Sudan University of Science and Technology

College of Graduate Studies

كلية الدراسات العليا

Spatial analysis model for Predication

Case study: spread of mosquitoes that transmit Dengue fever disease in the municipalities of Jeddah Province during 2018

نموذج التحليل المكاني للتنبؤ بظاهرة توالد البعوض الناقل حمى الضنك ببلديات محافظة جدة للعام ٢٠١٨م

A thesis submitted for Ph.D. in Statistics

Prepared by:

Farah Abdullah Mohammed El Hassan

Supervisor:

Dr. Ahmed Mohammed Abdulla Hamdi

Associate Professor of Statistics

Co. Supervisor:

Dr. Mohammed Elameen Essia Qurashi

Date M:2021

قال تعالى:

(إِنَّ اللَّهَ لا يَسْتَحْيِي أَنْ يَضْرِبَ مَثَلاً مَا بَعُوضَةً فَمَا فَوْقَهَا فَأَمَّا الَّذِينَ آمَنُوا فَيَعْلَمُونَ أَنَّهُ الْحَقُّ مِنْ رَبِّهِمْ وَأَمَّا الَّذِينَ كَفَرُوا فَيَقُولُونَ مَاذَا أَرَادَ اللَّهُ بِهَذَا مَثَلاً يُضِلُّ بِهِ كَثِيراً وَيَهْدِي بِهِ كَثِيراً وَمَا يُضِلُّ بِهِ إِلَّا الْفَاسِقِينَ)

صدق الله العظيم

البقرة، الآية(٢٦).

Dedication

To spirit of my Mother, Allah's Mercy upon her.

Acknowledgement

I really express my deep thanks gratitude to Allah for completing this research, also my full thanks to my supervisors Dr. Ahmed Mohammed Abdulla Hamdi, Dr. Mohammed Elameen Essia Qurashi, for their supervision. Special thanks to my teacher Dr. Mubarak Alhafian (Associate Professor of Statistics, King Abd-Alaziz University, KSA), also my great thanks to my friend Eng. Youssif Alsammany (Asma Environmental Solutions Company –Jeddah City) and all his colleges for help in data collection. And my college and special friend Mohammed Omer Mussa) for his supporting's.

Abstract

The method of Geographical Information System is used in building statistical samples that work to estimate phenomena affected by the surrounding environmental factors such as mosquitoes carrying dengue fever, which has a role in putting plans for decision makers to fight such mosquitoes, and that is by knowing its distribution pattern and spatial regression method.

The study aimed at examining mosquitoes carrying dengue fever reproduction patterns to assist the decision makers in Jeddah Municipality, and to know the geographical places where they concentrate inside University Municipality (the place of current study). The distribution pattern had been known in addition to the availability of spatial correlation for mosquitoes' reproduction, in addition to build spatial model for this phenomenon inside University municipality. The study also aimed at knowing the quality of geographical information system in specifying the places and reasons of spread of mosquitoes carrying dengue fever and other similar phenomenon inside Jeddah City. Two factors had been determined as independent variables; temperature and humidity which have impact on the reproduction of mosquitoes carrying dengue fever inside the municipality.

The data of average numbers of mosquitoes carrying dengue fever have been collected during the weeks (1 to 45) of the year 2018, in addition to the averages of temperature and humidity during the same period.

The most important results reached that the phenomena of mosquitoes carrying dengue fever reproduction is randomly distributed. Al-Thaghr and University districts are the most districts where dengue fever mosquitoes are reproduced. The distribution pattern is not spaced (closed) and not randomly distributed. The highest determination coefficient for all estimated models for weighted spatial regression was R^2 = 55%, i.e. 55% of variation occurred to mosquito reproduction can be resorted to the factors (temperature and humidity), and the rest 45% is resorted to other environmental factors such as (water tanking, sanitation, and other customs and traditions), since University municipality has various numbers of foreigners from different countries. The most important recommendation of the study is to concentrate on (University and Al-Thaghr districts) by taking many measurements to study the factors affecting the reproduction of mosquitoes carrying dengue fever to be generalized to all Jeddah municipalities. The research provides a good tool for the decision makers and assists them to determine the areas which need fighting operations more than the others.

ملخص الدراسة:

أسلوب نظم المعلومات الجغرافية يستخدم في بناء عينات إحصائية تعمل على دراسة وتقدير الظواهر المتأثرة بالعوامل البيئية المحيطة بها مثل ظاهرة البعوض الناقل لحمي الضنك، والذي بدوره يسهم في وضع الخطط لدى متخذي القرار لمكافحته، وذلك بمعرفة نمط التوزيع وكذلك أسلوب الانحدار المكاني.

هدفت الدراسة في التعرف على معرفة نمط تكاثر البعوض الناقل لحمى الضنك بغرض مساعدة متخذي القرار بأمانة جدة ومعرفة أماكن تركز الظاهرة بداخل بلدية الجامعة (مكان تطبيق الدراسة الحالية) فتم التعرف على نمط التوزيع ومدى توفر الارتباط المكاني لتكاثر البعوض، وأيضاً بناء نموذج مكاني لهذه الظاهرة بداخل بلدية الجامعة. والتعرف كذلك على مدى جودة نظم المعلومات الجغرافية في تحديد أماكن وأسباب انتشار ظاهرة البعوض والظواهر المشابهة لها بداخل مدينة جدة، تم تحديد عاملين هما درجتي الحرارة والرطوبة كمتغيرات مستقلة لها تأثير على توالد وتكاثر البعوض الناقل لحمى الضنك بداخل البلدية. تم تجميع بيانات لمتوسط أعداد البعوض الناقل لحمى الضنك ببلدية الجامعة خلال الأسابيع (١ إلى ٤٥) من العام ٢٠١٨م، بالإضافة لمتوسطي درجة الحرارة والرطوبة لنفس الفترة.

نجد أهم ما خرجت به الدراسة تمثل في أن ظاهرة توالد البعوض الناقل لحمى الضنك موزع عشوائياً، وأن الاحياء (الثغر والجامعة) من أكثر الاحياء يتكاثر بهما البعوض الناقل لحمى الضنك، وأن نمط التوزيع غير متباعد(متقارب) وغير موزع عشوائياً، وأن أعلى معامل تحديد لجميع نماذج الانحدار المكاني الموزون المقدر لأحياء بلدية الجامعة بلغ (٥٥%) أي أن ما يعادل ٥٥% من العوامل المؤثرة على توالد البعوض يعود إلى عاملي (الحرارة والرطوبة) بينما ٤٥% تعود إلى عوامل بيئية أخرى مثل (طريقة تخزين المياه والصرف الصحي بالإضافة إلى العادات والتقاليد الأخرى) سيما أن هذه البلدية يقطنها مجموعة كبيرة من المقيمين من شتى أنحاء العالم.

اهم ما أوصت به الدراسة يجب أن التركيز على الاحياء (الثغر والجامعة) بأجراء العديد من القياسات لمعرفة العوامل التي تؤدي لتكاثر البعوض الناقل لحمى الضنك والتي يمكن أن تعمم على جميع بلديات محافظة جدة.

Table of contents

List of tables

List of figure.

Chapter One

Introduction

1.0. Background

Statistical methods - at most - study the variables of different phenomena, to explore the relationship between them, and the degree of its positive or negative influence on each other, and how to deal with so that to decrease the negative phenomena and to support the positive factors by taking the suitable procedures for each variable. Also, to answer the questions set by the researchers and interested in the different fields, and why these phenomena occur? Its sizes? What are the most influential phenomena? Can we predict these phenomena in the future? That is by using the suitable methods according to the available data.

It is clear that spatial has a role in the spread of many phenomena increasingly or decreasingly, as example (rainfall amounts, crimes, economic growth per capita, spread of disease and epidemic, … etc.) a lot of these phenomena influenced by spatial, so the researcher opted to study the statistical methods concerned with spatial, which called Spatial Analysis. This method allows us to know the role of spatial in which the phenomena spread and how they related to it, and the shape of its spread in specific area, and that is by using the methods of Spatial Autocorrelation by studying its types, meanings and the methods of finding it. That is in addition to Spatial Autoregressive in comparison with Multiple Regression method.

Geographical data are like other data. Data distribution is considered the essence of data analysis whatever their type, Geographical or non-Geographical on which the analysis processes are based such as the distribution of certain patterns or the environmental phenomenon or demographic phenomenon or human conduct in different geographic areas. (Goodchild,1992).

For this reason, the need to know the patterns or forms taken by the phenomenon in a certain geographical place and study its reasons, or it is just a random distribution founded by luck and chance? Need to know the necessary conditions to specify the locations where certain specifications are available.

The studies do not include the spatial influence of the environmental phenomena and only become satisfied with their computing, analysis and circulation without giving attention to the fact that the area may be effected by the environmental factors, population density and other factors which assist generally in the spread of insects which lead to the difference in density from one place to another.

In addition of the change that occurred on the environment and the increase in population which effect negatively or positively on the natural environmental phenomenon within the extend geographical area whereas a need to show the effect of the spot phenomena protrudes on the increase or decrease of the phenomena on a specific geographical spot; and to what extend its difference on the other spots. This kind of regression is called multi spatial regression which is an attempt to find an equalization to the regression which enters dimensional spatial (spot's abscissas) within it, it cares of element value or that appeared in a specific geographical point or the spatial spot (x, y) whereas it finds the equalization of the regression appliances to apparent spots.(4)(Mauricio Bini, L., Diniz-Filho, J. A. F., Rangel, T. F., Akre, T. S., Albaladejo, R. G., Albuquerque, F. S., ... & Isabel Bellocq, M. (2009). Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography, 32(2), 193-204).

1-1: Study Problem.

Many of the phenomena are spread in certain areas and not others. That may be due to the geographical and environmental or demographic factors which contribute to the distribution and prevalence rates in these areas. A matter which requires study and develop solutions and future perceptions to them, so, many researches and studies concern with data collection regardless the effect of spatial and location, especially the geographical researches which employ both the geographical and descriptive statistical methods, while the spatial analysis models examine the relationship and the influence of spatial on the prevalence of these phenomena, so, the study problem lies therein for identifying the statistical models that the spatial component is considered in its calculations, identifying its statistical implications and comparing it with statistical methods that do not consider the spatial component as one of its variables by using the common statistical methods for comparison between the statistical models.

From all the above, it is clear that the study problem represents as an attempt to build a statistical model for spatial analysis showing the reality of spread of mosquitoes carry disease in Jeddah municipalities, and the most locations where it spreads widely, and how they are distributed and to identify the role of spatial in this spreading.

The fact that the descriptive spatial analysis method gives a pattern of the phenomenon increase and does not give an indication for the distribution type whether it is random or cluster, by exploring the mosquito breeding distribution patterns.

We find that phenomenon usually is affected by the environmental factors such as rains, temperatures, humidity and the like, but these factors are vary from an area to another the thing that affect the phenomenon (might refers to behavior attitude and people's tradition. all those reasons led to existence of complicated relationship when gathering samples that ignore the spatial dimension (the spot) for each single sample and does the result of that various responds according to the spot.

Also we find that each municipal has its own many subordinate factors regarding mosquitoes multiplication such as environmental factors and population demographic; whereas the need to use statistical methods that include the spot in the statistical operations came out which participate in estimating the numbers of mosquitoes that carry Dengue Fever in every single spot within the Municipal of the Campus's neighborhood.

The phenomena of mosquitoes' reproductions and multiplication, in particular that carrier to Dengue Fever, within the neighborhood of Jeddah Municipal could be moving between neighboring neighborhood because of the spatial neighboring with the abundance of all fighting methods within any single neighborhood. We find that each municipal has its own many subordinate factors regarding mosquitoes multiplication such as environmental factors and population demographic; whereas the need to use statistical methods that include the spot in the statistical operations came out which participate in estimating the numbers of mosquitoes that carry Dengue Fever in every single spot within the Municipal of the Campus's neighborhood.

1-2 Study importance:

The study importance lays where that it discusses the spatial statistical methods represented in the equalization of the multi spatial regression (Geographical weighted Regression), in addition to identifying the breeding pattern of mosquitoes that transmit dengue fever ,and showing the mechanism of using Geographical Informational Systems which is known by Modeling Spatial Relationship; through which diagnosing the accurate for the health positions to fight mosquitoes that carry Dengue Fever within the university Municipal. This spot is estimated through knowing to what extend these services are same comparing to different criterions using statistical analysis methods and GIS, which are totaled as follow:

1. Building a spatial model that represent the phenomena of mosquito's multiplication to each spot of the university Municipal neighborhood.
2. Showing the methods of diagnosing the qualities of the spatial samples and interpreting them.
3. Studying and determining the reasons that lead to mosquitoes carry Dengue Fever within the University Municipal in Jeddah City.

1-3 Objectives of the Study.

This research aims at achieving the following objectives:

- To build a spatial analysis model which can help accurately in the future prediction for mosquitoes spreading.
- To identify the probability distribution for mosquitoes reproduction
- To study the efficiency of spatial analysis models for predicting the spread of mosquitoes.
- To study the mathematical and statistical methods of spatial analysis.
- To configure database for the locations of reproduction and spreading of mosquitoes, regarding the number of locations and types of diseases they transmit.

- To identify environment of GIS software to build spatial statistic models and other spatial statistic.

1-4: Research Hypotheses.

- The models of spatial analysis of mosquitoes' reproduction is significant.
- There is a spatial correlation of the mosquito reproduction phenomenon in the university municipality.
- The distribution pattern of mosquito's reproduction phenomenon in the university municipality is random (spaced).
- There is relationship between mosquito's reproduction and environmental factors (temperature and moisture rate).
- Find the spatial correlation for mosquito's reproduction phenomenon in University municipality.
- Find the Distribution pattern of mosquito's reproduction in university municipality is moral.

1-5: Research Methodology.

In this research, the analytical method will be used, which represents in studying the spatial analysis model and to derive conclusions for the purpose of prediction. The descriptive method will be used for data presentation of the applied study such as some descriptive measurements and graphs to identify the general characteristics of the study data.

1-6: Study Boundaries.

Place Boundary: The place boundary of this research is the University municipality it's one of Jeddah city, which are fourteen municipalities.

Time Boundary: The time boundary of this research is the period (1 to 45) weeks from year 2018.

1-7 The Data.

The research is based on the daily data about mosquitoes transmit diseases in its phases (larval and adult); and use these data to spatial analysis and build a spatial model during the weeks (1 to 45) in year 2018.

1-8: Previous Studies and Researches.

Study of Montasir Ahmed Osman entitled (The Use of Geographic Information Systems for the Analysis of the Distribution of the spread of Malaria in Kassala State) 2015. This study aimed at using the geographic information systems and its applications for the analysis of the distribution of malaria across different localities of Kassala State and comparing the results with the results of the analysis using conventional statistical methods.

The GIS has been used to map Kassala with different localities, and the map had been provided with the annual incidents of malaria for each locality by using ArcGis and GeoDa. The statistical techniques provided by the two programs were used to test the random distribution of malaria across different localities of the state. Moran's I Test had been used to test the random distribution in addition to local and international (global) standards to test the existence of spatial auto-correlation for malaria incidence in the localities. The results showed that the spread of malaria in the state localities is random and does not have a special auto-correlation. These results differ from the results showed by goodness-of-fit test, where it showed that there were significant relationship between the incidence of malaria and localities.

Study of Aishah H. Azil and others entitled (Determining the spatial autocorrelation of dengue vector population:

The objective of this study to investigate the spatial autocorrelation of vector abundance from BG-Sentinel traps (BGSs) and double sticky ovitraps (SOs) collections , (b) investigate the potential of these collections for spatial interpolations (kriging).there was a significant positive autocorrelation for BGS2010 across the study site and at smaller spatial extent. Spatial autocorrelation for BGS2010 peaked at distance of 79 m, all other traps collections exhibited weak spatial autocorrelation or random patterns.

influences of mosquito sampling method, covariables and vector control, this study investigated spatial autocorrelation of female Aedes aegypti L. mosquito abundance from BG-Sentinel trap and sticky ovtirap collections in Cairns, north Queensland, Australia. BG-Sentinel trap collection in 2010 show a significant spatial auto correlation across the study site and over smaller spatial extent, while sticky ovitrap collections only indicate a non –significant, weak spatial autocorrelation. The BG-Sentinel trap collections were suitable for spatial interpolation using ordinary kriging and cokriging technique.

Study of Tahir Jumaa Taher Yousuf 2007 entitled (Spatial Analysis of Educational Services in Nablus city, by using Geographic Information Systems GIS Technology). This study aimed at studying the reality of educational services in Nablus city in terms of efficiency and distribution and its compatibility with the urban expansion and population growth in the city. The study based on the descriptive analytical approach through the application of GIS software and the statistical analysis program SPSS as well as it used some geographical models such as the link-neighbor and scope of **Study of Peter A. Ryan and others** entitled (Spatial Statistical Analysis of Adult Mosquito (Diptera: Culicidae) Counts: An Example Using Light Trap Data, in Redland Shire, Southeastern

Queensland, Australia,2004) Many mosquito control agencies use carbon dioxide-baited traps as surveillance tools for adult vector populations. However, decisions regarding the number and location of trap sites and the frequency of collections are often based on logistical issues, and not on the bionomics or spatial distribution of the target species. Therefore, with the aim of providing practical information for adult mosquito surveillance programs, weusedanarrayof81carbondioxide-andoctenol-baitedlightstraps to obtain weekly samples of adult mosquitoes in Redland Shire in southeastern Queens land, Australia. The spatial patterns of four different mosquito species were examined, and positive spatial autocorrelation in trap counts was evident for Ochlerotatus vigilax (Skuse), Coquillettidia linealis(Skuse), and Culex annulirostris Skuse, but not for the container species Ochlerotatus not oscriptus (Skuse). Of the three species that exhibited spatially correlated trap counts, the autocorrelation was greatest in Oc. vigilax at a lag distance of 0Ð1.5 km, with Moran's I values of 0.30-0.64. Moran's I indices were also positive and statistically significant ($P < 0.05$) at lag distances of 1.5Ð3.0 and 3.0Ð4.5 km on each of the 15 sampling occasions. However, at 3.0Ð4.5 km the Moran's I values were low, which indicated only weak spatial autocorrelation in trap counts. Universal kriging was used to estimate the numbers of each species at unstamped locations throughout the study area, and leave-one-out cross validation analyses indicated that this was a robust method for Cq.linealis and Oc. vigilax. In contrast, trap counts for the container-breeding species Oc. Not oscriptus were randomly distributed and the interpolated counts were not reliable. Comparisons of weekly contour maps of adult mosquito counts indicated a consistent spatial pattern for Oc. vigil ax and Cq. linealis. Particular geographic areas had consistently high or low numbers of vectors, and these patterns were stable from year to year. Definition of geographic areas with consistently high or low numbers of vectors may allow control activities to be focused in areas with the greatest risk of parvovirus transmission effect. The most important findings of this study are the randomness and weakness in the distribution of schools and kindergartens in Nablus city. That is because the city doesn't lean over pre-planning and its lack of compliance with planning standards. Moreover, the study pointed out the weakness of the efficiency and effectiveness of these services.

Study of Abdullah bin Hamid al-Qurashi entitled (the Spatial Distribution of Fire Incidents in Mecca City with an Indication of the Importance of the use of Geographic Information Systems in the Follow-up of Foundations' Safety) 2005. The study aimed at the possibility of reducing the foundations' incidents. That is after the spatial distribution of it was observed and the distribution was analyzed, with the indication of the capability of geographic information systems to follow the foundations' safety. Analytical, historical and geographical approaches were used, and the most important result came out of this study is that the change in administrative border between the civil defense departments in the Holy Capital for the year 1421 resulted in a somewhat remarkable balance in the number of foundations' incidents among themselves, and the most important causes of foundations fire incidents are electric faults.

Study of Qutaiba Saad Eddin Al-Maghribi entitled (The geographical distribution of diarrhea disease among children less than five years in Jeddah in 2002-2003 (1423AH), a study in medical geography. This study relied on the descriptive statistical method, where it showed the percentage of diarrhea infections among the children, which indicated that there is a clear geographical distribution of the incidence rates or the different patterns of diarrhea in children, and that the distribution is related to environmental and social level, which varies between districts. This reflects what has already been noted in previous studies that the districts located in south Jeddah have less social and economic levels if compared with districts lie in north Jeddah.

Study of Dr. Ali Abd Abbas Al-Azzawi entitled (Spatial Statistical Analysis in Geographic Information Systems). This study demonstrated the mathematical methods for spatial analysis and the mathematical fundamentals that the model bases on, in addition to a detailed explanation of Moran's I method, and examine the statistical significance of this test.

Study of P. Rogerson, Y. Sun entitled (Monitoring of Geographic Patterns: an Application to Crime Analysis). In this study, the application of spatial statistics was conducted such as Nearest-neighbor Statistic. It works on the comparison between the actual average of nearest distances and expected distances for the same distances, and explaining its mathematical formula, as well as the Cumulative sum methods which are statistics that had been designed to detect the differences occur in the average value, which is used widely in the fields of industry, especially in quality control.

1.9 The research gap:

We find most of the studies that dealt with the study of describing phenomena, that is, the pattern, while they were not interested in building spatial models to forecast phenomenon for each point or situations.

1.10 Research Structure.

This research contains five chapters. Chapter One includes an introduction, research problem, research importance, research objectives, research hypotheses, research domains, the data and research methodology as well as previous studies and researches.

Chapter Two contains the theoretical framework of the study, in which, we will identify Jeddah municipalities and explain the descriptive spatial statistics.

Chapter Three: explain the geographical models that will be used in this research.

Chapter Four: The methodology of statistical analysis and the methods of samples collection, and explain the tool of data collection.

Chapter Five: the statistical analysis, where the statistical analysis will be performed to prove/disprove the hypotheses of the study and to ensure the correctness and appropriateness of the estimated statistical models in prediction process.

Chapter Five: Findings summary and recommendations, where the results of the study would be derived, and the model would be used for prediction process, in addition to set the recommendations.

Chapter two

Methods

2-0 Introduction:

Over the past 10 years global positioning systems, geographical information systems (GIS), geo-statistics and remote sensing have been strongly promoted as tool for investigating the spatial component of the epidemiology of infectious diseases (O'Dwyer and Burton 1988).

The necessity for the spatial statistic of the social, environmental and other phenomena appeared because it became clear that the place or location in which the data is presented (map and descriptive data) for the phenomena existing in a designated place has an effect. May be the reading taken (non-spatial statistic) from a certain area due to its proximity where the statistic phenomenon appears in a density more than that in the said area. (Simpson, and Novak, 2013).

Geographical data are like other data. Data distribution is considered the essence of data analysis whatever their type, Geographical or non-Geographical on which the analysis processes are based such as the distribution of certain patterns or the environmental phenomenon or demographic phenomenon or human conduct in different geographic areas. (Goodchild,1992).

For this reason the need to know the patterns or forms taken by the phenomenon in a certain geographical place and study its reasons, or it is just a random distribution founded by luck and chance? Need to know the necessary conditions to specify the locations where certain specifications are available.

the studies do not include the spatial influence of the environmental phenomena and only become satisfied with their computing, analysis and circulation without giving attention to the fact that the area may be effected by the environmental factors, population density and other factors which assist generally in the spread of insects which lead to the difference in density from one place to another.

2-1 Descriptive Spatial Statistic:

Non-spatial data descriptive statistic methods are concerned with calculation of indicators that show the range of agglomeration or dispersion of this data at certain points which is known as dispersion and central tendency measures, they care only presenting and summarizing a big group of non-spatial data. A digital form only through description of the indicators or statistical dealings can describe them, which give a form about these data characteristics.

There are indicators and statistical dealings for the description and analysis of the spatial data (geographical locations) of the phenomena as to their spatial characteristics and spatial dispersion. The analysis of the spatial analysis is considered a basic complement of the analysis of the values of the phenomenon itself. Any phenomenon on the surface of the earth needs to be understood to analyze the locations, dimensions and volumes of its terms spatially. Central tendency measures provide information about the middle or intermediate center of a set of spatial distributions. The positioning of any phenomenon is determined spatially by: (Fotheringham, et al 2000).

2.1.1 Mean Center (X, Y) which is calculated in the following mathematical formula:

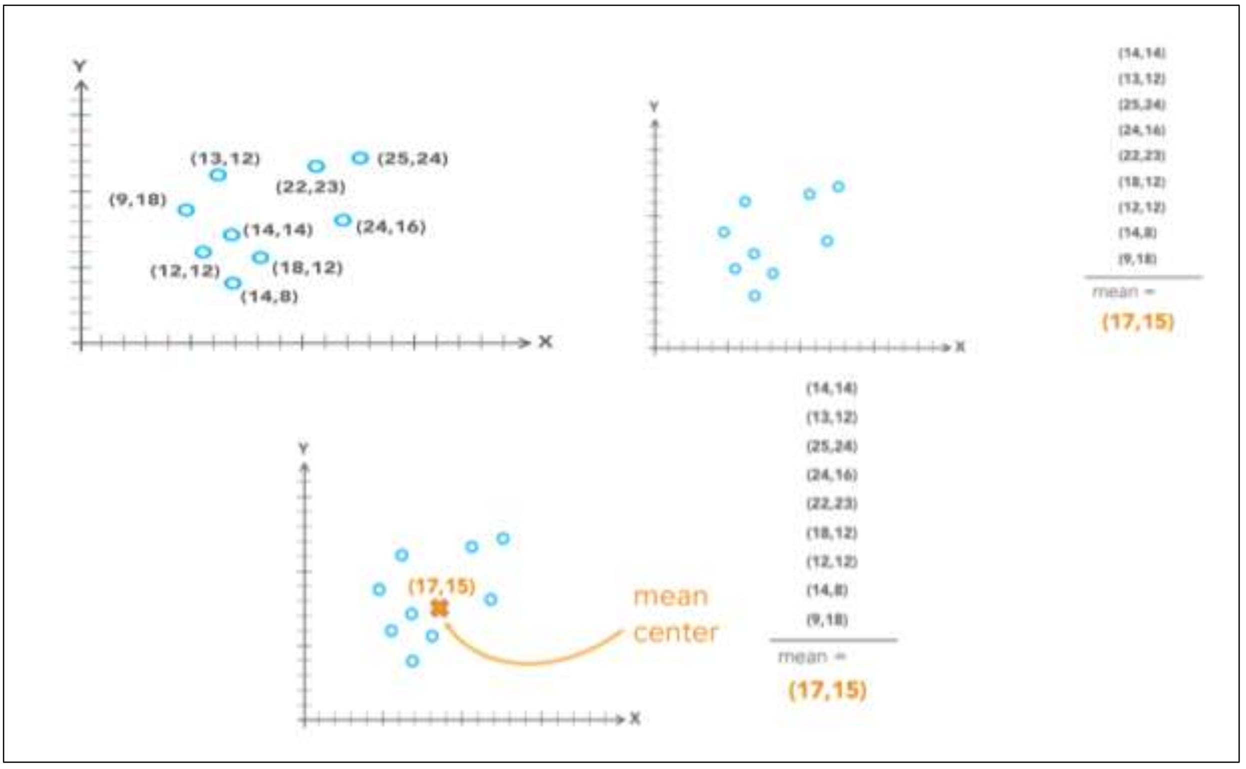
$$X = \frac{\sum_{i=1}^{n} X}{n} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (2.1)$$

$$Y = \frac{\sum_{i=1}^{n} Y}{n} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (2.2)$$

2.1.2Weighted Mean Center:

Mean center identifies the geographic center (or center of concentration) for a set features.(Lauren Bennett,2016).

Figure no (2.1) weighted mean center.



Source: esriurl.com/ spatialstats

Weighted Mean Center is calculated according to the following mathematical formula:

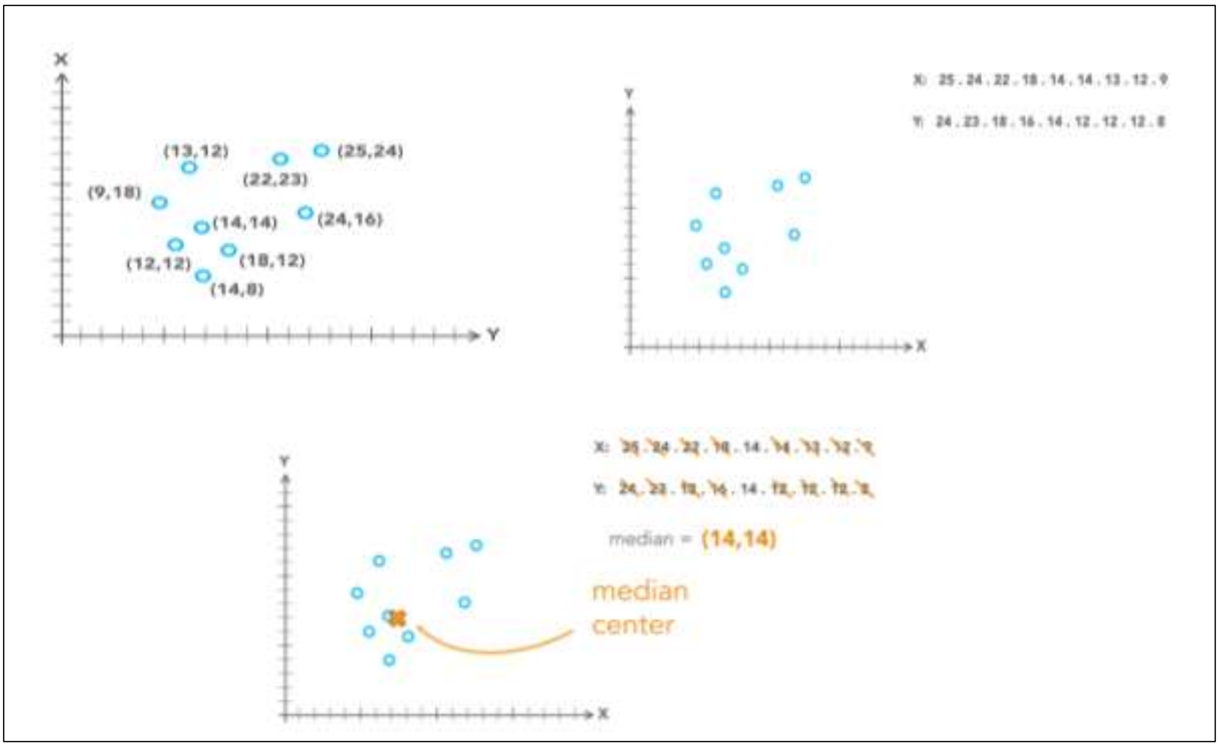
$$X = \frac{\sum_{i=1}^{n} WX}{\sum_{i=1}^{n} W} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2.3)$$

$$Y = \frac{\sum_{i=1}^{n} WY}{\sum_{i=1}^{n} W} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2.4)$$

2.1.3 Median Centre:

Median center is calculated by following the same steps of calculating the medium non-spatial data for each coordinate separately by arranging the coordinates in ascending or descending order then select the value which mediates these coordinates according to the following equations: (Webster, and Oliver, 2007.).

Figure no (2.2) median center.



Source: esriurl.com/ spatialstats

In the case of an odd number of coordinates:

$$mod = \frac{N+1}{2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (2.5)$$

There is one median value.

In the case of even coordinates:

$$\mathrm{mod}_1 = \frac{\mathrm{N}}{2}, \mathrm{mod}_2 = \frac{\mathrm{N}}{2} + 1 \ldots\ldots\ldots\ldots\ldots.. \ldots\ldots\ldots\ldots. (2.6)$$

There are two intermediate values and the median shall be as follows:

$$\frac{\mathrm{mod}_1 + \mathrm{mod}_2}{2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots... \ldots\ldots\ldots\ldots.. (2.7)$$

There are also many indicators that indicate the spatial dispersion of the phenomenon in the locations of the terms of a certain phenomenon:

2.1.4 Standard Distance:

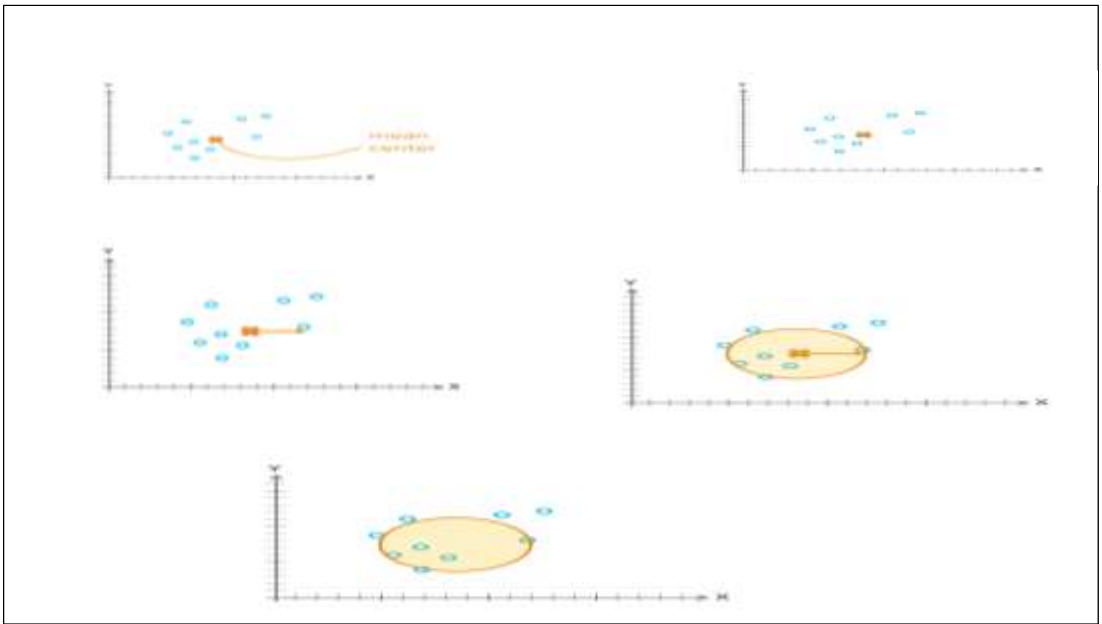
It is considered the corresponding of the standard deviation of non-spatial data and is used as an indicator of the extent of spacing or agglomeration of the values (the points) and is represented as a circle (standard circle) centered in the same position of the average center, and is calculated as follows (Haining, and Haining, 2003):

$$\mathrm{SD} = \sqrt{((\mathrm{x}-\bar{x})^2/n)} + \sqrt{((\mathrm{y}-\bar{y})^2/n)} \quad \ldots\ldots\ldots\ldots\ldots\ldots..\ldots\ldots (2.8)$$

2.1.5 Standard Distance Weighted:

$$\mathrm{SD_w} = \sqrt{\frac{\sum_{i=1}^{n} w_i (x_i-\bar{x})^2}{\sum_{i=1}^{n} w_i} + \frac{\sum_{i=1}^{n} w_i (y_i-\bar{y})^2}{\sum_{i=1}^{n} w_i}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.9)$$

Figure no (2.3) standard distance weighted.



Source: esriurl.com/ spatialstats

2.1.6 Directional Distribution

It is an oval shape (Ellipse) that expresses the properties of a directional distribution.

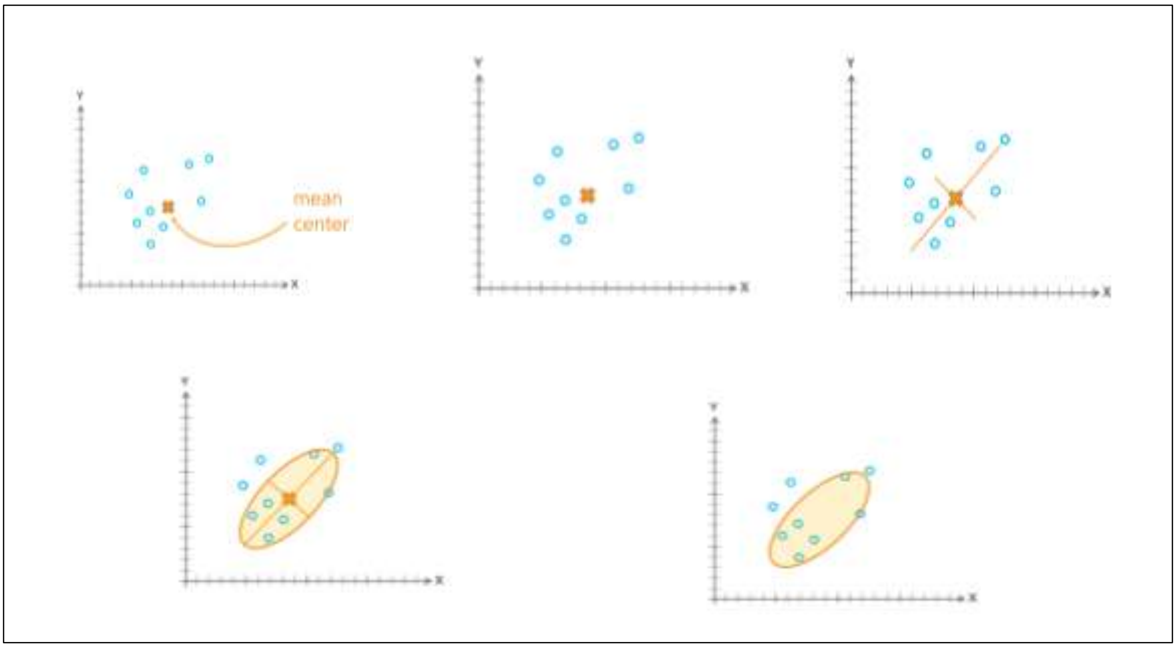
The center of this oval shape shall be applicable to the middle center (it is a figure showing the direction of the distribution of terms). Its larger center measures the value of the direction taken by most values of the phenomenon, based on the direction of the rotation. (Yager, and Hopkins,1993).

$$\sigma_{1,2} = \frac{\left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 \pm\right)\sqrt{\left(\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i^2\right)\sum_{i=1}^{n} x_i^2 + 4\left(\sum_{i}^{n} x_i y_i\right)^2}}{2n} \quad \ldots\ldots\ldots\ldots(2.10).$$

Also, we create standard deviational ellipses to summarize the spatial characteristics of geographic features: central tendency, dispersion, and directional trends.

To test the randomness of the distribution between the studied phenomenon points and to compare the distribution patterns for more than one area, the following criteria were used:

Figure (2.4) Directional Distribution.



Source: esriurl.com/ spatialstats

2.1.7 .Nearest Neighbor or Neighborhood relation

It is considered a common method spreading and used by geographers. its idea is based on giving a rough judgment of the spacing between points, aiming to reach a quantitative criterion to infer the spatial pattern distribution of the centers represented by these points and the nearest neighbor equation can be formulated as follows (Prasomphan, and Mase (2013):

$$R = 2x\sqrt{\frac{m}{n}} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (2.11)$$

Where:

R = Nearest neighbor.

X = Distance rate (total distance / number of distances).

M = number of points.

n = total distance.

The nearest neighbor coefficient (R) expresses the distribution pattern and its value ranges between ($0 \leq R \leq 2.15$). Accordingly three patterns of the main spatial distribution are specified with other patterns near to it, the patterns are:

Convergent distribution pattern:

The value of the coefficient $0 < R < 1$ is such that:

R = 0, the phenomenon point in this case is grouped in one location.

0>R>.5 The pattern in this case shall be convergent, and the closer to zero it shall be more convergent and the distribution pattern of data shall become in a form of cluster (clustered).

If the value of R 1 is 0.5, the pattern is convergent and moves towards the random pattern. (Aiba, et al (2013)

2.1.8 Randomization Distribution Pattern:

The value of the coefficient R = 1, the random pattern is a purely theoretical pattern and there is no distribution of geographic phenomena on the surface of the earth takes this pattern (Dixon, P. M. (2006).

Spacing Distribution Pattern:

The values of the nearest neighbor coefficient range from ($1 < R < 2.15$), so that if the value of R>1 is 2, the pattern shall be spaced unorganized, and if (R = 2) the divergent pattern shall be in this case resulting from a regular distribution of points. Their shape shall be square shape so that the closer the R value to 2.15, the greater the spacing between the distributing points. When the points are at their maximum spacing, their distribution shape in this case shall takes the shape of a hexago.(Clark and Evans,1954).

2.1.9 Spatial Auto-Correlation Coefficient (Moran index):

Statistician's experts use the spatial auto-correlation coefficient in the measurement of the similarity of contiguous phenomena, which depend on comparing the value related to each parameter with the average value of the structure, which is called the statistical value (Moran Index). (Schabenberger, andGotway,2017).

In this method, if the difference between the adjacent values is smaller than the difference between all the parameters, the similar values are aggregated. Geographical phenomena are normally associated with the values of spatially adjacent variables. When the values of a variable in a location is associated with the values of the same variable in a nearby location, this shows a self-correlation between the two variables in sometimes referred to as neighborhood effect or Contiguity. Here we ask, are similar phenomena in the place gather? Generally, the spatial self-correlation coefficient measures simultaneously the similarity between the spatial elements locations and their distinguished characteristics and it can be calculated according to the following mathematical model:

$$I=\frac{N}{\sum_1^i \sum_1^i \mathrm{Wij}} \frac{\sum_1^i \sum_1^i \mathrm{Wij}(x_{i-\bar{x}})((x_{j-\bar{x}})}{\sum_1^i (x_{i-\bar{x}})^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ldots.(2.12)$$

Where:

W_{ij}: spatial weight between i and j.

N: total number of terms.

Moran index is considered one of the important measures in detecting the self-correlation between the elements of the studied phenomenon and assess its spatial distribution pattern, is it dispersed or regular or random pattern?.

The value of the index ranges between (1-) to (+1) . If the index value is near to (+1). This indicates that the pattern is random, while the pattern is described as regular if its value reached zero or close to that.

The distribution pattern varies between grouping, regularity, and randomness depending on the index value.

The general framework for testing the hypotheses is considered a good outcome tool to judge the nature and pattern of the spatial distribution for the geographical phenomenon. Especially, the results of the measures used within the program (ArcGIS 10.3) depends entirely on the principles of testing hypotheses.(Hengl, (2009).

The issue necessitate first determining the initial hypothesis (nullity hypothesis) or the zero hypothesis, which stipulates that there is no existence of particular pattern of distribution, and the expected pattern is a random pattern produced by the act of chance or luck. In order to decide whether to accept or reject the above hypothesis in case of the using the coefficient, the zero theory recognizes that there is no aggregation or spatial agglomeration for the values of the geographical phenomena.

Test of distribution pattern significance:

Assurance shall be made that the differences are significant and the distribution pattern is having statistical index. Turning to Z test or the standard degree.

$$Z = \frac{A - B}{\frac{S}{\sqrt{N}}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . (2.13)$$

A: The real rate of distance.

B: theoretical rate of distance.

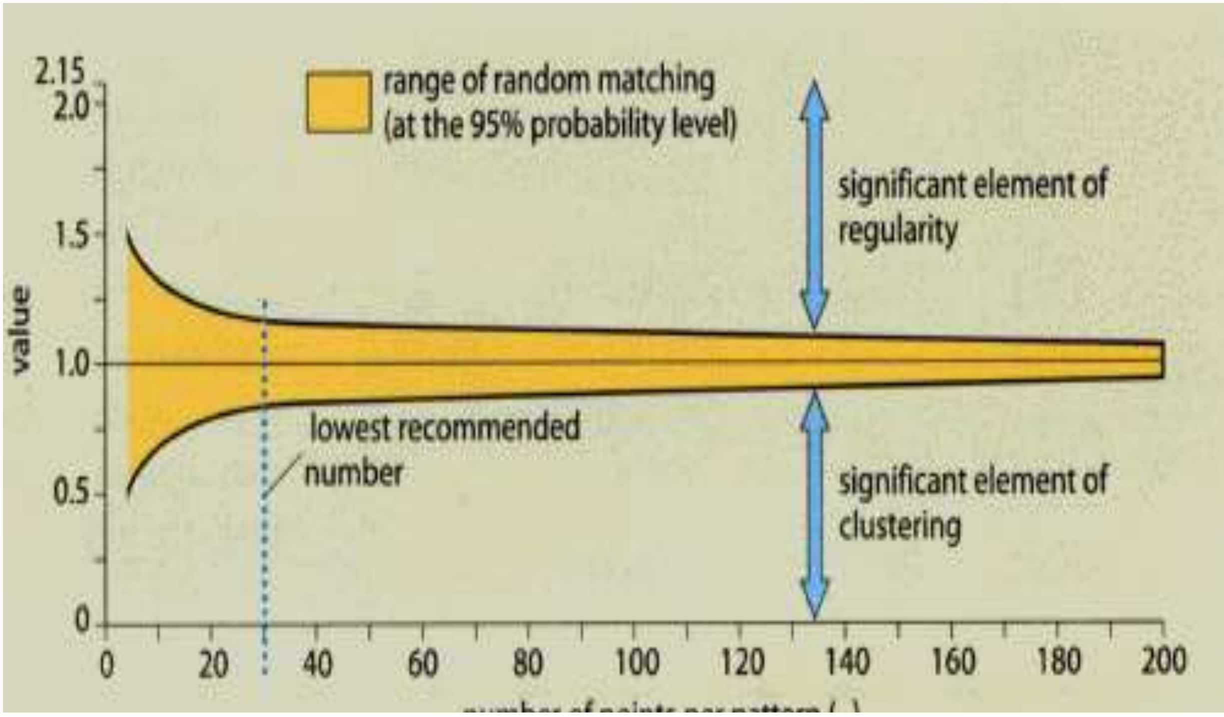
S: standard deviation.

N: Number of points.

The value of (Z) is found from the statistical tables where it is extracted and below the level (0.9, .95, .99) and the degree of freedom (n-1).

If the calculated Z value is less than the tabulated, the zero hypothesis (H0)shall be reject, i.e. the significant differences are having statistical index and the distribution of distances is not random, but if the opposite , i.e. the calculated differences are more than the tabulated ones and the distribution of distances shall be random. But when the value of (P) is small and the absolute value of (Z) is so great that it falls outside the required confidence level, if the value is less than (0), then the group of phenomena appears in diverging form. (Campbell, and Clarke, (1971).

Figure (2.5) Possible R values based on the number of points.



The figure below shows distances and values correlated:

Figure (2.6) distances and values correlated.

Source: esriurl.com/ spatialstats

2.2Spatial models.

Many of us involved in spatial analysis research are excited to see the explosive growth of interest in GIS. It is clear that the rapid growth of GIS has given a big boost to fundamental research in spatial analysis, and in many ways solidifies the future of the quantitative focus of the discipline.

There exist potential linkages between many aspects of spatial analysis and the new information processing, data handling, data storage and data display techniques available through GIS. Furthermore, there are strong emerging links to applications in other disciplines with parallel interests in spatial analysis (including ecology, archaeology, natural resources, landscape architecture and geodetic science, etc.) These allied fields have long used spatial analysis tech rüques (see, for example, Upham, 1979; Bartlett, 1975; and Diggle, 1983) but are now increasingly using 018 as a creative tool.

Against the background of this growth, and from the perspective of a spatial analyst looking out towards the next stages of development in GIS, two major directions which need attention are apparent: First, the traditional methods for displaying data about spatial situations, and for presenting the results of spatial analyses, need to be overhauled in view of the great advances in information processing technology. This line of attack is straightforward:

it suggests that new display techniques be added to the output of existing spatial analysis operations. Ideally, the analyst would use the GIS in query mode (Goodchild, 1987) to develop an improved understanding of the properties of a spatial system. While this is easy to state in principle, the actual implementation will stretch current capabilities and will require a rethinking of the conceptual bases for spatial analysis. One such example might arise if improved display techniques increased the accessibility of multiobjective programming, thereby encouraging analysts to take a multiobjective view, and therefore replacing existing uni-dimensional methods.

One obvious area for this to play a role is in locational analysis.

Second, spatial analysts need to help GIS end-users and providers to understand and to improve their sets of tools, and to enhance the appropriate levels of theory and modelling capability in real problem-solving situations.

This need is particularly acute if appropriate analytical methods are to lie at the foundation of developments in GIS. The capability to adapt existing al gorithms to the data structures in GIS is a critical component of this research.

The overall theme of this chapter is that when GIS integrates spatial analysis at a fundamental level, the full potential of spatial analysis will be unlocked. The case is made by demonstrating that there are potential new results and advances that are obtainable by linking spatial analysis and GIS.

2.3: Innovations in spatial Statistics and GIS applications:

The next few years promise an unprecedented opportunity for spatial analysts to promote and contribute to the spread of new tools to the GIS community. This section is designed to make some of the power of these types of tools apparent to the potential user. To attempt to make the discussion more concrete, we display some fundamental building blocks or entities in spatial analysis: i.e. Points, Lines and Areas. Associated with each of these classes of entities are new spatial analytical operations that need to be developed. Couclelis (1991) correctly points out that the focus on space as a container does not provide the platform for answering the complex types of spatial questions posed in planning; however, see more on the site and situation distinction she draws in later paragraphs. In each case the role of new exploratory tools, visualization, and space-time analyses can be seen in slightly different ways.

2.4 Space-time pattern recognition in point data sets:

Geographers have been slow to integrate both temporal and spatial dimension into GIS. There are formidable technical obstacles, but even at the fundamental research level, analysts have been unable to operationalize the role of space and time in simple models. A concrete example may help. A conceptual model of space and time acting as a constraint on human activity (Hagerstrand) has gone largely untested because of the difficulty of translating activity records from travel diaries into three dimensional diagrams called ‘prisms’ by Hagerstrand (however, see Miller, 1991). There have been studies of the importance of distance (Gatrell, 1983) and time (Parkes and Thrift, 1980; Goodchild and Janelle, 1984; and Janelle, Goodchild and lUinkenberg. 1988), but despite these impressive research efforts, it is uncommon to consider 67-combined time-space ‘reach’ as a quantitative tool in measuring transportation plans. Two exceptions include Villoria's (1989) dissertation, which measured the size of activity fields in an urban case study in the Philippines, and Kent's (1984) mapping of activity spaces from archaeological data.

2.5 A model for space-time data analysis:

Assume that the events are located at distinguishable locations, and that a simple date is used to keep track of the time of events. Let $W_{ij} = f(t_i, t_j)$ where t_i and t_j are the times of the events i and j, and further that i and j take place at $p_i = (x_i, y_j)$ and $p_j = (x_i, y_j)$ respectively. Supposing that W is a decreasing function of the time between the events:

$$W_{ij}=1/(e+|t_i-t_j|) \qquad \text{.....................................(2.14)}$$

Then, given thresholds 1) and T, classification of events in space and time can be simply represented in a two-by-two table (Table 4.1).

Table (4.1) Two by two, table.

SPACE		
TIME	d (pi , pj) < D	d (pi , pj) > D
Wij >T	CLOSE TOGETHER	FAR PARAT
Highly interactive	CONTEMPORANEOUS	CONTEMPORANEOUS
Wij <T	CLOSE TOGETHER	FAR APART
Weakly interactive	Time GAP	Time GAP

Source: from M. R. Kiouber and P. Musracchi (1970).

The statistical analysis of such 2-way tables is a fundamental one for recognizing whether or not there are significant patterns in space and time (see Knox, 1964). The strength of the temporal interaction is measured as follows: W increases as the event are closer together in time) and D increases as the events are further apart in space. To what extent are observations which are close together in time also spatially clustered? The problem is to recognize clusters or groups in the set. While there are many follow-up papers to Knox's pioneering analysis, the method is problematic because of the nonindependence of events (see Glick, 1979).

The next section develops a novel mathematical approach to the problem of space-time pattern analysis: one which is of great potential usefulness in identifying the spatial pattern and perhaps helping to uncover the underlying spatial process. Consider a set of n interacting points in a two-dimensional space. The levels of interactions between the observations are given exogenously, as functions of their temporal separation. Assume that the cluster means must be adjusted to reflect the interaction between the entities. For example, consider a system of n = n1 + n2 nodes such that the n1 subset is temporally linked, and the n subset is also highly interactive among themselves, and for the sake of illustration suppose that there are negligible interactions between the subsets. (That is the (n1 x n2) and (n2 x n1) subsystems contain only zero interactions.) If the n1 and n2 nodes are plotted graphically, it would be for tuitous if all the n1 and n2 nodes could be separated neatly into two easily identifiable spatial groups. Indeed, since there is no requirement of contiguity for the interacting entities, there is no guarantee that a cluster of points n1 should contain only adjacent nodes. While the conventional geostatistical clustering problem for several groups yields a partition with the property that all the observations which are closer to centroid A than to centroid B are assigned to the same group, this is not a property of the interacting cluster problem (see O'Kelly, 1992). O'Kelly (1992) has recently proposed a solution to this partitioning problem and this section briefly summarizes the method. It is required to cluster the n observations into p groups, so that the sum of squared deviations from the cluster means is as small as possible. Assume that the cluster means are adjusted to reflect the interaction between the entities. Further, since it is desirable to place highly interactive observations in the same group, it will be assumed that the penalty for assigning observation i to group g and observation *j* to group h is an increasing function of the distance between the group centroids and the interaction level W. Specifically, the 'cost' of assigning i to g and j to h is:

$P\{i(g),j(h)\}= W_{ij}(d_{ig}+d'_{gh}+d_{hj})$……………………………………….(2.15)

Where:

W_{ij} is the exogenous interaction effect,

d_{ig} is the squared distance from i to cluster center g.

d'_{gh} is the squared distance between the cluster centers.

d_{hj} is the squared distance from j to cluster center h.

Let (X_g,Y_g) be the centroid of group g for =1,…..,p. The distance are defined as: $d_{ig}=(X_i-X_g)^2+(Y_i-Y_g)^2$ for all i=1,…..,n and g=1,….p; and $d'_{gh}==(X_g-X_h)^2+(Y_g-Y_h)^2$ for all g and h=1,…,p.

MIN $T=\sum_i \sum_j W_{ij} \sum_g \sum_h K_{ijgh} D_{ijgh}$ … … … … … … … … … … … … (2.16)

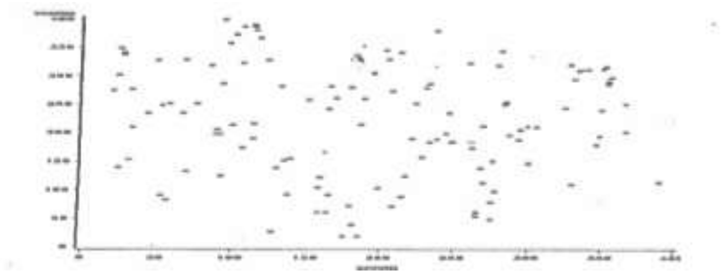
Where$K_{ijgh} = 1$ if i belongs to group g and j belongs to group h, and $K_{ijgh} = 0$ otherwise; and where $D_{ijgh}= d_{ig} + d'_{gh} + d_{hj}$. Note that the K, integer variables obey the following restrictions:

$K_{ijgh} = Xig\, Xij$……………………………………………………………………..(2.17)

Where the X values are allocation variables, that:

$\sum_i \text{Xij} = 1,$ *for all iand* Xij is either 0 or 1 … … … … . … … … … ….(2.18)

Figure (4.1) Point data set n = 125



Source: from M. R. Kiouber and P. Musracchi (1970).

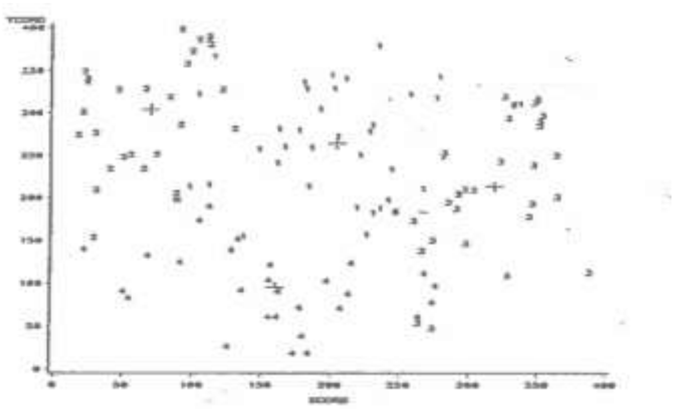
A useful property of the problem is that the use of a squared distance term yields a linear system of equations for the coordinates of the cluster centroids. These equations are derived and solved repeatedly, for a given set of cluster allocations. A sequential reallocation of the observations between the clusters is then performed: that is K_{ijgh} is initially assumed to be fixed and this is equivalent to starting the problem with a known partition of the observations between p groups. The model solved in O'Kelly (1992) discusses the iterative reassignment of observations to clusters.

As an example of the procedures explained in the previous paragraphs, consider the 125 points shown in Figure 4.1. The -observations are used because they present a convenient source for a set of x, y locations and a time stamp for each event. No substantive contribution to the original data context is attempted here, rather suppose that these are the locations of fires in a city, and we are interested to see if there are clusters of events in space, in time, or in space and time. For the sake of illustration, a simple set of interactions between entities is modeled as W_{ij} = 10/(EPS + t — t I) where EPS is a small constant (set to 0.1) to prevent division by zero if the two events occur at exactly the same time. The result of clustering 125 observations into 4 groups in shown in Figure 4.2 which shows conventional group centroids, using a 'X' symbol, and the cluster membership of the data points. No attempt has been made here to find the optimal number of groups, and it is recognized that many different partitions of the data could be produced by altering the initial partition, or by changing convergence criteria.

While conventional clustering has been in use for many years as a means of spatial pattern analysis (see Baxter, 1971, for early examples) the extension of clustering techniques to allow for interdependence between the clusters adds significant benefits in GIS applications such as pattern recognition - e.g. hypothesis testing for patterns of arson and crime.

There exists some strong potential applications of these tools, e.g. K-means program / ISODATA package; similar problems arise in the case of organizing large operations like the postal service, mail order deliveries and retail market delimitation. Both the OR questions (concerning the optimal flrtitioning) and the GIS question (concerning the display/visualization of large database) require further work.

Figure (4.2) Preliminary example: clustering into 4 groups.



Source: from M. R. Kiouber and P. Musracchi (1970).

2.6 Lines and flows: spatial interaction:

Consider the location of facilities, using an interaction model to drive the choice of sites. The probabilistic allocation of demand to these facilities is a classic problem in spatial interaction theory. Techniques to locate these facilities in a manner, which optimizes the spatial interaction properties, are well known (see Hodgson, 1978; and O'Kelly, 1987.) However, a new feature which is essential to these models, was omitted until the graphic capabilities of GIS revealed its importance. Spatial interaction concepts are linked to covering and optimal location models in the following way: it is apparent from the visualization tools that the role of spatial constraint in interaction models has not been fully integrated into locational analysis and indeed the interplay with the space-time prism model is a very interesting possibility. If a facility has a known range, and therefore can only cover a demand point in the fixed radius, then this fact ought to be built in to the spatial location model. The hybrid concept of probabilistic interaction towards facilities that have a covering radius introduces some subtle new spatial situations. Observe that there is no guarantee that a specific demand point has any facility within in its range. The demand from that origin is therefore uncovered, but in spatial interaction terms, some demand must be sent to an artificial destination to represent the uncovered demand.

To pull out the general point: spatial analysts are concerned to link methods from OR etc., and to develop novel visualization tools to help with analysis of the model. The algorithm does the work, and then the solutions are posted back to the GIS to validate and query the quality of the solution. Questions are clear are the correct assumptions built into the optimization tool? Are the key parameters and mechanisms represented in the model? What data/empirical estimates need to be adjusted as a result of the analysis? The entire model-building exercise is viewed as an iterative that cycles back to the beginning to check the validity of the model a step which is too often replaced by hand waving. The problem is a generic one whenever a complex model produces output in the domain of X, the sensitivities of outputs to changes in the key parameters, or to changes in the assumptions about the role of critical variables, must be investigated. The answers in this kind of research hinge on both a fundamental substantive understanding of the problem, and a mathematical insight to the model.

2.7 Spatial autocorrelation:

As another example of the marriage of spatial models and GIS, consider the example of spatial autocorrelation. The correct measurement of spatial autocorrelation is a necessity, but one which is open to a wide variety of subtle variation. Clearly, GIS allows these measures to be gathered rather easily, since adjacency is one of the properties of spatial entities that is maintained in GIS databases. This is possible because of the increased availability of adjacency '[acts' from topologically integrated databases. Even with spatial autocorrelation at hand, the analyst must decide how to make appropriate statistical use of the indices, and be cognizant of the role of units and scale of analysis on the results (Chou, 1991; Newsome, 1992). Also, the theory underlying the appropriate statistical operators for various measures of spatial autocorrelation needs to be fully understood by the end user (see, for example, Haining, 1978). There is little to be gained by making spatial autocorrelation one of the many descriptive statistics collected from a spatial database, unless the sophistication of the user is sufficient to make correct use of this information (see Getis, 1991; Chou, 1991).

2.8 Spatial situation vs. spatial site

A final example grows from the theme developed by Couclelis (1991) about the importance of site and situation and is discussed in terms of Fotheringham's competing destination (CD) models, and the possibility of enhanced measurement of CD effects in GIS.

An example of the kind of situation effect, which is widely known, is the Fotheringham (1983) measure of competition between destinations. The measure is important to spatial interaction models and can be expected to be an easily calculated variable in GIS implementations of spatial interaction models. Moreover, the GIS locus allows us to see that the conventional CD measure can be generalized once the ideas of neighborhood variables and adjacency are fully integrated.

The starting point is $A_{ij} = \sum_{k.j,k,I} W_{ik} f(d_{jk})$, which is the Fotheringham measure of CD for facility j, from i's perspective. There are three reasons that the accessibility of j is measured with an 'i' subscript (1) the calculation avoids the inclusion of i; (2) the attraction variable W is possibly calculated differently from each zonal perspective; and (3) the distance decay function is specific to zone i. Notice that this calculation includes all potential competitors of j, including some that might reasonably be expected to avoid interaction with 1. The following set of generalizations emerge, and these are measures which would be easy to calculate in a database that keeps track of the 'neighborhood' variables. Some care would have to be taken to ensure that the data structure allowed the efficient collection of the facts needed for the measures below. Two examples of each of three types of generalization are given here:

1. Restrictions on the list of competitors of j, from i's perspective:
   a) $A_{ij}=\sum_{k.j,}\ k \epsilon S(i)\ W_{ik}\ f_i(d_{jk})$, Where $S(i)=\{K/d_{ik} \geq L\}$

This measures the accessibility of j to those facilities ‘k’ which are further than L from i (e.g. consistent use of the masking radius used in Fotheringham, would set L = 160).

b) $A_{ij}=\sum_{k.j,}\, k\epsilon T(i)\ W_{ik}\ f_i(d_{jk})$, Where $T(i)=\{K/d_{ik}\leq d_{ij}\}$

This measures the accessibility of j to those facilities ‘K’ which are closer than d_{ij} to i (e.g. it could also be defined with inequality reversed).

2. Restrictions on the list of competitors of j, from j’s perspective:

(a)$A_{ij}=\sum_{k.j,}\, k\epsilon S(j)\ W_{ik}\ f_i(d_{jk})$, Where $S(j)=\{K/d_{jk}\leq Z\}$

This measures the accessibility of j to those facilities k’ which are closet than Z to j (e.g. this is called a traffic shadow effect, in Taaffe, 1956).

(b) $A_{ij}=\sum_{k.j,}\, k\epsilon T(j)\ W_{ik}\ f_i(d_{jk})$, Where $T(j)=\{K/D_{jk}=1\}$

This measures the accessibility of j to those facilities ‘k’ which are in the set of pairs for which 1)Jk = 1. These could be pairs of twin cities, or other close substitutes.

3. Restrictions on the tin of competitors of j, from (i,j)’s perspective

(a) $A_{ij}=\sum_{k.j,}\, k\epsilon S(i,j)\ W_{ik}\ f_i(d_{jk})$, Where $S(i,j)=\{K/d_{ij}+d_{kj}\leq R\}$

This measures the accessibility of j to those facilities ‘k’ which are in the set of pairs (j, k) for which the sum of distances from i to j and from k to j is less than some budget. This could he used to define an interaction space such as a corridor.

(b) $A_{ij}=\sum_{k.j,}\, k\epsilon T(i,j)\ W_{ik}\ f_i(d_{jk})$, Where $T(i,j)=\{K/d_{ik}+d_{kj}\leq R_1,\ d_{ik}+d_{kf}\leq R_2+d_{kf}\leq R_3\}$

This measures the accessibility of 3 to those facilities ‘k’ which are in the action space bounded by distances of R1, R2, and R3 from the fixed points i, j, k and f. These could be pairs of interactive points, or corridors pointing towards some particularly attractive alternative such as the city center 'f'.

Important research questions could then be answered in spatial analysis: such as further empirical evidence of the role of the competition effect, especially in the light of the generalized measurement of spatial competition. With these tools, realism is added to the theoretically justified measures of competition between alternatives proposed by Fotheringham (1983).

2.9 Progress via incremental improvements:

There are many spatial analytical tools which do not currently exist in GIS (and some which may be inherently impossible to implement because of data structures — see Couclelis, 1991; Goodchild, 1987). In addition GIS prompts us to see new ideas in spatial analysis, which otherwise might not be clear. But there are barriers to the realization of this potential. To get from the current level to sonic of the tools discussed here, major fundamental research will have to be done on some basic steps. These intermediate steps include:

1. Exploratory data analysis, to search for previously unseen processes within the data and test new analytical techniques for finding new instances of a sample pattern.
2. Systematic integration of existing models of spatial processes such as interaction and gravitational concepts, spatial autocorrelation measures etc. This includes investigation of the problem of ‘modifiable areal units’.

3. Spatial process and spatial patterns through, time also need to be investigated.

2.10 Explanatory data analysis:

Techniques for sifting through large data streams and innovative visualization techniques are needed to allow the analyst to assimilate the quantity of data presented (e.g. imaging systems; panel surveys; census products). The major problem becomes that of deciding what is important in the vast quantities of data which are generated in large models and GIS. The key technical advance will be in pattern recognition, which intelligently allows the user to sift through the data, reduce dimensionality, find patterns of interest, and then order the GIS to find other instances or similar occurrences. This sounds simple, but is difficult to implement when the size of the underlying database is of the order of multiple gigabytes.

Also in the arena of exploratory data analysis and categorical data analysis is the important point that large-scale data surveys in the social sciences are producing very large amounts of data. The quantity of these data may dis courage analysts from embarking on useful research. In some cases the quality of the data are also in question, especially in the marketing and retail analysis arena where the results of cluster analysis are used to estimate the micro demographics of spatial zones. Other high-quality examples which are currently available include the Annual Housing Survey, data collected from continuous work history, the Italian census as discussed by Open Shaw, the Cardiff panel data survey, etc. In the presence of such large data sets, the need for novel visualization techniques in the social sciences is just as pressing as in the physical realm.

2.11 Geometric pattern analysis:

At the earliest stage of quantitative analysis, simple small-scale analyses were performed. Typically, these data analyses considered small sample case studies, and involved a proof of concept kind of approach. The full 'industrial strength' scale up of these methods was constrained by the power of data processing technology, and by the lack of large digital database. Now, with the advent of digital databases, conventional spatial statistical studies using nearest neighbour analysis, quadrat analysis, or other spacing statistics (e.g. Gets and Boots, 1978; and Rogers, 1974) can be re-invigorated when matched to very large data sets, and to efficient computational geometry routines. A broader issue remains: are the types of questions and analyses carried out earlier still interesting? Are these techniques tractable for data sets of arbitrary size? Are heuristics needed? Do these breakthroughs allow us.to give trustworthy answers (Openshaw, 1987; Openshaw el aL, 1987) to reasonable questions about spatial patterns? Researchers need to re-examine the tool kit of spacing, quadrat and nearest neighbor techniques, and assess the impact of larger data volumes and improved computational technique on their applicability.

While short-run questions challenge us, spatial scientists must also aim for the integration of spatial models into GIS and anticipate the successful marriage of theory and practice, yielding solutions to problems that were previously thought to be unmanageable.

2.12 Methods of analysis points:

Suppose we have records of each of the customers of an insurance agent, including each customer's location. Perhaps these originated in an address list, and were subsequently converted to coordinates using the process known as geocoding or address-matching Plotted on a map they might look something. Several questions might occur to a market researcher hired by the insurance agent to study the state of the business and recommend strategies. For each there is a straightforward type of GIS analysis that can be used to provide the necessary information.

2.13 Distance

First, we might ask how far each customer is located from the agent's location. The agent might be interested in knowing how many customers do their business with the agent rather than with some other agent located closer to the customer. What is the average distance between each of the customers and the agent, and are there areas near the agent where advertising might increase the agent's market share? All of these questions require the ability to measure the distance between points, a commonly used function of a GIS.

To measure the distance between two points we need to know their locations in (x,y) coordinates, and many of them can be used to measure distances. The UTM coordinate system, for example, is based on measurements of coordinates in meters, so distances calculated in this coordinate system are easily understood. State Plane Coordinates are often expressed in feet. In these cases the distance between two points (x_A,y_A) and (x_C,y_C), representing the locations of the agent and a customer, are given by:

$$D = \sqrt{(x_A - x_c)^2 + (y_A - y_c)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.19)$$

This equation is derived from Pythagoras famous theorem, so it is also known as the Pythagorean distance, or the length of a straight line between the two points.

Pythagorean distance applies to points located on a plane, not on the curved surface of the Earth, and so should be used only when the points are close together. A good rule of thumb is that the study area should be no more than 500km across, because at that distance errors due to ignoring the Earth's curvature begin to approach 0.1%, even with the best-designed coordinate systems.

It is important to recognize that this problem of the Earth's curvature is not resolved by using latitude and longitude coordinates. Suppose we record location using latitude and longitude, and plug these into the Pythagorean equation as if they were y and x respectively (this is often called using un projected coordinates, and is also what happens if we specify the so-called Plate Carree or Cylindrical Equidistant projection). In this case, the coordinates are measured in degrees, so the value of D will also be in degrees. But one degree of latitude is not the same distance on the Earth's surface as one degree of longitude, except exactly on the Equator. At 34 degrees North or South, for example, the lines of longitude are only 85% as far apart as the lines of latitude, and that percentage drops all the way to zero at the poles. So the equation cannot be applied to latitude and longitude. Instead, we should use the equation for distance over the curved surface of the Earth, otherwise known as the Great Circle distance, because the shortest path between two places follows a Great Circle (a Great Circle is defined as the arc formed when the Earth is sliced through the two points and through the center of the Earth). If the point locations are denoted by (φ_A,λ_A) and (φ_C,λ_C), where φ

denotes latitude and λ denotes longitude, then the distance between them is given by:

$$D = R \cos^{-1}[\sin\varphi_A \sin\varphi_c - \cos\varphi_A \cos\varphi_c(\lambda_A - \lambda_c)] \ldots\ldots\ldots\ldots\ldots (2.20)$$

Where R denotes the radius of the Earth, or approximately 6378km

There are many other bases for measuring distance, because it is often necessary to allow for travel that must follow streets, or avoid barriers of one kind or another. So the actual distance traveled between two places may be much more than either of these formulae would predict.

2.14 Buffer:

Instead of asking how far one point is from another, we might turn the question around and identify all of the points within a certain distance of a reference point. For example, it might be interesting to outline on a map the area that is within 1km of the agent's location, and subsequent sections will cover several applications of this concept. In GIS the term buffer is used, and we say that the circle created by this operation constitutes the 1km buffer around the agent. Buffers can be created for any kind of object -- points, lines, or areas -- and are very widely used in GIS analysis.

By finding a buffer, and by combining it with other information using the methods discussed in this chapter, we could answer such questions as: how many customers live within 10km of the agent's location; what is the total population within 10km of the agent's location (based on accurate counts of residents obtained from the census); where are the people with the highest incomes within 10km of the agent's location?.

2.15 Points in polygons

The so-called point in polygon operation is another key feature of GIS analysis. Suppose we have a map of the tracts used by the census to publish population statistics. The point in polygon operation allows us to combine our point map with this map of areas, in order to identify the area that contains each point. For example, we could use it to identify the census tract containing each of the customer locations in our point data set. By counting customers in each tract, and comparing the totals to the known populations of each tract, we could get interesting data on market penetration by tract.

The point in polygon operation is actually very simple to execute, and so the method will be described briefly. One of the points in the figure will be used as an example. A line is drawn from the point, in this case diagonally upwards. The number of intersections between this line and the boundaries of each polygon is counted. In the example, there is one intersection with the boundary of Polygon A, 2 with the boundary of Polygon B, and two with the boundary of Polygon C. The polygon that contains the point is the only one with an odd number of boundary intersections (and it will always be true that exactly one polygon has an odd number -- all other polygons will have an even number). One useful way to think about the result of the point in polygon operation is in terms of tables.

2.16 Measurement of area

One of the strongest arguments for the use of GIS, as compared with manual methods of analysis of information shown on maps, is that computers make it easier to take measurements from maps. The measurement of area is in fact the strongest argument of this type. Suppose we need to measure an area from a map, such as the area of a particular class of land use. Perhaps our task is to measure the amount of land being used by industry in a specific city, and we are given a map showing the locations of industrial land use, in the form of appropriately colored areas. In a medium-sized city the number of such areas might be in the thousands, and so there would be thousands of measurements to be made and totaled. But even the measurement of a single area is problematic. Manually, we would have to use one of two traditional methods, known as dot counting and planimetry. Dot counting proceeds by overlaying a prepared transparent sheet covered by dots at a known density, and counting the number of such dots falling within the area of interest. To get a reliable estimate the density of dots needs to be high, and the dots need to be small, and it is easy to see how tedious and inaccurate this task can be. Planimetry proceeds by using a crude mechanical device to trace the outline of the area -- the result is read off a dial on the instrument, and again the process is tedious and error-prone. In short, manual methods are frustrating, and expensive in terms of the time required to obtain even a poor level of accuracy.

By contrast, measurement of area is extremely simple once the areas are represented in digital form as polygons. A simple calculation is made for each straight edge of the polygon, using the coordinates of its endpoints, and the calculations are summed around the polygon. Any vector GIS is able to do this quickly and accurately, and the accuracy of the result is limited only by the accuracy of the original digitizing.

The world’s first GIS, the Canada Geographic Information System, was developed in the mid-1960s precisely for this reason. A large number of detailed maps had been made of land use and other land properties, with the objective of providing accurate measurements of Canada’s land resource. But it would have taken decades, and a large workforce, to produce the promised measurements by hand, and the results would have been of disappointing accuracy. Instead, a GIS was developed, and all of the maps were digitized. In addition to straightforward measurement of area, the system was also used to produce measurements of combined areas from different maps, using the operation of polygon overlay, another important vector operation that is covered in the next section.

2.17 Polygon overlay:

Figure (4.4) shows a typical application of polygon overlay. The two input maps might represent a map of land use, and a map of county boundaries. A reasonable question to ask of a GIS in a case like this is: "how much land in County A is in agricultural use?" To answer the question, we must somehow compare the two maps, identifying the area that is both in County A and in agricultural land use, and measure this area. Polygon overlay allows this to be done. As a method, its history dates back to the 1960s, though efficient software to perform the operation on vector data sets was not developed until the late 1970s.

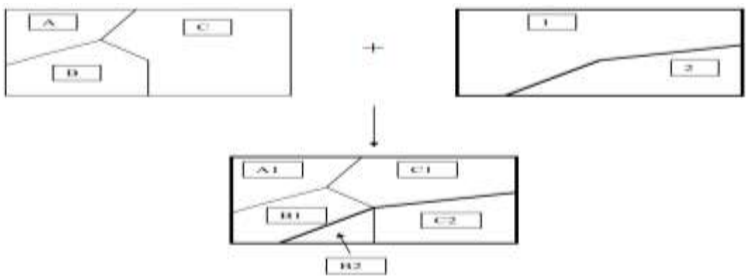
Figure (4.4) shows the result of the overlay operation, in the form of a new data set. Each of the polygons in the new data set represents a unique combination of the inputs, and the boundaries of the new polygons are combinations of the lines from the two inputs. The operation is actually reversible, since it is possible to recover both of the input data sets by deleting some of the attributes of the new polygons and merging polygons with the same remaining attributes. For example, by deleting land use attributes we recover the county data set. In other words, no information is lost or created in polygon overlay instead, the information is simply rearranged.

The new data set makes it easy to answer questions like the one we started with. To determine the amount of land in County A in agricultural land use, we simply identify all polygons in the new data set with those attributes, compute their areas, and total. Polygon overlay has a myriad of uses, of which this kind of query among the simplest. One popular use has to do with population statistics. Suppose a water supply agency wishes to estimate consumption in one of its service areas, and needs to know the total number of households in the area. The census provides statistics on total numbers of households, but for areas defined by the census, not those used by the water supply agency. So a simple way to estimate households is to overlay the service areas on the census areas, and measure the areas of overlap. Then the household counts from the census are apportioned based on these areas, to obtain estimates that are often remarkably accurate. For example, if one census area has 1,000 households, and 50% of its area lies in Service Area 1, and the remaining 50% lies in Service Area 2, then 500 households are allocated to each of these service areas. When all census counts are allocated in this fashion, the service area counts are summed.

Important to recognize the differences between overlay in vector GIS and overlay in raster GIS. The latter is a much simpler operation, and produces very different results, for different purposes. The discussion of raster analysis below includes the essentials of this version of the overlay operation.

2.18 Raster analysis

Figure (4.3) Polygon overlay.

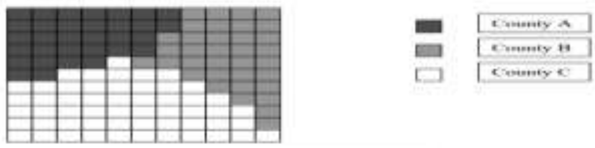


Raster GIS provides a very powerful basis for analysis, that is similar in many respects to the capabilities of other software that also relies on raster representations. For example, some of the raster operations described here will be familiar in concept to people who regularly use software for processing digital photographs or scanned documents, or for processing the images captured by remote sensing satellites. Vector data sets are similar in form to those used in computer-assisted design software (CAD) and in drawing software, but it is unusual to find comparable methods of analysis in these environments.

In raster GIS the world is represented as a series of layers, each layer dividing the same project area into the same set of rectangular or square cells or pixels. This makes it very easy to compare layers, since the pixels in each layer exactly coincide. But it also means that the ability represent detail is limited by the pixel size, since any feature on the Earth's surface that is much smaller than a single pixel cannot be represented. Smaller features can be represented by reducing the pixel size, but only at the cost of rapidly increasing data volume.

Each layer records the values of one variable or attribute, such as land use or county name, for each pixel. The recorded value might be the average value over the pixel, or the value at the exact center of the pixel, or the commonest value found in the pixel. If there are many attributes to record (the census, for example, reports hundreds of attributes for each county in the US), then separate layers must be created for each attribute. However, some raster GISs allow a more efficient solution to this particular problem, in which a single layer is used to record county ID for each pixel, and a related table gives the many attributes corresponding to that ID.

Figure (2.4) Raster cells to store the ID of polygons.



County	Population	Tax rate
A	1,998	5.5%
B	3,941	6.0%
C	15,227	6.75%

An excellent framework for thinking about raster analysis was provided some years ago by Dana Tomlin, who coined the term cartographic modeling (Tomlin, 1990). The basic operation types as following:

- Local operations, which process the contents of data sets pixel by pixel, performing operations on each pixel, or comparing the contents of the same pixel on each layer.
- Focal operations, which compare the contents of a pixel with those of neighboring pixels, using a fixed neighborhood (often the pixel's eight immediate neighbors).
- Zonal operations, which perform operations on zones, or contiguous blocks of pixels having the same values.
- Global operations, which are performed for all pixels.

2.19 Examples of operation types:

Local operations

The simplest kind of local operations occur when a single raster layer is processed. This is often done in order to apply a simple reclassification, such as: "All areas of Soil Classes 1, 3, and 5 are suitable building sites for residential development; all other areas are not." This could be operationalized by reclassifying the soil class layer, assigning 1 to all pixels that currently have Soil Class 1, 3, or 5, and 0 to all other pixels. Another might be: "Find all areas where the average July temperature is below 25 degrees Celsius." Again, this could be achieved by reclassifying a layer in which each pixel's attribute is that area's average July temperature. This kind of reclassification, using simple rules, is very common in GIS. Other local operations on a single layer include the application of simple numerical operations to attributes, for example, the July temperature layer might be converted to Fahrenheit by applying the formula:

$$F=9C/5+32 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.21)$$

GISs with well-developed raster capabilities, such as Idrisi, allow the user access to a wide range of operations of this type that create a new layer through a local operation on an existing layer. Other local operations operate on more than one layer. The raster equivalent of polygon overlay is an operation of this type, taking two or more input layers, and applying a rule based on the contents of a given pixel on each layer to create a new layer. The rules can include arithmetic operations, such as adding or subtracting, as well as logical operations such as:

"If the average July temperature is above 20 degrees celsius and soil type is Class 1 then the pixel is suitable for growing corn, so assign 1, otherwise assign 0." But notice how different this is from the vector equivalent. The operation is not reversible, since the new layer does not preserve all of the information in all of the input layers. Instead of reorganizing the inputs, raster overlay creates new information from them.

Focal operations:

Focal operations produce results by analyzing each pixel in relation to its immediate neighbors. Some of these operations work on all eight immediate neighbors, but others focus only on the four neighbors that share a common edge, and ignore the four diagonal neighbors. Sometimes we distinguish these options by referring to moves in the game of chess --

the eight-neighbor case is called the queen’s case, and the four-neighbor case is called the rook’s case.

Figure (4.5) two definition of a raster cell's neighborhood.

Among the most useful focal operations are the so-called convolutions, in which the output is similar to an averaging over the immediate neighborhood. For example, we might produce a new layer in which each pixel’s value is the average over the values of the pixel and its 8 queen’s case neighbors. Figure 4.8 shows a simple instance of this.

The result of this operation would be to produce an output layer that is smoother than the input, by reducing gradients, lowering peaks, and filling valleys. In effect, the layer has been filtered to remove some of the variation between pixels, and to expose more general trends. Repeated application of the convolution will eventually smooth the data completely. Convolutions are very useful in remote sensing, where they are used to remove noise from images. Of course averaging can only be used if the values in each pixel are numeric, and measured on continuous scales, so this operation would make no sense if the values represent classes of land, such as soil classes. But in this case it is possible to filter by selecting the commonest class in the neighborhood, rather than by averaging. Note also that special rules have to be adopted at the edge, where cells have fewer than the full complement of neighbors.

***Figure (4.6) Application of a convolution filter*.**

10	9	7	4	5
9	6	5	3	4
6	3	2	2	3

↓

8.5	7.5	5.7	4.7	4.0
7.2	6.3	4.6	3.9	3.5
6.0	5.2	3.5	3.2	3.0

If the input layer is a digital elevation model, with pixel values equal to terrain elevation, a form of local operation can be used to calculate slope and aspect. This is normally done by comparing each pixel's elevation with those of its eight neighbors, and applying simple formulae (given, for example, by Burrough and McDonnell, 1998). The result is not the actual slope and aspect at each pixel's central point, but an average over the neighborhood. Because of this, slope and aspect estimates always depend on the pixel size (often called the distance between adjacent postings), and will change if the pixel size changes. For this reason, it is always best to quote the pixel size when dealing with slope or aspect in a GIS.

Slope and aspect can also be used to estimate the pattern of surface water flow over a DEM, a very useful operation in determining watershed boundaries and other aspects of surface hydrology. Each pixel's elevation is compared to those of its eight neighbors. If at least one of the neighbors is lower, then the GIS infers that water will flow to the lowest neighbor. If no neighbor is lower, the pixel is inferred to be a pit, in which a shallow lake will form. If this rule is applied to every pixel in a DEM, the result is a tree-like network of flow directions (Figure 4.9), with associated watersheds. Many more advanced versions of this simple algorithm have been developed, along with a range of sophisticated methods for studying water flow (hydrology) using GIS (see Maidment and Djokic, 2000).

Figure (2.7) Digital elevation model.

Another powerful form of local operation forms the raster equivalent of the buffer operation, and also supports a range of other operations concerned with finding routes across surfaces. To determine a buffer on a raster, it is necessary only to determine which pixels lie within the buffer distance of the object. Figure 4.8 shows how this works in the cases of a point and a line. But suppose we make the problem a little more complex, by asking for pixels that are within a certain travel time of a given point, and allowing travel speed to be determined by a new layer. Figure 4.9 shows such a layer, and the result of determining how far it is possible to travel in given numbers of minutes. The operation is also known as spreading. Of course this assumes that travel is possible in all directions, but this would be appropriate in the case of an aircraft looking for the best route across the Atlantic, or a ship in the open ocean.

Figure (4.8) The buffer operation in it's raster from.

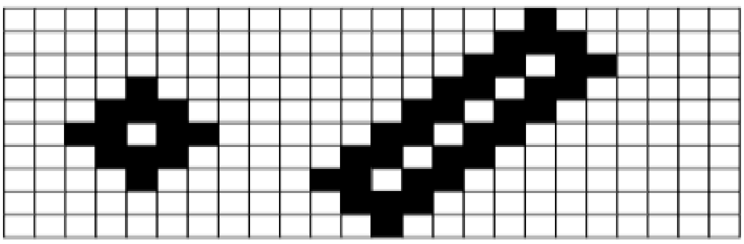
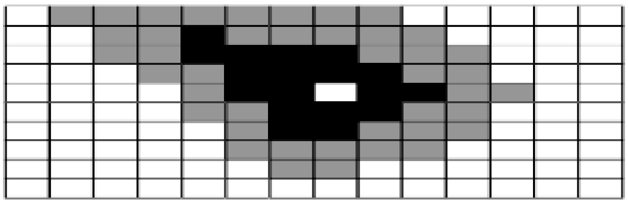


Figure (4.9) Illustration of a spreading operation over a variable friction.



Zonal operations

Rasters represent everything in pixels, so line objects and area objects are modeled as collections of pixel values, with no explicit linkage between them. Zonal operations attempt to address this, by focusing on operations that examine a layer as a collection of zones, each zone being defined by contiguous pixels of the same class. Zonal operations can be used to measure the areas of patches, or their perimeter lengths, and the results are returned as new layers in which each pixel is given the appropriate measure, evaluated over the zone of which the pixel is a part.

Global operations

Finally, global operations apply a simple manipulation to an entire layer, usually returning a single value. Examples include operations to determine the average of all pixel values, or the number of pixels having a specific value, or the numbers of pixels having each unique value, or the total area having each unique value. For example, a global measurement of area on a a soil layer with six classes might return six areas, plus the total area. Global operations produce summary statistics, so they are often the last operation in a sequence.

Optimization and design

Many applications of GIS analysis provide information that can be used to make decisions, but some methods come much closer to recommending decisions directly. These are the methods that focus on optimization, by finding the answers that best address a problem, and methods that focus on design. Ian McHarg, then Professor of Landscape Architecture at the University of Pennsylvania, who developed a method of overlaying maps to find the best locations for new developments (McHarg, 1969), advocated some of the earliest of these. Each map to be overlaid represented some specific issue, such as preservation of agricultural land, or cost of construction, so that when all maps were overlaid the best location would be the one with the least impact or cost. McHarg developed this as a manual method, using transparent sheets, but it is easily automated in a raster GIS.

Today, there are many methods of optimization and design available in GIS, in addition to the one already discussed. Many are concerned with optimum routing and scheduling of vehicles operating on a road network, a type of problem that is almost always implemented in vector GIS. In the simplest instance, the GIS is used to find the shortest path through a road and street network between a user-specified origin and destination. Millions of people use this kind of analysis daily when the access Web sites such as www.mapquest.com to request driving directions between two street addresses, and the same kind of analysis is possible in the in-vehicle navigation systems that are becoming increasingly common, especially in vehicles from car rental agencies

Much more sophisticated versions of the same basic idea are used by school bus authorities to design the best routes to pick up children and take them to school, by parcel delivery companies to design routes for their drivers, and by utility companies that schedule daily work orders for their maintenance staff. The term logistics is often used to describe this application of GIS. Another class of optimization problems is termed location-allocation. Here the issue is to find sites for activities that serve geographically distributed demand. Examples are retail stores, which must locate centrally with respect to their customers, schools and hospitals, fire stations and ambulance depots, and a host of other services. In all cases, the objective is to find locations that are best able to serve a dispersed population. The problem is known as location-allocation because the best solution involves both the location of one or more central facilities, and the allocation of demand to it, in the form of service areas (though in some cases this pattern of allocation is controlled by the system designer, as in the case of school districts, but in other cases it is determined by the behavioral choices of customers). In other versions of the problem there is no location as such, but instead an area must be divided into the best set of districts, for such purposes as voting, or the assignment of territories to sales staff.

We see many and different phenomenon in our everyday life effect by some factors regarding to increasing or decreasing, which cause statistical relationships which need suitable statistical method or methods to suit the type of data; whereas we find that the regression is one of these statistical methods which determined the form and direction of the relationship(1) (Seber, G. A., & Wild, C. J. (2003).; so it used widespread in many economical, sociological, environmental fields (2)(Rawlings, J. O et al . (2001). and it answers the relationship question between two variables or more; is this relationship positive one or negative? What are the most important elements or variables that effect in finding an arithmetical relationship between a group of variables? What are the Influential leads to specific result? Where does this relationship increase and where does it decrease? One of the two variables is called dependent while the other one is independent which is the cause in changing the dependent, in such case it calls (Simple linear Regression) which can be more than one variable, in this case it calls (Multiple linear Regression) (3) (Montgomery, D. C., et al. (2012) Furthermore the two types of regressions one of the advanced statistical method that insure accuracy in adherent for improving the research results through the optimal usage of data in finding casual relationship between the phenomena of this subject of the study.

Because of the change that occurred on the environment and the increase in population which effect negatively or positively on the natural environmental phenomenon within the extend geographical area whereas a need to show the effect of the spot phenomena protrudes on the increase or decrease of the phenomena on a specific geographical spot; and to what extend its difference on the other spots. This kind of regression is called multi spatial regression which is an attempt to find an equalization to the regression which enters dimensional spatial (spot's abscissas) within it, it cares of element value or that appeared in a specific geographical point or the spatial spot (x, y) whereas it finds the equalization of the regression appliances to apparent spots.(4)(Mauricio Bini, L et al . (2009).

2.20 Spatial Regression:

The spatial Regression Analysis is trying to find an answer to the relationship between tow variables or more, whereas this relationship a positive one or negative? The elements or variables that effect the mathematical relationship between the spatial variables and the factors or influential lead to a specific result. Spots variables increase and decrease.(5)(Tiefelsdorf, M. (2006).

It is a statistical method searches how to find regression equalization which enters the spatial dimension the axis (x , y) and we have a specific element changes within spatial spots with weighted abscissas and these values in these points represent the samples.

There are various types of the Spatial Regression Analysis concern of the relationship between a variable or more, whether these variables are positive (rejected) or negative (contrastive). The most important elements or variables that effect finding the Mathematical Relationship between two groups of spatial variables also the effective factors lead to a specific result such as increasing or decreasing within specific places. It is possible to compare multi spatial Regression to the common multi linearity regression as follow:

- The Common Multi Linear Regression is an equation for each study area, however the Spatial Regression is one changing equation to each part (a spot) from the study area (the value of variables are spatially changed).
- The Common Multi linear Regression its variable values are stable, while the data of Spatial Regression changes according to the spot.

2.21 Geographical Weighted Regression:

The Geographical Regression is considered to be one of the local regression models (or the favorite) most used one, whereas it is connected to one variable depends on independent variables groups and it is considered to be one of the most important regression's egression because it estimated as a marked one that connect every single independent variable on accredited one.(6)(Brunsdon, C., et al (1998).

The Geographical Weighted Regression is the advanced statistical technology that allows the operations sampled spatially. In other words, GWR implement the regression to each single site replacing the only regression for the whole study area, whereas there will be a spatial distribution to the GWR's results. There are three reasons at least we think that the relationship between the variables will vary from first to last place.(7)(Fotheringham, A. S., et al (2003).

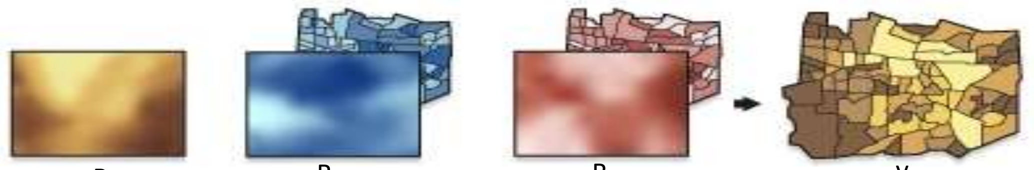
Firstly: there is no doubt, there is spatial difference to the samples because of the random variables samples that participate to this source which will be unconventional spatially, whereas these differences do not have great effect by itself but it should be taken in consideration.

Secondly: human attitude and behavior and residents'' practicing including their attitudes and favors, add to that their political and administrative issues, whereas these behaviors and residents' practice differ from place to another, and the result of that the necessity of taking care of the spatial dimension on the contrary to quantum analysis which ignore concerning spatial dimension without involving the spatial complications. The spatial treatment method relating to the spot is treated by respond to these complications and critics, which means responding to these critics.

Thirdly: Misconception to the model because of data gathering does not represent the reality usually; and any deletion leads to misrepresentation, that is why GWR is considered to be a method aids researchers and decision makers in distinguishing the local characteristics effects. Through using local model (the partial) and the earth querying model are not classical can easily obtain through which accurate local indications.

Weighted linear Regression analysis is used normally in sociological and economic studies also it uses in geographical and health studies.

Figure (4.10) geographical regression tests spatial treatments to every independent variable in a form of a map.



Source:

http://wiki.landscapetoolbox.org/doku.php/spatial_analysis_methods:geographically_weighted_regression.

According to the equation:

$$y_i = B_0 + \sum_{i=1}^{n} B_K x_{ki} + \varepsilon_i \ldots\ldots\ldots\ldots\ldots(2.22)$$

Where y_i is the independent variable, B_k the coefficients, X_{ki} the independent variables, and εi the error term.

GWR enable local variations (over space) in the estimation of coefficients, thus the regression coefficient B_k takes different values for each location. This method generates a separate regression equation for each observation, which can be expressed as follows:

Where y_i is the dependent variables, B_k the coefficients X_{ki} the independent variables; (ui , vi) co – ordinate location of i, and ε_i the error term.

The estimator of term takes the form of:

$$\acute{B} = (X^T W(u_i, v_i)X)^{-1}(X^T W(u_i, v_i)Y) \ldots\ldots\ldots\ldots.(2.23)$$

Where W (ui, vi) is square matrix of weights relative to the position of (ui, vi) in the study area. XT W (ui , vi) is the geographically weighted variance – covariance matrix (estimation requires its inverse to be obtained) and Y is the vector of the values of the Coefficients Standards Error.

The matrix W (u_i, v_i) matrix contains the geographical weights in its leading and diagonal and (0) in it's off-diagonal elements.

$$W(ui, vi) = \begin{pmatrix} W1\ (u1, v1) & 0 \\ \cdot & \cdot \\ 0 & Wi\ (ui, vi) \end{pmatrix}$$

Each equation is calibrated using the different weights of the observations contained in the data set. The assumption is that observations one near an anther have greater influence on each other's parameter estimates than observations farther apart, according to Table's law. The weight assigned to each observation is based on a distance decay function centered on observation i. (2) (8) (Bao, J., Shi, X., & Zhang, H. (2018).

2.22 Creating weights matrix

The new PROC SPATIALREG available in SAS 9.4M4 estimates spatial regression models (and also the classical regression model) proposed by Anselin (1988). The general form of these models is (Anselin, 1988):

$$Y= pW_{1y}+XB +u \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.24)$$

$$\text{Where } u = \lambda W_{2u}+\epsilon$$

where y is an × 1 vector of dependent variable, $W_1 n \times n$ and W_2 are matrices of spatial dependence X, is $n \times k$ an matrix of explicative variables, β is k × 1 vector of parameters model,μ is an n x 1 vector of error, ϵ is an n x1 vector of the remainder error and ρ and λ are spatial effects of parameters of same classical regression model when ρ and λ are both equal 0, i.e. , when there is no spatial dependence. Others models such as SAR (Spatial Auto Regressive Model) and SEM (Spatial Error Model) are derived from (1) just letting $\lambda = 0$ and $\rho = 0$, respectively. More details about spatial regression models and how to use (PROC COUNTREG) with spatial components can be found in Wu and Chvosta (2016) and how to use (PROC SPATIALREG) can be found in Wu and Chvosta (2017).

To estimate these spatial regression models is required to declare in the statement WMAT= the spatial weights matrix W, which is not created automatically by the SPATIALREG Procedure. Thus, this paper aims to show how creating some possible forms for the spatial weights matrix in order to estimate these spatial regression models.

The spatial weights matrix, also known as W matrix, plays an important part in the spatial modeling. Its purpose is representing, quantitatively, the spatial structure between regions. For n regions, $R_1, R_2, \dots, R_n$, the elements W_{ij} of the × W matrix, are representing some proximity measure between regions R_i and R_j , $_{I,\ j}$ = 1,2, … , n, and by definition, its diagonal W_{ij} is zero.

Some choices for measure the spatial proximity are:

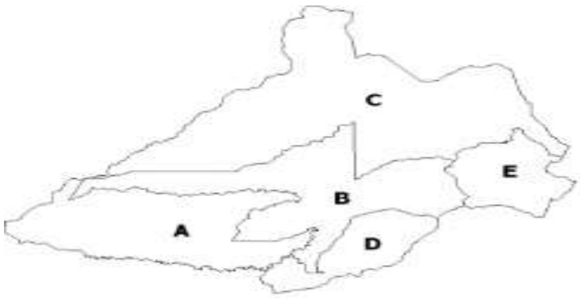
1) W_{ij}=1, if Ri shares border with Rj, and) Wij = 0 otherwise;
2)) W_{ij}=1, if Ri is one of the Nth nearest regions of Rj , and Wij = 0 otherwise;
3) W_{ij} = 1, if Ri is far from R_j by a distance of K miles (or kilometers), and Wij = 0 otherwise;

4) W_{ij} =1/(1 + d_{ij}), where dij is the distance between regions Ri and Rj . In this case it is necessary to force W_{ij} = 0; or you can just use W_{ij} = 1/ d_{ij}, for d_{ij} > 0.

5) W_{ij} = 1/(1 + t_{ij}), where tij is the distance between regions Ri and Rj . In this case it is necessary to force W_{ij} = 0; or you can just use W_{ij} = 1/ t_{ij}, for t_{ij} > 0.

The choices 4) and 5) are more realistic if they are computed using the network of highways, in order to capture road and/or traffic conditions, respectively.

An example of the choice1), also known as contiguity matrix, is in Figure 4.13. The representation has 5 regions, and it was attributed 1 for those regions which share borders and 0 otherwise. Note that region B shares borders with all others regions.



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\mathbf{W} = \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix} \end{array}
$$

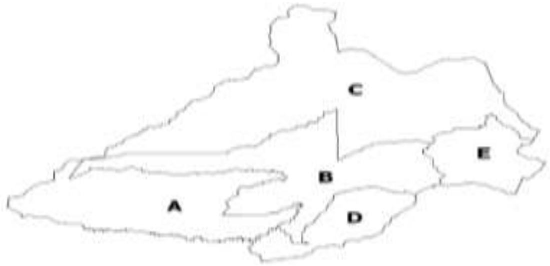
Figure (4.11) Example of a Contiguity Matrix

It is a common practice to use a row-standardized matrix instead of binary matrix, i.e., each row of this new matrix is summing to1. To do that, just compute each entrance of the matrix as:

$$w_{ij}^{*} = {w_{ij}}\Big/{\sum_{j=1}^{n} w_{ij}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.25)$$

The row-standardized matrix for Figure 1 is in Figure 4.12. We can see that the region B has 4 neighbors and because of that, the row-standardized matrix has values 0.25 for the neighbors, and the other regions has values 0.5 because they share border with only 2 regions:

Figure 4.12. Example of a Row-Standardized Contiguity Matrix



$$\mathbf{W}_s = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} 0 & 0.5 & 0 & 0.5 & 0 \\ 0.25 & 0 & 0.25 & 0.25 & 0.25 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \end{pmatrix} \end{array}$$

Table (2.2) : The difference between total statistical operations and the local (spatial)statistical.

Total statistic	**Local statistic**
Data are taken equally for the whole area	Search within spatial differences to total statistics
One value Statistic	Multi values statistic
Can't be represented by maps	Can be represented by maps
It is not a ally to GIS	It is a ally to GIS
Does not deal with place	Deals with place
excessive to resemblances to place	Excessive to difference on place
Searches symmetric phenomena to extract stabled equations	Searches local exceptions to extract what called hot spots

Fotheringham A. Stewart and others. Geographically Weighted Regression 2011 P6.

2.23:GWR's Interpretation and important factors:

Condition Numbers:

It is an standard for existence to what extend multi overlapping is between independence variables and the model's aptitude and quality; whereas if it is exceed 30th condition this implies that there is multi overlapping between independent variables which indicates incompetency and lack of quality for the model.

R-square (R^2) which determined the model's competency in interpreting the change within independence variables whereas its values vary between (0 to 1) when it becomes zero that reveals the weakness of the model performance and vice versa if it distancing from zero to one shows the model strength regarding interpreting the phenomena. The independence variables have an active effect in creating an effect on the dependent variable.

Predicted Values:

It is the calculated values to the dependent variable which resulted from the regression.

Residuals:-

The difference between the observed original value and the predicted to the Coefficients Standards Error whereas whenever residuals' value become tiny that showed the strength and quality of the model.

Coefficients Standards Error:

Values for measuring the quality of every single estimation to the regression regarding every single area whether separated or a geographical patch whereas these values are tiny the regression will be best.

The kind of the kernel carry out the regression operation, by using: adaptive value as an extend and calculated through determining the number of required neighbor.

Bandwidth Method: it determines the extend of the kernel carry out the regression operation so as to find the best suitable distance or the suitable one to the numbers of the neighbors. The user here has to choose one method out of the two for searching or sample process (9)(Guo, L., Ma, Z., & Zhang, L. (2008).

As generally, we found that the geographically weighted regression (GWR) performs a local of linear regression used to model spatially varying relationships as follows:

Figure (2.13) WGR for each feature (separate equation)

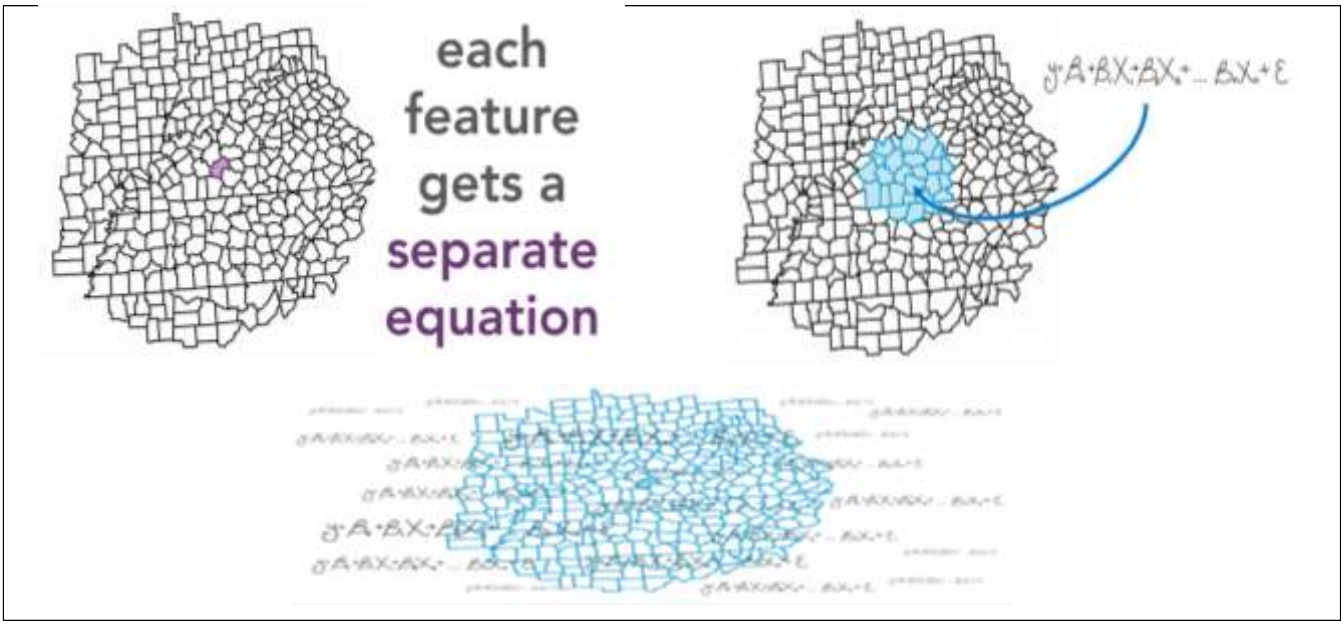
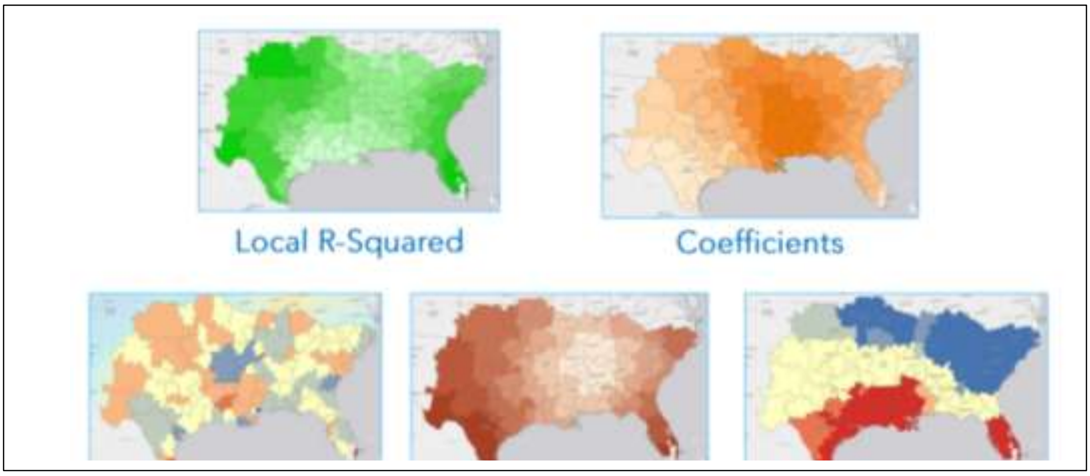


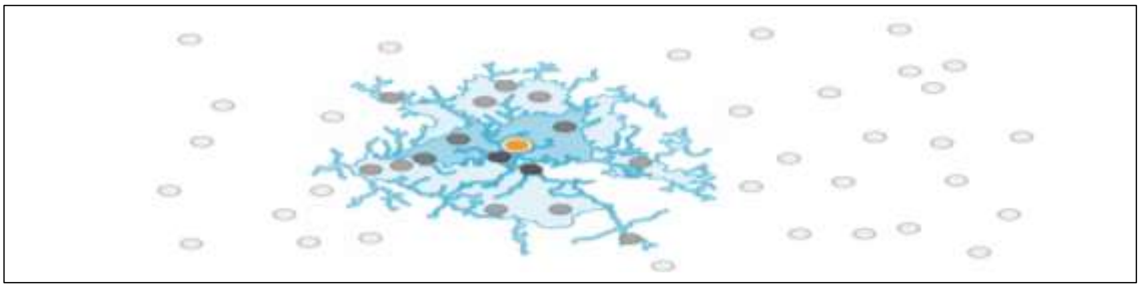
Figure (2.14) WGR's results.



2.24 Generate Network Spatial Weights:

Constructs a spatial weights matrix file (swm.) suing a Network dataset, defining feature spatial relationships in terms of the underlying network structure.

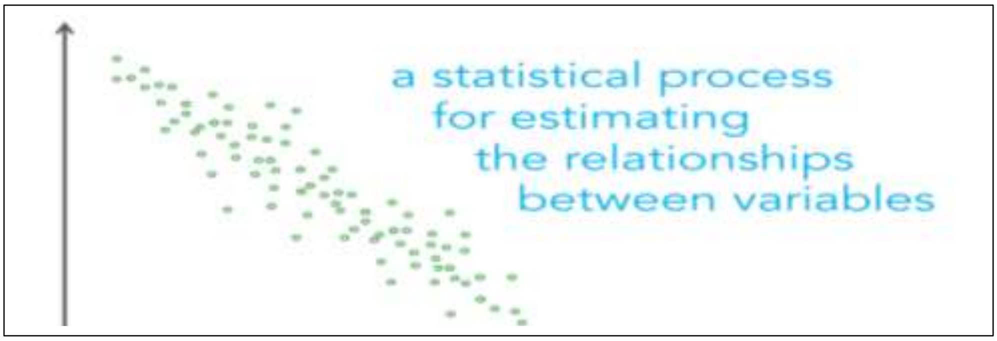
Figure (2.15) Generate Network Spatial Weights



2.25 Ordinary Least Squares:

Performs global linear regression to model a dependent variable in terms of its relationships to a test of explanatory variables:

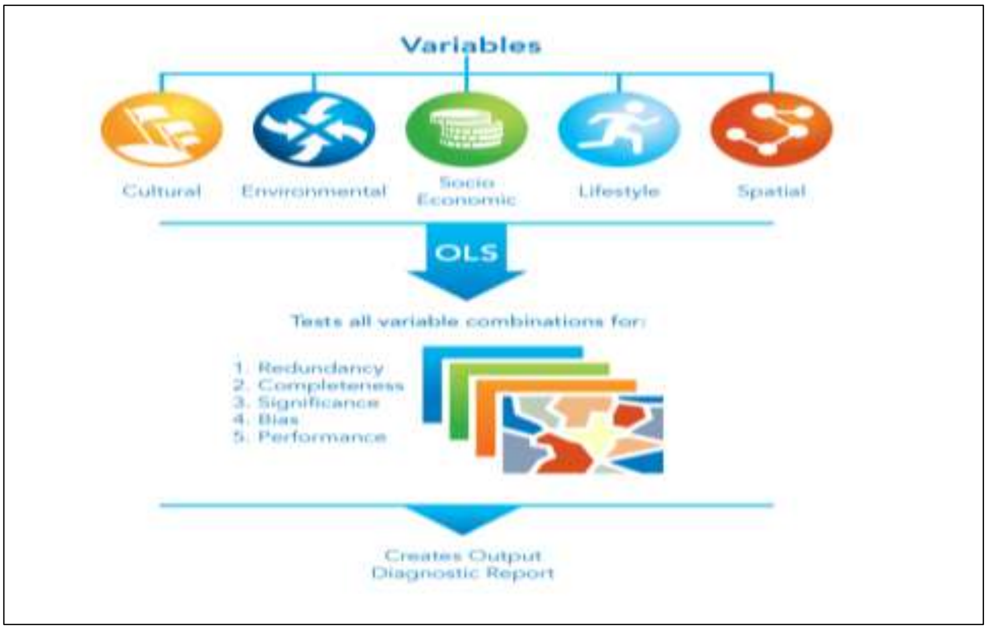
Figure (2.16) ordinary least square results.



2.26 Exploratory Regression:

Evaluates all possible combinations of the input candidate explanatory variables looking for properly specific OLS models.

Figure (2.17) OLS diagramme.



Chapter three

Applied aspect

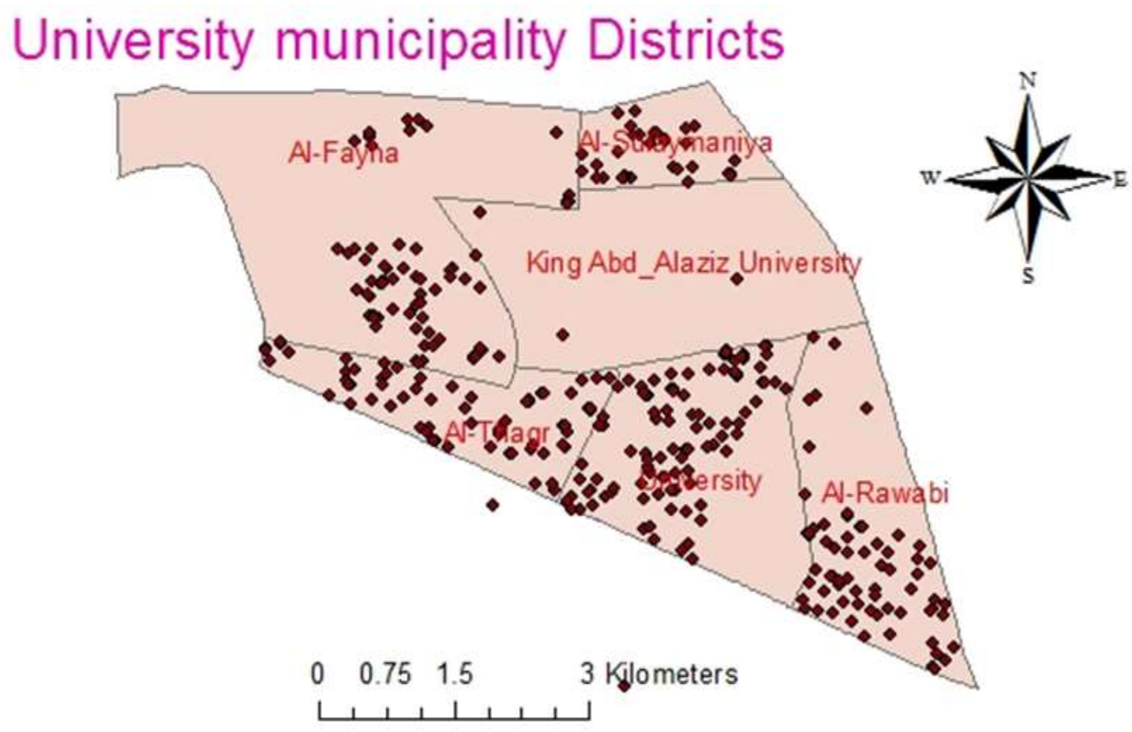
3.0 Introduction:

This chapter includes the applied aspect to what explained in the theoretical chapter and will describe the data, find the data patterns distribution, and build the Geographical weighted regression (GWR).

3.1 Data Set

Daily data for mosquitoes' reproduction were collected during the weeks (1-46) successively during 2018 for different University districts (Al-Thaghr, University, Al-Sulaymaniah, Al-Rawabi, Al-Fayha) using GPS technology for mosquitoes' reproduction pits during the said period.

Figure (3.1) mosquitos' distribution in university municipality.



Source: The researcher from applied study, GIS Software 2018

Spatial analysis in the GIS environment depends on the spatial characteristics of the phenomena sites under analysis. The spatial characteristics of any layer (Shape file) are represented in determining the type of used coordinates in the spatial signature, the type of used projector, as well as the earth model (reference) used in the layer. The researcher used the ArcGIS program since the program provides the possibility of non-spatial statistical analysis and spatial analysis of the data collected about the reproduction of mosquitoes in the university municipality as follows:

3.2 Descriptive Statistics:

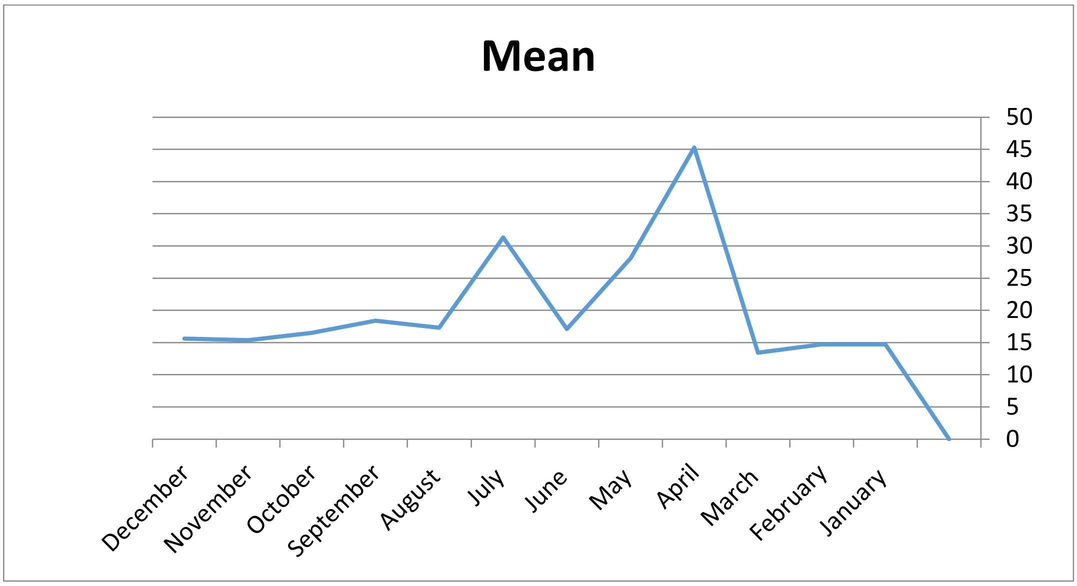
The table below shows calculation of mean and standard deviation, upper and lower bound of mean of mosquitoes' reproduction in university municipality districts:

Table 3.1: Non-spatial descriptive statistic for university municipality district per month.

Months	Mean	Standard deviation	95% Confidence Interval for Mean	
			Upper Bound	Lower Bound
January	14.7432	20.57547	17.4647	12.0218
February	14.7117	28.02161	18.4181	11.0053
March	13.4099	13.40522	15.1830	11.6368
April	45.2624	429.97816	102.2650	-11.7401
May	28.0724	173.73358	51.1044	5.0404
June	17.1176	25.30585	20.4725	13.7628
July	31.3122	134.00677	49.0776	13.5468
August	17.3258	31.22110	21.4648	13.1868
September	18.3937	27.61050	22.0540	14.7333
October	16.5023	22.79785	19.5246	13.4799
November	15.3575	30.52516	19.4042	11.3107
December	15.6380	21.83565	18.5328	12.7432

Source: the researcher from applied study, SPSS package, 2018.

Figure (3.2) (Non-spatial descriptive statistic for university municipality district per month.



Source: the researcher from applied study, Excel package, 2018.

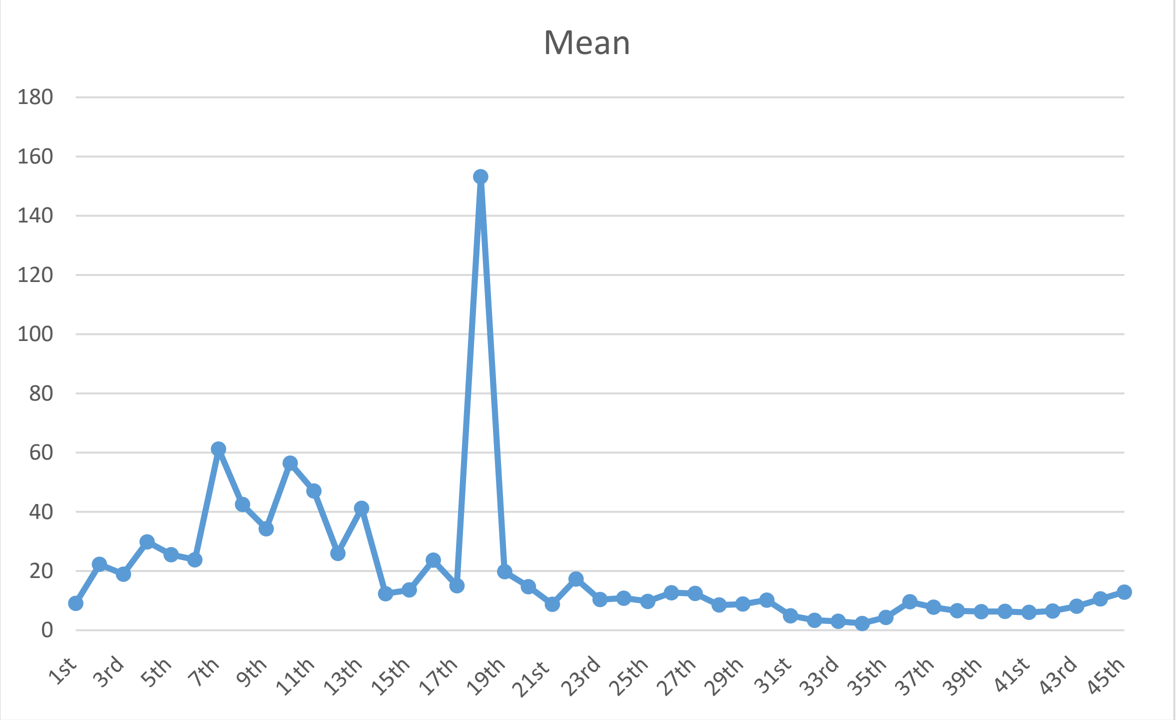
From table and figure above, it has been shown that according to the mean values in eleventh months, we found that April month recorded a high mean of reproduction (45.2624), flowed by July with mean (31.3122), May with mean (28.0724), September with mean (18.3937), August with mean (17.3258), June with mean (17.116), October with mean (16.5023), December with mean (15.6380), November with mean (15.3575), January with mean (14.7432), February with mean (14.7117) and the lowest reproduction was reported in March with (13.4099).

Table 3.2: Non-spatial descriptive statistic for university municipality district per week.

Week	Mean	Standard deviation	95% Confidence Interval for Mean		Week	Mean	Standard deviation	95% Confidence Interval for Mean	
			Upper Bound	Lower Bound				Upper Bound	Lower Bound
1st	9.1	11.2	12.0	6.1	24th	10.9	6.4	12.6	9.2
2nd	22.3	51.3	35.1	9.4	25th	9.8	6.4	11.4	8.2
3rd	19.0	16.6	23.1	14.9	26th	12.7	8.6	14.9	10.5
4th	29.9	32.2	38.1	21.7	27th	12.4	8.8	14.8	10.1
5th	25.6	22.2	31.3	19.9	28th	8.5	4.8	9.8	7.3
6th	23.9	19.4	29.1	18.6	29th	8.9	5.2	10.2	7.5
7th	61.2	229.7	125.8	-3.4	30th	10.2	10.7	12.9	7.5
8th	42.5	54.0	56.3	28.6	31st	4.9	3.8	6.0	3.9
9th	34.3	52.8	48.6	20.1	32nd	3.4	1.9	3.9	2.9
10th	56.5	163.5	96.1	16.9	33rd	3.1	1.4	3.5	2.7
11th	47.0	57.0	61.1	32.9	34th	2.3	1.2	2.7	1.9
12th	26.0	24.2	32.2	19.8	35th	4.4	2.7	5.2	3.6
13th	41.2	52.4	54.2	28.3	36th	9.6	7.1	11.5	7.8
14th	12.40	10.0	15.1	9.7	37th	7.8	6.6	9.6	6.0
15th	13.6	9.0	16.0	11.3	38th	6.6	6.6	8.4	4.8
16th	23.7	13.8	27.2	20.1	39th	6.3	4.1	7.4	5.2
17th	15.1	10.0	17.7	12.4	40th	6.4	4.5	7.6	5.1
18th	153.2	855.7	368.6	62.3	41st	6.1	5.5	7.6	4.6
19th	19.8	17.6	24.4	15.2	42nd	6.5	8.7	8.9	4.1
20th	14.7	11.9	17.9	11.6	43rd	8.1	12.6	11.7	4.5
21st	8.8	5.1	10.1	7.5	44th	10.6	28.8	18.4	2.8
22nd	17.4	15.6	21.5	13.2	45th	12.9	24.3	19.4	6.5
23rd	10.4	6.1	12.0	8.8					

Source: the researcher from applied study, SPSS package, 2018.

Figure (3.3): Non-spatial descriptive statistic for university municipality district per week.



Source: the researcher from applied study, Excel package, 2018.

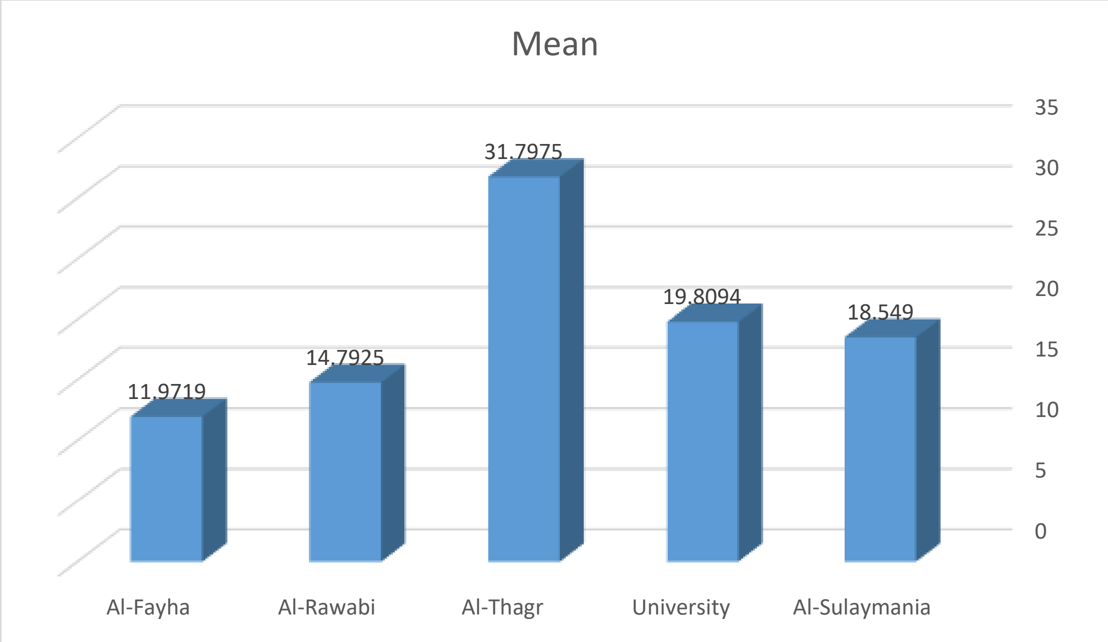
From table and figure above, it has been shown that according to the mean values in 45th week, we found that the month (18th) recorded the highest mean of reproduction (153.2), flowed by7th with mean (61.2), and the lowest mean (2.3) for 34th.

Table 3.3: statistic for university municipality district per districts. Non-spatial descriptive.

Districts	Mean	Standard deviation	95% Confidence Interval for Mean	
			Upper Bound	Lower Bound
Al-Sulaymania	18.55	27.38	21.93	15.17
University	19.81	91.45	25.68	13.94
Al-Thaghr	31.80	264.14	52.35	11.25
Al-Rawabi	14.79	20.13	16.55	13.03
Al-Fayha	11.97	11.91	13.28	10.66

Source: the researcher from applied study, SPSS package, 2018.

Figure (3.4) ***statistic for university municipality district per districts. Non-spatial descriptive***.



Source: the researcher from applied study, Excel package, 2018.

From table and figure above, it has been shown that according to the mean values in districts, we found that the district Al-Thaghr recorded the highest mean of reproduction (31.80), flowed by University with mean (19.81), Al-Sulaymania with mean (18.55), Al-Rawabi with (14.19), and the lowest mean (11.97) for Al-Fayha.

3.3: Spatial Analysis:

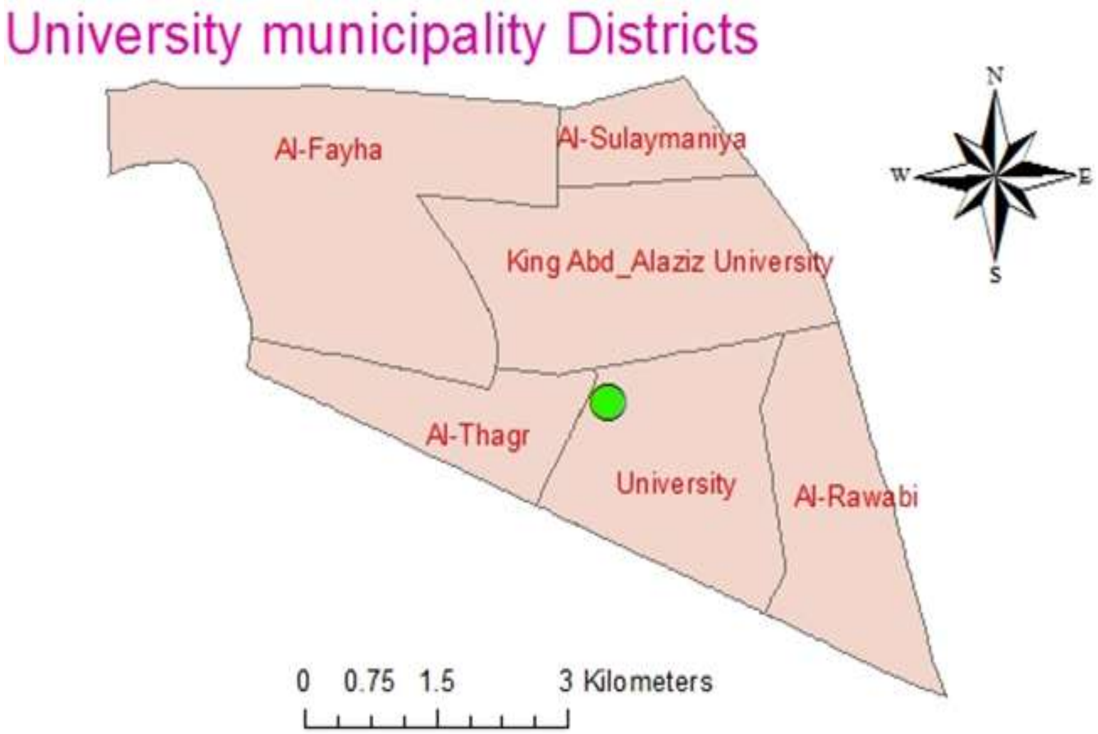
Necessary practical methods for entering data into ArcGIS program are followed, then spatial statistics to describe them spatially were performed and then to answer of the formulated hypotheses, as follows:

Mean Center:

The mean value of the two coordinates reached (39.244702, z = 20.660761 = y = 21.480965, x) by following the same steps of the normal calculation but for each coordinate separated from the other.

The value at this point is (20.660761), which is located inside the university district.

Figure 3.5: Mean center in University municipality.

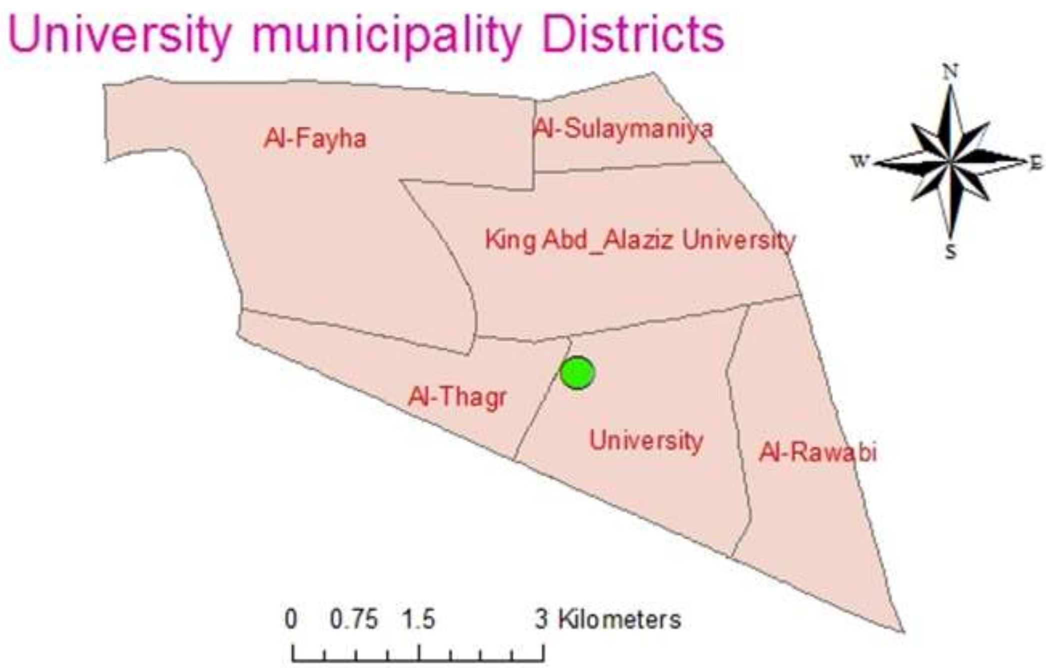


Source: The researcher from applied study, GIS Software 2018

Weighted Mean Centre

same steps of the mean center were followed, but the number of mosquitoes was chosen as the weight with the mean center and the result was (21.480224, Z = 982.82 = X = 39.244379, Y), which means that there is a difference in the weighted mean center compared to the center, where the value of Z reached at this Point (982.82) inside the university municipality

Figure 3.6 : Weighted mean center in University municipality



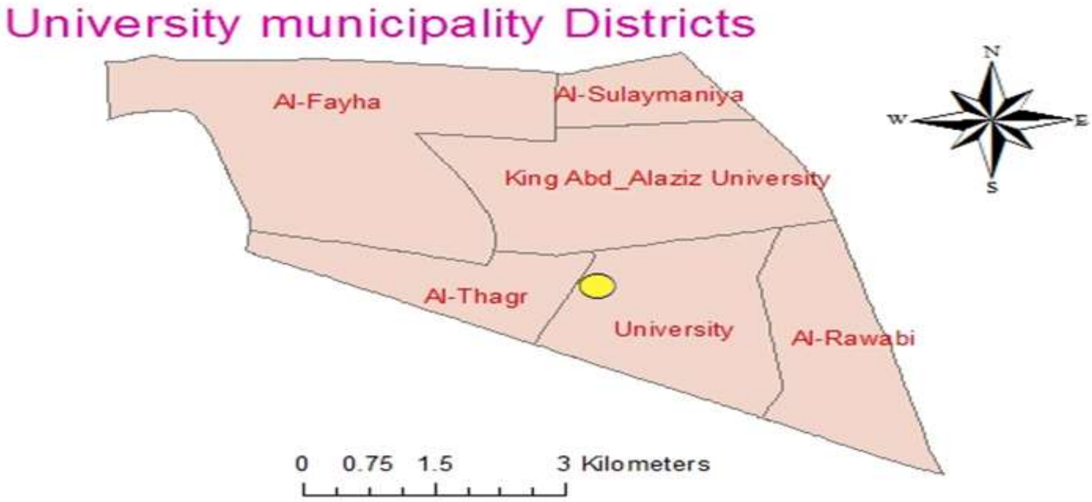
Source: The researcher from applied study, GIS Software 2018

It is clear from the shape of the figure 4 that the weighted mean of mosquito reproduction is located inside the university district near the Al-Thaghr district.

Median Center

All steps by which the median for non-spatial data is calculated are followed by arranging the variables ascending or descending and then calculating the median. However, for spatial data, each coordinate is calculated separately and the spatial median of the university municipality districts is (21.47878 = X = 39.242065, Y). as shown by figure (3.7):

Figure 3.7: The median center in University municipality



Source: The researcher from applied study, GIS Software 2018

It is clear from figure 5 that the location of the median is at the separating point between the university district and Al-Thaghr district and the bigger part of it is located inside the university district.

Standard Distance:

The standard distance value was calculated as in the following table:

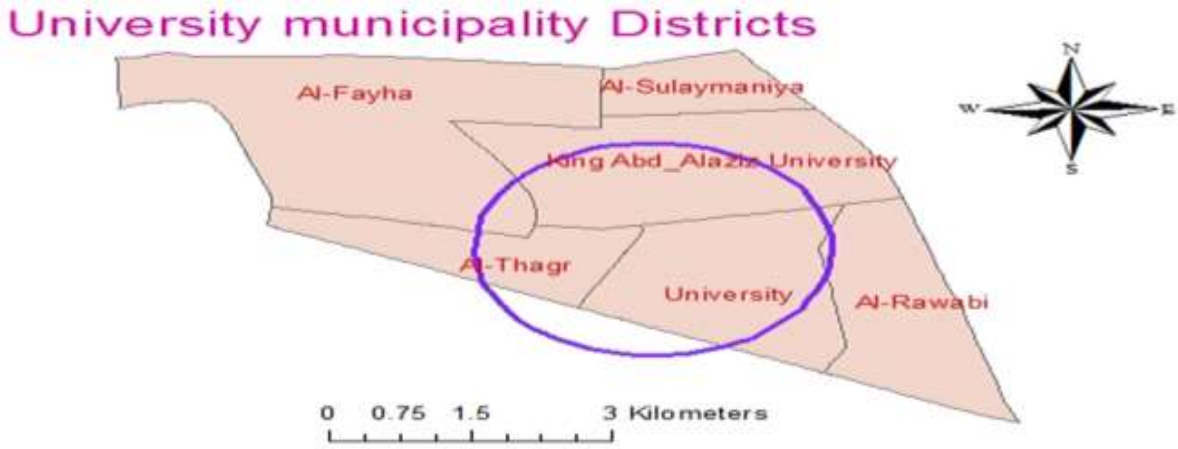
Table 3.4: Standard Distance of Mosquito Reproduction in the University Municipality

y Centre	x Centre	Standard Dist
21.480224	39.244379	0.017077

Source: The researcher from applied study, GIS Software 2018

Table 5.2 shows the value of the standard area for a small semi-circle is, which means that there is no dispersion in mosquitoes' reproduction and there is convergence in places of mosquito reproduction in the university municipality districts. i.e. spread in a convergent geographical locality.

Figure 3.8: The Standard Distance of Mosquito Reproduction in the University *Municipality*



Source: The researcher from applied study, GIS Software 2018

It is clear from the shape on the figure 5.6 map that the shape of the standard distance circle is widespread within the districts (university, Alt-Thaghr), which means that there is convergence in reproduction.

Distribution Direction

The following table shows the distribution direction statistics.

Table 3.5: Distribution Direction of Dengue Fever Mosquitos Transmitter in University Municipality.

Centre X	Centre Y	XStdDist.	YStdDist.	Rotation
39.244379	21.480224	.020441	.012861	120.6789

Source: The researcher from applied study, GIS Software 2018

Table 5.5 shows the coordinates of (x, y) as well as the radius of the standard distance in the direction of the coordinate (x Std. Dist. = 0.020441) and the radius of the standard distance in the coordinates (y Std. Dist. = 0.012861). We find that each value is small and the difference between them is small too, which means that the direction is about to take the form of a circle. Moreover, the rotation value reached (120.6789) and this value is located between the North and West directions (angle 90 and angle 180), which means that the direction of the rotation takes the direction towards Northwest, and that the extensions of this rotation is heading Southeast.

***Figure 3.9 : The distribution directions inside the university municipality*.**

Source: The researcher from applied study, GIS Software 2018

It is clear from the shape on figure 5.9 that the distribution direction of mosquitoes' reproduction sites is distributed in the South-West direction inside University municipality, and these extensions are located inside University and Al-Thaghr districts, and in small parts in each of (Al-Rawabi and Al-Fayha districts).

Hypothesis Testing:

Patterns analysis represented in the nearest neighbor analysis and Moran I analysis was used.

The First Hypothesis:

There is a spatial correlation of the mosquito reproduction phenomenon in University municipality.

Moran I statistic was calculated as follows:

Null hypothesis: The phenomenon of mosquitoes' reproduction is not spatially correlated (randomly distributed).

Alternative hypothesis: The phenomenon of mosquitoes' reproduction is spatially correlated (Non-randomly distributed).

The following table shows the results of Moran I test which is provided by ArcGIS program as shown in the following table:

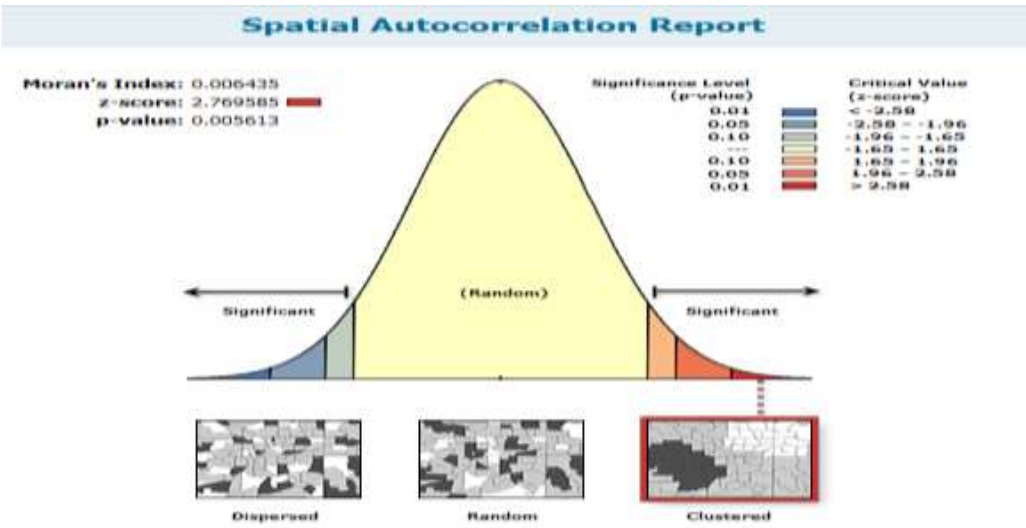
Table3.6: Moran's I Test

Global Moran's I Summary	
Index Moran's I	0.006435
Expected Index	-.000377
Variance	0.000006
Z-score	2.769585
p-value	0.005613

Source: The researcher from applied study, GIS Software 2018

Based on p-value (less than 0.01), there is sufficient evidence to reject null hypothesis and accordingly accept alternative hypothesis. This indicates that spread of mosquitoes in university municipality is (non-randomly distributed). Figure 3.10 shows these results.

Figure 3.10 : Shows the spatial Autocorrelation Report



Source: The researcher from applied study, GIS Software 2018

Through the index expressed for as Ratio between the observed or calculated distance divided by the expected distance, where the expected distance was based on a random distribution assumption by a similar number of points distributed over the same geographical area, and since the index reached (-.000377) less than one whole number, the geographical distribution of mosquito reproduction sites (Cluster). We find also the calculated Z value (2.769585) is greater than the (-2.58) and also the value of the calculated statistical index reached (0.005613) which is less than (0.01) which means acceptance of the alternative hypothesis, i.e. the phenomenon of mosquito reproduction in University municipality districts is clustered.

The Second Hypothesis:

The distribution pattern of mosquitoes' reproduction phenomenon in University municipality is random (spaced).

Null hypothesis: Spaced distribution pattern (non-clustered - random).

Alternative hypothesis: Non-spaced distribution pattern (clustered – random).

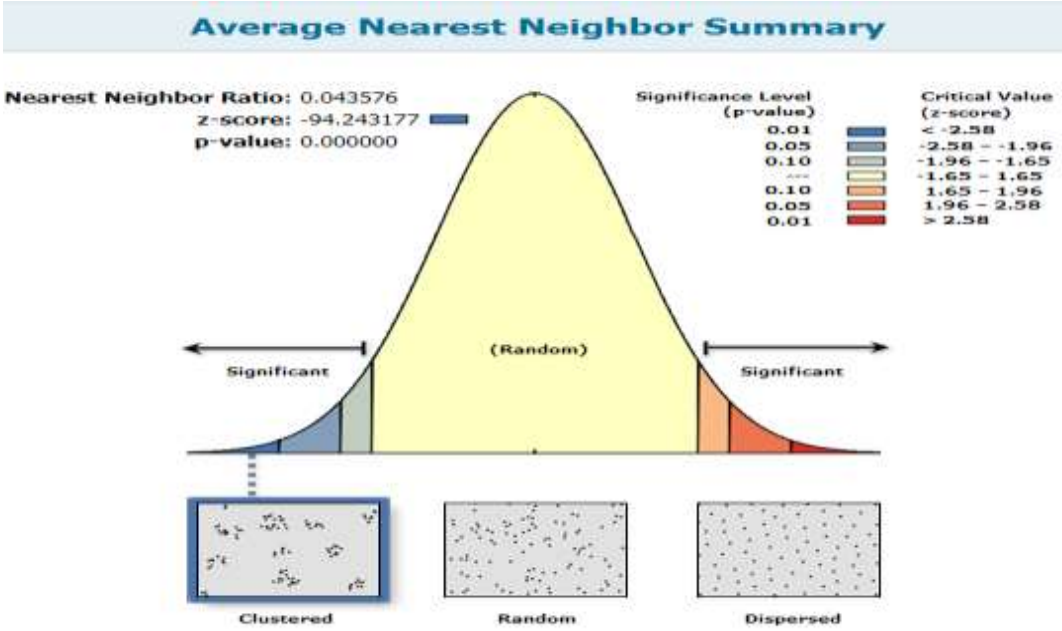
Table 3.7: Nearest neighbor results

Average Nearest Neighbor Summary	
Observed Mean Distance	2.709136
Expected Mean Distance	62.170181
Nearest Neighbor Ratio	0.043576
Z-score	-94.243177
p-value	0.00000

Source: The researcher from applied study, GIS Software 2018

P-value less than 0.05. This indicate that there is sufficient evidence to reject null hypothesis and accordingly accept alternative hypothesis (there is clustered in mosquitoes' spread at University municipality.

Figure 3.11 : Shows the result of average nearest



Source: The researcher from applied study, GIS Software 2018

This result means that the calculated mean of mosquito's reproduction is less than the expected mean of the same geographic area (random distribution).

This result means that the calculated mosquito-breeding rate is lower than the expected average breeding rate on the same geographical area (random distribution). i.e. Means that the Index is expressed as a Ratio between the observed distance or the calculated divided by the expected distance, where the expected distance depended on an assumption of random distribution by similar number of points distributed over the same geographical area, and since the index reached (0.043576) less than one whole number, the geographical distribution for mosquitoes' reproduction is cluster. Moreover, the calculated Z value (2.769585) is greater than the (-2.58) and the value of the calculated Z reached (-94.243177) which is more than (-2.58) and the value of the calculated statistical index reached (0.00) which is less than (0.01) which means acceptance of the alternative hypothesis, i.e. the phenomenon of mosquito reproduction in University municipality districts is clustered.

The third hypothesis:

The mosquito breeding phenomenon follows Log-logistic distribution.

Null hypothesis: Mosquito breeding phenomenon follows Log-logistic distribution.

Alternative hypothesis: Mosquito breeding phenomenon doesn't follow Log-logistic distribution.

We use Easy Fit software as follows:

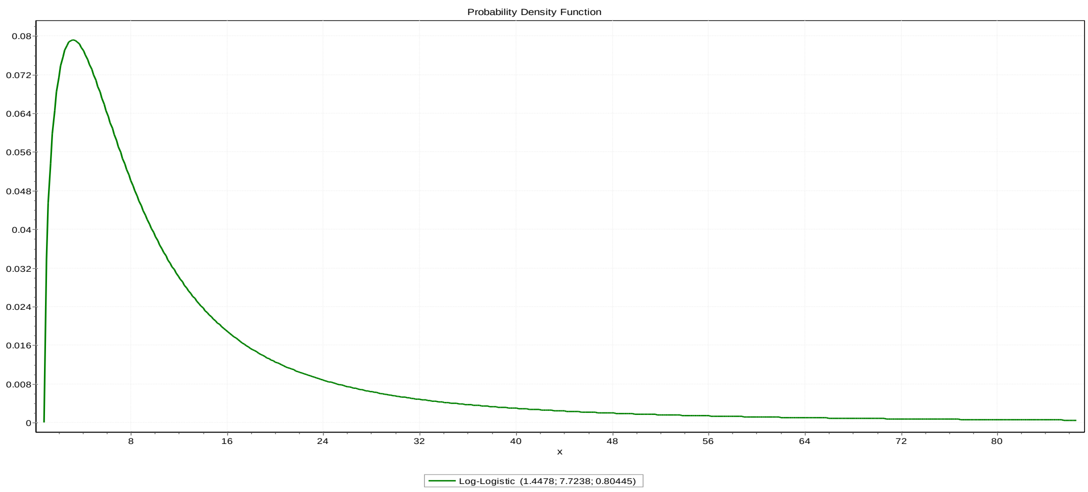
Table 3.8: Kolmogorov-Smirnov test.

Sample Size Statistic P-Value Rank	682 0.04106 0.19501 1				
A	0.2	0.1	0.05	0.02	0.01
Critical Value	0.04109	0.04683	0.052	0.05813	0.06238
Reject?	No	No	No	No	No

Source: The researcher from applied study, GIS Software 2018

Kolmogorov –Smirnov test has p-value (0.195101>0.05), this indicate that there is we accept null hypothesis; mosquito breeding phenomenon follows Log-logistic distribution.

Figure 3.12 : Mosquito breeding phenomenon distribution.

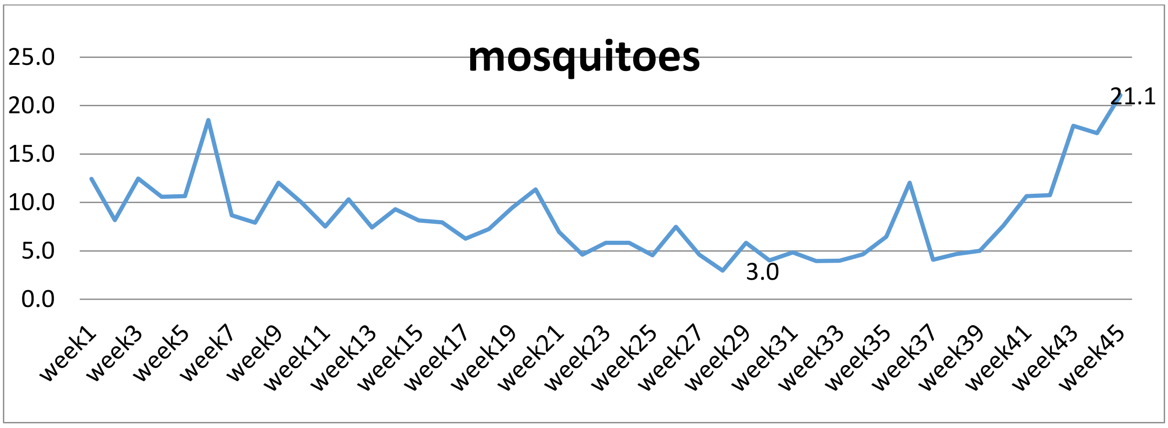


Source: The researcher from applied study, Easy fit Software 2018

3.4 Spatial Weighted Regression Results:

The data of the average of Mosquitoes' reproduction and multiplication were gathered within University neighborhood in (1 – 45) weeks of the year 2018; including the humidity and temperature within the same period of this study. The laborers and some employees working for Asma Company for Environmental Solution – which specialized in Environment Studies and Mosquitoes' Fight - carried out this part of the study regarding the type that carry Dengue Fever in Jeddah City.

Figure (3.13) The average of Mosquitoes' reproduction



Source: The researcher from applied study, Excel Software 2018

We can notice from Figure (3.13) that the spatial average for circulating the residuals vary between (21.1) in week 45th and (3) residuals in week 28th.

The reports that submitted to those in charge within Jeddah Secretariat were restricted on some descriptive statistics and showing the numbers only and the concentration places of reproductions. The researcher tends to use some advanced spatial statistical methods and knowing the effect of the spot regarding mosquitoes' breeding including some other spatial methods, whereas the data come as follows:

- The spatial average to the mosquitoes in the residuals within weeks (1 – 45) through the year as Coefficients Standard Error.
- The average to every week abscissas from week (1- 45) all the yearlong.
- The average of humidity for the weeks (1 – 45) all the yearlong 2018 was independent variable
- Average of temperature for the weeks (1- 45) all yearlong 2018 was independent variable

Which is spread as follows:

Table (3.9) :Descriptive Statistic to the Variables:

The variable	Greatest value	Less value	Average	Aberrance Apparatus
Number of Mosquitoes	21,078	2,95	8,53	4,12
Temperature	30,094	23,3	36,1	3,65
Humidity	68,7	34,2	51,43	8,01

Source: The researcher from applied study, ArcGIS Software 2018

3.5Results:

WGR was used to estimate the phenomena of mosquitoes carry Dengue Fever reproductions within University neighborhood accrediting the two variables (humidity and temperature) as independent variables according to following procedures:

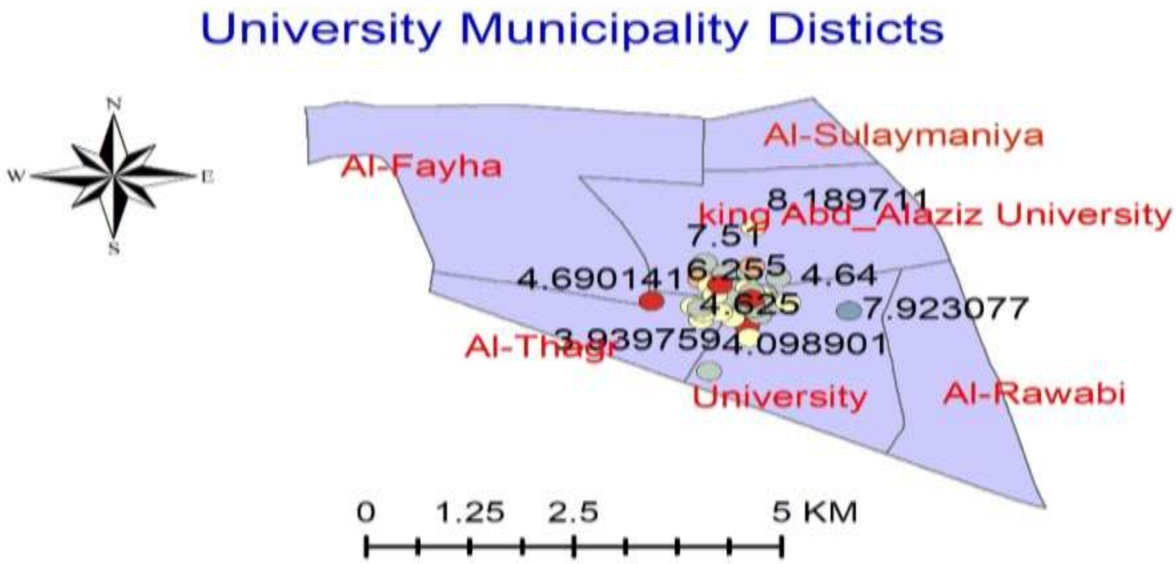
After following the special procedures regarding Geographical Data on Arc Map the Weighted Geographical Regression was calculated for each point of the mosquitoes' reproduction spots, (each separated equation to any single spot) does the result of that a number of 45 separated equations on a table consists of:

Observed values of independent variable which are the real value to the average numbers of the mosquitoes carry Dengue Fever within the neighborhood of University through the weeks (1– 45) the year 2018.

Weights:

Weights were calculated to every single geographic spot whereas least value was less than (2.954545) in the point (28) and the highest weight was on the spot (45) where the value reach (21.078292) distributed geographically as follows:

Figure (3.14) Mosquito's weights in university municipality districts.

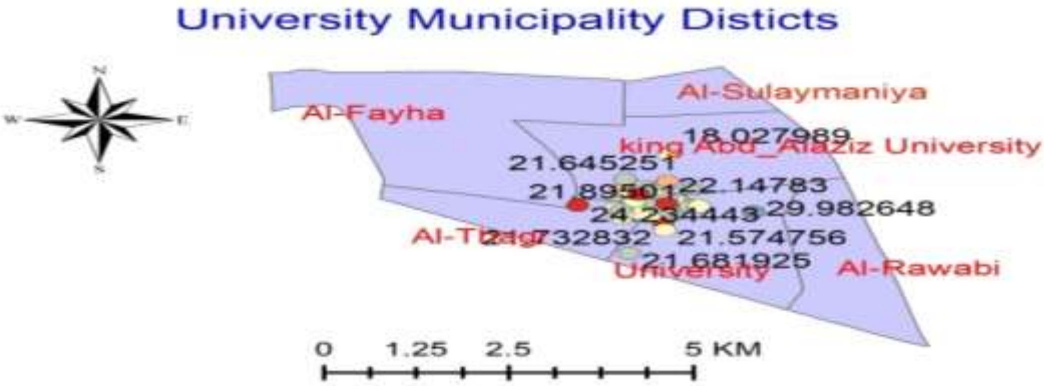


Source: The researcher from applied study, ArcGIS Software 2018.

Conditions Numbers:

It is a criterion to what extend is the linearity overlap to the independent variables; whereas we find that the condition numbers vary between (29m983 to18,028) less than (30) does the result of that means nothingness of linearity overlapping between the independent variables. Geographically distributed within Jeddah Municipal Neighborhood as follows:

Figure (3.15) Mosquito's condition numbers in university municipality districts.

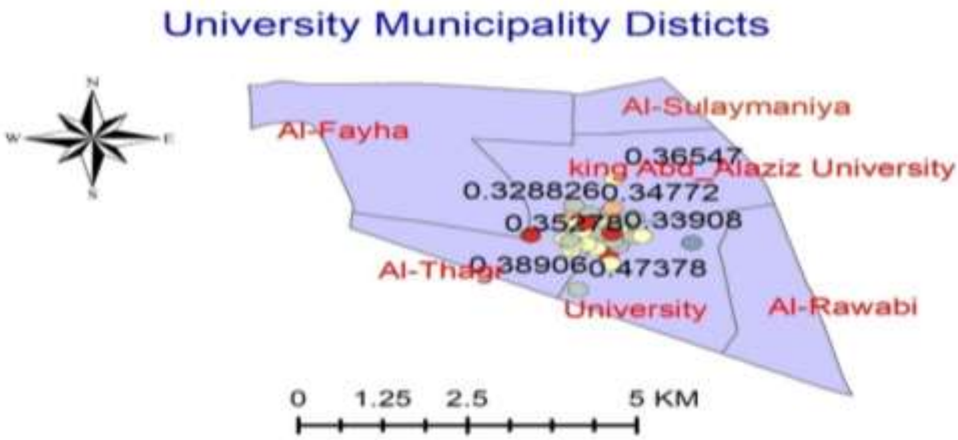


Source: The researcher from applied study, ArcGIS Software 2018.

R-Square:

We find that the values vary between (0.237416) at the point (8) as a least value and that means the average of humidity and temperature have effect on mosquitoes carry Dengue Fever within the neighborhood reached (24%) which is weak ratio. However, the highest value to it reached (0,610955) at point (45) which means that the proportion of temperature and humidity have effect on mosquitoes carry Dengue Fever reproductions (the residuals) at this point reached (64%) which is a high effect at this point. It geographically distributed as follows:

Figure (3.16) Mosquito's R square's in university municipality districts.

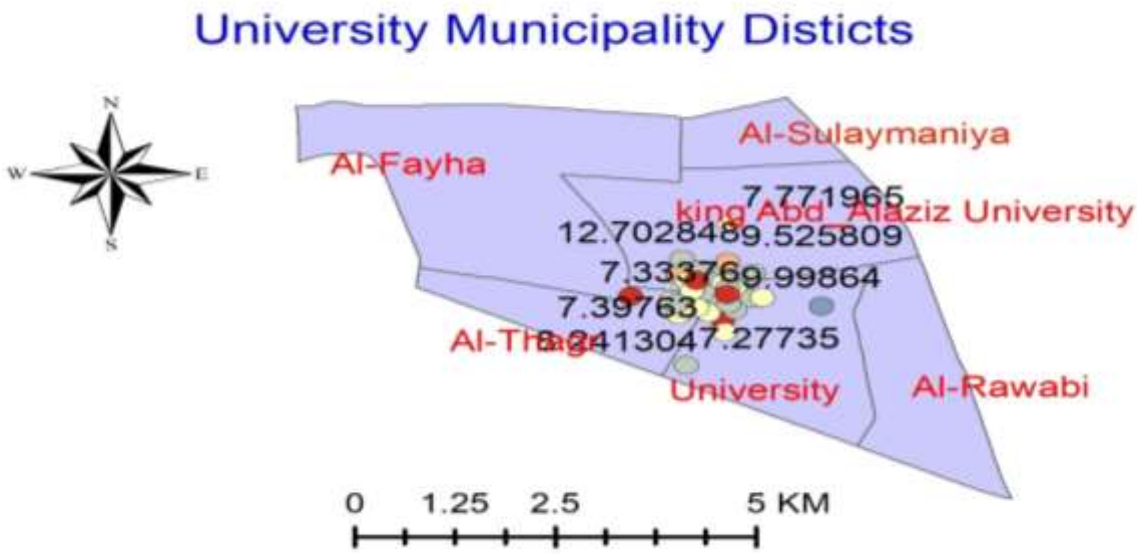


Source: The researcher from applied study, ArcGIS Software 2018.

Predicted Value: -

It is the predicted value to the numbers of mosquitoes carry Dengue Fever (the Residuals) which is resulted out of the regression equation for each single one, whereas we find that the highest value reached (17,758) at the point (45) in contrast we find least value reached (4.3.5) at the point (30). Geographically distributed as follows:

Figure (3.17) Mosquito's weights in university municipality districts.



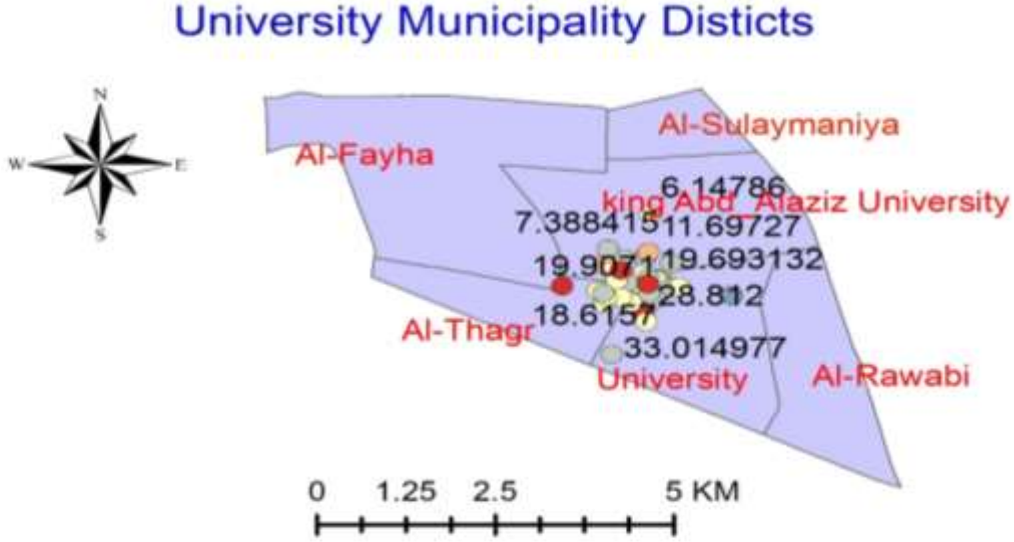
Source: The researcher from applied study, ArcGIS Software 2018.

Coefficients:

Intercept Coefficient:

The values vary from (6.148) and (23.015) which is an average of numbers of mosquitoes when taken in consideration the humidity and temperature, each of independent variables equal zero. Distributed geographically as follows:

Figure (3.18) Mosquito's intercept coefficient in university municipality districts

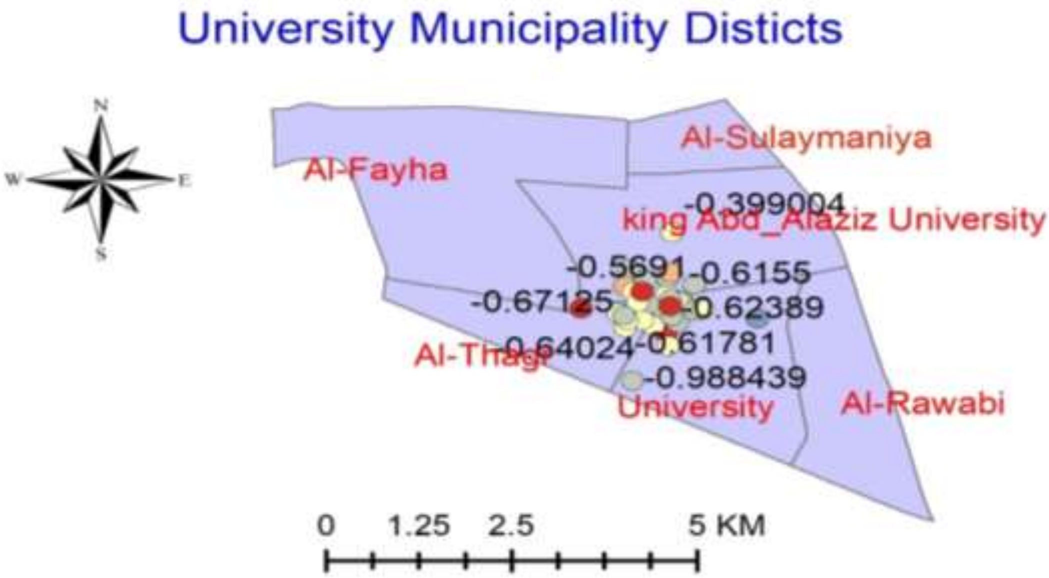


Source: The researcher from applied study, ArcGIS Software 2018.

Temperature Coefficient:

Its values vary from (0,98– 0,39) which means that the increase in temperature as a degree leads to decreasing in reproduction of mosquitoes carry Dengue Fever in an average of (0,98) as a least and as a highest a quantum decreases (.39) as a highest quantum. Geographically distributed as follows within the neighborhood of University municipal:

Figure (3.19) Mosquito's temperature coefficient in university municipality districts

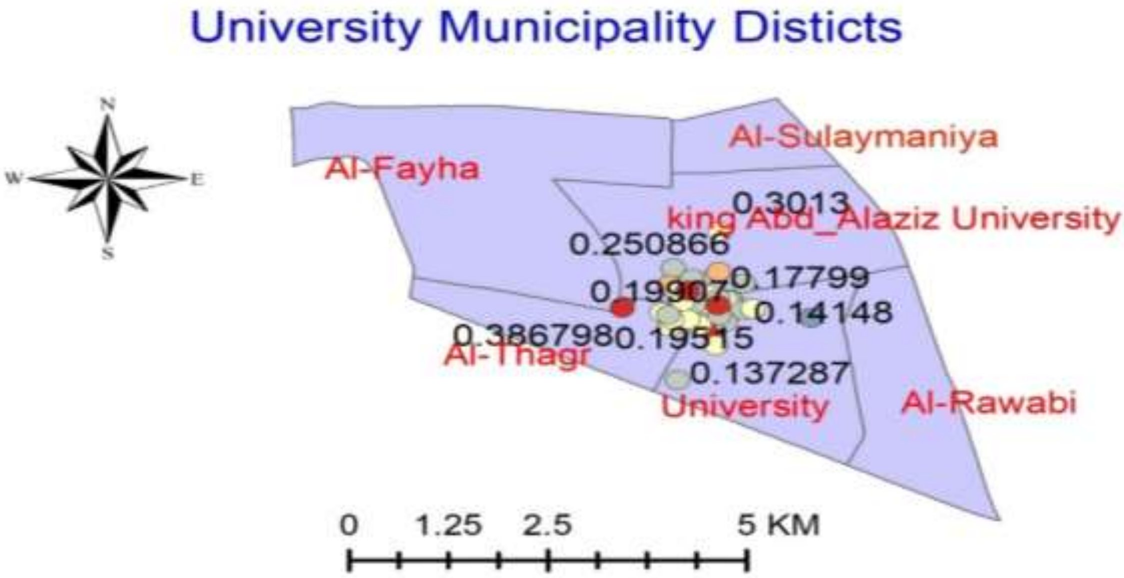


Source: The researcher from applied study, ArcGIS Software 2018

a. Humidity Coefficient:

Vary between (0.080461) as least and (0.386798) as a highest value to humidity. That means the increase in humidity as one degree leads to decrease in reproduction of mosquitoes carry Dengue Fever (0.0805) at the point (8) whereas increases in average regarding humidity a degree leads to increase in mosquitoes carry Dengue Fever multiplication (the adult) within Jeddah municipal by (0.3868) at the point (45) geographically distributed as follows:

Figure (3.20) Mosquito's temperature coefficient in university municipality districts

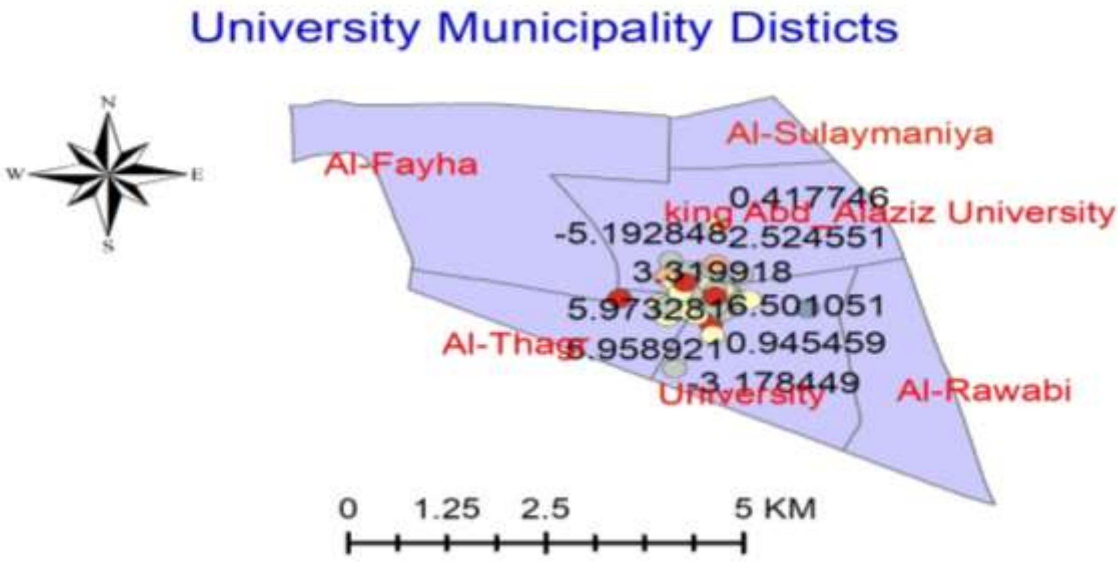


Source: The researcher from applied study, ArcGIS Software 2018.

Residuals:

It is the different between the real seen values and the expected output from the predicted operation. We find that the highest value to the difference reached (7,61) at the point (43) and the least value to the difference reached (-0,01) at the point (2)… geographically it is distributed as follows:

Figure (3.21) Mosquito's temperature coefficient in university municipality districts.

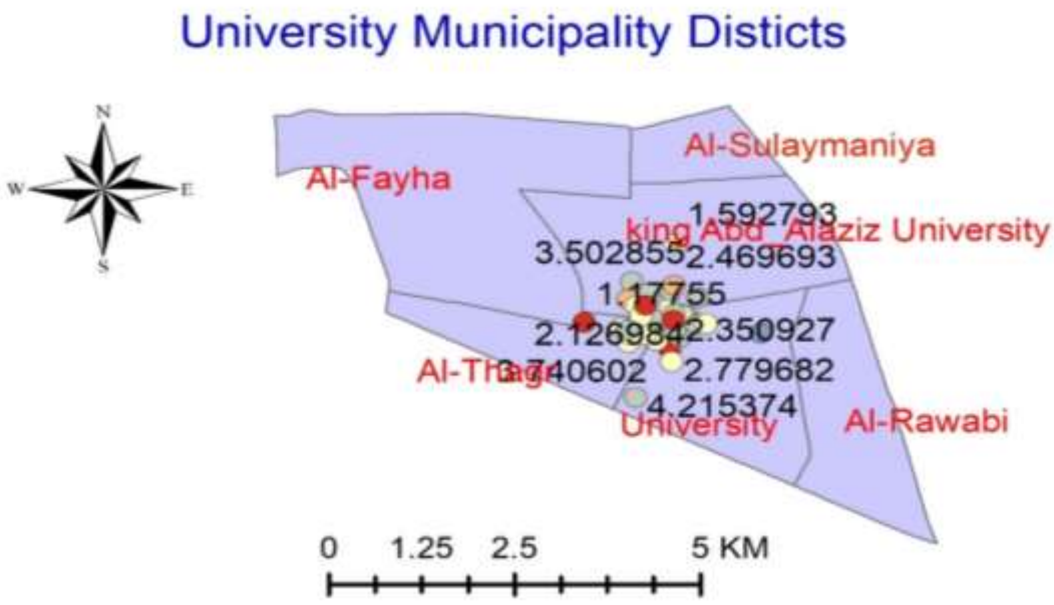


Source: The researcher from applied study, ArcGIS Software 2018.

Standard Error to the Sample:

We find that the least value reached (1,43) and this is an abdication that the regression at point (8) is considered to be the best equation because of the quality of the sample, while the greatest value to the standard error reached (2,95) at the point (12) which means that it was the weakest sample to the reproduction phenomena of mosquitoes carry Dengue Fever (larva period) within University municipal, geographically distributed as follows:

Figure (3.22) Mosquito's standard error to the sample in university municipality districts.



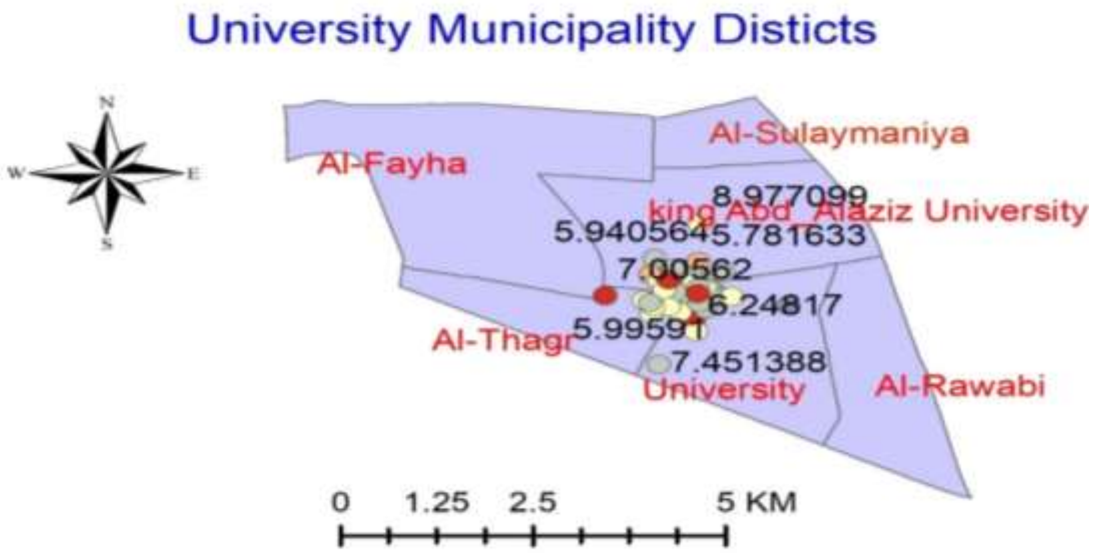
Source: The researcher from applied study, ArcGIS Software 2018.

Standard Error Intercept:

We find that the values vary between (7.27) at the point No (7) as the highest value, while the least value reached (5,01) at the point No (28).

We observe that the values are somehow raising, and geographically distributed as follows:

Figure (3.23) Mosquito's standard error intercept in university municipality districts.

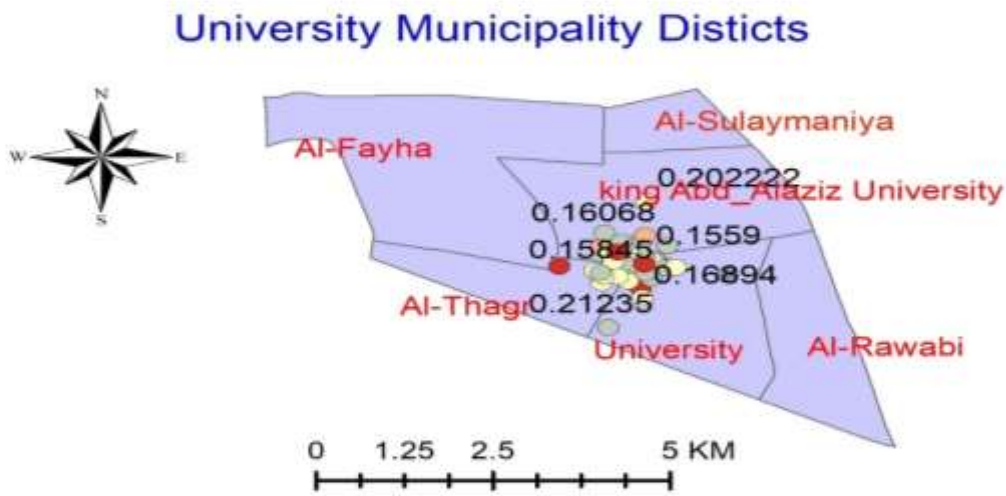


Source: The researcher from applied study, ArcGIS Software 2018.

Temperature Standard Error:

The Intercept Normative to Temperature reached (0,13030) at the point (16) as least value while the highest value reached (0,21634) at point (8), we observe that the values decrease and geographically distributed as follows:

Figure (3.24) Mosquito's temperature standard error in university municipality districts.

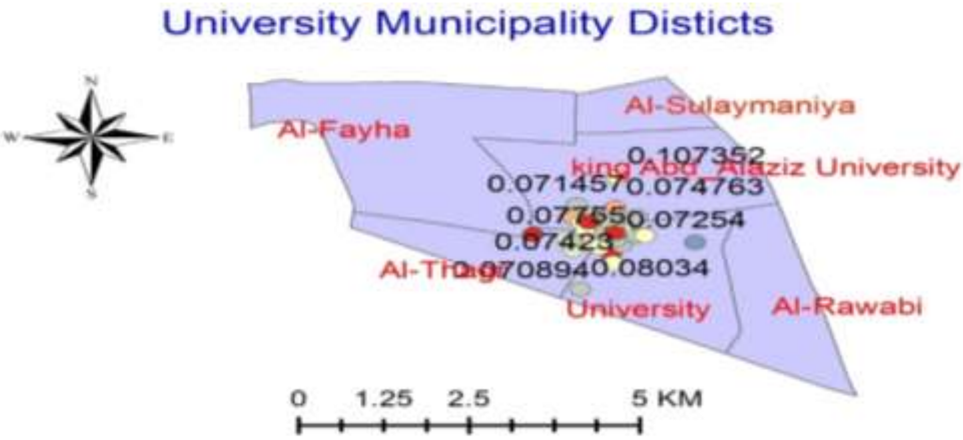


Source: The researcher from applied study, ArcGIS Software 2018.

Standard Error Humidity:

The Intercept Normative to humidity (0,05959) at point (43) as least value, whereas the highest value reached (0,08789(at point (7), we observed that the values reduced, and geographically distributed as follows:

Figure (3.25) Mosquito's standard error humidity in university municipality districts.

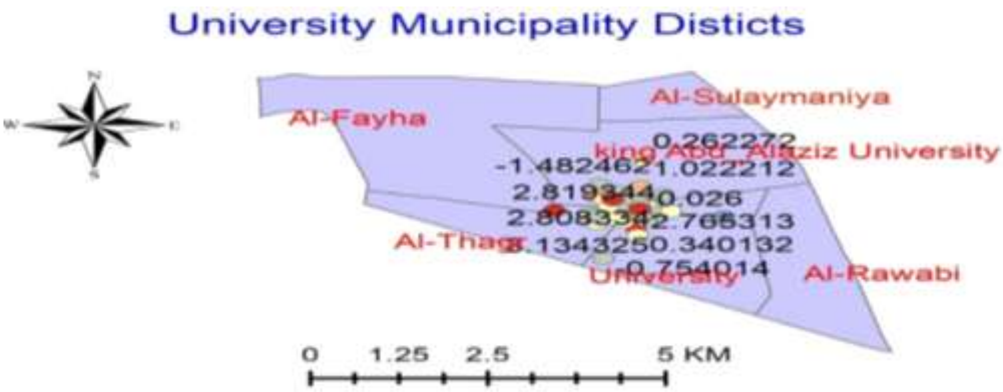


Source: The researcher from applied study, ArcGIS Software 2018.

Standard Residuals values:

The least value to the residuals reached (- 0,008) at point (2). However, the highest value reached (2,587) as biggest value to the residuals at point (44), and geographically distributed as follows.

Figure (3.26) Mosquito's standard error humidity in university municipality districts.



Source: The researcher from applied study, ArcGIS Software 2018.

Residuals Squares:

The value of total remaining residuals squares in the sample reached (342,474,32).

Effective Number:

Reflects the bargains between the discrepancy in the sample values and the value of aberrance regarding moduli estimation where it reached (7,43574).

Sigma Value: Value of the standard quadratic root for the major remaining or the value of normative aberrance to the remaining and they reached (3,01944).

AICs Modulus reached (235,405399).

R^2 Sample Modulus Competency (0,551687).

Formative corrected Modulus (the adjusted) reached (0,474879)

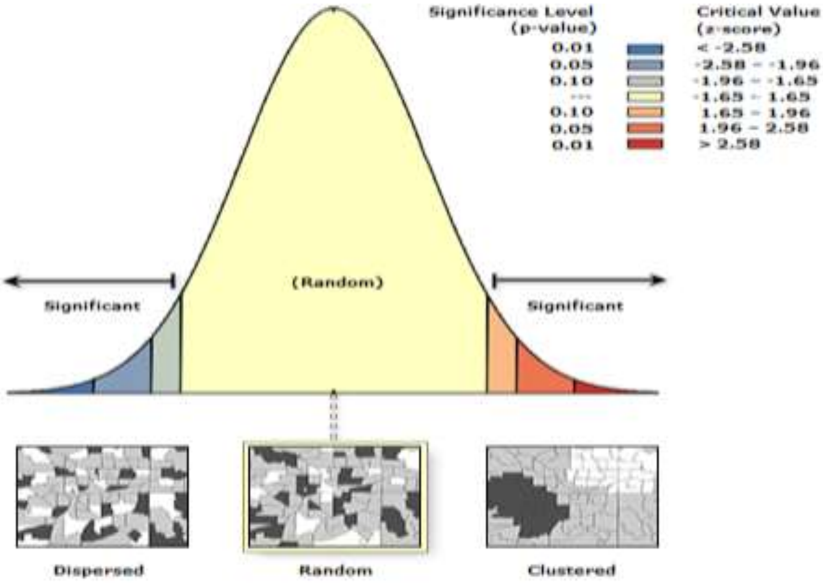
To attain to what extend the random errors mean something, a statistical modulus is used (Moran's Index) knowing to what extend is the concentration of the residuals.

Table (3.10) :Random of Errors Distribution Test.

Moran's Index	1,000
Z-score	70,9174
P-value	0,0000

Source: The researcher from applied study, ArcGIS Software 2018.

Figure (3.27) Mosquito's random of errors distribution Test.



Source: The researcher from applied study, ArcGIS Software 2018.

3.6 Test of the Quality of the Spatial Models:

To know the intangibleness of the conduct, the calculation (t) has been run to each modulus of the sample (B_0, B_1, B_2) according to the following equation.

The result was as the same as the attachments:

From table (t) the values of table had found statistically significant at level (0,05) and a degree (44), and whereas it reached (1,671) we find the transaction's values (t) (appendix), calculated to the fixed rim (B_0) maximum than the table one whereas it varies between (16,82) at the point (8) as a maximum value and (4,84) as a minimum value at the point (11) which means that the fixed rim is incorporeally at all points.

That why the comparison of temperature moduli at all points find the value scheduled to (t) greater than the accounted value to the (B_1) temperature at all points which means that there is no relationship of statistical denote to the temperature regarding mosquitoes' reproductions.

When comparing the humidity ratios at all points we find that the scheduled value to (t) is greater than calculated value for (B_2) humidity at all points and this means that there is no statistical relationship to the humidity behind mosquito's multiplication.

3.7 Compare between spatial models:

We will compare between three spatial models as table below:

Table (3.11): Compare between spatial models

Location	B	S.E	T	R-square
King Abd-Alaziz University	6.1479	8.977	0.684843	0.36
	-0.399	0.202	-1.97329	
	0.3013	0.107	2.8054	
University district	22.87	5.915	3.866441	0.3411
	-0.6698	0.165	-4.05939	
	0.1415	0.077	1.83053	
Al-Thaghr	11.891	7.006	1.697259	0.611
	-0.693	0.199	-3.47543	
	0.3868	0.084	4.599287	

Source: The researcher from applied study, ArcGIS Software 2018.

From table above we can built model for each three locations as:

$$\hat{y} = B_0 + B_1x_1 + B_2x_2$$

$$\hat{y} = number\ of\ mosqutios$$

$$x_1 = Averagetemperatures$$

$$x_2 = Average\ humidity$$

1. **King Abd-Alaziz University:**

$$\hat{y} = 6.1479 - 0.399x_1 + 0.313x_2$$

6.1479: is the averages number of mosquitoes, when (humidity and temperatures) zero or fixed, at King Abd-Al-aziz University area.

-.399: the increasing in temperatures one unit, leads to decrease in mosquitos breading with (39.9%), at King Abd-Al-aziz University area.

0.313: the increasing in humidity one unit, leads to increase in mosquitos breading with (31.3%), at King Abd-Al-aziz University area.

2. **University District:**

$$\hat{y} = 22.87 - 0.6698x_1 + 0.1415x_2$$

22.87: is the averages number of mosquitoes, when (humidity and temperatures) zero or fixed, at University district.

-.6698: the increasing in temperatures one unit, leads to decrease in mosquitos breading with (66.98%), at University district.

0.141: the increasing in humidity one unit, leads to increase in mosquitos breading with (14.15%), at University district.

3. Al-Thaghr:

$$\hat{y} = 11.891 - 0.693x_1 + .3863x_2$$

11.891: is the averages number of mosquitoes, when (humidity and temperatures) zero or fixed, at Al-Thaghr district.

-.693: the increasing in temperatures one unit, leads to decrease in mosquitos breading with (69.3%), at Al-Thaghr district.

0.3863: the increasing in humidity one unit, leads to increase in mosquitos breading with (38.63%), at Al-Thaghr district.

Table (3.12): shows these spatial models.

Points (districts)	**Spatial models**
King Abd-Alaziz University	$\hat{y} = 6.1479 - 0.399x_1 + 0.313x_2$
University district	$\hat{y} = 22.87 - 0.6698x_1 + 0.1415x_2$
Al-Thaghr	$\hat{y} = 11.891 - 0.693x_1 + .3863x_2$

Source: The researcher from applied study, ArcGIS Software 2018.

The table shows the relative importance for each model of three locations (Kin Abd-Alaziz University, University district and Al-Thaghr). It seems clearly that King Abd-Alaziz University area there is an equivalent (equal integer one) between the degree of temperature and humidity on their impact on mosquito reproduction. In University district residential, the temperature degree impact reached 5 times humidity degree and for Al-Thaghr district gets two times average. Therefore, we can conclude that from these models King Abd-Alaziz University area is the best among these areas because it doesn't face any environmental problems (sanitation, informal housing … etc.) that affect mosquito reproduction, then Al- Thaghr district area where the temperature and humidity degree are equivalent. This represents a weakness which means availability of such environmental problems, but the university district area got the third level which means there is more environmental problems that face this district, so we find that the relevant importance for independent variable reached five times.

3.8Discussion:

-The GWR produced to us the estimation of mosquito's numbers that carry Dengue Fever within University Municipal in Jeddah City, total number of 45 samples by numeral average of Geographical Variable.

-Determining Moduli to the spatial model which is estimated vary between (35%) to (57%) that means the interpreted power to the independence variables humidity and temperature degrees between weak to medium one.

- The value of total determined modulus (55%) that means the sample interpreted quantum of fifty-five percent of the variable occurred in the number of the mosquitoes carry Dengue Fever within University municipal.

-The results showed that the effect of temperature on mosquitoes' reproductions was negative, in other words, increase in temperature leads to decrease in the numbers of mosquitoes that carry Dengue Fever within University municipal.

-The result of the study showed that the humidity effect was positive whereas the increase in humidity leads to increase in the ratios of reproduction of the mosquitoes carry Dengue Fever within University municipal.

-Inference could be said that through the output maps with modulus of calibration that the ratios of the effect of stable variables (humidity and temperature) increase whenever we tend west to the municipal and decrease whenever we tend east.

-The results showed that there is no incorporeally effect to the variables (humidity and temperature ratios) in the phenomena of mosquitoes' breeding within the municipal.

-From the resulted map ratiocinate that by GWR that the spots need more fighting operation and that is for the reality of classification has to do with residuals; whenever these residuals are big this means that the spots need more studies and search to others environmental factors that assist in mosquitoes carry Dengue Fever reproduction, which are not included in the sample such as (numbers of swamps, water sanitation, ways of storing water, water produced from Air conditioners and etc. …which may return to the residents' customs and traditions within University municipals.

3.9Conclusions:

From the reality of the spatial analysis conducted for the mosquito's reproduction data in University municipality districts, it became clear that: the number of mosquitoes is denser in the districts (University and Al-Thaghr), the spatial mean is located inside University municipality. The mediator is located in the separating line between (University and Al-Thaghr) and the bigger part of the mediator location is located inside University district. The results also showed that the standard distance for mosquitoes' reproduction locations in University municipality is distributed on a non-verging range since the diameter of the standard distance circle is located inside the districts (University and Al-Thaghr). It became clear from the Directional Distribution results that the radii in the X and Y coordinates are very small and the rotation rate is moving towards North-West i.e. towards the districts (University and Al-Thaghr). The results of statistics of both the nearest neighbor and Moran I showed that the distribution pattern of mosquitoes' reproduction is convergent or clustered.

The results of spatial analysis of the phenomenon of mosquitoes' reproduction in University municipality concluded that mosquitos are more spreading in (University and Al-Thaghr) districts and that may be due to the high population density. The two districts are from the old districts, and suffer from problems in sewer system and generally lack of planning. Therefore, the Mosquitoes fighting Department must concentrate on (University and Al-Thaghr) districts when it is carrying out mosquitoes control operations. The study recommends the following: Fighting the factors, which lead to mosquito reproduction such as sewage systems maintenance. Fighting water leakage through water storage fighting. As it became clear that many of these areas population are storing water for use when the water supply is interrupted for several days which contributes to mosquito reproduction. Enact punitive laws and legislation against those who are proved that they help in water leakage.

-Depending Geographical Informational Systems in putting a policy to fight mosquitoes carry Dengue Fever within University municipal and the other municipals within Jeddah City in general.

-The importance of fighting mosquitoes outcropped according the plans the municipal set. Workshops and awareness colloquium should be launched to Jeddah residents to reduce mosquito's reproductions.

-The study discussed spots and areas of mosquitoes carry Dengue Fever reproduction within Jeddah Municipal whereas the study learns two factors could be interpreted which are humidity and temperature. The learners and researchers should study the other environmental factors of relation to mosquitoes that carry Dengue Fever reproduction in the entire municipal of Jeddah provinces and in particular University municipal.

References:

- Mitchell, A.. The ESRI Guide to GIS Analysis, Volume 2: Spatial measurements and Statistics. Redlands, CA: ESRI Press. 2005. 238 p.
-)ArcGIS9.3. (Oath, B. and Mitchell, A “ Getting Started with ArcGis”. ESRI. U.S.A. . 2001.P.341
- Jones B. Ch. “Geographical Information Systems and Computer Cartography”. Longman. Singapore. 1998.P.212.
- Dickinson, G.C. “Statistical Mapping”. Edward Arnold Ltd. London.1987.P.142
- Anselin, L. Local Indicators of Spatial Association – LISA. Geographical Analysis (1995).P.27
- Mitchell, A. 2005. The ESRI Guide to GIS Analysis, Volume 2: Spatial measurements and Statistics. Redlands, CA: ESRI Press. 238 p.
- Lee J, Wong WSD. Statistical analysis with ArcGIS.9.3. John Wiley & Sons, Inc., New York, USA; 2001. p. 192.
- Getis, A. and J. K. Ord Local Spatial Statistics: An Overview. Spatial Analysis: Modelling in a GIS Environment. P. Longley and M. Batty. Cambridge, Geoinformation International(1996).p.43
- Rogerson, P. A. The Application of New Spatial Statistical Methods to the Detection of Geographical Patterns of Crime. Applied GIS and Spatial Analysis. J. Stillwell and G. Clarke.Hoboken, New Jersey, United States, John Wiley and Sons. . (2004).p.56
- Anselin, L. (1995). Local Indicators of Spatial Association – LISA. eographical Analysis p.52
- ArcGIS 9.3 Desktop Help
- Gay Robinson, Methods&techniques in human geography,john wiley&Sons,Chichester,England,1998,pp.270-274.
- paulA.Longley,MichaelF.Goodchild,DavidJ.Maguire&Davidw.Rhind,GeographicInformationSystems&science,JohnWily&Sons,Chichester,England,pp.100-120.
- O’Dwyer, L. A., and D. L. Burton. 1998. Potential meets reality: GIS and public health research in Australia. Aust. N.Z. J. Public Health 22: 819-823.
- -Simpson, C. L., & Novak, L. L. (2013). Place matters: the problems and possibilities of spatial data in electronic health records. In AMIA Annual Symposium Proceedings (Vol. 2013, p. 1303). American Medical Informatics Association.

- -Webster, R., & Oliver, M. A. (2007). Geostatistics for environmental scientists. John Wiley & Sons (_.
- -Khormi, H. M., & Kumar, L. (2011). Examples of using spatial information technologies for mapping and modelling mosquito-borne diseases based on environmental, climatic and socio-economic factors and different spatial statistics, temporal risk indices and spatial analysis: A review. J Food Agr Environ, 9, 41-49 .
- -Goodchild, M. F. (1992). Geographical data modeling. Computers & Geosciences, 18(4), 401-408(_ .
- -Campbell, D. J., & Clarke, D. J. (1971). Nearest neighbour tests of significance for nonrandomness in the spatial distribution of singing crickets (Teleogryllus commodus (walker). Animal Behaviour, 19(4), 750-756.
- -Haining, R. P., & Haining, R. (2003). Spatial data analysis: theory and practice. Cambridge University Press.
- -Clark, P. J., & Evans, F. C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology, 35(4), 445-453.
- -Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). Quantitative geography: perspectives on spatial data analysis. Sage.
- -Yager, D. D., & Hopkins, C. D. (1993). Directional characteristics of tuberous electroreceptors in the weakly electric fish, Hypopomus (Gymnotiformes). Journal of Comparative Physiology A, 173(4), 401-414.
- -Prasomphan, S., & Mase, S. (2013). Generating prediction map for geostatistical data based on an adaptive neural network using only nearest neighbors. International Journal of Machine Learning and Computing, 3(1), 98.
- -Aiba, M., Katabuchi, M., Takafumi, H., Matsuzaki, S. I. S., Sasaki, T., & Hiura, T. (2013). Robustness of trait distribution metrics for community assembly studies under the uncertainties of assembly processes. Ecology, 94(12), 2873-2885.
- -Dixon, P. M. (2006). Nearest neighbor methods. Encyclopedia of environmetrics, 4..
- (Seber, G. A., & Wild, C. J. (2003). Nonlinear Regression. Hoboken. New Jersey: John Wiley & Sons, 62, 63.)
- 2.(Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (2001). Applied regression analysis: a research tool. Springer Science & Business Media.)

- (Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (Vol. 821). John Wiley & Sons.)
- Mauricio Bini, L., Diniz‐Filho, J. A. F., Rangel, T. F., Akre, T. S., Albaladejo, R. G., Albuquerque, F. S., ... & Isabel Bellocq, M. (2009). Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non‐spatial regression. Ecography, 32(2), 193-204.
- Tiefelsdorf, M. (2006). Modelling spatial processes: the identification and analysis of spatial relationships in regression residuals by means of Moran's I (Vol. 87). Springer.
- 6.Brunsdon, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighted regression. Journal of the Royal Statistical Society: Series D (The Statistician), 47(3), 431-443.)
- 7.Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons.
- 8.Bao, J., Shi, X., & Zhang, H. (2018). Spatial analysis of bikeshare ridership with smart card and POI data using geographically weighted regression method. IEEE Access, 6, 76049-76059.
- 9.Guo, L., Ma, Z., & Zhang, L. (2008). Comparison of bandwidth selection in application of geographically weighted regression: a case study. Canadian Journal of Forest Research, 38(9), 2526-2534.
- 10.Guo, L., Ma, Z., & Zhang, L. (2008). Comparison of bandwidth selection in application of geographically weighted regression: a case study. Canadian Journal of Forest Research, 38(9), 2526-2534.
- 11. Páez, A. (2005, May). Local analysis of spatial relationships: A comparison of GWR and the expansion method. In International Conference on Computational Science and Its Applications (pp. 162-172). Springer, Berlin, Heidelberg.)
- -Crime stat –A Spatial Statistics Program for Analysis of Crime Incident Locations – The National Institute of Justice Washington, DC, June 2013.
- -JOHN STILLWELL GRAHAM CLARKE , Applied GIS and Spatial Analysis-School of Geography, University of Leeds.
- 2-Illian J, Penttinen A, Stoyan H, Stoyan D Statistical analysis and modeling of spatial point patterns. Wiley, London. (2008).P.125

-)spacing)(3- Harvey, D.W. "Models of Evolution of spatial patterns in Human Geography" in Intergrated Models in Geography.Lit by Chorly and Hagget,London.1973.p.210
- Bartlett, M. S., 1975, The Statistical Analysis ofSpatial Pattern London: Chapman and Hall; New York: Wiley. Baxter, R. S., 1971, An urban atlas — Reading, prepared within the Urban Systems Study under the sponsorship of the Centre for Environmental Studies. Cambridge, University of Cambridge Department of Architecture. Chou,Y. H., 1991, Map resolution and spatial autocorrelation. Geographical Analysis, 21(3), 228—246.
- Couclelis, H., 1991, Requirements for planning.rclcvant 015: a spatial perspective, Papers in Regional Science, 70(1), 9—19. Diggle, P., 1983, Statistical Analysis of Spatial Point Patterns (Mathematics in biol 08)7), London: Academic Press. Fotheringham, A. 5., 1983, A new set of spatial interaction models: the theory of competing destinations, Environment and Planning A, 15, 15—36. Gaile, Gary, L and Con, I. Willmott (Eds.), 1984, Spatial Statistics and Models, Dordrecht; Boston: 1). Reidel Pub. Co.; Hingham, MA: distributed in the USA by Kluwer Boston Acndemic Publishers. Gatrell, Anthony C., 1983, Distance and Space a Geographical Perspective, Oxford: Carendon Press; New York: Oxford University Prez. Gefis, A. and B. Boots, 1978, ModeLs of Spatial Processes: an Approach to the Study of Point, Line, and Area Patterns. Cambridge (Eng.); New York: Cambridge University Press. Getis, A., 1991, Spatial interaction and spatial uutc,correlation: a cross-product approach, Environment and Planning A, 23, 1269—1277. Glick, B., 1979, Tests br space-time clustering used in cancer research, Geographical Analysis, 11(2), 202—207. Goodchild, M. F., 1987, A spatial analytical perspecLive on geographical information systems, tnt I. Geographical Infarniadan Syscwtc, 1(4), 327—334. Goodchild, M. F. and Janelle, D., 1984, ‘ftc city around the clock: space-time patterns of urban ecological structure, Environment and Planning A, 16, 807—820. Raining, Robert P., 1978, Specification and estimation problems in models of spatial dependence. Evanston, Ill.: Dept. of Geography, Northwestern University. Rodgson, M. 3., 1973, Towards more realistic allocation in location-ailootion models: an interaction approach, Environment and Planning A, 10, 1273—1285. maclIe, 0., M. F., Goodchild and Klinkcnberg. IL, 1988, Space.time diaries and travel characteristics for different levels of respondent aggregation, Environment and Planning A, 20, 891—906. Kent, 5., 1984, Analyzing activity areas: an

ethnoarchaeologcal study of the use of space, Albuquerque: University of New Mexico Press. Kiauber, M. R. and Mustacchi, P., 1970, Space-time clustering of childhood Leukemia in San Francisco. Cancer Research, 30, W69—1973. Knox, E. 0., 1964, The detection of space-time interactions. Applied Statistics, 13, 25-30. Macmillan, Bill, 1989, Remodelling Geography, Blackwell: Oxford, Miner, H., 1991, Modeling accessibility using space-time prism concepts within geo graphical information systems, mt. .i. of Geographical Informahon Systems, 5(3), 287—301. Newsome, T. H., 1992, Aleacuring Spatial Pattern in Ccnvac Units: Residential Segre gation in Franklin County, Ohio, Unpublished Ph.l). Dissertation, Department of Geography, The Ohio State University, Columbus, Ohio. Nyerges, T., 1991, Geographic inionnation abstractions: conceptual clarity for geo. graphic modeling, Environmunt and Planning it, 23, 1483—1499. O'Kelly, M. E., 1987, Spatial interaction based location.at!ocation models, Spatial Analysis and Location-Allocation Models, A. C bosh and 0. Rushton (Eds.), New York: van Nostrand Reinhold. O'Kelly, M. E, 1992, A clustering approach to the planar hub location problem, Annals of Operations Research, 40, 339—53, Openshaw, S., 1987, An automated geographical analysis system. Environment and Planning A, 19, 431. Openshaw, S., Charlton, M, Wymer, .C and Craft, A-, 1987, A Mark I Geographical Analysis Machine for the automated analysis of point data sets, Intl. Ceograph ical Information Systems, 1(4), 335—358.

Appendix:

ArcGIS Output:

OBJECT ID *	Shap e *	Observed MEAN	Condition Number	Local R2	Predic ted	Coefficient Intercept	Coefficient #1 RT	Coefficient #2 RP	Resid ual
1	Point	12.42	23.20	0.46	13.61	20.55	-0.66	0.16	-1.18
2	Point	8.19	19.18	0.41	8.20	11.38	-0.50	0.24	-0.01
3	Point	12.48	23.62	0.45	13.63	20.32	-0.65	0.16	-1.15
4	Point	10.58	22.94	0.45	12.70	16.75	-0.59	0.20	-2.13
5	Point	10.64	22.97	0.46	11.78	18.98	-0.64	0.18	-1.14
6	Point	18.51	23.41	0.47	11.56	24.55	-0.73	0.13	6.94
7	Point	8.68	23.27	0.46	12.16	21.30	-0.67	0.16	-3.49
8	Point	7.92	29.99	0.35	10.36	24.14	-0.53	0.00	-2.43
9	Point	12.05	24.45	0.46	11.55	23.00	-0.68	0.13	0.50
10	Point	9.96	22.93	0.45	10.82	16.47	-0.59	0.20	-0.86
11	Point	7.51	22.62	0.44	11.69	13.42	-0.53	0.23	-4.18
12	Point	10.34	23.17	0.46	9.05	22.52	-0.69	0.15	1.29
13	Point	7.43	23.40	0.45	7.70	20.46	-0.65	0.16	-0.27
14	Point	9.29	24.02	0.45	10.21	21.58	-0.66	0.15	-0.92

15	Point	8.13	22.92	0.47	9.23	19.57	-0.65	0.18	-1.10
16	Point	7.95	23.48	0.44	7.84	19.39	-0.63	0.17	0.11
17	Point	6.25	23.14	0.45	7.70	19.35	-0.63	0.17	-1.45
18	Point	7.24	23.74	0.45	7.41	21.75	-0.67	0.15	-0.17
19	Point	9.47	22.89	0.46	7.20	15.24	-0.57	0.22	2.27
20	Point	11.36	22.89	0.47	7.49	19.72	-0.66	0.19	3.87
21	Point	6.95	22.87	0.48	7.50	20.53	-0.68	0.18	-0.54
22	Point	4.61	23.96	0.45	6.98	21.22	-0.66	0.15	-2.37
23	Point	5.83	23.82	0.46	6.20	23.34	-0.70	0.13	-0.37
24	Point	5.83	23.82	0.46	4.78	23.34	-0.70	0.13	1.05
25	Point	4.55	23.84	0.45	4.90	21.03	-0.66	0.15	-0.36
26	Point	7.47	22.94	0.45	6.16	16.65	-0.59	0.20	1.31
27	Point	4.63	24.07	0.44	7.24	19.44	-0.63	0.16	-2.62
28	Point	2.95	23.09	0.43	5.01	16.93	-0.59	0.19	-2.05
29	Point	5.83	23.02	0.46	6.43	20.88	-0.67	0.17	-0.59
30	Point	4.01	22.99	0.46	3.54	18.18	-0.62	0.19	0.47
31	Point	4.83	24.47	0.46	6.31	23.43	-0.69	0.12	-1.48
32	Point	3.94	22.89	0.47	5.30	19.75	-0.66	0.19	-1.36
33	Point	3.99	22.87	0.44	5.85	15.86	-0.57	0.20	-1.86
34	Point	4.64	24.91	0.46	4.94	23.73	-0.69	0.12	-0.30
35	Point	6.44	23.83	0.47	7.46	24.00	-0.71	0.13	-1.01
36	Point	12.05	22.94	0.43	7.64	16.39	-0.58	0.19	4.41
37	Point	4.10	23.29	0.52	6.64	29.93	-0.89	0.12	-2.54
38	Point	4.69	22.91	0.48	5.52	18.17	-0.64	0.21	-0.83
39	Point	5.00	23.62	0.45	8.79	20.36	-0.65	0.16	-3.79
40	Point	7.57	23.66	0.46	8.81	22.70	-0.69	0.14	-1.23
41	Point	10.64	22.90	0.48	11.31	18.92	-0.65	0.19	-0.67
42	Point	10.74	23.39	0.48	8.95	26.40	-0.77	0.12	1.78
43	Point	17.92	23.01	0.45	10.31	17.21	-0.60	0.19	7.61
44	Point	17.14	23.64	0.45	9.54	21.24	-0.66	0.15	7.60
45	Point	21.08	23.56	0.57	15.42	15.33	-0.67	0.29	5.66

Table $(t_{\hat{B}})$

Residual	Standard Error	Standard Error Intercept	Standard Error Coefficient #1 RT	Standard Error Coefficient #2 RP	Std. Residual	Source ID	t (b0)	t(b1)	t(b2)
-1.18	2.74	5.10	0.13	0.06	-0.43	0.00	7.51	-0.13	1.23
-0.01	1.53	7.25	0.16	0.08	-0.01	1.00	7.44	-0.07	1.45
-1.15	2.72	5.04	0.13	0.06	-0.42	2.00	7.48	-0.13	1.21
-2.13	2.74	5.15	0.13	0.06	-0.78	3.00	6.12	-0.11	1.49
-1.14	2.62	5.20	0.13	0.06	-0.44	4.00	7.25	-0.12	1.36
6.94	2.70	5.29	0.14	0.06	2.58	5.00	9.11	-0.14	0.95
-3.49	2.84	5.12	0.13	0.06	-1.23	6.00	7.50	-0.13	1.18
-2.43	1.43	7.27	0.22	0.09	-1.69	7.00	16.82	-0.07	-0.01
0.50	2.82	5.28	0.14	0.06	0.18	8.00	8.17	-0.13	0.94
-0.86	2.91	5.14	0.13	0.06	-0.30	9.00	5.66	-0.11	1.50
-4.18	2.77	5.18	0.14	0.06	-1.51	10.00	4.84	-0.10	1.63
1.29	2.95	5.20	0.13	0.06	0.44	11.00	7.63	-0.13	1.11
-0.27	2.94	5.05	0.13	0.06	-0.09	12.00	6.95	-0.13	1.22
-0.92	2.89	5.13	0.13	0.06	-0.32	13.00	7.48	-0.13	1.09
-1.10	2.86	5.29	0.14	0.06	-0.38	14.00	6.83	-0.12	1.34
0.11	2.94	5.01	0.13	0.06	0.04	15.00	6.61	-0.13	1.28
-1.45	2.94	5.08	0.13	0.06	-0.49	16.00	6.59	-0.12	1.32
-0.17	2.92	5.11	0.13	0.06	-0.06	17.00	7.45	-0.13	1.11

2.27	**2.90**	**5.24**	**0.14**	**0.06**	**0.78**	**18.00**	**5.26**	**-0.11**	**1.58**
3.87	**2.91**	**5.35**	**0.14**	**0.06**	**1.33**	**19.00**	**6.77**	**-0.12**	**1.35**
-0.54	**2.92**	**5.40**	**0.14**	**0.06**	**-0.19**	**20.00**	**7.04**	**-0.13**	**1.31**
-2.37	**2.92**	**5.11**	**0.13**	**0.06**	**-0.81**	**21.00**	**7.27**	**-0.13**	**1.12**
-0.37	**2.90**	**5.21**	**0.13**	**0.06**	**-0.13**	**22.00**	**8.04**	**-0.13**	**0.99**
1.05	**2.83**	**5.21**	**0.13**	**0.06**	**0.37**	**23.00**	**8.25**	**-0.13**	**0.99**
-0.36	**2.85**	**5.08**	**0.13**	**0.06**	**-0.12**	**24.00**	**7.37**	**-0.13**	**1.15**
1.31	**2.87**	**5.16**	**0.13**	**0.06**	**0.46**	**25.00**	**5.80**	**-0.11**	**1.49**
-2.62	**2.85**	**5.21**	**0.13**	**0.06**	**-0.92**	**26.00**	**6.82**	**-0.12**	**1.19**
-2.05	**2.82**	**5.01**	**0.13**	**0.06**	**-0.73**	**27.00**	**6.00**	**-0.12**	**1.44**
-0.59	**2.71**	**5.20**	**0.13**	**0.06**	**-0.22**	**28.00**	**7.71**	**-0.13**	**1.24**
0.47	**2.74**	**5.15**	**0.13**	**0.06**	**0.17**	**29.00**	**6.63**	**-0.12**	**1.40**
-1.48	**2.86**	**5.31**	**0.14**	**0.06**	**-0.52**	**30.00**	**8.18**	**-0.13**	**0.91**
-1.36	**2.80**	**5.34**	**0.14**	**0.06**	**-0.49**	**31.00**	**7.05**	**-0.12**	**1.35**
-1.86	**2.85**	**5.02**	**0.13**	**0.06**	**-0.65**	**32.00**	**5.56**	**-0.11**	**1.52**
-0.30	**2.79**	**5.45**	**0.14**	**0.07**	**-0.11**	**33.00**	**8.52**	**-0.13**	**0.82**
-1.01	**2.85**	**5.26**	**0.14**	**0.06**	**-0.35**	**34.00**	**8.42**	**-0.14**	**0.94**
4.41	**2.72**	**5.07**	**0.13**	**0.06**	**1.62**	**35.00**	**6.03**	**-0.11**	**1.46**
-2.54	**2.36**	**6.42**	**0.17**	**0.07**	**-1.07**	**36.00**	**12.67**	**-0.14**	**0.75**
-0.83	**2.83**	**5.44**	**0.14**	**0.06**	**-0.29**	**37.00**	**6.42**	**-0.12**	**1.47**
-3.79	**2.74**	**5.04**	**0.13**	**0.06**	**-1.39**	**38.00**	**7.44**	**-0.13**	**1.21**
-1.23	**2.81**	**5.16**	**0.13**	**0.06**	**-0.44**	**39.00**	**8.09**	**-0.13**	**1.05**
-0.67	**2.55**	**5.36**	**0.14**	**0.06**	**-0.26**	**40.00**	**7.42**	**-0.12**	**1.40**
1.78	**2.78**	**5.47**	**0.14**	**0.06**	**0.64**	**41.00**	**9.51**	**-0.14**	**0.83**
7.61	**2.83**	**5.07**	**0.13**	**0.06**	**2.69**	**42.00**	**6.08**	**-0.12**	**1.45**
7.60	**2.94**	**5.08**	**0.13**	**0.06**	**2.59**	**43.00**	**7.23**	**-0.13**	**1.15**
5.66	**2.21**	**6.09**	**0.16**	**0.07**	**2.56**	**44.00**	**6.94**	**-0.11**	**1.80**

N	E	Mosquitos	Temperature	Humidity rate
21.48257	39.2452	9.196429	23.8	53.1
21.48236	39.24633	22.26563	24.7	38.1
21.4824	39.24525	19	24.9	59
21.48232	39.24526	29.90323	24.2	51.4
21.48268	39.24468	25.59016	23.3	41.5
21.48207	39.24417	23.85455	25.5	43.2
21.48226	39.24362	61.17647	27.7	60.1
21.48203	39.24519	42.5082	25.9	60.8
21.48258	39.24483	34.34546	27.2	55.6
21.48221	39.24537	56.47059	28.5	54.8
21.48264	39.24549	47	27.2	56
21.48302	39.24489	25.98361	29.8	49
21.48239	39.24518	41.24615	31	46.1
21.48192	39.24473	12.40351	30	59

21.4828	39.24473	13.63333	29.7	43
21.48197	39.24376	23.66102	30.7	46.4
21.48185	39.2479	15.0714	30.7	45
21.48199	39.24573	153.1587	31.2	44.8
21.47975	39.24554	19.76271	31.3	45.1
20	39.24474	14.73684	31	44.5
21.47861	39.24721	8.813559	31.5	46.2
21.47908	39.24551	17.36842	32	46
21.47916	39.24571	10.39286	33.3	45.9
21.47984	39.24577	10.88136	35.2	45.2
21.48036	39.2441	9.8	35.3	46.5
21.48247	39.24567	12.69492	34.9	50.3
21.48266	39.24327	12.4386	34	56.5
21.48053	39.24313	8.533333	35	45.7
21.47998	39.24527	8.870968	30.1	34.2
21.48008	39.24528	10.20313	36.1	40.8
21.4788	39.24519	4.947368	34.7	54.9
21.47941	39.24488	3.421053	35.8	49.8
21.48183	39.24261	3.056604	35.2	49.8
21.47856	39.24623	2.305556	35.4	47.9
21.47964	39.24687	9.649123	33.9	59
21.47984	39.24541	7.803571	35.3	60.7
21.47979	39.24556	6.607143	33.9	65.6
21.47996	39.24453	6.339286	33.2	72.1
21.48001	39.24562	6.363636	33.9	65.1
21.47822	39.24438	6.054545	33	63.1
21.48105	39.24239	6.54717	32.3	68.7
21.47939	39.24298	8.1	32	60.6
21.4807	39.23979	10.6	31.5	61.8
21.47988	39.23854	12.91228	30.6	56.4
21.47982	39.24277	14.5	27.5	64.4

Sup set of data:

YEAR	WEEKS	N	E	Mosquitos
2018	1	21.50449	39.26216	1076
2018	1	21.51001	39.26818	75
2018	1	21.49525	39.29548	122
2018	1	21.56608	39.29741	37
2018	1	21.48301	39.31401	178
2018	1	21.43672	39.34993	20
2018	1	21.42722	39.39494	24
2018	2	21.50449	39.26216	127
2018	2	21.51001	39.26818	83
2018	2	21.49525	39.29548	108
2018	2	21.56608	39.29741	942
2018	2	21.48301	39.31401	236
2018	2	21.43672	39.34993	4
2018	6	21.50449	39.26216	5057
2018	6	21.48553	39.29601	53
2018	6	21.49525	39.29548	1344
2018	6	21.56608	39.29741	1164
2018	6	21.48301	39.31401	1859
2018	6	21.45078	39.31884	10
2018	6	21.45208	39.32672	27
2018	16	21.501	39.25935	29
2018	16	21.4849	39.30319	0
2018	16	21.54739	39.29212	12
2018	16	21.56608	39.29741	0
2018	16	21.48301	39.31401	31
2018	16	21.45358	39.3194	12
2018	21	21.501	39.25935	49
2018	21	21.4849	39.30319	19
2018	21	21.54739	39.29212	43
2018	21	21.56608	39.29741	30
2018	21	21.48301	39.31401	4
2018	21	21.45078	39.31884	30
2018	21	21.45208	39.32672	34