Kohonen Map, GIS and the Analysis of Real Estate Sales

Eero CARLSON, Finland

Key words: Kohonen Map, GIS, Market Analysis, Appraisal.

ABSTRACT

Real estate sales prices are analyzed using GIS database containing cadastral and topographic information. Self-organizing map (SOM) introduced by Teuvo Kohonen is used for analysis and visualization. Object oriented data model includes *observation* (components of property including location and sales date), *valuation* (comparable sales in the neighborhood) and *group* (neighboring sales). Components of the *observations* are coded and scaled properly for reasonable comparison of sales. Geographic location is balanced with other components and the *observations* are organized as *valuations* and *groups* using SOM (Kohonen Map) for each category of real property (woodlots, agricultural, recreational land, one-family houses etc).

Kohonen Map organizes observations into **topological groups**. The topological order in the observation space is preserved. Resembling observations are found in the topological neighborhood. Small differences in the component values may be seen as small distances between the units in the Kohonen Map. Exceptional cases found their place more or less isolated within the ordered display of units showing the average values for components.

Kohonen Map organizes observations into **nearly equally sized groups**. The distribution of the observations is preserved. The symbols of the *groups* and the *valuations* are used as links to the matching observations and the values of the components can be seen in the list and the location of the real property can be shown on a detail map.

GIS database is used to get cadastral and topographic information. Parcels, plans, buildings, shoreline, network connections, land use, neighborhood, distances and other factors can be seen and extracted for analysis. Overview and detail maps are used to display *groups*, *valuations* and *observations*.

Object oriented data model includes *observation*, *valuation* and *group* in one super class. The components or factors are computed using methods defined in this super class. The methods return discrete or continuous numeric values and are used in the sales comparison and in the analysis of individual components. General information for each category of real property includes the description and scaling of the components, sales price index produced with Kohonen Map and other important data.

Time is managed by organizing observations as first *valuations* using Kohonen Map and approximate long-term average prices are found. Actual sales prices are then compared to the *valuations* and the difference is seen as the effect of time. Sales price index is formed and the sales prices are transformed to a given level. The *valuations* organized as a new Kohonen

Map give better estimates for the sales prices. The process is iterated once more and the final sales price index is produced. Structural market changes are recognized.

Location is managed by organizing neighboring observations as *groups* using Kohonen Map. Observed sales prices are compared to the matching *valuations* and the average value of the ratio and difference is computed for the *group*. These **local sales price undulations** are shown on the map and sales price topography can be visualized clearly. Local centers, attractive sites, shores, railways, highways, high voltage transmission lines and other local characteristics can be seen within the effect of a whole.

Local market is visualized as a local Kohonen Map. The *group* symbol is pointed on the geographic map and the neighboring observations are shown as an ordered display of units averaged from local observations only. The number of units may be the same as the number of observations. Small and old houses with oven or stove heating usually locate in one corner and large and new buildings with amenities and build-ups in the other. Another display is shown using the matching *valuations* computed from resembling observations in the surroundings. Scatter plots drawn according to principal components are helpful.

Individual components are analyzed by organizing selected observations as a one-dimensional Kohonen Map according to the selected component. Ad hoc methods returning a continuous or discrete numeric value can be used, too. The sales price ratios and differences are visualized as curves or bars and as tables. Local sales price undulations computed for the *groups* are taken into account in the comparison.

One-family houses in Helsinki capital region are analyzed. Ten components including location, sales price, floor area, year of construction, detail plan, parcel area, shore, building material and heating system are scaled properly for sales comparison. The 4750 observations are organized as 1428 *valuations* and 1134 *groups*. Special emphasis is given for location (neighborhood, local centers, shore, airport). The negative effects of highway (up to 15 percent), railway (up to 11 percent) and high voltage transmission line (up to 5 percent) are reported.

CONTACT

Eero Carlson
National Land Survey of Finland
P.O. Box 84
FIN-00521 Helsinki
FINLAND
Tel. + 358 20 541 5183
Fax + 358 20 541 5070
E-mail: eero.carlson@nls.fi
Web site: www.nls.fi

Kohonen Map, GIS and the Analysis of Real Estate Sales

Eero CARLSON, Finland

1. INTRODUCTION

Real estate market analysis may benefit huge amounts of information available in databases. The problem is how to understand the dependencies of the nearly unlimited number of factors changing over time. Same factors may have both positive and negative effects depending on the situations. Location is something very essential in real estate business. Location includes all positive and negative characteristics. All activities in the environment, landscape, view, smell, noise, houses, shops, offices, connections interact with each other. The neighborhood should be understood as a whole.

Exploratory data analysis using Self-Organizing Map (SOM or Kohonen Map) and huge amounts of data collected into Geographic Information Systems (GIS) may give some new insights. An ordered, generalized view from widely varying transactions and their features may help to understand complex dependencies.

Sales comparison approach may be realized in our minds as some self-organization process. Resembling things are categorized by only observing them. Self-organizing artificial neural network is a natural way to represent multi-dimensional information. Comparable sales are found and averaged values are produced as ordered representations.

2. KOHONEN MAP

Self-Organizing Map (SOM or Kohonen Map) is an unsupervised neural algorithm, method, tool or paradigm introduced by Teuvo Kohonen (Kohonen 1995). Kohonen Map has been widely used in exploratory data analysis (Kaski 1997), financial analysis (Deboeck 1998), bankruptcy prediction (Kiviluoto 1998) and real estate market analysis (Carlson 1991, 1997, 1998 and 2000). Data exploration, data mining, visualization and browsing of unknown data organized topologically may be helpful, if we are trying to find new structures and dependencies in huge amounts of data. Discoveries may happen in the topological neighborhood of known things. Some new and unexpected things may be found between familiar ones.

Kohonen Map is a usually two-dimensional hexagonal lattice of units adapting observations. Components are given as numeric vectors. The self-organization algorithm has two steps:

1. Locate the best matching unit.
2. Increase matching in the unit and in its topological neighborhood.

The distance between the observation vector and the reference vector (unit) is usually computed as a Euclidean distance. Vector components are coded and scaled properly for reasonable comparison. Reference vectors are then moved a given step towards the

observation vector in the topological neighborhood. The step and the neighborhood decrease during the process. The Kohonen Map adapts the observations repeatedly and the reference vectors (units) converge to certain average values observed in the unit and its six neighbors. Thus the identity of the unit is determined mainly from the observations matching the six neighbors.

Global topological order and distribution of observations is preserved. The neighborhood may be more than half of the map in the beginning and the units are distributed over the observation space. Dense places are represented with many units in valleys and sparse, exceptional observations find their place more or less isolated in ravines.

The distribution of observations is preserved. This is useful in browsing. The symbol of the unit can be used as a link to the list of the observations. Neighboring units give resembling observations and the direction may give some idea of the dependencies between the components.

3. GEOGRAPHIC INFORMATION SYSTEM

Huge amounts of data is collected into databases and organized as Geographic Information Systems (GIS). Cadastral and topographic data can be used for analysis. Real property, plans, buildings, road and other connections, shore, landscape, public and private services, shops and other physical and mental characteristics present in a given location may be found in the information system and may be used as a component for exploratory data analysis.

The units organized as Kohonen Maps may be included in a GIS. The observations of real property sales may be grouped with their comparables in the neighborhood and some idea of average sales prices can be given. The observations of real property sales may be grouped with their neighboring sales and some idea of the average sales price differences between the location and the comparables can be given.

4. OBJECT-ORIENTED DATA MODEL

Object-oriented design and programming gives many advantages for data analysis applications. Polymorphism and encapsulation allow classes to be created with clear and simple definitions. Inheritance allows methods to be written only once for a group of classes. Dynamic binding allows the use of ad hoc methods in the analysis. Classes and slots may be designed for easy and secure use of the objects recorded in the database.

Real estate sales data is organized in three classes inheriting from one super class. The slots or methods giving the component values are harmonized so that a given method is forced to give comparable values for the desired field or variable. For example ‘floor area’ is measured in same square meters in all instances. Three classes inheriting from one super class are:

1. *Observation*
   Components of real property including location and sales date. Continuous or discrete numeric values for the features. Relationship with the matching *valuation* and *group*.

2. *Valuation*
Comparable sales. Averaged values, including the sales price, for the components of real property in a group of resembling sales organized as a unit in a Kohonen Map. One observation belongs in a hexagonal neighborhood (usually seven units).

3. *Group*
Neighboring sales. Averaged values for the components of real property in a group of neighboring sales organized as a unit in a Kohonen Map. One observation belongs in a hexagonal neighborhood (usually seven units).

General information for each category of real property includes the description and scaling of the components used in the comparison of sales, residual errors (rmse) of the components after the organization of the Kohonen Map, sales price index produced using Kohonen Map and other important information.

5. *VALUATION*: COMPARABLE SALES

It is a cognitively highly demanding process to find comparable sales. Real properties are unique and they interact excessively with many things in the neighborhood. Expectations for future may differ for each buyer and seller in the market. Some features are important or necessary for one participant, but another can see these features as undesirable. Contradictory effects are realized in the sales prices changing over time.

5.1 Scaling of the Components

Kohonen Map is used to organize observations into groups of resembling observations. Components in the observation vector are coded and scaled properly for reasonable comparison. Euclidean distance is usually used for comparison. The scaling of the components is an important step in the process (Carlson 1997 and 1998). The topological order in the Kohonen Map can be slightly adjusted by changing the scaling factors. Strongly scaled components are represented with many units and respectively weakly scaled components are averaged more. Nevertheless the resulting averaged values in the units does not differ much because there are usually smaller and larger values in the neighboring units.

Good values to begin with are to normalize the minimum and maximum between 0 and 100. Other good values are to normalize the variance of the component values. Nevertheless, some exceptionally high values may have undesirable strong effect. Best values can be found by experience. The purpose of the grouping is taken into account. For example, one year in construction may cost average 1000 euros and one square meter in floor area may cost average 500 euros. Scaling factor value of 20 for the year of construction and 10 for the floor area may give desired results.

The point is how to tolerate the differences in the comparables. If the size of the building does not match, how far away or how old or new building is acceptable. Practical way may be to extend the distance about five times compared to the distance of neighboring

observations. Topological order may be forest, plan, parceling out, infrastructure, building and additional buildings. Topological order may be on lake, close to lake and far away from lake. Sales prices may not be symmetric on the left and on the right side and there may be some systematic deviations in certain situations.

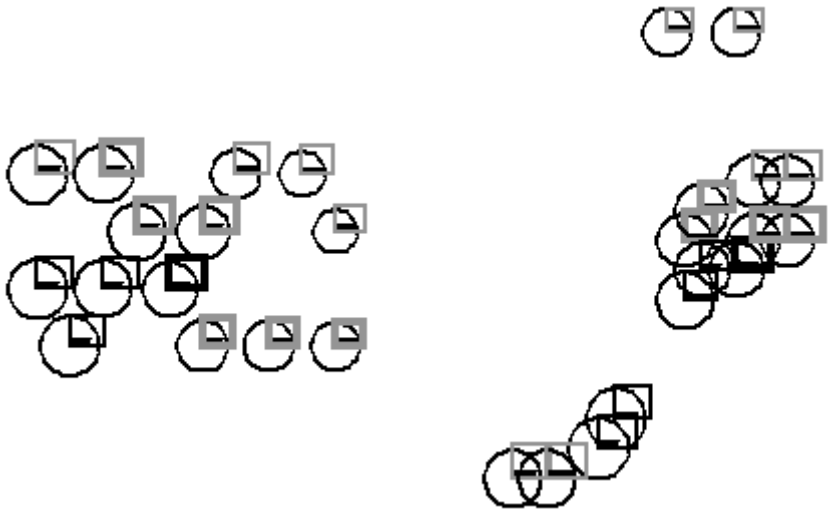

Figure 1. One *group* of 19 one-family houses in Suutarila is shown as a local Kohonen Map with 20 units (average 4.2 comparable sales per unit, left). Matching 15 *valuations* (average 22.3 comparable sales per unit) are shown from the 34 rows and 42 columns Kohonen Map for the 4750 observations in Helsinki capital region (right). Geographic maps are shown in Figure 2.

5.2 Symbol

Traditional cartographic knowledge can be used in visualization. The symbol for the *valuation* is designed using graphic variables. Strong and weak components should have strong and weak visual variables. Visually greater values are used for greater values of the components. *Location, size, intensity, saturation, texture and orientation* represent relationships of order and proportion. *Hue* also represents relationships of order, but in addition it represents strong relationship of kind. *Shape* represents only relationships of kind.

Real property symbol may have four parts:

1. Circle (sales price, category of real property, timber volumes as sectors)
2. Rectangle (building, floor area, year of construction, amenities, build-ups, condition)
3. Parallelogram (parcel, area, plan, rights)
4. Lake symbol (shore, size of lake, quality)

In the examples shown the rectangle is gray or black if the year of construction is before 1970 or after. The line is thick or thin if the building material is stone or wood. The line in the parallelogram is thick or thin if there is detail plan or not. Shore side lots are not shown.

5.3 Browsing the Sales

The symbol of the *valuation* is drawn in the hexagonal lattice with its neighbors. On the geographic map the symbol is drawn in the averaged location of the comparable sales. The symbol is used as a link to the list of the matching observations and all component values can be browsed and the location can be seen on a detail map.

Neighboring *valuations* are pointed and topological order and distribution of observations can be detected. All values of the components are changing simultaneously. The effect of an individual component is hardly seen because the effect of the whole is realized in the typical sales prices seen in the *valuations*.

6. TIME: STRUCTURAL MARKET CHANGES

Market changes may be strong over time. Expectations for future are changing: economy, income, physical and mental environment and neighborhood. All things are changing more or less. The problem is how to use old observations. It seems advisable to use old observations instead of not using them.

Kohonen Map is used to recognize structural market changes (Carlson 2000). Long-term observations are organized as *valuations* and approximate values for the sales prices are found. Actual sales prices are then compared to the sales prices in the *valuations* and the difference is seen as the effect of time. Sales price index is formed and the sales prices are transformed to a given level. The *valuations* organized as a new Kohonen Map give better estimates for the sales prices. The process is iterated once more and the final sales price index is produced.

Structural market changes can be analyzed by computing the average values of the components over time. Strong changes in sales prices may indicate strong structural changes. Cause and effect does interact. Moving average of the sales price differs from the sales price index. Deviations up to 15 percent units were observed in one-family houses in Helsinki capital region. Structural market changes should be taken into account.

7. *GROUP*: NEIGHBORING SALES

Location is something very essential in real estate business. Certain sites are attractive. There may be a beautiful view, shore and forests in the vicinity, good rail and road connections without noise and all kind of public and private services, schools, children's day care, shops and other attractions. Some sites are alone, close to motorways and to high voltage transmission lines. There may be poor connections and services. The quality of the investment may depend on the land value. Comparable sales are found more distant and the effect of location is averaged.

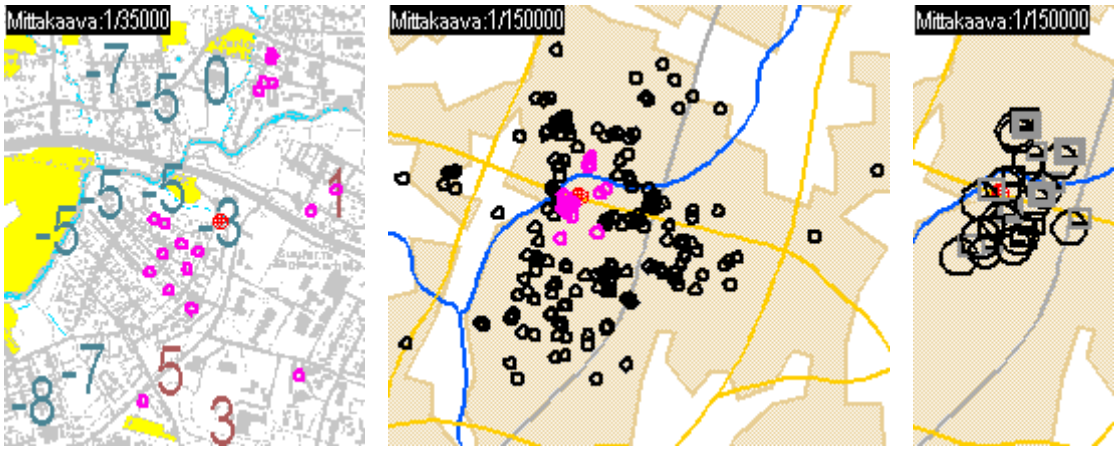

Figure 2. Local sales price undulations (percent) in Suutarila are shown on the map (left) and the neighboring one-family houses of a selected *group* are marked with pink circles. Comparable sales are found in a wider neighborhood (black circles, center). Matching *valuations* are shown (right). There are motorway, two high voltage transmission lines, flight track and boundary along the river within the *group*. Further there are railway and the center of the City of Vantaa in the neighborhood. Kohonen Maps are shown in Figures 1 and 3. © National Land Survey of Finland.

7.1 Local Sales Price Undulations

Sales price of the observation is compared to the sales price of the *valuation* matching the observation. Average ratio and difference in the *group* is computed. Comparable sales may be found in a distance five times compared to the distance of neighboring sales (Figure 2). The difference or ratio (percent) is drawn as *group* symbol and local sales price undulations can be clearly seen on the geographic map. One observation belongs in a hexagonal neighborhood (usually seven units) and the changes are smooth.

Sales price topography can be clearly visualized. Local centers are found. Sales prices may be typically 10 or 20 percent higher in the local centers compared to the surrounding prices and the overall changes can be visualized on the map. Shore, especially the seashore and other attractive sites may be seen as strong positive undulations. Respective negative undulations are found further away and in less attractive sites. Comparable sales are found in a wider neighborhood and all positive and negative effects are averaged.

7.2 Local Market Example

The observations of a *group* selected by pointing the symbol on the map are visualized as a local Kohonen Map. The number of units may be the same as the number of observations (Kaski 1997) and an ordered display of all sales in this neighborhood is shown. Small and old buildings are located in one corner of the local Kohonen Map and large and new houses with amenities and build-ups are located in another. Exceptional cases find their place more or less isolated in the topological order. Observations are browsed using the symbols of the local *valuations* averaged from usually four to seven local observations only.

Scatter plots may be used. Sales price is shown in the vertical axis and for example floor area or year of construction is shown on the horizontal axis (Figure 3). The symbol visualizes the average contents of the resembling observations and the values of the components can be seen in the list. Mutual effects of components may be observed, for example the floor area, the year of construction and the parcel area are usually changing together.

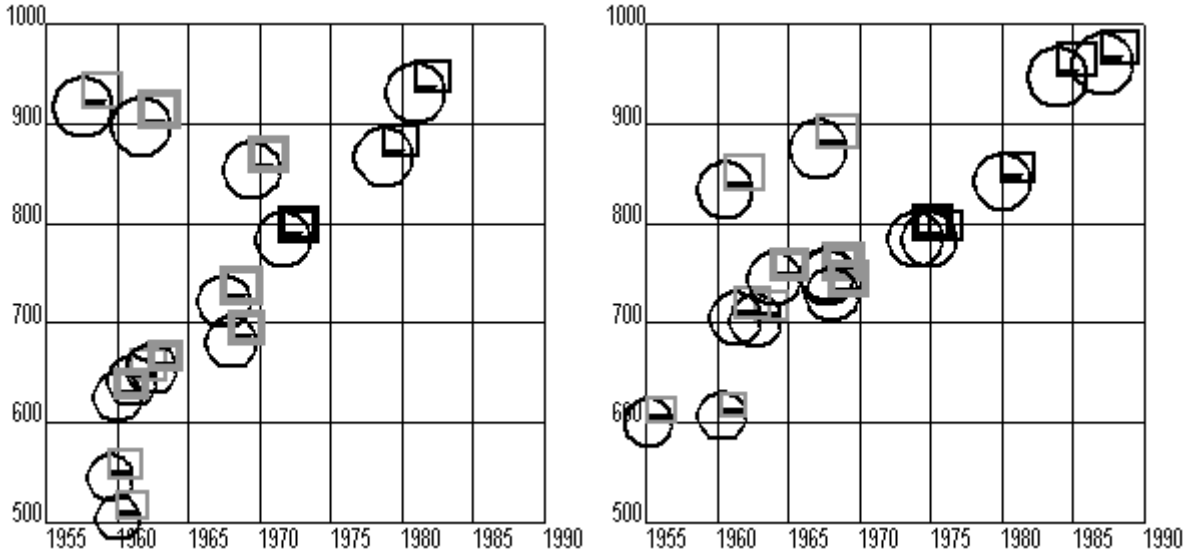

Figure 3. Scatter plots of one-family houses in Suutarila. Sales price is on the vertical axis and year of construction on the horizontal axis. Local Kohonen Map from 19 observations (left) shows a wider range of values for the year 1960 than the 15 *valuations* (right). The area of the rectangle means floor area, thick line means stone and gray means old. Maps are shown in Figures 1 and 2.

8. THE EFFECT OF INDIVIDUAL COMPONENTS

Real properties may change so that only one feature, one component is changing. A new road or high voltage transmission line will be constructed in the vicinity. The effect of an individual component is analyzed by organizing selected observations as a one-dimensional Kohonen Map according to the selected component. Ad hoc methods returning a continuous or discrete numeric value can be used, too. The sales price ratios and differences are visualized as curves or bars and as tables. Overall positive or negative effects shown as local price undulations in the *groups* are eliminated in the comparison.

Useful examples are the distances from a given point or a line. Local center, highway, railway, river and high voltage transmission line are analyzed. Sales date and disturbances in the market can be visualized. Planned railway and motorway may bee seen as positive and negative expectations years before the construction.

Principal components, for example the floor area and the year of construction in the building are usually adapted in the *valuations*. Nevertheless the smallest and largest values found their matching *valuations* inside the observation space and the deviations in the tails can be used to compute average values of the effect in both ends of the observed component values.

9. ONE-FAMILY HOUSES IN HELSINKI CAPITAL REGION

One-family houses are analyzed in period from the year 1985 to 1998 in Helsinki capital region. The 4750 observations are organized as 1428 *valuations* and as 1134 *groups*. Ten components including location, sales price, floor area, year of construction, detail plan, parcel area, shore, building material and heating system are scaled properly for sales comparison. Other components including detailed heating information, number of floors, types of buyer and seller, planned land use and other are analyzed. Distances from main roads, railways and high voltage transmission lines are used and the effects of these components are reported.

Local sales price undulations are computed and sales price topography is visualized on the geographic map. There are 34 rows and 42 columns in the Kohonen Map for *valuations* averaged from typically 17 to 27 resembling observations found in the distance of 952 and 941 meters (rmse) in x-north and y-east coordinates. Typical sales prices in the *valuations* are containing all the effects and characteristics decreasing or increasing the price in their restricted region. There are 27 rows and 42 columns in the Kohonen Map for *groups* averaged from typically 20 to 40 neighboring observations found in the distance of 189 and 188 meters (rmse) in x-north and y-east coordinates. Sales prices of the observations in the *groups* are compared to the averaged sales prices of their matching *valuations* in a region about 25 times larger and local undulations can be seen.

Certain attractive sites, shore, some islands in the sea and some rural lakes are seen as positive undulations of over 15 percent. Local centers are seen as positive undulations of about 10 to 15 percent. Respectively low sales prices are seen in between. Positive and negative undulations are mixed close to the Helsinki-Vantaa airport. Contradictory effects of development, good international connections and traffic noise can be seen. The vicinity of the airport is characteristic for a larger area and the *valuations* indicate the situation more or less. Local undulations may disappear especially more far away from the flight tracks.

Main roads are analyzed by computing the distance from the centerline of the traffic area. Good connections are valuable for residential sites. Land use planning and allocation of the areas for heavy traffic tries to find locations, which minimally conflict with housing areas. An overall negative effect of two percent is seen up to distance of 400 meters. Additional negative effects due to main road seem to be up to distance of 200 meters and up to 15 percent close to the road.

Railways are analyzed by computing the distance from the centerline of the traffic area. Good and fast connections due to commuter traffic are realized in real property values, especially in the vicinity of local railway stations. An overall positive effect of two percent is seen. Negative effects due to railway seem to be up to distance of 200 meters and up to 11 percent close to the railway.

High voltage transmission lines are analyzed by computing the distance from the line. Negative sales price undulations up to distance of one kilometer may indicate that the transmission lines are typically avoiding best housing areas. An overall negative effect of two

or three percent is seen up to distance of 500 meters. This average price difference indicates all positive and negative characteristics in the vicinity of the high voltage transmission lines and this is eliminated in the comparison of the sales prices according to the distance to the line. Within a distance of 60 meters the property values may have a negative effect of up to average 5 percent and two or three percent up to a distance of 150 meters.

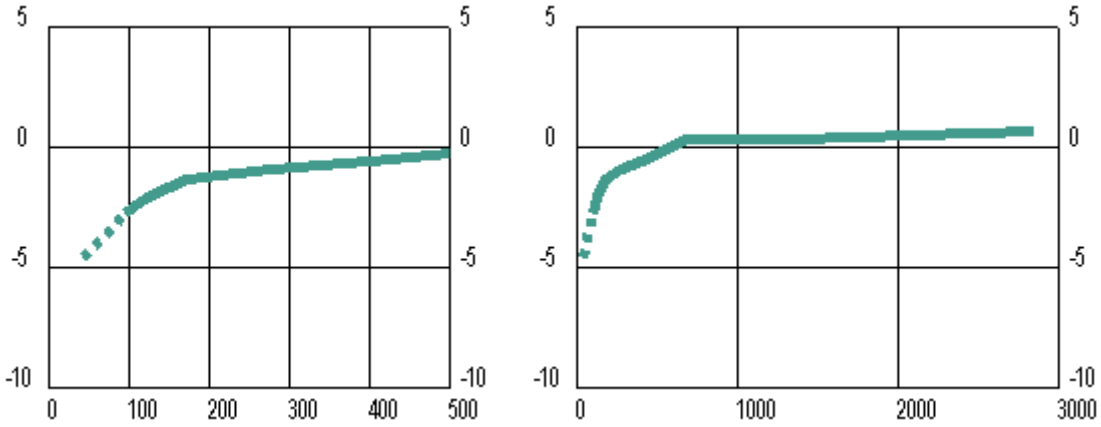

Figure 4. Distance (meters, horizontal axis) to high voltage transmission line and average sales price difference (percent, vertical axis) in one-family houses (4750 observations) in Helsinki capital region. Sales prices are compared to matching *valuations* in narrow regions according to the distance.

10. CONCLUDING REMARKS

Real estate market is analyzed by grouping comparable and neighboring sales using Self-Organizing Map (SOM, Kohonen Map). Comparable sales are organized as a topologically ordered hexagonal lattice using chosen properly coded and scaled components. Mutual information and all positive and negative characteristics in a given location are included in the typical averaged values of sales price and the components. Observations are browsed using the symbols for the units of the Kohonen Map. Resembling observations are found in the topological neighborhood.

Local sales price undulations are computed by comparing sales prices in the *groups* of neighboring observations to sales prices in the *valuations* computed from more distant comparable sales. Sales price topography is visualized clearly.

Real property is a whole. The whole is not the sum of its parts. Individual components are analyzed by using one-dimensional Kohonen Map organized according to a selected component or by a given method returning continuous or discrete numeric value. Overall effect computed as local sales price undulations is taken into account.

Measured components and features are more or less contradictory. Some idea of the dependencies may be given in the ordered display of exceptional cases. Averaged effects are computed in narrow regions for a given feature and visualized as curves or bars and as tables. Individual sales can be looked on the list and on the detail map.

REFERENCES

Carlson, E. (1991): Self-Organizing Feature Maps for Appraisal of Land Value of Shore Parcels. In: T. Kohonen, K. Mäkisara, O. Simula, J. Kangas (eds), *Artificial Neural Networks, Proceedings of the 1991 International Conference on Artificial Neural Networks (ICANN-91)*, (Espoo, 24-28 June 1991). North-Holland, Amsterdam, Vol. 2, pp. 1309-1312.

Carlson, E. (1997): Scaling and Sensitivity in Appraisal. In: *Proceedings of the Workshop of Self-Organizing Maps (WSOM'97)*, (Espoo, 4-6 June 1997). Helsinki University of Technology, Espoo, pp. 57-62.

Carlson, E. (1998): Real Estate Investment Appraisal of Land Properties using SOM. In: G. Deboeck, T. Kohonen (eds), *Visual Explorations in Finance*. Springer, London, pp. 117-127. (Japanese translation: Springer-Verlag Tokyo, July 1999, Russian translation: Alpina Publishing, Moscow, September 1999).

Carlson, E. (2000): Kohonen Maps, Real Estate Market and the Effect of Individual Components. In: H. Hyötyniemi (ed), *Proceedings of the Finnish Artificial Intelligence Symposium SteP 2000*, (Espoo, 28-31 August 2000). Finnish Artificial Intelligence Society, Vol. 3, pp. 227-234.

Deboeck, G. (1998): Investment Maps of Emerging Markets. In: G. Deboeck, T. Kohonen (eds), *Visual Explorations in Finance.* Springer, London, pp. 83-105.

Kaski, S. (1997): *Data Exploration Using Self-Organizing Maps*. Acta Polytechnica Scandinavia. Mathematics, Computing and Management in Engineering Series No. 82. DTech Thesis, Helsinki University of Technolgy, Espoo.

Kiviluoto, K. and Bergius, P. (1998): Maps for Analyzing Failures of Small and Medium-sized Enterprises. In: G. Deboeck, T. Kohonen (eds), *Visual Explorations in Finance.* Springer, London, pp. 59-71.

Kohonen, T. (1995): *Self-Organizing Maps*. Springer-Verlag, Berlin.

BIOGRAPHICAL NOTES

Eero Carlson obtained an Msc in Engineering at the Helsinki University of Technology in Finland in 1970. Since 1973 he has worked on the development of a finnish Geographic Information System (GIS) and since 1990 neural networks. His main interests are real estate appraisals with Self-Organizing Map (SOM, Kohonen Map) and the integration of SOM and GIS. Mr. Carlson is author of about 15 publications on neural networks.