ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

EXPLORING SPATIAL PATTERNS AND HOTSPOTS OF HEPATITIS A AND AMOEBIC DYSENTERY USING GIS AND GEOSTATISTICAL ANALYSIS IN TURKEY

M.Sc. THESIS

Ruusa-Magano DAVID

Department of Applied Informatics

Geographical Information Technologies Programme

Thesis Advisor: Assist.Prof. Dr. A. Özgür DOĞRU

MAY 2017

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

EXPLORING SPATIAL PATTERNS AND HOTSPOTS OF HEPATITIS A AND AMOEBIC DYSENTERY USING GIS AND GEOSTATISTICAL ANALYSIS IN TURKEY

M.Sc. THESIS

Ruusa-Magano DAVID
(706141019)

Department of Applied Informatics

Geographical Information Technologies Programme

Thesis Advisor: Assist.Prof. Dr. A.Özgür DOĞRU

MAY 2017

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKİYE'DE AMİPLİ DİZANTERİ VE HEPATİT A HASTALIKLARININ MEKANSAL DOKU VE SICAK NOKTALARININ CBS VE GEOİSTATİSTİK ANALİZLER İLE İNCELENMESİ

YÜKSEK LİSANS TEZİ

Ruusa-Magano DAVID
(706141019)

Bilişim Uygulamaları Anabilim Dalı

Coğrafi Bilgi Teknolojileri Programı

Tez Danışmanı: Yrd.Doç. Dr. A Özgür DOĞRU

MAYIS 2017

Ruusa-Magano DAVID, a M.Sc. student of İTU Graduate School of Science Engineering and Technology student ID 706141019, successfully defended the thesis/dissertation entitled "EXPLORING SPATIAL PATTERNS AND HOTSPOTS OF HEPATITIS A AND AMOEBIC DYSENTERY USING GIS AND GEOSTATISTICAL ANALYSIS IN TURKEY", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assist. Prof. Dr. Ahmet Özgür DOĞRU** ISTANBUL Technical University	.....
Jury Members :	**Prof.Dr. Dursun Zafer ŞEKER** ISTANBUL Technical University	.....
	Prof. Dr. Işıl BAVUNOĞLU ISTANBUL University	.....

Date of Submission : 05 May 2017
Date of Defense : 08 June 2017

To my Dad and Mom,

FOREWORD

I thank The Almighty Good Lord for his abundant grace, love and mercy upon my life. His love never fails and makes the impossible possible.

I have been fortunate to work and learn under the supervision of Assist.Prof.Dr. Ahmet Özgur Doğru, a dedicated and remarkable mentor, who provided me with data and equipment necessary in my research. His extensive support, patience, wisdom and knowledge has inspired me and guided me throughout this thesis work. Moreover, his vast experience in research and feedback during this study have made me better equipped in research than ever. I am so much humbled by his willingness to assist even in times he is busy. Once again, I thank him very much for translating my abstract into Turkish. I say thank you very much.

I am also grateful to Assoc.Prof.Dr. Çigdem Göksel for her support, patience and for always being so accommodating, I have learned a whole lot from her and I am sincerely grateful.

I would like to extend my gratitude to Kingsley Eloka Ngwu for always pushing me and motivating me to work as hard as I can. Lastly, I also would like to thank my department members, colleagues and friends for their help, contribution and involvement.

May 2017 Ruusa-Magano DAVID

TABLE OF CONTENTS

Page

ABBREVIATIONS

GIS	**:** Geographic Information Systems
HAV	**:** Hepatitis A
AD	**:** Amoebic Dysentery
SA	**:** Spatial Autocorrelation
LISA	**:** Local Indicators of Spatial Association
SPSS	**:** Statistical Package for the Social Sciences
EDA	**:** Exploratory Data Analysis
EBS	**:** Empirical Bayes Smoothing
WHO	**:** World Health Organization

SYMBOLS

HAV_i	**:** Hepatitis A incidence rate
AD_i	**:** Amoebic Dysentery incidence rate
p_i	**:** Total population at risk
o_i	**:** Observed incidence rate
N_t	**:** The total population at the end of the study area
N_o	**:** The total population at the beginning of the study area
P_t	**:** Population growth rate as a percentage
t	**:** The number of years of the study area
I	**:** Global Moran's I
n	**:** The number of provinces
S_o	**:** The aggregate of all spatial weights
w_{ij}	**:** The elememnt of the spatial weight matrix
x_i	**:** Observation for province i
x_j	**:** Observation for province j
$\bar{x}$	**:** The mean of xi
E(I)	**:** Expected value of province i
I_i	**:** Local Moran's I
$G_{i*(d)}$	**:** Local Getis-Ord Gi*(d)
w_i	**:** Sum of the weight wij
s^2	**:** Sample mean and variance

LIST OF TABLES

Page

LIST OF FIGURES

Page

EXPLORING SPATIAL PATTERNS AND HOTSPOTS OF HEPATITIS A AND AMOEBIC DYSENTERY USING GIS AND GEOSTATISTICAL ANALYSIS IN TURKEY

SUMMARY

Epidemics of Infectious diseases have been documented throughout history and recent global health reports show a continual vulnerability of a large number of people to infectious diseases such as Hepatitis A (HAV) and Amoebic Dysentery (A Dysentery or AD) in relation to their environment. Of particular concern is the epidemic outbreaks patterns of Hepatitis A and A Dysentery incidences in terms of their geographic distributions. Accurate statistics on the geographical distribution of different endemicities of Hepatitis A and A Dysentery, on the populations at risk, and on the implications of given levels of endemicity for morbidity and mortality are important for effective control of these diseases. These estimates can be obtained using Geographical Information Systems (GIS) in conjunction with geostatistical methods available to examine patterns of disease incidence, morbidity and mortality in space and time and to relate those patterns to the identification of potential causes of disease, such as environmental exposure, diet and unhealthy behaviours, economic or socio-demographic factors.

In this sense, spatial epidemiology, one of the tools for improving public health, is increasingly being used to assess health risks associated with environmental hazards. In most cases, epidemiological analyses are based on observations of disease occurrence in a population of people 'at risk'. Typically, we want to relate occurrence patterns between collections of people experiencing different levels of exposure to some factor having a putative impact on a person's risk of disease. In recent years, the research in environmental health issues is constantly becoming more and more effective owing to the use of both various information technology services and software. In this context, it's no doubt that the use of GIS and Geostatistical methods helps in identification and quantification of patterns in disease occurrences providing the first steps toward increased understanding and possibly, control of the diseases.

The thesis focused on providing an insight into the geographic distribution of Hepatitis A and A Dysentery incidence in terms of their temporal distribution, spatial patterns, hotspots and clusters identification in 81 provinces of Turkey. Hepatitis and Dysentery have been a public health burden in Turkey. These diseases are a significant cause of morbidity, even if the mortality rate is low. The endemicity of this infections varies due to sanitary and hygiene conditions and socio-economic differences among the countries and in various regions of the same country. Turkey is a middle endemic area with respect to Hepatitis A virus and A Dysentery infection. However, because of the geographical, economical, environmental and cultural differences among the provinces of Turkey, the frequency of this infection also varies due to sanitary / hygiene conditions and socio-economic differences in various regions of the country.

Within the scope of the study, first, spatial epidemiology was described and reviewed, different approaches in disease modeling were provided. Hepatitis A and A Dysentery

transmission and their burdens in Turkey were discussed. The benefits of disease mapping and smoothing techniques in spatial epidemiology were pointed out and the distinction between GIS and geostatistics were outlined. This research project made use of data obtained from the EnviroGRIDS project supported by the European Union within the 7th Framework Programme. In this context, the data for patients with Hepatitis A and A Dysentery at province level for 2001-2011 obtained from Turkish Ministry of Health and the census data for 2000-2011 provided by the State Statistical Institute of Turkey were used as a basis for further analysis. Visual data mining was performed using Boxplots which give general insights of HAV and AD morbidity rate and time series thematic maps which helped us identify space-time disparities. The global Moran's statistic demonstrated that there was clustering in the year 2005, 2006 and 2007 for Hepatitis A in Turkish population as a whole, 2002, 2003, 2006 and 2007 for Hepatitis A in Turkish children below 15 years and in 2006 and 2007 for A Dysentery.

After a general statistical analysis of the LISAs techniques which measured HAV and AD infections at the local level it was found that there is a decreasing trend in the morbidity rate of Hepatitis A and A Dysentery at epidemic level from 2001 to 2011. The LISA statistic showed that majority of the HAV and AD clusters and hotspots were relatively located in the Black Sea and Central Anatolia Region for all the years, covering the Kizilirmak, Yesilmak, Sevhan River Basin and Western Black Sea River Basins in Turkey. However, from 2006-2008, hotspots started to return slowly to the western part of Turkey and spreading to the Eastern Anatolia Region. The western provinces with higher HAV and AD morbidity rate could be found in the Marmara, Menderes, Western Black Sea and Gediz River Basins, while on the Eastern Anatolia Region high HAV and AD morbidity rate were observed in the Aras, Firat-Dicle and Van Golu River Basins. From 2009-2011, the hotspots and clusters attenuated significantly in HAV and AD infections. Also, it was noted that children under the age of 15 years were more infected with Hepatitis A. Specifically, age-group-5-9 revealed the highest HAV morbidity rate than the other age groups. Our approach of identifying and quantifying of space-time disparities, clustering, and hot spots provided useful and detailed information for guiding policy formulation to succinctly address the burden of Hepatitis A and A Dysentery and possibly, curtail the spread of these particular diseases in the Turkish population.

TÜRKİYE'DE AMİPLİ DİZANTERİ VE HEPATİT A HASTALIKLARININ MEKANSAL DOKU VE SICAK NOKTALARININ CBS VE GEOİSTATİSTİK ANALİZLER İLE İNCELENMESİ

ÖZET

Bulaşıcı hastalıklardan kaynaklanan salgınlar tarih boyunca belgelenmiştir ve güncel global sağlık raporları insanların çevresel etkileşimden kaynaklı Hepatit A ve Amipli Dizanteri (A Dizanteri) gibi bulaşıcı hastalıklardan etkilenmeye devam ettiğini göstermektedir. Bulaşıcı hastalıkların izlenmesi ve gerekli önlemlerin alınması hem tarihsel verilerin etkin bir şekilde değerlendirilmesi hem de söz konusu hastalıkların mekansal yayılımlarının detaylı olarak incelenmesi ile mümkün olmaktadır. 1854 yılında Londra'da ortaya çıkan kolera salgınından günümüze bu tür çalışmalarda haritalar etkin olarak kullanılmaktadır. Londra'da yaşanan salgında, salgının kaynağı, Dr. John Snow tarafından, haritalar aracılığı ile yapılan mekansal analizler kullanılarak belirlenmiş ve hastalığın daha fazla yayılması bu şekilde önlenmiştir. G ünümüzde benzer analizler bilgisayar ve bilişim teknolojilerinin de gelişimin ve konum verisinin daha hızlı ve etkin üretilmesinin sağladığı imkanlar ile çok daha etkin olarak gerçekleştirilebilmektedir. Bu kapsamda Coğrafi Bilgi Sistemleri (CBS) teknolojilerinin sağladığı olanaklar göz ardı edilemeyecek derecede önemlidir.

Bu çalışmanın konusu Hepatit A ve A Dizanteri hastalıklarının insidanslarının mekansal dağılımlarına bağlı coğrafi dokularını incelemektir. Hepatit A ve A Dizateri gibi hastalıkların farklı endemik ya da epidemik süreçlerine ilişkin coğrafi dağılımlarına dair doğru ve yeterli istatistiklerin olması bu hastalıkların bulaşıcılık, ölüm, olgu ya da yayılım hızlarına bağlı olarak etkin bir şekilde izlenmesi ve kontrol edilmesine olanak sağlamaktadır. Bu tür değerlendirmeler günümüzde Coğrafi Bilgi Sistemleri (CBS) ve geoistatistik/mekansal istatistik yöntemlerinin etkin kullanımları ile gerçekleştirilebilmektedir. Söz konusu yöntemler kullanılarak hastalıkların zaman ve mekandaki yayılımları ve bu yayılımların oluşturdukları mekansal dokular incelenebilmekte, hastalıkların sebepleri, etkenleri ve farklı sosyal çevrelerdeki hastalık davranışları ortaya konulabilmektedir.

Bu bağlamda, halk sağlığının geliştirilmesi için bir araç olan mekansal epidemioloji, çevresel etkilere bağlı sağlık risklerinin değerlendirilmesinde artarak kullanılmaktadır. Çoğunlukla epidemiyolojik analizler risk altındaki nüfusta görülen olguların gözlemlenmesine dayanmaktadır. Tipik olarak olguların dokuları farklı seviyelerede hastalığa maruz kalan hastalar ile kişilerin hastalanmasına etkisi olduğu varsyaılan faktörler ile ilişkilendirilmeye çalışılır. Son yıllarda bu tür çalışmalar gelişen bilgi ve mekansal bilişim teknolojilerinin kullanımı ile çok daha verimli ve güvenilir sonuçlar vermektedir. Bu kapsamda CBS ve geoistatistik yöntemlerin hastalıkların mekansal dokularının belirlenmesi, ölçümü ile bu hastalıkların doğalarının anlaşılmasına önemli katkılar koydukları şüphesizdir.

Bu tezde Türkiye'deki il bazlı Hepatit A ve A Dizanteri olgularının zaman ve mekandaki değişimleri; mekansal dağılım, mekansal doku, sıcak noktalar ve kümelenmeler bağlamlarında incelenerek, hastalıkların karakterlerinin anlaşılması

amaçlanmıştır. Bu hastalıklar, bir çok ülkede olduğu gibi, ölüm riski az olmasına rağmen görülme sıkılıkları ile Türkiye için de bir halk sağlığı problemidir. Hastalıkların endemisitesi temizlik ve hijyen koşullarındaki olumsuzluklardan bölgeler arası sosyo-ekonomik farklılıkların neden olduğu değişimlere kadar bir çok duruma bağlı olmakla birlikte Türkiye Hepatit A Virüsü (HAV) ve A Dizanteri enfeksiyonları bakımından orta endemik bir bölge olarak tanımlanabilmektedir. İl bazlı coğrafi, ekonomik, çevresel ve kültürel farklılıklar nedeniyle, bu hastalıkların görülme sıklıkları da ülke çapında değişiklikler göstermektedir.

Bu tez kapsamında öncelikle mekansal epidemiyoloji konusu tanımlanarak farklı yönlerden incelenmiş ve Türkiye'de Hepatit A ve A Dizanteri bulaşlarının etkileri değerlendirilmiştir. Hastalıkların haritalandırılması, verilerin yumuşatma tekniklerinin faydalarından bahsedilerek, CBS, mekansal istatistik analizler ve geoistatistik konuları hakkında bilgiler verilmiştir. Bu amaçla Avrupa Birliği 7. Çerçeve Programı kapsamında desteklenmiş olan EnviroGRIDS projesi verileri kullanılmıştır. T.C. Sağlık Bakanlığı tarafından resmi olarak sağlanan söz konusu veriler 2000 – 2011 yıllarını kapsayan Hepatit A ve A Dizanteri verilerini içermektedir. Bu verilere ek olarak Türkiye İstatistik Kurumu'ndan elde edilen il bazlı nüfus bilgileri proje kapsamında hastalık insidanslarının hesaplarında kullanılmıştır. Veri düzenlemesi sürecinde ilk aşamada A Dizanteri ve Hepatit A hastalıklarına ilişkin vaka sayı verileri nüfus verileri ile oranlanarak hastalıkların görülme hızları hesaplanmıştır. Veri düzensizliklerinden kaynaklı istatistik çalışmalarada çıkabilecek olası sorunları ortadan kaldırmak için hesaplanan görülme hızı verileri Empric Bayes Yumuşatması (Emprical Bayesian Smoothing – EBS) yöntemi kullanılarak yumuşatılmıştır. Daha sonra elde edilen yumuşatılmış insidans verileri belirli bir nüfus bazında ifade edilerek morbiditeler belirlenmiştir.

Bu aşamadan sonra düzenlenmiş veriler görsel ve istatistik analizlerden geçirilerek değerlendirmeler yapılmıştır. Bu kapsamda kutu diyagramları kullanılarak hastalıkların görülme hızları zamansal olarak incelenmiş, zaman serileri haritaları ile zaman mekansal değerlendirmeleri yapılmıştır Kutu diyagramı analizleri sonucunda Hepatit A ve A Dizanteri bulaşlarında 2009 – 2011 yılları arasında kayda değer bir düşüş görüldüğü net bir şekilde anlaşılmaktadır. Hastalıkların yayılımlarının mekansal olarak izlenmesi için ise zaman serileri haritaları yöntemi oluşturulmuştur. Zaman serileri hartiaları; aynı konunun aynı mekanda farklı zaman birimlerindeki (günlük, aylık, yıllık, mevsimlik, dönemlik, ve benzeri) değerleri dikkate alınarak tasarlanan tematik haritalardır. İşaretleştirme ve görselleştirme aşamasında etkin olarak kullanılan sınıflandırma teknikleri ile olayların ve olguların zamansal ve mekansal dağılımlarının statik haritalar ile analizinde etkin olarak kullanılmaktadırlar. Çalışma kapsamında üretilen A Dizanteri ve Hepatit A zaman serileri haritalarında morbidite değerleri endemik, hiperendemik ve epidemik olamk üzere üç temel grupta sınıflandırılmıştır. Gerçekleştirilen analizler çalışma bölgesindeki konu eidlen hastalıklar bakımında yüksek ve düşük riskli illerin belirlenmesinde kullanılmıştır. Zaman serileri haritalarında çalışma süresince her iki hastalığında görülmesinde azalan bir trend olduğu anlaşımıştır.

Çalışma kapsamında ayrıca lokal Moran I (LISA) ve local $G_i^*(d)$ istatistik yöntemleri ile hastalıkların istatistik ve mekansal kümelenmeleri farklı yaş grupları da dikkate alınarak lokal olarak araştırılmıştır. Moran I yöntemi ile veri grubundaki esas kümlenmeler ve ekstrem değerlere sahip uç değerli iller berlilenerek sıcak ve soğuk noktaların tayinleri yapılmıştır. Moran I yöntemine ek olarak lokal $G_i^*(d)$ yönteminde ise çalışmaya konu hastalıklar bakımından yüksek ve düşük

riskli illerin belirlenmesi yapılmıştır. Her iki istatistik yönteminde de ilk adım olarak mekansal ağırlık matrisleri çalışmaya konu illerin konumları dikkate alınarak emprik olarak belirlenen sınır değerler kullanılarak üretilmiştir. Söz konusu ağırlık matrsileri tüm analiz sonuçlarını etkileyen girdi parametrelerinden biridir. Çalışma kapsamında LISA için anlamlılık düzeyi 0.05 olarak kabul edilirken algoritme 999 permütasyon ile mekansal kümelenmelerin belirlenmesini yapmıştır. Birbirine yakın objelerin daha fazla benzerlikler göstereceği kabulünün farklı şekillerde geliştirilmesi ile oluşturulan bu yöntemler mekansal istatistik analizlerin temel yöntemeleri arasında değerlendirilmektedir.

Lokal mekansal otokorelasyon (LISA) çalışmaları sonucunda Hepatit A ve A Dizanteri hastalıklarının Türkiye'deki görülme hızlarının 2001-2011 yılları arasında azaldığı sonucuna varılmıştır. İstatistik sonuçlar da her iki hastalık için mekansal kümelenmelerin ve sıcak noktaların Karadeniz ve İç Anadolu bölgelerinde yoğunlaştıkları belirlenmiştir. Bu bölgelerde özellikle Kızılırmak, Yeşilırmak, Seyhan, Batı Karadeniz Akarsu havzalarında kümelenmelere rastlanmıştır. Çalışmanın genel sonuçları 2001-2003 yılları arasında batı bölgelerin, özellikle Marmara, susurluk ve Gediz Akarsu Havzalarındaki illerin, Hepatit A virüsünden diğer bölgelere oranla çok daha fazla etkilendiğini göstermiştir. Bununla birlikte, 2006-2008 yılları arasında sıcak noktaların Batı bölgelerinden Doğu Anadolu ve Güneydoğu Anadolu bölgelerine doğru hareket ettiği gözlenmiştir. Bu kapsamda Hepatat A ve A Dizanteri hastalıklarının yoğun olarak görüldüğü batı illerinin Marmara, Batı Karadeniz, Gediz ve Menderes Akarsu havzalarında; doğu illerinin ise Aras, Fırat-Dicle ve Van Gölü Akarsu havzalarında bulundukları tespit edilmiştir. 2009-2011 yılları arasında ise Hepatit A ve A Dizanteri sıcak nokta ve kümelenmelerinin dikkat çekici bir şekilde azaldıkları gözlenmiştir. Yaşa göre yapılan Hepatit A analizlerinde 15 yaş altı çocukların beklendiği gibi daha yüksek oranda enfekte oldukları, özellikle 5-9 yaş grubunda ise görülme sıklığının en yüksek orana ulaştığı belirlenmiştir. Gerçekleştirilen tez çalışması, uygulanan yöntemlerin sağlık politikalarının oluşturulması ve salgın hastalıklarla mücadelede kullanılabilecek nitelikte sonuçlar elde edilmesinde etkin bir araç olarak kullanılabileceğini göstermiştir

1. INTRODUCTION

Epidemics of Infectious diseases have been documented throughout history and recent global health reports show a continual vulnerability of a large number of people to infectious diseases such as Hepatitis A (HAV) and A Dysentery (AD) in relation to their environment (Nelson and Williams, 2013). Of recent century, infectious diseases have lost a lot of their threat to people's health as well as to the health of populations living in industrialized countries. Looking at it from a global perspective, however, infectious diseases still play a significant, and will continue to play, a significant role in public health since most of the regions of the world have not reached a level of modernization that is comparable with the industrialized world. Especially developing countries and countries in transition still face an enormous burden posed by communicable diseases on their population's health (Krèamer et al, 2010). Water-borne Hepatitis and Dysentery are infectious diseases which continue to be an important public health problem worldwide, in particular in intermediate endemic regions, hence the need to address them accordingly.

Enterically transmitted water-borne Hepatitis and Dysentery are recognized as a major public health problem in many developing countries. Hepatitis A virus and Hepatitis E virus are reported to be the most common cause of infectious hepatitis epidemics transmitted through water, especially in developing countries. Though both viruses generally lead to self-limiting symptomatic disease, fulminant hepatic failure with fatal outcome occurs in a small proportion of patients (Shankar et al, 2014). In this study, the main focus will be on Hepatitis A virus. Hepatitis A is a highly contagious liver infection caused by the Hepatitis A virus. Hepatitis A virus is resistant in the environment, and the infection is mainly transmitted by the fecal–oral route, either through contaminated food and water or through direct contact with an infected person.

Hepatitis A is generally a self-limiting but usually, it does not result in chronic infection. Normally, the majority of children with Hepatitis A do not have symptoms or have an unrecognized infection, however, more than 80% of adults with Hepatitis A have symptoms (Quarto and Chironna, 2004).

Dysentery may simply be defined as diarrhea containing blood and mucus in faeces. The illness also includes abdominal cramps, fever and rectal pain. There are two types of Dysentery, Bacillary Dysentery is caused by a gram-negative rod-shaped bacteria called Shigella. The other type of Dysentery which will be discussed in the paper is Amoebiasis also known as Amoebic Dysentery, this infection of the liver or intestine is caused by the parasite Entamoeba Histolytica with or without symptoms, normally found in the human intestinal tract and faeces. It is most serious in infants, the elderly, and those with impaired immune systems (Turkington and Ashby, 2007). In this study, the main focus will be on Hepatitis A virus and A Dysentery as mentioned above. The control and prevention of these enterically transmitted infections remain a major public health challenge especially in areas where access to clean safe drinking water and sufficient sanitation cannot be assured.

Viral Hepatitis has emerged as an important communicable disease in recent years. Viral Hepatitis is a group of systematic infectious diseases induced by various types of hepatitis virus, predominantly with liver damages (Lu and Zhou, 2015). Between the different types of viral Hepatitis (A–G), viral Hepatitis A is attributed as rarely fulminant diseases with no tendency for chronicity. Nevertheless, Hepatitis A is a worldwide endemic disease, whose real incidence is, owing to various reasons, difficult to estimate (Avonts et al, 1999). An estimated 1.4 million new cases of HAV infection emerge globally each year and it is estimated to kill 100,000 people each year (Lozano et al, 2013). Although HAV is generally perceived as a non-serious disease with low mortality rates in children, liver failure due to HAV occurs at all ages. Children are the key source of Hepatitis A disease and represent an important risk for susceptible adults, who may go on to suffer from prolonged disease.

In Turkey, it is one of the major health problems of the country and about two-thirds of patients visiting hospital have Hepatitis A (Avonts et al, 1999; Altinkaynak et al, 2008). In 1998, the Seropositivity of Hepatitis A was 68% in Western and 80% in eastern regions of Turkey (Ungan et al, 1998). Hepatitis A is mostly an asymptomatic infection of childhood. But, due to improving social conditions, there is a trend of getting infected in older ages which caused an increase in morbidity and mortality of the HAV in old people age group (Tahan et al, 2003). Various factors influence this, including large families, poor education, inappropriate human-waste disposal system, crowded day-care centers and international travel have been linked to outbreaks and

endemicity of HAV infection. High-risk groups are cleaning personnel in hospitals, staff in day-care centers, pediatric nurses, drug users, homosexuals, patients with the chronic liver disease, international travelers and consumers of high-risk foods such as bivalve shellfish from stagnant water (Tahan et al, 2003).

A Dysentery is a common life-threatening parasitic disease affecting 12% of the world population. It is the third leading cause of mortality due to parasitic infections worldwide, after malaria and schistosomiasis (Markell et al, 1999). It is predicted that around 480 million people are at risk for amoebiasis, mainly in tropical and subtropical areas, and it has mortality rates ranging between 40.000 and 110.000 annually (Markell et al, 1999; Farthing et al, 1996). According to WHO (1999-2000), in 1999 and 2000, an analysis of foodborne diseases was conducted in Turkey and a total of 84340 and 77515 cases were reported and A Dysentery was the second most frequently notified disease, comprising of 27% and 31% of reported cases in the two years, respectively. People living in developing countries have a higher risk and earlier age of infection than do those in developed regions. In addition to human suffering and loss of lives, Amoebiasis outbreak causes loss of manpower, disrupts socio-economic activities and subsequent economic damage in the affected areas (Jackson, 2000). Of particular concerns is the geographical distribution of Hepatitis A and A Dysentery prevalence in terms of their temporal distribution, spatial patterns and hotspots identification and its associated risk factors.

Spatial epidemiology concerns the analysis of the spatial/geographical distribution of the incidence of the disease (Lawson, 2013). In most cases, epidemiological analyses are based on observations of disease occurrence in a population of people “at risk.” Typically, we want to relate occurrence patterns between collections of people experiencing different levels of exposure to some factor having a putative impact on a person’s risk of disease. Integrating GIS with geostatistics can provide public health officials and policy makers with vital information needed to detect, manage Hepatitis A and A Dysentery epidemic outbreaks and predicting spatial distributions of these diseases.

Water-borne Hepatitis A and Dysentery have been a public health burden in Turkey. The ministry of public health of Turkey has been trying to monitor and control this disease for many years. Despite a number of studies that utilized the GIS technology for mapping Hepatitis A and A Dysentery in Turkey, there is no study Characterizing

their distributions in all the 81 provinces of Turkey. This thesis focuses on providing insight into the spatial patterns of Hepatitis A and A Dysentery based on hypothesis, which also leads to exposing previously unsuspected patterns leading to the formulation of additional theories by experimenting an innovative GIS-based approach.

1.1 Purpose of Thesis

The study focuses on the application of current spatial statistical methods to study the spatial epidemiology of Hepatitis A and A Dysentery in Turkey. The main objective of this study is to analyze and characterize the spatial and temporal distributions of Hepatitis A and A Dysentery epidemics in terms of their spatial patterns, clusters and hotspots identification by utilizing epidemiological data from 81 provinces of Turkey and advanced GIS supplemented with geostatistics.

1.1.1 Specific aims and objectives

- To describe the spatial variation in disease incidence for the formulation of aetiological hypotheses;
- Investigate the space-time diffusion dynamics of Hepatitis A and A Dysentery
- To identify areas of unusually high risk in order to take preventive action;
- Determine the spatial relationship between Hepatitis A and A Dysentery incidences and possibly get an insight on potential factors that may increase the risk of these infections using GIS and geostatistics.
- Identify whether the Hepatitis A infection in Turkish children under the age of 15 years has similar patterns with HAV infections in the whole Turkish population.
- To provide thematic maps of disease risk in a region to allow better resource allocation and risk assessment.

1.1.2 Research questions

The thesis will attempt to answer the following questions

- How are Hepatitis A and A Dysentery spatially and temporally distributed in Turkish provinces?

- Are the HAV and A Dysentery infections the same across Turkey? In other words, which regions of Turkey are high risks of Hepatitis A and A Dysentery?
- Is there a relationship between HAV infection in Turkish children (under 15 years) and HAV infections in the Turkish population as a whole?
- Can geostatistics and advanced GIS provide cycle of spatial correlation and can they be effective in identifying clustering and hotspots of the two infections

1.1.3 Outline of the thesis

This thesis is organized into five chapters.

Chapter 1- Introduction: The chapter outlines the problem and motivation of the research, establishes the overall focus and objectives of the project. It also gives an overview of the research questions and significance of the project.

Chapter 2-Literature review: This chapter comprises a review of important background literature on Hepatitis A and A Dysentery, Epidemiology, spatial analysis and geostatistical methods. It also includes discussion of previous research done on these infections in some part of Turkish provinces.

Chapter 3- Case Study (Hepatitis A and A Dysentery in Turkish population) and (Hepatitis A in Turkish children under the age of 15): The chapter provides details on the study area, description of the datasets and data sources, data preparation, research methodologies and implementation. It also describes the analysis procedures and tools applied in the study to try to address the questions outlined in the introduction and produce the outcomes. The results and analysis will be provided.

Chapter 4- Discussion: The discussion of the study will be outlined in this chapter.

Chapter 5- Conclusions and Recommendations: The chapter summarizes the project and the major findings, their implications, general recommendations and that of future research are identified.

2. LITERATURE REVIEW

2.1 Overview of Spatial Epidemiology

The broad definition of Public health can be given as collective actions to improve population health. In this sense, spatial epidemiology, considered one of the tools for improving public health, is progressively being used to assess health risks associated with environmental hazards. Spatial epidemiology concerns the analysis of the spatial/geographic distribution of the incidence of disease (Lawson, 2013). Person, place, time: these are the basic elements of outbreak investigations and epidemiology. Historically, however, the emphasis in epidemiologic research has been on person and time, paying little attention to the implications of place or space even though disease mapping has been done for over a hundred years (More and Carpender, 1999). The origin of spatial epidemiology dates back to 1854 with the eminent Dr. John Snow's study of London's cholera epidemic providing one of the most famous examples of spatial epidemiology. Snow believed that cholera was transmitted through contaminated drinking water, see (Figure 2.1) (Waller and Gotway, 2004). Snow's work showed us that public health measures, for instance, the improvement of water supplies and sanitation, can make enormous contributions to the health of populations, and that in many cases since 1850, epidemiological studies have recognized the appropriate measures to take (Bonita et al. 2006).

Spatial epidemiology enables you to better understand diseases or ill-health processes; investigate relationships between the environment and the presence of disease; conduct disease cluster analyses; predict disease spread; evaluate control alternatives; and basically do things an epidemiologist otherwise would have been unable to do and eradicate many errors that could have been committed (Carpenter, 2011). The development of Geographic Information Systems (GISs) and geostatistical methods over the last 20 years have provided a more powerful and rapid ability to examine spatial patterns and processes. Since many of these diseases have an environmental or geographical dimension, it is logical to use GIS technology to understand the etiology of diseases (Walker et al. 1996; Hales et al. 1999). The geography of a disease can

give valuable clues into an understanding of how culture, environment, and behavior interact with health and disease.

Figure 2.1: John Snow's 1854 cholera outbreak map of London (deaths shown as dots, water pumps as crosses.

The fundamental issue in using GIS and geostatistical techniques in epidemiology and health research is one of recognizing the spatial structure of a process, whether it is a cluster of health events or a spatial pattern of disease diffusion over time. To be specific, geostatistics can determine whether patterns are due to either random stochastic processes and/or variability in the estimated prevalence because of small population sizes for some units or are, in fact, caused by specific variables such as environmental heterogeneity (Bailey & Gatrell, 1995). The types of spatial issues that health researchers might be focused on include the patterns of morbidity and mortality; factors that are associated with these spatial patterns; the transmission of the disease and disease etiology; the location, the spatial distribution and regionalization of health care resources; factors associated with resource distribution and how they can be easily accessed and; spatial aspects of the interaction between disease and access to health care (Carpenter, 2011).

With regards to GIS and geostatistical applications in spatial epidemiology, the locations of the cohort population (where they are located) and their characteristics (what they are about, or the attributes) are the bases for further spatial analysis (Lai et al, 2009). Data for spatial analysis should comprise of two types of information.

The first class includes attributes of spatial features measured in interval or ratio variables such as; disease (essential) e.g. health or disease. The essential data provide the geographic context upon which to plot disease cases for visualization. The second class involves spatial data (additional) e.g census or demographic. The additional data on environmental or socio-demographic characteristics are needed to either augment visualization or support more in-depth analysis of the disease or health outcome (Goodchild, 1986; Lai et al. 2009). In bringing these two classes of information together, spatial analysis seeks to assess independence or association in values of attributes at the same or nearby locations or locations likely to experience spatial interaction.

Elliott et al. (2000) advised that the ideal data for spatial epidemiological research would consist of information on the population of a study area such as their individual characteristics, movements, personal exposures, and subsequent health records. However, it was recognized that a considerable investment in time and resources was required to obtain such a comprehensive data set. In most cases, three sources of disease data are available: those from hospital discharges, mortality and morbidity records, and surveillance or independent research. Because much of the available health and covariate data is collected for purposes other than spatial analysis, data integration and quality control are important precursors to the application of spatial-analytic techniques (Brown et al., 2010). Fortunately, GIS offers powerful ways of integrating and cross-validating data and this gives researchers the capacity for deciding earlier about feasibility.

A number of studies such as Toprak and Erdogan. (2008), Yiannakoulias et al. (2003), Chaikaew et al. (2009) and Tsai et al. (2009) utilized GIS and geostatistical methods in spatial epidemiology. Given the potential of these methods, the studies used locational information and their attributes values to detect and quantify patterns in public health data and to investigate the degree of association between putative risk factors and diseases. In Turkey, epidemiology of Hepatitis A and A Dysentery has undergone considerable changes. There are studies conducted on Hepatitis A and A Dysentery based on particular provinces.

Karadenis et al. (2017) had studied the seroprevalence of Hepatitis A in Istanbul for different age brackets using statistical analysis namely Chi-squared test and Fisher's exact test. According to their study, the prevalence of total anti-HAV antibodies was

64.8% for all patients, and patients older than 46 years had a significantly high seroprevalence in comparison with other age group. Another study contacted by Alhan et al. (2014) utilized statistical methods to analyze epidemiological shift of Hepatitis A in 11 years in Adana. They performed a study in 1998 and repeated the same study in 2009, from 1998 to 2009 the anti-HAV seroprevalence declined in children aged 2 to 16 years. Kurt et al. (2008) investigated the prevalence of amoebiasis in Izmir province and 59 of 2047 (2.9%) stool samples were found to be positive to E.histolyca/dispar.

2.2 Hepatitis A

Acute viral hepatitis is one of the most prevalent infectious diseases in the world. Early cases of epidemic hepatitis are generally attributed to the Hepatitis A virus, which represents a significant public health problem in many countries. By the middle of the 20th century, researchers recognized that viruses introduced by fecally contaminated food and water were responsible for HAV.

The associated condition was called infectious hepatitis, catarrhal jaundice or Hepatitis A. HAV is classified as the type species of the genus Hepatovirus of the family picornaviridae. The virus is a small (27nm) spherical RNA virus that lacks an outer envelope with an estimated infectious dose of 10–100 viral particles (Sánchez, 2013; Moore, 2006). Hepatitis is very common. According to the World Health Organization, there are more than 1.4 million new cases of Hepatitis A worldwide every year, and it is estimated to kill 100,000 people each year (Sánchez, 2013).

HAV is present worldwide, although the risk of transmission depends on the socioeconomic conditions of populations. In many developed countries, the progressive improvement of hygienic conditions and environmental sanitation has led to a much-reduced risk of acquiring the infection during infancy, when most cases are asymptomatic. In much of the developing world, HAV infection is hyperendemic, and the majority of persons are infected in early childhood. Outbreaks are rare because most infections occur among young children who generally remain asymptomatic (Quarto and Chironna, 2004). More than 70% of older children and adults infected with HAV are symptomatic, while in children younger than 6 years old, 70% of infections are asymptomatic. Symptomatic illness is characterized by fever, jaundice,

and dark urine, in addition to malaise, anorexia, and nausea, vomiting and diarrhea (Holleran, 2009).

Jaundice is the most common symptom. About 70% of people will suddenly develop Jaundice. The symptoms of jaundice include yellowing of the sclera, clay-colored stools, and severe itching. The liver is the only target organ of injury and it may become enlarged or tender. A virus passes through the stomach and into the intestines. Once ingested (taken in via food or drink), the Hepatitis A From there, the virus enters the blood. The portal vein carries the infected blood from the intestines to the liver. Once in the liver, Hepatitis A viruses invade the hepatocytes. The viruses start to churn out new viruses. Unlike most viruses, Hepatitis A does not burst from the cell and destroy it. Instead, the hepatocytes expel the new "daughter viruses" into the bile. The bile is released into the stool, which then carries huge quantities of the virus out of the body. Those viruses can then infect the next unsuspecting person. The symptoms can last one or two months (Goldsmith, 2010).

Unfortunately, the clinical signs and symptoms of Hepatitis A are indistinguishable from other types of acute viral hepatitis. Diagnosis is usually with immunoglobulin (ig) in blood, anti-Hepatitis A immunoglobulin M in the acute phase, and anti-Hepatitis A immunoglobulin G after 6 months. Hepatitis A infection rarely progresses to fulminant Hepatitis A, which can lead to death (Holleran, 2009). The Hepatitis A virus is very hardy. Because Hepatitis A virus lacks a lipid envelope, it is resistant to the action of chemical and physical agents in the environment. It can live in water as hot as 140°F (60°C). The virus can also survive in temperatures below freezing (32°F, or 0°C). The Hepatitis A virus can survive on surfaces such as bathroom and kitchen counters for more than a month at normal room temperatures. It can live in both freshwater and salt water for many months (Goldsmith, 2010). HAV is inactivated only by heating to 85°C for a few minutes, after sterilization (121°C for 20 minutes), or by ultraviolet and other chemical treatments (e.g., iodine, chlorine-containing compounds) at particular concentrations. It should be emphasized that shellfish from contaminated areas should be heated to 90°C for at least 4 minutes or steamed for 90 seconds to inactivate the virus (Quarto and Chironna, 2004; Lu and Zhou, 2015).

2.2.1 Hepatitis A transmission and source of infections

Outbreaks of Hepatitis A occur periodically throughout the world and is primarily passed via the oral-fecal route, which means that something we eat or drink has been contaminated by feces infected with the virus, which account for 2–7 % of the total disease burden. Nevertheless, this figure is probably an underestimate because a considerable proportion of cases (~68 %) remain uncharacterized (Goldsmith, 2010; Sánchez, 2013). In most instance, fecal–oral transmission via person-to-person contact within a household is the predominant way in which the disease spreads. Until very recently, in countries with high social and healthcare standards, Hepatitis A was seen as an infection typical of tourists traveling to highly endemic areas (Franco and Vitiello, 2003). People can get Hepatitis A when infected people do not thoroughly wash their hands after going to the bathroom. In restaurants, Hepatitis A spreads from dirty hands to food, dishes, or eating utensils. Especially, intake of uncooked fishery food is more likely to cause Hepatitis A (Lu and Zhou, 2015). Food handlers are not always to blame, however. The food itself may be contaminated with Hepatitis A long before it reaches the restaurant. For example, farmworkers with dirty hands can infect fresh produce.

The role of the person-to-person transmission is exemplified by the high transmission rates among young children in developing countries, in areas where crowding is common and sanitation is poor, in households Staes et al. (2000), and Day care centers are common places for Hepatitis A outbreaks (Desenclos and MacLafferty, 1993). When child care workers change diapers, they may pass the virus from one baby to another. Toddlers who do not wash their hands properly can easily pass Hepatitis A to playmates and caregivers. Outbreaks in elementary schools are not unusual. Through sewage discharge, HAV can contaminate food crops, natural watercourses, and soil. In the developing countries or regions, the water-borne transmission is an important factor causing endemic prevalence of Hepatitis A. Therefore, food and drinking water are considered major vehicles of HAV transmission to humans (Sánchez, 2013), Lu and Zhou, 2015). While Hepatitis A is not generally considered a sexually transmitted disease, it can be transmitted through sex. Although transmission via blood is not common, for a few days after infection, Hepatitis A can be passed by blood transfusions or by sharing needles for injecting drugs or by saliva. HAV may also be carried by flies, which serves as vectors (Moore, 2006; Goldsmith, 2010).

High prevalent countries/regions include Mexico, India, the Middle East, and Africa (Suchy et al, 2001). Young children have the highest rates of infection and are often the source of infection for others, primarily because infections in this age group are usually asymptomatic and standards of hygiene are generally lower among children than among adults (Plotkin et al, 2008). The patients at the acute stage and those with subclinical infection are the main sources of Hepatitis A infection. During 2 weeks before the onset of jaundice, the stool and blood possess the infectivity that persists for about 3 weeks (Lu and Zhou, 2015). The socioeconomic, environmental, demographic and climatic factors enhance the vulnerability of a population to infection and contribute to the epidemic spread of Hepatitis A. These factors includes the following: overcrowding, poor personal hygiene and improper sanitation, contamination of food and water, spending time in child care centers, travelling to endemic regions, male homosexual activities, intravenous drug abuse, inappropriate human-waste disposal system (Tahan et al, 2003; Suchy et al, 2001).

2.2.2 The burden of Hepatitis A in Turkey

As in many countries, HAV disease is the most common cause of acute viral hepatitis in most Turkish children. In Turkey, Hepatitis A is one of the major health problems of the country and about two-thirds of patients visiting hospital have Hepatitis A (Avonts et al, 1999; Altinkaynak et al, 2008). According to WHO (1999-2000), in 1999 and 2000, an analysis of foodborne diseases was conducted in Turkey and a total of 84340 and 77515 cases were reported and Hepatitis A was the third most frequently notified disease in all years, 14323 and 10435 cases were reported in 1999 and 2000 accordingly, comprising of 21.8% incidence rate in 1999 and 15.4% of incidence rate in 2000, respectively. Before the 90's, HAV disease was highly endemic in the Turkish population, with seroprevalences reaching around 80% in both children and adults. However, around the years of 1994 upwards, it can be seen that the morbidity and mortality due to HAV have been steadily decreasing, this can be a result of improved hygiene and sanitary conditions in the country (Badur and Nedret, 2010).

Today, at the national level, Turkey is a country of intermediate HAV endemicity. However, when considering regional incidences, a strong West/East gradient exists, with highly endemic regions found in the Eastern and South-Eastern parts of the

country with up to 23.6/100,000 incidences, in comparison to intermediate western regions where an average incidence rate of 10/100,000 can be found. In 1998 the Seropositivity of Hepatitis A was 68% in western and 80% in eastern regions of Turkey (Ungan et al, 1998). The seroprevalence rate among those under the age of 30 years is 71.3% and rates increase with age, from 42.7% among 1-year-olds to 91.1% at 25-29 years of age. Anti-HAV positivity was about 54.5% between the ages 3 to 10 and 100% over the age of 40 in Turkey (Tahan et al, 2003; Badur and Nedret, 2010). Even though there was a high rate of cases reported in 1998, this has improved drastically, Turkey has seen around a 50% fall in reported cases of HAV over the last ten years.

The pattern of HAV exposure in Turkey has been changing in recent years, due to altered eating habits (more fast food and exotic food consumption), increased day care attendance and changing migration patterns. A number of Studies from the 1990s in Turkey showed that more than half of all early adolescents had immunity. Some recent Studies conducted in the 2000s showed similar rates, with most studies indicating that about half of adolescents and more than 90% of adults over 30 years of ages immune. More people originating from highly endemic regions in eastern Turkey are moving to the western part of the country, with susceptible second generations who lack anti-HAV antibodies, traveling back to the endemic home region during holidays. Also, issues such as recent cases caused by contaminated tap water emphasize the need for continuing improvements in sanitation and in public education on hygiene practices to lower HAV susceptibility and endemic nature (Badur and Nedret, 2010).

2.3 Amoebic Dysentery

Amoebic Dysentery is caused by the protozoan parasite *Entamoeba histolytica,* and is characterized by inflammation and ulceration of the mucous membranes of the colon, and accompanied by diarrhea (Van den Berg & Viljoen, 1999). Humans are the only host of *E. histolytica* and there are no zoonotic reservoirs. (*(E. histolytica) has a simple lifecycle* existing as either infectious cyst form or amoeboid trophozoite stage. *E. histolytica cyst* is generally spherical, typically 10 to 16μm in diameter and is bordered by a refractile wall that may include chitin. It consists of four nuclei when mature, but contain one nucleus when immature, see Figure 2.2. They survive the acid of the stomach, travel through the small intestine, and, within the terminal ileum or colon,

excyst to form the trophozoite stage. Unlike the inert cysts, *E histolytica* trophozoites are highly motile, their size start from 20 to 40μm in diameter (Stanley, 2003).

Trophozoites may exit in the stool at times, but they cannot live outside the human host. On the other hand, the cysts can survive outside the host for many weeks or even months as they are protected by the wall, and their survival time could last longer especially in moist conditions such as water, soil and on foods (Stuart et al, 2009). The cysts can also survive for a lengthy period of time in normal environmental conditions, for instance, it can survive for more than 12 days in cool faeces and for few weeks in water. They are killed by drying at temperatures above 50 degree Celsius, freezing below -5 degree Celsius and standard treatments of water supplies (Gill & Beeching, 2011).

Entamoeba histolytica is an intestinal protozoan parasite that is responsible for invasive Amoebiasis in about 40–50 million people, causing around 40 000–100 000 deaths worldwide each year. Globally, A Dysentery is the third most common cause of death due to parasitic infection following malaria and schistosomiasis, as estimated by the World Health Organization (Solaymani-Mohammadi & Petri, 2008; Hegazi et al, 2013). About 10 percent of the world's population is estimated to be infected with *Entamoeba histolytica* and in many tropical countries the prevalence may reach 50%). The disease is often found in tropical regions of the world where hygiene and sanitation are often poor. However, about 90% of those infected are asymptomatic, while only 1% that may result in invasive/extraintestinal amoebiasis. It is not common in children under the age of 5 years. Individuals who suffer from damaged or impaired immunity e.g pregnancy, immunocompromised, or receiving corticosteroids may suffer more severe forms of this disease. Amoebic colitis has generally been regarded as an equal opportunity disease affecting children and adults of both sexes (although this notion has been challenged by data indicating a significant male predominance (Acuna-Soto et al, 2000). However, amoebic liver-abscess mainly affects men between the ages of 18 and 50, in whom rates are 3–20 times higher than other populations (Acuna-Soto et al, 2000; Stanley, 2003).

Asymptomatic carriers are common in endemic areas. In non-endemic areas, symptomless carriers should be treated with a luminal amoebicide which will reduce the risks of transmission and protect the patient from invasive Amoebiasis (Stuart et al, 2009). Invasive intestinal parasitic infection can then results into fever, chills,

bloody or mucous diarrhea, and abdominal discomfort, which are symptoms of fulminant dysentery. Symptomatic (invasive) Amoebiasis may be classified as either intestinal or extra-intestinal as illustrated in Figure 2.2. Intestinal forms of Amoebic Dysentery comprise of Amoebic Dysentery and non-Dysenteric Amoebic colitis. Even though extra-intestinal amoebiasis is not common, it is possible when the trophozoites spread to other organs of the body, example the liver and it could result in amoebic liver abscess (Url-1). Once contaminated foods or water is ingested, the cysts will move into the intestinal area. These cysts are protected from stomach acids and this is how they are able to evade destruction. As soon as its in the intestine, the cyst can release the amoebas by breaking open which then burrow into and cause damages to the intestinal walls (Url-2).

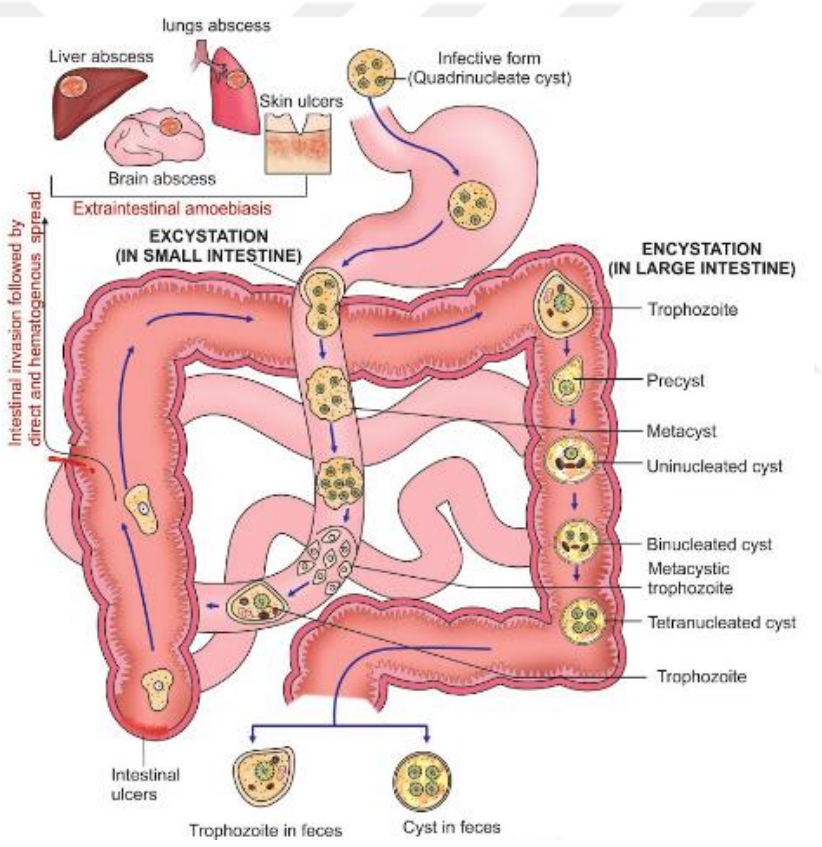

Figure 2.2: Life cycle of Entamoeba Histolyica (Sastry and Bhat, 2014).

2.3.1 Amoebic Dysentery transmission and source of infections

Water contaminated by A Dysentery cysts can cause regional epidemic outbreaks of the disease. A Dysentery is commonly transmitted through faecal-oral route, it the disease can be transmitted either directly being in contact with an infected person, (example by changing diaper) or indirectly by ingestion of cysts through eating or

drinking food or water contaminated with faeces (Hung et al, 2012). Human are the main reservoir. People infected are usually not showing any symptoms, and become carriers of Amoebic Dysentery and can excrete about 15 million cysts in the stool in a day. The incubation period varies from a few days to several months or years but is commonly 2- 4 weeks (Hawker et al, 2008; Heymann, 2008). Some groups of people are at high risk of infection such as immigrants coming from developing countries and travellers to these developing countries with inadequate sanitary conditions, peoples living in institutions with poor sanitation and men who have sex with men. Fomites and flies also have a role in the transmission. Autoinfection through improper cleaning of hands is also reported (Tilak, 2013). The disease can also spread by person-to-person contact. Food can be polluted with the protozoa through fecal contamination, such as when an infected food handler's do not wash their hands after using the bathroom (Turkington & Ashby, 2007). The main source of infections for Amoebic Dysentery is the patients with chronic conditions, at the recovery phase and those that are healthy but carrying cysts (Li, 2015).

2.3.2 The burden of Amoebic Dysentery in Turkey

In Turkey, the general incidences of A Dysentery rate were found to be 9 per cent in 1956. These figures were exceeded by those of a study done in 1958, in which incidence rates of 15 to 42 per cent were recorded (Cannan, 1962). According to WHO (1999-2000), in 1999 and 2000, an analysis of foodborne diseases was conducted in Turkey and a total of 84340 and 77515 cases were reported and AD was the second most frequently notified disease in all years, comprising of 27% and 31% of reported cases in the two years, respectively. It is reported that, in 1999, 22980 cases of Amoebiasis were notified with incidence rate of 34.9 per cent while in 2000, the reported cases increased to 23723 comprising of 35.1 percent incidence rate of Amoebiasis in Turkey (WHO, 1999-2000).

2.4 GIS and Spatial Analysis in Spatial Epidemiology

New diseases and epidemics spread through the world's population every year. The discipline of medical Geographic Information Systems provides a strong framework for our increasing ability to monitor these diseases and identify their causes. The transmission of infectious diseases is closely linked to the concepts of spatial and

spatio-temporal proximity, meaning the transmission of infections is more likely to occur if the at-risk individuals are close in a spatial and a temporal sense (Pfeiffer et al., 2008). The study of the geographical distribution of disease incidence and its relationship to potential risk factors (referred to here as 'Spatial epidemiology') has provided and continues to provide, rich ground for the application and development of statistical methods and models.

Over the last years, its clear that GIS and spatial analysis have had a long, useful and productive relationship. The benefits of GIS include the ability to handle repetitive tasks and rapidly compare spatial data from various sources and different spatial areas. GIS has been appreciated as the key to implementing different methods of spatial analysis, making them more available to a broader range of users, and it is utilized more widely in making effective decisions and in supporting scientific research (Goodchild and Longley, 1999). Spatial analysis can be expressed as a general ability to extract supplementary meaning as a result of manipulating spatial data into different forms to get this information. Spatial analysis is a set of methods useful when the data are spatial, in other words when the data are referenced to a 2-dimensional frame (Cromley and McLafferty, 2011).

GIS has proved to be useful for epidemiological research purposes, as it is concerned with the identification and explanation of the spatial structure, pattern, and process, and with the analysis and explanation of the links between humans and the environment. GIS helps to show regional variation in health problems, environmental risks and the use of health services and reveals abnormal patterns (Rytkönen, 2004). GIS are being used in spatial epidemiology to model where people live and their environments and make it easily possible to monitor residential distributions. Around the nineteenth and twentieth century, spatial analysis commonly took the custom of plotting the observed disease cases or rates.

Figure 2.3 below shows that advances in technology now allow not only disease mapping but also the application of spatial statistical methods, such as quadrant analysis and nearest neighbour analysis, cluster analysis and spatial autocorrelation, are commonly used to characterize spatial pattern of diseases and to test whether there is significant occurrence of clustering of disease in a particular area. Geographic Information System technologies together with modern statistical methods allow an integrated approach to address both the inference on the geographical distribution of

the disease and making a prediction at new locations (Osei FB. (2014). Health and population data can be examined together through the advanced application of GIS technologies and statistical methods in spatial epidemiology. Traditionally, many epidemiologists, public health professionals and GIS experts have used disease plotting or maps when examining the relationship between location, environment, and disease (Clarke et al, 1996).

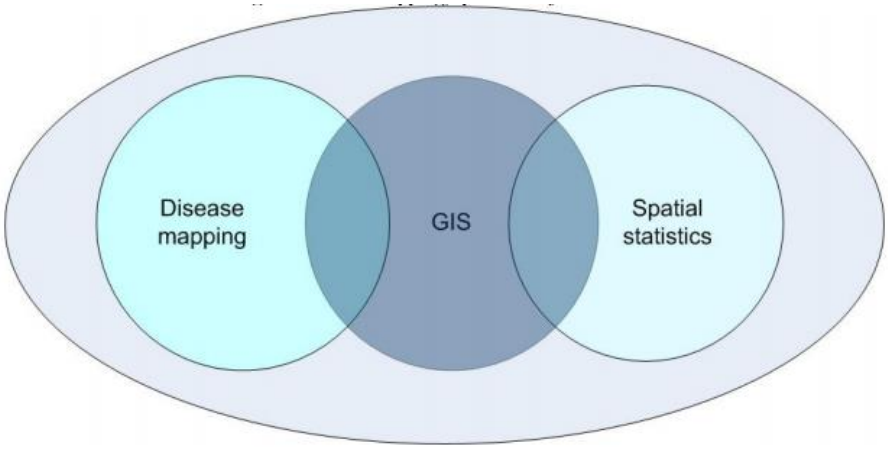

Figure 2.3: Disease mapping, spatial analysis and GIS (WHO, 2011).

2.5 Geostatistical Techniques

Advancement experienced in the field of spatial GIS, that is in conjunction with related geostatistical techniques have improved the use of spatial analysis in research related to health and environmental (Getis. 1996; Moore and Carpenter. 1999). When it comes to analysis of spatial health data, designing maps and visual inferences are not the only main focus. Giving a clear description or view of spatial patterns, the detection of disease clusters, if any and describing or predicting of disease risk are the main goals of spatial epidemiological analysis. Moreover, not every person experiencing a suspected causal exposure can contract the disease. This is the main reason why the analysis of public health data in most cases has to build from the statistical concept of each person partaking in a risk or having a probability of contracting a disease (Waller and Gotway, 2004). Spatial analysis in a GIS environment can be divided into three broad categories: visualization, exploration, and modelling as given in Figure 2.4.

Visualization is one of the first steps performed in any data analysis and is perhaps the most frequently spatial analysis method employed by many researchers or epidemiologists, it involves associating attribute values, for instance, morbidity rates, contamination estimates, or employment levels to locations computed according to a

coordinate system e.g longitude and latitude. Visualizing some characteristic of the effect such as morbidity rate provides an insight into the potential sources of the infection and it aids in coming up with hypotheses about possible relationships between environmental phenomena and health results (Jerrett et al, 2003).

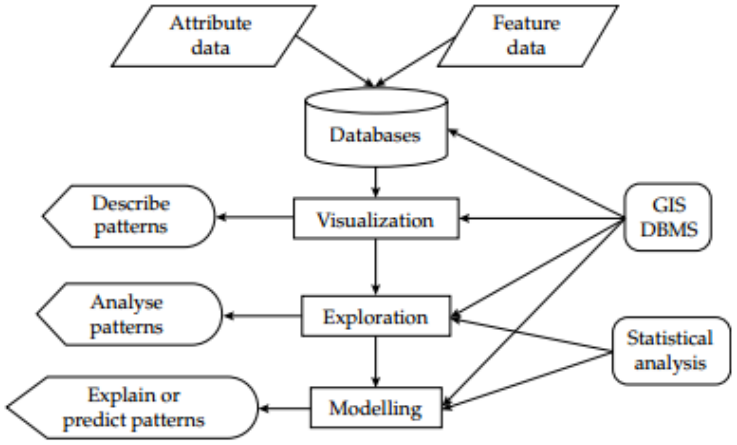

Figure 2.4: Conceptual framework of spatial epidemiology data analysis (Source: Pfeiffer et al, 2008, (Bailey and Gatrell, 1995)).

Exploring spatial associations is equally based on visualization, it involves exploring the datasets by using either spatial queries built on Boolean or set operators that may show collocation. These types of queries will highlight regions on the map that meet the specified conditions (Jerrett et al, 2003).

These aforementioned categories of spatial analysis techniques focus exclusively on exploring the spatial dimension of the data but abandon relevant issues, for instance, dealing with zones of different geographic magnitude, that could have statistical uncertainty and the modifiable areal unit problem. Other issues such as maps of incidence rates, which is defined as raw rates divided by the population at risk have been criticized and regarded as unreliable because of non-constant variance associated with heterogeneity in the sizes of the population at risk. This has led to the recommendation of making use of model-based methods such as Bayes or Empirical Bayes Smoothing when creating maps of relative risk. Most of these issues are dealt with in the modelling categories because they have the ability to affect not just map presentation but also succeeding statistical analyses, respectively.

Modelling spatial processes conveys an idea of cause-effect relationships using both spatial and non-spatial data sources to elucidate or predict spatial patterns. Modelling integrates both visualization and exploration techniques together with geostatistics

analysis in assessing if the spatial patterns apparent in the data have just randomly occurred by a chance or whether they display significant difference from random expectation (Jerrett et al, 2003).

Identification and quantification of any vulnerability to the disease, behaviors, and characteristics that may increase an individual's risk or probability of contracting a disease are the main goals of spatial analytical techniques. The central role of probabilities boosts the use of geostatistical methods to assess the differences in rates observed from different geographic areas. Spatial analytical models and techniques are widely employed in spatial epidemiology to detect spatial anomalies (hot spots) in disease regions (Tsai et al, 2009). Any type of spatial phenomena is examined using Geographic Information Systems technology, an efficient tool for decision making on epidemiological investigation, by providing integrated approach to disease control and surveillance at local, regional and global levels (Tim, 1995; Ulugtekin et al, 2007; Alkoy et al, 2007; Vopham et al, 2015).

Table 2.1 shows different methods for analyzing patterns and mapping clusters. This methods aids in summarizing the prominent characteristics of a spatial distribution, define statistically significant spatial clusters or spatial outliers, assess the overall patterns of clustering or dispersion, aggregate features according to their attribute similarities, identify a suitable scale of analysis as well as exploring the spatial relationships (Waller and Gotway, 2004; Béla, 2010). It is important to note that this is not a linear process because when presenting the results from exploration and modelling, it could require a return to visualization.

Table 2.1: A summary of the tools in the analyzing patterns and mapping clusters.

Tools	Description
Spatial Autocorrelation (global Moran's I)	Measures spatial autocorrelation clustering or dispersion based on feature locations and attribute values.
Cluster and outlier analysis (Anselin's Moran's I)	Given a set of weighted features, identifies clusters of high or low values as well as spatial outliers
Hot spot analysis (Getis-Ord *Gi* $*$)	Given a set of weighted features, identifies clusters of features with high values (hot spots) and clusters of features with low values (cold spots)

2.5.1 Cluster analysis

In spatial epidemiology, one of the important goals is to detect spatial 'clusters' of disease cases. A Spatial cluster is a group of cases or occurrences, lying close together, which are more numerous or denser than expected, relative to the background pattern of such cases, these occurrences are unlikely to have occurred by chance (Baddeley et al, 2015; Knox, 1989). A spatial cluster can be divided into two categories: clustering and cluster detection. Spatial data allows for the quick identification of any obvious patterns (regular, random, or clustered). Here we are focusing on the term 'clustering' which is used to describe the spatial aggregation of disease events, but as the observed spatial pattern may simply be a function of the distribution of the population at risk or of various risk factors, a more robust definition is the one proposed by Wakefield et al. (2000), that a disease is reflecting a clustered pattern if there is 'residual spatial variation in risk after known influences have been accounted for'.

Besag and Newell (1991) differentiate the different methods for analyzing clusters, it can either be specific or non-specific, even though epidemiologists generally choose to use the terms 'local' and 'global'. Global (non-specific) clustering methods are used to evaluate whether clustering is apparent throughout the study region but can not identify the exact location of clusters. They provide a single statistic that measures the degree of spatial clustering of the region, the statistical significance of which can then be assessed. The null hypothesis for global clustering methods is purely that 'no form of clustering exists' (i.e. random spatial dispersion). Global clustering methods involve techniques such as Moran's I and Geary's C. These techniques are similar in the essence that they compare adjacent area values in order to evaluate the level of large-scale clustering.

Local (specific) methods of cluster detection can identify the exact locations and the extent of clusters, with this in mind; it is of great importance to note that clustering applies to global methods of cluster analysis, while cluster detection refers to local methods of cluster analysis. Many local statistics have global counterparts that often are calculated as functions of local statistics. Clustering of infectious diseases can happen due to different types of reasons which can include the infectious spread of diseases, the presence of disease vectors in a particular location, the clustering of a risk factor or combination of different risk factors, or it could be the existence of a possible health hazards sources scattered all over the entire region, all of this issues can create

an increased risk of disease in its immediate locality. The identification and reporting of areas with an apparent increased incidence of the disease is known as a disease cluster alarm (Besag and Newell, 1991; Pfeiffer et al, 2008).

It is of great importance to investigate clustering of possible disease in epidemiology, with one of the aims being to determine whether the clustering is statistically significant and worthy of further investigation, or whether it is likely to be a chance occurrence, or is simply a reflection of the distribution of the population at risk. The statistical significance of clustering is especially important when studying the etiology of a disease, or when evaluating disease cluster alarms). The false identification of a cluster in any of these situations may lead to wasted resources while dismissing a genuine disease cluster can have serious consequences (Lawson and Kulldorff, 1999; Pfeiffer et al, 2008).

2.5.2 Hotspots analysis

Any area displaying "excess" or "unusual" risk (clustering), by some criterion, is a hot spot. (Lawson, 2009). Hotspot detection is normally essential, even when the global pattern is not clustered. Furthermore, clusters of cases may be randomly distributed but can also have an impact on the spread of an infectious disease. Identification of significant disease hotspots can advance our understanding of a disease in several ways, ıncluding suggesting potential risk factors for further ınvestigation, indicating likely disease transmission routes or informing surveillance and disease control effort (Lawson et al, 2016). In hot spot analysis, if a higher value is surrounded by similar magnitude of other high values, it is considered a hot spot (according to certain confidence intervals (CIs)). The cold spots are determined using the same principals. The values (or cluster values) between the statistically significant hot spots and cold spots are considered as random samples of a distribution (Arifin et al, 2016).

2.5.3 Disease mapping

Disease maps have been playing a significant descriptive role in epidemiology. Disease mapping provides (estimate) the true relative risk of a disease of interest across a geographical space (map). Disease mapping involves both the analysis and display of mortality or morbidity rates that show spatial patterns of health outcomes (Delmelle and Kanaroglou, 2015). In terms of the spreading of the disease, individuals closer or exposed to a contagious person or a tainted environmental location are considered

more susceptible to specific types of illnesses. Cartographic design and mapping techniques can draw attention to these locations by displaying an aggregation or lack of such events or patterns in space (Lai and Chan, 2008)

Disease maps show or display a visual summary of geographical risk, for example, the map of Snow in figure 3.1. Disease mapping also generates hypothesis by giving clues to causes of diseases or factors that influence spread by informal examination of exposure maps, components of spatial versus non-spatial residual variability may also give clues to the source of variability (Bailey., 2001). While maps are often used to good visual effect, they can obscure evidence, suggest bogus concentrations, and start false trails. But with careful use, GIS has the ability to sift rapidly the visual correlations between disease distributions and covers of explanatory variables and to combine such evidence with statistical or mathematical modelling in such a way that may both sharpen the value of disease maps and reduce errors of interpretation (Longley and Betty, 1996).

2.5.4 Smoothing

Maps of raw rates, disease counts divided by the total population at risk, have been criticized as unreliable due to non-constant variance associated with heterogeneity in base population size. Mapping disease mortality or morbidity, especially in the smaller geographical areas, or when the given disease is somewhat rare, may give rise to the problem of small numbers, which in turn produces unstable rates. Although the greater stability of rates may be achieved by choosing larger areas, a simple mapping of the raw data is unattractive in that it still yields sudden changes at geographical boundaries (Boyle et al, 1989). In such circumstances, it is advantageous to use smoothing methods as they produce stable estimates for cell-specific rates by borrowing strength from neighboring cells (Anselin et al, 2006). When using smoothing, we are in effect making a prior assumption that a rate estimate for a given area is better if it in some way makes a combination of data from the area itself and those from the surrounding areas.

Numerous smoothing techniques are used to get rid of variance instability in the disease rates (or proportions). Observed rates are usually extreme when the population at risk is too small (e.g., rural areas) or a disease to be analyzed is somewhat rare. Its ultimate goal is to produce a map representation of the important spatial effects

existing in the data, while at the same time, removing any distracting noise or extreme values. The resulting smoothed map should have increased precision without introducing significant bias (Pfeiffer et al, 2008).

The method used to analyze the data entirely depends on how the recording of the data has been conducted. In the case where the data occur as point locations (e.g. outbreaks of disease) kernel smoothing methods can be useful to facilitate visual analysis of the pattern. In the case of data representing, for example, the incidence of infection within administrative areas, Bayesian methods can be applied to take account of the uncertainty of the local measurement and spatial dependence between neighboring measurements (Pfeiffer et al, 2008). As smoothing methods are so numerous, a method needs to be chosen that ensures that features of interest to the reader are not lost. For example, Waller and McMaster (1997) propose a method for routine standardisation of disease rates and illustrate their approach using GIS-generated maps of Leukemia in New York State. Empirical Bayes Smoothing is another most popular smoothing method for modelling disease rates allow one to account for spatial autocorrelation, covariates and sources of error during map construction (Gattrel et al, 1996).

Empirical Bayes Smoothing. Many kinds of literature in disease mapping had often utilized Empirical Bayes Smoothing. There has been an increase in the application of Bayesian approaches in terms of identifying the hotspot and this has been found to decrease by 50 % the rates of false positives and false negatives hotspots in comparison to typical methods of hotspot identification such as classically based confidence intervals (Cheng & Washington, 2005).

An Empirical Bayes (EB) estimation of spatially-varying infectious disease risk, posterior risk can be estimated from a weighted combination of the local risk (also referred to as the likelihood) and the risk in neighboring areas, the latter representing the prior information. In other words, the crude rate is turned into a new variable that has a mean of zero and unit variance, thus avoiding problems with variance instability. The mean and variance used in the transformation are computed for each individual observation, thereby properly accounting for the instability in variance. The risk can be smoothed using a prior based on the global mean or on summarized data from the neighboring areas. The smoothed risk is more stable and has higher specificity (Pfeiffer et al, 2008; Anselin et al, 2006).

3. CASE STUDY

This chapter outlines the case study used in the research study, which in turn gives information on how the research was conducted. It describes the study area and explanation of the data sources followed by the steps used in preparing the data. The steps taken for the analysis of the data are described, the study employed Geographic Information System technology since it has emerged as the core of the geospatial technologies and has shown a big capability once integrated with geostatistical methods and other geospatial features. These methods are useful for identifying areas of high risk, and it aids in determining potential demographic and environmental factors causing diseases. Findings can be expected to inform the development of improved, more locally relevant public health strategies, which take into account socioeconomic and environmental characteristics of populations.

The aim of the case study is to provide insights into the geographic distributions of Hepatitis A cases from 2000 to 2011 and A Dysentery cases in the year 2000 to 2009 in the Turkish population, in terms of their geographical distributions, clusters and hotspots analysis. The research study will also analyze the spatial patterns of Hepatitis A morbidity rate in Turkish children under the age of 15 years. These children will be further considered in three age-groups (0-4, 5-9 and 10-14) will be from 2001 to 2011. The data of patients with Hepatitis A and A Dysentery at province level and the census data for 2000 to 2009 for A Dysentery and for 2000 to 2011 for Hepatitis A were used as a basis for further analysis. Time series maps were created with GIS to introduce the temporal and timely changes in the morbidity rate. Spatial analysis, using GIS and geostatistical methods were used to uncover the hidden patterns of Hepatitis A and A Dysentery in Turkey.

The spatial patterns of Hepatitis A and A Dysentery were measured using Global Moran's and additionally, the local indicators of spatial association (LISAs) Moran's I and Getis-Ord Gi* statistic was used to identify influential locations through clusters and hotspots detection of Hepatitis A and A Dysentery cases.

3.1 Description of the Study Area

The research work was conducted throughout the Republic of Turkey which lies between latitudes 35° and 43° N, and longitudes 25° and 45° E. The country is a transcontinental country in Eurasia stretching across the Anatolian peninsula in southwest Asia and a smaller portion on the Balkan Region of South-Eastern Europe. Turkey's total area, including lakes, occupies 779,452 km2 of which 755,688 km2 are in Southwest Asia and 23,764 km2 are in Europe. The country is divided into 81 administrative provinces and 25 river basins, located in 7 geographical regions: Aegean, Black Sea, Central Anatolia, Eastern Anatolia, Marmara, Mediterranean and Southeastern Anatolia. (see Figure 3.1) Each province is divided into districts, sub-districts and villages. There is a total of 923 districts.

Turkey has a young population structure, people within the 0–14 age group corresponds to 26 percent of the population, the 15–64 age group constitute 67 percent of the total population, and 65 years and higher of age correspond to 7 percent of the total population.

Figure 3.1: Turkish provinces and river basins.

3.2 Datasets

The study used three types of data as follows:

Health data: Tabular data used in the study included the total number of new Hepatitis A and A Dysentery cases and deaths at province level for the Turkish population as a

whole. The study further analyzed Hepatitis A in Turkish children under the age of 15. The children analysis also made use of tabular data, which consisted of the total number of new Hepatitis A cases and deaths at province level based on their age. The stated data were classified into 3 main age groups of children, including 0-4 years old, 5-9 years old and 10-14 years old. The age classification was done in compliance with the age classification of two national governmental authorities, Turkish Ministry of Health and State Statistical Institute, as they base their data collection using the same age group. All the stated data is provided by the Turkish Ministry of Health covered a period of 11 years from 2000 to 2011 for Hepatitis A and 9 years for A Dysentery.

Population data: Demographic census data for the year 2000 to 2009 for A Dysentery and 2000 to 2011 for Hepatitis A cases were also used in the study. The population data was supplied by the State Statistical Institute of Turkey. The missing data for the period of 2001-2006 were estimated using the geometric increase method.

Spatial Data: The spatial data used in the study included vector layer covering the administrative boundaries of the 81 provinces in Turkey. In addition to provinces, vector layers covering the river basins, main water bodies like seas, lakes and neighboring countries were used as reference data for producing user-friendly and more understandable thematic maps.

3.3 Data Preparation

Prior to analysis, all datasets (Hepatitis A and A Dysentery cases, demographic census data and spatial data) were linked in a GIS database for further spatial analysis. However, when shapefiles were loaded into ArcGIS had no coordinate system nor a reference ellipsoid. They were assigned to Lambert Conformal Conic, GSC European 1950. In data preparation, the first step was mapping annual incidence rate of Hepatitis A and A Dysentery for each province. The incidence rate for Hepatitis A and A Dysentery was calculated as the ratio of the number of observed HAV and A Dysentery to the total number of population at risk presented in equation 3.1 for Hepatitis A and 3.2 for A Dysentery.

$$HAV_i = \frac{o_i}{P_i} \quad \textbf{(3.1)}$$

$$AD_i = \frac{o_i}{P_i} \quad \textbf{(3.2)}$$

However, calculating the incidence rate this was could produce unreliable results as it could give an incorrect view of observed reality due to the heterogeneity of base population size. Particularly, calculating the incidence rate of the disease in the smaller geographical areas, or when the analyzed disease is a bit rare, may, in turn, produces unstable rates (Boyle, et al 1989). The corollary to this is that the rates may not completely represent the relative magnitude of the underlying risks in comparison with other counties that have a high population base. To lessen this discrepancy, this study applied Empirical Bayesian Smoothing (EBS) to the calculated incidence rate by shrinking the numbers of observed cases towards the overall mean. This is carried out to produce stable estimates for province-specific rates by borrowing strength from neighboring provinces (Anselin et al, 2006; Clayton, 1987). In this study, the HAV and AD incidence rate for each province during 2000–2011 were adjusted by EBS function and were converted to the Hepatitis A and A Dysentery morbidity rate by multiplying it by 100 000. The demographic census data obtained from state statistical institute constituted of only 2000, 2007-2011 as stated above.

Since the census data was used as the total population at risk, the geometrical increase method was applied to estimate the population for the missing years 2001 to 2006. The equation used is shown in equation 3.3 as follows:

$$N_t = N_0 \times (1 + p)^t \quad \textbf{(3.3)}$$

where;

Nt : the total population at the end of the study period,

$N_{0:}$ the total population at the beginning of the study period,

P: the population growth rate as percentage,

t : the number of years of the study period.

3.4 Applied Methodology and Software Used

The research comprises of 3 main steps: data acquisition/preparation, data analysis and the obtained results. As presented in the methodology, prior to data analysis, all datasets (Hepatitis and A Dysentery cases, demographic census data and spatial data) were linked in a GIS database for further spatial analysis. Data analysis was carried out with the aim to transform the input data into information to be utilized in knowledge construction and decision-making processes. The analysis included boxplots, time series maps, global and local spatial autocorrelation. The final stage of the study presents the results as plots, statistics or maps. The following Softwares were used in the study: ArcGIS (ArcMap 10.2) was used for time series map design and hotspot analysis. GeoDa 10.1.2.16 was used to conduct global and local Moran's I analysis.and SPSS was utilized to create Boxplots for visual analysis. The flowchart for the research study is shown in Figure 3.2.

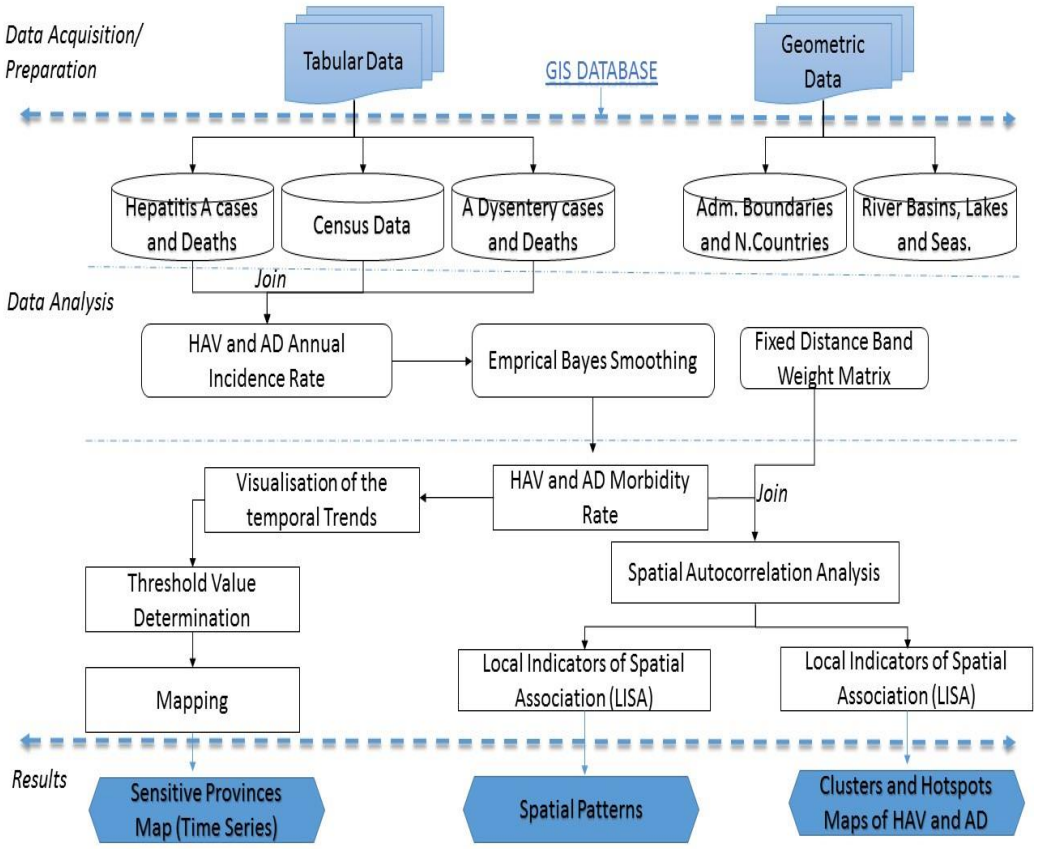

Figure 3.2: Flowchart of methodology.

3.5 Visual Data Mining of Hepatitis A and A Dysentery

Visualizing the spatial characteristics of the data is the primary step in any epidemiological analysis. Recent decades have seen the emergence of a set of techniques originally developed by John Tukey to display data in such a way that interesting features will become apparent. They are generically referred to as Exploratory Data Analysis approaches (EDA). Visualization of the data is perhaps one of the best powerful tools in this exploration process as EDA allow for exploration of the data in many different ways (Pfeiffer et al, 2008). Visualizing table data can be a very beneficial way to obtain further insight into the data, mostly generation of new questions and hypothesis.

The usage of boxplots or maps for visual displays of information will offer the epidemiologist with the foundation for generating hypotheses and, if necessary, assessing the fitness or predictive ability of the models (Pfeiffer,1996). Data visualization is essential when dealing with numerical data as it plays numerous important roles such as: (1) helping in formulating new hypotheses or to confirm existing hypotheses for quantitative data; and (2) guiding a statistical analysis of data and checks its validity; (3) It also aids in creating informative illustrations of the data, summarizing large amounts of quantitative information on a diagram for example on box plots (Badie et al, 2011). The case study covers data for the 81 provinces of Turkey. This data are large enough that it is not possible to explore the data by just looking at them in the form of a table, it often helps to obtain simple statistics on the data such as the mean of each feature. This kind of visualization can be of utmost importance in scientific data mining as it can assist our capabilities in drawing statistically sound and scientifically meaningful conclusions about the data. This is essential as well when it comes to communicating the findings to the target audience using plots of a disease distribution.

3.5.1 Boxplots

The box plot was initially introduced by Spear (1952) as a rane bar and later Tukey (1977) refers to them as the box-and whisker plot. Parallel boxplots give a summary on a variable's values, showing the basic order statistics of the data, for instance, the median, the upper and lower quartiles, and the minimum and maximum values. Together these statistics are beneficial in visually summarizing, understanding,

comparing numerous types of distributions.

The boxplots are arranged in such a way that the ends of the box are situated at the first and third quartiles, whereas the vertical bar inserted in the box is the median, that means that the median divides the data into quarters. Additionally, the entire length of the box is the interquartile range (IQR) (Kamath, 2009; Salkind, 2006) Boxplots could also indicate which observations if there is any, are outliers of the data. As defined by Tukey (1977), boxplots have the ability to plot individually that are outside or far-out values.

The length of the box gives information on the sample variability. In the case where the data being analyzed are symmetric, the median bar is located at the exact center of the box. Subsequently, the bar location gives information on the skewness of the data: If located in in the left half of the box, the data are skewed right. While if located in the right half of the box, the data are skewed left. The width of the box does not indicate anything. Meaning, if the whiskers are long relative to the size of the box, it portrays a long-tailed distribution. Whereas for a bell-shaped curve, in other words, a normal distribution, the whiskers are almost the same length as the box (Kvanli et al, 2005; Kamath, 2009; Young et al, 2011).

The box plot is commonly used when comparing various variables by placing the plots for each side by side with a uniform scale along the y-axis to allow the comparison (Kamath, 2009). Parallel Boxplot were created to represent the spread of Hepatitis A (HAV) and A Dysentery (AD) in the provinces of Turkey.The boxplots were used to display variation that exists in the annual HAV and AD in the period of 2000 to 2009, but this was done without necessary making any assumptions of the underlying statistical distributions. They were used as a handy method for comparing the distribution of annual Hepatitis A morbidity rate in Turkish provinces and the degree of dispersion (spreads), degree of skew and unusual values of the data (outliers) were also analyzed by using spacing between the different part of the box (Devore, 2015; Moreno-Día et al, 2012). Therefore this study, utilized the boxplots also to identify outliers, if any, in the study. This is essential because even one outlier can severely affect the values of the mean and the sample deviation. A boxplot is founded on the measures that could be resistant to the presence of the outliers.

3.5.2 Time series maps

The most important factor of the data in many fields that deal with scientific data is generally time. There is a necessity to understand the distributions of the patterns of changes and relationships among data attributes over both space and time especially when analyzing geo-spatial time-dependent data. A visualization approach can help in the exploration of the spatial and temporal distributions of geo-referenced data that include different time varying quantities. Analyzing the distributions can help in formulating initial Hypotheses about the patterns and even stimulate further inquiry regarding the attributes in a dataset (Thakur and Hanson, 2010).

Our approach in conducting the temporal analysis was to map HAV and AD morbidity rate for the years 2001 to 2009 and then try to visually discern whether or not spatial patterns are becoming more concentrated or more dispersed. In this research, the raw data, which is the health data and the population data were used for determining the annual incidence rates of the Hepatitis A and A Dysentery cases. Annual morbidity rates were then visualized as line graphs for the examination of the temporal trends of the diseases. Threshold values present the endemic or epidemic characteristics of the disease, therefore they are also significant for monitoring the spread of the disease in epidemiology.

Threshold values were set regarding the temporal trend of the disease in each province, as stated by (WHO, 1999). The thresholds were used to identify the risk levels of the provinces with high or low risk indicating the disease occurrence by the use of GIS application. The HAV and AD affected provinces were classified into three classes (endemic, hyper-endemic or epidemic) considering the determined threshold values. HAV and AD values below 10 were assigned to endemic class, HAV and AD values between 10 and 20 were considered hyper-endemic, whereas HAV and AD values above 20 were regarded as epidemic. The endemic level was stated as the level of the lowest risk of HAV occurrence, whereas hyper-endemic and epidemic levels were regarded as the highest risks of HAV occurrence. Subsequently, time series maps were created for determining the timely changes in HAV and AD morbidity rates.

3.6 Spatial Autocorrelation

Spatial Autocorrelation (SA) describes the similarity of nearby observations

considering simultaneously both locational and attribute information. SA is used to measure spatial dependency or spatial association between attribute values at a certain location, by evaluating the geographic pattern of features and determine whether they are clustered, dispersed, or random according to their associated attribute (Longley et al., 2003, Curtis, 2005).

Tobler's first law of geography encapsulates the above situation that everything is related to everything else, but nearby things are more likely to be related than distant things (Tobler, 1970). This could be further clarified by referring to spatial autocorrelation means the attribute values (say, morbidity) of proximal regions (say, metropolitan areas) are expected to be more clustered or more likely to have similar values than distant ones.

In the case, where the differences between neighboring regions are smaller than the differences between non-neighboring regions, then there is a positive spatial autocorrelation. On the other hand, a negative spatial autocorrelation proposes that dissimilar values are appearing as neighboring (or connected) regions, it could be high attribute values are closer to low attribute values, or low attributes values are closer to high attribute values (Chung et al, 2004). If the attribute values are randomly distributed across the study area that means that spatial autocorrelation does not exist. The concept of spatial randomness implies that values observed at a particular location do not depend on values observed at adjacent locations, in other words, the observed spatial pattern of values is equally likely to be similar as any other spatial pattern. Autocorrelation tests use point, line, or area features that have attribute values attached to them.

One important distinction in these autocorrelation tests is to determine whether they measure global or local autocorrelation (Jerrett et al, 2003). Global autocorrelation analysis assesses the study of the entire map pattern and particularly asks the question if the observed pattern displays clustering or not. On the contrary, local autocorrelation shifts the effort to explore within the global pattern to identify whether clusters or hot spots, if any, that could be either driving the overall clustering pattern or that reflect heterogeneities that depart from the global pattern (Mathur, 2015).

Global statistics normally answer the question: Is there statistically significant spatial clustering or dispersion? (yes/No)? Local statistics, on the other side, answer the question: where is the spatial clustering in the study area, or where spatial outliers are located (map)?

3.6.1 Global Autocorrelation (Moran's I)

Global spatial autocorrelation measures the overall clustering of the data in the study area. Global indices of spatial autocorrelation have been utilized in different studies to assess the degree to which similar objects tend to occur near each other (Rogerson,1999; Jackson and Waller, 2005). Global spatial autocorrelation methods use a spatial autocorrelation coefficient to analyze how clustered/dispersed observations are in space with respect to their attribute values (Waller and Gotway, 2004; Jackson and Waller, 2005; Lee and Wong, 2001). Global measures recapitulate the spatial association with respect to the entire region being analyzed. The applications of global indexes of spatial correlation in the evaluation of disease patterns always result in tests of *clustering* instead of tests to identify individual *clusters.*

Different geographic places hardly have similar characteristics throughout, making it crucial to take into account the characteristics of points in addition to their locations. In fact, it is not only that the locations of the diseases matter; but the situations of these locations or activities happening in that location are also of great importance (Lee & Wong, 2001).

Global autocorrelation tests measure the tendency, of the data across the study area, for higher values to cluster or associate more closely in space with other higher values than would be expected if the data in question were randomly distributed in the study area. (The same would hold for lower values clustering with other lower values) (Jerrett et al, 2003). In global tests for autocorrelation, it is expected that the relationship between neighboring or otherwise connected regions will remain the same everywhere in the study area (referred to as "stationary" or "structural stability"). For instance, the spatial autocorrelation between morbidity rates in metropolitan areas of Turkey would be the same at every other area in the country, implying that the relationship between the places was completely a function of the distance between these areas and not relative location.

A global measure of spatial autocorrelation is the most commonly used Moran's I Ratio. It measures the tendency of events to cluster or the extent to which points close together have similar values on average than those farther apart based on incident locations and their attribute values (e.g. disease cases) (Mitchel, 2005; Requia and Roig, 2015).

The Moran's I index of spatial autocorrelation deals with interval or ratio attribute data (Lee & Wong, 2001). Moran's I is used to assess the association among connected areas to identify patterns and measure the level of spatial clustering in these neighboring or otherwise connected districts (Boots and Getis, 1988). The formulation of the Global Moran's I are presented in equation 3.4 and equation 3.5, respectively:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad \textbf{(3.4)}$$

where;

n: the number of provinces (81 in the study),

wij: the element in the spatial weight matrix corresponding to the observation pair *i* and *j*,

xi and *xj*: observations for areas *i* and *j*,

$\bar{x}$: is the mean of *xi*

S_0: the aggregate of all the spatial weights, as represented by equation 3.5.

$$S_0 = \sum_i \sum_j w_{ij} \quad \textbf{(3.5)}$$

Moran's *I* focuses on each observation as a difference from the mean of all observations. Values for Moran's I theoretically range from 1 to −1. When neighboring regions tend to have similar values (i.e., the pattern is clustered), *I* will be positive (positive spatial autocorrelation). If neighboring regions tend to have different values (i.e., the pattern is regular), *I* will be negative (negative spatial autocorrelation) (Waller & Gotway, 2004). When there is no correlation between neighboring values, the expected value of *I* is as shown in equation 3.6.

$$E\ (I) = -\frac{1}{N-1} \quad \textbf{(3.6)}$$

Approaching zero as N increases. There are limitations to the use of Moran's I . The most substantial limitation when dealing with medical phenomena is that both measures are global in nature. Thus, the calculation of spatial autocorrelation is done for the entire study area.

For example, the values of the indices might indicate strong spatial autocorrelation but we are not provided with information as to where this cluster of similarity is. Such information is typically vital when dealing with health and risk (Gatrell, 2010). Measures dealing with localized measures of spatial autocorrelation are more appropriate for such investigation.

In the study, global Moran's I index was used to investigate the global pattern of Hepatitis A and A Dysentery in the Turkish population considering the locations of the cases and the annualized HAV and AD morbidity rate at each of these provinces. The method used a spatial autocorrelation coefficient to outline how clustered/dispersed/random the diseases are in the provinces with respect to their HAV and AD morbidity rates. The study used a conventional 0.05 significance threshold. Our decision of whether to fail to reject or reject the null hypothesis was determined by using a two-tailed test with the critical values of -1.96 and 1.96 for the .05 level. The null hypothesis is rejected, if the observed test statistic (z-value) is either greater than 1.96 or less than -1.96 (Rosenthal et al, 2011).

3.6.2 Local autocorrelation

In the case of positive global SA, we would expect similarly high values close to each other somewhere in the study area but a global index can suggest *clustering* but cannot reveal where those hot spots are located (Lee and Wong, 2001). In health research, global relationships are often of less interest than local relationships or clusters that may display nonstationarity. Therefore, local relationships and clusters are in the interest of health research. These issues made (Getis and Ord, 1992, Anselin (1995) and Ord and Getis, 1996) to take into account the local forms of the global indexes, termed *local indicators of spatial association* (LISAs).

In local autocorrelation, an arrangement whereby a low value is surrounded by high values and a high value surrounded by low values is referred to as a spatial outlier. It is encouraged that such outliers are explored further. They could be a data input error, or it could be a region that is unique from the rest of its neighbors (Waller and Gotway, 2004; Curtis and Leitner, 2006).

Anselin (1996) has formally defined the local indicator of spatial association (LISA) as any statistic that satisfies the following two requirements: 1. the LISA for each observation gives an indication of the extent of significant spatial clustering of similar values around that observation; 2. the sum of LISAs for all observations is proportional to a global indicator of spatial association. This connection defines LISAs as components of a global index and provides a means for partitioning a test of *clustering* (the global index) into a set of tests to detect *clusters* "hot spots" of high values or "cold spots" of low values. The general LISA can be used as the base for testing a null hypothesis of no local spatial association. For each location, LISA values allow for the measuring of its similarity with its neighbors and also test the significance of the value (Mathur, 2015). The most commonly used is local Moran's I and local Getis-Ord Gi* statistic (Anselin, 1995; Getis and Ord, 1996). Therefore, the local statistics can convey the nature of spatial dependency (e.g cluster of high values, cluster of low values, and high or low spatial outlier) in a given location, while also giving a global test (Jacquez, 2008).

3.6.2.1 Cluster Detection (Moran's I)

The most widely used family of LISAs is the local version of Moran's I. It detects local spatial autocorrelation in aggregated data by decomposing Moran‟s I statistic into contributions for each area within a study region.

The formulation of local Moran's I is presented in Equation 3.7. The notations in Equation 3.9 are as described for equation 3.4, but the corresponding values are from the local neighboring region. This local indicator identifies clusters of either similar or dissimilar infection morbidity values around a given locality (Pfeiffer et al, 2008).

$$I_i = \frac{(x_i - \bar{X})}{\frac{1}{n}\sum_i (x_i - \bar{x})^2} \sum_j w_{ij}(x_j - \bar{x}) \qquad \textbf{(3.7)}$$

where;

n: the number of provinces (81 in the study),

wij: the element in the spatial weight matrix corresponding to the observation pair *i* and *j*,

xi and *xj*: observations for areas i and j,

$\bar{x}$: is the mean of *xi*

3.6.2.2 Hotspot Detection (local Getis-Ord Gi*(d))

Unlike local Moran's *I* statistic, which measures the correlation between attribute values in adjacent areas, the G*i*(*d*) local statistic (Getis and Ord 1992; Getis and Ord, 1996) is an indicator of local clustering that measures how concentrated the spatially distributed attribute variable are. The statistic was very useful in the identification of hot spots areas in spatial data. The local Getis-Ord Gi*(*d*) equation is presented in equation 3.8 and 3.9 (Getis et al, 1992).

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)x_j - w_i\bar{x}}{s\sqrt{\frac{(ns_{1i} - w_i^2)}{(n-1)}}}, for\ all\ j \quad \textbf{(3.8)}$$

$$\bar{x} = \frac{1}{n}\sum_j x_j \;\; s_{1j} = \sum_j w_{ij}^2, \;\; s^2 = \frac{1}{n}\sum_j x_j^2 - \bar{x}^2 \quad \textbf{(3.9)}$$

where;

x: a measure of the prevalence rate of Hepatitis A for each age-group within a given polygon (in this case each administrative province),

wij: a spatial weight that defines neighboring administrative province *j* to *i*, *i* cannot be equal to *j*, *n*: total number of features (in this case provinces); and indicates that province *I* and *j* cannot be the same polygon,

wi: sum of the weights *wij*.

The polygons are categorized in seven types which are marked with different colors according to z value range in the legend.

In this study, two local measures of spatial association, local Moran's I and local $G_i^*(d)$ were employed for determining locations of Hepatitis A and A Dysentery clusters and hotspots at the local level.

The LISA (local Moran's I) was used to detect core clusters and outliers of provinces with extreme Hepatitis A and A Dysentery morbidity rate values that are not in any way associated with random variation, and subsequently to classify them into hotspots (high values neighboring other high values, HH), coldspots (low values next to other low values, LL) and spatial outliers (high values amongst low values (HL), or low values amongst high values (LH)) (Zulu et al., 2014). In either situation, the probability of obtaining our observed results, if the null hypothesis is true, the (*p*-value) for the location measured has to be smaller for the cluster or outlier to be considered as statistically significant. LISA enables differences to be made among provinces considering whether the pattern is statistically significant (0.05 level) which lead to rejection of the null hypothesis and indicates the existence of clusters. In this research study, local Moran's I analysis used 999 permutations to determine the dissimilarities among the spatial provinces.

Subsequent to Moran's I, the local Getis-Ord $G_i^*(d)$ was applied to provide supplementary information in indicating hotspots, high risk and cold spots of Hepatitis A morbidity rate with statistical significance across localities by determining the spatial dependence and relative magnitude between a specific province and neighboring provinces (Getis et al., 2003). The local Getis-Ord $G_i^*(d)$ statistic returns a z-score for each feature in the dataset. Developing the spatial weights w_{ij} is the first step to calculating local $G_i^*(d)$ and Moran's I statistics. Employing this measure results in a statistic whereby features not falling inside the determined threshold distance are not included in the hotspot analysis and this describes how similar spatial provinces are to one another based on distance. The neighboring provinces were allocated more weight. Values of high next to other high values shows a strongly positive spatial autocorrelation, whereas values of low next to low values indicates strongly negative spatial autocorrelation. In this context, all values in between, range from moderately positive to moderately negative (Getis and Ord, 1996).

3.6.3 Spatial weights and neighborhoods

Defining spatial association in order to determine the relevant neighborhood of the given location, that is the weights matrix, is an important aspect, as it is used to describe the spatial relationships of the districts, whereby those that are close in space are assigned greater weight in the computation than those that are distant (Moran, 1950). The weight matrix is fundamental to all autocorrelation statistics, especially when using areal data. A weighting system, therefore, has to be carefully considered in clustering analysis. Neighbors can be determined based on either adjacency (also known as contiguity) or distance (Pfeiffer et al, 2008).

Binary contiguity can be defined in different ways. In the case of the rook contiguity only shared boundaries are taken into account, if they share a common boundary, a one is entered into the matrix and if they do not share a common boundary, a zero is entered instead. The queen contiguity is an extension of the rook case and identifies neighbors by mutual boundaries and mutual nodes (corners).When using distance to define neighbors; polygons with their centroids situated within a stated distance range are regarded as neighbors. In terms of inverse distance, the relationship between the spatial components is defined by inverse distance weights (for e.g. 1/distance). Generally, distance-based matrices are more realistic and a better way to capture the spatial pattern in the data than contiguity-based matrices (Curtis & Leitner, 2006).

Since the spatial weights reflect the strength of the geographic relationship between observations in a neighborhood, then if it is assumed that a spatial relationship declines in strength as distance increases from any given site, then the $\boldsymbol{W}$ matrix will show that nearby areas are weighted more highly than sites that are far from one another. In the case of the epidemic of a disease and its transmissions such as Hepatitis A and A Dysentery would require the elements within the $\boldsymbol{W}$ matrix to reflect these effects. The weights are usually row-standardized to ensure that the row member for each observation sum to 1, with ones for contiguous neighbors and zero for all others, whether the data are raster or vector (Fischer & Getis, 2009). The formula for each weight is presented in equation 3.10.

$$W_{ij} = \frac{C_{\iota j}}{\sum_{J=1}^{N} C_{\iota j}} \qquad \textbf{(3.10)}$$

With

$$C_{ij} = 1$$

When *i* is linked to *j*, otherwise

$$C_{ij} = 0$$

3.7 Results and Analysis

3.7.1 Boxplots

The Visual Data mining processes results were represented with parallel boxplots to display variation in Hepatitis A and A Dysentery and for providing deeper insights into the distribution of these infections in the stated years. Figure 3.3 shows the annual Hepatitis A morbidity rate in each year (2000-2009) per 100 000 people. The observations are located at the low end of the box and the mean is greater than the median, yielding that the boxplots of every year showed a positively skewed distribution of annual Hepatitis A morbidity rate. The rate of distribution in the year 2009, 2010 and 2011 seems to have smaller variability than the other years. Although there seem to be a decline in the Hepatitis A morbidity rate in 2003 and 2004 in comparison to the earlier years, the Hepatitis A morbidity rate seemed to rise again from 2005. The boxplots clearly reveal that the largest distributions and the highest medians were found in the year 2001, 2002, 2005, 2006 and 2007, implying that people were more infected with Hepatitis A in these years. The results reveal some differences in the Hepatitis A annual morbidity rate, a significant decline can be noticed in the recent years (2009, 2010 and 2011), respectively, in comparison to the previous years.

In this study, we made use of the parallel boxplots to also identify abnormal values in the Hepatitis A morbidity rate. Figure 3.3 visibly illustrate a clear presence of mild outliers and extreme outliers in the HAV morbidity rate in each year. The outliers are represented by national id numbers, provinces such as Gaziantep (ID 27), Kilis (ID 79), Kutahya (ID 43) and Mardin (ID 47) constantly appeared as extreme outliers showing a high Hepatitis A morbidity rate in most years in comparison with other provinces.

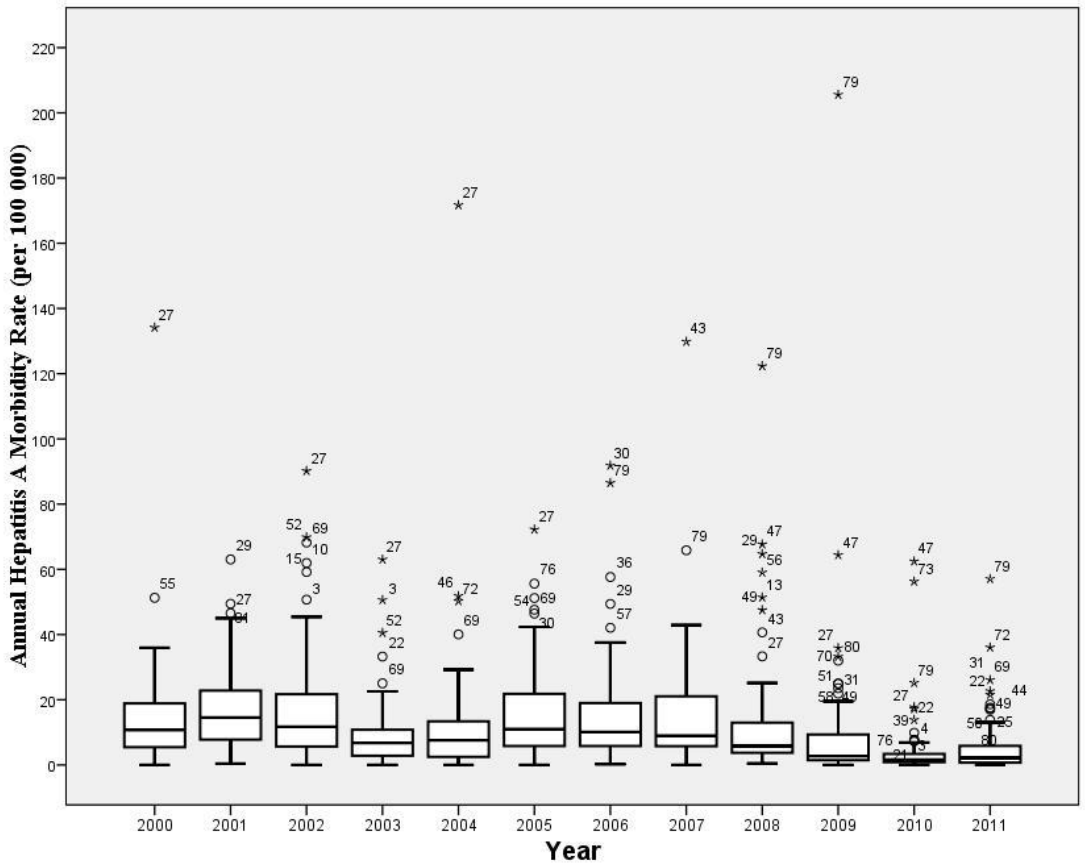

Figure 3.3: Parallel boxplots of annual HAV of 81 provinces of Turkey.

Figure 3.4 shows the parallel boxplots for Amoebic Dysentery (AD) in each year (2000-2009). The boxplots of every year show a positive skewed distribution of annual AD morbidity rate. The rate of distribution in 2008 and 2009 appear to be less variable than the other years, indicating that there is a significant decline in the rate of A Dysentery infection in the recent years (2008 and 2009) respectively, in comparison to the earlier years. AD morbidity rate was identified to be the highest in the year 2001, 2002, 2004, 2006 and 2007 as they have the largest distributions and highest medians. The A Dysentery annual morbidity rate has a large number of extreme and mild outliers in comparison with Hepatitis A morbidity rate. The parallel boxplots for A Dysentery morbidity rate showed extreme outliers as Gaziantep (27), Kilis (ID 79), Diyarbakir (ID 21) and Siirt (ID 56) in most years, among many others during the study period.

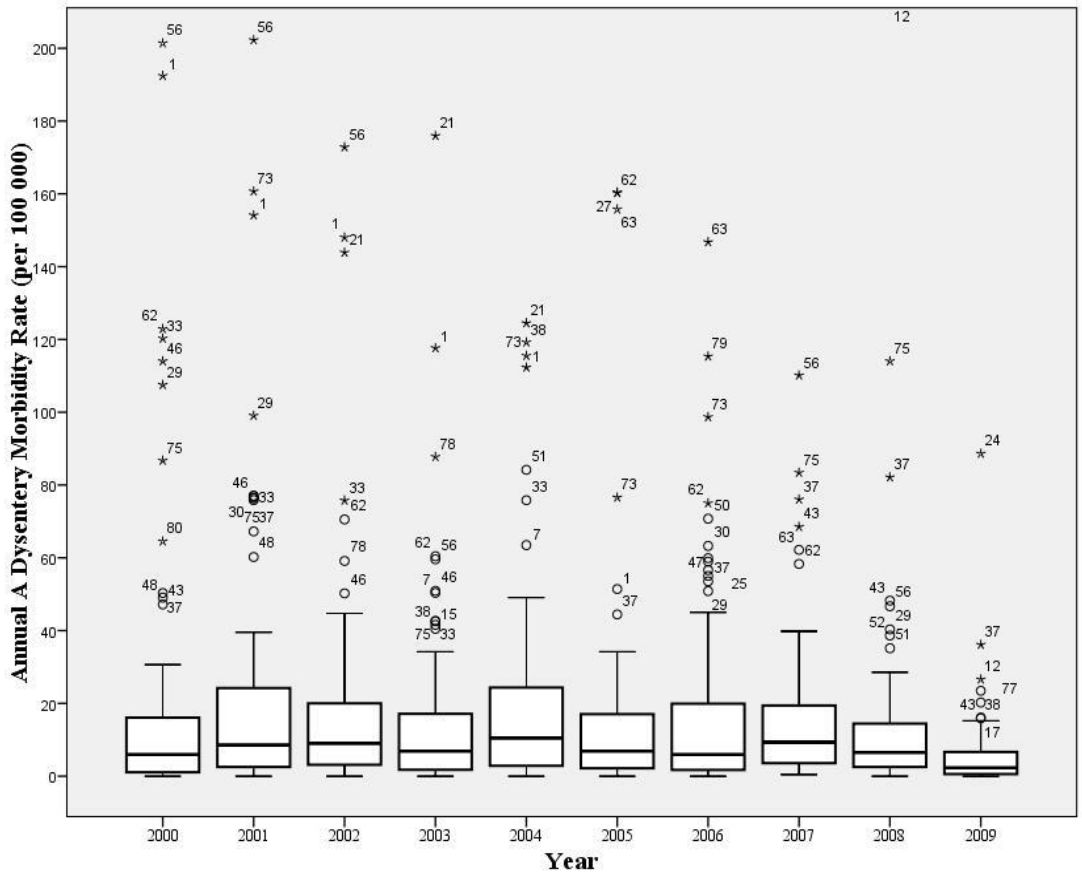

Figure 3.4: Parallel boxplots of annual AD of 81 provinces of Turkey.

Boxplots comparison of Hepatitis A infection between children under the age of 15 and people above the age of 15 years.

The annual parallel boxplots were created to measure the abundance of HAV infection between the Turkish children younger than 15 years and people above the age of 15 years. The comparison was done because it is generally noted that most Hepatitis A infections appear to be particularly common in infants and young children who generally remain asymptomatic (Quarto and Chironna, 2004).

Figure 3.5, 3.6 and 3.7 depict the parallel boxplots of HAV morbidity rates in each year (2001-2011) for children under the age of 15 and people above the age of 15 per 10 000 people. Both boxplots (in Figure 3.5 and 3.6) in all the years were positively skewed. In both categories, the highest infections of HAV morbidity rates are prominent in the years 2001, 2002, 2005, 2006 and 2007. Similarly, the rate of distribution in 2009, 2010 and 2011 appear to have the smallest variability than the rest of the years. This shows that a notable decline is experienced in the annual morbidity rate of recent years in comparison with the preceding years. Kilis (ID 79), Gaziantep (ID 27), Kutahya (ID 43), Batman (ID 72) and Sirnak (ID 73) frequently

showed a high morbidity rate of HAV, and thus they are classified as extreme outliers among others.

By Comparing the parallel boxplots of the two categories, it can be clearly seen that the Turkish children under 15 years are by far infected with Hepatitis A than the rest of the population as it is revealed in Figure 3.7.

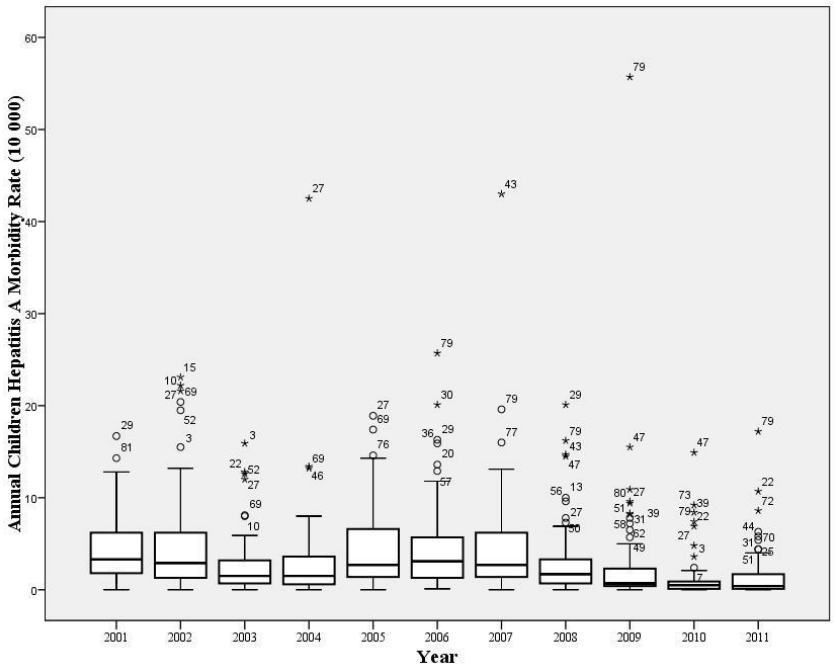

Figure 3.5: Parallel boxplots of annual HAV of children under the age of 15 years.

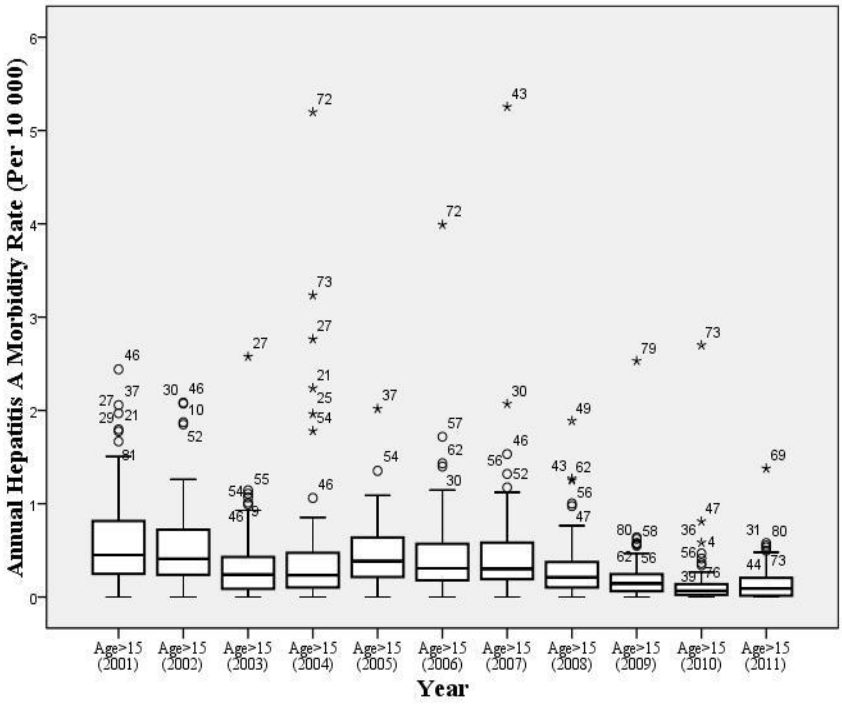

Figure 3.6: Parallel boxplots of annual HAV of people above the age of 15 years.

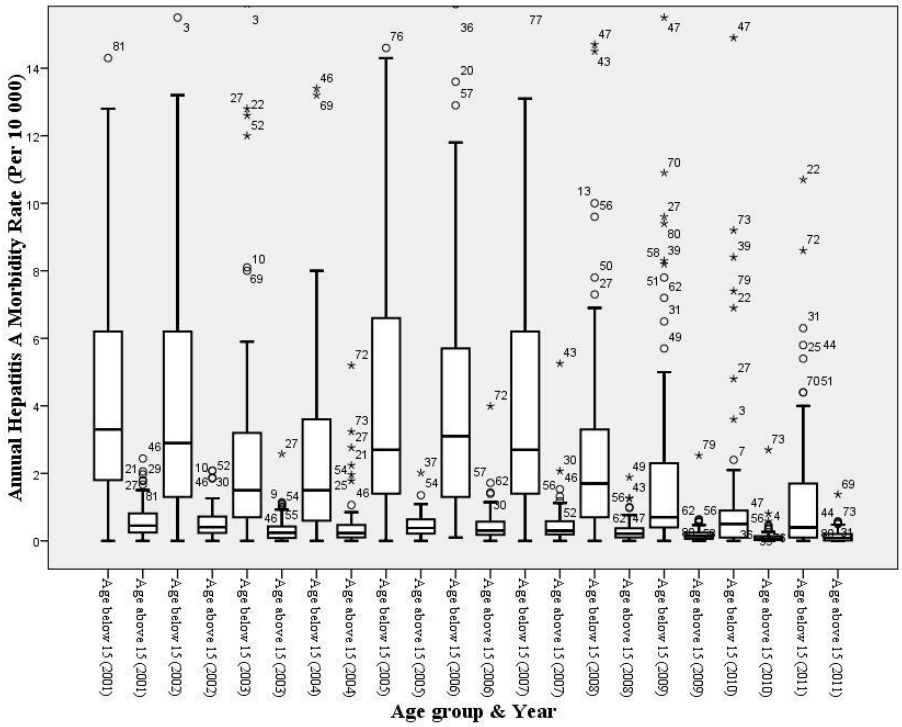

Figure 3.7: Parallel boxplots of annual HAV of children under and people above the age of 15 years.

Upon realizing that the children were more infected with Hepatitis A, they were further analyzed using parallel boxplots to depict the distribution of HAV in three age group (0-4, 5-9 and 10-14). Figure 3.8 shows the results of HAV morbidity rates for children under the age of 15 per 10 000 people in the years of 2001 through 2011. The boxplots of every year also showed a positive skewed distribution of annual Hepatitis A morbidity rate. According to the boxplots in figure 3.5, the boxplots exposed that the infection of HAV in Turkish children has decreased in recent years (2009-2011) than the other years. The largest distributions and medians can be found in 2001, 2002, 2005, 2006 and 2007, indicating that the children were more infected with Hepatitis A than the other years. Concerning the distribution of HAV in each age-group, the annual boxplots show that age-group 5-9 had the highest HAV infection in all the years, followed by children between 10 and 14 years. On the other hand, children under 5 years had the lowest infection, however that can be expected because generally, the HAV infections remain asymptomatics in infants. The extreme outliers can be seen in the same provinces such as Gaziantep (ID 27), Kiris (ID 79), Kutahya (ID 43) and Mardin (ID 47) with high Hepatitis A morbidity rate in most years.

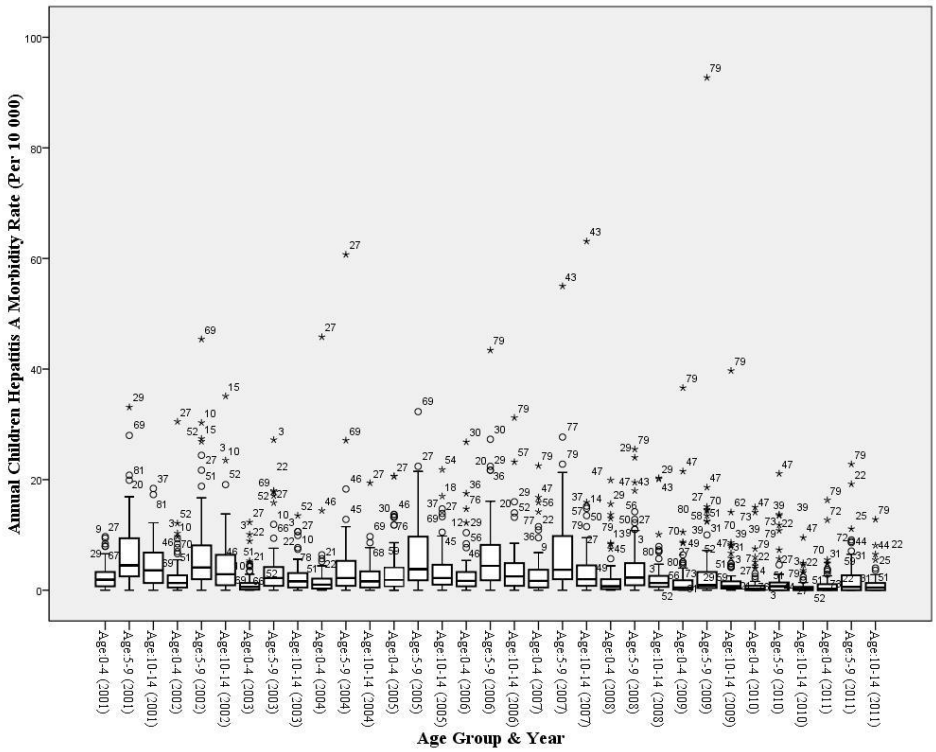

Figure 3.8: Parallel boxplots of annual HAV of all children in age group (0-4, 5-9 10-14) in Turkey.

3.7.2 Time series maps

Time series maps indicating the temporal changes in geographic distribution of Hepatitis A and A Dysentery morbidity rates in Turkish provinces were produced. The Hepatitis A and A Dysentery affected villages were classified and evaluated based on the determined threshold values, endemic (displayed in light pink), Hyper-endemic (shown in light orange) and epidemics (displayed in dark red) according to the occurrence of Hepatitis A or A Dysentery.

Figure 3.9 shows the results of the timely geographical distribution of Hepatitis A morbidity rate. The maps illustrated a decreasing trend in morbidity rates of Hepatitis A at epidemic level from 2001 to 201. However, it should be noted that the number of epidemic provinces was more pronounced in 2001, 2002, 2005, 2006 and 2007 for Hepatitis A cases as presented in Figure 3.9. This is also confirmed by the visual analysis of boxplots in Figure 3.3. Thematic maps were evaluated for the interpretation of the provinces with high and/or low risks according to the occurrence of both diseases. Provinces such as Gaziantep, Kilis, Hakkari, Manisa, Samsun and Sinop were repeatedly epidemic in most years, therefore they were assigned to high-risk provinces, whereas provinces such as Istanbul, Bilecik, Eskisehir, Artvin, Giresun were mostly

showing endemic tendency throughout the years and thus were considered low-risk areas. Most provinces that exhibited epidemic characteristics in earlier years became endemic, however, the Meriç-Ergene (e.g Edirne and Kirklareli), Firat-Dicle (e.g Bayburt, Batman, Mardin, Sirnak and Kilis) and Asi (e.g Hatay) River Basins provinces were still epidemic in the years 2010 and 2011.

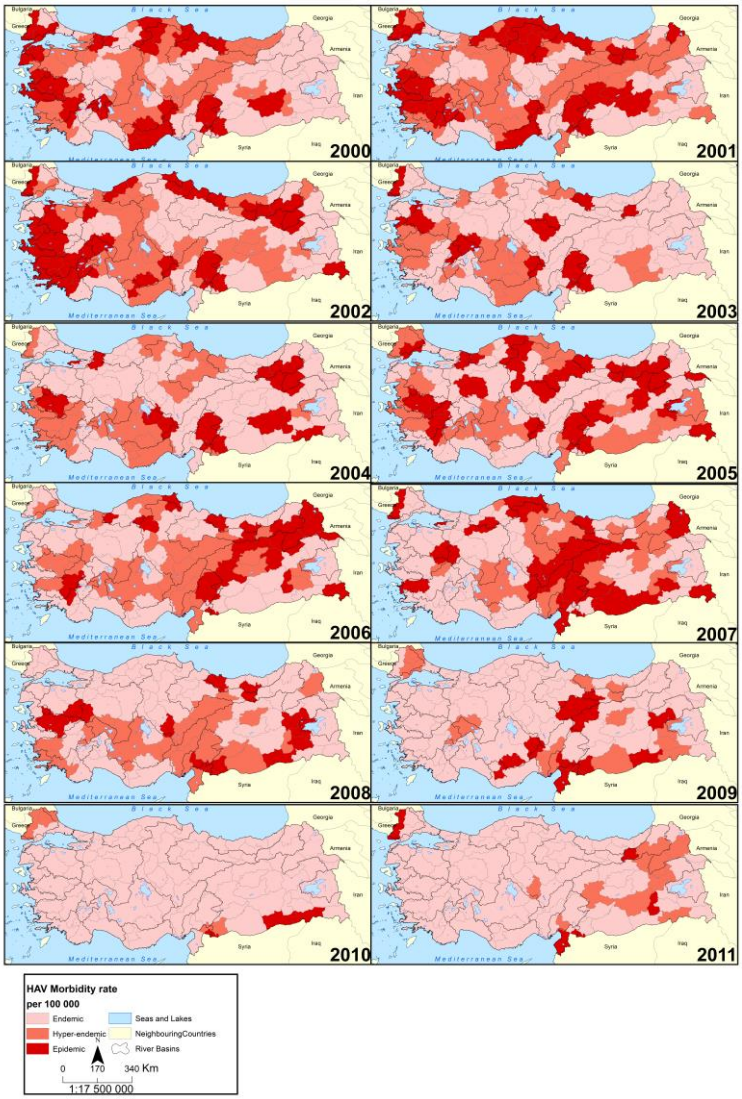

Figure 3.9: Time series maps of HAV morbidity rate in Turkey, from 2000 to 2011.

According to Figure 3.10, the time series maps for A Dysentery morbidity rate also shows a downward tendency at epidemic level from 2001 to 2009. The number of epidemic provinces decreases significantly after the year 2007 as seen in Figure 3.10 and this is can also be verified by the boxplots of A Dysentery morbidity rate as illustrated in Figure 3.4. The Mediterranean, southeast Anatolia, Central Anatolia and Aegean Regions were highly epidemic in 2000 to 2007, but there was a drastic improvement in these regions, most provinces that were high risks became endemic

(low risks provinces) in recent years 2008 and 2009. Provinces such as Istanbul, Kirklareli, Tekirdag, Agri, Rize Aydin were classified as low risk, while provinces such as Antalya, Mugla, Kastamonu, Ardahan, Sirnak, Kutahya, Malatya and Elazig were regarded as high-risk areas in terms of AD morbidity rate. Particularly, Kastamonu province recorded epidemic tendency since 2000 until 2009.

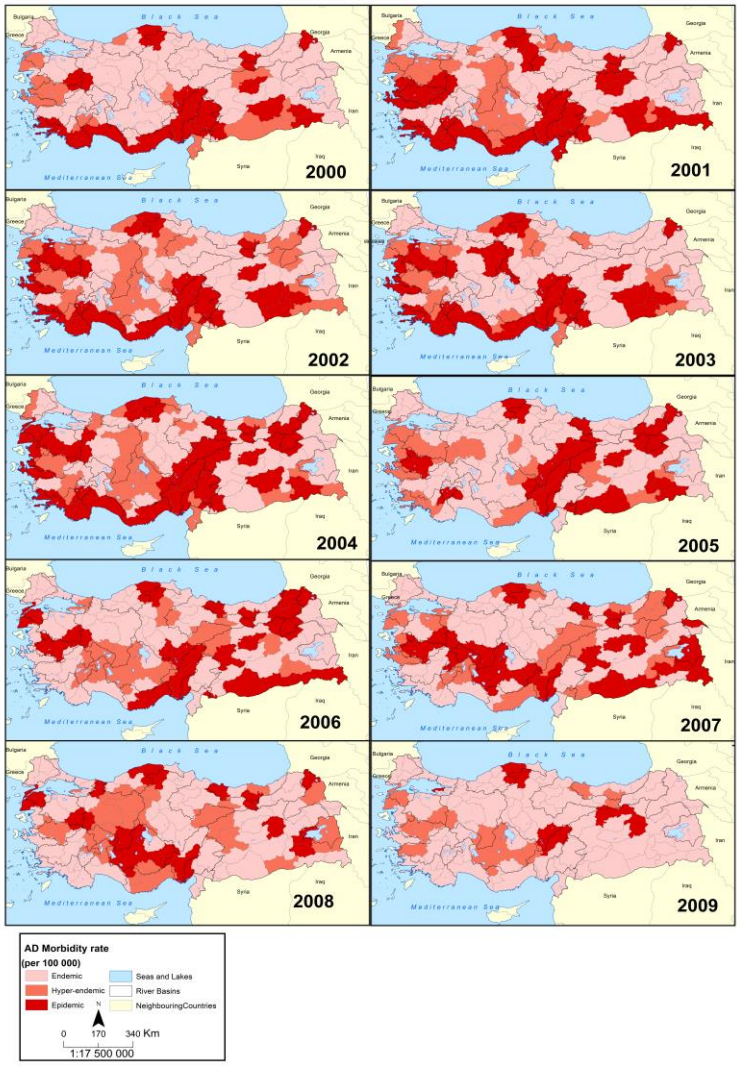

Figure 3.10: Time series maps of AD morbidity rate in Turkey, from 2000 to 2009.

Time series maps were also produced for all children below the age of 15 to evaluate the temporal geographical distribution of the HAV morbidity rate in children as given in Figure 3.11. The results depicted that most provinces revealed endemic characteristics during the course of the years, though some other provinces such as Kilis (ID 79), Gaziantep (ID 27), Edirne (ID 22) reflected high HAV morbidity rates. A decreasing trend of the morbidity rates at epidemic levels can be deduced from 2001 to 2011. Conversely, provinces with epidemic HAV characteristics were evident in the

years 2001, 2002, 2005, 2006 and 2007 as also shown by the box plots (see Figure 3.5).

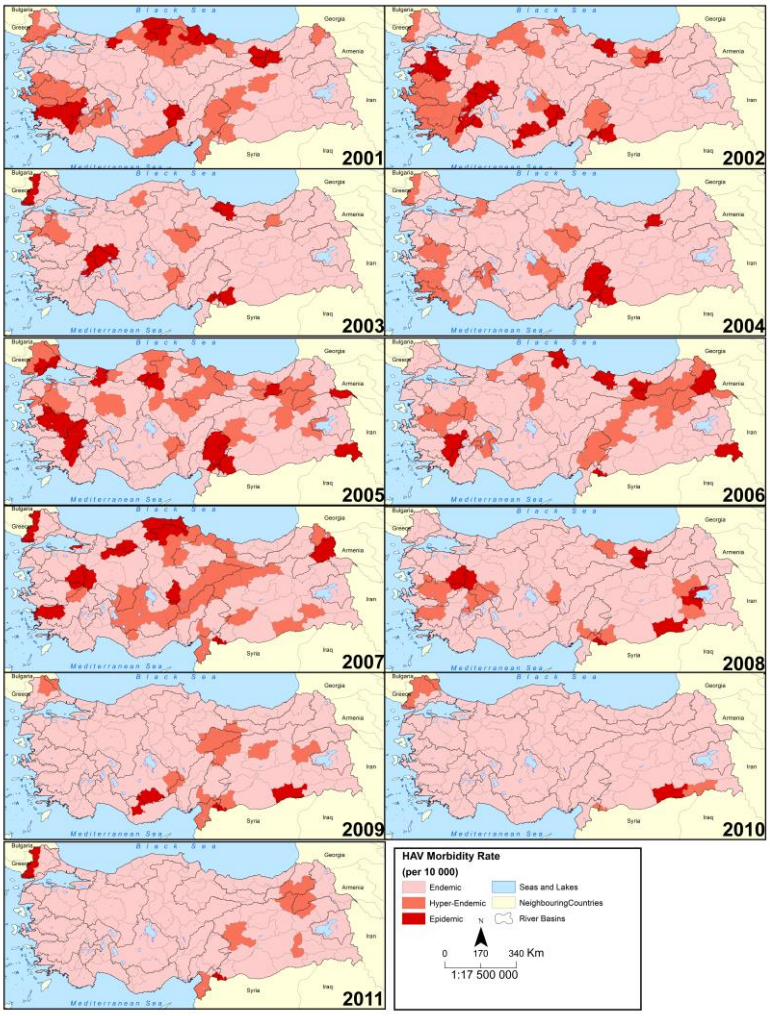

Figure 3.11: Time series maps of HAV morbidity rate for the children under the age of 15 years for 2001-2011.

3.7.3 Global autocorrelation Moran's I

The extent to which neighboring values of Hepatitis A and A Dysentery morbidity rates are correlated was measured using Moran's Index. The global Moran's I index evaluated whether the pattern expressed is clustered, dispersed, or random. As it can be seen in Table 3.1 and 3.2 Moran's I index were positive values in some years. Moran's I index was evaluated at the significance level of 0.05.

Table 3.1 shows the global spatial autocorrelation analysis of Hepatitis A morbidity rate in Turkish provinces for the year 2000 to 2011. The results reveal that at the province level, the spatial distribution of HAV morbidity rate was clustered in 2005, 2006 and 2007. The Moran's I value was 0.002 and had a z-score of 2.0 in 2005, in

2006 the Moran's I value was 0.02 with a z-score of 4.10 and finally in 2007 the Moran's I value was 0.009 and possessed a z-score of 2.9, respectively, implying that the distribution of the HAV affected provinces was somewhat spatially autocorrelated (low clustered) though the overall tendencies were not so strong. The z-score for 2005, 2006 and 2007 are greater than 1.96 and these values reject the null hypothesis of no spatial autocorrelation and accept the inverse perfect correlation. Conversely, for the year 2000, 2001, 2002, 2003, 2004, 2008, 2009, 2010 and 2011 the Moran's I index was expressed with negative values which indicated the negative spatial autocorrelation of the HAV phenomena investigated. The z-values of these years indicates statistical insignificance and thus these values suggested a tendency towards dispersion in HAV morbidity rate.

Table 3.1: Global (Moran's I) index of spatial association for Hepatitis A.

Year	Moran's I	z-score	Pattern
2000	-0.02	-1.2	dispersed
2001	-0.02	-1.8	dispersed
2002	-0.001	1.2	dispersed
2003	0.001	1.8	dispersed
2004	-0.01	-0.8	dispersed
2005	0.002*	2.0	clustered
2006	0.02*	4.10	clustered
2007	0.009*	2.9	clustered
2008	-0.01	-0.3	dispersed
2009	-0.03	-4.1	dispersed
2010	-0.02	-1.5	dispersed
2011	-0.03	-2.7	dispersed

Note *: The corresponding significance level (p-value) is < 0.05.

Table 3.2 present the global spatial autocorrelation analysis of A Dysentery morbidity rate of provinces of Turkey of the year 2000 to 2009. Moran's I statistic was 0.01 with a z –score of 2.6 and 2.4 for the year 2006 and 2007, respectively. These values were statistically significant at (significance < 0.05). The results show a positive spatial autocorrelation for these years and it indicated a geographical pattern in which the provinces of similar values tend to cluster. However, for the year 2000, 2001, 2002, 2003, 2004, 2005, 2008 and 2009 the Moran's I values were negatives, the null hypothesis was accepted, thus the overall spatial distribution of AD occurrence was not spatially auto-correlated.

Table 3.2: Global (Moran's I) index of spatial association for A Dysentery.

Year	Moran's I	z-score	pattern
2000	-0.01	-0.6	dispersed
2001	-0.01	0.03	dispersed
2002	-0.02	-1.4	dispersed
2003	-0.02	-0.9	dispersed
2004	-0.02	-1.3	dispersed
2005	-0.02	-2.6	dispersed
2006	0.01*	2.6	clustered
2007	0.01*	2.4	clustered
2008	-0.01	0.1	dispersed
2009	-0.01	0.04	dispersed

Note *: The corresponding significance level (p-value) is < 0.05.

The results of the global Moran's I statistic calculation for Hepatitis A morbidity rate in Turkish Children under the age of 15 for the year 2001-2011 are summarized in Table 3.3 below. It was found that the values of Moran's indices were positive (0.05 significance level) in 2002, 2003, 2006 and 2007 with Moran's I value of 0.005, 0.01 and 0.009, respectively, indicating clustered characteristic of the HAV morbidity rate in these years during the study period. The rest of the years (2001, 2004,2005, 2008-2011) exhibit a negative spatial autocorrelation for HAV morbidity rate in Turkish children. This indicated a map pattern in which the geographic units (which are provinces in our research) of similar values scattered throughout the map. This stated that feature values of these years were randomly distributed across the study area.

Table 3.3: Global (Moran's I) index of spatial association for Hepatitis A for children under the age of 15 years.

Year	Moran's I	z-score	pattern
2001	-0.01	-0.5	dispersed
2002	0.005*	2.07	clustered
2003	0.01*	3.05	clustered
2004	-0.01	0.01	dispersed
2005	-0.0004	1.22	dispersed
2006	0.009*	2.7	clustered
2007	0.009*	2.6	clustered
2008	-0.01	0.2	dispersed
2009	-0.04	-5.3	dispersed
2010	-0.04	-3.7	dispersed
2011	-0.04	-4.20	dispersed

Note *: The corresponding significance level (p-value) is < 0.05.

On the other hand, Table 3.4 a, b and c shows the global Moran's I of annualized morbidity rate of HAV in Turkish children categorized in three age-group (0-4, 5-9

and 10-14) for 2001, 2006 and 2011. Even though the study was carried out from 2001 to 2011, the study only presented the results for 2001, 2006 and 2011. These years were selected in such a way that 2001 will give the general overview of how the infections were at the beginning of the analysis, 2006 was chosen as it had the highest median and 2011 can outline how the HAV is in the recent years.The Moran's I statistic unveil that the index had statistically significant values in some years. In 2001, the HAV spatial pattern for the age-group of 10-14 (0.006) was spatially auto-correlated, but all other age-groups showed that the patterns were dispersed. The HAV morbidity rate in 2006 for children under the age of 10 were statistically significant (significance level < 0.05) implying that the distribution of the affected provinces with HAV was clustered in space with the Moran I values of 0.03 and 0.02. However, for children above 10 years old in 2006 and for all age-groups in 2011, Moran I values were all showing negative values, depicting the negative spatial autocorrelation of the HAV occurrences being investigated.

Table 3.4(a): Global (Moran's I) index of spatial association for HAV in the age-group 0-4.

Year	Moran's I	z-score	pattern
2001	-0.03	-2.20	dispersed
2006	0.03*	5.11	clustered
2011	-0.02	-1.41	dispersed

Table 3.4(b): Global (Moran's I) index of spatial association for HAV in the age group 5-9.

Year	Moran's I	z-score	pattern
2001	-0.01	-0.31	dispersed
2006	0.02*	4.31	clustered
2011	-0.05	-4.39	dispersed

Table 3.4(c): Global (Moran's I) index of spatial association for HAV in the age group 10-14.

Year	Moran's I	z-score	pattern
2001	0.006*	1.98	clustered
2006	-0.01	-0.25	dispersed
2011	-0.05	-4.79	dispersed

Note *: The corresponding significance level (p-value) is < 0.05.

3.7.4 Local cluster analysis (local Moran's I)

The localities with clusters of Hepatitis A and A Dysentery cases abundance were identified by using local Moran's I statistics. The study focused on univariate spatial distribution and the significant clusters or spatial outliers in the HAV or AD morbidity data. Figure 3.12 (a) displays the Moran scatterplot for the year 2006 showing the standardized values of HAV in each province to correlate the observed values and the spatial lag of HAV (formed by averaging all the values of neighboring provinces) and to detect outliers. The slope of the scatter plot corresponds to the value for Moran's I. The Moran scatterplot showed that most of the event points were located in the lower left (low-low) and upper-left (low-high) quadrants confirming the existence of both positive and negative spatial autocorrelations among the province based HAV morbidity rates, particularly for 2006.

The corresponding sample LISA cluster map shown in Figure 3.12 (b) provide essentially the same information as the four quadrants in the Moran's scatterplot but with the significant locations color coded by type of association. The LISA cluster map of HAV morbidity rate in 2006 is showing the locations with significant higher clusters (high surrounded by high), low clusters (low surrounded by low) and spatial outliers (respectively, high surrounded by low, and low surrounded by high). There were some outstanding clusters in some provinces in the Eastern Anatolia side of Turkey exhibited high clusters of high surrounded by high, while low clusters were found in the western and central region of Turkey; however central Turkey seem to have a mixture of the low and high cluster as well as spatial outliers. These clusters and outliers were statistically significant with a (p-value < 0.05).

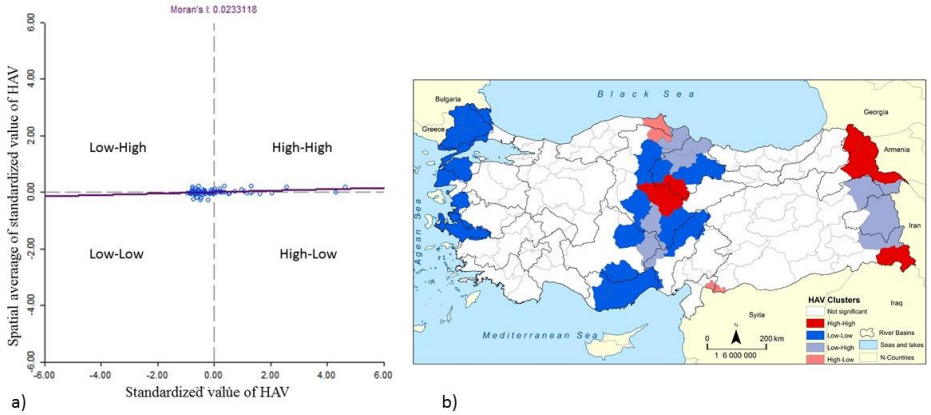

Figure 3.12: Moran scatterplot and LISA cluster map of HAV morbidity rate for p <0.05.

Local spatial statistics look for specific areas in an image that have clusters of similar or dissimilar values. The local Moran's index identified which provinces are clustered in terms of HAV and AD. Figure 3.13 shows the map of local Moran's I for Hepatitis A morbidity rate during 2000-2011. The results reveal that there were significant spatial clusters of HAV morbidity rate covering specific areas in each year. In 2000, 2001, 2002, 2003 and 2004 the clustered provinces with high morbidity rate were mostly concentrated in the western part of Turkey, particularly in the Aegean Region and a few provinces at the intersection of Marmara and Black Sea Region with high-high clusters. At the same time, these provinces with high HAV incidences such as Manisa (ID 45), Afyon (ID 3), Duzce (ID 81), Denizli (ID 20) and Isparta (ID 32) were surrounded by low-high provinces such as Eskisehir (ID 26), Ankara (ID 6) , Istanbul (ID 34) , Konya (ID 42) etc. indicating that these provinces had much higher HAV density than those in the neighbourhood. However, it should be noted that in the central of Turkey, there were significant low-low clusters of HAV morbidity rate surrounded by high-low outliers. Around these years, in the Eastern and South Eastern Anatolia Regions, there was no significant spatial pattern of HAV morbidity rate.

Interestingly, in 2005, the high-high clusters of HAV spread to the middle of Turkey to provinces that exhibited high-low and low-low type of relationships such as Nigde (ID 51), Yozgat (ID 66), Samsung (ID 55) and Tokat (ID 60). However, around the year 2006, the maps shows a clear spatial pattern of HAV further spreading to the provinces in the Eastern and South-Eastern Anatolia. The western part of Turkey was mostly exhibiting a low-low type of relationship indicating that the provinces had a low morbidity rate of HAV and surrounded by provinces with a similar low morbidity rate of HAV which implied clustering.

Provinces such as Kars (ID 36) and Kutahya (ID 43) has constantly shown the highest density of clustering (high-high clusters) in most recent years. While, in 2009 the densest clustering of HAV was in Sirnak (ID 73) Mus (ID 49), Bitlis (ID 13), and Afyon (ID 3) surrounded by provinces with the low-high type of relationship implying that many provinces had a low HAV morbidity rate than their counterparts. In 2011 most high HAV clusters are found in the Eastern and South Eastern Anatolia Region. From 2008-2011, southern provinces e.g Hatay (ID 31), Gaziantep (ID 27) , Osmaniye (ID 80) and Adana (ID 1) and some provinces in Marmara Region such as Edirne (ID 22), Tekirdag (ID 59) and Kirklareli (ID 39) provinces require a closer examination as

they show high-low relationship with its neighbours, depicting that these provinces have much higher HAV density than the surrounding neighbouring provinces.

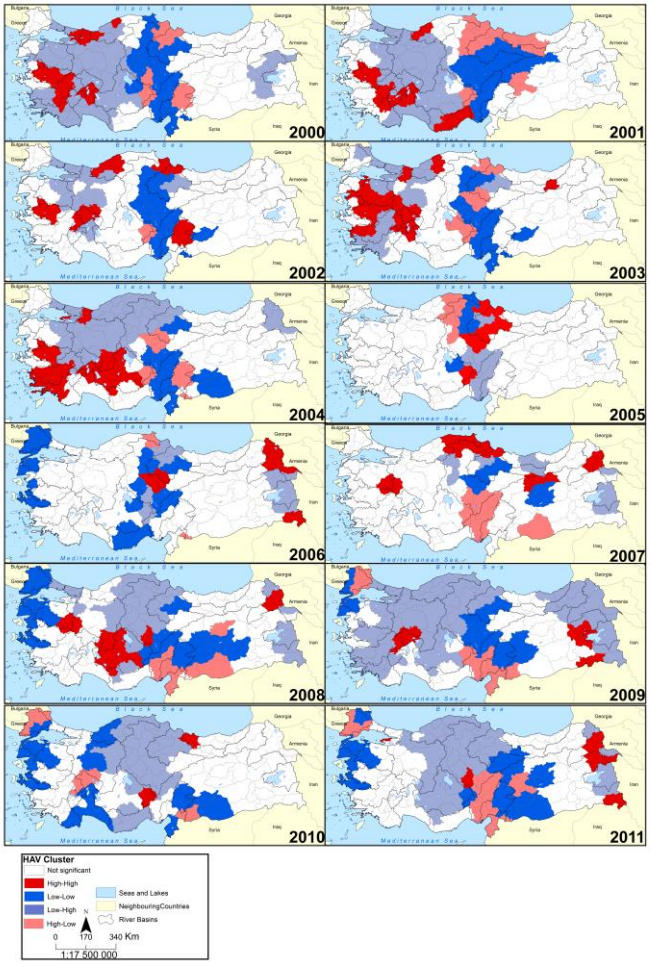

Figure 3.13: Clusters of Hepatitis A for 2000-2011.

The clusters of A Dysentery were found and illustrated in LISA cluster map in Figure 3.14. The spatial pattern of A Dysentery shows uniformity in most years. These maps shows that a in 2000 to 2009, the high-high positive spatial association (clusters) within the morbidity rate of A Dysentery were distributed across the region of Black Sea: Kastamonu (ID 37), Karabuk (ID 78), Central Anatolia: Kutahya (43), Ankara (ID 6), Nigde (ID 51), Nevsehir (ID 50), Mediterranean Region: Antalya (ID 7), Mugla (ID 48), Adana (ID 1) and in Eastern Anatolia Region: Ardahan (ID 75), Hakkari (ID 30), Artivin (ID 8). Provinces such as Kastamonu (ID 37) portrayed clusters of hotspots through all the years, expect in 2009. Low-high outliers can be seen in most location such as the western regions, Black Sea, Central and Eastern Anatolia Region illustrating that most of these provinces had a low AD density but surrounded by certain provinces with very high AD morbidity rate. It was also observed that the spatial pattern mostly spread toward the western side of the country in recent years, as

it can be seen in the maps of Figure 3.14, that around 2007-2009 the patterns in the Eastern and South-Eastern Anatolia Region of our study area was classified as insignificant.

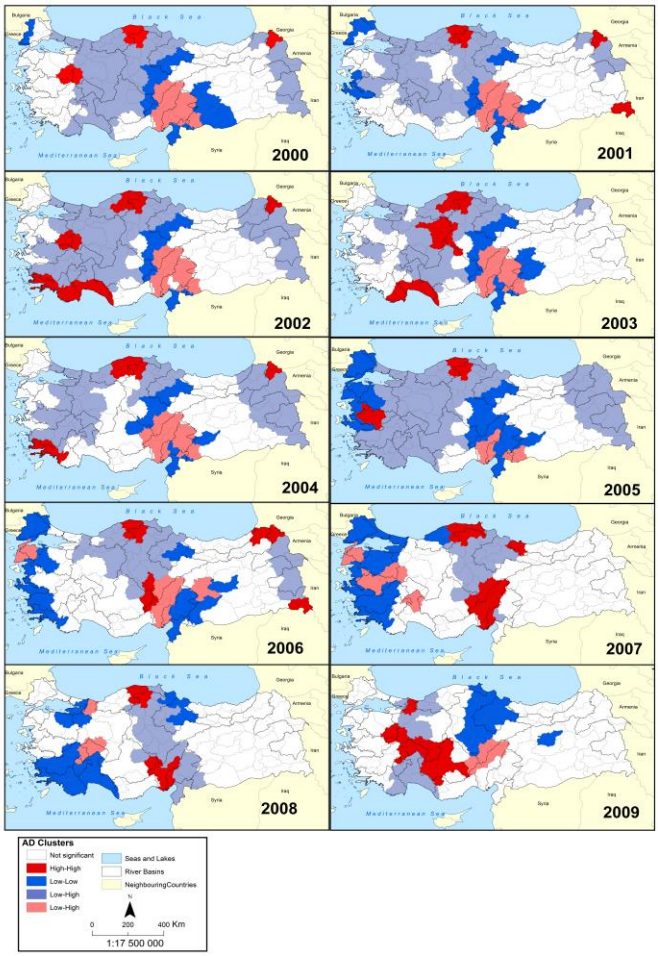

Figure 3.14: Clusters of A Dysentery for 2000-2009.

Figure 3.15 shows the clusters of HAV in children under the age of 15 during 2001 to 2011. We observe generally that cases of Hepatitis A affected most provinces in the country but only that the magnitude varies and it shows that the disease is spreading all over the study area. These maps show clear spatial patterns of Hepatitis A that were concentrated in the west, north and middle of the country in 2001-2004. The densest clustering of high-high clusters can be seen in the Aegean Region surrounded by provinces with the low-high type of relationship such as Istanbul (ID 34) among others. The middle of the country is covered in high-low outliers and clusters of low density. Around 2005-2006, the high-high clusters started migrating from the western to the eastern part of the country (Eastern Anatolia Region) in the provinces of Hakkari (ID 30), Kars (ID 36), Ardahan (ID 75) and Igdir (ID 76).

The maps for 2007 shows that Kilis (ID 79) and a few provinces in the western part of the country such as Yalova (ID 77), Kutahya (ID 43) and Kastamonu (ID 37) were still having high-high clusters. From 2008 to 2011, the high-high clusters were observed in some provinces in the western and also eastern part of the country; however, southern provinces has a new focus of low-low clusters surrounded by high-low outliers, meaning these provinces can be hotspots in the following years as they have high morbidity rate than their surrounding provinces.

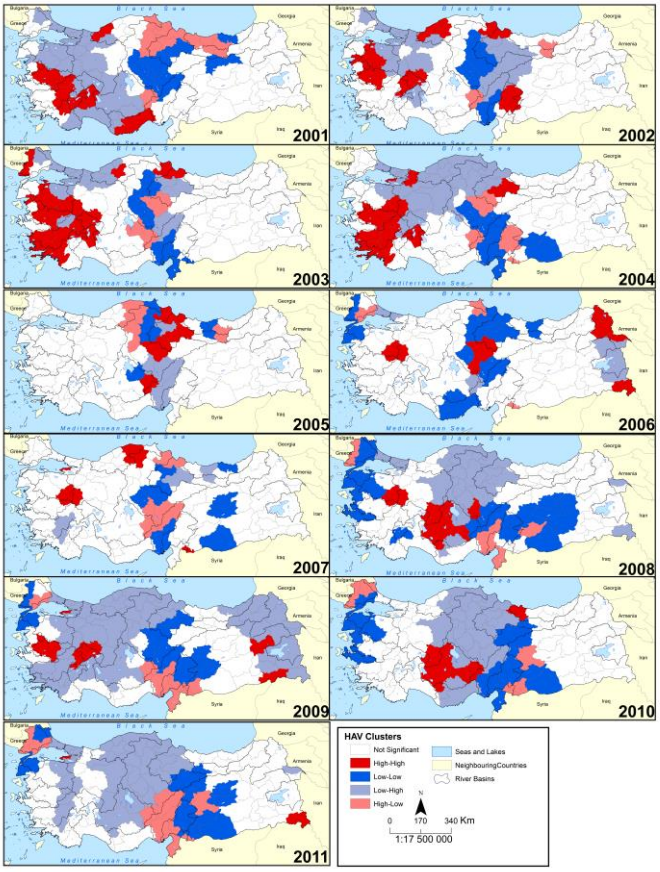

Figure 3.15: Clusters of A Hepatitis A in children under the age of 15 years from 2001-2011.

The locations for outstanding spatial clusters of HAV morbidity rate for children under the age of 15 for the three years (2001, 2006 and 2011) covering some areas are shown in Figure 3.16. In 2001, the clustered Provinces with high morbidity rates showing that the patterns of HAV were clustered in these areas were mostly concentrated in the northern and western part of Turkey. In all of the children age-groups most of the high-low outliers were found in the Black Sea Region (such as Sinop (ID 57), Samsun (ID 55), Ordu (ID 52) close to the low-low cluster like Sivas (ID 58) and Tokat (ID 60), which means that these provinces had much higher HAV concentration than the

neighboring provinces. Equally, the low-high outliers were closely found next to the high-high spatial clusters across the study area. The HAV started spreading to the provinces at the Eastern and Southern Anatolia around the year 2006. Moreover, the densest clustering (high-high values) was noticed in Hakkari (ID 30), Kahramanmaras (ID 46) and Sivas (ID 58) for the children under the age of 5 years, the similar trends were detected in Kilis (ID 79), Kars (ID 36), Ardahan (ID 75), Hakkari (ID 30) and Igdir (ID 76) for the children between the age of 5 and 15 years in 2006. Although the clusters of HAV morbidity rates in 2011 steadily attenuated in all age-groups, the high-high and high-low clusters can still be seen in some provinces in the western and south eastern provinces of Turkey.

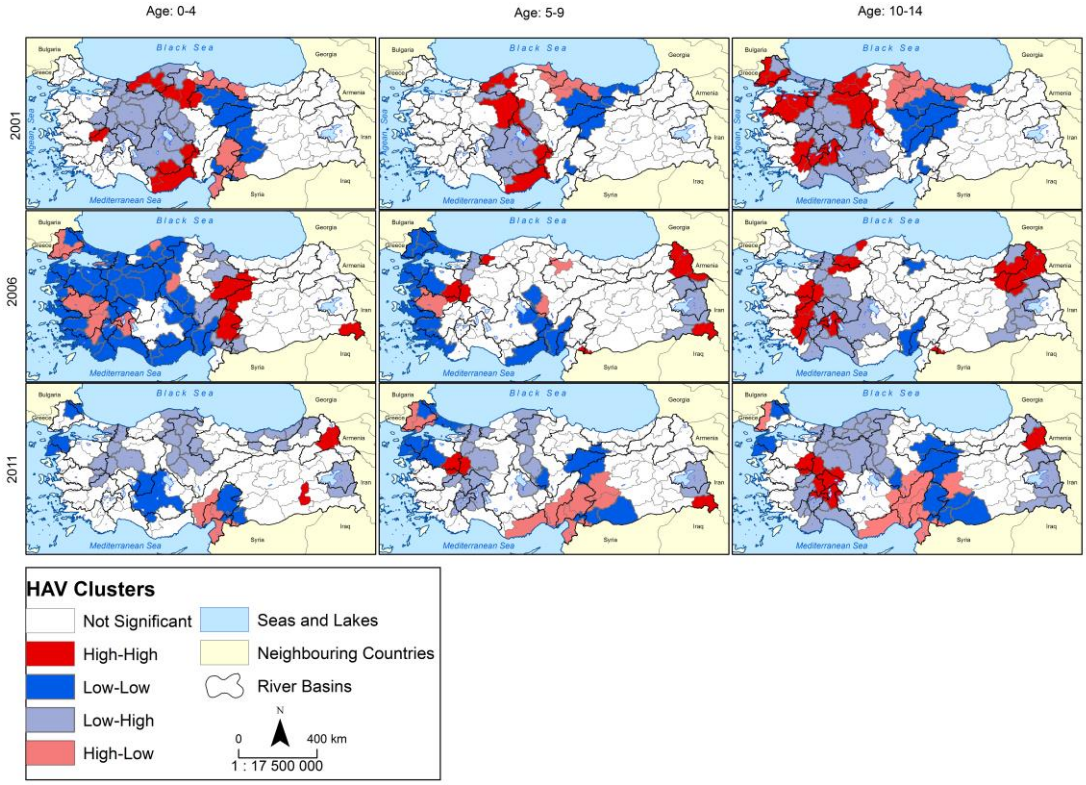

Figure 3.16: Clusters of A Hepatitis A in children for 2001, 2006 and 2011.

3.7.5 Hotspot detection (local Getis-Ord Gi* (d)

Local Getis-Ord Gi *(d) was used for further analysis of the high risks areas to justify the Anselin local Moran's I. It indicated hot spots locations giving out both high and low values hot spots of cases at province level at certain confidence intervals. The analysis results show that the larger the Z-score is, the more intense the clustering of high values (hot spot); and the smaller the Z-score is, the more intense the clustering of low values (cold spot). To be a statistically significant hot spot, a feature will have

a high value and be surrounded by other features with high values as well. In the maps (Figure 3.17, 3.18, 3.19 and 3.20), the red areas indicate statistically significant hotspots, while, blue areas represent significant cold spot areas.

According to Figure 3.17, the Getis-Ord Gi *(d) statistic results for Hepatitis A morbidity rate from 2001-2011 were in agreement with local Moran's I statistic results. The hotspots were mostly spread western part of Turkey around the year 2000-2003. Most of the hotspots with a confidence level of 99% can be found in the Black Sea Region and Central Anatolian Region for all the years, in provinces such as Samsun (ID 55), Amasya (ID 5), Tokat (60), Yozgat (ID 66) and Kayseri (ID 38). In 2006, the hotspot were intense and it showed that in this year most Turkish population had a high infection of Hepatitis A. The 99% hotspots were more prominent within the western part of Turkey, especially in the Aegean, Marmara and some part of Black Sea Region; however, they also started spreading mostly to the Eastern Anatolia Region, which was not the case in the previous years, this is also corresponding with the LISA cluster maps discussed earlier. In 2008 and 2009, the HAV hotspots seems to be concentrated in some provinces in the western, central and eastern side of Turkey, even though it can be noticed that in 2009, the hotspot of HAV morbidity rate attenuated in the western and eastern regions, with few less intense hotspots in comparison to 2008 and 2006. However, in the Central Anatolia Region, there can still be found high intense hotspots of (99%).

In 2010, the western part of the country and Hakkari (ID 30) were covered in cold spots suggesting that majority of provinces had low values of HAV morbidity rate; however in 2011, there was an increase of HAV hotspots areas in the western and eastern part of the country. But intense hotspots of 99% were still portrayed in the Black Sea and Central Anatolia Region in 2010 and 2011, just like in all the other years (see Figure 3.17).

The maps in Figure 3.18 show Getis-Ord Gi *(d) statistic results for A Dysentery cases abundance in 2000 to 2009. These maps shows the densest clustering of hotspots (99% confidence level) of A Dysentery cases that were mostly concentrated in some provinces such as Samsun (ID 55), Amasya (ID 5), Tokat (ID 60), Yozgat (ID 66), Nevsehir (ID 50)and Nigde (ID 51) and Kayseri (ID 38) situated in the Black Sea and Central Anatolia Region during 2000 to 2005 and some less intense hotspots in the Marmara and Eastern Anatolia Region in 2000, 2001 and 2005. In 2006 and 2007 the

hotspots were mostly spreading to the western part of Turkey: Mugla (ID 48), Aydin (ID 9) , Izmir (ID 35) , Istanbul (ID 34) and in the Eastern Anatolia Region: Hakkari (ID 30), Ardahan (ID 75), Agri (ID 4) etc. In 2006 and 2007 the hotspots were mostly spreading to the western part of Turkey: Mugla (ID 48), Aydin (ID 9) , Izmir (ID) , Istanbul (ID 34) and in the Eastern Anatolia Region: Hakkari (ID 30), Ardahan (ID 75), Agri (ID 4) etc. In recent years 2008 and 2009, the AD hotspots are portrayed in the central of the study area, with some less intense hotspot in the western part of the country.

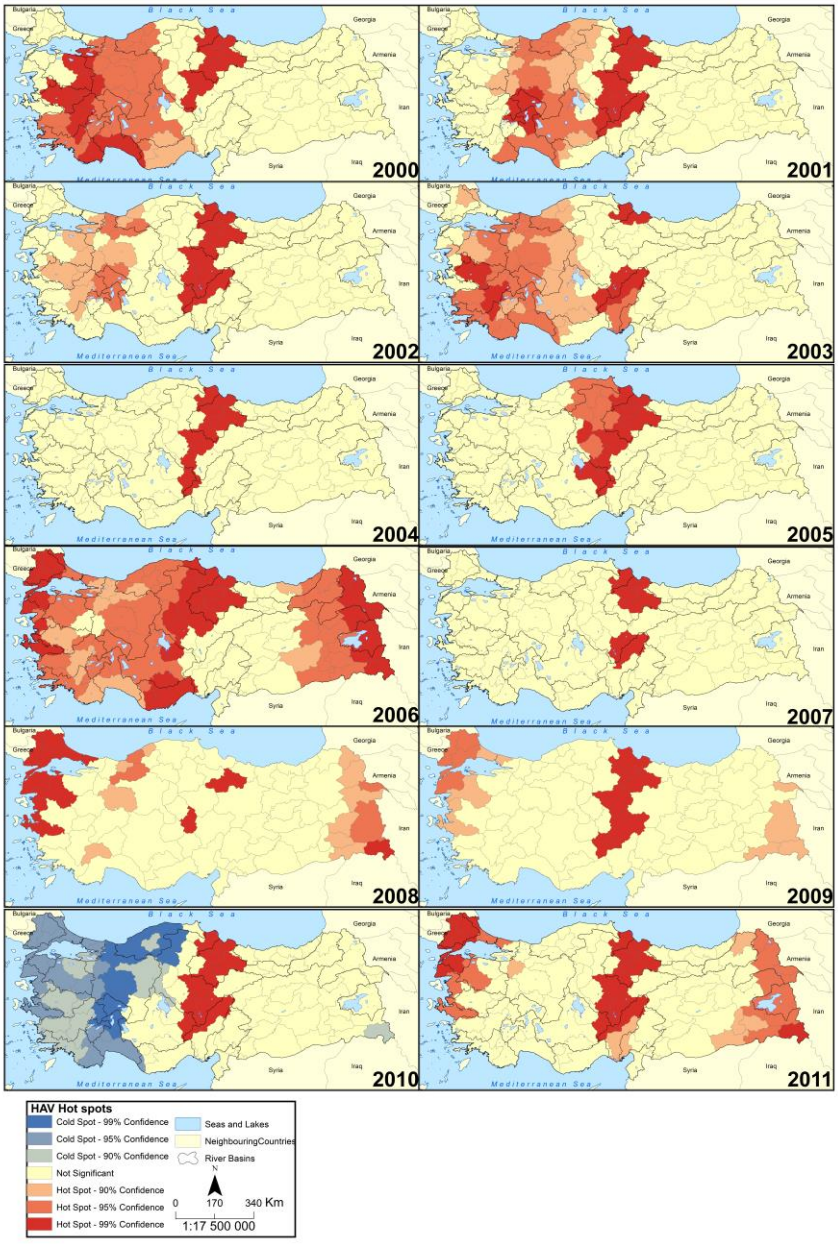

Figure 3.17: Hotspots of Hepatitis A for 2000-2011.

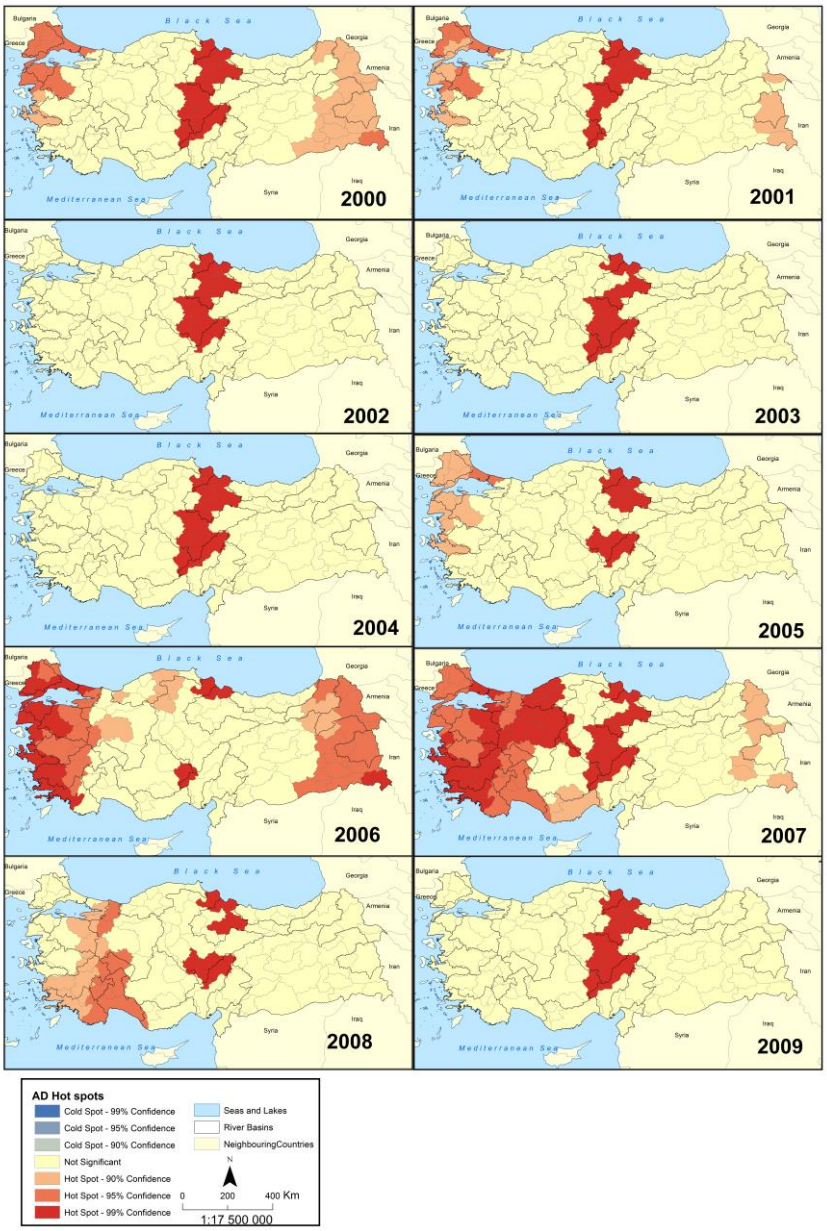

Figure 3.18: Hotspots of A Dysentery for 2000-2009.

As shown in Figure 3.19, the hotspots of Hepatitis A in Turkish children under the age of 15 were found and the results reveals that from 2001-2003 the hotspots of 99% confidence level were mostly distributed in the middle and western part of the country, and the eastern part showed less intense hotspots during the same years. In 2004-2005, hotspots were observed in the middle of the country; however, the hotspots significantly left the western part of the country but a less intense hotspot of 95% confidence level hotspots was portrayed in Aegean Region in the year 2004. For the year 2006, there was a slight increase in the HAV hotspots areas in western part of the country and hotspots areas shifted towards the Eastern Anatolia direction but the hotspots in the middle of the country were still persistent in this year. From 2007-2011,

the intense hotspots can still be observed in the middle of the country for all the years with a few less intense in the western part of the country in 2008-2010. However, in 2011, there are cold spots in the Marmara and Mediterranean Region.

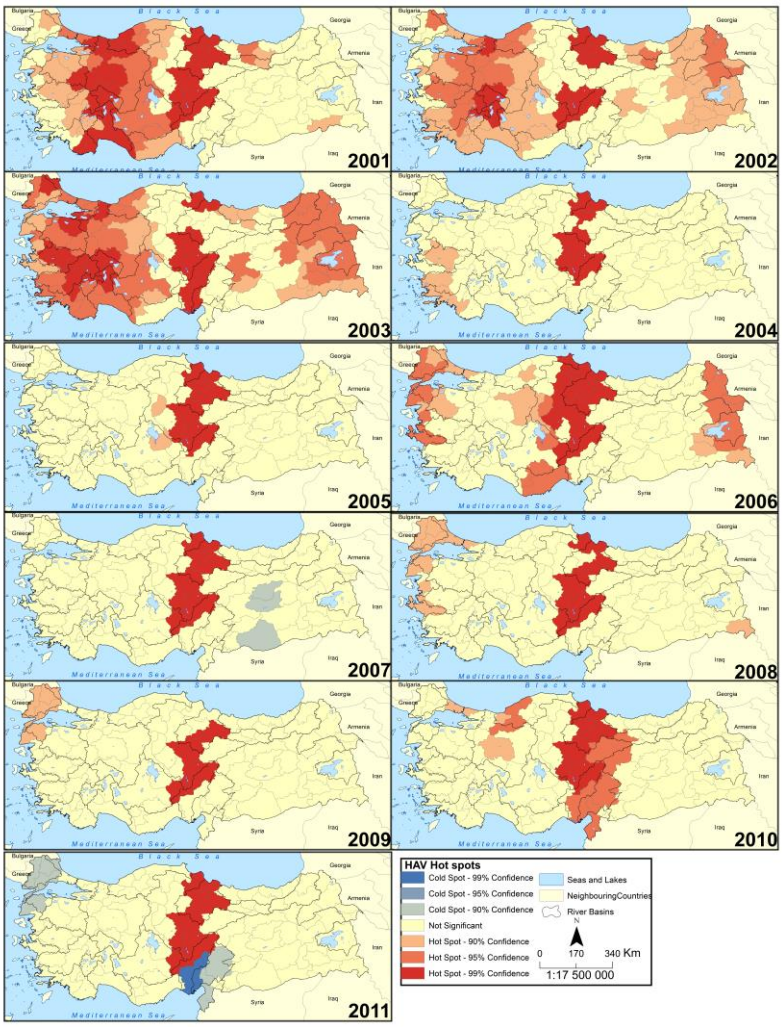

Figure 3.19: Hotspots of Hepatitis A in children under the age of 15 years for 2001-2011.

Getis-Ord Gi*(d) statistic results for children in age bracket of 0-4, 5-9 and 10-15 as illustrated in Figure 3.20 presented that the hotspots of HAV morbidity rate in children were mostly aggregated in the central and western parts of Turkey for the age-group of 5-9 and 10-14 in 2001, while age-group 0-4, there were a few stains of low-density clustering. In 2006, the western part of Turkey were covered in cold spots and the hotspots migrated to the Eastern and Southern Anatolia for the children under 10, with Hakkari (ID 30) and the nearby provinces having the confidence levels of 99% and 95% for the two age-group (0-4 and 5-9), respectively. In 2011, hotspots of morbidity rate drastically decreased, the hotspots in the Eastern Anatolia side also became less

intense (90% confidence level) for children under the age of 10 years. No hotspot clusters were detected in children between 10-14 years.

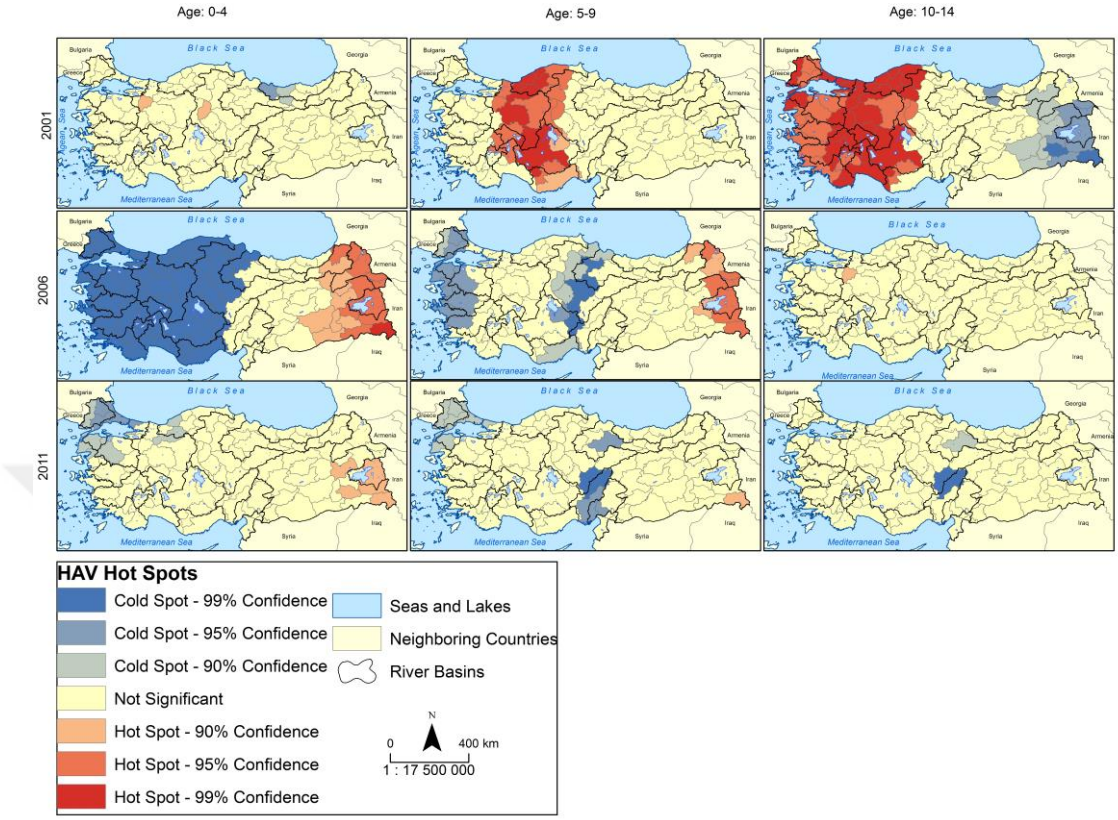

Figure 3.20: Hotspots of Hepatitis A in children for 2001, 2006 and 2011.

4. DISCUSSIONS

Due to the small number problem, visualization of raw risks from areas with low population or a small number of cases can be misleading. Estimates of Hepatitis A and A Dysentery incidence rates accounted for the variability in population distribution among provinces of Turkey. In this study, the smoothed risks did not change the raw pattern very much, except to make localized patterns more visually obvious for both Hepatitis A and A Dysentery. Apostolopoulos and Sönmez (2007) stated that Bayesian smoothing technique addresses the issue of heterogeneity in the population at risk by adjusting the raw incidence rate, taking into account information in the rest of the sample. Therefore it is a useful tool to use in explorative mapping of disease cases or incidence rates.

The study used boxplots and time-series maps as a means of visual data mining of HAV and AD morbidity rate. Visualizing HAV and AD data in this way was very useful as it displayed variation in the data and helped gain further insight into the distribution of the data. Boxplots hinted to us that there is a significant decline in Hepatitis A infection from 2009 to 2011. The same decreasing trend can be noticed also in the boxplots of A Dysentery annual morbidity rate in 2009. Boxplots were the best way to explore our data and to identify outliers in the HAV and AD morbidity rates. We were also able to identify which years had the highest infection considering the median and distribution of the data in the boxplots. It was also observed that Turkish children under the age of 15 years were most vulnerable to HAV, as they had the highest infection in comparison to those above the age of 15 years.

With regards to time series maps, the visual interpretation of spatial patterns can be strongly affected by the number and width of class intervals used to represent risk values. To reduce this potential bias, classification tools were used to classify and visualize the data geographically depending on the determined threshold values endemic, hyper-endemic, and epidemic levels by regarding the temporal trend of the disease in each province, as stated by (WHO, 1999). Time series maps unveil to us the non-uniformity in the distribution of the HAV and AD cases in both space and time

for the different years of study allowing for further statistical investigation. Time series analysis outlined sensitive provinces with the high and low risk of epidemics in the study area. In addition, time series maps showed a downward tendency in morbidity rates of Hepatitis A and A Dysentery at epidemic level from 2000 to 2011.

The extent to which neighboring values are correlated was measured using global Moran's I index. There was a positive spatial autocorrelation for HAV morbidity rate in the year 2005, 2006 and 2007 in the Turkish population as a whole, whereas, for children under the age of 15 years, Moran's I values were significant in the years 2002, 2003, 2006 and 2007; however, the patterns in the other years showed that the distribution of HAV morbidity rate was stochastic at space and time and that no clusters did exist. Consequently, the global Moran's I demonstrated a clustering pattern of AD morbidity rate in 2006 and 2007 but from 2000-2005 and 2008-2009, the patterns of AD were dispersed.

The LISAs were used to identify the HAV and AD clusters and hotspots in Turkish provinces. Since Hepatitis A and A Dysentery are waterborne communicable disease, spatial patterns were examined considering the coverages of the river basins in Turkey. The general results of the study were evaluated, and the patterns expose that from 2001-2003, the western regions were mostly affected by HAV, covering mostly the Marmara, Susurluk and Gediz River Basins. In all the years, the HAV morbidity rate were relatively high in the Black Sea Region and Central Anatolia Region, covering the Kizilirmak, Yesilmak and Sevhan River Basins in Turkey. Provinces such as Samsun (ID 55), Amasya (ID 5), Tokat (ID 60), Yozgat (ID 66) and Kayseri (ID 38) were the hotspots of these regions. From 2006, 2008-2011 the HAV cases were mostly spreading on the Eastern Anatolia side, with high HAV morbidity rate observed in the Aras, Firat-Dicle and Van Golu River Basins, and on the western part of Turkey, the hotspots were particularly in the Marmara, Menderes, Western Black Sea, Western Mediterranean and Gediz River Basins.

The findings of the HAV infections in the whole Turkish population confirms that of HAV infections in children under the age of 15 years, it can be seen that from 2001-2003 the intense hot spot locations for HAV in children were mostly located in the western river basins and eastern part of the country were covered in less intense hotspots. In general, the infections in the children also shows that the magnitude of the HAV morbidity rates in children appeared to be higher in the Aegean, Black Sea, and

Central Anatolia Region, especially in Gediz, Western Black Sea and Kizilirmak River Basins. The similar trends were found in the HAV infection of the whole population which indicated that the hotspots and clusters of HAV were almost concentrated in similar regions and river basins. Looking at the HAV results in details, we can see that young Turkish children under the age of 15 years have the highest infection, primarily because infections in this age group are usually asymptomatic and standards of hygiene are generally lower among children than among adults. Also, if not treated they are likely to be the agents of future epidemics.

In terms of AD clusters and hotspots patterns were still portrayed in the Black Sea and Central Anatolia Region covering Kizilirmak, Yesilmak and Western Black Sea River Basins, with provinces such as Samsun (ID 55), Amasya (ID 5), Tokat (ID 60), Yozgat (ID 66), Nevsehir(ID 50) and Nigde (ID 51), Kayseri (ID 38) and Kastamonu (ID 37) found to be hotspots in these regions in all the years. However, in 2006-2007 hotspots started spreading to the western part of Turkey and Eastern Anatolia Region and provinces with higher AD morbidity rate could be found in the Marmara, Menderes, Western Black Sea, Western Mediterranean and Gediz River Basins, while on the Eastern Anatolia side, high HAV morbidity rate were still observed in the Aras, Firat-Dicle and Van Golu River Basins. It was observed that the LISAs generated comparable results in detecting geographical areas of high-rate and low-rate clusters or hotspots across the study area. However, some contradictions were also noted. The local Getis-Ord Gi * statistic identified clusters in a more localized manner compared to local Moran's I statistic. The study noted that in most cases the high-low outliers in the local Moran's I generally becomes part of the hotspots in local Getis-Ord Gi * statistic. But this is not unexpected as it shows that the provinces have high incidences than the surrounding neighboring countries, which could be interpreted as having some tendency towards clustering. This might be due to their different criteria in determining the clusters.

The results are interesting because the two waterborne communicable diseases can be found almost in the same river basins which show that these infected regions and hotspots provinces can be considered as administrative units were intensive water treatment facilities is recommended. This could be the results of uncontrolled quality of water resources in those particular river basins, caused by the uncontrolled human-induced activities leading to deterioration of the environmental performance indicators

representing the case of environmental stressors, however, the quality of water in those specific river basins was not explored in the present study. The serious and strict measures began to be taken first with the enforcement of new law and regulations in connection with the River Basin Protection Plans introduced in 2009. The protection plans on the western part of the country were seen to completion in 2010, the rest were finished nearly at the end of 2013. This indicated that the transmission of infectious diseases is closely linked to the concepts of spatial and temporal proximity and hence transmission is more likely to occur if the at-risk individuals are close in a spatial and a temporal sense. The spatial distribution of Hepatitis A and A Dysentery is always correlated with socio-demographic factors, with environment and sanitation playing a huge role in the transmission. Therefore, clean water, food hygiene and environment condition are essential in controlling Hepatitis A and A Dysentery virus. It would be helpful to investigate the quality of water resources in the particular river basins identified in this study.

5. CONCLUSIONS AND RECOMMENDATION

5.1 Conclusions

This study utilized GIS in conjunction with geostatistics to explore the spatial patterns of Amoebic Dysentery for the year 2000 to 2009 and Hepatitis A from 2000 to 2011 in the Turkish population. The study addressed all the specific questions and objectives. Time series maps clearly represent the diffusion and temporal patterns of HAV and AD in Turkish provinces. The study showed that the spatial distribution was clustered in certain years 2005, 2006 and 2007 for Hepatitis A, 2006 and 2007 for A Dysentery and 2002, 2003, 2006 and 2007 for children under the age of 15 years. The spatial trends of the clusters and hotspots in the provinces were also highlighted and provinces with HAVand AD epidemics were identified and provinces with low level of HAV or AD infections but threatened by the neighboring provinces with high morbidity rate were highlighted. In considering the river basins, it could be seen that the epidemics of the disease were mainly in certain geographic region of Turkey covered by specific river basins. Water quality and quantity as the main stressors of HAV and AD epidemics should be examined in terms of integrated river basin management studies for providing an intensive treatment of contaminated water cycling in a single river basin in particular on the basis of the sewage treatment before discharge.

The study demonstrated that the use of GIS and geostatistics can take the study of disease cases beyond the mere visualization and exploration of disease maps to the modelling of disease case abundance for assessing whether observed patterns differ significantly from the spatial randomness of patterns and identification of spatial and spatio-temporal disease clusters and hotspots statistically. From a methodological standpoint, the study can point out that the use of LISA techniques was very beneficial in determining the locations of clusters, hotspots and outliers. These methods can be very useful if used complimentary rather than individually. The spatial clusters and hotspots detection techniques are more sophisticated in expansively recognizing risk

patterns of Hepatitis A and A Dysentery epidemic, rather than relying on annual cumulative incidence alone.

The findings, in terms of the presence of clusters and hotspots and how they are spreading, can help the provincial health officers and decision makers to control and predict the further HAV and AD spread over critical hotspot regions only rather than for a whole country. This may save time and most importantly cost and make public health department efforts more efficient. The results from the study can lay a foundation to pursue future investigations into the environmental factors responsible for the increased infections risk. The high morbidity rates of HAV and AD rates may be considered as a warning for the environmental performance indicators under deterioration. For example, it would be necessary to find out river basins with insufficient sewer systems as this could impose that these basins suffer mostly from the polluted well water with no water treatment facilities. On the positive side, the downward trend in the occurrence of waterborne diseases in recent years could be explained by the widespread application of best water treatment technologies within the regions of Turkey.

5.2 Recommendations

The government should investigate river basins identified with high morbidity rate for HAV and AD and contaminated basins could be analyzed based on provinces, an intensive province based treatment facilities could be initiated in each river basin.

The study has shown different regions where clusters and hotspots of HAV and AD are located across the countries. The decision makers and public health officers and other health-related organizations should consider these results when planning HAV and AD control measures. In particular, the HAV and AD infections Risk, clusters and hotspots maps should be updated on a regular basis for more effective control of these diseases.

Further studies can try to incorporate other technologies such as remote sensing in conjunction with geostatistics to investigate if water quality and quantity contribute to the high morbidity rate of the waterborne diseases or if other environmental factors such as refuse dump locations do have an effect on the spread of HAV and AD infections.

REFERENCES

Acuna-Soto, R., Maguire, J. H., & Wirth, D. F. (2000). Gender distribution in asymptomatic and invasive Amebiasis. The American journal of gastroenterology, 95(5), 1277-1283.

Alhan, E., Kozanoglu, B., Tumgor, G., Celik, U., Yaman, A., & Bozdemir, N. (2014). The epidemiological shift of hepatitis A in central Adana, Turkey. *Turk J Gastroenterol*, *25*(1), 6-8.

Alkoy, S., Ulugtekin, N., and Dogru, A.O. (2007). Monitoring vaccination coverage in Istanbul using the lot quality assurance sampling and geographic information system. The journal of international medical research, 35(2), pp. 242-252.

Altinkaynak, S., Selimoglu, M. A., Ertekin, V., & Kiliçarslan, B. (2008). Epidemiological factors affecting hepatitis A seroprevalence in childhood in a developing country. *Acta Paediatrica*, *97*, 128.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*(2), 93-115.

Anselin, L., Lozano, N., & Koschinsky, J. (2006). Rate transformations and smoothing. Urbana, 51, 61801.

Apostolopoulos, Y., & Sönmez, S. F. (2007). *Population mobility and infectious disease*. New York, NY:: Springer.

Arifin, S. N., Madey, G. R., & Collins, F. H. (2016). *Spatial agent-based simulation modeling in public health: design, implementation, and applications for malaria epidemiology* (pp. 167). John Wiley & Sons.

Avonts, D., Steylaerts, C., Van Royen, P., & Wens, J. (1999). European General practice research workshop. *Family Practice*, *16*(4), 454-462.

Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial point patterns: methodology and applications with R (*Pp 189). CRC Press.

Badie, B., Berg-Schlosser, D., & Morlino, L. (2011). *International encyclopedia of political science* (Vol. 1, pp 537-544). Sage.

Badur, S. and Nedret E. (2010, July 2). Viral Hepatitis. Burden and Prevention of Viral Hepatitis in Turkey Istanbul, Turkey, November 12-13, 2009, 18, 1-28.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis* (Vol. 413). Essex: Longman Scientific & Technical.

Bailey, T. C. (2001). Spatial statistical methods in health. *Cadernos de Saúde Pública*, *17*(5), 1083-1098.

Besag, J., & Newell, J. (1991). The detection of clusters in rare diseases. Journal of the Royal Statistical Society - Series A: Statistics in Society, 154, (1), 143-155.

Béla, M. (2010). Spatial Analysis|DigitálisTankönyvtár. Retrieved August 23, 2016, from: http://www.tankonyvtar.hu/hu/tartalom/tamop425/ 0027_SAN3/ ch01s02.

Bonita, R., Beaglehole, R., & Kjellström, T. (2006). Basic epidemiology (2nd ed, PP1-4). Geneva: World Health Organization.

Boots, B.N. & Getis, A. (1988). Point Pattern Analysis, Newburry Park, CA, Sage Publication

Boyle, P, Muir C S, Grundmann E. (1989). Cancer mapping. Berlin: Springer-Verlag.

Cannan, R.K. (1962). Amoebiasis. Section 1(infectious disease), (Chapter 1, 84-86). Tropical Health: A Report on a Study of Needs and Resources. Division of medical science, from:https://books.google.com.tr/books?id=wD8rAAAAYAAJ&printsec=frontcover#v=onepage&q&f=false

Carpenter, T. (2011). The spatial epidemiologic revolution: A look back in time and forward to the future. Spatial and Spatio-temporal Epidemiology, 2(3), 119-124.

Chaikaew, N., Tripathi, N. K., & Souris, M. (2009). Exploring spatial patterns and hotspots of diarrhea in Chiang Mai, Thailand. International Journal of Health GeographicsInt J Health Geogr, 8(1), 36. doi:10.1186/1476-072x-8-36

Cheng, W. & Washington, S. P. (2005). Experimental evaluation of hotspot identification methods. Accident Analysis and Prevention, 37(5), 870-881.

Chung, K., Yang, D., & Bell, R. (2004). Health and GIS: Toward Spatial Statistical Analyses. Journal of Medical Systems, 28(4), 349-360. doi:10.1023/b:joms.0000032850.04124.33.

Clarke, K. C., McLafferty, S. L., & Tempalski, B. J. (1996). On epidemiology and geographic information systems: a review and discussion of future directions. *Emerging infectious diseases*, *2*(2), 85.

Clayton, D. G.,& Kaldor, J. (1987). Empirical Bayesian estimates of age-standardized relative risks for use in disease mapping. Biometrics. 43:671-682. http://dx.doi.org/10.2307/2532003

Cromley, E.K. & McLafferty, S.L. (2011). *GIS and public health (*Chapter 1, pp.29). Guilford Press.

Curtis, A., Leitner, M. (2006). Geographic information systems and public health: Eliminating perinatal disparity (Pp. 178). Hershey: IRM Press.

Delmelle, A. P. E., & Kanaroglou, P. (2015). *Spatial analysis in health geography*. Ashgate Publishing, Ltd.

Desenclos, J.C., & MacLafferty, L. (1993) Community wide outbreak of hepatitis A linked to children in day care centers and with increased transmission in young adult men in Florida 1988–1989, J. Epidemiol. Commun. Health, 47: 269–273.

Devore, J. L. (2015). *Probability and Statistics for Engineering and the Sciences*. Cengage Learning.

Dogru, A. O., David, R. M., Ulugtekin, N., Goksel, C., Seker, D. Z., & Sözen, S. (2017). GIS based spatial pattern analysis: Children with Hepatitis A in Turkey. *Environmental Research, 156,* 349-357.

Elliott, P., Wakefield, J., Best, N., and Briggs, D. (2000). Spatial Epidemiology: Methods and applications. Oxford: Oxford University Press.

Farthing, M.J.G., Cevallos, A.M., & Kelly P. (1996). Intestinal protozoa. In G.C Cook (Ed).*Manson`s Tropical Diseases* (20th edition pp 1255-1298). WB Saunders Company.

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

Franco, E., & Vitiello, G. (2003). Vaccination strategies against hepatitis A in southern Europe. *Vaccine*, *21*(7), 696-697.

Gatrell Jay, D. (2010). Planning and Socioeconomic Applications.

Gatrell, A. C., Bailey, T. C., Diggle, P. J., and Rowlingson, B. S. (1996). Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology. Transactions of the Institute of British Geographers, 21(1), 256. doi:10.2307/622936.

Getis, A., Morrison, A. C., Gray, K., & Scott, T. W. (2010). Characteristics of the spatial pattern of the dengue vector, Aedes aegypti, in Iquitos, Peru. In *Perspectives on Spatial Data Analysis* (pp. 203-225). Springer Berlin Heidelberg.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical analysis*, *24*(3), 189-206.

Getis, A. & J. K. Ord. (1996). Local spatial statistics: An overview. In P. Longley and M. Batty (Eds.), Spatial Analysis: Modelling in a GIS Environment, (pp. 261–277). John Wiley & Sons.

Gill, G. V., & Beeching, N. (2011). *Lecture Notes: Tropical Medicine* (Vol. 67, pp.186). John Wiley & Sons.

Goldsmith, C. (2010). *Hepatitis (*Pp. 51-53). Twenty-First Century Books.

Goodchild, M.F & Longley, P.A. (1999). Future of GIS and spatial analysis. In P. Longley, M. Goodchild, D. Maguire, D. Rhind, *Geographical Information Systems (2nd edition*, pp 567- 580). John Wiley, New York.

Hales, S., Weinstein, P., Souares, Y. & Woodward, A. (1999). El Nino and the dynamic of vector-borne disease transmission', Environmental Health Perspectives, Vol. 107, No. 2, pp.99–102.

Hawker, J., Begg, N., Blair, I., Reintjes, R., & Weinberg, J. (2008). *Communicable disease control handbook* (pp55). John Wiley & Sons.

Hegazi, M. A., Patel, T. A., & El-Deek, B. S. (2013). Prevalence and characters of Entamoeba histolytica infection in Saudi infants and children admitted with diarrhea at 2 main hospitals at south Jeddah: a re-emerging serious infection with unusual presentation. *The Brazilian Journal of Infectious Diseases*, *17*(1), 32-40.

Heymann, D. L. (2008). *Control of communicable diseases manual* (No. Ed. 19). American Public Health Association.

Holleran, R. S. (2009). *ASTNA patient transport: Principles and practice (pp512-514).* Elsevier Health Sciences.

Hung, C. C., Chang, S. Y., & Ji, D. D. (2012). Entamoeba histolytica infection in men who have sex with men. *The Lancet infectious diseases*,*12*(9), 729-736.

Jackson, T. F. H. G. (2000). Epidemiology. *In* J. I. Ravdin (ed.), Amebiasis. (p. 47-63). Imperial College Press, London, United Kingdom.

Jackson, M.C. & Waller, L.A. (2005). Exploring Goodness -of-fit and spatial correlation using components of Tango's Index of spatial clustering. Geographical Analysis, 37(4):371-382.

Jacquez, G.M. (2008). Spatial cluster analysis. In: S. Fotheringham ., J. Wilson, (Ed.) The Handbook of Geographic information science (pp. 395-416). Oxford: Blackwell Publishing.

Jerrett, M., Burnett, R., Goldberg, M., Sears, M., Krewski, D., Catalan, R., & Finkelstein, N. (2003). Spatial Analysis for Environmental Health Research: Concepts, Methods, and Examples. Journal of Toxicology and Environmental Health, Part A, 66(16-19), 1783-1810. doi:10.1080/15287390306446.

Karadeniz, A., Alaşehir, E. A., Yeşilbağ, Z., Balıkçı, A., & Yaman, G. (2017). The seroprevalence of hepatitis A in Istanbul, Turkey. *Marmara Medical Journal*, *30*, 14-17.

Kamath, C. (2009). *Scientific data mining: a practical perspective (*pp210-211). Society for Industrial and Applied Mathematics.

Knox, E.G. (1989). Detection of clusters. In Elliott, P, Methodology of inquiries into disease clustering. Small Area Health Statistics Unit, London, pp.17-20

Krèamer, A., Kretzschmar, M.., & Krickeberg, K. (2010). *Modern Infectious Disease Epidemiology: Concepts, Methods, Mathematical Models* (pp3-4). Springer.

Kurt, O., Demirel, M., Ostan, I., Sevil, N. R., Mandiracioglu, A., Tanyuksel, M., & Dagci, H. (2008). Investigation of the prevalence of amoebiasis in Izmir province and determination of Entamoeba spp. using PCR and enzyme immunoassay. *The new microbiologica*, *31*(3), 393.

Kvanli, A., Pavur, R., & Keeling, K. (2005). *Concise managerial statistics* (pp84-86). Cengage Learning.

Lai, P. C., So, F. M., & Chan, K. W. (2008). *Spatial epidemiological approaches in disease mapping and analysis* (Chapter 1, pp9-13, 44-48). CRC Press. From

Lawson B.A. (2009). Disease Cluster Detection. Interdisciplinary Statistics Bayesian Disease Mapping Hierarchical Modeling Chapter 6. Pp 119-150. (). CRC Press.

Lawson, A.B., & Kulldorff, M. (1999). A review of cluster detection methods. In: A, Lawson, A, Biggeri, D, Böhning, E, Lesaffre, J.F, Viel, R, Bertollini, Disease Mapping and Risk Assessment in Public Health, (pp. 143–149). John Wiley & Sons, Chichester.

Lawson, A.B. (2013). *Statistical methods in spatial epidemiology* (Chapter 4, pp 41-43). John Wiley & Sons.

Lawson, A. B., Banerjee, S., Haining, R. P., & Ugarte, M. D. (2016). *Handbook of Spatial Epidemiology* (pp 453-454). Chapman and Hall/CRC.

Lee, J., & Wong, D. W. (2001). Point Descriptors. In Statistical Analysis with ArcView GIS, (pp. 78-82). New York: John Wiley.

Li, R., Li, H., & Qi, Z. (2015). Bacillary and Amebic Dysentery. In *Radiology of Infectious Diseases (Vol 2*, pp. 11-36). Springer Netherlands.

Longley, P. A., & Batty, M. (1996). *Spatial analysis: modeling in a GIS environment (*pp 322). John Wiley & Sons.

Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (2003). Geographic Information Systems and Science, John Wiley & Sons.

Lu, P. X., & Zhou, B. P. (2015). *Diagnostic Imaging of Emerging Infectious Diseases* (pp. 123). Springer.

Markell, E.K., John, D.T., Krotoskı, W.A. (1999). Lumen dwelling protozoa. In S. Ozmat, Markell and Voge's *Medical Parasitology* (8th edition, pp24-89). Saunders Company. Mexico.

Mathur, M. (2015). Spatial autocorrelation analysis in plant population: An overview. Journal of Applied and Natural Science, 7(1), pp. 501-513.

Mitchel, A. (2005). The ESRI Guide to GIS analysis, spatial measurements and statistics. ESRI Guide to GIS analysis, (Vol 2, pp. 204-208.

Moore, E. A. (2006). *Hepatitis: Causes, Treatments and Resources* (pp.12). McFarland.

Moore, D. A., & Carpenter, T. E. (1999). Spatial analytical methods and geographic information systems: Use in health research and epidemiology. Epidemiol. Rev. 21:143–161

Moran, P.A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17-23.

Moreno-Díaz, R., Pichler, F., & Quesada-Arencibia. (2012). *Computer Aided Systems Theory--EUROCAST 2011: 13th International Conference, Las Palmas de Gran Canaria, Spain, February 6-11, 2011, Revised Selected Papers* (Vol. 6927). Springer Science & Business Media.

Nelson, K. E., & Williams, C. (2013). *Infectious disease epidemiology* (pp. 3-5). Jones & Bartlett Publishers.

Osei, F. (2014). Current Statistical Methods for Spatial Epidemiology: A Review. *Austin Biometrics Biostat, 1*(2), 1-7.

Pfeiffer, D. U. (1996). Issues related to handling of spatial data. *Massey University, Palmerston North*, 309-329.

Pfeiffer, D. (2008). Spatial analysis in epidemiology (pp.1-15). Oxford: Oxford University Press.

Pfeiffer, D. (2008). Spatial analysis in epidemiology (pp. 32-49, 67-80). Oxford: Oxford University Press.

Plotkin, S. A., Orenstein, W. A., & Paul, A. (2008). Offit. *Vaccines andautism: Atale ofshifting hypotheses." Clinical Infectious Diseases48*, (4), 456-61.

Quarto, M., & Chironna, M. (2004). Hepatitis A: Sources in Food and Risk for Health. In V.R. Preedy, R.R Watson, *Reviews in food and nutrition toxicity*, (vol. *2*, Chapter 4, pp 91-118). CRC Press.

Requia, W., and Roig, H. (2015). Analyzing Spatial Patterns of Cardiorespiratory Diseases in the Federal District, Brazil. Health, 07(10), 1283-1293. doi:10.4236/health.2015.710143

Rogerson P. (1999). The detection of clusters using a spatial version of the Chi-Square Goodness-of-Fit Statistic. Geographical Analysis, 31(1):128-147.

Rosenthal, G., & Rosenthal, J. A. (2011). *Statistics and data interpretation for social work* (pp 227-237). Springer Publishing Company,

Rytkönen, M. J. (2004). Not all maps are equal: GIS and spatial analysis in epidemiology. International Journal of Circumpolar Health, 63(1).

Salkind, N. J. (2006). *Encyclopedia of measurement and statistics* (pp 111-113). Sage Publications.

Sánchez, G. (2013). *Hepatitis A virus in food: detection and inactivation methods* (p. 1-10). Springer.

Sastry, A. S., & Bhat, S. (2014). Essentials of medical parasitology (Pp. 27). New Delhi: Jaypee Brothers Medical (P) Ltd.

Shankar, P., Mishra, J. & Singh, S. (2014). Hepatitis A and E in Potable Water: A Threat to Health. In *Water and Health* (pp. 29-51). Springer India.

Solaymani-Mohammadi, S., & Petri Jr, W. A. (2008). Intestinal invasion by Entamoeba histolytica. In *Molecular Mechanisms of Parasite Invasion* (pp. 221-232). Springer New York.

Spear, M. E. (1952). *Charting statistics*. McGraw-Hil, New York.

Staes, C. J., Schlenker, T.L., Risk, I., Cannon, K.G., Harris, H., Pavia, A.T., Shapiro, C.N., and Bell, B.P. (2000). Sources of infection among person with acute hepatitis A and no identified risk factors during a sustained community-wide outbreak, Pediatrics, 106(4): E54.

Stanley, S. L. (2003). Amoebiasis. *The Lancet, 361*(9362), 1025-1034.

Stuart, M. C., Kouimtzi, M., & Hill, S. R. (2009). *WHO model formulary 2008* (pp.178). World Health Organization.

Suchy, F. J., Sokol, R. J., & Balistreri, W. F. (2001). *Liver disease in children* (pp365-370). Lippincott Williams & Wilkins.

Tahan, V., Ozdogan, O., & Tozun, N. (2003). Epidemiology of viral hepatitis in the Mediterranean basin. *Rocz Akad Med Bialymst, 48,* 11-7.

Tilak, KVGK. (2013). Amoebiasis. Section 1(infectious disease), (vol. 23, chapter 1, pp. 1-3). Muruganathan, A, JAYPEE BROTHERS MEDICAL PUBLISHERS PVT LTD. From http://www.apiindia.org/medicine_update_2013/chap01.pdf

Tim, U. S. (1995). The application of GIS in environmental health sciences: opportunities and limitations. *Environmental Research, 71*(2), 75-88.

Thakur, S., & Hanson, A. J. (2010). A 3D visualization of multiple time series on maps. In *Information Visualisation (IV), 2010 14th International Conference* (pp. 336-343). Jul 26, IEEE.

Tsai, P., Lin, M., Chu, C., & Perng, C. (2009). Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. BMC Public Health, 9(1). doi:10.1186/1471-2458-9-464.

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

Toprak, D., and Erdorgan, S. (2008). Spatial analysis of the distribution of typhoid fever in Turkey. *Int Arch Photogramm Remote Sens Spat Inf Sci*,*37*, 1367-1371.

Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley, Reading, MA.

Turkington, C., & Ashby, B. (2007). *The encyclopedia of infectious diseases* (pp. 5-9. Infobase Publishing.

Ulugtekin, N., Alkoy, S., Seker, D.Z. (2007). Use of Geographic Information Systems in an Epidemiological Study of Measles in Istanbul. The Journal of International Medical Research, Vol.35:1, pp.150-154.

Ungan, M., Yaman, H., Tahari, N. (1998). European General Practice Research Workshop, Crete, Greece.

Van den Berg, R. H., & Viljoen, M. J. (1999). *Communicable diseases: A nursing perspective (pp196-199)*. Pearson South Africa.

Vopham, T., Brooks, M.M., Yuan, J.M., Talbott, E.O., Ruddell, D., Hart, J.E., Chang, C.H.,& Weissfeld, J.L. (2015). Pesticide exposure and hepatocellular carcinoma risk: A case-control study using a geographic information system (GIS) to link SEER-Medicare and California pesticide data, Environmental Research, Vol. 143, Part A, November 2015, pp. 68-82.

Wakefield, J.C., Kelsall, J.E., & Morris, S.E. (2000). Clustering, cluster detection and spatial variation in risk. In P, Elliott, J.C, Wakefield, N.G, Best, D.J Briggs, Spatial Epidemiology – Methods and Applications (pp. 128–152). Oxford University Press, Oxford

Waller, L. A., & Gotway, C. A. (2004). Applied spatial statistics for public health data (Vol. 368). John Wiley & Sons.

Walker, D.H., Barbour, A.G., Oliver, J.H., Lane, R.S., Dumler, J.S., Dennis, T.D., Persing, D.H., Azad, A.F. & McSweegan, E. (1996). Emerging bacterial zoonotic and vector-borne diseases: Ecological and epidemiological factors', Journal of the American Medical Association, Vol. 275, No. 6, pp.463–469.

Waller, L. A., & McMaster, R. B. (1997). Incorporating indirect standardization in tests for disease clustering in a GIS environment. *Geographical Systems*, *4*, 327-342.

WHO. (1999). WHO Guidelines for epidemic preparedness and response to measles outbreaks, WHO/CDS/CSR/ISR/99.1. Geneva, Switzerland: May.

WHO Surveillance Programme. (1999-2000). WHO Surveillance Programme for Control of Foodborne Infections and Intoxications in Europe, Country Reports: Turkey (Vol. 8th, pp. 1-4, Rep. No. 8th). From http://www.bfr.bund.de/internet/8threport/CRs/tur.pdf

WHO. (2011). Technical guidance on water-related disease surveillance.

Yiannakoulias, N., Rowe, B. H., Svenson, L.W., Schopflocher, D. P., Kelly, K., & Voaklander, D. C. (2003). Zones of prevention: The geography of fall injuries in the elderly. *Soc. Sci. Med.* 57(11):2065–2073, 2003.

Young, F. W., Valero-Mora, P. M., & Friendly, M. (2011). *Visual statistics: seeing data with dynamic interactive graphics* (Vol. 914i Pp 193-195). John Wiley & Sons.

Zulu, L. C., Kalipeni, E., & Johannes, E. (2014). Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994-2010. *BMC infectious diseases*, *14*(1), 285.

Url<*http://www.chp.gov.hk/files/pdf/prevention_and_control_of_amoebic_dysentery_in_hong_kong.pdf/*>, date retrieved 29.06.2006.

Url-2<*https://www.boundless.com/microbiology/textbooks/boundless-microbiology-textbook/diseases-15/fungal-and-protozoan-diseases-of-the-digestive-system-185/amoebic-dysentery-amoebiasis-928-6849/*>, date retrived 24.03.2017.

CURRICULUM VITAE

Name Surname **:** Ruusa-Magano DAVID

Place and Date of Birth **:** Namibia, 10.07.1990

E-Mail **:** maganodavid2@gmail.com

EDUCATION **:**

- **B.Sc.** **:** April 2014, Polytechnic of Namibia, System administration and Networks, Department of Information Technology.
- **Diploma** **:** April 2010, Polytechnic of Namibia, Diploma in Information Technology, Department of Information Technology.

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2011-2013 Telecom Namibia, an intern in Geographical information Technology.
- 2013-2014 Telecom Namibia, IP/IT trainee.
- 2011 Best 3rd-year student in the program computer systems and Networks, Polytechnic of Namibia.

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Dogru A.O., David R.M., Ulugtekin N., Goksel C., Seker D.Z., Sozen S.** 2017: GIS-Based Spatial Pattern Analysis: Children with Hepatitis A in Turkey, Environmental Research 156 (2017) 349–357.
- **David, R.M., Dogru, A.O.** 2016: Geostatistical Analysis of Environmental Epidemiology. Proceedings of the Selçuk International Scientific Conference On Applied Sciences, pp.217-226, September 27-30, 2016, Antalya-Turkey.
- **Dogru A.O., David R.M., Ulugtekin N., Goksel C., Seker D.Z., Sozen S.** 2016: Spatial Pattern Analysis of Hepatitis A Cases in Children using Geostatistical Analysis of GIS, the 8th International Conference on Children's Health and Environment, Barcelona Spain, The Journal of Health and Pollution , pp.38-39.