XLI Incontro di Studio del Ce.S.E.T.: 535-545

Richard A. Borst

Tyler Technologies, Inc., USA
E-mail: *Rich.Borst@tylertech.com*

A Space-Time Model for Computer Assisted Mass Appraisal

Key words: *AVM, GWR, spatio-temporal*

An Automated Valuation Model (AVM) that seeks to attain predictive accuracy must take into account both spatial and temporal effects in the real estate market. A model structure that contains neither explicit spatial nor temporal variables is calibrated by a method that recognizes these variations in is calibration architecture. The method is conceptually similar to Geographically Weighted Regression (GWR) except that it extends into the temporal domain. The methodology is explained and results provided illustrating spatio-temporal variations in value.

1. Introduction

Automated Valuation Models (AVMs) are used extensively in government for property tax purposes and in business, primarily to establish a valuation for mortgage approval purposes. The challenges in developing models of the real estate market are many. Perhaps the most important factors in determining value are location and the time reference for the value being estimated. There are a number of methods that address either location or time in an AVM model specification. Market segmentation is used to account for large scale geographic variations in value (Borst and McCluskey 2008, p 21-54). Local variations are addressed by a number of methods including response surface analysis and comparable sales analysis (Borst and McCluskey 2007, p 107-130). The temporal aspects of value have been modeled in a number of ways including "dummy variables" for Month or year, linear splines based on month of sale, Fourier Expansions of the time domain and Time weighted Regression (Borst 2008, p 33-40) and Borst (2009, p 29-36). These methods generally require the specification of the functional form of the time adjustments.

The remainder of the article is devoted to presenting an alternative approach to the spatio-temporal AVM model specification. In particular a model is described that does not require the specification of the functional form of the location variables or the time variables. The next section describes Geographically Weighted Regression and how it is used in developing the spatio-temporal model.

2. Geographically Weighted Regression

In a typical application of Multiple Regression Analysis (MRA) one equation is calibrated on a given set of sales, each of which is weighted equally. There are variants of this technique that weight the points in the sample data set differently, such as Robust Regression, which seeks to down-weight or eliminate outliers in the data set. However, there is still only one equation developed for the entire data set, and there is no inherent spatial component to the weighting method. GWR, on the other hand is a computationally intensive technique that weights each point in the dataset based on its location. The introduction to the concepts involved in GWR often includes description of moving window regression. In this case the sample points within a fixed distance of a given point are included in the regression, all with equal weight, and all others are excluded.

GWR operates in a similar fashion; however the points do not receive equal weight. Instead, the weight is a function of location, and diminishes with the distance from the regression point. Figure 1 illustrates the concept of a weighting function based on the coordinates of the regression point and the data points near it. The peak of the surface is the regression point; any sample points under the surface would receive a weight based on the relative height of the surface at that point.

Two dimensional representations of the kernel density function are used to illustrate the concept of fixed and variable bandwidth weighting.

Figure 2 shows the spatial dimension along the *x* axis, and the weighting function is represented by the vertical height of the curve. The height of the curve at point *j*, given by W_{ij}, is the weight applied to point when point *j* is the regression point and d_{ij} is the distance between the regression point *i* and data point *j*. The bandwidth of the spatial kernel is a parameter that affects how the weight is computed as the distance between the regression point and other sample points increases.[1] It can be fixed or variable. This becomes an important consideration when the sample points are not regularly spaced. A fixed bandwidth could result in there being insufficient points considered in the regression if the bandwidth is too small.

Figure 3 shows two kernels. In the left part of the figure, the ten data points are more closely spaced than the ten points on the right. In this hypothetical example it can be seen that by appropriately adjusting the bandwidth, ten data points can be considered in each case.

The output of a GWR is quite voluminous. There are separate equations calibrated at each sample point along with corresponding diagnostic statistics. However, there are certain representations of the GWR model that are concise enough for presentation. Figure 4 depicts the spatial variation of a coefficient in the model relating to property size. The darker points have the higher values.

---

[1] In the Gaussian formulation of the kernel, the weighting function is given by the distance decay function.

Figure 1. Example Spatial Weighting Kernel.

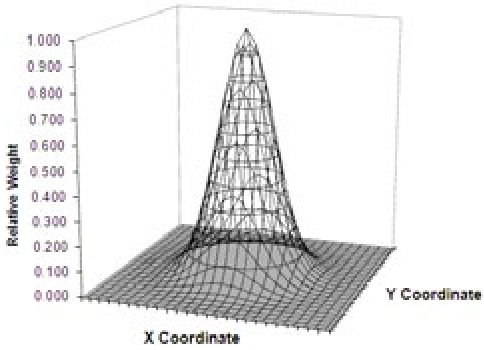


Figure 2. Two-Dimensional Kernel.

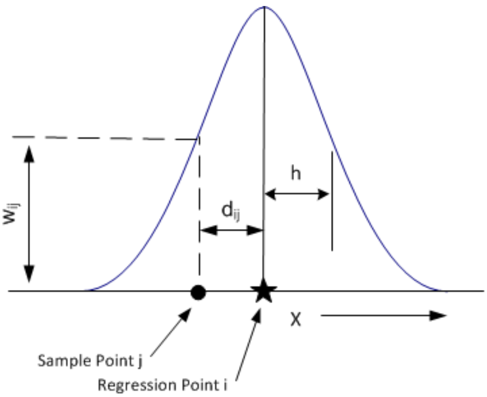
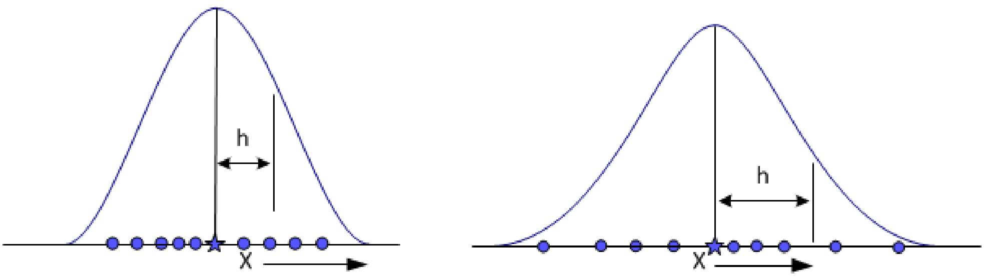


Figure 3. Variable Bandwidth Kernels.



3. Time Weighted Regression

There is a temporal analogue to GWR. It is based on devising a temporal kernel and using it to weight points according to the date of sale. There is a distinct difference between temporal weighting and geographic weighting. In GWR we are relying on the first law of geography, which according to Waldo Tobler is "everything is related to everything else, but near things are more related than distant things". This implies an isotropic view of space – more or less uniform regardless of direction. However with the time dimension there is directionality. To represent this concept a modified temporal kernel is used as shown in Figure 6.

Figure 4. Spatial Representation of Living Area Coefficient $/square foot.

Figure 5. Three Dimensional Representation of Living Area Coefficient.

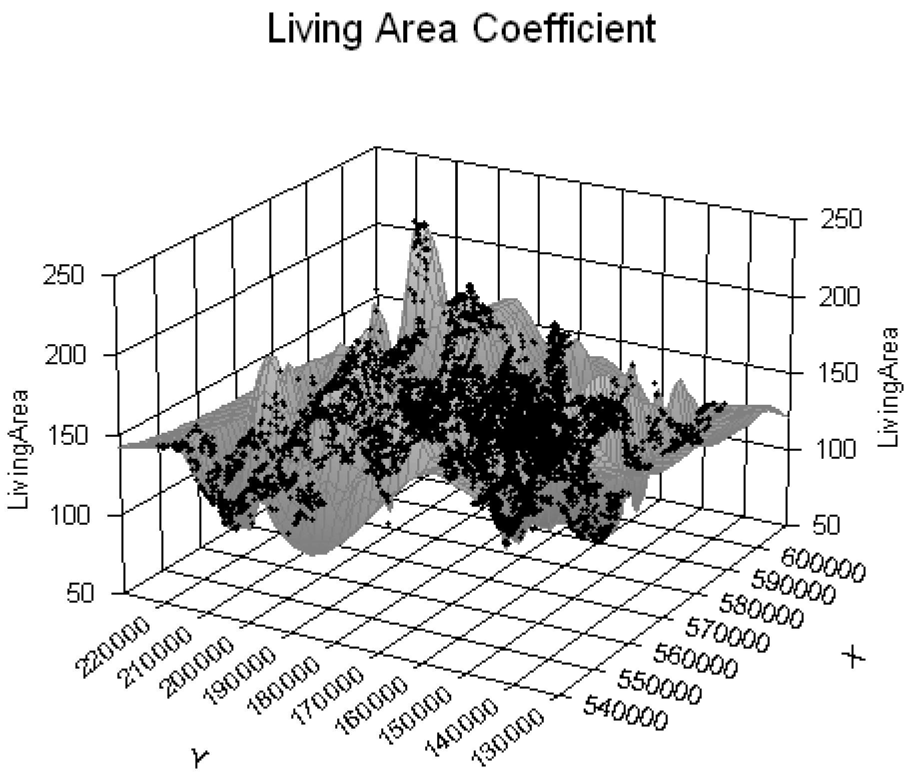
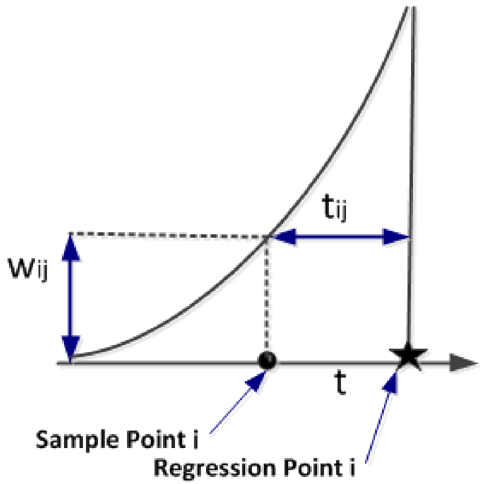


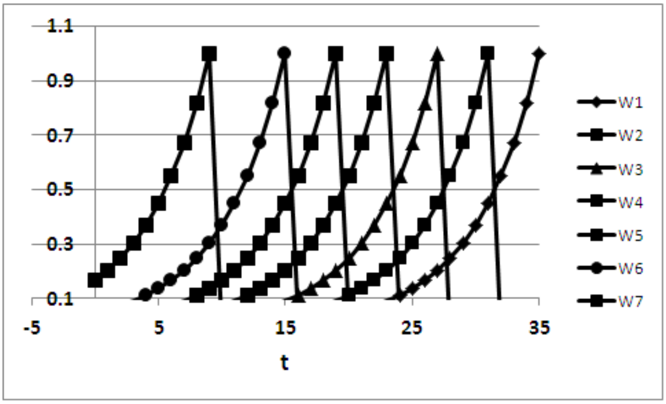
Figure 6. Temporal Weighting Kernel.



4. Implementation of Model

To calibrate the spatio-temporal model a dataset of 5,863 sales were obtained from a Midwestern U.S. county. The GWR tool in the ESRI product ArcMap was used to accomplish the spatial aspects of the model. The temporal component was achieved by replicating the data seven times and applying the temporal kernel represented by Figure 6. The actual weights are provided in Figure 7 where are the weights for the most recent sales expressed in months from a reference date. In this data the "month of sale" ranges from 0 to 35 covering a three year period.

Figure 7. Actual Temporal Weights.



An exponential decay function was used to compute the values of as shown in Table 1.

Table 1. Exponential Decay Functions for Wij.

Weight Set	Regression point t	W_{ij}
W_1	35	t>35, W_{ij}=0,t<=35, W_{ij}=Power(d_{i35},0.2)
W_2	31	t>31, W_{ij}=0,t<=31, W_{ij}=Power(d_{i31},0.2)
W_4	29	t>27, W_{ij}=0,t<=27, W_{ij}=Power(d_{i27},0.2)
W_3	23	t>23, W_{ij}=0,t<=23, W_{ij}=Power(d_{i23},0.2)
W_4	17	t>17, W_{ij}=0,t<=17, W_{ij}=Power(d_{i17},0.2)
W_5	13	t>13, W_{ij}=0,t<=13, W_{ij}=Power(d_{i13},0.2)
W_6	9	t>9, W_{ij}=0,t<=9, W_{ij}=Power(d_{i9},0.2)
W_7	5	t>5, W_{ij}=0,t<=5, W_{ij}=Power(d_{i5},0.2)

Figure 8. LocF for W_7.

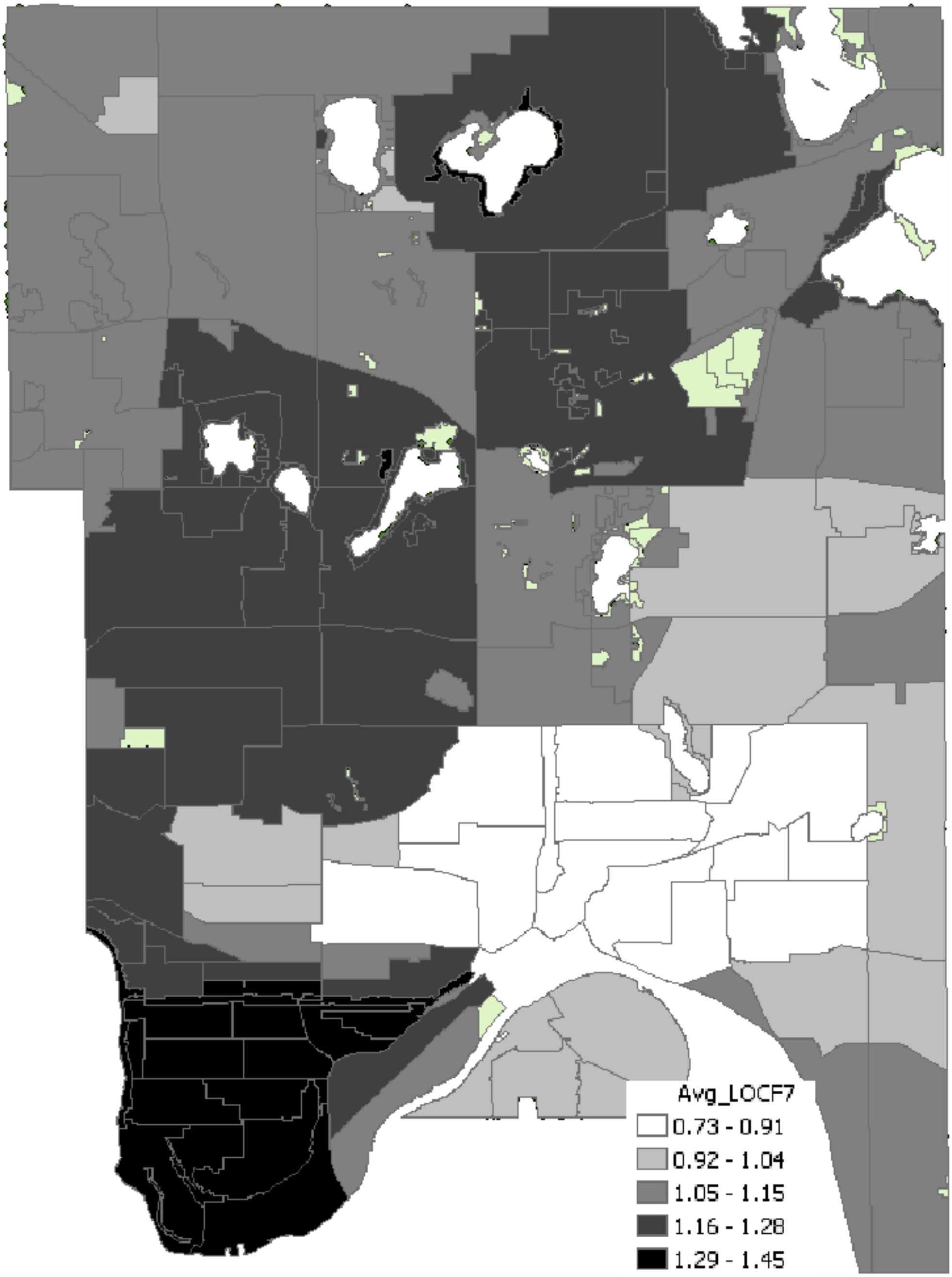


Figure 9. LocF for W1.

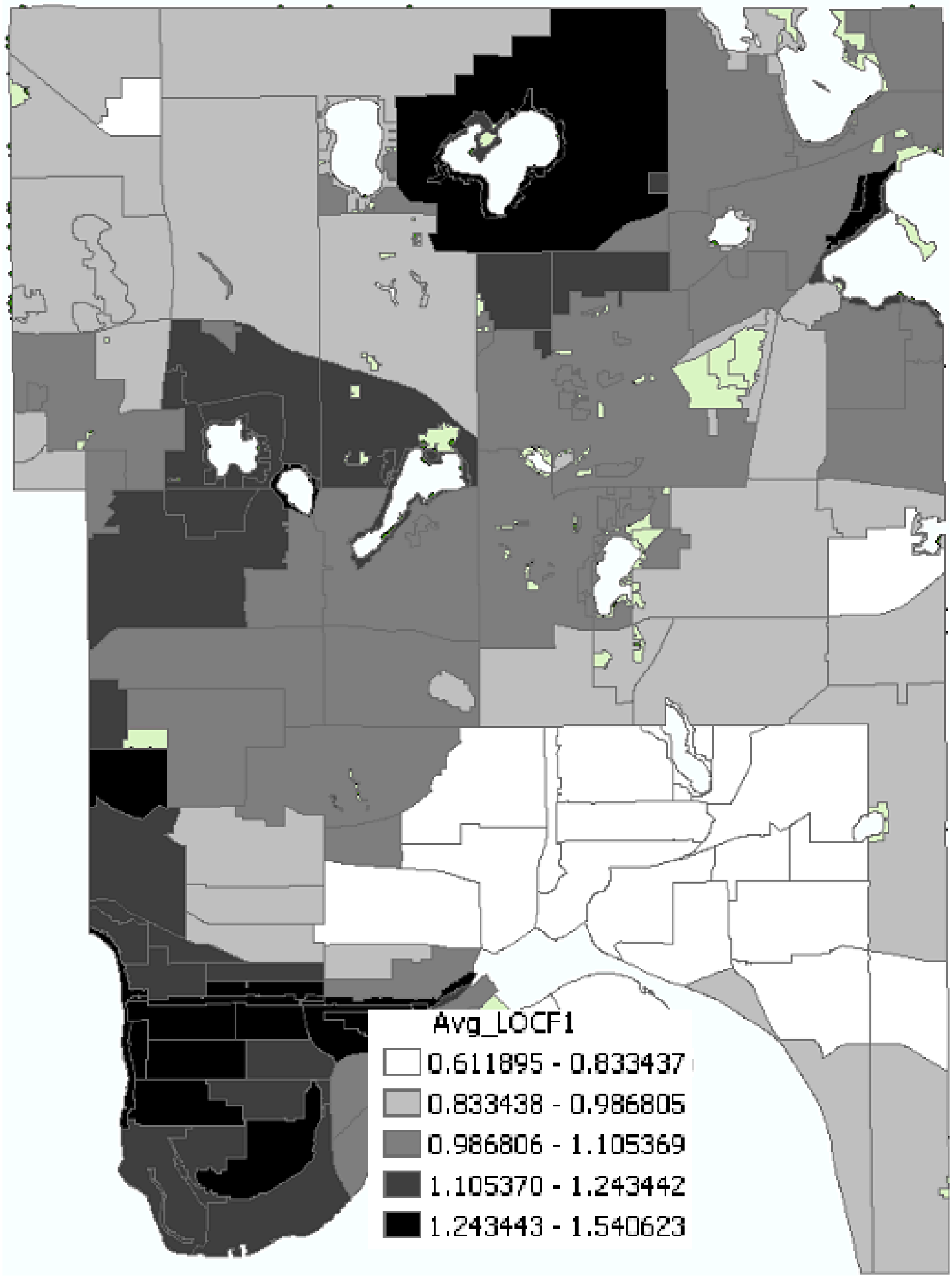


Figure 10. Maximum LocF by W_i.

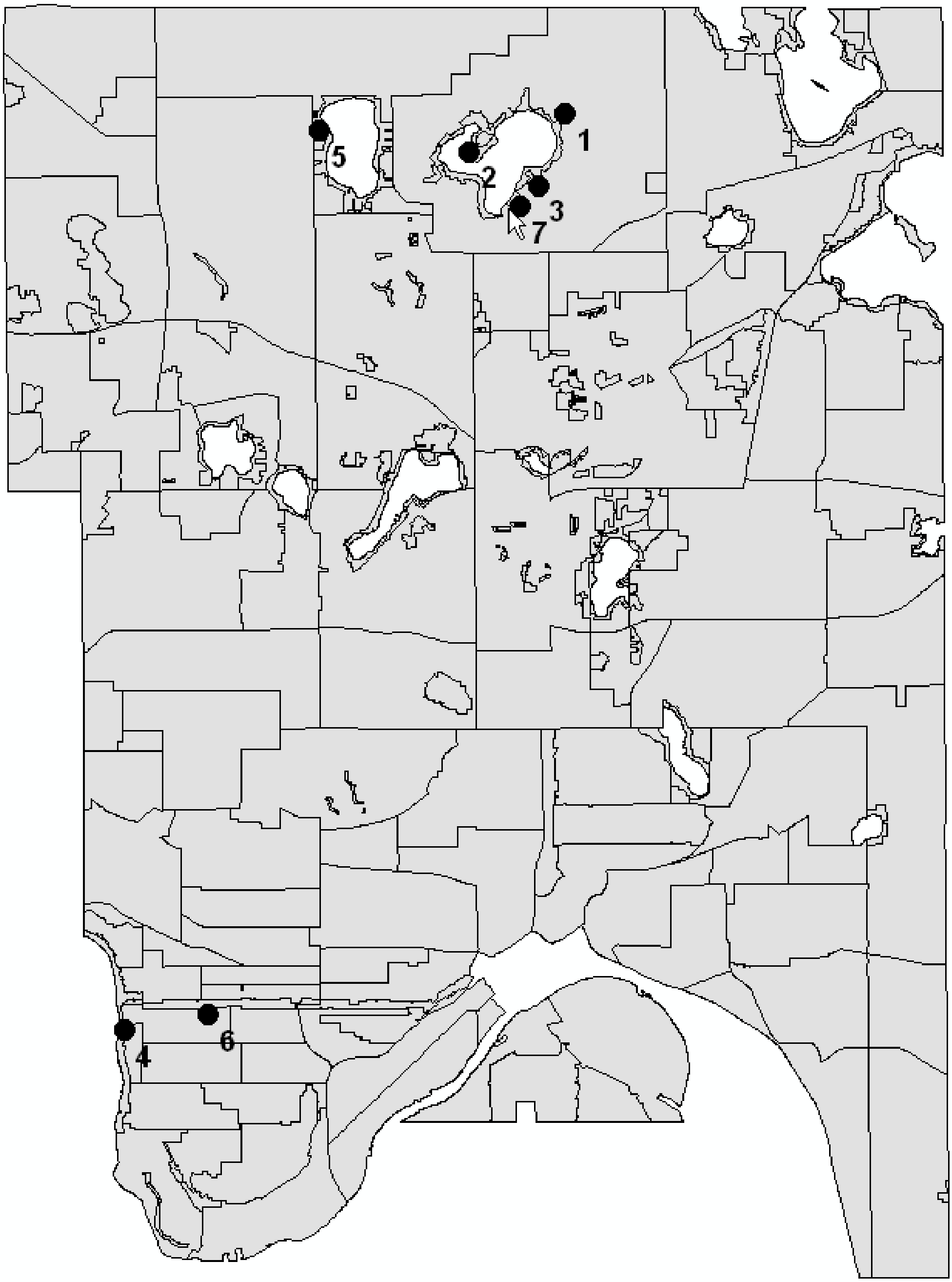


Figure 11. LocF Selected Models.

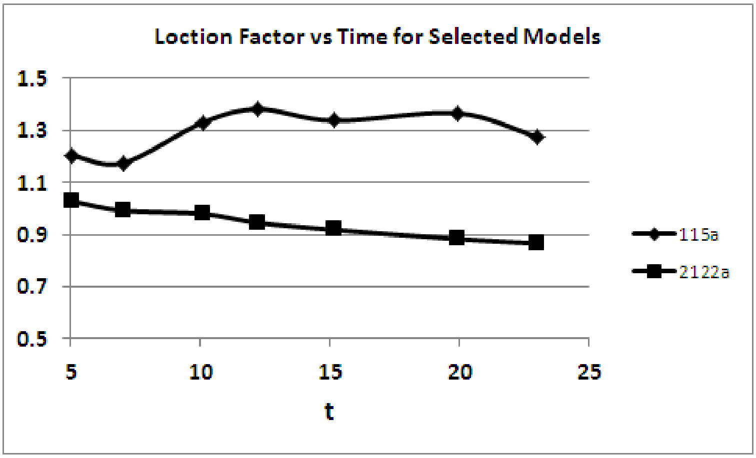
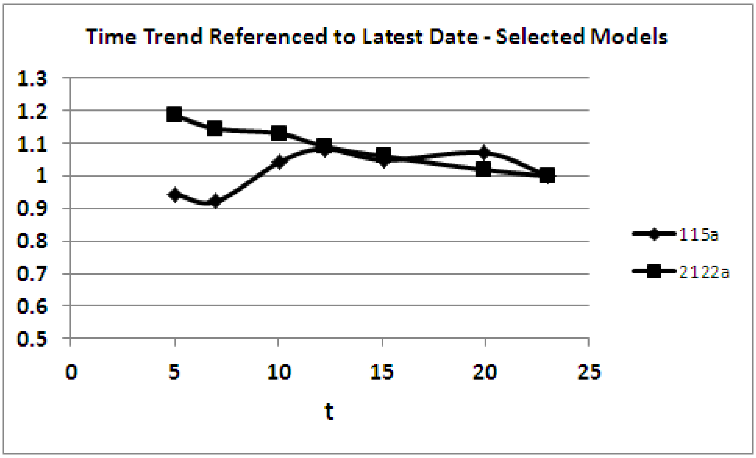


Figure 12. Time Trend Selected Models.



5. Presentation of Results

The spatio-temporal model is actually a set of seven tables of estimated coefficients and related goodness of fit statistics. The results, necessarily, need to be transformed so they may be more easily interpreted. To do this the value of the average home was computed for all data points across all seven models. This produces 7*5,683 values of interest. The average value of the average home for the set was used to normalize all values into a Location Factor (LocF). One way to review the results is to create a thematic map of LocF by data set. Figure 8 and Figure 9 provide thematic maps of LocF for and respectively the oldest sales vs.

the newest sales. A comparison of the two figures reveals the changing pattern of the peak location influence in the county.

In Figure 10 the maximum value of LocF is plotted. The numbers on the points in the figure represent each weight set from 1 to 7. Five of seven times the maximum is in the north central portion of the county, while two out of seven times it is in the southwest. These are two of the relatively higher valued areas of the county as seen from figures 8 and 9.

The county is divided into a number of "models" that are constructed to represent homogeneous and geographically contiguous areas. In fact they are used in the county's AVM. Two observations are evident. First the overall values of LocF are different for each model. Second, the trend over time is not the same shape for each model. This can be expressed in a slightly different way by referencing each trend line to the most recent data point. In Figure 12, the change in value over time for each model is quite clear.

6. Conclusions

An easy to implement methodology for developing a spatio-temporal model of real property values has been given. The methodology requires only the availability of a GWR program and a way to present thematic maps, a GIS.

With the spatio-temporal model a researcher can study the change in value patterns over time and space regardless of the cause of the change. A practitioner can derive and apply time corrections to sales data without having to specify a functional form of the model.

Bibliography

Borst Richard A. (2008). Evaluation of the Fourier transformation for modeling time trends in a hedonic model, *Journal of Property Tax Assessment & Administration 2008 v. 5 no. 4 (5)*, 34-40.

Borst Richard A. (2009). Time Weighted Regression: Modeling the Effect of Sale Date in a Hedonic Model, *Journal of Property Tax Assessment & Administration 2009 v. 6 no. 4 (3)*, 29-36.

Borst Richard, McCluskey William (2007). Comparative Evaluation of the Comparable Sales Method With Geostatistical Valuation Models, *Pacific Rim Property Research Journal Volume 13 Number 1 March (2007).* 106-130.

Borst Richard, McCluskey William J. (2008). Using geographically weighted regression to detect housing submarkets: modeling large-scale spatial variations in value, *Journal of Property Tax Assessment & Administration, 5:1, 21-54.*

Fotheringham A. Stewart, Brunsdon Chris, Charlton Martin (2002). *Geographically Weighted Regression,* John Wiley & Sons Ltd.