6 Analysing locational externality effects with the SOM: power line proximity and residential property prices[1]

In Chapter 5 the SOM was used for housing market analysis, as an alternative – and partly as a complement – to hedonic price models. It was shown that the method has added value in terms of isolating the effect of omitted variables and capturing market segmentation. The crucial difference from a more conventional approach was noted: the neural network adds a piece of qualitative thinking to the topic – even the marginal cases emerge.

The idea here is to demonstrate the potential of the SOM-based method for the impact analysis of a local externality capitalisation, the more locationally-specific variant of the SOM-based method of house price analysis. The exercise documented deals with the effects of proximity to high-voltage power transmission lines on single-family residences. It relates to the main study in two different ways:

(1) The power line represents a specific case of the effect of use restrictions on house prices (institutional aspect) and the indirect effect of the quality of the view (behavioural aspect) in the case of powerline proximity (see Section 3.1).

(2) The exercise demonstrates an alternative application based on the SOM, a comparison to a (hedonic) regression model and a suggestion of a hybrid model, combining the SOM with linear regression analysis (see Section 4.2 and App. E).

The aim of the exercise is to trace the effect on property value empirically within the neural network approach outlined in the previous two chapters. Furthermore, the study also tries to compare the approach with a more conventional method based on multiple regression analysis. As the measure of performance, we use the ability of each method to capture the assumed negative impacts on property values (as measured by transaction price) caused by proximity to a transmission line (as measured by distance and visibility).

The dataset comprises single-family property sale transactions during 1993-1997 in two Finnish towns, Jyväskylä and Järvenpää. The dataset allows the isolation of the price effect of the power line disturbance in accordance with the principles of hedonic price analysis – that is, through the market choices made in a price equilibrium by rational house purchasers with uniform preferences. The dataset is the same as that used in a regression-based study by Peltomaa (1998; with follow up in 2001). This fact permits comparison of our results with those previously obtained by Peltomaa in the study referred to above.

The chapter consists of five sections. Section 6.1 is a review of the literature on price impacts attributable to the proximity of a power line. Section 6.2 presents the analyses of the data of the target areas. Section 6.3 comprises a summary of the exercise and the conclusions drawn from it.

6.1 Power line proximity impacts on property value

6.1.1 Types of impact

The literature on power line compensation cases is largely in agreement on the determinants of property value. In a general sense, the nature of the power line disturbance may be either *aesthetic* (through a change in the landscape), or *psychological* (through health risks perceived by residents living nearby) (Delaney & Timons, cited in Peltomaa 2001). From an (urban) economics point of view, the nuisance a property owner suffers from a nearby power line is a clear case of a negative local externality effect and is potentially quantifiable in monetary terms.

According to the literature, proximity to a transmission line can reduce the value of a property in several ways (without having to distinguish between essentially aesthetic or psychological factors). Losses caused by a transmission line – on unimproved land with the potential for development – may be divided into two parts according to the nature of the loss (Dempsey 1981, p. 382-383 and p. 388):

(I) development or use damage is a direct reduction in the market value of a tract of undeveloped land through a decrease in the optimum development or use potential resulting from the acquisition, and/or an increase in the cost of developing the remaining land to its optimum use after the acquisition;

(II) proximity damage is an indirect reduction in market value of the property through proximity of the transmission line. This impact on property values can be divided into three interlinked parts (Kinnard & Dickey 1995, p. 24; Blinder 1981, p. 14-2):

(a) diminished price: the most obvious effect; it can be observed by comparing transaction prices paid for otherwise similar residences in and outside the vicinity of the transmission line; *(b) increased marketing time:* according to a frequently stated claim, this delay entails a real financial loss to the vendor; *(c) decreased sales volume:* diminishing sales could provide evidence for a decrease in the value of the residences brought about by the proximity of a transmission line.

Additionally, indirect monetary losses are caused when: (d) *a substantial part of the plot* is within the right of way of a power line; (e) *cheap residences* are built on plots within the right of way, or there are price concessions for such residences[2] (Clark & Treadway 1972, p. 20).

These points exemplify how property value is influenced negatively by a transmission line. The influence of the effect varies, depending on whether the cause is a direct restriction, or is proximity related. In the latter case, a distinction may be made depending on whether the relevant indicator is price reduction, marketing time, sales volume, share of plot within the right of way, or the price setting of new house building within the right of way. Compensation may also be justified on the basis of either direct substantial losses or indirect proximity-related losses in property value. The price is therefore only one of the factors affected by power line proximity.

6.1.2 Findings about the magnitude

Ten studies on the impact of power line proximity were briefly described by Peltomaa (1998, p. 4-8). Four of the studies were Finnish and six from the US. In one American paper (Kinnard et al. 1997), several studies from North America and New Zealand were summarised; the remainder of studies reported were individual cases. The studies (cited in Peltomaa 1998; Delaney & Timmons cited in Peltomaa 2001) are compared in Table 6.1.

As can be seen, in the sample of studies the methods differ and so do the results; no relationship can be found from the literature survey between approach (market data, or questionnaire) and findings (price reduction, or no effect). Deriving a homogeneous range of estimates from the studies is also difficult, because the operationalisation of the power line proximity effect varies so widely. Furthermore, when moving away from the transmission line, the possible impact seems to disappear quite rapidly. A substantial effect was only present below an estimated power line distance threshold which, depending on the study, lay between 15 and 90 metres. However, on the basis on findings from these studies, Peltomaa (1998, p. 8) concluded that there appeared to be a connection between proximity to a transmission line and a decrease in property value, but nothing certain could be said about the intensity of the relationship.

In his own study, Peltomaa (1998, p. 72-73) obtained no statistically significant, logical evidence supporting the hypothesis of a decrease in property value related to proximity to a transmission line. For possible policy implications, he suggested that compensation for value reduction determined for third parties (that is outside the expropriation relation) should be considered with reservation. There seemed to be some disagreement concerning the quantified extent of the effect and the reliability of the evidence.[3] Another finding – one not reported explicitly in his study – was that the influence of a power line on property values seemed to operate over a greater distance and had a flatter functional form than the results obtained in earlier studies indicated (Cajanus 1985, in particular).

Table 6.1 Comparison of studies on the impact of a power line on property value. (Sources: Peltomaa 1998; 2001)

Year and authors F=Finnish study; A= American study	Methods used	Findings
1972 Clark and Treadway (A);	Case studies (on sales transactions)	A significant price reduction only for raw residential land and for small commercial estates
1979 Colwell and Foley (A)	Regression models	The power line had no effect on the prices of single-family property above a distance of 60 metres, and a significant effect only below a distance of 15 metres
1981 Blinder (A)	Statistical tests and regression models	A small impact of the power line on the sales price of single-family property; price reductions of 2% were reported for properties with a 'tower behind the backyard' and reductions of 1% for 'other abutting lots' compared with 'non-abutting lots'
1981 Holmström (F)	?	The value of the area below the power line is 40-60% of a normal zoned area
1985 Cajanus (F)	Regression models	A significant price impact only for plots situated less than a distance of 30 metres from the power line
1986 Virtanen (F)	Analysis of the grounds for compensation	Similar to Holmström's study
1990 Colwell (A)	Regression models (same data as Colwell and Foley, with added variables)	Three results: (1) power line proximity has a negative impact on price, weakening with time; (2) having an easement clearly reduces the price; (3) a power line also has an influence on property prices if they do not have an easement
1992 Kung and Seagle (A)	Comparison of single-family property transaction prices followed by a questionnaire sent to the buyers	The comparison of prices did not show a price effect; according to the questionnaire, 53% of the respondents considered the power line a scenic drawback (however, 72% of these did not consider this had affected the price they paid), none of them considered it a health risk.
1992 Delaney & Timmons (A)	A questionnaire survey to property valuers	Reductions as high as 10% of the price were related to power line proximity
1997 Kinnard, Bond, Syms and DeLottie; Kinnard, Geckler and DeLottie (A)	Literature review of several studies from the US (incl. some cited above), Canada and New Zealand + a separate empirical study from Las Vegas (4269 transactions) and St Louis (1377 transact.)	Literature review: some negative impact below distances of 60-90 metres; empirical study: a 1.3-1.4% negative price effect for properties situated within 800 m.of a power line in Las Vegas, but not in St. Louis (possible reason is a more open landscape in Las Vegas)

(Continued)

Year and authors F=Finnish study; A= American study	Methods used	Findings
1998, 2001 Peltomaa (F)	Multiple regression analysis of two datasets: (1) a nation-wide set consisting of transactions during 1993-1997 (52,474 obs.); (2) a qualitatively better, but quantitatively scarcer sample from two towns in Finland: Jyväskylä (42 obs.) and Järvenpää (26 obs.), (see Table 6.2); (3) a questionnaire to owners in the sample (2)	(1) In most submodels the regression coefficients had an illogical sign; (2) the power line did not show a statistically significant price effect for any submodel of the target areas (without pooling the dataset); this finding applied to distance and visibility factors alike; in the Järvenpää areas the price reduction seemed more substantial than in the Jyväskylä areas, where factors other than power line proximity determine price; (3) no support for a hypothesised power line disturbance effect

The empirical literature is ambiguous about the sign, strength and nature of the effect. In this kind of setting, the neural network technique is often introduced as an alternative way of modelling the price effect (see App. E). The performance with the same dataset is then compared with the benchmark performance achieved by the multiple regression models.[4]

6.2 The analysis of the target areas

An SOM-based analysis was then conducted with the data for the target areas. Three datasets were used: one for each town, and one pooled dataset. The data is first discussed (Subsection 6.2.1). The processing with the SOM (6.2.2) and the LVQ classifier to determine the relative importance of the power line (6.2.3) is then described. Finally, price associations are reported which were identified through visual interpretation of the SOM-output and also on the basis of post-processing with regression analysis (6.2.4).

6.2.1 The data

The basic dataset consisted of transactions of single-family houses and plots from January 1993 to April 1997. The transactions were sampled from the Real Estate Purchase Register maintained by the National Land Survey of Finland (NLS)[5].

Not all the factors were obtained directly from the registers. The Land Information Centre of NLS calculated the shortest distance to the nearest high-voltage power line for each transaction from the coordinates of the centre of the property. The necessary power line data were taken from the numerical power line map maintained by Suomen Kantaverkko Oyj (Fingrid). With respect to the accuracy of the distances, please note: (1) as the measured accuracy for the centre of the property, 10 m. was given. However, this accuracy can only be considered reliable in planned areas (town or building plan). (2) In the numerical power line map the company's 'own' lines were digitised from the general map on historical grounds, in which case their accuracy was estimated to be +/- 20 m.. The accuracy of the remaining lines might be substantially poorer.

Suitable target areas were then selected from the nationwide parent dataset. The idea was to search for two residential areas which were as homogeneous as possible and where several transactions concerning single-family property in the proximity of a power line had occurred.[6] In both cases, the power line had been in place for a long time before the transaction took place. Since the aim of the study was to clarify the pure impact of a power line on property value, another criterion for selection was set up; no areas bordering major traffic routes or other significant sources of interference were included. The group of property transactions marked by the power line proximity (= the observations within 500 m. of the power line) was displayed on a map base on which the numerical power line sample was also shown.[7] Two suitable town plan residential areas from each towns were included in the final examination of the target areas:

in Järvenpää:

adjacent subareas Pajala and Sorto (near Helsinki)
adjacent subareas Jamppa and Peltola;

in Jyväskylä:

adjacent subareas Ristonmaa and Ristikivi (a relatively newly built area)
adjacent subareas Kuokkala and Kuokkalanpelto (built in the 1960s).

For the plots with a house, additional information about the building was obtained from the Building and Apartment Register kept by the Population Register (VTJ). Unbuilt plots were omitted from the analysis, because they were few in number. Finally, the target areas of Järvenpää included 26 and the target areas of Jyväskylä 42 observations. The descriptive statistics of the sample are presented in Table 6.2.

Table 6.2 A sample of single-family property transactions from the selected target areas from two Finnish towns. Descriptive statistics, 1993-1997

Variable	Mean	SD	Minimum	Median	Maximum
Continuous					
(1) Price	507985	156142	180000	500000	950000
(2) Gross area	132.64	40.313	42.000	132.00	232.00
(3) Price/Gr area	3958.8	1026.7	1761.3	3902.4	7857.1
(4) Net area	110.51	26.767	28.000	111.50	180.00
(5) Plot area	1029.9	385.62	477.00	921.50	2200.0
(6) Permitted	223.97	60.804	137.00	202.00	398.00
(7) Month	30.265	1.0000	15.413	31.500	53.000
(8) Age	20.971	16.357	0.0100	13.500	67.000
(9) Distance/Line	180.90	97.740	24.000	192.00	427.00
(10) Distance/P yl.	190.78	94.436	39.000	193.50	435.00
(11) Visibility	1.4559	0.6564	1.0000	1.0000	3.0000
(12) X	6.829E+06	93820	6.710E+06	6.902E+06	6.903E+06
(13) Y	3.421E+06	20334	3.394E+06	3.435E+06	3.438E+06
Discrete	N				
(14) Econ.build.	22				
(15) Water	66				
(16) Drain	66				

Variable descriptions
(1) Total transaction price (FIM)
(2) Gross floor area (square metres)
(3) Total transaction prcie/gross floor area
(4) Net floor area (square metres)
(5) Plot area (square metres)
(6) Permitted gross floor area (square metres)
(7) Month of the transaction (1=January 1993)
(8) Age of the reseidential building at the moment of transaction (years)
(9) Distance from the power line (metres)
(10) Distance from the nearest pylon (metres)
(11) Visibility of the power line (in scale [0,1])
(12) X: south-north –coordinates (metres)
(13) West-east –coordinates (metres)
(14) Economy building(E.g. shed, dummy)
(15) Water pipeline (dummy)
(16) Wastewater drain (dummy)

The observations from the target areas contained information from the VTJ about the buildings, and through terrain investigations about the line visibility (not visible, partially visible, totally visible) gained, and considerably more accurate power line distances. The distances were measured from a 1:2,000 scale base map. In Jyväskylä, the map extract was a print from the numerical map; in Järvenpää, the map was a conventional copy. The estimated error of a measurement compared with the actual distance was +/-2m. The distance was measured from the house wall closest to the power line to the center line of the power line or nearest pylon.

The accuracy of the power line proximity measures was analysed empirically within the selected areas in Jyväskylä and Järvenpää. The accuracy was observed (with a 99 % confidence interval) to be +/-14 m. in Järvenpää and +/- 16m. in Jyväskylä.

In the target area(s) of Jyväskylä, the residence closest to the power line was situated at a distance of 24 m. from the centre line, and at a distance of 39 m. from the pylon. The residences furthest from the power line were situated at a distance of 427 m. from the power line and 435 m. from the pylon. Either the power line or the pylon was clearly visible from four of the residences, and visible to some extent from a further eleven of them.

In the target area(s) of Järvenpää, the minimum distances were 28 m. from the power line and 51 m. from the pylon; the maximum distances were 336 m. to both power line and pylon. From two of the residences either the power line or the pylon was clearly visible they were visible to some extent from eight of them.

The target areas did not contain plots with a transmission line right of way on (that is, they were not in the closest theoretical proximity to the line area). From examination of the map, it would seem that not even the building restriction area of the power line reached any of the plots. As a final observation, Peltomaa (2001, p. 34-35) noted that, in the Jyväskylä-locations, the view from the garden of the house was blocked because the terrain contained greater differences in altitude and more vegetation coverage than in the Järvenpää case (cf. findings reported in Table 6.1).

6.2.2 Processing with the SOM

As already explained in chapters 4-5, neural network processing usually needs certain technical parameter adjustments. *Coding* refers to a preprocessing of the input data in such a way that the effect is measured in the most convenient manner. There are several examples of different ways of coding (see e.g. Evans et al. 1992). The coding should be in harmony with the character of the algorithm. For running the SOM a straight-lined metric distance (D) was inverted to a simple line effect ratio RRD = 100m / D (RRD = the reverse ratio of the distance). The proximity effect can be more readily perceived this way. Close to the power line the effect is substantial, decreasing rapidly until a distance of 100 metres away from it is reached. From 100 metres onwards it is assumed that, further decrease in value is only marginal. Alternatively, the straight metric distance was used.

As *scaling* has an impact emphasising of the variable in the organization process, the variables were initially normalised on the scale from 0 to 1. For generating maps of the target areas (Jyväskylä and Järvenpää areas combined, Jyväskylä areas alone, Järvenpää areas alone), the network parametres were chosen as follows:

Combined (68 observations[8])

size of the map	54 neurons (x=9, y=6)
neighbourhood function	bubble
observations/neighbourhood (approx.)	8.8
iterations, basic/fine-tune	2,700/27,000
initial learning rate, basic/fine-tune	0.04/0.01
initial radius, b./f.-t.	10/3
label (unweighted variable)	town and target area, house characteristics
quantization error (RMSE)	.465

Jyväskylä (42 obs.)

size of the map	24 neurons (x=6, y=4)
neighbourhood function	bubble
observations/neighbourhood (approx.)	12.3
iterations, basic/fine-tune	1,200/12,000
initial learning rate, basic/fine-tune	0.02/0.005
initial radius, b./f.-t.	7/3
label (unweighted variable)	target area, house characteristics
quantization error (RMSE)	.526

Järvenpää (26 obs.)

size of the map	12 neurons (x=4, y=3)
neighbourhood function	bubble
observations/neighbourhood (approx.)	13.0
iterations, basic/fine-tune	600/6,000
initial learning rate, basic/fine-tune	0.02/0.003
initial radius, b./f.-t.	5/2
label (unweighted variable)	target area, house characteristics
quantization error (RMSE)	.511

The capability of generalisation can be seen to be below the recommended 'rule of thumb' 20 observations/neighbourhood (see Section 4.3).

The resulting feature maps are illustrated as 1-d diagrams in the figures 6.1-3 below regarding the two layers for price (either total or per sq. m.) and the three layers for power line proximity (either line effect, line view, or pylon effect). For the time being, let us note here that the analysis needs to be split into two stages (as was the case in Chapter 5). To begin with, *segmentation* of the data with respect to the power line proximity factor is dealt with in Subsection 6.2.3. As the actual (2-d) feature map layers were not very informative, this analysis is based on the LVQ classifier – a method that so far has merely been proposed as a tool for checking the SOM analysis. By assigning two different types of labels to each observation and corresponding node, I show, how this method can be used for comparing the discriminative strength of two factors that are assumed relevant for the organisation of the dataset: the power line visibility variable and the locational identification.

After that, the *price associations* of the data is studied in Subsection 6.2.4. First, the simple bivariate diagrams illustrated in the figures above are interpreted. Subsequently, the resulting matrix is processed as smoothed data with a linear regression technique, as explained in Section 4.3. In this way, the price analysis becomes more rigorous in terms of quantitative statistical analysis and capable of comparison with benchmark results obtained regarding marginal adjustment factors (the monetary worth of the change in a given value factor, in this case a power line effect) .

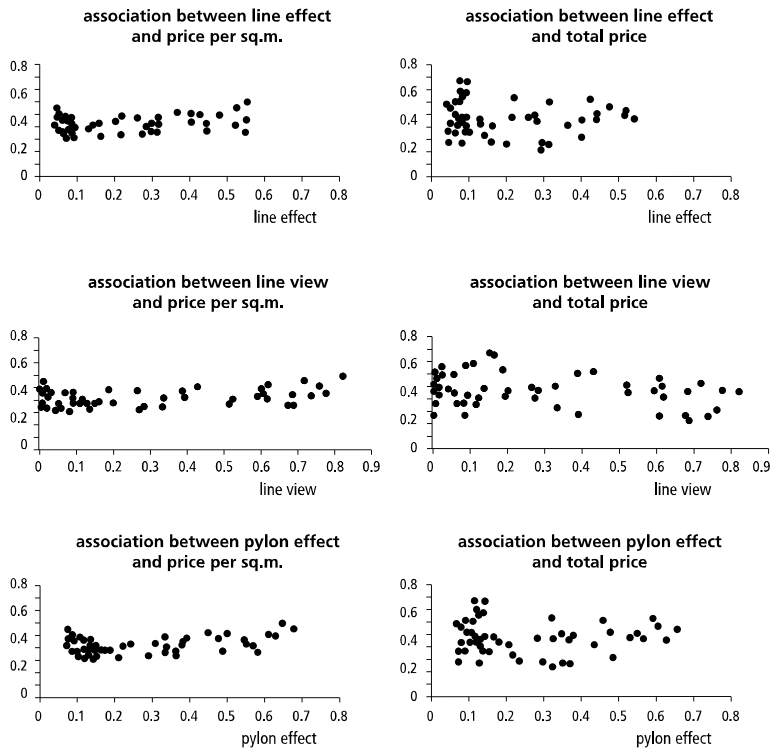

Figure 6.1 The association between proximity and price measures illustrated with processed data from the target areas in both towns. (Each observation (dot) is a neuron resembling a typical combination of property characteristics.)

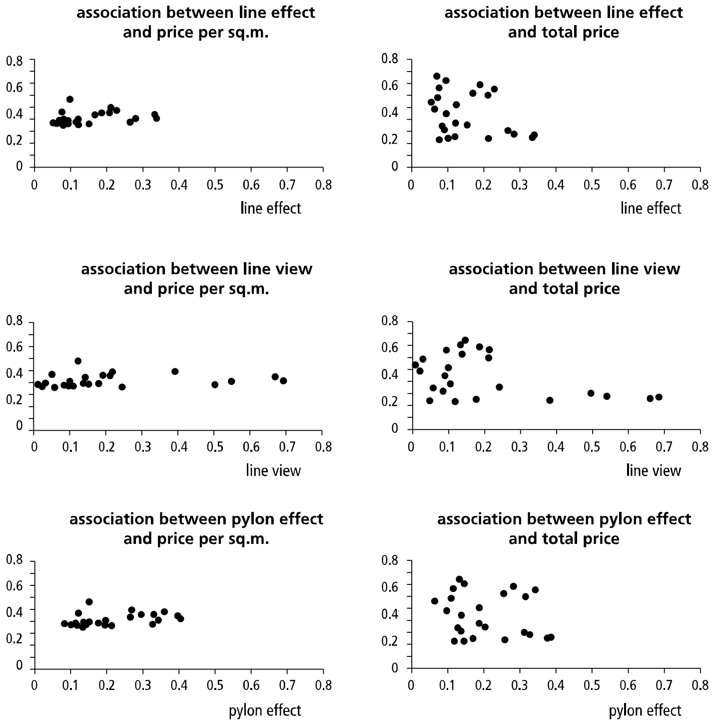

Figure 6.2 The association between proximity and price measures illustrated with processed data from the Jyväskylä target areas.

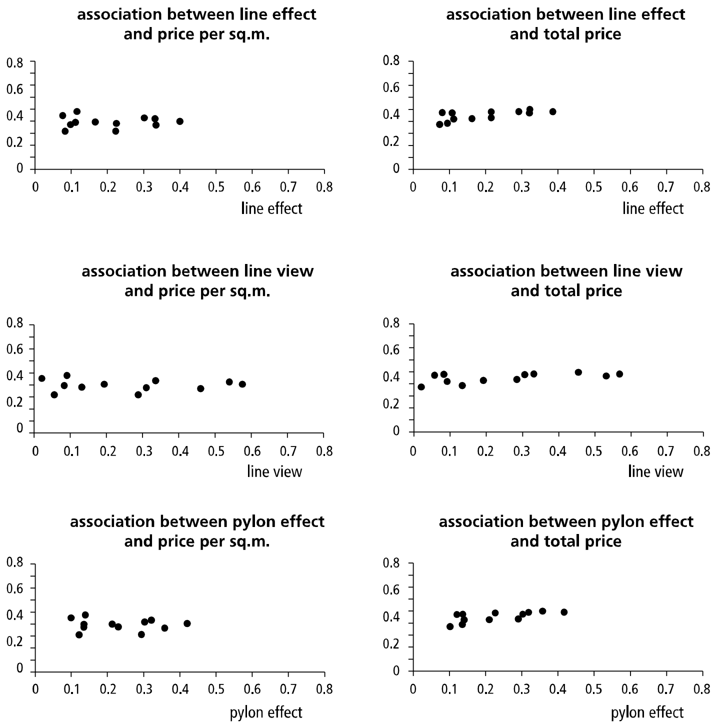

Figure 6.3 The association between proximity and price measures illustrated with processed data from the Järvenpää target areas.

6.2.3 Testing the segmentation with the LVQ

The six diagrams (in Figure 6.1 in particular) show a clustering of neurons based on the power line proximity effect. An interesting question now arises: what is the relevance of the power line factor for the organisation of the data in comparison with another important factor, such as location, which in the pooled sample ought to be of particular importance[9]. In the selected samples from the two towns, the resulting maps were therefore next post-processed with the LVQ in order to get some evidence about the relative importance of two discrete factors contributing to the clustering: location and visibility. However, the use of the LVQ was restricted here to its unsupervised classification properties (see earlier definitions in chapters 4-5).

Next the percentage ratio providing information about the classification performance (that is, the classification accuracy or the recognition accuracy, see App. G) was computed. As explained in Chapter 4 the classification accuracy is an alternative measure to Q when the aim is to evaluate the goodness of the maps. The classification accuracy tells us how large a proportion of the observations on average hit the corresponding nodes of the unsupervised map. In its unsupervised mode, the LVQ algorithm compares the label assigned to each observation with the label of the corresponding winner node in the map over the total sample. In theory, the classification may be correct (that is, it corresponds with the labels of the calibrated map) for all observations, in which case the classification accuracy is 100%. In such a case, the codebook vectors of all the labelled nodes can be distinguishable from each other, which means that, in practice, the categories used are mutually exclusive. When the classification accuracy is less than 100%, the categories do not differ enough in the composition of the input variables to be recognised by the algorithm. In other words: the observations of the dataset are 'too similar' and the classification task becomes too difficult for the algorithm.

In this case, the labels were simplified and used as *a priori* classes for the observations. We used two dummy variables as labelling criteria: first, a two-valued locational area dummy and second, a three valued line view dummy. In this way, the labels correspond to crucial information about each observation regarding these two attributes: location and power line view. The locational dummy corresponds to the town (Jyväskylä or Järvenpää) in the combined sample, and to the target area within each town in the two town-specific samples. The line view dummy is the variable (11) in Table 6.2 (see also figures 6.1-3), which has three values: no visibility (0); partial visibility (½); full visibility (1).

It is relevant to the research design that the other labelling criterion has three values while the other has only two values. The underlying idea is that the dummy with the higher classification accuracy would be the more substantial labelling criterion and thus the feature which better describes the real preferences dominant in the area. Furthermore, the locational dummy is expected to have a better classification accuracy, because it has fewer values than the line view dummy.[10] Hence, if visibility obtains a better classification accuracy than location, visibility is definitely the more dominant of these two features. The following classification accuracies were obtained:

Combined

	location (Jkl/Jpää)	100 %
	visibility (0,½,1)	92.65 %

Jyväskylä

	location (two areas)	80.95 %
	visibility (0,½,1)	85.71 %

Järvenpää

	location (two areas)	65.38 %
	visibility (0,½,1)	88.46 %

Additional runs were made with double scaled (that is, field range from 0 to 2) line effect, pylon effect, and line view variables. Three new feature maps were generated, with the three variables in question particularly emphasised in the organisational process of the network. The visibility label is then expected to show higher classification accuracies than above, and if the area location is related to power line proximity in a meaningful way, the locational classification would be improved as well (in the combined sample it is of course impossible to improve from 100%). In practice, this scaling procedure means that the purchaser's attitude towards the line proximity variables has been overemphasised. The following new classification accuracies were obtained:

Combined

	location (Jkl/Jpää)	100 %
	visibility (0,½,1)	97.06 %

Jyväskylä

	location (two areas)	78.57 %
	visibility (0,½,1)	100 %

Järvenpää

	location (two areas)	76.92 %
	visibility (0,½,1)	96.15 %

Some cautious conclusions can be drawn from these statistics about the classification of the observations based on the location and the line effect:

Combined

initially, the (macro) location seems to be the most important criterion
when the proximity effects are emphasised, the line effect attains the same relative significance as the (macro) location; the conclusion is that both factors are important.

Jyväskylä

the line effect is more important than the location

when the proximity effects are emphasised, the locational classification accuracy decreases; the conclusion is that the location (defined by the boundaries of the target areas) is not associated with the line proximity variables.

Järvenpää

the line effect is definitely more important than the location

when the proximity effects are emphasised, the locational classification accuracy increases; the conclusion is that location seems to be indirectly associated with the line proximity variables.

After a post-processing of the target area feature maps with the LVQ classifier, some conclusions can be drawn. The city in which houses are located is obviously the most dominant effect. Nevertheless, the proximity effect of power line is also undoubtedly an important determinant of the organisation of the data when the sample is restricted to cases where an impact is expected and the proximity factor is compared with the more general location factor. In all three cases the classification, and thus the clustering of the maps, was strongly associated with the proximity variables.

It is of interest to note that the nature of the association between location and proximity to a power line was different in the two target areas. In the Järvenpää areas these two factors appeared to ‘proxy’ for each other, which means that in this case a substantial proportion of the locational value is contributed to the perceived proximity to the power line. In the Jyväskylä areas, according to these results location has a more independent - role, with other factors contributing to the locational value and the effect of the power line being more site (or house) specific than neighbourhood specific in nature.

Some conclusions can be drawn from the results reached so far. First, when related to location, power line proximity (operationalised through 3 of the 16 input variables) is an important general feature contributing to the organisation of the feature map. Second, the relationship between powerline proximity and location differs in the two target areas presumably because of the terrain factor, in the Jyväskylä sample the nature and magnitude of the effect is related to each individual house rather than being a proxy for area location, whereas in the Järvenpää sample the opposite is the case: power line proximity is clearly related to the vicinity or whole neighbourhood in question. However, the question about the magnitudes of the power line impact still remains unanswered.

6.2.4 Looking for price-associations and post-processing with multiple regression analysis

To be able to draw conclusions about an association between house price and power line proximity, we return to the figures 6.1-3. The main results of the visual analysis of the feature maps are given below, using the same idea as in Fig. 5.6, namely to focus on the association between the three power line variables and the two price variables for the combined dataset and for each town-specific dataset in a one dimensional presentation of the map.

Combined sample (see Figure 6.1)

Contrary to expectations: at least there is not negative association between the three line proximity variables (line effect, pylon effect, line view) and the price per sq.m. of the property. There is possibly a small positive effect: the more intense the power line effect, the higher the price per unit.

A small logical association was found between the line view (but not the line effect, or the pylon effect) and the total price: the cheapest nodes also have a partial or full line view (a typical line view value is 0.7).

Jyväskylä (see Figure 6.2)

No association was found between any of the proximity variables and the price per sq.m. (although three distinct clusters emerged on the basis of their proximity effects: groups of nodes with no power line proximity effect, with some effect, and with a clear effect)

Partial association: for the strongest proximity effect the total price is always low and for the two most expensive nodes the proximity effect is relatively weak.

Järvenpää (see Figure 6.3)

No association was found between any of the proximity variables and the price per sq.m. (although the most expensive observations are situated furthest from the line or the pylon, they are also situated in the area closer to Helsinki)

Small illogical associations were found between the proximity variables and the total price: for all three power line variables (on the right of the diagram) there seems to be a linear relationship between an increase in power line proximity effect and an increase in price.

Visual interpretation of the map layers left in doubt the possibility of any association between the proximity and price variables within the data. In two of the six cases a speculative effect was detectable (combined/total price and Jyväskylä/total price). However, as the remaining four cases show no effects whatever, the results are still far from convincing that the SOM has managed to capture the hypothesised power line proximity effect. Besides, the visual analysis of the feature maps also suffered from some interfering factors. In both towns, the observations with the greatest proximity effect also had the smallest plots and thus the highest prices per unit. Another interfering effect was caused by the structural characteristics of the houses. In some cases they were the only *real* factors determining the price level of the property in question. There was no straightforward interpretation of the analysis, which is why post-processing of the neurons with OLS multiple regression analysis was undertaken as a final attempt to obtain some confirmation of the presence of a proximity effect.

Linear, loglinear and exponential semi-log models were tried – a standard procedure when the theoretical justifications are not clear (see Subsection 3.1.1). Five measures for power line effect were tried (straight, or reversed, distance to the power line or to the pylon, and visibility) and two measures for price. In this way we obtained 30 (=3*5*2) different price models of the selected target areas. Only the combined Jyväskylä/Järvenpää -case is concerned; the two town-specific cases did not contain a sufficient number of observations (= neurons) for reliable post-processing with linear regression analysis.[11]

In only 15 models was the price effect found significant. The differences between the magnitudes using straight line and reversed distance metric were substantial: even 5% of the average selling price (FIM 508,000, Table 6.2), using the calculations based on a 200 m. hypothetical distance effect described below. This variation is somewhat suspect, given the relatively small (maximum 400 m.) interval to the power line. Similarly, the calculated magnitudes for price changes differed between the models where total price and price per sq.m was used as the dependent variable. Therefore, the interpretation of the results is simplified by only showing six of the models (that is, the linear and exponential models with straight metric distance or visibility used in combination with price per sq.m.) are presented in Table 6.3.

Only the coefficients for the power line effect are presented. When the straight line distance to a power line or pylon is considered, the coefficients (a) of the models are expected to have a positive sign, whereas (a) is expected to have a negative sign in the models where visibility is used. Furthermore, when estimating the price effect of a marginal change in a proximity variable, the linear models yield amounts in FIM, whereas the exponential models yield the amounts in percentage terms. (The price changes estimated by the models can clearly be related to each other, as shown below.) Additional models were built with a sample containing twofold scaled proximity variables (see Section 6.2.3), but no statistically significant coefficients were obtained for them.

Table 6.3 Price models of the selected target areas with smoothed data

MODEL DESCRIPTION	COEFF. (a)	P-VALUE (a)
x = distance from the nearest pylon in metres		
Linear model, $a*x$, dep. var.= transaction price/gross floor area	2.02347	0.003
Exponential (semi-log) model, e^{x*a}, dep.	4.162E-04	0.005

var.= transaction price/gross floor area		
x = distance from the power line in metres		
Linear model, $a*x$, dep.var.= transaction price/gross floor area	2.54777	0.0001
Exponential (semi-log) model, e^{x*a}, dep. var.= transaction price/gross floor area	0.6225	0.0001
x = visibility of the power line in scale [0,1]		
Linear model, $a*x$, dep.var.= transaction price/gross floor area	-612.182	0.0009
Exponential (semi-log) model, e^{x*a}, dep. var.= transaction price/gross floor area	-0.14966	0.001

Because the capability of the network for generalisation in this case was modest (see 6.2.2), the 'compression' of the variables was minor. For instance, the field range of the line distance was between 42.11 m. and 248.5 m. with the smoothed data, while it was between 24 m. and 427 m. with the original data. (In order to interpret the results sensibly, the normalised price and line proximity variables were converted back to their original scale.)

Each line proximity variable was tried separately in the models, because the three line proximity variables were strongly multicollinear. The line proximity variables did not, however, correlate too strongly with other variables, so their coefficients can be considered fairly reliable even when detached from the models. (The variance inflation factors were low.)

The table shows that the regression coefficients are logical: positive and negative where expected. The real question of interest is *how substantial was the observed price effect of the line proximity and line view*. To answer that question we reconstruct a hypothetical situation and estimate the price changes for the 15 models. If we move from a 250 m. distance to a 50 m. distance from the power line or pylon, or if we have at least partial visibility, the estimated price reduction falls in the range of 5-14%, in relation to the average selling price. As noted above, the exact magnitude of the results depends on the functional form and which of the five proximity measures are used. Even so, these results did not yield absurd figures for any of the transformations. Also, in comparison with the findings reported in earlier studies (see Table 6.1), we may conclude that in the selected target areas the detected power line effect was substantial, but not unreasonable.

The original analysis by Peltomaa (1998) was based only on separate models for Järvenpää and Jyväskylä. However, encouraged by the good results reported above (and in Kauko & Peltomaa 1998), Peltomaa (2001, p. 19-35) subsequently conducted a new regression analysis with pooled data from the target areas. This analysis was the third attempt to quantify the power line proximity effect using the same basic dataset (i.e. Peltomaa 1998; 2001; and the study documented in this section). On this occassion, the results were similar to the analysis of the SOM-smoothed data above: logical and not too mutually contradictory. However, while the variation across models was still substantial, none of the results obtained in these studies are reliable enough to give a *clear* answer to the question, how much does proximity to the power line reduce property prices. Besides, as noted earlier, the nature of the effect differs i the two towns.

6.3 Summary, conclusions and further discussion

The chapter has made an attempt to address the classic problem of quantifying externality effects. One specific type of localised externality effect, the power line disturbance, was chosen for analysis. *Is the effect negligible or substantial in a particular spatial context?* Furthermore, the performance of a neural network approach was compared with that of the benchmark method standard multiple regression analysis. *Is there anything more to say about the method on the basis of this exercise*, in addition to the general conclusions presented in chapters 4 and 5 above?

The method involved visual interpretation of the feature maps, post-processing of the neurons with the LVQ classifier, and post-processing with OLS multiple regression analysis. Some coding, scaling and labelling routines were also tried in an attempt to obtain relevant findings with regard to determining the price reduction effect of power line proximity.

For the empirical question cited above, the results obtained with the data from the carefully selected target areas turned out to be reasonable: a substantial, but context-dependent reduction in price attributable to proximity to a power line was captured. Although the results differed between different submodels and also in comparison with the previous study by Peltomaa (1998), the overall results obtained with the SOM-based method proved to be informative and logical.

To proceed to the methodological conclusions, we cannot claim that the SOM-based method has an added value over a conventional regression method in the way suggested in the two previous chapters (and in Kauko & Peltomaa 1998), because similar results were obtained with plain multiple regression analysis simply by pooling the dataset (in Peltomaa 2001). On the other hand, because the results generated with the combined dataset had the same sign as the results from the regression analysis, the neural network could be considered a valid method.

At this point some words of caution are appropriate. With neural networks, one has to be prepared for ambiguous results. The maps from the selected areas in Jyväskylä and Järvenpää suffered from small sample sizes and consequently inadequate generalisation. Also, the decision to put two different areas into the same model is debatable. Nonetheless, these problems also apply to multiple regression analysis, and when comparing these two methods neural network enthusiasts often claim that even a small sample is all right if the quality of the data is good. Here, neuron maps were produced with as few as twenty-six observations. However, as shown, the post-processing approach does not differ in any substantial way from the regression approach.

Notes

1. The chapter is a focussed elaboration on Kauko & Peltomaa (1998).

2. Building cheap houses on sites close to a power line is not an externality in the strict sense and the causality could of course also be in the opposite direction: the power line is built close to cheap residences on purpose. This would, however be a more political argument and a somewhat sensitive statement without exact knowledge of the situation. Therefore, for the purposes of the study the positive argument certainly seems more appealing: that the powerline nuisance causes plots to be zoned for relatively cheaper housing; in other words, a powerline is treated as a negative externality that is internalised in the zoning and building decisions.

3. Peltomaa (1998) actually sent a questionnaire to the target area addresses, to support his results obtained with multiple regression analysis. The overall conclusion based on 18 returned questionnaires was that owners did *not* perceive any negative effect of a nearby (<200 m.) power line. In fact, the effect might even be positive: some plot owners might appreciate being situated adjacent to a power line instead of troublesome neighbours.

4. In a previous exploration with the SOM the power line proximity was included among the thirty variables participating in the computation. However, no visible effect was traced in that study (Kauko 1996), either.

5. The sampling was conducted by Juhani Väänänen from the Real Estate Information Centre of NLS.

6. One could argue that this way of sampling includes a selection bias usually discussed in econometric literature. Some people do not care about the negative effects and choose to reside close to the nuisance, which affects the ability of the method to measure the specific effect in question (e.g. Strand & Vågnes 2000).

7. The aim was to use small-scaled maps initially, and then gradually, to focus on larger-scaled maps. The search for target areas was conducted by Antti Heikkinen.

8. Note that in the original analysis by Peltomaa (1998) the two samples were not pooled as they were here. The idea of pooling came later, when trying to obtain more data for the analysis – especially for the SOM-regression modelling part in subsection 6.2.4.

9. These two towns are remote from each other, and do not resemble each other in any other way either (Hence the variable is a benchmark for our purposes: the presumably most important value factor.)

10. Classification with two labels is easier than with three labels. The more classes there are, the more difficult it *generally* becomes to recognise the appropriate class to which an observation vector should belong.

11. Because of the small sample size, the towns Jyväskylä and Järvenpää were put into the same model. With the original data no statistically significant difference were found between them when a (macro) locational dummy was used. Also, in the (equally scaled) feature maps they were quite well mixed. Another issue is of course that pooling together two different samples is questionable from the point of view of hedonic price theory (Peltomaa 2001, p. 21).