INVESTIGATION OF THE SPATIAL RELATIONSHIP OF MUNICIPAL SOLID WASTE GENERATION IN TURKEY WITH SOCIO-ECONOMIC, DEMOGRAPHIC AND CLIMATIC FACTORS

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

BY

SANİYE KESER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ENVIRONMENTAL ENGINEERING

FEBRUARY 2010

Approval of the thesis:

INVESTIGATION OF THE SPATIAL RELATIONSHIP OF MUNICIPAL SOLID WASTE GENERATION IN TURKEY WITH SOCIO-ECONOMIC, DEMOGRAPHIC AND CLIMATIC FACTORS

submitted by **SANİYE KESER** in partial fulfillment of the requirements for the degree of **Master of Science in Environmental Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen ____________________
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Göksel N. Demirer ____________________
Head of Department, **Environmental Engineering**

Assoc. Prof. Dr. Ayşegül Aksoy ____________________
Supervisor, **Environmental Engineering Dept., METU**

Assoc. Prof. Dr. H. Şebnem Düzgün ____________________
Co-Supervisor, **Mining Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. Gürdal Tuncel ____________________
Environmental Engineering Dept., METU

Assoc. Prof. Dr. Ayşegül Aksoy ____________________
Environmental Engineering Dept., METU

Prof. Dr. F. Dilek Sanin ____________________
Environmental Engineering Dept., METU

Dr. Tuba H. Ergüder ____________________
Environmental Engineering Dept., METU

Assoc. Prof. Dr. H. Şebnem Düzgün ____________________
Mining Engineering Dept., METU

Date: ____________________

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 :

Signature :

ABSTRACT

INVESTIGATION OF THE SPATIAL RELATIONSHIP OF MUNICIPAL SOLID WASTE GENERATION IN TURKEY WITH SOCIO-ECONOMIC, DEMOGRAPHIC AND CLIMATIC FACTORS

Keser, Saniye

M. Sc., Department of Environmental Engineering

Supervisor: Assoc. Prof. Dr. Ayşegül Aksoy

Co-Supervisor: Assoc. Prof. Dr. H. Şebnem Düzgün

February 2010, 123 pages

This thesis investigates the significant factors affecting municipal solid waste (MSW) generation in Turkey. For this purpose, both spatial and non-spatial techniques are utilized. Non-spatial technique is ordinary least squares (OLS) regression while spatial techniques employed are simultaneous spatial autoregression (SAR) and geographically weighted regression (GWR). The independent variables include socio-economic, demographic and climatic indicators. The results show that nearer provinces tend to have similar solid waste generation rate. Moreover, it is shown that the effects of independent variables vary among provinces. It is demonstrated that educational status and unemployment are significant factors of waste generation in Turkey.

Keywords: Municipal Solid Waste, Simultaneous Spatial Autoregression, Geographically Weighted Regression, Spatial Analysis

ÖZ

TÜRKİYE'DEKİ EVSEL KATI ATIK ÜRETİMİNİN SOSYO-EKONOMİK, DEMOGRAFİK VE İKLİMSEL FAKTÖRLERLE OLAN MEKANSAL İLİŞKİSİNİN İNCELENMESİ

Keser, Saniye

Yüksek lisans, Çevre Mühendisliği Bölümü

Tez danışmanı: Doç. Dr. Ayşegül Aksoy

Ortak tez danışmanı: Doç. Dr. H. Şebnem Düzgün

Şubat 2010, 123 sayfa

Bu çalışmada, Türkiye'de belediye katı atık üretimini etkileyen önemli etkenler araştırılmıştır. Bu amaçla, hem mekansal hem de mekansal olmayan teknikler kullanılmıştır. Mekansal olmayan teknik en küçük kareler yöntemi, mekansal olan teknikler ise eşzamanlı mekansal otoregresyon ve coğrafi ağırlıklandırılmış regresyondur. Bağımsız değişkenler sosyo-ekonomik, demografik ve iklimsel göstergeleri içermektedir. Sonuçlara göre birbirine yakın olan iller benzer katı atık üretim oranına sahip olma eğilimindedir. Ayrıca, bağımsız değişkenlerin etkisi iller arasında farklılık göstermektedir. Eğitim durumu ve işsizliğin, Türkiye'de katı atık üretimini etkileyen önemli faktörler olduğu gösterilmiştir.

Anahtar Kelimeler: Belediye Katı Atığı, Eşzamanlı Mekansal Otoregresyon, Coğrafi Ağırlıklandırılmış Regresyon, Mekansal Analiz

To My Aunt

ACKNOWLEDGEMENTS

I owe my deepest gratitude to my family for their endless patience and support.

I am sincerely grateful to my supervisor Assoc. Prof. Dr. Ayşegül Aksoy and co-supervisor Assoc. Prof. Dr. H. Şebnem Düzgün for broadening my horizons and for their valuable guidance.

I am also heartily thankful to my dear friends who have been always there for me.

TÜBİTAK is gratefully acknowledged for providing scholarship throughout my graduate education.

Lastly, I would like to thank to all those who provided any kind of support throughout the accomplishment of this thesis.

TABLE OF CONTENTS

ABSTRACT iv

ÖZ v

DEDICATION vi

ACKNOWLEDGEMENTS vii

TABLE OF CONTENTS viii

LIST OF TABLES x

LIST OF FIGURES xi

ABBREVIATIONS xii

LIST OF SYMBOLS xiii

CHAPTER

1. INTRODUCTION 1

2. LITERATURE SURVEY 5

2.1. Studies on Determination of Solid Waste Generation & Relating Them to Other Factors 5

2.2. Studies with SAR & GWR 17

3. METHODS 21

3.1. General Framework of Methodology 21

3.2. Data Collection 21

3.3. Multicollinearity Analysis 25

3.4. Non-Spatial Data Analysis 29

3.5. Spatial Data Analysis 30

3.5.1. Spatial Autocorrelation 30

3.5.2. Spatial Regression 34

3.5.2.1. Simultaneous Spatial Autoregression (SAR) 34

3.5.2.2. Geographically Weighted Regression (GWR) 36
3.5.3. OLS Regression with Eliminated Neighboring Effect 38
3.6. Validation 39
4.IMPLEMENTATION of METHODOLOGY & RESULTS 41
4.1. Data Collection 41
4.2. Multicollinearity Analysis 48
4.3. Non-Spatial Data Analysis 53
4.4. Spatial Data Analysis 55
4.4.1. Spatial Autocorrelation 55
4.4.2. Spatial Regression 58
4.4.2.1. Simultaneous Spatial Autoregression (SAR) 58
4.4.2.2. Geographically Weighted Regression (GWR) 61
4.4.3. OLS Regression with Eliminated Neighboring Effect 77
4.5. Validation 79
5. DISCUSSIONS 81
5.1. Data 81
5.2. Models 82
5.3. Significant Determinants of MSW Generation Rate 86
6. CONCLUSIONS & RECOMMENDATIONS FOR FUTURE WORK 91
6.1. Conclusions 91
6.2. Recommendations for Future Work 92
REFERENCES 93
APPENDICES
A. SUPPLEMENTARY INFORMATION 102
B. SCRIPTS 106
C. FIGURES 116
D. TABLES 121

LIST OF TABLES

TABLES

Table 1: Explanatory variables 43

Table 2: Iterations to obtain VIF< 4 for explanatory variables in DG1 48

Table 3: Correlation matrix... 49

Table 4: Explanatory variables in data groups 50

Table 5: VIFs of explanatory variables 51

Table 6: *r* values for explanatory variables 52

Table 7: OLS regression coefficient estimates 54

Table 8: Moran's I values for different spatial weighting scheme.. 56

Table 9: Monte Carlo test for Moran's I at spatial lag 4 57

Table 10: Moran's I values for OLS regression residuals.. 57

Table 11: Lagrange Multipliers test 58

Table 12: SAR_{lag} coefficient estimates.. 59

Table 13: SAR_{err} coefficient estimates 60

Table 14: Moran's I values for SAR residuals 61

Table 15: Summary of GWR coefficient estimates for DG1 62

Table 16: Summary of GWR coefficient estimates for DG2 62

Table 17: Summary of GWR coefficient estimates for DG3 63

Table 18: Moran's I values for GWR residuals 71

Table 19: OLS coefficient estimates for the selected subset.. 78

Table 20: Validation results 79

Table D.1: Observed MSW generation rates for 1990 and 2000 121

Table D.2: Summary statistics of explanatory variables.. 122

Table D.3: p-values of r values 123

LIST OF FIGURES

FIGURES

Figure 1: Flowchart of Methodology 22

Figure 2: Neighbors of Ankara based on (a) 1st nearest neighbors (lag 1); (b) 2nd nearest neighbors (lag 2); (c) 3rd nearest neighbors (lag 3); (d) 4th nearest neighbors (lag 4); (e) contiguity 33

Figure 3: (a) A spatial kernel; (b) GWR with fixed spatial kernels 37

Figure 4: (a) The choropleth map and (b) cartogram of MSW generation rate in Turkey for the year 2000 42

Figure 5: Distribution of explanatory variables in Turkey (2000) 46-47

Figure 6: Correlogram for MSW generation rate 56

Figure 7: GWR coefficient estimates for DG1 64

Figure 8: t-scores for GWR coefficient estimates of DG1 66

Figure 9: Significant GWR coefficient estimates for DG1 67

Figure 10: R^2 distribution of GWR (DG1) 72

Figure 11: t-score for GWR coefficient estimates of DG2 73

Figure 12: t-score for GWR coefficient estimates of DG3 74

Figure 13: Significant GWR coefficient estimates for DG2 75

Figure 14: Significant GWR coefficient estimates for DG3 76

Figure 15: Subset of the provinces chosen for OLS regression 77

Figure 16: MSW generation rates in Turkey: Observed and predicted (1990) 85

Figure 17: Difference between GWR and SAR coefficient estimates of DG1 88

Figure C.1: Provinces in Turkey 116

Figure C.2: Q-Q plots of the variables 117-120

ABBREVIATIONS

CBRT	Central Bank of the Republic of Turkey
EEA	European Environment Agency
EU	European Union
GDP	Gross Domestic Product
GIS	Geographical Information System
GWR	Geographically Weighted Regression
IMR	Infant Mortality Rate
MoEF	Ministry of Environment and Forest
MoIT	Ministry of Industry and Trade
MSW	Municipal Solid Waste
OID	Organized Industrial District
OLS	Ordinary Least Squares
RTCE	Related Total Consumer Expenditure
SAR	Simultaneous spatial autoregression
SAR_{err}	Simultaneous Spatial Autoregression Error Model
SAR_{lag}	Simultaneous Spatial Autoregression Lag Model
SPO	State Planning Organization
SWM	Solid Waste Management
TCA	Turkish Court of Accounts
TSMS	Turkish State Meteorological Service
TurkStat	Turkish Statistical Institute

LIST OF SYMBOLS

AIC	Akaike Information Criterion
H_o	Null hypothesis
$\mathbf{I}$	Identity matrix
LM	Lagrange multiplier
LM_{err}	Lagrange multiplier for spatial error model
LM_{lag}	Lagrange multiplier for spatial lag model
RLM	Robust Lagrange Multiplier
RLM_{err}	Robust Lagrange Multiplier for Error Model
RLM_{lag}	Robust Lagrange Multiplier for Lag Model
R^2	Coefficient of determination
R^2_{adj}	Adjusted coefficient of determination
R^2_k	Coefficient of determination when the explanatory variable k is the dependent variable in the regression equation
r	Pearson's product moment correlation coefficient
$RMSE$	Root mean squared error
VIF	Variance inflation factor
VIF_m	Variance inflation factor calculated with R^2 of the model
VIF_k	Variance inflation factor calculated with R^2_k of the model
α	Significance level
X_{k_1}, X_{k_2}	Vectors of explanatory variable k_1 and k_2
$\bar{X}_{k_1}, \bar{X}_{k_2}$	Means of explanatory variables k_1 and k_2
X	Matrix of explanatory variables in each province
Y	Vector of dependent variable in each province
$\bar{Y}$	Mean of the dependent variable
Y_i, Y_j	Dependent variables at data points i and j, respectively
β	Vector of regression coefficients of each explanatory variable
ϵ	Error vector
n	Number of data points

k	Number of explanatory variables
I	Moran's I
W	Spatial proximity matrix
w_{ij}	Element of the spatial proximity matrix
w^s_{ij}	Row standardized element of the spatial proximity matrix
ρ	Simultaneous autoregressive lag coefficient
γ	The coefficient for lagged explanatory variables
λ	Simultaneous autoregressive error coefficient
y_i	Dependent variable at regression point i
x_{ik}	Explanatory variable k at regression point i
ϵ_i	Error term at regression point i
(u_i, v_i)	Coordinates of regression point i
$\beta_k(u_i, v_i)$	Regression coefficient of explanatory variable k at regression point i
d_{ij}	Distance between the regression point i and data point j
b	Bandwidth of GWR kernel
Y'	Vector of predicted value of dependent variable
b_{OLS}	Vector of β coefficient estimates of OLS regression
b_{SAR}	Vector of β coefficient estimates of SAR_{lag} regression
B_{GWR}	Matrix of β coefficient estimates of GWR regression
$\hat{y}_i$	Predicted value of dependent variable at data point i

CHAPTER 1

INTRODUCTION

Wastes arise from our daily activities and waste generation gains pace with industrialization. Consequently, solid waste management (SWM) constitutes one of the major parts of environmental management. Solid wastes may cause serious environmental and health problems when improperly handled. Methods of SWM have evolved with concerns shifting towards sustainability. Managing solid wastes used to be composed of collection and disposal practices. These practices have been enhanced and new concepts were introduced in a systematic way with the sustainable environment approach. Today, reducing waste generation is in the front rank in SWM. This is supported by other concepts like reuse, recovery and recycling.

All the methods that SWM encompasses depend on the solid waste generation data for planning, design and evaluation. Collection vehicles, routes and frequencies are arranged according to solid waste generation rates. Capacities of landfills and incineration plants are determined making use of current and projected waste generation rates and they are designed accordingly. Actions that are taken to reduce solid waste to be disposed off are planned considering waste generation rates and also composition in many cases. The targets are expressed in different units of solid waste generation rates and the results are evaluated again with this variable. However, data on solid waste generation is very scarce in most of the countries due to lack of systematic monitoring and sampling although it is an essential variable for effective SWM.

Solid waste generation is affected by different aspects. This is why setting a general prototype model for solid waste generation is impossible. The characteristics of solid waste generation are unique to the region. The solid waste generation model for European Union (EU) countries, as an instance, may not be valid for Turkey. The differences regarding solid waste generation between countries stems from one very important determinant –consumption pattern, because waste generation is concomitant with consumption (Purcell & Magette, 2009). It is how much and what type goods we consume during daily activities that determine the amount and type of solid waste we generate. In other words, consumption behavior is the key factor affecting solid waste generation rate and composition. This behavior, in turn, is affected by many other factors. These range from environmental factors to economic and demographic factors. Whether these factors are originating naturally or enforced through fees, fines, etc., they are reflected in the attitude and behavior of people regarding both consumption and environment.

In Turkey, the Ministry of Environment and Forestry (MoEF) is in charge of compiling data on solid waste generation and enforcing municipalities to manage their solid wastes according to the regulations. According to MoEF, 45 % of the municipal solid wastes (MSW) were managed in compliance with the regulations in 2007. Turkish Statistical Institute (TurkStat) is the responsible institute in cooperation with MoEF for compiling MSW data. The data is collected based on administrative records. Efforts are made by MoEF to build up inventory for different waste streams as well. For example, the data on recycling of packaging materials has been recorded since 2005 in which the Regulation on the Control of Packaging Wastes came into force. (MoEF, 2008)

Although there have been attempts to have a database for solid waste generation, Turkey is still facing many deficiencies in this aspect. Yet, the municipalities should pay necessary attention and cooperate with the related institutions to determine waste generation rates and trends, and to create a reliable database. One of the ways of achieving this is to relate the MSW generation rates to socio-economic and other factors. The results of this study which offers a model valid for Turkey considering socio-economic, demographic and climatic differences between provinces may be a supporting tool. The versatility of the study in terms of variables may result in comprehensive projections. This model is also important in that it presents the determinants of solid waste generation in Turkey which is very useful in supporting decisions in action plans for solid waste reduction.

The aim of this theses work is to spatially analyze the solid waste generation rates with respect to socio-economic, demographic and climatic factors, and construct a model for the case of Turkey. The analyses were conducted in the bases of 81 provinces in Turkey. Data on MSW generation rates and explanatory variables were compiled from institutions and other studies. Data was subjected to both spatial and non-spatial statistical analyses. Spatial locations of data are of concern when applying spatial methods. In this study, area data is used in spatial analyses, i.e. the data is assigned to areal entities. These areal entities are irregular shaped polygons denoting provinces in Turkey.

The main spatial data analysis methods employed in this research are the Simultaneous Spatial Autoregression (SAR) and the Geographically Weighted Regression (GWR). These are similar to widely used linear regression models, yet they include spatial dependency. Spatial dependency is the dependency of a variable in nearby locations and it is of interest when the data is spatial. Since the results are affected by this phenomenon, spatial dependency should be incorporated in the models. In this respect, spatial models outperform non-spatial models. To obtain more reliable results, SAR and GWR were applied to MSW generation rates. Moreover, GWR gave access to more information since it is a local model. The relation of MSW generation rate to explanatory variables could be observed locally via GWR. From this point of view, it is more advantageous compared to other models for planning purposes which should consider the regional differences.

The determinants of the MSW generation rates were chosen from socio-economic and demographic variables as well as two environmental factors (i.e. temperature and rainfall). The analyses were carried out for the year 2000[1] due to ease of accessibility to data. As a result of the study, the factors affecting solid waste generation rates and the extent of their effect were determined.

The analyses were performed using a free software: R 2.8.0 (R-project, 2009). The R software provides an environment for a wide variety of statistical computations and

[1] The last General Population Census in Turkey was conducted in 2000 by TurkStat, in which socio-economic and demographic data were available as well. Address Based Population Registration System results for 2007 or 2008 were not used due to the inadequacy of the system to provide socio-economic variables.

graphics and can run on many platforms including Windows and MacOS. There are a high number of contributing packages in R to extend its usage for data manipulation, calculation and graphical display. MapInfo Professional 7.5, on the other hand, was used for mapping purposes and it is a commercial software running under Windows (MapInfo, 2009). This software helps to visualize the relationships between data and geography and performs comparatively simple data analyses.

A brief summary regarding related literature is given in Chapter 2 following this introduction. After the literature survey, Chapter 3 provides background information for the methods used. Then, Chapter 4 explains how these methods applied to MSW generation rate in Turkey and gives associated results. Discussions of the results are presented in Chapter 5. The report is ended in Chapter 6 with a conclusion section including main inferences.

CHAPTER 2

LITERATURE SURVEY

2.1. STUDIES ON DETERMINATION OF SOLID WASTE GENERATION & RELATING THEM TO OTHER FACTORS

Research on solid waste generation has a broad spectrum and it is still expanding. Solid waste generation as a subject can be dealt with from many perspectives. The purpose for investigating solid waste generation may be just for forecasting reasons. Then, the target is to find methods with better prediction power. On the other hand, the purpose of the researches shifts towards sustainability related issues due to decreasing handling capacity of waste generated and unwanted effects of inefficient SWM. The motives behind waste generation and the ways of using this information to reduce wastes at source and increase recycling are generally investigated with this purpose. Therefore, how and why waste is generated becomes of interest rather than focusing only on how much is generated.

Independent of the purpose of the study, the most differentiating feature of solid waste generation studies is the modeling technique. Some studies make use of temporal differences. The analysis may be carried out to detect seasonal variations or more complex methods can be used to model changes in waste generation over time. Data of frequent intervals within a year is needed to observe seasonal variations, which is generally obtained by sampling in the study area. Gómez, *et al.* (2009) carried out a detailed sampling survey in a city of Mexico by collecting and measuring various fractions of MSW. They measured the waste amount generated and identified the waste fractions at each day during a week in three different months of the year representing different seasons and found that waste generation was lower in winter for all socio-economic classes. They explained this trend by lower consumption of drinks, fresh foods and vegetables in winter months. Similar studies were also conducted in two cities of Nigeria for each month showing the variation in waste generation over the year (Afon, 2007; Afon and Okewole, 2007). Afon (2007) found that the waste generation was the highest in the harvest season due to consumption of farm products. Afon and Okewole (2007) also came up with the same result in a different city of Nigeria. Moreover, it was found that Saturdays had the highest waste generation in both studies since this was a work-free day on which general house cleaning and ceremonies took place. Buenrostro, *et al.* (2001) measured the waste deposited at the dumping site during more than a year in a Mexican city not only to show seasonal variation but also to determine any discrepancy between the generated and collected waste and as a result waste collected was found to be lower than the estimated generation. This result showed that waste measurements in the dumping ground were not reliable. Moreover, they found that waste generation was highest in the rainy season while it was lowest in dry season due to low content of humidity. Another survey conducted in a Greek island revealed the effect of tourism on waste generation by observing seasonal variation (Gidarakos, *et al.*, 2006). The waste generated was found to be higher in the seasons with increased number of tourists. In tourist regions, the fraction of glass increased which was especially composed of non-refillable bottles. On the other hand, plastic waste generation was higher in winter and autumn due to greenhouse construction.

Banar and Özkan (2008) gave an example of the same type of study in Turkey. They analyzed the change in percentage of waste components according to the season. The ash component of the solid waste generated was found to be highest in winter and autumn due to coal consumption for heating purposes. The share of food wastes increased in summer and spring since vegetable and fruit were consumed more in these seasons.

It is seen that seasonal variations play an important role in waste generation and convey essential information for waste management. However, continuous and frequent data is needed for studies investigating seasonal variations. It is important for the measurement periods to be representative of the related season if the sampling is not carried out for each month.

There are also more complicated temporal analyses which make use of discrete time series. Katsamaki, *et al.* (1998) applied a time series model called Box-Jenkins to daily MSW generation including autoregressive and moving average model types. They found a significant seasonality of length 5 (e.g. 5 days) explained by the collection pattern and showed that Box-Jenkins models estimated the waste generation and fluctuations adequately. Navarro-Esbrí, *et al.* (2002) also applied seasonal autoregressive and moving average model together with non-linear dynamics technique. They both used daily and monthly residential waste data and showed the good forecasting power of the model.

In time series analyses, the data should include values measured at equal time intervals. This interval may be one day or even a year. These analyses include regression techniques based on the dependency of a value on previous values in temporal extent and can be used for predictions with high performance especially in short-term projections.

The analyses based on time series are mostly carried out for prediction purposes and do not include variables other than amount or rate of waste generation. Besides, the majority of the research on solid waste generation tries to relate this variable to others. In spite of the possibility of using these methods also in prediction, the purpose is generally determining the driving factors of waste generation. In addition to the methods utilized, these studies differ from each other in the independent variables used.

The simplest method of bivariate analysis is to obtain correlation coefficients. Adedibu (1985) calculated correlation coefficients between different waste components and socio-economic variables regarding income, education and household size. Correlation analysis determines the relationship between two variables, but do not provide any information on which variable is dependent on the other. Although measuring correlation is a practical way to determine the relation between the possible determinants and different waste types, it cannot demonstrate the form of the relationship which can be defined as “the nature of the control that one variable exerts over the other” (Walford, 1995). Moreover, bivariate correlations are not satisfactory considering the complex dynamics of waste generation process since they do not give information about combined effects of more than one factor. This is why correlation coefficients are used as a supplementary tool in most of the studies which include other techniques as the main analysis.

One of the most used techniques is categorical comparison which can combine different independent variables to obtain a representative class. In these studies, the study region or the samples are partitioned into classes according to the attributes of independent variables. Then, these classes are compared according to the waste generation.

Buenrostro, *et al.* (2001) and Gómez, *et al.* (2009) obtained three socio-economic classes (e.g. low, middle and high) based on income level in different cities of Mexico. To determine the income level, the authors made use of the monthly wage per household information. As a result, they showed that waste generation rate increased with increasing socio-economic level. Paper and yard waste generation rate were found to be higher in the highest socio-economic level while low-income level generated higher plastics and batteries. Variations in other waste types among different socio-economic classes were also discussed. Another study in a city of Mexico showed that there was not a statistically significant relation between the hazardous household waste generation rate and income based socio-economic level (Buenrostro, *et al.*, 2008). On the other hand, different components of hazardous household waste were found to be related to income strata in this study. They further pointed out that waste generation frequency was higher in lower-income strata due to purchasing products in smaller containers. In this stratum, the residents also preferred short-life brands since they are cheaper. Banar and Özkan (2008) also compared the percentage of waste components between socio-economic levels stratified by income. They further calculated the correlation coefficients between income and waste components. Waste components except for food waste and ash were found to be positively related to income level. Negative correlation of ash was attributed to the usage of coal as a fuel in low-income group. Although food wastes were lower in amount in higher socio-economic level, the packaging wastes were higher and the results showed that moisture content decreased with increasing income level.

As a result, the relation of waste generation and composition is highly related to income. The direction of the relation depends on many other factors which are in essence related to different lifestyles. The differences can be observed in heating means, nutritional habits, preferences on purchased products and many other lifestyle habits. However, income may not be the sole indicator while observing the impact of different lifestyles.

The grouping of the samples can be based on other indicators. For example, a different stratification was carried out in United Kingdom by Emery, *et al.* (2003). They used a classification system based on dwelling types with the help of a geodemographic information system called ACORN. These dwelling types were terraced, semi-detached and council houses. The residents living in the modern semi-detached houses, which denoted the high prosperity level, were found to generate far more waste per household, which is consistent with outcome of other studies. These households produced more packaging materials due to higher consumption of comestibles. On the other hand, newspaper as a subcategory of paper wastes and glass waste generation were lower in these dwellings due to increased rate of recycling. Therefore, it can be concluded that the recycled materials were not included in the analyses which is not the case for all waste generation studies.

Dennison, *et al.* (1996) grouped the samples according to the household size rather than income and determined its effect on household waste generation per capita in Dublin. They found that 1-person households doubled the generation per capita compared to 4-person households. Although there was a decreasing trend in household size recently in the city, the waste generation per capita was determined to decrease in the last decade due to reduction in usage of solid fuels for domestic heating purposes. Geographic information system (GIS), moreover, was used in this study to visually represent the distribution of household size and per capita waste generation in terms of the wards of the city. This showed that adjacent wards had similar attributes of household size and waste generation rate. Based on these similarities, the authors proposed that “waste generation zones” could be a more practical than ward-based system.

A simultaneous stratification can also be employed to extract more information. As an instance, Ojeda-Benítez, *et al.* (2008) included stratification based on family typology in addition to the one based on income. Similar to previous studies, the waste generation per capita was found to be positively related to income. Furthermore, the households were grouped as extended, nuclear and mono-parental families. The waste generation per capita was highest in mono-parental families and lowest in the extended families. On the other hand, waste generation per family unit was highest in extended families in middle-income class. By this way, the effect of income could be decomposed. This demonstrated the effectiveness of considering other factors in addition to income. However, comparing the result in different strata may be challenging when there are more than two criteria for stratification.

Afon (2007) made use of ecological zones to determine the relation of solid waste generation to socio-economic indicators. These zones belong to different socio-economic levels due to historical development of cities in Nigeria. It was found that the waste generation rate increases with decreasing education, income and social status on contrary to other research results. This unexpected result was explained by the different food choices. Residents in high socio-economic class generated manufactured products while heavier organic wastes were produced in the low socioeconomic class. This pattern was also observed by Banar and Özkan (2008) in Eskişehir, Turkey. The food wastes were negatively related to income while there was a positive correlation between income and packaging wastes.

Another mostly used analysis technique in solid waste generation research is regression. Abu Qdais, *et al.* (1997) used bivariate (single) linear regression to determine the separate effects of income and household size on residential solid waste generation rate. Annual rental rates of the residents were used to infer the income level. They revealed that income had strong positive effect on residential solid waste generation rate while the effect of household size was not found that strong although it was observed that household size had a negative effect on waste generation rate. Hockett, *et al.* (1995), on the other hand, carried out a multivariate (multiple) linear regression and found that income was not a significant determinant of waste generation. As distinct from most of the studies, they also took retail sales and waste disposal fee as independent variables. Other variables included value added by manufacturing, construction costs per capita and percentage of urban population. Among these, they found that retail sales and disposal fee were significant determinants of solid waste generation rate. Retail sales affected waste generation per capita positively and this determinant was further investigated by dividing it into subclasses and carrying out a separate regression. As a result of this regression, the per capita sales of eating establishments were found to be a significant determinant of waste generation while the variable regarding retail sales of clothing stores had a zero coefficient. On contrary to retail sales, the coefficient for disposal fee was found to be negative. However, the mechanism that reduced waste generated with increasing disposal fee remained unexplained. In other words, the reduction in waste generation may be due to change in purchasing behavior or illegal dumping when the fee is increased. The negative effect of waste disposal pricing was also supported by the multiple non-linear regression analysis of Hibiki and Shimane (2006). They also pointed out that the households with low income group had lower price elasticity. Daskapoulos, *et al.* (1998) investigated the relation of MSW as annual tonnage to gross domestic product (GDP) in European countries and USA and further established a third degree polynomial equation relating the waste fractions to related total consumer expenditure (RTCE) via single non-linear regression. RTCE is the actual amount spent by the consumers on goods that are responsible for waste generation. The data covered a period of decades. As a result, they found that annual waste generation in tones was related to GDP and population positively.

Beigl, *et al.* (2004) and Bandara, *et al.* (2007) followed a different methodology by applying stratification and then applying regression analysis to each stratum separately. Beigl, *et al.* (2004) used more than one socio-economic indicator to create the categories for different cities in Europe and obtained regression equations for each category using a wider range of indicators. They included life expectancy and infant mortality rate (IMR) in the regression analysis of waste generation for the first time. The variables included in the final models were GDP, life expectancy, IMR, age structure and average household size since these were found to be significant. For very high and high prosperity level, GDP had a negative coefficient while IMR had a positive coefficient. For low and medium prosperity level, IMR and average household size affected waste generation rate in negative direction while life expectancy and percentage of population aged between 15 and 59 had a positive effect on waste generation. Bandara, *et al.* (2007), on the other hand, made use of property assessment tax value for stratification in a municipal area since it was the most reliable available data indicating the living standards. They presented the regression results for organic fraction of the waste and showed household size was a significant determinant with a negative relation to per capita generation for each stratum. For upper income group, the number of employed members in a household positively affected waste generation per capita. Other variables including number of motor vehicles owned, number of families in a household and property assessment tax value turned out to be insignificant.

As Parfitt and Flowerdew (1997) pointed out, data complexity increases with increasing sustainability efforts in waste analyses. This is why more complex methods have been emerged. One of them is system dynamics modeling technique indicating the dynamic nature of waste generation. One of the examples of applying system dynamics to waste generation is given by Sudhir, *et al.* (1997) to determine the impacts of various policy alternatives with giving attention to waste-pickers. They came up with a policy composed of three approaches. One of them was that waste-pickers should collect recyclables only after the formal body uses up its capacity. Second, funds should be allocated to collection, disposal and processing activities equally in proportion to the related requirements. Finally, these should be accompanied by user fees to recover the costs. Dyson and Chang (2005) benefited from system dynamics to predict the waste generation which can be used in site selection and capacity planning of MRFs. Among different driving factors, income per service center was chosen to be more appropriate since it incorporates other driving factors such as population per service center, population per household, income per household and historical amount generated.

There are also applications of fuzzy logic in solid waste generation. Chen and Chang (2000) developed grey fuzzy dynamic modeling for the prediction of urban solid waste generation to handle the case of limited samples. Karavezyris, *et al.* (2002)

combined system dynamics with fuzzy logic incorporating illegal disposal into the model. The model included the factors as environmental behavior, the amount of waste recycled, treatment price, waste generated and regulation regarding illegal disposal.

The studies above mainly target the factors affecting waste generation. On the other hand, the aim can be to determine the extent of waste generation by means of data on production, trade and consumption. The former type of studies refers to *'factor model'* and the latter is called *'input-output model'* (Beigl, *et al.*, 2004). Gay, *et al.* (1993), for example, proposed an input-output model by converting economic sales data into waste generation data. They defined the conversion factors for various economic activities and waste types based on material balance. Joosten, *et al.* (2000) and Patel, *et al.* (1998) conducted detailed materials flow analysis to investigate the plastics stream from production to waste generation. Input-output models are promising in terms of predicting generated waste amounts and composition with less effort compared to survey sampling. However, these models do not provide information about the determinants of consumption and waste generation process.

Additionally, there are studies which include determination of total waste generated based on specific waste generation rates and socio-economic variables for the purpose of planning and design of waste management systems. Karadimas and Loumos (2008) incorporated the area for different types of commercial activities in addition to population in GIS environment and estimated the waste generation and bin allocation accurately. The waste generation rate for each commercial activity type was determined by multiple regression analysis. The regression included the total area of each activity type as variables. Purcell and Magette (2009) similarly considered types of commercial activities while creating a GIS model to predict the generation of biodegradable MSW. However, for residential generation, they used two methods based on social class and household size and found that the latter was a better predictor. The MSW generation rates assigned to social classes, household sizes and commercial activity types were taken from the previous researches. Vijay, *et al.* (2005) conducted another study on waste collection making use of income and population density to estimate total waste generation precisely. These types of studies form a basis for the validation of the results of researches on significant determinants of solid waste generation rate. Achievement of high accuracy in predicting waste generated in these studies corroborates the significance of used variables as determinants of waste generation rate.

Another aspect of studies related to waste generation is behavioral and attitudinal analysis of the society. Results of those analyses can provide benefit in the prediction of solid waste generation and policy determination by enlightening the societal mechanisms between the waste generation and its determinants. The studies are carried out via questionnaires that are prepared to infer the respondents' concerns, attitudes and behaviors about solid waste issues. Correlation and regression analysis are generally included to interpret the questionnaire results. The differentiating feature of these studies is the variables used. As an instance, Ebreo, *et al.* (1999) investigated the links between environmentally responsible consumerism and psychological variables. They also included the socio-demographics and recycling behavior in the analysis. It was revealed that environmentally responsible consumerism was affected by socio-demographic variables such as age and gender. The higher concern for environment was found to be positively related to respondents' ratings for product attributes related to conservation and being kind to nature. Sterner and Bartelings (1999) conducted linear and non-linear regressions to analyze the determinants of waste disposal, composting and willingness to pay. The independent variables included behavioral, attitudinal, demographic, socio-economic and waste handling related variables. Composting kitchen waste, living area, age and difficulties in recycling were concluded to be significant determinants of waste generated. The study highlighted the importance of proper infrastructure and revealed that the economic incentive was not the sole driving force for MSW reduction. Godbey, *et al.* (1998) further did research on the effects of time use and availability of time on waste generation rate. They pointed out that waste generation could be affected by use of time due to its impact on habits. For example, time limits the willingness to repair broken appliances instead of discarding them and buying new ones. Recycling and composting behaviors as well as shopping behaviors and homemaking tasks are also dependent on perception of time. As seen, these researches focus on the waste generation mainly from a sociological point of view.

Recycling rate also attracts attention in these kinds of studies. Most of the aforementioned works try to include recycling in some way. Recycling may also be the sole focus of interest. For instance, Tsai (2008) focused on recycling and pointed out that recycling rate increased with increasing degree of social capital. Degree of social capital was represented by two indicators which were percentage of volunteers and number of social organizations. The relation was observed via a regression model. The model also included socio-economic variables as regional per capita income, percentage of employees with college degrees, funds allocated to environmental protection and community development, percentage of population below 14 years old and percentage of population above 65 years old. Income, education and population of elderly people were found to affect waste generation rate significantly in positive direction. It was also found that degree of social capital had significant affect when represented by percentage of volunteers.

Neither the techniques used nor the example studies are limited to the ones given here. The numbers of techniques can be increased through modification and combination of the above methods. Advancements in data handling and modeling also affect the methods in waste generation studies. Moreover, computerization led to improvements in these methods due to ability to handle advanced statistics (Chowdhury, 2009). Due to further improvements, sophisticated techniques can be utilized via available software.

Availability and reliability of data is essential independent of the method utilized. Most of the researchers conduct studies by collecting their own data via sampling in the study region. However, these kinds of studies require in depth sampling in the study area for which substantial amount of time and effort should be invested. Moreover, sampling should be carried out with caution so that the samples are representative of the whole population. If the extent of the study area is large including different administrative units, the data is generally compiled from institutions. For example, Beigl, *et al.* (2004) had to gather data from local representatives since the study covered 55 cities in 32 European countries.

Type of data determines the extent of the research. For example, seasonal variations cannot be observed by means of yearly data or aggregate generation data is not appropriate for composition studies. Moreover, the waste type investigated change from one study to another. Surveys at households target residential waste, while there are studies investigating both residential and commercial waste (e.g. Buenrostro, *et al.*, 2001; Karadimas & Loumos, 2008; Purcell & Magette, 2009). Besides the whole waste stream, the aim may be to examine a specific waste component such as paper waste (e.g. Bach, *et al.*, 2004), plastics (e.g. Patel, *et al.*, 1998; Joosten, *et al.*, 2000), organic fraction (e.g. Purcell & Magette, 2009), hazardous waste (e.g. Buenrostro, *et al.*, 2007; Buenrostro, *et al.*, 2008), etc.

As seen, the researches based on multivariate analyses differ from each other according to the independent variables involved. Demographic and socio-economic variables are used by a majority. Common demographic variables are population density, average household size and age structure. Socio-economic variables generally comprise income, educational level and occupational status. These may be denoted by many different indicators. As an instance, income can be represented by GDP at national level (e.g. Daskapoulos, *et al.*, 1998; Beigl, *et al.*, 2004) or annual rental rate of the property (e.g. Abu Qdais, *et al.*, 1997) instead of monthly income in terms of wage. Structural variables like unit-based pricing (e.g. Hockett, *et al.*, 1995; Hibiki & Shimane, 2006), density of collection sites (e.g. Bach, *et al.*, 2004) also appear in waste generation studies. Other variables involved are behavioral and climatic variables. Generally waste generation rate increases with increasing income and education level while it increases with decreasing average household size. Although these are the most common results, one can come up with different outcomes depending on the societal responses.

The results of each piece of work are unique to the region under study. The conditions and the mechanisms of waste generation vary between regions. This is especially important for factors, with which relationship of solid waste generation rate is vague. Although general rules are expected to hold, the effects of some independent variables may change among studies in different countries, even in different cities within a country. In other words, waste generation process varies not only temporally but also spatially. However, spatial variations in waste generation are rarely taken into consideration (e.g. Dennison, *et al.*, 1996) and the studies on waste generation do not employ spatial statistical techniques to analyze any possible spatial patterning.

2.2. STUDIES WITH SAR & GWR

SAR and GWR are two spatial regression techniques which incorporate spatial dependency. As a result, they are different from ordinary regression models which are generally put into practice while investigating solid waste generation as mentioned in the previous section. To include spatiality in waste generation investigation, SAR and GWR were used as regression techniques in this study. SAR is a global model since its results are valid for the whole study area. However, GWR explores the local variations and estimates the regression coefficients at local level.

Spatial regression techniques have been used in many disciplines among which econometrics, ecology and epidemiology stand out. While the real case applications are increasing in number, the improvement in theoretical background is still in progress.

Recently, Kissling and Carl (2008) tested the performance of different SAR models using artificial data sets by comparing model fit and residual spatial autocorrelation. They concluded that the model specification and type of spatial autocorrelation determine the performance of SAR model and for species distribution data they recommended spatial error model. This study is very helpful in understanding the mechanisms since it visually presents the data and error structure.

The real case applications of SAR generally take place in econometrics. Spatial econometrics is explained in detail by Anselin (1999) and Anselin (2002). For example, Yildirim, *et al.* (2009) and Yamanoğlu (2009) applied spatial error and lag models to investigate economic convergence in Turkey while Rey and Montouri (1999) analyzed convergence via these models in United States (US). Other examples may include investigations on employment growth (e.g. Shearmur, *et al.*, 2007) and welfare (e.g. Ayadi & Amara, 2008). Applications of SAR are not limited with economy. Other application areas include ecology, criminology and demography. SAR is used to investigate plant diversity. For example, Kreft and Jetz (2007) analyzed species-richness of vascular plants via SAR to examine the environmental and other possible determinants. In criminology, Baller, *et al.* (2001) used SAR to spatially analyze the factors affecting the homicide rates in US. As an example to usage of SAR in demographics, Sparks and Sparks (2009) investigated the relationship between mortality rates to socio-economic conditions using this model.

SAR was also used together with other autoregressive models to investigate the relation of lung cancer with SO_2 air pollution (Portnov, *et al.*, 2009). They found a positive relation between the areal density of lung cancer patients and SO_2 pollution levels. In fact, the number of studies spatially relating air pollution to human health conditions is noticeably high (Biggeri, *et al.*, 1996; Pope, *et al.*, 2002; Tağıl, 2007, etc.).

Another pollution related study making use of SAR lag model investigated the ground water NO_3 contamination with respect to land use types (Benson, *et al.*, 2006). The nitrate concentration was related to 14 land use types in different spatial aggregations. High nitrate concentrations in groundwater were associated with high percentage of potato, grain and hay coverage.

Applications of SAR model in environmental management problems are limited in spite of its convenience. This may be due to the complexity of the autoregressive models compared to other regression types. However, when it is about spatial dependency, autoregressive models are advantageous instruments to obtain reliable results and they are implemented widely in other disciplines. Details of SAR model and also other spatial data analysis techniques can be found in Bailey and Gatrell (1995) at introductory level.

GWR is a very recent technique and it is more user friendly compared to SAR due to its resemblance to ordinary regression formulation. GWR was introduced by Brunsdon, *et al.* (1998) for local modeling purposes. The technique was detailed by Fotheringham, *et al.* (2002). Brunsdon, *et al.* (1999) further developed a mixed GWR model. In mixed GWR, some of the variables may vary across the study area as in regular GWR while others may not which then have constant coefficient across the study area as in global models. Furthermore, after the introduction of GWR, it was subjected to conceptual investigations resulting in additional contributions (Leung, *et al.*, 2000; Wheeler, 2007; Wheeler & Tiefelsdorf, 2005).

GWR also took place in econometrics together with SAR. Shearmur, *et al.* (2007) investigated factors affecting the employment growth in Canada. Ayadi and Amara (2008) examined the local and global determinants of welfare in Tunesia. Yildirim, *et al.* (2009), on the other hand, employed GWR for the analysis of economic convergence and its determinants in Turkey. GWR further found place in studies regarding housing prices (e.g. Borst & Mccluskey, 2008; Cho, *et al.*, 2009). Cho, *et al.* (2009) tried to relate forest landscapes to housing prices while Borst and Mccluskey (2008) inspected the submarkets of housing.

Studies using GWR are increased with applications in different economic sectors. For example, Huang and Leung (2002) made use of GWR to analyze industrialization in China. They showed that there was a spatial variability in the relation of industrialization to various factors including GDP, percentage of urban labor, etc. On the other hand, Olgun and Erdoğan (2009) applied GWR to the wheat production in Eastern Anatolia to determine the effects of temperature, rainfall and humidity on crop yield. These effects were also found to be spatially varying. Both of these studies concluded that GWR outperformed ordinary least squares (OLS) regression. As it is seen, GWR can be applied in many areas to investigate the spatial variability of the relationship between the response and explanatory variables.

GWR also takes place in research related to environmental management. Even, it has a higher implementation potential compared to SAR. As mentioned, this may be due to its similarity to ordinary regression model except for spatially varying coefficients. Tu and Xia (2008), for example, examined the relation between water quality and land use indicators via GWR to determine local causes of water pollution. The model included one water quality parameter regressed on one land use indicator at each time and demonstrated spatially varying relation between land use type and water quality. Wentz and Gober (2007) inspected the factors affecting water consumption in addition to the water quality related studies (e.g. Wooldridge, *et al.*, 2006; Hudson, *et al.*, 2008; Tu & Xia, 2008). The water consumption was investigated in relation to household size, existence of swimming pools, lot size and prevalence of landscaping.

The examples of GWR application also exist in other areas of environmental management. For example, Mennis and Jordan (2005) applied GWR to air toxics release in an environmental equity investigation. The study investigated the spatially varying relationship between air toxic release and socio-economic factors including race, employment, etc. They indicated that relation between the air toxics release and a socio-economic variable can be mediated by another factor. Another example of GWR application was presented by Harms, *et al.* (2009) who examined denitrification in desert floodplains. The denitrification pattern in riparian systems was analyzed based on the characteristics of soil as well as characteristics of hydrologic vectors. Similar to SAR, GWR was also applied in ecological studies to investigate species richness (e.g. Svenning, *et al.*, 2009; Latimer, *et al.*, 2009; Terribile & Diniz-Filho, 2009; Keil, *et al.*, 2008).

GWR was implemented in a wide range of subjects in environment management especially compared to SAR. In fact, this may be perceived as an indication of prevalence of spatial non-stationarity in environmental data. At least, the studies given as examples above demonstrated the non-stationarity in their own data. However, the applications of both SAR and GWR are still related to limited subjects in environmental management. The spatial analysis in waste management does not go beyond exploration of waste generation and its determinants through geographical stratification and mapping for comparison purposes. However, waste generation is an appropriate concept to be analyzed in terms of non-stationarity considering the variety of possible factors of waste generation and complex dynamics lying behind these.

CHAPTER 3

METHODS

3.1. GENERAL FRAMEWORK OF METHODOLOGY

A brief description of the methodology will be given in this section. The flowchart of the methodology followed is given in Figure 1. The first and crucial step of this work is data collection. The data on both solid waste generation and socio-economic, demographic and climatic indicators were compiled in this step. The compiled data was observed through visualization techniques. Socio-economic, demographic and climatic variables were then subjected to multicollinearity analysis. This analysis ensures that these variables are independent from each other so that they can be input in the regression analyses.

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 MSW generation rate was primarily inspected. Then, SAR and GWR were employed, respectively. These regression techniques take spatial dependency into account. SAR gives global coefficients similar to OLS and local coefficients are obtained by applying GWR. Following GWR, OLS regression was again carried out among the provinces which are not accepted as neighbors to each other for the purpose of eliminating neighborhood effect. Finally, the estimated models were applied to a different data set for validation purposes.

3.2. DATA COLLECTION

Data collection may be the most problematic part of this type of studies. Gathering all necessary data and maintaining data reliability are essential although tedious. In

Turkey, a major source of data for a variety of indicators is TurkStat and most of the data is open to public on its website. TurkStat publishes data according to Classification of Statistical Regions (SRE). There are three levels in this classification system. Level 3 refers to 81 provinces in Turkey while Level 1 and Level 2 denote regions and sub-regions, respectively, obtained considering neighboring provinces and socio-economic and geographic similarities between them. Data for this study was compiled in terms of Level 3.

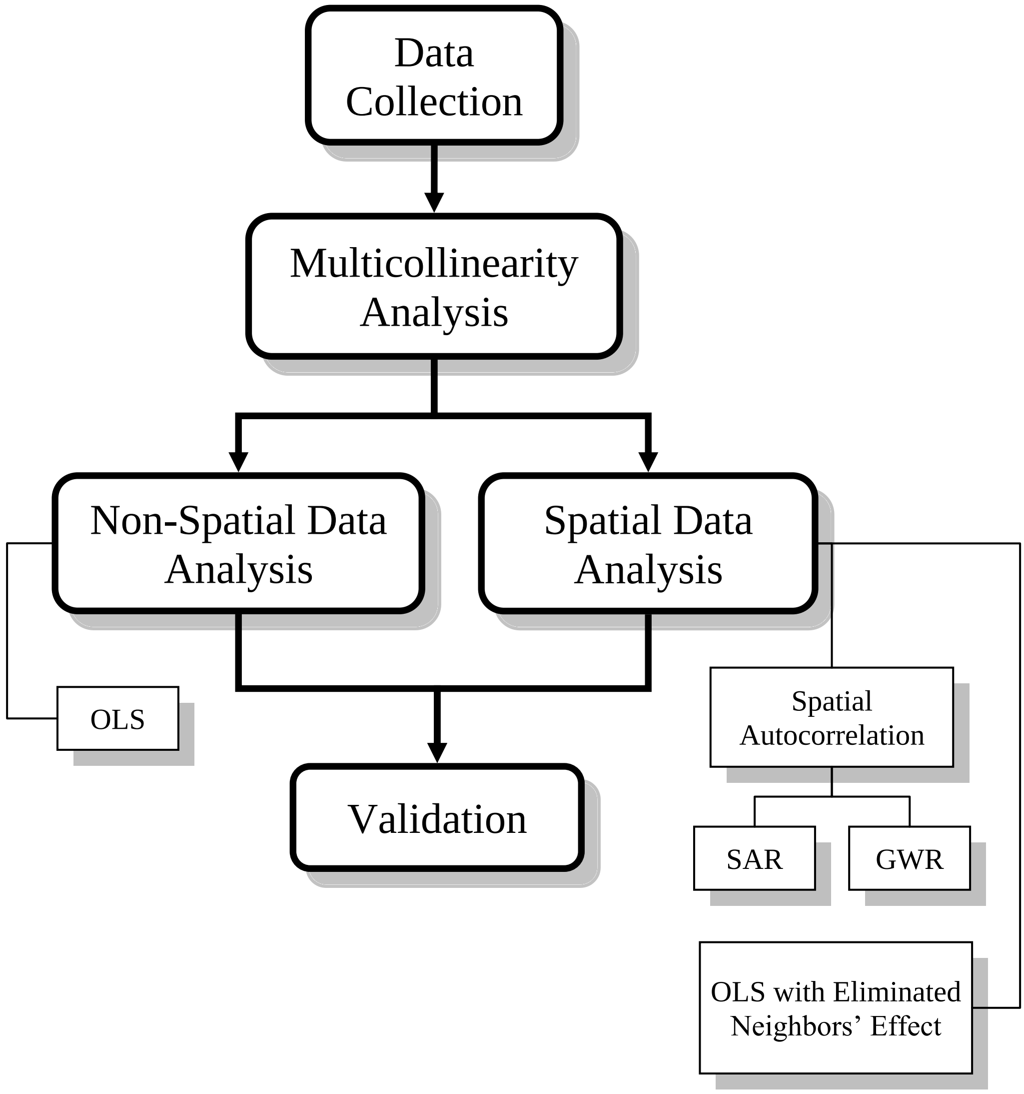


Figure 1: Flowchart of methodology

Solid waste generation rate was used as the dependent (response) variable of the regression analyses. Data on this variable was supplied from TurkStat (2009a). TurkStat has conducted surveys in all municipalities which have been quantifying their own data on yearly basis since 1994. However, there are missing data for some years due to technical restrictions and data is available only for the years 1994-1998, 2001-2004 and 2006. As a result, an interpolation was carried out to obtain MSW generation rate data for the year 2000 to bring the dependent variable into consonance with the explanatory variables in terms of time. In this method, a linear model was fit to the available MSW generation rates to obtain linear equations for each province which were then used to calculate MSW generation rate of each province for 2000. These were carried out by using MS Excel 2007. Data on a variety of socio-economic and demographic variables, on the other hand, are made available online by TurkStat only for the years that a census was compiled. Climatic variables are available on monthly basis by the Turkish State Meteorological Service (TSMS) for every year; however, these are not provided online.

Explanatory (predictor or independent) variables were composed of socio-economic, demographic and climatic variables. Although the socio-economic and demographic ones were mainly aimed to be investigated as the determinants of solid waste generation, temperature and rainfall were also included as environmental factors. Considered socio-economic and demographic variables are as follows: urbanization, population density, average household size, employed people in agricultural sector, employed people in industry sector, employed people in trade sector, literate women ratio, higher education graduates ratio, infant mortality rate, number of facilities in small Organized Industrial Districts (OIDs), number of manufacturing facilities, agricultural production value, number of dwellings, GDP per capita, asphalt-paved road ratio in rural areas, number of automobiles per 10000 people, number of motor vehicles per 10000 people and unemployment. Details of these variables will be given in Section 4.1.

The socio-economic and demographic explanatory variables were mostly compiled from a study conducted by State Planning Organization (SPO) of Turkey (Dinçer, *et al.*, 2003) in which diverse range of indicators was used to determine the socio-economic development ranking of provinces. The authors took most of them from the results of 2000 census compiled by TurkStat and collected the other variables from related institutions for the same year. Among these variables, the ones that were likely to be related to solid waste generation were selected based on background knowledge and previous studies. Another source for socio-economic and demographic variables was TurkStat website (TurkStat, 2009b) which provides census data. Temperature and rainfall data were supplied from TSMS. The data was gathered together in a MS Excel spreadsheet to be ready for multicollinearity analysis.

MapInfo Professional 7.5 software was used to create choropleth maps. Choropleth maps or thematic maps are means of displaying area data obtained by coloring or shading the areal units in accordance with their attribute values of interest (Bailey & Gatrell, 1995). ScapeToad-v11 was also used to obtain a cartogram of Turkey based on MSW generation rates attributed to the provinces (ScapeToad, 2008). Cartogram is a kind of choropleth map which, differently, rescales the areal unit according to the attribute and colors the polygons based on the size error. Calculation of size error is given in Appendix A.1. Expansions in a cartogram refer to high attribute values while contractions mean low attribute values in relative to the area of the polygon. The purpose of using a cartogram is to avoid misperception due to large areal units and to have a better visualization. The normality assumption of the variables was checked by obtaining Q-Q plots which compares the quantiles of a data set with a normal theoretical quantiles.

Another data set should be created which belongs to a different year for the validation step. This year was chosen as 1990 in which the previous census was taken. MSW generation rate for that year was derived by extrapolation using available data (TurkStat, 2009a) as done for the data of 2000. Higher education ratio and unemployment rate were calculated using the data on number of graduates and unemployed, respectively, obtained from TurkStat (2009c). Meteorological data was again supplied from TSMS. Other variables were compiled from another study of SPO (Dinçer, *et al.*, 1996). However, agricultural production value had to be corrected due to monetary depreciation. This correction was made based on exchange rate of American Dollar as given in Appendix A.2.

3.3. MULTICOLLINEARITY ANALYSIS

Among the explanatory variables compiled as given in the previous step, an elimination had to be carried out due to multicollinearity or in other words to satisfy the independency of the variables. Multicollinearity[1] exists when one of the explanatory variables has a linear relationship to another explanatory variable or to 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 (R^2) 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 (e.g. sample numbers is smaller than or equal to the number of explanatory variables). In case of perfect multicollinearity among explanatory variables, the regression analyses cannot be performed. Although multicollinearity is not an impediment to perform the analyses as long as it is not perfect, it still has some consequences that are determined by the extent of the correlation among the explanatory variables (Berry & Feldman, 1985).

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 & Shevlin, 2001), multicollinearity makes it hard to interpret the results of the analyses and it should be avoided as much as possible.

Multicollinearity can be caused by including several explanatory variables which are the indicators of the same phenomena or including an explanatory variable created by other explanatory variables which also take place in the equation. This may not be

[1] Multicollinearity and collinearity are generally used interchangeably. The former is preferred in this document.

easily detected in some cases; however there are symptoms that reveal multicollinearity. The sign of the coefficient estimates for explanatory variables may be implausible or a variable may be found to be insignificant in spite of its expected important effect on the dependent variable. High sensitiveness of the results to slight changes, such as excluding a variable or a sample, may also be an indication of multicollinearity. However, multicollinearity does not always result in these and there may be other motivations of these symptoms (Belsley, *et al.*, 1980). As a result, more concrete tests are conducted in this study.

Variance inflation factor (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.

VIF is calculated using the coefficient of determination (R_k^2) of the regression equation in which an explanatory variable denoted by *k* becomes the dependent variable while others are independent variables. The formula to calculate VIF is as follows (Belsley, *et al.*, 1980):

$$VIF_k = \frac{1}{1 - {R_k}^2} \tag{1}$$

As it is seen in the formula, VIF increases when R_k^2 or the correlation of one variable with others increases. Therefore, multicollinearity is indicated by high VIF. Statistically, square root of VIF reveals how much the standard errors are multipled as a result of multicollinearity (Miles & Shevlin, 2001). There is a wide range of cut-off VIFs one can come across in the literature. One mostly used threshold value for VIF is 10 over which multicollinearity affects the coefficient estimations seriously (Freund & Wilson, 1998). Another cut-off value often used is 4 which has a theoretical basis. Since square root of this value is 2, standard errors are doubled at this point and it is accepted that multicollinearity affects results beyond this value (Miles & Shevlin, 2001). Furthermore, Cho, *et al.* (2009) checks the existence of multicollinearity using the cut-off value of 10 for the highest VIF and also ensures that the mean of all VIFs is not substantially greater than 1. Although, threshold for VIF against multicollinearity varies in the literature, the general acceptance is that VIF should not exceed $1/(1\text{-}R^2)$ where R^2 is the coefficient of determination of the whole model (Freund & Wilson, 1998; Park, 2003).

The formula for computing r is obtained by dividing the covariance of the variables by the product of their standard deviation and given as follows:

$$r = \frac{\frac{\sum\left(X_{k_1} - \bar{X}_{k_1}\right)(X_{k_2} - \bar{X}_{k_2})}{n}}{\sqrt{\frac{\sum(X_{k_1} - \bar{X}_{k_1})^2}{n}}\sqrt{\frac{\sum(X_{k_2} - \bar{X}_{k_2})^2}{n}}} \tag{2}$$

where X_{k_1} and X_{k_2} are (nx1) vectors of any two explanatory variables with the means $\bar{X}_{k_1}$ and $\bar{X}_{k_2}$, respectively and n is the number of observation points or samples (Walford, 1995). r is obtained as a symmetric square matrix when there are more than two variables. By inspecting this matrix, the pairs of explanatory variables that are highly correlated can be detected easily. r ranges between -1 and +1. A value around zero indicates that there is no correlation between variables. The correlation emerges when r diverges from zero and approaches to 1 (or -1). The values close to 1 and -1 are the indicators of high correlation. If r has a minus sign, the variables are negatively related to each other. In other words, while one of the variables increases, the other one decreases. If r is positive, indicating positive correlation, the variables increase or decrease together.

A limit value for r should be determined for pair wise elimination of the coefficients. Generally, this predetermined value below which multicollinearity does not pose a problem is 0.80; however, this may not be always appropriate for all data sets depending on the sample size (Berry & Feldman, 1985) and therefore threshold value for r varies from one study to another. Harms, *et al.* (2009) does not include variables with r above 0.40 in the same model. Nevertheless, Mennis and Jordan (2005) did not discard any of the variables in spite of r values ranging between 0.50

and 0.60, but took these into consideration during statistical interpretation. Banar and Özkan (2008) decided that there is a good correlation between variables when *r* is over 0.70.

r gives us a hint about the shared variance. The square of this coefficient is the well-known statistic: R^2. By this way, we can learn about the variance in a variable shared with other variables and so about the redundant information (Miles & Shevlin, 2001; Walford, 1995). As an instance, the R^2 is around 0.64 if *r* is 0.80 for any two variables. This means that more than half of the variation in one variable can be explained by the variation in the other variable. This is why it is unnecessary and inaccurate to include two variables with high correlation in the same model.

The “car” package of R can be used to obtain VIFs and *r* can be calculated under “stats” package. Various data groups can be obtained by setting different cut-off values for VIF and *r*. Then, these data groups comprise different levels of multicollinearity. The package “leaps” is utilized to select the subset of variables of any length with the highest R^2. The function uses a branch-and-bound integer optimization algorithm to determine the best subset of variables. Since no consideration is given to multicollinearity in this selection, the packages should be used in conjunction with each other.

The remaining variables after multicollinearity analysis should be added to a shapefile for spatial analysis. 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.

3.4. NON-SPATIAL DATA ANALYSIS

An ordinary multiple regression comprises the non-spatial data analysis in this study. This type of regression is the simplest case for multivariate analysis and it is the one we are more familiar with. The formula for a multiple linear regression is given in matrix notation as follows (Bailey & Gatrell, 1995):

$$Y = X\beta + \epsilon \tag{3}$$

where Y (nx1) is the vector of dependent variable and X (nx(k+1)) is the matrix of independent variables with 1s in the first column. β ((k+1)x1) is a vector of regression coefficients or partial slopes and ϵ (nx1) is the error vector. n is the number of observation locations and (k+1) is the number of β coefficients to be estimated including the intercept.

The β coefficients are estimated by OLS method. The method involves choosing the coefficients resulting in the minimum sum of squared residuals (Bailey & Gatrell, 1995). Residual is the difference between the observed data value and the value predicted by the model. Sum of squared residuals is known as the unexplained sum of squares. R^2 is the ratio of explained sum of squares to the total sum of squares which is a measure of how well the model explains the variation in the response variable (Wesolowsky, 1976). The significance of R^2 is assessed by checking the p-value which should be smaller than the level of significance (α) to prove that the probability of making a Type I error is very small. Type I error is the error in hypothesis testing that occurs when the null hypothesis is rejected although it is correct (Griffith & Amrhein, 1997). Akaike Information Criterion (AIC) is another coefficient used in order to test the level of fit of the models. The model with a lower AIC is favored in terms of fitting the data. The formula for AIC calculation can be found in Appendix A.3.

T-statistics are estimated to test the significance of regression coefficient estimates and t-score is the ratio of the estimate to its standard error (Ge, 2009). Although the sign of the t-score determines the direction of the effect of the related variable, the absolute value of the t-score is important in terms of the magnitude of the effect. This is why a two-tailed test is performed for detecting associated significance of t-score.

OLS regression coefficient estimates given by R are also accompanied by their t-scores and associated p-values.

The variation in the mean value of a process in the study area is associated with the first order effects, while the spatial dependency of the deviations from that mean is attributed to the second order effects. Assumption of this type of regression (OLS) is that there are only first order effects in the process and so it does not model the second order variation. The results of the model regarding β coefficient estimates turn out to be invalid if second order effects or local effects are present. However, this type of effect is often encountered with areal data. One way to observe this effect is testing the error term (residuals of the model) against the spatial dependency which is particularly explained in the next section.

3.5. SPATIAL DATA ANALYSIS

3.5.1. Spatial Autocorrelation

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 (Gangodagamage, *et al.*, 2008).

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 (Zhou & Lin, 2008). Moran's I is calculated as follows:

$$I = \frac{n \sum_i \sum_{j \neq i} w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_i \sum_{j \neq i} w_{ij}) \sum_i (Y_i - \bar{Y})^2} \qquad (4)$$

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). $\bar{X}$ is the mean of the dependent variable. w_{ij} 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.

Spatial weights are created prior to spatial data analysis to be able to define proximity. There are many ways to construct the spatial proximity matrix based on the definition of the spatial neighbor and the weights assigned to these neighbors (Bivand, *et al.*, 2008). Being neighbors for areal units may depend upon the distance and contiguity. The distance can be determined based on the centroids of the areal units. Sharing a boundary is the criterion for being neighbors based on contiguity. The types of neighbors may be diversified depending on how to use the information on the distance and contiguity. Moreover, the matrices differ in terms of the weights assigned to the defined neighbors. The weight can be defined as a function of the distance between centers. The matrix can also be formed in a binary system. In a binary matrix, the elements are composed of 1s and 0s. Another type of matrix is obtained via row standardization. Row-standardized matrix is obtained by applying (Anselin, 2002):

$$w_{ij}^s = \frac{w_{ij}}{\sum_j w_{ij}} \qquad (5)$$

where w_{ij} is the element of the weight matrix based on binary system. In a row-standardized matrix, the weights are arranged so that the elements in a row add up to unity. In the end, regardless of the method, the element of the matrix takes the value of zero if it is on the diagonal or the neighboring requirement is not covered. For example, the contiguity matrix involves non-zero weights when two provinces share a boundary. In the case of the distance based matrix, nearest neighbors of a province are given non-zero weights based on the distance between centroids of polygons. Diagonal elements are zero to make sure that an areal entity is not a neighbor of itself. In the row-standardized contiguity matrix, it is expected that weights vary among the rows since number of neighbors may differ from one province 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.

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.). An example for the neighbors of Ankara is given in Figure 2.

Defining spatial weights or creating weighting matrices is very essential in spatial data analysis because it is how we can incorporate the spatiality into the models. The way how the matrix is created is important due to its power to affect the results. The matrix is used not only in Moran's I calculations but also in spatial regression analyses. Ultimately, the aim is to contribute to obtaining the proper model for the spatial process rather than demonstrating autocorrelation (Bivand, *et al.*, 2008). The autocorrelation remaining in the residuals can be also observed by calculating Moran's I with the weight matrix.

The "spdep" package of R involves commands for the Moran's I calculation and also for the construction of the weighting matrices importing shapefile and database file. Moran's I values calculated can be tested for their significance under the null hypothesis of randomization. "spdep" obtains the standard deviate of Moran's I statistic and compares it with the Normal distribution to obtain p-value (Bivand, *et al.*, 2008).

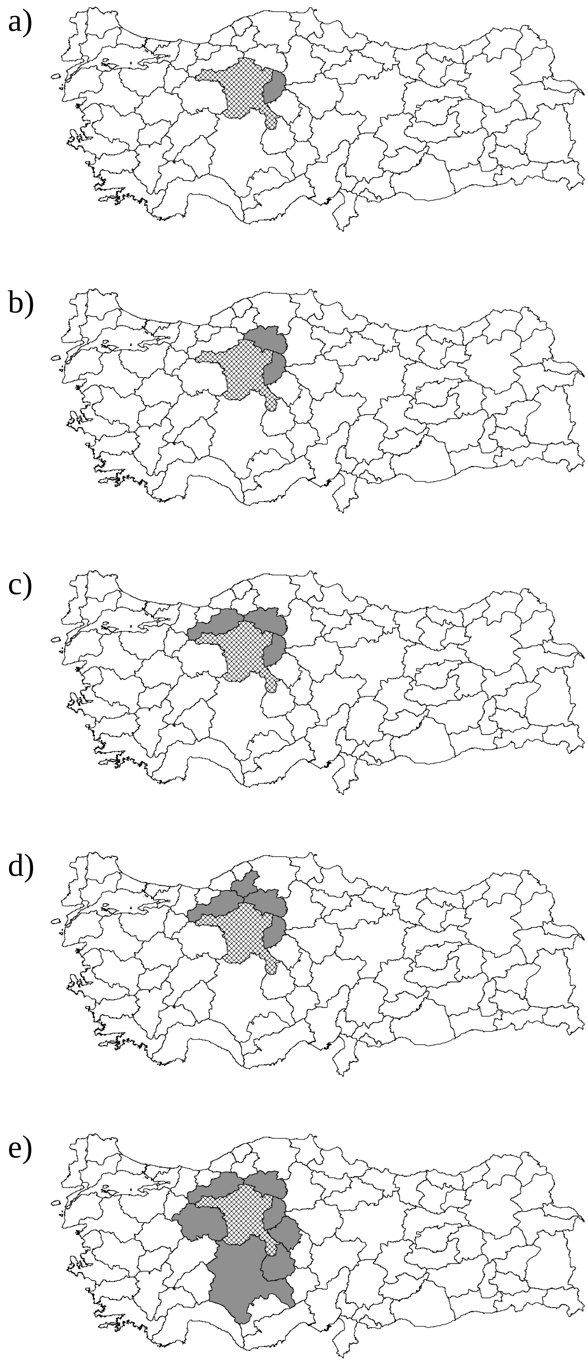


Figure 2: Neighbors of Ankara based on **(a)** 1st nearest neighbors (lag 1); **(b)** 2nd nearest neighbors (lag 2); **(c)** 3rd nearest neighbors (lag 3); **(d)** 4th nearest neighbors (lag 4); **(e)** contiguity

Another way of testing significance of Moran's I is the Monte Carlo test. In a Monte Carlo test, the samples are randomly selected among the possible arrangements developed by assigning the attribute values to different areal entity each time. Then, the Moran's I values for each case are ranked and the p-value is calculated according to the rank of the observed Moran's I.

3.5.2. Spatial Regression

The data lacks independency in case of presence of spatial autocorrelation. As a result, the prediction of the variable at one observation location is affected by the value of the variable at the nearby locations. This poses problems in standard statistical tests. In order to overcome these problems, either the autocorrelation structure of the data is removed prior to the analysis or it is incorporated into the statistical method (Legendre, 1993). SAR and GWR are models taking the spatial dependency into consideration satisfying the latter condition.

3.5.2.1. Simultaneous 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 model were investigated in this study: lag model (SAR_{lag}) and error model (SAR_{err}).

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

$$Y = X\beta + \rho WY + \epsilon \tag{6}$$

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, SAR_{lag} 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 & 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):

$$Y = X\beta \ + \rho WY \ + \ WX\gamma \ + \ \epsilon \tag{7}$$

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), SAR_{err} model is attained (Anselin, 1999). In the SAR_{err} 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 & Carl, 2008). SAR_{err} is formulated as (Bailey & Gatrell, 1995):

$$Y = X\beta + U \tag{7}$$

$$U = \lambda WU + \epsilon \tag{8}$$

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

$$Y = X\beta + \lambda WY - \lambda WX\beta + \epsilon \tag{9}$$

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 & Gatrell, 1995). As it is seen, SAR_{err} can be obtained from the spatial Durbin model by putting common factor constraint on the coefficient of spatially lagged explanatory variables such that $\gamma=-\lambda\beta$.

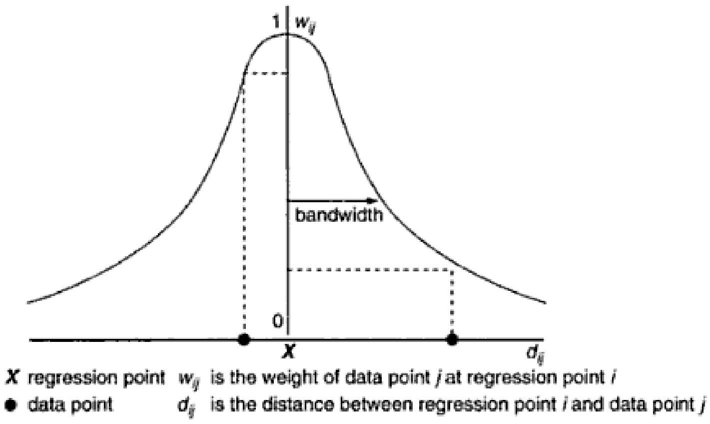
Simultaneous autoregressive coefficients (interaction parameters: ρ and λ) quantify the effect of neighboring observations and also they determine the direction of that effect (Düzgün & Kemeç, 2008). These are additional parameters to be estimated compared to the non-spatial regression model which only estimates the regression coefficients (β). Therefore, SAR models should estimate not only β but also interaction parameters which is a computationally intensive procedure. The models are fit by maximum likelihood in 'spdep' package (Bivand, *et al.*, 2008). p-value for R^2 can be calculated via F-distribution as given in Appendix A.4.

The means of generation of spatial autocorrelation determines which type of SAR model should be used. Under 'spdep', Lagrange multiplier (LM) tests can be conducted to choose between the SAR models. In these tests, LM_{err} and LM_{lag} are calculated to test against the null hypotheses of λ=0 and ρ=0, respectively. Robust LM_{err} (RLM_{err}) and robust LM_{lag} (RLM_{lag}) were also calculated to test against these null hypotheses but also considering non-zero ρ and λ, respectively (i.e. H_o: $\lambda=0$ assuming $\rho\neq0$ for RLM_{err} and H_o: $\rho=0$ assuming $\lambda\neq0$ for RLM_{lag}).

3.5.2.2. Geographically Weighted Regression (GWR)

The β coefficients estimated by OLS and SAR are valid for all provinces; 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 across the study area. GWR reveals these local relationships by moving a spatial kernel across the study area. A representation of kernel is given in Figure 3. The center of the kernel is located on the regression points (i). 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).

a)



b)

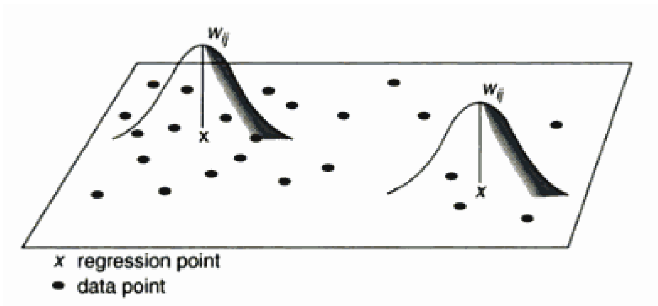


Figure 3: (a) A spatial kernel; **(b)** GWR with fixed spatial kernels (Fotheringham, *et al.*, 2002)

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

$$y_i = \beta_o(u_i, v_i) + \sum_k \beta_k(u_i, v_i)\, x_{ik} + \epsilon_i \tag{10}$$

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). Detailed model in matrix notation is given in Appendix A.5.

There is a package called “spgwr” contributing to R for GWR analysis. Using this package, the bandwidth of the kernel is optimized by cross-validation before computing GWR (see Appendix A.6). The kernel function can be chosen as bi-square or Gaussian. The kernel of Gaussian function distributes weights according to (Fotheringham, *et al.*, 2002):

$$w_{ij} = exp\left[-{}^{1}/_{2}\left(d_{ij}/b\right)^{2}\right] \tag{11}$$

where j is the data point, d_{ij} is its distance from the regression point i and b is bandwidth. 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):

$$w_{ij} = \begin{cases} \left[1-\left(d_{ij}/b\right)^{2}\right]^{2}, & d_{ij} < b \\ 0 \qquad\qquad\quad , & otherwise \end{cases} \tag{12}$$

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.

As a result of GWR, local coefficients for each province 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. The fit of GWR can be assessed by AIC. Moreover, local R^2 for each observation point is given along with the local coefficients.

3.5.3. OLS Regression with Eliminated Neighbors' Effect

Response variable associated with an areal unit is affected by the values of that variable in the nearby locations when there is spatial autocorrelation as explained in the previous sections. A subset of data locations can be taken for OLS regression so that they are not within each other's neighbors to eliminate the neighboring effect. The weighting matrix is utilized to eliminate the neighbors (i.e. pairs with non-zero weights) to obtain this subset. This is why this analysis is included together with spatial methods although the method itself (OLS regression) is not spatial.

The results can be compared with the analysis carried out with the whole population by observing the fit of the model and residual spatial autocorrelation. However, number of samples (i.e. data points) included in the regression is decreased although independent variables are the same. As a result, to compare the results, adjusted R^2 (R^2_{adj}) should be obtained using the formula given in Appendix A.7. R^2_{adj} enables the interpretation of results for the models of differing number of observation and/or independent variable. R^2 may be very high although there is not any linear relationship between the dependent and independent variables when the sample size is small compared to the number of independent variables (Wesolowsky, 1976).

3.6. VALIDATION

In the validation step, β coefficient estimates are tested with a data set belonging to a different year. The response variable is estimated with OLS coefficient estimates as follows:

$$Y' = Xb_{OLS} \tag{13}$$

where Y' is a vector of predicted MSW generation rates and X is the matrix (nx(k+1)) of variables compiled for the year 1990 with 1s in the first column while b_{OLS} is a vector ((k+1)x1) of OLS coefficient estimates. Rearrangement of SAR_{lag} model to estimate MSW generation rate is more complex since the dependent variable is found on both sides of the equation. SAR_{lag} model for validation is obtained via matrix operations as:

$$Y' = (\mathbf{I} - \rho W)^{-1} X b_{SAR} \tag{14}$$

where b_{SAR} is a vector ((k+1)x1) of SAR_{lag} coefficient estimates and $\mathbf{I}$ is an identity matrix (nxn). GWR model is similar to OLS equation. However, the coefficients are not the same for data points. The predicted values of the response variable using estimated GWR coefficients are found by first applying:

$$Y' = (X \otimes B_{GWR})1 \tag{15}$$

where B_{GWR} is a matrix (nx(k+1)) of estimated GWR coefficients and '$\otimes$' is an operator that computes element by element multiplication of matrices. 1 is a vector of 1s ((k+1)x1). The row sums of the matrix Y' (nx(k+1)) is then computed to obtain the vector Y (nx1) of predicted response variable.

The results can be analyzed by examining the correlation between the observed and predicted values which is accomplished by calculating r. Moreover, the root mean squared error (RMSE) should be assessed which is obtained by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{16}$$

where $\hat{y}_i$ is the estimator of y_i, and y_i is the observed response variable at location i.

CHAPTER 4

IMPLEMENTATION OF METHODOLOGY & RESULTS

4.1. DATA COLLECTION

According to the definition given by TurkStat, MSW is the residential, commercial and institutional waste collected by or on behalf of the municipalities also including the wastes originating from parks, gardens, market places and streets. Solid waste generation rates are in terms of kilograms of MSW daily generated per capita. The cartogram and choropleth map for MSW generation rate in Turkey[1] for the year 2000 are given in Figure 4. In Figure 4a, it is seen that low values are clustered in the south-east of the country and the solid waste generation rate tends to increase from east towards west. Moreover, in Figure 4b, the contractions of the areas mostly occur in the eastern part while expansions take place in the western and central parts. The former is caused by low values with respect to the area of the province and the latter is due to the high values considering the areal coverage. If we assume that the area is regularly distributed between provinces in Turkey, it can be said that the maps for MSW generation rate in Figure 4 are consistent with each other both indicating an increase in MSW generation rate in the east to west direction. In cartogram, it is seen that some areas are expanded and others are contracted. For example, the area for Yalova is enlarged while the area of Konya is reduced. By this way, overlooking the MSW generation rate at Yalova due to its smaller real area is avoided and it is prevented for Konya to look as if it has a very high waste generation rate due to its large area.

The cartogram report which details the results is given in Appendix A.8. The minimum and the maximum values for MSW generation rate are 0.27 kg/d-ca (Hakkari) and 3.25 kg/d-ca (Muğla), respectively. Average MSW generation rate is

[1] Turkey map with province names is given in Figure C.1 (Appendix C).

1.33 kg/d-ca in 2000 which is comparable to the average MSW generation rate of 1.43 kg/d-ca in EU countries for the same year (EEA, 2009). The observed MSW generation rates for all provinces belonging to 1990 and 2000 are given in Table D.1 (Appendix D).

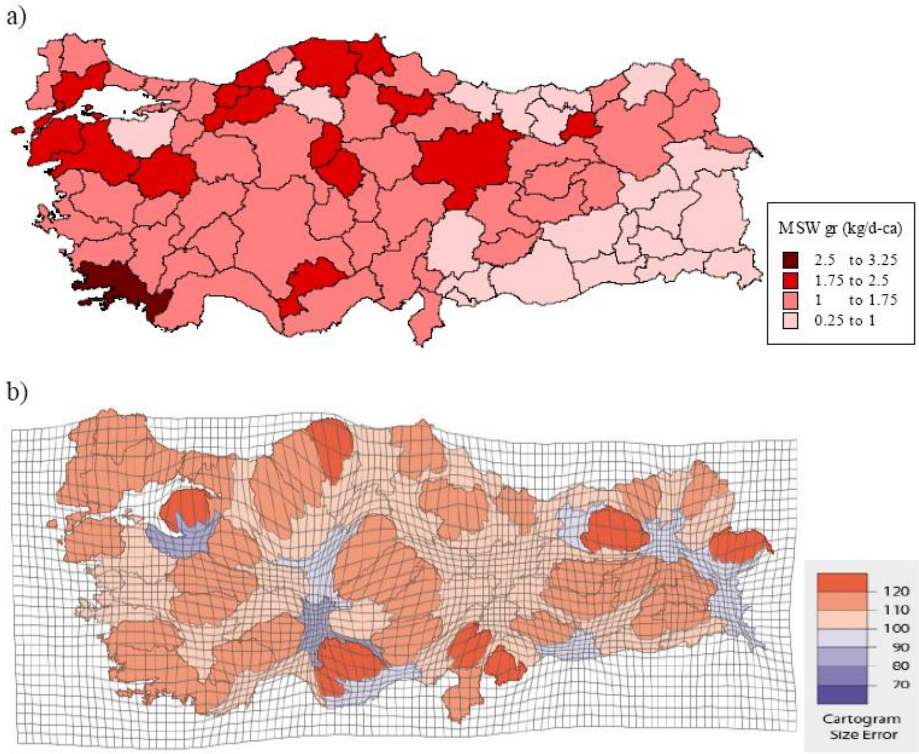


Figure 4: (a) The choropleth map and **(b)** cartogram of MSW generation rate in Turkey for the year 2000

Education level, income, employment type and average household size are mostly used variables in studies regarding solid waste generation. Bearing these in mind, the variables were selected among the ones used in Dinçer, *et al.* (2003). Since there was not a direct indicator of income, unemployment was additionally included as an explanatory variable. Climatic variables were also incorporated to include environmental indicators in the study. As a result, twenty explanatory variables were compiled among which socio-economic variables are in majority as seen in Table 1. The summary statistics of explanatory variables are given in Table D.2 (Appendix D).

Urbanization is one of the demographic variables used and it is defined as the percentage of the population living in the urban areas within the total population. Urban areas include province and district centers. Population density is a very well known demographic variable which is the population per one square kilometer. Average household size is another demographic variable which is the average number of people living in a household calculated by dividing total household population to the number of households.

Table 1: Explanatory variables

ID	Source	Variable name	Unit
1	SPO[i]	Urbanization	%
2	SPO	Population density	person/km^2
3	SPO	Average household size	person
4	SPO	Employed people in agricultural sector	%
5	SPO	Employed people in industry sector	%
6	SPO	Employed people in trade sector	%
7	SPO	Literate women ratio	%
8	SPO	Higher education graduates ratio	%
9	SPO	Infant mortality rate	‰
10	SPO	Number of facilities in small OIDs	number
11	SPO	Number of manufacturing facilities	number
12	SPO	Agricultural production value	million TL
13	SPO	Number of dwellings	number
14	SPO	GDP per capita	million TL
15	SPO	Asphalt-paved road ratio in rural areas	%
16	SPO	Number of automobiles per 10000 people	number
17	SPO	Number of motor vehicles per 10000 people	number
18	TurkStat	Unemployment	%
19	TSMS	Temperature	^oC
20	TSMS	Rainfall	mm

[i] SPO refers to Dinçer, *et al.*, 2003

The variables with ID number of 4, 5 and 6 are the percentages of the employed people in the agriculture, industry and trade sectors, respectively within the total labor force. These are indicators regarding employment in addition to unemployment. Unemployment is the percentage of unemployed people within the labor force. TurkStat (2009b) defines the unemployed people as follows:

> *The unemployed comprises all people 15 years of age and over who were not employed (neither worked for profit, payment in kind or family gain at any job even for one hour, who have no job attachment) during the reference period who have used at least one channels for seeking a job during the last three months and were available to start work within two weeks.*

There are two educational indicators. These are literate women ratio and the ratio of higher education graduates. The rationale behind the calculation of these two variables is similar. The former refers to the percentage of literate women within the women population aged 6 and over. The latter is the percentage of higher education graduates within the total population aged 22 and over. Higher education refers to any higher education of at least 2 years following secondary education. Theoretical age for this education is accepted as between 17 and 21 (TurkStat, 2009a).

IMR is the infant deaths per 1000 live births. Although this variable is a health indicator in its nature, it is a widely used indication of development. This is why Beigl, *et al.* (2004) introduced IMR as a variable in solid waste generation research.

Number of facilities in small OIDs and number of manufacturing facilities are industrial indicators. Small OIDS gathers small enterprises of tradesmen and craftsmen together (MoIT, 2009). In these variables, the number of all facilities operating in public sector and the ones in private sector with 10 or more employees are taken into consideration.

Agricultural production value is an agricultural indicator and given per rural population. The vegetal and animal products and livestock quantities are taken into consideration while determining the production value. Rural population is the population living in the rural areas which cover the settings other than province and district centers.

Number of dwellings is explicit by its name but its expected effect on solid waste generation is implicit. This variable shows alterations with increasing population, migration and urbanization due to social and economic development (Dinçer, *et al.*, 2003). In a way, it can also be linked to construction wastes.

GDP is the total value of produced goods and services. GDP per capita is a widely used variable reflecting economic development. It is important in this work because of its direct relation to consumption expenditures. Number of automobiles and motor vehicles per 10000 people are accepted as indicators of income and prosperity level (Dinçer, *et al.*, 2003) which are very suitable to assess economic development.

As an indicator related to infrastructure, asphalt-paved road ratio in rural areas was included to include the availability of access to rural areas by collection vehicles as a variable. This socio-economic variable can be considered as a structural variable in a solid waste generation study.

Temperature is the average of monthly temperature values for 2000 and rainfall is the sum of monthly precipitation giving annual precipitation for that year.

Choropleth maps for explanatory variables are presented in Figure 5. Similar to MSW generation rate, variation in explanatory variables tend to be in the east to west direction. This result is in line with the distribution of developed provinces in Turkey. The degree of development or development indexes of the provinces also change in the east-west direction (Dinçer, *et al.*, 2003). Temperature and rainfall are exceptions for this affinity since they are environmental variables. The Q-Q plots are given in Figure C.2 (Appendix C). It is seen that the normality holds around the mean (i.e. within one standard deviation) for all variables.

In the data set for validation, all variables belong to 1990 except for agricultural production value and asphalt-paved road ratio which were available for 1993 and 1994 in Dinçer, *et al.* (1996). The variable regarding facilities in small OIDs was unavailable. Moreover, there are administrative units which became provinces in 1990s. For these, the socio-economic and demographic variables were taken the same as the provinces they previously belonged to.

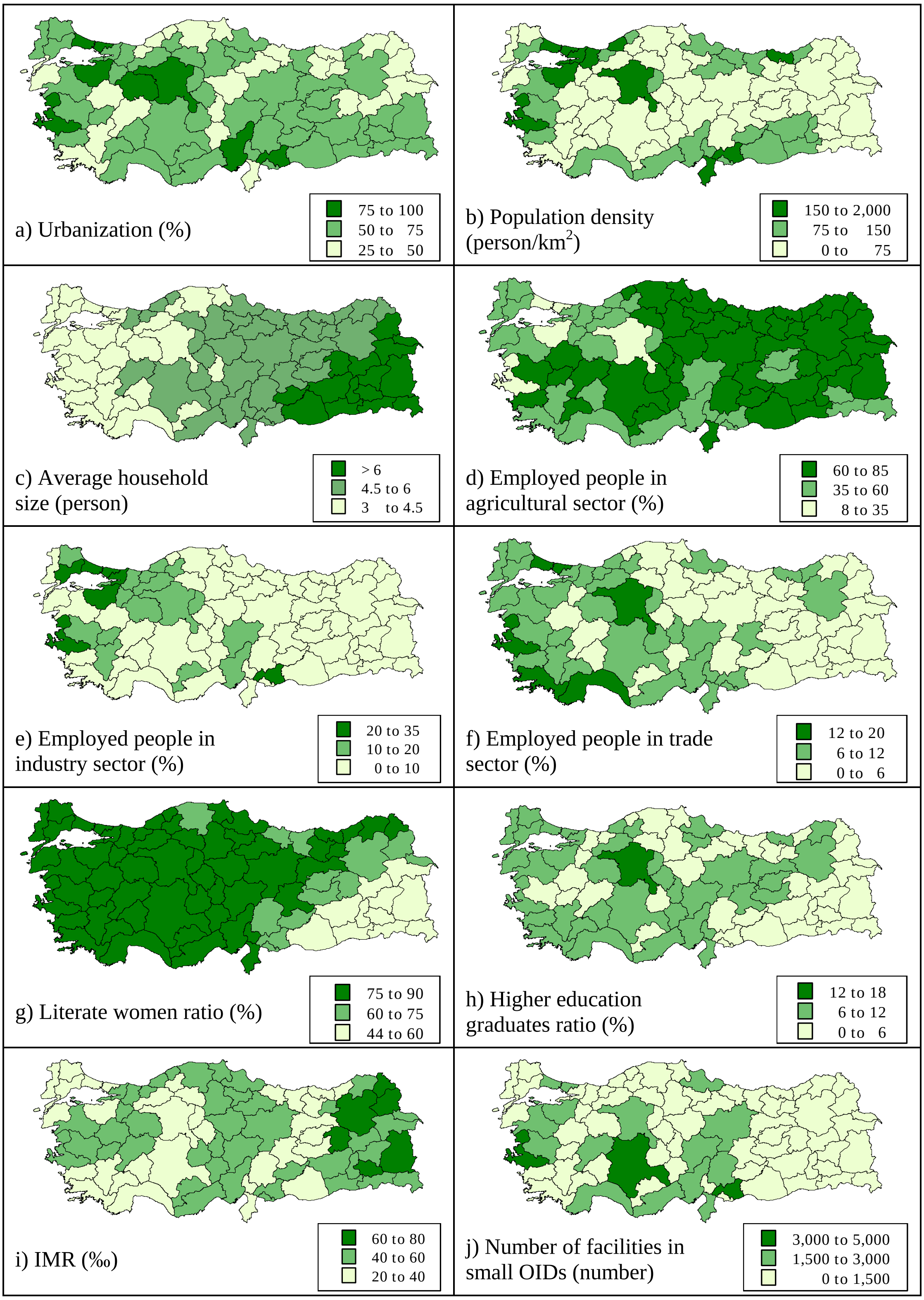


Figure 5: Distribution of explanatory variables in Turkey (2000)

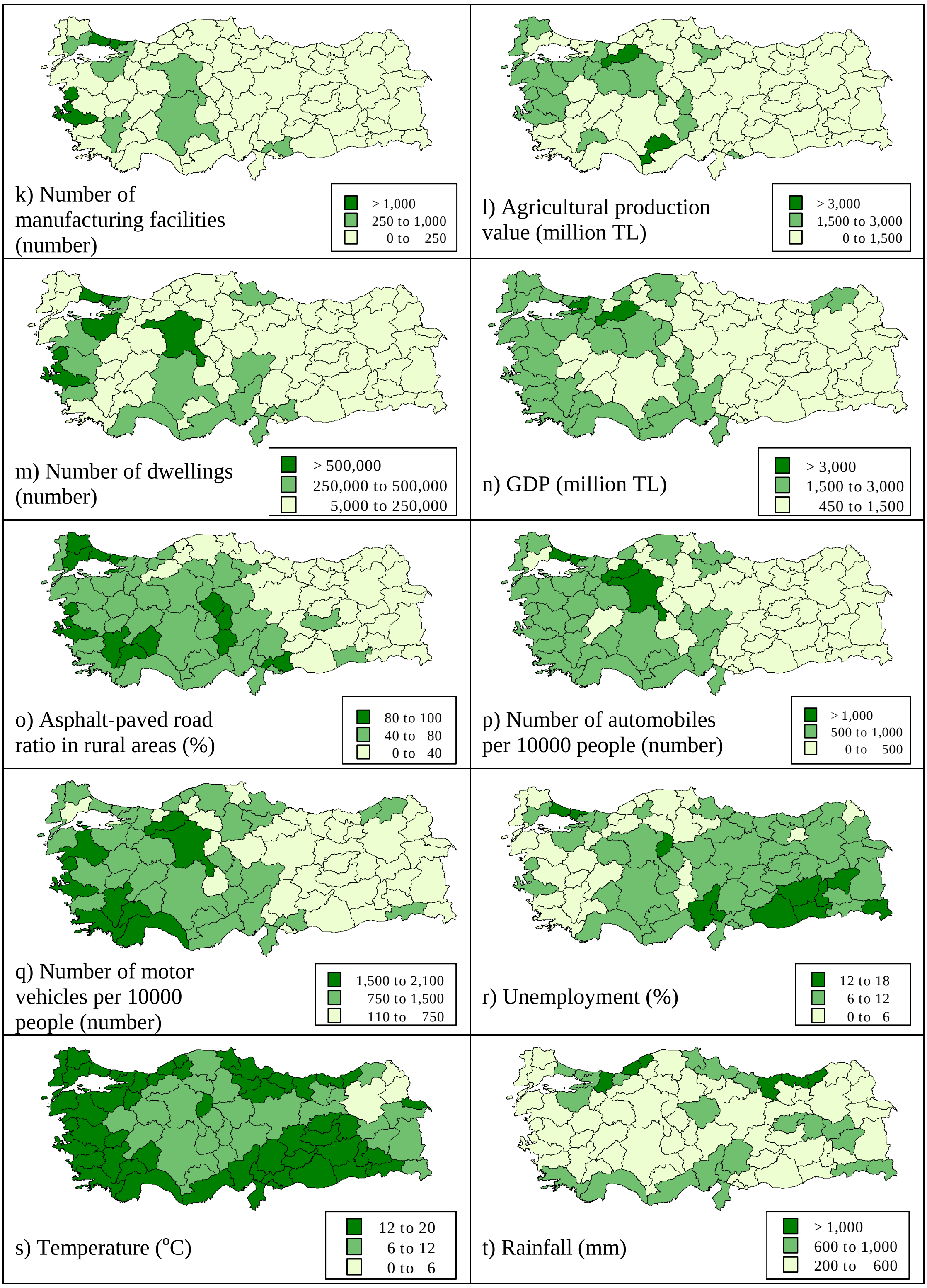


Figure 5 (cont'd)

4.2. MULTICOLLINEARITY ANALYSIS

The cut-off values for VIF and *r* were taken as 4 and 0.40, respectively, since these are the most stringent thresholds in the literature. VIF values were first calculated for 20 variables given in Table 1. The calculations were pursued in an iterative manner eliminating the variable with the highest VIF until the cut-off value of 4 was satisfied. At each step, the variable with the highest VIF was eliminated and VIFs were recalculated. The maximum VIF value dropped below 4 when nine explanatory variables were eliminated. The iteration steps of VIF analysis is given in Table 2. The variables remaining in the last iteration of VIF analysis were then checked by investigating correlation matrix which includes *r* values given in Table 3. Further eliminations were conducted, taking the threshold value for *r* as 0.40. In order to eliminate the least number of variables and still not exceed an *r* of 0.40, employed people in industry sector and GDP per capita were taken out. These variables were the only ones having *r* with p-value smaller than 0.01 (see Table D.3 in Appendix D for p-values of *r*) showing significant association between variables at 99 % confidence level. Finally, the multicollinearity analysis ended up with nine explanatory variables. These variables constituted the first data group (DG1) for which serious multicollinearity effects were eliminated.

Table 2: Iterations to obtain VIF< 4 for explanatory variables in DG1

Iteration #	# of variables entering	Max. VIF	Deleted variable (ID)	Average VIF
1	20	103.46	13	18.50
2	19	27.17	11	10.76
3	18	25.47	16	8.55
4	17	14.08	3	6.16
5	16	13.12	4	5.09
6	15	7.65	6	3.74
7	14	6.22	1	3.23
8	13	4.87	7	2.62
9	12	4.17	17	2.22
10	11	**2.87**	-	1.91

Table 3: Correlation matrix

ID	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]
[2]	0.41																		
[3]	-0.18	-0.17																	
[4]	-0.79	-0.52	0.46																
[5]	0.58	0.51	-0.61	-0.77															
[6]	0.59	0.54	-0.59	-0.76	0.67														
[7]	0.17	0.20	-0.92	-0.41	0.58	0.57													
[8]	0.55	0.38	-0.54	-0.76	0.45	0.80	0.55												
[9]	-0.26	-0.12	0.54	0.31	-0.34	-0.40	-0.55	-0.41											
[10]	0.45	0.31	-0.33	-0.38	0.38	0.57	0.35	0.42	-0.26										
[11]	0.52	0.94	-0.25	-0.63	0.60	0.61	0.28	0.52	-0.14	0.45									
[12]	0.18	-0.21	-0.45	-0.17	0.19	0.19	0.39	0.20	-0.20	0.19	-0.08								
[13]	0.54	0.92	-0.27	-0.63	0.54	0.68	0.29	0.60	-0.18	0.54	0.98	-0.07							
[14]	0.35	0.32	-0.70	-0.64	0.67	0.65	0.65	0.59	-0.42	0.30	0.38	0.39	0.38						
[15]	0.32	0.28	-0.58	-0.47	0.62	0.54	0.60	0.36	-0.33	0.40	0.34	0.38	0.34	0.56					
[16]	0.37	0.29	-0.75	-0.61	0.62	0.75	0.70	0.74	-0.47	0.49	0.44	0.44	0.49	0.74	0.55				
[17]	0.26	0.18	-0.73	-0.50	0.52	0.69	0.64	0.62	-0.49	0.47	0.30	0.53	0.36	0.70	0.58	0.94			
[18]	0.58	0.24	0.53	-0.22	-0.01	0.13	-0.48	0.09	0.10	0.09	0.21	-0.30	0.24	-0.18	-0.23	-0.24	-0.31		
[19]	0.33	0.21	-0.16	-0.33	0.29	0.46	0.04	0.24	-0.42	0.25	0.16	-0.01	0.22	0.29	0.22	0.21	0.30	0.33	
[20]	-0.18	0.11	-0.05	0.01	0.10	0.07	0.06	-0.04	-0.16	-0.11	0.00	-0.27	0.00	0.10	-0.17	0.05	0.05	0.01	0.27

Another data group (DG2) was obtained by loosening the thresholds to include some important explanatory variables, such as the average household size in the models. The aim was to include important and generally used variables in literature from each indicator category as much as possible. Therefore, personal judgment was incorporated in this part. This data group (DG2) was arranged so as to have nine explanatory variables similar to DG1 by taking the threshold values as 0.60 and 10 for *r* and VIF, respectively. Additionally, nine explanatory variables with the highest R^2 were chosen to constitute the third data group (DG3). This data group was constructed regardless of the multicollinearity restrictions, although multicollinearity coefficients were calculated for interpretation purposes.

The scripts regarding the construction of data groups are given in Appendix B.1 and B.2. Explanatory variables that form these data groups are listed in Table 4. There are three variables common for each data group. These are unemployment, IMR and temperature. Further similarities also exist between pairs of data groups.

Table 4: Explanatory variables in data groups

Variable name	**DG1**	**DG2**	**DG3**
Urbanization		✓	✓
Population density	✓		
Average household size		✓	✓
Employed people in industry sector			✓
Higher education graduates ratio	✓	✓	
IMR	✓	✓	✓
Number of facilities in small OIDs	✓		✓
Number of manufacturing facilities		✓	
Agricultural production value	✓	✓	
Asphalt-paved road ratio in rural areas	✓		✓
Number of automobiles per 10000 people			✓
Unemployment	✓	✓	✓
Temperature	✓	✓	✓
Rainfall	✓	✓	

VIF and *r* values are given in Table 5 and 6, respectively. As it is seen, although the cut-off was established as 4, VIF values are all below 2 for DG1 due to restrictions on *r*. *r* values for the variables in DG1 are below 0.40 with minor violations. Average VIF increases for DG2 since constraints are looser compared to DG1. DG2 satisfies the threshold values for VIF and *r* (10 and 0.60, respectively) without any violation. Although there was not any restriction for DG3, the coefficients for the variables in this data group mostly remain below the limits set for DG2 and average VIF does not increase noticeably. The only apparent violation is caused by the high correlation (*r* = -0.75) between average household size and availability of automobiles. Although there is not any serious indication of multicollinearity, relatively high coefficients should be borne in mind for proper interpretation of the analysis results.

Table 5: VIFs of explanatory variables

	VIF		
Variable name	**DG1**	**DG2**	**DG3**
Urbanization	-	4.40	4.31
Population density	1.55	-	-
Average household size	-	4.88	5.54
Employed people in industry sector	-	-	2.91
Higher education graduates ratio	1.60	2.24	-
IMR	1.55	1.76	1.75
Number of facilities in small OIDs	1.42	-	1.56
Number of manufacturing facilities	-	1.76	-
Agricultural production value	1.50	1.64	-
Asphalt-paved road ratio in rural areas	1.81	-	1.99
Number of automobiles per 10000 people	-	-	2.97
Unemployment	1.56	5.72	5.32
Temperature	1.69	1.64	1.65
Rainfall	1.32	1.34	-
Average	**1.56**	**2.82**	**3.11**

Table 6: *r* values for explanatory variables[i]

Variable ID	[1]	[2]	[3]	[5]	[8]	[9]	[10]	[11]	[12]	[15]	[16]	[18]	[19]
[2]	-												
[3]	-0.18	-											
[5]	0.58	-	-0.60										
[8]	0.55	0.38	-0.53	-									
[9]	-0.26	-0.12	0.54	-0.34	-0.41								
[10]	0.45	0.31	-0.33	0.38	0.42	-0.25							
[11]	0.52	-	-0.25	-	0.52	-0.14	-						
[12]	0.18	-0.21	-0.44	-	0.20	-0.20	0.19	-0.08					
[15]	0.32	0.28	-0.58	0.61	0.36	-0.33	0.40	-	0.38				
[16]	0.37	-	-0.75	0.61	-	-0.46	0.49	-	-	0.55			
[18]	0.58	0.24	0.53	-0.01	0.09	0.10	0.09	0.21	-0.30	-0.23	-0.24		
[19]	0.33	0.20	-0.16	0.29	0.24	-0.41	0.25	0.16	-0.01	0.22	0.21	0.33	
[20]	-0.18	0.11	-0.05	-	-0.04	-0.16	-0.11	0.00	-0.27	-0.17	-	0.01	0.27

[i] Only the coefficients for the variables that take place in the data groups are presented. Please refer to Table 1 on page 42 for the definitions of variables.

Higher *r* values pertain to average household size, number of automobiles and employed people in industry comparatively. This is why exclusion of these three variables from DG1 is natural. Urbanization also has high correlation with others to some extent. Temperature and rainfall possess lower *r* values compared to other variables. This indicates that these climatic variables are not correlated with socio-economic and demographic variables. Similarly, agricultural production value is not much correlated with other variables.

The matrix mostly seems to be coherent in terms of the relationships between explanatory variables except for agricultural production value. Agricultural activities generally coincide with underdeveloped regions (Dinçer, *et al.*, 2003). This fact is not supported by the relation of agricultural production value to other indicators of development. As an instance, average household size, high values of which also occur in underdeveloped regions, has a negative relation to this variable. However, it should be considered that there are other factors which, in turn, affect the explanatory variables. For example, proximity to markets and modernization of techniques in agriculture affect agricultural production positively (Dinçer, *et al.*, 2003) and these are consequences of development in a region. In addition, geographic condition of the province may affect the agricultural production.

4.3. NON-SPATIAL DATA ANALYSIS

In the context of non-spatial analysis, OLS regression was carried out with the help of "stats" package of R (see Appendix B.3 for the script). The results of OLS regression are given in Table 7. R^2 increases and AIC decreases from DG1 through DG3. This shows OLS fits DG3 better than the other data groups. The significance of R^2 is high for all data groups according to the p-values. However, there are differences between data groups in terms of β coefficient estimates of common variables. Minor variations may be attributed to the fact that the variables included in each data group differ highly since there is a small number of common variables. However, multicollinearity can be the source of minor and also major differences. VIF calculated with the models' R^2 (VIF_m) is not exceeded substantially in DG1 while most of the variables exceed this limit largely in DG2 and DG3 indicating multicollinearity (see Table 5).

The most discernible deviations are observed for unemployment and rainfall in both magnitude and sign of the β coefficient estimates. The difference in rainfall coefficient between DG1 and DG2 is not emphasized further since the value and the significance of this coefficient are very low. Nevertheless, unemployment is an important potential factor and it is expected to affect MSW generation rate negatively. This is reflected by the results for DG1 but the sign of the unemployment coefficient is implausible for DG2 and DG3. Another easily detectible difference is for β coefficient estimate regarding education level between DG1 and DG2. The expectation is that higher education graduates ratio is an important determinant for solid waste generation which is supported by the OLS regression for DG1. On the other hand, the result for DG2 does not meet the expectation due to low significance of the coefficient belonging to this variable. It is seen that the estimates for DG1 are more plausible, although the fit of the model is better for DG2 and DG3.

Table 7: OLS regression coefficient estimates

	Coefficient estimates[i]		
Variable name	**DG1**	**DG2**	**DG3**
Intercept	1.20 (0.017)	3.33 (0.000)	2.82 (0.000)
Urbanization	-	-1.85e-02 (0.016)	-1.52e-02 (0.035)
Population density	-3.81e-05 (0.886)	-	-
Average household size	-	-2.85e-01 (0.000)	-2.41e-01 (0.002)
Employed people in industry sector	-	-	-1.14e-02 (0.274)
Higher education graduates ratio	5.46e-02 (0.056)	2.27e-02 (0.469)	-
IMR	5.79e-04 (0.927)	4.77e-03 (0.450)	5.84e-03 (0.331)
Number of facilities in small OIDs	-5.55e-05 (0.317)	-	-7.10e-05 (0.175)
Number of manufacturing facilities	-	2.55e-05 (0.849)	-
Agricultural production value	7.54e-05 (0.429)	9.63e-05 (0.305)	-
Asphalt-paved road ratio in rural areas	5.10e-03 (0.033)	-	5.05e-03 (0.025)
Number of automobiles per 10000 people	-	-	4.78e-04 (0.046)
Unemployment	-5.93e-02 (0.002)	3.41e-02 (0.300)	3.81e-02 (0.208)
Temperature	-1.06e-02 (0.543)	-1.61e-02 (0.320)	-2.18e-02 (0.160)
Rainfall	1.16e-04 (0.474)	-9.01e-05 (0.556)	-
R^2 (p-value)	**0.38 (5.83e-05)**	**0.45 (1.24e-06)**	**0.50 (5.33e-08)**
VIF_m $(1/(1-R^2))$	**1.61**	**1.82**	**2**
AIC	**94.68**	**84.58**	**75.57**

[i] p-values are given in parenthesis.

It can be seen that the most significant determinant of MSW generation rate is unemployment if the results of DG1 are further assessed. Low p-value for this variable means that that probability to obtain this coefficient estimate by chance is very low. Considering unemployment as an indicator of income, its negative relation to solid waste generation rate is consistent with previous studies. Income mostly has positive relation to MSW generation rate. Increasing unemployment is an indication of decreasing income per person. As a result, decreasing income leads to less MSW generation rate also in Turkey. Asphalt-paved road ratio in rural areas and higher education graduates ratio are other important variables which are positively related to solid waste generation. The positive relation of education level is also an expected outcome. It can be said that education level leads to high level of waste generation due to improvements in life standards. Asphalt-paved road ratio, on the other hand, has not been used as a determinant of solid waste generation before. This relation is also logical since higher investment in physical infrastructure is associated with regional development (Dinçer, *et al.* 2003).

4.4. SPATIAL DATA ANALYSIS

4.4.1. Spatial Autocorrelation

Both contiguity and distance based weighting matrices were created for the provinces in Turkey using the shapefile which includes the coordinates of the city centers. Different matrices were obtained for different spatial lags determined by the number of nearest neighbors taken into consideration. All of the matrices were row standardized.

Moran’s I was calculated for each type of matrices and results were compared. The results for Moran’s I are given in Table 8 for both nearest neighbors at different lags and contiguity. The related script can be found in Appendix B.4. The correlogram is depicted as given in Figure 6. As expected, Moran’s I shows a decreasing trend at higher lags. Highest autocorrelation is observed between the fourth and sixth lags. Moran’s I based on contiguity falls in between the values for three and four nearest neighbors.

The fourth lag (four nearest neighbors) was chosen to perform the subsequent analysis. At this lag, spatial autocorrelation is significantly high. It can be used to demonstrate local variations with less smoothing compared to the sixth lag which gives a higher Moran’s I.

Table 8: Moran's I values for different spatial weighting scheme

Spatial lag	Moran's I	p-value
1	0.3035	0.1e-01
2	0.2842	0.2e-02
3	0.2963	7.3e-05
4	0.3216	1.4e-06
5	0.3172	1.4e-07
6	0.3287	2.1e-09
7	0.3045	1.5e-09
8	0.2885	6.5e-10
9	0.2875	5.7e-11
10	0.2878	3.3e-12
Contiguity	0.3029	3.3e-06

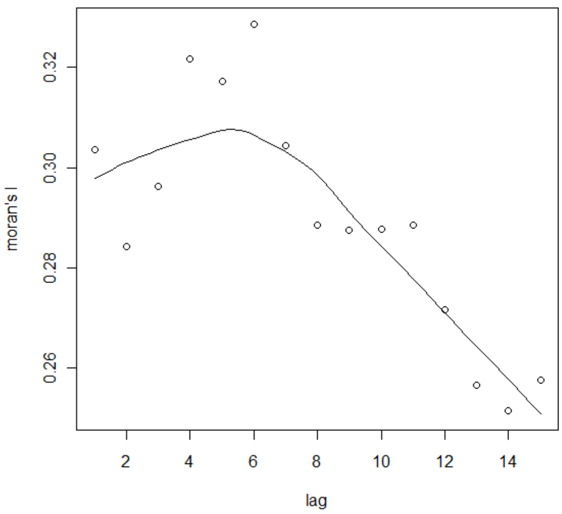


Figure 6: Correlogram for MSW generation rate

Moran's I shows that there is significant positive spatial autocorrelation for MSW generation rate in Turkey (I = 0.32, p-value < 0.001). This can be interpreted simply as that the provinces closer to each other tend to have similar MSW generation rate. This may be an inherent property of MSW generation rate while there may be spatial dependency in the response variable due to spatially structured explanatory variables.

Considering socio-economic similarities of provinces in the same region, the latter may be reasonable. On the other hand, MSW generation may also encompass spatial dependency inherently due to similar SWM systems in closer provinces.

Monte Carlo test results for Moran's I are given in Table 9. In this case, 9,999,999 simulations were randomly selected among 81! permutations. p-value obtained as a result of Monte Carlo test is slightly higher than p-value obtained under normality assumption; however, Moran's I remains significant at 99.9 %.

Table 9: Monte Carlo test for Moran's I at spatial lag 4

Monte Carlo simulations	
Number of simulations + 1	1e+07
Statistic (*I*)	0.3216
Observed rank	9,999,718
Alternative hypothesis	greater
p-value	2.82e-05

Moran's I for residuals of OLS regression are presented in Table 10. Although the Moran's I values for residuals are not very high and not very important at high significance levels, it can be accepted that there is still autocorrelation remaining in the residuals which can be further analyzed.

Table 10: Moran's I values for OLS regression residuals

	DG1	**DG2**	**DG3**
Moran's I	0.0086	0.0378	0.0313
p-value	0.2215	0.1279	0.1111

It should also be noted that residuals for DG1 are less autocorrelated compared to other data groups. This reveals that variables in DG1 explain the variation in MSW generation rate better. Since the variation explained is shared by correlated explanatory variables, less information is supplied by DG2 and DG3. Therefore, MSW generation rate process cannot be fully understood with the variables including multicollinearity.

4.4.2. Spatial Regression

4.4.2.1. Simultaneous Spatial Autoregression (SAR)

LM test was conducted for each data group (Appendix B.5). SAR_{lag} is preferred for DG1 considering the LM test which gives higher values for the lag model compared to the error model (Table 11). The robust statistic for the former (RLM_{lag}) is significant at 95 % for DG1. This condition is reversed for DG2 and DG3 for which LM_{err} (or RLM_{err}) is higher than LM_{lag} (or RLM_{lag}), albeit both insignificant. The affinity towards error model for DG2 and DG3 is an outcome of correlated errors due to less information given by the explanatory variables in these data groups as a result of multicollinearity compared to DG1. Further comparison should be made by checking the results of SAR for each data group. SAR models (SAR_{lag} and SAR_{err}) were run and the coefficients were obtained accompanied by p-values (Appendix B.6).

Table 11: Lagrange Multipliers test[i]

	DG1	DG2	DG3
LM_{err}	0.0134 (0.9078)	0.2579 (0.6115)	0.1767 (0.6742)
LM_{lag}	1.0994 (0.2944)	0.1062 (0.7445)	0.0128 (0.9100)
RLM_{err}	3.2656 (0.0708)	0.2546 (0.6139)	0.4566 (0.4992)
RLM_{lag}	4.3516 (0.0370)	0.1029 (0.7484)	0.2927 (0.5885)

[i] p-values are given in parenthesis.

Results of SAR_{lag} and SAR_{err} are given in Tables 12 and 13, respectively. As expected, SAR_{lag} fits DG1 slightly better than SAR_{err}. For DG2 and DG3, on the other hand, SAR_{err} gives lower AIC than SAR_{lag} showing that the former fits better; however, the difference between models is hardly detectable in terms of AIC. These results are also reflected in the values of autoregressive coefficients. The significance of ρ is higher than the significance of λ for DG1 and vice versa for DG2 and DG3.

Table 12: SAR_{lag} coefficient estimates

	Coefficient estimates[i]		
Variable name	**DG1**	**DG2**	**DG3**
Intercept	9.69e-01 (0.053)	3.21 (0.000)	2.78 (0.000)
Urbanization	-	-1.85e-02 (0.009)	-1.51e-02 (0.023)
Population density	-1.35e-05 (0.956)	-	-
Average household size	-	-2.77e-01 (0.000)	-2.39e-01 (0.002)
Employed people in industry sector	-	-	-1.14e-02 (0.241)
Higher education graduates ratio	5.08e-02 (0.053)	2.30e-02 (0.431)	-
IMR	1.10e-03 (0.852)	4.74e-03 (0.420)	5.83e-03 (0.296)
Number of facilities in small OIDs	-6.59e-05 (0.198)	-	-7.18e-05 (0.140)
Number of manufacturing facilities	-	2.58e-05 (0.836)	-
Agricultural production value	7.43e-05 (0.399)	9.66e-05 (0.267)	-
Asphalt-paved road ratio in rural areas	4.58e-03 (0.042)	-	5.01e-03 (0.017)
Number of automobiles per 10000 people	-	-	4.78e-04 (0.030)
Unemployment	-5.19e-02 (0.003)	3.49e-02 (0.254)	3.84e-02 (0.173)
Temperature	-1.01e-02 (0.534)	-1.60e-02 (0.292)	-2.17e-02 (0.138)
Rainfall	1.16e-04 (0.439)	-8.48e-05 (0.551)	-
Rho (p-value)	**0.1545 (0.291)**	**0.0463 (0.751)**	**0.0162 (0.911)**
R^2 (p-value)	**0.39 (2.94e-05)**	**0.45 (1.20e-06)**	**0.50 (5.7e-08)**
AIC	**95.56**	**86.48**	**78.56**

[i] p-values are given in parenthesis.

Table 13: SAR_{err} coefficient estimates

	Coefficient estimates[i]		
Variable name	**DG1**	**DG2**	**DG3**
Intercept	1.20 (0.010)	3.41 (0.000)	2.88 (0.000)
Urbanization	-	-1.93e-02 (0.006)	-1.59e-02 (0.017)
Population density	-2.96e-05 (0.905)	-	-
Average household size	-	-2.95e-01 (0.000)	-2.50e-01 (0.001)
Employed people in industry sector	-	-	-1.06e-02 (0.283)
Higher education graduates ratio	5.46e-02 (0.038)	2.17e-02 (0.457)	-
IMR	5.18e-04 (0.930)	4.64e-03 (0.439)	5.82e-03 (0.306)
Number of facilities in small OIDs	-5.90e-05 (0.253)	-	-7.67e-05 (0.113)
Number of manufacturing facilities	-	2.97e-05 (0.810)	-
Agricultural production value	7.58e-05 (0.391)	9.57e-05 (0.268)	-
Asphalt-paved road ratio in rural areas	5.14e-03 (0.021)	-	5.14e-03 (0.017)
Number of automobiles per 10000 people	-	-	4.68e-04 (0.033)
Unemployment	-5.80e-02 (0.001)	3.93e-02 (0.200)	4.28e-02 (0.129)
Temperature	-1.10e-02 (0.504)	-1.85e-02 (0.238)	-2.33e-02 (0.116)
Rainfall	1.20e-04 (0.427)	-6.82e-05 (0.634)	-
Lambda (p-value)	**0.0312 (0.883)**	**0.1011 (0.576)**	**0.0875 (0.636)**
R^2 (p-value)	**0.38 (4.75e-05)**	**0.45 (1.20e-06)**	**0.50 (5.7e-08)**
AIC	**96.66**	**86.27**	**78.35**

[i] p-values are given in parenthesis.

The regression coefficient estimates obtained for SAR_{lag} and SAR_{err} are similar to each other. There is not any conspicuous difference between the lag and error model coefficient estimates of variables.

Inconsistency between data groups due to multicollinearity in OLS regression results exists also in SAR results. The goodness of fit of the models also increases from DG1 towards DG3 as in the case of OLS.

Furthermore, spatial autocorrelation remained in the residuals were investigated to see the results of applying spatial regression. It is obvious in Table 14 that the spatial autocorrelation in the residuals is alleviated by applying SAR models compared to the Moran's I values given for OLS residuals in Table 10. The significance of the residual Moran's I diminishes for each of the data groups. From this point of view, SAR_{lag} gives better results for DG1 while SAR_{err} is better for DG2 and DG3. This outcome is consistent with the previous explanations. As it was said, correlated errors prevail for DG2 and DG3 due to less explanation power of variables in these data groups which is accounted for by multicollinearity on the contrary of DG1. This is also supported by the AIC of SAR models for each data group.

Table 14: Moran's I values for SAR residuals[i]

	DG1	DG2	DG3
SAR_{lag}	-0.0535	0.0151	0.0240
	(0.717)	(0.350)	(0.306)
SAR_{err}	-0.0003	-0.0015	0.0003
	(0.432)	(0.439)	(0.429)

[i] p-values are given in parenthesis.

4.4.2.2. Geographically Weighted Regression (GWR)

GWR was performed taking the city centers as regression points (see Appendix B.7). A fixed kernel was used, in other words the bandwidth was same for all regression points. Bi-square function was used as the weighting scheme. As a result of applying GWR, 81 regression coefficient estimates were obtained for each explanatory variable in each data group. The summary of these coefficients including minimum, maximum and quartiles are given in Tables 15, 16 and 17 for DG1, DG2 and DG3, respectively.

Table 15: Summary of GWR coefficient estimates for DG1

Variable name	Min.	1st Qu.	Median	3rd Qu.	Max.
Intercept	-0.9283	0.4051	0.6827	1.1430	2.1000
Population density	-0.0045	-0.0010	0.0000	0.0003	0.0077
Higher education graduates ratio	-0.0562	0.0344	0.0602	0.0888	0.3042
IMR	-0.0076	-0.0005	0.0079	0.0238	0.0507
Number of facilities in small OIDs	-0.0003	-0.0001	0.0000	0.0000	0.0002
Agricultural production value	-0.0002	0.0001	0.0002	0.0003	0.0006
Asphalt-paved road ratio (rural)	-0.0035	0.0019	0.0036	0.0067	0.0189
Unemployment	-0.4118	-0.0861	-0.0240	0.0274	0.0582
Temperature	-0.1312	-0.0782	-0.0403	-0.0152	0.0129
Rainfall	-0.0009	0.0000	0.0002	0.0005	0.0026

Bandwidth: 451.13 km

Average R^2: 0.62

AIC: 26.30

Table 16: Summary of GWR coefficient estimates for DG2

Variable name	Min.	1st Qu.	Median	3rd Qu.	Max.
Intercept	1.9260	2.5630	3.5700	5.1710	6.4240
Urbanization	-0.0529	-0.0415	-0.0194	-0.0003	0.0100
Average household size	-0.6229	-0.5231	-0.3460	-0.2294	-0.1707
Higher education graduates ratio	-0.0554	-0.0123	0.0139	0.0301	0.0336
IMR	0.0013	0.0049	0.0094	0.0132	0.0182
Number of manufacturing facilities	-0.0013	-0.0002	0.0000	0.0001	0.0002
Agricultural production value	0.0000	0.0001	0.0001	0.0001	0.0002
Unemployment	-0.0286	0.0057	0.0581	0.1319	0.1655
Temperature	-0.0787	-0.0638	-0.0372	-0.0199	-0.0059
Rainfall	-0.0002	-0.0001	-0.0001	-0.0001	-0.0001

Bandwidth: 986.04 km

Average R^2: 0.52

AIC: 48.94

Table 17: Summary of GWR coefficient estimates for DG3

Variable name	Min.	1st Qu.	Median	3rd Qu.	Max.
Intercept	1.4310	2.0860	3.0050	4.6150	6.5370
Urbanization	-0.0502	-0.0327	-0.0140	-0.0005	0.0074
Average household size	-0.6601	-0.4833	-0.2807	-0.1979	-0.1619
Employed people in industry	-0.0536	-0.0206	-0.0145	-0.0080	0.0028
IMR	0.0030	0.0060	0.0097	0.0138	0.0184
Number of facilities in small OIDs	-0.0001	-0.0001	-0.0001	0.0000	0.0000
Asphalt-paved road ratio (rural)	0.0019	0.0035	0.0043	0.0051	0.0070
Number of automobiles (/10000)	0.0002	0.0003	0.0004	0.0005	0.0010
Unemployment	-0.0109	0.0111	0.0492	0.1142	0.1706
Temperature	-0.0886	-0.0644	-0.0373	-0.0214	-0.0089

Bandwidth: 989.88 km

Average R^2: 0.54

AIC: 45.96

The results show that the regression coefficients vary over the study area. Even, the sign of the coefficient changes for most of the variables. As a result, it can be said that the effect of the explanatory variables on MSW generation rate differ from one province to another.

GWR fits DG1 far better than DG2 and DG3. GWR gives emphasis to the effects of explanatory variables in local basis and this is reflected by DG1 more since the variables in this data group have higher explanatory power. Moreover, the bandwidths for DG2 and DG3 are more than two times of the bandwidth for DG1. This leads to smoothing of the GWR coefficient estimates across the study area for DG2 and DG3 while the bandwidth obtained for DG1 helps local effects to be more highlighted.

Results for DG1 are interpreted in more detail since multicollinearity is vastly eliminated in this data set and so the results are more reliable for this data group. The GWR coefficient estimates of DG1 are mapped in Figure 7 where the variation in the coefficients over the study area can be observed.

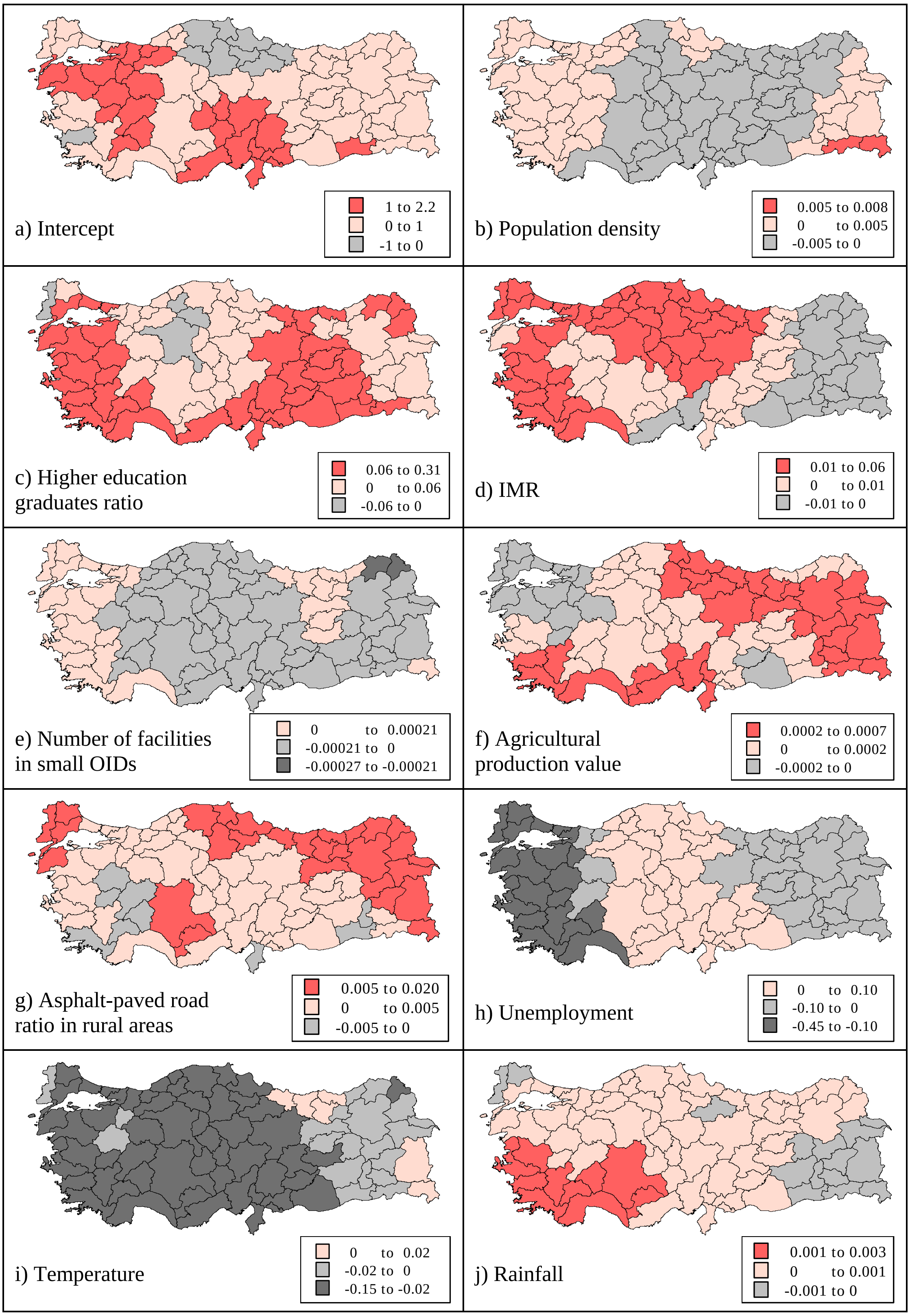


Figure 7: GWR coefficient estimates for DG1

Grey scale in Figure 7 denotes negative values while colored areas represent provinces with positive attribute values as in the subsequent figures. However, it should be noted that significances of these coefficients also vary across the country. This is why insignificant coefficient estimates should not be taken into consideration while interpreting the results.

t-scores[1] of the coefficient estimates are mapped in Figure 8. These maps are based on the significance level that associates with the t-values. The significances of β coefficients show great variations across the country. The coefficients can be found to be significant at 99.9 % confidence level in one province while it is found insignificant at confidence level of 90 % in another province.

t-values also make it possible to compare the significances of explanatory variables. Unemployment, temperature and higher education graduates ratio are important variables considering the number of provinces in which they are significant and also considering the level of this significance as visualized in Figure 8. Intercept and coefficients of population density and number of facilities in small OIDs have the lowest significance again visually comparing the maps.

The coefficients that are significant at 99 % confidence are presented in Figure 9. The direction of the effects of the explanatory variables can also be distinguished in this figure. When insignificant coefficient estimates at confidence level of 99 % are eliminated, the variation in the sign of the coefficients as pointed in the summary of GWR results (Tables 15, 16 and 17) is mitigated. Most of the variables have coefficients either positive or negative signed. Population density, number of facilities in small OIDs and unemployment are the only variables that have both negative and positive coefficients. Coefficients of the first two of these variables are very close to zero and have lower significance. This is why they are susceptible to changes in sign.

[1] Absolute values of t-scores were taken to avoid complexity.

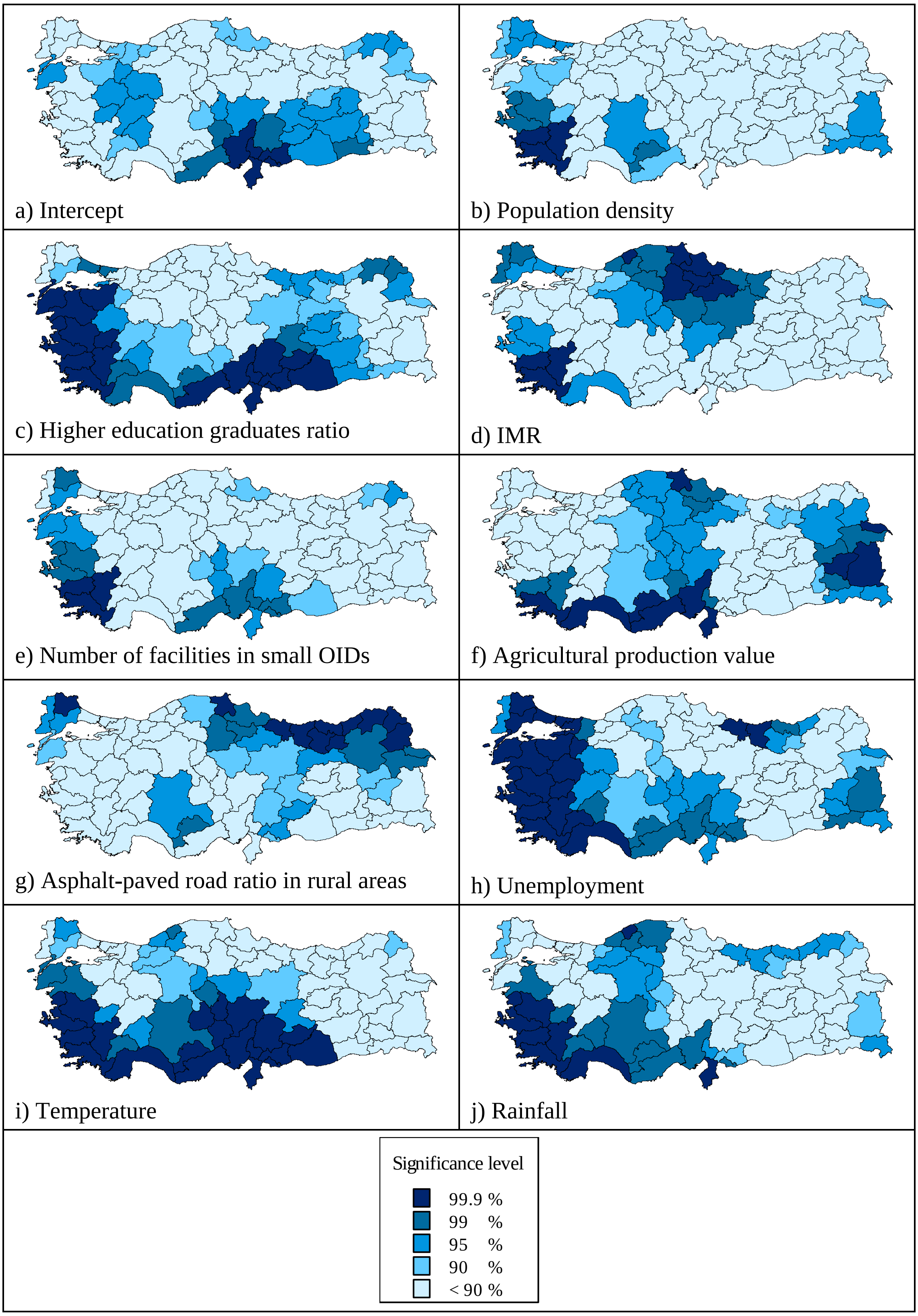


Figure 8: t-scores for GWR coefficient estimates of DG1

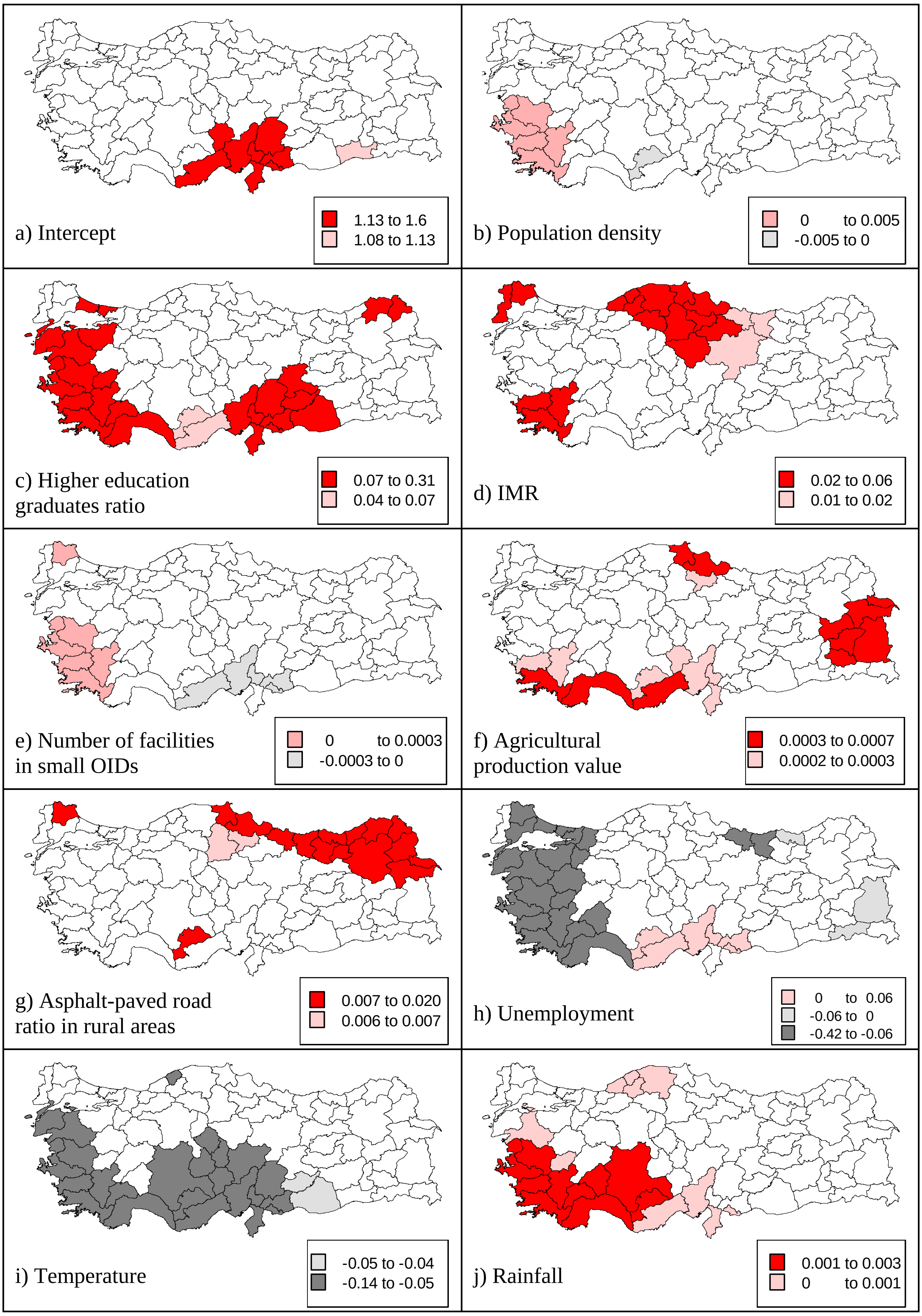


Figure 9: Significant GWR coefficient estimates for DG1

Population density is significant in the southwest and mostly has positive signed coefficients. This is not surprising since internal migrations occur towards developed provinces where waste generating activities are at higher levels. However, increase in population density may also cause decrease in per capita MSW generation rate since total waste generated is divided by the population to obtain generation per capita. Therefore, this variable may have positive or negative effect which compensate for each other. This is obviously the reason why it is significant in only few provinces. As a result, population density is not a robust explanatory variable to determine solid waste generation rate and its effect is difficult to explain as also pointed out by Linzner and Beigl (2005).

Higher education graduates ratio has a positive relation to MSW generation rate. This relation is significant in the west and south of Turkey. It is known that educational level is an essential indicator of economic and social development. Level of education determines the peoples' participation in social activities (Dinçer, *et al.*, 2003). Consequently, lifestyles of people change depending on their educational status. Moreover, most of the products appeal to educated people and manufacturers can reach these people easily. Therefore, educated people are more interested in and aware of these products and besides they have higher accessibility to these. This is why educational differences bring about different consumption patterns within the country. For example, reading newspapers, using electronics, access to internet are all positively related to educational status. Bach, *et al.* (2004) deduces that higher educational level causes higher amount of waste paper due to newspaper consumption. Moreover, computer usage and access to internet are more widespread among higher education graduates in Turkey (TurkStat, 2009d), which contributes to electronic waste generation. Correlation matrix (Table 3) also reveals that there is a strong positive relation (r = 0.74) between higher education graduates and number of automobiles per 10000 people. As a result, as ratio of higher education graduates within the population increases, MSW generation rate increases due to alteration in lifestyles and improvement in life standards.

IMR is also positively related to MSW generation rate and its significant coefficient estimates are located in the middle north (Middle Black Sea Region) and partly in the west. Positive effect of IMR on MSW generation rate is not an anticipated result. IMR is a widely-used indicator of socio-economic development level. It is found to be low in developed regions as a result of prevalence of health services, good economic condition and other factors like high educational level (Dinçer, *et al.*, 2003). This is why IMR is expected to affect MSW generation rate negatively as demonstrated by Beigl, *et al.* (2004). However, this is not the case in Turkey. The correlation between IMR and higher education graduates ratio is not very high (r = -0.41). Similarly, IMR is not significantly related to GDP per capita (r = -0.42). Therefore, other factors that are triggering IMR become important to explain its positive relation to MSW generation rate which can only be revealed by further investigation on IMR, especially in the related region.

Number of facilities in small OIDs has both negative and positive β coefficient estimates. The provinces having significant coefficient estimates are located in the central south and west. Coefficients of this variable are very close to zero and the number of provinces with significant coefficient estimates is not very high similar to population density. Industrial activities are known to produce high amounts of waste generation. However, small OIDs may not be adequate representative of industries due to the scale and type of enterprises. Besides, these establishments may actually help managing wastes so that generation of wastes declines compared to the situation in which small enterprises are independently dispersed within the provinces. As a result, this variable may have contradicting effects on solid waste generation and these effects in different directions may obscure each other as for the case of population density.

The significant β coefficient estimates of agricultural production value are positive and observed in the southern (Mediterranean Region) and eastern provinces and a few provinces in the middle north. It was mentioned that agricultural activities are generally associated with underdeveloped regions. As said, production value increases due to opportunities coming along with economic development. This is why positive relationship of agricultural production value to MSW generation rate is understandable. Moreover, farm products contribute to waste generation as Afon (2007) pointed. This contribution may differ from region to another depending on how much of these products are utilized. The type of the agricultural activities is also of importance. For example, greenhouses are common in the Mediterranean region and these may contribute to high amount of plastic waste generation as indicated by Gidarakos, *et al.* (2006).

Asphalt-paved road ratio in rural areas positively relates to MSW generation rate showing significance in the northeast. Asphalt-paved road ratio gives idea about accessibility to rural areas by collection vehicles. Consequently, total waste amount collected and included in the calculations increases, as the means of access to rural areas are enhanced. The northeast of the country is the region where urbanization is lower (see Figure 5a). In other words, population living in rural areas is in majority in the provinces of this region. Therefore, waste originating from rural areas becomes major determinant of the total MSW generation in the related province. This is why the coefficient for asphalt-paved road ratio in rural areas is found to be significant in this region.

Unemployment is an important economic indicator and it was used to reflect economic status of people instead of income for which the data was not available. Increase in unemployment results in reduction of buying power and, therefore, consumption. Accordingly, solid waste generation per capita experiences a decline. Poor economic condition of the public restricts consumption whatever the other circumstances are. As a result, unemployment is expected to affect MSW generation rate significantly and negatively. This is largely supported by the GWR results. However, there are some provinces where β coefficient estimates for unemployment is positive as seen in Figure 9h. Looking back at Figure 8h, it can be seen that these provinces have β coefficient estimates at lower significance among the ones in Figure 9h. This is also the case for number of facilities in small OIDs and population density. When level of significance is raised from 99 % to 99.9 %, the coefficients for these variables become single signed.

Temperature and rainfall also have significant coefficient estimates in a considerable number of provinces. The direction of effects of these environmental variables is not easily predictable as a result of their many-sided consequences. The results show that solid waste generation rate decreases as average annual temperature increases in a province. The provinces in the south and west coastlines stand out with respect to significance of temperature variable. One possible explanation lies behind the heating means. In these regions, central heating systems are not common. Firewood and coal are common means of heating instead of natural gas (TurkStat, 2009e) and temperature is an essential determinant of amount of fuel consumed and resultant waste. Rainfall represented by annual precipitation, on the other hand, positively affects MSW generation rate. One reasonable link between rainfall and solid waste generation is relevant to measurement of amount of waste collected. If the standard procedures are not followed while determining this amount, the differences may arise due to moisture content which adds up to weight of the waste.

GWR residuals are tested for spatial autocorrelation and the results are given in Table 18. The autocorrelation structure is almost completely removed by GWR for DG1 as revealed by the low significance of residual Moran's I for that data group. For DG2 and DG3, GWR works better than OLS in eliminating the spatial autocorrelation of residual; but, its performance is not as good as SAR. As seen, Moran's I of GWR residuals have higher significance than Moran's I of SAR residuals. The bandwidth again plays a role in these results determining the neighborhood of regression points.

Table 18: Moran's I values for GWR residuals

	DG1	DG2	DG3
Moran's I	-0.2091	-0.0343	-0.0209
p-value	0.9008	0.2665	0.1513

The distribution of local R^2 for GWR is given for DG1 in Figure 10. It is seen that model prediction is better in the south-west and south-east while around Central Anatolia lower prediction is observed. This map gives an idea about the representativeness of the model in different regions. According to Figure 9, β coefficients are not significant in the provinces with low R^2. Therefore, the explanatory variables included are not adequate to explain the variation in this region compared to other regions. Moreover, bandwidth also has an effect on the local fit of the model. Since bandwidth is fixed over the study area, number of data points included, which has an impact on the local fit, varies between regression points due to irregular areal units.

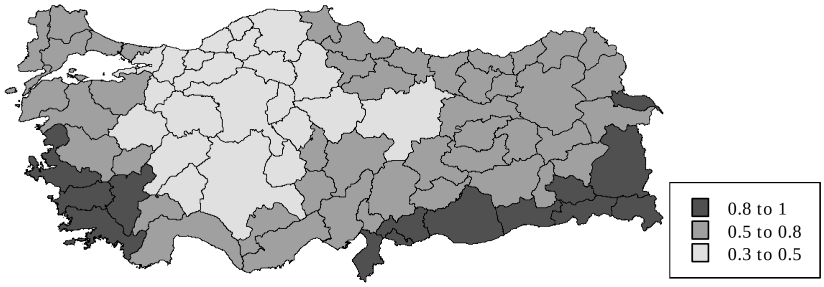


Figure 10: R^2 distribution of GWR (DG1)

t-scores of GWR coefficient estimates for DG2 and DG3 are presented via the maps in Figures 11 and 12, respectively. As seen, unemployment and temperature keep their significance also in these data groups. For both data groups, intercept become very significant for all provinces. Average household and urbanization, which are not included in DG1, also have highly significant coefficient estimates for most of the provinces. The maps in Figures 13 and 14 depict the coefficients that are significant at 99 % for DG2 and DG3, respectively. It is seen that a high number of variables does not have significant coefficient estimates in any province. Similar to SAR results, unemployment coefficients are positive on contrary to the expectation. It is highly probable that these are impacts of multicollinearity. Average household size and urbanization have negative coefficients. Although the results of these data groups are not reliable due to higher multicollinearity, the sign and significance of average household size coefficient is reasonable. This variable was also used in previous studies and found to have negative relation to solid waste generation rate (Linzner & Beigl, 2005; Beigl, *et al.*, 2004; Dennison, *et al.*, 1996; Sterner & Bartelings, 1999; Bandara, *et al.*, 2007). Basic needs of a household remain constant independent of the household size due to common usage and that covers the important part of consumption and so waste generated. That is why the per capita waste generation decreases as people living in a household increases.

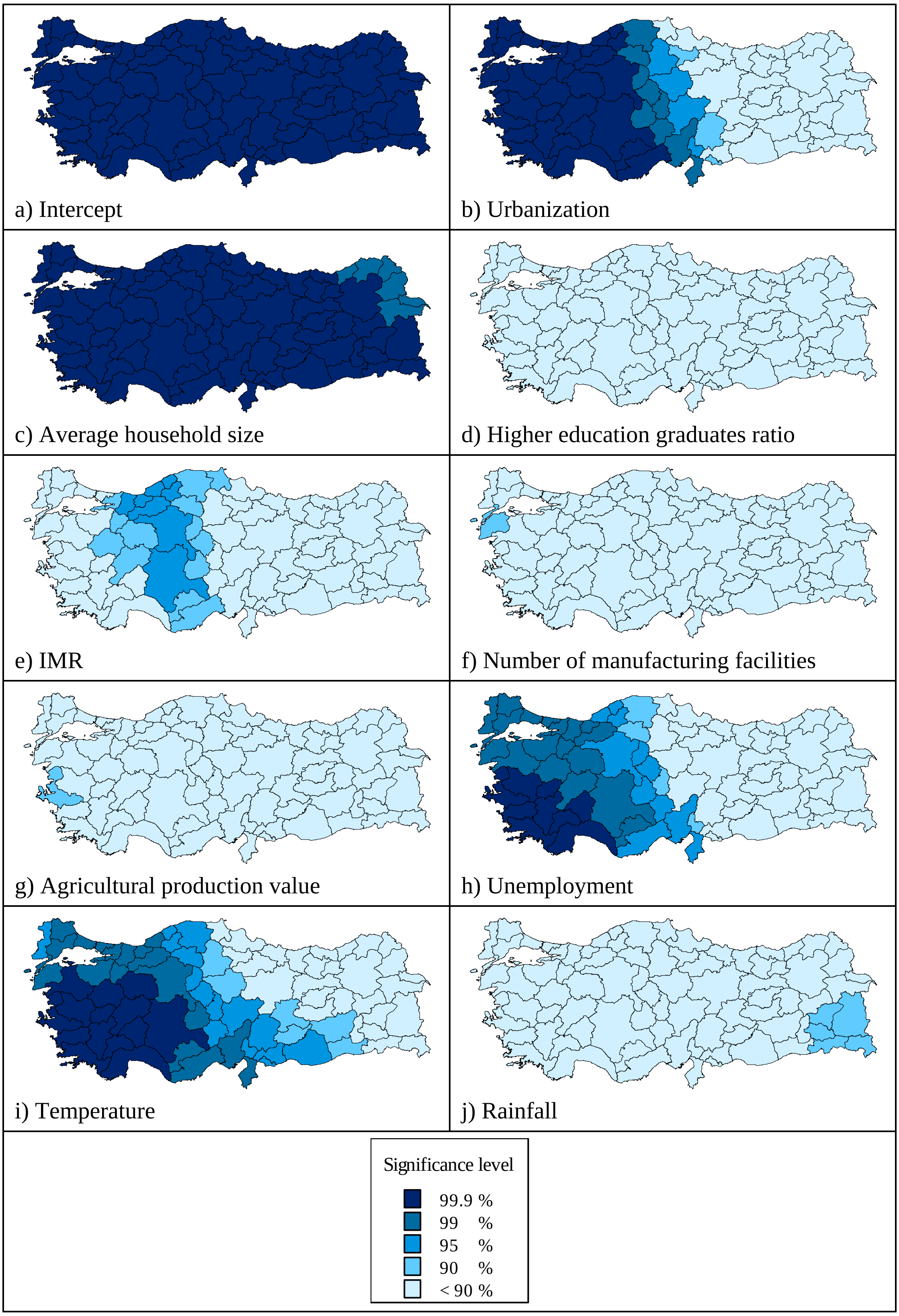


Figure 11: t-score for GWR coefficient estimates of DG2

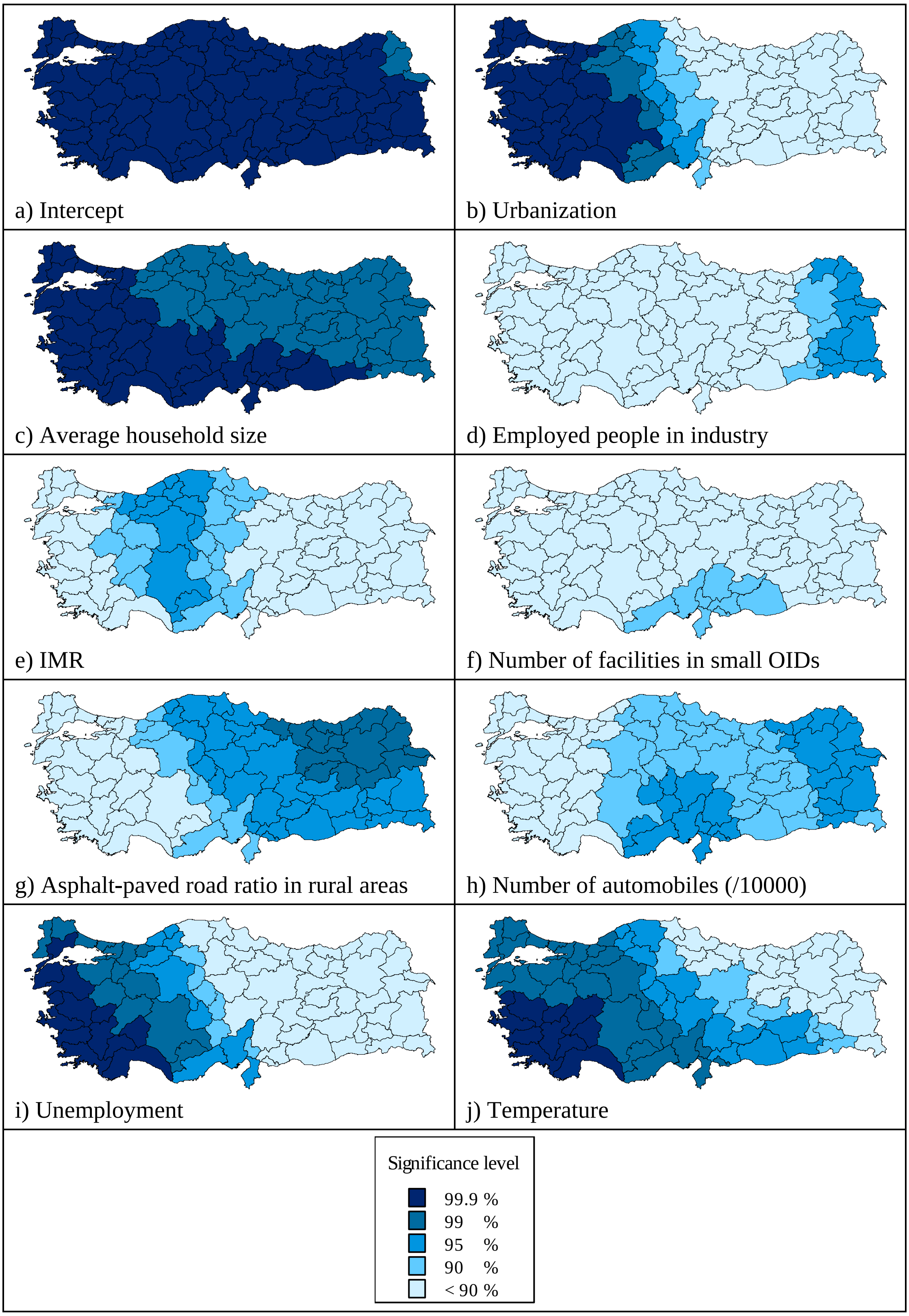


Figure 12: t-score for GWR coefficient estimates of DG3

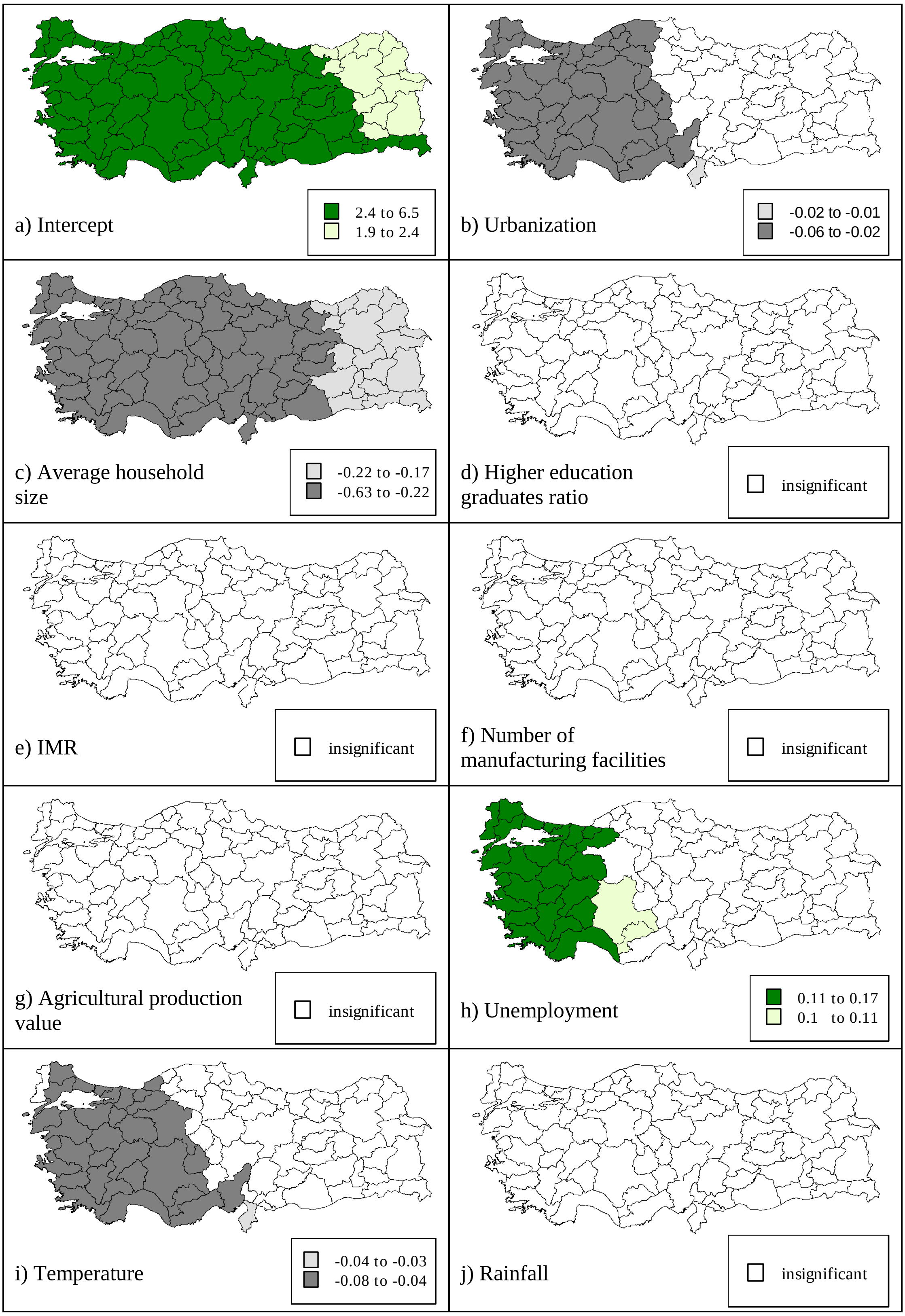


Figure 13: Significant GWR coefficient estimates for DG2

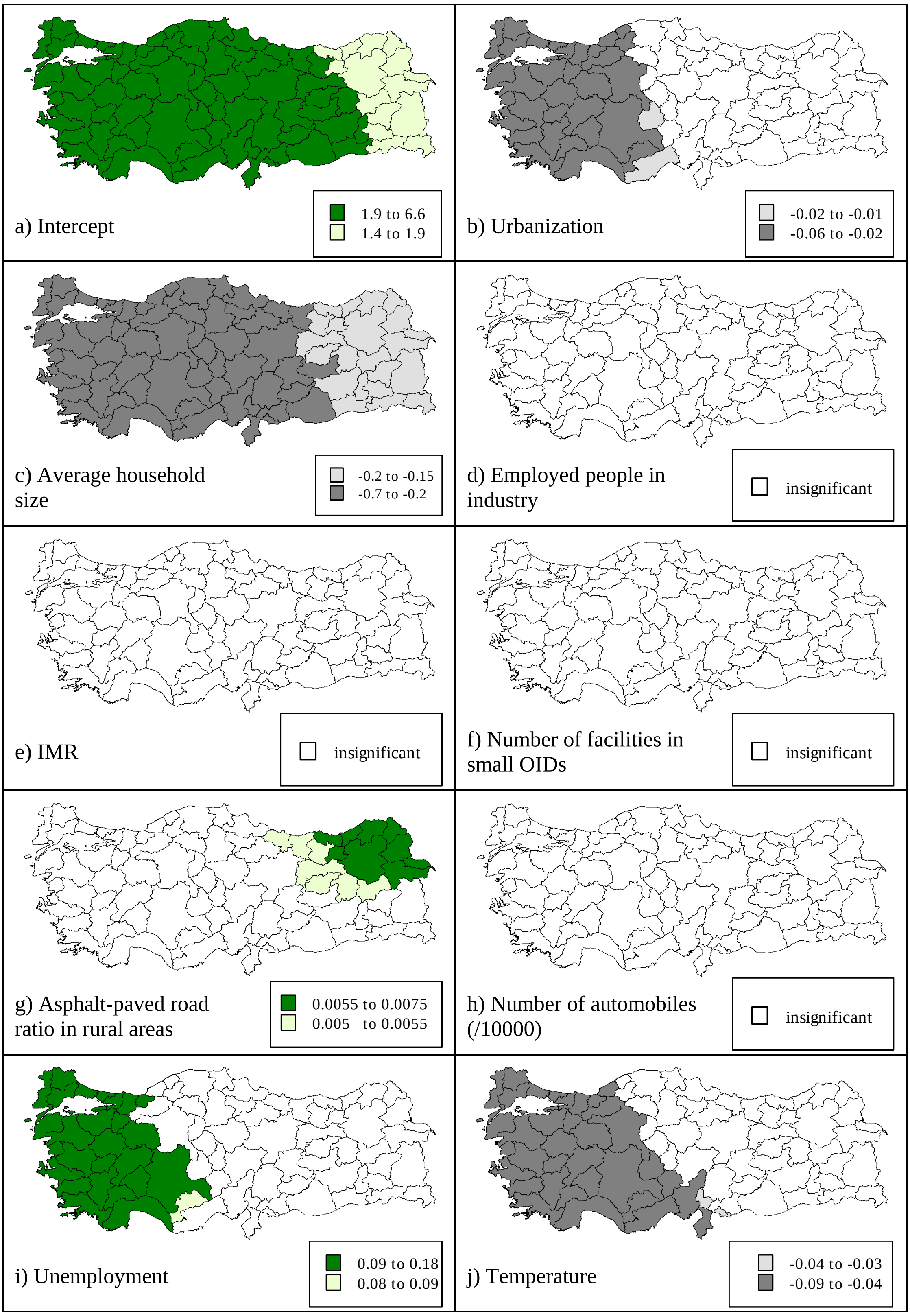


Figure 14: Significant GWR coefficient estimates for DG3

Average household size, furthermore, has a high negative correlation (r = -0.70) to GDP which is an important determinant of economic welfare and expected to affect solid waste generation rate positively (Linzner & Beigl, 2005; Beigl, *et al.*, 2004). However, average household size could not be included in DG1 due to its high correlation to many variables. It can be further investigated in a small sized data group. Urbanization coefficient, on the other hand, has an unexpected sign. As high ratio of urbanization is known to be an indication of socio-economic development, it is expected to affect MSW generation rate positively (Ordóñez-Ponce, *et al.*, 2004). On the other hand, the correlation between urbanization and unemployment is positive and it is not very low (r =0.58). Therefore, driving forces and socio-economic consequences of urbanization is of interest when interpreting the results. However, this should be carried out only after multicollinearity is eliminated as for average household size.

4.4.3. OLS with Eliminated Neighbors' Effect

OLS regression was carried out for the subset of provinces for only DG1 (see Appendix B.8 for the script). The subset of the provinces was created with the help of weight matrix determined in Section 4.4.1. This subset is made up of twenty provinces as seen in Figure 15. The results of OLS for this subset are presented in Table 19. Although R^2 is higher than the one for OLS regression of the whole population, the adjusted R^2 is very low since the number of samples is very small especially considering the number of independent variables.

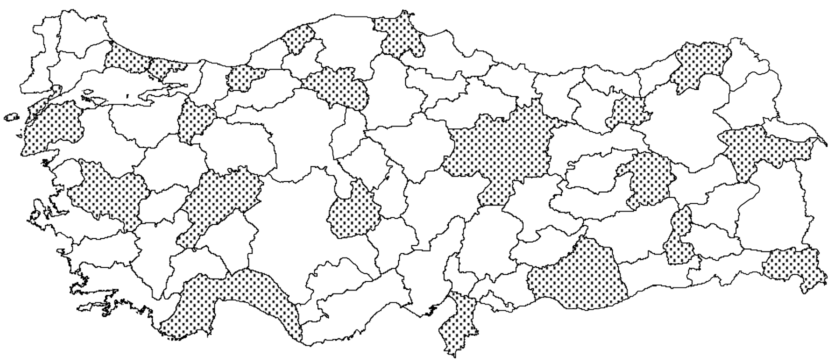


Figure 15: Subset of the provinces chosen for OLS regression

Table 19: OLS coefficient estimates for the selected subset

Variable name	Coefficient	p-value
Intercept	7.54e-01	0.620
Population density	1.98e-04	0.749
Higher education graduates ratio	4.80e-02	0.575
IMR	2.22e-02	0.327
Number of facilities in small OIDs	-1.83e-05	0.917
Agricultural production value	-1.99e-04	0.773
Asphalt-paved road ratio (rural)	1.78e-04	0.985
Unemployment	-1.19e-01	0.063
Temperature	4.95e-02	0.416
Rainfall	-4.92e-04	0.569

R^2 (p-value): 0.49 (0.465)
R^2_{adj}: 0.03
AIC: 34.13
Residual Moran's I: -0.1857 (0.7241)

The significances of the coefficients are also very low. Since the provinces are not within each other's neighborhood, the local effects of the variables are diminished which also alters the global effects. The only significant coefficient estimates pertains to unemployment, which has a confidence level of 90 %. This result is attributed to the high significance of unemployment at global scale since it is the only variable significant at 99 % in OLS regression computed with all provinces. Furthermore, the small sample size is an important factor of insignificant coefficient estimates. Small number of data points causes higher standard errors and lower t-values (Fotheringham, *et al.*, 2002). Multicollinearity which increases standard errors may also be the case for this subset of provinces since the sample size is small.

Moran's I of residuals has a p-value much higher than the Moran's I for the residuals of the whole population (see Table 10). When neighboring effect is eliminated, the same model with the same variables resulted in Moran's I of lower significance with recalculated β coefficients for a different set of provinces. Furthermore, the residuals of the previous OLS regression (Section 3.4) for this subset were subjected to Moran's I test (i.e. the β coefficients are the ones calculated for the whole population). It was found that the low significance of Moran's I prevails (I = -0.1675, p-value = 0.82). This means that the errors get correlated when all provinces are included due to neighboring effect which is a result of spatial dependency.

4.5. VALIDATION

The coefficients obtained as a result of OLS, SAR_{lag} and GWR analyses were tested for DG1 based on the year 1990. The data on facilities in small OIDs was not available for 1990. However, ignoring this parameter is assumed to be acceptable since this variable appeared to be insignificant in impacting the MSW generation rate.

The script for validation is given in Appendix B.9 and the results are given in Table 20 in terms of average predicted MSW generation rate, RMSE and r indicating the correlation between the observed and predicted MSW generation rate. OLS and SAR_{lag} give better predictions considering RMSE. However, the correlation results show that GWR gives predictions more consistent with the observed values although the deviation is higher. This is because GWR takes local variation into consideration and the effects of the explanatory variables are estimated more correctly especially in terms of direction.

Table 20: Validation results

Model	Ave. MSW generation rate (kg/d-ca)	RMSE	r
OLS	1.28	0.61	0.47
SAR_{lag}	1.31	0.60	0.49
GWR	1.56	0.78	0.59

Average observed MSW generation rate is 1.2 kg/d-ca for 1990. The average values for predicted MSW generation rates again show that OLS and SAR_{lag} give better predictions in terms of average value compared to GWR which considerably overestimates the average MSW generation rate.

RMSEs are high compared to the average MSW generation rate considering the error in predicting the total waste generated in a province. However, time interval between two data sets is critical. This is an important factor because waste management policies have marginally evolved in Turkey within this period. In 1991, Solid Waste Control Regulation entered in force. The waste collection service of the municipalities was first taxed in 1993 (i.e. sanitation tax) and National Environmental Action Plan came into effect in 1998. İZAYDAŞ was also founded in the same period of time (1996) (TCA, 2007). As a result, waste management in Turkey underwent a significant change between 1990 and 2000. This change may have led to alterations in solid waste generation process itself and also in the data quality. The former is related to source reduction attempts and the latter arises with the improvements in collection efficiency, preventing MSW commingling with industrial waste, etc.

CHAPTER 5

DISCUSSIONS

5.1. DATA

Data collection is a critical step for studies including data analysis as also observed in this work. The impediment in data analysis is generally the unavailability of reliable data. Socio-economic and especially demographic variables were substantially available for the years that a census was taken. Still, Dinçer, *et al.* (1996) and Dinçer, *et al.* (2003) were very helpful to bring together the necessary variables. Data on solid waste generation, on the other hand, was not adequate due to missing data for some years. The interpolation technique may not be efficient enough in spite of its coherency. Yet, it was utilized as a provisional remedy in order to be able to introduce spatial techniques to solid waste management. Moreover, MSW generation data was assumed to include waste separated as recyclables and the portion collected by scavengers. However, the data is based on measurements carried out by each municipality itself. Therefore, lack of standardization can lead to errors, which decreases the reliability of data. Furthermore, absence of waste composition data limited the extent of this research. Since the data is based on weight, the densities of different waste types gain importance in determining the results as previous studies demonstrated (Afon, 2007; Banar & Özkan, 2008). In fact, knowing the components of waste generated, which have a varying density range, would have been very helpful in interpreting the results.

In all types of regression analysis, the explanatory variables should be independent from each other. In other words, there should not be multicollinearity among these since dependency between explanatory variables leads to erroneous results. All possible determinants were compiled from available sources and subjected to elimination through multicollinearity analysis. It can be deduced from multicollinearity analysis results that the socio-economic and demographic variables are prone to multicollinearity. This could also have been observed among environment related variables if a high number of them were included in the analysis. As a result, examining multicollinearity should not be passed over for the sake of reliability of results, especially when there are a high number of explanatory variables.

The primary reason for creating three different data groups was to observe the effect of multicollinearity on results. As Afon and Okewole (2007) revealed, only a small portion of explained variance can be composed of unique contribution of each explanatory variable while the majority is shared by two or more variables. Consequently, fit of the model may be high although the variables included are not significant parameters. Similarly, OLS and SAR resulted in lower fit for the data group with eliminated multicollinearity (DG1). Moreover, coefficient estimates for some of the variables in data groups embodying multicollinearity (DG2 and DG3) were not found to be reasonable. As a result, high prediction power alone does not provide any evidence for the accuracy of the coefficient estimates for explanatory variables. Therefore, how well the model fits the data may be misguiding if multicollinearity exists.

5.2. MODELS

The models differ from each other in considering spatial dependency and also non-stationarity. OLS is a non-spatial technique so it does not account for spatial dependency while spatial models of SAR and GWR do. On the other hand, OLS and SAR are global models since they each result in a single coefficient estimate valid for all provinces for a variable, while GWR is a local model estimating coefficients specific to each province.

In spatial analysis, it is compulsory to construct a spatial proximity matrix. The features of the matrix are of high importance due to its effects on the results. The distances between city centers and state of sharing boundary were used as criteria of neighborhood and row standardization was applied.

Spatial autocorrelation structure of the MSW generation rate was inspected prior to regression analysis. The correlogram revealed that waste generation rate possess a spatial dependency within closer neighborhood and as the lag increases this dependency evaporates. The peak of Moran's I was observed around the fifth lag. The contiguity based Moran's I coincided with the value between the third and fourth lags. As a result, the fourth lag was chosen to be representative of the autocorrelation structure of MSW generation rate.

It was found that MSW generation rates belonging to provinces within closer neighborhood showed similarities in Turkey for the year 2000 due to positive spatial autocorrelation. This result is usual since, in Turkey, there is an obvious clustering of provinces according to their development level which in turn affects waste generation process. The idea of applying spatial regression is based on the fact that provinces which are in close neighborhood of each other tend to have similar MSW generation rates. However, Moran's I coefficient for residuals of OLS regression showed that the significance of remaining autocorrelation was not very high. Yet, it was not very low which would mean the spatial autocorrelation disappeared completely. The results of OLS residual autocorrelation also showed that as multicollinearity present in data groups increased, the significance of autocorrelation remained increased due to presence of redundant information.

When SAR was employed, it was seen that the fit of the model was very similar to OLS regression. This is not an extraordinary situation considering the moderate significance of spatial autocorrelation found in OLS residuals. Due to the same reason, the significance of autoregressive coefficients (ρ and λ) was found to be low. As a result, SAR model approximated non-spatial OLS model in terms of regression coefficient estimates and the estimated SAR coefficients did not differ from estimated OLS coefficients noticeably. If there were spatial autocorrelation at higher significance, the results would have been affected and spatial models would have been expected to surpass OLS regression.

GWR resulted in lower AIC for each data group compared to OLS and SAR models. GWR includes local variations and this is why it had higher goodness of fit compared to the global models. The improvement was remarkable especially for

DG1 so that it was the data group GWR fitted best. This was not the case for OLS and SAR which was explained by explanatory power of DG1 and its bandwidth. As bandwidth got higher, the results of GWR approached the results of global models.

The second quartiles or the medians of GWR coefficient estimates were near to the estimated OLS and SAR coefficients. This indicated that the local (GWR) coefficients are consistent with the global (OLS and SAR) coefficients. It can be deduced from this result that OLS and SAR coefficients are representative of the average and hinder variations.

Local R^2 also gave hint about varying relationships. The variables included were inefficient for some of the provinces as demonstrated by GWR results. Observing the variation in the coefficient estimates and the explanatory power of the independent variables can be helpful in dividing the country into regions for SWM. The effectiveness of GWR can be increased by applying an adaptive kernel. In this method, bandwidth of the kernel changes according to the density of data points. Local fit of the model can be improved by this way. It may also be good to apply a mixed GWR model as described in Brunsdon, *et al.* (1999).

The OLS model prior to which neighbors were eliminated resulted in less correlated residuals compared to the OLS regression with all provinces. That was another indicator of spatial dependency although it was not very significant to affect the coefficient estimates of models dramatically. It should be noted that the OLS regression gave very low goodness of fit when adjusted for the number of data points and explanatory variables.

According to the validation results, RMSE was lower for OLS and SAR. If the aim is predicting the amount of waste generated in Turkey, OLS and SAR are favorable. As seen in Figure 16, the predicted values of MSW generation rate are clustered around the trend line of observed MSW generation rates in 1990 for OLS and SAR. GWR, however, exhibits the distribution better in terms of peaks and troughs since it models MSW generation rate at local level while OLS and SAR are global models and represent the average.

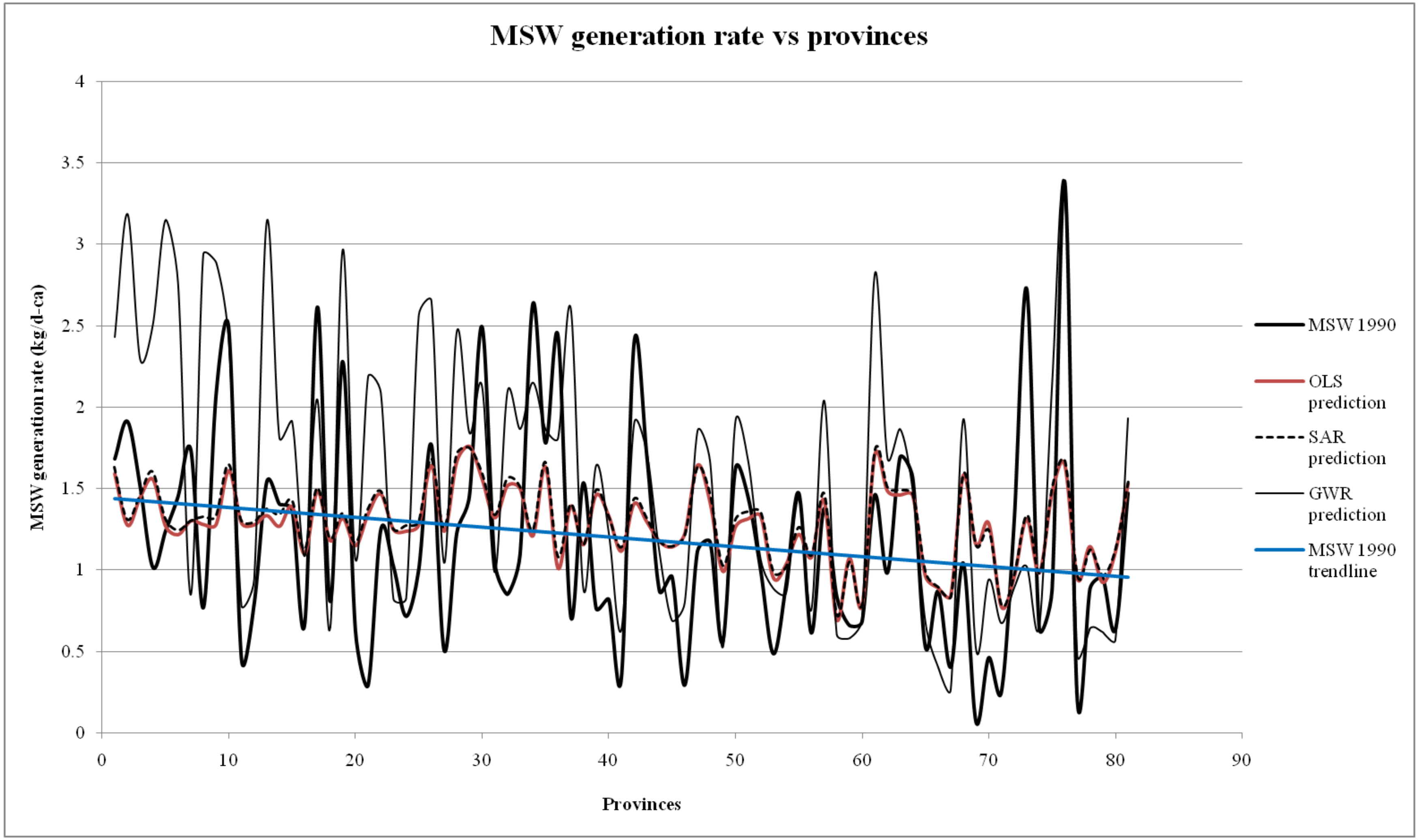


Figure 16: MSW generation rates in Turkey: Observed and predicted (1990)

Moreover, correlation between predicted and observed MSW generation rate was higher for GWR as found in validation step. This means it better demonstrates the variation of MSW generation rate among provinces arising from socio-economic, demographic and climatic differences. This result arises from local coefficients generated by GWR and it can be concluded that global coefficients are not representative at local level. Therefore, the effects of the factors included in this study spatially vary in Turkey. The differences between the models may further arise from spatial proximity matrix utilized. OLS and SAR models were run with the matrix based on four nearest neighbors to which equal weights were assigned. In GWR, the weights were assigned to each neighbor within the bandwidth regardless by the model itself. These weights were inversely related to the distance so that two data points were assigned equal weights if and only if they were equal distance away from the regression point. As said, setting a spatial proximity matrix is very essential. The method used by GWR can be applied to OLS and SAR to further inspect the effect of changing the method followed to obtain spatial proximity matrix. Other methods may also be experienced. For example, it would be interesting to make use of the distances of highway route between city centers.

As a result, GWR is rather an exploratory technique and global models might be more useful when the aim is to predict the MSW generation rate. GWR is of value when the effect of a variable in a specific province is in question. GWR results can be effectively used in determining waste reduction strategies.

5.3. SIGNIFICANT DETERMINANTS of MSW GENERATION RATE

A high number of explanatory variables were included in the analyses to find the significant factors. However, it is a best practice to find the best fitting model with minimum number of independent variables considering the difficulties in compiling reliable data. This study is guiding in terms of significant socio-economic, demographic and climatic factors.

Results of the models for DG2 and DG3 were found to be unreliable due to multicollinearity. This is why the coefficient estimates for DG1 should be taken into consideration to determine the significant factors affecting MSW generation rate in Turkey.

OLS and SAR results were similar to each other in terms of coefficient estimates. This similarity is also seen in Figure 16. The predicted MSW generation rates by OLS and SAR coincide with each other. The results show that unemployment is significantly related to MSW generation rate in Turkey. Higher education graduates ratio and asphalt-paved roads ratio are also significant determinants.

Some of the variables, which are not significant in global models, have high significance in specific regions. For example, temperature and rainfall are significant in the Aegean and Mediterranean Regions while they have high p-values globally (i.e. insignificant). In fact, temperature constitutes the most significant variables together with unemployment and higher education graduates ratio visually comparing the maps given in Figure 8. Similarly, IMR is noticeably insignificant in global models; however, its significance is high for provinces located in Middle Black Sea Region. Even, IMR is comparable to asphalt-paved roads ratio, which is significant in global models, in terms of number of provinces for which these variables are significant. Therefore, local coefficients can contradict with global coefficients in terms of level of significance and global results may be misleading in province scale.

As a result, the most outstanding variables are unemployment and higher education graduates ratio in terms of significance considering each of the models. As it is seen, the variables more directly related to consumption pattern significantly affect solid waste generation. On the other hand, the least significant variable as corroborated by both global and local methods is population density which is a demographic variable. Nevertheless, depending on the region within the country the significant and insignificant variables may change as pointed out by GWR.

For additional discussion, the differences between β coefficient estimates of GWR and SAR are demonstrated in Figure 17. The aim is to visually observe to what extent global effects of determinants deviate from the local effects.

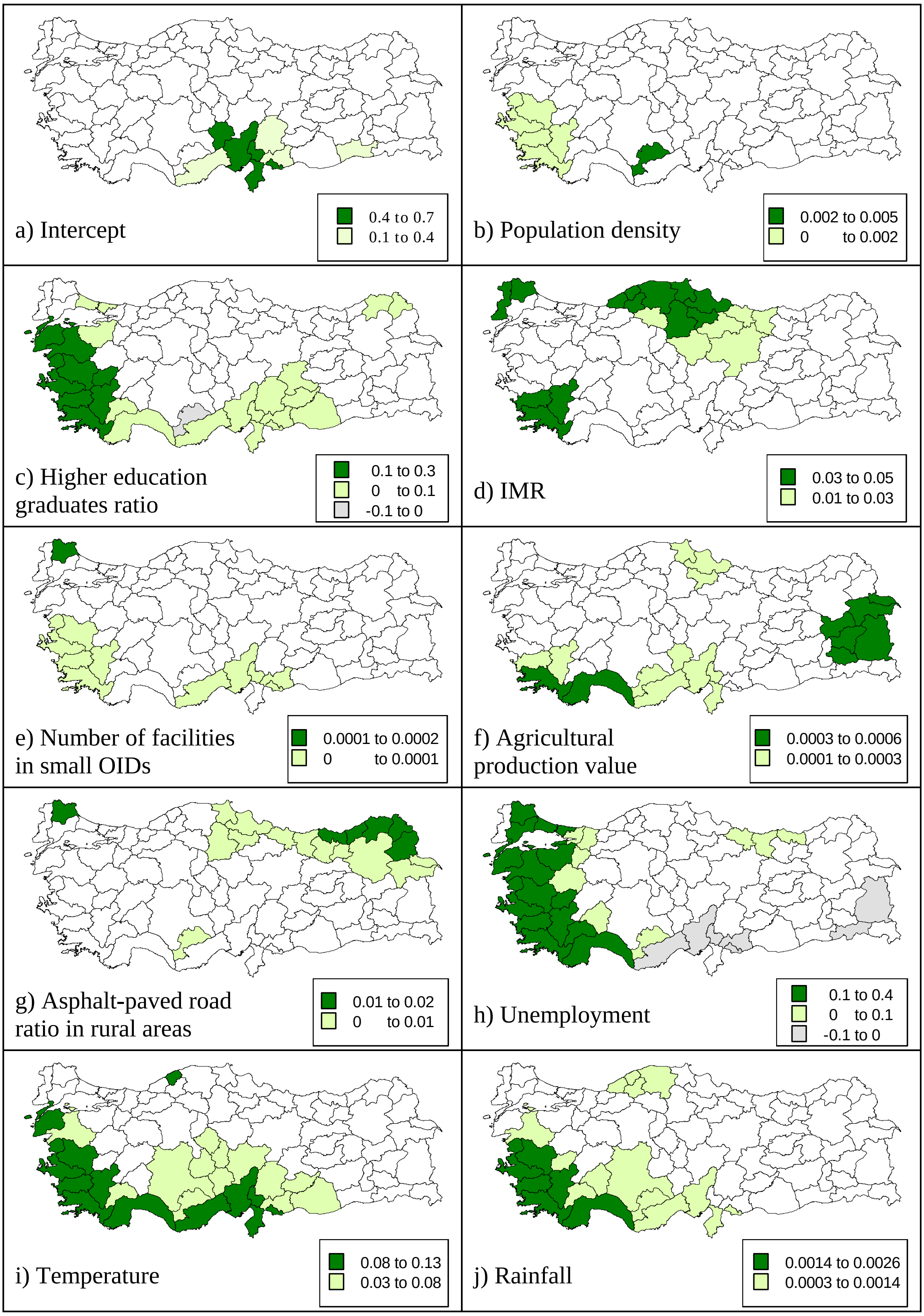


Figure 17: Difference between GWR and SAR coefficient estimates of DG1

The differences in Figure 17 were calculated by subtracting the absolute value of SAR coefficient estimate from the absolute value of GWR coefficient estimate. The negative values indicate that SAR overestimates the effect of the variable compared to GWR. Contrarily, the positive values refer to underestimation by SAR. The change in MSW generation rate as a result of unit change in any variable is higher for GWR compared to SAR for provinces where underestimation of SAR occurs. This is again a result of local variations in the effects of determinants.

In the maps given in Figure 17, it is seen that SAR underestimates the β coefficient estimates for the provinces which have significant coefficient estimates. Higher education graduates ratio and unemployment were found to have higher affect on MSW generation rate than predicted by SAR model especially in the western provinces. These maps indicate that SAR coefficient estimates may be inadequate at local level. Therefore, it is necessary to investigate local affects of the explanatory variables when local determinants are of interest.

Determining significant factors helps to understand the MSW generation process. Consequently, the methods detecting significant factors can be used as tools in attempts to reduce MSW waste generation. GWR is superior in that it appoints the regions where each factor is significant. For example, GWR revealed that education level is significant especially in the west and south coast line. The attempt should be to hamper the positive effect of high education level on solid waste generation in these regions. This can be accomplished via training programs, advertisements, publications, and etc. which raise awareness among educated people since their learning and innovation capacity is higher (Tsai, 2008). Meanwhile, the authorities should not waste effort for these programs in the eastern and central part since educational level is not a significant determinant in these regions. Moreover, high significance and negative impact of unemployment reveals that people tend to generate more waste as their buying power increases especially in the western part. Campaigns targeting the high-level consumers therefore gain importance with the purpose of increasing environmental responsibility among these consumers. In the eastern region and south coastline, the agricultural production value has significant effect on MSW generation rate. In these regions, precautions should be taken against generating farm product wastes. Training programs for farmers, fees, fines, etc. may be tools to reduce wastes arising from agricultural activities.

As a result, the measures for reducing solid waste generation in a region can be supported by the significant factors affecting waste generation in that region. Since GWR indicates regions where each variable is insignificant, unnecessary efforts caused by taking an insignificant factor into consideration can be avoided in these regions.

Significance of temperature and rainfall for some provinces designates the necessity of standardization in weight measurements of wastes to further reveal the effect of these variables and make use of them in SWM. The same result for asphalt-paved roads ratio, moreover, indicates that adding structural variables to the models might be beneficial.

Although DG2 and DG3 did not give dependable results, high significance of urbanization and average household size showed that these variables deserve further investigation.

CHAPTER 6

CONCLUSIONS & RECOMMENDATIONS FOR FUTURE WORK

6.1. CONCLUSIONS

Previous studies show that solid waste generation rates differ from one country to another. As a matter of fact, variations can be observed within a country, even within a city. These variations evidently arise from divergence of consumption patterns depending on socio-economic condition. The dependence on socio-economic factors varies by region as well in addition to the factor itself. This makes temporal and spatial variation of solid waste generation more complex. This complexity is further deepened by spatial dependency which is the case for most of the spatial data. Dennison, *et al.* (1996) highlighted the adjacent wards having similar waste generation rates. This can be the expectation considering the spatial distribution of administrative units in terms of development.

In this study, determinants of MSW generation rate -mostly socio-economic ones- were investigated paying attention to the possible spatial dependency. First, spatial autocorrelation structure of MSW generation rate was inspected. Then, both spatial and non-spatial analyses were carried out to explore the significant determinants of MSW generation rate and type of their effects. As a result, MSW generation rate was found to be spatially autocorrelated. However, this fact did not affect the results significantly at global scale. Local models, on the other hand, provided useful information on spatial variability of the determinants' effect on MSW generation rate. The results can be helpful in projection of waste amounts and also in planning of waste reduction activities.

Significant determinants of MSW generation rates in other countries may not be valid for Turkey. This is why this thesis study is of importance. Moreover, spatial dependency was considered in a study related to waste generation for the first time by applying SAR and GWR with this work. As a result, spatial analyses were introduced to SWM and the related methodology was constructed.

6.2. RECOMMENDATIONS FOR FUTURE WORK

The existing variables can be modified by means of transformation to strictly comply with the normality assumption. Besides, these can be replaced by alternative indicators symbolizing the same phenomenon. As an instance, ‘average years of education’ is an alternative to higher education graduates ratio. Purchasing power can also be represented by variables like average monthly income per household instead of unemployment. Structural variables, which are related to the waste management practices and policies can also be added and tested. These variables may stand for the variation in solid waste generation due to policy shifts and increase the validity of the models for long term.

A similar study in small scale can be performed to overcome complications in data. Sampling and surveying are efficient ways of data compiling as far as methods are appropriate. As a result, problems regarding availability and reliability of data are overcome since the researcher manages the data generating process. Composition data of solid waste can also be obtained for a detailed study. Survey questionnaires including attitude and behavior related questions further enable more informed interpretation of the results since this information help the researcher understand how people respond to the changes in the explanatory variables.

The models can also be improved through modifications. In the future work, weighting scheme similar to the bi-square decay function of GWR can be applied also for OLS and SAR considering the favorable AIC of GWR. On the other hand, GWR can be improved by using an adaptive kernel instead of fixed one. Local spatial autocorrelation can also be examined to stimulate outlier analysis. Moreover, there may be multicollinearity at local level although global multicollinearity is eliminated. Hence, local multicollinearity should be investigated prior to GWR analysis. Finally, the elasticity of GWR can be increased by applying mixed GWR.

REFERENCES

Abu Qdais, H., Hamoda, M. F. & Newham, J. (1997). Analysis of residential solid waste at generation sites. *Waste Management & Research, 15* (4), 395-406.

Adedibu, A. A. (1985). Spatial pattern of solid waste generation of Ilorin-Nigeria. *Environmental Monitoring and Assessment, 5* (3), 237-247.

Afon, A. O. (2007). An analysis of solid waste generation in a traditional African city: the example of Ogbomoso, Nigeria. *Environment and Urbanization, 19* (2), 527-537.

Afon, A. O. & Okewole, A. (2007). Estimating the quantity of solid waste generation in Oyo, Nigeria. *Waste Management & Research, 25* (4), 371-379.

Anselin, L. (1999). *Spatial Econometrics.* Center for Spatially Integrated Social Science. http://www.csiss.org/learning_resources/content/papers/baltchap.pdf. Last visited in December, 2009.

Anselin, L. (2002). Under the hood : Issues in the specification and interpretation of spatial regression models. *Agricultural Economics, 27* (3), 247-267 .

Ayadi, M. & Amara, M. (2008). Spatial patterns and geographic determinants of welfare and poverty in Tunisia. *ERF 15th Annual Conference: Equity and Economic Development.* 23-25 November 2008, Cairo, Egypt.

Bach, H., Mild, A., Natter, M. & Weber, A. (2004). Combining socio-demographic and logistic factors to explain the generation and collection of waste paper. *Resources, Conservation and Recycling, 41* (1), 65-73.

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

Baller, R. D., Anselin, L., Messner, S. F., Deane, G. & Hawkins, D. F. (2001). Structural covariates of US county homicide rates: Incorporating spatial effects. *Criminology, 39* (3), 561-590.

Banar, M. & Özkan, A. (2008). Characterization of the municipal solid waste in Eskisehir City, Turkey. *Environmetal Engineering Science, 25* (8), 1213-1219.

Bandara, N. J., Hettiaratchi, J. P., Wirasinghe, S. C. & Pilapiiya, S. (2007). Relation of waste generation and composition to socio-economic factors: a case study. *Environmental Monitoring and Assessment, 135* (1-3), 31-39.

Beigl, P., Wassermann, G., Schneider, F. & Salhofer, S. (2004). Forecasting municipal solid waste generation in major European cities. *iEMSs International Conference: ''Complexity and Integrated Resources Management.* 14-17 June 2004, Osnabrueck, Germany. http://www.iemss.org/iemss2004/pdf/regional/beigfore.pdf. Last visited in August, 2009.

Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: Wiley.

Benson, V. S., VanLeeuwen, J. A., Sanchez, J., Dohoo, I. R. & Somers, G. H. (2006). Spatial analysis of land use impact on ground water nitrate concentrations. *Journal of Environmental Quality, 35* (2), 421-432.

Berry, W. D. & Feldman, S. (1985). *Multiple Regression in Practice.* Beverly Hills: Sage Publications.

Biggeri, A., Barbone, F., Lagazio, C., Bovenzi, M. & Stanta, G. (1996). Air pollution and lung cancer in Trieste, Italy: Spatial analysis of risk as a function of distance from sources. *Environmental Health Perspectives, 104* (7), 750-754.

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

Borst, R. A. & Mccluskey, W. J. (2008). Using geographically weighted regression to detect housing submarkets: Modeling large-scale spatial variations in value. *Journal of Property Tax Assessment and Administration, 5* (1), 21-54 .

Brunsdon, C., Fotheringham, S. A. & Charlton, M. (1998). Geographically weighted regression - modelling spatial non- stationarity. *Journal of the Royal Statistical Society, Series D-The Statistician, 47* (3), 431-443.

Brunsdon, C., Fotheringham, S. A. & Charlton, M. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science, 39* (3), 497-524.

Buenrostro, D. O., Bocco, G. & Bernache, G. (2001). Urban solid waste generation and disposal in Mexico: a case study. *Waste Management & Research, 19* (2), s. 169-176.

Buenrostro, D. O., Márquez-Benavides, L. & Pinette, G. F. (2008). Consumption patterns and household hazardous solid waste generation in an urban settlement in México. *Waste Management, 28* (1), S2-S6.

Buenrostro, D. O., Ojeda-Benítez, S. & Márquez-Benavides, L. (2007). Comparative analysis of hazardous household waste in two Mexican regions. *Waste Management, 27* (6), 792-801.

CBRT (2009). *Exchange Rates.* Central Bank of the Republic of Turkey. http://www.tcmb.gov.tr/yeni/kurlar/kurlar_tr.php. Last visited in December, 2009.

Chen, H. V. & Chang, N.-B. (2000). Prediction analysis of solid waste generation based on grey fuzzy dynamic modeling. *Resources, Conservation and Recycling, 29* (1), 1-18.

Cho, S.-H., Kim, S. G., Roberts, R. K. & Jung, S. (2009). Amenity values of spatial configurations of forest landscapes over space and time in the Southern Appalachian Highlands. *Ecological Economics, 68* (10), 2646-2657.

Chowdhury, M. (2009). Searching quality data for municipal solid waste planning. *Waste Management, 29* (8), 2240-2247.

Daskapoulos, E., Badr, O. & Probert, S. D. (1998). Municipal solid waste: A prediction methodology for the generation rate and composition in the European Union countries and the United States of America. *Resources, Conservation and Recycling, 24* (2), 155-166.

Dennison, G. J., Dodd, V. A. & Whelan, B. (1996). A socio-economic based survey of household waste characteristics in the city of Dublin, Ireland: II. Waste Quantities. *Resources, Conservation and Recycling, 17* (3), 245-257.

Dinçer, B., Özaslan, M. & Kavasoğlu, T. (2003). *İllerin sosyo-ekonomik gelişmişlik sıralaması araştırması.* State Planning Organization, General Directorate of Regional Development and Structural Adjustment. http://www.dpt.gov.tr/DocObjects/Download /3116/2003-05.pdf. Last visited in February, 2009.

Dinçer, B., Özaslan, M. & Satılmış, E. (1996). *İllerin sosyo-ekonomik gelişmişlik sıralaması araştırması.* State Planning Organization, General Directorate of Regional Development and Structural Adjustment. http://ekutup.dpt.gov.tr/bolgesel/dincerb/il/. Last visited in November, 2009.

Düzgün, H. Ş. & Kemeç, S. (2008). Spatial and Geographically Weighted Regression. S. Shekhar, & H. Xiong, *Encyclopedia of GIS.* New York: Springer.

Dyson, B. & Chang, N. B. (2005). Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling. *Waste Management, 25* (7), 669-679.

Ebreo, A., Hershey, J. & Vining, J. (1999). Reducing solid waste: Linking recycling to environmentally responsible consumerism. *Environment and Behavior, 31* (1), 107-135.

EEA (2009). *Municipal Waste Generation: Assessment.* European Environment Agency. http://themes.eea.europa.eu/IMS/ISpecs/ISpecification20041007131809/ IAssessment1228140943395/view_content. Last visited in December, 2009.

Emery, A. D., Griffiths, A. J. & Williams, K. P. (2003). An in depth study of the effects of socio-economic conditions on household waste recycling practices. *Waste Management & Research, 21* (3), 180-190.

ESRI (1998). *ESRI Shapefile Technical Description.* ESRI. http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf. Last visited in November, 2009.

Fotheringham, S. A., Brunsdon, C. & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* Chichester, West Sussex, England; Hoboken, NJ: Wiley

Freund, R. J. & Wilson, W. J. (1998). *Regression Analysis: Statistical Modeling of a Response Variable.* San Diego, USA: Academic Press.

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

Gay, A. E., Beam, T. G. & Mar, B. W. (1993). Cost-effective solid waste characterization methodology. *Journal of Environmental Engineering, 119* (4), 631-644.

Ge, Z. (2009). Effectiveness of the t-test in multiple linear regression modeling of environmental systems. *Environmental Engineering Science, 26* (2), 277-383.

Gidarakos, E., Havas, G. & Ntzamilis, P. (2006). Municipal solid waste composition determination supporting the integrated solid waste management system in the island of Crete. *Waste Management, 26* (6), 668-679.

Godbey, G., Lifset, R. & Robinson, J. (1998). No Time to Waste: An exploration of time use, attitudes toward time, and the generation of municipal solid waste. *Social Research , 65* (1), 101-140.

Gómez, G., Meneses, M., Ballinas, L. & Castells, F. (2009). Seasonal characterization of municipal solid waste (MSW) in the city of Chihuahua, Mexico. *Waste Management, 29* (7), 2018-2024.

Griffith, D. A. & Amrhein, C. G. (1997). *Multivariate Statistical Analysis for Geographers.* Upper Saddle River, N.J.: Prentice Hall.

Harms, T. K., Wentz, E. A. & Grimm, N. B. (2009). Spatial heterogeneity of denitrification in semi-arid floodplains. *Ecosystems, 12* (1), 129-143.

Hibiki, A. & Shimane, T. (2006). Empirical study on determinants of household solid waste and the effect of the unit pricing in Japan. *Annual Meeting of Japanese Economic Association, Osaka City University.* 21-22 October 2006, Osaka, Japan.

Hockett, D., Lober, D. J. & Pilgrim, K. (1995). Determinants of per capita municipal solid waste generation in the Southeastern United States. *Journal of Environmental Management, 45* (3), 205-217.

Huang, Y. & Leung, Y. (2002). Analysing regional industrialisation in Jiangsu province using geographically weighted regression. *Journal of Geographical Systems, 4* (2), 233-249.

Hudson, N., Baker, A., Ward, D., Reynolds, D. M., Brunsdon, C., Carliell-Marquet, C. & Browning, S. (2008). Can fluorescence spectrometry be used as a surrogate for the Biochemical Oxygen Demand (BOD) test in water quality assessment? An example from South West England. *Science of the Total Environment, 391* (1), 149-158.

Joosten, L. A., Hekkert, M. P. & Worrell, E. (2000). Assessment of the plastic flows in The Netherlands using STREAMS. *Resources, Conservation and Recycling, 30* (2), 135-161.

Karadimas, N. V. & Loumos, V. G. (2008). GIS-based modelling for the estimation of municipal solid waste generation and collection. *Waste Management & Research, 26* (4), 337-346.

Karavezyris, V., Timpe, K.-P. & Marzi, R. (2002). Application of system dynamics and fuzzy logic to forecasting of municipal solid waste. *Mathematics and Computers in Simulation, 60* (3), 149-158.

Katsamaki, A., Willems, S. & Diamadopoulos, E. (1998). Time series analysis of municipal solid waste generation rates. *Journal of Environmental Engineering, 124* (2), 178-183.

Keil, P., Simova, I. & Hawkins, B. A. (2008). Water-energy and the geographical species richness pattern of European and North African dragonflies (Odonata). *Insect Conservation and Diversity, 1* (3), 142-150.

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

Kreft, H. & Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. *PNAS, 104* (14), 5925-5930 .

Latimer, A. M., Banerjee, S., Sang, H., Mosher, E. S. & Silander, J. A. (2009). Hierarchical models facilitate spatial analysis of large data sets: a case study on invasive plant species in the northeastern United States. *Ecology Letters, 12* (2), 144-154.

Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology, 74* (6), 1659-1673.

Leung, Y., Mei, C.L. & Zhang, W.X. (2000). Testing for spatial autocorrelation among the residuals of the geographically weighted regression. *Environment and Planning A, 32* (5), 871-890.

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

Linzner, R. & Beigl, P (2005). Analysing the waste generation in the city of Belgrade. *Proceedings of the International Symposium: "Environment for Europe".* 6-8 June 2005, Belgrade, Serbia and Montenegro.

MapInfo (2009). *MapInfo Professional.* Pitney Bowes Business Insight. http://www.pbinsight.com/products/location-intelligence/applications/mapping-analytical/mapinfo-professional/. Last visited in January, 2010.

Mennis, J. L. & Jordan, L. (2005). The distribution of environmental equity: exploring spatial nonstationarity in multivariate models of air toxic releases. *Annals of the Association of American Geographers, 95* (2), 249-268.

Miles, J. & Shevlin, M. (2001). *Applying Regression & Correlation: A Guide for Students and Researchers.* London ; Thousand Oaks, Calif.: Sage Publications

MoEF (2008). *Solid Waste Management Action Plan: 2008-2012.* Ministry of Environment and Forestry, General Directorate of Environmental Management. http://www.atikyonetimi.cevreorman.gov.tr/belge/atikeylemplani.pdf

MoIT (2009). *Small OIDs and Establishment Targets.* Ministry of Industry and Trade. http://www.sanayi.gov.tr/webedit/gozlem.aspx?sayfano=1429. Last visited in November, 2009.

Navarro-Esbrí, J., Diamadopoulos, E. & Ginestar, D. (2002). Time series analysis and forecasting techniques for municipal solid waste management. *Resources, Conservation and Recycling, 35* (3), 201-214 .

Ojeda-Benítez, S., Armijo-de Vega, C. & Marquez-Montenegro, M. Y. (2008). Household solid waste characterization by family socioeconomic profile as unit of analysis. *Resources, Conservation and Recycling, 52* (7), 992-999.

Olgun, M. & Erdoğan, S. (2009). Modeling crop yield potential of Eastern Anatolia by using geographically weighted regression. *Archives of Agronomy and Soil Science, 55* (3), 255-263.

Ordóñez-Ponce, E., Samarasinghe, S., & Torgerson, L. (2004). A model for assessing waste generation factors and forecasting waste generation using artificial neural networks: A case study of Chile. *Proceedings of 'Waste and Recycle 2004' Conference*. 21-24 September 2004, Fremantle, Australia.

Parfitt, J. P. & Flowerdew, R. (1997). Methodological problems in the generation of household waste statistics. *Applied Geography, 17* (3), 231-244.

Park, H. M. (2003). *Multicollinearity in Regression Models.* Jeeshim and KUCC625. http://www.masil.org/documents/multicollinearity.pdf. Last visited in November, 2009.

Patel, M. K., Jochem, E., Radgen, P. & Worrell, E. (1998). Plastic streams in Germany – an analysis of production, consumption and waste generation. *Resources, Conservation and Recycling, 24* (3-4), 191-215.

Pope, C. III, Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K. & Thurston, G. D. (2002). Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution. *JAMA, 287* (9), 1132-1141.

Portnov, B. A., Dubnov, J. & Barchana, M. (2009). Studying the association between air pollution and lung cancer incidence in a large metropolitan area using a kernel density function. *Socio-Economic Planning Sciences, 43* (3), 141-150.

Purcell, M. & Magette, W. L. (2009). Prediction of household and commercial BMW generation according to socio-economic and other factors for the Dublin region. *Waste Management, 29* (4), 1237-1250.

Rey, S. J. & Montouri, B. D. (1999). US regional income convergence: A spatial econometric perspective. *Regional Studies, 33* (2), 143-156.

ScapeToad (2008). *ScapeToad-v11.* Choros Laboratory. http://scapetoad.choros.ch/index.php. Last visited in December, 2009.

Shearmur, R., Apparicio, P., Lizion, P. & Polése, M. (2007). Space, time, and local employment growth: An application of spatial regression analysis. *Growth and Change, 38* (4), 696-722.

Sparks, P. J. & Sparks, C. S. (2009). An application of spatially autoregressive models to the study of US county mortality rates. *Population, Space and Place, 15* (0), 000-000.

Sterner, T. & Bartelings, H. (1999). Household waste management in a Swedish municipality: Determinants of waste disposal, recycling and composting. *Environmental and Resource Economics, 13* (4), 473-491.

Sudhir, V., Srinivasan, G. & Muraleedharan, V. R. (1997). Planning for sustainable solid waste management in urban India. *System Dynamics Review, 13* (3), 223-246.

Svenning, J.-C., Normand, S. & Flemming, S. (2009). Plio-Pleistocene climate change and geographic heterogeneity in plant diversity - environment relationships. *Ecography, 32* (1), 13-21.

Tağıl, Ş. (2007). Balıkesir'de hava kirliliğinin solunum yolu hastalıklarının mekansal dağılışı üzerine etkisini anlamada jeo-istatistik teknikler. *Coğrafi Bilimler Dergisi, 5* (1), 37-56.

TCA (2007). *Performance Audit Reports.* Turkish Court of Accounts. http://www.sayistay.gov.tr/rapor/perdenrap/2007/2007-1AtikYonetimi/2007-Atik_ Yonetimi _Raporu.pdf. Last visited in January, 2010.

Terribile, L. C. & Diniz-Filho, J. A. (2009). Spatial patterns of species richness in New World coral snakes and the metabolic theory of ecology. *Acta Oecologica, 35* (2), 163-173.

R-project (2009). *The R Project for Statistical Computing.* R Development Core Team. http://www.r-project.org/. Last visited in January, 2010.

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

Tsai, T.H. (2008). The impact of social capital on regional waste recycling. *Sustainable Development, 16* (1), 44-55.

Tu, J. & Xia, Z.-G. (2008). Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. *Science of the Total Environment, 407* (1), 358-378.

TurkStat (2009a). *Environment Statistics.* Turkish Statistical Institute. http://www.tuik. gov.tr/VeriBilgi.do?tb_id=10&ust_id=3. Last visited in January, 2010.

TurkStat (2009b). *Regional Statistics.* Turkish Statistical Institute. http://tuikapp. tuik.gov.tr/Bolgesel/menuAction.do?dil=en. Last visited in January, 2010.

TurkStat (2009c). *Databases: 1990 Census Results.* Turkish Statistical Institute. http://tuikapp.tuik.gov.tr/nufus90app/idari.zul. Last visited in January, 2010.

TurkStat (2009d). *ICT Usage Statistics.* Turkish Statistical Institute. http://www.turkstat.gov.tr/VeriBilgi.do?tb_id=60&ust_id=2. Last visited in December, 2009.

TurkStat (2009e). *Energy Statistics.* Turkish Statistical Institute. http://www.tuik.gov.tr/VeriBilgi.do?tb_id=11&ust_id=3. Last visited in December, 2009.

Vijay, R., Gupta, A., Kalamdhad, A. S. & Devotta, S. (2005). Estimation and allocation of solid waste to bin through geographical information systems. *Waste Management & Research, 23* (5), 479-484.

Walford, N. (1995). *Geographical Data Analysis.* Chichester, New York: Wiley.

Wentz, E. A. & Gober, P. (2007). Determinants of small-area water consumption for the city of Phoenix, Arizona. *Water Resources Management, 21* (11), 1849-1863.

Wesolowsky, G. O. (1976). *Multiple Regression and Analysis of Variance.* New York: Wiley.

Wheeler, D. C. (2007). Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning, 39* (10), 2464-2481.

Wheeler, D. & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems, 7* (2), 161-187.

Wooldridge, S., Brodie, J. & Furnas, M. (2006). Exposure of inner-shelf reefs to nutrient enriched runoff entering the Great Barrier Reef Lagoon: Post-European changes and the design of water quality targets. *Marine Pollution Bulletin, 52* (11), 1467-1479.

Yamanoğlu, K. B. (2009). Türkiye'de iller arası mali yakınsama. *EconAnadolu 2009: Anadolu International Conference in Economics.* 17-19 June 2009, Eskişehir, Turkey.

Yildirim, J., Öcal, N., & Özyildirim, S. (2009). Income inequality and economic convergence in Turkey: A spatial effect analysis. *International Regional Science Review, 32* (2), 221-254.

Zhou, X., & Lin, H. (2008). Moran's I. S. Shekhar, & H. Xiong, *Encyclopedia of GIS.* New York: Springer.

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology, 44* (1), 41-61.

APPENDIX A

SUPPLEMENTARY INFORMATION

A.1. Calculation of cartogram size error (ScapeToad, 2008)

$$Cartogram\ size\ error = \frac{A_{th}}{A} \times 100 \quad \text{(A.1)}$$

A_{th}: the theoretical area after rescaling
A: effective area of the polygon

A.2. Correction of agricultural production value

The exchange rates: (CBRT, 2009)

30th June 1993: 1 $ = 10860.24 TL
30th June 2000: 1 $ = 618985 TL

$$APV_c = APV \times \frac{618985}{10860.24} = APV \times 57 \quad \text{(A.2)}$$

APV_c: corrected agricultural production value for 1993
APV: agricultural production value in 1993

A.3. AIC calculation (Zucchini, 2000)

$$AIC = -2log\ (L) + 2(k+1) \quad \text{(A.3)}$$

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

A.4. F-value calculation for $\mathbf{R^2}$ (Wesolowsky, 1976)

$$F = \frac{n-(k+1)}{(k+1)-1} \cdot \frac{R^2}{1-R^2} \tag{A.4}$$

n: number of samples (observations)

k: number of explanatory variables

A.5. Detailed GWR model in matrix notation (Fotheringham, *et al.*, 2002)

$$Y = (\beta \otimes X)1 + \epsilon \tag{A.5a}$$

$\otimes$: 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

$$\beta = \begin{bmatrix} \beta_o(u_1,v_1) & \beta_1(u_1,v_1) & \cdots & \beta_k(u_1,v_1) \\ \beta_o(u_2,v_2) & \beta_1(u_2,v_2) & \cdots & \beta_k(u_2,v_2) \\ \cdots & \cdots & \cdots & \cdots \\ \beta_o(u_n,v_n) & \beta_1(u_n,v_n) & \cdots & \beta_k(u_n,v_n) \end{bmatrix} \tag{A.5b}$$

$$\hat{\beta}(i) = (X^T W(i) X)^{-1} X^T W(i) Y \tag{A.5c}$$

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

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

(u_n, v_n): coordinates of regression point n

$$W(i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix} \tag{A.5d}$$

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

A.6. Cross-validation of kernel bandwidth (Fotheringham, *et al.*, 2002)

$$CV = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(b)]^2 \tag{A.6}$$

$\hat{y}_{\neq i}(b)$: y_i fitted by omitting the point *i* in the calibration process

n: number of samples (observations)

A.7. Calculation of R^2_{adj} (Wesolowsky, 1976)

$$R^2{}_{adj} = 1 - \frac{n-1}{n-(k+1)}(1 - R^2) \tag{A.7}$$

R^2: coefficient of determination

n: number of samples (observations)

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

A.8. Cartogram report

```
CARTOGRAM COMPUTATION REPORT

CARTOGRAM PARAMETERS:
Cartogram layer: Turkey
Cartogram attribute: MSWGR
Attribute type: Population value
Transformation quality: 50 of 100
Cartogram grid size: 600 x 197
Diffusion grid size: 256
Diffusion iterations: 3

CARTOGRAM LAYER & ATTRIBUTE STATISTICS:
Number of features: 81
Attribute mean value: 1.3329629629629631
Attribute minimum value: 0.27
Attribute maximum value: 3.25
```

```
SIMULTANEOUSLY TRANSFORMED LAYERS:
None

CONSTRAINED DEFORMATION LAYERS:
None
CARTOGRAM ERROR
The cartogram error is a measure for the quality of the result.
Mean cartogram error: 99.5256799603899
Standard deviation: 9.129832245368211
25th percentile: 94.24596289914822
50th percentile: 100.10245407393785
75th percentile: 104.59496718169545
Features with mean error +/- 1 standard deviation: 63 of 81 (78%)

Computation time: 314 seconds
```

APPENDIX B

SCRIPTS

B.1. Constructing DG3

```
> library(leaps)
> filepath<-system.file("data", "theses_data.txt" , package="datasets")
> d<-read.table(filepath)
> mcl<-matrix(nrow=81,ncol=20)
> mcl[,1]<-d$V4;  mcl[,2]<-d$V5;  mcl[,3]<-d$V6;  mcl[,4]<-d$V7
> mcl[,5]<-d$V8;  mcl[,6]<-d$V9;  mcl[,7]<-d$V10;  mcl[,8]<-d$V11
> mcl[,9]<-d$V12;  mcl[,10]<-d$V13;  mcl[,11]<-d$V14;  mcl[,12]<-d$V15
> mcl[,13]<-d$V16;  mcl[,14]<-d$V17;  mcl[,15]<-d$V18;  mcl[,16]<-d$V19
> mcl[,17]<-d$V20;  mcl[,18]<-d$V21;  mcl[,19]<-d$V22;  mcl[,20]<-d$V23
> leaps(mcl, d$V3,int=TRUE, method=c("adjr2", "r2"), nbest=1)
> library(ltm)
> rcor.test(mcl,method = "pearson")
```

B.2. Multicollinearity Analysis

```
> library(car)
> filepath<-system.file("data", "theses_data.txt" , package="datasets")
> d<-read.table(filepath)
>
> #DG1
> vif(lm(d$V3 ~ d$V4 +d$V5 +d$V6 +d$V7 +d$V8 +d$V9 +d$V10 + d$V11
+d$V12 +d$V13 +d$V14 +d$V15 +d$V16 +d$V17 +d$V18 +d$V19 +d$V20
+d$V21 +d$V22 +d$V23))
```

```
> vif(lm(d$V3 ~ d$V4 +d$V5 +d$V6 +d$V7 +d$V8 +d$V9 +d$V10 + d$V11
+d$V12 +d$V13 +d$V14 +d$V15 +d$V17 +d$V18 +d$V19 +d$V20 +d$V21
+d$V22 +d$V23))
> vif(lm(d$V3 ~ d$V4 +d$V5 +d$V6 +d$V7 +d$V8 +d$V9 +d$V10 + d$V11
+d$V12 +d$V13 +d$V15 +d$V17 +d$V18 +d$V19 +d$V20 +d$V21 +d$V22
+d$V23))
> vif(lm(d$V3 ~ d$V4 +d$V5 +d$V6 +d$V7 +d$V8 +d$V9 +d$V10 + d$V11
+d$V12 +d$V13 +d$V15 +d$V17 +d$V18 +d$V20 +d$V21 +d$V22 +d$V23))
> vif(lm(d$V3 ~ d$V4 +d$V5 +d$V7 +d$V8 +d$V9 +d$V10 + d$V11 +d$V12
+d$V13 +d$V15 +d$V17 +d$V18 +d$V20 +d$V21 +d$V22 +d$V23))
> vif(lm(d$V3 ~ d$V4 +d$V5 +d$V8 +d$V9 +d$V10 + d$V11 +d$V12 +d$V13
+d$V15 +d$V17 +d$V18 +d$V20 +d$V21 +d$V22 +d$V23))
> vif(lm(d$V3 ~ d$V4 +d$V5 +d$V8 +d$V10 + d$V11 +d$V12 +d$V13 +d$V15
+d$V17 +d$V18 +d$V20 +d$V21 +d$V22 +d$V23))
> vif(lm(d$V3 ~ d$V5 +d$V8 +d$V10 + d$V11 +d$V12 +d$V13 +d$V15 +d$V17
+d$V18 +d$V20 +d$V21 +d$V22 +d$V23))
> vif(lm(d$V3 ~ d$V5 +d$V8 + d$V11 +d$V12 +d$V13 +d$V15 +d$V17 +d$V18
+d$V20 +d$V21 +d$V22 +d$V23))
> vif(lm(d$V3 ~ d$V5 +d$V8 + d$V11 +d$V12 +d$V13 +d$V15 +d$V17 +d$V18
+d$V21 +d$V22 +d$V23))
> # two variables eliminated according to correlation matrix
> a<-vif(lm(d$V3 ~ d$V5 + d$V11 +d$V12 +d$V13 +d$V15 +d$V18 +d$V21
+d$V22 +d$V23))
> mean(a)
>
> #DG2
> b<-vif(lm(d$V3 ~ d$V4 +d$V6 + d$V11 +d$V12 +d$V14 +d$V15 +d$V21
+d$V22 +d$V23))
> mean(b)
>
> #DG3
```

```
> c<-vif(lm(d$V3 ~ d$V4 +d$V6 + d$V8 +d$V12 +d$V13 +d$V18 +d$V19
+d$V21 +d$V22))
> mean(c)
>
> #correlation matrix
> data.cor<-cor(mcl,method = "pearson")
> cov2cor(data.cor)
```

B.3. OLS Regression

```
> filepath<-system.file("data", "theses_data.txt" , package="datasets")
> d<-read.table(filepath)
> #DG1
> lm_dg1<-lm(d$V3 ~ d$V5 + d$V11 +d$V12 +d$V13 +d$V15+ d$V18 +d$V21
+d$V22 +d$V23)
> summary(lm_dg1)
> AIC(lm_dg1)
>
> #DG2
> lm_dg2<-lm(d$V3 ~ d$V4 +d$V6 + d$V11 +d$V12 +d$V14 +d$V15 +d$V21
+d$V22 +d$V23)
> summary(lm_dg2)
> AIC(lm_dg2)
>
> #DG3
> lm_dg3<-lm(d$V3 ~ d$V4 +d$V6 + d$V8 +d$V12 +d$V13 +d$V18 +d$V19
+d$V21 +d$V22)
> summary(lm_dg3)
> AIC(lm_dg3)
```

B.4. Constructing Spatial Proximity Matrix and Calculating Moran's I

```
>library(spdep)
> filepath<-system.file("data", "theses_data.txt" , package="datasets")
> d<-read.table(filepath)
> # 4-nearest neighbors
> x<-matrix(nrow=81,ncol=2)
> x[,1]<-d$V1
> x[,2]<-d$V2
> knn<-knearneigh(x, k=4, longlat = TRUE)
> knn.nb<-knn2nb(knn, row.names = NULL, sym = FALSE)
> knn.listw<-nb2listw(knn.nb, style="W")
> # contiguity
> con<-readShapePoly(system.file("etc/shapes/Turkey.shp",package="spdep")[1])
> con.nb<-poly2nb(con)
> con.listw<-nb2listw(con.nb, style="W")
> # Moran's I for MSW generation rate
> mswgr<-matrix(nrow=81,ncol=1)
> mswgr[,1]<-d$V3
> # 4-nearest neighbors
> moran(mswgr, knn.listw, length(knn.nb), Szero(knn.listw))
> moran.test(mswgr,knn.listw)
> moran.mc(mswgr, listw=knn.listw, nsim=9999999) # monte carlo test
> # contiguity
> moran(mswgr, con.listw, length(con.nb), Szero(con.listw))
> moran.test(mswgr,con.listw)
> moran.mc(mswgr, listw=con.listw, nsim=9999999) # monte carlo test
> # Moran's I for OLS residuals
> # DG1
> lm.morantest(lm_dg1, nb2listw(knn.nb, style="W"))
> # DG2
> lm.morantest(lm_dg2, nb2listw(knn.nb, style="W"))
> # DG3
```

```
> lm.morantest(lm_dg3, nb2listw(knn.nb, style="W"))
```

B.5. LM Tests

```
> library(spdep)
> lm.LMtests(lm_dg1, nb2listw(knn.nb, style="W"), test=c("LMerr", "LMlag",
"RLMerr", "RLMlag"))
> lm.LMtests(lm_dg2, nb2listw(knn.nb, style="W"), test=c("LMerr", "LMlag",
"RLMerr", "RLMlag"))
> lm.LMtests(lm_dg3, nb2listw(knn.nb, style="W"), test=c("LMerr", "LMlag",
"RLMerr", "RLMlag"))
```

B.6. SAR Model

```
> #DG1
> sarlag1<-lagsarlm(d$V3 ~ d$V5 + d$V11 +d$V12 +d$V13 +d$V15 +d$V18
+d$V21 +d$V22 +d$V23, listw=nb2listw(knn.nb, style="W"), method="eigen",
quiet=FALSE)
> summary(sarlag1)
> SST<- sum((d$V3-mean(d$V3))^2)
> R2<-1-(sarlag1$SSE/SST)
> sarerr1<-errorsarlm(d$V3 ~ d$V5 + d$V11 +d$V12 +d$V13 +d$V15 +d$V18
+d$V21 +d$V22 +d$V23, listw=nb2listw(knn.nb, style="W"), method="eigen",
quiet=FALSE)
> summary(sarerr1)
> SST<- sum((d$V3-mean(d$V3))^2)
> R2<-1-(sarerr1$SSE/SST)
>
> #DG2
> sarlag2<-lagsarlm(d$V3 ~ d$V4 +d$V6 + d$V11 +d$V12 +d$V14 +d$V15
+d$V21 +d$V22 +d$V23, listw=nb2listw(knn.nb, style="W"), method="eigen",
quiet=FALSE)
```

```
> summary(sarlag2)
> SST<- sum((d$V3-mean(d$V3))^2)
> R2<-1-(sarlag2$SSE/SST)
> sarerr2<-errorsarlm(d$V3 ~ d$V4 +d$V6 + d$V11 +d$V12 +d$V14 +d$V15
+d$V21 +d$V22 +d$V23, listw=nb2listw(knn.nb, style="W"), method="eigen",
quiet=FALSE)
> summary(sarerr2)
> SST<- sum((d$V3-mean(d$V3))^2)
> R2<-1-(sarerr2$SSE/SST)
>
> #DG3
> sarlag3<-lagsarlm(d$V3 ~ d$V4 +d$V6 + d$V8 +d$V12 +d$V13 +d$V18
+d$V19 +d$V21 +d$V22, listw=nb2listw(knn.nb, style="W"), method="eigen",
quiet=FALSE)
> summary(sarlag3)
> SST<- sum((d$V3-mean(d$V3))^2)
> R2<-1-(sarlag3$SSE/SST)
> sarerr3<-errorsarlm(d$V3 ~ d$V4 +d$V6 + d$V8 +d$V12 +d$V13 +d$V18
+d$V19 +d$V21 +d$V22, listw=nb2listw(knn.nb, style="W"), method="eigen",
quiet=FALSE)
> summary(sarerr3)
> SST<- sum((d$V3-mean(d$V3))^2)
> R2<-1-(sarerr3$SSE/SST)
> # Moran’s I for residuals
> #DG1
> moran.test(sarlag1$residuals, knn.listw)
> moran.test(sarerr1$residuals, knn.listw)
> #DG2
> moran.test(sarlag2$residuals, knn.listw)
> moran.test(sarerr2$residuals, knn.listw)
> #DG3
> moran.test(sarlag3$residuals, knn.listw)
```

```
> moran.test(sarerr3$residuals, knn.listw)
```

B.7. GWR model

```
> library(spgwr)
> #DG1
> gwr.bw<-gwr.sel(d$V3 ~ d$V5 + d$V11 +d$V12 +d$V13 +d$V15 +d$V18
+d$V21 +d$V22 +d$V23,coords=cbind(d$V1,d$V2), longlat=TRUE,
gweight=gwr.bisquare)
> gwr.dg1<-gwr(d$V3 ~ d$V5 + d$V11 +d$V12 +d$V13 +d$V15 +d$V18 +d$V21
+d$V22 +d$V23,coords=cbind(d$V1,d$V2), bandwidth=gwr.bw, hatmatrix=TRUE,
gweight=gwr.bisquare)
> gwr.dg1
> gwr.dg1$SDF
> gwr.morantest(gwr.dg1, nb2listw(knn.nb, style="W"))
>
> #DG2
> gwr.bw<-gwr.sel(d$V3 ~ d$V4 +d$V6 + d$V11 +d$V12 +d$V14 +d$V15
+d$V21 +d$V22 +d$V23,coords=cbind(d$V1,d$V2), longlat=TRUE,
gweight=gwr.bisquare)
> gwr.dg2<-gwr(d$V3 ~ d$V4 +d$V6 + d$V11 +d$V12 +d$V14 +d$V15 +d$V21
+d$V22 +d$V23,coords=cbind(d$V1,d$V2), bandwidth=gwr.bw, hatmatrix=TRUE,
gweight=gwr.bisquare)
> gwr.dg2
> gwr.dg2$SDF
> gwr.morantest(gwr.dg2, nb2listw(knn.nb, style="W"))
>
> #DG3
> gwr.bw<-gwr.sel(d$V3 ~ d$V4 +d$V6 + d$V8 +d$V12 +d$V13 +d$V18 +d$V19
+d$V21 +d$V22,coords=cbind(d$V1,d$V2), longlat=TRUE, gweight=gwr.bisquare)
```

```
> gwr.dg3<-gwr(d$V3 ~ d$V4 +d$V6 + d$V8 +d$V12 +d$V13 +d$V18 +d$V19
+d$V21 +d$V22,coords=cbind(d$V1,d$V2), bandwidth=gwr.bw, hatmatrix=TRUE,
gweight=gwr.bisquare)
> gwr.dg3
> gwr.dg3$SDF
> gwr.morantest(gwr.dg3, nb2listw(knn.nb, style="W"))
```

B.8. OLS regression with eliminated neighbors' effect

```
> filepath<-system.file("data", "data_version2.txt" , package="datasets")
> d<-read.table(filepath)
> lm_dg1<-lm(d$V3 ~ d$V4 + d$V5 +d$V6 +d$V7 +d$V8+ d$V9 +d$V10 +d$V11
+d$V12)
> summary(lm_dg1)
> AIC(lm_dg1)
> x<-matrix(nrow=20,ncol=2)
> x[,1]<-d$V1
> x[,2]<-d$V2
> knn<-knearneigh(x, k=4, longlat = TRUE)
> data.nb<-knn2nb(knn, row.names = NULL, sym = FALSE)
> lm.morantest(lm_dg1, nb2listw(data.nb, style="W"))
> #Moran's I test for the subset with whole-population OLS coefficients
> ver2<-matrix(nrow=20,ncol=10)
> ver2[,1]<-v$V4
> ver2[,2]<-v$V5
> ver2[,3]<-v$V6
> ver2[,4]<-v$V7
> ver2[,5]<-v$V8
> ver2[,6]<-v$V9
> ver2[,7]<-v$V10
> ver2[,8]<-v$V11
> ver2[,9]<-v$V12
```

```
> ver2[,10]<-v$V13
> y=ver2%*%b
> res=d$V3-y
> moran.test(res, nb2listw(data.nb, style="W"))
```

B.9. Validation

```
> #lm validation
> b<-c(1.2, -3.809e-05, 5.460e-02, 5.792e-04, -5.553e-05, 7.540e-05, 5.096e-03, -
5.926e-02, -1.060e-02, 1.156e-04)
> filepath<-system.file("data", "data_val.txt" , package="datasets")
> v<-read.table(filepath)
> data1990<-matrix(nrow=81,ncol=10)
> data1990[,1]<-v$V1
> data1990[,2]<-v$V2
> data1990[,3]<-v$V3
> data1990[,4]<-v$V4
> data1990[,5]<-v$V5
> data1990[,6]<-v$V6
> data1990[,7]<-v$V7
> data1990[,8]<-v$V8
> data1990[,9]<-v$V9
> data1990[,10]<-v$V10
> y=data1990%*%b
> rmse=sqrt(sum((v$V11-y)^2)/81)
> cor(v$V11, y)
>
> #sarlag validation
> b<-c(9.6861e-01, -1.3445e-05, 5.0675e-02, 1.0970e-03, -6.5924e-05, 7.4311e-05,
4.5832e-03, -5.1938e-02, -1.0055e-02, 1.1558e-04)
> rho=0.15447
> filepath<-system.file("data", "theses_data.txt" , package="datasets")
```

```
> d<-read.table(filepath)
> x<-matrix(nrow=81,ncol=2)
> x[,1]<-d$V1
> x[,2]<-d$V2
> knn<-knearneigh(x, k=4, longlat = TRUE)
> data.nb<-knn2nb(knn, row.names = NULL, sym = FALSE)
> w<-nb2mat(data.nb)
> I<-diag(81)
> y=solve(I-rho*w)%*%data1990%*%b
> rmse=sqrt(sum((d$V3-y)^2)/81)
> cor(d$V3, y)
>
> #gwr validation
> filepath<-system.file("data", "gwrcoeff.txt" , package="datasets")
> c<-read.table(filepath)
> b<-matrix(nrow=81,ncol=10)
> b[,1]<-c$V1
> b[,2]<-c$V2
> b[,3]<-c$V3
> b[,4]<-c$V4
> b[,5]<-c$V5
> b[,6]<-c$V6
> b[,7]<-c$V7
> b[,8]<-c$V8
> b[,9]<-c$V9
> b[,10]<-c$V10
> y<-rowSums(data1990*b)
> rmse=sqrt(sum((d$V3-y)^2)/81)
> cor(d$V3, y)
```

APPENDIX C

FIGURES

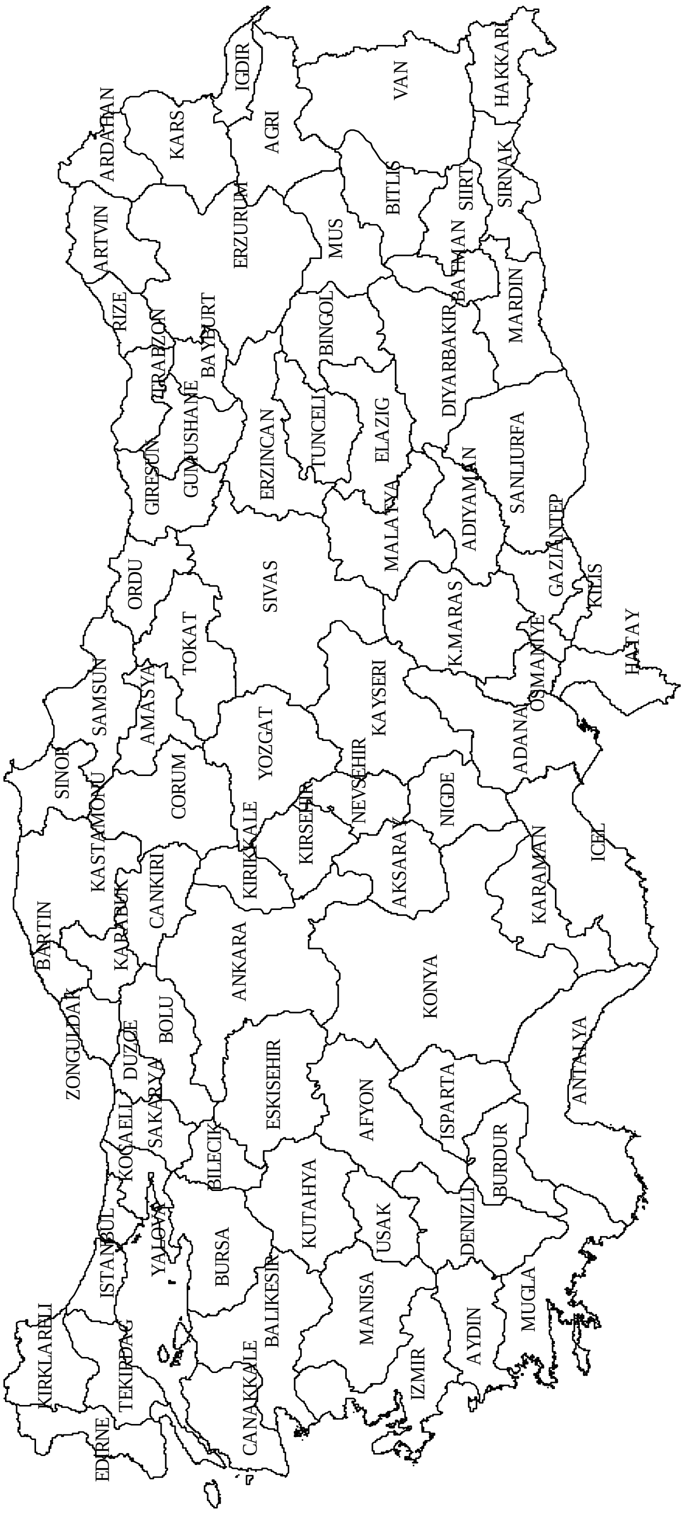


Figure C.1: Provinces in Turkey

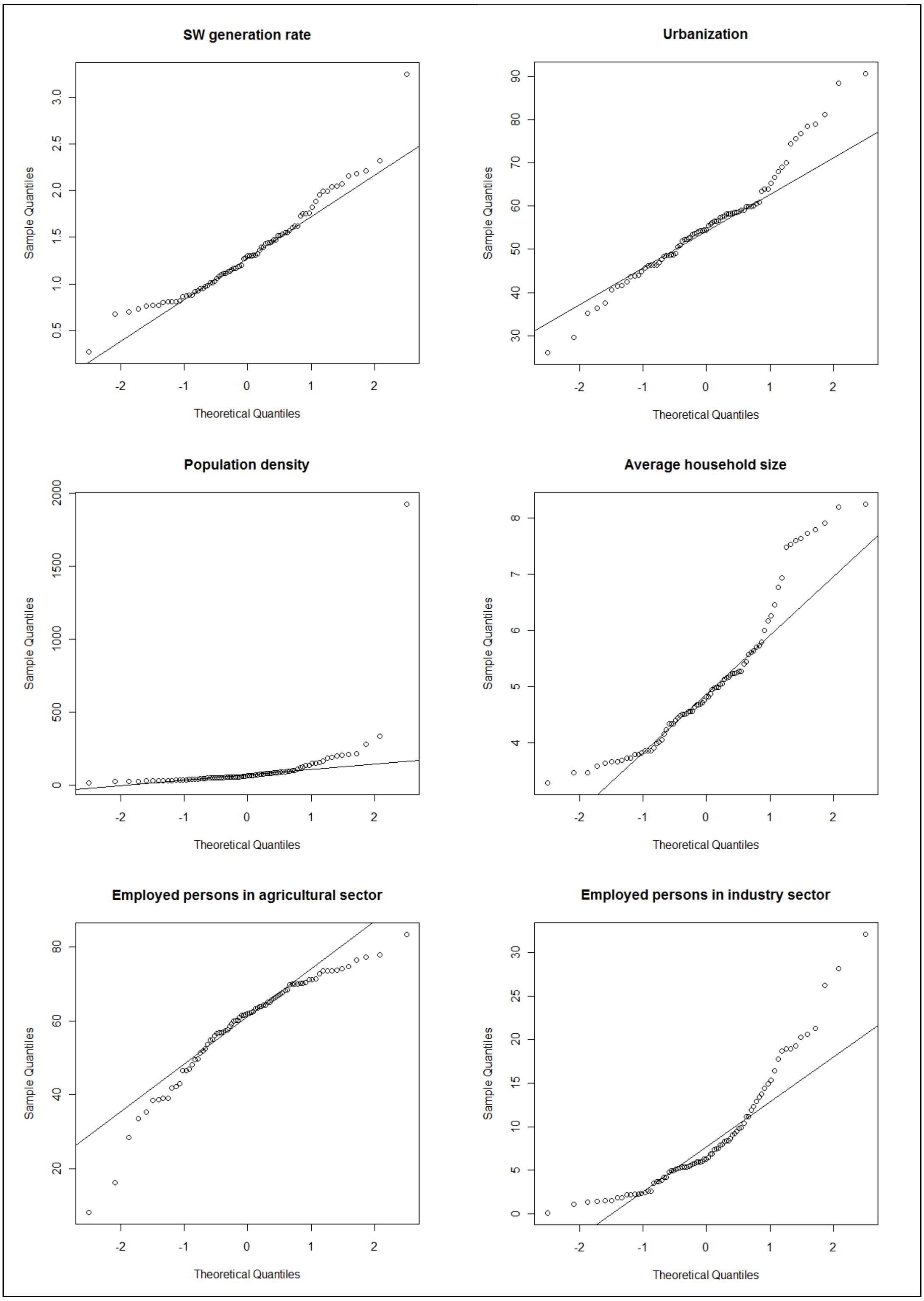


Figure C.2: Q-Q plots of the variables

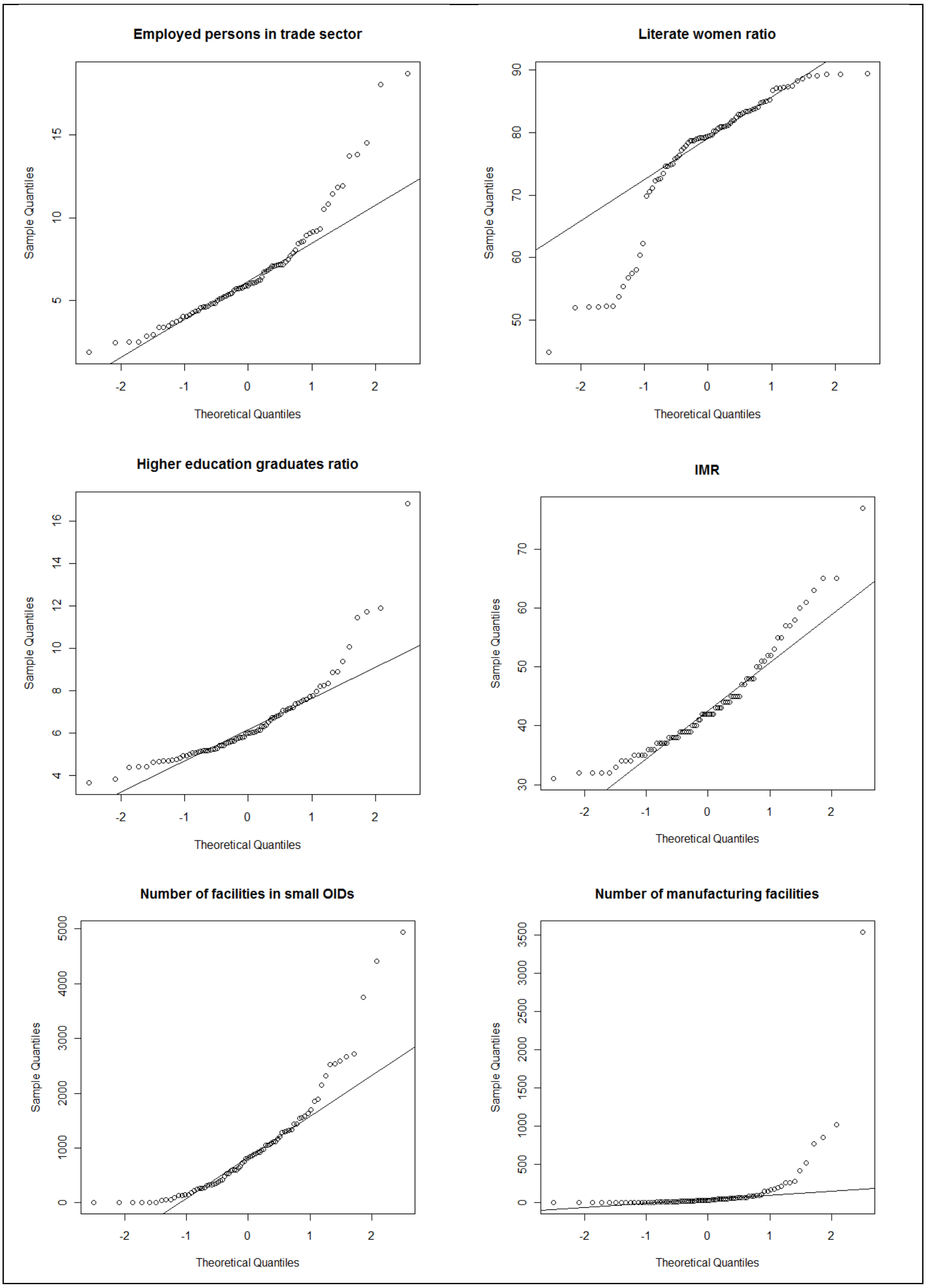


Figure C.2 (cont'd)

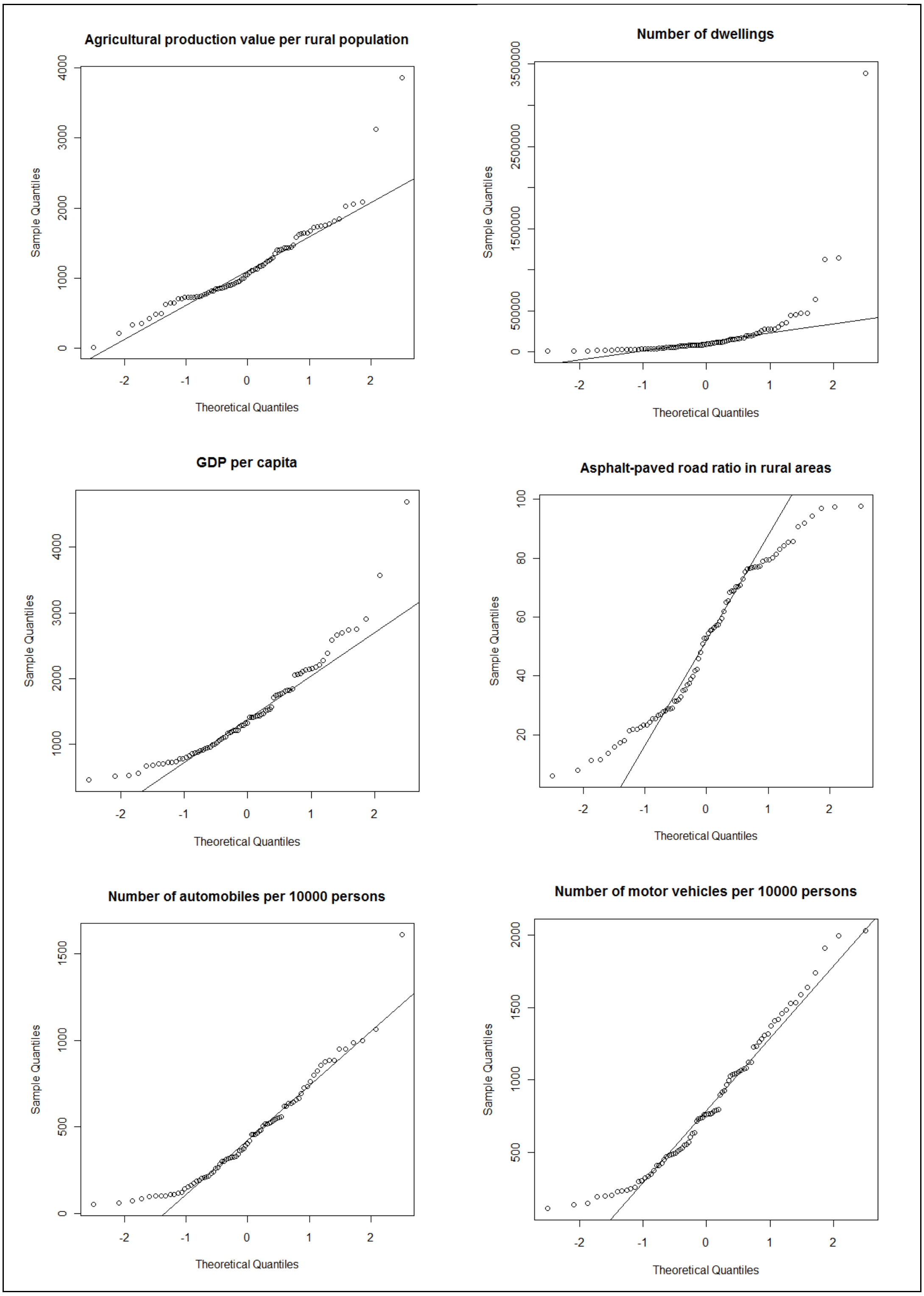


Figure C.2 (cont'd)

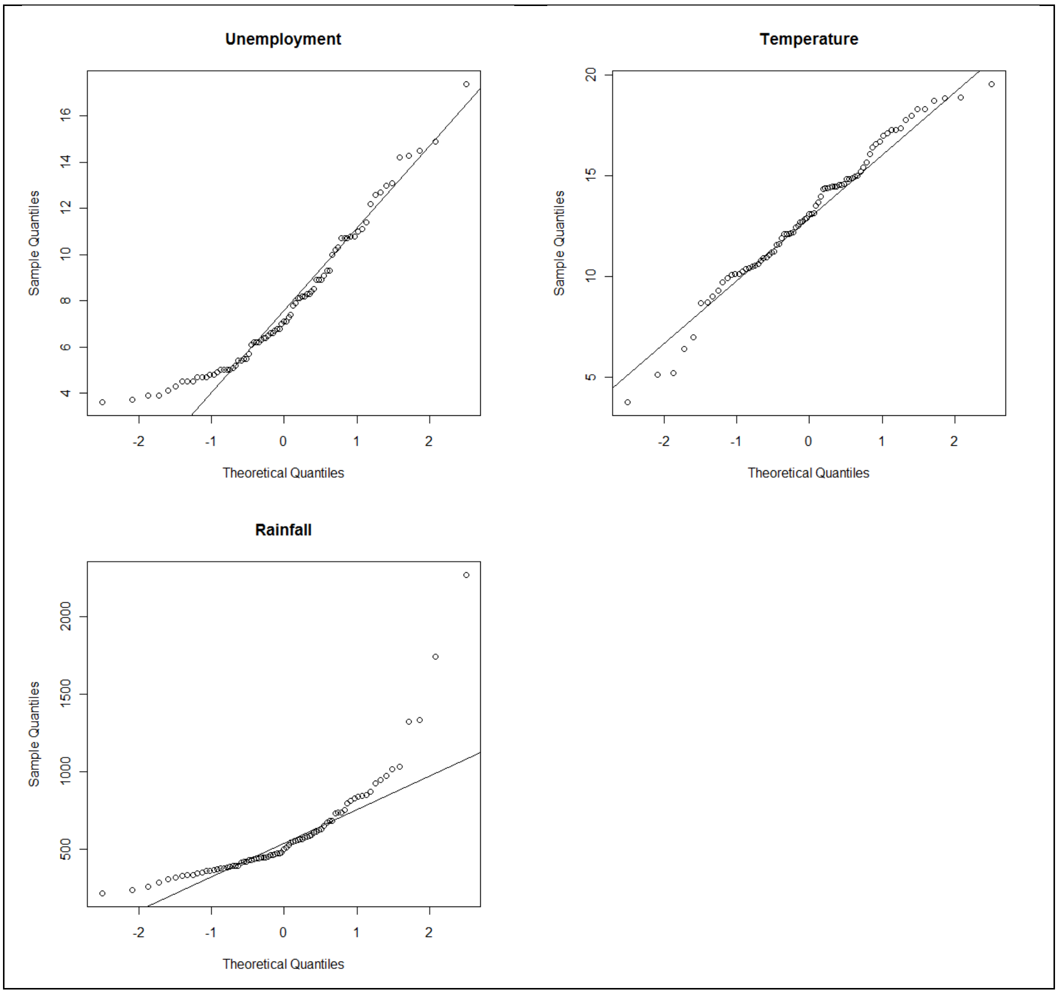


Figure C.2 (cont'd)

APPENDIX D

TABLES

Table D.1: Observed MSW generation rates for 1990 and 2000

Province	1990	2000	Province	1990	2000	Province	1990	2000
Adana	0.95	1.18	Edirne	1.02	1.39	Kütahya	2.40	1.99
Adıyaman	0.87	1.13	Elazığ	0.49	1.39	Malatya	0.91	1.17
Afyon	1.45	1.55	Erzincan	0.82	1.08	Manisa	1.18	1.30
Ağrı	0.34	0.73	Erzurum	0.72	1.01	Mardin	0.25	0.86
Aksaray	1.47	1.36	Eskişehir	0.77	1.01	Muğla	3.39	3.25
Amasya	2.28	2.32	Gaziantep	0.64	0.81	Muş	0.29	0.68
Ankara	1.46	1.60	Giresun	0.63	0.76	Nevşehir	1.12	1.57
Antalya	1.47	1.47	Gümüşhane	0.51	0.80	Niğde	1.57	1.73
Ardahan	1.74	1.17	Hakkari	0.07	0.27	Ordu	0.67	0.93
Artvin	0.46	0.88	Hatay	0.89	1.03	Osmaniye	1.29	1.14
Aydın	0.88	1.45	Iğdır	1.54	1.31	Rize	0.79	1.11
Balıkesir	2.49	1.99	Isparta	0.98	1.32	Sakarya	1.37	1.28
Bartın	1.24	1.55	İçel	0.64	1.06	Samsun	1.45	1.30
Batman	0.68	0.95	İstanbul	1.77	1.43	Siirt	0.53	0.77
Bayburt	1.05	1.75	İzmir	1.46	1.31	Sinop	1.91	1.75
Bilecik	1.10	1.44	K.Maraş	0.66	0.81	Sivas	2.63	1.95
Bingöl	0.96	1.19	Karabük	0.77	0.87	Şanlıurfa	0.41	0.77
Bitlis	0.61	0.92	Karaman	2.73	2.18	Şırnak	0.46	0.70
Bolu	1.25	1.76	Kars	1.02	1.30	Tekirdağ	2.48	2.21
Burdur	1.04	1.52	Kastamonu	1.52	1.88	Tokat	1.01	1.20
Bursa	0.85	0.97	Kayseri	1.62	1.62	Trabzon	0.81	0.82
Çanakkale	1.78	1.82	Kırıkkale	2.44	2.05	Tunceli	0.87	1.27
Çankırı	0.29	0.98	Kırklareli	1.68	1.53	Uşak	1.38	1.44
Çorum	1.54	1.62	Kırşehir	1.71	2.07	Van	0.56	0.88
Denizli	1.69	1.47	Kilis	0.20	0.81	Yalova	1.23	1.52
Diyarbakır	0.84	0.95	Kocaeli	1.41	1.11	Yozgat	0.72	1.15
Düzce	2.62	2.16	Konya	1.00	1.10	Zonguldak	2.08	2.04

Table D.2: Summary statistics of explanatory variables

Variable name	Minimum	Mean	Maximum
Urbanization	26	55	91
Population density	13	105	1928
Average household size	3.3	5.1	8.3
Employed people in agricultural sector	8	59	83
Employed people in industry sector	0.1	8.6	32.2
Employed people in trade sector	1.9	6.7	18.7
Literate women ratio	45	77	89
Higher education graduates ratio	3.7	6.5	16.9
Infant mortality rate	31	44	77
Number of facilities in small OIDs	0	1004	4947
Number of manufacturing facilities	0	137	3543
Agricultural production value	12	1161	3865
Number of dwellings	8079	200492	3393077
GDP per capita	453	1486	4696
Asphalt-paved road ratio in rural areas	6	51	98
Number of automobiles per 10000 people	51	453	1614
Number of motor vehicles per 10000 people	113	824	2033
Unemployment	3.6	7.9	17.4
Temperature	3.7	13.1	19.6
Rainfall	211	593	2269

Table D.3: p-values of *r* values

ID	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]
[2]	0.017																		
[3]	>0.999	>0.999																	
[4]	<0.001	<0.001	0.002																
[5]	<0.001	<0.001	<0.001	<0.001															
[6]	<0.001	<0.001	<0.001	<0.001	<0.001														
[7]	>0.999	>0.999	<0.001	0.017	<0.001	<0.001													
[8]	<0.001	0.045	<0.001	<0.001	0.003	<0.001	<0.001												
[9]	>0.999	>0.999	<0.001	0.486	0.18	0.025	<0.001	0.017											
[10]	0.003	0.482	0.256	0.053	0.052	<0.001	0.142	0.011	>0.999										
[11]	<0.001	<0.001	>0.999	<0.001	<0.001	<0.001	>0.999	<0.001	>0.999	0.003									
[12]	>0.999	>0.999	0.004	>0.999	>0.999	>0.999	0.04	>0.999	>0.999	>0.999	>0.999								
[13]	<0.001	<0.001	>0.999	<0.001	<0.001	<0.001	0.713	<0.001	>0.999	<0.001	<0.001	>0.999							
[14]	0.12	0.328	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.012	0.521	0.046	0.037	0.057						
[15]	0.365	0.917	<0.001	0.001	<0.001	<0.001	<0.001	0.113	0.256	0.022	0.17	0.051	0.222	<0.001					
[16]	0.075	0.713	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	<0.001	0.005	0.004	<0.001	<0.001	<0.001				
[17]	>0.999	>0.999	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.001	0.538	<0.001	0.087	<0.001	<0.001	<0.001			
[18]	<0.001	>0.999	<0.001	>0.999	>0.999	>0.999	0.001	>0.999	>0.999	>0.999	>0.999	0.621	>0.999	>0.999	>0.999	>0.999	0.383		
[19]	0.244	>0.999	>0.999	0.278	0.745	0.003	>0.999	>0.999	0.014	>0.999	>0.999	>0.999	>0.999	0.765	>0.999	>0.999	0.622	0.255	
[20]	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999