**ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY**

# UTILISING WEB-BASED GIS APPLICATION FOR SPATIAL CENTROGRAPHIC ANALYSIS OF EARTHQUAKE DATA

**M.Sc. THESIS**

**Boran ŞİMŞEK**

**Institute of Informatics**

**Geographical Information Technologies Programme**

**JUNE 2018**

**<u>ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY</u>**

# UTILISING WEB-BASED GIS APPLICATION FOR SPATIAL CENTROGRAPHIC ANALYSIS OF EARTHQUAKE DATA

**M.Sc. THESIS**

**Boran ŞİMŞEK**
**(706131004)**

**Institute of Informatics**

**Geographical Information Technologies Programme**

**Thesis Advisor: Asst. Prof. Dr. Caner GÜNEY**

**JUNE 2018**

**ISTANBUL TEKNİK ÜNİVERSİTESİ ★ BİLİŞİM ENSTİTÜSÜ**

# WEB CBS UYGULAMASIYLA DEPREM VERİLERİ ÜZERİNDE CENTROGRAPHY ANALİZİ

**YÜKSEK LİSANS TEZİ**

**Boran ŞİMŞEK**
**(706131004)**

**Bilişim Enstitüsü**

**Coğrafi Bilgi Teknolojileri Programı**

**Tez Danışmanı: Dr. Öğretim Üyesi Caner GÜNEY**

**HAZİRAN 2018**

Boran Şimşek, a M.Sc. student of ITU Graduate School of Science Engineering and Technology student ID 706131004, successfully defended the thesis entitled "UTILISING WEB-BASED GIS APPLICATION FOR SPATIAL CENTROGRAPHIC ANALYSIS OF EARTHQUAKE DATA", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

**Thesis Advisor :** **Asst. Prof. Dr. Caner GÜNEY** ..............................
Istanbul Technical University

**Jury Members :** **Asst. Prof. Dr. Caner GÜNEY** .............................
Istanbul Technical University

**Prof. Dr. Necla ULUĞTEKİN** ..............................
Istanbul Technical University

**Assoc. Prof. Dr. Aslı DOĞRU** ..............................
Bogazici University

**Date of Submission : 04 May 2018**
**Date of Defense : 07 June 2018**

***To my spouse,***

## FOREWORD

I would first like to thank my supervisor Asst. Prof. Dr. Caner GÜNEY for his never-ending supports, guidance, patience and understanding from the beginning of the 5-year long journey. I wish to extend my thanks to Prof. Dr. Dursun Zafer ŞEKER, Prof. Dr. Necla ULUĞTEKİN, Prof. Dr. Rahmi Nurhan ÇELİK ve Asst. Prof. Dr. Ahmet Özgür DOĞRU for their precious contributions during my journey in Master's programme. I would like to express my thanks to all scientists from the beginning of the evolution of the human for their very valuable lives that have been dedicated to understand the nature and the universe and to fight against the dark ages.

I would also like to thank to my parents, İbrahim Şimşek and Aliye Şimşek, and to all of my friends, especially Umut Tolğa ÖZCİVAN, Bayram ÖZDOĞAN, Onat GÜNGÖR, Gökçen KILIÇ, Muzaffer KARADEMİR, Enver ÇİÇEK, Seçkin ELÇİN and Tirebolu KIRKİKİ, who have never hesitated sharing their experiences, suggestions and contributions.

Finally, I must express my very profound gratitude to my spouse, Sinem Şimşek, for her continuous encouragement through the period of researching and writing this thesis, everlasting patience and precious scientific studies that keep inspiring me to continue studying.

The Web GIS application developed in this study has been published on a public domain so that anyone, who is interested in such analysis, can do the analysis on the earthquake data, which can be filtered with the provided parameters. The Web application can be reached at http://centro.boransimsek.com.

June 2018 Boran ŞİMŞEK

**TABLE OF CONTENTS**

**Page**

## ABBREVIATIONS

| | |
|---|---|
| **1NF** | : First Normalisation Form |
| **2NF** | : Second Normalisation Form |
| **3NF** | : Third Normalisation Form |
| **API** | : Application Program Interface |
| **BL** | : Business Layer |
| **CSS** | : Cascading Style Shhets |
| **DAL** | : Data Access Layer |
| **ED** | : Euclidean Distance |
| **EPSG** | : European Petroleum Survey Group |
| **GIS** | : Geographic Information Systems |
| **GPS** | : Global Positioning System |
| **GUI** | : Graphical User Interface |
| **HTML** | : Hypertext Markup Language |
| **HTTP** | : Hypertext Transfer Protocol |
| **IT** | : Information Technology |
| **JSON** | : JavaScript Object Notation |
| **MD** | : Mahalanobis Distance |
| **OSM** | : Open Street Map |
| **PL** | : Presentation Layer |
| **RDBMS** | : Relational Database Management System |
| **URL** | : Uniform Resource Locator |
| **USGS** | : United States Geological Survey |
| **UTM** | : Universal Transverse Mercator |
| **WGS** | : World Geodetic System |
| **WWW** | : World Wide Web |
| **XML** | : Extensible Markup Language |

## LIST OF TABLES

**LIST OF FIGURES**

**Page**

## UTILISING WEB-BASED GIS APPLICATION FOR SPATIAL CENTROGRAPHIC ANALYSIS OF EARTHQUAKE DATA

### SUMMARY

Geographical Information Systems (GIS) have become an essential tool for analyses and calculations because they handle all the events that can be displayed on a map in a digital environment. In this study, the earthquake data obtained by the sensors of the US Geological Survey (USGS) located in many regions of the world have been transferred to the geographical information system and analysed for the detection of the earthquakes' intensified regions via centrography method. The earthquake data of the USGS, which holds the information on earthquakes from the year of 1900 to the present day in a database and shared via web services, has been queried with different parameters and the obtained data has been analysed by using centrography in the geographical information system.

In this thesis, a web GIS application was needed in the first place, and modelling was done as to transform the structure of the earthquake data. Later on, the design of the web services to be queried for the earthquake data was investigated, and the code blocks were prepared to provide the necessary communication between the web GIS application and the web services and transform the model to apply the earthquake data. The earthquakes had been classified according to their magnitude while they were shown on the map and they were reflected on a thematic map. Thematic maps are the map types that are generated by classifying the phenomena on the map according to its characteristics and the representation of the phenomena can be based on either a colour scale or a size scale. On this count, the user, who checks the map, can easily distinguish the elements with different values because of the colours and the sizes of the markers. In this study, the earthquakes have been represented with different colours according to their magnitude, and it has been ensured that the users of the application could easily see where the earthquakes are more severe and weaker. Although the human beings may have a general idea of the density points on the map, it has no scientific or statistical significance. The aim of this study is to find the regions where earthquakes are clustered by statistical analysis instead of giving the initiative to the human to define the density regions. There are various methods in geographical information systems for this purpose. Some of them are heat map, centrography and clustering methods. In this study, an iterative algorithm has been developed using the centrography methods to identify the concentration zones of the earthquakes.

Centrography is basically a statistical method that aims to find out where a phenomenon is clustered, investigate whether there is any displacement or trend in the density over time and support the decision maker to choose the right decision for the future plans for the region based on the results. The centrography algorithm developed in this study tried to find out the areas, in which the earthquakes occurred frequently. However, since it is not enough to develop only the algorithm, the results obtained in the analyses have also been investigated. The results obtained for this purpose have also been compared with the results of the heat map method, which is another GIS

method. The results of the centrography method were expected to show the same zones that are displayed with high temperature value on the heat map. The results show that the centroids generated by the algorithm developed in this study covers the high temperature fields on the heat map.

**WEB CBS UYGULAMASIYLA DEPREM VERİLERİ ÜZERİNDE CENTROGRAPHY ANALİZİ**

# ÖZET

İnsanlık tarihiyle birlikte bilim, içinde yaşanılan doğayı ve evreni daha doğru anlama amacı gütmüştür. Bilimin birçok alanda, eldeki veriler ışığında yanlış çıkarımlar yaptığı su götürmez bir gerçek olsa da zaman içerisinde bilim, kendi yanlışlarını da düzeltmiştir ve her geçen gün doğru bilgiye daha çok yakınsamaktadır. Örneğin, önceleri durağanlığın bir gerçeklik olduğu ve bir cismi hareket ettirmek için kuvvetin gerektiğine inanılsa da ileriki bilimsel çalışmalar göstermiştir ki aslolan harekettir ve kuvvet, bir cismi durdurmak için gereklidir. Dünya da evrendeki tüm diğer maddeler gibi hareket ve kuvvet temeliyle oluşmuştur ve hareketini durmaksızın sürdürmüştür. Kıtaların her geçen gün yer değiştirmesi, rüzgarlar, dağların oluşumu ve dalgalar gibi günlük hayatta bulabileceğimiz birçok etkinlik, dünyanın hala aktif olduğuna birer işarettir. Bazıları doğal afet olarak adlandırılan bu hareketlilikler, yine dünyanın oluşumdan itibaren, çoğu insanlık tarafından yıkıcı olarak görülse de içinde yaşadığımız gezegenin bugünkü halini oluşturmuştur ve değiştirmeye de devam etmektedir. Önceleri insanlar, tıpkı diğer canlılar gibi ve evrimin de açıkladığı şekilde doğaya adapte olarak yaşamlarını sürdürmüşlerdir. Daha sonraları yerleşik hayata geçişle birlikte hızlı bir kentleşme başlamıştır. Kentleşme ile birlikte yer hareketlerinin insanlar üzerindeki özellikle yıkıcı etkisi önemli derecede dikkat çekmiştir. Yapılan tüm çalışmalar göstermiştir ki doğal afetleri önlemek mümkün değildir. Çünkü bu doğa olaylarının tamamı dünyanın dinamiğiyle alakalı ve gerçekleşmek durumundadır. Bunun öğrenilmesiyle birlikte insanlık, doğal afetlerin yıkıcı etkilerini en aza indirmek için çalışmalar yapmıştır. Gelişen teknolojiyle birlikte birçok doğal afetin gerçekleşmeden belli bir süre öncesinde fark edilip edilemeyeceği üzerine araştırmalar yapılmıştır. Örneğin, dünyanın herhangi bir yerinde tsunami oluşma ihtimali olduğunda, tsunamiden etkilenecek bölgelerdeki insanlar henüz tsunami ulaşmadan bilgilendirilmekte ve bu da can kayıplarının önüne geçmektedir. Ancak bu tür bir erken uyarı sisteminin geliştirilemediği ya da uyarı ile gerçekleşme süresi arasında yeterli sürenin bulunmadığı doğal afetler için istatiksel analizler büyük önem arz etmektedir. Bu sayede doğal afetlerin nerelerde daha çok gerçekleştiği, nerelerde daha çok yıkıcı etkiye sahip olduğu öngörülebilir duruma gelmektedir.

Coğrafi Bilgi Sistemleri, harita üzerinde gösterilebilecek tüm olguları dijital ortamda ele almasından ötürü analiz ve hesaplamalar için çok önemli araçlar haline gelmiştir. Bu çalışmada da dünyanın birçok bölgesine yerleştirilen sensörler sayesinde elde edilen deprem verilerinin web tabanlı Coğrafi Bilgi Sistemi ortamına aktarılarak yoğunlaştığı bölgelerinin tespiti için analiz yapılmıştır. 1900 yılından günümüze kadar her büyüklükten depremin bilgilerini bir veritabanında tutan ve web servisler aracılığıyla paylaşıma açan Amerikan Jeolojik Araştırma Enstitüsü'nden (USGS) deprem verileri farklı parametrelerle sorgulanarak verilerin web tabanlı harita üzerinde görselleştirilmesi yapılmıştır. Bu amaçla, öncelikle bir web tabanlı CBS uygulaması

ihtiyacı doğmuştur ve deprem verilerinin yapısına uygun olacak şekilde bir mekansal modelleme yapılmıştır. Daha sonra deprem verilerinin sorgulanacağı web servislerin tasarımı incelenerek, gerekli iletişimi sağlayacak, deprem verilerini uygulamadaki modele dönüştürecek kod blokları hazırlanmıştır. Depremler harita üzerinde gösterilirken büyüklüklerine göre sınıflandırılarak tematik bir harita üzerinden kullanıcılarla paylaşılmıştır. Bu uygulamaya http://centro.boransimsek.com adresinden ulaşılabilmektedir. Tematik haritalar, haritada incelenen olguların, kendine özgü özelliklerine göre sınıflandırılıp, kimi zaman renk skalası, kimi zaman da farklı büyüklükteki işaretlerle temsil edilmesiyle elde edilirler. Bu sayede haritayı inceleyen kişiler farklı değerlere sahip elemanları göz yardımıyla kolaylıkla ayırt edebilirler. Bu çalışmada da depremler büyüklüklerine göre farklı renklerle temsil edilmişlerdir ve uygulamayı kullanan kişilerin depremlerin nerelerde daha şiddetli ve nerelerde daha zayıf etkili olduklarını kolayca görmeleri sağlanmıştır. Her ne kadar insan genel bir bakışla haritadaki yoğunluk noktaları hakkında fikir sahibi olabilse de bunun hiçbir bilimsel ve istatistiksel zemini yoktur. Bu çalışmada da amaçlanan, yoğunluk bölgelerinin tespitini insan gözünün inisiyatifine bırakmak yerine, istatistiksel analizlerle depremlerin kümelendiği bölgeleri bulmaktır. Bu amaç için Coğrafi Bilgi Sistemlerinde çeşitli yöntemler vardır. Bunlardan bazıları ısı haritası, centrography ve kümeleme yöntemleridir. Bu çalışmada depremlerin yoğunlaştığı bölgelerin tayini için centrography yöntemleri kullanılarak iteratif bir algoritma geliştirilmiştir.

Centrography, temelde bir olgunun nerelerde kümelendiğini bulmayı amaçlayan ve zaman içerisinde bu yoğunlukta herhangi bir yer değiştirme ya da yönelim olup olmadığını araştıran, sonuçlarıyla da bölge hakkında ileriye dönük yapılacak planlar için doğru karar vermeye yardımcı olan bir istatiksel analiz yöntemidir. Bu çalışmada geliştirilen centrography algoritmasıyla depremlerin hangi bölgelerde sıkça gerçekleştiği bulunmaya çalışılmıştır. Ancak tek başına bu algoritmayı geliştirmek yeterli olmadığından, analizlerde elde edilen sonuçların da doğruluğu araştırılmıştır. Bu amaç doğrultusunda elde edilen sonuçlar, bir başka coğrafi bilgi sistemi yöntemi olan ısı haritası yöntemiyle de karşılaştırılmıştır. Isı haritaları çoğunlukla maviden kırmızıya doğru ilerleyen bir renk skalasıyla, incelenen olguların en seyrek olduğu yerleri mavi, en yoğun olduğu yerleri kırmızıyla gösteren haritalardır. Bu analizde de aynı deprem verileri ısı haritaları üzerine yansıtılarak yoğunlaştığı bölgeler kırmızı renklerle gösterilmiştir. Daha sonra geliştirilen centrography algoritmasıyla üretilen bölgelerin, ısı haritasında kırmızıyla gösterilen bölgelerle üst üste gelip gelmediği karşılaştırılmıştır. Analiz sonucunda görülmüştür ki iki farklı yöntem de aynı bölgeleri yoğunluk bölgeleri olarak belirlemiştir.

Bu çalışmada, depremlerin kümelendiği bölgelerin tayini dışında bir başka fikir verici sonuçla da karşılaşılmıştır. Büyüklüğü 7 ve üzeri olan depremler uzun bir zaman aralığında incelendiğinde görülmüştür ki, yıkıcı etkisi çok yüksek olan depremler çoğunluka dünyanın oluşumundan itibaren kıtaların birbirininden ayrılmasına sebep olan ana fay hatlarının çevresinde toplanmıştır. Bu depremler belli bir noktanın etrafında kümelenmedikleri için bu çalışmada geliştirilen yöntemlerle bulunamamışlardır. Ancak bu depremlerin konumları ile eğri uydurma yöntemi kullanıldığında görülmüştür ki, üretilen eğri, deprem verisi arttıkça ana fay hatlarına yakınsamaktadır.

Yapılan analizler sonucunda görülmüştür ki, San Francisco ve Alaska eyaletlerinde her gün yaklaşık 300 deprem gerçekleşmekte ve herhangi bir zaman kesiti alındığında, gerçekleşen tüm depremlerin yaklaşık yüzde yetmişinin bu bölgelerde bulunduğu görülmektedir. Ancak bu bölgelerde oluşan çoğu depremin büyüklüğü çok küçük olduğundan henüz yıkıcı etkiye sahip değildir. Bunun yanı sıra, yalnızca büyüklüğü yüksek depremler incelendiğinde görülmektedir ki Kuzey ve Güney Amerika kıtasının batı kıyısı boyunca ana fay hattı çevresinde zaman içerisinde çokça yıkıcı etkiye sahip depremler konumlanmıştır.

Bu çalışmanın amacı doğrultusunda, elde edilen sonuçlar da gösteriyor ki San Francisco ve Alaska bölgelerinde depremlerin yoğunlaşması, bölgelerin ana fay hattının yakınında olması ve bölgenin ileride oluşabilecek şiddetli depremlere gebe olması sebebiyle, bu bölgelerde inşa edilmiş ve edilecek tüm binaların bu riskler göz önünde bulundurularak gerekli çalışmaların yapılması ve karşılaşılabilecek hasarların en aza indirilmesi gerekmektedir. Ek olarak, bölgenin okyanus kıyısında olması sebebiyle olası şiddetli bir depremin, arkasından bir tsunamiyi de tetikleyebileceği göz önünde bulundurulmalıdır.

# 1. INTRODUCTION

## 1.1 Motivation

Since the earthquake is a highly destructive natural disaster, it is crucial to determine the most frequent and effective areas in order to be aware of the risks of the location. Earthquakes' nature is different than the other natural disasters such as hurricanes and volcanoes because the earthquakes are mostly non-forecastable disasters. It is possible to take precautions and actions before the forecastable natural disaster happens or reaches the place. However, it is difficult to take final actions to prepare for an earthquake before it happens. Particularly, there are several types of study focus on earthquakes. One type of the studies is the study that the geoscientists focus on the structure of the Earth and the interactions between the internal layers of the Earth in order to understand the earthquakes' reasons and the movements of the fault lines that cause earthquakes. As a result of these studies, the geoscientists are able to generate a map that shows the plates on the Earth, the fault lines that tend to split the plates with earthquakes and the movements of the plates and the fault lines in time series. Even though the studies of the geoscientists give the full picture of the lines that the earthquakes recur themselves along, the information that tells when and where the earthquake happens is still unknown. Another type of the studies on earthquakes mostly focus on the statistical analysis that aims to find the zones that the earthquakes happen frequently. In this thesis, the study focuses on statistical analysis on the earthquakes recorded since the beginning of the twentieth century in order to allow people to have more accurate location based decisions in related fields such as urbanisation and industry zone planning.

## 1.2 Purpose of the Thesis

This thesis aims to explain;

- How the centrography method, which is a GIS technique, can be applied on the earthquake data, which contains the earthquakes recorded since the year of

1900 and is provided by the United States Geological Survey (USGS) through the web services,

- How the spatial data can be adapted into a web GIS application,
- How the cluster zones of the earthquakes can be located on the web GIS application,

Although the centrography method has been only applied to the earthquake data in this thesis, it can be performed as it is used in this study, on any phenomenon that can be represented spatially such as natural disasters, human migration and the spread of the contagious diseases.

It is evident that the analyses done on a web GIS application are much more effective as the calculations are done by computers, visual as the information can be displayed on a digital map and useful for the decision-making stage.

## 2. GEOGRAPHICAL INFORMATION SYSTEMS (GIS)

There are two different approaches to define GIS, first of which is the technical approach. Geographical Information Science has been defined as a field of information science specialising on the fundamental issues arising from creation, handling, storage and use of geographical information (Longley et al., 2001). In other words, GIS is a technological approach to answer the question of "What is where and when?". The second approach is the theoretical approach, which defines GIS as a supportive solution for decision making according to the analysis done on the spatial data in order to take better action based on the analysis results.

### 2.1 Components of GIS

GIS consists of five key elements, which are hardware, software, data, people, and methods (Figure 2.1).


**Figure 2.1:** The components of GIS.

#### I) Hardware

Hardware is the computer that the GIS application runs on. With the development of technology, GIS software can be operated on various hardware such as centralised servers, stand-alone desktop computers and mobile devices.

**II) Software**

At this moment software refers to the GIS application's itself and its primary task is to provide the functions to store, manipulate, analyse and visualise the spatial data. A GIS application uses a database, which is designed with the rules of Relational Database Management System (RDBMS), in order to store the data. A GIS project needs a code development, which contains the implementation of the procedures to process and analyse the data.

**III) Data**

Since all the procedures in a GIS project are entirely based on the data, which will hereafter refer to spatial data, it is the most critical component amongst the others in GIS. The spatial data is represented in two different models as raster data and vector data. Raster data model is the transformation of an area into a grid. It consists of equal sized square cells such as pixels. Each cell stores a single value, which represents the characteristics of the spatial phenomenon based on the research field such as magnitude, elevation and spectral value (Figure 2.2).

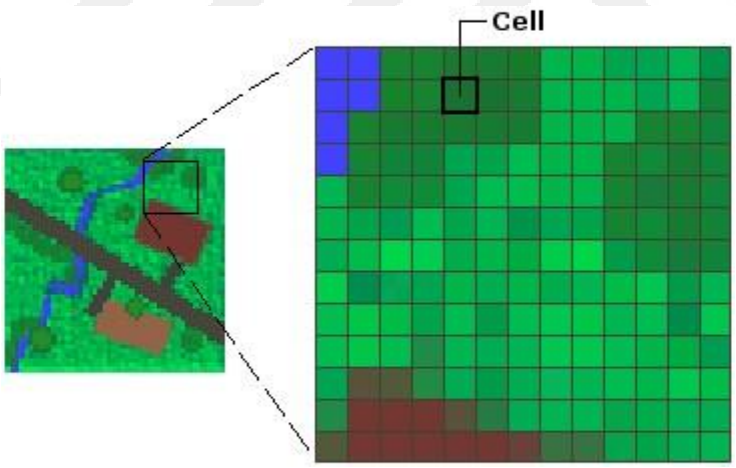


**Figure 2.2:** An example of a raster data model.

Since the raster model averages the values within a given pixel to acquire a single value, the more area included in per pixel means, the less precision for the related data values (Campbell and Shin, 2012, Figure 2.3). For example, if the cell's size is larger than the width of a road, the road may not exist within the raster dataset.

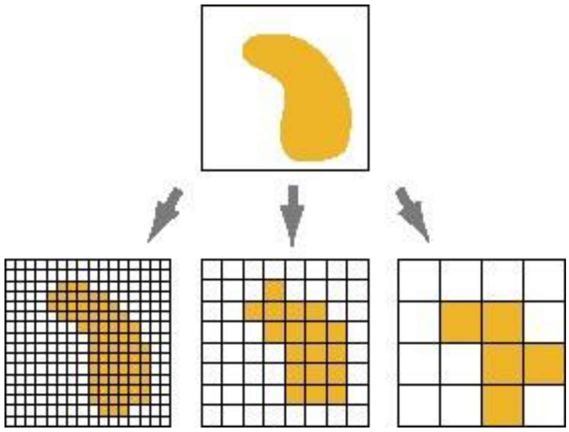

**Figure 2.3:** The loss of the precision according to the cell size.

There are several advantages of the raster data model such as having a simple structure, being a robust format for advanced spatial and statistical analysis, providing the ability to perform fast overlays with complex datasets and uniformly store points, lines and polygons. The raster data model fall under four main categories:

- Base maps
- Surface maps
- Thematic maps
- Attributes of a feature

On the other hand, the vector data model is the representation of an area by transforming the elements of the area into points, lines and polygons. For example, a digitised map can be considered as a vector data because, in the end, the data will contain the actual map points, lines and polygons as coordinates.

Since raster and vector data models both have advantages and disadvantages each other, it is crucial to define the data model will be used in a project based on the needs of the project. There is no restriction to define only one data model for a project. Raster data can be obtained easily by scanners, remote sensors and satellite images and gives quick information in pixels about the area. However, it needs higher resolution images or digitisation of the images when it comes to analysing discrete features such as polygons. In contrast, since the vector data has more precise values such as coordinates and elevation about the feature and can be easily transferred into a database structure, it is much more useful to make an analysis on discrete objects and easier to develop a software programme to analyse. However, it is a time-consuming process to gather a vector data.

**IV) People**

People is one of the leading factors of GIS because the people are the component that builds, manages and develops the GIS projects. GIS projects are a set of systems that need to run synchronised work of fieldworkers, database developers and administrators, data entry operators, data processor experts and end users. Therefore, a GIS project can succeed if the people, who work on any part of the system, are trained well. Also, if the GIS is institutionalised, the projects can be appropriately managed.

**V) Methods**

The components are required to be managed together in a synchronised manner in GIS. The methods in all phases of a project, which are the pre-project, project development and post-project, will determine the success or failure of the project in each of the data, resource, human, finance, risk, quality, standards and time components of the projects are managed in-house and together. Ensuring the sustainability and up-to-date of GIS can be achieved through methods to be determined and implemented.

## 2.2 Functions in GIS

A spatial analysis consists of several steps as a chain, where each step is done one by one till the end. These steps are defined as functions in GIS and each function defines the way how the data will be handled.

### 2.2.1 Data acquisition

Data acquisition is the most critical and time-consuming part of a GIS project because the entire project should be established on a well-organised dataset if the project is desired to be serviceable and adequate. There are several options for input data to be used in a GIS project such as digitised maps, images obtained from satellites, Global Positioning System (GPS) data and tabular data.

The tabular data is the most convenient input data amongst the other input data types for a GIS project because the tabular data format can be quickly transformed and integrated into a database structure. Although the data obtained from GPS is also another useful input and has accurate information, it can only be effective in small areas. The technique to obtain images from an aircraft or a satellite is called remote sensing in GIS, and in this technique, the devices use laser and radar altimeter in order to produce some aerial photographs to create topographic maps. In contrast to the GPS, vast areas can be reached by remote sensing technique. Another option to have an input data for a GIS project is the digitisation of maps. The process of digitising the paper-based maps is the most time-consuming process, and it needs significant human power to work on.

### 2.2.2 Data processing

Data processing is the step that the necessary parts of the dataset are extracted. When the necessary data is obtained, the data is transferred to the data storage component of the GIS project, which is mostly a database.

### 2.2.3 Data storage and management

The data acquired with one of the methods above must be either stored or managed to provide GIS project to process it. The most convenient way to store the data is using a database attached to the GIS project. In order to store the data efficiently, it is highly recommended to build a relational database model because of its very well organised structure and flexibility. In some projects, instead of storing the data, the data is directly retrieved from a data provider through services to be managed. In this thesis, the data is obtained from the web services of USGS to process the data and reflect the results in the web application.

### 2.2.4 Data querying and analysis

Data querying is the part that the necessary information is built by correlating the data stored in the database with the filters based on the analysis. The generated information is analysed on the GIS application, and the results are interpreted.

### 2.2.5 Data presentation

Data presentation is the stage, where the final information and the analysis results are displayed on the GIS application. In this stage, the analysis results are shown to the GIS users so that the users can deduce by the analysis.

## 2.3 Web GIS

Web GIS is a type of distributed information system, comprising at least a server and a client, where the server is a GIS server and the client is a web browser, desktop application or mobile application (ESRI, ArcGIS). In other words, web GIS can be defined as any GIS that uses web technology to communicate between a server and a client. There are some elements essential to web GIS. First of all, the GIS server should have a URL so that clients can access through the web. The client sends the request to the web GIS server, and the server performs the requested GIS operations and sends the response back to the client. All the communications rely on HTTP specifications. The response format sent back to the client can be in several formats such as XML, JSON, GeoJSON and HTML.

There are several advantages of web GIS regarding capability, accessibility and cost. First of all, since a web GIS application can be accessed via web browsers, it provides a cross-platform solution to the users. In other words, the compatibility is not an issue for a user to think about. It is entirely independent on the platform that is developed because when it is accessed via web browsers, the application is interpreted in a unified model thanks to the standards of web technologies. In contrast, the desktop GIS applications are fully dependent on the platform that they are developed, and it is much more costly to install a workstation to run the application. Secondly, a web GIS application can be reached by a large number of users, and this also reduces the cost and effort as averaged by the number of users. Finally, since the web GIS application can serve to multiple users at the same time, it provides better organised technological resource consumption while multi-tasking.

Nowadays, with the developments in web technologies, cloud computing also has taken its place in web GIS field. Cloud computing provides more powerful and 24/7 online servers than the stand-alone servers, where the large-scale computing requests can be handled easily. “For instance, a cloud-based GIS solution currently helps a medical supply company to provide the vital medical aids to the patients with the tighter time frame” as Victoria Kouyoumjian (2011), who is an IT strategies architect at ESRI, stated.

To develop a web GIS project, there are several steps to be followed, first of which is the modelling the database structure. When the database is designed, the application type must be decided carefully. For example, if the intent is to provide well-organised data to the clients, then developing a web service could be a better solution instead of developing an application with a user interface. If the intent is to allow users to make an analysis on a map, then developing a web application could be a useful approach.

### 2.3.1 Database modelling and RDBMS

Since it stores and provides the data, which is the vital component of a GIS project, the database represents the essential layer in a GIS project. In addition, it is often difficult to change the structure and the design of a database because the changes that are done on the database may cause a chain of modification in the application developed based on the database. Therefore, it is better to spare enough time to establish a well-organised database structure depends on the needs of the project. For this purpose, the relational database model could be an efficient solution with a better understanding of its normalisation methods such as first normalisation (1NF), second normalisation (2NF) and third normalisation (3NF) in order to reduce the duplications, avoid the anomalies and organise a relation based structure.

#### 2.3.1.1 First normal form (1NF)

The first normalisation method aims to organise all the tables, each column of which only keeps indivisible (atomic) value, in the database. Also, each value stored in a column should be of the same data type. In this method, each column is examined and if there is any column contains more than one value (Table 2.1), the data types of the values are analysed. If the values refer to a different feature, then unique columns are added for each value. If the data types of the values are same, then the row is duplicated for each value on the corresponding column until there is no column contains multiple values (Table 2.2).

**Table 2.1:** Unnormalised table.

| EarthquakeID | Location | Country | Magnitude | Category |
|---|---|---|---|---|
| 1 | 41.273, 41.411 | Georgia, Turkey | 4.8 | Cat. 2 |
| 2 | 39.167, 38.061 | Turkey | 3.6 | Cat. 1 |
| 3 | 13.084, 102.446 | Thailand, Cambodia | 4.2 | Cat. 2 |

**Table 2.2:** First normal form

| EarthquakeID | Latitude | Longitude | Country | Magnitude | Category |
|---|---|---|---|---|---|
| 1 | 41.273 | 41.411 | Georgia | 4.8 | Cat. 2 |
| 1 | 41.273 | 41.411 | Turkey | 4.8 | Cat. 2 |
| 2 | 39.167 | 38.061 | Turkey | 3.6 | Cat. 1 |
| 3 | 13.084 | 102.446 | Thailand | 4.2 | Cat. 2 |
| 3 | 13.084 | 102.446 | Cambodia | 4.2 | Cat. 2 |

#### 2.3.1.2 Second normal form (2NF)

Second normalisation method aims to eliminate the partial dependencies in the tables, and it can only be applied to the database, which is in first normal form. Partial dependency is defined as the functional dependency of a non-prime attribute, which is not a part of the primary key but is functionally dependent of an attribute that is fully dependent of the primary key in a table. To eliminate the partial dependency, the columns have partial dependency must be extracted to another table and the columns must be replaced with the ID values of the new table so that the relation stays in place. The “Country” column in Table 2.3 has the partial dependency with the “Latitude”, “Longitude” and “Magnitude” columns, where they form the primary key together in the table. Extracting the values in the “Country” column into the new table, the partial dependency can be eliminated (Table 2.4).

**Table 2.3:** Second normal form.

| EarthquakeID | Latitude | Longitude | CountryID | Magnitude | Category |
|---|---|---|---|---|---|
| 1 | 41.273 | 41.411 | 1 | 4.8 | Cat. 2 |
| 1 | 41.273 | 41.411 | 2 | 4.8 | Cat. 2 |
| 2 | 39.167 | 38.061 | 2 | 3.6 | Cat. 1 |
| 3 | 13.084 | 102.446 | 3 | 4.2 | Cat. 2 |
| 3 | 13.084 | 102.446 | 4 | 4.2 | Cat. 2 |

**Table 2.4:** Extraction of the partial dependent values.

| CountryID | CountryName |
|---|---|
| 1 | Georgia |
| 2 | Turkey |
| 3 | Thailand |
| 4 | Cambodia |

#### 2.3.1.3 Third normal form (3NF)

Thirds normalisation method aims to eliminate the transitive dependency in the tables. A transitive dependency is defined as the functional dependency between three or more columns, where one column relies on the second column through the third column called intermediate column. In Table 2.3, the “Category” column is fully dependent on the “Magnitude” column in order to define the earthquake’s category. Also, the “Magnitude” column is dependent on the “EarthquakeID” column because it indicates the basic information of an earthquake. In order to eliminate the dependencies, it is required to create a new table for the transitive column, keep all the fully dependent columns together with their primary keys in the individual tables and create a relation table between the two tables (Table 2.5, 2.6, 2.7, 2.8).

**Table 2.5:** Earthquake-Category relation table.

| EarthquakeCategoryID | EarthquakeID | CategoryID |
|---|---|---|
| 1 | 1 | 2 |
| 2 | 2 | 1 |
| 3 | 3 | 2 |

**Table 2.6:** Category table.

| CategoryID | Category |
|---|---|
| 1 | Cat. 1 |
| 2 | Cat. 2 |
| 3 | Cat. 3 |
| 4 | Cat. 4 |
| 5 | Cat. 5 |

**Table 2.7:** Earthquake table.

| EarthquakeID | Latitude | Longitude | Magnitude |
|---|---|---|---|
| 1 | 41.273 | 41.411 | 4.8 |
| 2 | 39.167 | 38.061 | 3.6 |
| 3 | 13.084 | 102.446 | 4.2 |

**Table 2.8:** Earthquake-Country relation table.

| EarthquakeCountryID | EarthquakeID | CountryID |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 1 | 2 |
| 3 | 2 | 2 |
| 4 | 3 | 3 |
| 5 | 3 | 4 |

### 2.3.2 Web services

Web service is a software system designed to support interoperable machine-to-machine interaction over a network (W3C Glossary, 2004). It enables machines to communicate with each other through the World Wide Web (WWW) protocols such as Hypertext Transfer Protocol (HTTP). More specifically, a web service has a request-response model (Figure 2.4), and it responds the requests with machine-readable texts in several formats such as Extensible Markup Language (XML) and JavaScript Object Notation (JSON).

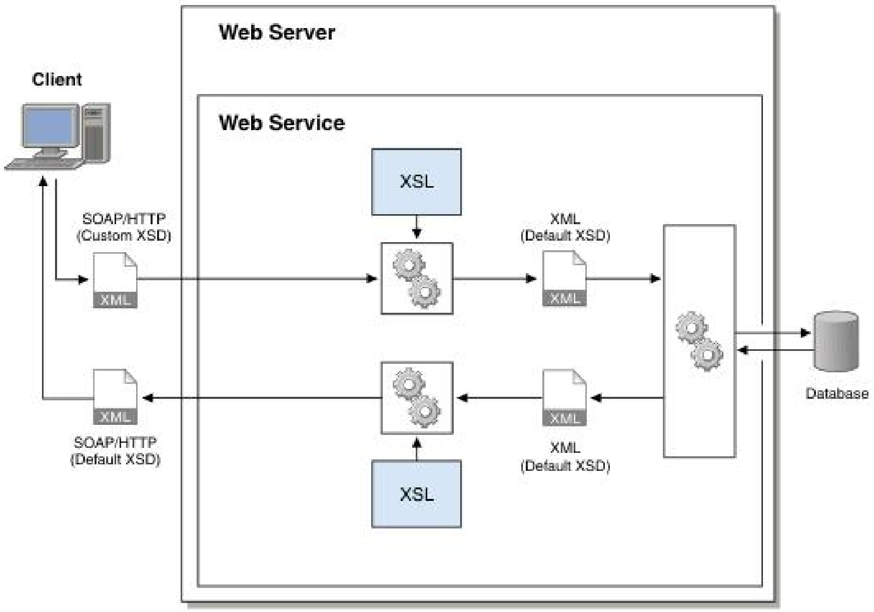


**Figure 2.4:** The diagram of a request-response structure in web services.

XML is the most common format used by the web services to receive the requests and send the response back. Since it is the text-based format and hierarchically organised, a human can easily interpret it. The other advantage of XML is that it provides position independence in the structure. On the other hand, JSON provides simpler structure than XML format and consumes less space, which allows faster data interchange between the systems.

For a while ago, another data format, which is called GeoJSON, has been started to be used for the GIS purposes. It is a format that keeps the same rules of JSON structure but more specialised for GIS data. In this thesis, GeoJSON format has been used for the data while requesting from the web services of USGS. The web application developed for the thesis has been adapted to extract the valuable information from the GeoJSON response.

### 2.3.3 Code development

Code development part of a GIS project is the part that the data stored in the database is processed and displayed on the screen with Graphical User Interface (GUI). Before starting the code development, the environment must be chosen based on the needs. There are several types of application such as console application, desktop application and web application. If the project is defined as a data provider, then it is also an option to develop a web service, from which the 3rd party client applications can request the data. In this thesis, a web application is decided to be developed because it is intended to request the data with some parameters from the web services of USGS and display the results of the analysis on a map. A web application consists of two different parts, which are back-end and front-end.

#### 2.3.3.1 Back-end development

Back-end development is the part of coding that is entirely hidden to the end users, and mostly this part is responsible for interacting the database, processing the data and providing the result data packages to the front-end side. For the back-end development, it is essential to decide which programming language to be used in the project. Although there is a wide range of programming languages, it is highly recommended to decide an object-oriented programming language. In this project, Microsoft's C# programming language has been used for the back-end development. As a framework, the version of 4.6.1 has been selected, and the web application has been developed on Microsoft's Visual Studio 2015 environment. The Microsoft's C# programming language is a fully object-oriented programming language, and it is fully compatible with the Microsoft's SQL Server 2014, which is the database server. To build the communication between the web application and the database, Entity Framework v6 has been used.

While developing the back-end side of a project, it is critical to define the responsibilities. In order to organise the responsibilities, a technical concept is recommended, which is called n-tier development. The n-tier development consists of three layers to separate the responsibilities: Data Access Layer (DAL), Business Layer (BL), Presentation Layer (PL). Data Access Layer is the part of the code, where the database communication standards of the project are configured. Also, this layer is the only layer, which communicates with the database. Business Layer is the layer that responsible for the communication between the Presentation Layer and the Data Access Layer. Most of the logical processes are handled in this layer before requesting the corresponding data from the DAL. The security configurations such as user authentications, timeout decision for the system can be implemented in this layer so that it reduces the workload on the DAL and the bandwidth consumption on the database server. The Presentation Layer is the layer that is responsible for the interpretation of the user's actions, transmitting the user's requests to BL and informs the user about the results.

#### 2.3.3.2 Front-end development

Front-end development is the part that the developer mostly focuses on the design and the implementation of the user interface. In a web application project, the user interface is the part that the user can see through the web browser such as Internet Explorer, Mozilla Firefox and Google Chrome. The responsibilities on the front-end development are also split into three concepts, which are called design or layout, visual properties and functionalities.

First, Hypertext Markup Language (HTML) is the basic web programming language that is used to prepare the layout of the web pages. Second, Cascading Style Sheets (CSS) is the specific style sheet language used to define the visual features and the abilities of the HTML elements. Finally, the JavaScript language is responsible for the actions done by the user on the web page and the communication between the front-end and the back-end. jQuery is the most popular JavaScript library that is developed to ease the front-end development.

In this thesis, the front-end side of the web application has been developed using the Microsoft's web programming language, ASP.NET. In addition to the ASP.NET's built-in libraries, the JavaScript/jQuery and CSS frameworks of Bootstrap have been used for the front-end development. Bootstrap is an open source toolkit widely used by web developers, especially the front-end developers, that contains very well organised JavaScript functions and responsive stylesheets. Also, the JavaScript library of MapBox has been integrated into the project as it provides the special GIS functions. The libraries of MapBox have been developed to be operational on the open source map, whose provider is Open Street Map (OSM).

# 3. CENTROGRAPHY IN GIS

Statistical analysis forms a substantial part of the spatial analysis techniques in Geographical Information Systems, where the spatio-temporal datasets are processed in order to have characteristic results about the inspected phenomenon such as tendency, frequency, dispersion and concentration areas. It is a fact that each statistical analysis has its useful methods that generates better results. In this study, the centrography method has been applied on earthquake datasets to analyse the dispersion and to find the concentration areas.

## 3.1 Clustering

Everitt et. al defines the clustering as “a technique that is concerned with exploring datasets to assess whether or not they can be summarised meaningfully in terms of a relatively small number of groups or clusters of objects or individuals which resemble each other and which are different in some respects from individuals in other clusters” (p. 13). Although there are several clustering methods that are used for different cases, the k-means clustering algorithm is defined as a standard clustering method. The standard algorithm is also called as Lloyd’s algorithm since the algorithm was first suggested by Stuart Lloyd in 1957. The algorithm consists of several steps to find the cluster centres. First of all, the number of clusters is chosen and the mean centre points are calculated for each cluster. Then, each point in the dataset is assigned to the closest cluster centre. When each point is associated to the cluster centres, then the new cluster centres are defined. Until the predefined threshold criterias are met, the points are assigned to the centres and new centres are defined.

The interface of a Web GIS application to display the clustering results can be defined by the developers as they want. Mostly, the cluster points displayed on the map show the number of points that are associated to them and the clustering algorithm is executed when the zoom level is changed (Figure 3.1).

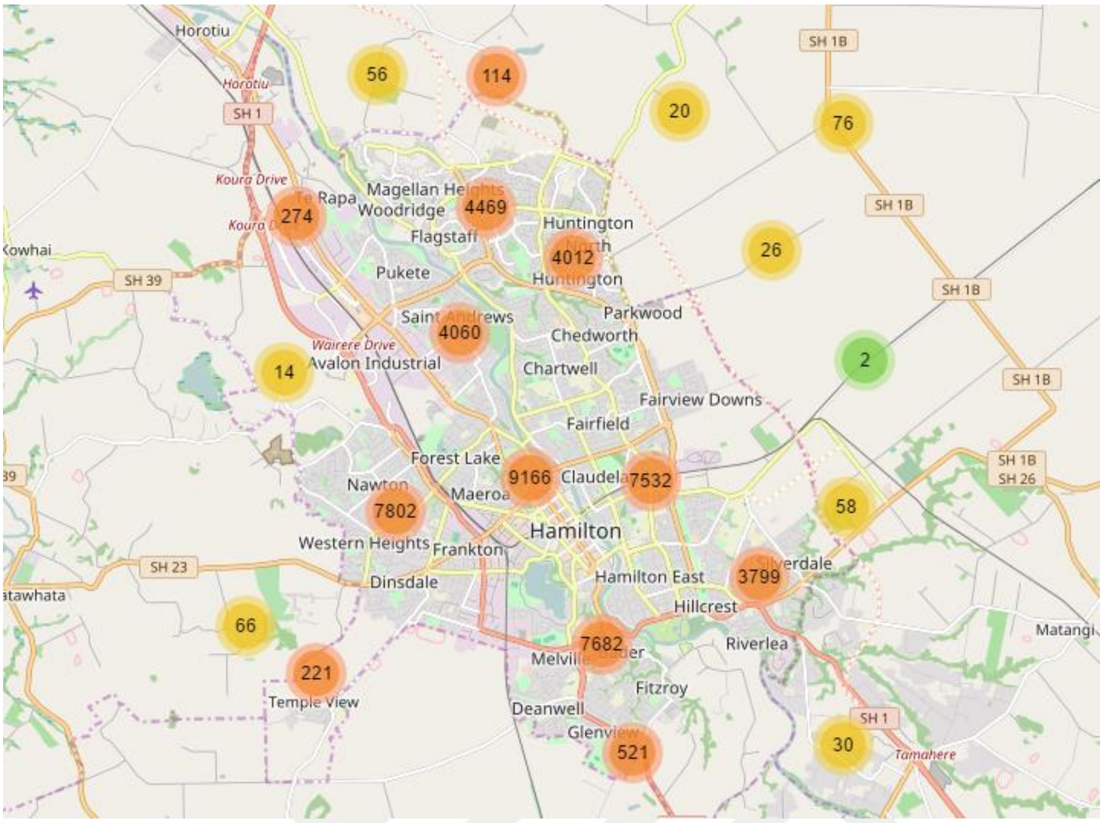


**Figure 3.1:** The representation of the cluster points on the map.

### 3.2 Centrography

Centrography is one of the Geographical Information Systems (GIS) techniques, which allows one to assess and measure the average location, dispersion, movements, and directional change of a phenomenon through time (LeBeau, 1987). As Kellerman stated that "these techniques include measures of central tendency equivalent to the mode, mean and median, and measures of dispersion such as the variance, standard deviation and analysis of variance transformed to two and three dimensions" (p. 3). The terms "centrography" and "centrogram", which are suggested by the Mendeleev Centrographical Laboratory, compactly express the group of geographical studies in this field of two-dimensional statistical analysis (Sviatlovsky and Eells, 1937). In the beginning, these analyses were bounded by the studies about human populations. For instance, the studies in 1937, which were about the movement of the centre of population in Europe and Northern America, showed that there was a contrary on the direction of movements. The movement in Europe had a tendency toward the east while the movement in Canada and the United States had toward the west. Also, the information was obtained that the movement in the northern hemisphere had been away from the Atlantic Ocean and toward the Pacific Ocean. In the light of this information, the studies suggested that the Pacific Ocean may become the concentration of the population. The result of the studies can be seen in Figure 3.2.

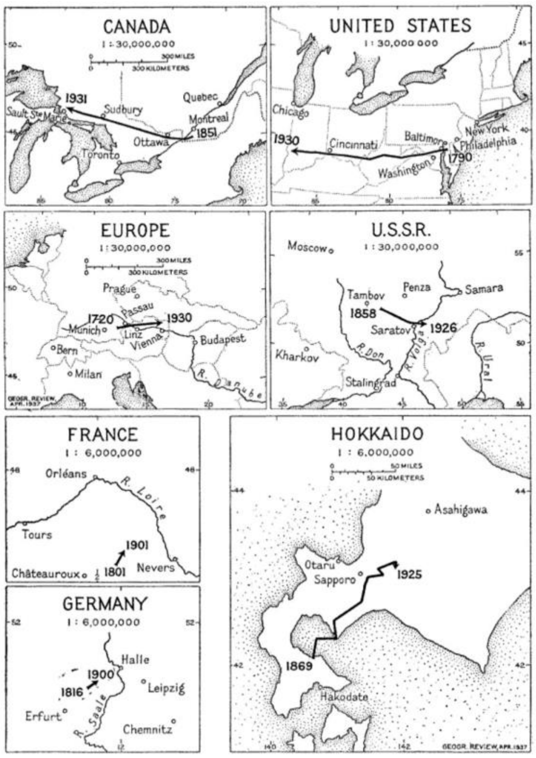


**Figure 3.2:** The trends of the centre of the population (Sviatlovsky and Eells, 1937).

Subsequent studies, which this method has been applied in, have been addressed different issues such as crops, natural disasters, criminology and a variety of regional distributions of other social and economic factors. In 1987, this method had been used by LeBeau, who was a professor at the Department of Criminology and Criminal Justice in Southern Illinois University, to analyse the 5-year dataset of lone-assailant rapes classified by type of offender (LeBeau, 1987). Mamuse et al. (2009) made a spatial centrographic analysis of mineral deposit clusters, using the komatiite-hosted Kambalda nickel sulphide deposit cluster (Figure 3.3). This study shows that the spatio-geometric partitions are likely locales for spatial analysis of nickel orebodies and endowment. The centrographic approach is potentially useful in mineral resource estimations and mineral exploration targeting and can be considered as a sample study in recent years.

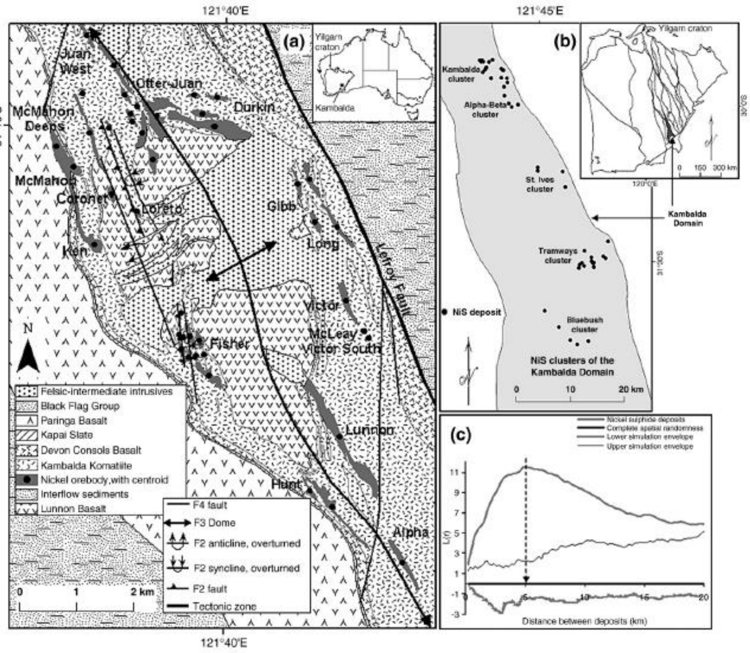


**Figure 3.3:** The komatiite clustering (Mamuse et al., 2009).

This technique aims to support the decision-making process based on the phenomenon's actual clustered points or trend through time. Centrography can also be applied to a non-geographic phenomenon, but if the focused data is spatial data, then it becomes a GIS technique. First of all, since the centrography focuses on the centralised area, the clustered points must be calculated based on the coordinates of the phenomenon. There are several ways to find the average locations such as distance-based analysis, which only concentrates on the locations, and density-based analysis that includes the effects of the typical information in addition to the location. On a 2-D flat plane, it is easy to find the average location by applying the standard mathematical operations. When it comes to an analysis of the locations of a phenomenon on the earth, which is projected onto a 2-D plane, then some effects of the projection applied must be taken in to account. In order to ease the analysis, the earth is transformed into a 2-D plane with several map projection methods.

### 3.3 Map Projection

Map projection is a mathematical transformation that aims to reshape the 3-dimensional earth on a 2-dimensional plane. It is evident that there will be some distortions and loss of the data during the process of the transformation. There are several projection types such as cylindrical, conical and azimuthal projection, each of which uses different reference lines to transform the earth on a surface. Hence, it is crucial to choose the correct projection type to be used in the analysis because each projection type has its advantages and disadvantages.

### 3.3.1 Conical map projection

Conical map projection is the projection method, where the inner surface of a cone is used as a 2-dimensional surface to reflect the sphere onto. In this method, a cone is placed on the top or the bottom of the globe, and the sides of the cone touching the sphere are well defined in order to have an accurate scale for the points along the tangent line (Figure 3.4). While the cylindrical projection methods are utilised to have a better result for the whole sphere, the conical projection methods are more useful to get a better result for a hemisphere.

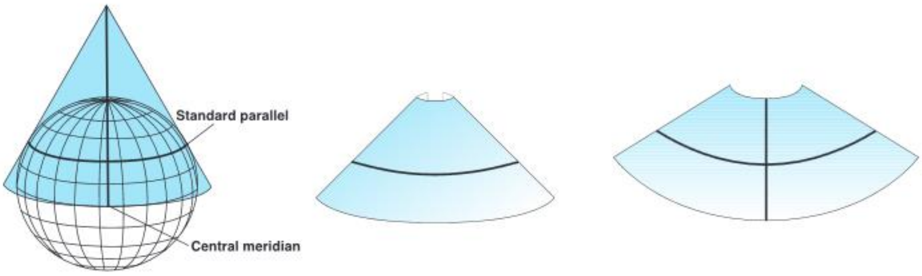


**Figure 3.4:** Representation of the conical map projection (Kennedy, 2000).

### 3.3.2 Azimuthal map projection

Azimuthal map projection, which is also known as planar projection, is another way of transforming a globe onto a flat plane according to an aspect. In this projection method, the contact point of the tangent line, which can be at either pole, the equator or any intermediate point on the globe, specifies the aspect for the projection (Figure 3.5). The simplest forms of the azimuthal projections are the polar aspects, in which all meridians are shown as straight lines radiating at their true angles from the centre, while parallels of latitude are circles, concentric about the pole (Snyder, 1982). One of the most known examples of the polar azimuthal equidistant projection can be the emblem of the United Nations.

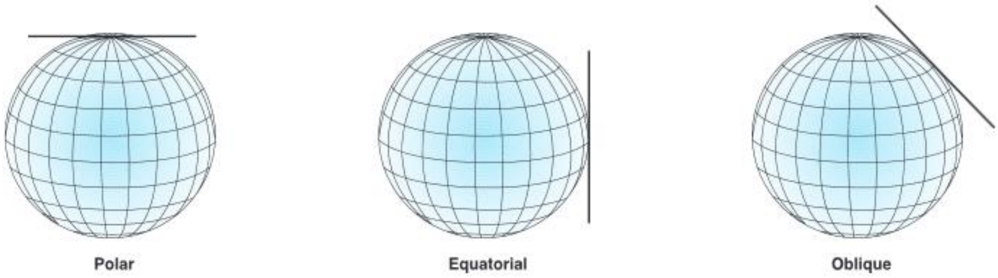


**Figure 3.5:** The aspects in azimuthal map projection (Kennedy, 2000).

### 3.3.3 Cylindrical map projection

A cylindrical projection is the transformation of the globe onto a cylindrical surface, in which the meridians are equally spaced while the spacing between parallel lines of latitude increases toward the poles. In this projection type, the sphere to be projected is encapsulated with a cylinder, and the sphere is expanded along the inner surface of the cylinder. Such projections are considered as conformal projections, which means that the shapes and the angles describe the spatial relationships are preserved. There are some cylindrical projection types, which are used according to the complexity of the situation, such as normal, transverse and oblique (Figure 3.6). In normal projection method, the equator is used as the tangent line. In contrast, a meridian is chosen for the sphere to be projected along with transverse projection method. In oblique projection method, the cylinder is rotated around a great circle line, which is different than the equator and a meridian.

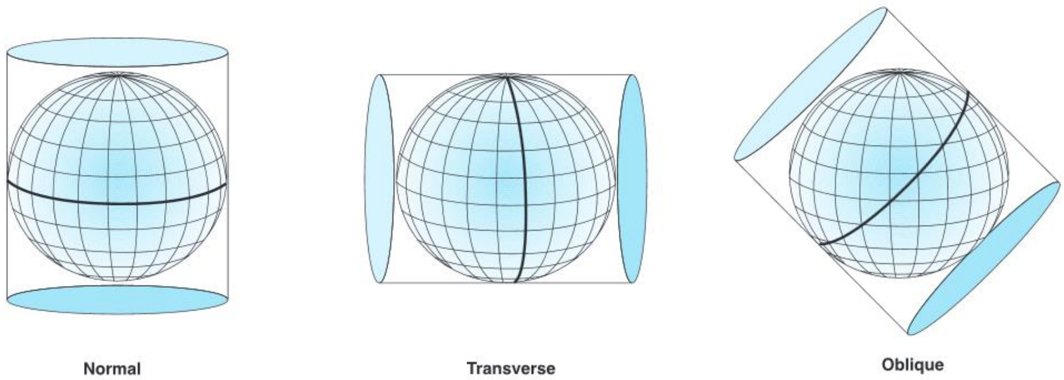


**Figure 3.6:** Cylindrical map projection types (Kennedy, 2000).

In this study, the transverse Mercator projection has been used because the map provider, Mapbox, supports the Web Mercator projection that is referred to as EPSG:3857 (WGS84 Web Mercator Projection). The Web Mercator projection is a sort of the Mercator projection that is mostly used in Web-based mapping applications. The primary difference between the Mercator and the Web Mercator projections is that the Web Mercator projection uses the spherical formulas while the large-scale Mercator projection uses the ellipsoidal formulas.

#### 3.3.3.1 Transverse Mercator Projection and UTM

Mercator projection, which is a type of cylindrical projection, is based on enclosing the globe by a visionary cylinder touching at the equator and projecting the earth onto the cylinder. Since the angles and small shapes on the globe project as the same angles and shapes on the map, this projection is a conformal projection. In a small neighbourhood of the equator, there are microscale errors on a Mercator map. However, if it is desired to map an area, which is along a meridian of longitude, the distortion would increase too much. In order to deal with this situation, the opinion of a Mercator projection, whose cylinder touches the earth along a meridian of longitude, reveals the Transverse Mercator projection. Since the Transverse Mercator projection is very accurate in narrow zones, it has become the basis for a global coordinate system called as Universal Transverse Mercator System (UTM). In UTM projection, the transverse cylinder rotates by $6^{\circ}$ increments, so it creates 60 projection zones as a strip. Each zone has a central meridian, which broadens north-south along the middle of the zone by $3^{\circ}$. In the UTM projection, all zones are projected differently by using different coordinate system (Figure 3.7). Therefore, they should not be combined because of the distortion.

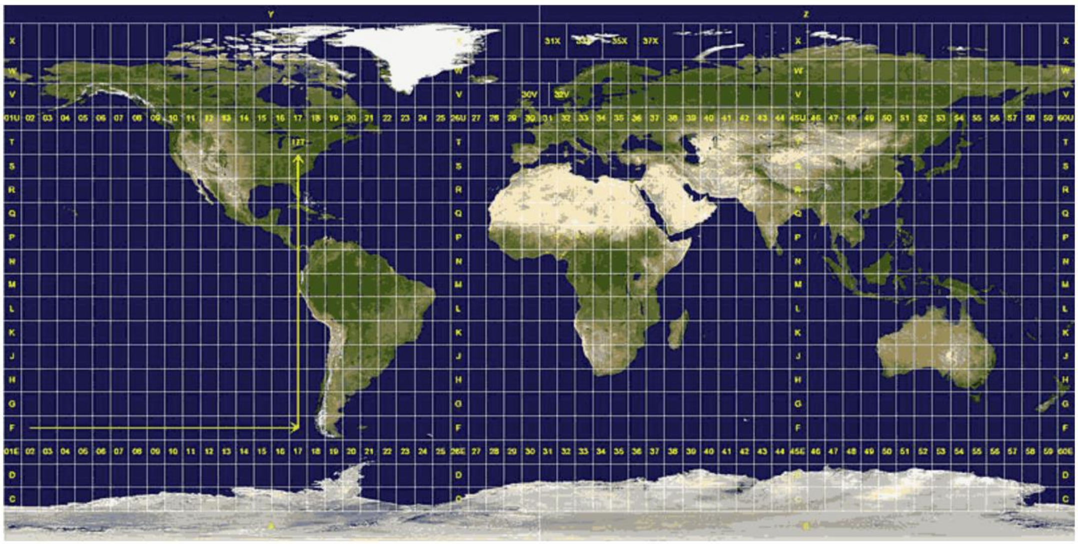

**Figure 3.7:** UTM grid zones on a projected map of the Earth.

### 3.4 Centrography Methods

Since the primary objective in centrography is to find the focal points or areas, there are several mathematical equations available to calculate the centre points according to the needs of the subject. Some of these methods are the mean centre, which is also known as spatial mean, weighted spatial mean and median centre. The final centre found with one of these methods represents the centre of the centroid. To locate the exact centroid, the radius needs to be defined after having the centre. To find the radius, there are several methods can be applied such as standard distance deviation and distance methods.

#### 3.4.1 Mean centre (Spatial mean)

The mean centre is a central or average location of a set of points (Wong and Lee, 2005). The calculation of the mean centre contains the summation of X and Y values of each coordinate individually and dividing each of these sums by the number of coordinates. This calculation generates the mean X and Y values, which produce the average coordinate of the spatial dataset. The formula of this calculation is as follows:

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (3.1)$$

The calculation of the mean centre for a sample dataset (Table 3.1), which has been used in this study, has been shown below (3.2, 3.3).

**Table 3.1:** Sample earthquake location dataset.

| Earthquake ID | Latitude (°) | Longitude (°) |
|---|---|---|
| 1 | 17.112 | 26.443 |
| 2 | 21.002 | 31.312 |
| 3 | 18.442 | 27.745 |
| 4 | 18.856 | 29.568 |
| 5 | 20.518 | 30.214 |
| 6 | 19.794 | 28.645 |
| 7 | 17.483 | 27.826 |
| 8 | 19.639 | 28.194 |
| 9 | 20.168 | 29.477 |
| 10 | 18.964 | 27.365 |

$$\sum_{i=1}^{10} X_i = 191.978^\circ \qquad \bar{X} = 19.198^\circ \tag{3.2}$$

$$\sum_{i=1}^{10} Y_i = 286.789^\circ \qquad \bar{Y} = 28.679^\circ \tag{3.3}$$

The spatial mean point of the sample dataset is located at the point of (19.198, 28.679).

### 3.4.2 Weighted mean centre

The weighted mean centre is the average location, which is produced by weighting coordinates with another variable. In this case, the coordinates of the weighted mean centre, ($\bar{X}_W, \bar{Y}_w$), are given by:

$$\bar{X}_W = \frac{\sum_{i}^{N} X_i w_i}{\sum_{i}^{N} w_i} \ , \ \bar{Y}_w = \frac{\sum_{i}^{N} Y_i w_i}{\sum_{i}^{N} w_i} \tag{3.4}$$

where "w" represents the weight for each value.

The calculation of the weighted mean centre of the sample dataset (Table 3.2) has been shown below (3.5, 3.6).

**Table 3.2:** Sample earthquake location and magnitude dataset.

| Earthquake ID | Latitude (°) | Longitude (°) | Magnitude (°) |
|---|---|---|---|
| 1 | 17.112 | 26.443 | 5.5 |
| 2 | 21.002 | 31.312 | 6.4 |
| 3 | 18.442 | 27.745 | 7.9 |
| 4 | 18.856 | 29.568 | 8.2 |
| 5 | 20.518 | 30.214 | 4.4 |
| 6 | 19.794 | 28.645 | 7.5 |
| 7 | 17.483 | 27.826 | 8.0 |
| 8 | 19.639 | 28.194 | 4.7 |
| 9 | 20.168 | 29.477 | 3.8 |
| 10 | 18.964 | 27.365 | 5.6 |

$$\sum_{i=1}^{10} X_i w_i = 1182.578^\circ \qquad \bar{X}_w = 19.074^\circ \tag{3.5}$$

$$\sum_{i=1}^{10} Y_i w_i = 1775.632^\circ \qquad \bar{Y}_w = 28.639^\circ \tag{3.6}$$

### 3.4.3 Median centre

The median centre for a dataset is the point, which the half of the values in the dataset are smaller while the other half is larger than this value. When the dataset contains two-dimensional values such as locations then the median centre is described as the point where 50% of the values fall east of north/south line and 50% fall west of that line, while an east/west line divides the values where 50% fall north of the line and 50% fall south of that line (Figure 3.8). The point, at which these lines intersect, is the median centre of the locations.

The median centre for a dataset is calculated by arranging in the values either ascending or descending order. The value at medium represents the median value. If the size of the dataset is even, the average of the two values at middle gives the median value. For a dataset contains two-dimensional values such as locations, the median centre is calculated by finding median values of each component of the points in the dataset individually and generating a new point with these median values.

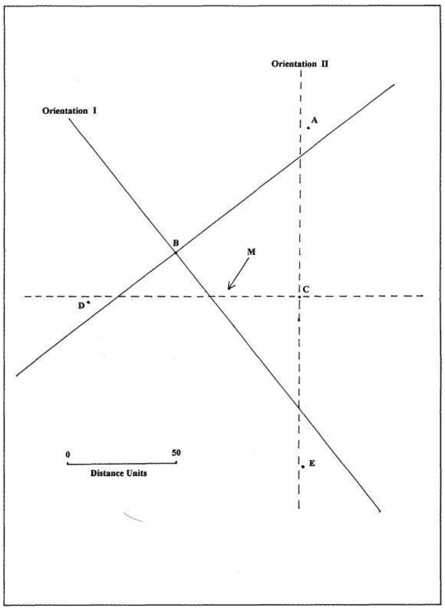


**Figure 3.8:** Representation of the median centre (Stoddard R., 1993).

### 3.4.4 Standard distance deviation

Standard distance deviation (also known as standard distance and analogous to standard deviation unclassical statistics) (Furfey, 1927; Bachi, 1957), is a distance that is applied as the radius of a circle known as standard distance circle (SDC) centred at the spatial mean centre of the spatial point pattern (Mamuse et. al, 2009). A large standard distance deviation indicates that the data points can spread far from the mean centre and a small standard distance deviation indicates that they are clustered tightly around the mean centre. The formula to calculate the standard distance deviation is given by:

$$\sigma_{\bar{X}\bar{Y}} = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \bar{X})^2 + \sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N}} \tag{3.7}$$

where $\bar{X}$ and $\bar{Y}$ values represent the latitude and longitude components of the centre point, which is calculated with one of the average finding methods such as mean centre, median centre and weighted mean centre.

The standard distance deviation calculated for the sample dataset (Table 3.1) indicates that the centralised area has a radius of 1.838 (3.8).

$$\sigma_{\bar{X}\bar{Y}} = 1.838 \tag{3.8}$$

### 3.4.5 Distance methods

Since the centrography aims to locate the concentration areas of a dataset, calculating the distances between the phenomens and the centre points is indispensable in order to draw an area. Therefore, it is crucial to choose the most accurate distance method to have more precise results. In this study, Euclidean and Mahalanobis distance methods have been applied and the results of each method have been compared.

#### 3.4.5.1 Euclidean distance

Euclidean distance (ED) is the straight-line between two points in Cartesian coordinates. To calculate the Euclidean distance between two points in 2-dimensional Cartesian coordinates, the following equation has been applied:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (3.9)$$

where $d$ is the Euclidean distance, $x$ and $y$ values are the coordinate components of the points. In this thesis, the Euclidean distance has been applied to obtain the distance between the mean centre point and the earthquake locations. The distance results have been used to define the furthest earthquake to the mean centre. The furthest earthquake in each iteration has been eliminated from the dataset since the aim is to find a concentration area. The Euclidean distance method has been applied on the dataset (Table 3.1) and the first two earthquakes are given as the two furthest earthquakes to the mean centre (Table 3.3).

**Table 3.3:** Euclidean distance result table.

| Earthquake ID | Latitude (°) | Longitude (°) | Euclidean Distance (°) |
|---|---|---|---|
| 1 | 17.112 | 26.443 | 3.057 |
| 2 | 21.002 | 31.312 | 3.192 |
| 3 | 18.442 | 27.745 | 1.201 |
| 4 | 18.856 | 29.568 | 0.952 |
| 5 | 20.518 | 30.214 | 2.024 |
| 6 | 19.794 | 28.645 | 0.597 |
| 7 | 17.483 | 27.826 | 1.915 |
| 8 | 19.639 | 28.194 | 0.655 |
| 9 | 20.168 | 29.477 | 1.256 |
| 10 | 18.964 | 27.365 | 1.334 |

#### 3.4.5.2 Mahalanobis distance

Mahalanobis distance (MD), which is also called as normalised Euclidean distance, is a method that takes the correlation in the data into account while calculating the distance. The correlation in the data is calculated using the inverse of the variance-covariance matrix of the dataset (De Maesschalck et al., 2000). In this method, the variance-covariance matrix is constructed based on the dataset. To calculate the variance-covariance matrix, the following equation is applied:

$$X_c = X - \bar{X} \quad (3.10)$$

$$C_x = \frac{1}{(n-1)}(X_c)^T (X_c) \quad (3.11)$$

where $X$ is the matrix consists of the $n$ data values, $\bar{X}$ is the matrix consists of the mean values, $C_x$ is the variance-covariance matrix. When the variance-covariance matrix is calculated, the inverse of the matrix is calculated to be used in distance equation. The equation to calculate Mahalanobis distance is as follows:

$$d_i = \sqrt{(X_i - \bar{X})C_x^{-1}(X_i - \bar{X})^T} \quad (3.12)$$

where $C_x^{-1}$ is the inverse of the variance-covariance matrix. The equation gives a diagonal matrix, where the distance values for each earthquake are located at the diagonal of the matrix.

The calculations done on the sample earthquake dataset in Table 3.1 gives different result than the Euclidean distance method. While the furthest earthquake is found as the same in both methods, the second furthest earthquake found with Mahalanobis distance method is different than the one found with Euclidean distance method (Table 3.4).

**Table 3.4:** Mahalanobis distance result table.

| Earthquake ID | Latitude (X) | Longitude (Y) | Mahalanobis Distance (°) | Euclidean Distance (°) |
|---|---|---|---|---|
| 1 | 17.112 | 26.443 | 1.752 | 3.057 |
| 2 | 21.002 | 31.312 | 1.907 | 3.192 |
| 3 | 18.442 | 27.745 | 0.681 | 1.201 |
| 4 | 18.856 | 29.568 | 1.674 | 0.952 |
| 5 | 20.518 | 30.214 | 1.145 | 2.024 |
| 6 | 19.794 | 28.645 | 0.966 | 0.597 |
| 7 | 17.483 | 27.826 | 1.801 | 1.915 |
| 8 | 19.639 | 28.194 | 1.288 | 0.655 |
| 9 | 20.168 | 29.477 | 0.829 | 1.256 |
| 10 | 18.964 | 27.365 | 1.478 | 1.334 |

The comparison of the results obtained by the Mahalanobis and Euclidean distance methods show that the methods sort the earthquakes in different order. For example, while the Euclidean distance method shows the earthquake assigned with the ID of 3 is the closest to the mean centre, the Mahalanobis distance method gives the earthquake assigned with the ID of 6 is the closest to the mean centre. On the other hand, the Euclidean distance method gives the earthquake assigned with the ID of 7 is the second furthest to the mean centre whereas the Mahalanobis distance method says the earthquake assigned with the ID of 1 is the second furthest earthquake. The reason of the difference between the distance methods is the difference in the principle of the methods. The Mahalanobis distance takes the correlation in the data into account while finding the distances. In other words, the distance between the same earthquake and the mean centre might be found different for a different dataset, where the mean centre has the same value. Therefore, it is obvious that the Mahalanobis distance method is fully dependent of the correlation of the dataset. However, the Euclidean distance method takes the points into account individually so that each point in the dataset has its own effect on the calculation of the distance, except the calculation of the mean centre. In other words, the distance between the same point and the mean centre will be always same in a different dataset, where the mean centre has the same value. Therefore, it is crucial to choose the most proper method to have correct results.

In this study, both the Mahalanobis and the Euclidean distance methods have been applied separately in the algorithm to compare the results. It’s been understood that the centroids obtained by applying the Mahalanobis distance method have greater error margins than the ones calculated by the Euclidean distance. The centroids obtained by both distance methods separately have been displayed on the map to see the error margins (Figure 3.9, 3.10).

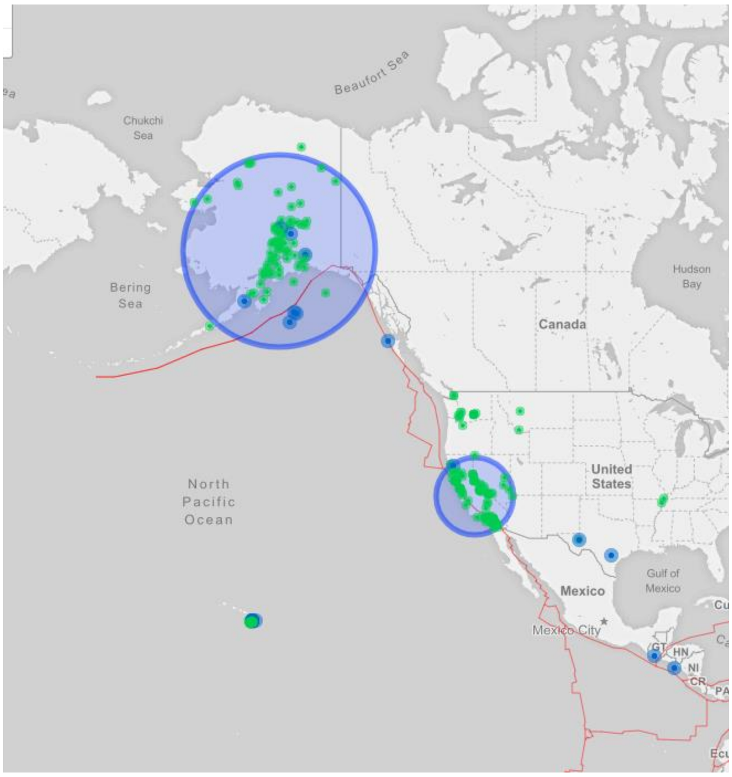


**Figure 3.9:** The centroids calculated by the Euclidean distance method.

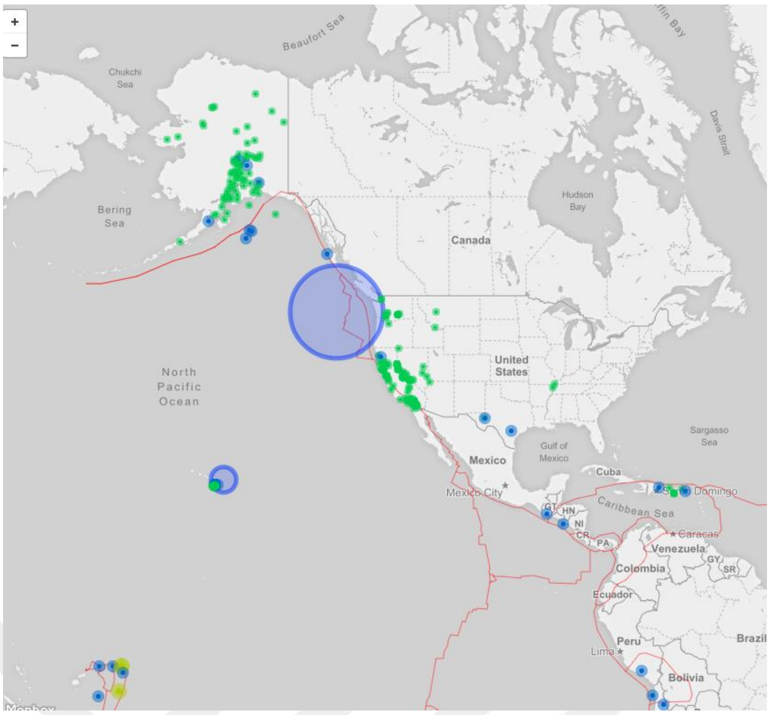


**Figure 3.10:** The centroids calculated by the Mahalanobis distance method.

The results show that the centroids calculated by the Mahalanobis distance method do not fully cover the earthquakes seem they are shifted because the method doesn't takes the exact distance into account as it is done in the Euclidean distance method. Instead, the Mahalanobis distance method calculates the relative distances based on the correlation between the earthquakes and the mean centres.

# 4. CASE-STUDY: CENTROGRAPHY ANALYSIS ON EARTHQUAKES

## 4.1 Data Gathering and Data Modelling

In this study, the spatial-temporal earthquake data including magnitude has been gathered through the web services of the United States Geological Survey (USGS). The web services provide the real-time data, and they are reachable from any platform because of being open data platform. The functions and the parameters that can be used to filter the data are shared in the manifests of the web services (Figure 4.1). The web service provides several parameters such as time interval, minimum magnitude and latitude and longitude based region definition to query the earthquake data.

**Time**

All times use ISO8601 Date/Time format. Unless a timezone is specified, UTC is assumed. Examples:

- *2018-04-08*, Implicit UTC timezone, and time at start of the day (00:00:00)
- *2018-04-08T15:51:22*, Implicit UTC timezone.
- *2018-04-08T15:51:22+00:00*, Explicit timezone.

| parameter | type | default | description |
|---|---|---|---|
| `endtime` | String | present time | Limit to events on or before the specified end time. NOTE: All times use ISO8601 Date/Time format. Unless a timezone is specified, UTC is assumed. |
| `starttime` | String | NOW - 30 days | Limit to events on or after the specified start time. NOTE: All times use ISO8601 Date/Time format. Unless a timezone is specified, UTC is assumed. |
| `updatedafter` | String | null | Limit to events updated after the specified time. NOTE: All times use ISO8601 Date/Time format. Unless a timezone is specified, UTC is assumed. |

**Location**

Requests that use both rectangle and circle will return the intersection, which may be empty, use with caution.

**Rectangle**

Requests may use any combination of these parameters.

| parameter | type | default | description |
|---|---|---|---|
| `minlatitude` | Decimal [-90,90] degrees | -90 | Limit to events with a latitude larger than the specified minimum. NOTE: min values must be less than max values. |
| `minlongitude` | Decimal [-360,360] degrees | -180 | Limit to events with a longitude larger than the specified minimum. NOTE: rectangles may cross the date line by using a minlongitude < -180 or maxlongitude > 180. NOTE: min values must be less than max values. |
| `maxlatitude` | Decimal [-90,90] degrees | 90 | Limit to events with a latitude smaller than the specified maximum. NOTE: min values must be less than max values. |
| `maxlongitude` | Decimal [-360,360] degrees | 180 | Limit to events with a longitude smaller than the specified maximum. NOTE: rectangles may cross the date line by using a minlongitude < -180 or maxlongitude > 180. NOTE: min values must be less than max values. |

**Figure 4.1:** The parameters defined in the manifest of the web services of the USGS to filter the earthquake data (USGS API Documentation, 2018).

In order to request and analyse the data, a web GIS application has been developed in this study. First of all, the earthquake data model has been designed and implemented in the web application based on the response structure in GeoJSON format that is described in the API documentation of the web service (Figure 4.2). The GeoJSON response has been analysed and parsed into indivisible objects according to the relationship between the data types in the response. For example, the metadata and the spatial data has been separated, and the location information has been extracted to the geometry class, which is included by the feature class that also includes the properties.

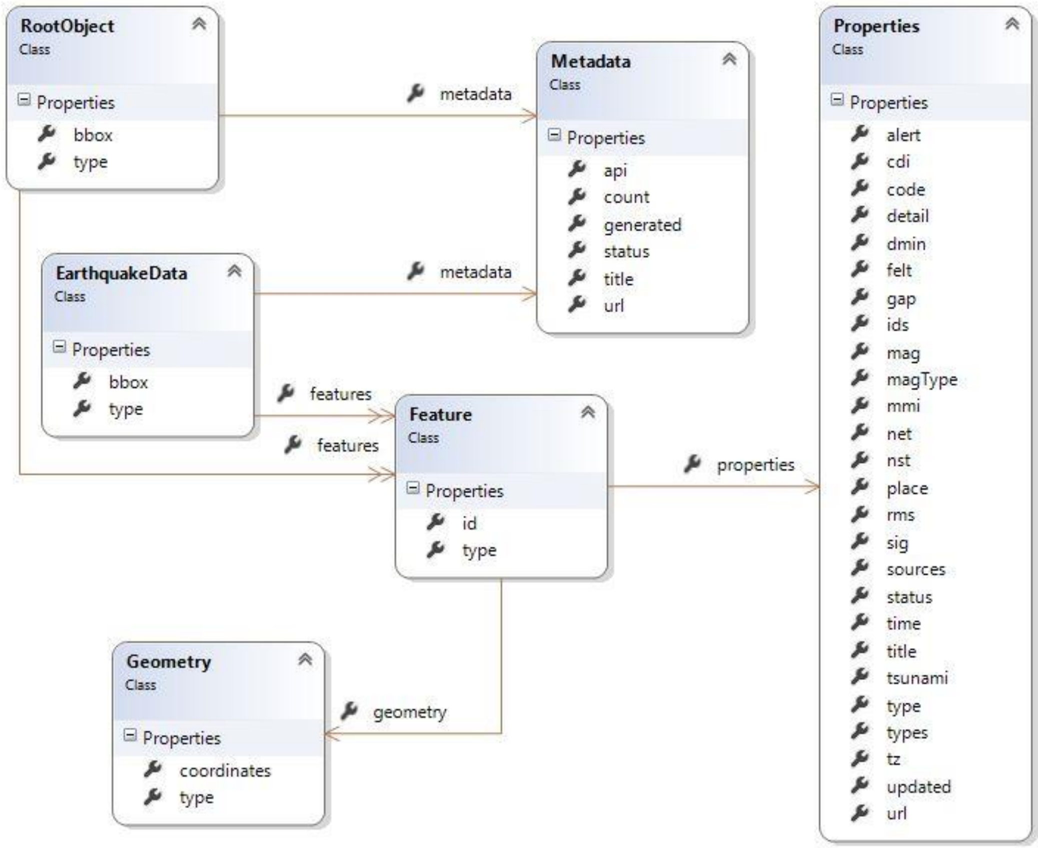


**Figure 4.2:** The class diagram of the earthquake data in web GIS application.

## 4.2 Designing the Web GIS Application

A panel has been designed in the web application to define the parameters as input values to be sent within the HTTP request package to the web service in order to get the filtered earthquakes. In this study, the parameters are limited with the date, location and magnitude fields (Figure 4.3).

To display the map on the web page, the JavaScript and CSS libraries of the Mapbox have been integrated into the project. The map has been configured the be a thematic map as to display the earthquakes with the icons having distinct colours and sizes based on the magnitude information. The colour and magnitude scale has been added to the bottom-right corner of the map as a legend.


**Figure 4.3:** The parameters input panel and the map in the web GIS application.

On the back-end development, first, a function that interacts with the web services of USGS has been developed to query the data and parse the response. To build the connection, an instance of the WebClient class, which is defined in the .NET Framework 4.6.1, has been created and the generated URL that contains the address of the web service and the parameters has been provided to the WebClient object so that the necessary information can be requested. Secondly, the GeoJSON response has been parsed to the collection object that contains the earthquake data objects, and the collection has been transmitted to the front-end functions. On the front-end side, a feature layer has been created to bind the earthquake list on the map. First, a feature object that consists of type, geometry and property objects has been created for each earthquake in the list, and while creating the feature object, only the necessary information such as coordinates, magnitude and the title has been extracted from the earthquake object to lessen the memory load. Also, icons are defined for each earthquake based on the magnitude. When the feature layer is filled with the earthquakes, then they are displayed on the map (Figure 4.4).

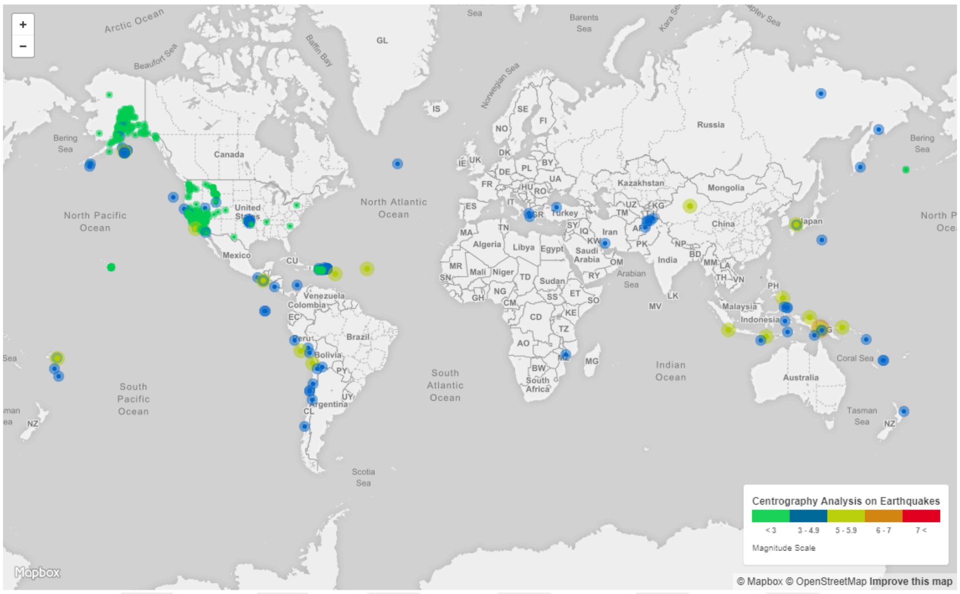


**Figure 4.4:** The representation of a sample earthquake dataset on the map.

## 4.3 Implementation of the Centrography Algorithm

An algorithm has been developed for this study to apply centrography methods and find the concentration areas. The algorithm consists of an iterative method, where the non-centralised earthquake subset, which is the full dataset for the first iteration, is processed in each iteration. First of all, the mean centre point is calculated for the full set of the earthquake at the beginning of the iteration. Then, each earthquake's location is compared with the weighted mean centre regarding distance. The furthest earthquake to the mean centre is eliminated from the copy of the earthquake list, which will be considered as the centralised earthquake list at the end of the iteration. To determine the furthest earthquake from the mean centre, the distance between the mean centre and the earthquakes must be calculated for each earthquake. Although there are a number of distance calculation methods, the Euclidean and Mahalanobis distance methods have been used to calculate the distances. The process of the determination and elimination of the furthest earthquake is repeated until the 40%, which is defined as a threshold in this study, of the earthquakes remain in the final list. The final list is considered as the list of the earthquakes that are concentrated within an area. After the determination of the final list, the earthquakes in the final list are excluded from the initial list in order to avoid being reprocessed. Then, the initial iteration is re-executed to find the other centroids. Besides, each concentration areas found in each iteration are expected to pass a confidence test to be considered as correct centroids. To determine whether a centroid is correct or not, the radius of the centroid is compared with the maximum distance between the earthquakes in the initial set. If the radius of the centroid is equal to or smaller than the 2% of the maximum distance, then the centroid is tagged as correct and displayed on the map (Figure 4.5, 4.6).

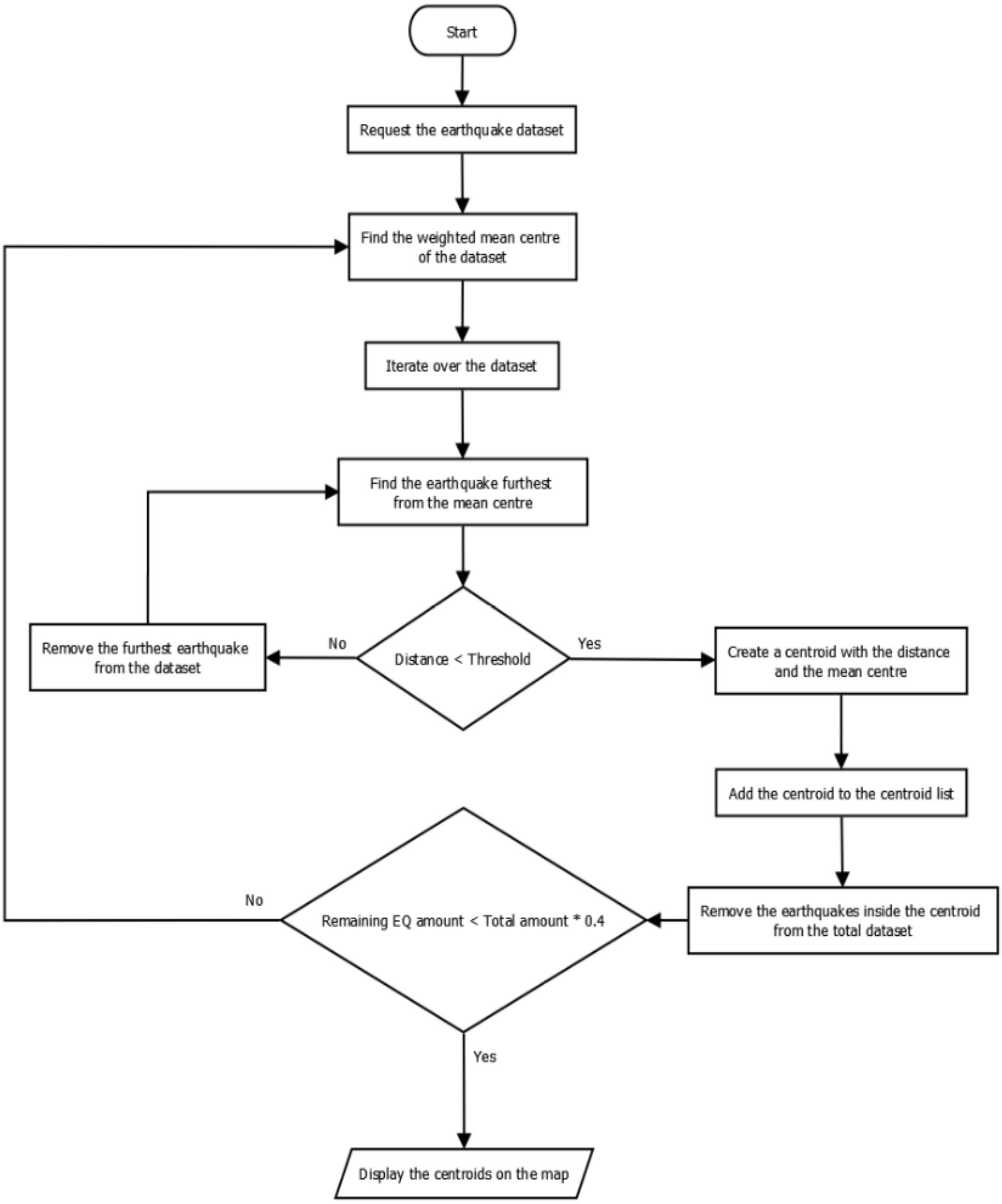


**Figure 4.5:** The workflow of the centrography algorithm.

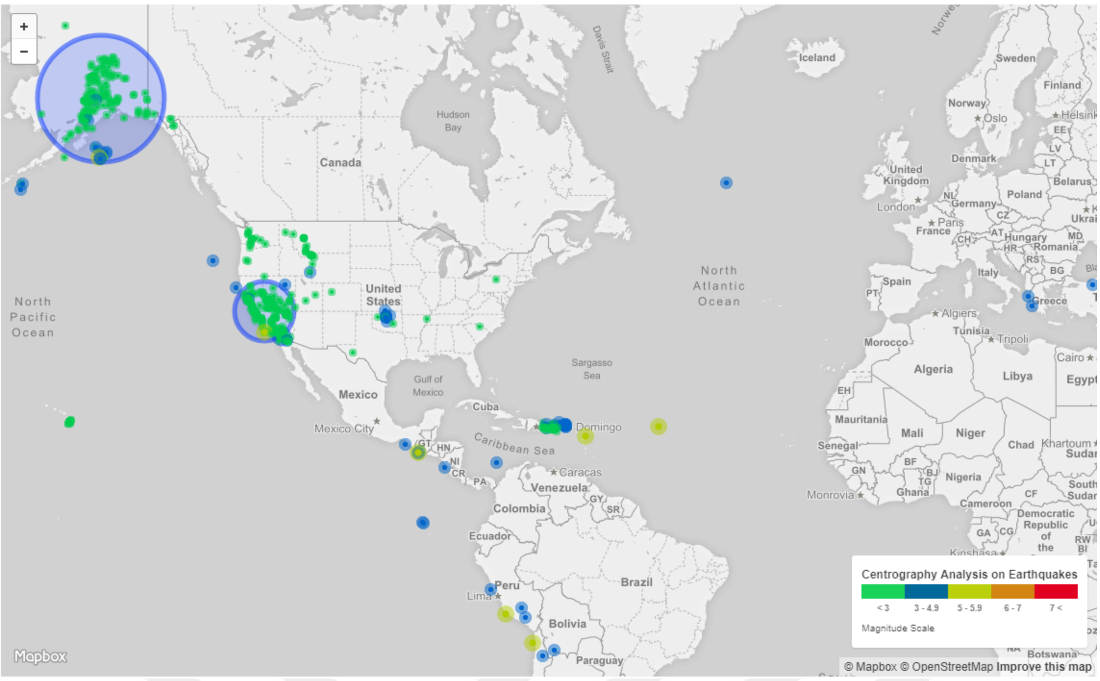


**Figure 4.6:** The representation of the centroids on the map.

Since all map projection methods have distortions and errors, it is essential to define how to calculate the real distance between two points on a projected map. The distance between consecutive longitudes differs according to the latitude on a 2-D map, onto which the earth is projected with the transverse method that has also been used in this study. The Haversine Formula (4.1, 4.2, 4.3) has been applied to the latitude and longitude ranges to calculate the exact distance. The Haversine Formula is:

$$a = \sin^2(\frac{\Delta\varphi}{2}) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2(\frac{\Delta\lambda}{2}) \tag{4.1}$$

$$c = 2 \cdot \arcsin(\min(1, \sqrt{a})) \tag{4.2}$$

$$d = R \cdot c \tag{4.3}$$

where $\varphi$ is the latitude, $\lambda$ is the longitude and $R$ is the radius of the Earth, which is 6371 km. The unit of the latitude and longitude values have been converted to radian unit. This calculation produces the distance in km. The radius' of the centroids have been calculated by the standard distance deviation method, which produces the result in radian. The Haversine Formula has also been applied to the radiuses of the centroids to have the exact length in km.

### 4.4 Centrography with Curvilinear Approach

Centrography methods have been mostly used to locate the concentration regions with the circular or eliptical approach. However, it is not always possible to find any centroid for a dataset if the spatial data is dispersed on a wide region. In case the data is spread along a pattern and if a curve can be fit for the pattern, then the shortest distance between the phenomenon's locations and the curve can be taken into account to define the curvilinear centroids.

In this study, it's been noticed after several deep inspections with different filters that the earthquakes, which have a magnitude greater than 7.0, are spread along a pattern. First of all, another algorithm has been developed to generate a 5$^{th}$ degree polynomial curve in order to analyse the pattern of the dispersion. After the analysis of the curve, it is been detected that the curve's pattern seems similar to the major fault lines that are also called the tectonic plate boundaries. Then, the GeoJSON information of the tectonic plate boundaries has been also displayed on the map so that the fitted curves and the major fault lines can be compared against each other (Figure 4.7, 4.8).

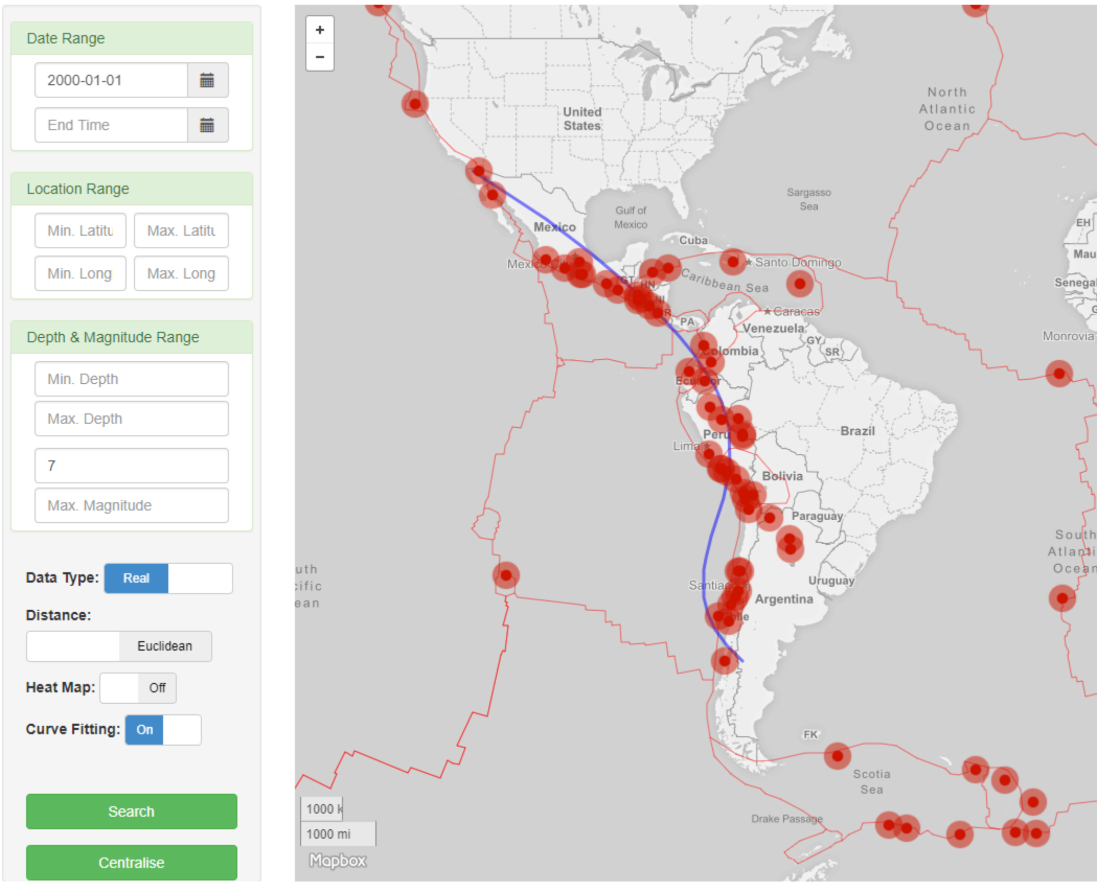


**Figure 4.7:** The fitted curve (blue) and the major fault lines (red) for America plates.

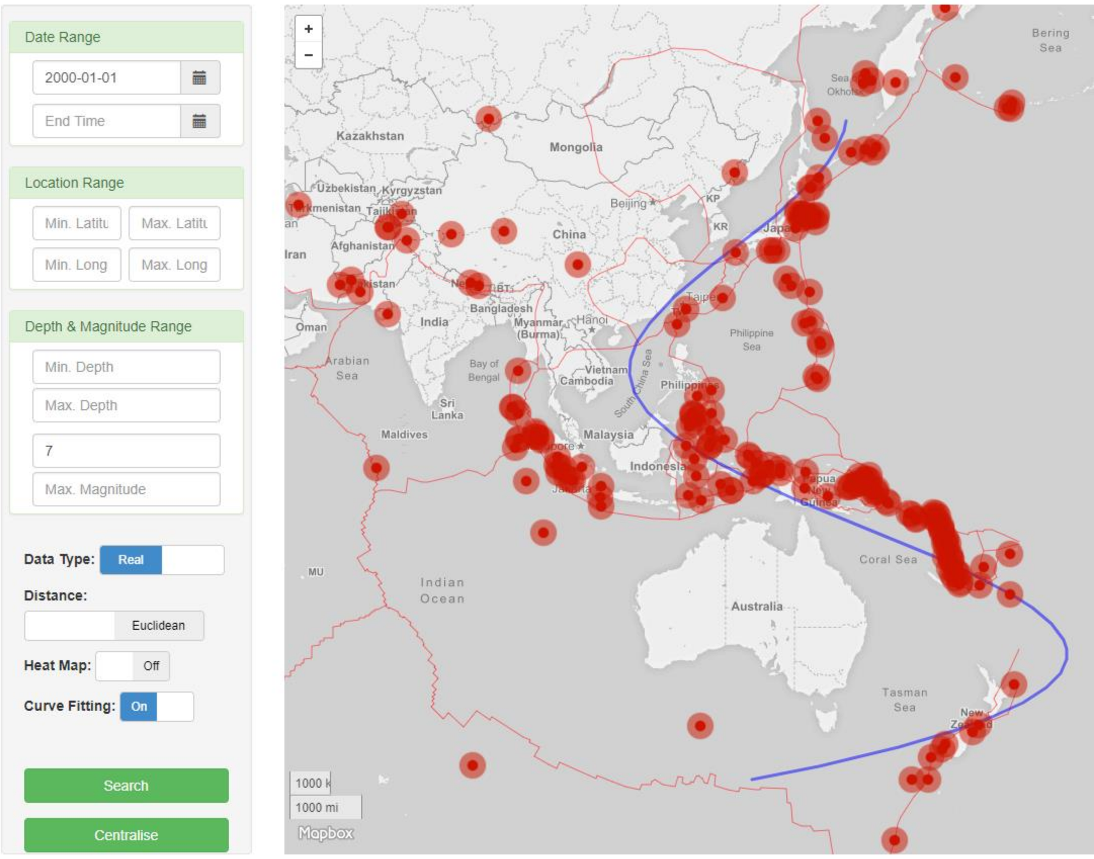


**Figure 4.8:** The fitted curve (blue) and the major fault lines (red) for Asia plates.

## 4.5 Conformity Test and Analysis of the Results

To test the accuracy of the concentration areas found by the centrography algorithm applied in this study, the results are shown in several histograms such as scatter chart and distance distribution graph. It's been aimed to analyse the distribution of the distances between the earthquakes and the centroids so that the accuracy of the centrography algorithm can be tested. On the other hand, the scatter chart has been generated to see the relative dispersion by shifting the axes, where the centroid centres are located at (0,0). The locations of the earthquakes have also been shifted accordingly.

First of all, when the central zones are defined (Figure 4.9) after the iterations, the distance between each earthquake and the mean centres has been calculated. The distance values have been added to the histogram to obtain a distance distribution graph. If there is more than one centralised zone, the distance distribution histograms have been generated for each zone separately (Figure 4.10, 4.11).

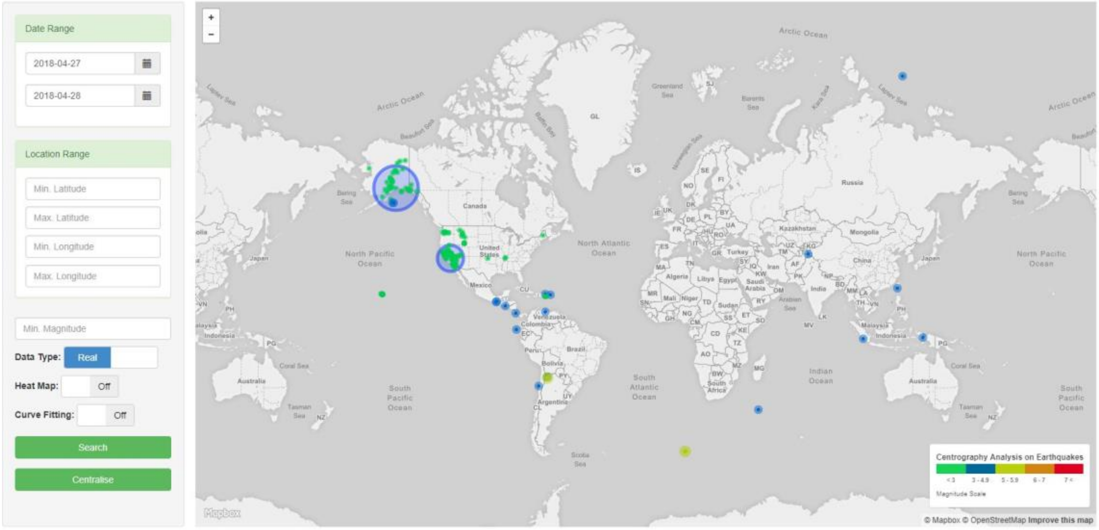


**Figure 4.9:** The centroids for the earthquakes between 27.04.2018 and 28.04.2018.

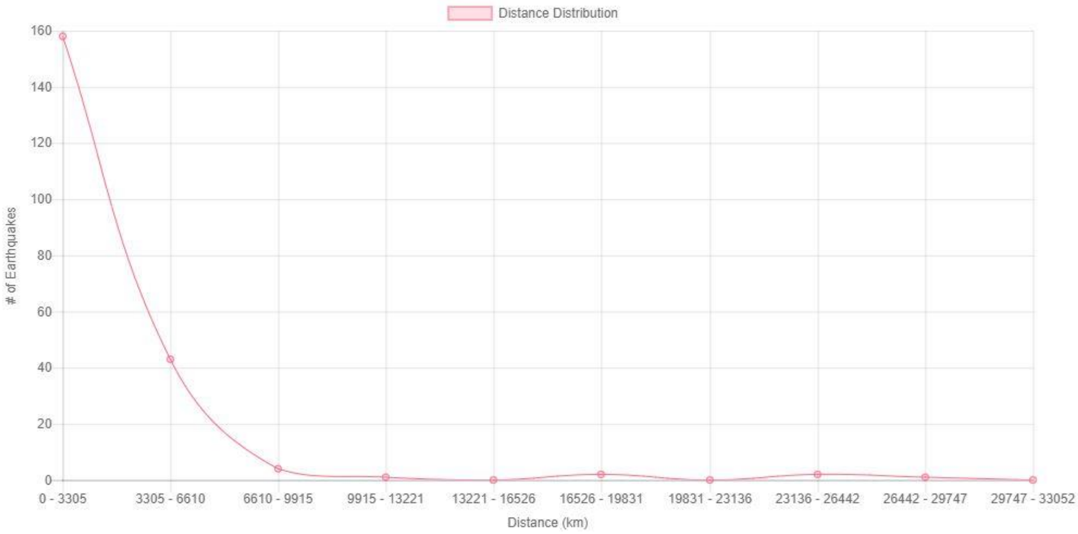


**Figure 4.10:** The distance distribution for the first centroid.

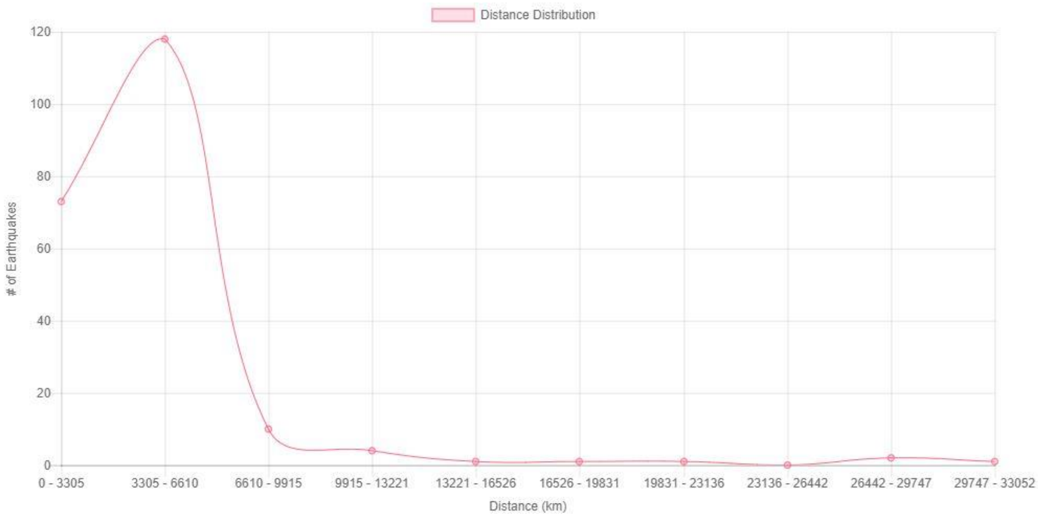


**Figure 4.11:** The distance distribution for the second centroid.

The first histogram (Figure 4.10) shows that 75% of the earthquakes in the full dataset are in 3305 km range from the first centroid.

Secondly, the centrography algorithm has been compared with the heat map results. Heat map is a thematic map type that is mostly used to distinguish the concentration zones by finding a correlation based on the distance between the phenomenons to generate a clustering points. In heat maps, the transition between the markers on the map is smoother than the clustered maps. Basically, a heat map uses a colour scale to show the frequency difference. In most of the heat maps, the colour scale is defined as from blue to red, from fewer elements to more elements respectively. In this study, the red zones represent the concentration zones for the earthquakes (Figure 4.12).

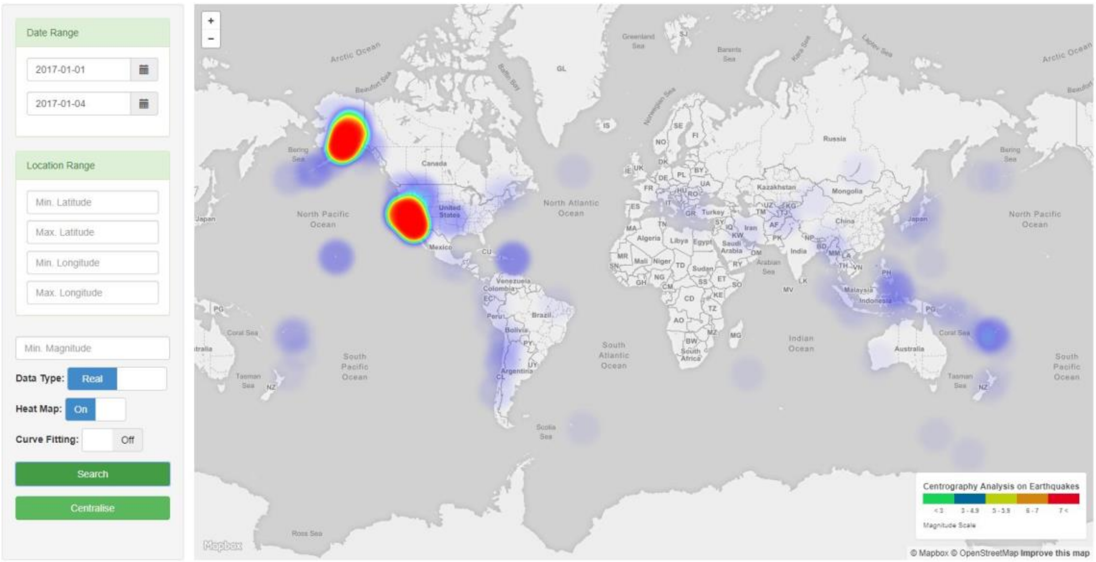


**Figure 4.12:** The heat map representation of the earthquakes.

To test the precision of the centroids, heat map is a practical tool because the centroids are simply expected to be positioned at the same location with the high temperature zones on the map. The centroids generated by the centrography algorithm developed in this study are located as covering the high temperature zones on the heat map (Figure 4.13), which means that the centroids can be considered as accurate.

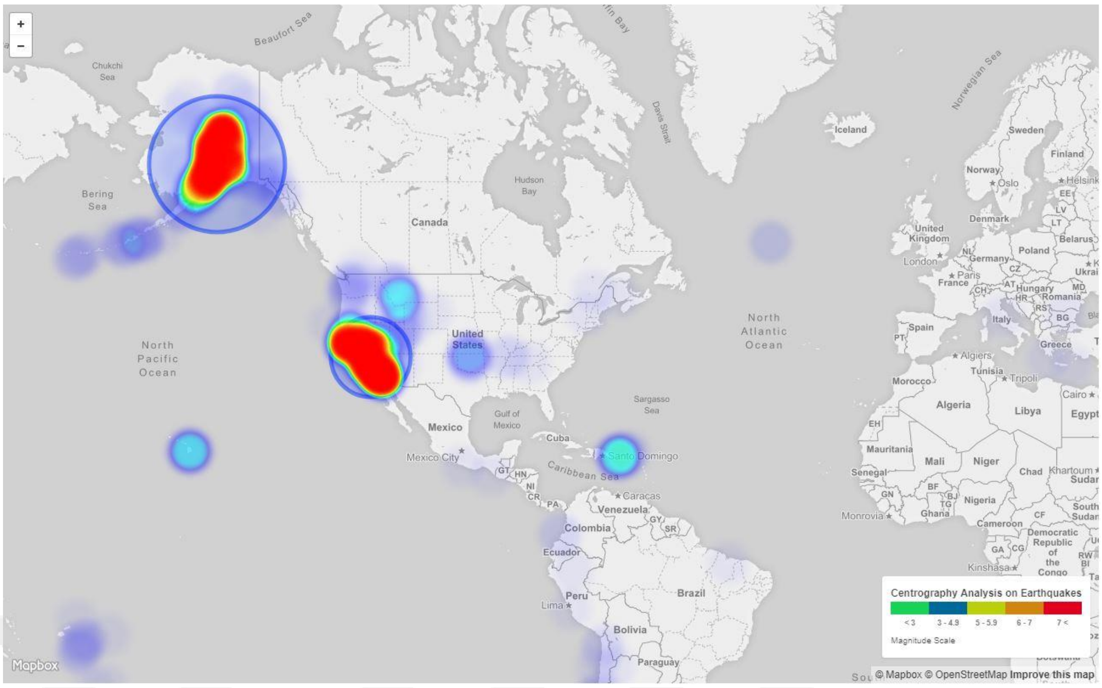


**Figure 4.13:** The location of the centroids on the heat map.

The Web GIS application developed in this study has been published on a public domain so that anyone, who is interested in such analysis, can do the analysis on the earthquake data, which can be filtered with the provided parameters. The Web application can be reached at http://centro.boransimsek.com.

## 5. CONCLUSIONS

In this study, the earthquakes happened since the year of 1900 have been projected on a map and an analysis have been done to locate the concentration areas in order to have an idea about where the earthquakes are more frequent. To have a better argument, the earthquakes have been queried with a number of different parameters such as time intervals, magnitude, depth and location ranges. Once the earthquake data is obtained and displayed on the map, the centrography algorithm has been applied on the earthquake dataset to show the concentration zones. The results show the plate contains the states of San Francisco and Alaska in the United States has the most of the earthquakes (Figure 5.1).

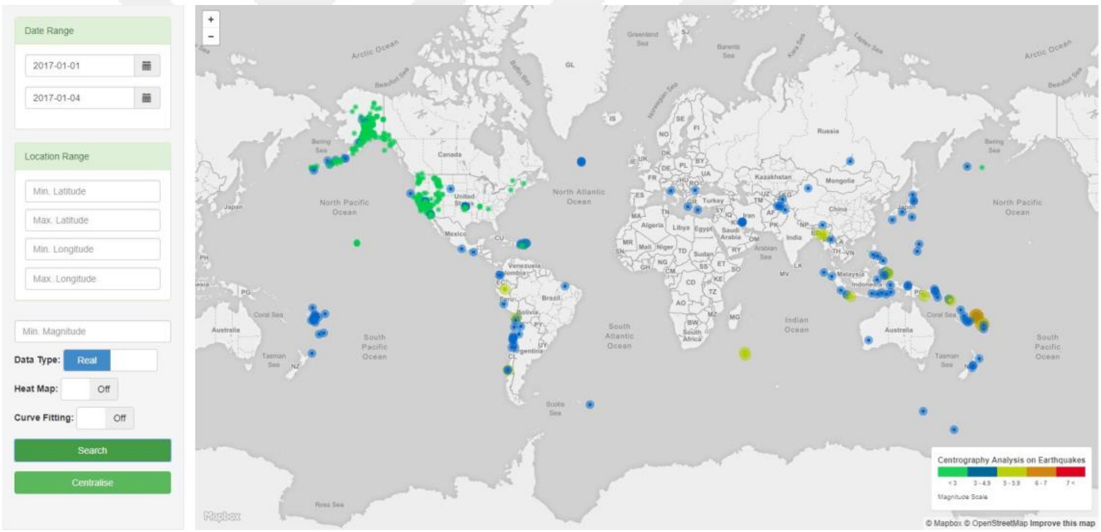


**Figure 5.1:** The distribution of the earthquakes happened between 01.01.2017 and 04.01.2017.

However, the magnitudes of the earthquakes are mostly less than 3.0, which means that the earthquakes are not highly destructive as the ones happen on the plate that contains Australia and Philippines. In addition, the plate between Alaska and San Francisco averagely holds more than 300 earthquakes per day. It can be easily dedus that the plate is notably active in terms of geological dynamism.

The earthquakes, which have magnitude greater than 7.0, have been analysed with a different approach than the circular centrography methods. The idea of this approach was to find out a pattern that gives a better opinion about the earthquakes dispersion, where a circular centroid is not possible to define. First of all, the major fault lines that form the tectonic plates have been displayed on the map as polylines to see whether such destructive earthquakes are located close to the major fault lines or not. To compare the dispersion of the earthquakes with the tectonic plate boundaries, a 5$^{th}$ degree polynomial curve fitting algorithm has been developed in the web GIS application. The polynoms that are generated by the fitting algorithm have been displayed on the map as polylines as well (Figure 5.2).

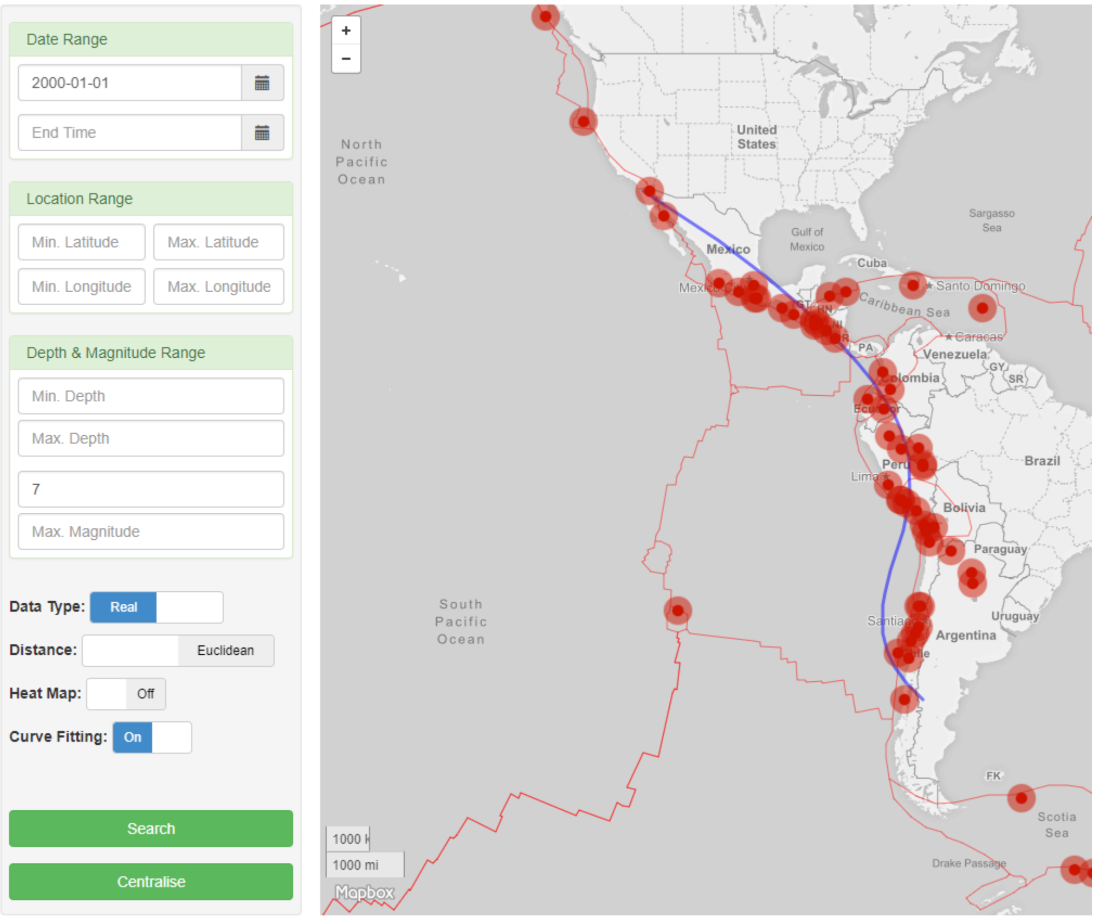


**Figure 5.2:** The fitted curve (blue) and the fault line (red).

The distances between points on the polynoms and the points on the major fault lines have been calculated along the curves by step-by-step iteration and the distances have been shown on the histogram (Figure 5.3).

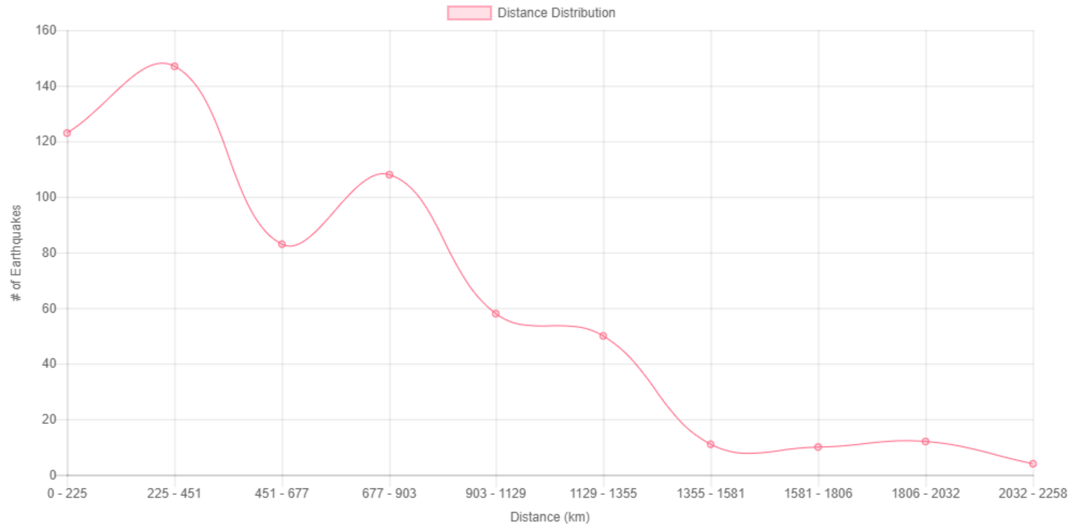


**Figure 5.3:** The distance distribution for the polynoms and the major fault lines.

The results show that 60% of the earthquakes, which have magnitude greater than 7.0, are dispersed in 500 kilometers neighbourhood of the major fault lines. It can be predicted that the destructive earthquakes tend to recur themselves at the locations, which are close to the major fault lines.

This stuıdy also shows that the curvilinear centrography approach can be useful to define a pattern for the cases that finding a circular centroid is not possible.

## REFERENCES


**Bachi, R.** (1957). Statistical Analysis of geographical Series, *Bulletin de l'Institut International de Statistique, **36***, 229-269.

**Campbell, J. E., and Shin, M.** (2012). *Geographic Information System Basics* (v. 1). Washington: FlatWorld.

**De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L.** (2000). The Mahalanobis distance, *Chemometrics and Intelligent Laboratory Systems, **50***, 1-18.

**Everitt, B. S., Landau, S., Leese M. and Stahl D.** (2011). *Cluster Analysis 5th ed.* United Kingdom: Wiley.

**Furfey, P. H.** (1927). Note on Lefever's Standard Deviational Ellipse, *American Journal of Sociology, **33***, 94-98.

**Kellerman, A.** (1981). Centrographic Measures in Geography, *Concepts and Techniques in Modern Geography, **32***, 1-33.

**Kennedy, M. and Kopp, S**. (2000). *Understanding Map Projections*. New York: Environmental Systems Research Institute (ESRI).

**Kouyoumjian, V.** (2011). *GIS in the Cloud*. New York: Environmental Systems Research Institute (ESRI).

**LeBeau, J. L.** (1987). The methods and measures of centrography and the spatial dynamics of rape, *Journal of Quantitative Criminology, **3***, 125-141.

**Longley, P., Goodchild, M., Maguire, D., Rhind, D.** (2001). *Geographical Information Systems and Science*, New York: Wiley.

**Mamuse, A., Porwal A., Kreuzer, O., & Beresford, S.** (2009). A new method for spatial centrographic analysis of mineral deposit clusters, *Ore Geology Reviews, **36***, 293-305.

**Snyder, J. P.** (1982). *Map Projections Used by the U.S. Geological Survey* (v. 2). Geological Survey Bulletin 1532. Washington, DC: United States Government Printing Office.

**Stoddard, R.** (1993). Where is the Median Center?, *Journal of Geography*, ***92***, 236-237.

**Sviatlovsky, E. E., and Eells, W. C.** (1937). The Centrographical Method and Regional Analysis, *Geographical Review*, ***27***, 240-254.

**United States Geological Survey.** (2018). *API Documentation – Earthquake Catalog*. Retrieved from https://earthquake.usgs.gov/fdsnws/event/1/.

**W3C Glossary.** (2004). *Web Services Glossary*. Retrieved from https://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/.

**Wong, D. W. S., and Lee, J.** (2005). *Statistical Analysis of geographic Information with ArcView GIS and ArcGIS*. New York: Wiley.

**CURRICULUM VITAE**

**Name Surname** : Boran Simsek

**Place and Date of Birth** : Istanbul – 11.02.1988

**E-Mail** : simsekbo@itu.edu.tr

**EDUCATION** :

- **B.Sc.** : 2013, ITU, Mathematics Engineering

**PROFESSIONAL EXPERIENCE AND REWARDS:**

- 2008-2009 Software Developer, Telegami Telesales
- 2011-2013 Software Engineer, Adayazılım
- 2013-2014 Software Specialist, Sampas
- 2015-… MIDAS Project Officer, IOM (The UN Migration Agency)

**PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:**

- **Simsek B., Seker D. Z., Guney C.,** 2016: Centroid Analysis for Earthquakes in Turkey. 37th Asian Conference on Remote Sensing (ACRS), October 17-21, 2016 Colombo, Sri Lanka.