

PONTIFICAL CATHOLIC UNIVERSITY OF RIO GRANDE DO SUL  
FACULTY OF INFORMATICS  
GRADUATE PROGRAM IN COMPUTER SCIENCE

**SPATIO-TEMPORAL DATA MINING  
IN PALAEOGEOGRAPHIC DATA WITH  
A DENSITY-BASED CLUSTERING ALGORITHM**

DAIANE HEMERICH

Dissertation presented as partial requirement for obtaining the Master's degree in Computer Science from Pontifical Catholic University of Rio Grande do Sul.

Supervisor: Prof. Dr. Duncan Dubugras Alcoba Ruiz

Porto Alegre

2014



## Dados Internacionais de Catalogação na Publicação (CIP)

H488s Hemerich, Daiane  
Spatio-temporal data mining in palaeogeographic data with a density-based clustering algorithm / Daiane Hemerich. - Porto Alegre, 2014.  
103 p.

Diss. (Mestrado) – Fac. de Informática, PUCRS.  
Orientador: Prof. Duncan Dubugras Alcoba Ruiz.

1. Informática. 2. Mineração de Dados (Informática).  
3. Paleogeografia. I. Alcoba Ruiz, Duncan Dubugras. II. Título.

CDD 005.74


**Ficha Catalográfica elaborada pelo  
Setor de Tratamento da Informação da BC-PUCRS**



Pontifícia Universidade Católica do Rio Grande do Sul  
FACULDADE DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

### TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO

Dissertação intitulada "*Spatio - Temporal Data Mining in Paleogeographic Data with Density-Based Clustering Algorithm*" apresentada por Daiane Hemerich como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, aprovada em 20/03/2014 pela Comissão Examinadora:

  
Prof. Dr. Duncan Dubugras Alcoba Ruiz -  
Orientador

PPGCC/PUCRS


  
Prof. Dr. Paulo Henrique Lemelle Fernandes -

PPGCC/PUCRS

Prof. Dr. Altigran Soares da Silva -

UFAM

Homologada em 22/05/2014, conforme Ata No. 008 pela Comissão Coordenadora.

  
Prof. Dr. Luiz Gustavo Leão Fernandes  
Coordenador.

**PUCRS**

Campus Central  
Av. Ipiranga, 6681 - P32 - sala 507 - CEP: 90619-900  
Fone: (51) 3320-3611 - Fax (51) 3320-3621  
E-mail: [ppgcc@pucrs.br](mailto:ppgcc@pucrs.br)  
[www.pucrs.br/facinf/pos](http://www.pucrs.br/facinf/pos)

## DEDICATION

This work is dedicated to my family.

*"Trust yourself to know the way  
that will prove true in the end."*

Bob Dylan

"Trust Yourself"



## ACKNOWLEDGEMENTS

They say "happiness is only real when shared", and I really can't see otherwise. The achievement of this objective is a full happiness to me, which does not deserve only to be shared, but also assigned to everyone responsible. And I have so much to thank to everyone who took part of this journey with me, in one way or another.

First of all, many thanks to my supervisor Prof. Duncan Ruiz, for his support and guidance during these two years of Masters course. Thank you for trusting on my potential and giving me the opportunity to develop my postgraduate research at PUCRS, as well as to extend it in University of Trento (Italy). Thanks also to Prof. Paulo Fernandes, for the opportunity and support, and to the domain experts Maria Alejandra Gomes Pivel (oceanografist) and Tiago Fischer (geologist), for all the help during the work on Paleoprospec Project.

I would also like to thank Prof. Fabio Casati and his Social Informatics Research Group, with whom I had the opportunity to interact during my exchange in University of Trento. Certainly the experience acquired in this exchange period aggregated much to my work and to my appreciation for research.

Thanks to all the professors with whom I had the honor to learn and build the knowledge which allowed me to keep going further, until today.

A special thanks to my colleagues of Research Group in Business Intelligence (GPIN) from PUCRS, whose help was essential to my growth during the research period. In particular, thanks to Christian Quevedo, Luciano Blomberg, Renata De Paris, Juliano Varella de Carvalho and Juliano Silveira, whose support, reviews and advices were very important. More than colleagues and co-workers, you are truly great friends. I consider myself a very lucky person for having met you.

I would also like to thank my colleagues since graduate school, Eli Maruani, Lucas Oleksinsky, Aníbal Solon and Viviane Lara, as well as all my friends with whom I had the chance to spend this period of my Masters in Porto Alegre, and somehow helped me in my work. Thanks also to all the friends I have made during my period abroad (Kostas, Zeca, Giuliano, Umut, Begüm, Anna, Tom, Mayessa, just to name a few), as well as to my closest friends Camila, Verônica, Tanise, Tinho, Gerso and 2D. I am pretty sure that all my old friends I kept and every new friend I've made during this period are part of my reward for having been a good girl and eaten all the vegetables when I was a kid. You guys are awesome! \o/

Last, but not least, my most sincere thanks to the most important people of my life: my parents, Jairo and Ana, and my brother Jardel, as well as all my family, whose support and comprehension were essential for me to get to this stage of my journey. Thank you for every minute of concern, every word of support, every day missing me so far away. You mean everything to me.





# **SPATIO-TEMPORAL DATA MINING IN PALAEOGEOGRAPHIC DATA WITH A DENSITY-BASED CLUSTERING ALGORITHM**

## **ABSTRACT**

The usefulness of data mining and the process of Knowledge Discovery in Databases (KDD) has increased its importance as grows the volume of data stored in large repositories. A promising area for knowledge discovery concerns oil prospection, in which data used differ both from traditional and geographical data. In palaeogeographic data, temporal dimension is treated according to the geologic time scale, while the spatial dimension is related to georeferenced data, i.e., latitudes and longitudes on Earth's surface. This approach differs from that presented by spatio-temporal data mining algorithms found in literature, arising the need to evolve the existing ones to the context of this research. This work presents the development of a solution to employ a density-based spatio-temporal algorithm for mining palaeogeographic data on the Earth's surface. An evolved version of the ST-DBSCAN algorithm was implemented in Java language making use of Weka API, where improvements were carried out in order to allow the data mining algorithm to solve a variety of research problems identified. A set of experiments that validate the proposed implementations on the algorithm are presented in this work. The experiments show that the solution developed allow palaeogeographic data mining by applying appropriate formulas for calculating distances over the Earth's surface and, at the same time, treating the temporal dimension according to the geologic time scale.

**Keywords:** spatio-temporal data mining, palaeogeographic data, geographic distance calculations, geologic time scale.

# **MINERAÇÃO DE DADOS PALEOGEOGRÁFICOS ESPAÇO- TEMPORAIS COM ALGORITMO DE AGRUPAMENTO BASEADO EM DENSIDADE**

## **RESUMO**

O uso da mineração de dados e do processo de descoberta de conhecimento em banco de dados (Knowledge Discovery in Databases (KDD)) vem crescendo em sua importância conforme cresce o volume de dados armazenados em grandes repositórios. Uma área promissora para descoberta do conhecimento diz respeito à prospecção de petróleo, onde os dados usados diferem tanto de dados tradicionais como de dados geográficos. Nesses dados, a dimensão temporal é tratada de acordo com a escala de tempo geológico, enquanto a escala espacial é relacionada a dados georeferenciados, ou seja, latitudes e longitudes projetadas na superfície terrestre. Esta abordagem difere da adotada em algoritmos de mineração espaço-temporal presentes na literatura, surgindo assim a necessidade de evolução dos algoritmos existentes a esse contexto de pesquisa. Este trabalho apresenta o desenvolvimento de uma solução para uso do algoritmo de mineração de dados espaço-temporais baseado em densidade ST-DBSCAN para mineração de dados paleogeográficos na superfície terrestre. O algoritmo foi implementado em linguagem de programação Java utilizando a API Weka, onde aperfeiçoamentos foram feitos a fim de permitir o uso de mineração de dados na solução de problemas de pesquisa identificados. Como resultados, são apresentados conjuntos de experimentos que validam as implementações propostas no algoritmo. Os experimentos demonstram que a solução desenvolvida permite a mineração de dados paleogeográficos com a aplicação de fórmulas apropriadas para cálculo de distâncias sobre a superfície terrestre e, ao mesmo tempo, tratando a dimensão temporal de acordo com a escala de tempo geológico.

**Palavras-Chave:** mineração de dados espaço-temporais, dados paleogeográficos, cálculos de distância geográfica, escala de tempo geológico.

## LIST OF FIGURES

Figure 1 - South Atlantic research area presented in Robinson projection, in which is possible to notice how large the region is, being thus more liable to distortions in distance calculations.....	19
Figure 2 - Example of the Earth's surface represented by Clarke 1866 spheroid, in which one degree of longitude at the Equator is equal to 111.321 km, while at 60° latitude it is only 55.802 km [IBM Corporation, 2005]......	25
Figure 3 - Representation of parallels and meridians on the Earth's surface.....	26
Figure 4 – Geodesic Coordinate System.....	27
Figure 5 - Representation of the semi-major (a) and semi-minor (b) axis of an ellipsoid of revolution. ....	28
Figure 6 – Comparing the position of the cylinder in Mercator Projection and UTM Projection. ....	29
Figure 7 - Representation of 6° longitude zones in UTM system.....	29
Figure 8 – UTM Projection of the Earth's surface.....	30
Figure 9 - Version of the International Chronostratigraphic Chart given by The International Commission on Stratigraphy, used in this research [ICS, 2012]. ....	31
Figure 10 - Process of cluster analysis [XU & WUNSCH II, 2005]. ....	32
Figure 11 - Classification of points in dense regions by DBSCAN algorithm. ....	34
Figure 12 - ST-DBSCAN pseudo code [BIRANT & KUT, 2006]......	37
Figure 13 - Spherical triangle solved by the law of Haversines.....	40
Figure 14 - Comparison between WGS84 ellipsoid and a sphere of identical volume. ....	40
Figure 15 - Example of ARFF file used to test the algorithm, where the attributes (index, longitude, latitude, millions of years and bathymetry) are described. ....	46
Figure 16 - Structure of a KML file generated by a test execution of the algorithm. ....	49
Figure 17 - Visualization of the KML file generated by the execution of the algorithm, plotted on Google Earth.....	49
Figure 18 - Example of CSV file generated by the execution of the algorithm.....	50
Figure 19 - Example of output log presented along the execution of the algorithm, where the comparison between each object of the cluster is shown on the screen.....	50
Figure 20 - Region of interest selected to build the dataset <i>noaa_bathymetry</i> . ....	54
Figure 21 - KML file plotted on Google Earth showing the 42 clusters formed using Haversine formula on the full area of <i>noaa_bathymetry</i> dataset.....	57
Figure 22 - KML file plotted on Google Earth showing the 42 clusters formed using Haversine formula on the full area of <i>noaa_bathymetry</i> dataset.....	59
Figure 23 - Sample of clusters from Vincenty Experiments which had a different formation in size and elements from Haversine Experiments. ....	59
Figure 24 - Elements clustered in Haversine Experiments in two groups, while in Vincenty Experiments the distribution of the same points was made into four groups. ....	60
Figure 25 - Clusters formed using Haversine formula on the selected area of filtered <i>noaa_bathymetry</i> dataset. ....	60
Figure 26 - Clusters formed using Vincenty Formula on the selected area of filtered <i>noaa_bathymetry</i> dataset. ....	60

- Figure 27 - Clusters formed using Vincenty formula on the selected area of filtered *noaa\_bathymetry* dataset. .... 61
- Figure 28 - KML file plotted on Google Earth showing clusters formed using Haversine formula on the selected area of filtered *noaa\_bathymetry* dataset. Cluster 1 is identified by yellow color; cluster 2 is identified by blue color. Values are bathymetry of each respective point. .... 61
- Figure 29 -KML file plotted on Google Earth showing clusters formed using Vincenty formula on the selected area of filtered *noaa\_bathymetry* dataset. Cluster 1 is identified by yellow color; cluster 2 is identified by dark blue color; cluster 3 is identified by green color; cluster 4 is identified by light blue color. Values are bathymetry of each respective point. .... 62
- Figure 30 - KML file plotted on Google Earth showing clusters formed using Euclidean formula on the selected area of filtered *noaa\_bathymetry* dataset. Cluster 1 is identified by yellow color; cluster 2 is identified by blue color. Values are bathymetry of each respective point. .... 62
- Figure 31 - KML file plotted on Google Earth showing the 37 clusters formed using Euclidean formula on the full area of *noaa\_bathymetry* dataset..... 65

## LIST OF TABLES

Table 1	Notation and description of Vincenty formula.....	38
Table 2	Epochs of the Geologic Time Scale defined by [GRADSTEIN et al., 2012] and their corresponding intervals (in Ma), implemented on the solution developed in this work.....	41
Table 3	Parameters used for the set of experiments on <i>noaa_bathymetry</i> dataset.....	52
Table 4	List of clusters found in the selected execution using Haversine formula, containing the label of each cluster, for which are listed the elements by their index, followed by the total number of elements.....	52
Table 5	List of clusters found in the selected execution using Vincenty formula, containing the label of each cluster, for which are listed the elements by their index, followed by the total number of elements.....	54
Table 6	Examples of points (P1 and P2) from filtered <i>noaa_bathymetry</i> dataset, the Euclidean, Haversine and Vincenty distances between them and the difference between the results of Haversine x Euclidean, Euclidean x Vincenty and Vincenty x Haversine (in meters).....	60
Table 7	List of clusters found in the selected execution using Euclidean formula, containing the label of each cluster, for which are listed the elements by their index, followed by the total number of elements.....	60
Table 8	Comparison between the results of the executions using the three distance calculation formulas (Haversine, Vincenty and Euclidean), regarding number of clusters, number of points clustered and number of points not clustered.....	63
Table 9	Parameters used for the set of experiments on the filtered <i>muller_bathymetry</i> dataset, maintaining the same values for <i>eps1</i> , <i>eps2</i> , <i>minPts</i> and <i>threshold</i> , and varying among <i>Eocene</i> , <i>Paleocene</i> , <i>Upper Cretaceous</i> and <i>Lower Cretaceous</i> as nominal parameters, and among 40, 60, 70 and 100 as numeric parameters.....	65
Table 10	Results of the executions between the four nominal parameters chosen to test the temporal dimension.....	66
Table 11	Results of the executions between the four numerical parameters chosen to test the temporal dimension.....	68
Table 12	Objects of the clusters formed by the execution of Haversine formula on the filtered <i>muller_bathymetry</i> dataset, using <i>Eocene</i> as <i>eps3</i> parameter.....	78
Table 13	Objects of the clusters formed by the execution of Haversine formula on the filtered <i>muller_bathymetry</i> dataset, using <i>Paleocene</i> as <i>eps3</i> parameter.....	81
Table 14	Objects of the clusters formed by the execution of Haversine formula on the filtered <i>muller_bathymetry</i> dataset, using <i>Upper Cretaceous</i> as <i>eps3</i> parameter.....	83
Table 15	Objects of the clusters formed by the execution of Haversine formula on the filtered <i>muller_bathymetry</i> dataset, using 40 as <i>eps3</i> parameter.....	87

Table 16	Objects of the clusters formed by the execution of Haversine formula on the filtered <i>muller_bathymetry</i> dataset, using 60 as <i>eps3</i> parameter.....	90
Table 17	Objects of the clusters formed by the execution of Haversine formula on the filtered <i>muller_bathymetry</i> dataset, using 70 as <i>eps3</i> parameter.....	92

## ABBREVIATIONS

3D	Three-Dimensional
ARFF	Attribute-Relation File Format
CENPES	Petrobras Research and Development Centre
CSV	Comma-Separated Values
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
E	East
GIS	Geographic Information System
GKD	Geographic Knowledge Discovery
GPS	Global Positioning System
GRS80	Geodetic Reference System 1980
HTML	Hypertext Markup Language
IBGE	Instituto Brasileiro de Geografia e Estatística
KDD	Knowledge Discovery in Databases
Km	Kilometres
KML	Keyhole Markup Language
M	Meters
MY	Millions of years
N	North
NOAA	National Oceanic and Atmospheric Administration
PUCRS	Pontifical Catholic University of Rio Grande do Sul
S	South
SAD69	1969 South American Datum
SIRGAS2000	Sistema de Referência Geocêntrico para a América do Sul 2000
UML	Unified Modeling Language
UTM	Universal Transverse Mercator
W	West
WGS84	World Geodetic System 1984
XML	eXtensible Markup Language





## SUMMARY

1. INTRODUCTION .....	18
1.1 Motivation.....	20
1.2 Objectives .....	21
1.2.1 Main Objective .....	21
1.2.2 Specific Objectives .....	22
1.3 Research Methodology .....	22
1.4 Dissertation Overview .....	22
2. BACKGROUND .....	24
2.1 Coordinate Systems .....	24
2.1.1 Geographic Coordinate System .....	25
2.1.2 Geodesic Coordinate System .....	26
2.1.3 Cartesian Coordinate System.....	28
2.2 Geologic Time Scale.....	30
3. SPATIO-TEMPORAL DATA MINING.....	32
3.1 Cluster Analysis.....	32
3.1.1 Spatio-Temporal Cluster Analysis.....	33
3.1.2 Density-Based Clustering Methods .....	33
4. METHODOLOGY .....	36
4.1 Software.....	36
4.1.1 Weka Data Mining Package.....	36
4.1.2 ProGrid .....	37
4.1.3 Google Earth.....	38
4.2 Method Applied to Handle Spatial Dimension.....	38
4.2.1 Haversine Formula.....	39
4.2.2 Vincenty Formula .....	41
4.3 Method Applied to Handle Temporal Dimension.....	43
5. ST-DBSCAN TO HANDLE PALAEOGEOGRAPHIC DATA .....	46
5.1 Input Files .....	46
5.2 Definition of Parameters .....	47
5.3 Execution .....	47
5.4 Output .....	48
6. EXPERIMENTAL RESULTS.....	52
6.1 Hardware.....	52
6.2 Datasets.....	52

6.3 Experiments .....	54
6.3.1 Experiment 1 - Comparing Haversine, Vincenty and Euclidean distance calculation formulas .....	55
6.3.1.1 Final Considerations about Experiment 1 .....	66
6.3.2 Experiment 2 – Testing Haversine formula on a dataset until 140 MY.....	67
6.3.2.1 Testing Haversine formula on a dataset until 140 MY given an Epoch (name) as parameter.....	68
6.3.2.2 Testing Haversine formula on a dataset until 140 MY given a numerical age as parameter.....	70
6.3.2.3 Final Considerations about Experiment 2 .....	71
7. FINAL CONSIDERATIONS .....	72
7.1 Main Contributions .....	74
7.2 Future Works .....	76
REFERENCES .....	78
APPENDIX A.....	82

# 1. INTRODUCTION

Geographic or georeferenced data are those located on the Earth's surface, and geographically represented. The volume of this type of data has increased considerably over the last few years. This is due to the fast advances in telecommunications and data collection via satellite, remote sensing systems and other environmental monitoring devices [MILLER & HAN, 2009]. This increasing stimulates the demand for knowledge discovery applied to geographical data, thus constituting the Geographical Knowledge Discovery (GKD) field, of which the spatial data mining is an integral step. Geographic data present strong dependency between themselves. This spatial dependency is stated on the first law of geography invoked by [TOBLER, 1970], which states:

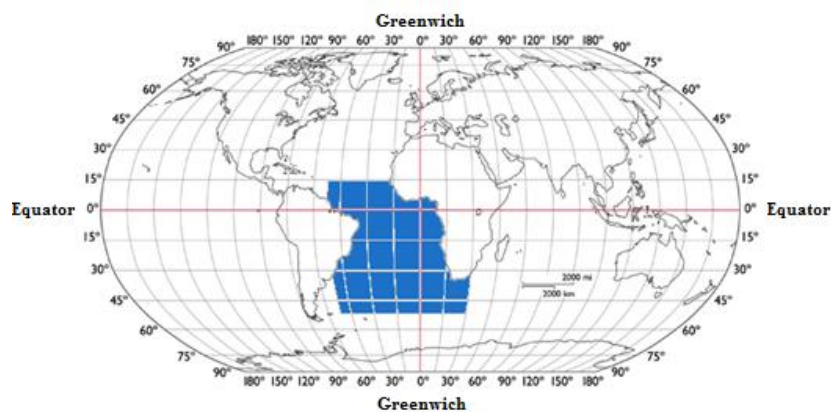
*"Everything is related to everything else, but near things are more related than distant things".*

Considering this, traditional data mining algorithms often have a poor performance when applied to geographic data, arising the need of developing new techniques and tools, or improvements on these traditional algorithms, in order to address this specific kind of data [TAN *et al.*, 2009]. One of the main characteristics of spatial data mining that justifies the need for developing specific techniques is the treatment of distances and measures on the Earth's surface. In geographic data, the dimensions are directly linked, and there should be a special process of calculating distances on the coordinate system [MILLER & HAN, 2009]. Different approaches of reference coordinate systems require a proper method for distance calculations, which takes into account the shape of the Earth and its parameters. Depending on the reference coordinate system adopted, there may occur spatial distortion, a problem inherent to any type of projection, which is the conversion of coordinates in the Earth's surface to coordinates in a plane. Every type of projection causes distortion, and as a result, there are several types of projections that limit a type of distortion, but, on the other hand, increase other types [SAMPLE & IOUP, 2010]. Since these distortions may cause errors in the calculations of distances, there is the need for considering and treating them.

Still, some types of georeferenced spatial data, such as palaeogeographic data, present variation in time. This addition of the temporal dimension increases the complexity of GKD, raising the need for a specific treatment for time. Palaeogeographic data differ from other types of spatio-temporal data in their time approach, presented in millions of years, included in the geological time scale. This approach presents a non-linear structure of time in subdivisions, based on events, not necessarily equidistant from each other. According to [OGG *et al.*, 2008], one of the greatest benefits of the geologic time scale is that it provides to geologists a common and precise language to discuss and reveal the history of Earth. Therefore, it is of interest of a domain expert (geologist, oceanologist, among others) that this temporal dimension is treated according to the geological divisions included in the range of time of interest to the research, which, in our study, ranges from Cretaceous to the present.

Clustering is an unsupervised data mining task present in literature that presents promising algorithms for multidimensional databases with the characteristics of this study. It groups together similar objects, without any prior knowledge about the dataset. Density-based clustering method is approached in this research due to its efficiency when applied on large databases, and also for its ability in identifying groups of arbitrary shape, a characteristic of the database in study. One of the most representative algorithms in spatial data mining, DBSCAN (Density Based Spatial Clustering of Applications with Noise) [ESTER *et al.*, 1996], has a variation for dealing with spatio-temporal data, called ST-DBSCAN [BIRANT & KUT, 2007], which presents the same advantages in discovering groups. However, ST-DBSCAN does not consider two important characteristics inherent to palaeogeographic data, described below.

1. When dealing with areas on the Earth's surface, it is important to take into account the shape of the Earth in order to apply the most suitable geographic distance calculation formula, and, thus, to find well-formed groups of similar objects inside the database. Different approaches of reference system coordinates present different parameters and different ways to calculate distances, and all these characteristics are relevant when seeking for accuracy in the results. Figure 1 illustrates the area addressed in this research in Robinson projection, which uses coordinate systems in latitude / longitude as reference system. This figure shows South Atlantic region of dimensions ranging from -50 to 15 of latitude, and from -70 to 30 of longitude, which is a broad area of the globe, liable to spatial distortions in results of distance calculations.



**Figure 1** - South Atlantic research area presented in Robinson projection, in which is possible to notice how large the region is, being thus more liable to distortions in distance calculations.

The process to compensate this spatial distortion, which tends to accentuate in spatial distance calculations in traditional algorithms, is not found on the density-based clustering for spatio-temporal data mining found in literature, being one of the proposals presented in this work.

2. Besides the challenge presented by the distortion in map projections and distance calculations on Earth's surface, another characteristic is part of the definition of palaeogeographic data, regarding the temporal dimension. Palaeogeographic data are represented in millions of years, according to the geological time scale, which is different from the typical approach used by spatio-temporal data mining algorithms found in the literature (usually in days, months, years or hours, minutes and seconds). When working with data from a time period covering several subdivisions of geologic time, it is of interest to the domain expert that time is represented respecting this approach.

Thus, this work presents the solution developed to address this research problem, both in relation to the spatial distortion in distance calculations and considering the temporal dimension according to the geological time scale, producing better data mining results and facilitating the work of the domain expert.

## 1.1 Motivation

The extraction of interesting and useful patterns in spatial databases is more difficult than data mining on conventional data, due to the complexity of the relations in spatial data [SHEKHAR & HUANG, 2001]. Still, some types of spatial data, such as palaeogeographic data, besides being georeferenced, also present variation in time, being, thus, a type of spatio-temporal data. This addition of the temporal dimension increases the complexity of GKD process, since time is a dimension which must be treated specifically, because it presents different semantics than the space: time is directional and has unique scale and granularity properties [RODDICK & LEES, 2001].

This work takes place within the context of Paleoprospec Project, a project of PUCRS and CENPES (Petrobras Research and Development Centre). The Paleoprospec Project has gathered in a spatio-temporal database information from the remote past on the formation of the ocean floor, since the separation of South America from Africa by continental drift, 140 million years ago. Each million of year can be represented as a set of observation points, moving in geographic space (latitude and longitude) according to time (millions of years) varying properties (pressure, temperature, bathymetry<sup>1</sup>, etc). It is expected to identify on this large spatio-temporal database which is the probability of finding sedimentary deposits rich in organic materials in South Atlantic.

As related works, among the different data mining tasks described in literature that have a version for spatial data, which are described in [KOPERSKI et al., 1996], [ESTER et al., 2000] and [ESTER et al., 2001], there is clustering. In our scenario of research, in which the aim is to identify

---

<sup>1</sup> The term "bathymetry" originally referred to the ocean's depth relative to sea level, although it has come to mean "submarine topography," or the depths and shapes of underwater terrain [NOAA, 2014].

groups of arbitrary formats and similar characteristics over space and time, this seemed to be the most suitable task.

Several spatial clustering algorithms can be found as examples of these tasks, as versions of algorithms for traditional data. These algorithms generally are based on tasks that require the pre-determination of clusters, and as output, present this set of clusters in ellipsoidal or circular shapes. Partitional clustering algorithms based on k-means and k-medoids such as PAM [THEODORIDIS & KOUTROUMBAS, 2006], CLARA [KAUFMAN & ROUSSEEUW, 1990], CLARANS [NG & HAN, 2002], and EM [DEMPSTER *et al.*, 1977], for example, tend to find clusters in such formats, being thus not suitable for treating our specific problem. Among the proposals in the literature, density-based algorithms are useful for discovering groups of similar characteristics identified by dense regions on the database. One of the most representative algorithms using this method is DBSCAN, which has a good efficiency in big spatial databases and finds clusters of arbitrary shapes, without the pre-determination of the number of clusters [ESTER *et al.*, 1996]. There are also other density-based spatial algorithms with these characteristics, such as DBRS [WANG & HAMILTON, 2003] (which also considers the non-spatial attributes in the dataset), DENCLUE [HINNEBURG & KEIM, 1998] or OPTICS [ANKERST *et al.*, 1999]. Among these, DBSCAN is the algorithm with a proposal of adaptation for spatio-temporal data, the ST-DBSCAN [BIRANT & KUT, 2007], which identifies clusters of arbitrary shape, with no need to previously identify the number of groups. It also considers the non-spatial attributes in the clustering process, being thus suitable to be applied on our problem of research. However, despite being the most promising algorithm found in literature to solve the research problem presented, ST-DBSCAN does not consider two important characteristics inherent to palaeogeographic data related to temporal and spatial dimensions, as described in the previous section. This was the motivation for the development of an evolved version of ST-DBSCAN algorithm to deal with palaeogeographic data and its particularities, presented in this dissertation and detailed in the next sections.

## **1.2 Objectives**

### **1.2.1 Main Objective**

The main objective of this work is to develop a solution for clustering algorithms that minimizes the distortions that occur in distance calculations with georeferenced data through coordinate reference system. In addition, this solution aims to address the temporal dimension considering the granularity of geologic time, enabling the domain expert to place the research according to this approach of time. This solution can be applied to spatio-temporal data mining algorithms, enabling an increase in quality of the results from geographic distance calculations, and facilitating the work of the domain expert by adjusting the approach of time according to geological time, typical of palaeogeographic data.

### 1.2.2 Specific Objectives

The specific objectives of this research are:

- To study coordinate systems and spatial distance calculation formulas for data mining algorithms, in order to analyse which are suitable to mitigate the problem of spatial distortion in projections and distance calculations;
- To identify a geologic time scale proposed in the literature and the most suitable method to apply this treatment of the temporal dimension;
- To develop a solution to reduce the problem of spatial distortion from distance calculations in density-based algorithms, which at the same time deals with temporal dimension according to the geological time scale.
- To perform experiments with palaeogeographic data, in order to validate the solution implemented.

### 1.3 Research Methodology

The following are activities performed during the development of this research.

- A study on spatio-temporal data mining and clustering algorithms, in order to choose and implement an algorithm found on literature. Also, the review of the literature to identify relevant aspects about spatial distance calculations and a suitable approach of time to handle the granularity of geological time scale.
- Development of a solution regarding the spatial dimension, since it was found that Euclidean distance calculation used by ST-DBSCAN algorithm is not the most suitable to calculate distances on large areas of the Earth's surface.
- Development of the solution regarding the temporal dimension, through a new approach to handle the granularity of the geologic time scale.
- Execution of a set of experiments to validate the solution implemented.

### 1.4 Dissertation Overview

This dissertation is organized in seven sections, as follows:

- Chapter 2 introduces the major geosciences concepts necessary for a better comprehension of this research. It starts with an overview of coordinate reference systems, pointing out the different approach for the representation of the Earth presented by each one, and closes with an explanation the geologic time scale.

- Chapter 3 describes the basic aspects of spatio-temporal data mining. It starts with an overview of KDD process, situating GKD process in this context. After that, there is an overview about clustering analysis, focusing on density-based clustering method.
- Chapter 4 describes the methodology used during the development of this research in order to implement, adapt and execute the solutions for the problems identified. It is divided into three parts. First, an overview on Weka Data Mining Software, Google Earth and ProGrid converter. Next, the method applied to handle spatial dimension, where a subsection focus on distance calculations, specifically on the choices for this work (Haversine and Inverse Vincenty Formulas). Finally, it is presented the method applied to handle temporal dimension.
- Chapter 5 presents the first results of this research, i.e., the solutions developed in this work on ST-DBSCAN algorithm. It includes details of input files, the definition of parameters, the execution process and the outputs.
- Chapter 6 presents the second set of results of this research. From the experiments performed, two situations are presented, using the solutions developed in this work and defined in Chapter 5. The first set of experiments reports the executions that aim to validate the solution for the spatial dimension, comprising the formulas implemented. The second set of experiments are intended to validate the solution for the temporal dimension using both approaches implemented.
- Chapter 7 summarizes the conclusions of the study which have led to this dissertation. Additionally, it draws the main contributions and gives directions for future works.

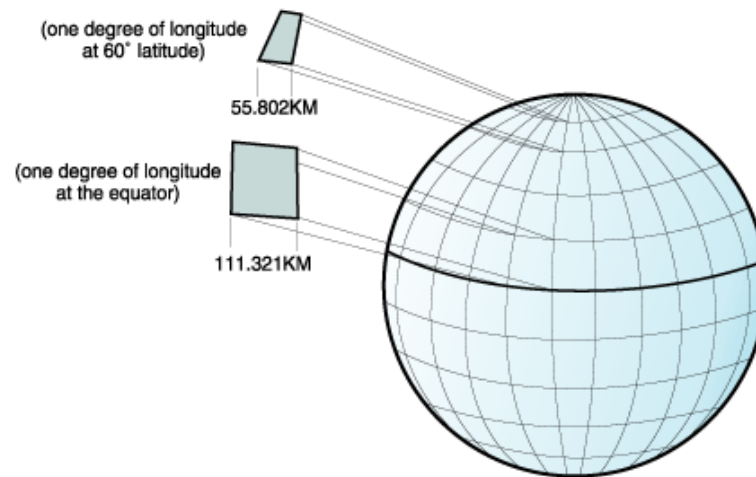


## 2. BACKGROUND

This chapter introduces the major concepts necessary for a better comprehension of this research. It starts with an overview of geographic coordinate reference systems, which is followed by an explanation on the geologic time scale. After the overview on the main characteristics of palaeogeographic data, it is focused on the computational concepts of spatio-temporal data mining, emphasizing the density-based clustering method.

### 2.1 Coordinate Systems

Coordinates are mathematical elements used to determine the position of a point on a surface, being it an ellipsoid, a sphere or a plane. The ellipsoid of revolution, generated by an ellipse rotated around a minor axis, is the geometry that best approximates to the actual shape of the Earth. For representations at very small scales - smaller than 1:5,000,000, the difference between the equatorial radius and the polar radius has an insignificant value, allowing to represent the shape of the Earth, in some applications, as a sphere. However, to represent the land surface more accurately, the ellipsoid of revolution is used by choosing the appropriate model depending on the considered region. Still, it is possible to perform the projection of these coordinates in a plane, a process that causes distortions in several senses. Among all these possible representations of the Earth's surface and the references used for location, it is important to be aware of the peculiarities presented in methods of distance calculation on these surfaces. Because the Earth is divided into imaginary lines named parallels and meridians, it is possible to observe differences in distances between meridians on the globe, and also the distortions that the projection of these imaginary lines on a flat surface can cause. Above and below the Equator, the circles defining the parallels of latitude get gradually smaller until they become a single point at the North and South Poles where the meridians converge. As the meridians converge toward the poles, the distance represented by one degree of longitude decreases to zero. Figure 2 illustrates an example using Clarke 1866 spheroid, in which one degree of longitude at the Equator is equal to 111,321 km, while at 60° latitude it is only 55,802 km. Since degrees of latitude and longitude do not have a standard length, it is necessary to consider the reference system in which geographic data are presented, in order to apply the most appropriate distance formula.



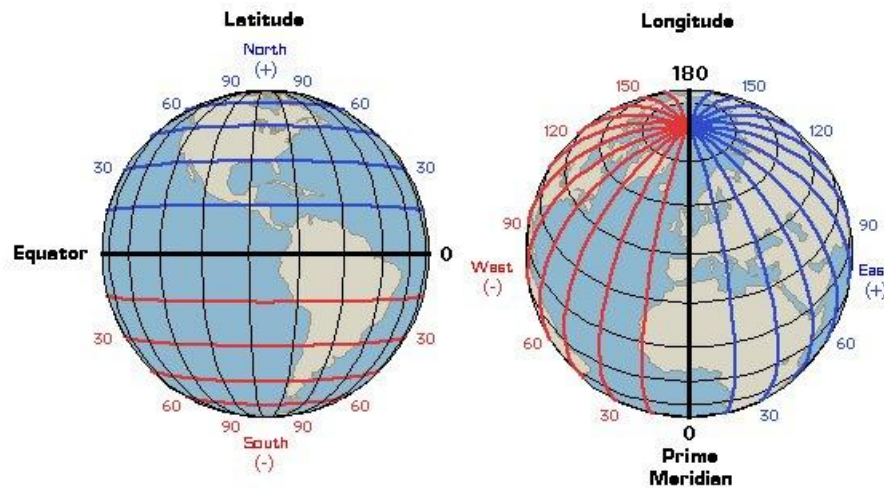
**Figure 2** - Example of the Earth's surface represented by Clarke 1866 spheroid, in which one degree of longitude at the Equator is equal to 111.321 km, while at 60° latitude it is only 55.802 km [IBM Corporation, 2005].

In the next subsections are presented the concepts of systems based on these types of surfaces, in the form of Geographic, Geodesic and Cartesian coordinate systems.

### 2.1.1 Geographic Coordinate System

The Geographic Coordinate System is based on Earth spherical geometry and axis of rotation. The poles are defined as the points of intersection between the Earth axis of rotation and the surface of the sphere [SANTOS, 1989]. This mapping systems comprises imaginary lines drawn on the globe, as shown in Figure 3, described as:

- Parallels: horizontal lines parallel to the Equator, measuring latitude. To the north of the Equator, in the northern hemisphere, the coordinate values increase, varying from 0° to +90°. To the south of the Equator in the southern hemisphere, the measures decrease, ranging from 0° to -90°.
- Meridians: semicircular lines, i.e. 180 degrees, ranging from the North Pole to the South Pole, measuring longitude. Each meridian has an antimeridian, or an opposite meridian forming with it a circle. The Earth is divided into two hemispheres, east and west, by the Greenwich meridian (0°) and its antimeridian (180°). To the east, the coordinate values increase, varying from 0° to 180°. To the west, the values decrease, varying from 0° to -180°.



**Figure 3** - Representation of parallels and meridians on the Earth's surface.

To measure distances between two points given in latitude/longitude considering the spherical shape of the Earth, it is necessary to use a formula which takes into account these factors, in order to decrease distortions on the results.

### 2.1.2 Geodesic Coordinate System

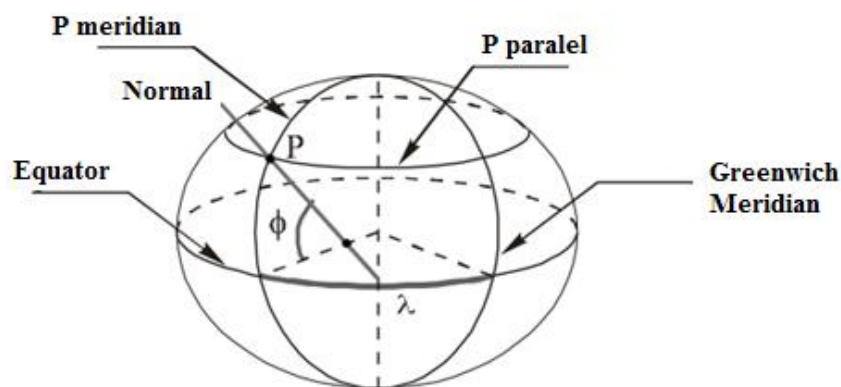
Despite the spherical format be widely used for representation of the globe in many applications that have a small scale, the reference surface that most closely resembles the actual shape of the Earth is an oblate ellipsoid irregular along the line of poles. Thus, the ellipsoid of revolution is a three-dimensional (3D) geometric figure generated by the rotation of a semi-ellipse about its minor axis, which imitates the shape and rotational motion of the Earth. The ellipsoid of best adjustment varies according to the location of the area studied, so there are several theoretical reference surfaces to represent the Earth. Due to this fact, each region tends to adopt a specific *datum*<sup>2</sup>. The different dimensions for various ellipsoids result from variation in precision in geodetic measurements (measurements of the locations of the Earth). Differences also arise because the curvature of the Earth's surface is not uniform, due to irregularities in the gravitational field. Therefore, the ellipsoid that best fits on the Earth's surface depends not only on how the measurements are made, but also where they are done.

In Brazil, until the late 1970s, the International ellipsoid of Hayford and Córrego Alegre-MG was used as the origin of coordinates. Since 1977, SAD69 (1969 South American Datum) was adopted, which presents the vertex Chuá-MG as the origin of coordinates, and the reference ellipsoid as recommended by the International Astronomical Union, approved in 1967 by International Association of Geodesy. With the advent of GPS (Global Positioning System), it has been common to

<sup>2</sup> In geodesy, *geodetic datum* is the term used to describe a set of parameters that constitute the reference of a certain geographic coordinate system, namely, the reference ellipsoid, defined by measures of the semi-major axis and semi-minor axis and its position on the globe [GASPAR, 2005].

use the global planimetric WGS84 datum, whose ellipsoid is adopted for global mapping. But from 2014 on, SIRGAS2000 will be the official datum used in Brazil. The SIRGAS (Geocentric Reference System for South America) was created in October 1993 with the participation of countries of South America, represented by their national agencies, with the main objective to establish a geocentric reference for South America. The adoption of SIRGAS follows a current trend, taking into account the potential of GPS and facilities for users, because with this geocentric system, the coordinates obtained from GPS can be applied directly on the cartographic surveys, avoiding the need for change and integration between the two reference systems [DALAZOANA & FREITAS, 2002].

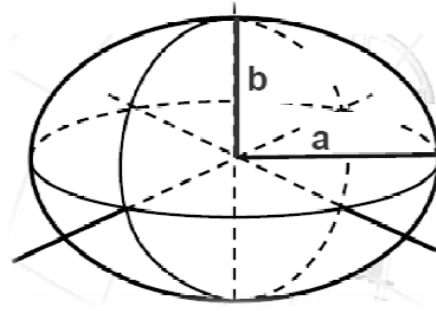
The geodetic coordinate system follows the same logic based on imaginary lines drawn on the surface, called parallels and meridians, from which the latitude and longitude of a point are measured. According to Figure 4, the definitions of geodetic coordinates of any point P on the surface of the ellipsoid are:



**Figure 4** – Geodesic Coordinate System

- Latitude: it is defined geodetic latitude of a place as being the angle between the normal of the place and the plane of the Equator. The latitude ranges from  $0^\circ$  to  $90^\circ$  being considered negative in the southern hemisphere.
- Longitude: it is defined longitude of a location as the dihedral angle between the Greenwich meridian plane and the plane of the place's meridian, or the angular distance from the Equator since the origin meridian (Greenwich) until this meridian. The length varies from  $0^\circ$  to  $180^\circ$  being considered negative in the western hemisphere.

As shown in Figure 5, the shape of an ellipsoid of revolution is determined by the parameters of the ellipse that generates the ellipsoid when it is rotated on its minor-axis.



**Figure 5** - Representation of the semi-major ( $a$ ) and semi-minor ( $b$ ) axis of an ellipsoid of revolution.

The semi-major axis of an ellipse,  $a$ , is identified as the equatorial radius of the ellipsoid; the semi-minor axis of the ellipse,  $b$ , is identified as the polar distance (from the center). These two measures specify the shape of the ellipsoid, but in practice geodesic publications classify reference ellipsoids by their semi-major axis and the inverse flattening,  $1/f$ . The flattening,  $f$ , is the measure of how much the axis of symmetry is compressed relative to the equatorial radius:

$$f = a - b / a$$

The geodesic system SIRGAS2000, used in Brazil, adopts the ellipsoid of revolution GRS80 (Geodetic Reference System 1980), whose semi-major axis value is 6,378,137m and flattening  $1/298.25722101$ .

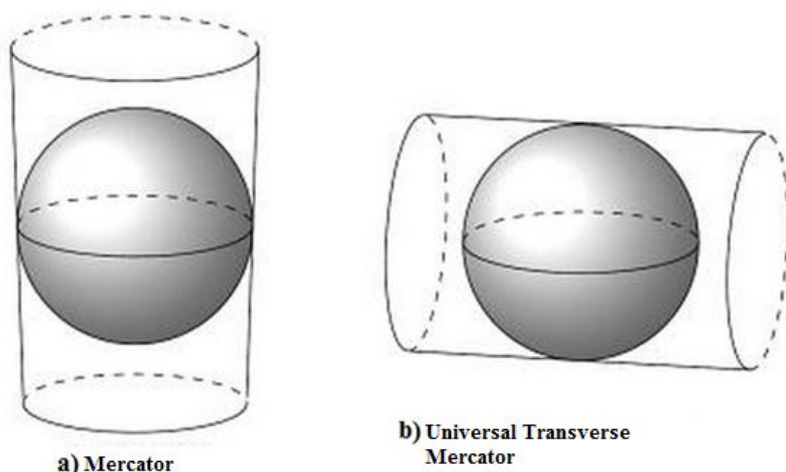
### 2.1.3 Cartesian Coordinate System

Because the geographic coordinate system, although unequivocally finding points on the surface of the ellipsoid, has proved impractical for working with plane maps, Cartesian plane coordinate systems associated to map projections were established. As a system used for map projections, it is susceptible to distortions, in distance, direction, shape or area. According to [MARTONNE, 1953], every geographic map is a deformation of the surface and the figures which are observed. Among these figures, the most important are the spherical coordinates: meridians and parallels. The problem of projections consists in finding for the coordinate lines a principle by which the deformation is as small as possible. In Cartesian coordinate system, the coordinates are presented in metric units, instead of degrees.

In this context, Mercator projection is a cylindrical conformal projection that touches the globe on the Equator, preserving the angles and shapes of the map. By these properties, it is traditionally used in mapping and cartography. Being a cylindrical projection, the east-west deformation increases as increases latitude. Furthermore, the space between adjacent parallels

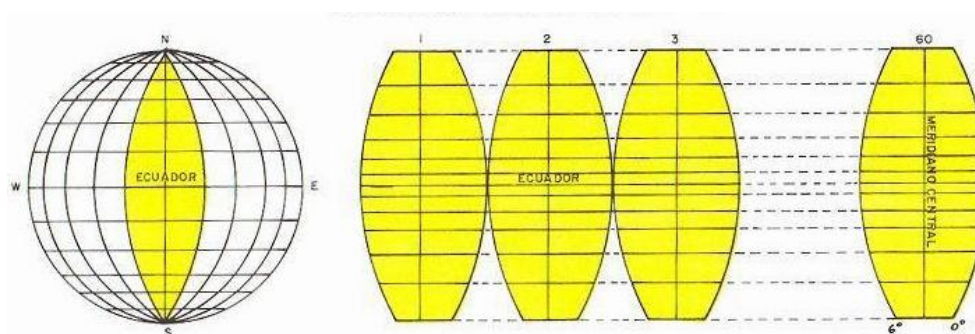
increases with latitude, being the east-west distortion accompanied by a north-south deformation. Thus, the scale of the projection increases also with the latitude, becoming infinite at the poles.

UTM (Universal Transverse Mercator) projection is a particular case of Mercator Projection, in which the cylinder is placed in transverse position, according to Figure 6.



**Figure 6** – Comparing the position of the cylinder in Mercator Projection and UTM Projection.

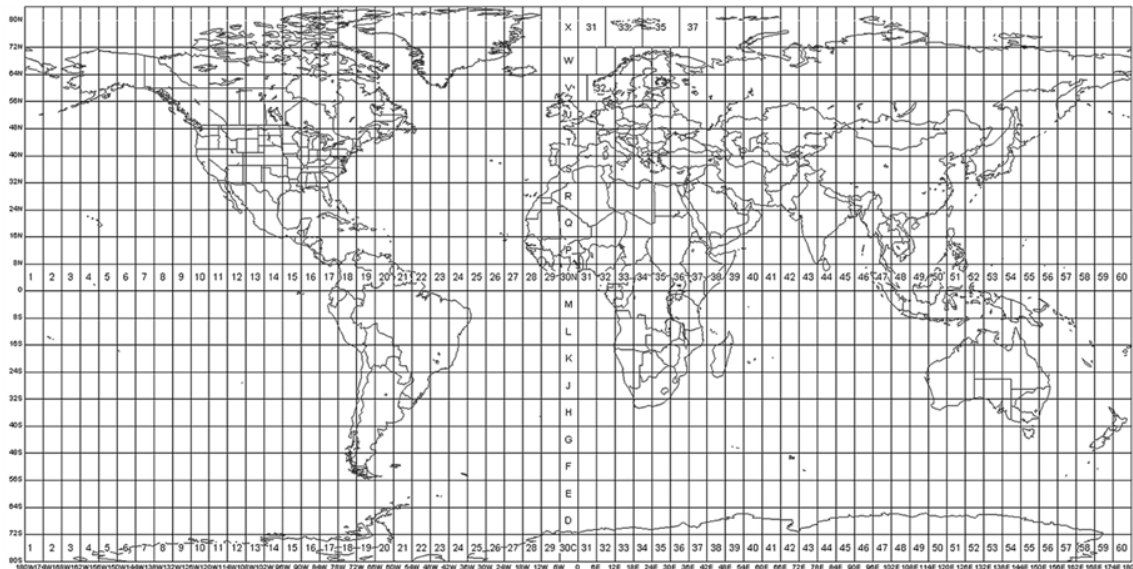
Because the cylinder axis is perpendicular to the Earth polar axis, this is a conformal projection, which keeps the angles and shapes of the small areas, occurring deformation only in distances. Therefore, it is a projection often used in maps of small areas, such as aerial and satellite imagery [SAMPLE & IOUP, 2010]. As shown in Figure 7, the terrestrial ellipsoid is divided into 60 zones of  $6^\circ$  of longitude, numbered from Greenwich's antemeridian, following from west to east until it finds its point of origin. This subdivision in zones of a determined longitude is done in order not to exceed certain acceptable limits of deformation [LIBAUT, 1962].



**Figure 7** - Representation of  $6^\circ$  longitude zones in UTM system.

There are only two axes:  $E(x)$ , representing latitude, for east-west coordinates (abscissa), and  $N(y)$ , that represents longitude, for north-south coordinates (ordinate). In the southern hemisphere, the distances from the axis  $N(y)$  start at 10,000,000 on the Equator and decrease southward to 0; the axis

E(x) starts on 500,000 increasing to the East and decreasing to the West. In the Northern Hemisphere, the coordinates axis E(x) behave in the same way, while axis N(y) has its origin in the Equator and increases to the north. Figure 8 shows the UTM projection of the Earth's surface with its division of 60 zones.



**Figure 8** – UTM Projection of the Earth's surface.

Each UTM coordinate corresponds to a geographic coordinate, which is curved, transformed by mathematical procedures in plane coordinates. As the cylinder touches the Equator, the map has less distortion in this region, but great distortion at high latitudes. In addition, each zone of this system presents a unique system of distances, with its origin defined by the intersection of the central meridian of the zone and the Equator. Thus, the calculation of distances in large areas covering multiple UTM zones becomes a non-trivial task, being necessary to use reprojections between coordinate systems and treating distortions to improve the quality of calculations.

## 2.2 Geologic Time Scale

The geologic time scale is a non-linear and hierarchical division of time, divided into eons, eras, periods, epochs and ages. According to [OGG *et al.*, 2004], the geologic time scale has three components:

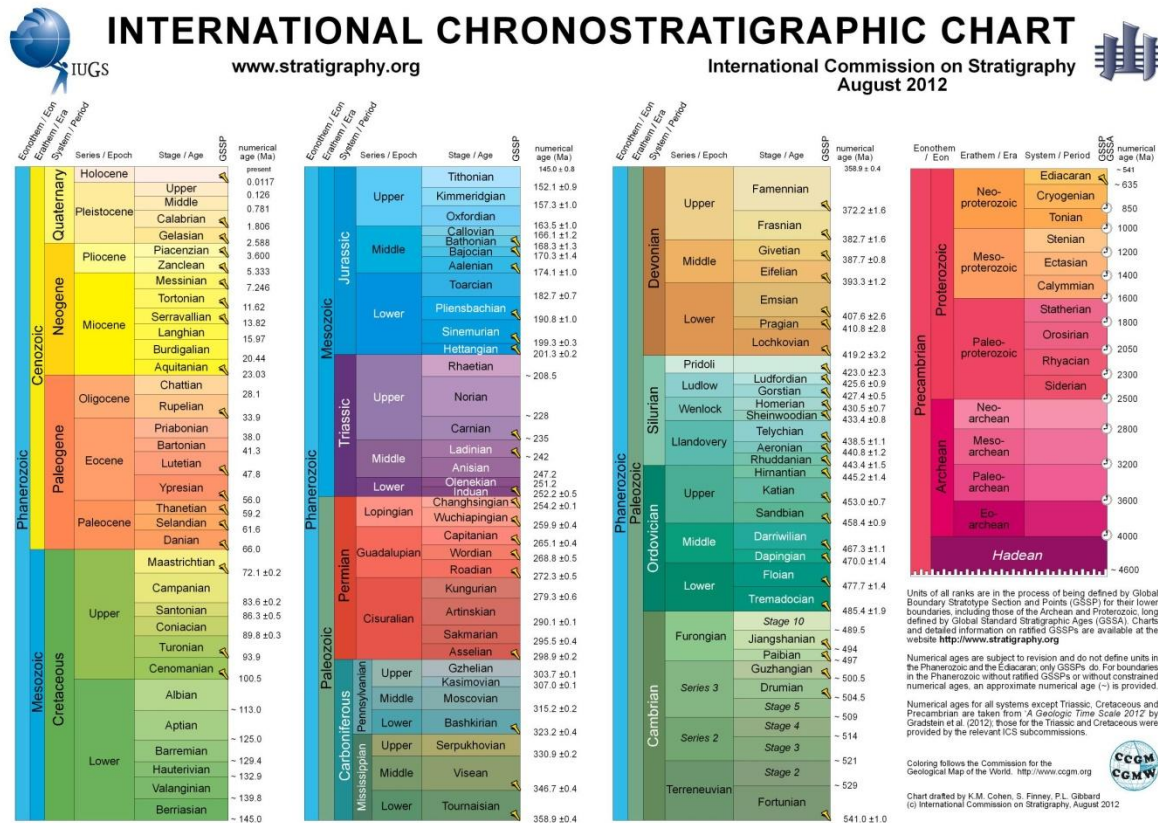
- The stratigraphic division and correlation in international global registry of rocks;
- The means of measuring the linear time or elapsed durations from the rock record;
- The methods to effectively combine the two scales.

There are several methods for reading the geological record, as an interpretation of faults and other structures, looking for evidence of uplift or erosion, construction of deductions about the



environments in which the sediments were deposited, reconstruction of the original conditions of the rocks that were deformed and metamorphosed [GROTZINGER *et al.*, 2007]. The stratigraphy is the most used method to interpret geological events from records in the rocks.

As knowledge about the history of Earth is constantly increasing, simultaneously the geological time scale receives modifications [OGG *et al.*, 2008]. Figure 9 shows the version of the geologic timescale used in this research.



**Figure 9 -** Version of the International Chronostratigraphic Chart given by The International Commission on Stratigraphy, used in this research [ICS, 2012].

Through geologic time scale, it is possible to situate the geologic processes that shape the Earth’s surface and give structure to its interior. On this time scale, major events occurred, such as continental drift, sediment build up, oil formation, continents, oceans and mountains moving in great distances. One of the essential works of the geologist is to understand the patterns and rates of these movements [GROTZINGER *et al.*, 2007], situating the research according to this scale.

In the next chapter, the computational concepts used in this research are presented, since cluster analysis until the density-based method chosen to be applied on the spatio-temporal data of this research.



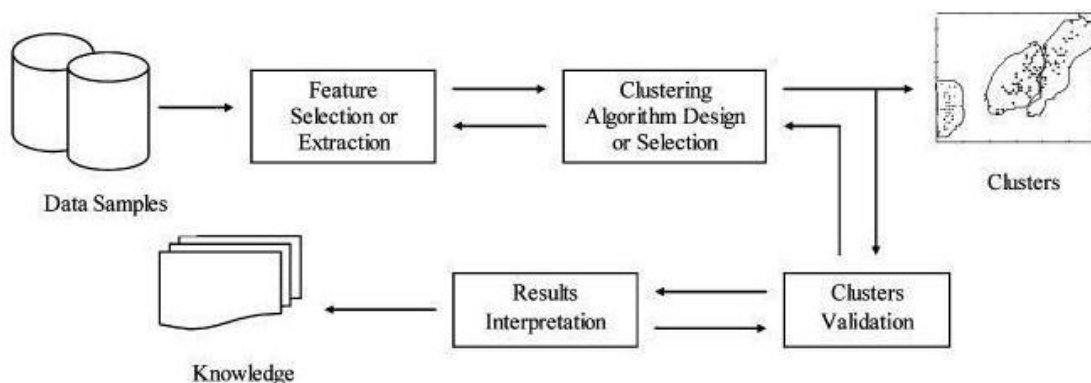
### 3. SPATIO-TEMPORAL DATA MINING

This chapter describes the basic aspects of spatio-temporal data mining. It presents an overview on cluster analysis, focusing on spatio-temporal cluster analysis and density-based method.

#### 3.1 Cluster Analysis

Cluster analysis is the process of partitioning a dataset into subsets, each subset being a group, so that objects within a group are similar to each other but different from objects in other groups [HAN & KAMBER, 2006]. Clustering is defined as an unsupervised classification method [CHEN *et al.*, 1996], applied when there is no class to be predicted, but it is wanted to divide the instances into natural groups [WITTEN & FRANK, 2000]. Thus, it does not rely on predefined classes and class-labelled training examples, being a form of learning by observation, rather than learning by examples. In data mining, efforts have focused on finding methods for efficient and effective cluster analysis in large databases. Active themes of research focus on the scalability of clustering methods, the effectiveness of methods for clustering complex shapes and types of data, high-dimensional clustering techniques, and methods for clustering mixed numerical and categorical data in large databases [HAN & KAMBER, 2001].

The process of cluster analysis presented by [XU & WUNSCH II, 2005], which can be seen in Figure 10, is similar to the process of knowledge discovery defined by [FAYYAD *et al.* 1996].



**Figure 10** - Process of cluster analysis [XU & WUNSCH II, 2005].

According to [XU & WUNSCH II, 2005], the first step in this process is crucial to get quality on results. Pre-processing techniques are applied in order to adequate the data so that they can serve as input to clustering algorithms. The second step defines similarity metrics that will be used to calculate the similarity between objects and the choice of clustering method, depending on the types of data and the problem being addressed. The cluster algorithm chosen will be run on the dataset, dividing objects into groups.

In [HAN & KAMBER, 2001], cluster methods are divided into five main categories: partitional, hierarchical, grid-based, model-based and density-based methods.

The density-based method seems to be the most appropriate to identify groups of arbitrary shape, which includes noise and outliers values [HAN, 2012], as is the case of a database comprising spatio-temporal palaeogeographic data. The next subsection provides details on this method, preceded by a description of spatio-temporal clustering analysis.

### 3.1.1 Spatio-Temporal Cluster Analysis

Despite being more complex, spatial data mining is based on the same strategy of non-spatial data mining, being the characteristic aspects of a set of objects identified and classified according to these sets of objects [NG & HAN 1994]. Spatio-temporal data is composed by a set of spatio-temporal sequences, in which each element of the sequence is represented by its spatial and temporal attributes  $(x_1, x_2, \dots, x_n, t)$ , where  $x_i$ ,  $1 \leq i \leq n$ , is a spatial attribute, and  $t$  is a temporal attribute. In this repository of our scenario of research, we aim to find groups of data objects with similar characteristics, in clusters presented in different formats (oval, elongated, concave, etc). In addition, we have to consider that in geographic spatio-temporal data mining, uncertainty is involved at each step, from data pre-processing through data conceptualization until the extraction of knowledge [SHUA *et al.*, 2008]. These characteristics (uncertainty, high-dimensionality, big database, possibility of clusters with different formats, aim of finding groups of similar characteristics) were the features by which we have chosen the spatio-temporal data mining task. This task can be classified as: (i) Segmentation, (ii) Dependency Analysis, (iii) Deviation and outlier analysis, (iv) Trend Discovery and (v) Generalization and Characterization. Among them, cluster analysis of spatio-temporal data, so called regionalization, is used to analyse the spatial variability of one or more physical variables and to decompose a large complex area into smaller homogeneous regions [KALYANI, 2012].

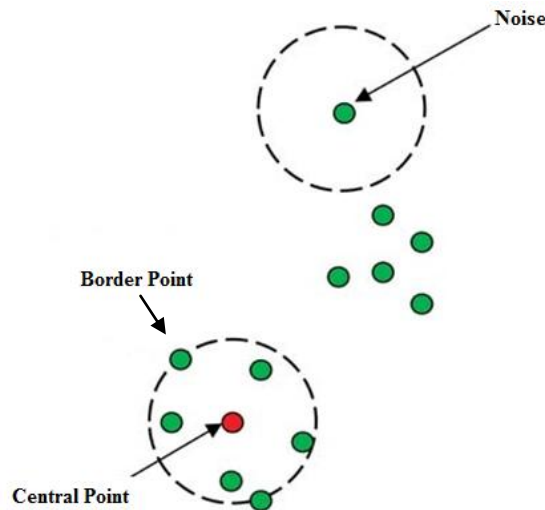
### 3.1.2 Density-Based Clustering Methods

The strategy of density-based clustering method is to find groups of arbitrary shape as dense regions in the database, separated by sparse regions [HAN, 2012]. Thus, a group is formed based on the density of its neighbourhood in the multidimensional space.

From the clustering algorithms based on density found in the literature, DBSCAN stands for its efficiency when applied to large spatial databases [TAN *et al.*, 2009; HAN, 2012; ESTER *et al.*, 1996]. This algorithm is able to find groups in arbitrary formats, since the format of a cluster in a spatial database can be spherical, linear, elongated, among others [ESTER *et al.*, 1996]. For this, it uses a centre-based approach to identify areas of density. Its input parameters are a dataset  $X$ , the neighbourhood size  $Eps$  and a minimum number of points  $MinPts$  required to form a cluster. With

these input parameters, all data from the dataset  $X$  are visited and the points densely connected form clusters, being the neighbourhood of each point denoted by  $NEps(xi) = \{xj \in X \mid d(xi, xj) \leq Eps\}$ .

After identified dense regions, the points are classified according to their position, as shown in Figure 11. A point can be classified as central point (if  $|NEps(xi)| \geq MinPts$ ), boundary point (if it belongs to the neighbourhood of a central point  $xj$ ,  $xi \in NEps(xj)$ ), or noise.



**Figure 11** - Classification of points in dense regions by DBSCAN algorithm.

It is possible to recognize each cluster because each one has a typical density of points that is considerably greater than the density outside of the cluster. Thus, the density within the areas of noise is less than the density in any cluster [ESTER *et al.*, 2001].

Of all the density-based spatial algorithms studied, DBSCAN has a variation to handle spatio-temporal data, called ST-DBSCAN [BIRANT & KUT, 2007]. Despite having the same logic regarding the identification of central points and their density-reachable neighbourhood  $Eps$ , above a minimum value of  $MinPts$ , ST-DBSCAN has a second  $Eps$  parameter, thus presenting parameters  $Eps1$  and  $Eps2$ .  $Eps1$  is used for spatial values to measure the closeness between two geographical points, while  $Eps2$  is used for non-spatial values. To address the temporal dimension, the spatio-temporal data are first filtered retaining only temporal neighbours - objects whose dimension is on consecutive days in the same year or the same day in consecutive years - and their corresponding spatial values. There is no treatment of time in addition to this, which considers days and years in the moment of the filtering. Furthermore, there is no treatment of spatial distortions in the results of geographic distance calculations, since ST-DBSCAN algorithm uses Euclidean formula to calculate the closeness of two geographical points on Earth's surface, given in latitude/longitude.



## 4. METHODOLOGY

This section describes the methodology used during the development of this research in order to implement, evolve and execute the solution to the problem of research. It is divided into three parts. First, an overview on Weka Data Mining Package, ProGrid and Google Earth. After, the methodology applied to handle spatial dimension, where a subsection focus on distance calculations, specifically on the choices for this work (Haversine [SINNOT, 1984] and Inverse Vincenty [VINCENTY, 1975] Formulas). Finally, it is presented the method applied to handle the temporal dimension.

### 4.1 Software

Weka Data Mining Package was used to implement ST-DBSCAN algorithm as a module, using Java programming language. Google Earth was used to visualize the results of the experiments. ProGrid was used to convert data from decimal degrees to UTM coordinate system. A brief description of these software is presented below.

#### 4.1.1 Weka Data Mining Package

The Weka workbench is a collection of state-of-the-art machine learning algorithms and data pre-processing tools. Weka was developed at the University of Waikato in New Zealand, and the name stands for *Waikato Environment for Knowledge Analysis*. As well as a wide variety of learning algorithms, it includes a wide range of pre-processing tools. This diverse and comprehensive toolkit is accessed through a common interface [WITTEN & FRANK, 2000]. Because it is an open-source software project, it is vastly used in academic environment, also due to the fact that it is possible to implement code using its package, available online for download.

After defining ST-DBSCAN as the selected algorithm to be used in this research, we started to study its implementation from the pseudo code provided by [BIRANT & KUT, 2006], as can be seen in Figure 12.

```

Algorithm ST_DBSCAN (D, Eps1, Eps2, MinPts, Δε)
// Inputs:
// D={o1, o2, ..., on} Set of objects
// Eps1 : Maximum geographical coordinate (spatial) distance value.
// Eps2 : Maximum non-spatial distance value.
// MinPts : Minimum number of points within Eps1 and Eps2 distance.
// Δε : Threshold value to be included in a cluster.
// Output:
// C={C1, C2, ... Ck} Set of clusters

Cluster_Label = 0
For i=1 to n // (i)
  If oi is not in a cluster Then // (ii)
    X=Retrieve_Neighbors(oi, Eps1, Eps2) // (iii)

    If |X| < MinPts Then
      Mark oi as noise // (iv)
    Else //construct a new cluster (v)
      Cluster_Label = Cluster_Label + 1

      For j=1 to |X|
        Mark all objects in X with current Cluster_Label
      End For

      Push(all objects in X) // (vi)

      While not IsEpmty()
        CurrentObj = Pop()
        Y= Retrieve_Neighbors(CurrentObj, Eps1, Eps2)

        If |Y| >= MinPts Then
          For All objects o in Y // (vii)
            If (o is not marked as noise or it is not in a cluster) and
              |Cluster_Avg() - o.Value| <= Δε Then
              Mark o with current Cluster_Label
              Push(o)
            End If
          End For
        End If
      End While
    End If
  End For
End Algorithm

```

Figure 12 - ST-DBSCAN pseudo code [BIRANT & KUT, 2006].

In order to implement the code, we downloaded the Weka package. Thus, ST-DBSCAN code was implemented inside the package as a module. For the implementation, Java language was used, in NetBeans development tool [NetBeans, 2013].

#### 4.1.2 ProGrid

ProGrid [IBGE, 2008] is a free software launched by the Brazilian Institute of Geography and Statistics (Instituto Brasileiro e Geografia e Estatística - IBGE) for coordinate conversion. This program works as a desktop application, runs on Microsoft Windows environment, and was developed to allow the coordinate transformation between the official reference systems used in Brazil: Córrego Alegre, SAD69 and SIRGAS2000.

The types of coordinates used associated with each reference are:

- Córrego Alegre (1961): latitude / longitude and UTM;
- Córrego Alegre (1970/1972): latitude / longitude and UTM;
- SAD69 Classic Network: latitude / longitude and UTM;
- SAD69/96 Classic Network: latitude / longitude and UTM;

- SAD69 Doppler / GPS Technique: latitude / longitude / geometric height, Cartesian (X, Y, Z) and UTM.
- SIRGAS2000: latitude / longitude / geometric height, Cartesian (X, Y, Z) and UTM.

Using this program, it is possible to convert a range of coordinates simultaneously by listing the coordinates in an input file. The results can also be presented on an output file produced by the execution, as well as to be shown on the screen.

### 4.1.3 Google Earth

Google Earth is a virtual globe, map and geographical information software whose function is to present a 3D model of the globe, built from a mosaic of satellite images obtained from multiple sources, aerial images (photographed from aircraft) and 3D GIS. The basic version is free and available for download [Google Earth, 2013]. Using the Google Earth user interface, it is possible to create, read and analyse KML (Keyhole Markup Language) files. KML is a file format used to display geographic data on Google Earth and Google Maps. Just as web browsers display Hypertext Markup Language (HTML) files, Earth browsers such as Google Earth display KML files. It uses a tag-based structure with nested based elements and attributes encoded using eXtensible Markup Language (XML). XML is a widely used standard for creating structured documents. Like the HTML files used in web pages, tags are enclosed in  $\langle \rangle$  and each opening  $\langle \text{tag} \rangle$  is matched by a  $\langle / \text{tag} \rangle$  indicating the end of the tagged section. Custom icons can be added to placemarks within the KML file. The guides for the development of a KML file are available online on the documentation [KML Documentation, 2013].

## 4.2 Method Applied to Handle Spatial Dimension

As it was stated previously in this work, ST-DBSCAN uses Euclidean formula to calculate distances on Earth's surface using latitude / longitude coordinates. Since the shape of the Earth is an ellipsoid, also treated in some cases as a sphere, any calculation of distance between two points projected on its surface should consider the spatial distortion that different results may generate, due to inadequate formulas.

The Euclidean distance, used by ST-DBSCAN and most of spatio-temporal data mining algorithms, is used to measure the similarity (or dissimilarity) between objects. Its formula is given by

$$d(i,j)=\sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2}$$

, where  $i=(x_{i1}, x_{i2}, \dots, x_{in})$  and  $j=(x_{j1}, x_{j2}, \dots, x_{jn})$  are two  $n$ -dimensional data objects. This formula is used to calculate distances in the Euclidean space. Euclidean  $n$ -space, sometimes called

Cartesian space or simply  $n$ -space, is the space of all  $n$ -tuples of real numbers,  $(x_1, x_2, \dots, x_n)$ . Although this distance can be generalized to higher dimensions, it specifically applies to spaces of two and three dimensions. Thus, using Euclidian distance to calculate points in latitude/longitude, being them points projected on Earth's surface, which is an ellipsoid or a sphere, will skew the results causing spatial distortion.

Also there is the fact that, according to [ESTER *et al.*, 1996], the shape of the neighbourhood is determined by the choice of the distance function for two points  $p$  and  $q$ , denoted by  $dist(p,q)$ . For instance, when using Manhattan distance in 2D space, the shape of the neighbourhood is rectangular. Given the fact that in our research it is more likely to find clusters of arbitrary shape, the choice of a distance function is an important factor to take into account. Aiming at attenuating this problem, we sought for appropriated formulas to calculate distances in latitude/longitude, and thus be applied in our research. The formulas we found and implemented on the algorithm are described in the following sections.

#### 4.2.1 Haversine Formula

This equation, important in navigation, uses longitude and latitude to determine great-circle distances between two points, considering the shape of the Earth as a sphere [SINNOTT, 1884]. The Haversine formula is given by:

$$a = \sin^2(\Delta\varphi/2) + \cos(\varphi_1).\cos(\varphi_2).\sin^2(\Delta\lambda/2)$$

$$c = 2.\text{atan2}(\sqrt{a}, \sqrt{1-a})$$

$$d = R.c$$

, where  $\varphi$  is latitude,  $\lambda$  is longitude,  $R$  is Earth radius (mean radius = 6,371km). In this case, angles need to be in radians to pass to trigonometric functions. This formula is based on the law of Haversines, which can be explained as follows:

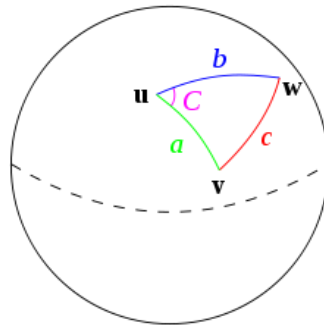
Given a unit sphere, a "triangle" on the surface of the sphere is defined by the great circles connecting three points  $u$ ,  $v$ , and  $w$  on the sphere. If the lengths of these three sides are  $a$  (from  $u$  to  $v$ ),  $b$  (from  $u$  to  $w$ ), and  $c$  (from  $v$  to  $w$ ), and the angle of the corner opposite  $c$  is  $C$ , then the law of haversines states:

$$\text{haversin}(c) = \text{haversin}(a - b) + \sin(a) \sin(b) \text{haversin}(C)$$

Since this is a unit sphere, the lengths  $a$ ,  $b$ , and  $c$  are simply equal to the angles (in radians) subtended by those sides from the centre of the sphere (for a non-unit sphere, each of these arc lengths

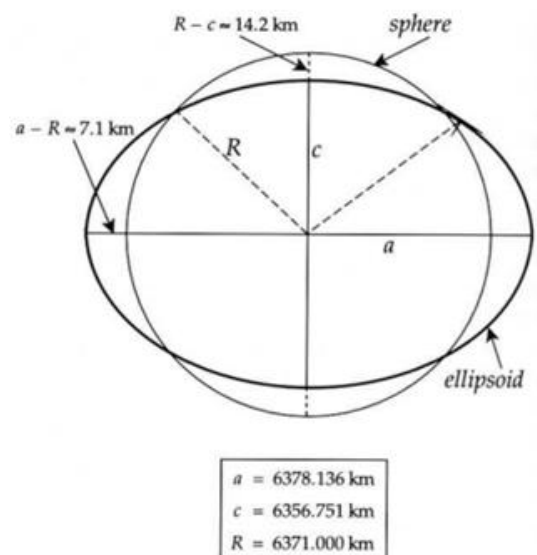


is equal to its central angle multiplied by the radius of the sphere). Figure 13 shows the application of this law on a sphere.



**Figure 13** - Spherical triangle solved by the law of Haversines.

This way, we have chosen Haversine formula to be used as method for calculation of distances in ST-DBSCAN algorithm, instead of the Euclidean formula. However, while studying Haversine formula and the possibility to use it in order to calculate distances in our scenario of research, we noticed that this formula only assumes the shape of the Earth as a sphere. But, actually, the shape of the Earth is not a perfect sphere, but an oblate ellipsoid of revolution, also called an oblate spheroid. This is an ellipse rotated about its shorter axis. Figure 14 illustrates the comparison between the shape of one of the several ellipsoids currently in use, the global ellipsoid WSG84, and a sphere, making it possible to notice the difference between these two shapes.



**Figure 14** - Comparison between WGS84 ellipsoid and a sphere of identical volume.

Errors from assuming spherical geometry might be up to 0.55% crossing the Equator, though generally below 0.3%, depending on latitude and direction of travel. Considering this, we identified in the literature Vincenty formula, which is the more accurate option to calculate distances on ellipsoids, described in the next section.

### 4.2.2 Vincenty Formula

Vincenty formula is a very precise method for calculating distances based on semi-major axis and flattening parameters of the ellipsoid of revolution. There are differences in this case regarding the reference ellipsoid in use, being this the reason why the ellipsoid must be chosen first. This solution, based on series [RAINSFORD, 1955], has two methods, being the first (direct) method intended to compute the location of a point which is a given distance and azimuth (direction) from another point. The second (inverse) method computes the geographical distance and azimuth between two given points, being the most appropriate to our case.

Thus, according to the formula given by [VINCENTY, 1975], it is defined the notation presented in Table 1.

**Table 1 - Notation and description of Vincenty formula**

Notation	Description
$\varphi 1, \varphi 2$	latitude of two points
$L1, L2$	longitude of two points
$L = L2 - L1$	difference in longitude of two points;
$U1 = \arctan[(1 - f) \tan \varphi 1],$ $U2 = \arctan[(1 - f) \tan \varphi 2]$	reduced latitude (latitude on the auxiliary sphere)
$\lambda 1, \lambda 2$	longitude of the points on the auxiliary sphere
$\Sigma$	arc length between points on the auxiliary sphere
$A$	length of semi-major axis of the ellipsoid (radius at Equator)
$f$	flattening of the ellipsoid
$A$	azimuth at the Equator
$\alpha 1, \alpha 2$	forward azimuths at the points
$b = (1 - f) a$	length of semi-minor axis of the ellipsoid (radius at the poles)
$S$	ellipsoidal distance between the two points

, given the coordinates of the two points ( $\phi 1, L1$ ) and ( $\phi 2, L2$ ), the inverse problem finds the azimuths  $\alpha 1, \alpha 2$  and the ellipsoidal distance  $S$ . Thus, it is possible to calculate  $U 1, U 2$  and  $L$ , setting the initial value of  $\lambda = L$ , and then iteratively evaluate the following equations until  $\lambda$  converges.

$$\sin\sigma = \sqrt{[(\cos U_2 \cdot \sin\lambda)^2 + (\cos U_1 \cdot \sin U_2 - \sin U_1 \cdot \cos U_2 \cdot \cos\lambda)^2]}$$

$$\cos\sigma = \sin U_1 \cdot \sin U_2 + \cos U_1 \cdot \cos U_2 \cdot \cos\lambda$$

$$\sigma = \text{atan2}(\sin\sigma, \cos\sigma)$$

$$\sin\alpha = \cos U_1 \cdot \cos U_2 \cdot \sin\lambda / \sin\sigma$$

$$\cos^2\alpha = 1 - \sin^2\alpha \text{ (trig identity; §6)}$$

$$\cos 2\sigma_m = \cos\sigma - 2 \cdot \sin U_1 \cdot \sin U_2 / \cos^2\alpha$$

$$C = f/16 \cdot \cos^2\alpha \cdot [4 + f \cdot (4 - 3 \cdot \cos^2\alpha)]$$

$$\lambda' = L + (1 - C) \cdot f \cdot \sin\alpha \cdot \{\sigma + C \cdot \sin\sigma \cdot [\cos 2\sigma_m + C \cdot \cos\sigma \cdot (-1 + 2 \cdot \cos^2 2\sigma_m)]\}$$

When  $\lambda$  has converged to the desired degree of accuracy (10–12 corresponds to approximately 0.06mm), evaluate the following:

$$u^2 = \cos^2\alpha \cdot (a^2 - b^2) / b^2$$

$$A = 1 + u^2 / 16384 \cdot \{4096 + u^2 \cdot [-768 + u^2 \cdot (320 - 175 \cdot u^2)]\}$$

$$B = u^2 / 1024 \cdot \{256 + u^2 \cdot [-128 + u^2 \cdot (74 - 47 \cdot u^2)]\}$$

$$\Delta\sigma = B \cdot \sin\sigma \cdot \{\cos 2\sigma_m + B/4 \cdot [\cos\sigma \cdot (-1 + 2 \cdot \cos^2 2\sigma_m) - B/6 \cdot \cos 2\sigma_m \cdot (-3 + 4 \cdot \sin^2\sigma) \cdot (-3 + 4 \cdot \cos^2 2\sigma_m)]\}$$

$$s = b \cdot A \cdot (\sigma - \Delta\sigma)$$

$$\alpha_1 = \text{atan2}(\cos U_2 \cdot \sin\lambda, \cos U_1 \cdot \sin U_2 - \sin U_1 \cdot \cos U_2 \cdot \cos\lambda)$$

$$\alpha_2 = \text{atan2}(\cos U_1 \cdot \sin\lambda, -\sin U_1 \cdot \cos U_2 + \cos U_1 \cdot \sin U_2 \cdot \cos\lambda)$$

A "geodetic curve" is the result that means how to get from one point on an ellipsoid to another. The ellipsoidal distance is the distance, in meters, between the two points along the surface of the ellipsoid. The azimuth is the direction of travel from the starting point to the ending point, while the reverse azimuth is the direction back from the endpoint. To apply this formula, it is required to identify the correct reference ellipsoid for the area studied. This is due to the necessity to specify the length of semi-major axis of the ellipsoid (radius at Equator). In our case, the reference ellipsoid used is SIRGAS2000.

Thus, Vicenty formula could be used instead of Euclidean distance in the scenario of research, being even more effective than Haversine formula. However, since this formula is based on the measures of the axis of the ellipsoid, it cannot be applied to data whose temporal dimension is presented in millions of years ago, since the shape of the ellipsoid changed along with the evolution of the planet in millions of years. This way, this formula is intended to be applied only in data from the present time, in geologic terms.

### **4.3 Method Applied to Handle Temporal Dimension**

The Geologic Time Scale is a structure which represents all the millions of years since the formation of the planet divided according to geological events [TEIXEIRA *et al.*, 2009]. This structure is divided into Eons, Periods, Epochs and Ages. According to discussions with the experts of Paleoprospec, the Ages of the scale are seen as regional boundaries for geological events, such as Maastrichtian, for example, which ranges from 66.0 MY to 72.1 MY [GRADSTEIN *et al.*, 2012], and it was delimited in 1849 based on sites on rock strata of the Chalk Group close to the Dutch city of Maastricht. These strata are now classified as the Maastricht Formation - both formation and stage derive their names from the city, which occurs only in the Netherlands and Belgium.

In addition to the fact that the divisions of Ages are very specific to particular locations, there is also the need to consider that there are many proposals for Geologic Time Scales in the literature, which diverge in relation to the millions of years of divisions between ages. Thus, considering these points and also given the fact that Epoch is a broader period of interest for the domain expert, we decided for this approach to be implemented on our solution, addressing the Epochs of the Geologic Time Scale. The structure was implemented according to the definitions of [GRADSTEIN *et al.*, 2012], since this is the version of Geologic Time Scale used in Paleoprospec Project. These Epochs and their respective intervals (in MY) are presented in Table 2.

**Table 2 - Epochs of the Geologic Time Scale defined by [GRADSTEIN *et al.*, 2012] and their corresponding intervals (in MY), implemented on the solution developed in this work.**

<b>EPOCH</b>	<b>INTERVAL (in MY)</b>
Holocene	from 0 to 0.0117
Pleistocene	from 0.018 to 2.588
Pliocene	from 2.589 to 5.333
Miocene	from 5.334 to 23.03
Oligocene	from 23.04 to 33.9
Eocene	from 34.0 to 56.0
Paleocene	from 56.1 to 66.0
Upper Cretaceous	from 66.1 to 100.5
Lower Cretaceous	from 100.6 to 145.0
Upper Jurassic	from 145.1 to 163.5
Middle Jurassic	from 163.6 to 174.1
Lower Jurassic	from 174.2 to 201.3
Upper Triassic	from 201.4 to 235
Middle Triassic	from 235.1 to 247.2
Lower Triassic	from 247.3 to 252.2
Lopongian	from 252.3 to 259.9
Guadalupian	from 260 to 272.3
Cisuralian	from 272.4 to 298.9
Upper Pennsylvanian	from 299 to 307.0
Middle Pennsylvanian	from 307.1 to 315.2
Lower Pennsylvanian	from 315.3 to 323.2
Upper Mississippian	from 323.3 to 330.9
Middle Mississippian	from 331 to 346.7
Lower Mississippian	from 346.8 to 358.9
Upper Devonian	from 359 to 382.7
Middle Devonian	from 382.8 to 393.3
Lower Devonian	from 393.4 to 419.2
Pridoli Silurian	from 419.3 to 423
Ludlow Silurian	from 423.1 to 427.4
Wenlock Silurian	from 427.5 to 433.4
Llandovery Silurian	from 433.5 to 443.4
Upper Ordovician	from 443.5 to 458.4
Middle Ordovician	from 458.5 to 470.0
Lower Ordovician	from 470.1 to 485.4
Furongian Cambrian	from 485.5 to 497
Series 3 Cambrian	from 497.1 to 509
Series 2 Cambrian	from 509.1 to 521
Terreneuvian Cambrian	from 521.1 to 541.0
Neoproterozoic Proterozoic	from 541.1 to 1000
Mesoproterozoic Proterozoic	from 1000.1 to 1600
Paleoproterozoic Proterozoic	from 1600.1 to 2500
Neoproterozoic Archean	from 2500.1 to 2800
Meosoarchean Archean	from 2800.1 to 3200
Paleoarchean Archean	from 3200.1 to 3600
Eoarchean Archean	from 3600.1 to 4000
Hadean	from 4000.1 to 4600

Thus, the solution developed was based on the definition of a third parameter *eps3*, to handle temporal dimension. According to this definition, the user has the option of declaring the *eps3* as a numerical age, as well as a specific Epoch, from which it is wanted to split the objects of the database into similar groups. This approach allows the algorithm to handle two different types of data, being possible for the user to inform a parameter of type numeric, as well as a string. Both the possibilities decrease domain knowledge needed to inform this parameter, making it more comfortable to the user.

Although this solution does not make use of a distance formula to measure the temporal similarity, as happens with other *eps* parameters in density-based algorithms, this filter groups the objects by considering them similar if they occur at the same epoch in geological time. The term *eps3* facilitates the understanding of the three approaches considered to cluster objects: *eps1*, which is the maximum geographic distance that two points should have in order to be considered belonging to a same cluster; *eps2*, which is the maximum distance between two values for a given property, such as bathymetry, for points to be considered similar enough to belong to the same group, and *eps3*, the epoch by which objects are considered similar.

This chapter showed the methodology used to develop the solution to be implemented on ST-DBSCAN in order to address the problems of spatial and temporal dimensions in paleogeographical data. The next chapter shows the solution developed in details.

## 5. ST-DBSCAN TO HANDLE PALAEOGEOGRAPHIC DATA

This chapter describes the first results of this research. It presents the solutions implemented on ST-DBSCAN algorithm in order to mine palaeogeographic data. In Chapter 6, we used these solutions to obtain experimental results.

As mentioned in the Introduction, the foremost objective of this study is to develop a solution for clustering algorithms to minimize the distortions that occur in distance calculations with georeferenced data through geographic coordinate system, as well as to address the temporal dimension considering the granularity of geologic time, enabling the domain expert to place the research according to this approach.

### 5.1 Input Files

The file read by the program can be a CSV (comma-separated values) format, or a text file which respects this structure, being only necessary to inform its extension (.txt, .csv, .arff, etc). Since the algorithm ST-DBSCAN was implemented inside Weka, it is enabled to read the structure of an ARFF (Attribute-Relation File Format) file, an ASCII text file that describes a list of instances sharing a set of attributes. This is a standard way of representing datasets that consist of independent, unordered instances and do not involve relationships among instances [WITTEN & FRANK, 2000], used by Weka. ARFF files have two distinct sections, being the first the Header information, which is followed by Data information. The Header of the ARFF file contains the name of the relation, a list of the attributes (the columns in the data), and their types. Figure 15 shows an example of ARFF file used to test the algorithm.

```
1 @relation teste
2
3 %NOAA's dataset
4 %index, longitude, latitude, millions of years, bathymety
5
6 @attribute INDEX numeric
7 @attribute LON numeric
8 @attribute LAT numeric
9 @attribute MA numeric
10 @attribute BAT numeric
11
12 @data
13 1,-38,-25,0,-3978
14 2,-38,-24.9,0,-3940
15 3,-38,-24.8,0,-3911
16 4,-38,-24.7,0,-3888
17 5,-38,-24.6,0,-3869
18 6,-38,-24.5,0,-3849
19 7,-38,-24.4,0,-3870
20 8,-38,-24.3,0,-3923
21 9,-38,-24.2,0,-3934
22 10,-38,-24.1,0,-3757
23 11,-38,-24,0,-3639
24 12,-38,-23.9,0,-3545
25 13,-38,-23.8,0,-3574
26 14,-38,-23.7,0,-3609
27 15,-38,-23.6,0,-3619
28 16,-38,-23.5,0,-3640
29 17,-38,-23.4,0,-3642
30 18,-38,-23.3,0,-3694
31 19,-38,-23.2,0,-3726
32 20,-38,-23.1,0,-3692
```

**Figure 15** - Example of ARFF file used to test the algorithm, where the attributes (index, longitude, latitude, millions of years and bathymetry) are described.

The structure of the input file follows the sequence: index, longitude, latitude, millions of years and bathymetry. Lines that begin with a % are comments. The @RELATION, @ATTRIBUTE and @DATA declarations are case insensitive.

## 5.2 Definition of Parameters

Parameters in traditional ST-DBSCAN are defined according to the heuristic presented by [ESTER *et al.*, 1996], which determines the parameters *eps* and *minPts* of the "thinnest" cluster in the database, to measure spatial closeness according to the result of the Euclidean formula.

In the solution developed in this research, the spatial calculations return a value in meters between one point and the other. The distance between the properties will vary according to what kind of property is being handled at the time, and the value of comparison to consider them similar or not can vary from property to property (which is the same case of the *threshold*). Finally, the temporal dimension depends on the epoch of interest of the research. For all these reasons, the determination of the parameters in the improvements made in ST-DBSCAN algorithm depends more on the specific case of research, than on a heuristic to determine the parameters in general. Thus, the spatial parameter *eps1*, as well as the non-spatial parameter *eps2*, the *threshold*, and the temporal parameter *eps3* are determined by the domain expert.

## 5.3 Execution

At the execution time, the neighbourhood of each object from the database is retrieved by the function *Retrieve\_Neighbours*, which forms a list of neighbours of a point according to the definition of similarity. In this moment, similarity is measured at the same time in three approaches: geographical distance, similar values of the properties, and temporal similarity according to the epoch in the Geologic Time Scale. Thus, a call of *Retrieve\_Neighbours* returns the objects that have a distance less than *eps1* and *eps2*, and is in the range of *eps3*, forming the *eps-neighbourhood*.

If the total number of returned points in *eps-neighbourhood* is smaller than the minimum of points *minPts* determined by the user, objects are assigned as noise, and returned to the database. Otherwise, if the size of the neighbourhood defined by the three types of similarity between the objects is greater than or equals to the minimum number of points *minPts*, the object is assigned as belonging to a cluster, of which the label is also assigned to the object. Thus, objects initially assigned as *NONE* (not of any type) are assigned as *CLUSTER* or *NOISE*. After, function *Retrieve\_Neighbours* is called to apply the *threshold* test on the noise objects. If the difference between the average value of the cluster and the new coming value is smaller than the *threshold*, then the point is placed to the current cluster.



The geographical distance can be measured by three possible formulas: Euclidean, Haversine and Vincenty formulas. When selecting the Haversine distance, the radius of the Earth is specified as 6,372.8 kilometres, and the conversion of the values  $lat1$ ,  $lon1$ ,  $lat2$ ,  $lon2$ , which in our database are expressed in decimal degrees, to Radians. The great circle distance between two points is expressed in the same units as R. The intermediate result  $c$  is the great circle distance in radians.

When using the Vincenty distance calculation formula, it is necessary to inform the parameters for the ellipsoid being used. These parameters are the values of semi-major axis  $a$ , semi-minor axis  $b$ , and *inverse flattening*. Since the parameters of the ellipsoid changed along the formation of the ellipsoid (the shape of the Earth changed from millions of years ago until now), it is advisable to apply this formula only to data from the present. The result will be the distance, in meters, between one point and another on the ellipsoid used. The intermediate result  $c$  is the angular distance in radians, and  $a$  is the square of half the chord length between the points.

In the non spatial calculation, the properties of each object are compared, by applying the Euclidean distance calculation, which in this case is the absolute value of their numerical difference, given by:

$$\sqrt{(p_x - q_x)^2} = |p_x - q_x|$$

In order to determine temporal neighbours, the value of MY given for each object on the dataset is compared to the *eps3* value given by the user - or the correspondent string, depending on the type of value informed. This filter allows temporal neighbours to be retrieved by their occurrence in the same epoch, according to the geologic time scale (despite not using for this a distance function).

This way, non-spatial, spatial and temporal characteristics of the objects are used by the time of execution of the algorithm, in the data mining step of the KDD process, to group together similar objects.

## 5.4 Output

The output of the execution is a list containing the set of clusters (and their labels), with the index of each object assigned to each cluster. Besides the list of clusters presented on the screen, two files are generated as output for further use to evaluate the results.

The first of them is a KML file, generated to allow the visualization of the groups formed directly on Google Earth. KML is an XML file format for representing data within Google Earth, Google Maps, and other geographical visualization tools. Figure 16 shows an example of KML structure from a test execution of the algorithm.

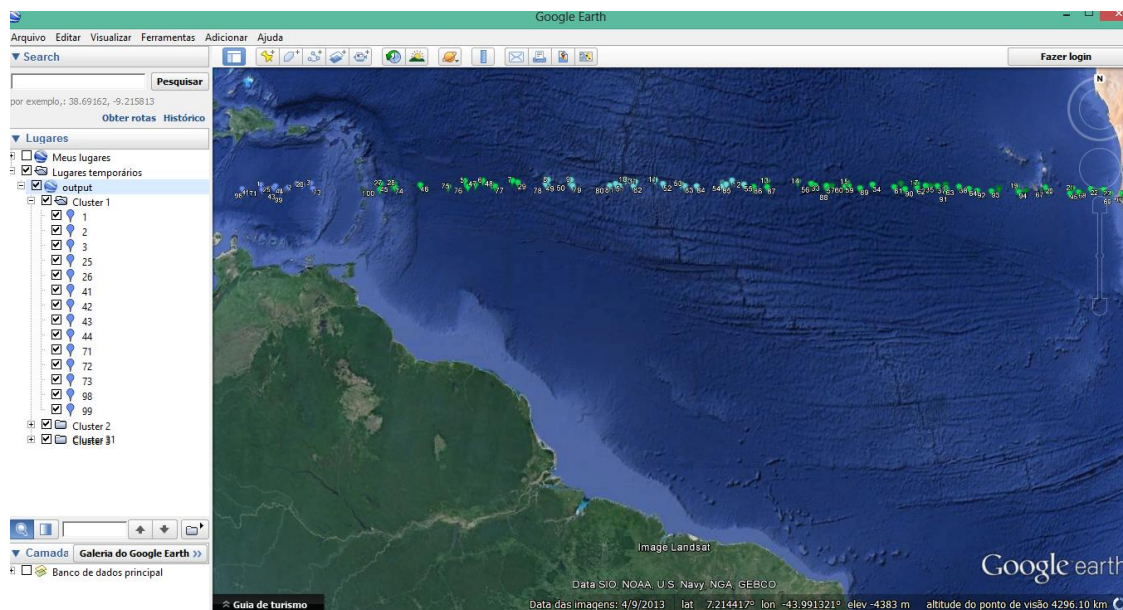
```

1 <?xml version="1.0" encoding="UTF-8" ?><?xml xmlns="http://www.opengis.net/kml/2.2" ?>
2 <Document><name>output</name>
3 <Folder><name>Cluster 1</visibility>1</visibility></name>
4 <Placemark><name>1</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
5 </href></Icon></Style></Style><Point><coordinates>-68.3,15.0,0</coordinates></Point></Placemark>
6 <Placemark><name>2</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
7 </href></Icon></Style></Style><Point><coordinates>-65.5,15.0,0</coordinates></Point></Placemark>
8 <Placemark><name>3</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
9 </href></Icon></Style></Style><Point><coordinates>-64.5,15.0,0</coordinates></Point></Placemark>
10 <Placemark><name>25</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
11 </href></Icon></Style></Style><Point><coordinates>-67.1,14.9,0</coordinates></Point></Placemark>
12 <Placemark><name>26</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
13 </href></Icon></Style></Style><Point><coordinates>-65.1,14.9,0</coordinates></Point></Placemark>
14 <Placemark><name>41</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
15 </href></Icon></Style></Style><Point><coordinates>-69.8,14.8,0</coordinates></Point></Placemark>
16 <Placemark><name>42</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
17 </href></Icon></Style></Style><Point><coordinates>-67.6,14.8,0</coordinates></Point></Placemark>
18 <Placemark><name>43</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
19 </href></Icon></Style></Style><Point><coordinates>-67.2,14.8,0</coordinates></Point></Placemark>
20 <Placemark><name>44</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
21 </href></Icon></Style></Style><Point><coordinates>-66.1,14.8,0</coordinates></Point></Placemark>
22 <Placemark><name>71</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
23 </href></Icon></Style></Style><Point><coordinates>-68.1,14.7,0</coordinates></Point></Placemark>
24 <Placemark><name>72</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
25 </href></Icon></Style></Style><Point><coordinates>-66.3,14.7,0</coordinates></Point></Placemark>
26 <Placemark><name>73</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
27 </href></Icon></Style></Style><Point><coordinates>-64.3,14.7,0</coordinates></Point></Placemark>
28 <Placemark><name>98</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
29 </href></Icon></Style></Style><Point><coordinates>-69.3,14.6,0</coordinates></Point></Placemark>
30 <Placemark><name>99</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/blue.png
31 </href></Icon></Style></Style><Point><coordinates>-66.6,14.6,0</coordinates></Point></Placemark>
32 </Folder>
33 <Folder><name>Cluster 2</visibility>1</visibility></name>
34 <Placemark><name>4</name><visibility>1</visibility><Style><IconStyle><Icon><href>http://www.google.com/mapfiles/ms/micons/green.png
35 </href></Icon></Style></Style><Point><coordinates>-66.6,14.6,0</coordinates></Point></Placemark>
36 </Folder>
37 </Document>

```

**Figure 16** - Structure of a KML file generated by a test execution of the algorithm.

When opened on the visualization software, each object of the cluster is represented by a custom icon, being the icons of one cluster different from the icons used to represent objects in the other clusters, according to the example showed by Figure 17. The visualization allowed by plotting this file is useful to easier identify similar regions geographically.



**Figure 17** - Visualization of the KML file generated by the execution of the algorithm, plotted on Google Earth.

The second file is a CSV format, in which each line contains the label of the cluster, followed by the index of the instance assigned to that cluster, along with its longitude, latitude and epoch, as it is shown in Figure 18.

```

1 1:1:-68.3:15.0
2 1:2:-65.5:15.0
3 1:3:-64.5:15.0
4 1:25:-67.1:14.9
5 1:26:-65.1:14.9
6 1:41:-69.8:14.8
7 1:42:-67.6:14.8
8 1:43:-67.2:14.8
9 1:44:-66.1:14.8
10 1:71:-68.1:14.7
11 1:72:-66.3:14.7
12 1:73:-64.3:14.7
13 1:98:-69.3:14.6
14 1:99:-66.6:14.6
15 2:4:-59.3:15.0
16 2:5:-55.0:15.0
17 2:6:-54.0:15.0
18 2:7:-52.4:15.0
19 2:13:-39.3:15.0
20 2:14:-37.7:15.0
21 2:15:-35.1:15.0
22 2:16:-32.1:15.0
23 2:17:-31.2:15.0
24 2:18:-29.5:15.0
25 2:19:-25.0:15.0
26 2:20:-23.0:15.0
27 2:21:-21.0:15.0
28 2:22:-19.6:15.0
29 2:23:-18.5:15.0
30 2:24:-16.2:15.0
31 2:27:-59.9:14.9
32 2:28:-59.1:14.9
33 2:29:-52.1:14.9
34 2:32:-57.0:14.6

```

**Figure 18** - Example of CSV file generated by the execution of the algorithm.

In this format, every instance of every cluster is listed, in a format which can be further used to interpret the results, as well as to be applied on new algorithms and tools available in data mining software such as Weka.

Besides these two output files, a log can be seen on the screen, reporting the comparison between all the objects of the dataset. This comparison is presented following the structure:

*Point1;Point2;SpatialDistance;NoSpatialDistance*

, in which *Point1* is the first point, *Point2* is the second point, *SpatialDistance* is the result of the distance calculation between *Point1* and *Point2*, and *NoSpatialDistance* is the result of the calculation between the non-spatial values of *Point1* and *Point2*. Figure 19 shows an example of the output log presented on the screen along the execution.

```

Saída - ST-DBSCAN (run) | Usos
1:2;11131.884502145049;38.0
1:3;22263.769004290098;67.0
1:4;33395.65850643515;90.0
1:5;44527.53800857984;109.0
1:6;55659.4225107256;129.0
1:7;66791.30701287065;108.0
1:8;77923.19151501499;55.0
1:9;89055.07601716004;44.0
1:10;100186.96051980544;221.0
1:11;111318.84502145048;339.0
1:12;122450.72962359554;439.0
1:13;133582.6140257406;404.0

```

**Figure 19** - Example of output log presented along the execution of the algorithm, where the comparison between each object of the cluster is shown on the screen.

This output log allows further analysis of the results and can be also used on tools such as Excel, in order to compare distance results between points, executions of different formulas and parameters, etc.

This chapter showed the solution implemented on ST-DBSCAN in order to address the problems of spatial and temporal dimensions in palaeogeographical data. It includes details of input files, the definition of parameters, the execution process and the outputs. Chapter 6 shows the results obtained from the algorithm executions.

## 6. EXPERIMENTAL RESULTS

This chapter presents the set of results of this research regarding the experiments run. It starts with the description of the resources used for the execution of the experiments (hardware and datasets for input), followed by the description of each set of executions.

### 6.1 Hardware

The experiments were executed in two machines with the following configuration:

- One desktop having processor Intel(R) Core(TM) i7-2600, 12.0 GB RAM memory and Linux Ubuntu Operating System;
- One laptop having processor Intel(R) Core(TM) i7-3517U, 6.0 GB RAM memory and Windows 8 Operating System.

Both machines used NetBeans to run the experiments. The laptop was used to run experiments testing different distances, while the desktop was used to test the mining process with temporal dimension included (due to the bigger size of the dataset and to the better memory management feature offered by Linux).

### 6.2 Datasets

Two datasets were used to run the set of experiments, as follows:

*muller\_bathymetry*: this dataset was obtained by selecting a region of interest and running the program *Deriva.m*, a Matlab code which uses models of the theory on continental drift to generate a mapping of points and their positions in the past. The chosen region covers the area of oceanic platform next to the Campos Basin, in the Brazilian coast, ranging from -18 to -22 latitude, and -33 to -36 longitude. This region was chosen due to the fact that the Campos Basin is the largest oil province in Brazil, accounting for over 80% of national oil production and also has the largest proven reserves already identified and classified in Brazil. From the area selected, the model proposed by [NÜRNBERG & MÜLLER, 1991] and the respective Euler poles were used to run the program *Deriva.m* and generate the mapping of the points of this region to the past, until 140 MY. After generated the points, bathymetric data from the work of [MÜLLER *et al.*, 2008] were added to the correspondent points in the correspondent MY, constructing the dataset. The final result was a dataset containing 115,946 instances, of the points derived by the region chosen moving in time and presenting valid bathymetry. Columns are presented in the following structure:

INDEX	LON_DRIFT	LAT_DRIFT	MY	LON_ORIGIN	LAT_ORIGIN	BAT
-------	-----------	-----------	----	------------	------------	-----

, being:

INDEX: the index of the object;

LON\_DRIFT: spatial dimension in longitude of the object in the given MY according to the continental drift, in decimal degrees.

LAT\_DRIFT: spatial dimension in latitude of the object in the given MY according to the continental drift, in decimal degrees.

MY: temporal dimension in millions of years;

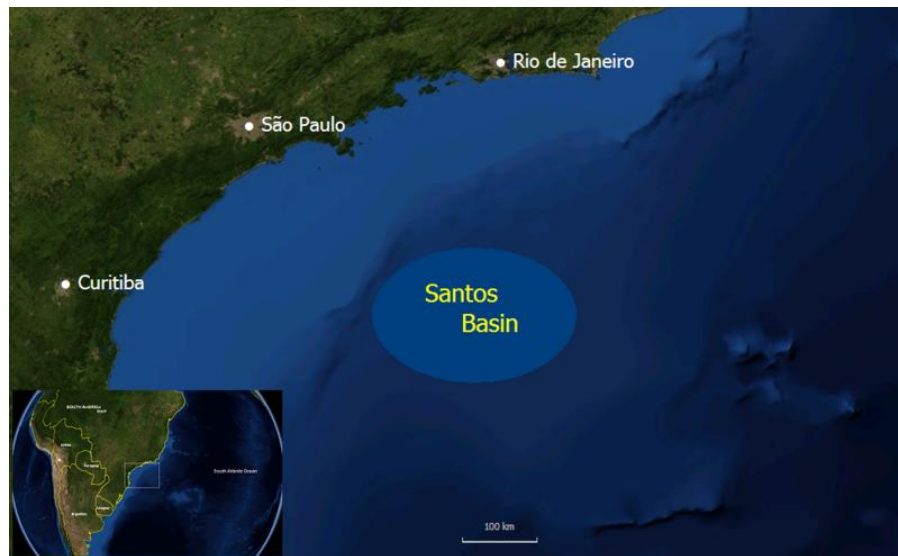
LON\_ORIGIN: spatial dimension in longitude of the object in the present (0 MY), in decimal degrees.

LAT\_ORIGIN: spatial dimension in latitude of the object in the present (0 MY), in decimal degrees.

BAT: the non-spatial property (in this case, bathymetry).

The region selected goes further from the limits of the basin, in order to cover the area of the islands next, and also because the bathymetric data from [MÜLLER *et al.*, 2008] have a lack of values for the continental coast, while these data is available for the ocean crust.

- *noaa\_bathymetry*: this dataset was extracted from National Geophysical Datacenter of NOAA (National Oceanic and Atmospheric Administration) [NOAA, 2013], with 1776 instances, aiming to obtain bathymetric data from the continental crust of Santos Basin, in the Brazilian Coast, as shown in Figure 20. This is a region of interest for research since in this basin are located oil fields in production and large reserves to be exploited.



**Figure 20** - Region of interest selected to build the dataset *noaa\_bathymetry*.

The temporal dimension in this dataset is situated only on the present (0 MY). It is used to test the Vincenty formula, given the fact that this formula uses parameters of semi-major and semi-minor axis of the ellipsoid, values which changed over time (millions of years) during the formation of the ellipsoid as it is today. The spatial dimension is presented in latitude and longitude, ranging from -20 to -25 latitude, and -38 to -42 longitude, in decimal degrees. Columns follow the structure below:

INDEX	LON	LAT	MY	BAT
-------	-----	-----	----	-----

, being:

INDEX: the index of the object;

LON: spatial dimension in longitude of the object in decimal degrees;

LAT: spatial dimension in latitude of the object in decimal degrees;

MY: temporal dimension in millions of years (in this case, 0).

BAT: the non-spatial property (in this case, bathymetry).

### 6.3 Experiments

The set of experiments performed on the datasets described and their results selected to be on this dissertation are presented in the next subsections.

### 6.3.1 Experiment 1 - Comparing Haversine, Vincenty and Euclidean distance calculation formulas

The first set of experiments was carried in order to analyse the results generated by clustering the objects from the dataset by Haversine formula. Then, the algorithm was executed on the same dataset using the same parameters from the previous execution, changing the distance calculation to the Vincenty formula. After, the coordinates were converted from decimal degrees to UTM in order to test the dataset using the Euclidean formula. Finally, the results were analysed.

For this set of experiments, *noaa\_bathymetry* dataset was used, being the temporal dimension of the instances only from 0 MY, given the fact that Vincenty distance is applied on data from the present.

The first set of experiments was carried using as parameters values ranging from 10,000 to 50,000 for *eps1*, from 10 to 100 for *eps2*, 10 to 50 for the *minPts*, 10 to 125 for the *threshold*, and *eps3* was maintained as *Holocene*, since this is the present epoch. From the set of results analysed, one was chosen given the formation of the clusters, distribution of its elements and coherence of the parameters with the region and characteristics presented.

Table 3 presents the parameter of the chosen experiment, set according to the domain expert and considering the size of the dataset.

**Table 3 - Parameters used for the set of experiments on *noaa\_bathymetry* dataset**

<i>eps1</i>	<i>eps2</i>	<i>eps3</i>	<i>minPts</i>	<i>threshold</i>
50000	100	Holocene	10	125

For the region approached, a variation of 100m in bathymetry and 50,000m in geographic location between points are plausible values to consider two points as having similar characteristics and belonging to the same group.

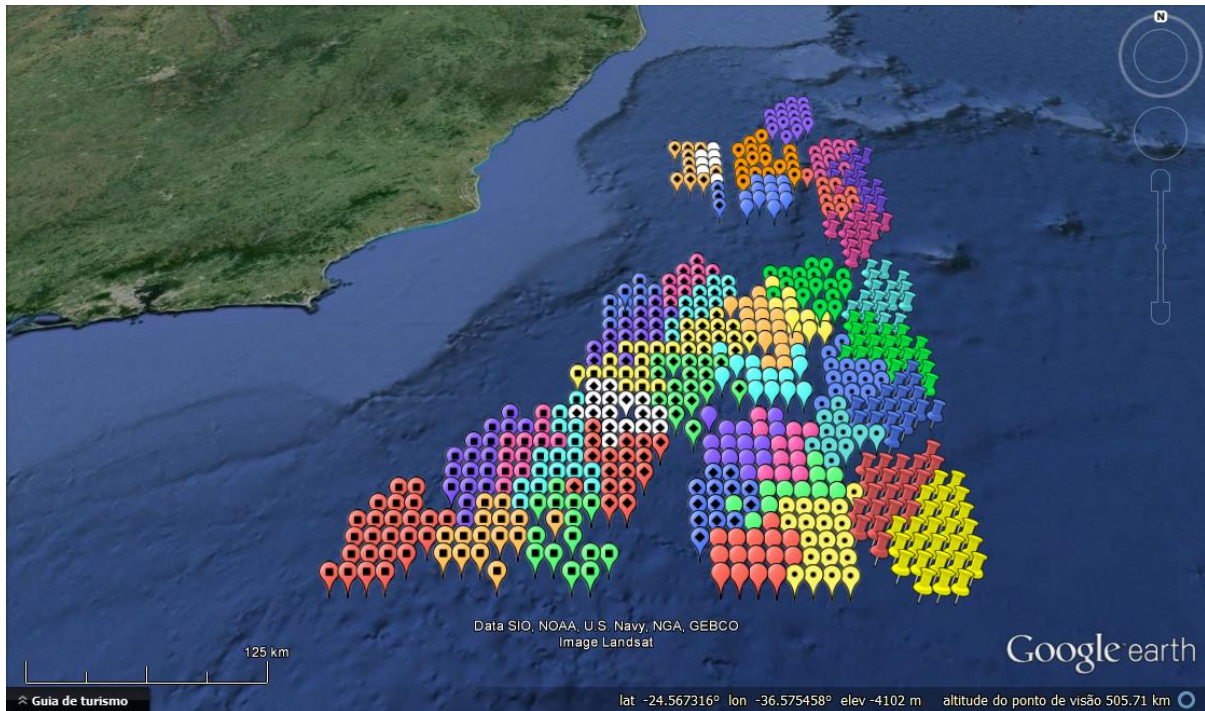
The result of the execution using Haversine formula was a total of 601 points clustered, distributed into 42 groups. Table 4 shows the distribution of the points clustered.



**Table 4 - List of clusters found in the selected execution using Haversine formula, containing the label of each cluster, for which are listed the elements by their index, followed by the total number of elements.**

CLUSTER	ELEMENTS	TOTAL
1	2 - 3 - 4 - 5 - 6 - 7 - 51 - 52 - 53 - 54 - 55 - 56 - 57 - 99 - 100 - 101 - 102 - 103 - 104 - 149 - 150 - 151 - 152 - 200 - 201 - 202	26
2	12 - 62 - 63 - 64 - 65 - 109 - 110 - 111 - 112 - 158 - 159 - 160 - 161 - 209 - 210 - 260 - 261	17
3	14 - 15 - 16 - 17 - 66 - 67 - 68 - 113 - 114 - 115 - 116 - 164 - 165 - 166 - 167 - 214 - 215 - 216 - 217 - 265	20
4	21 - 22 - 23 - 69 - 71 - 72 - 74 - 117 - 118 - 119 - 120 - 121 - 122 - 168 - 169 - 218	16
5	29 - 78 - 79 - 80 - 124 - 125 - 126 - 127 - 128 - 173 - 174 - 175 - 176 - 180 - 226 - 227	16
6	31 - 32 - 33 - 37 - 81 - 82 - 83 - 84 - 85 - 86 - 87 - 131 - 132 - 134 - 182 - 183	16
7	58 - 59 - 105 - 106 - 153 - 154 - 155 - 156 - 203 - 204 - 205 - 206 - 252 - 253 - 254 - 255 - 256 - 257 - 304 - 305	20
8	88 - 133 - 135 - 136 - 184 - 185 - 186 - 232 - 233 - 234 - 235	11
9	129 - 130 - 178 - 179 - 181 - 228 - 229 - 230 - 231 - 283	10
10	239 - 240 - 241 - 242 - 289 - 290 - 291 - 292 - 293 - 339 - 341 - 342 - 343 - 391 - 392 - 393	16
11	162 - 163 - 211 - 212 - 213 - 262 - 263 - 264 - 266 - 313 - 314 - 315 - 316 - 363	14
12	220 - 221 - 269 - 270 - 271 - 273 - 320 - 322 - 323 - 324 - 372 - 373 - 374 - 424 - 425 - 474 - 476	17
13	208 - 259 - 309 - 310 - 311 - 312 - 359 - 360 - 361 - 362 - 411	11
14	306 - 352 - 353 - 354 - 355 - 356 - 403 - 404 - 405 - 406 - 407 - 454 - 455 - 456 - 457 - 458 - 505 - 508 - 509	19
15	334 - 335 - 336 - 337 - 386 - 436 - 438 - 439 - 441 - 487 - 489 - 490 - 491 - 537 - 538 - 539 - 541	17
16	318 - 319 - 369 - 370 - 420 - 421 - 422 - 423 - 473 - 475	10
17	382 - 383 - 384 - 432 - 433 - 434 - 435 - 484 - 485 - 486 - 534 - 535	12
18	357 - 358 - 408 - 409 - 410 - 459 - 510 - 560 - 561 - 610 - 662	11
19	415 - 417 - 418 - 466 - 467 - 517 - 518 - 568 - 569 - 570 - 621 - 671 - 673	13
20	419 - 468 - 469 - 470 - 471 - 472 - 519 - 521 - 522 - 523 - 524 - 525 - 573 - 574 - 575 - 626	16
21	412 - 460 - 461 - 462 - 463 - 511 - 512 - 513 - 562 - 565 - 566	11
22	514 - 515 - 563 - 564 - 613 - 614 - 615 - 616 - 665 - 666 - 667 - 716 - 717 - 719	14
23	506 - 507 - 556 - 557 - 558 - 559 - 607 - 608 - 609 - 658 - 659 - 660 - 709 - 710 - 711	15
24	636 - 637 - 638 - 639 - 641 - 642 - 643 - 691 - 692 - 693	10
25	572 - 622 - 623 - 624 - 674 - 675 - 676 - 724 - 725 - 726 - 775 - 776 - 777	13
26	611 - 612 - 661 - 663 - 664 - 712 - 713 - 714 - 715 - 762 - 763 - 764 - 765	13
27	619 - 672 - 721 - 722 - 723 - 769 - 771 - 772 - 773 - 774 - 821 - 822 - 823 - 825 - 873 - 875 - 876	17
28	627 - 628 - 677 - 678 - 679 - 727 - 728 - 729 - 778 - 779 - 827 - 828	12
29	640 - 690 - 694 - 741 - 742 - 743 - 744 - 745 - 792 - 796	10
30	680 - 730 - 731 - 732 - 780 - 781 - 782 - 830 - 831 - 832	10
31	829 - 878 - 879 - 880 - 881 - 929 - 930 - 931 - 977 - 979 - 980 - 1024 - 1025 - 1065 - 1066	15
32	870 - 919 - 920 - 921 - 967 - 968 - 969 - 970 - 971 - 1013 - 1014 - 1015 - 1016 - 1058 - 1059 - 1060 - 1094	17
33	871 - 872 - 922 - 923 - 924 - 972 - 974 - 1017 - 1018 - 1019 - 1020 - 1021 - 1061 - 1062 - 1063 - 1099	16
34	826 - 874 - 877 - 925 - 926 - 927 - 928 - 975 - 976 - 978 - 1022 - 1023 - 1064 - 1102	14
35	882 - 932 - 933 - 934 - 981 - 982 - 983 - 1026 - 1027 - 1028	10
36	1010 - 1051 - 1052 - 1055 - 1089 - 1091 - 1092 - 1093 - 1123 - 1126 - 1127 - 1150 - 1151 - 1153	14
37	1056 - 1057 - 1095 - 1096 - 1097 - 1098 - 1100 - 1128 - 1129 - 1131 - 1132 - 1154 - 1155 - 1156 - 1179 - 1180	16
38	1125 - 1177 - 1178 - 1201 - 1203 - 1204 - 1205 - 1226 - 1227 - 1228 - 1229 - 1249 - 1250 - 1271 - 1272	15
39	1130 - 1157 - 1158 - 1159 - 1181 - 1182 - 1183 - 1206 - 1207 - 1208 - 1209	11
40	1184 - 1210 - 1211 - 1230 - 1231 - 1232 - 1233 - 1234 - 1251 - 1252 - 1253 - 1254 - 1255 - 1256 - 1275 - 1276 - 1277	17
41	1273 - 1292 - 1293 - 1310 - 1311 - 1312 - 1313 - 1314 - 1326 - 1327 - 1328 - 1329 - 1330 - 1331 - 1342 - 1343 - 1344 - 1345 - 1346 - 1357 - 1358 - 1359 - 1360 - 1371	24
42	1130 - 1157 - 1158 - 1159 - 1181 - 1182 - 1183 - 1207 - 1208 - 1209	10

The visualization of the results from Haversine execution can be seen in Figure 21, which shows the KML file plotted on Google Earth.



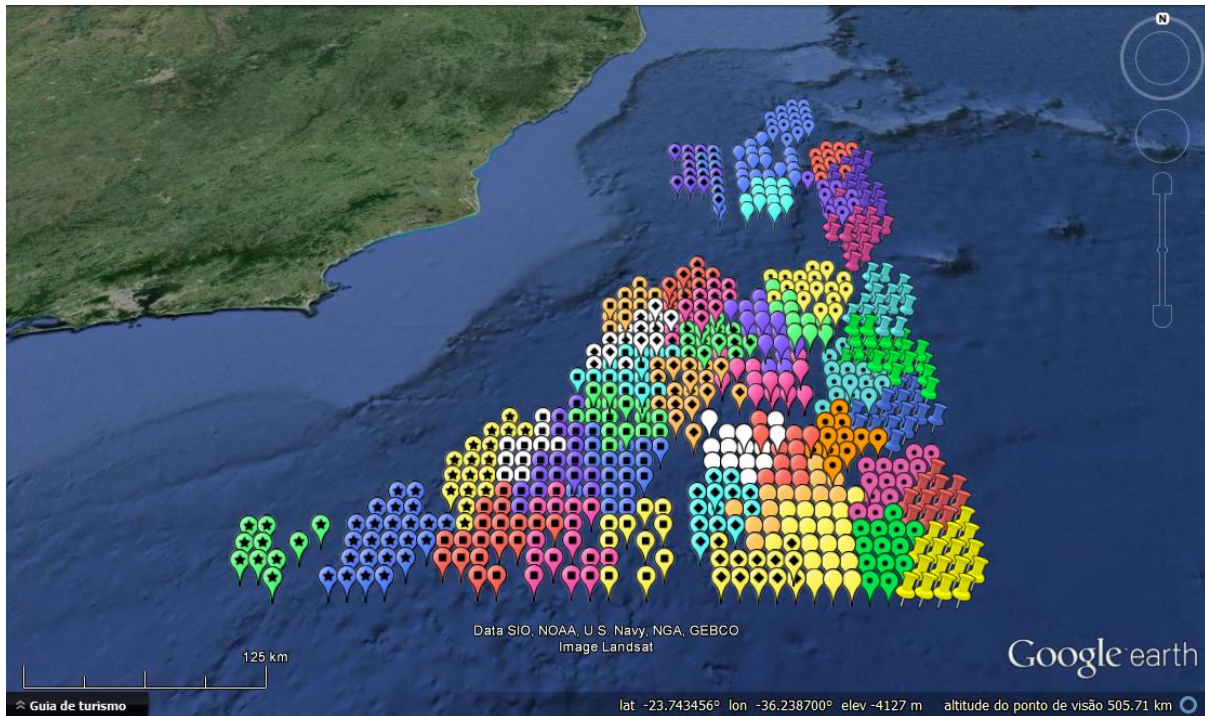
**Figure 21** - KML file plotted on Google Earth showing the 42 clusters formed using Haversine formula on the full area of *noaa\_bathymetry* dataset.

The same parameters from the Haversine experiment presented in Table 3 were maintained in the Vincenty experiment. The result of the execution using Vincenty formula was a total of 631 points clustered, distributed into 45 groups. Table 5 shows the distribution of the points clustered.

**Table 5 - List of clusters found in the selected execution using Vincenty formula, containing the label of each cluster, for which are listed the elements by their index, followed by the total number of elements.**

CLUSTERS	ELEMENTS	TOTAL
1	2 - 3 - 4 - 5 - 51 - 52 - 53 - 54 - 99 - 100 - 101 - 102 - 149 - 150 - 151 - 199	16
2	12 - 62 - 63 - 64 - 65 - 109 - 110 - 111 - 112 - 158 - 159 - 160 - 161 - 209 - 210 - 260 - 261	17
3	14 - 15 - 16 - 17 - 66 - 67 - 68 - 113 - 114 - 115 - 116 - 164 - 165 - 166 - 167 - 214 - 215 - 216 - 217 - 265	20
4	21 - 22 - 23 - 69 - 71 - 72 - 74 - 117 - 118 - 119 - 120 - 121 - 122 - 168 - 169 - 218	16
5	29 - 78 - 79 - 80 - 124 - 125 - 126 - 127 - 128 - 173 - 174 - 175 - 176 - 180 - 226 - 227	16
6	31 - 32 - 33 - 37 - 81 - 82 - 83 - 84 - 85 - 86 - 87 - 131 - 132 - 134 - 182 - 183	16
7	6 - 7 - 55 - 56 - 57 - 58 - 103 - 104 - 105 - 153 - 154	11
8	59 - 106 - 107 - 155 - 156 - 204 - 205 - 206 - 254 - 255 - 256 - 257 - 305	13
9	88 - 133 - 135 - 136 - 184 - 185 - 186 - 232 - 233 - 234 - 235	11
10	129 - 130 - 178 - 179 - 181 - 228 - 229 - 230 - 231 - 283	10
11	239 - 240 - 241 - 242 - 289 - 290 - 291 - 292 - 293 - 339 - 341 - 342 - 343 - 391 - 392 - 393	16
12	152 - 200 - 201 - 202 - 203 - 250 - 251 - 252 - 253 - 301 - 302 - 303 - 304	13
13	162 - 163 - 211 - 212 - 213 - 262 - 263 - 264 - 266 - 313 - 314 - 315 - 316 - 363	14
14	220 - 221 - 269 - 270 - 271 - 273 - 320 - 322 - 323 - 324 - 372 - 373 - 374 - 424 - 425 - 474 - 476	17
15	208 - 259 - 309 - 310 - 311 - 312 - 359 - 360 - 361 - 362 - 411	11
16	306 - 352 - 353 - 354 - 355 - 356 - 403 - 404 - 405 - 406 - 407 - 454 - 455 - 456 - 457 - 458 - 508 - 509 - 558	19
17	334 - 335 - 336 - 337 - 386 - 436 - 438 - 439 - 441 - 487 - 489 - 490 - 491 - 537 - 538 - 539 - 541	17
18	318 - 319 - 369 - 370 - 420 - 421 - 422 - 423 - 473 - 475	10
19	382 - 383 - 384 - 432 - 433 - 434 - 435 - 484 - 485 - 486 - 534 - 535	12
20	357 - 358 - 408 - 409 - 410 - 459 - 510 - 559 - 560 - 561 - 609 - 610 - 662	13
21	415 - 417 - 418 - 466 - 467 - 517 - 518 - 568 - 569 - 570 - 621 - 671 - 673	13
22	419 - 468 - 469 - 470 - 471 - 472 - 519 - 521 - 522 - 523 - 524 - 525 - 573 - 574 - 575 - 626	16
23	412 - 460 - 461 - 462 - 463 - 511 - 512 - 513 - 562 - 565 - 566	11
24	514 - 515 - 563 - 564 - 613 - 614 - 615 - 616 - 665 - 666 - 667 - 716 - 717 - 719	14
25	505 - 506 - 507 - 556 - 557 - 607 - 608 - 658 - 659 - 709 - 710 - 711	12
26	636 - 637 - 638 - 639 - 641 - 642 - 643 - 691 - 692 - 693	10
27	572 - 622 - 623 - 624 - 674 - 675 - 676 - 724 - 725 - 726 - 775 - 776 - 777	13
28	611 - 612 - 661 - 663 - 664 - 712 - 713 - 714 - 715 - 762 - 763 - 764 - 765	13
29	619 - 672 - 721 - 722 - 723 - 769 - 771 - 772 - 773 - 774 - 821 - 822 - 823 - 825 - 873 - 875 - 876	17
30	627 - 628 - 677 - 678 - 679 - 727 - 728 - 729 - 778 - 779 - 827 - 828	12
31	640 - 690 - 694 - 741 - 742 - 743 - 744 - 745 - 792 - 796	10
32	680 - 730 - 731 - 732 - 780 - 781 - 782 - 830 - 831 - 832	10
33	829 - 878 - 879 - 880 - 881 - 929 - 930 - 931 - 977 - 979 - 980 - 1024 - 1025 - 1065 - 1066	15
34	863 - 865 - 866 - 913 - 915 - 917 - 966 - 1009 - 1011 - 1012	10
35	870 - 919 - 920 - 921 - 967 - 968 - 969 - 970 - 971 - 1013 - 1014 - 1015 - 1016 - 1058 - 1059 - 1060 - 1094	17
36	871 - 872 - 922 - 923 - 924 - 972 - 974 - 1017 - 1018 - 1019 - 1020 - 1021 - 1061 - 1062 - 1063 - 1099	16
37	826 - 874 - 877 - 925 - 926 - 927 - 928 - 975 - 976 - 978 - 1022 - 1023 - 1064 - 1102	14
38	882 - 932 - 933 - 934 - 981 - 982 - 983 - 1026 - 1027 - 1028	10
39	1010 - 1051 - 1052 - 1055 - 1089 - 1091 - 1092 - 1093 - 1123 - 1126 - 1127 - 1149 - 1150 - 1151 - 1153	15
40	1056 - 1057 - 1095 - 1096 - 1097 - 1098 - 1100 - 1128 - 1129 - 1131 - 1132 - 1154 - 1155 - 1156 - 1179 - 1180	16
41	1125 - 1152 - 1177 - 1178 - 1201 - 1203 - 1204 - 1205 - 1206 - 1225 - 1226 - 1227 - 1228 - 1229 - 1249 - 1250 - 1271 - 1272	18
42	1130 - 1157 - 1158 - 1159 - 1181 - 1182 - 1183 - 1207 - 1208 - 1209	10
43	1184 - 1210 - 1211 - 1230 - 1231 - 1232 - 1233 - 1234 - 1251 - 1252 - 1253 - 1254 - 1255 - 1256 - 1275 - 1276 - 1277	17
44	1273 - 1292 - 1293 - 1310 - 1311 - 1312 - 1313 - 1314 - 1326 - 1327 - 1328 - 1329 - 1330 - 1331 - 1342 - 1343 - 1344 - 1345 - 1346 - 1357 - 1358 - 1359 - 1360 - 1371	24
45	1388 - 1400 - 1411 - 1412 - 1426 - 1427 - 1428 - 1440 - 1441 - 1442	10

The visualization of the results from Vincenty execution can be seen in Figure 22, which shows the KML file plotted on Google Earth.



**Figure 22** - KML file plotted on Google Earth showing the 42 clusters formed using Haversine formula on the full area of *noaa\_bathymetry* dataset.

The main difference between the experiments on the same full dataset using the two formulas with the same parameters was noticed on clusters 1, 7, 8 and 12 of Vincenty experiment, whose labels are shown in Figure 23, which differentiated from the clusters formed in size and elements. These elements in Haversine Experiment were distributed on two clusters, 1 and 7, instead of four, according to Figure 24.

```

cluster 1
2 - 3 - 4 - 5 - 51 - 52 - 53 - 54 - 99 - 100 - 101 - 102 - 149 - 150 - 151 - 199

cluster 7
6 - 7 - 55 - 56 - 57 - 58 - 103 - 104 - 105 - 153 - 154

cluster 8
59 - 106 - 107 - 155 - 156 - 204 - 205 - 206 - 254 - 255 - 256 - 257 - 305

cluster 12
152 - 200 - 201 - 202 - 203 - 250 - 251 - 252 - 253 - 301 - 302 - 303 - 304

```

**Figure 23** - Sample of clusters from Vincenty Experiments which had a different formation in size and elements from Haversine Experiments.

```

cluster 1
2 - 3 - 4 - 5 - 6 - 7 - 51 - 52 - 53 - 54 - 55 - 56 - 57 - 99 - 100 - 101 - 102 - 103 - 104 - 149 - 150 - 151 - 152 - 200 - 201 - 202

cluster 7
58 - 59 - 105 - 106 - 153 - 154 - 155 - 156 - 203 - 204 - 205 - 206 - 252 - 253 - 254 - 255 - 256 - 257 - 304 - 305

```

**Figure 24** - Elements clustered in Haversine Experiments in two groups, while in Vincenty Experiments the distribution of the same points was made into four groups.

The detailed analysis was carried by filtering the dataset only to the specific region of these clusters, which were formed in a different way from the rest of the cluster results. The specific region covered by the filtered dataset ranged from -38 to -38,6 and from -25 to -24,2. New executions were run on this filtered dataset using the same parameters from Table 2. The results from Haversine and Vincenty Formula are shown on Figures 25 and 26, respectively (labels are different from those presented on Figures 23 and 24, which shown the same clusters with the same objects, due do the fact that the dataset is now filtered and, thus, the indexes are reordered). Still, we analysed the results produced by Euclidean formula using the same parameters on this filtered dataset. In order to analyse the distance in the same units from the other distance calculation formulas, and also because Euclidean formula is intended to be applied on a plane surface, it is necessary to project the coordinates to UTM system. This conversion was made using ProGrid coordinate converter. The clusters formed by the Euclidean distance calculation formula are shown in Figure 27.

```

cluster 1
2 - 3 - 4 - 5 - 6 - 7 - 10 - 11 - 12 - 13 - 14 - 15 - 16 - 19 - 20 - 21 - 22 - 23 - 24 - 28 - 29 - 30 - 31 - 38 - 39 - 40

cluster 2
17 - 18 - 25 - 26 - 32 - 33 - 34 - 35 - 41 - 42 - 43 - 44 - 48 - 49 - 50 - 51 - 52 - 53 - 58 - 59

```

**Figure 25** - Clusters formed using Haversine formula on the selected area of filtered *noaa\_bathymetry* dataset.

```

cluster 1
2 - 3 - 4 - 5 - 10 - 11 - 12 - 13 - 19 - 20 - 21 - 22 - 28 - 29 - 30 - 37

cluster 2
6 - 7 - 14 - 15 - 16 - 17 - 23 - 24 - 25 - 32 - 33

cluster 3
18 - 26 - 27 - 34 - 35 - 42 - 43 - 44 - 50 - 51 - 52 - 53 - 59

cluster 4
31 - 38 - 39 - 40 - 41 - 46 - 47 - 48 - 49 - 55 - 56 - 57 - 58

```

**Figure 26** - Clusters formed using Vincenty Formula on the selected area of filtered *noaa\_bathymetry* dataset.



```

cluster 1
2 - 3 - 4 - 5 - 10 - 11 - 12 - 13 - 19 - 20 - 21 - 22 - 28 - 29 - 30 - 31 - 37 - 38 - 39

cluster 2
6 - 7 - 14 - 15 - 16 - 17 - 23 - 24 - 25 - 32 - 33

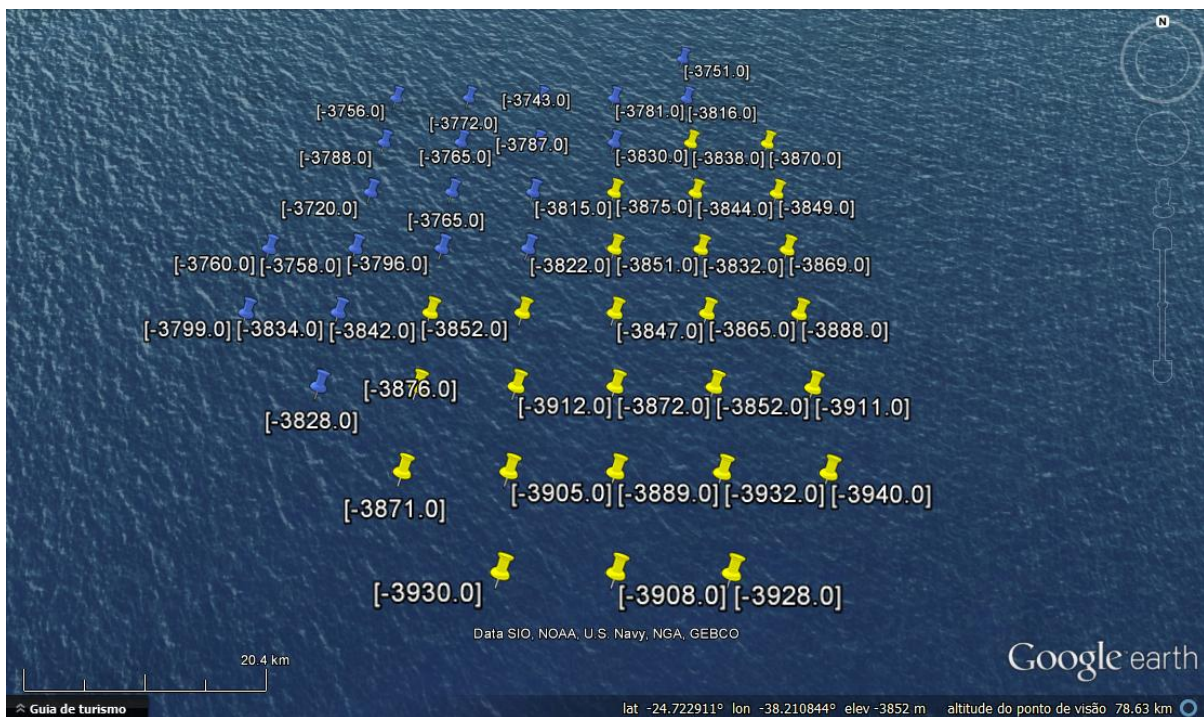
cluster 3
18 - 26 - 27 - 34 - 35 - 42 - 43 - 44 - 51 - 52 - 53

cluster 4
40 - 41 - 46 - 47 - 48 - 49 - 50 - 55 - 56 - 57 - 58 - 59

```

**Figure 27** - Clusters formed using Vincenty formula on the selected area of filtered *noaa\_bathymetry* dataset.

Figures 28, 29 and 30 show the output KML file plotted on Google Earth by Haversine, Vincenty and Euclidean formulas, respectively, where it is also possible to see the bathymetry values of each point.



**Figure 28** - KML file plotted on Google Earth showing clusters formed using Haversine formula on the selected area of filtered *noaa\_bathymetry* dataset. Cluster 1 is identified by yellow color; cluster 2 is identified by blue color. Values are bathymetry of each respective point.



**Figure 29** -KML file plotted on Google Earth showing clusters formed using Vincenty formula on the selected area of filtered *noaa\_bathymetry* dataset. Cluster 1 is identified by yellow color; cluster 2 is identified by dark blue color; cluster 3 is identified by green color; cluster 4 is identified by light blue color. Values are bathymetry of each respective point.



**Figure 30** - KML file plotted on Google Earth showing clusters formed using Euclidean formula on the selected area of filtered *noaa\_bathymetry* dataset. Cluster 1 is identified by yellow color; cluster 2 is identified by blue color. Values are bathymetry of each respective point.



The analysis on the distances calculated was made using Excel and database techniques, where the comparison between all the objects showed divergences between the results of the three formulas. Table 6 shows some examples from the executions in the filtered *noaa\_bathymetry* dataset, where a sample of the full comparison between the results is presented, showing the differences in results of distance calculations. These differences can form different dense regions on the database, depending on the formula used.

**Table 6 - Examples of points (P1 and P2) from filtered *noaa\_bathymetry* dataset, the Euclidean, Haversine and Vincenty distances between them and the difference between the results of Haversine x Euclidean, Euclidean x Vincenty and Vincenty x Haversine (in meters).**

P1	P2	lon/ lat P1	lon/ lat P2	UTM converted Euclidean Distance	Haversine Distance	Vincenty Distance	Haversine x UTM converted Euclidean	UTM converted Euclidean x Vincenty	Vincenty x Haversine
1	18	-38/-25	-38,1/-24,2	89164.92	89628.39	89190.411	463.4699874	25.48793605	437.9821
9	10	-38/-24,2	-38,1/-25	89165.57	89628.39	89190.411	462.8194593	24.83740795	437.9821
18	10	-38,1/ -24,2	-38,1/-25	88587.89	89055.08	88613.582	467.1838591	25.68984199	441.494
18	19	-38,1/ -24,2	-38,2/-25	89162.86	89628.39	89190.411	465.537977	27.5559257	437.9821
27	10	-38,2/ -24,2	-38,1/-25	89162.83	89628.39	89190.411	465.5670643	27.58501293	437.9821
27	19	-38,2/ -24,2	-38,2/-25	88584.5	89055.08	88613.582	470.5789875	29.08497033	441.494
27	28	-38,2/ -24,2	-38,3/-25	89161.18	89628.39	89190.411	467.217459	29.23540771	437.9821
36	19	-38,3/ -24,2	-38,2/-25	89160.78	89628.39	89190.411	467.6173425	29.63529117	437.9821
36	28	-38,3/ -24,2	-38,3/-25	88584.15	89055.08	88613.582	470.9279431	29.43392598	441.494
45	28	-38,4/ -24,2	-38,3/-25	89159.77	89628.39	89190.411	468.6267607	30.64470937	437.9821

The region approached in this filtered dataset is small, but in larger areas of the globe, these differences in distances calculation results are more likely to increase, and may cause significant divergences between clusters formed by different formulas.

Since the experiment using Euclidean formula on the filtered dataset presented a considerably different number of groups from the results generated by Haversine and Vincenty distance calculations, the next experiments were run on the full dataset, which had its coordinates converted to UTM system by ProGrid software, using Euclidean calculation formula and the same parameters from the previous executions.

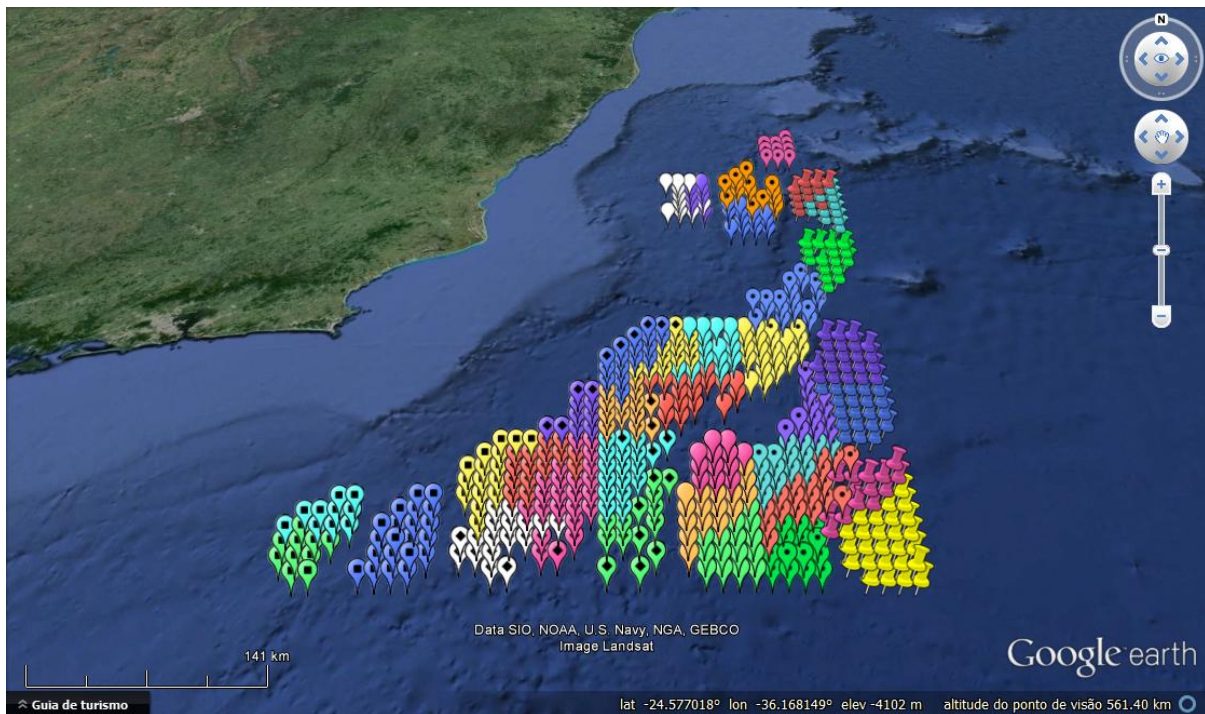
The result of the execution using Euclidean formula on the full dataset was a total of 548 points clustered, distributed into 37 groups. Table 7 shows the distribution of the points clustered.



**Table 7 - List of clusters found in the selected execution using Euclidean formula, containing the label of each cluster, for which are listed the elements by their index, followed by the total number of elements.**

CLUSTERS	ELEMENTS	TOTAL
1	51 - 52 - 53 - 54 - 55 - 56 - 57 - 99 - 100 - 101 - 102 - 103 - 104 - 149 - 150 - 151 - 152 - 153 - 199 - 200 - 201 - 202 - 203 - 251 - 252 - 253 - 256	27
2	62 - 63 - 64 - 65 - 109 - 110 - 111 - 112 - 113 - 159 - 160 - 161 - 162 - 163 - 209 - 210 - 211 - 212 - 213 - 264	20
3	78 - 79 - 80 - 124 - 125 - 126 - 127 - 128 - 173 - 174 - 175 - 176 - 178 - 226 - 227 - 228	16
4	82 - 83 - 84 - 85 - 86 - 87 - 131 - 132 - 134 - 182 - 183	11
5	58 - 59 - 105 - 106 - 154 - 155 - 156 - 204 - 205 - 206 - 254 - 257 - 304 - 305	14
6	66 - 67 - 68 - 69 - 114 - 115 - 116 - 117 - 118 - 164 - 165 - 166 - 167 - 168 - 214 - 215 - 216 - 217 - 218 - 265 - 266	21
7	88 - 133 - 135 - 136 - 184 - 185 - 186 - 232 - 233 - 234 - 235	11
8	239 - 240 - 241 - 242 - 290 - 291 - 292 - 341 - 342 - 343	10
9	207 - 255 - 258 - 306 - 307 - 308 - 309 - 356 - 357 - 358 - 407 - 408 - 409 - 458 - 459 - 507 - 508 - 509	18
10	260 - 261 - 262 - 263 - 311 - 312 - 313 - 314 - 315 - 316 - 362 - 363 - 413	13
11	221 - 222 - 223 - 270 - 271 - 274 - 275 - 322 - 323 - 324 - 325 - 372 - 373 - 374 - 424 - 425 - 474 - 476	18
12	352 - 353 - 354 - 355 - 403 - 404 - 405 - 406 - 454 - 455 - 456 - 457 - 505 - 506	14
13	259 - 310 - 359 - 360 - 361 - 410 - 411 - 412 - 460 - 461 - 462 - 510 - 511 - 512 - 513	15
14	318 - 319 - 320 - 368 - 369 - 418 - 420 - 421 - 422 - 473	10
15	334 - 335 - 336 - 337 - 386 - 436 - 438 - 439 - 441 - 489 - 490 - 491 - 539 - 541	14
16	370 - 419 - 468 - 469 - 470 - 471 - 472 - 519 - 521 - 522 - 523 - 524	12
17	382 - 383 - 384 - 433 - 434 - 435 - 484 - 485 - 486 - 487 - 535 - 537 - 538	13
18	556 - 557 - 558 - 559 - 560 - 607 - 608 - 609 - 610 - 658 - 659 - 660 - 709 - 710 - 711	15
19	572 - 573 - 574 - 575 - 622 - 623 - 624 - 626 - 674 - 675 - 676 - 677 - 724 - 728 - 775	15
20	561 - 562 - 563 - 611 - 612 - 661 - 662 - 663 - 712 - 713 - 714 - 761 - 762 - 763 - 764 - 765	16
21	564 - 613 - 614 - 615 - 616 - 664 - 665 - 666 - 667 - 715 - 716 - 717	12
22	636 - 637 - 638 - 639 - 641 - 642 - 643 - 691 - 692 - 693	10
23	569 - 570 - 619 - 621 - 672 - 673 - 721 - 722 - 723 - 771 - 772 - 773 - 774 - 822 - 823 - 825 - 876	17
24	640 - 690 - 694 - 741 - 742 - 743 - 744 - 745 - 792 - 796	10
25	725 - 726 - 727 - 776 - 777 - 778 - 779 - 826 - 827 - 828 - 877 - 928	12
26	829 - 830 - 878 - 879 - 880 - 881 - 929 - 930 - 931 - 976 - 977 - 978 - 979 - 980 - 1023 - 1024 - 1025	17
27	817 - 863 - 865 - 866 - 867 - 868 - 913 - 915 - 917 - 965 - 966 - 1009 - 1011 - 1012	14
28	820 - 870 - 919 - 920 - 921 - 922 - 967 - 968 - 969 - 970 - 971 - 972 - 1013 - 1014 - 1015 - 1016 - 1017 - 1018	18
29	872 - 874 - 923 - 924 - 925 - 926 - 927 - 974 - 975 - 1019 - 1020 - 1021 - 1022	13
30	1055 - 1056 - 1057 - 1058 - 1059 - 1060 - 1091 - 1092 - 1093 - 1094 - 1095 - 1098 - 1123 - 1126 - 1127 - 1128 - 1129 - 1150 - 1151 - 1153 - 1154 - 1155 - 1179	23
31	1061 - 1062 - 1063 - 1064 - 1099 - 1100 - 1101 - 1102 - 1132 - 1159	10
32	1096 - 1097 - 1130 - 1131 - 1156 - 1157 - 1158 - 1180 - 1181 - 1182 - 1183 - 1206 - 1207 - 1208 - 1209	15
33	1125 - 1152 - 1177 - 1178 - 1201 - 1203 - 1204 - 1205 - 1225 - 1226 - 1227 - 1228 - 1229 - 1249 - 1250 - 1271 - 1272	17
34	1184 - 1210 - 1230 - 1231 - 1232 - 1233 - 1234 - 1251 - 1252 - 1253 - 1254 - 1255 - 1256 - 1275 - 1276 - 1277	16
35	1310 - 1311 - 1312 - 1313 - 1314 - 1326 - 1327 - 1328 - 1329 - 1330 - 1331 - 1342 - 1343 - 1344 - 1345 - 1346 - 1357 - 1358 - 1359 - 1360 - 1371	21
36	1400 - 1411 - 1412 - 1413 - 1425 - 1426 - 1427 - 1428 - 1440 - 1441	10
37	1388 - 1389 - 1390 - 1401 - 1402 - 1403 - 1414 - 1415 - 1429 - 1442	10

The visualization of the results from Euclidean execution can be seen in Figure 31, which shows the KML file plotted on Google Earth.



**Figure 31** - KML file plotted on Google Earth showing the 37 clusters formed using Euclidean formula on the full area of *noaa\_bathymetry* dataset.

The groups resulting from the Euclidian distance calculation formula were the most divergent among the executions on the full dataset, being the most different number and formation of clusters among the three formulas. This can be justified by the fact that it uses a different approach to calculate distances from the other two formulas, and the differences caused by this characteristic are accentuated in large areas. While Haversine and Vincenty use geographic data in decimal degrees converted to Radius on trigonometric functions, these geographic data has to be converted to UTM to be possible to apply Euclidean formula on the plane UTM Cartesian projection, in which distances are calculated in metres. The differences between the Euclidean results of the filtered dataset were more similar to Haversine and Vincenty because the region approached in this particular dataset is very small, although it covers two UTM zones (23 and 24).

From the results presented by the set of executions on the two datasets (full and filtered), it can be also noted that the clustering execution using Euclidean formula led a bigger amount of objects not clustered, while the amount of points clustered and points left by Haversine and Vincenty formulas were similar. The comparison between these and other characteristics of the results from the three methods for execution (by Haversine, Vincenty and Euclidean formulas) on the full dataset is presented on Table 8.

**Table 8 -Comparison between the results of the executions using the three distance calculation formulas (Haversine, Vincenty and Euclidean), regarding number of clusters, number of points clustered and number of points not clustered.**

	<b>Number of Clusters</b>	<b>Number of Points Clustered</b>	<b>Number of points not Clustered</b>
<b>Haversine</b>	<b>42</b>	601	853
<b>Vincenty</b>	<b>45</b>	631	823
<b>Euclidean</b>	<b>37</b>	548	906

Since the parameters set for spatio-temporal data mining are restrictive, and, besides the geographical closeness between objects, they are also clustered by the similarity between the values of properties, the amount of not clustered points left by the three executions was high. Still, it is desirable that the execution leaves to be clustered as less points as possible and, in this case, the Vincenty Formula was the method which clustered more points.

Thus, as shown in this subsection, Vincenty distance calculation formula showed to be the most suitable used on the calculation of points on Earth's surface in our experiments. The next subsection presents a final discussion and points out other important characteristics of the executions.

### **6.3.1.1 Final Considerations about Experiment 1**

The main goal of this section was to present a set of experiments run on *noaa\_bathymetry* dataset and on its filtered version, in order to analyse the use of three different distance calculation formulas to cluster groups of objects of similar characteristics. The characteristics used to carry out this comparison were the number of clusters formed, the formation of the clusters (and similarity among them) and the number of objects clustered by using each formula.

Vincenty calculation formula showed to work well both on small and big regions. As stated previously, it has a high level of accuracy, by using parameters of the ellipsoid of revolution on which is based the approached region and trigonometric functions in order to be the most precise as possible in the results. Haversine formula presented similar results to Vincenty for the largest region (full dataset), while in small regions it seems to present significantly different results from Vincenty, which is the most precise of the three. Finally, the Euclidean formula showed similar cluster results to Vincenty for small regions, while it had a significantly different behaviour on larger regions of the globe.

From the results above, it can be concluded that, in our experiments with *NOAA\_dataset*, Euclidean distance formula is the less advisable to be used, mainly in larger regions. This is due to the fact that it does not consider the shape of the Earth in its process of calculation, being necessary to convert the coordinates to UTM and, thus, to apply the formula considering a plane surface. However, on large regions, more than one UTM zone is approached, and since the calculation between UTM

zones is a non-trivial task, the results can present distortion. Besides, the incidence of not clustered points was the higher between the three formulas. On the other hand, Haversine formula showed to work better on larger than on smaller regions, because it considers the radius of the Earth on its calculations. However, points clustered and the clusters found were less than the Vincenty formula, which has shown more accuracy and ability in finding groups of similar characteristics and clustering as more points as possible.

### 6.3.2 Experiment 2 – Testing Haversine formula on a dataset until 140 MY

The main goal of this experiment was to test the solution implemented related to the temporal dimension of palaeogeographic data. The dataset used was *muller\_bathymetry*, since this dataset presents variation in time, generated by Deriva.m, from a region of interest for research defined in the Brazilian coast.

The region approached on this dataset presents 2,478 points, from which the set of points of their location in the past was derived. After mapping each point, the locations were combined to bathymetric data of the points in the corresponding MY. In order to carry out a sample analysis and test the validity of the temporal solution on a smaller dataset, containing a sample from the 115,946 instances of the full dataset, a small region was chosen, ranging from -18.4 to -19 latitude and from -33 to -33.9 longitude. In this small region, experiments were run, using parameter values ranging from 10000 to 50000 for *eps1*, from 100 to 500 for *eps2*, from 10 to 50 for the *minPts*, from 125 to 525 for the *threshold*, and *eps3* varying among *Eocene*, *Paleocene*, *Lower Cretaceous*, *Upper Cretaceous* - as nominal parameters -, 40, 60, 70 and 100 - as numeric parameters -, since these are the epochs of interest for the research on Paleoprospec Project.

Haversine formula was chosen to be applied in this experiment, due to the fact that it considers the shape of the Earth as a sphere and, thus, calculates the great-circle distance between two points using the Earth radius as parameter. Vincenty formula, although presenting a higher accuracy, could not be used in this case, because the temporal dimension of the dataset ranges until 140 MY, and the parameters of the ellipsoid used by Vincenty formula are valid only for the present. Still, since the region goes further from SIRGAS2000 region, the decimal degrees of latitude and longitude could not be converted to UTM by ProGrid. Thus, this dataset was tested using only Haversine Formula.

From the set of results analysed, four were chosen given the formation of the clusters, distribution of its elements and coherence of the parameters with the region and characteristics presented. Table 9 presents the parameters set on the algorithm for the four experiments, set according to the domain expert and considering the size and characteristics of the dataset varying only *eps3*.

**Table 9 - Parameters used for the set of experiments on the filtered *muller\_bathymetry* dataset, maintaining the same values for *eps1*, *eps2*, *minPts* and *threshold*, and varying among *Eocene*, *Paleocene*, *Upper Cretaceous* and *Lower Cretaceous* as nominal parameters, and among 40, 60, 70 and 100 as numeric parameters.**

<i>eps1</i>	<i>eps2</i>	<i>eps3</i>	<i>minPts</i>	<i>threshold</i>
30000	500	<i>Eocene / Paleocene / Upper Cretaceous / Lower Cretaceous / 40 / 60 / 70 / 100</i>	20	525

For the region approached, a variation of 500m in bathymetry and 30000m in geographic location between points are plausible values to consider two points as having similar characteristics and belonging to the same group, according to the domain experts from Paleoprospec Project. The *minPts* was set considering the size of the dataset, and the *threshold* considering the *eps2*.

Thus, defined the parameters, the set of executions was run. The solution implemented allows the option to declare the third parameter *eps* as a name of the Epoch, as well as by the number of the MY, from which points will be grouped according to the epoch in which they are located. In order to validate the solution for both options, experiments were carried using the same dataset and maintaining the same values of the other parameters than *eps3*. The next two subsections describe the execution and the results for each one of the possibilities.

### 6.3.2.1 Testing Haversine formula on a dataset until 140 MY given an Epoch (name) as parameter

In these experiments, the parameters used were the name of the epochs of interest for the research, being *Eocene*, *Paleocene*, *Lower Cretaceous*, *Upper Cretaceous*.

The result of the execution using parameter *Eocene* as *eps3* was a total of 1607 points distributed into 27 groups. This Epoch ranges from 34 to 56 MY, according to the reference used in our research [GRADSTEIN *et. al*, 2012]. The list of clusters and distribution of elements into the groups is presented on Table 12 in Appendix A.

The second value for *eps3* parameter used, *Paleocene*, presented a total of 700 points clustered into 12 groups. In our implementation, the Paleocene ranges from 56.1 to 66 MY. The list of clusters formed considering this Epoch and distribution of elements into the groups is presented on Table 13 in Appendix A.

The third value, *Upper Cretaceous*, resulted in 39 clusters of similar characteristics, grouping 1571 objects. This Epoch ranges from 66.1 to 100.5 MY, in our implementation. The result of the execution can be seen in on Table 14 in Appendix A.

Finally, the fourth value, *Lower Cretaceous*, did not find any group of similar characteristics considering the spatial, non-spatial and temporal characteristics above the minimum number of points. Thus, there was no cluster formed.

Table 10 presents a comparison of the executions using the four different parameters for the temporal dimension.

**Table 10 - Results of the executions between the four nominal parameters chosen to test the temporal dimension.**

<i>eps3</i>	Number of Clusters	Number of Points Clustered
<b>Eocene</b>	27	1607
<b>Paleocene</b>	12	700
<b>Upper Cretaceous</b>	39	1571
<b>Lower Cretaceous</b>	0	0

From the comparison presented on Table 10, it is possible to validate the solution implemented for temporal dimension, grouping temporal neighbours by Epoch, defined by its name. This can be done by filtering in the dataset the total number of points belonging to each Epoch, and then comparing to the total number of points clustered, analysing their non-spatial values. The difference between the groups formed by each parameter is the geographic location along the continental drift, which must be less than parameter *eps1*. The parameter which determines the last rule to be observed for a point to be considered a member of the cluster, or be classified as *NOISE*, is *eps2*.

For Eocene Epoch, the total number of objects in the dataset whose MY ranges from 34 to 56 MY is 1610. The difference between bathymetry values of these objects ranges from -4480 to -5026. The difference between these bathymetry values exceeds the parameter *eps2*. Thus, the algorithm found 1607 objects belonging to the Eocene Epoch, except the points which did not fit the parameter *eps2*.

For Paleocene Epoch, the total number of objects in the dataset whose MY ranges from 57 to 66 MY is 700. The difference between bathymetry values of these objects ranges from -4705 to -4296. The difference between these bathymetry values does not exceed the parameter *eps2*. Thus, the algorithm clustered all the 700 objects belonging to the Eocene Epoch.

For Upper Cretaceous Epoch, the total number of objects in the dataset whose MY ranges from 67 to 100.5 MY is 1577. The difference between bathymetry values of these objects ranges from -3119 to -4461. The difference between these bathymetry values exceeds the parameter *eps2*. Thus, the algorithm clustered all the 1571 objects belonging to the Eocene Epoch, except the points which, compared to other points of the dataset, exceed the parameter *eps2*.

For Lower Cretaceous Epoch, there are no objects in the dataset whose MY ranges from 107 to 145. This is probably due to the fact that, for the specific continental drift model from which the locations in the past were generated, this region chosen near Campos Basin only appeared after this Epoch, when the seafloor in this region was formed from the Mid-Atlantic Ridge. Thus, the algorithm did not find any cluster of objects belonging to this epoch.

The following section continues this validation by testing the execution using the temporal dimension as Epochs from the geologic time scale, being the parameters defined as the number of the MY of interest.

### 6.3.2.2 Testing Haversine formula on a dataset until 140 MY given a numerical age as parameter

After validated the solution regarding temporal dimension given a name of an Epoch as a parameter, experiments were run on the same dataset in order to validate the other option to declare the temporal parameter, being a million of year as a number, from which points will be considered neighbours if they belong to the same Epoch. The parameters *eps1*, *eps2*, *minPts* and *threshold* were the same from the previous executions, and for each epoch declared previously as parameter *eps3*, a correspondent million of year was assigned.

In the first execution, the value declared as *eps3* parameter was 40. The total number of objects clustered in this execution was 1610, the same as the result of the execution for Eocene Epoch.

In the second execution, the value declared as *eps3* parameter was 60. The total number of objects clustered in this execution was 700, the same as the result of the execution for Paleocene.

In the third execution, the value declared as *eps3* parameter was 70. The total number of objects clustered in this execution was 1571, the same as the result of the execution for Upper Cretaceous Epoch.

In the fourth execution, the value declared as *eps3* parameter was 100. No clusters were formed by this execution, which was the same result of the execution for Upper Cretaceous Epoch.

The comparison between the results of the executions is presented on Table 11.

**Table 11 - Results of the executions between the four numerical parameters chosen to test the temporal dimension.**

<i>eps3</i>	Number of Clusters	Number of Points Clustered
40	27	1607
60	12	700
70	39	1571
100	0	0

The clusters formed by the execution of these four experiments are presented on Tables 15, 16 and 17 in Appendix A. Analysing these tables, it is possible to look at the elements present in each cluster of the executions using numeric *eps3* parameter and conclude that they formed the same clusters from the correspondent executions using numerical *eps3* parameters.

From the description of these experiments, it was possible to validate the implementation that handles temporal dimension according to epochs from the geologic time scale, in which the parameter *eps3* is declared as a numerical age. Next subsection presents further discussions on the executions.

### **6.3.2.3 Final Considerations about Experiment 2**

The main goal of this set of experiments was to validate the solution implemented to handle temporal dimension of palaeogeographic data considering the structure of the geologic time scale, by considering temporal neighbours objects belonging to the same epoch.

From the first approach, it was concluded that the implementation successfully groups objects of similar characteristics according to spatial, non-spatial and also temporal dimensions, being this latter declared in parameter *eps3* as a name of the epoch. These experiments run testing three different epochs identified all the similar objects, except for the objects which presented difference in non-spatial values greater than parameter *eps2*. Thus, this solution showed to be effective to its end.

The second set of experiments was carried on the same dataset, using the same parameters, but changing the name of the epoch to a correspondent numeric value of an age belonging to the specific epoch. The results of this set of experiments showed the same results from the previous one, thus validating the solution implemented to cluster objects according to the epochs of the geologic scale by both approaches.



## 7. FINAL CONSIDERATIONS

Spatio-temporal data mining is an emerging research area dedicated to the development and application of novel computational techniques for the analysis of spatio-temporal databases [ANDRIENKO *et al.*, 2006]. It has increased its importance as the amount of data stored in geographic databases has grown, being these data collected from several sources, such as telecommunications, mobile computing applications, web traffic records, historical sources, and so on. This increasing in geographic data stored had highlighted the importance of GKD. The process of GKD is based on the idea that there are new and useful knowledge to be discovered in the unprecedented amount of georeferenced data that have been collected, recorded and shared by researchers, public agencies and the private sector. However, traditional methods are not appropriate for geographic knowledge discovery, with the need for adaptation or development of new tools [MILLER & HAN, 2009], in order to analyse spatio-temporal data and its features.

For this reason, this dissertation presents a set of improvements on ST-DBSCAN algorithm to mine palaeogeographic data. Solutions were developed, implemented and tested, and compared to the results of the traditional approach. Results obtained are very promising.

Chapter 2 described the theoretical principles to better understand the problem at study. It provided explanations on the types of representation for georeferenced data, based on different surfaces. It also made an overview on geologic time scale and its granularity, pointing out how particular is this approach to be used as temporal dimension. Chapter 3 presented a revision about clustering analysis, particularly the density-based method, and spatio-temporal data mining in general. Its purposes were to provide a better comprehension of the resources and methods explained in Chapter 4 and used to build the solution to the research problem.

Chapter 5 and 6 presented the results obtained during the development of this research. The solutions implemented were presented in details in Chapter 5, from the pre-processing step to the output that allowed the analysis of the results. From these implementations, the experimental results were generated and showed on Chapter 6. The experiments compared the three distance calculation formulas presented in this research, showing in the results that Vincenty formula is able to find more groups of similar characteristics, leaving less points to be clustered, followed by Haversine formula. The same experiments showed that Euclidean formula is the one which presents less groups formed, leaving more points to be clustered.

Experiments testing the three distance formulas could only be carried on datasets whose temporal dimension is situated on the present (0 MY), since Vincenty formula uses parameters of semi-major and semi-minor axes of the ellipsoid, values which changed over time (millions of years)

during the formation of the ellipsoid as it is today. Thus, these experiments were carried on *noaa\_bathymetry* dataset, which contains data of the continental crust of Santos Basin in 0 MY.

The result of the experiments on this full dataset showed a considerable similarity between clusters formed by Haversine and Vincenty formulas, while Euclidean formula formed completely different groups. Apart from a set of points situated on a small region, clusters formed by Haversine and Vincenty formulas were the same. But in this small region, points were distributed into 4 clusters by Vincenty formula, and into 2 clusters by Haversine formula. The dataset was filtered to only this small region, and data was converted to UTM, in order to test Euclidean distance in meters on the Cartesian plan of UTM projection. The execution of the algorithm on the filtered dataset using Euclidean formula showed similar results to Vincenty formula.

From the results of the executions on *NOAA\_dataset*, it was possible to conclude that Vincenty formula presented the most accurate results in distance calculations, clustering more points in a bigger number of groups, for both small and broad areas. Haversine formula showed to be more effective on large areas, in which almost all the clusters were formed with the same elements than the execution using Vincenty formula. On the other hand, Euclidean formula showed results similar to Vincenty formula on small areas, while on large areas the results were the most divergent among the formulas tested. This can be due to the fact that calculations in UTM system are made inside each zone, and the full dataset covers several UTM zones. Indeed, the filtered dataset in which Euclidean formula results were better is covered by only two UTM zones. Thus, in broader areas of the Earth's surface, our experiments showed that Euclidean formula is not the most suitable to be applied. In this case, Haversine and Vincenty clustered more points in a bigger number of groups, showing to be more effective. Still, Vincenty formula showed the best results not only for big, but also for small areas.

Another limitation regarding the use of Euclidean formula in distance calculations is the conversion of data from broad areas to UTM system. On the second set of experiments to test the temporal solution, it was used *muller\_dataset*, in which data from a specific region near Campos Basin was mapped to 140 MY ago, according to the continental drift approach described by [NÜRNBERG & MÜLLER, 1991]. The geographical location of objects in the past was placed further from the limits of SIRGAS2000 region, the *datum* in use in Brazil. Thus, it was impossible to convert the dataset using ProGrid software. By other methods, it would be possible to convert the dataset object by object, by defining the specific UTM zone of each object. Since the dataset has 115,946 instances (full version, being 6,057 instances on the filtered version), this task could not be completed. Thus, this conversion to UTM of a dataset in which the geographical points are presented in a region covering several UTM zones becomes a challenge.

Haversine formula was the only applied in this experiment to test the temporal dimension, not only because of the conversion of decimal degrees to UTM that limits the use of Euclidean formula, but also because Vincenty formula, although presenting a higher accuracy, uses the parameters of the ellipsoid, which are valid only for the present. The results showed that the solution implemented is effective in filtering as temporal neighbors objects which occurred in the same epoch of the geologic time scale. However, a limitation of this method, which is suggested as a future work, is to determine the temporal similarity of objects by a distance function, allowing similar objects belonging to different epochs to be clustered in the same group.

Therefore, the foremost contribution on chapter 6 was to confirm that the solution implemented addressed the problem of research, by examples of executions on datasets of different characteristics. The results show that this implementation is effective in finding groups of similar characteristics according to the spatial, non-spatial and temporal dimensions of palaeogeographic data, considering appropriate formulas to calculate distances on Earth's surface and the temporal dimension according to the geologic time scale.

## 7.1 Main Contributions

According to [HAN & KAMBER, 2006], clustering is a challenging field of research in which its potential applications pose their own special requirements. The main contributions of this research to the addressed problems are presented below, according to a list of typical requirements of clustering in data mining, presented by [HAN & KAMBER, 2001], describing the contributions made by this work in each sense:

- Ability to deal with different types of attributes: Many algorithms are designed to cluster interval-based (numerical) data. However, applications may require clustering other types of data, such as binary, categorical (nominal), and ordinal data, or mixtures of these data types [HAN & KAMBER, 2001]. Palaeogeographic databases tend to present more than one type of data, since time dimension is presented in millions of years, or because some properties can be presented as categorical values. The solution presented in this work handles different types of data, according to the needs of the data mining task and the definition of the parameters by the user. Thus, our solution is able to deal with different types of attributes.
- Discovery of clusters with arbitrary shape: Many clustering algorithms determine clusters based on Euclidean or Manhattan distance measures. Algorithms based on such distance measures tend to find spherical clusters with similar size and density. However, a cluster could be of any shape. It is important to develop algorithms that can detect clusters of arbitrary shape [HAN & KAMBER, 2001]. In this sense, the solution developed in this work

to address spatial dimension uses different formulas to calculate distances and find similarities between objects, being able to detect clusters of arbitrary shapes, as showed in figures of experiments presented on Chapter 6.

- Minimal requirements for domain knowledge to determine input parameters: Many clustering algorithms require users to input certain parameters in cluster analysis (such as the number of desired clusters). The clustering results can be quite sensitive to input parameters. Parameters are often difficult to determine, especially for datasets containing high-dimensional objects. This not only burdens users, but it also makes the quality of clustering difficult to control [HAN & KAMBER, 2001]. The density-method chosen to our research does not require the predetermination of the number of clusters, but only the minimum number of points required to a group of objects be classified as such. Also, the *eps* values for similarity measures in temporal, spatial and non-spatial dimensions are determined based on the opinion of the domain expert and according to the needs of the research, given the measures of the distance calculation formulas in kilometres, as well as the epoch or MY of interest.
- Ability to deal with noisy data: Most real-world databases contain outliers or missing, unknown, or erroneous data. Some clustering algorithms are sensitive to such data and may lead to clusters of poor quality. In this sense, ST-DBSCAN has an improvement on DBSCAN by defining a parameter *threshold* by which objects can be grouped based on the average of the non-spatial values of clusters formed, decreasing the number of noise values in the mining process. Also the temporal approach in epochs makes it possible to group a larger number of objects in a cluster than in specific ages, since the border values between one age and another is not a consensus in the literature, and the definition of a broader period such as the Epochs can considerably low noisy values.
- High dimensionality: A database or a data warehouse can contain several dimensions or attributes. Many clustering algorithms are good at handling low-dimensional data, involving only two to three dimensions. Human eyes are good at judging the quality of clustering for up to three dimensions. Finding clusters of data objects in high dimensional space is challenging, especially considering that such data can be sparse and highly skewed [HAN & KAMBER, 2001]. Palaeogeographic datasets are formed by multi-dimensional instances, including at least a time dimension in millions of years, a georeferenced spatial dimension, and properties. The analysis of palaeogeographic data carried out in our experiments showed the ability of our improvements in handling high dimensional datasets, taking into account temporal, spatial and non-spatial dimensions to cluster similar objects.

- Constraint-based clustering: Real-world applications may need to perform clustering under various kinds of constraints. Suppose that your job is to choose the locations for a given number of new automatic banking machines (ATMs) in a city. To decide upon this, you may cluster households while considering constraints such as the city's rivers and highway networks, and the type and number of customers per cluster. A challenging task is to find groups of data with good clustering behaviour that satisfy specified constraints [HAN & KAMBER, 2001]. In this sense, one of the improvement we carried on ST-DBSCAN to cluster palaeogeographic data is the definition of an age according to a given Epoch in the geologic time scale. Thus, a cluster is formed if it presents similar characteristics regarding the properties of the data, if the instances are geographically close to each other, and if they are situated in the same geological Epoch, which could be a constraint, defined by the user.
- Interpretability and usability: Users expect clustering results to be interpretable, comprehensible, and usable. That is, clustering may need to be tied to specific semantic interpretations and applications. It is important to study how an application goal may influence the selection of clustering features and methods [HAN & KAMBER, 2001]. One of the features implemented in our solution is the visualization of the data clustered in maps. This way, clusters are easily analysed regarding their locations, being their instances presented in specific colours in order to be different from other clusters.

Thus, the aim of this work is a solution for spatio-temporal data mining with density-based clustering algorithms that takes into account the characteristic of palaeogeographic data: both in relation to distortions in distance calculations, and the approach of the geological time. It aims to increase the quality in the results of distance calculation and thus reflect an increased efficiency of the spatio-temporal data mining algorithm, as well as facilitate the work of domain experts situating their research in an approach that takes into account the granularity of geologic time. The development of this solution demonstrates the scientific relevance of this topic. It was developed aiming to meet a set of specifications found in literature on what is expected from a good clustering algorithm, from which almost all the characteristics were met, except those defined in the next section.

## 7.2 Future Works

From the work developed during this Master's research, it is possible to identify some possibilities and directions to work in the future. From the list presented in the previous section on typical requirements of clustering in data mining, it was not possible to address two items, due to the need for more time to study and develop an appropriate approach in this sense. The first characteristic, scalability, gives respect to the fact that many clustering algorithms work well on small data sets containing fewer than several hundred data objects; however, a large database may contain millions of

objects. Clustering on a sample of a given large data set may lead to biased results. Highly scalable clustering algorithms are needed [HAN & KAMBER, 2001]. In this sense, the solution developed showed to work well on datasets of hundreds of thousand objects, although the execution on bigger datasets, with more than a million instances, presented scalability problems. An alternative for optimization could be to consider a Map Reduce or similar technique to handle Big Data, seeking for more efficiency in the data mining process using a parallel approach. The item regarding the second characteristic listed but still not addressed in this research, incremental clustering and insensitivity to the order of input records, states that some clustering algorithms cannot incorporate newly inserted data (i.e., database updates) into existing clustering structures and, instead, must determine a new clustering from scratch. Some clustering algorithms are sensitive to the order of input data. That is, given a set of data objects, such an algorithm may return dramatically different clustering depending on the order of presentation of the input objects. It is important to develop incremental clustering algorithms and algorithms that are insensitive to the order of input [HAN & KAMBER, 2001]. Typically, palaeogeographic databases are fed by the execution of palaeoclimatic and palaeogeographic reconstruction and simulation models, and new objects can be included on the database coming from new executions. The treatment of new input data is another task to be addressed.

Other works aimed to be developed in the future include the total inclusion of the adapted ST-DBSCAN algorithm on Weka Package, developing an interface, which will allow to the user directly connect to the database and make use of other features of Weka in the pre-processing and data mining processes. Also other characteristics can be included in the possible *no\_spatial* values, making it possible to find groups of similar geographic, temporal and of more than one characteristic, such as bathymetry and age of the rock, or pressure, or depth.

Finally, another need for future study should be concentrated on the improvement of the temporal approach. Implementing the full UML diagram provided by [COX & RICHARDS, 2005], based on the notations of ISO 19108 [ISO, 2002], for instance, would make possible to consider the full hierarchy of the geologic time scale, since the Epoch is the only temporal subdivision considered in the current version. Still, the use of a distance function to measure temporal similarity could be more consistent with density-based clustering, also allowing clusters that traverse different ages and epochs.

## REFERENCES

ANDRIENKO, G. L.; MALERBA, D.; MAY, M.; TEISSEIRE, M., Mining spatio-temporal data. In: *Journal Intelligent Information Systems*, 2006, pp. 187-190.

ANKERST, M.; BREUNIG, M.; KRIEGEL, H.P.; SANDER, J., OPTICS: Ordering Objects to Identify the Clustering Structure. In: *Proceedings of International Conference on Management of Data ACM SIGMOD*, 1999, pp. 49-60.

BIRANT, D.; KUT, A., ST-DBSCAN: An algorithm for clustering spatial-temporal data. In: *Data & Knowledge Engineering*, 2007, pp. 208-221.

CHEN, M.; HAN, J.; YU, P. S., Data Mining: An Overview from Database Perspective, *IEEE Transactions on Knowledge and Data Engineering*, 1996, pp. 866-883.

COX, S. J. D.; RICHARD, S. M., A formal model for the geologic time scale and global stratotype section and point, compatible with geospatial information transfer standards. In: *Geosphere*, 2005, pp. 119-137.

DALAZOANA, R.; FREITAS, S.R.C., Efeitos na cartografia devido à evolução do Sistema Geodésico Brasileiro e adoção de um referencial geocêntrico. In: *Revista Brasileira de Cartografia* 54, 2002, pp. 66-76.

DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B., Maximum Likelihood from Incomplete Data via the EM Algorithm, In: *Journal of the Royal Statistical Society*, 1977, pp. 1-38.

ESTER, M.; FROMMELT, A.; KRIEGEL, H. P.; SANDER, J., Spatial Data Mining: Database Primitives, Algorithms and Efficient DBMS Support, In: *Data Mining and Knowledge Discovery 4*, 2000, pp. 193-216.

ESTER, M.; KRIEGEL, H. P.; SANDER, J.; XU, X., A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226-231.

ESTER, M.; KRIEGEL, H. P.; SANDER, J., Algorithms and Applications for Spatial Data Mining, In: *Geographic Data Mining and Knowledge Discovery*, London and New York, UK and USA: Taylor and Francis, 2001.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P., From data mining to knowledge discovery - a review. In: *Advances in knowledge discovery*, 1996, pp. 1-33.

GASPAR, J. A., *Cartas e Projecções Cartográficas*, Lisboa: Lidel, 2005, 332p.

GOOGLE EARTH, Google Developers, 2013. Accessible on <http://www.google.com.br/intl/pt-BR/earth/>. Accessed on November 2013.

GRADSTEIN, F.M.; OGG, J.G.; SCHMITZ, M.D., *et. al*, *The Geologic Time Scale 2012*. Boston: Elsevier, 2012.

GROTZINGER, J.; JORDAN, T. H.; PRESS, F.; SIEVER, R., *Understanding Earth*. New York: W.H. Freeman & Co., 2007, 549p.

HAN, J.; KAMBER, M., *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann, 2001, 550p.

HAN, J.; KAMBER, M., *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann, 2nd edition, 2006.

HAN, J.; KAMBER, M.; PEI, J., *Data Mining: Concepts and Techniques*, San Francisco: Morgan Kaufmann, 3rd edition, 2012, 703p.

HAN, J.; KAMBER, M.; & TUNG, A. K. H., Spatial clustering methods in data mining: A survey. In: Miller, H. J., Han, J., Ed; *Geographic data mining and knowledge discovery*, London and New York, UK and USA: Taylor and Francis, 2001, pp. 33–50.

IBGE, ProGrid, 2008. Accessible on: [ftp://geofpt.ibge.gov.br/aplicativos/transformacao\\_coordenadas/](ftp://geofpt.ibge.gov.br/aplicativos/transformacao_coordenadas/). Accessed on: December, 2013.

ICS, International Chronostratigraphical Chart, International Commission on Stratigraphy, 2012. Accessible on: <http://stratigraphy.org/ICSchart/ChronostratChart2012.pdf>. Accessed on April 2012.

IBM Corporation, DB2(R) Information Center, 2005. Accessible on: <http://publib.boulder.ibm.com/infocenter/db2luw/v8/index.jsp?topic=/com.ibm.db2.udb.doc/opt/csb3022a.htm>. Accessed on June 2013.

ISO, ISO 19108:2002: Geographic information - Temporal schema, 2002, 48p.

KALYANI, D., A Survey on Spatio-Temporal Data Mining. In: *International Journey of Computer Science and Network*, 2012, pp. 140-147.

KAUFMAN, L.; ROUSSEUW, P. J., Clustering Large Applications (Program CLARA). In: *Finding Groups in Data: An Introduction to Cluster Analysis*, Hoboken: John Wiley & Sons, 1990, 368p.

KML Documentation Introduction, Google Developers, 2013. Accessible on <https://developers.google.com/kml/documentation/>. Accessed on November 2013.

KOPERSKI, K.; ADHIKARY, J.; HAN, J., Spatial data mining: progress and challenges. In: *Proceedings ACM SIGMOD Workshop on Research Issues on Mining and Knowledge Discovery*, 1996, pp.1-10.

LIBAUT, A., *La cartographie*. Paris: Presses Universitaire de France, 1966, 937p.

MARTONNE, E., *Panorama da Geografia*. Lisboa: Edições Cosmos, 1953, 679p.

MILLER, H.; HAN, J., Geographic data mining and knowledge discovery: An overview. In: *Geographic data mining and knowledge discovery*, London and New York, UK and USA: Taylor and Francis, 2009, pp. 1-26.

MÜLLER, R.D.; SDROLIAS, M., GAINA; C. STEINBERGER, B.; HEINE, C., Long-Term Sea-Level Fluctuations Driven by Ocean Basin Dynamics, *Science*, 2008, pp. 1357-1362.

NetBeans. Oracle Corporation, 2013., Accessible on <https://netbeans.org/>. Accessed on April 2013.

NG, R.; HAN, J., Efficient and Effective Clustering Methods for Spatial Data Mining, In: *Proceedings of 20th International Conference on Very Large DataBases*, 1994, pp. 144-155.



NG, R.; HAN, J., CLARANS: A Method for Clustering Objects for Spatial Data Mining. In: *IEEE Transactions on Knowledge and Data Engineering*, 2002, pp. 1003-1016.

NOAA - National Oceanic and Atmospheric Administration, National Geophysical Datacenter, 2013. Accessible on <http://maps.ngdc.noaa.gov/>. Accessed on December 2013.

NOAA - National Oceanic and Atmospheric Administration, 2014. Accessible on <http://oceanservice.noaa.gov/facts/bathymetry.html>. Accessed on January 2014.

NÜRNBERG, D.; MÜLLER, R.D., The Tectonic Evolution of the South Atlantic from Late Jurassic to Present. In: *Tectonophysics*, 1991, pp. 27-53.

OGG, J.G; GRADSTEIN, F.M., Geologic Time Scale 2004 - why, how, and where next!, In: *Lethaia* 37, 2004, pp. 175-181.

OGG, J.G.; GRADSTEIN, F.M.; SMITH, A.G., A Geologic Time Scale 2004. Cambridge: Cambridge University Press, 2004.

OGG, J. G.; OGG, G.; GRADSTEIN, F. M., The Concise Geologic Time scale. Cambridge University Press, 2008, 150p.

RAINSFORD, H. F., Long geodesics on the ellipsoid. In: *Geodesy*, 1955, pp. 12–22.

RODDICK, J. F.; LEES, B. G., Paradigms for spatial and spatio-temporal data mining. In: *Geographic Data Mining and Knowledge Discovery*, London and New York, UK and USA: Taylor and Francis, 2001, pp. 30-49.

SAMPLE, J. T.; IOUP, E., Tile-Based Geospatial Information Systems: Principles and Practices. New York: Springer, 2010, 238p.

SANTOS, M. C. S. R., Manual de fundamentos cartográficos e diretrizes gerais para elaboração de mapas geológicos, geomorfológicos e geotécnicos, São Paulo: Secretaria da Ciência, Tecnologia e Desenvolvimento Econômico, Programa de Desenvolvimento de Recursos Minerais - Pró-Minério, Instituto de Pesquisas Tecnológicas, 1989, 60p.

SHEKHAR, S.; HUANG, Y., Discovering spatial co-location patterns: A summary of results. In: *Advances in spatial and temporal databases, proceedings, lecture notes in computer science*, 2001, pp. 236–256.

SHUA, H; ZHUB, X.; DAIC, S., Mining Association Rules In Geographical Spatio-Temporal Data. In: *The International Archives of the Photometry, Remote Sensing and Spatial Information Sciences*, 2008, pp. 225-228.

SINNOT, R.W., Virtues of the Haversine, In: *Sky and Telescope*, 1984, 159p.

TAN, P. N.; STEINBACH, M.; KUMAR, V., Introdução ao Data Mining. São Paulo: Ciência Moderna, 2009, 928p.

TEIXEIRA, W. *et. al*, Decifrando a Terra. São Paulo: Nacional, 2009, 624p.

THEODORIDIS, S.; KOUTROUMBAS, K., Pattern Recognition. Orlando: Academic Press, 2006, 635p.

TOBLER, W., A computer movie simulating urban growth in the Detroit region. In: *Economic Geography* 46, 1970, pp. 234-240.

VINCENY, T., Direct and Inverse Solutions of Geodesics on the Ellipsoid with application of nested equations, 1975, pp. 88–93.

WANG, X.; HAMILTON, H. J., DBRS: A Density-Based Spatial Clustering Method with Random Sampling. In: *Proceedings of the 7th PAKDD*, 2003, pp. 563-575.

WITTEN, I.; FRANK, E., Data mining: practical machine learning tools and techniques with java implementations. San Francisco: Morgan Kaufmann, 2000, 371p.

XU, R.; WUNSCH II, D., Survey of Clustering Algorithms, In: *IEEE Transactions on Neural Networks*, 2005, pp. 645-678.

**APPENDIX A**  
**CLUSTERS FORMED IN EXPERIMENT 2**

**Table 12 - Objects of the clusters formed by the execution of Haversine formula on the filtered *muller\_bathymetry* dataset, using *Eocene* as *eps3* parameter.**

<b>Clusters - <i>Eocene</i></b>
cluster 1
32 - 33 - 116 - 117 - 201 - 202 - 287 - 288 - 373 - 374 - 460 - 461 - 548 - 549 - 635 - 636 - 723 - 724 - 812 - 813 - 897 - 898 - 981 - 982 - 1066 - 1067 - 1152 - 1153 - 1238 - 1239 - 1324 - 1325 - 1412 - 1413 - 1499 - 1500 - 1586 - 1587 - 1675 - 1676 - 1760 - 1761 - 1846 - 1847 - 1932 - 1933 - 1934 - 2018 - 2019 - 2020 - 2104 - 2105 - 2106 - 2191 - 2192 - 2193 - 2277 - 2278 - 2279 - 2365 - 2366 - 2452 - 2453 - 2454 - 2540 - 2541 - 2542 - 2627 - 2712 - 2796 - 2797 - 2798 - 2882 - 2884 - 2969 - 2970 - 3055 - 3056 - 3057 - 3143 - 3144 - 3230 - 3231 - 3318 - 3319 - 3320 - 3407 - 3408 - 3578 - 3750 - 4010 - 4096 - 4274 -
cluster 2
34 - 35 - 118 - 119 - 203 - 204 - 289 - 290 - 375 - 376 - 462 - 463 - 464 - 550 - 551 - 637 - 638 - 639 - 725 - 726 - 727 - 814 - 815 - 899 - 900 - 983 - 984 - 1068 - 1069 - 1154 - 1155 - 1240 - 1241 - 1326 - 1327 - 1328 - 1414 - 1415 - 1501 - 1502 - 1503 - 1588 - 1589 - 1590 - 1677 - 1678 - 1679 - 1762 - 1763 - 1848 - 1849 - 1935 - 1936 - 2021 - 2022 - 2107 - 2108 - 2109 - 2194 - 2195 - 2280 - 2281 - 2367 - 2368 - 2369 - 2455 - 2456 - 2543 - 2544 - 2545 - 2629 - 2714 - 2800 - 2886 - 2972 - 2973 - 3058 - 3059 - 3145 - 3146 - 3147 - 3232 - 3233 - 3234 - 3321 - 3322 - 3409 - 3410 - 3752 - 3838 - 3925 - 4012 - 4013 - 4100 - 4188 -
cluster 3
36 - 37 - 120 - 121 - 122 - 205 - 206 - 207 - 291 - 292 - 377 - 378 - 465 - 466 - 552 - 553 - 554 - 640 - 728 - 816 - 901 - 902 - 985 - 986 - 1070 - 1071 - 1072 - 1156 - 1157 - 1242 - 1243 - 1329 - 1330 - 1416 - 1417 - 1504 - 1591 - 1764 - 1765 - 1850 - 1851 - 1937 - 1938 - 2023 - 2024 - 2110 - 2111 - 2196 - 2282 - 2457 - 2715 - 2802 - 2887 - 2888 - 2974 - 3060 - 3411 - 3583 - 3666 - 3839 - 3926 - 4879 -
cluster 4
48 - 49 - 132 - 133 - 217 - 218 - 303 - 304 - 389 - 390 - 476 - 477 - 564 - 565 - 651 - 652 - 739 - 740 - 741 - 827 - 828 - 829 - 912 - 913 - 997 - 1082 - 1083 - 1168 - 1254 - 1340 - 1428 - 1515 - 1602 - 1604 - 1690 - 1691 - 1776 - 1862 - 1948 - 2381 - 2642 -
cluster 5
50 - 134 - 135 - 219 - 220 - 305 - 306 - 391 - 392 - 478 - 479 - 566 - 567 - 653 - 654 - 742 - 830 - 831 - 915 - 999 - 1000 - 1084 - 1085 - 1170 - 1171 - 1256 - 1257 - 1342 - 1343 - 1430 - 1431 - 1517 - 1518 - 1605 - 1693 - 1694 - 1778 - 1779 - 1864 - 1951 - 2037 - 2038 - 2123 - 2124 - 2210 - 2211 - 2296 - 2297 - 2383 - 2384 - 2471 - 2559 - 2560 - 2644 - 2729 - 2901 - 2987 - 3074 - 3161 - 3249 - 3337 - 3338 - 3425 - 3426 - 3510 - 3595 - 3767 - 3853 - 4027 - 4114 - 4203 - 4291 - 4893 - 4980 - 5069 - 5157 -
cluster 6
51 - 52 - 136 - 137 - 221 - 222 - 307 - 308 - 393 - 394 - 480 - 481 - 568 - 655 - 743 - 916 - 917 - 1001 - 1002 - 1086 - 1087 - 1172 - 1258 - 1259 - 1344 - 1345 - 1432 - 1519 - 1606 - 1780 - 1866 - 1867 - 1952 - 1953 - 2039 - 2125 - 2126 - 2212 - 2298 - 2385 - 2472 - 2646 - 2647 - 2731 - 2732 - 2817 - 2903 - 2990 - 3076 - 3163 - 3250 - 3512 - 3598 - 3683 - 3770 - 3856 - 3942 - 4029 - 4722 - 4808 -
cluster 7
40 - 41 - 42 - 124 - 125 - 126 - 209 - 210 - 211 - 294 - 295 - 296 - 381 - 382 - 469 - 556 - 905 - 906 - 989 - 990 - 1074 - 1075 - 1159 - 1160 - 1161 - 1767 - 1768 - 1769 - 1854 - 1855 - 1940 - 1941 - 2027 - 2113 - 2633 - 2634 - 2635 - 2719 - 2720 - 3500 - 3585 - 3586 -

## cluster 8

43 - 44 - 128 - 129 - 213 - 214 - 298 - 299 - 384 - 385 - 472 - 560 - 908 - 909 - 993 - 1078 - 1163 -  
1164 - 1250 - 1336 - 1771 - 1772 - 1857 - 1858 - 1943 - 1944 - 2030 - 2031 - 2637 - 2638 - 2722 -  
2723 - 2808 - 2894 - 3503 - 3504 - 3588 - 4369 - 4370 - 5235 -

## cluster 9

38 - 39 - 123 - 208 - 293 - 379 - 380 - 467 - 730 - 903 - 904 - 987 - 988 - 1073 - 1158 - 1244 - 1245 -  
1331 - 1418 - 1766 - 1852 - 1853 - 1939 - 2025 - 2284 - 2632 - 2717 - 2718 - 2803 - 2889 - 3498 -  
3755 - 4364 - 5230 -

## cluster 10

127 - 212 - 297 - 383 - 470 - 471 - 558 - 645 - 733 - 907 - 991 - 992 - 1076 - 1077 - 1162 - 1248 - 1334  
- 1509 - 1596 - 1684 - 1770 - 1856 - 1942 - 2028 - 2029 - 2115 - 2201 - 2462 - 2636 - 2721 - 2806 -  
2892 - 3065 - 3152 - 3239 - 3501 - 3502 - 3587 - 3672 - 3758 - 4018 - 4367 - 4368 - 4453 - 4538 -  
4624 - 4710 - 5233 - 5489 -

## cluster 11

45 - 46 - 130 - 215 - 216 - 300 - 386 - 387 - 473 - 474 - 561 - 648 - 736 - 910 - 911 - 994 - 995 - 1079 -  
1080 - 1165 - 1251 - 1337 - 1338 - 1773 - 1774 - 1859 - 1860 - 1945 - 1946 - 2032 - 2118 - 2204 -  
2639 - 2724 - 2725 - 2809 - 2810 - 3505 - 3590 - 3675 - 3676 - 4371 - 4456 -

## cluster 12

286 - 372 - 459 - 546 - 547 - 633 - 634 - 721 - 722 - 809 - 810 - 811 - 1151 - 1237 - 1323 - 1410 - 1411  
- 1497 - 1498 - 1584 - 1585 - 1672 - 1673 - 1674 - 2190 - 2276 - 2363 - 2364 - 2450 - 2451 - 2538 -  
2539 - 3228 - 3316 - 3404 - 3405 - 4182 -

## cluster 13

47 - 131 - 301 - 302 - 388 - 475 - 562 - 563 - 649 - 650 - 738 - 826 - 996 - 1081 - 1166 - 1167 - 1252 -  
1253 - 1339 - 1426 - 1427 - 1513 - 1514 - 1601 - 1689 - 1775 - 1861 - 1947 - 2033 - 2119 - 2120 -  
2206 - 2292 - 2379 - 2380 - 2467 - 2555 - 2640 - 2641 - 2726 - 2811 - 2897 - 2983 - 3070 - 3333 -  
3506 - 3507 - 3591 - 3592 - 3677 - 3763 - 3849 - 4372 - 4457 - 4458 - 4543 - 5238 - 5323 - 5409 -

## cluster 14

468 - 555 - 642 - 643 - 731 - 819 - 1246 - 1332 - 1419 - 1506 - 1507 - 1593 - 1594 - 1682 - 2026 -  
2112 - 2198 - 2199 - 2285 - 2372 - 2459 - 2460 - 2547 - 2548 - 2804 - 2890 - 2976 - 3062 - 3063 -  
3149 - 3150 - 3237 - 3325 - 3413 - 3414 - 3499 - 3584 - 3669 - 3670 - 3756 - 3842 - 3928 - 4015 -  
4016 - 4102 - 4103 - 4191 - 4279 - 4280 - 4365 - 4450 - 4535 - 4536 - 4621 - 4622 - 4708 - 4794 -  
4881 - 4882 - 4968 - 4969 - 5056 - 5057 - 5145 - 5231 - 5316 - 5401 - 5486 - 5487 - 5572 - 5573 -  
5659 - 5746 - 5747 - 5833 - 5834 - 5921 - 5922 - 6010 -

## cluster 15

557 - 644 - 732 - 820 - 821 - 1247 - 1333 - 1420 - 1421 - 1508 - 1595 - 1683 - 2114 - 2200 - 2286 -  
2287 - 2373 - 2374 - 2461 - 2549 - 2550 - 2805 - 2891 - 2977 - 2978 - 3064 - 3151 - 3238 - 3326 -  
3327 - 3415 - 3671 - 3757 - 3843 - 3844 - 3929 - 3930 - 4017 - 4104 - 4105 - 4192 - 4193 - 4281 -  
4366 - 4451 - 4452 - 4537 - 4623 - 4709 - 4795 - 4796 - 4883 - 4970 - 4971 - 5058 - 5059 - 5146 -  
5147 - 5232 - 5317 - 5318 - 5402 - 5403 - 5488 - 5574 - 5575 - 5660 - 5661 - 5748 - 5835 - 5923 -  
6011 -

## cluster 16

559 - 646 - 734 - 822 - 823 - 1249 - 1335 - 1422 - 1423 - 1510 - 1597 - 1685 - 2116 - 2202 - 2288 -

2375 - 2376 - 2463 - 2551 - 2807 - 2893 - 2979 - 2980 - 3066 - 3153 - 3240 - 3328 - 3329 - 3416 -  
3417 - 3673 - 3759 - 3845 - 3846 - 3931 - 3932 - 4019 - 4106 - 4194 - 4195 - 4282 - 4283 - 4454 -  
4539 - 4625 - 4711 - 4797 - 4798 - 4884 - 4885 - 4972 - 5060 - 5061 - 5148 - 5149 - 5234 - 5319 -  
5320 - 5404 - 5405 - 5490 - 5576 - 5662 - 5663 - 5749 - 5750 - 5836 - 5837 - 5925 - 6013 -

cluster 17

53 - 54 - 138 - 223 - 918 - 919 - 1003 - 1088 - 1173 - 1781 - 1782 - 1868 - 1954 - 2040 - 2648 - 2733 -  
2818 - 2904 - 3513 - 3514 - 3599 - 3684 - 4379 - 4380 - 4464 - 4465 - 4550 - 4636 - 5245 - 5246 -  
5331 - 5416 - 5501 -

cluster 18

647 - 735 - 824 - 1424 - 1511 - 1598 - 1686 - 1687 - 2117 - 2203 - 2289 - 2290 - 2377 - 2464 - 2552 -  
2895 - 2981 - 3067 - 3154 - 3155 - 3241 - 3242 - 3330 - 3418 - 3589 - 3674 - 3760 - 3761 - 3847 -  
3933 - 4020 - 4107 - 4108 - 4196 - 4284 - 4455 - 4540 - 4626 - 4627 - 4712 - 4713 - 4799 - 4886 -  
4973 - 4974 - 5062 - 5150 - 5236 - 5321 - 5406 - 5491 - 5577 - 5578 - 5664 - 5751 - 5838 - 5839 -  
5926 - 5927 - 6014 - 6015 -

cluster 19

641 - 729 - 817 - 818 - 1505 - 1592 - 1680 - 1681 - 2197 - 2283 - 2370 - 2371 - 2458 - 2546 - 2630 -  
2631 - 2716 - 2801 - 2975 - 3061 - 3148 - 3235 - 3236 - 3323 - 3324 - 3412 - 3495 - 3496 - 3497 -  
3581 - 3582 - 3667 - 3668 - 3753 - 3754 - 3840 - 3841 - 3927 - 4014 - 4101 - 4189 - 4190 - 4277 -  
4278 - 4361 - 4362 - 4363 - 4447 - 4448 - 4449 - 4532 - 4533 - 4534 - 4618 - 4619 - 4620 - 4705 -  
4706 - 4707 - 4791 - 4792 - 4793 - 4880 - 4966 - 4967 - 5055 - 5143 - 5144 - 5227 - 5228 - 5229 -  
5313 - 5314 - 5315 - 5398 - 5399 - 5400 - 5483 - 5484 - 5485 - 5570 - 5571 - 5656 - 5657 - 5658 -  
5744 - 5745 - 5831 - 5832 - 5920 - 6008 - 6009 -

cluster 20

914 - 998 - 1169 - 1255 - 1341 - 1429 - 1516 - 1603 - 1692 - 1777 - 1863 - 1949 - 1950 - 2035 - 2036 -  
2122 - 2208 - 2209 - 2294 - 2295 - 2382 - 2469 - 2470 - 2557 - 2558 - 2643 - 2728 - 2813 - 2814 -  
2899 - 2900 - 2986 - 3072 - 3073 - 3159 - 3160 - 3247 - 3248 - 3335 - 3336 - 3423 - 3424 - 3509 -  
3594 - 3679 - 3680 - 3765 - 3766 - 3852 - 3938 - 3939 - 4025 - 4026 - 4113 - 4201 - 4202 - 4289 -  
4290 - 4374 - 4375 - 4376 - 4460 - 4461 - 4545 - 4546 - 4631 - 4632 - 4633 - 4718 - 4719 - 4804 -  
4805 - 4891 - 4892 - 4979 - 5067 - 5068 - 5155 - 5156 - 5240 - 5241 - 5326 - 5327 - 5411 - 5412 -  
5496 - 5497 - 5582 - 5583 - 5584 - 5669 - 5670 - 5756 - 5757 - 5758 - 5843 - 5844 - 5845 - 5932 -  
5933 - 5934 - 6020 - 6021 - 6022 -

cluster 21

737 - 825 - 1425 - 1512 - 1599 - 1600 - 1688 - 2205 - 2291 - 2378 - 2465 - 2466 - 2553 - 2554 - 2896 -  
2982 - 3068 - 3069 - 3156 - 3243 - 3331 - 3419 - 3420 - 3762 - 3848 - 3934 - 3935 - 4021 - 4022 -  
4109 - 4197 - 4285 - 4541 - 4542 - 4628 - 4714 - 4800 - 4887 - 4888 - 4975 - 5063 - 5151 - 5237 -  
5322 - 5407 - 5408 - 5492 - 5493 - 5579 - 5665 - 5752 - 5840 - 5928 - 6016 -

cluster 22

2034 - 2121 - 2207 - 2293 - 2468 - 2556 - 2727 - 2812 - 2898 - 2984 - 2985 - 3071 - 3158 - 3245 -  
3246 - 3334 - 3422 - 3508 - 3593 - 3678 - 3764 - 3850 - 3851 - 3937 - 4024 - 4111 - 4112 - 4199 -  
4200 - 4288 - 4373 - 4459 - 4544 - 4630 - 4716 - 4717 - 4803 - 4890 - 4977 - 4978 - 5065 - 5066 -  
5154 - 5239 - 5324 - 5325 - 5410 - 5495 - 5581 - 5667 - 5668 - 5755 - 5842 - 5930 - 5931 - 6019 -

cluster 23

2626 - 2711 - 2883 - 3492 - 3577 - 3662 - 3663 - 3749 - 3835 - 3836 - 3922 - 4009 - 4097 - 4185 -  
4273 - 4358 - 4443 - 4444 - 4528 - 4529 - 4615 - 4616 - 4701 - 4702 - 4787 - 4788 - 4875 - 4876 -

4962 - 4963 - 5051 - 5139 - 5140 - 5224 - 5309 - 5310 - 5394 - 5395 - 5480 - 5481 - 5566 - 5567 -  
5652 - 5653 - 5740 - 5741 - 5827 - 5828 - 5916 - 6004 - 6005 -

cluster 24

2628 - 2713 - 2799 - 2885 - 2971 - 3493 - 3494 - 3579 - 3580 - 3664 - 3665 - 3751 - 3837 - 3923 -  
3924 - 4011 - 4098 - 4099 - 4186 - 4187 - 4275 - 4276 - 4359 - 4360 - 4445 - 4446 - 4530 - 4531 -  
4617 - 4703 - 4704 - 4789 - 4790 - 4877 - 4878 - 4964 - 4965 - 5052 - 5053 - 5054 - 5141 - 5142 -  
5225 - 5226 - 5311 - 5312 - 5396 - 5397 - 5482 - 5568 - 5569 - 5654 - 5655 - 5742 - 5743 - 5829 -  
5830 - 5917 - 5918 - 5919 - 6006 - 6007 -

cluster 25

1865 - 2645 - 2730 - 2815 - 2816 - 2902 - 2988 - 2989 - 3075 - 3162 - 3511 - 3596 - 3597 - 3681 -  
3682 - 3768 - 3769 - 3854 - 3855 - 3940 - 3941 - 4028 - 4115 - 4116 - 4204 - 4292 - 4377 - 4378 -  
4462 - 4463 - 4547 - 4548 - 4549 - 4634 - 4635 - 4720 - 4721 - 4806 - 4807 - 4894 - 4895 - 4981 -  
4982 - 5070 - 5158 - 5242 - 5243 - 5244 - 5328 - 5329 - 5413 - 5414 - 5415 - 5498 - 5499 - 5500 -  
5585 - 5586 - 5587 - 5671 - 5672 - 5673 - 5759 - 5760 - 5846 - 5847 - 5935 - 6023 -

cluster 26

2968 - 3054 - 3141 - 3142 - 3229 - 3317 - 3406 - 3748 - 3834 - 3920 - 3921 - 4007 - 4008 - 4094 -  
4095 - 4183 - 4184 - 4270 - 4271 - 4272 - 4614 - 4700 - 4786 - 4873 - 4874 - 4960 - 4961 - 5048 -  
5049 - 5050 - 5136 - 5137 - 5138 - 5479 - 5565 - 5651 - 5738 - 5739 - 5825 - 5826 - 5913 - 5914 -  
5915 - 6001 - 6002 - 6003 -

cluster 27

3157 - 3244 - 3332 - 3421 - 3936 - 4023 - 4110 - 4198 - 4286 - 4287 - 4629 - 4715 - 4801 - 4802 -  
4889 - 4976 - 5064 - 5152 - 5153 - 5494 - 5580 - 5666 - 5753 - 5754 - 5841 - 5929 - 6017 - 6018 -

**Table 13 - Objects of the clusters formed by the execution of Haversine formula on the filtered *muller\_bathymetry* dataset, using *Paleocene* as *eps3* parameter.**

**Clusters - *Paleocene***

cluster 1

55 - 56 - 139 - 140 - 224 - 225 - 310 - 311 - 396 - 397 - 483 - 484 - 571 - 572 - 658 - 659 - 746 - 747 -  
748 - 834 - 835 - 836 - 920 - 921 - 1004 - 1005 - 1089 - 1090 - 1175 - 1176 - 1261 - 1262 - 1347 -  
1348 - 1435 - 1436 - 1522 - 1523 - 1609 - 1610 - 1611 - 1698 - 1699 - 1783 - 1784 - 1869 - 1870 -  
1955 - 1956 - 1957 - 2041 - 2042 - 2043 - 2128 - 2129 - 2214 - 2215 - 2301 - 2302 - 2388 - 2389 -  
2475 - 2476 - 2477 - 2564 - 2565 - 2650 - 2735 - 2819 - 2820 - 2821 - 2905 - 2906 - 2907 - 2992 -  
2993 - 3078 - 3079 - 3080 - 3165 - 3166 - 3167 - 3253 - 3254 - 3255 - 3341 - 3342 - 3343 - 3430 -  
3431 - 3516 - 3601 - 3773 - 3859 - 3944 - 3945 - 3946 - 4031 - 4032 - 4033 - 4119 - 4120 - 4121 -  
4207 - 4208 - 4209 - 4295 - 4296 - 4297 - 4639 - 4725 - 4899 - 4986 - 5073 - 5074 - 5075 - 5161 -  
5162 - 5163 - 5764 - 5851 - 5940 - 6028 -

cluster 2

57 - 141 - 142 - 226 - 227 - 312 - 313 - 398 - 399 - 485 - 486 - 573 - 574 - 660 - 661 - 749 - 750 - 837 -  
838 - 922 - 1006 - 1007 - 1091 - 1092 - 1177 - 1178 - 1263 - 1264 - 1349 - 1350 - 1437 - 1438 - 1524 -  
1525 - 1612 - 1613 - 1700 - 1701 - 1785 - 1786 - 1871 - 1872 - 1958 - 1959 - 2044 - 2045 - 2130 -  
2131 - 2216 - 2217 - 2218 - 2303 - 2304 - 2390 - 2391 - 2392 - 2478 - 2479 - 2566 - 2567 - 2652 -  
2737 - 2822 - 2823 - 2908 - 2909 - 2994 - 2995 - 3081 - 3082 - 3168 - 3169 - 3256 - 3257 - 3344 -  
3345 - 3432 - 3433 - 3603 - 3775 - 4035 - 4211 -

cluster 3

58 - 59 - 143 - 144 - 228 - 229 - 314 - 315 - 400 - 401 - 487 - 488 - 575 - 576 - 662 - 663 - 751 - 752 -  
839 - 840 - 923 - 924 - 1008 - 1009 - 1093 - 1094 - 1179 - 1180 - 1265 - 1266 - 1351 - 1352 - 1439 -  
1440 - 1526 - 1527 - 1614 - 1615 - 1702 - 1703 - 1787 - 1788 - 1873 - 1874 - 1960 - 1961 - 2046 -  
2047 - 2132 - 2133 - 2219 - 2220 - 2305 - 2306 - 2393 - 2394 - 2480 - 2481 - 2568 - 2569 - 2654 -  
2739 - 2825 - 2911 - 3084 - 3171 - 3259 - 3347 -

cluster 4

60 - 61 - 145 - 146 - 230 - 231 - 316 - 317 - 402 - 403 - 489 - 490 - 577 - 578 - 664 - 665 - 753 - 841 -  
925 - 926 - 1010 - 1011 - 1095 - 1096 - 1181 - 1182 - 1267 - 1268 - 1353 - 1354 - 1441 - 1442 - 1528 -  
1529 - 1616 - 1704 - 1789 - 1790 - 1875 - 1876 - 1962 - 1963 - 2048 - 2049 - 2134 - 2135 - 2221 -  
2222 - 2307 - 2308 - 2395 - 2482 - 2570 - 2656 - 2741 - 2827 - 2913 - 3086 - 3173 -

cluster 5

62 - 63 - 64 - 147 - 148 - 232 - 233 - 318 - 404 - 491 - 927 - 928 - 929 - 1012 - 1013 - 1097 - 1098 -  
1183 - 1269 - 1355 - 1791 - 1792 - 1877 - 1878 - 1964 - 2050 - 2136 - 2658 - 2743 -

cluster 6

309 - 395 - 482 - 569 - 570 - 656 - 657 - 744 - 745 - 832 - 833 - 1174 - 1260 - 1346 - 1433 - 1434 -  
1520 - 1521 - 1607 - 1608 - 1695 - 1696 - 1697 - 2127 - 2213 - 2299 - 2300 - 2387 - 2474 - 2562 -  
2563 - 2991 - 3077 - 3252 - 3340 - 3428 - 3429 - 3771 - 3857 - 3943 - 4118 - 4206 - 4897 - 4984 -  
5072 - 6026 -

cluster 7

2386 - 2473 - 2561 - 3164 - 3251 - 3339 - 3427 - 4030 - 4117 - 4205 - 4293 - 4294 - 4723 - 4809 -  
4896 - 4983 - 5071 - 5159 - 5160 - 5588 - 5674 - 5761 - 5848 - 5849 - 5936 - 5937 - 6024 - 6025 -

cluster 8

2649 - 2734 - 3515 - 3600 - 3685 - 3686 - 3772 - 3858 - 4381 - 4382 - 4466 - 4467 - 4551 - 4552 -  
4637 - 4638 - 4724 - 4810 - 4811 - 4898 - 4985 - 5247 - 5248 - 5332 - 5333 - 5417 - 5418 - 5502 -  
5503 - 5504 - 5589 - 5590 - 5675 - 5676 - 5762 - 5763 - 5850 - 5938 - 5939 - 6027 -

cluster 9

2651 - 2736 - 3517 - 3602 - 3687 - 3688 - 3774 - 3860 - 3861 - 3947 - 4034 - 4122 - 4210 - 4298 -  
4299 - 4383 - 4468 - 4469 - 4553 - 4554 - 4640 - 4641 - 4726 - 4727 - 4812 - 4813 - 4900 - 4901 -  
4987 - 4988 - 5076 - 5077 - 5164 - 5165 - 5249 - 5334 - 5335 - 5419 - 5420 - 5505 - 5506 - 5591 -  
5592 - 5677 - 5678 - 5765 - 5766 - 5852 - 5853 - 5941 - 5942 - 6029 - 6030 -

cluster 10

2653 - 2738 - 2824 - 2910 - 2996 - 2997 - 3083 - 3170 - 3258 - 3346 - 3434 - 3435 - 3518 - 3519 -  
3604 - 3605 - 3689 - 3690 - 3776 - 3777 - 3862 - 3863 - 3948 - 3949 - 4036 - 4037 - 4123 - 4124 -  
4212 - 4213 - 4300 - 4301 - 4384 - 4385 - 4470 - 4471 - 4555 - 4556 - 4642 - 4643 - 4728 - 4729 -  
4814 - 4815 - 4902 - 4903 - 4989 - 4990 - 5078 - 5079 - 5166 - 5167 - 5250 - 5251 - 5336 - 5337 -  
5421 - 5422 - 5507 - 5508 - 5593 - 5594 - 5679 - 5680 - 5767 - 5768 - 5854 - 5855 - 5943 - 5944 -  
6031 - 6032 -

cluster 11

2655 - 2740 - 2826 - 2912 - 2998 - 2999 - 3085 - 3172 - 3260 - 3348 - 3436 - 3520 - 3521 - 3606 -  
3607 - 3691 - 3692 - 3778 - 3779 - 3864 - 3865 - 3950 - 3951 - 4038 - 4039 - 4125 - 4126 - 4214 -  
4302 - 4386 - 4387 - 4472 - 4473 - 4557 - 4558 - 4644 - 4645 - 4730 - 4731 - 4816 - 4817 - 4904 -  
4905 - 4991 - 4992 - 5080 - 5168 - 5252 - 5253 - 5338 - 5339 - 5423 - 5424 - 5509 - 5510 - 5595 -  
5596 - 5681 - 5682 - 5769 - 5770 - 5856 - 5857 - 5945 - 6033 -



cluster 12  
 2657 - 2742 - 2828 - 2914 - 3000 - 3522 - 3523 - 3524 - 3608 - 3609 - 3693 - 3694 - 3780 - 3866 -  
 3952 - 4388 - 4389 - 4390 - 4474 - 4475 - 4559 - 4560 - 4646 - 4732 - 4818 - 5254 - 5255 - 5256 -  
 5340 - 5341 - 5425 - 5426 - 5511 - 5597 - 5683

**Table 14 - Objects of the clusters formed by the execution of Haversine formula on the filtered *muller\_bathymetry* dataset, using *Upper Cretaceous* as *eps3* parameter.**

<b>Clusters - <i>Upper Cretaceous</i></b>
cluster 1
65 - 149 - 150 - 234 - 235 - 320 - 321 - 406 - 407 - 493 - 494 - 581 - 582 - 668 - 669 - 757 - 845 - 846 - 930 - 1014 - 1015 - 1099 - 1100 - 1185 - 1186 - 1271 - 1272 - 1357 - 1358 - 1445 - 1446 - 1532 - 1533 - - 1620 - 1708 - 1709 - 1793 - 1794 - 1879 - 1880 - 1965 - 1966 - 1967 - 2051 - 2052 - 2053 - 2138 - 2139 - 2224 - 2225 - 2226 - 2310 - 2311 - 2312 - 2398 - 2399 - 2485 - 2486 - 2487 - 2574 - 2575 - 2660 - 2745 - 2829 - 2831 - 2915 - 2917 - 3088 - 3090 - 3175 - 3177 - 3263 - 3351 - 3353 - 3441 -
cluster 2
66 - 67 - 68 - 151 - 152 - 236 - 237 - 322 - 323 - 408 - 409 - 495 - 496 - 583 - 670 - 671 - 758 - 759 - 847 - 931 - 932 - 933 - 1016 - 1017 - 1101 - 1102 - 1187 - 1188 - 1273 - 1274 - 1359 - 1360 - 1447 - 1534 - 1535 - 1621 - 1622 - 1710 - 1795 - 1881 - 1968 - 2054 - 2140 - 2227 - 2313 - 2400 - 2576 - 2661 - 2746 - 2918 - 3004 - 3178 - 3265 - 3526 - 3527 - 3612 - 3697 - 3784 - 3870 - 3956 - 4044 - 4131 - 4219 -
cluster 3
69 - 153 - 238 - 239 - 324 - 410 - 497 - 584 - 760 - 848 - 934 - 1018 - 1103 - 1104 - 1189 - 1275 - 1361 - 1448 - 1623 - 1711 - 1797 - 1883 - 1884 - 1970 - 2056 - 2142 - 2228 - 2402 - 2489 - 2577 - 2663 - 2834 - 2920 - 3006 - 3355 - 3443 - 3529 - 3786 - 3872 - 4309 -
cluster 4
71 - 155 - 240 - 326 - 412 - 499 - 586 - 673 - 761 - 850 - 936 - 1020 - 1105 - 1191 - 1277 - 1363 - 1450 - 1537 - 1624 - 1713 - 1799 - 1885 - 1971 - 1972 - 2057 - 2058 - 2144 - 2230 - 2316 - 2403 - 2490 - 2491 - 2579 - 2835 - 2836 - 2921 - 2922 - 3008 - 3094 - 3181 - 3268 - 3356 - 3357 - 3444 - 3445 - 3788 - 3874 - 4222 - 4310 - 4311 -
cluster 5
72 - 156 - 241 - 327 - 413 - 500 - 587 - 674 - 762 - 851 - 937 - 1021 - 1106 - 1192 - 1278 - 1364 - 1451 - 1538 - 1625 - 1714 - 1800 - 1886 - 2231 - 2317 - 2404 - 2666 - 2751 - 3182 - 3269 - 3532 - 3617 - 3702 - 4048 - 4135 - 4223 -
cluster 6
73 - 157 - 242 - 328 - 414 - 501 - 588 - 675 - 763 - 852 - 938 - 1022 - 1107 - 1193 - 1279 - 1365 - 1452 - 1539 - 1626 - 1715 - 1801 - 1887 - 2232 - 2318 - 2405 - 2581 - 2667 - 3183 - 3533 - 4049 -
cluster 7
74 - 158 - 243 - 329 - 415 - 502 - 589 - 676 - 764 - 853 - 939 - 1023 - 1108 - 1194 - 1280 - 1366 - 1453 - 1540 - 1627 - 1716 - 1802 - 1888 - 1975 - 2061 - 2147 - 2233 - 2319 - 2406 - 2493 - 2494 - 2582 - 2668 - 3011 - 3097 - 3184 - 3271 - 3534 - 3963 - 4050 - 4829 - 4916 -
cluster 8

75 - 159 - 244 - 330 - 416 - 503 - 590 - 677 - 765 - 854 - 940 - 1024 - 1109 - 1195 - 1281 - 1367 - 1454  
- 1541 - 1628 - 1717 - 1803 - 1889 - 2234 - 2320 - 2407 - 2583 - 2669 - 3185 -

cluster 9

76 - 160 - 245 - 331 - 417 - 504 - 591 - 678 - 766 - 855 - 941 - 1025 - 1110 - 1196 - 1282 - 1368 - 1455  
- 1542 - 1629 - 1718 - 1804 - 1890 - 1977 - 2063 - 2149 - 2235 - 2321 - 2408 - 2495 - 2496 - 2584 -  
2670 - 3013 - 3099 - 3186 - 3273 - 3536 - 3965 - 4052 - 4831 - 4918 -

cluster 10

77 - 161 - 246 - 332 - 418 - 505 - 592 - 679 - 767 - 856 - 942 - 1026 - 1111 - 1197 - 1283 - 1369 - 1456  
- 1543 - 1630 - 1719 - 1805 - 1891 - 2236 - 2322 - 2409 - 2671 - 3187 -

cluster 11

78 - 162 - 247 - 333 - 419 - 506 - 593 - 680 - 768 - 769 - 857 - 943 - 1027 - 1112 - 1113 - 1198 - 1284 -  
1370 - 1457 - 1544 - 1631 - 1632 - 1720 - 1806 - 1892 - 2237 - 2323 - 2410 - 2411 - 2498 - 2586 -  
2672 - 3188 - 3538 - 4054 -

cluster 12

79 - 163 - 248 - 334 - 420 - 507 - 594 - 681 - 770 - 858 - 944 - 1028 - 1199 - 1285 - 1371 - 1458 - 1545  
- 1633 - 1721 - 1807 - 1893 - 2238 - 2324 - 2673 - 2758 - 3189 - 3276 - 3539 - 4055 -

cluster 13

80 - 164 - 249 - 335 - 421 - 508 - 595 - 682 - 859 - 945 - 1029 - 1114 - 1200 - 1286 - 1372 - 1459 -  
1546 - 1722 - 1808 - 1894 - 1981 - 2067 - 2153 - 2239 - 2325 - 2588 - 2674 - 2845 - 2931 - 3017 -  
3103 - 3366 - 3454 - 3540 - 3797 - 3883 - 3969 - 4232 - 4320 - 5186 -

cluster 14

81 - 165 - 250 - 336 - 422 - 509 - 596 - 683 - 771 - 772 - 860 - 946 - 1030 - 1115 - 1201 - 1287 - 1373 -  
1460 - 1547 - 1634 - 1635 - 1723 - 1809 - 1895 - 1982 - 2068 - 2154 - 2240 - 2326 - 2413 - 2414 -  
2500 - 2501 - 2589 - 2846 - 2932 - 3018 - 3104 - 3191 - 3278 - 3279 - 3367 - 3455 - 3798 - 3884 -  
4233 - 4321 -

cluster 15

82 - 166 - 251 - 337 - 423 - 510 - 597 - 684 - 773 - 861 - 947 - 1031 - 1116 - 1202 - 1288 - 1374 - 1461  
- 1548 - 1636 - 1724 - 1810 - 1896 - 1983 - 2069 - 2155 - 2241 - 2327 - 2415 - 2502 - 2590 - 2847 -  
2933 - 3019 - 3105 - 3280 - 3368 - 3456 - 3799 - 4234 -

cluster 16

83 - 167 - 252 - 338 - 424 - 511 - 598 - 685 - 774 - 862 - 948 - 1032 - 1117 - 1203 - 1289 - 1375 - 1462  
- 1549 - 1637 - 1725 - 1811 - 1897 - 1984 - 2070 - 2156 - 2242 - 2328 - 2503 - 2591 - 2848 - 2934 -  
3281 - 3369 - 3457 - 4147 - 4235 -

cluster 17

84 - 168 - 169 - 253 - 254 - 339 - 425 - 512 - 599 - 600 - 686 - 687 - 775 - 863 - 949 - 1033 - 1034 -  
1118 - 1119 - 1204 - 1290 - 1376 - 1463 - 1464 - 1550 - 1551 - 1638 - 1726 - 1812 - 1898 - 1985 -  
2071 - 2157 - 2243 - 2329 - 2330 - 2416 - 2417 - 2504 - 2592 - 2935 - 3021 - 3107 - 3194 - 3282 -  
3370 - 3458 - 3801 - 4236 - 4324 -

cluster 18

340 - 426 - 513 - 514 - 601 - 688 - 776 - 864 - 1205 - 1291 - 1377 - 1378 - 1465 - 1552 - 1639 - 1727 -  
1986 - 2072 - 2158 - 2244 - 2331 - 2418 - 2505 - 2593 - 2850 - 2936 - 3371 - 3459 - 3802 - 4237 -

## cluster 19

319 - 405 - 492 - 579 - 580 - 666 - 667 - 754 - 755 - 756 - 842 - 843 - 844 - 1184 - 1270 - 1356 - 1443 -  
1444 - 1530 - 1531 - 1617 - 1618 - 1619 - 1705 - 1706 - 1707 - 2137 - 2223 - 2309 - 2397 - 2484 -  
2573 - 3001 - 3087 - 3262 - 3439 - 3781 - 3867 - 3953 - 4041 - 4128 - 4217 - 4305 -

## cluster 20

70 - 154 - 325 - 411 - 498 - 585 - 672 - 849 - 935 - 1019 - 1190 - 1276 - 1362 - 1449 - 1536 - 1712 -  
1798 - 2143 - 2229 - 2315 - 2578 - 3007 - 3787 - 5176 -

## cluster 21

1796 - 1882 - 1969 - 2055 - 2141 - 2314 - 2401 - 2488 - 2662 - 2747 - 2748 - 2832 - 2833 - 2919 -  
3005 - 3091 - 3092 - 3179 - 3266 - 3354 - 3442 - 3528 - 3613 - 3614 - 3698 - 3699 - 3785 - 3871 -  
3957 - 3958 - 4045 - 4132 - 4220 - 4308 - 4393 - 4394 - 4395 - 4478 - 4479 - 4480 - 4564 - 4565 -  
4650 - 4651 - 4652 - 4736 - 4737 - 4738 - 4823 - 4824 - 4910 - 4911 - 4997 - 4998 - 5086 - 5174 -  
5175 - 5259 - 5260 - 5261 - 5344 - 5345 - 5346 - 5430 - 5431 - 5515 - 5516 - 5517 - 5601 - 5602 -  
5603 - 5688 - 5689 - 5775 - 5776 - 5862 - 5863 - 5951 - 6039 - 6040 -

## cluster 22

1813 - 1899 - 2679 - 2764 - 2849 - 3022 - 3108 - 3195 - 3545 - 3630 - 3715 - 3888 - 3974 - 4061 -  
4148 - 4325 - 4411 - 4496 - 4581 - 4582 - 4668 - 4754 - 4840 - 4927 - 5014 - 5102 - 5103 - 5191 -  
5277 - 5362 - 5447 - 5533 - 5619 - 5705 - 5792 - 5879 - 5880 - 5967 - 5968 - 6056 -

## cluster 23

1973 - 2059 - 2145 - 2492 - 2580 - 2752 - 2837 - 2923 - 3009 - 3095 - 3270 - 3358 - 3446 - 3618 -  
3703 - 3789 - 3875 - 3961 - 4136 - 4224 - 4312 - 4398 - 4399 - 4483 - 4484 - 4569 - 4655 - 4741 -  
4827 - 4914 - 4915 - 5001 - 5002 - 5090 - 5178 - 5264 - 5265 - 5349 - 5350 - 5435 - 5520 - 5606 -  
5692 - 5693 - 5779 - 5780 - 5866 - 5867 - 5955 - 6043 -

## cluster 24

1974 - 2060 - 2146 - 2753 - 2838 - 2924 - 3010 - 3096 - 3359 - 3447 - 3619 - 3704 - 3790 - 3876 -  
3962 - 4137 - 4225 - 4313 - 4400 - 4485 - 4570 - 4656 - 4742 - 4828 - 5003 - 5091 - 5179 - 5266 -  
5351 - 5436 - 5521 - 5607 - 5694 - 5781 - 5868 - 5956 - 6044 -

## cluster 25

1976 - 2062 - 2148 - 2755 - 2840 - 2926 - 3012 - 3098 - 3361 - 3449 - 3621 - 3706 - 3792 - 3878 -  
3964 - 4139 - 4227 - 4315 - 4402 - 4487 - 4572 - 4658 - 4744 - 4830 - 5005 - 5093 - 5181 - 5268 -  
5353 - 5438 - 5523 - 5609 - 5695 - 5696 - 5783 - 5870 - 5958 - 6046 -

## cluster 26

1978 - 2064 - 2150 - 2497 - 2585 - 2757 - 2842 - 2928 - 3014 - 3100 - 3275 - 3363 - 3451 - 3623 -  
3708 - 3794 - 3880 - 3966 - 4141 - 4229 - 4317 - 4404 - 4489 - 4574 - 4660 - 4746 - 4832 - 4920 -  
5007 - 5095 - 5183 - 5270 - 5355 - 5440 - 5525 - 5611 - 5697 - 5698 - 5785 - 5872 - 5960 - 6048 -

## cluster 27

1979 - 2065 - 2151 - 2843 - 2929 - 3015 - 3101 - 3364 - 3452 - 3624 - 3709 - 3795 - 3881 - 3967 -  
4142 - 4230 - 4318 - 4405 - 4490 - 4575 - 4661 - 4747 - 4833 - 4921 - 5008 - 5096 - 5184 - 5271 -  
5356 - 5441 - 5526 - 5612 - 5699 - 5786 - 5873 - 5961 - 6049 -

## cluster 28

1980 - 2066 - 2152 - 2412 - 2499 - 2587 - 2759 - 2844 - 2930 - 3016 - 3102 - 3190 - 3277 - 3365 -

3453 - 3625 - 3710 - 3796 - 3882 - 3968 - 4056 - 4143 - 4231 - 4319 - 4406 - 4491 - 4576 - 4662 -  
 4748 - 4749 - 4834 - 4835 - 4922 - 5009 - 5097 - 5185 - 5272 - 5357 - 5442 - 5527 - 5613 - 5614 -  
 5700 - 5787 - 5874 - 5962 - 6050 -

cluster 29

1728 - 1814 - 1900 - 2506 - 2594 - 3109 - 3196 - 3283 - 3284 - 3372 - 3460 - 3716 - 3975 - 4062 -  
 4149 - 4150 - 4238 - 4326 - 4841 - 4928 - 5015 - 5016 - 5104 - 5192 - 5706 - 5793 - 5969 - 6057 -

cluster 30

2396 - 2483 - 2571 - 2572 - 3174 - 3261 - 3349 - 3350 - 3437 - 3438 - 4040 - 4127 - 4215 - 4216 -  
 4303 - 4304 - 4733 - 4819 - 4906 - 4907 - 4993 - 4994 - 5081 - 5082 - 5169 - 5170 - 5171 - 5598 -  
 5684 - 5771 - 5772 - 5858 - 5859 - 5946 - 5947 - 6034 - 6035 - 6036 -

cluster 31

2659 - 2744 - 2830 - 2916 - 3002 - 3003 - 3089 - 3176 - 3264 - 3352 - 3440 - 3525 - 3610 - 3611 -  
 3695 - 3696 - 3782 - 3783 - 3868 - 3869 - 3954 - 3955 - 4042 - 4043 - 4129 - 4130 - 4218 - 4306 -  
 4307 - 4391 - 4392 - 4476 - 4477 - 4561 - 4562 - 4563 - 4647 - 4648 - 4649 - 4734 - 4735 - 4820 -  
 4821 - 4822 - 4908 - 4909 - 4995 - 4996 - 5083 - 5084 - 5085 - 5172 - 5173 - 5257 - 5258 - 5342 -  
 5343 - 5427 - 5428 - 5429 - 5512 - 5513 - 5514 - 5599 - 5600 - 5685 - 5686 - 5687 - 5773 - 5774 -  
 5860 - 5861 - 5948 - 5949 - 5950 - 6037 - 6038 -

cluster 32

2664 - 2749 - 3093 - 3180 - 3267 - 3530 - 3615 - 3700 - 3873 - 3959 - 4046 - 4133 - 4221 - 4396 -  
 4481 - 4566 - 4653 - 4739 - 4825 - 4912 - 4999 - 5087 - 5262 - 5347 - 5432 - 5518 - 5604 - 5690 -  
 5777 - 5864 - 5952 - 6041 -

cluster 33

2665 - 2750 - 3531 - 3616 - 3701 - 3960 - 4047 - 4134 - 4397 - 4482 - 4567 - 4568 - 4654 - 4740 -  
 4826 - 4913 - 5000 - 5088 - 5089 - 5177 - 5263 - 5348 - 5433 - 5434 - 5519 - 5605 - 5691 - 5778 -  
 5865 - 5953 - 5954 - 6042 -

cluster 34

2675 - 2760 - 3541 - 3626 - 3711 - 3970 - 4057 - 4144 - 4407 - 4492 - 4577 - 4663 - 4664 - 4750 -  
 4836 - 4923 - 5010 - 5098 - 5099 - 5187 - 5273 - 5358 - 5443 - 5528 - 5529 - 5615 - 5701 - 5788 -  
 5875 - 5963 - 5964 - 6051 - 6052 -

cluster 35

2676 - 2761 - 3192 - 3542 - 3627 - 3712 - 3885 - 3971 - 4058 - 4145 - 4322 - 4408 - 4493 - 4578 -  
 4665 - 4751 - 4837 - 4924 - 5011 - 5100 - 5188 - 5274 - 5359 - 5444 - 5530 - 5616 - 5702 - 5789 -  
 5876 - 5965 - 6053 -

cluster 36

2677 - 2762 - 3020 - 3106 - 3193 - 3543 - 3628 - 3713 - 3800 - 3886 - 3972 - 4059 - 4146 - 4323 -  
 4409 - 4494 - 4579 - 4666 - 4752 - 4838 - 4925 - 5012 - 5101 - 5189 - 5275 - 5360 - 5445 - 5531 -  
 5617 - 5703 - 5790 - 5877 - 5966 - 6054 -

cluster 37

2678 - 2763 - 3544 - 3629 - 3714 - 3887 - 3973 - 4060 - 4410 - 4495 - 4580 - 4667 - 4753 - 4839 -  
 4926 - 5013 - 5190 - 5276 - 5361 - 5446 - 5532 - 5618 - 5704 - 5791 - 5878 - 6055 -

cluster 38

2754 - 2839 - 2925 - 3272 - 3360 - 3448 - 3535 - 3620 - 3705 - 3791 - 3877 - 4051 - 4138 - 4226 -  
 4314 - 4401 - 4486 - 4571 - 4657 - 4743 - 4917 - 5004 - 5092 - 5180 - 5267 - 5352 - 5437 - 5522 -  
 5608 - 5782 - 5869 - 5957 - 6045 -

cluster 39

2756 - 2841 - 2927 - 3274 - 3362 - 3450 - 3537 - 3622 - 3707 - 3793 - 3879 - 4053 - 4140 - 4228 -  
 4316 - 4403 - 4488 - 4573 - 4659 - 4745 - 4919 - 5006 - 5094 - 5182 - 5269 - 5354 - 5439 - 5524 -  
 5610 - 5784 - 5871 - 5959 - 6047

**Table 15 - Objects of the clusters formed by the execution of Haversine formula on the filtered *muller\_bathymetry* dataset, using 40 as *eps3* parameter.**

**Clusters - 40**

cluster 1

32 - 33 - 116 - 117 - 201 - 202 - 287 - 288 - 373 - 374 - 460 - 461 - 548 - 549 - 635 - 636 - 723 - 724 -  
 812 - 813 - 897 - 898 - 981 - 982 - 1066 - 1067 - 1152 - 1153 - 1238 - 1239 - 1324 - 1325 - 1412 -  
 1413 - 1499 - 1500 - 1586 - 1587 - 1675 - 1676 - 1760 - 1761 - 1846 - 1847 - 1932 - 1933 - 1934 -  
 2018 - 2019 - 2020 - 2104 - 2105 - 2106 - 2191 - 2192 - 2193 - 2277 - 2278 - 2279 - 2365 - 2366 -  
 2452 - 2453 - 2454 - 2540 - 2541 - 2542 - 2627 - 2712 - 2796 - 2797 - 2798 - 2882 - 2884 - 2969 -  
 2970 - 3055 - 3056 - 3057 - 3143 - 3144 - 3230 - 3231 - 3318 - 3319 - 3320 - 3407 - 3408 - 3578 -  
 3750 - 4010 - 4096 - 4274 -

cluster 2

34 - 35 - 118 - 119 - 203 - 204 - 289 - 290 - 375 - 376 - 462 - 463 - 464 - 550 - 551 - 637 - 638 - 639 -  
 725 - 726 - 727 - 814 - 815 - 899 - 900 - 983 - 984 - 1068 - 1069 - 1154 - 1155 - 1240 - 1241 - 1326 -  
 1327 - 1328 - 1414 - 1415 - 1501 - 1502 - 1503 - 1588 - 1589 - 1590 - 1677 - 1678 - 1679 - 1762 -  
 1763 - 1848 - 1849 - 1935 - 1936 - 2021 - 2022 - 2107 - 2108 - 2109 - 2194 - 2195 - 2280 - 2281 -  
 2367 - 2368 - 2369 - 2455 - 2456 - 2543 - 2544 - 2545 - 2629 - 2714 - 2800 - 2886 - 2972 - 2973 -  
 3058 - 3059 - 3145 - 3146 - 3147 - 3232 - 3233 - 3234 - 3321 - 3322 - 3409 - 3410 - 3752 - 3838 -  
 3925 - 4012 - 4013 - 4100 - 4188 -

cluster 3

36 - 37 - 120 - 121 - 122 - 205 - 206 - 207 - 291 - 292 - 377 - 378 - 465 - 466 - 552 - 553 - 554 - 640 -  
 728 - 816 - 901 - 902 - 985 - 986 - 1070 - 1071 - 1072 - 1156 - 1157 - 1242 - 1243 - 1329 - 1330 -  
 1416 - 1417 - 1504 - 1591 - 1764 - 1765 - 1850 - 1851 - 1937 - 1938 - 2023 - 2024 - 2110 - 2111 -  
 2196 - 2282 - 2457 - 2715 - 2802 - 2887 - 2888 - 2974 - 3060 - 3411 - 3583 - 3666 - 3839 - 3926 -  
 4879 -

cluster 4

48 - 49 - 132 - 133 - 217 - 218 - 303 - 304 - 389 - 390 - 476 - 477 - 564 - 565 - 651 - 652 - 739 - 740 -  
 741 - 827 - 828 - 829 - 912 - 913 - 997 - 1082 - 1083 - 1168 - 1254 - 1340 - 1428 - 1515 - 1602 - 1604 -  
 1690 - 1691 - 1776 - 1862 - 1948 - 2381 - 2642 -

cluster 5

50 - 134 - 135 - 219 - 220 - 305 - 306 - 391 - 392 - 478 - 479 - 566 - 567 - 653 - 654 - 742 - 830 - 831 -  
 915 - 999 - 1000 - 1084 - 1085 - 1170 - 1171 - 1256 - 1257 - 1342 - 1343 - 1430 - 1431 - 1517 - 1518 -  
 1605 - 1693 - 1694 - 1778 - 1779 - 1864 - 1951 - 2037 - 2038 - 2123 - 2124 - 2210 - 2211 - 2296 -  
 2297 - 2383 - 2384 - 2471 - 2559 - 2560 - 2644 - 2729 - 2901 - 2987 - 3074 - 3161 - 3249 - 3337 -  
 3338 - 3425 - 3426 - 3510 - 3595 - 3767 - 3853 - 4027 - 4114 - 4203 - 4291 - 4893 - 4980 - 5069 -  
 5157 -

## cluster 6

51 - 52 - 136 - 137 - 221 - 222 - 307 - 308 - 393 - 394 - 480 - 481 - 568 - 655 - 743 - 916 - 917 - 1001 -  
1002 - 1086 - 1087 - 1172 - 1258 - 1259 - 1344 - 1345 - 1432 - 1519 - 1606 - 1780 - 1866 - 1867 -  
1952 - 1953 - 2039 - 2125 - 2126 - 2212 - 2298 - 2385 - 2472 - 2646 - 2647 - 2731 - 2732 - 2817 -  
2903 - 2990 - 3076 - 3163 - 3250 - 3512 - 3598 - 3683 - 3770 - 3856 - 3942 - 4029 - 4722 - 4808 -

## cluster 7

40 - 41 - 42 - 124 - 125 - 126 - 209 - 210 - 211 - 294 - 295 - 296 - 381 - 382 - 469 - 556 - 905 - 906 -  
989 - 990 - 1074 - 1075 - 1159 - 1160 - 1161 - 1767 - 1768 - 1769 - 1854 - 1855 - 1940 - 1941 - 2027 -  
2113 - 2633 - 2634 - 2635 - 2719 - 2720 - 3500 - 3585 - 3586 -

## cluster 8

43 - 44 - 128 - 129 - 213 - 214 - 298 - 299 - 384 - 385 - 472 - 560 - 908 - 909 - 993 - 1078 - 1163 -  
1164 - 1250 - 1336 - 1771 - 1772 - 1857 - 1858 - 1943 - 1944 - 2030 - 2031 - 2637 - 2638 - 2722 -  
2723 - 2808 - 2894 - 3503 - 3504 - 3588 - 4369 - 4370 - 5235 -

## cluster 9

38 - 39 - 123 - 208 - 293 - 379 - 380 - 467 - 730 - 903 - 904 - 987 - 988 - 1073 - 1158 - 1244 - 1245 -  
1331 - 1418 - 1766 - 1852 - 1853 - 1939 - 2025 - 2284 - 2632 - 2717 - 2718 - 2803 - 2889 - 3498 -  
3755 - 4364 - 5230 -

## cluster 10

127 - 212 - 297 - 383 - 470 - 471 - 558 - 645 - 733 - 907 - 991 - 992 - 1076 - 1077 - 1162 - 1248 - 1334  
- 1509 - 1596 - 1684 - 1770 - 1856 - 1942 - 2028 - 2029 - 2115 - 2201 - 2462 - 2636 - 2721 - 2806 -  
2892 - 3065 - 3152 - 3239 - 3501 - 3502 - 3587 - 3672 - 3758 - 4018 - 4367 - 4368 - 4453 - 4538 -  
4624 - 4710 - 5233 - 5489 -

## cluster 11

45 - 46 - 130 - 215 - 216 - 300 - 386 - 387 - 473 - 474 - 561 - 648 - 736 - 910 - 911 - 994 - 995 - 1079 -  
1080 - 1165 - 1251 - 1337 - 1338 - 1773 - 1774 - 1859 - 1860 - 1945 - 1946 - 2032 - 2118 - 2204 -  
2639 - 2724 - 2725 - 2809 - 2810 - 3505 - 3590 - 3675 - 3676 - 4371 - 4456 -

## cluster 12

286 - 372 - 459 - 546 - 547 - 633 - 634 - 721 - 722 - 809 - 810 - 811 - 1151 - 1237 - 1323 - 1410 - 1411  
- 1497 - 1498 - 1584 - 1585 - 1672 - 1673 - 1674 - 2190 - 2276 - 2363 - 2364 - 2450 - 2451 - 2538 -  
2539 - 3228 - 3316 - 3404 - 3405 - 4182 -

## cluster 13

47 - 131 - 301 - 302 - 388 - 475 - 562 - 563 - 649 - 650 - 738 - 826 - 996 - 1081 - 1166 - 1167 - 1252 -  
1253 - 1339 - 1426 - 1427 - 1513 - 1514 - 1601 - 1689 - 1775 - 1861 - 1947 - 2033 - 2119 - 2120 -  
2206 - 2292 - 2379 - 2380 - 2467 - 2555 - 2640 - 2641 - 2726 - 2811 - 2897 - 2983 - 3070 - 3333 -  
3506 - 3507 - 3591 - 3592 - 3677 - 3763 - 3849 - 4372 - 4457 - 4458 - 4543 - 5238 - 5323 - 5409 -

## cluster 14

468 - 555 - 642 - 643 - 731 - 819 - 1246 - 1332 - 1419 - 1506 - 1507 - 1593 - 1594 - 1682 - 2026 -  
2112 - 2198 - 2199 - 2285 - 2372 - 2459 - 2460 - 2547 - 2548 - 2804 - 2890 - 2976 - 3062 - 3063 -  
3149 - 3150 - 3237 - 3325 - 3413 - 3414 - 3499 - 3584 - 3669 - 3670 - 3756 - 3842 - 3928 - 4015 -  
4016 - 4102 - 4103 - 4191 - 4279 - 4280 - 4365 - 4450 - 4535 - 4536 - 4621 - 4622 - 4708 - 4794 -  
4881 - 4882 - 4968 - 4969 - 5056 - 5057 - 5145 - 5231 - 5316 - 5401 - 5486 - 5487 - 5572 - 5573 -  
5659 - 5746 - 5747 - 5833 - 5834 - 5921 - 5922 - 6010 -

## cluster 15

557 - 644 - 732 - 820 - 821 - 1247 - 1333 - 1420 - 1421 - 1508 - 1595 - 1683 - 2114 - 2200 - 2286 -  
2287 - 2373 - 2374 - 2461 - 2549 - 2550 - 2805 - 2891 - 2977 - 2978 - 3064 - 3151 - 3238 - 3326 -  
3327 - 3415 - 3671 - 3757 - 3843 - 3844 - 3929 - 3930 - 4017 - 4104 - 4105 - 4192 - 4193 - 4281 -  
4366 - 4451 - 4452 - 4537 - 4623 - 4709 - 4795 - 4796 - 4883 - 4970 - 4971 - 5058 - 5059 - 5146 -  
5147 - 5232 - 5317 - 5318 - 5402 - 5403 - 5488 - 5574 - 5575 - 5660 - 5661 - 5748 - 5835 - 5923 -  
6011 -

## cluster 16

559 - 646 - 734 - 822 - 823 - 1249 - 1335 - 1422 - 1423 - 1510 - 1597 - 1685 - 2116 - 2202 - 2288 -  
2375 - 2376 - 2463 - 2551 - 2807 - 2893 - 2979 - 2980 - 3066 - 3153 - 3240 - 3328 - 3329 - 3416 -  
3417 - 3673 - 3759 - 3845 - 3846 - 3931 - 3932 - 4019 - 4106 - 4194 - 4195 - 4282 - 4283 - 4454 -  
4539 - 4625 - 4711 - 4797 - 4798 - 4884 - 4885 - 4972 - 5060 - 5061 - 5148 - 5149 - 5234 - 5319 -  
5320 - 5404 - 5405 - 5490 - 5576 - 5662 - 5663 - 5749 - 5750 - 5836 - 5837 - 5925 - 6013 -

## cluster 17

53 - 54 - 138 - 223 - 918 - 919 - 1003 - 1088 - 1173 - 1781 - 1782 - 1868 - 1954 - 2040 - 2648 - 2733 -  
2818 - 2904 - 3513 - 3514 - 3599 - 3684 - 4379 - 4380 - 4464 - 4465 - 4550 - 4636 - 5245 - 5246 -  
5331 - 5416 - 5501 -

## cluster 18

647 - 735 - 824 - 1424 - 1511 - 1598 - 1686 - 1687 - 2117 - 2203 - 2289 - 2290 - 2377 - 2464 - 2552 -  
2895 - 2981 - 3067 - 3154 - 3155 - 3241 - 3242 - 3330 - 3418 - 3589 - 3674 - 3760 - 3761 - 3847 -  
3933 - 4020 - 4107 - 4108 - 4196 - 4284 - 4455 - 4540 - 4626 - 4627 - 4712 - 4713 - 4799 - 4886 -  
4973 - 4974 - 5062 - 5150 - 5236 - 5321 - 5406 - 5491 - 5577 - 5578 - 5664 - 5751 - 5838 - 5839 -  
5926 - 5927 - 6014 - 6015 -

## cluster 19

641 - 729 - 817 - 818 - 1505 - 1592 - 1680 - 1681 - 2197 - 2283 - 2370 - 2371 - 2458 - 2546 - 2630 -  
2631 - 2716 - 2801 - 2975 - 3061 - 3148 - 3235 - 3236 - 3323 - 3324 - 3412 - 3495 - 3496 - 3497 -  
3581 - 3582 - 3667 - 3668 - 3753 - 3754 - 3840 - 3841 - 3927 - 4014 - 4101 - 4189 - 4190 - 4277 -  
4278 - 4361 - 4362 - 4363 - 4447 - 4448 - 4449 - 4532 - 4533 - 4534 - 4618 - 4619 - 4620 - 4705 -  
4706 - 4707 - 4791 - 4792 - 4793 - 4880 - 4966 - 4967 - 5055 - 5143 - 5144 - 5227 - 5228 - 5229 -  
5313 - 5314 - 5315 - 5398 - 5399 - 5400 - 5483 - 5484 - 5485 - 5570 - 5571 - 5656 - 5657 - 5658 -  
5744 - 5745 - 5831 - 5832 - 5920 - 6008 - 6009 -

## cluster 20

914 - 998 - 1169 - 1255 - 1341 - 1429 - 1516 - 1603 - 1692 - 1777 - 1863 - 1949 - 1950 - 2035 - 2036 -  
2122 - 2208 - 2209 - 2294 - 2295 - 2382 - 2469 - 2470 - 2557 - 2558 - 2643 - 2728 - 2813 - 2814 -  
2899 - 2900 - 2986 - 3072 - 3073 - 3159 - 3160 - 3247 - 3248 - 3335 - 3336 - 3423 - 3424 - 3509 -  
3594 - 3679 - 3680 - 3765 - 3766 - 3852 - 3938 - 3939 - 4025 - 4026 - 4113 - 4201 - 4202 - 4289 -  
4290 - 4374 - 4375 - 4376 - 4460 - 4461 - 4545 - 4546 - 4631 - 4632 - 4633 - 4718 - 4719 - 4804 -  
4805 - 4891 - 4892 - 4979 - 5067 - 5068 - 5155 - 5156 - 5240 - 5241 - 5326 - 5327 - 5411 - 5412 -  
5496 - 5497 - 5582 - 5583 - 5584 - 5669 - 5670 - 5756 - 5757 - 5758 - 5843 - 5844 - 5845 - 5932 -  
5933 - 5934 - 6020 - 6021 - 6022 -

## cluster 21

737 - 825 - 1425 - 1512 - 1599 - 1600 - 1688 - 2205 - 2291 - 2378 - 2465 - 2466 - 2553 - 2554 - 2896 -  
2982 - 3068 - 3069 - 3156 - 3243 - 3331 - 3419 - 3420 - 3762 - 3848 - 3934 - 3935 - 4021 - 4022 -  
4109 - 4197 - 4285 - 4541 - 4542 - 4628 - 4714 - 4800 - 4887 - 4888 - 4975 - 5063 - 5151 - 5237 -  
5322 - 5407 - 5408 - 5492 - 5493 - 5579 - 5665 - 5752 - 5840 - 5928 - 6016 -

<p style="text-align: center;">cluster 22</p> <p>2034 - 2121 - 2207 - 2293 - 2468 - 2556 - 2727 - 2812 - 2898 - 2984 - 2985 - 3071 - 3158 - 3245 - 3246 - 3334 - 3422 - 3508 - 3593 - 3678 - 3764 - 3850 - 3851 - 3937 - 4024 - 4111 - 4112 - 4199 - 4200 - 4288 - 4373 - 4459 - 4544 - 4630 - 4716 - 4717 - 4803 - 4890 - 4977 - 4978 - 5065 - 5066 - 5154 - 5239 - 5324 - 5325 - 5410 - 5495 - 5581 - 5667 - 5668 - 5755 - 5842 - 5930 - 5931 - 6019 -</p>
<p style="text-align: center;">cluster 23</p> <p>2626 - 2711 - 2883 - 3492 - 3577 - 3662 - 3663 - 3749 - 3835 - 3836 - 3922 - 4009 - 4097 - 4185 - 4273 - 4358 - 4443 - 4444 - 4528 - 4529 - 4615 - 4616 - 4701 - 4702 - 4787 - 4788 - 4875 - 4876 - 4962 - 4963 - 5051 - 5139 - 5140 - 5224 - 5309 - 5310 - 5394 - 5395 - 5480 - 5481 - 5566 - 5567 - 5652 - 5653 - 5740 - 5741 - 5827 - 5828 - 5916 - 6004 - 6005 -</p>
<p style="text-align: center;">cluster 24</p> <p>2628 - 2713 - 2799 - 2885 - 2971 - 3493 - 3494 - 3579 - 3580 - 3664 - 3665 - 3751 - 3837 - 3923 - 3924 - 4011 - 4098 - 4099 - 4186 - 4187 - 4275 - 4276 - 4359 - 4360 - 4445 - 4446 - 4530 - 4531 - 4617 - 4703 - 4704 - 4789 - 4790 - 4877 - 4878 - 4964 - 4965 - 5052 - 5053 - 5054 - 5141 - 5142 - 5225 - 5226 - 5311 - 5312 - 5396 - 5397 - 5482 - 5568 - 5569 - 5654 - 5655 - 5742 - 5743 - 5829 - 5830 - 5917 - 5918 - 5919 - 6006 - 6007 -</p>
<p style="text-align: center;">cluster 25</p> <p>1865 - 2645 - 2730 - 2815 - 2816 - 2902 - 2988 - 2989 - 3075 - 3162 - 3511 - 3596 - 3597 - 3681 - 3682 - 3768 - 3769 - 3854 - 3855 - 3940 - 3941 - 4028 - 4115 - 4116 - 4204 - 4292 - 4377 - 4378 - 4462 - 4463 - 4547 - 4548 - 4549 - 4634 - 4635 - 4720 - 4721 - 4806 - 4807 - 4894 - 4895 - 4981 - 4982 - 5070 - 5158 - 5242 - 5243 - 5244 - 5328 - 5329 - 5413 - 5414 - 5415 - 5498 - 5499 - 5500 - 5585 - 5586 - 5587 - 5671 - 5672 - 5673 - 5759 - 5760 - 5846 - 5847 - 5935 - 6023 -</p>
<p style="text-align: center;">cluster 26</p> <p>2968 - 3054 - 3141 - 3142 - 3229 - 3317 - 3406 - 3748 - 3834 - 3920 - 3921 - 4007 - 4008 - 4094 - 4095 - 4183 - 4184 - 4270 - 4271 - 4272 - 4614 - 4700 - 4786 - 4873 - 4874 - 4960 - 4961 - 5048 - 5049 - 5050 - 5136 - 5137 - 5138 - 5479 - 5565 - 5651 - 5738 - 5739 - 5825 - 5826 - 5913 - 5914 - 5915 - 6001 - 6002 - 6003 -</p>
<p style="text-align: center;">cluster 27</p> <p>3157 - 3244 - 3332 - 3421 - 3936 - 4023 - 4110 - 4198 - 4286 - 4287 - 4629 - 4715 - 4801 - 4802 - 4889 - 4976 - 5064 - 5152 - 5153 - 5494 - 5580 - 5666 - 5753 - 5754 - 5841 - 5929 - 6017 - 6018</p>

**Table 16 - Objects of the clusters formed by the execution of Haversine formula on the filtered *muller\_bathymetry* dataset, using 60 as *eps3* parameter**

<b>Clusters - 60</b>
<p style="text-align: center;">cluster 1</p> <p>55 - 56 - 139 - 140 - 224 - 225 - 310 - 311 - 396 - 397 - 483 - 484 - 571 - 572 - 658 - 659 - 746 - 747 - 748 - 834 - 835 - 836 - 920 - 921 - 1004 - 1005 - 1089 - 1090 - 1175 - 1176 - 1261 - 1262 - 1347 - 1348 - 1435 - 1436 - 1522 - 1523 - 1609 - 1610 - 1611 - 1698 - 1699 - 1783 - 1784 - 1869 - 1870 - 1955 - 1956 - 1957 - 2041 - 2042 - 2043 - 2128 - 2129 - 2214 - 2215 - 2301 - 2302 - 2388 - 2389 - 2475 - 2476 - 2477 - 2564 - 2565 - 2650 - 2735 - 2819 - 2820 - 2821 - 2905 - 2906 - 2907 - 2992 - 2993 - 3078 - 3079 - 3080 - 3165 - 3166 - 3167 - 3253 - 3254 - 3255 - 3341 - 3342 - 3343 - 3430 - 3431 - 3516 - 3601 - 3773 - 3859 - 3944 - 3945 - 3946 - 4031 - 4032 - 4033 - 4119 - 4120 - 4121 - 4207 - 4208 - 4209 - 4295 - 4296 - 4297 - 4639 - 4725 - 4899 - 4986 - 5073 - 5074 - 5075 - 5161 - 5162 - 5163 - 5764 - 5851 - 5940 - 6028 -</p>



## cluster 2

57 - 141 - 142 - 226 - 227 - 312 - 313 - 398 - 399 - 485 - 486 - 573 - 574 - 660 - 661 - 749 - 750 - 837 -  
838 - 922 - 1006 - 1007 - 1091 - 1092 - 1177 - 1178 - 1263 - 1264 - 1349 - 1350 - 1437 - 1438 - 1524 -  
1525 - 1612 - 1613 - 1700 - 1701 - 1785 - 1786 - 1871 - 1872 - 1958 - 1959 - 2044 - 2045 - 2130 -  
2131 - 2216 - 2217 - 2218 - 2303 - 2304 - 2390 - 2391 - 2392 - 2478 - 2479 - 2566 - 2567 - 2652 -  
2737 - 2822 - 2823 - 2908 - 2909 - 2994 - 2995 - 3081 - 3082 - 3168 - 3169 - 3256 - 3257 - 3344 -  
3345 - 3432 - 3433 - 3603 - 3775 - 4035 - 4211 -

## cluster 3

58 - 59 - 143 - 144 - 228 - 229 - 314 - 315 - 400 - 401 - 487 - 488 - 575 - 576 - 662 - 663 - 751 - 752 -  
839 - 840 - 923 - 924 - 1008 - 1009 - 1093 - 1094 - 1179 - 1180 - 1265 - 1266 - 1351 - 1352 - 1439 -  
1440 - 1526 - 1527 - 1614 - 1615 - 1702 - 1703 - 1787 - 1788 - 1873 - 1874 - 1960 - 1961 - 2046 -  
2047 - 2132 - 2133 - 2219 - 2220 - 2305 - 2306 - 2393 - 2394 - 2480 - 2481 - 2568 - 2569 - 2654 -  
2739 - 2825 - 2911 - 3084 - 3171 - 3259 - 3347 -

## cluster 4

60 - 61 - 145 - 146 - 230 - 231 - 316 - 317 - 402 - 403 - 489 - 490 - 577 - 578 - 664 - 665 - 753 - 841 -  
925 - 926 - 1010 - 1011 - 1095 - 1096 - 1181 - 1182 - 1267 - 1268 - 1353 - 1354 - 1441 - 1442 - 1528 -  
1529 - 1616 - 1704 - 1789 - 1790 - 1875 - 1876 - 1962 - 1963 - 2048 - 2049 - 2134 - 2135 - 2221 -  
2222 - 2307 - 2308 - 2395 - 2482 - 2570 - 2656 - 2741 - 2827 - 2913 - 3086 - 3173 -

## cluster 5

62 - 63 - 64 - 147 - 148 - 232 - 233 - 318 - 404 - 491 - 927 - 928 - 929 - 1012 - 1013 - 1097 - 1098 -  
1183 - 1269 - 1355 - 1791 - 1792 - 1877 - 1878 - 1964 - 2050 - 2136 - 2658 - 2743 -

## cluster 6

309 - 395 - 482 - 569 - 570 - 656 - 657 - 744 - 745 - 832 - 833 - 1174 - 1260 - 1346 - 1433 - 1434 -  
1520 - 1521 - 1607 - 1608 - 1695 - 1696 - 1697 - 2127 - 2213 - 2299 - 2300 - 2387 - 2474 - 2562 -  
2563 - 2991 - 3077 - 3252 - 3340 - 3428 - 3429 - 3771 - 3857 - 3943 - 4118 - 4206 - 4897 - 4984 -  
5072 - 6026 -

## cluster 7

2386 - 2473 - 2561 - 3164 - 3251 - 3339 - 3427 - 4030 - 4117 - 4205 - 4293 - 4294 - 4723 - 4809 -  
4896 - 4983 - 5071 - 5159 - 5160 - 5588 - 5674 - 5761 - 5848 - 5849 - 5936 - 5937 - 6024 - 6025 -

## cluster 8

2649 - 2734 - 3515 - 3600 - 3685 - 3686 - 3772 - 3858 - 4381 - 4382 - 4466 - 4467 - 4551 - 4552 -  
4637 - 4638 - 4724 - 4810 - 4811 - 4898 - 4985 - 5247 - 5248 - 5332 - 5333 - 5417 - 5418 - 5502 -  
5503 - 5504 - 5589 - 5590 - 5675 - 5676 - 5762 - 5763 - 5850 - 5938 - 5939 - 6027 -

## cluster 9

2651 - 2736 - 3517 - 3602 - 3687 - 3688 - 3774 - 3860 - 3861 - 3947 - 4034 - 4122 - 4210 - 4298 -  
4299 - 4383 - 4468 - 4469 - 4553 - 4554 - 4640 - 4641 - 4726 - 4727 - 4812 - 4813 - 4900 - 4901 -  
4987 - 4988 - 5076 - 5077 - 5164 - 5165 - 5249 - 5334 - 5335 - 5419 - 5420 - 5505 - 5506 - 5591 -  
5592 - 5677 - 5678 - 5765 - 5766 - 5852 - 5853 - 5941 - 5942 - 6029 - 6030 -

## cluster 10

2653 - 2738 - 2824 - 2910 - 2996 - 2997 - 3083 - 3170 - 3258 - 3346 - 3434 - 3435 - 3518 - 3519 -  
3604 - 3605 - 3689 - 3690 - 3776 - 3777 - 3862 - 3863 - 3948 - 3949 - 4036 - 4037 - 4123 - 4124 -  
4212 - 4213 - 4300 - 4301 - 4384 - 4385 - 4470 - 4471 - 4555 - 4556 - 4642 - 4643 - 4728 - 4729 -

4814 - 4815 - 4902 - 4903 - 4989 - 4990 - 5078 - 5079 - 5166 - 5167 - 5250 - 5251 - 5336 - 5337 -  
 5421 - 5422 - 5507 - 5508 - 5593 - 5594 - 5679 - 5680 - 5767 - 5768 - 5854 - 5855 - 5943 - 5944 -  
 6031 - 6032 -

cluster 11

2655 - 2740 - 2826 - 2912 - 2998 - 2999 - 3085 - 3172 - 3260 - 3348 - 3436 - 3520 - 3521 - 3606 -  
 3607 - 3691 - 3692 - 3778 - 3779 - 3864 - 3865 - 3950 - 3951 - 4038 - 4039 - 4125 - 4126 - 4214 -  
 4302 - 4386 - 4387 - 4472 - 4473 - 4557 - 4558 - 4644 - 4645 - 4730 - 4731 - 4816 - 4817 - 4904 -  
 4905 - 4991 - 4992 - 5080 - 5168 - 5252 - 5253 - 5338 - 5339 - 5423 - 5424 - 5509 - 5510 - 5595 -  
 5596 - 5681 - 5682 - 5769 - 5770 - 5856 - 5857 - 5945 - 6033 -

cluster 12

2657 - 2742 - 2828 - 2914 - 3000 - 3522 - 3523 - 3524 - 3608 - 3609 - 3693 - 3694 - 3780 - 3866 -  
 3952 - 4388 - 4389 - 4390 - 4474 - 4475 - 4559 - 4560 - 4646 - 4732 - 4818 - 5254 - 5255 - 5256 -  
 5340 - 5341 - 5425 - 5426 - 5511 - 5597 - 5683

**Table 17 - Objects of the clusters formed by the execution of Haversine formula on the filtered *muller\_bathymetry* dataset, using 70 as *eps3* parameter.**

**Clusters - 70**

cluster 1

65 - 149 - 150 - 234 - 235 - 320 - 321 - 406 - 407 - 493 - 494 - 581 - 582 - 668 - 669 - 757 - 845 - 846 -  
 930 - 1014 - 1015 - 1099 - 1100 - 1185 - 1186 - 1271 - 1272 - 1357 - 1358 - 1445 - 1446 - 1532 - 1533  
 - 1620 - 1708 - 1709 - 1793 - 1794 - 1879 - 1880 - 1965 - 1966 - 1967 - 2051 - 2052 - 2053 - 2138 -  
 2139 - 2224 - 2225 - 2226 - 2310 - 2311 - 2312 - 2398 - 2399 - 2485 - 2486 - 2487 - 2574 - 2575 -  
 2660 - 2745 - 2829 - 2831 - 2915 - 2917 - 3088 - 3090 - 3175 - 3177 - 3263 - 3351 - 3353 - 3441 -

cluster 2

66 - 67 - 68 - 151 - 152 - 236 - 237 - 322 - 323 - 408 - 409 - 495 - 496 - 583 - 670 - 671 - 758 - 759 -  
 847 - 931 - 932 - 933 - 1016 - 1017 - 1101 - 1102 - 1187 - 1188 - 1273 - 1274 - 1359 - 1360 - 1447 -  
 1534 - 1535 - 1621 - 1622 - 1710 - 1795 - 1881 - 1968 - 2054 - 2140 - 2227 - 2313 - 2400 - 2576 -  
 2661 - 2746 - 2918 - 3004 - 3178 - 3265 - 3526 - 3527 - 3612 - 3697 - 3784 - 3870 - 3956 - 4044 -  
 4131 - 4219 -

cluster 3

69 - 153 - 238 - 239 - 324 - 410 - 497 - 584 - 760 - 848 - 934 - 1018 - 1103 - 1104 - 1189 - 1275 - 1361  
 - 1448 - 1623 - 1711 - 1797 - 1883 - 1884 - 1970 - 2056 - 2142 - 2228 - 2402 - 2489 - 2577 - 2663 -  
 2834 - 2920 - 3006 - 3355 - 3443 - 3529 - 3786 - 3872 - 4309 -

cluster 4

71 - 155 - 240 - 326 - 412 - 499 - 586 - 673 - 761 - 850 - 936 - 1020 - 1105 - 1191 - 1277 - 1363 - 1450  
 - 1537 - 1624 - 1713 - 1799 - 1885 - 1971 - 1972 - 2057 - 2058 - 2144 - 2230 - 2316 - 2403 - 2490 -  
 2491 - 2579 - 2835 - 2836 - 2921 - 2922 - 3008 - 3094 - 3181 - 3268 - 3356 - 3357 - 3444 - 3445 -  
 3788 - 3874 - 4222 - 4310 - 4311 -

cluster 5

72 - 156 - 241 - 327 - 413 - 500 - 587 - 674 - 762 - 851 - 937 - 1021 - 1106 - 1192 - 1278 - 1364 - 1451  
 - 1538 - 1625 - 1714 - 1800 - 1886 - 2231 - 2317 - 2404 - 2666 - 2751 - 3182 - 3269 - 3532 - 3617 -  
 3702 - 4048 - 4135 - 4223 -

cluster 6

73 - 157 - 242 - 328 - 414 - 501 - 588 - 675 - 763 - 852 - 938 - 1022 - 1107 - 1193 - 1279 - 1365 - 1452  
- 1539 - 1626 - 1715 - 1801 - 1887 - 2232 - 2318 - 2405 - 2581 - 2667 - 3183 - 3533 - 4049 -

cluster 7

74 - 158 - 243 - 329 - 415 - 502 - 589 - 676 - 764 - 853 - 939 - 1023 - 1108 - 1194 - 1280 - 1366 - 1453  
- 1540 - 1627 - 1716 - 1802 - 1888 - 1975 - 2061 - 2147 - 2233 - 2319 - 2406 - 2493 - 2494 - 2582 -  
2668 - 3011 - 3097 - 3184 - 3271 - 3534 - 3963 - 4050 - 4829 - 4916 -

cluster 8

75 - 159 - 244 - 330 - 416 - 503 - 590 - 677 - 765 - 854 - 940 - 1024 - 1109 - 1195 - 1281 - 1367 - 1454  
- 1541 - 1628 - 1717 - 1803 - 1889 - 2234 - 2320 - 2407 - 2583 - 2669 - 3185 -

cluster 9

76 - 160 - 245 - 331 - 417 - 504 - 591 - 678 - 766 - 855 - 941 - 1025 - 1110 - 1196 - 1282 - 1368 - 1455  
- 1542 - 1629 - 1718 - 1804 - 1890 - 1977 - 2063 - 2149 - 2235 - 2321 - 2408 - 2495 - 2496 - 2584 -  
2670 - 3013 - 3099 - 3186 - 3273 - 3536 - 3965 - 4052 - 4831 - 4918 -

cluster 10

77 - 161 - 246 - 332 - 418 - 505 - 592 - 679 - 767 - 856 - 942 - 1026 - 1111 - 1197 - 1283 - 1369 - 1456  
- 1543 - 1630 - 1719 - 1805 - 1891 - 2236 - 2322 - 2409 - 2671 - 3187 -

cluster 11

78 - 162 - 247 - 333 - 419 - 506 - 593 - 680 - 768 - 769 - 857 - 943 - 1027 - 1112 - 1113 - 1198 - 1284 -  
1370 - 1457 - 1544 - 1631 - 1632 - 1720 - 1806 - 1892 - 2237 - 2323 - 2410 - 2411 - 2498 - 2586 -  
2672 - 3188 - 3538 - 4054 -

cluster 12

79 - 163 - 248 - 334 - 420 - 507 - 594 - 681 - 770 - 858 - 944 - 1028 - 1199 - 1285 - 1371 - 1458 - 1545  
- 1633 - 1721 - 1807 - 1893 - 2238 - 2324 - 2673 - 2758 - 3189 - 3276 - 3539 - 4055 -

cluster 13

80 - 164 - 249 - 335 - 421 - 508 - 595 - 682 - 859 - 945 - 1029 - 1114 - 1200 - 1286 - 1372 - 1459 -  
1546 - 1722 - 1808 - 1894 - 1981 - 2067 - 2153 - 2239 - 2325 - 2588 - 2674 - 2845 - 2931 - 3017 -  
3103 - 3366 - 3454 - 3540 - 3797 - 3883 - 3969 - 4232 - 4320 - 5186 -

cluster 14

81 - 165 - 250 - 336 - 422 - 509 - 596 - 683 - 771 - 772 - 860 - 946 - 1030 - 1115 - 1201 - 1287 - 1373 -  
1460 - 1547 - 1634 - 1635 - 1723 - 1809 - 1895 - 1982 - 2068 - 2154 - 2240 - 2326 - 2413 - 2414 -  
2500 - 2501 - 2589 - 2846 - 2932 - 3018 - 3104 - 3191 - 3278 - 3279 - 3367 - 3455 - 3798 - 3884 -  
4233 - 4321 -

cluster 15

82 - 166 - 251 - 337 - 423 - 510 - 597 - 684 - 773 - 861 - 947 - 1031 - 1116 - 1202 - 1288 - 1374 - 1461  
- 1548 - 1636 - 1724 - 1810 - 1896 - 1983 - 2069 - 2155 - 2241 - 2327 - 2415 - 2502 - 2590 - 2847 -  
2933 - 3019 - 3105 - 3280 - 3368 - 3456 - 3799 - 4234 -

cluster 16

83 - 167 - 252 - 338 - 424 - 511 - 598 - 685 - 774 - 862 - 948 - 1032 - 1117 - 1203 - 1289 - 1375 - 1462  
- 1549 - 1637 - 1725 - 1811 - 1897 - 1984 - 2070 - 2156 - 2242 - 2328 - 2503 - 2591 - 2848 - 2934 -  
3281 - 3369 - 3457 - 4147 - 4235 -

## cluster 17

84 - 168 - 169 - 253 - 254 - 339 - 425 - 512 - 599 - 600 - 686 - 687 - 775 - 863 - 949 - 1033 - 1034 -  
1118 - 1119 - 1204 - 1290 - 1376 - 1463 - 1464 - 1550 - 1551 - 1638 - 1726 - 1812 - 1898 - 1985 -  
2071 - 2157 - 2243 - 2329 - 2330 - 2416 - 2417 - 2504 - 2592 - 2935 - 3021 - 3107 - 3194 - 3282 -  
3370 - 3458 - 3801 - 4236 - 4324 -

## cluster 18

340 - 426 - 513 - 514 - 601 - 688 - 776 - 864 - 1205 - 1291 - 1377 - 1378 - 1465 - 1552 - 1639 - 1727 -  
1986 - 2072 - 2158 - 2244 - 2331 - 2418 - 2505 - 2593 - 2850 - 2936 - 3371 - 3459 - 3802 - 4237 -

## cluster 19

319 - 405 - 492 - 579 - 580 - 666 - 667 - 754 - 755 - 756 - 842 - 843 - 844 - 1184 - 1270 - 1356 - 1443 -  
1444 - 1530 - 1531 - 1617 - 1618 - 1619 - 1705 - 1706 - 1707 - 2137 - 2223 - 2309 - 2397 - 2484 -  
2573 - 3001 - 3087 - 3262 - 3439 - 3781 - 3867 - 3953 - 4041 - 4128 - 4217 - 4305 -

## cluster 20

70 - 154 - 325 - 411 - 498 - 585 - 672 - 849 - 935 - 1019 - 1190 - 1276 - 1362 - 1449 - 1536 - 1712 -  
1798 - 2143 - 2229 - 2315 - 2578 - 3007 - 3787 - 5176 -

## cluster 21

1796 - 1882 - 1969 - 2055 - 2141 - 2314 - 2401 - 2488 - 2662 - 2747 - 2748 - 2832 - 2833 - 2919 -  
3005 - 3091 - 3092 - 3179 - 3266 - 3354 - 3442 - 3528 - 3613 - 3614 - 3698 - 3699 - 3785 - 3871 -  
3957 - 3958 - 4045 - 4132 - 4220 - 4308 - 4393 - 4394 - 4395 - 4478 - 4479 - 4480 - 4564 - 4565 -  
4650 - 4651 - 4652 - 4736 - 4737 - 4738 - 4823 - 4824 - 4910 - 4911 - 4997 - 4998 - 5086 - 5174 -  
5175 - 5259 - 5260 - 5261 - 5344 - 5345 - 5346 - 5430 - 5431 - 5515 - 5516 - 5517 - 5601 - 5602 -  
5603 - 5688 - 5689 - 5775 - 5776 - 5862 - 5863 - 5951 - 6039 - 6040 -

## cluster 22

1813 - 1899 - 2679 - 2764 - 2849 - 3022 - 3108 - 3195 - 3545 - 3630 - 3715 - 3888 - 3974 - 4061 -  
4148 - 4325 - 4411 - 4496 - 4581 - 4582 - 4668 - 4754 - 4840 - 4927 - 5014 - 5102 - 5103 - 5191 -  
5277 - 5362 - 5447 - 5533 - 5619 - 5705 - 5792 - 5879 - 5880 - 5967 - 5968 - 6056 -

## cluster 23

1973 - 2059 - 2145 - 2492 - 2580 - 2752 - 2837 - 2923 - 3009 - 3095 - 3270 - 3358 - 3446 - 3618 -  
3703 - 3789 - 3875 - 3961 - 4136 - 4224 - 4312 - 4398 - 4399 - 4483 - 4484 - 4569 - 4655 - 4741 -  
4827 - 4914 - 4915 - 5001 - 5002 - 5090 - 5178 - 5264 - 5265 - 5349 - 5350 - 5435 - 5520 - 5606 -  
5692 - 5693 - 5779 - 5780 - 5866 - 5867 - 5955 - 6043 -

## cluster 24

1974 - 2060 - 2146 - 2753 - 2838 - 2924 - 3010 - 3096 - 3359 - 3447 - 3619 - 3704 - 3790 - 3876 -  
3962 - 4137 - 4225 - 4313 - 4400 - 4485 - 4570 - 4656 - 4742 - 4828 - 5003 - 5091 - 5179 - 5266 -  
5351 - 5436 - 5521 - 5607 - 5694 - 5781 - 5868 - 5956 - 6044 -

## cluster 25

1976 - 2062 - 2148 - 2755 - 2840 - 2926 - 3012 - 3098 - 3361 - 3449 - 3621 - 3706 - 3792 - 3878 -  
3964 - 4139 - 4227 - 4315 - 4402 - 4487 - 4572 - 4658 - 4744 - 4830 - 5005 - 5093 - 5181 - 5268 -  
5353 - 5438 - 5523 - 5609 - 5695 - 5696 - 5783 - 5870 - 5958 - 6046 -

## cluster 26

1978 - 2064 - 2150 - 2497 - 2585 - 2757 - 2842 - 2928 - 3014 - 3100 - 3275 - 3363 - 3451 - 3623 -  
3708 - 3794 - 3880 - 3966 - 4141 - 4229 - 4317 - 4404 - 4489 - 4574 - 4660 - 4746 - 4832 - 4920 -

5007 - 5095 - 5183 - 5270 - 5355 - 5440 - 5525 - 5611 - 5697 - 5698 - 5785 - 5872 - 5960 - 6048 -

cluster 27

1979 - 2065 - 2151 - 2843 - 2929 - 3015 - 3101 - 3364 - 3452 - 3624 - 3709 - 3795 - 3881 - 3967 -  
4142 - 4230 - 4318 - 4405 - 4490 - 4575 - 4661 - 4747 - 4833 - 4921 - 5008 - 5096 - 5184 - 5271 -  
5356 - 5441 - 5526 - 5612 - 5699 - 5786 - 5873 - 5961 - 6049 -

cluster 28

1980 - 2066 - 2152 - 2412 - 2499 - 2587 - 2759 - 2844 - 2930 - 3016 - 3102 - 3190 - 3277 - 3365 -  
3453 - 3625 - 3710 - 3796 - 3882 - 3968 - 4056 - 4143 - 4231 - 4319 - 4406 - 4491 - 4576 - 4662 -  
4748 - 4749 - 4834 - 4835 - 4922 - 5009 - 5097 - 5185 - 5272 - 5357 - 5442 - 5527 - 5613 - 5614 -  
5700 - 5787 - 5874 - 5962 - 6050 -

cluster 29

1728 - 1814 - 1900 - 2506 - 2594 - 3109 - 3196 - 3283 - 3284 - 3372 - 3460 - 3716 - 3975 - 4062 -  
4149 - 4150 - 4238 - 4326 - 4841 - 4928 - 5015 - 5016 - 5104 - 5192 - 5706 - 5793 - 5969 - 6057 -

cluster 30

2396 - 2483 - 2571 - 2572 - 3174 - 3261 - 3349 - 3350 - 3437 - 3438 - 4040 - 4127 - 4215 - 4216 -  
4303 - 4304 - 4733 - 4819 - 4906 - 4907 - 4993 - 4994 - 5081 - 5082 - 5169 - 5170 - 5171 - 5598 -  
5684 - 5771 - 5772 - 5858 - 5859 - 5946 - 5947 - 6034 - 6035 - 6036 -

cluster 31

2659 - 2744 - 2830 - 2916 - 3002 - 3003 - 3089 - 3176 - 3264 - 3352 - 3440 - 3525 - 3610 - 3611 -  
3695 - 3696 - 3782 - 3783 - 3868 - 3869 - 3954 - 3955 - 4042 - 4043 - 4129 - 4130 - 4218 - 4306 -  
4307 - 4391 - 4392 - 4476 - 4477 - 4561 - 4562 - 4563 - 4647 - 4648 - 4649 - 4734 - 4735 - 4820 -  
4821 - 4822 - 4908 - 4909 - 4995 - 4996 - 5083 - 5084 - 5085 - 5172 - 5173 - 5257 - 5258 - 5342 -  
5343 - 5427 - 5428 - 5429 - 5512 - 5513 - 5514 - 5599 - 5600 - 5685 - 5686 - 5687 - 5773 - 5774 -  
5860 - 5861 - 5948 - 5949 - 5950 - 6037 - 6038 -

cluster 32

2664 - 2749 - 3093 - 3180 - 3267 - 3530 - 3615 - 3700 - 3873 - 3959 - 4046 - 4133 - 4221 - 4396 -  
4481 - 4566 - 4653 - 4739 - 4825 - 4912 - 4999 - 5087 - 5262 - 5347 - 5432 - 5518 - 5604 - 5690 -  
5777 - 5864 - 5952 - 6041 -

cluster 33

2665 - 2750 - 3531 - 3616 - 3701 - 3960 - 4047 - 4134 - 4397 - 4482 - 4567 - 4568 - 4654 - 4740 -  
4826 - 4913 - 5000 - 5088 - 5089 - 5177 - 5263 - 5348 - 5433 - 5434 - 5519 - 5605 - 5691 - 5778 -  
5865 - 5953 - 5954 - 6042 -

cluster 34

2675 - 2760 - 3541 - 3626 - 3711 - 3970 - 4057 - 4144 - 4407 - 4492 - 4577 - 4663 - 4664 - 4750 -  
4836 - 4923 - 5010 - 5098 - 5099 - 5187 - 5273 - 5358 - 5443 - 5528 - 5529 - 5615 - 5701 - 5788 -  
5875 - 5963 - 5964 - 6051 - 6052 -

cluster 35

2676 - 2761 - 3192 - 3542 - 3627 - 3712 - 3885 - 3971 - 4058 - 4145 - 4322 - 4408 - 4493 - 4578 -  
4665 - 4751 - 4837 - 4924 - 5011 - 5100 - 5188 - 5274 - 5359 - 5444 - 5530 - 5616 - 5702 - 5789 -  
5876 - 5965 - 6053 -

cluster 36

2677 - 2762 - 3020 - 3106 - 3193 - 3543 - 3628 - 3713 - 3800 - 3886 - 3972 - 4059 - 4146 - 4323 -  
4409 - 4494 - 4579 - 4666 - 4752 - 4838 - 4925 - 5012 - 5101 - 5189 - 5275 - 5360 - 5445 - 5531 -  
5617 - 5703 - 5790 - 5877 - 5966 - 6054 -

cluster 37

2678 - 2763 - 3544 - 3629 - 3714 - 3887 - 3973 - 4060 - 4410 - 4495 - 4580 - 4667 - 4753 - 4839 -  
4926 - 5013 - 5190 - 5276 - 5361 - 5446 - 5532 - 5618 - 5704 - 5791 - 5878 - 6055 -

cluster 38

2754 - 2839 - 2925 - 3272 - 3360 - 3448 - 3535 - 3620 - 3705 - 3791 - 3877 - 4051 - 4138 - 4226 -  
4314 - 4401 - 4486 - 4571 - 4657 - 4743 - 4917 - 5004 - 5092 - 5180 - 5267 - 5352 - 5437 - 5522 -  
5608 - 5782 - 5869 - 5957 - 6045 -

cluster 39

2756 - 2841 - 2927 - 3274 - 3362 - 3450 - 3537 - 3622 - 3707 - 3793 - 3879 - 4053 - 4140 - 4228 -  
4316 - 4403 - 4488 - 4573 - 4659 - 4745 - 4919 - 5006 - 5094 - 5182 - 5269 - 5354 - 5439 - 5524 -  
5610 - 5784 - 5871 - 5959 - 6047

