Received: 17 March 2020 | Revised: 23 October 2020 | Accepted: 3 December 2020

DOI: 10.1111/rsp3.12382

ORIGINAL ARTICLE

House price valuation of environmental amenities: An application of GIS-derived data

Liv Osland[1] | **John Östh**[2,3] | **Viggo Nordvik**[4]

[1]Western Norway University of Applied Sciences, Bergen, Norway

[2]Uppsala University, Uppsala, Sweden

[3]The Open University, Heerlen, the Netherlands

[4]OsloMet – Oslo Metropolitan University, Oslo, Norway

Correspondence
John Östh, Uppsala Universitet, Social and economic geography, Box 513, Uppsala 751 05, Sweden.
Email: john.osth@kultgeog.uu.se

Abstract

Hedonic house price models are frequently used to improve our understanding of local housing markets. In recent years, rich registers containing details about home-qualities and neighbourhood characteristics have successfully been coupled with spatial qualities such as job-accessibility or distances to transport. Additional data sources provided by Open data communities, NGOs, data created by governmental agencies on regional national and international level has the potential of being very useful for analysing housing prices. However, the recent methodological advances in GIS and spatial analysis have not been extensively applied. We expand the hedonic price modelling toolbox with geo-coded free data on environmental amenities. We specifically include local measures describing the view-shed, and more varied specifications of access or dominance of green and blue amenities, in addition to urban public-type service and sport facilities. The GIS-derived data is used to study how the variables should be specified and to study their ability to improve even well specified hedonic price models. To our knowledge, this paper is the first to combine all the listed environmental properties in a hedonic model, and at the same time controlling for a large number of other important local neighbourhood characteristics.

KEYWORDS
GIS, hedonic price method, Oslo

JEL CLASSIFICATION
R32; C32

1 | INTRODUCTION

There is an increased awareness that environmental amenities could be important for the wellbeing of city residents, both leisure-wise, health-wise and for aesthetic satisfaction. They may also be important for firms for instance within the tourist industry. Recent trends of urbanization, densification and expansion of cities lay pressure on a range of environmental amenities. At the same time, there is an increased awareness that one should preserve, for example, green areas in cities and metropolitan areas. There is, hence, a trade-off between densification of urban areas on the one hand, and preservation on the other. In order to ensure sustainable protection in the long run, or to study the cost and benefits of preserving these amenities, it is necessary to increase our knowledge about their value in the public urban landscape. For instance, what is the value of a house located closer to a park or other open areas, and what is the value of residing closer to rivers in urban areas? Most environmental or natural amenities are public or quasi-public local goods and do not directly have an observable market price. This feature makes the valuation far from straight forward. There exists no explicit market for these goods, and the prices or market values have to be estimated.

One of the most important valuation approaches goes via housing markets and the hedonic house price model (Kuminoff, Parmeter, & Pope, 2010). This method is a revealed preference method, based on the assumption of heterogeneous housing and theories of capitalization. The basic idea is that housing prices are expected to be higher in areas with a greater share of *inter alia* attractive environmental amenities, and lower in areas characterized by environmental dis-amenity, all else equal. In contrast to markets for housing, the market for amenities is implicit. By implicit we mean that no direct transaction or monetary exchange take place for the amenities as such (Oxley, 2004). However, via the total market price for housing, and by using statistical methods one may estimate the implicit prices for the relevant amenities. We may disentangle which attributes are perceived, valued or capitalized into housing prices. Obviously, this information is useful for firms that value houses, public authorities, tax appraisers, banks and households. It may lay the ground for studying distributional issues for dwellings located at different places, and even across cohorts. In the event of conflicts of interest, for example, between neighbours regarding obstructions of views, estimation of losses and gains could be calculated.

The aim of this paper is three-fold. We demonstrate types of information on environmental amenities that can be obtained from open sources, and show that the impact that environmental amenities have on a home with a specific location can be 'decomposed' into distance to an environmental asset and measures of spatial domination and vistas. Finally, we estimate the impact the different dimensions of the amenities have on home prices within a frame of a transparent and well specified hedonic price model.

The given data could be linked to any building by the matching of geocoded data, or co-ordinates. Availability of data has improved recently, and is increasingly being used in neighbouring fields. Recent methodological advances in GIS and spatial analysis makes it possible to test a greater range of spatial economic approaches and theories and to use more sophisticated measures of chosen amenities, compared to, for example, an approach based on using dummy variables and subjective statements about view amenities (see for instance, Li, Zhang, Li, Ricard, Meng, & Zhang, 2015). In contrast to surveys which frequently base their result on fewer observations (Baranzini & Schaerer, 2011) using GIS allows for using large samples of observations. We, therefore, aim at studying the applicability of GIS-based open source methods in relation to a heterogeneous good, such as housing. An important advantage of these data is that they in general are not confined to tabled counts on an administrative level such as a municipality or census tracts, but geocoded and therefore useful for integration in disaggregate modelling. GIS data is commonly provided in the same format regardless of geography[1] and is often made available freely by Open data communities, NGOs, or by governmental agencies on regional national and international levels. In addition, recent methodological advances in GIS and spatial analysis have not extensively been employed in the development of hedonic price models.

When it comes to the empirical specification of the hedonic price function, there exists very little theoretical basis for choosing specific types (e.g., Butler, 1982). Previous research show that the prediction of overall housing prices is robust to specification error of the estimated function (Green Richard & Malpezzi, 2003). However, if the aim is to study individual coefficients, and the impact of characteristics external to the house itself, it could be more important to include a relative comprehensive number of variables, given that the external neighbourhood variables, more than the traditional structural housing characteristics, may correlate with included or more importantly, excluded variables.

An important strength of our analysis is that we start with will a well-specified and very comprehensive empirical hedonic price model both when it comes to the number of explanatory variables and when it comes to the number of observations. So, our starting point is the results found in Nordvik, Liv, Thorsen, and Thorsen (2019). This paper has exploited a rich availability of structural variables, accessibility, and neighbourhood characteristics, such as measures of neighbourhood income, segregation and diversity according to immigrant background. The mentioned paper also contains an extensive treatment of the spatial correlation in house prices and uses a conditional autoregressive spatial model accounting for the possibility that the intercept may change continuously in the geography. Other approaches and discussions for treating spatial dependence and spatial heterogeneity are found in Anselin and Lozano-Gracia (2009); Anselin (2013); Crespo and Grêt-Regamey (2013); Dubé and Legros (2014). In this paper we do not focus on particularly on issues related to spatial correlation in order to keep the models transparent.

When choosing where to reside, environmental amenities are important. Nordvik et al. (2019) does not contain any variables related to environmental amenities, which is the focus here. In contrast to what is common in most previous literature, this study include data on environmental amenities collected using GIS technology. This is one important reason why we could allow for a large number of observations, since the collection of data is highly simplified. See the overview of the literature found in, for example, Bourassa, Hoesli, and Sun (2004); and Brander and Koetse (2011). Moreover, the precision of the environmental variables has probably improved, which contribute to a reduction in potential biases because of measurement errors.

Hence, we explore the role of space and geography with specific attention to six different types of amenity variables: *ocean view*, *natural reserves*, *parks*, *lakes*, *retail and recreational facilities or arenas*. So, our definition of environmental amenities contains variables that are inherently linked to nature and green or blue areas. However, our definition is also broader than that. In line with recent research, we also include more urban *public-type of service facilities* in our definition (Schaeffer & Dissart, 2018). Hence, we define environmental amenities as perceived *local place-based* attributes to be included in consumers' utility functions and in the production functions of suppliers. According to Schaeffer and Dissart (2018), this definition is typical for the economic literature.

There are various indicators or ways of including the mentioned six amenity-variables into the price function. The specifications include the relatively commonly used measure of *distance* to each amenity that makes use of the cartesian distance between each property and the nearest amenity in each category,[2] and the more uncommon methods including *measures of spatial domination* of the variables in the area surrounding each building (see for instance Kong, Yin, & Nakagoshi, 2007), and *vista measures* regarding the oceanic view from each building (see also

[1]Geography markup language is standard for meta-data rescription and data treatment—more information available at: http://www.ogc.org/standards/gml

[2]Cartesian distance is used as a proxy for distance to each kind of amenity since network-distance, cost or time distances cannot be estimated due to lack of data on mobility behaviour on disaggregate level.

Hamilton & Morgan, 2010). Very few studies, with the exception of for instance Cavailhès et al. (2009), have been combining all three types of methods, namely, distance-, domination- and vista-methods in price modelling. Two recent studies introduce information of views in a similar way as we do in a hedonic price model, but do not utilize information on socio-demographics (Fu, Jia, Zhang, Li, & Zhang, 2019; Ye, Xie, Fang, Jiang, & Wang, 2019). However, to our knowledge, no study has analysed the combined effects of both a wide range of GIS-derived amenity variables with sociodemographic composition of neighbourhood, location specific attributes such as accessibility and property specific attributes. It is hence, model fit, and explanatory power of the different indicators which is important, in addition to individual implicit prices and their relative importance.

Endogeneity or selection issues can rarely be ruled out in empirical analyses of housing markets. One reason for this is the basic feature of spatial fixity. In contrast to most other goods, people move to their dwelling, and housing is generally not transported to people. In this way, measures of neighbourhood characteristics like the proportion of low-income families or certain ethnic minorities may be the result of the market process which in turn may generate a certain price level. So reverse causality could be a problem when studying the impact of these neighbourhood characteristics, and was dealt with in Nordvik et al. (2019), by using time-lagged specification of variables. Arguably, reverse causality is not as relevant for the studied variables in this paper. Natural amenities do not in general exist in certain places because overall prices are high or low. We therefore do not expect reverse causality to be a problem for most of the new introduced variables.

However, households are not allocated randomly into neighbourhoods. Therefore, the estimated prices may not reflect average values, and this self-sorting of households may potentially cause biases. Places with attractive parks, views, etc., could be reflecting preferences realizable for those with higher income, rather than the willingness to pay for average households. The hedonic price function may not provide information about the preferences for low income groups, which may have a high taste for view, and who may only afford to live in areas without access to environmental amenities. However, this problem is accounted for, as far as possible, via a highly comprehensive model specification, where we account for a large number of demographic and socio-economic neighbourhood characteristics, physical structure-related characteristics of housing, in addition to variables measuring access to workplaces and the city centre.

Finally, if there are large changes in the area, it is usually not be reasonable to assume that the estimated hedonic price function is *stable over time* (see e.g., Freeman, 2003), and one may frequently have to estimate different functions before and after the change representing two different equilibria, rather than movements along a stable hedonic house price function. In the case of marginal changes in the amenities, affecting a restricted area and a limited number of people, and the market being in equilibrium, the estimated implicit prices may be interpreted as marginal willingness to pay (Freeman, 2003). The studied environmental variables rarely change significantly over at least shorter time periods (a few years), although public or private investment projects like new transport infrastructure, densification, or larger construction projects may dramatically change the supply of the mentioned environmental amenities. For this reason, the period we are studying is relatively short (2009–2012), and we assume that there are no major changes which may cause the hedonic price function to shift during this period.

The structure of the paper is as follows: In the next section we present relevant findings from the literature. Section 3 presents the study area, the data and how the data has been derived. Section 4 presents the hedonic model to be estimated and the empirical results. Finally, we conclude in Section 5.

2 | VALUATION OF ENVIRONMENTAL AMENITIES IN HEDONIC HOUSE PRICES STUDIES: AN OVERVIEW OF THE LITERATURE

The theoretical framework for studying the capitalization of amenities on housing prices are given by Rosen (1974), Lancaster (1966) and Tiebout (1956). Even before Rosen (1974), there is a history of studying whether environmental amenities or disamenities capitalize into housing prices (see for instance, Ridker & Henning, 1967). Given the existence of a large number of papers, we aim at referring to existing meta-analysis for the relevant amenities. But before we study related literature for each type of amenity, we start off with some general perspectives from the previous literature.

2.1 | Environmental amenities, some general aspects

There exist a large number of environmental attributes, which could be appealing for households. A comprehensive and general presentation of related research is found in Schaeffer and Dissart (2018). Their focus is on definitions, categorizations of the amenities and a presentation of which individual indicators or measures are included in the empirical research papers. They show that there exist a diverse number of definitions, with two types of convergences. First, most papers seem to agree about the materiality and spatial fixity of environmental amenities (Schaeffer & Dissart, 2018). Second, their impacts are *local* and subjective in nature, and also mainly important for people in contrast to firms.

Moreover, Schaeffer and Dissart (2018) identify 16 classes of environmental amenities. Our paper includes amenities, which could belong to six of these classes, depending on how they are mutually defined. Note also that some of the mentioned categories relates to climate, ecology and agriculture which have little or effect on a local urban housing market. There are also amenities that are multidimensional (e.g., index-based), and some represent subjective statements rather than more objective measures. These categories are not equally relevant for our purpose, given that we mainly measure use and/or aesthetic values for individual house owners and apply data from one single and unified geographical *urban* market, in contrast to rural areas. One important category that we do not have information about is *negative or positive externalities*. See Melissa and Kiel (2001) for a review of the impact of selected negative externalities on housing prices. In some areas, negative externalities such as traffic noise could be important, in other areas, positive externalities such as well-kept gardens could be more dominating. If the excluded externalities correlate with the studied amenities, it may bias estimation result. Given that the externalities could be both positive and negative, is difficult to predict the direction of any bias. A less comprehensive approach than ours, could possibly deal with this issue in more detail.

Panduro and Veie (2013) propose a classification of what they call green space. Their list contains eight classes: Lake, park, nature, common area for houses, common area for apartments, sports fields, agriculture fields and green buffer zones. In their review of the empirical literature, they point to varying results for the same class of green space. The main explanations are that definitions and methods of measurement differ. Moreover, most of the literature focus on only a few classes of green space, and study their impact by modelling *either proximity to* or *size* of green space. In their paper, Panduro and Veie (2013) perform an empirical study from Aalborg, Denmark. As expected, they find significant heterogeneity among the *different* categories of green space. The economic price impact of parks and lakes are significantly positive and high. Sports fields and nature areas have very limited values (nature areas are larger spaces of green, with tree covers and lakes, smaller gravel roads and paths). The differences between nature areas and parks is the better maintenance of parks (Panduro & Veie, 2013).

2.2 | Studies of individual amenities

For decades environmental amenities and dis-amenities of *urban blue or visible water*, *have* been studied empirically in relation to housing prices. Still Chen, Xun, and Junyi (2019, p. 2) seem to be the first to perform a meta-analysis of these results. Based on research from various sciences, they claim that urban rivers, streams and riverine environs are “the most influential natural components and defining features of many cities.” This is also to some extent confirmed in the meta-analysis of Schaeffer and Dissart (2018) in the sense that they find that water-related indicators are present in a large proportion of the reviewed papers.

Chen et al. (2019) identify two main research lines. The first relates to the willingness to pay for the *recreational use and aesthetic value* of proximity to urban rivers or by the *visible* aspects of rivers. The second line of research study whether *degraded rivers* negatively influence property prices. It is the first approach, which is relevant for this paper.

The results from the meta-analysis suggest that proximity to river, water quality and river view, respectively, all have a significant positive impact on housing prices. The last-mentioned variable, *river view*, has the highest economic impact. *Proximity to river* has the lowest economic impact of the three. However, it is this variable which has achieved most attention in the research literature (likely due to the relative ease of estimation and use of river proximity in modelling). The results do not appear to depend on whether one account for spatial effects, and they are also the same in a range of different geographical contexts. However, the results vary over time, and the economic impact is higher after 2000. Increased scarcity and reduction in negative externalities related to rivers are the mentioned reasons for this. Finally, Chen et al. (2019) study the impact on estimated prices of various contextual factors. The results show that the willingness to pay for urban river-related amenities is higher for affluent households. Therefore, they are characterized as normal goods. In contrast, there is a negative relation between population density and the willingness to pay for rivers as an amenity.

Brand and Koetse (2011) performs a meta-analysis of valuation studies the published the last 30 years focusing on *urban open space*. They define urban open space as urban parks, forests, green spaces such as golf courses and sports arenas, undeveloped and agricultural land at the urban periphery (Brand and Koetse, 2011). The cited hedonic studies estimate the effect on housing prices of increased distance to open space. Alternative indicator such as the size of open space or the overall percentage of open space is not studied. The reason for this choice is that the most studies include distance measures rather than measures of size.

Moreover, most studies focus on the impact of urban *parks*. The majority of the studies are recent (after 2000), and the vast majority are from the United States. Results from the meta-analysis show that increased distance to *open space* clearly has a negative impact on housing prices. The further the distance to open space the smaller is the price effect of being located closer to open space. Assuming average values for relevant variables, the increase in housing prices is 0.1% when moving 10 metres closer to open space. Distance to *urban parks* has similar impact on housing prices as the distance to open space. One important result is that there are no statistically significant differences of estimates in high income areas versus low income areas. The value of open space increases the more densely populated an area is. Specification of the hedonic function has an impact on the result. The estimated prices are higher for linear models in comparison to non-linear models. Finally, there are significant regional differences between the estimated results.

Recreational facilities and arenas that we are studying is a very diverse amenity category, which contain sports facilities, designated recreational areas, outdoor baths and cross-country tracks. So, to some extent the above-mentioned results documented for *open spaces* is also relevant here. There are a number of studies focusing on professional sports arenas (see e.g., Feng and Humphreys, 2012). We have, however, not been able to find any published papers studying the type of recreational amenities that we include here. The only exception is Panduro and Veie (2013). They find that sports fields (which are not professional) have limited impact on housing prices.

When it comes to *natural reserves*, not many results have been found in the literature. One exception is the Finnish study by Tyrväinen (1997). She studies the value of urban forests in a hedonic house price model, using data from North Karelia, in eastern Finland. She included three different variables in this respect: The first is distance to the nearest wooded recreation area used for jogging, skiing or walking. The second is distance to smaller forested area, those located nearest a given house. The third is based on density or the relative amount of forested areas. The distance to the nearest forested area was not significant. However, the two other measures were. Closeness to the recreational area and the amount of forest area had a positive impact on housing prices. As mentioned above, Panduro and Veie (2013) find a more limited, yet significant impact at the 10% level of natural reserves.

Obviously, the impact of a *view* can be both negative and positive, depending on type of view (street view, mountain view, industry, etc.). According to Bourassa et al. (2004) the majority of the studies focusing on view, use data from the United States (as was the case for open space). Most studies focus on water view, albeit a number of studies do not specify type of view because of problems of finding clear definitions, and view is most frequently measured by a single dummy variable. In a more recent paper studying housing market in Minnesota, Sander and Polasky (2009) stress that there exists no general agreement regarding the size of the impact of views on housing prices. This is also because there is a large variation in methodological approaches. So, results vary at different places at various points in time. However, it is an open question whether studies using the GIS methodology could improve matters. The GIS methodology could imply more standardized approaches. *Inter alia*, it could be easier to distinguish between variations in view quality, and one may not have to rely on subjective visual inspections.

To our knowledge, there exist no meta-analysis focusing on the value of *ocean view* in particular. However, some literature summaries exist. Kwong, Yiu, Wong, and Lai (2003) focus on studies where view is the prime focus. They find that various types of view contribute *significantly* to explain housing prices in all the papers they cite. Bourassa et al. (2004) however, highlight that some studies find *insignificant* results, and suggest that this finding could be due to measurement errors, weak definitions of view and small samples.

Among the mentioned papers in Kwong et al. (2003), only Benson, Hansen, Schwartz, and Smersh (1998) focus on ocean view. Benson et al. (1998) distinguish between the qualitative variation of ocean view, which according to their literature is a novelty. Moreover, previous results have shown modest impact on housing prices. Misspecification could be one explanation, given that differences in view quality has regularly not been included.

In line with our approach, Benson et al. (1998) use GIS technology and a number of dummy variables. They distinguish between five different qualities of ocean view, where the 5th category is no ocean view. Quality of ocean view contributes positive and significantly to explain housing prices. The impact ranges from 8.2% (poor partial ocean view) to 58.9% (unhindered ocean view). The highest value of 58.9% is only relevant for houses which is also located very close to the water front (0.1 miles). The difference between full partial view and superior partial view is not very high, and commands an increase in price of 29% and 31%, respectively. The house price impact reduces substantially with increased distance to the ocean. Bourassa et al. (2004) find that this is a general result for attractive views.

The empirical analysis performed in Bourassa et al. (2004) is from Auckland, New Zealand, and it distinguishes between variations in view quality. They find that a wide view of waterside on average increases house values by as much as 59%. The impact reduces with distance from the coast. The magnitude of the impact also rests on the attractiveness and beauty of neighbourhoods.

Whether closeness to *retail areas* has a positive or negative impact on housing prices have been studied for a long time. This is similar to what is studied in the access space trade off theory of Alonso. However, distance to a central business district (CBD) is a global variable, affecting the overall area. The impact of retail services on the other hand is local. The results may be positive or negative, see Osland and Thorsen (2013) for an overview some related literature. However, this may be due to confounding factors such as negative externalities. We are not aware of any study which focus on retail specified as a dominance variable.

3 | STUDY AREA AND DATA

Oslo is the capital of Norway, located close to a fjord and with historical ties to the inlands. Together with its neighbouring municipalities, it forms a natural labour and housing market – often called Greater Oslo. The exact delimitation of the Greater Oslo region is not obvious. In this paper, we define Greater Oslo as Oslo municipality and the nine closest surrounding municipalities. In total, this region contains about one fifth of the Norwegian population (1.1 million inhabitants). Over the last couple of decades, Greater Oslo experienced a rather strong growth both in terms of population, housing prices and in terms of economic activity. See Nordvik et al. (2019) for further details. Moreover, the Norwegian housing sector is market based, and two out of three households are homeowners, according to Statistics Norway. The market has shown a remarkable stability over time, in the sense that there are no major institutional changes or alterations in market regulations in the study period. For more details on the region, see, for example, Nordvik et al. (2019).

Our data consists of combined time series and cross-sectional data on actual sales prices and characteristics of transacted units including location, dating from 2009 to 2012. We base our estimations on 99,852 housing transactions, and 73% of the transactions are from the Oslo municipality. Our data-source is Finn.no, which is a net-based sale portal. Around 70% of all sales in the study area is included. Information on many control variables stem from a range of different public registers, such as the public population register, tax register and registers of the Central Welfare Agency of Norway. These registers have complete population coverage. In contrast to the distances of the environmental amenities which are the focus here, distances to CBD and labour market zones (control variables) are measured by the shortest travelling time by car. We account for speed limits, but not for the within-zone travelling time. See also Nordvik et al. (2019) for more details regarding the data, including summary statistics.

3.1 | Creation of new data using GIS

We make use of GIS to create data estimating the view from each property towards the Oslo fjord, for the estimation of distances to specific land-cover amenities and finally we estimate how dominant, in terms of area share, the selected land-covers are in the bespoke neighbourhood surrounding each sold property. In the subsequent sections, methods and settings are discussed at some length.

3.1.1 | Vista metrics and view

Oslo is situated in the inner part of the Oslo fjord, which is a deep Atlantic Ocean fjord that ends where the hills and mountains raise. This sets the scene for some spectacular views over the ocean. In order to estimate the view over the Oslo fjord, we employ a 3D GIS-technology (view shed analysis) in which a high-resolution open-data elevation raster is used to create a digital elevation model of the urban area. To quantify the view, we position four vista buoys in the ocean and measure the count of buoys that can be seen from any part of the greater Oslo region taking into account the 3D-landscape that facilitates or restricts visibility. The result is a high-resolution raster image that holds vista buoy counts for each m^2. By using a spatial join method (*extract values to point*, in ArcGIS), we associate the co-ordinates for each residence sold in the greater Oslo region to a vista buoy count value that is used as a proxy for ocean view. It should be noted that the rendered vista-results have no information about the factual view (i.e., determined the orientation of windows, trees blocking view, or how view is changing between floors in a multi-story building) but is a proxy for unrestricted ground-level view. In Figure 1, the result is illustrated as a map showing the locations with ocean view. The four vista buoys are shown as dark-blue hexagons in the water.

This method generates three variables, denoting how many vista buoys are visible. The majority of the observations have no view (69% of the observations), around 9% of the observations has fjord view in the sense that one or two vista buoys are visible, respectively, 7,8% has the view of three, and 4% has the view of four vista buoys. Hence, it is a very specific type of view that we study. Comparisons across studies are hence, not necessarily valid. Our study bears some resemblance to a study using Google Street View to demonstrate variations in views of urban greenery (Li et al., 2015).

3.1.2 | Distance to, and domination of land-cover in the neighbourhood

In cases where geo-coded information is used to relate house pricing to amenities it is commonplace to measure distance to amenities from each property but more complex spatial patterns are usually omitted (Kong et al., 2007;

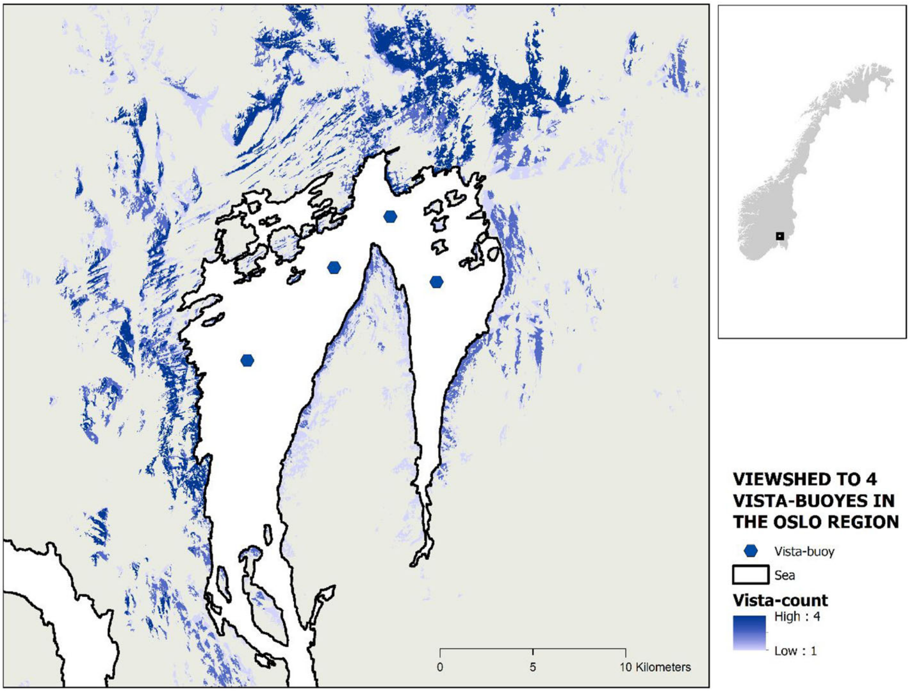

FIGURE 1 Vista buoy distribution and blue colour indicating the number of buoys that can be seen from different locations

Melichar & Kaprová, 2013). However, using only distance measures will not generate a measure of how dominating the amenity is in the neighbourhood. We propose a model in which both distance (Cartesian distance since we lack information about modes of transport and mobility patterns among the buyers[3]) to the nearest observation of an amenity is used, as well as a new measure of how dominating the amenity is in the bespoke neighbourhood. Domination is here understood as the land-share of the surrounding from each sales-location that belongs to any of the amenities, where the domination is expressed as a ratio of area occupied of the amenity. This approach is similar to that described by Kong et al. (2007). The size of the area surrounding each property is set to a circle with a radius of 500 m.[4]

Using geocoded data derived during spring 2019 from OSM (Open Street Map) we selected five land-cover types (amenities) for further analysis:[5] *Natural reserves*, *parks*, *sweet water* (lakes and larger streams), *retail areas*, and *recreational facilities*. Natural reserves are areas, which the municipalities have decided to protect from development and interference. All listed natural reserves and parks were included in the analysis. However, the sweet water was restricted to waterbodies of at least 50 m^2 in size (smaller streams and ponds were consequently deselected).[6] All the amenities have a specified code.

The retail areas were constituted by areas that were dominated by retail activities which means that isolated shops are excluded. Finally, the recreational areas were mainly constituted by sports-facilities and outdoor-activity

[3]A relevant study to cite in this respect is Heyman and Sommervoll (2019) who introduce relative location such as walking distance to metro and parks.

[4]Using a radius of 500 m we are close to *classic* specifications of the walkable neighbourhood which in many ways influenced urban planning, see for instance Perry (1929/1998) in which a planned neighbourhood is 160 acres.

[5]Each amenity has a specific code in the data source.

[6]The Oslo region has a very large number of smaller water bodies (as much of the Nordic region) as a result of climate, bedrock and soil (morraine hummock). Most of the excluded smaller water bodies are spread throughout forests or along constructions and have no recreational value.

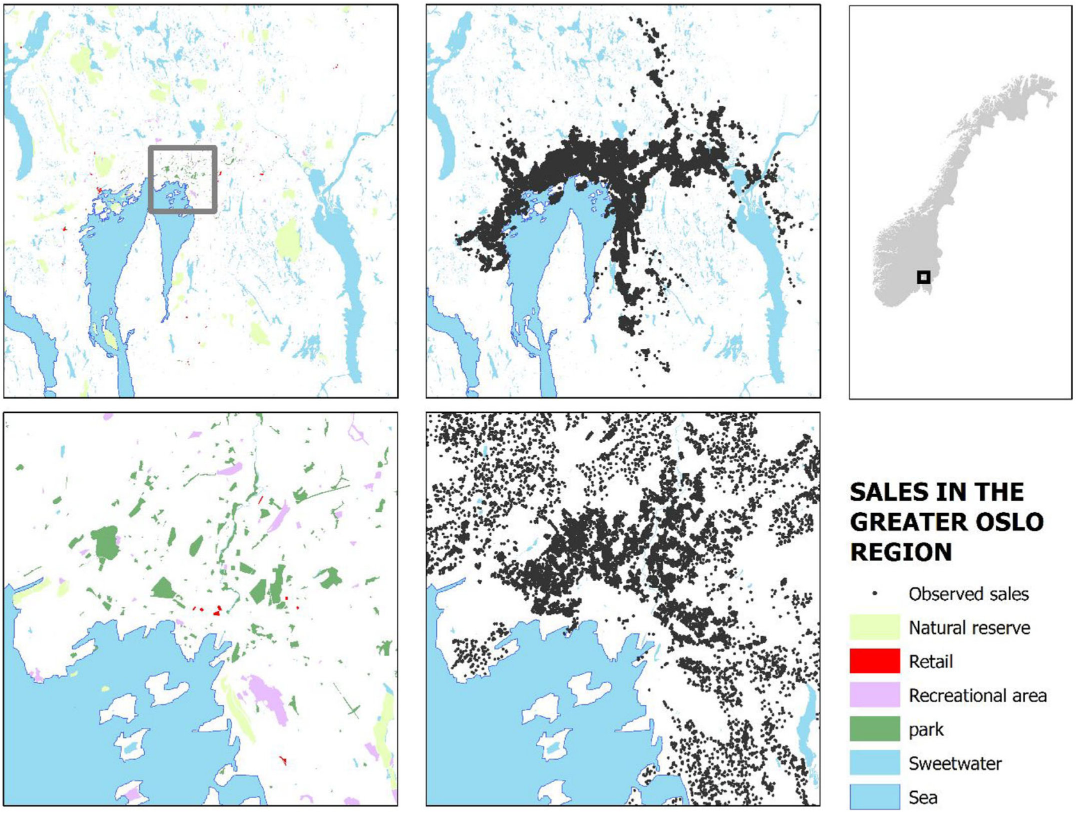

FIGURE 2 Data from open street maps showing the spatial distribution of measured amenities (left) and observed sales in the region (right). Top maps show the entire region, bottom row shows central parts of Oslo. Top right map indicates the location of the study-area in Norway

arenas such as walking/running lanes, public baths, and public squares, etc. The data-selection was conducted between February and April in 2019 which is later in time compared to the sales data used in analysis. In addition, being an open-community map-service OSM have been known to have quality variations in certain parts. In order to reduce potential errors created from the use of erroneous map-material we have manually compared OSM to Street-view imagery and vice versa in random locations around the urban area. Most of the land-cover features are over-time stable but minor alterations to the OSM cover of retail and recreation may deviate over time creating a mismatch between amenities and sales. However, we expect any effect to be of a too small magnitude to create inference problems (Figure 2).

Each of the above described polygon shape-files are converted to raster data with a spatial resolution of 100 m^2 (10m × 10 m). Using focal statistics (ArcGIS is used for the GIS analyses) we measure the share of the surrounding area of a radius of 500 m from each 100 m^2 unit in the greater Oslo region that is containing the land-cover features described above.[7] In addition, we measure the Euclidian distance between all sold residential properties and the five land-cover types. This means that we have information about the proximity to the selected amenities as well as a measure of the domination of the amenity in the bespoke neighbourhood.

[7]Since the data input contains 100 m^2 units the 500 m radius circle is not perfectly round but rugged. The maximum count of units within any 500 m radius is therefore 7,845 and not ~7,853.98. The share is calculated as the observed count of units occupied by the specific land-cover divided by 7,845, and finally multiplied by a 100 to create a percentage-value.

The domination variables represent the percentage of land-share from each sales-location that belongs to any of the amenities. The surrounding area is defined as 500 m around each sale. Distances are measured as Cartesian distance (metres). NR represents natural reserves.

Table 1 shows that the average distance to natural reserves, and its standard deviation, is larger than for all the other amenities. The domination variables in the table should be interpreted as percentages. We see that in the overall study area it is the park-amenity which is most dominant, as measured by the average domination of the amenity in each pixel (having a radius of 500 m around each sale). The retail variable is measured as the least dominant variable.

4 | THE HEDONIC MODEL TO BE ESTIMATED AND ESTIMATION RESULTS

4.1 | The empirical model specification

The starting point for the empirical analysis is results from Nordvik et al. (2019). We refer to this model as the base model (BM). BM is specified in equation 1. We include a full list of variables in the Appendix.

$$logP_{it} = \beta_0 + \beta_1(\log\mathbf{DWELLING})_{it} + \beta_2(\log\mathbf{LOCATION})_{it} + \beta_3(\log\mathbf{SOCIOECDEM})_{it} + \beta_4(\mathbf{ENVIRONAMENITIES})_{it} + \sum_{t=10}^{12}\beta_1(\mathbf{YEARDUM})_{it} + \in_{it}. \tag{1}$$

Our dependent variable is the natural logarithm of housing prices. Including the year dummies, we distinguish between five types of explanatory variables. The *first* relates to characteristics of the house itself, such as the size of the house, age and type of dwelling. The *second* relates to characteristics of the location, which includes access to workplaces and anisotropic specification of distance to the city centre. These variables are *global* in nature, which implies that they are specified to connect the overall study area. We have also included a range of *local*

TABLE 1 Summary statistics for some of specifications of the studied variables (n = 99,852)

Variable	Mean	Std. dev.	Min	Max
View				
OceanView	0.684163	1.168391	0	4
Distance				
ParkDist	498.4518	742.1293	0	10401.34
LakeDist	1243.623	852.7225	0	6310.704
NRDist	2520.363	1283.476	0	10563.47
RETDist	1567.802	1251.675	0	12217.1
REC Dist	594.6465	598.368	0	9064.156
Domination				
Park domination	5.143665	6.699789	0	46.06756
Lake domination	0.814401	2.899294	0	52.07138
NR domination	0.376739	2.746821	0	75.3601
Retail domination	0.356897	2.130895	0	35.90822
RecDomination	1.33275	2.860265	0	41.86106

characteristics of neighbourhoods. These characteristics are either physical, such as the proportion of various house types in each neighbourhood. A number of sociodemographic characteristics is also accounted for, and BM is particularly well specified in this respect. We include the proportion of various age groups, the proportion of children and households moving out of neighbourhoods, education level of neighbours, measures of income and the proportion receiving social and national security benefits. Finally, segregation of neighbourhoods has been studied, and we control for proportions of immigrants, measured by country of origin for specific groups of immigrants. The proportion of native Norwegians is important, as is diversity with respect to country background of neighbours. Both these clearly contribute significantly positive to explain variation in housing prices. Time or year dummies are included to capture the pure change in housing prices over the studied period. As mentioned in Section 3, a neighbourhood is defined by clusters of census tracts (1,440 in all), making up 182 neighbourhoods. We include a dummy variable for each neighbourhood (except one) in each estimated model to account for the time invariant idiosyncratic characteristics. These variables do not appear in Equation 1.

4.2 | Modelling procedures, results and discussion

The paper has a specific attention to six different types of amenity variables (environ amenities in Equation (1): *ocean view*, *natural reserves*, *parks*, *lakes*, *retail and recreational facilities or arenas*. These variables are described in Section 3. The variables could be included in alternative ways. The variables natural reserves, park, lake, retail and recreational facilities are available as domination variables and by distance from home to each amenity. The distance variables are included by taking the natural logarithm of the variables. We test whether their squared terms also should be included. None of the squared terms of these variables is significant. Ocean view is accounted for by including four dummy variables, each of which indicates different levels of view. No view is the reference group. In one of the reported models, we have merged the two dummy variables, which represent the best view into one category. This is explained further below.

Almost all variables added separately to the BM, are statistically significant at the 5% level, and the coefficients have the expected sign. Natural reserve dominance, is one exception, and contributes slightly negative to explain housing prices. We obtain the same result for this variable in more complete model specifications. If we specify it as a distance variable, it is not statistically significant, perhaps because the average distance to natural reserves is relatively high, and may, hence, not have an impact on housing prices. There are several possible explanations for the result. First, by definition, and according to the empirical literature, the environmental amenities are local (Schaeffer & Dissart, 2018). From Table 1 we see that the average Cartesian distance is quite long. In this respect, this amenity is perhaps not characterized as local, and second, it could be more or less equally available from many residential areas in our study area. The same result of no significance is found for the recreational variable, in both types of variable specifications. One possible explanation for this, which is different from the above explanation related to natural reserve, is that there could be negative externalities related to having recreational facilities close to a house, such as parking of cars in the surrounding and litter. Among the dominance variables, park and retail are the most significant ones. Included as distance variables, distance to retail and lakes are the most significant variables in relation to housing prices.

In the second step, we augment the model with one and one variable in turn; starting with the variables, which have the highest white-robust *t*-values. The results are presented in Table 1, and show that parks should be included by its dominance variable. The user value and/or aesthetic value of larger parks measured by dominance is what capitalized into housing prices, rather than living closer to parks in general.

In contrast to park, lake and retail should be specified as distance variables. Including the variables as distances rather than as domination variables, increases the explanatory power of the hedonic model. Hence, it is access to retail, which is most important for housing prices, rather than the size of the retail centres. Note that we do not consider multicollinearity to be a major problem in this paper, even for the distance variables. The reason is the large number of observations, and the fact that the results are very stable even when including and excluding variables. Moreover, the Pearson correlation coefficients between distance to park and distance to retail and lakes is 0.317 and 0.573, respectively, which is not specifically high.

All the dummy variables of view contribute significantly to explain housing prices. We have tested if there are significant differences between coefficients of different types of view. The *p*-values of a standard robust *t*-test, testing the null hypothesis of equality between the variables (1) and TwoOceanDum; and (3) and FourOceanDum, respectively, were 0.06 and 0.2. These results show that the mentioned pair of variables could be merged. However, measures of the predictive power of the model (SRMSE) point to the inclusion of all four dummy variables separately. When including four variables SRMSE is 0.2415, when including three dummy variables SRMSE gets the value 0.2841, no matter which of the mentioned pair variables we combine. Hence, we have included all the four view variables in M2Best, which is the preferred model specifications.

A comparison of M2Best with the BM, show that there is a clear and significant increase in explanatory power when including the four types of environmental variables. The value of the likelihood ratio test statistic is 546, which by far exceeds the critical value of a chi-square distribution having 7 degrees of freedom. An important result is that all the estimated coefficients estimated in M2Best which is also included in the BM are within a 95% confidence region of the BM. So, the results are highly consistent and stable across model specifications and indicates that the BM is a valid starting point.

In Figure 3 we illustrate the variation in housing prices of a standard house as distance to lake vary. A standard house is mainly defined by including average values of control variables, zero values for the other environmental amenities and sold in 2012. The estimated parameter related both to lake and to retail is −0.017. This implies that the predicted housing prices are around NOK 804,000 (13.7%) higher in proximity to lakes, in comparison to the furthest distance from the lake (around 6.3 km). The decrease in prices are steepest for distances until 2 km from the house. These results are comparable to the variable distance to retail. A house located in areas where park dominance is highest, gives a price increase of NOK 500,000 (7.3%) in comparison to a house located in areas where the park dominance is lowest. Regarding the view, the price increase is around NOK 260,000 (3.8%) for a standard house that has a view, according to category 3. The basis is a standard house that does not have a view at all. In all these results confirm that the environmental variables included have, a significant impact on housing prices both statistically and economically. In total they contribute to improve the predictions of our model (Table 2).

The GIS-derived part of the regression results is interesting in the sense that it put more focus on the importance of (certain) amenities around properties, which in turn can have implications for urban planning

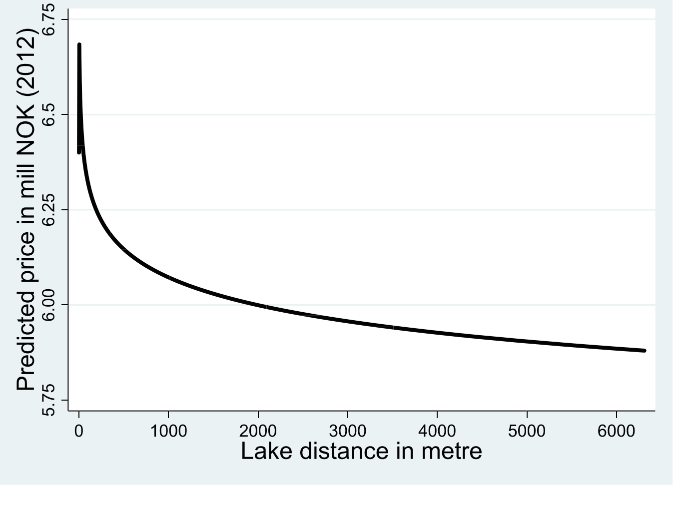

FIGURE 3 Predicted variation in the price of a standard house, assuming change in proximity to lake

TABLE 2 Regression results of alternative model specifications

Variables	(1) BM	(2) M2(all)	(3) M2(best)	(4) M2(3View)
View				
OneOceanDum		0.005*(1.903)	0.005*(1.905)	
TwoOceanDum		0.011***(4.152)	0.011***(4.152)	
ThreeOceanDum		0.037***(12.288)	0.037***(12.3)	0.037***(12.208)
FourOceanDum		0.031***(7.211)	0.031***(7.238)	0.031***(7.166)
OneTwOceanDum				0.008***(3.816)
Distance				
lnLakeDist		−0.018***(−11.487)	−0.017***(−11.464)	−0.017***(−11.404)
lnRETDist		−0.017***(−10.042)	−0.017***(−10.061)	−0.017***(−10.144)
lnNRDist		0.001(0.239)		
lnRECDist		−0.001(−0.551)		
Domination				
ParkDomination		(8.588)	(8.704)	(8.697)
Observations	99,852	99,852	99,852	99,852
Adj. R-squared	0.848	0.849	0.849	0.849
Log likelihood	22,323	22,596	22,596	22,594
SRMSE	0.2849	0.2842	0.2415	0.2841
Moran's I	0.1203		0.1163	0.1163
Z (Moran's I)	191.85		186.4761	185.4761

Notes: Robust *t*-statistics in parentheses,
$^{***}p < 0.01$,
$^{**}p < 0.05$,
$^{*}p < 0.1$. SRMSE is the standardized root mean square error, which could be used to compare model variants (see Knutsen & Fotheringham, 1986). A value of 0 indicates that the model predictions are perfect. Calculating the SRMSE implies transforming the dependent variable. The transformation follows Wooldridge (2003), using an estimator of the residual variance that is unbiased. Dummy variables for each neighbourhood (182 in all minus one), are included in all model alternatives. M2(3View) contains only three different view variables. Moran's I for residuals is estimated for all sales in the greater Oslo region using an inverse distance decay function, and Cartesian distances, to deter more distant sales. Estimations were conducted in ArcMap Pro.

(Ye et al., 2019). As shown by Kong et al. (2007), more complex neighbourhood measures such as the domination of amenities (not just distance to) are important for pricing. In addition, being able to see green and blue amenities increase the value of properties. These results are in line with earlier findings by Hamilton and Morgan (2010) and Fu et al. (2019). However, this study shows that the results remain also after controlling for social and demographic factors which suggests that planners more actively could make room for urban green and amenity rich views in poorer areas.

5 | CONCLUSIONS

Our study is in no way the first study to make use of GIS-generated data to better understand pricing of houses, but it is, to our knowledge, the first to combine detailed variables describing the listed properties, the socio-demographics of the neighbourhood, the location and accessibility to the urban area, as well as several variables describing distances and domination of urban amenities including retail, park, water, etc., as well as view of ocean from any of the properties.

The results show that the new GIS-related variables from Open Source Data are highly useful when predicting house prices in the area, also after taking into account the wide range of additional variables. Environmental amenities are important in the model, both from a statistical and from an economic market point of view. Endogeneity is probably not a major a problem, given *inter alia* the stability of the results across alternative specifications of the models, and given the large number of control variables, including fixed effects dummy variables for neighbourhoods. Note, however, that we cannot completely rule out that there is some selection bias related to our results. For instance, close to areas where parks dominate, neighbourhoods may be more well-kept and dominated by wealthy households. We also find that it is park dominance, rather than distance to parks, which is of importance. Proximity to retail and lakes seems to be the most important distance variables. It is access to these variables which is important, not how dominant they are in the neighbourhood. Having a view to the ocean has a significant positive implicit price. Note, however, that view is defined very specific. So, a nice or ugly view close to homes in general is not accounted for. Recreational facilities such as public sports arenas or public baths do not contribute to explain housing prices. The same holds for the natural reserves. This last-mentioned result is in line with an agreement in the literature, that the impact of environmental amenities in relation to housing prices is usually local. On average, the distance to natural reserves is probably too long to have an impact on housing prices, and in our study area, it may be available and accessible from a large number of residential areas. The findings also suggest that planning for green and blue amenities could be used actively to increase the value of properties in areas that are poor or in other ways marginalized.

The analysis illustrate that it may be important to access information on several environmental benefits using the new data sources. First, the variables could be specified in different ways. Hence, we could test how the model specification of variables. This feature reduces the potential for misspecification in the estimated house price models. Obtaining information for larger geographical areas on larger dataset is also central. This is mainly important for valid inferences and when studying a more complete set of environmental variables. Multicollinearity is a common problem when several amenities are specified as distance variables. Using a large number of observations could be one reason why multicollinearity does not seem to be a problem in our analyses. In future research, we think more advanced spatial panel data models could be estimated, especially for estimating spatial autocorrelation (see for instance, López, Chasco, & Gallo, 2015; Chasco, Le Gallo, & López, 2018),

Finally, the new data sources may also provide useful descriptive statistics, and we may perform more comprehensive distribution analyses. As an example, we may study distribution of environmental goods in different types of neighborhoods with different socio-economic characteristics. So, a combination of the newly available open data sources with high quality register data opens for more a variety of distributional analyses in future research.

ACKNOWLEDGEMENT

This paper is based on research funded by the Norwegian Research Council, Grant 217210/H2.

ORCID

John Östh https://orcid.org/0000-0002-4536-9229

REFERENCES

Anselin, L. (2013). *Spatial econometrics: Methods and models*. Springer Science & Business Media.

Anselin, L., & Lozano-Gracia, N. (2009). Spatial hedonic models. In T. C. Mills & K. Patterson (Eds.), *Palgrave handbook of econometrics: Volume 2: Applied econometrics* (pp. 1213–1250). London: Palgrave Macmillan. https://doi.org/10.1057/9780230244405_26

Baranzini, A., & Schaerer, C. (2011). A sight for sore eyes: Assessing the value of view and land use in the housing market. *Journal of Housing Economics*, *20*(3), 191–199. https://doi.org/10.1016/j.jhe.2011.06.001

Benson, E. D., Hansen, J. L., Schwartz, A. L. Jr., & Smersh, G. T. (1998). Pricing residential amenities: The value of a view. *Journal of Real Estate Finance and Economics*, *16*(1), 55–73.

Bourassa, S. C., Hoesli, M., & Sun, J. (2004). What's in a view? *Environment and Planning A*, *36*, 1427–1450. https://doi.org/10.1068/a36103

Brander, L. M., & Koetse, M. J. (2011). The value of urban open space: Meta-analyses of contingent valuation and hedonic pricing results. *Journal of Environmental Management*, *92*(10), 2763–2773. https://doi.org/10.1016/j.jenvman.2011.06.019

Butler, R. V. (1982). The specification of hedonic indexes for urban housing. *Land Economics*, *58*(1), 96–108. https://doi.org/10.2307/3146079

Carruthers, J. I., & Mundy, B. (2006). *Environmental valuation*. Aldershot, Hampshire, England: Interregional and intraregional perspectives. Ashgate Publishing Limited.

Cavailhès, J., Brossard, T., Foltête, J. C., Hilal, M., Joly, D., Tourneux, F. P., … Wavresky, P. (2009). GIS-based hedonic pricing of landscape. *Environmental and Resource Economics*, *44*(4), 571–590. https://doi.org/10.1007/s10640-009-9302-8

Chasco, C., Le Gallo, J., & López, F. A. (2018). A scan test for spatial groupwise heteroscedasticity in cross-sectional models with an application on houses prices in Madrid. *Regional Science and Urban Economics*, *68*, 226–238. https://doi.org/10.1016/j.regsciurbeco.2017.10.015

Chen, Y. W., Xun, L., & Junyi, H. (2019). Environmental amenities of urban rivers and residential property values: A global meta-analysis. *Science of the Total Environment*, *693*. https://doi.org/10.1016/j.scitotenv.2019.133628

Crespo, R., & Grêt-Regamey, A. (2013). Local hedonic house-Price modelling for urban planners: Advantages of using local regression techniques. *Environment and Planning. B, Planning & Design*, *40*(4), 664–682. https://doi.org/10.1068/b38093

Dubé, J., & Legros, D. (2014). Spatial econometrics and the hedonic pricing model: What about the temporal dimension? *Journal of Property Research*, *31*(4), 333–359. https://doi.org/10.1080/09599916.2014.913655

Feng, X., & Humphreys, B. R. (2012). The impact of professional sports facilities on housing values: Evidence from census block group data. *City, Culture and Society*, *3*(3), 189–200. https://doi.org/10.1016/j.ccs.2012.06.017

Freeman, A. M. (2003). *The measurement of environmental and resource values theory and methods*. New York: RFF Press.

Fu, X., Jia, T., Zhang, X., Li, S., & Zhang, Y. (2019). Do street-level scene perceptions affect housing prices in Chinese megacities? An analysis using open access datasets and deep learning. *PLoS ONE*, *14*(5), e0217505. https://doi.org/10.1371/journal.pone.0217505

Green Richard, K., & Malpezzi, S. (2003). *A primer on U.S. housing markets and housing policy (Areuea monograph series).*, Washington: Urban Institute press.

Hamilton, S. E., & Morgan, A. (2010). Integrating lidar, GIS and hedonic price modeling to measure amenity values in urban beach residential property markets. *Computers, Environment and Urban Systems*, *34*(2), 133–141.

Heyman, V., & Sommervoll, D. E. (2019). House prices and relative location. *Cities*, *95*, 102373. https://doi.org/10.1016/j.cities.2019.06.004

Knutsen, D. C., & Fotheringham, A. (1986). Matrix comparison, goodness-of-fit, and spatial interaction modeling. *International Regional Science Review*, *10*(2), 127–147. https://doi.org/10.1177/016001768601000203

Kong, F., Yin, H., & Nakagoshi, N. (2007). Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: A case study in Jinan City, China. *Landscape and Urban Planning*, *79*(3–4), 240–252. https://doi.org/10.1016/j.landurbplan.2006.02.013

Kuminoff, N. V., Parmeter, C. F., & Pope, J. C. (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? *Journal of Environmental Economics and Management*, *60*(3), 145–160. https://doi.org/10.1016/j.jeem.2010.06.001

Kwong, C. W., Yiu, C. Y., Wong, S. K., & Lai, L. W. C. (2003). Hedonic price modelling of environmental attributes: A review of the literature and a Hong Kong case study. In L. W. C. Lai & F. T. Lorne (Eds.), *Understanding and Implementing Sustainable Development* (pp. 87–110). Hauppauge: Nova Publishers.

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

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google street view and a modified green view index. *Urban Forestry & Urban Greening*, *14*(3), 675–685. https://doi.org/10.1016/j.ufug.2015.06.006

López, F. A., Chasco, C., & Gallo, J. L. (2015). Exploring scan methods to test spatial structure with an application to housing prices in Madrid. *Papers in Regional Science*, *94*(2), 317–346.

Melichar, J., & Kaprová, K. (2013). Revealing preferences of Prague's homebuyers toward greenery amenities: The empirical evidence of distance–size effect. *Landscape and Urban Planning*, *109*(1), 56–66. https://doi.org/10.1016/j.landurbplan.2012.09.003

Melissa, B., & Kiel, K. (2001). A survey of house Price hedonic studies of the impact of environmental externalities. *Journal of Real Estate Literature: 2001*, *9*(2), 117–144.

Nordvik, V., Liv, O., Thorsen, I., & Thorsen, I. S. (2019). Capitalization of neighbourhood diversity and segregation. *Environment and Planning A, Economy and Space*, *51*(8), 1775–1799. https://doi.org/10.1177/0308518X19861108

Osland, L., & Thorsen, I. (2008). Effects on housing prices of urban attraction and labor market accessibility. *Environment and Planning A*, *40*, 2490–2509. https://doi.org/10.1068/a39305

Osland, S., & Thorsen, I. (2013). Spatial impacts, local labour market characteristics and housing prices. *Urban Studies*, *50*(10), 2063–2083.

Oxley, M. (2004). *Economics, planning and housing*. Palgrave, London: Macmillan International Higher Education.

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

Perry, C. (1929/1998). *The neighbourhood unit (1929) reprinted Routledge/Thoemmes* (Vol. 1998). London.

Ridker, R. G., & Henning, J. A. (1967). The determinants of residential property values with special reference to air pollution. *The Review of Economics and Statistics*, *49*(2), 246–257. https://doi.org/10.2307/1928231

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55. https://doi.org/10.1086/260169

Sander, H. A., & Polasky, S. (2009). The value of views and open space: Estimates from a hedonic pricing model for Ramsey County, Minnesota, USA. *Land Use Policy*, *26*(3), 837–845. https://doi.org/10.1016/j.landusepol.2008.10.009

Schaeffer, Y., & Dissart, J. C. (2018). Natural and environmental amenities: A review of definitions, measures and issues. *Ecological Economics*, *146*, 475–496. https://doi.org/10.1016/j.ecolecon.2017.12.001

Tiebout, C. (1956). A pure theory of local expenditures. *Journal of Political Economy*, *65*, 416–424.

Tyrväinen, L. (1997). The amenity value of the urban forest: An application of the hedonic pricing method. *Landscape and Urban Planning*, *37*(3–4), 211–222. https://doi.org/10.1016/S0169-2046(97)80005-9

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review*, *93*(2), 133–138.

Ye, Y., Xie, H., Fang, J., Jiang, H., & Wang, D. (2019). Daily accessed street greenery and housing price: Measuring economic performance of human-scale streetscapes via new urban data. *Sustainability*, *11*(6), 1741. https://doi.org/10.3390/su11061741

How to cite this article: Osland L, Östh J, Nordvik V. House price valuation of environmental amenities: An application of GIS-derived data. *Reg Sci Policy Pract*. 2021;1–21. https://doi.org/10.1111/rsp3.12382

APPENDIX

TABLE A1 Regression results including all variables except dummy variables for neighbourhood

Variables	(1) BM	(2) M2(all)	(3) M2Best	(4) M2(3View)
LIVAREA (DET)	0.601*** (63.140)	0.600*** (63.160)	0.600*** (63.179)	0.600*** (63.184)
LIVAREA (SMA)	0.642*** (86.770)	0.641*** (86.731)	0.641*** (86.719)	0.641*** (86.720)
LIVAREA (APA)	0.771*** (350.638)	0.770*** (351.684)	0.770*** (351.745)	0.770*** (351.638)
CONYEAR45	0.001 (0.214)	0.006 (1.353)	0.006 (1.348)	0.006 (1.389)
CONYEAR 57	−0.027*** (−9.902)	−0.024*** (−9.115)	−0.024*** (−9.127)	−0.024*** (−9.112)
CONYEAR78	−0.024*** (−6.734)	−0.022*** (−6.295)	−0.022*** (−6.295)	−0.022*** (−6.263)
CONYEAR89	−0.001 (−0.373)	−0.000 (−0.026)	−0.000 (−0.034)	−0.000 (−0.013)
CONYEAR90	0.047*** (12.982)	0.048*** (13.397)	0.048*** (13.393)	0.049*** (13.420)
CONYEAR0_6	0.126*** (39.965)	0.129*** (40.805)	0.129*** (40.807)	0.129*** (40.791)
CONYEARREC	0.128*** (38.802)	0.129*** (39.179)	0.129*** (39.180)	0.129*** (39.183)
COOPDUM	−0.000 (−0.184)	−0.001 (−0.517)	−0.001 (−0.497)	−0.001 (−0.414)
SHAREDUM	−0.006** (−2.275)	−0.010*** (−3.768)	−0.010*** (−3.780)	−0.010*** (−3.767)
YEARDUM10	0.083*** (46.182)	0.083*** (46.174)	0.083*** (46.174)	0.083*** (46.172)
YEARDUM11	0.182*** (105.829)	0.182*** (105.933)	0.182*** (105.924)	0.182*** (105.906)
YEARDUM12	0.251*** (136.806)	0.250*** (136.833)	0.250*** (136.807)	0.250*** (136.784)
DETACHDUM	0.918*** (18.632)	0.922*** (18.744)	0.922*** (18.748)	0.922*** (18.759)
TERRDUM	0.648*** (17.911)	0.649*** (17.955)	0.650*** (17.958)	0.649*** (17.951)
SEMI-DETDUM	0.657*** (17.556)	0.659*** (17.610)	0.659*** (17.612)	0.659*** (17.605)
TIMECBDW	−0.024** (−2.557)	−0.019** (−2.026)	−0.019** (−2.003)	−0.019** (−1.961)
TIMECBDN	−0.076*** (−5.315)	−0.073*** (−5.054)	−0.072*** (−5.030)	−0.072*** (−5.021)
TIMECBDS	−0.013 (−1.147)	−0.013 (−1.155)	−0.013 (−1.151)	−0.012 (−1.100)
ACCESSIBILITY	−0.021*** (−3.258)	−0.022*** (−3.537)	−0.022*** (−3.518)	−0.022*** (−3.446)
INC(0711)MED	0.099*** (7.054)	0.110*** (7.925)	0.110*** (7.930)	0.110*** (7.910)
INC(0711)SD	0.072*** (24.902)	0.070*** (24.130)	0.070*** (24.223)	0.070*** (24.226)
SOCSECBEN	0.001 (1.080)	0.000 (0.711)	0.000 (0.697)	0.000 (0.700)
AGE1019	−0.000 (−0.119)	−0.003 (−0.702)	−0.003 (−0.709)	−0.003 (−0.708)
AGE2030	−0.016*** (−2.775)	−0.015*** (−2.801)	−0.015*** (−2.799)	−0.015*** (−2.776)
AGE3040	0.046*** (7.040)	0.044*** (6.879)	0.044*** (6.894)	0.044*** (6.895)
AGE4050	−0.029*** (−4.527)	−0.024*** (−3.663)	−0.024*** (−3.663)	−0.024*** (−3.648)
AGE5060	0.028*** (4.556)	0.028*** (4.799)	0.028*** (4.803)	0.028*** (4.783)
AGE 6070	0.012*** (3.781)	0.013*** (4.260)	0.013*** (4.244)	0.013*** (4.291)
AGE70+	0.009*** (5.164)	0.007*** (4.162)	0.007*** (4.176)	0.007*** (4.207)
RDIage	−0.024*** (−7.909)	−0.024*** (−7.816)	−0.024*** (−7.828)	−0.024*** (−7.853)
HIGHEDU	−0.008*** (−3.408)	−0.009*** (−4.033)	−0.009*** (−4.028)	−0.009*** (−4.048)
LOWEDU	−0.004 (−0.685)	−0.002 (−0.321)	−0.002 (−0.314)	−0.002 (−0.336)
5OUTMOVE	0.004*** (4.214)	0.003*** (3.568)	0.003*** (3.554)	0.003*** (3.563)
OUTMOVE	0.015*** (3.041)	0.015*** (3.190)	0.015*** (3.195)	0.015*** (3.154)
INCL120	−0.140*** (−9.443)	−0.151*** (−10.198)	−0.151*** (−10.260)	−0.151*** (−10.223)

TABLE A1 (Continued)

Variables	(1) BM	(2) M2(all)	(3) M2Best	(4) M2(3View)
CASHCARE	−0.002** (−2.499)	−0.001* (−1.709)	−0.001* (−1.720)	−0.001* (−1.741)
NATSECBEN	−0.022*** (−7.864)	−0.022*** (−7.603)	−0.021*** (−7.597)	−0.021*** (−7.585)
DETPROP	−0.001*** (−3.063)	−0.001*** (−2.892)	−0.001*** (−2.918)	−0.001*** (−2.838)
APAPROP	0.011*** (5.571)	0.011*** (5.781)	0.011*** (5.782)	0.011*** (5.754)
BLOCKPROP	−0.001* (−1.697)	−0.001 (−1.540)	−0.001 (−1.531)	−0.001 (−1.546)
SMALLPROP	0.001 (0.680)	0.000 (0.275)	0.000 (0.279)	0.000 (0.280)
NORWEGIAN	0.070*** (8.498)	0.078*** (9.494)	0.078*** (9.507)	0.078*** (9.487)
DIV01	0.028*** (7.488)	0.027*** (7.425)	0.027*** (7.405)	0.028*** (7.476)
POLAND	0.000 (0.049)	0.000 (0.088)	0.000 (0.119)	0.000 (0.091)
RUSSIA	0.000 (0.204)	0.000 (0.192)	0.000 (0.186)	0.000 (0.164)
OCEANIA	0.001*** (6.054)	0.001*** (5.354)	0.001*** (5.352)	0.001*** (5.298)
SOMALIA	−0.001** (−2.237)	−0.000** (−1.966)	−0.000** (−1.993)	−0.000** (−1.982)
TURKEY	−0.002*** (−7.765)	−0.002*** (−7.358)	−0.002*** (−7.354)	−0.002*** (−7.396)
CENTRALASIA	−0.002*** (−5.589)	−0.002*** (−5.596)	−0.002*** (−5.585)	−0.002*** (−5.580)
DEVOLOPING	−0.006*** (−7.101)	−0.006*** (−6.803)	−0.006*** (−6.776)	−0.006*** (−6.804)
WESTERN	0.013*** (4.018)	0.014*** (4.288)	0.014*** (4.312)	0.014*** (4.258)
ParkDomination		0.002*** (8.588)	0.002*** (8.704)	0.002*** (8.697)
lnLakeDist		−0.018*** (−11.487)	−0.017*** (−11.464)	−0.017*** (−11.404)
lnRETDist		−0.017*** (−10.042)	−0.017*** (−10.061)	−0.017*** (−10.144)
OneOceanDum		0.005* (1.903)	0.005* (1.905)	
TwoOceanDum		0.011*** (4.152)	0.011*** (4.152)	
ThreeOceanDum		0.037*** (12.288)	0.037*** (12.300)	0.037*** (12.208)
FourOceanDum		0.031*** (7.211)	0.031*** (7.238)	0.031*** (7.166)
lnNRDist		0.001 (0.239)		
lnRECDist		−0.001 (−0.551)		
OneTwoOceanDum				0.008*** (3.816)
Constant	9.646*** (48.513)	9.801*** (49.370)	9.797*** (49.443)	9.793*** (49.453)
Observations	99,852	99,852	99,852	99,852
Adjusted R-squared	0.848	0.849	0.849	0.849
Log likelihood	22,323	22,596	22,596	22,594
SRMSE	0.2849	0.2842	0.2415	0.2841
Moran's I	0.1156***			

Notes: Robust t-statistics in parentheses.
***p < 0.01,
**p < 0.05,
*p < 0.1.

TABLE A2 Definition of control variables

Dwelling attributes	
DETACHDUM =	1 if the observation is a detached house, 0 if it is not (dummy)
TERRDUM =	1 if the observation is a terraced house, 0 if it is not (dummy)
SEMI-DETDUM =	1 if the observation is a semi-detached house, 0 if it is not (dummy)
LIVEAREA (DET) =	The living area of a detached house, measured in square meters
LIVEAREA (SMA) =	The living area of a small house, measured in square meters
LIVEAREA (APA) =	The living area of an apartment, measured in square meters
CONYEAR45 =	1 if the house was constructed in 1940–1950, 0 if not (dummy)
CONYEAR60 =	1 if the house was constructed in 1951–1970, 0 if not (dummy)
CONYEAR75 =	1 if the house was constructed in 1971–1980, 0 if not (dummy)
CONYEAR85 =	1 if the house was constructed in 1981–1990, 0 if not (dummy)
CONYEAR95 =	1 if the house was constructed in 1991–2000, 0 if not (dummy)
CONYEAR03 =	1 if the house was constructed in 2001–2006, 0 if not (dummy)
CONYEARREC =	1 if the house was constructed in 2007-, 0 if not (dummy)
COOPDUM =	1 if the building is a part of a housing cooperative, 0 if not (dummy)
SHAREDUM =	1 if the building is a condominium, 0 if not (dummy)
Yeardummies	
YEARDUM10 =	1 if the house/apartment was sold in 2010, 0 if not (dummy)
YEARDUM11 =	1 if the house/apartment was sold in 2011, 0 if not (dummy)
YEARDUM12 =	1 if the house/apartment was sold in 2012, 0 if not (dummy)
Location	
TIMECBDW =	The travelling time in minutes from the Oslo city centre, to the west
TIMECBDN =	The travelling time in minutes from the Oslo city centre, to the north
TIMECBDS =	The travelling time in minutes from the Oslo city centre, to the south
ACCESSIBILITY =	Labour market accessibility, defined by $\text{ACCESSIBILITY}_i = \sum_{k=1}^{w} D_k \exp(\beta_e d_{ik})$, where D_k represents the number of jobs (employment opportunities) in destination (zone) k (Osland & Thorsen, 2008).
Economic and sociodemographic characteristics	
AGE1019 =	The proportion of the population in a zone in the age group of 10–19
AGE2030 =	The proportion of the population in a zone in the age group of 20–30
AGE3040 =	The proportion of the population in a zone in the age group of 30–40
AGE4050 =	The proportion of the population in a zone in the age group of 40–50
AGE5060 =	The proportion of the population in a zone in the age group of 50–60
AGE6070 =	The proportion of the population in a zone in the age group of 60–70
AGE70 + =	The proportion of the population in a zone in the age group of 70+
RDIAGE =	a measure of the diversity of age groups in a zone
5OUTMOVE =	The proportion of households with children under 5 years old that has moved out of the zone
OUTMOVE =	The proportion of households that has moved out of the zone
HIGHEDU=	The proportion of the zonal population in the age group 30–49 with a bachelor's and/or a master's degree
LOWEDU =	The proportion of the zonal population in the age group 30–49 with no more than secundary education
INC(0711)MED =	The average income in the period 2007–2011 of the median male income earner in the zone

TABLE A2 (Continued)

INC(0711)SD =	The standard deviation of the average 2007–2011 incomes for male income earners in the zone
INCGT120 =	The proportion of the population in the zone with income higher than 120,000 NOK
SOCSECBEN =	The proportion of the population in the zone receiving social security benefits
NATSECBEN =	The proportion of the population in the zone receiving National Security benefits
CASHCARE =	The proportion of the households in a zone receiving cashcare for small children
DETPROP =	The proportion of detached houses in the zone
APAPROP =	The proportion of apartments in the zone
BLOCKPROP =	The proportion of blocks of flats in the zone
SMALLPROP =	The proportion of small houses in the zone
Country background	
NORWEGIAN =	The proportion of the population in the zone with background from Norway
POLAND =	The proportion of the population in the zone with background from Poland
RUSSIA =	The proportion of the population in the zone with background from Russia
OCEANIA =	The proportion of the population in the zone with background from Oceania
SOMALIA =	The proportion of the population in the zone with background from Somalia
TURKEY =	The proportion of the population in the zone with background from Turkey
CENTRALASIA =	The proportion of the population in the zone with background from Central Asia
DEVELOPING =	The proportion of the population in the zone with background from developing countries (except Somalia)
WESTERN =	The proportion of the population in the zone with background from western countries (except Norway, Poland, and Russia)
NON-WESTERN =	The proportion of the population in the zone with background from non-western countries
DIV_k =	The diversity of the population in a zone, w.r.t. country background