Measuring Residential Location Preference in Greater London

***Abstract*——**This paper establishes an index to measure residential location preference in Greater London by considering housing psychology. It consists of six factors: housing price, crime rate, accessibility of public transportation & the density of road network, green & blue cover, and average income. The feasibility of this index is tested through ordinary least square regression. Besides, its positive spatial autocorrelation is found using Moran's I statistics, and local indicators of spatial association (LISA) are applied to find the local spatial cluster. The geospatial detector explores the spatial heterogeneity of the index and proves that there are interactive effects of various factors on residential satisfaction.

***Keywords*——**residential location preference, spatial autocorrelation, spatial heterogeneity, interactive effects

1. Context

London is facing the housing crisis that accommodation supply is inadequate(Gallent et al., 2017), while there are significant increases in house prices for the last twenty years(Travers et al., 2016). It allows a young generation to afford the rising cost of housing in London rigorously. It is necessary for individuals to find high-quality residential choices, and there is a great deal of research on this topic(Jim and Chen, 2007, Guo and Peeta, 2020, Scheuer et al., 2021). Public house purchase intention is explored in (Ghazali et al., 2020, Song and Zhang, 2020, Kurniawan et al., 2020), which is helpful to have a proper understanding of residential preference. Besides, housing choice may be beneficial for the government to implement policies that cater to the mass market to solve this issue.

This paper aims to find the main factors of residential preference by considering public housing psychology and establish an index to evaluate it in each London Borough. In order to suggest to help solve accommodation issues, preference inequality needs to be found, and the influences of factors should be tested and quantified.

2. Data and Methodology

2.1 Factors selection

There are multi-source factors that have a combined effect on the housing options. Factors are selected by considering the psychology of buying and selling a house as well as data validation to reflect most of residents' preferences. Therefore, the theory of Maslow's Hierarchy of Needs(Maslow, 1958) can be utilized to explain their housing psychologies and find the leading indicators(Zavei and Jusan, 2012, Baqutayan et al., 2015). Maslow described a hierarchical relationship of different human needs in this theory that contains five levels of needs: (a) fundamental needs, (b) safety needs, (c) belongingness and love needs, (d) esteem needs and (e) self-actualization needs. These demands exist simultaneously, and one of which plays a decisive role. But the key need is constantly changing with increasing age of human, and the reason is that people always feel unsatisfied with their accomplishment. Therefore, indicators of housing choice can be collected using this approach and literature review, as shown in Table 1.

Table 1: The hierarchies of factors

Needs	Factors	Reasons
Basic needs ⬇	Housing price(Karsten, 2007, Morrow-Jones et al., 2004)	Housing prices play a crucial role in choosing a residence. Affordable house is a foundational requirement, so it should be the basic needs.
Safety needs ⬇	Crime rate(Morrow-Jones et al., 2004, Kovacs-Györi et al., 2019)	Personal safety has a negative relationship with the crime rate. Crime is a threat to security.
Belongingness and love needs ⬇	Accessibility of public transports & the density of road network(Morrow-Jones et al., 2004)	Accessibility decides people's moving area, and high accessibility can ensure a person gets in touch with their families and friends, fulfilling their belongingness and love needs.
Esteem needs ⬇	The sized area of green and blue infrastructures(Jiang et al., 2015, Kovacs-Györi et al., 2019)	Esteem needs are spiritual requirements. Greenland and blueland, like a park, are good places to relieve stress and cultivate their hobbies.
self-actualization needs	Household income(Guo and Hardin, 2014, Morrow-Jones et al., 2004)	Prosperities positively correlate to income. Self-actualization needs may be met when reaching the highest income level. Additionally, purchasing the best house is determined by income.

These data sets are explored based on the Boroughs map in Greater London, and their sources are shown in Table 2.

Table 2: Data source

Data	Time	Source
Housing sales	2020	https://data.london.gov.uk/dataset/uk-house-price-index
Housing price	2020	https://data.london.gov.uk/dataset/uk-house-price-index
Crime rate	2020	https://data.police.uk/data/
Accessibility of public transports	2015	https://data.london.gov.uk/dataset/public-transport-accessibility-levels
Density of road network	2020	https://www.gov.uk/government/statistical-data-sets/road-length-statistics-rdl
Green & blue infrastructure cover rate	2020	https://data.london.gov.uk/dataset/green-and-blue-cover
Average income	2017-2018	https://data.london.gov.uk/dataset/average-income-tax-payers-borough

2.2 Method

This work develops an approach to the housing measurement based on these multi-source data and uses non-spatial and spatial methods to test its feasibility.

A. The Measurement of Residential Location Preference

The housing preference index (HPI) is proposed to compare residents' preferences in different London Boroughs. There is a hypothesis that all the factors have a cumulative impact on housing choice, and they compose the HPI. Besides, the weight of each element needs to be set according to their contributions to the index. The hierarchy of housing psychology of the mass can be a rule to establish each weight, as shown in Table 3. It should be noted that people may not prefer to live in an area with a high housing price or a high crime rate, so these factors have negative weights.

Table 3: The Factors and weights

The hierarchy of needs	Factors	Weights
Basic needs	Housing price	30%
Safety needs	Crime rate	25%
Belongingness and love needs	Accessibility	10%
	Density of road network	10%
Esteem needs	Green & blue cover	15%
self-actualization needs	Income	10%

Therefore, the HPI can be calculated by the following equation:

$$HPI_n = \sum_{j=0}^{n} (\pm) w_{nj} x_{nj}$$

where HPI_n presents the housing preference index in the Borough n, and it is expected to have a positive relationship with housing sales; x_{nj} presents the j-th factors in the Borough n, while w_{nj} presents the weight of the j-th factors in the Borough n. Additionally, due to different units used to measure variables, the data need to be nondimensionalized by using the min-max normalization and takes the following form:

$$x'_j = \frac{x_j - x_{j,min}}{x_{j,max} - x_{j,min}}$$

where x_j and x'_j present the value and non-dimensional value of the element j; $x_{j,min}$ and $x_{j,max}$ are the smallest and largest amount of the factor j.

B. Ordinary Least Squares (OLS) Regression

After calculating the HPIs in London Boroughs, their validity needs to be proved, so the evidence can be found using the OLS regression, which can measure the linear association between the HPIs and the housing sales status of corresponding Boroughs in 2020. So the idea is that a set of n pairs of data points can be fitted by a linear function, and the value of housing sales are predicted with the values of HPIs and the following regression line:

$$\hat{S} = \hat{\beta}_0 + \hat{\beta}_1 \cdot HPI$$

Two parameters are the intercept $\hat{\beta}_0$ and the slope $\hat{\beta}_1$ of the equation. However, this paper does not expect to find the best fit line to make a prediction of the sales but to validate the index. So the intercept $\hat{\beta}_0$ can be ignored at the first attempt, and if there is no substantial evidence in the regression, it will be reconsidered. The residuals ε, errors between actual values and predicted values, is an indicator to estimate the performance of the linear model, while the least square method is to minimize the squared prediction residuals by the expression:

$$\varepsilon = \sum_n (S_n - \hat{S}_n)^2 = \sum_n [S_n - (\hat{\beta}_0 + \hat{\beta}_1 \cdot HPI)]^2$$

The minimum value of errors is obtained when its derivatives fade away(Abdi, 2007). The estimation of $\hat{\beta}_0$ and $\hat{\beta}_1$ is computed by calculating the derivatives of ε with respect to each of which when the value of every derivative is zero. The final equation can be simplified as an expression with the average value of sales (S_{mean}) and HPIs (HPI_{mean}):

$$\hat{\beta}_0 = S_{mean} - \hat{\beta}_1 HPI_{mean}$$

$$\hat{\beta}_1 = \frac{\sum(S_n - S_{mean})(HPI_n - HPI_{mean})}{\sum(HPI_n - HPI_{mean})^2}$$

C. Spatial Autocorrelation

Spatial autocorrelation(Griffith, 1987) reflects the strength of the interdependence between factors in different areas. Moran's I statistic is utilized to explore the clusters of sites with similar HPIs or dispersion of those with dissimilar HPIs. If the value of Moran's I increases, there will be a strengthening spatial autocorrelation of HPIs in London Boroughs. Moran's I denoted as I can be computed as:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} d_i d_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \sum_{i=1}^{n} {d_i}^2}$$

where n is the total number of factors, d_i and d_j present the deviations of the i-th and j-th Borough's HPI to their means: $(HPI_i - \overline{HPI})$ and $\left(HPI_j - \overline{HPI}\right)$, $w_{i,j}$ is the spatial weight between factor i and j.

Additionally, LISA indicates the local spatial autocorrelation of HPIs as follows:

$$LISA_i = z_i \sum_j w_{ij} z_j \, , z_i = \frac{HPI_i - \overline{HPI}}{\sqrt{\frac{1}{n} \sum (HPI_i - \overline{HPI})^2}}$$

According to LISA, HPIs can be classified into four groups, as shown in Table 4.

Table 4: Description of LISA clusters

	Cluster	Meaning
1	The High-High (HH) cluster	High value surrounded by high values
2	The Low-Low (LL) cluster	Low value surrounded by low values
3	The High-Low (HL) cluster	High value surrounded by low values
4	The Low-High(LH) cluster	Low value surrounded by high values

D. Geospatial Detector

Spatial heterogeneity(De Marsily et al., 2005) is another geographical phenomenon implying unequal spatial variance of factors. A geospatial detector(Wang et al., 2016) can be employed to examine the determinants of spatial heterogeneity between stratified areas, as spatial stratified heterogeneity. Its primary assumption is that if there is an association between two variables, they will follow a similar spatial distribution. It computes the proportion of the within sum of variance in each sampling zone to the global variance in the entire study area, which is defined as follows:

$$q = 1 - \frac{1}{N\sigma^2} \sum_{h=1}^{l} N_h \sigma_h^2$$

where σ^2 and σ_h^2 represent the variance of the factor in the whole study area and that in the zone h respectively, N is the number of the units. The value of q-statistics identifies the strength of the spatial stratified heterogeneity of response variable HPI and measures the contribution of explanatory factors to the HPI. Besides, the geospatial detector is able to estimate the interactive influences between every two factors to the HPI, which is defined as the interaction detector(Wang et al., 2016).

3. Findings & Discussion

3.1 An assessment of residential preference

The HPIs for London Boroughs are calculated based on six factors, as described in section 2.1. Before exploring this index, it needs to be tested with respect to real-life housing sale status by the OLS regression, and their distributions are shown in Fig.1(a) & (b), respectively. In Table 5, the regression result illustrates that they are positively correlated with each other, and there is a significant correlation due to the small p-value (<0.05). As described in section 2.2 A, the intercept of the regression line is not considered because a good Pearson correlation coefficient has been obtained, at 0.66, and this step is not aimed at predicting real estate sales but to find a relationship between them. Therefore, the HPI can be an indicator to reflect housing preference in different Boroughs.

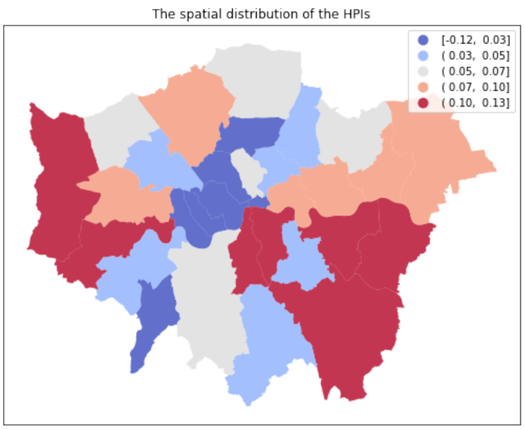

Fig.1(a): The spatial distribution of HIPs;

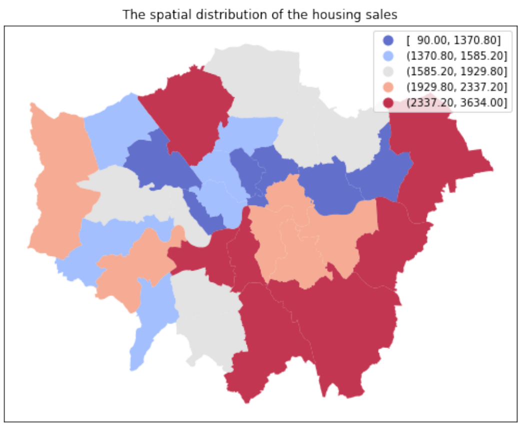

Fig.1(b): The spatial distribution of housing sales

Table 5: The result of OLS regression

	coeffience	Standard error	t	P>|t|	[0.025	0.975]
HPI	20370	2596.564	7.846	0.000	15100	25700

Fig. 1(a) shows the spatial distribution of the index with five sets based on natural breaks. The southeastern and western Boroughs have the highest HPIs, while their values are lowest in the central part of London. To be specific, the average value of HPIs is 0.0555, while the Boroughs who obtain the top five HPIs are Southwark, Bromley, Greenwich, Hillingdon, and Hounslow, at 0.1347, 0.1192, 0.1149, 0.1126, and 0.1086, respectively. There are smaller indexes in Ealing, Havering, Barking & Dagenham, Newham, and Tower Hamlets, which show the mass prefers to live in the east and west part of London.

In contrast to these places, City of London(-0.1242), Kensington & Chelsea(-0.0805), Westminster(-0.0083), Haringey(-0.0077) and Camden(-0.0054) have negative values of HPIs. In other words, housing in the central London Boroughs and Haringey is not recommended.

3.2 Spatial distribution of the HPI

According to Fig. 1(a) & (b), it seems that there is a spatial autocorrelation of HPIs in London. The results of the Global Moran's I prove this hypothesis statistically, as shown in Fig 2 (a) & (b). Its value is 0.1643 with a p-value at 0.04 (<0.05) and the z-score at 2.01 (>1.96), which implies that HPIs do not follow the complete spatial random (CSR) distribution. There are spatial clusters consisted of the regions which have a similar value of HPIs.

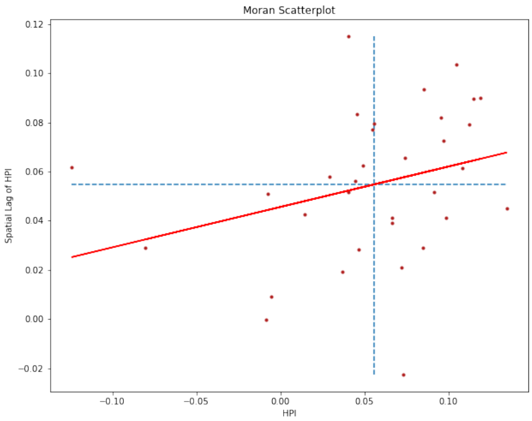

Fig.2 (a): The scatterplot of Global Moran's I;

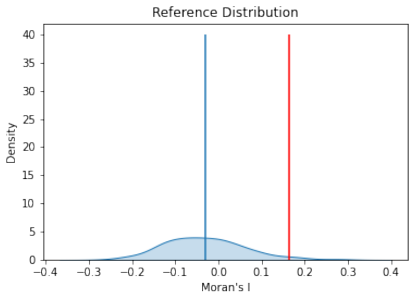

Fig.2(b): The reference ditribution

Additionally, the relationship between each Borough and its neighbors is varied based on the LISA cluster map, as shown in Fig.3. The HH cluster contains Barking & Dagenham, Greenwich, Bexley, and Bromley, while the LL cluster comprises three areas: Brent, Camden, and Westminster. Also, Islington, Wandsworth (HL), and Lewisham (LH) have an opposite association with nearby regions.

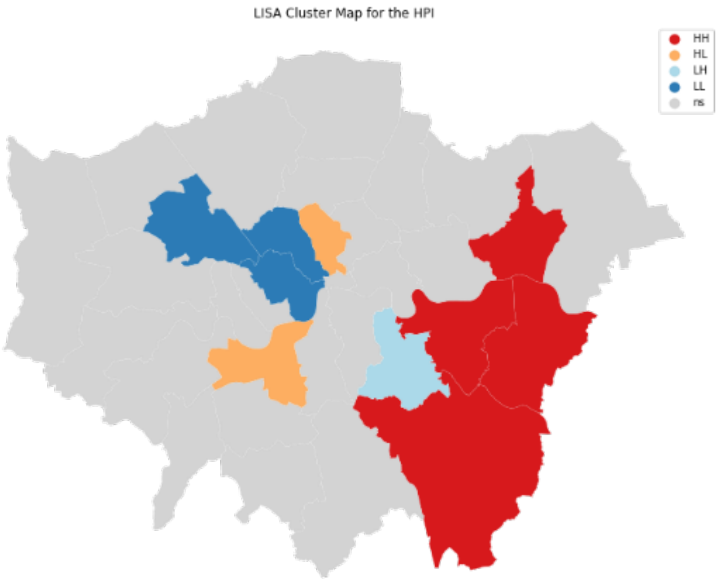

Fig.3: The LISA cluster map

3.3 Spatial stratified heterogeneity of the HPIs

The result of the geospatial detector confirms the spatial stratified heterogeneity of HPIs, as shown in Table 6. To be specific, since p-values of all factors are less than 0.1, they are statistically significantly explanative to HPIs, and their distributions are shown in Fig.4 (a)~(f). average income, green & blue cover, housing price may have more significant impacts on this index. However, the crime rate in 2020 may affect the index marginally, and its p-value will not be significant if its threshold is set as 0.05. The reason may be that this work does not consider the impact of the Covid-19 pandemic on the crime rate.

Also, the accessibility of public transportation sees the same issue, and it can be tested with a new data set like the accessibility in 2020 in the future work.

Table 6: The result of factor detector

	Housing price	Crime rate	The density of road network	Accessibility of public transport	Green & blue cover rate	Average income
q statistic	0.64	0.34	0.49	0.40	0.66	0.66
p value	0.000	0.081	0.027	0.058	0.000	0.000

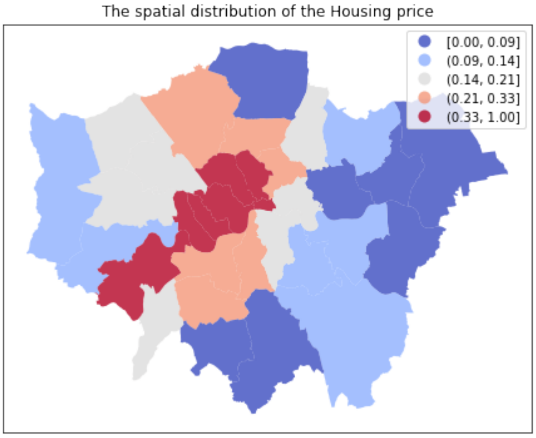

Fig.4 (a): The spatial distribution of the Housing price

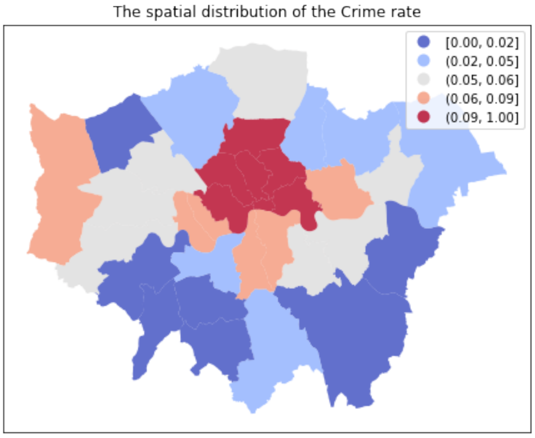

Fig.4 (b): The spatial distribution of the Crime rate

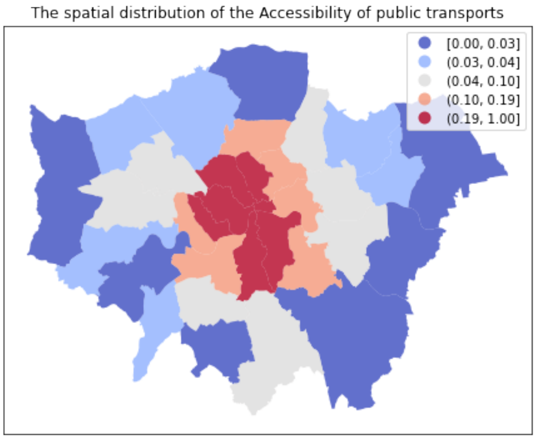

Fig.4 (c): The spatial distribution of the Accessibility of public transports

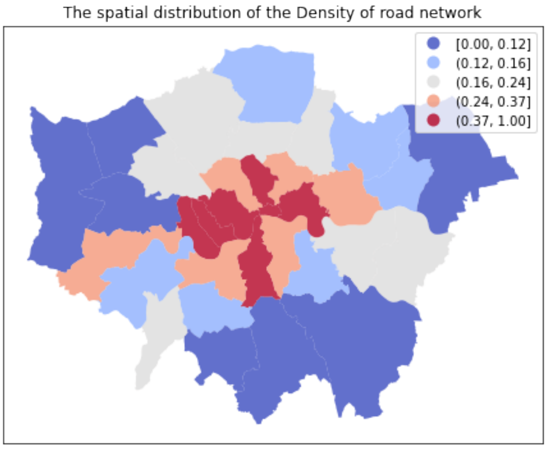

Fig.4 (d): The spatial distribution of the Density of road network

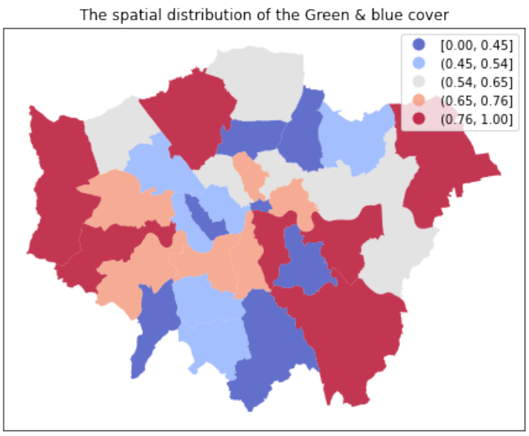

Fig.4 (e): The spatial distribution of the Green & blue cover

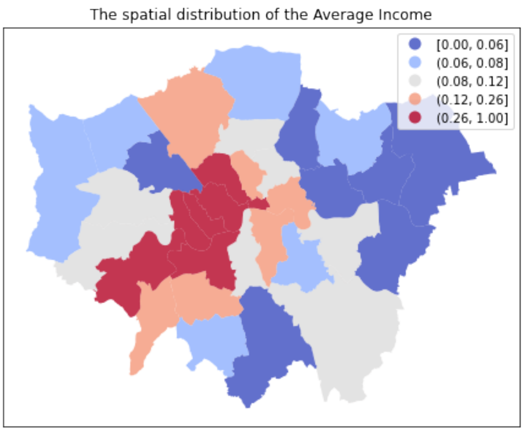

Fig.4 (f): The spatial distribution of the Average income

According to the result of the interaction detector (Table 7), there is strong positive interaction of two factors. The interactive effect of housing price and green & blue cover is the greatest, and the combined q value is 0.94. Specifically, the Boroughs of the HH cluster have high housing prices and low green & blue cover.

Table 7: The result of factor detector

	Housing price	Crime rate	The density of road network	Accessibility of public transport	Green & blue cover rate	Average income
Housing price	0.6431					
Crime rate	0.7448	0.344				
The density of road network	0.80	0.50	0.49			
Accessibility of public transport	0.76	0.41	0.65	0.40		
Green & blue cover rate	0.94	0.67	0.82	0.74	0.66	
Average income	0.75	0.68	0.71	0.71	0.88	0.66

4. Summary & Recommendations

This paper proposes the index of residential location preference, the HPI, to qualify the housing choice in London Boroughs. The factors are selected using the theory of Maslow's Hierarchy of Needs, including the housing price, crime rate, accessibility of public transportation, the density of road network, green & blue cover, and average income.

The HPI has a positive linear correlation with sales status, and it is feasible to use the HPI to reflect residential location preference. This index is subject to a positive spatial autocorrelation in London. Specifically, there is a low HPIs cluster of the central London Boroughs and Haringey, and government needs to focus on the housing problem in these regions. On the contrary, housing in Brent, Camden, and Westminster is recommended, but it is crucial to control the unusual increasing tendency of real estate prices in these areas and maintain their market advantages. Besides, the increasing gap between housing price and green & blue cover may cause stronger stratified heterogeneity of HPIs, and there are interactive influences of different factors on housing choice. Therefore, the combined effects of comprehensive factors should be considered to tackle the housing crisis in London.

References

ABDI, H. 2007. The method of least squares. *Encyclopedia of Measurement and Statistics. CA, USA: Thousand Oaks*.

BAQUTAYAN, S. M. S., ARIFFIN, A. S. B. & RAJI, F. 2015. Describing the need for affordable livable sustainable housing based on Maslow's theory of need. *Mediterranean Journal of Social Sciences,* 6**,** 353-353.

DE MARSILY, G., DELAY, F., GONçALVèS, J., RENARD, P., TELES, V. & VIOLETTE, S. 2005. Dealing with spatial heterogeneity. *Hydrogeology Journal,* 13**,** 161-183.

GHAZALI, E. M., NGIAM, E. Y.-L. & MUTUM, D. S. 2020. Elucidating the drivers of residential mobility and housing choice behaviour in a suburban township via push–pull–mooring framework. *Journal of Housing and the Built Environment,* 35, 633-659.

GRIFFITH, D. A. 1987. Spatial autocorrelation. *A Primer (Washington, DC, Association of American Geographers)*.

GUO, S. & HARDIN, W. G. 2014. Wealth, composition, housing, income and consumption. *The Journal of Real Estate Finance and Economics,* 48, 221-243.

GUO, Y. & PEETA, S. 2020. Impacts of personalized accessibility information on residential location choice and travel behavior. *Travel Behaviour and Society,* 19, 99-111.

JIANG, B., LARSEN, L., DEAL, B. & SULLIVAN, W. C. 2015. A dose–response curve describing the relationship between tree cover density and landscape preference. *Landscape and Urban Planning,* 139, 16-25.

JIM, C. Y. & CHEN, W. Y. 2007. Consumption preferences and environmental externalities: A hedonic analysis of the housing market in Guangzhou. *Geoforum,* 38, 414-431.

KARSTEN, L. 2007. Housing as a way of life: Towards an understanding of middle-class families' preference for an urban residential location. *Housing studies,* 22, 83-98.

KOVACS-GYöRI, A., CABRERA-BARONA, P., RESCH, B., MEHAFFY, M. & BLASCHKE, T. 2019. Assessing and representing livability through the analysis of residential preference. *Sustainability,* 11, 4934.

KURNIAWAN, C., DEWI, L. C., MAULATSIH, W. & GUNADI, W. 2020. Factors Influencing Housing Purchase Decisions of Millennial Generation in Indonesia. *International Journal of Management,* 11.

MASLOW, A. H. 1958. *A Dynamic Theory of Human Motivation,* Cleveland, OH, US, Howard Allen Publishers.

MORROW-JONES, H. A., IRWIN, E. G. & ROE, B. 2004. Consumer preference for neotraditional neighborhood characteristics. *Housing Policy Debate,* 15, 171-202.

SCHEUER, S., HAASE, D., HAASE, A., WOLFF, M. & WELLMANN, T. 2021. A glimpse into the future of exposure and vulnerabilities in cities? Modelling of residential location choice of urban population with random forest. *Natural Hazards and Earth System Sciences,* 21, 203-217.

SONG, Y. & ZHANG, C. 2020. City size and housing purchase intention: Evidence from rural–urban migrants in China. *Urban Studies,* 57, 1866-1886.

TRAVERS, T., SIMS, S. & BOSETTI, N. 2016. Housing and inequality in London. *Centre for London*.

WANG, J.-F., ZHANG, T.-L. & FU, B.-J. 2016. A measure of spatial stratified heterogeneity. *Ecological Indicators,* 67, 250-256.

ZAVEI, S. J. A. P. & JUSAN, M. M. 2012. Exploring housing attributes selection based on Maslow's hierarchy of needs. *Procedia-Social and Behavioral Sciences,* 42, 311-319.