Is your neighbour's balcony cheaper than yours?

Investigating the effect of incorporating geographical data on the predictive performance of hedonic pricing models in Rotterdam's rental property market

Willem Amesz, Sem van Embden, Anna Grefhorst & Wenske Tuk

April 2024

Abstract

A common problem in the housing market is the accurate valuation of properties, considering the multitude of attributes that influence prices. This paper considers traditional hedonic pricing models, such as Ordinary Least Squares with forward selection, Least Absolute Shrinkage and Selection Operator (LASSO) and the popular Random Forest model, as well as their geographically weighted counterparts. Specifically, we focus on a Geographically Weighted Regression (GWR) and a Geographically Weighted Random Forest (GWRF), evaluating their comparable predictive performance in the context of the rental housing market in Rotterdam. Using a unique dataset including ward-specific characteristics, we aim to assess the effect of incorporating geographical data into traditional pricing models. Through empirical analysis and comparison of predictive performance metrics, we provide insights into the strengths and limitations of these approaches. We find there is an improvement in predictive power in the use of the Geographically Weighted Regression over the standard Ordinary Least Squares method, but no substantial improvement in the use of a Geographically Weighted Random Forest versus a standard Random Forest model in the city of Rotterdam. Finally, this paper uses Partial Dependence, LIME, and ALE methods to interpret the Random Forest model and finds there are different preferences for features per ward in Rotterdam.

1 Introduction

The city of Rotterdam boasts distinct wards, also known as districts, each with its own identity: while Kralingen is known for its vibrant student life and villas, Delfshaven is characterized by its canal houses, antique shops and authenticity. The motivation behind this research lies in the desire to enhance our understanding of housing market dynamics, particularly in this city. Traditional hedonic pricing models, such as Ordinary Least Squares (OLS) with forward selection, Least Absolute Shrinkage and Selection Operator (LASSO) regressions and Random Forests (RF) have been employed to estimate rental prices based on various features of properties. However, most of the models in the literature do not include geographical variables, for example, the safety or social cohesion of a ward. In a big city such as Rotterdam, where different wards possess distinct characteristics, these supplementary variables could be used to improve the predictions of hedonic pricing models. Furthermore, analysis of the interpretability becomes more interesting, since the features of the wards can be investigated separately. First, this research implements a Geographically Weighted Regression (GWR), which takes nearby variables into account when estimating a local model for each property. In addition, the RF is extended by introducing again a local model for every data point and this is called the Geographically Weighted Random Forest (GWRF). In the end, we compare the models' predictive performances and yield a substantial difference in mean squared error and mean absolute error for the GWR compared to LASSO and OLS with forward selection, but not for the RF compared to the GWRF. Possible explanations for GWR outperforming LASSO and OLS are the presence of spatial nonstationarity in the data or overfitting of the models. Spatial nonstationarity indicates that the effect of features differs depending on the geographical location of the property, i.e. the coefficients of explanatory variables are geographically varying. Furthermore, this research analyses which variables are important in the price of a property in the different districts by implementing interpretable machine learning methods. In this way, this paper investigates the behaviour of features on a global level but also looks into whether features are more valuable than others in different wards. For example, if your neighbour's balcony is more valuable than yours. We find that the most important features behave in an interpretable way. Moreover, our findings indicate that the interpretation varies locally among wards, leading to certain wards being more impacted by specific features than others.

This research is relevant because it aims to tackle the intricate dynamics and subtleties inherent to Rotterdam's housing market. The city's diverse districts, each with distinct characteristics, create variations in housing demand, preferences, and ultimately prices. Ignoring these geographic differences can lead to inaccurate estimations of rental prices and hinder effective decision-making for both renters and landlords.

Furthermore, as Rotterdam is a major urban centre with significant economic and social importance, enhancing the understanding of housing market dynamics in the city is essential for policymakers, urban planners, real estate professionals, and residents alike. Namely, it will help understand the different districts better, including their needs and preferences. By providing insights into how geographic variations influence housing prices, this research can inform better policy decisions, urban development strategies, and real estate investments, ultimately contributing to the city's overall economic vitality and livability. Therefore, the relevance of this research lies in its potential to address real-world challenges and improve decision-making processes in the housing market of Rotterdam, specifically on a district level.

The remainder of this paper is constructed as follows: Section 2 discusses relevant literature and the contribution of our study. Section 3 explains the source and structure of the data and provides summary statistics. The different models that are used in this paper are provided in Section 4, along with an explanation of the assessment and interpretability. Results on the predictive accuracy of the models, the importance of the different regressors, and interpretation are presented in Section 5. Finally, Section 6 concludes our research and gives an overview of our findings.

2 Literature

2.1 Hedonic Pricing models

Hedonic Pricing models have been widely used in previous literature, especially in determining housing prices (Chau and Chin (2003)). The idea is to explain price differences by looking at the inherent characteristics of houses. This is mostly used in the housing market due to the way real estate has multiple of these characteristics, such as number of rooms, location of the property, or the presence of a balcony. The attributes that are mostly used in hedonic pricing literature are listed in Chau and Chin (2003). These include locational, structural, and neighbourhood characteristics which are attributes of geographical location, property characteristics, and ward characteristics respectively. The idea is to explain the prices not by inspecting demand and supply behaviour, but by viewing the price as the willingness to pay for these attributes (Freeman III (1979)). In Wen et al. (2005), researchers applied hedonic pricing models on the housing in a single city: Hangzhou, just as we apply our models to Rotterdam specifically. This research tries to gather as many of the variables specified in Chau and Chin (2003) since these attributes are seen as the most important in this field.

2.1.1 Linear Specification

The way these characteristics have been used varies across the literature. The simplest version of a hedonic pricing model is the linear approach. This approach can be seen in the earliest works on hedonic pricing such as Griliches (1971), where they applied hedonic pricing models to see what impact a change in the quality of an automobile has on the price. Later research from Lancaster (1966) and Rosen (1974) builds on the linear hedonic price framework focusing on consumer theory and product differentiation. Ordinary Least Squares (OLS) is most commonly used in the linear specification. This method provides clear interpretability but suffers from some limitations due to the assumptions that may not hold in this framework.

Extensions of the standard linear model have been exercised in the literature, including penalized regressions (Xin and Khalid (2018)), such as Ridge and LASSO, and different Least Squares methods, such as Weighted Least Squares (Owusu-Ansah (2011)). The features in these models are often standardized due to multicollinearity issues, using semi-log or Box-Cox transformations (Atkinson et al. (2021)).

2.1.2 Geographically Weighted Regression

A limitation of using OLS for hedonic pricing models is the lack of flexibility to allow for geographically differing relations in the data. Spatial approaches emerged as an extension to linear models, aiming to capture local variations and spatial interactions. Spatial dependence theory states that values of the observations have more similarities with observations that are closer geographically than those that are far away (Lee et al. (2019)). This idea interferes with standard estimation procedures and has consequences for the results. Authors Getis (2007) and Kim et al. (2020) argue that not accounting for these spatial effects not only causes OLS assumptions of linearity, independence and homoscedasticity to fail but also failure to identify local variations among the variables, which can exhibit biased estimation results (Kim et al. (2020), Anselin (1988)).

To overcome these geographic consequences, spatial models were introduced. One of the most used methods is Geographically Weighted Regression (GWR) (Brunsdon et al. (1996)). This method has been used in several fields including food safety (Lee et al. (2019)), environmental justice (Gilbert and

Chakraborty (2011)) and typical hedonic pricing (Kim et al. (2020)) for example. Models using GWR often perform better (Kim et al. (2020)) than OLS, but still have drawbacks. The non-Gaussian distribution of the data is still present. This means that a transformation of the data is often required. Furthermore, GWR cannot capture complex non-linear behaviour. A way to solve this is proposed by Basile and Mínguez (2018) and requires a spline method to simulate a non-linear relationship in the data. Another issue is a lack of formal statistical inference. GWR does not have a statistical framework, because it is a collection of local spatial regressions (Wheeler (2021)).

Other extensions in this field examine the bandwidths in a GWR. To allow for a flexible bandwidth Yang (2014) proposed an algorithm to find the optimal bandwidth for the GWR model. Furthermore, there exist other spatial methods not concerning regression, but these methods seem to have multiple restrictions we cannot deal with in our research.

In conclusion, while spatial regressions have shown improved explanatory power, they still face challenges such as stationarity assumptions and difficulties in capturing complex nonlinear patterns and interaction effects. These models can outperform relatively simple OLS models but do suffer from some of the limitations OLS models also face.

2.1.3 Machine Learning Approaches

Non-parametric approaches such as certain machine learning methods have gained traction due to their flexibility in modeling nonlinear relationships without imposing distributional assumptions. Methods such as Random Forests (RF) (Potrawa and Tetereva (2022), Hong et al. (2020)) often show better results than linear methods. Random forests can also be combined with penalized methods such as Ridge, as seen in Cao et al. (2022), which reduces overfitting.

Random Forests showed better performance than spatial methods in terms of predictive accuracy (Credit (2022)), showing an advantage over traditional models. Most studies using ML methods aim at enhancing predictive performance, especially black-box models such as neural networks (Neloy et al. (2019), Abidoye and Chan (2018)). Another advantage of RF models is that interpretability is not entirely lost compared to other black-box models. Machine learning models have shown promise in surpassing traditional hedonic regression models in predictive accuracy, also in the real estate market (Peterson and Flanagan (2009)).

A spatial extension of the Random Forest is the Geographically Weighted Random Forest (GWRF). This local spatial analysis method was introduced by Georganos and Kalogirou (2022) and proved an effective exploratory tool to visualize the geographically varying relationship between dependent and independent variables. In addition to this, it allows for a deeper analysis of spatial nonstationarity and proved to be better in terms of accuracy than a traditional Random Forest model. There are multiple existing applications of GWRF. Luo et al. (2022) use the GWRF to explain the relationship between socioeconomic factors and poverty in China, with as a result that this relationship varies with space at the county level, while Wu et al. (2024) successfully outperforms the Random Forest in predicting traffic crashes in London.

2.1.4 Interpretability

Addressing the problem of interpretability for black-box models has only been studied in the recent past. This is because explainable methods were only recently developed. Globally, partial dependence plots (PDP) can explain the behaviour of features of a model sufficiently (Zhao and Hastie (2021)). A more unbiased alternative to PD plots is the so-called Accumulated Local Explanation (ALE) plot (Apley and Zhu (2020)). This method is often used when the features in the machine learning model are highly correlated.

For a specific local interpretation, most literature utilizes the Local Interpretable Model-agnostic Explanation (LIME) method (Ribeiro et al. (2016)). LIME can provide local explanations of models and is therefore very useful for black-box models and therefore will also be used in this paper.

2.2 Contribution

Our study aims to extend the existing literature by proposing a comprehensive research framework that addresses some limitations of previous approaches. Our research draws inspiration from Potrawa and Tetereva (2022) by collecting similar data from the same city, namely Rotterdam, and constructing a hedonic pricing model. We extend this research by adding ward-specific variables, such as safety and physical indices further explained in Section 3.2. While there is extensive literature on employing a GWR model on a city, this research differs from the rest due to its rich dataset and inclusion of the GWRF model, which is less prominent in the literature. Furthermore, we extend the research by Wu et al. (2024), in which the authors use a GWR model as well as a GWRF model to predict the number of crashes in traffic in London. Finally, this paper focuses mainly on predictive performance, while we also concentrate on interpretability using partial dependence plots and LIME. Specifically, we implement a Random Forest and look at the interpretability of this model on a district level, extending the ideas of Potrawa and Tetereva (2022).

3 Data

Housing data for Rotterdam was collected from funda.nl, a prominent housing website in the Netherlands, using the Python package funda-scraper 1.1.1. Our final dataset contains 1193 listings of apartments and houses that were sold from January 2023 to February 2024. This dataset provides valuable insights into the housing market trends in Rotterdam during this period. For more information on the location and surroundings of a property, we use the Google Maps geocoding API. Furthermore, we obtained data about the quality of living in the neighbourhood and other location-based statistics of the different districts in Rotterdam from the municipality website. The following sections will elaborate more on each dataset.

The resulting dataset contained data on rental price per month we transformed using natural logarithm ($log(price)$), which is plotted per observation in Figure 1, living area ($living\_area$), number of rooms ($room$), number of bedrooms ($bedroom$) and property age ($house\_age$). From the description of the properties, more variables were extracted using text analysis resulting in the variables of floor level ($floor$), presence of balcony (dummy: $balcony$, 1 if balcony present 0 else), presence of garden (dummy: $garden$, 1 if garden present 0 else) and property type, either new construction or not, (dummy: $building\_dummy$, 1 if the property is a house 0 if the property is an apartment).

3.1 Funda data

Some observations of the scraped data were duplicates because some rental places had been placed on the website multiple times within the scraped timeframe. These duplicates were deleted from the dataset, and the most recent information about the rental place retained. Furthermore, very few observations did not include a price and therefore these were deleted from the dataset as well. In addition, Funda advertisements are not consistent as they suffer from human errors when entering information. Sometimes not all information is provided or wrongly entered into the website. This affects the information on the number of bedrooms and floor level. The missing data on the number

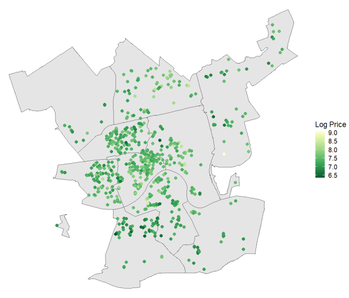


Figure 1: The $log(price)$ of each observation.

of bedrooms were manually inserted, as a few observation points were missing this information. Furthermore, about 250 offerings failed to note the floor level of the property. After text analysis of the description, the missing information on floor levels was manually inserted. For the remaining 92 observations with missing data we created a dummy equal to 1 if the floor was missing, and 0 otherwise ($floor\_dummy\_missing$). To still be able to include these observations in our dataset, we set the floor level equal to 2 for these observations, which is the median of the floor levels. After removing the duplicates, the resulting dataset includes a total of $N = 1193$ observations available for the analysis and regressions in this paper.

3.2 Rotterdam municipality data

We requested the data set of ward profiles in Rotterdam from the Rotterdam municipality website[1]. The data for 2024 was analysed since the rental properties from the data were also rented to tenants during this period, but the same data is also available for the years 2014, 2016, 2018, 2020 and 2022. The data set was based on a questionnaire: in 2023, 30.000 residents of Rotterdam participated in the study. Half of the participants answered questions about social and physical subjects, while the other half addressed safety matters. The resulting physical index describes, for example, the number of vacant buildings, the public space, and the overall living experience, while the safety index describes, among other things, the number of burglaries and nuisances in a ward.

Finally, the social index is based on questions related to self-sustainability, participation in the neighbourhood, and co-reliance. The data contains this information for Rotterdam's 14 wards, shown in Figure 2 with the physical, safety, and social index respectively. However, we exclude Hoek van Holland, Rozenburg, Hoogvliet, and Pernis from our analysis since Funda does not classify these wards as Rotterdam. The number of observations we obtained from Funda per ward can be seen in Table 1. We note that the ward Overschie has only 18 observations, while the Centrum ward evidently has the most. We take this difference into account when discussing our results, but having more observations for this ward would have been better for the analysis. In the circles in Figure 2, a dark green colour means the index score is far above 130, while yellow corresponds to an index score below 89. Considering that this figure is publicly available and it is safe to assume people research the neighbourhood they want to move to, these indices could provide variable insight into reasons prices in the wards differ. The index scores are computed as follows: first, index scores are calculated for the different themes within the social, physical, and safety categories where the city average of 2014 is chosen as the base year with an index score of 100. The themes are measured in both an objective and a subjective manner. For example, the theme "living objective" describes the quality of the homes, while "living subjective" describes how content the residents are with their homes. More information about this can be found on the website (in Dutch) [2]. Next, the standard deviation is used to calculate the distance from 100, with one standard deviation equal to 40 index points. The category index (e.g. social index) is then the average of the index scores of the themes belonging to this category. The description of the themes

District	**# Observations**
Centrum	326
Charlois	88
Delfshaven	130
Overschie	18
Prins Alexander	70
Noord	112
IJsselmonde	56
Feijenoord	186
Kralingen-Crooswijk	125
Hillegersberg-Schiebroek	82

Table 1: Number of observations per district.

[1] https://wijkprofiel.rotterdam.nl/nl/2024/dataset
[2] https://wijkprofiel.rotterdam.nl/nl/2024/hulp/themas

can be found in the Appendix.

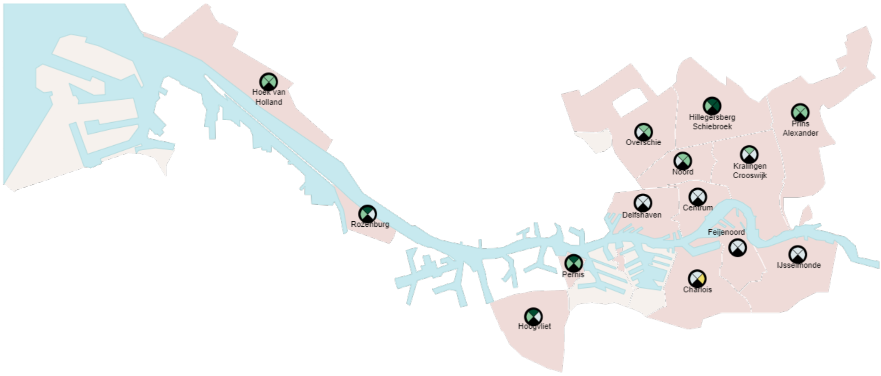


Figure 2: The 14 wards of Rotterdam with from left to right the physical, safety, and social index. Source: *https://wijkprofiel.rotterdam.nl/nl/2024/rotterdam*.

3.3 Google API Enrichment

To increase the number of explanatory variables available per observation, our dataset is enhanced with the distances to useful destinations. We first extracted the coordinates for each observation's location using Nominatim, an open-source Geocoder. Then, using the Google Places API, we obtained the location of the nearest type of destination we were searching for. Having retrieved this location we then used the Google Distance Matrix API to get the travel distance in meters and the biking distance in minutes to this location. We have specifically decided to only use the biking distance because of the following reasons; firstly, biking in the city is usually the quickest option for short distances and is a more applicable real-world measure, and secondly, if we were to then also use the travel time by, for example, car this would likely be highly correlated with the bicycle travel time. Table 2 shows the included destinations that could influence the price together with an explanation.

Point of interest	Explanation
Rotterdam Centre (Stadhuis)	The city centre is often considered a hub of activity, with a high concentration of businesses, shops, restaurants, and cultural attractions. Proximity to the city centre can indicate convenience.
Primary schools	Access to education is a significant factor for families when choosing a residential area. Proximity to primary schools can be attractive to families with young children.
Grocery stores	Access to grocery stores is essential for daily living. Properties located near grocery stores offer convenience for residents, saving time and effort on shopping trips.
Subway	Proximity to public transportation, such as subway stations, might be highly desirable for many urban residents.
Night clubs	Proximity to nightlife and entertainment venues like nightclubs could be a negative factor in making a decision on a rental property for families.
Public transport hubs	In addition to subway stations, proximity to other public transport hubs such as bus terminals or train stations could be important.
Gym	Access to fitness facilities like gyms is increasingly valued by urban residents who prioritize health and wellness. We do note that some more luxury apartments might have a gym included in their building, which would not show up using the Google Places API.
Parks	Proximity to parks and green spaces are associated with various benefits. Properties located near parks may command higher rental prices.

Table 2: The added variables from Google API.

Summary statistics on select variables are given in Table 3 below, for more details see the Appendix. It is interesting to see that we have much variation in price in our dataset, indicated by the high standard deviation. In general, the dataset contains properties built in the 1980s that are not more than a 10-minute bike ride away from the city centre.

	price	*log(price)*	*house_age*	*floor*	*safety_index*	*traveltime_stadhuis*
Mean	1707.93	7.385	51.91	4.588	99.212	683.842
Median	1595	7.375	43	2	97.986	580
St. Dev	647.73	0.333	42.578	6.62	14.408	465.092
Min	571	6.347	0	0	81.227	72
Max	8500	9.048	370	43	127.460	2533
Kurtosis	13.459	1.186	1.326	8.592	-0.770	1.021
Skewness	2.413	0.360	0.721	2.775	0.347	1.084

Table 3: Summary statistics of some variables of the resulting dataset.

4 Methodology

To investigate the presence of differences in feature effects across Rotterdam and the value of adding ward-specific data to our dataset, this paper compares the results of global and spatial methods. Before performing the estimation, the continuous predictor variables will be scaled and the dependent variable, rental price, transformed using the natural logarithm to normalize the distribution.

4.1 Ordinary Least Squares Regression (OLS)

As a benchmark model, an OLS regression is employed. This regression will be compared to its equivalent which allows for geographical differences. This OLS regression is done with forward selection of variables. By using this method, only significant features are selected to be included in the model. Moreover, the results from this regression can be used to analyse the importance of the variables. This model does not allow for geographically differing effects of the features.

4.2 LASSO Regression (LASSO)

The LASSO regression is comparable to the OLS regression with forward selection, in the way that it also performs variable selection. LASSO essentially is an extension of the OLS regression, where not only the sum of squared residuals but also the coefficients of the regression are minimized:
$\mathcal{L}_{LASSO} = \sum_{i=1}^{n}(y_i - \sum_{j=1}^{m} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{m} |\beta_j|$. By "pushing" the coefficients towards zero, LASSO selects only the most important variables. Therefore, allowing the importance ranking of variables based on the absolute value of their coefficients. The LASSO regression will be used to compare traditional methods, i.e. OLS and LASSO, to similar linear methods that incorporate geographical information, i.e. GWR.

4.3 Geographically Weighted Regression (GWR)

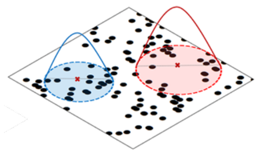


Figure 3: Adaptive kernel: constant number of nearest neighbours[3].

To assess the effects of location on the variable effects, we compare the OLS and LASSO models to a Geographically Weighted Regression (GWR). This method estimates a model for each rental property in the studied area, assigning weights to the data of surrounding observations using a kernel. The kernel is the neighbourhood in which the local model operates and is based on a bandwidth that includes the observations within radius r of the data point (in meters) in the case of a fixed kernel, or the n nearest neighbours in the case of an adaptive kernel. There are multiple options for kernel shapes; Gaussian, exponential, bisquare, tricube and boxcar. All kernel shapes, except boxcar, assign more weight to observations that are geographically nearby than observations that are further away. The shape of the kernel is set to Gaussian, as this shape is most commonly used in the literature and weighs observations according to distance to the considered observation. The kernel bandwidth considered for the GWR is a so-called "adaptive" bandwidth rather than a "fixed" one, meaning its radius is different for each observation ensuring a fixed sample size for each local regression. For example; instead of considering all properties within a 500-meter radius ("fixed" bandwidth), we consider the 50 closest observations for each regression ("adaptive" bandwidth). From now on, this paper refers to an adaptive (fixed) kernel as a kernel with an (fixed) adaptive bandwidth. See Figure 3 for a visualization of the adaptive kernel: the red circle is larger than the blue circle but contains the same number of observations. The choice is based on the paper of Suryowati et al. (2021), which showed that adaptive kernels for GWR have better results than fixed kernels. The bandwidth of the kernel is estimated using cross-validation selecting the bandwidth resulting in the lowest sum of squared residuals. Moreover, for each regression performed by the GWR, we use forward selection to select relevant predictors, based on the AICc, a version of the Akaike information criterion that has a correction for small sample sizes. For the implementation of this method we use the R package "GWModel", introduced by Gollini et al. (2013).

4.4 Random Forest (RF)

Made out of multiple decision trees, the Random Forest (RF) is a popular machine learning algorithm to perform regression analyses. The fact that it combines decision trees to make one final prediction leads to a reduced risk of overfitting considering that all trees are uncorrelated, and it is easy to measure the importance of certain regressors on the outcome. This method uses a process known as bootstrapping and aggregation: The process begins with arranging the training data, after which it is organised into samples known as bootstrap samples. The algorithm builds uncorrelated decision trees on these bootstrap samples, which reduces the risk of bias. These trees are trained individually and yield different regressions. If we combine the predictions of all trees into one value (aggregation), we have a Random Forest. The combination of bootstrapping (creating multiple random samples from the training data) and aggregation (combining predictions of models) gives Random Forest its power and versatility in predictive modelling tasks. Potrawa and Tetereva (2022) have illustrated the potential of a Random Forest in hedonic pricing models. Therefore, we will use the non-parametric model to assess the differences between a local and global model.

The regression equation for the Random Forest can hence be described as

$$Y_i = \alpha x_i + e_i, \tag{1}$$

[3]https://gistbok.ucgis.org/bok-topics/geographically-weighted-regression-framework

where Y_i is the dependent variable for observation i, αx_i is the nonlinear function of the Random Forest based on a set of explanatory variables and e_i is the error term for the given observation.

The Random Forest has several hyperparameters that need to be set. Through a grid search on the tuning data, we acquire the number of variables that are tried at each split of every tree, which is called "mtry". For more details on this tuning procedure, we refer to the Appendix and Section 4.6. In addition to this, one needs to decide the number of trees in the forest and the minimum number of data points in the terminal nodes. We implement this method using the "randomForest" package in R.

4.5 Geographically Weighted Random Forest (GWRF)

Our dataset consisting of different wards in Rotterdam might suffer from spatial non-stationarity, meaning that the true relationship between the house prices and the attributes can vary across space. Being an aspatial model, the traditional Random Forest model cannot capture these geographical relationships which we suspect to occur within the differing wards of Rotterdam. The primary contrast between a Geographically Weighted Regression and a Geographically Weighted Random Forest lies in flexibility. GWRF proves advantageous for datasets with many predictors, owing to the robustness of the Random Forest algorithm in handling high dimensionality.

In practice, for each home in our dataset, a local Random Forest is fitted using data on observations based on how geographically nearby they are using a kernel and bandwidth. See an explanation for kernel and bandwidth in Section 4.3. Unlike the GWR method, we examine both the fixed and adaptive bandwidth for the GWRF model and determine the best method. In this research, we label the optimal bandwidth the one with the lowest out-of-bag mean squared error on the tuning set. The GWRF creates a forest of t trees for every data point i in the training set. In this case, the regression equation is equal to

$$Y_i = \alpha(u_i, v_i)x_i + e_i, \tag{2}$$

where Y_i is the log price of the i^{th} house, $\alpha(u_i, v_i)x_i$ is the predictive function of a Random Forest model calibrated on location i, and u_i, v_i are the spatial coordinates of the home in Rotterdam. In turn, the predictions of the GWRF model are given by $\alpha(u_i, v_i)x_i$ where the closest available GWRF model is used to make predictions for the data points in the testing set. Another interesting aspect of the GWRF model is that we can put different weights on the local and global Random Forest models: we merge the GWRF global and local model predictions with the as in Wu et al. (2024). For the predictions, this means that

$$P_i = wP_{local,i} + (1 - w)P_{global}, \tag{3}$$

where P_i is the predicted value for the i^{th} property and w is the weight between 0 and 1 that is put on the local model for that observation. In this paper, we put weights of $0.25, 0.5, 0.75,$ and 1 on the local model and investigate the performance on the test set.

4.6 Model Assessment

To assess the model performance, the data is split into two subsets; 15% for tuning parameters for Random Forests, and 85% for training and testing. This tuning dataset is excluded from training and testing for every model. To assess the predictive performance of the models considered in this paper, k-fold cross-validation is used. This means that we split the data that is left after tuning in k parts, where $\frac{1}{k}$ share of the data is used as a testing set and $1 - \frac{1}{k}$ share is used for training. In turn, each observation has been used once in the testing set and $k - 1$ times in the training set. By using k-fold cross-validation, we reduce the randomness in the performance and test the ability of the models to generalize to new data sets. In this study, The choice of $k = 10$ is used for the regressions, as this value is commonly used in the literature and ensures a sufficiently large test set for each fold. The cross-validation method accounts for the effects that the choice of training and test data sets might have on the performance of the models. To assess the performances of the models we average the performance measures of each model for each fold. Note that the variables are scaled in each iteration of the k-fold method. As performance measures, we use the Mean Absolute Error (MAE), Mean Squared Error (MSE), absolute percentage prediction error, and R^2, which are defined as

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y_i}|}{n}, \text{ and} \tag{4}$$

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y_i})^2}{n}, \tag{5}$$

$$Absolute\ percentage\ prediction\ error = |\frac{y_i - \hat{y_i}}{y_i}| * 100\% \tag{6}$$

where y_i is the observed logarithm of the price of the home, $\hat{y_i}$ is the prediction of the logarithm of the price and n is the size of the test set.

4.7 Interpretability

After assessing the predictive performances of all models considered in this paper, one of the models is analysed, using Partial Dependence plots (PDP), Accumulated Local Explanation (ALE) plots, and the Local Interpretable Model-agnostic Explanation (LIME) method.

PDP provide a global understanding of how the price changes as a function of a specific predictor variable which could help increase interpretability. In addition to PDP, ALE plots are considered, as they are more reliable in correlated data environments, which we are likely to have in our research. PDP isolate the effect of the investigated features by averaging out the effects of all other features. This is done by inserting all possible values (from the dataset) for the feature of interest and making predictions. Lastly, all these predictions are averaged and plotted against the dependent variable. An issue that might arise in the case of correlated data, is that to construct the PDP, predictions are made on unrealistic observations. For example, it might consider an observation (property), with 10 rooms and a rental price of 1 euro. This is where ALE plots come into play. ALE plots only consider data points with "realistic" characteristic combinations. Additionally, rather than averaging the predictions the plot calculates the change in predictions in intervals. Thus, decreasing bias and improving accuracy. We will illustrate the effects of the most interesting variables based on expected versus realized patterns and interpretability.

The other method that will be used is LIME, a technique used to explain the predictions of models on a local level. We can obtain explanations for individual predictions, making it easier to understand why a particular prediction was made. This is done by approximating the black-box model on a local level with an interpretable model, therefore allowing for interpretation on a local level. These explanations can be presented in terms of feature importance, showing which features contribute the most to a specific prediction. An observation from each ward will be investigated with LIME, such that a comparison can be made between wards regarding interpretability and some conclusions can be drawn for each ward specifically through LIME. These observations per ward will be constructed by taking the average of the score of the features for each observation in the same ward in the test set. This will give an overview of the average effects of features of rental properties in each ward. The goal is to explain the feature differences on a local level and spot differences between wards using this method.

Moreover, the three linear models, OLS, LASSO, and GWR, are analysed to assess variable importance. This is done based on the resulting coefficients of the variables from each regression. Again, as done for the linear models, the five most important variables are identified.

5 Results

As mentioned in Section 4.6, we have made multiple partitions of test and training data sets according to a k-fold algorithm. Firstly, this section discusses how the bandwidth for the geographically weighted models is selected. Afterwards, the performance of these models is then evaluated based on their predictive capability according to the measures also discussed in Section 4.6. Next, the performance per ward is analysed. Finally, the importance of the variables and interpretation thereof are discussed.

5.1 Tuning the bandwidth

First and foremost, this research uses the tuning set with all the variables included to obtain our bandwidth. The procedure for choosing the bandwidth has four steps: first, a bandwidth is chosen based on the lowest cross-validation score, which is the sum of the squared residuals. The bandwidth needs to be tuned because a bandwidth that is too small will not include enough observations for robust regression, while a bandwidth that is too big will not be able to differentiate between local effects, as it includes too many observations to differ significantly from a global model. With this optimal bandwidth, the aim is to choose the best GWR model. Next, for this bandwidth, the GWR algorithm selects the optimal model by trying all possible combinations of the variables, selecting the model with the lowest AICc. Then the chosen optimal model is again used to find a new optimal bandwidth using the same cross-validation technique. And lastly, the resulting model and bandwidth are used to predict on the test set of that fold. Step 1 is executed before the k-fold cross-validation, steps 2, 3 and 4 are repeated in each fold of the k-fold. The selected model regression for each of the k folds is presented in the Appendix. Figure 4 shows the bandwidth selection resulting from the cross-validation for each fold (iteration). This figure shows that different bandwidths are chosen for each iteration.

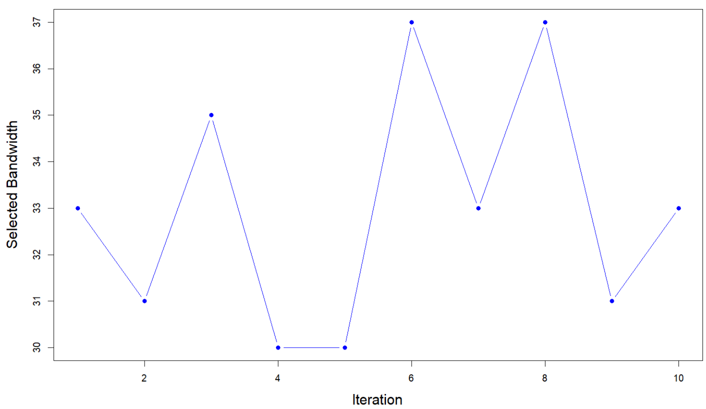


Figure 4: Selected bandwidth for each of the k iterations for the GWR.

In the case of the Geographically Weighted Random Forest, we estimate a GWRF for the tuning set to determine the best bandwidth. Note, however, that the best bandwidth for the tuning set is not necessarily the best bandwidth for the training set. For choosing the bandwidth of the adaptive kernel for the GWRF, bandwidths from 10 to 170 nearest neighbours with steps of 10 neighbours are investigated. The best bandwidth is chosen based on the out-of-bag Mean Squared Error of the forest made on the tuning set. For the fixed kernel, bandwidths from 60 to 600 meters are investigated, again with steps of 10 meters. As shown in Figure 5, the best kernel width for the tuning set is 40. For the fixed kernel with 200 trees, this is a kernel of 90 as shown in Figure 6.

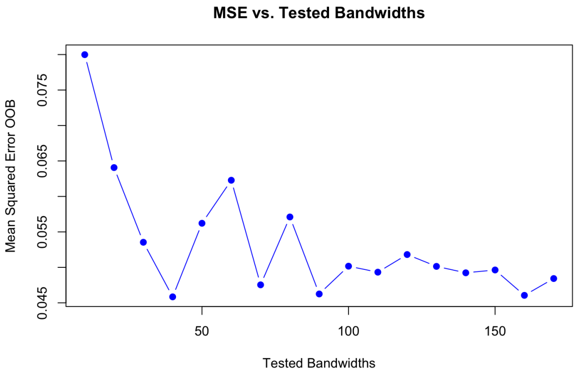


Figure 5: The tested bandwidths for the GWRF adaptive kernel.

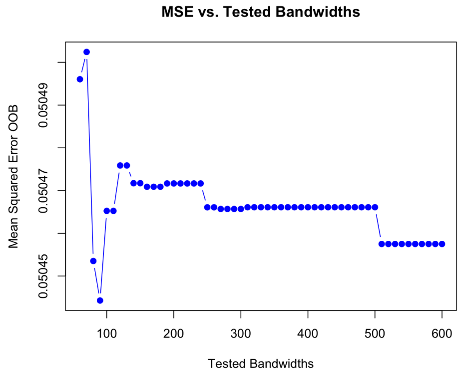


Figure 6: The tested bandwidths for the GWRF fixed kernel.

In Figures 5 and 6 Out Of Bag MSE (OOB MSE) values are shown for different bandwidths for both an adaptive and fixed kernel. Interestingly, the difference in OOB MSE for the tuning set differs more for different bandwidths in the adaptive kernel model than in the fixed kernel model. A possible explanation is that for neighbourhoods with fewer observations such as Overschie, the adaptive kernel may take into account observations from different neighbourhoods: if you consider, for example, a point in Overschie, but take the 50 nearest neighbours into account when estimating the local GWRF regression, these observations will come from Kralingen-Crooswijk and Hillegersberg-Schiebroek. One observes for the adaptive kernel that especially the lowest bandwidth of 10 has the highest OOB MSE. For an aberrant observation, the 10 nearest neighbours might come from a different neighbourhood, while with a fixed bandwidth of 10 meters, no other properties will be weighted into the local model. If the 10 nearest observations are included with an equal weight of $\frac{1}{10}$, the prediction might be inaccurate. As weighing observations that are geographically far as heavily as observations that are close, the model might be trained on information that is less relevant for that location. In contrast, for the Geographically Weighted Regression these observations will be given less weight due to the Gaussian kernel. Therefore weighing the data on its geographical relevance, and thus including only relevant data. Consequently, the fixed kernel results in the best performance for GWRF, and an adaptive kernel for GWR.

5.2 Predictive performance comparison

The focus of this paper is on the predictive capability of the models, and therefore Mean Absolute Error (MAE) and Mean Squared Error (MSE) are the main performance measures. In Table 4 an overview is shown of the MAE, MSE and in-sample R^2 for all models.

Model	MAE	MSE	In-sample R^2
OLS (forward selection)	0.1611	0.0446	0.6373
LASSO	0.1607	0.0447	0.5374
GWR (adaptive kernel)	**0.1450**	0.0387	0.6397
RF	0.1470	0.0395	0.5798
GWRF (fixed kernel)	0.1465	**0.0383**	**0.7020**

Table 4: Performance measures of all models.

From the MSE and MAE values it can be concluded that the models all perform quite well, as the differences are relatively small. Looking at the linear regression models, OLS (LASSO) and GWR, we find that the GWR gives the most accurate predictions, with noticeably smaller MAE and MSE values. When comparing the Random Forest models, RF and GWRF, the MSE value of GWRF is smaller than that of the RF. However, dissimilar to the linear models, the values appear to differ only slightly.

One explanation for why GWRF improves upon RF by only a small amount while GWR improves upon

LASSO with a noticeable amount is that the RF model approaches the optimal predictive performance for this dataset. Therefore, incorporating geographical information into the model can only show a small improvement, as there is little room for improvement left. If RF and GWRF are compared on a bigger and more complex dataset, the predictive differences might prove to be bigger.

Finally, this subsection evaluates the different weights put on the local Geographically Weighted Random Forest.

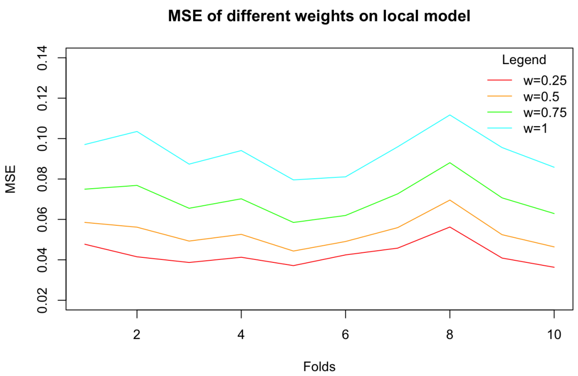


Figure 7: MSE of the adaptive kernel. (GWRF)

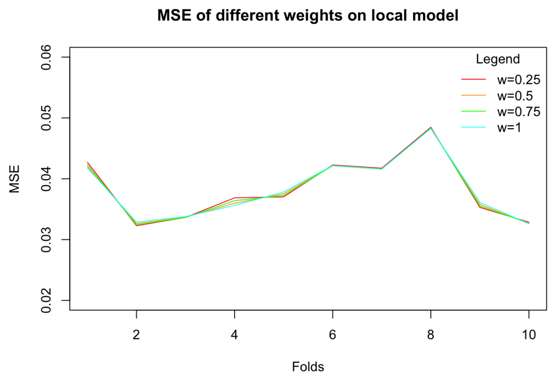


Figure 8: MSE of the fixed kernel. (GWRF)

In Figures 7 and 8 the MSE values for different weights on local and global models of the GWRF are shown for both the fixed and adaptive kernel, and plotted for each fold executed in the K fold algorithm. For the adaptive kernel, the more weight put on the local model, the worse the performance. As explained before, the adaptive kernel is sensitive to outliers and data with different sampling frequencies. Therefore, the local models do not perform well. Instead, for the fixed kernel, all models perform relatively the same. This can be derived from Table 4, considering that the MSE for the RF and GWRF are relatively close. Fusing the predictions from global and local models, in this case, does not lead to an immense improvement in predictive accuracy. However, it can be concluded that the adaptive kernel for the GWRF model is not appropriate for modelling the prices of rental properties in Rotterdam.

5.3 Performance per ward

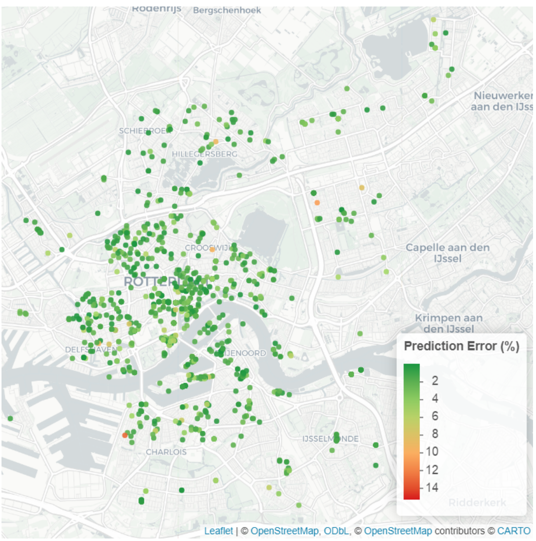


Figure 9: Percentage error for each observation for the GWR.

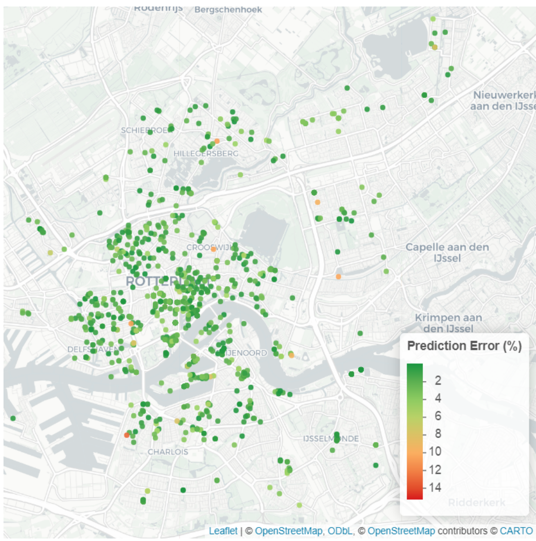


Figure 10: Percentage error for each observation for the GWRF.

This subsection compares and explains the differences in performance per ward. Shown in Figure 11 - see Appendix for a bigger figure and a plot of all coefficients for each observation - are the absolute percentage prediction errors for every ward for every model, averaging all percentage prediction errors of all observations within the wards.

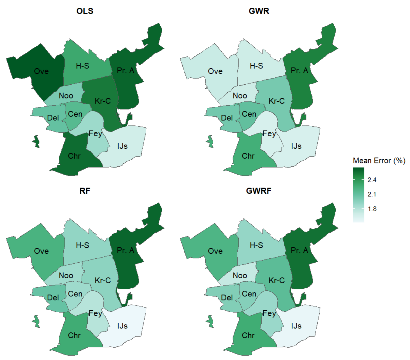


Figure 11: Percentage error per ward for the four models. For the abbreviations, see Appendix.

In addition, in Figure 9 and 10 all observations excluding the tuning set are plotted for GWR and GWRF, and coloured according to their absolute percentage prediction error. Note the red-coloured observation in Charlois in Figure 9. This can be considered an outlier because it has a high error mainly because of the unique combination of a bigger living area combined with a lower price, in which this observation has the highest deviation. Moving on, Table 1 shows there are relatively fewer observations in certain districts, and this scarcity of observations for certain wards has different effects on GWRF and GWR. This difference is highlighted in the wards Prins Alexander and Overschie and could have several explanations. First, as GWR is tuned to an adaptive bandwidth, it considers the closest e.g. 30 observations for the regressions. Therefore, it includes observations that are geographically far from the considered rental property thus ensuring sufficient local sample size for the local regressions. Also, GWR weighs these observations according to geographical distance to the considered observation. This means giving more weight to nearby properties than those far away, thus enabling the method to use an adaptive kernel without incorporating geographically far observations in such a manner that they obstruct the locality of the model. On the other hand, GWRF is tuned to a fixed bandwidth, thus including observations within the given bandwidth as a radius. This results in only including geographically close observations, although subsequently, its individual regressions are not certain to have sufficient local sample size. In particular, Figure 10 confirms that some individual errors of GWRF are bigger than those of GWR for wards with few observations. This is in line with what Figure 11 shows.

The results for OLS and GWR yield noticeable differences. Overall, the mean absolute error for GWR is smaller than for OLS, as shown in Table 4. More interesting to see is that errors are more evenly distributed for the GWR model whereas for the OLS model, some wards have evidently bigger errors than others. Consequently, this shows that the GWR can differentiate between locations of properties and implement this information in its predictions, improving its predictive performance. The biggest improvements are achieved in the wards indicated by Ove (Overschie), Chr (Charlois) and Kr-C (Kralingen-Crooswijk). The dataset contains only 18 observations for Overschie. Therefore it could be that the global OLS model focuses on fitting observations in more data-rich wards such as Centrum, while the GWR estimates local models for these smaller parts of the data set and can thus capture the specific patterns in these wards better. Another explanation is that GWR overfits the data, and therefore makes more accurate predictions in general.

When comparing the RF and GWRF models, there are no noticeable differences in Figure 11. This is supported by the results in Table 4, which indicate that the difference between the predictive performances of the global Random Forest model and the Geographically Weighted Random Forest model with full weight on the local component is small. Therefore, the figure does not show differences: the incorporation of geographical information in the model improves predictions by only a relatively small amount, which is again supported by Table 4.

5.4 Explanation of differences in predictive performance

The fact that GWR outperforms OLS on predictive performance with forward selection can be attributed to different factors. Firstly, a possible explanation is that the GWR is a more complex model than OLS. The fact that GWR has more tuning parameters, means that it could overfit the data, and explain the noise in it. This would mean that the GWR model would not outperform OLS on different datasets, and could be shown in further research. This would also explain why RF performs well at predictive tasks. Assuming the high predictive performance of GWR is due to overfitting, Random Forest models are outperforming the linear models. This could indicate that the true data-generating process could be non-linear, as Random Forest models are able to capture such relationships. Then the fact that GWRF only slightly outperforms the RF could again be explained by overfitting.

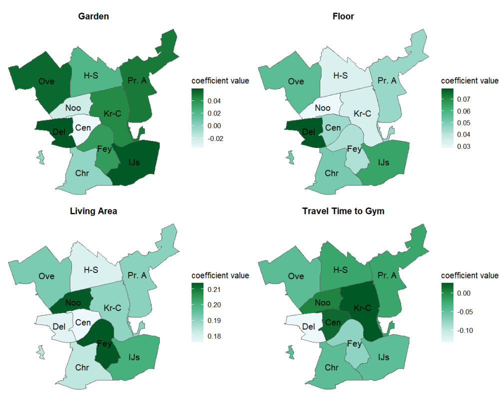


Figure 12: Average coefficients for various variables for each ward. For the abbreviations, see Appendix.

Secondly, the differences in predictive performance could be due to the presence of spatial non-stationarity in the data. This would mean that the effects of explanatory variables on the dependent variable differ per location. GWR would then outperform OLS, as it is able to capture these geographically varying coefficients. To investigate this explanation, we have calculated the average coefficient from GWR for various explanatory variables for each ward, and shown them in Figure 12 - see Appendix for a bigger plot - the darker the green, the more positive the coefficient value is, note the different scales for the plots. Interestingly the estimated coefficients vary a lot throughout the wards. Especially the variable *garden* varies a lot, considering the relatively small coefficients of the ward Centrum. This could be because renters aiming to live in the centre of Rotterdam value location higher than having a garden. Thus resulting in a lower coefficient for that variable.

These two explanations could be both applicable at the same time. For the second explanation, Figure 12 shows possible evidence. However, regardless of the explanation, on this specific dataset, the geographical models outperform the traditional models.

5.5 Variable importance

All linear regression models considered in this paper allow for variable importance evaluation. The forward selection of OLS and GWR, and the LASSO regression provide us with sparse models by performing variable selection among the large set of explanatory variables. As the variables have been scaled, the size of the absolute coefficients can be used as a proxy to rank variables on importance; the higher the absolute value, the more important the variable. See Appendix for the details of the rankings, an overview of the results of GWR is given in Figure 13. This figure shows the absolute average coefficients of the variables from GWR, ranked on size, thus on the left are the most important variables and on the right are the least important variables. Interestingly, the rankings are almost identical for LASSO and OLS, with only minor differences for GWR. Overall, the most important variables are living area, safety index, social index, travel time to the city centre, and floor. All models agree that living area is decidedly the most important variable, with its absolute coefficient being more than twice the size of the second most important variable, for each model. This is in line with the expectation that living area is the most important factor for rental prices, as it greatly contributes to the fundamental utility that people gain from properties. It is also noteworthy that adding safety index and social index seem to be significant in determining the price of a rental property, and therefore adding them to the regression proved to be useful.

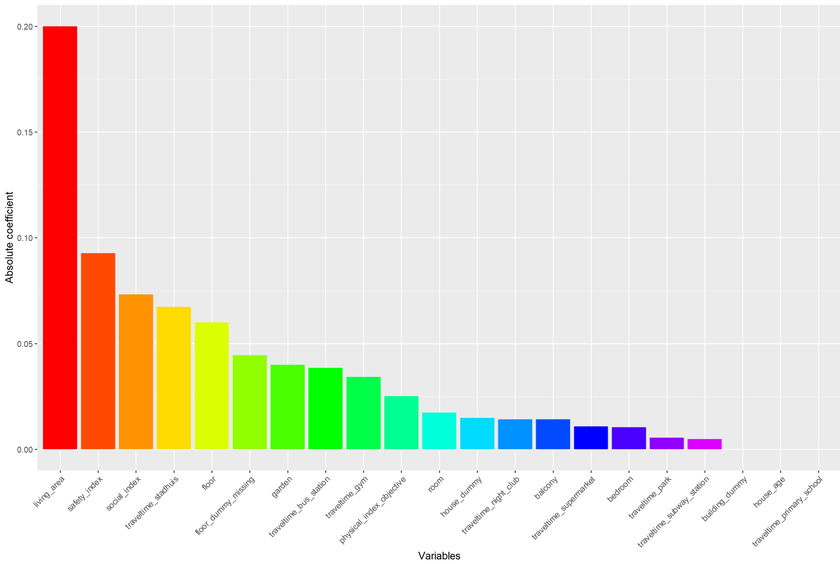


Figure 13: GWR variable importance.

It is interesting to see that the coefficient for safety index is negative (see Appendix), while also being among the most important variables for each model. This would mean that the higher the safety index, the lower the price, so properties in safer areas would be cheaper. This can be the result of observations in two wards: Centrum and Prins Alexander. Centrum has the lowest safety index and Prins Alexander has the highest safety index out of all the wards considered in this paper. As shown in Figure 1, Centrum contains many properties with relatively high prices, while in Prins Alexander there are many properties with relatively low rental prices. Thus causing the relationship of $safety\_index$ on rental price to be negative.

5.6 Interpretability

Now that we know which model performs best, we try to interpret one of the models and ascertain the inner workings of the model. As shown in Table 4, GWRF and GWR are the models with the best predictive performance. However, as the RF model performs almost as well as GWRF and GWR, and allows for more detailed and intuitive interpretation, this model will be interpreted in the following section.

The results that capture the global effects are obtained through Partial Dependence plots. ALE plots are used to check the validity of the PDP. These graphs are only made for the most important variables that are established in the previous section since these are the ones that significantly impact the model's predictions.

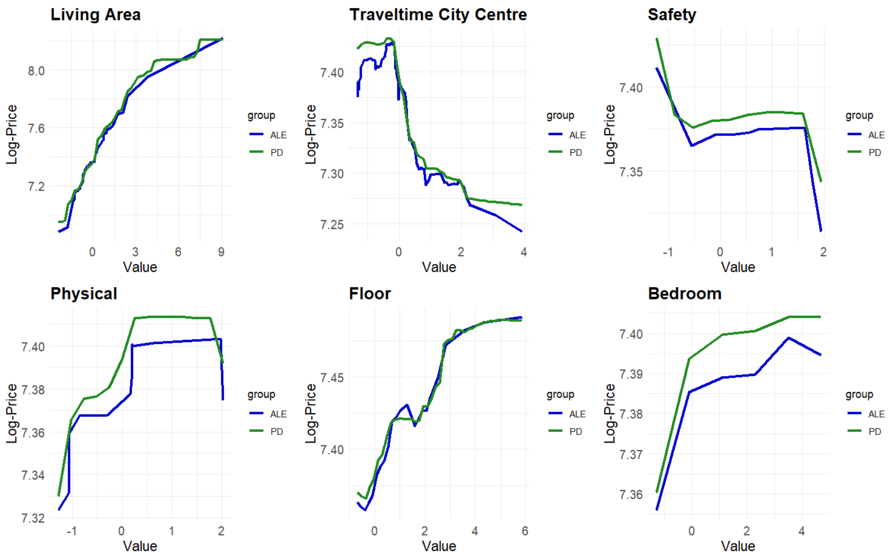


Figure 14: PD and ALE plots of six important variables.

The results in Figure 14 show that the influence of the most important features is consistent with what each feature entails. For example, the PDP of the living area feature shows that for a higher living area, the price will go up. This is also the case for the floor, bedroom and physical index features, except the bedroom and physical index features show a slower increase after passing the zero value. This indicates for the bedroom feature that the marginal effect on the price of having an extra bedroom is relatively low according to our model. The physical index feature also has a clear interpretation: a generally better living experience is also reflected in the price. It also shows that when the area code has a higher than average score, the difference with the average matters relatively little. This is also the case for the safety feature while looking at the average values. The price is only affected by the low and high values of this feature. The safety feature itself is negatively correlated with the price, which was also concluded from the other models. The same conclusion as in Section 5.5 can be drawn concerning the reason why the safety feature behaves in this manner, but the PDP shows that the negative impact only occurs for relatively high values of the index. Which would be in line with the given explanation. The plot of the travel time to the city centre feature shows that the price drops when the biking trip to the centre is lengthy, which is to be expected.

When examining the difference between the PDP and ALE plots, they appear to be quite similar. Especially for the living area, travel time, and floor features the similarities stand out. For the other features the ALE plots seem to be the same shape as the PDP but shifted upwards. A possible explanation is the nature of the features, as the physical and safety features are both measures for entire zip code areas and thus have similar values across observations. The shift in these features may be attributed to the cumulative nature of the ALE calculation. As the ALE plot accumulates the local effects of the feature across the data distribution, any consistent bias or pattern in the feature, such as the uniformity of values within zip code areas, becomes magnified. Overall, the ALE plots show that the possible correlation between features does not interfere with the interpretability results.

Next, we look at the LIME model to try to give interpretation to single observations. As discussed in the methodology the single observations that are chosen are the average features of each ward and explaining these features can give insights about the wards specifically.

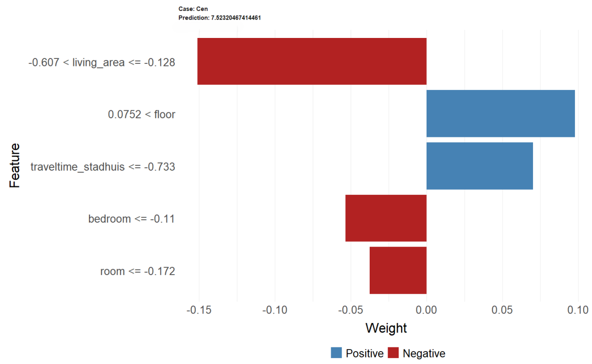


Figure 15: LIME for observations in ward “Centrum”.

One of the results of LIME is shown in Figure 15. This figure shows how a prediction is made for an individual observation of the ward “Centrum”. The prediction is made by attributing weights to the values the features obtain. In this case, the living area is lower than a certain threshold (-0.128), so this negatively impacts the level of the prediction. The specific thresholds are determined based on the significance of the features, when checking alterations of the local model. On the other hand, there is the travel time to the city centre, which is also below the average value, but this feature positively affects the price, as was also concluded when discussing the PD plots. The five most important features are shown and this figure gives insight into how a single prediction for this ward is constructed.

The results of a similar analysis for the other nine wards are depicted on the horizontal axis in Figure 16, with all of the abbreviations of the wards in the Appendix. On the vertical axis, the different importance inequalities are shown, similar to Figure 15, except now including them for all ward observations. One observes that features such as travel time to the city centre and living area are again very important since more weight is put on these features for most observations. We do notice some slight differences between the other feature importance techniques that are used. The safety and physical indices do not have any outstanding importance, compared to the other feature importance methods. This could be because LIME creates an interpretable local model based on a single observation, while the safety and physical indices continue to be ward-specific data and could therefore be less important on a local level.

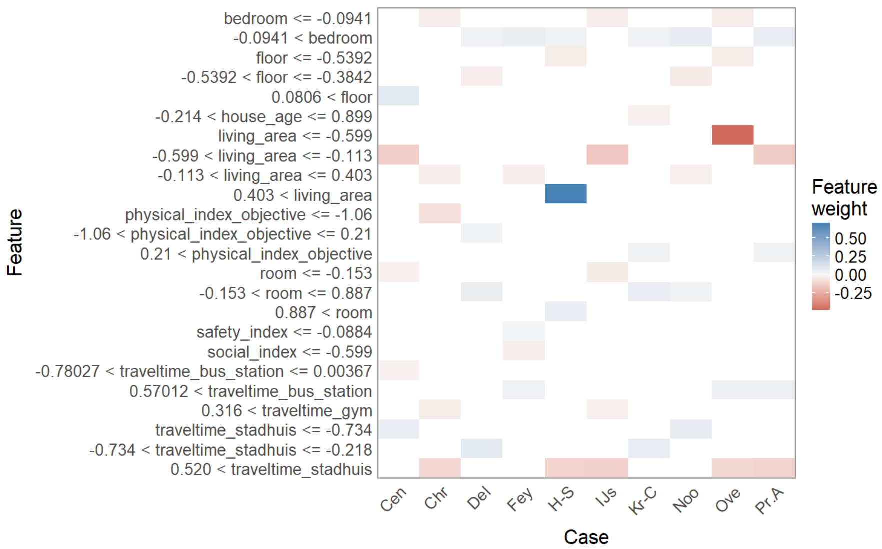


Figure 16: LIME for each ward.

Another thing to note is the differences between the wards. Because each observation is a generalization of a ward, we can paint a general picture of each ward using the local interpretation. Generally, the location-based features such as travel time to the centre correspond with the location of the wards. This entails that wards that are close to the centre, such as Centrum and Delfshaven have a positive weight, while more remote wards such as Charlois and IJsselmonde have a negative weight.

The other wards that are distinctive are Hillegersberg-Schiebroek and Overschie. These wards have a remarkably positive and negative weight for the living area feature, respectively. This indicates that living area is vitally important in these wards and could mean that on average properties in Hillegersberg are larger and properties in Overschie are smaller in size. This figure shows on average the characteristics of each ward and the feature importance that is attributed to each feature.

We do note, however, that averaging the observations in each ward for performing LIME can interfere with the "local" aspect of LIME. In LIME analysis an observation is often imputed from a real observation using the distribution of each of the features. Because this research aims to explain the entire ward's behaviour compared to a single observation, our interpretation might be over-simplified. However, the average vector per ward that is created can be viewed as a single average observation and the question remains: How much information concerning the entire ward is encapsulated in this average? We believe that it gives sufficient information regarding the ward, but this can be discussed and investigated in further research.

In general, this LIME analysis could be useful in determining the needs of a resident of a certain ward. By leveraging such insights, policymakers, urban planners, and real estate professionals can tailor interventions and investments to better serve the diverse communities across Rotterdam, since this figure shows that the prediction for each ward is based on different features, with different weights. Balancing these needs could provide better insights for these organizations and individuals. For example, if new houses were to be built in the area H-S, the municipality or project manager in control could know what the needs are of the general residents in this area. After obtaining useful information regarding the pricing intricacies of the ward, the municipality or project manager could play into this information and decide what kind of houses to build for the rental market. In this way, they can control the demographic of the ward, ensuring that the image or the (ward-specific) characteristics of the ward stay the same or change them. This could lead to a process called gentrification and there exists some debate whether this process is beneficial for the districts in a city or not (Slater (2011)). Even so, when stakeholders understand the preferences and needs of residents well, they can plan urban development more effectively. They can aim for results that can improve the quality of life, while still preserving what makes each ward special.

6 Conclusion

In literature, conventional hedonic pricing models have often been used to gauge rental prices by considering different property features. Nonetheless, these models frequently fail to account for spatial nonstationarity. In this study, it has been shown that models that allow for spatial nonstationarity (GWR and GWRF), for the city of Rotterdam, outperform standard linear models such as OLS with forward selection and LASSO, and the Random Forest model on predictive power. It is found that this fact can be attributed to either the capability to take spatial nonstationarity into account or the overfitting of the model.

Firstly in this paper, data for Rotterdam was gathered from the popular housing site Funda and enhanced with crime, social, and physical index data from the municipalities and travel times using the Google Places API. This data is unique in the way that it incorporates information about the ward in which the properties are located, thus providing new insight into the effects of such wards' characteristics on the rental prices in Rotterdam. In the methodology, two common hedonic pricing models and their geographically weighted counterparts were discussed. In our results, we find that incorporating geographical information into conventional hedonic pricing models does improve the predictive accuracy of a Random Forest model, but only by a small amount. Nevertheless, for linear models, the performances do differ with a relatively big amount. The effects of incorporating geographical information into a Random Forest model might be more prominent when investigating a more sizeable and diverse area than Rotterdam. Additionally, most prior machine learning studies focused on housing markets have neglected to interpret and evaluate the covariates utilized in their models. However, in this study, the issue of interpretability was tackled through the application of LIME, Partial Dependence plots, and ALE. In general, we find that the global behaviour of features coincides with our expectations. PDP and ALE make the interpretation of the features easier for a Random Forest Model. Similarities of PDP and ALE indicate minimal impact of correlation between features. The LIME method gives us insight into the local explanation of the model. By looking at the average observation of each ward, we can paint a picture of the priorities of each ward. We find that the importance of features varies locally among wards. This results in wards being more impacted by specific features than others. This paper finds that living area, safety index, social index, and travel time to the city centre are the most important variables, and hence adding ward-specific data gives valuable insight into Rotterdam's rental market.

The approach proposed in this paper has certain limitations that can be identified. Firstly, the data gathered on the rental houses and apartments is, while extensive, still not optimal, since it includes a relatively low set of observations to draw conclusions. Moreover, the data used and results drawn do not take into account information on time, inflation or general housing availability in Rotterdam. As time and inflation affect future rental prices, the results of the models in this paper need to be adjusted for further prediction uses. Also, as there is currently a housing crisis in the Netherlands, this scarcity could be affecting housing prices. However, as the time frame in which this data is collected is relatively small, the effect of these limitations on the conclusion of this paper is limited.

Additionally, this paper focuses solely on the city of Rotterdam. Further research is necessary on different and more extensive datasets to investigate the possibility of extrapolation of our conclusions. Additionally, further research could investigate the reasons why geographical hedonic pricing models improves their predictive performance over their aspatial versions.

The conclusion of this research is as follows: your neighbour does not pay less for their balcony compared to you, but your friend who lives in another ward does.

Appendix

Explanation of categories

Physical index

- Living: the quality of the homes and the buildings in the area
- Public Space: image quality of the public space and traffic safety in the neighbourhood
- Environment: air quality and sound quality
- Services: the proximity of various facilities in the area

Social index

- Self-reliance: the collection of resources available to an individual to be able to (continue to) participate in society. This includes, among other things, the level of income, the absence of debt problems, the education level and the health of residents.
- Social cohesion: the active commitment of citizens (whether or not in an organized context) to support and care for others, as well as the collective commitment of citizens to realize public, and social interests.
- Participation: labour market position, volunteer work, residents' initiatives and policy participation, leisure activities, such as culture, sports, club life and going out.
- Bonding: the involvement and responsibility that residents feel and take for their immediate living environment.

Safety index

- Theft
- Violence
- Robbery
- Vandalism
- Nuisance

Summary statistics

	Mean	Median	St Dev	Min	Max	Kurtosis	Skewness
price	1707.93	1595	647.73	571	8500	13.459	2.413
log_price	7.385	7.375	0.333	6.347	9.048	1.186	0.36
house_dummy	0.07	0	0.254	0	1	9.493	3.388
building_dummy	0.034	0	0.18	0	1	24.969	5.189
room	3.174	3	0.994	1	9	3.687	1.211
bedroom	2.098	2	0.843	1	7	2.325	1.032
living_area	90.961	86	32.42	13	382	10.537	2.137
house_age	51.91	43	42.578	0	370	1.326	0.721
floor	4.588	2	6.62	0	43	8.592	2.775
floor_dummy_missing	0.077	0	0.267	0	1	8.09	3.174
balcony	0.652	1	0.476	0	1	-1.594	-0.64
garden	0.246	0	0.431	0	1	-0.613	1.178
traveltime_primary_school	130.987	106	108.583	1	699	9.696	2.682
safety_index	99.212	97.986	14.408	81.227	127.46	-0.77	0.347
social_index	97.894	94.048	12.2	80.175	131.618	1.601	1.325
physical_index_objective	103.275	105.13	8.613	92.381	120.936	-0.351	0.641
traveltime_night_club	245.884	197	203.726	1	1924	6.79	2.011
traveltime_subway_station	298.867	234	241.516	14	1426	4.857	2.007
traveltime_gym	102.102	71	130.036	0	1277	49.958	6.07
traveltime_supermarket	105.3	78	108.196	1	925	23.924	3.842
traveltime_bus_station	330.705	328	205.122	5	1368	3.943	1.356
traveltime_park	112.406	93	92.681	1	1219	48.245	4.767
traveltime_stadhuis	683.842	580	465.092	72	2533	1.022	1.084

Table 5: Summary statistics dataset used for analysis.

Tuning

As some of our models require tuning, we reserve 15% of the available data for this purpose. This data will not be used in either the training data or the test data. Both the global Random Forest and the Geographically Weighted Random Forest are tuned on this data, to find the optimal number of variables per split (mtry) of all trees. We tune this parameter by creating a Random Forest for mtry between 1 and 21, which are the minimum and maximum possible number of variables and select the best Random Forest based on the lowest Out-of-bag Mean Squared Error. We find that the optimal value is 13 for the global random forest and the geographical random forest. Note however, that the best mtry for the tuning set is not necessarily the best mtry for the training set. We keep the number of trees for RF at 500 and 200 trees for GWRF considering the costly computation, and the minimum number of data points in the terminal nodes, which we leave at the default value of 5 given by the "randomForest" package in R.

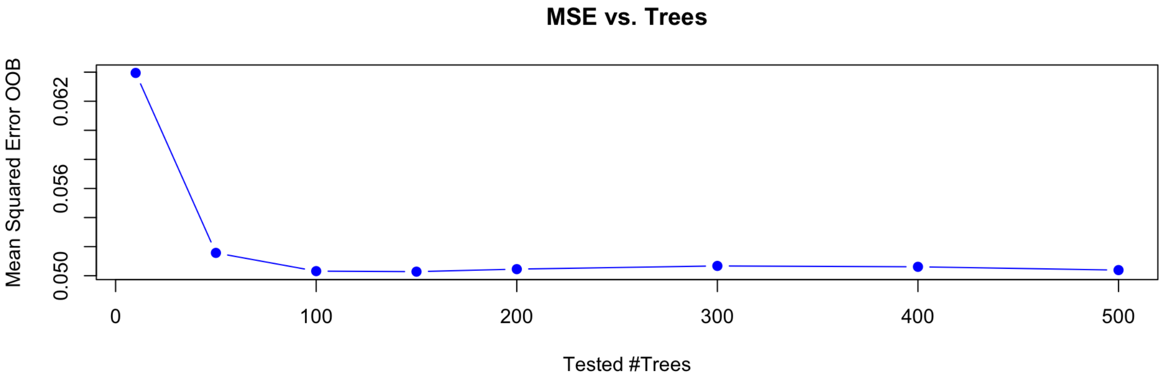


Figure 17: The influence of the number of trees on the OOB MSE for GWRF.

Results OLS, LASSO and GWR coefficients and variable importance

OLS with forward selection		LASSO		GWR	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
Intercept	7.3578	Intercept	7.3644	Intercept	7.3664
living_area	0.1926	living_area	0.1872	living_area	0.2000
social_index	0.1011	social_index	0.0838	social_index	0.0732
building_dummy	0.0712	floor	0.0552	floor	0.0599
floor	0.0582	building_dummy	0.0518	garden	0.0400
traveltime_bus_station	0.0397	traveltime_bus_station	0.0382	traveltime_bus_station	0.0385
garden	0.0364	balcony	0.0351	physical_index_objective	0.0251
balcony	0.0324	garden	0.0316	room	0.0173
room	0.028	room	0.0204	traveltime_night_club	0.0142
traveltime_primary_school	0.0126	traveltime_primary_school	0.0147	balcony	0.0141
bedroom	0.0024	bedroom	0.0095	traveltime_supermarket	0.0108
traveltime_subway_station	0.0011	physical_index_objective	0.0072	bedroom	0.0104
house_dummy	0.0000	traveltime_supermarket	0.0037	traveltime_park	0.0054
house_age	0.0000	traveltime_subway_station	0.0019	traveltime_subway_station	0.0048
physical_index_objective	0.0000	traveltime_night_club	0.0010	building_dummy	0.0000
traveltime_night_club	0.0000	traveltime_park	-0.0024	house_age	0.0000
traveltime_supermarket	0.0000	house_age	-0.0078	traveltime_primary_school	0.0000
traveltime_park	0.0000	house_dummy	-0.0127	house_dummy	-0.0148
traveltime_gym	-0.026	traveltime_gym	-0.0265	traveltime_gym	-0.0342
traveltime_stadhuis	-0.058	traveltime_stadhuis	-0.0628	floor_dummy_missing	-0.0445
floor_dummy_missing	-0.0732	safety_index	-0.0803	traveltime_stadhuis	-0.0674
safety_index	-0.0957	floor_dummy_missing	-0.0957	safety_index	-0.0927

Table 6: Coefficients from OLS with forward selection, LASSO and GWR.

Regression #	house_dummy	building_dummy	room	bedroom	living_area	house_age	floor	floor_dummy_missing	balcony	garden	safety_index	physical_index_objective	social_index
1	X		X		X	X	X	X	X	X	X	X	X
2	X		X		X		X	X	X	X	X	X	X
3	X		X		X	X	X	X	X	X	X	X	X
4	X		X		X		X	X	X	X	X	X	X
5	X		X		X		X	X		X	X	X	X
6	X		X		X		X	X	X	X	X	X	X
7	X		X		X		X	X		X	X	X	X
8	X			X	X		X	X		X	X	X	X
9	X		X		X		X	X		X	X	X	X
10	X		X		X		X	X		X	X	X	X

Regression #	traveltime_primary_school	traveltime_night_club	traveltime_subway_station	traveltime_gym	traveltime_supermarket	traveltime_bus_station	traveltime_park	traveltime_stadhuis	**AICc**
1		X	X	X	X	X	X	X	-345.0597
2		X	X	X	X	X	X	X	-385.3139
3		X	X	X	X	X	X	X	-405.0620
4		X	X	X	X	X	X	X	-403.7255
5		X	X	X	X	X	X	X	-401.4105
6		X	X	X	X	X	X	X	-390.5252
7		X	X	X	X	X	X	X	-435.4917
8		X	X	X	X	X	X	X	-448.9418
9		X	X	X	X	X	X	X	-358.1662
10		X	X	X	X	X	X	X	-422.5784

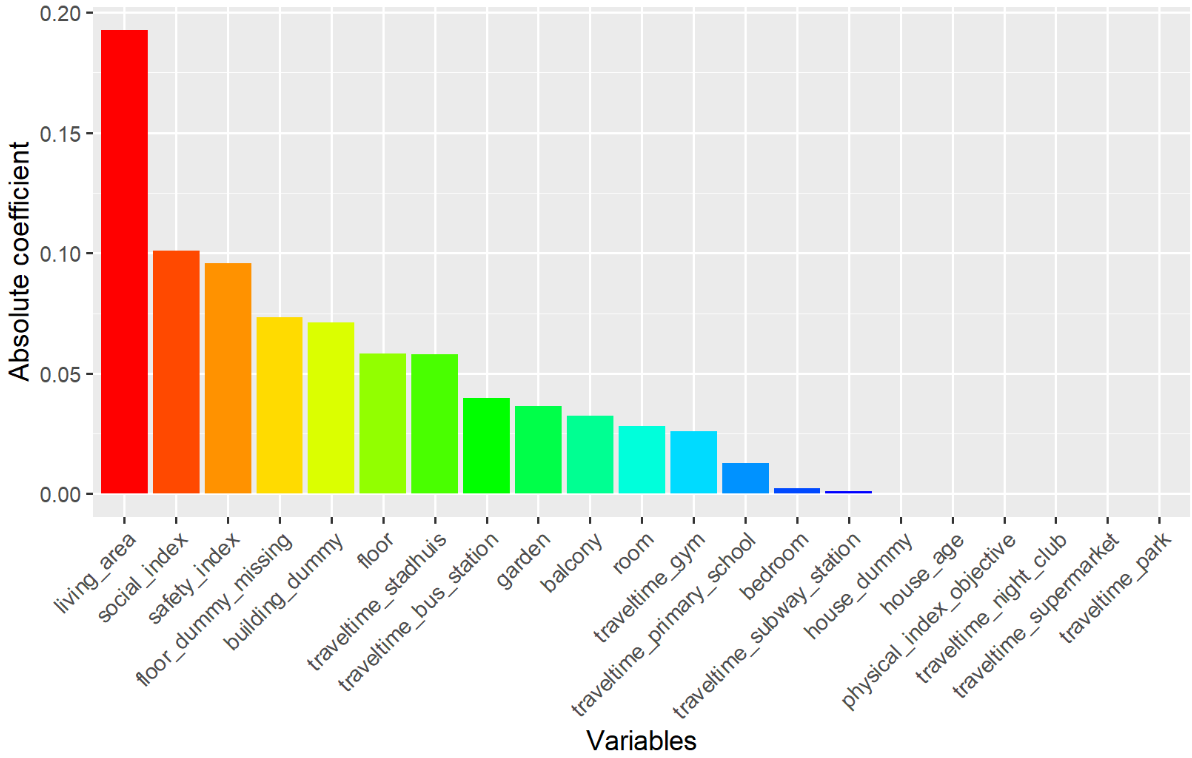
Table 8: Selected independent variables in the GWR regression for each iteration of K-fold algorithm, with corresponding AICc.

OLS with forward selection		LASSO		GWR	
Variable	Abs coef	Variable	Abs coef	Variable	Abs coef
Intercept	7.3578	Intercept	7.3644	Intercept	7.3664
living_area	0.1926	living_area	0.1872	living_area	0.2000
social_index	0.1011	floor_dummy_missing	0.0957	safety_index	0.0927
safety_index	0.0957	social_index	0.0838	social_index	0.0732
floor_dummy_missing	0.0732	safety_index	0.0803	traveltime_stadhuis	0.0674
building_dummy	0.0712	traveltime_stadhuis	0.0628	floor	0.0599
floor	0.0582	floor	0.0552	floor_dummy_missing	0.0445
traveltime_stadhuis	0.0580	building_dummy	0.0518	garden	0.0400
traveltime_bus_station	0.0397	traveltime_bus_station	0.0382	traveltime_bus_station	0.0385
garden	0.0364	balcony	0.0351	traveltime_gym	0.0342
balcony	0.0324	garden	0.0316	physical_index_objective	0.0251
room	0.0280	traveltime_gym	0.0265	room	0.0173
traveltime_gym	0.0260	room	0.0204	house_dummy	0.0148
traveltime_primary_school	0.0126	traveltime_primary_school	0.0147	traveltime_night_club	0.0142
bedroom	0.0024	house_dummy	0.0127	balcony	0.0141
traveltime_subway_station	0.0011	bedroom	0.0095	traveltime_supermarket	0.0108
house_dummy	0.0000	house_age	0.0078	bedroom	0.0104
house_age	0.0000	physical_index_objective	0.0072	traveltime_park	0.0054
physical_index_objective	0.0000	traveltime_supermarket	0.0037	traveltime_subway_station	0.0048
traveltime_night_club	0.0000	traveltime_park	0.0024	building_dummy	0.0000
traveltime_supermarket	0.0000	traveltime_subway_station	0.0019	house_age	0.0000
traveltime_park	0.0000	traveltime_night_club	0.0010	traveltime_primary_school	0.0000

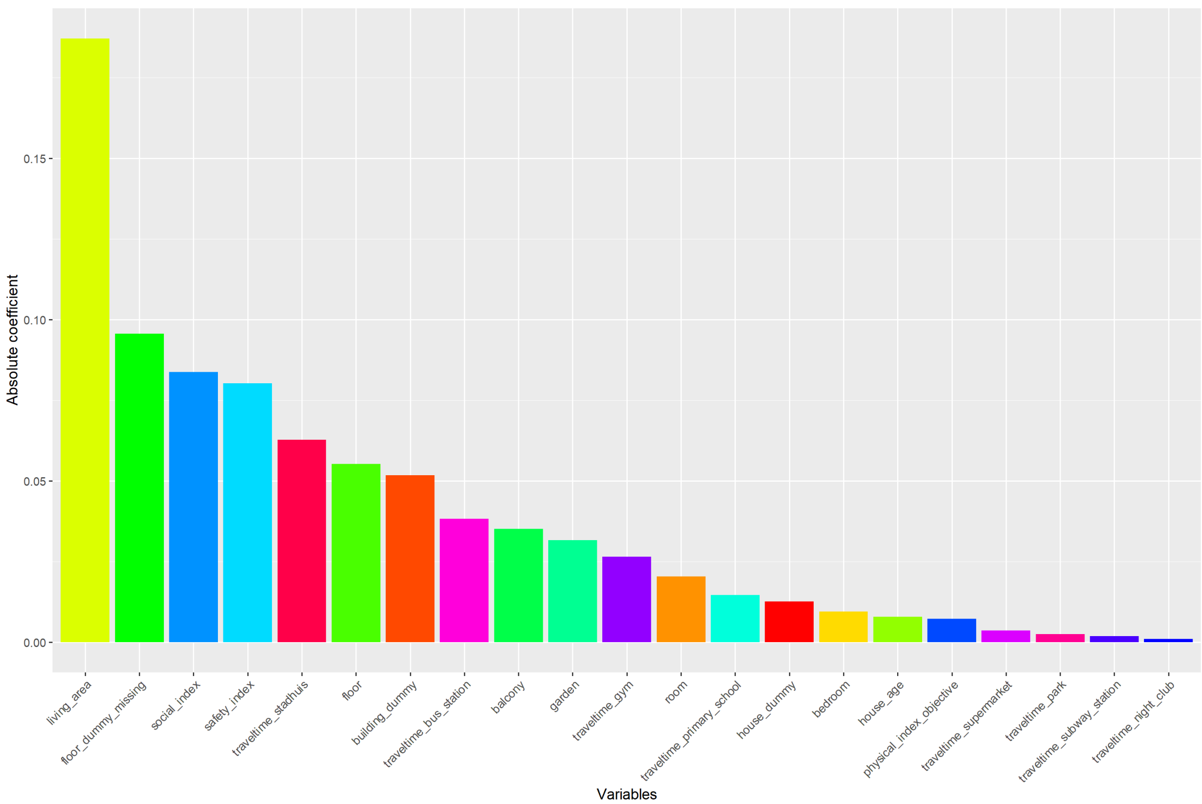
Table 7: Ranked variables of OLS with forward selection, LASSO and GWR by absolute coefficient size.

Abbreviated	Full name
Cen	Centrum
Del	Delfshaven
Ove	Overschie
Noo	Noord
H-S	Hillegersberg-Schiebroek
Kr-C	Kralingen-Crooswijk
Fey	Feijenoord
IJs	IJsselmonde
Pr. A	Prins Alexander
Chr	Charlois

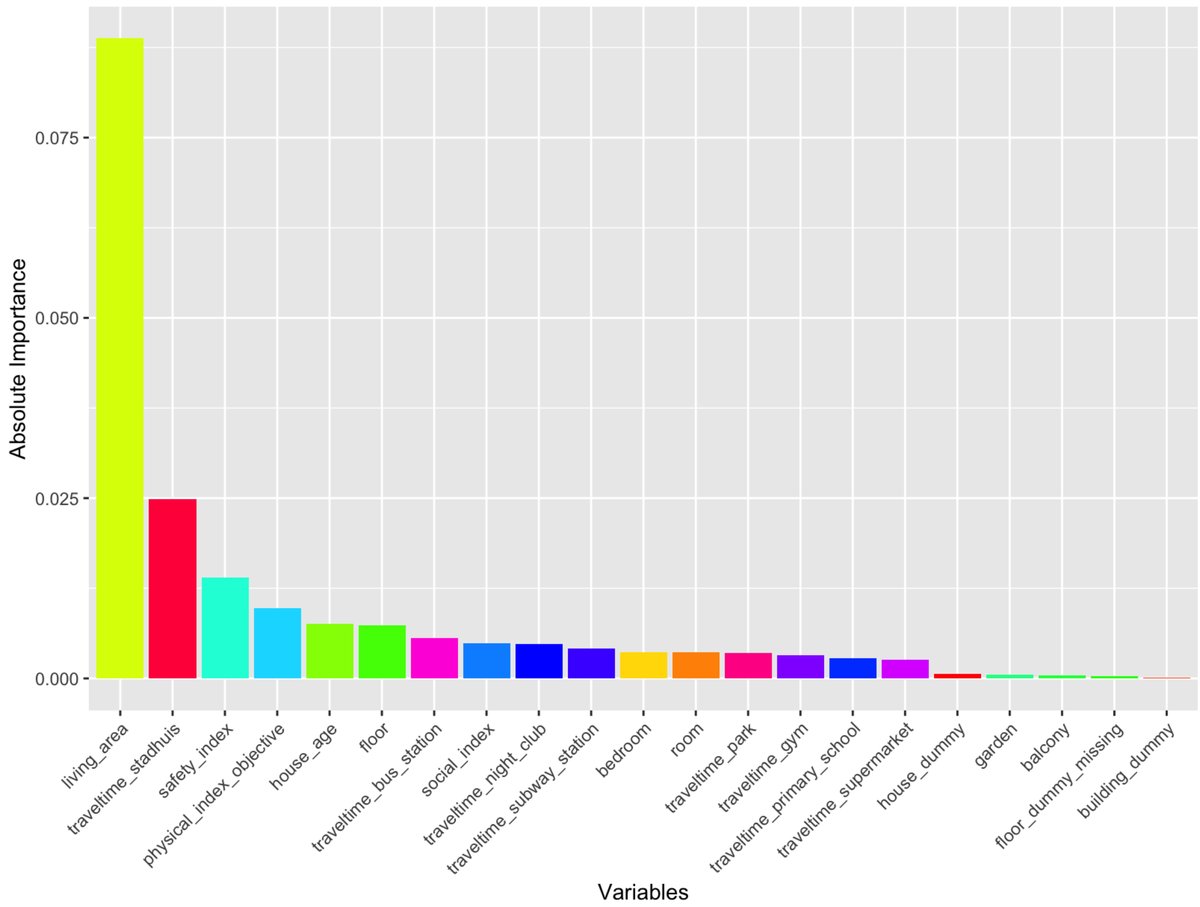
Table 9: Abbreviation of ward names.



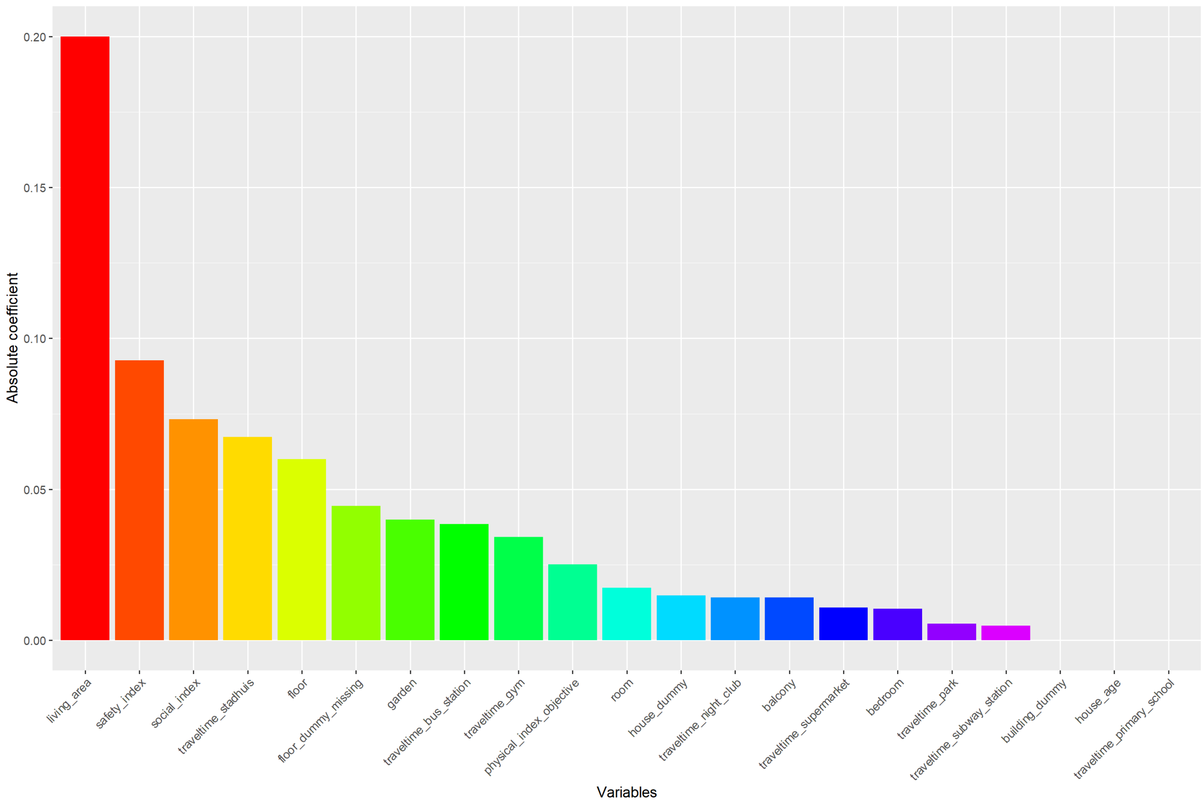
OLS (forward selection)



LASSO


Absolute Importance
0.000
0.025
0.050
0.075
living_area
traveltime_stadhuis
safety_index
physical_index_objective
house_age
floor
traveltime_bus_station
social_index
traveltime_night_club
traveltime_subway_station
bedroom
room
traveltime_park
traveltime_gym
traveltime_primary_school
traveltime_supermarket
house_dummy
garden
balcony
floor_dummy_missing
building_dummy
Variables

GWRF



GWR

Figure 18: Variable importance for the three linear regression models.

Percentage error plots GWR and GWRF

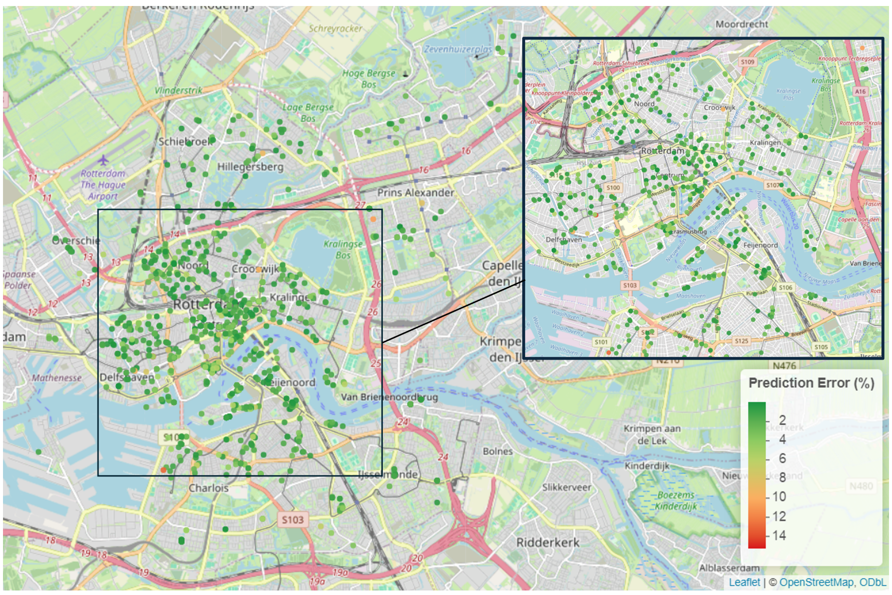


Figure 19: Percentage error for each observation
for the GWR.

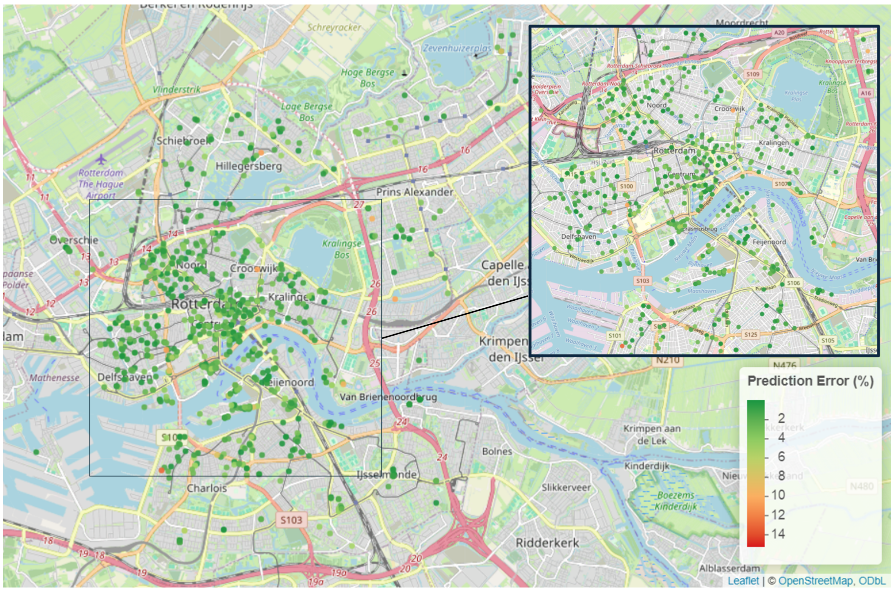


Figure 20: Percentage error for each observation
for the GWRF.

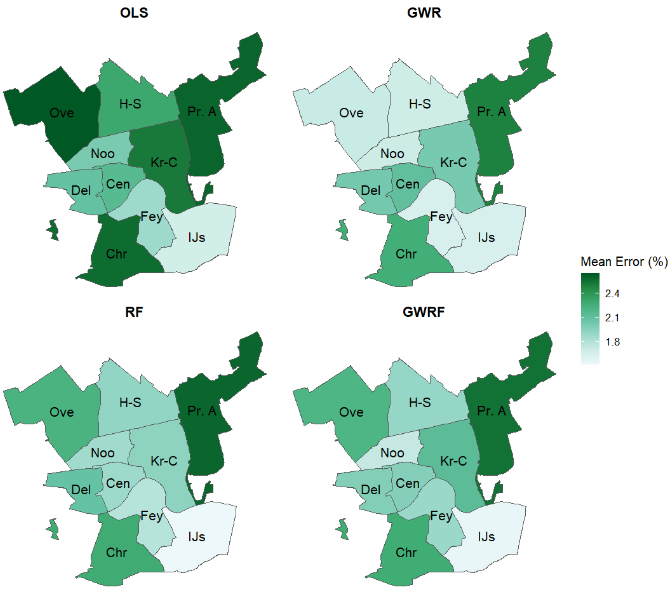


Figure 21: Percentage error per ward, for the four models. For the abbreviations, see Table 9.

Coefficients per ward

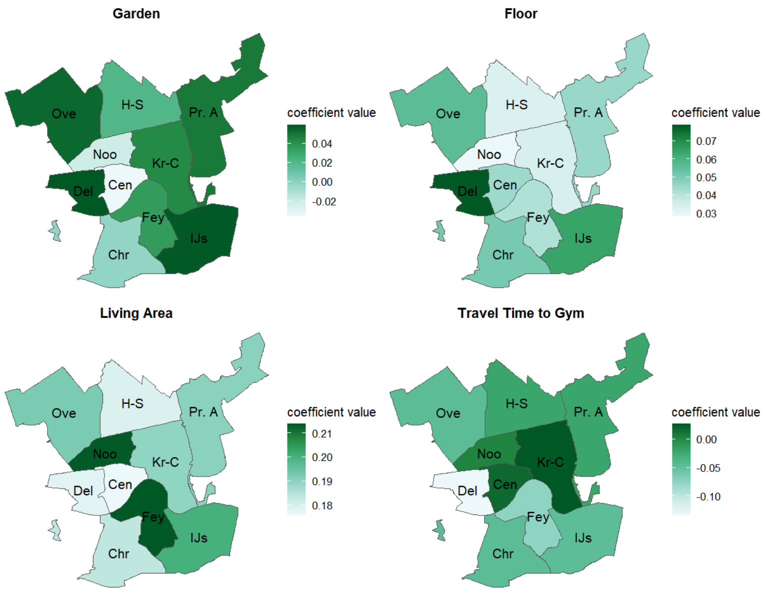


Figure 22: Average coefficients for various variables for each ward

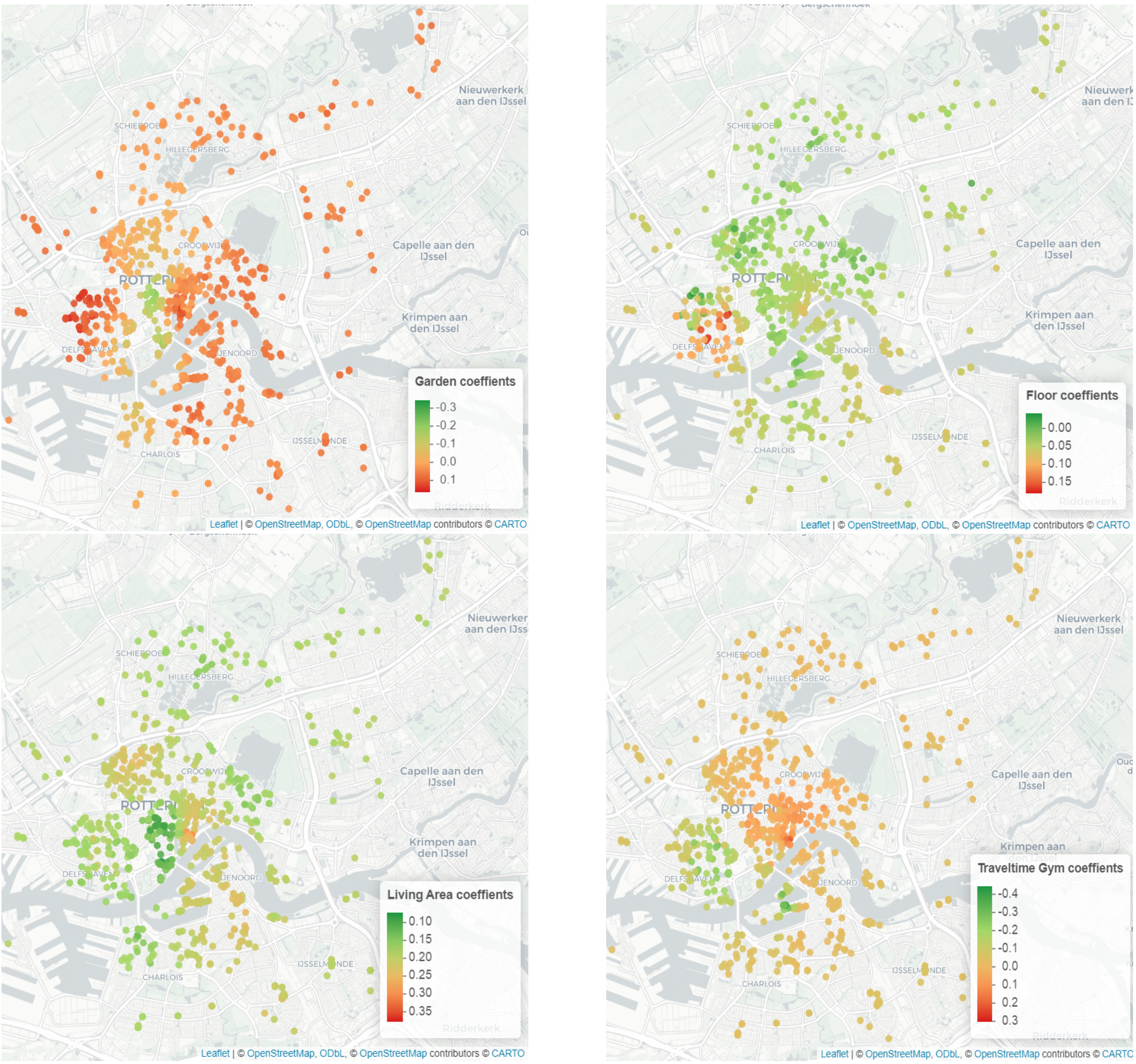


Figure 23: Coefficient value of some variables per observation for the GWR.

References

Abidoye, R. B. and Chan, A. P. (2018). Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal*, 24(1):71–83.

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical analysis*, 20(1):1–17.

Apley, D. W. and Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4):1059–1086.

Atkinson, A. C., Riani, M., and Corbellini, A. (2021). The box–cox transformation: Review and extensions.

Basile, R. and Mínguez, R. (2018). Advances in spatial econometrics: Parametric vs. semiparametric spatial autoregressive models. *The Economy as a Complex Spatial System: Macro, Meso and Micro Perspectives*, pages 81–106.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4):281–298.

Cao, Y., Li, H., and Yang, Y. (2022). Combining random forest and multicollinearity modeling for index tracking. *Communications in Statistics-Simulation and Computation*, pages 1–12.

Chau, K. W. and Chin, T. (2003). A critical review of literature on the hedonic price model. *International Journal for Housing Science and its applications*, 27(2):145–165.

Credit, K. (2022). Spatial models or random forest? evaluating the use of spatially explicit machine learning methods to predict employment density around new transit stations in los angeles. *Geographical Analysis*, 54(1):58–83.

Freeman III, A. M. (1979). The hedonic price approach to measuring demand for neighborhood characteristics. *The economics of neighborhood*, pages 191–217.

Georganos, S. and Kalogirou, S. (2022). A forest of forests: a spatially weighted and computationally efficient formulation of geographical random forests. *ISPRS International Journal of Geo-Information*, 11(9):471.

Getis, A. (2007). Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, 37(4):491–496.

Gilbert, A. and Chakraborty, J. (2011). Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in florida. *Social Science Research*, 40(1):273–286.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., and Harris, P. (2013). Gwmodel: an r package for exploring spatial heterogeneity using geographically weighted models. *arXiv preprint arXiv:1306.0413*.

Griliches, Z. (1971). Hedonic price indexes for automobiles: an econometric analysis of quality change. In *Price indexes and quality change: Studies in new methods of measurement*, pages 55–87. Harvard University Press.

Hong, J., Choi, H., and Kim, W.-s. (2020). A house price valuation based on the random forest approach: the mass appraisal of residential property in south korea. *International Journal of Strategic Property Management*, 24(3):140–152.

Kim, J., Yoon, S., Yang, E., and Thapa, B. (2020). Valuing recreational beaches: A spatial hedonic pricing approach. *Coastal Management*, 48(2):118–141.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy*, 74(2):132–157.

Lee, Y., Pennington-Gray, L., and Kim, J. (2019). Does location matter? exploring the spatial patterns of food safety in a tourism destination. *Tourism Management*, 71:18–33.

Luo, Y., Yan, J., McClure, S. C., and Li, F. (2022). Socioeconomic and environmental factors of poverty in china using geographically weighted random forest regression model. *Environmental Science and Pollution Research*, pages 1–13.

Neloy, A. A., Haque, H. S., and Ul Islam, M. M. (2019). Ensemble learning based rental apartment price prediction model by categorical features factoring. In *Proceedings of the 2019 11th International conference on machine learning and computing*, pages 350–356.

Owusu-Ansah, A. (2011). A review of hedonic pricing models in housing research. *Journal of International Real Estate and Construction Studies*, 1(1):19.

Peterson, S. and Flanagan, A. (2009). Neural network hedonic pricing models in mass real estate appraisal. *Journal of real estate research*, 31(2):147–164.

Potrawa, T. and Tetereva, A. (2022). How much is the view from the window worth? machine learning-driven hedonic pricing model of the real estate market. *Journal of Business Research*, 144:50–65.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1):34–55.

Slater, T. (2011). Gentrification of the city. *The New Blackwell companion to the city*, pages 571–585.

Suryowati, K., Ranggo, M. O., Bekti, R. D., Sutanta, E., and Riswanto, E. (2021). Geographically weighted regression using fixed and adaptive gaussian kernel weighting for maternal mortality rate analysis. In *2021 3rd International Conference on Electronics Representation and Algorithm (ICERA)*, pages 115–120.

Wen, H.-Z., Sheng-hua, J., and Xiao-yu, G. (2005). Hedonic price analysis of urban housing: An empirical research on hangzhou, china. *Journal of Zhejiang University-Science A*, 6(8):907–914.

Wheeler, D. C. (2021). Geographically weighted regression. In *Handbook of regional science*, pages 1895–1921. Springer.

Wu, D., Zhang, Y., and Xiang, Q. (2024). Geographically weighted random forests for macro-level crash frequency prediction. *Accident Analysis & Prevention*, 194:107370.

Xin, S. and Khalid, K. (2018). Modelling house price using ridge regression and lasso regression. *International Journal of Engineering & Technology*, 7:498.

Yang, W. (2014). *An extension of geographically weighted regression with flexible bandwidths.* PhD thesis, University of St Andrews.

Zhao, Q. and Hastie, T. (2021). Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, 39(1):272–281.