the height of vegetation around the properties, the property value starts to fluctuate, showing that the effects of the height of vegetation on property value are not strictly monotonic. The feature importance analysis of the RF model also illustrates that distance to greenhouses and recreational centres, as two types of urban green environments (classified based on the services), have the highest importance in property value prediction in Alkmaar. However, the RF model have an error percentage of 12.1%, meaning that the RF model predict the actual value with a 12.1% error range.

OLS modelling for the effects of 3D factors on property value also illustrates that a height of vegetation and a view of vegetation at a distance of 25m positively correlated with property value, while sunlight and a view of the sky have a negative correlation with property value. The error percentage of the validation of this model is 17.01%, indicating that the property model based on the 3D factors predicts the property value 83% times accurately.

In conclusion, comparing the RF model developed by the combination of 2D factors and types of green environments with the OLS model developed by 3D determines that the RF model explains the property value more accurately than the OLS model. However, modelling based on the 3D factors was based on the limited sample points, and analysis by a huge number of sample points might make the comparison of these two property value models more accurate.

6.2. Ethical considerations

In this research, the property value data provided by the Netherlands' real estate valuation (Waardering onroerende zazen (WOZ)) from the municipality of Alkmaar. The WOZ value is open-source data calculated annually for all properties by the municipalities. Moreover, the data for other steps of this research are all open-source data without any restrictions to finding and sharing them.

6.3. Recommendations for future research

Recommendations for future research are mentioned in the following:

- It will be better to classify urban green environments by a qualitative approach, considering factors such as feeling of safety, quietness, cleanliness, and maintenance to find how these factors affect the value of property.
- The results of the current study should be compared with the people's perceived property value to evaluate the model. This can be conducted through an interview with residents of several properties, finding to what extent the impact of types of green environments on property value, based on the developed model, is similar to residents' perception of the property value.
- Based on the literature, there are various approaches to the classification of green environments, such as the shape of green environments, which can be considered in future research to explore how these new types affect the value of properties.
- There are other machine learning algorithms, including SVM or CNN, which can also be applied to find the most suitable, accurate, and well-fitted model.
- It will be better to find the correlation of types of green environments in combination with other 2D and 3D factors with the real estate property value instead of the WOZ property value. The property value prediction model will be more accurate since the real estate market value is the actual property value.
- In future research, the data regarding 3D factors should be acquired for all the properties. Then, by the combination of the 2D and 3D factors with different types of urban green environments, a concise and more authentic property value model will be developed. This model finds the effects of each type of urban green environment on property value more precisely.

7. REFERENCES

Aryal, J., Sitaula, C., & Aryal, S. (2022). NDVI Threshold-Based Urban Green Space Mapping from Sentinel-2A at the Local Governmental Area (LGA) Level of Victoria, Australia. *Land*, *11*(3). https://doi.org/10.3390/land11030351

Barrow, D. K., & Crone, S. F. (2013). Crogging (cross-validation aggregation) for forecasting - A novel algorithm of neural network ensembles on time series subsamples. *Proceedings of the International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN.2013.6706740

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Cao, K., Diao, M., & Wu, B. (2019). A Big Data–Based Geographically Weighted Regression Model for Public Housing Prices: A Case Study in Singapore. *Annals of the American Association of Geographers*, *109*(1), 173–186. https://doi.org/10.1080/24694452.2018.1470925

Centraal Bureau voor de Statistiek. (2023). *Alkmaar*. https://www.cbs.nl/nl-nl/publicatie/2011/50/alkmaar

Chen, Y., Jones, C. A., Dunse, N. A., Li, E., & Liu, Y. (2023). Housing Prices and the Characteristics of Nearby Green Space: Does Landscape Pattern Index Matter? Evidence from Metropolitan Area. *Land*, *12*(2). https://doi.org/10.3390/land12020496

Czembrowski, P., & Kronenberg, J. (2016). Hedonic pricing and different urban green space types and sizes: Insights into the discussion on valuing ecosystem services. *Landscape and Urban Planning*, *146*, 11–19. https://doi.org/10.1016/j.landurbplan.2015.10.005

Degerickx, J., Hermy, M., & Somers, B. (2020). Mapping functional urban green types using high resolution remote sensing data. *Sustainability (Switzerland)*, *12*(5). https://doi.org/10.3390/su12052144

Del Giudice, M. (2023). Individual and group differences in multivariate domains: What happens when the number of traits increases? *Personality and Individual Differences*, *213*. https://doi.org/10.1016/j.paid.2023.112282

Dell'Anna, F., Bravi, M., & Bottero, M. (2022). Urban Green infrastructures: How much did they affect property prices in Singapore? *Urban Forestry and Urban Greening*, *68*. https://doi.org/10.1016/j.ufug.2022.127475

Derkzen, M. L., van Teeffelen, A. J. A., & Verburg, P. H. (2015). REVIEW: Quantifying urban ecosystem services based on high-resolution data of urban green space: An assessment for Rotterdam, the Netherlands. *Journal of Applied Ecology*, *52*(4), 1020–1032. https://doi.org/10.1111/1365-2664.12469

Donovan, G. H., Champ, P. A., & Butry, D. T. (2007). Wildfire Risk and Housing Prices: A Case Study from Colorado Springs. *Land Economics*, *83*(2), 217–233. https://doi.org/10.3368/le.83.2.217

Efthymiou, D., & Antoniou, C. (2013). How do transport infrastructure and policies affect house prices and rents? Evidence from Athens, Greece. *Transportation Research Part A: Policy and Practice*, *52*, 1–22. https://doi.org/10.1016/j.tra.2013.04.002

Faryadi, S., & Taheri, S. (2009). Interconnections of Urban Green Spaces and Environmental Quality of Tehran. *Int. J. Environ. Res*, *3*(2), 199–208. https://doi.org/10.1016/j.scs.2009.03.001

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics 2* (5th ed.). SAGE Edge.

Fleming, D., Grimes, A., Lebreton, L., Maré, D., & Nunns, P. (2018). Valuing sunshine. *Regional Science and Urban Economics*, *68*, 268–276. https://doi.org/10.1016/j.regsciurbeco.2017.11.008

Gemeente Alkmaar. (2017). *Omgevingsvisie Alkmaar 2040*. https://alkmaar.raadsinformatie.nl/document/5782852/1#search=%22omgevingsvisie%22

Gupta, K., Kumar, P., Pathan, S. K., & Sharma, K. P. (2012). Urban Neighborhood Green Index - A measure of green spaces in urban areas. *Landscape and Urban Planning*, *105*(3), 325–335. https://doi.org/10.1016/j.landurbplan.2012.01.003

Ho, W. K. O., Tang, B. S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research*, *38*(1), 48–70. https://doi.org/10.1080/09599916.2020.1832558

Hoang, N. D., & Tran, X. L. (2021). Remote Sensing-Based Urban Green Space Detection Using Marine Predators Algorithm Optimized Machine Learning Approach. *Mathematical Problems in Engineering*, *2021*. https://doi.org/10.1155/2021/5586913

Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., & Cai, Z. (2019). Monitoring housing rental prices based on social media:An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. *Land Use Policy*, *82*, 657–673. https://doi.org/10.1016/j.landusepol.2018.12.030

Hui, E. C. M., Zhong, J. W., & Yu, K. H. (2012). The impact of landscape views and storey levels on property prices. *Landscape and Urban Planning*, *105*(1–2), 86–93. https://doi.org/10.1016/j.landurbplan.2011.12.002

Jiao, L., & Liu, Y. (2010). Geographic Field Model based hedonic valuation of urban open spaces in Wuhan, China. *Landscape and Urban Planning*, *98*(1), 47–55. https://doi.org/10.1016/j.landurbplan.2010.07.009

Jim, C. Y., & Chen, W. Y. (2010). External effects of neighbourhood parks and landscape elements on high-rise residential value. *Land Use Policy*, *27*(2), 662–670. https://doi.org/10.1016/j.landusepol.2009.08.027

Kaloustian, N., & Bechtel, B. (2016). Local Climatic Zoning and Urban Heat Island in Beirut. *Procedia Engineering*, *169*, 216–223. https://doi.org/10.1016/j.proeng.2016.10.026

Kim, H. G., Hung, K. C., & Park, S. Y. (2015). Determinants of Housing Prices in Hong Kong: A Box-Cox Quantile Regression Approach. *Journal of Real Estate Finance and Economics*, *50*(2), 270–287. https://doi.org/10.1007/s11146-014-9456-1

Lee, H., Lee, B., & Lee, S. (2020). The unequal impact of natural landscape views on housing prices: Applying visual perception model and quantile regression to apartments in Seoul. *Sustainability (Switzerland)*, *12*(19). https://doi.org/10.3390/su12198275

Liang, X., Liu, Y., Qiu, T., Jing, Y., & Fang, F. (2018). The effects of locational factors on the housing prices of residential communities: The case of Ningbo, China. *Habitat International*, *81*, 1–11. https://doi.org/10.1016/j.habitatint.2018.09.004

Liebelt, V., Bartke, S., & Schwarz, N. (2019). Urban green spaces and housing prices: An alternative perspective. *Sustainability (Switzerland)*, *11*(13). https://doi.org/10.3390/su11133707

Lu, B., Charlton, M., & Fotheringham, A. S. (2011). Geographically Weighted Regression using a non-Euclidean distance metric with a study on London house price data. *Procedia Environmental Sciences*, *7*, 92–97. https://doi.org/10.1016/j.proenv.2011.07.017

Ludwig, C., Hecht, R., Lautenbach, S., Schorcht, M., & Zipf, A. (2021). Mapping public urban green spaces based on openstreetmap and sentinel-2 imagery using belief functions. *ISPRS International Journal of Geo-Information*, *10*(4). https://doi.org/10.3390/ijgi10040251

Luttik, J. (2000). The Value of Trees, Water and Open Space as Reflected by House Prices in the Netherlands. *Landscape and Urban Planning*, *48*, 161–167. https://doi.org/10.1016/S0169-2046(00)00039-6

Mathey, J., Hennersdorf, J., Lehmann, I., & Wende, W. (2021). Qualifying the urban structure type approach for urban green space analysis – A case study of Dresden, Germany. *Ecological Indicators*, *125*. https://doi.org/10.1016/j.ecolind.2021.107519

Morano, P., Guarini, M. R., Tajani, F., Di Liddo, F., & Anelli, D. (2019). Incidence of Different Types of Urban Green Spaces on Property Prices. A Case Study in the Flaminio District of Rome (Italy). *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11622 LNCS*, 23–34. https://doi.org/10.1007/978-3-030-24305-0_3

Mosammam, H. M., Nia, J. T., Khani, H., Teymouri, A., & Kazemi, M. (2017). Monitoring land use change and measuring urban sprawl based on its spatial forms: The case of Qom city. *The Egyptian Journal of Remote Sensing and Space Science*, *20*(1), 103–116. https://doi.org/10.1016/J.EJRS.2016.08.002

Nijskens, R., Lohuis, M., Hilbers, P., & Heeringa, W. (2019). Hot Property: The Housing Market in Major Cities. In *Hot Property: The Housing Market in Major Cities*. Springer International Publishing. https://doi.org/10.1007/978-3-030-11674-3

NVM. (2024). *Regionale analyse COROP-regio Alkmaar en omgeving*. https://www.nvm.nl/media/zmsh1eci/regionale-analyse-alkmaar-en-omgeving-3e-kwartaal-2023.pdf

Panduro, T. E., & Veie, K. L. (2013). Classification and valuation of urban green spaces-A hedonic house price valuation. *Landscape and Urban Planning*, *120*, 119–128. https://doi.org/10.1016/j.landurbplan.2013.08.009

Park, B., & Kwon Bae, J. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications*, *42*(6), 2928–2934. https://doi.org/10.1016/j.eswa.2014.11.040

Peschardt, K. K., Schipperijn, J., & Stigsdotter, U. K. (2012). Use of Small Public Urban Green Spaces (SPUGS). *Urban Forestry and Urban Greening*, *11*(3), 235–244. https://doi.org/10.1016/j.ufug.2012.04.002

Piaggio, M. (2021). The value of public urban green spaces: Measuring the effects of proximity to and size of urban green spaces on housing market values in San José, Costa Rica. *Land Use Policy*, *109*. https://doi.org/10.1016/j.landusepol.2021.105656

Rodríguez, J. D., Pérez, A., & Lozano, J. A. (2010). Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 569–575. https://doi.org/10.1109/TPAMI.2009.187

Saptutyningsih, E., Ekonomi, F., Muhammadiyah, U., Jalan, Y., & Selatan, L. (2013). Impact of air pollution on property values: A Hedonic Price study. *Jurnal Ekonomi Pembangunan*, *14*(1), 52–65.

Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications*, *36*, 2843–2852. https://doi.org/10.1016/j.eswa.2008.01.044

Sirmans, G. S., Macpherson, D. A., & Zietz, E. N. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, *13*(1), 3–43. https://www.jstor.org/stable/44103506

Song, M., Wang, S., Wu, J., & Yang, L. (2011). A new space-time correlation coefficient and its comparison with Moran's Index on evaluation. *Management Decision*, *49*(9), 1426–1443. https://doi.org/10.1108/00251741111173925

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, *134*, 93–101. https://doi.org/https://doi.org/10.1016/j.eswa.2019.05.028

Stessens, P., Canters, F., Huysmans, M., & Khan, A. Z. (2020). Urban green space qualities: An integrated approach towards GIS-based assessment reflecting user perception. *Land Use Policy*, *91*. https://doi.org/10.1016/j.landusepol.2019.104319

Stewart, I. D., Oke, T. R., & Krayenhoff, E. S. (2014). Evaluation of the "local climate zone" scheme using temperature observations and model simulations. *International Journal of Climatology*, *34*(4), 1062–1080. https://doi.org/10.1002/joc.3746

Troy, A., & Grove, J. M. (2008). Property values, parks, and crime: A hedonic analysis in Baltimore, MD. *Landscape and Urban Planning*, *87*(3), 233–245. https://doi.org/https://doi.org/10.1016/j.landurbplan.2008.06.005

United Nations. (2022). *Envisaging the Future of Cities*. https://unhabitat.org/sites/default/files/2022/06/wcr_2022.pdf

Usman, H., Lizam, M., & Bint Burhan, B. (2021). A review of spatial economics in explicit location modelling of commercial property market. *Journal of the Malaysian Institute of Planners*, *19*, 438–448.

Waltert, F., & Schläpfer, F. (2010). Landscape amenities and local development: A review of migration, regional economic and hedonic pricing studies. *Ecological Economics*, *70*(2), 141–152. https://doi.org/10.1016/j.ecolecon.2010.09.031

Wang, Y., de Groot, R., Bakker, F., Wörtche, H., & Leemans, R. (2017). Thermal comfort in urban green spaces: a survey on a Dutch university campus. *International Journal of Biometeorology*, *61*(1), 87–101. https://doi.org/10.1007/s00484-016-1193-0

Wen, H., Xiao, Y., & Zhang, L. (2017). Spatial effect of river landscape on housing price: An empirical study on the Grand Canal in Hangzhou, China. *Habitat International*, *63*, 34–44. https://doi.org/10.1016/j.habitatint.2017.03.007

Wen, H., Zhang, Y., & Zhang, L. (2014). Do educational facilities affect housing price? An empirical study inHangzhou, China. *Habitat International*, *42*, 155–163. https://doi.org/10.1016/j.habitatint.2013.12.004

Wing, C. K., & Chin, T. (2003). A Critical Review of Literature on the Hedonic Price Model. *International Journal for Housing Science and Its Applications*, *27*, 145–165.

Wittowsky, D., Hoekveld, J., Welsch, J., & Steier, M. (2020). Residential housing prices: impact of housing characteristics, accessibility and neighbouring apartments–a case study of Dortmund, Germany. *Urban, Planning and Transport Research*, *8*(1), 44–70. https://doi.org/10.1080/21650020.2019.1704429

Wolch, J. R., Byrne, J., & Newell, J. P. (2014). Urban green space, public health, and environmental justice: The challenge of making cities "just green enough." *Landscape and Urban Planning*, *125*, 234–244. https://doi.org/10.1016/j.landurbplan.2014.01.017

Wu, J., Wang, M., Li, W., Peng, J., & Huang, L. (2015). Impact of Urban Green Space on Residential Housing Prices: Case Study in Shenzhen. *Journal of Urban Planning and Development*, *141*(4). https://doi.org/10.1061/(asce)up.1943-5444.0000241

Wüstemann, H., & Kolbe, J. (2015). Estimating the Value of Urban Green Space: A hedonic Pricing Analysis of the Housing Market in Cologne, Germany. *SFB 649 Discussion Paper 2015-002*, *2015*.

Yang, L., Chen, Y., Xu, N., Zhao, R., Chau, K. W., & Hong, S. (2020). Place-varying impacts of urban rail transit on property prices in Shenzhen, China: Insights for value capture. *Sustainable Cities and Society*, *58*. https://doi.org/10.1016/j.scs.2020.102140

Yao, W., Poleswki, P., & Krzystek, P. (2016). Classification of urban aerial data based on pixel labelling with deep convolutional neural networks and logistic regression. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *41*, 405–410. https://doi.org/10.5194/isprsarchives-XLI-B7-405-2016

Ying, Y. (2019). *Assessment of 2D and 3D methods for property valuation using remote sensing data at the neighborhood scale in Xian, China*. https://essay.utwente.nl/83578/1/ying.pdf

Ying, Y., Koeva, M., Kuffer, M., Asiama, K. O., Li, X., & Zevenbergen, J. (2021). Making the third dimension (3d) explicit in hedonic price modelling: A case study of Xi'an, China. *Land*, *10*(1), 1–26. https://doi.org/10.3390/land10010024

Zambrano-Monserrate, M. A., Ruano, M. A., Yoong-Parraga, C., & Silva, C. A. (2021). Urban green spaces and housing prices in developing countries: A Two-stage quantile spatial regression analysis. *Forest Policy and Economics*, *125*. https://doi.org/10.1016/j.forpol.2021.102420

Zheng, B., & Li, J. (2022). Evaluating the Annual Effect of the Sky View Factor on the Indoor Thermal Environment of Residential Buildings by Envi-met. *Buildings*, *12*(6). https://doi.org/10.3390/buildings12060787

8. APPENDIX

Appendix 1

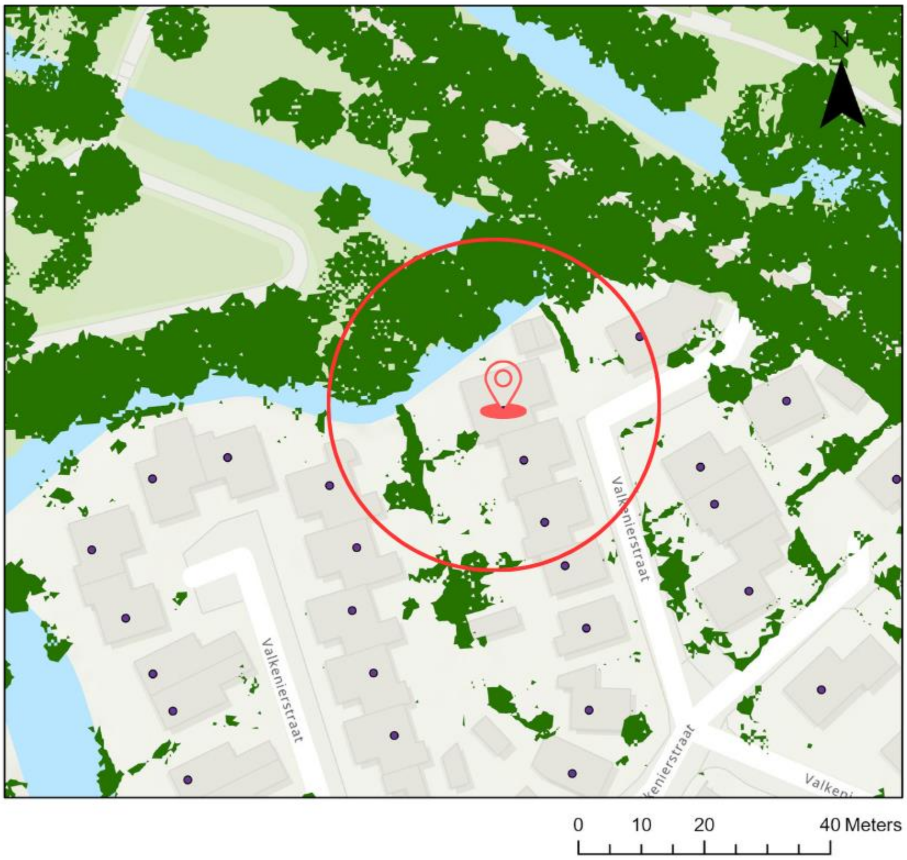


Figure 25: A buffer zone of 25m around a sample property in Alkmaar

Source: Author, 2024.

Appendix 2

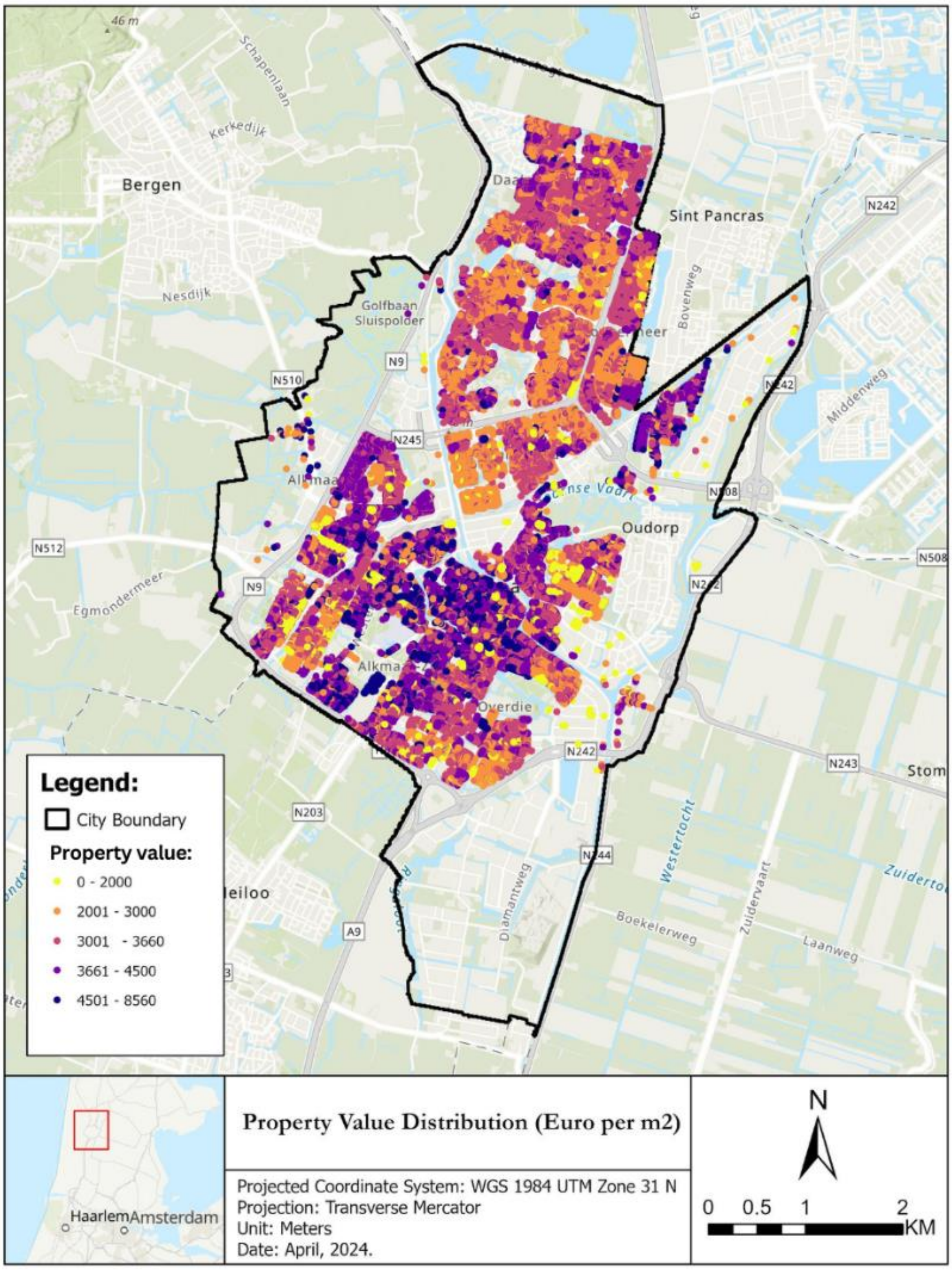


Figure 26: Property value distribution (Euro/m2) in Alkmaar.

Source: Authror, 2024.

Appendix 3

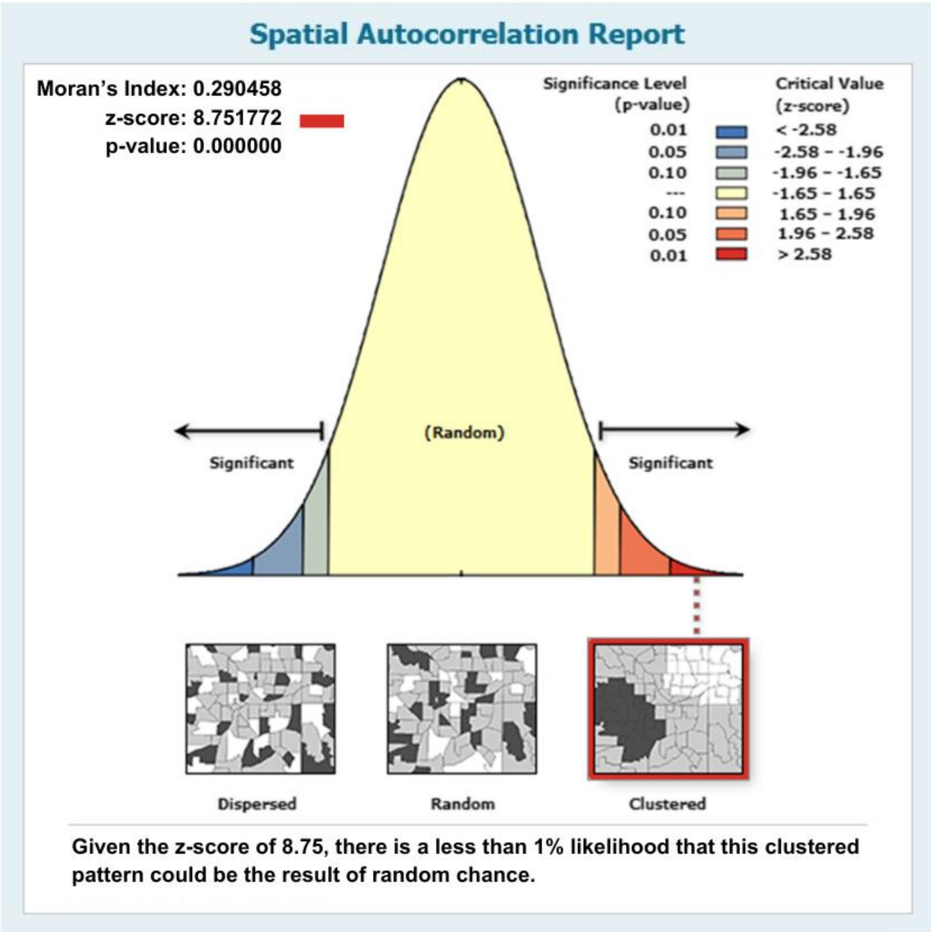


Figure 27: Spatial autocorrelation report of property value in Alkmaar

Source: Author, 2024.

Appendix 4

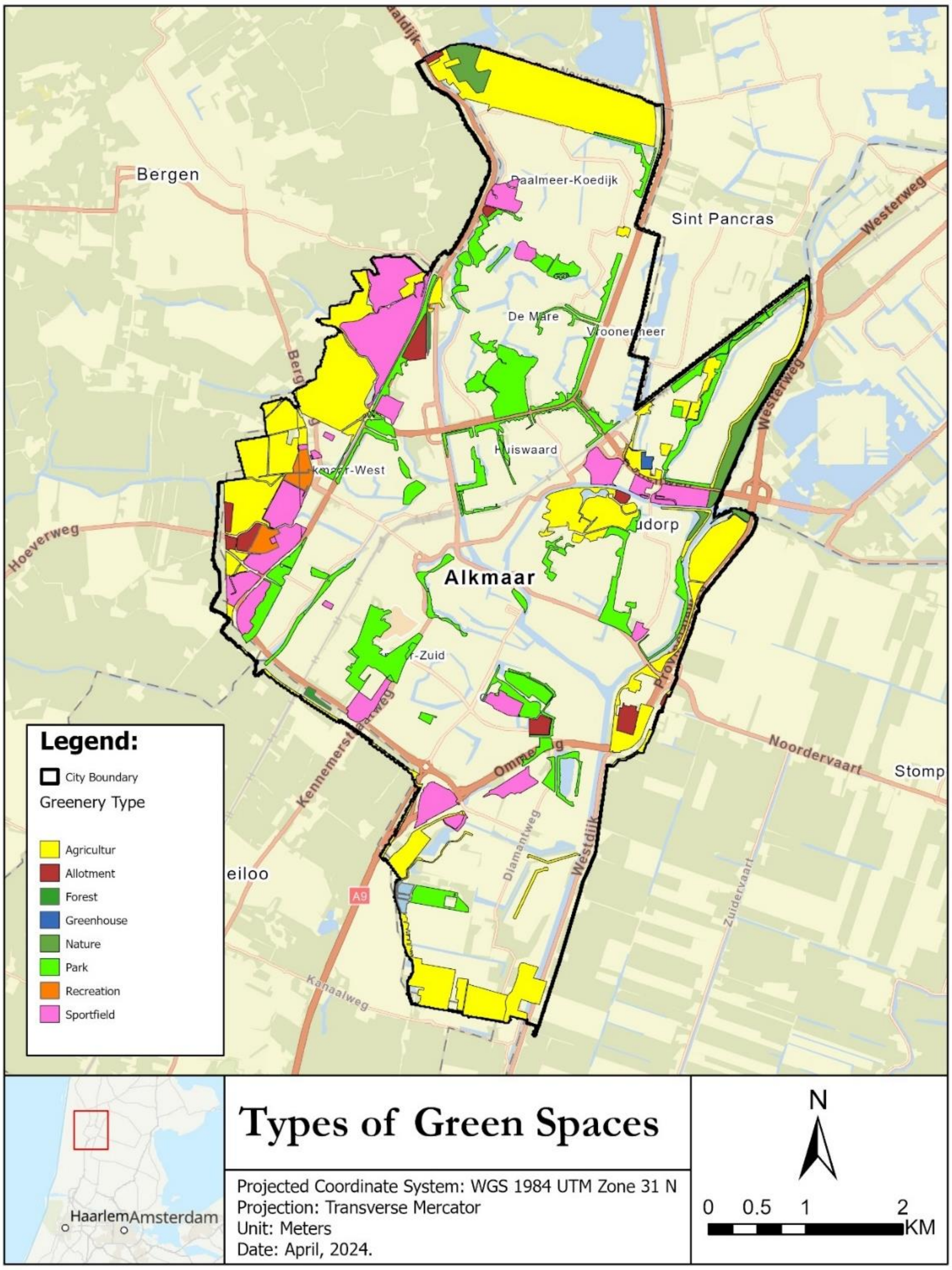


Figure 28: A map of the classification of green environments based on services they provide in Alkmaar

Source: Author, 2024.

Appendix 5

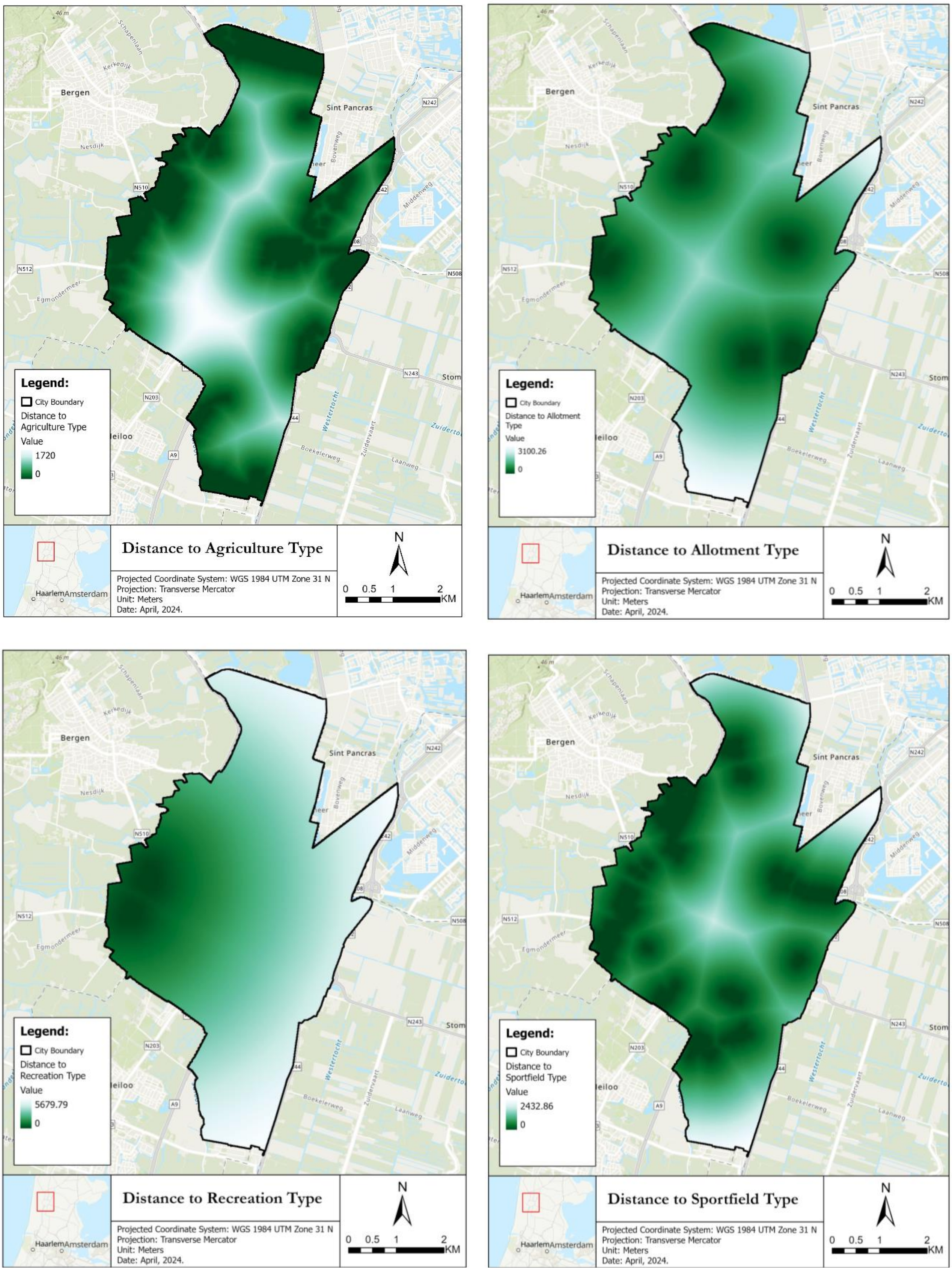


Figure 29: Maps of Euclidean distance to four types of green environments classified by services.

Source: Author, 2024.

Appendix 6

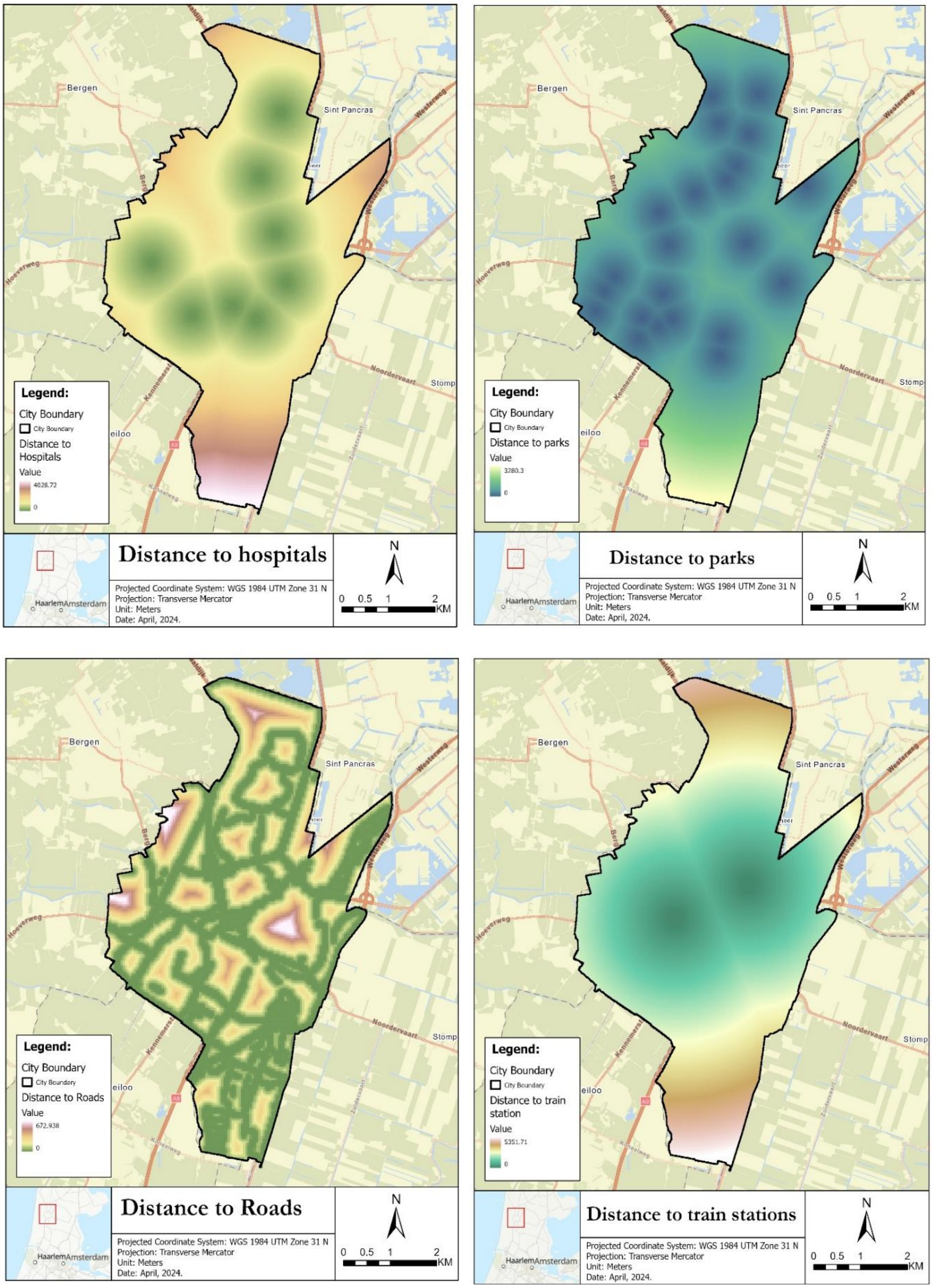


Figure 30: A map of the spatial analysis of 2D factors

Source: Author, 2024.

Appendix 7

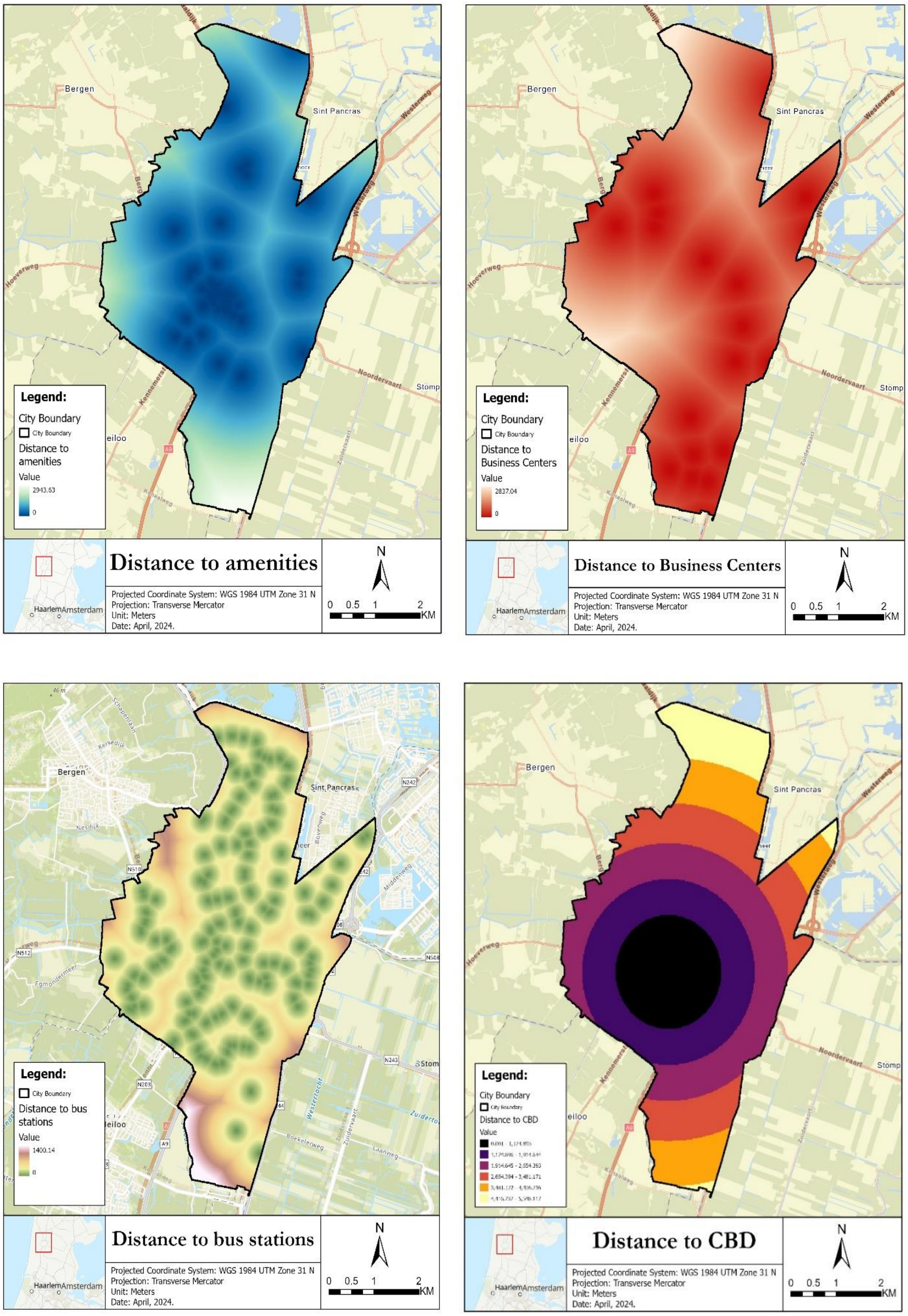


Figure 31: Map of the spatial analysis of 2D factors.

Source: Author, 2024.

Appendix 8

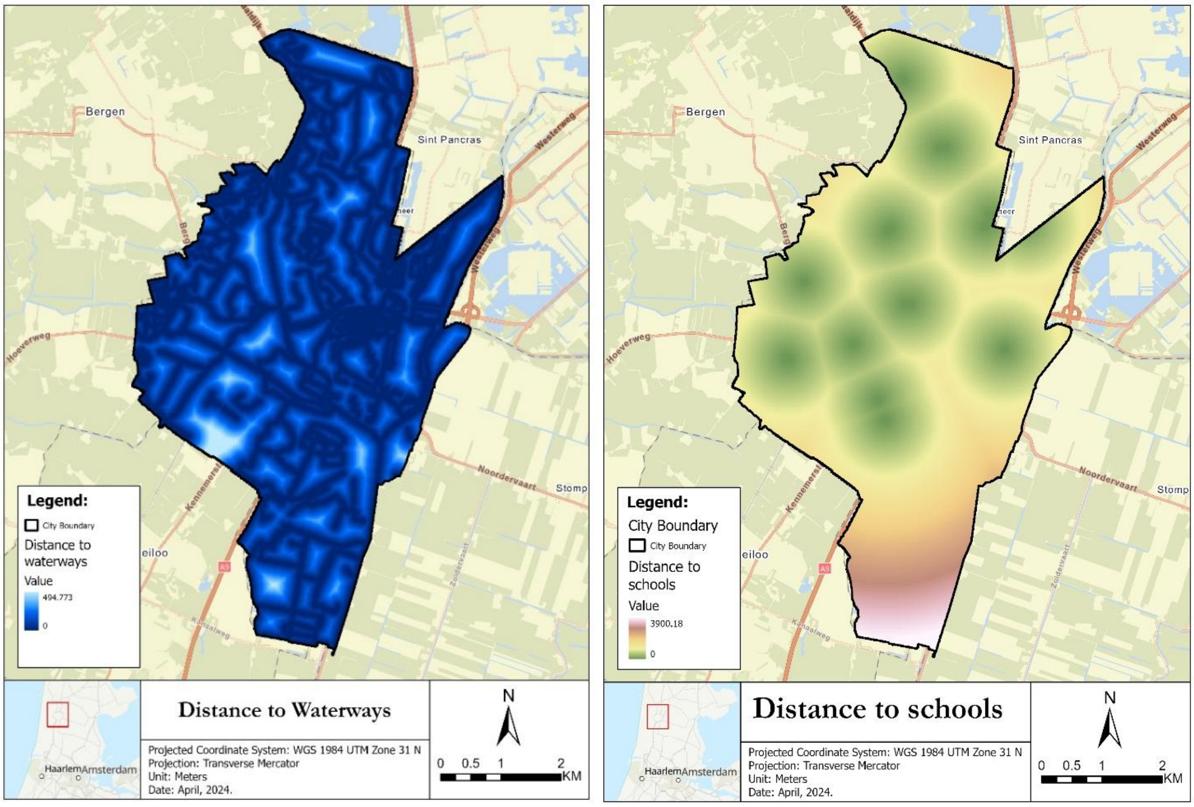


Figure 32: Map of the spatial analysis of 2D factors.

Source: Author, 2024.

Appendix 9

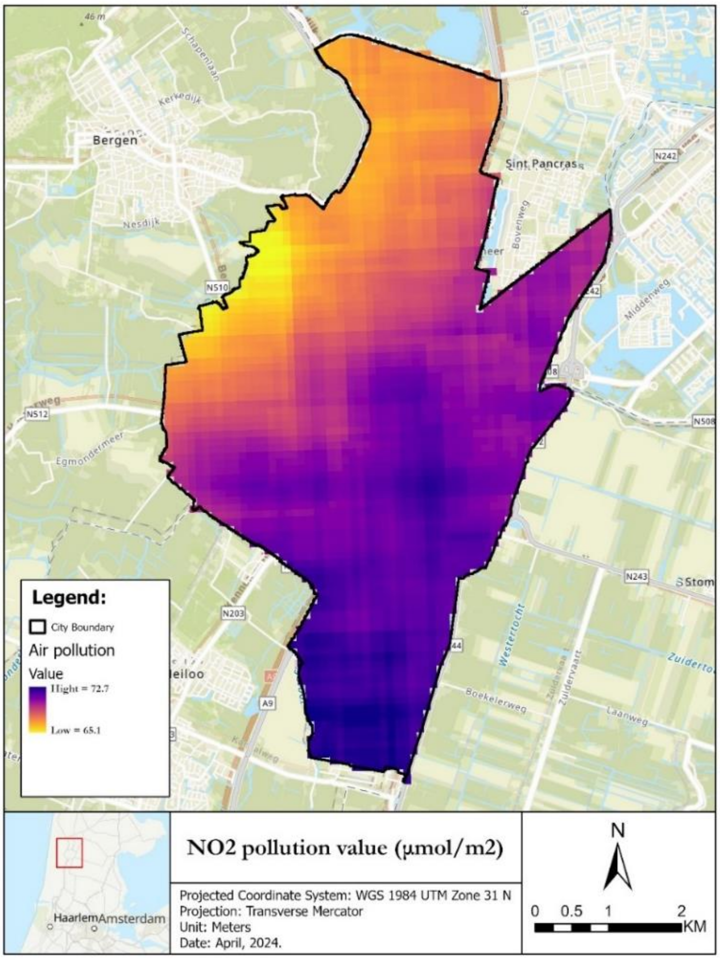


Figure 33: Map of the spatial distribution of No_2 pollutant.

Source: Author, 2024.

```
// Define the geographical boundary for Alkmaar, Netherlands
var alkmaarBoundary = ee.Geometry.Rectangle([4.6841, 52.6127, 4.7748, 52.6500]);
// Import the Sentinel-5P NO2 dataset
var dataset = ee.ImageCollection('COPERNICUS/S5P/NRTI/L3_NO2')
                  .select('NO2_column_number_density')
                  .filterDate('2022-01-01', '2022-01-31');

// Get the mean NO2 values for January 2023
var no2Mean = dataset.mean();

// Set visualization parameters for NO2
var visParams = {
  min: 0,
  max: 0.0002,
  palette: ['black', 'blue', 'purple', 'cyan', 'green', 'yellow', 'red']
};

// Add the NO2 layer to the map
Map.setCenter(4.753375, 52.632381, 10); // Center the map over Alkmaar
Map.addLayer(no2Mean, visParams, 'Mean NO2');

// Export the image, specifying scale and region
Export.image.toDrive({
  image: no2Mean,
  description: 'Alkmaar_NO2_Median',
  scale: 1000,
  region: alkmaar,
  fileFormat: 'GeoTIFF'
});
```

Figure 34: Coding for mapping No_2 distribution in GEE

Appendix 10

Python language coding for RF

```
import geopandas as gpd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
from sklearn.metrics import r2_score, mean_absolute_error
from sklearn.metrics import mean_squared_error
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.optimizers import Adam
```

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, r2_score
from sklearn.pipeline import make_pipeline
```

```
data = gpd.read_file(r'D:\thesis\Alkmaar\Spatial_factors\SHP_price\Pro_price_2dfactors_edited_final20.shp')

# predictors and the target
X = data[['size_nom', 'GE_siz_nom', 'Type3_norm', 'Type2_norm', 'Type1_norm', 'DFS_norm', 'D_altm_nom', 'D_agr_norm', 'D_GH_norm'
y = data['propertypr']

X = X.astype(float)
y = y.astype(float)

# split the data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# RF Model
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# SVM Model
svm_model = SVR(kernel='rbf')
svm_model.fit(X_train, y_train)

# ANN Model
ann_model = Sequential([
    Dense(128, activation='relu', input_shape=(X_train.shape[1],)),
    Dense(64, activation='relu'),
    Dense(1)
])
ann_model.compile(optimizer=Adam(), loss='mean_squared_error')
ann_model.fit(X_train, y_train, epochs=50, batch_size=10)
```

```
# Prediction
y_pred_rf = rf_model.predict(X_test)
y_pred_svm = svm_model.predict(X_test)
y_pred_ann = ann_model.predict(X_test).flatten()

# R-squared and MAE and MSE
print("RF R-squared:", r2_score(y_test, y_pred_rf))
print("RF MAE:", mean_absolute_error(y_test, y_pred_rf))
print("RF MSE:", mean_squared_error(y_test, y_pred_rf))
print("SVM R-squared:", r2_score(y_test, y_pred_svm))
print("SVM MAE:", mean_absolute_error(y_test, y_pred_svm))
print("ANN R-squared:", r2_score(y_test, y_pred_ann))
print("ANN MAE:", mean_absolute_error(y_test, y_pred_ann))
```

```
import matplotlib.pyplot as plt

# predictors and the target
X = data[['size_nom', 'GE_siz_nom', 'Type3_norm', 'Type2_norm', 'Type1_norm', 'DFS_norm', 'D_altm_nom', 'D_agr_norm', 'D_GH_norm'
y = data['propertypr']

# Split the data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# feature importances
feature_importances = model.feature_importances_

# visualization
feature_importances_df = pd.DataFrame({
    'Feature': X.columns,
    'Importance': feature_importances
}).sort_values(by='Importance', ascending=False)

# Plot
plt.figure(figsize=(10, 6))
plt.barh(feature_importances_df['Feature'], feature_importances_df['Importance'])
plt.xlabel('Importance')
plt.ylabel('Feature')
plt.title('Feature Importance in Predicting Property Price')
plt.gca().invert_yaxis()
plt.show()

# Print
print(feature_importances_df)
```

Appendix 11

Python language coding for RF

```
from sklearn.inspection import PartialDependenceDisplay


#  predictors and target
X = data[['size_nom', 'GE_siz_nom', 'Type3_norm', 'Type2_norm', 'Type1_norm', 'DFS_norm', 'D_altm_nom', 'D_agr_norm', 'D_GH_norm'
y = data['propertypr']


model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X, y)

# partial dependence plot for 'DFS_norm'
features = ['DFS_norm']  # the feature for the x-axis
PartialDependenceDisplay.from_estimator(model, X, features, grid_resolution=50)

plt.xlabel('DFS_norm')
plt.ylabel('Predicted Property Value')
plt.title('Partial Dependence Plot of Property Value vs DFS_norm')
plt.show()
```

```
# variable
X = data[['size', 'housetyp_', 'D_bus', 'D_air', 'D_school', 'D_road', 'D_hospital', 'D_amenitie', 'D_CBD', 'D_train', 'D_park',
y = data['propertypr']


# Train the RF model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X, y)

# partial dependence plot for 'DFS_norm'
features = ['size']  # the feature for the x-axis
PartialDependenceDisplay.from_estimator(model, X, features, grid_resolution=50)

# Customize the plot
plt.xlabel('Size of property')
plt.ylabel('Predicted Property Value')
plt.title('Partial Dependence Plot of Property Value vs Size of property')
plt.show()
```

```
X = data[['size', 'housetyp_', 'D_bus', 'D_air', 'D_school', 'D_road', 'D_hospital', 'D_amenitie', 'D_CBD', 'D_train', 'D_park',
y = data['propertypr']


model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X, y)


features = ['NDVI']  # the feature for the x-axis
PartialDependenceDisplay.from_estimator(model, X, features, grid_resolution=50)


plt.xlabel('Density of green environments')
plt.ylabel('Predicted Property Value')
plt.title('Partial Dependence Plot of Property Value vs Density of green environments')
plt.show()
```

```
X = data[['size', 'housetyp_', 'D_bus', 'D_air', 'D_school', 'D_road', 'D_hospital', 'D_amenitie', 'D_CBD', 'D_train', 'D_park',
y = data['propertypr']


model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X, y)


features = ['DFS']  # the feature for the x-axis
PartialDependenceDisplay.from_estimator(model, X, features, grid_resolution=50)


plt.xlabel('Height of green environments')
plt.ylabel('Predicted Property Value')
plt.title('Partial Dependence Plot of Property Value vs Height of green environments')
plt.show()
```

```
features = ['D_CBD']  # the feature for the x-axis
PartialDependenceDisplay.from_estimator(model, X, features, grid_resolution=50)


plt.xlabel('Distance to CBD')
plt.ylabel('Predicted Property Value')
plt.title('Partial Dependence Plot of Property Value vs Distance to CBD')
plt.show()
```

```
features = ['D_recreat']
PartialDependenceDisplay.from_estimator(model, X, features, grid_resolution=50)


plt.xlabel('Distance to recreational green environments')
plt.ylabel('Predicted Property Value')
plt.title('Partial Dependence Plot of Property Value vs Distance to recreational green environments')
plt.show()
```

```
features = ['D_amenitie']
PartialDependenceDisplay.from_estimator(model, X, features, grid_resolution=50)


plt.xlabel('Distance to amenities')
plt.ylabel('Predicted Property Value')
plt.title('Partial Dependence Plot of Property Value vs Distance to amenities')
plt.show()
```

Appendix 12

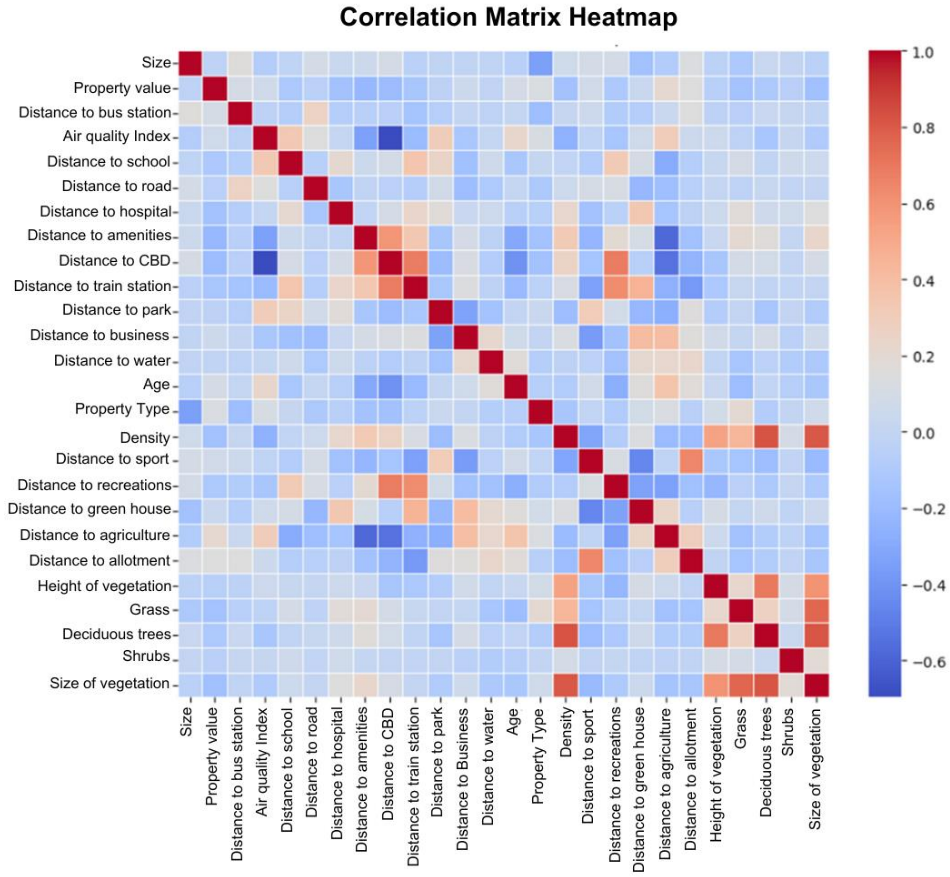


Figure 35: The correlation matrix heatmap of the RF model

Source: Author, 2024.

Appendix 13

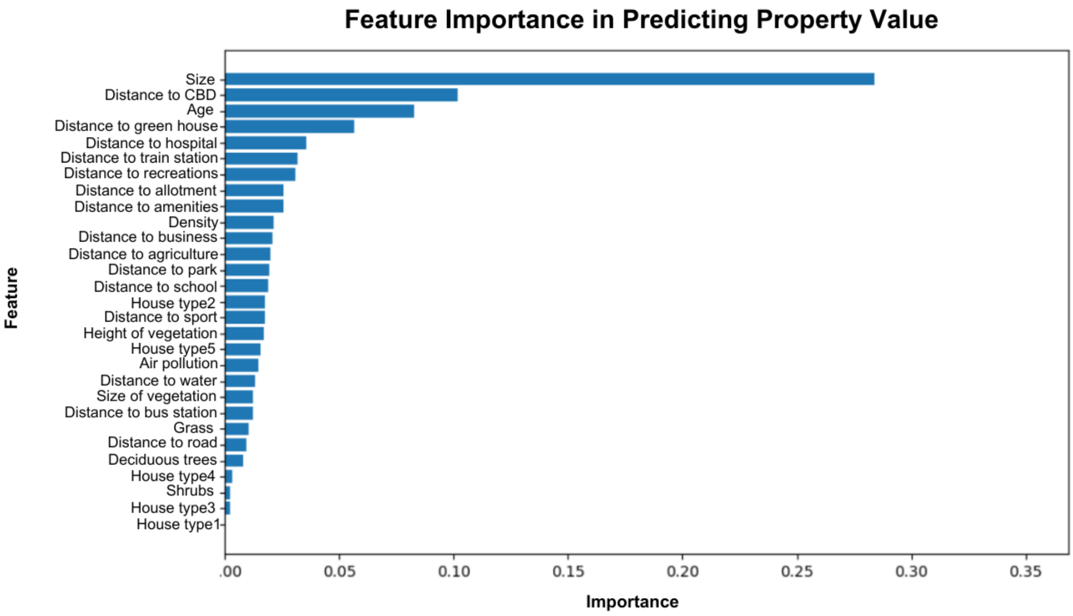
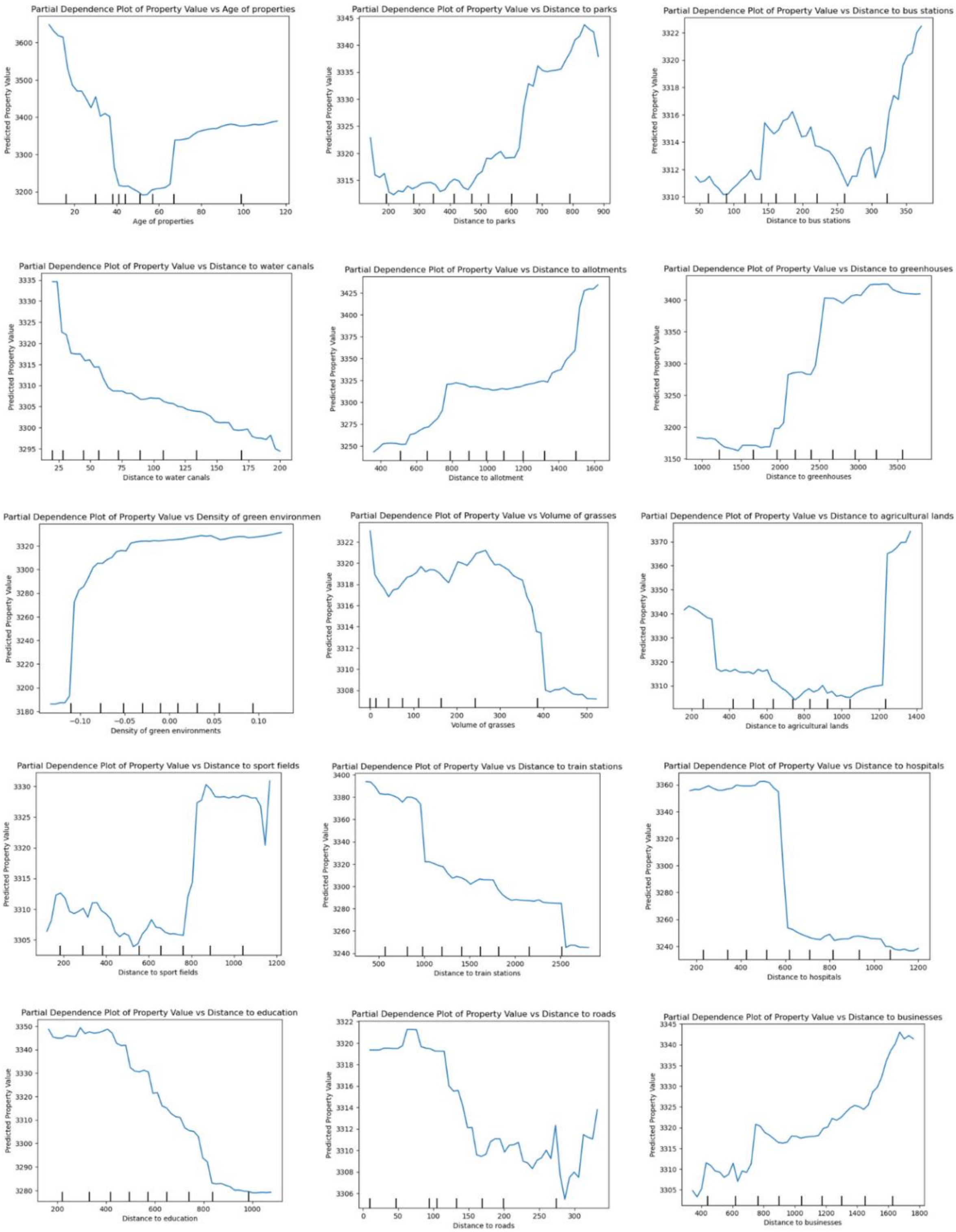


Figure 36: Feature importance in predicting property value in the RF model

Source: Author, 2024.

Appendix 14

The partial dependence plot of property value vs other variables in the RF model



Appendix 15

Table 19: Actual and predicted property value analysis

ID	*WOZ*	*Predicted by 3D factors*	*Actual property value*	*MAE-Woz*	*MAE-funda*
1	3071	3393.46903	3823	322.469034	429.530966
2	4227	3996.72732	5112	-230.272676	1115.272676
3	3106	3439.11536	3765	333.115364	325.884636
4	4587	3265.26585	5639	-1321.73415	2373.734146
5	2800	3153.91342	3340	353.913424	186.086576
6	2764	3278.93929	3423	514.939294	144.060706
7	3739	2978.98899	4485	-760.011006	1506.011006
8	3780	3335.73307	4483	-444.266926	1147.266926
9	3780	3626.25813	4470	-153.741866	843.741866
10	3794	3030.78138	4381	-763.218616	1350.218616
11	3385	3117.15453	4275	-267.845466	1157.845466
12	3438	3386.90858	3937	-51.091416	550.091416
13	3606	2509.73648	4245	-1096.26352	1735.263516
14	5142	4173.79325	5893	-968.206746	1719.206746
15	3395	3058.77155	4095	-336.228446	1036.228446
16	3200	4060.84796	4120	860.847964	59.152036
17	2985	2990.11245	3620	5.112454	629.887546
18	2957	3696.97905	3747	739.979054	50.020946
19	3000	3081.42224	3451	81.422244	369.577756
20	1000	2868.55764	1310	1868.55764	-1558.557644
21	2714	3083.11099	3274	369.110994	190.889006
22	1000	3043.31753	1420	2043.31753	-1623.317534
23	4026	3460.47684	4716	-565.523156	1255.523156
24	3651	3220.66662	4225	-430.333376	1004.333376
25	3651	3283.86577	4437	-367.134226	1153.134226
26	2552	3338.58185	3155	786.581854	-183.581854
27	3234	3067.13027	3866	-166.869726	798.869726
28	3542	3189.67795	4319	-352.322046	1129.322046
	Avg: 3290.214286	Avg: 3290.367982	Avg: 3965.214286	Avg deviation: 591.2296505	Avg deviation: 674.8463035

Source: Author, 2024.