Available online at www.humg.edu.vn

ESASGD 2016

GIS-IDEAS (2016)

Conference Title: International Conference on GeoInformatics for Spatial-Infrastructure Development in Earth & Allied Sciences (GIS-IDEAS)

Integrating GIS and Spatial Statistics to study Hanoi's Urban Structure and Housing Market

Pham Khanh Chi*, Tran Hung

GeoViet Consulting Co., Ltd., Hanoi, Vietnam

Abstract

This paper studies the urban structure of the Hanoi City through exploring and analyzing spatial patterns and spatial segmentation of the Hanoi's housing prices using GIS and spatial statistics based on the dataset from the 2014 house price survey of 1000 households located within the 4th Ring Road. Combined with spatial variables (determinants) derived from the Hanoi GIS database, the computed Moran I and Getis-Ord Gi* confirms the clustered urban structure and the emergence of a number of urban centers – new 'status' poles as a result of Hanoi's rapid expansion. Further, the nature of spatial dependence and their influence in residential market behavior and outcomes are explored revealing the geographic segmentation of Hanoi's housing market. Geographically Weighted Regression (GWR) modeling is then, adopted to account for collinearity and spatial heterogeneity, which significantly improves the house price estimation as compared to traditional hedonic modeling. The paper also discusses how GIS with powerful data integration, mapping and visualization functionality can aid with empirical research investigations. Integration of GIS and spatial statistics can provide powerful spatial integrated platform and analytical framework for a spatial decision-support-system (SDSS), for example, in support of housing development programs in Hanoi and other large cities of Vietnam.

Keywords: Urban structure, GIS, housing sub-markets, spatial segmentation, hedonic modeling, spatial statistics and SDSS.

1. Introduction

As housing makes up the major part of the land of cities, the distribution and development of housing in urban areas determine the structure of the cities as the ways in which cities are formed and have evolved. Much residential location theory explains residential location patterns using polar structures, where one or several poles represent the highest points of certain kinds of social status held by the residential population relating to notions of wealth, political power, business, culture, ethnicity, education, etc. The geographic location of a house determines access to employment, shopping, and recreation; neighbors and neighborhood characteristics; proximity to environmental amenities; and the level and quality of public services… (Can, 1998). Its role can be examined in two inter-related ways – through *adjacency effects* and *neighborhood characteristics*. Differences in household incomes and preferences, along with systematic spatial variations across neighborhoods, lead to spatial segmentation. Therefore, the study of urban housing and its relationship with urban structure requires application of appropriate techniques of spatial statistics and spatial econometrics in addressing the two-dimensional nature of spatial interaction and the prevalence of spatial dependence in the cross-sectional housing data. This paper has a two-fold purpose. First, this paper aims to explore and analyze spatial patterns and spatial segmentation of the Hanoi's housing prices and its key determinants market based on GIS and spatial statistics.

* Corresponding author. Tel.: +84 901 776685.
E-mail address: chi.geoviet@gmail.com.

And at the same time, incorporating spatial heterogeneity into traditional hedonic equations, it attempts to further improve the prediction accuracy of housing prices in a GIS-based streamlined spatial analytical framework, which can be arguably served as building blocks of a spatial decision-support-system (SDSS) for the implementation of housing development programs in Hanoi and other large cities of Vietnam.

2. Hanoi's Housing Data and Methods

2.1. Background: SQTO-based research and data

In the Spring 2014, a research program was initiated by the Vinaconex R&D JSC with funding from the Ministry of Construction aiming at establishing a methodology of establishing the House Price Index for the nascent housing market of Vietnam. Based on the status-quality trade off (SQTO) theory developed by Phe and Wakely (2000), that research is designed to extensively explore tangible and intangible components of the value of residential properties, which would then form a framework for regular data gathering of the housing market. Specifically, the SQTO seeks to explain urban development dynamics and residential location choices using housing status to describe the social desirability of housing in a particular area and dwelling quality to describe physical characteristics such as floor area, number of bathrooms, number of storeys, etc.

A detailed house price survey of nearly 1,000 households (located within the 4th Ring Road, see Figure 1) was undertaken to support the SQTO analyses in Hanoi, Vietnam. The target population was private households and previously public houses transferred to sitting tenants were also included. The survey used a 400m x 400m sampling grid and collected data for 245 variables, reflecting both tangible and intangible housing attributes. This data has been used to segment housing markets and to identify significant predictor variables using a stepwise hedonic regression approach (Phe *et al.*, 2015) for the dependent variable HPRICVND (the house price in millions of Vietnamese Dong). The most notable feature of the big picture as seen from the tentative analysis of the survey results is that there is no unified housing market for the whole surveyed area, the fact that gives legitimacy to a statement that there is no unified housing market for the whole city of Hanoi. As complex organism, the market of housing in cities is deeply local and segregated: The whole city represents a series of inter-connected submarkets, either based on quality levels, environmental, human capital characteristics or geographical locations. The 30 variables originally selected as input to this study's regression analyses were reduced to 6 most statistically–significant variables as being those most related to price-forming influences according to the SQTO theory (Table 1).

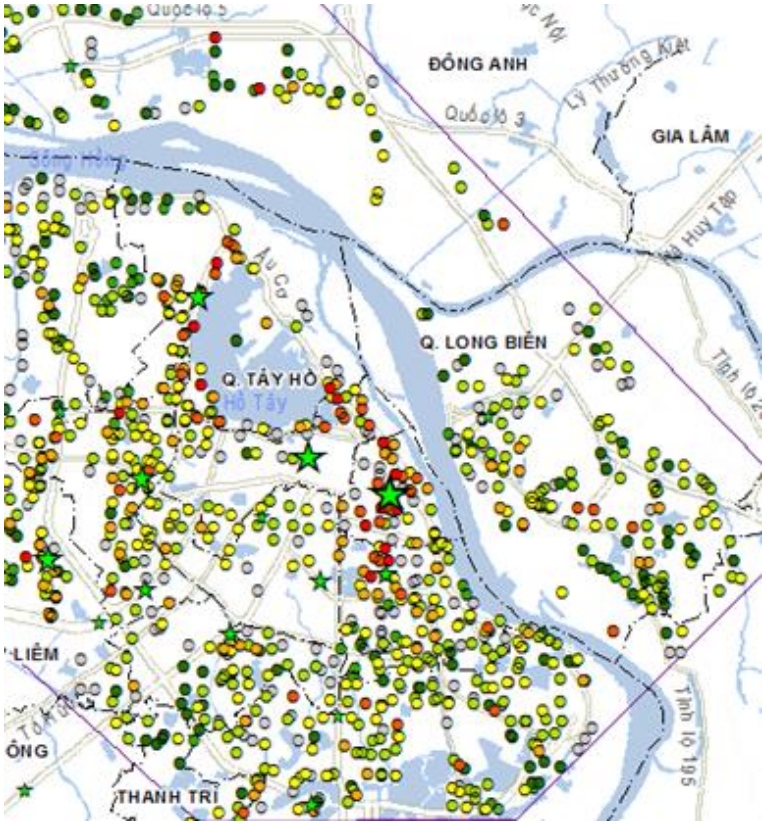

Fig. 1. Location of 1000 house sampling points colored with price level (VND/m^2)

Table 1. Household price and significant predictor variables

Variable	Description	Type
HPRICVND	Price of house in Millions of Vietnamese Dong (grouped into 5 levels)	Dependent
HPRICPSM	Price of house in Millions of Vietnamese Dong per square meters	Dependent
GFA	Total floor area (incl. mezzanine) (m^2)	Predictor - Tangible
PLOTAREA	Total plot area (m^2)	Predictor - Tangible
CENTDISR	Measured distance to Hanoi Centre (m)	Predictor - Intangible
DISCENDI	Perceived travel time to the Centre District (min)	Predictor - Intangible
PLUMBING	Plumbing Quality (Good, Fair, Poor)	Predictor - Intangible
EDYEARS	Time in education of the interviewees (years)	Predictor - Intangible

2.2. GIS-based Spatial Statistics Methods

To further findings of the SQTO-based classical hedonic analysis above, GIS-based spatial analysis and spatial statistics methods are adopted for this study using ArcGIS Desktop software. A multi-stage analysis was undertaken to understand the Hanoi's urban structure and spatial patterns of its housing market consisting of the following stages:

- GIS-based construction of spatial variables: Survey data are mapped as a point layer into the Hanoi GIS database, which is a comprehensive set of various layers such as basemap, transportation, urban physical and social infrastructures, land use, administrative boundary, emerging urban (residential, commercial or industrial) centers and populations, etc. Using GIS data integration capability, a number of derivative exploratory spatial variables are generated including population density, road density and distance-based variables (e.g., computation of distance from a house to the city center or a local / neighborhood centers…). This is used to visualize and map exploratory analysis results and provide explicit information on spatial relationships (generating spatial weight matrix for spatial hedonic models, Anselin 1998).

- Exploring spatial dependence on housing market with geostatistical approach and analyzing spatial patterns of housing markets based on exploratory techniques of spatial statistics of potential housing markets' determinants / components. The spatial interpolation via kriging method would then prepare map of spatial distribution of housing prices in geographical space, which can be visualized alongside with kriging surfaces of population & road density to see the relationship as well as analyzed structure of spatial variability as semivariance function of range distance. The global (Moran I) and local (Getis-Ord Gi*) measures of spatial autocorrelation, on the other hand, of house price and its potential determinants can be calculated and mapped to explore the clustering spatial patterns and to identify local hot spots (e.g., local neighborhoods with high housing prices), cold spots (e.g., neighborhoods with low housing prices) or spatial outliers.

- Modeling spatial relationships of housing markets. Following classical hedonic model, OLS residuals are examined for its collinearity and non-stationarity, where model diagnostics of strong spatial autocorrelation of potential predictors leads to the application of geographically weighted regression (GWR) model for the Hanoi's cross-sectional housing data with better specification tests for spatial dependence and more efficient model estimation. In overview, a GWR approach is one that uses a moving kernel that passes through the study area. At each location being considered, data under the kernel are used to make a local calculation of some kind, such as a regression. The data are weighted by their distance to the kernel centre and in this way GWR approaches construct a series of models at discrete locations in the study area (Comber *et al.*, 2016).

3. Empirical Results

3.1. Hanoi's urban development, urban structure and spatial pattern of housing market

After long and turbulent years of war and centralized economic policy, Hanoi is gradually becoming a fast-growing metropolis with a substantial improvement to the transport infrastructure and the mushrooming of commercial development and construction of high-rise office / commercial blocks near the center. As most cities, Hanoi conforms to a polar structure, in which one or several poles represent the highest points of certain kinds of social status, recognized by a given proportion of the population. In 1990x, the city remains basically monocentric, which can be explained by the perceived importance of the traditional downtown area ("Khu Pho co" and French Quarter), as a commercial quarter and especially as a cultural focus. There are few indications that new sub-centers are emerging in terms of population density in 2009 (Figure 2b), which is becoming more

obvious by 2014 in terms of road density and house price (Figure 2a). The inner city area is expanding and new urban centers emerge in northern area of Ho Tay, western area of My Dinh and southern area of Linh Dam. These maps also show that spatial distribution of house price quite resembles the spatial patterns of population and road infrastructure distribution.

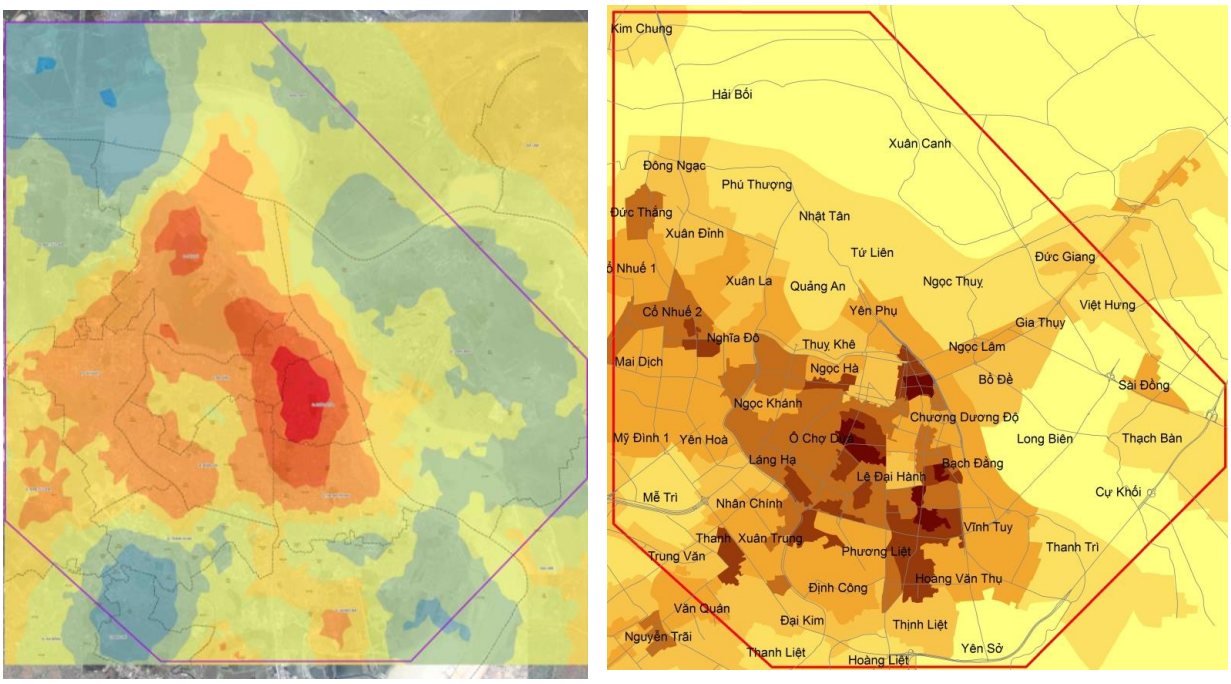

Fig. 2. (a) Kriging of the house price 2014 per sq. m (mil VND/m^2); (b) Map of commune-level population density 2009

The spatial patterns of house price and its potential determinants are quantified by the global and local measures of spatial autocorrelation. The computed global Moran indices with incremental distances show strong, significant clustered spatial patterns of population density in 2009 (Moran I =0.88 at 0.5km reducing to 0.35 at 6km with p-value < 0.001) and road density in 2014 (Moran I = 0.55 – 0.2 with p-value < 0.01). The z-score of the Moran I reaches maximum for both variables at 6km distance, can be explained as the perceived size of the inner city area of Hanoi, where the neighborhood and adjacency ("pull") effects appear particularly strong. However, the computation of spatial autocorrelation measure for house price shows z-score = 0.61 with p-value = 0.54 (> 0.05), meaning the spatial dependence of house price in close-by neighborhoods is statistically insignificant. The price of a house seems not significantly affected by the price of the house nearby in small neighborhoods, perhaps depends on other factors such as building quality, accessibility, etc. This combined with the fact that potential determinants are collinear and highly clustered can suggest the application of geographically weighted regression (GWR) model as the most appropriate as compared with other spatial regression model such as spatial autoregressive lag (SAR) model or spatial error (SEM) model.

Further, the spatial clustering patterns of the housing market in Hanoi and its determinants are explored using local Getis-Ord G* statistics. The computed z-score > 2.0 (with p-value < 0.05) for EDYEARS discovers few local hot spots around My Dinh and Thanh Xuan areas (i.e., local neighborhoods of new urban quarters with highly-educated house owners) and cold spots around Minh Khai – Trương Dinh or Gia Lâm, Dong Anh areas (i.e., local neighborhoods of ancient villages with low-educated house owners). Similarly, the computed z-score with p-value < 0.1 for PLUMBING discovers a local hot spot of poor water supply at the few communes of Hoang Mai district, which certainly affects the residential choice of people and consequently the house price. All these measures show the existence of housing spatial submarkets segmentation as well as spatial interactions among house price's potential determinants' that may describe impact on the residents in terms of their market behaviors (Chi and Hung, 2016).

3.2. GIS-based spatial modeling of Hanoi's housing market

The results of OLS regression using ArcGIS Desktop 10.2 confirms significant variables (same as those shown in Table 1 from the analysis by *Phe et al.* in 2015) and shows that house price is significantly explained by variability of house physical properties (GFA and PLOTAREA) and its spatial locations (CENTDISR and DISCENDI). Overall, the model is statistically valid ($R^2_{Adjusted}$ = 0.31 with F-Statistic = 84.33; p < 0.00001).

However, the random test for OLS residual shows z-score = 13.23 with p-value < 0.00001 indicating the inadequacy of the OLS model, possibly due to collinearity and spatial dependence of predictor variables (as explored above).

A basic GWR is run using Spatial Statistics Tool in ArcGIS Desktop 10.2 on the variables identified as being significant predictors of house price in the OLS regression above. Overall, the resulted GWR model shows much improvement in the model-fit ($R^2_{Adjusted}$ = 0.54; AIC = 14,140.64) with model residual tested as random (z-score = -0.31; p-value = 0.76). Considerable spatial variation is observed in many of the derived coefficient estimates, suggesting that spatial structure in the data (i.e., non-stationary relationships between house price and housing attributes) is present. However for a reliable assessment of this observed non-stationarity, collinearity needs to be assessed and addressed for all of the localised regressions of the GWR model. The GWR collinearity diagnostics on local correlations amongst pairs of predictors show local Conditional Number (CN) varying between 6.45 and 29.64, indicating no evidence of significant local collinearity. In addition, the computed local R^2 also shows a good fit (of more than 50%) for most of the housing data points, except for some points in the southern suburb areas of Linh Dam and Vinh Tuy.

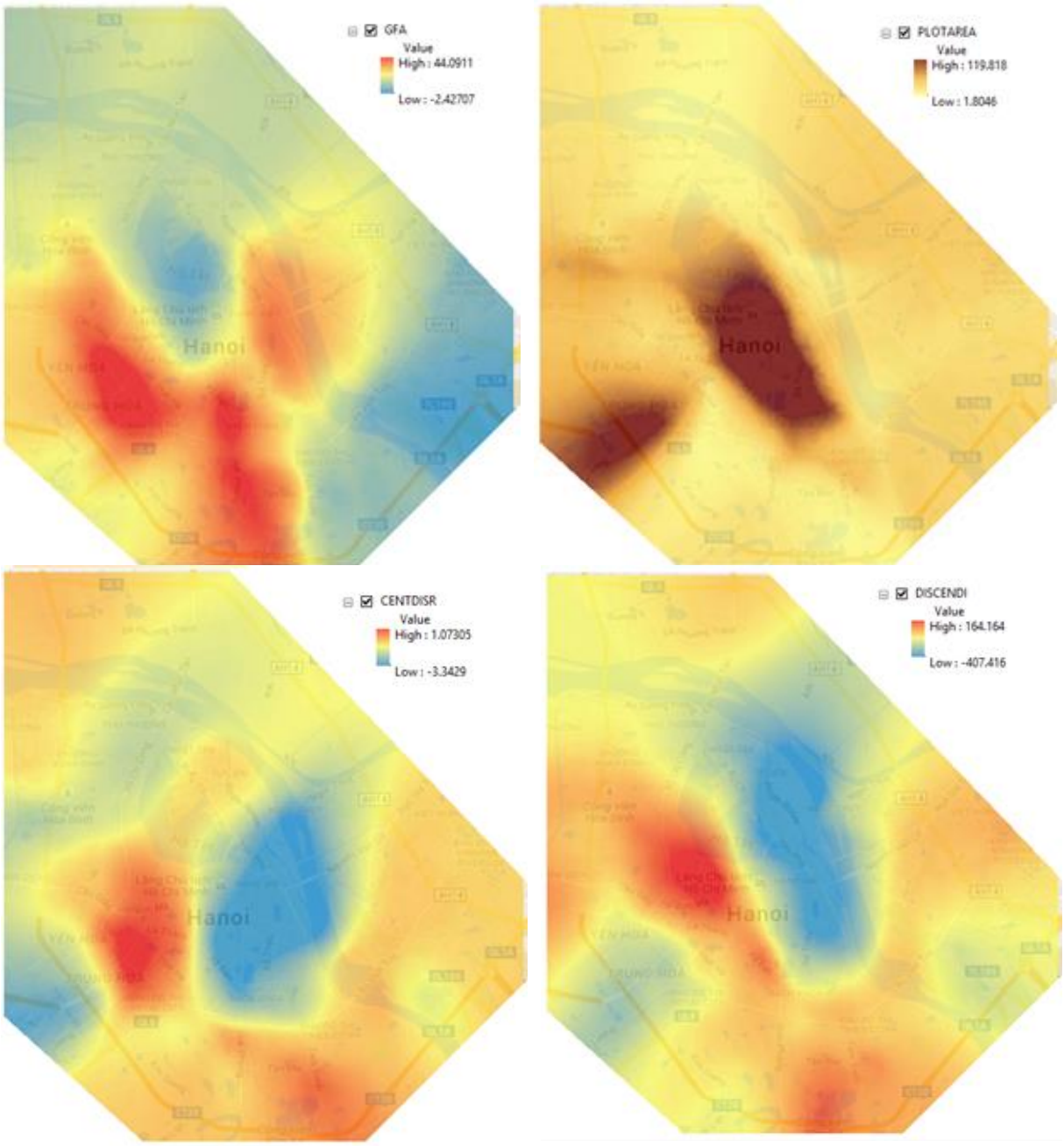

Fig. 3. Map of contribution of 4 most significant predictor variables on house price in Hanoi (a) GFA; (b) PLOTAREA; (c) CENTDISR; and (d) DISCENDI

The spatially distributed variables are combined using the spatially distributed coefficient estimates arising from the locally compensated GWR in order to construct the model of house price. The predicted house price is estimated a linear function of its spatial locations (CENTDISR and DISCENDI) and house's physical properties (GFA and PLOTAREA) for each of data points which are clearly varying from neighborhood to neighborhood in Hanoi. Figure 3 shows the spatial distribution of coefficient estimates indicating the importance levels or contribution of each of those predictor variables in determining the house price. Again, the resulted GWR model confirms clear segmentation of Hanoi's housing market with different spatial submarkets from city center to suburb areas with quite different residential choices / market behaviors. This shows distinct patterns of predicted house prices reflecting not only the downtown poles that one would expect but also distinct poles of different levels of house price in areas around the centre (e.g., Ho Tay, My Dinh, Thanh Xuan, Ha Dong, Linh Dam...). In the city core area, the house price largely depends on PLOTAREA – quickly increases with increasing land plot area (Figure 3b). This pattern is sharply changing when moving from center toward immediate outer areas, where the house price is much more explained by the GFA (Figure 3a). This makes sense as with bigger land plots available, residents in new urban development area invest more on quality of newly built housing. Related to intangible components, the house price appears to less depend on the distance to the city centre (CENTDISR - i.e., Ho Guom) for the houses within the city core area (Figure 3c), while the price is quickly decreases with the perceived travel time to the Centre District (DISCENDI, Figure 3d) for the houses in the new urban development neighborhoods – emerging urban growth poles (e.g., My Dinh, Ha Dong…).

4. Concluding Remarks and Next Steps

The analytical results in this paper demonstrated that the integration of GIS and spatial statistics helps exploring and analyzing the urban structure and spatial patterns of Hanoi's housing market with its potential determinants. Especially, addressing the spatial dependence issue and utilizing the spatial information (implicitly inherited within the urban and housing data), a combination of exploratory and confirmatory spatial statistical approaches have revealed the emergence of a number of urban centers – new 'status' poles as a result of Hanoi's rapid expansion with associated spatially housing sub-markets. The residential market behaviors are taken into account locally, which is resulted in improved house price estimation through GWR modeling. Using a GWR approach more generally has advantages over other spatially explicit approaches because it is fundamentally concerned with understanding the spatial variation in relationships between the response and predictor variables and in so doing understanding how model predictions are arrived at. The findings also show some indication on the changing urban landscape with urban housing market due to the rapid development of the Hanoi metropolis.

In terms of analytical methods, future work is needed to improve the findings in a number of directions, such as application of mixed GWR approach, use of network-based distance matrix and include more predictor variables for GW analyses. In terms of data, future work will need to collect additional data points, repeat data collection over time combined with development of dynamic and continuous measures related to house price (e.g., through real estate transition data, social media, etc.) in order to develop a continuous monitoring of the housing market. With the powerful data integration, mapping and visualization functionality (demonstrated in this paper), GIS will provide comprehensive platform for spatial analysis research in exploring and modeling of geographically referenced data. A prototype of Web-based GIS system with comprehensive urban database of Hanoi is developed and operated by the GeoViet Consulting Co. Ltd. at http://360.geovietmap.com can well serve as spatial integrated platform and analytical framework for a future spatial decision-support-system (SDSS) in support of housing development programs in fast-growing Hanoi.

References

Anselin L., 1998. GIS Research Infrastructure for Spatial Analysis of Real Estate Markets. Journal of Housing Research, Vol. 9, Issue 1 pp. 113 - 133

Can, A., 1998. GIS and Spatial Analysis of Housing and Mortgage Markets. Journal of Housing Research, Vol. 9, Issue 1 pp. 61 – 86

Chi P. K., Hung T., 2016. GIS và thống kê không gian trong nghiên cứu cấu trúc không gian đô thị (Part II & III in Vietnamese) - http://geoviet.vn/goc-ky-thuat/vn/

Comber A., Phe H. H., Hung T., Chi P.K., 2016. Modelling hedonic house price in Hanoi: SQTO theory with localised regressions adapted for collinear effects. Paper submitted to Urban Studies.

Phe H. H. *et al.*, 2015. Status- Quality Trade-off Theory and Housing Policy: The segmentation of Hanoi City's housing market and the development challenges. Paper presented for the RMIT Seminar Series.

Phe, H. H., Wakely, P., 2000. Status, quality and the other trade-off: Towards a new theory of urban residential locations. Urban Studies, 37(1), pp. 7 – 35.