



*International Journal of Engineering and Geosciences (IJEG),
Vol; 4, Issue; 3, pp. 094-105, October, 2019, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.458430*

REDUCTION OF MASS APPRAISAL CRITERIA WITH PRINCIPAL COMPONENT ANALYSIS AND INTEGRATION TO GIS

Fatma Bünyan Ünel^{1*} and Şükran Yalpir²

¹Mersin University, Engineering Faculty, Department of Geomatics Engineering, Mersin, Turkey.
(fatmabunel@mersin.edu.tr); **ORCID 0000-0002-9949-640X**

²Konya Technical University, Engineering Faculty, Department of Geomatics Engineering, Konya, Turkey.
(syalpir@ktun.edu.tr); **ORCID 0000-0001-9366-5770**

*Corresponding Author, Received: 09/09/2018, Accepted: 27/02/2019

ABSTRACT: In real estate, mass appraisal is a very important subject in the valuation of two and more properties. It can be of benefit in a number of fields including taxation, banking transactions, expropriation, etc. The base problem is which criteria to use for mass appraisal. Because the number of criteria and the criteria themselves vary according to people, regions and countries, they are uncertain. They should be optimum in order to save on time, labour and cost. The aim of this study is to reduce the criteria by determining which ones affect the plot value. A survey which was answered by a total of 2,531 participants was conducted in Turkey. Principal Component Analysis (PCA), one of the criteria analysis methods, was applied to the survey data. The number of criteria was reduced to 14 with separation and to 30 according to the results of PCA. But they decreased in the model verification when criteria data for 558 samples were collected in the Konya study area. An index of the neighbourhood and locational features of these criteria was created by using GIS. Three models were established using Multiple Regression Analysis (MRA) and the performance of the models was examined. The prediction values and the market value were integrated into the GIS to compare the spatial distributions of plot values.

Keywords: *Mass Appraisal, Principal Component Analysis (PCA), Geographic Information Systems (GIS), index.*

1. INTRODUCTION

Real estate valuation is of great significance in many countries, and some of their revenues are acquired through taxes collected from the real estate. The economic value of real estate is also used for taxation, insurance and expropriation as well as for such management of real estate as land management, planning, urban restructuring and zoning. The need for mass valuation emerges in these processes and such traditional methods as comparison, review and cost remain inefficient in this regard, due to the fact that they cannot yield a solution. Relevant criteria should first be identified in order for mass valuation to be carried out. In academic and practice studies, different criteria have been used and universal criteria are not found (Yalpir & Ünöl, 2016). There are International Valuation Standards but no standard criteria are globally accepted. However, since the number of criteria is more than necessary this poses a problem. Using all the current criteria is time-consuming, labour-intensive and costly. The optimum criteria to be used for mass appraisal can be determined through research and analysis.

The literature related to the criteria affecting the value of real estate was examined and it was especially seen that criteria studied such as neighbourhood greenspace (Li, 2010), zebra mussels (Henry, 2013), shale gas (Muehlenbachs, Spiller, & Timmins, 2014), or state forests (Smigielski, 2014) had impacts on property values. Zavadskas et al. (2017) mentioned that the sustainability criteria for real estate play a more and more significant role in the different evaluation processes. In a study by Sdino, Rosasco, Torrieri, and Oppio (2018), the real estate characteristics selected to appraise the real estate value in Italy were grouped as extrinsic, intrinsic and technological. The weights of each of the characteristics were calculated with AHP.

Due to the lack of studies on criteria, statistical and advanced real estate valuation methods were taken into consideration and real estate types and their criteria were discussed. In one study, Kauko (2002) applied methods by considering Finland and Helsinki on separate scales through different criteria. He used YSA and the hedonic method for the buildings he chose in the areas by using 16 criteria. In the valuation of buildings, Lynn employed the methods of multiple regression, nonparametric regression, and YSA with a total of 83 criteria, 66 of which were composed of local features. Schulz (2003) compared the success stories by applying linear regression and hedonic regression with the criteria related to buildings after examining German valuation regulations. Modern valuation methods are compared with the regression and hedonic methods, which are usually easy to use and which yield high accuracy in the mass appraisal of real estate. The studies that deal with local criteria at length display an analytic approach based on the development of local features and identified techniques for locational modelling (McCluskey & Borst, 2007). A local weight matrix was used to produce a locational correlation, and locational were analysed through the hedonic method (M. Kryvobokov & Wilhelmsson, 2007); the relationship between the valuation of real estate and green areas, surface water, the effects of noise and view features was researched through correlation analysis (Cellmer, Senetra, & Szczepanska, 2012), and the relationship between the determinants

related to the value of buildings in terms of both the urban environment and the structural features of the building market was analysed in another study (Zoppi, Argiolas, & Lai, 2015).

In the literature review, it was observed that the studies were conducted in a method-based manner first by conducting research concerning the method. It was generally seen that different criteria were used when the applications of the modern valuation methods were developed and compared with regression. However, it was also observed that the criteria affecting the value of real estate were considered on a local basis for these applications, and the criteria research was conducted with all the criteria involved. Rather, the criteria were examined for them to become a basis for the method. In each application, the criteria were observed to be in differing numbers. Because the literature research was conducted on different countries and different areas, it was confirmed that the criteria varied according to these different countries and areas. Besides, studies were generally conducted on buildings in terms of real estate, but there are also studies, although few, in which values related to commercial buildings, were also considered (Bender, Din, Hoesli, & Laakso, 1999; Marko Kryvobokov, 2005, 2006; Son, 2012). It was seen in these applications that the market value, the unit value and the rent value were used as independent variables (El-Gohary, 2004; Son, 2012; Zoppi et al., 2015).

The plot is treated as a form of real estate in this study, and the criteria affecting the plot value are determined from the findings of the literature review. The number of criteria being very high makes the application of mass appraisal very difficult. Also, using all the criteria for mass appraisal does not seem economical and appropriate because collecting, arranging and analysing the data on the criteria takes a long time and requires labour and expense. In addition, the criteria demonstrate significant differences between young and old people (Unel, Yalpir, & Gulnar, 2017).

The purpose of this study is first to determine, at an optimal level, the criteria that can be used for mass appraisal and then to create a geographical data model with neighbourhood and local indices, thus standardizing the criteria. Reduced criteria were obtained by applying Principal Component Analysis (PCA) to the surveys carried out earlier. Since GIS is a powerful tool (Amil, 2018), the neighbourhood and location indices were generated by using ArcGIS software on the data corresponding to the reduced criteria belonging to local and locational features. Three models were obtained in total through the application of Multiple Regression Analysis (MRA) for the verification of the model by using all the criteria and the reduced criteria. Prediction values were calculated from the models and their performance analyses were conducted for a comparison with the market value. The model in which the dependent variable was best explained by independent variables in MRA turned out to be Model 1, which was obtained with a combination of all the criteria. The model closest to Model 1 was seen to be Model 2, which was obtained through local indices from the criteria reduced with PCA.

2. MATERIALS AND METHOD

The outline of this study can be presented as in the following flow chart. First, it defines the problem of

which criteria affect the plot value. It then continues with the criteria and their analysis, model verification, and performance analysis (Fig. 1).

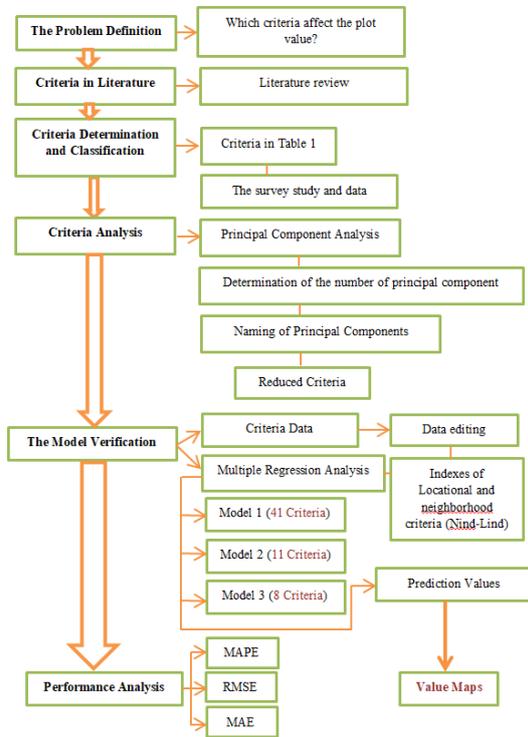


Figure 1 The study diagram

2.1 Study Area and Market Sampling

The central neighbourhoods of the city of Konya were identified as the study area, and information regarding local features of the neighbourhoods involved was obtained for each neighbourhood on a small scale. The facilities on the map relating to the criteria under the location features were arranged in ArcGIS software in a plot-based manner, each in different layers, and maps were generated in a vector format. Furthermore, the values of the plots in the central neighbourhoods on which buying and selling transactions were conducted were obtained, and a total of 558 market samplings were collected. Also, data about legal and physical features were collected, and they were arranged in a matrix format. In the verification of the model, the plot values were taken as dependent variables and the other information about the plots was taken as independent variables. Data that could not be accessed were disregarded, and not included in the model.

It should not be expected for the distribution of market samplings on the map to be homogeneous for Konya, as the density of buildings increases towards the city centre and it becomes impossible to find plots for construction. While there are few samplings in the city centre, it is seen that the number of samplings increases with distance from the city (Fig. 2).

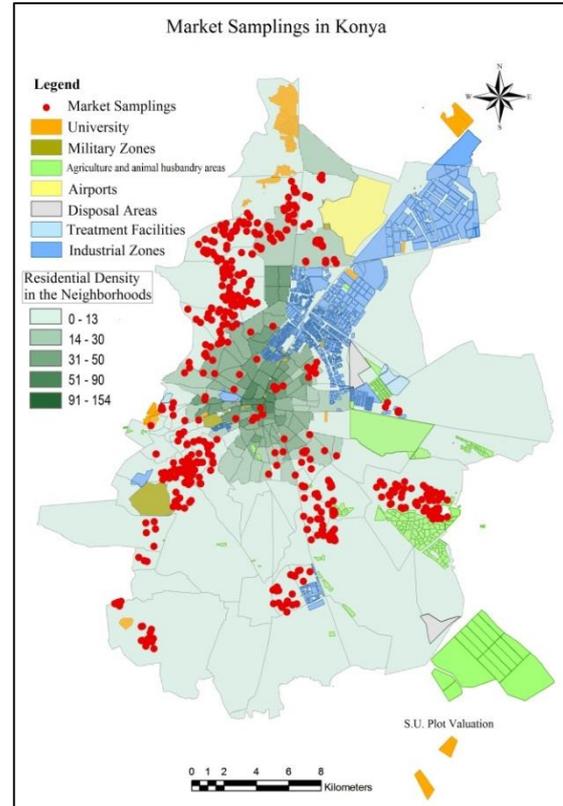


Figure 2 Market Sampling (Ünel, 2017)

2.2 Criteria Affecting the Value of the Plots

In this study, the plots and the criteria that affected their value were identified through the literature review. The criteria were compiled as main, sub and criterion, and they were grouped under the main heading as legal, physical, locational and local features (Table 1). The criteria were arranged in a survey form and the surveys were carried out on 2,531 participants in previous studies (Ünel, 2017). The survey consists of 116 questions related to the criteria. The questions were asked of experts working on real estate valuation, and citizens who played a role in real estate purchases and/or sales in Ankara, Konya, and Kayseri in the Central Anatolia Region. Finally, data collection was conducted in two steps: first the survey data and then the data about the plots were collected. PCA was employed to reduce the criteria for the survey data, while MRA was used to verify the model for the other data.

2.3 Principal Component Analysis

Principal Component Analysis (PCA) is a method of choosing an attribute used to describe the original data with a small cluster of data and with the lowest possible number of samples; it is a statistical analysis with multiple variables (Özdemir, 2010). PCA is a tool that is used in studies carried out in many fields such as psychology, education, quality, agriculture, chemistry, mapping, photographic sciences, market research, economy, anatomy, biology, forest sciences and genetics (M. Çilli & Arıtan, 2010).

PCA converts interrelated variables into a small unrelated theoretical structure called the principal

component (PC). PCA is used to discover and interpret the dependence that exists between variables as well as examining the possible relations between individuals (Timm, 2002). PCA can be used independently; it can also be used as a technique for data collection for other analyses (Sangün, 2007). The mathematical steps of PCA are as follows:

1. The matrix form of the data cluster
2. The mean values of the data cluster
3. Variance covariance matrix
4. The eigenvalue and the eigenvector
5. Determining the number of principal components

Judging from the fact that an x-series is obtained by measuring each of a number of variables such as X_1, X_2, \dots, X_p on n number of participants, the matrix formed by the raw data is shown on the $n \times p$ dimension as an X data matrix in Eq. (1) (Sangün, 2007):

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix} \quad (1)$$

The mean values (\bar{X}_j) of the data cluster are obtained by using Eq. (2) for each variable.

$$\bar{X}_j = \frac{\sum_{i=1}^n X_{ij}}{n} \quad (2)$$

- i: 1, 2, 3, ..., n (the number of participants),
j: 1, 2, 3, ..., p (the number of variables)

While variance measures the divergence of the points on a variable from the average, covariance measures how much each of them changes from the average in relation to each other (Farag & Elhabian, 2009). The variance–covariance matrix of the data cluster is formed in a $p \times p$ dimensional way Eq. (5) by using the variance Eq. (3) and covariance Eq. (4) equalities. One of the most important aspects that differentiate PCA from factor analysis is that PCA is processed through variance–covariance analysis (MathWorks, 2014; Minitab, 2014; Sangün, 2007). The matrixes of variance $\mathit{var}(X)$, covariance $\mathit{cov}(X_1, X_2)$ and variance–covariance cov are shown as

$$\mathit{var}(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})}{(n-1)} \quad (3)$$

$$\mathit{cov}(X_1, X_2) = \frac{\sum_{i=1}^n (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{(n-1)} \quad (4)$$

$$\mathit{cov} = \begin{bmatrix} \mathit{var}(X_1) & \mathit{cov}(X_1, X_2) & \mathit{cov}(X_1, X_3) & \dots & \mathit{cov}(X_1, X_p) \\ \mathit{cov}(X_2, X_1) & \mathit{var}(X_2) & \mathit{cov}(X_2, X_3) & \dots & \mathit{cov}(X_2, X_p) \\ \mathit{cov}(X_3, X_1) & \mathit{cov}(X_3, X_2) & \mathit{var}(X_3) & \dots & \mathit{cov}(X_3, X_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathit{cov}(X_p, X_1) & \mathit{cov}(X_p, X_2) & \mathit{cov}(X_p, X_3) & \dots & \mathit{var}(X_p) \end{bmatrix} \quad (5)$$

(Farag & Elhabian, 2009; Gaborski, 2010; Polat, 2008; Sangün, 2007).

The eigenvalues and the eigenvectors of the variance–covariance matrix are found; the eigenvectors which have the biggest eigenvalue correspond to the variables that have the strongest relation in the data cluster (Gaborski, 2010). Therefore, the eigenvectors found here are accepted as the new principal components (Orhan, 2013).

Table 1 Criteria affecting the value of the plot (*subheadings are in bold and italic*) (Yalpir & Unel, 2017).

NO	A. LEGAL FEATURES	Q38	<i>Proximity to Educational Institutions</i>	Q78	Proximity to underpass/ overpass
Q1	<i>Property Conditions</i>	Q39	Proximity to Pre-schools	Q79	<i>Proximity to Unsanitary Areas</i>
Q2	Full Ownership	Q40	Proximity to High Schools	Q80	Proximity to was disposal areas
Q3	Shared Ownership	Q41	Proximity to Higher Education Institutions	Q81	Proximity to treatment facilities
Q4	<i>Zoning Status</i>	Q42	Proximity to courses	Q82	Proximity to natural gas and tube filling facilities
Q5	The Gross Floor Area	Q43	<i>Proximity to Public Institutions</i>	Q83	Proximity to petrol stations
Q6	Total Construction Area	Q44	Proximity to governorships	Q84	Proximity to base stations
Q7	The number of floors \geq 10	Q45	Proximity to Municipalities	Q85	Proximity to energy transmission lines
Q8	The number of floors $<$ 10	Q46	Proximity to Courthouse	Q86	Proximity to underdeveloped areas
Q9	Detached Building	Q47	Proximity to Jailhouse	Q87	Proximity to marsh areas
Q10	Attached Buildings	Q48	<i>Proximity to Security Units</i>	Q88	Proximity to natural disaster areas
Q11	<i>Legal Restraints</i>	Q49	Proximity to Police Stations	Q89	Proximity to not improved river areas
Q12	Right of Mortgage	Q50	Proximity to Military Zones	Q90	<i>Proximity to Industrial Zones</i>
Q13	Easement	Q51	Proximity to Fire Departments/ 112 Emergency	Q91	<i>Proximity to Graveyards</i>
Q14	Annotation of Lease	Q52	<i>Proximity to Attraction Centres</i>	Q92	<i>Proximity to Worship Places</i>
Q15	<i>Plot Area</i>	Q53	<i>Proximity to Shopping Centres</i>	Q93	<i>Proximity to Business Centres</i>
NO	B.PHYSICALFEATURES	Q54	Proximity to Hypermarkets	Q94	<i>Proximity to Parking Areas</i>
Q16	<i>The location of the plot</i>	Q55	Proximity to mini-markets	Q95	<i>The View From The Plot</i>
Q17	Corner parcel	Q56	Proximity to open/closed bazaars	Q96	Mountain, valley, etc. views
Q18	Intermediate parcel	Q57	Proximity to commercial enterprises	Q97	Lake, river, stream, etc. view
Q19	<i>Geometric Structure</i>	Q58	<i>Proximity to Cultural Centres</i>	Q98	City view
Q20	Length of the Frontage	Q59	Proximity to cinemas/theatres	NO	D.NEIGHBOURHOOD FEATURES
Q21	The number of frontage	Q60	Proximity to historical sites and touristic attractions	Q99	<i>Population density</i>
Q22	Geometric shape	Q61	<i>Proximity to Entertainment Centres</i>	Q100	Education Level
Q23	<i>Technical Infrastructure Services</i>	Q62	Proximity to fairs, concert areas, etc.	Q101	Level of income
Q24	Water supply	Q63	Proximity to sport facilities	Q102	Immigrant receiving

Q25	Electricity, sewer, natural gas, and telephone	Q64	Proximity to stadium/hippodrome	Q103	Criminal Rate
Q26	Solid waste collection service	Q65	Proximity to entertainment venues	Q104	Neighbourliness Relations
Q27	Storm drainage	Q66	Proximity to Green Areas	Q105	Homeowner/tenant
Q28	Unpaved road	Q67	Proximity to forest/copses	Q106	The Surrounding Environment
Q29	Asphalt road	Q68	Proximity to recreation areas	Q107	The favourite neighbourhood
Q30	The Road Condition	Q69	Proximity to parks	Q108	Residential Density
Q31	The Periphery Road	Q70	Proximity to playgrounds	Q109	Development potential
Q32	Road width \geq 10	Q71	Proximity to Public Transportation Points	Q110	Purchasing and selling mobility of real estate
Q33	Road width $<$ 10 metre	Q72	Proximity to airports	Q111	Underground, soil, and aboveground features
Q34	The Slope of The Plot	Q73	Proximity to railway stations	Q112	Slope of the neighbourhood
NO	C. LOCATIONAL FEATURES	Q74	Proximity to coach station	Q113	Geological condition
Q35	Proximity to Health Facilities	Q75	Proximity to tramway, subway and metrobus stations	Q114	Climate Condition
Q36	Proximity to health centre, village clinic, etc.	Q76	Proximity to bus stops	Q115	Air Pollution
Q37	Proximity to State/Private Hospitals	Q77	Proximity to shared taxi routes	Q116	Noise Pollution

The number of components is equal to the number of p variables at the beginning. However, the number of the components chosen must be smaller than p . A curve is drawn to combine the points on which components and the eigenvalues are shown. The breakpoint at which the eigenvalues start to become parallel to the horizontal axis gives the number of components. The number of components can also be identified depending on what percentage of the total variance the first components represent. The acceptance of this percentage ranges in the literature from 70% to 95% (M. Çilli, 2007; Semmlow, 2004).

2.4 Multiple Regression Analysis

In multivariate regression analysis, the independent variables are used to explain the change in the dependent variable simultaneously. The model of the MRA analysis,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i \quad (6)$$

y_i : Dependent variable (value of real estate)
 $x_{i1}, x_{i2}, \dots, x_{ik}$: Independent variables (share, area, TAKS, KAKS, number of floors, etc.)
 u_i : Corruption or error term.
 β_0 : Constant
 $\beta_1, \beta_2, \dots, \beta_k$: Coefficients of variation

can be demonstrated by a general formula in Eq. (6). The F test in the result of analysis and R^2 are the important concepts to be checked first. The F test is a test with ANOVA to examine whether the regression model is significant. The level of significance corresponding to the value of F resulting from the ANOVA test helps to determine whether the model created is appropriate. When the result of the F test is significant ($p < 0.05$), it is interpreted that the model contributes significantly to explain the dependent variable. The R^2 value indicates what percentage of the variance in the dependent variable is explained by the independent variable (Altunışık, Coşkun, Bayraktaroğlu, & Yıldırım, 2010). The closer this value is to 1, the better the model is explained by the independent variables (Yalprı, 2007).

2.5 Performance Analysis

The results of the mean absolute error (MAPE) Eq. (7), the root mean square error (RMSE) Eq. (8) and the mean absolute error (MAE) Eq. (9) were used in the performance analyses to compare the model values and the market values obtained from the criteria. The performance of the model is investigated with these error rates especially in the studies in which the method is developed for real estate valuation (Fernandez-Martinez, Fernandez-Ceniceros, Sanz-Garcia, Lostado-Lorza, & Martinez-De-Pison-Ascacibar, 2011; Kavas, 2014; Kuşan, Aytekin, & Özdemir, 2010; Lin, 2010; Lughofer, Trawinski, Trawinski, Kempa, & Lasota, 2011; Saraç, 2012).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad \begin{array}{l} y_i : \text{Market values,} \\ \hat{y}_i : \text{Model values,} \\ i : 1, 2, 3, \dots, n \\ n : \text{Number of} \\ \text{selected samples} \end{array} \quad (9)$$

3. RESULTS AND DISCUSSION

3.1 Results of Principal Component Analysis

Principal component analysis is a reduction analysis and is applied to the survey data by using Minitab 17 software. A covariance–variance matrix was preferred because of the fact that raw data were used. The data belonging to legal, physical, location and neighbourhood features were analysed separately in the general sampling group.

Because the principal components were composed of PCA, the main heading and most subheadings were not included in the process of analysis. However, the plot itself was taken as a criterion because there were not any criteria under certain headings such as its size and its slope. A total of 96 criteria were subjected to analysis. The survey data were composed of 96 (p) criteria and they

were measured on 2,474 (n) participants, giving the matrix of the data cluster a size of $96 \times 2,474$ which consisted of the raw data. Because the legal features had data about 12 sub-criteria, a total of 12 components were calculated. The percentages of the eigenvalues, variance and cumulative variance were given for each component (Table 2). Although the number of components was 12 according to the raw data, it must be lower than 12 according to PCA. Accordingly, the number of components with a significance level of 75 (0.746) % according to the cumulative percentages of variances is 4.

Table 2 Eigenvalue analysis of variance–covariance matrix

Component	The Eigen Value analysis of variance covariance matrix		
	Eigen Value	Rate	Cumulative
PC1	28.068	.407	.407
PC2	9.769	.142	.549
PC3	7.557	.110	.658
PC4	6.048	.088	.746
PC5	4.496	.065	.811
PC6	3.371	.049	.860
PC7	3.224	.047	.907
PC8	1.872	.027	.934
PC9	1.517	.022	.956
PC10	1.218	.018	.973
PC11	0.963	.014	.987
PC12	0.870	.013	1.000

Considering that eigenvalues could be 1 and higher than 1, the number of components is 10 (Table 2). The number of components is 2 according to the graphics composed of the numbers of eigenvalues and of components (Fig. 3). This being so, it was concluded that the number of components should be taken as 2 according to the graphics with the lowest number of components.

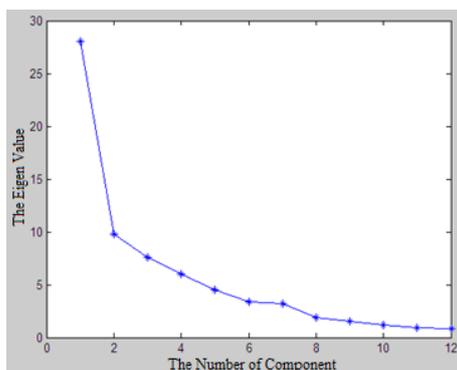


Figure 3 Graphics of eigenvalues–the number of components

PCA was applied to the sub-criteria of the legal features, and the analysis was repeated after removing the criteria that remained below 30% (Q2, Q7, Q8, Q9, and Q10) and those that were composed of a single criterion (Q3). According to the results of the repeated analysis, the cumulative variance percentage of two components was 70 (0.699) % and therefore the analysis was understood to be significant (Table 3).

Table 3 Principal components of the legal features

Component Question No	Mean	Standard Deviation	Legal Restraints	Zoning Status
Q14	-1.64	3.60	.579	
Q13	-2.47	3.34	.575	
Q12	-2.68	3.41	.573	
Q6	3.81	2.04		-.539
Q15	3.83	2.06		-.488
Q5	3.89	1.82		-.466
Eigen Value			26.099	7.048
Cumulative			.550	.699
Reliability			.813	.647
N			3	3

PCA was applied to the sub-criteria of the physical features, and it was seen that Q28 had high loads in two components. The analysis was repeated after leaving Question 28 out. The criteria that remained below 30% (Q17, Q22, Q24, and Q25), a criterion which was discordant with the criteria in the component (Q18) and those which were the only criterion in a component (Q31, Q32, Q33, and Q34) were removed, and then the analysis was repeated. According to the repeated analysis, the cumulative variance percentage of the components was 72 (0.716) %. It was understood that the analysis was significant (Table 4).

Table 4 Principal components of the physical features

Component Question No	Mean	Standard Deviation	Infrastructure Services	Status of Frontage
Q29	3.45	2.14	.463	
Q27	3.54	2.00	.454	
Q26	3.66	1.84	.426	
Q21	3.74	2.12		-.579
Q20	3.75	2.02		-.515
Eigen Value			9.697	5.036
Cumulative			.471	.716
Reliability			.733	.754
N			3	2

PCA was applied to the sub-criteria of the locational features. The criteria that remained below 30% (Q36, Q37, Q39, Q40, Q41, Q42, Q49, Q51, Q52, Q54, Q55, Q56, Q57, Q59, Q60, Q63, Q67, Q68, Q69, Q70, Q74, Q75, Q76, Q77, Q78, Q87, Q88, Q89, and Q98), the criteria that were discordant with the criteria in the component (Q46, Q47, Q50, Q91, Q92, Q93, and Q94), the one which was the only criterion in a component (Q90) and that with the highest load into components (Q62) were removed, and then the analysis was repeated at least 15 times. The rate of cumulative variance percentage of the five components was 75 (0.748) % in the last analysis. The analysis was understood to be significant (Table 5).

PCA was applied to the sub-criteria of neighbourhood features. The criteria that remained below 30% (Q100, Q103, Q104, Q105, Q107, and Q109) were removed. Then the analysis was repeated. According to the results of the repeated analysis, those discordant with the criteria in the component (Q110, Q113, and Q114), those which were the only criterion in one component (Q101 and Q102) and the one with a high load in two criteria (Q112) were removed. The analysis was repeated. According to the last analysis, the cumulative variance value of the two components was 78 (0.781) %. This showed that the analysis was significant (Table 6).

Table 5 Principal components of the locational features

Component Question No	Mean	Standard Deviation	Unsanitary Areas	Entertainment Areas	Transportation Networks	The View from the Plot	Public Institutions
Q83	-2.90	2.75	-.370				
Q82	-3.58	2.21	-.336				
Q84	-3.44	2.27	-.335				
Q85	-3.37	2.23	-.334				
Q81	-3.56	2.26	-.329				
Q86	-3.32	2.27	-.321				
Q80	-3.70	2.23	-.306				
Q64	1.64	2.92		.408			
Q65	1.77	2.83		.381			
Q72	1.28	3.28			-.704		
Q73	2.43	2.69			-.380		
Q96	3.28	2.26				-.509	
Q97	3.43	2.18				-.495	
Q45	2.88	2.23					.522
Q44	2.81	2.22					.508
Eigen Value			28.010	22.425	9.687	8.079	6.186
Cumulative			.281	.507	.604	.685	.748
Reliability			.925	.683	.672	.843	.879
N			7	2	2	2	2

Table 6 Principal components of neighbourhood features

Component Question No	Mean	Standard Deviation	Population/Residential Density	Air/Noise Pollution
Q108	0.70	3.61	.727	
Q99	0.90	3.50	.662	
Q116	-3.53	2.37		.676
Q115	-3.53	2.37		.675
Eigen Value			18.346	10.177
Cumulative			.502	.781
Reliability			.599	.910
N			2	2

3.1.1 The Wording of the principal components

A total of 96 criteria were subjected to PCA. As a result of the repeated analysis carried out after removing those that remained below 30%, any which was the only criterion in a component and those that were discordant with the other criteria, 30 criteria remained. These criteria were composed of 11 principal components. A total of 11 principal components, 2 from legal features, 2 from physical features, 5 from locational features and 2 from neighbourhood features, were formed according to PCA (Table 7).

Great attention was paid to the wording of the components in accordance with the criteria in the number of the questions under the components.

Table 7 Principal components at a total of 30 criteria

No of Order	Main Heading	Component No	The Wording of the Main Components	The number of Items	Question No
1	Legal Features	TB1	Legal Restraints	3	Q14, Q13, and Q12.
2		TB2	Zoning Status	3	Q6, Q15, and Q5.
3	Physical Features	TB1	Infrastructure Services	3	Q29, Q27, and Q26.
4		TB2	The Status of the Frontage	2	Q21 and Q20.
5	Locational Features	TB1	Unsanitary Areas	7	Q83, Q82, Q84, Q85, Q81, Q86, and Q80.
6		TB2	Entertainment Areas	2	Q64 and Q65.
7		TB3	Transportation Networks	2	Q72 and Q73.
8		TB4	The View from the Plot	2	Q96 and Q97.
9		TB5	Public Institutions	2	Q45 and Q44.
10	Neighbourhood Features	TB1	Population and Residential Density	2	Q108 and Q99.
11		TB2	Air and Noise Pollution	2	Q116 and Q115.

3.1.2 Separation of the principal components

It is necessary to separate some of the principal components because they involve criteria which require specific information about the plot. The components of zoning and frontage must be separated into sub-criteria. The components of infrastructure services can also be considered separately, depending on whether such services as water, electricity, sewage system, natural gas, telephone, etc. are available. However, because such a separation is impossible due to a lack of data, these criteria were taken as a single criterion and were not separated (Table 8). At the end of this process, a total of 14 reduced criteria were established.

Table 8 Separation of certain components

THE COMPONENTS NOT PUT THROUGH SEPARATION		THE COMPONENTS PUT THROUGH SEPARATION	
No	Factors	No	Zoning Status
1	Legal Restraints	1	The Gross Floor Area
2	Infrastructure Services	2	Total Construction Area
3	Unsanitary Areas	3	The Size of the Plot
4	Entertainment Areas	4	Length of the Frontage
5	Transportation Networks	5	The number of frontage
6	The View from the Plot		
7	Public Institutions		
8	Population and Residential Density		
9	Air and Noise Pollution		Total=14

3.2 Formation of the Indexes

Because of the relationship of the criteria about neighbourhood and locational features with the map and because of the difference between evaluation scales, indexes were formed in two basic steps:

- Neighbourhood index (Nind)
- Locational index (Lind)

The formation process is the process of obtaining a separate value for a standard neighbourhood and for the location by combining the value that the criteria of neighbourhood and locational features take on the map and the percentage of the responses from the survey with ArcGIS. The graphical data models were generated with GIS, and index maps were created. These provide that criteria are objectively appraised, standardized, and facilitated in the process of real estate valuation.

The numerical data of the criteria about neighbourhood features and their markings on the map were prepared in a vector environment. Raster format maps with pixel values classified between the standard point ranges are required in order to form an index. The maps of the criteria about neighbourhood features were generated in a standard way in the raster format. The neighbourhood index was obtained by using the weights of the criteria reduced in PCA, combining them and classifying them in a range from 1 to 10 (Fig. 4).

A proximity analysis was conducted by considering accessible distances to facilities of the locational features with the criteria reduced in PCA. The vector format map obtained as a result of this analysis was converted into standard ranges in a raster format. The criteria of locational features reduced with PCA were combined depending on their weight, and a locational index was generated with their classification ranging between 1 and 10 (Fig. 5). Locational and neighbourhood indexes were obtained by applying similar procedures to all the criteria. However, PC4 was not taken into consideration because it contained features such as mountain, lake, river, etc., and these were non-existent in the study area.

3.3 Results of Regression and Performance Analyses

MRA was applied by using market sampling data about a total of 41 criteria composed of legal, physical and neighbourhood features and locational indexes, and Model 1 was formed out of all the criteria. With the application of MRA, Model 2 and Model 3 were obtained by using the criteria reduced from PCA. The data about the gross floor area, total construction area and the size of the plot derived from the legal features, and the data about the infrastructure services, the length of the frontage and the number of the frontage are the same in Model 2 and Model 3.

The neighbourhood index which involved the density of population and of buildings and air and noise pollution were the same in both models and responded to a single value. However, the data about locational features varied. While the indexes of unsanitary areas, entertainment, transportation and public institutions were used to form the locational index in Model 2, the locational index which responded to a single value was used in Model 3. Also, legal restraints about market samplings and the criteria composed of the view status were disregarded and not included in the analysis.

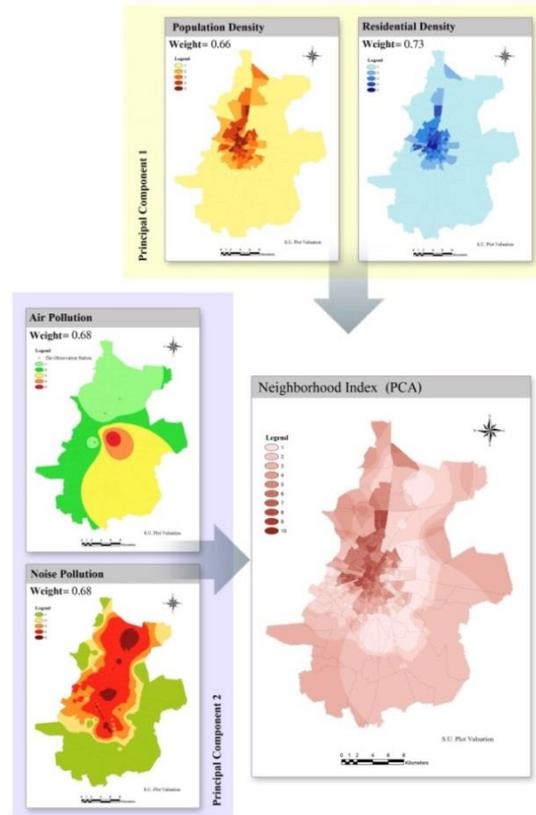


Figure 4 Neighbourhood index with the criteria reduced according to PCA

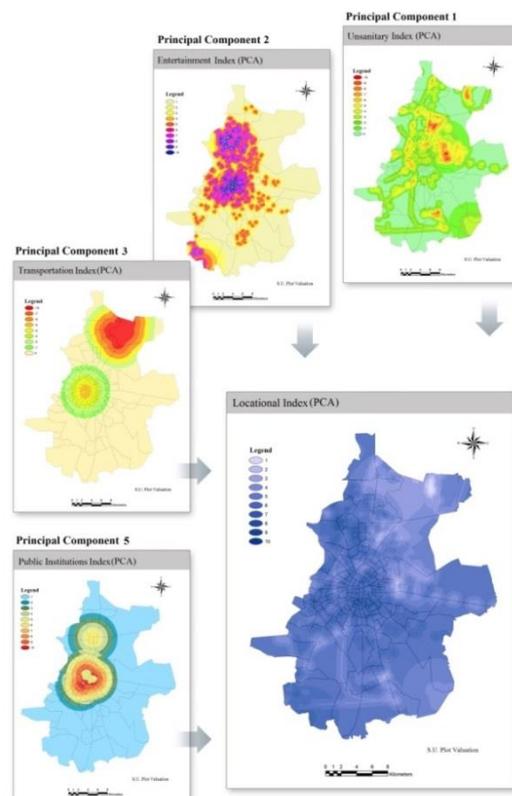


Figure 5 Locational index with the criteria reduced according to PCA

Prediction values were calculated by using mathematical models of Model 1, Model 2 and Model 3. Performance analyses were carried out in order to compare market values with the values obtained from the models, and these were examined (Table 9).

The R^2 value was calculated as 0.815 for Model 1, obtained with all the criteria. Because this was close to 1, it is the best model to explain the dependent variable. The second best model was Model 2 with 11 criteria. It is seen that the R^2 success rate (0.712) of Model 2 is closest to that of Model 1. The MAPE, RMSE and MAE error rates of this model are also closest to those of Model 1. The MAPE and MAE error rates of Model 3 are higher than the others. As the number of criteria decreases, it is observed that error rates increase (Table 9).

Table 9 The results of performance

Models	The number of criteria	Standard Deviation	R^2	$y=ax$	MAPE	RMSE	MAE
Model 1	41	0.030	0.815	$y=0.9223x$	0.368	0.014	0.009
Model 2	11	0.029	0.712	$y=0.8538x$	0.389	0.017	0.010
Model 3	8	0.028	0.677	$y=0.8397x$	0.447	0.017	0.011

3.4 Value Maps

Value maps were produced to see the locational distribution of the values. The market values and the prediction values obtained from Model 1, Model 2 and Model 3 were associated with market samplings in ArcGIS by using GIS. Value maps were generated with the kriging method as the geo-statistical method. Because the value maps were location-based, comparison and analysis were seen to be easier. When the locational distribution of the values was examined, it was established that Model 1 and Model 2 were closest to the market values. Because it contained 11 criteria, Model 2 is more appropriate for mass appraisal. In addition, it was observed that particularly the lowest and the highest values differed from one map to another (Fig. 6).

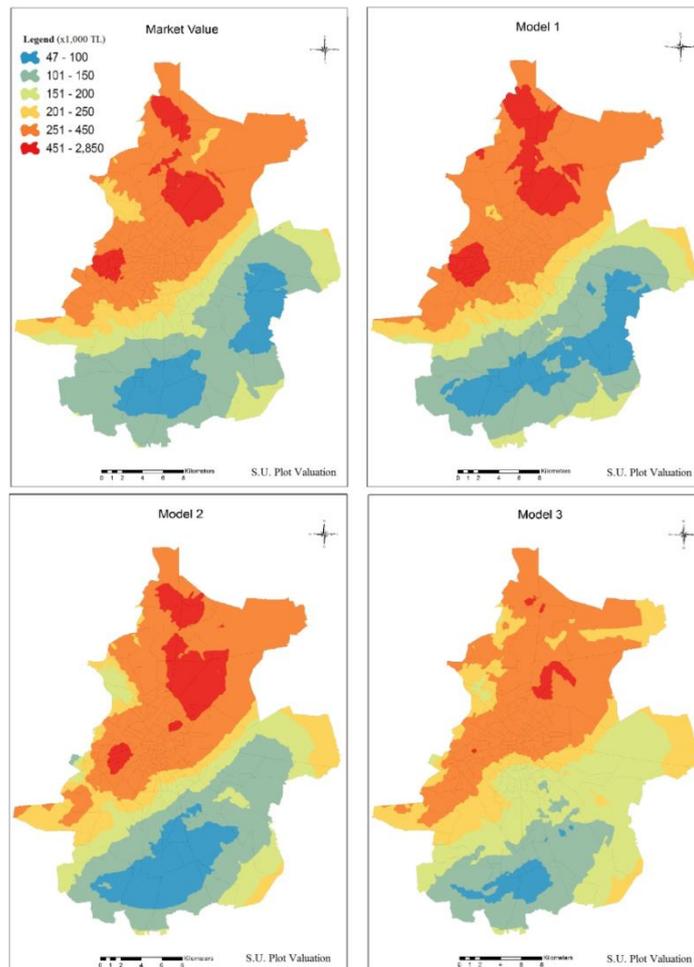


Figure 6 Value maps

While the total number of criteria was 116 in the survey, the number of reduced criteria after PCA was 30. The number of criteria used later to generate an index for model verification was 41 in Model 1, 11 in Model 2 and

8 in Model 3. The criteria used in Model 2 were the reduced criteria as a result of PCA. Locational and neighbourhood criteria were obtained in the form of an index and were turned into an objective form (Fig. 7).

THE REDUCED CRITERIA IN MODEL 2

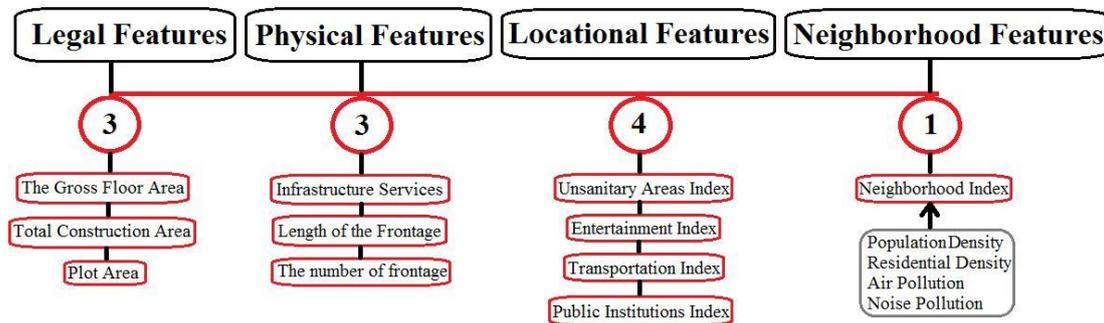


Figure 7 Reduced criteria used in model 2

4. CONCLUSION

It is observed that criteria change according to the study area concerned in the literature. When they are listed as a table, there are a large number of criteria that affect the plot value. Using all of them for mass appraisal is not considered to be economical. The optimum criteria and the number of criteria should be determined by taking into consideration the structural characteristics of the real estate in the country of interest. In addition, they should be obtained as independent from dependent variables such as price, letting value.

The criteria were reduced by using PCA and the results of criteria analysis were objectively evaluated. Neighbourhood and locational indexes were created by generating a geographical data model on ArcGIS with criteria for neighbourhood and locational features. Model verification was conducted by applying MRA to the data of the criteria of the plots. When the performances of Model 1 (all criteria), Model 2 (11 criteria) and Model 3 (8 criteria) were examined, Model 2 is closer to the market value than Model 3. In addition, it was concluded that the inexplicable part of Model 2 (nearly 29 %) resulted from the unstable economy, annuity, political situations and unavailable data.

Collecting, editing, and getting ready for the analysis of data relating to criteria are difficult and time-consuming processes. The reduced criteria save time, labour and expense in the processes. They will be a basis for mass appraisal and a system of real estate valuation. The application will be facilitated by standardizing real estate valuation in Turkey. In future, criteria will be separately defined according to all real estate types. Analyses should be carried out in critical regions where the criteria have changed, such as whether a property has sea view or not.

ACKNOWLEDGEMENT

The authors would like to thank the pollsters for their efforts and the participants who patiently completed the questionnaire in the data collection phase of the survey. This study is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) with 115Y769 Project Number and Selcuk University BAP coordinator with the project of 15101008.

REFERENCES

- Altunışık, R., Coşkun, R., Bayraktaroğlu, S., & Yıldırım, E. (2010). Sosyal Bilimlerde Araştırma Yöntemleri SPSS Uygulamalı (6. ed.). Sakarya: Sakarya Yayıncılık.
- Amil, T. A. (2018). Determining Of Different Inundated Land Use in Salyan Plain During 2010 The Kura River Flood Through GIS and Remote Sensing Tools. *International Journal of Engineering and Geosciences (IJEG)*, 3(3), 80-86. doi:10.26833/ijeg.412348
- Bender, A., Din, A., Hoesli, M., & Laakso, J. (1999). Environmental quality perceptions of urban commercial real estate. *Journal of Property Investment & Finance*, 17(3), 280 - 296.
- Cellmer, R., Senetra, A., & Szczepanska, A. (2012). The effect of environmental factors on real estate value. Paper presented at the FIG Working Week 2012, Rome, Italy.
- Çilli, M. (2007). İnsan hareketinin modellenmesi ve benzeşiminde Temel Bileşenler Analizi yönteminin kullanılması. Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü, Doktora Tezi, Ankara.
- Çilli, M., & Arıtan, S. (2010). Temel Bileşenler Analizi yardımı ile elde edilen daha az sayıda değişken kullanılarak farklı hızlarda insan koşusunun fourier tabanlı modelinin oluşturulması. *Hacettepe J. of Sport Sciences*, 21(1), 1-12.
- El-Gohary, M. (2004). Property valuation model effect of traffic noise on property value. Retrieved from
- Farag, A. A., & Elhabian, S. (2009). A Tutorial on Principal Component Analysis. Retrieved from <http://dai.fmph.uniba.sk/courses/ml/sl/PCA.pdf>
- Fernandez-Martinez, R., Fernandez-Ceniceros, J., Sanz-Garcia, A., Lostado-Lorza, R., & Martinez-De-Pison-Ascacibar, F. J. (2011). Dimensionality reduction in a database related with viticulture crops using wrapper techniques. *International Journal of Mathematical Models And Methods In Applied Sciences*, 5(5), 866-873.

- Gaborski, R. S. (2010). Principal Components Analysis. Retrieved from http://www.cse.psu.edu/~rcollins/CSE586Spring2010/lectures/pcaLectureShort_6pp.pdf
- Henry, E. N. N. (2013). From turtle populations to property values: The effects of lakeshore residential development and The Invasive Zebra Mussel. Michigan State University, PhD, Michigan.
- Kauko, T. J. (2002). Modelling the locational determinants of house prices: neural network and value tree approaches. Utrecht University, PhD, Netherlands.
- Kavas, S. (2014). Konut fiyatlarının çok kriterli bir karar destek modeli ile tahmin edilmesi İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Kryvobokov, M. (2005). Estimating the weights of location attributes with the Analytic Hierarchy Process in Donetsk, Ukraine. *Nordic Journal of Surveying and Real Estate Research*, 2(2), 5-29.
- Kryvobokov, M. (2006). Mass valuation of urban land in Ukraine: From normative to a market-based approach. Real Estate and Construction Management School of Architecture and The Built Environment Royal Institute of Technology, PhD, Stockholm.
- Kryvobokov, M., & Wilhelmsson, M. (2007). Analysing location attributes with a hedonic model for apartment prices in Donetsk, Ukraine. *International Journal of Strategic Property Management*, 11, 157-178.
- Kuşan, H., Aytakin, O., & Özdemir, İ. (2010). The use of fuzzy logic in predicting house selling price. *Expert Systems with Applications*, 37, 1808-1813. doi:10.1016/j.eswa.2009.07.031
- Li, Q. C. (2010). Neighborhood greenspace's impact on residential property values: understanding the role of spatial effects. Faculty of the USC Graduate School University of Southern California, PhD, California.
- Lin, C. C. (2010). Critical analysis and effectiveness of key parameters in residential property valuations. State University of New York, The Faculty of The Graduate School of The University at Buffalo, PhD, New York.
- Lughofer, E., Trawinski, B., Trawinski, K., Kempa, O., & Lasota, T. (2011). On employing fuzzy modeling algorithms for the valuation of residential premises. *Information Sciences*, 181, 5123-5142. doi:10.1016/j.ins.2011.07.012
- MathWorks. (2014). İstatistik, Faktör Analizi. Retrieved from <http://www.mathworks.com/products/demos/statistics/factorandemo.html>
- McCluskey, W. J., & Borst, R. A. (2007). Specifying the effect of location in multivariate valuation models for residential properties: A critical evaluation from the mass appraisal perspective. *Property Management*, 25(4), 312-343. doi:10.1108/02637470710775185
- Minitab. (2014). What are the differences between principal components analysis and factor analysis? Retrieved from <http://support.minitab.com/en-us/minitab/17/topic-library/modeling-statistics/multivariate/principal-components-and-factor-analysis/differences-between-pca-and-factor-analysis/>
- Muehlenbachs, L., Spiller, E., & Timmins, C. (2014). The housing market impacts of shale gas development. Retrieved from Washington, New York:
- Orhan, U. (2013). Makine öğrenmesi. Retrieved from <http://bmb.cu.edu.tr/uorhan/DersNotu/Ders11.pdf>
- Özdemir, A. T. (2010). Erken ventriküler kasılmalarda YSA tabanlı bir sınıflandırıcının FPGA ile gerçekleştirilmesi. Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Doktora tezi, Kayseri.
- Polat, K. (2008). Biyomedikal sinyallerde veri ön-işleme tekniklerinin medikal teşhiste sınıflama doğruluğuna etkisinin incelenmesi. Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, Konya.
- Sangün, L. (2007). Temel bileşenler analizi, ayırma analizi, kümeleme analizleri ve ekolojik verilere uygulanması üzerine bir araştırma. Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Doktora tezi, Adana.
- Saraç, E. (2012). Yapay sinir ağları metodu ile gayrimenkul değerlendirme. İstanbul Kültür Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Schulz, M. A. R. (2003). Valuation of properties and economic models of real estate markets. Humboldt-University, Wirtschaftswissenschaftlichen Fakultät, PhD, Berlin.
- Sdino, L., Rosasco, P., Torrieri, F., & Oppio, A. (2018). A Mass Appraisal Model Based on Multi-Criteria Evaluation: an Application to the Property Portfolio of the Bank of Italy. *NMP2018*, 128 (7), 1-10.
- Semmlow, J. L. (2004). Biosignal and biomedical image processing, MATLAB-Based applications
- Smigielski, E. (2014). Effect of New York State Forests on residential property values. State University of New York, MSc, Syracuse, New York.
- Son, K. (2012). Regression model predicting appraised unit value of land in San Francisco county from number of and distance to public transit stops using GIS. Texas A&M University, PhD, Texas.
- Timm, N. H. (2002). Applied multivariate analysis Retrieved from http://hbanaszak.mjr.uw.edu.pl/TempTxt/Shorts/_XeL3THpJaw.pdf

Unel, F. B., Yalpir, S., & Gulnar, B. (2017). Preference Changes Depending on Age Groups of Criteria Affecting the Real Estate Value. *International Journal of Engineering and Geosciences (IJEG)*, 2(2), 41-51.

Ünel, F. B. (2017). Taşınmaz Değerleme Kriterlerine Yönelik Coğrafi Veri Modelinin Geliştirilmesi. Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, Konya.

Yalpir, S., & Unel, F. B. (2017). Use of Spatial Analysis Methods in Land Appraisal; Konya Example. Paper presented at the 5th International Symposium on Innovative Technologies in Engineering and Science (ISITES2017), Baku, Azerbaijan.

Yalpir, Ş. (2007). Bulanık mantık metodolojisi ile taşınmaz değerlendirme modelinin geliştirilmesi ve uygulaması: Konya örneği. Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, Konya.

Yalpir, Ş., & Ünel, F. B. (2016). Türkiye 'de ve Uluslararası çalışmalarda arsa değerlemede kullanılan kriterlerin irdelenmesi ve Faktör Analizi ile azaltım. *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 16(025502), 303-322. doi:10.5578/fmbd.28134

Zavadskas, E. K., Bausys, R., Kaklauskas, A., Ubarte, I., Kuzminske, A., & Gudiene, N. (2017). Sustainable market valuation of buildings by the single-valued neutrosophic MAMVA method. *Applied Soft Computing*, 57, 74–87.

Zoppi, C., Argiolas, M., & Lai, S. (2015). Factors influencing the value of houses: Estimates for the city of Cagliari, Italy. *Land Use Policy*, 42, 367–380.