INVESTIGATION OF HOUSING VALUATION MODELS BASED ON SPATIAL AND NON-SPATIAL TECHNIQUES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ERTUĞRUL BOZA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
GEODETIC AND GEOGRAPHIC INFORMATION TECHNOLOGIES

MAY 2015

Approval of the thesis:

INVESTIGATION OF HOUSING VALUATION MODELS BASED ON SPATIAL TECHNIQUES

submitted by **ERTUĞRUL BOZA** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Geodetic and Geographic Information Technologies Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver _________________
Dean, Graduate School of **Natural and Applied Sciences**

Assoc. Prof. Dr. Uğur Murat Leloğlu _________________
Head of Dept., **Geodetic & Geographic Infor. Technologies**

Prof. Dr. Şebnem H. Düzgün _________________
Supervisor, **Mining Engineering Dept., METU**

Prof. Dr. Ali Türel _________________
Co-supervisor, **City and Regional Planning Dept., METU**

Examining Committee Members:

Assoc. Prof. Dr. Nurünnisa Usul _________________
Civil Engineering Dept., METU

Prof. Dr. Şebnem H. Düzgün _________________
Mining Engineering Dept., METU

Assoc. Prof. Dr. Çiğdem Varol _________________
City and Regional Planning Dept., Gazi University

Assoc. Prof. Dr. Arzu Erener _________________
Geodesy and Geoinformation Engineering Dept., Kocaeli Uni.

Assoc. Prof. Dr. Volkan Çağdaş _________________
Geomatic Engineering Dept., Yıldız Technical University

Date: 28.05.2015

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Ertuğrul BOZA

Signature :

ABSTRACT

INVESTIGATION OF HOUSING VALUATION MODELS BASED ON SPATIAL AND NON-SPATIAL TECHNIQUES

Boza, Ertuğrul
Ph.D., Department of Geodetic and Geographic Information Technologies
Supervisor: Prof. Dr. Şebnem H. Düzgün
Co-supervisor: Prof. Dr. Ali Türel

May 2015, 307 pages

The aim of this thesis is to develop hedonic housing valuation models based on spatial (SAR-simultaneous spatial autoregression and GWR - geographically weighted regression) and non-spatial (OLS - ordinary least squares) techniques, to compare the performances of these models and to investigate significant factors affecting housing value. The developed housing valuation models were tested at the Çankaya and Keçiören districts of Ankara province, Turkey.

The results of the analyses revealed that significant spatial non-stationarity exists between the dependent and independent variables. A semi-logarithmic hedonic model was used in order to interpret the coefficients easily and minimize the problem of heteroscedasticity. The results show that Area, Security and Distance to Shopping Center are common significant factors for both Çankaya and Keçiören districts in Ankara. Other important factors are the Type of Property and Distance to Subway for Çankaya and the Floor and Household variables for Keçiören.

The SAR and the GWR spatial models gave a better approximation to the observed house values than the traditional non-spatial regression model. The SAR model showed the best performance in Çankaya and the GWR model indicated high performance in Keçiören. The GWR maps displayed the variation of the coefficients of each variable clearly.

Keywords: Hedonic Housing Pricing Method, GIS-based Housing Valuation, Spatial and Non-spatial Housing Valuation

ÖZ

MEKANSAL VE MEKANSAL OLMAYAN TEKNİKLERE DAYALI KONUT DEĞERLEME MODELLERİNİN İNCELENMESİ

Boza, Ertuğrul
Doktora, Jeodezi ve Coğrafi Bilgi Teknolojileri Bölümü

Tez Yöneticisi: Prof. Dr. Şebnem H. Düzgün

Ortak Tez Yöneticisi: Prof. Dr. Ali Türel

Mayıs 2015, 307 sayfa

Bu tezin amacı, mekânsal (eşzamanlı mekânsal otoregresyon ve coğrafi ağırlıklandırılmış regresyon) ve mekânsal olmayan (en küçük kareler) tekniklerine dayalı hedonik konut değerleme modelleri geliştirmek, bu modellerin performanslarını karşılaştırmak ve konut değerini etkileyen önemli faktörleri araştırmaktır. Geliştirilen konut değerleme modelleri Ankara ilinin Çankaya ve Keçiören ilçelerinde test edilmişlerdir.

Analizlerin sonuçları, bağımlı ve bağımsız değişkenler arasında önemli mekânsal değişimin varlığını ortaya çıkarmıştır. Katsayıların yorumunu kolaylaştırmak ve değişen varyans sorununu minimize etmek için yarı-logaritmik hedonik model kullanılmıştır. Sonuçlar, Alan, Güvenlik ve Alış-Veriş Merkezine Uzaklık faktörlerinin hem Çankaya hem de Keçiören için ortak önemli faktörler olduğunu göstermiştir. Çankaya için Mülkiyet Tipi ve Metroya Uzaklık ve Keçiören için Dairenin Katı ve Hanehalkı konut değerine etki eden diğer önemli faktörlerdir.

Mekânsal ve mekânsal olmayan modellerin performansları gerçek dünya uygulamaları ile test edilmiştir. SAR ve GWR mekânsal modelleri, mekânsal olmayan geleneksel regresyon modeline göre daha iyi performans göstermektedir. GWR haritaları çalışma alanları boyunca her bir değişkenin katsayılarındaki değişkenliği açıkça göstermektedir.

Anahtar kelimeler: Hedonik Konut Fiyat Metodu, CBS Tabanlı Konut Değerleme, Mekânsal ve Mekânsal Olmayan Konut Değerleme

To My Family

Berin, Selin and Seçil

In love, gratitude, and respect.

ACKNOWLEDGEMENTS

I first thank Allah (SWT) for his infinite mercies, and strength for enabling me to complete my doctoral degree successfully. I pray to Allah (SWT) to bless the education I acquired and make it beneficial not only to myself and my family but to the Muslim Umma and humanity in general, amen.

I would like to express my deep heartfelt thanks to my supervisor, Prof. Dr. Şebnem H. Düzgün for her keen and intelligent guidance throughout the research. Her great personality, patience and thoughtfulness made this study possible. Special thanks go to my co-supervisor Prof. Dr. Ali Türel for his outstanding contributions. I also extend my appreciation to my thesis committee and jury members, Assoc. Prof. Dr. Nurünnisa Usul, Assoc. Prof. Dr. Çiğdem Varol, Assoc. Prof. Dr. Arzu Erener and Assoc. Prof. Dr. Volkan Çağdaş. Their thoughtful comments and suggestions made a significant improvement of my dissertation.

First, I must thank God for giving me the ability to think and write. Next, I would like to thank my beautiful wife, Seçil, for her support and patience throughout this dissertation. I also want to thank my lovely daughters Selin and Berin for making my life so special. I wish to express my sincere gratitude to my beloved wife Seçil Boza for her unconditional love, morale support and prayers. I am also deeply indebted to all my father, mother, brothers and sister for their support and prayers. Especially, I owe a great deal of appreciation to my father Ali Boza and brother Zeki Boza.

I am also indebted to Hasan Adsız Kahraman for their friendship and assistance during the geocoding of my work. I also thank to my colleagues Leyla Bilen Kazancı, Ercan Boyar and Özhan Yılmaz for their guidance to non-spatial analysis. Especially, I extend my deep appreciation to Şinasi Bayraktar who gave the support to obtain the dataset necessary for my thesis believing that it would be useful for the real estate sector.

I would like to sincerely acknowledge to the Middle East Technical University and the Turkey Valuation Experts Association with support in obtaining the necessary data for my PhD studies. I am also grateful to the General Directorate of Civil Registration and Nationality of the Republic of Turkey for the opportunity and support given to me to pursue this great achievement.

Thanks are also due to all the academic staff of Geodetic and Geographic Information Technologies (GGIT) department. Finally, my deepest thanks are due to all those who contributed towards the successful completion of this research and my studies in general.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

ABBREVIATIONS

ANOVA	Analysis of Variance
CBD	Central Business District
CL	Confidence Level
CLT	Central Limit Theorem
E	Euclidean Distance
GIS	Geographic Information System
GWR	Geographically Weighted Regression
IVSC	International Valuation Standards Council
HPI	House Price Index
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
METU	Middle East Technical University
ML	Maximum Likelihood
MRA	Multiple Regression Analysis
N	Network Route Distance
OLS	Ordinary Least Square
RLMerr	Robust Lagrange Multiplier for Error Model
RLMlag	Robust Lagrange Multiplier for Lag Model
RMSE	Root Mean Square Error
SAC	Spatial Autocorrelation
SAR	Simultaneous Spatial Autoregression
SARerr	Simultaneous Spatial Autoregression Error Model
SARlag	Simultaneous Spatial Autoregression Lag Model
S.E.	Standard Error
SH	Spatial heterogeneity
TIN	Triangulated Irregular Network
TAVE	Turkish Association of Valuation Experts
TurkStat	Turkish Statistical Institute

LIST OF SYMBOLS

AIC	Akaike Information Criterion
AICc	Corrected Akaike Information Criterion
Ho	Null hypothesis
LnP	Logarithm of house value based on market
R^2	Coefficient of determination
adj R^2	Adjusted coefficient of determination
VIF	Variance inflation factor
r	Pearson's product moment correlation coefficient
α	Significance level
ε	Error vector
n	Number of data points
σ	Spatial autoregressive coefficient (Spatial Error Model)
λ	Spatial autoregressive coefficient (Spatial Lag Model)
k	Number of explanatory variables
I	Moran's I matrix
W	Spatial weighted matrix
p	Simultaneous autoregressive lag coefficient
y	Dependent variable (value)

CHAPTER 1

INTRODUCTION

1.1 Purpose and Scope of the Research

The market value of real estate is an important variable in terms of the economy of a country. There are a number of different types of real estate, such as house, land, office, store, hotel, and so on. Thus, different methods exist for their valuation and there are numerous factors that affect their value. In this thesis, it was focused on housing valuation because it is important for monitoring the economic conditions of a country. The place and importance of housing market within the general economy of countries has been receiving growing attention. House prices have boomed worldwide from time to time and sharp changes that emerged in a short time in housing prices have led to economic crisis. Changes in housing prices are considered to be an important indicator of economic vulnerability. Therefore, the efforts aimed at monitoring for these changes at local, regional, national or international economy level are currently on the increase. Countries pay bigger attention to investigating house price determinants and tracking to variation on it. However, the determinants of housing market vary from neighborhood to neighborhood, district to district, city to city and country to country. House values are needed in many applications including finance, taxation, expropriation, mortgage, rent/lease, zoning and urban transformation (Yetgin and Lepkova, 2005; Uluçay and Tecim, 2009). Capital market institutions, banks, private people, individual and institutional investors and courts need house values for their activities in the country. On the other hand, housing as a substantial investment tool is in competition with other assets. Especially in Turkey, there is a widely-held perception among people that housing is one of the best investment instruments. The main instruments of investment in Turkey are foreign currency, gold, deposit accounts, land, housing and stock market. However, people generally prefer investment on housing or land when the housing market offers a favorable investment opportunity in Turkey. A widespread belief among people of Turkey says that investment in real estate never makes its investor lose. Therefore, real estate is always regarded as a reliable investment tool by most investors. In cases against expropriation in Turkey, there are numerous judicial decisions of both the European Court of Human Rights and the Council of State of Turkey as to the rejection of expropriation. The main reason of these decisions is that the expropriation price of a disputed house is not determined by objective criteria which affect its value, such as locational, structural and environmental characteristics. For these reasons, housing valuation plays a major role in the capital market of Turkey.

Housing valuation is the process of determining the value of a house at a certain time considering its internal and external factors. Determination of a house value depends on a number of physical, economic, social and environmental characteristics, which is known as the hedonic approach. There are numerous studies to investigate the relationship between house prices and housing characteristics using hedonic based regressions. Housing valuation is important for monitoring the economic conditions of the country. Therefore, many methods have been applied in the housing market in order to construct house price indices (HPI). Since monitoring the developments in house prices is an important factor for monetary policy decisions, the determination of real market prices for houses is also important for promoting financial stability (Kaya et al., 2013). In the real estate valuation literature, residential valuation or dwelling valuation is used in the same sense as housing valuation. House value is stated in the literature as market value, market price, benefit value, income value or real value. Price is the amount asked, offered or paid for a good or a service (International Valuation Standards Council; IVSC, 2005:25). Unlike price, value is an estimate of the likely price to be paid at a given time (IVSC, 2005:26). The term market value is usually interchangeable with open market value or real value. However, the market price of a house may not be equal to its market value. Market price is the price agreed upon by a willing buyer and a willing seller (Baum and Crosby, 1995). The transaction of goods depends on a reasonable time on which buyers and sellers have agreed. The market price of a house may indicate a more rapid change compared to its market value. If there is no compulsion to buy or sell, value and price are the same. Market value of a house is the determination of price based on a detailed analysis of its bundles according to structural, environmental, and locational factors (Freeman, 1979; Harrison and Rubinfeld, 1978; Bowes and Ihlanfeldt, 2001; Shultz and King, 2001). Various combinations of these variables have been used to model house values in the housing valuation literature. Therefore, the potential impacts of each of these comparable internal and external factors on house value must be specified separately. For this purpose, numerous housing valuation estimation models have been developed. In the housing valuation literature, hedonic pricing methods and their combinations are the most frequently preferred models. These predictive models for housing valuation are based on various spatial or non-spatial valuation techniques. The hedonic pricing method, a typical form of regression analysis, was adopted to determine the regression coefficients for housing values in this study. In statistical terms, there is a major weakness in empirical hedonic price models since they do not take into account the underlying spatial dependency across neighborhoods (Dubin, 1998).

In order to cope with this weakness of hedonic pricing methods, the Spatial Autoregressive (SAR) model and the Geographically Weighted Regression (GWR) model are employed within the context of our research.

The SAR is a global model, whereas the GWR is a local version of spatial regression. To obtain more reliable results, the SAR and the GWR models were applied to house values. The GWR provided opportunities to get more information on the data since it is a local model (Erener and Duzgun, 2012). In recent years, local analysis has been the most preferred one among spatial analyses types (Gao and Asami, 2005). Especially the GWR is proposed for large sample areas because the coefficients vary with location.

There are a limited number of studies on the housing valuation in Turkey. Empirical studies on housing valuation were performed in some parts of İstanbul (Keskin, 2008; Ozus, 2009; Ozus et al., 2007; Koramaz and Dokmeci, 2012; Topcu and Kubat, 2009; Bulut, et al., 2010), and in some other provinces such as Ankara (Ayan and Erkin, 2014; Gultekin and Yamamura, 2002; Kaya, 2012), İzmir (Celik and Yankaya, 2006), Trabzon (Yomralioglu and Nisanci, 2004), Erzurum (Yilmaz et al., 2008) and Konya (Yalpir and Unel, 2014; Yalpir and Özkan, 2011).

In previous studies, the data used for housing valuation was obtained from real estate agents, questionnaires and public institutions. Mostly the hedonic pricing method and the sales comparison method were used in these early empirical studies. However, to the author's best knowledge, this is the first study on housing valuation that takes into account both spatial and non-spatial statistical techniques in Turkey.

There is still a considerable need for further studies on housing valuation in Turkey in order to see how housing prices vary across the country and determine local and global factors affecting the house value. Besides, there is a need to investigate the validity of the findings in the dwelling valuation literature for Turkey.

In terms of macroeconomic studies, the two most important factors that influence the value of a house are supply and demand. Economic, social and structural differences between urban areas affect supply and demand level on the housing market. It is difficult to predict the demand for housings. Instead, a house can be decomposed into its internal and external characteristics such as the number of bedrooms, the area, the floor, the size of lot, or the distance to the city center and prices can be estimated for each of them separately. This approach is known as the hedonic price method. This method provides significant advantages for the studies on microeconomics (Malpezzi et al., 1987). In the studies of the hedonic pricing method, location is considered as the most important factor affecting the market value of a house (Kiel and Zabel, 2008; Archer et al., 1996). Besides being the key factor in the field of valuation of housing, studies have recently shown that it is also an important determinant to explain the rental price changes (Ustaoglu et al., 2013). These constitute the basic determination of house prices. In order to detect the influence of social, economic, structural and locational differences on housing valuation, two study areas were selected that could reflect this difference. On the other hand, the attributes for the case studies were determined for this purpose.

1.2 The Main Contributions

The aim of this thesis is to construct global and local housing valuation models using spatial and non-spatial techniques, to compare the performances of these models in two districts with a different economic, social and environmental texture and to evaluate the findings in terms of international and national findings. In this context, the most appropriate models and attributes for housing valuation were determined for Turkey. For this purpose, the hedonic model, the OLS (Ordinary Least Square), the spatial lag, the spatial error and the GWR (Geographically Weighted Regression) were examined and compared for modeling the housing valuation by means of a case study in Ankara, the capital city of Turkey. The proposed models were validated using the house price data in Çankaya and Keçiören districts in Ankara.

As a result, this thesis makes distinct contributions in the field of housing valuation. This study, which uses spatial and non-spatial techniques, has proved the effects of social, economic, structural and locational differences on housing valuation by examining two study areas which reflect these differences. On the other hand, the variables used in the case studies were determined and generated for this purpose.

The secondary contributions of this thesis are threefold. First of all, the dwelling valuation models were constructed based on spatial and non-spatial statistical techniques for the first time in Turkey. The second contribution is that it was studied with the data collected according to the national real estate valuation standards (Turkish Capital Market Standards). In other words, again for the first time, standard data was used for a doctoral thesis. The final one is that theft events related to the work areas were obtained from some insurance and security companies in Ankara. Therefore, the effects of theft events on house values were investigated for the first time in Turkey.

This methodology will provide a more robust model of housing valuation for Turkey. The methodology is suggested for studies on housing estimation models and for selection of factors that affect house values in Turkey.

1.3 Outline of the Thesis

This thesis is structured in six chapters. Chapter 2 gives a review on the spatial and non-spatial housing valuation models. A brief summary regarding the variables used widely for building the housing predictive models is given and some important findings are listed in Chapter 2. Chapter 3 presents the general methodology adopted for this thesis. Chapter 4 explains how these methods are applied to the residential valuation using the case studies for the Çankaya and Keçiören districts, Ankara, in

Turkey and gives the associated results. Chapter 5 discusses the results of the implementation of the methodology based on the case studies. Finally, Chapter 6 presents the conclusion of the study and the future outlook section.

CHAPTER 2

LITERATURE SURVEY

2.1 Studies on Housing Valuation Based Models

The housing market affects the financial stability both directly and indirectly. Therefore, the determination of the housing price is important with regard to the economic structure of countries. There are numerous studies to build estimation models for house values and determine the factors affecting housing values. Since the focus in this thesis was directed towards the estimation models of housing valuation based on spatial and non-spatial techniques, special attention was paid to studies on this subject. The literature on housing valuation was examined in terms of two basic concepts in this thesis: housing valuation models and factors affecting housing values. In most of the studies reviewed in this chapter, theoretical analyses are buttressed by empirical research.

Determining the demand for housing is a challenging empirical problem. Many approaches have been developed to tackle this problem. Pagourtzi et al., (2003) classified housing valuation methods as traditional and advanced. The most widely used traditional methods are the sales comparison method, the cost method and the income capitalization method (Karagöl, 2007; Jaffe and Sirmans, 1995). All these methods have been used for a long time in most of the developed countries such as France, the U.K., Germany, Canada, the U.S., and in developing countries such as Brazil, Iran, Malaysia, and Turkey. They all focus on different aspects of the real estate object. The review of these methods will be explained in turn below with their pros and cons. The sales comparison method estimates the value of a house by comparing the sales prices of similar houses sold in similar locations within a recent period of time. It is a simple and widely used method in the residential housing market. Although it is one of most commonly used valuation methods, its power of estimation is strongly dependent on the existence of enough comparable sales and the quality of these. Another limitation is that reflecting the prices of previous sales to the present day is rather difficult. The cost method is based on estimating the current cost of construction and subtracting the physical, functional and environmental depreciation from the cost value. Market value is predicted by comparing the house being valued to similar houses that have recently been sold. This method is particularly useful in valuing new houses in the residential market. The cost approach is the best alternative when there is a lack of information in the sales comparison and the income approach. It is also accepted in the real estate valuation literature as the most difficult one of the three methods. An initial limitation with the cost approach is the assumption that value is derived through costs minus depreciation. It is difficult to measure the value of the depreciation and appreciation in the cost method. The income method is particularly common in commercial real estate appraisal and in business appraisal. It focuses directly on the value of the property to the individual concerned. The net income of the constructed real property consists of building and land income (Gür et al., 2002). The expected future cash flows are taken into account in the income approach. The value according to the income approach is the present value of future cash flows. This method disregards the actual market prices for property by ignoring the comparable sales analysis.

The weaknesses of these traditional methods, which are the existence of depreciation, the deviation from highest and the best use that would distort the income and the lack of comparable samples, have led researchers to develop new housing valuation methods called advanced methods. These methods include hedonic, artificial intelligence, neural network, fuzzy logic, expert systems and genetic algorithm methods. As hedonic models will also be used in this thesis, the literature studies related these models are discussed in more detail below.

Each house is unique in terms of its location and other characteristics. In other words, there are no two houses that are completely identical to each other. Because of this heterogeneity, it is not possible to mention a housing valuation model applicable all over the world. Although there is no single unified model for housing valuation, in the housing valuation literature, the developed models are generally based on the hedonic methodology (Karagöl, 2007). In the hedonic model, the internal and external characteristics of a house are separately taken into consideration. The model describes a market equilibrium produced by the interaction between demand and supply in the urban housing market. The hedonic hypothesis is that goods are valued for their utility-bearing attributes or characteristics. In this context, the hedonic pricing model aims at explaining the specific contribution of each attribute of a house on its overall price (Can, 1990; Can, 1992; Dubin, 1998). In the hedonic regression, the value of individual characteristics cannot be directly monitored. In other words, the hedonic regression estimates give the implicit values of each structural characteristics, neighborhood characteristics, and environmental characteristics. Therefore, this analysis is used to estimate the relative contribution of individual variables to the total values of the housing (Kestens et al., 2006).

It is a commonly used method to estimate dwelling valuations by means of regression analyses (Ustaoğlu, 2003). The hedonic modelling requires the use of the linear regression and the OLS regression method to estimate the implicit prices of each variable or characteristic (Farber and Yeates, 2006). If there is one or more than one independent variable, the first method that may come to mind is the multiple regression analysis method. The ordinary least squares or the multiple linear regression analysis is used in the majority of hedonic research.

The theoretical foundation of the hedonic models is based on the work by Lancaster (1966) and Rosen (1974). The general form of hedonic price function is as follows:

$$P = \alpha_0 + \sum \alpha_i Z_i + \varepsilon_i \qquad [1]$$

where:

P_i = A house value
α = Coefficients
Z_i = A vector of housing characteristics variables
ε = Random error

The relationship between housing price and its characteristics can be classified into 3 categories as follows (Chin and Chau, 2003):

$$\text{House Value/Price} = f\,(L, S, N) + \varepsilon \qquad [2]$$

where L denotes locational attributes, S denotes structural attributes and N denotes neighborhood attributes and ε represents an independent and normally distributed error term.

The structural factors include the lot size, the total square feet of living space (area), room size, floor, the age of building, the number of bedrooms and the number of bathrooms, the availability of balcony, kitchen, toilet, etc. The neighborhood factors include security services, crime rate and pollution. And finally, locational factors consist of accessibility to jobs, accessibility to schools, accessibility to public transport, accessibility to hospitals, accessibility to shopping centers, accessibility to subway, and so on. It is possible to encounter different hedonic price formulas in the literature, which are based on urban generation, demographic, socio-cultural and micro- and macro-economic factors.

Empirical researchers usually refer to three possible model specifications: linear, semi-log, and log-log. In the linear functional form, it is assumed that each attribute is obtained independently from the other attributes in the model. The estimated coefficients present the actual magnitude of attribute prices. This functional form is rarely used in practice. In log-linear or semi-logarithmic functional form, the dependent variable (in this case housing values) is logged. In other words, taking only the logarithm of one side of a regression (dependent variable) is called semi-log transformation (Adair et al., 1996). The most common functional form recommended in the hedonic literature is the semi-logarithmic form. Bello and Moruf (2010), in their study on housing valuation, state that semi-log functional forms of hedonic price models are the best fit data with respect to the coefficient of determination (R^2). The resulting coefficient estimates enable users to calculate the percentage change in value for a one-unit change in the given variable (Sirmans et al., 2005). The semi-logarithmic hedonic equation minimizes the problem of heteroscedasticity (Ottensmann et al., 2008). It has some advantages related to the easy interpretation of coefficients. Moreover, it reduces the effect of non-linear relationship between market price and the explanatory variables (Malpezzi, 2003). In log-log functional form, all continuous variables on the left hand side and the right hand side in the model equation are logged (Fik et al., 2003; Pearson et al., 2002). The log transformation is only applicable when all the observations in the data set are positive. A disadvantage of the log-log model is that it is much more difficult to fit the data than a straight line.

There are no clear rules or guidance which will help to make a choice among hedonic functional forms in the hedonic literature.

It is explained below how to interpret different hedonic regression models (Bello and Moruf, 2010). In the following formulations, y represents the dependent variable, x is the independent variable, β is the y-intercept, β is the slope coefficient, and ln(y) and ln(x) represent the natural logarithm of y and x, respectively. ε denotes an error term.

(1) Linear form: $y = \beta + \beta x + \varepsilon$

In this functional form β represents the change in y (in units of y) that will occur as x changes one unit.

(2) Semi-log form: $\ln(y) = \beta + \beta x + \varepsilon$

In this functional form β is interpreted as follows. A one-unit change in x will cause a β (100) % change in y, e.g., if the estimated coefficient is 0.03, it means that a one-unit increase in x will generate a 3% increase in y.

(3) Log-log (double-log) form: $\ln(y) = \beta + \beta \ln(x) + \varepsilon$

In this functional form, β is the elasticity coefficient. A one-percent change in x will cause a β% change in y, e.g., if the estimated coefficient is a -3, this means that a 1% increase in x will generate a -3% decrease in y.

Zabel and Kiel (2000) used a data set of properties in four cities from 1974 to 1991 in order to estimate demand equations for air quality. They compared three main specifications of the hedonic equation: the linear, log-linear, and log-log models. The linear is found to be the worst, but the other two forms yield relatively similar results.

Traditional hedonic models are used widely in housing valuation studies; however, they do not consider spatial relationships between variables (Dubin, 1992). In other words, the spatial autocorrelation (SAC) and the spatial heterogeneity (SH) are two main challenges in the hedonic modeling (Helbich et al., 2013). Some leading researchers in the field of spatial statistics such as Cassetti (1972), Anselin (1988;

1990), Can (1990), Dubin (1998), and Fotheringham et al., (1998) are in consensus that a hedonic model based on only the OLS is inefficient. Therefore, they suggested making use of the spatial characteristics of variables to improve the efficiency of the models. Various spatial valuation techniques have emerged to determine spatial effects. The SAR and the GWR models have been widely used to control spatial effects (Efthymiou and Antoniou, 2013). These two regression techniques will also be used to investigate the effect of spatial relationships in the housing valuation in this study. Particularly the GWR will be used to demonstrate the spatial variation in the study areas and to estimate the regression coefficients at different points in space. In contrast to the GWR, the results of the SAR are valid for the whole study areas. The GWR model may be regarded as the one which accounts for the spatial variation in house prices the best (Helbich et al., 2013). The GWR technique is a newly developed statistical methodology useful for modeling spatial non-stationarity among regressed relationships. These advanced spatial techniques, which incorporate geographic information systems, enhance the possibilities of handling location in the hedonic-based housing valuation analysis. Another notable empirical study based on the hedonic and advanced spatial techniques are listed with the data used and their findings in Table 1.

Table 1: Review of notable papers for housing valuation models

Purpose	Model Used and Comparison Criteria of Models	Variables Used	Data Used/ Study Area	Findings	Source
An alternative empirical evaluation method for regression models	Hedonic, SAR and GWR. Empirical evaluation of hedonic, spatial dependency and GWR models	Floor area, lot area, dist.to.nearest station,age,dist.to.central city area,quality of nearby buildings (dummy), sun shine duration in hour, proximity to park, greenery in the neighborhood (dummy).	The data set has a sample of 190 properties. Western Tokyo	The result shows that spatial dependency model, the GWR model, and the mixed model is significantly better than the basic hedonic model (OLS)	Gao, et al., (2006)

Table 1: Review of notable papers for housing valuation models (continued)

Purpose	Model Used and Comparison Criteria of Models	Variables Used	Data Used/ Study Area	Findings	Source
Examining whether there are omitted variables in regressions on housing valuation	Hedonic and GWR. Global regression (coefficients are stationary over space) and local regression (regression coefficients are vary over space). R-square and AIC	Floor area, lot area,distance to nearest station, age,landscaping, proximity to center of city (min), lot frontage, number of parks, continuity to park (dummy), greenery (dummy), density of population.	The data set has a sample of 190 properties. Western Tokyo	The global model is convenient for western Tokyo data set, but the GWR model is slightly better.GWR was predict the area-associated variables stronger than global regressions done. Proximity to a large park has a significant positive effect on the house prices.	Gao and Asami (2005)
Investigate the relationship between public transport accessibility and residential land value	Hedonic house price model, Spatial Autoregressive Regression (SAR), GWR and Moving Window Regression (MWR). Akaike information criterion is used to compare global hedonic regression model and GWR models.	Area, age, sale data, size, quality, distance to mall, distance to water, distance to green, floor number, public transport availability, higher education institutes.	1000 flats observation. Riga, Letonia.	The GWR regressions are significantly better than global hedonic regressions. Every new transport route and bus stop will increase flat prices for places outside the city center. There is no significant relationship between house values and transport accessibility.	Dmitry (2009)

Table 1: Review of notable papers for housing valuation models (continued)

Purpose	Model Used and Comparison Criteria of Models	Variables Used	Data Used/ Study Area	Findings	Source
				The coefficients of distances to water, supermarkets, and higher education institutes are negative. In contrast these, the coefficients of area and new project have positive sign.	
Comparison of Localized Regression Models in a Hedonic House Price Context	Hedonic house price model, Spatial Autoregressive Regression (SAR), GWR and Moving Window Regression (MWR) R^2 (the coefficient of determination), SSE (Sum of Squares of Error) and pseudo-R^2 (the squared correlation coefficient between the observed and the predicted values).	Area, age, sale data, size, quality, distance to the downtown, distance to mall, income, price paid for a house in any neighbourhood (PC_FOR)	The data set consists of 19,007 housing sales taking place between July 2000 and June 2001 in the City of Toronto. Sales prices (normalized through a logarithmic transformation). Toronto-japan	The SAR model provides an improvement over the OLS hedonic model. GWR residuals are better than the SAR model and Moving Window Regression (a special case of a GWR, only weight matrix differ from GWR).	Farber and Yeates, 2006

Table 1: Review of notable papers for housing valuation models (continued)

Purpose	Model Used and Comparison Criteria of Models	Variables Used	Data Used/ Study Area	Findings	Source
				The three variables having the most impact on variation in house prices are the area of the house (positive),the age of the property (negative) and distance to downtown (negative). Distance to mall has a negative effect on house values.	
Comparison the quality of prediction for several models	To compare the effectiveness of OLS, Spatial Expansion, spatial lag, spatial error and GWR	Flat area, number of rooms, floor, building type, year of construction, and the presence of the garage and location	Wroclaw	The geographically weighted regression is the best fit to the data among the presented methods	Chrostek, and Kopczewska, 2013
Investigate the effects of attribute reducing on real-estate valuation	Multiple regressions. R^2 is used to compare the models. Considering R^2 value closer to 1 is the best model.	9 factors used as input: the size of the house, floor information, facade, the age of the building, road conditions, the distance to public transport, the distance to education sites,	190 flats data collected from land agencies. 171 data are used for data modeling and 19 data are used for testing the models.	The model with reduced attributes (9 attributes) has better performance than the model without reduced (14 attributes) the attributes.	Bulut et al., (2011)

Table 1: Review of notable papers for housing valuation models (continued)

Purpose	Model Used and Comparison Criteria of Models	Variables Used	Data Used/ Study Area	Findings	Source
		The distance to health centers, the distance to parks and 1 output: market value.	Konya, Turkey		

Two important components for housing valuation are parameter estimation and model selection. The review of some notable studies published recently in the housing valuation literature, and the most important variables that affect housing valuation are summarized below.

Table 2: Summaries of notable findings on housing valuation

Findings/Results	Source
Using the real observed values of the sample as a measure to evaluate the spatial and non-spatial models shows that none of the proposed spatial dependency model, the GWR model, and the mixed model is significantly better than the basic hedonic model (OLS).	Gao et al., 2006
The global model is convenient for western Tokyo data set, but the GWR model is slightly better. The GWR predicted the area-associated variables were stronger than the global regressions. Proximity to a large park has a significant positive effect on the house prices.	Gao and Asami, 2005
The SAR model provides an improvement over the OLS hedonic model. GWR residuals are better than the SAR model and Moving Window Regression (a special case of a GWR, only weight matrix differs from the GWR). On the other hand, according to this study, the three variables creating the biggest impact on variation in house prices are the area of the house (positive), the age of the property (negative), and distance to downtown (negative). Distance to mall has a negative effect on house values.	Farber and Yeates, 2006
Geographically weighted regression is the best fit to the data among the presented methods.	Chrostek and Kopczewska, 2013
The model with reduced attributes (nine attributes) has better performance than the model without reduced (14 attributes) attributes.	Bulut et al., 2011

Table 2: Summaries of notable findings on housing valuation (continued)

Findings/Results	Source
The most important structural variable is floor area. Floor area, the size of the dwelling, the number of rooms and bathrooms are positively related to the price of housing.	Karagöl, 2007
The results show that the main variable influencing the price is the living area of the dwelling. Other statistically significant variables are the size of the balcony, the number of bathrooms, the age of the building, the existence of elevator and the existence of a small storeroom.	Morancho, 2003
Land area, main floor area and position are more significant factors affecting housing value.	İsmail et al., 2008
GWR models and Spatial Expansion Methods were used to analyze based on 11,732 transactions in 2000 houses in Tucson. Important variables used were dwelling area, air conditioning, number of rooms, structural quality of the dwelling, age of the dwelling, number of floors, number of bathrooms, interior quality of the dwelling and presence of a garage.	Bitter et al., 2007
The results show that the main variable influencing the price is the living area of the dwelling. Other statistically significant variables are the size of the balcony, the number of bathrooms, the age of the building, the existence of elevator and the existence of a small storeroom.	Morancho, 2003
Structural characteristics (square feet, lot size, bedrooms, bathrooms, and central air conditioning) are generally significant for most counties; age has a significant negative effect on price in most counties; a garage and in-ground pool significantly increase price although an above ground pool adds little value; a family room and dining room tended to be valued across countries.	Sirmans and Macpherson, 2003b
Location, market conditions, micro and macro-economic dynamics and building features are the most influential factors affecting the market values of residential properties. Crime levels, security, and accessibility/proximity of the property to centers of interests are the first important factors affecting housing values. Besides these, population density, size and the number of rooms are the second most influential factors on the house values.	Mbachu and Lenono, 2005
In a study conducted in Ankara, the existence of a negative relationship between homeownership and supply shopping center has been obtained.	Ozuduru and Varol, 2011
High education ratio in the district, the number of rooms, floor level and car parking have a significant positive impact to the price level. Higher educational attainment is supposed to be correlated with higher incomes. The number of rooms has a significant positive impact on the house prices.	Lehner, 2011

Table 2: Summaries of notable findings on housing valuation (continued)

Findings/Results	Source
The most important factors affecting house prices are floor area, sea view and heat insulation, respectively. Although the most important factor at the metropolitan level is sub-market, the other variables vary from one district to another.	Özüş et al., 2007
Neighborhood churches have negative impact on the values of nearby residential properties	Babawale, 2011
The number of bathrooms and bedrooms has an important effect on house values.	Neelawala, 2010
Among structural attributes, housing size and floor level are commonly found to affect house prices positively. Floor level is often expected to be positive due to better views and less polluted environment. Conversely, building age is found to be adversely affecting property value. The effects of the availability of shopping centers and sports stadiums have positive impact on nearby housing property prices.	Ki and Jayantha, 2010
The variation of the property value is explained by floor area, the presence of air conditioning, the presence of a fireplace, the number of bathrooms, age of the property, and type of surface.	Yu et al., 2007

The following chart was created on the most commonly used housing valuation parameters based on over 50 studies reviewed in the empirical housing valuation literature (Figure 1).

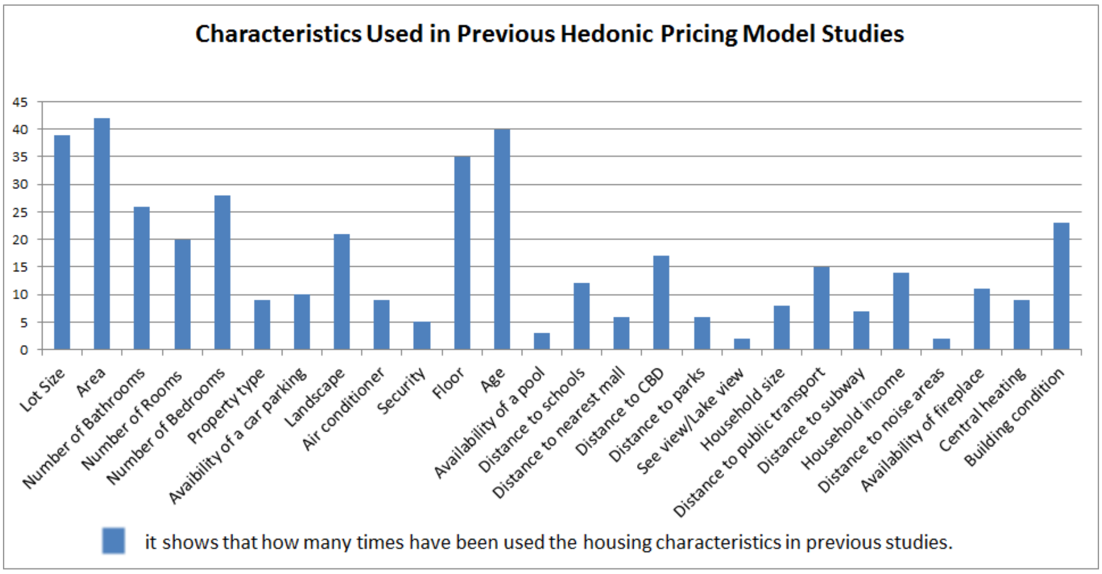


Figure 1: Important characteristics used in some previous housing valuation studies

In the literature of empirical housing valuation, there are two main research topics: determining key parameters that affect housing value and building a model capable of accurate estimation. Therefore, various housing valuation methods/models have been developed. Studies conducted to determine significant parameters as to the area of housing valuation and an analysis of academic studies chosen with regard to the methodologies that were used.

Yu et al., (2007) indicate that the structural attributes of housing and the neighborhood environment conditions are sufficient to construct a reliable housing valuation model. Floor size, air conditioner, fireplace and the number of bathrooms were positively associated with house values, whereas house age was negatively related with them. He states that the hedonic house price model is a powerful econometric tool in capturing important determinants of house prices/values. He shows the existence of significant non-stationary relationships between house values and all the selected structural and neighborhood attributes of housing using the GWR. Akaike Information Criterion and the ANOVA test were used to show that the GWR provides a significant improvement over the global OLS model. In other words, mapping GWR results showed that local modeling techniques are more robust than the global ones. This study relies on only the six structural and neighborhood attributes, so future studies should attempt to verify these results. The author also refers to this point because the GWR result revealed that important determinants are possibly missing.

Samapatti and Tay (2002) conducted a study to identify the hedonic factors and their impacts on the new house prices in small, medium and large developments using multiple regression analyses. They found that the structural, locational and neighborhood characteristics are important determinants of prices for small-sized developments and the locational attributes have a significant impact on house prices for the medium and large-sized developments. Also, proximity to the CBD, total floor area and road condition are important price factors. Aslan (2012) concluded that the real estate values can be determined more efficiently, economically and objectively using multi-criteria decision analysis and analytic hierarchy process integrated with GIS.

McMillen (2008) used quantile hedonic price function instead of the OLS to identify the different market segments and their implicit prices. He explains two main reasons for using the quantile hedonic price function as follows. Firstly, there is no limit to explain the mean of the dependent variable in contrast to the OLS. Secondly, it can be possible to explain the determinants of the dependent variable at any point of the distribution of the dependent variable. In terms of the hedonic price functions, quantile regression indicates the weight of each housing characteristics on the distribution of housing prices. According to OLS results, house prices increase with lot size, building area and the number of bathrooms and decline with age. The results of this study show that the effect of housing characteristics on housing value/price can be better explained by estimating quantile regressions. The results also indicate that the location and housing characteristics do not clarify the changes in the distribution of house prices. Higher-priced houses have certain housing characteristics different from lower-priced houses. However, there are some critics on this method that do not take into account spatial effects on data (Zietz et al., 2008). In a recent study performed by Bekar and Akay (2014), quantile hedonic regression and spatial dependence was taken into account together. For spatial analysis, a weight matrix is constructed according to the k-nearest neighbor criteria based on the Euclidean distances calculations. According to their findings of the spatial quantile regression model, although the effect of space does not have an impact on housing prices at the locations where housing prices are low, it has an increasing importance on the high-priced housings.

In the dwelling valuation literature, the effects of accessibility on housing values/prices have been measured using distance and/or travel times. The importance of accessibility for hedonic models has been examined by some researchers. Adair et al., (2000) tested in a monocentric city whether the effect of accessibility is on housing values. The results indicate that transportation accessibility has a limited impact on housing prices but it is not a significant factor to explain the changes in house prices. They found that household income limits housing choice. Kestens et al., (2006) tried to determine the effects of household income, the previous tenure status, and the age of household, the living area, the age of the property, the social status of the neighborhood (the percentage of university degree holders in the Census tract), and accessibility on house prices. One of the findings is that accessibility is one of the most significant factors on housing value (Case and Mayer, 1996). The other important findings are that educational attainment of the buyer and the household income are also important determinants for housing values because of their preference to maintain neighborhood homogeneity (Goodman and Thibodeau, 2003). In this study, the semi-log functional form of hedonic modeling has been performed using OLS specification.

Another study investigates the effect of accessibility on housing values in a different way. Chin and Foong (2006) investigated the effect of accessibility to prestigious schools (travel time and distance to prestigious primary schools and secondary schools) on the value of housing properties using a hedonic housing price model. The other important variables used in this study are price, floor area (m^2), floor, age, the availability of swimming pool, the distance to the major shopping district (km), the distance to central business district (km), and the distance to the nearest subway station (km). Some studies like Visser et al., (2008) demonstrated that characteristics of the residential environment (in this case, the accessibility to employment opportunities) can explain the regional variations of house prices.

Accessibility in the sense of distance to a point (km or mile) has been used in order to indicate the spatial effects on housing valuation recently. Bae et al., (2003) investigated the impact of the construction of a new subway line in Seoul on nearby residential property values via a hedonic pricing regression analysis. The hedonic model constructed for the study is a function of structure, neighborhood and accessibility variables. The important findings of this study are that floor space (area) is the most important structural variable; the heating system has an insignificant impact on house values. The most important result is that the distance from the Line5 subway station was significant in 1989, 1995 and 1997, but not in the year 2000. They interpreted the reason why the effect of subway station on housing valuation is insignificant in the year 2000 as follows: These price impacts are measured for four years (1989, 1995, 1997 and 2000), which correspond to the announcement of the subway, a year during construction, the completion date, and 3 years after its opening. After the announcement of the subway done in 1989, homeowners living near this subway line began to take advantage of it up to the completion date in 1997.

However, 3 years after its opening (in 2000), the advantage of the subway disappeared, since the economic rent provided by the subway line was already taken by householders living closer to the subway line.

There has been little empirical research related to the effects of subway station on housing values in Turkey. A similar study was conducted by Yankaya and Çelik (2005) to determine the impacts of İzmir (Turkey) subway on house values using the linear and log-linear functions of the hedonic price method. The data set used in this study was obtained from some real estate agents in İzmir through questionnaire. The findings are that the proximity to subway stations is a statistically significant determinant and the impact of transport investment on real estate values depends on transport costs, total vehicle time and the distance to the nearest station. However, the results show that there was no effect of the bus transportation on real estate values in İzmir. Parallel to these findings, Chen and Hao (2008) found that the availability of a subway increases the housing value very sharply.

Hedonic model does not explicitly take into account the spatial effects (spatial dependency and heterogeneity) among observations. Therefore, researchers have used other methods to take spatial effects into account. For this purpose, they have been done either by integrating hedonic methods with other well-known methods or by using a method which is completely different from the hedonic method. Recently, spatial statistical analysis has been used widely in the field of housing valuation. To predict housing sale prices, Wheeler et al., (2014) investigated the use of the Bayesian methods for hedonic price analysis. In this study, the linear regression model using ordinary least square was used besides the Bayesian method to explain and predict housing sale price and the GWR to detect spatial heterogeneity. A logarithmic transformation was applied to housing sale price to increase the linearity with log housing sale price. The authors assert that the Bayesian model performed much better than the linear regression model for both the estimation and prediction of hedonic prices. They reported that the Bayesian model has a very high goodness-of-fit and predictive power of the spatially varying coefficients than the GWR. The major disadvantage of this method, since it is a simulation-based estimation technique, is that there is too much computational burden. However, the GWR has been used widely to show spatial variation in parameter estimates and exhibits spatial patterns of variables (Brunsdon et al., 1998). Bitter et al., (2007) compared the spatial expansion method and the GWR to examine spatial heterogeneity in housing attribute prices. The spatial expansion method is a global method, which means that parameters vary over space, whereas the GWR is a local model and parameters change at every observation point. They found that the GWR provides more explanatory power and estimation accuracy than the spatial expansion method.

The artificial intelligence-based (AI) methodologies such as expert systems, fuzzy logic, artificial neural networks and genetic algorithms have increasingly been used in housing valuation field. These advanced methodologies are known as knowledge-based methods.

Larraz (2011) tried to develop an expert system to valuate residential properties automatically and online for Spain. This expert system produces an online report called the residential properties valuation report, which can be used to evaluate each of the residential properties in Spain. The basic components of the system are property characteristics, environment, the neighborhood property values and distances to the focal property. Kriging methods, which take into account the spatial dependence among housing prices, were used to valuate these properties. The main drawback of the proposed system is that the effect of time is ignored in housing prices. Therefore, it is difficult to track of price changes for the same housing. The most important variables used for this expert system are area, age, the number of rooms, the number of bathrooms, floor, available elevator, heating, basement, swimming pool and garage.

In order to specify the determinants of house prices in Turkey, Selim (2009) compared the hedonic regression and artificial neural network models in respect of the prediction performance. The 2004 Household Budget Survey Data was used in this study for the analyses. This data was collected through questionnaires and the quality of the data depends on the accuracy of the answers given by the respondents. According to the comparison results, he claims that ANN can be used to predict the house prices in Turkey as a better alternative for hedonic models. However, it is very difficult to provide the sustainability of the data based on questionnaires for online valuation systems. Din et al., (2001) compared the linear and non-linear models

(ANN) using eight environmental and four structural variables. The comparison results show that the ANN and linear model price indices have similar shapes. The other result is that the linear models do not obtain the effect of environmental factors very precisely and the ANN model is not sufficiently robust. The ANN was also compared with the SAR model by Mimis et al., (2013). The result of their study supports the superiority of the ANN in housing valuation.

Besides the ANN models in recent years, fuzzy logic-based models have been used as alternative tools to estimate housing values.

Kuşan et al., (2010) used fuzzy logic models to predict house selling prices using a small dataset (160). The data was obtained through questionnaires. According to their testing results in the model, the predictions of the model are very close to real price values. Also, using a small dataset consisting of 120 housing values, a comparison study with fuzzy logic and hedonic approach (based on multiple regression) was carried out by Yalpır and Özkan (2011). The results of this study show that the fuzzy-based model predicted the market prices with 87% accuracy and the MRA-based model with 83% accuracy.

However, it is not possible to say that the ANN was proven to be a superior housing estimation model because this analysis based on the ANN was performed with limited housing transaction data set (120). It is necessary that these results be checked with larger data sets for a further research.

The ANN, fuzzy logic and multiple regression-based models were compared by Lokshina et al., (2003). According the results of their study, the ANN and fuzzy logic can be used to estimate the real estate price. Besides, the performance of the multiple regression application to predict house prices is quite well. The applicability of fuzzy clustering methods in housing market segmentation was studied by Liu et al., (2006). They integrated some features of the fuzzy logic and the ANN theories under the fuzzy neural network (FNN) to compensate the weaknesses of one theory with advantages of the other. They assert that the FNN prediction model based on hedonic price theory is highly convenient for the estimation of real estate values and decision-making jobs.

The study conducted by Türel (1981) is particularly remarkable because its topic is similar to that of this dissertation and in both studies the case study was about the same city (Ankara). He investigated the spatial differentiation of housing prices in Ankara. In this study, the functioning of the housing market, changes in prices over time and the causes of spatial differentiation of price in Ankara were examined, while ignoring submarkets, which are a form of specialized subdivision of a market. The results indicate the existence of a spatial variation of prices for Ankara and this supports the housing market segmentation hypothesis, which says the housing price structure is different in each segment. The size of dwelling units, central heating, hot water and lift were used as three important structural attributes of housing and represented by dummy variables. The location of a dwelling unit is defined by two variables; distance to the CBD and distance to the workplace of the household head. One of the important findings of this study is that social agglomeration used as the main neighborhood factor has effects on the price of housing. The effect of the edge of building was observed to be in accord with the results of most of the previous studies in that they are insignificant but negative. Another finding is that housing prices will increase rapidly yet in a non-predetermined manner. At the end of the study, the author concludes that there is a significant difference between the south and the north sides of the study area.

The findings of this study will provide important contributions to this dissertation in terms of the research topic and the target study area (city of Ankara, Turkey). However, this dissertation takes only the distance to the CBD variable into consideration because the data related to the household head for this case study area was not available. The other difference is that in the former study the distance to the CBD variable was determined according to two central business districts, namely Ulus and Kızılay. In those years when the mentioned study was performed, Ulus was the historical business and administrative center and Kızılay was the new business center in Ankara.

There are two major shortcomings of the previous investigation. First, it does not take spatial effects (autocorrelation and heterogeneity) into consideration. Moreover, it does not take the advantage of GIS. Over the last 35 years, Kızılay has gained a bigger importance and it has almost become the leading center of Ankara in terms of business and administration. For this reason, distance measurements will be derived between sample houses and Kızılay. Unfortunately, temporal data for accessibility could not be used because of the lack of adequate data. On the other hand, spatial variation of attribute values (spatial patterns) across space and spatial dependencies (autocorrelation) between variables can be detected using spatial techniques (the SAR and the GWR) differently from the previous study. To enable the identification of house positions on the map based on the longitude and latitude of houses, GIS technologies will be used in this thesis. GIS provides an efficient tool to measure both the linear distance (Euclidean) and the network route distance to derive proximity measurements to focal points such as the CBD. The spatial non-stationarity in the case study areas will be mapped using a GIS to reveal spatial patterns.

Recently some empirical studies performed on housing valuation have focused to investigate the influence of only a few or single characteristics on housing value. Although these studies have been carried out mostly by developed countries, there are limited-size studies on this subject in developing countries including Turkey.

Yusuf and Resosudarmo (2009) investigated the effects of air pollution on property value. The results indicate that urban air pollutants have a negative association with property value as expected. The authors state that there are a few studies relating to the impact of air pollution on house prices in developing countries. The main reason of this can be the lack of available data on air pollution. In Turkey, air quality measurements have been done locally but these data cannot be kept in a common database environment. Lewis et al., (2008) investigated the influence of being close to or distant from a dam site on property values using a semi-logarithmic functional form of hedonic property value methods. The findings of this study suggested that the value of a property closer to a dam decreases and the removal of a dam increases the values of nearby properties. Vichiensan et al., (2011) found that urban railway has a great influence on the area around stations. Bin (2011) showed that proximity to shoreline has a strong positive effect on property values. It was investigated by Kestens et al., (2006) whether location and property choice vary depending on the household profile. They specified the household profile according to the household type, age, educational attainment, income, and the previous tenure status of the buyers. They found that some characteristics of the buyer’s household have a direct influence on property prices, such as income and age. They also concluded that the educational attainment of the buyer and the household income are significant determinants of house prices. In other words, higher income and highly-educated households may be willing to pay more for housing to maintain neighborhood homogeneity. This finding partially confirms the hypothesis of Goodman and Thibodeau (2003) that only higher income households may prefer to pay more for housing to maintain neighborhood homogeneity. Case and Mayer (1996) reported that a household with school-age children would probably be willing to pay more for housing in a city with successful schools than households with no children. A household with no children would not be willing to pay for such schools. The influence of airports and airport light paths on housing prices was examined by Rahmatian and Cockerill (2004) using three functional forms. The results show that the semi-log model has the highest R^2 among the other functional forms and house prices increase when distance to airport increases.

The General Directorate of Land Registry and Cadastre (TKGM) started a project called the Land Registry and Cadastre Modernization Project funded by the World Bank in 2008. The project consists of four components and one of these components is property valuation (for taxation using mass appraisal approach). The aim of this component is to investigate and develop the policy and institutional options for the property valuation function in Turkey in line with the best international practices. As the background and project rationale, it is expressed that property valuation for taxation is less developed in Turkey than in similar economies. A survey study made in Europe confirms that a large majority (84%) of respondent countries either have (72%) or were developing (14%) mass valuation systems for taxation purposes. It has three sub-components: policy development (proposals on legal, institutional and technical arrangements); pilot implementation; and capacity building.

Separate committees dealing with the subject (parameters, legal, administrative) were established by TKGM. Reports were produced on the work of each committee. Recently, TKGM has announced its willingness to construct a Property Valuation Database integrated into the TAKBIS system (the computerized Land Registry Software for Turkey).

However, there is no property valuation law in Turkey that would assign institutional responsibilities or provide the framework for property valuation guidelines and standards. Moreover, the main focus of property valuation is related to taxation. The Ministry of Finance is responsible for determining taxation ratio and collecting it in Turkey. Also, land registry transactions in Turkey are based on seller and buyer declaration, so they may reflect the real market. To construct an online property valuation database, firstly a responsible institution should be determine and an inter-institutional agreement should be provided.

2.2 Evaluation of the Literature Review

Empirical research has primarily focused on identifying house characteristics that influence house values/prices the most using different methods. The parameters affecting housing value in the literature are based on various combinations of structural, environmental or neighborhood parameters, and different results are observed as to the size and signs of parameters in line with the location where the study has been carried out and the data used. Generally, structural characteristics such as the presence of a lift, a service room, a car parking, a terrace, a balcony, a basement, a garden, a pool, car park, private security, central heating and so on are used as dummy variables.

It is seen in studies conducted in recent years on housing valuation that the focal point is different for developing countries and those countries whose transition to private property and free market economy is relatively new. Within this context, studies are carried out by considering numerous parameters, such as area, parcel size, age, income, distance to certain points, the number of rooms, lift, private security, school, hospital, work place, pool, car park, in certain combinations.

Since valuation systems have not been established well in developing countries like China, Russia, Iran, Malaysia and Turkey, the studies on parameters affecting values and the most appropriate valuation models have not been concluded yet.

Developed countries have brought the studies to determine the key parameters that affect house values to a significant level in recent years. Therefore, it is observed that they make researches in order to investigate the influence of more specific parameters such as air pollution, traffic congestion, landscape, water bodies, noise impact, presence of open spaces, the amenity of urban green spaces, waste, train and fresher air on house value. The topics that are studied have the characteristic of an indication about the development level of the country.

In studies as to the models, on the other hand, it seems harder to make a clear-cut distinction. Studies are conducted in both the developing and developed countries using certain methods. In other words, it is seen that both traditional and advanced valuation methods are used in the developing and developed countries for housing valuation. When considered from this point of view, the most significant difference emerges from the way the data is obtained and its quality. While data as to housing can be obtained from online databases in developed countries, it is mostly based on sources like questionnaires, data collection from the field, and inquiries from real estate agents in developing countries.

Since the process of valuation takes into consideration the open market price determined according to particular features of a dwelling in a certain period of time, it needs to be performed dynamically. For this purpose, it is important that databases are formed about the structural, environmental, economic and locational characteristics of a dwelling and a dynamic valuation system based on these databases is devised.

CHAPTER 3

METHODOLOGY

3.1 General Framework of the Methodology

House valuation models can be grouped from different perspectives: traditional and advanced, global and local, spatial and non-spatial, knowledge-driven, data-driven and rule-based and tree-based (hierarchical). In spatial sense, global modelling tries to model the spatial relationships (spatial autocorrelation or spatial dependence) among the data whereas local modelling focus to model varying relationships spatially among the data (spatial heterogeneity or spatial non-stationary). In this section, a brief description of the methodology is given. The flowchart of the followed methodology is described in Figure 2.

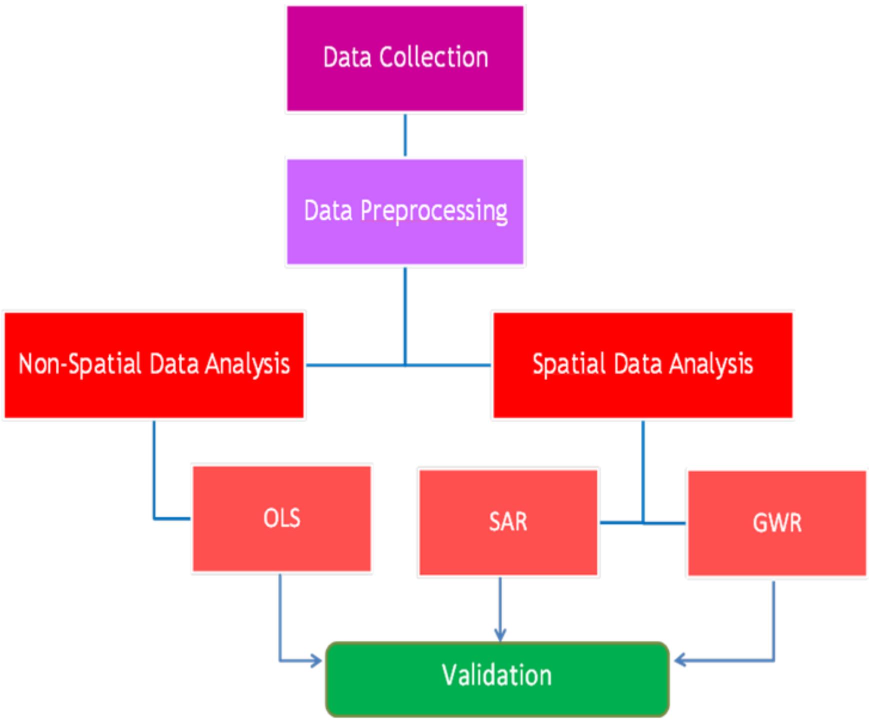


Figure 2: Flowchart of methodology for two case studies

The methodology of this study is composed of four sequential steps including data collection, data preprocessing, non-spatial and spatial analysis and validation. The first step of this work is data collection. The house values, structural and locational characteristics of houses, addresses of houses, socio-economic and demographic data were collected. Also, the urban maps and orthophoto images for Keçiören and Çankaya were obtained.

In the second step, the house values, structural, locational characteristics and addresses of houses and socio-economic and demographic were compiled and geocoded. The compiled and geocoded data were observed through visualization techniques.

In the third step, the dataset cover three years of data (2010-2012), so all 2010 and 2011 housing prices were deflated with Public Fixed Capital Investments and Foreign Currency Deflators published by Ministry of Development, Turkey[1] .

All housing prices in the dataset were transformed 2012 prices in Turkish Lira. Transformation factor is calculated as follows considering Table 3:

Table 3[1]: Public Fixed Capital Investments and Foreign Currency Deflators (2015=1,0000000)

Year	Housing Sector	Year	Housing Sector	Year	Housing sector
1963	0,000000628	1982	0,000028228	2001	0,223139714
1964	0,000000657	1983	0,000036696	2002	0,313146907
1965	0,000000702	1984	0,000054894	2003	0,381854380
1966	0,000000736	1985	0,000079631	2004	0,437123860
1967	0,000000815	1986	0,000110881	2005	0,471362771
1968	0,000000848	1987	0,000161581	2006	0,553052113
1969	0,000000896	1988	0,000315932	2007	0,591711270
1970	0,000000946	1989	0,000462768	2008	0,678656493
1971	0,000001090	1990	0,000715844	2009	0,635638476
1972	0,000001252	1991	0,001234004	***2010***	***0,673626718***
1973	0,000001403	1992	0,001930105	***2011***	***0,772230640***
1974	0,000001745	1993	0,003234909	***2012***	***0,810163279***
1975	0,000002064	1994	0,006890517	2013	0,854288485
1976	0,000002517	1995	0,011401661	2014	0,943396226
1977	0,000003508	1996	0,020235492	**2015**	**1,000000000**
1978	0,000005061	1997	0,038650032	2016	1,053000000
1979	0,000008279	1998	0,064755319	2017	1,105650000
1980	0,000017441	1999	0,097181272		
1981	0,000022400	2000	0,138681994		

The deflator factor for 2010-2012 (multiplication factor for transformed price from 2010 to 2012) = 0,810163279/0,673626718

[1] http://www2.kalkinma.gov.tr/kamuyat/2015/rehber/2015-2017-genelge-rehber.pdf

The deflator factor for 2011-2012 (multiplication factor for transformed price from 2011 to 2012) = 0,810163279/0,772230640

The deflator factor for 2012-2015 (multiplication factor for transformed price from 2012 to 2015) = 1,000000000/0,810163279

Hedonic pricing model based on OLS analysis was performed and then spatial regression analyses were done. The spatial dependence model, also known as the spatial lag (SARlag) model, and the spatial error model (SARerr) were carried out to construct global spatial models. Finally, GWR was applied to build local spatial model.

In the fourth and final step, performance of each model was tested using validation data sets. The mean absolute error (MAE), the mean absolute percentage error (MAPE), the mean-squared error (MSE) and root mean square error (RMSE) were used to measure the performances of the models. On the other hand, the performances of the models were also tested with the sales information obtained from the most popular internet real estate sites and some real estate agencies in Çankaya and Keçiören. The predictions of the OLS, SAR and GWR models were deflated from 2012 to 2015 considering Table 3.

The impact of variables selected was empirically examined on housing values using a hedonic-pricing model across two districts in Ankara, Turkey. The study included a combination of analytical and spatial models and techniques including GIS functions to measure the distance of houses to the selected site.

3.2 Data Collection

The process of obtaining the data was difficult, time consuming and tedious. The most problematic side of empirical based studies is mostly data collection. Although tedious and time consuming, gathering all necessary data and maintaining data reliability are essential. The most important step in this study was to obtain needed data from relevant organizations. In Turkey, a major source of data for a variety of indicators is Turkish Statistical Institute (TurkStat). Data on census, natural gas usage, water usage, the number of schools, the number of teachers and students for Çankaya and Keçiören districts were obtained from TurkStat. House values were used as dependent variables in the regression. House prices and structural data for these houses were supplied from Turkish Association of Valuation Experts (TAVE). TAVE is responsible for creating professional rules and valuation standards for real estate appraisal companies in Turkey. According to existing Capital Markets Board of Turkey legislations, the real estate appraisal companies without the consent of the customer can not disclose valuation reports. Therefore, the original data needed for this study was obtained from TAVE officially. In this context, 609 observations for Çankaya and 656 observations for Keçiören district of Ankara were obtained. The data belonging to 1265 flats have been collected by TAVE from real estate appraisal companies in Ankara. The data covers the three year period between 2010 and 2012. The attributes of data were listed in Table 4.

Table 4: Variables used by real estate appraisal companies in Turkey

Variable	**Variable**
Province	Quality of housing construction
District	Building house style
District /village	Security
Cadastral map no	Car parking
Island number	Pool
Lot no	Lift
Block no	Heating system
Floor no	# of saloons
Detached house	# of rooms
Type of title deed	# of kitchens
Properties of house	# of bathrooms
Street	# of balconies
Site / apartment	Lot area
Age	Value based on current use area
# Of floors in the building	Date of valuation report

The urban maps for case studies were obtained from Çankaya and Keçiören municipalities. Data set for Keçiören comprises household attribute apart from above-mentioned features. Household data is not available for Çankaya Municipality. Since locations of the buildings in Turkey are not yet geocoded, buildings needed to be geocoded. Geocoding is about adding x, y coordinates to point locations represented by these pieces of information (Paterson and Boyle, 2002). The data being geocoded must include information about their locations. City name, district, neighborhood, island and parcel, building number, house number, street name, street number and street direction and postal code are widely used in order to geocoding the data in housing valuation literature (Clapp, 2003; Pavlov, 2000; Basu and Thibodeau, 1998).

The data for theft events in the districts of Çankaya and Keçiören were collected from some insurance and security companies in Ankara. The Point of Interest (POI) data (schools, hospitals, shopping malls, subway, main transportation roads and bus stops in Ankara) were obtained from Ankara Metropolitan Municipality.

As seen in Table 4, there are many locational, environmental, structural, demographic, and socio-economic factors to take into account in housing valuation. This makes it difficult to analyze and construct a reliable valuation model. Location is the most indispensable factor to take into consideration when constructing a housing valuation model (Kiel and Zabel, 2008, Pagourtzi et al., 2006).

3.3 Data Preprocessing

Housing valuation reports provided by TAVE were not in digital form, so they were entered manually in an Excel spreadsheet. On the other hand, locations of the buildings in Turkey are not yet geocoded. Therefore, they must be geocoded on urban maps manually using a GIS tool. The data for theft events were geocoded on Çankaya and Keçiören digital urban maps. Distances to the POI and Kızılay (most people living in Ankara accepted that Kızılay is still the area where the heart of Ankara beats), Şelale and Etlik (very important two centers of Keçiören district) were derived for each house. All distance measurements were carried out both in Euclidean and the shortest network route using ArcGIS. The number of thefts was determined using buffer radii of 500 and 1000 meters around each sample.

In the data sets, some variables were coded as dummy. A dummy variable or indicator variable is an artificial variable created to represent an attribute with two or more distinct categories (Gujarati, 1970). A dummy variable, in other words, is a numerical representation of the categories of a nominal or ordinal variable. If the categorical variable has n categories one uses $n - 1$ dummy variables (Suits, 1957).

Consequently, the data on house properties including their values were geocoded on fundamental base maps and plans belonging to Çankaya and Keçiören.

As can be seen in Figure 3, observations for Çankaya case study area are concentrated along north–south axis namely older settlement regions. There are a few observations through western part of the district which are relatively new residential areas such as Ümitköy, Mustafa Kemal, Mutlukent, Beytepe neighborhoods.

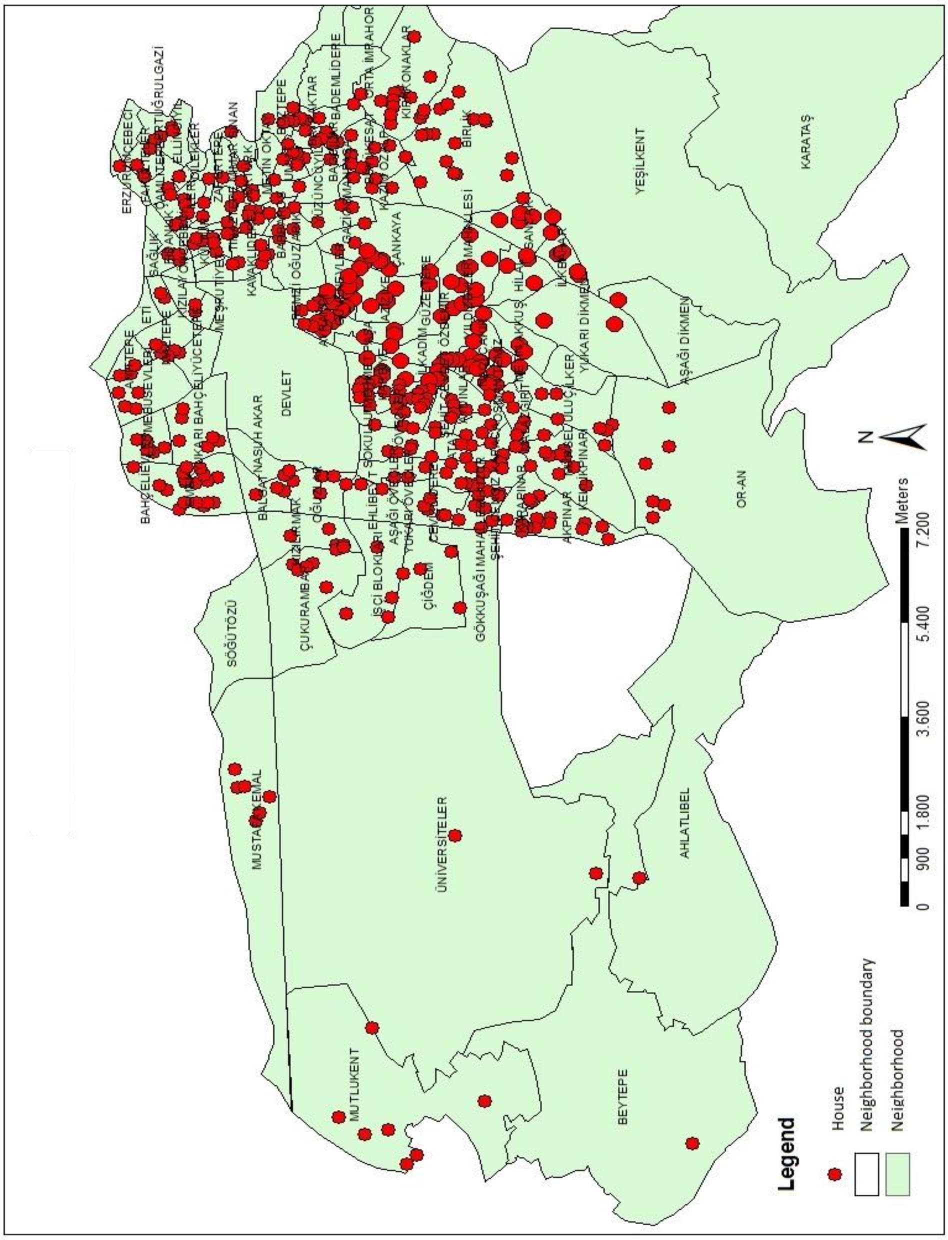


Figure 3: Study site: Distribution of houses for Çankaya

Observations for Keçiören district, in Figure 4, are concentrated along east-west except a few at the north, which are in Karşıyaka, Hisar and Karakaya neighborhoods.

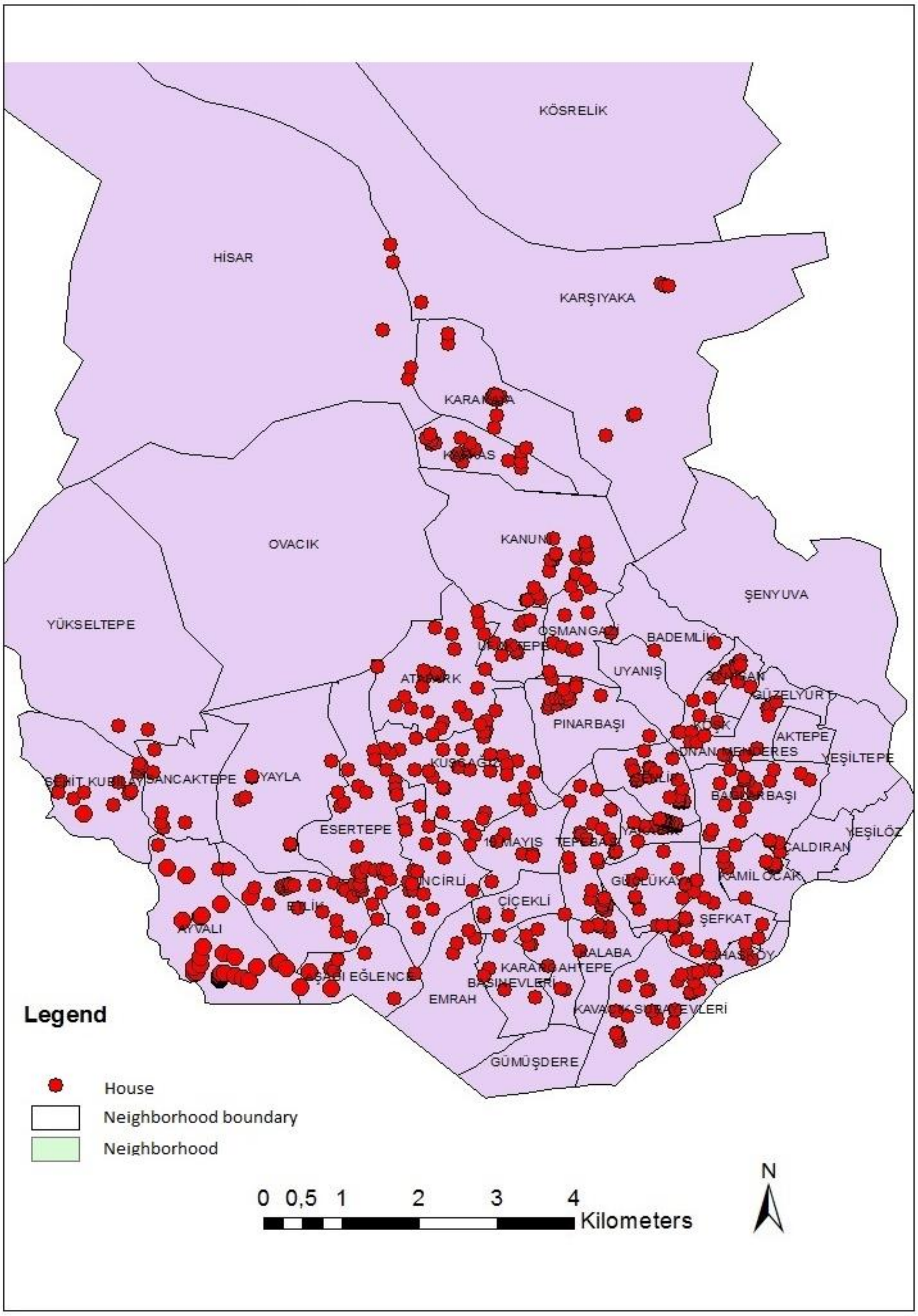


Figure 4: Study site: Distribution of houses for Keçiören

Thiessen Polygons

Tobler (1970) Law states that all things (houses in that case) are related to each other, space and spatial relations have been explored by researchers. Researchers have applied two major types of approaches to expose these relations: contiguity or distance. To build contiguity relationships among houses Thiessen polygon technique will be used in this study. The other names of the technique are Voronoi diagrams and Dirichlet tessellations.

Figure 5: Thiessen polygons created around the observation points for Çankaya

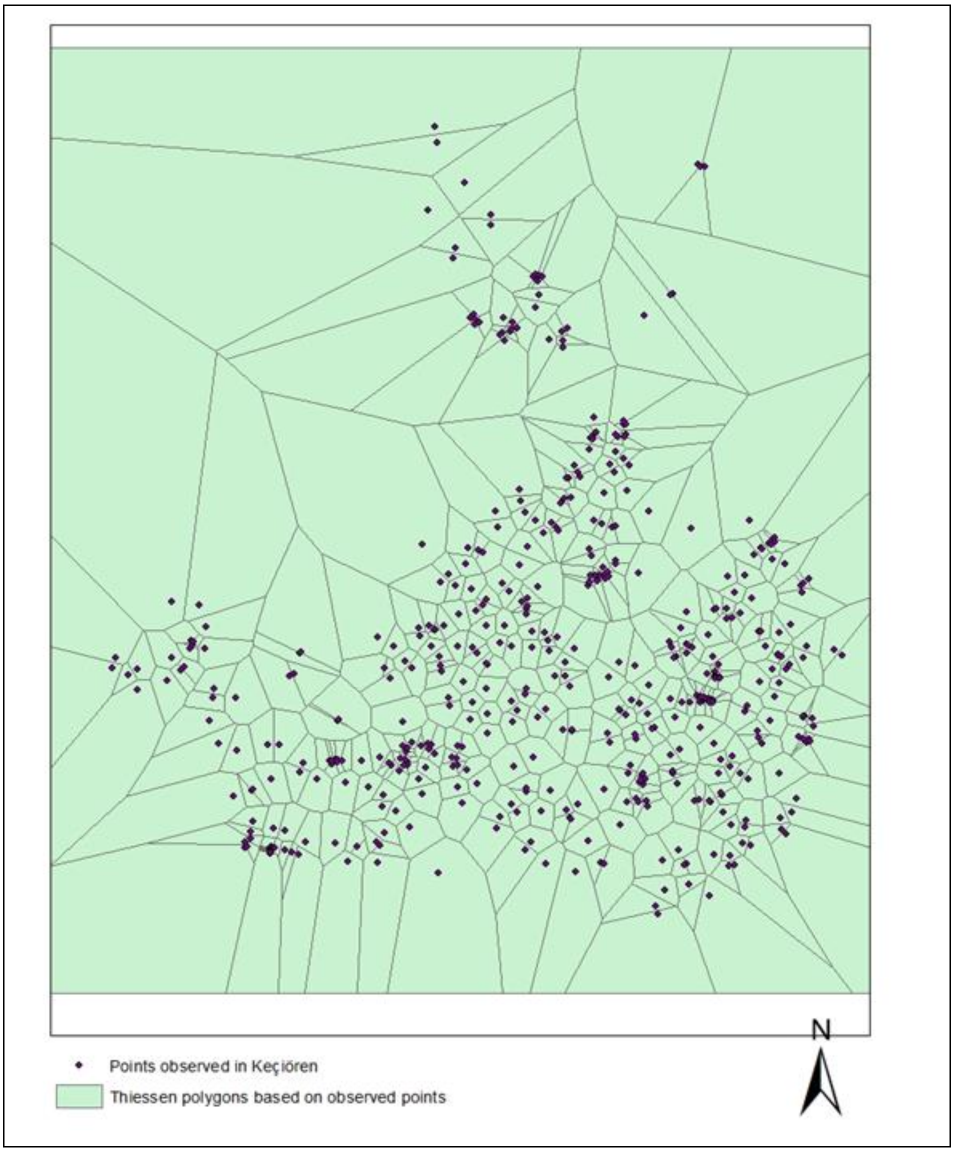


Figure 6: Thiessen polygons created around the observation points for Keçiören

Thiessen polygons are created as follows: All points are triangulated into a triangulated irregular network (TIN) that meets the Delaunay criterion. The perpendicular bisectors for each triangle edge are generated, forming the edges of the Thiessen polygons. The locations at which the bisectors intersect determine the locations of the Thiessen polygon vertices.

Thiessen polygons are generated around each point representing observed values. Figure 5 and Figure 6 show the points which were transformed to polygons for Çankaya and Keçiören respectively. This provides the advantages of coverage of all analyzed points and the establishment of neighborhood relations based on contiguity (Kryvobokov, 2013).

3.4 Proposed Model

First, OLS regression was applied to the data to obtain global coefficients without any respect to spatial dependency and to compare the results of spatial and non-spatial models. In spatial analysis, the spatial autocorrelation of the house values was primarily inspected. Spatial autocorrelation measures the degree to which near and distant things are related. SAR and GWR regression techniques take spatial dependency and heterogeneity into account. While SAR gives global coefficients similar to OLS, GWR provides local coefficients. Finally, the estimated models for housing valuation based on OLS, SAR and GWR were constructed and these estimation models were tested with different data sets for validation purposes.

3.5 Model Evaluation Criteria

Traditionally, statistical testing criteria such as R^2, Maximum Likelihood value, Akaike's Information Criterion are used to evaluate regression models. Meanwhile, the null hypothesis, that the contribution of a relationship is zero, is investigated based on t-test or F-test. In this study, R^2, the Akaike information criterion (AIC) (Akaike, 1974), Corrected Akaike's Information Criterion (AICc) (Hurvich and Tsai,1993), the log likelihood, and Schwarz Information Criterion (SIC) which is also known as Bayesian Information Criterion (BIC) (Koehler, 1988) are used to measure the suitability (Gayawan and Ipinyomi, 2009) of the models. The following criteria for selection of the best model are used widely in literature (Beal, 2007; Burnham and Anderson, 2004):

- max's the R^2/adjusted R^2
- max's the log likelihood
- min's the AIC
- min's the SIC

R^2 cannot be used alone to determine the goodness of fit of the model, since it does not demonstrate whether the predicted regression coefficients are statistically different from zero. It is also not convenient as a measure of fit in comparing spatial models.

To overcome this problem AIC/AICc, BIC/SIC and the log likelihood model selection criteria are considered together. According to Anselin and Getis (2010); the best model in the group compared is the one that maximizes the log likelihood and minimizes AIC and SIC scores. The model with the highest log-likelihood has the best fit. In addition to this, the lower the AIC and the SIC values, the better the model. Overfitting occurs when the R^2 and the log-likelihood increases with additional variables. This over-fitting can be corrected by employing the AIC or the SIC. The AIC and SIC are more commonly used than the adjusted R^2. Each of the two has certain (but different) theoretical properties that make them appealing (Hough et al., 2010).

3.5.1. Akaike information criterion

The Akaike information criterion is a measure of the relative quality of a statistical model for a given set of data. For any statistical model, the AIC value is expressed as follows (Akaike, 1987; Bozdogan, 2000; Amin et al., 2012):

$$AIC = 2k - 2\ln(L) \qquad [3]$$

where n is the sample size, k is the number of parameters used in the model, and L is the maximized value of the likelihood function for the model. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value.

3.5.2. Corrected Akaike's Information Criterion (AICc)

AICc is AIC with a correction for finite sample sizes:

$$AICc = AIC + 2k(k + 1)/(n - k - 1) \qquad [4]$$

where n denotes the sample size and k is the number of parameters used in the model. Thus, AICc is another version of AIC with a greater penalty for extra parameters. In academic literature, if k is large it is suggested to use AICc rather than AIC (Cavanaugh, 1997).

3.5.3. Maximum Likelihood

The Maximum Likelihood Estimation (MLE) is a method of predicting the parameters of a model. This estimation method is one of the most widely used in literature. The method of MLE selects the values of the model parameters that maximizes the likelihood function (Hurlin, 2013). Therefore, estimation method is affected as little as possible by sampling error.

The log likelihood fuction is defined as follow (Pace et al., 1998):

$$\ell_N (\theta; x_1, .., x_n) = \sum_{i=1}^{N} \ln f_X (x_i; \theta) \qquad [5]$$

where

N=number of variables

$x_1, \ldots\ldots x_N =$ random variables

θ = unknown parameter

It is generally accepted that the best model is the one that has higher log likelihood value whereas lower AIC and SIC values.

3.5.4. Schwarz Information Criterion (SIC)

Schwarz Information Criterion (SIC), also known as Bayesian Information Criterion (BIC) is an alternative widely used criterion to AIC. The Bayesian Information Criterion (BIC) which essentially replaces the term 2k in the AIC with the expression k+klnN (Chatfield, 2013). It is based, in part, on the likelihood function and it is closely related to the AIC (Beal, 2007).

$$BIC=-2\ln L + k\ln(n) \qquad [6]$$

where,

x = the observed data;

n = the sample size;

k = the number of free parameters to be estimated. If the model under consideration is a linear regression, k is the number of regressors, including the intercept;

L = the maximized value of the likelihood function of the model.

As with the AIC, minimizing the BIC is intended to give the best model.

3.6 Non-Spatial Data Analysis

Non-spatial (OLS regression) is the most widely used method for fitting linear statistical models. The OLS is more commonly named linear regression which is applied simple (the simple OLS regression) or multiple (the multiple OLS regression) depending on the number of explanatory variables (Craven and Islam, 2011). The OLS approach to multiple linear regressions which was introduced by Gauss is the simplest type of prediction in statistics (Weisberg, 2005). The OLS model is estimated where the resulting coefficients are global, i.e., the coefficients are constant over the study area.

In this study, OLS regression method is used as the first method. OLS regression minimizes the sum of the squared residuals. In general, a model fits the data well if the differences between the observed values and predicted values of the model are small and unbiased.

Regression analysis of the hedonic price models is used to understand how the typical value of the dependent variable changes when any one of the independent variables is changed, while the other independent variables are remained fixed (Bin, 2004).

In this study, linear multiple regression analysis (an extended type of OLS) is to used since there are many variables to be used to construct the housing valuation model. In other words, the dependent and independent variables are regressed using properties of known prices to determine the established relationships (coefficients) between the two types of variables (Adair and McGreal, 1996). Then a housing valuation OLS model is constructed according to the determined coefficients. Generally, significance level is denoted by α in statistics and 0.1, 0.05, and 0.01 are common significance levels. To take advantage of an OLS analysis, a number of assumptions listed below must be satisfied.

Linearity: The assumption is the relationship between the predictors and the dependent variable is linear. In this test, the linearity assumption is checked by examining correlations between continuous variables and scatter diagrams of the dependent variable versus independent variables.

Normality: Another assumption of linear regression is that the residuals are normally distributed. The null hypothesis is that the data is normally distributed and the alternative hypothesis is that the data is not normally distributed. Because the sample size is sufficiently large (N>50), the normality assumption is accepted by the central limit theorem. In probability theory, the central limit theorem (CLT) states conditions under which the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed (Rice, 2006).

Multicollinearity: The weak correlations among the independent variables are desirable. Multicollinearity increases the standard errors of the coefficients. The use of the variance inflation factor (VIF) is the most reliable way to examine multicollinearity. The VIF is the reciprocal of the Tolerance stated by O'brien, (2007) as follows:

$$VIF=1/\ (1-R^2) \qquad [7]$$

where R^2 is correlation coefficient.

As a rule of thumb, if any of the VIF is greater than 10 (a lower limit of 5 is deemed to be very conservative) there is a multicollinearity problem. If there are two or more variables with VIF values around or greater than 10, this shows evidence of serious multicollinearity. To solve this problem one of these variables must be removed from the regression model. However, a tolerance of less than 0.20 or 0.10 and/or a VIF 10 and above indicates a multicollinearity problem (Lin and Wen, 2011).

Among the explanatory variables compiled as given in the previous step, an elimination to be carried out due to multicollinearity or in other words to satisfy the independency of the variables. Multicollinearity exists when one of the explanatory variables has a linear relationship with another explanatory variable or with the combination of other explanatory variables. If this linear relationship is perfect (i.e.

the linear relationship of two explanatory variables have a coefficient of determination equal to 1), it is called perfect or extreme multicollinearity. Although perfect multicollinearity is a rare case, there is a risk of artificially obtaining it when the data set is very small. In case of perfect multicollinearity among explanatory variables, the regression analyses cannot be performed.

Multicollinearity increases standard errors and so uncertainty of the coefficient estimates in the regression, resulting in lower significance of coefficient estimates for explanatory variables and larger confidence intervals. This leads to insignificant coefficient estimates of explanatory variables although the overall equation is significant. Since it is impossible to differentiate between the effects of explanatory variables when they covary (Miles and Shevlin, 2001), multicollinearity makes it hard to interpret the results of the analyses and it should be avoided as much as possible.

VIF and Pearson's product moment correlation coefficient (r) were used to detect and eliminate multicollinearity. r is used for detecting bivariate association while VIF enables us to analyze multivariate correlations. In other words, the analysis based on r uncovers multicollinearity caused by correlation of only two variables. Nevertheless, a variable may have correlation with not only another variable but also combination of more than one variable which also accounts for multicollinearity as it can be deduced from the definition and this is overcome by inspecting VIFs.

After multicollinearity analysis, a shapefile is constructed to perform spatial analyses. Shapefile is a data format developed by ESRI in which the features are composed of points, lines or polygons and any information can be attached to these features as an attribute (ESRI, 1998). The shapefile includes polygons representing the areal units of interest and the data belonging to each areal unit composed of dependent and independent variables and also coordinates of the centers of these units. Any modification on data can be carried out through a database file linked to the shapefile and viewed as a spreadsheet. The shapefile can be imported into R for subsequent data analysis.

Constant variance: Homoscedasticity (constant variance) is considered to be the most important assumption that must be met in linear regression. Homoscedasticity means that the variance of errors is the same across all levels of the independent variables. The points should be equally distributed around the mean. If the variance of errors differs among independent variables there is heteroscedasticity. In other words, the variance of the error term is constant this is called as homoscedasticity whereas the error terms do not have constant variance, this is called as heteroscedastic.

It can lead to serious distortion of findings and seriously weaken the analysis thus increasing the possibility of a Type I error, which is rejecting the null hypothesis although it is true.

3.7 Spatial Regression

The value of a property in one location in a hedonic price analysis may be affected by property values in other locations. Ignoring this spatial effect or spatial dependence may cause the simple OLS estimation to be either inconsistent or inefficient. Spatial regression analysis is used to explore spatial relationships in a dataset and includes new variables to increase the power of explanation of the model. This analysis provides to see which factors are more important to explain the spatial variation and patterns in dataset. However, there are two important points that should be considered in spatial regression namely the spatial lag model and the spatial error model. In the former, a house value depends on both its characteristics and on its neighboring house values. The spatial lag model is an appropriate tool to measure neighborhood spillover effects. It assumes that the spatially weighted sum of neighborhood housing prices (the spatial lag) enters as an explanatory variable in the specification of housing price formation. These are spatial dependency (also known spatial autocorrelation) and spatial heterogeneity (also known spatial non-stationary). The null hypothesis of autocorrelation is that values observed at one location do not depend on values observed at neighboring locations. As a result, the estimation of the variable at one observation location is affected by the value of the variable at the nearby locations. However, defining a proper weight matrix is a crucial issue in statistical regression analysis. In standard statistical tests, the presence of spatial autocorrelation is misleading or can lead to inaccurate estimates of test performance. In order to overcome this problem, the SAR was used in order to take account the spatial dependence of the dependent variable (house values). On the other hand, GWR was used to capture the spatially varying impacts of some independent variables (on per house) as a local analysis. If the relationships among regression variables do not change through space, the global spatial regression is convenient to model the data; otherwise, the GWR is an appropriate technique to modelling the data (Matthews et al., 2012).

3.7.1. Creating Spatial Weight Matrices

Before running the multiple regressions to represent a spatial structure, it is needed to create some weight matrices (Chi and Zhu, 2008). The spatial weight matrix is the basic tool used to model the spatial relationships among features in a dataset. Therefore, it is necessary especially to create weight matrices for analyses such as spatial lag, spatial error and GWR.

The spatial weight matrix, W_{ij}, shows whether any pair of observations are neighbours. *i= {1,..,n}* and *j= {1,.. ,n}, n* denotes the number of observations, W_{ij} reflects the spatial influence of unit *j* on unit *i*.

Although there are a number of ways to define spatial weight matrices, the most widely used in practice are boundary based, distance based and kernel based. Spatial contiguity weights indicate whether spatial units share a boundary or not. If *i* and *j* units share a boundary W_{ij} =1 otherwise W_{ij} =0.

d_{ij} is a distance between each pair of spatial units i and j. If distance itself is an important criterion of spatial influence, and if d denotes a threshold distance (or bandwidth) beyond which there is no direct spatial influence between spatial units, then the corresponding radial distance weight matrix, W, has spatial weights of the form.

If $0 <= d_{ij} < = d$ then $W_{ij} = 1$

$d_{ij} > d$ then $W_{ij} = 0$

The type of weight matrix is determined by the definition of the spatial neighbors considering data structures (raster or vector) and data distribution pattern. In this study, W matrices were created based on contiguity (Rook) and distance for both districts.

The spatial weight matrix is usually standardised, such that every row of the matrix is summed to 1. In a row-standardized matrix, the weights are arranged so that the elements in a row add up to unity. In the row-standardized contiguity matrix, it is expected that weights vary among the rows since number of neighbors may differ from one district to another. However, for distance based matrices, the weights are equal within each matrix since the number of neighbors is fixed for each spatial lag. Defining spatial weights or creating weighting matrix is very essential in spatial data analysis because it is how we can incorporate the spatiality into the models.

Sharing a boundary is the criterion for being neighbors based on contiguity. In this study, in order to create a contiguity relationship the type of point data was converted to Thiessen polygons. The contiguity weight matrix was row-standardized (Getis and Aldstadt, 2010) using GeoDa and R software. This means that the row elements for each observation sum to 1, with zero on the diagonal and some non-zero off-diagonal elements.

Although many approaches are available to define a spatial weight matrix, there is not any agreement on which one is the best in literature.

3.7.2. Spatial Autocorrelation

Spatial autocorrelation is a phenomenon where values of a variable show regular pattern over space. Spatial autocorrelation refers to a situation where the OLS residuals exhibit a regular pattern over space.Spatial autocorrelation is the spatial dependency of a variable over the study area. One can say there is spatial autocorrelation when the variable is spatially distributed according to a systematic pattern. Tobler's first law of geography implicitly refers to spatial autocorrelation stating that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

Autocorrelation can be characterized as positive, zero or negative. Positive autocorrelation exists when nearby observation locations exhibit similar variable values. On the contrary, close observational units may have dissimilar values. Then, this pattern is called negative autocorrelation. Zero spatial autocorrelation means that the nearby values are not related to each other or simply that there is no spatial autocorrelation (Griffith, 1987).

The dependency between houses that share the same economic, social, and environmental properties is natural. Moran's I and Geary's C spatial autocorrelation statistics are used to determine the degree of dependency among observations in the study area. Moran's I and Geary's C are well known global techniques to estimate the overall degree of spatial autocorrelation for a dataset. Moran coefficient (Moran's I) is a coefficient to quantify spatial autocorrelation. Moran's I lies between approximately -1 and 1 and takes value of zero when the variable is randomly distributed rather than having a spatial pattern. Spatial autocorrelation is positive when the coefficient has a positive sign and there is negative spatial autocorrelation if the coefficient is negative. The strength of the interdependency increases when Moran's I deviates from zero and gets closer to -1 and 1 for negative and positive autocorrelation, respectively (Oden, 1995).

A correlogram is useful to determine the scale at which spatial autocorrelation is generated. In a correlogram, Moran's I values are depicted versus the spatial lag. For a distance based neighborhood criterion, spatial lag may be the distance at equal intervals or the number of nearest neighbors taken into consideration (e.g. first nearest neighbors for spatial lag 1, second nearest neighbors for spatial lag 2, etc.).
Spatial correlation can also be demonstrated using a Moran scatter plot (Figure 7). This provides more detail visually about the type of spatial autocorrelation and spatial pattern (Fischer and Getis, 2009). The Moran scatter plot is introduced by four different quadrants regarding association between each attribute value of a space and its neighbors. Moran's Index is calculated for each region as follows:

$$I = \frac{n \sum_i \sum_{j \neq i} w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_i \sum_{j \neq i} w_{ij}) \sum_i (Y_i - \bar{Y})^2} \qquad [8]$$

where n is the number of observation locations and Y refers to the dependent variable with subscripts i and j denoting areal observation units (Gangodagamage et al., 2008). $\bar{Y}$ is the mean of the dependent variable. wij is the element of a weighting matrix W (nxn), which includes weights for each pair of observation locations. This weighting matrix is called spatial proximity matrix and makes it possible to convert proximity definitions (e.g. close, nearby, far, etc.) into mathematical terms so that it can be incorporated into the formulation. Other names that are used to designate the matrix are spatial connectivity matrix, spatial link matrix, geographic weights matrix, etc.

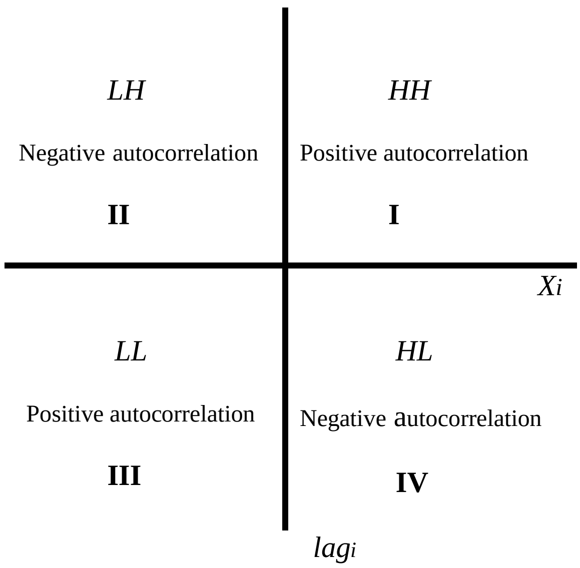


Figure 7: The Moran Scatter Plot

Quadrant High-High (H-H) displays the spaces with a high value of the variable surrounded by spaces with high values. The spaces have positive values namely the values are above the overall average value.

Quadrant Low-Low (L-L) displays the spaces with a low value surrounded by spaces with low values. The spaces have negative values namely the values are below the overall average value.

Quadrant Low-High (L-H) shows the spaces with low value (negative) surrounded by spaces with high values (positive).

Quadrant High-Low (H-L) shows the spaces with high value (positive) surrounded by spaces with low values (negative).

3.8 Spatial Autoregression (SAR)

SAR estimates the coefficients based on the fact that the dependent variable in an observation location is affected by the dependent variable of neighboring observations in addition to the effects of explanatory variables (Lichstein *et al.*, 2002). Two types of SAR models were investigated in this study: lag model (SAR_{lag}) and error model (SAR_{err}).

Simultaneous autoregressive coefficients (interaction parameters: ρ and λ) quantify the effect of neighboring observations and also they determine the direction of that effect (Düzgün and Kemec, 2008). These are additional parameters to be estimated compared to the non-spatial regression model which only estimates the regression coefficients (β). Therefore, SAR models should estimate not only β but also interaction parameters, which is a computationally intensive procedure.

The spatial lag model is equivalent to spatial simultaneous autoregressive lag model (Anselin, 1988). A spatial-lag hedonic price model can be written as follows:

$$\text{House Value} = \rho W + X1\beta1 + X2\beta2 + \varepsilon, \qquad [9]$$

where ρ is a spatial autocorrelation parameter, W is a $n \times n$ spatial weight matrix (where n is the number of observations), X1 is a matrix with observations on structural characteristics, X2 is a matrix with observations on location characteristics, with ε assumed to be a vector of independent and identically distributed error terms. Typically, the definition of neighbors used in the weights matrix is based on a notion of distance decay or contiguity.

In case of SARlag model, the autoregressive structure is encompassed only in the response variable due to its inherent properties. SARlag model can be written in the following form (Bailey and Gatrell, 1995):

$$Y = X\beta + \rho WY + \epsilon \qquad [10]$$

Similar to non-spatial regression notation, Y, β and ϵ are the vectors of dependent variable, regression coefficients and errors, respectively, while X is the matrix of independent variables. W is the spatial proximity matrix, which is detailed in the previous section. ρ is the simultaneous autoregressive (lag) coefficient. In addition to an ordinary regression, SARlag involves ‘ρWY’ term which indicates that the response variable in a location is affected by the value of response variable in the neighboring locations (Sparks and Sparks, 2009). If the weight matrix is row-standardized, this term averages the response variable in the neighbors.

Spatially lagged explanatory variables are introduced into the spatial lag model to obtain a model which is known as spatial Durbin (mixed) model. Using the same notation, mixed model is obtained as (Bivand et al., 2008):

$$Y = X\beta + \rho WY + WX\gamma + \epsilon \quad [11]$$

where γ is the coefficient for lagged explanatory variables. If this coefficient is constrained so that it is equal to the negative of the product of autoregressive coefficient and the regression coefficient (i.e. common factor constraint), SARerr model is attained (Anselin, 1999).

3.9 Spatial Error Model

In case when spatial dependence is present in the error term, a spatial autoregressive specification for this dependence is usually assumed. This is called spatial error model (SEM) and can be formulated as follows (Anselin, 2001):

$$P = X1\beta1 + X2\beta2 + \varepsilon, \quad [12]$$

$$\varepsilon = \lambda W\varepsilon + u, \quad [13]$$

where λ is the spatial autoregressive coefficient, W is the spatial weight matrix, and u is assumed to be a vector of identically distributed errors. This model is a special case of a regression specification with a non-spherical error variance-covariance matrix. Therefore, W now pertains to shocks in the unobserved variables (the errors) but not to the explanatory variables of the model (X). Consequently, the price at any location is a function of the local characteristics but also of the omitted variables at neighboring locations.

In the SARerr model, the autocorrelation is reflected by the correlated errors. This may be due to lacking an important explanatory variable so that the explanatory variables included are not adequate to explain the variation in the response variable.

The inherent autocorrelation structure of the response variable itself may also lead to correlated residuals (Kissling and Carl, 2008). SARerr is formulated as (Bailey and Gatrell, 1995):

$$Y=X\beta+U \quad [14]$$

$$U=\lambda WU+\epsilon \quad [15]$$

where λ is the simultaneous autoregressive (error) coefficient. By rearranging, SARerr model can be rewritten as:

$$Y=X\beta+\lambda WY-\lambda WX\beta+\epsilon \quad [16]$$

The first term ($X\beta$) introduces the general trend in the formulation. 'λWY' is the term for spatially lagged response variable and it incorporates the neighboring values of the response variable. The general trend in the neighboring locations is further included via the third term ($\lambda WX\beta$) in the formula (Bailey and Gatrell, 1995). As it is seen, SARerr can be obtained from the spatial Durbin model by putting common factor constraint on the coefficient of spatially lagged explanatory variables such that $\gamma=-\lambda\beta$.

3.10 Geographically Weighted Regression Model

GWR, which is increasingly used in geography and other disciplines, is one of several spatial regression techniques to explore the spatial relationships of variables locally. The main idea of GWR is to estimate parameters for every regression point by using observations in a given neighborhood. For this purpose a weight matrix must be created regarding Tobler observation namely everything is related to everything else, but near things are more related than distant things. This is known as the first law of geography.

GWR is an extension of OLS regression in which the parameters are allowed to vary spatially. Variations in relationships among parameters coefficients through space are referred to as spatial non-stationarity. It builds a local regression equation for each feature in the dataset. GWR constructs these separate equations by incorporating the dependent and explanatory variables of features falling within the bandwidth of each target feature. The shape and size of the bandwidth is dependent on user input for the kernel type, bandwidth method, distance, and number of features.

GWR is formulated similar to an ordinary regression; however, the *β* coefficients are site specific in this model. The GWR model is formed as:

$$y_i = \beta_o(u_i, v_i) + \sum_k \beta_k(u_i, v_i)\, x_{ik} + \epsilon_i \quad [17]$$

where *i* denotes the regression point where model is calibrated and (u_i,v_i) refers to the coordinates of point *i* (Fotheringham *et al.*, 2002).

Using a spatial kernel is important to make the geographic weighting in the model. The chosen kernel type and bandwidth methods can be change the results of GWR model (Lin and Wen, 2011). Predicted parameters in GWR depend on the weighting function of the kernel selected (Propastin and Kappas, 2006). The difficulty to select an appropriate is the major drawback of GWR.

The bandwidth is key coefficient in the kernel, which controls the size of the kernel. Bandwidths can be considered as smoothing functions of the local parameter estimations (Sharma et al., 2011). The Kernel function can be chosen as bi-square or Gaussian. The Kernel of Gaussian function distributes weights according to (Fotheringham et al., 2002):

$$w_{ij} = exp\left[-{}^{1}/_{2}\left(d_{ij}/b\right)^2\right] \quad [18]$$

where *j* is the data point, d_{ij} is the distance between regression point *i* and data point *j* and *b* is bandwidth. As *b* becomes larger the closer will be the model solution to that of OLS and when *b* is equal to the maximum distance between points in the system, the two models will be equal. In this stuation, GWR becomes equivalent to OLS. At the regression point, the weight is equal to 1 and it decreases as the distance increases.

When bi-square function is chosen, the weights are assigned according to (Fotheringham *et al.*, 2002):

$$w_{ij} = \begin{cases} \left[1 - \left(d_{ij}/b\right)^2\right]^2, & d_{ij} < b \\ 0 \quad , & otherwise \end{cases} \qquad [19]$$

According to this weighting function, the data points within bandwidth *b* are weighted through a near-Gaussian function. It should be noted that the weight given to a data point decreases as d_{ij} increases and data points beyond *b* are not included in the calibration at point *i* since they take zero weights.

GWR is a very powerful analytical tool and has the ability to reveal spatially varying patterns in the determinants of value in a hedonic model formulation (McCluskey et al., 2007). The β coefficients estimated by OLS and SAR are valid for both districts; consequently, they are global coefficients. GWR, on the contrary, estimates coefficients specific to each areal unit which are then called local coefficients. In other words, the relation of the dependent variable to the explanatory variables varies through the study area.

As a result of GWR analysis, local coefficients for each districts and associated standard errors are obtained. t-values, then, can be obtained by dividing each β coefficient estimate by its standard error. Coefficients and associated t-values should be observed via choropleth maps to explore the varying relationships between the dependent and independent variables. This allows seeing where GWR predicts well and where it predicts poorly.

Adaptive Gaussian

$$w_{ij} = \exp(-d_{ij}^2 / \theta_{i(k)}^2) \qquad [20]$$

where,

i is the regression point index,

j is data point (location index),

wij is the weight value of observation at location j for estimating the coefficient at location i,

dij is the Euclidean distance between i and j,

θ is a fixed bandwidth size defined by a distance metric measure,

θi(k) is an adaptive bandwidth size defined as the k th nearest neighbor distance.

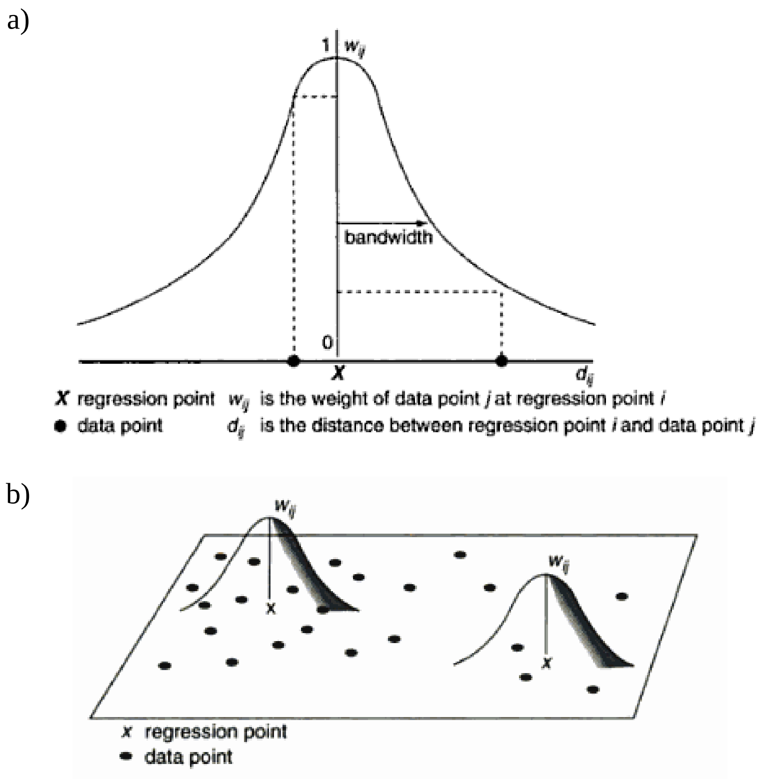


Figure 8: (a) A spatial kernel**;** (b) GWR with fixed spatial kernels (Fotheringham, *et al.,* 2002)

GWR reveals these local relationships by moving a spatial kernel across the study area (Charlton et. al., (2009). A representation of kernel is given in Figure 8. The center of the kernel is located on the regression points (*x*). At each regression point, local coefficients are estimated and the model is calibrated for that point according to a weighting scheme. The function of the kernel modifies the weights given to each data point according to its distance from the regression point. Higher weights are assigned to the data points closer to the regression point and the weight given decreases as moving away from the regression point. The data points to be used in the model calibration each time are determined by the bandwidth - the base radius - of the kernel (Fotheringham, *et al.,* 2002).

3.11 Validation

Validation is the process of assessing how well housing valuation models perform against real data. The purpose of validation is to test the signs and significance of the attributes and goodness of fit of the models. Before starting regression analysis data set was separated into a training set and testing set (validation set), most of the data was used for training, and a smaller portion of the data was used for testing. This technique is called cross-validation. Gao *et al.*, (2006) recommended to use the cross-validation technique and to test the prediction power of models using the information of observed data. The regression equation generates a straight line (regression line) showing the best approximation of the given set of data, i.e., to see how well the observed prices can be predicted with the test samples.

There are a number of error metrics that can be used to compare the performance of models. The mean absolute error (MAE), the mean absolute percentage error (MAPE), the mean-squared error (MSE) and root mean square error (RMSE) have been the most commonly used error metrics in the literature. The model which has smaller MAE, MAPE, MSE and RMSE values denote a smaller prediction error and thus it is considered as a more accurate model. In other words, the model which gives low values for the most of these error metrics is selected as the best performing model. These four most common measures of predictive accuracy are computed to evaluate the performance of the spatial and non-spatial models in this study. These measures are calculated using equations (21), (22), (23) and (24) (Sujjaviriyasup and Pitiruek, 2013; Ostertagová and Ostertag, 2012). y_{pred} is the predicted value, y_{obs} is the observed value and *n* is the number of observations used in computing accuracy measure.

The MSE measures the squared difference between predicted and actually observed values. It gives considerably more weight to large errors than smaller ones.

$$\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}(y_{pred}-y_{obs})^2 \qquad [21]$$

RMSE measures the square root of the mean of the square differences between estimations and real prices. The RMSE is equal to the root of MSE and has the advantage of being measured in the same unit as the predicted variable. The smaller the error, the better the forecasting ability of that model according to the RMSE criterion. The RMSE metric, which is the square root of MSE, can be compared with the standard error of the regression. Rule of thumb: an RMSE around two or more times higher than the standard error indicates a weak forecasting performance.

$$\text{RMSE}=\sqrt{MSE} \qquad [22]$$

The MAE is also measured in the same unit as the predicted variable, but gives less weight to large predict errors than the MSE and RMSE (Chappell et al., 2012). The MAE is an absolute measure and this is its biggest disadvantage.

$$\text{MAE}= \frac{1}{n}\sum_{i=1}^{n}|y_{pred}-y_{obs}| \qquad [23]$$

The MAPE measures the forecast quality independent of the unit of measurement of the variable.

$$\text{MAPE}=\frac{1}{n}\sum_{i=1}^{n}|\frac{(y_{pred}-y_{obs})}{y_{obs}}| \qquad [24]$$

The MAPE is often useful for purposes of reporting, because it is expressed in generic percentage terms. One of the drawbacks of the MAPE is that if there are zero values there will be a division by zero. Therefore, the MAPE can only be computed with positive data (Makridakis and Hibon, 1995).

MAPE is the most useful measure to compare the accuracy of forecasts between different items or products since it measures relative performance. If MAPE calculated value is less than 10 %, it is interpreted as excellent accurate forecasting, between 10–20 % good forecasting, between 20–50 % acceptable forecasting and over 50 % inaccurate forecasting.

To make a decision on which forecasting methods are most accurate method is related to the purpose of forecasting (Makridakis, 1993).

CHAPTER 4

IMPLEMENTATION OF METHODOLOGY

The proposed stochastic methodology is illustrated with a case study from Ankara, which is the capital of Turkey. The implementation process is divided into three phases, (1) Geocoding, (2) Spatial and non-spatial analysis, and (3) Cross validation of housing valuation models were carried out to assess how well the models perform against real data.

When the housing dataset that consists of 1265 flats were geocoded, 101 out of 609 observations for Çankaya and 78 out of 656 observations for Keçiören did not match with the urban maps of Çankaya and Keçiören respectively, hence, these records were removed from the datasets. 508 pieces of data for Çankaya and 578 pieces for Keçiören were geocoded on the urban maps. These 508 pieces were separated into two parts, the training data (459) and the testing data (49) for Çankaya. On the other hand, 578 pieces of data for Keçiören were separated as training (522), and testing data (56). The observations for testing were selected randomly and were not used for the analyses in the study except the validation process.

4.1 The Study Area

In this thesis, an empirical study was carried out in the city of Ankara, the capital of Turkey. Two districts were selected with different characteristics in order to identify the most significant factors that affect the housing value. The maps showing study areas and the geographical distribution of data (houses) can be seen in Figure 9. The total number of neighborhoods is 127 for Çankaya, and 51 for Keçiören. Out of these, 76 neighborhoods for Çankaya and 37 for Keçiören were used in this study because of the availability of data.

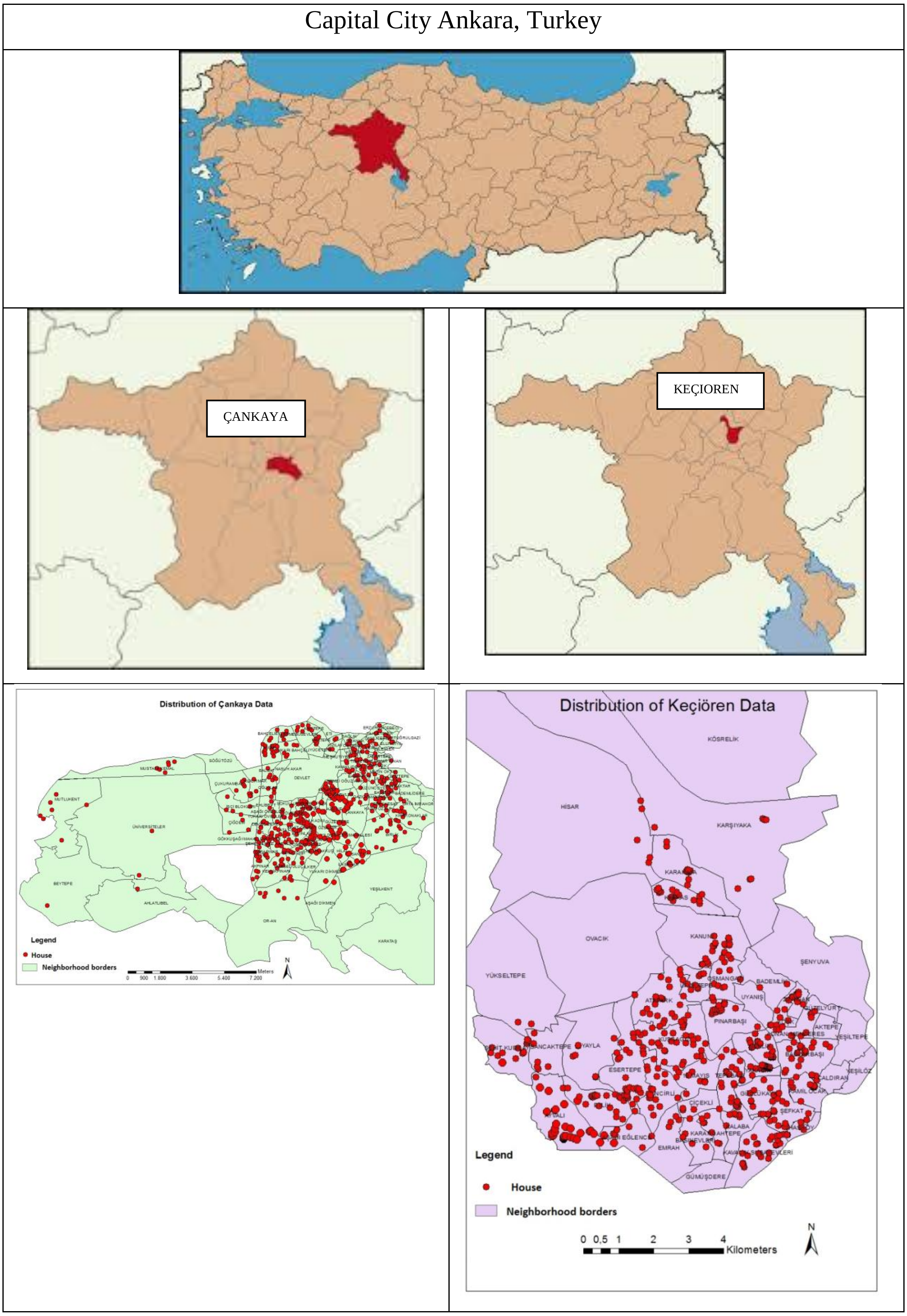


Figure 9: Study area: the bottom left figure shows Çankaya study area and the bottom right figure illustrates Keçiören study area.

The reason why the districts of Çankaya and Keçiören of Ankara were chosen as the case study areas is that Çankaya is a business, cultural and central metropolitan district. On the other hand, Keçiören is the most populated and the second largest district.

According to the 2012 census, the population of the urban center was 832,075. The district covers an area of 268 km^2. The number of housings produced in the 2002-2011 period was 63,873 and the average price per square meter of these housings was 1251 Turkish Liras (1 US dollar is equal to 2,35 Turkish Liras as of the first quarter of 2015). The population growth rate compared to the 2012 data is 2.30%. The numbers of natural gas and water subscriptions in 2012 were 403.544 and 346.677 respectively. Natural gas consumption per dwelling in 2012 was 2165 cubic meters. In 2012, the average annual income for Ankara was 4387 Turkish Liras. The poverty rate (the annual household income is below 4387 TL) is 2,0%. The numbers of teachers, students and classrooms for primary and secondary schools (private and public) in the 2012-2013 term were 7206, 100,143, and 3550 respectively.

On the other hand, Keçiören is the second largest metropolitan district of Ankara with a surface area of 190 km^2, and a population of 840,809 according to the 2012 census. The number of housings produced in the 2002-2011 period was 84,984 and the average price per square meter of these housings was 747 Turkish Liras. The numbers of natural gas and water subscriptions in 2012 were 258,678 and 300,372 respectively. The natural gas consumption per dwelling in 2012 was 1106 cubic meters. The poverty is 6.7%. The numbers of teachers, students and classrooms for primary and secondary schools (private and public) during the term 2012-2013 are 6889, 134,067, and 2900 respectively.

Considering these statistical indicators, it can be said that both districts possess very different economic, demographic and social characteristics. The number of students per teacher for Çankaya is fewer than that of Keçiören. Moreover, there are more classrooms in Çankaya than in the Keçiören district. The poverty rate for Çankaya is less than the one in Keçiören. Apart from this, the numbers of subscriptions for natural gas and water and the in consumptions per housing are higher for the Çankaya district. The Çankaya district has a larger surface area and the number of neighborhoods is greater than Keçiören. Also, the average price per square meter of the housings in Çankaya is greater than that of Keçiören. An underground subway in Ankara serves between the city center and the western part of the city, which belongs to Çankaya. However, the construction work for Keçiören underground system is still underway. Consequently, the residents in Çankaya in general have higher income and better educational and transport facilities compared to those in Keçioren.

4.2 Variables used by real estate appraisal companies in Turkey

The data set used in this thesis, which is listed in Table 5, covers all the variables listed in Table 4, which are determined by the Capital Markets Board. Moreover, the data about theft events, which was collected from insurance companies in Ankara, was used for the first time in this study. The data was gathered together in an MS Excel spreadsheet so that it would be ready for the regression analysis. On the other hand, the GIS is used to measure some distances from the house to the point of interest.

Table 5: All candidate housing valuation variables used in the thesis.

Variable	Variable Name	Category/Type	Description
v1	Value	*Market-based Characteristic; Numeric*	Dependent variable; the variable whose values were wanted to be predicted.
v2	Property type	*Dummy*	A qualitative variable with the value 1 for apartments, 0 for sites
v3	Number of floors	*Structural/physical: numeric*	The number of floors in the building
v4	Building age	*Structural/physical: numeric*	Building age; age of the dwelling expressed in years.
v5	Construction type	*Structural/physical: numeric*	1 concrete, 0 masonry
v6	Construction quality	*Structural/physical: numeric*	1 good, 0 moderate
v7	Type of house	*Structural/physical: numeric*	1 apartment, 0 detached
v8	Type of deed	*Structural/physical: numeric*	Condominium 1 otherwise 0
v9	Floor	*Structural/physical: numeric*	The floor of dwelling
v10	Private Security	*Structural/physical: numeric*	Dummy variable equal to 1 if security exists and otherwise
v11	Car park	*Structural/physical: numeric*	Dummy variable equal to 1 if car park exists and otherwise 0

Table 5: All candidate housing valuation variables used in the thesis (continued)

Variable	Variable Name	Category/Type	Description
v12	Swimming pool	*Structural/physical: dummy*	Dummy variable equal to 1 if swimming pool exists and otherwise 0
v13	Lift	*Structural/physical: dummy*	1 if lift exists otherwise 0
v14	Heating type	*Structural/physical: dummy*	1 if central heating system exists otherwise 0
v15	Number of rooms	*Structural/physical: numeric*	The number of rooms
v16	Number of living rooms	*Structural/physical: numeric*	The number of living rooms
v17	Number of kitchens	*Structural/physical: numeric*	The number of kitchens
v18	Number of bathrooms	*Structural/physical: numeric*	The number of bathrooms
v19	Number of balconies	*Structural/physical: numeric*	The number of balconies
v20	Floor area	*Structural/physical: numeric*	Square meters of living area: measured in usable square meters.
v21	Valuation date	*Market-based Characteristic/ Date: dd/mm/year*	The date on which a valuation report was prepared by an appraisal company for a house.
v22	Distance to health center (network)	*Environmental: numeric*	Distance to the nearest health center via network expressed in meters.
v23	Distance to shopping mall (network)	*Environmental: numeric*	Distance to the nearest shopping malls via network, measured by the shortest path method and expressed in meters.
v24	Distance to the nearest school (network)	*Environmental: numeric*	Distance to the nearest school via network, measured by the shortest path method and expressed in meters.
v25	Distance to park (network)	*Environmental: numeric*	Distance to the nearest park via network, measured by the shortest path method and expressed in meters.
v26	Distance to Şelale (network)	*Environmental: numeric*	This data exists only for the Keçiören district, measured by the shortest path method and expressed in meters.
v27	Distance to Etlik (network)	*Environmental: numeric*	This data exists only for the Keçiören district, measured by the shortest path method and expressed in meters.
v28	Distance to Şelale (Euclidean)	*Environmental: numeric*	This data exists only for the Keçiören district, expressed in meters.
v29	Distance to Etlik (Euclidean)	*Environmental: numeric*	This data exists only for the Keçiören district, expressed in meters.
v30	Distance to nearest bus stop (Euclidean)	*Environmental: numeric*	Distance to the nearest bus stop via Euclidean expressed in meters.

Table 5: All candidate housing valuation variables used in the thesis (continued)

Variable	Variable Name	Category/Type	Description
v31	Distance to subway (Euclidean)	*Environmental: numeric*	Distance to the nearest subway via Euclidean expressed in meters. This data exists for only the Çankaya district.
v32	Distance to health center (Euclidean)	*Environmental: numeric*	Distance to the health center via Euclidean expressed in meters.
v33	Distance to shopping mall (Euclidean)	*Environmental: numeric*	Distance to the nearest shopping mall via Euclidean expressed in meters.
v34	Distance to the nearest school (Euclidean)	*Environmental: numeric*	Distance to the nearest school via Euclidean expressed in meters.
v35	Distance to the nearest park (Euclidean)	*Environmental: numeric*	Distance to the nearest park via Euclidean expressed in meters.
v36	Distance to Kızılay (Euclidean)	*Environmental: numeric*	Distance to Kızılay via Euclidean expressed in meters.
v37	Distance to main transport routes (Euclidean)	*Environmental: numeric*	Distance to main transport routes via Euclidean expressed in meters.
v38	Theft_500 m	*Locational Characteristics (Amenities) of Neighborhoods: numeric*	The number of theft events within a 500 m radius buffer for a house.
v39	Theft_1000 m	*Locational Characteristics (Amenities) of Neighborhoods:numeric*	The number of theft events within a 1000 m radius buffer for a house.
v40	Household size	Demographic: numeric	The number of people living at per house (household): This data exists only for the Keçiören district.
v41	Ratio of higher educated person in neighborhood to district	*Social: ratio*	The ratio of higher educated people in the neighborhood to the total population in the district
v42	Ratio of primary educated person to district	*Social: ratio*	The ratio of primary educated people in the neighborhood to the total population in the district
v43	The number of people in a hectare	*Social: ratio*	Population in one hectare in the district
v44	The number of persons per building in the neighborhoods	*Social: ratio*	The ratio of the population in the neighborhood to the number of houses in the neighborhood

Table 5: All candidate housing valuation variables used in the thesis (continued)

Variable	Variable Name	Category/Type	Description
v45	Rate of theft events in neighborhoods to district	*Social: ratio*	The ratio of the number of total theft events in the neighborhood to the number of total theft in the district

When Tables 6 and 7 are examined, it can be seen that the mean age of buildings for Çankaya is greater than that of for Keçiören. In other words, the observations for Çankaya consist of older houses. According to the housing literature, house age is negatively related with house value hence house prices decline with age. The mean of area variable for Çankaya is also greater than that of for Keçiören. The numbers of theft events in a 500 m and 1000 m radius for Keçiören are nearly twice as many as for Çankaya. The number of people in a hectare for Keçiören is also twice as high as in Çankaya. Considering these findings, it can be said that the number of crime events in Çankaya is lower than in Keçiören and the number of higher educated people in the Çankaya district is higher than in Keçiören. These results clearly show that these two districts are distinctly different from each other in social and cultural terms.

Table 6: Descriptive statistics of candidate variables for Çankaya

Variable Name	Variable	Minimum	Maximum	Mean	Std. Deviation
Ln Value	v1	10.82	13.82	12.19	.54
Property type	v2	0	1	.12	.33
Building floor size	v3	2	34	7.03	3.64
Building age	v4	1	58	21.87	14.78
Construction type	v5	0	1	.97	.18
Construction quality	v6	0	1	.37	.48
Type of house	v7	0	1	.91	.28
Type of deed	v8	0	1	.73	.44
Floor	v9	-1	17	2.18	2.65
Private Security	v10	0	1	.06	.23
Carpark	v11	0	1	.62	.49
Swimming pool	v12	0	1	.02	.14
Lift	v13	0	1	.42	.49
Heating type	v14	0	1	.75	.44

Table 6: Descriptive statistics of candidate variables for Çankaya (continued)

Variable Name	Variable	Minimum	Maximum	Mean	Std. Deviation
Number of rooms	v15	1	8	3.26	.99
Number of living rooms	v16	1	2	0.78	.14
Number of kitchens	v17	1	2	0.62	.22
Number of bathrooms	v18	1	2	0.66	.58
Number of balconies	v19	0	3	1.42	.89
Floor area	v20	50	460	137.61	55.12
Valuation date	v21	2010	2012		
DistancetoMall (N)	v22	500	10000	4564.27	2492.94
DistancetoHealthCente (N)	v23	100	5000	3042.48	1801.12
DistancetoNearestSchool (N)	v24	100	2000	715.47	345.08
DistancetoNearestPark (N)	v25	100	1000	680.61	287.58
Distance to nearest bus stop (E)	v30	100	5000	279.52	299.52
Distance to subway (E)	v31	200	10000	5105.12	3279.26
Distance to health center (E)	v32	100	10000	1158.28	1192.39
Distance to shopping mall (E)	v33	100	10000	2214.27	892.53
Distance to the nearest school (Euclidean)	v34	100	5000	334.64	283.01
Distance to the nearest park (Euclidean)	v35	100	5000	406.75	377.43
Distance to Kızılay (Euclidean)	v36	500	20000	4928.10	3207.91
Distance to main transport routes (Euclidean)	v37	100	500	137.47	68.92
Theft_500 m	v38	1	1945	313.65	360.02
Theft_1000 m	v39	1	4425	1013.80	925.32
Ratio of higher educated people in neighborhood to district	v41	.14	.45	.28	.05
Ratio of primary educated people to district	v42	.05	.17	.10	.03
The number of people in a hectare	v43	10.48	1388.68	122.97	175.65
The number of persons per building in the neighborhoods	v44	.67	24.13	10.22	4.51
Ratio of theft events in neighborhoods to district	v45	.02	8.02	1.31	1.56

Note: E denotes Euclidean distance and N denotes network distance.

Table 7: Descriptive statistics of candidate variables for Keçiören

Variable Name	**Variable**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Ln Value	v1	10,80	13,20	11,68	0,43
Property type	v2	0	1	0,11	0,32
Building floor size	v3	3	26	7,33	4,35
Building age	v4	1	48	13,82	14,02
Construction type	v5	0	1	0,95	0,21
Construction quality	v6	0	1	0,37	0,48
Type of house	v7	0	1	0,94	0,24
Type of deed	v8	0	1	0,54	0,50
Floor	v9	-1	14	1,88	2,57
Private Security	v10	0	1	0,05	0,21
Carpark	v11	0	1	0,65	0,48
Swimming pool	v12	0	1	0,00	0,06
Lift	v13	0	1	0,41	0,49
Heating type	v14	0	1	0,94	0,24
Number of rooms	v15	1	5	3,08	0,75
Number of living rooms	v16	1	2	0,34	0,17
Number of kitchens	v17	1	2	0,31	0,09
Number of bathrooms	v18	1	2	0,27	0,47
Number of balconies	v19	0	3	1,06	0,80
Floor area	v20	60	384	118,38	35,61
Valuation date	v21	2010	2012		
DistancetoMall (N)	v22	100	10000	2891,19	1940,58
DistancetoHealthCente (N)	v23	3000	10000	8287,36	2441,99
DistancetoNearestSchool (N)	v24	200	1000	983,72	99,63
DistancetoNearestPark (N)	v25	100	1000	512,55	249,60
Distance to Şelale (N)	v26	500	5000	4298,85	1342,41
Distance to Etlik (N)	v27	200	5000	4531,03	1139,28
Distance to Şelale (E)	v28	300	10000	4560,92	2642,30
Distance to Etlik (E)	v29	100	10000	4871,65	2378,66
Distance to nearest bus stop (E)	v30	100	2000	242,53	187,29
Distance to health center (E)	v32	100	5000	611,78	546,56
Distance to shopping mall (E)	v33	250	10000	3257,18	1895,25

Table 7: Descriptive statistics of candidate variables for Keçiören (continued)

Variable Name	Variable	Minimum	Maximum	Mean	Std. Deviation
Distance to the nearest school (Euclidean)	v34	100	5000	374,14	408,25
Distance to the nearest park (Euclidean)	v35	100	2000	271,65	219,86
Distance to Kızılay (Euclidean)	v36	7500	20000	10809,39	4348,49
Distance to main transport routes (Euclidean)	v37	100	5000	214,56	281,26
Theft_500 m	v38	1	3680	715,44	670,99
Theft_1000 m	v39	1	8212	2669,13	2222,60
Household size	v40	1	10	2,74	0,83
Ratio of higher educated people in neighborhood to district	v41	0,00	0,25	0,16	0,06
Ratio of primary educated people to district	v42	0,13	0,21	0,18	0,02
The number of people in a hectare	v43	32,29	1494,12	224,36	373,62
The number of persons per building in the neighborhoods	v44	4,17	20,21	15,46	3,65
Ratio of theft events in neighborhoods to district	v45	0,20	7,95	2,95	2,09

Note: E denotes Euclidean distance and N denotes network distance.

Creating a correlation matrix

Before starting the regression analysis, a correlation analysis was performed among all the continuous variables. Correlation is a term referring to the strength of a relationship between two variables, which are dependent (y) and independent (X). Pearson's r is used to measure the linear correlation/dependence between the variables y and X. Numerous guidelines exist to interpret the correlation coefficients (r) of Pearson. A strong, or high, correlation means that two or more variables have a strong relationship with each other, while a weak, or low, correlation means the variables are hardly related. There are certain differences to determine the limits of high or low correlation among these. This difference in assessing the significance of the correlation coefficient originates from different disciplines such as social sciences, engineering, medicine etc. A common scale used to detect Pearson r is the one proposed by Hinkle et al., (1998) as given in Table 8.

Table 8: Scale for evaluation of correlation coefficients

Value of r (social science)	Strength of relationship	Value of r (*)	Strength of relationship
0.90<=r	Strong/very high	-1.0 to -0.5 or 0.5 to 1.0	Strong
0.70<r< 0.90	High	-0.5 to -0.3 or 0.3 to 0.5	Moderate
0.50<r<0.70	Moderate	-0.3 to -0.1 or 0.1 to 0.3	Weak
0.30<r< 0.50	Low	-0.1 to 0.1	None or very weak
r<=0.30	Weak/little		

(*) For Mathematics, Natural and Applied Sciences and Medical Sciences

The correlation analyses between continuous variables are carried out and the correlation matrices are created (see Tables 38 and 39).

The variables of Property type, Building floor size, Building age (negative), Construction quality, Type of house (negative), Floor, Private Security, Carpark, Swimming pool, Lift, Heating type (negative), The number of rooms, The number of living rooms, The number of kitchens, The number of bathrooms, The number of balconies, Area, Distance to Mall (Network route distance), Distance to Health Center (negative) (N), Distance to the Nearest School (N), Distance to the Nearest Park (N), Distance to nearest bus stop (Euclidean distance), Distance to subway (E), Distance to health center (E), Distance to shopping mall (E) (negative), Distance to the nearest school (E), Distance to the nearest park (E), Distance to Kızılay (E), Distance to main transport routes (E), Theft_1000 m (negative), The number of people in a hectare, and The ratio of theft events in neighborhoods to district are significant at the 1 percent significance level. The variables of The ratio of higher educated people in the neighborhood to district and The ratio of primary educated people to district (negative) are significant at the 5 percent significance level. The variables of Construction type, Type of deed, Valuation date, Theft_500 m, The number of persons per building in the neighborhood are insignificant. This means that these variables do not have any significant effect on housing values in Çankaya district at the 5 percent significance level and at the 1 percent significance level.

On the other hand, for the Keçiören district, the variables of Building floor size, Building age, Type of house, Floor, Private Security, Carpark, Lift, Heating type, The number of rooms, The number of bathrooms, The number of balconies, Floor area, Distance to the Nearest Park (N), Distance to Şelale (N), Distance to Şelale (E), Distance to Etlik (E), Distance to the nearest bus stop (E), Distance to shopping mall (E), Distance to Kızılay (E), Theft_500 m, Theft_1000 m, Household size, the Ratio of higher educated people in the neighborhood to district, the Ratio of primary educated people to district, The number of people in a hectare, and The number of persons per building in the neighborhood are significant at the 1 percent significance level. However, the variables of Property type, Construction type, Construction quality, Type of deed, Swimming pool, Number of living rooms, Number of kitchens, Valuation date, Distance to Mall (N), Distance to Health Center (N), Distance to the Nearest School (N), Distance to Etlik (N), Distance to health center (E), Distance to the nearest school (E), Distance to the nearest park (E), Distance to main transport routes (E), and The ratio of theft events in neighborhood to district are insignificant both at the 5 and at the 1 percent significance levels for the Keçiören district.

The variables of (private) Security (r =0,536), The number of rooms (r =0,571), The number of bathrooms (r =0,601) and Floor area (r =0,756) have the highest correlations with the dependent variable (house value) for the Çankaya district. However, for Keçiören, the variables of Building floor size (r =0,574), Floor (r =0,631), Private Security (r =0,720), The number of rooms (r =0,525), The number of bathrooms (r =0,584), and Floor area (r =0,712) have the highest correlations with house values.

Floor area, The number of bathrooms and The number of rooms have a major impact on house values in Çankaya. However, two variables (Floor and The number of floors) are highly effective on the dependent variable in Keçiören. Correlation coefficients can range from -1.00 to +1.00. The value of -1.00 represents a perfect negative correlation while a value of +1.00 represents a perfect positive correlation.

A value of zero means that there is no relationship between the variables being tested. Referring to correlation matrix for Çankaya (Table 38); the independent variables The number of rooms and The number of bathrooms are both correlated with Floor area, and these variables are highly correlated with each other. In this case, the statisticians tell that only one of the variables may contribute significantly to the model and the other variables are dropped from the model. In the Çankaya case, Floor area has a high correlation with both The number of rooms (r =0,789) and The number of bathrooms (r =0,655). Also it can be seen in the correlation matrix for Keçiören (Table 39) that there is a high correlation between The number of rooms and The number of bathrooms (r =0,625). Likewise, in the Keçiören case, the variable of Floor area has a high correlation with both The number of rooms (r =0,721) and The number of bathrooms (r =0,545). However, the correlation is close to the value of the high correlation limit (r =0,491). Since the variable of The number of bathrooms has a bigger effect on the dependent variable than the variable of The number of rooms, the latter one was removed from the list in both models for two districts.

After the correlation analysis, several regressions using the stepwise approach were carried out to test different models based on the remaining variables and tested for their statistical significance (Eckert, 1990). Stepwise regression actually does multiple regression a number of times, each time eliminating the weakest correlated variable. Whereas the only continuous variables were entered as input values for the correlation analysis, both the continuous and discrete variables together were used as input values for stepwise regression in the SPSS 17. At this stage, in order to establish the best house valuation model, numerous combinations were tested with independent variables. The criteria of the best model for housing valuation are to have high coefficients of correlation as much as possible and no insignificant coefficients. The adjusted R^2 was used to measure the explanatory power of the regression.

The most appropriate candidate variables to establish housing valuation model were selected regarding the results of stepwise regression analysis and correlation matrix for both districts. In this context, the variables v2, v10, v20, v31 and v33 were selected for Çankaya and the variables v9, v10, v20, v33 and v44 were selected for

Keçiören to test the assumptions of the OLS and to construct appropriate housing valuation models. The OLS regression analyses were performed based on these candidate variables. The descriptive statistics related to these variables are illustrated in Tables 9 and 10.

Table 9: Descriptive Statistics for Çankaya OLS Model

	Property Type	**Security**	**Area**	**DisttoSubway**	**DisttoMall**	**LnValue**
Minimum	0	0	50	200	100	10.82
Maximum	1	1	288	10000	10000	13.68
Mean	.10	.05	127.04	5029.47	2223.07	12.19
Std. Deviation	0.31	0.21	39.87	3281.47	896.35	0.54

Table 10: Descriptive Statistics for Keçiören OLS Model

	Floor	**Security**	**Area**	**DisttoMall**	**Householdsize**	**LnValue**
Minimum	-1	0	60	250	4.17	10.80
Maximum	14	1	240	10000	20.21	13.20
Mean	1.80	0.04	115.21	3316.73	15.58	11.68
Std. Deviation	2.52	0.20	28.81	1887.23	3.62	0.43

As can be seen in Tables 9-10, the variables Security (private), Area and Distance to Mall are common for both districts. The average housing value (LnValue) and the area size is greater for Çankaya than for Keçiören. Although there were around 40 candidate factors for each case study, only five could be used for each case. The principal reason for this is that the variables which were expected to have an impact on house value the most were not included in the model since there were also high correlations among themselves.

There is a need to verify that these variables have met the regression assumptions. Otherwise, the results can be misleading. As mentioned before, these assumptions are (a) linearity, (b) normality, (c) multicollinearity, (d) autocorrelation, and (e) homoscedasticity. The results of the tests of these assumptions were evaluated below briefly. The first step in an OLS regression analysis is to check the residual plots in order to validate the model. A predictive error, which is the difference between an observed value and its expected value, must be unpredictable. In the OLS regression, random errors are assumed to produce residuals that are normally distributed. The residuals should not be either systematically high or low. Therefore, the residuals should fall in a symmetrical pattern and have a constant spread throughout the range.

As it can be seen in Figures 10 and 11, the scatterplots showed that the interaction between regression standardized residual and regression standardized predicted values are not constant for both districts.

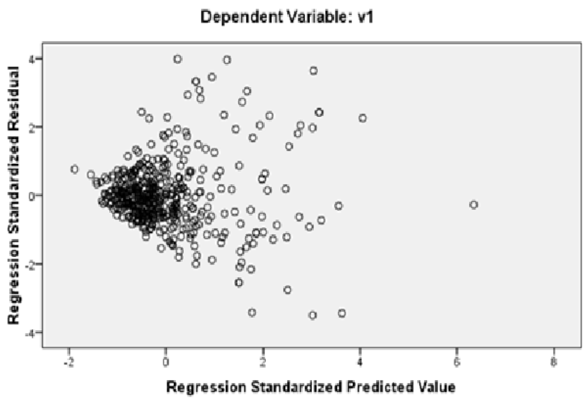


Figure 10: OLS scatterplots for Çankaya before transformation and outlier elimination

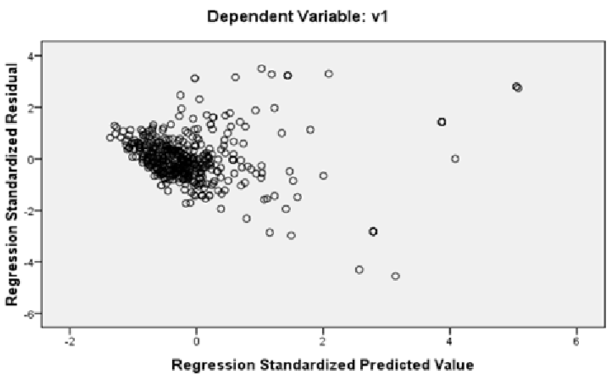


Figure 11: OLS scatterplots for Keçiören before transformation and outlier elimination

This violates the null hypothesis of homoscedasticity, which assumes the existence of constant variance among the residuals. When there are large departures from homogeneity, transformation is frequently advised by authors of statistics textbooks (e.g. Cohen et al., 2013) in order to produce more meaningful results. Since the homogeneity of variance is violated, the dependent variable Value was transformed as LnValue and some outliers were extracted from the dataset. If the transformed variable (LnValue) demonstrates homoscedasticity (equal variance), it is used instead of the untransformed variable (Value) in the analyses. This type of transformation is called semi-logarithmic technique in the literature. Semi-logarithmic technique helps to minimize the problem of heteroscedasticity and normalize the model. Some outliers were eliminated from data sets; therefore, the numbers of data for Çankaya and Keçiören decreased by 45 (from 459 to 414) and 20 (from 522 to 502), respectively. After the semi-log transformation, stepwise regression was re-tested with reduced datasets again for homoscedasticity. According to the repeated regression results, the logarithmic transformation of the dependent value reduced the resulting heterogeneity of variance as shown in Figure 12 for Çankaya and in Figure 13 for Keçiören. These visual results indicate that the assumption of homoscedasticity is not violated for both districts any more.

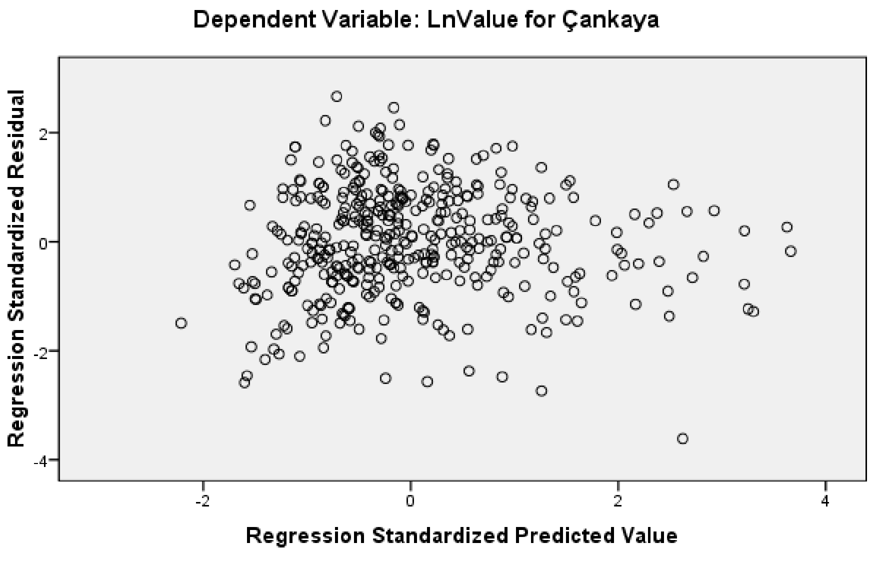


Figure 12: OLS scatterplots for Çankaya after log transformation of dependent variable

.

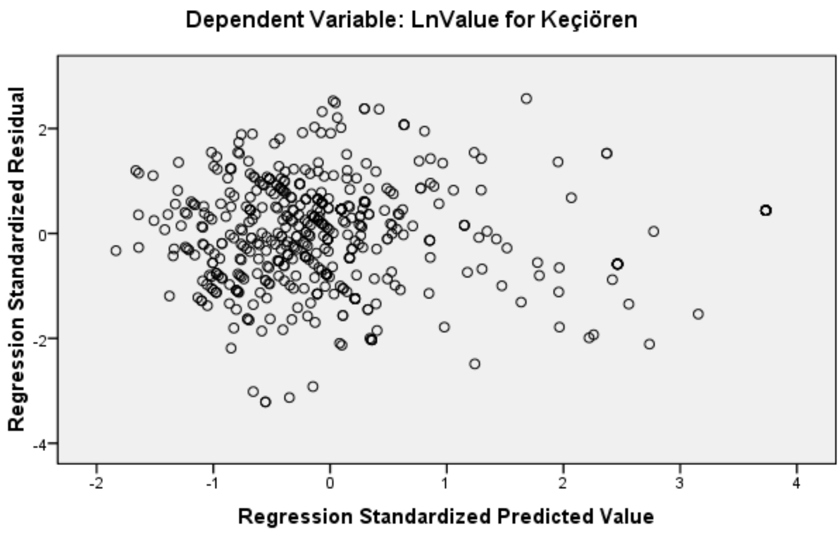


Figure 13: OLS scatterplots for Keçiören after log transformation of dependent variable.

The histograms in Figure 14 for Çankaya and in Figure 15 for Keçiören demonstrate that samples are approximately normally distributed.

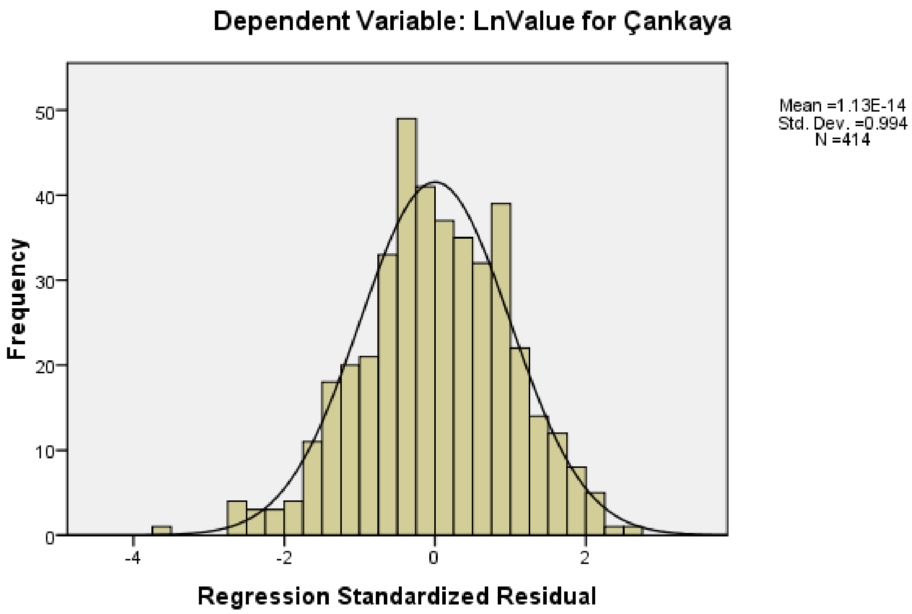


Figure 14: OLS histogram results for Çankaya after transformation

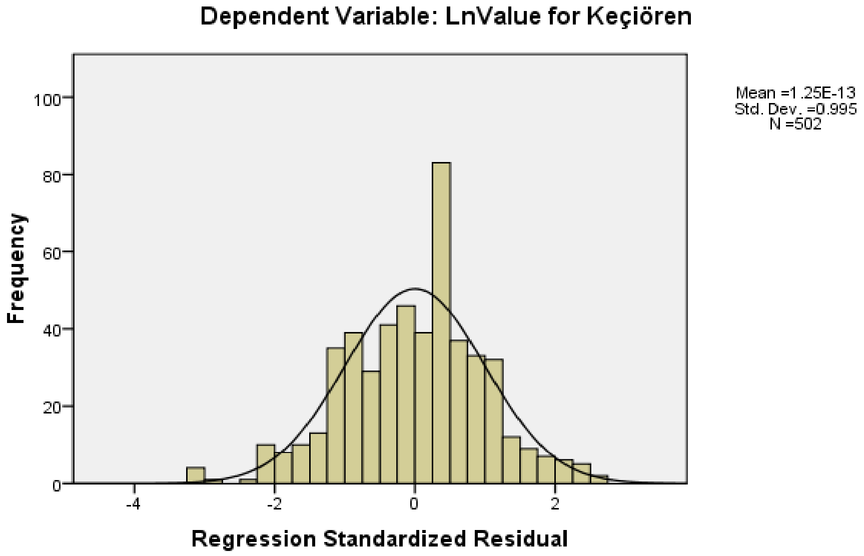


Figure 15: OLS histogram results for Keçiören after transformation

However, as can be seen in Figure 15, the data for Çankaya shows a better normal distribution than Keçiören's.

The relationship between the predictor variables and the outcome variable is assumed to be linear in a linear regression analysis. If the relationship between the predictors and the outcome variable is not linear, then the regression analysis will tend to underestimate the true relationship. An effective method of assessing the linearity is to examine a plot of function of standardized predicted values against standardized residuals (Montgomery et al., 2012). A plot of function of standardized predicted values against standardized residuals was used to check for linearity (Figure 16).

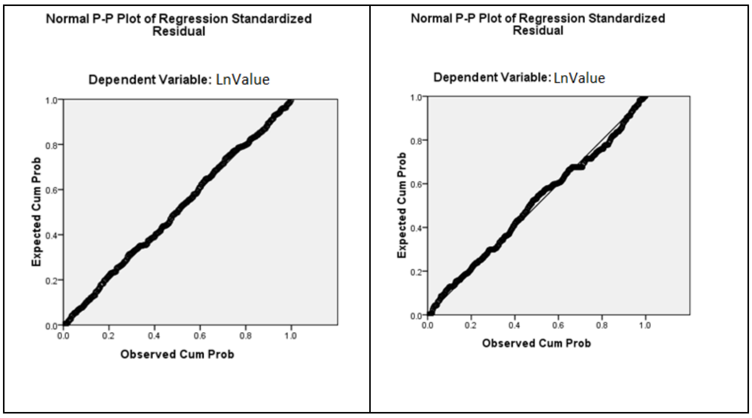


Figure 16: Assessing linearity using standardized residuals for Çankaya (left side) and Keçiören (right side)

Another way to check the existence of linearity assumption is the ANOVA analysis (Tables 11 and 12). According to these tables, the F-tests are highly significant, thus it can be assumed that there is a linear relationship between the variables in both models.

Table 11: OLS ANOVA for Çankaya

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	72.481	5	14.496	260.932	.000
Residual	22.667	408	.056		
Total	95.147	413			

Table 12: OLS ANOVA for Keçiören

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	73.247	5	14.649	355.944	.000
Residual	20.414	496	.041		
Total	93.661	501			

In determining which variables should be included in the regression equation, steps were taken in the application process to minimize the issue of multicollinearity. The presence of multicollinearity is detected using the variance inflation factor (VIF) of each variable after a preliminary regression procedure.

Table 13: OLS Coefficients for Çankaya

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
Constant	11.161	.051		218.157	.000		
Property type (v2)	.205	.042	.130	4.843	.000	.805	1.242
Private Security (v10)	.460	.063	.201	7.363	.000	.784	1.275
Floor area (v20)	.009	.000	.751	28.893	.000	.864	1.157
Distance to subway (E) (v31)	-1.198E-5	.000	-.082	-3.186	.002	.884	1.132
Distance to shopping mall (E) (v33)	-8.623E-5	.000	-.161	-6.551	.000	.966	1.035

Dependent Variable: LnValue

Table 14: OLS Coefficients for Keçiören

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
Constant	10.884	.061		178.985	.000		
Floor (v9)	.040	.005	.235	7.814	.000	.486	2.060
Private Security (v10)	.487	.062	.226	7.908	.000	.540	1.851
Floor area (v20)	.009	.000	.577	24.611	.000	.799	1.252
Distance to shopping mall (E) (v33)	-4.946E-5	.000	-.216	-10.187	.000	.978	1.022
The number of persons per building in the neighborhoods (v44)	-.008	.003	-.069	-3.245	.001	.975	1.025

Dependent Variable: LnValue

Regarding the rule of thumb, it can be said that there is no multicollinearity among the variables in Table 13 for Çankaya and Table 14 for Keçiören. This is because all the VIF values are smaller than 10 and the tolerance scores are bigger than 0.2. The regression coefficients are also shown in Table 13 and Table 14. The results indicate that the coefficients of V10 (security) have a high influence on house values for both districts.

The Durbin-Watson values for Çankaya and Keçiören are 1.63 and 1.76 respectively, which are between the two critical values of $1.5 < \text{critic value} < 2.5$ and therefore it can be assumed that there is no first order linear autocorrelation in both multiple linear regression datasets (Ho, 2006). The R^2 values were investigated to see the data which is the best fit regression line. In general, the higher the R^2, the better the model fits the data. An R^2 of 1 indicates that the regression line perfectly fits the data. In order that R^2 increases with every predictor added to a model, statisticians suggest using the adjusted coefficient of determination (adjusted R^2) instead of the R^2. The reason of this is that the adjusted R^2 shows the percentage of variation in the dependent variable. The coefficients of correlation (R value) for Çankaya and Keçiören are 0.873 (87.3%) and 0.884 (88.4%) respectively. The coefficients of determination (R^2) are 0.762 (76.2%) for Çankaya and 0.782 (78.2%) for Keçiören.

The adjusted R^2 values for Çankaya and Keçiören models are 75.9% and 78.0% respectively.

The final result in the multiple regressions is a regression equation between the dependent variable and several independent variables, which is established on the estimates of the regression coefficients.

The regression model for Çankaya district is

$$V1 = 11.161 + 0.205*V2 + 0.460*V10 + 0.009*V20 - 0.00001198*V31 - 0.00008623*V33 + \varepsilon$$

and for Keçiören is

$$V1 = 10.884 + 0.040*V9 + 0.487*V10 + 0.009*V20 - 0.00004946*V33 - 0.008*V44 + \varepsilon$$

The regression equations show that the Property Type (V2) and Security (V10) for Çankaya and the Floor Level (V9) and Security (V10) for Keçiören have the highest impact on the dependent variable (V1). In addition, all these coefficients have positive signs. The area variable (V20) is a common factor for both equations and has the same influence in both equations. However, the variables of Distance to Subway (V31), Distance to Shopping center (V33) and Size of Household (V44) have negative coefficients.

Figures 17 and 18 show the spatial pattern of the house values for Çankaya and Keçiören respectively using natural break classification with 5 subclasses. The dark brown points indicate the highest housing prices and the light yellow points display the lowest housing prices.

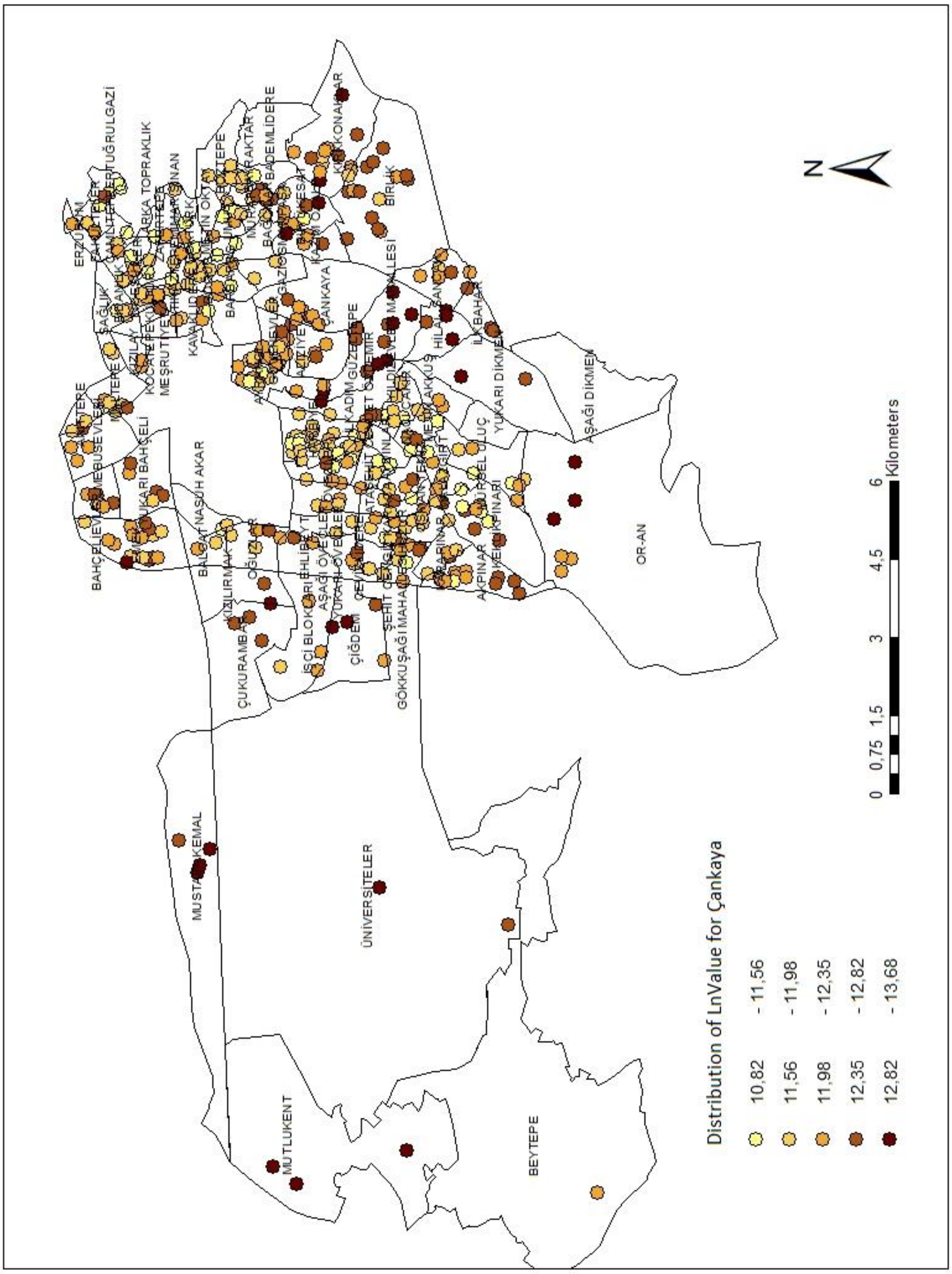


Figure 17: Spatial distribution of the dependent variable for Çankaya

The dark brown points (houses) are concentrated in the northwest of Çankaya. These houses are located in newly urbanized, the fastest growing and expensive residential areas of the two districts. Such areas for the Çankaya district are Mutlukent, Üniversiteler, Mustafa Kemal, Çiğdem, Oran, Dikmen, Yukarı Bahçeli, Özalp, Yıldızevler, Hilal, Yenikonaklar ve Çukurambar (Figure 17). High valued houses in Keçiören are concentrated in the south side of the district. These are Kavacık Subayevleri, Pınarbaşı, Tepebaşı and Etlik (Figure 18).

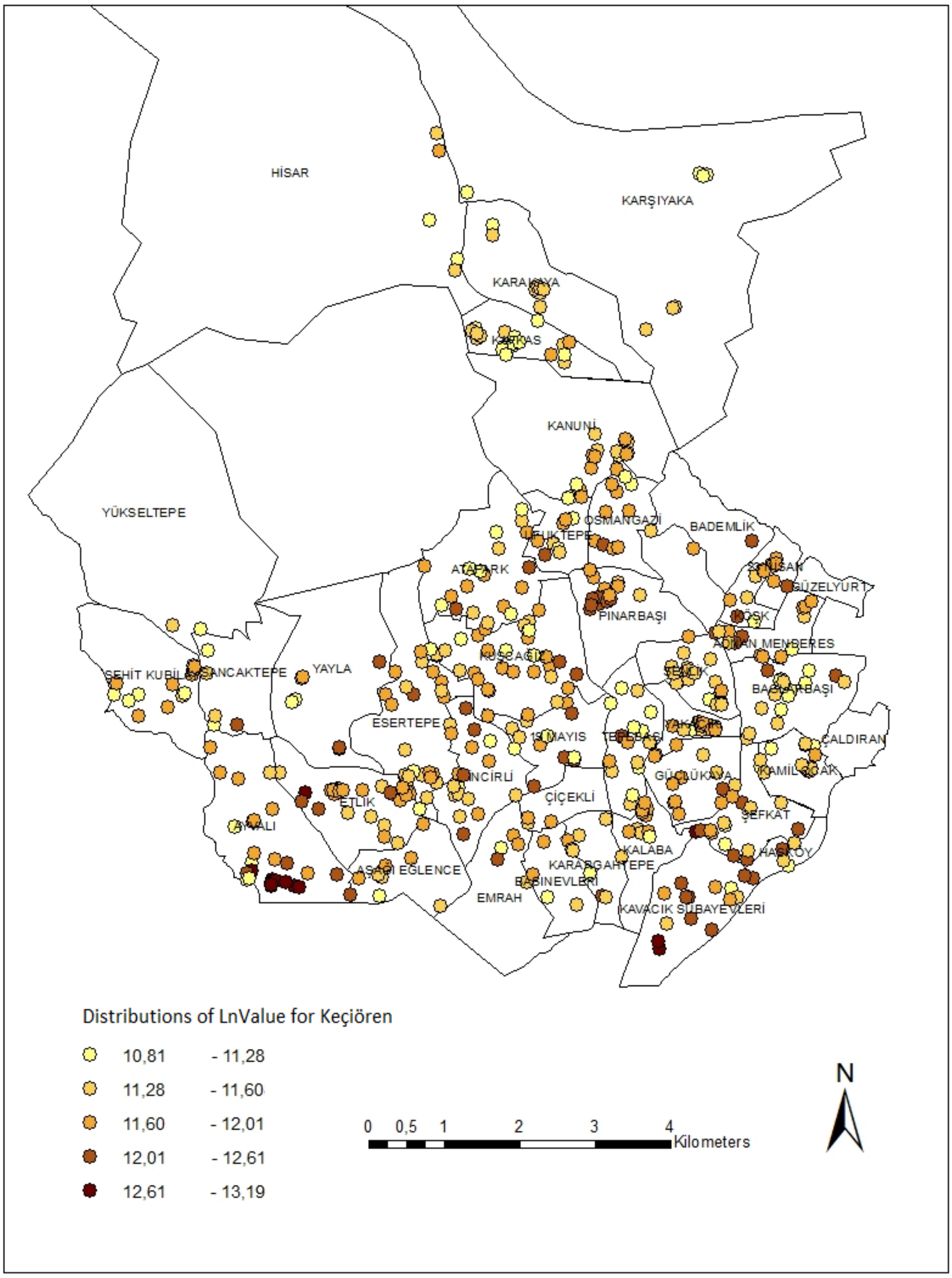


Figure 18: Spatial distribution of the dependent variable for Keçiören

4.3 Spatial Regression Analyses

After the classical OLS estimation, spatial regressions were performed for two districts in order to determine whether autocorrelation exists and to investigate whether non-stationarity exists in different parts of the study area. In order to perform the spatial regression analysis, weight matrices were generated for both districts. These weight matrices were formed based on contiguity (Rook) and distance (threshold).

First, spatial models, namely the spatial lag and the spatial error, were performed by rook and threshold weighted indices. The results of the SARlag and the SARerr for Çankaya are given in Table 15. Also, the results obtained for the log likelihood, the AIC/AICc, the Schwarz criterion and the R^2 indices were presented in Table 15. It can be seen that the AIC (-79.51) and the SIC (-55.36) have the lowest values in the spatial error model; the second lowest value of the AIC (-72.39) and the SIC (-44.21) belongs to the spatial lag model.

Table 15: Comparison of spatial regression models for Çankaya

Model Type	Spatial Lag	Spatial Lag	Spatial Error	Spatial Error
Weight Matrix	Rook	Threshold	Rook	Threshold
Log likelihood	43.19	13.96	45.76	14.06
Akaike info criterion	-72.39	-13.91	-79.51	-16.11
Schwarz criterion	-44.21	14.27	-55.36	8.04
R^2	0.7973	0.7620	0.8089	0.7622
Lag coeff. (Rho)	0.299603	0.228737	0.522748	-0.731435

The additional indicators (W_LnValue and Lambda) reflect the spatial dependence inherent in both the SARlag and the SARerr models. The R^2 value is accepted by statisticians to be a bit problematic with spatial models, so the log-likelihood, the AIC, and the Schwarz are preferred to compare the spatial models in this model. When the model performance parameters are compared regarding Table 15, it can be seen that while the SARerr model with rook weighted has the highest R^2 (0.80) and log likelihood (45.76) values, it has the lowest AIC (-13.91) and SIC (-55.36) values. However, looking at the explanatory variables (signs and magnitudes) it looks like the W_LnValue and Constant variables lost significance in the SARlag models with threshold weighted (Table 16).

Table 16: Spatial lag model weighted by threshold outputs for Çankaya

SPATIAL LAG MODEL-MAXIMUM LIKELIHOOD (Threshold 120 m)				
Variable	Coefficient	Std.Error	z-value	Probability
W_LnValue	0.2287371	0.499482	0.4579487	**0.6469894**
Constant	8.398186	6.03E+00	1.39E+00	**0.1640251**
Property Type	0.202216	4.22E-02	4.79E+00	0.0000017
Security	0.4591625	0.06204791	7.400128	0
Area	0.009039835	0.00031055	29.10906	0
Dist.to Subway	-1.20E-05	3.73E-06	-3.219619	0.0012837
Dist.to Shopping Centers	-8.76E-05	1.33E-05	-6.57302	0

Also, the probability of Lambda variable in the SARerr with threshold weighted (Table 17) and Distance to Subway parameter in the SARerr model with rook weighted (Table 18) lost their significance.

Table 17: Spatial error model weighted by threshold outputs for Çankaya

SPATIAL ERROR MODEL-MAXIMUM LIKELIHOOD (Threshold 120 m)				
Variable	Coefficient	Std.Error	z-value	Probability
Constant	11.15798	0.04957666	225.0651	0
Property Type	0.2088799	0.04184738	4.991469	0.0000006
Security	0.4665908	0.06213008	7.509902	0
Area	0.009040927	0.00031079	29.09042	0
Dist.to Subway	-1.18E-05	3.74E-06	-3.142454	0.0016755
Dist.to Shopping Centers	-8.57E-05	1.31E-05	-6.550683	0
Lambda	-0.7314355	1.1743	-0.6228694	**0.5333702**

Table 18: Spatial error model weighted by rook outputs for Çankaya

SPATIAL ERROR MODEL-ROOK WEIGHTED-MAXIMUM LIKELIHOOD				
Variable	Coefficient	Std.Error	z-value	Probability
Constant	11.19391	0.06061545	184.6709	0
Property Type	0.1127316	0.03889331	2.898483	0.0037499
Security	0.4458239	0.05698054	7.824143	0
Area	0.008526369	0.00027996	30.45518	0
Distance to Subway	8.95E-08	5.83E-06	0.01534441	**0.9877573**
Distance to Shopping Centers	-9.53E-05	1.65E-05	-5.776501	0
Lambda	0.5227483	0.05365398	9.742954	0

Consequently, W_LnValue and Constant parameters in SARlag model with threshold weighted are insignificant (Table 16). Likewise, Lambda parameter in SARerr model with threshold weighted is insignificant (Table 17). Therefore, the SARlag model with rook weighted was chosen as a best fit model for Çankaya dataset (Table 19).

Table 19: Spatial lag model weighted by rook case outputs for Çankaya

SPATIAL LAG MODEL-MAXIMUM LIKELIHOOD ESTIMATION				
Variable	Coefficient	Std.Error	z-value	Probability
W_LnValue (Rho)	0.2996025	0.03749697	7.990047	0
Constant	7.566364	0.4518824	16.7441	0
Property Type (V2)	0.1558621	0.03888198	4.008596	0.0000611
Security (V10)	0.4088057	0.05782128	7.070161	0
Area (V20)	0.008488432	0.000296714	28.60816	0
Dist.to Subway (V31)	-1.64E-05	3.52E-06	-4.658283	0.0000032
Dist.to Shopping Centers (V33)	-5.47E-05	1.29E-05	-4.242461	0.000022

Table 19 shows that five indicators, Constant, Property Type, Security, Area, and Rho are positively related to house value, while two indicators, Distance to Shopping Centers and Distance to Subway, are negatively related and all the p-values (probability) are smaller than 0.05 at the 95 percent confidence level. This indicates the existence of a statistically significant relationship between the dependent and the independent variables for the Çankaya dataset.

On the other hand, the results of the SARlag and the SARerr models for Keçiören are presented in Table 20. According to this table, the SARerr model with rook weighted has the lowest AIC (-231.21) and SIC (-205.89) values and the highest log likelihood (121.60).

Table 20: Comparison of spatial regression models for Keçiören

Model Type	Spatial Lag	Spatial Lag	Spatial Error	Spatial Error
Weight Matrix	Rook	Threshold	Rook	Threshold
Log likelihood	98.83	102.23	121.60	107.48
Akaike info criterion	-183.65	-190.46	-231.21	-202.95
Schwarz criterion	-154.12	-160.93	-205.89	-177.64
R^2	0.7885	0.7916	0.8159	0.7974
Lag coeff. (Rho)	0.0628	0.5911	0.4680	0.9312

Unlike Çankaya, the SARerr model is more appropriate for the Keçiören dataset. In this model, the coefficient on the spatially correlated errors Lambda (λ) appeared (Table 21) as an additional indicator for Keçiören regression and its coefficient (λ) is 0.4680. Thus, it can be said that it has a positive effect and it is highly significant (0.05 > p).

Table 21: Spatial error model weighted by Rook case outputs for Keçiören

Variable	Coefficient	Std.Error	t-Statistic	Probability
Constant	10,91859	0,070186	155,5673	0
Floor Level	0,046818	0,005092	9,194052	0
Security	0,386422	0,070096	5,51277	0
Area	0,008277	0,000333	24,84348	0
Dist. to Shopping Centers	-4,27E-05	7,21E-06	-5,922702	0
Household	-0,009359	0,003203	-2,921717	0,003481
Lambda	0,467973	0,053305	8,779109	0

Table 21 shows that five indicators, Constant, Floor Level, Security, Area and Rho, are positively related to the house prices, while two indicators, Distance to Shopping Centers and Household, are negatively related and all the p-values (probability) are smaller than 0.05 at the 95 percent confidence level. This indicates the existence of a statistically significant relationship between the dependent and the independent variables for the Keçiören dataset.

The Multicollinearity Condition Number (MCN) is widely used to detect the existence of multicolinearity. If the condition numbers are greater than 30, the regression is said to have significant multicollinearity (Paris, 2001). Therefore, the multicollinearity value of the model below 30 is not suggestive of multicollinearity. In this study, the MCNs for Çankaya and Keçiören are 10.82 and 16.88 respectively. This means that multicollinearity is not a problem for both datasets.

The low probabilities of the Breusch-Pagan and Koenker-Bassett tests in Table 22 point to the inexistence of heteroscedasticity for Çankaya at the 0.05 confidence interval. Unlike Çankaya, these results indicate that there is heteroskedasticity at the 0.05 confidence level but not at the 0.01 confidential interval in the Keçiören dataset.

Table 22: Indicators for spatial regression statistic

TEST	DF	VALUE	PROBABILITY
Breusch-Pagan test (Çankaya)	5	5.34	0.38
Koenker-Bassett test (Çankaya)	5	5.10	0.40
Breusch-Pagan test (Keçiören)	5	16.40	0.01
Koenker-Bassett test (Keçiören)	5	14.39	0.01

4.4 GWR Analysis

The GWR analysis was performed to examine how the local parameter estimates vary over space (the spatial patterns). The adjusted coefficient of determination (the adjusted R2) and the ANOVA were used to compare the OLS and the GWR models. The AIC and the SIC were also used for model comparison (Fotheringham et al., 2002).

The concept here is to determine which model could interpret the data better. The summary results of the GWR are listed in Table 23 for Çankaya.

The outputs in Table 23 showed that the GWR model was more suitable than the OLS model because the former could explain 79 percent of the total model variation with the decreased AIC (-55.88).

Table 23: OLS and GWR outputs for Çankaya

Model Type	OLS	GWR
Weight Matrix		Rook
Akaike info criterion	-15.7787	-55.8791
R^2	0.7618	0.8124
Adjusted R^2	0.7589	0.7929

Moreover, the result of Table 24 (ANOVA table for Çankaya) showed that the residuals decreased from 22.677 to 19.005. This means that the GWR model improved (3.66) significantly the results of the OLS model.

Table 24: GWR ANOVA for Çankaya

Source		SS	DF	MS	F
Global	Residuals	22.667	408		
GWR	Improvement	3.661	25.485	0.144	
GWR	Residuals	19.005	382.515	0.05	2.891493

For Keçiören, the outputs in Table 25 showed that the GWR model was more suitable than the OLS model since the GWR model could explain 82 percent of the total model variation with the decreased AIC (-274.13).

Table 25: OLS and GWR outputs for Keçiören

Model Type	OLS	GWR
Weight Matrix		Rook
Akaike info criterion	-170.99	-274.13
Schwarz criterion	-145.678	-163.06
R^2	0.7820	0.8363
Adjusted R^2	0.7800	0.8245

Moreover, the result of Table 26 (ANOVA table for Keçiören) showed that the residuals decreased from 22.41 to 15.33. This means that the GWR model improved (5.08) significantly the results of the OLS model.

Table 26: GWR ANOVA for Keçiören

Source		SS	DF	MS	F
Global	Residuals	20.414	496		
GWR	Improvement	5.084	27.854	0.183	
GWR	Residuals	15.329	468.146	0.033	5.574627

The results of GWR were mapped in order to demonstrate the spatial variation across space. Mennis (2006) published a paper on how to perform the mapping of the GWR results. The method proposed by George Jenks in 1967 is the most popular one to represent the spatial characteristic of values. This classification method, also known as the natural breaks classification method, is based on maximizing the variance between classes and it minimizes the variance within the same classes. This method is also called the goodness of variance fit (Konecny et al., 2010). In this study, the data are classified into five classes to facilitate the visual exploration of the mappable results of the GWR. In this way, the differences between the created classes were emphasized as clusters.

All the maps that were generated for both districts are given in between Figures 61 and 114. The maps between the OLS and the SAR, between the SAR and the GWR and the OLS and the GWR are named as comparison maps by Erener and Düzgün (2010). The comparison maps help to understand the variation of the coefficients at local scale. In this way, the regions where the SAR and the GWR models over- or underestimate the house values were identified.

For the sake of clarity, the dark brown points within the red circle represent clustering of high coefficient values and the light yellow points within the blue circle demonstrate clustering of low coefficient values on the map. This representation sometimes formed a circle, sometimes an ellips and sometimes an axis according to the spatial distribution. The ArcGIS software was used to create the choropleth maps. Choropleth maps or thematic maps are means of displaying areal data obtained by coloring or shading the areal units in accordance with their attribute values of interest (Bailey and Gatrell, 1995).

T-values generated by the GWR were masked out regarding Table 27 in order to demonstrate only those areas where t values are significant. In this framework, all the local GWR results were mapped considering Table 27. In this table, the first column shows t range for one tail, the second and third columns indicate α and Z values corresponding to the value t and the last column represents the range of insignificant values. According to this table, the data values smaller than the lower limit of the interval or those greater than the upper limit of interval lying between insignificant intervals were masked out.

Table 27: Confidence level scale used for mapping of GWR results

T one tail	**α**	**Z**	**Insignificant Range**
t.90	0.10	Z.80	$-1.290<\alpha<1.290$
t.95	0.05	Z.90	$-1.660<\alpha<1.660$
t.99	0.01	Z.98	$-2.364<\alpha<2.364$
t.9995	0.0005	Z.99.9	$-3.390<\alpha<3.390$

The GWR coefficient estimates were mapped for each data location. The estimated GWR coefficients for Çankaya (intercept, property type, security, area, distance to subway and distance to shopping center) were mapped in Figures 19-24. Similarly, the estimated GWR coefficients for Keçiören (intercept, floor level, security, area, distance to shopping center and household) were mapped in Figures 25-30.

Four different confidence intervals (90%, 95%, 99% and 99.95%) were used to show the spatial patterns in space. The maps of coefficients and t-values were produced for each of the confidence intervals. When interpreting the results of the GWR, the map which reflects the spatial variation the best was preferred.

In Figure 19, the houses with high Intercept (often labeled the constant) coefficients were shown by the red circle, while those with low Intercept coefficients were indicated by the blue circle. The neighborhoods of İlkadım, Güzeltepe, Sokullu Mehmet Paşa, Şehit Cevdet Özdemir, Aziziye, Yıldızevler, Hilal and Sancak were clustered by the high Intercept coefficients. In contrast to this, the neighborhoods of Aydınlar, Ata, Osman Temiz, Karapınar, Akpınar, Malazgirt, Mürsel Uluç, Huzur and Keklikpınarı were clustered by the low Intercept coefficients.

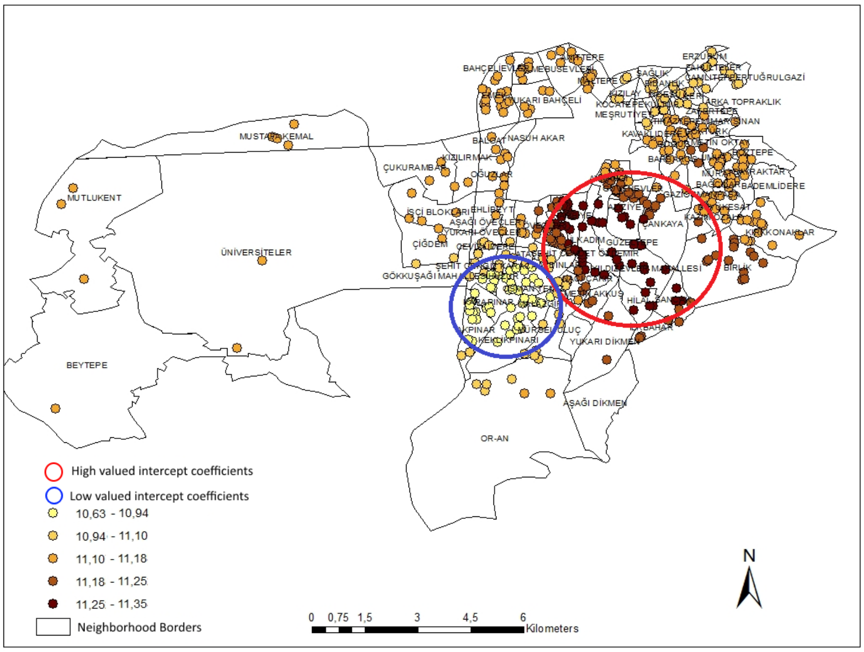


Figure 19: Distribution of GWR "Intercept" coefficients at the 95 % confidence level for Çankaya

In Figure 20, while the houses with high Property Type coefficients were shown by the red circle, the rest of the houses (light yellow) consists of low Property Type coefficients. The neighborhoods of Fidanlık, Öncebeci, İleri, Kültür, İncesu, Arka Topraklık, Zafertepe, Tınaztepe, Göktürk and Seyranbağları were clustered by high Property Type coefficient values.

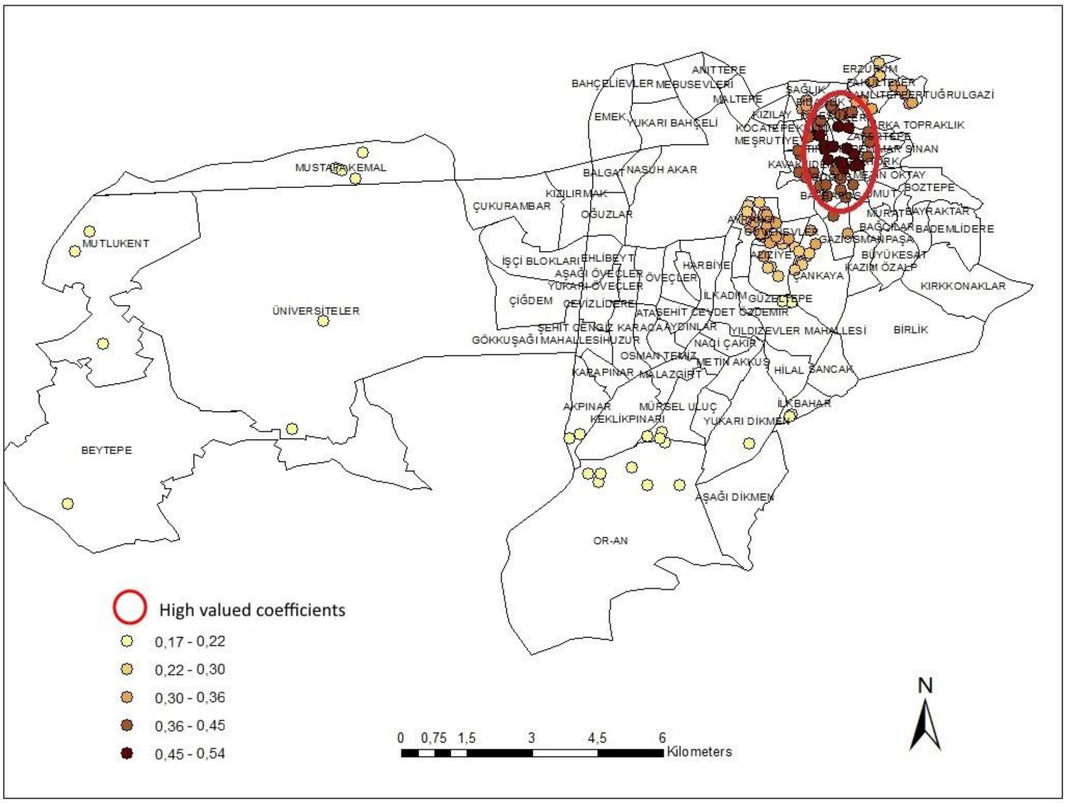


Figure 20: Distribution of GWR “Property Type” coefficients at the 95 % confidence level for Çankaya

In Figure 21, while the houses with high Security (available) coefficients were shown by the red circle, those with low coefficients were indicated by the blue circle. The neighborhoods of Mimar Sinan, Metin Oktay, Göktürk, Umut, Muhsin Ertuğrul, Murat, Bayraktar, Bağcılar, Bademlidere Büyükesat, Kazım Özalp, Kırkkonaklar and Birlik were clustered by high security values. In contrast to this, the neighborhoods of Gökkuşağı, Huzur, Karapınar, Malazgirt, Akpınar, Mürsel Uluç, Keklik Pınarı and the northern part of Oran were clustered by low security values.

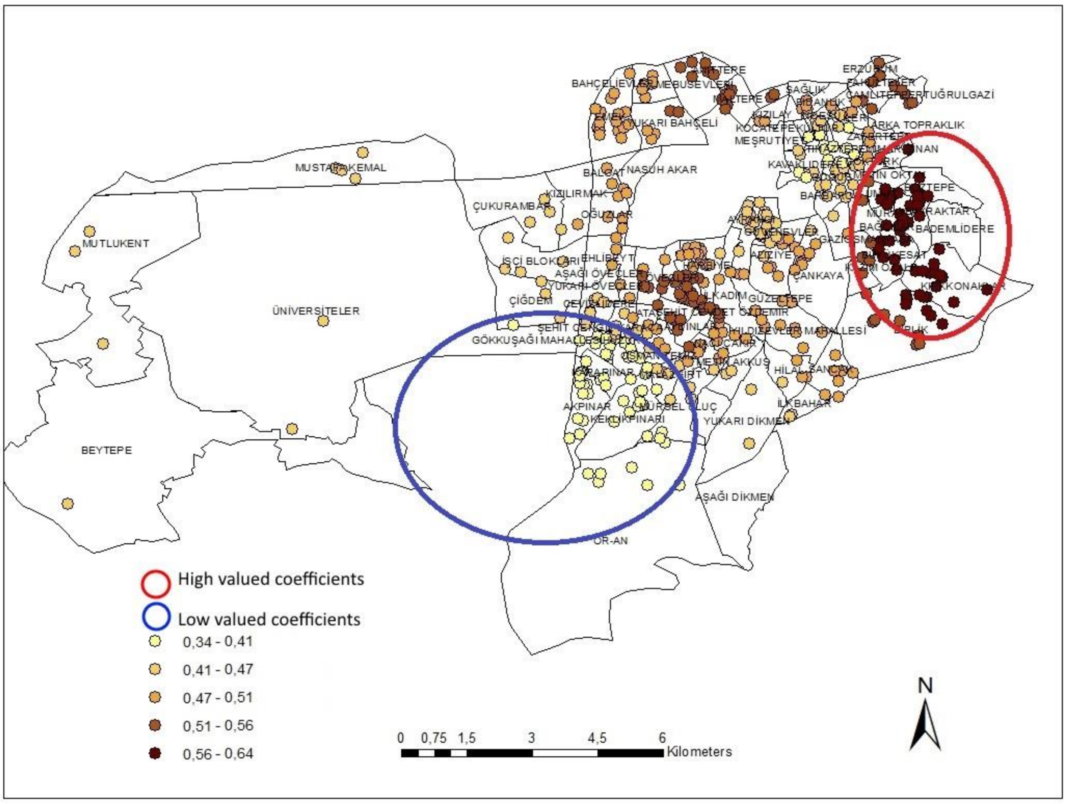


Figure 21: Distribution of GWR "Security" coefficients at the 95 % confidence level for Çankaya

In Figure 22, while the houses with high Area coefficients were shown by the red circle, the houses with low coefficients were indicated by the blue circle. The neighborhoods of Sağlık, Fidanlık, Öncebeci, İleri, İncesu, Ehlibeyt, Aşağı Öveçler, Yukarı Öveçler, Çiğdem, Cevizlidere, Şehit Cevdet Özdemir, Şehit Cengiz Karaca, Ata, Aydınlar, Gökkuşağı, Huzur, Karapınar, Akpınar, Malazgirt, Osman Temiz, Keklik Pınarı and Mürsel Uluç were clustered by high Area coefficients. In contrast to this, the neighborhoods of Çankaya, Kazım Özalp, Büyükesat, Kırkkonaklar, Güzeltepe, Yıldızevler and Birlik were clustered by low Area values.

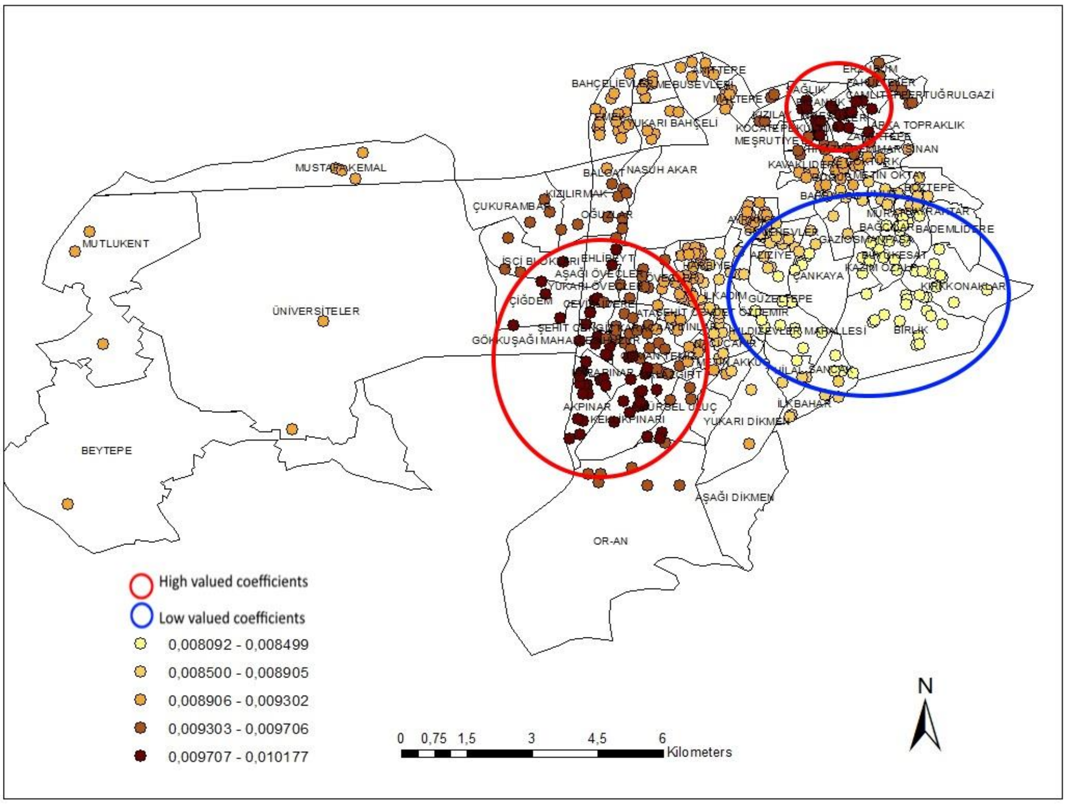


Figure 22: Distribution of GWR "Area" coefficients at the 95 % confidence level for Çankaya

In Figure 23, while the houses with high Distance to Subway coefficients were represented as dark brown points, the houses with low Distance to Subway coefficients were indicated as light yellow points. The neighborhoods of Mutlukent, Mustafa Kemal, Üniversiteler, Beytepe, Çiğdem and Ehlibeyt were clustered by high Distance to Subway coefficient values. In contrast to this, the neighborhoods of Öncebeci, İncesu, Arka Topraklık, Zafertepe, Doğuş and Göktürk were clustered by low Distance to Subway coefficient values. The red line shows the Bahçelievler-Kızılay- Ümitköy underground subway line.

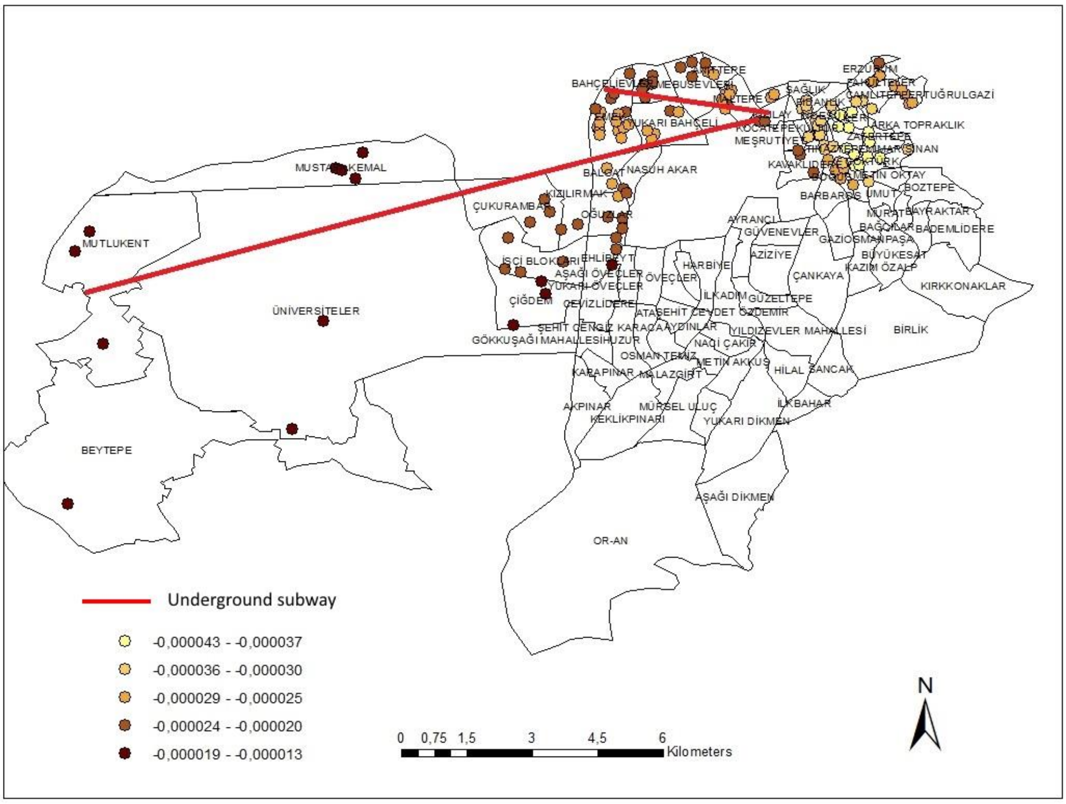


Figure 23: Distribution of GWR "Distance to Subway" at the 95 % confidence level for Çankaya

In Figure 24, while the houses with high Distance to Shopping Center coefficients were represented as dark brown points, the houses with low Distance to Shopping Center coefficients were indicated as light yellow points. The neighborhoods of Mustafa Kemal, Mutlukent, Üniversiteler, Beytepe, Çukurambar, Kızılırmak, Çiğdem, İşçi Blokları, Oğuzlar, Gökkuşağı, Akpınar, Huzur, Ata, Osman Temiz, Bahçeli, Mebusevler, Anıttepe, Kızılay, Kocatepe, Maltepe, Meşrutiyet, Yukarı Bahçeli, Gazi Osman Paşa, Büyük Esat, Bademlidere and Kırkkonaklar were clustered by high Distance to Shopping Center coefficient values. In contrast to this, the neighborhoods of Harbiye, Sokullu Mehmet Paşa, Şehit Cevdet Özdemir, İlkadım, Güzeltepe and Yıldızevler were clustered by low-valued Distance to Shopping Centers coefficients.

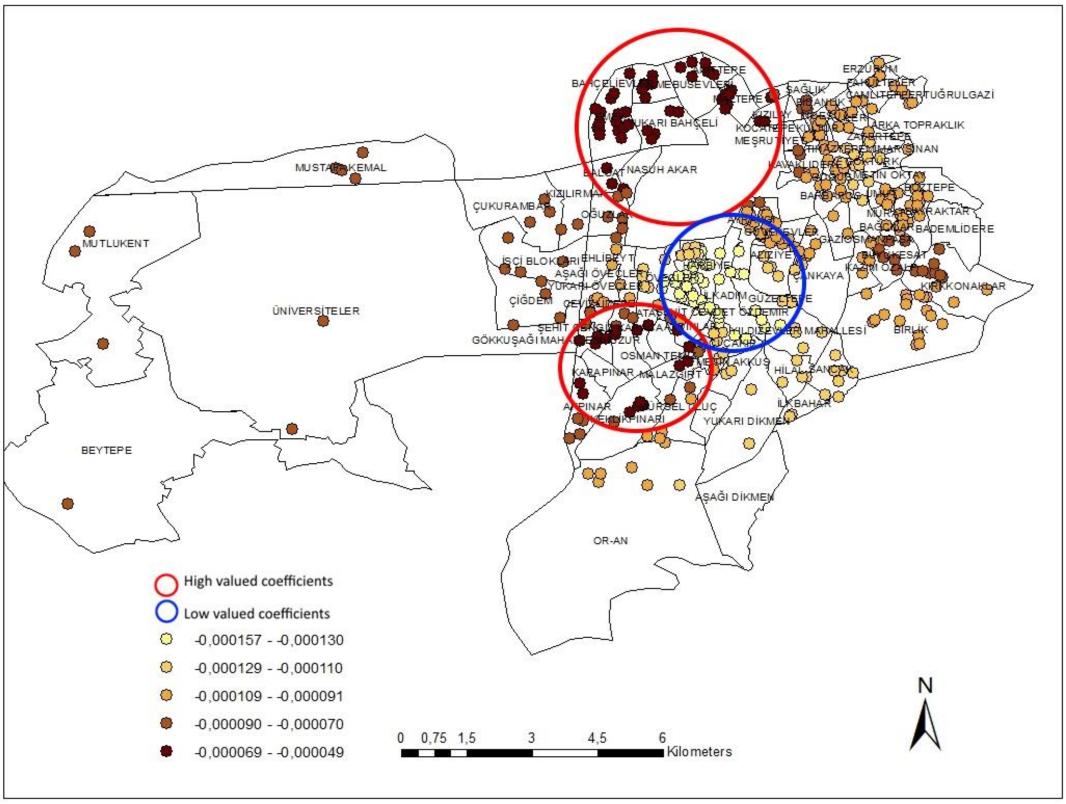


Figure 24: Distribution of GWR "Distance to Shopping Center" at the 95 % confidence level for Çankaya

In Figure 25, while the high-valued Intercept coefficients were shown by the red circle, the low-valued Intercep coefficients were indicated by the blue circle. The neighborhoods on the northern side of the district, namely Karşıyaka, Karakaya, and Kafkas, and the neighborhoods to the southeast, namely Kavacık Subay Evleri, Çaldıran, Hasköy, Şefkat, Bağlarbaşı and Kamil Ocak, were clustered by high Intercept coefficients. In contrast to this, the neighborhoods between two high clustered areas, which are Atapark, Ufuktepe, Osmangazi, Bademlik, Pınarbaşı, Kuşcağız, Köşk, Adnan Menderes and Şenlik, were clustered by low Intercept coefficient values.

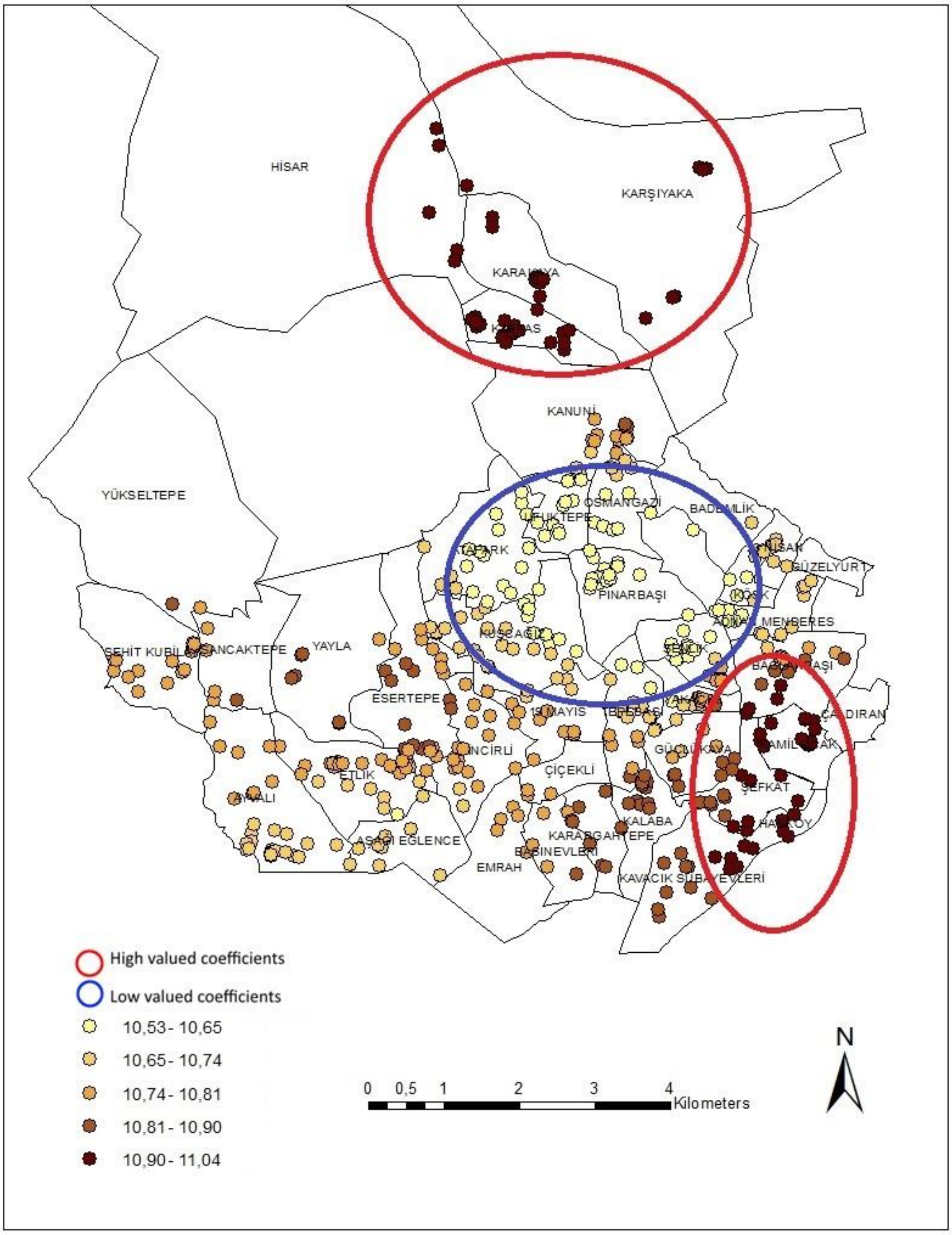


Figure 25: Distribution of GWR “Intercept” coefficients at the 95 % confidence level for Keçiören

In Figure 26, while the high-valued Floor Level coefficients were shown by the red circle, the low-valued Floor Level coefficients were indicated by the blue circle. The neighborhoods of Kavacık Subay Evleri, Şevkat, Hasköy, Güçlükaya and Kalaba were clustered by the houses with high Floor Level values. In contrast to this, the neighborhoods of Karşıyaka, Kafkas, Karakaya and Kanuni were clustered by low Floor Level coefficient values.

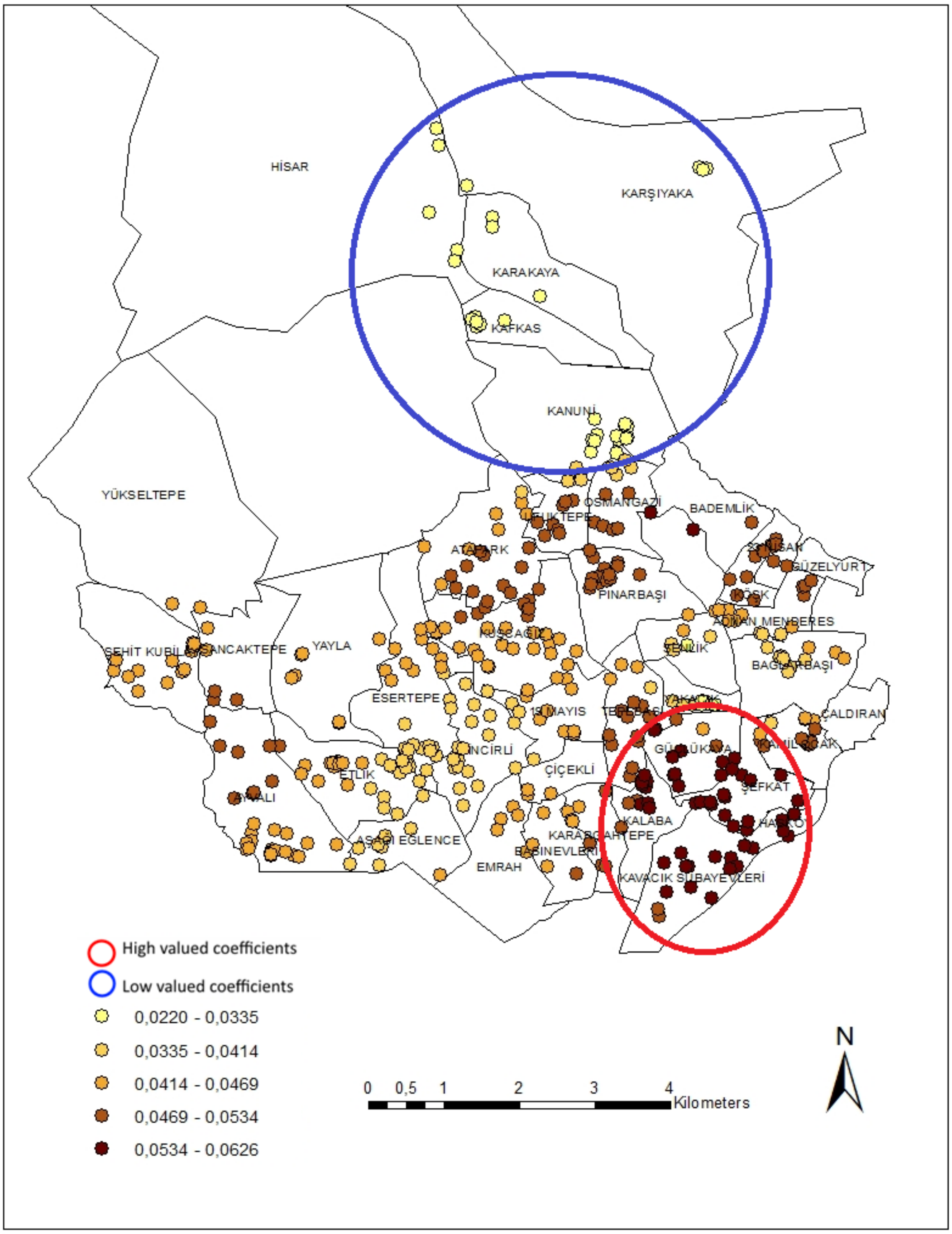


Figure 26: Distribution of GWR "Floor" coefficients at the 95 % confidence level for Keçiören

In Figure 27, while the high-valued Security (available security) coefficients were shown by the red circle, the low-valued Security coefficients were indicated by the blue axis, which runs northwest to southeast. Karakaya, Kafkas and İncirli neighborhoods were clustered by high Security values. In contrast to this, Atatürk, 19 Mayıs, Kuşcağız, Tepebaşı, Kalaba and Kavacık Subay Evleri neighborhoods were clustered by low Security values.

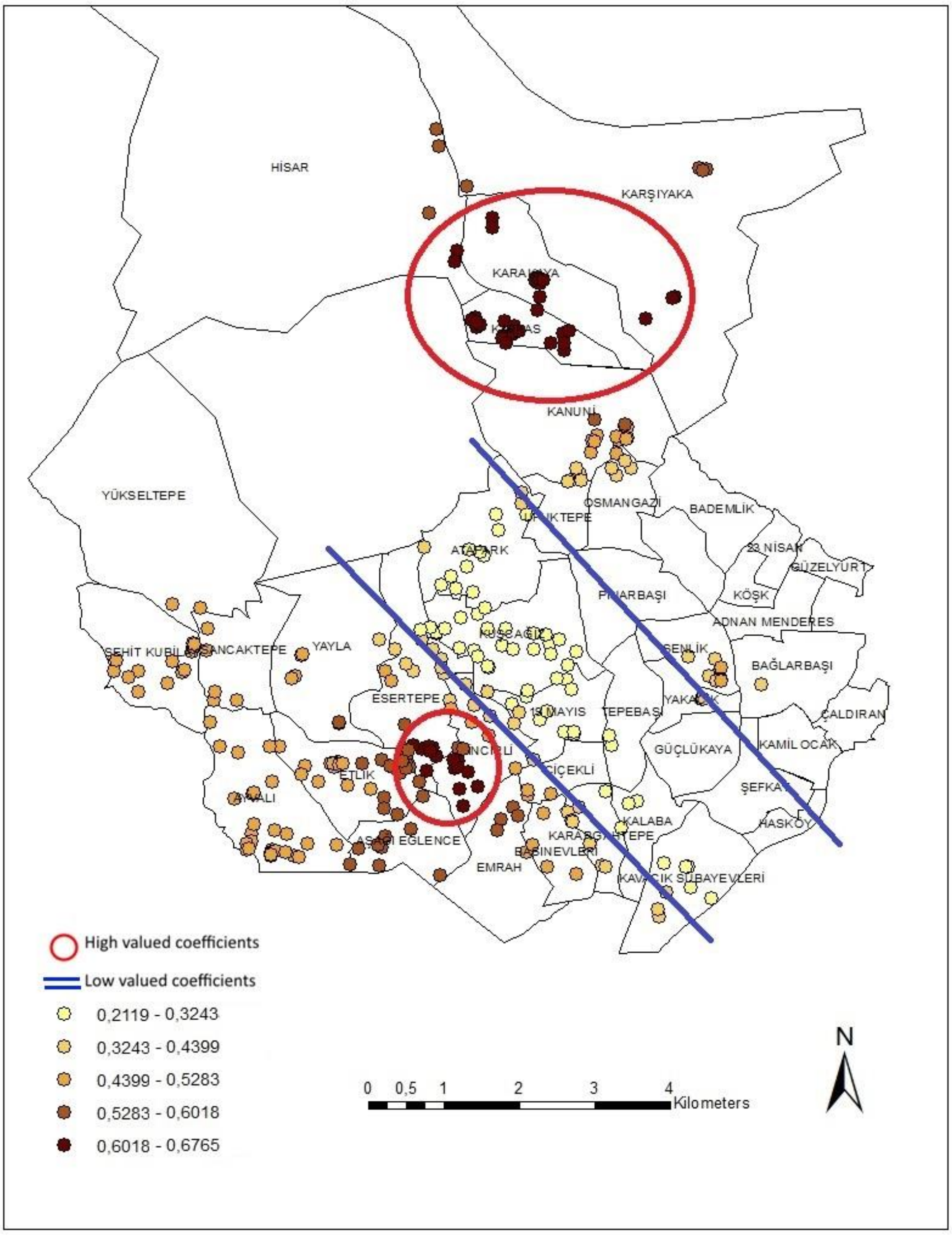


Figure 27: Distribution of GWR “Security” coefficients at the 95 % confidence level for Keçiören

In Figure 28, while the high-valued Area coefficients were shown by the red circle, the low-valued Area coefficients were indicated by the blue circle. The neighborhoods of Şenlik, Bağlarbaşı, Kamil Ocak and Güçlükaya were clustered by high area coefficient values. In contrast to this, Atatürk, Kuşcağız, İncirli and Esertepe neighborhoods were clustered by low area coefficient values.

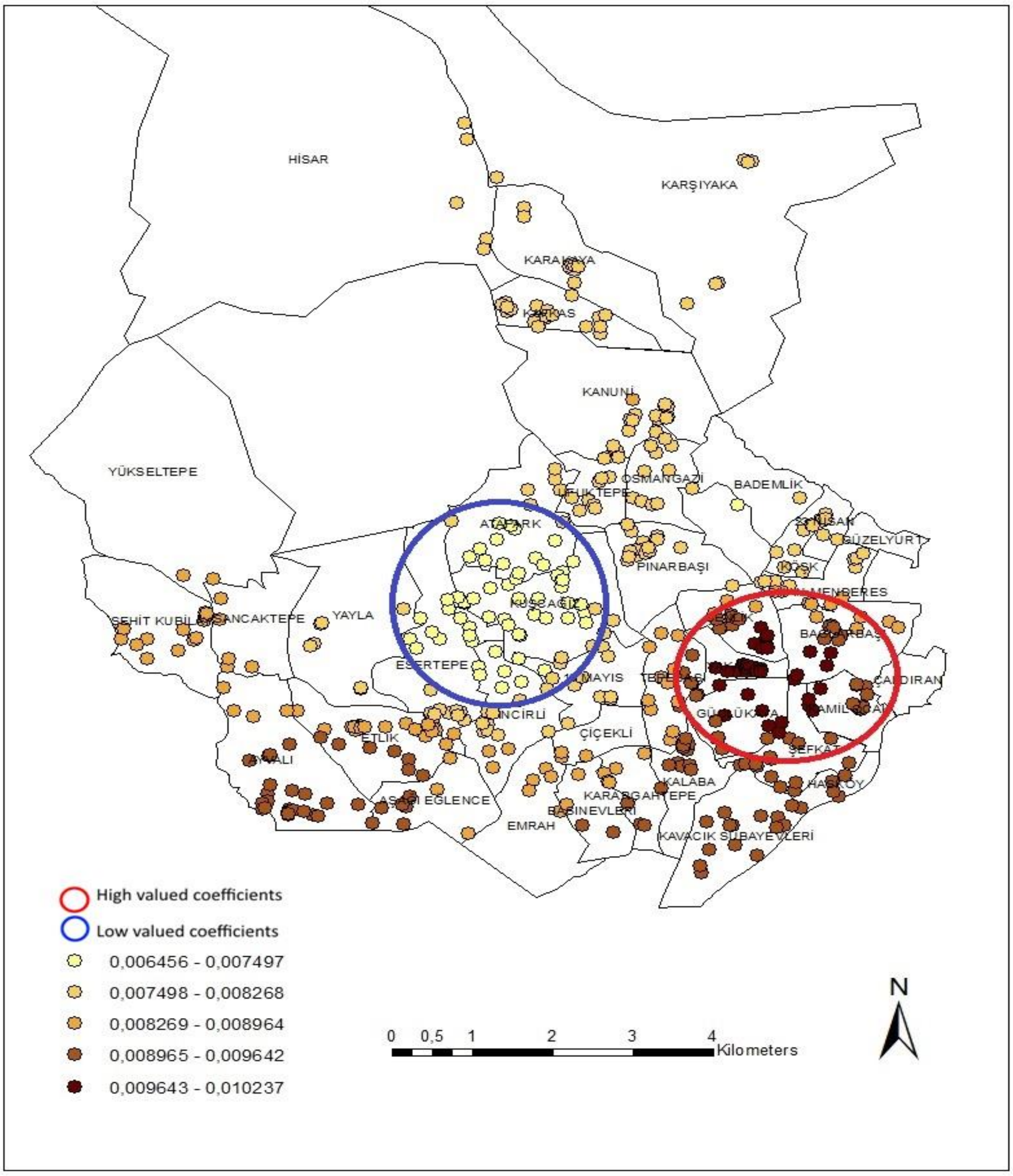


Figure 28: Distribution of GWR “Area” coefficients at the 95 % confidence level for Keçiören

In Figure 29, while the high-valued Distance to Shopping Center coefficients were shown by the red circles, the low-valued Distance to Shopping Center coefficients were indicated by the blue circle. The neighborhoods of Atapark, 19 Mayıs, Tepebaşı Kamil Ocak and Yakacık were clustered by high Distance to Shopping Center coefficient values. In contrast to this, Ayvalı and Etlik neighborhoods were clustered by low Distance to Shopping Center coefficient values.

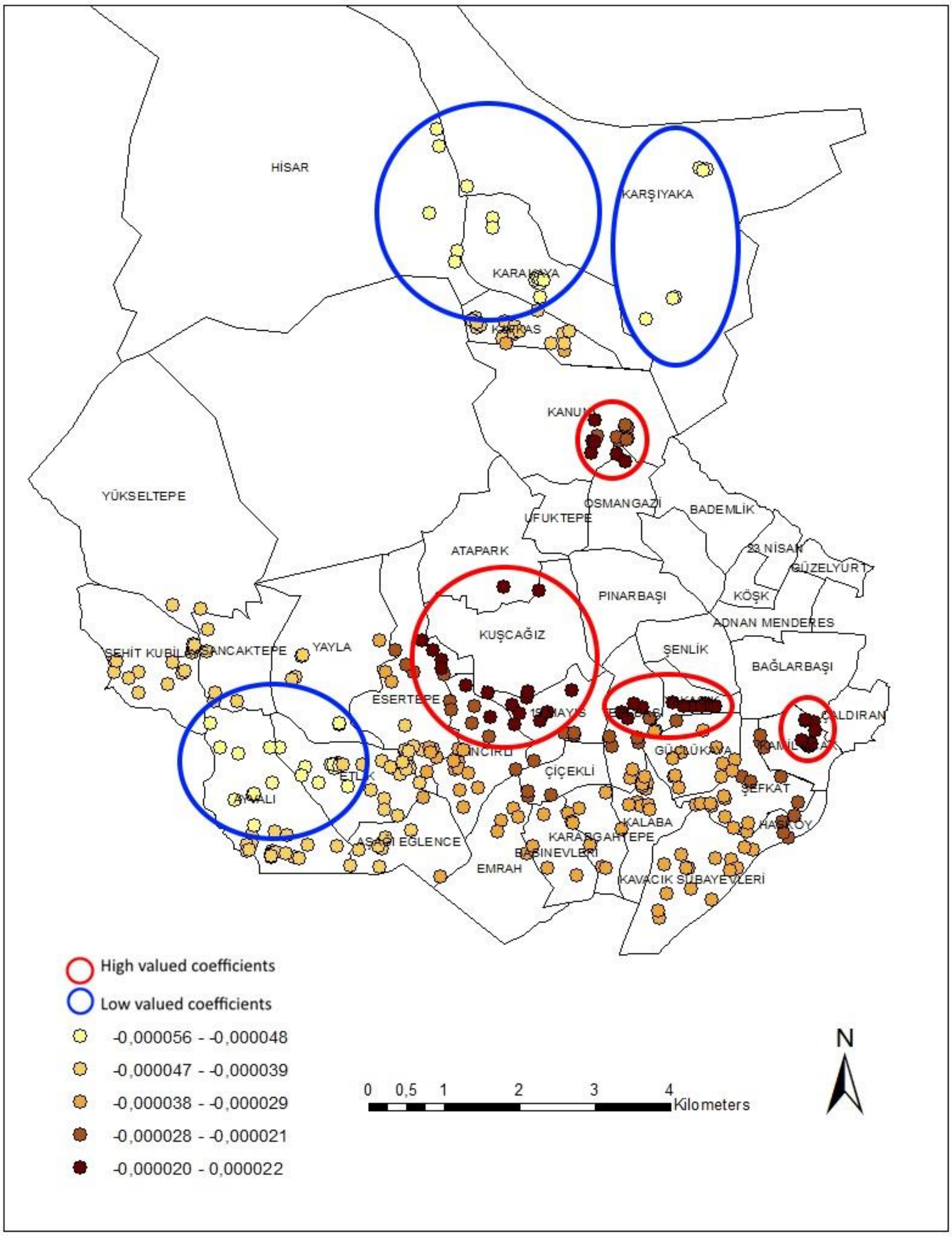


Figure 29: Distribution of GWR “Distance to Shopping Center” coefficients at the 95 % confidence level for Keçiören

In Figure 30, while the high-valued Household coefficients were shown by the red circle, the low-valued Household coefficients were indicated by the blue circle. The neighborhoods of Kanuni, Bağlarbaşı, Karargah Tepe, Tepebaşı and Basınevler were clustered by high Household coefficient values. In contrast to this, the neighborhoods of Kamil Ocak, Şevkat and Hasköy were clustered by the low-valued Household coefficient values.

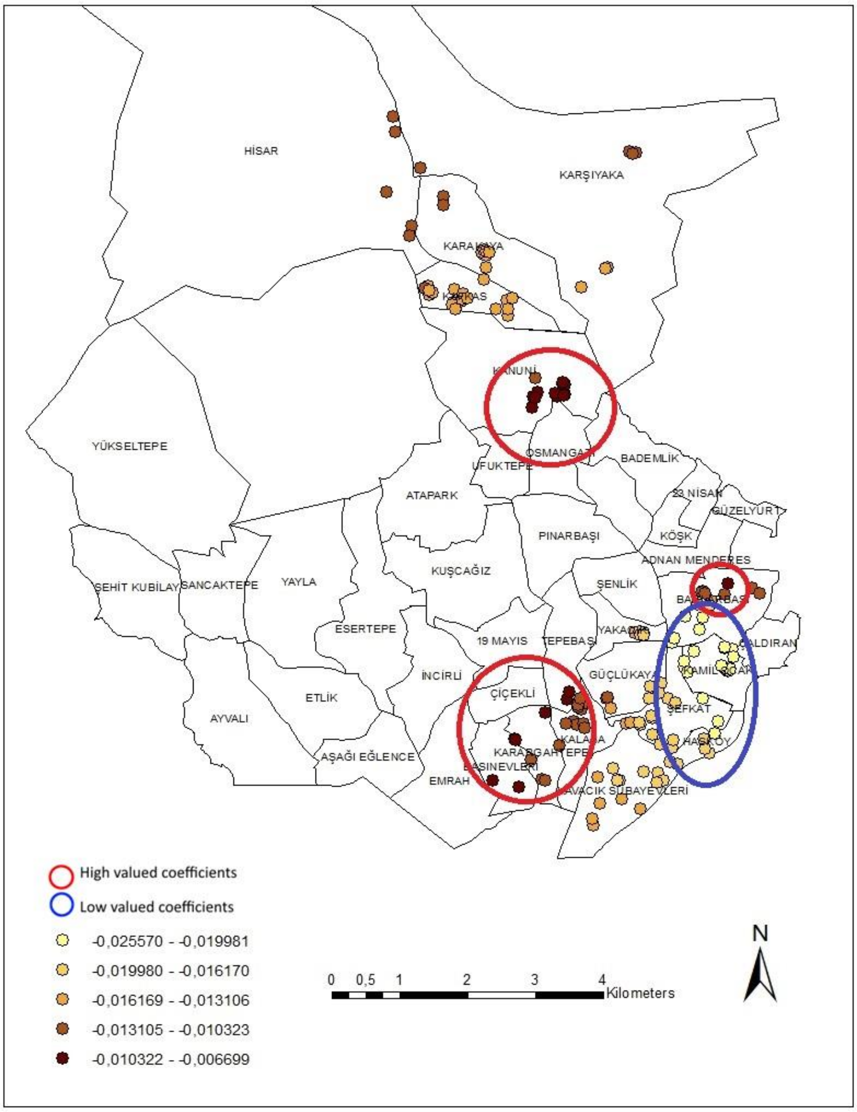


Figure 30: Distribution of GWR Household coefficients at the 95 % confidence level for Keçiören

As can be seen in Figure 31, the residuals demostrate slightly positive or negative clustered areas for Çankaya data.

The map for the residuals was generated according to the difference between the observed and predicted housing values computed by the GWR. In the map, the brown and dark brown points represent under predictions (where the actual house values are higher than the model predicted) and the light brown points display over predictions (actual house values are lower than the predicted).

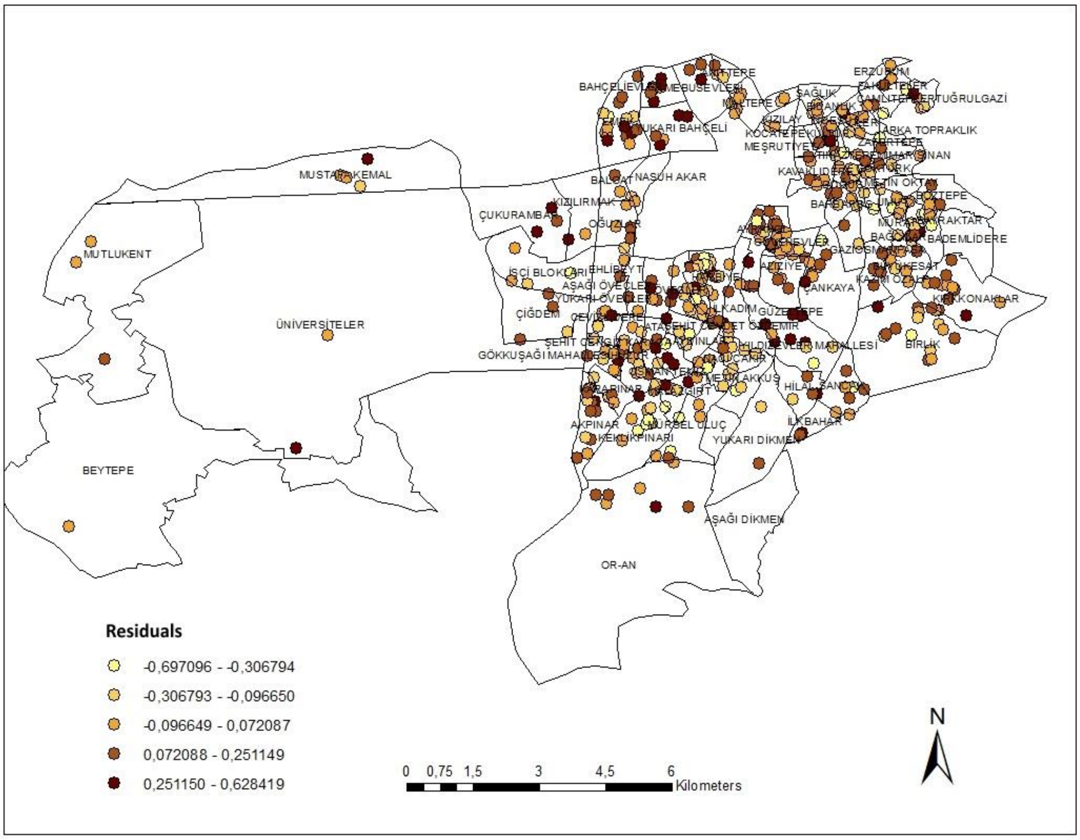


Figure 31: GWR residuals map (between observed and predicted house values) for Çankaya

In Figure 32, it can be seen that the residuals demostrate slightly positive or negative clustered areas for Keçiören data. The map for residuals was produced according to the difference between the observed and predicted housing values computed by the GWR.

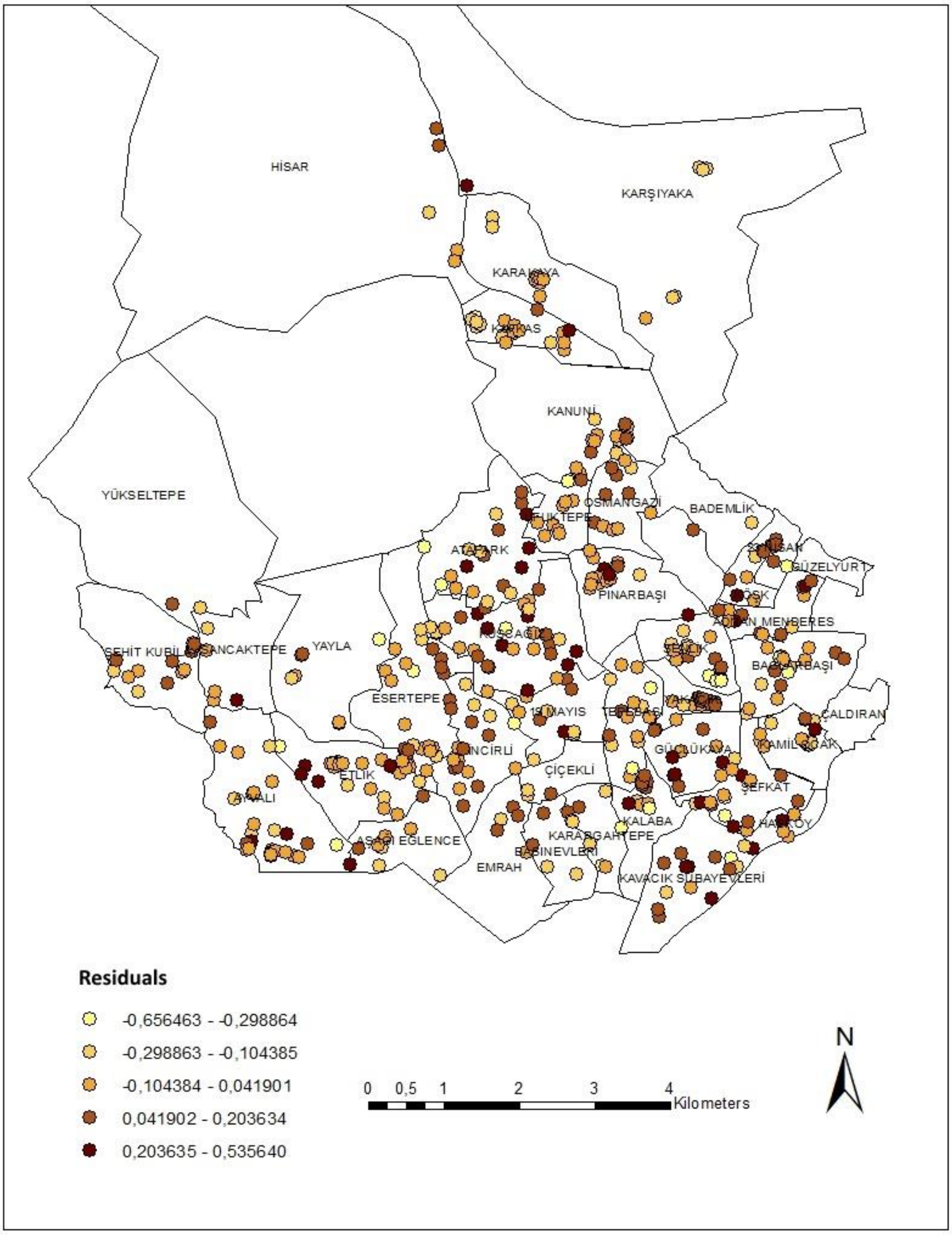


Figure 32: GWR residuals map for Keçiören

A more formal way of detecting residual spatial autocorrelation is to use a spatial correlogram. A spatial correlogram is a plot of a statistic called Moran's I as a function of distance. The resulting correlogram, which is shown in Figure 33 for Çankaya, indicates the presence of a positive autocorrelation at short distance classes of the coefficients at the statistical level α=0.05. Moran plot (Figure 7) has four quadrants of the graphs which identify the local spatial relationship between a space with a high-valued one and its neighbors. The type of association between Quadrants I and III is positive spatial autocorrelation, which is called as spatial clusters, while the one between Quadrant II and IV is negative spatial autocorrelation, called as spatial outliers. In this context, Quadrants for Çankaya can be seen in Figure 33; the red highlighted regions have high values of variables and have also neighbors with high values (41 houses). As indicated in the legend, the blue area is low-low in the same scheme (42 houses), while the pale blue regions are low-high (5 houses) and the pink areas are high-low (12 houses). The strongly colored regions are therefore those that contribute significantly to a positive global spatial autocorrelation outcome, while the paler colors contribute to a negative autocorrelation outcome.

Also, the correlogram in Figure 34 for Keçiören indicates both the presence of a positive autocorrelation and a negative autocorrelation at short distance classes of the coefficients at the statistical level α=0.05. In the Moran correlogram for Çankaya, the autocorrelation coefficient is positive up to lag 6 (distance class 6) and decreases up to lag 6. It decreases again from distance lag 7 to 8. The positive similarity between the samples decreases up to distance class 6 and then the negative similarity decreases up to distance class 8.

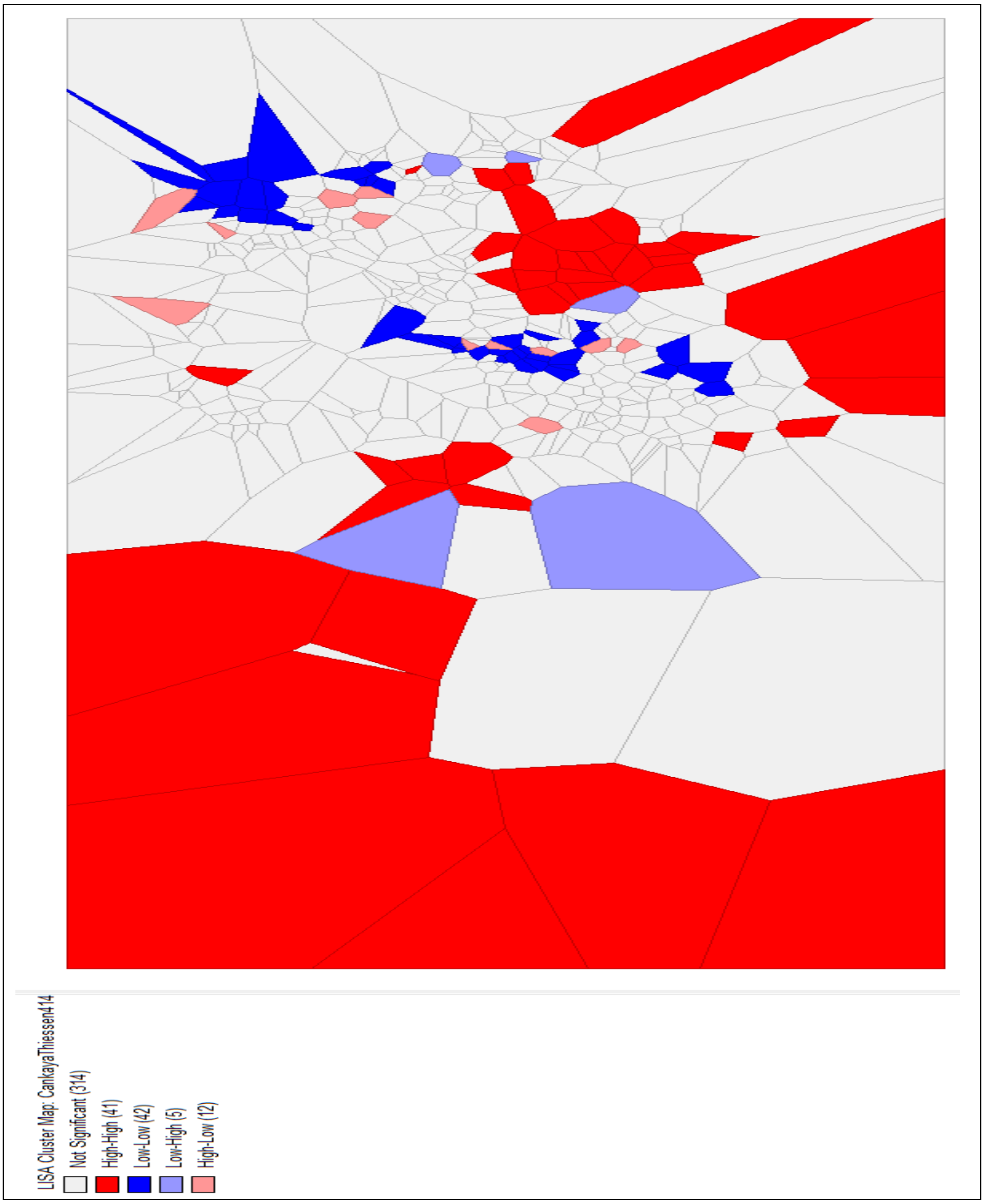


Figure 33: Cluster map for Çankaya

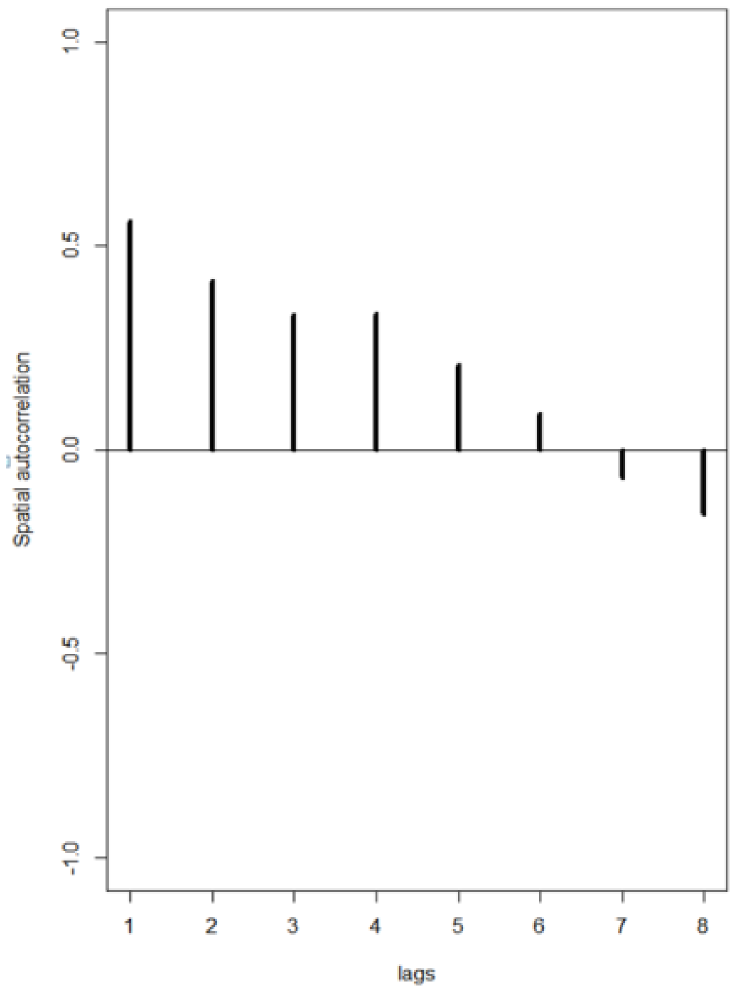


Figure 34: Moran's I spatial correlogram for Çankaya

The classifications of Quadrants for Keçiören are 30 houses (high-high) with red highlight, 50 houses (low-low) with blue highlight, 6 houses (low-high) with pale blue highlight and 8 houses (high-low) with pink highlight in Figure 35.

In the Moran correlogram for Keçiören (Figure 36), the autocorrelation coefficient increases up to lag 2 (distance class 2) and decreases up to lag 3. It increases again from distance lag 3 to 5. The high similarity between the samples decreases up to distance class 3 and then the low similarity increases up to distance class 5.

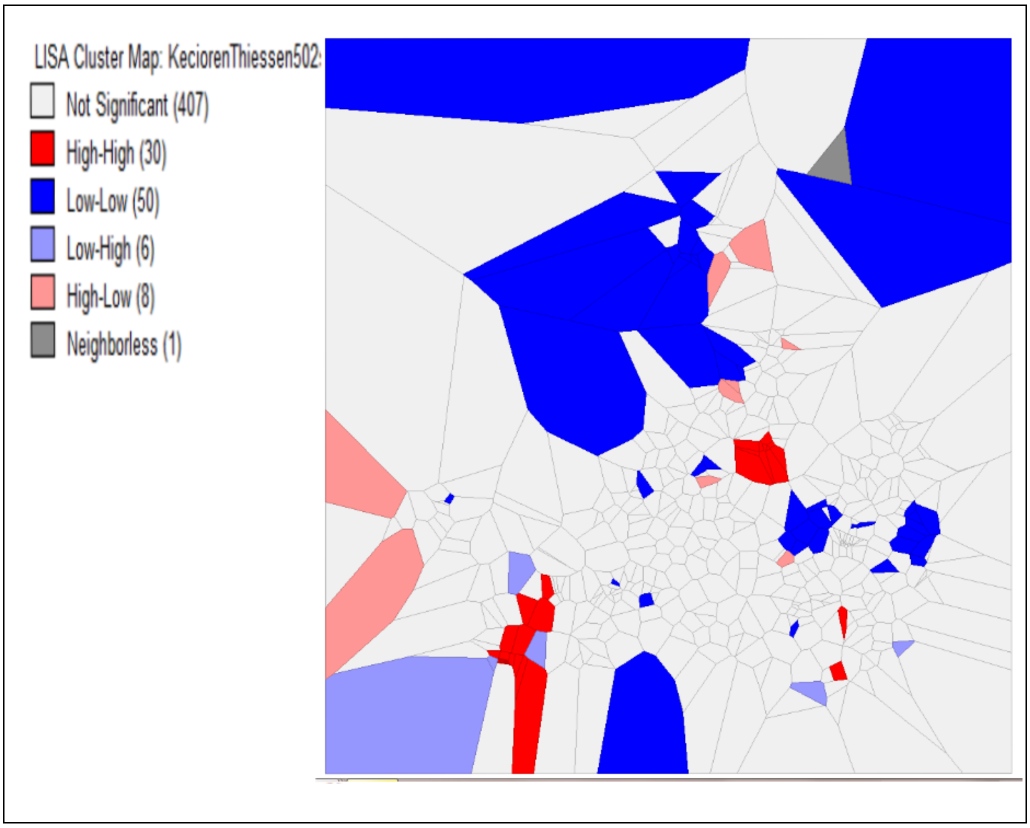


Figure 35: Cluster map for Keçiören

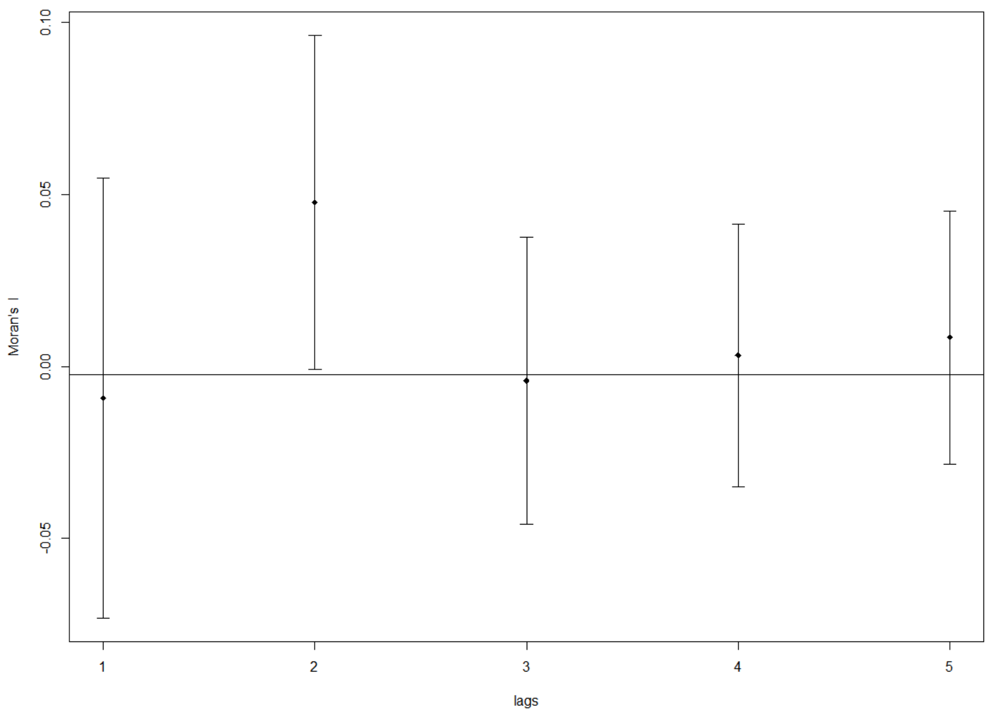


Figure 36: Moran's I spatial correlogram for Keçiören

One of the goodness-of-fit measures is local R^2, which indicates how well the local regression model fits the values of observed dependent variable (value). It varies between 0.0 and 1.0, and high values (close to 1) indicate that the local model performs well. In order to see where the GWR estimates well and where it estimates poorly, the values of local R^2 were mapped for two districts (Figure 37 and 38). Local R^2 values vary between 0.71 and 0.81 for Çankaya.

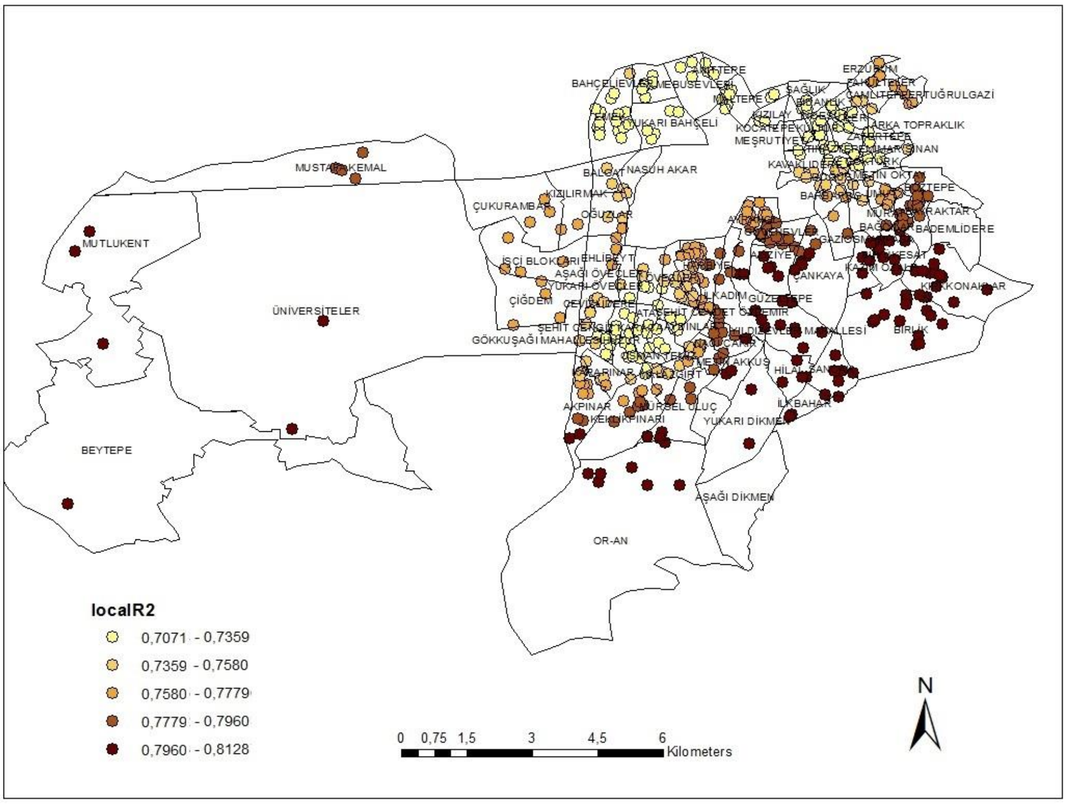


Figure 37: Distribution of GWR local R^2 values for Çankaya

On the other hand, the values of Local R^2 change between 0.61 and 0.93 for Keçiören. Since all these local R^2 values are high, that is to say, close to 1, it can be said that the local models display a good performance in both districts.

However, while the local R^2 interval for Çankaya is 10 units, this interval for Keçiören is 36. This shows that the model goodness-of-fit for Çankaya is better than Keçiören.

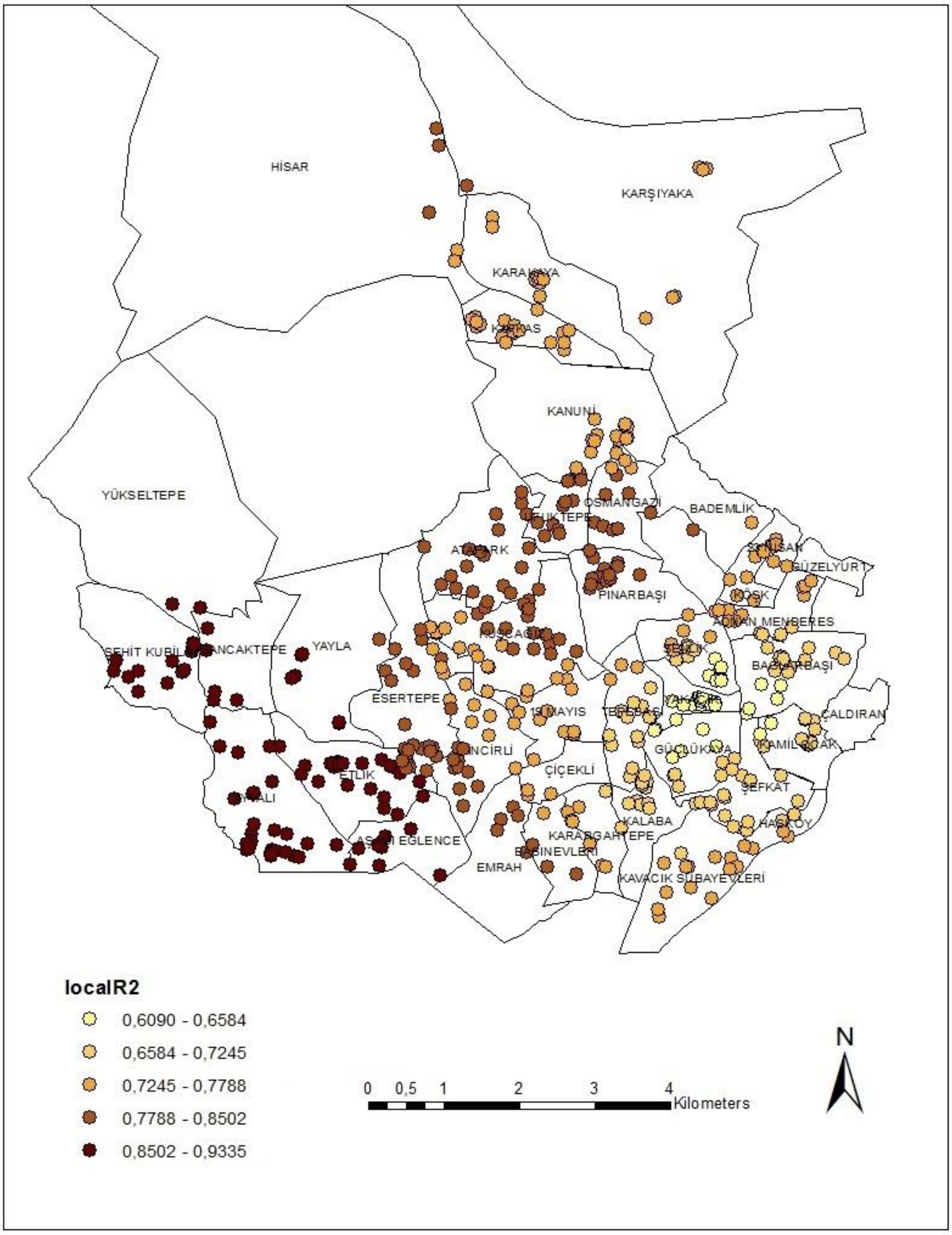


Figure 38: Distribution of GWR local R² values for Keçiören

4.5 Validation

The model selection among competing models is one of the important tasks in regression analyses. There are a large number of criteria to select the best valuation model for which the predicted values tend to be the closest to the true expected values.

In this section, two different approaches were used to determine the best fit model: linear line and mathematical metrics based on the measures of errors. The estimation of OLS, SAR and GWR models were tested using the test data set in Excel. It is assumed that the relationship between the observed and the predicted values is linear. Regression line was used as a way of visually depicting the relationship between the observed and the predicted variables in the graphs below. The predictive power of the model is measured by the proximity to regression line of the predicted values. Theoretically, if a model could predict 100% of the observed value, all the prediction points would fall on the fitted regression line. The smaller the differences between the observed values and the model's predicted values are, the better the predictive power of the model gets. R^2 close to 1 is also a good indicator for this. The test results of each model for Çankaya are shown separately in Figure 39. In this figure, the y axis displays the observed values of houses and the x axis shows the predictions of the model. Considering the R^2 values, it is seen that the SAR model has the highest R^2 value (R^2=0.80). The GWR gives the second highest R^2 value (R^2=0.72).

Figure 40 shows comparison of the prediction of the models with observed house values (test data set). Looking at the graph, it can be seen that the predictions of the SAR model are closer to the regression line. This means that the performance of the SAR model is better than the others for Çankaya data set.

In the Figures 39 and 41, the horizontal axis shows observed values and vertical axis indicates predicted values made by the model. The values in the both axes of the Figures 39, 40, 41 and 42 are denominated in Turkish Lira.

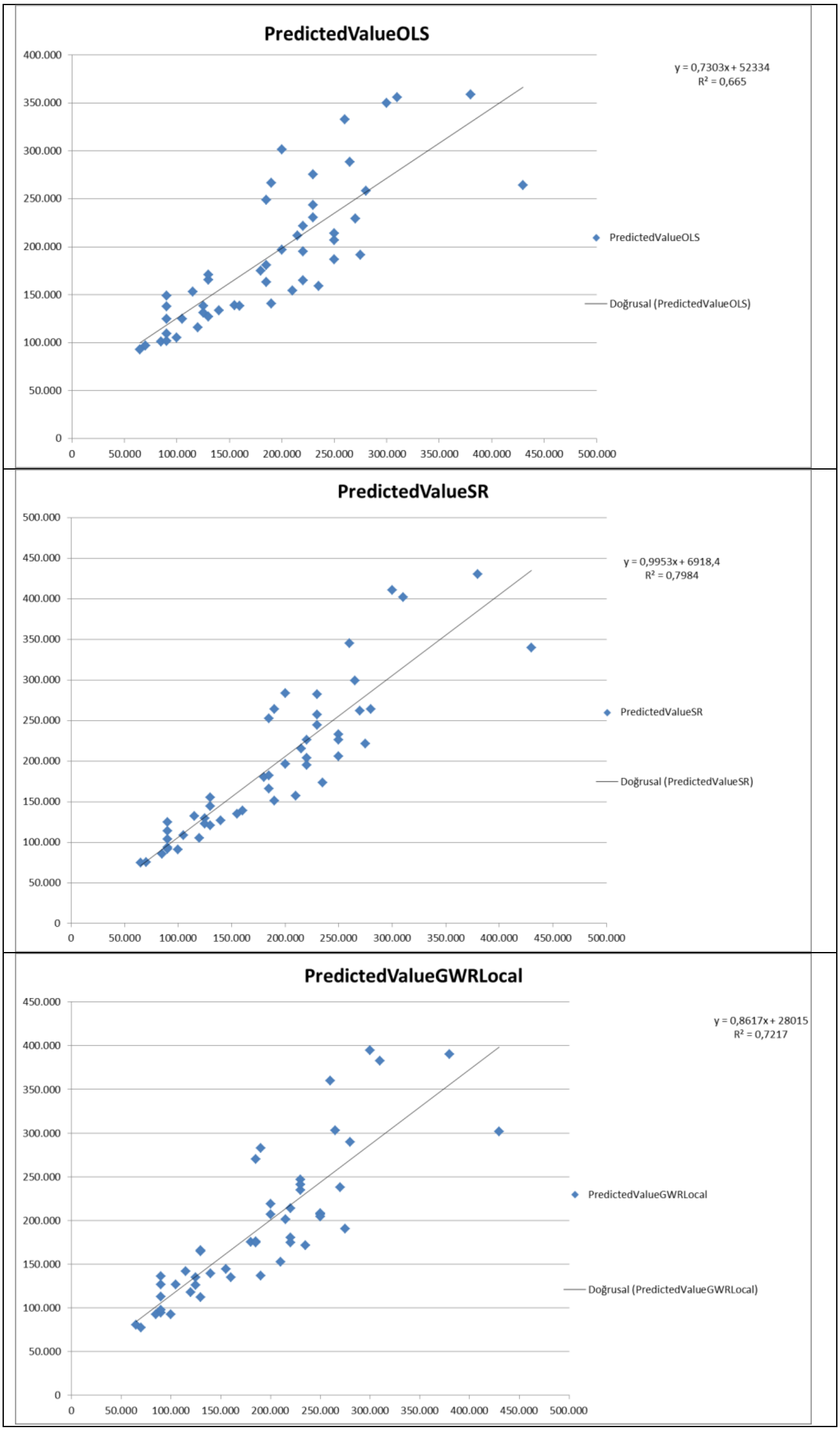


Figure 39: The result of predictions of OLS, SAR and GWR models separately for Çankaya

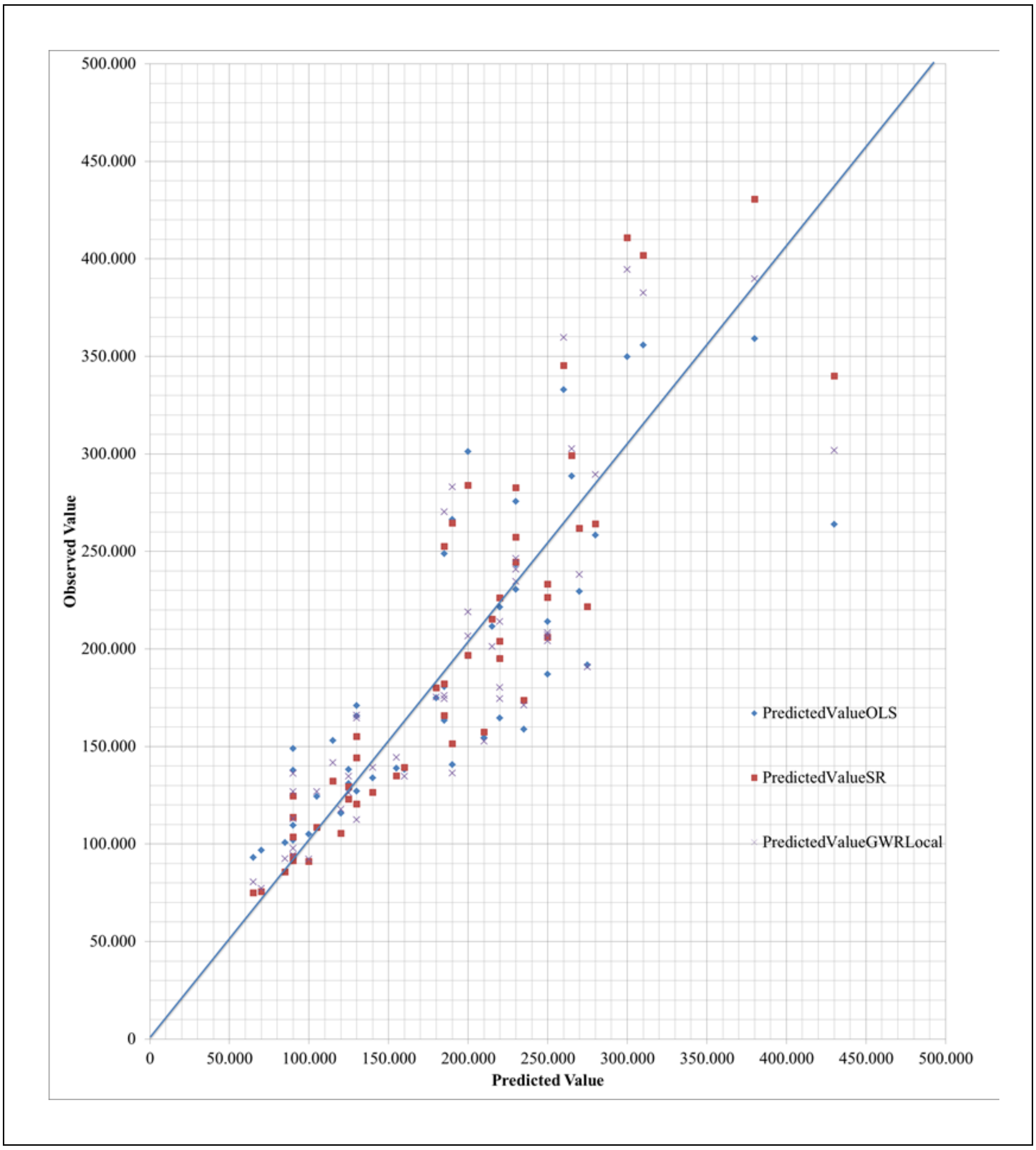


Figure 40: Comparison of predictions of OLS, SAR and GWR models for Çankaya

Contrary to the outputs of Çankaya, as can be seen in Figure 41 the highest R^2 value belongs to the GWR model (R^2=0.79) for Keçiören. The SAR model gives the second highest R^2 value (R^2=0.73).

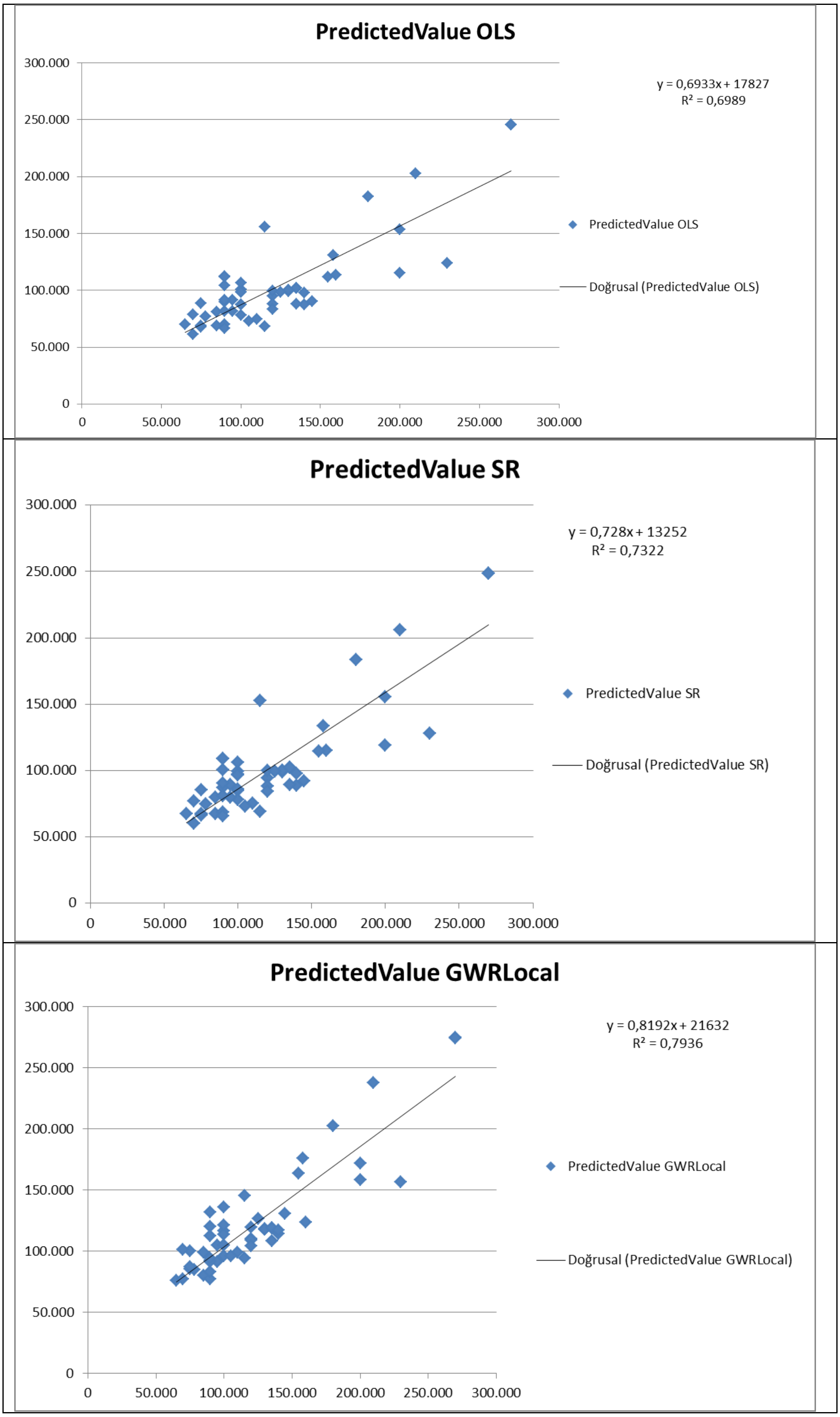


Figure 41: The result of predictions of OLS, SAR and GWR models separately for Keçiören

Figure 42 shows comparison of the predictions of models with the regression line. This graph shows that the GWR predictions are closer to the linear equation line (blue line) than the predictions of other models. In other words, the GWR model has the best performance to predict the house values in Keçiören.

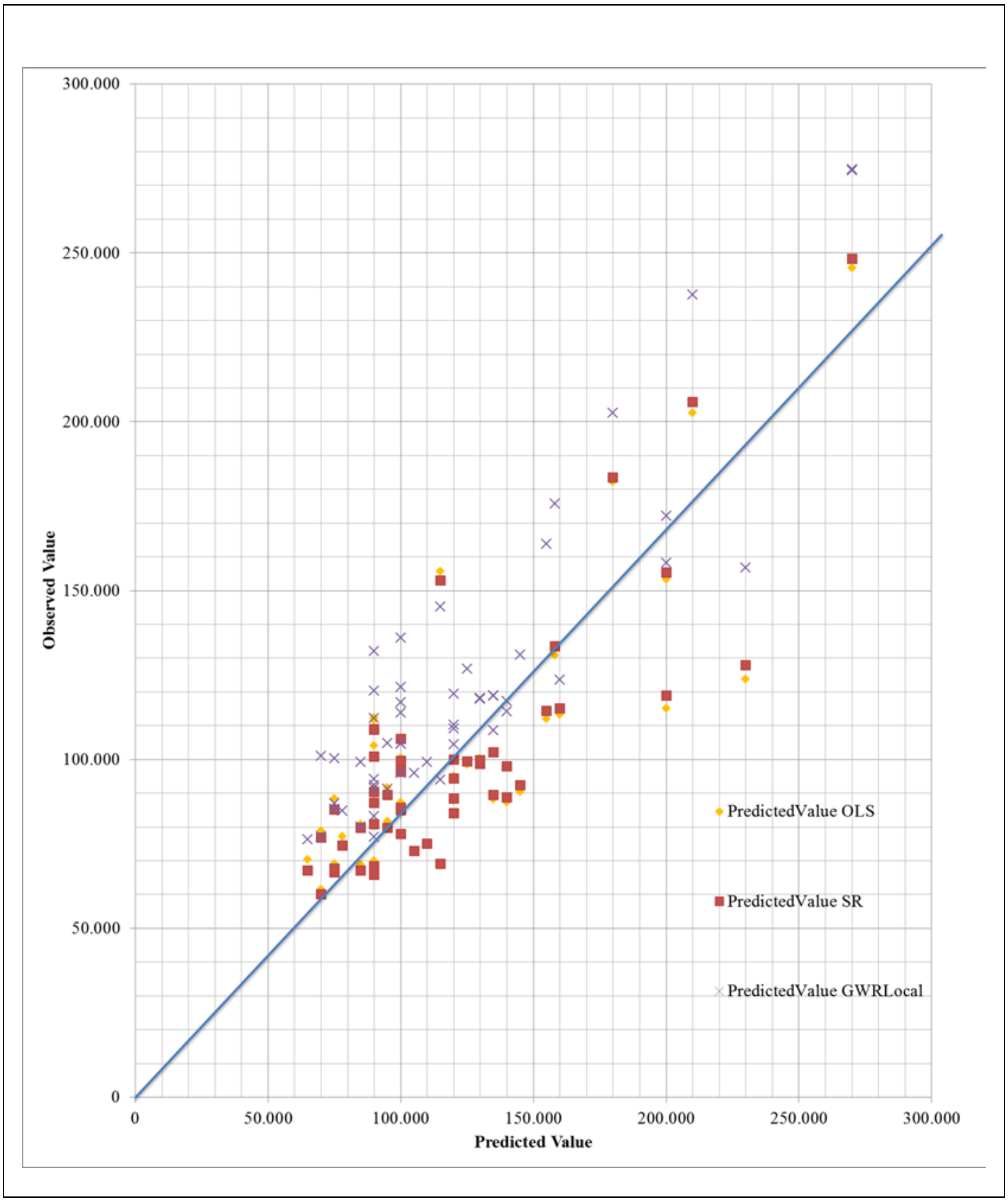


Figure 42: Comparison of predictions of OLS, SAR and GWR separately for Keçiören

Consequently, the incorporation of the spatial relationships has significantly improved the simple OLS model; the spatial relationships revealed by the spatial models above are significant.

The other way to test the model performance is to use mathematical metrics. In this study, the model performances were measured and compared based on the four most common measures of predictive accuracy, namely (Mean squared error), RMSE (Root mean squared error), MAE (Mean absolute error) and MAPE (Mean absolute percentage error). The model which has minimum MSE, RMSE, MAE, and MAPE values is the best one. Table 28 shows that the SAR models have the lowest error values. This means that the SAR models have the the best performance to estimate the housing values for Çankaya.

Table 28: Evaluation of the model performance based on four different kinds of errors for Çankaya

Model	MSE	RMSE	MAE	MAPE
PredictedValueOLS	2.196.005.657	46861.56	35128.30	19.85
PredictedValueSR	1.643.756.216	40543.26	28914.50	14.30
PredictedValueGWR	1.966.767.456	44348.25	32260.90	16.95

For Keçiören, as can be seen in Table 29 the GWR has the minimum error measures. This implies that the GWR is the best predicted method to estimate housing values in Keçiören.

Table 29: Evaluation of the model performance based on four different kinds of errors for Keçiören

Model	MSE	RMSE	MAE	MAPE
PredictedValueOLS	1.024.887.466	32013.86	24.619	19
PredictedValueSR	967.237.106	31100.44	24.170	19
PredictedValueGWR	437.645.384	20919.98	16.442	14

CHAPTER 5

RESULTS AND DISCUSSION

The OLS model itself gave poor results, apparently because it does not take into account the spatial dependency of the underlying variables. Therefore, several SAR models and the GWR approach were tested.

In this chapter, the implementation results of the proposed methodology are presented and discussed. In this study, the spatial and non-spatial statistical techniques were applied for the two largest districts in Ankara, Turkey. The findings that were revealed at the municipal level (Çankaya and Keçiören) were assessed. In the first step, a hedonic regression model was estimated by means of ordinary least squares. A correlation analysis was carried out using a large set of variables in order to determine the highly correlated variables. During the investigation of a consistent and unbiased global hedonic housing valuation model, many variable combinations were tested and some of them had to be disregarded. When testing the assumption of homoscedasticity, unequal variances were encountered. Therefore, a logarithmic transformation was applied to the dependent variable (house value) and some outliers were eliminated from two datasets. For this reason, the size of sample fell from 459 to 414 for Çankaya and from 522 to 502 for Keçiören. Consequently, the OLS model has been estimated with a semi-logarithmic specification. The model of Çankaya consists of three structural/housing characteristics and two accessibility characteristics. Furthermore, the model of Keçiören consists of three structural/housing characteristics, one accessibility characteristic and one social characteristic.

According to Figure 43, a value of 0,3629 for Moran's I indicates the existence of spatial autocorrelation in OLS residuals for Çankaya data set.

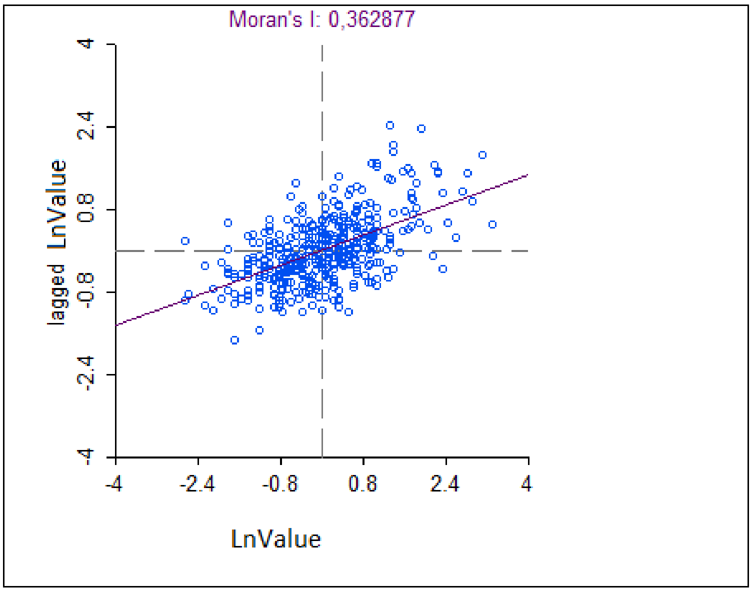


Figure 43: OLS Moran's I for Çankaya

Considering Figure 44, it can be concluded for Lagged Residuals that the Moran's I test statistic is 0.0703. This indicates that including the spatially lagged variable (W_LnValue) term in the model has minimized all spatial autocorrelation in the Çankaya data set.

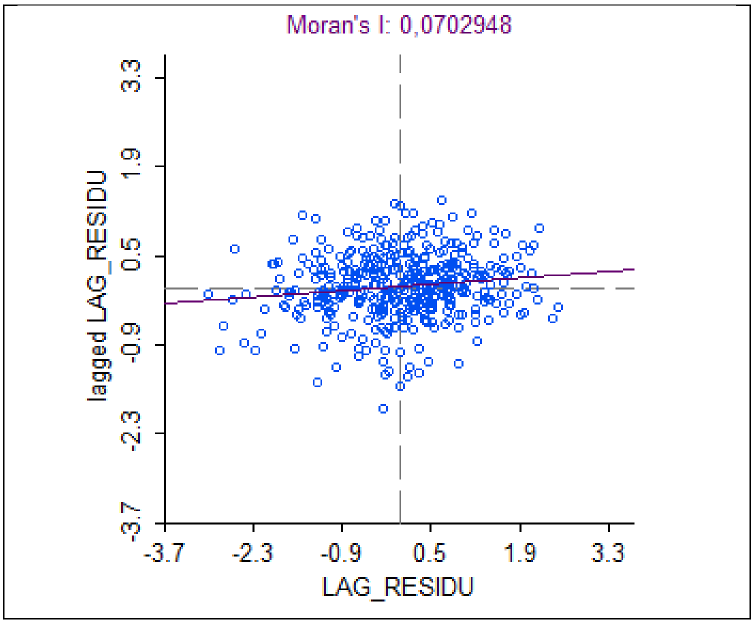


Figure 44: SAR Moran's I for Çankaya

As can be seen in Figure 45 Moran's I value for the OLS model is 0,3728. After spatial regression analysis (Figure 46) the Moran's I test statistic fell from 0,3728 to -0.0161, close to zero. This indicates that including the spatially autoregressive error term in the model has minimized all spatial autocorrelation in the Keçiören data set.

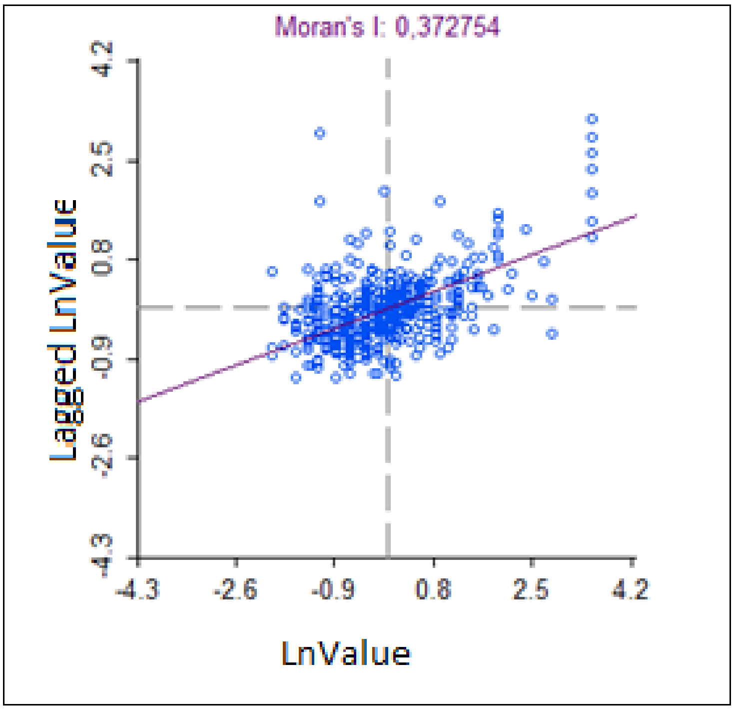


Figure 45: OLS Moran's I for Keçiören

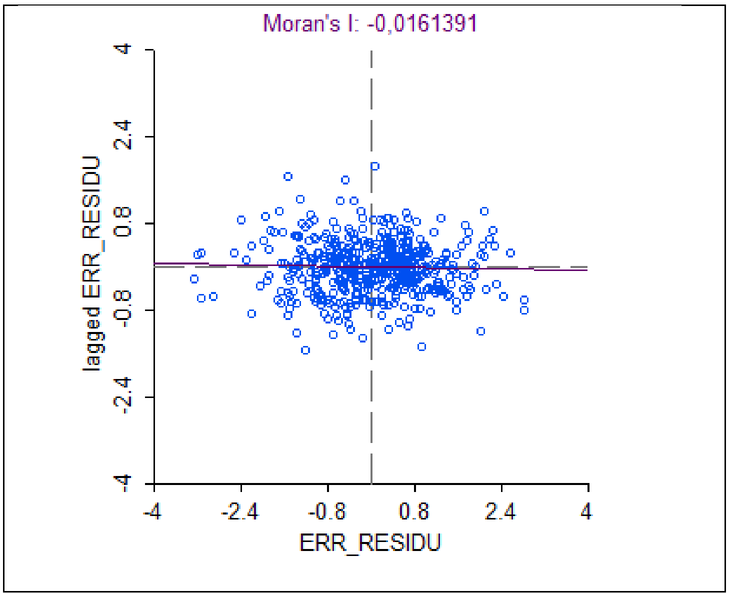


Figure 46: SAR Moran's I for Keçiören

After the OLS regression, the spatial regression analysis was performed to examine the spatial relationships of the explanatory variables determined by the OLS analysis. When the results of the OLS and the SAR models for both districts, which are given in Table 30 for Çankaya and Table 31 for Keçiören, are compared, it can be seen that the spatial models yield improvement to the original OLS models. In other words, the spatial regression improved fitting of the general model, as indicated in higher values of R^2 and log likelihood for both districts. The SARerr model R^2 (0.81) is higher than the SARlag model R^2 (0.80) and the SARerr model AIC (-79.51) is lower than the SARlag AIC (-72.39). Comparing the measures of models goodness of fit indicates that the spatial error model fits the data better for Çankaya. However, the SARerr model is insignificant at the 95% confidence interval. Therefore, the SARlag model was assumed to be the best fit model for the Çankaya dataset. The GWR model is the second best fit model for the Çankaya dataset.

Table 30: Evaluation of OLS, SAR and GWR based on model diagnostics for Çankaya

Çankaya Model	OLS	SARlag	GWR
R^2	0.760	0.80	0.80
Adj R^2	0.759		0.793
AICc	-14.03	**-72.39**	-55.88
Schwarz		-44.21	-31.62
Log likelihood	-28.31	43.19	101.29
Moran's I	0.36	0.07	
Intercept	11.16	7.57	
Property Type	0.205	0.16	
Security	0.46	0.41	
Area	0.009	0.008	
DistanceToSubway	-0.012	-0.00002	
DistanceToShoppingCenter	-0.086	-0.00006	
SARlag (Rho)		0.299	

In contrast to Çankaya, the GWR model is assumed to be the best fit model for Keçiören data set (Table 31). The SARerr model is the second best fit model for the Keçiören dataset.

Table 31: Evaluation of OLS, SAR and GWR based on model diagnostics for Keçiören

Keçiören Model	OLS	SARerr	GWR
R^2	0.78	0.82	0.84
Adj R^2	0.78		0.82
AICc	-170.99	-231.21	**-271.08**
Schwarz	-145.68	-205.89	-244.23
Log likelihood	91.49	121.60	326.76
Moran's I	0.37	-0.02	
Intercept	10.88	10.92	
Floor	0.04	0.047	
Security	0.48	0.386	
Area	0.009	0.008	
DistanceToShoppingCenter	-0.05	-0.00004	
Household	-0.008	-0.009	
SARerr (λ)		0.47	

The main outputs from the GWR are demonstrated in Table 32 and Table 33 for Çankaya and Keçiören, respectively. These consist of a set of local parameter estimates for each independent variable, namely minimum, maximum, median, range (between upper and lower quartiles), upper and lower quartiles and interquartile R of the both data sets. This is helpful to determine the degree of spatial non-stationarity in a relationship by comparing the range of the GWR parameter estimates with a confidence interval around the OLS estimate of the equivalent parameter. According to the results for Çankaya, Intercept changes from 10.63 to 11.35, Property Type from 0.10 to 0.55, Security from 0.35 to 0.64, Area from 0.008 to 0.01, Distance to Subway from -0.00004 to 0.00001 and Distance to Shopping Center from -0.000016 to 0.00002.

Table 32: GWR summary statistics for varying (Local) coefficients for Çankaya (Variables are significant on the 95% level)

Variable	Intercept	Property Type	Security	Area	Distance to Subway	Distance to Mall
Minimum	10.632314	0.099866	0.347722	0.008093	-0.000043	-0.000156
Maximum	11.352787	0.550409	0.642575	0.010187	0.00001	0.000019
Range	**0.720473**	**0.450543**	**0.294853**	**0.002094**	**0.000054**	**0.000175**
Lwr Quartile	11.090576	0.156858	0.44683	0.008701	-0.000023	-0.00011
Median	11.159886	0.186096	0.490057	0.009106	-0.000008	-0.000095
Upper Quartile	11.201441	0.279589	0.52708	0.009557	-0.000004	-0.000072
Interquartile R	0.110865	0.12273	0.08025	0.000857	0.000019	0.000038

Looking at the change of coefficients for Keçiören in Table 33 for Keçiören, the Intercept value changes from 10.53 to 11.04, Floor from 0.015 to 0.06, Security from 0.079 to 0.068, Area from 0.006 to 0.010, Distance to Shopping Center from -0.00006 to 0.00002 and Household from -0.026 to 0.006.

Table 33: GWR summary statistics for varying (Local) coefficients for Keçiören (Variables are significant on the 95% level)

Variable	Intercept	Floor	Security	Area	Distance to Mall	Household
Minimum	10.535858	0.015711	0.079245	0.006456	-0.000056	-0.02557
Maximum	11.040666	0.062643	0.676591	0.010237	0.000023	0.00598
Range	**0.504808**	**0.046933**	**0.597345**	**0.003781**	**0.000079**	**0.031551**
Lwr Quartile	10.693549	0.038444	0.209317	0.007913	-0.000041	-0.013001
Median	10.781972	0.044206	0.335605	0.008493	-0.000025	-0.002681
Upper Quartile	10.834391	0.049413	0.504893	0.009225	-0.000005	0.001096
Interquartile R	0.140842	0.010969	0.295575	0.001313	0.000036	0.014097

The comparison between Range (GWR) and 2 x S.E. (standard errors for the OLS) values for each parameter can give an idea about the variation across space. The tables including the OLS coefficients and standard errors that are required for comparison are shown below (Table 34 for Çankaya and Table 35 for Keçiören).

Table 34: OLS coefficients and standard errors for Çankaya

Variable	Estimate	Standard Error	2 x S.E.
Intercept	11.161343	0.051128	0.102256
Property Type	0.204332	0.042289	0.084578
Security	0.459808	0.06247	0.12494
Area	0.009037	0.000313	0.000626
DistanceToSubway	-0.000012	0.000004	0.000008
DistanceToMall	-0.000086	0.000013	0.000026

Table 35: OLS coefficients and standard errors for Keçiören

Variable	Estimate	Standard Error	2 x S.E.
Intercept	10.884241	0.060817	0.121634
Floor	0.040342	0.005163	0.010326
Security	0.486597	0.061534	0.123068
Area	0.008664	0.000352	0.000704
DistancetoMall	-0.000049	0.000005	0.00001
Household	-0.008232	0.002537	0.005074

If the Range value of a parameter in the GWR is greater than the value of 2 standard errors of the corresponding parameter in the OLS, this suggests that the relationship might be non-stationary.

In this context, according to Table 36 the interquartile ranges of the local estimates are much greater than corresponding 2 x S.E. of the global estimates, which indicates a non-stationary relationship for Çankaya data set.

Table 36: Comparison for non-stationarity in Çankaya data set

	Intercept	Property Type	Security	Area	DistanceToSubway	DistanceToMall
Range (GWR)	0.719262	0.449737	0.290826	0.002086	0.000053	0.000176
2 X S.E. (OLS)	0.102256	0.084578	0.12494	0.000626	0.000008	0.000026

Also, according to Table 37 the interquartile ranges of the local estimates are much greater than corresponding 2 x S.E. of the global estimates, which indicates also a non-stationary relationship for Keçiören.

Table 37: Comparison for non-stationarity in Keçiören data set

	Intercept	Floor	Security	Area	DistanceToMall	Household
Range (GWR)	0.504808	0.046933	0.597345	0.003781	0.000079	0.031551
2 X S.E. (OLS)	0.121634	0.010326	0.123068	0.000704	0.00001	0.005074

The importance of security, area and distance to shopping center parameters are common for both districts.

The impact of the theft incidents on housing values was investigated in this study for the first time in Turkey. It was shown that the impact of theft events on house values has a very limited extent. The main reason for this is that a potential dwelling buyer cannot access the information as to theft incidents officially.

Also, the influence of distance to certain places from houses and that of access to public transport were examined using spatial techniques. The effect of the ratio of population density to per building was searched.

The GWR provided facilities to see the effects of factors on housing valuation visually and how their parameters varied spatially in the study area. In other words, the GWR provided opportunities to examine the spatial structure of the non-stationary spatial processes. The results of the GWR in Table 32 for Çankaya and Table 33 for Keçiören show that the regression coefficients change considerably over the study area. This case indicates that the significance of the factors that have an effect on housing valuation changes depending on the location of the interaction of these factors. As a result, it can be said that the effect of the explanatory variables on house values differs from one neighborhood to another.

Evaluation of the Results of the Comparison (difference coefficients) Maps

The differences between the OLS and the GWR coefficient estimates were calculated by subtracting the absolute value of the GWR coefficient estimate from the absolute value of the OLS coefficient estimate. Likewise, the differences between the SAR and the GWR coefficient estimates were calculated by subtracting the absolute value of the GWR coefficient estimate from the absolute value of the SAR coefficient estimate. Positive values indicate that the OLS/SAR overestimates the effect of the variable compared to the GWR. In contrast, negative values refer to the underestimation by OLS/SAR. For each figure below, dark brown points illustrate high-valued coefficients (overestimations) and light yellow points demonstrate low-valued coefficients (underestimations). Also, the maps at the top display coefficients computed by the GWR, the maps show the difference between the OLS and the GWR coefficient estimates and the difference between the SAR and the GWR coefficient estimates.

For additional discussion, the differences between β coefficient estimates of the OLS/SAR and the GWR are demonstrated in Figures 47-52 for Çankaya and in Figures 53-58 for Keçiören. The aim of these comparisons is to visually observe to what extent the global effects (OLS/SAR) of the determinants deviate from the local effects (GWR) (Keser et al., 2012). Each of the independent variables below was evaluated separately. The neighborhoods with high-valued coefficients and those with low-valued coefficients were listed for each explanatory variable.

According to Figure 47, the neighborhoods with high-valued Intercept coefficients estimated by the GWR are Aziziye, Çankaya, Güzeltepe, Yıldızevler, Şehit Cevdet Özdemir, Güvenevler, İlkadım, and Naci Çakır. The neighborhoods with low-valued Intercept coefficients estimated by the GWR are Şehit Cengiz Karaca, Osman Temiz, Huzur, Karapınar, Malazgirt, Akpınar, Mürsel Uluç and Keklikpınarı. The neighborhoods with high-valued Intercept coefficients estimated by the OLS/SAR are Gökkuşağı, Huzur, Karapınar, Malazgirt, Mürsel Uluç, Akpınar and Keklikpınarı. Finally, the neighborhoods with low-valued Intercept coefficients estimated by the OLS/SAR are İlkadım, Güzeltepe, Şehit Cevdet Özdemir, Yıldızevler, Hilal, Sancak and Naci Çakır.

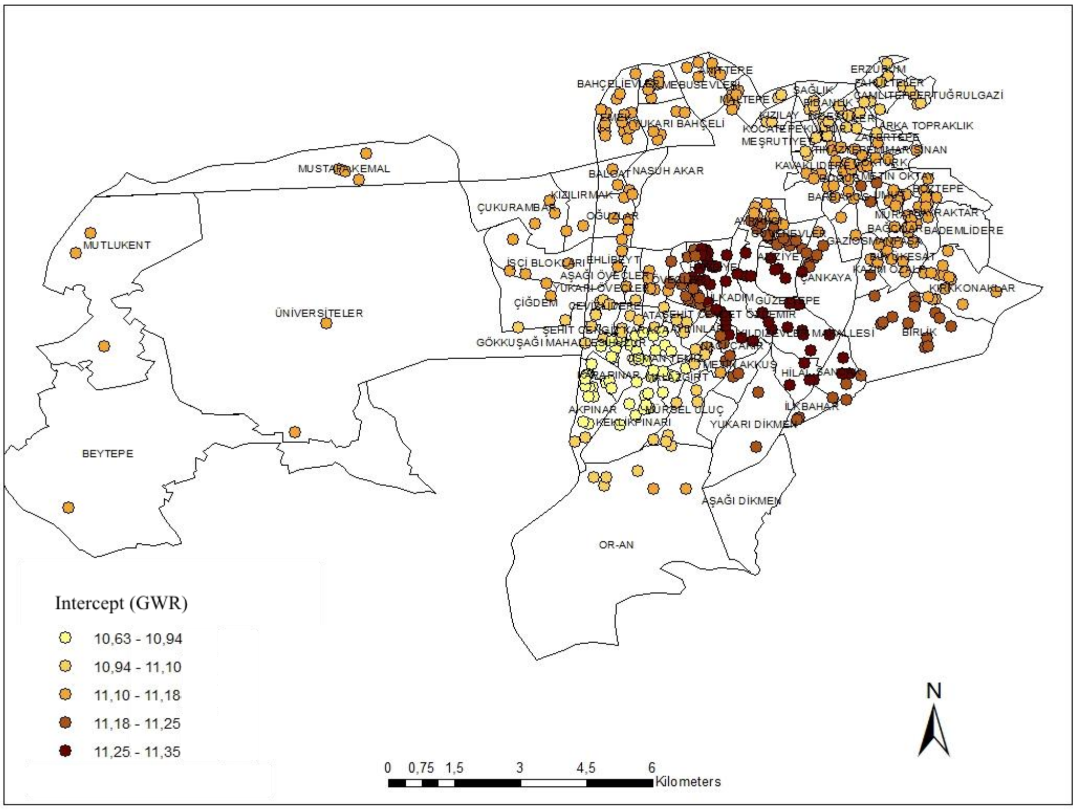
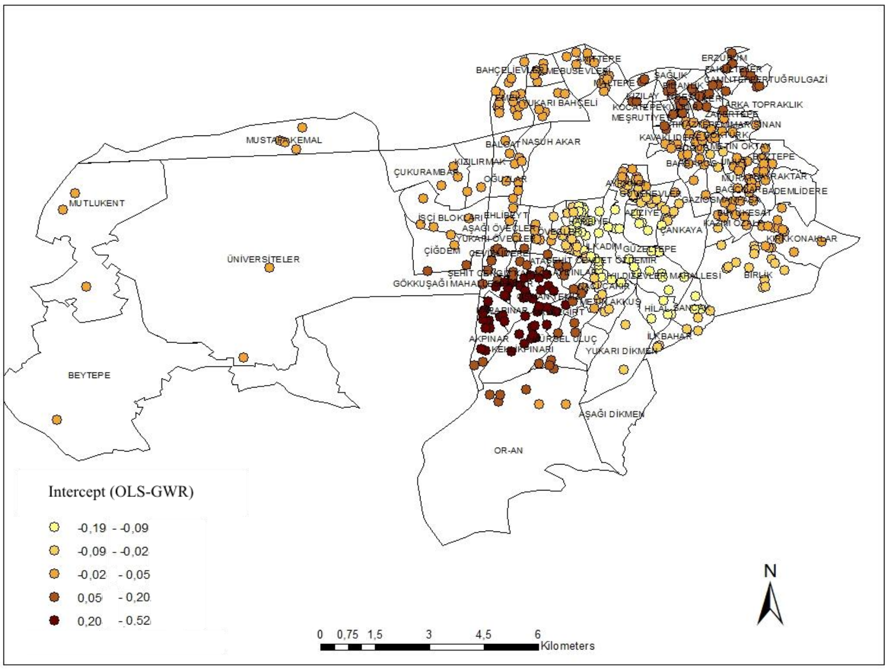
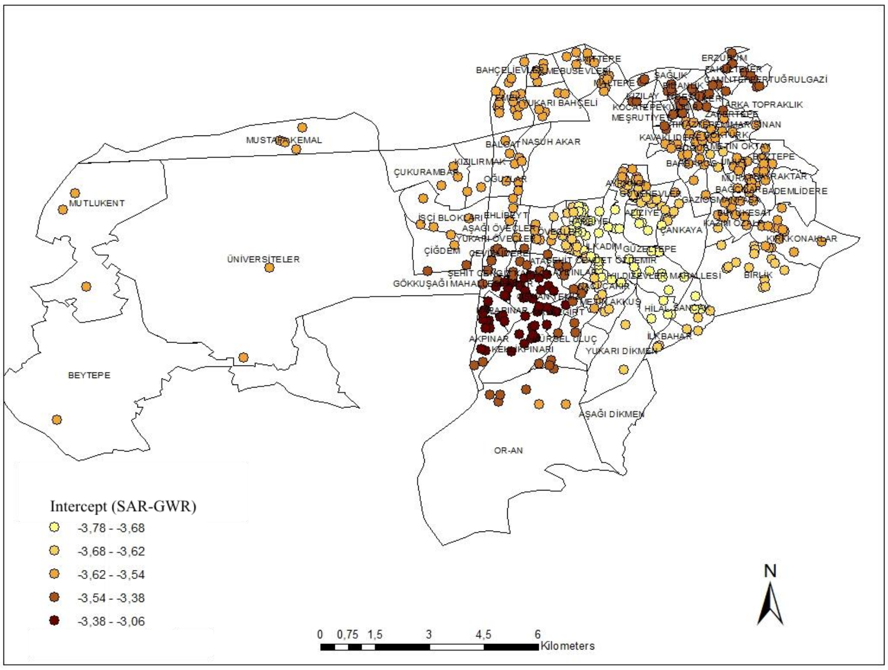


Figure 47: The results of Intercept coefficients (GWR) at the 95% confidence level for Çankaya



(a)



(b)

Figure 47 (continued): a) Differences of Intercept coefficients OLS-GWR and b) Differences of Intercept coefficients SAR-GWR at the 95% confidence level for Çankaya

According to Figure 48, the neighborhoods with high-valued Property Type coefficients estimated by the GWR are Fidanlık, Öncebeci, İleri, Arka Topraklık, Zafertepe, Mimar Sinan, Seyranbağları, Meşrutiyet, Tınaztepe, Kavaklıdere, Esat, Doğuş, Metin Oktay, Küçükesat, Barbaros, Göktürk and İncesu. The neighborhoods with low-valued Property Type coefficients estimated by the GWR are Emek, Yukarı Bahçeli, Balgat, Nasuh Akar, Kızılırmak, Çukurambar, Oğuzlar, İşçi Blokları, Ehlibeyt, Aşağı Öveçler, Yukarı Öveçler, Çiğdem, Cevizlidere, Şehit Cengiz Karaca, Ata, Gökkuşağı, Huzur and Osman Temiz. The neighborhoods with high-valued Property Type coefficients estimated by the OLS/SAR are Emek, Yukarı Bahçeli, Balgat, Nasuh Akar, Kızılırmak, Çukurambar, Oğuzlar, İşçi Blokları, Ehlibeyt, Yukarı Öveçler, Çiğdem, Cevizlidere, Şehit Cengiz Karaca, Ata, Karapınar, Gökkuşağı, Huzur and Osman Temiz. Lastly, the neighborhoods with low-valued Property Type coefficients estimated by the OLS/SAR are Fidanlık, Öncebeci, İleri, Arka Topraklık, Zafertepe, Mimar Sinan, Seyranbağları, Tınaztepe, Kavaklıdere, Esat, Doğuş, Metin Oktay and Küçükesat.

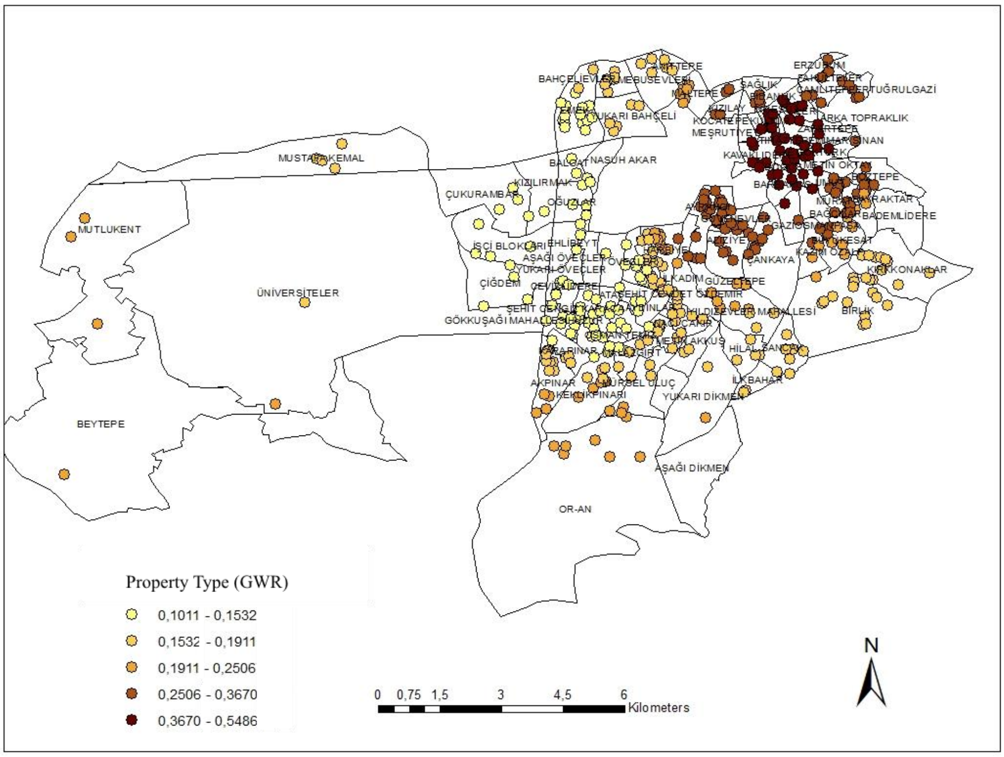
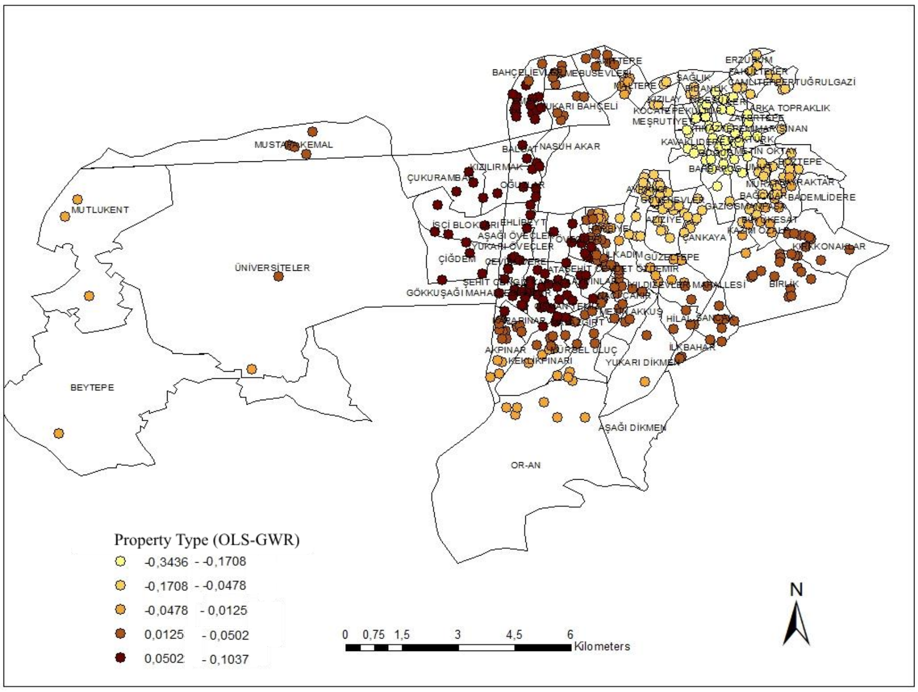
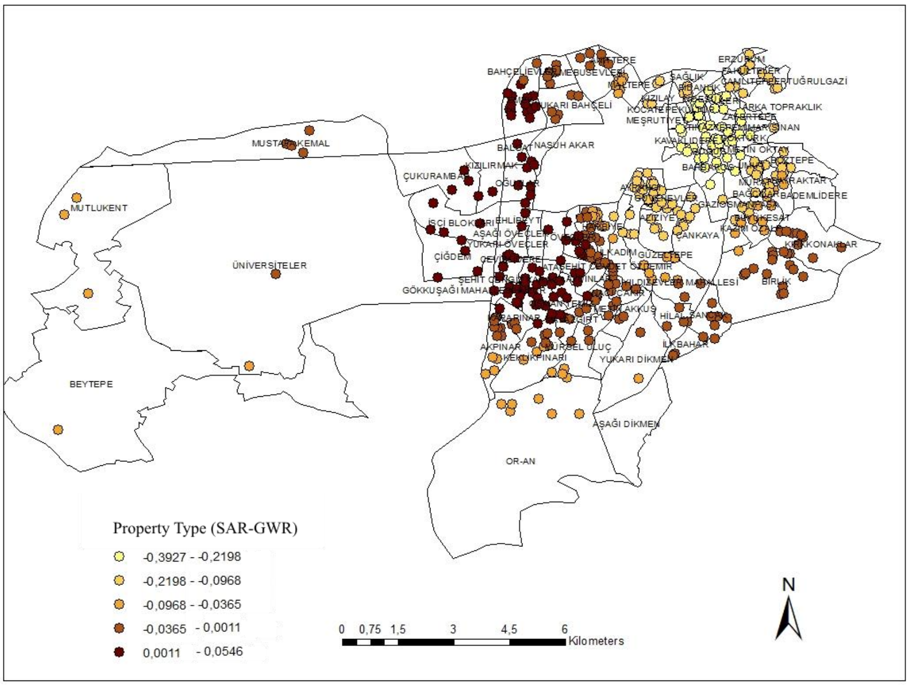


Figure 48: The results of Property Type coefficients (GWR) at the 95% confidence level for Çankaya



(a)



(b)

Figure 48 (continued): a) Differences of Property Type coefficients OLS-GWR and b) Differences of Property Type coefficients SAR-GWR at the 95% confidence level for Çankaya

According to Figure 49, the neighborhoods with high-valued Security coefficients estimated by the GWR are Metin Oktay, Büyükesat, Bayraktar, Bağcılar, Bademlidere, Kazım Özalp, Kırıkkonaklar and Birlik. The neighborhoods with low-valued Security coefficients estimated by the GWR are Meşrutiyet, Seyranbağları, Tınaztepe, Göktürk, Kavaklıdere, Çiğdem, Gökkuşağı, Huzur, Karapınar, Akpınar, Keklikpınarı, Mürsel Uluç, Oran and Osman Temiz. The neighborhoods with high-valued Security coefficients estimated by the OLS/SAR are Meşrutiyet, Seyranbağları, Tınaztepe, Göktürk, Kavaklıdere, Çiğdem, Gökkuşağı, Huzur, Karapınar, Akpınar, Keklikpınarı, Mürsel Uluç and Oran. The neighborhoods with low-valued Security coefficients estimated by the OLS/SAR are Boztepe, Murat, Bayraktar, Bağcılar, Bademlidere, Büyükesat, Kazım Özalp and Kırkkonaklar.

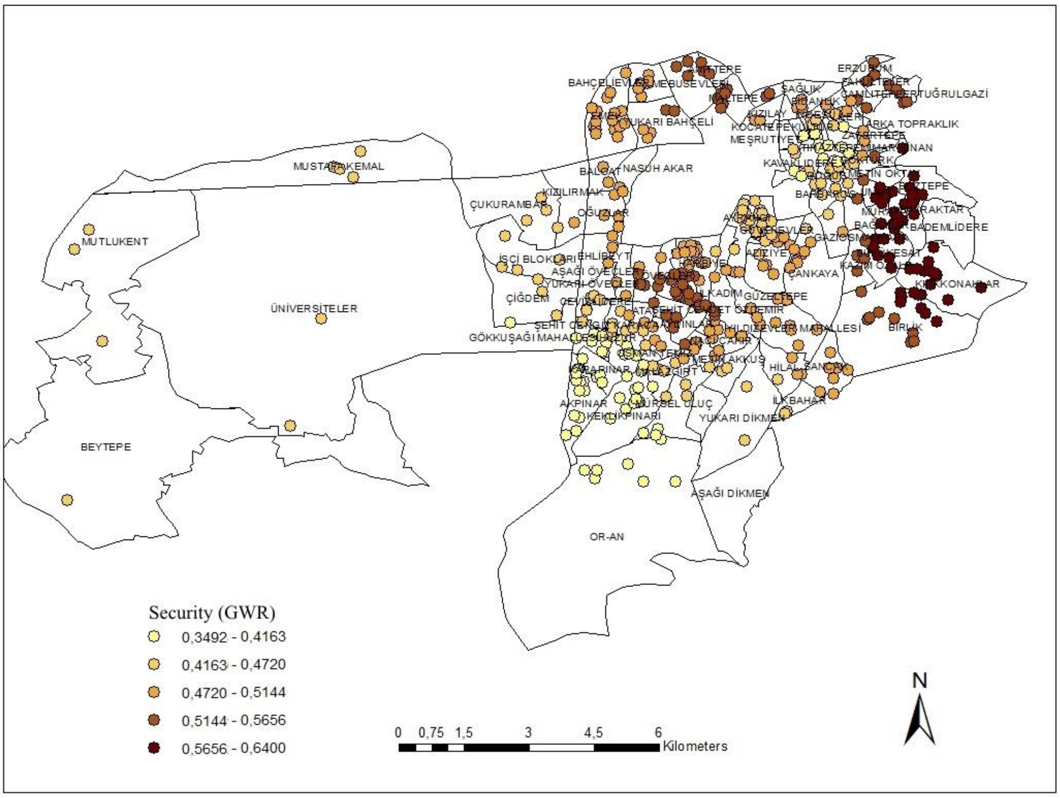
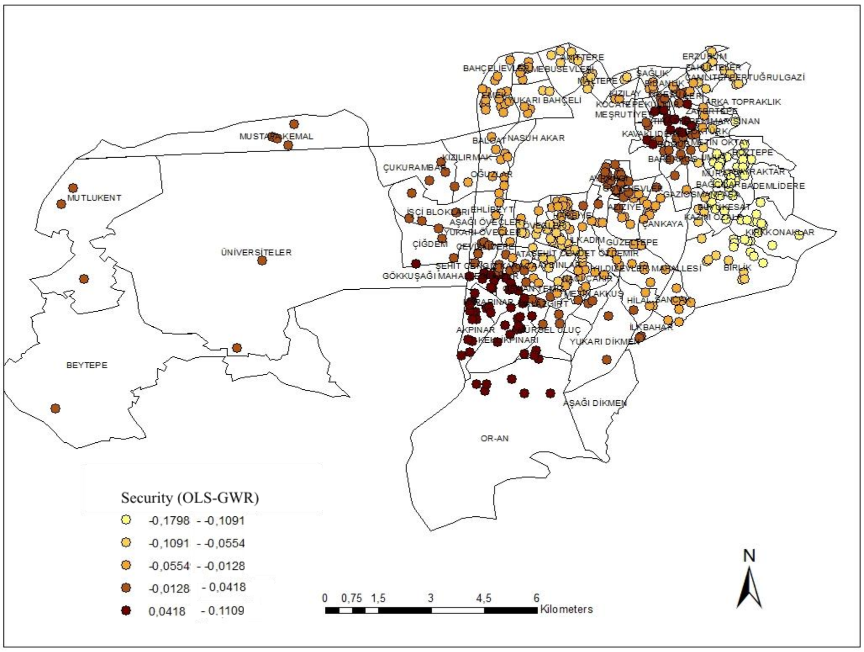
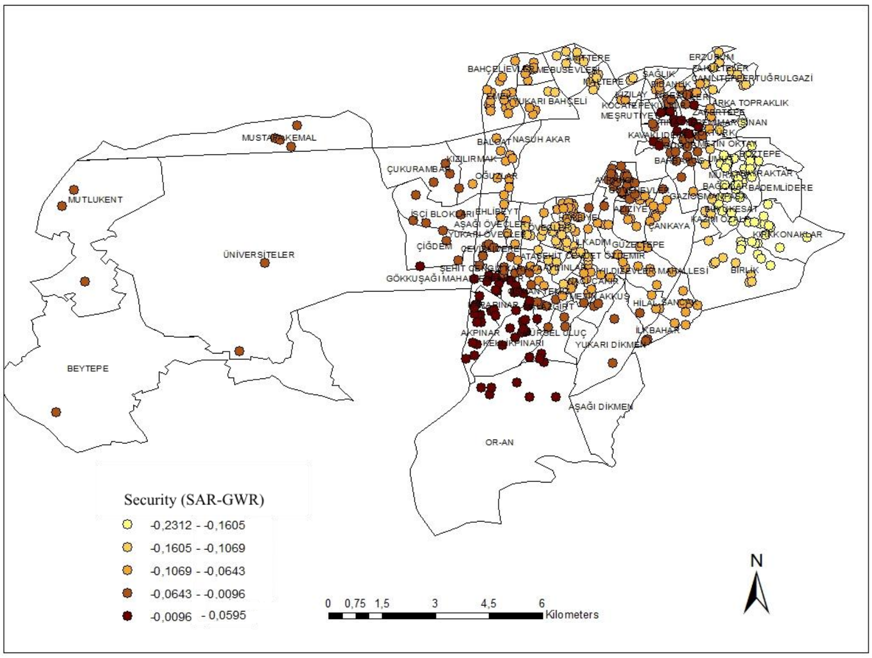


Figure 49: The results of Security coefficients (GWR) at the 95% confidence level for Çankaya



(a)



(b)

Figure 49 (continued): a) Differences of Security coefficients OLS-GWR and b) Differences Security coefficients SAR-GWR at the 95% confidence level for Çankaya

According to Figure 50, the neighborhoods with high-valued Area coefficients estimated by the GWR are Fidanlık, Öncebeci, İleri, İncesu, Kültür, İşçi Blokları, Arka Topraklık, Ehlibeyt, Çiğdem, Yukarı Öveçler, Cevizlidere, Şehit Cengiz Karaca, Gökkuşağı, Karapınar, Akpınar and Keklikpınarı. The neighborhoods with low-valued Area coefficients estimated by the GWR are Çankaya, Güzeltepe, Yıldızevler, Hilal, Sancak, Birlik, Kırkkonaklar, Kazım Özalp, Büyükesat and Gaziosmanpaşa. The neighborhoods with high-valued Area coefficients estimated by the OLS/SAR are Murat, Bayraktar, Bağcılar, Bademlidere, Çankaya, Güzeltepe, Yıldızevler, Sancak, Kırkkonaklar, Kazım Özalp, Büyükesat and Gaziosmanpaşa. The neighborhoods with low-valued Area coefficients estimated by the OLS/SAR are Çamlıdere, Fidanlık, Öncebeci, İleri, İncesu, Arka Topraklık, Zafertepe, Kültür, İşçi Blokları, Çiğdem, Yukarı Öveçler, Cevizlidere, Şehit Cengiz Karaca, Gökkuşağı, Karapınar, Akpınar and Keklikpınarı and Mürsel Uluç.

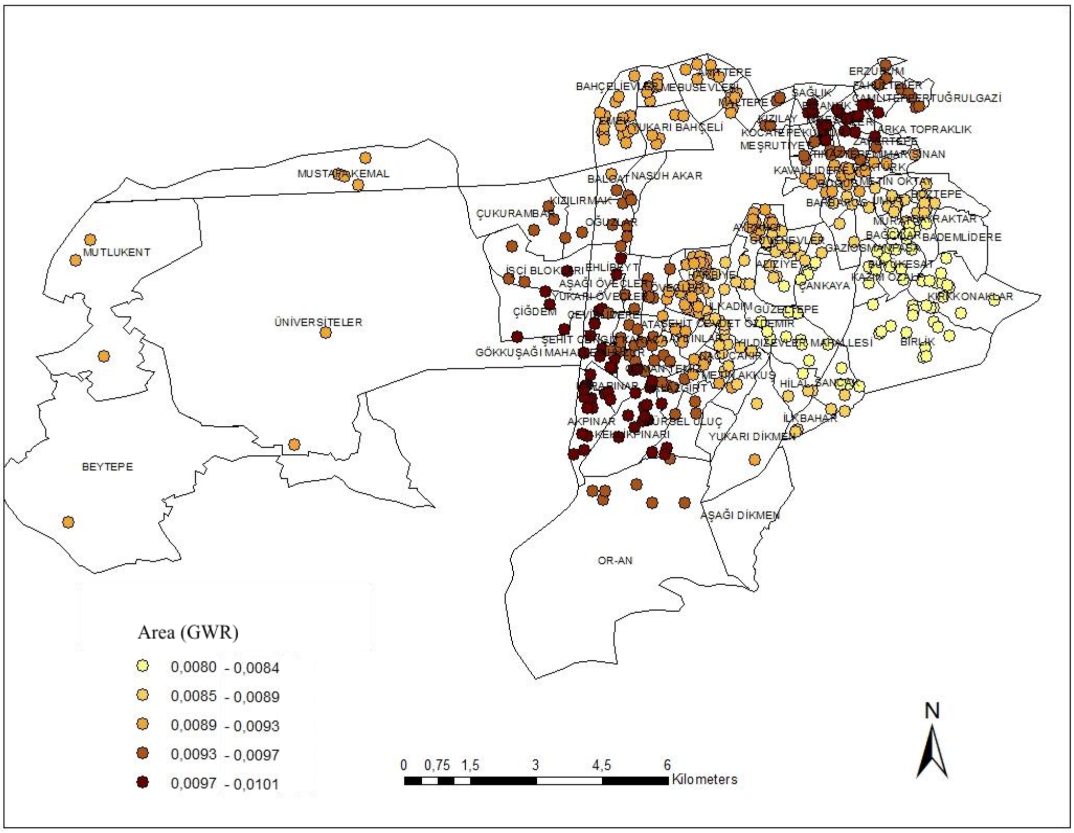
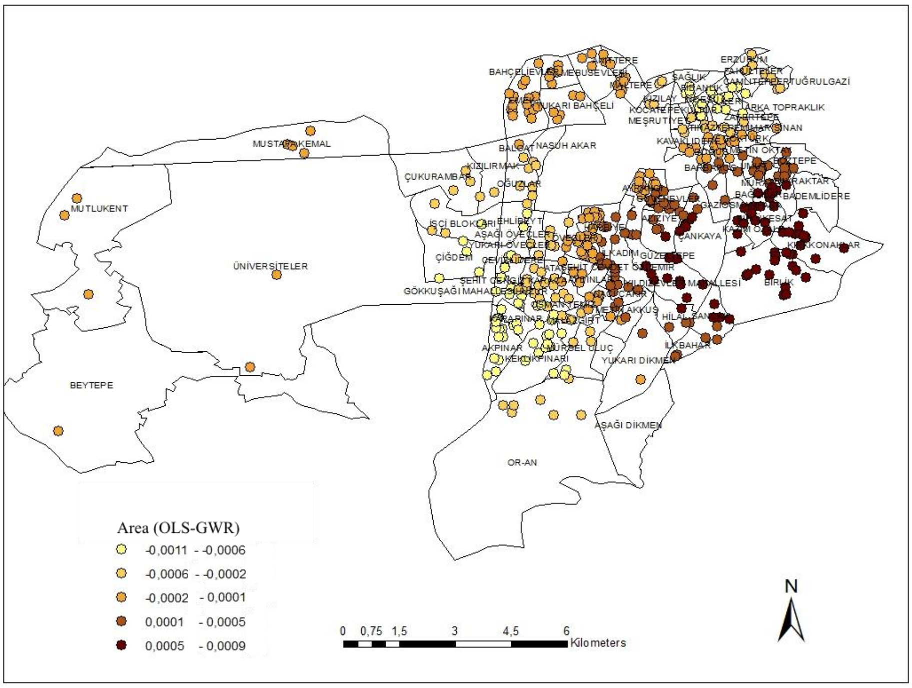
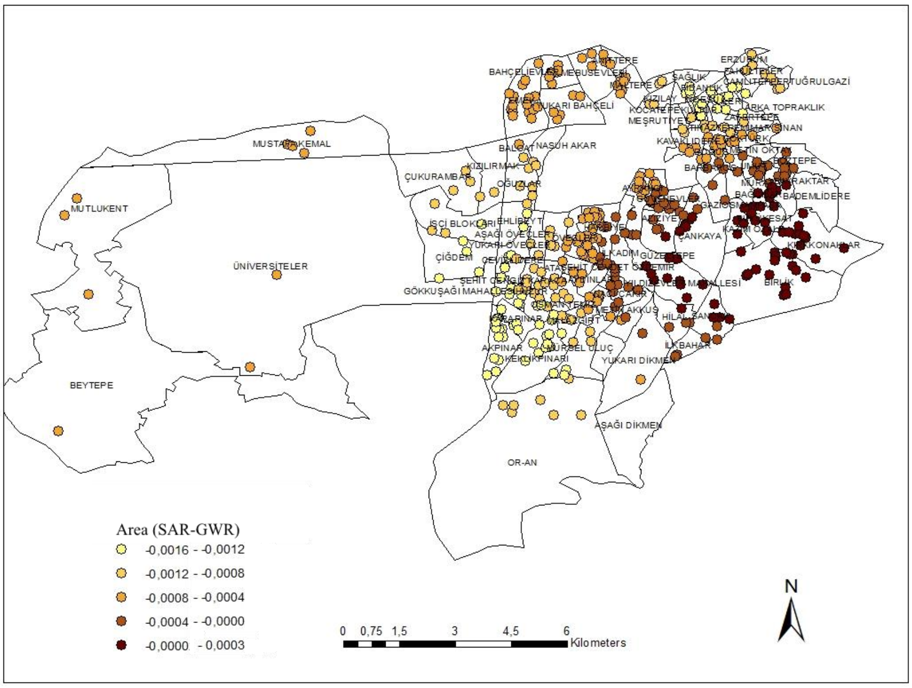


Figure 50: The results of Area coefficients (GWR) at the 95% confidence level for Çankaya



(a)



(b)

Figure 50 (continued): a) Differences of Area coefficients OLS-GWR and b) Differences of Area coefficients SAR-GWR at the 95% confidence level for Çankaya

According to Figure 51, the neighborhoods with high-valued DistanceToSubway coefficients estimated by the GWR are Beytepe, Mutlukent, Üniversiteler, Çiğdem, Aşağı Öveçler, Harbiye and Ayrancı. The neighborhoods with low-valued DistanceToSubway coefficients estimated by the GWR are İleri, Arka Topraklık, Zafertepe, Mimar Sinan, Göktürk and Mürsel Uluç. The neighborhoods with high-valued and low-valued DistanceToSubway coefficients estimated by the OLS/SAR are the same as the GWR.

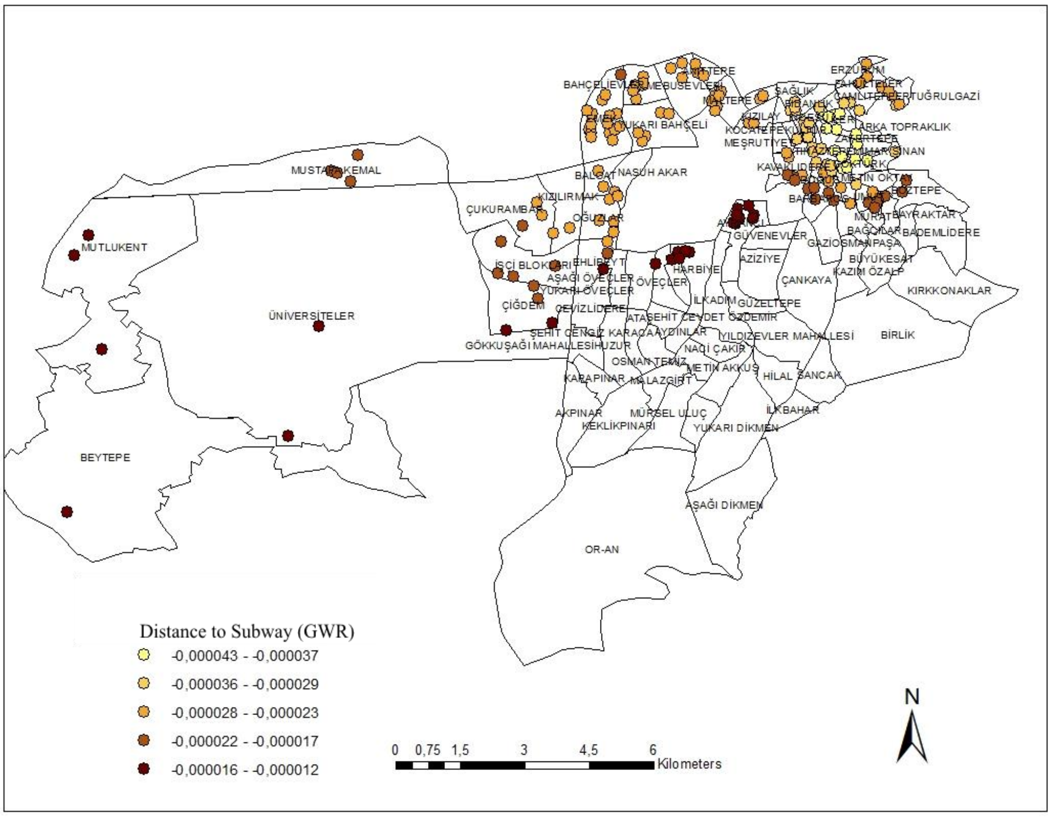
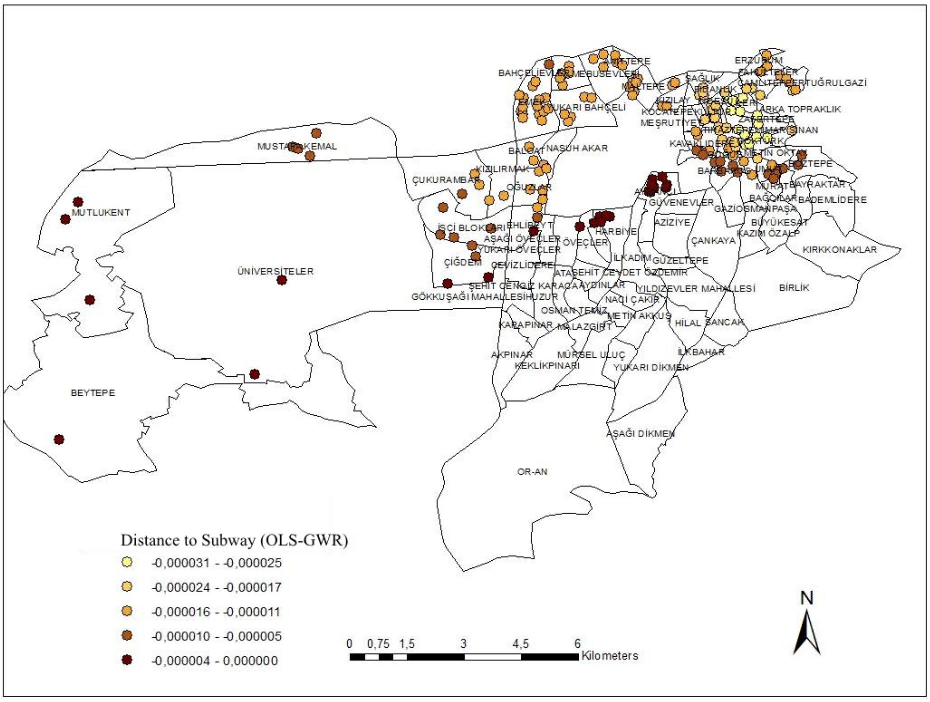
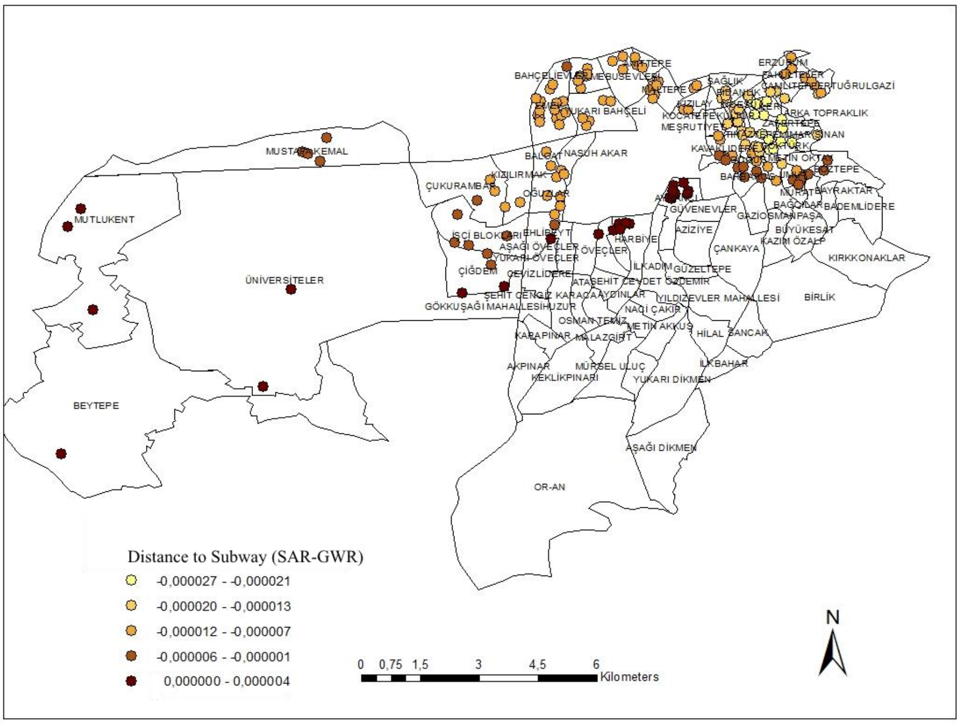


Figure 51: The results of DistanceToSubway coefficients (GWR) at the 95% confidence level for Çankaya



(a)



(b)

Figure 51 (continued): a) Differences of DistanceToSubway coefficients OLS-GWR and b) DistanceToSubway coefficients SAR-GWR at the 95% confidence level for Çankaya

According to Figure 52, the neighborhoods with high-valued Distance to Shopping Center coefficients estimated by the GWR are Bahçelievler, Anıttepe, Mebusevler, Emek, Yukarıbahçeli, Kızılay, Maltepe, Balgat, Gökkuşağı, Huzur, Osman Temiz, Karapınar, Malazgirt and Mürsel Uluç. The neighborhoods with low-valued Distance to Shopping Center coefficients estimated by the GWR are Harbiye, Sokullu Mehmet Paşa, Şehit Cevdet Özdemir, İlkadım, Naci Çakır and Yıldızevler. The neighborhoods with high-valued and low-valued Distance to Shopping Center coefficients estimated by the OLS/SAR are the same as the GWR.

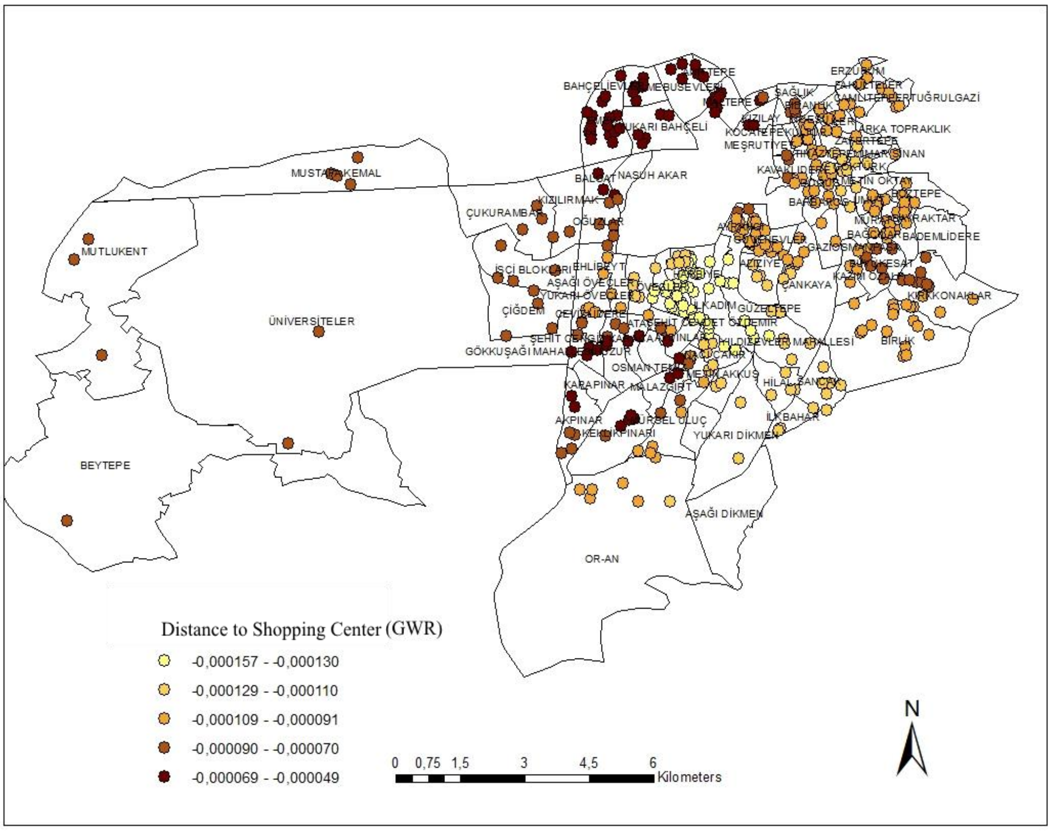
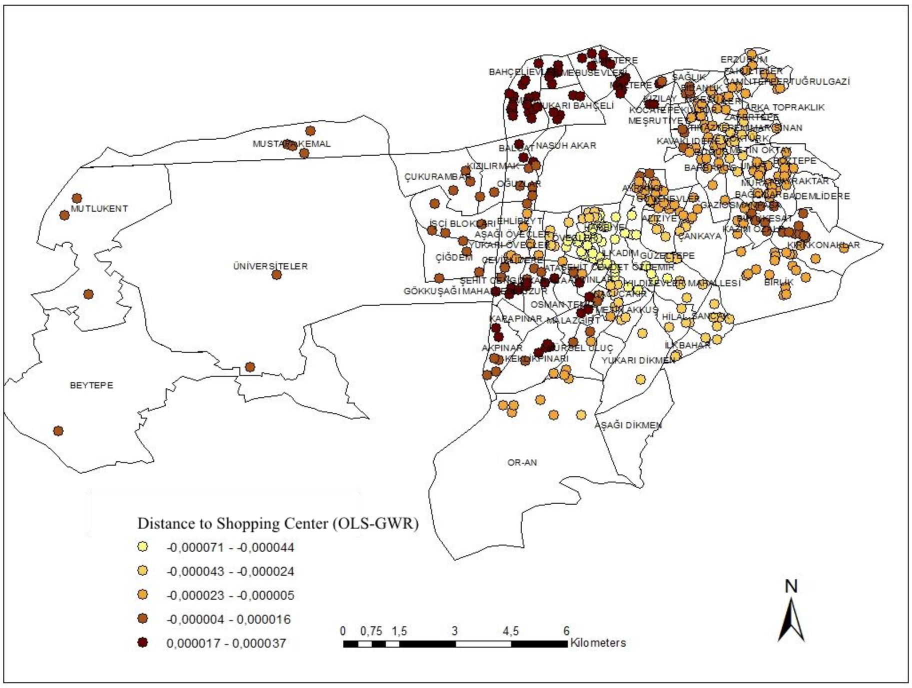
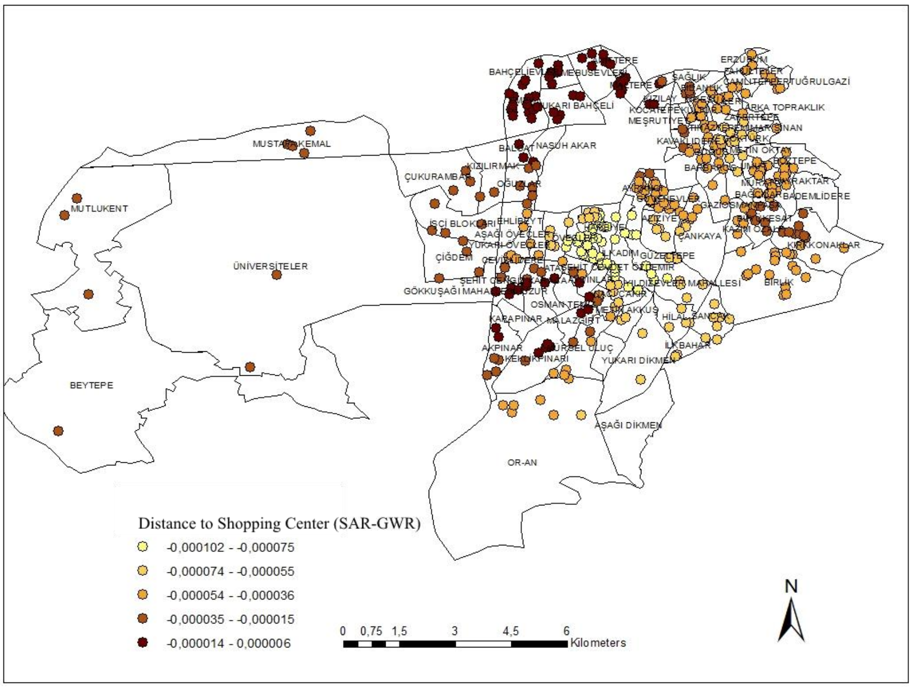


Figure 52: The results of DistanceToShoppingCenter coefficients (GWR) at the 95% confidence level for Çankaya



(a)



(b)

Figure 52 (continued): a) Differences of DistanceToMall coefficients OLS-GWR and b) DistanceToMall coefficients SAR-GWR at the 95% confidence level for Çankaya

According to Figure 53, the neighborhoods with high-valued Intercept coefficients estimated by the GWR are Hisar, Karakaya, Karşıyaka, Kafkas, Bağlarbaşı, Kamil Ocak, Şefkat, Hasköy, and Kavacık Subayevleri. The neighborhoods with low-valued Intercept coefficients estimated by the GWR are Osmangazi, Ufuktepe, Bademlik, Köşk, Adnan Menderes, Şenlik, Pınarbaşı, Atapark and Kuşcağız. The neighborhoods with high-valued Intercept coefficients estimated by OLS/SAR are Osmangazi, Ufuktepe, Bademlik, Köşk, Adnan Menderes, Şenlik, Pınarbaşı, Atapark, Kuşcağız, Yakacık and Tepebaşı. The neighborhoods with low-valued Intercept coefficients estimated by the OLS/SAR are Hisar, Karakaya, Karşıyaka, Kafkas, Bağlarbaşı, Kamil Ocak, Şefkat, Hasköy, Kavacık Subayevleri and Çaldıran.

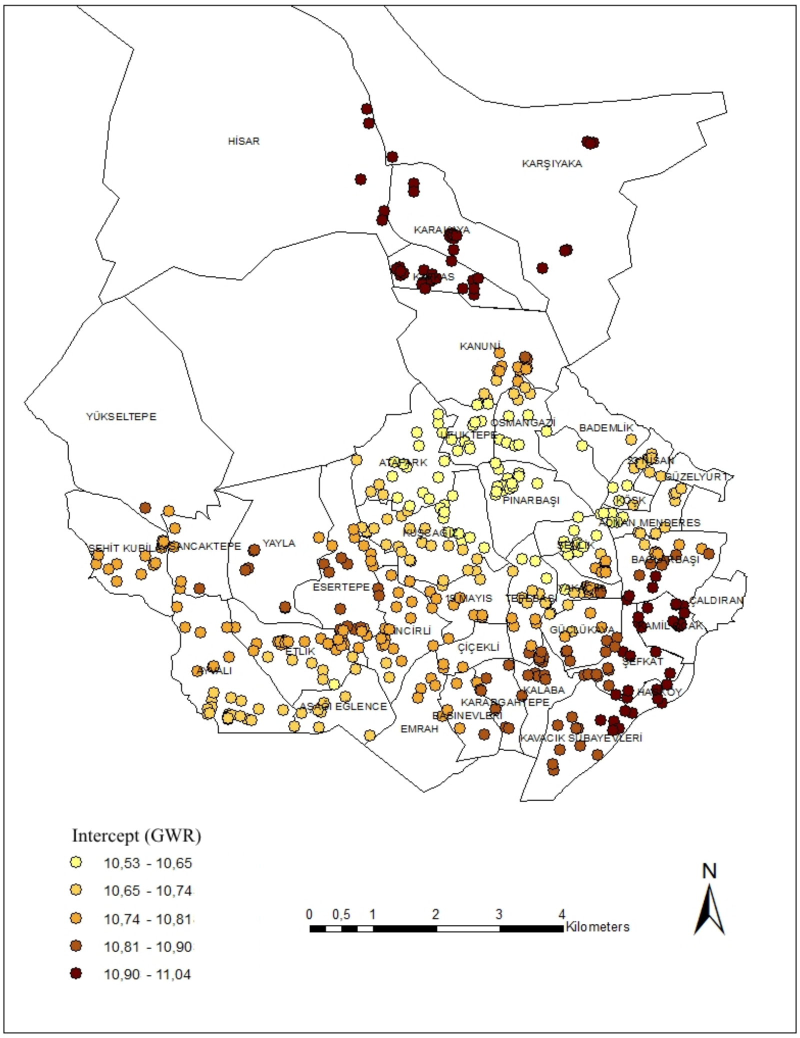
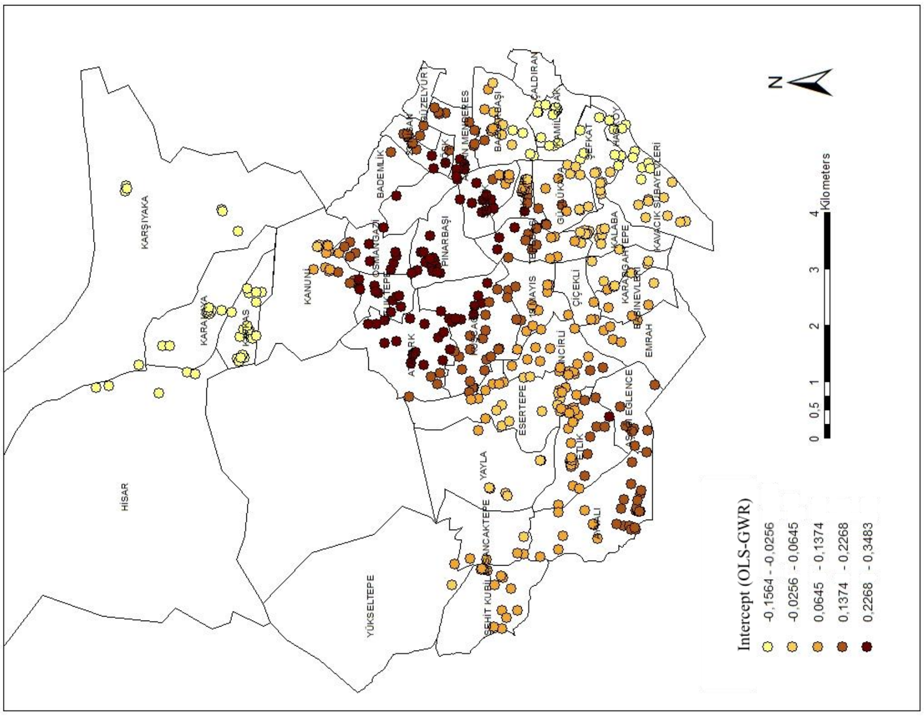
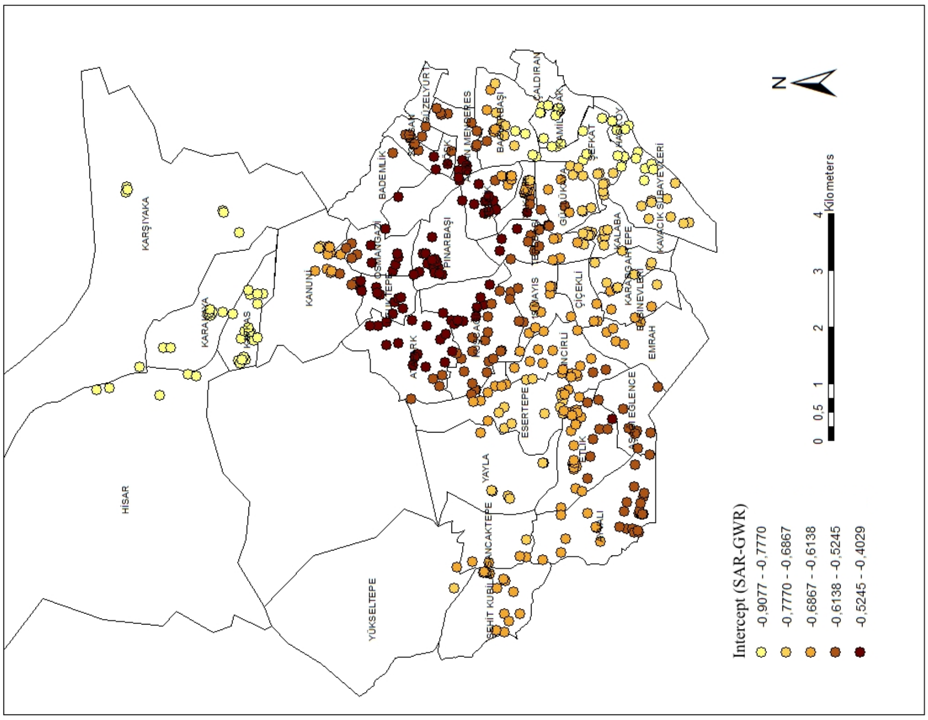


Figure 53: The results of Intercept coefficients (GWR) at the 95% confidence level for Keçiören



(a)



(b)

Figure 53 (continued): a) Differences of Intercept coefficients OLS-GWR and b) Differences of Intercept coefficients SAR-GWR at the 95% confidence level for Keçiören

According to Figure 54, the neighborhoods with high-valued Floor coefficients estimated by the GWR are Osmangazi, Bademlik, Güçlükaya, Şefkat, Hasköy and Kalaba. The neighborhoods with low-valued Floor coefficients estimated by the GWR are Hisar, Karşıyaka, Karakaya, Kafkas, Kanuni, Bağlarbaşı, Şenlik and Yakacık. The neighborhoods with high-valued Floor coefficients estimated by the OLS/SAR are Hisar, Karşıyaka, Karakaya, Kafkas, Kanuni, Şenlik and Yakacık. The neighborhoods with low-valued Floor coefficients estimated by the OLS/SAR are Güçlükaya, Şefkat, Hasköy, Kavacık Subayevleri and Kalaba.

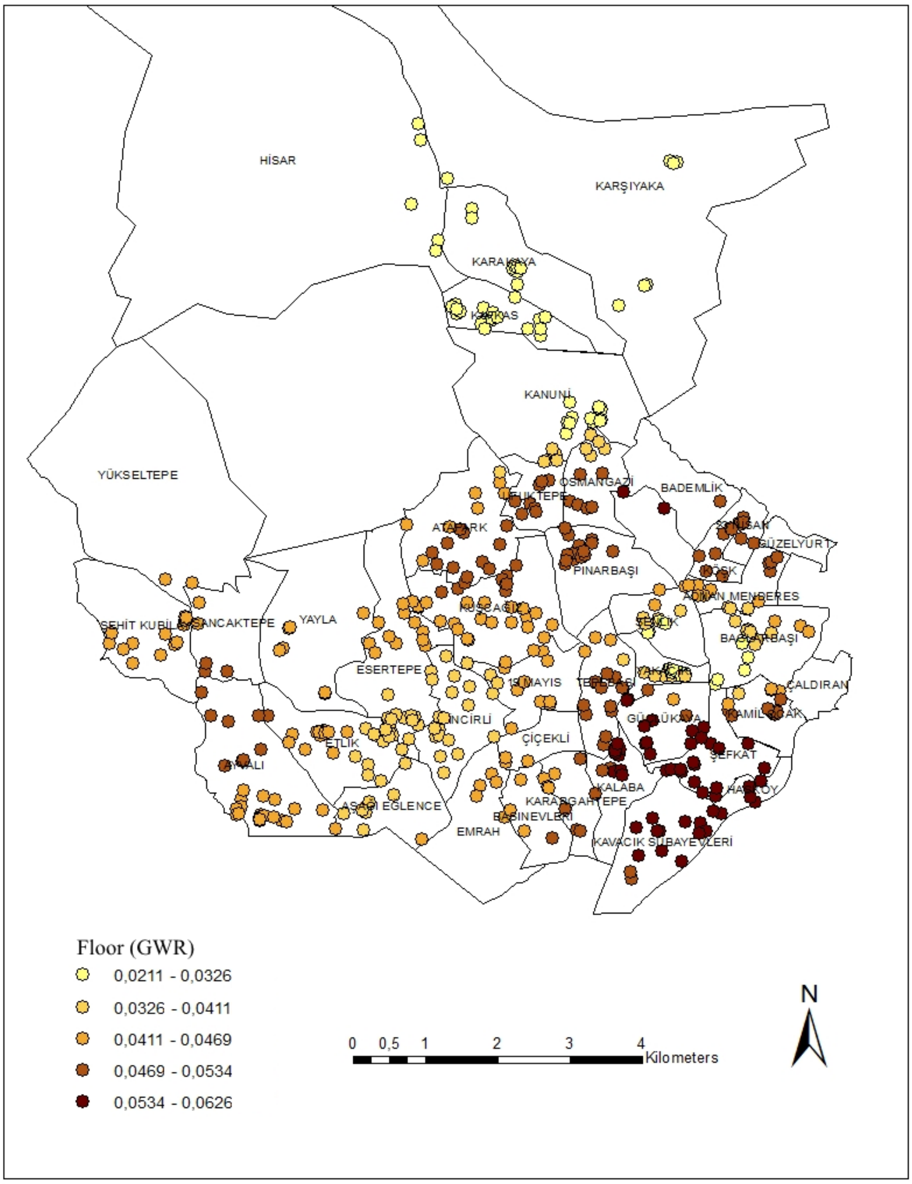
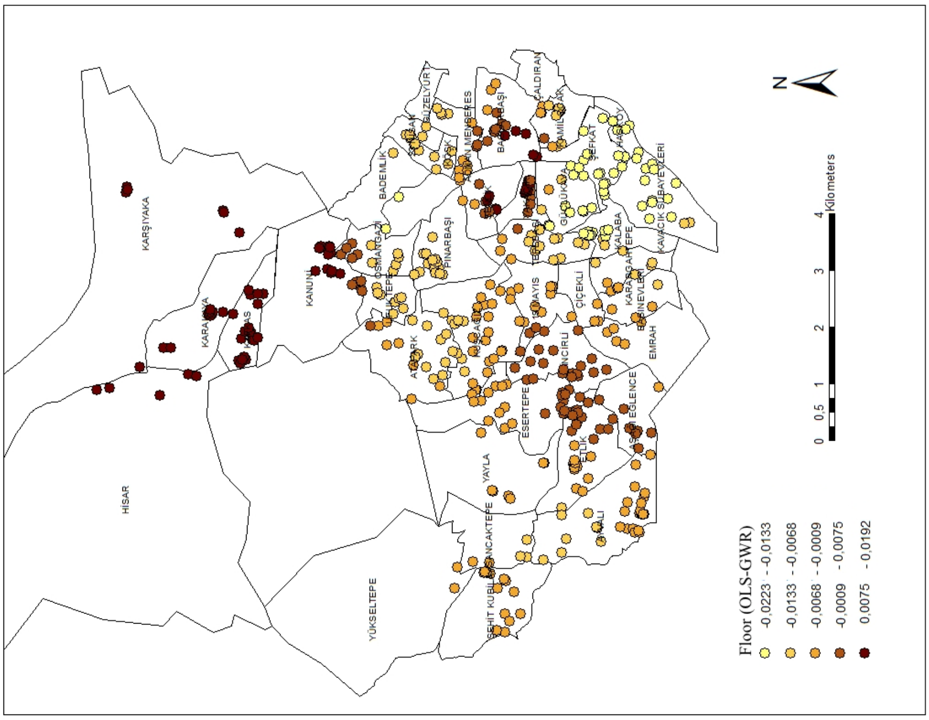
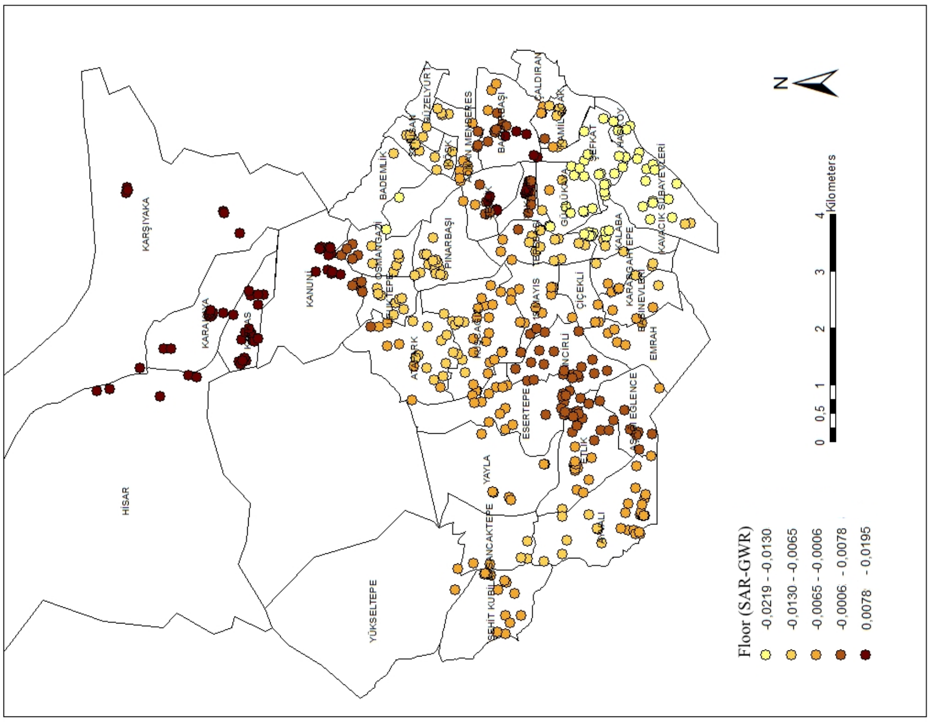


Figure 54: The results of Floor coefficients (GWR) at the 95% confidence level for Keçiören



(a)



(b)

Figure 54 (continued): a) Differences of Floor coefficients OLS-GWR and b) Differences of Floor coefficients SAR-GWR at the 95% confidence level for Keçiören

According to Figure 55, the neighborhoods with high-valued Security coefficients estimated by the GWR are Hisar, Karşıyaka, Karakaya, Kafkas, Kanuni, Esertepe, Etlik, İncirli and Emrah. The neighborhoods with low-valued Security coefficients estimated by the GWR are Osmangazi, Bademlik, 23 Nisan, Güzelyurt, Köşk, Pınarbaşı, Ufuktepe, Kuşcağız, Atapark, Tepebaşı, Güçlükaya, Kalaba, Hasköy, Kavacık Subayevleri and Kamil Ocak. The neighborhoods with high-valued Security coefficients estimated by the OLS/SAR are Osmangazi, Bademlik, 23 Nisan, Pınarbaşı, Ufuktepe, Kuşcağız, Atapark, Tepebaşı, Güçlükaya, Kalaba and Kavacık Subayevleri. The neighborhoods with low-valued Security coefficients estimated by the OLS/SAR are Hisar, Karşıyaka, Karakaya, Kafkas, Kanuni, Esertepe, Etlik, İncirli and Emrah.

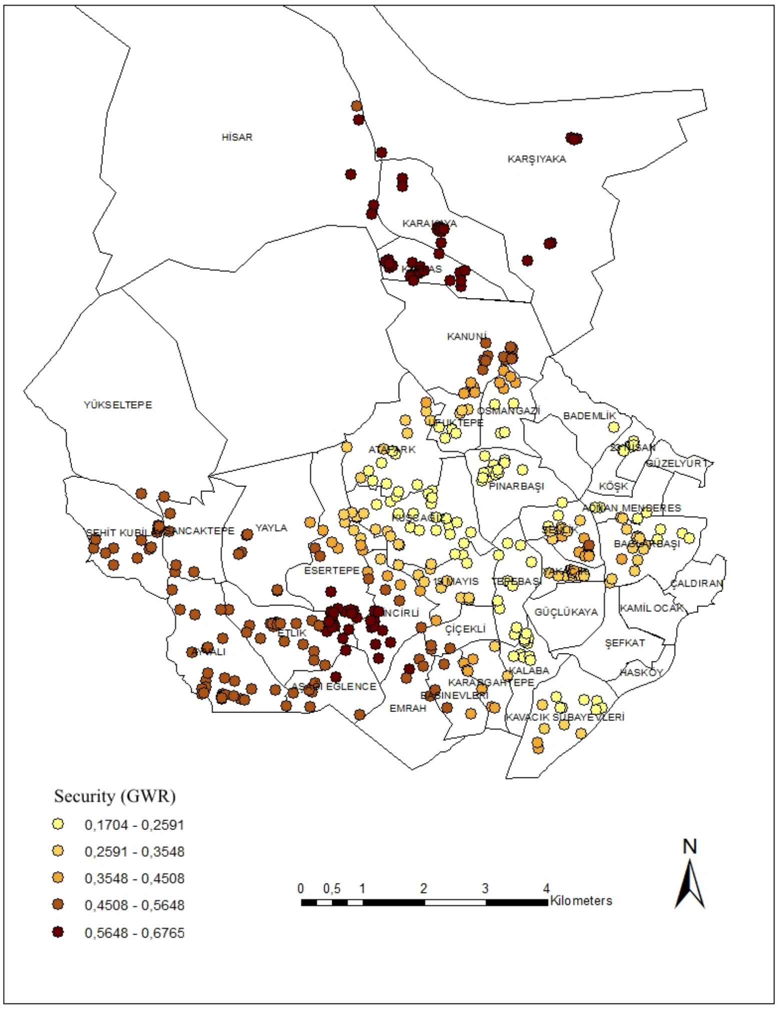
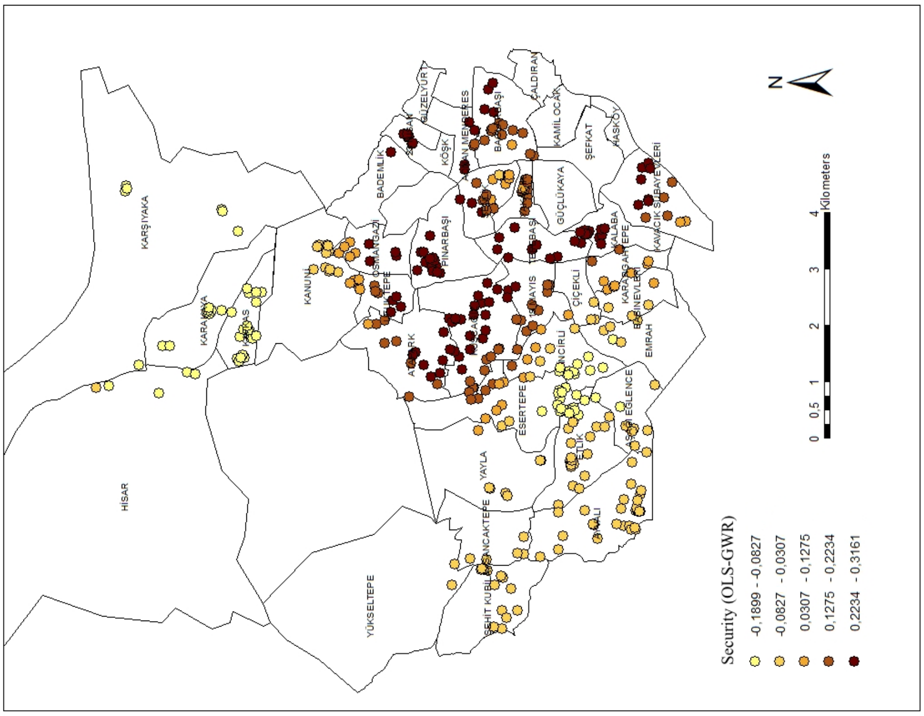
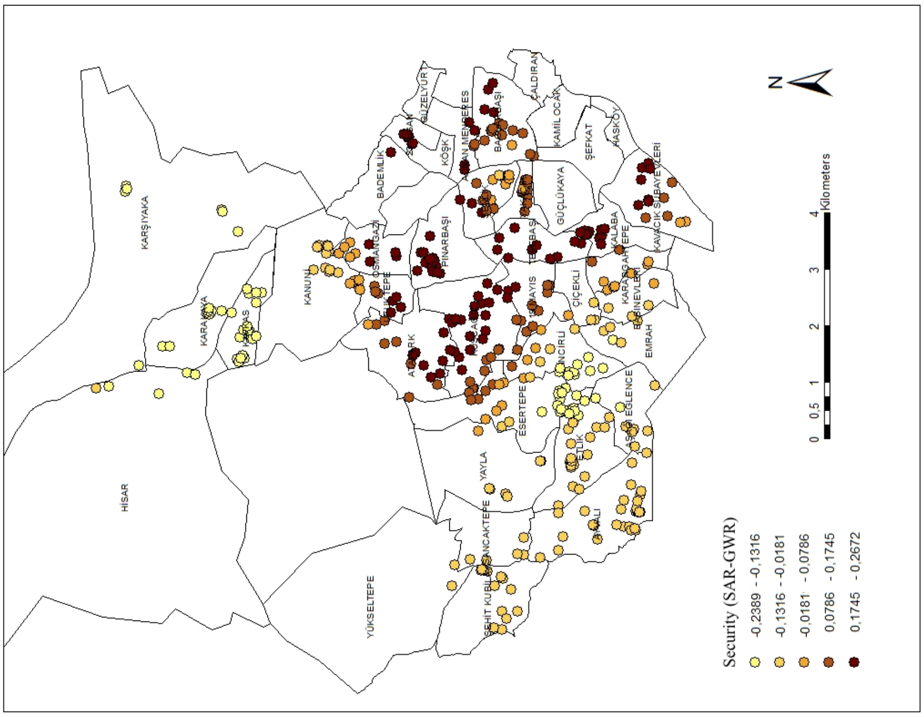


Figure 55: The results of Security coefficients GWR at the 95% confidence level for Keçiören



(a)



(b)

Figure 55 (continued): a) Differences of Security coefficients OLS-GWR and b) Differences of Security coefficients SAR-GWR at the 95% confidence level for Keçiören

According to Figure 56, the neighborhoods with high-valued Area coefficients estimated by the GWR are Şenlik, Bağlarbaşı, Yakacık, Güçlükaya and Kamil Ocak. The neighborhoods with low-valued Area coefficients estimated by the GWR are Atapark, Esertepe and Kuşcağız. The neighborhoods with high-valued Area coefficients estimated by the OLS/SAR are Atapark, Esertepe, İncirli and Kuşcağız. The neighborhoods with low-valued Area coefficients estimated by the OLS/SAR are Şenlik, Bağlarbaşı, Yakacık, Güçlükaya and Kamil Ocak.

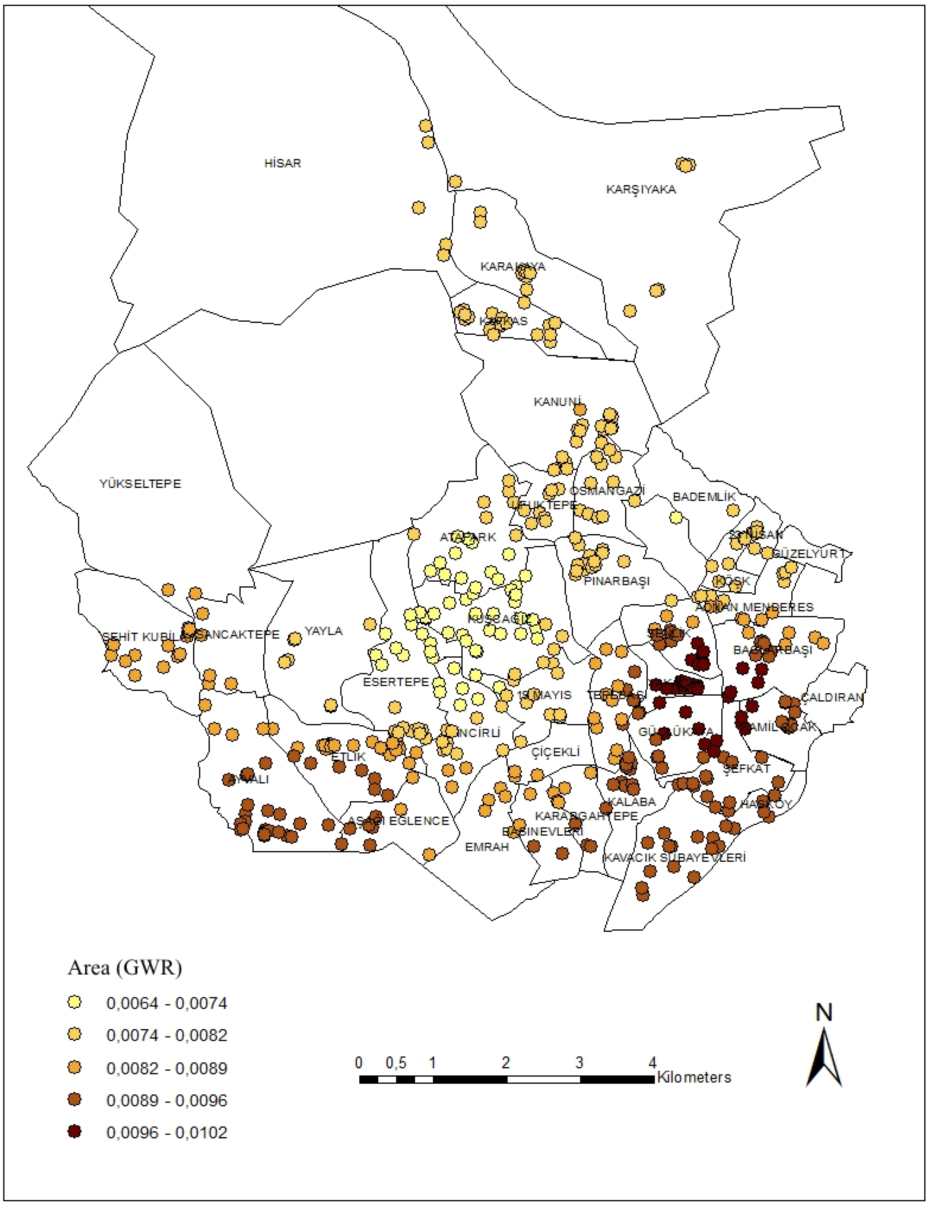
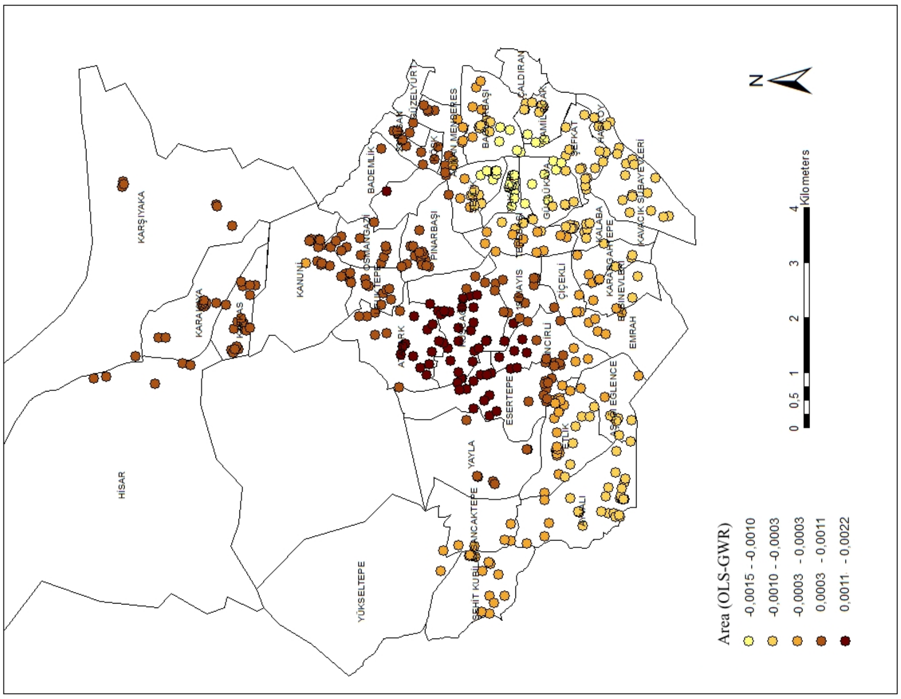
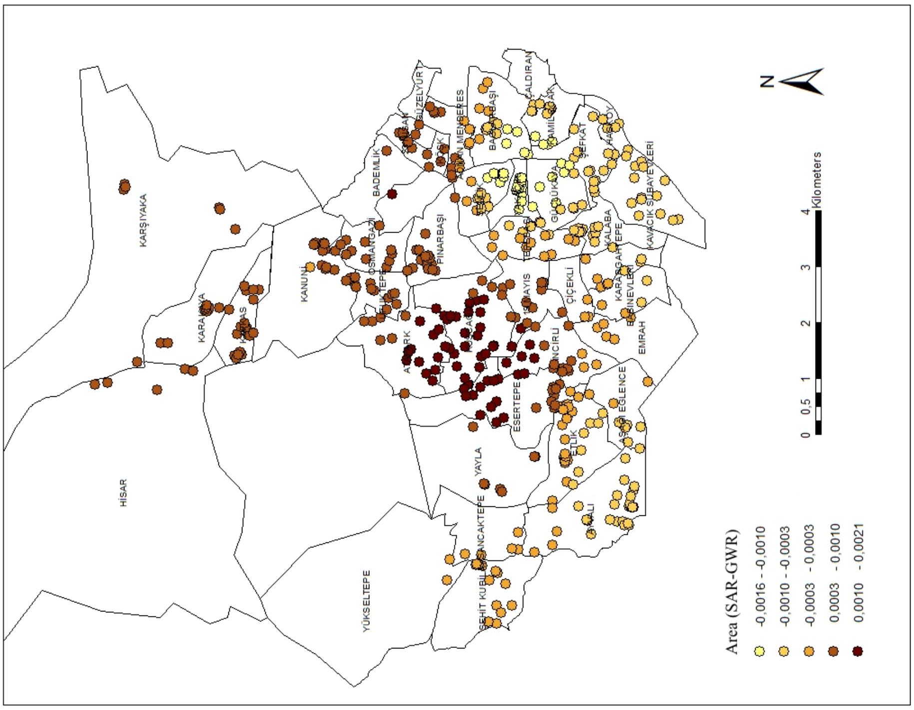


Figure 56: The results of Area coefficients (GWR) at the 95% confidence level for Keçiören



(a)



(b)

Figure 56 (continued): a) Differences of Area coefficients OLS-GWR and b) Differences of Area coefficients SAR-GWR at the 95% confidence level for Keçiören

According to Figure 57, the neighborhoods with high-valued Household coefficients estimated by the GWR are Kanuni, Karargahtepe, Bağlarbaşı, Kalaba and Basınevleri. The neighborhoods with low-valued Household coefficients estimated by the GWR are Kamil Ocak, Şevkat and Hasköy. The neighborhoods with high-valued Household coefficients estimated by the OLS/SAR are Kanuni, Karargahtepe, Bağlarbaşı, Tepebaşı and Basınevleri. Household coefficients for Yükseltepe and Kuşcağız neighborhoods are also estimated by the SAR model as high-valued coefficients differ from the OLS model. The neighborhoods with low-valued Household coefficients estimated by the OLS/SAR are Şevkat, Hasköy and Kamil Ocak.

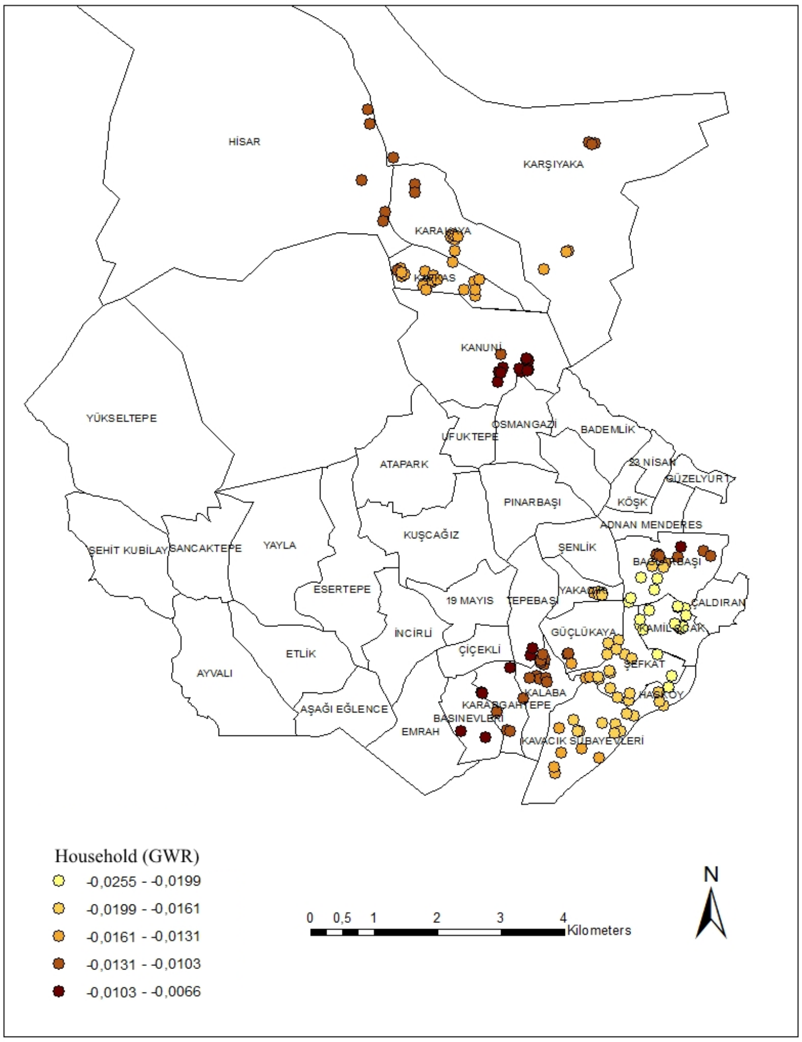
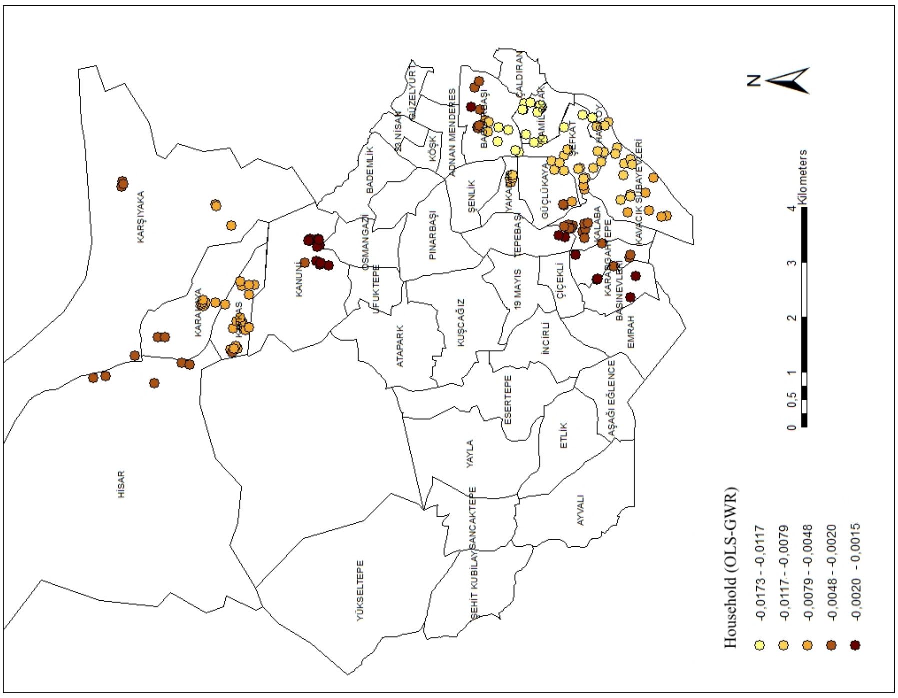
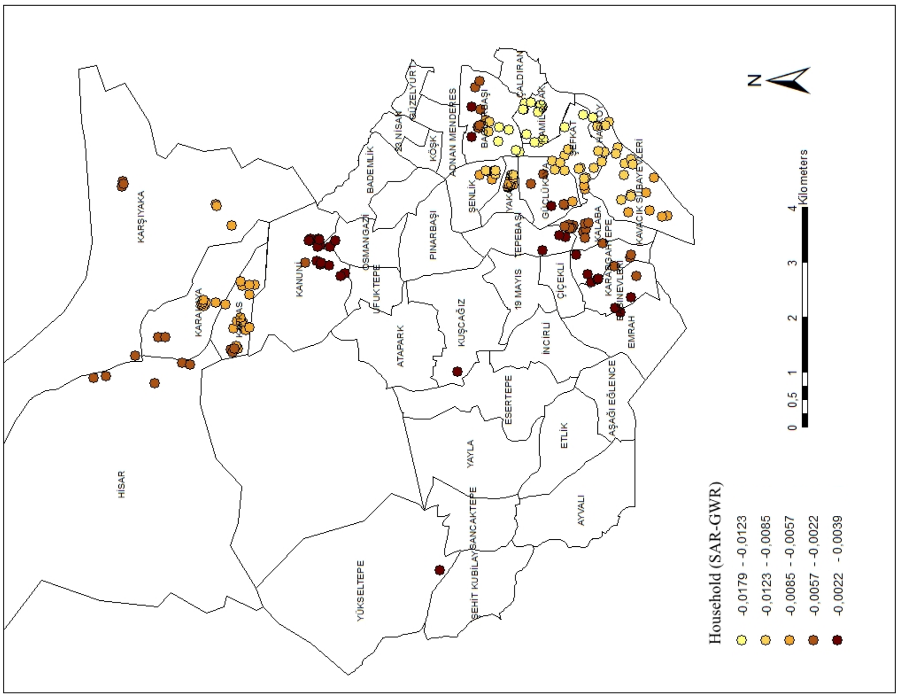


Figure 57: The results of Household coefficients (GWR) at the 95% confidence level for Keçiören



(a)



(b)

Figure 57 (continued): a) Differences of Household coefficients OLS-GWR and b) Differences of Household coefficients SAR-GWR at the 95% confidence level for Keçiören

According to Figure 58, the neighborhoods with high-valued Distance to Shopping Center coefficients estimated by the GWR are Kanuni, Esertepe, 19 Mayıs, Tepebaşı, Yakacık and Kamil Ocak. The neighborhoods with low-valued DistanceToShoppingCenter coefficients estimated by the GWR are Hisar, Karşıyaka, Karakaya, Ayvalı and Etlik. The neighborhoods with high-valued and low-valued DistanceToShoppingCenter coefficients estimated by the OLS/SAR are the same as the GWR.

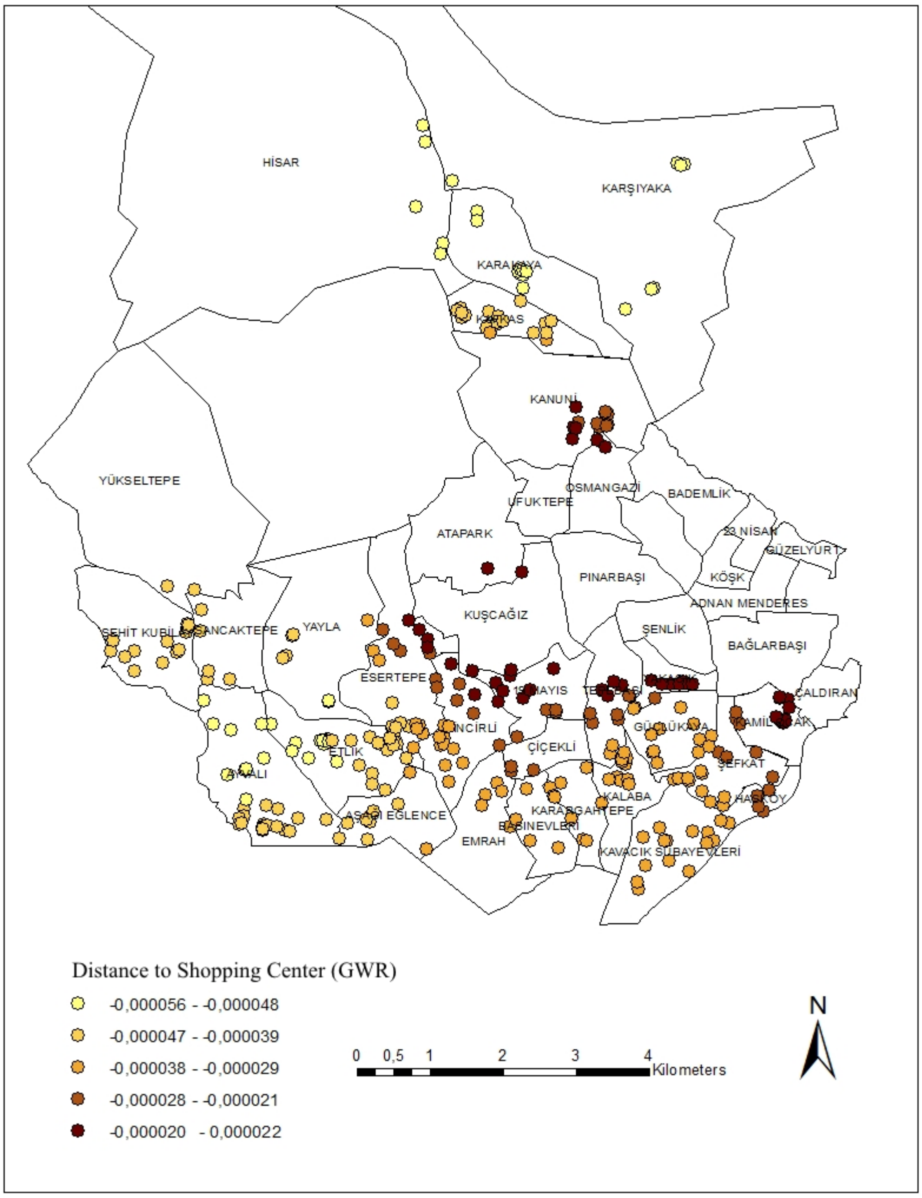
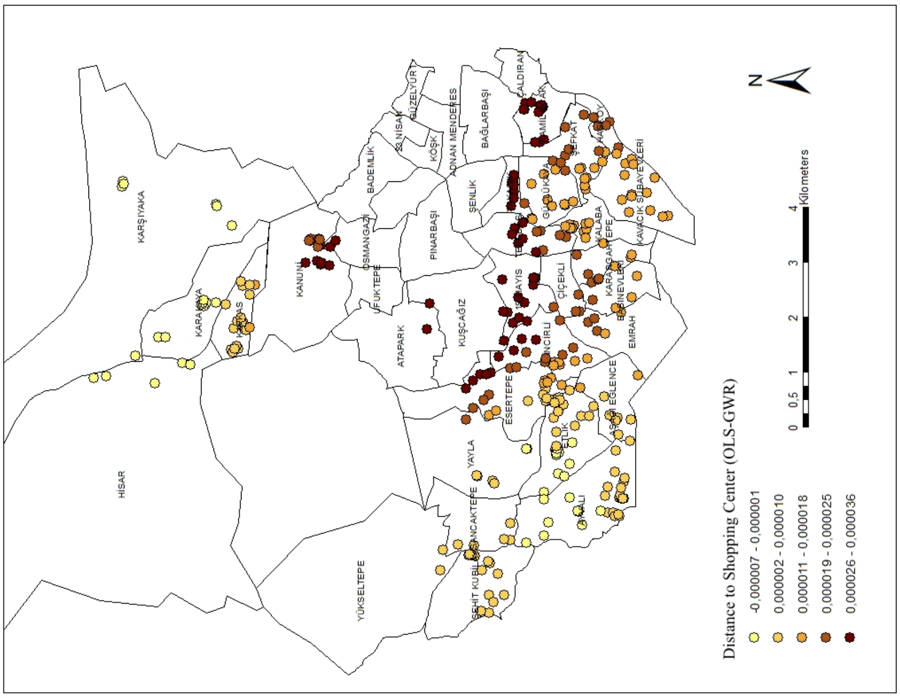
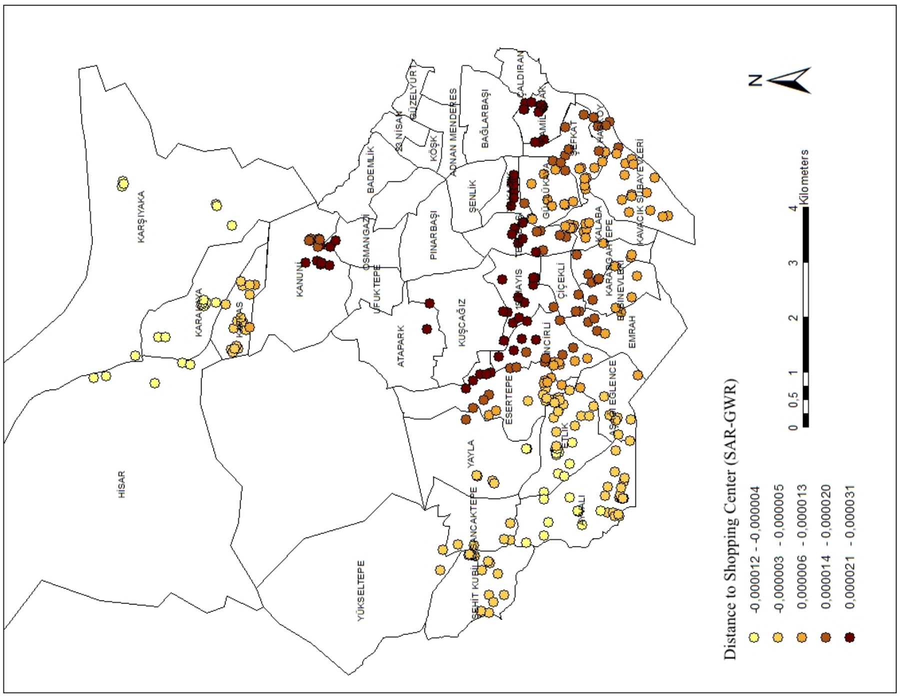


Figure 58: The results of Distance to Shopping Center coefficients (GWR) at the 95% confidence level for Keçiören



(a)



(b)

Figure 58 (continued): a) Differences of Distance to Shopping Center coefficients OLS-GWR and b) Distance to Shopping Center coefficients SAR-GWR at the 95% confidence level for Keçiören

As it is seen in figures above, the coefficient difference maps for the OLS-GWR and for the SAR-GWR are similar to each other in terms of coefficient estimates. The neighborhoods with high- or low-valued coefficients predicted by the OLS and the SAR global models coincide with each other. However, the resulting maps of the GWR show the opposite clustering maps of the coefficient differences for the OLS/SAR-GWR in point of the neighborhoods with high-valued and low-valued.

When the difference maps of the OLS-GWR (the maps that show the difference of the GWR coefficients from the OLS coefficients) and the difference maps of the SAR-GWR (the maps that show the difference of the GWR coefficients from the SAR coefficients) are compared, high clustering or low clustering is observed in the same place. When the GWR results are checked, however, a clustering that is opposite to the clustering of the difference maps of global models and the GWR is observed. That is to say, in places where the differences between global models and the GWR are high, low-valued clustering (small differences between the coefficients) is encountered, while in places where the differences of global models and the GWR are low, high-valued GWR clustering (high differences between the coefficients) is observed. The coefficients of Distance to Shopping Center for both districts and those of Distance to Subway estimated by the OLS, the SAR and the GWR showed the same clustering structure. In other words, the neighborhoods with high-valued clustering are the same for these three models or vice versa.

SUMMARY

- The most appropriate candidate variables for establishing the housing valuation model were selected with regard to the correlation matrix for both districts. A correlation matrix shows the amount of variability that is shared between two variables and what they have in common. Each variable is compared to another variable in the correlation matrix (Table 38 and Table 39). Of those independent variables among which there is a high correlation, the one that had the biggest influence on the dependent variable was kept in the analysis (as candidate variable) and the other one was removed.

- To determine the goodness-of-fit of the OLS models for both districts, first the R^2 values and the adjusted R^2 values were investigated. The results showed that the R^2 values are 0.762 and 0.782, and the adjusted R^2 values are 0.759 and 0.780 for Çankaya and Keçiören, respectively. These values indicated that 76% of the variation in housing value for Çankaya can be explained by the variance in Property Type, Security, Area, Distance to Subway and Distance to Mall. On the other hand, 78% of the variation in housing value for Keçiören can be explained by the variance in Floor, Security, Area, Distance to Mall and Household. Spatial models increased the R^2 values from 76% to 80% (both the SAR and the GWR) for Çankaya and from 78% to 82% (the SAR) and 84% (the GWR) for Keçiören.

- To see the difference between a global regression model (OLS) and a local regression model (GWR), the adjusted R^2 values and the AIC values were checked. The Adjusted R^2 value increased from 0.759 (the OLS result) to 0.793 (the GWR result) for Çankaya and from 0.78 (the OLS result) to 0.82 (the GWR result) for Keçiören. On the other hand, the AIC/AICc value decreased from -14.03 to -55.88 for Çankaya and from -170.99 to -271.08 for Keçiören.

- It can be stated that 79% of the variance in housing value for Çankaya can be explained by the variance in Property Type, Security, Area, Distance to Subway and Distance to Shopping Center variables. However, 82% of the variance in housing value for Keçiören can be explained by the variance in Floor, Security, Area, Distance to Shopping Center and Household variables.

- However, these values also indicate that some variables might be potentially omitted from the models because they still cannot explain 20% and 16% of the variation in housing value for Çankaya and Keçiören, respectively.

- The darker brown points in the maps show where the coefficients of the variables are greater and the light yellow points in the maps show lower coefficients. The darkest brown areas for each map demonstrate where the variable is an important factor in determining housing values. The light yellow colors for each map demonstrate those areas where the variable is not an important factor in determining housing prices.

- The highest coefficients of the Intercept variable were distributed more generally throughout the southeastern neighbourhoods of the Çankaya district.
- The highest coefficients of the Property Type variable were distributed more generally throughout the northeast neighbourhoods of the Çankaya district.
- The highest coefficients of the Security variable were distributed more generally throughout the east neighbourhoods of the Çankaya district.
- The highest coefficients of the Area variable were distributed more generally throughout the central and northern neighbourhoods of the Çankaya district.
- The highest coefficients of the Distance to Subway variable were distributed more generally throughout the central and western neighbourhoods of the Çankaya district.
- The highest coefficients of the Distance to Shopping Center variable were distributed more generally throughout the central and northern neighbourhoods of the Çankaya district.
- The highest coefficients of the Intercept variable were distributed more generally throughout the northern and southeastern neighbourhoods of the Keçiören district.
- The highest coefficients of the Floor variable were distributed more generally throughout the southeastern neighbourhoods of the Keçiören district.
- The highest coefficients of the Security variable were distributed more generally throughout the southeastern neighbourhoods of the Keçiören district.

- The highest coefficients of the Area variable were distributed more generally throughout the southeastern neighbourhoods of the Keçiören district.

- The highest coefficients of the Distance to Shopping Center variable were distributed more generally throughout the central neighbourhoods and the neighbourhoods from the center towards the east in the Keçiören district.

- The highest Distance to Household variable coefficients were distributed more generally throughout the central and southern neighbourhoods of the Keçiören district.

- While the areas with dark red pattern demonstrate the places where there is positive clustering, the blue color represents a pattern of negative clustering.

- The variables that were used most frequently in housing valuation in certain important studies in the literature were shown in Figure 1 graphically. When the variables in Figure 1 and those used in this thesis are compared, it can be seen that the Area and Floor variables show similarity to their uses in the literature. That is to say, the variables of Area and Floor, which are used in housing valuation the most and which affect housing values the most, appear as the most significant variables in this study as well. Nonetheless, the variables of Property Type, Available Security (private), Distance to Subway, Distance to Mall and Household, which do not come to the fore much in terms of usage and impact in the housing valuation literature, have become prominent in this thesis study.

- According to the results of this study, the housing value has a strong relations with the Area and Floor variables. Thanks to this result, it was able to infer that an increase in per square meter has a strong increase in housing value.

- The results confirm the previous studies in the literature, suggesting that Distance to Mall and Distance to Subway have a negative effect on house values.

- The results of this study support the previous studies which report the superiority of the local model (GWR) over the global model (OLS) in the field of housing valuation. This superiority is mainly due to the consideration of the spatial variation of the relationship across the study area.

- The results also promote that the local approach provides a better solution to the problem of spatially autocorrelated errors in spatial modelling.

- The results support the assumption that local modelling significantly improves the accuracy and prediction power of the model.

- The resulting spatial variation in the pattern of relationships shows that the strength of relationship decreases from north to south.

- In this study, the GWR model provided smaller errors than the OLS did.

- The housing value models based on the GWR and the SAR showed a better performance statistically when compared to those based on the OLS. Therefore, in the analysis the hedonic models of housing value, global and local spatial models were recommended.

- As a result, if a prospective housing valuation is going to be performed with the models built in this study, it will be a proper approach to use the spatial autoregressive model, which is a global model, in the Çankaya district and to use the GWR, a local model, in the Keçiören district.

- The global spatial model which is based on five parameters in the model used for Çankaya generally functions well. This might be related to the homogeneity of the Çankaya data set. In a housing valuation to be performed in the center and peripheries of Keçiören, it is necessary to use a local model because the variable coefficients demonstrate a huge variety. This shows that the Keçiören data set, in contrast to Çankaya, has a more heterogeneous structure. However, it is not possible to conclude that the urban fabric in Keçiören is more heterogeneous compared to Çankaya by taking only the results of this study into consideration. It is necessary that analyses be carried out using a broader data set for this purpose.

CHAPTER 6

CONCLUSIONS

It can be asserted that the main objective of this thesis, which is the evaluation of spatial and non-spatial hedonic techniques for housing valuation, was successfully achieved for the Çankaya and Keçiören districts. Three models were used to predict the log of housing values: the OLS model, the spatial autoregressive model and the geographically weighted regression model. The results of the SAR and the GWR models are compared with those of the OLS model. The best fitting housing valuation models were chosen based on the choice criteria, that is to say, the R^2, the Akaike Information Criterion, the Schwarz Information Criteria, and the log likelihood. The minimum AIC value was adopted as the basic criterion to determine the best fit housing valuation model.

The improvement of the fittings of OLS models was detected for the presence of spatial dependency and heterogeneity relationships. Generally, spatial lag and spatial error models are used to investigate the presence of spatial dependency. Since the traditional hedonic model do not consider the spatial relations in the data sets, the SAR models were used to explore the spatial dependency in the data set and include the spatial variables into the models. On the other hand, the GWR is the most popular method to detect the presence of heterogeneity. To determine the spatial variations in the data set, the GWR model proposed by Fotheringham et al., (1998) was adopted in this study. The results of the GWR were mapped to see the spatial variation in the study areas (from Figure 61 to Figure 114).

The results of the model predictions were validated by two different methods: regression line and error methods. The proximity of the model predictions to the regression line is used as a measure of the predictive power of the models. The other method used to find out the most suitable model is based on error computations. The results of error models were compared based on the four most common measures of predictive accuracy, namely MSE (Mean squared error), RMSE (Root mean squared error), MAE (Mean absolute error) and MAPE (Mean absolute percentage error). The outputs of these two methods indicated that spatial lag model is the best fit model for Çankaya data set and the GWR is the best estimation model for Keçiören data set.

The techniques of spatial regression and geographically weighted regression were employed to examine the spatial dependence and heterogeneity. The analyses demonstrate that the global and local models can be used for both data sets. The difference between the estimation powers of all these models is not large in this study. However, the SAR and the GWR spatial models have shown better performance, especially in terms of model performance and estimation accuracy than the ordinary least squares estimates. These findings are consistent with some previous studies in the housing valuation literature. The studies report that the spatial models (spatial lag, the spatial error and the GWR) outperform the OLS in terms of the goodness of fit and explanatory power (Long et al., 2007; Bitter et al., 2007; Vichiensan et al., 2011; Yu et al., 2007; Propastin and Kappas, 2008; De Bruyne and Van Hove, 2013).

The result maps of the GWR model showed how the coefficients of each parameter changed spatially. The GWR can bring about significant benefits in creating a housing valuation model and index maps for Turkey. This means taking into account the spatial variability could be an important tool for designing and evaluating house valuation strategies in Turkey.

It is possible to construct global and local house value prediction models throughout Turkey using the methodology chosen for this study and to build housing index maps or house value maps based on these models.

The housing value is regressed to a function of the structural, environmental and location attributes of the houses. This approach is known as the hedonic regression approach and it is widely used in house price prediction models. The address for every dwelling unit in the datasets was geocoded at the building level and matched with a wide set of spatial variables. The dataset includes dummy variables with information about available facilities such as balcony, swimming pool, lift, car park, building quality, property type, heating system and security (private) in the dwelling unit. This research found that most of the characteristics have a significant influence on housing values. However, in this thesis only 5 out of 45 characteristics for each district were used because of multicollinearity. They are classified into two groups according to their impact degree for Çankaya and three groups for Keçiören. The structural variables (property type, security, area) and the accessibility variables (distance to subway and distance to shopping centers) have a significant impact on housing values for the Çankaya dataset. The structural variables (floor, security, and area), the accessibility variable (distance to shopping centers) and the social/demographic variable (size of household) have significant effects on housing values for the Keçiören dataset.

These findings are supportive to the results in the housing valuation literature, in which value is a function of location, structure, social, demographic and neighboring characteristics (Selim, 2011; Yankaya, 2004; Özüş et al., 2007; Adair et al., 2000; Huang et al., 2010; Vichiensan et al., 2011).

However, there were some limitations to this study. The parameters of Security (private), Distance to Subway, Distance to Shopping Center used in this study can only be utilized in a few metropolitan cities. In many regions of Turkey there are no shopping centers, underground systems or residential sites with private security. In this case, it will be necessary to establish housing valuation models using other parameters unique to each region. For this purpose, more research needs to be done on housing valuation in Turkey. As a further work, it is suggested to carry out research considering the variables and method used in this thesis.

The models in this study were also tried to be tested through real world applications and given in Appendix E. House prices in certain neighborhoods in Çankaya and Keçiören were obtained from real estate agents and the prices of houses with similar characteristics were taken from Internet sites for estate sale and compared with the model prediction results in this study. The model prediction results can perform predictions with a maximum deviation of 20% from the real market values. On the other hand, it is also seen that there is not a vast difference between house prediction values of the models built for this study.

In interviews with real estate agents in Çankaya and Keçiören, it was stated that the factors that affect housing values the most in the real world are the location of the house, its area, the number of floors, facade, age and transportation facilities to the center, respectively. Houses with a large area, high floors (except the entrance and the top floor) and south facade are in great demand by the buyers. The prices of those houses that possess these characteristics are higher than the mean neighborhood prices and are directly related to the income status of buyers. Within this context, it is of great importance that the income level of the buyer is used as a parameter in housing valuation models. Nevertheless, it is not possible yet to get access to the databases as to the income or to retrieve data about income from the relevant institution. This will only be possible through an online central real estate valuation system.

The points below were determined in the field study carried out to compare the housing prices in locations where the case study was performed and the results of model predictions.

Old buildings (40 years old or older) are sold for surprisingly high prices. Owners of many old properties prefer to keep their existing housings with and expectation of urban transformation. This is clearly seen in the neighborhoods of Tınaztepe, Çankaya and Şevkat, Keçiören. Advertisements of real estate for sale are encountered quite rarely in these neighborhoods.

The contractors collect old buildings and construct new buildings in place of them and the owners of old properties are given new ones that are more valuable. In this case, it is not possible to talk about urban transformation, because in urban transformation an interaction between urban texture, infrastructure and transportation systems and social spaces should be provided. However, currently only the old buildings are demolished in many cases and new buildings are constructed at the same location. This is called urban renewal in the field of planning. The age of building becomes the most important factor that affects the housing value.

Huge price gaps were found between properties with similar features on both sides of a street in the same neighborhood. The real estate agents attribute this to the discrepancies in education, culture and income. It is stated that usually people with a low educational background and low income reside in buildings where shoes are kept outside the flat doors and carpets and similar things are hanged on the balconies.

Therefore, when performing a housing valuation study in Ankara, including the data on education and income in the analyses can contribute significantly to the housing valuation model.

Recently a different way of housing construction can be seen in Ankara. The ground floors of the buildings which are constructed on commercial lots are designed as shopping centers, while the higher floors are constructed as housing units. In some cases, one part of the building is constructed as flats, while the other part is constructed as workplace (office). Since the VAT is 1% for flats and 18% for workplaces, the prices of flats in a building with similar features vary. Consequently, it is necessary that information be obtained as to whether flats in buildings that have been constructed on commercial lots are housing or workplaces.

Each neighborhood has a particular center. Estate agents denote that proximity to the neighborhood center is more important than proximity to the city center. It will be beneficial in housing valuation studies to be conducted in Ankara to determine the center of each neighborhood and to include the proximity to this center in the analysis as a variable.

Recommendations for Future Work

The data about housing values used in this study consist of only those values determined up to a certain standard. Thus, it can be regarded a limitation of the thesis. Apart from the valuation firms, there are also several other sources of house prices such as real estate agents, web sites for real estate sale, and estate deed sales. It will be useful to test the methodology of this study through the data to be obtained from other sources.

According to the housing valuation literature, these variables except Area and Floor are not the most important determinants on housing valuation. This study should be performed for other districts of Ankara and also for Turkey. The parameters that affect house values might be different in each neighborhood or district; therefore efforts must be made to determine appropriate parameters using the methodology of this study.

Valuation takes into account the open market price determined according to certain characteristics of a given house in a given period of time. Housing valuation is a difficult task due to the great variability of affecting internal and external variables. A great amount of up-to-date information is needed in order to make accurate estimations. For this purpose, databases related to the structural, environmental, economic, social and locational characteristics of the house need to be built. In fact there are numerous databases in Turkey which might serve this purpose, such as The Central Civil Registration System (MERNIS), the Land Registry and Cadastre Information System (TAKBIS), Spatial Address Registration System (UAVT/MAKS), Building Inspection System, Turkey Geology Database (TJVT), Orthophoto (in 1/1000 and 1/5000 scale), and road and rail transport databases. A dynamic central housing valuation system based on these databases should be constructed for Turkey. This online system should take into consideration both supply-side determinants (variables) such as housing incentives, the arrangements of The Mass Housing Law, zoning regulations, urban transformation, infrastructure policies, tax arrangements for housing sector and other regulations, and also demand-side determinants such as income, demographic variables, taxes, bank housing credit interest regulations, tax cuts and advantages in addition to structural, neighboring and environmental factors. The developments in information technology including GIS are able to provide this information to buyers, sellers, planners, valuers and decision makers.

REFERENCES

Adair, A., McGreal, S., Smyth, A., Cooper, J., and Ryley, T. (2000). House prices and accessibility: The testing of relationships within the Belfast urban area. Housing studies, 15(5), 699-716.

Adair, A. S., Berry, J. N. and McGreal, W. S. (1996). Hedonic modelling, housing submarkets and residential valuation. Journal of Property Research, 13(1), 67-83.

Akaike, H. (1974). A New Look at the Statistical Model Identification, I.E.E.E. Transactions on Automatic Control, AC 19: 716-723.

Akaike, H. (1987). Factor analysis and AIC. Psychometrika, 52, 317-332.

Amin, A., Colman, A., and Grunske, L. (2012). An approach to forecasting QoS attributes of web services based on ARIMA and GARCH models. In Web Services (ICWS), 2012 IEEE 19th International Conference on (pp. 74-81). IEEE.

Anselin, L. (2001). Spatial econometrics. A companion to theoretical econometrics, 310330.

Anselin, L. (1999). Spatial Econometrics. School of Social Sciences University of Texas, Dallas.

Anselin L. Spatial econometrics: methods and models. Vol. 4. Springer Science & Business Media, 1988.

Anselin, L., and Getis, A. (2010). Spatial statistical analysis and geographic information systems. In Perspectives on Spatial Data Analysis (pp. 35-47). Springer Berlin Heidelberg.

Archer, W. R., Gatzlaff, D. H., and Ling, D. C. (1996). Measuring the importance of location in house price appreciation. Journal of urban Economics, 40(3), 334-353.

Ayan, E., and Erkin, H. C. (2014). Hedonic Modeling for a Growing Housing Market: Valuation of Apartments in Complexes. International Journal of Economics and Finance, 6(3), p188.

Arslan, Y. (2014). Interest rate fluctuations and equilibrium in the housing market. The BE Journal of Macroeconomics, 14(1), 173-204.

Babawale, G. K., The Impact of Neighbourhood Churches on House Prices- Department of Estate Management, University of Lagos, Nigeria- www.ccsenet.org/jsd Journal of Sustainable Development Vol. 4, No. 1; February 2011.

Bae, C. H. C., Jun, M. J., and Park, H. (2003). The impact of Seoul's subway Line 5 on residential property values. Transport policy, 10(2), 85-94.

Bailey, T. C., Gatrell, A. C. (1995). Interactive Spatial Data Analysis. Essex: Longman Scientific and Technical.

Basu, S., Thibodeau, T. G. (1998). Analysis of spatial autocorrelation in house prices. The Journal of Real Estate Finance and Economics, 17(1), 61-85.

Baum, A. and Crosby, N. (1995). Property investment appraisal (pp. 60-108). London: Routledge.

Beal, D. J. (2007). Information criteria methods in SAS for multiple linear regression models. 15th Annual SouthEast SAS Users Group (SESUG) Proceedings, South Carolina, Paper SA05.

Bekar, E. and Akay, Ç. E. (2014). Modelling Housing Prices in Istanbul Applying the Spatial Quantile Regression, The Empirical Economics Letters,13(8), ISSN 1681 8997.

Bello, A. K., and Moruf, A. (2010). Does the Functional Form Matter in the Estimation of Hedonic Price Model for Housing Market?. The Social Sciences, 5(6), 559-564.

Bin, O., Poulter, B., Dumas, C. F., and Whitehead, J. C. (2011). Measuring the impact of sea-level rise on coastal real estate: A hedonic property model approach. Journal of Regional Science, 51(4), 751-767.

Bin, O. (2004). A prediction comparison of housing sales prices by parametric versus semi-parametric regressions. Journal of Housing Economics, 13(1), 68-84.

Bitter, C., Mulligan, G. F., and Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. Journal of Geographical Systems, 9(1), 7-27.

Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V. (2008). Applied Spatial Data Analysis with R. New York; London: Springer.

Bowes, D. R., and Ihlanfeldt, K. R. (2001). Identifying the impacts of rail transit stations on residential property values. Journal of Urban Economics, 50(1), 1-25.

Bozdogan, H. (2000). Akaike's information criterion and recent developments in informational complexity. Journal of Mathematical Psychology, 44, 62-91.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. (1998). Spatial nonstationarity and autoregressive models. Environment and Planning A, 30(6), 957-973.

Bulut, B., Allahverdi, N., Kahramanli, H., and Yalpir, S. (2011). A Residential real-estate valuation model with reduced attributes. International Journal Of Mathematical Models And Methods In Applied Sciences, 3 (5), 586-593.

Bulut, B., Allahverdi, N., Yalpir, Ş., and Kahramanli, H. (2010). Effects of attribute reducing on real-estate valuation. transport, 265(86.9), 10-0.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. Sociological methods and research, 33(2), 261-304.

Can, A. (1992). Specification and estimation of hedonic housing price models. Regional science and urban economics, 22(3), 453-474.

Can, A. (1990). The measurement of neighborhood dynamics in urban house prices. Economic Geography 66, 254–272.

Case, K. E., and Mayer, C. J. (1996). Housing price dynamics within a metropolitan area. Regional Science and Urban Economics, 26(3), 387-407.

Cassetti, E. (1972). Generating models by the expansion method: applications to geographical research. Geographical Analysis, 4, 81-91.

Cavanaugh, J. E. (1997). Unifying the derivations for the Akaike and corrected Akaike information criteria. Statistics and Probability Letters, 33(2), 201-208.

Celik, H. M., and Yankaya, U. (2006). The impact of rail transit investment on the residential property values in developing countries: the case of Izmir subway, Turkey. Property management, 24(4), 369-382.

Chappell, A., Renzullo, L., and Haylock, M. (2012). Spatial uncertainty to determine reliable daily precipitation maps. Journal of Geophysical Research: Atmospheres (1984–2012), 117(D17).

Charlton, M., Fotheringham, S., and Brunsdon, C. (2009). Geographically weighted regression. White paper. National Centre for Geocomputation. National University of Ireland Maynooth.

Chen, J., and Hao, Q. (2008). The impacts of distance to CBD on housing prices in Shanghai: a hedonic analysis. Journal of Chinese Economic and Business Studies, 6(3), 291-302.

Chi, G., and Zhu, J. (2008). Spatial regression models for demographic analysis. Population Research and Policy Review, 27(1), 17-42.

Chin, H. C., and Foong, K. W. (2006). Influence of school accessibility on housing values. Journal of urban planning and development, 132(3), 120-129.

Chin, T. L. and Chau, K. W. (2003). A critical review of literature on the hedonic price model, International Journal for Housing and Its Applications 27 (2), 145-165.

Chrostek, K., and Kopczewska, K. (2013). Spatial Prediction Models for Real Estate Market Analysis. Ekonomia/Uniwersytet Warszawski, (35), 25-43.

Clapp, J. M. (2003). A semiparametric method for valuing residential locations: application to automated valuation. The Journal of Real Estate Finance and Economics, 27(3), 303-320.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2013). Applied multiple regression/correlation analysis for the behavioral sciences. Routledge.

Craven, B. D. and Islam S. M. N. Ordinary least-squares regression. SAGE Publications, 2011.

De Bruyne, K., and Van Hove, J. (2013). Explaining the spatial variation in housing prices: an economic geography approach. Applied Economics, 45(13), 1673-1689.

Din, A., Hoesli, M., and Bender, A. (2001). Environmental variables and real estate prices. Urban Studies, 38(11), 1989-2000.

Dmitry, P. (2009). Public transport accessibility and residential land value: a case study of Riga. Proceedings of the 9th International Conference “Reliability and Statistics in Transportation and Communication”, 21–24 October, Riga, Latvia.Dubin, R. A. (1998). Spatial autocorrelation: a primer. Journal of Housing Economics 7, 304–327.

Dubin, R. A. (1998). Predicting House Prices Using Multiple Listings Data, Journal of Real Estate Finance and Economics, 17(1), 35-59.

Dubin, R. A. (1992). Spatial autocorrelation and neighborhood quality. Regional science and urban economics, 22(3), 433-452.

Düzgün, H. Ş. B., and Kemeç, S. (2008). Spatial regression and geographically weighted regression for spatial prediction. the Encyclopedia of Geographical Information Science, S. Shekhar and H. Xiong (Eds.) New York: Springer.

Efthymiou, D., and Antoniou, C. (2013). How do transport infrastructure and policies affect house prices and rents? Evidence from Athens, Greece. Transportation Research Part A: Policy and Practice, 52, 1-22.

Erener, A. and Düzgün, H. S. B. (2012). Landslide susceptibility assessment: what are the effects of mapping unit and mapping method? Environmental Earth Sciences, 66(3), 859-877.

Erener, A., and Düzgün, H. S. B. (2010). Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of More and Romsdal (Norway). Landslides, 7(1), 55-68.

ESRI (1998). ESRI Shapefile Technical Description. ESRI. http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf. Last visited in January, 2015.ESRI (2010). ArcGIS Desktop Help. Environmental Systems Research Institute (ESRI), Redlands, California.

Farber, S. and Yeates, M. (2006). A Comparison of Localized Regression Models in a Hedonic House Priee Context. Canadian Journal of Regional Science. 3, 405-420.

Fik, T. J., Ling, D. C., and Mulligan, G. F. (2003). Modeling spatial variation in housing prices: a variable interaction approach. Real Estate Economics, 31(4), 623-646.

Fischer, M. M. and Getis, A. (2009). Handbook of applied spatial analysis: software tools, methods and applications. Springer.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships; Wiley: New York, NY, USA.

Fotheringham, A. S., Charlton, M.E., and Brunsdon, C. (1998). Geographical weighted regression: a natural evolution of the expansion method for spatial data analysis. Environment and Planning A30,1905–1927.

Freeman III, A. M. (1979). Hedonic prices, property values and measuring environmental benefits: a survey of the issues. The Scandinavian Journal of Economics, 154-173.

Gangodagamage, C., Zhou, X., and Lin, H. (2008). Spatial Autocorrelation. S. Shekhar, and H. Xiong, Encyclopedia of GIS. New York: Springer.

Gao, X., Asami, Y., and Chung, C. F. (2006). An empirical evaluation of spatial regression models. Computers and Geosciences 32: 1040–1051.

Gao, X., Asami, Y. (2005). Influence of Spatial Features on Land and Housing Prices. Tsinghua Science and Technology. 10(3):344-353.

Gayawan, E. and Ipinyomi, R.A. (2009). A comparison of Akaike, Schwarz and R square criteria for model selection using some fertility models, Australian Journal of Basic and Applied Sciences, 3(4): 3524-3530.

Getis, A., and Aldstadt, J. (2010). Constructing the spatial weights matrix using a local statistic. In Perspectives on spatial data analysis (pp. 147-163). Springer Berlin Heidelberg.

Goodman, A. C, Thibodeau T. G. (2003). Housing market segmentation and hedonic prediction accuracy. J Housing Econ 12:181–201.

Griffith, D. A. (1987). Spatial autocorrelation. A Primer. Washington DC: Association of American Geographers.

Gujarati, D. (1970). Use of dummy variables in testing for equality between sets of coefficients in linear regressions: a generalization. The American Statistician, 24(5), 18-22.

Gultekin, B., and Yamamura, E. (2002). Predicting Housing Prices in Central Ankara, Turkey Based on Spatial Dependence Analysis. 33(1), 217-227.

Gür, M., Çağdaş, V., and Demir, H. A. General Overview Real Estate Valuation Applications in Turkey, FIG XXII International Congress Washington, D.C. USA, April 19-26 2002.

Harrison Jr, D., and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. Journal of environmental economics and management,5(1), 81-102.

Helbich, M., Brunauer, W., Vaz, E., and Nijkamp, P. (2013). Spatial heterogeneity in hedonic house price models: the case of Austria. Urban Studies, 0042098013492234.

Hinkle, D.E., Wiersma, W., and Jurs, S.G. (1998). Applied statistics for the behavioral sciences. Boston: Houghton Mifflin Company, (4th ed.), 749 pps.

Ho, R. (2006). Handbook of univariate and multivariate data analysis and interpretation with SPSS. CRC Press.

Hough, E. C., Nell, W. T., Maine, N., Groenewald, J. A., and van der Rijst, M. (2010, September). Best fit model selection for spatial differences (regression) in the profitability analysis of precision phosphate (P) application to winter cereals in Precision Agriculture (PA). In 2010 AAAE Third Conference/AEASA 48th Conference, September 19-23, 2010, Cape Town, South Africa (No. 96642). African Association of Agricultural Economists (AAAE) & Agricultural Economics Association of South Africa (AEASA).

Huang, B., Wu, B., and Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. International Journal of Geographical Information Science, 24(3), 383-401.

Hurlin, C. (2013). Maximum Likelihood Estimation. http://www.univ-orleans.fr/deg/masters/ESA/CH/Chapter2_MLE.pdf (17.05.2015)

Hurvich, C. M., and Tsai, C. L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. Journal of time series analysis, 14(3), 271-279.

Hwang, S. N. (2003). Environmental amenities and disamenities, and housing prices; using GIS techniques (Doctoral dissertation, Texas A&M University).

İsmail, S., Buyong, T., Sipan İ., Hashim, M.G., Navaneethan, R. (2008). Spatial Hedonic Modelling (SHM) For Mass Valuation . International Real Estate Research Symposium IRERS 2008.

Jaffe, A. J., and Sirmans, C. F. (1995). Fundamentals of real estate investment. Prentice hall.

Karagöl, T. (2007). A study of housing prices in Ankara. Msc. Thesis, Department of City and Regional Planning, Ankara - Turkey.

Kaya, A., Bozkurt, A. T., Bastan, E. M., & Ayanoglu, Ö. A. (2013). Constructing a house price index for Turkey. IFC Bulletins chapters, 36, 153-171.Keskin, B. (2008). Hedonic analysis of price in the Istanbul housing market. International Journal of Strategic Property Management, 12(2), 125-138.

Kaya, A. (2012). Determination of The Factors Affecting Housing Prices in Turkey by using Hedonic Pricing Model. Specialist Qualification Thesis, Central Bank of Republic of Turkey.

Keskin, B. (2008). Hedonic analysis of price in the Istanbul housing market. International Journal of Strategic Property Management, 12(2), 125-138.

Kestens, Y., Thériault, M., and Des Rosiers, F. (2006). Heterogeneity in hedonic modelling of house prices: looking at buyers' household profiles. Journal of Geographical Systems, 8(1), 61-96.

Ki, C. O., and Jayantha, W. M. 2010. The Effects of Urban Redevelopment on Neighbourhood Housing Prices, International Journal of Urban Sciences,14(3), 276-294.

Kiel, K. A., and Zabel, J. E. (2008). Location, location, location: The 3L Approach to house price determination. Journal of Housing Economics, 17(2), 175-190.

Kissling, W. D., and Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. Global Ecology and Biogeography, 17(1), 59-71.

Koehler, A. B., and Murphree, E. S. (1988). A comparison of the Akaike and Schwarz criteria for selecting model order. Applied Statistics, 187-195.

Konecny, M., Zlatanova, S., and Bandrova, T. L. (2010). Geographic information and cartography for risk and crisis management. Springer, Berlin.

Koramaz, T. K. (2013, November). A Dualistic Manner In Spatial Analysis: Location And Sector Based Investigations For Housing Price Determinants. In ERSA conference papers (No. ersa13p1062). European Regional Science Association.

Koramaz, T. K., and Dokmeci, V. (2012). Spatial determinants of housing price values in Istanbul. European Planning Studies, 20(7), 1221-1237.

Kryvobokov, M. (2013). Hedonic price model: defining neighbourhoods with Thiessen polygons, International Journal of Housing Markets and Analysis, 6: 79-97.

Kuşan, H., Aytekin, O., and Özdemir, İ. (2010). The use of fuzzy logic in predicting house selling price. Expert Systems with Applications, 37(3), 1808-1813.

Lancaster, K. J. (1966). A new approach to consumer theory. The journal of political economy, 132-157.

Larraz, B. (2011). An Expert System for Online Residential Properties Valuation. Review of Economics and Finance, 1, 69-82.

Lehner, M. (2011). Modelling housing prices in Singapore applying spatial hedonic regression. Master of Science Thesis Spatial Development and Infrastructure Systems. Institute for Transport Planning and Systems (IVT), ETH Zurich, Zurich.

Lewis, L. Y., Bohlen, C., and Wilson, S. (2008). Dams, dam removal, and river restoration: A hedonic property value analysis. Contemporary Economic Policy, 26(2), 175-186.

Lichstein, J. W., Simons, T. R., Shriner, S. A., and Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs, 72 (3), 445-463.

Lin, C.H., Wen, T.(2011). Using Geographically Weighted Regression (GWR) to Explore Spatial Varying Relationships of Immature Mosquitoes and Human Densities with the Incidence of Dengue, Int. J. Environ. Res. Public Health, 8: 2798-2815.

Liu, J. G., Zhang, X. L., and Wu, W. P. (2006). Application of fuzzy neural network for real estate prediction. In Advances in Neural Networks-ISNN 2006 (pp. 1187-1191). Springer Berlin Heidelberg.

Lokshina, I.; Hammerslag, M.; and Insinga, R. (2003). Applications of artificial intelligence methods for real estate valuation and decision suppor. In Hawaii international conference on business.

Long, F., Páez, A., and Farber, S. (2007). Spatial effects in hedonic price estimation: a case study in the city of Toronto. Working paper series, Centre for Spatial Analysis, Canada.

Makridakis, S. and Hibon, M. (1995). Evaluating Accuracy (or Error) Measures, INSEAD Working Paper 95/18/TM.

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. International Journal of Forecasting, 9(4), 527-529.

Malpezzi, S. (2003). "Hedonic pricing models: A selective and applied review", in T. O'Sullivan and K. Gibb, Housing Economics and Public Policy, Blackwell Science, Oxford.

Malpezzi, S., Ozanne, L., and Thibodeau, T. G. (1987). Microeconomic estimates of housing depreciation. Land Economics, 372-385.

Matthews, S. A., and Yang, T. C. (2012). Mapping the results of local statistics: Using geographically weighted regression. Demographic Research. 26(6):151-166.

Mbachu, J. I. and Lenono, N. (2005). Factors influencing market values of residential properties. Queensland University of Technology Research Week, pp.891-903.

McCluskey, William J., Borst, Richard A. (2007). Specifying the effect of location in multivariate valuation models for residential properties, Emerald Group Publishing Limited, 25(4):312-343.

McMillen, D. P. (2008). Changes in the distribution of house prices over time: Structural characteristics, neighborhood, or coefficients?. Journal of Urban Economics, 64(3), 573-589.

Mennis, J. (2006). Mapping the results of geographically weighted regression.The Cartographic Journal, 43(2), 171-179.

Miles, J. and Shevlin, M. (2001). Applying regression and correlation: A guide for students and researchers. Sage.

Mimis, A., Rovolis, A., and Stamou, M. (2013). Property valuation with artificial neural network: the case of Athens. Journal of Property Research, 30(2), 128-143.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2012). Introduction to linear regression analysis (Vol. 821). John Wiley & Sons.

Morancho, A.B., (2003). A Hedonic Valuation Of Urban Green Areas. Elsevier Science, 66(2003), 35-41.

Neelawala, P. (2010). Impacts of major roads on property values: An analysis of an existing and a proposed road corridor project- Thesis for completion of Master of Business (Economics)-Queensland University of Technology.

O’brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality and Quantity, 41(5), 673-690.

Oden, N. (1995). Adjusting Moran's I for population density. Statistics in Medicine, 14(1), 17-26.

Ostertagová, E., and Ostertag, O. (2012). Forecasting using simple exponential smoothing method. Acta Electrotechnica et Informatica, 12(3), 62-66.

Ottensmann, J. R., Payton, S., and Man, J. (2008). Urban location and housing prices within a hedonic model. Journal of Regional Analysis and Policy, 38(1), 19-35.

Ozuduru, B. H. and Varol Ç. (2011). Spatial Statistics Methods in Retail Location Research: A Case Study of Ankara, Turkey. Procedia Environmental Sciences 7:287-292.

Özüş, E. (2009). Determinants of office rents in the Istanbul Metropolitan Area. European Planning Studies, 17(4), 621-633.

Özüş, E., Dökmeci, V., Kiroğlu, G., Eğdemir, G. (2007). Spatial Analysis of Residential Prices in İstanbul. European Planning Studies 15(5), 705-721.

Pace, R. K., Barry, R., and Sirmans, C. F. (1998). Spatial statistics and real estate. The Journal of Real Estate Finance and Economics, 17(1), 5-13.

Pagourtzi, E., Nikolopoulos, K., and Assimakopoulos, V. (2006). Architecture for a real estate analysis information system using GIS techniques integrated with fuzzy theory. Journal of Property Investment and Finance, 24(1), 68-78.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., and French, N. (2003). Real estate appraisal: a review of valuation methods. Journal of Property Investment and Finance, 21(4), 383-401.

Paris, Q. (2001). Multicollinearity and maximum entropy estimators. Economics Bulletin, 3(11), 1-9.

Paterson, R. W. and Boyle, K. J. (2002). Out of sight, out of mind? Using GIS to incorporate visibility in hedonic property value models. Land economics, 78(3), 417-425.

Pavlov, A. D. (2000). Space varying regression coefficients: a semi-parametric approach applied to real estate markets. Real Estate Economics, 28(2), 249-283.

Pearson, L. J., Tisdell, C., and Lisle, A. T. (2002). The impact of Noosa National Park on surrounding property values: An application of the hedonic price method. Economic Analysis and Policy, 32, 155-172.

Propastin, P. A., & Kappas, M. (2008). Reducing uncertainty in modeling the NDVI-precipitation relationship: a comparative study using global and local regression techniques. GIScience & Remote Sensing, 45(1), 47-67.

Rahmatian, M., and Cockerill, L. (2004). Airport noise and residential housing valuation in southern California: A hedonic pricing approach. International Journal of Environmental Science and Technology, 1(1), 17-25.

Rice, J. (2006). Mathematical statistics and data analysis. Cengage Learning.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. The journal of political economy, 34-55.

Samapatti, S., and Tay, L. (2002). An Hedonic Price Model of New Housing in Indonesia. Pacific Rim Property Research Journal, 8(3), 203-11.

Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. Expert Systems with Applications, 36(2), 2843-2852.

Selim, S. (2011). Determinants of House Prices in Turkey: A Hedonic Regression Model= Türkiye'de Konut Fiyatlarının Belirleyicileri: Hedonik Regresyon Modeli. Doğuş Üniversitesi Dergisi, 9(1), 65-76.

Sharma, V., Kilic, A., Kabenge, I., and Irmak, S. (2011). Application of GIS and Geographically Weighted Regression to Evaluate the Spatial Non-Stationarity Relationships between Precipitation vs. Irrigated and Rainfed Maize and Soybean Yields.

Sirmans, G. S., Macpherson, D. A., and Zietz, E. N. (2005). The composition of hedonic pricing models. Journal of real estate literature, 13(1), 1-44.

Sirmans, G., S., and Macpherson, D., A. (2003b). The Value of Housing Characteristics, National Association of Realtors, National Center for Real Estate Research, Chicago.

Shultz, S. D., and King, D. A. (2001). The use of census data for hedonic price estimates of open-space amenities and land use. The Journal of Real Estate Finance and Economics, 22(2-3), 239-252.

Sparks, P. J. and Sparks, C. S. (2010). An application of spatially autoregressive models to the study of US county mortality rates. Population, Space and Place, 16(6), 465-481.

Suits, D. B. (1957). Use of dummy variables in regression equations. Journal of the American Statistical Association, 52(280), 548-551.

Sujjaviriyasup, T., and Pitiruek, K. (2013). Hybrid ARIMA-Support Vector Machine Model for Agricultural Production Planning. Applied Mathematical Sciences, 7(57), 2833-2840.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic Geography, 46 (2), 234-240.

Topcu, M., and Kubat, A. S. (2009). The analysis of urban features that affect land values in residential areas. In Proceedings of the 7th International Space Syntax Symposium, KTH, Stockholm (Vol. 26, No. 1, p. 26).

Türel, A. (1981). Ankara'da Konut Fiyatlarının Mekansal Farklılaşması. ODTÜ Mimarlık Fakültesi Dergisi, 7(1), 97-109.

Uluçay, Ç., and Tecim, V. (2009). Avrupa birliği ve e-devlet öncelikleri açisindan cbs tabanli değerleme haritalarinin üretimi. Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 1(1), 15-28.

Ustaoglu, E., Türel, A., and Güzel, A. (2013). Evaluating the Functional Utility Characteristics of Office Buildings: A Case Study of the Office Market in Ankara, Turkey. Journal of Real Estate Literature, 21(1), 79-103.

Vichiensan, V., Malaitham, S., and Miyamoto, K. (2011). Hedonic analysis of residential property values in Bangkok: Spatial dependence and nonstationarity effects. Journal of the Eastern Asia Society for Transportation Studies, 9, 886-899.

Visser, P., Van Dam, F., and Hooimeijer, P. (2008). Residential environment and spatial variation in house prices in the Netherlands. Tijdschrift voor economische en sociale geografie, 99(3), 348-360.

Weisberg, S. (2005). Applied linear regression (Vol. 528). John Wiley and Sons.

Wheeler, D. C. (2014). Geographically weighted regression. In Handbook of Regional Science (pp. 1435-1459). Springer Berlin Heidelberg.

Yalpir, Ş., and Unel, B. F. (2014). Multiple Regression Analysis for house valuation using GIS: a case study. Electrical Engineering and Information Technology, 63, 353.

Yalpir, Ş., and Özkan, G. (2011). Fuzzy logic methodology and multiple regressions for residential real-estates valuation in urban areas. Scientific Research and Essays, 6(12), 2431-2436.

Yankaya, U., and Çelik, H. M. (2005). İzmir Metrosunun Konut Fiyatları Üzerindeki Etkilerinin Hedonik Fiyat Yöntemi ile Modellenmesi. Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 20(2), 61-79.

Yetgin, F., and Lepkova, N. (2005). Housing Policies in Turkey and Lithuania (No. eres 2005_357). European Real Estate Society (ERES).

Yılmaz, H., Bulut, Z., Temurlenk, M. S., and Yesil, P. (2008). Determination of the Impact of Parks and Playgrounds on House Prices in the City of Erzurum, Turkey. International Journal of Natural and Engineering Sciences, 2(1), 47-51.

Yomralioglu, T., & Nisanci, R. (2004). Nominal asset land valuation technique by GIS. In Proceedings of the FIG Working Week.

Yu, D., Wei, Y. D., and Wu, C. (2007). Modeling spatial dimensions of housing prices in Milwaukee, WI. Environment and Planning B: Planning and Design, 34(6), 1085-1102.

Yusuf, A. A., and Resosudarmo, B. P. (2009). Does clean air matter in developing countries' megacities? A hedonic price analysis of the Jakarta housing market, Indonesia. Ecological Economics, 68(5), 1398-1407.

Zabel, J. E., and Kiel, K. A. (2000). Estimating the demand for air quality in four US cities. Land Economics, 174-194.

Zietz, J., Zietz, E. N., and Sirmans, G. S. (2008). Determinants of house prices: a quantile regression approach. The Journal of Real Estate Finance and Economics, 37(4), 317-333.

APPENDIX A

SUPPLEMENTARY INFORMATION

A.1. AIC calculation (Zucchini, 2000)

$$AIC=-2\log(L) + 2(k + 1) \tag{A.1}$$

L: likelihood
k: number of explanatory variables(1 is added to include the intercept)

A.2. Hedonic Pricing Formula (Bello and Moruf, 2010)

$$P= f(L, S, N) \tag{A.2}$$

A.3. Hedonic Pricing Formula (Bello and Moruf, 2010)

$$Price = Function(L, P, T) + \varepsilon \tag{A.3}$$

A.4. Hedonic Regression Formula (Sirmans et al, 2005)

$$LnValue = \alpha + \beta_{1i} S + \beta_{2i} N + \beta_{3i} E + \varepsilon_i \tag{A.4}$$

A.5. Hedonic Pricing Formula (Hwang, 2003).

$$P = \alpha + \beta 1X1+\beta 2X2+\beta 3X3 + e \tag{A.5}$$

A.6. Spatial Lag Formula

$Y = Xß + \sigma WY + e$

$Y = Xß + \sigma WY + WX\gamma + e$
$Y = Xß + U$
$U = \lambda WU + e$
$Y = Xß + XWY - XWXß + e$

A.7. GWR Formula

$$y_i = \beta_o(u_i, v_i) + \sum_k \beta_k(u_i, v_i)\, x_{ik} + \epsilon_i$$

A.8. Detailed GWR model in matrix notation (Fotheringham, *et al.*, 2002)

$$Y = (\beta \otimes X)1 + \epsilon$$

$\otimes$: a logical operator that multiplies each element of β with the corresponding element of X

Y: a vector of dependent variable (nx1)

X: a matrix of independent variables (nx(k+1))

β: a matrix of local coefficients (nx(k+1))

ϵ : a vector of errors (nx1)

1: a vector of 1s ((k+1)x1)

n: number of data points

k: number of explanatory variables

$$\beta = \begin{bmatrix} \beta_o(u_1, v_1) & \beta_1(u_1, v_1) & \cdots & \beta_k(u_1, v_1) \\ \beta_o(u_2, v_2) & \beta_1(u_2, v_2) & \cdots & \beta_k(u_2, v_2) \\ \cdots & \cdots & \cdots & \cdots \\ \beta_o(u_n, v_n) & \beta_1(u_n, v_n) & \cdots & \beta_k(u_n, v_n) \end{bmatrix}$$

$$\hat{\beta}(i) = (X^T W(i) X)^{-1}\, X^T W(i) Y$$

$\hat{\boldsymbol{\beta}}(\boldsymbol{i})$: estimation of β at location i

$W(i)$: weight matrix for location i

(u_n, v_n): coordinates of regression point n

$$W(i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix}$$

w_{in}: weight given to data point n in the calibration of the model for location i

A.9. Calculation of R^2adj (Wesolowsky, 1976)

$$R^2{}_{adj} = 1 - \frac{n-1}{n-(k+1)}(1-R^2)$$

R^2: coefficient of determination

n: number of samples (observations)

k: number of explanatory variables(1 is added to include the intercept)

APPENDIX B

FIGURES

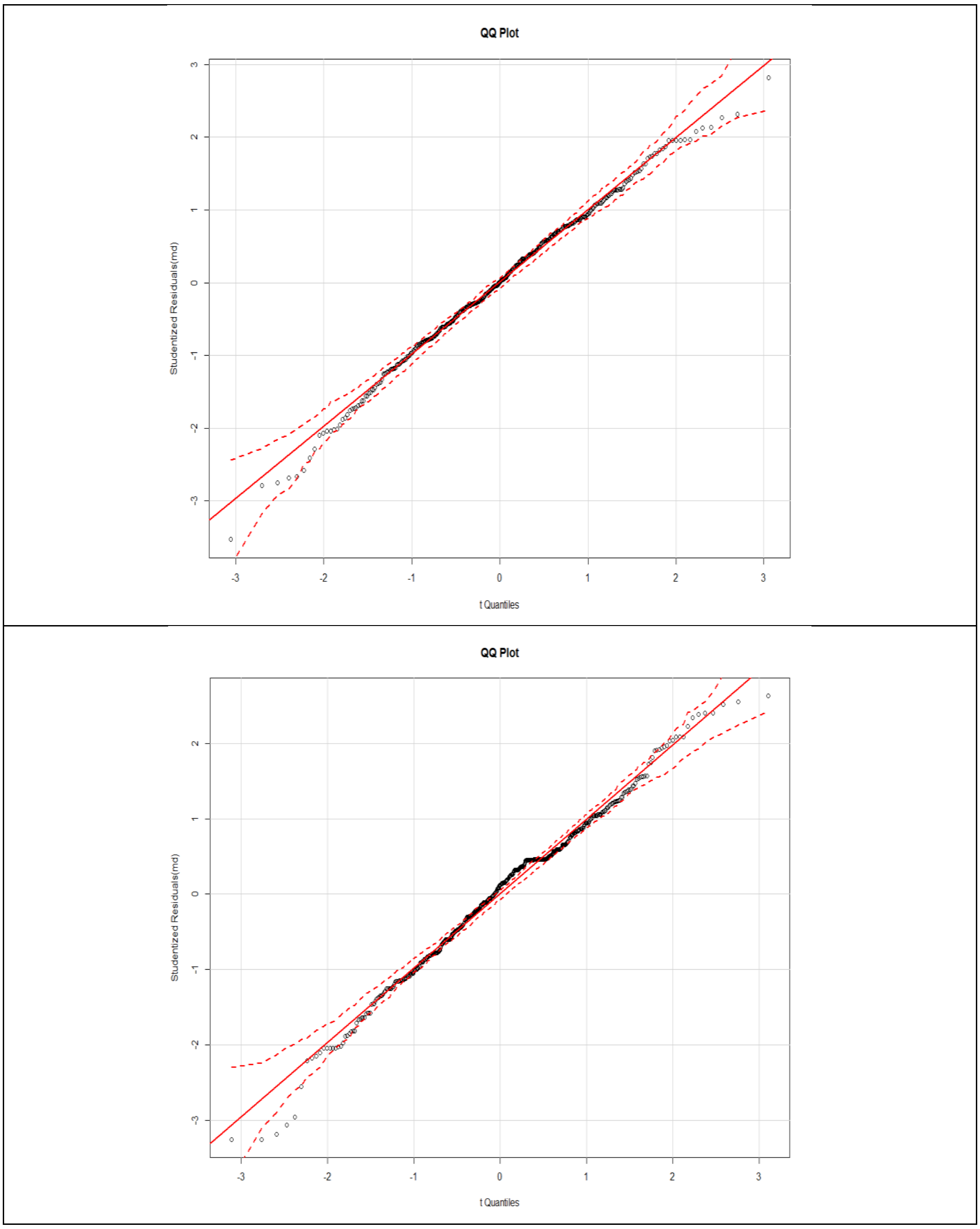
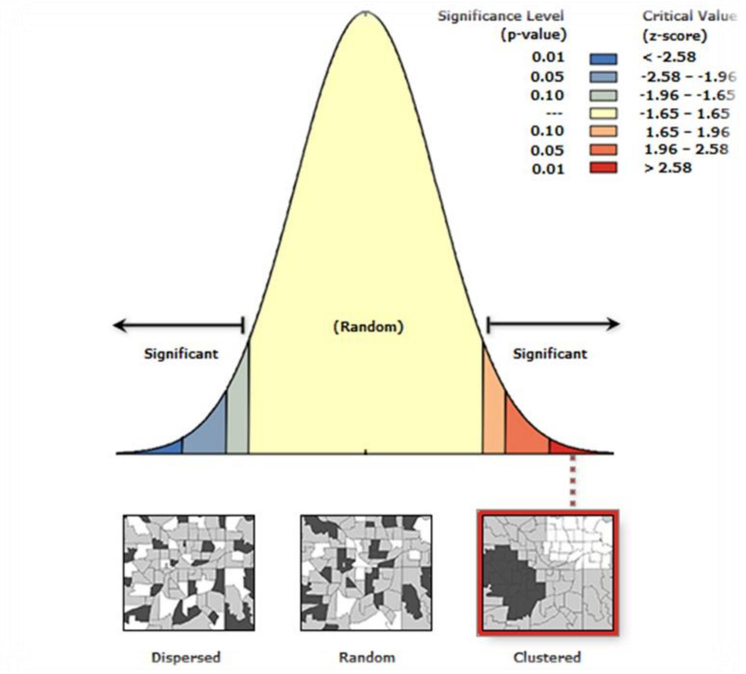


Figure 59: LnValue QQ-plot displaying normal distribution results after transformation for Çankaya (top) and Keçiören (bottom)

Çankaya



Keçiören

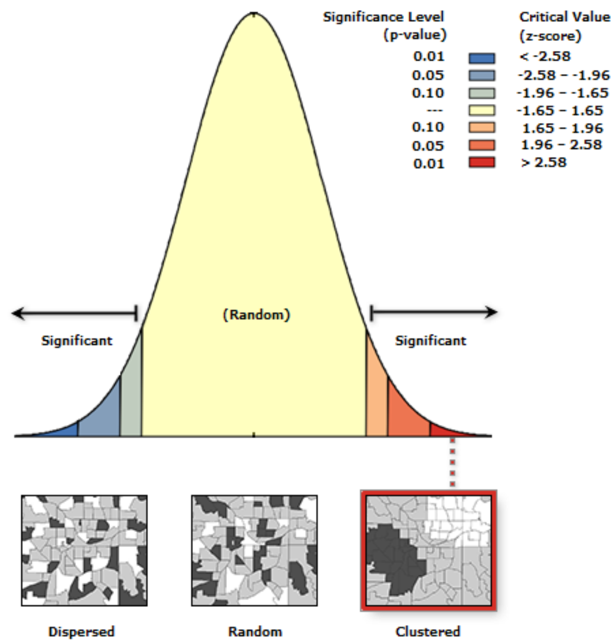


Figure 60: Moran's I Results

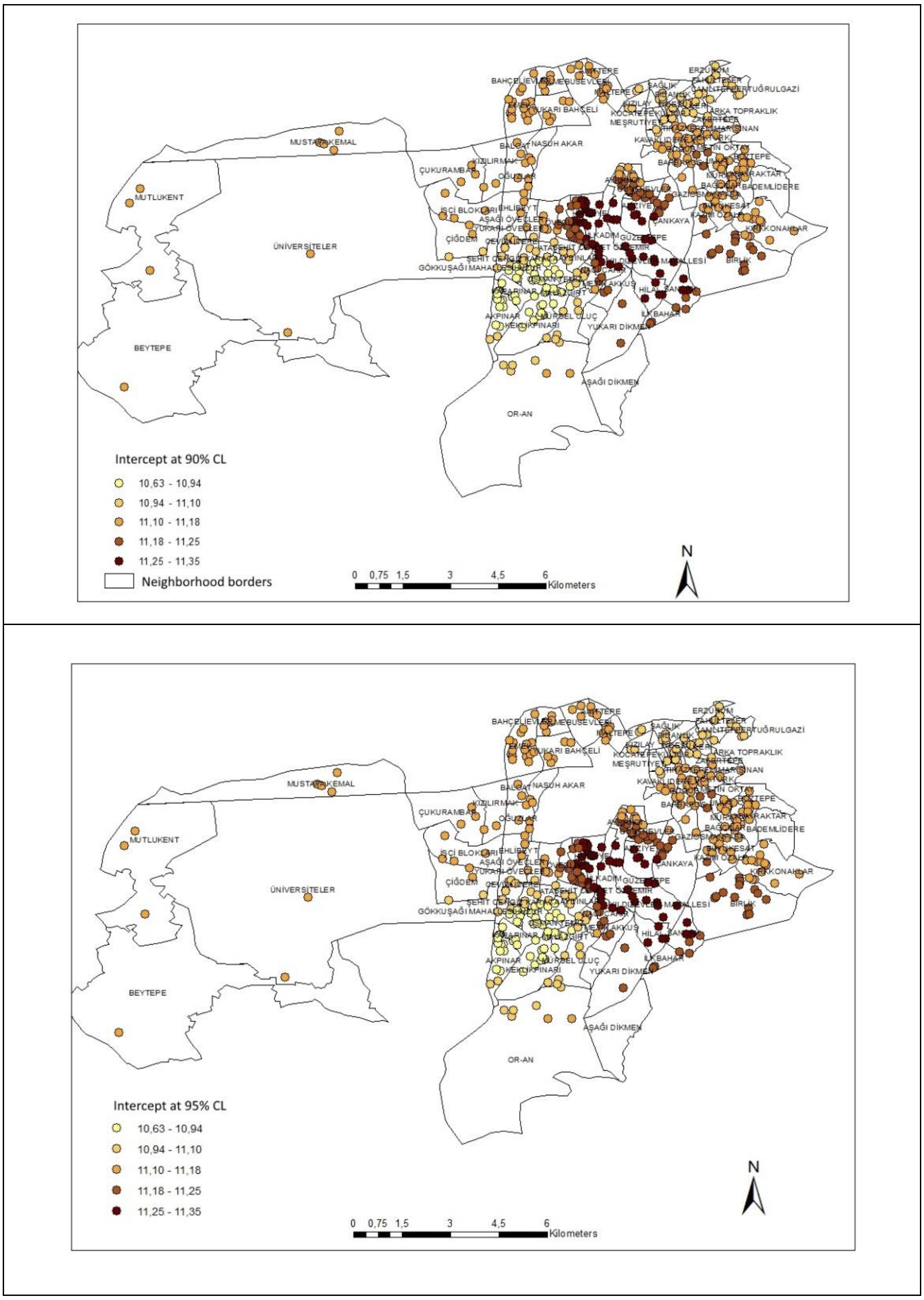


Figure 61: GWR Results for Intercept Coefficients for Çankaya

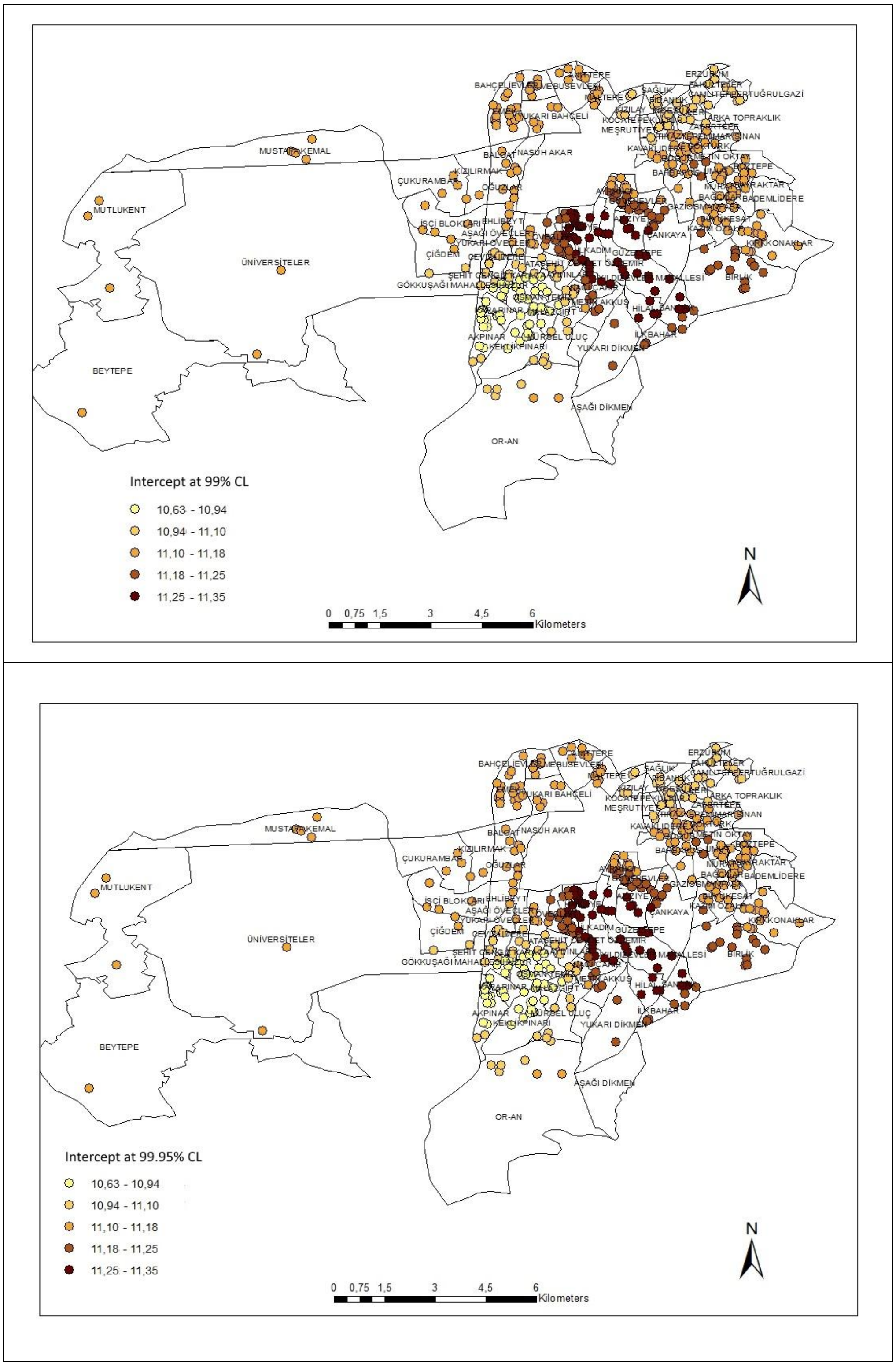


Figure 61 (continued): GWR Results for Intercept Coefficients for Çankaya

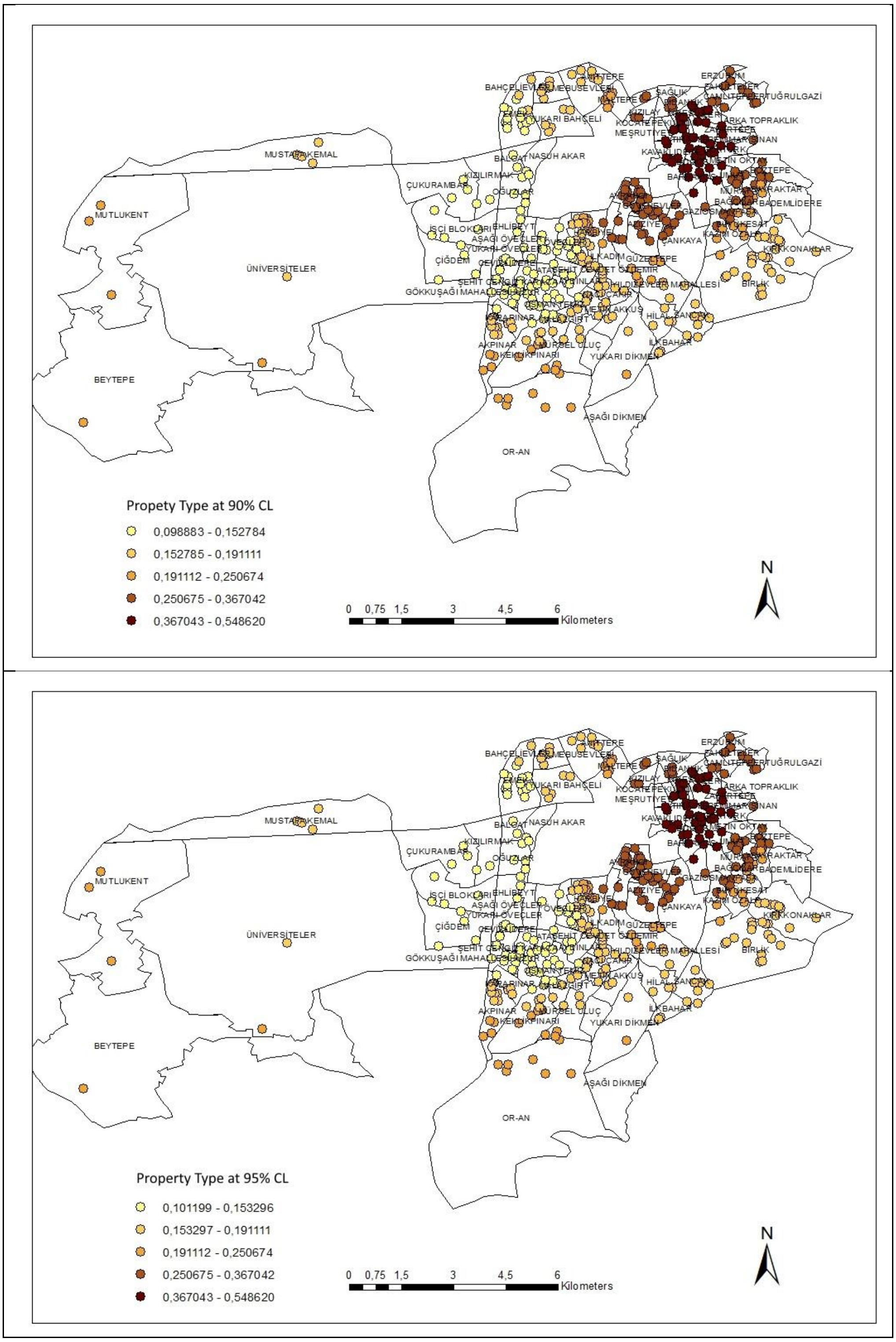


Figure 62: GWR Results for Property Type Coefficients for Çankaya

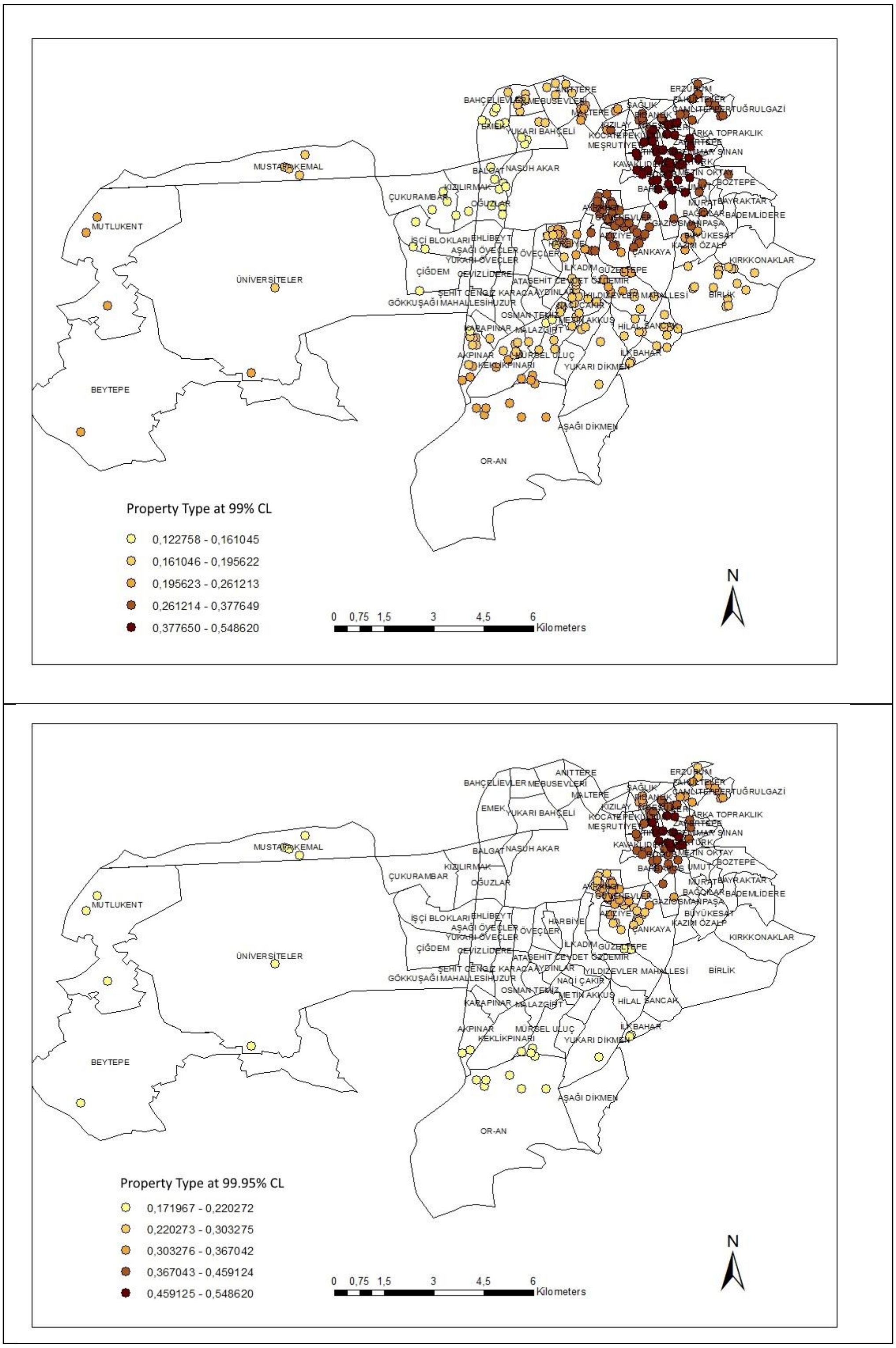


Figure 62 (continued): GWR Results for Property Type Coefficients for Çankaya

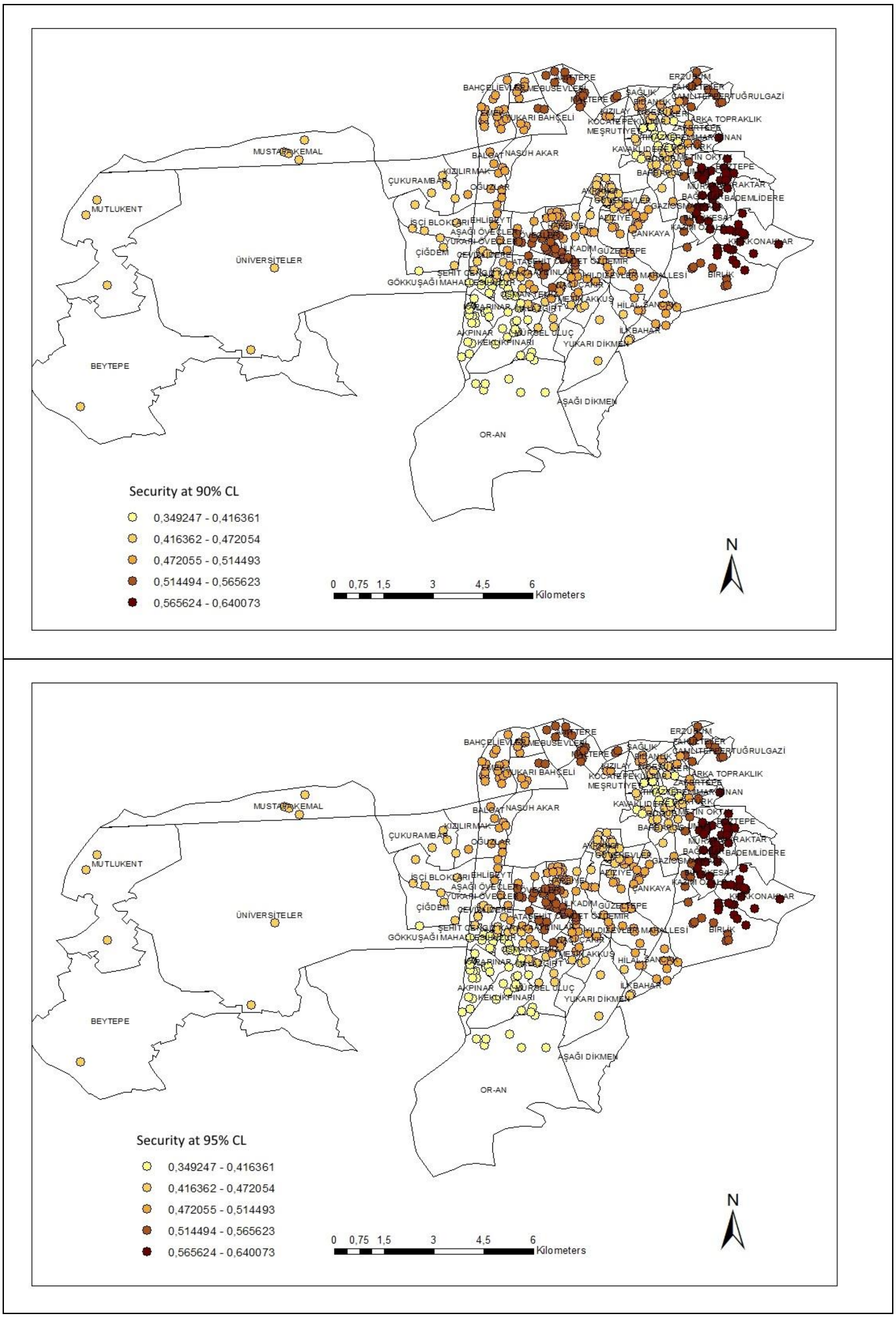


Figure 63: GWR Results for Security Coefficients for Çankaya

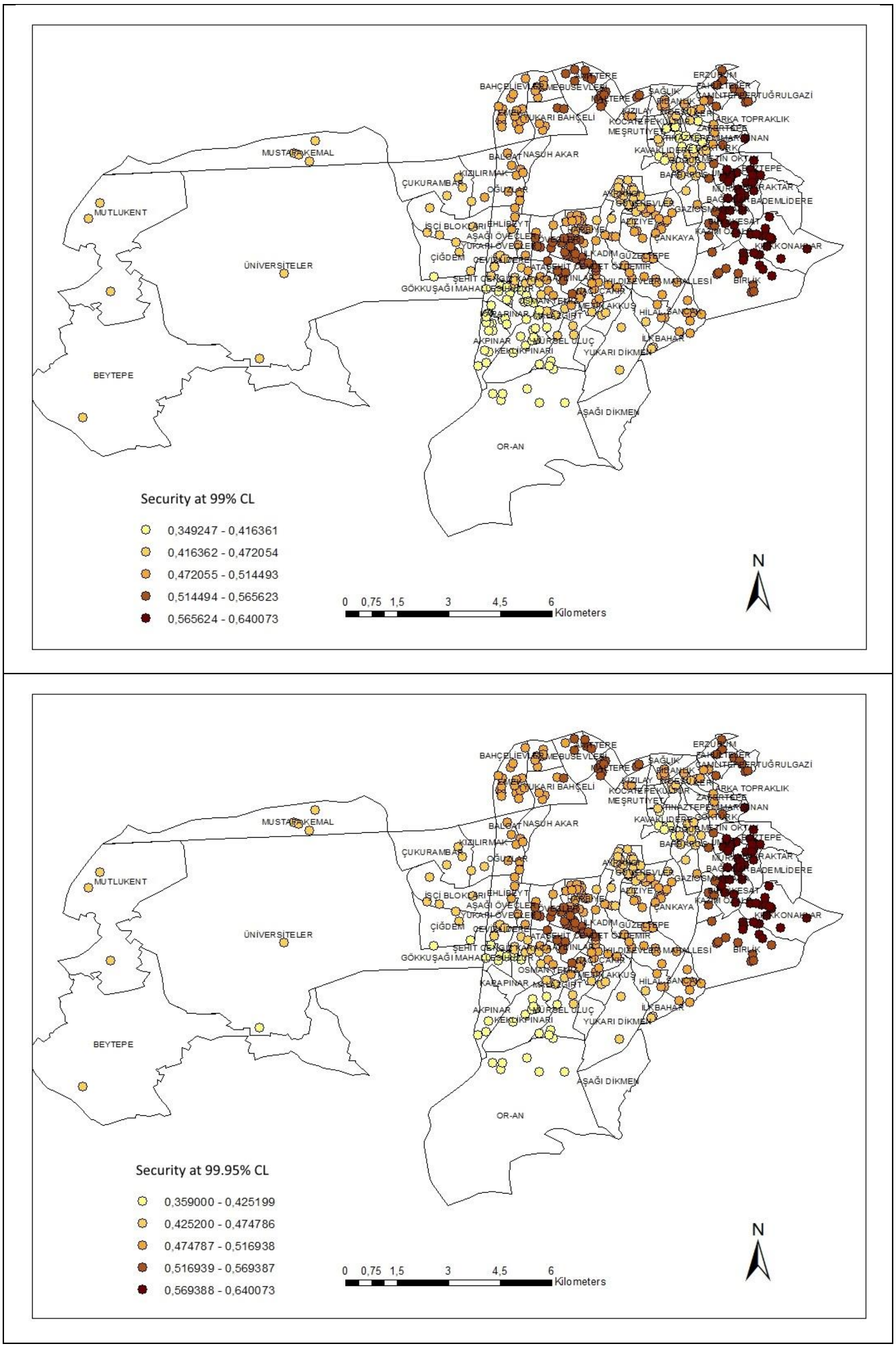


Figure 63 (continued): GWR Results for Security Coefficients for Çankaya

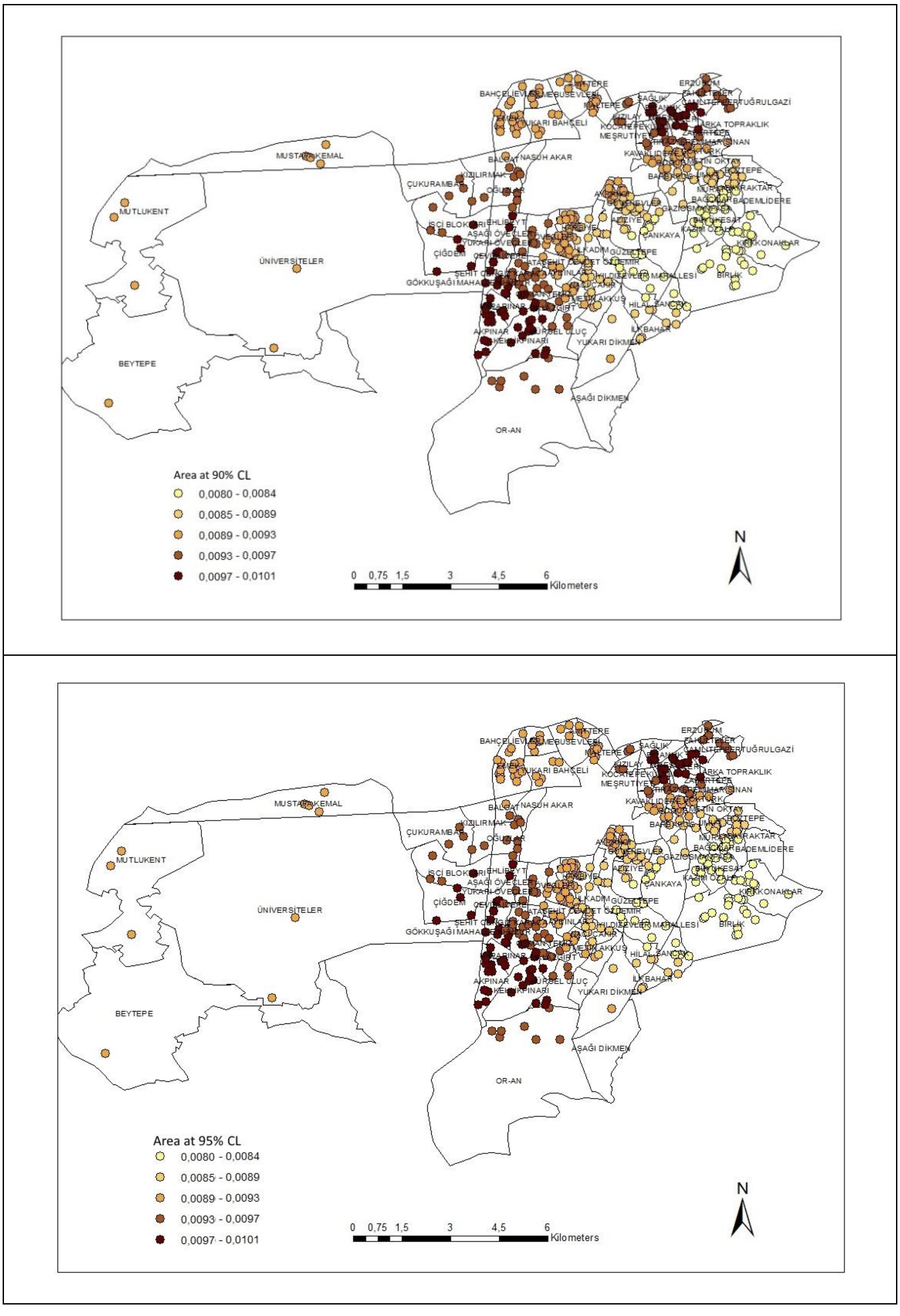


Figure 64: Results for Area Coefficients for Çankaya

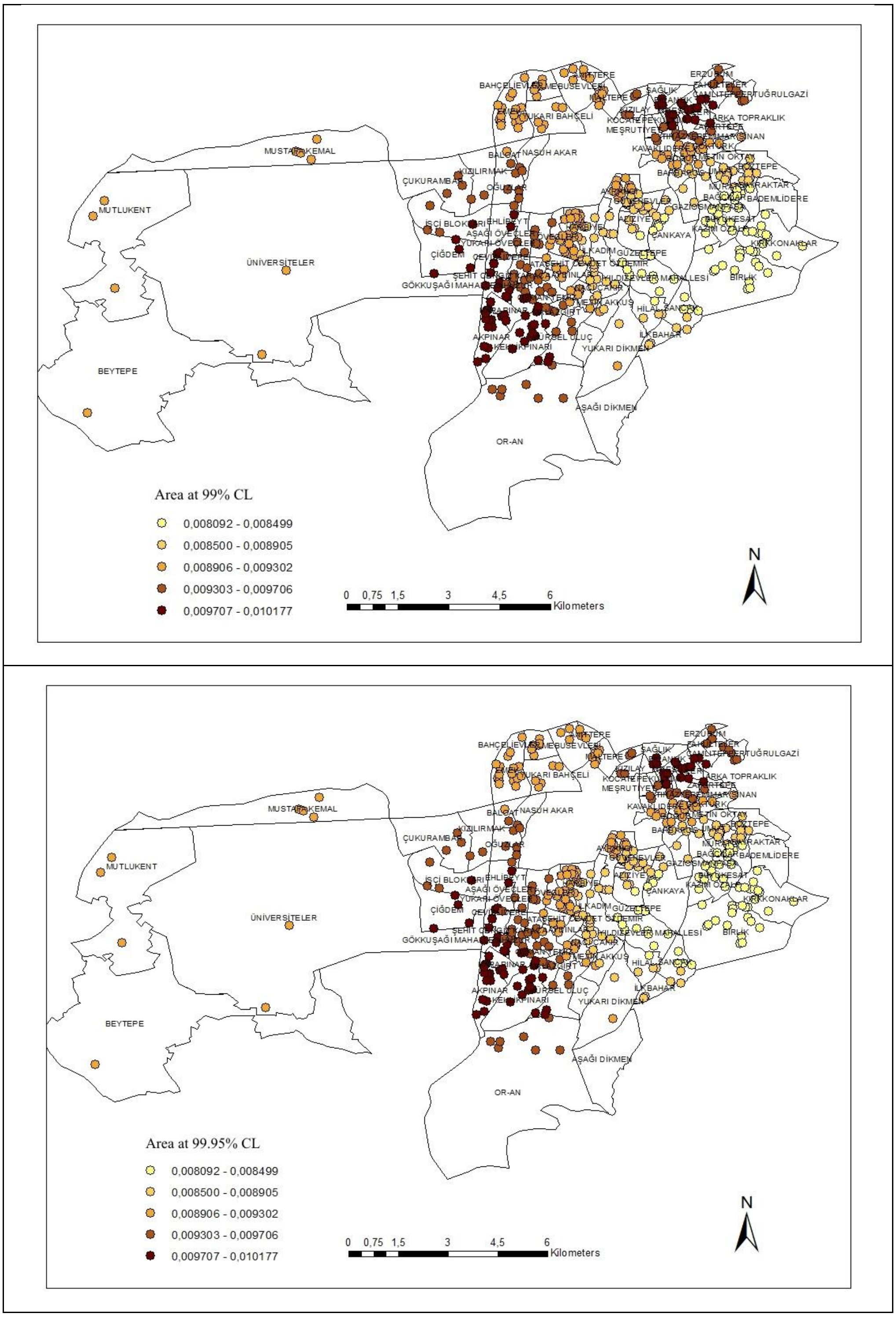


Figure 64 (continued): Results for Area Coefficients for Çankaya

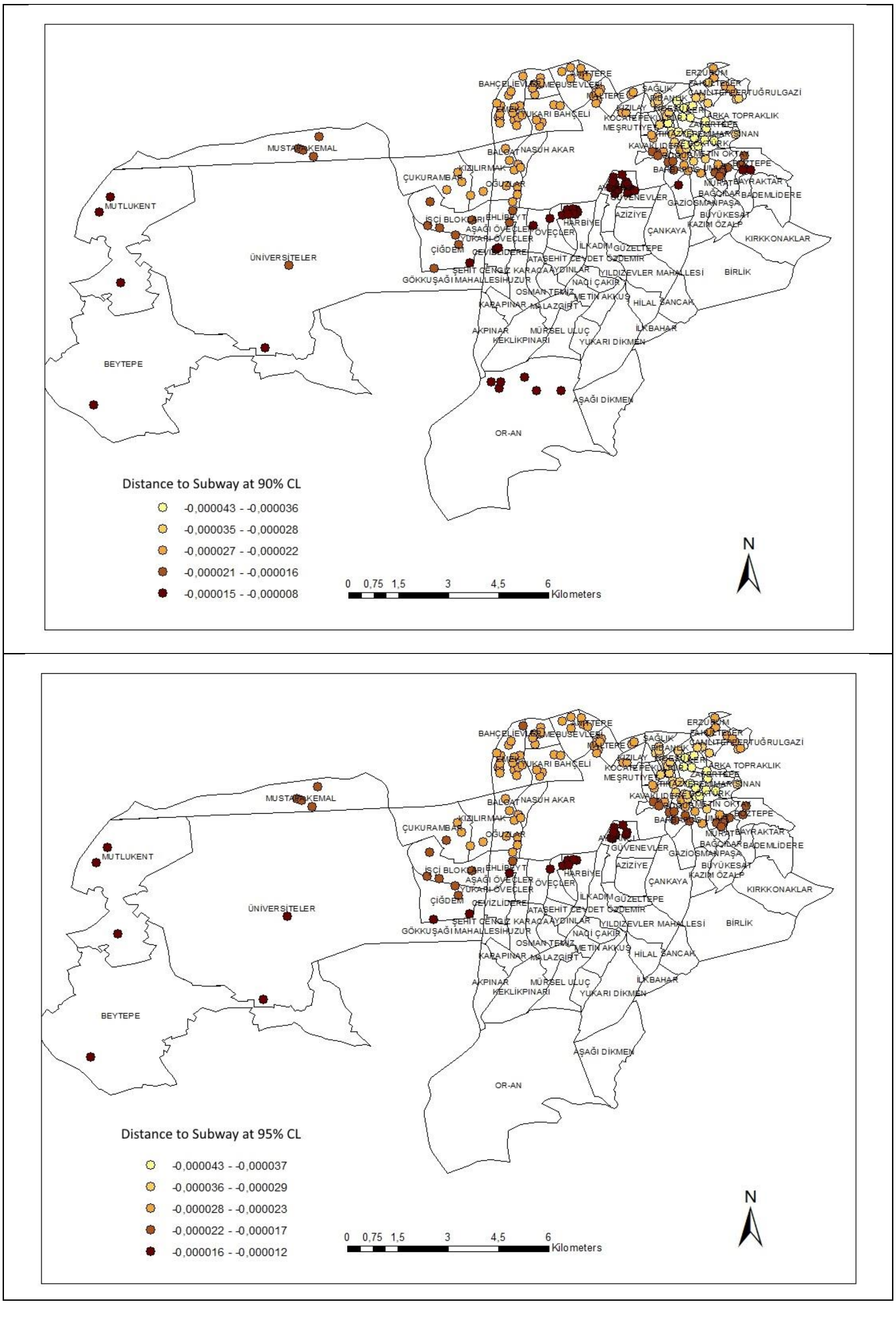


Figure 65: GWR Results for Distance to Subway Coefficients for Çankaya

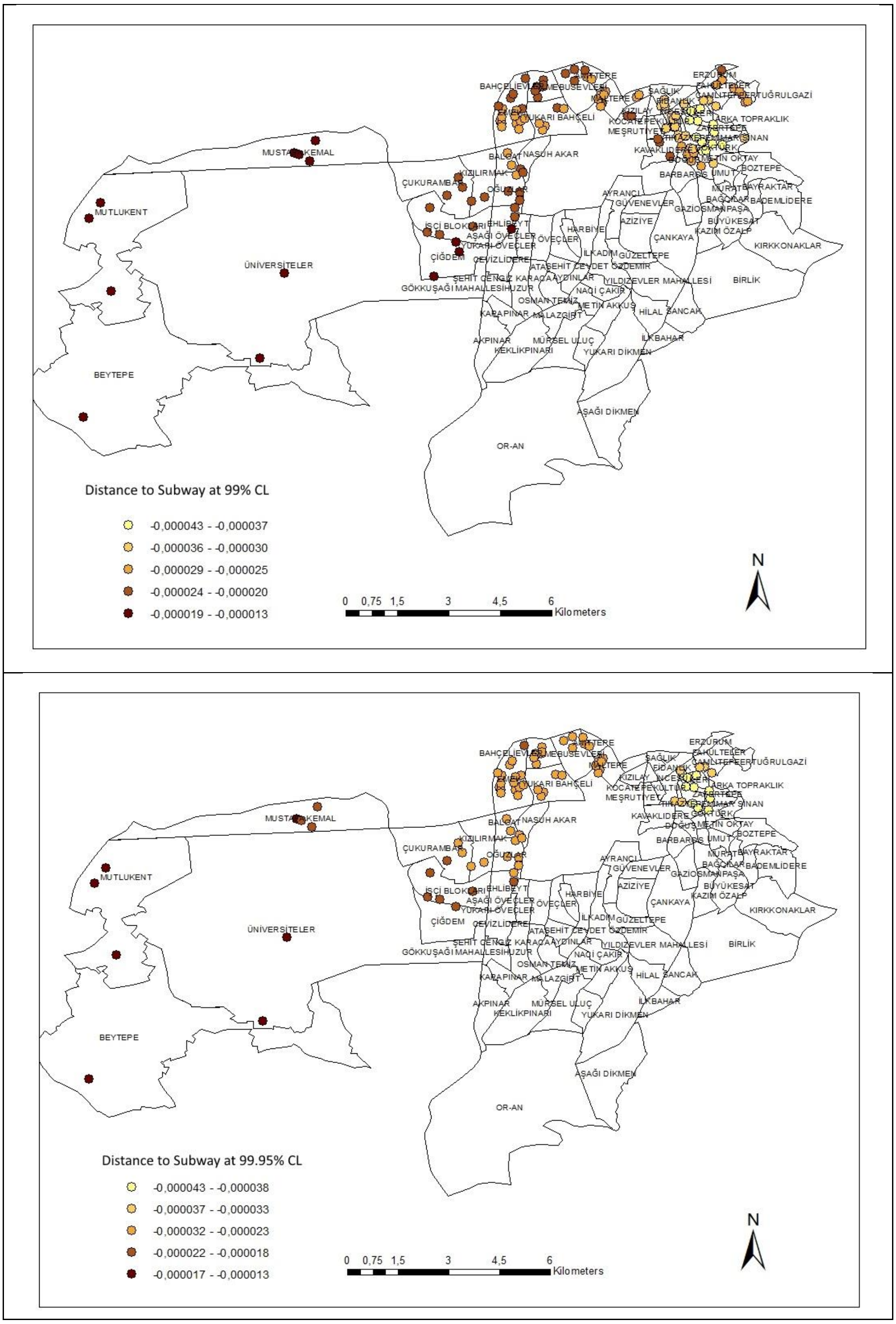


Figure 65 (continued): GWR Results for Distance to Subway Coefficients for Çankaya

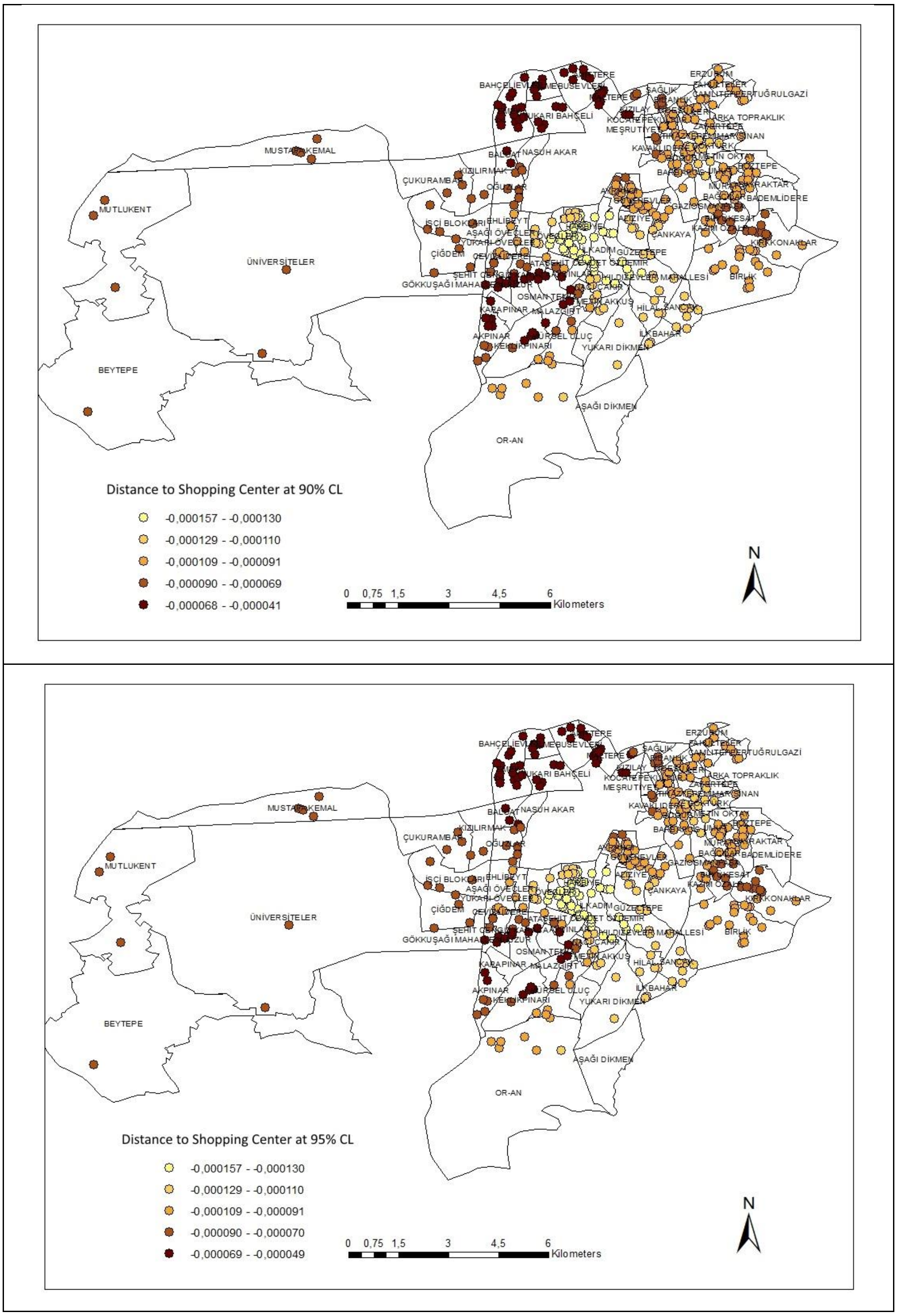


Figure 66: GWR Results for Distance to Shopping Center Coefficients for Çankaya

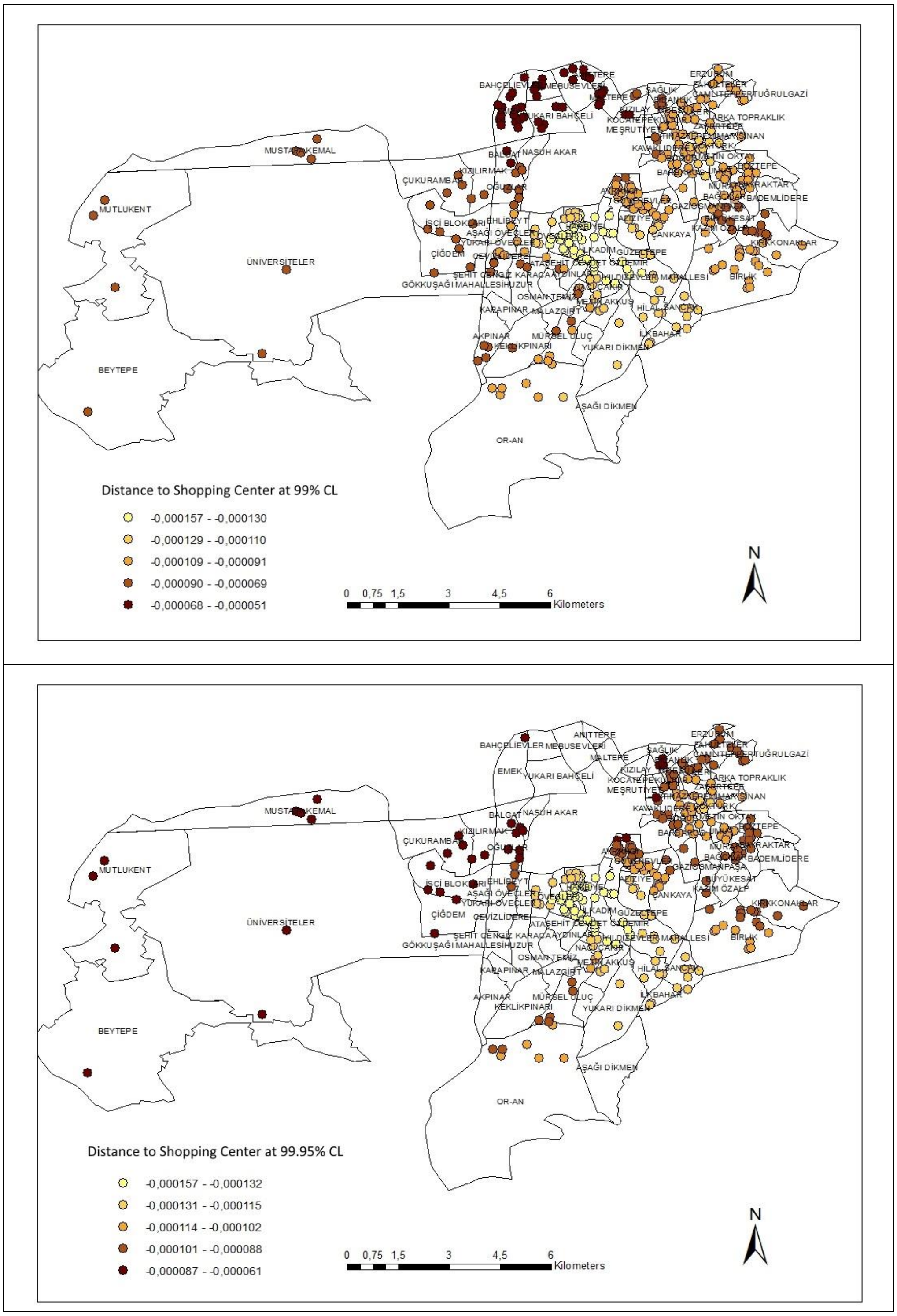


Figure 66 (continued): GWR Results for Distance to Shopping Center Coefficients for Çankaya

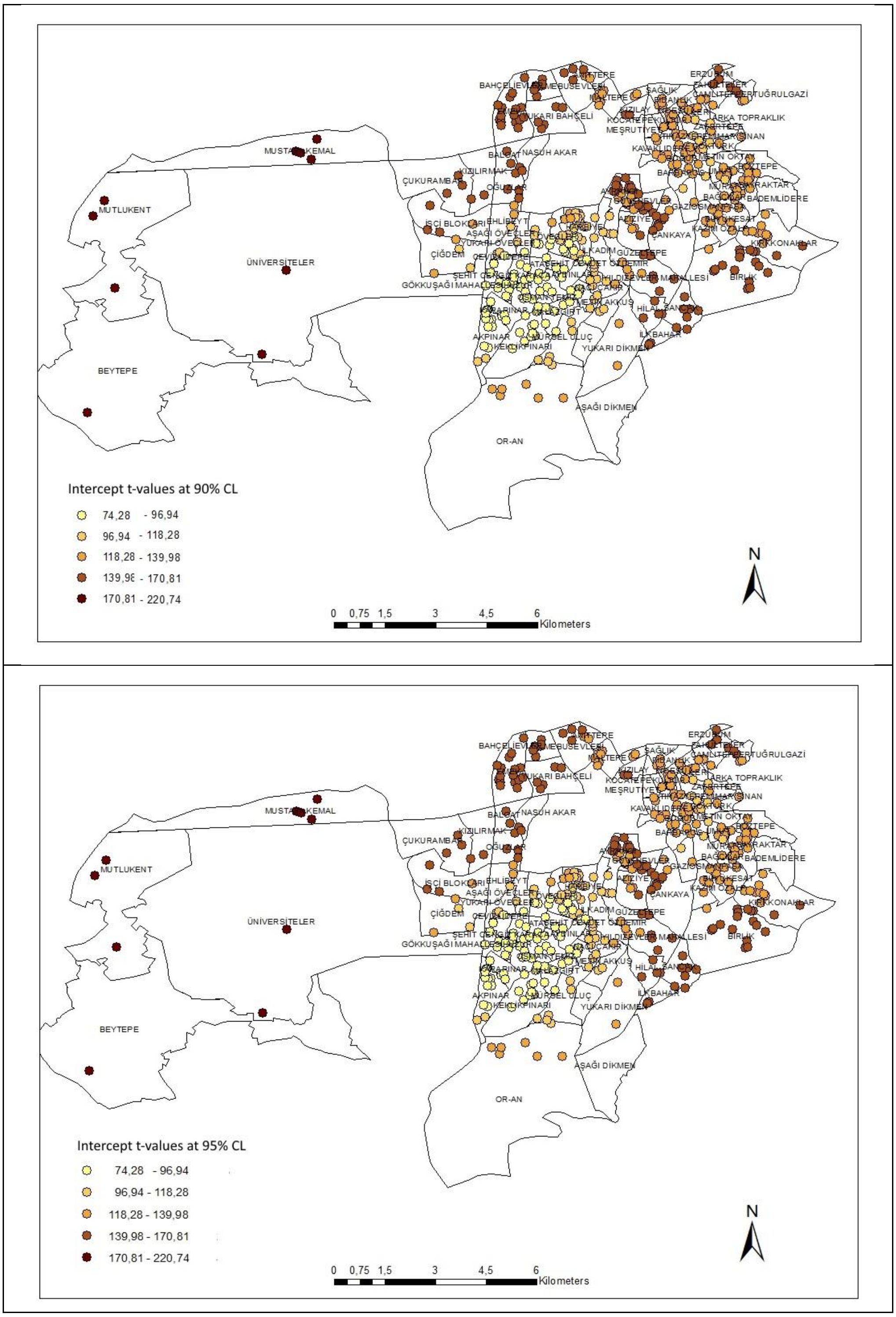


Figure 67: GWR t-Values for Intercept Coefficients for Çankaya

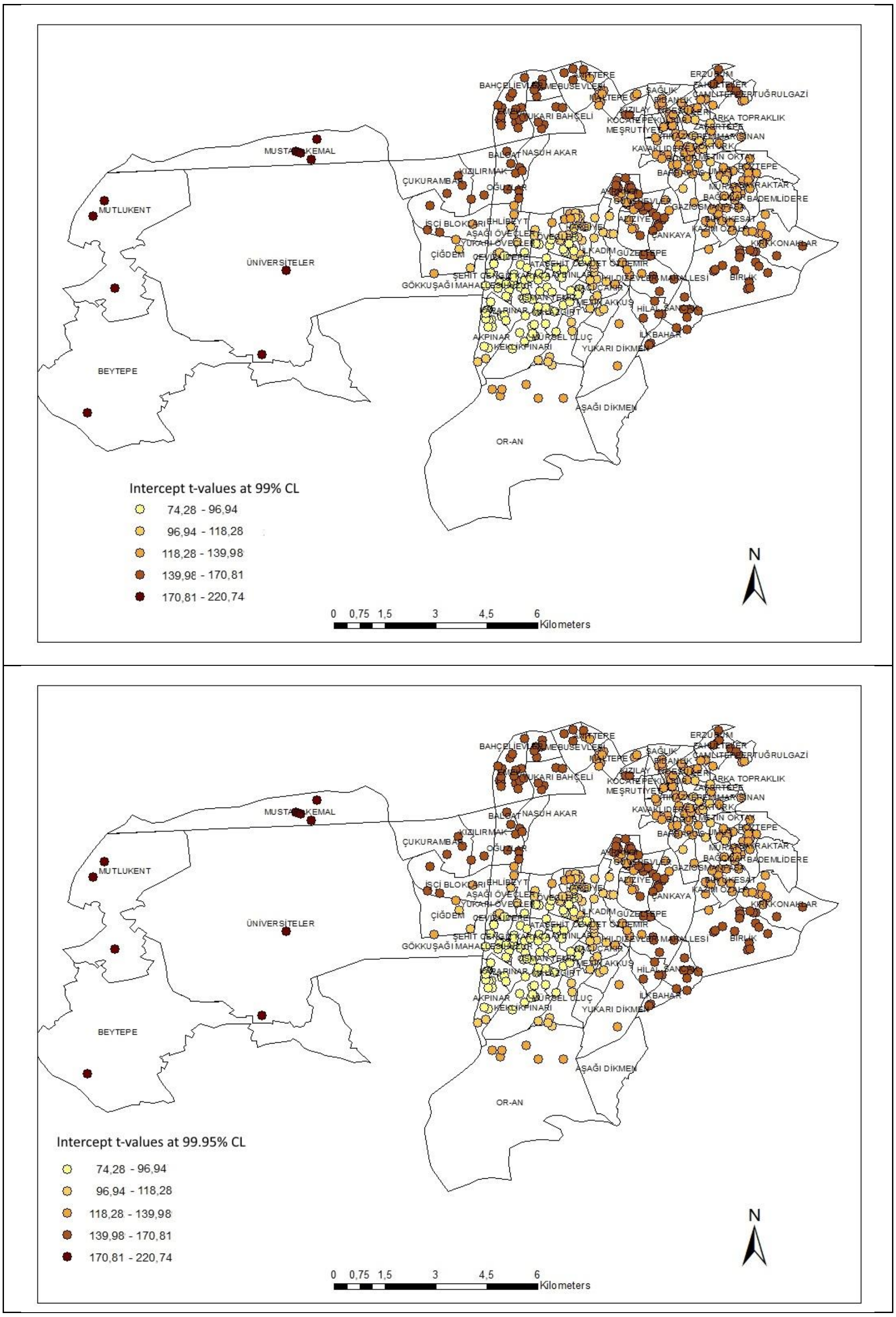


Figure 67 (continued): GWR t-Values for Intercept Coefficients for Çankaya

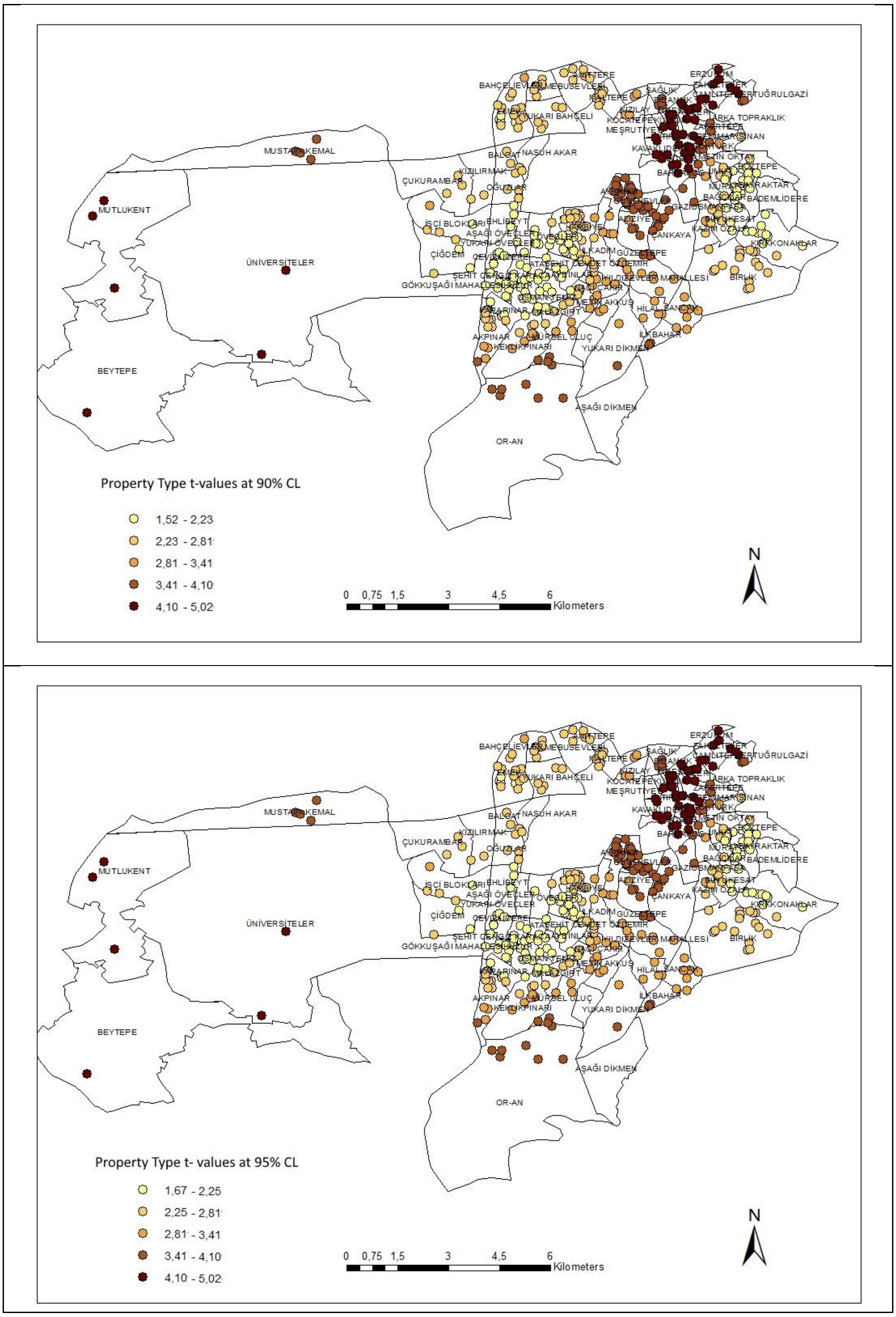


Figure 68: GWR t-Values for Property Type Attribute for Çankaya

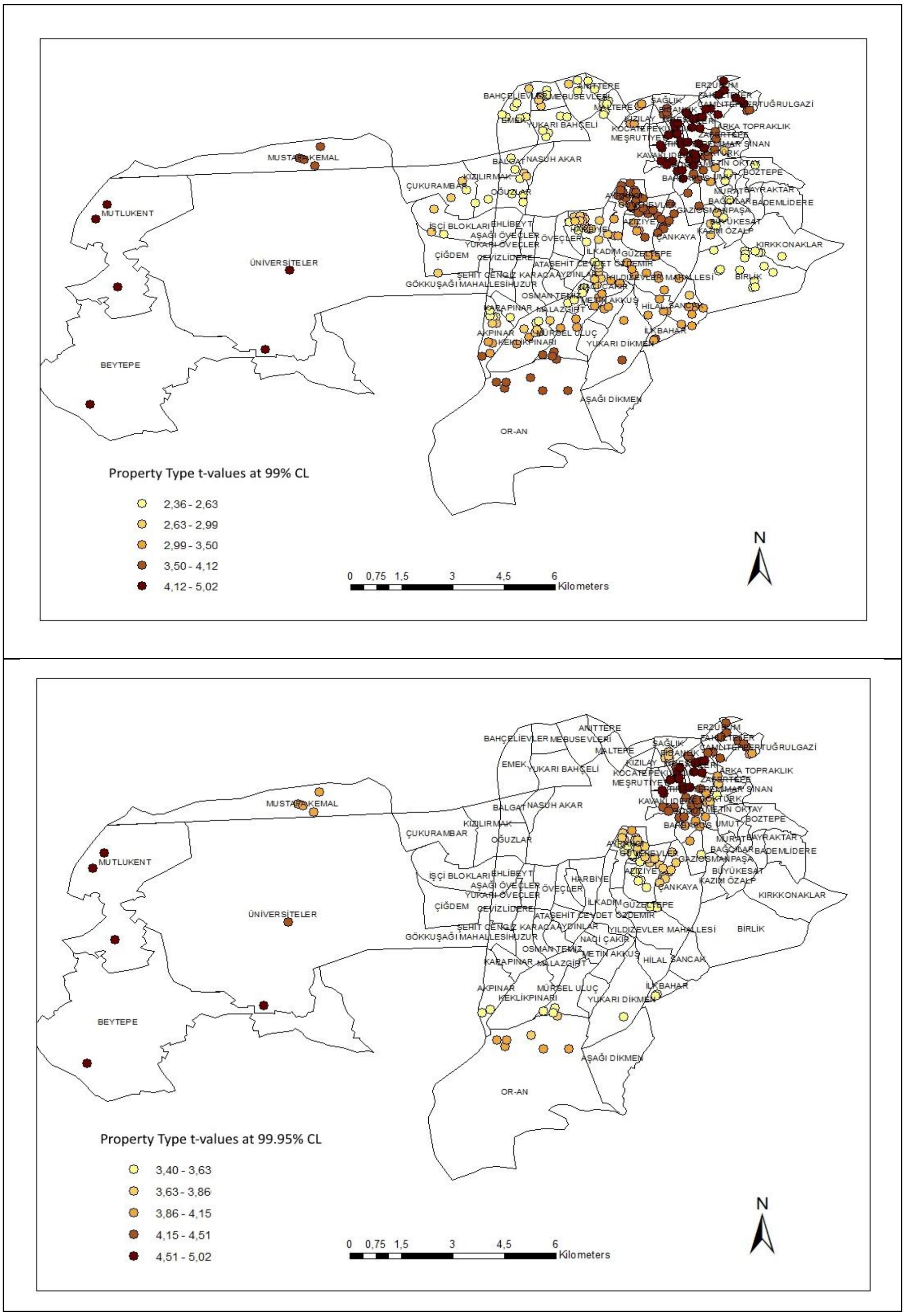


Figure 68 (continued): GWR t-Values for Property Type Attribute for Çankaya

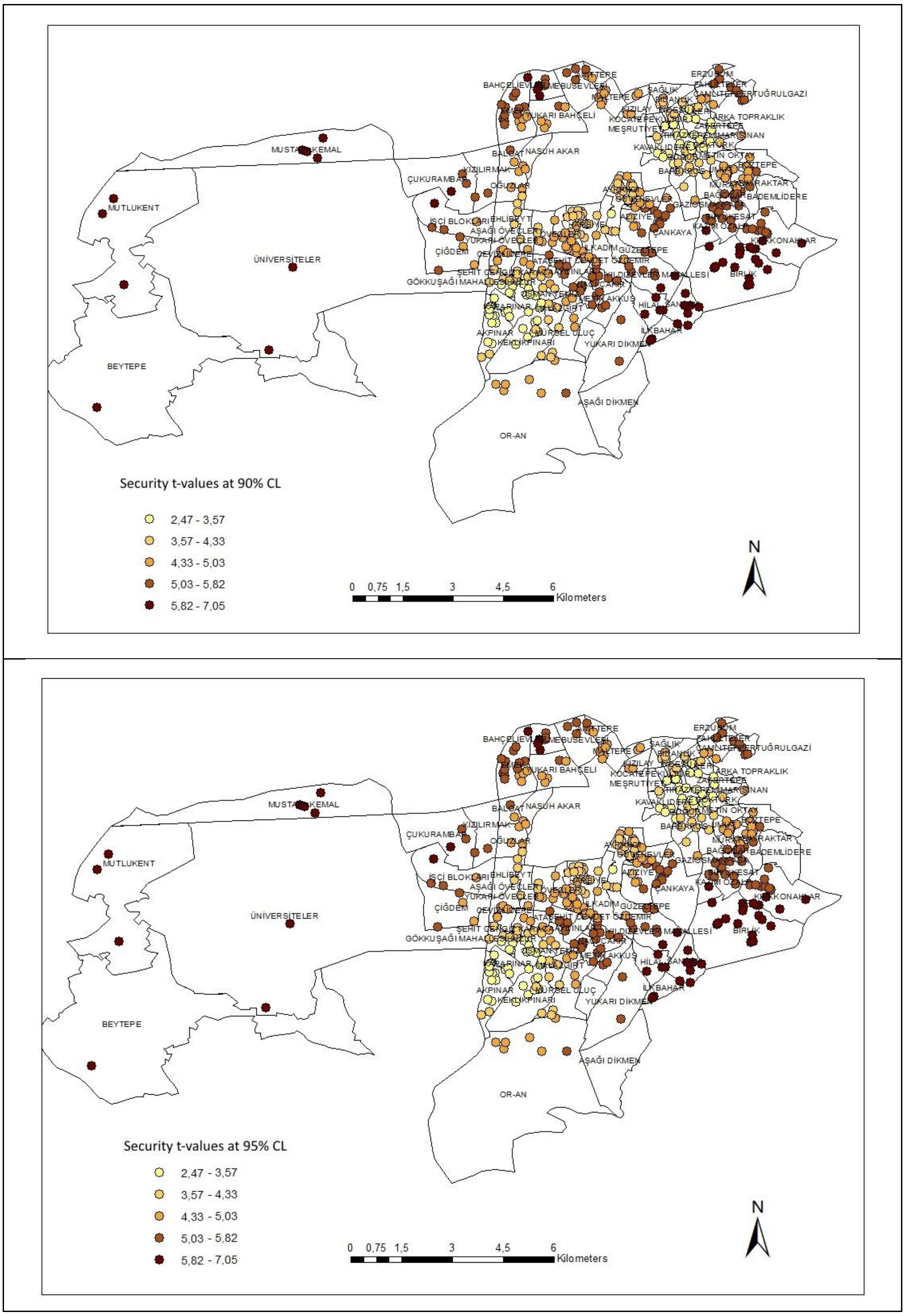
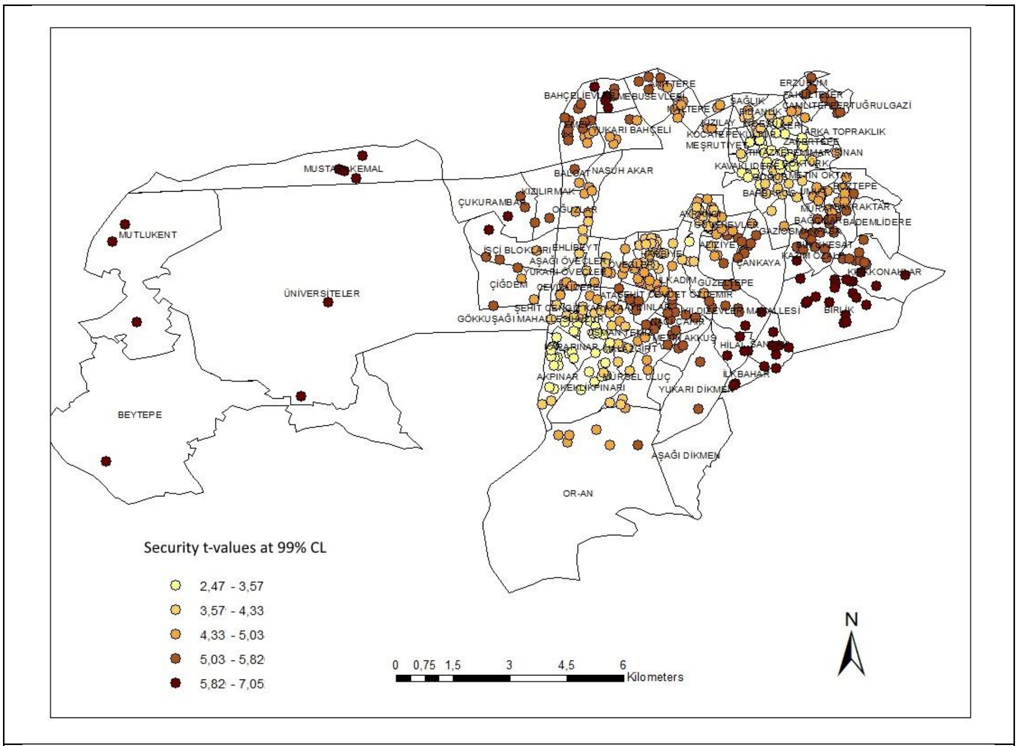


Figure 69: GWR t-Values for Security Attribute for Çankaya



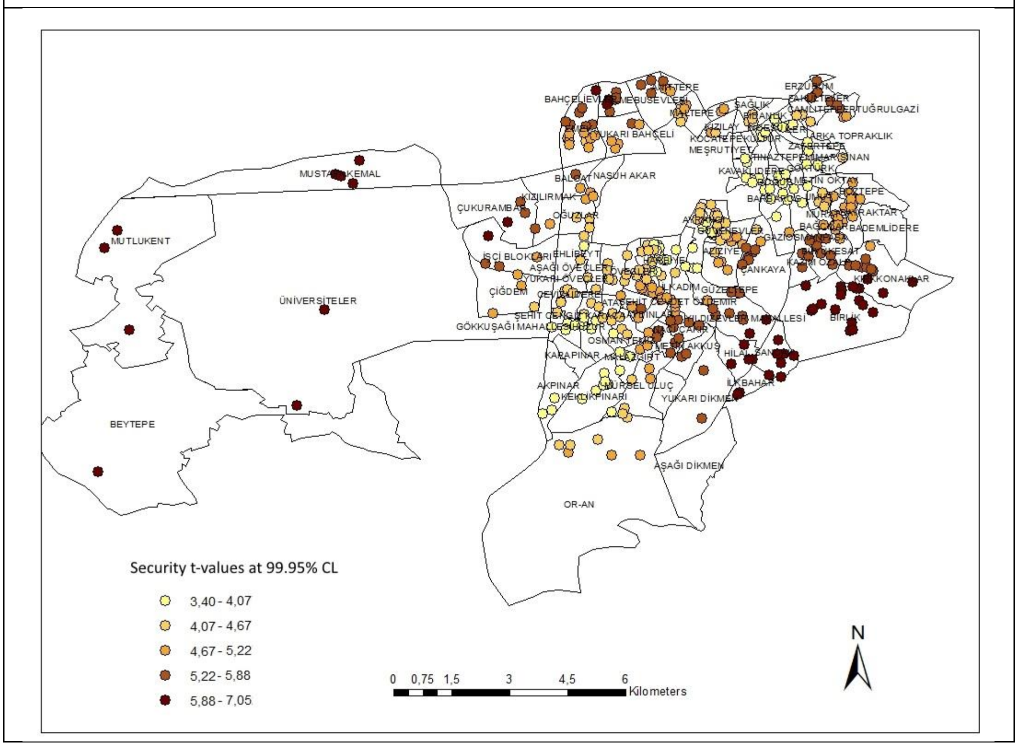


Figure 69 (continued): GWR t-Values for Security Attribute for Çankaya

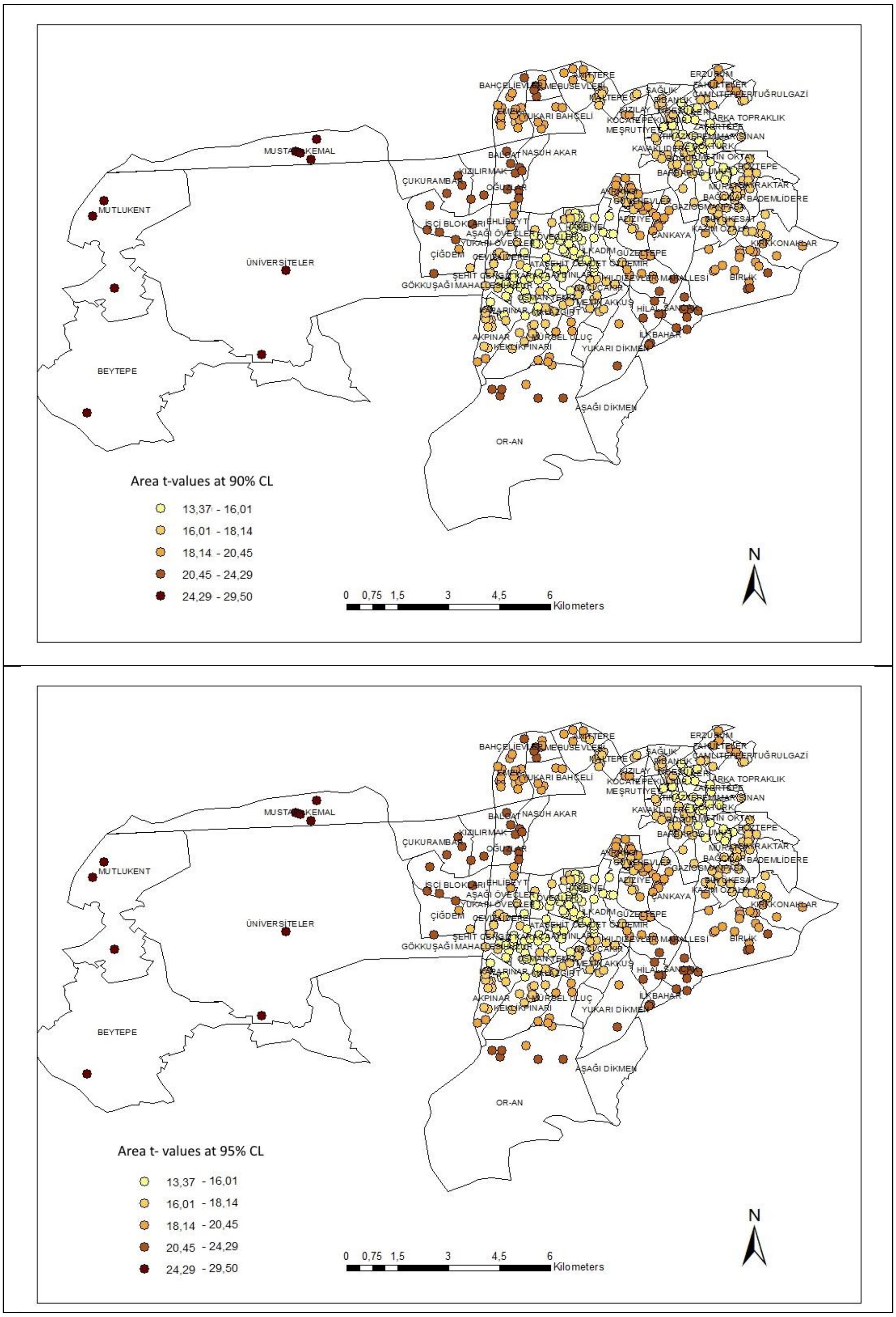


Figure 70: GWR t-Values for Area Type Attribute for Çankaya

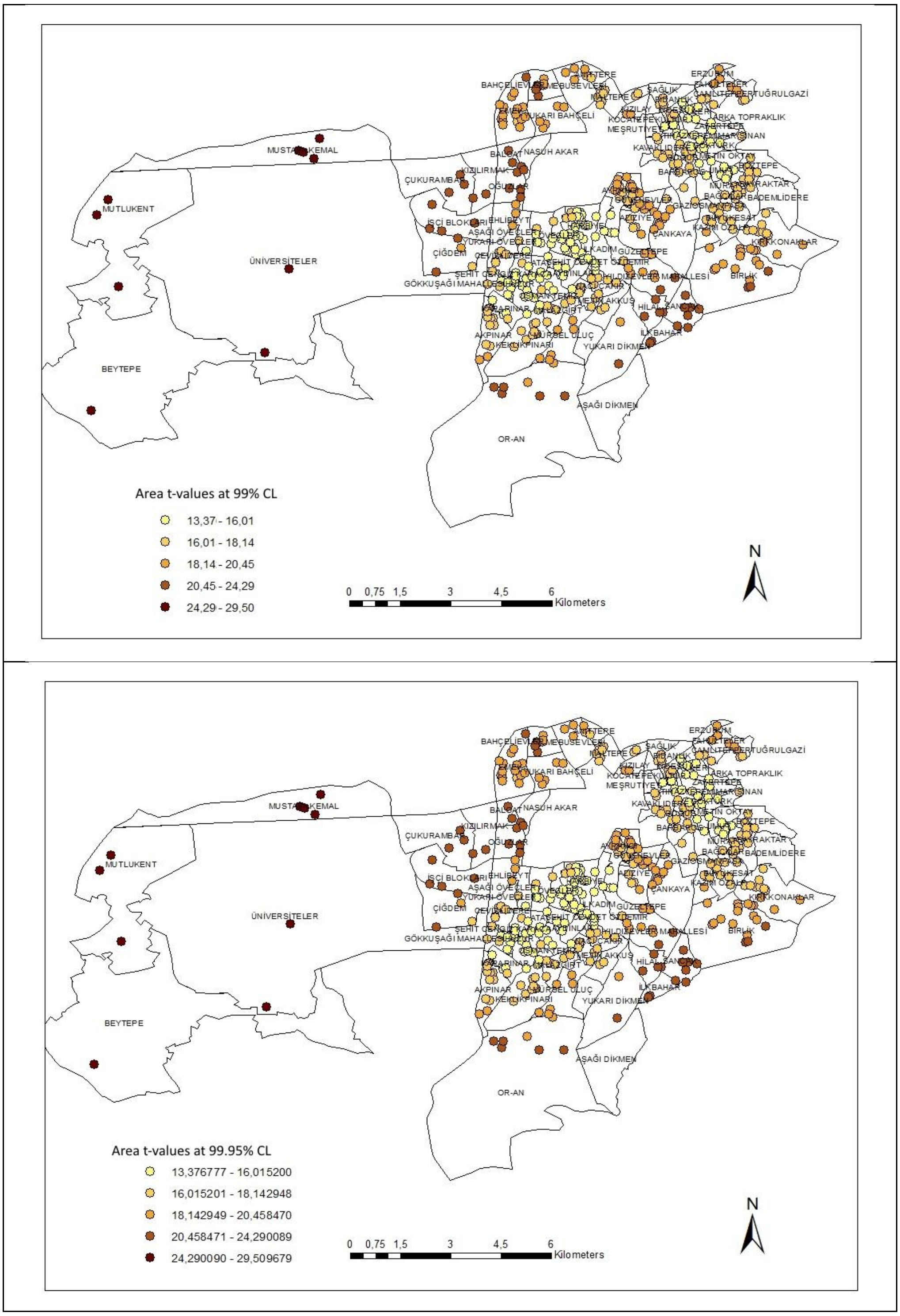


Figure 70 (continued): GWR t-Values for Area Type Attribute for Çankaya

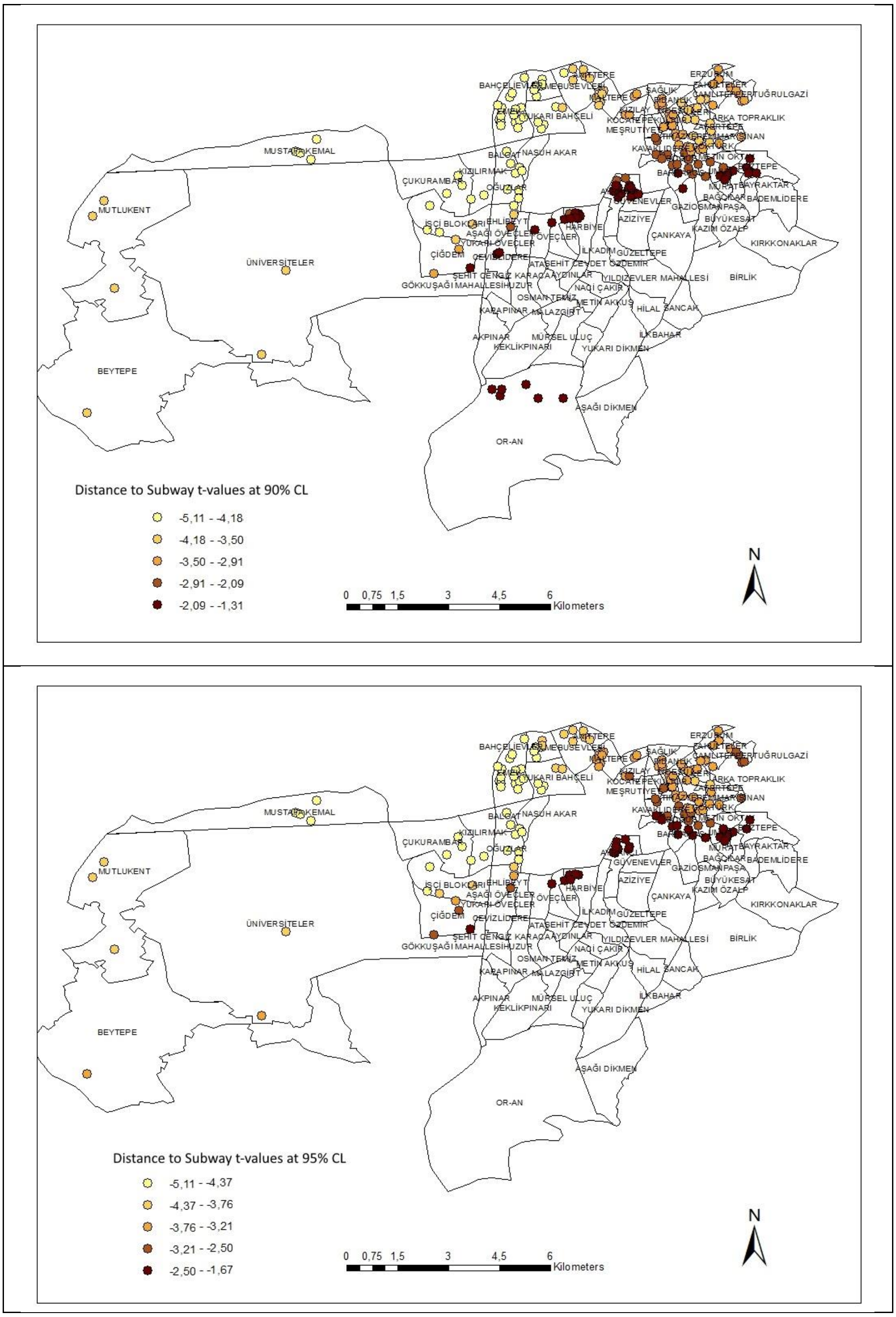


Figure 71: GWR t-Values for Distance to Subway Attribute for Çankaya

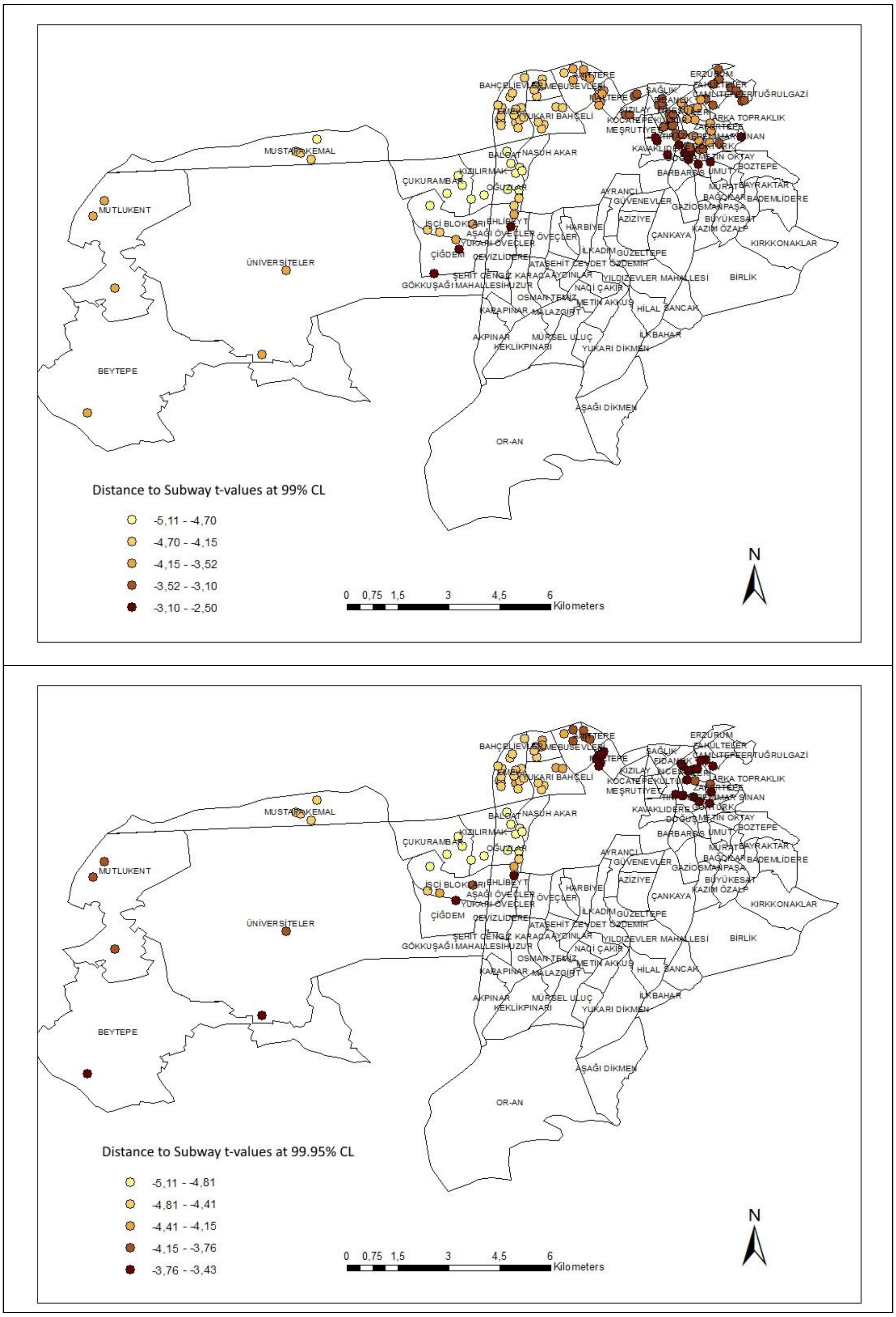


Figure 71 (continued): GWR t-Values for Distance to Subway Attribute for Çankaya

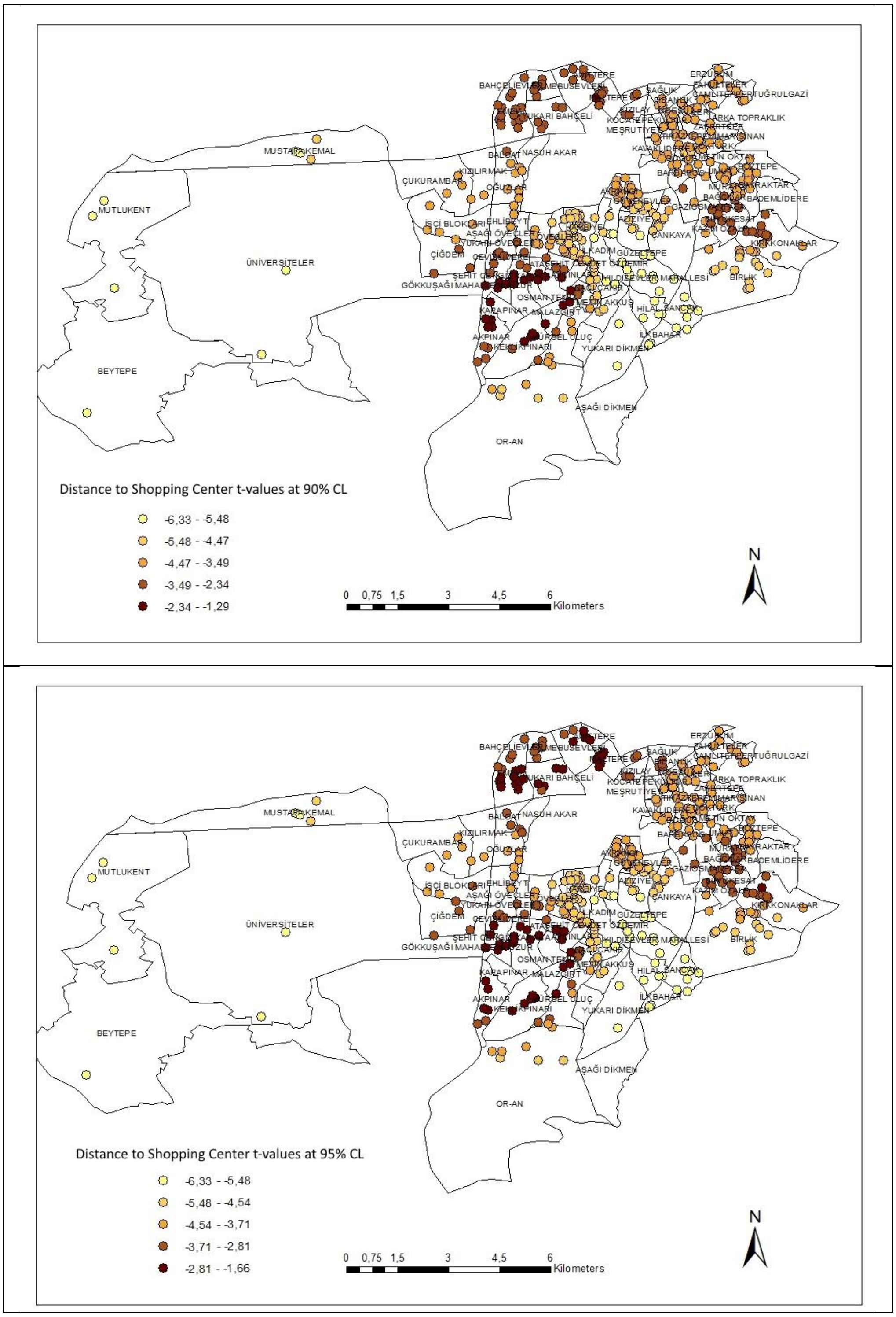
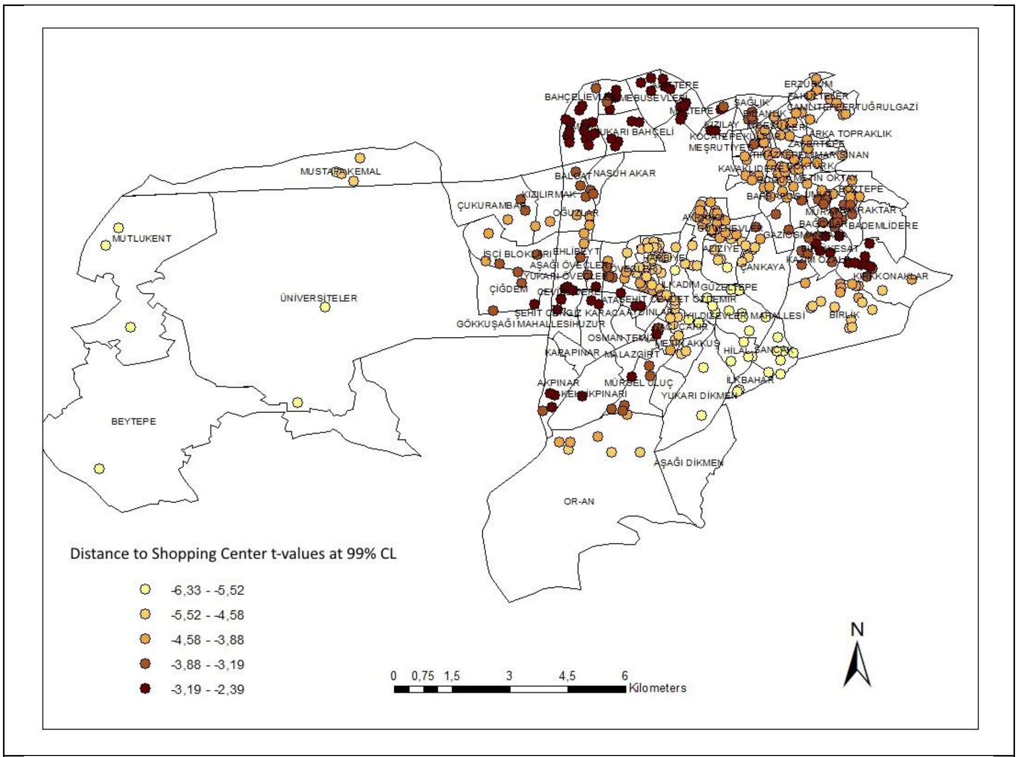


Figure 72: GWR t-Values for Distance to Shopping Center Attribute for Çankaya



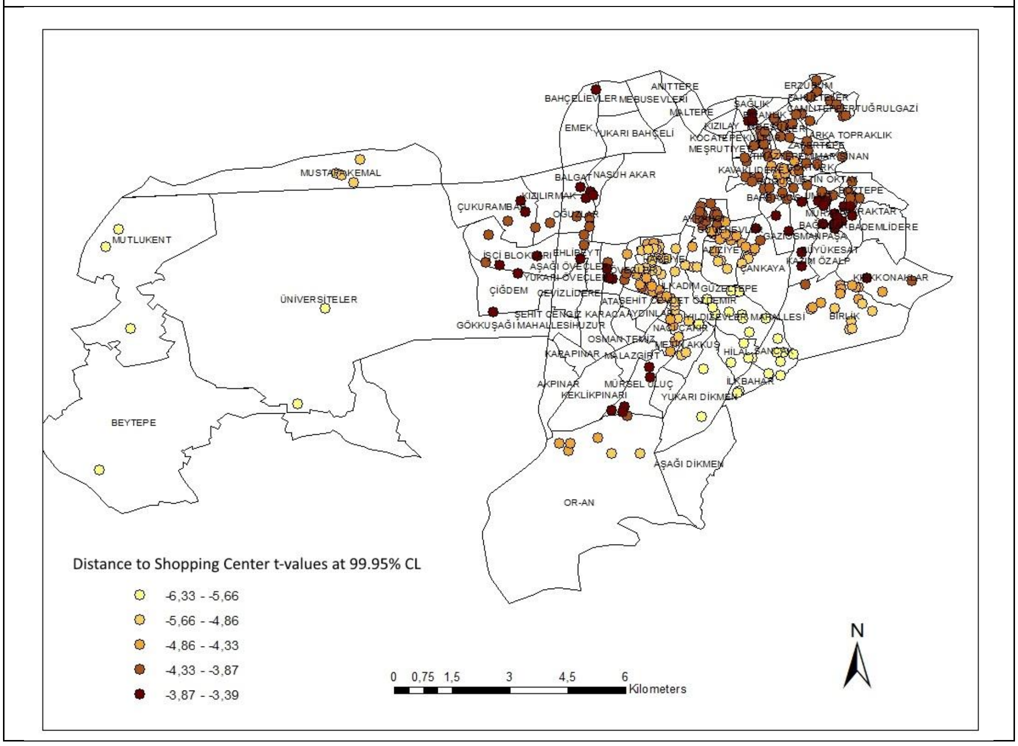


Figure 72 (continued): GWR t-Values for Distance to Shopping Center Attribute for Çankaya

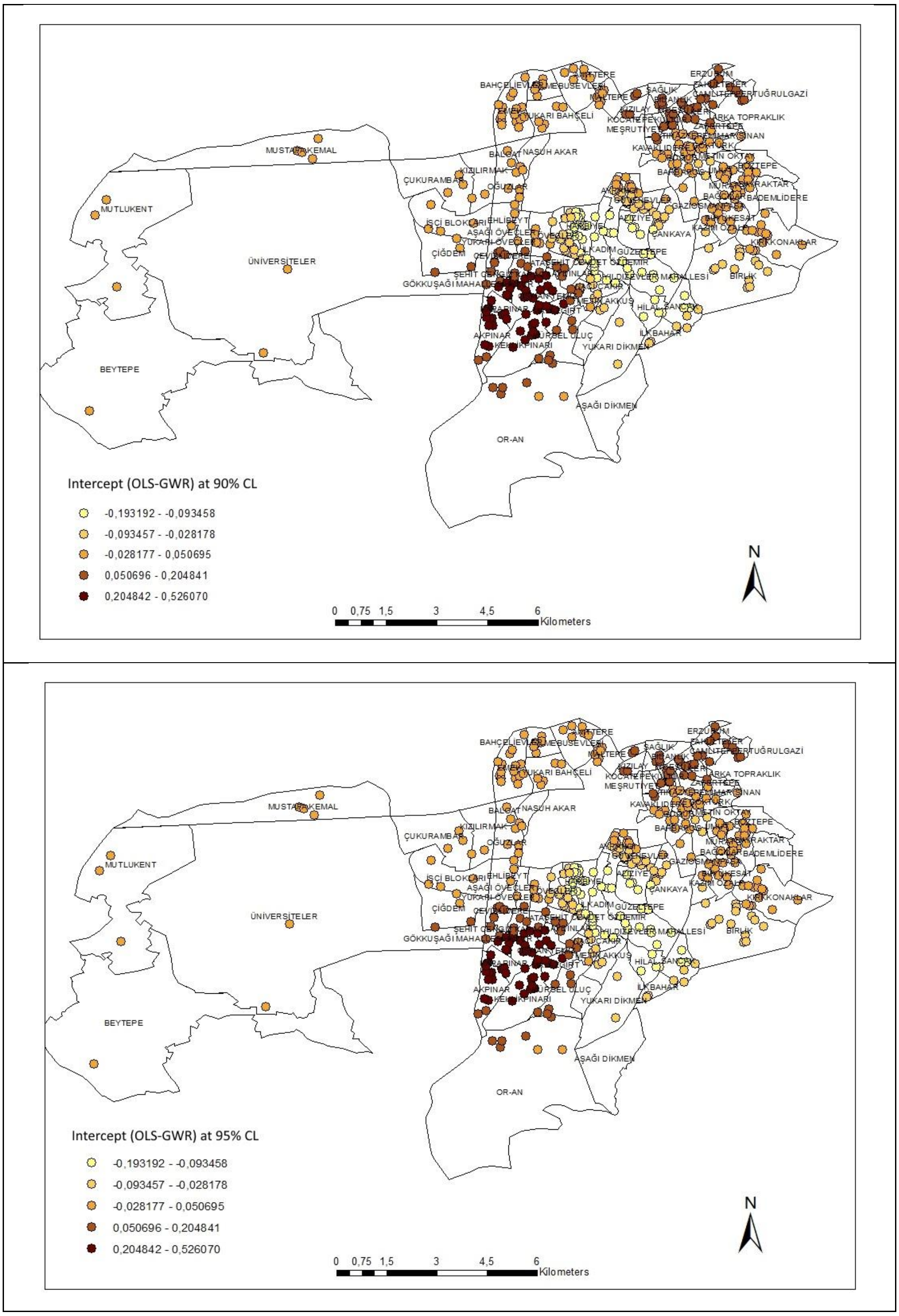


Figure 73: OLS-GWR Comparison Maps for Intercep Coefficients for Çankaya

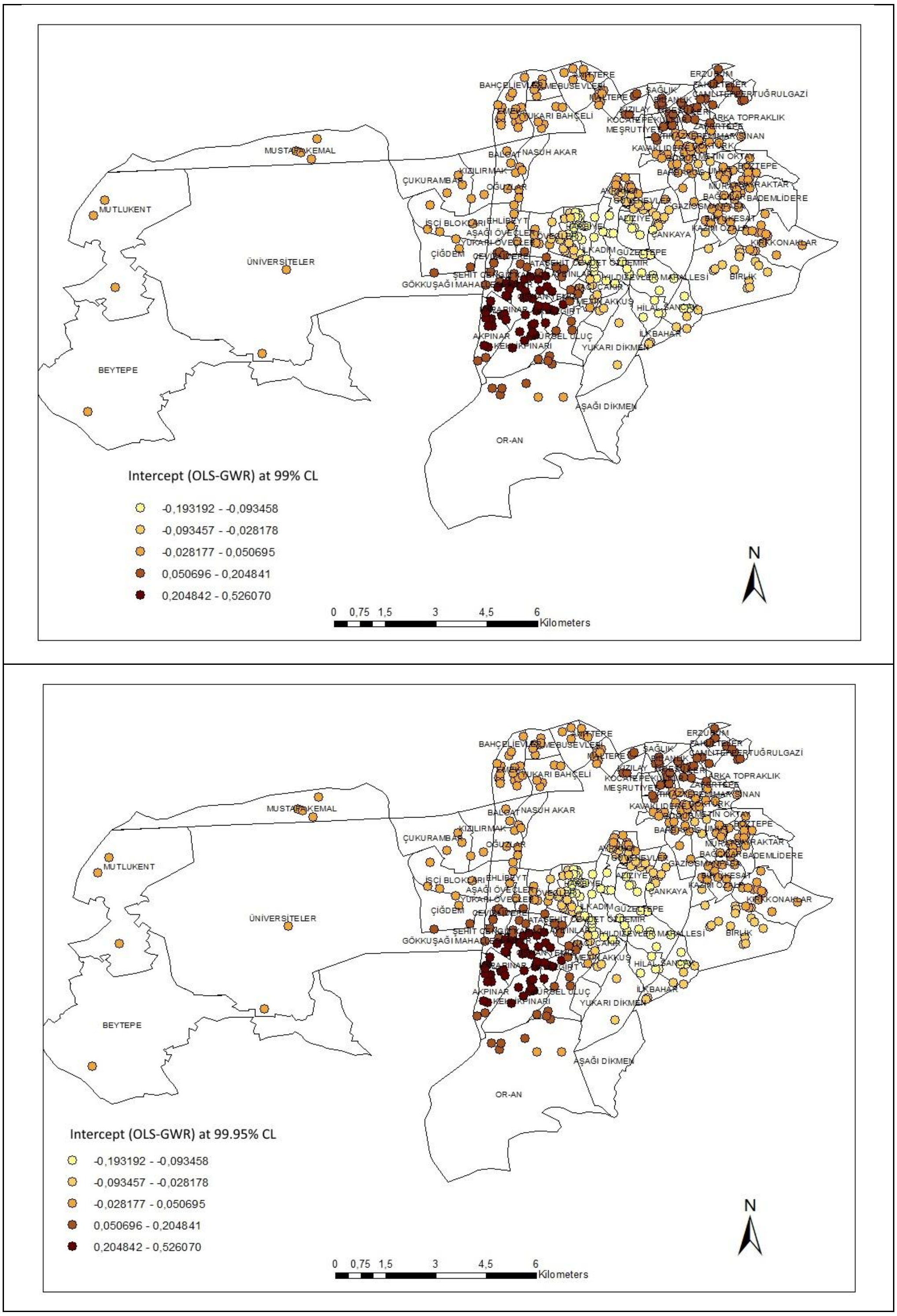


Figure 73 (continued): OLS-GWR Comparison Maps for Intercep Coefficients for Çankaya

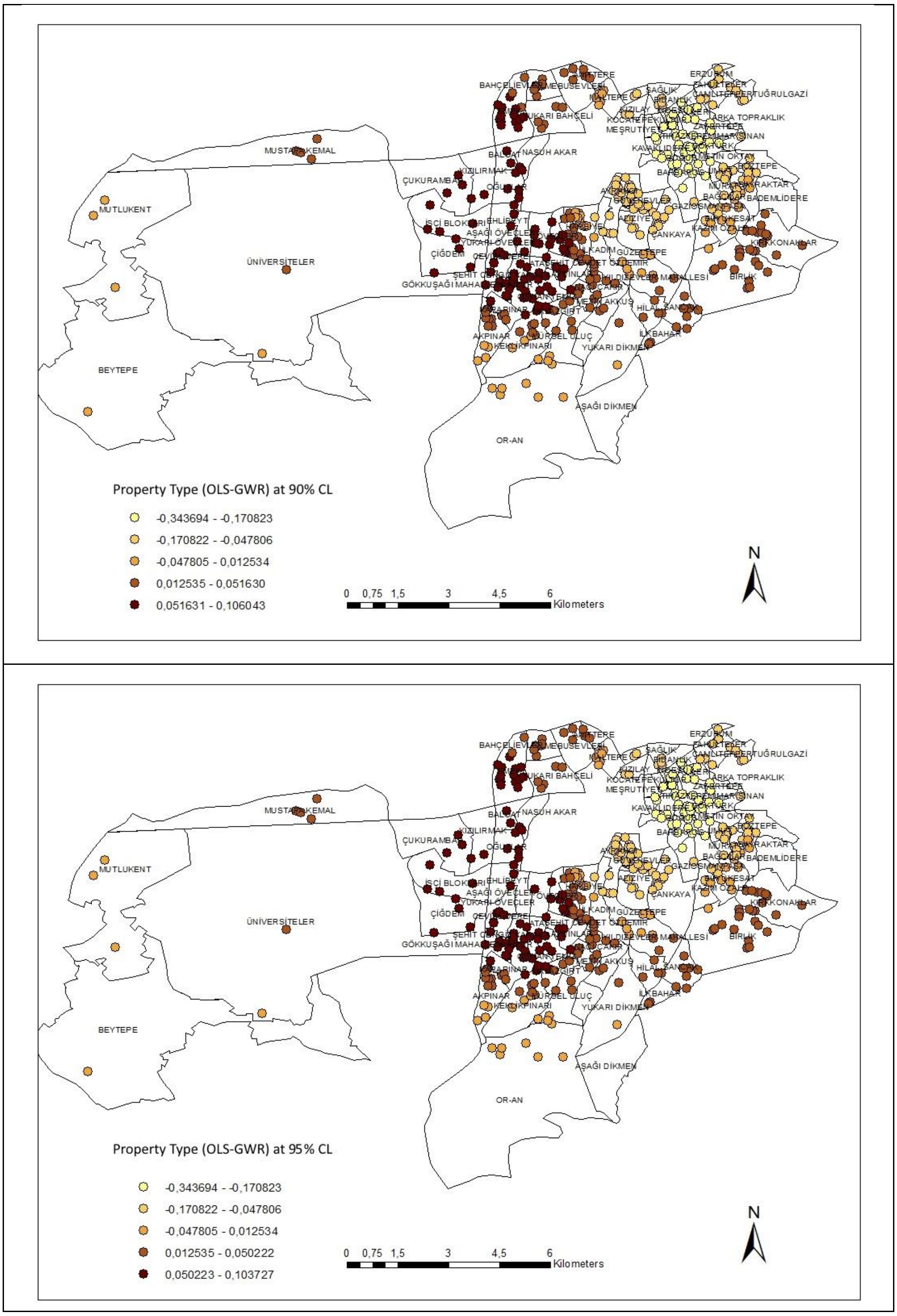


Figure 74: OLS-GWR Comparison Maps for Property Type Coefficients for Çankaya

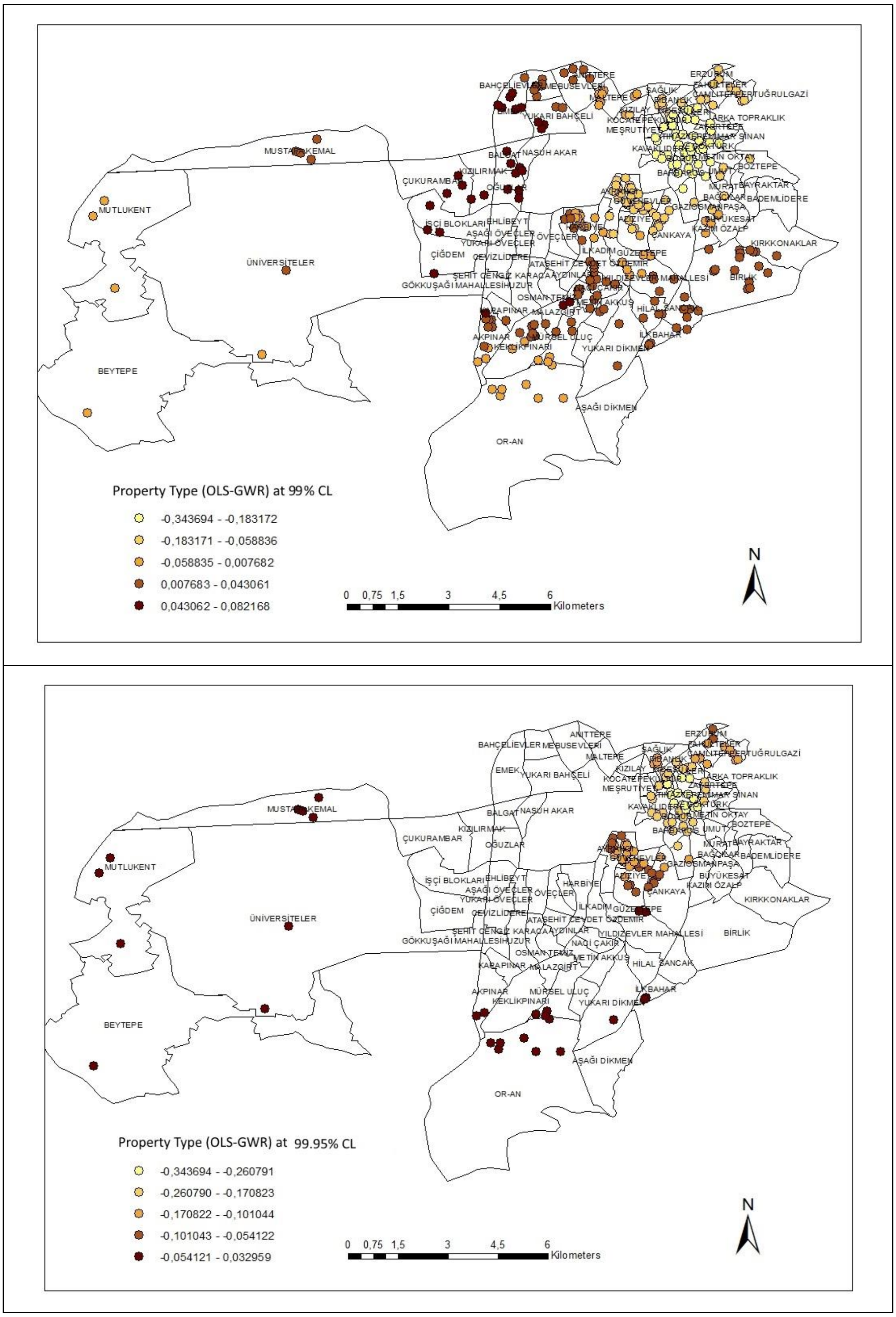


Figure 74 (continued): OLS-GWR Comparison Maps for Property Type Coefficients for Çankaya

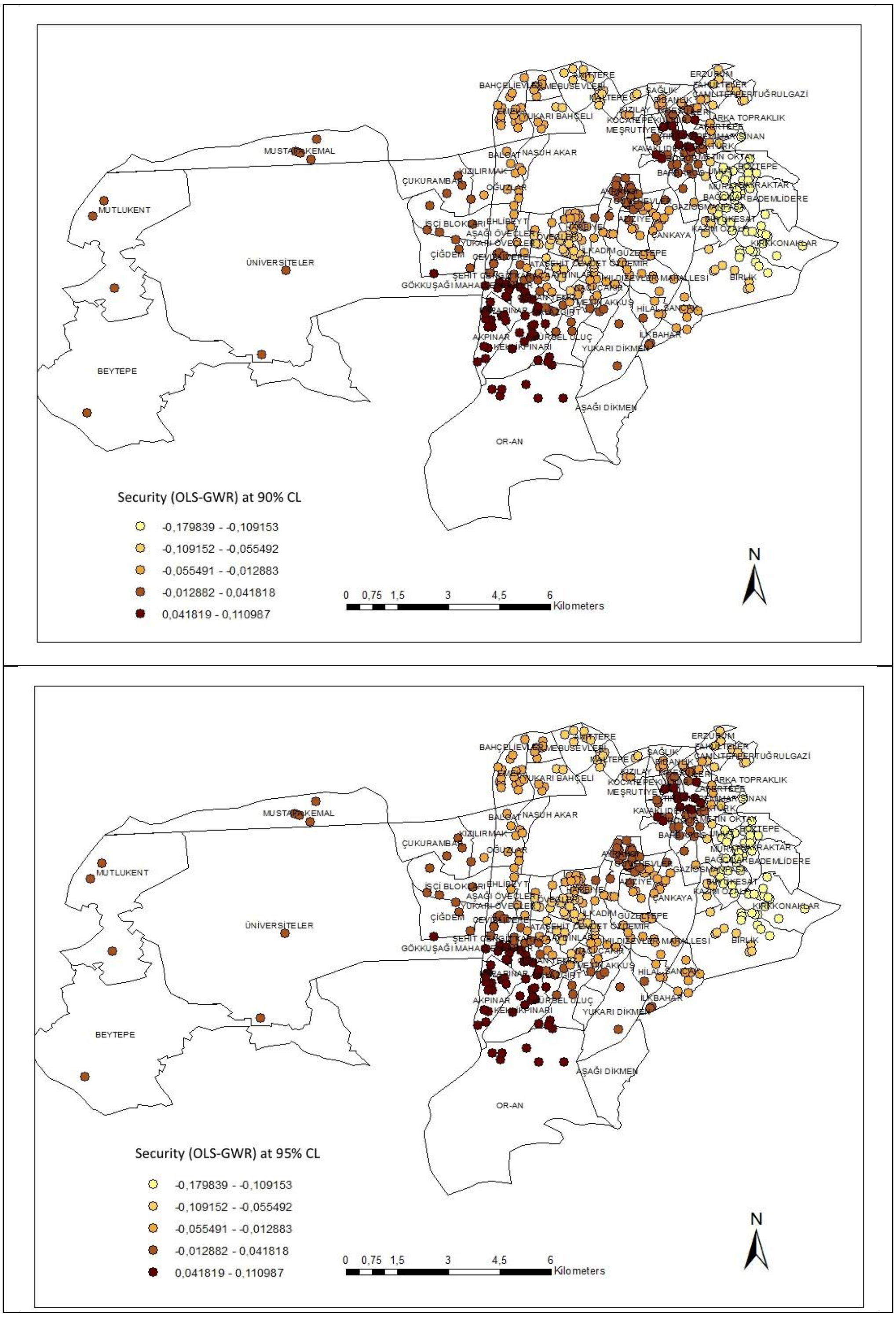


Figure 75: OLS-GWR Comparison Maps for Security Coefficients for Çankaya

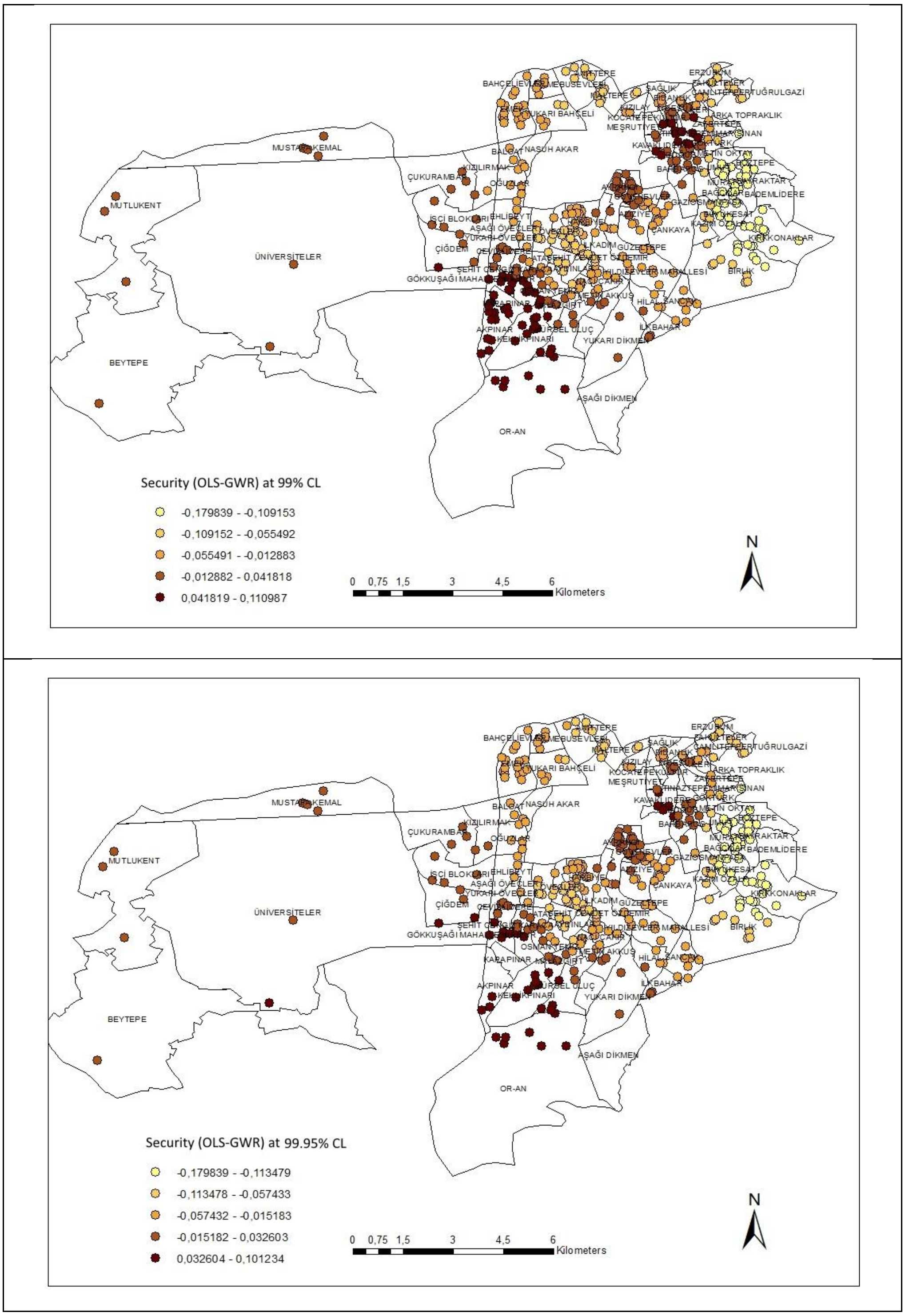


Figure 75 (continued): OLS-GWR Comparison Maps for Security Coefficients for Çankaya

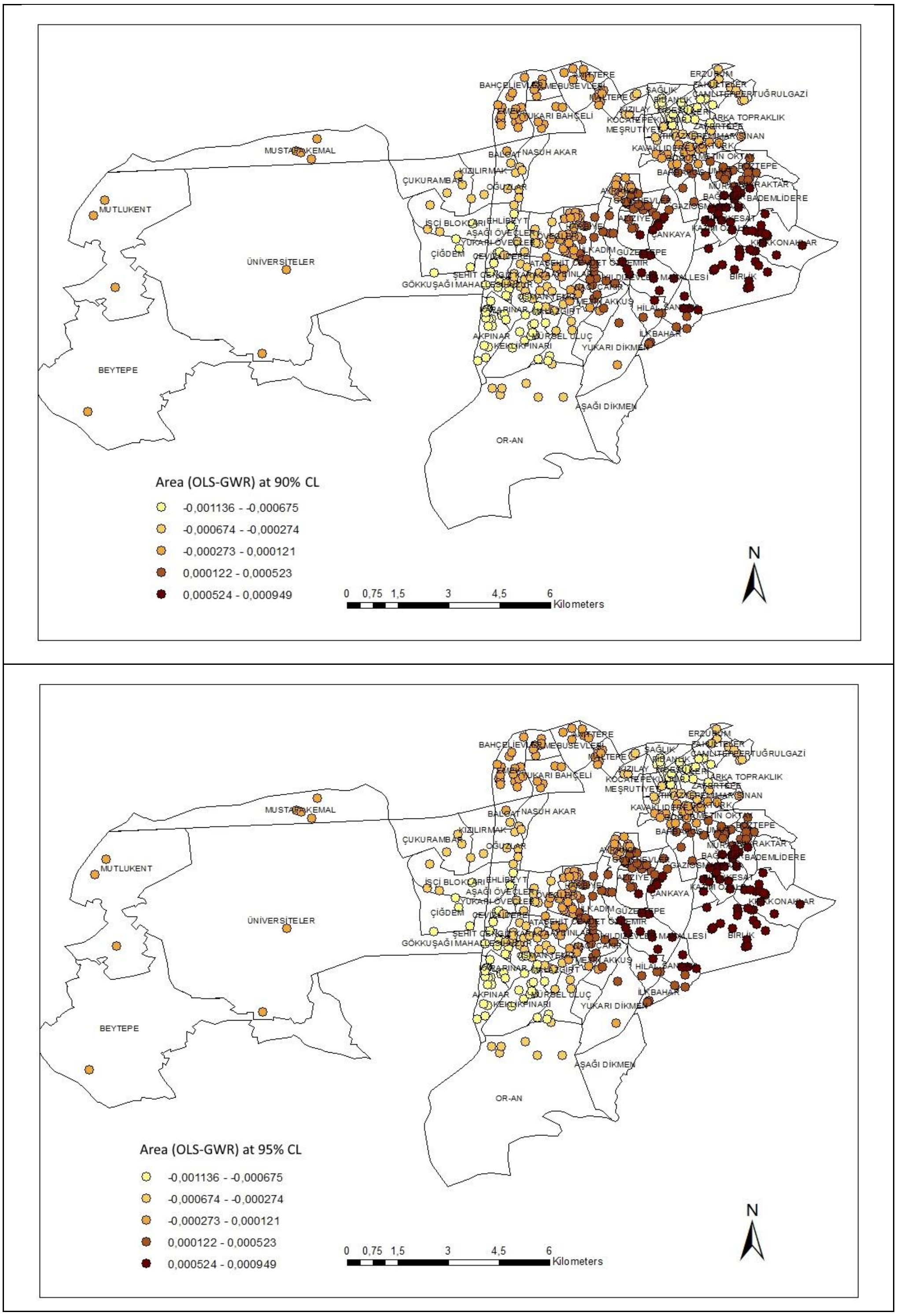


Figure 76: OLS-GWR Comparison Maps for Area Coefficients for Çankaya

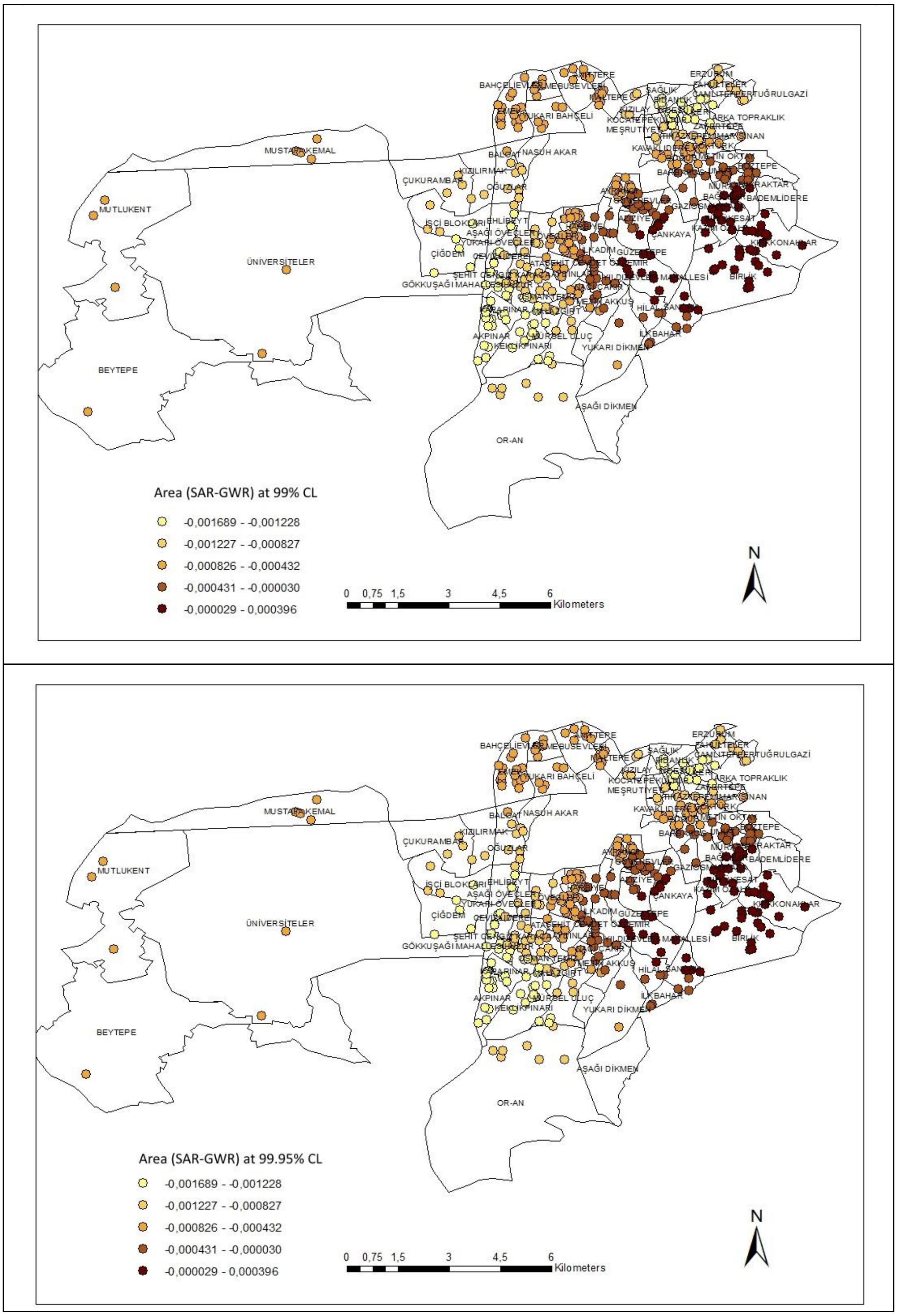


Figure 76 (continued): OLS-GWR Comparison Maps for Area Coefficients for Çankaya

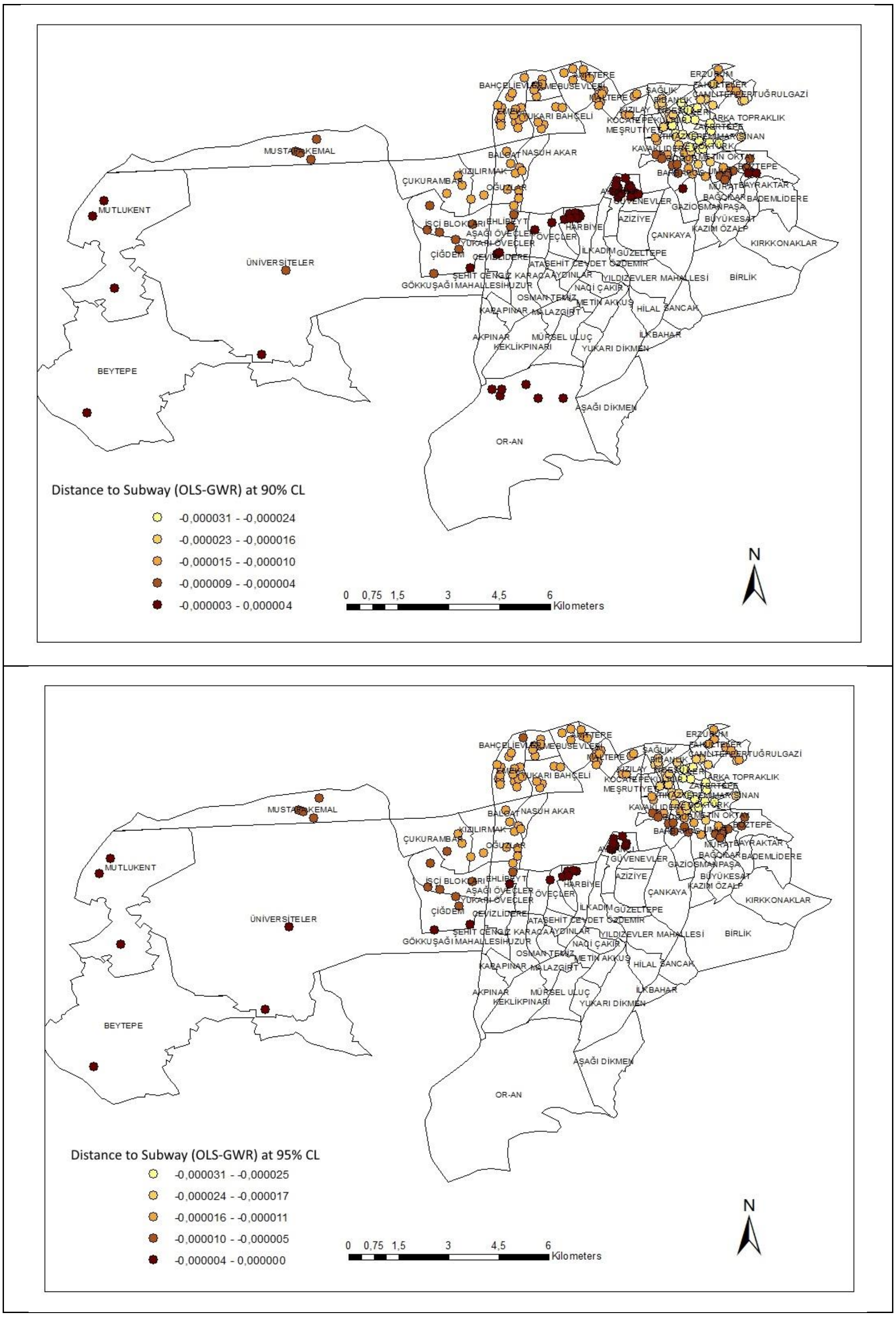


Figure 77: OLS-GWR Comparison Maps for Distance to Subway Coefficients for Çankaya

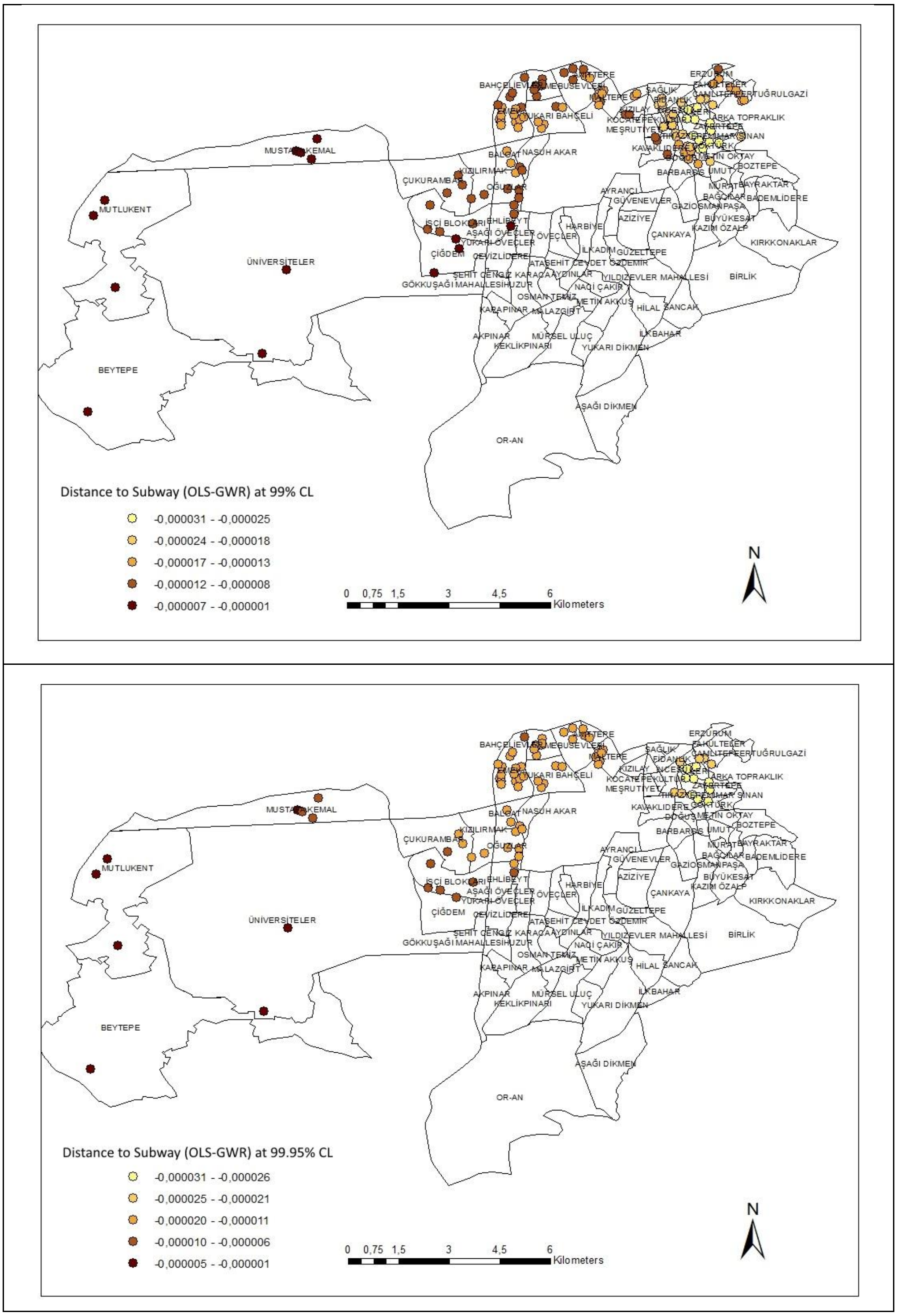


Figure 77 (continued): OLS-GWR Comparison Maps for Distance to Subway Coefficients for Çankaya

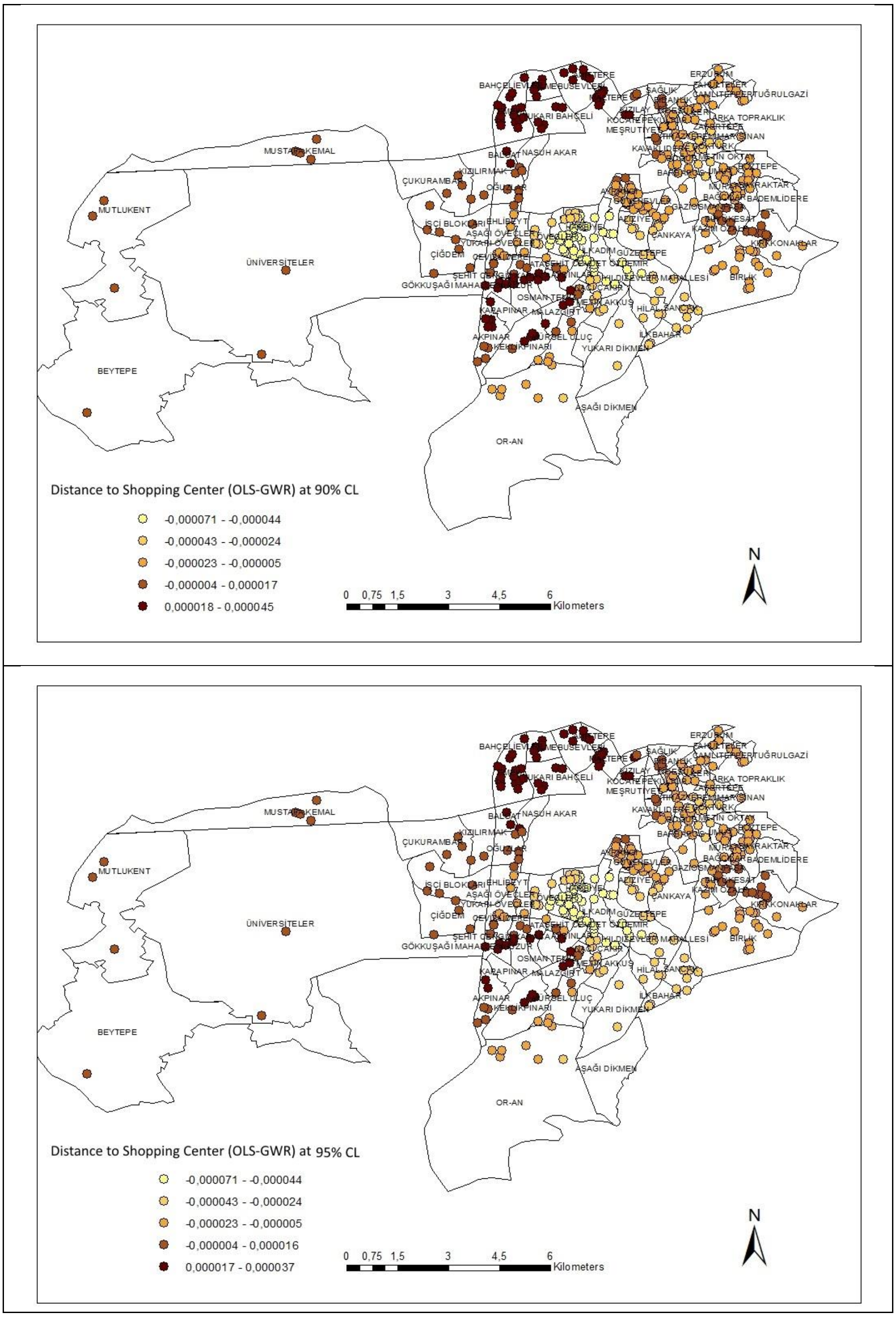


Figure 78: OLS-GWR Comparison Maps for Distance to Shopping Center Coefficients for Çankaya

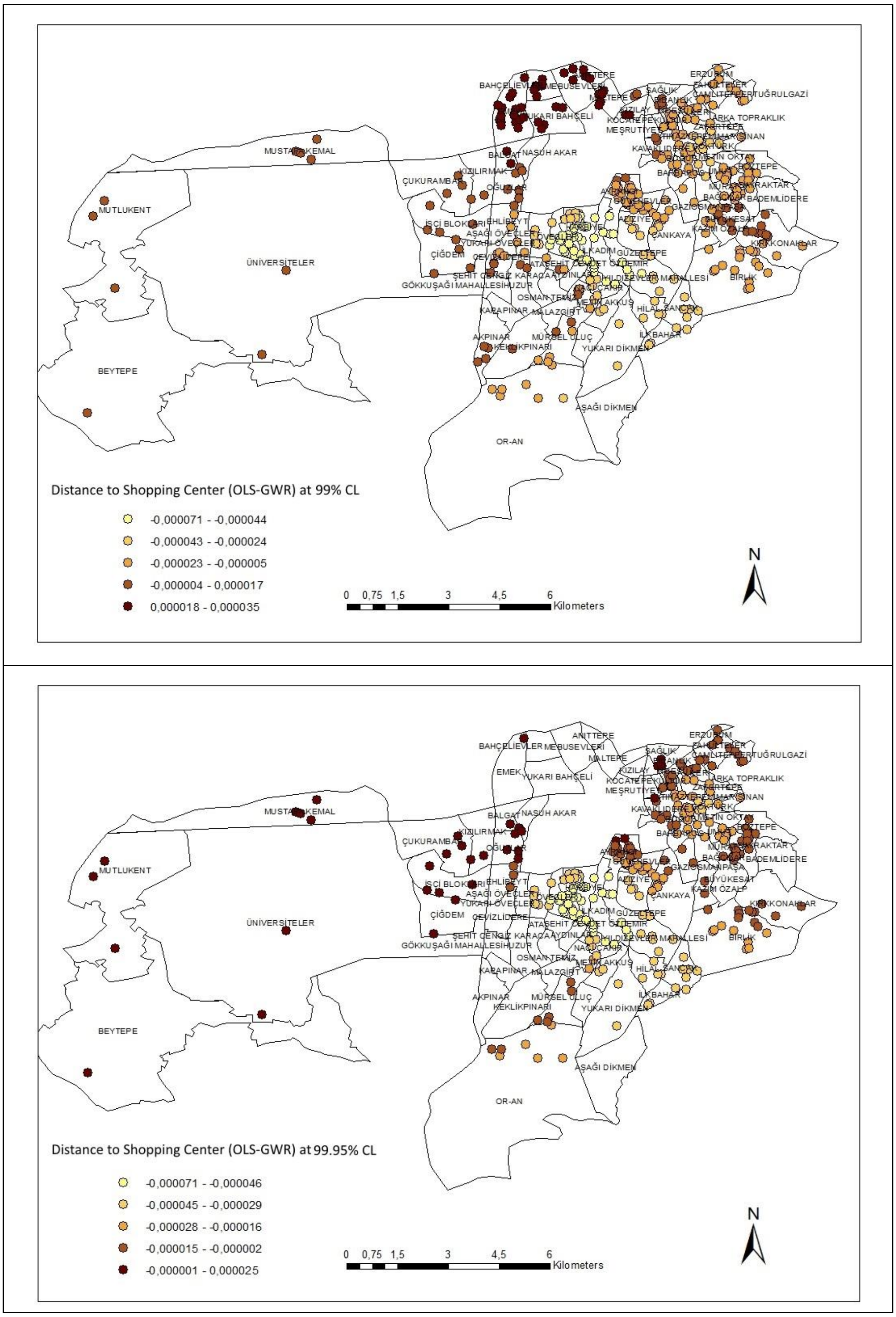


Figure 78 (continued): OLS-GWR Comparison Maps for Distance to Shopping Center Coefficients for Çankaya

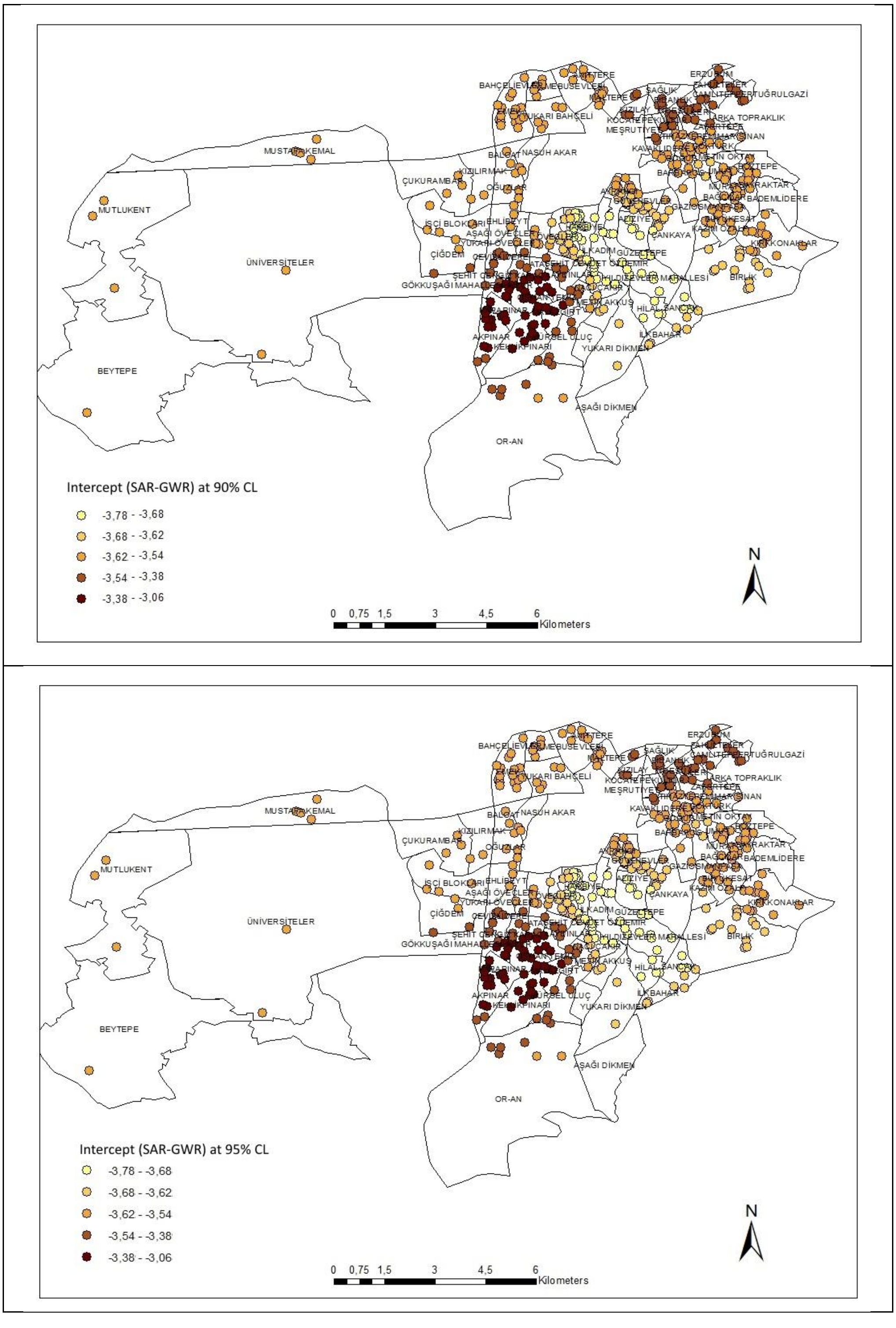


Figure 79: SR-GWR Comparison Maps for Intercep Coefficients for Çankaya

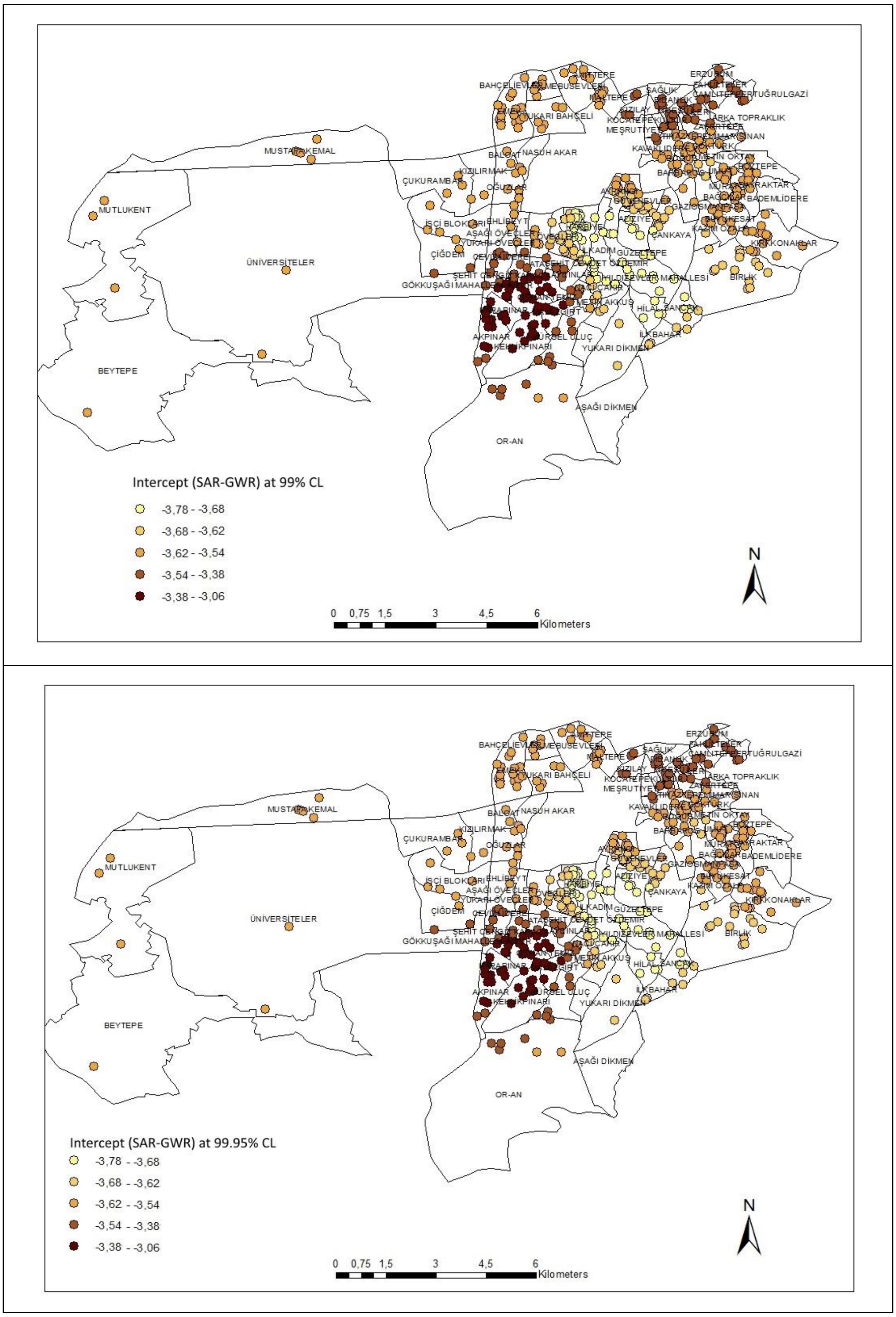


Figure 79 (continued): SR-GWR Comparison Maps for Intercep Coefficients for Çankaya

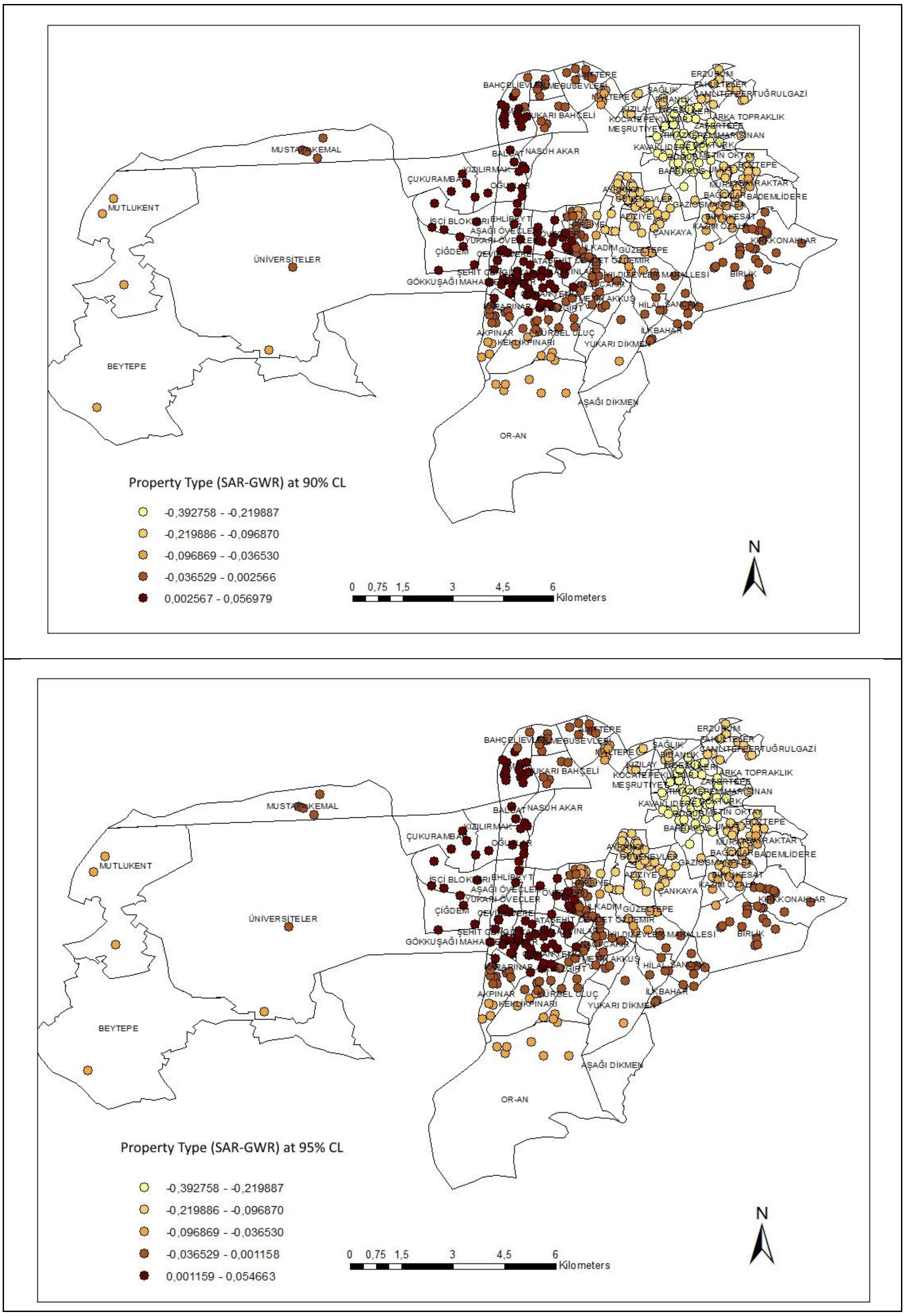


Figure 80: SR-GWR Comparison Maps for Property Type Coefficients for Çankaya

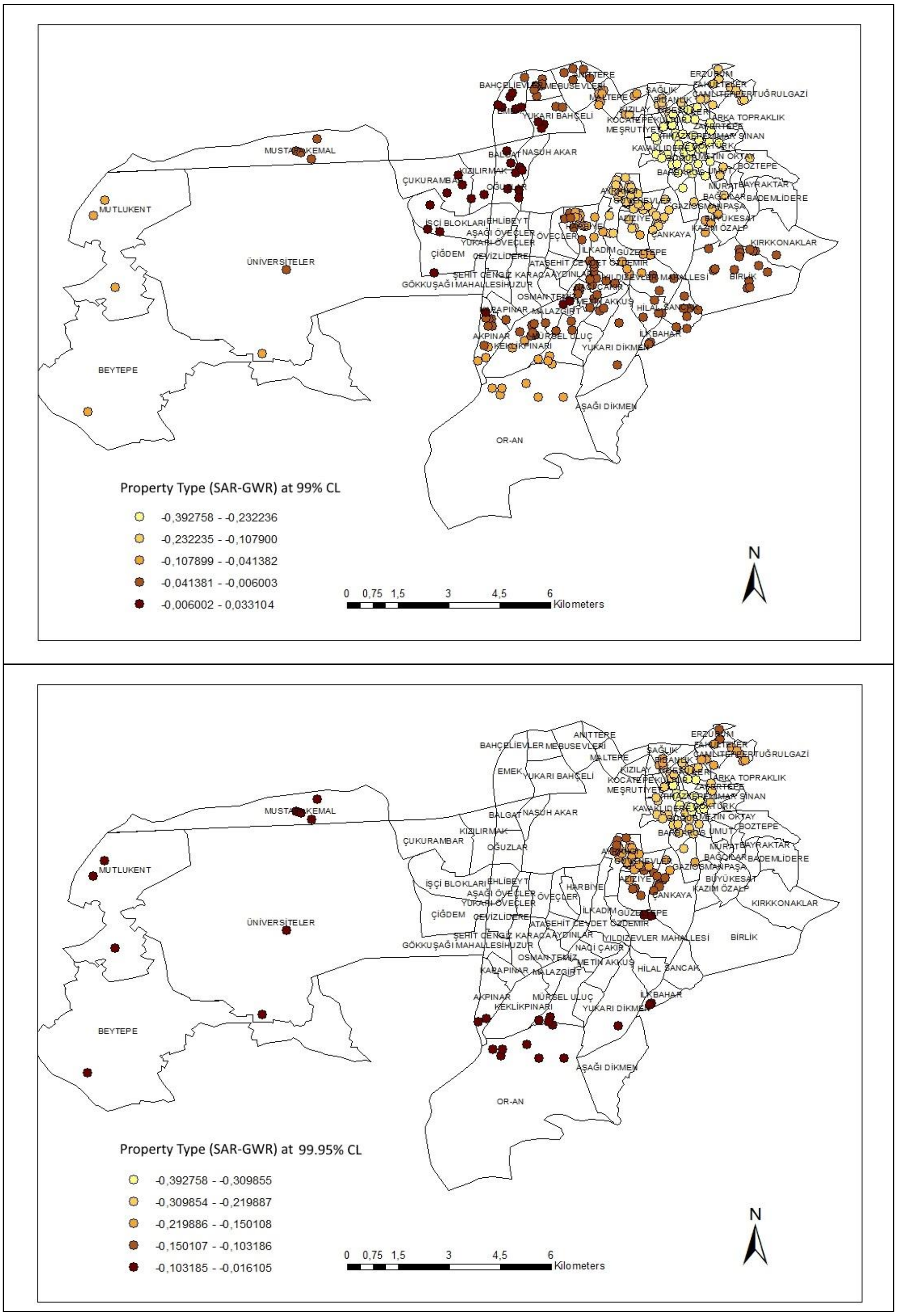


Figure 80 (continued): SR-GWR Comparison Maps for Property Type Coefficients for Çankaya

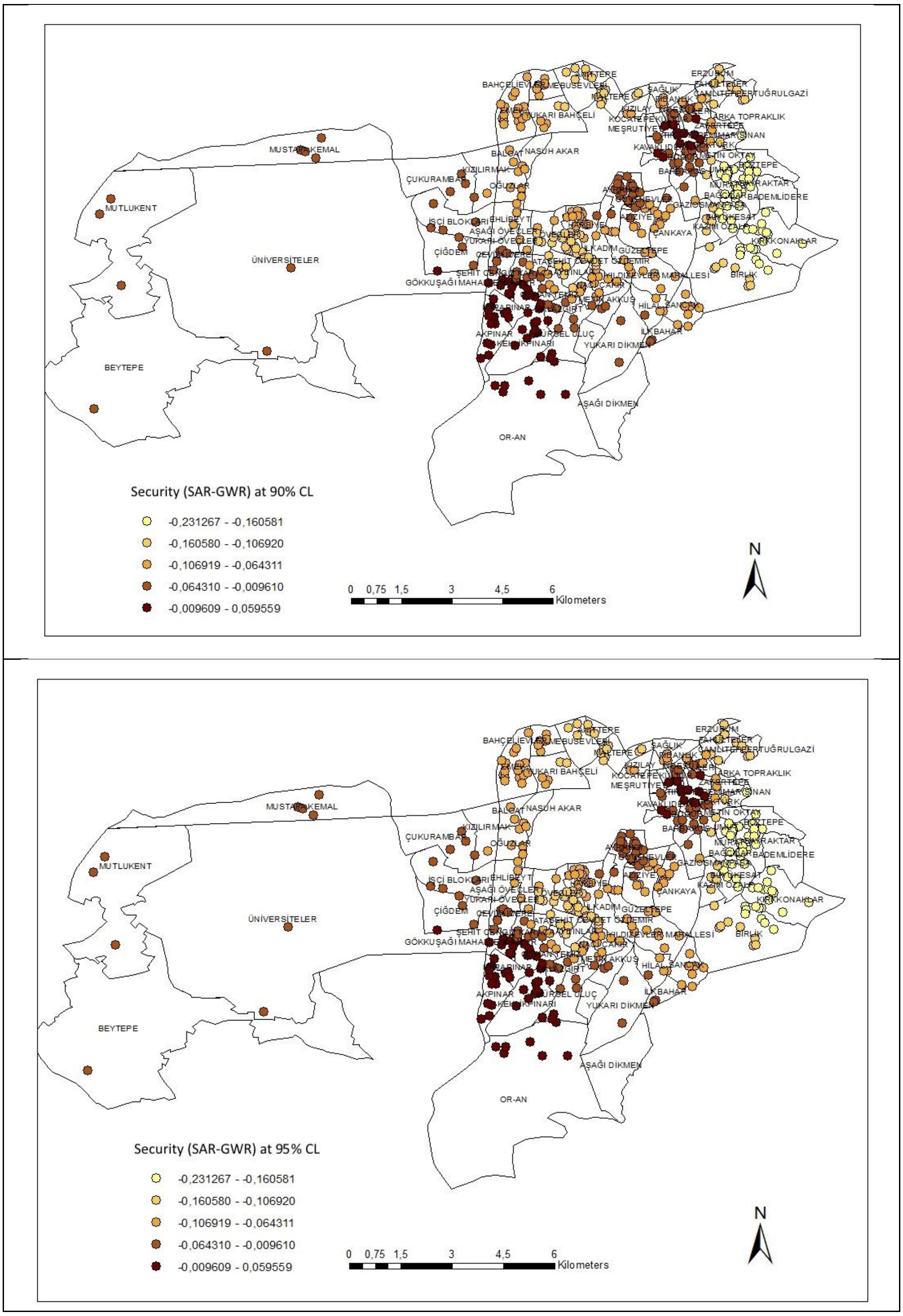


Figure 81: SR-GWR Comparison Maps for Security Coefficients for Çankaya

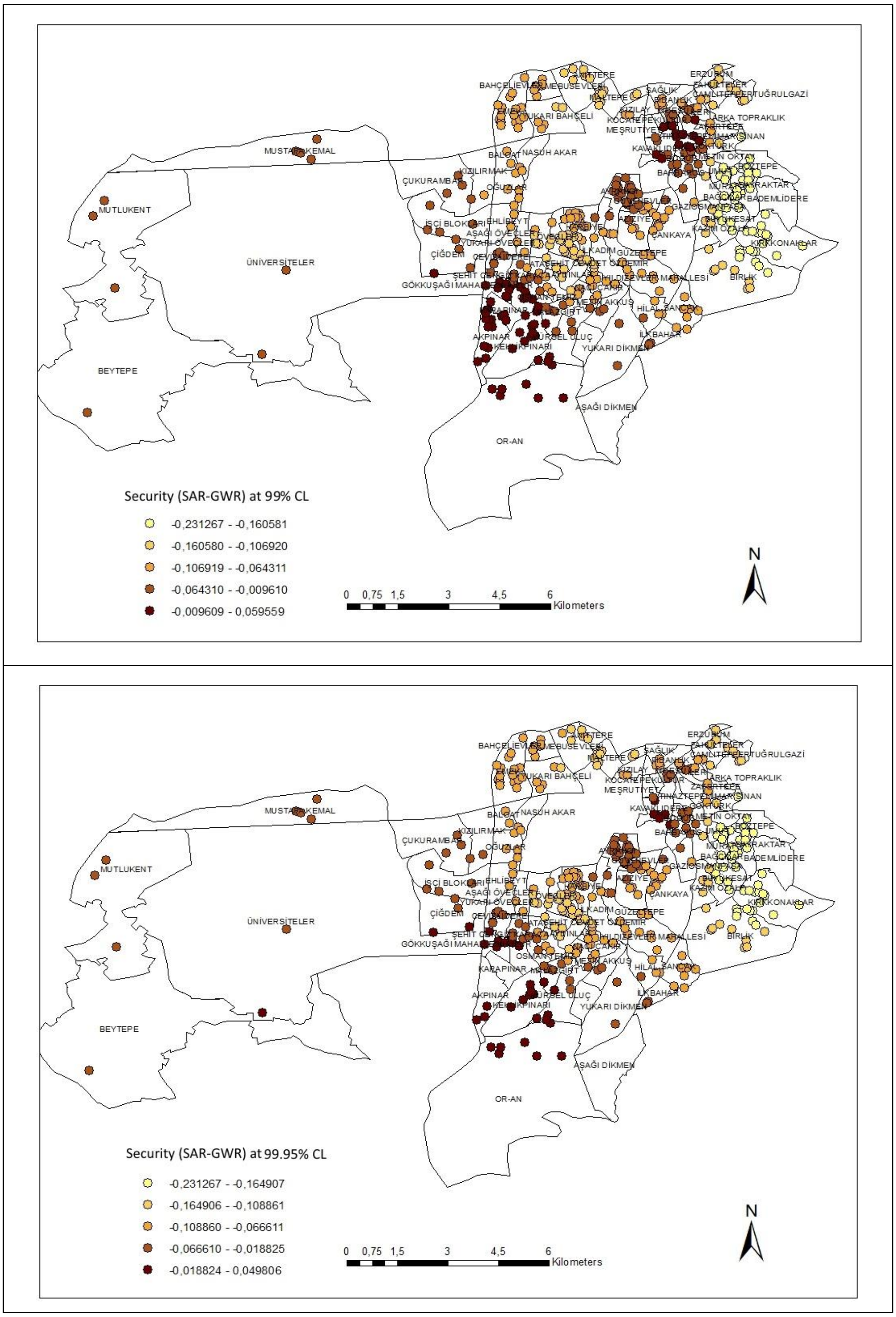
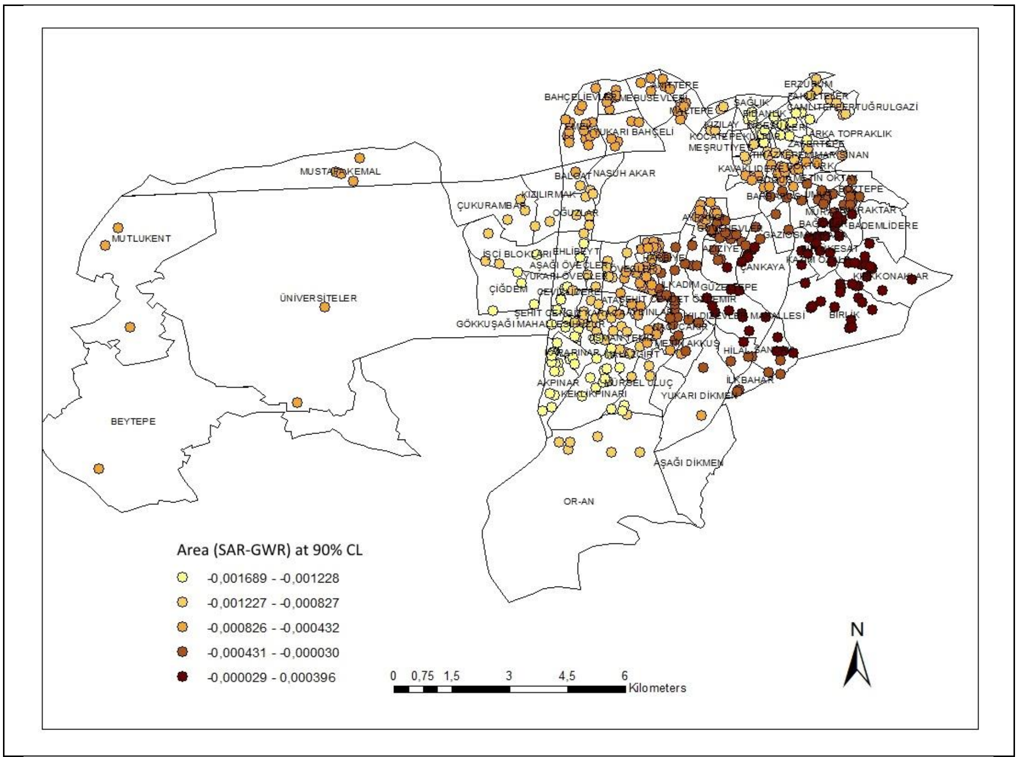


Figure 81 (continued): SR-GWR Comparison Maps for Security Coefficients for Çankaya



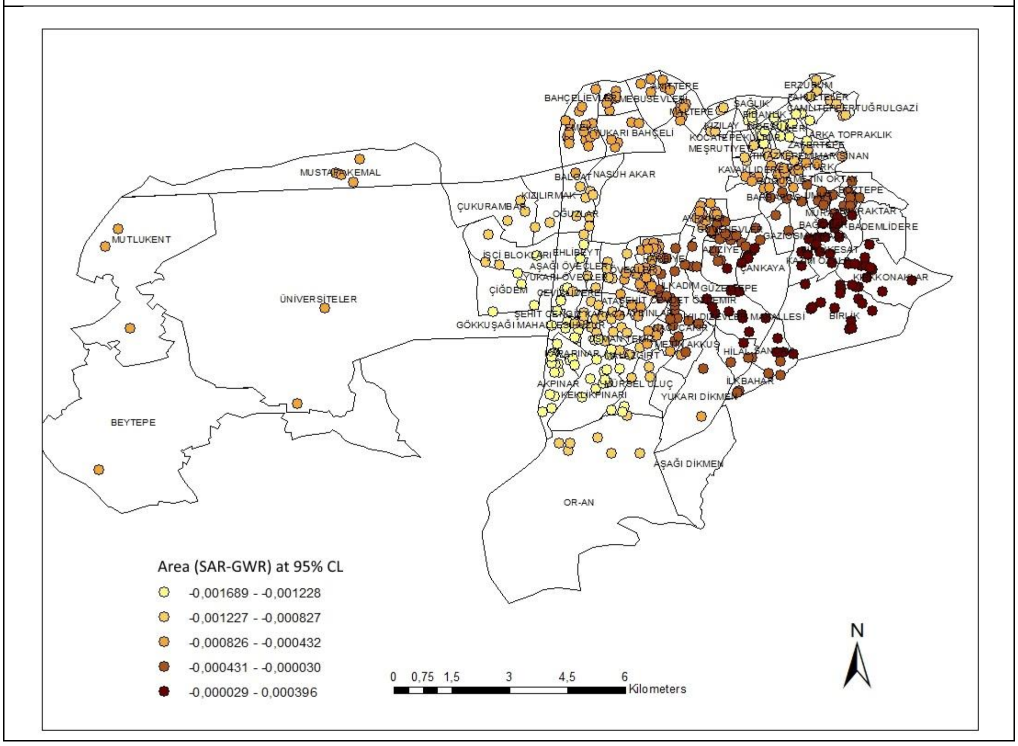


Figure 82: SR-GWR Comparison Maps for Area Coefficients for Çankaya

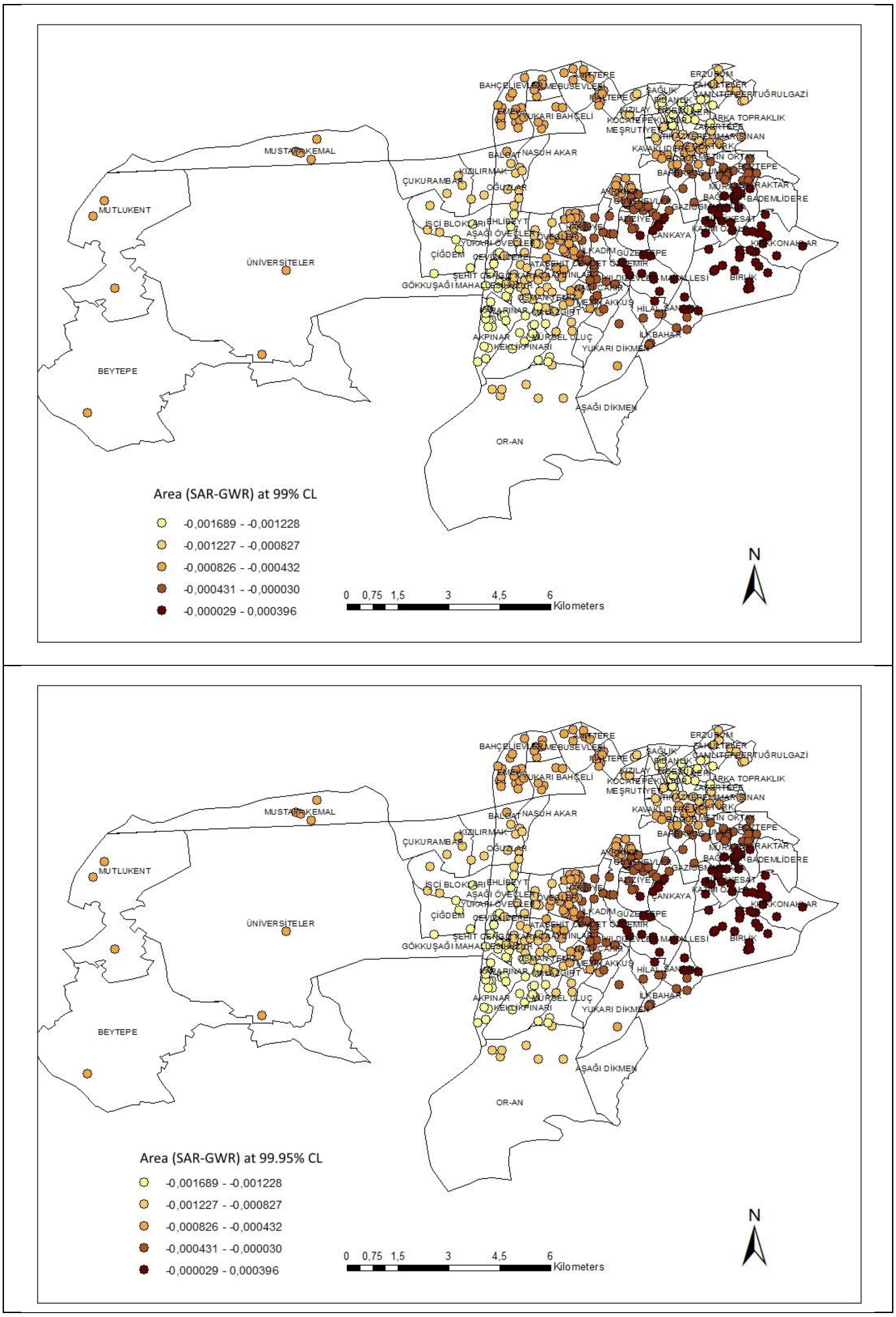


Figure 82 (continued): SR-GWR Comparison Maps for Area Coefficients for Çankaya

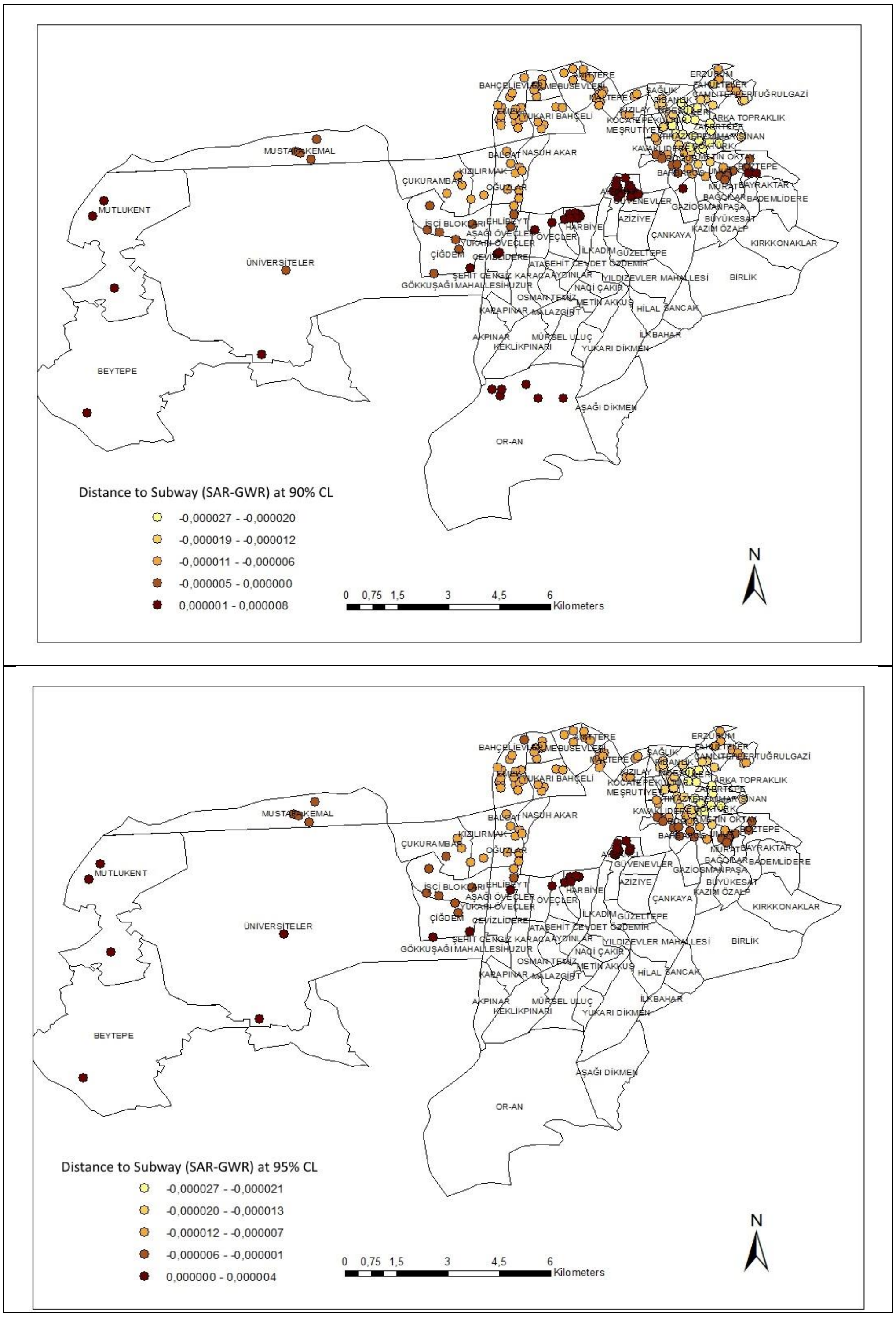


Figure 83: SR-GWR Comparison Maps for Distance to Subway Coefficients for Çankaya

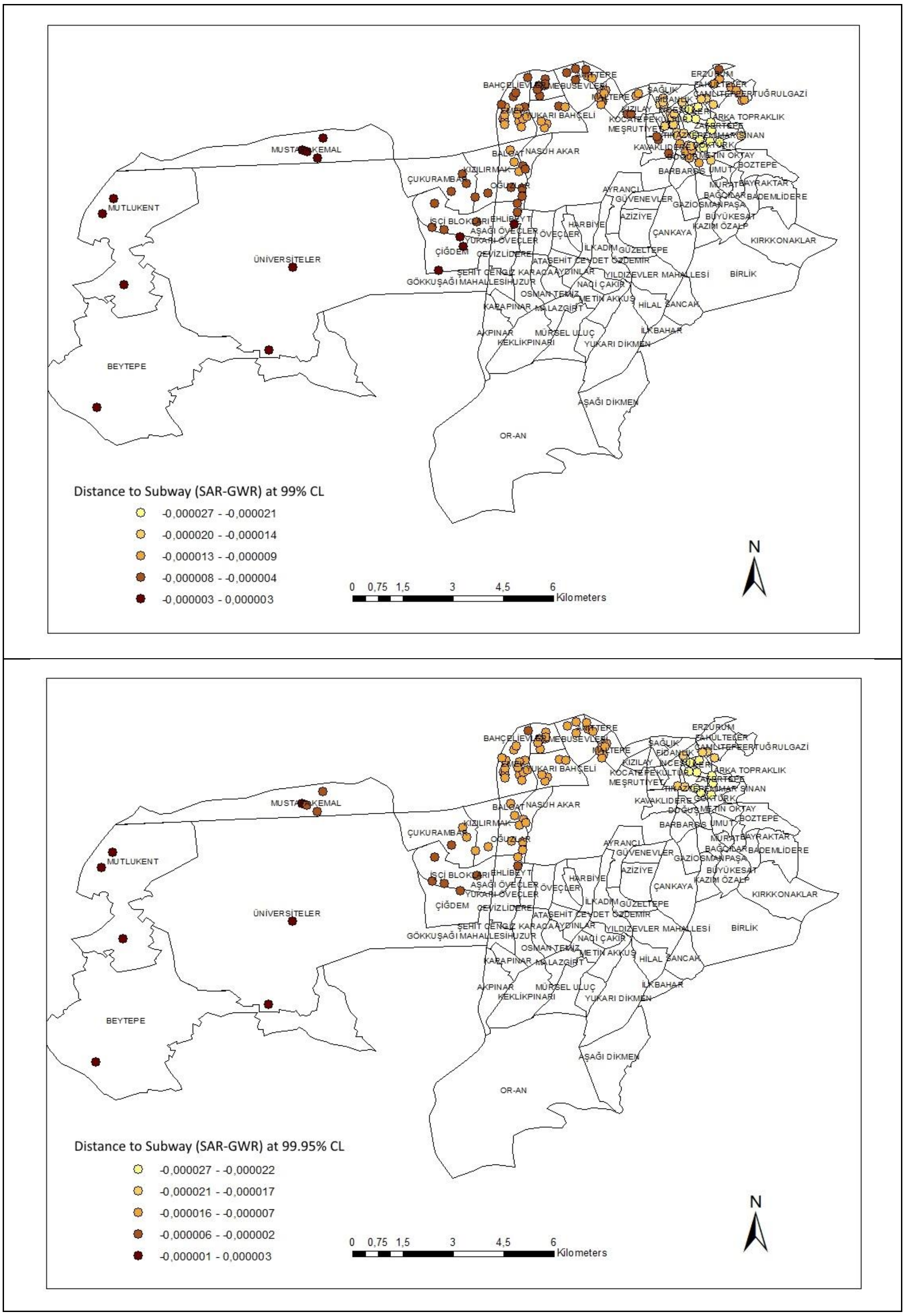


Figure 83 (continued): SR-GWR Comparison Maps for Distance to Subway Coefficients for Çankaya

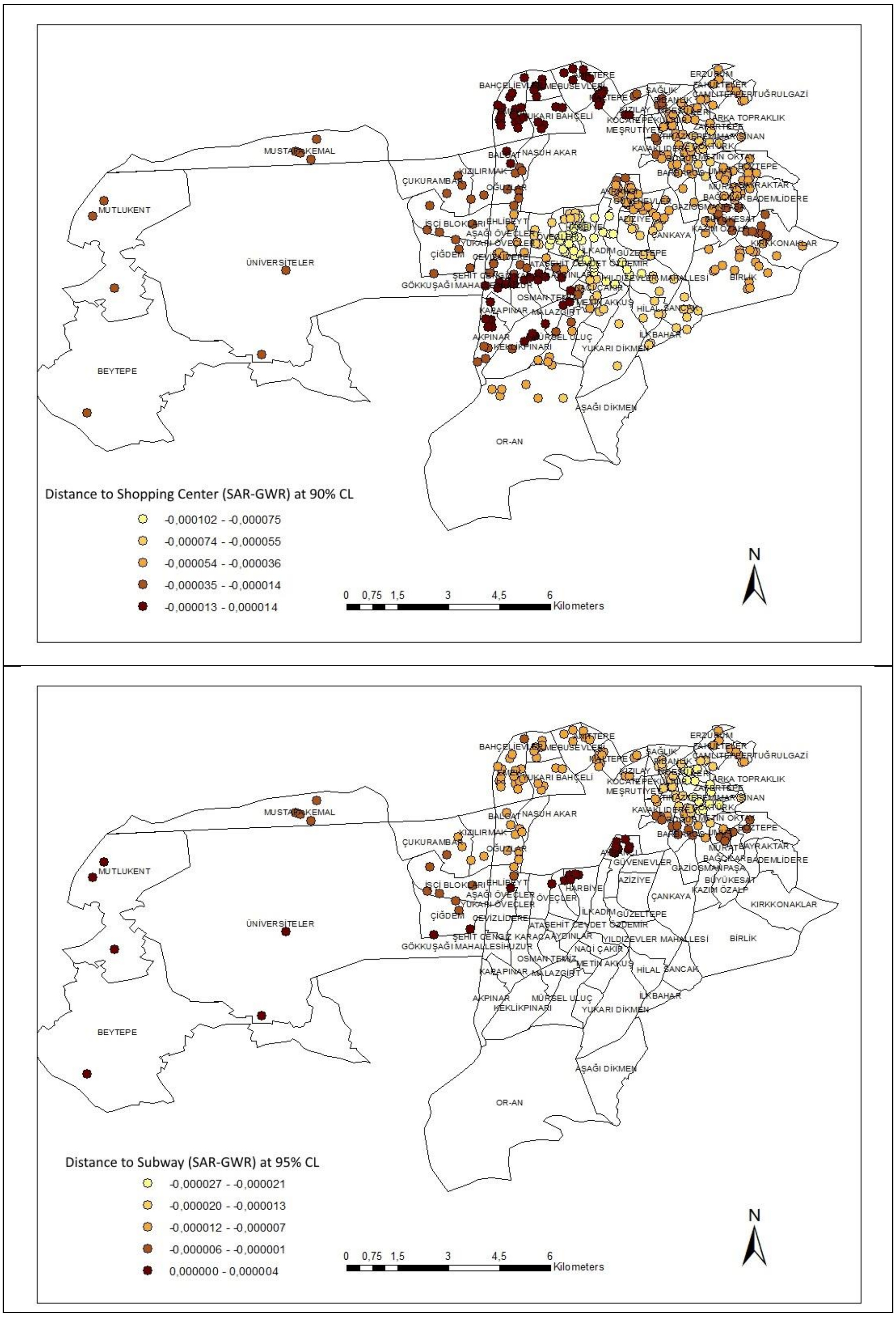


Figure 84: SR-GWR Comparison Maps for Distance to Shopping Centers Coefficients for Çankaya

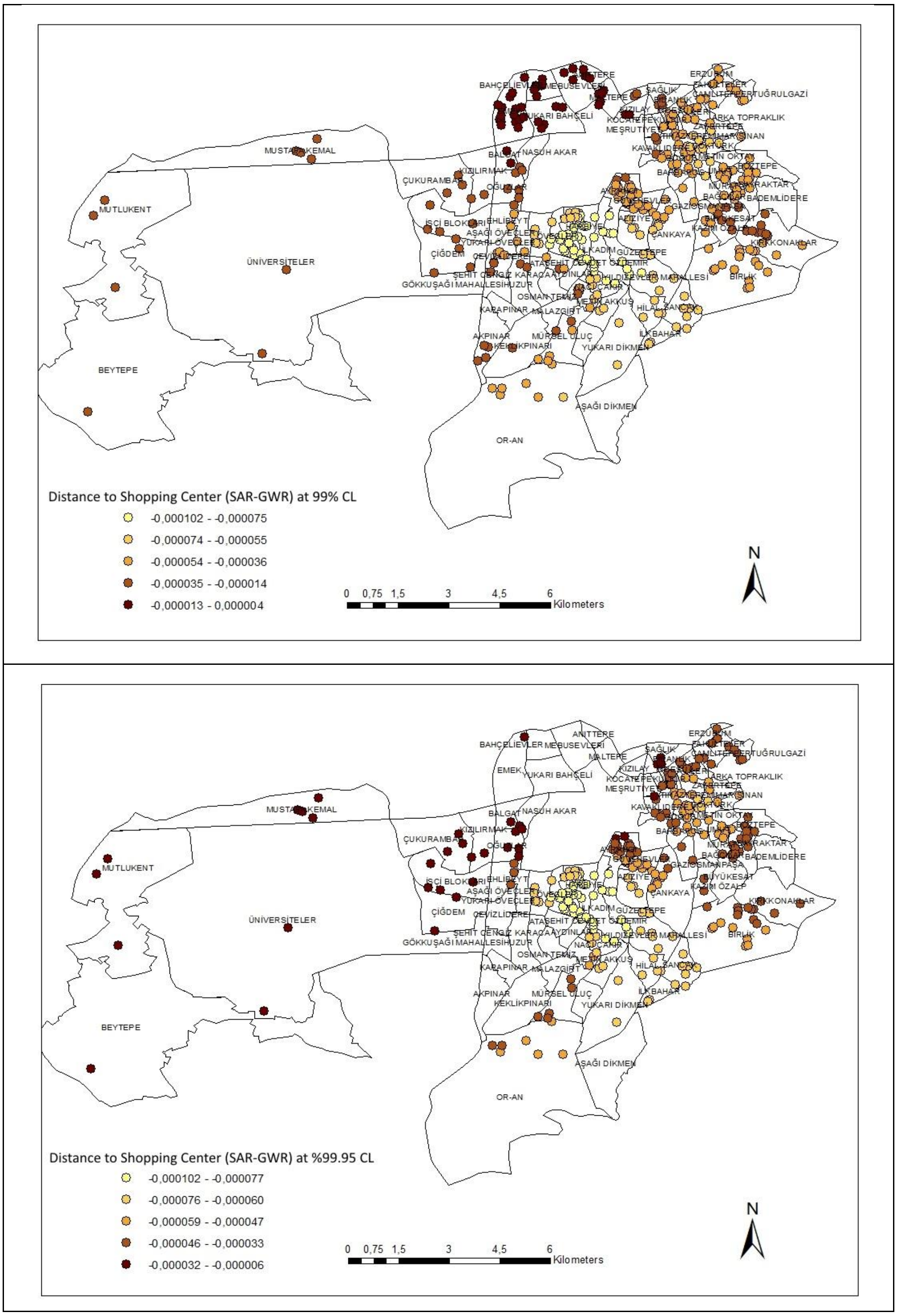


Figure 84 (continued): SR-GWR Comparison Maps for Distance to Shopping Centers Coefficients for Çankaya

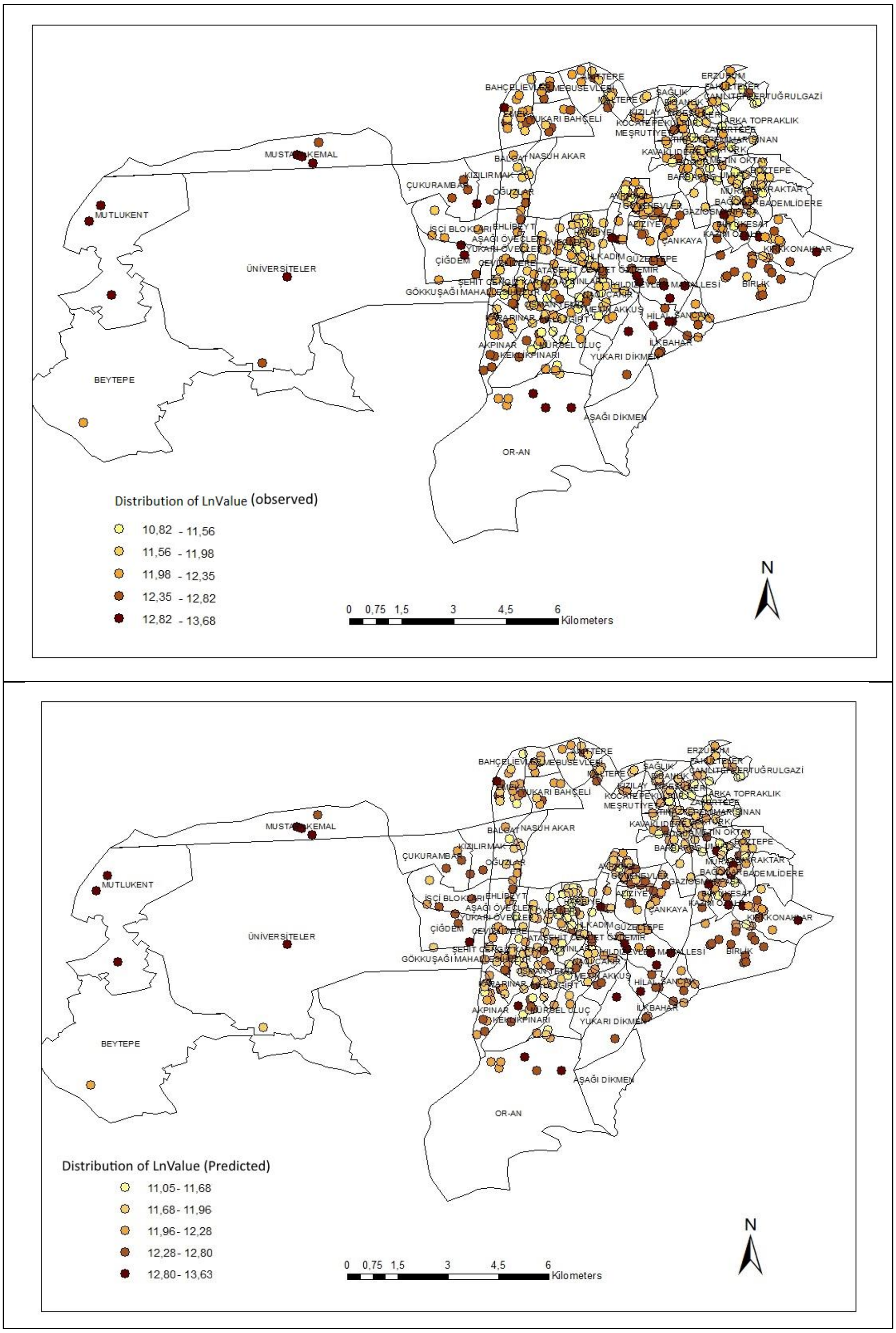


Figure 85: Distribution of Observed and Predicted LnValue Variables for Çankaya

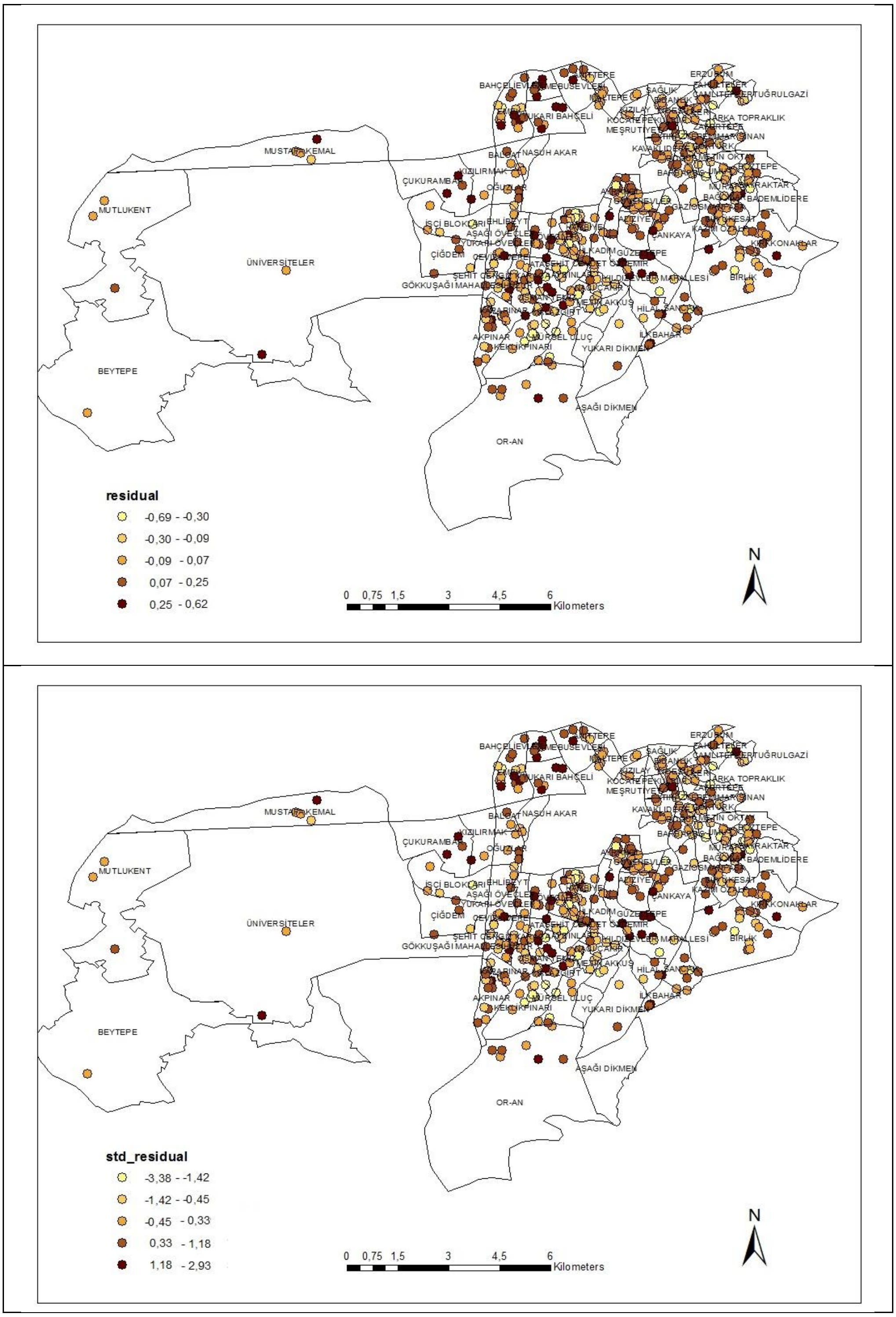


Figure 86: GWR Map for Distribution of Residual and Standard Residuals for Çankaya

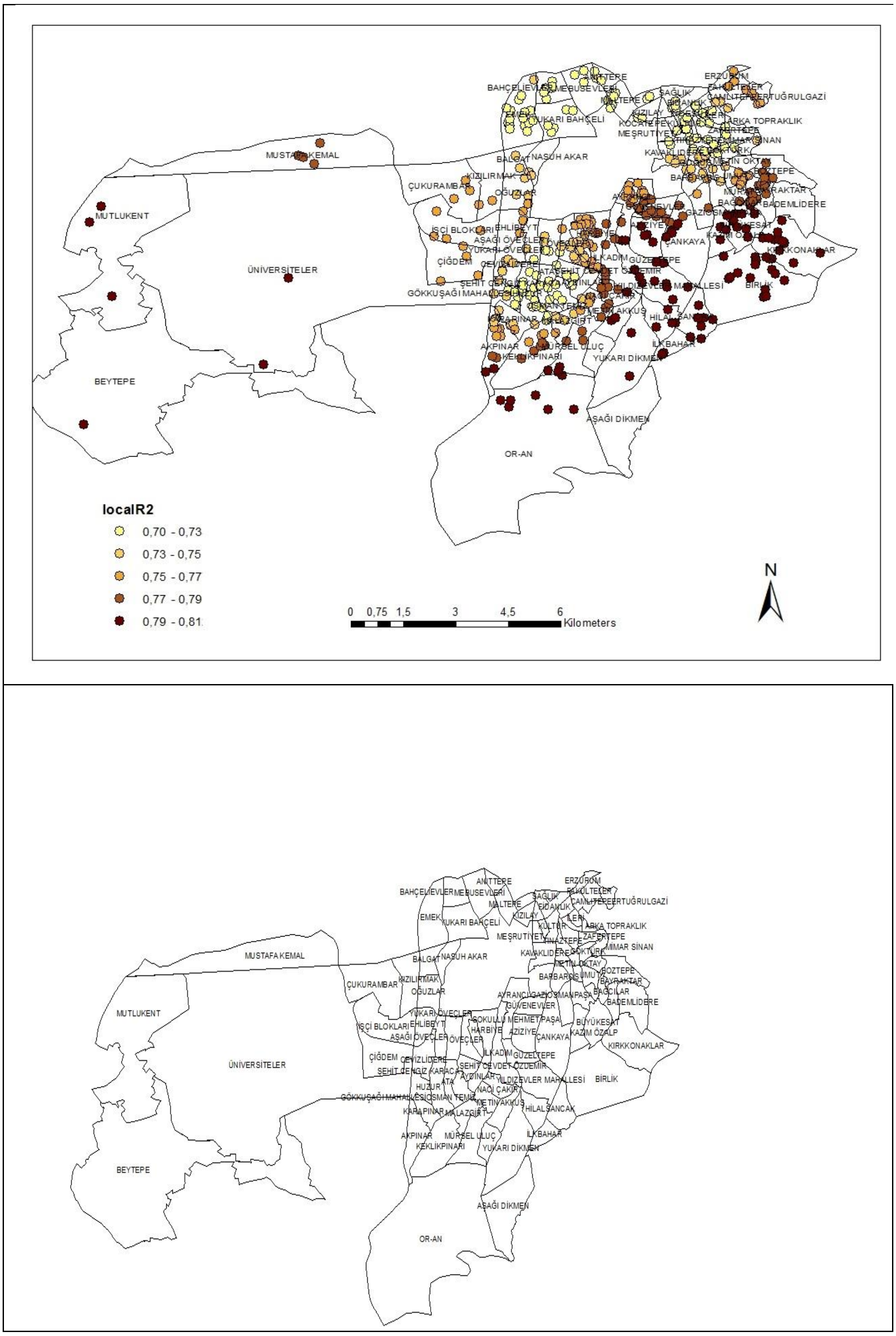


Figure 87: GWR Map for Local R² and Neighborhoods for Çankaya

Figure 88: GWR Results for Intercept Coefficients for Keçiören

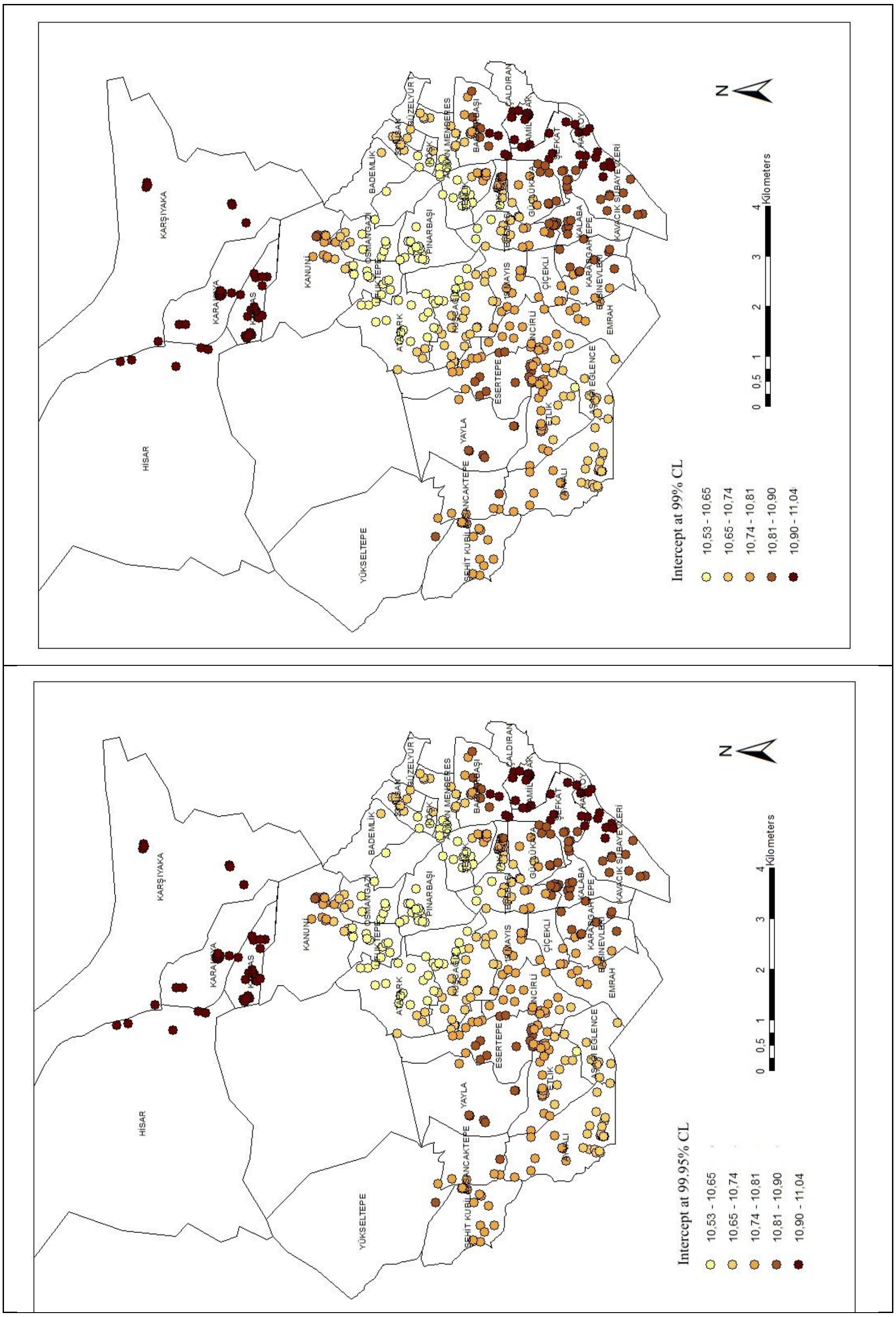


Figure 88 (continued): GWR Results for Intercept Coefficients for Keçiören

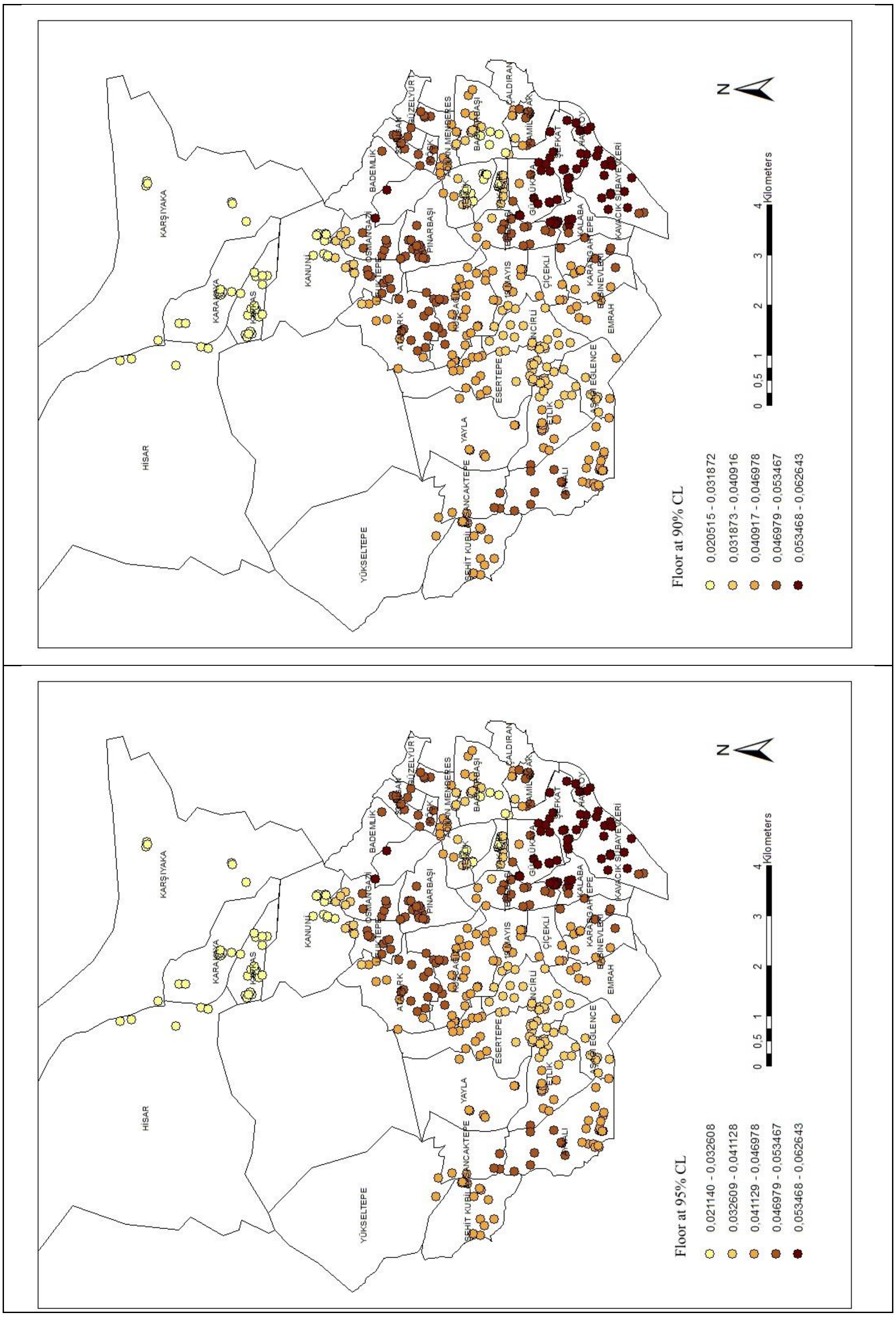


Figure 89: GWR Results for Floor Coefficients for Keçiören

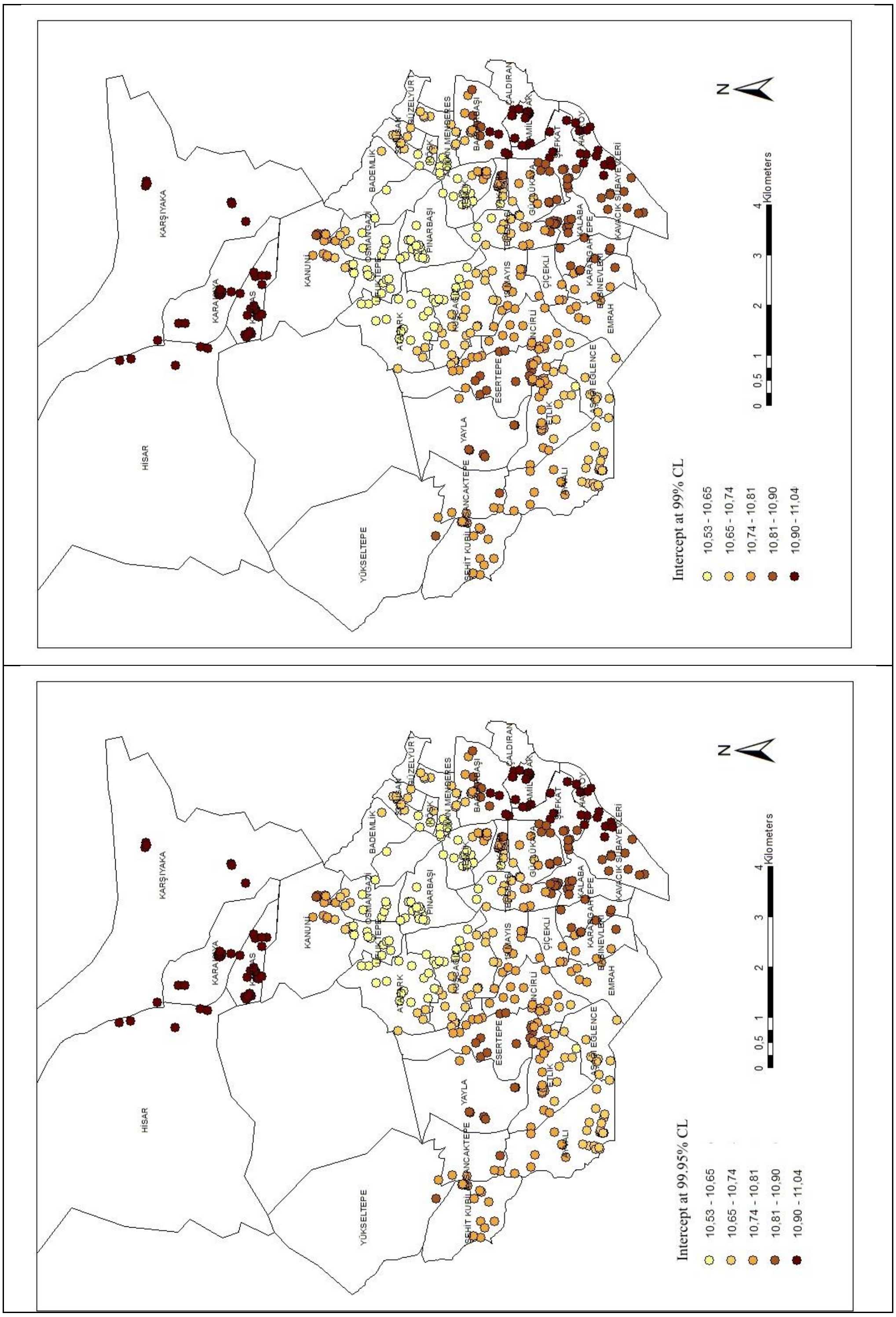


Figure 89 (continued): GWR Results for Floor Coefficients for Keçiören

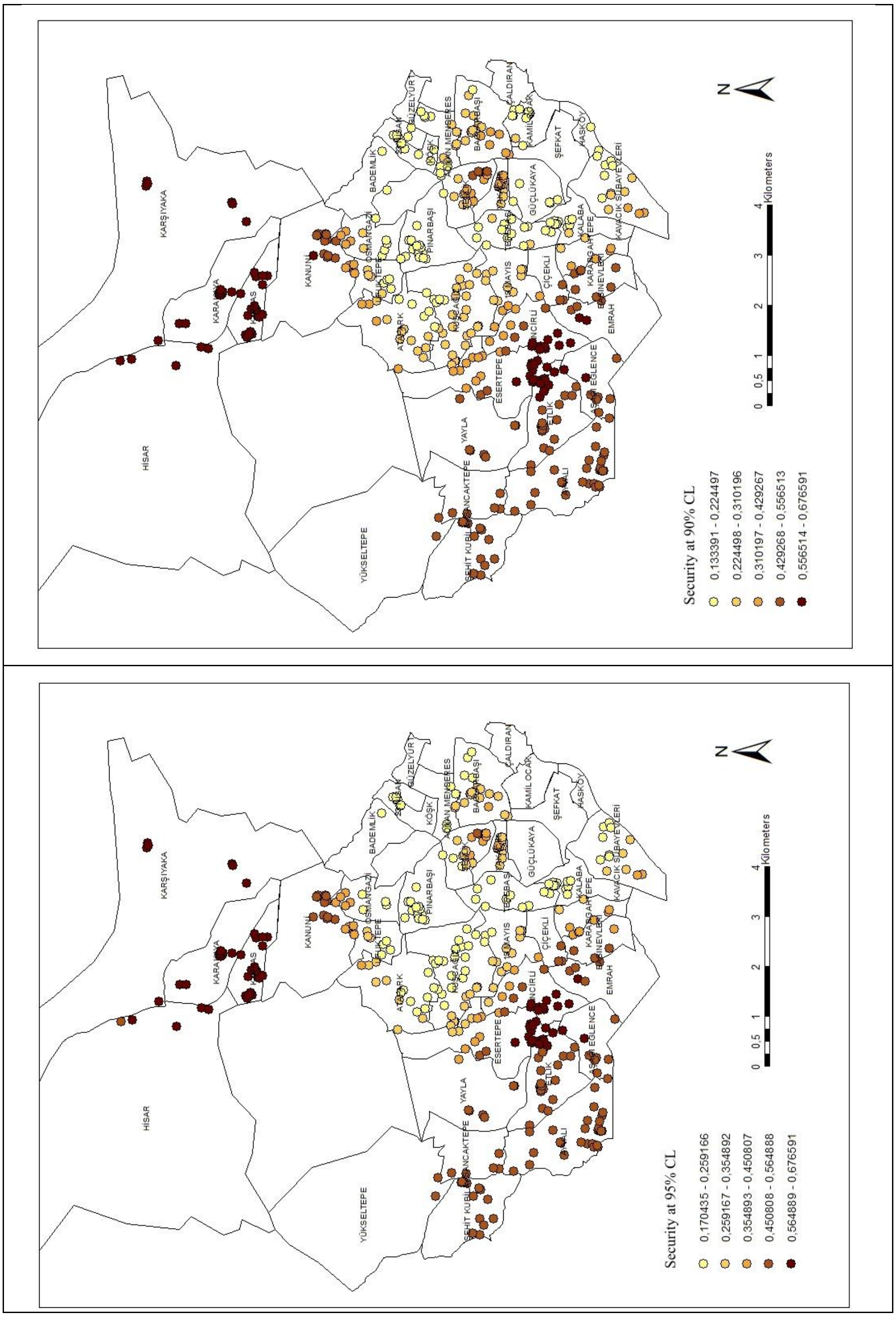


Figure 90: GWR Results for Security Coefficients for Keçiören

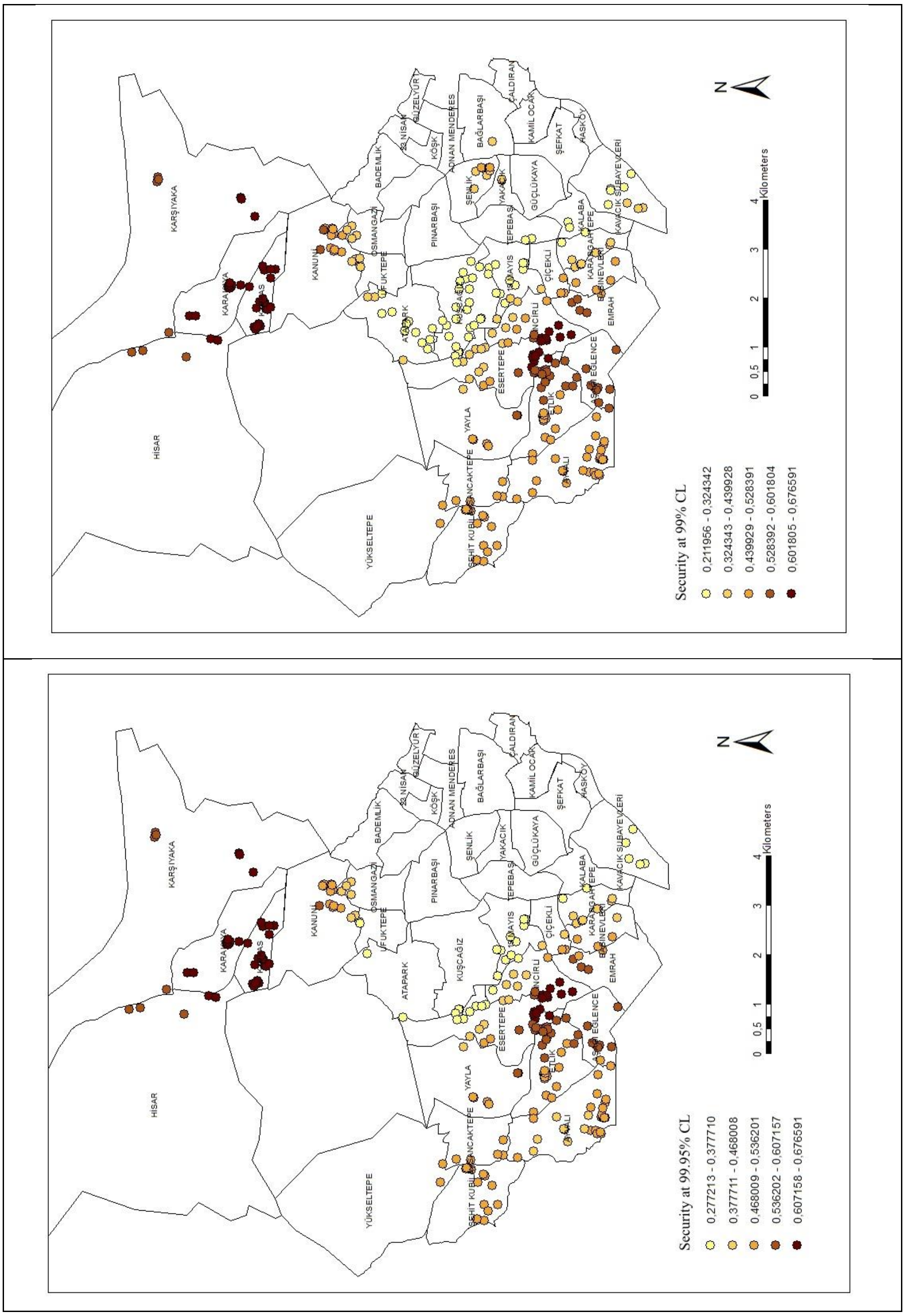


Figure 90 (continued): GWR Results for Security Coefficients for Keçiören

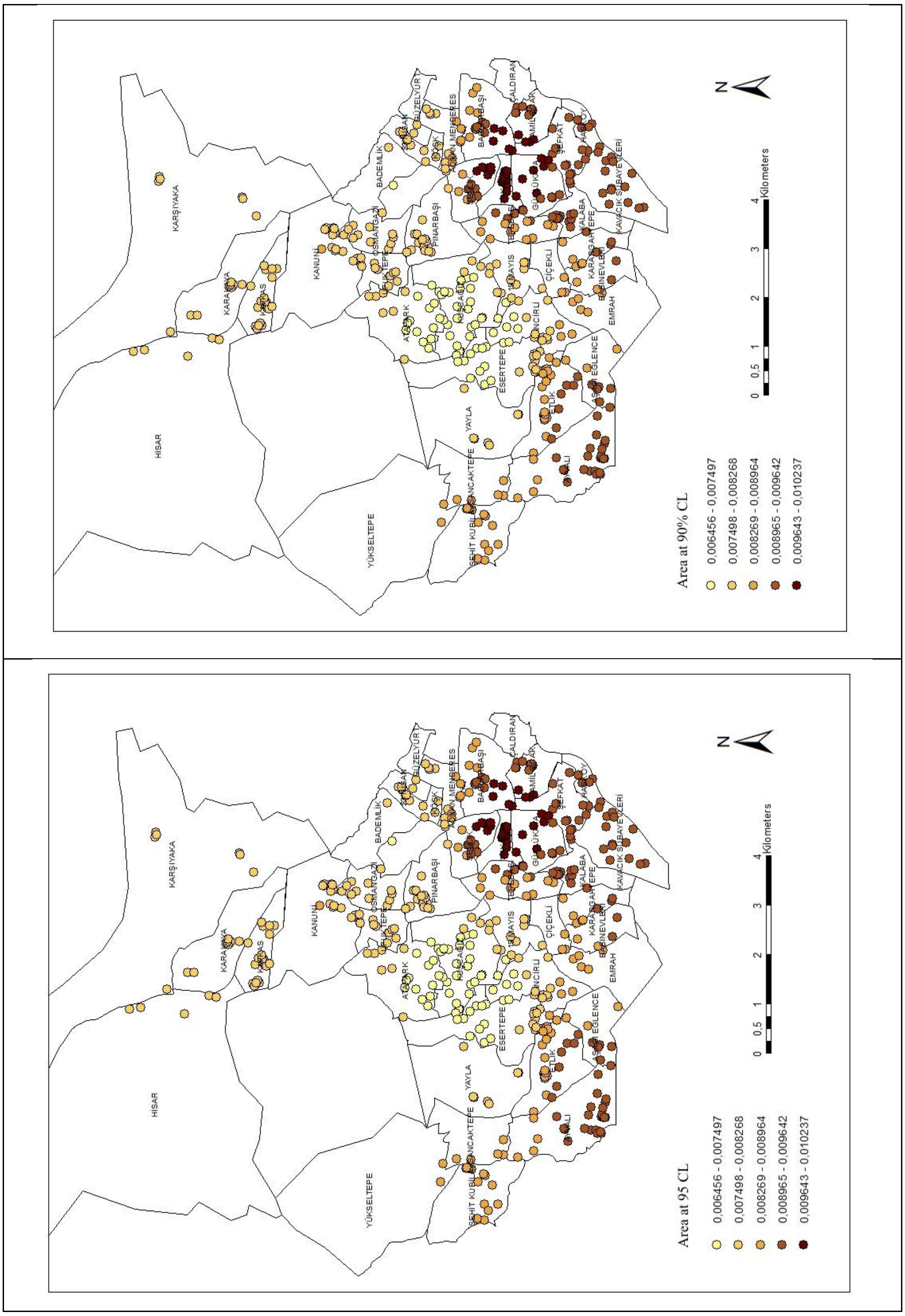


Figure 91: GWR Results for Area Coefficients for Keçiören

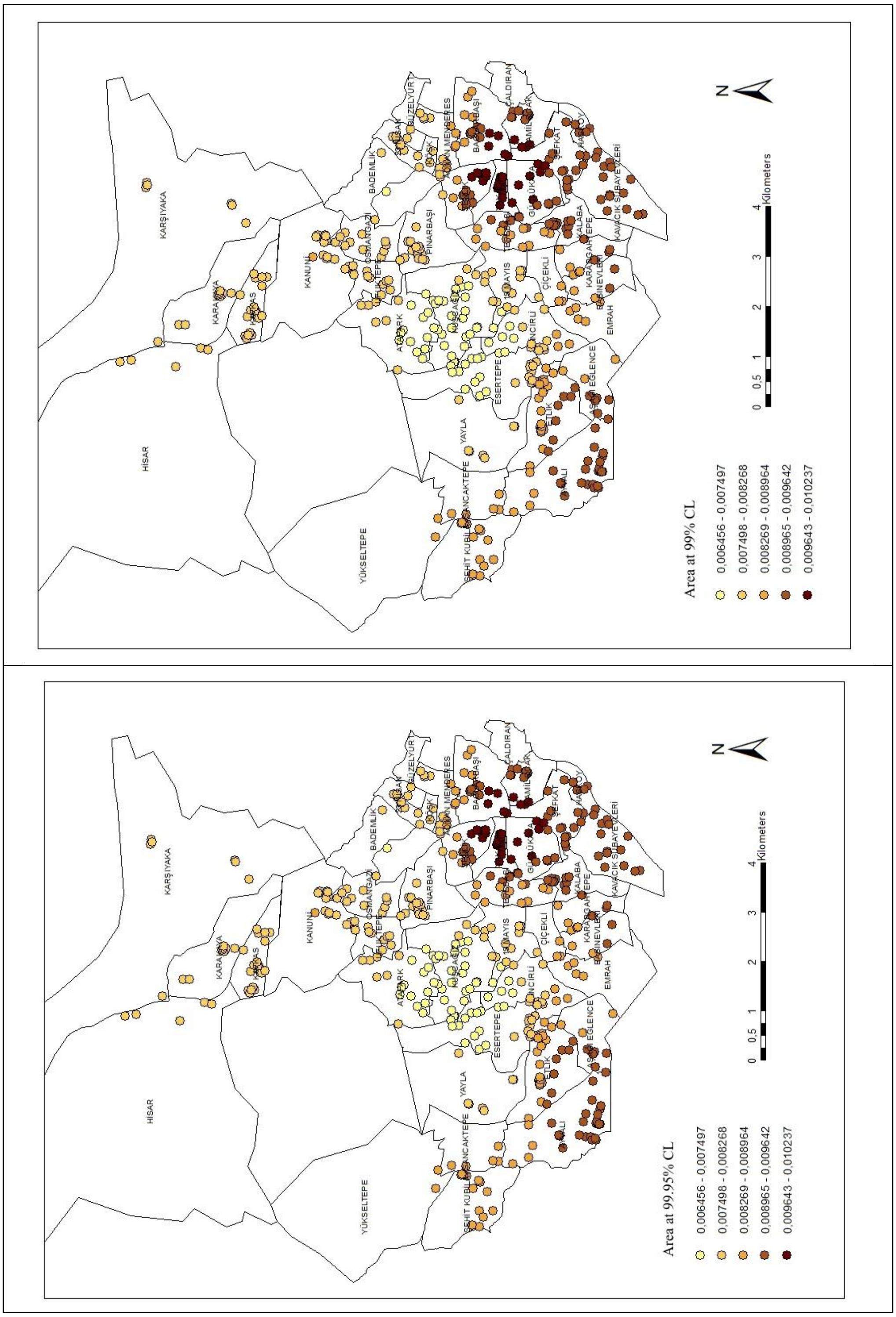


Figure 91 (continued): GWR Results for Area Coefficients for Keçiören

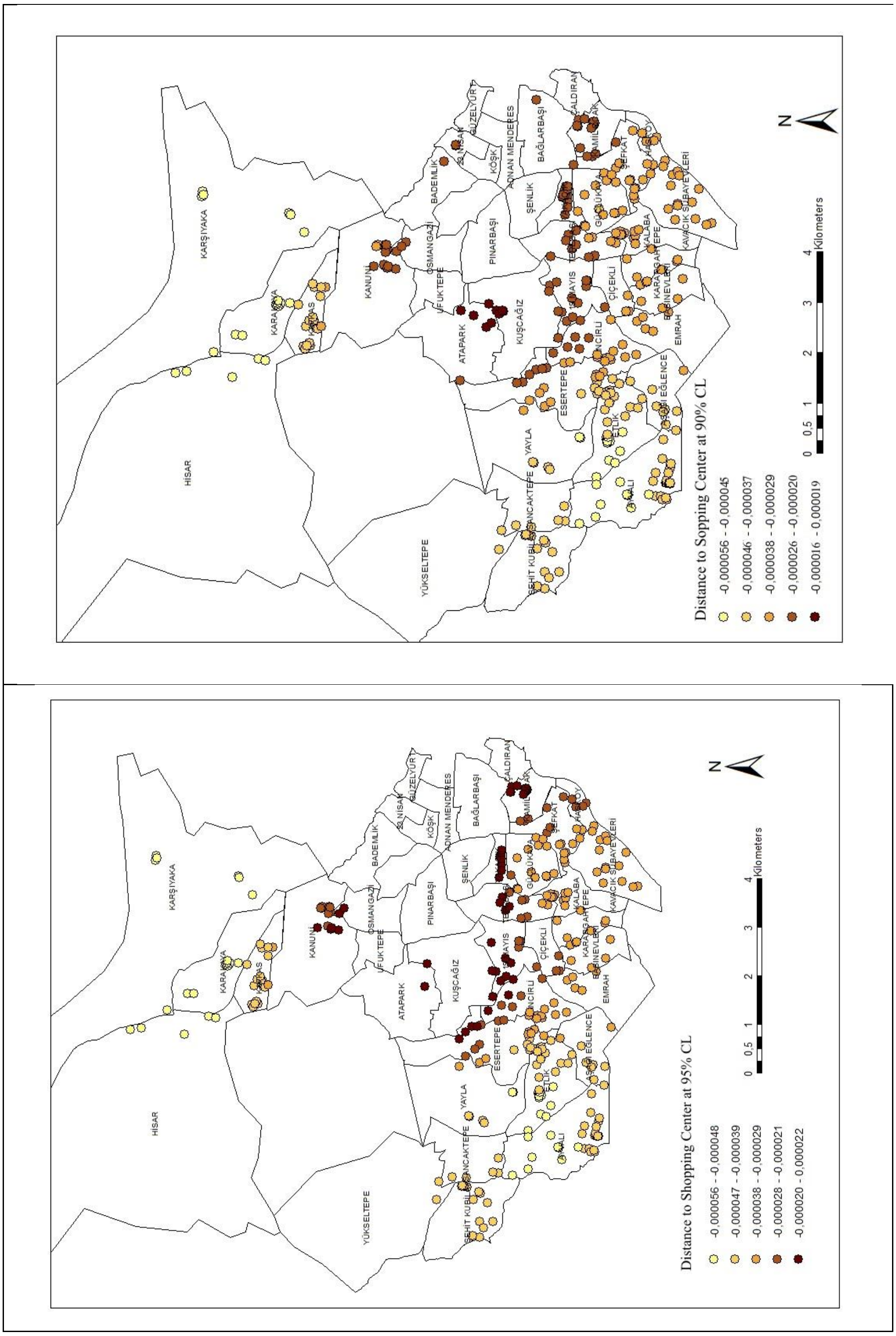


Figure 92: GWR Results for Distance to Shopping Centers Coefficients for Keçiören

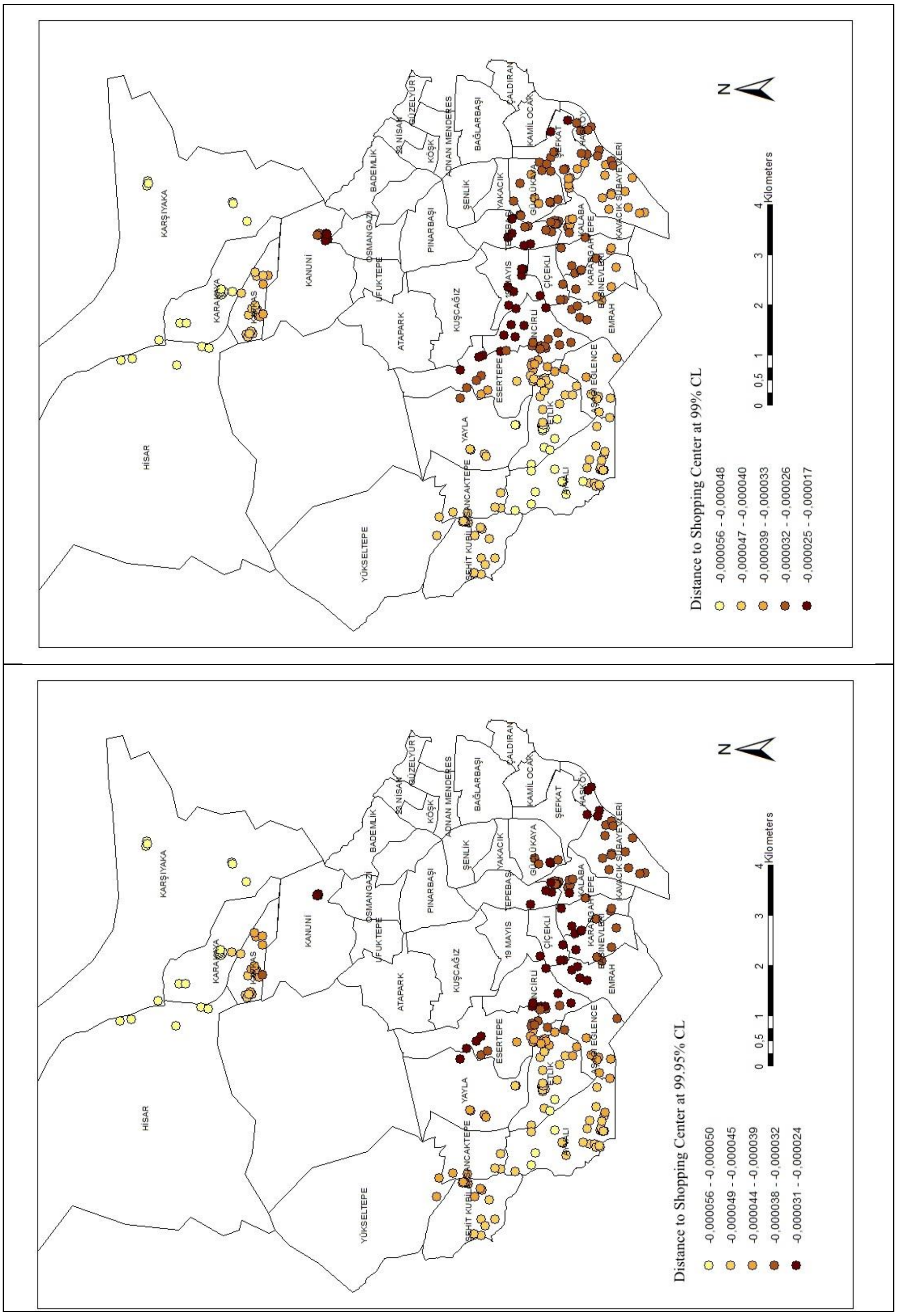


Figure 92 (continued): GWR Results for Distance to Shopping Centers Coefficients for Keçiören

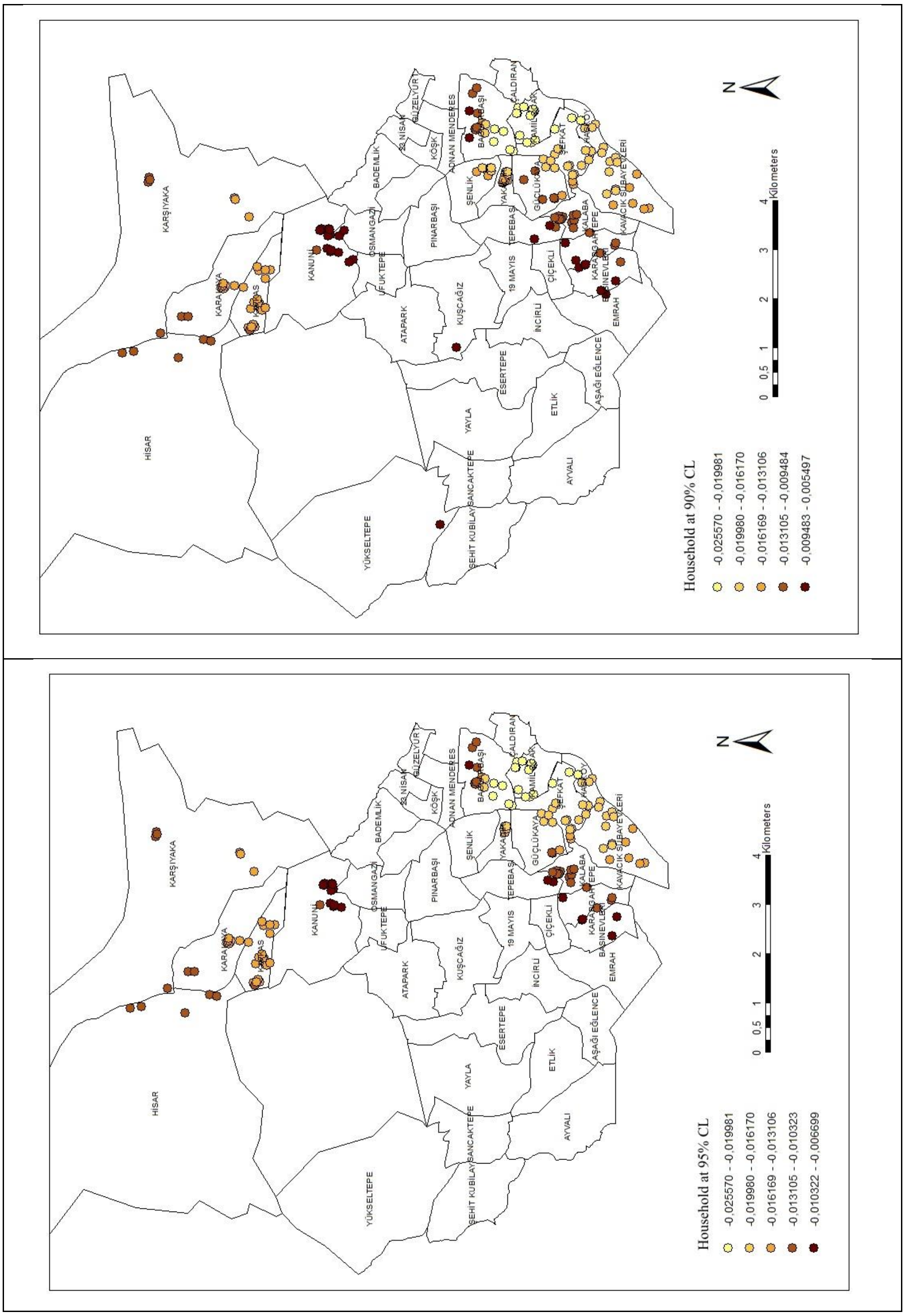


Figure 93: GWR Results for Household Coefficients for Keçiören

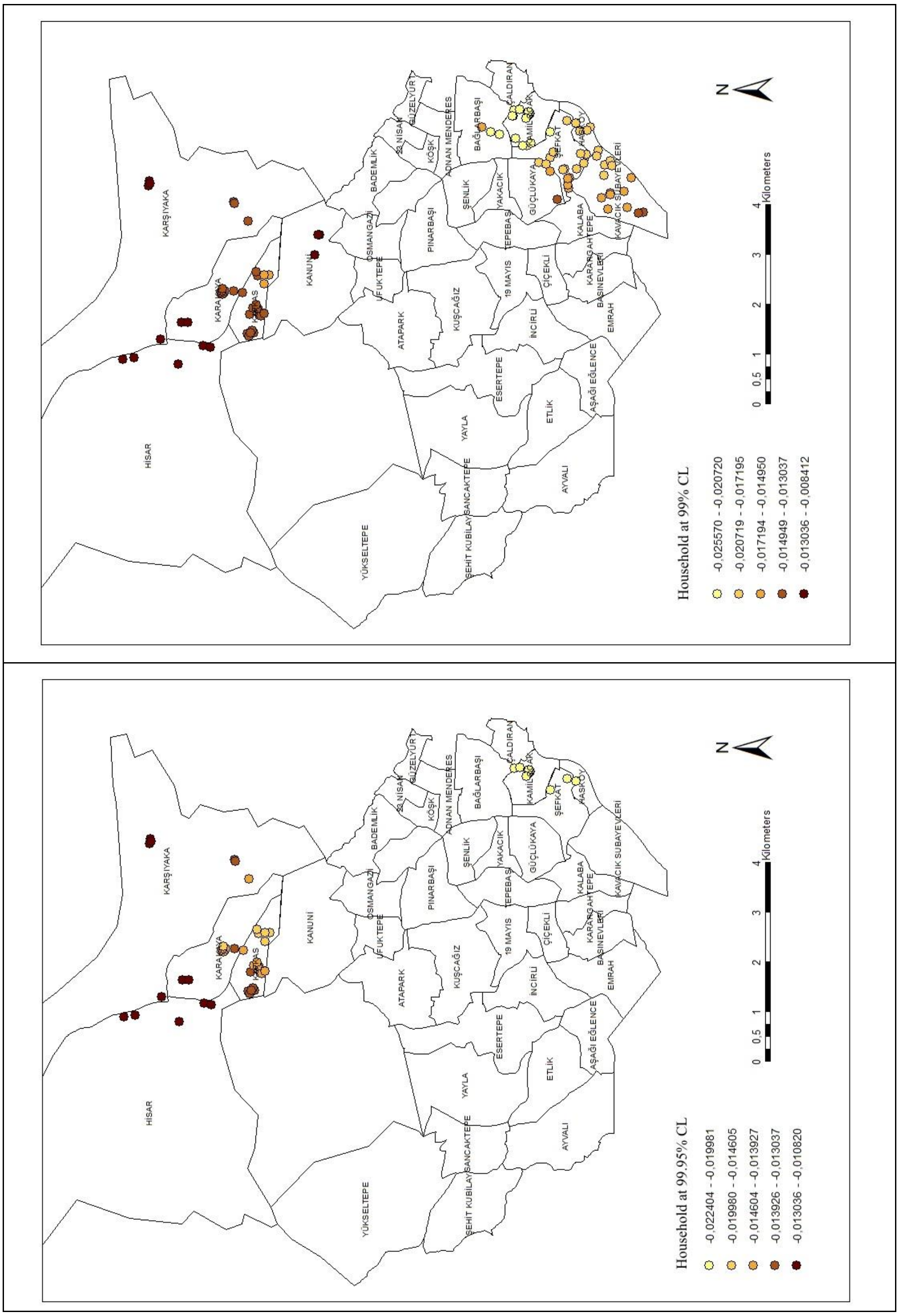


Figure 93 (continued): GWR Results for Household Coefficients for Keçiören

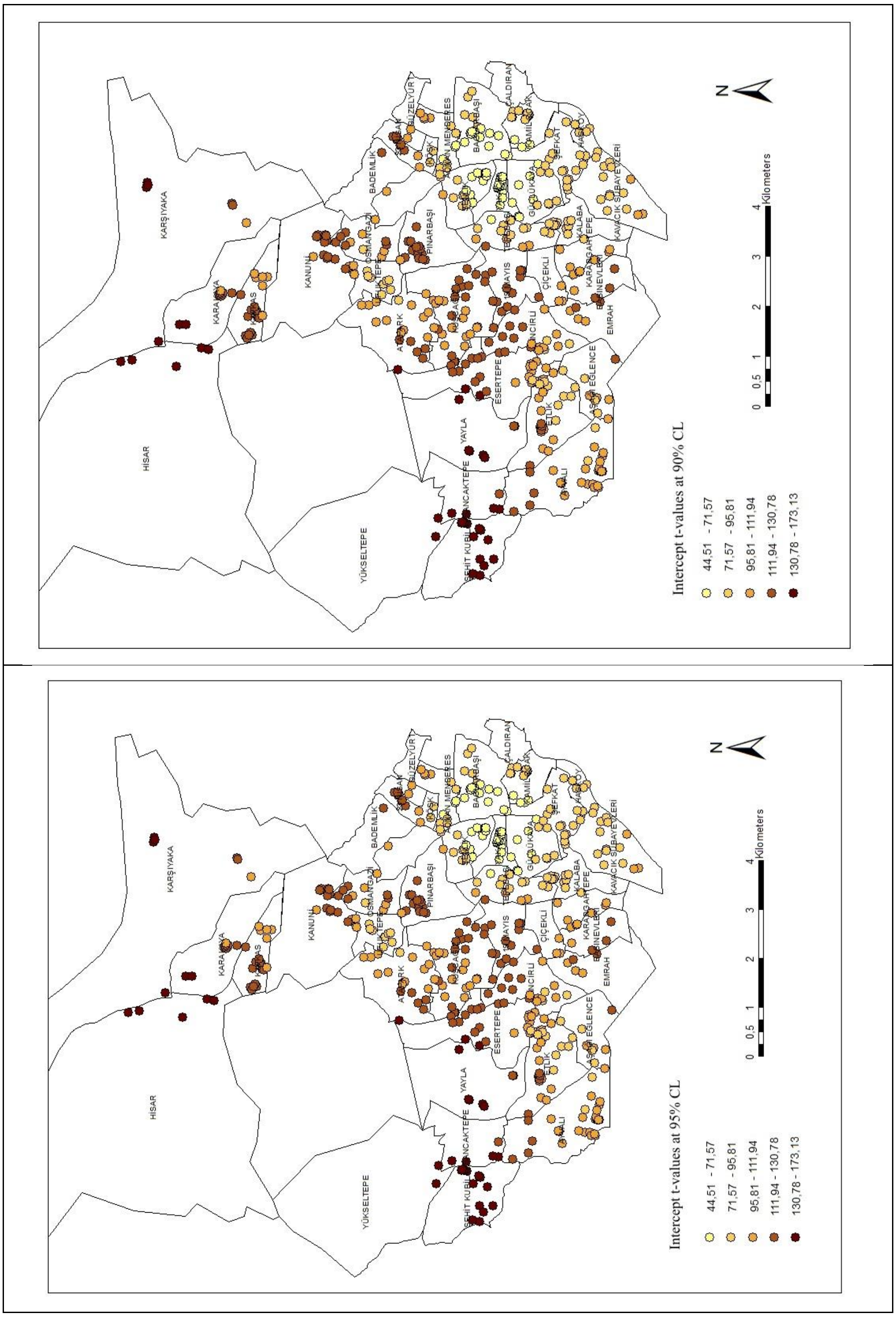


Figure 94: GWR t-Values for Intercept Coefficients for Keçiören

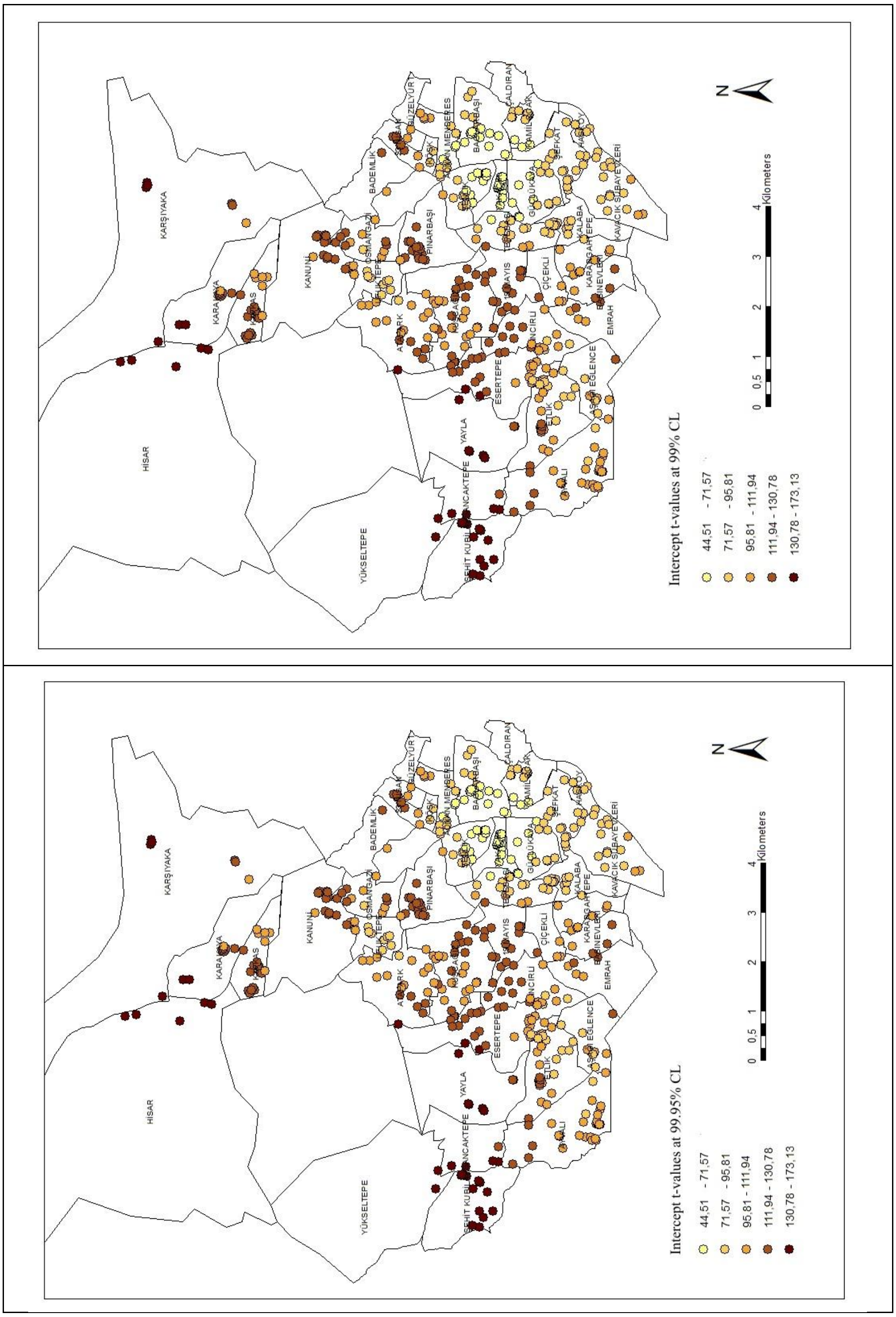


Figure 94 (continued): GWR t-Values for Intercept Coefficients for Keçiören

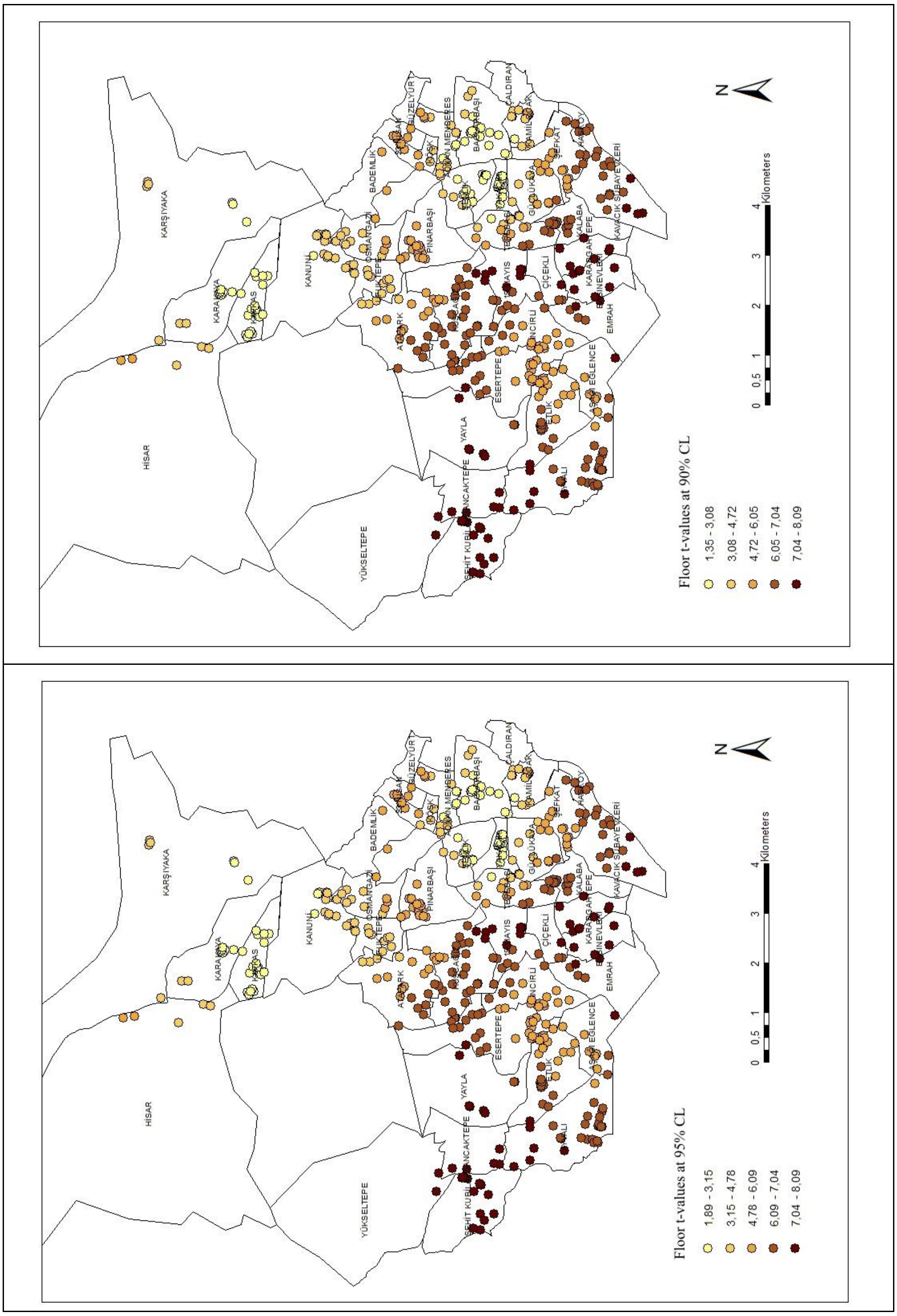


Figure 95: GWR t-Values for Floor Attribute for Keçiören

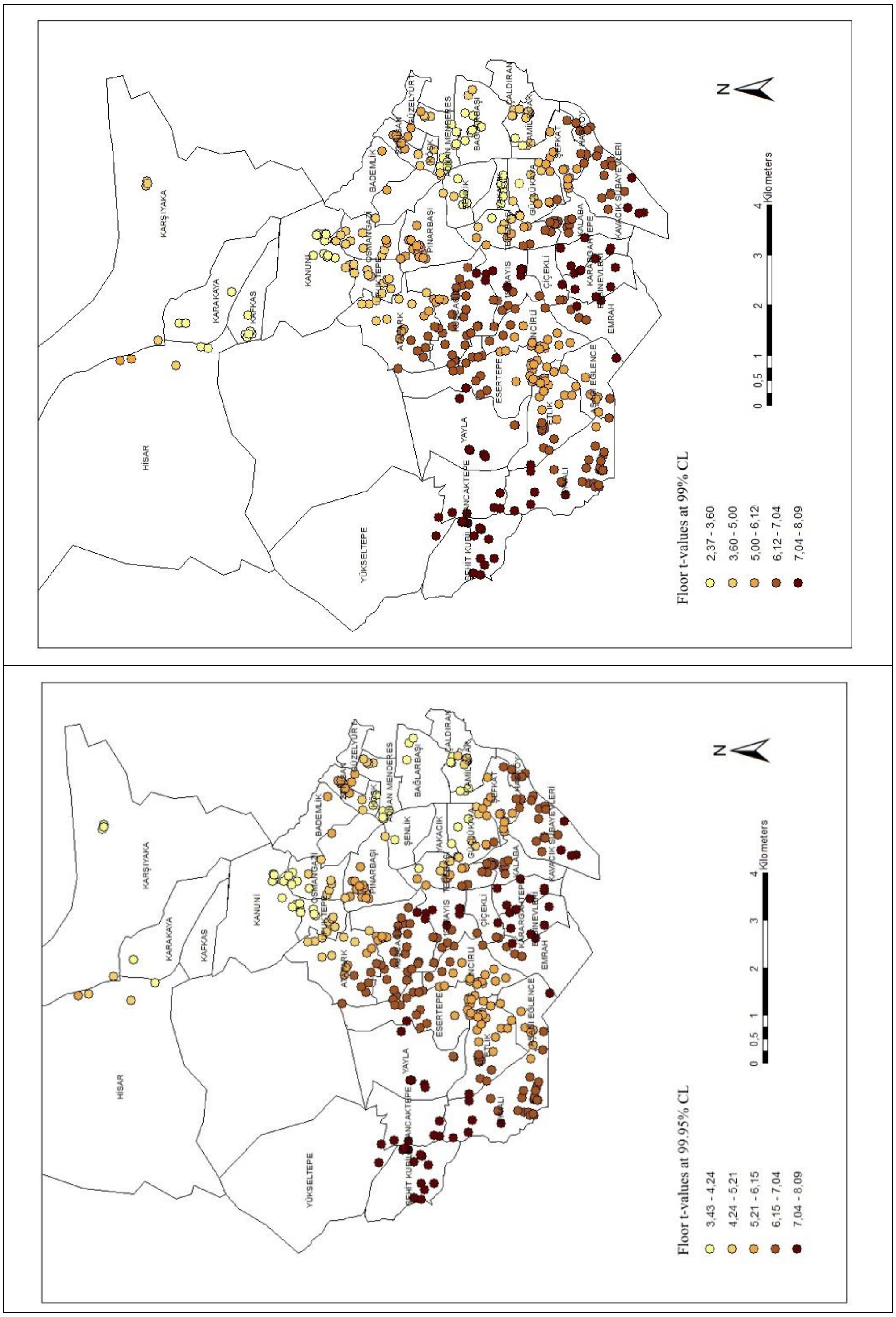


Figure 95 (continued): GWR t-Values for Floor Attribute for Keçiören

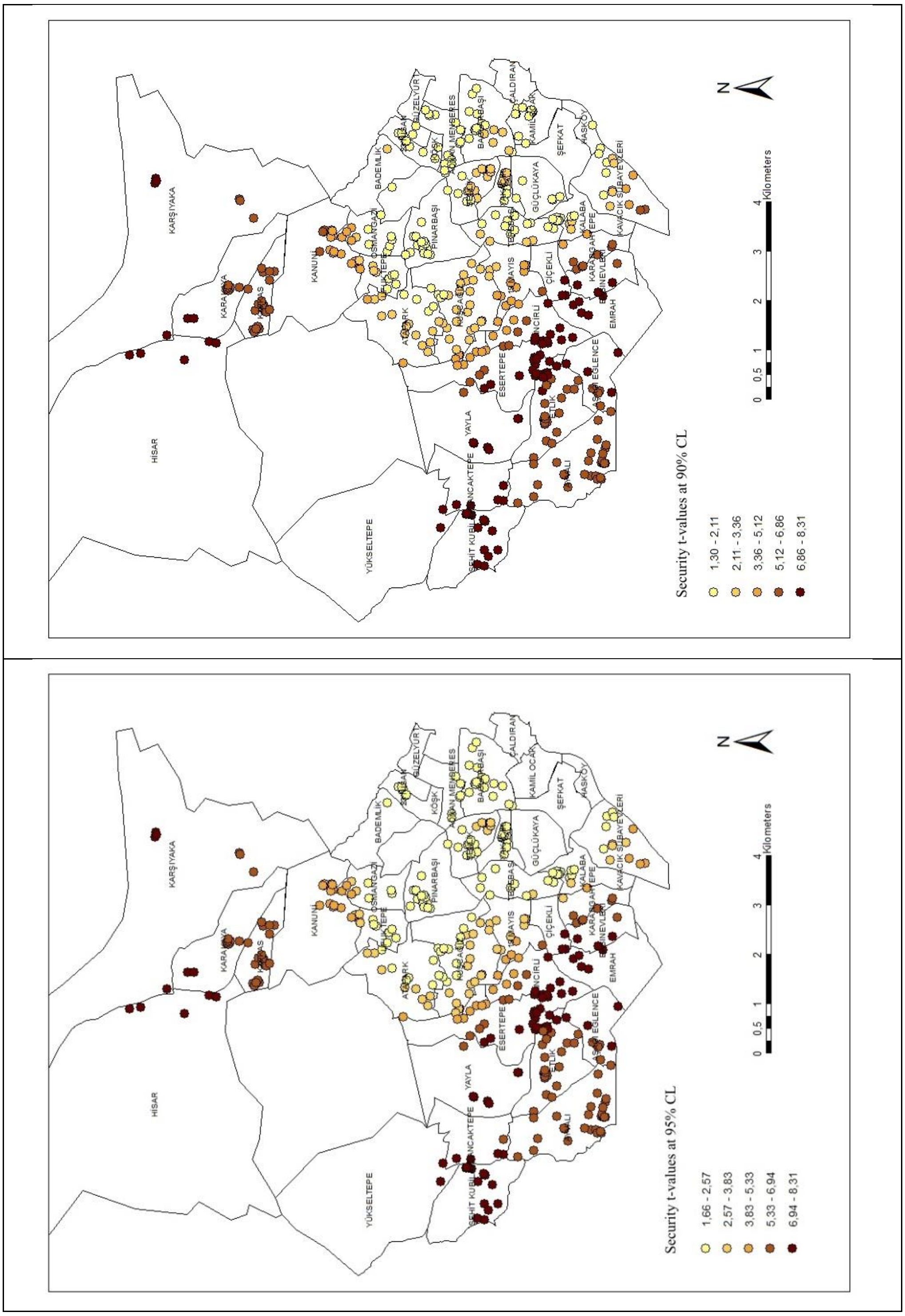


Figure 96: GWR t-Values for Security Attribute for Keçiören

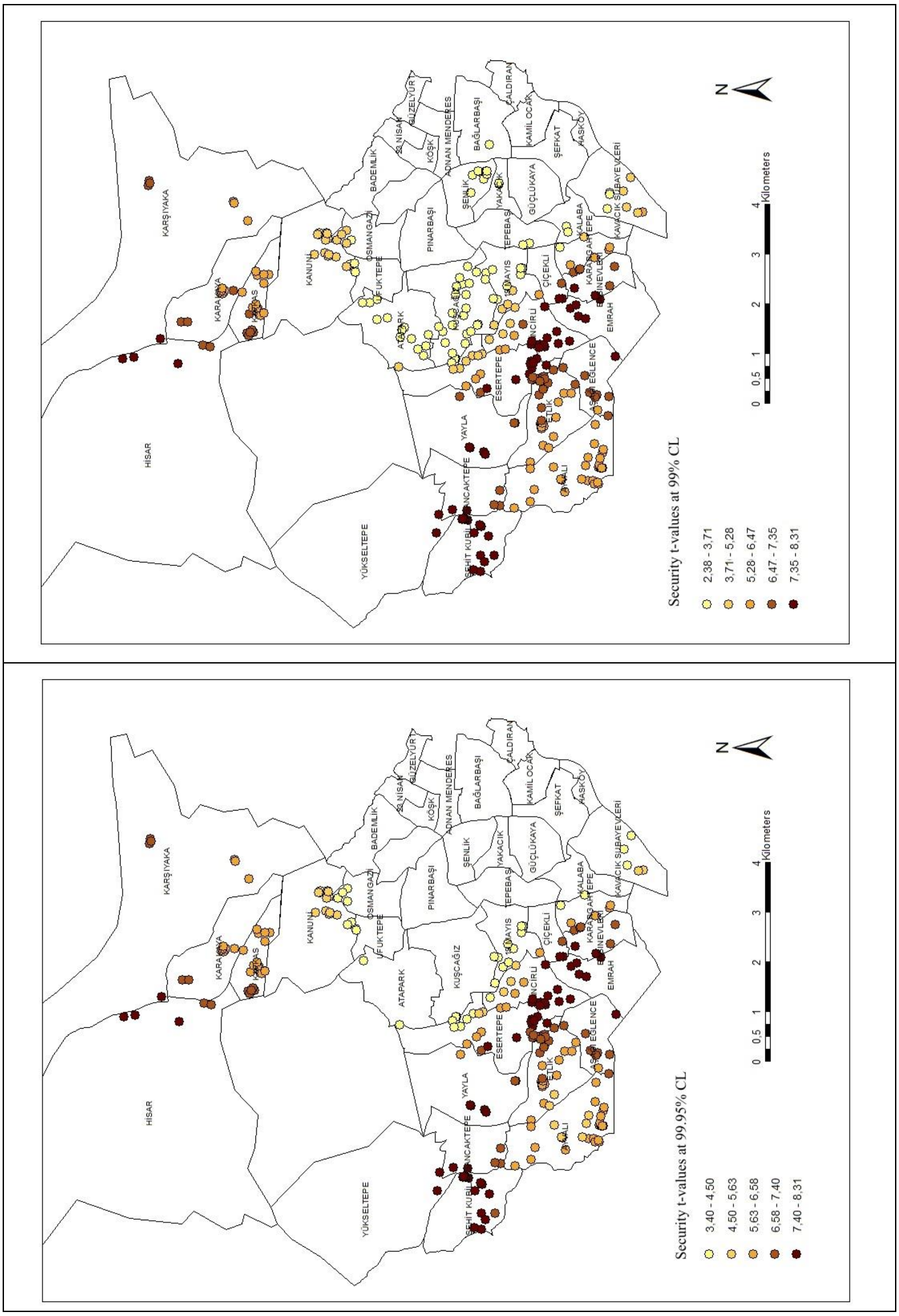


Figure 96 (continued): GWR t-Values for Security Attribute for Keçiören

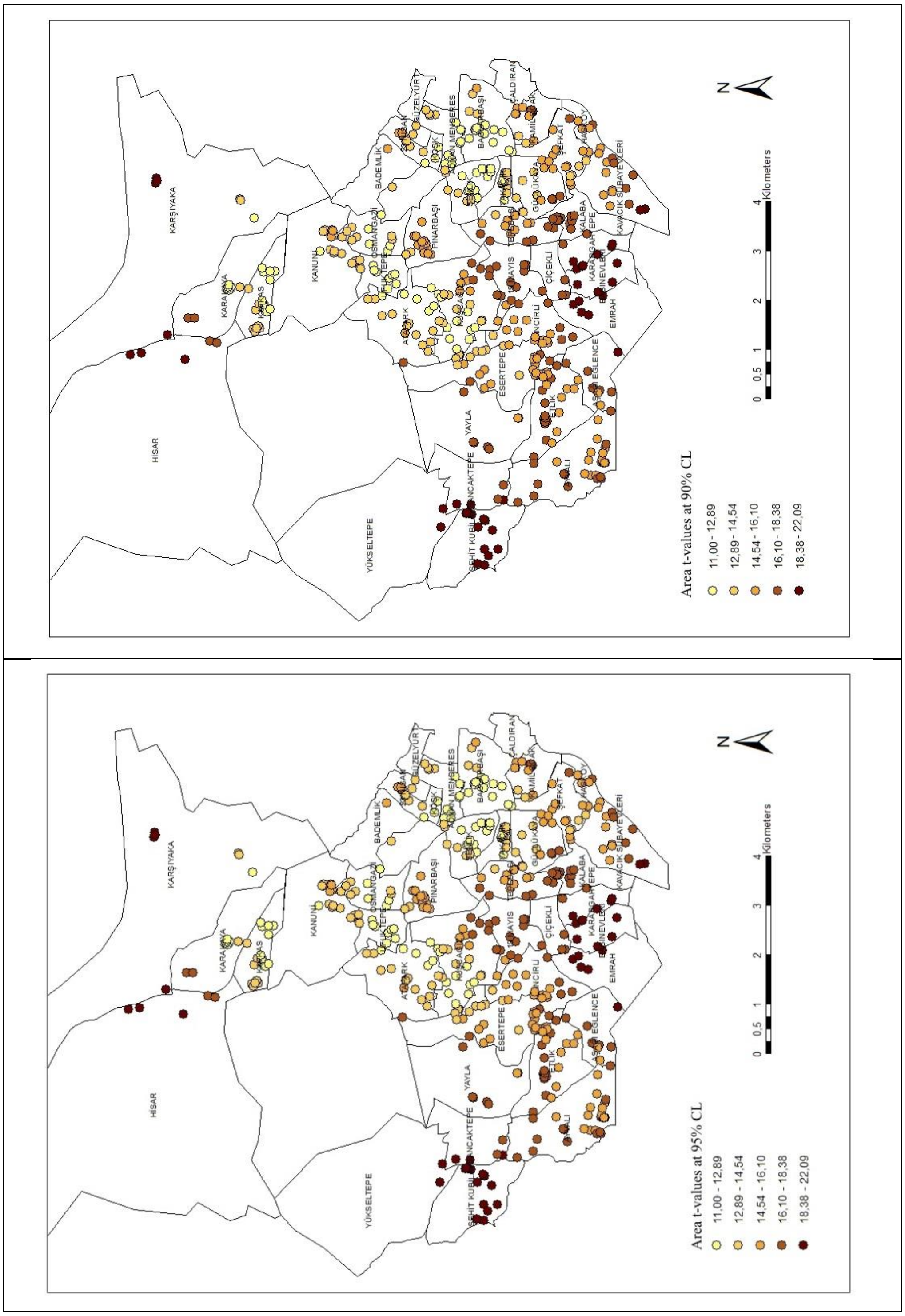


Figure 97: GWR t-Values for Area Type Attribute for Keçiören

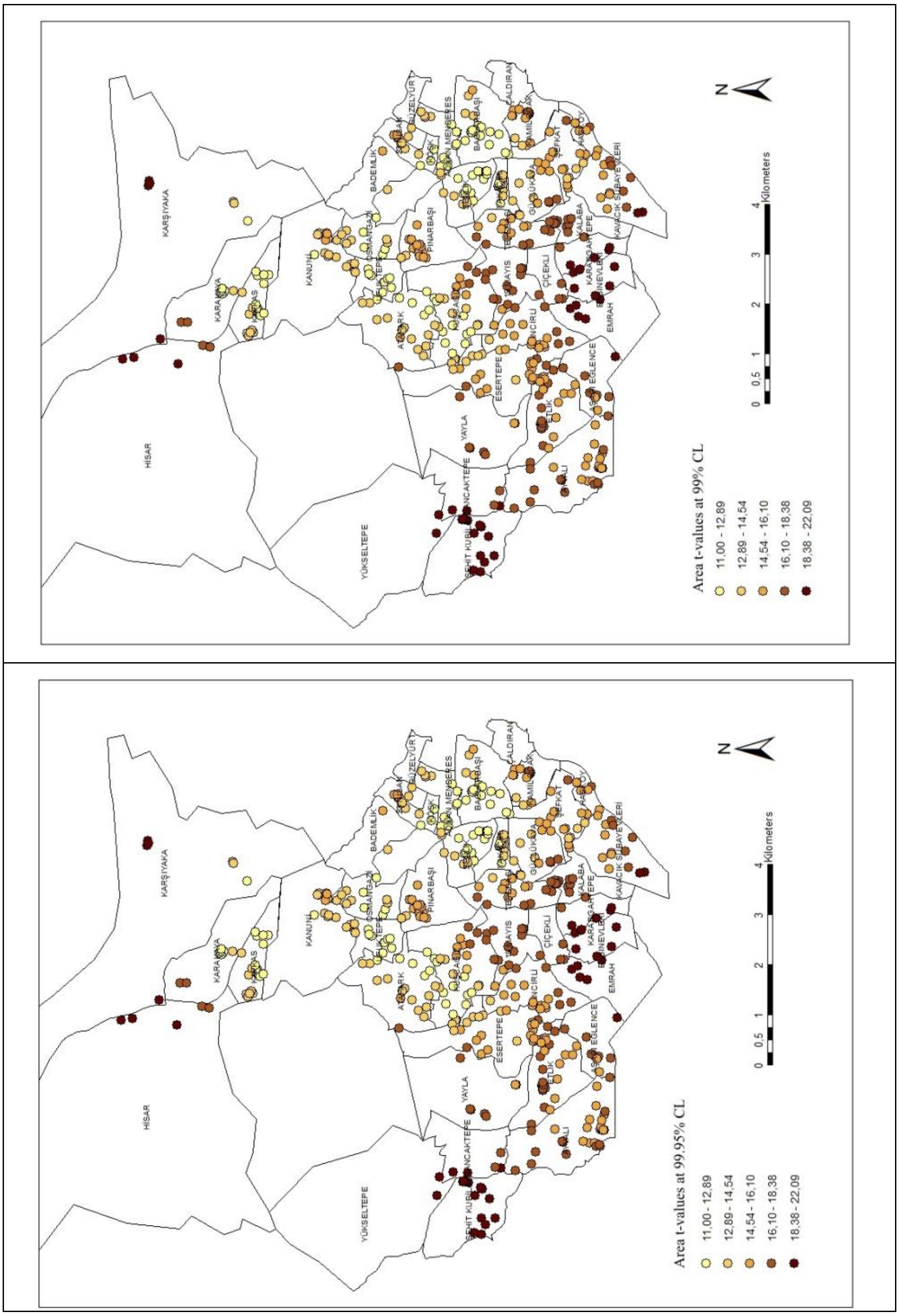


Figure 97 (continued): GWR t-Values for Area Type Attribute for Keçiören

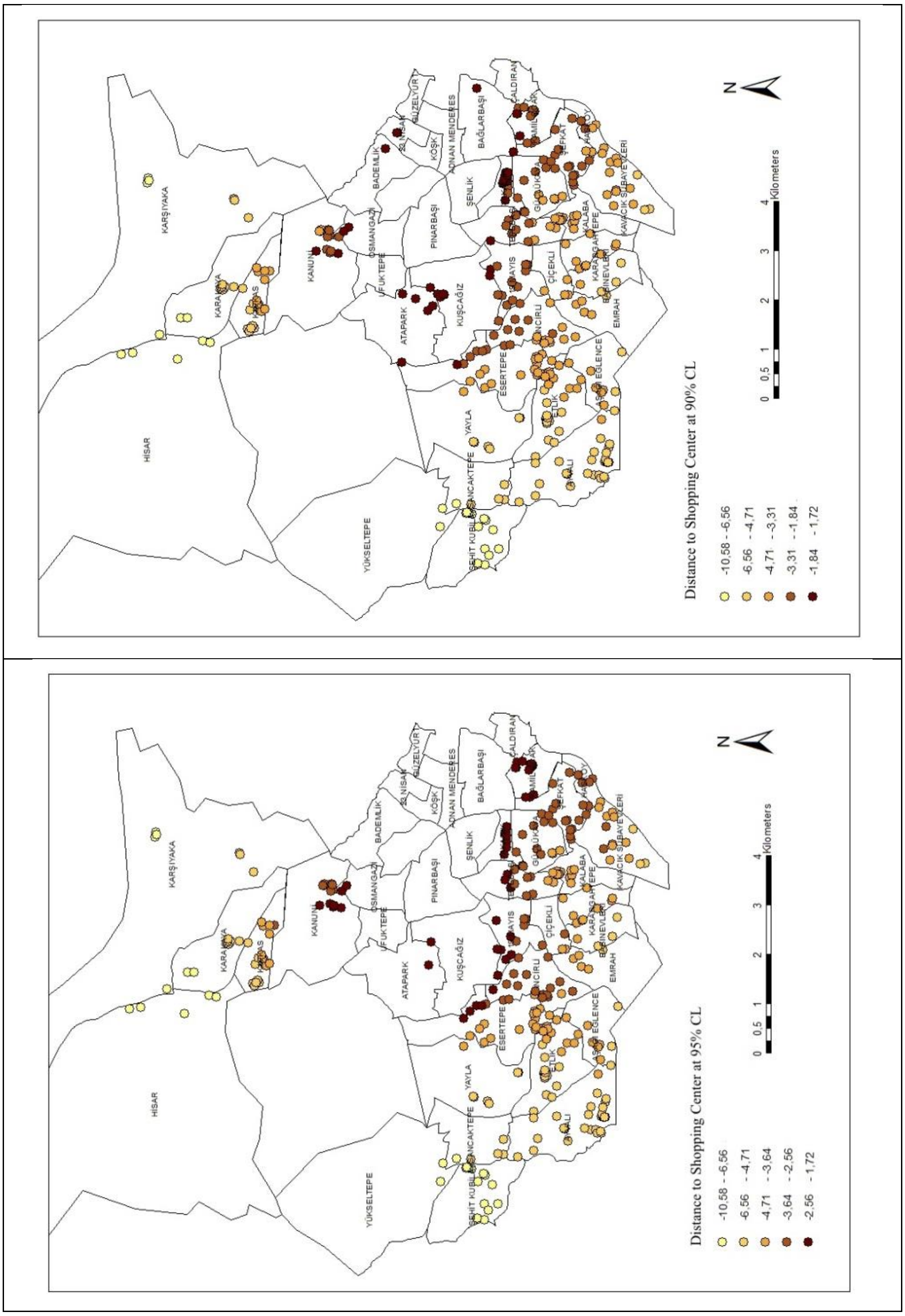


Figure 98: GWR t-Values for Distance to Shopping Center Attribute for Keçiören

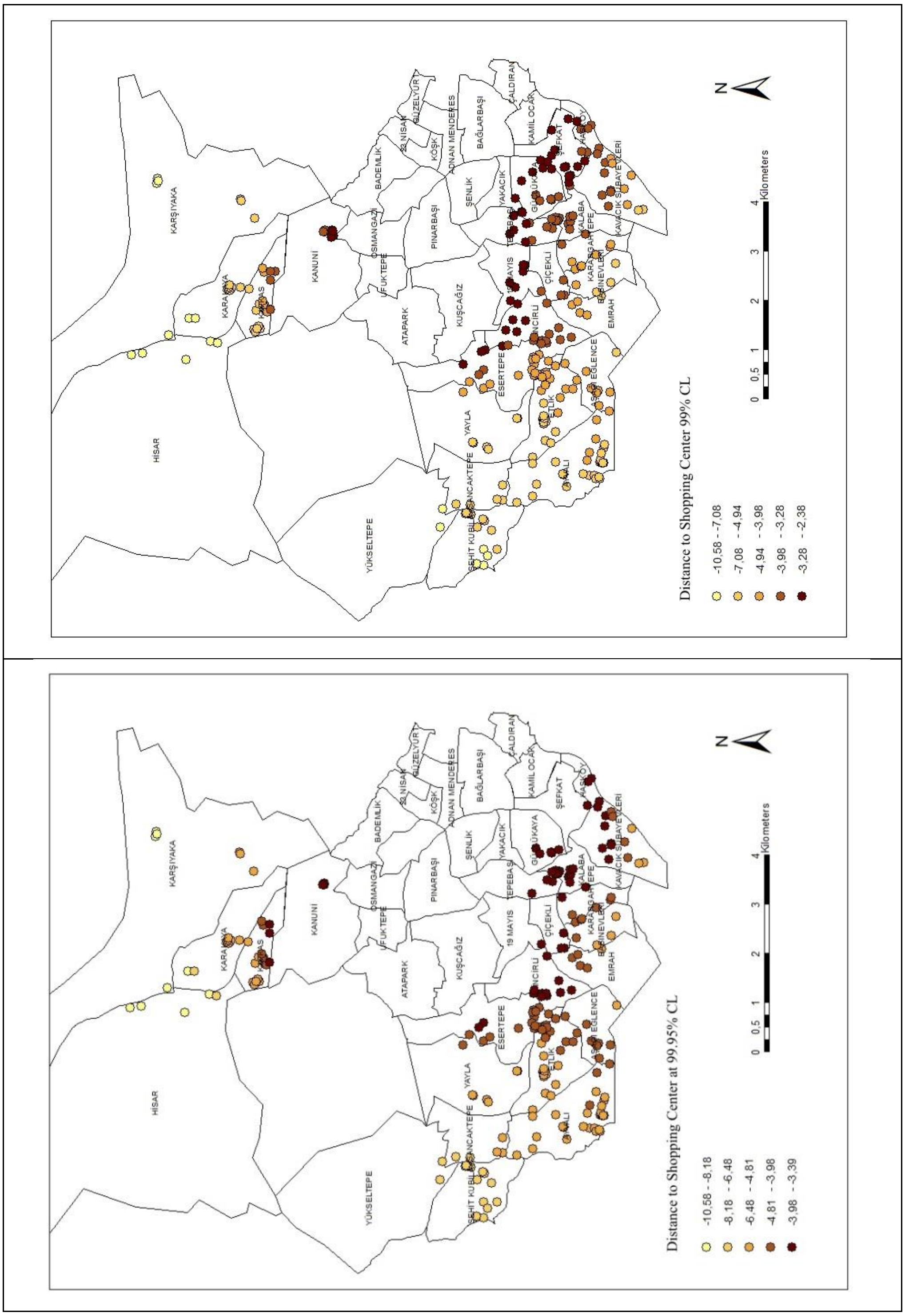


Figure 98 (continued): GWR t-Values for Distance to Shopping Center Attribute for Keçiören

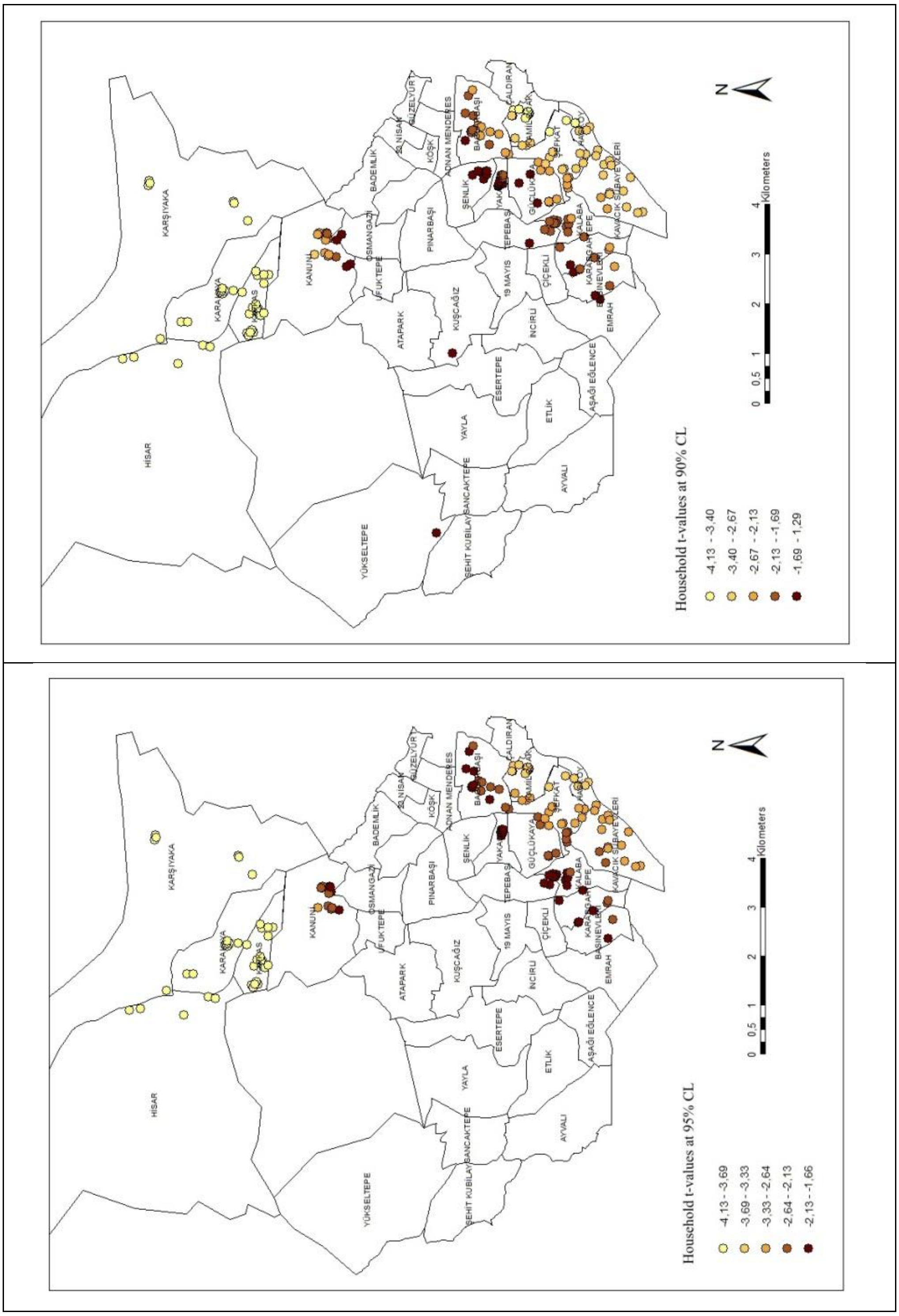


Figure 99: GWR t-Values for Household Attribute for Keçiören

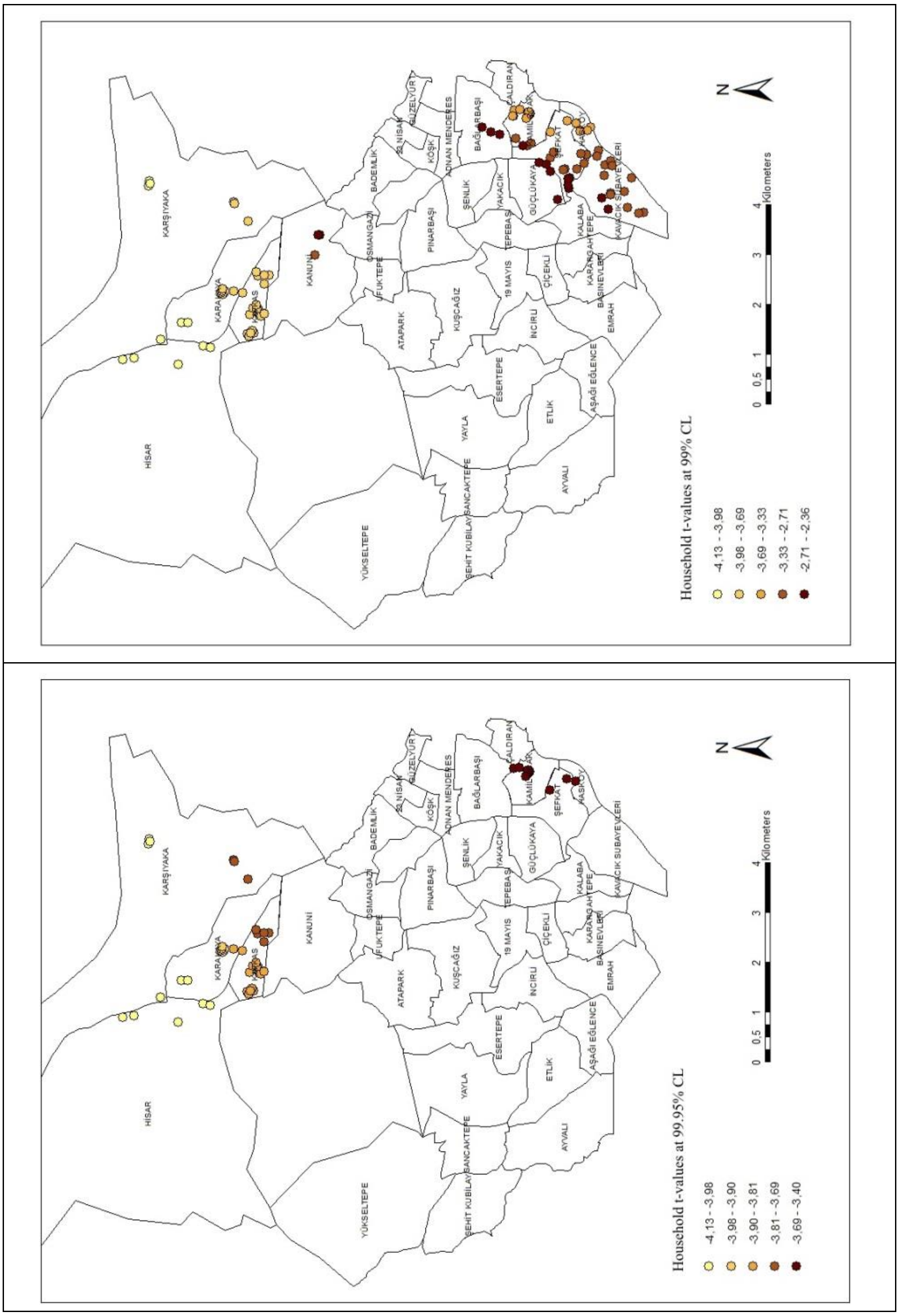


Figure 99 (continued): GWR t-Values for Household Attribute for Keçiören

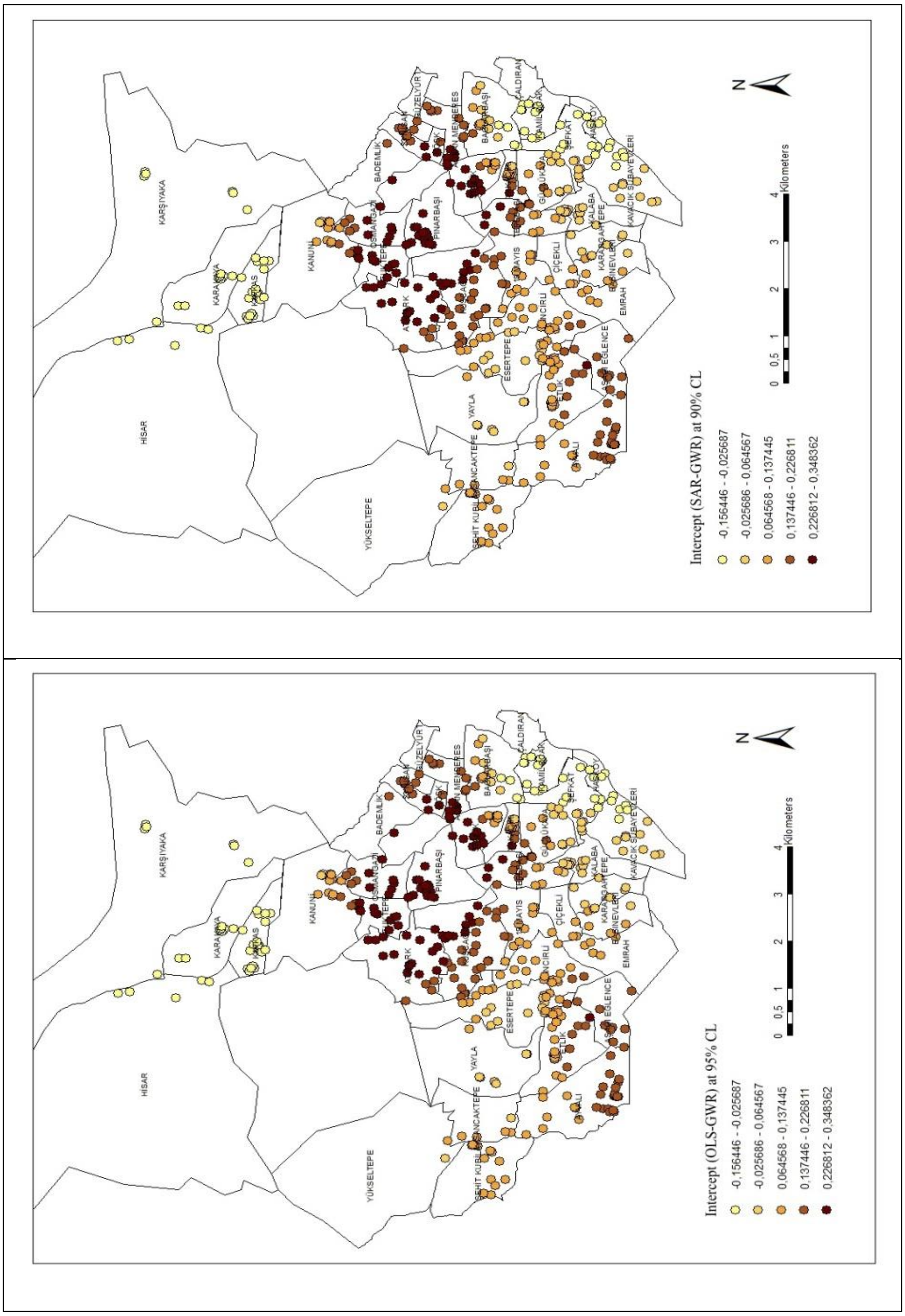


Figure 100: OLS-GWR Comparison Maps for Intercept Coefficients for Keçiören

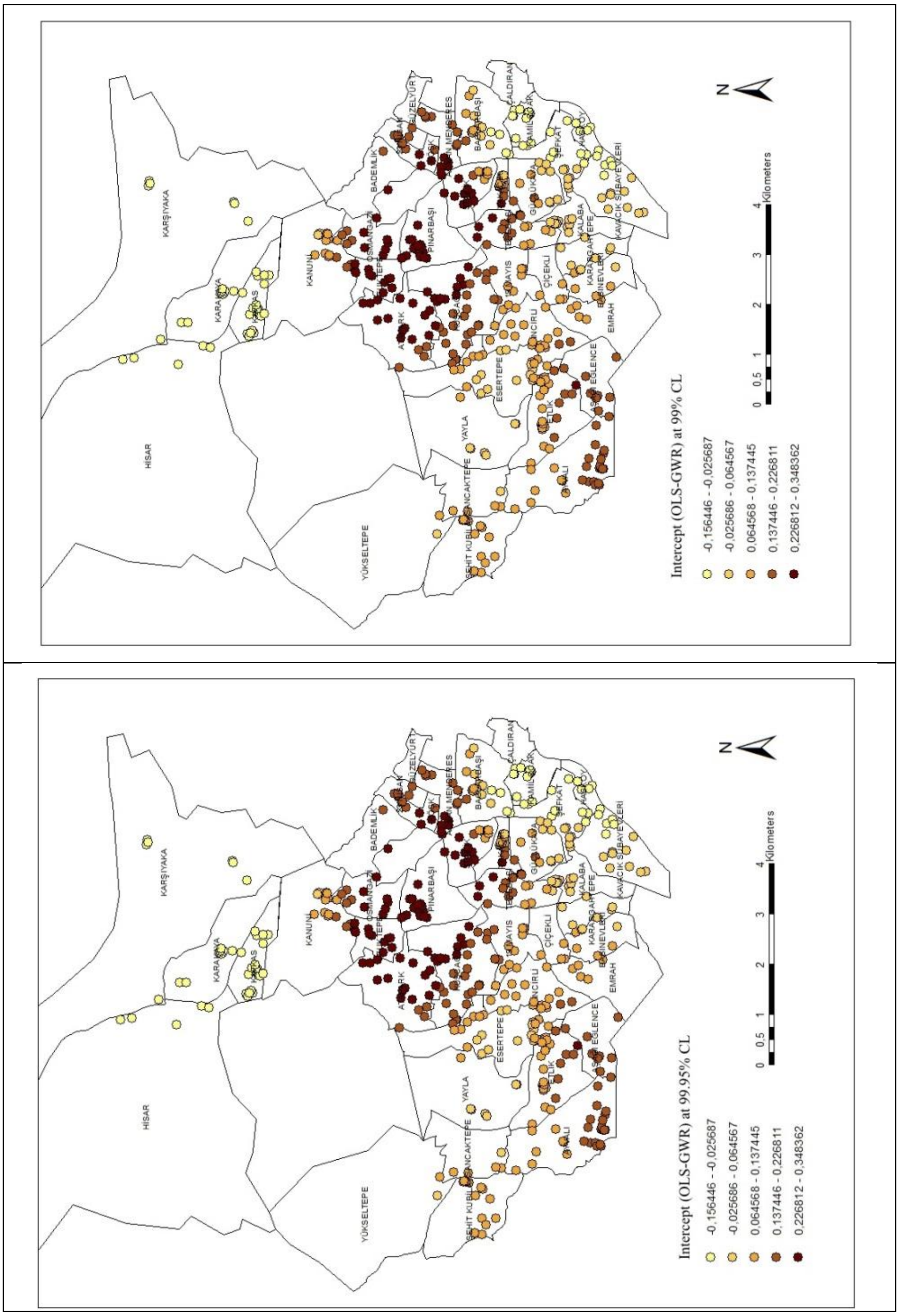


Figure 100 (continued): OLS-GWR Comparison Maps for Intercept Coefficients for Keçiören

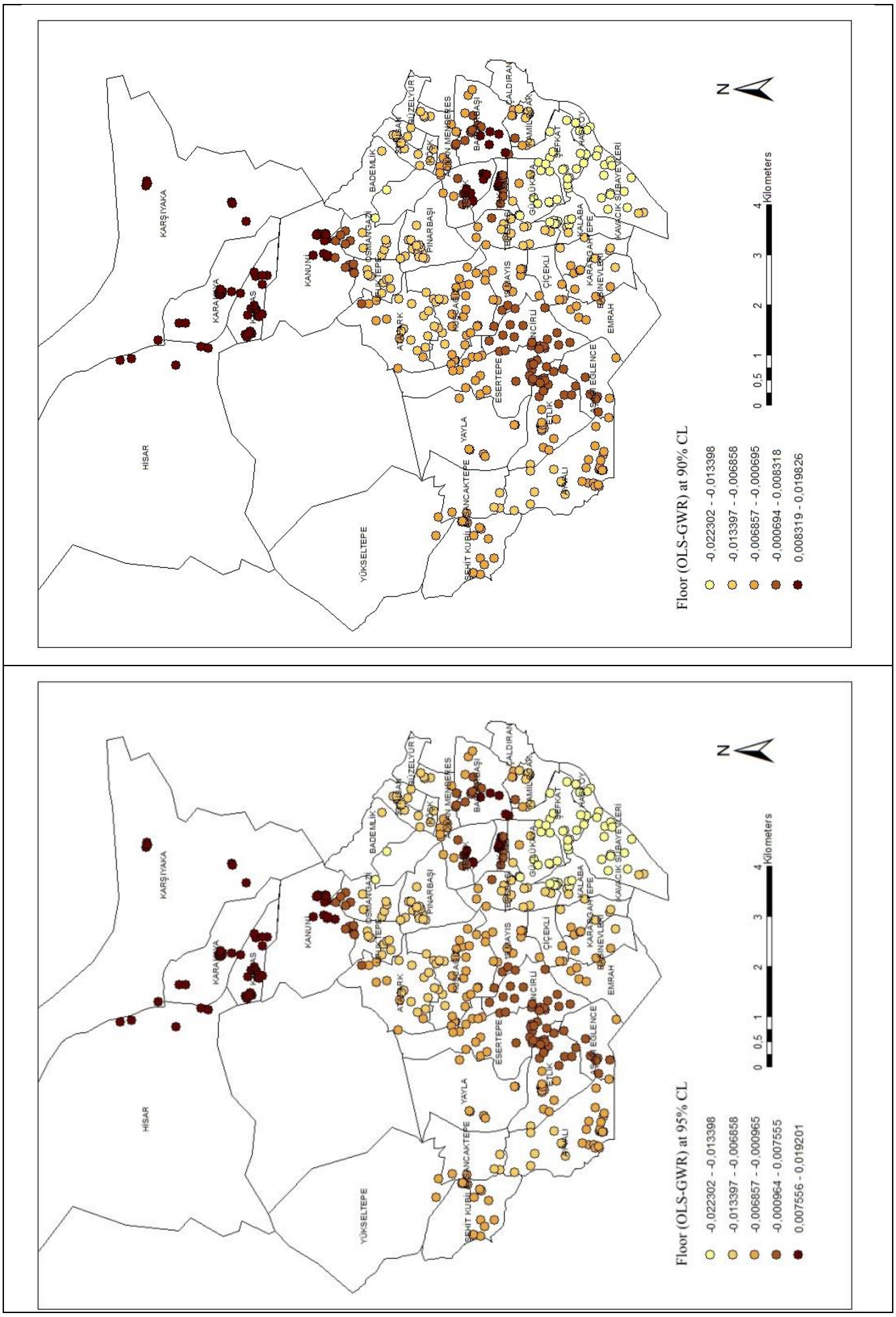


Figure 101: OLS-GWR Comparison Maps for Floor Coefficients for Keçiören

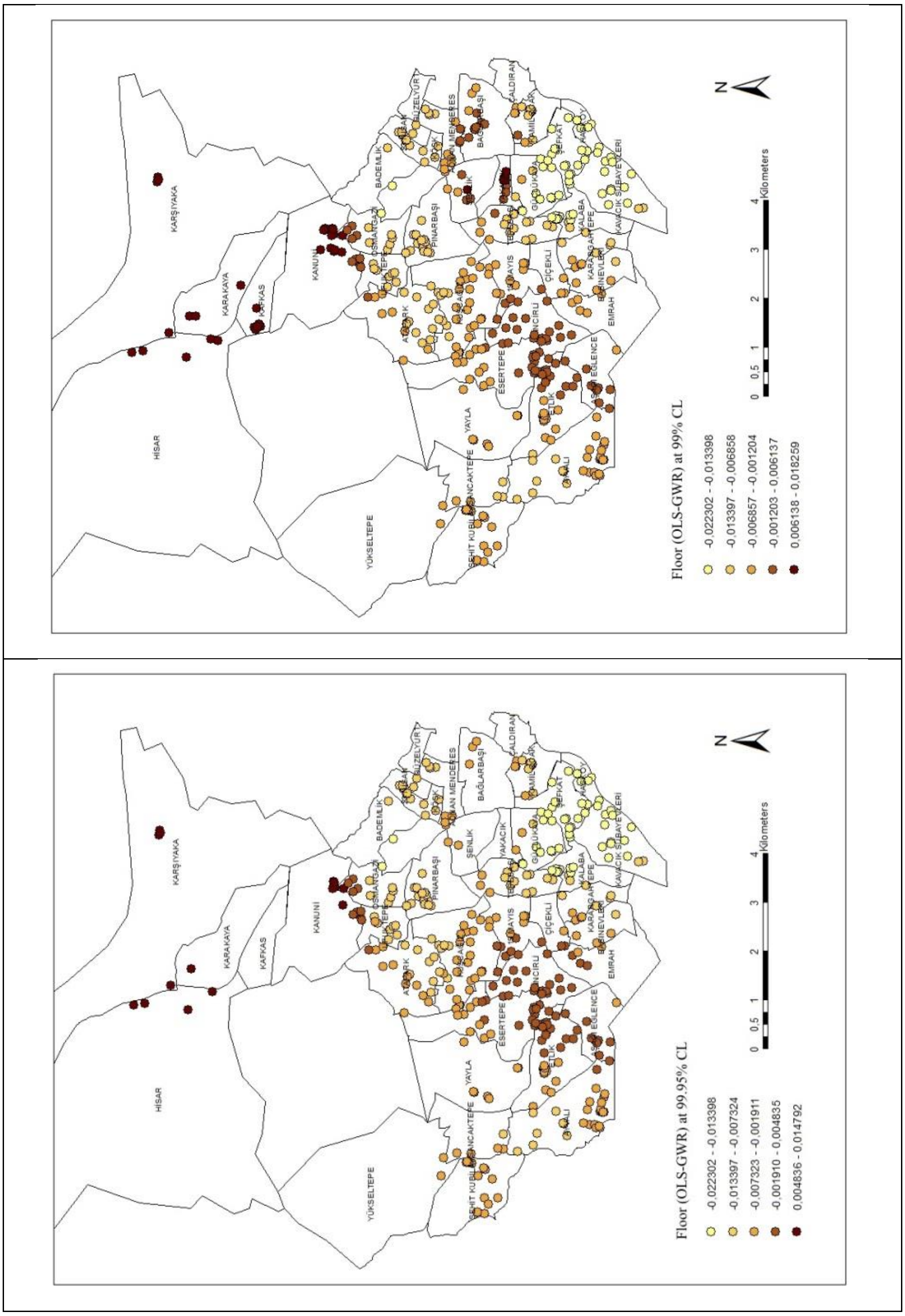


Figure 101 (continued): OLS-GWR Comparison Maps for Floor Coefficients for Keçiören

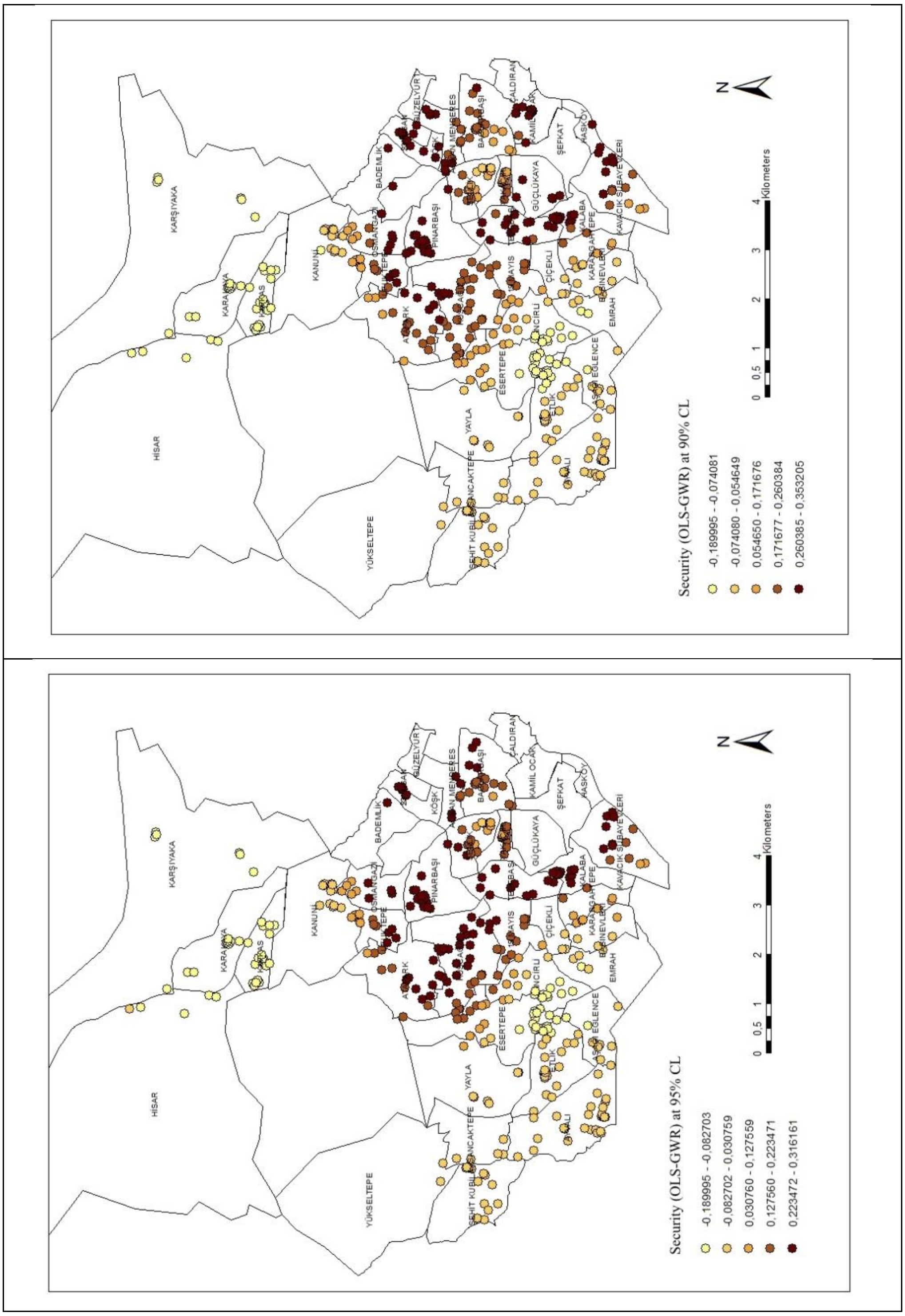


Figure 102: OLS-GWR Comparison Maps for Security Coefficients for Keçiören

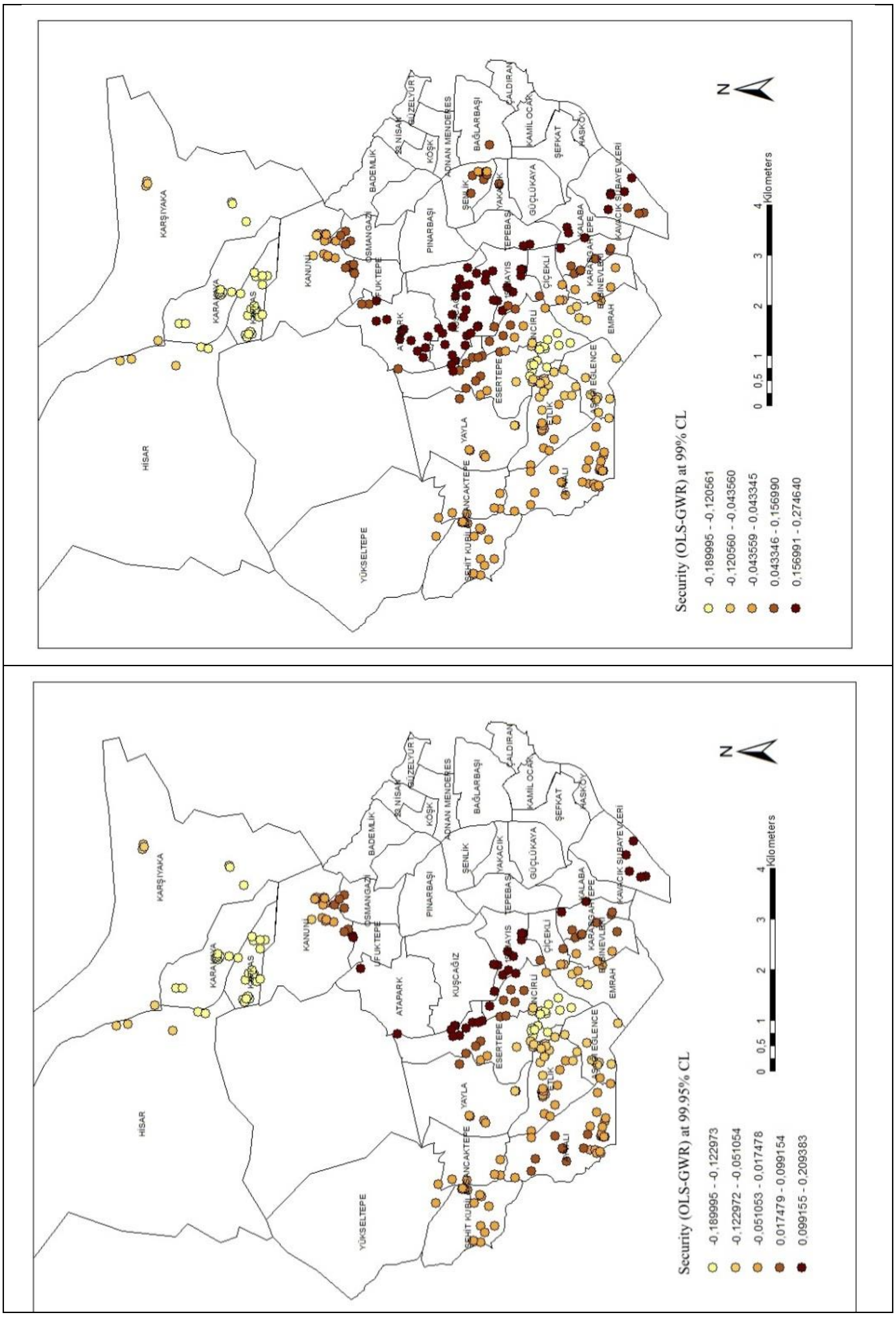


Figure 102 (continued): OLS-GWR Comparison Maps for Security Coefficients for Keçiören

Figure 103: OLS-GWR Comparison Maps for Area Coefficients for Keçiören

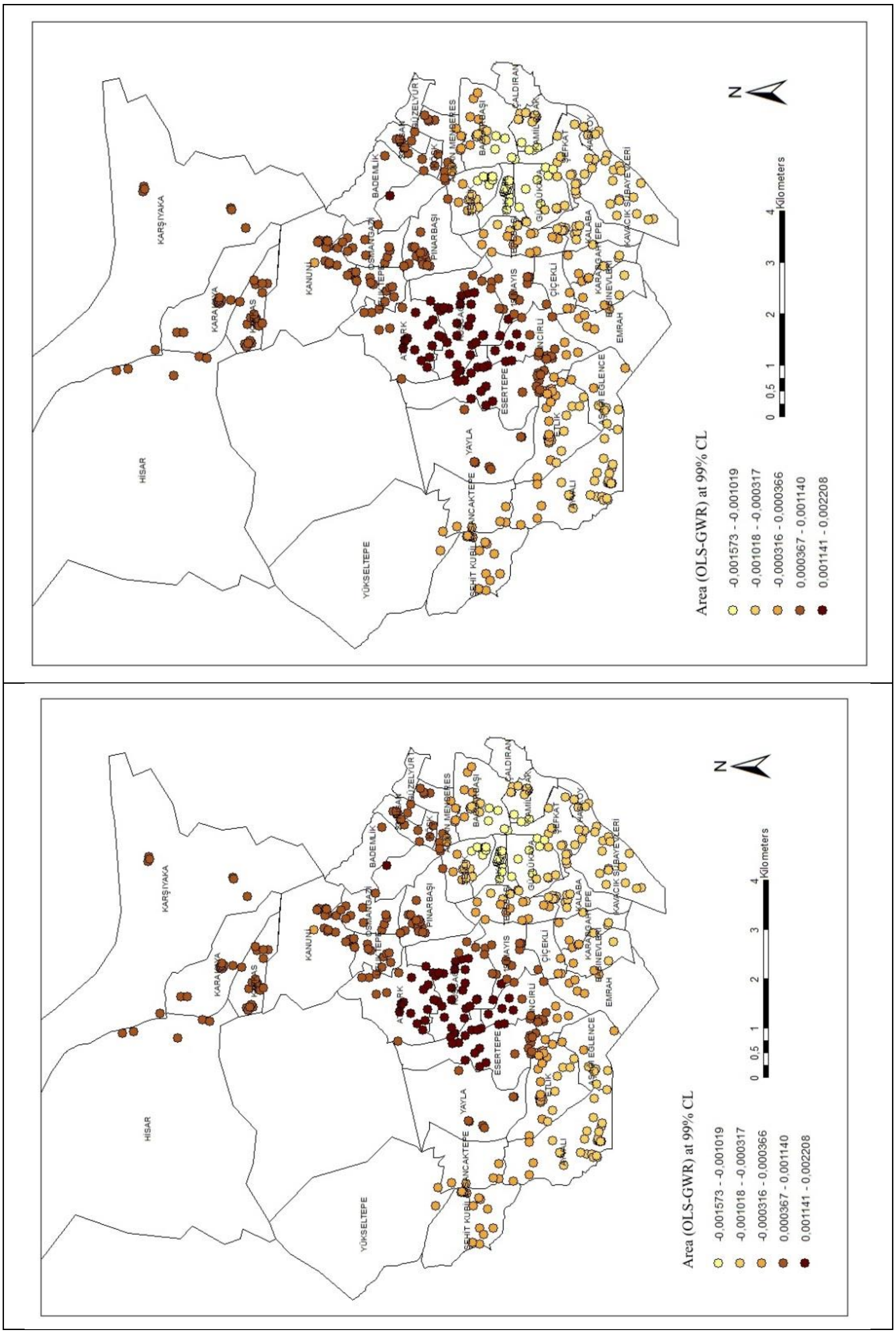


Figure 103 (continued): OLS-GWR Comparison Maps for Area Coefficients for Keçiören

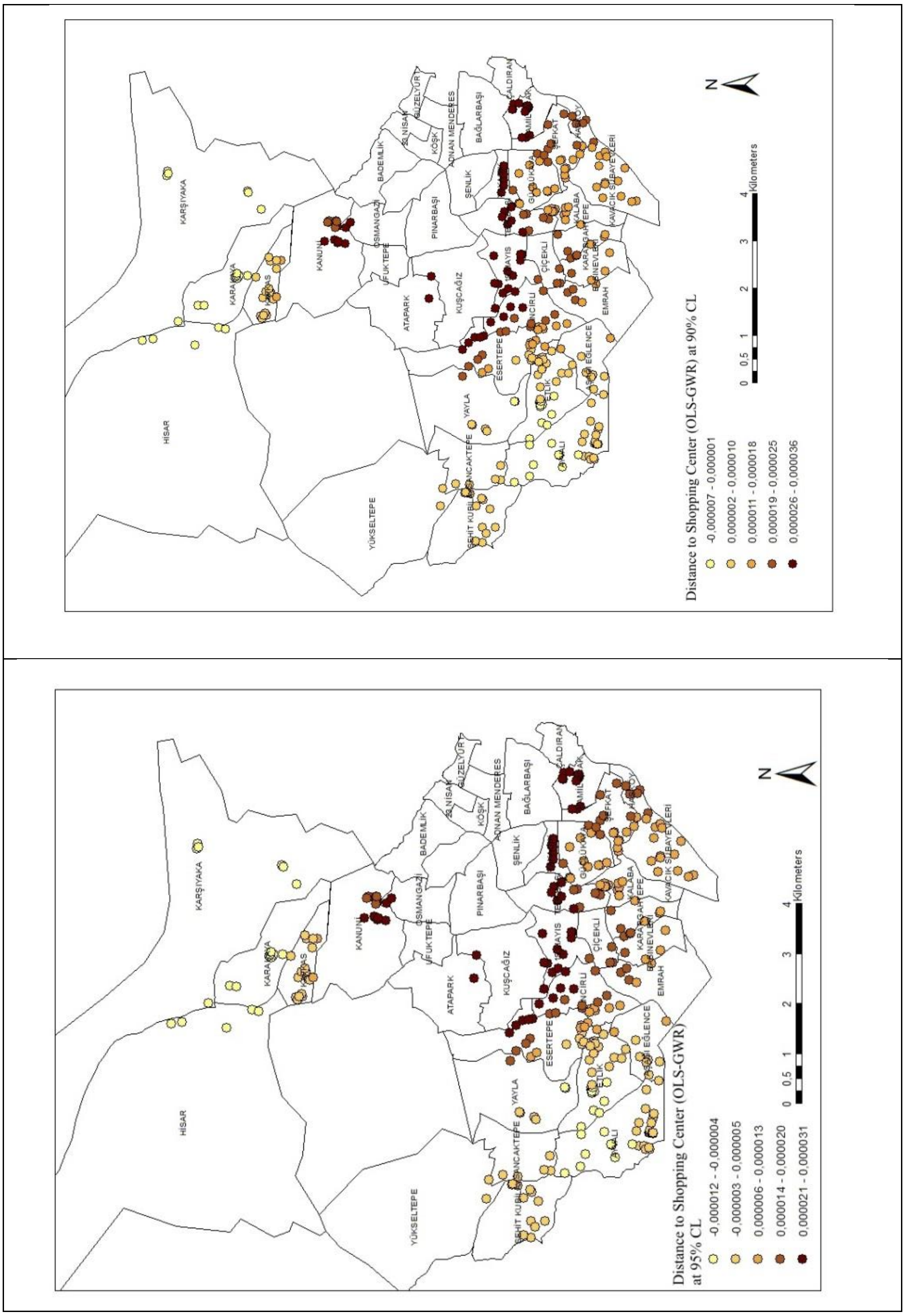


Figure 104: OLS-GWR Comparison Maps for Distance to Shopping Centers Coefficients for Keçiören

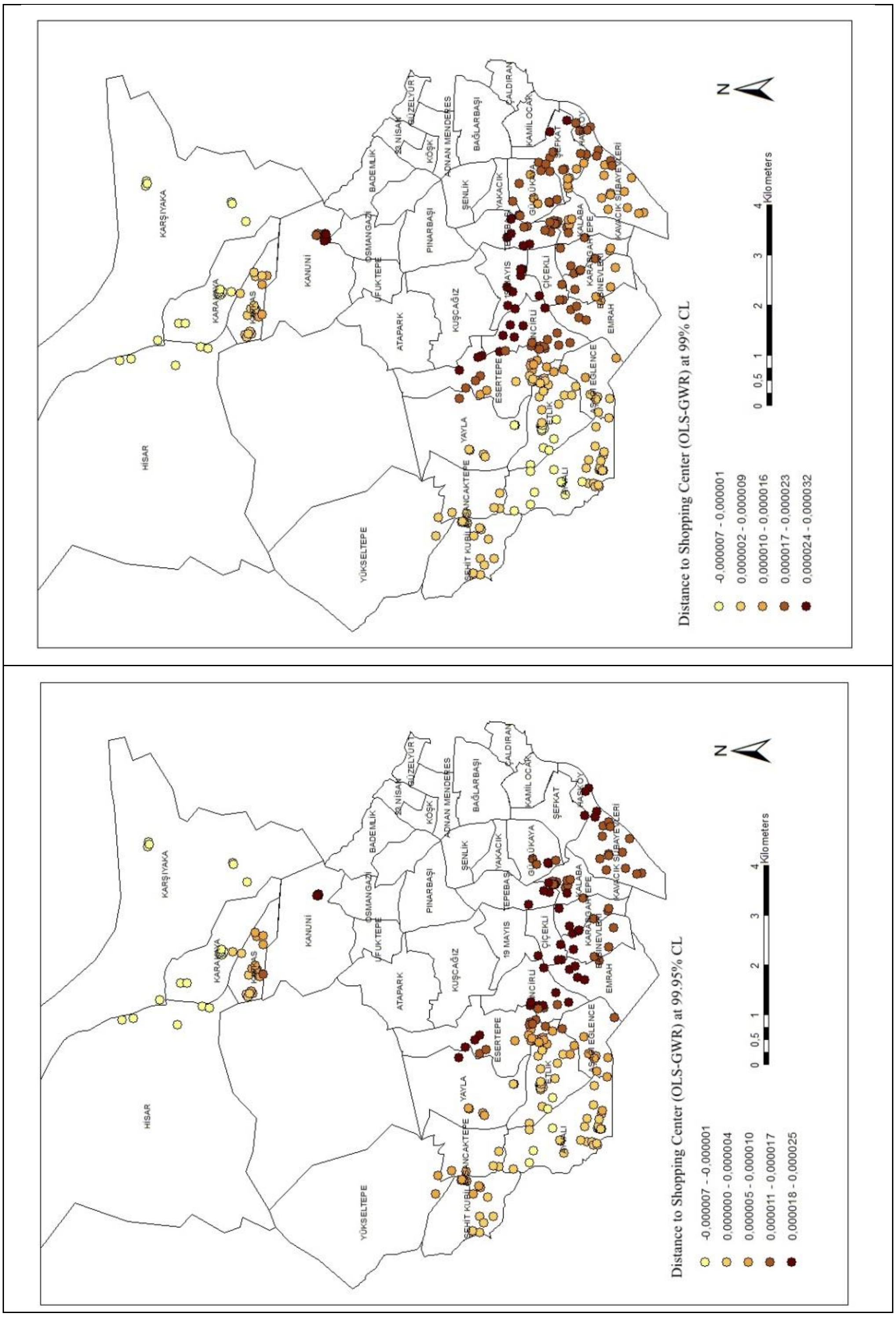


Figure 104 (continued): OLS-GWR Comparison Maps for Distance to Shopping Centers Coefficients for Keçiören

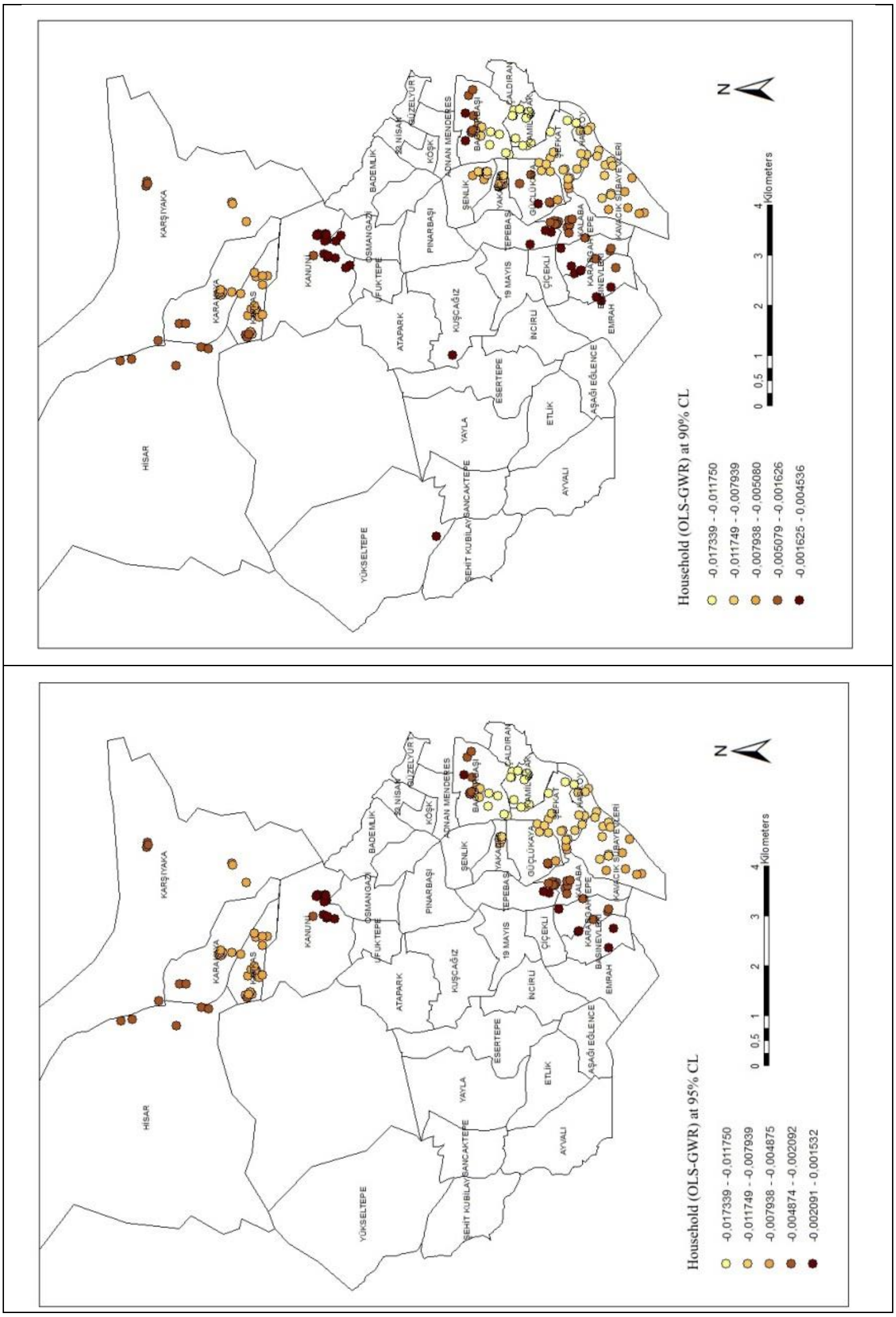


Figure 105: OLS-GWR Comparison Maps for Household Size Center Coefficients

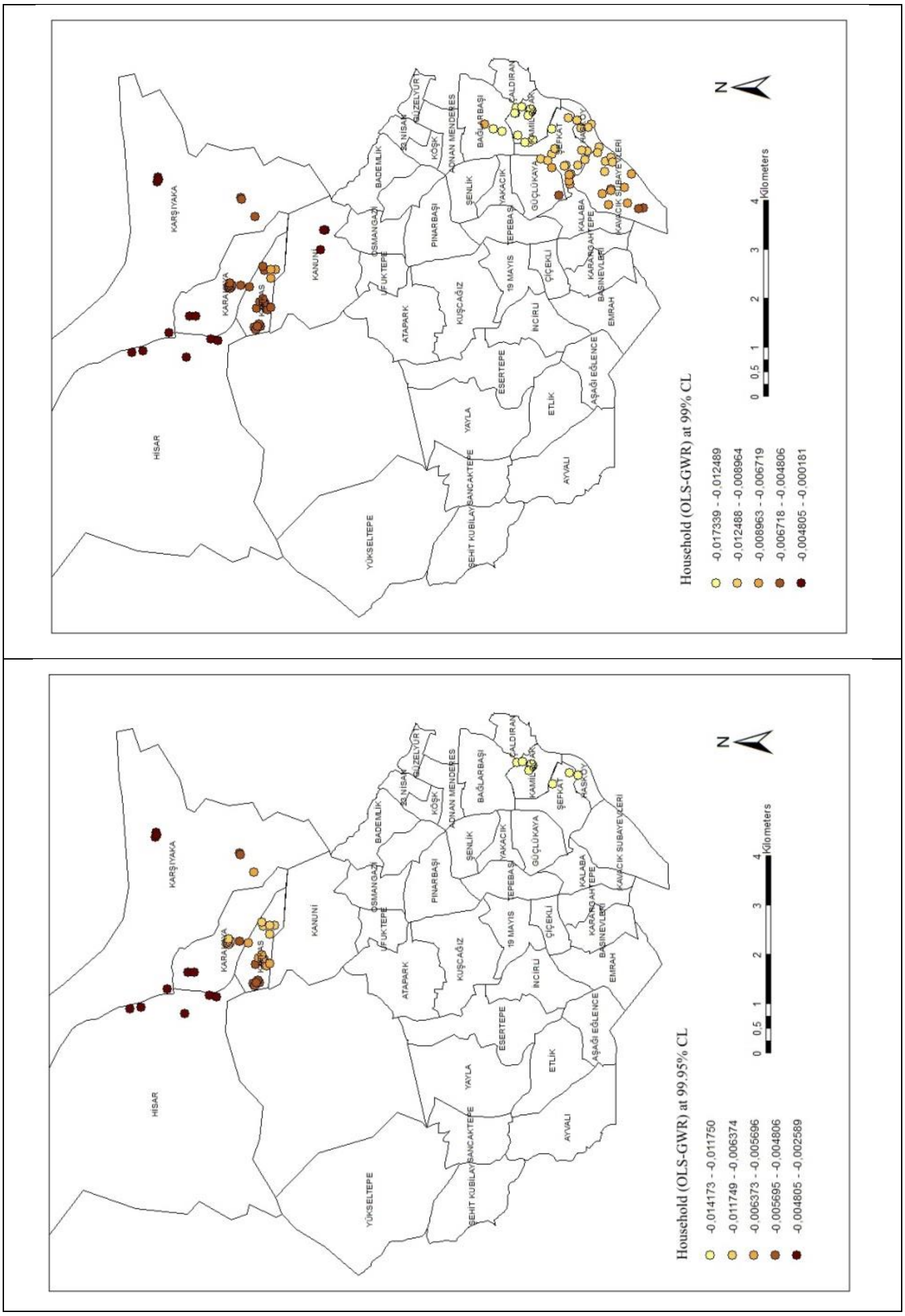


Figure 105 (continued): OLS-GWR Comparison Maps for Household Size Center Coefficients

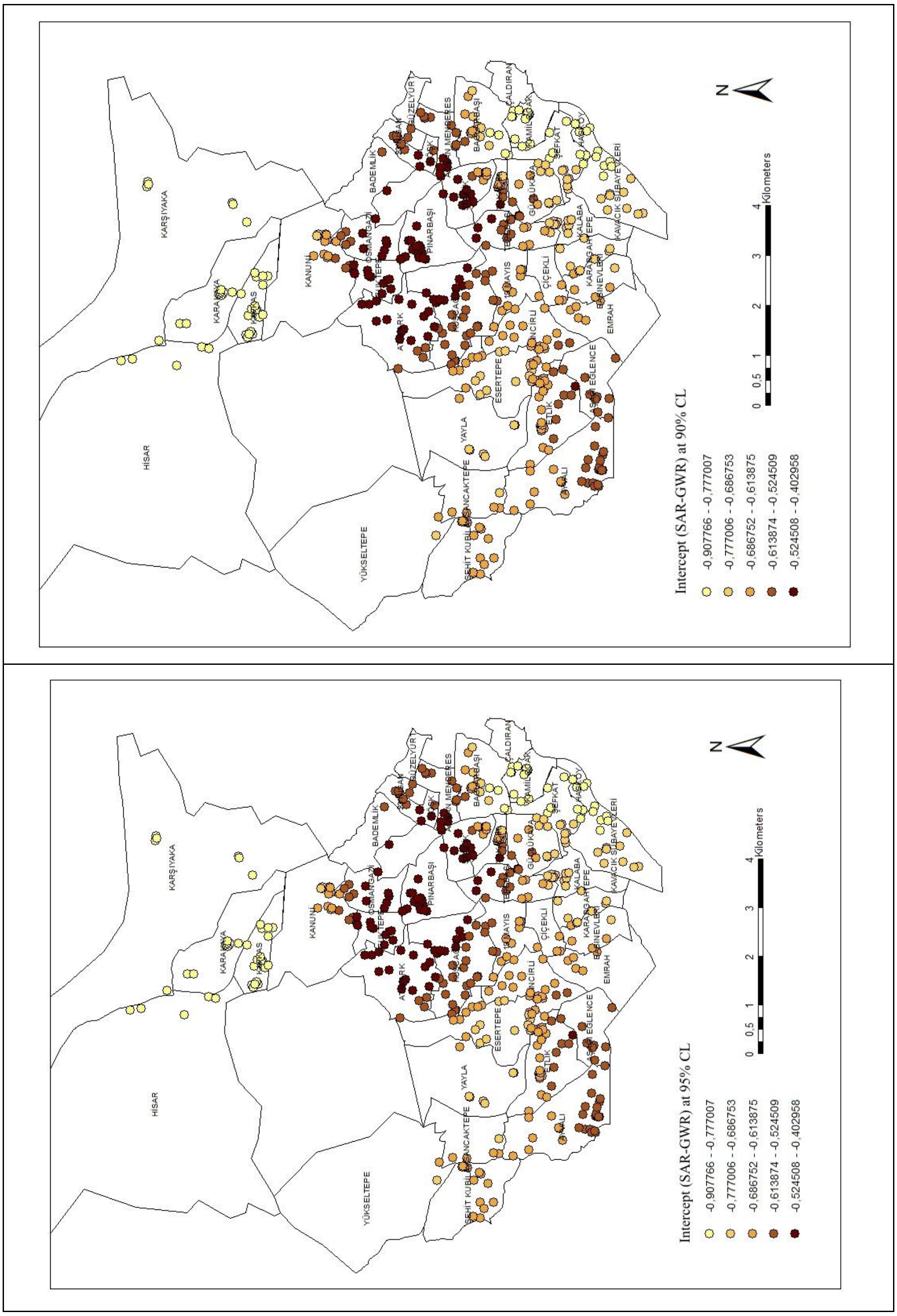


Figure 106: SR-GWR Comparison Maps for Intercept Coefficients for Keçiören

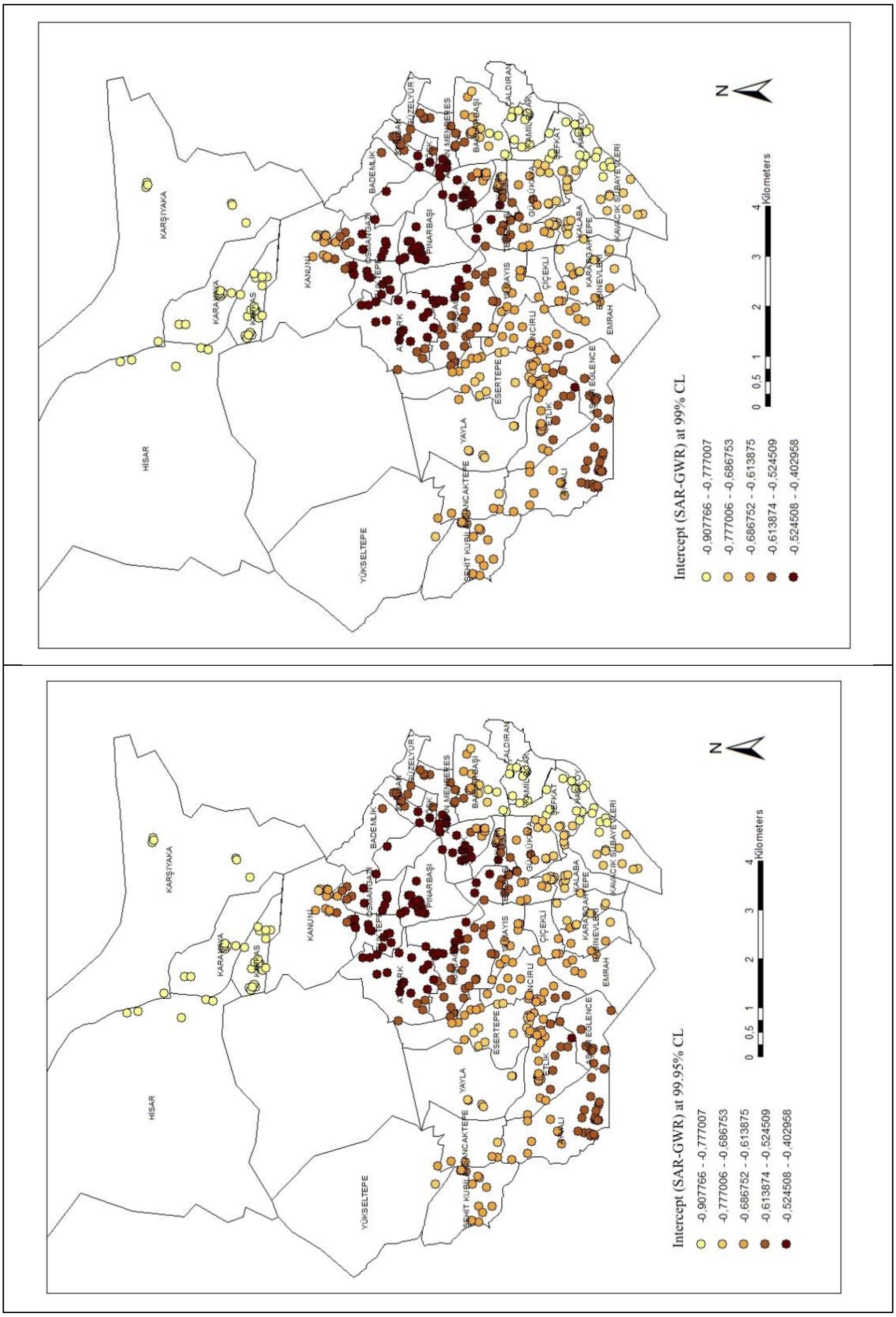


Figure 106 (continued): SR-GWR Comparison Maps for Intercept Coefficients for Keçiören

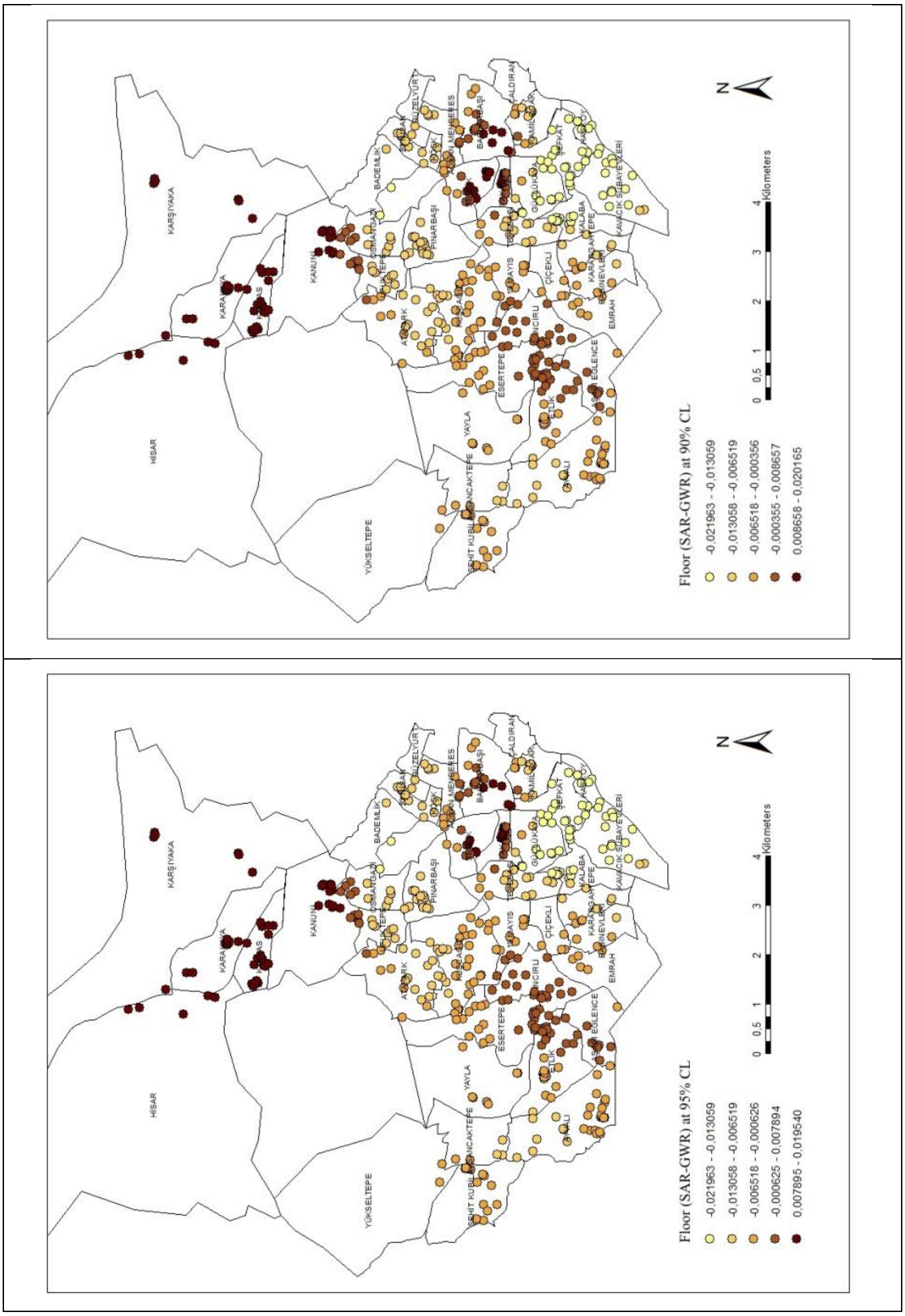


Figure 107: SR-GWR Comparison Maps for Floor Coefficients for Keçiören

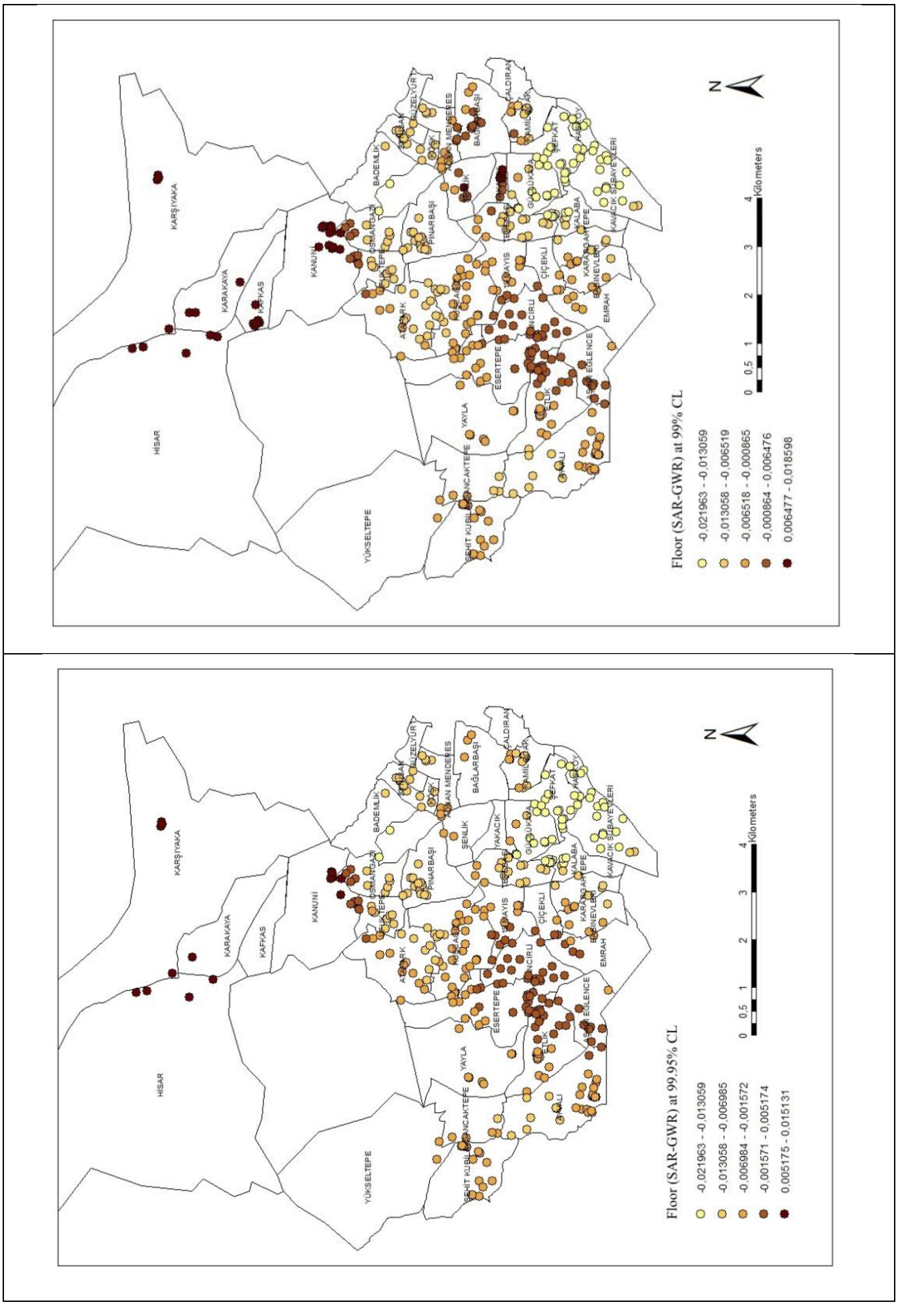


Figure 107 (continued): SR-GWR Comparison Maps for Floor Coefficients for Keçiören

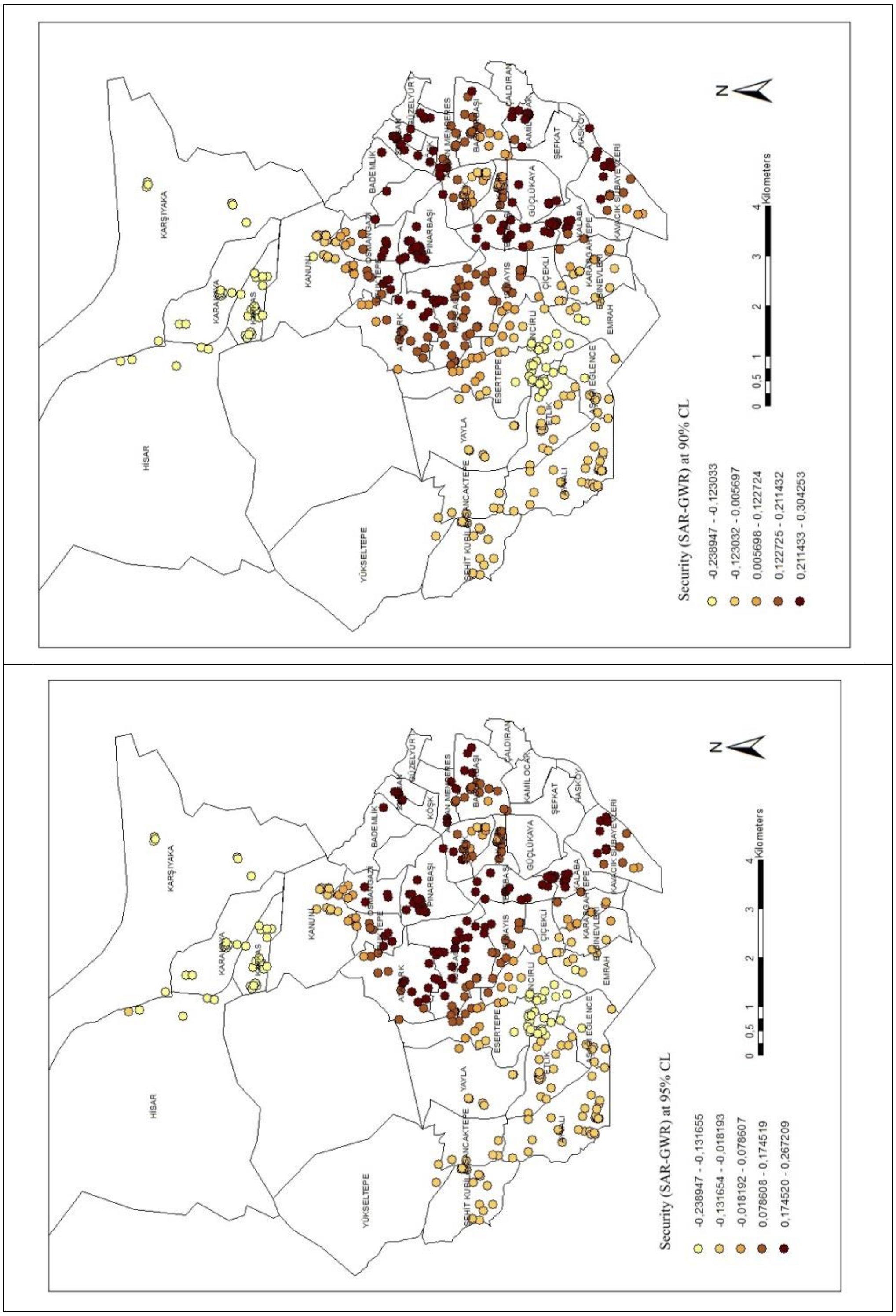


Figure 108: SR-GWR Comparison Maps for Security Coefficients for Keçiören

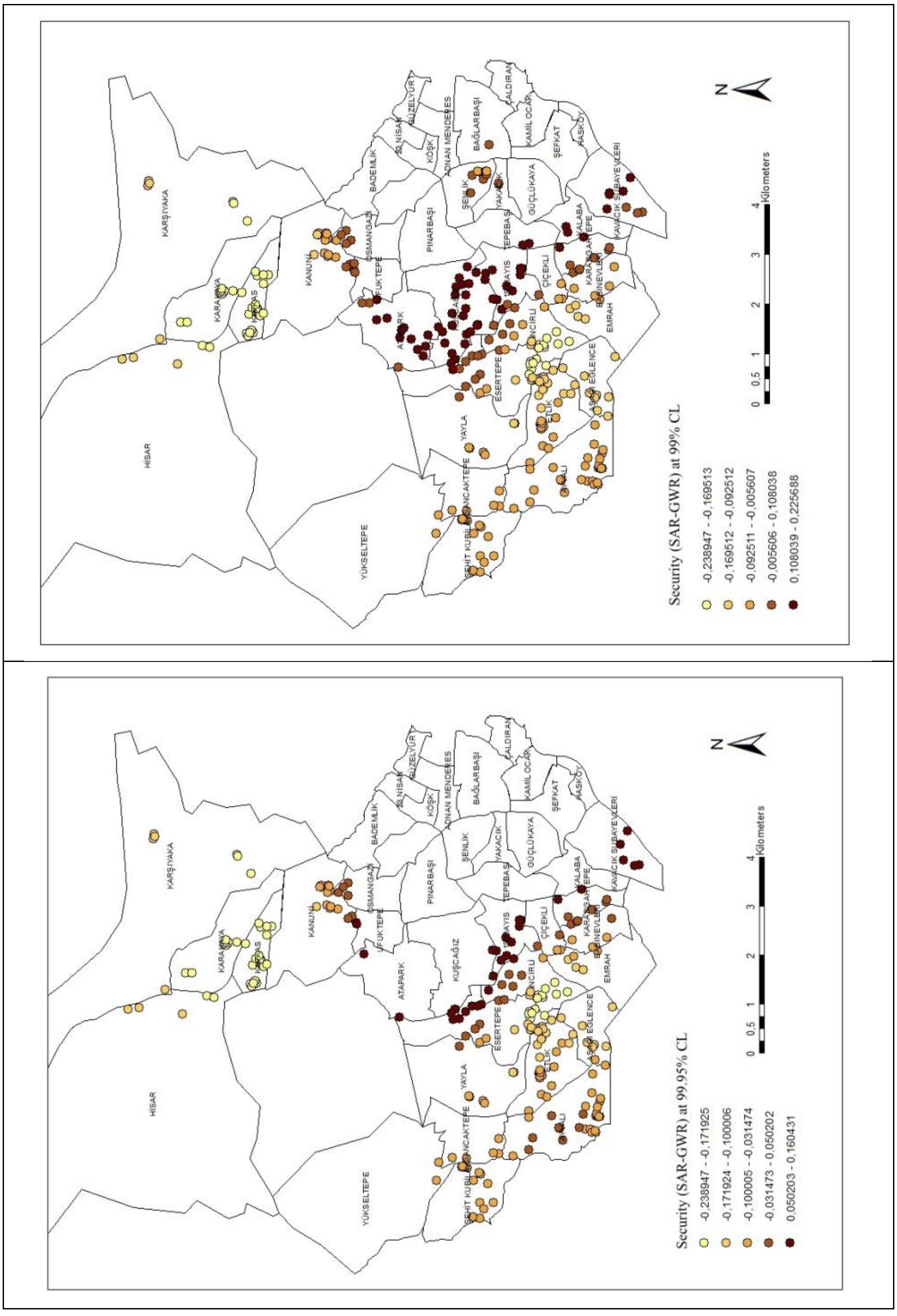


Figure 108 (continued): SR-GWR Comparison Maps for Security Coefficients for Keçiören

Figure 109: SR-GWR Comparison Maps for Area Coefficients for Keçiören

Figure 109 (continued): SR-GWR Comparison Maps for Area Coefficients for Keçiören

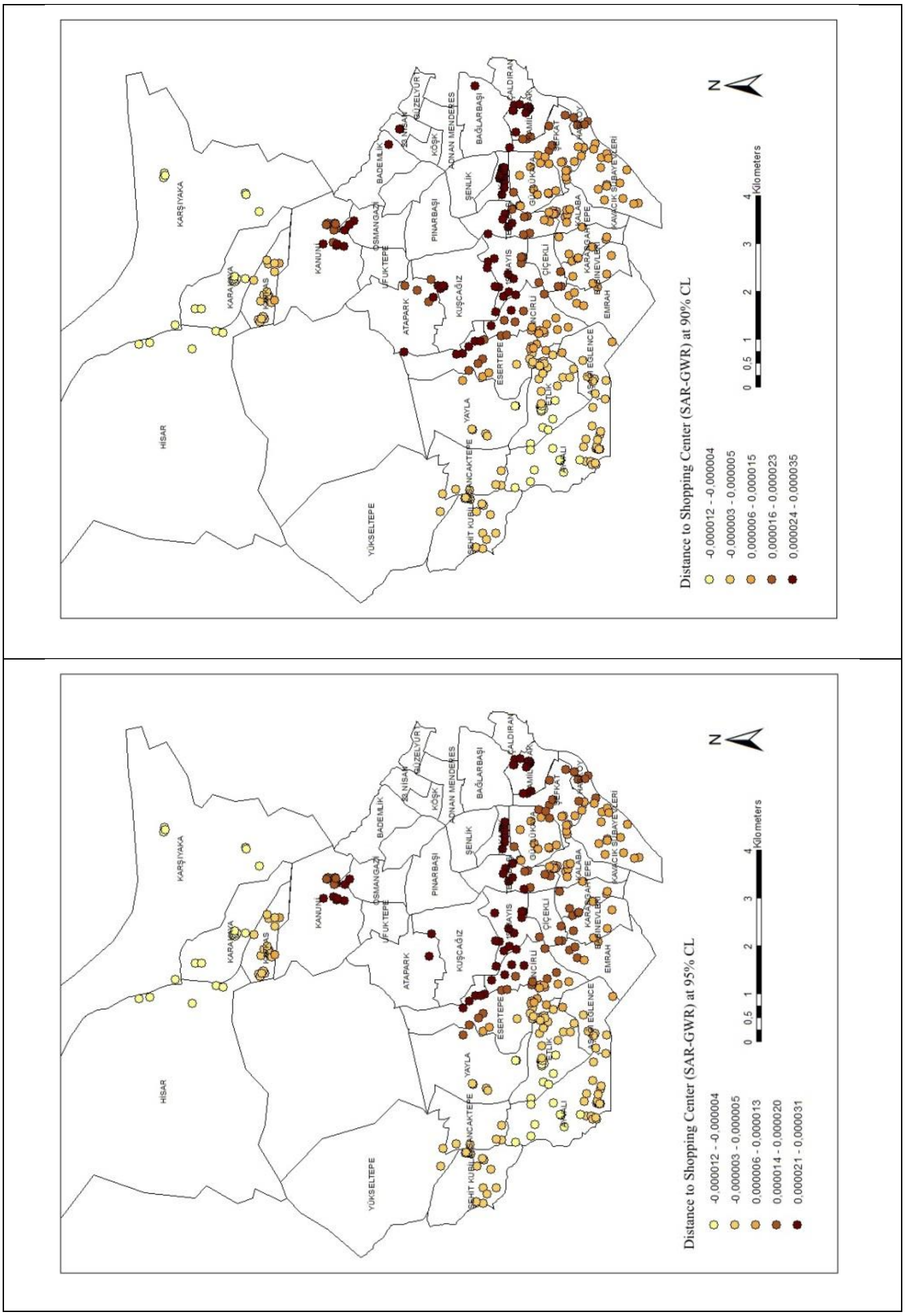


Figure 110: SR-GWR Comparison Maps for Distance to Shopping Centers Coefficients for Keçiören

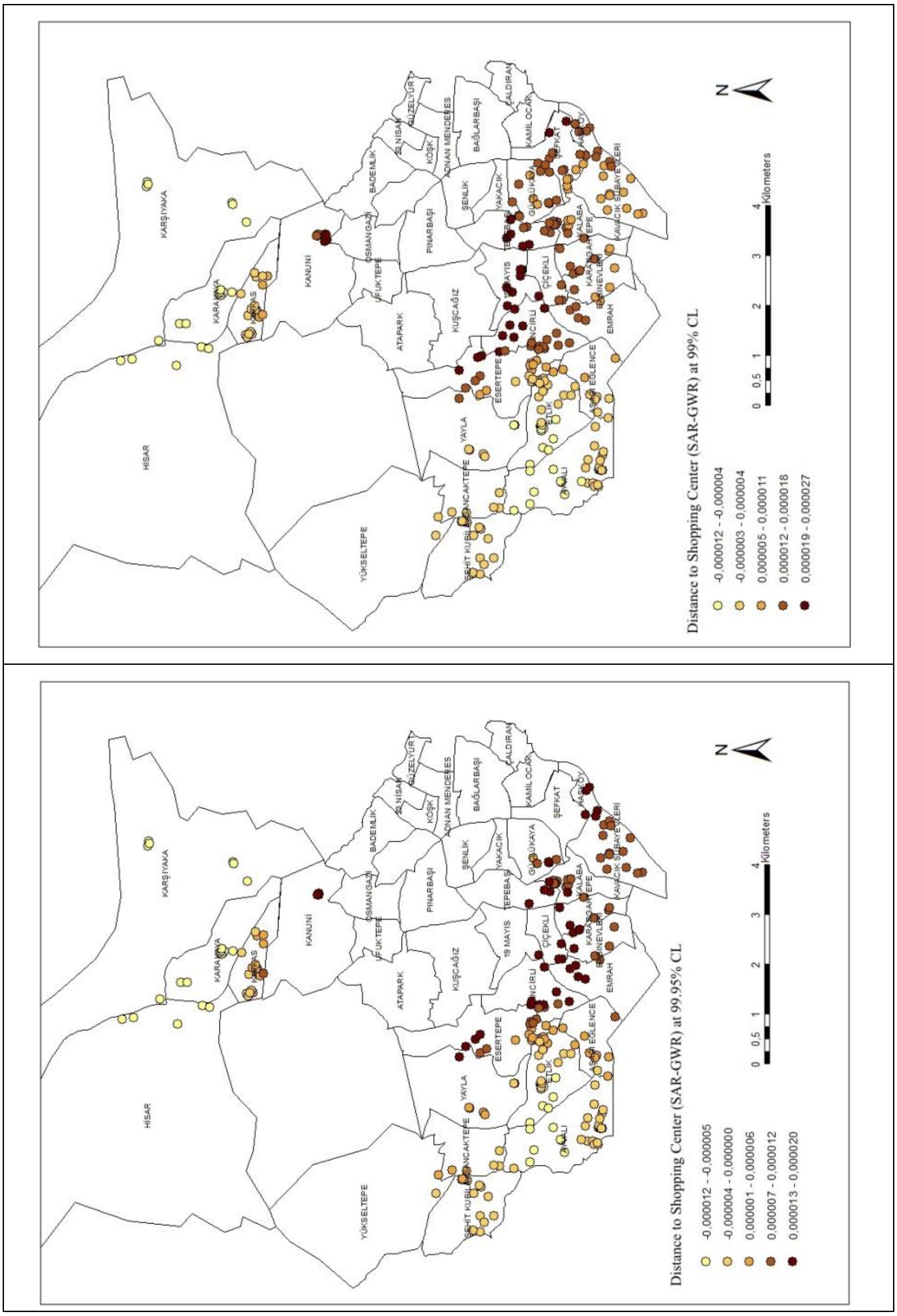


Figure 110 (continued): SR-GWR Comparison Maps for Distance to Shopping Centers Coefficients for Keçiören

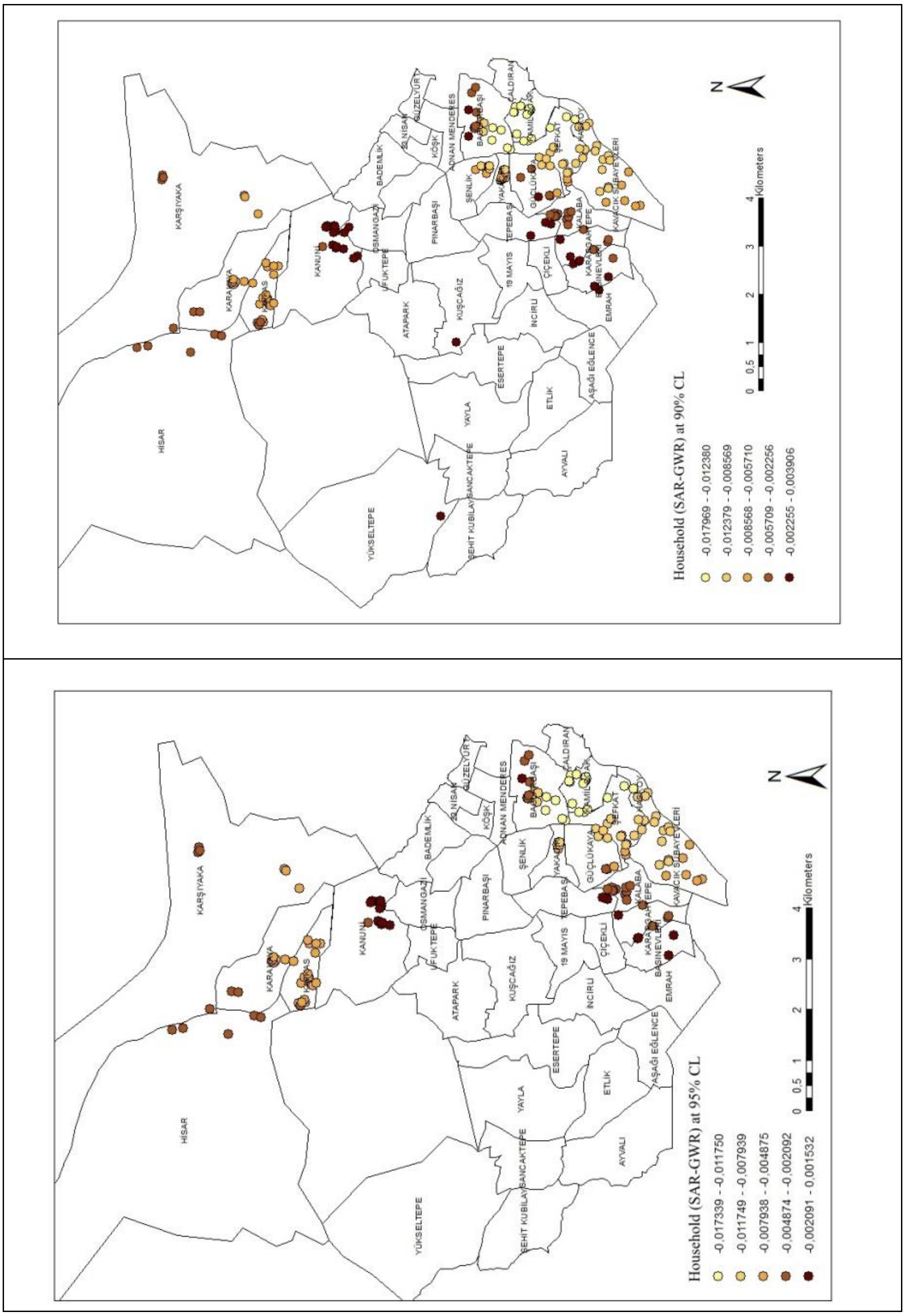


Figure 111: SR-GWR Comparison Maps for Household Size Coefficients for Keçiören

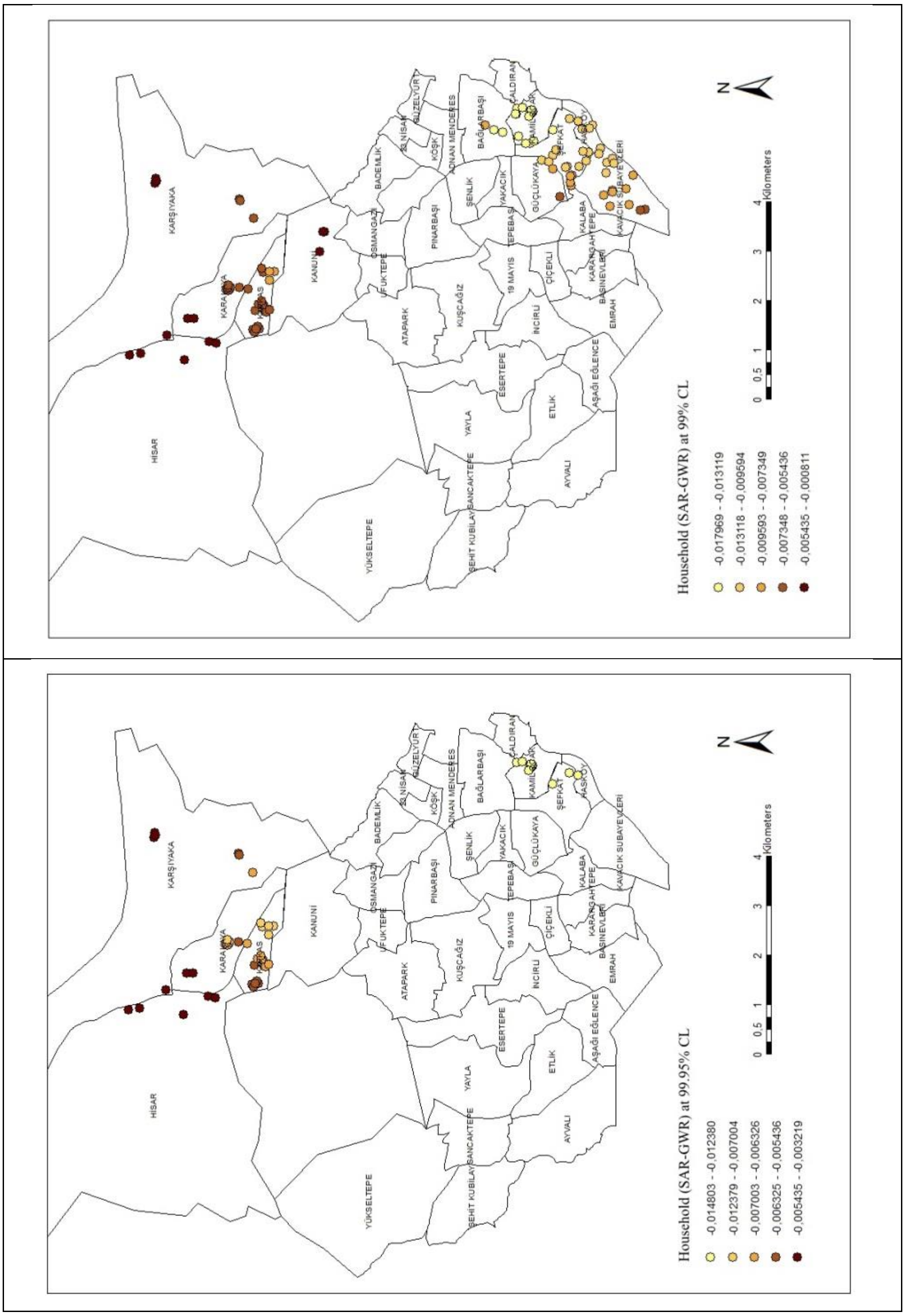


Figure 111 (continued): SR-GWR Comparison Maps for Household Size Coefficients for Keçiören

Figure 112: Distribution of Observed and Predicted LnValue Variables for Keçiören

Figure 113: GWR Map for Distribution of Residuals and Standard Residuals for Keçiören

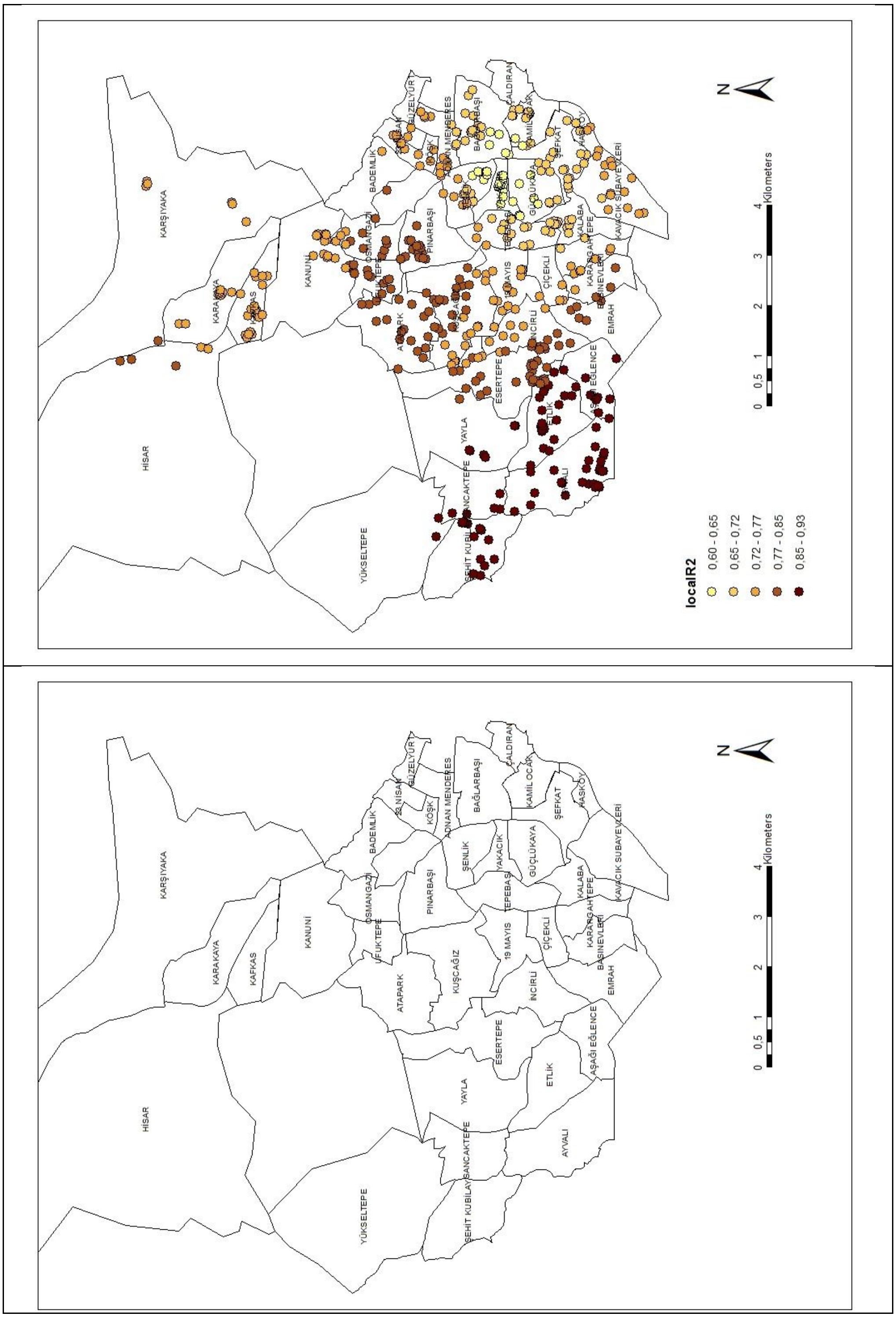


Figure 114: Local R^2 and Neighborhoods for Keçiören

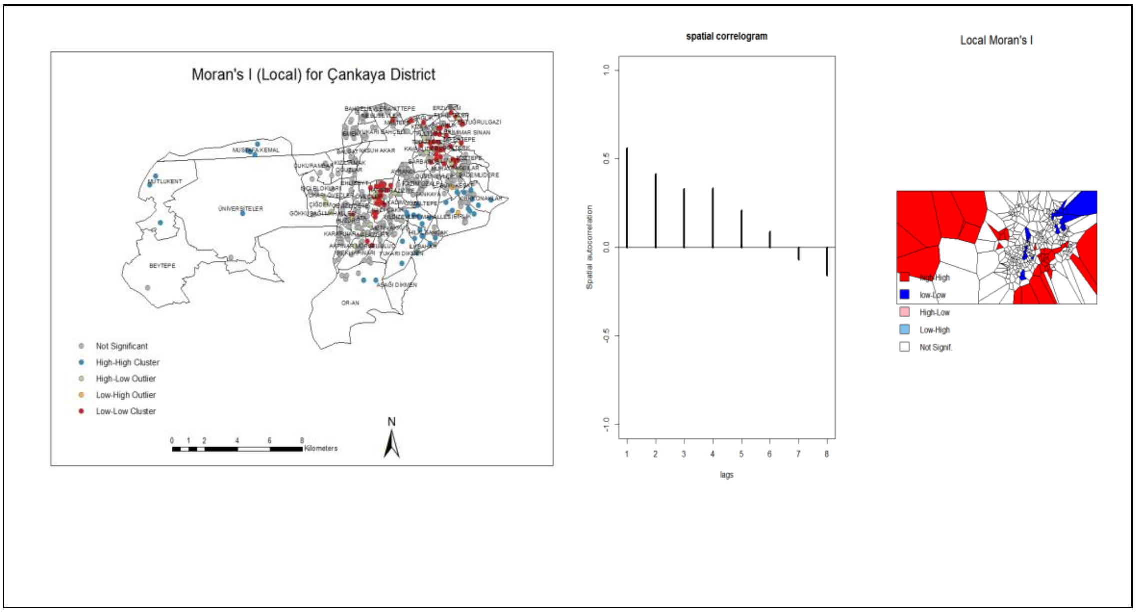


Figure 115: Spatial Correlogram (Local Moran's I) for Çankaya

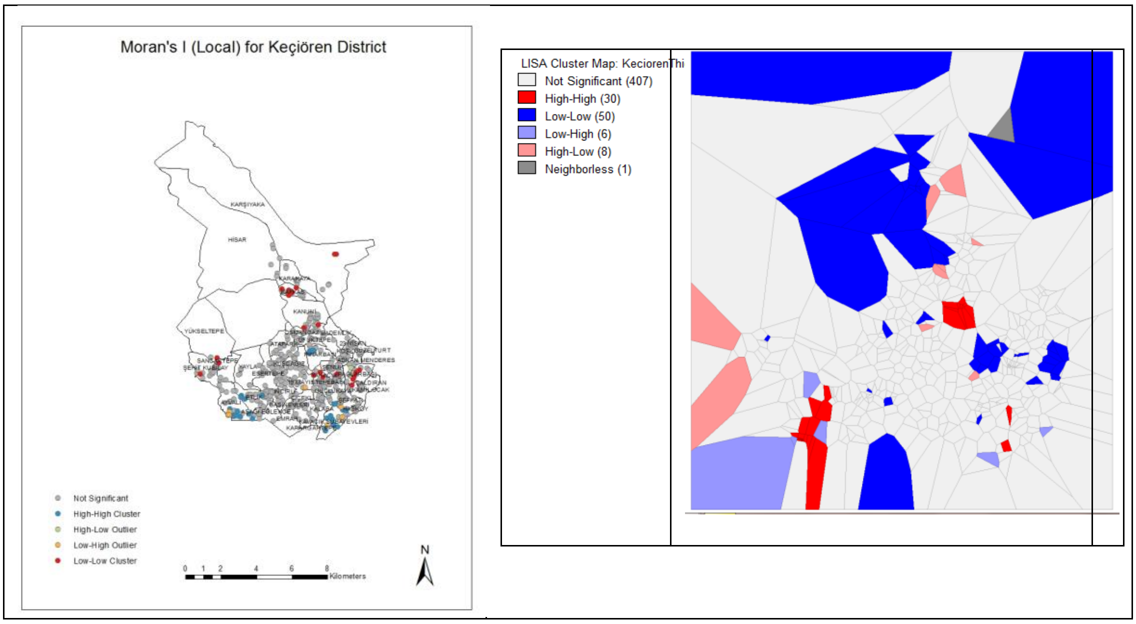


Figure 116: Spatial Correlogram (Local Moran's I) for Keçiören

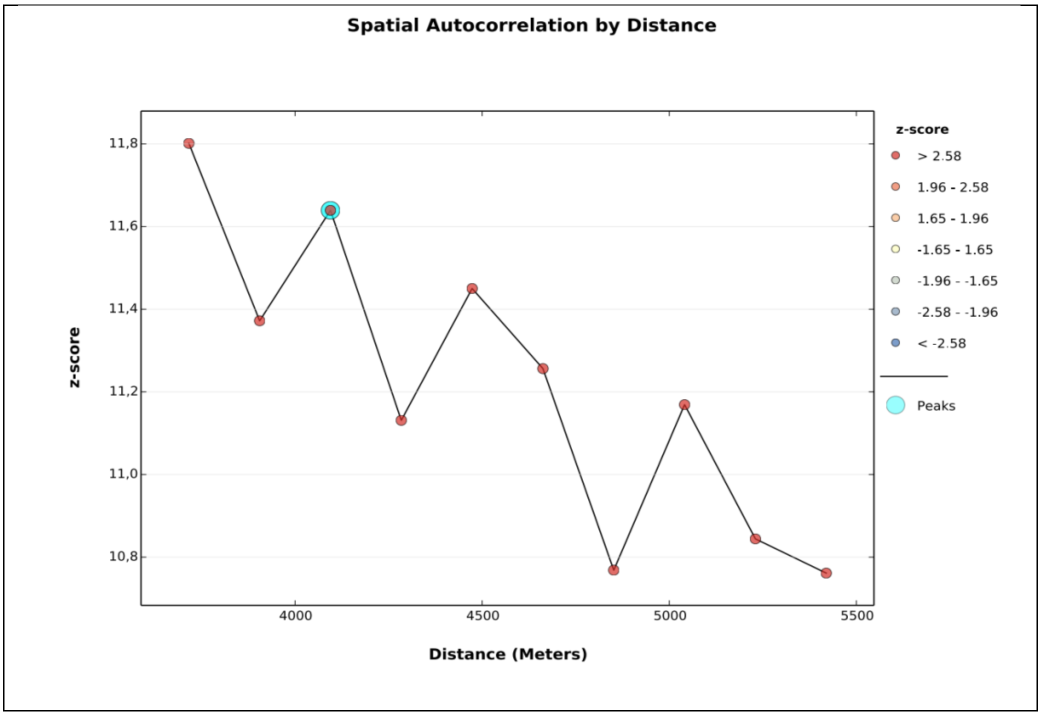


Figure 117: Spatial Autocorrelation by Distance for Çankaya

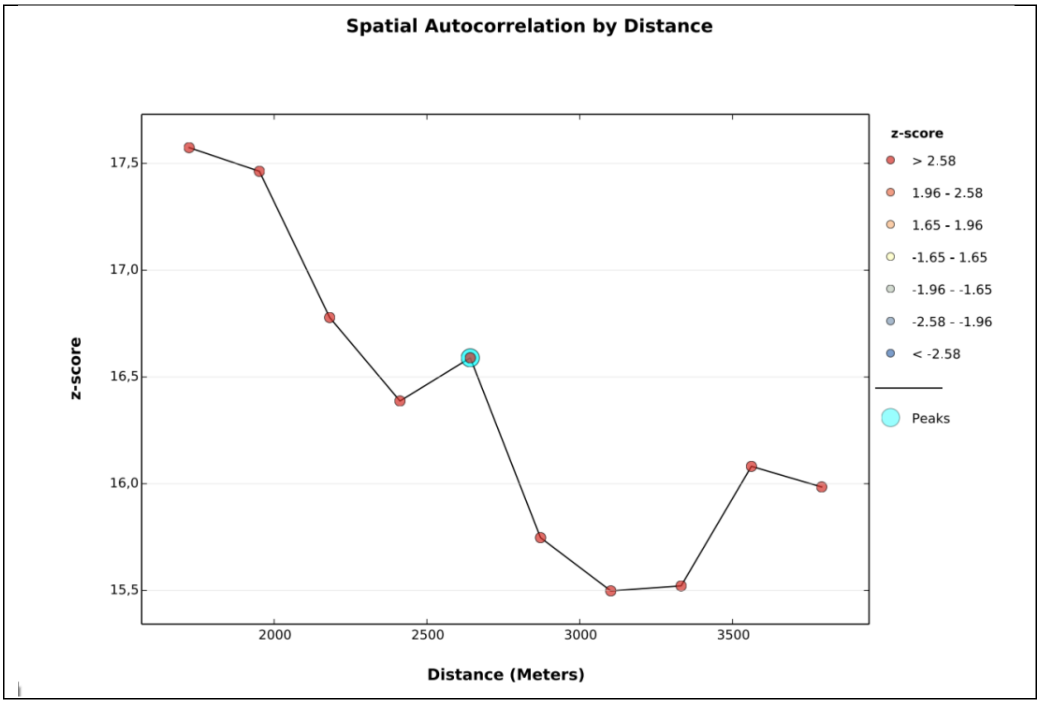


Figure 118: Spatial Autocorrelation by Distance for Keçiören

APPENDIX C

TABLES

Table 38: Correlation matrix for Çankaya

Correlation Matrix for Çankaya

	v1	v3	v4	v20	v15	v17	v18	v22	v30	v31	v33	v32	v34	v35	v36	v38	v39	v45	v44	v41	v42	v43
v1	1																					
v3	.424**	1																				
v4	-.347**	-.262**	1																			
v9	.427**	.605**	-,070																			
v20	.756**	.185**	-.405**	1																		
v16	.209**	-,066	-,089	.302**																		
v15	.571**	,075	-.283**	.789**	1																	
v17	.196**	,005	-,066	.235**	.217**	1																
v18	.601**	.197**	-.411**	.655**	.625**	.151**	1															
v22	.173**	,034	-.408**	.236**	.211**	.126**	.302**	1														
v30	.179**	.097*	-.183**	.190**	,063	,045	.163**	,078	1													
v31	.155**	.195**	-.545**	.289**	.191**	,058	.269**	.437**	.107*	1												
v33	-.262**	-.150**	-.178**	-,060	-,017	-,013	-.130**	.234**	.328**	,046	1											
v32	.239**	.191**	-.302**	.254**	.126**	-,001	.141**	.314**	.556**	.267**	.143**	1										
v34	.134**	,062	-.110*	.187**	,046	,011	,017	,054	.740**	.109*	.285**	.474**	1									
v35	.140**	.151**	-.270**	.248**	.124**	-,020	,084	,090	.634**	.204**	.358**	.531**	.677**	1								
v36	.329**	.219**	-.475**	.342**	.236**	,019	.298**	.499**	.434**	.695**	.142**	.552**	.340**	.440**	1							
v38	,007	-,028	.239**	-.099*	-,059	,004	-,065	-,035	-.119*	-.391**	-.124**	-.230**	-,089	-,066	-.303**	1						
v39	-.097*	-.092*	.312**	-.171**	-.098*	-,003	-.109*	-,083	-.135**	-.422**	-.148**	-.299**	-,064	-.100*	-.380**	.801**	1					
v45	.128**	-,061	-,023	,063	,078	-,005	,054	.298**	-,040	-.108*	-,074	,026	,042	-,003	-,008	.510**	.460**	1				
v44	-,039	.142**	-.208**	,036	,014	-,065	,057	,012	-,045	.249**	.144**	-,064	-,001	.135**	,063	-.153**	-.203**	,012	1			
v41	.120*	-,037	.528**	-,055	-,049	-,006	-,040	-.290**	-.139**	-.493**	-.310**	-.216**	-,056	-.138**	-.344**	.375**	.402**	.304**	-,021	1		
v42	-.359**	-.115*	-.156**	-.210**	-.102*	-,017	-.207**	,020	-.149**	.096*	.320**	-.176**	-.121**	-.123**	-.167**	-.184**	-.131**	-.291**	.110*	-.648**	1	
v43	.279**	.140**	-.209**	.191**	.122**	-,031	.231**	.168**	.438**	.214**	-.094*	.440**	.121**	.187**	.559**	-.155**	-.189**	,078	-.195**	-,087	-.326**	1

Table 39: Correlation matrix for Keçiören

Correlation Matrix for Keçiören

	v1	v4	v3	v9	v40	v15	v18	v19	v20
v1	1								
v4	-.150**	1							
v3	.574**	-.116**	1						
v9	.631**	-,064	.569**	1					
v40	.105*	-,009	,083	,025	1				
v15	.525**	-.127**	.147**	.381**	,051	1			
v18	.584**	-.163**	.282**	.377**	,053	.491**	1		
v19	.472**	-,069	.499**	.444**	,038	.428**	.212**	1	
v20	.712**	-.204**	.259**	.428**	,075	.721**	.545**	.455**	1
v37	-,067	-,076	,007	,012	-,064	-,044	-,009	,025	-,027
v30	-.116**	-.117**	,073	,056	-.136**	-,040	-,056	,041	-,012
v33	-.313**	-.347**	-.170**	-,021	-,074	-,082	-,064	-.091*	-.111*
v32	,008	-.186**	,052	.090*	-.104*	,021	-,047	,068	,031
v34	-,041	-.125**	-,013	,071	-.163**	,014	,017	,036	,023
v35	-,078	-.110*	-,026	.091*	-.182**	,008	-,025	,011	,020
v36	-.263**	-.315**	-,083	-,034	-,013	-,031	-,071	-,003	-,072
v38	-.159**	.177**	-,056	-.098*	,048	-.160**	-.098*	-.111*	-.151**
v39	-.135**	.226**	,043	-.089*	,060	-.120**	-.101*	-,020	-.102*
v29	-.208**	-.260**	-.090*	-,048	-,072	-,061	-,019	-,034	-,028
v28	-.183**	-.330**	-,020	,071	-.090*	-,021	-,007	,039	-,072
v26	-.123**	-.314**	.159**	,037	-,003	-,023	,020	.163**	-,026
v27	,046	-.177**	,047	-,005	,004	,029	,083	,069	,061
v23	-,027	-.248**	.091*	-,042	,041	-,041	.099*	,024	-,001
v22	-,044	-.161**	,050	,054	-.147**	,009	,001	.135**	-,031
v45	,038	,029	,043	,061	,027	-,084	,069	-,043	,000
v44	-.134**	.253**	,066	,069	,035	-,075	-.148**	,066	-.105*
v41	.179**	.339**	,000	-,063	.115**	-,021	-,020	-,021	,046
v42	-.225**	-.223**	,010	,027	,008	-,028	-,076	-,026	-.156**
v43	-.153**	-.180**	-,026	,067	-.144**	,010	-,005	,050	-,012

	v37	v30	v33	v32	v34	v35	v36	v38	v39	v29	v28
v1											
v4											
v3											
v9											
v40											
v15											
v18											
v19											
v20											
v37	1										
v30	.581**	1									
v33	.249**	.316**	1								
v32	.477**	.445**	.276**	1							
v34	.721**	.679**	.307**	.707**	1						
v35	.610**	.749**	.292**	.522**	.733**	1					
v36	.140**	.294**	.571**	.345**	.309**	.338**	1				
v38	-,051	-,063	,050	-.220**	-.167**	-,030	-.126**	1			
v39	-,061	-,060	,036	-.243**	-.190**	-.090*	-.167**	.783**	1		
v29	.184**	.288**	.478**	.328**	.331**	.369**	.791**	,010	-,016	1	
v28	.169**	.279**	.677**	.361**	.305**	.284**	.625**	-.318**	-.401**	.515**	1
v26	,044	.155**	.586**	.182**	.086*	,085	.349**	-,006	-,022	.134**	.621**
v27	,076	,063	.193**	,013	.102*	.128**	.305**	.307**	.311**	.569**	,071
v23	.141**	.136**	.227**	,036	.157**	.150**	.413**	.262**	.228**	.601**	.210**
v22	.308**	.272**	.258**	.456**	.398**	.271**	.326**	-.412**	-.492**	.318**	.664**
v45	-,077	-.096*	,063	-.201**	-.151**	-.087*	-.278**	.620**	.631**	-,076	-.250**
v44	,032	.181**	-,012	,043	,037	.138**	-,005	.278**	.370**	-,005	-.115**
v41	-.175**	-.349**	-.519**	-.387**	-.386**	-.385**	-.722**	.282**	.372**	-.647**	-.800**
v42	,030	.091*	.439**	.188**	.124**	,081	.462**	-,013	-,028	.292**	.561**
v43	.223**	.444**	.373**	.408**	.424**	.472**	.600**	-.291**	-.360**	.626**	.636**

	v26	v27	v23	v22	v45	v41		v43
v1								
v4								
v3								
v9								
v40								
v15								
v18								
v19								
v20								
v37								
v30								
v33								
v32								
v34								
v35								
v36								
v38								
v39								
v29								
v28								
v26								
v27		1						
v23		.618**						
v22		,068		1				
v45		.258**		-.395**				
v44		-,026		-,026				
v41		-.244**		-.621**		1		
v42		.198**		.326**		-.642**	1	
v43		.122**		.598**		-.798**	.151**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 40: Correlations between Dependent and Independent Variables

Variables	LnValue	Property Type	Security	Area	Distance to Subway	Distance to Shopping Centers
LnValue	1.000	.296	.437	.808	.179	-.270
PropertyType	.296	1.000	.417	.131	.185	.009
Security	.437	.417	1.000	.230	.121	-.118
Area	.808	.131	.230	1.000	.295	-.109
Distance to Subway	.179	.185	.121	.295	1.000	.057
Distance to Shopping Center	-.270	.009	-.118	-.109	.057	1.000

Table 41: Correlations of Çankaya Explanatory Variables

Variables	LnValue	Floor	Security	Area	Distance Shopping Centers	Household size
LnValue	1.000	.630	.590	.765	-.257	-.059
Floor	.630	1.000	.669	.435	.002	.099
Security	.590	.669	1.000	.325	-.098	.029
Area	.765	.435	.325	1.000	-.039	-.049
Distance to Shopping Centers	-.257	.002	-.098	-.039	1.000	-.038
Household size	-.059	.099	.029	-.049	-.038	1.000

Table 42: Model Summaries

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
Çankaya	.873	.762	.759	.236	1.630
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**Durbin-Watson**
Keçiören	.884	.782	.780	.203	1.758

Table 43: OLS ANOVA Result for Çankaya

Çankaya Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	72.481	5	14.496	260.932	.000
Residual	22.667	408	.056		
Total	95.147	413			

Table 44: OLS ANOVA Result for Keçiören

Keçiören Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	73.247	5	14.649	355.944	.000
Residual	20.414	496	.041		
Total	93.661	501			

Table 45: Coefficients for OLS for Çankaya

Çankaya Model	Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
(Constant)	11.161	.051		218.157	.000		
Property Type	.205	.042	.130	4.843	.000	.805	1.242
Security	.460	.063	.201	7.363	.000	.784	1.275
Area	.009	.000	.751	28.893	.000	.864	1.157
Distance to Subway	-1.198E-5	.000	-.082	-3.186	.002	.884	1.132
Distance to Shopping Centers	-8.623E-5	.000	-.161	-6.551	.000	.966	1.035

Table 46: Coefficients for OLS for Keçiören

Keçiören Model	Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
(Constant)	10.884	.061		178.985	.000		
Floor	.040	.005	.235	7.814	.000	.486	2.060
Security	.487	.062	.226	7.908	.000	.540	1.851
Area	.009	.000	.577	24.611	.000	.799	1.252
Distance to Shopping Centers	-4.946E-5	.000	-.216	-10.187	.000	.978	1.022
Household Size	-.008	.003	-.069	-3.245	.001	.975	1.025

Table 47: OLS Moran’s I results for Çankaya and Keçiören

Model	Çankaya	Keçiören
Moran's Index:	0,36	0,37
Expected Index:	-0,002421	-0,001996
Variance:	0,000874	0,000706
z-score:	12,126255	15,055132
p-value:	0,000000	0,000000

Table 48: Global Moran's I Summary by distance for Çankaya

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
3716,13	0,081418	-0,002421	0,00005	11,801566	0,0000
3905,39	0,071809	-0,002421	0,000043	11,371996	0,0000
4094,65	0,069361	-0,002421	0,000038	11,639386	0,0000
4283,91	0,064442	-0,002421	0,000036	11,131127	0,0000
4473,18	0,060938	-0,002421	0,000031	11,450185	0,0000
4662,44	0,053564	-0,002421	0,000025	11,256274	0,0000
4851,7	0,046644	-0,002421	0,000021	10,768442	0,0000
5040,97	0,044837	-0,002421	0,000018	11,168988	0,0000
5230,23	0,039316	-0,002421	0,000015	10,844203	0,0000
5419,49	0,036629	-0,002421	0,000013	10,761329	0,0000

Table 49: Global Moran's I Summary by distance for Keçiören

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
1721,79	0,144697	-0,001996	0,00007	17,573914	0,0000
1951,81	0,122224	-0,001996	0,000051	17,463006	0,0000
2181,83	0,102237	-0,001996	0,000039	16,778398	0,0000
2411,85	0,08602	-0,001996	0,000029	16,38739	0,0000
2641,88	0,075676	-0,001996	0,000022	16,589451	0,0000
2871,9	0,061972	-0,001996	0,000017	15,747419	0,0000
3101,92	0,05417	-0,001996	0,000013	15,498093	0,0000
3331,94	0,048121	-0,001996	0,00001	15,52179	0,0000
3561,97	0,044757	-0,001996	0,000008	16,081462	0,0000
3791,99	0,040038	-0,001996	0,000007	15,984883	0,0000

Table 50: Comparison of the results of three models for Çankaya

Model	OLS	SARlag	GWR
R^2	0,762	0,797	0,8124
Adj R^2	0,759		0,7929
AICC	-15,7787	-72,3873	-55,8791
Schwarz	8,37654	-44,2062	
Log likelihood	13,8893	43,1936	
Moran's I	0,36	0,07	
Intercept	11,161	7,5663	
β1(V2)	0,205	0,1559	
β2(V10)	0,46	0,4088	
β3(V20)	0,009	0,0084	
β4(V31)	-1,20E-02	-1,64E-05	
β5(V33)	-8,62E-02	-5,47E-05	
SARlag (Rho)		0,2996	

Table 51: Comparison of the results of three models for Keçiören

Model	OLS	SARerr	GWR
R^2	0,7820	0,8159	0,7618
Adj R^2	0,78		0,7583
AICC	-170,99	-231,21	-13,5028
Schwarz	-145,68	-205,89	14,4024
Log likelihood	91,49	121,6	-27,78
Moran's I	0,37	-0,02	
Intercept	10,88	10,9186	
β1(V9)	0,04	0,0468	
β2(V10)	0,487	0,3864	
β3(V20)	0,009	0,0083	
β4(V33)	-0,049	-4,27E-05	
β5(V44)	-.008	-0,0093	
SARerr (λ)		0,4679	

APPENDIX D

SCRIPTS for R

D.1. OLS Regression

```
 > summary(model)
Checking Normality
> qqnorm(resid(md)) # A quantile normal plot - good for checking normality
> qqline(resid(md)

> # Assessing Outliers
> outlierTest(md) # Bonferonni p-value for most extreme obs

> NY_nb <- read.gal("C:\\Users\\PC\\Desktop\\GWR\\cankaya\\ShapeFile\\Cankaya
414_Rook.GAL", region.id=NULL, override.id=FALSE)
> summary(NY_nb)

> plot(NY8, border="grey20")
> plot(NY_nb, coordinates(NY8), pch=19, cex=0.6, add=TRUE)

> summary(dsts0)

# Normality of Residuals
# qq plot for studentized resid

> qqPlot(md, main="QQ Plot") #qq plot for studentized resid
> # distribution of studentized residuals
> library(MASS)
> sresid <- studres(md)
> hist(sresid, freq=FALSE,main="Distribution of Studentized Residuals")
> Xmd<-seq(min(sresid),max(sresid),length=40)
> Ymd<-dnorm(Xmd)
> lines(Xmd, Ymd)
> # Evaluate homoscedasticity
> plot(fitted(md), studres(md))
> abline(0,0)
> # non-constant error variance test
> ncvTest(md)

> # plot studentized residuals vs. fitted values
> spreadLevelPlot(md)
> # Evaluate Collinearity
```

```
> vif(md) # variance inflation factors
> sqrt(vif(md)) > 2 # problem?

> # Test for Autocorrelated Errors
> durbinWatsonTest(md)
> # Global test of model assumptions
> library(gvlma)
> gvmodel <- gvlma(md)
> summary(gvmodel)
> ba <- read.gwt2nb("C:\\Users\\PC\\Desktop\\GWR\\cankaya\\ShapeFile\\Threshold
_0_119622.GWT", region.id=OBJECTID)
> ba
```

D. 2. S A R MODEL

```
> SAR=spautolm(dC$LnValue ~ dC$EMLAKTIPI_ +dC$GUVENLIK+dC$YUZO
LCUMU+dC$E_MetroUza+dC$E_AVMUzakl , data=dC, family = "SAR", nb2listw
(NY_nb))

> summary(SAR)

> SAR1=spautolm(dC$LnValue ~ dC$EMLAKTIPI_ +dC$GUVENLIK+dC$YUZO
LCUMU+dC$E_MetroUza+dC$E_AVMUzakl , data=dC, family = "SAR", nb2listw
(ba))

> summary(SAR1)

> summary(CAR)  (rook)

> summary(CAR1)  (threshold)

> SARresCor <- sp.correlogram(NY_nb, residuals(SAR), order = 5, method = "I",zer
o.policy =TRUE)

> SARMt=moran.test(residuals(SAR), nb2listw(NY_nb))
> SARMt


> SARMt1=moran.test(residuals(SAR), nb2listw(ba))
> SARMt1
```

D.3 SPATIAL AUTOCORRELATION

MORANS'I

```
> moran.test(dC$LnValue, nb2listw(snbk1))
```

```
> sp.correlogram(snbk1, dC$LnValue, order=3, method="I",zero.policy=TRUE)
> Mt=moran.test(residuals(md), nb2listw(NY_nb))
> Mt #Result: p-value < 2.2e-16, deviation from expected value is significant, means
autocorrelation.

> moran.plot(residuals(md), nb2listw(NY_nb)) #Result: Trend induces spatial autoco
rrelation
> Mt1=moran.test(residuals(md), nb2listw(ba))
> Mt1 #Result: p-value < 2.2e-16, deviation from expected value is significant, mean
s autocorrelation.

> #999 Monte—Carlo  simulation  of Moran' s I #
> morpermLnValue<-moran.mc(dt$LnValue, dt_nbr_w, 999)  # W
> morpermLnValue1<-moran.mc(dt$LnValue, dt_nbr_wb, 999)  # B
> morpermLnValue

> morpermLnValue1

> MyMoran999

> MyMoran9999 <- moran.mc(dt$LnValue, listw = dt_nbr_w, nsim = 9999)
> MyMoran9999

> MyMoran99999 <- moran.mc(dt$LnValue, listw = dt_nbr_w, nsim = 99999)
> MyMoran99999
# plot Moran's I

hist(MyMoran999$res, breaks = 50)

> hist(MyMoran9999$res, breaks = 50)
> hist(MyMoran99999$res, breaks = 50)
> # Plotting Moran's I (looking for outliers...)
> mp <- moran.plot(dt$LnValue, dt_nbr_w, labels = as.character(dt$CNTY_ST), xla
b = "Percent PRICE", ylab = "Lag of Percent PRICE")
> #Another way plotting  Moran's I scatter
> par (mfrow=c(1 ,2))
> spc<- moran.plot(dt$LnValue,dt_nbr_w,spChk=NULL, labels=NULL, xlab="DEG
ER",ylab="spatially lagged DEGER", quiet=NULL, pch=19, main="Moran scatterpl
ot, I=0.3551, p=0.0000")
> dt$sLnValue <- scale(dt$LnValue)
> dt$lag_sLnValue <- lag.listw(dt_nbr_w, dt$sLnValue)
> plot(x = dt$sLnValue, y = dt$lag_sLnValue, main = "Moran Scatterplot LnValue")
> abline(h = 0, v = 0)
> abline(lm(dt$lag_sLnValue ~ dt$sLnValue), lty = 3, lwd = 4, col = "red")
> cspc <- sp.correlogram(dt_nbr, dt$LnValue, order=8,
method="corr", zero.policy=TRUE)
> plot(cspc, main="spatial correlogram")
```

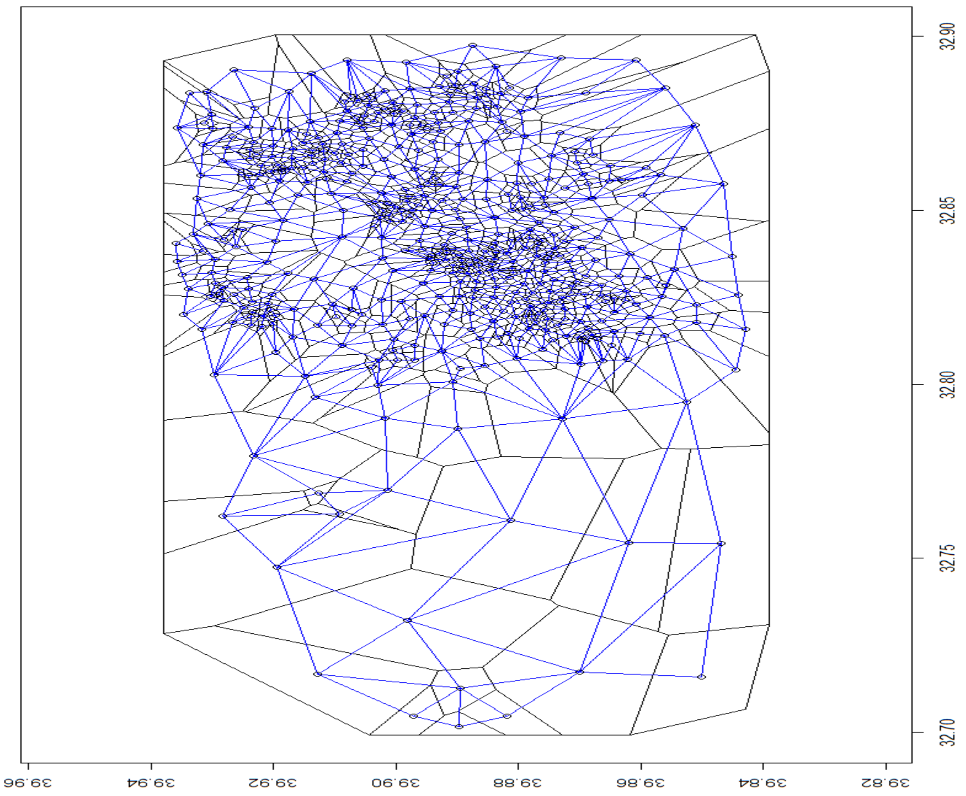


Figure 119: First order Rook contiguity neighbours for Çankaya

Neighbour list object:

Number of regions: 414
Number of nonzero links: 2190
Percentage nonzero weights: 1.27774
3
Average number of links: 5.289855
Link number distribution:
1 2 3 4 5 6 7 8 9 10

1 2 30 27 74 116 111 90 28 12 7 1
1 2
2 least connected regions:
374 501 with 1 link
2 most connected regions:
313 391 with 13 links

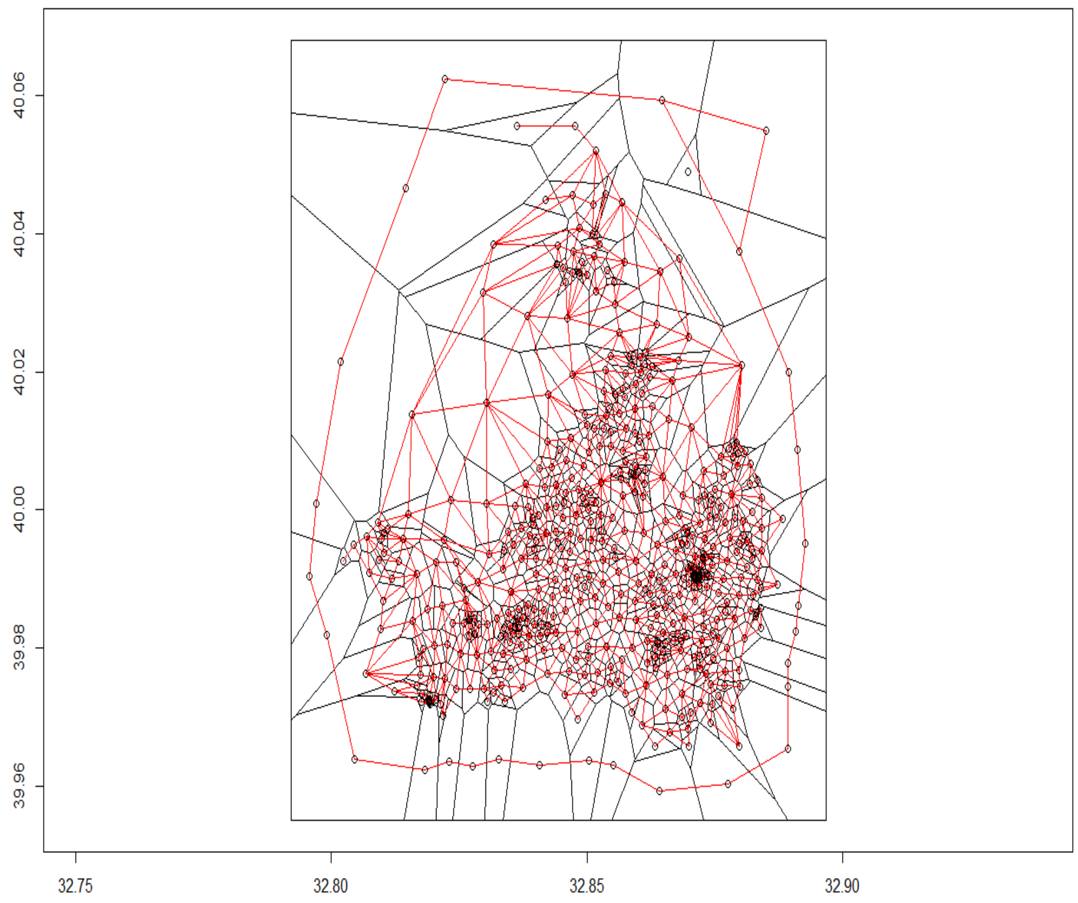


Figure 120: First order Rook contiguity neighbours for Keçiören

Neighbour list object:

Number of regions: 502
Number of nonzero links: 2766
Percentage nonzero weights: 1.097602
Average number of links: 5.50996
1 region with no links: 59

Link number distribution:

0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2	30	27	74	116	111	90	28	12	7	1	1	2

2 least connected regions:
374 501 with 1 link
2 most connected regions:
313 391 with 13 links

APPENDIX E

REAL-WORLD APPLICATIONS TO TEST THE MODELS

	
Aşağı Eğlence	**Etlik**
Area:110 m2 Floor: 1 Age:15	Area: 120 m2 Floor:2 Age: 10
Real estate agent price: 175.000 TL	Real estate agent price:200.000
OLS : 166.218 TL	OLS: 203.825 TL
SAR : 168.623 TL	SAR: 189.363 TL
GWR: 161.299 TL	GWR: 191.189 TL
Website: 165.000 TL	Website:195.000 TL
	
Esertepe	**Basınevleri**
Area: 120 m2 Floor:2 Age: 10	Area: 130 m2 Floor:3 Age: 10
Real estate agent price:180.000	Real estate agent price:190.000
OLS: 187.685 TL	OLS: 203.007 TL
SAR: 174.367 TL	SAR: 187.572 TL
GWR: 174.298 TL	GWR: 186.936TL
Website:175.000 TL	Website:185.000 TL

Figure 121: Comparison of realtors prices with the OLS, SAR and GWR Model Predictions Deflated by 2015 for Keçiören

	
Bağlarbaşı	**Adnan Menderes**
Area: 110 m2 Floor:2 Age: 7	Area: 110 m2 Floor:2 Age: 7
Real estate agent price:150.000 TL	Real estate agent price:150.000 TL
OLS: 156.768 TL	OLS: 168.150 TL
SAR: 146.667 TL	SAR: 157.301 TL
GWR: 141.283 TL	GWR: 153.283 TL
Website:142.000 TL	Website:148.000 TL
	
Ayvalı	**Kavacık Subay Evleri**
Area: 120 m2 Floor:2 Age: 10	Area: 120 m2 Floor:2 Age: 10
Real estate agent price:180.000	Real estate agent price:180.000
OLS: 187.685 TL	OLS: 187.685 TL
SAR: 174.367 TL	SAR: 174.367 TL
GWR: 174.298 TL	GWR: 174.298 TL
Website:175.000 TL	Website:175.000 TL

Figure 121 (continued): Comparison of realtors prices with the OLS, SAR and GWR Model Predictions Deflated by 2015 for Keçiören

	
Incesu	**Türközü**
Area: 120 m2 Floor:2 Age: 2	Area: 110 m2 Floor:1 Age: 10
Real estate agent price:310.000	Real estate agent price:245.000
OLS: 287.685 TL	OLS: 230.633 TL
SAR: 301.362 TL	SAR: 244.024 TL
GWR: 292.290 TL	GWR: 248.117 TL
Website:310.000 TL	Website:240.000 TL
	
Gaziosmanpaşa	**Bahçelievler**
Area: 100 m2 Floor:1 Age: 12	Area: 100 m2 Floor:2 Age: 50
Real estate agent price:355.000	Real estate agent price:370.000
OLS: 366.770 TL	OLS: 357.442 TL
SAR: 355.367 TL	SAR: 376.605 TL
GWR: 343.298 TL	GWR: 379.893 TL
Website:350.000 TL	Website:370.000 TL

Figure 122: Comparison of realtors prices with the OLS, SAR and GWR Model Predictions Deflated by 2015 for Çankaya

	
Mebusevler	**Esat**
Area: 125 m2 Floor:1 Age: 47	Area: 130 m2 Floor:2 Age: 38
Real estate agent price:375.000	Real estate agent price:280.000
OLS: 385.740 TL	OLS: 292.830 TL
SAR: 379.362 TL	SAR: 276.344 TL
GWR: 358.104 TL	GWR: 277.210 TL
Website:370.000 TL	Website:270.000 TL

Figure 122 (continued): Comparison of realtors prices with the OLS, SAR and GWR Model Predictions Deflated by 2015 for Çankaya

GWR Model Prediction: 400.132 TL, OLS Model Prediction: 385.270 TL

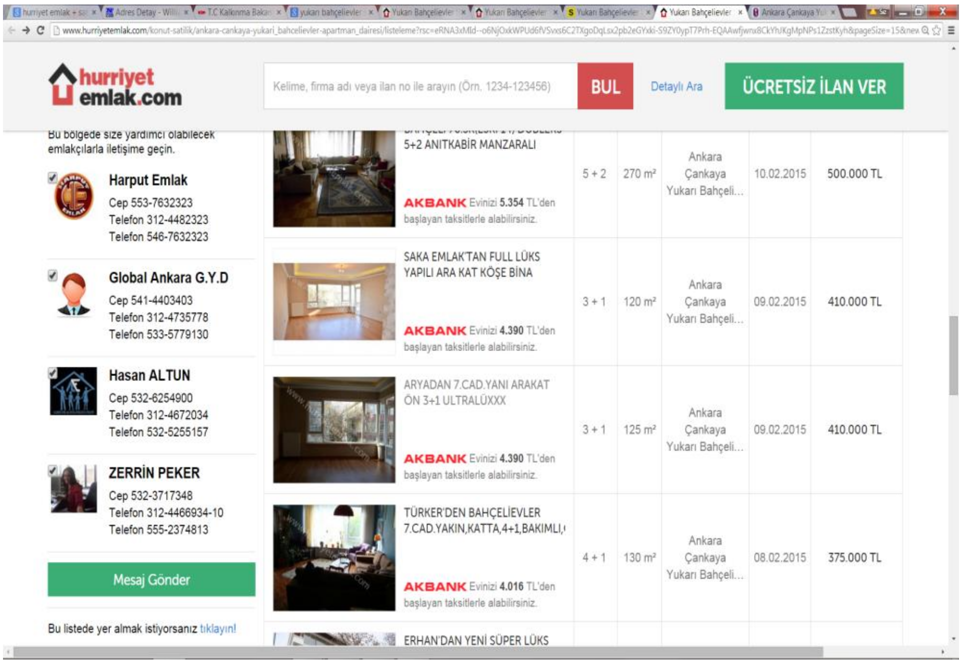


Figure 123: Comparison of popular real estate website prices with the OLS and GWR Model Predictions for Çankaya (Yukarı Bahçelievler Neighborhood)

GWR Prediction: 654.346 TL OLS Prediction: 630.204 TL

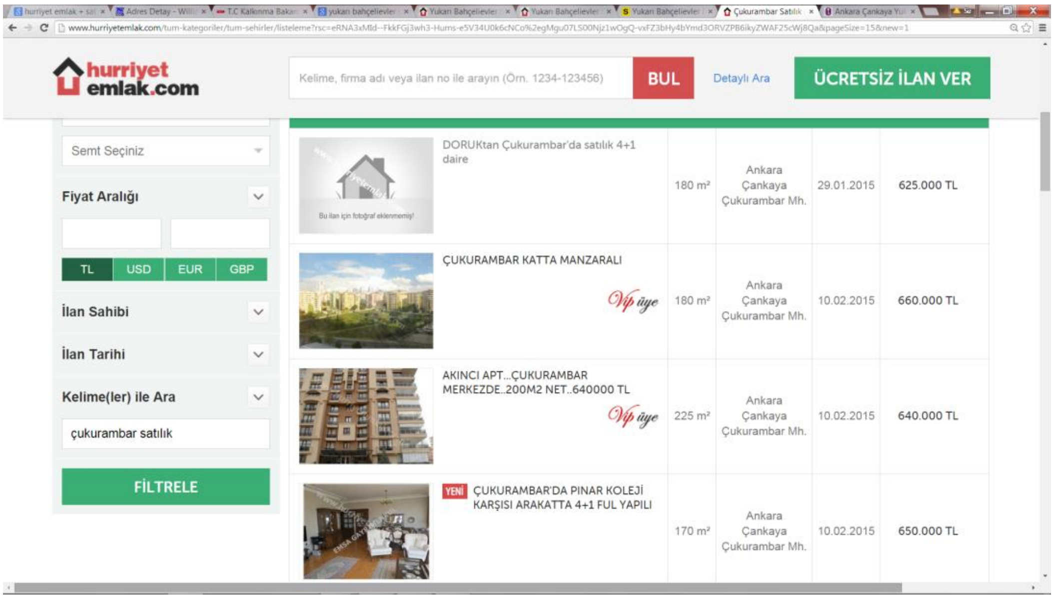


Figure 124: Comparison of popular real estate website prices with the OLS and GWR Model Predictions for Çankaya (Çukurambar Neighborhood)

GWR Model Prediction: 586.417 TL, OLS Model Prediction: 574.698 TL

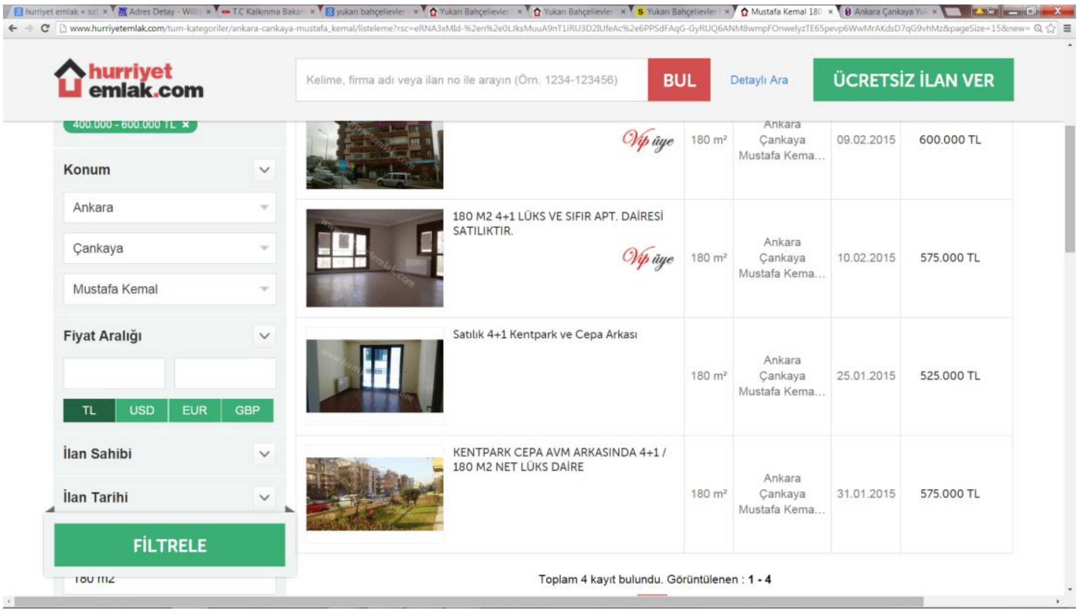


Figure 125: Comparison of popular real estate website prices with the OLS and GWR Model Predictions for Çankaya (Mustafa Kemalpaşa Neighborhood)

GWR Model Prediction: 435.668 TL, OLS Model Prediction: 433.479 TL

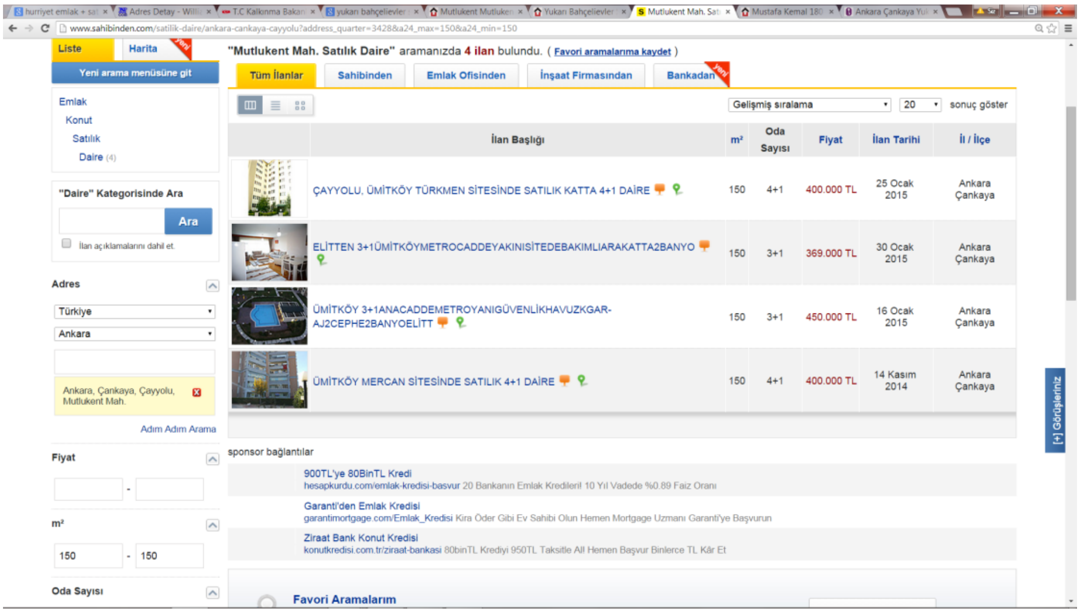


Figure 126: Comparison of popular real estate website prices with the OLS and GWR Model Predictions for Çankaya (Ümitköy- Mutlukent Neighborhood)

GWR Model Prediction: 236.007 TL, OLS Model Prediction: 211.419 TL

Figure 127: Comparison of popular real estate website prices with the OLS and GWR Model Predictions for Çankaya (Ayrancı Neighborhood)

GWR Model Prediction: 175.320 TL, OLS Model Prediction: 184.167 TL

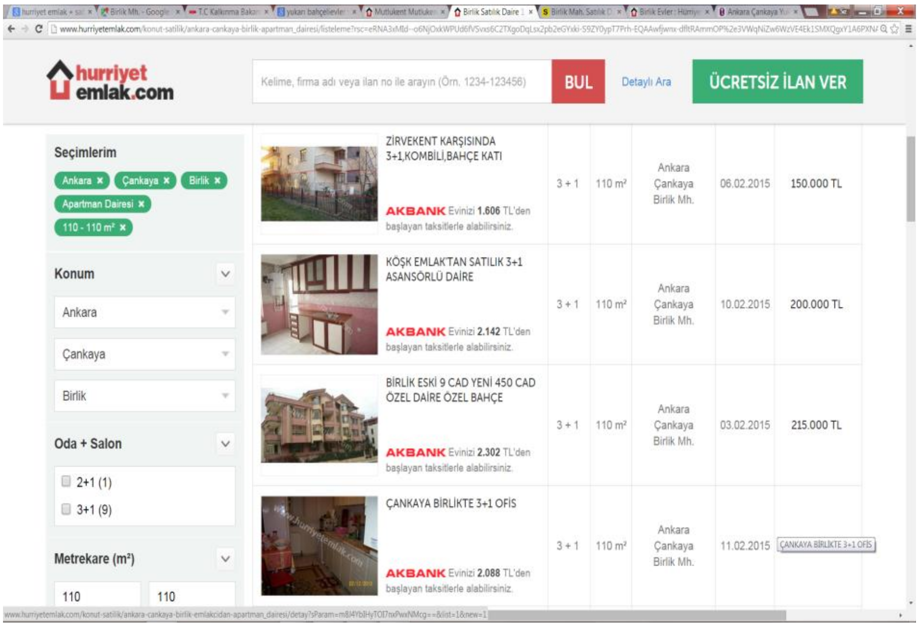


Figure 128: Comparison of popular real estate website prices with the OLS and GWR Model Predictions for Çankaya (Birlik Neighborhood)

GWR Model Prediction: 187.325 TL, OLS Model Prediction: 176.841 TL

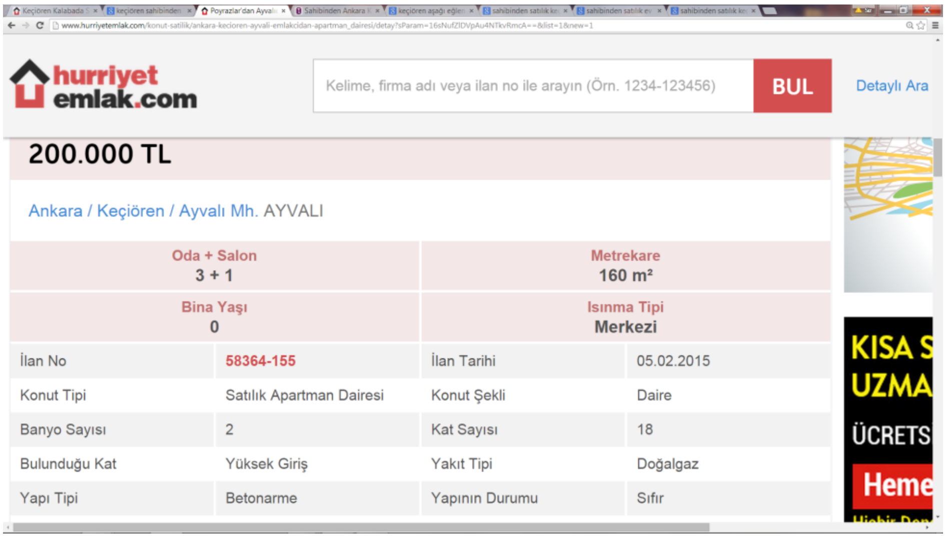


Figure 129: Comparison of popular real estate website prices with the OLS and GWR Model Predictions for Keçiören (Ayvalı Neighborhood)

GWR Model Prediction: 243.566 TL, OLS Model Prediction: 266.110 TL

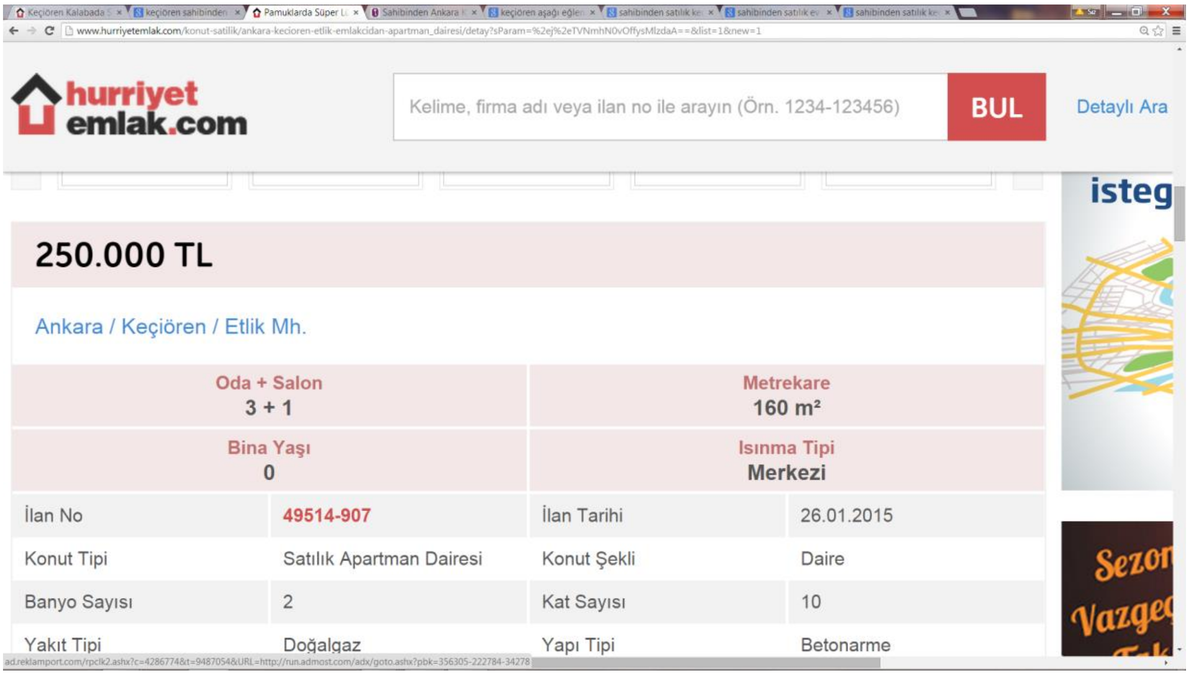


Figure 130: Comparison of popular real estate website prices with the OLS and GWR Model Predictions for Keçiören (Etlik Neighborhood)

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name	: Boza, Ertuğrul
Email	: ebosza@gmail.com
Title	: Phd.Student
Nationality	: Turk
Date of Birth	: 27 June 1967
Marital Status	: Married
Address	: Ministry of Development. Information Society Dept. Ankara Turkey
Phone	: 0 505 507 97 25

ACADEMIC QUALIFICATIONS

Degree **Institude** **Year of Graduation**

Ph.D METU, Geodetic and Geographic Inf. Tech. Dept. 2008-

MS METU, Informatics Institute.Software Management 2002 – 2005

BS YTU, Geodesy and Photogrammetry Eng. Dept. 1990- 1994

EMPLOYEMENT

Year	Place	Enrollment
2006- :	Ministry of Development:	GIS Specialist
1987 – 2005	General Directorate of Land Registry and Cadastre:	Engineer