REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

THE EFFECTS OF STATIC AND DYNAMIC WEIGHTING ON SPATIAL AUTOCORRELATION

Ahmet Furkan EMREHAN

DOCTOR OF PHILOSOPHY THESIS
Department of Statistics
Program of Statistics

Supervisor
Assist. Prof. Dr. Doğan YILDIZ

June, 2022

REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

THE EFFECTS OF STATIC AND DYNAMIC WEIGHTING ON SPATIAL AUTOCORRELATION

A thesis submitted by Ahmet Furkan EMREHAN in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** is approved by the committee on 30.06.2022 in Department of Statistics, Program of Statistics.

Assist. Prof. Dr. Doğan YILDIZ
Yildiz Technical University
Supervisor

Approved By the Examining Committee

Assist. Prof. Dr. Doğan YILDIZ, Supervisor
Yildiz Technical University

Prof. Dr. Müjgan TEZ, Member
Marmara University

Prof. Dr. Ali Hakan BÜYÜKLÜ, Member
Yildiz Technical University

Assoc. Prof. Dr. Atıf Ahmet EVREN, Member
Yildiz Technical University

Prof. Dr. Ali Karun NEMLİOĞLU, Member
İstanbul University

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I haven't falsified and/or fabricated research data and results of the study and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of The Effects of Static And Dynamic Weighting on Spatial Autocorrelation supervised by my supervisor, Assist. Prof. Dr. Doğan YILDIZ. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Ahmet Furkan EMREHAN

Signature

Dedicated to to mothers of children with special needs,
parents who do not discourage their children's curiosity,
and my late, hilarious uncle "Pasha"

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Doğan YILDIZ and my monitoring committee members Prof. Dr. Müjgan TEZ and Prof. Dr. Ali Hakan BÜYÜKLÜ for all their valuable help and advice with this PhD. I would like to thank Assoc. Prof. Dr. İbrahim DEMİR for his substantial informing me for registration "special student program" of Statistics Department to be first stage of my PhD Journey. I would like to thank Prof. Dr. Ulviye BAŞER for directing and encouraging me to Statistics. And I would like to thank my study-mate Mehmet Şamil GÜNEŞ for his precious support and friendship. I would like to thank my workmate Fatih Mustafa KORKMAN for his contribution to forming Turkey's Provincial Contiguity Matrix and his genial and always-positive guidance for coding process.

The last but not the least I feel a deep sense of gratitude to my wife, struggle-mate and best friend Derin EMREHAN for her invaluable support and contributions to my life and vision. I would like to express my sincere thanks to my parents Ali Osman and Nermin EMREHAN, for the fact that they always encourage me to keep going after my curiosities.

Ahmet Furkan EMREHAN

TABLE OF CONTENTS

LIST OF SYMBOLS

$B_X(\theta)$	Bounded Continuous Index
$b_X(\theta)$	Bounded Continuous Strict Index
$b^l_X(\theta)$	Piecewise Linear Bounded Continuous Strict Index
I	Global Moran
I^θ	Conditional Global Moran
I_i	Local Moran
I_i^θ	Conditional Local Moran
I_i^0	Minimum Local Moran
$I_i^{0.5}$,I_i^M	Median Local Moran
I_i^1	Maximum Local Moran
L_i^θ	Spatial Theta Lag
W	Contiguity Matrix
W^θ	Conditional Contiguity Matrix

LIST OF ABBREVIATIONS

BCI	Bounded Continuous Index
BCSI	Bounded Continuous Strict Index
CAM	Conditional Autoregressive Model
GIS	Geographic Information System
GWR	Geographically Weighted Regression
MAM	Moving Average Model
OLS	Ordinary Least Squares
SAC	Spatial Autocorrelation
SEM	Spatial Error Model
SLM	Spatial Lag Model
SLX	Spatial Lagged X Model

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The Effects of Static And Dynamic Weighting on Spatial Autocorrelation

Ahmet Furkan EMREHAN

Department of Statistics
Doctor of Philosophy Thesis

Supervisor: Assist. Prof. Dr. Doğan YILDIZ

Spatial Autocorrelation, whose tools detect whether data have a spatial attribution or not, is one of the central issues of Spatial Analysis. Data to be examined in Spatial Autocorrelation should be contiguity or distance based. The tools are developed to measure spatial Autocorrelation in the course of time: Moran's I, Geary's c, Getis-Ord G, and Joint-Count Statistics. Moran's I is the most popular among these measures because it is easily interpretable and it is applicable to both contiguity and distance based data. To detect spatial autocorrelation Moran's I is frequently used on contiguity-based data in the literature.

Spatial units are linked to each other by means of contiguity or inverse distance matrix. The matrix determines the neighbourhood set. Moran's I is based on the correlation between the observation of the spatial unit and its spatial Lag. The spatial Lag means the spatial effect, or more clearly, the effect of the neighbourhood set in this context. Spatial Lag is considered as the sum or weighted sum of the observations in the neighbourhood set. The mean of the neighbourhood set is widely used in the literature as a weighted sum thanks to the row-standardization of the contiguity matrix.

The purpose of the dissertation is to use minimum, median and maximum of the observations of the neighbourhood set as a certain spatial lag at the detection of spatial effect. In doing so, the concepts of minimum, median and maximum are extended to quantiles of the neighbourhood sets. And the quantiles are modelled in a conceptual consistency as particular spatial lags by the help of a continuous function on [0,1] interval. Minimum, median and maximum are exceptional cases of quantiles. Thereby regardless of their sizes, the distributions of all neighbourhood sets are examined by means of quantiles and their particular spatial lags. Spatial theta lag denotes the spatial Lag based on quantiles. This approach creates a continuous index on [0,1] The conditional Contiguity Matrix is defined as the matrix to create Spatial Theta Lag.

The effect of Spatial Theta Lag over Spatial Autocorrelation is examined employing Global ve Local Moran's I as measure and Pseudo p, FDR, and Bonferroni Bound as statistical criteria, by dint of striking examples; Unhappiness Rate of People Aged 25-34 in 2013 and Voting Turnout in 2015 ,. Statistically significant clusters appeared in the mean based Spatial Autocorrelation are investigated in the light of minimum and maximum based spatial autocorrelation. The change of statistical significance levels of Spatial Theta Lag on [0,1] interval are shown. Finally, the suggestions are presented.

Keywords: Spatial Autocorrelation, Moran's I, Quantile, Continuous Index, Randomization, Pseudo-p, Contiguity Matrix

YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

ÖZET

Statik ve Dinamik Ağırlıklandırmanın Mekansal Otokorelasyona Etkisi

Ahmet Furkan EMREHAN

İstatistik Anabilim Dalı
Doktora Tezi

Danışman: Dr.Öğr.Üye. Doğan YILDIZ

Verinin mekânsal yönünü olup olmadığını tespit eden araçları açısından Mekansal Otokorelasyon, Mekansal Analizin merkezi konularından biridir. Mekansal Otokorelasyonu incelenecek veri ya komşuluk ya uzaklık bazlı olmalıdır. Zaman içinde Mekansal Otokorelasyonu ölçmek amaçlı bazı araçlar geliştirilmiştir: Moran's I, Geary's c, Getis-Ord G, Joint-Count İstastiği. Bu ölçümler arasından Moran's I kolay yorumlanabilmesi sebebiyle en popüler araçtır ve hem komşuluk bazlı hem de uzaklık bazlı verilere uygulanabilir. Literaturde Komşuluk bazlı verilerde mekânsal otokorelasyon tespiti için sıkça kullanılır.

Mekansal birimler birbirine komşuluk yada ters uzaklık matrisi aracılığı ile birbirine bağlanır. Komşuluk kümesi bu matris tarafından belirlenir. Moran's I, mekânsal verinin gözlemi ile onun mekansal gecikmesi arasındaki korelasyonu esas alır. Bu bağlamda mekansal gecikme, mekansal etkiyi veya daha somut olarak komşuluk kümesinin etkisini kast etmektedir. Mekansal gecikme, komşu birimlerdeki gözlemlerin ya düz toplamları ya da ağırlıklı toplamı olarak kabul edilir. Komşuluk kümesinin ortalaması, komşuluk matrisinin satır standartlaştırılmasıyla, ağırlıklı toplam olarak ifade edilebilir. Bu method literaturde geniş bir biçimde kullanılır.

Bu tezin önerisi mekansal etkinin tespitinde komşuluk kümesindeki gözlemlerin minimum, medyen ve maksimumun da özel bir mekansal gecikme olarak kullanılmasıdır. Bunu yaparken minimum, medyen ve maksimum kavramlarını komşuluk kümesinin kantillerine genişletilir. Ve kantilleri [0,1] aralığında sürekli bir fonksiyon yardımıyla, özel bir mekansal gecikme olarak modellenir. Bu modelde minimum, medyen ve maksimum, kantillerin özel halleridir. Böylelikle kantiller ve onların özel mekansal gecikmeleri aracılığı ile, bütün komşuluk kümelerinin dağılımları boyutlarına bakmaksızın kavramsal bir tutarlılık içinde mekansal bir modelde ele alınabilir. Mekansal theta gecikmesi, kantil bazlı mekansal gecikmeleri ifade eder. Bu yaklaşım aynı zamanda sürekli bir index yaratır. Koşullu komşuluk matrisi ise mekansal theta gecikmesini yaratacak olan kantile bağlı bir komşuluk matrisi olarak tanımlanır.

Mekansal theta gecikmesinin, mekansal otokorelasyon üzerindeki etkisi ölçüm olarak Global ve Local Moran's I ve istatiksel kriter olarak Pseudo p, FDR ve Bonferroni Sınırı kullanılarak araştırılmıştır. Tezdeki yaklaşımın sonuçlarını daha iyi gösterebilmek için, bu yaklaşım TUIK'in il bazlı 2013 yılı 25-34 Yaşındakilerin Mutsuzluk ve 2015 yılı Seçime Katılım Oranlarına uygulanmıştır. Ortalama bazlı mekansal otokorelasyonda oluşan istatiksel olarak anlamlı kümelenmeler, minimum, medyen ve maximum bazlı mekansal otokorelasyon ışığında incelenmiştir. Mekansal theta gecikmesinin [0,1] aralığı üzerinde istatiksel önem düzeylerinin değişimi gösterilmiştir. Son olarak önerilerde bulunulmuştur.

Anahtar Kelimeler: Mekansal Otokorelasyon, Moran's I, Kantil, Sürekli İndex, Randomizasyon, Pseudo-p, Komşuluk Matrisi

1
Introduction

1.1 Literature Review

Spatial Autocorrelation(SAC) is described as a degree of autocorrelation of spatial units with their surrounding spatial units in an area to be examined. The concept of the surrounding can be extended to all spatial units in the area like in distance-based studies; however, the degree of autocorrelation degree decreases as the distance increases. These spatial units can be a point, like observation points, or polygons, like provinces. If the spatial unit is a point, then the distance-based study in question. If the spatial unit is a polygon, then the contiguity-based study in question. Conversion from points to polygons, and vice versa, are available through certain transformations like Voronoi Diagrams like Centroid Approach. One should keep in mind that used terms and concepts in SAC literature can vary with the researcher's choice or scientific parlance.

With regard to the early period of literature on SAC, Geo-economics and Geo-sociology research enhanced the studies and dominated the literature at the beginning. For example, diplomat and economist Gardner Ackley investigates how the sales scatter over the market region [1]. Walter Isard, the founder of Peace Science-Economics and Regional Science, discusses and asserts the importance of the spatial economic attribution together with the temporal effect [2]. Furthermore, Isard pens down two articles in which spatial distribution analysis with mathematical structures is used in the context of spatial analysis [3][4]. Iyer makes a crucial theoretical contribution to the development of the SAC concept by studying the distribution of points on a lattice structure [5].

Moran's I, the most used measure for SAC, is proposed by Patrick Alfred Pierce Moran, an Australian statistician. But Moran's approach is not so extensive as the usage of Moran's I at his time. He concentrates on only structures including contiguity property among its components [6][7]. Robert Charles Geary, an Irish statistician, develops Geary's c, another SAC measure, as a projection of the Wilcoxon signed-rank test approach by dealing with the contiguity problem. [8]. Thanks to Andrew Cliff and Keith Ord's studies, usages of Moran's I and Geary's c gained great recognition by other branches of sciences. Cliff and Ord use Moran's studies to deal with measuring SAC like residual analysis [9][10]. Moreover, Arthur Getis and Ord introduce a new SAC measure for distance base [11] [12]. Getis and Ord's article leads up to considering SAC on a local scale, in other words, in the aspect of the unit. Getis also represents SAC and interaction by using Cross-Product Approach [13]. Furthermore, Getis and Ord also propose a Global SAC measure [14]. Geary's study on contiguity ratio is used in SAC analysis by Atsuyuki Okabe under Regional Science [15]. By force of geographical boundary problem, extreme values in SAC are examined [16]. SAC books written by John Odland and Micheal F. Goodchild are reference guides in terms of an informative and comprehensive introduction to Global SAC and its mathematical tools [17][18]. They can be said to represent perspicuous concepts for SAC analysis extracting relatively abstruse studies by that time.

Spatial Econometrist Luc Anselin extends SAC measures, called global SAC measure, to a level of a particular unit, the local SAC measure. That measure is coined as LISA, Local Indicators of Spatial Association, by Anselin. Unlike Moran and Geary, local SAC measures gain a great recognition among researchers relevant to SAC with Anselin's conspicuous expressions [19]. Moreover, the pros and cons of SAC are discussed by ecologist Pierre Legendre [20]. Similar to the previously mentioned study [16], Getis and Ord investigate extreme values of G statistics by using data of AIDS cases in San Francisco [12] and Boots and Dufournaud develop an algorithm for finding extreme values deals for SAC [21]. Unwin studies the potential of Geary's c statistics for SAC. He obtains positive outcomes [22].

In literature, spatial weights having central importance for SAC are examined. Equal weight effects are examineda as a particular case [23]. Unlike simple methods generating the map, algorithmic methods are used for the determination of more realistic spatial weights, especially superficial data [24][25]. In order to eliminate model misspecification, latent negative SAC is examined by using Thiessen Polygons with centroid Approach [26]. Methods based eigenvalues are used for determining the degree of SAC [27]. By amalgamating advanced linear algebraic methods, spatial filtering, as a concept, is developed under SAC literature with the help of properties of eigendecomposition of weight matrix [28]. Furthermore, eigendecomposition based methods are developed for SAC [29][30]. As an advance of spectral theory, related field to eigendecomposition eigenvector maps is used in studies for SAC Analysis[31]. By matrix decomposition based on eigenvalues, information scientist Yuzo Maruyama proposes an alternative method to Moran's I [32].

Strikingly theoretical study for Spatial Analysis can be an application for another branch of study. In this context, it is an excellent example that computer scientist Fernando Diaz shows the existence of an upper boundary for Moran's I by using Cauchy-Schwartz inequality [33]. The contribution of SAC analysis to Regression analysis is a significant level by the medium of the spatial lag and weight. Spatial lag is used for evaluating regression models. But the study deal with spatial lag as a weighted average [34]. Some transformation is suggested for SAC to maintain a realistic model [35]. Spatial weight functions are examined [36] . SAC is employed as a tool for the solution of problems generated by linear regression analysis [37]. Spatial lag may not be used for only autoregressive models. An iterative method including a fractal approach is proposed for spatial cross-correlation, in other words, bivariate correlation, by environmental scientist Yanguang Chen [38]. Bayesian hierarchical modeling is examined in case SAC exists over data relevant to the religious segregation in Northern Ireland. Unsurprisingly Autoregressive model is used in this study [39]. Moran's I under model assumptions are represented by evolutionary scientist Emmanuel Paradis [40]. Fractal methods also make a contribution to SAC for improving over continuous models [36]. Through Moran's I and Geary's c, regression models with SAC are tested with the Bootstrap method [41]. SAC methods also are available in computer programming languages relevant to Statistics. For example, Spatialpack package involving SAC is developed for R for SAC analysis [42]. In consideration of present computational advances, SAC problem [43] represented by Cliff and Ord is scrutinized again [44]. In the presence Heavy tailed SAC, Kreuzer proposes a tSAR model to determine fitted residuals for SAR. [45]. Furthermore, visualization of SAC is another issue in the literature [46].

In addition to these studies, historical development of SAC [47][48], by mentioning quantitative revolution [49], SAC in introductionary general concept [50][51] and Local predictions of SAC in introductionary general concept [52] are papers theoretically contributing SAC methods in general context.

Economics and SAC Analysis make a great contribution to each other. One of the reasons for this is the fact that Anselin, a pioneer of SAC, is also econometrics and writer of reference guides in that field [53]. Bodson and Peeters employ the SAC measures to explain the Belgian Labor-Demand function [54]. Fischer and Griffith study on Patent Citation Data of Europe by using SAC to understand the regional distribution of research-development activities over EU [55]. Economic activities in rural areas of Ethiopia, an east African country, and Nigeria, a West African country [56] and the development of 98 countries between 1965 and 1995 [57] are compared in terms of SAC. SAC is also used for measuring and evaluating regional growth in the EU by employing panel data methods [58][59]. In order to shed light on the variability of unemployment in the central European region, Moran's I, as a SAC measure, is employed [60]. SAC analysis help to understand foreign aid effects in conflict areas, in the example of Darfur, South Sudan, and central Ethiopia by employing Global and Local (LISA) Moran's I [61]. Furthermore, under the sponsorship of the European Central Bank, zonal perimeters of economic performance are examined [62]. Kantar and Aktaş examine the dispersion of the Provincial Unemployment Rate of Turkey, and local SAC measures are used [63]. Taking the level of urbanization and per capita GRP (2000-2013) into consideration, the economic levels of 29 regions in China are investigated by using SAC methods with the Durbin-Watson test as a residual analysis [64]. Local SAC method is used for agricultural operations in the USA with clustering approach and determining hot spots [65]. Haung and Wei examine the effects of investments in China by employing general Getis-Ord statistics [66][67][68]. In order to develop tourism plans, tourist arrivals to Serbia are scrutinized by using Global and Local SAC methods [69]. Uslu represents a comprehensive study on Spatio-temporal analysis of provincial income distribution of Turkey at period 1992-2013 [70]. Distinctively Real Estate Appraisal is a main field in which SAC is often used [71]. As examples of Real-Estate studies, Basu and Thibodeau apply the SAC with other spatial methods to the real estate market in Dallas, TX [72], and Hsieh also investigates Spatial Dependence among housing prices in Taiwan [73]. There are also functions, like Moran's and Geary's c, developed for spatial econometrics in R, programming language[74]. Getis emphasizes on contributions of SAC into spatial econometrics [75].

Medicine is another field in which SAC methods are used. Wartenberg uses SAC in combination with Principal Component Analysis for the spatial distribution of blood groups [76]. Cook adopts a primitive form of SAC to analyze cardiovascular mortality data in British towns [77]. In the example of the AIDS epidemic in San Francisco, Getis and Ord provide an insight into the extreme values of SAC [12]. To unveil spatial dynamics of the spread of disease, SAC is discussed [78]. Breast cancer cases in Kentucky are examined with SAC as residual analysis of Loglinear regression model [79]. Distance-based SAC model is used for gender discrimination of cancer incidences in Saudi Arabia [80]. Spatial attributes of health events in general concept are analyzed using SAC and forming clusters associated with SAC [81]. Viral foodborne diseases in Toronto, Canada, are examined by dint of general Getis-Ord statistics [82]. Moreover, Richardson and Joyaux mention SAC methods in epidemiology [83].

Biology and its subbranches are among the most influenced sciences by Spatial Analysis and SAC. Jumar's study of the diversity of marine benthos population opens up the horizon for studies of biology, and in particular oceanography, in recognition of SAC and Analysis [84]. Furthermore, the significant contribution of Roger R. Sokal to Biology and related subbranches by SAC and Spatial Analysis can never be ignored when looking at the literature. Sokal introduces SAC into Biology and related science in both methodological [85] in the example of blood group and tree distribution[86] and Local SAC measures into Biology and related sciences by applying lilies and gopher distributions in Colorado USA [87]. Sokal also studies over selection and migration of genes [88] and expounds on local SAC methods for biological models by using permutational methods to overcome random distribution problems for hypothesis tests [89]. In conclusion, Sokal is said to make researchers interested in biological sciences acquainted with global and local SAC methods through his exhaustive studies. Smouse proposes the distance method to be used in SAC with the intent to develop studies about genetics, including multivariate data [90]. Dispersal of an enzyme is examined by using dispersal of a plant, Centaurea jacea, as marker [91]. Mutational transformation of genotypes based on microsatellite data [92]. Genetic diversity is one of the most-SAC-used fields. Local species differentiation of Kangaroo Rats is examined with respect to genetics and dispersal of population [93]. Diversification of gene [94] and spatial constraints of gene pools [95] are scrutinized with SAC. Genetic differentiation of Red Snapper (Lutjanus campechanus), a fish living in Northern Gulf of Mexico [96], Seal counts in Lazarev Sea Antartica (with ANOVA) [97], spatial heterogeneity of amphibia population in Southern Brazil [98]and of fungicide resistance of Botrytis cinerea in vineyard and Botrytis squamosa in onion field [99], dispersal of foot-and-mouth disease in animals in China 2010-2016 [100] are examined by using SAC.

There is no doubt that Ecology is one of the most using SAC and other spatial methods. Canadian biologist Pierre Legendre has a great contribution to that fact and can be considered a pioneer in spatial ecology. This example is similar to Anselin's studies in spatial econometrics. There exist extensive and introductory, ecological studies in literature [101][102][103]. Along with those, there exist incentive papers that represent practical advantages of SAC for ecological studies [104][105]. Neotropical migrant sonbirds in the Southern Appalachian Mountains [106], the ecological determinants of Medfly populations on orchards in Greece [107], dynamics of several species of ladybugs population in Cheaskespeake and Delaware Canal [108], watershed ecology in the Coastal Plain in Maryland [109], fire ecology as a theoretical study [110], estimating forest bird population with emphasising spatial resolution and autologistics modelling [111], birds occurrence and vegetation interactions in the Central Great Basin [112], species distributions for theoretical contributions to ecological modelling [113][114][34], sampling methods of survey for vegetation behavior North Carolina [115], unfasting the balance and neutral and niche processes [116], plant species richness and environment on Germany [35], coral reef ecology on Central Bommie in the Great Barrier Reef [105], heterogeneity of Bird species flocks in Patuxent Maryland [117], the development of model for subdivided populations of migration schemes, by using Fourier transforms [118], simulated data from mule deer population [119], prediction of endangered swift parrot population in Tasmania [120], heterogeneity of grasshopper in Atelier Plaine and Val de Sevre [121], Marin ecology in example of benthic habitat mapping [122], fungal community in northern Mississippi uplands ecosystem [123], climatic niche and its range size [124] and data mining strategies for data including various characters, like phylogenetic, spatial, temporal and hierarchical [125] are examined by employing SAC methods.

There exists a considerable number of studies about forestry using SAC in literature. Natural pine stands in Georgia [126], local forest growth in several areas of U.S.A. [127], Spatio-temporal dynamics of tree transpiration in southern Wyoming [128], three forests in Ontario Canada, by using comparative regression models, [129]. Forest biomasses and their carbon stocks on a global scale [130] are scrutinized by using SAC.

Agriculture is one of the most-used scientific branches using spatial analysis. Inherently that fact is valid for SAC analysis. Modeling of sulfate concentration in Atlantic Highlands [131], soil properties in northwest Poland [132], examination of between crop yield and weather in 1964 counties of U.S.A. [133], the outcome of land consolidation in Lubelskie Voivodeship Poland, as an application to agricultural policy [134]. Soil properties, like clay, potassium, etc., in Keller peninsula Antarctica [135] are examples of SAC in agricultural studies.

SAC analysis gives a visible contribution to also Risk Analysis studies. Avalanche risk in Bridger Range in U.S.A [136], catastrophic landslide risk as a comparative study [137], fires in Serbia, with outlier analysis [138], debris slide risk in the Austrian Central Alps [139] and determination of risk and place of landslide in Myanmar [140] uses SAC methods.

Performance prediction of information retrieval in computer science [33]; spatial segmentation of black population in Newyork [141] and participation in 2007 French Presidential and 2010 Regional elections [142] as an introductionary paper for social sciences and crime analysis [143]; in order to maintain the efficiency of transportation in route distribution analysis in the urban area of Cosenza Italy [144], growth of the urban area in a river delta in China with panel data [145], crash frequencies in Nevada freeways [146] in transportation and accident Analyzes; daily precipitation in Peribonca River in Quebec [147] and Spatio-temporal thermal pattern Analysis in Iran [148] in Meteorology and Climatology; Chemistry of stream water in Maryland

U.S.A. [149] and heavy metal and nitrogen deposition on extended Europe [150] in environmental sciences; analysis of regional linguistic differentiation [151] and avoiding the autocorrelated residual problem for modeling language richness with spatial attributes [152] in linguistics ; permeability prediction for porous rocks, with methods based on image processing [153], surface waves over geophysical systems, like hydrothermal system [154][155][156] in Geoscience; testing SAC in the panel data model [157] in statistics employ SAC methods. Andy Mitchel presents computational methods for SAC by introducing GIS software, ArcGIS [158] . Anselin and his team developed a unique software tool for spatial autocorrelation [159].

1.2 Objective of the Thesis

Spatial Lag holds the principal role in spatial autocorrelation and regression studies. In contiguity based studies Spatial Lag is extensively employed as the mean of the neighbourhood set. Furthermore there is a strong tendency about the mean of the neighbourhood set as an indicator of the spatial effect. But empirically, the most of size of the neighbourhood set is small in the contiguity based studies. This fact gives rise to occur a representation problem for the neighbourhood set, particularly in the cases where the neighbourhood set includes extreme observations.

The classical approach to spatial autocorrelation assumes that extreme values in the set create an extreme mean, and then a statistically significant spatial unit manifests itself in the analysis. In such a case, the cause is extreme observation, and the effect is the mean of the observations in the neighbourhood set, which is extreme. The classical approach is based on the effect and embeds it into the spatial autocorrelation model. Moreover, the mean of a neighbourhood set, including extreme minimum and maximum observations, may not represent the reality of the neighbourhood.

The main objective of the dissertation is to use the minimum, median and maximum of the neighbourhood set as particular spatial lags to deliver an approach to be a solution for these problems mentioned above. Furthermore, the dissertation aims to create a quantile based structure over the neighbourhood set by a bounded continuous index. The minimum, median and maximum of the neighbourhood set are particular quantiles which are indexed in a bounded continuous set. Consequently, all neighbourhood sets are consistently and conceptually unified.

The basic approach of the dissertation will enhance the understanding of the spatial mechanism.

1.3 Hypothesis

The dissertation focuses on some specific observations in the neighbourhood in the alternative to the mean. We investigate whether the global spatial autocorrelation based on the neighbourhood set's minimum, median and maximum are statistically significant. In addition, we examined the local spatial autocorrelation for each spatial unit based on the neighbourhood set's minimum, median and maximum. Particularly the question of whether the local spatial autocorrelation based on the maximum and minimum of the neighbourhood set can cause the occurrence of the statistically significant spatial units is one of the focuses of the dissertation. Whether the spatial autocorrelation based on the quantiles of the neighbourhood set, which are constructed a bounded continuous index, is a question whose answer is in the scope of the dissertation. Due to violation of the assumption of Normality, Randomization and pseudo-p methods are used to evaluate results statistically with FDR (False Discovery Rate) and Bonferroni Bound conditions.

2
Spatial Analysis

2.1 Definitions of Spatial Analysis

To define a branch of science, it is essential to determine what the branch includes and excludes. If a field, like "Spatial Analysis", involving techniques galore, like statistical, mathematical or image-based, and being in the intersection of many fields, such as geography, statistics and mathematics, is in question, the definition of the field becomes more sophisticated attempt.

It is an obvious fact that Spatial Analysis emerges from Geographical Sciences. Therefore all definitions for "Spatial Analysis" include geographical references. In literature, one can find many definitions for Spatial Analysis :

"An approach to geography that emphasizes investigating the spatial distribution of phenomena and the factors influencing observed distribution patterns." [160]

"The application of quantitative methods in locational analysis within human geography and sometimes used a synonym for that portion of the discipline that concentrates on the geometry of the landscape"[161]

"A type of geographical analysis which seeks to explain patterns of human behavior and its spatial expression in terms of mathematics and geometry, that is, locational analysis." [162]

Even though these definitions cover the field being relevant to Spatial Analysis to a great extent, they do not involve the non-geographical fields in which the spatial analysis is used. There is a matter to be discussed; how Spatial Analysis should be defined. All definitions lay between two marginal approaches: one side is to define regardless of all application literature, and that means a method as a combination of geometry, statistics, and mathematical analysis. Another is to define by using possible all fields using the methods of Spatial Analysis. To comprehend what extent Spatial Analysis spreads, the latter is indubitably better.

In literature, as seen above, most definitions tend to mention just geographical sciences on account of historical background; however, one can find a relatively better definition, as seen below. "Analysis can be defined as spatial if the locations of the objects in the space matter by affecting the analysis results. "... To a geographer, spatial analysis almost always implies that the space of interest is geographic space-the surface and near surface of the earth- **but methods of spatial analysis might also, in principle, be applied to phenomena distributed within the space of the human brain, along the linear space of the human genome, or on the surface of the another planet**." Micheal Goodchild [163]

All studies under Spatial Analysis can be classified into three classes, prediction, correlation, and pattern analysis

2.2 Historical Background

In the historical course, it can be said that the studies,that would be evaluated in the scope of spatial Analysis, are emerged from the fact that the quantitative methods gain importance in geographical sciences. The oldest known study ascribed to Spatial Analysis is research overlapping the cholera cluster and water pump layers over an epidemic map by Dr. John Snow, English Physician in 1854 Broad Street Cholera Outbreak in London [164] [161]. Moreover, he is a pioneer of the germ theory, which would be recognized afterward by academia. Then Spatial Analysis gives a shred of evidence supporting the germ theory at the beginning. I would like to lay emphasis on the fact that the first studies of Spatial Analysis involve Cartographical Analyses in accordance with a qualitative or quantitative attribution in the method. Apart from this, the development of spatial sampling methods and the endeavor the description of a particular field by using spatial sampling are the fundamental issues of the first studies of spatial Analysis. Unsurprisingly Ronald Fisher is also another pioneer of Spatial Analysis. Particularly Fisher's studies on agriculture is one of the first studies in spatial sampling and autocorrelation [165]. The splendid boom in the use and availability of computers in the 1970s made multivariable statistical methods more visible and comprehensively used in Spatial Analysis than studies in the past. Still, this progress tends to initialize in late of the 1940s [166]. It goes without saying that Social-Geography has an outstanding contribution to Spatial Analysis as a sub-branch of Geography at the earlier stages. That fact paves the way for the usage of many multivariate statistical methods in Geography through sociology in pioneer studies of Spatial Analysis, in a modern concept [166] [167]. Classification of Socioeconomic differences in British Towns by using Factor Analysis [168], application of statistical methods to Urban System with geographical aspect [169] and grouping cities [170] are considered to be stimulating studies in developing era of Spatial Analysis. Spatial Analysis uses not only statistical but also applied mathematical methods. For the purpose of modeling spatial interaction, partial differential equations are used [171].

Of early application fields, Sociology is an essential place for Spatial Analysis. Lefever investigates the effects of location over social factors by using spot map [172]. One of the fields in which the Spatial Analysis era is applied in early is the Environment. Sutton studies a pattern of flying pollutants emitted from the factory's chimney by using archetypal methods from Spatial Analysis [173]. As one of the first studies in spatial economics, Samuelson combines linear programming and spatial prices [174]. For the purpose of determination of spatial characteristics of Housing Deterioration, Spatial pattern methods are used with Logit models [175] . Genetics is a field that firstly uses Spatial Analysis, too. Birdsell looks for the potential of Spatial Analysis to assist the research in Racial anthropology as an analytical approach [176]. That continues afterward. Heywood scrutinizes the genetic variation of plants by using spatial distribution [177]. It is essential to be said that Ecology is another trend in the application field of Spatial Analysis in the early phase. Clark and Evans, researchers in quantitative ecology, developed a spatial measure by using Nearest Neighbor Distance which would help the emergence of Geostatistical Methods and Concepts, like Variogram and Kriging [178]. Archaeological Studies also take advantage of tools of Spatial Analysis. Hodder and Orton spatially apply a combination of nearest neighbor analysis, regression analysis, and surface analysis to the distribution of archaeological findings in a study area [179]. Spatial Analysis is used for also Astronomy. One investigates magnetic fields of sunspot [180]. In the past, it is evident that the overall tendency in geographical sciences is to apply data on points on the relevant map. The spatial aspect of the data was ignored. Kriging is also used with SAC methods [181].

The fact that GIS Tools develop rapidly makes a great contribution to Spatial Analysis. Likewise, the exact opposite direction of the fact is also true. That is to say, methods of Spatial Analysis gives impetus to Geographical Information System in a geocomputational context. This is a reciprocal relation. On the contrary to the classical approach in geoeconomics, a potentially significant role of Spatial attribution over the phenomena lies behind Spatial Analysis as a latent assumption. That fact attracts criticism due to negligence of free-will of economic agent [162].

In a general sense, Spatial Analysis is used for the detection of patterns, clusters, correlations, associations, dependencies, interactions, and anomalies in spaces being hidden or partially distinguishable for the researcher. Methods of Spatial Analysis are applied depending upon the necessities and purposes of the study. In this respect, Spatial Analysis has multifaceted, flexible, and feracious characteristics. Moreover, Spatial Analysis provides the researcher an opportunity for broader interpretations of the matter that he or she scrutinizes. Spatial Analysis contributes to statistical methods, like regression analysis, regarding the explanation of the models. For example, the Spatial Lag concept increases the explanatory strength of the regression model.

Spatial Analysis is also used and considered in explanatory and confirmatory contexts as a subbranch of statistics. Spatial Analysis is able to discover patterns, anomalies, abstractions, and associations. Spatial Analysis can be used for testing a suggested theory relevant to geographical space as confirmatory statistics.

One should never forget that Cartography and Surveying constitute irreplaceable fundamentals of Spatial Analysis. Thereby each technique used in cartography study to be the basis of spatial analysis have a potential to change the direction of the research. On account of its nature having fractal geometry, measuring the coastline of Britain with a yardstick of various lengths can give as an example. This case is directly relevant to the resolution problem. Projection problems may have a great influence over Spatial Analysis when the study is on a large scale in the same way. All forms of these problems can be studied under model misspecification.

Line, area, volume, point, surface, and lattices are types of spatial characterization. They are also crucial with regard to passing to GIS. At this point, one should keep in mind that spatial characterizations can be transformed into each other with the allowance of spatial analysis in the study. For example, a spatial structure of seemingly areal data can be transformed into a lattice structure without losing any required information.

The meaning of "Space" may not always be used in a geographical context. That is to say; Space does not always mean "Geographical Space". That is why Spatial Analysis is too comprehensive to contract into Geographical Science. Spatial Analysis is also used in different fields, like Computer Science [33] and Cellular Biology [91].

In the historical process, the contribution of Regional science to Spatial Analysis can not be neglected. Regional science is described as a hybrid discipline merging neo-classical economics and quantitative methods in order to analyze human geography and to create more rational urban planning and policy [161]. The fact that Walter Isard, the father of regional science, focuses on mathematical models [162] puts spatial analysis into the center of regional science as a fundamental method all along. In this respect, it is remarkable that the copyright of the book "Spatial Autocorrelation", the first and reasonable comprehensive book in that field [17], belongs to Regional Research Institute in West Virginia University USA, as a Regional

Science Resource, with Sage Publications.

Spatial Analysis tends to inhold the most valuable and used method in Quantitative Geography. Furthermore, Spatial Analysis acts as a connection between Qualitative attributions and Quantitative Geography.

2.3 Branches of Spatial Analysis

All studies in Spatial Analysis can be classified into three branches with their combinations: Geostatistics, Lattice Models, and Pattern Analysis [182]. It is crucial to state that three branches can not be distinctively separated in studies involving Spatial Analysis. For example, Variogram in Geostatistics [129] [150][105] and Ripley-K function in Point Pattern Analysis [183] [184][185] is used for measuring and understanding SAC , one of the central themes in Lattice Models.

2.3.1 Geostatistics

The main focus of Geostatistics is to predict unmeasured spatial elements on the map. Data in Geostatistics is examined point-based. In other words, informative attribution belongs to a spatial point. Models in Geostatistics tend to be constructed under the assumption of the presence of isotropy and stationary. Furthermore, Variogram, semivariogram, and covariogram in Geostatistics, are based on Cartesian representation of distance of two points by distance the attributions of them. Kriging and its all versions, such as universal and block Kriging, are mathematical modeling of the variograms constructed using non-linear regression models for the purpose of more efficient, realistic, and data-driven interpolation.

In the historical process, Mining and geoeconomics attempts have a chief role in developing Geostatistical methods, like Variogram and Kriging. Daniel G. Krige makes an excellent contribution to using mathematical models for interpolation spatial data [186]. Whereas Kriging is invented as an idea by Daniel G. Krige, Georges Matheron represents Krige method in a mathematical form [187].

2.3.2 Point Pattern

On the other hand, unlike Geostatistics and Lattice Data analysis, Point Pattern analysis focuses on only where incidents occur or exist in the study area. Magnitude (interval and rate) and categorical (ordinary and non-ordinary) attribution of spatial unit (point) are out of study interest. Point Pattern analysis seeks an answer to the question of how incidents scatter over the study area: dispersed, randomly, or clustered. Furthermore, provided that the study area is divided into several subareas, like Quadrat Analysis, outcomes of Point Pattern analysis are considered to be a tool for measuring SAC [183][184][185].

Point Pattern Analysis involves many methods. Quadrat Analysis is based on Kolmogorov-Smirnov Hypothesis tests to detect whether two spatial scattering of incidents is statistically acceptable equal to each other or not. Apart from that, Quadrat Analysis is a tool for determining how incidences are scattered, dispersed, randomly, or clustered in keeping with point pattern analysis, its parent branch [188][189][190]. Nearest Neighbour Analysis is another sub-branch of point pattern analysis. Nearest Neighbour analysis tries to find the answer to the question of whether spatial scattering is random or not. It is invented by Philip J. Clark and Francis C. Evans [178]. Finally, Ripley K-function is the best known and most used method in Point Pattern analysis. Moreover, both nearest neighbor analysis and Ripley K-function are based on the distance among the points of the incidences by employing surrounding concentric circles for each spatial unit. Ripley's K-function is an extension of Nearest Neighbour analysis already. Ripley K-function is proposed by Brian David Ripley [191]. The most remarkable property of the Ripley K-function is to bring spatial point sub-patterns to the light by comparing them with the Poisson distribution. In doing so, iterative methods are used [192][193]. As an extension of point pattern analysis, the spatial pattern uses many mathematical methods, like Partial Differential Equations [171].

2.3.3 Lattice Models

First of all, lattice data can be considered as a projection of graph theory into Spatial Analysis. Unlike the other two branches of spatial data, Spatial Models on Lattices supposes that spatial units are of dynamical attributes. That is to say; spatial units can dynamically influence each other. Because of that fact, the Lattice Model is often used for temporal analysis. In the nature of data in lattice models, autocorrelation, the main theme of my dissertation on spatial data, and spatial dependence, interaction and etc., are examined under that branch of spatial analysis. In this respect, spatial models on lattices are a very fertile field. Lattice data can connect each other in two ways: distance (for point data) and contiguity (for lattice or areal data in polygon shape). Owing to that fact, Spatial sampling or Spatial resolution is very important for direction of pertinent studies [182]. Contiguity based studies can be based on county or provincial borders [194] [195] [196] [197] , or on quadrats [198] [199]. Tessellation, generalization of quadrat, is useful for Geographical Science, also SAC studies by employing Voronoi diagram, also called Thiessen Polygons [200][27][26][201].

Spatial analysis still preserves its righteous assertive position in geographical sciences as the main bridge between mathematical sciences and geography. Furthermore, the development of computation and data visualization provide an incredible speed of flourishing for spatial analysis, particularly in the last decades.

2.4 Spatial Data

First of all, it is important to emphasize the fact that although Statistical concepts are easily and perceptibly defined, they are depicted in many names in compliance with the field in which they are used, like geography, and econometrics. That case is also valid for spatial analysis. There is no convention over naming of particularly some spatial statistics concepts, like lattice data [182] - areal data [202]. Remarkably, the areal data concept is recently more popular than lattice data, particularly after advances in GIS. Spatial Data can be decomposed into two distinct parts, observation, and location of observation. Data, in a spatial analysis context, can be defined to be an observation at a spatial unit, whereas the location of data can be defined to be the address of a spatial unit. On that note, The address is a general concept to express where the point is. Not only is the address a standard Euclidean coordinate, but also it can be defined as a node like in graph theory, as a polygon or a point of which only distances with other spatial units in a defined non-regularly space are known like in the point pattern analysis. Data in spatial analysis is nothing but statistical data. Then it can be univariate or multivariate, categorical or continuous [17], real-value or nominal, finally nominal, ordinal, ratio, and interval like in the well-known classification of statistical data. Location of data is a prerequisite for spatial data. That is why data location is an input to pass the spatial attribution of the unit into the model. Data location is relevant to not only the coordinate of the spatial unit but also neighborhood, like in lattice data studies.

It is a trivial fact that spatial data is as old as data on the maps. One should not forget that all maps on which data is found can be interpreted as spatial data. However, all maps on which data is found can not be said to be used for scientific purposes. That fact should be kept in mind when the historical background of spatial data is examined. The findings of spatial data in modern meaning can be traced back to the 17th century. Edmond Halley, an English geophysicist and mathematician put on climatic data on the map for meteorological and commercial purposes [203]. Moreover some of first contributions into the literature belong to pioneers of statistics [204][205][206][207][208][209] . Choosing model and plot configuration are a essential matter and then not missed in burgeoning period of spatial data literature [210][211][212]. It should be uttered that Agriculture is the most motivator field to grow the field.

2.4.1 Mathematical Aspect of Spatial Data

Taking Spatial Analysis literature into consideration, it stands to reason that the mathematical aspect of spatial data is said to be not presented or founded as it should be. Because the overwhelming majority of studies and books in literature concentrate on empirical and practical goals rather than scientific concerns, nevertheless, it is possible to find some studies trying to present mathematical aspects of spatial data [213][182]. Fundamentally spatial data can be defined in the mathematical form below

Definition 2.1. Spatial Data

Let D be a set of the index indicating spatial property and X be a set of statistical data

$$\begin{aligned} Z : D &\rightarrow X \\ s &\rightarrow Z(s) \end{aligned}$$

where $Z(s)$ binding function between spatial property and statistical data or observation. Then spatial data is said to be $(s, Z(s)) \in D \times X$.

The essence of the matter in spatial data is that the characteristics of the domain, in other words, "index of data" determine the type of spatial data. In classical literature, by virtue of mathematical generalization concept, D-set of location information of data is considered a subset of R^n, Real numbers with n-dimension [213][182]. Spatial Data, D-set of location information of data, exists over a space homeomorphic to real numbers R and its Cartesian spaces with n-dimension, commonly R^2 and R^3. In other words, Most of the studies are confined to spaces with two or three dimensions, and it should be stated that models with three dimensions can be projected into models with two dimensions without crucial data loss in most cases. This fact makes mathematical operations over spatial data available. Then it is worthy of note to classify spatial data in a dimensional way on account of mathematical structure. However, the spatial modeling processes in the studies push the dimensional aspect of spatial data into the background. Related to X, Data can be any observation with spatial property.

Another point worth mentioning is the matter of whether a distinction among spatial data types exists or not. At this point, it is essential to determine the context of spatial data. Spatial Data can be classified in many ways. Dimension and usage are the classifications examined in this dissertation. Point Data can be considered a fundamental and essential concept for spatial data analysis because the length and surface data types are nothing but data configured, modeled, and generated by interpolation of point data. Furthermore, historically even the pioneer study of spatial analysis is based on point data [164]. In fact, point data is the founding concept of all branches of spatial analysis. However, point data is insufficient to explain all spatial relations. That is why mathematical and computational operations need value and location of the spatial unit in different concepts, like length and areal data.

Length and surface data are extensions of point data spanned by interpolation processes. But in this case, what is represented with point data has prime importance. The phenomena represented with point data may not possess the same characteristics. The presence of potential continuity of data must be determined. If the scattering pattern of "a specific event", examined in point pattern analysis, comes into question, then point data cannot be said to have continuity properties topologically.

Otherwise, if the scattering measurement of "a magnitude" is a matter of study, then point data can be extended to the length or surface data for the purpose of spatial interpolation. This case is commonly used in Geostatistics. Another issue of spatial data is whether it is continuous data or not. Continuous data can be defined as a data set that can be generated at every possible point by means of using interpolation of known points in the region. Apart from that definition, continuous data can be mathematically defined as a domain that can construct a surjective (onto) and injective (one-to-one) function to real numbers set.

Discrete data is a spatial sample in specific region coordinates; moreover, that can be considered a seed data set to generate continuous data. Continuous and Discrete mapping is based on these kinds of data.

Definition 2.2. Point Data

Point Data can be mathematically defined as

$$s_i \in D \subset R^n \text{ where } i \in \{1, 2, ..., n\}$$

Point data can represent contiguity-based data through convenient transforms, like Voronoi Diagrams, as also called as Thiessen polygons [27] [26].

Definition 2.3. Length Data

Length Data is defined a data array to describe a structure with one-dimension below

$$s_r \in D \subset R \text{ where } r \in R$$

s_r is assumed to be homeomorphic to real numbers R.

Rivers is said to be the best-known example of a phenomenon to be modeled by using length data. In this example, The points, observation stations on the river, are used for the points that do not include metering points through the medium of interpolation, using methods like Inverted Dimension Weighting (IDW) as usual. Thus each point on the river can be modeled by the known points. This is a regular length data process. Moreover, length data can also contain direction properties, like rivers.

Definition 2.4. Surface Data Surface Data can be defined as below

$$s = (s_1, s_2) \in D \subset R^2$$

D is assumed to be a manifold and homeomorphic to R^2.

As in Length Data, Surface Data is defined a data matrix to describe a structure with two dimensions. Air pollution is the best-known example of surface data. Similarly, each point on the surface can be determined by the known points through many methods, like Kriging and inverse distance weighting (IDW).

One should note that the length data and surface data are nothing but a particular arrangement of a group of point data on the ground of necessities of computation. In addition to that, Dimensional aspect of spatial data is almost completely related to geostatistical data on the ground of availability of interpolation [72][80][214][215] .

2.4.2 Usage in Model Aspect of Spatial Data

Unlike other science branches, modeling and usage of data in Spatial Analysis can not be confined to the definition or representation of data. For this reason, it is vital to classify the data types in conformity with the usage.

Geostatistics is based on prediction models by employing interpolation methods, like Kriging and IDW. The points of which value is known, in other saying, the measured points are always limited. The reason for that case can be financial or geographic constraints. Then the conditions make the prediction models, obtained by projections of the known values, necessary for the study. Thus the field, for example, a river or air pollution of a city, on which the analysis is done can be realistically modeled in many ways. Geostatistics includes all methods in this sense. Point Pattern analysis is directly related to scattering “a certain event category” over the study area. Because of the fact that Interpolation is out of the study, the dimensional aspect of spatial data is not essential. In this respect, data in point pattern analysis is simpler than other peer branches under spatial analysis.

2.4.3 Data in Lattice Data Analysis

Lattice data is a data type to which SAC Analysis is chiefly applied. So lattice data is to be evaluated in this dissertation. Unlike Geostatistics and Point Pattern Analysis, SAC Analysis focuses on the dynamic structure, the spatial units interacting with each other via lattice data. The spatial units can be in point form, for example, weather condition monitoring systems [148], or polygon form, for example, provincial data [70][216]. For this reason, it is essential to choose the data type, point, or polygon, in the beginning, taking the data and the goal of the study into consideration. SAC Analysis of point data structures is based on the distances among the spatial units. Through a threshold value on the inverse of the measured distance, the effect created by spatial units can be confined to a reasonable area. This case makes a neighborhood for each spatial data. Without threshold value, point data in Spatial Autocorrelation is said to be similar to fully connected networks in graph theory. SAC on polygon data is based on the contiguity among the spatial units. Concordantly spatial data can be considered to relate only to neighboring polygons at the first level. Moreover, there exist some methods $n-th$ degree neighborhood generated recursively with contiguity relation. Definition of boundary is a central theme for spatial analysis with contiguity based. It is assumed that the boundaries of polygons are of permeability, a borrowed term from soil science literature. It is also called the spillover effect. On the other hand, permeability is interpreted as the presence of boundary effects or contiguity, which is needed for SAC.

Some studies are approaching the field through a combination of contiguity-based and distance-based literature. For instance, Griffith uses five alternative methods based on the buffer zone, or in other words, outer perimeter, for the purpose of estimation of autoregressive models [217]. It is another significant point that the points in the buffer zone of the boundary are more dependent on the effects from the other side of the boundary than interior points. The smaller area or polygon is chosen as the spatial unit; the more substantial dependencies is observed [17]. As is seen, much research, including spatial contiguity-based data in social sciences, is based on administrative or demographic regions, like states, provinces, precincts [70][216]. And most of them have irregular shapes owing to historical, political, and geographical reasons. Aside from these facts, Grid–Lattice structure is introduced for much sensitive research in order to overcome the problems arising from irregular shapes of administrative regions. The reason for that can be political reasons. Particularly gerrymandering is a root cause of irregular shapes of administrative regions. This structure brings more effective mathematical modeling into use to analyze the study field. This arrangement provides the superiority of spatial neighborhood stationary, that is, each point or polygon has the same distance for point data, or adjacency, for contiguity-based data, relation in the chosen region, introduced by Tobler [218]. Another issue for Grid-Lattice Structure is the determination of the size of the grid. In this regard, Spatial Resolution is the size of the quadrat used in the research. The issue of Spatial Resolution is connected with the subject and budget of research. Tobler proposes the square root of the average area belonging to subregions for the purpose of assignment of spatial resolution. For example, Tobler offers 43 kilometers for the grid size in order to analyze data based on conterminous states in the USA, except Hawaii and Alaska by definition[219]. As a matter of fact, it is hard to detect a pattern in low spatial resolution. Particularly the processes pertinent to commuting, human interaction, or disease contagion do not manifest themselves in low spatial resolution [17]. Fortunately, recently dazzling improvements in GIS Technology provide high spatial resolution thanks to pixels in image processing. Sample Size for Spatial Analysis is another point to be emphasized.

Sample size in spatial analysis is directly relevant to fulfilling normal distribution assumption. Furthermore, it is essentially mentioned that SAC statistics, such as Moran's I, involve normal distribution assumptions. In literature, to provide a normal distribution assumption, Cliff and Ord propose a minimum of fifty observations [220]. However, it is difficult to suggest a minimum sample size due to the structure of the weighting matrices, to put in a different way, the location of spatial units [220]. Nevertheless, a small sample size can generate misleading results for categorical spatial data. Finally, in the literature, Cliff and Ord offer some methods for small samples in order to validate normal distribution assumption. Once for all, the spatial analysis involving contiguity-based units is intrinsically based on census. Therefore some primary issues for general statistics, such as sample size, verification of assumption, and estimation parameters, are secondary issues for spatial analysis. Particularly the instances in spatial autocorrelation are generated by randomization process.

Although it is desired to study with a regular grid structure, there does not exist a satisfying number of cases for the study owing to the properties of that field. Nevertheless, there exist some enlightening studies. For example, Sibert represents a study on the fluctuation of urban borders of Detroit by using bi-dimensional SAC [221]. Moreover, Gatrell examined the scattering of populations over Southern Germany by using Christaller's Data [222] through two-dimensional autocorrelation functions [223].

It is obvious that SAC Analysis for nominal data is more straightforward than for numerical data. However, there exist some points to regard as categorical data analysis. Binary classification based on continuous data has the potential to mislead the researcher by overestimating SAC when class B (Black) and W (White) are interpreted above and below the mean for the variable [17]. That fact is scrutinized in the following sections. Geostatistics and Point Pattern Analysis are not suited for Categorical Data in common sense.

3
Spatial Autocorrelation

Spatial Autocorrelation (SAC) can be considered an integrated form of autoregressive models in Statistics into Spatial Analysis. It reveals the relationship among spatial units. That is a basic definition based on autoregressive model studies. In a functional context, SAC is a measure to indicate whether values on a map are clustered or not. In this respect, it is also used Point Pattern Analysis for the problem mentioned in the previous sentence. However, unlike Point Pattern Analysis, SAC analysis focuses on a case-based event and all spatial data types in Spatial Analysis literature.

Moran's I and Geary's care the most used measures for SAC. Apart from these, General Getis-Ord and Joint-Count statistics are also employed for that purpose. Moreover, some statistical methods, like the Mantel test, are also used for that purpose.

3.1 Definitions of Spatial Autocorrelation

Spatial Autocorrelation (SAC) is first mentioned by Geographer M. F. Dacey (1932-2006) in the late 1950s [224]. Furthermore, W. L. Garrison (1924-2015), a scholar in one of the pioneers in Quantitative Revolution in Geography during the 1950s and 1960s, and E. Ullman (1912-1976), proposing gravity models for trade models, have a significant contribution to the sprouting of SAC as a concept in regional science at first, in spatial analysis at last [183]. One should state that even though Moran and Geary introduce fundamental measures [6][7][8] in SAC literature, A. Getis, and J.K. Ord make these measures popular in studies [183].

It is beyond doubt that the emergence of spatial autocorrelation as a concept is an output of studies over residuals in regression analysis, like standard autocorrelation analysis. As is known, the presence of autocorrelation in residuals violates the basic assumption of regression analysis, the independence of residuals. Furthermore, the presence of autocorrelation can cause other modeling problems. Data including autocorrelation generates a biased estimation of standard error, particularly in the difference of means of two sets [225]. In that regard, data in spatial analysis provides more tools for researchers due to the contiguity property of spatial units than classical statistics. Moreover, some definitions and remarks on SAC can be found in the literature. Some of them are represented below:

"Spatial Autocorrelation refers to similar behavior in space, unlike the temporal case, space may be two- or even three-dimensional. A general statement Waldo Tobler, often termed Tobler's first law of geography, asserts that SAC is positive for nearly all geographic phenomena." - Micheal Goodchild [163]

"A clustering pattern in the spatial pattern of some variable which seems to be due to the very fact the occurrences are physically close together, that is, that they are in geographical proximity. They are not independent of each other but somehow linked. In other words, the data are spatially dependent." [162]

"The presence of spatial pattern in a mapped variable due to geographical proximity. The most common form of SAC is where similar values for a variable (such as county income levels) tend to cluster together in adjacent observation-units or REGIONS, so that on average across the map the values for neighbours are more similar than would occur if the allocation of values to observation-units were the result of purely random mechanism. This is positive SAC. Negative SAC is where neighboring regions are significantly dissimilar; more general and complicated forms of autocorrelation can be defined."-Less Hepple [161]

"This occurs when the observations of a variable are mapped and the resulting SPATIAL PATTERN shows that neighboring values in that pattern are either more alike or more dissimilar than would be the case if the pattern were due to RANDOM processes." [160]

Of those definitions, Less Hepple gives the most explanatory and inclusive definition. In both social and natural science, the existence of SAC manifests itself in many ways. For example, in the economy, prices of goods do not vary in a considerable spatial scope, clearly, stability of prices in the region says this fact. Then it can be said that positive autocorrelation exists. The same conditions are valid for the real-estate market. Unless there exists a formidable barrier, like a mountain, geographical climate, rainfall, and temperature, data have positive autocorrelation. That is to say, the transition of weather conditions between adjacent neighboring regions is such as to be smooth. Then almost heterogeneous clusters occur on the map.

On the other hand, it would be an excellent example of the existence of negative autocorrelation that the assignment of ATM of a small-scale bank is of scattered pattern in a region. Similarly, the distribution of airports in a country is expected to show negative autocorrelation because accessibility of airports in a country is provided in that way. Then almost homogeneous distribution occur on the map.

Weighting has a vital role in SAC because weighting attaches spatial units within a mathematical model. Configuration of weighting reflects the degree and meaning of SAC. Weighting is also the creator of spatial lag for the spatial effect to be examined. Within this context, lag is defined as a measure involving the value of neighboring units described in spatial modeling. That lag can be based on distance or contiguity. That dissertation focuses on the potential use of the interaction of spatial lag and weightings. SAC is examined under two approaches, global and local SAC. Global SAC had directly meant SAC until Anselin introduced Local Indicators of Spatial Association (LISA), in other words Local SAC, to the literature in 1995 [19]. Local SAC deals with the spatial association of the sole spatial unit with its neighboring units, whereas Global SAC indicates the degree of spatial association in the whole structure.

3.2 Relations To Other Concepts

In spatial analysis literature relevant to SAC, some concepts , like spatial association, dependence, interaction and interdependence, are mentioned [75][47][57]. Despite the elusive nature of determination for boundaries among terms in literature, nuances turn out in relevant studies even a little. The reason for this connotation or the likes of these terms dominates SAC literature. At that point, it is indispensable to elucidate the relation of SAC with this concept in order to help to determine the boundaries of the actual scope of the term, SAC in related fields. In this dissertation, studies are classified into four categories: including and using SAC methods, mentioning SAC, mentioning the related term, like spatial association, dependence, interaction, and interdependence in SAC literature, and not mentioning SAC.

3.2.1 Spatial Association

Although the term, Spatial Association has no explicit definition in fundamental encyclopedic dictionaries in geographical science literature [160][161][162][163], it is frequently used in studies relating to spatial analysis. In this sense, Spatial Association must be considered as only adjective and noun separately, rather than the united term "adjective+noun", like "SAC". Nonetheless, one can find some references to spatial association:

“Spatial association links information sets, social processes, and problems to geographic coordinates and regions. For example, maps of environmental quality and human health can be overlaid to examine correlations that may suggest clues for further research." – Donald Janelle and Micheal Goodchild [163]

The first sentence of the definition can be attributed to “georeferenced data”, but the following example in the definition provides insight into spatial association by touching upon correlation. In order to determine the degree of Spatial Association, SAC measures are mostly used. Strikingly, leading studies of SAC use spatial association as an explanatory concept. Local SAC has also been described as “Local Indicators of Spatial Association-LISA” ever since Luc Anselin published the introductory text of Local SAC [19]. Introductory articles of Getis-Ord [14] and Geary’s c statistics[22] are developed under spatial association title . In literature, the relation between baleen whales and euphausia superba population in Antarctica by employing spatial regression analysis [226], carbon emission density of forest floor in southeastern China [227], performance analysis of landscape genome models [228] are studies investigating and using spatial association with SAC methods. Moreover SAC studies deal with spatial association in the study fields [80][57][128][69][60][142][70][42][25].

Moreover, some spatial association studies takes only SAC into consideration without using methods [229][230][231]. However, many studies examines spatial association in a bounded field, there is no any mention of the term, SAC [232][233][234][235][236][237].

3.2.2 Spatial Dependence

One topic in which SAC methods are used in Spatial Analysis is Spatial Dependence. Spatial dependence is defined below

“Spatial dependence is a characteristic of distribution of geographic data, that is, data where the location of observations is explicitly taken into account. It combines the notion of attribute similarity with that of locational similarity. Not only is there dependence (correlation) between observations for a given variable, but also this dependence shows a spatial structure such as closer locations being more similar than locations that are farther apart.- Luc Anselin" [163]

In the light of the definition above, it is clear that SAC statistics are used to determine the presence of spatial dependence. Furthermore, Unwin and Hepple tend to define Spatial Dependence as a lack of statistical dependence in spatial data [238]. SAC can also indicate the degree of spatial dependence. When one dilates upon spatial dependency, the studies only considering SAC outnumbers ones using SAC methods. Anselin uses spatial dependence and SAC [143]. Regional unemployment in Australia for 1991 and 2001 by using distance-based methods [239], irregular natural forest formation in new south wales Australia [240], ecological vegetation models by employing Moran's ı and Geary's c [241], savannah ecosystem analysis through lions and their preys through the agency of water dependency [242], commercial property analysis in the Netherlands as Spatio-temporal approach [243], regional economic growth in US BEA Economic Areas in the period 1969 to 2009 [244] are investigated with SAC methods in order to determine existence and degree of spatial dependence. SAC literature also includes studies with spatial dependence [146][61][45][105][57][129][121][147][132][127][135]. Furthermore in literature one can find studies mentioning SAC [245][246][247] or spatial correlation[248][249][250][251] but not using traditional methods of SAC. Apart from that classifications, There exist spatial dependence studies including having no reference to SAC or correlation , even only as a concept [252][253][254][255].

3.2.3 Spatial Interaction

Spatial Interaction is another topic related to SAC. Edward Louis Ullman, the American geographer, made spatial Interaction more popular by writing an introductory book, "Geography as spatial interaction," in 1980. According to Leslie Hepple, a British geographer, Ullman's goal was to unify the scope of geographical science through that concept, but he failed. Many researchers use the term for more specific fields [256][161]. Spatial Interaction is a concept used in architecture and sociological contexts, like urban planning and economics, rather than in traditional quantitative geographical sciences. For this reason, many definitions of spatial Interaction involve some traces from social sciences. Definitions of spatial interactions are given below:

“The interdependence of areas; the movement of people, capital, goods, information, ideas, etc. between places. Ulmman has suggested that the degree of spatial interaction between places is conditioned by three factor (i) Complementarity; (ii) intervening opportunity (see Intervening Opportunity Theory); (iii) transferability .” [160]

“Spatial interactions refer to movements among places-migrations among nations, traffic flows within cities, commodity flows among regions, and so forth" - Donald Janelle and Micheal Goodchild [163]

"(as spatial interaction theory) The view that the movement of persons between places can be expressed in terms of the attributes, such as population or employment rates, of each place.” [162]

"Spatial interaction...to indicate the interdependence between geographical regions,... as complementary to the more traditional emphasis on relationships (people and their social context, people and environment) within each individual region.- Leslie Hepple" [161]

SAC methods are seen in geographical studies relevant to spatial interaction. In some studies, spatial interaction is used as a synonym of SAC [53] [143]. The problems caused by the urbanization of Yahara Watershed, Wisconsin, USA [257] deal with SAC methods in order to detect spatial interaction. One study about spatial interaction interestingly attributes to SAC as a spatial lag autocorrelation [258]. Studies including spatial interaction can be found in SAC literature[27][38][175][141][55][52][144][75]. Some studies in literature touch upon SAC, not in a profound manner [259].

However overwhelming majority of studies in spatial interaction literature does not have any mention of SAC [260][261][262][263][264]

3.2.4 Spatial Interdependence

Spatial Interdependence is one of the fields in which SAC methods are employed. Even though there is no clear definition of spatial Interdependence in fundamental encyclopedic dictionaries in geographical science literature [160] [161] [162] [163], there is a noteworthy number of studies in that field. AFDC (Aid to Families with Dependent Children) benefits in Missouri USA 1989-1990 with various models [265], growth of agglomeration economies[266], and daily household travel conducted in Shaoxing City China 2007 [267] are spatial interdependence studies employing SAC methods. Moreover, certain articles in SAC literature use spatial Interdependence as a field[57] [70]. Some studies mention SAC [268] [269][270]. Nevertheless, many studies deal with spatial interdependences without mentioning SAC [271] [272] [273].

3.3 Weightings

Weighting has an essential role in Spatial Analysis. Because weighting binds each other spatial units on a map or a space, in other words, weight is the fundamental relation to configuring spatial units. In this regard, weighting is considered to be an abstraction of the relation among spatial units. If weight is used as a concept in Spatial Analysis, it is more convenient to call spatial weight in order to attribute spatial property in the study.

Spatial weighting can be defined in conformity with the context in which research studies. Contiguity and distance are the primary factors in creating spatial weight. Whereas distance is a standard Euclid measure on a map, as usual, contiguity grounds how to determine boundary weight among spatial units. Thus contiguity can be different from one study to another study. Subsequent to the concept of boundary, spatial weighting functions are a vital issue for SAC. That is why the choosing of spatial weight function directly influences the SAC measure. First of all, it is indispensable to touch upon Tobler's famous quote, " first law of geography: everything is related to everything else, but near things are more related than distant things" [274].

As is known, location and contiguity are the most significant attributions of spatial units naturally on account of being "spatial". The majority of studies in the spatial analysis are based on that trite phrase. A contiguity matrix is a rectangular array constructed in order to represent the contiguity, or the adjacency in the context of polygon units, among spatial units in the mathematical model. It must be noted that the contiguity matrix is also called as "Adjacency Matrix" in many studies [85][39][134][22]. It is defined for spatial units based on contiguity as

$$w_{ij} = \begin{cases} w(i,j)(\neq 0) & \textit{units i is adjacent to unit j} \\ 0 & otherwise \end{cases} \tag{3.1}$$

$w(i,j)$ is a non-zero determined by choosing to depend upon the scope of the study. In most of the studies, contiguity matrix is in a binary form such as

$$w_{ij} := \begin{cases} 1 & \textit{units j is a neighbor of the unit i} \\ 0 & otherwise \end{cases} \tag{3.2}$$

To interpret the spatial lag easily, one use row-standardized contiguity matrix such that

$$w^s_{ij} := \begin{cases} \frac{1}{n_i} & \textit{units j is a neighbor of the unit i} \\ 0 & otherwise \end{cases} \tag{3.3}$$

where n_i is the size of neighborhood set, or the count of neighbors.

In distance based data, spatial weighting is defined in general form

$$w_{ij} := \begin{cases} 0 & i = j \\ f(d_{i,j}) & otherwise \end{cases} \tag{3.4}$$

Where d_{ij} is euclidean distance between unit i and j in the form $d_{ij} := |s_i - s_j|$, s_i

and s_i are coordinates of unit i and j, respectively. And function $f(d_{ij})$ is inversely proportional to d_{ij}. In many researchers employ distance based spatial matrix in this concrete form below

$$w_{ij} := \begin{cases} 0 & i = j \\ \frac{1}{d_{ij}} & otherwise \end{cases} \tag{3.5}$$

In literature, many spatial weighting functions are developed, as they will be defined in this dissertation. Before anything else, the form of spatial weights is related to characteristics of data, contiguity, or distance. In data based on distance, by its nature, a few spatial weights are offered.

In addition to this form, Tobler lists these technics for spatial weights as

$$w_{ij} = \begin{cases} 1 & d_{ij} \leq t \\ 0 & otherwise \end{cases} \tag{3.6}$$

where t threshold value. These spatial Weights are based on polygonwise data. Cliff and Ord offers spatial weights such as

$$w_{ij}(\alpha, \beta) = \frac{p_{ij}^{\alpha}}{d_{ij}^{\beta}} \tag{3.7}$$

Where p_{ij} is the proportion of the boundary of unit i that is adjacent to unit j and d_{ij} is the distance between centers of unit i and unit j [275] . α and β are predetermined positive numbers. In addition to these approaches, measured distances in the study field are divided into "distance classes" [91][152][116][80]. Distance class is also called as "Diameter class" [129].

Weighting is also crucial for local SAC analysis based on the permutational method mentioned in the relevant section. If one chooses equal weights for each neighbor, variance remains constant for all permutational reshuffling. The permutational method and pseudo p-value do not give any information about SAC for the study area. As distances among spatial units tend to be different, the permutational method for distances based studies is said to give more significant results than for contiguity-based studies.

As an advanced technique, neighboring relation is expanded into higher-order neighboring structures. Thus every possible pair of the spatial unit indicate an n-order spatial weighting for suitable n. Herein it is expected that SAC statistics decrease, or in another saying, decay into 0 for Moran's I or 1 for Geary's c, as the n-order increases.

This fact validates Tobler's law. The n-order spatial weightings are easily obtained by matrix algebra [225].

The Uniform grid structure represents a rectangle consisting of $m\ x\ n$ squares. A structure comprised of quadrats is substantially equivalent to a uniform grid structure practically. It provides simplicity for the researcher in view of computation and interpretation. Uniform grid structure has two approaches to the contiguity concept. Whereas rook contiguity is grounded on only the existence of common boundaries, in other saying, edges in graph theory, Queen contiguity is based on the presence of common points together with common boundaries in rook contiguity. Thus, queen contiguity conceptually includes rook contiguity by definition, as in chess.

Moreover, regression model based on rook contiguity is defined as [85] [86] [17]

$$x_{h,v} = Rx_{h,v} + \epsilon \tag{3.8}$$

where

$$Rx_{h,v} = a + b_{left}x_{h-1,v} + b_{right}x_{h+1,v} + b_{down}x_{h,v-1} + b_{up}x_{h,v+1} \tag{3.9}$$

Regression Model based on Queen Contiguity With Grid Lattice Structure is defined [85]as

$$x_{h,v} = Qx_{h,v} + \epsilon \tag{3.10}$$

where

$$Qx_{h,v} = a + b_{left,down}x_{h-1,v-1} + b_{right,down}x_{h+1,v-1} + b_{left,up}x_{h-1,v+1} + b_{right,up}x_{h+1,v+1} + Rx_{h,v} \tag{3.11}$$

The regression model does not include spatial units staying outside, neighbors of units on the border of structure. This model is based on the Cartesian coordinates system. However, this model can be applied for inclined diamond shape coordinate systems in a general sense. Both rook [129][157] and queen [133][65][113][140] contiguity are used in SAC literature. Moreover there are also comparative studies for both approaches [60][14]. In many studies, Quadrats are very compatible with these contiguity forms [276][129][102].

Spatial Weighting is used for epidemic diseases research as the first studies [277]. Matula and Sokal introduce spatial weights for Biological fields [278]. There is a comparison of five spatial weights for spatial diffusion of epidemic diseases among subregions of Ibadan by using contiguity properties [279]. Correlograms are also an enlighting tool for SAC analysis. Sokal and Mennozi use spatial correlogram to examine the opinion of demic diffusion of early Middle Eastern farmers into Europe [280]. Computational aspect of spatial weighting is used for GIS [281][159]. Alternative methods for distance-based [24] and contiguity based [25] spatial weight are examined in literature.

3.4 Spatial Lag

Spatial Lag is one of the central concepts of spatial analysis. Defined in a general sense, Spatial Lag is a function of units in a neighborhood. Hence the definition of spatial Lag can be defined as

$$Wz_i = f(W, z_i) \tag{3.12}$$

Where W is the spatial weight matrix and Z is the vector for the magnitude of units in the space. In this form, it should not be thought that the units contributing to the spatial Lag are from only adjacent or notably near neighborhoods. Depending on the model, each unit can be related to each at a certain level. However the concept of spatial lag is confined in linear combination of units in neighborhood in literature. Then spatial lag is transformed into the form

$$Wz_i = \sum_{i=1}^{n} w_{ij} x_j \tag{3.13}$$

The conceptual origin of the spatial Lag is found in time series analysis. Whereas Lag in time series is attributed to data coming from the previous time unit of an incident, Lag in spatial analysis refers to data in the vicinity of a spatial unit, point, or polygon. Each study, including SAC, can be considered a spatial lag study on account of the central role of spatial Lag in literature. In addition to that fact, spatial Lag is latently used to estimate the value at a point by using IDW [182]. Lee introduces a new lag statistics combining non-spatial correlation Pearson-r and SAC measure Moran's I by using Spatial Smoothing Scalar as an example of using spatial lag [282].

It is notable research that Junior and his co-author's study on the classification of whether tumors are benign and malignant by using SAC measures, Moran's I and Geary's c. Moreover, This study has a great contribution to spatial analysis literature in connection with spatial analysis in micro scales [283].

One of the goals of the dissertation is to reveal the meaningful conceptual nuances among alternative spatial lag approaches.

3.5 Spatial Regression

First of all, one should emphasize on the fact that one of fundamental motivation underlying SAC studies is violation of regression assumption: independence and identical distribution of residual errors [81][139][157][110][34][55][102]. To put it in a different way, SAC, as a concept, is a output of residual analysis emerged from regression analysis. That is why the nuance between the regression model with spatial factor and regression model without spatial property is unpredictable in most cases [284]. Given that fact, it is indispensable to deal with SAC in case of the presence of autocorrelation among residuals with the spatial property.

Concordantly there are two approaches to overcome the problem of assumption violation. One of them is to put spatial effect into the regression model, Spatial Lag Model. Another is to use spatial effect for modeling residual.

Some studies criticize spatial regression models. [285][286][287]. Because the reason for autocorrelation manifested itself spatially may not be spill-over effects as a primary result of SAC. It can be a common third factor. Some studies emphasize that Spatial weight matrice may not reflect the realities relevant to the relation of social science in some cases [288].

Spatial Lag Model (SLM) is defined to be a model in which spatial lag is directly employed. That can be formalized as

$$y = \rho W y + \beta x + \varepsilon \tag{3.14}$$

Where x independent variable, y dependent variable,Wy spatial lag for y, ε residual error and ρ spatial coefficient [11][53][56].

If $\rho = 0$, then the regression model can be said to be spatially independent. SLM is also termed as Simultaneous Autoregressive Model (SAM) [289][290][291].

The spatial Error Model (SEM) is based on the question of whether the residuals of the model are autocorrelated to each other or not. Then Spatial Error Model is depicted as

$$y = \beta x + \varepsilon \qquad\qquad \varepsilon = \lambda W \varepsilon + \zeta \tag{3.15}$$

Where x independent variable, y dependent variable, ε error, $W\varepsilon$ spatial lag for ε and λ spatial error coefficient and ζspatially uncorrelated error term [11][53][292][56]. In order to determine coefficients in two approaches, Ordinary Least Squares (OLS) is used as if it is basic regression model in question.

Spatial regression models are frequently used in spatial analysis literature. Furthermore, Anselin, one of the pioneers of spatial econometrics, attaches great importance to spatial regression [143]. Cancer cases in Saudi Arabia [80], regional economic performance in EU [62], voting participation in France [142] and urban transportation in Calabria Italy [144] are topics that Spatial Lag Model is used. In addition to those, some studies include both models. Reef ecology [105] and non-farm enterprise performance in rural areas from Ethiopia and Nigeria [56] are studied, including both the spatial lag model and spatial error model. These studies also represent an informative comparison between the two models. Theoretical studies about regression models also illuminate application fields [23][55][52][45]. Some theoretical studies involve a comparison of other regression models in spatial analysis [34][129]. One should note that regression models used in spatial analysis can not be confined in spatial regression models mentioned above. Geographically Weighted Regression (GWR), Conditional Autoregressive Model (CAM) and Moving Average Model (MAM) are used models in that field [289][290][293][291]. GWR is a central method in Geostatistics [294][295][296][297]. However many SAC studies do not exclude GWR [80][138][129][50][148].

3.6 Measures For Global Spatial Autocorrelation

In Spatial Analysis, it is an important question of how to measure the interaction among spatial units. For the purpose of the answer to these questions, SAC is introduced, and some measures are developed. Moreover, it is another question in what scope that measure is evaluated, that is to say, in global or in local. Whereas Global SAC is relevant to a whole region consisting of spatial points or polygons, Local SAC pertains to interaction with its neighboring units of a spatial unit.

Global SAC depends upon what measure and method are chosen. One of the goals of the thesis is an attempt to elucidate how and to what degree a variety of measures and methods to be chosen differentiates measures of SAC. Prior to mentioning these measures, it is indispensable to set forth the concepts of contiguity (or adjacency in some references) and spatial weight matrices.

Whether spatial data verifies asymptomatically normal distribution assumption or not is another influencing factor in global SAC measures, for Moran's I and Geary's c. Even though expected values of Moran's I and Geary's c, the variance of these measures differentiate. If spatial data may be supposed to violate asymptomatically normal distribution assumption, then it is thought to be randomly distributed, independent of any classical distribution model. In the latter approach, data with random structure covers the former approach, data with normal distribution, and formulas for data with the random structure are generally used to compute global SAC studies in literature by virtue of practical reasons. Rare studies consider that nuance [220][18][298][299].

In consideration of SAC, it is critical to determine whether autocorrelation is positive or negative. Negative SAC refers to a dispersed structure in the region. Then the spatial structure is similar to chessboard structure, in another saying, heterogeneous structure. In almost fragmentation of field with the considerable size, it can be found dichotomous variables in almost equal numbers or quantities. Positive SAC means the existence of a polarized structure. In other words, spatial structure with positive autocorrelation is said to be clustered structure, in other saying, homogeneous structure. In this case, spatial units are positively correlated by definition. Apart from these cases, the absence of SAC indicates a random structure. In this case, every sub-regions of the spatial area have unique characteristics. It is another question of how to hypothesize over SAC. It is essential to state that a contiguity-based study is based on population, whereas a distance-based study is based on a sample. This is an important point in the evaluation of expectations in studies, including SAC analysis.

Many theoretical studies on SAC supporting and enriching spatial analysis and its applications are made. Legendre discusses whether SAC is advantageous or not in some studies [20]. In order to ameliorate the regression model, Cliff and Ord scrutinize residuals using SAC [9]. Apart from these studies, latent negative SAC is investigated [26], and local estimation on simulated econometric data is made by SAC [52].

One must bear in mind that global SAC tests, for example, Moran's I , Geary's c, are based on normal distribution assumption [220].

Apart from these SAC measures (Moran's I and Geary's, etc.), Hubert introduces some statistics to test the significance of SAC by applying methods including permutation [300].

3.6.1 Moran's I

Firstly it can be said that Moran's I is the most used SAC measure. This statistics is based on the studies of Australian Statistician Patrick Alfred Pierce Moran, who enriched geometric probability theory and its applications to evolutionary genetics and to population, [6][7] . Moran's I defined as follows

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{3.16}$$

Moran's I is a correlation value between data and spatial effect of data, in other words, spatial lag. Then Moran's I is located between -1 and 1. If Moran's I is negative, Then It means negative SAC, in other words, dispersed structure. Likewise, the case of Moran's I being positive means to the existence of positive SAC. If Moran's I is close or equal to 0, then it means the absence of autocorrelation and the existence of random structure.

Moran's I is with normal distribution such that The expected value of Moran's I is

$$E(I) = \frac{-1}{n-1} \tag{3.17}$$

And the variance of Moran's I for data independent of asymptomatically normal distribution assumption is [175][220][18],

$$V(I) = E(I^2) - E(I)^2 = \frac{nP_1 - P_2P_3}{(n-1)(n-2)(n-3)W} - E(I)^2 \tag{3.18}$$

where

$$P_1 = (n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2 \tag{3.19}$$

$$P_2 = \frac{m_4}{m_2^2} \tag{3.20}$$

$$P_3 = (n^2 - n)S_1 - 2nS_2 + 6S_0^2 \tag{3.21}$$

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \tag{3.22}$$

$$S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij} + w_{ji})^2 \tag{3.23}$$

$$S_2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji} \right)^2 \tag{3.24}$$

$$m_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \qquad m_4 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n} \tag{3.25}$$

Variance for data with asymptomatically normal distribution assumption [18],

$$V(I) = \frac{n^2 S_1 - nS_2 + 3S_0^2}{S_0^2(n^2 - 1)} - E(I)^2 \tag{3.26}$$

One should keep in mind that the variance equation (3.18) is used in applications unless otherwise stated. The reason for this is a tendency to choose more realistic models.

It is essential to note that Moran's I is nothing but the slope of the regression line through the Moran scatter plot for standardized data [19][12][143][26][63]. In order to understand that fact, one should take spatial lag and weight into consideration. As is known, The Moran scatter plot is a representation of the product space of variable and spatial lag. That property of Moran's I brings visual explanation and interpretation power to SAC. Incontrovertibly it is easily observed that studies including SAC analysis have a propensity for employing Moran's I. Anselin is one of the first scholars to mention the property [19][301][22][57]. Moreover, it is noted that the Moran's I is an indicator of association between variable and average of neighboring values [22]. The slope coefficient is considered as unstandardized Moran coefficient [50] and as SAC Index (SAI) [64]. The interpretation of linear regression slope in a similar way to Moran's I is also scrutinized under influential data analysis [302][303][304]. These studies are examples of Moran's I method in non-spatial analyses. Extension of SAC methods into non-spatial studies will be mentioned ahead. Moreover, the slope is used for detection of low SAC in the Digital Elevation Model (DEM) for the geographic purposes [26] and is interpreted as Moran's I value in unemployment rates in Turkey [63].

Moran's I, as a global autocorrelation measure, can be decomposed into LISA [19], as a local autocorrelation measure [22]. In other words, Moran's I can unite global and local approaches under the same model. Geary's c , another SAC measure, does not possess such a property.

In keeping with Tobler's law [274], many studies show that Moran's I tends to be zero as distance increases [276][35] [129] [106][113][140]. However there are some exceptions [102] [152].

Moran's I have some different namings, like SAC in literature. Moran Coefficient (MC) is another names for Moran's I in Griffith's studies [26][50][201]. SAC index (SAI) [38] are equivalent indexes of Moran's I. Applications Applied SAC literature teems with the studies using Moran's I . Moran's is used for biology and genetics [93][94][90][91][97][100], forestry [126][129], economy [53][55][60][56][69] [70][63], ecology [106][104][118][105][117][120], medicine [279][79][80], meteorology and Climatology [147][113][148] , computer science[33], risk Analysis [137][140], real estate market analysis [73], transportation [144], environment [150], agriculture [132][133][134], social science [305][142] and linguistics [152].

Evidently, Moran's I is the most used and referenced measure for SAC in literature. Our dissertation brings Moran's I into focus in local and global approaches.

3.6.2 Geary's c

Geary's c is another measure of SAC. This statistic was created by Irish Statistician Robert Charles Geary, known for his studies on Regression Analysis and Economic Theory, in 1954 [8]. Geary's c can be considered as a extension of the von Neumann ratio in spatial concept [306][8][41]. Geary's c tends to be based on the nuance among adjacent spatial units. Furthermore, one can say that Geary's c is a projection of the approach of Durbin-Watson Statistics in time series analysis. Geary's c is defined as

$$c = \frac{n-1}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - x_j)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3.27}$$

Geary's c is always positive and varies on a scale from 0 to 2 by definition. Interpretation of Geary's c is different from Moran's I. As Moran's I value increases, this means to be positive autocorrelation. But Geary's c is the opposite. If Geary's c is below 1, the spatial structure possesses positive autocorrelation. If Geary's c is above 1, then spatial structure keeps negative autocorrelation. If Geary's c is close to 1, it is of random SAC.

Moreover, Geary's c can be transformed into the alternative index c' to interpret Geary's c like Moran's I.

$$c' = 1 - c \tag{3.28}$$

Geary's c is also with normal distribution. And the expected value for Geary's c is

$$E(c) = 1 \tag{3.29}$$

And Variance of Geary's c is given for data independent of asymptomatically normal distribution assumption below

$$V(c) = \frac{T_1 + T_2 + T_3}{4n(n-2)(n-3)S_4} \tag{3.30}$$

$$T_1 = 4(n-1)S_1[n^2 - 3n + 3 - (n-1)P_2] \tag{3.31}$$

$$T_2 = -(n-1)S_2[n^2 - 3n - 6 - (n^2 - n + 2)P_2] \tag{3.32}$$

$$T_3 = 4S_0^2[n^2 - 3 - (n-1)^2 P_2] \tag{3.33}$$

Variance of Geary's c for data with asymptomatically normal distribution assumption [18],

$$V(c) = \frac{(2S_1 + S_2)(n-1) - 4S_0^2}{2(n+1)S_0^2} \tag{3.34}$$

S_1, S_2, P_2 and S_0 are given in equations respectively (3.23), (3.24), (3.20) and (3.22).

In comparison to studies including Moran's I, the studies in which Geary's c is used constitute a tiny part of SAC. Linear regions, in other saying, ordered spatial units [15], spatial effects of regression residuals [22], SAC tests with Bootstrap [41] are examined with Geary's c. In addition to those studies, There exist studies mentioning Geary's c in a superficial or introductory manner [84][94][90][147].

3.6.3 General Getis-Ord

Getis-Ord General G statistics is a meaningful measure for Global SAC and is also found as a tool in ArcGIS, despite its minimal recognition. General Getis-Ord G statistics can be deemed as an extension of Local Getis-Ord statistics. General Getis-Ord statistics focus on the degree of clustering on the map as a whole. In this context, it is more informative and comparable to other Global SAC measures for general Getis Ord statistics is a measure for SAC. Some studies in the literature include comparison or examination together with Global Moran's I [66][67][68]. Some study uses also Local Moran's I (LISA) [145]. The general G Statistics is defined as

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j} \qquad where \qquad \forall j \neq i \tag{3.35}$$

The Estimation Value of G-Statistics is

$$E(G) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}{n(n-1)} = \frac{S_0}{n(n-1)} \qquad where \qquad \forall j \neq i \tag{3.36}$$

And The variation of G-Statistics is [14][158]

$$V(G) = E(G^2) - E(G)^2 \tag{3.37}$$

Where $E(G^2)$ can be computed as [14]

$$E(G^2) = \frac{A+B}{C} \tag{3.38}$$

where

$$A = D_0\left(\sum_{i=1}^{n} x_i^2\right)^2 + D_1\sum_{i=1}^{n} x_i^4 + D_2\left(\sum_{i=1}^{n} x_i\right)^2 \sum_{i=1}^{n} x_i^2 \tag{3.39}$$

$$B = D_3\left(\sum_{i=1}^{n} x_i\right)\sum_{i=1}^{n} x_i^3 + D_4\left(\sum_{i=1}^{n} x_i\right)^4 \tag{3.40}$$

$$C = n(n-1)(n-2)(n-3)\left[\left(\sum_{i=1}^{n} x_i\right)^2 - \sum_{i=1}^{n} x_i^2\right]^2 \tag{3.41}$$

$$D_0 = (n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2 \tag{3.42}$$

$$D_1 = -[n(n-1)S_1 - 2nS_2 + 6S_0^2] \tag{3.43}$$

$$D_2 = -[2nS_1 - (n+3)S_2 + 6S_0^2] \tag{3.44}$$

$$D_3 = 4(n-1)S_1 - (n+1)S_2 + 8S_0^2 \tag{3.45}$$

$$D_4 = S_1 - S_2 + S_0^2 \tag{3.46}$$

S_0,S_1 and S_2 are given in equations respectively (3.22), (3.23) and (3.24).

In order to eliminate the condition $\forall j \neq i$ in (3.35), an equivalent form of General Getis-Ord statistics can be used in computations, defined below

$$G = \frac{(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}x_i x_j) - (\sum_{i=1}^{n} w_{ii}x_i^2)}{(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j) - (\sum_{i=1}^{n} x_i^2)} \tag{3.47}$$

Moreover, since SAC deal with the relation between the spatial unit and its neighbors, excluding itself, the overwhelming majority of studies in SAC literature assume a weighting matrix such that

$$w_{ii} = 0 \quad i = 1, 2, \ldots, n \tag{3.48}$$

General Getis-Ord statistics are said to employ exact weighting that other SAC measures use. Thereby components of expected value and variances like S_0,S_1, and S_2 are identical to components given other SAC measures.

Unlike Moran's I and Geary's c, General Getis-Ord statistics are not construed in a specific interval. That is to say; there are no boundaries for General Getis-Ord. So null hypothesis is focused rather than the parameter. In order to interpret statistics, it must be standardized as follows

$$z_G = \frac{G - E(G)}{\sqrt{V(G)}} \tag{3.49}$$

In contrast with Local Getis-Ord Statistics, General Getis-Ord Statistics is not prevalent in literature. Foreign direct investments in China [66] effect of foreign direct investment over metropolitan area in Wuhan China [67], urban expansion in Yangtze river delta, China by regression models and panel data [145], Salmonella Enteritidis infection in Toronto, Canada between 2007-2009 [82] and spatial inequality emerged from foreign direct investments in China [68] are examined by using General Getis-Ord statistics.

3.6.4 Joint-Count Statistics

SAC Approach is developed for also categorical data. This methodological expansion helps us to understand whether the spatial effect of categorical properties over a region exists or not. However, it cannot be expressed that SAC is used for all types of categorical data. Due to practical conceptual reasons, like focusing on boundaries between regions with two different or same classes, SAC for categorical data is only applied to data with two categories, such as black-white, high-low, presence-absence, true-false, etc. However, some studies for cases including more than two categories exist in literature [220]. SAC analysis for is carried out through Black-White Categorization [85] [17]. Spatial sampling is an important issue for the determination of confidence interval and test of hypothesis for categorical data. For this reason, joint-count statistics are divided into sampling with replacement and sampling without replacement.

First of all, It is required the fundamental variables for categorical data. x_i is a categorical property of $i-th$ region and is defined as

$$x_i = \begin{cases} 1 & if\ \ region\ i\ \ is\ black \\ 0 & otherwise\ (white) \end{cases} \tag{3.50}$$

The statistics used in spatial categorical data are defined as

$$BB = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}x_i x_j \tag{3.51}$$

$$BW = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - x_j)^2 \tag{3.52}$$

BB is a statistic to measure the spatial association of property under research, whereas BW is a statistic to measure spatial disassociation. These two statistics are assumed to have a normal distribution with mean and variance depending upon sampling with or without replacement. As for WW statistics, that can be calculated with the reconfiguration of x_i. Moreover, these statistics are termed the Joint-Count Statistics by virtue of the usage of the count of the boundary between regions with the same or different properties [17]. Cliff and Ord study on effects of BB and BW statistics in certain cases [277][17].

a) Sampling with replacement

This case is based on the apriori assumption of p, the probability of black-region. Clearly, the probability of two black regions coming side by side becomes p^2. Thus Expected Values and Variances for BB and BW are

$$E(BB) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}p^2 \tag{3.53}$$

$$E(BW) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}p(1-p) \tag{3.54}$$

$$V(BB) = \frac{1}{4}p^2(1-p)[S_1(1-p)+S_2p] \tag{3.55}$$

$$V(BW) = \frac{1}{4}p(1-p)\{4S_1+S_2[1-4p(1-p)]\} \tag{3.56}$$

where

$$S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij}+w_{ji})^2 \tag{3.57}$$

$$S_2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right)^2 \tag{3.58}$$

$$S_3 = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)^2 \tag{3.59}$$

b) Sampling with replacement

This case is based on sampling from a population with a hypergeometric distribution. In this case number of black and white regions are known. Thus

$$n_b = \sharp(black\ regions) \tag{3.60}$$

$$n_w = \sharp(white\ regions) \tag{3.61}$$

And then

$$n = n_b + n_w \tag{3.62}$$

Expected Values and Variances for BB and BW are

$$E(BB) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} p_{bb} \tag{3.63}$$

$$E(BW) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} p_{bw} \tag{3.64}$$

$$V(BB) = \frac{1}{4}S_1 p_{bb} + (S_1 - 2S_2)p_{bbb} + (S_3 + S_1 - S_2)p_{bbbb} - E(BB)^2 \tag{3.65}$$

$$V(BW) = \left(S_2 - \frac{3}{2}S_1\right)p_{bw} + 4(S_3 + S_1 - S_2)p_{bbww} - E(BW)^2 \tag{3.66}$$

Where

$$p_{bb} = \frac{n_b}{n}\frac{n_b - 1}{n-1} \tag{3.67}$$

$$p_{bw} = \frac{n_b}{n}\frac{n_w}{n-1} \tag{3.68}$$

$$p_{bbb} = \frac{n_b}{n}\frac{n_b - 1}{n-1}\frac{n_b - 2}{n-2} \tag{3.69}$$

$$p_{bbbb} = \frac{n_b}{n}\frac{n_b - 1}{n-1}\frac{n_b - 2}{n-2}\frac{n_b - 3}{n-3} \tag{3.70}$$

$$p_{bbww} = \frac{n_b}{n}\frac{n_b - 1}{n-1}\frac{n_w}{n-2}\frac{n_w - 1}{n-3} \tag{3.71}$$

And S_1, S_2 and S_3 are defined in equations (3.57), (3.58) and (3.59) .Studies including SAC measures for categorical data, as usual, joint-count statistics, are not frequent. Nonetheless, some studies use joint-count statistics and relevant hypotheses. Epidemic characteristic of Dengue fever and Dengue hemorrhagic fever in Sukhothai province, Thailand [307] and economic development level of provinces of Poland [308].

3.7 Measures For Local Spatial Autocorrelation

In SAC, it is another issue whether a chosen spatial unit statistically significantly differentiates from the units in its neighborhood or not. Because global SAC measures do not give any information about a certain unit, they reflect overall spatial structure. To that end, some statistics and hypotheses are developed. These are measures obtained from certain global autocorrelation measures. Furthermore global SAC can be decomposed into local SAC, in some measures [19][22]. Some study on local SAC considers permutational assumptions, like total, conditional and

Local SAC is used in many fields of literature. Rosenberg uses Local SAC to explore epidemiological maps of cancer in western and southern Europe. It is one of the evident results in the study that Denmark has high overall homogeneous cancer mortality, whereas Southern Italy has low overall homogeneous cancer mortality [309]. Junior uses Local SAC to discriminate tissues in mammograms as benign or malignant [283]. Tiefelsdorf investigates bladder cancer incidences in 219 counties of the former German Democratic Republic and the migration problem in spatial epidemiology by using global and local SAC and spatial clustering methods [310].

Due to the fact that moments of Local Moran and Local Geary's tend to violate normal distribution assumptions, relevant hypothesis tests based on p-value cannot be carried out like global SAC measures [19][89][158]. As a result of that case, examined spatial unit may be randomly significantly different than its neighbors. The hypothesis test may not give a statistically reliable result. The permutational method is used for the purpose of solution of this problem [89][158]. The permutational method is based on the rearrangement of values of spatial units in the neighborhood of chosen spatial team by using permutation. Each rearrangement generates a p-value. The numbers of p-values generated by permutation are compared with the p-value generated by actual data. Thus pseudo p-value is computed as

$$p = \frac{L+1}{M+1} \tag{3.72}$$

where L number of p-values generated by permutation which is equal or less than p-value generated by the actual value and M number of all possible permutations [311][312]. The pseudo p-value is a helpful tool for avoiding violation of the normal distribution assumption problem mentioned above. Because if pseudo p-value for a spatial unit indicates an extreme situation, then the spatial unit is considered a genuinely local extreme. But many studies, The permutational method is also used for global SAC measures as an alternative method [311].

3.7.1 Local Moran's I

Local Moran's I, or "Local Indicators of Spatial Association"-LISA [19][38] as an original name, is the most known local SAC measure in the literature. That is why Local Moran's I is a local projection of Global Moran's I. Local Moran's I is conceived by Belgian geographical economist Luc Anselin [19] inspired by Global Moran's I invited by Australian statistician Moran [6][7]. Local Moran's I determines whether clusterings are formed among contagious spatial units , or not [313]. In addition to that property, determination of outliers in the study area is another availability of local Moran's I. In doing so, Local Moran's I make the maps more informative and salient in terms of the statistically significance of each unit [50]. Local Moran's I is seen as a measure for deviation from spatial randomness [143], or for similarity or difference from its neighbors of a particular unit [314]. Local Moran's I can also be used to detect the center of statistically significant clustering [75].

Local Moran's I, as is evident from its name, is a measure based on global Moran's I. Local Moran's I For ith unit is defined as

$$I_i = z_i \sum_{i=1}^{n} w_{ij} z_j \qquad where \qquad z_i = \frac{(x_i - \bar{x})}{s} \tag{3.73}$$

The expected value of Local Moran's I is

$$E(I_i) = \frac{-w_i}{n-1} \qquad where \qquad w_i = \sum_{j=1}^{n} w_{ij} \qquad j \neq i \tag{3.74}$$

where

$$w_i = \sum_{j=1}^{n} w_{ij} \qquad j \neq i \tag{3.75}$$

The variance for Local Moran's I is

$$V(I_i) = \frac{w_{i(2)}(n-b_2)}{n-1} - \frac{2w_{i(kh)}(2b_2-n)}{(n-1)(n-2)} - \frac{w_i^2}{(n-1)^2} \tag{3.76}$$

where

$$w_{i(2)} = \sum_{j=1}^{n} w_{ij}^2 \qquad j \neq i \tag{3.77}$$

$$b_2 = \frac{n \sum_{i=1}^{n} z_i^4}{\left(\sum_{i=1}^{n} z_i^2\right)^2} \tag{3.78}$$

$$2w_{i(kh)} = \sum_{k=1}^{n} \sum_{h=1}^{n} (w_{ik} w_{ih}) \qquad k,h \neq i \tag{3.79}$$

In order to augment interpretation advantages, weighting matrices can be chosen as row standardized form, below mentioned

$$\sum_{j=1}^{n} w_{ij} = 1 \tag{3.80}$$

That form is to be discussed in the weighting section. However row standardized form is not requisite for studies including Local Moran's I [19][188]. In keeping with Global Moran's I, Anselin proposes [19]

$$w_{ii} = 0 \tag{3.81}$$

Local Moran's I is thought of as a decomposition [22] or disaggregation of Global Moran's I[110]. As technical advantages, Local Moran's I represent how local trends differentiate from the global trend in Moran scatter plot [50]. It should not be forgotten that the determination of hot and cold spots depends upon the prespecified level of significance [144]. Many studies employ Local Moran's as an almost unique tools of LISA [73][80][138] [60][305]. Moreover, Local Moran's I is called Local Moran Index [142]. Some study uses Local Moran without calling its name [142].

Local Moran's I is applied for spatial clustering of higher prices and lower prices in real estate markets [73], determination of hot ve cold spots of cancer incidences over Saudi Arabia by dealing with R^2 coefficient of determination as a value of spatial unit [80], risk analysis of fires in Serbia [138], regional decomposition of unemployment in Central Europe by using lognormal distribution [60], performance of farming activities and non-farming enterprise in rural Africa [56], voting rates general election on county-scale in Turkey 2011 [305], meteorological patterns of Iran [148], provincial unemployment rates in Turkey [63], organic agricultural operations in the USA with bivariate methods [65], tourism performance of Serbia [69], agricultural consolidation effects in Lubelskie Poland [134], voting participance level in French Presential Elections [142], provincial income inequality in Turkey [70].

3.7.2 Local Geary's c

In literature, the local version of Geary's c is also constructed [19]. Notably, local Geary's c is a measure based on global Geary's c for local scale. Local Geary's c indicates the degree of variability over spatial units [313].

Local Geary's c is defined as

$$c_i = \frac{1}{m_2} \sum_{j=1}^{n} w_{ij}(z_i - z_j)^2 \tag{3.82}$$

Expected value and Variance of c_i are [19][89]

$$E(c_i) = \frac{2nw_i}{n-1} \tag{3.83}$$

$$V(c_i) = \left(\frac{n}{n-1}\right)(w_i^2 + w_{i(2)})(3 + b_2) - E(c_i)^2 \tag{3.84}$$

and w_i, $w_{i(2)}$ and b_2 defined respectively in equations (3.75), (3.77) and (3.78)

By reason of that distribution of local Geary's c is undetermined , local Geary's c is hardly ever used in SAC literature. Thereby local Geary's c remains recondite for many researchers. However, Sokal and his co-authors use Local Geary's c based on permutational methods for a biological model with simulations [89] [87].

3.7.3 Getis-Ord Local Statistics

It should be first noted that Getis-Ord Statistics can be applied to point data by its invented approach. American Geographer Arthur Getis develop this statistics and British Computer Scientist J. Keith Ord [14] [12]. This is also coined as Getis-Ord Spatial Analysis in literature. The fundamental motivation for Getis-Ord Statistics is the question of whether the relative values are clustered over the same location or not. In other words, Getis-Ord statistics dilates on whether the tendency of clustering exits or not and if it exists, where it is [224]. Unlike Moran's I and Geary's c, Getis-Ord Statistics represents the locations of clusters that are classified into two classes: hotspots and coldspots, if they exist. Hot Spots and Cold Spots are two concepts given by Getis-Ord G(d) Statistics. "Hot Spots" is defined as a cluster with high values, whereas "Cold Spots" is defined as a cluster with low values.

The Spatial Weight firstly used in Getis and Ord is defined below [14]

$$w(d)_{ij} = \begin{cases} 1 & |s_i - s_j| \leq d \\ 0 & otherwise \end{cases} \tag{3.85}$$

Where s_i is the coordinate of i.th point. $|.|$ is the standard absolute value, in other words standard norm of R-real numbers. It is another important fact that this spatial weight depends upon d-distance [315]. d-distance can be considered as a threshold value, like in natural sciences and neural network studies.

Getis-Ord's $G_i(d)^*$ is defined as

$$G_i(d)^* = \frac{\sum_{j=1}^{n} w(d)_{ij} x_j}{\sum_{j=1}^{n} x_j} \tag{3.86}$$

For this form, it must be noted that [316]

$$w_{ij} = 0 \qquad where \qquad i = j \tag{3.87}$$

Then $G_i(d)^*$ is represented as $G_i(d)$, without asterisk. Differently from 3.87, if spatial weight is defined as

$$w_{ij} = d_{ij}^{-\alpha} \qquad where \qquad d_{ij} = |s_i - s_j| \quad and \quad \alpha > 0 \tag{3.88}$$

Then in the case above it should be [183].

$$w_{ij} = 1 \qquad where \qquad i = j \tag{3.89}$$

Herein one lays emphasis on that Getis and Ord firstly propose values of w_{ij} as 0 or 1, that is to say, binary codes [14]. As it stands, Getis-Ord statistics can be used for structures with contiguity property.Moreover in most geographical studies α is tend to be equal to 2 [188][224].

The approaches to spatial weight are investigated in the following chapters. Regardless of the definition of spatial weight, Getis-Ord statistics for Local SAC are defined as

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij} x_j - \bar{x} \sum_{j=1}^{n} w_{ij}}{\sqrt{SF}} \tag{3.90}$$

where

$$S = \frac{\sum_{j=1}^{n} x_j^2}{n} - \bar{x}^2 \quad and \quad F = \frac{n \sum_{j=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij}\right)^2}{n-1} \tag{3.91}$$

The Getis-Ord Local Statistics G_i^* is assumed to have a standard normal distribution. Thus hypothesis test is applied in keeping with standard normal distribution. Whereas high and positive values show hotspots, low and negative values signify coldspots.

In literature, Getis-Ord local statistics are used in Hotspot Analysis. In order to detect hotspots, Getis-Ord Local statistics is used in spatial risk analysis of forest fire in Serbia 2000-2013 [138], in analysis of regional linguistic differentiations [151], in analysis of general election 2011 results of Turkey in provincial and county-based scale [305], in urban crime analysis [315], in spatial descriptive studies for transportation in Italy [144], in climatological studies in Iran [148] in modelling of forest growth model [127] and in spatial analysis for epidemic animal disease in China [100].

3.8 Statistical Evaluation of Spatial Autocorrelation

3.8.1 Problems of Classical Hypothesis Tests

With tests of a classical hypothesis based on standard distributions, statistical evaluations of spatial autocorrelation have some problems. A prominent example of these problems is that parameters of the assumed distribution, like mean and variance, are in only approximate and asymptotic form [19]. This problem manifests itself in local SAC rather than global SAC. But different from other local SAC measures, it should be noted that the distributions of G_i and G_i^* are compatible with standard distributions [19]. Particularly small data sets tend to violate the assumed distribution due to spatial (or also temporal) autocorrelation [317].

An approach based on randomization is proposed in order to overcome these problems in SAC literature. Furthermore, SAC studies and software adopt the approach as a standard statistical evaluation. In this doctoral dissertation, classical methods based on assumed distributions and randomization are used and compared to each other to understand global and local SAC.

3.8.2 Permutational Distribution (Randomization)

Randomization is a method fundamentally based on the random relocation of observed values over spatial units employing permutation [19][317] [318]. Observed values and contiguity of spatial units (W in mathematical modeling) are invariant under randomization. The invariance of observed values can be considered to be an assumption of randomization. However, in comparison to classical hypothesis tests based on standard distributions, randomization has less restrictive assumptions and is conceptually more convenient for spatial data.

Randomization is applied in many ways as required by the aim of the study. Two of them, total and conditional randomization, are the most frequently used methods. Apart from these randomization types, restricted within the region and sequential randomizations are used . Complete randomization is the random relocation of “all” observed values over spatial units and is used for global SAC measures [317]. Complete randomization is also called “total randomization” [87][89] or just

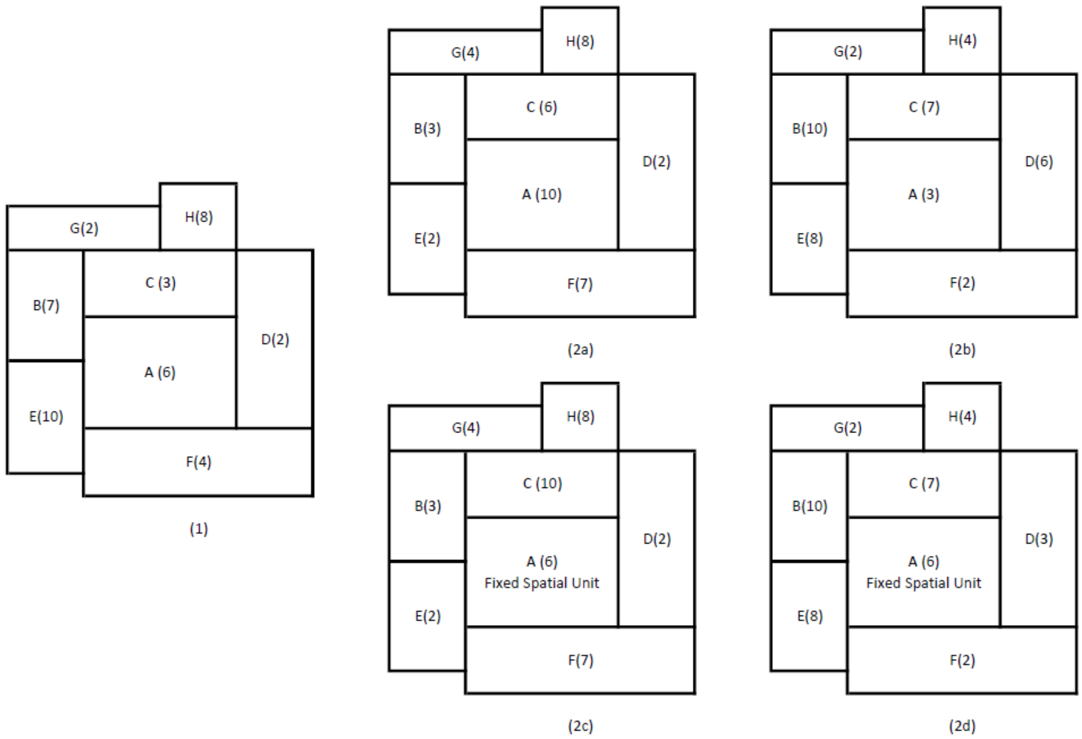

Figure 3.1 Randomizations of a hypothetical contiguity-based map , (1) Original Observations, (2a-2b) instances of Total Randomization, (2c-2d) instances of Conditional Randomization

"randomization" [318].

Conditional randomization is the random relocation of "all but one" observed values. That observed value, except for the relocation process, is the observed value of the focused spatial unit. In this context, the condition is holding that value fixed. Conditional randomization is used for local SAC measures [87][89] [19].

Illustrations of randomizations based on contiguity and distance are given the Figures 3.1 and 3.2, respectively.

Random relocation of observed values and randomization created a parameter, like Global and Local SAC measures each time. The set of these parameters generated by randomization yields a distribution. This distribution is called "permutation distribution" [319] [320] and is used as "empirical reference distribution" [19] or "reference distribution" [317], [321] [89], [101]. Statistical evaluation of actual parameters (like global and local SAC measures) through the distribution is called the "randomization test" [111],[102] and "Randomization Hypothesis" [87][89], [18][19], [51].

Statistical evaluation of the actual parameter through randomization is directly related to the position of the actual parameter in the reference distribution. The closer the position is to extreme points (minimum and maximum) of the distribution, the more significant the parameter is. The distribution of (total and conditional) randomization used for the detection of (global and local) spatial autocorrelation tends to violate the assumption of standard statistical distributions. Pseudo p-value is an indicator to help determine statistical significance. As a side note, the empirical reference distribution for Global Moran's I is observed to converge with the normal distribution during the study of the dissertation.

Lastly, the combinational method can be used for the randomization of data with contiguity instead of permutational method. Because in the contiguity matrix defined in the equations 3.2 and 3.3, all weightings in a row is equal, then same n_i elements with different order generate the same value. However, this fact is not valid for distance based spatial weighting matrix because weightings in a row may not be equal.

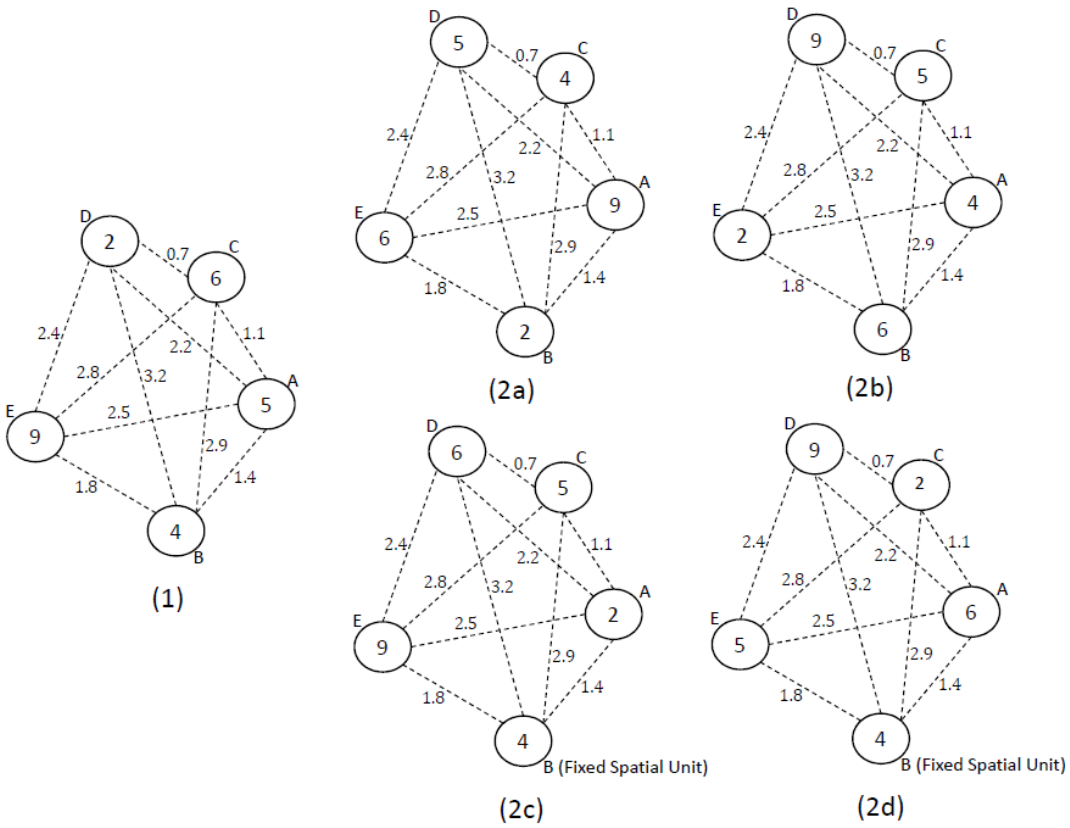

Figure 3.2 Randomizations of a hypothetical distance-based map, (1) Original Observations, (2a-2b) instances of Total Randomization, (2c-2d) instances of Conditional Randomization

3.8.3 Pseudo P-Value

One should note that the distribution of hypothetical values generated by the randomization process is not necessary to be in the form of standard statistical distribution. Global SAC measures tend to form a Normal distribution, whereas Local SAC measures tend to violate standard statistical distributions. As a consequence, hypothesis tests may not be suitable for the statistical evaluation of SAC measures, particularly in local SAC. Statistical evaluation of actual parameters by means of randomization is developed by force of finding a solution to this problem.

Randomization process is essential for understanding distribution and its general tendency (expected value, or “mean” in a more practical context) of hypothetical values. Statistical evaluation is based on the position in this reference distribution generated by randomization. From a computational perspective, each randomization process generates a unique distribution due to different initial values for the permutation. Nevertheless, these distributions are identical if permutations are selected independently and in an unbiased way. Then the position of the actual parameter does not change such that a statistical significance occurs. The pseudo p-value indicates how the actual parameter is close to the extreme points (in other words, two ends) of reference distribution. The pseudo p-value is computed as

$$Pseudo-p = \frac{r+1}{m+1} \tag{3.92}$$

r is a number of values being equal and more extreme values of the reference distribution than the actual parameter. m is a number of elements of the reference distribution. The term “+1” in the numerator and the denominator of the fraction in 3.92 express the fact that the actual parameter is added to the reference distribution [322].

Even though no statistical significance occurs, the pseudo p-value is not invariant under each randomization. But pseudo p-values for the same actual parameters are close to each other. Then pseudo p values should be evaluated as a level rather than a point. Furthermore, if the size of the distribution is increased, the closeness decreases. The sensitivity of pseudo p-value is an important question, particularly in local SAC measures. In case pseudo p-value is in the form 10^{-n} this shows the insufficiency of the size of the distribution.

The introductory study [19] and first [87][89] studies of Local SAC use pseudo p-value for statistical evaluation. Rey employs pseudo p for understanding the spatial mechansim of the USA at the state level [323]. Pseudo p is also used in multivariate spatial association [321]. Induction studies for global and local SAC give place to pseudo p values for statistical evaluation [101].

3.8.4 False Discovery Rate (FDR)

False Discovery Rate (FDR), is based on an application of pseudo p-value on spatial units particular local SAC measures. FDR is also called as FDR correction [121][69] [324]. FDR is computed for each spatial unit

$$FDR_{r_k} = \frac{k * \alpha}{n} \tag{3.93}$$

Where r_k is index of the element with $k-th$ rank based on pseudo-p values of spatial units, α is the confidence level. If FDR_{r_k} value is less than pseudo p-value of spatial unit r_k, then the spatial unit i is statistically significant, providing FDR condition. In literature, FDR is applied to ecological [121], touristic [69] and theoretical (with principal component analysis (PCA))[321] studies.

3.8.5 Bonferroni Bound

As is known, Bonferroni Bound (or correction) is a frequently used approach for statistical evaluation of (usually pairwise) comparisons. In a spatial context, Bonferroni bound is based on spatial units. Each spatial unit is compared to (usually) the mean of its neighbourhood set through distribution generated by the randomization process [19]. Therefore count of comparisons is the count of spatial units. Bonferroni Bound is determined by this count and selected p-value. Bonferroni Bound is simply computed as

$$\frac{\alpha}{n} \tag{3.94}$$

Where n is a count of spatial units. Bonferroni bound is called in many ways like (mostly) “Bonferroni correction” [102] [181][228][27][96] and “Bonferroni limit”[325]. In literature, Bonferonni Bound is applied to many fields like theoretical [321] biological [87][102] , ecological [181][104], and genetics [94] to avoid type 1 error.

4
Method

4.1 Motivation

The spatial effect for contiguity based studies is examined with classical methods in the literature. These methods are limited in two ways: the representation of the neighbourhood set as the mean and the domination of central tendency.

The modelling of spatial effect is a central issue in SAC literature. The Spatial Effect is the mean of the neighbourhood set in many contiguity-based studies. As is known, the mean may not be a good representative of a set in cases the outliers exist, or the distribution of data violates the assumption of the normal distribution. The smaller the size of samples (the neighbourhood set in this context) is, the less the mean represents the set. However, in literature, the median is employed as the representative value of the neighbourhood set [326]. Moreover, GeoDa also uses the median for the same purpose as "Median Local Moran" [327][328].

Another issue is the domination of the central tendency to represent the neighbourhood set. The spatial effect of the unit is conceived as the central tendency (mean and median) of the neighbourhood set in the literature of contiguity-based studies. However, the mean and median may not reflect the spatial effect on the unit. As a striking example for this case, Mexico and Canada are the neighbours of the USA, having $63,123, and have $8,326 and $43,560 per capita GDP 2020, respectively [329]. The mean and median of the neighbourhood set of (or the spatial lag of) the USA is $25,943. But there is no a country of that per capita GDP 2020 is near $25,943 in the neighborhood set. Then the spatial lag, the mean or median of the neighbourhood set in this context, is insufficient for modelling spatial effect.

Furthermore, the spatial relation between the minimum of the neighbourhood set (Mexico) and the spatial unit (USA) creates an illegal immigration problem, which is different from the one other developed countries face in the USA. This example teaches that the spatial effect should not be restricted to the mean or median of the neighbourhood set in all contiguity-based studies. The minimum and maximum can be used as particular spatial lags for the examination of the spatial effect.

The usage of the minimum, median and maximum of the neighbourhood set is extended to the quartiles of the set. Furthermore, quantiles, in a general form, can be employed for the same purpose. Quantiles are essential because they standardize the sets regardless of their sizes. The question is whether the quantiles are embedded in an interval as continuous space. A continuous index can reach all quantiles of the set in that way. Quantiles are described as the section of the neighbourhood set. Thereby, the distributions of sets, "neighbourhood sets" in this context, with different sizes are reflected in SAC models.

This dissertation focuses on two issues and one improvable area mentioned above. The conditional approach indicates the section of the continuous form, interval, of the neighbourhood set. In this context, the sections are equivalent to a generalized form of the quantiles.

4.2 Bounded Continuous Strict Index (BCSI)

Standardization of the sets with different sizes is important for the comparison of them. Mean and Quantiles are the known measures based on value. Moreover, the standardization can be extended to the whole set level. To realize this standardization, the closed intervals are used. Because all intervals are homeomorphic to each other. Thereby all discrete set is to be matched to a convenient interval, the convenient interval is matched to the interval $[0,1]$.

Definition 4.1. Continuous Form Let $X = \{x_1, x_2, x_3, \ldots, x_n\}$ be a discrete set with n elements. And $i = 1, 2, 3 \ldots, n$ original discrete (countable) index of X. Then the continuous form of the discrete set X is defined as

$$X^c := [min(X), max(X)] \tag{4.1}$$

In case $min(X) = max(X)$, then $X^c := [min(X), max(X)] = \{min(X)\}$

Continuous form of X is the interval of which the ends are minimum and maximum of X. This approach makes all discrete sets considered in the convenient convenient intervals, as a concept. And all closed intervals is continuously mapped by the interval $[0,1]$.

Definition 4.2. Bounded Continuous Index (BCI) Let be B_X a function such that

$$\begin{aligned} B_X : [0,1] &\rightarrow X^c \\ \theta &\rightarrow B_X(\theta) \end{aligned}$$

and B_X satisfies three conditions below:

1. B_X is continuous
2. $B_X(0) = min(X)$ and $B_X(1) = max(X)$ (Boundary Condition)
3. $B'_{X+}(\delta) \geq 0$ for $\delta \in [0,1)$ and $B'_{X-}(\delta) \geq 0$ for $\delta \in (0,1]$ (Non-negative Slope)

B'_{X+} and B'_{X-} are right and left derivatives of B_X. Then B_X is called as Bounded Continuous Index (BCI) for the set X

A coarse linear interpolation between minimum and maximum of X can be given as

$$B_X(\theta) = min(X) + \theta(max(X) - min(X)) \tag{4.2}$$

Furthermore, a BCI including median of X ($med(X)$) can be constructed. Then it should be $B_X(0.5) := med(X)$ because $B_X(\theta)$ corresponds to quantiles of X. A parabolic function can be used for $B_X(\theta)$ such that

$$B_X(\theta) = a\theta^2 + b\theta + min(X) \tag{4.3}$$

where $a = 2min(X) - 4med(X) + 2max(X)$ and $b = -3min(X) + 4med(X) - max(X)$. The model defined in 4.3 satisfies all conditions of BCI if and only if

$$3min(X) + max(X) \leq 4med(X) \leq min(X) + 3max(X) \tag{4.4}$$

Otherwise the model in 4.3 does not satisfy third condition of BCI (Non-negative Slope). a and b can be written respectively in terms of ranges such that

$$a = 2(d_2 - d_1) \qquad b = 3d_1 - d_2 \tag{4.5}$$

where $d_1 = med(X) - min(X)$ and $d_2 = max(X) - med(X)$. Likewise the condition in 4.4 can be written in terms of ranges such that

$$\frac{d_2}{3} \leq d_1 \leq 3d_2 \tag{4.6}$$

Based on this example 4.3 and its condition 4.4, pure interpolation is not sufficient for the construction of BCI. Behaviour of non-linear model may violates 3rd condition of

BCI. Then a constrain must be created, as in 4.4, to overcome this violation problem. Instead of one-piece function as in 4.3, a piecewise function is preferable because it is more flexible and more manageable than one-piece function.

BCI is a coarse approach for the modelling of distribution of the set. Because the condition of the BCI is based on the minimum and maximum as boundary conditions. BCI functions of the different set with same minimum and maximum may be same (in other saying, equal). A specific BCI can be constructed in such that it characteristically represents the distribution of the set by employing and generalizing quantile method. Bounded Continuous Strict Index is an answer of this problem. In literature, some studies use this approach for the practical purposes in the different fields and contexts [330][331].

Definition 4.3. Bounded Continuous Strict Index (BCSI) Let be b_X a BCI and X can be written as $X = \{x_{r^1}, x_{r^2}, x_{r^3}, \ldots, x_{r^n}\}$ where x_{r^k} index of the element with k-th rank. Then $min(X) = x_{r^1} \leq x_{r^2} \leq x_{r^3} \leq \cdots \leq x_{r^n} = max(X)$ and b_X satisfies the condition below

$$b_X(\theta_k) = x_{r^k} \qquad where \quad \theta_k := \frac{k-1}{n-1}, \quad k = 1, 2, 3, \ldots, n \tag{4.7}$$

b_X is strictly dependent upon the distribution of X. Image of θ in X^c can be represented as continuous index r_θ by means of X such that

$$x^c_{r^\theta} := b_X(\theta_k) \qquad where \quad r^\theta := (n-1)\theta + 1 \tag{4.8}$$

By the condition of BCSI 4.7 we get

$$x^c_{r^{\theta_k}} := b_X(\theta_k) = x_{r^k} \tag{4.9}$$

BCSI binds continuous index $x^c_{r^{\theta_k}}$ in X^c with x_{r^k} in X. Indexing can be made by both $[0,1]$ (via θ) and $[0,n]$ (via r_θ). BCSI of X gives this schematic below

$$\begin{gathered}[0,1] \rightarrow [1,n] \rightarrow X^c \\ \theta \rightarrow r_\theta \rightarrow x^c_{r_\theta}\end{gathered}$$

However, these remarks are directly relevant to the results of the condition of BCSI and an convenient model should be created for determination of $x^c_{r_\theta}$ where $\theta \neq \theta_k$ or the part $X^c \setminus X$. In this context, BCSI should be in the general form below

$$x^c_{r_\theta} := b_X(\theta) = \begin{cases} x_{r^{k_1}}, & \theta = \theta_{k_1} \\ f_X(\theta), & \theta_{k_2} < \theta < \theta_{k_2+1} \end{cases} \tag{4.10}$$

where $k_1 = 1,2,3,\dots,n$ and $k_2 = 1,2,3,\dots,n-1$. In an exceptional case $n = 1$, BCSI is defined , consistently with BCI, as

$$x^c_{r^\theta} := b_X(\theta) = x_{r^1} = x_1 \tag{4.11}$$

The choice of $f_X(\theta)$ gets more important, as the sample size decreases. Piecewise-linear approach is both more interpretable and easier computable than others. Piecewise linear BCSI is contructed below

$$b^l_X(\theta) = \begin{cases} x_{r^{k_1}}, & \theta = \theta_{k_1} \\ (1-r^*_\theta)x_{r^{k_2}} + r^*_\theta x_{r^{k_2+1}}, & \theta_{k_2} < \theta < \theta_{k_2+1} \end{cases} \tag{4.12}$$

where $r^*_\theta = r_\theta - k_2$, $k_1 = 1,2,3,\dots,n$ and $k_2 = 1,2,3,\dots,n-1$. To illustrate piecewise linear BCSI, let X have 7 elements then $x^c_{r^{0.6}} = b^l_X(0.6) = x^c_{4.6} = 0.4x_{r^4} + 0.6x_{r^5}$.

Piecewise-linear $b^l_X(\theta)$ is the linear interpolation between closest ranks (r^{k_2}, r^{k_2+1}) method is chosen for the open intervals $(x_{r^{k_2}}, x_{r^{k_2+1}})$, then BCSI satisfies

$$b^l_X(\frac{s}{q}) = \text{the s/q-quantile}, s = 0,1,2,3,\dots,q \tag{4.13}$$

From the equation 4.13, quantiles include maximum and minimum of the set X. And Piecewise-linear BCSI $b^l_X(\theta)$ is equivalent to the best-known and the most-used form of *Quantile Function*. θ plays a role to indicate a quantile. In this context, $0-th$ and $1-th$ quantiles indicate minimum and maximum of X, respectively. This approach is used in the statistics and machine learning literature [332][333][334].

Thereby all numerical data sets with one-dimension can be continuously indexed with the closed interval $[0,1]$ thanks to BCSI regarless of their sizes. Moreover BCSI ensures to model the distribution of the set. Because of these facts, BCSI is applicable to neighborhood sets having different sizes. A specific spatial lag given by BCSI applied to the neighborhood set can help to understand spatial effect.

4.3 Spatial Theta Lag

In this section, the application of a particular BCI to neighborhood sets having different sizes is examined. So the focus of the dissertation is a family of sets X_i indexed by i, instead of X.

Definition 4.4. Spatial Theta Lag Let X be a set comprised of n spatial units, W be the contiguity matrix of X, $X_i := \{x_j | w_{ij} \neq 0\}$ be the the set of observations from the neighborhood set of the spatial unit indexed by i. Then Spatial Theta Lag based on

BCI is defined as

$$L_i^\theta := B_{X_i}(\theta) = \sum_{j=1}^{n} f_{ij}(\theta) w_{ij} x_j \tag{4.14}$$

Where $f_{ij}(\theta)$ are appropriate functions not violating the conditions of BCI.

Spatial Theta Lag defined in 4.14 is written in terms of a specific contiguity matrix and data.

Definition 4.5. Conditional Contiguity Matrix A specific contiguity matrix is written for L_i^θ is Spatial Theta Lag of X_i such that

$$L_i^\theta = \sum_{j=1}^{n} f_{ij}(\theta) w_{ij} x_j = \sum_{j=1}^{n} w_{ij}^\theta x_j \tag{4.15}$$

Where

$$w_{ij}^\theta := f_{ij}(\theta) w_{ij} \tag{4.16}$$

then $W^\theta := \{w_{ij}^\theta\}$ is called as Conditional Contiguity Matrix. And in this context $W := \{w_{ij}\}$ is called as Base Contiguity Matrix

Contiguity matrices are independent from "data" in classical spatial autocorrelation literature. Contrary to this approach, conditional contiguity matrices are strictly dependent to "data".

X_i can be represented as $X_i := \{x_j | w_{ij} \neq 0\} = \{x_{s_1}, x_{s_2}, x_{s_3}, \ldots, x_{s_{n_i}}\}$ where $j = s_1, s_2, s_3, \ldots, s_{n_i}$ are original indices of units in neighborhood set of the spatial unit indexed by i. And n_i is the size of X_i. Moreover the element of X_i can be labeled by i and rank of the unit in $X_i = \{x_{r_i^1}, x_{r_i^2}, x_{r_i^3}, \ldots, x_{r_i^{n_i}}\}$ such that

$$x_{r_i^1} \leq x_{r_i^2} \leq x_{r_i^3} \leq \cdots \leq x_{r_i^{n_i}} \tag{4.17}$$

where r_i^k is the original index of spatial unit with $k-th$ rank in X_i. Then one concludes that

$$x_{r_i^1} = min(X_i) \qquad\qquad x_{r_i^{n_i}} = max(X_i) \tag{4.18}$$

To reflect the boundary condition of BCI, $f_{ij}(0)$ and $f_{ij}(1)$ should be chosen as

$$f_{ij}(0) = \begin{cases} 1, & j = r_i^1 \\ 0, & otherwise \end{cases} \qquad\qquad f_{ij}(1) = \begin{cases} 1, & j = r_i^{n_i} \\ 0, & otherwise \end{cases} \tag{4.19}$$

then the conditional weightings w_{ij}^0 and w_{ij}^1 are computed as

$$w_{ij}^0 := f_{ij}(0)w_{ij} = \begin{cases} 1, & j = r_i^1 \\ 0, & otherwise \end{cases} \tag{4.20}$$

$$w_{ij}^1 := f_{ij}(1)w_{ij} = \begin{cases} 1, & j = r_i^{n_i} \\ 0, & otherwise \end{cases} \tag{4.21}$$

w_{ij}^0 and w_{ij}^1 denote *Conditional Contiguity for Minimum* and *Conditional Contiguity for Maximum*, respectively. Spatial Theta Lag may be examined with the tools from classical Spatial Autocorrelation. Then this analysis is called as Conditional Spatial Autocorrelation because Spatial Theta Lag is based on the condition of θ. Furthermore, conditional contiguity matrix given by θ may be used in classical tool of SAC. As a particular case of SAC measure, Global Conditional Moran's I with conditional contiguity matrix may be defined as

$$I^\theta := C\frac{\sum_{i=1}^n x_i L_i^\theta}{\sum_{i=1}^n x_i^2} = C\frac{\sum_{i=1}^n x_i B_{X_i}(\theta)}{\sum_{i=1}^n x_i^2} = C\frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}^\theta x_i x_j}{\sum_{i=1}^n x_i^2} \tag{4.22}$$

where

$$C := \frac{n}{\sum_{i,j} w^\theta}$$

In a similar way, Conditional Local Moran's I is defined as

$$I_i^\theta := x_i L_i^\theta = x_i B_{X_i}(\theta) = x_i \sum_{j=1}^n w_{ij}^\theta x_j \tag{4.23}$$

Finally one can find some traces for median based spatial autocorrelation. The concept "median smoother" proposed by Wall and Devine is weakly related to the median of the neighborhood set in this context [326]. Haining uses and discusses the median smoother for measuring spatial effect [51]. Anselin defines median based spatial autocorrelation by using Local Moran's and calls its as Median Local Moran "I_i^M" [328]. Median Local Moran is written in the form of Piecewise Linear BCSI (a specific form of BCI) such that

$$I_i^M = I_i^{0.5} = x_i L_i^{0.5} = x_i b_{X_i}^l(0.5) = x_i \sum_{j=1}^n w_{ij}^{0.5} x_j \tag{4.24}$$

$w_{ij}^{0.5}$ is called as *Conditional Contiguity for Median*. As an extension of Anselin's conceptualization, I_i^0 and I_i^1 can be called as "Minimum Local Moran" and "Maximum

Local Moran", respectively.

4.4 Spatial Theta Lag Based on Piecewise Linear BCSI

As a particular case, Piecewise Linear BCSI $b^l_{X_i}(\theta)$, a specific form of BCI $B_{X_i}(\theta)$, can be chosen for Spatial Theta Lag L^θ_i. Thereby L^θ_i generated by $b^l_{X_i}(\theta)$ can reflect the quantiles of X_i to the Spatial Autocorrelation and Regression Models. Conditonal contiguity matrices $w^\theta_{i,j}$ can be determined thanks to the open form of $b^l_X(\theta)$ in definition 4.12 such that

$$w^\theta_{ij} := \begin{cases} 1 - \{r_{i,\theta}\}, & j = r_i^{k^\theta} \\ \{r_{i,\theta}\}, & j = r_i^{k^\theta+1} \\ 0, & otherwise \end{cases} \tag{4.25}$$

where $r_{i,\theta} := (n_i - 1)\theta + 1$, $k^\theta := \lfloor r_{i,\theta} \rfloor$, $\{\}$ is fractional part function and $\lfloor\rfloor$ is floor function. By the definitions of $\{\}$ and $\lfloor\rfloor$, one gets

$$r_{i,\theta} = k^\theta + \{r_{i,\theta}\} \tag{4.26}$$

In in compliance with the definition of w^θ_{ij} in 4.25, L^θ_i generated by $b^l_{X_i}(\theta)$ is written as

$$L^\theta_i = b^l_{X_i}(\theta) = \sum_{j=1}^{n} w^\theta_{ij} x_j \tag{4.27}$$

w^θ_{ij} for Anselin's *Median Local Moran* I^M_i in equation 4.24 can be shown in exact form (without fractional part function $\{\}$ and floor function $\lfloor\rfloor$) by means of Median property and Conditional Contiguity for Spatial Theta Lag based on BCSI $b^l_{X_i}(\theta)$ in 4.25 such as

$$w^{0.5}_{ij} = \begin{cases} 1, & j = r_i^{p+1} \; and \; n_i = 2p+1 \\ 0.5, & j = r_i^{p} \; and \; n_i = 2p \\ 0.5, & j = r_i^{p+1} \; and \; n_i = 2p \\ 0, & otherwise \end{cases} \tag{4.28}$$

where p is a natural number.

4.5 An Inference On the neighbors of Minimum and Maximum

When one uses the minimum (or maximum) of the neighborhood set as a specific spatial lag, one can make an estimation for pseudo p values of Local Moran of them (Minimum and Maximum Local Moran's) with some conditions. Because the distribution of Local Moran's directly depends on the distribution of the spatial lag as seen in 3.73 and 4.23. This fact allows to make an inference for Minimum and Maximum Local Moran's. If a spatial unit (with k index) satisfies these conditions

1. $x_k \neq 0$
2. $min(X - \{x_k\}) \in X_k$ for I_k^0 or $max(X - \{x_k\}) \in X_k$ for I_k^1
3. $n > n_k + s$

then

$$R = \frac{(n-1)!}{(n-1-n_k)!} \qquad N = \frac{(n-1-s)!}{(n-1-s-n_k)!} \tag{4.29}$$

where R is the number of all permutations for the neighborhood set with n_k elements of the spatial unit from $X - \{x_k\}$, N for I_k^0 is the number of the permutations which creates instances of I_k^0 greater (or less) than the observed I_k^0 for positive (or negative) Minimum Local Moran. Likewise, N for I_k^1 is the number of the permutations which creates instances of I_k^1 less (or greater) than the observed I_k^1 for positive (or negative) Maximum Local Moran, s is defined as

$$s := \#\{j|x_j = min(X - \{x_k\})\} \text{ for } I_k^0 \; or \; s := \#\{j|x_j = max(X - \{x_k\})\} \text{ for } I_k^1 \tag{4.30}$$

where # denotes the count of the elements of the set. One can conclude that

$$\text{Pseudo p value for } I_k^0 \text{ or } I_k^1 = 1 - \frac{N}{R} = 1 - \prod_{v=1}^{s} \frac{n - n_k - v}{n - v} \tag{4.31}$$

Because

$$min(X - \{x_k\}) \in X_k \iff L_k^0 = min(X - \{x_k\}) \tag{4.32}$$

or

$$max(X - \{x_k\}) \in X_k \iff L_k^1 = max(X - \{x_k\}) \tag{4.33}$$

No permutation can generate an instance (a spatial theta lag) less than the observed L_k^0. So if one extracts the rate, "N/R" ,of the number of the permutation generating instances equal to the observed L_k^0, "N", to all possible permutations "R" from 1, one gets pseudo-p value for $I_k^0 := x_k L_k^0$ by the relation 4.33. One makes similar inference for pseudo-p value for $I_k^1 := x_k L_k^1$ by the relation 4.33. Furthermore, minimum and/or maximum are not repetitive in many cases.Then $s = 1$ and equation 4.31 becomes

$$\text{Pseudo p value for } I_k^0 \text{ or } I_k^1 = \frac{n_k}{n-1} \tag{4.34}$$

As an side note, Because the spatial units are based on contiguity, one uses the combinational method instead of the permutational method then one gets same result such that

$$R^* = \frac{(n-1)!}{(n-1-n_k)!n_k!} \qquad N^* = \frac{(n-1-s)!}{(n-1-s-n_k)!n_k!} \tag{4.35}$$

then

$$\frac{N^*}{R^*} = \frac{\frac{(n-1-s)!}{(n-1-s-n_k)!n_k!}}{\frac{(n-1)!}{(n-1-n_k)!n_k!}} = \frac{\frac{(n-1-s)!}{(n-1-s-n_k)!}}{\frac{(n-1)!}{(n-1-n_k)!}} = \frac{N}{R} \tag{4.36}$$

As is seen in 4.36 combinational approach is equivalent to permutational approach for pseudo-p value computation for I_k^0 and I_k^1.

5
Application

5.1 Data and Descriptions

The method is based on contiguity relation among the spatial units. So one should choose lattice or polygon data, like countries, provinces, precincts. Provinces of Turkey have an appropriate structure and is used for the method. Turkey has 81 provinces and 198 boundaries. Provinces are coded by *Vehicle Registration Numbers*, the numbers widely used for symbolized the provinces.

Table 5.1 The count of Provinces with n_i neighbors

n_i	Counts	n_i	Counts	n_i	Counts
1	1	4	16	7	5
2	5	5	22	8	4
3	11	6	14	9	3

As to the obsevarations, 136 datasets, from "*TUIK*" based on the provinces of Turkey are scrutinized to show the effect of "*Spatial Theta Lag*" strikingly [335]. Two of them, "*Unhappiness Rate of People Aged 25-34 in 2013*" and "*Voting Turnout Rate in 2015*", are selected. Unhappiness Rate of People Aged 25-34 in 2013 have the largest gap between Conditional Global Moran's I^0 and I^1 and Voting Turnout Rate in 2015 is the data in which the greatest number of significant provinces (in sense of Conditional Local Moran) occur. Basic descriptions of these data are given with provinces codes in table 5.2. From the table 5.2, $x_i = median(x) = 89.1$ satisfies where x=*Voting Turnout Rate in 2015* and $i = 31, 55, 71$. To ease interpretation, these data are transformed into $z - score$ (or *Standard Score*) such that

$$z_i = \frac{x_i - \overline{x}}{\sigma_x} \tag{5.1}$$

where $\overline{x}$ and σ_x are the mean and the standard deviation of X. Standard Score is used for the application. The observations are clustered with Jenks Natural Breaks Optimization such that each cluster has minimum variance and maximum distance among the centers of the clusters. Optimal number of the clusters for the

Table 5.2 Descriptive Statistics of Data

Descriptive	Unhappiness Rate Ages 25-34	Voting Turnout
Min(i)	3.7(32)	77.1(4)
Median(i)	8.11(35)	89.1(31,55,71)
Mean(i)	9.12	88.16
Max(i)	25.18(21)	93.1(45)
Standard Deviation	3.38	3.19

observations,Unhappiness Rate of People Aged 25-34 in 2013 and Voting Turnout Rate in 2015, are 5 and 6, respectively. Spatial distribution of the clustered observations in Standard Score (Z-score) are shown in Figures 5.1 and 5.2.

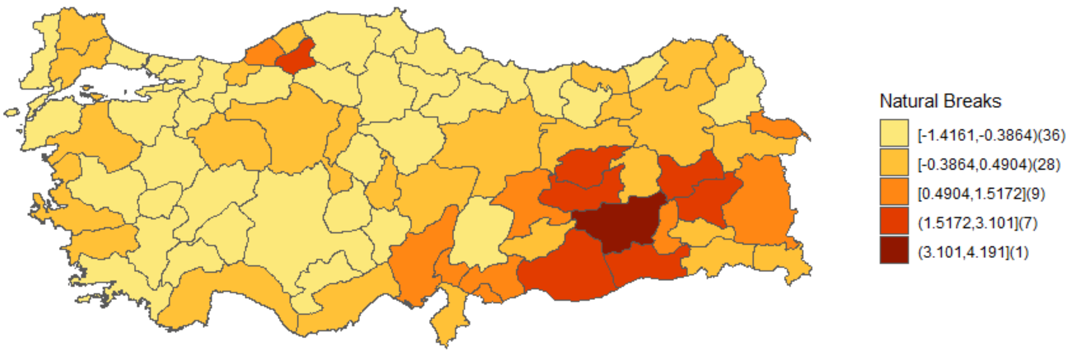

Figure 5.1 Unhappiness Rate of People Aged 25-34 in 2013

As is seen in the Figure 5.2, Voting Turnout Rate in 2015 has a strong spatial autocorrelation on the east-west axis. Unhappiness Rate of People Aged 25-34 in 2013 is seen to have an almost isotropic clustered, particularly in the vicinity of Diyarbakır (i=21).

In terms of interpretability, Spatial Theta Lag Based on Piecewise Linear BCSI $b^l_{X_i}(\theta)$ is to be used in the application of the method proposed in the dissertation. Because $b^l_{X_i}(\theta)$ is considered as a specific form of quantile function. Then Spatial Theta Lag and its Conditional Contiguity Matrix are defined in the expressions 4.27 and 4.25. Spatial Theta Lag $L_i\theta$ is computed based on z-scores.

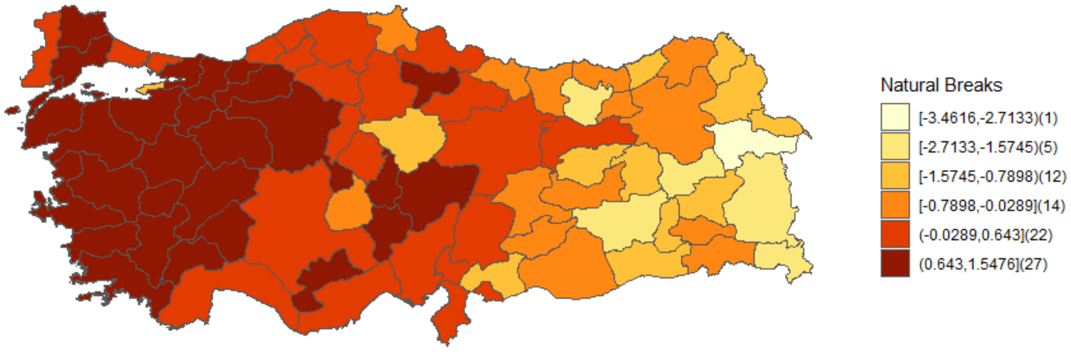

Figure 5.2 Voting Turnout Rate in 2015

To illusrate the Spatial Theta Lag province Nigde (i=51) and its neighborhood set are used. Unhappiness Rate of People Aged 25-34 in 2013 of Nigde and its z-score are 6.84 and -0.5964, respectively. L_i denotes the average of z-scores of neighborhood set of Spatial Unit with $i-code$, used in classical spatial autocorrealtion studies as a spatial lag. Then $L_{51} = -0.1004$. The observations of the spatial units in the neighborhood set in a ranked (or sorted) structure ($x_{r_{51}^k}$) and their z-scores ($z_{r_{51}^k}$) for Nigde are given in table 5.3.

By the Table 5.3, the values of Spatial Theta Lag are computed by means of the definition of Piecewise Linear BCSI in 4.27 and 4.25. The computations are made for $\theta = 0, 0.1, 0.2, \ldots, 0.9, 1$. The results are shown step-by-step in the table 5.4. The Graph of L_i^{θ} is given in 5.3.

Table 5.3 Unhappiness Rate of People Aged 25-34 in 2013 for neighborhood set of Nigde (i=51)

Provinces	Adana	Mersin	Kayseri	Konya	Nevsehir	Aksaray
$\boldsymbol{k}$	6	5	4	2	3	1
$\boldsymbol{r_{51}^k}$	1	33	38	42	50	68
$\boldsymbol{x_{r_{51}^k}}$	11,88	10,61	8,81	7,03	8,44	5,67
$\boldsymbol{z_{r_{51}^k}}$	0.7192	0.3877	-0.0822	-0.5468	-0.1788	-0.9018

To elucidate spatial theta lag, members chosen from group by n_i and their scaled neighborhood sets are presented in Figure 5.4. Whereas lines indicate scaled

Table 5.4 Values of Spatial Theta Lag for Nigde (i=51) by 0.1 incerement through interval [0,1]

θ	$r_{i,\theta}$	k	$k+1$	r_i^k	r_i^{k+1}	$w_{ir_i^k}$	$w_{ir_i^{k+1}}$	$z_{r_i^k}$	zr_i^{k+1}	L_i^θ
0	1	1	2	68	42	1	0	-0.9018	-0.5468	-0.9018
0.1	1.5	1	2	68	42	0.5	0.5	-0.9018	-0.5468	-0.7243
0.2	2	2	3	42	50	1	0	-0.5468	-0.1788	-0.5468
0.3	2.5	2	3	42	50	0.5	0.5	-0.5468	-0.1788	-0.3628
0.4	3	3	4	50	38	1	0	-0.1788	-0.0822	-0.1788
0.5	3.5	3	4	50	38	0.5	0.5	-0.1788	-0.0822	-0.1305
0.6	4	4	5	38	33	1	0	-0.0822	0.3877	-0.0822
0.7	4.5	4	5	38	33	0.5	0.5	-0.0822	0.3877	0.15275
0.8	5	5	6	33	1	1	0	0.3877	0.7192	0.3877
0.9	5.5	5	6	33	1	0.5	0.5	0.3877	0.7192	0.55345
1	6	6	ND	1	NA	1	ND	0.7192	-	0.7192

neighborhoods, points on the lines indicates values of provinces in neighborhood. Graph is optimized up to maximization of interval between vertical z_i and horizontal L_i^θ points by permutational method.

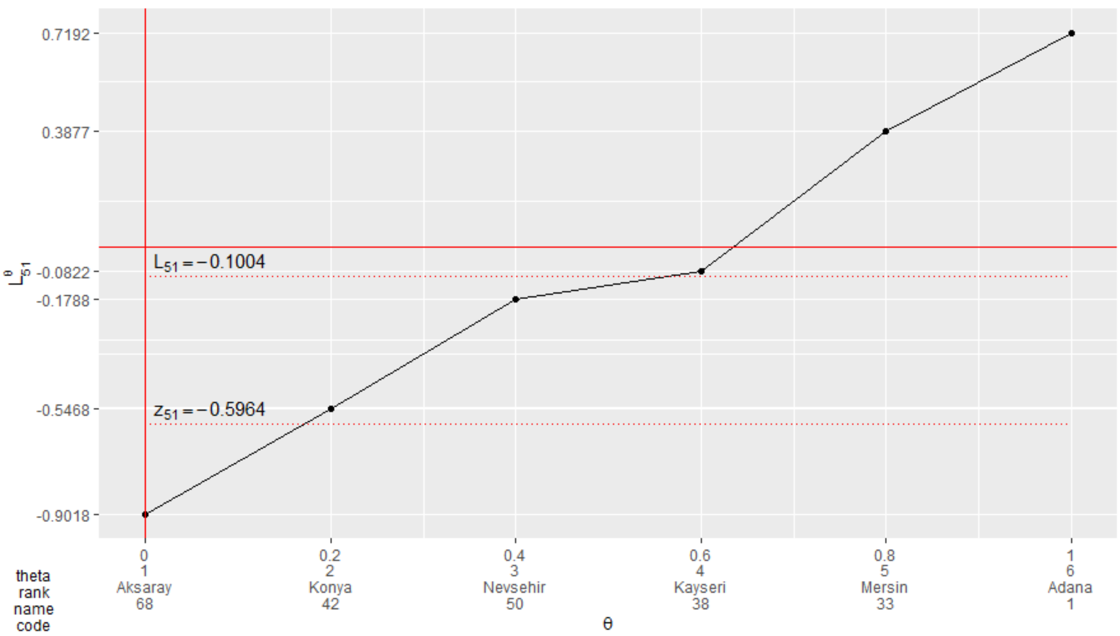

Figure 5.3 L_{51}^θ based on z-scores of Unhappiness Rate of People Aged 25-34 in 2013

5.1.1 Dealing with Repetitive Observations

As a example of occurence of with repetitive observations in neighborhood set, Samsun (i=55), with 5 neighbors, based on Voting Turnout Rate in 2015 presented in Table 5.5. The observation (x_{55}) and z-score (z_{55}) of Samsun are 89.1 and 0.2953, respectively.

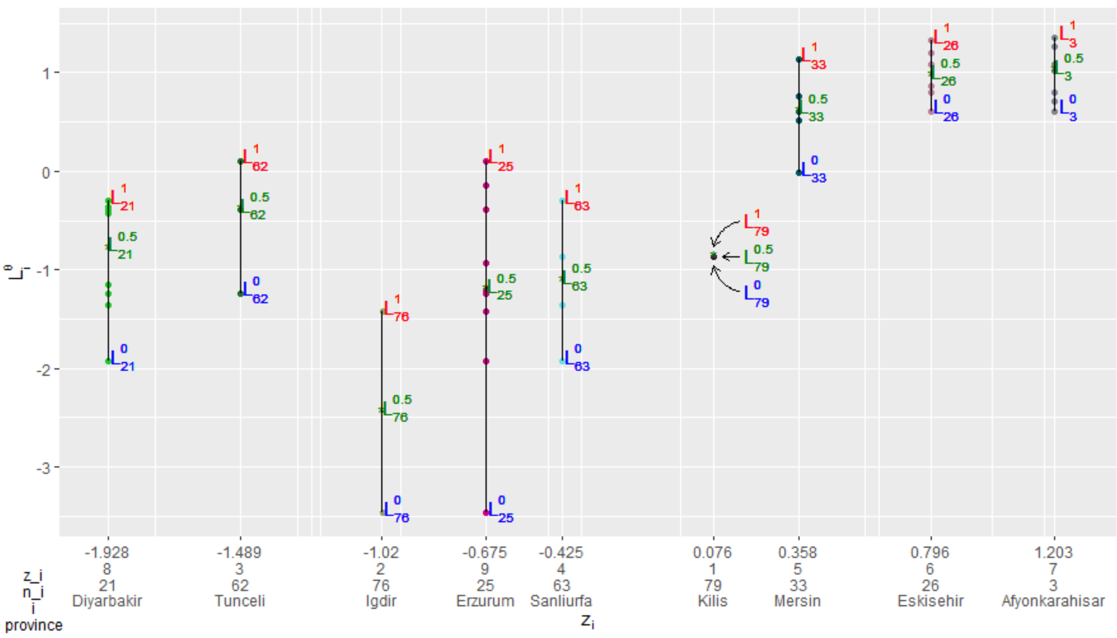

Figure 5.4 Graph of L_i^θ at Voting Turnout rate in 2015 (Labels indicate only ordinates of points)

Mean of z-score of neighbors (L_{55}) is 0.1826 . As is seen, the observation of Ordu (i=52) and Sinop (i=57) are equal. In this case code of these provinces are selected as a secondary criteria for ranking. Even if reverse index of these provinces ($r_{55}^1 = 57$ *and* $r_{55}^2 = 52$) are chosen as secondary sorting criteria, spatial theta lag function L_{55}^θ does not change as seen in Figure 5.5. If one say in mathematical, spatial theta lag preserves invariant against changing rank of equal values. Furthermore, L_i^θ is constant between the repetitive observations. The example of this fact can be seen in Figure 5.5 such that

$$L_{55}^\theta = L_{55}^0 = L_{55}^{0.25} = x_{r_{55}^1} = x_{r_{55}^2} = -0.4248 \tag{5.2}$$

where $0 \le \theta \le 0.25$.

Table 5.5 Voting Turnout Rate in 2015 for neighborhood set of Nigde (i=51)

Provinces	Amasya	Corum	Ordu	Sinop	Tokat
$\boldsymbol{k}$	5	3	1	2	4
$\boldsymbol{r_{55}^k}$	5	19	52	57	60
$\boldsymbol{x_{r_{55}^k}}$	91.2	89.3	86.8	86.8	89.6
$\boldsymbol{z_{r_{55}^k}}$	0.9527	0.3579	-0.4248	-0.4248	0.4518

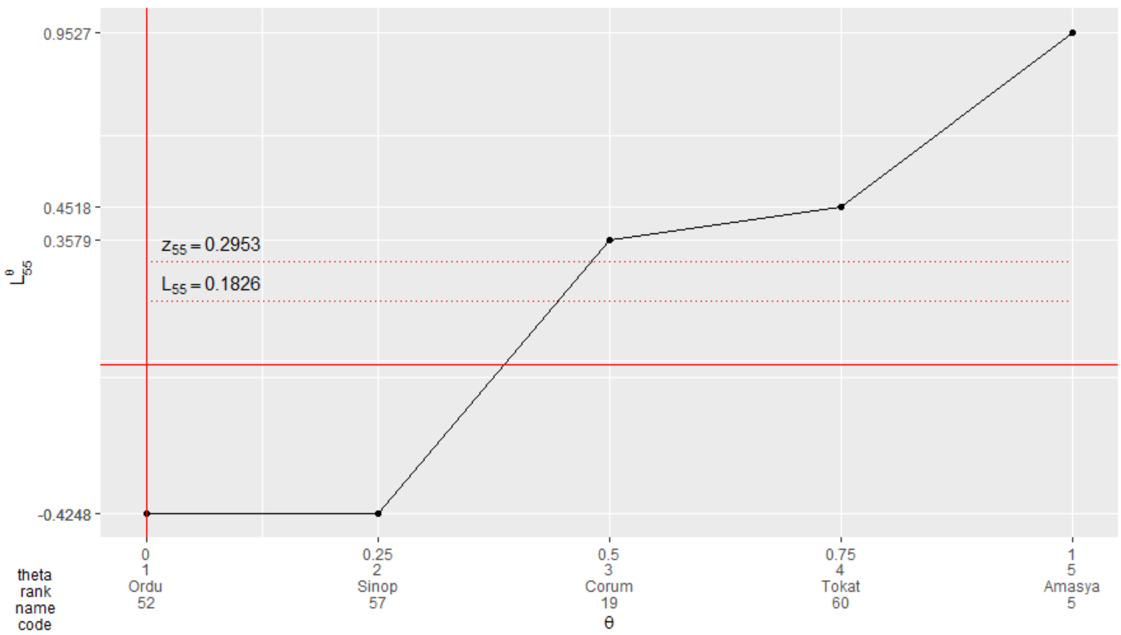

Figure 5.5 Graph of L^{θ}_{55} at Voting Turnout rate in 2015

5.2 Conditional Global Moran

Global Autocorrelation values of spatial theta lag with z-score of is examined by using Global Moran' I. Due to CPU and time constraints, permutation numbers for Figure 5.6 and Table 5.6 are chosen as 9999 and 999999, respectively. Owing to the same reason, increments is determined as 0.1 and 0.02 for Table 5.6 and Figure 5.6, respectively . All computed points are statistically significant at level $pseudo - p < 0.0001$.

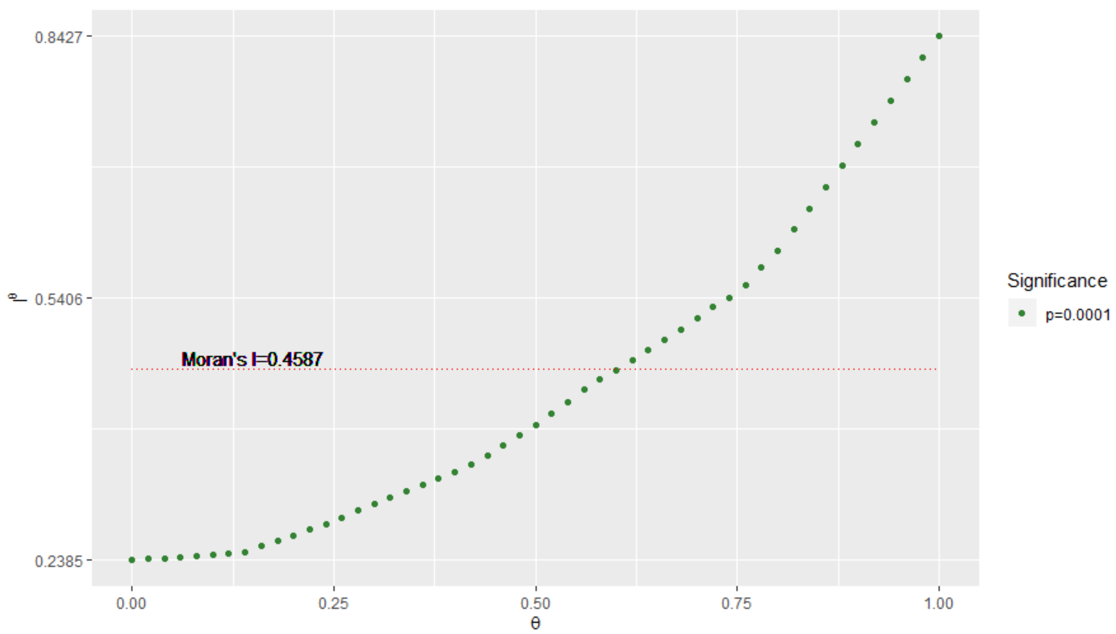

Figure 5.6 Graph of I^{θ} values for z-scores of Unhappiness Rate Ages 25-34 in 2013

There is no differentiation on significance levels in Figure 5.6. The reason for this

Table 5.6 Global Moran's I Unhappiness Rate of People Aged 25-34 in 2013 under Standard and Conditional Contiguity

Standard Contiguity		
Moran's I		**pseudo-p**
0.4587		0.000001
Conditional Contiguity		
θ	I^{θ}	**pseudo-p**
0	0.2385	0.000093
0.1	0.2453	0.000068
0.2	0.2675	0.000041
0.3	0.3043	0.000020
0.4	0.3397	0.000009
0.5	0.3939	0.000007
0.6	0.4583	0.000001
0.7	0.5179	0.000001
0.8	0.5952	0.000001
0.9	0.7189	0.000001
1	0.8427	0.000001

is directly related to number of permutation. Pseudo p value depends on number of permutations by definition. 9999 permutations is not sufficient to see true level of relevant pseudo p values for this data but to show all pseudo p values below 0.0001. Likewise 999999 permutations sufficent for determination of true pseudo p value level of I^{θ} less than $\theta = 0.6$, but not for that of I^{θ} equal and greater than $\theta = 0.6$. In such cases, the fact that I^{θ} is included to permutations ensures pseudo p value being greater than zero with one example, itself.

Furthermore in order to understand I^{θ} for this case , it can be modeled by using a parabolic function such that

$$I^{\theta} \approx \beta_0 + \beta_1 \theta^2 \tag{5.3}$$

With ordinary least square, β_0 and β_1 are computed as 0.2434 and 0.5816 and their p-values are below 0.001, respectively. It should be keep in mind that parabolic model in approximate equation 5.3 pertains to Unhappiness Rate Ages 25-34 in 2013. Naturally different spatial data can entail different model.

5.3 Conditional Local Moran

Effect of spatial theta lag over Local Spatial Autocorrelation is also examined for specific theta values by employing Voting Turnout in 2015. Some values of neighborhood set such as L_i^0,$L_i^{0.5}$,L_i^1 and L_i, are of importance in terms of intrepretability. In this part, significance of Local Moran's I values, as LISA, of them are examined. Not only significance level (0.01) but also FDR (False Discovery Rate) and Bonferroni Bound (BfB, as a abbreviation) conditions relevant to significance level (0.01) are employed for determination significance in order to avoid false positives (or Type 1 error) [321]. For all computation of pseudo p of I_i^0,$I_i^{0.5}$,I_i^1 and I_i , 999999 permutations are used for all provinces. Moreover the parameters for I_i^0,$I_i^{0.5}$,I_i^1 and I_i are generated by same value set (with n_i elements) for each permutation for province i, not separately.

5.3.1 Minimum Local Moran vs. Maximum Local Moran

Significant spatial units of Local Moran's I for Minimum Local Moran I_i^0 and Maximum Local Moran I_i^1 are distincly separated from each other. Significant spatial units of Local Moran's I for Local Minimum I_i^1 are overlapped with High-High clusters of classic LISA mean based I_i, whereas those of Local Moran's I for Local Minimum I_i^1 are overlapped with Low-Low clusters of same analysis.

Figure 5.7 LISA Cluster Map for I_i^0

As is seen, significant units of local minimum are clustered in West of Turkey in Figure 5.7 , those of local maximum are clustered in East of Turkey in Figure 5.8. Results about both of them is presented in Table 5.7[1] That fact is consistent with Figure 5.2. Voting Turnouts of Western provinces of Turkey are “spatially” higher than other provinces. Consequently Local minimum of a Western provinces is higher than that

[1]* It satisfies also only FDR condition, ** It satisfies both FDR and Bonferroni bound conditions

Table 5.7 Provinces of which pseudo p of I_i^0 and I_i^1 are below 0.01, and their z-scores

i	Province	z_i	I_i^0	Pseudo-p	I_i^1	Pseudo-p
3	Afyonkarahisar	1.2032	0.732	0.000349*	1.6361	0.168715
4	Agri	-3.4616	6.9974	0.145116	2.3374	0.000056**
7	Antalya	0.5144	0.1841	0.009232	0.6512	0.272578
9	Aydin	0.7962	0.5342	0.009511	1.2322	0.049929
10	Balikesir	0.9841	0.6603	0.002721	1.5229	0.062174
11	Bilecik	1.3284	1.0161	0.00139	1.432	0.330822
12	Bingol	-1.2388	2.3878	0.33025	-0.1331	0.004497
13	Bitlis	-1.3327	4.6132	0.062014	0.274	0.005899
14	Bolu	0.7962	0.1354	0.007355	1.0577	0.273736
15	Burdur	1.2658	0.6512	0.005964	1.7212	0.121678
20	Denizli	1.3597	1.0401	0.000316	2.1043	0.074966
21	Diyarbakir	-1.9275	3.7154	0.349581	0.5774	0.00010**
24	Erzincan	0.1075	-0.1768	0.540588	0.0183	0.000536*
25	Erzurum	-0.6752	2.3374	0.112281	-0.0726	0.000222*
26	Eskisehir	0.7962	0.4844	0.001249	1.0577	0.211048
36	Kars	-1.4266	4.9384	0.050283	0.9633	0.001994
43	Kutahya	1.078	0.8246	0.000071**	1.6683	0.087255
45	Manisa	1.5476	1.0384	0.000764	2.1043	0.074334
47	Mardin	-1.364	2.6291	0.282657	0.2805	0.005907
49	Mus	-1.9275	6.6723	0.074915	1.3015	0.000058**
54	Sakarya	0.9841	0.7527	0.001384	1.3073	0.177743
56	Siirt	-0.2056	0.4157	0.178402	0.1324	0.000694*
64	Usak	1.0154	1.0945	0.000023**	1.5714	0.049859
65	Van	-2.0215	6.9974	0.062172	0.4157	0.006053
66	Yozgat	-1.1762	-0.2	0.009116	-1.1206	0.325914
72	Batman	-1.1448	2.2067	0.282371	0.2354	0.005937
76	Igdir	-1.0196	3.5295	0.025083	1.4546	0.008617

of another province outside of West of Turkey. Likewise similar inference is made for Eastern provinces.

When interpreting Figure 5.7 and 5.8, all spatial units satisfying Bonferroni Bound condition also satisfies FDR condition by their definitions. 4 and 2 provinces for Local Moran's I for Local Minimum (I_i^0), and for 6 and 3 provinces for Local Moran's I for Local Maximum(I_i^1) satisfying FDR and Bonferroni Bound conditions, respectively are detected. Even though Siirt (i=56) are isolated, all the other provinces satisfying FDR and Bonferroni Bound conditions form clusters. Yozgat (i=66) is isolated from the cluster in Figure 5.7. z_{66} (-1.1762) and L_{66}^0 (0.1701) are considered to be spatially low (due to the skewness of the empirical distrirbution is negative (-0.9409)) and high (under randomization with 999999 permutation) respectively. Erzincan (i=24),

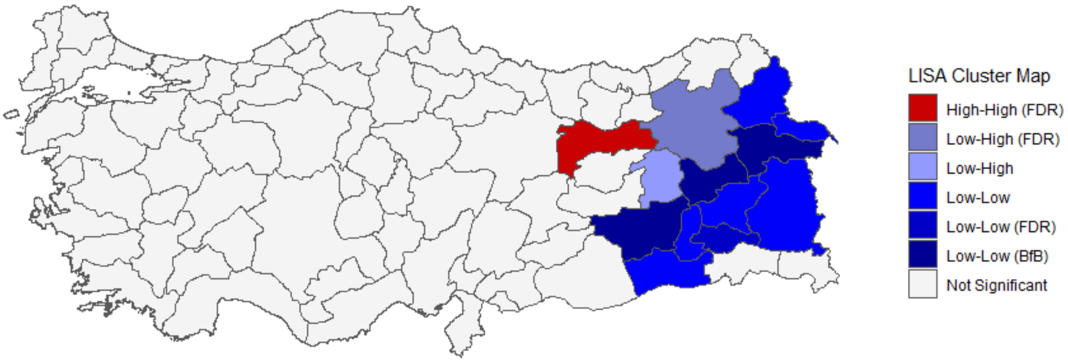

Figure 5.8 6 LISA Cluster Map for I_i^1

Bingol(i=12) and Erzurum (i=25) can be similarly interpreted.

The results of Conditional Local Moran's validate the inference on the neighbors of Minimum and Maximum. Agri (i=4) is the province with minimum observation in Table 5.2 and $z_4 = -3.4616$. Relevant Theoretical and empirical Pseudo-p values of neigbours of Agri are given below

Table 5.8 Theoretical and Empirical Pseudo p for I_k^0 of the provinces in the neighborhood set of Agri (i=4)

k	Province	n_k	Theoretical Pseudo p	Empirical Pseudo p
13	Bitlis	5	0.062500	0.062014
25	Erzurum	9	0.112500	0.112281
36	Kars	4	0.050000	0.050283
49	Mus	6	0.075000	0.074915
65	Van	5	0.062500	0.062172
76	Igdir	2	0.025000	0.025083

Likewise Manisa (i=45) is the province with maximum observation in Table 5.2 and $z_{45} = 1.5476$. Relevant Theoretical and empirical Pseudo-p values of neigbours of Manisa are given below

As minimum and maximum observations are not repetitive ($s = 1$) Theoretical Pseudo

Table 5.9 Theoretical and Empirical Pseudo p for I_k^1 of the provinces in the neighborhood set of Manisa (i=45)

k	Province	n_k	Theoretical Pseudo p	Empirical Pseudo p
4	Aydın	4	0.050000	0.049929
10	Balıkesir	5	0.062500	0.062174
20	Denizli	6	0.075000	0.074966
35	Izmir	3	0.037500	0.037560
43	Kutahya	7	0.087500	0.087255
64	Usak	4	0.050000	0.049859

p values are computed in Tables 5.8 and 5.9 thanks to the equation 4.34, such that

$$\text{Pseudo p value for } I_k^0 \text{ or } I_k^1 = \frac{n_k}{80} \tag{5.4}$$

where n_k is the count of neighbors of province labbeled by k

5.3.2 Local Moran vs. Median Local Moran

It is obivous that LISA cluster map of $I_i^{0.5}$ (or Anselin's I_i^M) and I_i combine significants of I_i^0 and I_i^1 in a certain extent by preserving spatial significance properties (i.e. High-High, Low-Low). One can say that High-High and Low-Low clusters are polarized in east-west direction in both cluter map of $I_i^{0.5}$ and I_i in Figure 7.

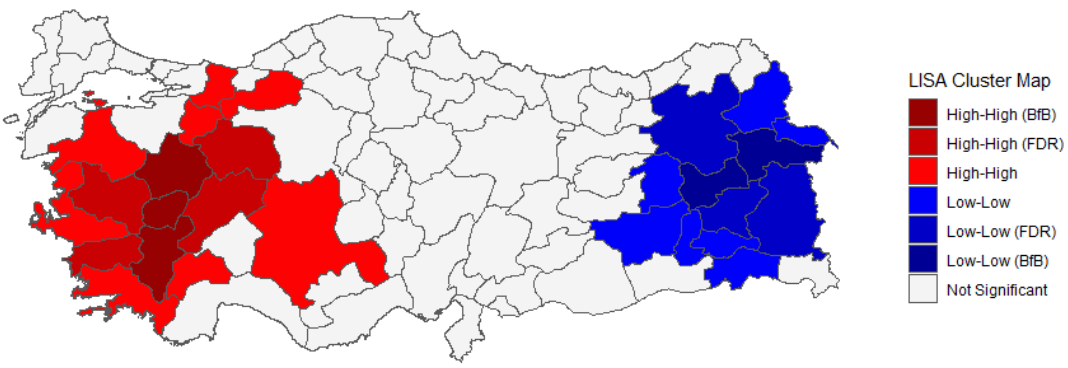

Figure 5.9 LISA Cluster Map for I_i

LISA cluster of Median Local Moran $I_i^{0.5}$ have no province satisfying FDR and

Table 5.10 Provinces of which pseudo p of I_i and $I_i^{0.5}$ are below 0.01, and their z-scores

i	Province	z_i	I_i	Pseudo-p	$I_i^{0.5}$	Pseudo-p
3	Afyonkarahisar	1.2032	1.1733	0.000186*	1.2217	0.000742
4	Agri	-3.4616	4.848	0.00008**	4.7758	0.001306
9	Aydin	0.7962	0.8645	0.001433*	0.8458	0.006093
10	Balikesir	0.9841	0.956	0.002075	0.7835	0.084095
11	Bilecik	1.3284	1.1742	0.005519	1.0577	0.083675
12	Bingol	-1.2388	1.3019	0.006888	1.3406	0.012014
13	Bitlis	-1.3327	2.3352	0.000133*	2.5688	0.001242
14	Bolu	0.7962	0.5872	0.005369	0.6589	0.015952
15	Burdur	1.2658	1.1505	0.003951	0.9682	0.139609
16	Bursa	0.7649	0.5452	0.020685	0.7527	0.003635
20	Denizli	1.3597	1.4942	0.000021**	1.5083	0.000154
21	Diyarbakir	-1.9275	1.724	0.005201	1.5127	0.018801
25	Erzurum	-0.6752	0.7965	0.00029*	0.8153	0.002996
26	Eskisehir	0.7962	0.7794	0.000625*	0.7711	0.004983
35	Izmir	0.671	0.7443	0.004444	0.6603	0.039721
36	Kars	-1.4266	2.2697	0.002011	1.5886	0.024439
42	Konya	0.6084	0.4103	0.007874	0.4653	0.08555
43	Kutahya	1.078	1.1765	0.000011**	1.0945	0.000711
45	Manisa	1.5476	1.5229	0.000256*	1.5471	0.001445
48	Mugla	0.7649	0.7527	0.005313	0.7886	0.008005
49	Mus	-1.9275	3.1421	0.000049**	2.4783	0.002942
54	Sakarya	0.9841	0.9314	0.003072	0.9068	0.022203
56	Siirt	-0.2056	0.2676	0.003701	0.274	0.01275
64	Usak	1.0154	1.3171	0.000015**	1.3012	0.000254
65	Van	-2.0215	3.2129	0.000372*	2.694	0.008868
72	Batman	-1.1448	1.5472	0.002432	1.5616	0.008968
73	Sirnak	-0.6439	0.9489	0.003881	1.09	0.001818
76	Igdir	-1.0196	2.492	0.002211	2.492	0.002239

Bonferroni Bound conditions in 5.10 Results about both of them is presented in Table 5.10[2]. That is why local median is not affected by extreme values of neighborhood set. Furthermore that fact also makes number of significant provinces of $I_i^{0.5}$ lower than those of I_i in Table 5.11. Unlike that of $I_i^{0.5}$, LISA Cluster of Local Moran's I_i, or in other words "Local Moran's I for Local Mean" in this context, have a great deal provinces satisfying FDR and Bonferroni Bound conditions due to local means affected by extreme values of neighborhood set in Figure 5.10 and Table 5.11.

LISA cluster of I_i cover those of $I_i^{0.5}$ except one province. In this case (Bursa, i=16)

[2]* It satisfies also only FDR condition, ** It satisfies both FDR and Bonferroni bound conditions

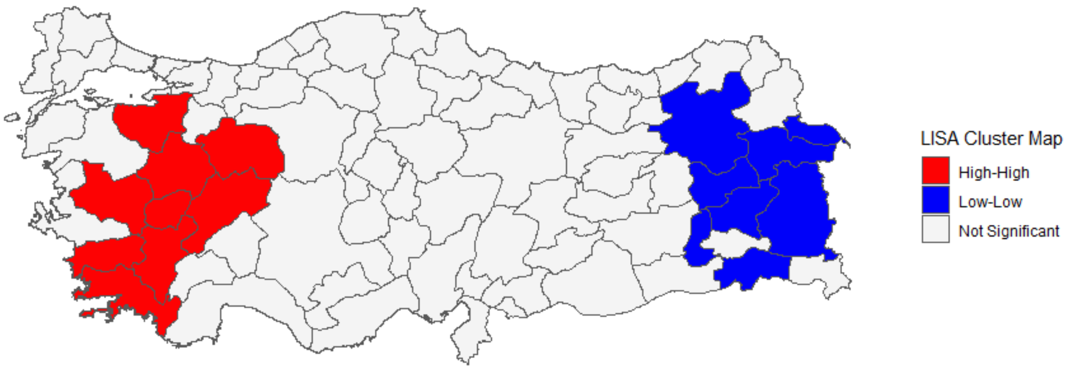

Figure 5.10 LISA Cluster Map for $I_i^{0.5}$

pseudo p value of $I_{16}^{0.5}$ (0.003635) is far lower than that of more I_{16} (0.020685) from Table 5.10.

5.3.3 On Overlapping Clusters

Overlapping is another issue for clusters of significants for I_i^0,I_i^1,$I_i^{0.5}$ and I_i. Relations of significants of I_i and $I_i^{0.5}$ with I_i^0 and I_i^1 are examined in Table 5.11. Clusters of I_i cover those of I_i^0 and I_i^1 in a great proportion. However coverage of clusters of $I_i^{0.5}$ for those of I_i^0 and I_i^1 are so limited due to its stable structure (no change against extreme values).

Most of significant provinces of I_i are common with those of I_i^0 and I_i^1 in Table 5.11: 23 of 27 (signifcants for I_i) for $p < 0.01$, 7 of 12 for FDR and 4 of 5 for Bonferroni Bound conditions. These results can contribute to explanation of occurrence of significant spatial units of I_i.

5.3.4 Conditional Local Moran Through [0,1]

Behaviour of I_i^θ and its change of pseudo p value through θ are also examined. At first glance I_i^θ is nothing but linear transformation of Spatial Theta Lag L_i^θ by definition $(I_i^\theta := z_i L_i^\theta)$. Thus I_i^θ has a structure composed of broken lines with all non-positive or all non-negative slopes.

Two provinces, Agri and Kutahya, with opposite characteristics are examined for

Table 5.11 Overlapping table between significants of $I_i - I_i^{0.5}$ and those of $I_i^0 - I_i^1$

Local Morans		I_i^0			I_i^1		
		p (14)	**FDR (4)**	**BfB (2)**	**p (13)**	**FDR (6)**	**BfB (3)**
I_i	**p (27)**	12	4	2	11	5	3
	FDR (12)	7	4	2	5	3	2
	BfB (5)	3	3	2	2	2	2
$I_i^{0.5}$	**p (17)**	7	4	2	7	3	2
	FDR (0)	0	0	0	0	0	0
	BfB (0)	0	0	0	0	0	0

illustration. Agri (i=4) is an eastern province of which z-score and spatial theta lags values are low, whereas Kutahya (i=43) is a western province of which z-score and spatial theta lags are high. Change of I_4^θ and I_{43}^θ are numerically given in Table 5.12 with 0.1 increment and graphically illustrated in Figures 5.11 and 5.12 with 0.01 increment by using 999999 permutations.

Pseudo p values of I_4^θ decreases and those of I_4^θ increases when θ increases from 0 (Local Minimum) to 1 (Local Maximum) as a general fact in Table 5.12. But some exceptional cases (visible small discontinuities in pseudo p values) may occur in this interval: Interestingly whereas pseudo p for $I_4^{0.60}$ is 0.001097, pseudo p values for $I_4^{0.58}$ $I_4^{0.59}$ $I_4^{0.61}$ and $I_4^{0.62}$ are 0.000883, 0.000900, 0.000521 and 0.000337, respectively. The reason of that phenomenon may be an effect of a possible bias in permutations in micro level. Other interesting case even though pseudo p value of I_4^1 is 0.087591, that of $I_4^{0.99}$ is 0.012955. Those cases in pseudo p values may be considered to be small jumpings (or visible discontinuities). If one decreases increments (i.e. 0.001, 0.0001), such cases can be interpreted better.

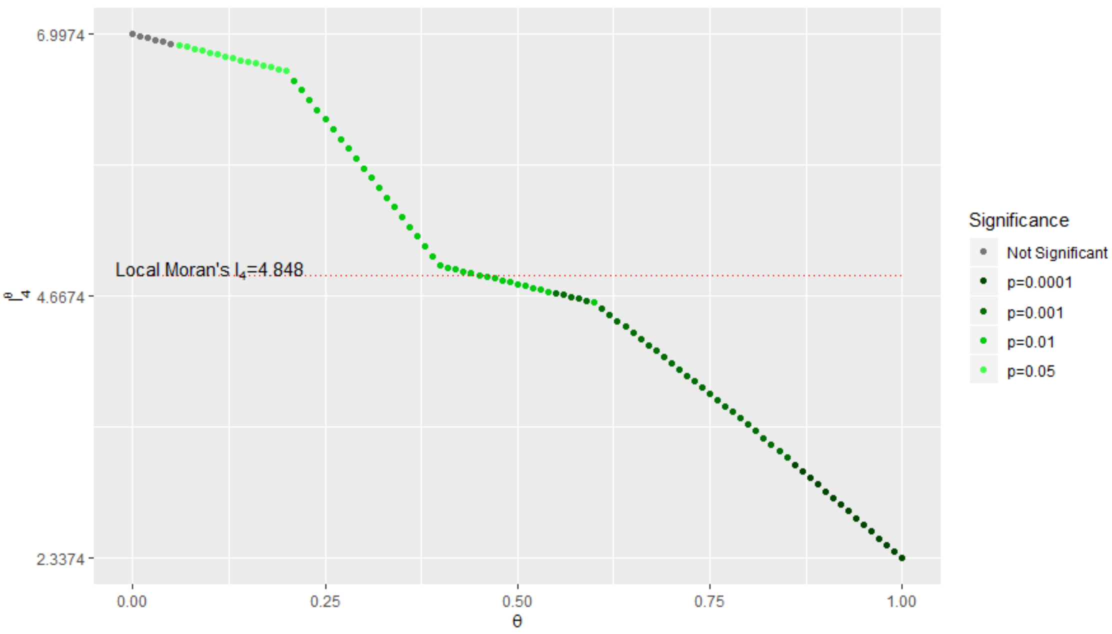

Figure 5.11 I_4^{θ} through [0,1]

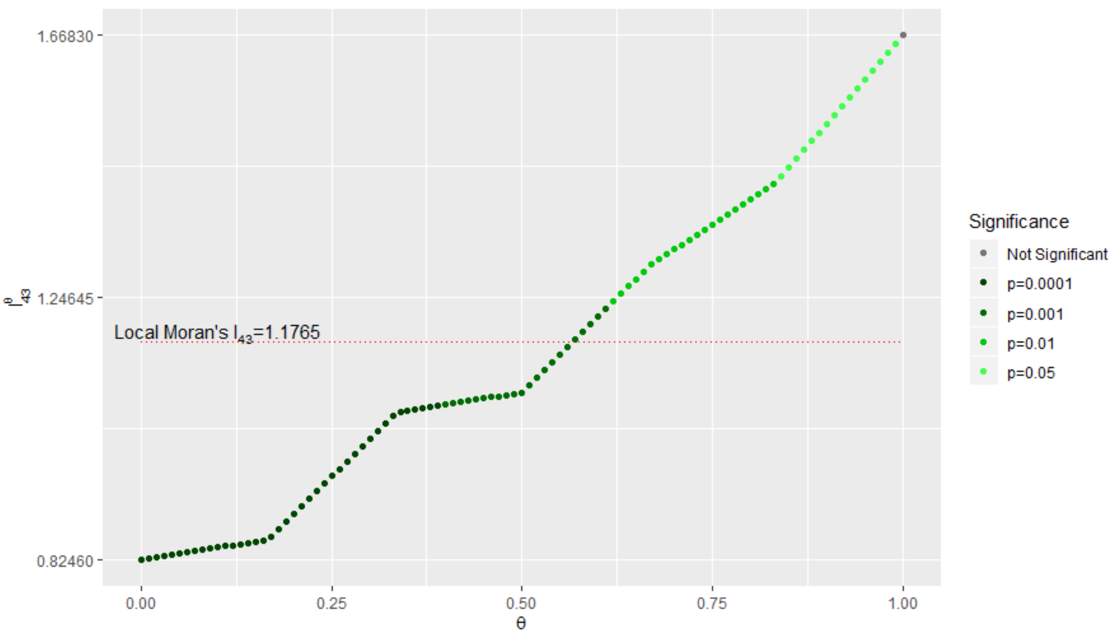

Figure 5.12 I_{43}^{θ} through [0,1]

Table 5.12 Change of I_4^θ and I_{43}^θ and their pesudo-p values through [0,1]

I_4		pseudo-p	I_{43}	Pseudo-p
4.8480		0.000080	1.1765	0.000011
θ	I_4^θ	Pseudo-p	I_{43}^θ	Pseudo-p
0	6,9974	0.144288	0.8246	0.000077
0.1	6,8349	0.026541	0.8448	0.000053
0.2	6,6723	0.026622	0.8988	0.000076
0.3	5,8053	0.005002	1,0203	0.000032
0.4	4,9384	0.007608	1,0743	0.000110
0.5	4,7758	0.001329	1,0945	0.000744
0.6	4,6132	0.001097	1,216	0.000860
0.7	4,0714	0.000309	1,324	0.002757
0.8	3,5295	0.000642	1,405	0.003101
0.9	2,9334	0.000069	1,5265	0.012778
1	2,3374	0.000076	1,6683	0.087591

6 Suggestions

As seen in the application chapter, Spatial Theta Laga and Coniditonal Contiguity Matrix give impressive results in both global and local spatial autocorrelation when they are applied to convenient data. In spatial autocorrelation studies, global moran's I based on spatial theta lags, like minimum, median and maximum of the neighbourhood set, should be examined in addition to the classical approach based on the mean.

The examination of the spatial autocorrelation based on quantiles of the neighbourhood set, called spatial theta lag, is important. Because in global scale, these autocorrelations may be stronger than classical spatial autocorrelation. Furthermore, clusters that occurred in the classical approach can be decomposed into the clusters that occurred under the spatial autocorrelation minimum and maximum, as shown in the application chapter. Thereby the reasons for the occurrence of the clusters in the classical approach (mean based) can be explained with the help of the clusters based on the minimum, and maximum local moran's I. The clusters or spatial units, which can not be explained in this way, can be explained by focusing general level of elements of the neighbourhood set, instead of extreme observations in the set.

Another suggestion is to focus on changing the significant level of conditional local moran's I at a sub-interval of [0,1] for a spatial unit instead of a pointwise approach. And these levels can be compared to each other. Furthermore discontinuities of pseudo p of conditional local moran's I may be another issue to be studied.

Lastly, Spatial theta lag model can be integrated into Spatial Lag models (SAR). The spatial theta regression model can be defined as

$$Y_i = \beta_{\theta,0} + \beta_{\theta,1} x_i + \beta_{\theta,2} L_i^{\theta} + \epsilon_{\theta,i}$$

Where

$$L_i^{\theta} = \sum_{i=1}^{n} w_{ij}^{\theta} Y_i$$

and $0 \leq \theta \leq 1$. In a similar way, spatial theta lag can be expanded to other spatial regression models, Spatially Lagged X (SLX) and Spatial Error Model (SEM). It should be kept in mind that models are up to θ.And performance of models must be optimized by adjusting θ.

REFERENCES

[1] G. Ackley, "Spiethoff's views on the business cycle," 1954.

[2] W. Isard, "The general theory of location and space-economy," *The Quarterly Journal of Economics*, vol. 63, no. 4, pp. 476–506, 1949.

[3] ——, "Location and space-economy," 1956.

[4] ——, "Regional science, the concept of region, and regional structure," *Papers in Regional Science*, vol. 2, no. 1, pp. 13–26, 1956.

[5] P. Krishna Iyer, "The first and second moments of someprobability distributions arising from points on a lattice and their application," *Biometrika*, vol. 36, pp. 135–141, 1949.

[6] P. A. Moran, "The interpretation of statistical maps," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 10, no. 2, pp. 243–251, 1948.

[7] ——, "A test for the serial independence of residuals," *Biometrika*, vol. 37, no. 1/2, pp. 178–181, 1950.

[8] R. C. Geary, "The contiguity ratio and statistical mapping," *The incorporated statistician*, vol. 5, no. 3, pp. 115–146, 1954.

[9] A. Cliff, K. Ord, "Testing for spatial autocorrelation among regression residuals," *Geographical analysis*, vol. 4, no. 3, pp. 267–284, 1972.

[10] A. Cliff, R. L. Martin, J. Ord, "Evaluating the friction of distance parameter in gravity models," *Regional Studies*, vol. 8, no. 3-4, pp. 281–286, 1974.

[11] L. Anselin, A. Getis, "Spatial statistical analysis and geographic information systems," *The Annals of Regional Science*, vol. 26, no. 1, pp. 19–33, 1992.

[12] J. K. Ord, A. Getis, "Local spatial autocorrelation statistics: Distributional issues and an application," *Geographical analysis*, vol. 27, no. 4, pp. 286–306, 1995.

[13] A. Getis, "Spatial interaction and spatial autocorrelation: A cross-product approach," *Environment and Planning A*, vol. 23, no. 9, pp. 1269–1277, 1991.

[14] A. Getis, J. K. Ord, "The analysis of spatial association by use of distance statistics," in *Perspectives on Spatial Data Analysis*, Springer, 2010, pp. 127–145.

[15] A. Okabe, "A note on geary's spatial contiguity ratio," *Geographical Analysis*, vol. 8, no. 3, pp. 315–318, 1976.

[16] P. De Jong, C. Sprenger, F. Van Veen, "On extreme values of moran's i and geary's c," *Geographical Analysis*, vol. 16, no. 1, pp. 17–24, 1984.

[17] J. Odland, *Spatial autocorrelation*. Sage Publications Newbury Park, CA, 1988, vol. 9.

[18] M. F. Goodchild, *Spatial autocorrelation*. Geo Books, 1986, vol. 47.

[19] L. Anselin, "Local indicators of spatial association—lisa," *Geographical analysis*, vol. 27, no. 2, pp. 93–115, 1995.

[20] P. Legendre, "Spatial autocorrelation: Trouble or new paradigm?" *Ecology*, vol. 74, no. 6, pp. 1659–1673, 1993.

[21] B. Boots, C. Dufournaud, "A programming approach to minimizing and maximizing spatial autocorrelation statistics," *Geographical Analysis*, vol. 26, no. 1, pp. 54–66, 1994.

[22] A. Unwin, "Geary's contiguity ratio," *Economic and Social Review*, vol. 27, no. 2, pp. 145–159, 1996.

[23] B. H. Baltagi, "Random effects and spatial autocorrelation with equal weights," *Econometric Theory*, vol. 22, no. 5, pp. 973–984, 2006.

[24] L. La Tona, A. Mazza, M. Mucciardi, "A distance weight matrix determination for superficial spatial data," 2005.

[25] J. Aldstadt, A. Getis, "Using amoeba to create a spatial weights matrix and identify spatial clusters," *Geographical analysis*, vol. 38, no. 4, pp. 327–343, 2006.

[26] D. A. Griffith, "Hidden negative spatial autocorrelation," *Journal of Geographical Systems*, vol. 8, no. 4, pp. 335–355, 2006.

[27] S. Dray, P. Legendre, P. R. Peres-Neto, "Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (pcnm)," *ecological modelling*, vol. 196, no. 3-4, pp. 483–493, 2006.

[28] D. A. Griffith, *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*. Springer Science & Business Media, 2013.

[29] M. Tiefelsdorf, D. A. Griffith, "Semiparametric filtering of spatial autocorrelation: The eigenvector approach," *Environment and Planning A*, vol. 39, no. 5, pp. 1193–1221, 2007.

[30] D. A. Griffith, P. R. Peres-Neto, "Spatial modeling in ecology: The flexibility of eigenfunction spatial analyses," *Ecology*, vol. 87, no. 10, pp. 2603–2613, 2006.

[31] A. Brind'Amour, S. Mahévas, P. Legendre, L. Bellanger, "Application of moran eigenvector maps (mem) to irregular sampling designs," *Spatial Statistics*, vol. 26, pp. 56–68, 2018.

[32] Y. Maruyama, "An alternative to moran's i for spatial autocorrelation," *arXiv preprint arXiv:1501.06260*, 2015.

[33] F. Diaz, "Performance prediction using spatial autocorrelation," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, 2007, pp. 583–590.

[34] C. F. Dormann, "Assessing the validity of autologistic regression," *ecological modelling*, vol. 207, no. 2-4, pp. 234–242, 2007.

[35] I. Kühn, "Incorporating spatial autocorrelation may invert observed patterns," *Diversity and Distributions*, vol. 13, no. 1, pp. 66–69, 2007.

[36] Y. Chen, “On the four types of weight functions for spatial contiguity matrix,” *Letters in Spatial and Resource Sciences*, vol. 5, no. 2, pp. 65–72, 2012.

[37] S. Dray, “A new perspective about moran’s coefficient: Spatial autocorrelation as a linear regression problem. moran,” *Geographical Analysis*, vol. 43, no. 2, pp. 127–141, 2011.

[38] Y. Chen, “A new methodology of spatial cross-correlation analysis,” *PloS one*, vol. 10, no. 5, e0126158, 2015.

[39] D. Lee, J. Minton, G. Pryce, “Bayesian inference for the dissimilarity index in the presence of spatial autocorrelation,” *Spatial Statistics*, vol. 11, pp. 81–95, 2015.

[40] E. Paradis, “Moran’s autocorrelation coefficient in comparative methods,” 2019.

[41] A. Shchiptsova, *Testing spatial autocorrelation with the bootstrap*, 2016.

[42] F. Osorio, R. Vallejos, F. Cuevas, “Spatialpack: Package for analysis of spatial data,” *R package version 0.2-3, URL: CRAN. R-project. org/package= SpatialPack*, 2014.

[43] A. D. Cliff, J. K. Ord, *The problem of spatial autocorrelation*. University, 1968.

[44] R. Bivand, ““the problem of spatial autocorrelation:” forty years on,” 2018.

[45] A. Kreuzer, T. Erhardt, T. Nagler, C. Czado, “Heavy tailed spatial autocorrelation models,” *arXiv preprint arXiv:1707.03165*, 2017.

[46] R. Beecham, J. Dykes, W. Meulemans, A. Slingsby, C. Turkay, J. Wood, “Map lineups: Effects of spatial structure on graphical inference,” *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 391–400, 2016.

[47] A. Getis, “A history of the concept of spatial autocorrelation: A geographer’s perspective,” *Geographical Analysis*, vol. 40, no. 3, pp. 297–309, 2008.

[48] D. A. Griffith, “Spatial structure and spatial interaction: 25 years later,” *Review of Regional Studies*, vol. 37, no. 1, pp. 28–38, 2007.

[49] R. P. Haining, “Spatial autocorrelation and the quantitative revolution,” *Geographical Analysis*, vol. 41, no. 4, pp. 364–374, 2009.

[50] D. A. Griffith, “Spatial autocorrelation,” *International encyclopedia of human geography*, vol. 2009, pp. 308–316, 2009.

[51] R. P. Haining, R. Haining, *Spatial data analysis: theory and practice*. Cambridge University Press, 2003.

[52] F. López, J. Mur, A. Angulo, “Local estimation of spatial autocorrelation processes,” in *Progress in Spatial Analysis*, Springer, 2010, pp. 93–116.

[53] L. Anselin, “Spatial econometrics,” *A companion to theoretical econometrics*, vol. 310330, 2001.

[54] P. Bodson, D. Peeters, “Estimation of the coefficients of a linear regression in the presence of spatial autocorrelation. an application to a belgian labour-demand function,” *Environment and Planning A*, vol. 7, no. 4, pp. 455–472, 1975.

[55] M. M. Fischer, D. A. Griffith, "Modeling spatial autocorrelation in spatial interaction data: An application to patent citation data in the european union," *Journal of Regional Science*, vol. 48, no. 5, pp. 969–989, 2008.

[56] N. Owoo, W. Naudé, "Non-farm enterprise productivity and spatial autocorrelation in rural africa: Evidence from ethiopia and nigeria," 2014.

[57] M. T. Ramirez, A. M. Loboguerrero, "Spatial dependence and economic growth: Evidence from a panel of countries," *Borradores de Economia Working Paper*, no. 206, 2002.

[58] G. Arbia, G. Piras, "Convergence in per-capita gdp across european regions using panel data models extended to spatial autocorrelation effects," 2005.

[59] G. Arbia, R. Basile, G. Piras, "Using spatial panel data in modelling regional growth and convergence," 2005.

[60] V. Nosek, P. Netrdová, "Measuring spatial aspects of variability. comparing spatial autocorrelation with regional decomposition in international unemployment research," *Historical Social Research/Historische Sozialforschung*, pp. 292–314, 2014.

[61] S. Van Weezel, "A spatial analysis of the effect of foreign aid in conflict areas," *http://aiddata. org/aiddata-working-paper-series*, 2015.

[62] S. Ozyurt, S. Dees, *et al.*, "Regional dynamics of growth in the european union: To what extent spatial spillovers matter?" In *ERSA conference papers*, European Regional Science Association, 2015.

[63] Y. M. Kantar, S. G. Aktaş, "Spatial correlation analysis of unemployment rates in turkey," *Journal of Eastern Europe Research in Business and Economics*, vol. 1, no. 9, 2016.

[64] Y. Chen, "Spatial autocorrelation approaches to testing residuals from least squares regression," *PloS one*, vol. 11, no. 1, e0146865, 2016.

[65] I. J. Marasteanu, E. C. Jaenicke, "Hot spots and spatial autocorrelation in certified organic operations in the united states," *Agricultural and Resource Economics Review*, vol. 45, no. 3, pp. 485–521, 2016.

[66] H. Huang, Y. D. Wei, "Spatial-temporal patterns and determinants of foreign direct investment in china," *Erdkunde*, pp. 7–23, 2011.

[67] H. Huang, Y. D. Wei, "Intra-metropolitan location of foreign direct investment in wuhan, china: Institution, urban structure, and accessibility," *Applied geography*, vol. 47, pp. 78–88, 2014.

[68] ——, "Spatial inequality of foreign direct investment in china: Institutional change, agglomeration economies, and market access," *Applied Geography*, vol. 69, pp. 99–111, 2016.

[69] U. Stankov, T. Armenski, M. Klauco, V. Pavluković, M. Cimbaljević, N. Drakulić-Kovačević, "Spatial autocorrelation analysis of tourist arrivals using municipal data: A serbian example," *Geographica Pannonica*, vol. 21, no. 2, pp. 106–114, 2017.

[70] Ç. L. Uslu, "Provincial income inequality and spatial autocorrelation across turkish provinces: 1992-2013.," *Sosyoekonomi*, vol. 25, 2017.

[71] T. J. Fik, D. C. Ling, G. F. Mulligan, "Modeling spatial variation in housing prices: A variable interaction approach," *Real Estate Economics*, vol. 31, no. 4, pp. 623–646, 2003.

[72] S. Basu, T. G. Thibodeau, "Analysis of spatial autocorrelation in house prices," *The Journal of Real Estate Finance and Economics*, vol. 17, no. 1, pp. 61–85, 1998.

[73] B.-M. Hsieh *et al.*, "A study on spatial dependence of housing prices and housing submarkets in tainan metropolis, taiwan," European Real Estate Society (ERES), Tech. Rep., 2011.

[74] R. Bivand, "Spatial econometrics functions in r: Classes and methods," *Journal of geographical systems*, vol. 4, no. 4, pp. 405–421, 2002.

[75] A. Getis, "Reflections on spatial autocorrelation," *Regional Science and Urban Economics*, vol. 37, no. 4, pp. 491–496, 2007.

[76] D. Wartenberg, "Multivariate spatial correlation: A method for exploratory geographical analysis," *Geographical Analysis*, vol. 17, no. 4, pp. 263–283, 1985.

[77] D. G. Cook, S. J. Pocock, "Multiple regression in geographical mortality studies, with allowance for spatially correlated errors.," *Biometrics*, vol. 39, no. 2, pp. 361–371, 1983.

[78] R. J. Marshall, "A review of methods for the statistical analysis of spatial patterns of disease," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 154, no. 3, pp. 421–441, 1991.

[79] G. Lin, T. Zhang, "Loglinear residual tests of moran's i autocorrelation and their applications to kentucky breast cancer data," *Geographical Analysis*, vol. 39, no. 3, pp. 293–310, 2007.

[80] K. Al-Ahmadi, A. Al-Zahrani, "Spatial autocorrelation of cancer incidence in saudi arabia," *International journal of environmental research and public health*, vol. 10, no. 12, pp. 7207–7228, 2013.

[81] G. Jacquez, "Spatial clustering and autocorrelation in health events," *Handbook of Regional Science*, pp. 1311–1334, 2014.

[82] C. Varga, D. L. Pearl, S. A. McEwen, J. M. Sargeant, F. Pollari, M. T. Guerin, "Area-level global and local clustering of human salmonella enteritidis infection rates in the city of toronto, canada, 2007–2009," *BMC infectious diseases*, vol. 15, no. 1, p. 359, 2015.

[83] S. Richardson, C. Guihenneuc-Jouyaux, "Impact of cliff and ord (1969, 1981) on spatial epidemiology," *Geographical Analysis*, vol. 41, no. 4, pp. 444–451, 2009.

[84] P. A. Jumars, D. Thistle, M. Jones, "Detecting two-dimensional spatial structure in biological data," *Oecologia*, vol. 28, no. 2, pp. 109–123, 1977.

[85] R. R. Sokal, N. L. Oden, "Spatial autocorrelation in biology: 1. methodology," *Biological journal of the Linnean Society*, vol. 10, no. 2, pp. 199–228, 1978.

[86] ——, "Spatial autocorrelation in biology: 2. some biological implications and four applications of evolutionary and ecological interest," *Biological Journal of the Linnean Society*, vol. 10, no. 2, pp. 229–249, 1978.

[87] R. R. Sokal, N. L. Oden, B. A. Thomson, “Local spatial autocorrelation in biological variables,” *Biological Journal of the Linnean Society*, vol. 65, no. 1, pp. 41–62, 1998.

[88] R. R. Sokal, G. M. Jacquez, M. Wooten, “Spatial autocorrelation analysis of migration and selection.,” *Genetics*, vol. 121, no. 4, pp. 845–855, 1989.

[89] R. R. Sokal, N. L. Oden, B. A. Thomson, “Local spatial autocorrelation in a biological model,” *Geographical Analysis*, vol. 30, no. 4, pp. 331–354, 1998.

[90] P. E. Smouse, R. Peakall, “Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure,” *Heredity*, vol. 82, no. 5, pp. 561–573, 1999.

[91] O. J. Hardy, S. Vanderhoeven, P. Meerts, X. Vekemans, *et al.*, “Spatial autocorrelation of allozyme and quantitative markers within a natural population of centaurea jacea (asteraceae),” *Journal of Evolutionary Biology*, vol. 13, no. 4, pp. 656–667, 2000.

[92] S. Pálsson, “Isolation by distance, based on microsatellite data, tested with spatial autocorrelation (spaida) and assignment test (spassign),” *Molecular Ecology Notes*, vol. 4, no. 1, pp. 143–145, 2004.

[93] P. M. Waser, L. F. Elliott, “Dispersal and genetic structure in kangaroo rats,” *Evolution*, vol. 45, no. 4, pp. 935–943, 1991.

[94] G. Bertorelle, G. Barbujani, “Analysis of dna diversity by spatial autocorrelation.,” *Genetics*, vol. 140, no. 2, pp. 811–819, 1995.

[95] O. N. Bjørnstad, A. Iversen, M. Hansen, “The spatial structure of the gene pool of a viviparous population of poa alpina-environmental controls and spatial constraints,” *Nordic Journal of Botany*, vol. 15, no. 4, pp. 347–354, 1995.

[96] E. Saillant, S. C. Bradfield, J. R. Gold, “Genetic variation and spatial autocorrelation among young-of-the-year red snapper (lutjanus campechanus) in the northern gulf of mexico,” *ICES Journal of Marine Science*, vol. 67, no. 6, pp. 1240–1250, 2010.

[97] J. Ferguson, M. Bester, “The treatment of spatial autocorrelation in biological surveys: The case of line transect surveys,” *Antarctic Science*, vol. 14, no. 2, pp. 115–122, 2002.

[98] S. Iop, V. M. Caldart, T. Santos, S. Z. Cechin, “What is the role of heterogeneity and spatial autocorrelation of ponds in the organization of frog communities in southern brazil,” *Zoological Studies*, vol. 51, no. 7, pp. 1094–1104, 2012.

[99] H. Van der Heyden, P. Dutilleul, L. Brodeur, O. Carisse, “Spatial distribution of single-nucleotide polymorphisms related to fungicide resistance and implications for sampling,” *Phytopathology*, vol. 104, no. 6, pp. 604–613, 2014.

[100] J. Ma, J. Xiao, X. Gao, B. Liu, H. Chen, H. Wang, “Spatial pattern of foot-and-mouth disease in animals in china, 2010–2016,” *PeerJ*, vol. 5, e4193, 2017.

[101] M. Mathur, “Spatial autocorrelation analysis in plant population: An overview,” *Journal of Applied and Natural Science*, vol. 7, no. 1, pp. 501–513, 2015.

[102] P. Legendre, M. J. Fortin, "Spatial pattern and ecological analysis," *Vegetatio*, vol. 80, no. 2, pp. 107–138, 1989.

[103] W. D. Koenig, J. M. Knops, "Testing for spatial autocorrelation in ecological studies," *Ecography*, vol. 21, no. 4, pp. 423–429, 1998.

[104] M. Valcu, B. Kempenaers, "Spatial autocorrelation: An overlooked concept in behavioral ecology," *Behavioral Ecology*, vol. 21, no. 5, pp. 902–905, 2010.

[105] S. Hamylton, "Five practical uses of spatial autocorrelation for studies of coral reef ecology," *Marine Ecology Progress Series*, vol. 478, pp. 15–25, 2013.

[106] J. W. Lichstein, T. R. Simons, S. A. Shriner, K. E. Franzreb, "Spatial autocorrelation and autoregressive models in ecology," *Ecological monographs*, vol. 72, no. 3, pp. 445–463, 2002.

[107] D. Nestel, B. Katsoyannos, E. Nemny-Lavy, Z. Mendel, N. Papadopoulos, B. Barnes, "Spatial analysis of medfly populations in heterogeneous landscapes," in *Proceedings of the 6th International Symposium on fruit flies of economic importance*, Isteg Scientific Publications Irene, South Africa, 2004, pp. 35–43.

[108] P. C. Tobin, "Estimation of the spatial autocorrelation function: Consequences of sampling dynamic populations in space and time," *Ecography*, vol. 27, no. 6, pp. 767–775, 2004.

[109] R. S. King *et al.*, "Spatial considerations for linking watershed land cover to ecological indicators in streams," *Ecological applications*, vol. 15, no. 1, pp. 137–153, 2005.

[110] A. L. Bataineh, B. P. Oswald, M. Bataineh, D. Unger, I.-K. Hung, D. Scognamillo, "Spatial autocorrelation and pseudoreplication in fire ecology," *Fire Ecology*, vol. 2, no. 2, pp. 107–118, 2006.

[111] M. G. Betts, A. Diamond, G. Forbes, M.-A. Villard, J. Gunn, "The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence," *Ecological Modelling*, vol. 191, no. 2, pp. 197–224, 2006.

[112] E. Fleishman, R. Mac Nally, "Patterns of spatial autocorrelation of assemblages of birds, floristics, physiognomy, and primary productivity in the central great basin, usa," *Diversity and Distributions*, vol. 12, no. 3, pp. 236–243, 2006.

[113] G. Carl, I. Kühn, "Analyzing spatial autocorrelation in species distributions using gaussian and logit models," *ecological modelling*, vol. 207, no. 2-4, pp. 159–170, 2007.

[114] C. F. Dormann, "Effects of incorporating spatial autocorrelation into the analysis of species distribution data," *Global ecology and biogeography*, vol. 16, no. 2, pp. 129–138, 2007.

[115] S. C. Goslee, "Behavior of vegetation sampling methods in the presence of spatial autocorrelation," *Plant Ecology*, vol. 187, no. 2, pp. 203–212, 2006.

[116] J. A. F. Diniz-Filho, T. Siqueira, A. A. Padial, T. F. Rangel, V. L. Landeiro, L. M. Bini, "Spatial autocorrelation analysis allows disentangling the balance between neutral and niche processes in metacommunities," *Oikos*, vol. 121, no. 2, pp. 201–210, 2012.

[117] B. J. Mattsson, E. F. Zipkin, B. Gardner, P. J. Blank, J. R. Sauer, J. A. Royle, “Explaining local-scale species distributions: Relative contributions of spatial autocorrelation and landscape heterogeneity for an avian assemblage,” *PLoS One*, vol. 8, no. 2, e55097, 2013.

[118] O. Hössjer, “Spatial autocorrelation for subdivided populations with invariant migration schemes,” *Methodology and Computing in Applied Probability*, vol. 16, no. 4, pp. 777–810, 2014.

[119] J. M. Northrup, M. B. Hooten, C. R. Anderson Jr, G. Wittemyer, “Practical guidance on characterizing availability in resource selection functions under a use–availability design,” *Ecology*, vol. 94, no. 7, pp. 1456–1463, 2013.

[120] M. H. Webb *et al.*, “Location matters: Using spatially explicit occupancy models to predict the distribution of the highly mobile, endangered swift parrot,” *Biological Conservation*, vol. 176, pp. 99–108, 2014.

[121] B. Gauffre *et al.*, “Spatial heterogeneity in landscape structure influences dispersal and genetic structure: Empirical evidence from a grasshopper in an agricultural landscape,” *Molecular ecology*, vol. 24, no. 8, pp. 1713–1728, 2015.

[122] V. Lecours, R. Devillers, D. C. Schneider, V. L. Lucieer, C. J. Brown, E. N. Edinger, “Spatial scale and geographic context in benthic habitat mapping: Review and future directions,” *Marine Ecology Progress Series*, vol. 535, pp. 259–284, 2015.

[123] A. J. Craig, S. Woods, J. D. Hoeksema, “Influences of host plant identity and disturbance on spatial structure and community composition of ectomycorrhizal fungi in a northern mississippi uplands ecosystem,” *Fungal Ecology*, vol. 24, pp. 7–14, 2016.

[124] V. Boucher-Lalonde, D. J. Currie, “Spatial autocorrelation can generate stronger correlations between range size and climatic niches than the biological signal—a demonstration using bird and mammal range maps,” *PloS one*, vol. 11, no. 11, e0166243, 2016.

[125] D. R. Roberts *et al.*, “Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure,” *Ecography*, vol. 40, no. 8, pp. 913–929, 2017.

[126] R. L. Czaplewski, R. M. Reich, W. A. Bechtold, “Spatial autocorrelation in growth of undisturbed natural pine stands across georgia,” *Forest Science*, vol. 40, no. 2, pp. 314–328, 1994.

[127] M. A. Wulder, J. C. White, N. C. Coops, T. Nelson, B. Boots, “Using local spatial autocorrelation to compare outputs from a forest growth model,” *ecological modelling*, vol. 209, no. 2-4, pp. 264–276, 2007.

[128] J. D. Adelman, B. E. Ewers, D. S. MacKay, “Use of temporal patterns in vapor pressure deficit to explain spatial autocorrelation dynamics in tree transpiration,” *Tree physiology*, vol. 28, no. 4, pp. 647–658, 2008.

[129] L. Zhang, Z. Ma, L. Guo, “An evaluation of spatial autocorrelation and heterogeneity in the residuals of six regression models,” *Forest Science*, vol. 55, no. 6, pp. 533–548, 2009.

[130] M. Rejou-Mechain *et al.*, "Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks," *Biogeosciences Discussions*, vol. 11, p. 5711, 2014.

[131] K. M. Irvine, A. I. Gitelman, J. A. Hoeting, "Spatial designs and properties of spatial correlation: Effects on covariance estimation," *Journal of agricultural, biological, and environmental statistics*, vol. 12, no. 4, pp. 450–469, 2007.

[132] J. Dlugosz, M. Kobierski, A. Piotrowska-Dlugosz, D. Gozdowski, "Field-scale spatial autocorrelation of some sodium and potassium forms in a luvisol humic horizon," *Journal of Elementology*, vol. 19, no. 2, 2014.

[133] S. D. Yun, B. M. Gramig, M. S. Delgado, R. J. Florax, "Does spatial correlation matter in econometric models of crop yield response and weather?" Tech. Rep., 2015.

[134] M. Dudzińska, "The study of spatial autocorrelation of the land consolidation in lubelskie voivodeship," in *Environmental Engineering. Proceedings of the International Conference on Environmental Engineering. ICEE*, Vilnius Gediminas Technical University, Department of Construction Economics . . ., vol. 10, 2017, pp. 1–7.

[135] A. G. d. L. Moraes, M. R. Francelino, W. d. Carvalho Junior, M. G. Pereira, A. Thomazini, C. E. G. R. Schaefer, "Environmental correlation and spatial autocorrelation of soil properties in keller peninsula, maritime antarctica," *Revista Brasileira de Ciência do Solo*, vol. 41, 2017.

[136] R. S. Mullen, K. W. Birkeland, "Mixed effect and spatial correlation models for analyzing a regional spatial dataset," in *Proceedings of the 2008 International Snow Science Workshop, Whistler, British Columbia*, Citeseer, 2008, pp. 421–425.

[137] V. Dakos, E. H. van Nes, R. Donangelo, H. Fort, M. Scheffer, "Spatial correlation as leading indicator of catastrophic shifts," *Theoretical Ecology*, vol. 3, no. 3, pp. 163–174, 2010.

[138] V. Gajović, B. Todorović, "Spatial and temporal analysis of fires in serbia for period 2000–2013," in *International Conference "Natural Hazards–Links between Science and Practice," J. Geogr. Inst. Jovan Cvijic*, vol. 63, 2013, pp. 297–312.

[139] T. Heckmann, K. Gegg, A. Gegg, M. Becht, "Sample size matters: Investigating the effect of sample size on a logistic regression susceptibility model for debris flows," *Natural Hazards and Earth System Sciences*, vol. 14, no. 2, p. 259, 2014.

[140] A. Mondini, "Measures of spatial autocorrelation changes in multitemporal sar images for event landslides detection," *Remote Sensing*, vol. 9, no. 6, p. 554, 2017.

[141] C.-M. Lee, D. P. Culhane, "A perimeter-based clustering index for measuring spatial segregation: A cognitive gis approach," *Environment and Planning B: Planning and Design*, vol. 25, no. 3, pp. 327–343, 1998.

[142] M. Saib, "Spatial autocorrelation in voting turnout," *Journal of Biometrics and Biostatistics*, vol. 8, no. 5, pp. 1–4, 2017.

[143] L. Anselin, J. Cohen, D. Cook, W. Gorr, G. Tita, “Spatial analyses of crime,” *Criminal justice*, vol. 4, no. 2, pp. 213–262, 2000.

[144] G. Mazzulla, C. Forciniti, “Spatial association techniques for analysing trip distribution in an urban area,” *European Transport Research Review*, vol. 4, no. 4, pp. 217–233, 2012.

[145] J. Gao, Y. Wei, W. Chen, K. Yenneti, “Urban land expansion and structural change in the yangtze river delta, china,” *Sustainability*, vol. 7, no. 8, pp. 10 281–10 307, 2015.

[146] E. Mulokozi, H. H. Teng, “Investigating the existence of second order spatial autocorrelation in crash frequency across adjacent freeway segments,” *Journal of transportation technologies*, vol. 6, no. 05, p. 286, 2016.

[147] M. Khalili, R. Leconte, F. Brissette, “Stochastic multisite generation of daily precipitation data using spatial autocorrelation,” *Journal of hydrometeorology*, vol. 8, no. 3, pp. 396–412, 2007.

[148] G. F. Ghalhari, A. D. Roudbari, “An investigation on thermal patterns in iran based on spatial autocorrelation,” *Theoretical and applied climatology*, vol. 131, no. 3-4, pp. 865–876, 2018.

[149] E. E. Peterson, A. A. Merton, D. M. Theobald, N. S. Urquhart, “Patterns of spatial autocorrelation in stream water chemistry,” *Environmental Monitoring and Assessment*, vol. 121, no. 1-3, pp. 571–596, 2006.

[150] W. Schröder, R. Pesch, H. Harmens, H. Fagerli, I. Ilyin, “Does spatial auto-correlation call for a revision of latest heavy metal and nitrogen deposition maps?” *Environmental Sciences Europe*, vol. 24, no. 1, p. 20, 2012.

[151] J. Grieve, “The use of spatial autocorrelation statistics for the analysis of regional linguistic variation,” 2011.

[152] N. Pettorelli, C. Duncan, H. J. Owen, S. T. Turvey, “Spatial autocorrelation and congruence in the distribution of language and mammal richness: A reply to cardillo et al.(2015),” *Proceedings of the Royal Society B: Biological Sciences*, vol. 282, no. 1809, p. 20 150 591, 2015.

[153] A. Torabi, H. Fossen, B. Alaei, “Application of spatial correlation functions in permeability estimation of deformation bands in porous rocks,” *Journal of Geophysical Research: Solid Earth*, vol. 113, no. B8, 2008.

[154] V. C. Tsai, M. P. Moschetti, “An explicit relationship between time-domain noise correlation and spatial autocorrelation (spac) results,” *Geophysical Journal International*, vol. 182, no. 1, pp. 454–460, 2010.

[155] H. Nakahara, “Formulation of the spatial autocorrelation (spac) method in dissipative media,” *Geophysical Journal International*, vol. 190, no. 3, pp. 1777–1783, 2012.

[156] A. P. Lamb, K. van Wijk, L. M. Liberty, T. D. Mikesell, “The spatial cross-correlation method for dispersive surface waves,” *Geophysical Journal International*, vol. 199, no. 1, pp. 1–10, 2014.

[157] N. Debarsy, C. Ertur, “Testing for spatial autocorrelation in a fixed effects panel data model,” *Regional Science and Urban Economics*, vol. 40, no. 6, pp. 453–470, 2010.

[158] A. Mitchell, *The esri guide to gis analysis (volume 2). redlands*, 2005.

[159] L. Anselin, I. Syabri, Y. Kho, “Geoda: An introduction to spatial data analysis,” *Geographical analysis*, vol. 38, no. 1, pp. 5–22, 2006.

[160] M. Witherick, S. R. Ross, R. J. Small, *A modern dictionary of geography*. Arnold London, 2001.

[161] D. Gregory, R. Johnston, G. Pratt, M. Watts, S. Whatmore, *The dictionary of human geography*. John Wiley & Sons, 2011.

[162] S. Mayhew, *A dictionary of geography*. Oxford quick reference, 2015.

[163] B. Warf, *Encyclopedia of human geography*. Sage, 2006.

[164] J. Snow, *On the mode of communication of cholera*. John Churchill, 1855.

[165] R. A. Fisher, “The wave of advance of advantageous genes,” *Annals of eugenics*, vol. 7, no. 4, pp. 355–369, 1937.

[166] E. Shevky, W. Bell, “Social area analysis; theory, illustrative application and computational procedures.,” 1955.

[167] M.-F. Cicéri, “Méthodes d’analyse multivariée dans la géographie anglo-saxonne, évaluation des techniques et des applications,” Ph.D. dissertation, 1974.

[168] W. Scott *et al.*, “British towns: A statistical study of their social and economic differences,” *Centre for Urban Studies (United Kingdom) eng no. 2*, 1961.

[169] B. J. L. Berry, F. E. Horton, J. O. Abiodun, *Geographic perspectives on urban systems: with integrated readings*. Prentice Hall, 1970.

[170] R. H. Smith, *City classification handbook: Methods and applications*, 1973.

[171] E. E. Holmes, M. A. Lewis, J. Banks, R. Veit, “Partial differential equations in ecology: Spatial interactions and population dynamics,” *Ecology*, vol. 75, no. 1, pp. 17–29, 1994.

[172] D. W. Lefever, “Measuring geographic concentration by means of the standard deviational ellipse,” *American Journal of Sociology*, vol. 32, no. 1, pp. 88–94, 1926.

[173] O. Sutton, “The theoretical distribution of airborne pollution from factory chimneys,” *Quarterly Journal of the Royal Meteorological Society*, vol. 73, no. 317-318, pp. 426–436, 1947.

[174] P. A. Samuelson, “Spatial price equilibrium and linear programming,” *The American economic review*, vol. 42, no. 3, pp. 283–303, 1952.

[175] J. Odland, R. Barff, “A statistical model for the development of spatial patterns: Applications to the spread of housing deterioration,” *Geographical Analysis*, vol. 14, no. 4, pp. 326–339, 1982.

[176] J. B. Birdsell, “Some implications of the genetical concept of race in terms of spatial analysis,” in *Cold Spring Harbor Symposia on Quantitative Biology*, Cold Spring Harbor Laboratory Press, vol. 15, 1950, pp. 259–314.

[177] J. S. Heywood, “Spatial analysis of genetic variation in plant populations,” *Annual Review of Ecology and Systematics*, vol. 22, no. 1, pp. 335–355, 1991.

[178] P. J. Clark, F. C. Evans, "Distance to nearest neighbor as a measure of spatial relationships in populations," *Ecology*, vol. 35, no. 4, pp. 445–453, 1954.

[179] I. Hodder, C. Orton, "Spatial analysis in archaeology," 1976.

[180] P. SJ, "The spatial analysis of magnetic fields in sunspots," *Monthly Notices of the Royal Astronomical Society*, vol. 120, no. 5, pp. 412–429, 1960.

[181] M.-J. Fortin, P. Drapeau, P. Legendre, "Spatial autocorrelation and sampling design in plant ecology," in *Progress in theoretical vegetation science*, Springer, 1990, pp. 209–222.

[182] N. A. Cressie, *Statistics for Spatial Data: Revised Edition*. John Wiley & Sons, 1993.

[183] M. M. Fischer, A. Getis, *Handbook of applied spatial analysis: software tools, methods and applications*. Springer Science & Business Media, 2009, pp. 255–27.

[184] T. Bailey, A. Gatrell, "Interactive spatial data analysis (vol. 413). essex: Longman scientific & technical," *Bartier, PM, & Keller, CP*, pp. 795–799, 1996.

[185] A. Getis, J. Franklin, "Second-order neighborhood analysis of mapped point patterns," *Ecology*, vol. 68, no. 3, pp. 473–477, 1987.

[186] D. G. Krige, "A statistical approach to some basic mine valuation problems on the witwatersrand," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 52, no. 6, pp. 119–139, 1951.

[187] G. Matheron, *Traité de géostatistique appliquée. 1 (1962)*. Editions Technip, 1962, vol. 1.

[188] D. W.-S. Wong, J. Lee, *Statistical analysis of geographic information with ArcView GIS and ArcGIS*, G 70.212. L43 2005. John Wiley & Sons Hoboken, NJ, 2005.

[189] P. Greig-Smith, "Pattern in vegetation," *Journal of Ecology*, vol. 67, no. 3, pp. 755–779, 1979.

[190] C. Amrhein, D. Griffith, "A model for statistical quality control of spatial data on gis," 1991.

[191] B. D. Ripley, "The second-order analysis of stationary point processes," *Journal of applied probability*, vol. 13, no. 2, pp. 255–266, 1976.

[192] J. Conolly, M. Lake, *Geographical information systems in archaeology*. Cambridge University Press, 2006.

[193] B. F. Manly, *Randomization, bootstrap and Monte Carlo methods in biology*. Chapman and Hall/CRC, 2018.

[194] N. A. Cressie, T. Read, "Do sudden infant deaths come in clusters?," 1985.

[195] N. Cressie, T. R. Read, "Spatial data analysis of regional counts," *Biometrical Journal*, vol. 31, no. 6, pp. 699–719, 1989.

[196] N. Cressie, N. H. Chan, "Spatial modeling of regional variables," *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 393–401, 1989.

[197] R. Haining, "A spatial model for [southwestern nebraska and northwestern kansas] high plains agriculture [corn and wheat yields, statistical estimation].," *Annals of the Association of American Geographers*, 1978.

[198] W. Mercer, A. Hall, "The experimental error of field trials," *The Journal of Agricultural Science*, vol. 4, no. 2, pp. 107–132, 1911.

[199] W. Federer, C. Schlottfeldt, "The use of covariance to control gradients in experiments," *Biometrics*, vol. 10, no. 2, pp. 282–290, 1954.

[200] B. Boots, M. Tiefelsdorf, "Global and local spatial autocorrelation in bounded regular tessellations," *Journal of Geographical Systems*, vol. 2, no. 4, pp. 319–348, 2000.

[201] D. A. Griffith, G. Arbia, "Detecting negative spatial autocorrelation in georeferenced random variables," *International Journal of Geographical Information Science*, vol. 24, no. 3, pp. 417–437, 2010.

[202] T. Dusek, "Spatially aggregated data and variables in empirical analysis and model building for economics," *Cybergeo: European Journal of Geography*, 2004.

[203] E. Halley, "An historical account of the trade winds, and monsoons, observable in the seas between and near the tropicks, with an attempt to assign the physical cause of the said winds," *Philosophical Transactions of the Royal Society of London*, vol. 16, no. 183, pp. 153–168, 1753.

[204] Student, "On the error of counting with a haemacytometer," *Biometrika*, pp. 351–360, 1907.

[205] J. Papadakis, "Méthode statistique pour des expériences sur champ," *Bull. Inst. Amel. Plantes a Salonique*, vol. 23, pp. 13–29, 1937.

[206] M. Bartlett, "The characteristic function of a conditional statistic," *Journal of the London Mathematical Society*, vol. 1, no. 1, pp. 62–67, 1938.

[207] M. S. Bartlett, *An introduction to stochastic processes: with special reference to methods and applications*. CUP Archive, 1978.

[208] R. A. Fisher, F. Yates, *et al.*, "Statistical tables for biological, agricultural aad medical research.," *Statistical tables for biological, agricultural aad medical research.*, 1938.

[209] F. Yates, W. Cochran, "The analysis of groups of experiments," *The Journal of Agricultural Science*, vol. 28, no. 4, pp. 556–580, 1938.

[210] H. F. Smith, "Report of a preliminary statistical investigation of flowering dates of plants recorded in the phenological reports of the royal meteorological society," *Quarterly Journal of the Royal Meteorological Society*, vol. 64, no. 273, pp. 23–46, 1938.

[211] ——, "An empirical law describing heterogeneity in the yields of agricultural crops," *The Journal of Agricultural Science*, vol. 28, no. 1, pp. 1–23, 1938.

[212] P. Whittle, "On stationary processes in the plane. biometrika 41 434–449," *Mathematical Reviews (MathSciNet): MR67450 Zentralblatt MATH*, vol. 58, pp. 41–3, 1954.

[213] E. Vanmarcke, "Random fields," *Random Fields, by Erik Vanmarcke, pp. 372. ISBN 0-262-72045-0. Cambridge, Massachusetts, USA: The MIT Press, March 1983.(Paper)*, p. 372, 1983.

[214] U. Mishra, W. J. Riley, “Alaskan soil carbon stocks: Spatial variability and dependence on environmental factors,” *Biogeosciences*, vol. 9, no. 9, pp. 3637–3645, 2012.

[215] M. A. Islam *et al.*, “Utilization of open source spatial data for landslide susceptibility mapping at chittagong district of bangladesh—an appraisal for disaster risk reduction and mitigation approach,” *International Journal of Geosciences*, vol. 8, no. 4, pp. 577–598, 2017.

[216] A. F. Emrehan, D. YILDIZ, M. ŞAMİL GÜNEŞ, “A spatio-temporal approach to national natural resources: The change of provincial water use over turkey.,” *Sigma: Journal of Engineering & Natural Sciences/Mühendislik ve Fen Bilimleri Dergisi*, vol. 36, no. 2, 2018.

[217] D. A. Griffith, “The boundary value problem in spatial statistical analysis,” *Journal of regional science*, vol. 23, no. 3, pp. 377–387, 1983.

[218] W. R. Tobler, “Cellular geography,” in *Philosophy in geography*, Springer, 1979, pp. 379–386.

[219] W. Tobler, “Application of image processing techniques to map processing,” in *International Symposium on Spatial Data Handling, Zurich, Switzerland*, 1984, pp. 140–145.

[220] A. Cliff, J. Ord, *Spatial processes: Models and applications (pion limited)*, 1981.

[221] J. L. Sibert, “Spatial autocorrelation and the optimal prediction of assessed values.,” 1975.

[222] W. Christaller, *Central places in southern Germany*. Prentice Hall, 1966.

[223] A. Gatrell, “The autocorrelation structure of central-place populations in southern germany,” *Statistical applications in the spatial sciences*, pp. 111–126, 1979.

[224] K. M. Çubukçu, *Planlamada ve coğrafyada temel istatistik ve mekansal istatistik*. Nobel, 2015.

[225] P. Haggett, A. D. Cliff, A. Frey, “Locational analysis in human geography.,” *Tijdschrift Voor Economische En Sociale Geografie*, vol. 68, no. 6, 1977.

[226] J. A. Santora, C. S. Reiss, V. J. Loeb, R. R. Veit, “Spatial association between hotspots of baleen whales and demographic patterns of antarctic krill euphausia superba suggests size-dependent predation,” *Marine Ecology Progress Series*, vol. 405, pp. 255–269, 2010.

[227] W. J. Fu, P. K. Jiang, G. M. Zhou, K. L. Zhao, “Using moran’s i and gis to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern china,” *Biogeosciences*, vol. 11, no. 8, pp. 2401–2409, 2014.

[228] S. Stucki *et al.*, “High performance computation of landscape genomic models including local indicators of spatial association,” *Molecular ecology resources*, vol. 17, no. 5, pp. 1072–1089, 2017.

[229] S. Getzin, T. Wiegand, S. P. Hubbell, “Stochastically driven adult–recruit associations of tree species on barro colorado island,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 281, no. 1790, p. 20 140 922, 2014.

[230] B. A. Miller, S. Koszinski, M. Wehrhan, M. Sommer, "Comparison of spatial association approaches for landscape mapping of soil organic carbon stocks," *Soil*, vol. 1, no. 1, p. 217, 2015.

[231] T. M. Anderson *et al.*, "The spatial distribution of african savannah herbivores: Species associations and habitat occupancy in a landscape context," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1703, p. 20 150 314, 2016.

[232] A. Appice, M. Ceci, A. Lanza, F. A. Lisi, D. Malerba, "Discovery of spatial association rules in geo-referenced census data: A relational mining approach," *Intelligent Data Analysis*, vol. 7, no. 6, pp. 541–566, 2003.

[233] K. J. McManus, D. A. Stephens, N. M. Adams, S. A. Islam, P. S. Freemont, M. J. Hendzel, "The transcriptional regulator cbp has defined spatial associations within interphase nuclei," *PLoS computational biology*, vol. 2, no. 10, e139, 2006.

[234] P. Lidji, R. Kolinsky, A. Lochy, J. Morais, "Spatial associations for musical stimuli: A piano in the head?" *Journal of Experimental Psychology: human perception and performance*, vol. 33, no. 5, p. 1189, 2007.

[235] T. Suzuki, H. Hashimoto, H. Ishihara, T. Kasai, H. Kunoh, J. Takada, "Structural and spatial associations between fe, o, and c in the network structure of the leptothrix ochracea sheath surface," *Appl. Environ. Microbiol.*, vol. 77, no. 21, pp. 7873–7875, 2011.

[236] H. Saiz, C. Alados, Y. Pueyo, "Plant–plant spatial association networks in gypsophilous communities: The influence of aridity and grazing and the role of gypsophytes in its structure," *Web Ecology*, vol. 14, no. 1, pp. 39–49, 2014.

[237] P. L. James *et al.*, "Metal worker's lung: Spatial association with mycobacterium avium," *Thorax*, vol. 73, no. 2, pp. 151–156, 2018.

[238] D. J. Unwin, L. W. Hepple, "The statistical analysis of spatial series," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 23, no. 3-4, pp. 211–227, 1974.

[239] W. Mitchell, A. Bill, *et al.*, "Spatial dependence in regional unemployment in australia," *Future That Works, A*, p. 312, 2004.

[240] J. C. Fox, H. Bi, P. K. Ades, "Modelling spatial dependence in an irregular natural forest," *Silva Fennica*, vol. 42, no. 1, p. 35, 2008.

[241] J. Miller, J. Franklin, R. Aspinall, "Incorporating spatial dependence in predictive vegetation models," *Ecological modelling*, vol. 202, no. 3-4, pp. 225–242, 2007.

[242] W. F. de Boer *et al.*, "Spatial distribution of lion kills determined by the water dependency of prey species," *Journal of Mammalogy*, vol. 91, no. 5, pp. 1280–1286, 2010.

[243] S. Özyurt, "Spatial dependence in commercial property prices: Micro evidence from the netherlands," 2014.

[244] U. Lim, D. Kim, "Toward sustainable economic growth: A spatial panel data analysis of regional income convergence in us bea economic areas," *Sustainability*, vol. 7, no. 8, pp. 9943–9959, 2015.

[245] M. A. Murphy, J. S. Evans, S. A. Cushman, A. Storfer, "Representing genetic variation as continuous surfaces: An approach for identifying spatial dependency in landscape genetic studies," *Ecography*, vol. 31, no. 6, pp. 685–697, 2008.

[246] M. Castro, R. Paleti, C. R. Bhat, "A spatial generalized ordered response model to examine highway crash injury severity," *Accident Analysis & Prevention*, vol. 52, pp. 188–203, 2013.

[247] H. Ming, L. Kuan-Pin, "Testing for spatial dependence in a two-way fixed effects panel data model," *Documento de trabajo Regional Research Institute, West Virginia University*, 2013.

[248] A. AghaKouchak, E. Habib, A. Bárdossy, "Modeling radar rainfall estimation uncertainties: Random error model," *Journal of Hydrologic Engineering*, vol. 15, no. 4, pp. 265–274, 2009.

[249] S. Lamsal, R. C. Cobb, J. H. Cushman, Q. Meng, D. M. Rizzo, R. K. Meentemeyer, "Spatial estimation of the density and carbon content of host populations for phytophthora ramorum in california and oregon," *Forest Ecology and Management*, vol. 262, no. 6, pp. 989–998, 2011.

[250] C. Gülser, I. Ekberli, F. Candemir, "Spatial variability of soil physical properties in a cultivated field," *Eurasian Journal of Soil Science*, vol. 5, no. 3, pp. 192–200, 2016.

[251] R. Delsol, M. Loreau, B. Haegeman, "The relationship between the spatial scaling of biodiversity and ecosystem stability," *Global ecology and biogeography*, vol. 27, no. 4, pp. 439–449, 2018.

[252] J. Pfeuffer, J. C. McCullough, P.-F. Van de Moortele, K. Ugurbil, X. Hu, "Spatial dependence of the nonlinear bold response at short stimulus duration," *Neuroimage*, vol. 18, no. 4, pp. 990–1000, 2003.

[253] F. J. Areal, K. Balcombe, R. Tiffin, "Integrating spatial dependence into stochastic frontier analysis," *Australian Journal of Agricultural and Resource Economics*, vol. 56, no. 4, pp. 521–541, 2012.

[254] D. Cooley *et al.*, "A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects," *Revstat*, vol. 10, no. 1, pp. 135–165, 2012.

[255] M. Kabir, M. Salam, D. Paul, M. Hossain, N. Rahman, M. Latif, "Geo-statistical models for determining spatial variation and spatial dependency of soil arsenic in bangladesh," *Journal of the National Science Foundation of Sri Lanka*, vol. 45, no. 2, 2017.

[256] R. R. Boyce, "Geography as spatial interaction," *Seattle: U. of Washington*, 1980.

[257] J. Qiu, M. G. Turner, "Spatial interactions among ecosystem services in an urbanizing agricultural watershed," *Proceedings of the National Academy of Sciences*, vol. 110, no. 29, pp. 12 149–12 154, 2013.

[258] C. Bhat, "A new spatial (social) interaction discrete choice model accommodating for unobserved effects due to endogenous network formation," *Transportation*, vol. 42, no. 5, pp. 879–914, 2015.

[259] H. N. aus Stadthagen, "Spatial and functional determinants of long-term fecundity in serotinous shrub communities," Ph.D. dissertation, Universität Hohenheim, 2016.

[260] J. van de Koppel, M. Rietkerk, "Spatial interactions and resilience in arid ecosystems," *The American Naturalist*, vol. 163, no. 1, pp. 113–121, 2004.

[261] P. Silveira, V. Silva, T. Dentinho, "Spatial interaction model with land and water use: An application to terceira island," in *Proceedings 45th ISOCARP Congress*, vol. 17, 2009.

[262] H. Sjödin, "Population-level consequences of spatial interactions," Ph.D. dissertation, Umeå universitet, 2013.

[263] D. Leithinger, S. Follmer, A. Olwal, H. Ishii, "Shape displays: Spatial interaction with dynamic physical form," *IEEE computer graphics and applications*, vol. 35, no. 5, pp. 5–11, 2015.

[264] J. Yin, A. Soliman, D. Yin, S. Wang, "Depicting urban boundaries from a mobility network of spatial interactions: A case study of great britain with geo-located twitter data," *International Journal of Geographical Information Science*, vol. 31, no. 7, pp. 1293–1313, 2017.

[265] R. J. Franzese, J. C. Hays, "Empirical models of spatial interdependence," 2008.

[266] C. Guevara *et al.*, "Growth agglomeration effects in spatially interdependent latin american regions," *Journal of Applied Economic Sciences (JAES)*, vol. 11, no. 45, pp. 1350–1355, 2016.

[267] C. Xu, C. Wang, W. Wang, J. Bao, M. Yang, "Investigating spatial interdependence in e-bike choice using spatially autoregressive model," *Promet-Traffic&Transportation*, vol. 29, no. 4, pp. 351–362, 2017.

[268] S. J. Cook, J. C. Hays, R. J. Franzese, "Model specification and spatial interdependence," in *Annual Meeting of the American Political Science Association, San Francisco, CA*, 2015.

[269] D. Van den Poel, P. Baecke, "The value of neighborhood information in prospect selection models investigating the optimal level of granularity," 2013.

[270] M. Hanel, P. Máca, "Spatial variability and interdependence of rain event characteristics in the czech republic," *Hydrological processes*, vol. 28, no. 6, pp. 2929–2944, 2014.

[271] G. J. Hewings, Y. Okuyama, M. Sonis, "Economic interdependence within the chicago metropolitan area: A miyazawa analysis," *Journal of Regional Science*, vol. 41, no. 2, pp. 195–217, 2001.

[272] H. Jiang, N. Serban, "Clustering random curves under spatial interdependence with application to service accessibility," *Technometrics*, vol. 54, no. 2, pp. 108–119, 2012.

[273] A. Niedermayr, M. Kapfer, J. Kantelhardt, "Using econometric models to analyse the spatial distribution of oil pumpkin cultivation in austria," Tech. Rep., 2016.

[274] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic geography*, vol. 46, no. sup1, pp. 234–240, 1970.

[275] A. D. Cliff, J. K. Ord, "Spatial and temporal analysis: Autocorrelation in space and time," *Quantitative geography: a British view*, vol. 1, pp. 104–10, 1981.

[276] B. Epperson, "Spatial autocorrelation of genotypes under directional selection.," *Genetics*, vol. 124, no. 3, pp. 757–771, 1990.

[277] A. Cliff, J. Ord, "Model building and the analysis of spatial pattern in human geography," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 37, no. 3, pp. 297–328, 1975.

[278] D. W. Matula, R. R. Sokal, "Properties of gabriel graphs relevant to geographic variation research and the clustering of points in the plane," *Geographical analysis*, vol. 12, no. 3, pp. 205–222, 1980.

[279] H. Adesina, "Identification of the cholera diffusion process in ibadan, 1971," *Social Science & Medicine*, vol. 18, no. 5, pp. 429–440, 1984.

[280] R. R. Sokal, P. Menozzi, "Spatial autocorrelations of hla frequencies in europe support demic diffusion of early farmers," *The American Naturalist*, vol. 119, no. 1, pp. 1–17, 1982.

[281] L. Anselin, *Spatial externalities*, 2003.

[282] S.-I. Lee, "Developing a bivariate spatial association measure: An integration of pearson's r and moran's i," *Journal of geographical systems*, vol. 3, no. 4, pp. 369–385, 2001.

[283] G. B. Junior, A. C. de Paiva, A. C. Silva, A. C. M. de Oliveira, "Classification of breast tissues using moran's index and geary's coefficient as texture signatures and svm," *Computers in Biology and Medicine*, vol. 39, no. 12, pp. 1063–1072, 2009.

[284] C. M. Beale, J. J. Lennon, J. M. Yearsley, M. J. Brewer, D. A. Elston, "Regression analysis of spatial data," *Ecology letters*, vol. 13, no. 2, pp. 246–264, 2010.

[285] M. D. Partridge, M. Boarnet, S. Brakman, G. Ottaviano, "Introduction: Whither spatial econometrics?" *Journal of Regional Science*, vol. 52, no. 2, pp. 167–171, 2012.

[286] D. P. McMillen, "Issues in spatial data analysis," *Journal of Regional Science*, vol. 50, no. 1, pp. 119–141, 2010.

[287] J. Pinkse, M. E. Slade, "The future of spatial econometrics," *Journal of Regional Science*, vol. 50, no. 1, pp. 103–117, 2010.

[288] L. Corrado, B. Fingleton, "Where is the economics in spatial econometrics?" *Journal of Regional Science*, vol. 52, no. 2, pp. 210–239, 2012.

[289] D. A. Griffith, F. Csillag, "Exploring relationships between semi-variogram and spatial autoregressive models," *Papers in Regional Science*, vol. 72, no. 3, pp. 283–295, 1993.

[290] M. Tiefelsdorf, *Modelling spatial processes: the identification and analysis of spatial relationships in regression residuals by means of Moran's I*. Springer, 2006, vol. 87.

[291] J. P. Elhorst, "Applied spatial econometrics: Raising the bar," *Spatial economic analysis*, vol. 5, no. 1, pp. 9–28, 2010.

[292] T. De Graaff, R. J. Florax, P. Nijkamp, A. Reggiani, “A general misspecification test for spatial regression models: Dependence, heterogeneity, and nonlinearity,” *Journal of Regional Science*, vol. 41, no. 2, pp. 255–276, 2001.

[293] S. Shekhar, H. Xiong, *Encyclopedia of GIS*. Springer Science & Business Media, 2007.

[294] C. Brunsdon, S. Fotheringham, M. Charlton, “Geographically weighted regression,” *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 47, no. 3, pp. 431–443, 1998.

[295] C. Brunsdon, A. Fotheringham, M. Charlton, “Geographically weighted summary statistics—a framework for localised exploratory data analysis,” *Computers, Environment and Urban Systems*, vol. 26, no. 6, pp. 501–524, 2002.

[296] A. S. Fotheringham, C. Brunsdon, M. Charlton, *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2003.

[297] M. Charlton, S. Fotheringham, C. Brunsdon, “Geographically weighted regression,” *White paper. National Centre for Geocomputation. National University of Ireland Maynooth*, 2009.

[298] M. S. Rosenberg, C. D. Anderson, “Passage: Pattern analysis, spatial statistics and geographic exegesis. version 2,” *Methods in Ecology and Evolution*, vol. 2, no. 3, pp. 229–232, 2011.

[299] G. BRATIC, “Validation of the high-resolution global land cover maps in lombardy,” 2018.

[300] L. J. Hubert, R. G. Golledge, C. M. Costanzo, “Generalized procedures for evaluating spatial autocorrelation,” *Geographical Analysis*, vol. 13, no. 3, pp. 224–233, 1981.

[301] L. Anselin *et al.*, “Estimation methods for spatial autoregressive structures.,” *Estimation methods for spatial autoregressive structures.*, no. 8, 1980.

[302] R. D. Cook, “Detection of influential observation in linear regression,” *Technometrics*, vol. 19, no. 1, pp. 15–18, 1977.

[303] D. C. Hoaglin, R. E. Welsch, “The hat matrix in regression and anova,” *The American Statistician*, vol. 32, no. 1, pp. 17–22, 1978.

[304] K. Belsley, E. Kuh, “Welsch (1980). regression diagnostics,” *New York, NY: WileyBelsleyRegression Diagnostics1980*, 1993.

[305] O. Kaya, E. Toroglu, F. Adiguzel, “The spatial analysis of the parties voting rate on the district scale at the general election in 2011,” *JOURNAL OF GEOGRAPHY-COGRAFYA DERGISI*, no. 31, pp. 1–13, 2015.

[306] J. Von Neumann, R. Kent, H. Bellinson, B. t. Hart, “The mean square successive difference,” *The Annals of Mathematical Statistics*, vol. 12, no. 2, pp. 153–162, 1941.

[307] K. Nakhapakorn, S. Jirakajohnkool, “Temporal and spatial autocorrelation statistics of dengue fever.,” 2006.

[308] M. B. Pietrzak, J. Wilk, T. Kossowski, R. Bivand, “The identification of spatial dependence in the analysis of regional economic development-join-count test application,” in *Proceedings of the 8th professor Aleksander Zelias international conference on modelling and forecasting of socio-economic phenomena*, Foundation of the Cracow University of Economics, Cracow, 2014, pp. 135–144.

[309] M. S. Rosenberg, R. R. Sokal, N. L. Oden, D. DiGiovanni, “Spatial autocorrelation of cancer in western europe,” *European Journal of Epidemiology*, vol. 15, no. 1, pp. 15–22, 1999.

[310] M. Tiefelsdorf, “Some practical applications of moran’s i’s exact conditional distribution,” *Papers in Regional Science*, vol. 77, no. 2, pp. 101–129, 1998.

[311] L. Anselin. “Global spatial autocorrelation (1).” (2018), [Online]. Available: `https://geodacenter.github.io/workbook/5a_global_auto/lab5a.html` (visited on 11/18/2019).

[312] ——, “Local spatial autocorrelation (1).” (2019), [Online]. Available: `https://geodacenter.github.io/workbook/5a_global_auto/lab5a.html` (visited on 11/18/2019).

[313] L. H. G. documentation center. “Local spatial statistics.” (2018), [Online]. Available: `https://www.harrisgeospatial.com/docs/LocalSpatialStatistics.html` (visited on 10/17/2019).

[314] D. O’SULLIVAN, D. Unwin, *Geographic information analysis wiley: Hoboken*, 2003.

[315] M. Craglia, R. Haining, P. Wiles, “A comparative evaluation of approaches to urban crime pattern analysis,” *Urban Studies*, vol. 37, no. 4, pp. 711–729, 2000.

[316] J. Le Gallo, C. Ertur, “Exploratory spatial data analysis of the distribution of regional per capita gdp in europe, 1980–1995,” *Papers in regional science*, vol. 82, no. 2, pp. 175–201, 2003.

[317] M.-J. Fortin, G. M. Jacquez, “Randomization tests and spatially auto-correlated data,” *Bulletin of the Ecological Society of America*, vol. 81, no. 3, pp. 201–205, 2000.

[318] S. T. Dr. Luc Anselin. “Randomization.” (2022), [Online]. Available: `https://geodacenter.github.io/glossary.html#r` (visited on 05/11/2022).

[319] B. D. Ripley, *Spatial statistics*. John Wiley & Sons, 2005, pp. 111–112.

[320] L. Salmaso, F. Pesarin, *Permutation tests for complex data: theory, applications and software*. John Wiley & Sons, 2010.

[321] L. Anselin, “A local indicator of multivariate spatial association: Extending geary’s c,” *Geographical Analysis*, vol. 51, no. 2, pp. 133–150, 2019.

[322] S. T. Dr. Luc Anselin. “Pseudo-p.” (2022), [Online]. Available: `https://geodacenter.github.io/glossary.html#p` (visited on 05/11/2022).

[323] S. J. Rey, “Spatial empirics for economic growth and convergence,” *Geographical analysis*, vol. 33, no. 3, pp. 195–214, 2001.

[324] E. Mitchell Andy. “Fdr.” (2022), [Online]. Available: https://desktop.arcgis.com/en/arcmap/10.5/tools/spatial-statistics-toolbox/what-is-a-z-score-what-is-a-p-value.htm (visited on 05/11/2022).

[325] J. s. Ord, “Spatial autocorrelation: A statistician’s reflections,” in *Perspectives on Spatial Data Analysis*, Springer, 2010, pp. 165–180.

[326] P. A. Wall, O. J. Devine, “Interactive analysis of the spatial distribution of disease using a geographic information systems,” *Journal of Geographical Systems*, vol. 2, no. 3, pp. 243–256, 2000.

[327] L. Anselin, X. Li, J. Koschinsky, “Geoda, from the desktop to an ecosystem for exploring spatial data,” *Geographical Analysis*, 2021.

[328] S. T. Dr. Luc Anselin. “Median local moran.” (2022), [Online]. Available: https://geodacenter.github.io/workbook/6b_local_adv/lab6b.html (visited on 05/20/2022).

[329] U. N. S. Division. “Select all countries, "gdp, per capita gdp - us dollars", and "2020" to generate the table.” (2022), [Online]. Available: https://unstats.un.org/unsd/snaama/Basic (visited on 05/20/2022).

[330] D. R. Pleydell, S. Chrétien, “Mixtures of gams for habitat suitability analysis with overdispersed presence/absence data,” *Computational statistics & data analysis*, vol. 54, no. 5, pp. 1405–1418, 2010.

[331] K. G. Logan, J. D. Nelson, C. Osbeck, J. D. Chapman, A. Hastings, “The application of travel demand management initiatives within a university setting,” *Case Studies on Transport Policy*, vol. 8, no. 4, pp. 1426–1439, 2020.

[332] J. H. Kim, W. B. Powell, R. A. Collado, “Quantile optimization for heavy-tailed distribution using asymmetric signum functions,” *Princeton University*, 2011.

[333] A. A. Rosenberg, S. Vedula, Y. Romano, A. M. Bronstein, “Fast nonlinear vector quantile regression,” *arXiv preprint arXiv:2205.14977*, 2022.

[334] D. P. Huttenlocher, R. H. Lilien, C. F. Olson, “View-based recognition using an eigenspace approximation to the hausdorff measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 951–955, 1999.

[335] T. S. Institute. “Provincal data of turkey.” (2022), [Online]. Available: https://data.tuik.gov.tr/Search/Search?text=illere%20g%C3%B6re (visited on 06/13/2022).

PUBLICATIONS FROM THE THESIS

Contact Information: matahmet@gmail.com

Papers

1. Emrehan, A. F., YILDIZ, D., GÜNEŞ, M. Ş. (2018). A spatio-temporal approach to national natural resources: the change of provincial water use over Turkey. Sigma Journal of Engineering and Natural Sciences, 36(2), 539-551. Accepted: 25.05.2018

Conference Papers

1. A Spatial Approach To The Relationship Between Water Consumption And Provincial Management Capacity , 10. International Statistics Days Conference, Giresun University, 07.11.2016

2. An Infrastructural Approach to Spatial Autocorrelation, Ankara 10. International Statistics Congress, Turkish Statistics Association, 06.11.2017