

# The economic value of streets: mix-scale spatio-functional interaction and housing price patterns

Yao Shen   Kayvan Karimi

Space Syntax Laboratory, Bartlett School of Architecture, University College London,  
England, United Kingdom

## ABSTRACT

Location factors are vital elements for describing housing price variation. However, limited studies have explicitly illustrated the relationship between urban design and the heterogeneity of housing price patterns. This article specifically evaluates how the interactions between the spatial layouts and land-use system at various scales through street network affect the valuation of the residential properties and the segmentation of housing markets in a network-based Mixed-scale Hedonic Model (MHM) where the submarkets pattern are determined and annotated by the spatially varying estimates on streets. The application of the delivered method in the case of Shanghai City, China, confirms the necessity of using the non-Euclidean distance metric and represent the coexistence between the stationarity and the non-stationarity of the introduced street accessibility variables. The results provide evidence that the impacts of street accessibility measures on the local levels showcase significant spatial variation. It is common for all the places that the properties located on the streets with the higher levels of angular closeness, smaller values of angular betweenness and longer angular distance to the nearby land-uses at the larger scales will be bided higher. It is proven that our delineation of submarket performs better in prediction accuracy than the traditional submarket specifications. The detected submarkets pattern yields that reachable land-use diversity at the pedestrian level is not a preferred factor in the housing submarkets located in the developed city centres. The signs of the price effects of the angular distance to local land-uses distinguish the developing submarkets as two main groups with different degrees of geometrical walkability. It is suggestive that continuously developing pedestrian-oriented neighbours in the walkable areas could contribute to decelerating the growth of house price in Chinese cities. The productions of this study can enrich the understanding of the socioeconomic effects of urban design with greater spatial precision across submarkets.

## KEYWORDS

Housing price; Spatial configuration; Land-use system; Submarket regionalisation;  
Network accessibility; Spatial autocorrelation; Space syntax; Urban design

## 1. INTRODUCTION

As the vital factors reflecting the economic externality, location variables capture the geographical characteristics of city properties. The connotations of location can span from environmental quality, landscape comfort, socioeconomic features, to travelling impedance, etc. In urban studies, accessibility is a critical concept that represents the *potential of opportunities for interaction* in different locations (Hansen, 1959). It has been widely adopted for assessing urban place's advantages and related policy distributions (e.g., Geurs et al., 2015, Stewart, 1948 and Geurs and Van Wee, 2004). Its origins can be traced back to the monocentric Alonso–Muth–Mills model (Anas, Arnott, & Small, 1998). The effectiveness of accessibility measures has been proven in the housing price analysis for describing the process that how location advantages are capitalised into residential property values (e.g., Adair et al., 2000, Boyle and Kiel, 2001, Smersh and Smith, 2000 and Song and Knaap, 2003). Related research focusing on the regional determinants or the distance factors in the analysis of housing price patterns were typically conducted based on the assumption that built environment is homogeneous across the landscape (Batty, 2009). Consequently, the influence of the fine-grained design of built environment upon the housing price variation is over-simplified or overlooked. The absence of the proper consideration of spatial disparity of urban space thereby constraining the possibility of explicitly understanding the socioeconomic impacts of urban design.

Accessibility is formed in urban spatial configurations which can be divided into two subsystems: the spatial configuration and the functional system. Efforts have been made on measuring accessibility symmetrically based on these two structures respectively without proper interlinks between one another (Karlström & Mattsson, 2009). In the field of transport geography, the land-use system is one of the main focuses; the accumulated chances and the travelling energy/cost expenditure to them are conceptualised as the metrical gravity to urban opportunities (Wegener, 2004). Related measures have been widely discussed in housing studies to investigate the impacts of the transport transits (Bowes & Ihlanfeldt, 2001), school quality (Haurin & Brasington, 1996), transport investments (Banister and Berechman, 2003 and Henneberry, 1998) and the attractiveness of landscape (Luttik, 2000) on housing appreciation. By contrast, configurational studies use public space as the opportunity landscape and emphasise on the cognitive efforts beyond the metric distance cost that are required to interconnect segmented spaces, which provides an opportunity of scrutinising how the topo-geometrical properties of urban grids differentiate urban places and implying the informational efforts for human movement (Hillier and Hanson, 1984, Hillier, 2003, Kim and Penn, 2004 and Kwan, 2000). The topo-geometrical nature of urban movement is reflected in the process how people recognise the value of location externality because location conditions are sensed by individuals through other people's observable movement rather than through the 'unreadable' built environment directly (Hillier, 1996). As the fundamental of public goods, the geometrical properties of urban grids will be priced to maximise the efficiency of accessibility allocation (Webster, 2010). Recent concern has been given to investigate the relationship between street network centralities and the property value patterns and the impacts of urban form on submarket delineation (e.g., Chiaradia et al., 2009, Enström and Netzell, 2008, Law, 2017, Law et al., 2013, Law et al., 2015, Matthews and Turnbull, 2007, Shen and Karimi, 2015 and Xiao et al., 2016). Understanding the linked nature between these spatial and functional structures is critical for the cities in the process of rapid urban growth and generation (Tallon, 2013), where the

change of spatial and functional elements are increasingly significant, such as the cities in current urbanised China.

Yet, successful marriage between the geometrical and the configurational models with consistent theoretical and methodological foundations is still sparse, thereby restricting our knowledge with regard to the socioeconomic consequences of urban design. In recent time, it was argued that accessibility of urban form and its related attractions can be reconceptualised as 'spatial capital' to assess the performativity of urban space (Marcus, 2010 and Berghauser Pont and Marcus, 2014). It is shown that adding geographical accessibility into the space syntax model can enhance the predictability of the statistical regression model for predicting human pedestrian patterns (Ozbil, Peponis, & Stone, 2011) and describing the distribution of commercial frontages (Porta et al., 2012, Scoppa and Peponis, 2015 and Sevtsuk, 2014). As a reflection on the criticism of configurational measures as a one-dimensional approach for simplifying the impacts of urban attractions (Ratti, 2004), recent methodological developments of network accessibility have addressed the role of the buildings' attractiveness in the configurational analysis (Stähle et al., 2005 and Sevtsuk and Mekonnen, 2012). Lately, through incorporating the place attractiveness and the cognitive cost, Shen and Karimi (2016) introduced a more comprehensive framework of characterising and segmenting streets based on the delivered urban function connectivity measures, including the accessible density, diversity and the delivery efficiency with social media check-ins data and validated its methodological novelty of predicting socioeconomic performance. In this study, we unravel the price effects of location advantages that are measured by space syntax centralities and urban function connectivity indices, respectively, on the property values.

In hedonic price theory, the urban property is priced for its inherent utility-bearing characteristics (Lancaster, 1966 and Rosen, 1974) including its structural features, location situations and the neighbourhood effects (Bourassa et al., 2007 and Dubin, 1988). The existence of spatial heterogeneity of housing price pattern indicates that property value is not self-existent but closely related to its surrounding property values. In this sense, the spatial autocorrelation between property values can hardly be captured by the structural or locational variables; therefore, the neighbourhood effects should be taken into account in the hedonic regression models (e.g., Goodman, 1978, Hwang and Thill, 2009 and Hui et al., 2014). Due to the recent development of the local regression methods, for instance, notably the locally weighted regression (LWR) method (Cleveland & Devlin, 1988), geographically weighted regression (GWR) (Brunsdon et al., 1996 and Fotheringham et al., 2003), spatial autoregressive models (Kelejian & Prucha, 1998), etc., addressing the neighbourhood impacts is suggested to be an effective way of controlling the spatial variation of hedonic price functions, improving the prediction accuracy (e.g., Goodman & Thibodeau, 1998) and generating the reasonable submarkets (Bourassa et al., 1999 and Bourassa et al., 2007; Helbich et al., 2013 and Helbich et al., 2013). Nevertheless, it was found that not all variables vary geographically, and sometimes only certain parameters influence housing prices based on spatial locations (Wei & Qi, 2012). Thus, spatially homogeneous factors and heterogeneous elements should be considered simultaneously so that spatial knowledge of property valuation can be advanced. Moreover, the priority of adopting the landscapes of local estimates in housing submarket segmentation process has been distinguished from conventional ways to define submarkets, for instance, demarcating the submarkets according to the constant marginal prices (Goodman & Thibodeau, 2007) or with distinct

patterns of observed characteristics (Palm, 1978), as its independence of the exogenously predefined units and effectiveness for improving the accuracy of the hedonic models for the defined submarkets (Helbich, Brunauer, Vaz & Nijkamp, 2013). Against this background, the mixed-hedonic model (MHM) is recognised as an efficient regression approach to address the neighbourhood effects by combining the standard GWR model and the conventional hedonic model to capture the complex interactions between variables and to define the emergent submarket patterns in a data-driven procedure.

Housing hedonic studies can be conducted with different variables settings on different scales, which might lead to the different conclusions with the modifiable areal unit problems (MAUPs). Conventionally, the MHM was employed with the regional indicators and the Euclidean distance metrics beyond the sampling landscape (Fotheringham et al., 2003), which are not particularly suitable for studying the impacts of built environments within non-Euclidean metrics on an intra-city scale (Lu et al., 2011 and Lu et al., 2014). Urban streets, the places where the real urban economy occurs (Hillier and Hanson, 1984, Hillier and Hanson, 1997, Hillier and Vaughan, 2007 and Jacobs, 1961), have been paid very few attention in hedonic regression models. In fact, they are not only the urban elements that deliver spatial and functional convenience but also the proper spatial units for the high-resolution analysis of housing price pattern. Additionally, the network distance along the streets explicitly portrays the actual spatial interactions between properties, and its contribution to optimising the statistical performance of the GWR has been confirmed in recent studies on housing price distribution (Lu et al., 2014, Lu, Charlton, Brunsdon, & Harris, 2016).

In this study, we apply a network-based MHM framework to explore the role that the fine-scaled spatio-functional interaction plays in the observed spatial heterogeneity and housing market segmentation of Central Shanghai. The scope of this research is three-folded: 1) to examine whether spatial accessibility and function connectivity metrics through street network impact the housing price patterns; 2) to investigate whether the effects of spatial accessibility and function connectivity indices perform uniformly across space; 3) to study the contribution of spatio-functional interaction through streets at different scales to forming the rational submarkets distributions. Check-in point-of-interest (POIs) and the road networks are employed to compute space syntax centrality and function connectivity metrics through streets, respectively, which are further used as location variables in the proposed network-based MHM with the non-Euclidean distance metric. Based on the local coefficients generated from the detected local parameters in the network-based MHM instead of raw data, we use a network-constrained k-means clustering analysis to identify the housing submarkets. In this respect, the effects of spatio-functional interplay through streets on the spatial heterogeneity and the submarket delineation of the house price patterns can be explicitly revealed.

Ongoing research is constructed in the following four sections. The subsequent section presents the research framework and the computation of the spatial and functional accessibility measures, followed by a brief introduction of the data and study area. The empirical results are interpreted in the fourth section. This article concludes with the related discussion, potential implications, and suggestions for the future efforts.

## 2. THE METHOD

### 2.1. Research design

The study design in this article is illustrated in Fig. 1. This framework includes a stepwise research procedure, which generally follow the methods conducted by Bourassa et al. (2007) and Helbich, Brunauer, Hagenauer and Leiner (2013) with different variables settings, definition of spatial relationship, and basic unit for analysis. In the first step (step a), the required data are obtained from websites through data mining processes. The house price sample is randomly split into two parts for cross-validation: one part is the training data for regression analysis and the remaining is the test data for the validation and the annotation of the generated submarkets. In step b and c, spatial and functional network accessibility measures are computed and then assigned to the corresponding residential properties based on their exact locations on roads. In the following step, these location features and structural features are used as the inputs in the proposed MHM to generate the street coefficient maps that are interpolated with spatial constrains that are defined by the connecting relationship between road segments. At the last stage, a data-driven community detection method with network connectivity constraints is adopted to group the street network as the submarkets.

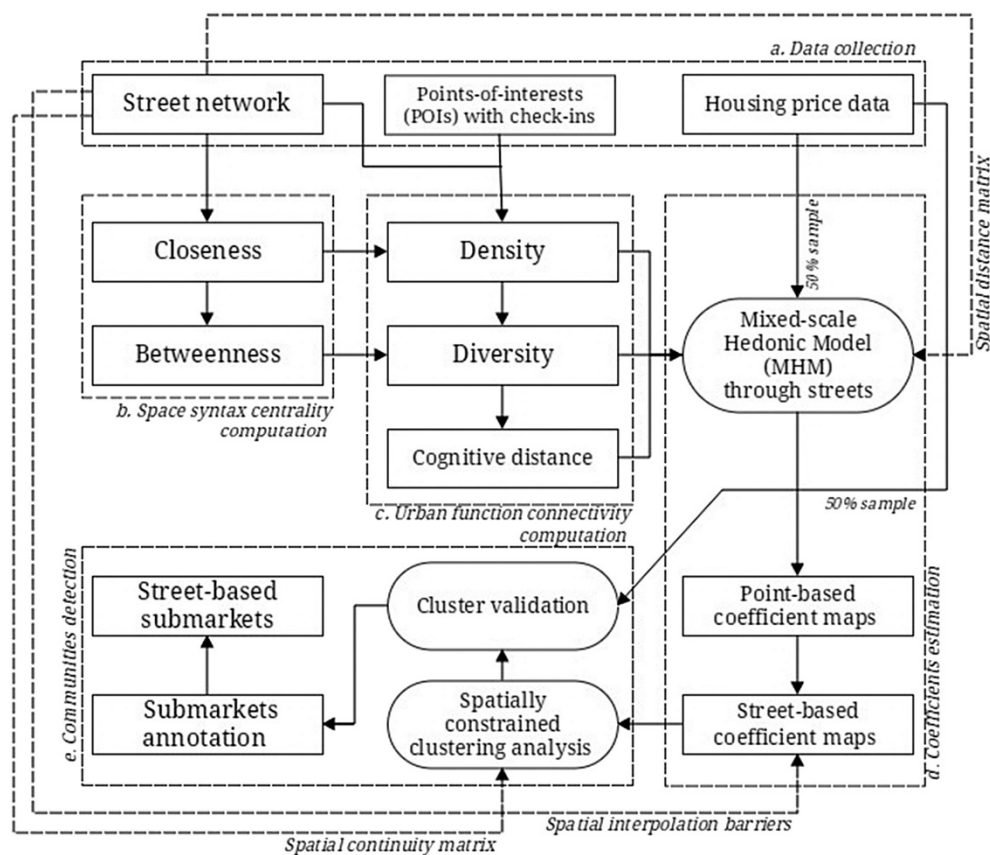


Fig. 1.

The network-based mixed-scale hedonic model (MHM) for estimating the coefficient maps of local factors and detecting the housing submarkets patterns.

## 2.2. Indexing streets with the neighbouring spatial and functional context

As argued in the previous context, street accessibility measures based on a graph representation of built environment can provide finer-scaled maps of location advantage with the non-Euclidean settings, which are also fit for the need of explicitly describing the features of the neighbouring urban design situations. The urban built environment here is conceptualised as an opportunities landscape where the individual public spaces and the land-use amenities are treated as spatial and functional opportunities through streets respectively. Accordingly, two groups of the network accessibility measures are used in this study: the spatial accessibility measures that are the space syntax centrality indices, including integration and choice, and functional accessibility measures that are the urban function connectivity indices, containing density, diversity and the cognitive distance (Fig. 2). Typical radii, including 500 m, 1,000 m, 2,500 m, 5,000 m, and 10,000 m, are used to represent the spatial and functional accessibilities across scales. In our experiments, these radii can successfully summarise the spatial accessibilities at various scales by controlling the potential multicollinearity and the boundary effects.

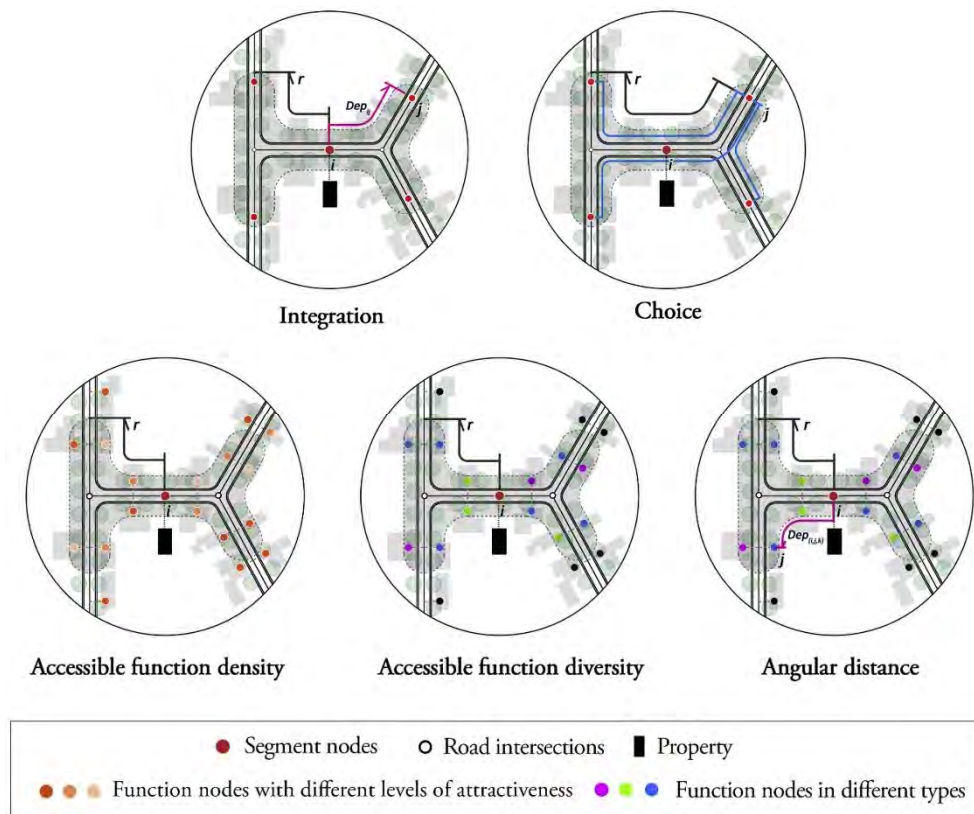


Fig. 2.

Graphic illustration of the space syntax centralities (integration and choice) and the function connectivity indices (accessible function density, accessible function diversity and angular distance) within the reachable area for a segment node  $i$  where the property was located in. The reachable area is defined by a metric radius  $r$ . The angular distance from the segment node  $i$  to another segment node  $j$  is  $Dep(i,j)$ ; and  $Dep(i,j,k)$  denotes to the angular distance between the segment node  $i$  and the function node  $j$  of specific land-use type  $k$ .

### 2.2.1. Street-based spatial accessibility – space syntax centrality indices

The so-called space accessibility indices are computed according to space syntax theory and include both segmental angular integration and choice (Hillier & Iida, 2005, pp. 475–490; Hillier et al., 2012 and Turner, 2001a). The platform used to calculate these two types of spatial accessibility measures is Depthmap, which was developed by Turner (2001b). The spatial accessibility indices assigned to the residential properties are the integration and choice values of their nearest segments. Angular integration can be understood as an adjusted closeness in graph analysis and it represents the angular shallowness between segments, which can be formally defined as follows:

$$INT_{(i,r)} = \frac{(N_i - 1)}{\sum_{j=1}^J Dep_{(i,j)}}, \{dis_{(i,j)} \leq r\} \dots\dots\dots (1)$$

In this equation, the angular integration ( $INT_{i,r}$ ) at the radius  $r$  is represented as the reciprocal of the mean angular depth ( $Dep_{(i,j)}$ ) from segment  $i$  to all the reachable segments  $j$  within a buffer zone defined by a pre-identified radius  $r$ .

The so-called choice measurement ( $CHO_{(i,r)}$ ) in space syntax is similar to the concept of betweenness in graph analysis but it uses the angular change at each road intersection as the expenditure to find the shortest path (Turner, 2001a). It can be defined as the number of times ( $n_{jk}$ ) that the focused segment  $i$  has been passed through in the angular shortest paths from segment  $j$  to  $k$  within a reachable area defined by a fixed radius  $r$ . From a mathematical perspective, this definition can be represented as:

$$CHO_{(i,r)} = \sum_{k=1}^K n_{jk}, \{dis_{(i,j)} \leq r; dis_{(i,k)} \leq r\} \dots\dots\dots (2)$$

### 2.2.2. Street-based functional accessibility – urban function connectivity measures

There are three principal aspects of urban function connectivity: density, diversity and cognitive distance to the reachable complementary amenities (Shen & Karimi, 2016). The accessible function density index measures the accumulation of scored urban activities from each street within a defined radius through the reachable shortest paths. Assuming there are  $K$  types of active land-uses, the accessible function density for the segment node  $i$  at radius  $r$  would be aggregated as  $DEN_{(i,r)}$ :

$$DEN_{(i,r)} = \sum_{k=1}^K \sum_{j=1}^J O_{(j,k)} \times W_{(j,k)}, \{dis_{(i,j)} \leq r\} \dots\dots\dots (3)$$

This summation takes into account the function nodes assigned to street segment edges and weighted based on the social media check-in scores. In the equation above,  $r$  is the defined radius,  $O(j,k)$  refers to the potential opportunities that can be supplied and  $W_{(j,k)}$  is the specified weight for the function node  $j$  in type  $k$  to reflect its popularity. In this study, we use the normalised amount of social media check-ins assigned to each POI as its weight on

the demand side, which can be presented as  $W_{(j,k)} = \frac{\log C_{(j,k)}}{\log C_k^{max}}$ , where  $\log C_{(j,k)}$  represents the log-normalised check-ins for the specific function node  $j$  in type  $k$ , and  $\log C_k^{max}$  denotes to the log-normalised value of the maximum check-ins for all the function nodes. Due to lacking of the reliable corresponding information about the land-use from the supply perspective for calibration, we assume that the supply of the function is even across locations.

The accessible function diversity index measures the balance degree of all reachable weighted urban activities from the original street within a given radius. In this study, information entropy is applied to measure the diversity of urban function nodes from segment node  $i$  at the radius  $r$ , and represented as  $DIV(i,r)$  ( $DIV(i,r) = [0, 1]$ ). Further, a normalisation transformation is applied to enable the different types of activities to be comparable by converting the absolute density to a relative density by dividing the accessible weighted density in type  $k$  for each segment node by the maximum value of the accessible density of land-use of the same type at the same radius for all the segment nodes within the

study area ( $NaDEN_{(i,k,r)} = \frac{DEN_{(i,k,r)}}{DEN_{(k,r)}^{max}}$ ). The presence probability ( $P(i,k,r)$ ) of the function nodes in type  $k$  at radius  $r$  for segment nodes  $i$  is measured by its empirically observed frequency of normalised density ( $NaDEN(i,k,r)$ ) among all  $K$  types of land-uses.

The computation of accessible diversity can be formally represented as follows.

$$DIV_{(i,r)} = \frac{-\sum_{k=1}^K P_{(i,k,r)} \times \ln(P_{(i,k,r)})}{\ln(K)}, \quad \{dis_{(i,j)} \leq r\} \dots\dots\dots (4)$$

$$P_{(i,k,r)} = \frac{NaDEN_{(i,k,r)}}{\sum_{k=1}^K NaDEN_{(i,k,r)}} \dots\dots\dots (5)$$

The cognitive distance (angular step depth) index measures the mean angular cost to all the reachable urban activities from the original street within a given radius through the shortest paths. It reveals the cognitive efforts required to reach all accessible functions from the original street segments beyond the same energy expenditure that is measured in the light of the metric length of the streets. This index can be formally expressed in the following equation.

$$DIS_{(i,r)} = \frac{\sum_{k=1}^K \sum_{j=1}^J Dep_{(i,j,k)}}{N_{(i,r)}}, \quad \{dis_{(i,j)} \leq r\} \dots\dots\dots (6)$$



In the equation above,  $Dep(i,j,k)$  shows the angular step depth from segment node  $i$  to function node  $j$  in type  $k$  within the buffer area defined by the radius  $r$ , and  $N(i,r)$  is the summation of the accessible functions at the same radius.

### 2.3. The network-based MHM

The purely local regression model may not always be the best approach to explore the relationship between response and explanatory variables. An appropriate solution to address this issue is to apply a regression model in which both local and global effects are properly defined and positioned. The MHM is such a model, in which the coefficients that are proven to be non-fluctuant across locations will be kept constant, thus improving prediction efficiency. Therefore, the purely local hedonic model (LHM) can be extended into a mix-scale model (MHM) to reflect real spatial complexity in housing price modelling. A standard MHM can be formulated as follows:

$$y_i = \sum_{g=1}^{k_a} \beta_g x_{ig}(a) + \sum_{j=1}^{k_b} \beta_j(u_i, v_i) x_{ij}(b) + \varepsilon_i \dots\dots\dots (7)$$

where  $k_a$  and  $k_b$  denote the total number of global and local parameters of the variables, respectively;  $x_{ig}(a)$  represent the global variables;  $x_{ij}(b)$  represent the local variables; and  $\beta_g$  is the  $g$ th parameter associated with the global explanatory variables at all locations while  $\beta_j(u_i, v_i)$  is the coefficient for the  $j$ th local variable at the location  $i$  with a pair of coordinates  $(u_i, v_i)$ .

In the MHM, the spatial relationship between the observed samples is defined through a spatial distance matrix that is relatively independent from the GWR model. Therefore, converting the conventional MHM to a network-based MHM can be achieved by using a pre-specified non- Euclidean distance matrix to replace the default Euclidean distance matrix. In this study, the network distance matrix is computed using the cumulative walking distance along the shortest path between any two residential locations in the sample. The adaptive Gaussian kernel function is applied to estimate the weight of the neighbourhood effects and the minimised cross-validation (CV) score is used to determine the optimised bandwidth for regression.

The adoptability of the MHM relies on a calibration procedure in which a multiple stepwise regression algorithm is used to test the geographical variability of each variable. This process is undertaken literally through the model comparisons between all pairs of the fitted GWR model, namely, the purely local GWR and a modified model in which only the  $k$ th coefficient is fixed globally. By comparing the difference of fit-goodness in regression models measured by the reduction of the Akaike Information Criterion (AIC), we can then determine which local factors should be regarded as global ( Nakaya, Fotheringham, Brunsdon, & Charlton, 2005).

The network-based inverse distance weighting (IDW) interpolation method is adopted to produce the network-based coefficient maps for the detected local variables (Okabe &

Sugihara, 2012). We argue that urban street-based maps are closer to the reality of local impact distributions as urban public spaces are unevenly distributed as the carriers of urban daily life.

#### **2.4. Network-constrained cluster analysis via principal component analysis**

The network constrained k-means clustering analysis is employed to delineate the structure of the emerging submarkets. In order to control the potential correlations between the local coefficient patterns, principal component analysis (PCA) is employed to generate the latent variables for maintaining the most useful information for the further clustering analysis. Spatial continuity of the submarkets is critical for delineating the localised submarket pattern where related policy can be allocated with explicit spatial boundaries. Some approaches to mitigate this issue have been used and discussed, including using factor analysis (Watkins, 1999), taking into account the coordinates of spatial units in clustering analysis (Wu & Sharma, 2012), adopting spanning trees to constrain the proximity of units (Assunção, Neves, Câmara, & da Costa Freitas, 2006), etc.. As the detected housing submarkets in this study are network-based, the spatially constrained rule for the community detection is considered as the node continuity for the network groups that contain the contiguous polyline features. That is, only edges/segments that share a node can be considered as the part of the same community in the cluster analysis. Lastly, we use the average out-of-sample prediction error (the root mean square error (RMSE)) and AIC of hedonic models for all the clusters to identify the optimised number of submarkets. The whole process for delineating street-based submarkets is data-driven from the bottom up without the prior knowledge.

### **3. STUDY AREA, DATA AND MODEL SPECIFICATION**

#### **3.1. Study area**

The central areas in the Shanghai Metropolitan Area (SMA) is selected for the empirical study (Fig. 3). Shanghai is the economic centre and one of the biggest municipalities in China, along with Beijing, Tianjin, and Chongqing. Given its geographical location, Shanghai was the city that experienced the earliest modernisation in China after the Second Opium War. Since the Open Door Policy was implemented in 1979, Shanghai has grown significantly and has transformed into a mega city. At present, the city has 23 million residents living in the 6,340 km<sup>2</sup> administrative area. We did not examine the entire metropolitan area in this case study; instead, we focused only on housing prices at the intra-urban scale within the central area, where census units are considerably smaller and highly urbanised. The central area in SMA serves as an important case study for this article because of the complexity of its built environment and the maturity of its housing market.

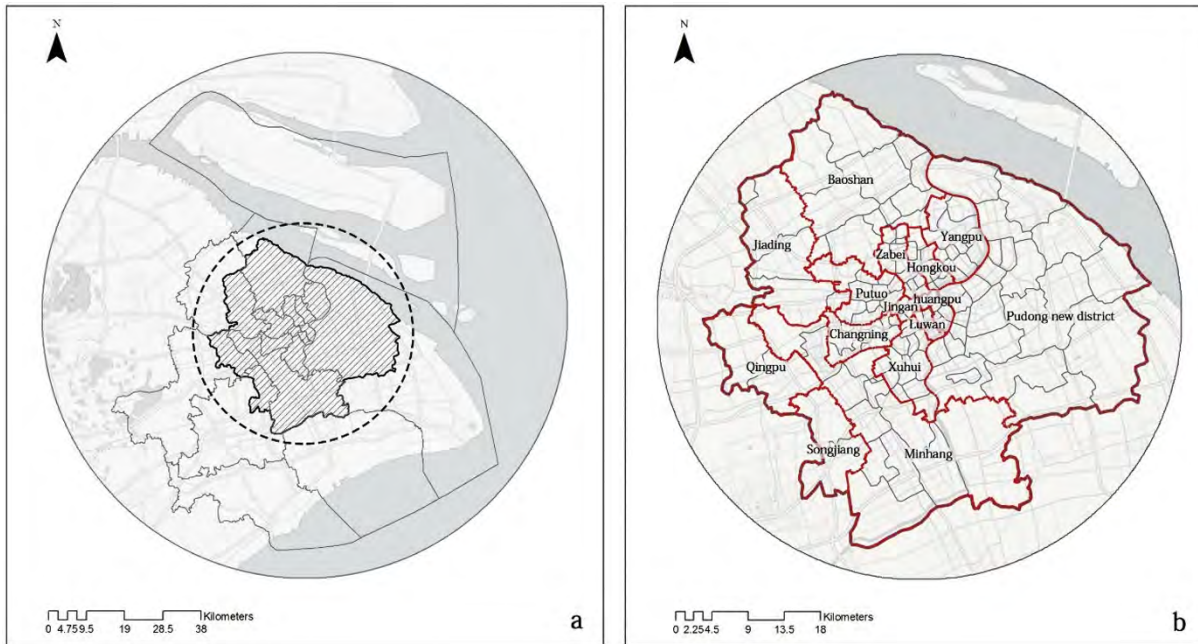


Fig. 3.

Study areas in the Shanghai Metropolitan Area (SMA): (a) the location of the study area in the SMA; (b) the administrative districts in Central Shanghai.

### 3.2. Data

#### 3.2.1. Housing price data

The housing price dataset was collected from the websites of local property sales agents using a data crawling approach; hence, we use asking price data rather than the actual transaction records for housing sales. With the consideration of the unavailability of the official housing data in China, these datasets are alternatively reliable with sufficient details and a large coverage. Previous studies have also validated these records by comparing them to actual selling prices of houses and have asserted the reliability of these datasets (Chandler & Disney, 2014). Our datasets cover the period from October 2014 to December 2014. The sample size is 8,262 with 3,818 unique locations after removing duplicate records. Fig. 4 shows that the housing price distribution generally follows a steep trend of decay from the downtown area to the suburban areas. The housing price distribution indicates that the highest price cluster emerges at the city centre, and the decay trend is extremely significant because of the high value of the alpha parameter ( $\alpha = 5$ ), indicating that the house price in Shanghai follows a relaxed scaling law that there are far more properties with lower price than the properties with higher price.

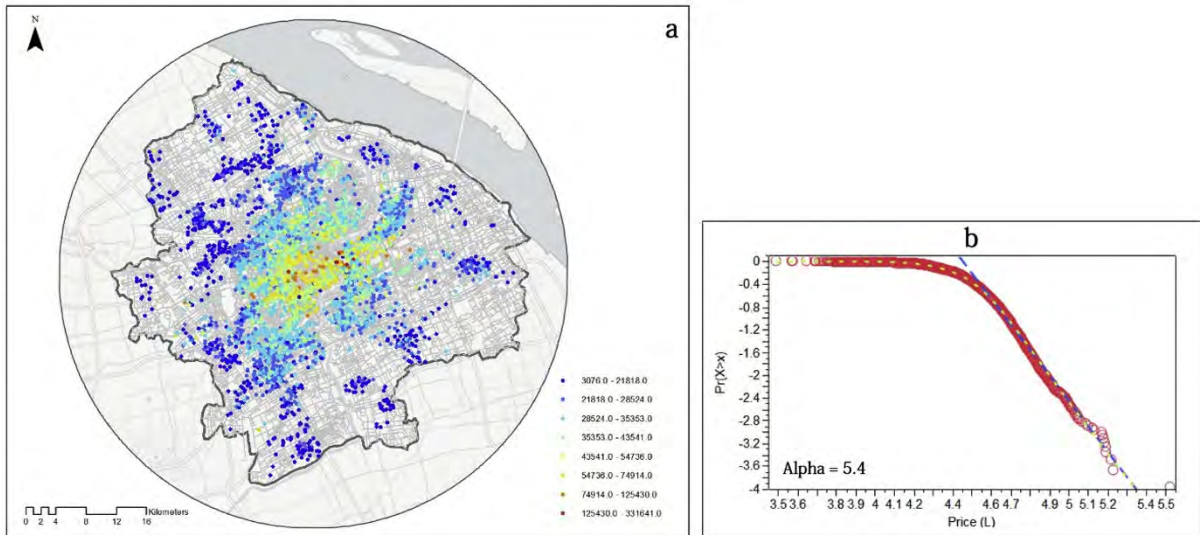


Fig. 4.

Housing price map of central area in Shanghai (a) and the complementary cumulative distribution function (CCDF) pattern of housing prices (b).

### 3.2.2. Road network and check-in POIs

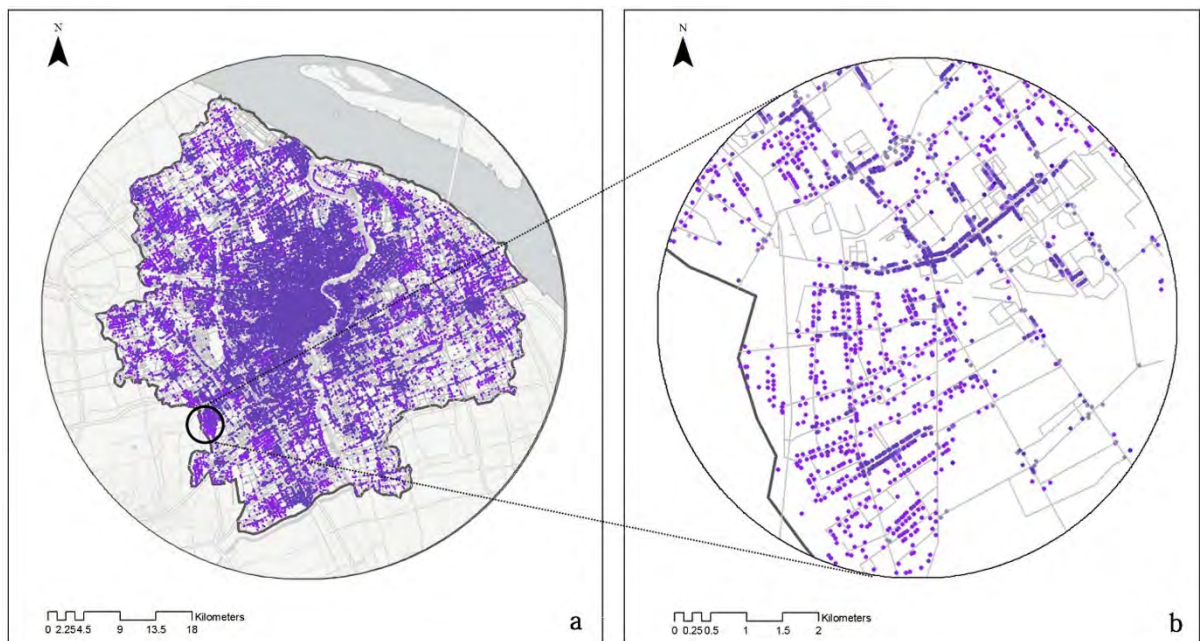


Fig. 5.

Roadwork and POIs in the study area: (a) road network overlapped with POIs in the study area; (b) road network and check-in POIs in the sample area (the darkness of POIs denotes to the check-in scores in the social media).

The road network and the scored POIs data are collected from a navigation company and a social media service provider in China, respectively. The segmental map has 92,920 segments after being converted from an axial map, whereas the POI dataset consists of 191,035 point-based land-use venues with social media check-in records. The check-in POIs are reclassified into 11 main types of active land-use for reflecting the complementarity among land-uses (Hess, Moudon, & Logsdon, 2001). The defined active land-uses include retail, catering, hotel, office, recreation, public service, park, education, hospital, culture, and transportation. All land-uses are used to compute accessible function density, whereas only the defined active land-uses are employed to accessible function diversity. A sample of the road network and POIs datasets are shown in Fig. 5.

### 3.3. Variables structure design

The variable design in the various hedonic equations serves for verifying different theoretical assumptions. In this study, from an urban design perspective, the overarching assumption is that the contextual design of the spatial network and land-use system influence the house valuation process across space. Although some other factors have been evidently studied to impact the housing price in Chinese cities, e.g., green space viewing (e.g., Jim & Chen 2006), distance to CBD (e.g., Mok, Chan, & Cho, 1995), distance to tube stations (e.g., Wu, Gyourko, & Deng, 2012), school accessibility (e.g., Wen, Zhang, & Zhang, 2014), managing cost (e.g., Chen, 1996), etc., these factors are not included as variables in the hedonic model used in this research because they either form the independent markets by their own factors or have been reflected by the network accessibility measures to a large extent (Law et al., 2013). Therefore, we use the asking house price per square metre as the dependent variable with two main types of explanatory variables including the structural variables and the location variables for describing the contextual characteristics of the housing market in the study area. The structural variables, such as the useable floor size, typology, the number of bedrooms, the number of bathrooms, etc., are used as the inputs to the proposed model. The location variables include the space syntax centrality measures and the function connectivity metrics at different levels. The detailed definitions of all variables included in this study and their expected effects on housing price are given in Table 1. The factors describing spatial comfort and convenience are expected to be positive variables. Consequently, floor area, the number of rooms, the degree of depreciation, property's direction (indicating the design-related energy efficiency (Kroll & Cray, 2010)), floor level and the neighbour's greenness are the augmenting factors, only the floor area ratio – the indicator of development intensity – is expected to be a suppressing structural factor. In the family of space syntax centrality measures, the angular integration, capturing the spatial convenience by indexing the space relatedness, is supposed to be a positive factor, while the angular choice is expected to be negatively related to house price as it represents the possible places where traffic congestion might happen with potential noise and air pollution. Accessible function density and diversity are assumed to have positive impacts on house price growth, and the cognitive distance measure is considered as a suppressive factor since they illustrate the richness and the shallowness of the existence of daily amenities for the desired convenience in people's minds.

Table 1.

Variables and their definitions used in the hedonic model of residential properties in Central Shanghai.

Variables	Unit/coding	Type	Definitions	Expected sign
<i>House price variable</i>				
HSP <sup>¶</sup>	RMB	Numeric	The asking price of the residential property	NA
<i>House structural variables</i>				
SIZE <sup>¶</sup>	m <sup>2</sup>	Numeric	The total unusable floor area	+
TYPE	0/1 (0=no; 1=yes)	Dummy	Existence of detached house	+
BEDR <sup>¶</sup>	Count	Numeric	Quantity of bedrooms	+
LIVR <sup>¶</sup>	Count	Numeric	Quantity of living rooms	+
BATR <sup>¶</sup>	Count	Numeric	Quantity of bathrooms	+
AGE <sup>¶</sup>	Count	Numeric	Age of the house in years	+
ORIT	0/1 (0=no; 1=yes)	Dummy	Existence of north-south exposure	+
FLOR <sup>¶</sup>	Count	Numeric	The floor where the flat locates	+
FAR <sup>¶</sup>	-	Numeric	The floor area ratio of the residential community	-
GR	%	Numeric	The site green intensity of the residential community	+
<i>Location variables</i>				
<i>Space syntax centrality variables</i>				
INT %radius% <sup>¶</sup>	-	Numeric	The angular closeness of street network at a fixed radius	+
CHO %radius% <sup>¶</sup>	Count	Numeric	The angular betweenness of street network at a fixed radius	-
<i>Urban function connectivity variables</i>				
DEN %radius% <sup>¶</sup>	Count	Numeric	The accessible function density of accessible POIs at a fixed radius	+
DIV %radius%	%	Numeric	The accessible function diversity of accessible POIs at a fixed radius	+
DIS %radius% <sup>¶</sup>	Count	Numeric	The mean angular step depth to the accessible POIs at a fixed radius	-

Note: '¶' indicates that a logarithmic transformation is applied. '+' denotes to the augmenting effect and '-' refers to the suppressing effect. %radius% includes 500m, 1,000m, 2,500m, 5,000m, and 10,000m

## 4. EMPIRICAL RESULTS IN CENTRAL SHANGHAI

### 4.1. Street-based MHM results

The analysis of the residuals generated by the global hedonic model in Shanghai implies that a significant spatial autocorrelation has been empirically discovered (Moran's Index = 0.367, z-score = 25.128,  $p < 0.001$ ), which further implies the potential spatial instability of the global hedonic modelling of housing patterns and confirms the necessity of addressing the neighbourhood effects in the MHM.

The results of the stepwise ordinary least squares (OLS) analysis, the standard MHM with straight line distance and the network-based MHM are reported in Table 2. Minimising the AIC is applied as the criteria with a threshold of the generalised variance inflation factor for the variables ( $VIF < 3$ ) in the stepwise variable selection procedure for producing a parsimonious OLS model. Only 14 variables are captured as the critical factors accounting for the house price variation in the final model. House orientation, the number of living rooms, and the FAR are the insignificant factors in the global hedonic model. The function accessibility metrics at the global scale and the spatial accessibility indices at the local scales (except for the choice at 1,000 m) are absent from the model, indicating that Shanghai is still in the process of urban growth in which the spatial adjustment of the global centre have not been sufficiently followed by the land-use transformation at the city-wide scale.

Table 2.

Regression results for the house price variation in Central Shanghai

Variable	OLS	Standard MHM	Street-based MHM		
	Beta coefficients	Mean. estimate	Mean. estimate	Min. estimate	Max. estimate
Intercept		0.018	0.021	-2.164	2.658
BEDR	-0.094***	-0.095	-0.100	-1.629	0.864
BATR	0.158***	0.110	0.119	-0.481	1.502
SIZE	0.256***	0.201	0.218	-1.258	1.646
DIS500	0.014***	0.012	0.019	-0.512	0.467
DIV500	-0.049*	-0.036	-0.052	-1.983	0.877
DEN2500	0.299***	0.254	0.241	-1.824	2.787
CHO1000	-0.046***	-0.015	-0.011	-1.864	1.180
TYPE	-0.038***				
FLOR	0.042***				
AGE	-0.100***				
GR	0.085***				
DIS2500	0.166***				
INT10000	0.369***				
CHO10000	-0.071***				
AIC	7632.549	6752.267	6103.285		
AIC reduction from OLS	0	1280.282	1529.264		
Adjusted R <sup>2</sup>	56.958%	69.216%	75.479%		

Significant: '\*\*\*' – 0.001, '\*\*' – 0.01, '\*' – 0.05; Note: VIFs (variance inflation factors) for all the variables are below 3 in the OLS model.

By comparing the performance of these particular forms of hedonic models, the effectiveness of employing the network-based MHM is illustrated. The explanatory power of street-based MHMs is 18.521% higher than that of the global model. Moreover, the goodness-of-fit of the standard MHMs is also improved when the Euclidean distance metric is replaced by the network distance. Therefore, it suggests that the network-based version of MHM is statistically favoured because the travelling distance can describe the neighbourhood effects more explicitly.

The optimised bandwidths for the standard MHM and the network-based MHM are 108 and 116 respectively. The reduction of AIC values demonstrates that 7 variables in the total 14 variables are better assumed to local in both specified MHMs. The stationary counterparts of MHM show the clearly identical performance across Central Shanghai. Specifically, the newly-built properties with higher green ratios are more likely to be preferred in Shanghai's housing market regardless of the houses' geographical locations. Simultaneously, for all the properties, space syntax integration has a significant impact on property values at the large scale (10,000 m). Furthermore, the exhibited negative impact of the choice measure at the large scale (10,000 m) and the positive effects of the cognitive distance to urban function at the semi-local scale (2,500 m) are both statistically significant without spatial variation, which implies that the disamenity impacts of the over-crowded human movements and the traffic congestion on housing price patterns are spatially homogeneous. This result validates the results in other studies that integration and choice perform positively and negatively in the housing market (Law, Chiaradia, & Schwander, 2012; Shen and Karimi, 2015 and Xiao et al., 2016) and further suggests an additional finding that the angular proximity to existing land-uses at the semi-local level would be not favoured for the whole housing market. The coexistence of the stationary and non-stationary coefficients demonstrates that the Shanghai's housing markets are economically interconnected based on the global factors on the one hand, but they are differentiated relying on the local factors on the other hand.

The model fit for the network-based MHM varies from 0.20 to 0.95 across the study space. Fig. 6 illustrates the interpolated coefficient maps of 7 non-stationary variables which exhibit both negative and positive effects on housing price. The structural variables quantifying the indoor layouts including the floor area, the number of bedrooms and the number of bathrooms can generally distinguish the city centres from its context, and the violent fluctuation of the effects mainly occur in the city centre. In the city centre, floor size has the highest positive effect except for the Jinan district where bigger houses are not preferred for the potential buyers. By holding the effect of floor size, the design of the indoor layouts represents different effects for various places. Additional bathrooms are favoured by the buyers with more budgets in Jinan and Huangpu districts where housing price are highest in Shanghai. This effect is reduced for other locations with the negative elasticity in Changning district and the area that is located in the west end of Pudong new district. Also, the significantly negative impacts of the number of bedrooms are recognised to be more pronounced in the Huangpu district, but its positive influence can be observed in other places. The spatial performance of local structural factors illustrates that the design features of residential properties are the key dimension reflecting the varying balance between the housing locational supply and associated demands in different urban areas.

The effects of spatial design and land-use distributions at the local and semi-local radii vary across the street landscape, illustrating the externality of housing prices at different scales. Most areas are characterised by the positive influence of accessible urban diversity at 500 m, angular choice at 1,000 m and the urban density at 2,500 m, although the rise and fall of the local coefficients are dramatically significant in the central districts. By contrast, the angular distance to urban land-uses at the local scale is recognised as a negative factor for many areas, particularly for the relatively historically developed areas on the right bank of the Huangpu River. House buyers are more likely to pay more for a home located on a street that is configurationally proximal to the land-uses at the pedestrian level in Pudong new district



where most of the properties are designed in the large gated communities with more cognitive distance from the main roads in nature.

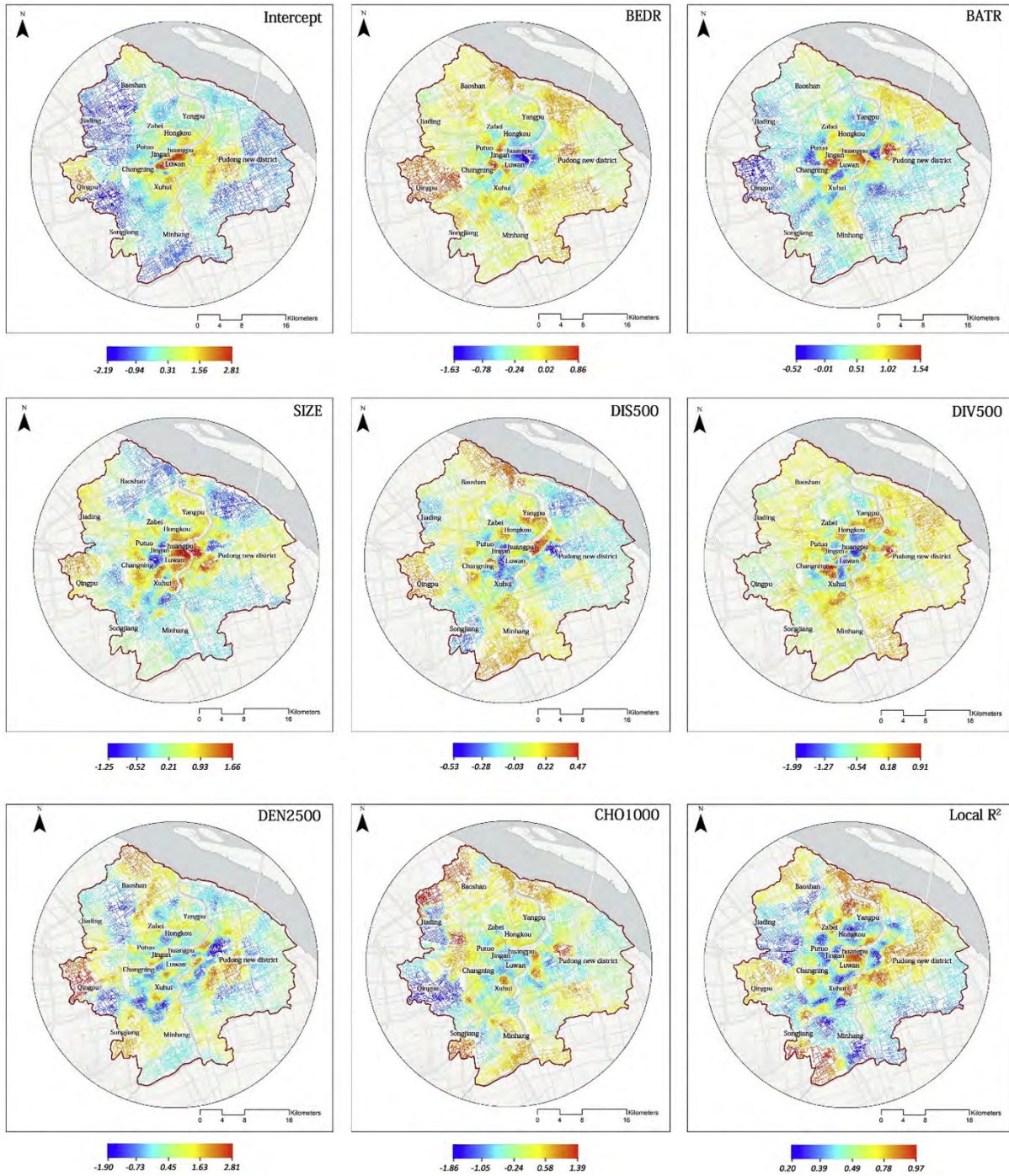


Fig. 6.  
The interpolated coefficient maps of the local variables in the street-based MHM.

## 4.2. Street-based submarket detection

The coefficients landscapes of 7 local variables are used in a network-constrained k-means clustering analysis to segment the entire study area into submarkets. Before that, we employ PCA to determine the principle components (PCs) to reduce the potential multicollinearity among the coefficients of all local variables. 6 PCs are finally selected to present 84% of the information in the original dataset. As shown in Fig. 7, the optimised number of clusters for the projected study area is 10 because of the minimised scores of the out-of-sample AIC and RMSE. The overall pattern of the street-based housing submarkets identified and represented in Fig. 8.

Compared with the submarkets determined based on the assumed spatial units such as grid cells or census wards, the spatial partitions of submarkets in Central Shanghai in this study reflect the real spatial connectivity relationship without the scattered distributions. The rivers, hills, and large parks without dense road connections are considered as the natural boundaries of the detected submarkets. Only the submarket 8 spans two areas that are connected by many bridges, which can be validated by the administrative boundaries that cover these two areas as well, suggesting that spatial policy might also matter in this submarket. Nevertheless, the non-correspondence between census districts and the generated street-based submarkets in other regions implies that submarkets are formed in a sophisticated manner and are difficult to be captured if the predefined spatial units are used.

The partitioned submarkets patterns show significant differences concerning the compositions of the local estimates (Fig. 9). Table 3 summarises the final annotation of each submarket by combining the observed place characteristics and the significant estimates of local variables. The adjusted R-square value varies from 0.54 to 0.71. The Chow test confirms that the 10 submarkets based on detected coefficients of the stationary variables are statistically differentiable at 1% confidence level (Table 4). In comparison with other transitional segmented models in terms of the weighted standard error (WSE) reduction, our alternative specification improves 17.15% the performance of the city-wide model, while the segmented model based on administrative boundary at the district level reduces 10.88% of the baselined WSE value (Table 5). Moreover, the building-type and orientation submarkets do not pass the significance threshold of 5%. This finding is supportive of the evidence that the spatially varying impacts of the design of the built environment and the property provide the more essential information for segmenting housing submarkets than the existing administrative districts, although the spatial allocation of local policy in different districts is still influential.

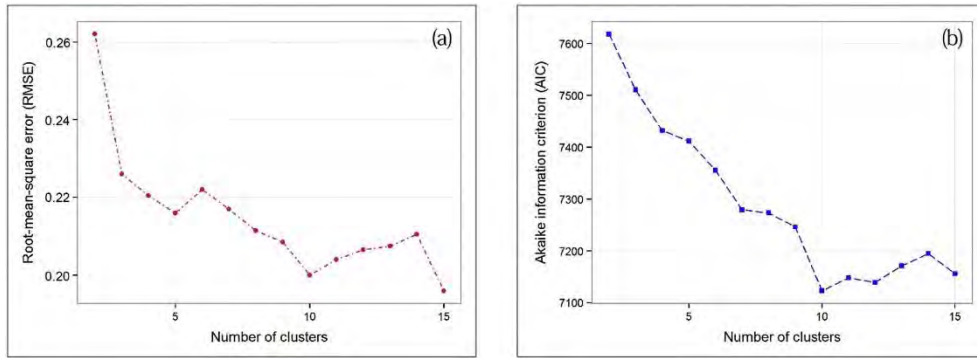


Fig. 7. Calibration of the optimised number of clusters based on the AIC and RMSE values in the test models

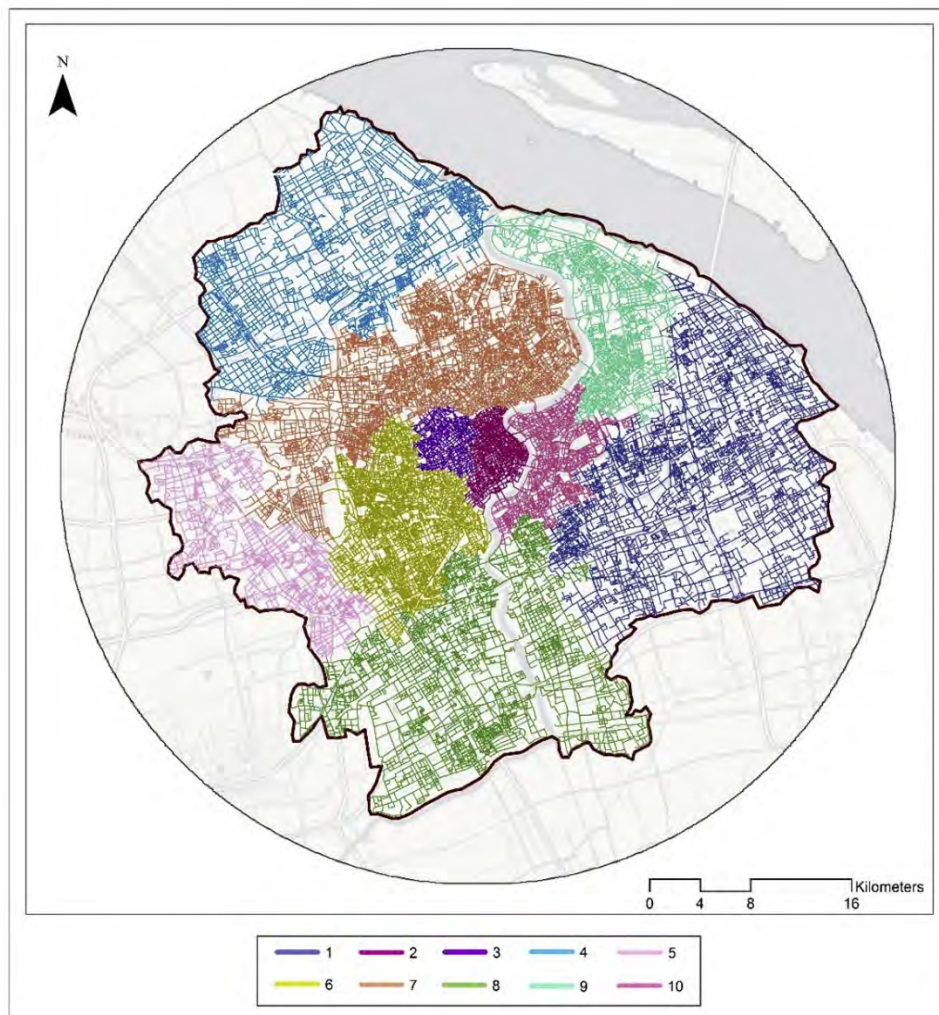


Fig. 8. Detected submarkets in Central Shanghai.

Compared with the submarkets determined based on the assumed spatial units such as grid cells or census wards, the spatial partitions of submarkets in Central Shanghai in this study reflect the real spatial connectivity relationship without the scattered distributions. The rivers, hills, and large parks without dense road connections are considered as the natural boundaries of the detected submarkets. Only the submarket 8 spans two areas that are connected by many bridges, which can be validated by the administrative boundaries that cover these two areas as well, suggesting that spatial policy might also matter in this submarket. Nevertheless, the non-correspondence between census districts and the generated street-based submarkets in other regions implies that submarkets are formed in a sophisticated manner and are difficult to be captured if the predefined spatial units are used.

The partitioned submarkets patterns show significant differences concerning the compositions of the local estimates (Fig. 9). Table 3 summarises the final annotation of each submarket by combining the observed place characteristics and the significant estimates of local variables. The adjusted R-square value varies from 0.54 to 0.71. The Chow test confirms that the 10 submarkets based on detected coefficients of the stationary variables are statistically differentiable at 1% confidence level (Table 4). In comparison with other transitional segmented models in terms of the weighted standard error (WSE) reduction, our alternative specification improves 17.15% the performance of the city-wide model, while the segmented model based on administrative boundary at the district level reduces 10.88% of the baselined WSE value (Table 5). Moreover, the building-type and orientation submarkets do not pass the significance threshold of 5%. This finding is supportive of the evidence that the spatially varying impacts of the design of the built environment and the property provide the more essential information for segmenting housing submarkets than the existing administrative districts, although the spatial allocation of local policy in different districts is still influential.

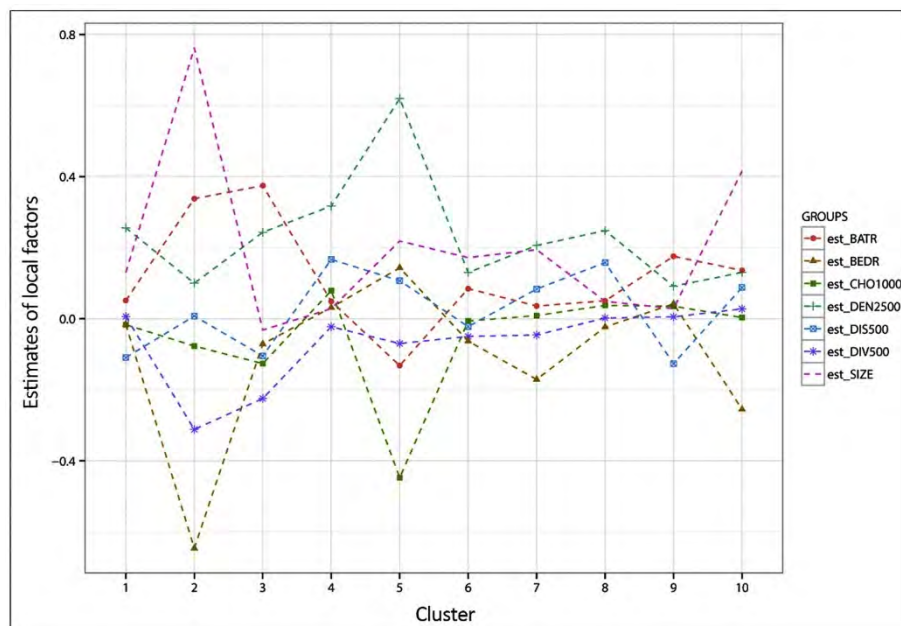


Fig. 9.

The estimates of local factors in every submarket based on the test data.

Table 3.

Annotation of the emerging submarkets by using the significant local estimates.

Observed characteristics	Submarkets	Coefficients of the significant local factors		Average house price (RMB per square metre)	Adj. R <sup>2</sup>
		Positive factors (p < 0.05)	Negative factors (p < 0.05)		
Central city centre (West bank)	2	SIZE, BATR, DEN2500	BEDR, DIV500, CHO1000	50,736	0.71
Developed city centre (West bank)	3	BATR, DEN2500	DIV500, CHO1000	47,909	0.54
Modern business region (East bank)	10	SIZE, BATR, DEN2500	BEDR	39,903	0.72
Developed sub-centre region (West bank)	6	SIZE, DEN2500, BATR		36,783	0.69
Developed sub-centre region (West bank)	7	DEN2500, SIZE, DIS500		31,095	0.63
Developing region (East bank)	9	BATR, DEN2500	DIS500	29,995	0.59
Developing region (East bank)	1	DEN2500, SIZE	DIS500	25,934	0.63
Developing region (East & west bank)	8	DEN2500, DIS500		25,340	0.64
Developing region (West bank)	5	DEN2500, SIZE, BEDR, DIS500	CHO1000, BATR	22,865	0.63
Developing region (West bank)	4	DEN2500, DIS500		19,824	0.65

Table 4.

Chow test for the test dataset.

Submarket	Obs.	Submarket								
		1	2	3	4	5	6	7	8	9
1	317									
2	233	11.755 <sup>¶</sup>								
3	234	18.843 <sup>¶</sup>	5.972 <sup>¶</sup>							
4	361	23.254 <sup>¶</sup>	22.932 <sup>¶</sup>	25.078 <sup>¶</sup>						
5	130	24.641 <sup>¶</sup>	11.845 <sup>¶</sup>	11.361 <sup>¶</sup>	4.883 <sup>¶</sup>					
6	902	18.126 <sup>¶</sup>	12.587 <sup>¶</sup>	8.549 <sup>¶</sup>	32.824 <sup>¶</sup>	17.037 <sup>¶</sup>				
7	819	14.385 <sup>¶</sup>	10.849 <sup>¶</sup>	8.553 <sup>¶</sup>	22.601 <sup>¶</sup>	14.436 <sup>¶</sup>	8.056 <sup>¶</sup>			
8	488	6.961 <sup>¶</sup>	10.455 <sup>¶</sup>	14.246 <sup>¶</sup>	15.549 <sup>¶</sup>	14.809 <sup>¶</sup>	9.797 <sup>¶</sup>	9.185 <sup>¶</sup>		
9	302	6.879 <sup>¶</sup>	9.816 <sup>¶</sup>	14.603 <sup>¶</sup>	32.245 <sup>¶</sup>	32.221 <sup>¶</sup>	28.626 <sup>¶</sup>	26.629 <sup>¶</sup>	16.669 <sup>¶</sup>	
10	345	9.791 <sup>¶</sup>	2.347 <sup>¶</sup>	8.365 <sup>¶</sup>	26.78 <sup>¶</sup>	18.60 <sup>¶</sup>	43.84 <sup>¶</sup>	38.00 <sup>¶</sup>	16.62 <sup>¶</sup>	5.039 <sup>¶</sup>

Note: ¶ denotes to the significance at 1% level.

Apart from the overall pattern for describing housing market, the emerging individual submarket can be identified with the influence features - the estimates of the significant local factors. By using average housing price of 30,000 RMB per square metre as the threshold, we preliminarily group the submarkets to two main categories: the developed submarkets and

the developing submarkets. The spatial and functional configurations in developed submarkets have a longer history and are denser than that in the developing submarkets, which suggests that the defined submarkets patterns based on the coefficients in the network-based MHM can capture the spatial characteristic of built environment.

The estimates of local structural factors significantly contribute to distinguishing the submarkets. The impact variations of Floor area size is more critical for developed city areas with larger coefficients, illustrating the status quo in contemporary large Chinese cities where the house floor area are strongly demanded due to the shortage of the housing supply and the rapid growth of urban population. The most significantly positive effect of floor size is found in the submarket 2 where the Chinese old town and the oldest colonial areas are located in. Furthermore, the number of bathrooms is recognised as positively related to housing price in many submarkets (2, 3, 6, 9, and 10), with the exception of submarket 5 where this effect turn to be negative and the additional number of bed rooms are more favoured.

Table 5.

Weighted standard error analysis of the submarket segmentation based on the test data.

Definitions of submarkets	Standard error	Percentage of reduction
City-wide model	0.239	
Definition with house orientation	0.238	0.42%
Definition with house type	0.234	2.09%
Definition with administrative districts	0.213	10.88%
Definition with local impacts of urban configuration and the structural features	0.198	17.15%

By holding the effects of structural variables, the design of the built environment is also recognised to be crucial elements for differentiating the submarkets. Accessible function density at the semi-local scale (2,500 m) is found to be the positive factor for all the submarkets, but its importance is inconsistent between submarkets. For the submarkets that are highly urbanised, the effects of semi-local perceived function agglomeration are relatively smaller; they turn to be more significant when the submarkets in the developing regions are the new focuses. This implies that the demand strength of urban services is dramatically increased for the urbanising areas due to the urban growth and regeneration. The accessible urban diversity at the pedestrian scale (500 m) is statistically significant with a negative sign in the submarkets (2 and 3) within the historic inner city. This can be explained by the fact that the urban amenities have been already pronouncedly clustered and mixed and the extra mixture of urban functions might lead to long-time noise thereby losing the locational enjoyment. Table 3 also shows a significant negative relationship between space syntax choice and housing values at the radius of 500 m in the submarkets 2, 3 and 5, indicating that local traffic in these submarkets will depress the house value.

Another important element for grouping submarkets is the significant influences exerted by the angular distance to local land-use at the radius 500 m. Being similar to the performance of other detected local factors, the sign and the significance of its impacts vary differently

between the submarkets. By grouping the submarkets according to their relative locations on the Huangpu River that passes through the study area in the middle, we find that the developing submarkets on the west bank are the regions with the positive impacts whereas this tendency is reversed in the submarkets on the east bank. This result might be related to the existing features of the spatial layout and the land-use system in these two regions, because the areas that we described as the developing submarkets on the west bank are relatively self-organised and more walkable with less angular depth steps to the local functions (2.9 steps in average), while developing submarkets on the east bank are newly planned after 1980s with the logic of modern urban design and more car-dependent with more angular depth to the locally neighbouring attractions (4.6 steps in average).

### **4.3. Related policy implications**

In the empirical study of Central Shanghai, it is standard for all the places that the properties located on the streets with bigger angular closeness, smaller values of angular betweenness at the large scale and longer angular distance to the nearby land-uses at the mesoscale are bid higher by the potential purchasers. For the planning and design practice, this finding reveals that the city centres with higher street segment density and connectivity are the areas where the property values are higher, whereas the traffic corridors and the global high streets are both the negative factors decreasing house value. The performance of the detected global location variables has been confirmed in other empirical studies, which implies that these regular patterns can benefit the urban planners and designers in practice for estimating the fine-resolution price effects of spatial interventions without the prior knowledge of the urban context. There is also a consensus on the more valued properties in Shanghai in terms of their structural design. For the housing market as a whole, the more newly built flats on the higher floor with better green coverage add more premiums to property value. This observed relations, however, might be different across cities since the preference on the properties' structural design is related to the socioeconomic conditions. Identifying the price impacts of these variables for a particular city can still contribute to making proper planning and design decisions on the design of housing supply.

Given the two modes of equilibriums observed in the developing submarkets on the two banks of Huangpu River in terms of the existing distinguished geometric properties of the local land-use network and the reversed signs of the associating price effects, we find that the pedestrian-oriented neighbours on the west bank with less angular depth steps to the local functions shape the hunger for the spatial privacy, whereas the gated communities with more angular depth steps to local amenities on the east foster the desire of spatial convenience to daily destinations. Within the process of urban growth, the clustering of land-use will reduce the cognitive cost to the urban attractions naturally thereby making the housing price in the large-scale gated submarkets grow faster as the demand on spatial convenience there will be met. By contrast, the hunger of spatial privacy in the pedestrian-friendly submarkets will be increasingly strong if the walkability to local amenities in this area has not been changed via the spatial interventions. These findings yield a suggestive evidence that continuously developing the pedestrian-friendly neighbourhoods and securing the publicity of urban space in the walkable areas could be an alternative way to against the excessive growth of the housing price in Chinese cities. This is can also be confirmed by the

detected globally positive effect of the angular distance to land-uses at the mesoscale for all the submarkets (see Table 2). However, it is noteworthy that reducing the cognitive cost of reaching local functions in the submarkets where the existing large-scale gated communities clustered might increase the risk of speeding up the house price growth till the new equilibrium with the desire for spatial privacy is formed. Besides, the impacts of the structural variables should also be taken into account in the action plan for stabilising house market. For example, fixing the floor size ceilings of the newly-built residential properties would be a useful measure to decelerate the growth of the house price in the city centre.

For framing a comprehensive guidance for designing the cities with stable housing economy, we propose an ideal model to achieve an understanding of the role of urban configurations in changing home prices (Fig. 10). The interaction between the spatial network and the land-use system provides various types of co-presence among various network accessibilities at different levels. The overall effects will be perceived by homebuyers, developers, and designers, who will select a suitable home or design and build ideal homes for an improved price in the market. Housing price patterns, including spatial heterogeneity and the related segmental submarkets, are simultaneously influenced by these complex urban interactions. The entire process could be triggered by the plans of the urban configuration and the design of the property plans.

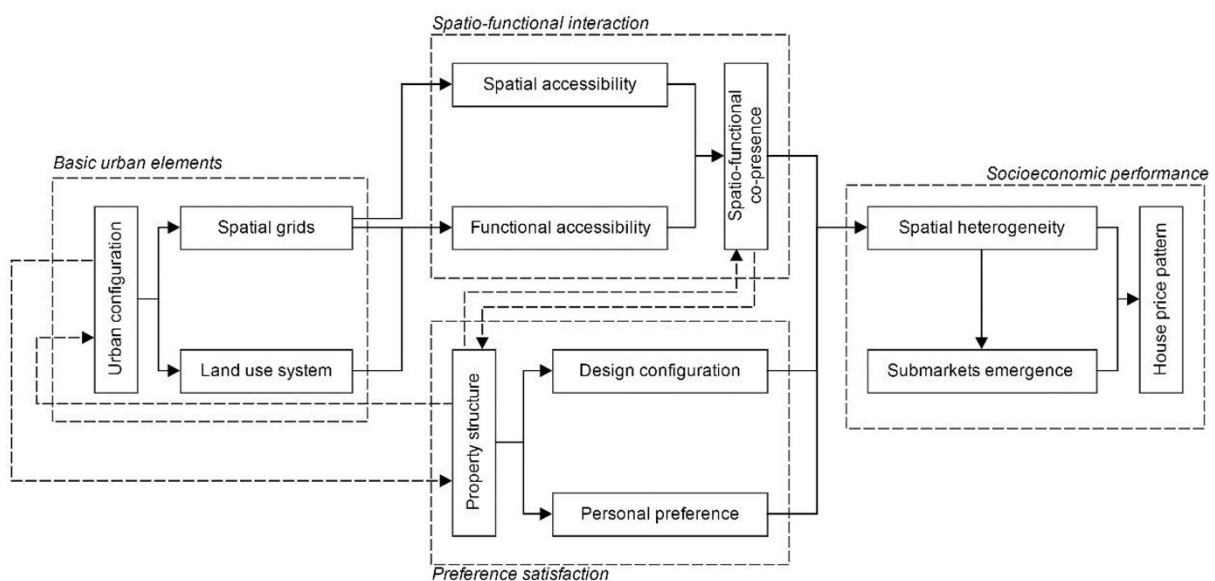


Fig. 10.

The chart describing the process that how the spatio-functional interaction influence the housing price pattern with property's structural design.

## 5. DISCUSSION

Quantifying the relationship between built environment and the spatial variation of the housing price is increasingly important for urban design and the related strategy distribution.



This demand is more pronounced for the cities within the process of rapid urban growth and regeneration. With the aim to advance the current knowledge of the spatial heterogeneity of housing price and the submarket regionalisation captured from the urban design perspective, we use the network-based MHM where the dimensions of the built environment layouts are measured by the street network centralities and the function connectivity measures instead of the distance features of some specific attractions to reflect the perceived features of the spatial design and land-use allocations. The varied influences of the street accessibility measures at different levels are then adopted to identify the submarket distribution with the estimates of the local structural variables.

The empirical analysis of the housing price pattern in Central Shanghai shows that the utilisation of the metric distance and the recognition of the local variables can improve the predictability of the spatial statistic model. The scales of the street accessibility measures are related to the discrimination between the global and local variables in the network-based MHM. The global location variables are basically at the larger scales, and the local location variables are on the smaller scales. In addition to the known effects of the global angular closeness and betweenness indices in other studies, the geometric properties of land-use system are also crucial for describing house price patterns and generating the submarket segmentation. The obtained submarkets based on the local coefficients maintain spatially contiguous boundaries that are more realistic and the performance of the intra-submarket global hedonic models are also enhanced. For instance, the angular distance to the active urban functions at the radius 2,500 m is discovered as a global force impacting positively the overall pattern of the house price. The cognitive distance to the pedestrianly accessible functions, on the other hand, is a vital factor for submarket segmentation with the local structural variables and other location variables, such as accessible function density at the mesoscale and the local urban diversity. By comparing the detected equilibriums in the developing submarkets, this study yields that designing pedestrian-oriented neighbourhoods in the walkable areas could be helpful for stabilising the house price growth in the process of rapid urban expansion and renewal. It is proven that our delineation of submarket performs better in prediction accuracy than the traditional submarket specifications. The aforementioned results support the argument that through segmenting submarkets based on the influences of urban form, function and house design, urban designers and decision makers can identify the localised areas for the specialised strategies according to the influences of specific features in different submarkets.

Although the empirical results of this study indicate the merits of the proposed method and the importance of urban design, this study nonetheless has several limitations. First, only the claimed housing price dataset is used in this research, which may not refer to the real price of housing transactions. Second, more empirical tests in different cases are required to verify the effectiveness of the proposed approach and the robustness of the related conclusions. Finally, other variables can be considered for this framework to identify the real determinants on house values. All these limitations will be addressed in future work to enhance the efficiency and accuracy of the proposed approach and to produce more insightful findings.

## REFERENCE

- Adair, A., McGreal, S., Smyth, A., Cooper, J., & Ryley, T. (2000). House prices and accessibility: The testing of relationships within the Belfast urban area. *Housing studies*, 15(5), 699-716.
- Anas, A., Arnott, R., & Small, K. A. (1998). Urban spatial structure. *Journal of economic literature*, 36(3), 1426-1464.
- Assunção, R. M., Neves, M. C., Câmara, G., & da Costa Freitas, C. (2006). Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science*, 20(7), 797-811.
- Banister, D., & Berechman, J. (2003). *Transport investment and economic development*. Routledge.
- Batty, M. (2009). Accessibility: in search of a unified theory. *Environment and Planning B: Planning and Design*, 36(2), 191-194.
- Berghauer Pont, M., & Marcus, L. (2014). Innovations in measuring density: From area and location density to accessible and perceived density. *NA*, 26(2).
- Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). Spatial dependence, housing submarkets, and house price prediction. *The Journal of Real Estate Finance and Economics*, 35(2), 143-160.
- Bourassa, S. C., Hamelink, F., Hoesli, M., & MacGregor, B. D. (1999). Defining housing submarkets. *Journal of Housing Economics*, 8(2), 160-183.
- Boyle, M., & Kiel, K. (2001). A survey of house price hedonic studies of the impact of environmental externalities. *Journal of real estate literature*, 9(2), 117-144.
- Bowes, D. R., & Ihlanfeldt, K. R. (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics*, 50(1), 1-25.
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4), 281-298.
- Chandler, D., & Disney, R. (2014). The housing market in the United Kingdom: Effects of house price volatility on households. *Fiscal Studies*, 35(3), 371-394.
- Chen, A. (1996). China's urban housing reform: price-rent ratio and market equilibrium. *Urban Studies*, 33(7), 1077-1092.
- Chiaradia, A., Hillier, B., Barnes, Y., & Schwander, C. (2009). Residential Property Value Patterns in London: Space Syntax spatial Analysis, In Proceedings of the 7th International Space Syntax Symposium, Stockholm: KTH.
- Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association*, 83(403), 596-610.
- Dubin, R. A. (1988). Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *The Review of Economics and Statistics*, 466-474.
- Enström, R., Netzell, O., (2008). Can space syntax help us in understanding the intraurban office rent pattern? Accessibility and rents in downtown Stockholm. *The Journal of Real Estate Finance and Economics*, 36(3), pp.289-305.
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

- Geurs, K., Montis, A., & Reggiani, A. (2015). Recent advances and applications in accessibility modelling. *Computers, environment and urban systems*, 49, 82-85.
- Geurs, K. T., & Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport geography*, 12(2), 127-140.
- Goodman, A. C. (1978). Hedonic prices, price indices and housing markets. *Journal of Urban Economics*, 5(4), 471-484.
- Goodman, A. C., Thibodeau, T. G., (1998). Housing market segmentation. *Journal of housing economics*, 7(2), 121-143.
- Goodman, A. C., & Thibodeau, T. G. (2007). The spatial proximity of metropolitan area housing submarkets. *Real Estate Economics*, 35(2), 209-232.
- Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of planners*, 25(2), 73-76.
- Haurin, D. R., & Brasington, D. (1996). School quality and real house prices: Inter-and intrametropolitan effects. *Journal of Housing Economics*, 5(4), 351-368.
- Henneberry, J. (1998). Transport investment and house prices. *Journal of Property Valuation and Investment*, 16(2), 144-158.
- Helbich, M., Brunauer, W., Hagenauer, J., & Leitner, M. (2013). Data-driven regionalization of housing markets. *Annals of the Association of American Geographers*, 103(4), 871-889.
- Helbich, M., Brunauer, W., Vaz, E., & Nijkamp, P. (2013). Spatial heterogeneity in hedonic house price models: the case of Austria. *Urban Studies*, 0042098013492234.
- Hess, P. M., Moudon, A. V., & Logsdon, M. G. (2001). Measuring land use patterns for transportation research. *Transportation research record*, (1780), 17-24.
- Hui, E., Zhong, J., & Yu, K. (2014). Heterogeneity in Spatial Correlation and Influential Factors on Property Prices of Submarkets Categorized by Urban Dwelling Spaces. *Journal of Urban Planning and Development*, 10.1061/(ASCE)UP.1943-5444.0000270, 04014047.
- Hwang, S., & Thill, J. C. (2009). Delineating urban housing submarkets with fuzzy clustering. *Environment and Planning B: Planning and Design*, 36(5), 865-882.
- Hillier, B. (1996). Cities as movement economies. *Urban Design International*, 1(1), 41-60.
- Hillier, B. (2003). The architectures of seeing and going: or, are cities shaped by bodies or minds? And is there a syntax of spatial cognition?.
- Hillier, B., & Hanson, J. (1984). *The social logic of space*. Cambridge university press.
- Hillier, B., & Hanson, J. (1997). The reasoning art. In *Proceeding of the First International Space Syntax Symposium*. London.
- Hillier, B., & Iida, S. (2005). Network and psychological effects in urban movement. In *Spatial information theory* (pp. 475-490). Springer Berlin Heidelberg.
- Hillier, B., Yang, T., & Turner, A. (2012). Normalising least angle choice in Depthmap-and how it opens up new perspectives on the global and local analysis of city space. *Journal of Space syntax*, 3(2), 155-193.
- Hillier, B., & Vaughan, L. (2007). The city as one thing. *Progress in Planning*, 67(3), 205-230.
- Jacobs, J. (1961). *The Death and Life of Great American Cities*. Random House Digital, Inc.
- Jim, C. Y., & Chen, W. Y. (2006). Impacts of urban environmental elements on residential housing prices in Guangzhou (China). *Landscape and Urban Planning*, 78(4), 422-434.
- Karlström, A., & Mattsson, L. G. (2009). Place, space syntax and attraction-accessibility. *Space Syntax Sempozyumu*, 5, 104(1).

- Kelejian, H.H., Prucha, I.R., (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1), pp.99-121.
- Kim, Y. O., & Penn, A. (2004). Linking the spatial syntax of cognitive maps to the spatial syntax of the environment. *Environment and Behavior*, 36(4), 483-504.
- Kroll, C. A., & Cray, A. F. (2010). Hedonic Valuation of Residential Resource Efficiency Variables. *A Review of the Literature, Berkeley: University of California, Berkeley*.
- Kwan, M. P. (2000). Analysis of human spatial behavior in a GIS environment: Recent developments and future prospects. *Journal of Geographical Systems*, 2(1), 85-90.
- Lancaster, K. J. (1966). A new approach to consumer theory. *The journal of political economy*, 132-157.
- Law, S., (2017). Defining Street-based Local Area and measuring its effect on house price using a hedonic price approach: The case study of Metropolitan London. *Cities*, 60, pp.166-179.
- Law, S., Chiaradia, A., Schwander, C., (2012), January. Towards a multimodal space syntax analysis: A case study of the London street and underground network. In *Published in Proceedings of the 8th International Space Syntax Symposium, Santiago de Chile, Chile* (pp. 3-6).
- Law, S., Karimi, K., Penn, A., & Chiaradia, A. (2013). Measuring the influence of spatial configuration of the housing market in metropolitan London. In *Proceedings of the Ninth International Space Syntax Symposium, Seoul*.
- Law, S., Karimi, K., & Penn, A., (2015). An empirical study on applying community detection methods in defining spatial housing submarkets in London. 10th Space Syntax Symposium. London, UCL.
- Lu, B., Charlton, M., & Fotheringham, A. S. (2011). Geographically weighted regression using a non-Euclidean distance metric with a study on London house price data. *Procedia Environmental Sciences*, 7, 92-97.
- Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science*, 28(4), 660-681.
- Lu, B., Charlton, M., Brunsdon, C., Harris, P., (2016). The Minkowski approach for choosing the distance metric in geographically weighted regression. *International Journal of Geographical Information Science*, 30(2), pp.351-368.
- Luttik, J. (2000). The value of trees, water and open space as reflected by house prices in the Netherlands. *Landscape and urban planning*, 48(3), 161-167.
- Marcus, L. (2010). Spatial capital. *The Journal of Space Syntax*, 1(1), pp.30-40.
- Matthews, J.W., Turnbull, G.K., (2007). Neighborhood street layout and property value: The interaction of accessibility and land use mix. *The journal of real estate finance and economics*, 35(2), pp.111-141.
- Mok, H. M., Chan, P. P., & Cho, Y. S. (1995). A hedonic price model for private properties in Hong Kong. *The Journal of Real Estate Finance and Economics*, 10(1), 37-48.
- Nakaya, T., Fotheringham, A.S., Brunsdon, C. and Charlton, M., (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in medicine*, 24(17), pp.2695-2717.
- Okabe, A., & Sugihara, K. (2012). *Spatial analysis along networks: statistical and computational methods*. John Wiley & Sons.

- Ozbil, A., Peponis, J., & Stone, B. (2011). Understanding the link between street connectivity, land use and pedestrian flows. *Urban Design International*, 16(2), 125-141.
- Palm, R. (1978). Spatial segmentation of the urban housing market. *Economic Geography*, 210-221.
- Porta, S., Latora, V., Wang, F., Rueda, S., Strano, E., Scellato, S., & Latora, L. (2012). Street centrality and the location of economic activities in Barcelona. *Urban Studies*, 49(7), 1471-1488.
- Ratti, C., (2004). Space syntax: some inconsistencies. *Environment and Planning B: Planning and Design*, 31(4), 487-499.
- Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1), 34-55.
- Scoppa, M. D., & Peponis, J. (2015). Distributed attraction: the effects of street network connectivity upon the distribution of retail frontage in the City of Buenos Aires. *Environment and Planning B: Planning and Design*, 42(2), 354-378.
- Sevtsuk, A. (2014). Location and Agglomeration The Distribution of Retail and Food Businesses in Dense Urban Environments. *Journal of Planning Education and Research*, 34(4), 374-393.
- Sevtsuk, A., & Mekonnen, M. (2012). Urban network analysis. *Revue internationale de géomatique-n*, 287, 305.
- Smersh, G. T., & Smith, M. T. (2000). Accessibility changes and urban house price appreciation: a constrained optimization approach to determining distance effects. *Journal of Housing Economics*, 9(3), 187-196.
- Shen, Y. & Karimi, K., (2015). Understanding the roles of urban configuration on spatial heterogeneity and submarket regionalisation of house price pattern in a mix-scale hedonic model. In *Proceedings of the 10th Space Syntax Symposium*, London, University College London.
- Shen, Y., & Karimi, K. (2016). Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities*, 55, 9-21.
- Song, Y., & Knaap, G. J. (2003). New urbanism and housing values: a disaggregate assessment. *Journal of Urban Economics*, 54(2), 218-238.
- Ståhle, A., Marcus, L. & Karlström, A., (2005). Place Syntax: Geographic accessibility with axial lines in GIS. In *Fifth international space syntax symposium* (pp. 131-144). Techne Press.
- Stewart, J. Q. (1948). Demographic gravitation: evidence and applications. *Sociometry*, 11(1/2), 31-58.
- Tallon, A. (2013). *Urban Regeneration in the UK*. Routledge.
- Turner, A. (2001a). Angular analysis. In *Proceedings of the 3rd Space Syntax Symposium*, Atlanta, University of Michigan.
- Turner, A. (2001b). A program to perform visibility graph analysis. In *Proceedings of the 3rd Space Syntax Symposium*, Atlanta, University of Michigan.
- Xiao, Y., Webster, C., & Orford, S. (2016). Identifying house price effects of changes in urban street configuration: An empirical study in Nanjing, China. *Urban Studies*, 53(1), 112-131.
- Watkins, C., (1999). Property valuation and the structure of urban housing markets. *Journal of Property Investment & Finance*, 17(2), pp.157-175.

- Webster, C. (2010). Pricing accessibility: Urban morphology, design and missing markets. *Progress in Planning*, 73(2), 77-111.
- Wei, C. H., & Qi, F. (2012). On the estimation and testing of mixed geographically weighted regression models. *Economic Modelling*, 29(6), 2615-2620.
- Wegener, M. (2004). Overview of land-use transport models. *Handbook of transport geography and spatial systems*, 5, 127-146.
- Wen, H., Zhang, Y., & Zhang, L. (2014). Do educational facilities affect housing price? An empirical study in Hangzhou, China. *Habitat International*, 42, 155-163.
- Wu, C., Sharma, R., (2012). Housing submarket classification: The role of spatial contiguity. *Applied Geography*, 32(2), 746-756.
- Wu, J., Gyourko, J., & Deng, Y. (2012). Evaluating conditions in major Chinese housing markets. *Regional Science and Urban Economics*, 42(3), 531-543.