DELINEATION OF GEOSPATIAL RESIDENTIAL REAL ESTATE SUBMARKET BOUNDARIES

TONY LOCKWOOD
University of South Australia

ABSTRACT

The objective of this study was to develop and assess a methodology for deriving the geospatial dimensions of residential real estate submarkets. This was achieved through the analysis of marketplace behaviour with respect to the underlying dimensions of the residential real estate geography. Importantly, the methodology makes no prior assumptions about where the spatial boundaries might be as they were empirically derived from the data alone. The contribution made lies in the use of the principal components of the underlying real estate geography as linearly independent variables in a geographically weighted hedonic model. This allows changing patterns in the parameter estimates of the principal components to define the geospatial submarket boundaries.

INTRODUCTION

This study was based on two premises. The first was that when a dwelling is sold, the commodity traded is a piece of residential real estate geography (REG) comprising a complex bundle of both spatial and structural attributes. The second was the recognition in the methodology of the importance of 'location' in real estate analysis.

The price of the various components comprising the real estate geography varies across geographical space in a continuous fashion. It is the variability of the price patterns of these components across space that is defined, in this study, to be the geospatial submarket identifier.

By using principal component analysis (PCA) to quantify the REG as the dimensions of a large number of both structural and spatial attributes, two benefits are derived. First, a large amount of data can be accounted for in a relatively small number of principal components. Second, these principal components can then be used as linearly independent surrogate property characteristics in a geographically weighted hedonic regression model providing the opportunity to use the variation in their parameter estimates to indicate submarket boundaries.

There are many ways of defining residential submarkets and this is one of the identified problems with current submarket research (Adair et al. 1996b). However, there appears to be an emerging sense from the literature that a workable delineation of submarkets should be based on:

- Data analysis and not on *a priori* knowledge (Bourassa et al. 1999).
- A recognition that a submarket may be composed of both structural and spatial components simultaneously (Adair et al. 1996a; Watkins 2001).
- The underlying residential structure of the study area (Maclennan & Tu 1996).
- The concept of substitutability constrained by price (Bourassa et al. 2003; Grigsby 1963). Submarkets are an economic entity and therefore should be defined with reference to the marketplace (Pryce 2004).

The definition of a residential spatial submarket adopted in this study accommodates all these elements, thus adhering to residential submarket theory, and is termed a *geospatial residential submarket.* It was defined for this study as:

> A geographic area within which the market prices of the individual components of the underlying residential real estate geography (REG) have a predefined homogeneous pattern.

The submarket identifier is price. In particular, it is the price pattern of the continuously changing individual hedonic parameter estimates of the underlying real estate geography (REG) principal components across geographical space. The determination of a particular set of boundaries delineated along that continuous surface is defined by the user for a particular purpose.

The importance of understanding the residential submarket structure is present in three broad areas of residential property analysis. Firstly, in the formulation of housing policy. As Pryce and Evans (2007) point out, people are administered in spatial units and funds are allocated along these spatial lines. Therefore to understand the spatial structure of the market will help better align policy administration with market place structures. Also, both Maclennan & Tu (1996) and Meen & Meen (2003) stress the importance of policy-makers using an understanding of market structures in the policy formulation process. Secondly, that planning interventions should be based on a market view of the housing sector (Barker 2004; Bates 2006) if they are to be appropriate to the development process. Thirdly, the mass appraisal function used to support a property taxation policy in many jurisdictions needs to understand spatial market structures to better derive individual property market values (McCluskey et al. 2002; Watkins 1999).

Potentially, this presents an opportunity for the mass appraisal process, as part of the broader Land Administration System, to provide an analysis of the real estate market structure needed to support wider government housing policy issues. This increases the importance of the mass appraisal process beyond the sole traditional provision of a tax base to one of supporting the wider functions of a spatially enabled Land Administration System.

LITERATURE REVIEW

The use of various submarket identifiers and analytical techniques in defining residential real estate submarkets has been well reviewed by several authors (Kauko 2000; Borst & McCluskey 2008; Watkins 2001). Although there has been general agreement that submarkets exist, there is not such consensus as to their definition and Watkins offers this as a possible reason for the lack of using submarket structure in housing market analysis. For the purpose of this study, the various types of analytical techniques may be regarded as falling into three broad groups.

The first group sought to confirm that certain *a priori* spatial and or structural boundaries were better than none at all in terms of establishing more homogeneous housing markets. This was not an attempt to find the optimum solution, but rather to confirm that smaller submarkets did exist and were each more homogenous than the global market. Various submarkets are selected *a priori*, to be either structural and or spatial and then tested using hedonic housing price differences to see if the results were significantly improved in the smaller submarkets (Adair et al. 1996b). The advantage of this method is that submarkets can be quickly identified in a cost-effective manner and if market difference is significant, they may be used to generate, in the case of the mass appraisal function, more reliable predictive value models than if submarkets were not used (Adair et al. 1996b; Goodman & Thibodeau 2007). There is also concern expressed in the literature that if hedonic pricing models do not recognise the existence of submarkets (however delineated), they may be subject to aggregation bias (Watkins 1999).

The second group of analytical techniques recognises the importance of location in residential real estate value and attempts to include it as part of the housing hedonic price modelling process, thus removing the need to be aware of submarket boundaries (Bourassa et al. 2007; Clapp 2003; Figueroa 1999; Fik et al. 2003; Gallimore et al. 1996; McCluskey et al. 2002; Pryce & Evans 2007; Tu et al. 2007). The attraction of these analytical techniques is that they recognise every property as having a unique location factor and therefore accounting for location on an individual property basis. This correctly identifies the effect due to location as a continuous geographic surface. A more recent approach taken by Borst & McCluskey (2008) use the geographically weighted regression response to signal market segmentation, again making no assumption as to where the spatial delineations may be, instead relying on the data alone for its determination.

However, the disadvantage of the these two groups is that they can not explain location in terms of the underlying real estate geography that includes various location attributes such as amenity, accessibility, socio-economic and environmental indicators. In addition, some surface interpolators used to generate the value gradients may not suitably detect the changes that may be exist in reality. For example, as Clapp (2003) points out, there appears to be a potential problem using polynomials of sufficiently high degree to capture the flexible value surface. The disadvantage of the ‘blind’ residual model approach (Figueroa 1999; Gallimore et al. 1996) is the uncertainty as to the degree of model error that may be present in the residual used as a proxy for location.

The third group again focuses on the delineation of submarket boundaries from the data alone, also without the need to assume *a priori* spatial boundaries. These data include product group attributes and consumer group attributes; sometimes used separately or together. In this group, the most common analytical technique appears to be the use of principal component analysis (PCA) to group attributes that represent the dimensions of marketplace. These components can then be represented spatially either using cluster analysis (Bourassa et al. 1999; Bourassa et al. 2003) or using geostatistical interpolation (Cano-Guervos et al. 2003) to give more homogenous districts. The principal components of the structural attributes are termed product groups by Maclennan & Tu (1996) and tested as being significantly different using the accepted test by Schnare and Struyk (1976) and thereby redefined as submarkets. A drawback with the principal component approach is that it is not necessarily related to the current marketplace and therefore difficult to be viewed as an optimal submarket delineation as an economic entity. However, the advantage in the approach of this group is the recognition of the underlying housing structure is deemed an important element to be included in submarket delineation. The use of PCA as an analytical tool to quantify underlying geographical structures is not new. Shevky and Bell (1955) used the technique to demonstrate three broad social constructs of urban residential structure may be presented as socio economic status, familism and ethnicity. Similar factorial ecology studies evolved in Australia during the 1970s and 1980s describing the essence of Australian urban residential structure in similar fashion. This study recognizes an opportunity to express the underlying real estate geography in terms of the market place in order to reveal its market structure.

STUDY AREA

The study area is basically the metropolitan area of Adelaide, capital city of South Australia. With a population of just over 1 million as at the 2001 census, a total of approximately 440,000 residential properties (excluding flats) were used in the study. The study area includes 343 suburbs contained in 19 local government areas.

DATA

The attribute data was collected for each of the approximately 440,000 residential properties in the study area. The variables chosen were based on previous studies that have investigated those property attributes having most appropriately been found to contribute to value. These include studies covering both spatial and structural attributes recognizing that, together, these constitute the REG that is traded in the market place. Accessibility attributes (Kestens et al. 2004), amenity attributes (Chhetri et al. 2006), socio economic (Jackson et al. 2007) and structural attributes (Rossini & Kershaw 2005) have all been shown to influence property value.

Altogether, 50 variables were identified as representing both the structural and spatial attributes of residential property in the study area. These data were from various sources, and as far as possible, taken at a common data. The adopted study date was August 2001, as the 2006 quinquennial Census of Population data for Australia was not available at the time of the study. A summary of data is shown in Table 1.

Table 1: Data characteristics: summary

Variable name	Variable type	Data source
Single dwelling	Boolean (0,1)	Valuer General (VG)
Multiple dwelling	Boolean (0,1)	VG
Home unit	Boolean (0,1)	VG
Rural living (non-primary	Boolean (0,1)	VG
production)	Boolean (0,1)	VG
Dwelling construction – brick	Boolean (0,1)	VG
Dwelling construction – stone	Boolean (0,1)	VG
Dwelling construction – rendered		VG
Dwelling area	Boolean (0,1)	VG
Dwelling condition	Continuous standardised variable	VG
Dwelling added value	Continuous standardised variable	VG
Land (site) area	Continuous standardised variable	VG
	Continuous standardised variable	VG
Road distance of dwelling from:		
◆ GP surgery	Continuous standardised variable	The University of Adelaide
◆ primary school	Continuous standardised variable	The University of Adelaide
◆ secondary school	Continuous standardised variable	The University of Adelaide
◆ major shops	Continuous standardised variable	The University of Adelaide
◆ urban shops	Continuous standardised variable	The University of Adelaide
◆ CBD	Continuous standardised variable	The University of Adelaide
Amenity value	Continuous standardised variable	Department of Environment and Heritage

Variable name	Variable type	Data source
Household size		Australian Bureau of Statistics
◆ Small	Continuous standardised variable	(ABS)
◆ Average	Continuous standardised variable	ABS
◆ Large	Continuous standardised variable	ABS
Household tenure		
◆ Owned	Continuous standardised variable	ABS
◆ Mortgaged	Continuous standardised variable	ABS
◆ Rental	Continuous standardised variable	ABS
Household income		
◆ Low	Continuous standardised variable	ABS
◆ Below average	Continuous standardised variable	ABS
◆ Average	Continuous standardised variable	ABS
◆ Above average	Continuous standardised variable	ABS
◆ High	Continuous standardised variable	ABS
Length of same occupancy		
◆ 1 year	Continuous standardised variable	ABS
◆ 5 years	Continuous standardised variable	ABS
Individual place of birth		
◆ NW Europe	Continuous standardised variable	ABS
◆ SE Europe	Continuous standardised variable	ABS
◆ SE Asia	Continuous standardised variable	ABS
◆ NE Asia	Continuous standardised variable	ABS
◆ Australia	Continuous standardised variable	ABS
Individual status		
◆ Married	Continuous standardised variable	ABS
◆ Sole parent	Continuous standardised variable	ABS
◆ Lone	Continuous standardised variable	ABS
◆ Dependent children	Continuous standardised variable	ABS
Individual age		
◆ 0 to 20 years	Continuous standardised variable	ABS
◆ 21 to 34 years	Continuous standardised variable	ABS
◆ 35 to 54 years	Continuous standardised variable	ABS
◆ 55 to 65 years	Continuous standardised variable	ABS
◆ Greater than 65 years	Continuous standardised variable	ABS
Languages spoken at home		
◆ English only	Continuous standardised variable	ABS
◆ English and another	Continuous standardised variable	ABS
Employment status		
◆ Not in labour force	Continuous standardised variable	ABS
◆ Unemployed	Continuous standardised variable	ABS
Education level		
◆ tertiary	Continuous standardised variable	ABS
Total number of variables	**50**	

Property transaction prices

The transaction price data forms a fundamental data set used in this study. It contains the sale price of approximately 7,000 properties, across the whole study area, which sold between July and September 2001 and are considered (through analysis of the Valuer General) to represent the market at the date of the study, namely August 2001. The corresponding factor scores for each of the principal components, derived in stage 1, for each of the sale properties were added to this data set.

METHODOLOGY

The methodology undertook the study in two stages. The first stage was concerned with quantifying the dimensions of the underlying real estate geography for the whole study area using principal component analysis. The second stage was concerned with using the derived principal components from stage 1 as independent variables in a hedonic geographically weighted regression model.

Stage 1

Derivation of the structure of the REG fulfills one of the elements identified in the introduction as one of the emerging requirement for submarket delineation. Principal Component Analysis (PCA) has been reported in the real estate literature as an appropriate methodology for achieving this and was adopted in this study.

Based on an orthogonal rotation, with the number of components chosen based on a combination of their respective eigenvalues and the screen plot interpretation, factor scores were derived for each component and assigned as independent surrogate property attributes to each of the approximate 440,000 properties within the study area as shown in equation 1.

$$Y_{i} = b_1X_1 + b_2X_2 + \ldots + b_nX_n + \text{error} \qquad (1)$$

where:
Y_i is the i^{th} (1 to n) component
b_i is the factor score coefficient
X_i is the original variable.

However, PCA is a methodology with acknowledged limitations that have to be addressed if results can be confidently accepted. The use of PCA is not an exact science. Judgments have to be made by the researcher throughout the analysis, ranging from the appropriate data to be used, to the appropriate rotation and interpretation methods to be employed. This will always be a caveat on the final results. However, subject to this, if the analyst is faithful to the research objectives and the results ‘make sense’, then the methodology can provide valuable insights and significantly contribute to the research objectives. These limitations, together with the PCA methodology followed in this study, are discussed in more detail by Lockwood & Coffee (2006).

Through better managed land related information, an ever increasing amount of appropriate data describing the REG are becoming available. The advantage of this approach is that it allows a necessarily large number of individual property attributes that describe the complex underlying REG to be incorporated into its description. PCA has the potential to provide a manageable number of statistically independent surrogate property characteristics that represent this complex structure.

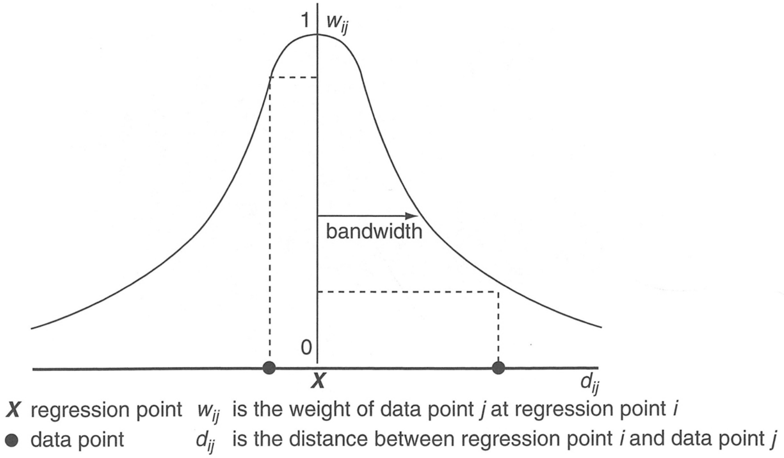
Stage 2

This stage recognises the last emerging element identified in the introduction necessary to delineation submarket boundaries, namely relating the underlying structure to the market place. This was achieved in this study using a hedonic geographically weighted regression model with the surrogate property characteristics quantified in stage 1 representing the underlying structure (REG) as the independent variables.

The use of Geographical Weighted Regression (GWR) is a relatively new technique in the housing market literature. However, it has intuitive appeal as it specifically addresses the question of spatial non-stationarity among the independent variables in the regression model. The essence of geographically weighted regression methodology (Fotheringham et al. 2002) is that it does not assume spatial stationarity of the independent variables and allows local variation to be captured where it exists. This is exactly the reason for using it in this study; indeed, it is the central part of the analysis as it is the difference in the local variation across space which forms the basis to the definition of the geospatial submarket boundaries in this study, allowing their delineation without reference to any existing spatial boundary.

Conceptually, the GWR model may be represented as shown in Figure 1. The spatial kernel weights the sales used according to their respective distances from the regression point. The weighting can be flexible in terms of its functional form. One such form would be a Gaussian function shown in Figure 1, but a bi-square weight function is also used. The bandwidth can be either fixed as the model moves across geographical space, including only those data points which happen to fall within the defined bandwidth or it can be adaptive, allowing the kernel to adapt to include an optimal number of data points or nearest neighbours.

Figure 1: A spatial kernel



The weighting function (w) used in the Gaussian function above where:

$w_{ij} = [\exp(-1/2(d_{ij}/b)^2]$

where d_{ij} is the distance of the data point from the regression point and b is the bandwidth.

Two GWR models are constructed in this study, one performing local regressions at the data points (sale property locations generating 7143 regression points, one for each sale location), and a second performing the local regressions over a regular 300 metre grid generating 9075 regression points. Information from the first GWR model is needed to calibrate the second GWR model. This includes the determination of the optimum bandwidth which gives the optimum number of nearest neighbours used in the adaptive kernel derived from the Akaike Information Criteria (AIC) statistic calculated using GWR software (v.3.0.18 from National University of Ireland) as it moves across the regular grid. This is described in equation 2.

$$\text{AIC} = 2n\log_e(\hat{\sigma}) + \log_e(2\pi) + n\left\{\frac{n + tr(S)}{n - 2 - tr(S)}\right\} \qquad (2)$$

where:
n is the number of observations and $\hat{\sigma}$ is the estimated standard deviation of the error term tr(S) is the trace of the HAT matrix of the GWR.

This optimisation process not only takes into account the goodness of fit of any given model, but also penalises models for having a greater number of parameters so as to favour models that are parsimonious and also fit the data well.

The general GWR model used in this study may be represented by the following form set out in equation 3.

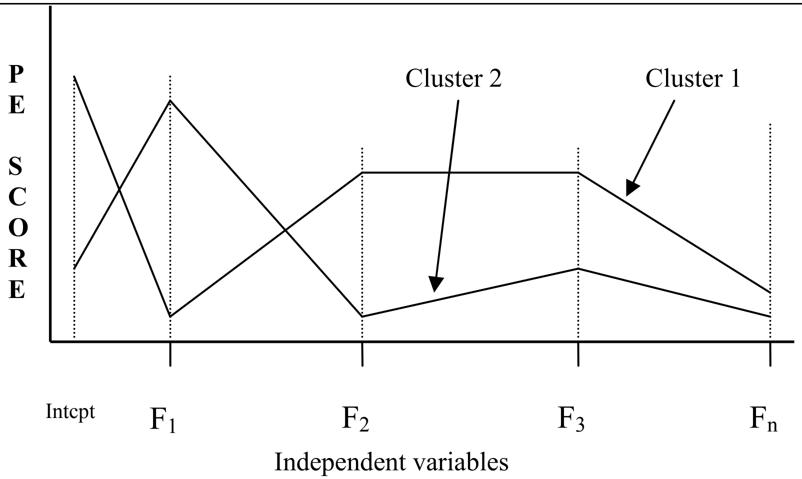
$$\text{Log (sale price)} = \beta_0 (u,v) + \beta_1 (F_1)(u,v) + \ldots + \beta_n (F_n)(u,v) + \text{error}(u,v) \qquad (3)$$

where:
(u,v) are the location coordinates of the sale β_0 to β_n are the parameter estimates
$F_1 \ldots F_n$ are the surrogate property characteristics (principal components from stage 1).

In this study, equation 3 is recalibrated at each point on a regular 300 metre grid set up over the study area, yielding 9075 sets of regression coefficients ($\beta_1 \ldots \beta_n$ in equation 3). The GWR methodology followed in this study employs an adaptive kernel as being more appropriate in the use of real estate data, as sales are not necessarily evenly distributed over the study area (Borst & McCluskey 2008). In reality, the value offered by the regression coefficients is a portion of a continuous surface that can either be approximated by an interpolated grid or a recalibration of the GWR model in a uniform manner on a regular grid over the entire study area. It is considered more realistic to rely on the recalibration than on an interpolated surface in this study, but this may be a topic that requires further investigation. A regular 300 metre grid was considered appropriate for this study, as the accuracy of the data may not warrant a higher density grid.

Cluster analysis was then used to group together the submarket identifiers into appropriate homogeneous geospatial submarkets. There are many different types of clustering algorithms and the important question of choosing the most appropriate for the task must be based on an understanding of their differences. The submarket theory suggests that areas of similarity are areas of homogeneity in which consumers are indifferent to choice based on price. Therefore, the within cluster homogeneity needs to be maximised while the between cluster heterogeneity needs to be maximised. In this case, the price parameter estimates show the market's value of each REG dimension and therefore the closer the price pattern amongst the regression points within a cluster, the more homogeneous the market's interpretation of the REG. The most appropriate clustering algorithm for achieving this was considered to be Ward's hierarchical method, because it minimises the within group variance and maximises the between group variance; closely simulating the submarket theory and is also supported in the submarket literature as being appropriate (Bourassa et al. 1999).

Figure 2: Cluster signatures



Note: Cluster 1 & 2 lines (cluster signatures) represent examples of groups of regression points with similar price patterns.
PE SCORE is the mean parameter estimate score for each principal component (including the intercept - see equation 3) of the group of similar regression points.

The algorithm clusters the like patterns of the parameter estimates (PE) as shown conceptually in Figure 2. For each resulting cluster, the factors with strongest market influence on the formation of the cluster can be seen in both the magnitude and direction.

The question of interpretability of parameter estimates in this manner has been raised in the literature (Borst & McCluskey 2008), and of more particular interest for data in this study, even with independent variables that were derived as principal components (Wheeler & Tiefelsdorf 2005). The concern lies in the multicollinearity that may exist among local regression coefficients which, it is suggested, may invalidate any meaningful interpretation of the spatial patterns. This is of obvious concern, as this study suggests it is appropriate to view this spatial dependency as a source of information rather than error. The resolution of this is a topic of ongoing research and discussion.

RESULTS

Stage 1 – structure of the real estate geography

A summary of the preferred PCA solution is shown in Table 2. Together, the 10 principal components explain approximately 71% of the variance in the original variables and shown in Table 2.

Table 2: Principal components of original 50 variables

Factor	Description	Main contributing variables (correlation between variable and factor) Varimax rotation (from 50 variables)	Variance explained by factor	Cumulative variance explained
1	Families (young, large – average income)	◆ Household size; large (.88); average (.75); small (-.93) ◆ Dependent children (.52) ◆ Lone person household (-.83) ◆ Tenure; mortgaged (.77) ◆ Household income; above-average (.65) ◆ Age structure; 0 to 20 yrs (.82); 35 to 54 yrs. (.53); > 65 yrs (-.76)	15.9%	15.9%
2	Families (disadvantaged)	◆ Household income; below average (.76); average (.41); high (-.89) ◆ Not in labour force (.46) ◆ Education; tertiary (-.86) ◆ Distance; CBD (.45) ◆ Amenity (-.41) ◆ Dwelling area (-.40)	10%	25.9%
3	Ethnicity (Australian born)	◆ English only spoken (.94); English & another language (-.96) ◆ Place of birth – Australia (.67); SE European (-.79); SE Asia (-.66); NE Asia (-.42)	9.2%	35.1%
4	Families (older & well established)	◆ Same dwelling 5 yrs ago (.81); 1 yr ago (.78) ◆ Owned dwelling (.73) ◆ Married (.44) ◆ Age structure; 21 to 34 yrs (-.77)	8.2%	43.3%
5	Dwelling tenure (low income, rental)	◆ Tenure; rental (.62); ◆ Multiple dwelling (.58); ◆ Household income; low (.53); ◆ Unemployment (.68) ◆ Sole parent (.57)	7.2%	50.5%
6	Dwelling proximity (poor accessibility)	◆ Distance from; GP surgery (.81); secondary school (.79); major shops (.68); urban shops (.59); primary school (.66)	6.1%	56.6%
7	Dwelling type (redevelopment potential)	◆ dwelling type; single (.72); home unit (-.73) ◆ dwelling added value (-.63) ◆ dwelling condition (-.46) ◆ dwelling wall construction; stone (.41);	4.6%	61.2%

8	Dwelling type (brick)	◆ dwelling wall construction; brick (.82); rendered (-.81)	3.3%	64.5%
9	Dwelling type (rural living)	◆ dwelling type; rural living (.89) ◆ Site area (.86)	3.2%	67.7%
10	Ethnicity (older European born)	◆ NW European born (.56); ◆ Age 55 to 65 yrs. (.44)	3.2%	70.9%

Although not shown in this paper, the spatial distribution of these factors across the study area made intuitive sense supporting the acceptability of this result.

Stage 2 – relating the REG structure to the market

The results of the hedonic GWR using the surrogate property characteristics derived in stage 1 as independent variables are shown in Table 3.

Table 3: Hedonic GWR results

GWR results **Dependent variable is log sale price**		**Global model**		**Local model (adaptive kernel)**			
	Independent variable	*R^2*	*AIC*	*R^2*	*AIC*	*Bandwidth*	*ANOVA F-statistic*
No. of cases = **7143**	No. of variables =10 F1 to F10	0.618	-7791.98	0.72	-9889.54	1842	30.02

Notes: F-test with d_1, d_2 degrees of freedom (df) where d_1 denotes the df for the global model and d_2 the df for the local model. The null hypothesis is that the local model represents no significant improvement over the global model.

The importance of the results shown in Table 3 are the R^2 and the AIC statistics that indicate the global model is not as good at predicting a result as the local model, meaning there is an effect due to location and hence the possible existence of spatial submarkets. Also, a Monte Carlo simulation was run as part of the GWR software options, indicating that the variation in each of the local parameter estimates is not due to sampling variation at the 0.1% level; suggesting that the local models are displaying something of spatial interest. This was supported by Moran's I calculation for each local parameter estimate, indicating less than a 1% likelihood that the spatial cluster pattern could be a result of random chance.

Therefore, pursuant to the results shown in Table 1, the second GWR model was calibrated over the regular 300 metre grid with an adaptive kernel (bandwidth 1842). This produced 9,075 sets of regression points each with 11 parameter estimates that were clustered and geographically plotted as described in the methodology. The optimum cluster solution representing geospatial submarket boundaries is determined by the user;

however a minimum 10-cluster solution is presented along with examples of higher numbered cluster solutions (150 and 330) in order to demonstrate the results of this study. The background in both Figures 3 and 4 is an interpolated sale price surface (Inverse Distance Weighted interpolation using the same 7,143 sales selected for this analysis). This is to provide contrast between value boundaries and geospatial submarket boundaries, visually demonstrating that one does not necessarily follow the other.

Figure 3 shows a small number of clusters (geospatial submarkets) exhibiting a relatively low degree of homogeneity in the patterns of the parameter estimates. This gives a broad overview of the market structure dividing the whole study area into only 10 submarkets. The geographic size of the geospatial submarkets appear smallest in the older (central) parts of the study area, perhaps because they contain a more heterogeneous residential make up, causing submarket boundaries to change more rapidly than in the newer, more homogeneous, areas in the northern and southern geographic portions of the study area. The submarket signatures show the mean parameter estimate value (standardized) for each of the surrogate property characteristics and the intercept. These are displayed from left to right across the X-axis from the intercept, F1 (factor 1) to F10 (factor 10) for each of the 10 submarkets. Each of the 10 submarket signatures in Figure 3 present obviously different patterns indicating those different factors that affect the submarket formation in those different geographic locations.

As the homogeneity of the parameter estimates increases, so does the resulting number of clusters. Figure 4 is an example of a 150 cluster solution delineated by the heavy dark line and a 330 cluster solution delineated by the lighter internal lines (numbered 111,112 120 & 127). The example in Figure 4 is a smaller geographic area taken from cluster 4 (shown in Figure 3 on the eastern side of the study area). Similar submarket signatures can be seen between all these cluster solutions in so far as they are dominated by factor 5 (F5) and factor 7(F7). The similarity is more pronounced in the smaller geographic areas shown in Figure 4. As can be seen from Table 2, these are dimensions of the REG describing low income, rental accommodation in dwellings with high redevelopment potential. Based on local and anecdotal knowledge, this makes sense.

The submarket structure is a continuous complex surface representing the 10 components of the underlying real estate geography. The definition of a particular set of boundaries delineated along that continuous surface is a matter for the particular user and their application. Broad planning studies may be interested in the lesser number of submarket delineations showing the overall picture as exampled in Figure 3. On the other hand, the mass appraisal function would want submarkets containing enough sales data for the construction of predictive models. In each case, the underlying market drivers can be seen from their respective submarket signatures.

Figure 3: Submarket structure

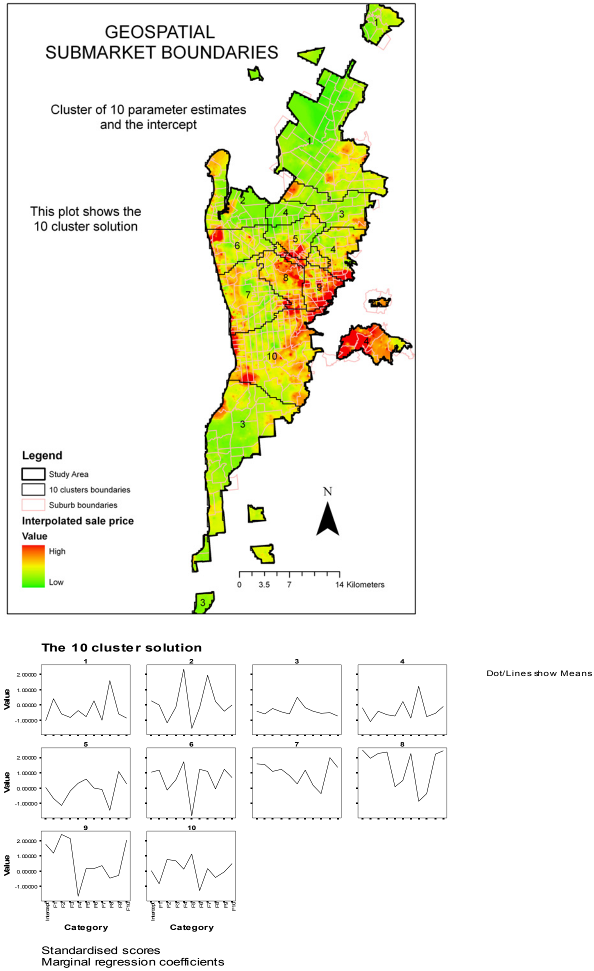
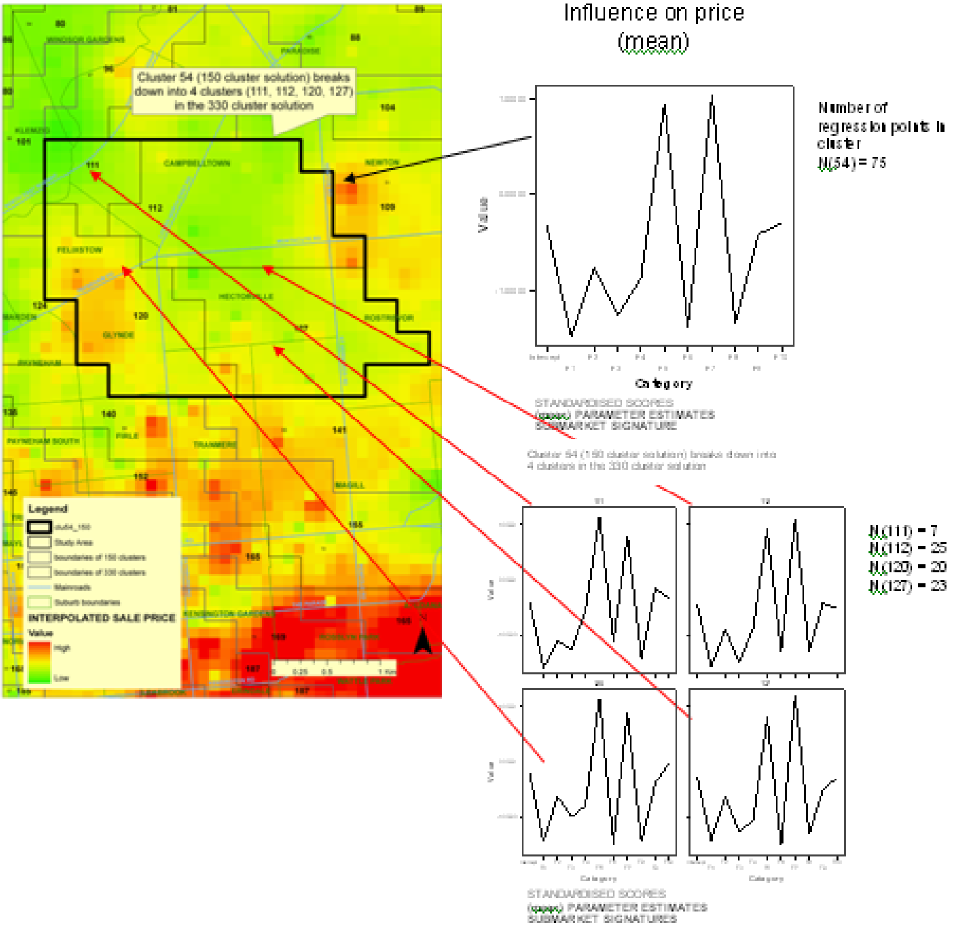


Figure 4: Submarket structure



CONCLUSION

The literature suggests that submarket delineation should contain a number of elements in order to satisfy the needs of various land related professionals in relating their activities to an understanding of the underlying real estate market structure. These elements, summarized in the introduction, included the need to recognize that submarket delineation should be derived from the data alone, incorporate both spatial and structural attributes and should reflect the underlying residential market structure of the area in which submarkets are to be detected. This study has attempted to incorporate all these elements.

The use of PCA has allowed the large number of both spatial and structural property attributes necessary to describe the underlying real estate geography to be quantified into a sensible number of surrogate property characteristics. The use of GWR allowed these characteristics to be related to the market, detecting price movement to indicate submarket change. This methodology is limited by the original data collected, both in terms of quality and quantity and by the interpretation of results. It is dependent on the input of the land related professional for the management of the whole process.

This is an experimental methodology with additional research needed to be undertaken to more properly understand the interpretation of the resulting parameter estimate patterns. This is important as it may add another dimension to understanding submarket structure recognized as important in housing market analysis.

REFERENCES

Adair, A, Berry, J & McGreal, S 1996a, 'Valuation of residential property: Analysis of participant behaviour', *Journal of Property Valuation and Investment*, vol.14, pp.20–35.

——1996b, 'Hedonic modelling, housing submarkets and residential valuation', *Journal of Property Research*, vol.13, pp.67–83.

Barker, K 2004, *Review of housing supply delivering stability: Securing our future housing needs (Final report – recommendations)*, HM Treasury, London.

Bates, L 2006, 'Does neighborhood really matter? Comparing historically defined neighborhood boundaries with housing submarkets', *Journal of Planning Education and Research*, vol.26, pp.5–17.

Borst R. & McCluskey W. (2008) 'Using Geographically Weighted Regression to detect housing submarkets: modeling large-scale spatial variations in value', *Journal of Property Tax Assessment & Administration Vol 5 Issue 1.*

Bourassa, S, Cantoni, E & Hoesli, M 2007, 'Spatial dependence, housing submarkets, and house price prediction', *Journal of Real Estate Finance and Economics*, vol.35, pp.143–60.

Bourassa, S, Hameling, F, Hoesli, M & McGregor, B 1999, 'Defining housing submarkets', *Journal of Housing Economics*, vol.8, pp.160–83.

Bourassa, SC, Hoesli, M & Peng, VS 2003, 'Do housing submarkets really matter?', *Journal of Housing Economics*, pp.12–28.

Cano-Guervos, R, Chica-Olmo, J & Hermoso-Gutierrez, J 2003, 'A geo-statistical method to define districts within a city', *Journal of Real Estate Finance and Economics*, vol.27, pp.61–85.

Chhetri, P, Stimson, R & Western, J 2006, 'Modelling the factors of neighbourhood attractiveness reflected in residential location decision choices', *Chiikigaku Kenkyu (Studies in Regional Science)*, vol.36, pp.393–417.

Clapp, J 2003, 'A semiparametric method for valuing residential locations: Application to automated valuation', *Journal of Real Estate Finance and Economics*, vol.27, pp.303–20.

Figueroa, R 1999, 'Modelling the value of location in Regina using GIS and spatial autocorrelation statistics', *Assessment Journal*, vol.6, pp.29–38.

Fik, T, Ling, D & Mulligan, G 2003, 'Modelling spatial variation in housing prices: A variable interaction approach', *Real Estate Economics*, vol.31, pp.623–46.

Fotheringham, A, Brunsdon, C & Charlton, M 2002, *Geographically weighted regression the analysis of spatially varying relationships*, Wiley.

Gallimore, P, Fletcher, M & Carter, M 1996, 'Modelling the influence of location on value', *Journal of Property Valuation and Investment*, vol.14, pp.6–19.

Goodman, A & Thibodeau, T 2007, 'The spatial proximity of metropolitan area housing submarkets', *Real Estate Economics*, vol.35, pp.209–32.

Grigsby, W 1963, *Housing markets and public policy,* University of Pennsylvania Press.

Jackson, E, Kupke, V & Rossini, P 2007, *The relationship between socio-economic indicators and residential property values in Darwin*, Thirteenth Annual Pacific-Rim Real Estate Society Conference, Fremantle, Western Australia.

Kauko, T 2000, *Modelling location in house prices: A literature review of the empirical research methodology*, Urban Research Centre Utrecht (URU), Utrecht.

Kestens, Y, Theriault, M & Des Rosiers, F 2004, 'The impact of surrounding land use and vegetation on single-family house prices', *Environmental and Planning B*, vol.31, pp.539–67.

Lockwood, A & Coffee, N 2006, 'Residential living structure as a basis for the spatial delineation of residential submarkets', *Pacific Rim Property Research Journal*, vol.12, pp.350–68.

Maclennan, D & Tu, Y 1996, 'Economic perspectives on the structure of local housing systems', *Housing Studies*, vol.11, pp.387–405.

McCluskey, WJ, Deddis, WG & Lamont, IG 2002, *The application of spatially derived location factors within a GIS environment*, PRRES conference, Christchurch, New Zealand, <http://business.unisa.edu.au/prres>.

Meen, D & Meen, G 2003, 'Social behaviour as a basis for modelling the urban housing market: A review', *Urban Studies*, vol.40, pp.917–35.

Pryce, G 2004, *The micro and macro effects of the location of new housing supply: Interim technical report for the Office of the Deputy Prime Minister*, March 2004.

Pryce, G & Evans, G 2007, *Identifying submarkets at the sub-regional level in England*, Department for Communities and Local Government, London, <www.communities.gov.uk/index.asp?id=1510297>.

Rossini, P. & Kershaw, P. 2005, *Using a Hybrid Automated Valuation Model to Estimate Capital and Site Values*, Eleventh Annual Pacific-Rim Real Estate Society Conference, Melbourne, Australia.

Schnare, A & Struyk, R 1976, 'Segmentation in urban housing markets', *Journal of Urban Economics*, vol.3, pp.146–66.

Shevky, E & Bell, W 1955, *Social area analysis: Theory, illustrative application and computational procedures*, Stanford University Press.

Tu, Y, Sun, H & Yu, S 2007, 'Spatial autocorrelations and urban housing market segmentation', *Journal of Real Estate Finance and Economics*, vol.34, pp.385–406.

Watkins, C 1999, 'Property valuation and the structure of urban housing markets', *Journal of Property Investment and Finance*, vol.17, pp.157–75.

——2001, 'The definition and identification of housing submarkets', *Environmental and Planning A*, vol.33, pp.2235–53.

Wheeler, D. & Tiefelsdorf, M. (2005) 'Multicollinearity and correlation among local regression coefficients in geographically weighted regression' J *Geograph Syst* 7: pp.161-187.

Email address of contact author: tony.lockwood@unisa.edu.au